Casimir invariants, characteristic identities, and Young diagrams for color algebras and superalgebras

H. S. Green
Department of Mathematical Physics, University of Adelaide, Box 498, G.P.O., Adelaide, South Australia 5001

P. D. Jarvis
Department of Physics, University of Tasmania, Box 252C, G.P.O., Hobart, Tasmania 7001, Australia

(Received 17 March 1982; accepted for publication 5 November 1982)

The generalized commutation relations satisfied by generators of the general linear, special linear, and orthosymplectic color (super) algebras are presented in matrix form. Tensor operators, including Casimir invariants, are constructed in the enveloping algebra. For the general, special linear and orthosymplectic cases, eigenvalues of the quadratic and higher Casimir invariants are given in terms of the highest-weight vector. Correspondingly, characteristic polynomial identities, satisfied by the matrix of generators, are obtained in factorized form. Classes of finite-dimensional representations are identified using Young diagram techniques, and dimension, branching, and product rules for these are given. Finally, the connection between color (super) algebras and generalized particle statistics is elucidated.

PACS numbers: 02.10.Sp, 02.20.Hj

I. INTRODUCTION AND MAIN RESULTS

To abstract Lie algebras and superalgebras[1,2] there correspond concrete representations in terms of matrices satisfying certain specified commutation and anticommutation relations. A natural generalization of such structures involves matrix brackets wherein the minus (or plus) sign of the (anti) commutator is replaced by a complex phase. These so-called color algebras and superalgebras were introduced by Lukierski, Rittenberg, and Wyler,[3–5] and their formal properties studied by Scheunert.[6] The name derives from the analogy with the concept of color symmetry in particle physics, including its parastatistics[7] realization.[8] In fact, algebras of this type have appeared in connection with earlier work on parastatistics,[9] and more particularly with the generalized (modular) statistics of Green.[10,11]

The theory of Lie superalgebras has shown that their classification theory and formal aspects,[12] representation theory,[13] and other properties can be developed similarly to ordinary Lie algebras. Diagram techniques are available[14,15] for classes of finite-dimensional representations, and the usual product, dimension, and branching rules apply.[14] One of the aims of the present paper is to consider, for classes of color algebras and superalgebras, the Casimir invariants and characteristic identities for generators, following earlier work on Lie superalgebras[16,17] and Lie algebras.[18,19] This problem has already been considered in a more abstract context by Agrawala,[20] but we shall present some explicit results adapted to the physical applications, in particular to quantum chromodynamics.

A color algebra or superalgebra is defined as follows.[3] Given are a vector space L and a grading ( ) which associates (homogeneous) elements $X, Y, Z,$ of L with elements (X), (Y), (Z) of an abelian group Γ. Further, there is a scalar commutation factor $u_{(\ ),(\ )}$ with the properties

$$u_{(X),(Y)}u_{(Y),(X)} = 1,$$

$$u_{(X),(Y)+(Z)} = u_{(X),(Y)}u_{(X),(Z)},\qquad (1)$$

$$u_{(X)+(Y),(Z)} = u_{(X),(Z)}u_{(Y),(Z)}.$$

L becomes a color algebra or superalgebra when endowed with a product [,] satisfying

$$([X,Y]) = (X) + (Y),$$

$$[X,Y] = -u_{(X),(Y)}[Y,X],\qquad (2)$$

$$[X[Y,Z]] = [[X,Y],Z] + u_{(X),(Y)}[Y,[X,Z]].$$

Obviously, $u_{(X)} \equiv u_{(X),(X)} = \pm 1$; further properties are $u_{(X),(Y)} = u_{(Y),-(X)}$ and $u_{0,(X)} = u_{(X),0} = 1$. If $u_{(X)} = +1$ for all homogeneous X, L is called a color algebra; otherwise, it is called a color superalgebra. In terms of a basis $\{X_a\}$ of L, we have

$$[X_a,X_b] = C_{ab}{}^{c}X_c,$$

and the conditions (1) become

$$C_{ab}{}^{c} = \delta_{(a)+(b),(c)}C_{ab}{}^{c},$$

$$[X_a,X_b] = -u_{(a),(b)}[X_b,X_a],\qquad (1')$$

$$c_{ab}{}^{d}C_{dc}{}^{e}u_{(c),(a)} + C_{bc}{}^{d}C_{da}{}^{e}u_{(a),(b)} + C_{ca}{}^{d}C_{db}{}^{e}u_{(b),(c)} = 0.$$

Examples of gradings and commutation factors are given by Rittenberg and Wyler.[3] The case $\Gamma = \mathbb{Z}_2$ and $u_{(X),(Y)} = (-1)^{(X)(Y)}$ corresponds to Lie superalgebras. If $\Gamma = \mathbb{Z}_2 \oplus \mathbb{Z}_2$ with $(X) = (X_1,X_2)$ and $(Y) = (Y_1,Y_2)$, one may take, for example,

$$u_{(X),(Y)} = \exp\{2\pi i[\tfrac{1}{2}(X_1Y_2 - X_2Y_1) + (X_1Y_1 + X_2Y_2)]\},$$

for a color superalgebra. In general, given several grading types (degrees) and a multiplet $(m) = (m_0, m_1,\cdots)$ of whole numbers, graded endomorphisms of the space $\mathbb{C}^M$, $M = \Sigma_\alpha m_\alpha$, form a color (super)algebra, which we shall denote gl((m)), the general linear color (super)algebra. For $\Gamma = \mathbb{Z}_2$ or $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ we would then have gl(m_0/m_1), gl$(m_{00}/m_{01}/m_{10}/m_{11})$, and so on.

The main results of the paper are as follows. In Sec. II we present in explicit form the defining relations of gl$((m))$,

 0022-2488/83/071681-07$02.50

sl((m)), and osp((m)) with the generators in two-index notation $E^i{}_j$. For sl((m)) the generators are projected from those of gl((m)) by means of a trace condition, while for osp((m)) the projection is effected through the introduction of a graded symmetrical bilinear form (metric tensor). In each case it is shown that the matrix powers

$$\begin{aligned}
&\hat{\delta}^i{}_j = \delta^i{}_j u_{(j)},\\
&(E^1)^i{}_j = \hat{\delta}^i{}_k u_{(k)} E^k{}_j = E^i{}_j,\\
&(E^{p+1})^i{}_j = (E^p)^i{}_k u_{(k)} E^k{}_j,
\end{aligned} \tag{3}$$

with a summation convention on the index k, are tensor operators in the enveloping algebra, transforming like $E^i{}_j$ itself. The traces

$$C_p = (E^p)^k{}_k \tag{4}$$

are the required Casimir invariants.

Weight vectors for gl((m)) and sl((m)) are defined in Sec. III by identifying a set of simultaneously diagonalizable generators (Cartan subalgebra). In those representations where a highest-weight vector exists, the eigenvalue of C_2 is calculated in terms of its components. The corresponding polynomial characteristic identity, satisfied by the matrix of generators $\hat{E}^i{}_j = E^i{}_j u_{(j)}$, in such representations, may be derived by tensor operator projection techniques[16,18] or by infinitesimal character considerations.[20,21] Both methods are discussed, and the identity for gl((m)) and sl((m)) obtained in the factorized form

$$\prod_i (\hat{E} - \alpha_i) = 0 \tag{5}$$

with the α_i given in terms of the components of the highest-weight vector and combinatorial factors. These polynomial identities are the generalization for color (super) algebras of the Cayley–Hamilton identity for ordinary matrices (for Lie algebras, see also Gould[22]). Their existence, as Agrawala[20] has shown, is a consequence of the structure of the quadratic Casimir invariant, but follows also from general arguments[23] concerning matrices over an associative algebra. One consequence of (5) worth pointing out is that there are only a finite number of Casimir invariants of the trace form. For, if M is the degree of the characteristic identity, it follows that higher matrix powers may be rewritten in terms of lower ones; the same goes for the traces $C_{M+1}, C_{M+2}, \cdots$.

In Sec. IV, a tensor calculus for gl((m)) and sl((m)) is developed using the vector operator formalism.[14] The notion of a color-graded permutation is made precise, and Young diagrams are defined. Examples are given to show that the usual product and branching rules apply. In addition, explicit formulae are given for the dimension and graded dimension of representations corresponding to the lowest diagrams, independently of the type of color (super) algebra in question.

In Sec. V a particle interpretation is introduced via color annihilation and creation operators, and the connection with generalized (modular) statistics outlined. The "modules" of the latter (composite operators obeying ordinary statistics) correspond to zero-graded vectors in the associated representation space.

Finally, in Sec. VI, as an application of the characteristic identities, explicit formulae are derived for the eigenvalues of the quadratic and higher Casimir invariants, in terms of the highest weights λ, for gl((m)) and osp((m)).

II. DEFINING RELATIONS AND TENSOR OPERATORS

As discussed in the Introduction, the color general linear algebra is generated by graded endomorphisms of the space $\mathbb{C}^M$. Relative to a fixed basis, these are given in terms of M^2 elementary matrices

$$(e^i{}_j)_p{}^q = \delta^i{}_p \delta_j{}^q, \tag{6}$$

satisfying

$$e^i{}_j e^k{}_l = \delta^k{}_j e^i{}_l.$$

Here $M = \Sigma_\alpha m_\alpha$ and the grading is $(e^i{}_j) = (j) - (i)$. The color (super) algebra arises from the bracket

$$[e^i{}_j, e^k{}_l] = e^i{}_j e^k{}_l - u_{j-i,l-k} e^k{}_l e^i{}_j,$$

dropping the brackets in $u_{(\ ),(\ )}$ where no confusion arises. From (6), the defining relations of gl((m)) in two-index notation are

$$[e^i{}_j, e^k{}_l] = \delta^k{}_j e^i{}_l - u_{j-i,l-k} \delta^i{}_l e^k{}_j. \tag{7}$$

Going to the dual space $\mathbb{C}^{M*}$, the matrices

$$(\bar{e}^i{}_j)^p{}_q = -u_{j-i,i} \delta_j{}^p \delta^i{}_q, \tag{8}$$

the negative color (super) transpose of the $e^i{}_j$, also satisfy the gl((m)) algebra (7) (and correspond to the contragredient representation).

Equations (6)–(8) allow the identification of tensor operators corresponding to the contravariant vector, adjoint, and covariant vector representations, respectively. The transformation rules are

$$[E^i{}_j, V^k] = \delta^k{}_j V^i, \tag{9}$$

$$[E^i{}_j, X^k{}_l] = \delta^k{}_j X^i{}_l - u_{j-i,l-k} \delta^i{}_l X^k{}_j, \tag{10}$$

$$[E^i{}_j, V_k] = -u_{j-i,k} \delta^i{}_k V_j, \tag{11}$$

respectively (for formal definitions, see Ref. 17 and references cited), where the $E^i{}_j$ provide a representation of gl((m)).

It is easily verified from (10) that if $X^i{}_j$, $Y^k{}_l$ are adjoint tensor operators, then so is $Z^i{}_j = X^i{}_k u_k Y^k{}_j$, and that, furthermore, the traces $Z^k{}_k$ commutes with all generators $E^i{}_j$. Thus the matrix powers (3) are adjoint tensor operators, and the traces (4) are the Casimir invariants, viz., $C_1 = E^k{}_k$, $C_2 = E^i{}_j u_j E^j{}_i$, and so on.

The above formalism for gl((m)) may readily be carried over to subalgebras. Indeed any color (super) algebra with defining relations

$$[X_a, X_b] = C_{ab}{}^c X_c$$

may be embedded in some gl((m)) by the identification

$$X_a = C_{ab}{}^c e^{*b}{}_c,$$

where $e^{*b}{}_c$ is a conjugate of $e^b{}_c$ satisfying

$$[e^{*a}{}_b, e^{*c}{}_d] = \delta^a{}_d e^{*c}{}_b - u_{b-a,d-c} \delta^c{}_b e^{*a}{}_d,$$

(where $u^*_{i,j} = u_{j,i}$, and one may take $e^{*a}{}_b = e^b{}_a$). A wide class of subalgebras may arise as fixed point sets of color-graded (super) algebra homomorphisms; in the $\mathbb{Z}_2$ superalgebra case, this covers[17] all the classical series (up to trivial factors).

We shall consider two examples, the special linear and orthosymplectic cases. For sl((m)), consider the matrices

$$a^i_j = e^i_j - \delta^i_j u_j (e^k_k / \hat{M}), \tag{12}$$

where summation on k is understood, and $\hat{M} = \Sigma_\alpha m_\alpha u_\alpha$ is assumed to be nonzero. The a^i_j are the $M^2 - 1$ generators of sl((m)) and satisfy

$$a^k_k = 0,$$
$$[a^i_j, a^k_l] = \delta^k_j a^i_l - u_{j-i,l-k} \delta^i_l a^k_j, \tag{13}$$

that is, formally the same as for gl((m)).

In the osp((m)) case the space $\mathbb{C}^M$ possesses a nonsingular bilinear form (metric) with components g_{ij} relative to the basis and dual basis. We require of g that it be color-graded symmetrical and even, that is,

$$g_{ij} = u_{j,i} g_{ji}$$
and
$$g_{ij} = \delta_{(i),(j)} g_{ij}. \tag{14}$$

The generators of osp((m)) are

$$s_{ij} = g_{ki} e^k_j - u_{j,i} g_{kj} e^k_i = -u_{j,i} s_{ji}, \tag{15}$$

and satisfy the defining relations

$$[s_{ij}, s_{kl}] = g_{jk} s_{il} - u_{j,i} g_{ik} s_{jl} - u_{l,k} g_{jl} s_{ik} + u_{j,i} u_{l,k} g_{il} s_{jk}. \tag{16}$$

Clearly, the generators s_{ij} are not all independent; in fact, the symmetry $s_{ij} = -u_{j,i} s_{ji}$ indicates that there are $\Sigma_\alpha \frac{1}{2} m_\alpha (m_\alpha - u_\alpha) + \Sigma_{\alpha \neq \beta} m_\alpha m_\beta$, or $\frac{1}{2}(M^2 - \hat{M})$, generators of osp((m)).

If the contravariant form g^{ij} is defined by $g^{ij} g_{jk} = \delta^i_k$, and a matrix $s^i_j = u_i g^{ik} s_{kj}$ corresponding to s_{ij}, it is easy to verify that the (generalized) traces of all powers of s^i_j are osp((m)) invariants, and, in particular, that $u_i s^i_j s^j_i$ is the fundamental quadratic invariant.

III. CHARACTERISTIC IDENTITIES FOR gl((m)) AND sl((m))

The derivation of Casimir eigenvalues and characteristic identities for gl((m)) and sl((m)) requires the underlying Lie algebra and root system to be identified. From (6), the grading vector (e^i_j) of e^i_j is zero whenever $(i) = (j)$. Clearly, the m_α^2 matrices $\{e^i_j \mid (i) = (j) = \alpha\}$ generate gl(m_α) and commute with generators of gl(m_β) for $\alpha \neq \beta$. Thus the underlying Lie algebra is Π_α gl(m_α) and a suitable Cartan subalgebra is $\{e^i_i$ (no sum)$\}$. Weights can be lexically ordered if an ordering $<$ on Γ is adopted; within each grading type α we label indices $(\Sigma_{\beta<\alpha} m_\beta) + 1 \leqslant i \leqslant (\Sigma_{\beta<\alpha} m_\beta) + m_\alpha$. Then it follows that e^i_j is raising if $i > j$, and lowering if $i < j$. The Casimir-invariant operator $C_2 = e^i_j u_j e^j_i$ can then be normal-ordered using the defining relations (7). In a finite-dimensional representation π_λ with highest weight λ and components

$$E^i_i v = \lambda_i v, \quad E^i_i = \pi_\lambda(e^i_i),$$

on the highest weight vector, we find

$$C_2 = \sum_i \lambda_i u_i (\lambda_i + M_i + 1 - 2i), \tag{17}$$

where

$$M_i = 2i + 1 = u_i \Big(\sum_{j<i} \delta^j_j u_j - \sum_{j>i} \delta^j_j u_j \Big)$$

or

$$M_i - 2 \sum_{j<i} m_j + m_i + u_i \Big(\sum_{j<i} u_j m_j - \sum_{j>i} u_j m_j \Big). \tag{18}$$

The derivation of a characteristic polynomial identity for generators E^i_j satisfying (7) follows the ordinary superalgebra[16,17] and Lie algebra[18,19] cases. A vector operator V^j satisfying (9) is assumed invariantly decomposable into a set of components V^j_k, each of which changes one of the highest-weight components by one and commutes with the remainder:

$$[\lambda_i, V^j_k]_- = \delta_{ik} V^j_k. \tag{19}$$

Thus

$$[C_2, V^j_k]_- = u_k (2\lambda_k + M_k - 2k) V^j_k, \tag{20}$$

from (17). However, from the definition [see (4)]

$$C_2 = E^i_j u_j E^j_i,$$

and using (9), one has

$$C_2, V^j_k = 2E^j_m u_m V^m_k - \hat{M} V^j_k, \tag{21}$$

or, from (20) and (21),

$$\{E^j_m u_m - [\lambda_m + \tfrac{1}{2}(M_m + u_m \hat{M}) - m] \delta^j_m u_m\} V^m_k = 0.$$

Finally, since this is true for every component of V^i, we have as a matrix equation

$$\prod_{i=1}^{M} \{\hat{E} - [\lambda_i + \tfrac{1}{2}(M_i + u_i \hat{M}) - i] \hat{\delta}\} = 0, \tag{22}$$

where $\hat{E}^i_j = E^i_j u_j$ and $\hat{\delta}^i_j = \delta^i_j u_j$. Similar considerations for a covariant vector operator lead to the related matrix characteristic identity

$$\prod_{i=1}^{M} \{\bar{E} - [\lambda_i + \tfrac{1}{2}(M_i - u_i \hat{M}) + 1 - i] \bar{\delta}\} = 0, \tag{23}$$

where $\bar{E}_i^j = E^j_i u_i$ and $\bar{\delta}_i^j = \delta^j_i u_i$.

An alternative derivation[20,21,24] of (22) and (23) starts from a representation

$$\pi(e) = \pi_\lambda(e) \times 1 + 1 \times \pi_\phi(e),$$

where λ is as before and ϕ is either the fundamental ρ or its contragredient ρ^*. Taking matrix elements of the operator

$$E_\phi = \tfrac{1}{2}[\pi(C_2) - \pi_\lambda(C_2) \times 1 - 1 \times \pi_\phi(C_2)],$$

with respect to a standard basis of $\mathbb{C}^M$ or $\mathbb{C}^{M*}$, one finds that E_{ρ^*} and E_ρ turn out to be just the matrices $\hat{E}$ and $\bar{E}$, respectively. The identities are just the characteristic equations for $\hat{E}$ and $\bar{E}$, their eigenvalues being obtained by repeated use of (17) for λ, ϕ and the constituents of $\lambda \times \phi$. The results are subject to multiplicity criteria in general.[16,25] There may be some repeated factors in the identity, which can be omitted. There are also complications for infinite-dimensional and incompletely reducible representations. However, classes of completely reducible finite-dimensional representations will be identified in the next section, where the criteria specified above are satisfied.

IV. REPRESENTATIONS AND DIAGRAM TECHNIQUES

Finite-dimensional representations of gl((m)) [and also sl((m)) and osp((m))] may be investigated using the tensor operator formalism of the previous sections, analogously to the superalgebra case.[14] For simplicity, we take only contravariant operators (corresponding to Kronecker products of the fundamental representation on $\mathbb{C}^M \times \mathbb{C}^M \times \cdots$); covariant and mixed tensors can be similarly treated. The result is that diagram methods still apply, with appropriate modifications for the color grading.

The rank-1 case is just the vector operator already introduced, transforming as

$$[E^i{}_j, V^k] = \delta^k{}_j V^i,$$

corresponding to the fundamental irreducible representation, to which we assign the diagram $\square$ or $\{1\}$. A rank-2 tensor satisfies

$$[E^i{}_j, V^k] = \delta^k{}_j V^{il} + u_{j-i,-k}\delta^l{}_j V^{ki},$$

and it is readily shown, using the distributivity and symmetry of the commutation factor, that the color-graded symmetrical and antisymmetrical parts,

$$V^{ij}_{\pm} = \tfrac{1}{2}(V^{ij} \pm u_{i,j} V^{ji}), \quad V^{ij}_{\pm} = \pm u_{i,j} V^{ji},$$

transform identically to V^{ij} and thereby provide a decomposition into symmetrized parts denoted as usual by $\square\square$ and $\begin{smallmatrix}\square\\\square\end{smallmatrix}$, or $\{2\}$ and $\{1^2\}$, respectively.

For higher-rank tensors, the appropriate symmetrizations[14] are with respect to the color-graded permutation group (permutations plus factors $u_{i,j}$ for transpositions). Results which depend only upon the permutation group carry through to the color (super) algebra case. For example, the usual Littlewood–Richardson product rule holds, and the branching rule[14,26] $\mathrm{gl}((m+n)) \supset \mathrm{gl}((m)) \times \mathrm{gl}((n))$ is

$$\{\lambda\} = \sum_{\xi} \{\lambda/\xi\} \times \{\xi\}, \tag{24}$$

as usual, where the summation extends over all Young tableaux $\{\xi\}$ by which the tableau $\{\lambda\}$ is divisible. An extension of this rule to $\mathrm{gl}((m)) \supset \Pi_\alpha\, \mathrm{gl}((m_\alpha))$, where $(m_\alpha) = (m_{\alpha 0}, m_{\alpha 1}, m_{\alpha 2}, \cdots)$ is defined by $m_{\alpha\beta} = \delta_{\alpha\beta} m_\alpha$, yields the branching rule $\mathrm{gl}((m)) \supset \Pi_\alpha\, \mathrm{gl}(m_\alpha)$, using the fact that a tableau $\{\lambda\}$ of the color (super) algebra $\mathrm{gl}((m_\alpha))$ coincides with the tableau $\{\lambda\}$ of the Lie algebra $\mathrm{gl}(m_\alpha)$, or its conjugate tableau $\{\tilde\lambda\}$, according as $u_\alpha = +1$ or $u_\alpha = -1$, respectively.

Dimensions of representations corresponding to color-graded tableaux can be worked out from the $\Pi_\alpha \mathrm{gl}(m_\alpha)$ branching rule for each type of grading and corresponding color (super) algebra. However, it is of some interest to have dimension formulae $D\{\lambda\}$ which depend only upon the tableau. Easiest is the graded dimension $\hat D\{\lambda\}$, the difference between the dimensions of the even and odd subspaces (a vector is even or odd according as its grading α satisfies $u_\alpha = +1$ or $u_\alpha = -1$, respectively). For the fundamental representation, this is obviously $\hat M = \Sigma_\alpha u_\alpha m_\alpha$. For general tableaux, it is a polynomial in $\hat M$: as for the superalgebra case,[15] this polynomial is identical to that for the dimension of the same tableau in ordinary $\mathrm{gl}(\hat M)$. Up to rank 4, we therefore have[26]

$$\begin{aligned}
&\hat D\{1\} = \hat M, \quad \hat D\{2\} = \hat M(\hat M+1)/2,\\
&\hat D\{1^2\} = \hat M(\hat M-1)/2, \quad \hat D\{3\} = \hat M(\hat M+1)(\hat M+2)/6,\\
&\hat D\{21\} = \hat M(\hat M^2-1)/3, \quad \hat D\{1^3\} = \hat M(\hat M-1)(\hat M-2)/6,\\
&\hat D\{4\} = \hat M(\hat M+1)(\hat M+2)(\hat M+3)/24, \quad \hat D\{31\} = \hat M(\hat M+2)(\hat M^2-1)/8,\\
&\hat D\{2^2\} = \hat M^2(\hat M^2-1)/12, \quad \hat D\{21^2\} = \hat M(\hat M-2)(\hat M^2-1)/8,\\
&\hat D\{1^4\} = \hat M(\hat M-1)(\hat M-2)(\hat M-3)/24.
\end{aligned} \tag{25}$$

For the total dimension $D\{\lambda\}$ (the sum of the dimensions of the even and odd subspaces), the polynomials are functions of both M and $\hat M$, as foreshadowed by the osp((m)) dimension (corresponding to $D\{1^2\}$) already given. The following formulae, derived for the $\mathbb{Z}_2$ case $\mathrm{gl}(m_0/m_1)$, can be verified by taking particular examples for other gl((m)):

$$\begin{aligned}
&D\{1\} = M, \quad D\{2\} = (M^2+\hat M)/2,\\
&D\{1^2\} = (M^2-\hat M)/2, \quad D\{3\} = (M^3+3M\hat M+2M)/6,\\
&D\{21\} = (M^3-M)/3, \quad D\{1^3\} = (M^3-3\hat M M+2M)/6,\\
&D\{4\} = (M^4+6\hat M M^2+[8M^2+3\hat M^2]+6\hat M)/24, \quad D\{31\} = (M^4+2\hat M M^2-\hat M^2-2\hat M)/8,\\
&D\{22\} = (M^4+3\hat M^2-4M^2)/12, \quad D\{21^2\} = (M^4-2\hat M M^2-\hat M^2+2\hat M)/8,\\
&D\{1^4\} = (M^4-6\hat M M^2+[8M^2+3\hat M^2]-6\hat M)/24.
\end{aligned} \tag{26}$$

The similarity between (25) and (26) is seen by expanding (25) in powers of $\hat M$. $D\{\lambda\}$ is obtained from $\hat D\{\lambda\}$ by replacing each power of $\hat M$ by an appropriate homogeneous expression in M and $\hat M$, with the same overall coefficient. These formulae may also be derived from a generalization of the (super) trace formulae for supercharacters, given by Balantekin and Bars,[15] to color superalgebras. The general formula

$$D\{\lambda\} = \sum_{(\rho)} \chi^\lambda_{(\rho)} h_{(\rho)} (M^{\rho_1+\rho_3+\cdots})(\hat M^{\rho_2+\rho_4+\cdots})/r!,$$

where $\chi^{\lambda}_{(\rho)}$ is the character of the symmetric group S_r for the irreducible representation λ and class $(\rho) = (1^{\rho_1} 2^{\rho_2} \cdots)$, and $h_{(\rho)}$ the class order, follows by appropriate substitution in the corresponding S-function power sum expansion.[27,28]

As an example of the above formulae, we have the following products and dimension checks in gl(2/1) ($\mathbb{Z}_2$ superalgebra):

$$\square \times \square = \square\square + \begin{array}{l}\square\\\square\end{array}$$
$$3 \times 3 = 5 + 4,$$
$$\square\square \times \square = \begin{array}{l}\square\square\\\square\end{array} + \square\square\square$$
$$5 \times 3 = 8 + 7,$$
$$\begin{array}{l}\square\\\square\end{array} \times \square = \begin{array}{l}\square\square\\\square\end{array} + \begin{array}{l}\square\\\square\\\square\end{array}$$
$$4 \times 3 = 8 + 4,$$

whereas in gl(2/1/1/1) ($\mathbb{Z}_2 \oplus \mathbb{Z}_2$ superalgebra) we have

$$\square \times \square = \square\square + \begin{array}{l}\square\\\square\end{array}$$
$$5 \times 5 = 13 + 12,$$
$$\square\square \times \square = \begin{array}{l}\square\square\\\square\end{array} + \square\square\square$$
$$13 \times 5 \quad 40 + 25,$$
$$\begin{array}{l}\square\\\square\end{array} \times \square \quad \begin{array}{l}\square\square\\\square\end{array} + \begin{array}{l}\square\\\square\\\square\end{array}$$
$$12 \times 5 \quad 40 + 20.$$

In the case of a mixed tensor representations (corresponding to tensor operators $V^{ij\cdots}{}_{pq\cdots}$) it can be shown that traces with respect to the invariant tensor δ^i_j are tensor operators of lower rank. It follows that a resolution may be made into traceless mixed tensors

$$\{\bar{\lambda}\} \times \{\mu\} = \sum_{\zeta} \{\overline{\lambda/\zeta}; \mu/\zeta\}$$

in the usual manner.[26] However, as in the superalgebra case, these representations, and their Kronecker products, may not be completely reducible: For example, this is the case for the representation $\{\bar{1};1\}$ whenever $\widehat{M} = 0$.

V. PARTICLE INTERPRETATION AND MODULAR STATISTICS

A particle interpretation for color (super) algebra arises from postulated color creation and annihilation operators satisfying

$$\begin{aligned}
[c_i, c_j] &= c_i c_j \pm u_{(i),(j)} c_j c_i = 0, \\
[c^i, c^j] &= c^i c^j \pm u_{(i),(j)} c^j c^i = 0, \\
[c_i, c^j] &= c_i c^j \pm u_{(i),-(j)} c^j c_i = \delta_i^j.
\end{aligned} \tag{27}$$

In modular or generalized statistics[10,11] analogous relations arise, and it is convenient to treat both cases together. For modular Fermi (Bose) statistics, start with a set of fermion (boson) operators $a^{\alpha}{}_i$, $a^j{}_{\beta}$ [as in the previous sections, α and (i) are grading vectors, and i a general affix], such that

$$\begin{aligned}
[a^{\alpha}{}_i, a^{\beta}{}_j] &= a^{\alpha}{}_i a^{\beta}{}_j \pm a^{\beta}{}_j a^{\alpha}{}_i = 0, \\
[a^i{}_{\alpha}, a^j{}_{\beta}] &= a^i{}_{\alpha} a^j{}_{\beta} \pm a^j{}_{\beta} a^i{}_{\alpha} = 0, \\
[a^{\alpha}{}_i, a^j{}_{\beta}] &= a^{\alpha}{}_i a^j{}_{\beta} \pm a^j{}_{\beta} a_i{}^{\alpha} = \delta^{\alpha}{}_{\beta} \delta_i{}^j.
\end{aligned} \tag{28}$$

Then there always exists a (unitary) permutation operator U_α such that $U_\alpha U_\beta = U_{\alpha+\beta}$, $U_0 = 1$, and

$$U_\alpha a_i{}^{\beta} U_{-\alpha} = a_i^{\beta+\alpha}, \quad U_\alpha a^i{}_{\beta} U_{-\alpha} = a^i_{\beta+\alpha}. \tag{29}$$

The following construction is due essentially to Kleeman.[29] Let

$$b_i{}^{\alpha} = a_i^{(i)-\alpha} U_{(i)}, \quad b^i{}_{\alpha} = U_{-(i)} a^i_{(i)-\alpha}.$$

Using (28) and (29), it follows that

$$\begin{aligned}
b_i{}^{\alpha} b_j{}^{\beta} &\pm b_j^{\beta-(i)} b_i^{\alpha+(j)} = 0, \\
b^i{}_{\alpha} b^j{}_{\beta} &\pm b^j_{\beta-(i)} b^i{}_{\alpha+(j)} = 0, \\
b_i{}^{\alpha} b^j{}_{\beta} &\pm b^j_{\beta-(i)} b_i^{\alpha-(j)} = \delta^{\alpha}{}_{\beta} \delta_i{}^j.
\end{aligned} \tag{30}$$

These are just the relations satisfied by the ansatz components in modular Fermi (Bose) statistics[10,11] (where the defining relations are independent of color).

Further simplification, and the connection with color (super) algebra, follow when there exists a commutation factor (thus far, the discussion has been for any grading). For , define

$$c_i = \sum_{\alpha} [u_{\alpha,(i)}]^{1/2} b_i{}^{\alpha}, \quad c^i = \sum_{\alpha} [u_{(i),\alpha}]^{1/2} b^i{}_{\alpha},$$

and note that

$$\begin{aligned}
\sum_{\alpha} [u_{\alpha,(i)}]^{1/2} b_i^{\alpha+(j)} &= \sum_{\alpha} [u_{-(j),(i)}]^{1/2} [u_{\alpha+(j),(i)}]^{1/2} b_i^{\alpha+(j)} \\
&= [u_{(i),(j)}]^{1/2} c_i
\end{aligned}$$

and

$$\sum_{\beta} [u_{\beta,(j)}]^{1/2} b_j^{\beta-(i)} = [u_{(j),-i}]^{1/2} c_j = [u_{(i),(j)}]^{1/2} c_j.$$

Therefore, using (30),

$$\begin{aligned}
c_i c_j &= \sum_{\alpha,\beta} [u_{\alpha(i)}]^{1/2} [u_{\beta,(j)}]^{1/2} b_i{}^{\alpha} b_j{}^{\beta} \\
&= \mp \sum_{\alpha,\beta} [u_{\beta,(j)}]^{1/2} [u_{\alpha,(i)}]^{1/2} b_j^{\beta-(i)} b_i^{\alpha+(j)} \\
&= \mp u_{(i),(j)} c_j c_i,
\end{aligned} \tag{27'}$$

or

$$[c_i, c_j] = c_i c_j \pm u_{(i),(j)} c_j c_i = 0.$$

This relation, and similar ones satisfied by $[c^i, c^j]$ and $[c_i, c^j]$, just reproduce the postulated relations (27) for color creation and annihilation operators.

Given (27), gl((m)) generators satisfying (7) are obviously

$$E^i{}_j = c^i c_j. \tag{31}$$

In a Fock representation, with vacuum

$$c_i |\rangle = 0, \tag{32}$$

the state $c^i| >$, $c^i c^j| >$, $\cdots$ are obviously totally color-graded symmetrical [or antisymmetrical, with the $+$ sign in (27)]. General tableaux can be built up by considering several species c^i, $c^{j'}$, $c^{k''}$, $\cdots$ of color operators. Of particular interest in modular statistics are composite operators, built from the b_i, b^i, where (i) is fixed, obeying ordinary Bose or Fermi statistics (corresponding to physical fields). In general, these so-called "modules" are just vectors of weight $(i) = 0$. In the color superalgebra case, one has for example $c^i c_i$ or $c^{i_1}c^{i_2}\cdots c^{i_N}$, where N is the characteristic of the grading group.

It is important to note that the restriction to finite-dimensional irreducible representations of the E^i_j defined in (31) does not necessarily imply the same restriction for the c_i and c^i, which can be regarded as shift operators between contiguous irreducible representations of the E^i_j. This is sufficiently illustrated by some of the simplest of applications, e.g., to bosons, where the momenta are limited to a finite set. We may, in fact, suppose in a model theory that the momentum states are discrete and compact.

To complete the physical picture, it is desirable to assign a meaning to the invariants λ_k of gl((m)) and the vectors V^j_k defined in (19), which are projected from the creation operators c^j by maximal factors of the characteristic identity (22). The simplest applications, to particles satisfying parastatistics,[30] are sufficient to show that the characteristic identity contains redundant factors wherever particles of the same type are present, so that the degree of the minimal polynomial depends on the number of kinds of particles, e.g., the order of the parastatistics, but is independent of the number of momentum states. If we examine the factors

$$\hat{E} - [\lambda_k + \tfrac{1}{2}(M_k + u_k\hat{M}) - k\,]\hat{\delta},$$

of (22), we see that the u_k have the value -1 for half-odd-integral spin and $+1$ for integral spin. Also, if we assume that the m_k are all equal to some simple multiple of the number of momentum states, the M_k defined by (18) depend only on color and spin. It follows that each distinct factor of the characteristic identity (apart from the trivial factor $\hat{E}$, which is present even when only fermions or bosons are considered) is associated with particles of a particular color and spin. Thus, λ_k may be interpreted as the maximum number of particles of type (k), when values of λ_m for $(m) < (k)$ have been assigned. Also, the eigenvectors V^j_k of the matrix $\hat{E}$ represent creation operators for particles of type (k). The use of color algebras may lead, in the future, to the development of field theories in which elementary particles of different types are represented by the same field variable.

VI. HIGHER CASIMIR INVARIANTS

We have already seen that many results, based on techniques developed originally for the study of finite-dimensional Lie algebras, can be generalized in a straightforward way to the color algebras and superalgebras. This is true also of the method of calculation of the higher Casimir invariants in terms of the highest weights λ. For ordinary Lie algebras, the essential results in their simplest form were obtained originally by Louck and Biedenharn[31]; easier derivations, based on the use of the characteristic identities, were given by Okubo[32] and Green.[19,33] The following generalization follows the method of the latter.

Let us write the characteristic identity (22) for gl((m)) in the form

$$\prod_{i=1}^{M} (\hat{E} - \alpha_i) = 0, \quad \alpha_i = u_i\lambda_i + \sum_{j>i} u_i. \tag{33}$$

Then, as for ordinary matrices, the operator matrices defined by

$$P_i = \prod_{j\neq i}^{M} (\hat{E} - \alpha_j)/(\alpha_i - \alpha_j) \tag{34}$$

are projective idempotents, and any power $\hat{E}^q$ of the matrix $\hat{E}$ can be expressed in the form $\Sigma_i \alpha_i^q P_i$, so that the higher Casimir invariants C_q are given by

$$C_q = \mathrm{ctr}(\hat{E}^q) = \sum_i \alpha_i{}^q \mathrm{ctr}(P_i), \tag{35}$$

where $\mathrm{ctr}(A) = (A\hat{\delta})^k{}_k$ and $\hat{\delta}^i_j = \delta^i_j u_j$, as previously. The calculation of the Casimir invariants is thereby reduced to the determination of the color traces

$$T_i = \mathrm{ctr}\left(\prod_{j\neq i}^{M} (\hat{E} - \alpha_j)\right). \tag{36}$$

Now, it is evident from the definition of the Casimir invariants that they are invariant under the transformation $E^j{}_k \to E^{\mathrm{P}j}{}_{\mathrm{P}k}$, $u_j \to u_{\mathrm{P}j}$, where (Pj) denotes any permutation of the affixes (j). Under this transformation, $\alpha_j \to \alpha_{\mathrm{P}j}$, and it follows that C_q, expressed as a polynomial in the α_j and u_j, is symmetric in the pairs (α_j, u_j). It also follows that T_i, expressed as a polynomial in the α_j and u_j, is symmetric in all such pairs with the exception of (α_i, u_i). Since T_i is of degree $M-1$ in the α_j, if we can show that it has a factor $(\alpha_i - \alpha_j - u_j)$, it must have the form

$$T_i = c_i \prod_{j\neq i} (\alpha_i - \alpha_j - u_j), \tag{37}$$

where c_i is independent of the representation and can depend only on the u_j.

The existence of a factor $(\alpha_i - \alpha_j - u_j)$ of T_i is established by the same argument as is applicable for simply graded algebras.[31] That T_i can depend only on differences of the eigenvalues α_i of $\hat{E}$ is already evident from the fact that P_i is invariant under the transformation $\hat{E}^j{}_k \to \hat{E}^j{}_k + \mu\delta^\lambda{}_\mu$, $\alpha_j \to \alpha_j + \mu$. But the projection operator P_i may be expressed in the form $V_i\bar{V}_i$ (i.e., $P_i{}^j{}_k = V_i{}^j\bar{V}_{ik}$), where $V_i{}^k$ is a vector operator as defined in (9) and $\bar{V}_{ik}$ is its dual, suitably normalized. Also, according to (19), V_i increases the eigenvalue of α_i in any irreducible representation by one unit, and $\bar{V}_i$ decreases the eigenvalue of λ_i by one unit. But as $\lambda_i \geqslant \lambda_{i+1}$ when $u_i = u_{i+1}$, P_i must vanish in a representation with $\lambda_i = \lambda_{i+1}$. Similarly, $\lambda_i + \lambda_{i+1} \geqslant 0$ when $u_i = -u_{i+1}$. So T_i must vanish whenever $\lambda_i u_i = \lambda_{i+1}u_{i+1}$, i.e., when $\alpha_i = \alpha_{i+1} + u_{i+1}$, and must have the factor $(\alpha_i - \alpha_{i+1} - u_{i+1})$.

The multiplier c_i in (37) is easily found from comparing the value of C_0 given by (35) with its known value:

$$\sum_{i=1}^{M} c_i \prod_{j\neq i} \frac{\alpha_i - \alpha_j - u_j}{\alpha_i - \alpha_j} = \sum_{i=1}^{M} u_i.$$

As this must be an identity (independent of the values assumed by the α_j in a particular representation), we verify that

$$c_i = u_i. \tag{38}$$

The values found for C_1 and C_2 by substitution in (35):

$$C_1 = \sum_i u_i\left(\alpha_i - \sum_{j>i} u_j\right),$$

$$C_2=\sum_i u_i\Big(\alpha_i-\sum_{j>i}u_j\Big)\Big(\alpha_i-\sum_{j<i}u_j\Big), \tag{39}$$

agree with those found in Sec. III when expressed with the help of (37) in terms of highest weights. Conversely, if the eigenvalues of the invariants of C_q are known in any irreducible representation, the values of the α_j and hence of the λ_j may be determined. This suggests a way of extending the definition of the λ_j to representations other than the finite-dimensional representations, which formed the basis of our discussion in the foregoing.

The Casimir invariants of osp((m)) can be found in a similar way. The value of the quadratic invariant in terms of highest weights $\mu_1,\mu_2,...,\mu_H$, where $H=[M/2]$, is obtained by the elementary method used for osp(m/n) in Ref. 16, and can be expressed in a form

$$C_2=2\sum_{i=1}^{H}\mu_i\Big(\mu_i u_i+\sum_{j>1}u_j-\sum_{j<i}u_j-1\Big), \tag{40}$$

strikingly similar to the corresponding result for gl((m)). The characteristic identity obtained with the help of this result is (cf. Sec. III)

$$\begin{aligned}&\prod_{i=1}^{H}(S-\beta_i)(S-\overline{\beta}_i)=0,\\ &\beta_i=\mu_i u_i+\sum_{j>i}u_j-1,\\ &\overline{\beta}_i=-\mu_i u_i+\sum_{j<i}u_j,\end{aligned} \tag{41}$$

where S is the matrix of operators s^i_j as defined in Sec. II, restricted to a finite-dimensional representation. Again the most easily calculated invariants are the

$$\begin{aligned}T_i&=\mathrm{ctr}\Big[(S-\overline{\beta}_i)\prod_{k\neq i}(S-\beta_j)(S-\overline{\beta}_j)\Big],\\ \overline{T}_i&=\mathrm{ctr}\Big[(S-\beta_i)\prod_{j\neq i}(S-\beta_j)(S-\overline{\beta}_j)\Big],\end{aligned} \tag{42}$$

from which it is easy to determine the C_q. The form of T_i and $\overline{T}_i$ is again completely determined, apart from a factor u_i, from consideration of symmetry, together with the inequalities satisfied by the highest weights; the detailed argument is parallel to that given in Ref. 33. For even M, we obtain

$$\begin{aligned}T_i&=u_i(\beta_i-\overline{\beta}_i)\prod_{j\neq i}(\beta_i-\beta_j-u_j)(\beta_i-\overline{\beta}_j-u_j),\\ \overline{T}_i&=u_i(\overline{\beta}_i-\beta_i)\prod_{j\neq i}(\overline{\beta}_i-\beta_j-u_j)(\beta_i-\overline{\beta}_j-u_j).\end{aligned}$$

For odd M, the additional u_M must have the value $+1$, and additional factors $(\beta_i-\beta_M)$ and $(\overline{\beta}_i-\beta_M)$ are required in T_i and $\overline{T}_i$, respectively. But, in addition, the factors $\beta_i-\overline{\beta}_i$ and $\overline{\beta}_i-\beta_i$ must be replaced by $\overline{\beta}_i-\beta_i-2u_i$, respectively, to ensure that T_i and $\overline{T}_i$ vanish when λ_i is changed to λ_i+1, in representations such that $\lambda_i=\lambda_{i+1}$.

VII. CONCLUSION

In this paper we have obtained a large number of new results for color algebras and superalgebras. These include the explicit forms of the quadratic and higher Casimir invariants for gl((m)) and osp((m)) in Secs. III and VI, the characteristic identities satisfied by the matrices of generators of these algebras, the Young diagrams for gl((m)) in finite-dimensional representations in Sec. IV, and the application to particles of different colors described in Sec. V.

ACKNOWLEDGMENTS

One of the authors (P.D.J.) wishes to acknowledge the support of a Queen Elizabeth Fellowship for part of this research and to thank Professor B. G. Wybourne for informative discussions.

[1]L. Corwin, T. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975).
[2]V. Rittenberg, "A Guide to Lie Superalgebras," in *Lecture Notes in Physics*, **79**, edited by P. Kramer and A. Rieckers (Springer-Verlag, Berlin, 1978), and references therein.
[3]V. Rittenberg and D. Wyler, Nucl. Phys. B **139**, 189 (1978).
[4]V. Rittenberg and D. Wyler, J. Math. Phys. **19**, 2193 (1978).
[5]J. Lukierski and V. Rittenberg, Phys. Rev. D **18**, 385 (1978).
[6]M. Scheunert, J. Math. Phys. **20**, 712 (1979).
[7]H. S. Green, Phys. Rev. **90**, 270 (1953).
[8]O. W. Greenberg, Phys. Rev. Lett. **13**, 598 (1964).
[9]A. J. Bracken and H. S. Green, J. Math. Phys. **14**, 1784 (1973).
[10]H. S. Green, Aust. J. Phys. **28**, 115 (1975).
[11]H. S. Green, Aust. J. Phys. **29**, 483 (1976).
[12]M. Scheunert, "The Theory of Lie Superalgebras," in *Lecture Notes in Mathematics*, **716** (Springer-Verlag, Berlin, 1979), and references therein.
[13]V. G. Kac, "Representations of classical Lie Superalgebras," in *Lecture Notes in Mathematics*, **676**, edited by A. Dold and R. Eckmann (Springer-Verlag, Berlin, 1978).
[14]P. H. Dondi and P. D. Jarvis, J. Phys. A **14**, 547 (1981).
[15]A. Baha Balantekin and I. Bars, J. Math. Phys. **22**, 1810 (1981); **22**, 1149 (1981).
[16]P. D. Jarvis and H. S. Green, J. Math. Phys. **20**, 2115 (1979).
[17]M. K. Murray and P. D. Jarvis, "Casimir Invariants and Characteristic Identities for 'Strange' Superalgebras," preprint, University of Tasmania (to appear in J. Math. Phys.).
[18]A. J. Bracken and H. S. Green, J. Math. Phys. **12**, 2099 (1971).
[19]H. S. Green, J. Math. Phys. **12**, 2107 (1971).
[20]V. K. Agrawala, Hadronic J. **4**, 444 (1981).
[21]K. Hannabuss, Mathematical Institute, University of Oxford, preprint 1972 (unpublished).
[22]M. D. Gould, J. Austral. Math. Soc. Ser. B **20**, 290 (1978).
[23]Y. Lehrer-Ilamed, Bull. Res. Counc. Isr. Sect. A **5**, 197 (1956).
[24]D. M. O'Brien, A. Cant, and A. L. Carey, Ann. Inst. H. Poincaré **26**, 405 (1977).
[25]S. A. Edwards (unpublished).
[26]R. C. King, J. Phys. A **8**, 429 (1975).
[27]See, e.g., B. G. Wybourne, *Symmetry Principles and Atomic Spectroscopy* (Wiley, New Tork, 1970).
[28]P. H. Butler and R. C. King, J. Math. Phys. **14**, 1176 (1973).
[29]R. Kleeman, "Aspects of Modular Quantisation," University of Adelaide preprint, Dec. 1981.
[30]A. J. Bracken and H. S. Green, Nuovo Cimento A **9**, 349 (1972).
[31]J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970).
[32]S. Okubo, J. Math. Phys. **16**, 528 (1975).
[33]H. S. Green, J. Aust. Math. Soc. **19**(b), 129 (1975).

Character generators for unitary and symplectic groups

Kenneth Baclawski[a]
Northeastern University, Boston, Massachusetts 02115

(Received 22 July 1982; accepted for publication 2 September 1982)

Combinatorial algorithms for computing the character generators of U(n), SU(n), and Sp($2n$, $\mathbb{C}$) are described. These algorithms produce relatively compact, nested expressions for the character generators. Moreover, the terms appearing in these expressions all have positive coefficients. This feature is not shared by the expression for the character generator which uses the Weyl character formula.

PACS numbers: 02.20. + b

I. INTRODUCTION

The purpose of this paper is to describe a combinatorial algorithm for computing the character generators of the unitary groups U(n) and SU(n) and of the symplectic groups Sp($2n$, $\mathbb{C}$). This algorithm is based on a unique method of labelling every basis vector of the weight spaces of the irreducible representations. In a sequel[1] to this paper, we give a recursive algorithm for computing the action of any element of one of the Lie groups mentioned above on one of the basis vectors. For SU(n) or U(n) these labels are equivalent to Gel'fand patterns[2] and to the usual Young tableaux associated with Sl(n, $\mathbb{C}$). For Sp($2n$, $\mathbb{C}$) our labels are equivalent to the modified Gel'fand patterns based on the branching rules of Hegerfeldt[3] and to the tableaux of King.[4]

One feature of our approach is that it can be extended to other Lie groups. This paper can be regarded as an introduction to the more general (and more complicated) algorithms needed to deal with arbitrary reductive Lie groups. For example, our algorithm has been extended to the exceptional Lie group G_2 by Baclawski and Towber.[5]

The character generator was first defined by Patera and Sharp,[6] who gave an expression for it using the Weyl character formula. However, this expression involves a great deal of cancellation so they posed the question of finding expressions that do not involve any cancellations. Such an expression was found for the groups SU(n) by Stanley.[7] However, his formula rapidly becomes intractably large even for small groups such as SU(5). Subsequently, King[4] extended Stanley's method to Sp($2n$, $\mathbb{C}$). We compare the various known methods of computing the character generator in Sec. IV. For example, our algorithm yields a formula for the character generator of Sp(6, $\mathbb{C}$) that is 52 times smaller than the formula of King. More dramatically, in the case of SU(7), our formula is 65 times smaller than the Patera–Sharp formula and over a *million* times smaller than Stanley's formula.

The use of tableau methods for studying group representations has a long history going back to Young. For an overview of this field see Baclawski (Ref. 1, Introduction; Ref. 8, Sec. 5) and DeConcini, *et al.* (Ref. 9, Chapter III).

II. THE CHARACTER GENERATOR

Let G be a semisimple complex Lie group. We have in mind the two examples Sl(n, $\mathbb{C}$) and Sp($2n$, $\mathbb{C}$). If we are given a finite-dimensional irreducible representation (irrep)

$$\rho: G \to \mathrm{Gl}(V),$$

we define the *character* of ρ to be the function

$$\chi_\rho: G \to \mathbb{C}$$

such that $\chi_\rho(\sigma) = \mathrm{Tr}(\rho(\sigma))$, for $\sigma \in G$. If G is a subgroup of Gl(m, $\mathbb{C}$), then $\chi_\rho(\sigma)$ is determined by the set of eigenvalues $\{\xi_1, \ldots, \xi_m\}$ of σ. Thus we may regard χ_ρ as a symmetric function of $\xi_1, \ldots, \xi_m$.

If G is a proper subgroup of Gl(m, $\mathbb{C}$) then the eigenvalues of an element of G will not be independent of one another, and so it is conventional to view χ_ρ as a function of certain expressions $\alpha_1, \alpha_2, \ldots, \alpha_l$, of a subset of l of the eigenvalues, where l is the rank of G. Although many conventions appear in the literature, there is one that seems to be most popular: the Dynkin basis. It has the advantage that it is intrinsic to the group, i.e., does not require that the group be specified as a subgroup of Gl(m, $\mathbb{C}$).

Here are the conventions for Sl(n, $\mathbb{C}$) and Sp($2n$, $\mathbb{C}$).

(1) $\mathrm{Sl}(n, \mathbb{C}) = \{\sigma \in \mathrm{Gl}(n, \mathbb{C}) \mid \det(\sigma) = 1\}$. The rank is $l = n - 1$, and χ_ρ is a function of

$$\alpha_1 = \xi_1,$$
$$\alpha_2 = \xi_1\xi_2,$$
$$\vdots$$
$$\alpha_{n-1} = \xi_1\xi_2\cdots\xi_{n-1}.$$

(2) $\mathrm{Sp}(2n, \mathbb{C}) = \{\sigma \in \mathrm{Sl}(2n, \mathbb{C}) \mid {}^t\sigma A \sigma = A\}$, where A is the block-diagonal matrix

$$\begin{pmatrix} \begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix} & & & \mathbf{0} \\ & \begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix} & & \\ & & \vdots & \\ \mathbf{0} & & & \begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix} \end{pmatrix}$$

The rank is $l = n$. The eigenvectors of $\sigma \in \mathrm{Sp}(2n, \mathbb{C})$ may be listed as $\xi_1, \xi_1^{-1}, \xi_2, \xi_2^{-1}, \ldots, \xi_n, \xi_n^{-1}$, and χ_ρ is a function of

$$\alpha_1 = \xi_1,$$
$$\alpha_2 = \xi_1\xi_2,$$
$$\vdots$$
$$\alpha_n = \xi_1\xi_2\cdots\xi_n.$$

[a] Supported by NSF Grant MCS 79-03029.

Although $\mathrm{Gl}(n, \mathbb{C})$ is itself not semisimple, we can treat it in a similar manner, using the convention that χ_ρ is a function of $\alpha_1 = \xi_1, \alpha_2 = \xi_2, ..., \alpha_n = \xi_n$. Our techniques apply to any reductive Lie group over the complex numbers.

A *weight vector* of V is a vector v that is an eigenvector of every diagonal matrix, and in particular, if $\sigma = \mathrm{diag}(\xi_1, ..., \xi_m)$, then $\rho(\sigma)v$ will be the vector ξv, for a number ξ which will be a function of $\xi_1, ..., \xi_m$. One can show that ξ will always have the form $\alpha_\mu = \alpha_1^{\mu_1} \cdots \alpha_l^{m_l}$ for some l-tuple of integers $\mu = (\mu_1, ..., \mu_l)$. We will call μ the *weight* of v (in the Dynkin basis). The subspace of V consisting of all weight vectors of weight μ (in other words, the eigenspace belonging to α^μ) is denoted by V_μ and called the *weight space* of weight μ. Its dimension is called the *multiplicity* of μ in V. One can show that V is the direct sum of its weight spaces (i.e., every vector of V is a unique linear combination of weight vectors), and hence that

$$\chi_\rho = \sum_\mu \dim (V_\mu)\alpha^\mu.$$

In other words, χ_ρ is the generating function of the multiplicities of the representation ρ. To distinguish this incarnation of the character from the original definition, one calls this the *formal character* of ρ.

So far we have not made use of the fact that ρ is irreducible. One can show that the irreps can themselves be labelled by l-tuples of integers. If ρ has label $\lambda = (\lambda_1, ..., \lambda_l)$, then we write $V(\lambda)$ for its representation space. The label λ is the *highest weight* of ρ and has these properties:

(1) $\dim V_\lambda(\lambda) = 1$, i.e., λ occurs with multiplicity 1 as a weight of the representation labelled by it; (2) λ is the maximum weight, among all weights occurring with multiplicity 1 or more in $V(\lambda)$, with respect to a certain partial order on weights;

(3) two irreps $V(\lambda)$ and $V(\lambda')$ are isomorphic if and only if $\lambda = \lambda'$;

(4) if G is simply connected, then every l-tuple λ of nonnegative integers occurs as a highest weight of some irrep.

For a group such as $\mathrm{Gl}(n, \mathbb{C})$ which is not semisimple, one can still label irreps with l-tuples of integers, but they may not be nonnegative. However, (1) and (3) above will still hold. For $\mathrm{Gl}(n, \mathbb{C})$ the n-tuples λ that occur as irrep labels are those that satisfy $\lambda_1 \geqslant \lambda_2 \geqslant ... \geqslant \lambda_\mu$.

We come at last to the *character generator*. This is the generating function of the characters of all irreps of G:

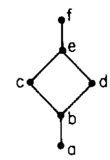
$$\begin{aligned} F_G &= F_G(\alpha; A) = F_G(\alpha_1, ..., \alpha_l; A_1, ..., A_l) \\ &= \sum_\lambda \chi_\lambda(\alpha) A^\lambda \\ &= \sum_\lambda \sum_\mu \dim [V_\mu(\lambda)] \alpha^\mu A^\lambda. \end{aligned}$$

One may regard the character generator as the generating function of the multiplicities of all weights in all irreps of G. It was introduced by Patera and Sharp.[6] As noted by them and as we shall see, it is closely related to the "labelling problem," i.e., the problem of finding unique labels for the basis vectors of the irreps of G.

We end with the remark that the finite-dimensional representation theory of $\mathrm{SU}(n)$ coincides with that of $\mathrm{Sl}(n, \mathbb{C})$ under restriction. Similarly, the representation theory of $\mathrm{U}(n)$ coincides with that of $\mathrm{Gl}(n, \mathbb{C})$.

III. PARTIALLY ORDERED SETS

The combinatorial structure we utilize for our computations is the partially ordered set (poset). A *poset* is a set P together with a reflexive, transitive, antisymmetric relation, written, "$\leqslant$". Here is an example of a poset:



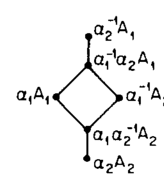
In this poset $a \leqslant e$ and $b \leqslant d$, but c and d are not comparable: $c \nleqslant d$ and $d \nleqslant c$. We will use the usual notation for open and closed intervals in a poset, e.g., $[x, y] = \{z | x \leqslant z \leqslant y\}$. The diagrams we will use to depict a poset, as above, is called its *Hasse diagram*. The elements are the vertices of this diagram and the edges are the covering relations, where y is said to *cover* x in P if the closed interval $[x, y]$ has exactly two elements.

A *multichain* of a poset P is a finite sequence (possibly empty) $x_1, x_2, ..., x_r \in P$ such that $x_1 \leqslant x_2 \leqslant ... \leqslant x_r$. Note that repetitions are allowed. Certain special cases are of special interest. A *chain* is a multichain which has no repeated elements. A *maximal chain* is a chain that is not properly contained in another chain, and a chain $x_1 < x_2 < \cdots < x_r$ of P is said to be *saturated* if it is a maximal chain in the closed interval $[x_1, x_r]$. A covering relation could also be defined as a two-element saturated chain. A three-element saturated chain is called a *link*. The sets of all chains of the various kinds described above are denoted as follows:

Notation	Set
$\mathrm{M}(P)$	multichains of P
$\mathrm{C}(P)$	chains of P
$\mathrm{Max}(P)$	maximal chains of P
$\mathrm{Cov}(P)$	covering relations of P
$\mathrm{Link}(P)$	links of P

The largest cardinality of an element of $\mathrm{Max}(P)$ is called the *rank* of P, denoted $r(P)$. The posets we will consider have the property that every maximal chain has $r(P)$ elements.

A *labelling* of a poset is a map $\ell: P \rightarrow R$ into a ring R of *labels*. In other words, we attach a label to every element of P, and we require that labels can be multiplied and added. The labelled poset we will use for SU(3) is the following:



The labels above are in the ring of polynomials in the variables $\alpha_1, \alpha_1^{-1}, \alpha_2, \alpha_2^{-1}, A_1$, and A_2.

Given a multichain $\mathbf{x} = (x_1 \leqslant \cdots \leqslant x_r) \in \mathrm{M}(P)$ in a labelled poset P, we define the label of $\mathbf{x}$ to be $\ell(\mathbf{x}) = \Pi_{i=1}^r \ell(x_i)$. The label of the emply multichain is 1. We will show that the character generator of SU(3) is given by $\Sigma_{\mathbf{x}\in M(P)} \ell(\mathbf{x})$, where P is the labelled poset drawn above; moreover, the multichains of P furnish labels of the weight vectors of the irreps of SU(3). More generally, we will associate a labelled poset, called a *fundamental poset* with each group U(n), SU(n), and Sp($2n$, $\mathbb{C}$), such that the corresponding formula holds. The notation we will employ is the following:

Group	Rank	Fundamental poset
SU(n)	$n-1$	$\mathbf{A}(n-1)$
U(n)	n	$\mathbf{P}(n)$
Sp($2n$, $\mathbb{C}$)	n	$\mathbf{C}(n)$

SU(n)

The elements of $\mathbf{A}(n-1)$ are the strictly increasing sequences $b_1 < b_2 < \ldots < b_l$, such that $1 \leqslant b_1 < \cdots < b_l \leqslant n$ and $1 \leqslant l < n$. The partial order on $\mathbf{A}(n-1)$ is given as follows: $(b_1 < b_2 < \cdots < b_l) \leqslant (c_1 < \cdots < c_m)$ if and only if

(1) $l \geqslant m$,

(2) $b_i \leqslant c_i$ for every $i \leqslant m$.

Thus "smaller" in this poset means "longer sequence but smaller entries." The labels we attach to the element of P are given by

$$\ell(b_1 < \cdots < b_l) = A_l \prod_{i=1}^{l} f(b_i),$$

where

$$\begin{aligned} f(k) &= \alpha_1, \quad \text{if } k = 1, \\ &= \alpha_k \alpha_{k-1}^{-1} \quad \text{if } 1 < k < n, \\ &= \alpha_{n-1}^{-1} \quad \text{if } k = n. \end{aligned}$$

The element of $\mathbf{A}(n-1)$ may appear as more familiar objects if one writes them as columns

$$\begin{array}{|c|}\hline b_1 \\ \hline b_2 \\ \hline \vdots \\ \hline b_l \\ \hline \end{array}$$

rather than rows. Later we will see that the elements of $\mathbf{A}(n-1)$ are column Young tableaux.

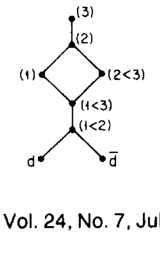
U(n)

We included this example to show how to deal with reductive Lie groups that are not semisimple. The fundamental poset $\mathbf{P}(n)$ is obtained from $\mathbf{A}(n-1)$ by adjoining two new elements and relabelling. As a poset, $\mathbf{P}(n) = \mathbf{A}(n-1) \cup \{d, \bar{d}\}$ where

(1) for every $B \in \mathbf{A}(n-1)$, $B \geqslant d, \bar{d}$,

(2) d and $\bar{d}$ are not comparable.

For example, here is $\mathbf{P}(3)$:



We label $\mathbf{P}(n)$ as follows:

$$(1)\ \ell(b_1 < \cdots < b_l) = \prod_{i=1}^{l} (A_i \alpha_{b_i}),$$

$$(2)\ l(d) = \prod_{i=1}^{n} (A_i \alpha_i),$$

$$(3)\ l(\bar{d}) = \prod_{i=1}^{n} (A_i \alpha_i)^{-1} = l(d)^{-1}.$$

Note that d acts like the sequence $(1 < \ldots < n)$ while $\bar{d}$ acts like the "inverse" of d. In fact, d corresponds to the one-dimensional irrep $\mathrm{Gl}(n, \mathbb{C}) \xrightarrow{\det} \mathbb{C}$, while $\bar{d}$ corresponds to $\det^{-1}$.

Sp($2n$, $\mathbb{C}$)

As a poset $\mathbf{C}(n)$ is a subposet of $\mathbf{A}(2n-1)$, corresponding to the fact that Sp($2n$, $\mathbb{C}$) is a subgroup of Sl($2n$, $\mathbb{C}$):

$$\mathbf{C}(n) = \{B \in \mathbf{A}(2n-1) \mid B \geqslant (1 < 3 < \cdots < (2n-1))\}.$$

The labelling of $\mathbf{C}(n)$ is given by

$$\ell(a_1 < \cdots < a_l) = A_l \prod_{i=1}^{l} g(a_i),$$

where

$$\begin{aligned} g(k) &= \alpha_j \alpha_{j-1}^{-1} \quad \text{if } k = 2j-1, \\ &= \alpha_j^{-1} \alpha_{j-1} \quad \text{if } k = 2j, \end{aligned}$$

and where $\alpha_0 = 1$.

Our main result, which will be proved in Sec. V, is the following:

Theorem 1: If P is one of the labelled posets defined above and if G is the corresponding group, then the character generator of G is given by

$$F_G(\alpha; A) = \sum_{\mathbf{x}\in \mathbf{M}(P)} \ell(\mathbf{x}).$$

IV. LINKABILITY

The concept of linkability, introduced by Gessel,[10] is a convenient tool for computing sums of the kind in Theorem 1. In this section we define this concept and exhibit linkings of the posets defined in Sec. II.

Let P be a finite, labelled poset. It is convenient for algorithmic purposes to introduce two additional elements to P, denoted $\hat{0}$ and $\hat{1}$, such that for every $p \in P$ we have $\hat{0} < p < \hat{1}$. We write $\hat{P}$ for $P \cup \{\hat{0}, \hat{1}\}$. We extend the labelling to $\hat{P}$ by defining $\ell(\hat{0})$ and $\hat{\ell}(1)$ to be 0.

Definition: A *linking* of P is a partition of $\mathrm{Link}(\hat{P})$ into two disjoint subsets $\mathrm{Link}^+(\hat{P})$ and $\mathrm{Link}^-(\hat{P})$ such that, for every pair $x < y$ in $\hat{P}$, there exists a *unique* maximal chain $(x_0 < x_1 < \cdots < x_n) \in \mathrm{Max}([x, y])$, every link of which is in $\mathrm{Link}^+(\hat{P})$.

We call $\mathrm{Link}^+(\hat{P})$ the set of *ascending links* of $\hat{P}$ (for reasons that will be made clear later), the remaining links being the *descending links*. A saturated chain $\mathbf{x} = (x_0 < x_1 < \cdots < x_n)$ is said to be *ascending* if all its links are in $\mathrm{Link}^+(\hat{P})$. Thus a linking of P is a choice of a subset $\mathrm{Link}^+(\hat{P}) \subseteq \mathrm{Link}(\hat{P})$ such that there is a unique ascending chain from any $x \in \hat{P}$ to any $y > x$. More generally, the *descent set* of $\mathbf{x}$, denoted $\mathrm{D}(\mathbf{x})$, is the set

$\mathrm{D}(\mathbf{x}) = \{x_i \mid 0 < i < n \quad \text{and} \quad (x_{i-1} < x_i < x_{i+1}) \in \mathrm{Link}^+(\hat{P})\}$.

The importance of linkability for us is the following:

Theorem 2: Let P be a labelled, linked poset. Then

$$\sum_{\mathbf{x}\in \mathbf{M}(P)} \ell(\mathbf{x}) = \sum_{\mathbf{y}\in \mathrm{Max}(\hat{P})} \frac{\Pi_{y_i\in \mathrm{D}(\mathbf{y})}\,\ell(y_i)}{\Pi_i\,[1-\ell(y_i)]}.$$

Note that every term in the right-hand side above is a product of labels and of geometric series in a label. Thus *no* cancellation is involved in the evaluation of this expression. This is a key requirement of a "good" expression for a character generator. The formulas of Stanley[7] and King[4] coincide with formula (3.1) for the cases $P = \mathbf{A}(n-1)$ and $P = \mathbf{C}(n)$, respectively.

Proof: The *support* of a multichain is the set of elements that occur at least once. This defines a map Supp: $\mathrm{M}(\hat{P}) \to \mathrm{C}(\hat{P})$. It is an easy exercise to show that for a maximal chain $\mathbf{y}$, the generating function which enumerates the multichains whose support lies between $\mathrm{D}(\mathbf{y})$ and $\mathbf{y}$, i.e.,

$$\sum_{\substack{\mathbf{x}\in \mathrm{M}(\hat{P})\\ \mathrm{D}(\mathbf{y})\subseteq \mathrm{Supp}(\mathbf{x})\subseteq \mathbf{y}}} \ell(\mathbf{x})$$

coincides with the product

$$\prod_{y_i\in \mathrm{D}(\mathbf{y})} \left(\frac{\ell(y_i)}{1-\ell(y_i)}\right) \prod_{y_i\notin \mathrm{D}(\mathbf{y})} \left(\frac{1}{1-\ell(y_i)}\right).$$

Thus we need only show that every chain $\mathbf{x}\in \mathrm{C}(\hat{P})$ occurs exactly once between some maximal chain $\mathbf{y}$ and its corresponding descent set $\mathrm{D}(\mathbf{y})$.

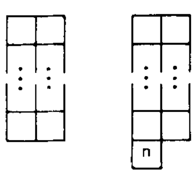
Accordingly let $\mathbf{x} = (x_1 < \cdots < x_n)$ be any chain of $\hat{P}$. By definition of a linking, there are unique ascending chains from $\hat{0}$ to x_1, from x_1 to x_2, ..., from x_n to $\hat{1}$. Concatenating these together yields a maximal chain $\mathbf{y}$ whose descents occur in a subset of $\{x_1, x_2, ..., x_n\}$. Thus there exists a maximal chain $\mathbf{y}$ such that $\mathrm{D}(\mathbf{y}) \subseteq \mathbf{x} \subseteq \mathbf{y}$.

Conversely, let $\mathbf{z}$ be any maximal chain such that $\mathrm{D}(\mathbf{z}) \subseteq \mathbf{x} \subseteq \mathbf{z}$. Since $\mathrm{D}(\mathbf{z}) \subseteq \mathbf{x}$, the restrictions of $\mathbf{z}$ to the intervals $[\hat{0}, x_1]$, $[x_1, x_2]$, ..., $[x_n, \hat{1}]$ are all ascending chains. By the uniqueness of ascending chains, we have that $\mathbf{z}$ coincides with the maximal chain $\mathbf{y}$ constructed above. QED

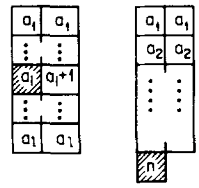
It remains to find linkings for the posets $\mathbf{A}(n-1)$, $\mathbf{P}(n)$, and $\mathbf{C}(n)$. These linkings are derived from a method due to Stanley[11] that applies to any finite distributive lattice.

SU(n); A(n − 1)

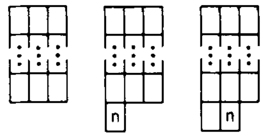
The poset $\hat{\mathbf{A}}(n-1)$ has two new elements $\hat{0}$, $\hat{1}$ which will be identified with the sequence $(1 < 2 < \cdots < n)$ and the empty sequence, respectively. The elements of Co-v($\hat{\mathbf{A}}(n-1)$) have two possible shapes:



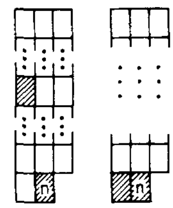
(In the second possible shape above, the last column could be empty.) In both types of covering relation exactly one row has an unequal pair of entries. We indicate the changed entry with shading as follows:



Now the elements $\mathrm{Link}(\hat{\mathbf{A}}(n-1))$ have three possible shapes:

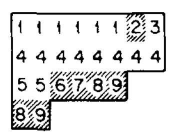


The elements of $\mathrm{Link}^+(\hat{\mathbf{A}}(n-1))$ are the ones for which the shaded (changing) entries are on the same level or ascend from left to right. For example, a link having the third shape above may be either an ascent or descent:



descent ascent

The uniqueness of the ascending chain is visually obvious as this example of a saturated chain in $\mathbf{A}(8)$ illustrates:



U(n); P(n)

The only distinction between $\hat{\mathbf{A}}(n-1)$ and $\hat{\mathbf{P}}(n)$ is the "insertion" of d and $\bar{d}$ between $\hat{0}$ and $(1 < 2 < \cdots < n-1)$. Thus there are four elements of $\mathrm{Link}(\hat{\mathbf{P}}(n))$ which have no corresponding elements in $\mathrm{Link}(\hat{\mathbf{A}}(n-1))$. All of these are defined to be ascents except for the link

$$(\hat{0} < \bar{d} < (1 < 2 < \cdots < n-1)).$$

Sp(2n, ℂ); C(n)

Since $\mathbf{C}(n)$ is an interval of $\mathbf{A}(2n-1)$, we may transfer the linking of $\mathbf{A}(2n-1)$ to that of $\mathbf{C}(n)$, by treating $\hat{0}\in\hat{\mathbf{C}}(n)$ as being the element $(1 < 3 < \cdots < 2n-1 < 2n)$ in $\mathbf{A}(2n-1)$.

V. ALGORITHIMS FOR COMPUTING THE CHARACTER GENERATORS

The "local" nature of a linking allows us to rewrite formula (3.1) in a "recursive" fashion which is the basis for the formula we derive for the character generator. Let P be a labelled, linked poset. For an element $b\in P$, we would like to restrict (3.1) to the closed interval $[\hat{0}, b]$. Unfortunately, for a maximal chain $\mathbf{y} = (y_0 < \cdots < y_m) \in \mathrm{Max}([\hat{0}, b])$, there is no way in general to determine whether $b = y_m$ ought to be considered an ascent or descent of $\mathbf{y}$.

We deal with this problem as follows. Let $(b < c) \in \mathrm{Cov}(\hat{P})$. Define

$$T(b<c) = \sum_{\mathbf{y}\in \mathrm{Max}([\hat{0},\, b])} \frac{\prod_{y_i \in \mathrm{D}(\mathbf{y},c)} \ell(y_i)}{\prod_i [1-\ell(y_i)]} \quad \text{if } b \neq \hat{0}$$

$$= 1 \quad \text{if } b = \hat{0},$$

where $\mathrm{D}(\mathbf{y},c) = \{\, y_i \mid \text{either } 0 < i < m \text{ and } (y_{i-1} < y_i < y_{i+1}) \in \mathrm{Link}^-(\hat{P}) \text{ or } i = m \text{ and } (y_{m-1} < y_m < c) \in \mathrm{Link}^-(\hat{P})\}$. The following is the recursive form of formula (3.1):

Lemma 3: Let $(c<d)\in \mathrm{Cov}(\hat{P})$ be a covering such that $c > \hat{0}$. Then

$$T(c<d) = \sum_{(b<c)\in \mathrm{Cov}(\hat{P})} \frac{g(b,c,d)}{1-\ell(c)}\, T(b<c),$$

where

$$g(b,c,d) = \begin{cases} 1 & \text{if } (b<c<d)\in \mathrm{Link}^+(\hat{P}) \\ \ell(c) & \text{if } (b<c<d)\in \mathrm{Link}^-(\hat{P}) \end{cases}.$$

Proof: This follows immediately from the observation that a sum over $\mathrm{Max}([\hat{0},c])$ for $c > \hat{0}$ is equivalent to a double sum of the form

$$\sum_{(b<c)\in \mathrm{Cov}(\hat{P})} \; \sum_{\mathbf{y}\in \mathrm{Max}([\hat{0},b])} . \qquad \text{QED}$$

In other words, $T(c<d)$ is a linear combination of the set of $T(b<c)$ such that b is covered by c, using only two possible coefficients: either $1/[1-\ell(c)]$ or $\ell(c)/[1-\ell(c)]$. Combining Lemma 3 with Theorem 2, we obtain the following recursive algorithm for the character generator:

Algorithm 4: Let P be a labelled, linked poset. Then:

(Basis) for every $(\hat{0}<b)\in \mathrm{Cov}(\hat{P})$, we have

$$T(\hat{0}<b) = 1;$$

(Recursion) for every $(c<d)\in \mathrm{Cov}(\hat{P})$ such that $c > \hat{0}$, we have

$$T(c<d) = \sum_{(b<c)\in \mathrm{Cov}(\hat{P})} \frac{g(b,c,d)}{1-\ell(c)}\, T(b<c);$$

(Conclusion)

$$\sum_{x\in \mathbf{M}(P)} \ell(x) = \sum_{(b<\hat{1})\in \mathrm{Cov}(\hat{P})} \mathrm{T}(b<\hat{1}).$$

The algorithm above furnishes an expression for the character generator in terms of subexpressions. This is the same method utilized by symbolic manipulation languages, such as MACSYMA, for displaying large formulas. This method of displaying the character generator uses one subexpression for every element $(c<d)$ of $\mathrm{Cov}(\hat{P})$ and each such subexpression has as many terms as there are elements covered by c. The total "size" of the expression thus obtained for the character generator is therefore essentially $|\mathrm{Link}(\hat{P})|$. Using the same unit of size, the original formula (3.1) for the character generator would have total size equal to $r(P)|\mathrm{Max}(\hat{P})|$. In Table I we list these numbers for the posets $\mathbf{A}(n-1)$ and $\mathbf{C}(n)$ using small values of n. For comparison we have also included the size of the Patera–Sharp formula, which is $n|\mathrm{W}(G)|$, where $\mathrm{W}(G)$ is the Weyl group of G.

There are, in addition, a few more simplifications that can be applied to Algorithm 4. The first is that it often happens that subexpressions $T(c<d)$ and $T(c<d')$ will coincide even when $d \neq d'$. This happens precisely when

$$g(b,c,d) = g(b,c,d') \quad \text{for every } b.$$

Since $g(b,c,d)$ takes only two values for a given c, this kind of coincidence often occurs. After performing this simplification, we can perform one more: If a given subexpression $T(c<d)$ (and those that coincide with it) appears just once as a subexpression of another, then substituting the formula for $T(c<d)$ where it appears will decrease the overall size of the formula for the character generator. Still other simplifications are possible, but these tend to be more *ad hoc*.

Applying Algorithm 4 and the simplifications described above, we computed the following formulas for the character generators of SU(3), SU(4), SU(5), and Sp(4, $\mathbb{C}$). For typographical simplicity we employed the substitutions

$$A = A_1, \quad \alpha = \alpha_1,$$
$$B = A_2, \quad \beta = \alpha_2,$$
$$C = A_3, \quad \lambda = \alpha_3, \quad \text{etc.}$$

SU(3) (compare with Patera and Sharp[6])

$$F_{\mathrm{SU}(3)} = \frac{1}{(1-A\beta^{-1})} \cdot \frac{1}{(1-A\alpha^{-1}\beta)} \left(\frac{1}{1-A\alpha} + \frac{B\alpha^{-1}}{1-B\alpha^{-1}} \right) \frac{1}{(1-B\alpha\beta^{-1})} \cdot \frac{1}{(1-B\beta)}.$$

TABLE I. Cardinalities of some structures related to the posets $\mathbf{A}(n)$ and $\mathbf{C}(n)$.

G	P	$r(P)$	$\lvert P\rvert$	$\lvert\mathrm{Cov}(P)\rvert$	$\lvert\mathrm{Link}(\hat{P})\rvert$	$n\lvert\mathrm{W}(G)\rvert$	$\lvert\mathrm{Max}(\hat{P})\rvert$	$r(P)\lvert\mathrm{Max}(\hat{P})\rvert$
SU(3)	$\mathbf{A}(2)$	5	6	6	8	18	2	10
SU(4)	$\mathbf{A}(3)$	9	14	18	26	96	12	98
SU(5)	$\mathbf{A}(4)$	14	30	46	76	600	286	4 004
SU(6)	$\mathbf{A}(5)$	20	62	110	208	4 320	33 592	671 840
SU(7)	$\mathbf{A}(6)$	27	126	254	544	35 280	23 178 480	625 818 960
SU(n)	$\mathbf{A}(n-1)$	$\binom{n+1}{2}-1$	2^n-2	$(n+1)2^{n-2}-2$	$\left[\binom{n+2}{2}-2\right]2^{n-3}$	$n\cdot n!$	$\dfrac{\binom{n+1}{2}!}{\prod_{i=0}^{n-1}\prod_{j=i+1}^{n}(i+j)}$	
Sp(4, $\mathbb{C}$)	$\mathbf{C}(2)$	6	9	11	16	16	5	30
Sp(6, $\mathbb{C}$)	$\mathbf{C}(3)$	12	34	59	107	144	462	5 544

SU(4) (compare with Stanley[7])

$$F_{\mathrm{SU}(4)} = \frac{1}{(1-A\lambda^{-1})}\frac{1}{(1-A\beta^{-1}\lambda)}\left\{\frac{1}{1-A\alpha^{-1}\beta}\right.$$
$$\times\left[\left(\frac{1}{1-A\alpha}+\frac{B\alpha^{-1}\beta\lambda^{-1}}{1-B\alpha^{-1}\beta\lambda^{-1}}\right)E_3\right.$$
$$\left.+\frac{E_4}{1-B\alpha^{-1}\beta\lambda^{-1}}\right]$$
$$\left.+\frac{B\beta^{-1}(E_3+E_4)}{(1-B\beta^{-1})(1-B\alpha^{-1}\beta\lambda^{-1})}\right\},$$

where

$$E_1 = \frac{1}{(1-C\beta\lambda^{-1})(1-C\lambda)},$$
$$E_2 = \frac{1}{(1-B\alpha\beta^{-1}\lambda)}\left(\frac{1}{(1-B\beta)}+\frac{C\alpha\beta^{-1}}{1-C\alpha\beta^{-1}}\right)E_1,$$
$$E_3 = \frac{E_2}{(1-B\alpha\lambda^{-1})},$$
$$E_4 = \frac{1}{(1-B\alpha^{-1}\lambda)}\left[\beta\alpha^{-1}\lambda E_2\right.$$
$$\left.+\frac{C\alpha^{-1}E_1}{(1-C\alpha^{-1})(1-C\alpha\beta^{-1})}\right].$$

SU(5)

$$F_{\mathrm{SU}(5)}$$
$$= \frac{1}{(1-A\delta^{-1})(1-A\lambda^{-1}\delta)}\left\{\frac{1}{1-A\beta^{-1}\lambda}\left[\frac{1}{1-A\alpha^{-1}\beta}\right.\right.$$
$$\times\left(\frac{E_{16}}{1-A\alpha}+\frac{B\alpha^{-1}\beta\delta^{-1}E_{16}+E_{15}}{1-B\alpha^{-1}\beta\delta^{-1}}\right)$$
$$\left.+\frac{B\beta^{-1}\lambda\delta^{-1}E_{17}+E_{18}}{1-B\beta^{-1}\lambda\delta^{-1}}\right]$$
$$\left.+\frac{B\lambda^{-1}(E_{17}+E_{18})}{(1-B\lambda^{-1})(1-B\beta^{-1}\lambda\delta^{-1})}\right\},$$

where

$$E_1 = \frac{1}{(1-D\delta)(1-D\lambda\delta^{-1})},\quad E_2 = \frac{E_1}{1-D\beta\lambda^{-1}},$$
$$E_3 = \frac{1}{1-C\beta\lambda^{-1}\delta}\left(\frac{1}{1-C\lambda}+\frac{D\beta\lambda^{-1}}{1-D\beta\lambda^{-1}}\right)E_1,$$
$$E_4 = \frac{D\alpha\beta^{-1}E_2}{1-D\alpha\beta^{-1}},\quad E_5 = \frac{E_3}{1-C\beta\delta^{-1}},$$
$$E_6 = \frac{C\alpha\beta^{-1}\delta E_3+E_4}{1-C\alpha\beta^{-1}\delta},\quad E_7 = \frac{E_5+E_6}{1-C\alpha\beta^{-1}\lambda\delta^{-1}},$$
$$E_8 = \frac{1}{1-B\alpha\beta^{-1}\lambda}\left(\frac{E_5}{1-B\beta}\right.$$
$$\left.+\frac{C\alpha\beta^{-1}\lambda\delta^{-1}E_5+E_6}{1-C\alpha\beta^{-1}\lambda\delta^{-1}}\right),$$
$$E_9 = \frac{E_7}{1-C\alpha\lambda^{-1}},$$
$$E_{10} = \frac{1}{1-C\alpha^{-1}\delta}\left[\frac{C\alpha^{-1}\delta(E_3+E_4)}{1-C\alpha\beta^{-1}\delta}\right.$$
$$\left.+\frac{D\alpha^{-1}E_2}{(1-D\alpha^{-1})(1-D\alpha\beta^{-1})}\right],$$
$$E_{11} = \frac{C\alpha^{-1}\lambda\delta^{-1}E_7+E_{10}}{1-C\alpha^{-1}\lambda\delta^{-1}},$$
$$E_{12} = \frac{1}{1-B\alpha\lambda^{-1}\delta}\left(E_8+\frac{C\alpha\lambda^{-1}E_7}{1-C\alpha\lambda^{-1}}\right),$$
$$E_{13} = \frac{B\alpha^{-1}\lambda E_8+E_{11}}{1-B\alpha^{-1}\lambda},$$
$$E_{14} = \frac{C\alpha^{-1}\beta\lambda^{-1}(E_9+E_{11})}{1-C\alpha^{-1}\beta\lambda^{-1}},$$
$$E_{15} = \frac{B\alpha^{-1}\beta\lambda^{-1}\delta E_{12}+E_{13}+E_{14}}{1-B\alpha^{-1}\beta\lambda^{-1}\delta},$$
$$E_{16} = \frac{E_{12}}{1-B\alpha\delta^{-1}},$$
$$E_{17} = \frac{E_{15}+E_{16}}{1-B\alpha^{-1}\beta\delta^{-1}},$$
$$E_{18} = \frac{1}{1-B\beta^{-1}\delta}\left[\frac{B\beta^{-1}\delta(E_{12}+E_{13}+E_{14})}{1-B\alpha^{-1}\beta\lambda^{-1}\delta}\right.$$
$$\left.+\frac{C\beta^{-1}(E_9+E_{11})}{(1-C\beta^{-1})(1-C\alpha^{-1}\beta\lambda^{-1})}\right].$$

Sp(4, $\mathbb{C}$) (compare with King[4])

$$F_{\mathrm{Sp}(4)} = \frac{1}{(1-A\alpha\beta^{-1})(1-A\alpha^{-1}\beta)}\left[\frac{B(E_1+E_2)}{(1-B)(1-B\beta^{-1})}\right.$$
$$+\frac{1}{1-A\alpha^{-1}}\left(\frac{E_1}{1-A\alpha}+\frac{E_2}{1-B\beta^{-1}}\right.$$
$$\left.\left.+\frac{B\beta^{-1}E_1}{1-B\beta^{-1}}\right)\right],$$

where

$$E_1 = \frac{1}{(1-B\beta)(1-B\alpha^2\beta^{-1})},$$
$$E_2 = \frac{B\alpha^{-2}\beta}{(1-B\beta)(1-B\alpha^{-2}\beta)}.$$

All the formulas above were computed by hand and then checked using a computer program written in Pascal VAX-11/780 for the computer, which was used to find the character generators of SU(n), for $n \leqslant 8$, and Sp($2n$, $\mathbb{C}$), for $n \leqslant 4$.

VI. YOUNG TABLEAUX

In this section we describe the relationship between the posets described in Sec. II and the classical Young tableaux. A *Young tableau* for Sl(n, $\mathbb{C}$) [or for SU(n)] is an array of positive integers of this form:

$$\begin{array}{llll} a_{11} & a_{12} & & \cdots a_{1\eta_1} \\ a_{21} & a_{22} & \cdots a_{2\eta_2} & \\ \vdots & \vdots & & \\ a_{m1} & a_{m2} & \cdots a_{m\eta_m}. & \end{array}$$

such that the following conditions hold:

(1) The number of rows is at most $n-1$, i.e., $m<n$.

(2) The lengths of the rows form a (weakly) decreasing sequence, $\eta_1 \geqslant \eta_2 \geqslant \cdots \geqslant \eta_m > 0$, called the *shape* of the tableau.

(3) The entries in any row form a (weakly) increasing sequence $0 < a_{k1} \leqslant a_{k2} \leqslant \cdots \leqslant a_{k\eta_k} \leqslant n$.

(4) The entries in any column form a strictly increasing sequence $0 < a_{1j} < a_{2j} < \cdots$.

Each Young tableau is a basis vector in the irrep whose highest weight is the n-tuple $(\lambda_1, \lambda_2, ..., \lambda_{n-1})$, where $\lambda_i = \eta_i - \eta_{i+1}$. (For the purpose of this definition we use the convention that $\eta_{m+1} = \eta_{m+2} = \cdots 0$.) We may recover $\eta_1, ..., \eta_m$ via the formula $\eta_i = \lambda_i + \lambda_{i+1} + \cdots + \lambda_{n-1}$. The *content* of a Young tableau is the n-tuple whose k th component is the number of times that k appears as an entry in the tableau. If a tableau T has content $c = (c_1, ..., c_n)$, then the weight of T (as a basis vector of an irrep) is $(c_1 - c_2, c_2 - c_3, ..., c_{n-1} - c_n)$.

To relate Young tableaux with the poset $\mathbf{A}(n-1)$, we observe that the elements of $\mathbf{A}(n-1)$ correspond precisely with the *column tableaux*, i.e., those for which η_i is either 0 or 1. Furthermore, the partial order of $\mathbf{A}(n-1)$ is defined so that the columns of an arbitrary Young tableau form a multichain of $\mathbf{A}(n-1)$, and every multichain is so obtained. It follows from this that Theorem 1 holds for the groups $\mathrm{SU}(n)$.

The irreps of $\mathrm{U}(n)$ differ from those of $\mathrm{SU}(n)$ only in the following respect: each irrep of $\mathrm{U}(n)$ may be multiplied by an arbitrary integral power of the one-dimensional irrep given by det: $\sigma \rightarrow \det(\sigma) \in \mathbb{C}$, to yield a new irrep of $\mathrm{U}(n)$. Conversely, all irreps obtained as above coincide under restriction to the subgroup $\mathrm{SU}(n)$. It follows that Theorem 1 holds for the groups $\mathrm{U}(n)$.

Finally, the concept of a tableau for the symplectic groups has been developed by King[4] and his description is equivalent to our definition of $\mathbf{C}(n)$. In retrospect, however, one could have obtained $\mathbf{C}(n)$ by applying the Baird and Biedenharn[12] correspondence between Young tableaux and Gel'fand patterns to the branching rules for $\mathrm{Sp}(2n, \mathbb{C})$ obtained by Hegerfeldt.[3] This completes our proof of Theorem 1.

ACKNOWLEDGMENTS

The author would like to thank Jiři Patera and Robert Sharp for suggesting this problem to me and for their help with my work. I would also like to thank Roland Haggkvist and Herbert Wilf for discussions concerning the algorithmic aspects of partially ordered sets.

[1]K. Baclawski, "Recursive algorithms for unitary and symplectic group representations," SIAM J. Alg. Disc. Meth., 1982, to be published.

[2]I. Gel'fand and M. Tseitlin, "Finite-dimensional representations of the group of unimodular matrices," Dokl. Akad. Nauk SSSR **71**, 825–8 (1950) (in Russian).

[3]G. Hegerfeldt, "Branching theorem for the symplectic groups," J. Math. Phys. **8**, 1195–6 (1967).

[4]R. King, "Weight multiplicities for the classical groups," in *Group Theoretical Methods in Physics*, Lecture Notes in Physics **50** (Springer-Verlag, New York, 1975), pp. 490–9; "The character generator of Sp(2k)," C. R. Math. Pep. Acad. Sci. Can. **3**, 149 (1981).

[5]K. Baclawski and J. Towber, "The shape-algebra and standard bases for G_2," submitted to Am. J. Math., 1982.

[6]J. Patera and R. Sharp, "Generating functions for characters of group representations and their applications," in *Lecture Notes in Physics* (Springer-Verlag, New York, 1979), Vol. 94, pp. 175–83.

[7]R. Stanley, "The character-generator of SU(n)," J. Math. Phys. **21**, 2321–6 (1980).

[8]K. Baclawski, "Rings with lexicographic straightening law," Adv. Math. **39**, 185–213 (1981).

[9]C. DeConcini, D. Eisenbud, and C. Procesi, "Hodge Algebras," preprint, Brandeis University, 1981.

[10]I. Gessel (unpublished).

[11]R. Stanley, "Finite lattices and Jordan-Hölder sets," Alg. Univ. **4**, 361–71 (1974).

[12]G. Baird and L. Biedenharn, "On the representations of semisimple Lie groups. II," J. Math. Phys. **4**, 1449–66 (1963).

Missing label operators in the reduction $\mathcal{O}(p)\downarrow\mathcal{O}(p-2)\times\mathcal{O}(2)$

Adam M. Bincer
Physics Department, University of Wisconsin, Madison, Wisconsin 53706

(Received 23 August 1982; accepted for publication 21 January 1983)

I consider the "missing label" problem for basis vectors of an $\mathcal{O}(p)$ representation corresponding to a group reduction chain with links $\mathcal{O}(p)\downarrow\mathcal{O}(p-2)\times\mathcal{O}(2)$. A chain with these links is required if the basis vectors are to be of definite weight. I obtain two different sets of missing label operators, which together with the Casimir operators of group and subgroups from a complete set of labeling operators whose eigenvectors provide a canonical basis in the $\mathcal{O}(p)$ representation space. The problem is solved for both the even- and odd-dimensional orthogonal groups.

PACS numbers: 02.20.Qs

1. INTRODUCTION

In two remarkable papers published in 1950, Gel'fand and Zetlin[1] explicitly constructed canonical bases for all finite-dimensional irreducible representations of the unitary and orthogonal groups. The Gel'fand–Zetlin patterns that label a basis vector consist of an array of integers or semi-integers which can be put into a one-to-one correspondence with the eigenvalues of the Casimir operators of the group and its subgroups in a canonical chain. The links in this chain consist of

$$\mathrm{U}(p)\downarrow\mathrm{U}(p-1)\times U(1) \tag{1.1}$$

in the case of the unitary groups, and of

$$\mathcal{O}(p)\downarrow\mathcal{O}(p-1) \tag{1.2}$$

in the case of the orthogonal groups.

It might be argued that the Gel'fand–Zetlin construction for the orthogonal groups suffers from two defects, as compared to their construction for the unitary groups. First, from the group-theoretical point of view, the orthogonal groups in even and odd dimensions are very different, yet both kinds are used in the chain (1.2), as opposed to only unitary groups being involved in the chain (1.1). Second, the presence of the U(1) in (1.1) means that at every link of the chain a U(1) is peeled off and, as a result, the basis vectors are eigenvectors of the weight generators. In contrast, the basis vectors in the Gel'fand–Zetlin scheme for the orthogonal groups do not have definite weight.[2]

Both of the above objections may be eliminated by constructing a basis using a chain of subgroups with links

$$\mathcal{O}(p)\downarrow\mathcal{O}(p-2)\times\mathcal{O}(2). \tag{1.3}$$

However, whereas the reduction (1.2) is multiplicity-free, the reduction (1.3) is not, and one is faced with the so-called "missing label problem," whose solution requires the determination of an appropriate number of missing label operators.

In this paper, I present two different solutions to the missing label operators problem. One solution is given by the set of operators

$$(G^{2k+1})^n_n-(G^{2k+1})^{\bar n}_{\bar n}, \tag{1.4}$$

the other solution by

$$(G^{2k+2})^n_n+(G^{2k+2})^{\bar n}_{\bar n}, \tag{1.5}$$

where G^r is essentially the rth power of the generators and where $1\leqslant k\leqslant n-1$ for the reduction $\mathcal{O}(2n+1)\downarrow\mathcal{O}(2n-1)\times\mathcal{O}(2)$, while $1\leqslant k\leqslant n-2$ for the reduction $\mathcal{O}(2n)\downarrow\mathcal{O}(2n-2)\times\mathcal{O}(2)$.

These two solutions are fully analogous to those found[3] in the corresponding problem involving the reduction

$$\mathrm{Sp}(2n)\downarrow\mathrm{Sp}(2n-2)\times\mathrm{U}(1) \tag{1.6}$$

for the symplectic groups.

This paper is organized as follows. In Sec. 2 the notation is explained, and the symmetric and antisymmetric tensors, as well as the invariants, are introduced. In Sec. 3 the missing label operators for the reduction $\mathcal{O}(2n+1)\downarrow\mathcal{O}(2n-1)\times\mathcal{O}(2)$ are explicitly constructed and their Hermiticity and invariance properties are verified. In Sec. 4 the crucial commutativity property is proved and in Sec. 5 the independence property is proved. In Sec. 6 the modifications needed for the reduction $\mathcal{O}(2n)\downarrow\mathcal{O}(2n-2)\times\mathcal{O}(2)$ are described.

2. GENERATORS AND TENSORS OF $\mathcal{O}(2n+1)$

I denote the generators of $\mathcal{O}(2n+1)$ by G^a_b with the indices ranging from $-n$ to $+n$, zero included. In the Racah[4] basis their commutation relations are

$$[G^a_b,G^c_d]=\delta^c_bG^a_d-\delta^a_dG^c_b+\delta^{\bar b}_dG^c_{\bar a}-\delta^c_{\bar a}G^{\bar b}_d, \tag{2.1}$$

where

$$\bar a\equiv -a. \tag{2.2}$$

These G's obey the antisymmetry

$$G^a_b=-G^{\bar b}_{\bar a}, \tag{2.3}$$

so that the numbers of independent generators (order of the group) in $n(2n+1)$. The Cartan subalgebra is generated by the n generators

$$G^a_a,\quad 1\leqslant a\leqslant n, \tag{2.4}$$

hence the rank of $\mathcal{O}(2n+1)$ is n and therefore n labels are needed to specify an irrep of $\mathcal{O}(2n+1)$.

I define an $\mathcal{O}(2n+1)$ tensor operator T to be an object with $(2n+1)^2$ components T^a_b obeying the commutation relations

$$[G^a_b,T^c_d]=\delta^c_bT^a_d-\delta^a_dT^c_b+\delta^{\bar b}_dT^c_{\bar a}-\delta^c_{\bar a}T^{\bar b}_d. \tag{2.5}$$

Such a tensor is reducible into a symmetric and antisymme-

 0022-2488/83/071695-05$02.50

tric part:

$$S^a_b = \tfrac{1}{2}(T^a_b + T^{\bar b}_{\bar a}) = S^{\bar b}_{\bar a}, \tag{2.6}$$

$$A^a_b = \tfrac{1}{2}(T^a_b - T^{\bar b}_{\bar a}) = -A^{\bar b}_{\bar a}, \tag{2.7}$$

and the symmetric tensor may be reduced further into a traceless part

$$S^a_b - 2(2n+1)^{-1}\mathscr{C}\delta^a_b \tag{2.8}$$

and the invariant

$$\mathscr{C} \equiv \frac{1}{2}\sum_{a=\bar n}^{n} S^a_a. \tag{2.9}$$

It is readily verified that positive integer powers of the generators, G^k, defined by

$$(G^k)^a_b = \sum_{c=n}^{n} G^a_c (G^{k-1})^c_b, \quad (G^0)^a_b = \delta^a_b \tag{2.10}$$

are tensor operators. In particular, for $k=1$ the generators themselves are seen to be the components of an antisymmetric tensor operator.

The n invariants

$$\mathscr{C}_{2n+1}(k) = \frac{1}{2}\sum_{a=\bar n}^{n} (G^{2k})^a_a, \quad 1 \leqslant k \leqslant n, \tag{2.11}$$

are the well-known Casimir operators of $\mathscr{O}(2n+1)$ whose eigenvalues may be used to label an irrep.

3. MISSING LABEL OPERATORS IN THE REDUCTION $\mathscr{O}(2n+1)\downarrow\mathscr{O}(2n-1)\times\mathscr{O}(2)$

I choose the $\mathscr{O}(2n-1)$ subalgebra to be the one obtained by omitting from the range of the indices the values n and $\bar n$, while the $\mathscr{O}(2)$ subalgebra consists of the single generator

$$G^n_n = -G^{\bar n}_{\bar n}. \tag{3.1}$$

As stated earlier, the number of labels needed to specify an irrep of $\mathscr{O}(2n+1)$ is n. The additional number of labels needed to uniquely specify a vector within an irrep is given by half the difference between order and rank, i.e., n^2. In the reduction $\mathscr{O}(2n+1)\downarrow\mathscr{O}(2n-1)\times\mathscr{O}(2)$, the subgroup $\mathscr{O}(2n-1)$ provides a number of labels equal to half the sum of its order and rank, i.e., $n(n-1)$, and $\mathscr{O}(2)$ provides one label. Thus the number of missing labels is

$$n-1. \tag{3.2}$$

It is clear that in the reduction $\mathscr{O}(2n+1)\downarrow\mathscr{O}(2n-1)\times\mathscr{O}(2)$ the basis vectors may be taken as simultaneous eigenvectors of the n Casimir operators $\mathscr{C}_{2n+1}(k)$ of $\mathscr{O}(2n+1)$, of the $n-1$ Casimir operators $\mathscr{C}_{2n-1}(k)$ of $\mathscr{O}(2n-1)$, and of G^n_n. This reduction will be the desired first link in the formation of a chain of subgroups leading to a canonical basis provided the basis vectors are simultaneously eigenvectors of an additional $n-1$ operators, the so-called missing label operators (MLO), with the following properties:

(a) Hermiticity. Choosing the MLO to be Hermitian ensures that eigenvectors belonging to different eigenvalues are orthogonal.

(b) Invariance. The MLO must be scalars under the $\mathscr{O}(2n-1)\times\mathscr{O}(2)$ subgroup.

(c) Commutativity. The MLO must commute with all the $\mathscr{C}_{2n+1}(k)$ and with each other to ensure the existence of simultaneous eigenvectors.

(d) Independence. The MLO must be polynomially independent[5] of all the $\mathscr{C}_{2n+1}(k)$, of all the $\mathscr{C}_{2n-1}(k)$, of G^n_n, and of each other.

I assert that the operators

$$A(k)^n_n, \quad 1 \leqslant k \leqslant n-1 \tag{3.3}$$

satisfy all the requirements to yield a set of $n-1$ MLO. Here the $A(k)$ are antisymmetric $\mathscr{O}(2n+1)$ tensor operators obtained by antisymmetrizing odd powers of the generators:

$$A(k)^a_b = \tfrac{1}{2}\{(G^{2k+1})^a_b - (G^{2k+1})^{\bar b}_{\bar a}\}. \tag{3.4}$$

Similarly, the operators

$$S(k)^n_n, \quad 1 \leqslant k \leqslant n-1 \tag{3.5}$$

satisfy all the requirements to yield *another* set of $n-1$ MLO. Here the $S(k)$ are symemetric $\mathscr{O}(2n+1)$ tensor operators obtained by symmetrizing even powers of the generators:

$$S(k)^a_b = \tfrac{1}{2}\{(G^{2k+2})^a_b + (G^{2k+2})^{\bar b}_{\bar a}\}. \tag{3.6}$$

The proof is as follows. In a unitary representation the generators satisfy

$$G^{a\dagger}_b = G^b_a. \tag{3.7}$$

Consequently,

$$(G^k)^{a\dagger}_b = (G^k)^b_a \tag{3.8}$$

and, therefore,

$$A(k)^{n\dagger}_n = A(k)^n_n, \quad S(k)^{n\dagger}_n = S(k)^n_n. \tag{3.9}$$

Next it is seen from Eq. (2.5) that $A(k)^n_n$ and $S(k)^n_n$ commute with G^n_n and with all G^a_b, $|a| \neq n$, $|b| \neq n$; hence they are $\mathscr{O}(2n-1)\times\mathscr{O}(2)$ scalars.

Thus the Hermiticity and invariance requirements are satisfied. The remaining requirements are more complicated and are taken up in the next two sections.

4. COMMUTATIVITY

Since the $A(k)^n_n$ and $S(k)^n_n$ are constructed out of the generators of $\mathscr{O}(2n+1)$, they commute with the $\mathscr{C}_{2n+1}(k')$, which are $\mathscr{O}(2n+1)$ scalars. The proof that

$$[A(k)^n_n, A(k')^n_n] = 0, \quad 1 \leqslant k,k' \leqslant n-1, \tag{4.1}$$

$$[S(k)^n_n, S(k')^n_n] = 0, \quad 1 \leqslant k,k' \leqslant n-1 \tag{4.2}$$

is more involved and proceeds as follows.

Define the antisymmetric $\mathscr{O}(2n+1)$ tensor operators

$$N(k)^a_b \equiv \tfrac{1}{2}\{(G^k)^a_b - G(^k)^{\bar b}_{\bar a}\}, \tag{4.3}$$

the symmetric $\mathscr{O}(2n+1)$ tensor operators

$$M(k)^a_b \equiv \tfrac{1}{2}\{(G^k)^a_b + (G^k)^{\bar b}_{\bar a}\}, \tag{4.4}$$

and the $\mathscr{O}(2n+1)$ invariants

$$D(k) \equiv \frac{1}{2}\sum_{a=\bar n}^{n} (G^k)^a_a. \tag{4.5}$$

Making use of Eqs. (2.3), (2.5), and (2.10), I have

$$
\begin{aligned}
(G^k)^{\bar b}_{\bar a} &= \sum_{c=\bar n}^{n} G^{\bar b}_{c}(G^{k-1})^{c}_{\bar a} \\
&= \sum_{c=\bar n}^{n} \{ [G^{\bar b}_{c},(G^{k-1})^{c}_{\bar a}] + (G^{k-1})^{c}_{\bar a} G^{\bar b}_{c} \} \\
&= 2n(G^{k-1})^{\bar b}_{\bar a} + (G^{k-1})^{a}_{b} \\
&\quad - 2D(k-1)\delta^a_b - \sum_{c=\bar n}^{n} (G^{k-1})^{\bar c}_{\bar a} G^{c}_{b},
\end{aligned} \tag{4.6}
$$

from which it follows by induction on k that

$$
(G^k)^{\bar b}_{\bar a} = (-)^k (G^k)^a_b + \sum_{j=0}^{k-1} g_j(k)(G^j)^a_b, \tag{4.7}
$$

where the $g_j(k)$ are some functions of the $\mathscr{O}(2n+1)$ invariants. Therefore,

$$
N(k) = \sum_{j=0}^{k} n_j(k) G^j, \quad M(k) = \sum_{j=0}^{k} m_j(k) G^j, \tag{4.8}
$$

where $n_j(k)$ and $m_j(k)$ are some functions of the $\mathscr{O}(2n+1)$ invariants. Noting that

$$
G^k G^{k'} = G^{k+k'} = G^{k'} G^k, \tag{4.9}
$$

I conclude that

$$
(XY)^a_b = (YX)^a_b, \tag{4.10}
$$

where X, Y are any linear combinations of powers of the generators [such as $N(k)$, or $M(k)$, for example].

Now let A be an arbitrary $\mathscr{O}(2n+1)$ antisymmetric tensor operator. Then

$$
\begin{aligned}
2[N(k)^n_n, A^n_n] &= \sum_{r=1}^{k} \sum_{a=\bar n}^{n} \sum_{b=\bar n}^{n} \{ (G^{k-r})^n_a [G^a_b, A^n_n](G^{r-1})^b_n - (G^{k-r})^{\bar n}_a [G^a_b, A^n_n](G^{r-1})^b_{\bar n}) \} \\
&= \sum_{r=1}^{k} \{ (G^{k-r}A)^n_n (G^{r-1})^n_n - (G^{k-r})^n_n (AG^{r-1})^n_n + (G^{k-r})^{\bar n}_n (AG^{r-1})^n_{\bar n} - (G^{k-r}A)^{\bar n}_n (G^{r-1})^n_{\bar n} \\
&\quad + (G^{k-r})^n_{\bar n} (AG^{r-1})^{\bar n}_n - (G^{k-r}A)^n_{\bar n} (G^{r-1})^{\bar n}_n + (G^{k-r}A)^{\bar n}_{\bar n} (G^{r-1})^{\bar n}_{\bar n} - (G^{k-r})^{\bar n}_{\bar n} (AG^{r-1})^{\bar n}_{\bar n} \} \\
&= \sum_{r=1}^{k} \{ [(G^{k-r}A)^n_n,(G^{r-1})^n_n] + [(G^{k-r}A)^{\bar n}_{\bar n},(G^{r-1})^{\bar n}_{\bar n}] + [(G^{r-1})^{\bar n}_n,(AG^{k-r})^n_{\bar n}] + [(G^{r-1})^n_{\bar n},(AG^{k-r})^{\bar n}_n] \},
\end{aligned} \tag{4.11}
$$

where the last step involves the use of Eq. (4.10) and the observation that $k-r$ and $r-1$ may be interchanged whenever convenient. If I now denote by $\widetilde{A}$ the antisymmetric and by $\widetilde{S}$ the symmetric part of the tensor $G^{k-r}A$, I arrive at the final result

$$
[N(k)^n_n, A^n_n] = \sum_{r=1}^{k} \{ [\widetilde{A}^n_n, N(r-1)^n_n] + [\widetilde{S}^n_n, M(r-1)^n_n] + \tfrac{1}{2}[M(r-1)^{\bar n}_n, \widetilde{S}^n_{\bar n}] + \tfrac{1}{2}[M(r-1)^n_{\bar n}, \widetilde{S}^{\bar n}_n] \}. \tag{4.12}
$$

By the same procedure, I get

$$
[M(k)^n_n, S^n_n] = \sum_{r=1}^{k} \{ [\hat{S}^n_n, M(r-1)^n_n] + [\hat{A}^n_n, N(r-1)^n_n] + \tfrac{1}{2}[\hat{S}^n_{\bar n}, M(r-1)^{\bar n}_n] + \tfrac{1}{2}[\hat{S}^{\bar n}_n, M(r-1)^n_{\bar n}] \}, \tag{4.13}
$$

where S is an arbitrary $\mathscr{O}(2n+1)$ symmetric tensor operator, and $\hat{S}$ is the symmetric, $\hat{A}$ the antisymmetric part of the tensor $G^{k-r}S$. Lastly,

$$
[M(k)^n_{\bar n}, S^{\bar n}_n] + [M(k)^{\bar n}_n, S^n_{\bar n}] = 4 \sum_{r=1}^{k} \{ [\hat{S}^n_n, M(r-1)^n_n] + [N(r-1)^n_n, \hat{A}^n_n] \}. \tag{4.14}
$$

Noting that (here I denote the invariant unit tensor $I^a_b = \delta^a_b$) $N(0) = 0$, $N(1) = G$, $M(0) = I$, $M(1) = 0$, it now follows from Eqs. (4.12), (4.13), and (4.14) by induction on k that for any $k \geqslant 0$,

$$
[N(k)^n_n, A^n_n] = 0, \tag{4.15}
$$

$$
[M(k)^n_n, S^n_n] = 0, \tag{4.16}
$$

$$
[M(k)^n_{\bar n}, S^{\bar n}_n] + [M(k)^{\bar n}_n, S^n_{\bar n}] = 0, \tag{4.17}
$$

where A is an arbitrary antisymmetric, S an arbitrary symmetric, $\mathscr{O}(2n+1)$ tensor operator.

Noting that the various operators defined previously are related to the ones defined in this section by

$$
\begin{aligned}
&A(k) = N(2k+1), \quad S(k) = M(2k+2), \\
&\mathscr{C}_{2n+1}(k) = D(2k),
\end{aligned} \tag{4.18}
$$

it is seen that the desired Eqs. (4.1) and (4.2) are special cases of Eqs. (4.15) and (4.16). This completes the proof of commutativity.

To close this section I note that, in general,

$$
[A(k)^n_n, S(k')^n_n] \neq 0 \tag{4.19}
$$

and, therefore, the MLO must be formed from *either* the symmetric *or* the antisymmetric tensors, but not both.

5. INDEPENDENCE

It was shown by Green[6] and Nwachuku and Rashid[7] that the generators of $\mathscr{O}(2n+1)$ satisfy a characteristic polynomial identity of degree $2n+1$.[8] It follows that G^{2n+1} is not polynomially independent, but may be expressed in terms of the lower powers and the invariant unit tensor I, and that the

$$
G^k, \quad 1 \leqslant k \leqslant 2n \tag{5.1}
$$

are all polynomially independent. Thus the $2n$ tensors of Eq. (5.1) provide a basis for all $\mathscr{O}(2n+1)$ tensors formed out of the generators.

Clearly, these powers of the generators may be used to form $2n$ symmetric and $2n$ antisymmetric tensors, the $M(k)$ and $N(k)$ of the previous section. But it follows from Eq. (4.7) that

$$n_k(k) = \tfrac{1}{2}[1 - (-)^k], \quad m_k(k) = \tfrac{1}{2}[1 + (-)^k], \tag{5.2}$$

and therefore the $N(k)$ for even k, the $M(k)$ and $D(k)$ for odd k, are not independent but can be expressed in terms of operators of lower degree. Thus there are n polynomially independent antisymmetric tensors given by

$$A(k), \quad 0 \leqslant k \leqslant n-1, \tag{5.3}$$

n polynomially independent symmetric tensors given by

$$S(k), \quad 0 \leqslant k \leqslant n-1, \tag{5.4}$$

and n polynomially independent invariants given by

$$\mathscr{C}_{2n+1}(k), \quad 1 \leqslant k \leqslant n. \tag{5.5}$$

Now the independence of two tensors means independence of their corresponding components. Noting that

$$A(0)^n_n = G^n_n, \tag{5.6}$$

I conclude that the $n-1$ entities

$$A(k)^n_n, \quad 1 \leqslant k \leqslant n-1, \tag{5.7}$$

and the n entities

$$S(k)^n_n, \quad 0 \leqslant k \leqslant n-1, \tag{5.8}$$

are independent of each other and of G^n_n.

As far as dependence on $\mathscr{C}_{2n+1}(k)$ and $\mathscr{C}_{2n-1}(k)$ is concerned, I note that the $A(k)^n_n$ are components of *antisymmetric* tensors, whereas the Casimirs are traces of *symmetric* tensors; hence the $A(k)^n_n$ cannot possibly the expressed as a polynomial in these Casimirs. This completes the proof that the $n-1$ antisymmetric entities given by Eq. (5.7) or (3.3) are satisfactory candidates for MLO.

The $S(k)^n_n$, being symmetric, could possibly be dependent on the Casimirs. In fact, it is easy to show that

$$2S(0)^n_n = \mathscr{C}_{2n+1}(1) - \mathscr{C}_{2n-1}(1) + G^n_n G^n_n, \tag{5.9}$$

i.e., that $S(0)^n_n$ *is* polynomially dependent on G^n_n and the quadratic Casimirs of group and subgroup. This, however, is the only such dependence, as can be seen by performing similar manipulations on the Casimirs of higher degree. Thus, at the quartic level, one has

$$\begin{aligned} 8S(1)^n_n = {} & 2\mathscr{C}_{2n+1}(2) - 2\mathscr{C}_{2n-1}(2) - (G^n_n)^4 \\ & + [\mathscr{C}_{2n+1}(1) - \mathscr{C}_{2n-1}(1)] \\ & \times [\mathscr{C}_{2n+1}(1) - \mathscr{C}_{2n-1}(1) - 2G^n_n G^n_n] \\ & + 8A(1)^n_n G^n_n + 4S(0)^n_{\bar{n}} S(0)^{\bar{n}}_n \\ & + \text{lower degree terms} \end{aligned} \tag{5.10}$$

and the presence of the new entity $S(0)^n_{\bar{n}} S(0)^{\bar{n}}_n$ prevents $S(1)^n_n$ from being dependent.

Hence the $n-1$ symmetric entities given by Eq. (3.5) [or Eq. (5.8) with $k=0$ omitted] are satisfactory candidates for MLO.

6. MISSING LABEL OPERATORS IN THE REDUCTION $\mathscr{O}(2n)\downarrow\mathscr{O}(2n-2)\times\mathscr{O}(2)$

Two modifications are necessary to convert the discussion of orthogonal groups in odd dimensions to that in even dimensions. First, in the Racah basis, $\mathscr{O}(2n+1)$ becomes $\mathscr{O}(2n)$ by simply *excluding zero* from the range of the indices. All the concepts discussed in the previous sections apply to the present case once the above change in the range of the indices is made. It follows, in particular, that the order of $\mathscr{O}(2n)$ is $n(2n-1)$, the rank is still n, and the number of missing labels in the reduction $\mathscr{O}(2n)\downarrow\mathscr{O}(2n-2)\times\mathscr{O}(2)$ is

$$n-2. \tag{6.1}$$

Repeating all the arguments of the previous sections leads to the conclusion that the MLO may be taken to be either

$$A(k)^n_n, \quad 1 \leqslant k \leqslant n-1 \tag{6.2}$$

or

$$S(k)^n_n, \quad 1 \leqslant k \leqslant n-2, \tag{6.3}$$

where the upper limit on k in Eq. (6.3) reflects the fact that the characteristic polynomial identity[6,7] in $\mathscr{O}(2n)$ is of degree $2n$ and, therefore, $S(n-1)$ is *not* a polynomially independent tensor.

The second modification has to do with the fact that the complete set of polynomially independent Casimir operators for $\mathscr{O}(2n)$ is not

$$\mathscr{C}_{2n}(k), \quad 1 \leqslant k \leqslant n, \tag{6.4}$$

[which is the analog of Eq. (5.5)] but is instead

$$E_{2n} \text{ and } \mathscr{C}_{2n}(k), \quad 1 \leqslant k \leqslant n-1. \tag{6.5}$$

Here E_{2n} is given by

$$E_{2n} = \sum_{i=1}^{n} \sum_{a_i, b_i = \bar{n}}^{n} \epsilon_{a_1\bar{b}_1 a_2\bar{b}_2 \cdots a_n\bar{b}_n} G^{a_1}_{b_1} G^{a_2}_{b_2} \cdots G^{a_n}_{b_n} \tag{6.6}$$

with ϵ the invariant antisymmetric tensor in $2n$ dimensions.

Consequently, the previous discussion of polynomial independence must be modified by the replacement of $\mathscr{C}_{2n}(n)$ and $\mathscr{C}_{2n-2}(n-1)$ by E_{2n} and E_{2n-2}, respectively. Since the entities given by Eqs. (6.2) and (6.3) were constructed without the ϵ tensor, they cannot depend on E_{2n} or E_{2n-2} linearly. They cannot depend on E^2_{2n} or E^2_{2n-2} either because E^2_{2p} is expressible as a polynomial in the $\mathscr{C}_{2p}(k)$, $1 \leqslant k \leqslant p$ [note that $\mathscr{C}_{2p}(p)$ is *included*]; hence the original proof of independence applies.

This leaves as the only candidate for dependency $E_{2n}E_{2n-2}$, which is of degree $2n-1$ in the generators. Since $A(k)^n_n$ is of degree $2k+1$ and $S(k)^n_n$ of degree $2k+2$ in the generators, all of the entities given by Eq. (6.2) with $k=n-1$ omitted and all of the entities given by Eq. (6.3) are of too low a degree to depend on $E_{2n}E_{2n-2}$. This completes the proof that the $n-2$ antisymmetric entities of Eq. (6.2) with $k=n-1$ omitted, or the $n-2$ symmetric entities of Eq. (6.3), are satisfactory sets of MLO.

It is perhaps worth noting that consistency of the formalism demands that $A(n-1)^n_n$ be dependent on $E_{2n}E_{2n-2}$ lest the set of antisymmetric MLO be too large. Indeed, explicit calculation for $n=3$ gives

$$\begin{aligned} 2A(2)^3_3 = {} & \frac{1}{4!} E_6 E_4 + 2A(1)^3_3 \mathscr{C}_6(1) \\ & + G^3_3\{\mathscr{C}_6(2) - [\mathscr{C}_6(1)]^2\} \\ & + \text{lower degree terms.} \end{aligned} \tag{6.7}$$

ACKNOWLEDGMENTS

This research was supported in part by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation, and in part by the Department of Energy under Contract DE-AC02-76ER00881.

[1] I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk SSSR **71**, 825 (1950); **71** 1017 (1950).

[2] This distinction between the Gel'fand–Zetlin treatments of the unitary and orthogonal groups was emphasized to me by J. Patera (private communciation).

[3] A. M. Bincer, J. Math. Phys. **21**, 671 (1980).

[4] G. Racah, Springer Tracts Modern Phys. **37**, 28 (1965), Eq. (74). The relation between Racah's and my notation is $X_{ik} = G^{i}_{k}$.

[5] The statement "X is polynomially independent of all X_i, $i = 1,2,3,\cdots$" means that X cannot be expressed as a polynomial in the X_i.

[6] H. S. Green, J. Math. Phys. **12**, 2106 (1971).

[7] C. O. Nwachuku and M. A. Rashid, J. Phys. A **11**, 457 (1978).

[8] This is true for the most general irrep. For certain special cases, the degree of the characteristic polynomial identity can be lower.[6,7]

Construction of N-dimensional indecomposable representations for scale and special conformal transformations

Meiun Shintani
Research Institute for Theoretical Physics, Hiroshima University, Takehara, Hiroshima 725, Japan

(Received 27 April 1982; accepted for publication 6 August 1982)

The theories of N-dimensional indecomposable representations for the group of semidirect products of D (dilatation group) and K (special conformal group) are investigated through the studies on space-time inversions.

PACS numbers: 02.20.Qs

1. INTRODUCTION

The N-dimensional representations of the scale transformations first appeared in Ref. 1 in connection with Wilson's short distance expansion, and later have been studied in detail by some authors[2,3] and applied to explain the experiments observed in the deep Euclidean region. Among other, Dell'Antonio[2] used them to show that the presence of logarithmic singularities in the operator product expansion is not inconsistent with the exact scale invariance. His result was incorporated into the so called quasicanonical field theory by Brandt *et al.*[3,4] to understand the (exact or appropriate) canonical Bjorken scaling. In spite of the extensive uses of the N-dimensional representations, the mathematical aspects of them have not been studied so far. Among other things, to construct the theories of the N-dimensional indecomposable representation (NIR for short) is of fundamental significance, and this becomes the main purpose of the present paper.

In a previous article[5] we investigated the space-time inversions as an origin of scale and conformal transformations, and then clarified the relationship between the two. Group theoretical studies on the space-time inversions enabled us to yield a unified way of the description to deal with them. In fact on the basis of some rules we deduce the one-dimensional (irreducible) representation with respect to the inversions for a certain field, and using it we then constructed the representation theories for both transformations. Herein it is demonstrated that the method (reviewed in Sec. 2) is a simple and powerful one which also applies to the case when any element of a set of local fields $\{\phi_i(x); i=1,\ldots,N, N\geqslant 2\}$ mixes nontrivially under the space-time inversion, i.e., the set forms a N-dimensional indecomposable multiplet (NIM for short). Following the unified description of the scale and conformal transformations together with some theorems in the matrix theory,[6] we actually construct the (triangular) matrix representations for both of them. It is pointed out that the general matrix representations of the space-time inversions inevitably lead to the coordinate-dependent scale transformations, the appearance of which is an aspect peculiar to our theory. Only one choice of the free parameters appearing in the matrix can give rise to the "global" scale transformations, which turns out to be of the forms seen in the Dell'Atonio's paper.[2]

In Sec. 2 the theories of the one-dimensional (irreducible) representation for the inversions, scale and conformal transformations are reviewed. In Sec. 3 along the line illustrated in Sec. 2 we construct the theories of the N-dimensional indecomposable representation for those transformations. In Sec. 4 we mention the relationship between the NIR and the "preferred" field in the theory of nonlinear realization for the conformal group.[4,7]

2. THE THEORIES OF IRREDUCIBLE REPRESENTATION ($N=1$)

This section is for the most part devoted to a review of Ref. 5, but is made fresh with an introduction of the "preferred" field $\sigma(x)$.

A. Space-time inversion

The space-time inversion $\mathscr{I}_a$ is known as a discontinuous transformation defined by

$$\mathscr{I}_a : x^\mu \to y^\mu = a x^\mu / x^2, \tag{2.1}$$

where "a" is a real and nonzero dimensionless parameter, and its magnitude is termed a "radius" of the inversion. Hereafter we take $a>0$ for convenience. In order to construct the representation theory for the dynamical variables with respect to (2.1), we employ the following two rules: (i) $\dim\hat{F}(y) = -\dim F(x)$ for physical quantity $F(x)$ and its inverted quantity $\hat{F}(y)$ defined by $\hat{F}(y) \equiv \mathscr{I}_a(F(x))$, and (ii) $(\mathscr{I}_a)^2 = I$ (I identity operator). On the basis of these two rules we shall derive, for a local scalar field $\phi(x)$ with the mass dimension l, the transformation property in such a manner that they are homomorphic to the coordinate transformation (2.1). From (i) we set

$$\hat{\phi}(y) \equiv c(x^2)^l \phi(x), \tag{2.2}$$

and further

$$\hat{\hat{\phi}}(z) = c(y^2)^l \hat{\phi}(y), \tag{2.3}$$

where c is a dimensionless constant to be determined. By considering (ii), that is, $\hat{\hat{\phi}}(z) = \phi(x)$ and $z^\mu = x^\mu$, we have $c = \pm a^{-l}$, which suggests

$$\hat{\phi}(y) = \pm (x^2/a)^l \phi(x). \tag{2.4}$$

In case of $l=0$, in addition to (2.4) with $l=0$, there occurs another possibility of putting

$$\hat{\sigma}(y) = \sigma(x) + C(x^2, a), \tag{2.5}$$

where $\sigma(x)$ is termed a "preferred' field[7] and the $C(x^2,a)$ is a

 0022-2488/83/071700-05$02.50

dimensionless c-number function to be determined. In the manner similar to the above, we obtain the equation for C

$$C(x^2,a)+C(y^2,a)=0. \tag{2.6}$$

By taking into account the relation of x and y, i.e., $x^2y^2=a^2$, we deduce

$$C(x^2,a)=\sum_{m=-\infty}^{+\infty} c_m(\ln x^2/a)^{(2m+1)} \quad (m \text{ integer}), \tag{2.7}$$

with the dimensionless constants c_m's, whence under the $\mathscr{I}_a$

$$\hat{\sigma}(y)=\sigma(x)+\sum_{m=-\infty}^{+\infty} c_m(\ln x^2/a)^{(2m+1)}. \tag{2.8}$$

B. Scale and conformal transformations

Next let us obtain the transformation properties of the scalar fields under both scale and conformal transformations. Before doing this, we must explain that both scale and conformal transformations can be built of the two inversions with different radii, and the translation.

Now let us consider a composite transformation consisting of the following transformations:

(1) inversion $\mathscr{I}_a$

$$\mathscr{I}_a: x^\mu \to y^\mu = ax^\mu/x^2,$$

(2) translation $\mathscr{T}_c$ by a constant 4-vector c^μ

$$\mathscr{T}_c: y^\mu \to u^\mu = y^\mu + c^\mu,$$

and (3) inversion $\mathscr{I}_b$

$$\mathscr{I}_b: u^\mu \to z^\mu = bu^\mu/u^2.$$

The successive product of these transformations forms a composite transformation given by

$$x^\mu \to z^\mu = \lambda\Omega^{-1}(x,c/a)(x^\mu+(c^\mu/a)x^2), \tag{2.9}$$

where $\lambda=b/a$ and $\Omega(x,c)=1+2cx+c^2x^2$. Note that the transformation (2.9) is decomposed into the following transformations:

(1)′ conformal transformation $\mathscr{K}_{c/a}$ characterized by c'^μ $(=c^\mu/a)$

$$\mathscr{K}_{c/a}: x^\mu \to y'^\mu = \Omega^{-1}(x,c/a)(x^\mu+(c^\mu/a)x^2),$$

and (2)′ scale transformation $\mathscr{D}_{b/a}$ by a scale factor $\lambda=b/a$

$$\mathscr{D}_{b/a}: y'^\mu \to z^\mu = \lambda y'^\mu = \lambda\Omega^{-1}(x,c/a)[x^\mu+(c^\mu/a)x^2].$$

From the above argument it follows that (2.9) is equivalent to the expression

$$\mathscr{I}_b\mathscr{T}_c\mathscr{I}_a=\mathscr{D}_\lambda\mathscr{K}_{c/a}, \tag{2.10}$$

or for the reverse order of (1)′ and (2)′ to

$$\mathscr{I}_b\mathscr{T}_c\mathscr{I}_a=\mathscr{K}_{c/b}\mathscr{D}_\lambda. \tag{2.11}$$

It is easily verified from (2.10) and (2.11) that the set of the composite transformations (2.9), denoted by G, satisfies the group axioms (not abelian) and is equivalent to the semidirect product of D (dilatation group) and K (special conformal group), i.e.,

$$G \cong D \circledS K. \tag{2.12}$$

Here the K is an invariant subgroup of G.

Now consider the two special cases of equation (2.10):

Case (a) $c^\mu=0$

Equation (2.10) [or (2.11)] becomes

$$\mathscr{I}_b\mathscr{I}_a=\mathscr{D}_\lambda \quad (\lambda=b/a),$$

which is a pure scale transformation with a scale factor λ.

Case (b) $\lambda=1$ $(a=b)$

Equation (2.10) [or (2.11)] turns out to be

$$\mathscr{I}_a\mathscr{T}_c\mathscr{I}_a=\mathscr{K}_{c/a},$$

which is a (special) conformal transformation characterized by the parameter $c'^\mu=c^\mu/a$. We, therefore, come to know that the study of the group G allows us to treat the scale and conformal transformations in a unified manner. In the first part of this section, we derived the transformation properties for the scalar field under the inversions. Using the results, we are now in a position to obtain the transformation properties for the ϕ or the σ under G.

A scalar field $\phi(x)$ transforms as $\phi^{(1)}(y)=\pm(x^2/a)^l\phi(x)$, $\phi^{(2)}(u)=\phi^{(1)}(y)$, and $\phi^{(3)}(z)=\pm(u^2/b)^l\phi^{(2)}(u)$, respectively, under (1), (2), and (3). Hence the representation of G for $\phi(x)$

$$\phi^{(3)}(z)=\lambda^{-l}[\Omega(x,c/a)]^l\phi(x). \tag{2.13}$$

For $\sigma(x)$, it transforms as $\sigma^{(1)}(y)=\sigma(x)+\sum_{m=-\infty}^{+\infty}c_m(\ln x^2/a)^{(2m+1)}$, $\sigma^{(2)}(u)=\sigma^{(1)}(y)$, and $\sigma^{(3)}(z)=\sigma^{(2)}(u)+\sum_{m=-\infty}^{+\infty}c_m(\ln u^2/b)^{(2m+1)}$, respectively, under (1), (2), and (3). The the representation of G for $\sigma(x)$ becomes

$$\sigma^{(3)}(z)=\sigma(x)+(\ln\lambda^{-1}+\ln\Omega(x,c/a)) \times\Big\{c_0+\sum_{m=1}^{\infty}(c_m f_{2m}(x^2/a,u^2/b) + c_{-m}\,g_{2m}(x^2/a,u^2/b))\Big\}, \tag{2.14}$$

where

$$\left.\begin{aligned} g_{2m}(x,y) &\equiv (\ln x\cdot\ln y)^{1-2m}f_{2m-2}(x,y)\\ \text{and}\quad f_{2m-2}(x,y) &\equiv \sum_{k=1}^{2m}(-1)^k x^{2m-k}y^k \end{aligned}\right\}. \tag{2.15}$$

When $c_\mu=0$ in (2.14), $\sigma(x)$ is shifted by the coordinate-dependent function, that is, $\ln\lambda^{-1}\{c_0+\sum_{m=1}^{\infty}(c_m f_{2m}(x^2/a,\lambda^{-1}x^2/a)+c_{-m}\,g_{2m}(x^2/a,\lambda^{-1}x^2/a))\}$, as long as $c_m\neq 0$ for $m\neq 0$. In this case we also note that the transformation property with $c_0\neq 0$ and $c_m=0$ $(m=\pm 1,\pm 2,\cdots)$ in (2.14) is identified with that appearing in Ref. 8. The field $\sigma(x)$ is intimately related to the representation theory of the NIM for the inversion. This is discussed in the final section. The above argument can be repreated for other quantities such as line element, vector fields, spinor fields, etc.[5]

3. THE N-DIMENSIONAL INDECOMPOSABLE REPRESENTATION ($N\geqslant 2$)

Let us consider a set of the N real scalar fields $\{\phi_1,\ldots,\phi_N, N\geqslant 2\}$ with the equal mass dimension l and denote it by $\Phi(x)$. From rule (i) in Sec. 2 we may assume that each of them transforms under inversion $\mathscr{I}_a$ defined by (2.1) as

$$\phi_i(x) \to \hat{\phi}_i(x) \equiv \left(\frac{x^2}{a}\right)^l \sum_{j=1}^{N} L_{ij}(x^2,a)\phi_j(x) \quad (i=1,\dots,N), \tag{3.1}$$

where the $L(x^2,a)$ is the real and dimensionless $N \times N$ matrix whose elements may be a function of x^2 and a. Unless $L_{N\times N} = L'_{M\times M} \oplus L''_{(N-M)\times(N-M)}$ for $M<N$, Eq. (3.1) forces the members of $\Phi(x)$ to be nontrivially mixed among them through the inversion. In this case we shall say that the matrix L is a N-dimensional indecomposable representation (NIR), and the set $\Phi(x)$ forms a N-dimensional indecomposable multiplet (NIM) for the inversion. Applying the procedure outlined in Sec. 2 for the scalar field ($N=1$) to the case of the NIM($N \geqslant 2$), we shall construct the NIR for the group G as well as for the inversion. Consequently, it is seen that each element of $L(x^2,a)$ is not completely determined, and then it generates the coordinate-dependent scale transformations (a kind of local scale transformations). However, if we confine the NIR to the ones leading to the "global" scale transformations, we arrive at the unambiguous representation of $L(x^2,a)$.

A. Space-time inversion

If in Eq. (3.1) the operation of the $\mathscr{I}_a$ to the $\hat{\phi}_i(y)$ is repeated, the $\hat{\phi}_i(y)$ is transformed into the $\hat{\hat{\phi}}_i(z)$ as

$$\hat{\hat{\phi}}_i(z) \equiv \left(\frac{y^2}{a}\right)^l \sum_{j=1}^{N} L_{ij}(y^2,a)\hat{\phi}_j(y) \quad (i=1,\dots,N). \tag{3.2}$$

By rule (ii) in Sec. 2, $z=x$ and $\hat{\hat{\phi}}_i(z) = \phi_i(x)$, which implies

$$\sum_{j=1}^{N} L_{ij}(y^2,a)L_{jk}(x^2,a) = \delta_{ik} \quad (i,k=1,\dots,N). \tag{3.3}$$

From the theorem in the matrix theory[6] the L can be decomposed into the product of the two real $N \times N$ matrices as

$$(L)_{ij} = (RC)_{ij}, \tag{3.4}$$

where the R is a real orthogonal matrix and the C a triangular matrix. It is known that choosing all the diagonal elements of C being positive is always possible, and for a given L this yields a unique decomposition of (3.4). By rewriting $R^T\Phi(y)$ by $\Phi(y)$, Eq. (3.1) can be cast into the form

$$\phi_i(x) \to \hat{\phi}_i(y) \equiv \left(\frac{x^2}{a}\right)^l \sum_{j=1}^{N} C_{ij}(x^2,a)\,\phi_j(x) \quad (i=1,\dots,N), \tag{3.1'}$$

with

$$C_{ij} = 0 (1<i<j<N) \text{ and } C_{ii}>0 \quad (i=1,\dots,N). \tag{3.5}$$

Instead of (3.3), the C satisfies

$$\sum_{j=1}^{N} C_{ij}(y^2,a)C_{jk}(x^2,a) = \delta_{ik} \quad (i,k=1,\dots,N). \tag{3.3'}$$

By setting $i=k$ in (3.3') and using the properties of C_{ij} [i.e., dim $C=0$ and Eq. (3.5)], we derive

$$C_{ii} = 1, \quad (i=1,\dots,N). \tag{3.6}$$

Let us further restrict the representation of C to the ones such that for the off-diagonal elements

$$C_{ij} = C_{i+1,j+1} \quad (1<j<i<N-1). \tag{3.7}$$

In this representation, if all the C_{i1}'s ($i=2,\dots,N$) are given somehow, the matrix C is determined. To this end, we need only solve Eq. (3.3)' combined with (3.5), (3.6) and (3.7). The $(N-1)$ equations for the C_{i1}'s ($i=2,\dots,N$) can now be written down as

$$\sum_{j=1}^{N} C^{(N)}_{i-j+1,1}(y^2,a)C^{(N)}_{j,1}(x^2,a) = 0 \quad (i=2,\dots,N). \tag{3.8}$$

Here the superscript (N) of $C^{(N)}$ means that $C^{(N)}$ is the $N \times N$ matrix. The results are summarized separately for the cases of (a) $N=$ odd and (b) $N'=N+1=$ even as follows:

(a) $N=$ odd

$$C^{(N)}(x^2,a) = \begin{bmatrix} 1, & 0\dots & & 0 \\ c_2(\ln x^2/a)^{2m+1}, & 1, & 0, & \\ c_3(\ln x^2/a)^{2(2m+1)}, & c_2(\ln x^2/a)^{(2m+1)}, & 1, \quad 0. & \\ \vdots & \vdots & \ddots & \\ c_{N-1}(\ln x^2/a)^{(N-2)(2m+1)}, & & & 0 \\ c_N(\ln x^2/a)^{(N-1)(2m+1)}, & c_{N-1}(\ln x^2/a)^{(N-2)(2m+1)}, & \dots & 1 \end{bmatrix} \tag{3.9}$$

(b) $N'=$ even $=N+1$

$$C^{(N')}(x^2,a) = \left[\begin{array}{c|c} C^{(N)}(x^2,a) & 0 \\ \hline c_{N'}(\ln x^2/a)^{(2m'+1)}, c_N(\ln x^2/a)^{(N-1)(2m+1)}, c_{N-1}(\ln x^2/a)^{(N-2)(2m+1)},\dots & 1 \end{array}\right]. \tag{3.10}$$

In the above, for simplicity of the argument as a solution of Eq. (2.6) for C_{21}

$$C^{(N)}_{21}(x^2,a) = c_2(\ln x^2/a)^{(2m+1)} \quad (m \text{ integer}), \tag{3.11}$$

has been used. Each of the coefficients indexed by even numbers (i.e., $c_2, c_4, \dots, c_{N-3}, c_{N-1}, c_{N'}$), and integers m or m' cannot be determined by solving Eq. (3.8), and then may be arbitrarily chosen. On the other hand each of the coefficients c_{2k+1} ($k=1,2,\dots,[(N-1)/2]$ is expressible as a polynomial of c_{2l}'s ($1<l<k$). When $N=7$ for example, $c_3=(1/2!)c_2^{\,2}$, $c_5 = c_2(2c_4 - \frac{1}{4}c_2^{\,3})$ and $c_7 = \frac{1}{2}\{2c_2c_6 + c_4^{\,2} - c_2^{\,3}(2c_4 - \frac{1}{4}c_2^{\,3})\}$. It is noted that the power index m' of $C_{N'1}(x^2,a)$ is not made equal to m until we pass to $N''=N'+1$ (odd). This has already been reflected in Eq. (3.9).

B. Scale and conformal transformations

Along the line illustrated in Sec. 2, let us investigate the representation theory of the group $G \cong D \circledast K$. The set of scalar fields $\Phi(x)$ transforms successively as $\phi_i^{(1)}(y) = (x^2/a)^l \cdot \Sigma_{j=1}^N C_{ij}^{(N)}(x^2,a)\phi_j(x), \phi_i^{(2)}(u) = \phi_i^{(1)}(y)$ and $\phi_i^{(3)}(z) = (u^2/b)^l \cdot \Sigma_{j=1}^N C_{ij}^{(N)}(u^2,b)\phi_j^{(2)}(u)$, respectively, under (1) $\mathscr{I}_a$, (2) $\mathscr{T}_c$ and (3) $\mathscr{I}_b$. Hence the transformation property of ϕ_i under G is

$$\phi_i^{(3)}(z) = \lambda^{-l}\Omega^l(x,c/a)\sum_{k=1}^N \hat{G}_{ik}^{(N)}(x^2,a;u^2,b)\phi_k(x),$$
$$\hat{G}_{ik}^{(N)}(x^2,a;u^2,b) \equiv \sum_{j=1}^N C_{ij}^{(N)}(u^2,b)C_{jk}^{(N)}(x^2,a). \tag{3.12}$$

In order to simplify the arguments, we consider only the case of N = odd. Of course the same argument is repeated for the case of N = even. Then $\hat{G}_{i1}^{(N)}(x^2,a;u^2,b)$ for odd N is given by

$$\hat{G}_{i1}^{(N)}(x^2,a;u^2,b) = \sum_{k=1}^{i} c_{i-k+1}c_k(\ln u^2/b)^{(i-k)(2m+1)}(\ln x^2/a)^{(k-1)(2m+1)}, \tag{3.13}$$

with $c_1 = 1$. It is easily checked that the $\hat{G}^{(N)}$ is a triangular matrix satisfying the same properties, i.e., (3.5), (3.6) and (3.7), as the matrix $C^{(N)}$ does. By putting $c_\mu = 0$ in Eq. (3.12), we arrive at the transformation property of $\phi_i(x)$ under the $\mathscr{D}_\lambda$ as

$$\phi_i^{(3)}(y = \lambda x) = \lambda^{-l}\sum_{k=1}^N \hat{G}_{ik}^{(N)}(x^2,a;y^2,b)\phi_k(x), \tag{3.12'}$$

with

$$\hat{G}_{i1}^{(N)}(x^2,a;y^2,b) = \sum_{k=1}^{i} c_{i-k+1}c_k(\ln y^2/b)^{(i-k)(2m+1)}(\ln x^2/a)^{(k-1)(2m+1)}. \tag{3.13'}$$

From the above equations we find that the representation of D generally yields the coordinate-dependent scale transformations, which is unusual, though it may be interpreted as a local version of the scale transformations. The only way out of this is to put (3.13)′ equal to

$$\left.\begin{aligned}\hat{G}_{i1}^{(N)}(x^2,a;y^2,b) &= c_i(\ln y^2/b + \ln x^2/a)^{(i-1)(2m+1)}\\ &\text{or} = c_i(\ln\lambda^{-1})^{(i-1)(2m+1)} \text{ for } i = 1,\ldots,N.\end{aligned}\right\} \tag{3.14}$$

By solving one by one the above $(N-1)$ equations for m and c_i $(i = 2,\ldots,N)$, we deduce

$$m = 0 \text{ and } c_i = \frac{(c_2)^{(i-1)}}{(i-1)!} \quad (i = 2,\ldots,N), \tag{3.15}$$

which suggests

$$\left.\begin{aligned}&C_{ij}^{(N)}(x^2,a) = (c_2)^{(i-j)}(\ln x^2/a)^{(i-j)}/(i-j)! \text{ for } i>j,\\ &\text{otherwise } C_{ij}^{(N)}(x^2,a) = 0.\end{aligned}\right\} \tag{3.16}$$

Since this leads to the usual (i.e., "global") scale transformations, under the $\mathscr{D}_\lambda$, the $\Phi(x)$ transforms as

$$\Phi^{(3)}(\lambda x) = \lambda^{-l}\begin{bmatrix} 1, & 0 & \cdots & & 0\\ c_2\ln\lambda^{-1}, & 1, & 0. & & \vdots\\ \frac{(c_2\ln\lambda^{-1})^2}{2!}, & c_2\ln\lambda^{-1}, & 1. & \ddots & \vdots\\ \vdots & & \ddots & \ddots & 0\\ \frac{(c_2\ln\lambda^{-1})^{N-1}}{(N-1)!}, \cdots & & & , c_2\ln\lambda^{-1}, & 1,\end{bmatrix}\Phi(x) \tag{3.17}$$

or equivalently in a compact expression

$$\Phi^{(3)}(\lambda x) = \lambda^{-l}\cdot\exp\{c_2\ln\lambda^{-1}A^{(N)}\}\Phi(x), \tag{3.17'}$$

where the $N \times N$ matrix $A^{(N)}$ is given by

$$A^{(N)} \equiv \begin{bmatrix} 0 & \cdots & & 0\\ 1 & \ddots & & \\ 0 & \ddots & \ddots & \vdots\\ 0 & \cdots 0 & 1 & 0\end{bmatrix}. \tag{3.18}$$

Here use has been made of the nilpotency property of $A^{(N)}$, i.e.,

$$(A^{(N)})^k = 0 \text{ for } k \geqslant N \quad (k \text{ integer}). \tag{3.19}$$

The expression (3.17) with $c_2 = 1$ coincides with the one in Ref. 2 or 4. Similarly, the transformation property of Φ under the $\mathscr{K}_c$ is obtained by putting $a = b = 1$ in (3.12) together with (3.16) as

$$\Phi^{(3)}(z) = \Omega^l(x,c)\exp\{c_2\ln\Omega(x,c)\cdot A^{(N)}\}\cdot\Phi(x). \tag{3.20}$$

It is suggested that from now on one should use the compact forms (3.17)′ and (3.20) instead of the matrix expressions such as (3.17).

As was seen in the above, our method is a simple and convenient one in a sense that one may treat NIR's $(N \geqslant 2)$ as well as the irriducible representation $(N = 1)$ for both of the scale and conformal transformations one at a time.

4. CONCLUDING REMARKS

1) It is seen that in general the matrix form of the NIR of the space-time inversion does not pass to the "global" scale transformation, but with a unique choice of the free parameters the global transformation (3.17)′ can be reproduced. The same argument for the preferred field $\sigma(x)$ is repeated.

2) We mention the connection between the σ-field and the NIM. With the use of the recipe in Sec. 3 it transforms

$$\sigma(x) \to \sigma^{(3)}(\lambda x) = \sigma(x) + \ln \lambda^{-1}, \qquad (4.1)$$

under the $\mathscr{D}_\lambda$,

$$\sigma(x) \to \sigma^{(3)}(z) = \sigma(x) + \ln \Omega(x,c), \qquad (4.2)$$

under the $\mathscr{K}_c$. Here we have normalized the factor c_0 appearing in (2.8) to unity. Now define a set of new fields by

$$\phi_i(x) \equiv [(\sigma(x))^{(i-1)}/(i-1)!]\, \phi_1(x) \quad (i = 1,\ldots,N), \quad (4.3)$$

where the field $\phi_1(x)$ is an ordinary field with mass dimension l which transforms under $G \simeq K \circledS D$ as (2.13). Then, it is found[4] that with the aid of (2.13), (3.21) and (3.22) the transformation property for the set of fields $\{\phi_i(x)\}$ under $\mathscr{D}_\lambda(\mathscr{K}_c)$ reproduces (3.17) [(3.20)] for the NIM. This observation associates the quasicanonical quantum field theory developed by Brandt *et al.* with the effective Lagrangian theories[7,9] using the field $\phi_1(x)$ and the σ-field as the Goldstone boson. Further, the net effect due to the σ-field in the large N limit occurs as the factor $\exp[\sigma(x)]$ and under the scale transformation (4.1) it transforms as if it were a scalar field with mass dimension one. This suggests that one puts

$$\exp[\sigma(x)] = b\chi(x), \qquad (4.4)$$

where b is the parameter with dim $b = 1$ and the field $\chi(x)$ is an ordinary scalar field with dim $\chi(x) = -1$ but does not coincide with ϕ_1. If one uses the field $\chi(x)$ instead of the σ-field, the theory turns out to be the work of Nambu and Freund.[10]

3) As in the above the role of the NIR in the field theories may be investigated by studying that of the σ-field. The role of the σ-field has been widely disscussed.[7,8,9] Here we shall not repeat it. The thing we point out is that without artificially imposing the conditions like $\langle \phi \rangle_0 \neq 0$ the presence of the σ-field in a scale invariant theory automatically implies the spontaneous breaking of the scale transformation at the tree level. This is due to the transformation law of σ, i.e., $\langle [\sigma(x), D] \rangle_0 = i (\neq 0)$. This situation is similar to the presence of R-transformation in QED[11] which implies that the photons are the Goldstone bosons.

4) The method in this article can be applied to the construction of the NIR's for the spinors, vectors and other higher rank tensors.

[1] P. Otterson and W. Zimmermann, Comm. Math. Phys. **24**, 107 (1972).
[2] G. F. Dell'Antonio, Nuovo Cimento A **12**, 756 (1972).
[3] R. A. Brandt and Wing-chiu Ng, Nuovo Cimento A **13**, 1025 (1973).
[4] R. A. Brandt and Wing-chiu Ng, Phys. Rev. D **9**, 1025 (1974).
[5] M. Shintani, Hiroshima University Preprint RRK 80–11, 1980, submitted to Prog. Theor. Phys. (in revised form).
[6] S. Helgason, *Differential Geometry, Lie Groups and Symmetric Spaces* (Academic, New York, 1978), Chap. VI.
[7] S. Ferrara and A. F. Grillo, Nuovo Cimento Lett. **2**, 177 (1971).
[8] A. Salam and J. Strathdee, Phys. Ref. **184**, 1760 (1969).
[9] B. Zumino, *Lectures on Elementary Particles and Quantum Field Theory*, edited by S. Deser *et al.* (MIT Press, Cambridge, MA, 1970).
[10] P. G. O. Freund and Y. Nambu, Phys. Rev. **174**, 1741 (1968).
[11] R. A. Brandt and Wing-chiu Ng, Phys. Rev. D **10**, (1974), 4198.

Casimir invariants, characteristic identities, and tensor operators for "strange" superalgebras

P. D. Jarvis and M. K. Murray[a)]
Department of Physics, University of Tasmania, G. P. O. Box 252C, Hobart, Tasmania 7001, Australia

(Received 16 June 1981; accepted for publication 5 November 1982)

We define a class of (super) subalgebras of gl(m/n) realized as the set of fixed points of a (graded) endomorphism of gl(m/n). This class includes the superalgebras gp(m) and gq(m) [related to the so-called "strange" simple superalgebras p(m) and q(m)], as well as osp(m/n). General covariant, contravariant, and mixed tensor operators are defined for this class in terms of appropriate module homomorphisms. Traces of certain tensors give the usual sequence of Casimir invariants. For gp(m), these are shown to vanish identically, while for gq(m), eigenvalues of the quadratic and cubic Casimir invariants are derived in terms of highest weights and a polynomial characteristic identity is exhibited.

PACS numbers: 02.20.Qs, 02.20.Sv

I. INTRODUCTION

Lie superalgebras[1] arise in a number of physical contexts such as supersymmetry and the quantization of some classical systems. A survey of some of their areas of application can be found in Corwin, Ne'eman, and Sternberg.[2]

The definition of Lie superalgebras raises the question of defining an object called a "Lie supergroup" in analogy to the relationship between Lie algebras and Lie groups. Definitions of these and supermanifolds in general can be found in Sternberg,[3] Kostant,[4] Rogers,[5] Batchelor,[6] and other papers referred to in these.

Kac[7] has provided a classification of the finite-dimensional Lie superalgebras (see also Scheunert[8]). A Lie superalgebra $L = L_0 \oplus L_1$ is called classical if it is simple and the action of L_0 on L_1 is a completely reducible representation of the Lie algebra L_0.

Among the classical Lie superalgebras classified by Kac are $p(m)$ and $q(m)$ defined as (A, B, C, D are $m \times m$ matrices):

$$p(m) = \{(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}) | A + D^T = B - B^T = C + C^T = 0, \quad \mathrm{tr}A = 0\} \leqslant \mathrm{gl}(M/m) \tag{1}$$

and

$$q(m) = \tilde{q}(m)/\text{center of } \tilde{q}(m), \tag{2}$$

where

$$\tilde{q}(m) = \{(\begin{smallmatrix} A & B \\ B & A \end{smallmatrix}) | \mathrm{tr}B = 0\} \leqslant \mathrm{gl}(m/m). \tag{3}$$

We shall be concerned in the following with the related Lie superalgebras gp(m) and gq(m) defined as

$$\mathrm{gp}(m) = \{(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}) | A + D^T = B - B^T = C + C^T = 0\} \geqslant p(m) \tag{4}$$

and

$$\mathrm{gq}(m) = \{(\begin{smallmatrix} A & B \\ B & A \end{smallmatrix})\} \geqslant \tilde{q}(m). \tag{5}$$

Note that whereas $p(m)$ is a subalgebra of gp(m), $q(m)$ is realized as a factor algebra of $\tilde{q}(m)$ and hence not, *a priori*, a subalgebra[9] of matrices.

In Sec. II we review the basic definitions of graded vector spaces, Lie superalgebras, and (super) modules.[9] A general definition of a tensor operator is given and also the particular case of contra- and covariant vector operators. These are defined also in Cant and Hurst[10] and Hannabuss.[11] The definition is in accord with the intuitive idea of "vector" or "matrix" of operators, as originally formulated by Racah and Wigner. When a particular product of these is well defined, we can define the sequence of Casimir invariants as defined in Jarvis and Green.[12]

The class of subalgebras which are fixed points of graded automorphisms is defined in Sec. III. With some constraints on the automorphisms, these all have a well-defined product mentioned above. As gp(m) and gq(m) are in this class, this allows us to carry over the derivation of the Casimir invariants and characteristic identities following the previous treatment[12] of gl(m/n), sl(m/n), and osp(m/n).

The Casimir invariants of gp(m) are of interest in that they vanish identically. This, however, prevents the derivation of a characteristic identity. In the case of gq(m), the quadratic invariant vanishes. However, the eigenvalue of the cubic invariant is obtained explicitly in terms of the highest weight, and a characteristic identity is presented.

II. BASIC DEFINITIONS AND EXAMPLES

For completeness and convenience we present here the basic definitions and examples used throughout.

A vector space V is said to be a Z_2-graded vector space (GVS) if it is decomposed into a direct sum $V = V_0 \oplus V_1$. Here, as elsewhere, we regard 0 and 1 as elements of Z_2 and, in general, add modulo 2. An element x of V_i is said to be homogeneous of degree i, denoted $|x| = i$.

If f: $V \rightarrow W$ is a linear map between GVS's V and W, then f is said to be a homomorphism of degree k if $f(V_i) \subset W_{i+k}$ for all $i = 0$ and 1. Unspecified maps are assumed to have degree 0.

A Lie superalgebra is a GVS L with a bilinear map [ ]: $L \times L \rightarrow L$ satisfying the three properties

[a)] Present address: The Mathematical Institute, University of Oxford, 24–29 St. Giles, Oxford OX7 3LB, U.K.

 0022-2488/83/071705-06$02.50

$$[x,y] = -(-1)^{|x||y|}[y,x], \tag{6}$$

$$[x[y,z]] = [[x,y]z] + (-1)^{|x||y|}[y[x,z]], \tag{7}$$

$$[L_i,L_j] \subseteq L_{i+j} \tag{8}$$

for all x, y, z homogeneous elements of L.

The standard GVS is $\mathbb{C}^{m+n}$ with the grading given by

$$\mathbb{C}^{m+n} = \mathbb{C}^m \oplus \mathbb{C}^n \tag{9}$$

so that

$$\mathbb{C}_0^{m+n} = \mathbb{C}^m \quad \text{and} \quad \mathbb{C}_1^{m+n} = \mathbb{C}^n.$$

If V is any GVS, then End$V$ is a Lie superalgebra with grading by the degree of the map and bracket defined as

$$[f,g] = f \circ g - (-1)^{|f||g|} g \circ f \tag{10}$$

for homogeneous elements and the linear extension for nonhomogeneous elements. End$V$ is denoted gl(V) as a Lie superalgebra or gl(m/n) if $V = \mathbb{C}^{m+n}$.

We give gl(m/n) the standard basis of matrices $e^A{}_B$ defined as

$$(e^A{}_B)_X{}^Y = \delta^A_X \delta^Y_B \tag{11}$$

for $A,B = 1,...,m+n$.

The two-index notation used by Jarvis and Green[12] is defined as follows. We grade the labels A as

$$(A) = 0 \quad \text{if} \quad A = 1,...,m,$$

and (12)

$$(A) = 1 \quad \text{if} \quad A = m+1,...,m+n;$$

then

$$|e^A{}_B| = (A) + (B). \tag{13}$$

It is also useful to employ the notation

$$\left|\begin{matrix} A_{11}A_{12}\cdots \\ A_{21}A_{22} \\ \vdots \end{matrix}\right| = (-1)^{\{(A_{11})+(A_{21})+\cdots\}\{(A_{12})+\cdots\}\cdots\}}. \tag{14}$$

The natural analog of an L-module for L a Lie algebra is the L-(super) module where L is a Lie superalgebra. V is an L-module if equipped with a map from $L \times V$ to V denoted $(x,y) \mapsto xv$ such that

$$x(\lambda u + \mu v) = \lambda(xu) + \mu(xv), \tag{15}$$

$$(\lambda x + \mu y)v = \lambda(xv) + \mu(yv), \tag{16}$$

$$[x,y]v = x(yv) - (-1)^{|x||y|}(y)(xv) \tag{17}$$

$$x(V_i) \subseteq V_{i+|x|} \tag{18}$$

for $x, y \in L$, $u, v \in V$, and $\lambda, \mu \in \mathbb{C}$.

A morphism of L-modules V and W is a homomorphism ϕ of the GVS's such that

$$\phi(xv) = x\phi(v) \tag{19}$$

for all $x \in L$ and $v \in V$.

If V is an L-module, we can construct the following two important L-modules:

(i) V^* the contragredient module, where V^* is the dual space and the action is given by

$$(xf)(v) = -(-1)^{|x||f|} f(xv) \quad \text{for} \quad f \in V^*; \tag{20}$$

(ii) gl(V) = End$V$ with the action given by

$$(xf)(v) = xf(v) - (-1)^{|x||f|} f(xv) \quad \text{for} \quad f \in \text{End}V. \tag{21}$$

We can define a representation of a Lie superalgebra to be a homomorphism (of degree 0) of L into some gl(m/n). The usual relationship between representations and L-modules exists.

If $\phi: L \to$ gl(m/n) is a representation, then $\mathbb{C}^{m+n}$ is an L-module and we define $-\phi^{ST}$ to be the representation related to the contragredient module $(\mathbb{C}^{m+n})^*$.

Then, if a typical element of gl(m/n) is $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$, it can be shown that

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{ST} = \begin{pmatrix} A^T & -C^T \\ B^T & D^T \end{pmatrix}, \tag{22}$$

and for homogeneous X, $Y \in$ gl(m/n) we have

$$(XY)^{ST} = (-1)^{|X||Y|} Y^{ST} X^{ST}. \tag{23}$$

III. TENSOR OPERATORS

A general definition[10,11] of tensor operators is as follows. If V and W are L-modules, then a tensor operator is module morphism from V into gl(W). Thus we consider L and a representation $\pi: L \to$ gl(M/N), V an L-module, and apply the above definition. That is, a tensor operator is a map $T: V \to$ gl(M/N) such that if $v \in V$ and $x \in L$,

$$[\pi(x), T(v)] = T(xv). \tag{24}$$

Three cases are of particular importance. Take $L \leqslant$ gl(m/n) with the standard basis $e^A{}_B$, $\mathbb{C}^{m+n}$ with the standard basis δ_A, such that $(\delta^A)_X = \delta^A{}_X$, and $\mathbb{C}^{m+n*}$ with the dual basis δ_B. Then a tensor operator is a linear map: X:gl$(m/n) \to gl(M/N)$ such that if $X(e^C{}_D) = X^C{}_D$ and $x \in L$, then

$$[\pi(x), X^C{}_D] = X([x, e^C{}_D]). \tag{25}$$

A (contravariant) vector operator is a linear map $V: \mathbb{C}^{m+n} \to$ gl(M/N) such that if $V(\delta^A) = V^A$ and $x \in L$, then

$$[\pi(x), V^A] = V(x\delta^A). \tag{26}$$

Using the contragredient module we define a (covariant) vector operator using a linear map $V: \mathbb{C}^{m+n*} \to$ gl(M/N) such that if $V(\delta_A) = V_A$ and $x \in L$, then

$$[\pi(x), V_A] = V(x\delta_A). \tag{27}$$

Now for $L =$ gl(m/n) we have the usual ideas of a "matrix of operators," a "column of operators," and a "row of operators," required to transform as

$$[E^A{}_B, X^C{}_D] = \delta^C{}_B X^A{}_D - \left[\begin{smallmatrix} A & C \\ B & D \end{smallmatrix}\right] \delta^A{}_D X^C{}_B, \tag{28}$$

$$[E^A{}_B, V^C] = \delta^C{}_B V^A, \tag{29}$$

and

$$[E^A{}_B, V_D] = -\left[\begin{smallmatrix} A & \\ B & D \end{smallmatrix}\right] \delta^A{}_D V_B. \tag{30}$$

These are precisely formulae (25), (26), and (27) for this particular case, with $x = e^A{}_B$, $\pi(x) = E^A{}_B$, as can be seen from

$$[e^A{}_B, e^C{}_D] = \delta^C{}_B e^A{}_D - \left[\begin{smallmatrix} A & C \\ B & D \end{smallmatrix}\right] \delta^A{}_D e^C{}_B, \tag{31}$$

$$(e^A{}_B \delta^C)_D = \delta^C{}_B (\delta^A)_D, \tag{32}$$

and

$$
\begin{aligned}
(e^A{}_B\delta_D)^C &= (e^A{}_B\delta_D)(\delta^C) \\
&= -(-1)^{|e^A{}_B||\delta_D|}\delta_D(e^A{}_B\delta^C) \\
&= -[{}^A_B{}^D]\delta_D(\delta^C{}_B\delta^A) \\
&= -[{}^A_B{}^D]\delta^A{}_D(\delta_B)^C,
\end{aligned}
\tag{33}
$$

using the definition of the contragredient module, (20).

Using the canonical isomorphism $V \otimes V^* \simeq \mathrm{gl}(V)$, given by $(\delta^A,\delta_B) \mapsto e^A{}_B$ in this case, the definitions of vector operators give rise to tensor operators. Clearly, we can define higher-order tensor operators as morphisms of the form

$$X:(\otimes^k V) \otimes (\otimes^l V^*) \to \mathrm{gl}(W)$$

for L-modules V and W.

IV. CASIMIR INVARIANTS

Considering the definition of a tensor operator we can often show that if $L \leqslant \mathrm{gl}(m/n)$ and X and Y are tensor operators, then so is

$$W^A{}_B = X^A{}_E[E]Y^E{}_B. \tag{34}$$

Here, as elsewhere, we sum over repeated indices. We note also from the definition that $X^A{}_A$ is an invariant.

So we can define[12] the sequence of Casimir invariants C_p as follows:

$$(\hat{E}^0)^C{}_D = \delta^C{}_D[D], \tag{35}$$

$$(\hat{E}^{p+1})^C{}_D = (\hat{E}^p)^C{}_E[E]E^E{}_D \quad (\text{sum over } E), \tag{36}$$

and

$$C_p = (\hat{E}^p)^C{}_C. \tag{37}$$

Here the $E^A{}_B$ are the images of $e^A{}_B$ under some representation. For generators of a subalgebra suitably labelled we can do likewise.

Now from Kac[7] we have the analogue of Schur's lemma for superalgebra as follows: If $\pi:L \to \mathrm{gl}(m/n)$ is an irreducible representation of L and $X \in \mathrm{gl}(m/n)$ such that $[X,L]=0$, then either

(i) X is a scalar, or

(ii) $\exists\, A$, a nondegenerate matrix permuting $\mathbb{C}_0^{m+n}$ and $\mathbb{C}_1^{m+n}$ such that $A^2 = I$ and $X = \lambda I + \mu A$ for $\lambda,\mu \in \mathbb{C}$.

Note that case (ii) applies only if $m = n$. If $m = n$, there is a question of whether the Casimir invariants are still scalars. If the tensor operators are of degree 0 (as we have implied), then as $E^A{}_A$ is of degree 0, so also is $X(E^A{}_A) = X^A{}_A$. Thus $X^A{}_A$ cannot have an odd component A, and the Casimir invariants C_p are still scalars.

V. SUBALGEBRAS AS FIXED POINT SETS OF HOMOMORPHISMS

A number of useful subalgebras can be defined in the following way. Let $\phi:L \to L$ be a homomorphism; then

$$L|_\phi = \{x \mid \phi(x) = x\} \tag{38}$$

is a subalgebra. For $L|_\phi$ to be graded, that is, a superalgebra, it suffices that ϕ be a graded homomorphism. If we require also that $\phi^2 = id_L$, then

$$L|_\phi = \{x + \phi(x) \mid x \in L\}. \tag{39}$$

So if $\phi:\mathrm{gl}(m/n) \to \mathrm{gl}(m/n)$, a homomorphism such that $\phi^2 = id$, $\mathrm{gl}(m/n)|_\phi$ is generated by

$$
\begin{aligned}
f^A{}_B &= e^A{}_B + \phi(e^A{}_B) &(40)\\
&= e^A{}_B + \phi^{AY}{}_{BX}e^X{}_Y \quad (\text{sum over } X \text{ and } Y). &(41)
\end{aligned}
$$

Then the commutation relations of the $f^A{}_B$'s are

$$
\begin{aligned}
[f^A{}_B, f^C{}_D] = {}& \delta^C{}_B f^A{}_D - [{}^A_B{}^C_D]\delta^A{}_D f^C{}_B \\
& + \phi^{AC}{}_{BX} f^X{}_D - [{}^A_B{}^C_D]\phi^{AY}{}_{BD} f^C{}_Y,
\end{aligned}
\tag{42}
$$

noting that $\phi^{XY}{}_{AB} = 0$ unless $(A)+(B) = (X)+(Y)$ as ϕ is graded. We observe also that $id + \phi$ is a tensor operator.

Some examples of these subalgebras are the following:

(i) $\mathrm{gp}(m) = \{(\begin{smallmatrix} A & B \\ C & D\end{smallmatrix}) \mid A + D^T = B - B^T = C + C^T = 0\} \leqslant \mathrm{gl}(m/m)$

or

$$\mathrm{gp}(m) = \{X \mid X^{\mathrm{ST}}H + HX = 0\} = \mathrm{gl}(m/m)|_p, \tag{43}$$

where

$$p(X) = -H^{-1}X^{\mathrm{ST}}H, \quad H = (\begin{smallmatrix} 0 & I \\ -I & 0\end{smallmatrix}). \tag{44}$$

(ii) $\mathrm{gq}(m) = \{(\begin{smallmatrix} A & B \\ B & A\end{smallmatrix})\} = \mathrm{gl}(m/m)|_q$, (45)

where

$$q(\begin{smallmatrix} A & B \\ C & D\end{smallmatrix}) = (\begin{smallmatrix} D & C \\ B & A\end{smallmatrix}). \tag{46}$$

(iii) $\mathrm{gl}(m/n) = \mathrm{gl}(m/n)|_{id}$.

(iv) $\mathrm{osp}(m/n) = \{(\begin{smallmatrix} A & B \\ C & D\end{smallmatrix}) \mid A^T + A = D^TJ + JD = B - C^TJ = 0\} \leqslant \mathrm{gl}(m/n)$

or

$$\mathrm{osp}(m/n) = \{X \mid X^{\mathrm{ST}}G + GX = 0\} = \mathrm{gl}(m/n)|_\phi, \tag{47}$$

where

$$
\phi(X) = -G^{-1}X^{\mathrm{ST}}G, \quad G = \begin{pmatrix} I & 0 \\ 0 & J\end{pmatrix},
$$

$$
J = \begin{pmatrix} 0 & I_h \\ -I_h & 0\end{pmatrix}, \quad n = 2h. \tag{48}
$$

Returning to the general case, we can define tensor operators for $\mathrm{gl}(m/n)|_\phi$. Using the definition, we see that Y is a tensor operator if it transforms as

$$
\begin{aligned}
[F^A{}_B, Y^C{}_D] = {}& \delta^C{}_B Y^A{}_D - [{}^A_B{}^C_D]\delta^A{}_D Y^C{}_B \\
& + \phi^{AC}{}_{BX} Y^X{}_D - [{}^A_B{}^C_D]\phi^{AY}{}_{BD} Y^C{}_Y.
\end{aligned}
\tag{49}
$$

The contravariant vector operator transforms as

$$[F^A{}_B, W^C] = \delta^C{}_B W^A + \phi^{AC}{}_{BX} W^X \tag{50}$$

and the convariant as

$$[F^A{}_B, W_D] = -[{}^A_B{}^D]\delta^A{}_D W_B - [{}^A_B{}^D]\phi^{AY}{}_{BD} W_Y. \tag{51}$$

Now if X, Y are two tensor operators for $\mathrm{gl}(m/n)|_\phi$, we can form W as follows:

$$W:\mathrm{gl}(m/n) \to \mathrm{gl}(M/N), \quad e^A{}_B \mapsto X^A{}_E[E]Y^E{}_B. \tag{52}$$

Then by considering $x = k^B{}_A E^A{}_B \in \mathrm{gl}(m/n)|_\phi$ it can be shown that W is an adjoint tensor operator for $gl(m/n)|_\phi$. So for all the examples the Casimir invariants are well defined.

VI. CARTAN SUBALGEBRAS

We note that the $e^A{}_A$ (no sum) span a Cartan subalgebra for $gl(m/n)$, and it is natural to consider if the $f^A{}_A$ span a Cartan subalgebra of $gl(m/n)|_\phi$. A tedious analysis shows that they do so as long as there are no B, A such that $\phi(E^A{}_B) = \lambda E^B{}_A$ for some $\lambda \in \mathbb{C}$; this occurs in the osp(m/n) case but not for gp(m) and gq(m). For osp(m/n),[12] a suitable set must take into account the choice of metric. As usual,[12] weights will be lexically ordered and highest weights[7] assigned correspondingly. [For convenience we define the components (μ_A) of the highest weight by $\mu_A = \mu(f^A{}_A)$, even when the $f^A{}_A$ are not all linearly independent.]

VII. CASIMIR INVARIANTS OF gp(m) AND gq(m)

Using the above results, we define the generators of gp(m)

$$\pi^A{}_B = e^A{}_B + p(e^A{}_B) = e^A{}_B - [\bar{A}B] e^{\bar{B}}{}_{\bar{A}}. \tag{53}$$

Here $\bar{A} = A + m \pmod{2m}$. Then the commutation relations are

$$[\pi^A{}_B, \pi^C{}_D] = \delta^C{}_B \pi^A{}_D - [{}^{A}_{B}{}^{C}_{D}] \delta^A{}_D \pi^C{}_B - [\bar{A}B] \delta^C{}_{\bar{A}} \pi^{\bar{B}}{}_D + [\bar{A}B][{}^{A}_{B}{}^{C}_{D}] \delta^{\bar{B}}{}_D \pi^C{}_{\bar{A}} \tag{54}$$

and we have the symmetry

$$\pi^A{}_B = -[\bar{A}B] \pi^{\bar{B}}{}_{\bar{A}}. \tag{55}$$

For a general tensor opeator X, (54) still holds, but with π's replaced by X's, except for $\pi^A{}_B$. If we take an irreducible representation of gp(m), say $\Pi(\pi^A{}_B) = \Pi^A{}_B$, then the same commutation relations hold, and we can define the Casimir invariants C_p. These are constructed recursively through

$$(\hat{\Pi}^1)^A{}_B = \Pi^A{}_B, \tag{56}$$

$$(\hat{\Pi}^k)^A{}_B = \Pi^A{}_E [E] (\hat{\Pi}^{k-1})^E{}_B, \quad \text{etc.,} \tag{57}$$

and

$$C_p = (\hat{\Pi}^p)^A{}_A. \tag{58}$$

Let $c,d = 1,\dots,m$. We observe that $\Pi^c{}_d$ is a raising operator if $c<d$ and lowering if $c>d$, $\Pi^{\bar{c}}{}_{\bar{d}}$ is the converse, $\Pi^{\bar{c}}{}_d$ is always lowering, and $\Pi^c{}_{\bar{d}}$ is always raising.

The $(\hat{\Pi}^k)^A{}_A$ (no sum) have weight zero so act as scalars when applied to a fixed highest weight vector. We denote this value by C^k_A. Let (μ_A) be the components of the highest weight [in view of (55), $\mu_{\bar{a}} = -\mu_a$, for $a = 1,\dots,m$]. Using the definitions and normal-ordering the raising and lowering operators, we can derive the following general recursion relations (here $a,b = 1,\dots,m$):

$$C^{k+1}_a = (\mu_a - a - 1) C^k_a - C^k_{\bar{a}} - \sum_{b>a} C^k_b - \sum_b C^k_{\bar{b}}, \tag{59}$$

$$C^{k+1}_{\bar{a}} = (\mu_a - a + 1) C^k_{\bar{a}} + \sum_{b<a} C^k_{\bar{b}}, \tag{60}$$

and therefore

$$C^{k+1}_a + C^{k+1}_{\bar{a}} = (\mu_a - a - 1)(C^k_a + C^k_{\bar{a}}) - \sum_{b>a} (C^k_b + C^k_{\bar{b}}). \tag{61}$$

Now we know that $C^1_a + C^1_{\bar{a}} = \mu_a + \mu_{\bar{a}} = 0$, so that (61) implies that, for all $k = 1,2,\dots$, we have

$$C^k_a + C^k_{\bar{a}} = 0. \tag{62}$$

Thus, from (62), $C_k = 0$ for all $k = 1,2,\cdots$. This result can also be confirmed directly from the definition (58) by an inductive argument, making use of the symmetry property (55) of the matrix of generators.

The generators of gq(m) are defined in the light of (47) by

$$q^A{}_B = e^A{}_B + q(e^A{}_B) = e^A{}_B + e^{\bar{B}}{}_{\bar{A}}, \tag{63}$$

where again $\bar{A} = A + m \pmod{2m}$. The commutation relations are

$$[q^A{}_B, q^C{}_D] = \delta^C{}_B q^A{}_D - [{}^{A}_{B}{}^{C}_{D}] \delta^A{}_D q^C{}_B + \delta^C{}_{\bar{B}} q^{\bar{A}}{}_D - [{}^{A}_{B}{}^{C}_{D}] \delta^{\bar{A}}{}_D q^C{}_{\bar{B}}, \tag{64}$$

and there is the symmetry

$$q^A{}_B = q^{\bar{A}}{}_{\bar{B}}. \tag{65}$$

For a general tensor operator X, (64) still holds, but with q's replaced by X's, except for $q^A{}_B$. If we take an irreducible representation of gq(m), say $Q(q^A{}_B) = Q^A{}_B$, then the same commutation relations hold, and we can define the Casimir invariants C_p. These are constructed recursively through

$$(\hat{Q}^1)^A{}_B = Q^A{}_B, \tag{66}$$

$$(\hat{Q}^k)^A{}_B = Q^A{}_E [E] (\hat{Q}^{k-1})^E{}_B, \quad \text{etc.,} \tag{67}$$

and

$$C_p = (\hat{Q})^A{}_A. \tag{68}$$

We observe that $Q^c{}_d$ is a raising operator if $c<d$, and lowering if $c>d$ and $Q^c{}_{\bar{d}}$ is also a raising operator if $c<d$, and lowering if $c>d$, while $Q^c{}_{\bar{c}}$ has zero weight. If the highest weight is $(\mu_a/\mu_{\bar{a}})$, it is easily seen that $\mu_a = \mu_{\bar{a}}$, and, using the definition (68) and normal-ordering the raising and lowering operators, we can derive the following eigenvalues of the C_p (acting on a highest weight vector, and hence for the corresponding irreducible representation):

$$C_1 = 2 \sum_{a=1}^m \mu_a, \tag{69}$$

$$C_2 = 0, \tag{70}$$

$$C_3 = 2 \sum_{a=1}^m \mu_a (\mu_a^2 - \mu_a - 2 \sum_{b>a} \mu_b). \tag{71}$$

The result (70) can also be confirmed directly from the definition, making use of the symmetry property (65) of the matrix of generators. In fact a similar argument shows that, in general,

$$C_{2k} = 0 \tag{72}$$

for gq(m).

VIII. CHARACTERISTIC POLYNOMIAL IDENTITIES

For the discussion of characteristic identities for generators of gp(m) and gq(m), rather than use the vector operator techniques used in the earlier[12] derivation of identities for gl(m/n), sl(m/n), and osp(m/n), we follow the treatment of Hannabuss[11] for the Lie algebra case, and also some work of Edwards and Gould.[13]

Let L be a superalgebra, and let V_λ and V_μ be finite-dimensional L-modules with highest weights λ and μ and corresponding representations π_λ and π_μ. Let $z \equiv C_2$ be the universal (quadratic) Casimir invariant and consider the operator

$$Z = \tfrac{1}{2}[\pi_\lambda \otimes \pi_\mu(z) - \pi_\lambda \otimes 1(z) - 1 \otimes \pi_\mu(z)]. \tag{73}$$

Clearly, Z is even, and by construction

$$\pi_\lambda \otimes \pi_\mu(x) Z = Z \pi_\lambda \otimes \pi_\mu(x), \tag{74}$$

for all $x \in L$. In an appropriate basis, Z turns out to be essentially the matrix of generators in the representation μ in the various cases. The characteristic equation of Z thus provides the characteristic identity for the superalgebra.

In practice, we take λ to be the fundamental representation ρ or its contragredient ρ^*. For example, in the latter case, consider Z acting on vectors of the form $\delta_A \otimes v$, where $v \in V_\mu$ and $\{\delta_A\}$ is the standard basis for $\mathbb{C}^{m+n\,*}$, introduced above, and take $L = \mathrm{gl}(m/n)$. Then from (33) and (36) we find {defining $Z(\delta_Y \otimes v) = (1 \otimes Z^X{}_Y)(\delta_X \otimes v) = [X^X_Y]\delta_X \otimes Z^X{}_Y v$}:

$$[\pi_{\rho^*} \otimes \pi_\mu(e^A{}_B)]^X{}_Y = -[{}^{AA}_{B}]\delta^A{}_Y \delta_B{}^X + \delta^X{}_Y E^A{}_B,$$

$$Z^X{}_Y = -E^X{}_Y[XY],$$

where

$$E^X{}_Y = \pi_\mu(e^X{}_Y).$$

In the Lie algebra case, $V_\lambda \otimes V_\mu$ is completely reducible,[14] and, from (74), Z must be a scalar on each irreducible constituent. Moreover, its value can be easily computed from standard formulae, and the characteristic equation written down. However, in the superalgebra case, $V_\lambda \otimes V_\mu$ is not in general completely reducible. However, Z is even and, restricted to each irreducible factor in the associated composition series, must by (74) and Schur's lemma, again, be a scalar, with eigenvalue given by standard formulae. If the irreducible factor occurs with multiplicity m, then the characteristic equation includes m repeated factors for this eigenvalue.

Edwards and Gould[13] have shown that the irreducible factors occurring correspond to irreducible modules with highest weight $\delta + \mu$, where δ is a weight of V_λ; furthermore, the maximum multiplicity that $\delta + \mu$ can occur in the composition series is just the multiplicity of the weight δ in V_λ, independent of μ.

For example, in gl(m/n) the $(m+n)$ weights δ of V_{ρ^*} are $(0,0,\ldots,-1,\ldots,0)$, with multiplicity 1, and the characteristic identity reads[12] in general

$$\prod_{A=1}^{m+n} (X - [A](\mu_A + m - [A]n - A)) = 0, \tag{75}$$

satisfied by the matrix $X^A{}_B = E^A{}_B[B]$. [The corresponding Eq. (31) of Ref. 12 contains an error, corrected in (75).]

For gp(m), $C_2 = 0$, and indeed $C_p = 0$ in general, so that no central element of the trace form is available. For gq(m), $C_2 = 0$, but $C_3 \neq 0$, so that we take $z = C_3$ and consider the operator

$$Z = \tfrac{1}{6}[\pi_{\rho^*} \otimes \pi_\mu(z) - \pi_{\rho^*} \otimes 1(z) - 1 \otimes \pi_\mu(z)]. \tag{76}$$

Acting on vectors of the form $\delta_A \otimes v$, with $v \in V_\mu$ and $\{\delta_A\}$ the standard basis of $\mathbb{C}^{m+m\,*}$, Z becomes essentially the square of the matrix of generators,

$$X^A{}_B = Q^A{}_E[E]Q^E{}_B[B],$$

where $Q^A{}_B = \pi_\mu(q^A{}_B)$. The m weights δ of ρ^* have components $(0,0,\ldots,-1,\ldots,0)$, with multiplicity 2. Evaluating $\pi_{\delta+\mu}(C_3)$ and $\pi_\mu(C_3)$ using (71) in terms of the m components (μ_a) of μ, the characteristic equation becomes a matrix identity for $X^A{}_B$. In general, we have

$$\prod_{a=1}^{m} [X - \mu_a(\mu_a - 1)]^2 = 0. \tag{77}$$

IX. CONCLUSIONS

We have shown that a large class of (super) subalgebras of gl(m/n) can be realized as fixed-point sets of module homomorphisms, including gl(m/n), sl(m/n), osp(m/n), and gp(m) and gq(m), related to the simple superalgebras $p(m)$ and $q(m)$. Natural definitions of tensor and vector operators have been introduced which permit of a unified treatment of all these cases. In particular, previous work[12] on Casimir invariants and characteristic identities for the classical superalgebras has been extended to the previously unstudied cases of gp(m) and gq(m).

For gp(m), the Casimir invariants C_p were shown to vanish identically. For gq(m), C_p vanishes for even p, and eigenvalues of C were given in terms of the highest weight components. A polynomial characteristic identity was derived for a matrix $Q^A{}_E[E]Q^E{}_B$ quadratic in the generators.

The vanishing of all C_p for gp(m) shows the structure of its enveloping algebra to be particularly special: However, one must look to other methods[15] than those used here for the characteristic identities, which from general arguments should still exist.[15–17] On the other hand, the identity derived for gq(m) is a new type of result in that it derives from a central element of the enveloping algebra which is cubic in the generators, as the usual construction involving the universal (quadratic) Casimir invariant fails [as it must for gp(m), gq(m) and $p(m)$, $q(m)$ since the Killing forms vanish].

The tensor calculus allows Young diagram methods for the classical superalgebras[18–20] to be excluded to gp(m) and gq(m), and likewise (by trace projections) to $p(m)$ and $\tilde{q}(m)$. Representation of $q(m)$ correspond to those "zero triality" representations of $\tilde{q}(m)$, wherein the center is trivially represented.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. D. M. O'Brien, Dr. A. Cant, Dr. A. L. Carey, and Dr. S. A. Edwards for helpful discussions. M. K. M. acknowledges the support of an Australian Commonwealth Postgraduate Award during this research.

P. D. J. acknowledges the support of a Queen Elizabeth II Fellowship. We thank the referee for correcting several points in an earlier version of the paper.

[1]We adopt here the terminology of Kac rather than "graded Lie algebras" as, in general, Lie superalgebras are not Lie algebras.

[2]L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975).
[3]S. Sternberg, "Some Recent Results on Supersymmetry," in *Lecture Notes in Mathematics* **570**, edited by A. Dold and B. Eckman (Springer-Verlag, Berlin, 1977).
[4]B. Kostant, "Graded Manifolds, Graded Lie Theory and Pre-quantization," in *Lecture Notes in Mathematics* **570**, Ref. 3.
[5]A. Rogers, J. Math. Phys. **21**(6), 1352 (1980).
[6]M. Batchelor, Trans. Am. Math. Soc. **253**, 329 (1979).
[7]V. G. Kac, Adv. in Math. **26**, 8 (1977).
[8]M. Scheunert, "Theory of Lie Superalgebras," in *Lecture Notes in Mathematics* **716** (Springer-Verlag, Berlin, 1979).

[9]Here, as elsewhere, for simplicity we call subsuperalgebras, subalgebras and supermodules, modules.

[10]A. Cant and C. A. Hurst, J. Aust. Math. Soc. Ser. B **20**, 446 (1978).
[11]K. Hannabuss, "Characteristic Equations for Semisimple Lie Groups," preprint Math. Inst., Oxford, 1972.
[12]P. D. Jarvis and H. S. Green, J. Math. Phys. **20**, 2115 (1979).
[13]S. A. Edwards and M. D. Gould (unpublished).
[14]J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer-Verlag, Berlin, 1967).
[15]D. M. O'Brien, A. Cant, and A. L. Carey, Ann. Inst. H. Poincaré **26**, 405 (1977).
[16]M. D. Gould, Ann. Inst. H. Poincaré A **32**, 203 (1980).
[17]Y. Ilamed, Bull. Res. Council Isr. Sect. A. **5**, 197 (1956).
[18]P. H. Dondi and P. D. Jarvis, Z. Phys. C **4**, 201 (1980).
[19]P. H. Dondi and P. D. Jarvis, J. Phys. A **14**, 547 (1981).
[20]A. Baha Balantekin and I. Bars, J. Math. Phys. **22**, 1810 (1981).

Inequalities and local uncertainty principles

John F. Price
School of Mathematics, University of New South Wales, Kensington, New South Wales, 2033, Australia

(Received 15 June 1982; accepted for publication 17 September 1982)

Inequalities for Fourier transforms are developed which describe local uncertainty principles in the sense that if the uncertainty of momentum is small, then so is the probability of being localized at any point. They give estimates for essentially all states and lead to lower bounds for Hamiltonians.

PACS numbers: 02.30. − f, 03.65.Bz

1. INTRODUCTION

Local uncertainty principles assert that when the uncertainty of the momentum is small, the probability of being localized at any point is very small. From this follows the classical uncertainty principle that the position uncertainty is large. Faris[1] has recently presented a number of inequalities supporting local uncertainty principles and used them to find lower bounds on the total energy of quantum mechanical systems.

The starting point for this paper is the following local uncertainty principle inequality[1]: in dimension $k\geqslant 3$,

$$\text{prob}\{|q-c|\leqslant d\}\leqslant(2/\hbar(k-2))^2 d^2(\Delta p)^2 \tag{1.1}$$

for all $d\geqslant 0$ and c in $\mathbb{R}^k$, where p and q are the momentum and position observables and $\Delta p=\langle(p-\langle p\rangle)^2\rangle^{1/2}$, the uncertainty of momentum. For this inequality to provide an effective bound on the localization of position, we need Δp to be finite and preferably not too large. In a similar vein, for the Heisenberg uncertainty principle, $\Delta p.\ \Delta q\geqslant k\hbar/2$, to have any content we need Δp and Δq to be finite.

Taking the states as unit vectors in $L^2(\mathbb{R}^k)$, the finiteness of Δp and Δq imposes restrictions on their vanishing rates at infinity and the vanishing rates of their Fourier transforms. For example, if a state ψ is of polynomial form at infinity, that is, of the form $\psi\sim|x|^{-\alpha}$ as $|x|\rightarrow\infty$, finiteness of Δq requires $\alpha>(k+2)/2$. Finiteness of Δp imposes a similar restriction on the transform of ψ. On the other hand, membership of $L^2(\mathbb{R}^k)$ only requires $\alpha>k/2$. The uncertainty of momentum (or position) can be infinite if, roughly speaking, the wave function (or its Fourier transform) oscillates very rapidly or sharply. Even when ψ is the (suitably normalized) characteristic function of the unit ball, the uncertainty of momentum is infinite.

Here we define a family of measures of uncertainty which generalize the usual notion, but first some notation. For (Lebesgue measurable) functions f on $\mathbb{R}^k$ define $\|f\|_t=(\int|f(x)|^t\,dx)^{1/t}$ when $1\leqslant t<\infty$, and when $t=\infty$, $\|f\|_\infty=\text{ess sup}\{|f(x)|:x\in\mathbb{R}^k\}$. (Unless stated otherwise, all integrals are over $\mathbb{R}^k$.) The Fourier transform $\hat f$ of f is defined as

$$\hat f(y)=\int f(x)e^{-2\pi ixy}\,dx,$$

where $xy=x_1y_1+\cdots+x_ky_k$.

Definition 1: For each pair (t,θ), with $1\leqslant t\leqslant\infty$ and $\theta>0$, define uncertainties of q,p for the system in state ψ by

$$\Delta_{t,\theta}q=\|\,|x-\langle q\rangle|^\theta\psi\|_t,$$
$$\Delta_{t,\theta}p=\|\,|2\pi\hbar y-\langle p\rangle|^\theta\hat\psi\|_t.$$

Remarks: (i) Whenever the state of a system is described by a wave function ψ, then $|\psi|^2$ is the probability distribution of its position and $|\hat\psi|^2$ of its momentum. Hence $\Delta_{t,\theta}q$ and $\Delta_{t,\theta}p$ are measures of the "spread" of these distributions about $\langle q\rangle$ and $\langle p\rangle$, respectively. When $t=2$ and $\theta=1$ we recover the classical uncertainties, that is, $\Delta_{2,1}q=\Delta q$ and $\Delta_{2,1}p=\Delta p$.

(ii) If $\psi\sim|x|^{-\alpha}$ is a state (and hence $\alpha>k/2$), then we can always find t,θ such that $\Delta_{t,\theta}q<\infty$, namely those t,θ satisfying $1\leqslant t\leqslant\infty$, $\theta>0$ and $\alpha>\theta+k/t$. This situation also holds for $\hat\psi$.

Whenever $1\leqslant t\leqslant\infty$, define $t'=t(t-1)^{-1}$ and $t^\#=2t(t-2)^{-1}$ (with $1'=\infty$, $\infty'=1$, and $2^\#=\infty$). The main result of the paper is:

Theorem 1: Suppose that $1\leqslant t\leqslant\infty$, $\theta\geqslant 0$ and the dimension is k. There exists a constant K such that for all states,

$$(\text{prob}\{|q-c|\leqslant d\})^{1/2}\leqslant Kd^{\theta-k/t^\#}\Delta_{t,\theta}p \tag{1.2}$$

for all $d\geqslant 0$ and $c\in\mathbb{R}^k$ if and only if $k/t^\#<\theta<k/t'$ or $(t,\theta)=(1,0)$ or $(2,0)$.

Remarks: (i) Replacement of states ψ by their "normalized dilates" $D_\lambda\psi$, where $D_\lambda\psi(x)=\lambda^{k/2}f(\lambda x)$, shows that only the power $\theta-k/t^\#$ is possible for d in (1.2).

(ii) Suppose $\theta=1$ and $t=2$. When $k\geqslant 3$, inequality (1.2) is (1.1) above apart from the constant. If $k=1$ or 2, $\theta\geqslant k/t'$ and so (1.2) is not possible as was pointed out by Faris.[1]

In Sec. 2, the main inequality (1.2) is proved under the conditions described in Theorem 1 and in Sec. 3 these conditions are shown to be the best possible. Inequalities with more general weights, that is, more general measures of uncertainty, are considered in Sec. 4, while applications to lower estimates of Hamiltonians are given in Sec. 5. (See Faris[1] for more details of this approach. Also Lieb[2] uses a Sobolev inequality as a local uncertainty principle to obtain results on the stability of matter.) In Cowling and Price[3] conditions are given on s, $t\in[1,\infty]$ and $\theta,\phi\geqslant 0$ so that

$$\Delta_{s,\theta}p\,.\,\Delta_{t,\phi}q\geqslant C_1\quad\text{or}\quad\Delta_{s,\phi}p+\Delta_{t,\phi}q\geqslant C_2$$

for constants C_1, $C_2>0$. Special cases of these results follow from Theorem 1 and are given as Corollary 1 of Sec. 2 and subsequent remarks.

 0022-2488/83/071711-04$02.50

2. THE MAIN INEQUALITY

In the Schrödinger representation the wave functions ψ describing the states of a k-dimensional system are unit vectors in $L^2(\mathbb{R}^k)$. The observables of position $\{q_j : j = 1,...,k\}$ and momentum $\{p_j : j = 1,...,k\}$ are given by the operators $\psi \to x_j \psi$ and $\psi \to -i\hbar D_j$. Hence, apart from using distributions as a starting point, the following result is equivalent to Theorem 1. (Also see the first remark after Definition 1.)

Before stating the result, we remind the reader of two standard definitions. *Tempered distributions* were introduced by Schwartz and form the dual of the space of rapidly decreasing, infinitely differentiable functions $\mathscr{S}$. *Locally square integrable functions* f satisfy $\int_K |f(x)|^2 dx < \infty$ for every bounded closed set K.

Theorem 1′: Suppose that $1 \leqslant t \leqslant \infty$, $\theta \geqslant 0$ and $k \in \{1, 2, ...\}$.

(i) Let t, θ, k satisfy $k/t^{\#} < \theta < k/t'$ or $(t, \theta) = (1, 0)$ or $(2, 0)$. Given $b \in \mathbb{R}^k$, suppose that f is a tempered distribution generated by a locally integrable function and that $\| |x-b|^\theta f \|_t < \infty$. Then $\hat{f}$ is locally square integrable. Moreover, there exists a constant $K_1 = K_1(t, \theta, k)$ independent of b and f such that

$$\left(\int_{|y-c|<d} |\hat{f}(y)|^2 dy \right)^{1/2} \leqslant K_1 d^{\theta - k/t^{\#}} \| |x-b|^\theta f \|_t \tag{2.1}$$

for all $c \in \mathbb{R}^k$ and $d > 0$.

(ii) If $\theta \geqslant k/t'$ [except for $(t, \theta) = (1, 0)$] or $\theta \leqslant k/t^{\#}$ [except for $(t, \theta) = (2, 0)$], no such inequality is possible.

Constants: Let $\omega_k = 2\pi^{k/2}/\Gamma(k/2)$ and $\Omega_k = \pi^{k/2}/\Gamma\{(k+2)/2\}$ be the surface area and volume, respectively, of the unit ball in $\mathbb{R}^k$. (When $k = 1$, redefine the area as 2.) Under the hypotheses of (i), define

$$\lambda_1 = \Omega_k^{1/2} \omega_k^{1/t'} (k - \theta t')^{-1/t'}, \tag{2.2}$$

$$\lambda_2 = \begin{cases} \Omega_k^{-1/t^{\#}} & \text{if} \quad 1 \leqslant t \leqslant 2, \\ \omega_k^{1/t^{\#}} (\theta t^{\#} - k)^{-1/t^{\#}} & \text{if} \quad 2 < t \leqslant \infty. \end{cases} \tag{2.3}$$

The method described below gives the constant in (2.1) as $K_1 = \lambda_1 + \lambda_2$. This means that the constant in (1.2) is $K = K_1/(2\pi\hbar)^\theta = (\lambda_1 + \lambda_2)/(2\pi\hbar)^\theta$.

Proof of Theorem 1′ (i): Without loss of generality we assume that $b = c = 0$ since the general case reduces to this by replacing f with the function $f(x-b) e^{2\pi i c(x-b)}$.

The case of $\theta = 0$ and $1 \leqslant t \leqslant 2$ is easily disposed with. If $\| |x|^0 f \|_t = \| f \|_t < \infty$, then $\| \hat{f} \|_{t'} \leqslant \| f \|_t$ by the Hausdorff–Young inequality and so f is locally square integrable. Furthermore, by Hölder's inequality with $r = t'/2$,

$$\begin{aligned} \left(\int_{|y|\leqslant d} |\hat{f}(y)|^2 dy \right)^{1/2} &= \left(\int_{|y|\leqslant d} |\hat{f}(y)|^2 \cdot 1 \, dy \right)^{1/2} \\ &\leqslant \left(\int_{|y|\leqslant d} |\hat{f}(y)|^{2r} dy \right)^{1/2r} \\ &\quad \times \left(\int_{|y|\leqslant d} 1^{r'} dy \right)^{1/2r'} \\ &= \left(\int_{|y|<d} |\hat{f}(y)|^{t'} dy \right)^{1/t'} \\ &\quad \times (d^k \Omega_k)^{-1/t^{\#}}. \end{aligned} \tag{2.4}$$

(This step, or minor variants of it, will be used frequently in the sequel; it will simply be referred to as Hölder's inequality with parameter $t'/2$. Also we won't bother to state separately the case $t'/2 = \infty$.) Hence, using Hausdorff–Young,

$$\left(\int_{|y|<d} |\hat{f}(y)|^2 dy \right)^{1/2} \leqslant \Omega_k^{-1/t^{\#}} d^{-k/t^{\#}} \| f \|_t$$

as required.

From now on suppose that $1 \leqslant t \leqslant \infty$, $\theta \geqslant 0$, $k \in \{1, 2, ...\}$ and $k/t^{\#} < \theta < k/t'$. Let f be a tempered distribution generated by a locally integrable function and suppose that $\| |x|^\theta f \|_t < \infty$. Let B be the closed unit ball in $\mathbb{R}^k$ with center 0. Clearly f can be decomposed as $\phi + g$, where ϕ is a distribution with support in B and g is a function which is continuous on B and which satisfies $\| |x|^\theta g \|_t < \infty$. Assume that $1 \leqslant t \leqslant 2$. Since $g \in L^t$, $\hat{g} \in L^{t'}$ and so is locally square integrable. Hence, so is $\hat{f} = \hat{\phi} + \hat{g}$ because $\hat{\phi}$ is analytic.

Now suppose $2 < t \leqslant \infty$ and let B' denote the complement of B. By Hölder's inequality with parameter $t/2$,

$$\left(\int_{B'} |g(x)|^2 dx \right)^{1/2} \leqslant \left(\int_{B'} |x|^{\theta t} |g(x)|^t dx \right)^{1/t} \left(\int_{B'} |x|^{-\theta t^{\#}} dx \right)^{1/t^{\#}}. \tag{2.5}$$

The last integral is finite since $\theta > k/t^{\#}$ and $g \in L^2$. Hence once again, $\hat{f} = \hat{\phi} + \hat{g}$ is locally square integrable.

Let $E = \{x \in \mathbb{R}^k : |x| \leqslant a\}$ and $F = \{y \in \mathbb{R}^k : |y| \leqslant d\} = dB$. Denote the characteristic or indicator function of E by χ_E. Since $f = f\chi_E + f\chi_{E'}$, $\hat{f} = (f\chi_E)^\wedge + (f\chi_{E'})^\wedge$ and so

$$\| \hat{f} \chi_F \|_2 \leqslant \| (f\chi_E)^\wedge \chi_F \|_2 + \| (f\chi_{E'})^\wedge \chi_F \|_2. \tag{2.6}$$

We estimate separately the last two integrals, beginning with the first:

$$\begin{aligned} \| (f\chi_E)^\wedge \chi_F \|_2 &\leqslant \| (f\chi_E)^\wedge \|_\infty \Omega_k^{1/2} d^{k/2} \\ &\leqslant \| f\chi_E \|_1 \Omega_k^{1/2} d^{k/2} \\ &\leqslant \| |x|^\theta f \chi_E \|_t \| |x|^{-\theta} \chi_E \|_{t'} \Omega_k^{1/2} d^{k/2}, \end{aligned}$$

where the first and third steps are by Hölder's inequality (with parameters ∞ and t) and the second by the Hausdorff–Young inequality. Since $\theta < k/t'$, $\| |x|^{-\theta} \chi_E \|_{t'}$ exists and equals $\omega_k^{1/t'} (k - \theta t')^{-1/t'} a^{-\theta + k/t'}$. Hence,

$$\| (f\chi_E)^\wedge \chi_F \|_2 \leqslant \lambda_1 d^{k/2} a^{-\theta + k/t'} \| |x|^\theta f \|_t. \tag{2.7}$$

Turning to the last integral of (2.5), suppose first that $1 \leqslant t \leqslant 2$ (and hence that $2 \leqslant t' \leqslant \infty$). In the following sequence of steps, the first is by Hölder's inequality with parameter $t'/2$:

$$\begin{aligned} \| (f\chi_{E'})^\wedge \chi_F \|_2 &\leqslant \| (f\chi_{E'})^\wedge \chi_F \|_{t'} (d^k \Omega_k)^{-1/t^{\#}} \\ &\leqslant \| f\chi_{E'} \|_t (d^k \Omega_k)^{-1/t^{\#}} \\ &\leqslant \| |x|^\theta f \chi_{E'} \|_t \| |x|^{-\theta} \chi_{E'} \|_\infty (d^k \Omega_k)^{-1/t^{\#}} \\ &\leqslant \lambda_2 a^{-\theta} d^{-k/t^{\#}} \| |x|^\theta f \chi_{E'} \|_t. \end{aligned} \tag{2.8}$$

Now suppose that $2 < t \leqslant \infty$:

$$\begin{aligned} \| (f\chi_{E'})^\wedge \chi_F \|_2 &\leqslant \| (f\chi_{E'})^\wedge \|_2 = \| f\chi_{E'} \|_2 \\ &\leqslant \| |x|^\theta f \|_t \lambda_2 a^{-\theta + k/t^{\#}}, \end{aligned} \tag{2.9}$$

where we argue as for (2.5).

Substitution of (2.7), (2.8), and (2.9) in (2.6) with $a = 1/d$ yields the required inequality (2.1) with $K_1 = \lambda_1 + \lambda_2$. ■

Uncertainty principles: As indicated in the introduction, we would expect local uncertainty principles to lead to classical uncertainty principles. In our case this implication is straightforward, but we point out that more general inequalities have recently been established.[3]

Corollary 1: Suppose that t, θ, and k satisfy the hypotheses of Theorem 1' (i). Suppose also that $2\leqslant s\leqslant\infty$ and $\phi\geqslant 0$ satisfy $\phi > k/s^{\#}$. Then there exists a constant K_2 such that

$$\|f\|_2\leqslant K_2(\|\,|x|^{\theta}f\|_t + \|\,|y|^{\phi}\hat{f}\|_s) \tag{2.10}$$

for all f in L^2.

Proof: With notation as in the preceding proof and arguing as for (2.5),

$$\left(\int_{F'}|\hat{f}(y)|^2dy\right)^{1/2}\leqslant \text{const}\|\,|y|^{\phi}\hat{f}\|_s$$

since $\phi > k/s^{\#}$. Hence, by combining with (2.1),

$$\begin{aligned}\|f\|_2 = \|\hat{f}\|_2 &\leqslant \|\hat{f}\chi_{F'}\|_2 + \|\hat{f}\chi_F\|_2\\ &\leqslant (\lambda_1+\lambda_2)d^{\,\theta-k/t^{\#}}\|\,|x|^{\theta}f\|_t\\ &\quad+\text{const}\times\|\,|y|^{\phi}\hat{f}\|_s\\ &\leqslant K_2(\|\,|x|^{\theta}f\|_t + \|\,|y|^{\phi}\hat{f}\|_s),\end{aligned}$$

as required. ■

Remarks: (i) The parameter d can be varied to minimize K_2 (but even so K_2 is far from the best possible).

(ii) Using dilates of f it can be shown[3] that (2.10) is equivalent to

$$\|f\|_2\leqslant K_3\|\,|x|^{\theta}f\|_t^{\alpha}\|\,|y|^{\phi}\hat{f}\|_s^{1-\alpha},$$

where α is defined by $\alpha(\theta - k/t^{\#}) = (1-\alpha)(\phi - k/s^{\#})$ and $K_3 = K_2\alpha^{-\alpha}(1-\alpha)^{\alpha-1}$. Hence, with the notation of Sec. 1,

$$(\Delta_{t,\theta}q)^{\alpha}(\Delta_{s,\phi}\,p)^{1-\alpha}\geqslant\hbar K_3^{-1}, \tag{2.11}$$

where t, s, θ, and ϕ satisfy the requirements of the preceding corollary and α is as shown. When $p=q=2$ and $\theta,\phi>0$, this was shown by Hirschman.[4]

The following corollary will be useful in Sec. 5.

Corollary 2: Given $r\in[1,\infty)$, select $\theta\in(0,k/2r)$. Suppose $v\in L^r\cap L^{\infty}$ and for each $a>0$ define

$$K_4 = a^{-\theta}\{a^{k/2r}[\omega_k/(k-2\theta r)]^{1/2r}\|v\|_r^{1/2} + \|v\|_{\infty}^{1/2}\}.$$

Then for all $f\in L^2$ and $b\in\mathbb{R}^k$

$$\left(\int|\hat{f}(y)|^2|v(y)|dy\right)^{1/2}\leqslant K_4\|\,|x-b|^{\theta}f\|_2.$$

Proof: Without loss of generality, suppose that $b=0$ and $v\geqslant 0$. Since $f = f\chi_E + f\chi_{E'}$, where $E=\{x:|x|\leqslant a\}$,

$$\|\hat{f}v^{1/2}\|_2\leqslant\|(f\chi_E)\hat{\ }v^{1/2}\|_2 + \|(f\chi_{E'})\hat{\ }v^{1/2}\|_2. \tag{2.12}$$

As in the proof of Theorem 1'(i), we estimate the two parts separately. For the first integral, the Hausdorff–Young inequality between two applications of Hölder's inequality with parameters r' and $(r+1)/r$ shows that

$$\begin{aligned}\|(f\chi_E)\hat{\ }v^{1/2}\|_2 &\leqslant\|\,|x|^{\theta}f\chi_E\|_2\|\,|x|^{-\theta}\chi_E\|_{2r}\|v\|_r^{1/2}\\ &\leqslant\|\,|x|^{\theta}f\|_2[2\omega_k/(k-2\theta r)]^{1/2r}a^{-\theta+k/2r}\|v\|_r^{1/2},\end{aligned}$$

using the fact that $\theta < k/2r$. Now argue as in the derivation of (2.8) with $t=2$:

$$\begin{aligned}\|(f\chi_{E'})\hat{\ }v^{1/2}\|_2 &\leqslant\|f\chi_{E'}\|_2\|v\|_{\infty}^{1/2}\\ &\leqslant\|\,|x|^{\theta}f\chi_{E'}\|_2\|\,|x|^{-\theta}\chi_{E'}\|_{\infty}\|v\|_{\infty}^{1/2}\\ &\leqslant\|\,|x|^{\theta}f\|_2a^{-\theta}\|v\|_{\infty}^{1/2}.\end{aligned}$$

The required result is arrived at by substituting these two estimates into (2.12). ■

3. COUNTEREXAMPLES

In this section we collect together the counterexamples necessary to establish Theorem 1'(ii). Throughout we suppose that $\theta\geqslant 0$, $1\leqslant t\leqslant\infty$, and $k\in\{1,2,\ldots\}$.

Counterexample 1: Suppose $\theta > k/t'$. Then no inequality of the form

$$\left(\int_{|y|\leqslant d}|\hat{f}(y)|^2dy\right)^{1/2}\leqslant\text{const}\times d^{\,\theta-k/t^{\#}}\|\,|x|^{\theta}f\|_t \tag{3.1}$$

is possible for all f in $\mathscr{S}$.

Proof: Choose $f\in\mathscr{S}$ with $\hat{f}(0)>0$. Then

$$\begin{aligned}&d^{\,k/t^{\#}-\theta}\left(\int_{|y|\leqslant d}|\hat{f}(y)|^2dy\right)^{1/2}\\ &\quad= d^{\,k/t'-\theta}\left(d^{-k}\int_{|y|\leqslant d}|\hat{f}(y)|^2dy\right)^{1/2}\\ &\quad\qquad\to\infty\quad\text{as}\quad d\to 0.\end{aligned}$$ ■

The following case is more delicate.

Counterexample 2: No inequality of the form (3.1) is possible if $\theta = k/t'$ and $t>1$.

Proof: Let α be any function in $C^{\infty}(\mathbb{R})$ satisfying $\alpha(s)=0$ for $s\leqslant\frac{1}{2}$, $\alpha(s)=1$ for $s\geqslant 1$ and $0\leqslant\alpha\leqslant 1$. For $\epsilon>0$ define $f_{\epsilon}(x) = \alpha(|x|)|x|^{-k}(\log(2+|x|))^{-1}\exp(-\epsilon|x|)$ on $\mathbb{R}^k$. Let g_{ϵ} denote its Fourier transform. By Theorem 4 of Wainger,[5] $g(y)=\lim_{\epsilon\to 0+}g_{\epsilon}(y)$ is defined for all $y\neq 0$, is infinitely differentiable except at $y=0$, and as $y\to 0$,

$$\begin{aligned}g(y) &= 2\pi^{k/2}\Gamma(k/2)^{-1}\int_1^{1/|y|}(r\log(2+|r|))^{-1}dr\\ &\quad+o\int_1^{1/|y|}(r\log(2+|r|))^{-1}dr.\end{aligned}$$

In particular,

$$g(y)\sim\log\log(1/|y|)\quad\text{as}\quad y\to 0. \tag{3.2}$$

Let $f(x)=\lim_{\epsilon\to 0+}f_{\epsilon}(x)$. Evidently $f, f_{\epsilon}\in L^2$ and $\|f-f_{\epsilon}\|_2\to 0$ as $\epsilon\to 0+$, so that $\|g-g_{\epsilon}\|_2\to 0$ as well. Hence, $g=\hat{f}$.

Assume $\theta = k/t'$, divide both sides of (3.1) by $d^{k/2}$, and substitute the function f we have just described. Then $\|\,|x|^{\theta}f\|_t<\infty$ provided $1<t\leqslant\infty$, whereas

$$\left(d^{k}\int_{|y|\leqslant d}|g(y)|^2dy\right)^{1/2}\to\infty\quad\text{as}\quad d\to 0+$$

by virtue of (3.2). ■

Counterexample 3: If $\theta\leqslant k/t^{\#}$ with $2<t\leqslant\infty$, then (3.1) is not possible.

Proof: If $\theta < k/t^{\#}$, then for each $f\in\mathscr{S}$ the left side of (3.1) can be made arbitrarily small by choosing d sufficiently large. This provides a contradiction since as $d\to\infty$, the left side tends to $\|f\|_2$.

Now assume $\theta = k/t^{\#}$ with $2 < t \leqslant \infty$. If we suppose that (3.1) is valid we arrive at

$$\|f\|_2 \leqslant \text{const} \times \| \, |x|^{k/t^{\#}} f \|_t .$$

This is contradicted by substituting $f = f_n$, where $f_n(x) = |x|^{-k/2}$ for $1 \leqslant |x| \leqslant n$ and 0 otherwise, and letting $n \to \infty$. ∎

4. GENERAL WEIGHTS

Suppose that w: $\mathbb{R}^k \to \mathbb{R}^+$ is a continuous function. In this section we mention modifications of the proof of Theorem 1′(i) to establish conditions on w to ensure the existence of a function α: $\mathbb{R}^+ \to \mathbb{R}^+$ satisfying

$$\alpha(d) \to 0 \quad \text{as} \quad d \to 0+, \tag{4.1}$$

$$\left(\int_{|y-c|<d} |\hat{f}(y)|^2 dy \right)^{1/2} \leqslant \alpha(d) \| w(x-b) f \|_t \tag{4.2}$$

for all functions f in L^2, and all b, $c \in \mathbb{R}^k$, and $d > 0$.

Given $t \in [1, \infty]$, suppose that w satisfies the following conditions:

(i) $\|w^{-1} \chi_{d^{-1}B}\|_{t'} = o(d^{-k/2})$ as $d \to 0+$,

(ii) $\|w^{-1} \chi_{(d^{-1}B)'}\|_\infty = o(d^{k/t^{\#}})$ as $d \to 0+$ if $1 \leqslant t \leqslant 2$,

(iii) $w^{-1} \in L^{t^{\#}}$ if $2 < t \leqslant \infty$.

If $1 \leqslant t \leqslant 2$, define

$$\alpha(d) = \Omega_k^{1/2} d^{k/2} \|w^{-1} \chi_{d^{-1}B}\|_{t'} + \Omega_k^{-1/t^{\#}} d^{-k/t^{\#}} \|w^{-1} \chi_{d^{-1}B}\|_\infty .$$

Otherwise define

$$\alpha(d) = \Omega_k^{1/2} d^{k/2} \|w^{-1} \chi_{d^{-1}B}\|_{t'} + \|w^{-1} \chi_{(d^{-1}B)'}\|_{t^{\#}} .$$

By following through the argument for the proof of Theorem 1′(i) we get:

Theorem 2: Let $t \in [1, \infty]$ and suppose w satisfies the preceding conditions. For α as defined above, Eqs. (4.1) and (4.2) are valid.

5. ESTIMATES FOR HAMILTONIANS

In this section we use our inequalities to develop conditions on potentials v: $\mathbb{R}^k \to \mathbb{R}$ which ensure that $\langle H \rangle \geqslant 0$, or at least that $\langle H \rangle$ is bounded below, where H is the Hamiltonian $H(p, q) = p^2/2m + v(q)$. There is considerable overlap between our results and those of Faris.[1]

The most straightforward approach is to assume that v is bounded with support in a ball and apply Theorem 1. More interesting results may be obtained from Corollary 2 of Theorem 1′. We require $v \in L^r \cap L^\infty$ for some $r \in [1, \infty)$ so that, although no singularities are allowed, decay at infinity can be arbitrarily mild. The main inequality (5.1) is valid for all dimensions but only provides nonbinding conditions (that is, $\langle H \rangle \geqslant 0$) for $k \geqslant 3$.

The following lemma is established by elementary calculus.

Lemma: Suppose $\eta(t) = \mu_1 t - \mu_2 t^\theta$ for $t \in \mathbb{R}^+$ where $\mu_1, \mu_2 \geqslant 0$ and $0 < \theta \leqslant 1$. Then, if $\theta < 1$,

$$\min\{\eta(t): t \in \mathbb{R}^+\} = -(1-\theta)(\theta^\theta \mu_2 / \mu_1^\theta)^{1/(1-\theta)}$$

and $\eta(t) > 0$ if and only if $t > (\mu_2/\mu_1)^{1/(1-\theta)}$. Otherwise $\eta(t) \geqslant 0$ for all $t \geqslant 0$ provided $\mu_2 \geqslant \mu_1$.

Suppose that $1 \leqslant r < \infty$, $v \in L^r \cap L^\infty$, and $0 < \theta \leqslant 1$ with $\theta < k/2r$. Let K_m be the minimum of the constant K_4 in Corollary 2 of Sec. 2 as a ranges over $(0, \infty)$. Hölder's inequality shows that $\| \, |x|^\theta f \|_2 \leqslant \|x f\|_2^\theta \|f\|_2^{1-\theta}$ if $0 < \theta < 1$ and so, from Corollary 2,

$$|\langle v \rangle| \leqslant K_m^2 \| y \hat{\phi} \|_2^{2\theta} = K_m^2 (2\pi\hbar)^{-2\theta} \langle p^2 \rangle^\theta$$

for all states ϕ. Since $\langle H \rangle = \langle p^2 \rangle / 2m + \langle v(q) \rangle \geqslant \langle p^2 \rangle / 2m - \langle v(q) \rangle$ if v is real valued,

$$\langle H \rangle \geqslant \langle p^2 \rangle / 2m - K_m^2 (2\pi\hbar)^{-2\theta} \langle p^2 \rangle^\theta . \tag{5.1}$$

In combination with the preceding lemma, this gives:

Theorem 3: Suppose that $r \in [1, \infty)$, v is real valued in $L^r \cap L^\infty$ and $\theta \in (0, 1]$ satisfies $\theta < k/2r$.

(i) If $\theta = 1$, $\langle H \rangle \geqslant 0$ for all states provided $K_m^2 \leqslant 2\pi^2 \hbar^2 / m$.

(ii) If $0 < \theta < 1$, $\langle H \rangle \geqslant -(1-\theta) \times (K_m^2 (\theta m / 2(\pi\hbar)^2)^\theta)^{1/(1-\theta)}$ and $\langle H \rangle \geqslant 0$ for all states satisfying $\langle p^2 \rangle \geqslant (2m K_m^2 / (2\pi\hbar)^{2\theta})^{1/(1-\theta)}$.

[1] W. G. Faris, J. Math. Phys. **19**(2), 461 (1978).
[2] E. H. Lieb, Rev. Mod. Phys. **48**, 553 (1976).
[3] M. G. Cowling and J. F. Price (submitted to SIAM J. Math. Anal.).
[4] I. I. Hirschman, Jr., Amer. J. Math. **79**, 152 (1957).
[5] S. Wainger, Mem. Amer. Math. Soc. **59** (1965).

Superposition rules for nonlinear coupled first-order differential equations

George L. Strobel
Physics Department, The University of Georgia, Athens, Georgia 30602

(Received 16 November 1982; accepted for publication 11 February 1983)

Pairs of coupled nonlinear differential equations of the polynomial type have been studied. No higher powers than quadratic were considered. Lie's theorem provides a superposition rule exists if certain operators generate a finite-dimensional Lie algebra. The Lie algebras possible were divided into twenty categories. For each case the general form of the coupled differential equations is obtained. The coupled differential equations are then separated where possible. Solutions are obtained expressed in terms of a finite number of particular solutions.

PACS numbers: 02.30.Jr

Consider a system of two first-order differential equations of the nonautonomous type.

$$\dot{x} = \xi(x,y,t), \qquad (1)$$
$$\dot{y} = \eta(x,y,t),$$

where $\dot{x}$ denotes dx/dt. We are interested in determining superposition rules for such equations. Lie's theorem[1] provides that a superposition rule exists for Eq. (1) if and only if the functions $\xi(x,y,t)$ and $\eta(x,y,t)$ have the separated form:

$$\xi = \sum_{i=i} \xi_i(x,y) Z_i(t), \qquad (2)$$
$$\eta = \sum_{i=i}^{l} \eta_i(x,y) Z_i(t),$$

and where the differential operators

$$Y_i \equiv \xi_i \frac{\partial}{\partial x} + \eta_i \frac{\partial}{\partial y} \qquad (3)$$

generate a finite-dimensional Lie algebra. This means the operators Y_i must satisfy the commutator relation

$$[Y_i, Y_j] = \sum_k C_{ijk} Y_k, \qquad (4)$$

where C_{ijk} are the structure constants of the Lie group.

We now consider the case where $\xi_i(x,y)$ and $\eta_i(x,y)$ are polynomials containing no higher power in x and y than quadratic.[2] We seek to determine all the distinct sets of differential equations that satisfy Eq. (4). A superposition principle involving a finite number of particular solutions will be sought for each case.

For future convenience, we now list the six constant and linear operators L_i, the six quadratic operators Q_i and also eight cubic operators C_i (see Table I). We are considering only quadratic or lower operators to be present in the Y_i operators of Eq. (3), but cubic terms appear in the commutators of certain quadratic operators. From a table of commutators of the linear and quadratic operators, we will determine all the sets of operators that generate finite-dimensional Lie algebras. The associated differential equations and their superposition rule will then be deduced.

In Table II the commutator is presented for the operators of lower than cubic power in two dimensions. The commutator is defined as $[A,B] = AB - BA$.

We now consider commutators of operators, both of which are quadratic. Inspection of the table of commutators shows that Q_1 and Q_6 commute. Any Lie algebra containing two or more quadratic operators must have the quadratic operators commute. This follows from inspection of the commutator table, as all nonvanishing commutators of quadratic operators involve cubic operators. As a result, the commutators of the quadratic operators must commute to form a finite-dimensional Lie algebra. Other pairs of quadratic operators commute beside Q_1 and Q_6. The quadratic operators that commute are

$$[Q_1, Q_6] = 0, \qquad (5)$$
$$[Q_3, (2Q_2 + Q_6)] = 0, \qquad (6)$$
$$[Q_4, (Q_1 + 2Q_5)] = 0, \qquad (7)$$

TABLE I. Linear quadratic and cubic operators.

	Constant or linear operator	Quadratic operator	Cubic operator
i	L_i	Q_i	C_i
1	$\frac{\partial}{\partial x}$	$x^2\frac{\partial}{\partial x}$	$x^3\frac{\partial}{\partial x}$
2	$x\frac{\partial}{\partial x}$	$xy\frac{\partial}{\partial x}$	$x^2y\frac{\partial}{\partial x}$
3	$y\frac{\partial}{\partial x}$	$y^2\frac{\partial}{\partial x}$	$xy^2\frac{\partial}{\partial x}$
4	$\frac{\partial}{\partial y}$	$x^2\frac{\partial}{\partial y}$	$y^3\frac{\partial}{\partial x}$
5	$x\frac{\partial}{\partial y}$	$xy\frac{\partial}{\partial y}$	$x^3\frac{\partial}{\partial y}$
6	$y\frac{\partial}{\partial y}$	$y^2\frac{\partial}{\partial y}$	$x^2y\frac{\partial}{\partial y}$
7			$xy^2\frac{\partial}{\partial y}$
8			$y^3\frac{\partial}{\partial y}$

TABLE II. Commutators of linear and quadratic operators.

	L_1	L_2	L_3	L_4	L_5	L_6	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6
L_1	0	L_1	0	0	L_4	0	$2L_2$	L_3	0	$2L_5$	L_6	0
L_2	$-L_1$	0	$-L_3$	0	L_5	0	Q_1	0	$-Q_3$	$2Q_4$	Q_5	0
L_3	0	L_3	0	$-L_1$	L_6-L_2	$-L_3$	$2Q_2$	Q_3	0	$2Q_5-Q_1$	Q_6-Q_2	$-Q_3$
L_4	0	0	L_1	0	0	L_4	0	L_2	$2L_3$	0	L_5	$2L_6$
L_5	$-L_4$	$-L_5$	L_2-L_6	0	0	L_5	$-Q_4$	Q_1-Q_5	$2Q_2-Q_6$	0	Q_4	$2Q_5$
L_6	0	0	L_3	$-L_4$	$-L_5$	0	0	Q_2	$2Q_3$	$-Q_4$	0	Q_6
Q_1	$-2L_2$	$-Q_1$	$-2Q_2$	0	Q_4	0	0	$-C_2$	$-2C_3$	$2C_5$	C_6	0
Q_2	$-L_3$	0	$-Q_3$	$-L_2$	Q_5-Q_1	$-Q_2$	C_2	0	$-C_4$	$2C_6-C_1$	C_7-C_2	$-C_3$
Q_3	0	Q_3	0	$-2L_3$	Q_6-2Q_2	$-2Q_3$	$2C_3$	C_4	0	$2C_7-2C_2$	C_8-2C_3	$-2C_4$
Q_4	$-2L_5$	$-2Q_4$	Q_1-2Q_5	0	0	Q_4	$-2C_5$	C_1-2C_6	$2C_2-2C_7$	0	C_5	$2C_6$
Q_5	$-L_6$	$-Q_5$	Q_2-Q_6	$-L_5$	$-Q_4$	0	$-C_6$	C_2-C_7	$2C_3-C_8$	$-C_5$	0	C_7
Q_6	0	0	Q_3	$-2L_6$	$-2Q_5$	$-Q_6$	0	C_3	$2C_4$	$-2C_6$	$-C_7$	0

$$[(Q_1 + Q_5), (Q_2 + Q_6)] = 0, \tag{8}$$

$$[(q_1 + nQ_3 + 2Q_5), (2Q_2 + Q_6 + Q_4/n)] = 0. \tag{9}$$

In Eq. (9), n is a constant, independent of x and y, and not equal to zero.

By systematically exchanging x and y, Eqs. (6) and (7) are seen to transform into each other. Equation (7) is ignored in the following because the coupled differential equations and their superposition rule found for Eq. (6) can be converted, upon exchange of x and y into the differential equation and superposition rule for Eq. (7).

Certain combinations of the linear operators can be added to these pairs of commuting quadratic operators to form a finite-dimensional Lie algebra. No other combination of quadratic operators can be added to the quadratic pairs. The sets of operators forming finite-dimensional Lie algebras with two commuting quadratic operators are

(I) $Q_1, Q_6, L_1, L_2, L_4, L_6,$

(II) $Q_3, 2Q_2 + Q_6, L_1, L_2, L_3, L_4, L_6,$

(III) $Q_1 + Q_5, Q_2 + Q_6, L_1, L_2, L_3, L_4, L_5, L_6,$

(IV) $Q_1 + nQ_3 + 2Q_5, 2Q_2 + Q_6 + Q_4/n, L_2 + L_6$ if $n \neq 1$,

(V) $Q_1 + nQ_3 + 2Q_5, 2Q_2 + Q_6 + Q_4/n, L_1, L_2, + L_6, L_4, L_3 + L_5$ if $n = 1$.

The sets of operators forming a finite-dimensional Lie algebra with one quadratic operator included are

(VI) $Q_1, L_1, L_2, L_4, L_6,$

(VII) $Q_2, L_2, L_4, L_6,$

(VIII) $Q_3, L_1, L_2, L_3, L_4, L_6,$

(IX) $Q_1 + Q_5, L_1, L_2, L_4, L_5, L_6,$

(X) $Q_1 + Q_4, L_1, L_4, L_2 + L_5, L_2 + L_6,$

(XI) $Q_2 + Q_3, L_2 + L_3, L_2 + L_6, L_3 - L_6, L_1 - L_4,$

(XII) $n_2Q_2 + n_3Q_3 + n_2Q_6, L_1, L_3, L_2 + L_6,$

(XIII) $n_1Q_1 + n_3Q_3 + 2n_1Q_5 + n_6Q_6, L_1, L_2 + L_6,$

(XIV) $n_1Q_1 - 2n_3Q_2 + n_3Q_3 + n_4Q_4 - 2n_4Q_5 + (n_1 - n_3 + n_4)Q_6, L_{1+}L_4, L_2 + L_6,$

(XV) $n_1Q_1 + 2n_3Q_2 + n_3Q_3 + n_4Q_4 + 2n_4Q_5 + (n_4 + n_3 - n_1)Q_6, L_1 - L_4, L_2 + L_6,$

(XVI) $n_1Q_1 + (n_1 + n_3)Q_2 + n_3Q_3 + n_3Q_4 + (n_1 + n_3)Q_5 + n_1Q_6, L_1 + L_4, L_2 + L_6, L_3 + L_5,$

(XVII) $Q_1 - 2Q_2 + Q_3 - Q_4 + 2Q_5 - Q_6, L_1 + L_4, L_2 + L_6, L_3 + L_5, L_1 - L_4,$

(XVIII) $n_1Q_1 + (n_2 - n_1)Q_2 - n_2Q_3 + n_2Q_4 + (n_1 - n_2)Q_5 - n_1Q_6, L_1 - L_4, L_2 + L_6, L_3 + L_5,$

(XIX) $Q_1 + 2Q_2 + Q_3 + Q_4 + 2Q_5 + Q_6, L_1 + L_4, L_1 - L_4, L_2 + L_6, L_3 + L_5,$

(XX) $n_1Q_1 + n_2Q_2 + n_3Q_3 + n_4Q_4 + n_5Q_5 + n_6Q_6, L_2 + L_6.$

The systematic interchange of x and y in the operator sets (VI)–(XIII) will result in similar Lie algebras. We ignore such similar Lie algebras because the corresponding differential equations can be analyzed by interchanging x and y in the cases considered. In cases (XII)–(XX), the n_i stand for constant arbitrary coefficients. The last 15 cases have a single linear combination of quadratic operators as one element. These cases are distinguished from each other by the different linear operators with which they form a Lie algebra. The operator $L_2 + L_6$ will form a finite-dimensional Lie algebra with any linear combination of the quadratic operators listed in Table II. Such combinations are treated collectively in set (XX). Cases where another linear operator, in addition to $L_2 + L_6$, can form a Lie algebra are separately treated.

Finite-dimensional Lie algebras with no quadratic operators can also be developed. But the resulting differential equations are linear, and the corresponding linear superposition rules are presumed to be so well known to not be of interest here. We now proceed to develop nonlinear superposition rules for each case permitted by Lie's theorem.

Case (I): The differential equations corresponding to this case are

$$\dot{x} = a + bx + cx^2, \tag{10}$$

$$\dot{y} = f + gy + hy^2. \tag{11}$$

These equations are recognized as uncoupled Riccati equations. In this and what follows a, b, c, f, g, and h stand for arbitrary functions of time, t, the independent variable. These were referred to as $Z_i(t)$ in Eq. (2). P and K stand for arbitrary constants of integration. Particular solutions to a differential equation will have a subscript appended to them. The superposition rule for this case is

$$(x - x_1)(x_3 - x_2)/(x - x_2)(x_3 - x_1) = P. \tag{12}$$

and

$$(y - y_1)(y_3 - y_2)/(y - y_2)(y_3 - y_1) = K. \tag{13}$$

The general solution to Eqs. (10) and (11) is expressed in a superposition rule requiring six particular solutions. Lie's theorem has been shown correct with no work on our part for this case. We will not be so fortunate for the other cases. However, it is well to point out one valuable aspect of Lie's theorem. If the conditions of Lie's theorem are satisfied, one is encouraged to continue seeking a superposition rule with a finite number of particular solutions. If the operators do not form a finite-dimensional Lie algebra, one need not even attempt to seek a superposition principle.

Case (II): The differential equations for this Lie algebra are

$$\dot{x} = a + bx + cy + fy^2 + 2gxy, \tag{14}$$

$$\dot{y} = h + gy + gy^2. \tag{15}$$

This pair of equations is coupled. One could proceed by solving the $\dot{y}$ equation as a separate Riccati equation as

$$(y - y_1)(y_3 - y_2)/(y - y_2)(y_3 - y_1) = P. \tag{16}$$

Then with y known from Eq. (16), the $\dot{x}$ equation becomes the linear inhomogeneous equation:

$$\dot{x} + (b + 2gy)x + (a + cy + fy^2) \equiv Qx + R. \tag{17}$$

The general solution to Eq. (17) can be written as

$$x = Ke^{\int Q\,dt} + e^{\int Q\,dt} \int \mathrm{Re}^{-\int Q\,dt'}\,dt, \tag{18}$$

but Lie's theorem states we can do better than Eq. (18)! What is difficult with Eq. (18) as it stands? The general solution for y involves an arbitrary integration constant, P, R, and Q appearing in the solution for x, Eq. (18), dependent on P; thus Eq. (18) would have to be integrated for each value of P!

To obtain a superposition rule with a finite number of particular solutions, let

$$y = u + y_1, \tag{19}$$

where y_1 is a given particular solution to Eq. (15). Then u satisfies

$$\dot{u} = (q + 2gy_1)u + gu^2, \tag{20}$$

which has a superposition rule

$$(1/u) = (1/u_1) + C(1/u_2 - 1/u_1). \tag{21}$$

Here u_1 and u_2 are particular solutions of Eq. (20) and C is an arbitrary constant of integration.

We now define

$$x = vu^2 \tag{22}$$

and substitute this into Eq. (14). Using Eqs. (19) and (20), the result can be written as

$$\dot{v} = \left[\frac{A}{u^2} + \frac{c + 2fy_1}{u} + f\right] + [b - 2(q + 2gy_1)]v \equiv R' + Q'v, \tag{23}$$

where we have defined

$$A \equiv a + cu_1 + fy_1^2. \tag{24}$$

The general solution of Eq. (23) is

$$v = K'e^{\int Q'\,dt} + e^{\int Q'\,dt} \int R' e^{-\int Q'\,dt}\,dt. \tag{25}$$

The integrand Q' is independent of C, the arbitrary constant of integration for u, that appears in Eq. (21). The other integrand, R', has the simple form

$$R' \equiv R_2 C^2 + R_1 C + R_0, \tag{26}$$

as can be seen from comparing Eqs. (21) and (23). R_0, R_1, and R_2 are functions of time, but are independent of C. The general solution v to Eq. (25) depends on two constants of integration, K and C, where we drop the primes. Thus we now indicate this dependence by $v \equiv v_{KC}$.

Then we have the linear superposition rule, from Eqs. (25) and (26), that

$$v_{KC_0} = v_{K0} + \left(\frac{C_0^2}{C^2 - C}\right)(v_{0C} - v_{00} - Cv_{01} + Cv_{00}) + \left(\frac{C_0}{C^2 - C}\right)[C^2 v_{01} - C^2 v_{00} - v_{0C} + v_{00}]. \tag{27}$$

Thus v has a superposition rule with four particular solutions. Then a knowledge of u_1, u_2 completely determines x via

$$x = vu^2, \tag{22}$$

where u is given by Eq. (21). Thus, x and y can be determined in terms of three particular solutions of Eq. (15), two particular solutions of Eq. (20), and four particular solutions of Eq. (23).

Case (III): The differential equations corresponding to this Lie algebra are

$$\dot{x} = ax + by + gx^2 + fxy + h, \tag{28}$$

$$\dot{y} = cx + dy + gxy + fy^2 + q. \tag{29}$$

TABLE III. Cases separable by a linear transformation.

Case	Differential equation	Superposition rule	Comment
(IV)	$\dot{x} = ax + bx^2 + b\lambda^2 y^2 + 2cxy$	$\frac{1}{S} = \frac{1}{S_1} + P\left(\frac{1}{S_2} - \frac{1}{S_1}\right)$	$\alpha = 1$
	$\dot{y} = ay + 2bxy + cy^2 + cx^2/\lambda^2$	$\frac{1}{m} = \frac{1}{m_1} + k\left(\frac{1}{m_2} - \frac{1}{m_1}\right)$	λ = constant not 0 or 1
(V)	$\dot{x} = ax + bx^2 + by^2 + 2cxy + dy + h$	Riccati for S	$\alpha = 1$
	$\dot{y} = ay + 2bxy + cx^2 + cy^2 + dx + f$	Riccati for m	$\lambda = 1$
(X)	$\dot{x} = ax^2 + bx + cx + g$	Riccati for x	$m = x - y$ only transformation
	$\dot{y} = ax^2 + bx + cy + d$	$m = m_1 + K(m_2 - m_1)$	
(XI)	$\dot{x} = axy + ay^2 - by + cx + dx + dy + f$	$y = y_1 + K(y_2 - y_1)$	$S = x + y$ only transformation
	$\dot{y} = by + cy - f$	$S_{pk} = PS_{10}(S_{11}/S_{10})^k$	
(XIII)	$\dot{x} = ax + b(n_1x^2 + n_3y^3) + c$	Riccati for s for both values of α	$\alpha = \frac{-n_6 \pm (n_6^2 + 4n_1n_3)^{1/2}}{2n_3}$
	$\dot{y} = ay + b(2n_1xy + n_6y^2)$		$\lambda = 1$
(XIV)	$\dot{x} = ax + b(n_1x^2 - 2n_3xy + n_3y^2) + c$	Riccati for S for both values of λ	$\alpha = 1$
	$\dot{y} = ay + b(n_4x^2 - 2n_4xy + [n_1 + n_4 - n_3]y^2) + c$		$\lambda = \frac{-n_1 - n_4 \pm [(n_1 + n_4)^2 - 4n_3n_4]^{1/2}}{2n_4}$
(XV)	$\dot{x} = ax + b(n_1x^2 + 2n_3xy + n_3y^2) + c$	Riccati for S for both values of λ	$\alpha = 1$
	$\dot{y} = ay + b(n_4x^2 + 2n_4xy + [n_3 + n_4 - n_1]y^2) - c$		$\lambda = \frac{n_4 - n_1 \pm [(n_4 - n_1)^2 + 4n_3n_4]^{1/2}}{2n_4}$
(XVII)	$\dot{x} = ax + b(x^2 - 2xy + y^2) + cy + d + g$	$S = S_1 + K(S_2 - S_1)$	$\alpha = 1$
	$\dot{y} = ay - b(x^2 - 2xy + y^2) + cx + d - g$	Riccati for m	$\lambda = 1$
(XIX)	$\dot{x} = ax + b(x^2 + 2xy + y^2) + cy + d + g$	Riccati for S	$\alpha = 1$
	$\dot{y} = ay + b(x^2 + 2xy + y^2) + cx + d - g$	$m = m_1 + P(m_2 - m_1)$	$\lambda = 1$
(XX)	$\dot{x} = ax + b(n_1x^2 + n_2xy + n_3y^2)$	Riccati for S	$\alpha/\lambda = \frac{2n_1 - n_5 \pm [(n_5 - 2n_1)^2 + 8n_2n_4]^{1/2}}{2n_2}$
	$\dot{y} = ay + b(n_4x^2 + n_5xy + n_6y^2)$	Riccati for x	

This pair of equations can be written in the form

$$\dot{U}_i = A_i + \sum_j B_{ij} U_j + \sum_j C_j U_j U_i, \tag{30}$$

where

$$U_1 = x, \quad U_2 = y.$$

Thus Eqs. (28) and (29) are a set of coupled Riccati equations of the projective type in two dimensions. The superposition rule has been given by Anderson.[3]

Other cases: In Table III we present the differential equations for the cases that uncouple using the transformation

$$\begin{aligned} S &= \alpha x + \lambda y, \\ m &= \alpha x - \lambda y. \end{aligned} \tag{31}$$

Upon using these transformations, uncoupled equations for S and m are obtained, and the resulting superposition rules are also indicated in the table. P and K are arbitrary constants of integration. The letters a–h, and also q denote arbitrary known coefficients appearing in the differential equations that can be functions of time. The letter n denotes a known constant coefficient. The Riccati superposition rule for x is analogous to Eq. (12).

A product transformation will separate other cases. The corresponding differential equations and the resulting superposition rules are in Table IV.

Finally, the remaining cases are already uncoupled. These cases are listed in Table V, along with the corresponding superposition rules that apply.

TABLE IV. Cases separable by a product transformation.

Case	Differential equation	Superposition rule	Comment
(IX)	$\dot{x} = ax^2 + gx + f$	Riccati in x	$y = v(x - x_1)$
	$\dot{y} = axy + by + cx + d$	linear in v	
(XII)	$\dot{x} = ax + b(n_2xy + n_3y^2) + cy + d$	$\frac{1}{y} = \frac{1}{y_1} + k\left(\frac{1}{y_2} - \frac{1}{y_1}\right)$	$x = vy$
	$\dot{y} = ay + bn_2y^2$	linear in v	
(XVI)	$\dot{x} = ax + b[n_1x^2 + (n_1 + n_3)xy + n_3y^2] + cy + d$	Riccati in S	$S = x + y$
	$\dot{y} = ay + b[n_3x^2 + (n_1 + n_3)xy + n_1y^2] + cx + d$	linear in v	$v(S - S_1) = \frac{(n_1 - n_3)(x - y)}{(n_1 + n_3)}$
(XVIII)	$\dot{x} = ax + b[n_1x^2 + (n_2 - n_1)xy - n_2y^2] + cy + d$	Riccati in S	$S = x - y$
	$\dot{y} = ay + b[n_2x^2 + (n_1 - n_2)xy - n_1y^2] + cx - d$	linear in v	$v(S - S_1) = \frac{(n_1 + n_2)(x + y)}{(n_1 - n_2)}$

TABLE V. Uncoupled cases.

Case	Differential equation	Superposition rule	Comment
(I)	$\dot{x} = a + bx + cx^2$	Riccati in x	
	$\dot{y} = f + gy + hy^2$	Riccati in y	
(VI)	$\dot{x} = ax^2 + dx + f$	Riccati in x	This is case (i) with $h = 0$
	$\dot{y} = by + c$	linear in y	
(VII)	$\dot{x} = axy + dx$	linear in y	
	$\dot{y} = by + c$	once y known, linear in x	
(VIII)	$\dot{x} = ay^2 + dy + gx + f$	linear in y	
	$\dot{y} = by + c$	once y known, linear in x	

SUMMARY

Lie's theorem provides that a superposition rule exists for a set of differential equations if certain operators generate a finite-dimensional Lie algebra. Pairs of coupled nonlinear differential equations of the polynomial type have been studied here. No polynomial terms of higher power than the quadratic were considered. As a result, the Lie algebras possible can be divided into 20 cases. Five of these cases contain various pairs of commuting quadratic generators. The remaining cases each have a single quadratic generator, plus various linear generators. If two quadratic generators are present in a given Lie algebra, they must commute with each other. This helps limit the number of Lie algebras to be considered. For each case, the general form for the coupled differential equations is written down. The differential equations contain an arbitrary function of time, the independent variable, for each generator of the Lie algebra. Changes of the dependent variables are sought for each case, to produce uncoupled differential equations. Superposition rules for these uncoupled differential equations then permit the solutions of the coupled equations to be determined.

ACKNOWLEDGMENT

I wish to thank the physics department at UAB, Birmingham, Alabama for their hospitality where a portion of this work was done.

[1]S. Lie and G. Scheffers, *Vorlesungen uber Continuierliche Gruppen* (Teubner, Leipzig, 1893).

[2]E. H. Kerner, J. Math. Phys. **22**, 1366–1371 (1981); Carmen Chicone and Tian Jinghuang, Amer. Math. Monthly **89** (3), 167–177 (1982).

[3]R. L. Anderson, Lett. Math. Phys. **4**, 1–7 (1980).

Nonlinear evolution equations and nonabelian prolongations

M. Leo, R. A. Leo, G. Soliani, and L. Solombrino
Dipartimento di Fisica dell'Università, Lecce, Italy and Istituto Nazionale di Fisica Nucleare, Italy
L. Martina
Dipartimento di Fisica, Università della Calabria, Arcavacata di Rende, Cosenza, Italy

(Received 27 September 1982; accepted for publication 18 November 1982)

A systematic analysis of the class of nonlinear evolution equations $u_t + u_{xxx} + \phi(u, u_x) = 0$ is carried out within the Estabrook–Wahlquist prolongation scheme.

PACS numbers: 02.30.Jr, 11.10.Lm

I. INTRODUCTION

In the study of nonlinear evolution (NLE) equations based on the Estabrook–Wahlquist (EW) prolongation scheme[1] one finds that all NLE equations which are completely integrable admit a nonabelian prolongation.[1–3] This fact suggests that there is a connection between complete integrability and existence of nonabelian prolongations. Since a general theory of such a correspondence is not available at present, the accumulation of cases might open the way towards the construction of the theory.[4,5]

In this context, this paper is devoted to a prolongation analysis of the class of equations

$$u_t + u_{xxx} + \phi(u, u_x) = 0, \tag{1.1}$$

where $\phi(u, u_x)$ denotes a function of the variables u and u_x. Precisely, we determine all the functions $\phi(u, u_x)$ such that Eq. (1.1) admits a nonabelian prolongation structure. Such a function turns out to be of polynomial form in u_x (see Ref. 6).

On the one hand, we obtain known NLE equations which can be solved by the inverse scattering method, and whose associated nonabelian Lie algebras are presumably infinite-dimensional (as it happens, for example, for the Korteweg–de Vries equation[7]). On the other, we provide some new NLE equations with which one can associate *finite*-dimensional nonabelian Lie algebras.

A notable feature of our results is the fact that some NLE equations belonging to the class (1.1) have a prolongation structure whose pseudopotential cannot be of the first kind.[8]

In Sec. II we derive a set of equations which are essential to treat Eq. (1.1). In Sec. III and IV we perform a systematic analysis of the fundamental equations found in Sec. II, and obtain explicit forms of $\phi(u, u_x)$ such that Eq. (1.1) admits a nonabelian prolongation structure. Section V deals with some concluding remarks, and Appendices A, B, and C contain details of calculations.

II. BASIC EQUATIONS

Let us introduce the set of variables $\{y^i\}$ defined by the prolongation equations

$$y^i_x = F^i(u, y^j), \tag{2.1a}$$

$$y^i_t = G^i(u, u_x, u_{xx}, y^j), \tag{2.1b}$$

where $i, j = 1, 2, ..., N$.

We recall that when the set of variables $\{y^i\}$ has one element only, say y, it will be called "pseudopotential of the first kind."[8]

The integrability conditions for Eqs. (2.1), which assure that $y^i_{xt} = y^i_{tx}$, are given by[4]

$$D_x G^i = D_t F^i \quad (i = 1, 2, ..., N), \tag{2.2}$$

where

$$D_x = u_x \frac{\partial}{\partial u} + u_{xx} \frac{\partial}{\partial u_x} + u_{xxx} \frac{\partial}{\partial u_{xx}} + F^j \frac{\partial}{\partial y^j}, \tag{2.3}$$

and

$$D_t = u_t \frac{\partial}{\partial u} + G^j \frac{\partial}{\partial y^j}. \tag{2.4}$$

Taking account of (1.1), Eq. (2.2) becomes

$$(F^i_u + G^i_p)r + F^i_u \phi + [F, G]^i + G^i_u z + G^i_z p = 0, \tag{2.5}$$

where $z = u_x$, $p = u_{xx}$, $r = u_{xxx}$, and

$$[F, G]^i = F^j \frac{\partial G^i}{\partial y^j} - G^j \frac{\partial F^i}{\partial y^j}. \tag{2.6}$$

In the following we shall omit for simplicity the indexes, i.e., we shall write F instead of F^i, F_u instead of F^i_u, and so on.

From the requirement that Eq. (2.5) be identically satisfied for any value of the independent variables u, z, p, and r, one has

$$\tfrac{1}{2}z^3 F_{uuu} + \tfrac{3}{2}z^2 [F, F_{uu}] + z([F, [F, F_u]] - L_u) - [F, L] + F_u \phi(u, z) = 0, \tag{2.7}$$

where L is a function of integration depending on u and y only, related to G by[4]

$$G = -pF_u + \tfrac{1}{2}z^2 F_{uu} + z[F, F_u] - L. \tag{2.8}$$

Since (2.7) holds for any z, one deduces that $\phi(u, z)$ is a third-degree polynomial in the z variable, namely,

$$\phi(u, z) = g(u) + h(u)z - \tfrac{3}{2}k(u)z^2 - \tfrac{1}{2}m(u)z^3. \tag{2.9}$$

Substituting (2.9) in (2.7) and equating to zero the coefficients of the powers of z, we have the set of basic equations:

$$F_{uuu} = mF_u, \tag{2.10a}$$

$$[F, F_{uu}] = kF_u, \tag{2.10b}$$

$$L_u = [F, [F, F_u]] + hF_u, \tag{2.10c}$$

$$[F, L] = gF_u. \tag{2.10d}$$

In the following, we shall perform a full analysis of these equations in order to have some information about the functions F, L, and G [see (2.8)], which depend on u and y,

 0022-2488/83/071720-11$02.50

and the coefficients m, k, h, and g of the polynomial (2.9), which depend on the variable u only. This goal can be achieved distinguishing essentially two cases, which is based on the assumption that F_u and F_{uu} may be linearly independent vector functions, or not.

III. CASE WHERE F_u AND F_{uu} ARE LINEARLY INDEPENDENT

Let us suppose that F_u and F_{uu} are linearly independent. In order to treat the basic equations (2.10), it is convenient to distinguish the subcases (I) $m_u \neq 0$, $k \neq 0$; (II) $m = \text{const} \neq 0, k = 0$; (III) $m = 0, k = \text{const} \neq 0$; (IV) $m = 0$, $k = 0$; (V) $m_u \neq 0$, $k = 0$. (We have seen that the remaining subcases, such as for example $m = 0$, $k_u \neq 0$, fall into the preceding ones).

First we deal with the first subcase

(I) $m_u \neq 0$, $k \neq 0$

Then, the following relations hold (see Appendix A):

$$[F_u, F_{uu}] = 0, \tag{3.1}$$

$$k m_u = 2 m k_u, \tag{3.2a}$$

$$2k_u^2 = k k_{uu}, \tag{3.2b}$$

$$g_u + (k_u/k)g - (2k_u/m_u)h_u = 0, \tag{3.3a}$$

$$g_{uu} - (k_{uu}/m_u)h_u = 0. \tag{3.3b}$$

We point out that the commutator relation (3.1) is not valid when the pseudopotential is of the first kind. In fact, in this case it turns out that F_u and F_{uu} are proportional. This contradicts our assumption that F_u and F_{uu} are linearly independent.

From Eq. (3.2b) one has

$$k = \tilde{k}e^{-w}, \tag{3.4}$$

where

$$e^w = u - u_0, \tag{3.5}$$

$\tilde{k}$ and u_0 being arbitrary constants.

On the other hand, Eq. (3.2a) gives

$$m = \tilde{m}e^{-2w}, \tag{3.6}$$

where $\tilde{m}$ is a constant.

Then, Eq. (2.10a) yields

$$F_{www} - 3F_{ww} + (2 - \tilde{m})F_w = 0, \tag{3.7}$$

where $F_w = F_u e^w$.

The characteristic equation associated with (3.7) has the roots $\lambda = 0$ and $\lambda_{1,2} = \frac{1}{2}[3 \pm (1 + 4\tilde{m})^{1/2}]$. We can therefore distinguish three cases, specifically,

$$\tilde{m} \neq 2, \ -\tfrac{1}{4} \tag{3.8a}$$

$$\tilde{m} = 2, \tag{3.8b}$$

$$\tilde{m} = -\tfrac{1}{4}. \tag{3.8c}$$

Case (a)

For brevity's sake, we shall handle in some detail only the case (a), for which Eq. (3.7) admits the solution

$$F = e^{\lambda_1 w}A + e^{\lambda_2 w}B + C, \tag{3.9}$$

where $\lambda_1 \neq \lambda_2$ ($\lambda_1, \lambda_2 \neq 0$) and A, B, C are arbitrary vector functions of the variables y's only.

Inserting (3.9) in Eqs. (3.1) and (2.10b), we find the commutator relations

$$[A, B] = 0, \tag{3.10a}$$

$$[A, C] = [\tilde{k}/(1 - \lambda_1)]A, \tag{3.10b}$$

$$[B, C] = [\tilde{k}/(1 - \lambda_2)]B. \tag{3.10c}$$

Using (3.4), Eqs. (3.3a) and (3.3b) yield

$$g = \tilde{g}e^w + \hat{g}e^{-w} \tag{3.11}$$

and

$$h = (\tilde{m}/\tilde{k})e^{-2w} + \tilde{h}, \tag{3.12}$$

where $\tilde{g}$, $\hat{g}$ and $\tilde{h}$ are constants.

Now we shall resort to Eq. (2.10c) in order to derive $L(u, y)$.

In virtue of (3.9), (A5) (see Appendix A), and (3.12), integrating (2.10c) with respect to u, we have

$$L = \left[\frac{\tilde{k}^2}{(\lambda_1 - 1)^2} + \tilde{h} \right] e^{\lambda_1 w}A + \left[\frac{\tilde{k}^2}{(\lambda_2 - 1)^2} + \tilde{h} \right] e^{\lambda_2 w}B + \frac{\tilde{m}\hat{g}}{\tilde{k}} \left(\frac{\lambda_1}{\lambda_1 - 2} e^{(\lambda_1 - 2)w}A + \frac{\lambda_2}{\lambda_2 - 2} e^{(\lambda_2 - 2)w}B \right) + D, \tag{3.13}$$

where D is a vector function of integration of the y's only.

Substituting (3.13) in (2.10d), with the help of (3.10) we obtain a relation which is a linear combination of the independent quantities $e^{\lambda_1 w}$, $e^{\lambda_2 w}$, $e^{(\lambda_1 - 2)w}$, $e^{(\lambda_2 - 2)w}$, and 1, through coefficients which are vector functions of the y's only.

Equating to zero these coefficients, we obtain the commutator relations

$$[C, D] = 0, \tag{3.14a}$$

$$[A, D] = \left[-\frac{\tilde{k}}{\lambda_1 - 1} \left(\frac{\tilde{k}^2}{(\lambda_1 - 1)^2} + \tilde{h} \right) + \tilde{g}\lambda_1 \right] A, \tag{3.14b}$$

$$[B, D] = \left[-\frac{\tilde{k}}{\lambda_2 - 1} \left(\frac{\tilde{k}^2}{(\lambda_2 - 1)^2} + \tilde{h} \right) + \tilde{g}\lambda_2 \right] B. \tag{3.14c}$$

The set of commutator relations (3.10) and (3.14) defines a finite-dimensional 4D nonabelian Lie algebra with 2D abelian derived algebra,[9] which is the *prolongation algebra* associated with the NLE equation

$$u_t + u_{xxx} + \frac{\tilde{g}}{u - u_0} + \hat{g}(u - u_0) + \left[\tilde{h} + \frac{\tilde{m}\hat{g}}{\tilde{k}} \frac{1}{(u - u_0)^2} \right] u_x - \frac{3}{2} \frac{\tilde{k}}{u - u_0} u_x^2 - \frac{\tilde{m}}{2(u - u_0)^2} u_x^3 = 0. \tag{3.15}$$

This equation is obtained from Eq. (1.1) substituting (3.11), (3.12), (3.4), and (3.6) in (2.9), having in mind (3.5).

To give further insights into the algebra (3.14), let us consider the quantity

$$\Delta = \begin{vmatrix} \dfrac{\tilde{k}}{\lambda_1 - 1} & \dfrac{\tilde{k}}{\lambda_2 - 1} \\ \dfrac{\tilde{k}}{\lambda_1 - 1}\left[\dfrac{\tilde{k}^2}{(\lambda_1 - 1)^2} + \tilde{h}\right] - \tilde{g}\lambda_1 & \dfrac{\tilde{k}}{\lambda_2 - 1}\left[\dfrac{\tilde{k}^2}{(\lambda_2 - 1)^2} + \tilde{h}\right] - \tilde{g}\lambda_2 \end{vmatrix}$$

$$= -(1 + 4\tilde{m})^{1/2} \frac{\tilde{k}}{\tilde{m}^3} (\tilde{k}^3 + 2\tilde{g}\tilde{m}^2). \tag{3.16}$$

We may distinguish the following cases: (i) $\tilde{k}^3 \neq -2\tilde{g}\tilde{m}^2$ and (ii) $\tilde{k}^3 = -2\tilde{g}\tilde{m}^2$. In the case (i), we have $\Delta \neq 0$ and the algebra (3.14) can be written as

$$\begin{aligned} &[C', B] = B, \quad [D', A] = A, \\ &[C', A] = 0, \quad [D', B] = 0, \\ &[B, A] = 0, \quad [D', C'] = 0, \end{aligned} \tag{3.17}$$

where the (linearly independent) elements C' and D' are given by

$$C' = \frac{1}{\Delta}\left\{\frac{\tilde{k}}{\lambda_1 - 1} D - \left[\frac{\tilde{k}}{\lambda_1 - 1}\left(\frac{\tilde{k}^2}{(\lambda_1 - 1)^2} + \tilde{h}\right) - \tilde{g}\lambda_1\right] C\right\}, \tag{3.18}$$

and

$$D' = \frac{1}{\Delta}\left\{-\frac{\tilde{k}}{\lambda_2 - 1} D + \left[\frac{\tilde{k}}{\lambda_2 - 1}\left(\frac{\tilde{k}^2}{(\lambda_2 - 1)^2} + \tilde{h}\right) - \tilde{g}\lambda_2\right] C\right\}. \tag{3.19}$$

The algebra (3.17) is the direct sum of two 2D nonabelian algebras which correspond to the standard form II (b) according to Jacobson's classification.[9]

In the case (ii), we have $\Delta = 0$ and the algebra (3.14) takes the form

$$\begin{aligned} &[A, B] = 0, \quad [D'', A''] = 0, \\ &[C'', A] = A, \quad [D'', B] = 0, \\ &[C'', B] = [(\lambda_1 - 1)/(\lambda_2 - 1)]B, \quad [D'', C''] = 0, \end{aligned} \tag{3.20}$$

where

$$C'' = \frac{\lambda_1 - 1}{\tilde{k}} C \tag{3.21}$$

and

$$D'' = D - \left[\frac{\tilde{k}^2}{(\lambda_2 - 1)^2} + \tilde{h} - \tilde{g}\frac{\lambda_2}{\tilde{k}}(\lambda_2 - 1)\right] C. \tag{3.22}$$

The algebra (3.20) is the direct sum of a 1D algebra $\mathbb{R}$ and a 3D algebra of the standard form III (d).[9]

Case (b)

Concerning this case ($\tilde{m} = 2$), exploiting a procedure similar to the one used for the case (a), we find an NLE equation given again by (3.15) with $\tilde{m} = 2$, and the set of commutator relations

$$[A, B] = 0, \tag{3.23a}$$

$$[A, C] = -\tfrac{1}{2}\tilde{k}A, \tag{3.23b}$$

$$[B, C] = \tilde{k}B, \tag{3.23c}$$

$$[C, D] = \tilde{g}B, \tag{3.24a}$$

$$[A, D] = [3\tilde{g} - \tfrac{1}{2}\tilde{k}(\tfrac{1}{4}\tilde{k}^2 + \tilde{h})]A, \tag{3.24b}$$

$$[B, D] = \tilde{k}(\tilde{h} + \tilde{k}^2)B. \tag{3.24c}$$

In order to classify this finite dimensional Lie algebra, let us put $\tilde{D} = D + \tilde{g}/\tilde{k}B$. Then the commutator relations (3.24) become

$$[C, \tilde{D}] = 0, \tag{3.25a}$$

$$[A, \tilde{D}] = [3\tilde{g} - \tfrac{1}{2}\tilde{k}(\tfrac{1}{4}\tilde{k}^2 + \tilde{h})]A, \tag{3.25b}$$

$$[B, \tilde{D}] = \tilde{k}(\tilde{h} + \tilde{k}^2)B. \tag{3.25c}$$

In analogy to the case (a), let us introduce the quantity

$$\Delta = \begin{vmatrix} \tfrac{1}{2}\tilde{k} & -\tilde{k} \\ \tfrac{1}{2}\tilde{k}(\tfrac{1}{4}\tilde{k}^2 + \tilde{h}) - 3\tilde{g} & -\tilde{k}(\tilde{k}^2 + \tilde{h}) \end{vmatrix} = -3\tilde{k}(\tilde{g} + \tfrac{1}{8}\tilde{k}^3). \tag{3.26}$$

Let us deal with the cases (i) $\tilde{k}^3 \neq -2\tilde{g}\tilde{m}^2 = -8\tilde{g}$ and (ii) $\tilde{k}^3 = -8\tilde{g}$.

In the case (i), the commutator relations (3.23) and (3.25) take the form (3.17), where C' and D' are defined by

$$C' = (1/\Delta)[-\tilde{k}(\tilde{k}^2 + \tilde{h})C + \tilde{k}\tilde{D}] \tag{3.27}$$

and

$$D' = (1/\Delta)\{-[\tfrac{1}{2}\tilde{k}(\tfrac{1}{4}\tilde{k}^2 + \tilde{h}) - 3\tilde{g}]C + \tfrac{1}{2}\tilde{k}\tilde{D}\}. \tag{3.28}$$

In the case (ii), the commutator relations (3.23) and (3.25) have the form (3.20) where the quantity $(\lambda_1 - 1)/(\lambda_2 - 1)$ takes the value -2, and C'' and D'' are given by

$$C'' = (2/\tilde{k})C \tag{3.29}$$

and

$$D'' = \tilde{D} - (\tilde{h} + \tilde{k}^2)C. \tag{3.30}$$

Case (c)

For $\tilde{m} = -\tfrac{1}{4}$ we obtain the set of commutator relations

$$\begin{aligned} &[A, B] = 0, \quad [C, D] = 0, \\ &[A, C] = -2\tilde{k}(A - 2B), \quad [A, D] = (\tfrac{3}{2}\tilde{g} - 8\tilde{k}^3 - 2\tilde{k}\tilde{h})A + (48\tilde{k}^3 + 4\tilde{k}\tilde{h} + \tilde{g})B, \\ &[B, C] = -2\tilde{k}B, \quad [B, D] = (\tfrac{3}{2}\tilde{g} - 8\tilde{k}^3 - 2\tilde{k}\tilde{h})B. \end{aligned} \tag{3.31}$$

We shall distinguish the cases (i) $\tilde{k}^3 \neq -2\tilde{g}\tilde{m}^2 = -\tilde{g}/8$ and (ii) $\tilde{k}^3 = -\tilde{g}/8$.

In the case (i), the algebra (3.31) reads

$$[A, B] = 0, \quad [D', C'] = 0,$$
$$[A, C'] = B, \quad [D', A] = A, \tag{3.32}$$
$$[B, C'] = 0, \quad [D', B] = B,$$

where

$$C' = [8\tilde{k}(8\tilde{k}^3 + \tilde{g})]^{-1}[2\tilde{k}D - (8\tilde{k}^3 + 2\tilde{k}\tilde{h} - \tfrac{3}{2}\tilde{g})C] \tag{3.33}$$

and

$$D' = [8\tilde{k}(8\tilde{k}^3 + \tilde{g})]^{-1}[-4\tilde{k}D + (48\tilde{k}^3 + 4\tilde{k}\tilde{h} + \tilde{g})C]. \tag{3.34}$$

The algebra (3.32) is a nonabelian 4D algebra with a 2D abelian derived algebra.[9]

In the case (ii), the algebra (3.31) becomes

$$[C'', A] = A - 2B, \quad [D'', A] = 0,$$
$$[C'', B] = B, \quad [D'', B] = 0, \tag{3.35}$$
$$[A, B] = 0, \quad [D'', C''] = 0,$$

where

$$C'' = (1/2\tilde{k})C \tag{3.36}$$

and

$$D'' = (8\tilde{k}^3 + 2\tilde{k}\tilde{h} - \tfrac{3}{2}\tilde{g})^{-1}D - (1/2\tilde{k})C. \tag{3.37}$$

The algebra (3.35) is the direct sum of a 1D algebra $\mathbb{R}$ and a 3D algebra of the form III (d).[9]

(II) $m = \text{const} \neq 0$, $k = 0$

If we assume that m is a constant different from zero and $k = 0$, the basic equations (2.10) take the form

$$F_{uuu} = mF_u, \tag{3.38a}$$
$$[F, F_{uu}] = 0, \tag{3.38b}$$
$$L_u = [F, [F, F_u]] + hF_u, \tag{3.38c}$$
$$[F, L] = gF_u. \tag{3.38d}$$

From (3.38a) one gets

$$F = Ae^{\lambda u} + Be^{-\lambda u} + C, \tag{3.39}$$

where $\lambda^2 = m$ and A, B, and C are vector functions of integration depending on the y's only.

Using the abbreviation

$$D \equiv [B, A], \tag{3.40}$$

one has

$$[F_u, F_{uu}] = -2\lambda^3 D. \tag{3.41}$$

Substituting from (3.39) in (3.38b), one finds

$$[C, A] = [C, B] = 0. \tag{3.42}$$

In order to derive $L(u, y)$, let us introduce (3.39) in (3.38c). We obtain

$$L_u = 2\lambda\{e^{\lambda u}[A, D] + e^{-\lambda u} \times [B, D]\} + \lambda h(e^{\lambda u}A - e^{-\lambda u}B), \tag{3.43}$$

which can be integrated to give

$$L = 2\{e^{\lambda u}[A, D] - e^{-\lambda u} \times [B, D]\} + \lambda(h_1 A - h_2 B) + E, \tag{3.44}$$

where $E \equiv E\{y^i\}$ is a vector function of integration and

$$h_1 = \int^u h(t)e^{\lambda t}\,dt, \quad h_2 = \int^u h(t)e^{-\lambda t}\,dt. \tag{3.45}$$

With the help of (3.44), Eq. (3.38d) yields

$$2\{e^{2\lambda u}[A, [A, D]] - e^{-2\lambda u}[B, [B, D]]\} + \lambda(h_1 e^{-\lambda u} + h_2 e^{\lambda u})D + e^{\lambda u}[A, E] + e^{-\lambda u}[B, E] + [C, E] = \lambda g(e^{\lambda u}A - e^{-\lambda u}B). \tag{3.46}$$

Taking now the commutator of (3.46) with D, we obtain a relation which allows us to determine $g(u)$ in the nontrivial case where $[A, D]$ and $[B, D]$ are not zero.

After some manipulations, we have

$$g = \alpha_0 + \alpha_1 e^{-\lambda u} + \alpha_2 e^{\lambda u}, \tag{3.47}$$

where α_0, α_1, and α_2 are constants.

A similar procedure provides

$$h = \gamma_0 + \gamma_1 e^{2\lambda u} + \gamma_2 e^{-2\lambda u} + \gamma_3 u, \tag{3.48}$$

where γ_0, γ_1, γ_2, and γ_3 are constants.

By virtue of (3.47), (3.48), and (3.45), taking equal to zero the coefficients in front of the functions of the variable u such as $e^{\lambda u}$, $e^{-\lambda u}$, and so on, which are mutually independent, from (3.46) we are led to the following commutator relations:

$$2[A, [A, D]] = -\tfrac{4}{3}\gamma_1 D + \lambda\alpha_2 A, \tag{3.49a}$$
$$2[B, [B, D]] = -\tfrac{4}{3}\gamma_2 D + \lambda\alpha_1 B, \tag{3.49b}$$
$$[A, E] = \lambda\alpha_0 A, \tag{3.49c}$$
$$[B, E] = -\lambda\alpha_0 B, \tag{3.49d}$$
$$[C, E] = (2/\lambda)\gamma_3 D + \lambda\alpha_1 A - \lambda\alpha_2 B. \tag{3.49e}$$

We have shown that if γ_1 and γ_2 are supposed different from zero, then the constants α_0, α_1, α_2, and γ_3 are vanishing (see Appendix B). Thus Eqs. (3.49) read

$$[A, [A, D]] = -\tfrac{2}{3}\gamma_1 D, \tag{3.50a}$$
$$[B, [B, D]] = -\tfrac{2}{3}\gamma_2 D, \tag{3.50b}$$
$$[A, E] = 0, \tag{3.50c}$$
$$[B, E] = 0, \tag{3.50d}$$
$$[C, E] = 0, \tag{3.50e}$$

where D is given by (3.40). These relations, together with the following ones (see Appendix B),

$$[A, C] = [B, C] = [E, D] = [C, D] = 0, \tag{3.50f}$$

constitute the nonabelian prolongation Lie algebra associated with the NLE equation [see (1.1), (2.9), (3.47), and (3.48)]:

$$u_t + u_{xxx} + (\gamma_0 + \gamma_1 e^{2\lambda u} + \gamma_2 e^{-2\lambda u})u_x - \tfrac{1}{2}mu_x^3 = 0. \tag{3.51}$$

Putting $u = W/2\lambda$, Eq. (3.51) becomes

$$W_t + W_{xxx} + (\gamma_0 + \gamma_1 e^{W} + \gamma_2 e^{-W})W_x - \tfrac{1}{8}W_x^3 = 0. \tag{3.52}$$

Equation (3.52) was deduced first by Calogero and Degasperis through the spectral transform method.[10] It possesses an infinite set of conserved quantities, Bäcklund transformations[4] and multisoliton solutions. Furthermore, we point out that there exist a transformation which links (3.52) to the modified Korteweg–de Vries equation

$$u_t + u_{xxx} + (c_1 + c_2 u^2)u_x = 0, \tag{3.53}$$

where c_1 and c_2 are constants. Precisely, one can prove[11,12] that if $W(x, t)$ satisfies the equation

$$W_t + W_{xxx} + [\gamma - \tfrac{3}{8}b^2(c_1 e^{W/2} + c_2 e^{-W/2})^2]W_x - \tfrac{1}{8}W_x^3 = 0, \tag{3.54}$$

where γ, b, c_1, and c_2 are constants, then the function $u(x, t)$ expressed by

$$u = W_x + b(c_1 e^{W/2} + c_2 e^{-W/2}) \tag{3.55}$$

fulfills the modified Korteweg–de Vries equation

$$u_t + u_{xxx} + (\gamma - \tfrac{3}{8}u^2)u_x = 0. \tag{3.56}$$

(III) $m = 0$, $k = \mathrm{const} \neq 0$

When $m = 0$ and k is a constant different from zero, from Eqs. (2.10a) and (2.10b) we obtain

$$F = Au^2 + Bu + C, \tag{3.57}$$

and

$$[B, A] = kA, \tag{3.58}$$

$$[C, A] = \tfrac{1}{2}kB, \tag{3.59}$$

and A, B, and C depend on the y's only.

Furthermore, one can easily see that

$$L_u = u([B, D] + kD) + [C, D] + hF_u, \tag{3.60}$$

where $D \equiv [C, B]$.

Putting

$$\psi_{uu} = h \tag{3.61}$$

and recalling that in our case $F_{uuu} = 0$, Eq. (3.60) yields

$$L = \tfrac{1}{2}u^2([B, D] + kD) + [C, D]u + \psi_u F_u - \psi F_{uu} + E, \tag{3.62}$$

where $E \equiv E\{y^i\}$ is a vector function of integration.

Substituting (3.62) in (2.10d), we have

$$\begin{aligned} &\tfrac{1}{6}u^3([B, [B, D]] - k^2 D) \\ &\quad + \tfrac{1}{2}u^2([B, [C, D]] + k[C, D] + 2[A, E]) \\ &\quad + u([C, [C, D]] + [B, E]) \\ &\quad + [C, E] + \psi_u(ku^2 A + kuB + D) \\ &\quad - k\psi(2Au + B) = g(2Au + B). \end{aligned} \tag{3.63}$$

Then, taking the commutator of A with (3.63), we are led to the relation

$$\frac{\partial^3}{\partial u^3}[-\psi_u(k^2 uA + \tfrac{1}{2}k^2 B) + (k^2\psi + kg)A] = 0, \tag{3.64}$$

which gives

$$\psi = \alpha_0 + \alpha_1 u + \alpha_2 u^2 + \alpha_3 u^3, \tag{3.65}$$

$$g = \beta_0 + \beta_1 u + \beta_2 u^2 + 2k\alpha_3 u^3 \tag{3.66}$$

since A and B are linearly independent [see (3.58)]. The quantities α_i and β_i are constants.

Inserting (3.65) and (3.66) in (3.63) and equating to zero the coefficients of powers of u, we obtain $\alpha_3 = 0$ and the commutator relations:

$$[B, [B, D]] = k^2 D + 4\beta_2 A, \tag{3.67a}$$

$$3[B, [B, D]] + k[C, D] + 2[A, E'] = 4\beta_1 A, \tag{3.67b}$$

$$[C, [C, D]] + [B, E'] = \beta_2 B - (2\beta_2/k)D, \tag{3.67c}$$

$$[C, E'] = 0, \tag{3.67d}$$

where

$$\begin{aligned} E' = E &- (2/k)(\beta_0 + k\alpha_0)A + \alpha_1 B \\ &+ (2/k)(\beta_2 - k\alpha_2)C. \end{aligned} \tag{3.68}$$

We have proved in Appendix C that $\beta_2 = 0$. Furthermore, for a one-dimensional prolongation, where the set of variables $\{y^i\}$ has one element only, one can see that β_1 must be zero in order to avoid inconsistency. In this case the commutator relations (3.67) (with $\beta_2 = \beta_1 = 0$) define a nonabelian (infinite-dimensional[7]) Lie algebra associated with the equation [see (3.61), (3.65), and (2.9)]:

$$u_t + u_{xxx} + \beta_0 + 2\alpha_2 u_x - \tfrac{3}{2}ku_x^2 = 0, \tag{3.69}$$

which can also be written as the Korteweg–de Vries equation

$$W_t + W_{xxx} + (2\alpha_2 - 3kW)W_x = 0, \tag{3.70}$$

through the transformation $u_x = W$. The problem as to whether this result holds also for a pseudopotential of any kind is under investigation.

(IV) $m = k = 0$

In the case where both m and k are zero, from the fundamental equations (2.10) we find

$$F = \tfrac{1}{2}u^2 A + uB + C, \tag{3.71}$$

and

$$\begin{aligned} L = \tfrac{1}{2}u^2[A, [C, B]] &+ u[C, [C, B]] \\ &+ \tilde{h}A + \hat{h}B + D, \end{aligned} \tag{3.72}$$

where A, B, C, and D depend on the y's only, and

$$\tilde{h} = \int^u h(t)t\,dt, \quad \hat{h} = \int^u h(t)\,dt. \tag{3.73}$$

Inserting (3.71) and (3.72) in (2.10d), after some manipulations we obtain the set of relations:

$$\begin{aligned} &\tfrac{3}{2}u^2[B, [B, E]] + u\{3[B, [C, E]] + [A, D]\} \\ &\quad + \{[B, D] + [C, [C, E]]\} + hE \\ &\quad = (g_u u + g)A + g_u B, \end{aligned} \tag{3.74a}$$

$$\begin{aligned} &3u[B, [B, E]] + \{3[B, [C, E]] + [A, D]\} \\ &\quad + h_u E = (2g_u + ug_{uu})A + g_{uu}B, \end{aligned} \tag{3.74b}$$

$$3[B, [B, E]] + h_{uu}E = (3g_{uu} + ug_{uuu})A + g_{uuu}B, \tag{3.74c}$$

$$h_{uuu}E = (4g_{uuu} + ug_{uuuu})A + g_{uuuu}B, \tag{3.74d}$$

where

$$E \equiv [C, B]. \tag{3.75}$$

The following subcases are possible:

(i) $E = 0$, (3.76)

(ii) $E \neq 0$, $\quad h_{uuu} \neq 0$, (3.77)

(iii) $E \not\equiv 0$, $\quad h_{uuu} = 0$. (3.78)

Subcase (*i*): We have proved that Eq. (1.1) becomes

$$u_t + u_{xxx} + g_0 + g_1 u + hu_x = 0, \tag{3.79}$$

where $h \equiv h(u)$ is an arbitrary function such that $h_{uuu} \neq 0$, and g_0 and g_1 are constants of integration. The Lie algebra associated with Eq. (3.79) is closed and is given by

$$[A, B] = [A, C] = [C, B] = 0, \tag{3.80a}$$

$$[A, D] = 2g_1 A, \tag{3.80b}$$

$$[B, D] = g_0 A + g_1 B, \tag{3.80c}$$

$$[C, D] = g_0 B. \tag{3.80d}$$

The commutator relations (3.80) define a 4D Lie algebra with a 2D abelian derived algebra.[9]

Subcase (ii): We have obtained the equation [see (1.1)]:

$$u_t + u_{xxx} + \tilde{g}/u + g_1 u + g_0 + (a/u^2 + h_1 u + h_0)u_x = 0, \tag{3.81}$$

where $\tilde{g}$, g_1, g_0, a, h_1, and h_0 are constants. The Lie algebra associated with (3.81) is closed and is defined by

$$[A, B] = [A, C] = 0, \tag{3.82a}$$

$$[C, B] = -(\tilde{g}/a)B, \tag{3.82b}$$

$$[A, D] = 2g_1 A + h_1(\tilde{g}/a)B, \tag{3.82c}$$

$$[B, D] = g_0 A + (g_1 + \tilde{g}/a + \tilde{g}^3/a^3)B, \tag{3.82d}$$

$$[C, D] = \tilde{g}A + g_0 B. \tag{3.82e}$$

The algebra (3.82) is a 4D Lie algebra with a 2D abelian derived algebra.[9]

Subcase (iii): Equation (1.1) takes the form

$$u_t + u_{xxx} + g_0 + g_1 u + \tfrac{1}{2}g_2 u^2 + (h_0 + h_1 u + \tfrac{1}{2}h_2 u^2)u_x = 0, \tag{3.83}$$

where g_i and h_i are constants.

The Lie algebra associated with Eq. (3.83) is given by the commutator relations

$$[C, B] \equiv E \neq 0, \tag{3.84a}$$

$$[A, B] = [A, C] = [A, E] = 0, \tag{3.84b}$$

$$[C, D] = g_0 B, \tag{3.84c}$$

$$[B, [B, E]] = g_2 A - \tfrac{1}{3}h_2 E, \tag{3.84d}$$

$$[A, D] + 3[B, [C, E]] = 2g_1 A + g_2 B - h_1 E, \tag{3.84e}$$

$$[B, D] + [C, [C, E]] = g_0 A + g_1 B - h_0 E. \tag{3.84f}$$

Choosing $g_0 = g_1 = g_2 = 0$, the algebra (3.84) reduces to the prolongation algebra associated with the modified Korteweg–de Vries equation [see (3.83)], which is known to be a completely integrable NLE equation.[1]

At present we can say nothing about the possibility that some constants g_i may be chosen different from zero. In fact, it could happen that the symmetry properties of the commutator relations (3.84) *imply* that all the constants g_i must be zero [in analogy to what occurs, for example, for the algebra (3.49)].

(V) $m_u \neq 0$, $k = 0$

This case can be treated in strict analogy to the case (I) ($m_u \neq 0$, $k \neq 0$). One can see that

$$[F, F_u] = 0, \tag{3.85}$$

$$[F_u, F_{uu}] = 0, \tag{3.86}$$

$$L_u = hF_u. \tag{3.87}$$

Since F_u and F_{uu} are supposed to be linearly independent, Eq. (3.86) entails that the pseudopotential defined by the prolongation equations (2.1) cannot be of the first kind.

After some manipulations, employing (3.85), (3.86), and (3.87), Eq. (2.10d) yields

$$m = \tilde{m}e^{-2W}, \tag{3.88}$$

$$g = \tilde{g}e^{W}, \tag{3.89}$$

where $W = \ln(u - u_0)$ and $\tilde{m}$, $\tilde{g}$, u_0 are constants.

We shall distinguish the following subcases: (i) $\tilde{m} \neq 2$, $-\frac{1}{4}$; (ii) $\tilde{m} = 2$, and (iii) $\tilde{m} = -\frac{1}{4}$.

Subcase (i)

One obtains the following solution of Eq. (2.10a):

$$F = e^{\lambda_1 W}A + e^{\lambda_2 W}B + C, \tag{3.90}$$

where A, B, and C depend on the y's only, and λ_1 and λ_2 ($\lambda_1 \neq \lambda_2$; $\lambda_1, \lambda_2 \neq 0$) are roots of the characteristic equation associated with (3.7).

Equation (1.1) takes the form

$$u_t + u_{xxx} + \tilde{g}(u - u_0) - [\tilde{m}/2(u - u_0)^2]u_x^3 + hu_x = 0, \tag{3.91}$$

where h is an arbitrary function of u.

The Lie algebra associated with this equation is finite-dimensional and is given by the commutator relations:

$$[A, B] = [C, A] = [C, B] = 0, \tag{3.92a}$$

$$[A, D] = \tilde{g}\lambda_1 A, \tag{3.92b}$$

$$[B, D] = \tilde{g}\lambda_2 B, \tag{3.92c}$$

$$[C, D] = 0, \tag{3.92d}$$

where D is a (vector) function of integration depending on the variables y's only.

The algebra (3.92) is the direct sum of a 1D algebra $\mathbb{R}$ and a 3D algebra of the standard form III(d) according to Jacobson's classification.[9]

Subcase (ii)

One has

$$F = e^{3W}A + WB + C. \tag{3.93}$$

The Lie algebra associated with Eq. (3.90) (for $\tilde{m} = 2$) is finite-dimensional and reads

$$[A, B] = [C, A] = [C, B] = 0, \tag{3.94a}$$

$$[A, D] = 3\tilde{g}A, \tag{3.94b}$$

$$[B, D] = 0, \tag{3.94c}$$

$$[C, D] = \tilde{g}B. \tag{3.94d}$$

The algebra (3.94) is a 4D algebra with a 2D abelian derived algebra.[9]

Subcase (iii)

One finds

$$F = e^{3W/2}A + We^{3W/2}B + C. \tag{3.95}$$

The Lie algebra associated with Eq. (3.90) (for $\tilde{m} = -\frac{1}{4}$) is finite-dimensional and turns out to be

$$[A, B] = [C, A] = [C, B] = 0, \tag{3.96a}$$

$$[A, D] = \tilde{g}(\tfrac{3}{2}A + B), \qquad (3.96b)$$

$$[B, D] = \tfrac{3}{2}B, \qquad (3.96c)$$

$$[C, D] = 0. \qquad (3.96d)$$

The algebra (3.96) is the direct sum of a 1D algebra and a 3D algebra of the standard form III(d).[9]

IV. CASE WHERE F_u AND F_{uu} ARE LINEARLY DEPENDENT

If F_u and F_{uu} are linearly dependent, then $[F_u, F_{uu}] = 0$. We can put

$$F_{uu} = l(u)F_u. \qquad (4.1)$$

First let us consider the case $l \neq 0$.

Taking account of (4.1), Eq. (2.10a) yields

$$m = l^2 + l_u. \qquad (4.2)$$

From (2.10b) and (4.1) we get

$$l[F, F_u] = kF_u. \qquad (4.3)$$

Then, with the help of (4.3), Eq. (2.10b) gives

$$k = \alpha l, \qquad (4.4)$$

where α is a constant of integration.

Now let us introduce a function $f(u)$ such that

$$f_{uu}/f_u = l. \qquad (4.5)$$

Inserting (4.5) in (4.1) and integrating, we obtain

$$F = Af + B, \qquad (4.6)$$

where A and B are vector functions of the y's only.

Substitution from (4.6) in (4.3) yields

$$[B, A] = \alpha A. \qquad (4.7)$$

Now, subtracting the first from the second derivative with respect to u of Eq. (2.10d), we are led to the relation

$$2[F_u, L_u] + [F, L_{uu}] - l[F, L_u]$$
$$= g_{uu}F_u + 2g_uF_{uu} + gF_{uuu} - l(g_uF_u + gF_{uu}). \qquad (4.8)$$

The commutators on the left-hand side can be readily calculated. In fact, using (4.3), we have

$$[F, [F, F_u]] = \alpha^2 F_u. \qquad (4.9)$$

Inserting (4.9) in (2.10c), we find

$$L_u = (\alpha^2 + h)F_u, \qquad (4.10)$$

so that

$$[F_u, L_u] = 0. \qquad (4.11)$$

On the other hand, exploiting (4.10), we can write

$$[F, L_{uu}] = (\alpha^2 + h)\alpha l F_u + \alpha h_u F_u \qquad (4.12)$$

and

$$[F, L_u] = (\alpha^2 + h)\alpha F_u. \qquad (4.13)$$

Putting now (4.11), (4.12), and (4.13) in (4.8), in virtue of (4.2) we obtain the condition

$$\alpha h_u = g_{uu} + \frac{d}{du}(gl), \qquad (4.14)$$

which gives (for $g \neq 0$)

$$\alpha h = g_u + gl + \beta, \qquad (4.15)$$

where β is a constant of integration.

The function L can be explicitly calculated from (4.10) with the help of (4.5), (4.6), and (4.15). We have

$$L = (\alpha^2 + \beta/\alpha)Af + (A/\alpha)(gf_u) + C, \qquad (4.16)$$

where C depends on the y's only.

Substitution from (4.16) in (2.10d) yields the commutator relations

$$[B, C] = 0 \qquad (4.17)$$

and

$$[A, C] = -(\alpha^3 + \beta)A, \qquad (4.18)$$

which define, together with the relation (4.7), a nonabelian *closed* (finite-dimensional) Lie algebra, which can be put in the standard form III(c) according to Jacobson's classification.[9]

From (4.2), (4.4), and (4.15) we see that Eq. (1.1) can be written as

$$u_t + u_{xxx} + g + (1/\alpha)(g_u + gl + \beta)u_x - \tfrac{3}{2}\alpha l u_x^2$$
$$- \tfrac{1}{2}(l^2 + l_u)u_x^3 = 0, \qquad (4.19)$$

where g and l are arbitrary functions of the variable u.

Special cases

A. Case $k = 0$, $l \neq 0$, $(\alpha = 0)$

We have found that (4.6) holds again. Equation (4.7) becomes

$$[B, A] = 0. \qquad (4.20)$$

From (4.6), (4.20), and (2.10b) we have

$$[F, F_u] = 0 \qquad (4.21)$$

and

$$[F, F_{uu}] = 0. \qquad (4.22)$$

Furthermore, from (4.10) we get

$$L_u = hF_u = hAf_u, \qquad (4.23)$$

which yields

$$L = A\psi(u) + C, \qquad (4.24)$$

where

$$\psi_u = hf_u \qquad (4.25)$$

and C is a (vector) function of integration.

Inserting (4.24) in (2.10d), we get

$$f[A, C] + [B, C] = gAf_u. \qquad (4.26)$$

Now from (4.15) (for $\alpha = 0$) we obtain

$$gf_u = -\beta f + \gamma, \qquad (4.27)$$

where γ is an arbitrary constant.

Substitution from (4.27) in (4.26) gives

$$[A, C] = -\beta A \qquad (4.28)$$

and

$$[B, C] = \gamma A. \qquad (4.29)$$

The commutator relations (4.28), (4.29), and (4.20), define a *closed* nonabelian Lie algebra. When $\beta \neq 0$, this algebra can be put in Jacobson's standard form III(c), while when $\beta = 0$ this algebra becomes of the type III(b).[9] This algebra can be associated with the nonlinear evolution equation

$$u_t + u_{xxx} + g + hu_x - \tfrac{1}{2}mu_x^3 = 0, \tag{4.30}$$

where m is given by (4.2), g is such that

$$g_u + gl = -\beta, \tag{4.31}$$

and h is an arbitrary function of the variable u.

B. Case $l = 0$

In this case, from (4.2) and (4.4) we have $m = 0$ and $k = 0$. Furthermore, Eq. (4.1) yields

$$F = Au + B, \tag{4.32}$$

where the (vector) functions of integration A and B depend on the y's only. We have [see (4.32)]

$$[F, F_u] = [B, A], \tag{4.33}$$

and

$$[F, [F, F_u]] = u[A, [B, A]] + [B, [B, A]]. \tag{4.34}$$

Inserting (4.34) in (2.10c), we obtain

$$L_u = u[A, [B, A]] + [B, [B, A]] + hA. \tag{4.35}$$

Differentiating (2.10d) with respect to u, we find

$$2[F_u, L_u] + [F, L_{uu}] = g_{uu} F_u. \tag{4.36}$$

Then, substitution from (4.35) in (4.36) yields

$$3u[A, [A, [B, A]]] + 3[B, [A, [B, A]]] + h_u [B, A] = g_{uu} A, \tag{4.37}$$

which gives

$$h_{uuu} [B, A] = g_{uuuu} A. \tag{4.38}$$

From (4.38) we get

$$h_{uuu} [A, [B, A]] = 0. \tag{4.39}$$

We have to distinguish two cases, namely

$$\text{(i)} \quad h_{uuu} = 0, \quad [A, [B, A]] \neq 0 \tag{4.40}$$

and

$$\text{(ii)} \quad h_{uuu} \neq 0, \quad [A, [B, A]] = 0. \tag{4.41}$$

First we consider the case (i). From (4.40) and (4.38) we obtain

$$h = 3\alpha_1 u^2 + 2\alpha_2 u + \alpha_3 \tag{4.42}$$

and

$$g = \beta_1 u^3 + \beta_2 u^2 + \beta_3 u + \beta_4, \tag{4.43}$$

where α_i and β_i are constants.

Inserting (4.42) and (4.43) in (4.35), we find

$$L = \tfrac{1}{2}u^2[A, [B, A]] + u[B, [B, A]] + (\alpha_1 u^3 + \alpha_2 u^2 + \alpha_3 u)A + C, \tag{4.44}$$

where C is a vector function which depends on the y's only.

With the help of (4.44), Eq. (2.10d) leads to the following set of commutator relations:

$$[A, [A, [B, A]]] + 2\alpha_1 [B, A] = 2\beta_1 A, \tag{4.45a}$$

$$3[B, [A, [B, A]]] + 2\alpha_2 [B, A] = 2\beta_2 A, \tag{4.45b}$$

$$[A, C] + [B, [B, [B, A]]] + \alpha_3 [B, A] = \beta_3 A, \tag{4.45c}$$

$$[B, C] = \beta_4 A. \tag{4.45d}$$

Choosing all the constants β_i equal to zero, the algebra (4.45) becomes the prolongation algebra associated with the modified Korteweg–de Vries equation:

$$u_t + u_{xxx} + (3\alpha_1 u^2 + 2\alpha_2 u + \alpha_3) u_x = 0. \tag{4.46}$$

We notice that the algebra (4.45) can be formally obtained from the algebra (3.84) by putting $A \equiv 0$.

Furthermore, we recall that the algebra (3.84) has been obtained assuming that F_u and F_{uu} are linearly independent, while the algebra (4.45) has been derived in the case where $F_{uu} = 0$. In the first case, unlikely to what happens for the second case, the pseudopotential defined by Eqs. (2.1) cannot be of the first kind.

Concerning the case (ii), from (4.37) we get

$$h_u [B, A] = g_{uu} A, \tag{4.47}$$

from which one has

$$g_u = \gamma h + \delta, \tag{4.48}$$

where γ and δ are constants.

Putting now

$$\tilde{h}_u = h, \tag{4.49}$$

Eqs. (4.35) and (4.41) yield

$$L = u[B, [B, A]] + A\tilde{h} + C, \tag{4.50}$$

where C is a vector function of integration of the variables $\{ y^i \}$ and $\tilde{h}$ is given by

$$\tilde{h} = (1/\gamma)(g - \delta u) + \eta, \tag{4.51}$$

η being a constant and $\gamma \neq 0$.

With the help of (4.47) and (4.48) we find

$$[B, A] = \gamma A. \tag{4.52}$$

Then, substitution from (4.52) in (4.50) yields

$$L = [(\gamma^2 - \delta/\gamma)u + (g/\gamma + \eta)] A + C. \tag{4.53}$$

Inserting (4.53) in (2.10d), we get the commutator relations

$$[C, A] = (\gamma^3 - \delta)A, \tag{4.54}$$

$$[C, B] = \gamma\eta A. \tag{4.55}$$

We point out that the algebra defined by (4.52), (4.54), and (4.55) is a *closed* nonabelian Lie algebra which can be put in the standard form III(c) according to Jacobson's classification.

The NLE equation with which this prolongation algebra is associated is given by

$$u_t + u_{xxx} + g(u) + h(u) u_x = 0, \tag{4.56}$$

where the functions g and h have to be chosen in such a way that the constraint (4.48) is satisfied.

V. CONCLUDING REMARKS

Concerning the systematic analysis of the class of NLE equations (1.1), carried out within the Estabrook–Wahlquist prolongation scheme, we have found that all the equations which admit nonabelian prolongations, are such that the functions $\phi(u, u_x)$ are polynomials in the variable u_x.

The equations whose associated nonabelian algebras are infinite-dimensional are likely the only equations which are completely integrable. In this connection, we notice that the NLE equations whose nonabelian Lie algebras are finite-dimensional are not solvable by the inverse scattering method, since the structure constants of their prolongation algebras do not contain free parameters. We have seen also that for any equation of this kind, the search of nontrivial Bäcklund transformations of the type $\psi = \psi(u, u_x, u_{xx}, y)$, where ψ and u are required to satisfy the original equation, fails. However, this result does not imply, *a priori*, that one might not find for these equations nontrivial Bäcklund transformations using mathematical tools more suitable than those employed in this work.[4,5,12,13]

We point out that one can have a similar situation for NLE equations which may be solvable through the inverse scattering method, although they have nontrivial Bäcklund transformations within the scheme adopted in this paper. A notable example is offered by the Harry–Dym equation,[14] $u_t + u^3 u_{xxx} = 0$, which is completely integrable.[15] It is worth to mention the fact that the Harry–Dym equation admits a nonabelian (presumably infinite-dimensional) prolongation algebra whose a quotient algebra is given by the algebra of SL(2, $\mathbb{R}$).[16]

ACKNOWLEDGMENT

The authors are very grateful to professor F. A. E. Pirani for his interest in this work and for many helpful conversations.

APPENDIX A

In this appendix we shall derive the relations (3.1), (3.2), and (3.3), assuming that $m_u \neq 0$ and $k \neq 0$. To this end, let us differentiate Eq. (2.10d) twice with respect to u. We obtain

$$[F_{uu}, L] + 2[F_u, L_u] + [F, L_{uu}] = g_{uu}F_u + 2g_u F_{uu} + mgF_u. \tag{A1}$$

In order to calculate the commutators on the left-hand side, let us start from

$$[F_u, L] = -(1/k)[L, [F, F_{uu}]], \tag{A2}$$

where Eq. (2.10b) has been used.

Exploiting now the Jacobi identity, we get

$$[L, [F, F_{uu}]] = -[F, [F_{uu}\, L]] - [F_{uu}\, [L, F]]. \tag{A3}$$

Then, substitution from (A3) in (A2) and the use of (A1) yield

$$\begin{aligned}[F_u, L] &= (1/k)g_{uu}[F, F_u] + 2g_u F_u + (mg/k)[F, F_u] \\ &\quad - (2/k)[F, [F_u, L_u]] - (1/k)[F, [F, L_{uu}]] \\ &\quad + (g/k)[F_u, F_{uu}].\end{aligned} \tag{A4}$$

Differentiating (2.10b) twice with respect to u, with the help of (2.10a) one obtains

$$[F, F_u] = (k_{uu}/m_u)F_u + (2k_u/m_u)F_{uu}, \tag{A5}$$

from which one has

$$[F, [F, F_u]] = (k_{uu}/m_u)[F, F_u] + (2k_u k/m_u)F_u. \tag{A6}$$

Taking now the commutator of F_u with L_u [see (2.10c)] and using (A6) and (A5), one gets

$$\begin{aligned}[F_u, L_u] &= [F_u, [F, [F, F_u]]] \\ &= (k_{uu}/m_u)[F_u, [F, F_u]] \\ &= 2(k_{uu}k_u/m_u^2)[F_u, F_{uu}].\end{aligned} \tag{A7}$$

Now let us deal with $[F, L_{uu}]$. From (2.10c) we have

$$\begin{aligned}[F, L_{uu}] &= (2k_u/m_u)[F, [F_u, F_{uu}]] + k[F, [F, F_u]] \\ &\quad + h_u[F, F_u] + hkF_u.\end{aligned} \tag{A8}$$

With the help of (2.10c), the quantity (A8) becomes

$$[F, L_{uu}] = (2k_u/m_u)[F, [F_u, F_{uu}]] + kL_u + h_u[F, F_u]. \tag{A9}$$

Resorting then to the Jacobi identity, we can write

$$\begin{aligned}[F, [F_u, F_{uu}]] &= -[F_{uu}, [F, F_u]] \\ &= (k_{uu}/m_u)[F_u, F_{uu}],\end{aligned} \tag{A10}$$

where (A5) has been employed.

Insertion of (A10) in (A9) yields

$$[F, L_{uu}] = (2k_u k_{uu}/m_u^2)[F_u, F_{uu}] + kL_u + h_u[F, F_u], \tag{A11}$$

from which one has [see (A5) and (A10)]

$$\begin{aligned}&[F, [F, L_{uu}]] \\ &= (2k_u k_{uu}/m_u^2)[F, [F_u, F_{uu}]] + k[F, L_u] \\ &\quad + h_u[F, [F, F_u]] \\ &= (2k_u k_{uu}^2/m_u^3)[F_u, F_{uu}] + k[F, L_u] \\ &\quad + (k_{uu}/m_u)h_u[F, F_u] + (2kk_u/m_u)h_u F_u.\end{aligned} \tag{A12}$$

Introducing this expression into (A4), we obtain [see (A7) and (A10)]

$$\begin{aligned}[F_u, L] &= \left(\frac{g_{uu}}{k} + \frac{mg}{k} - \frac{k_{uu}}{km_u}h_u\right)[F, F_u] \\ &\quad + \left(\frac{g}{k} - \frac{6}{k}\frac{k_u k_{uu}^2}{m_u^3}\right)[F_u, F_{uu}] \\ &\quad + \left(2g_u - \frac{2k_u}{m_u}h_u\right)F_u - [F, L_u].\end{aligned} \tag{A13}$$

Comparing then (A13) with the relation

$$[F_u, L] + [F, L_u] = g_u F_u + gF_{uu}, \tag{A14}$$

which is readily obtained from (2.10d), we are led to the expression

$$\begin{aligned}&\left(\frac{g_{uu}}{k} + \frac{mg}{k} - \frac{k_{uu}h_u}{km_u}\right)[F, F_u] \\ &\quad + \left(\frac{g}{k} - \frac{6}{k}\frac{k_u k_{uu}^2}{m_u^3}\right)[F_u, F_{uu}] \\ &\quad + \left(g_u - \frac{2k_u}{m_u}h_u\right)F_u = gF_{uu}.\end{aligned} \tag{A15}$$

Now from (2.10a) and (2.10b) one has

$$[F_u, F_{uu}] = \left(k_u - \frac{mk_{uu}}{m_u}\right)F_u + \left(k - \frac{2mk_u}{m_u}\right)F_{uu}. \tag{A16}$$

Then, from (A10) with the help of (A5) and (2.10b), we get

$$k\left(k - \frac{2mk_u}{m_u}\right)F_u + \frac{1}{m_u}(2k_u^2 - kk_{uu})F_{uu} = 0. \tag{A17}$$

Substitution from (A5) and (A16) in (A15) yields

$$\left[\frac{k_{uu}}{km_u}\left(g_{uu}+mg-\frac{k_{uu}h_u}{m_u}\right)+\frac{1}{k}\left(g-6\frac{k_uk_{uu}^2}{m_u^3}\right)\right.$$
$$\times\left(k_u-\frac{mk_{uu}}{m_u}\right)+g_u-\left.\frac{2k_u}{m_u}h_u\right]F_u$$
$$+\left[\frac{2k_u}{km_u}\left(g_{uu}+mg-\frac{k_{uu}h_u}{m_u}\right)\right.$$
$$+\frac{1}{k}\left(g-6\frac{k_uk_{uu}^2}{m_u^3}\right)$$
$$\times\left(k-\frac{2mk_u}{m_u}\right)-\left.g\right]F_{uu}=0. \qquad \text{(A18)}$$

If one supposes that the vector functions F_u and F_{uu} are linearly independent, Eqs. (A17) and (A18) provide the constraints

$$m_uk=2mk_u, \qquad \text{(A19)}$$

$$2k_u^2=kk_{uu}, \qquad \text{(A20)}$$

and (after slight manipulation)

$$g_u+\frac{k_u}{k}g-\frac{2k_u}{m_u}h_u=0, \qquad \text{(A21)}$$

$$g_{uu}-\frac{k_{uu}}{m_u}h_u=0. \qquad \text{(A22)}$$

We notice that (A16) becomes

$$[F_u,F_{uu}]=0. \qquad \text{(A23)}$$

APPENDIX B

Here we shall prove that if

$$\text{(i) } [A,D]\neq 0,\quad [B,D]\neq 0 \qquad \text{(B1)}$$

and

$$\text{(ii) } \gamma_1\neq 0,\quad \gamma_2\neq 0, \qquad \text{(B2)}$$

then the constants α_0, α_1, α_2, and γ_3 are zero.

In doing so, first we notice that from the Jacobi identity

$$[A,[B,E]]=-[B,[E,A]]-[E,[A,B]], \qquad \text{(B3)}$$

with the help of (3.49d) and (3.49c) one finds

$$[E,D]=0. \qquad \text{(B4)}$$

Now, taking account of (B4) and (3.49c) one has

$$[E,[A,D]]=[D,[A,E]]=-\lambda\alpha_0[A,D]. \qquad \text{(B5)}$$

From (3.49a), employing (B4) and (3.49c), we obtain

$$2[E[A,[A,D]]]=-\lambda^2\alpha_2\alpha_0A. \qquad \text{(B6)}$$

The use of the Jacobi identity and (B5) yields

$$[E,[A,[A,D]]]=-2\lambda\alpha_0[A,[A,D]]. \qquad \text{(B7)}$$

Furthermore, we can write

$$[E,[B,D]]=\lambda\alpha_0[B,D] \qquad \text{(B8)}$$

and

$$[E,[B,[B,D]]]=2\lambda\alpha_0[B,[B,D]]. \qquad \text{(B9)}$$

Now substituting (B7) in (B6), we obtain

$$4\lambda\alpha_0[A,[A,D]]=\lambda^2\alpha_2\alpha_0A. \qquad \text{(B10)}$$

Finally, comparing (B10) with (3.49a) we get

$$\tfrac{8}{3}\alpha_0\gamma_1D=\lambda\alpha_2\alpha_0A, \qquad \text{(B11)}$$

where γ_1 and A are supposed to be nonvanishing.

From (B11) we deduce that α_0 must be zero in order that the assumption (i) is not contradicted.

Then, taking the commutator of (3.49e) with A and B, respectively, we are led to the relations

$$2\gamma_3[A,D]=-\lambda^2\alpha_2D \qquad \text{(B12)}$$

and

$$2\gamma_3[B,D]=-\lambda^2\alpha_1D. \qquad \text{(B13)}$$

From (B12) we obtain

$$2\gamma_3[A,[A,D]]=-\lambda^2\alpha_2[A,D]. \qquad \text{(B14)}$$

Then, substitution from (B14) in (3.49a) yields

$$-\lambda^2\alpha_2[A,D]+\tfrac{4}{3}\gamma_1\gamma_3D=\lambda\gamma_3\alpha_2A. \qquad \text{(B15)}$$

Taking account of (B12), Eq. (B15) gives rise to the relation

$$(\lambda^4\alpha_2^2+\tfrac{8}{3}\gamma_1\gamma_3^2)D=2\lambda\gamma_3^2\alpha_2A, \qquad \text{(B16)}$$

from which we deduce that $\alpha_2=0$ and $\gamma_3=0$, since $\gamma_1\neq 0$ and $[A,D]\neq 0$ by hypothesis.

In a similar manner, from Eqs. (3.49b) and (B13) one can see that $\alpha_1=0$ and $\gamma_3=0$.

APPENDIX C

Here we shall prove that $\beta_2=0$. In doing so, let us take the commutator of Eq. (3.67b) with C. We obtain

$$2[C,[A,E']]$$
$$+k[C,[C,D]]+3[C,[B,[C,D]]]=2\beta_1kB. \qquad \text{(C1)}$$

Furthermore, from the Jacobi identity, (3.67c), (3.59), and (3.67d), we have

$$[C,[A,E']]=\tfrac{1}{2}k[B,E']$$
$$=\tfrac{1}{2}k[C,[C,D]]+\tfrac{1}{2}k\beta_1B-\beta_2D. \qquad \text{(C2)}$$

Substitution from (C2) in (C1) yields

$$3[C,[B,[C,D]]]=\beta_1kB+2\beta_2D. \qquad \text{(C3)}$$

Taking the commutator of (C3) with A and using (3.58), we have

$$[A,[C,[B,[C,D]]]]=-\tfrac{1}{3}\beta_1k^2A-\tfrac{1}{3}\beta_2k^2B, \qquad \text{(C4)}$$

where the relation

$$[A,D]=-\tfrac{1}{2}k^2B \qquad \text{(C5)}$$

has been used.

Exploiting the Jacobi identity, Eq. (C4) becomes

$$[C,[A,[B,[C,D]]]]$$
$$-\tfrac{1}{2}k[B,[B,[C,D]]]=-\tfrac{1}{3}\beta_1k^2A-\tfrac{1}{3}\beta_2k^2B. \qquad \text{(C6)}$$

Furthermore, we get

$$[A,[B,[C,D]]]=[B,[A,[C,D]]]-k[A,[C,D]]$$
$$=-\tfrac{1}{2}k[B,[B,D]]+\tfrac{1}{2}k^3D. \qquad \text{(C7)}$$

Inserting (C7) in (C6), we obtain

$$-\tfrac{1}{2}k[C,[B,[B,D]]]+\tfrac{1}{2}k^3[C,D]-\tfrac{1}{2}k[B,[B,[C,D]]]$$
$$=-\tfrac{1}{3}\beta_1k^2A-\tfrac{1}{3}\beta_2k^2B. \qquad \text{(C8)}$$

Taking now the commutator of (3.67a) with C, we have

$$[C,[B,[B,D]]]=k^2[C,D]+2\beta_2kB. \tag{C9}$$

Substituting then (C9) in (C8), we find

$$[B,[B,[C,D]]]=\tfrac{2}{3}\beta_1kA-\tfrac{4}{3}\beta_2kB. \tag{C10}$$

Now from (3.67b) we have

$$[B,[A,E']]+\tfrac{3}{2}[B,[B,[C,D]]] + \tfrac{1}{2}k[B,[C,D]]=2\beta_1kA. \tag{C11}$$

Since [see (3.67c)]

$$[A,[B,E']]+[A,[C,[C,D]]]=-\beta_1kA+\beta_2kB \tag{C12}$$

and

$$[B,[A,E']]-[A,[B,E']]=k[A,E'], \tag{C13}$$

subtracting (C12) from (C11), in virtue of (C13) we get

$$k[A,E']+\tfrac{3}{2}[B,[B,[C,D]]] + \tfrac{1}{2}k[B,[C,D]]-[A,[C,[C,D]]] = 3\beta_1kA-\beta_2kB. \tag{C14}$$

Now, using (3.67b) and the relation

$$[A,[C,[C,D]]]=-\tfrac{1}{2}k^2[C,D]-k[B,[C,D]], \tag{C15}$$

which arises from the Jacobi identity and from

$$[A,[C,D]]=-\tfrac{1}{2}k[B,D]-\tfrac{1}{2}k^2D, \tag{C16}$$

Eq. (C14) leads to the relation

$$[B,[B,[C,D]]]=\tfrac{2}{3}\beta_1kA-\tfrac{2}{3}\beta_2kB. \tag{C17}$$

Finally, comparing (C17) with (C10), we obtain $\beta_2=0$.

[1]H. D. Wahlquist and F. B. Estabrook, J. Math. Phys. **16**, 1 (1975).
[2]H. D. Wahlquist and F. B. Estabrook, J. Math. Phys. **17**, 1293 (1975).
[3]F. A. E. Pirani, D. C. Robinson, and W. F. Shadwick, *Local Jet Bundle Formulation of Bäcklund Transformations* (D. Reidel, Dordrecht, 1979).
[4]M. Leo, R. A. Leo, L. Martina, F. A. E. Pirani, and G. Soliani, Physica D **4**, 105 (1981); M. Leo, R. A. Leo, G. Soliani, and L. Martina, "Prolongation analysis of the cylindrical Korteweg–de Vries equation," Phys. Rev. D **26**, 809 (1982).
[5]J. Harnad and P. Winternitz, Preprint CRMA-952, Montréal, Québec, Canada, 1980.
[6]R. Dodd and A. Fordy, Preprint DIAS-STP-82-08, Dublin, Ireland, 1982.
[7]W. F. Shadwick, J. Math. Phys. **21**, 454 (1980).
[8]J. Corones and F. J. Testa, in *Lecture Notes in Mathematics* **515**, edited by R. M. Miura (Springer-Verlag, Berlin, 1976).
[9]N. Jacobson, *Lie Algebras* (Wiley, New York, 1962), Chap. I, Par. 4.
[10]F. Calogero and A. Degasperis, J. Math. Phys. **22**, 23 (1981).
[11]A. Degasperis, *Lecture Notes in Physics* **120** (Springer-Verlag, Berlin, 1980), p. 16; F. Magri, *Lecture Notes in Physics* **120** (Springer-Verlag, Berlin, 1980), p. 233.
[12]K. M. Tamizhman and M. Lakshmanan, Phys. Lett. A **90**, 159 (1982).
[13]G. L. Lamb, Jr., J. Math. Phys. **15**, 2157 (1974).
[14]M. D. Kruskal, *Lecture Notes in Physics* **38** (Springer-Verlag, Berlin, 1975), p. 310; M. Wadati, K. Konno, and Y. H. Ichikawa, J. Phys. Soc. Jpn. **47**, 1698 (1979).
[15]M. Wadati, Y. H. Ichikawa, and T. Shimizu, Prog. Theor. Phys. **64**, 1959 (1980).
[16]G. Tondo, Bacc. thesis, Lecce, 1982.

Characterization of canonical Bose–Fermi systems by "anti-Hermitian" symplectic forms

Hideo Nakajima[a)]
Department of Physics, Dalhousie University, Halifax, Nova Scotia, B3H 3J5, Canada

(Received 29 October 1981; accepted for publication 30 June 1982)

A complexification of graded manifold theory is given, following Kostant's procedure, but with a reality concept defined by "classical" correspondent to Hermitian conjugation in quantum mechanics. Presented herein are definitions of graded manifolds with "Hermite" coordinates and of "Hermiticity" on graded differential forms and graded vector fields, all in the coordinate independent way, and characterization of "classical" Bose–Fermi systems by graded symplectic forms ω which are, here, "anti-Hermitian" nonsingular closed 2-forms of z_2 grading 0. Also given are Frobenius' theorem on the graded manifold with "Hermiticity," and Darboux's theorem, $\omega = \Sigma_k dp_k \wedge dq_k + i\,\Sigma_j(\epsilon_j/2)ds_j \wedge ds_j$, where all coordinates are "Hermite," $p_k{}^\dagger = p_k$, $q_k{}^\dagger = q_k, s_j{}^\dagger = s_j$. Naive quantization procedures fit in with these systems.

PACS numbers: 02.40. + m, 03.20. + i, 03.65. − w, 11.30.Pb

1. INTRODUCTION

Utility of "anticommuting c-numbers" is now widely appreciated. In an attempt to construct a unified theory of elementary particles, using the supergravity theory, they are used as parameters of superspace with Bose and Fermi dynamical operators. A characteristic aspect in the usage is that they are treated, not separately from (commuting) c-numbers, but in a combined manner so that Bose and Fermi dynamical variables are transformed and mixed with each other. The superspace is viewed as such. Another utility of "anticommuting c-numbers" is, as is known, that they may be considered as "classical" variables, which are to be quantized in a conventional manner, to describe Fermi or spin degrees of freedom.[1-4] Also in quantization, it is preferable to treat Bose (c-number) and Fermi ("anticommuting c-numbers") coordinates together, not separately, from a viewpoint of mixed coordinate transformations. As such a manifold as the superspace, Kostant defined an abstract manifold called a graded manifold where a local coordinate system consists of even and odd Z_2 graded algebra elements, and the even coordinates themselves are not real numbers, but more or less abstract ones, although there exists an algebra isomorphism between the even coordinates and real functions.[5] Kostant investigated in detail the graded versions of tangent space and of differential forms, and further characterized "classical" canonical Bose–Fermi systems as graded symplectic manifolds which are graded manifolds with graded symplectic forms. In his discussion on graded symplectic manifolds, all elements are considered as real ones forming real algebra, which implies that complex conjugation is understood, to define reality concept, in the background. So far as "classical" Bose–Fermi system, graded symplectic manifold, is concerned, the above reality concept allows simple structure for the system. Here in this paper, however, we take an alternative choice for the reality concept, defined by "Hermitian" conjugation, which will appear below, for the following reason: Suppose quantized Bose–Fermi systems are obtained somehow and we will see the relation between the quantized systems and the "classical" counterparts by putting the Planck constant equal to 0, which is the conventional way in physics. Then all operators in the quantized systems come to form graded commutative algebras with "Hermiticity." We consider these systems (or algebraically isomorphic systems) directly as classical systems.[6] Here we note this reality concept is defined not by complex conjugation, but by "Hermitian" conjugation, where no distinction between them exists in the case when Bose (Z_2 even) system is considered, but in the case when Fermi (Z_2 odd) elements are involved, the distinction exists. The main purpose of this paper is to realize a kind of complexification of Kostant's theory such that we can define graded manifolds with "Hermite" coordinates, and further the concept of Hermiticity also on graded differential forms and graded vector fields, all in the coordinate independent way, and characterize "classical" canonical Bose–Fermi systems by graded symplectic forms which are now "anti-Hermitian" nonsingular closed 2-forms with Z_2 grading 0. Having obtained graded symplectic forms, we can proceed along the same line of Kostant, and define graded Poisson brackets. At this point, an important theorem, Darboux's theorem, is to be demonstrated; that is, if ω is an "anti-Hermitian" nonsingular closed 2-form with Z_2 grading 0, then there exists a "Hermitian" canonical coordinate system $(q_1, p_1,\dots, q_m, p_m, s_1,\dots,s_n)$ such that $\omega = \Sigma_{k=1}^m dp_k \wedge dq_k + i\Sigma_{j=1}^n(\epsilon_j/2)ds_j \wedge ds_j$, where $\epsilon_j = \pm 1$. In order to prove the above theorem, presented is Frobenius' theorem on the graded manifold[7] with "Hermiticity," which, important by itself, reads as; let $\{\omega_i\}$ $(i = 1,\dots,r)$ be a set of linearly independent "Hermitian" 1-forms with homogeneous Z_2 grading, and if there exists 1-forms θ_{ji}, such that $d\omega_i = \omega_j \wedge \theta_{ji}$, then there exists a set of df_i $(i = 1,..,r)$ of homogeneous linearly independent "Hermitian" 1-forms, and a nonsingular $r\times r$ matrix g_{ji} such that $\omega_i = (df_j)g_{ji}$. In the course of showing Darboux's theorem, we have a lemma which says if there exists a set of Hermitian 0-forms, $\{h_i\}$ $(i = 1,\dots,k \leqslant m + n)$ such that dh_i's are homogeneous linearly independent 1-forms, and $\{h_i\}$ satisfies Poisson bracket relations in such a

[a)] On leave of absence from Utsunomiya University, Department of Engineering Mathematics, Utsunomiya, Japan.

way as a subset of canonical coordinate system does, then there exists an extension of $\{h_i\}$ to a canonical coordinate system. One will find the above form of ω fits in with the naive conventional quantization procedure.[8] Now we can understand canonical transformations in terms of differential forms as in Bose systems cases. Furthermore, we can answer to a basic problem of constrained canonical Bose–Fermi systems, that is, existence of constraints in "standard" form,[9] which will appear in a separate paper. In Sec. 2, a graded manifold theory is given with emphasis on concept of "Hermiticity" of graded algebra elements, graded differential forms, and graded vector fields, respectively, which are nontrivially modified parts from Kostant's work, and Frobenius' theorem is given. We ask readers to see his work for the basic concepts and definitions, more in detail, of graded manifolds and graded differential forms and so forth, which are not necessarily repeated here. Many notations and terminology are borrowed from it. In Sec. 3, we state Darboux's theorem and a related lemma and survey characterization of canonical Bose–Fermi system by the fundamental 2-form. Section 4 is devoted to some discussions on canonical transformations and on an inter-relation of a conventional quantization procedure and the "classical" system presented here, with emphasis on "Hermiticity." The Appendix is devoted to the proof of Darboux's theorem.

2. GRADED MANIFOLD THEORY WITH "HERMITICITY"

A. Graded manifolds with "Hermiticity"

Here we will show that by introducing "Hermiticity" into a graded commutative algebra over C (complex number) in the course of Kostant's construction of graded manifolds, the concept of graded manifolds with "Hermite" local coordinate systems with complex sheaf, can be defined. We begin with summing up necessary definitions.

A: Sheaf of graded Z_2 commutative algebra over C (complex number) on X (a real manifold of dim m).

Sheaf A over X: Correspondence for any open set $U\subset X$; $U\to A(U)$, with $A(U)$ an abstract set, such that

(i) $\exists$ restriction map $\rho_{U,V}$; $A(U)$, for open sets $\forall U\supset V$.

(ii) $\rho_{V,W}\cdot\rho_{U,V}=\rho_{U,W}$ if $W\subset V\subset U$.

(iii) Let $\cup_{i\in\Lambda} U_i$ be any open convering of any open set U. If $\rho_{U,U_i}(f)=\rho_{U,U_i}(g)$ for all $i\in\Lambda$, $f,g\in A(U)$, then $f\equiv g\in A(U)$.

(iv) If $\exists\ h_i\in A(U_i)$ such that $\rho_{U_i,U_i\cap U_j}(h_i)=\rho_{U_j,U_i\cap U_j}(h_j)$ for all $i,j\in\Lambda$, then uniquely $\exists\ h\in A(U)$ such that $\rho_{U,U_i}(h)=h_i$.

If $A(U)$ has an algebraic structure, it is assumed in the following the restriction maps are morphisms of the algebra structure.

Graded commutative algebra $A(U)$: $A(U)=A(U)_0\oplus A(U)_1$, $A(U)_0$; a set of even elements, $A(U)_1$; a set of odd elements. Notation for Z_2 grading; $|f|=0$ if $f\in A(U)_0$, $|f|=1$ if $f\in A(U)_1$. An element $f\in A(U)$ for which $|f|$ is defined, is called (Z_2) homogeneous. For homogeneous f, g, $fg=gf(-1)^{|f||g|}$ which implies f and g graded commutative, and $|fg|=|f|+|g|$.

We assume that for any open set $U\subset X$ there is a homomorphism of graded algebra, $\tilde{}$:

$$A(U)\to\mathscr{C}^\infty(U),\quad f\to\tilde{f},\tag{2A.1}$$

which commutes with restriction maps, where $\mathscr{C}^\infty(U)$ denotes a set of complex valued C^∞ functions. $\mathscr{C}^\infty(U)$ is graded in the sense $\mathscr{C}^\infty(U)=(\mathscr{C}^\infty(U))_0$, so that $\tilde{f}=0$ if $f\in(A(U))_1$.

Function factor $\mathscr{C}(U)$ of $A(U)$: A subalgebra $\mathscr{C}(U)\subset A(U)_0$, $\mathscr{C}(U)\ni 1_U$ and the map,

$$\mathscr{C}(U)\to\mathscr{C}^\infty(U),\quad f\to\tilde{f},\tag{2A.2}$$

is an algebra isomorphism.

Exterior factor $D(U)$ of $A(U)$: A subalgebra $D(U)\subset A(U)$ $D(U)$ is generated by 1_U and n algebraically independent odd elements for some n, where n is assumed a maximum of those, dim $D(U)=2^n$, where algebraically independent odd elements $(s_1,\dots,s_k)$;

$$s_i\in A(U)_1\quad(i=1,\dots,k)\ \text{and}\ s_1s_2\cdots s_k\neq 0.\tag{2A.3}$$

Splitting factor $(\mathscr{C}(U),D(U))$ for $A(U)$: $\mathscr{C}(U)$; a function factor of $A(U)$. $D(U)$; an exterior factor of $A(U)$ and the map,

$$\mathscr{C}(U)\otimes_C D(U)\to A(U),\quad f\otimes w\to fw,\tag{2A.4}$$

is a linear isomorphism.

A-splitting neighborhood U of odd dim n: An open set $U\subset X$ such that there exists a splitting factor $(\mathscr{C}(U),D(U))$ for $A(U)$, where dim $D(U)=2^n$.

Graded manifold (X,A) of dim (m,n) *with complex sheaf A*: X; m dimensional ordinary manifold. For any nonempty open set $U\subset X$, $\exists$ covering of U of A-splitting neighborhoods of odd dim n.

$\mathscr{C}^\infty(U)$ is a complexification of $C^\infty(U)$ as a vector space, and there exists a unique decomposition $\mathscr{C}^\infty(U)=C^\infty(U)\oplus iC^\infty(U)$. Therefore, once $\mathscr{C}(U)$ given $\exists C(U)\cong C^\infty(U)$ such that $\mathscr{C}(U)=C(U)\oplus iC(U)$ (unique decomposition). We call $C(U)$ as a "real" function factor, which is a subalgebra over R.

A-splitting neighborhood has a covering of coordinate neighborhoods U, where LCS (local coordinate system) $(u_1,\dots,u_m)$, $u_i\in C^\infty(U)$, represents $\forall f\in C^\infty(U)$ as $f=f(u)$ with a suitable $f(\cdot)\in C^\infty(R^m)$, i.e., $C^\infty(U)\cong C^\infty(R^m)$ locally. U is also an A-splitting neighborhood, and $r_i\in C(U)$ $(i=1,\dots,m)$ such that $\tilde{r}_i=u_i$ and thus one may adopt a notation for all $f\in C(U)$ such that $f=f(r)$ if $\tilde{f}=f(u)$ where $f(\cdot)\in C^\infty(R^m)$. Let $D(U)$ be generated by 1_U and odd elements $\{s_j; j=1,\dots,n\}$ which are called odd coordinates. The above $\{r_i; i=1,\dots,m\}$ is called even coordinates.

A-CS (A-coordinate system): $(r,\dots,r_m;s_1,\dots,s_n)$; $\{r_i\}$ even coordinates and $\{s_j\}$ odd coordinates, defined above. $\forall f\in A(U)$ can be uniquely written in terms of $(r_i;s_j)$ as

$$f=\sum_{(\nu)\leqslant n}(f_\nu(r)+ig_\nu(r))s^\nu,\tag{2A.5}$$

where ν denotes $(\nu_1,\nu_2,\dots,\nu_n)$ with ν_j equal to 1 or 0 and $(\nu)\equiv\sum_{j=1}^n\nu_j\leqslant n$, and s^ν denotes $s_1^{\nu_1}s_2^{\nu_2}\cdots s_n^{\nu_n}$.

A-coordinate neighborhood: An open set $U\subset X$, which is an A-splitting neighborhood and is a coordinate neighborhood of X, i.e., a neighborhood admitting A-CS.

Once given even coordinates $\{r_i\}$, then $\exists$ unique "real" function factor $C(U)$ containing $\{r_i\}$, but there may be another "real" function factor $C'(U)$ containing $\{h_i\}$ even if $\tilde{h}_i = \tilde{r}_i = u_i$, i.e., $h_i = r_i + z_i, z_i \in A^1(U)$, a set of nilpotent elements.

Note that as for complex structure of $A^1(U)$ part, it is still an open question. Therefore in the sense that $C(U)$ is not uniquely determined owing to its $A^1(U)$ part, i.e., splitting $A(U) = \mathscr{C}(U) \oplus A^1(U)$ is not unique, the question of complex structure of $C(U)$ is also open. Yet, once $C(U)$ is given, we may call $C(U)$ as a "real," not "Hermite" in the sense defined later, function factor with respect to $C(U) \simeq C^\infty(U)$.

On transformations of A-CS, we have the following proposition.

Proposition 1: Let U admit A-CS $(r_i; s_j; i = 1,...,m; j = 1, ..., n)$, where $(\tilde{r}_i)$ is a LCS for U. Let $r_i' \in A(U)_0$, $s_j' \in A(U)_1$ such that

$$\begin{cases} r_i' = a_i(r) + \sum\limits_{\substack{(\nu) \geq 2 \\ (\nu)\ \text{even}}} b_{i\nu}(r) s^\nu \quad (i = 1,...,m), \\ s_j' = \sum\limits_{k=1}^{n} c_{jk}(r) s_k + \sum\limits_{\substack{(\nu) \geq 3 \\ (\nu)\ \text{odd}}} d_{j\nu}(r) s^\nu \quad (j = 1,...,n). \end{cases} \tag{2A.6}$$

If $a_i(r) \in C(U)$; $b_{i\nu}(r)$, $c_{jk}(r)$, $d_{j\nu}(r) \in \mathscr{C}(U)$ and

$$\det \widetilde{\frac{\partial(r',s')}{\partial(r,s)}} = \det \begin{pmatrix} \widetilde{\partial_j a_i} & 0 \\ 0 & \tilde{c}_{jk} \end{pmatrix} \neq 0, \tag{2A.7}$$

where derivation may be of left or right, holds everywhere in U, then U admits A-CS $(r_i'; s_j'; i = 1,...,m; j = 1,...,n)$.

Proof is accomplished in three steps. Show U admits firstly A-CS $(r_i'' \equiv a_i(r); s_j)$, utilizing the invertibility of ordinary coordinate transformation and $C(U) \cong C^\infty(U)$; secondly A-CS $(r_i''; s_j')$, utilizing iterative process and nilpotency; and finally A-CS $(r_i'; s_j')$, utilizing Taylor's expansion and iterative process with nilpotency.

Proposition 1 amounts to

Corollary 2: If $(\widetilde{r_i'})$ is a LCS for U and $(\widetilde{\partial^n(s_1' \cdots s_n')/\partial s_1 \cdots \partial s_n})$ is nonvanishing everywhere in U then $(r_i'; s_j')$ is A-CS.

Now we introduce complex structure on $A(X)$.

"*Hermitian*" *conjugation*†: Antiautomorphism of graded algebra, $^\dagger: A(U) \to A(U)$, which commutes with restriction map, such that $\forall a, b \in A(U)$, $(a+b)^\dagger = a^\dagger + b^\dagger$, $(ab)^\dagger = b^\dagger a^\dagger$, $(ca)^\dagger = c^* a^\dagger$, where $c \in C$ (complex number), $(a^\dagger)^\dagger = a$, with $(\tilde{a}^\dagger) = \tilde{a}^*$ where $*$ is the complex conjugation.

$$(A(U)_0)^\dagger = A(U)_0,\ (A(U)_1)^\dagger = A(U)_1 \text{ and } 1_U{}^\dagger = 1_U$$

can be easily seen.

Remark 3: $\dagger$ induces an linear isomorphism;

$$(\mathscr{C}(U))^\dagger \otimes_C (D(U))^\dagger \to A(U),$$

$$f \otimes w \to fw, \tag{2A.8}$$

where $f \in (\mathscr{C}(U))^\dagger$, $w \in (D(U))^\dagger$, and $(D(U))^\dagger$ is an exterior factor generated by algebraically independent $\{s_j{}^\dagger\}$ $(i = 1,...,n)$ and 1_U.

The above follows from $(A(U))^\dagger = A(U)$ and linear isomorphism; $\mathscr{C}(U) \otimes_C D(U) \mapsto A(U), f \otimes w \to fw$, where $f \in \mathscr{C}(U)$, $w \in D(U)$.

Remark 4; The mapping $\sim$ is an algebra isomorphism,

$$\sim : (\mathscr{C}(U))^\dagger \to \mathscr{C}^\infty(U), f \to \tilde{f}. \tag{2A.9}$$

Thus $(\mathscr{C}(U))^\dagger$ is a function factor.

One first notes $(\mathscr{C}(U))^\dagger$ is a subalgebra of $A(U)_0$ containing 1_U, and finds, a mapping $\tilde{\dagger}^*: (\mathscr{C}(U))^\dagger \to \mathscr{C}^\infty(U)$ gives an algebra isomorphism, since so is the map $\sim: \mathscr{C}(U) \mapsto \mathscr{C}^\infty(U)$ and the mapping $\tilde{\dagger}^* \equiv \tilde{\ }$, which finishes the proof.

Thus one has

Remark 5: If $(\mathscr{C}(U), D(U))$ is a splitting factor for $A(U)$, then $(\mathscr{C}(U))^\dagger$, $(D(U))^\dagger$ is, also.

Let $\mathscr{C}(U) = C(U) \oplus iC(U)$, a direct sum such that an algebra isomorphism $\sim: C(U) \mapsto C^\infty(U)$, (both, R-module understood). Then $(\mathscr{C}(U))^\dagger = (C(U))^\dagger \oplus i(C(U))^\dagger$ is a direct sum, where $\sim: (C(U))^\dagger \to C^\infty(U)$, an algebra isomorphism, i.e., $(C(U))^\dagger$ is a "real" function factor as $C(U)$. One should note here that for $f \in C(U)$, $\widetilde{f^\dagger} = \tilde{f}^* = \tilde{f}$ holds, yet, not necessarily, $f^\dagger = f$ and even $(C(U))^\dagger = C(U)$ follow. For example, let $\widetilde{r_i{}^\dagger} = \widetilde{r_i}$, $s_i{}^\dagger = s_i$, $r_i' \equiv r_i + s_1 s_2$, then $r_i'^\dagger \neq r_i'$ with $\widetilde{r_i'^\dagger} = \widetilde{r_i'}$, and consider a function factor $C(U)$ containing r_i' $(i = 1,2)$ with dimension of U, 2, then $(C(U))^\dagger \neq C(U)$.

We say $a \in A(U)$ is "*Hermite*" if $a^\dagger = a$, "*anti-Hermite*" if $a^\dagger = -a$, and $C(U)$ is a "*Hermite*" *function factor* if $\forall\ a \in C(U)$ is "Hermite".

The above situation implies a "real" function factor is not necessarily a "Hermite" one.

We know from definitions, a graded manifold (X, A) has a covering of A-coordinate neighborhoods U, and let $(r_i; s_j$: $i = 1,...,m; j = 1,...,n)$ be A-CS on U. Then $\{\widetilde{r_i{}^\dagger} = \tilde{r}_i = u_i\}$ (real LCS on U). Let $r_i' \equiv \frac{1}{2}(r_i + r_i{}^\dagger)$, then $\tilde{r}_i' = \tilde{r}_i = u_i$ and one finds $(r_i'; s_j)$ is A-CS from Corollary 2. Note $r_i'^\dagger = r_i'$ ("Hermite"). Let $s_j{}^r \equiv \frac{1}{2}(s_j + s_j{}^\dagger)$ ("Hermite"), $s_j{}^a \equiv \frac{1}{2}(s_j - s_j{}^\dagger)$ ("anti-Hermite"), then $s_j = s_j{}^r + s_j{}^a$ and $s_1 \cdots s_n = \Sigma_{\alpha_i = r,a} s_1{}^{\alpha_1} s_2{}^{\alpha_2} \cdots s_n{}^{\alpha_n}$ which gives

$$1 = \sum_{\alpha_i = r,a} \left(\frac{\partial^n}{\partial s_n \cdots \partial s_1} s_1{}^{\alpha_1} \cdots s_n{}^{\alpha_n} \right). \tag{2A.10}$$

It is impossible for all terms to vanish at any point in U. The U has an open covering of U_λ, $\lambda \in A$, such that on U_λ, $\lambda \in A$, there exists an A-CS $(r_i'; s_j{}^{\alpha_j^\lambda})$ and, thus, a "Hermite" A-CS $(r_i'; s_j')$.

We say an A-CS is a HCS ("*Hermite*" *coordinate system*) if all coordinates are "Hermite", and call a graded manifold (X, A) as *one with* "*Hermiticity*", if "Hermitian" conjugation is defined on A. Thus the definition of a graded manifold (X, A) with "Hermiticity" amounts to say

Proposition 6: A graded manifold (X, A) with "Hermiticity" has a covering of neighborhoods, each of which admits HCS.

Owing to $C(U) \subset A(U)_0$, we have (2A.11)

Remark 7: Let $C(U)$ be a "real" function factor containing "Hermite" local coordinates (r_i), then $C(U)$ is a "Hermite" function factor, i.e., for all $f \in C(U), f^\dagger = f$ and $(C(U))^\dagger = C(U)$.

Note again here the above $C(U)$ forms a subalgebra over R.

Remark 8: $\forall a \in A(U)$ is uniquely decomposed into "Hermitian" and "anti-Hermitian" parts, i.e., $A(U) = A^h(U) \oplus iA^a(U)$, (direct sum), where $A^h(U) = \{(a + a^\dagger)/2 | a \in A(U)\}, A^a(U) = \{(a - a^\dagger)/(2i) | a \in A(U)\}$. And $A^h(U) = A^a(U)$, where $A^h(U)$; R-module, and $A(U) \cong C \otimes_R A^h(U)$; R-linear isomorphism.

Note $A^h(U)$ is not a subalgebra even over R.

Remark 9: Let $D(U)$ be an exterior factor, a subalgebra over C, generated by algebraically independent "Hermitian" odd elements $(s_j : j = 1,...,n)$ and 1_U. Then $D(U) = D^h(U) \oplus iD^h(U) \cong C \otimes_R D^h(U)$ (direct sum), where $D^h(U)$ denotes a "Hermite" R-module whose basis is given by 2^n elements $i^{((\nu)/2)((\nu)-1)} s^\nu$, $((\nu) = 0,1,..,n)$, where ν denotes a set $(\nu_1, \nu_2,...,\nu_n)$, $\nu_j = 0$ or 1, and $(\nu) = \Sigma_j \nu_j$ and $s^\nu = s_1^{\nu_1} \cdots s_n^{\nu_n}$ with for $(\nu) = 0$, $s^\nu = 1_U$.

$D^h(U)$ is not a subalgebra over R.

Proposition 10: Let $(\mathscr{C}(U), D(U))$ be a splitting factor for $A(U)$ such that $\mathscr{C}(U) = C(U) \oplus iC(U) \cong C \otimes_R C(U)$, and $C(U)$ be a "Hermite" function factor and $D(U)$ be as in Remark 9. Then there exists a R-linear isomorphism: $C(U) \otimes_R D^h(U) \to A^h(U)$ with an algebra isomorphism $\sim$: $C(U) \to C^\infty(U)$, where $A(U) = A^h(U) \oplus iA^h(U)$, i.e., using HCS explicitly, we have for all $f \in A(U)$, unique expression, $f = \Sigma_\nu (f_\nu^{\,1}(r) + i f_\nu^{\,2}(r)) \, i^{((\nu)/2)((\nu)-1)} s^\nu$, where $f_\nu^1, f_\nu^2 \in C(U)$, $r_i^\dagger = r_i$, $s_j^\dagger = s_j$.

Now we see the assumption of "antiautomorphism" on a graded manifold (X, A) of dim (m, n) with complex sheaf A, results in saying any elements of A can be uniquely expressed locally as above.

We again emphasize the difference between the complexification here and that of Kostant, i.e., † and *, where $(ab)^\dagger = b^\dagger a^\dagger$ and $(ab)^* = a^* b^*$, which changes reality concept drastically, especially in the exterior factor; one can say that "real" means no appearance of i in * case, and suitable appearance of i in † case,[10] using real coordinates by definition in both cases, and that the real part forms a subalgebra over R in * case, and does not in † case.

B. "Hermiticity" on derivations on the graded manifold

Here we show it is possible to define "Hermiticity" on a space of derivations. For that purpose we have to define a C-linear mapping, transposition$'$: $\mathrm{Der}^L A(U) \to \mathrm{Der}^R A(U)$, $\mathrm{Der}^R A(U) \to \mathrm{Der}^L A(U)$ and C-antilinear mapping "Hermitian" conjugation†: $\mathrm{Der}^L A(U) \to \mathrm{Der}^R A(U)$, $\mathrm{Der}^R A(U) \to \mathrm{Der}^L A(U)$, where $\mathrm{Der}^L A(U)$ and $\mathrm{Der}^R A(U)$ denote a set of left-derivations and one of right-derivations, respectively. For X ($\in \mathrm{Der}^L A(U)$ or $\mathrm{Der}^R A(U)$; hereafter denoted, $\in \mathrm{Der}^{(R)}_{\;L} A(U)$) if X is invariant under successive operations of $'$ and †, then X is called "Hermite", and if X changes to $-X$, then "anti-Hermite".

We begin with summing up necessary definitions.

Left-derivation X, with quadratic form $x\,y$, of $A(U)$: $X \in \mathrm{End}\,A(U)$, $A(U)$: C-module, where $\mathrm{End}\,A(U)$ [endomorphism of $A(U)$] is a graded vector space over C of homomorphism of graded vector space $A(U)$ over C into itself. $X = X_0 + X_1$, and for homogeneous $x, y \in A(U)$ and homogeneous $X_j (j = 0,1)$, it holds that

$$X_j(xy) = (X_j(x))y + (-1)^{j|x|} x(X_j(y)). \tag{2B.1}$$

Right-derivation X, with quadratic form $x\,y$, of $A(U)$: $X \in \mathrm{End}A(U)$, $A(U)$: C-module, $X = X_0 + X_1$, and for homogeneous $x, y \in A(U)$ and homogeneous X_j, it holds that

$$X_j(xy) = x(X_j(Y)) + (-1)^{j|y|}(X_j(x))y. \tag{2B.2}$$

$\mathrm{Der}^L A(U)$: A set of all left derivations of $A(U)$

$\mathrm{Der}^R A(U)$: A set of all right derivations of $A(U)$

Note that $\mathrm{Der}^L A(U)$ is a left $A(U)$-module, and $\mathrm{Der}^R A(U)$ right one.

Transposition$'$: $\mathrm{Der}^{(R)}_{\;L} A(U) \to \mathrm{Der}^{(L)}_{\;R} A(U)$: $'$ is C-linear, for $\mathrm{Der}^{(R)}_{\;L} A(U) \ni X$, $X' = X_0' + X_1'$, and $X' \in \mathrm{End}\,A(U)$ and for homogeneous X_j, x,

$$X_j'(x) = (-1)^{j(1+|x|)} X_j(x), \tag{2B.3}$$

from which one finds if $X \in \mathrm{Der}^L A(U)$ ($X \in \mathrm{Der}^R A(U)$), then $X' \in \mathrm{Der}^R A(U)$ ($X' \in \mathrm{Der}^L A(U)$).

It is shown that $\mathrm{Der}^{(R)}_{\;L} A(U)$ forms a GLA (graded Lie algebra) with graded commutator, bilinear operation $[X, Y]$, where for homogeneous X, $Y \in \mathrm{Der}^{(R)}_{\;L} A(U)$,

$$[X,Y] \equiv X \circ Y - (-1)^{|X||Y|} Y \circ X, \in \mathrm{Der}^{(R)}_{\;L} A(U), \text{ i.e.,}$$

$$[X,Y] = -(-1)^{|X||Y|}[Y, X] \text{ and Jacobi identity,}$$

$$\sum_{\text{cyclic perm}} (-1)^{|X||Z|}[X,[Y,Z]] = 0 \tag{2B.4}$$

We note even if we change Lie bracket of $\mathrm{Der}^R A(U)$, as $[X, Y]' \equiv (-1)^{|X||Y|}[X,Y]$, then $\mathrm{Der}^R A(U)$ forms a GLA, and then we can say the transposition$'$ defines a GLA isomorphism.

Remark 11: For $X \in \mathrm{Der}^{(R)}_{\;L} A(U)$, $(X')' = X$. $(X_1 + X_2)' = X_1' + X_2'$, $[Y,Z]' = (-1)^{|Y||Z|}[Y', Z',]$ where $X_1, X_2, Y, Z \in \mathrm{Der}^{(R)}_{\;L} A(U)$, and Y, Z, homogeneous.

As we know $\mathrm{Der}^L A(U)$ and $\mathrm{Der}^R A(U)$ are left and right $A(U)$-module, let us see how $'$ works:

Remark 12: Let $X \in \mathrm{Der}^{(R)}_{\;L} A(U)$, $a \in A(U)$ and both homogeneous, and $a \circ$ denotes left (right) multiplication, and $a'\circ$, right (left) multiplication. Then

$$(a \circ X)' = (-1)^{|a|(1+|X|)} a' \circ X'. \tag{2B.5}$$

Note the sign factor $(-1)^{|a|(1+|X|)}$ is not $(-1)^{|a||X|}$, and so $'$ operation should not be mistaken for left–right module structure change.

In explicit use of an A-CS (ξ^μ) $(\mu = 1,...,m+n)$, $\mathrm{Der}^L A(U) \ni X$ is uniquely written,

$$X = X^\mu \partial_\mu, \text{ where } \partial_\mu(\xi^\nu) = 1_U \delta_\mu{}^\nu, X^\mu \in A(U). \quad (2B.6)$$

Then $\partial_\mu{}'(\xi^\nu) = 1_U \delta_\mu{}^\nu$ or $\partial_\mu{}' \equiv \overleftarrow{\partial}_\mu$, and

$$X' = \overleftarrow{\partial}_\mu X^\mu(-1)^{|X|(1+|\mu|)}, \text{ where } |\mu| \equiv |\xi^\mu|. \quad (2B.7)$$

Now let us define similarly a C-antilinear operation, "*Hermitian*" *conjugation*†: $\mathrm{Der}^{(R)}_{L} A(U) \to \mathrm{Der}^{(L)}_{R} A(U)$, by:

For all $X \in \overset{\mathrm{L}}{\mathrm{Der}^{(\mathrm{R})}} A(U), {}^\forall x \in A(U), X^\dagger(x) = (X(x^\dagger))^\dagger$. (2B.8)

In fact, we see the above relation defines a C-antilinear mapping from $\overset{\mathrm{L}}{\mathrm{Der}^{(\mathrm{R})}} A(U)$ to $\overset{\mathrm{R}}{\mathrm{Der}^{(\mathrm{L})}} A(U)$.

We easily see from the definition,

Remark 13: For $\overset{\mathrm{L}}{\mathrm{Der}^{(\mathrm{R})}} A(U) \ni X$, $(X^\dagger)^\dagger = X$.

We note, for $\overset{\mathrm{L}}{\mathrm{Der}^{(\mathrm{R})}} A(U) \ni X, Y$, $(X+Y)^\dagger = X^\dagger + Y^\dagger$, $(\lambda X)^\dagger = \lambda^* X^\dagger$, where $\lambda \in C$, $[X, Y]^\dagger = [X^\dagger, Y^\dagger]$.

Remark 14: Mappings † and $'$, successively operated on $\overset{\mathrm{L}}{\mathrm{Der}^{(\mathrm{R})}} A(U)$, commute, i.e., ${}'\circ{}^\dagger\circ X \equiv X^{\dagger\prime} = X^{\prime\dagger} \equiv {}^\dagger\circ{}'\circ X$, $X \in \overset{\mathrm{L}}{\mathrm{Der}^{(\mathrm{R})}} A(U)$.

Now we have

Remark 15: The C-antilinear mapping $({}'\circ{}^\dagger)$ is a bijection: $\overset{\mathrm{L}}{\mathrm{Der}^{(\mathrm{R})}} A(U) \to \overset{\mathrm{L}}{\mathrm{Der}^{(\mathrm{R})}} A(U)$ such that $[X, Y]^{\dagger\prime} = (-1)^{|X||Y|}[X^{\dagger\prime}, Y^{\dagger\prime}]$ for homogeneous X, Y.

Here we come to define "*Hermiticity*" *on* $\overset{\mathrm{L}}{\mathrm{Der}^{(\mathrm{R})}} A(U)$. We call $X \in \overset{\mathrm{L}}{\mathrm{Der}^{(\mathrm{R})}} A(U)$ is "Hermite" if $X^{\dagger\prime} = X$, and "anti-Hermite" if $X^{\dagger\prime} = -X$.

Remark 16: $\overset{\mathrm{L}}{\mathrm{Der}^{(\mathrm{R})}} A(U) \ni X$ is uniquely decomposed into "Hermitian" and "anti-Hermitian" parts.

The proof is as usual.

Let us see component expressions of † and $^{\dagger\prime}$ where we note if (ξ^μ) $(\mu = 1,...,m+n)$ is an A-CS, then so is $(\xi^{\mu\dagger})$ $(\mu = 1,...,m+n)$, which follows from Remark 5.

Component expression of $X^\dagger$:

Let $\mathrm{Der}^L A(U) \ni X = X^\mu(\partial/\partial\xi^\mu)$ in the A-CS (ξ^μ), where $X^\mu = X^\mu(\xi) \in A(U)$, then $X^\dagger = (\overleftarrow{\partial}/\partial\xi^{\mu\dagger})X^{\mu\dagger}$ in the A-CS $(\xi^{\mu\dagger})$, where if (ξ^μ) is a HCS, then $X^\dagger = (\overleftarrow{\partial}/\partial\xi^\mu)X^{\mu\dagger}$.

Component expression of $X^{\dagger\prime}$:

Let homogeneous $X \in \mathrm{Der}^L A(U)$, $X = X^\mu(\partial/\partial\xi^\mu)$ in the A-CS (ξ^μ), then $X^{\dagger\prime} = (-1)^{|X|(1+|\mu|)}X^{\mu\dagger}(\partial/\partial\xi^{\mu\dagger})$, where if (ξ^μ) is a HCS, then $X^{\dagger\prime} = (-1)^{|X|(1+|\mu|)}X^{\mu\dagger}(\partial/\partial\xi^\mu)$.

Component conditions in the HCS (ξ^μ) for "Hermiticity" of X: Let $X \in \mathrm{Der}^L A(U)$ $(\in \mathrm{Der}^R A(U))$, $X = X^\mu(\partial/\partial\xi^\mu)$ $(X = \overleftarrow{\partial}/\partial\xi^\mu X^\mu)$, and X homogeneous, then $X^{\mu\dagger} = (-1)^{|X|(1+|\mu|)}X^\mu$ in both cases.

For example, we recommend readers to check consistency in the following case by coordinate transformation; let ξ^μ be a HCS (r, s^1, s^2), where $(\partial/\partial\xi^\mu)$ all "Hermite", and ξ'^μ be a HCS (r', s'^1, s'^2), where $(\partial/\partial\xi'^\mu)$ all "Hermite", as $r' = r + is^1 s^2, s'^1 = s^1, s'^2 = s^2$.

C. "Hermiticity" on graded differential forms

In order to define "Hermiticity" on graded differential forms, we need to introduce graded differential forms both of left and right versions, as in the case of derivations. Definition of graded differential forms (left version) was given clearly by Kostant in his graded manifold theory[5], and for the purpose here, we may consider, his definition is essentially sufficient to start our discussion, only by replacing his graded manifold with the one here. And for the right version, the parallel definition with complete inversion of left side right, will work. For completeness of discussion, we review the definitions, in a little different way utilizing a graded antisymmetrization operation $\mathfrak{A}$, as in many textbooks for ordinary differential forms.

$\overset{\mathrm{L}}{\mathrm{Der}^{(\mathrm{R})}} A(U)$: GLA and a left (right) module over the graded commutative algebra $A(U)$, understood left (right) multiplication operation. As noted by Kostant in general, however, we may define right (left) module structure for $\mathrm{Der}^L A(U)$ $(\mathrm{Der}^R A(U))$ by considering them as graded vector space, as follows; for homogeneous $X \in \mathrm{Der}^L A(U)$, $a \in A(U)$, $Xa \equiv (-1)^{|a||X|} aX \in \mathrm{Der}^L A(U)$. Similarly, for $\mathrm{Der}^R A(U)$, left module structure is defined. Hereafter, the module structure of $\overset{\mathrm{L}}{\mathrm{Der}^{(\mathrm{R})}} A(U)$ is understood as above.

$\overset{\mathrm{L}}{T^{(\mathrm{R})}}(U)$: Tensor algebra over $\overset{\mathrm{L}}{\mathrm{Der}^{(\mathrm{R})}} A(U)$ over $A(U)$ such that for X, $Y \in \mathrm{Der}^L A(U)$ and $x \in A(U)$, $xX \otimes Y = x(X \otimes Y)$, $X \otimes Yx = (X \otimes Y)x$, $Xx \otimes Y = X \otimes xY$. By iterative tensor products, we define $T^L(U)$, similarly $T^R(U)$, replacing $\mathrm{Der}^L A(U)$ with $\mathrm{Der}^R A(U)$. Then $\overset{\mathrm{L}}{T^{(\mathrm{R})}} A(U)$ is a bigraded (Z_+, Z_2) algebra, where Z_+ denotes a set of non-negative integer.

$\overset{b\,\mathrm{L}}{T^{(\mathrm{R})}}(U)$: Homogeneous space in that $b \in Z_+$, where $\overset{0\,\mathrm{L}}{T^{(\mathrm{R})}}(U) \equiv A(U)$.

$\overset{\mathrm{L}}{I^{(\mathrm{R})}}(U)$: Two sided bigraded ideal in $\overset{\mathrm{L}}{T^{(\mathrm{R})}}(U)$ generated by all elements in $\overset{2\,\mathrm{L}}{T^{(\mathrm{R})}}(U)$ of the form

$$X \otimes Y + (-1)^{|X||Y|} Y \otimes X \text{ (graded symmetric)}, \quad (2C.1)$$

where X, $Y \in \overset{\mathrm{L}}{\mathrm{Der}^{(\mathrm{R})}} A(U)$.

$\overset{b\,\mathrm{L}}{I^{(\mathrm{R})}}(U)$: $\overset{b\,\mathrm{L}}{I^{(\mathrm{R})}}(U) \equiv \overset{\mathrm{L}}{I^{(\mathrm{R})}}(U) \cap \overset{b\,\mathrm{L}}{T^{(\mathrm{R})}}(U)$.

$\overset{b\,\mathrm{L}}{J^{(\mathrm{R})}}(U)$: A quotient space, $\overset{b\,\mathrm{L}}{J^{(\mathrm{R})}}(U) \equiv \overset{b\,\mathrm{L}}{T^{(\mathrm{R})}}(U) / \overset{b\,\mathrm{L}}{I^{(\mathrm{R})}}(U)$.

Graded antisymmetrization operator $\mathfrak{A}$:

$\mathfrak{A}$: $\overset{\mathrm{L}}{T^{(\mathrm{R})}}(U) \to \overset{\mathrm{L}}{T^{(\mathrm{R})}}(U)$, is a $A(U)$-linear mapping as follows. Hereafter as far as Proposition 19, superscripts$\overset{\mathrm{L}}{^{(\mathrm{R})}}$ are deleted, no mixing of L and R understood. Let $\{X_\mu\}$ ($\mu = 1,\ldots,n$) be a basis for $\mathrm{Der}\,A(U)$, where $|X_\mu| \equiv |\mu| = 0, 1$. For all $Z \in T^n(U)$,

$$Z = Z^{\mu_1\cdots\mu_n} X_{\mu_1} \otimes \cdots \otimes X_{\mu_n} \; [\text{simply} \equiv Z^\mu(\otimes X)_\mu]. \quad (2\mathrm{C}.2)$$

Let $\mathfrak{A}$ be (left and right) $A(U)$-linear mapping: $T^n(U) \to T^n(U)$, as

$$\mathfrak{A}Z \equiv \frac{1}{n!} Z^\mu \sum_\tau \epsilon(\tau;\mu)(\otimes X)_{\tau\mu}, \quad \tau\text{: permutation}, \quad (2\mathrm{C}.3)$$

where $\epsilon(\tau;\mu)$ is given;

$$\begin{pmatrix} (\mu) = (\mu_1, \mu_2, \ldots, \mu_n) \\ \\ (\tau\mu) = (\mu_n, \ldots, \mu_2, \ldots, \mu_1) \end{pmatrix},$$

$$\epsilon(\tau;\mu) \equiv \prod_{\substack{\text{crossing} \\ (i,j)}} \epsilon(i,j),$$

$$\epsilon(i,j) \equiv (-1)^{(1 + |\mu_i||\mu_j|)}. \quad (2\mathrm{C}.4)$$

It is easily seen from the above illustration that $\epsilon(\tau;\mu)$ is invariant under any deformations of the lines such that crossings be interchanging crossings, with fixed initial and final order, which reads for any permutations, τ_1, τ_2,

$$\epsilon(\tau_2\tau_1;\mu) = \epsilon(\tau_2;\tau_1\mu)\epsilon(\tau_1;\mu). \quad (2\mathrm{C}.5)$$

From $\epsilon(I;\mu) = 1$ for all μ,

$$1 = \epsilon(\tau^{-1};\tau\mu)\epsilon(\tau;\mu). \quad (2\mathrm{C}.6)$$

Thus we have

$$\epsilon(\tau^{-1};\tau\mu) = \epsilon(\tau;\mu). \quad (2\mathrm{C}.7)$$

Note, before proceeding, the above definition is independent of a basis $\{X^\mu\}$ chosen, and note consistency of (left and right) $A(U)$-linearity.

From the above grouplike property of the signature factor, we easily obtain

Remark 17: $\mathfrak{A} \circ \mathfrak{A} = \mathfrak{A}$, i.e., $\mathfrak{A}$ is a projection, defined on $T^n(U)$.

Remark 18: $I^n(U) = (I - \mathfrak{A}) \circ T^n(U)$.

The proof goes firstly by noting $\mathfrak{A} \circ I^n(U) = 0$, which is easy and implies $I^n(U) \subset (I\text{-}\mathfrak{A}) \circ T^n(U)$, and secondly by noting any permutation τ can be decomposed into transpositions of neighboring two elements, $(I\text{-}\mathfrak{A}) \circ T^n(U) \subset I^n(U)$.

Thus we have from decomposition,

$$T^n(U) = \mathfrak{A} \circ T^n(U) \oplus I^n(U) \;\; \text{(direct sum)}. \quad (2\mathrm{C}.8)$$

Proposition 19: $J^n(U) = \mathfrak{A} \circ T^n(U)$.

Note, in component expression,

$$(\mathfrak{A}Z)^\mu = \Sigma_\tau \frac{1}{n!} \epsilon(\tau;\mu) Z^{\tau\mu} \;\; \text{(graded antisymmetric)}. \quad (2\mathrm{C}.9)$$

Next we define $A(U)$-linear mappings: $T^{n\overset{L}{(R)}}(U) \to A(U)$, whose set is denoted as $\mathrm{Hom}^{\overset{L}{(R)}}_{A(U)}(T^{n\overset{L}{(R)}}(U), A(U))$ with natural properties:

For $\omega \in \mathrm{Hom}^L_{A(U)}(T^{nL}(U), A(U))$, $X \in T^{nL}(U)$, we denote as $\langle X|\omega\rangle \in A(U)$; $a \in A(U)$, $\langle aX|\omega\rangle = a\langle X|\omega\rangle$ ($A(U)$-linear property); $a \in A(U)$, $\langle X|\omega\rangle a = \langle X|\omega a\rangle$. For the right version, invert all left side right, for example, $\langle\omega|X\rangle \in A(U)$.

We write $\mathrm{Hom}^{\overset{L}{(R)}}_{A(U)}(T^{n\overset{L}{(R)}}(U), A(U)) \equiv L^{n\overset{L}{(R)}}(U,A)$.

$L^{n\overset{L}{(R)}}(U,A)$ is Z_2 graded, such that $\langle X|\omega\rangle \in A(U)_k$ for $k = |\omega| + |X|$ (left version). Left and right $A(U)$-module structure is as usual, for homogeneous $a \in A(U)$, $\omega \in L^{n\overset{L}{(R)}}(U,A)$, $a\omega \equiv \omega a(-1)^{|a||\omega|}$, which coincides with $\langle X|a\omega\rangle = \langle Xa|\omega\rangle$ for the left version, $\langle\omega a|X\rangle = \langle\omega|aX\rangle$ for the right.

Let $\{X_\mu\}$ ($\mu = 1,\ldots,$n) be a basis for $\mathrm{Der}^L A(U) = T^{1L}(U)$, where always bases are taken Z_2 homogeneous. We take for a basis of $L^{1L}(U,A)$ such that

$$\langle X_\mu|\alpha^\nu\rangle = \delta_\mu{}^\nu 1_U. \quad (2\mathrm{C}.10)$$

Then $L^{1L}(U,A) \ni \omega = \alpha^\mu\omega_\mu$, $Y = Y^\mu X_\mu$, and $\langle Y|\omega\rangle = Y^\mu\omega_\mu$, (the right version, similar).

Bigraded (Z_+, Z_2) tensor algebra $L^{\overset{L}{(R)}}(U,A)$ of $L^{n\overset{L}{(R)}}(U,A)$ over $A(U)$ can be considered as $L^{\overset{L}{(R)}}(U,A) = \bigoplus_{n=0}^{\infty} L^{n\overset{L}{(R)}}(U,A)$, and $L^{n\overset{L}{(R)}}(U,A) \ni \omega = (\otimes\alpha)^\mu\omega_\mu$ where μ denotes $(\mu_1,\ldots,\mu_n)$ as before.

Now we consider $\mathrm{Hom}^{\overset{L}{(R)}}_{A(U)}(J^{n\overset{L}{(R)}}(U), A(U))$, where $J^{n\overset{L}{(R)}}(U) = \mathfrak{A}T^{n\overset{L}{(R)}}(U)$. Let $(\otimes X)_\mu \in T^{nL}(U)$, $(\otimes\alpha)^\nu \in L^{nL}(U,A)$, and $\langle(\otimes X)_\mu|(\otimes\alpha)^\nu\rangle = 1_U \Pi^n_{i=1}\delta_{\mu_i}{}^{\nu_i}(-1)^{\sum_{i<j}|\nu_i||\mu_j|}$ then, one finds

$$\langle\mathfrak{A}(\otimes X)_\mu|(\otimes\alpha)^\nu\rangle = \sum_\tau \frac{1}{n!}\epsilon(\tau;\nu)\langle(\otimes X)_\mu|(\otimes\alpha)^{\tau\nu}\rangle, \quad (2\mathrm{C}.11)$$

and thus defining similarly $\mathfrak{A}$ on $L^{n\overset{L}{(R)}}(U,A)$, and from $\mathfrak{A} \circ \mathfrak{A} = \mathfrak{A}$,

$$\langle\mathfrak{A}X|\omega\rangle = \langle X|\mathfrak{A}\omega\rangle = \langle\mathfrak{A}X|\mathfrak{A}\omega\rangle. \quad (2\mathrm{C}.12)$$

The right version is similar. Noting the dimension of $\mathrm{Hom}^{\overset{L}{(R)}}_{A(U)}(J^{n\overset{L}{(R)}}(U), A(U))$ is the same as that of $J^{n\overset{L}{(R)}}(U)$, we have

$$\mathrm{Hom}^{\overset{L}{(R)}}_{A(U)}(J^{n\overset{L}{(R)}}(U), A(U)) = \mathfrak{A}L^{n\overset{L}{(R)}}(U,A) \equiv \Omega^{n\overset{L}{(R)}}(U,A), \quad (2\mathrm{C}.13)$$

which is called a set of left (right) n-forms. Note the mapping $\mathfrak{A}$ does not change Z_2 sign, and so $\Omega^{n\overset{L}{(R)}}(U,A) = \Omega^{n\overset{L}{(R)}}(U,A)_0 \oplus \Omega^{n\overset{L}{(R)}}(U,A)_1$ where $\Omega^{n\overset{L}{(R)}}(U,A)_i = \mathfrak{A}L^{n\overset{L}{(R)}}(U,A)_i$. We put $\Omega^{0\overset{L}{(R)}}(U,A) \equiv \mathrm{A(U)}$, $\Omega^{\overset{L}{(R)}}(U,A) \equiv \bigoplus_{n=0}^{\infty}\Omega^{n\overset{L}{(R)}}(U,A)$. Both $\Omega^{\overset{L}{(R)}}(U,A)$ will have the structure of a bigraded (Z_+, Z_2) commutative algebra over $A(U)$, respectively, if we suitably define a product, the graded exterior product, on $\Omega^{\overset{L}{(R)}}(U, A)$.

Exterior product $\wedge$: Let $\Omega^{n_1\overset{L}{(R)}}(U, A) \ni \omega_1 = \mathfrak{A}\omega_1$, $\Omega^{n_2\overset{L}{(R)}}(U, A) \ni \omega_2 = \mathfrak{A}\omega_2$, and put

$$\omega_1 \wedge \omega_2 \equiv \frac{(n_1 + n_2)!}{n_1! n_2!} \mathfrak{A}(\omega_1 \otimes \omega_2). \tag{2C.14}$$

Remark 20: If $\omega_1 \in \Omega^{n_1\overset{L}{(R)}}(U, A)_{|\omega_1|}, \omega_2 \in \Omega^{n_2\overset{L}{(R)}}(U, A)_{|\omega_2|}$, then $\omega_1 \wedge \omega_2 \in \Omega^{n_1 + n_2\overset{L}{(R)}}(U, A)_{|\omega_1| + |\omega_2|}$, and $\omega_1 \wedge \omega_2 = \omega_2 \wedge \omega_1 (-1)^{|\omega_1| \cdot |\omega_2|}$ (bigraded commutative), where $|\omega_i| = (n_i, |\omega_i|) (i = 1,2)$, $|\underset{\sim}{\omega}_1| \cdot |\underset{\sim}{\omega}_2| = n_1 n_2 + |\omega_1||\omega_2|$.

The proof can be done from the definition of $\mathfrak{A}$ and the grouplike property of its signature factor. Here one should note Z_2 odd 1-forms commute, which makes the dimension of $\Omega^{\overset{L}{(R)}}(U, A)$ infinity.

Remark 21: Let $\{\alpha^\mu\}$ $(\mu = 1,...,n)$ be a basis for $L^{1\overset{L}{(R)}}(U)$, i.e., $\Omega^{1\overset{L}{(R)}}(U, A)$, then

$$(\wedge \alpha)^\mu \equiv \alpha^{\mu_1} \wedge ... \wedge \alpha^{\mu_n} = n! \mathfrak{A}(\otimes \alpha)^\mu.$$

If $\Omega^{n\overset{L}{(R)}}(U, A) \ni \omega = \mathfrak{A}\omega = \mathfrak{A}(\otimes \alpha)^\mu \omega_\mu$, then $\omega = ((\wedge \alpha)^\mu / n!)\, \omega_\mu$, where ω_μ is graded antisymmetric.

Let $\omega \in \Omega^{n^L}(U, A) = \mathrm{Hom}^{n^L}_{A(U)}(J^{n^L}(U), A(U))$, then, whose value on $\mathfrak{A}(\otimes X)^\mu \in J^{n^L}(U)$, is

$$\langle \mathfrak{A}(\otimes X)_\mu | \omega \rangle = \langle (\otimes X)_\mu | \mathfrak{A}\omega \rangle = \langle (\otimes X)_\mu | \omega \rangle, \tag{2C.15}$$

and ω is considered as a graded antisymmetric n-linear mapping from $\mathrm{Der}^L A(U)$ to $A(U)$. Thus the following notation is often used;

$$\langle (\otimes X)_\mu | \omega \rangle \equiv \langle X_{\mu_1}, X_{\mu_2}, ..., X_{\mu_n} | \omega \rangle. \tag{2C.16}$$

The right version is similar.

With the above identification, the equivalence between the definition of $\Omega^L(U, A)$ here and that of $\Omega(U, A)$ by Kostant, is, of course, seen. Further the definitions of the exterior product, here and by Kostant, are identified, also.

We define the exterior differentiation $d^{\overset{L}{(R)}}$, the interior differentiation $i_X{}^{\overset{L}{(R)}}$ and the Lie differentiation $\theta_X{}^{\overset{L}{(R)}}$, in the left and right versions, acting on $\Omega^{\overset{L}{(R)}}(U, A)$. We sum up properties and definitions.

$\Omega^{\overset{L}{(R)}}(U, A)$: bigraded with respect to (Z_+, Z_2).

$\mathrm{End}^{\overset{L}{(R)}} \Omega^{\overset{L}{(R)}}(U, A)$ is also bigraded with respect (Z_+, Z_2). Thus $u \in \mathrm{End}^{\overset{L}{(R)}} \Omega^{\overset{L}{(R)}}(U, A)$ is of bidgree $(c, |u|) \equiv |\underset{\sim}{u}|$, if $u(\Omega^{n\overset{L}{(R)}}(U, A)_i) \subset \Omega^{n + c\overset{L}{(R)}}(U, A)_{i + |u|}$ for all $(n, i) \in (Z_+, Z_2)$.

$\overset{L\text{-}}{(R\text{-})}$ Derivation $u^{\overset{L}{(R)}}$ of bidegree $|\underset{\sim}{u}^{\overset{L}{(R)}}|$: $u^{\overset{L}{(R)}} \in \mathrm{End}^{\overset{L}{(R)}} \Omega^{\overset{L}{(R)}}(U, A)$ of bidegree $|\underset{\sim}{u}|$ such that for all $\alpha, \beta \in \Omega^L(U, A)$ of bidegree $|\underset{\sim}{\alpha}|$, $|\underset{\sim}{\beta}|$, respectively,

$$u^L(\alpha \wedge \beta) = u^L \alpha \wedge \beta + (-1)^{|u^L| \cdot |\alpha|} \alpha \wedge u^L \beta, \tag{2C.17}$$

and for all $\alpha, \beta \in \Omega^R(U, A)$ of bidegree $|\underset{\sim}{\alpha}|$, $|\underset{\sim}{\beta}|$, respectively,

$$u^R(\alpha \wedge \beta) = \alpha \wedge u^R \beta + (-1)^{|u^R| \cdot |\beta|} u^R \alpha \wedge \beta. \tag{2C.18}$$

$\overset{L\text{-}}{(R\text{-})}$ derivation of $\Omega^{\overset{L}{(R)}}(U, A)$ has a left (right) $\Omega^{\overset{L}{(R)}}(U, A)$ module structure naturally with the left (right) exterior product. We denote a set of $\overset{L\text{-}}{(R\text{-})}$ derivation of $\Omega^{\overset{L}{(R)}}(U, A)$ as $\mathrm{Der}^{\overset{L}{(R)}} \Omega^{\overset{L}{(R)}}(U, A)$.

Remark 22: $\mathrm{Der}^{\overset{L}{(R)}} \Omega^{\overset{L}{(R)}}(U, A)$ is a GLA with a bigraded commutator; for homogeneous $u_1, u_2, u_3 \in \mathrm{Der}^{\overset{L}{(R)}} \Omega^{\overset{L}{(R)}}(U, A)$,

$$[u_1, u_2] \equiv u_1 \circ u_2 - (-1)^{|u_1| \cdot |u_2|} u_2 \circ u_1 \tag{2C.19}$$

is a $\overset{L\text{-}}{(R\text{-})}$ derivation of bidegree $|\underset{\sim}{u}_1| + |\underset{\sim}{u}_2|$, where $[u_1, u_2] = (-1)^{|u_1| \cdot |u_2|}[u_2, u_1]$ and

$$\sum_{\substack{\text{cyclic}\\ \text{perm}}} (-1)^{|u_1| \cdot |u_3|}[u_1, [u_2, u_3]] = 0 \tag{2C.20}$$

Proposition 23: There exists a unique $\overset{L\text{-}}{(R\text{-})}$ derivation $d^{\overset{L}{(R)}}$ of bidegree $(1, 0)$ on $\Omega^{\overset{L}{(R)}}$ such that

(i) $\langle X | d^L f \rangle \equiv Xf$, for all $X \in \mathrm{Der}^L A(U)$, $\forall f \in A(U)$. $(\langle d^R f | X \rangle \equiv X \circ f$, for all $X \in \mathrm{Der}^R A(U)$, $\forall f \in A(U)$.)

$$\text{(ii) } d^{\overset{L}{(R)}} \circ d^{\overset{L}{(R)}} = 0. \tag{2C.21}$$

The proof is done in a few steps. See Ref. 5.

Corollary 24: Let $\{\xi^\mu\}$ $(\mu = 1,...,m + n)$ be a A-CS $\{r_i; s_j : i = 1,...,m, j = 1,...,n\}$. Then $\langle \partial_\mu | d^L \xi^\nu \rangle = \partial_\mu \xi^\nu = \delta_\mu{}^\nu 1_U$, and since $\{\partial_\mu\}$ is a basis of $\mathrm{Der}^L A(U)$, $\{d\xi^\mu\}$ is a basis for $\Omega^{1L}(U, \mathrm{A})$. Thus $\forall \omega \in \Omega^L(U, A)$ is uniquely expressed as

$$\omega = \sum_{\mu, \nu} (\wedge dr)^\mu \wedge (\wedge ds)^\nu \omega_{\mu\nu}, \tag{2C.22}$$

where $(\wedge dr)^\mu \equiv dr_{\mu_1} \wedge dr_{\mu_2} \wedge ... \wedge dr_{\mu_k}$, $1 \leqslant \mu_1 < \mu_2 < ... < \mu_k \leqslant m, (\mu) \equiv k$, and $(\wedge ds)^\nu \equiv (ds_1)^{\nu_1} \wedge (ds_2)^{\nu_2} \wedge ... \wedge (ds_n)^{\nu_n}$, $\nu_j = 0, 1$, $(\nu) \equiv \Sigma_{j=1}^n \nu_j$, and $\omega_{\mu\nu} \in A(U)$. $(\wedge dr)^\mu \wedge (\wedge ds)^\nu \in \Omega^{(\mu) + (\nu)L}(U, A)$. $d\omega = (\wedge dr)^\mu \wedge (\wedge ds)^\nu (d\omega_{\mu\nu})(-1)^{(\mu) + (\nu)}$.

Hereafter we delete $\wedge$.

$\overset{L\text{-}}{(R\text{-})}$ interior differentiation, $i_X{}^{\overset{L}{(R)}}$, by $X \in \mathrm{Der}^{\overset{L}{(R)}} A(U)$: Let X be homogeneous $X \in \mathrm{Der}^{\overset{L}{(R)}} A(U)_{|X|}$, and $\omega \in \Omega^{n+1\overset{L}{(R)}}(U, A)$, then a graded antisymmetric n-linear form on $\mathrm{Der}^{\overset{L}{(R)}} A(U)$ is uniquely given by putting for all $X_i \in \mathrm{Der}^{\overset{L}{(R)}} A(U)_{|X_i|}$ $(i = 1,...,n)$,

$$\langle X_1,\ldots,X_n|i_X{}^L\omega\rangle \equiv (-1)^{|X|\sum_{i=1}^{n}|X_i|}\langle X,X_1,\ldots,X_n|\omega\rangle$$

$$\left(\langle i_X{}^R\omega|X_n,\ldots,X_1\rangle \equiv \langle\omega|X_n,\ldots,X_1,X\rangle(-1)^{|X|\sum_{i=1}^{n}|X_i|}\right), \tag{2C.23}$$

which extends for all $X_i\in\mathrm{Der}^{\overset{L}{(R)}}A(U)$ and makes

$$\langle aX_1,\ldots,X_n|i_X{}^L\omega\rangle = a\langle X_1,\ldots,X_n|i_X{}^L\omega\rangle$$

$$\left(\langle i_X{}^R\omega|X_n,\ldots,X_1a\rangle = \langle i_X{}^R\omega|X_n,\ldots,X_1\rangle a\right) \tag{2C.24}$$

valid. For all $f\in\Omega^{0\overset{L}{R}}(U,A) = A(U)$, we put $i_Xf = 0$.

Proposition 25: Let $i_X{}^{\overset{L}{(R)}}\omega$ be defined as above, where $X\in\mathrm{Der}^{\overset{L}{(R)}}A(U)_{|X|}$, $\omega\in\Omega^{n+1\overset{L}{(R)}}(U,A)$, then $i_X{}^{\overset{L}{(R)}}\omega \in\Omega^{n\overset{L}{(R)}}(U,A)$ and a $\overset{L\text{-}}{(R\text{-})}A(U)$-linear mapping $i_X{}^{\overset{L}{(R)}}$: $\Omega^{\overset{L}{(R)}}(U,A)\to\Omega^{\overset{L}{(R)}}(U,A)$ is defined, which turns out to be a $\overset{L\text{-}}{(R\text{-})}$derivation of bidegree $(-1,|X|)$.

For the proof, one uses the definition of $i_X{}^L(\alpha\beta)$ as above with the definition of the exterior product, and has $i_X{}^L(\alpha\beta) = (i_X\alpha)\beta + (-1)^{|i_X{}^L|\cdot|\alpha|}\alpha(i_X{}^L\beta)$, where $|\underset{\sim}{i_X}{}^L| = (-1,|X|)$, and the right version is similar.

We put, for all all $X\in\mathrm{Der}^{\overset{L}{(R)}}A(U)$,

$$i_X{}^{\overset{L}{(R)}} \equiv i_{X_0}{}^{\overset{L}{(R)}} + i_{X_1}{}^{\overset{L}{(R)}} \text{ where } X = X_0 + X_1. \tag{2C.25}$$

We denote $a^L\circ$ as left multiplication by $a\in A(U)$ and $a^R\circ$ as right one. Then we have, for all $a\in A(U)$ and for all $X\in\mathrm{Der}^LA(U)$,

$$i_{aX}{}^L = a^L\circ i_X{}^L \equiv a(i_X{}^L)\in\mathrm{Der}^L\Omega^L(U,A), \tag{2C.26}$$

and for $X\in\mathrm{Der}^RA(U)$

$$i_{Xa}{}^R = a^R\circ i_X{}^R \equiv (i_X{}^R)a\in\mathrm{Der}^R\Omega^R(U,A). \tag{2C.27}$$

$\overset{L\text{-}}{(R\text{-})}$ Lie differentiation, $\theta_X{}^{\overset{L}{(R)}}$, by $X\in\mathrm{Der}^{\overset{L}{(R)}}A(U)$:

$$\theta_X{}^{\overset{L}{(R)}} \equiv \left[d^{\overset{L}{(R)}}, i_X{}^{\overset{L}{(R)}}\right] = d^{\overset{L}{(R)}}\circ i_X{}^{\overset{L}{(R)}} + i_X{}^{\overset{L}{(R)}}\circ d^{\overset{L}{(R)}}, \tag{2C.28}$$

which is a derivation of bidegree $(0,|X|)$ for homogeneous $X\in\mathrm{Der}^{\overset{L}{(R)}}A(U)_{|X|}$, from Remark 22.

Forming graded commutators by two of the exterior differentiation, the interior differentiation and the Lie differentiations, we see

Remark 26: $\mathrm{Der}^{\overset{L}{(R)}}\Omega^{\overset{L}{(R)}}(U,A)\supset\overset{L\text{-}}{(R\text{-})}\Omega^{\overset{L}{(R)}}(U,A)$ submodule whose basis is given by $\{d^{\overset{L}{(R)}}, i_X{}^{\overset{L}{(R)}}, \theta_Y{}^{\overset{L}{(R)}}; X, Y\in\mathrm{Der}^{\overset{L}{(R)}}A(U)\}$ forms a graded Lie subalgebra: Here we delete $\overset{L}{(R)}$ with no mixing of $\overset{L}{(R)}$ understood.

$$[i_X,i_Y] = 0,\quad [d,d] = 0,$$

$$[i_X,d] = \theta_X,\quad [d,\theta_X] = 0,$$

$$[i_X,\theta_Y] = i_{[X,Y]},\quad [\theta_X,\theta_Y] = \theta_{[X,Y]}. \tag{2C.29}$$

Using those relation above, we have,

Proposition 27: Let $\omega\in\Omega^{n\overset{L}{(R)}}(U,A)$, homogeneous $X, X_i \in\mathrm{Der}^{\overset{L}{(R)}}A(U)$ $(i = 1,\ldots,n)$, then

$$X\langle Y_1,\cdots,Y_n|\omega\rangle = \sum_{j=1}^{n}(-1)^{|X|\sum_{i<j}|Y_i|}\langle Y_1,\ldots,[X,Y_j],\ldots,Y_n|\omega\rangle$$

$$+(-1)^{|X|\sum_{i=1}^{n}|Y_i|}\langle Y_1,\ldots,Y_n|\theta_X{}^L\omega\rangle. \tag{2C.30}$$

$$\left(\begin{aligned}&X\langle\omega|Y_n,\ldots,Y_1\rangle^{11}\\ &= \sum_{j=1}^{n}(-1)^{|X|\sum_{i<j}|Y_i|}\langle\omega|Y_n,\ldots,[X,Y_j],\ldots,Y_1\rangle + (-1)^{|X|\sum_{i=1}^{n}|Y_i|}\langle\theta_X{}^R\omega|Y_n,\ldots,Y_1\rangle\end{aligned}\right). \tag{2C.31}$$

Proposition 28: Let $\omega\in\Omega^{n\overset{L}{(R)}}(U,A)$, $X_i\in\mathrm{Der}^{\overset{L}{(R)}}A(U)_{|X_i|}$ $(i = 1,\ldots,n+1)$, then

$$\langle X_1,\ldots,X_{n+1}|d^L\omega\rangle = \sum_{j=1}^{n+1}(-1)^{j-1+|X_j|\sum_{i<j}|X_i|}X_j\langle X_1,\ldots,\hat{X}_j,\ldots,X_{n+1}|\omega\rangle$$

$$+\sum_{k<m}(-1)^{d_{k,m}}\langle[X_k,X_m],X_1,\ldots,\hat{X}_k,\ldots,\hat{X}_m,\ldots,X_{n+1}|\omega\rangle, \tag{2C.32}$$

where $d_{k,m}\equiv|X_k|(\Sigma_{i<m}|X_i|) + |X_m|(\Sigma_{i<m}|X_i|) + |X_k||X_m| + k + m$, and $\hat{X}_j$ implies X_j is missing.

$$\left(\begin{aligned}&\langle d^R\omega|X_{n+1},\ldots,X_1\rangle = \sum_{j=1}^{n+1}(-1)^{j-1+|X_j|\sum_{i<j}|X_i|}X_j\langle\omega|X_{n+1},\ldots,\hat{X}_j,\ldots,X\rangle\\ &+\sum_{k<m}(-1)^{d_{k,m}}\langle\omega|X_{n+1},\ldots\hat{X}_m,\ldots\hat{X}_k,\ldots,X_1,[X_k,X_m]\rangle\end{aligned}\right). \tag{2C.33}$$

Afterwards, we make use of (2C.32 and 2C.33) for $n = 1$ case, in the discussion of complete integrability of "differential equations" in connection of Frobenius' theorem. For that purpose we write it down explicitly; let notations be as in Proposition 28, then

$$\langle X_1,X_2|d\omega\rangle = X_1\langle X_2|\omega\rangle - (-1)^{|X_1||X_2|}X_2\langle X_1|\omega\rangle - \langle [X_1,X_2]|\omega\rangle. \tag{2C.34}$$

In this subsection C, thus far, given are essentially nothing but reviews from Kostant's paper.

Now on our graded manifold with "Hermiticity", we define a new concept "*Hermiticity*" *on* $\Omega^{\overset{L}{(R)}}(U, A)$. For that purpose, we define a *C-antilinear mapping* † and a *C*-linear mapping ^{t}.

$$^\dagger: \Omega^{n\overset{L}{(R)}}(U,A) \to \Omega^{n\overset{R}{(L)}}(U,A)$$

such that $\alpha\in\Omega^{n\overset{L}{(R)}}(U,A)$, $X_i\in\mathrm{Der}^{\overset{L}{(R)}}A(U)$ $(i=1,\dots,n)$, and we put as definition,

$$\langle\alpha^\dagger|X_n,\dots,X_1\rangle \equiv \langle X_1^\dagger,\dots,X_n^\dagger|\alpha\rangle^\dagger, \quad (\langle X_1,\dots,X_n|\alpha^\dagger\rangle \equiv \langle\alpha|X_n^\dagger,\dots,X_1^\dagger\rangle^\dagger). \tag{2C.35}$$

That $\alpha^\dagger\in\Omega^{n\overset{R}{(L)}}$ is justified by confirming, from (2C.35), the properties,

$$\langle\alpha^\dagger|X_n,\dots,X_{i+1},X_i,\dots,X_1\rangle = \langle\alpha^\dagger|X_n,\dots,X_i,X_{i+1},\dots,X_1\rangle(-1)^{|X_i||X_{i+1}|+1}, \tag{2C.36}$$

and

$$\langle\alpha^\dagger|X_n,\dots,X_1a\rangle = \langle\alpha^\dagger|X_n,\dots,X_1\rangle a, \tag{2C.37}$$

where $a\in A(U)$, and for $\alpha\in\Omega^{n^R}(U,A)$, similarly. † on $\Omega^{0\overset{L}{R}}(U,A) = A(U)$ is understood as that defined before.

We have an alternative definition † equivalent to the above.

Remark 29: $^\dagger:\Omega^{n\overset{L}{(R)}}(U,A)\to\Omega^{n\overset{R}{(L)}}(U,A)$ is defined such that for $n=0$, † is defined as † on $A(U)$; for $n=1$, †: $\Omega^{n\overset{L}{(R)}}(U,A)\to\Omega^{n\overset{R}{(L)}}(U,A)$ by

$$\langle\alpha^\dagger|X\rangle\equiv\langle X^\dagger|\alpha\rangle^\dagger, (\langle X|\alpha^\dagger\rangle\equiv\langle\alpha|X^\dagger\rangle^\dagger), \tag{2C.38}$$

where $\alpha\in\Omega^{1\overset{L}{(R)}}(U,A)$, $X\in\mathrm{Der}^{\overset{R}{(L)}}A(U)$. For all $\alpha\in\Omega^{m\overset{L}{(R)}}(U,A)$, $\forall\beta\in\Omega^{k\overset{L}{(R)}}(U,A)$, we put

$$(\alpha\beta)^\dagger = \beta^\dagger\alpha^\dagger. \tag{2C.39}$$

Then since $\Omega^{\overset{L}{(R)}}(U,A)\ni\gamma$ is expressed in the A-CS (ξ^μ) as

$$\gamma = (d\xi)^\mu(\gamma_\mu/(\mu)!), \tag{2C.40}$$

where abbreviated notation is used as in Remark 21, and $(d\xi)^\mu\in\Omega^{(\mu)\overset{L}{(R)}}(U,A)$ understood, thus we have defined $\gamma^\dagger$ for all $\gamma\in\Omega^{\overset{L}{(R)}}(U,A)$. The definition of † here is equivalent to † before.

Proof can be performed by showing $(\alpha\beta)^\dagger = \beta^\dagger\alpha^\dagger$ from the definition (2C.35) with the definition of the exterior product.

We sum up properties of † as a proposition.

Proposition 30: †: $\Omega^{\overset{L}{(R)}}(U,A)\to\Omega^{\overset{R}{(L)}}(U,A)$ gives a bigraded algebra anti-isomorphism, i.e., for all $\alpha,\beta\in\Omega^{\overset{L}{(R)}}(U,A)$,

$$(\alpha+\beta)^\dagger = \alpha^\dagger+\beta^\dagger,\ (\alpha\beta)^\dagger = \beta^\dagger\alpha^\dagger,\ (\alpha^\dagger)^\dagger = \alpha \tag{2C.41}$$

and † keeps bidegrees unchanged. Let $f\in A(U)$, then

$$(d^{\overset{L}{(R)}})^\dagger = d^{\overset{R}{(L)}}f^\dagger. \tag{2C.42}$$

Next we define a *C-linear mapping, transposition* ^{t} *on* $\Omega^{\overset{L}{R}}(U, \mathrm{A})$.

$$^t: \Omega^{n\overset{L}{(R)}}(U,A)\to\Omega^{n\overset{R}{(L)}}(U,A):$$

For $n=0$, $a\in\Omega^{0\overset{L}{R}}(U,A) = A(U)$, we put $a^t = a$, and for $n\geqslant 1$, let $\omega\in\Omega^{n\overset{L}{(R)}}(U,A)_{|\omega|}$ and $X_i\in\mathrm{Der}^{\overset{R}{(L)}}A(U)_{|X_i|}$ $(i=1,\dots,n)$, we put

$$\langle\omega^t|X_n,\dots,X_1\rangle\equiv\langle X_n^t,\dots,X_1^t|\omega\rangle(-1)^{\left(\sum_{i=1}^n|X_i|\right)(|\omega|+1)}.$$
$$(\langle X_1,\dots,X_n|\omega^t\rangle\equiv\langle\omega|X_1^t,\dots,X_n^t\rangle(-1)^{\left(\sum_{i=1}^n|X_i|\right)(|\omega|+1)}). \tag{2C.43}$$

That $\omega^t\in\Omega^{n\overset{R}{(L)}}(U,A)$ is justified by confirming, from (2C.43), the properties.

$$\langle\omega^t|X_n,\dots,X_{i+1},X_i,\dots,X_1\rangle = \langle\omega^t|X_n,\dots,X_i,X_{i+1},\dots,X_1\rangle(-1)^{|X_i||X_{i+1}|+1} \tag{2C.44}$$

and

$$\langle\omega^t|X_n,\dots X_1a\rangle = \langle\omega^t|X_n,\dots,X_1\rangle a, \tag{2C.45}$$

where $a\in A(U)$, and for $\omega\in\Omega^{n^R}(U,A)$, similarly.

Proposition 31: ^{t}: $\Omega^{\overset{L}{(R)}}(U,A)\to\Omega^{\overset{R}{(L)}}(U,A)$ gives a bigraded algebra isomorphism, i.e., for all $\alpha,\beta\in\Omega^{\overset{L}{(R)}}(U,A)$,

$$(\alpha+\beta)^t = \alpha^t+\beta^t, (\alpha\beta)^t = \alpha^t\beta^t, (\alpha^t)^t = \alpha \tag{2C.46}$$

and ^{t} keeps bidegrees unchanged. Let $f\in A(U)$, then

$$f^t = f, \text{ and } (d^{\overset{L}{(R)}}f)^t = d^{\overset{R}{(L)}}f \tag{2C.47}$$

Proof can be performed by using the definition (2C.43), firstly showing $(a\omega)^t = a\omega^t$, and secondly using the definition of the exterior product and showing $(\alpha\beta)^t = \alpha^t\beta^t$ for α, $\beta\in\Omega^{\overset{L}{(R)}}(U,A)$.

Similarly as in case of †, we can give an alternative definition of ^{t}, utilizing the above properties.

Remark 32: † and ′ on $\overset{L}{\Omega}{}^{(R)}(U,A)$ commute, i.e., for $\alpha\in\overset{L}{\Omega}{}^{(R)}(U,A)$,

$$(\alpha^\dagger)' = (\alpha')^\dagger. \qquad (2C.48)$$

For proof, we make use of Remark 14, and the definitions (2C.35) and (2C.43).

Proposition 33: $'\circ^\dagger$: $\overset{L}{\Omega}{}^{(R)}(U,A)\to\overset{L}{\Omega}{}^{(R)}(U,A)$ gives a bigraded algebra antiautomorphism which is C-antilinear and bidegree preserving, i.e., for all $\alpha,\beta\in\overset{L}{\Omega}{}^{(R)}(U,A)$,

$$(\alpha+\beta)^{\dagger\prime}=\alpha^{\dagger\prime}+\beta^{\dagger\prime},(\alpha\beta)^{\dagger\prime}=\beta^{\dagger\prime}\alpha^{\dagger\prime},(\alpha^{\dagger\prime})^{\dagger\prime}=\alpha \qquad (2C.49)$$

For all $f\in\overset{0\,L}{\Omega}{}^{(R)}(U,A)$, $f^{\dagger\prime}=f^\dagger$ and $(\overset{L}{d}{}^{(R)}f)^{\dagger\prime}=\overset{L}{d}{}^{(R)}f^\dagger$. (2C.50)

Now we reach a concept, "Hermiticity" on $\overset{L}{\Omega}{}^{R}(U,\mathrm{A})$, which is our main concern of this subsection.

"Hermiticity" on $\overset{L}{\Omega}{}^{(R)}(U,A)$: We call $\omega\in\overset{L}{\Omega}{}^{(R)}(U,A)$ is "Hermite" if $\omega^{\dagger\prime}=\omega$, and "anti-Hermite" if $\omega^{\dagger\prime}=-\omega$.

Remark 34: $\overset{L}{\Omega}{}^{(R)}(U,A)\ni\omega$ is uniquely decomposed into "Hermitian" and "anti-Hermitian" parts, and

$$\overset{n\,L}{\Omega}{}^{R}(U,A)_k=\overset{n\,L}{\Omega}{}^{(R)}(U,A)^h_k\oplus i\overset{n\,L}{\Omega}{}^{(R)}(U,A)^h_k \quad (k=0,1), \qquad (2C.51)$$

where $\overset{n\,L}{\Omega}{}^{(R)}(U,A)^h_k$ denotes a set of "Hermitian" forms of bidegree (n,k).

The proof is as usual.

Further we will define *"Hermiticity" on* $\overset{L}{\mathrm{Der}}{}^{(R)}\overset{L}{\Omega}{}^{(R)}(U,A)$ in the same way as we did on $\overset{L}{\mathrm{Der}}{}^{(R)}A(U)$.

"Hermitian" conjugation †: $\overset{L}{\mathrm{Der}}{}^{(R)}\overset{L}{\Omega}{}^{(R)}(U,A)\to\overset{R}{\mathrm{Der}}{}^{(L)}\overset{R}{\Omega}{}^{(L)}(U,A)$: Let $u\in\overset{L}{\mathrm{Der}}{}^{(R)}\overset{L}{\Omega}{}^{(R)}(U,A)$ and $\alpha\in\overset{R}{\Omega}{}^{(L)}(U,A)$, and we put as definition,

$$u\alpha^\dagger\equiv(u\alpha^\dagger)^\dagger. \qquad (2C.52)$$

That $u^\dagger\in\overset{R}{\mathrm{Der}}{}^{(L)}\overset{R}{\mathrm{Der}}{}^{(L)}(A,U)$ is justified by confirming $u^\dagger$ defined as (2C.52) satisfies (2C.18) [or (2C.17)].

Remark 35: Let $u\in\overset{L}{\mathrm{Der}}{}^{(R)}\overset{L}{\Omega}{}^{(R)}(U,A)$, $\alpha\in\overset{L}{\Omega}{}^{(R)}(U,A)$ and $\overset{L}{\alpha}{}^{(R)}\circ$ denote left (right) multiplication (exterior product) then $\overset{L}{\alpha}{}^{(R)}\circ u\in\overset{L}{\mathrm{Der}}{}^{(R)}\overset{L}{\Omega}{}^{(R)}(U,A)$ [$\binom{L\text{-}}{(R\text{-})}\overset{L}{\Omega}{}^{(R)}(U,A)$-module property of $\overset{L}{\mathrm{Der}}{}^{(R)}\overset{L}{\Omega}{}^{(R)}(U,A)$].

$$(\overset{L}{\alpha}{}^{(R)}\circ u)^\dagger=(\alpha^\dagger)\overset{R}{{}^{(L)}}\circ u^\dagger\in\overset{R}{\mathrm{Der}}{}^{(L)}\overset{R}{\Omega}{}^{(L)}(U,A). \qquad (2C.53)$$

Let $u_1,u_2\in\overset{L}{\mathrm{Der}}{}^{(R)}\overset{L}{\Omega}{}^{(R)}(U,A)$, then

$$(u_1+u_2)^\dagger=u_1{}^\dagger+u_2{}^\dagger,\ [u_1,u_2]^\dagger=[u_1{}^\dagger,u_2{}^\dagger],(u^\dagger)^\dagger=u, \qquad (2C.54)$$

and † keeps bidegree unchanged.

Remark 36: (i) $(\overset{L}{d}{}^{(R)})^\dagger=\overset{R}{d}{}^{(L)}$, (ii) $(\overset{L}{i}_X{}^{(R)})^\dagger=\overset{R}{i}_{X^\dagger}{}^{(L)}$, (iii) $(\overset{L}{\theta}_X{}^{(R)})^\dagger=\overset{R}{\theta}_X{}^{(L)}$, where $X\in\overset{L}{\mathrm{Der}}{}^{(R)}A(U)$.

We proceed to transposition ′ on $\overset{L}{\mathrm{Der}}{}^{(R)}\overset{L}{\Omega}{}^{(R)}(U,A)$.

Transposition ′: $\overset{L}{\mathrm{Der}}{}^{(R)}\overset{R}{\Omega}{}^{(L)}(U,A)\to\overset{R}{\mathrm{Der}}{}^{(L)}\overset{R}{\Omega}{}^{(L)}(U,A)$:

Let homogeneous $u\in\overset{L}{\mathrm{Der}}{}^{(R)}\overset{L}{\Omega}{}^{(R)}(U,A)$ of bidegree $|\underline{u}|$ and $\alpha\in\overset{m\,R}{\Omega}{}^{(L)}(U,A)_{|\alpha|}$ of bidegree $|\underline{\alpha}|=(m,|\alpha|)$, then we put a C-linear mapping ′ such that

$$u'\alpha\equiv(u\alpha')'(-1)^{|\underline{u}|\cdot(|\underline{\alpha}|+\underline{1})}, \qquad (2C.55)$$

where $\underline{1}=(1,1)$, which extends for $\alpha\in\overset{R}{\Omega}{}^{(L)}(U,A)$. That $u'\in\overset{R}{\mathrm{Der}}{}^{(L)}\overset{R}{\Omega}{}^{(L)}(U,A)$ is justified by confirming (2C.18) [or (2C.17)].

Remark 37: Let notation be as in Remark 35. Then it holds

$$(\overset{L}{\alpha}{}^{(R)}\circ u)'=(-1)^{|\underline{\alpha}|\cdot(|\underline{u}|+\underline{1})}\alpha'\overset{R}{{}^{(L)}}\circ u'\in\overset{R}{\mathrm{Der}}{}^{(L)}\overset{R}{\Omega}{}^{(L)}(U,A), \qquad (2C.56)$$

and ′ keeps bidegree unchanged. Let $v_1,v_2\in\overset{L}{\mathrm{Der}}{}^{(R)}\overset{L}{\Omega}{}^{(R)}(U,A)$ be homogeneous. $(u_1+u_2)'=u_1{}'+u_2{}',(u')'=u$, $[v_1,v_2]'=(-1)^{|\underline{v}_1||\underline{v}_2|}[v_1{}',v_2{}']$. (2C.57)

Note (2C.56) is a generalization of (2B.5), which is included in (2C.56) by putting $u=\theta_X$ with restriction to $\Omega^0(U,A)$, and putting $\alpha=a\in\Omega^0(U,A)$.

Remark 38:

(i) $(\overset{L}{d}{}^{(R)})'=-\overset{L}{d}{}^{(R)}$, (ii) $(\overset{L}{i}_X{}^{(R)})'=\overset{L}{i}_{X'}{}^{(L)}$, (iii) $(\overset{L}{\theta}_X{}^{(R)})'=\overset{R}{\theta}_{X'}{}^{(L)}$, where $X\in\overset{L}{\mathrm{Der}}{}^{(R)}A(U)$.

For proofs of (i), (ii) of Remarks 36 and 38, we first note bidegrees unchanged, and so it is sufficient to check the effect on $\overset{0\,L}{\Omega}{}^{(R)}(U,A)$ and $\overset{1\,L}{\Omega}{}^{(R)}(U,A)$. As for $\overset{L}{\theta}_X{}^{(R)}$ we use its definition with (2C.54) or (2C.57)

Remark 39: † and ′ on $\overset{L}{\mathrm{Der}}{}^{(R)}\overset{L}{\Omega}{}^{(R)}(U,A)$ commute, i.e., for † $u\in\overset{L}{\mathrm{Der}}{}^{(R)}\overset{L}{\Omega}{}^{(R)}(U,A)$,

$$u^{\dagger\prime}=u^{\prime\dagger}. \qquad (2C.58)$$

Proposition 40: $'\circ^\dagger$: $\overset{L}{\mathrm{Der}}{}^{(R)}\overset{L}{\Omega}{}^{(R)}(U,A)\to\overset{L}{\mathrm{Der}}{}^{(R)}\overset{L}{\Omega}{}^{(R)}(U,A)$

gives a bidegree preserving C-antilinear isomorphism such that for $u \in \mathrm{Der}^{\overset{L}{(R)}}\Omega^{\overset{L}{(R)}}(U, A)$, and homogeneous v, $w \in \mathrm{Der}^{\overset{L}{(R)}}\Omega^{\overset{L}{(R)}}(U, \mathrm{A})$,

$$(u^{\dagger\prime})^{\dagger\prime} = u \quad \text{and} \quad [v,w]^{\dagger\prime} = (-1)^{|v|\cdot|w|}[v^{\dagger\prime}, w^{\dagger\prime}]. \tag{2C.59}$$

Now we define "Hermiticity" on $\mathrm{Der}^{\overset{L}{(R)}}\Omega^{\overset{L}{(R)}}(U, A)$.

"Hermiticity" on $\mathrm{Der}^{\overset{L}{(R)}}\Omega^{\overset{L}{(R)}}(U, A)$: We call $u \in \mathrm{Der}^{\overset{L}{(R)}}\Omega^{\overset{L}{(R)}}(U, A)$ is "Hermite" if $u^{\dagger\prime} = u$, and "anti-Hermite" if $u^{\dagger\prime} = -u$.

Remark 41: $\mathrm{Der}^{\overset{L}{(R)}}\Omega^{\overset{L}{(R)}}(U, A) \ni u$ is uniquely decomposed into "Hermitian" and "anti-Hermitian" parts, and

$$\mathrm{Der}^{\overset{L}{(R)}}\Omega^{\overset{L}{(R)}}(U, A)_{(m,k)} = (\mathrm{Der}^{\overset{L}{(R)}}\Omega^{\overset{L}{(R)}}(U, A))^h_{(m,k)} \oplus i(\mathrm{Der}^{\overset{L}{(R)}}\Omega^{\overset{L}{(R)}}(U, A))^h_{(m,k)}, \tag{2C.60}$$

where (m,k) denotes the bidegree with integer m and $k = 0$, 1, and h denotes a "Hermitian" space.

Proof goes as usual.

We combine the definitions and write down the properties:

Remark 42: For homogeneous $u \in \mathrm{Der}^{\overset{L}{(R)}}\Omega^{\overset{L}{(R)}}(U, A)$ and $\alpha \in \Omega^{\overset{L}{(R)}}(U, A)$,

$$(u\alpha)^{\dagger\prime} = (-1)^{|u|\cdot(|\alpha|+1)}u^{\dagger\prime}\alpha^{\dagger\prime}. \tag{2C.61}$$

(i) $d^{\overset{L}{(R)}\dagger\prime} = -d^{\overset{L}{(R)}}$, (ii) $(i_X{}^{\overset{L}{(R)}})^{\dagger\prime} = (i_{X^{\dagger\prime}}{}^{\overset{L}{(R)}})$, (iii) $(\theta_X{}^{\overset{L}{(R)}})^{\dagger\prime} = \theta_X{}^{\dagger\prime\overset{L}{(R)}}$,

where $X \in \mathrm{Der}^{\overset{L}{(R)}}A(U)$.

If we note $i_{X+Y}{}^{\overset{L}{(R)}} = i_X{}^{\overset{L}{(R)}} + i_Y{}^{\overset{L}{(R)}}$ by definition, we have the following remark.

Remark 43: Decomposition into "Hermitian" and "anti-Hermitian" parts is as follows:

$$i_X = i_{X^h} + i_{X^a}, \quad \theta_X = \theta_{X^h} + \theta_{X^a}, \tag{2C.62}$$

where $\mathrm{Der}A(U) \ni X = X^h + X^a$, and X^h and X^a are "Hermite" and "anti-Hermite", respectively. In this remark we delete all $^{\overset{L}{(R)}}$, understanding no mixing of $^{\overset{L}{(R)}}$.

It is useful to write down the following as a remark from a practical point of view.

Remark 44: If $\alpha \in \Omega^n(U, A)$ is "Hermite" i.e., $\alpha^{\dagger\prime} = \alpha$, then

$$(d\alpha)^{\dagger\prime} = (-1)^n d\alpha, \tag{2C.63}$$

where no mixing of $^{\overset{L}{(R)}}$ understood, we have deleted $^{\overset{L}{(R)}}$.

It is easy from Remark 42.

D. Poincaré's lemma and Frobenius' theorem

Before going into discussions on Poincaré's lemma and Frobenius' theorem, let us clarify a concept of "linear independence" of 1-forms, $\{\alpha^i\}$ $(i = 1,\ldots,k+p)$ and of derivations, $\{X_i\}$ $(i = 1,\ldots,k+p)$, and also let us see "Hermiticity" condition of them in matrix form.

Let the total dimension of the graded manifold (X, A) be (m, n) where even dimension m, odd n, and U be an A-coordinate neighborhood and $(\xi^\mu; \mu = 1,\ldots,m+n)$ be a A-CS, where $\{\xi^\mu\}$ $(\mu = 1,\ldots,m)$ even coordinates and $\{\xi^\mu\}$ $(\mu = m+1,\ldots,m+n)$ odd ones. Suppose $\alpha^i \in \Omega^{1\overset{L}{(R)}}(U, \mathrm{A})_0$ for $i = 1,\ldots,k$, $\alpha^i \in \Omega^{1\overset{L}{(R)}}(U, \mathrm{A})_1$ for $i = k+1,\ldots,k+p$ and $X_i \in \mathrm{Der}^{\overset{L}{(R)}}A(U)_0$ for $i = 1,\ldots,k$, $X_i \in \mathrm{Der}^{\overset{L}{(R)}}A(U)_1$ for $i = k+1,\ldots,k+p$. In the matrix notation, we always write $(\alpha^i; i = 1,\ldots k+p)$ in a row vector (α) for $\alpha^i \in \Omega^{1L}(U, \mathrm{A})$ and in a column vector (α) for $\alpha^i \in \Omega^{1R}(U, A)$, and $(X_i;$ $i = 1,\ldots,k+p)$ in a column vector (X) for $X_i \in \mathrm{Der}^L A(U)$ and in a row vector (X) for $X_i \in \mathrm{Der}^R A(U)$. We can express, in the left version, unambiguously,

$$(\alpha)_i = (d\xi)_j M_{ji}, \quad (X)_i = N_{ij}(\partial)_j, \tag{2D.1}$$

where

$$M = \begin{pmatrix} M^{++} & M^{+-} \\ M^{-+} & M^{--} \end{pmatrix}, \quad N = \begin{pmatrix} N^{++} & N^{+-} \\ N^{-+} & N^{--} \end{pmatrix} \tag{2D.2}$$

and M^{++}; $(m\times k)$, M^{+-}; $(m\times p)$, M^{-+}; $(n\times k)$, M^{--}; $(n\times p)$, and N^{++}; $(k\times m)$, N^{+-}; $(k\times n)$, N^{-+}; $(p\times m)$, N^{--}; $(p\times n)$, and all elements of $M^{+-}, M^{-+}, N^{+-}, N^{-+}$ are Z_2 odd, while others are Z_2 even. We call such matrices like M, as $(m+n)\times(k+p)$ matrices and such like N as $(k+p)\times(m+n)$. In the right version,

$$(\alpha) = M(d\xi), \ (X) = (\partial)N. \tag{2D.3}$$

We call (α) and (X) are linearly independent if rank $\tilde{M} = k+p$ and if rank $\tilde{N} = k+p$, everywhere in U, respectively, where $\tilde{M}_{ij} = (\tilde{M}_{ij})$, $\tilde{N}_{ij} = (\tilde{N}_{ij})$ and note $\tilde{M}^{+-} = \tilde{M}^{-+} = 0$ and $\tilde{N}^{+-} = \tilde{N}^{-+} = 0$. The concept of linear independence is independent of A-CSs used.

We define for $\alpha^i \in \Omega^{1L}(U, A)$,

$$(\alpha)^T \equiv (\alpha'), \tag{2D.4}$$

where $\alpha'^i \in \Omega^{1R}(U, A)$ and thus (α') is a column vector, then for $\alpha^i \in \Omega^{1R}(U, A)$,

$$(\alpha)^{T^{-1}} = (\alpha'). \tag{2D.5}$$

Similarly, for $X_i \in \mathrm{Der}^L A(U)$,

$$(X)^T \equiv (X'), \tag{2D.6}$$

and for $X_i \in \mathrm{Der}^R A(U)$,

$$(X)^{T^{-1}} \equiv (X'). \tag{2D.7}$$

For a $(m'+n')\times(k'+p')$ matrix M, where $M_{ij} \in A(U)$, we define transposition^T: $M \rightarrow M^T$,

$$M^T=\begin{pmatrix} M^{++t} & -M^{-+t} \\ M^{+-t} & M^{--t}\end{pmatrix}, \tag{2D.8}$$

where t denotes the ordinary transposition, and of course

$$M^{T^{-1}}=\begin{pmatrix} M^{++t} & M^{-+t} \\ -M^{+-t} & M^{--t}\end{pmatrix}. \tag{2D.9}$$

We also define for $\alpha^i \in \overset{\text{L}}{\Omega}{}^{1(\text{R})}(U, A)$.

$$(\alpha)^\dagger \equiv (\alpha^\dagger), \tag{2D.10}$$

for $X_i \in \overset{\text{L}}{\text{Der}}{}^{(\text{R})}(U, A)$

$$(X)^\dagger \equiv (X^\dagger), \tag{2D.11}$$

where column or row vectors should be understood properly, and for matrices M,

$$(M^\dagger)_{ij} \equiv (M^t)_{ij}{}^\dagger. \tag{2D.12}$$

We consider matrix multiplication MN, for a while, only for such matrices, e.g., M; $(k'+p')\times(m'+n')$, N; $(m'+n')\times(q'+r')$, and then MN; $(k'+p')\times(q'+r')$.

Remark 45:

$$(MN)^T=N^TM^T,\quad (MN)^\dagger=N^\dagger M^\dagger,$$
$$(MN)^{T\dagger}=M^{T\dagger}N^{T\dagger},\ M^{T\dagger}=M^{\dagger T^{-1}}$$
$$\text{and } (M^{T\dagger})^{T\dagger}=M, \tag{2D.13}$$

where T may be replaced with T^{-1}.

Remark 46: Let notations be as in (2D.1) [or (2D.3)], i.e., $\alpha^1 \in \overset{\text{L}}{\Omega}{}^{1(\text{R})}(U, \text{A})$, $X_i \in \overset{\text{L}}{\text{Der}}{}^{(\text{R})}(U, \text{A})$, then

$$(\alpha)^T=M^T(d\xi)^T,\ ((\alpha)^{T^{-1}}=(d\xi)^{T^{-1}}M^{T^{-1}}), \tag{2D.14}$$

$$(X)^T=(\partial)^TN^T,\ ((X)^{T^{-1}}=N^{T^{-1}}(\partial)^{T^{-1}}), \tag{2D.15}$$

$$(\alpha)^{T\dagger}=(d\xi)^{T\dagger}M^{T\dagger},\ ((\alpha)^{T^{-1}}=M^{T^{-1}\dagger}(d\xi)^{T^{-1}\dagger}), \tag{2D.16}$$

$$(X)^{T\dagger}=N^{T\dagger}(\partial)^{T\dagger},\ ((X)^{T^{-1}\dagger}=(\partial)^{T^{-1}\dagger}N^{T^{-1}\dagger}). \tag{2D.17}$$

For the proof, use Proposition 31, (2B.7), Proposition 30 and the component expression of $X^\dagger$.

Hereafter we work only in the left version, though translation to the right version is trivial.

We call those matrices like M and N in (2D.1) as *component matrices* (CM) of (α) and of (X), in the A-CS (ξ^μ), and denote a set of $(m'+n')\times(k'+p')$ matrices as $\mathscr{M}(m'+n')\times(k'+p')$, and we call M is a "*Hermitian*" *component matrix* (HCM) if $M\in\mathscr{M}$ and $M^{T\dagger}=M$, an "*anti-Hermitian*" *component matrix* (AHCM) if $M\in\mathscr{M}$ and $M^{T\dagger}=-M$. We define a set of HCM$\in\mathscr{M}(m'+n')\times(k'+p')$ as $\mathscr{H}(m'+n')\times(k'+p')$.

Remark 47: Let $(\alpha)=(d\xi)M$, $(X)=N(\partial)$, in HCS (ξ^μ). Then all α^i are "Hermite" if and only if $M\in\mathscr{H}$, and all X_i are "Hermite" if and only if $N\in\mathscr{H}$.

Remark 48: Let $M\in\mathscr{H}(m'+n')\times(k'+p')$, $N\in\mathscr{H}(k'+p')\times(r'+s')$, then $MN\in\mathscr{H}(m'+n')\times(r'+s')$.

Remark 49: Let $\{\alpha^i\}=(i=1,\ldots,k+p;\ k\leqslant m,\ p\leqslant n)$ be linearly independent homogeneous 1-forms, i.e., $(\alpha)=(d\xi)M$, where $M\in\mathscr{M}(m+n)\times(k+p)$ and rank $\tilde{M}=k+p$. Then $\{\alpha^i\}$ can be extended to linearly independent homogeneous 1-forms $\{\alpha^\mu\}$ $(\mu=1,\ldots,m+n)$ such that $(\alpha)=(d\xi)M'$, $M'=\in\mathscr{M}(m+n)\times(m+n)$, rank $\tilde{M}'=m+n$. We could have all extended parts be "Hermite". As for $\{X_i\}$, the similar statement is valid.

Remark 50: Let $\{df^i\}$ $(i=1,\ldots,k+p;\ k\leqslant m,\ p\leqslant n)$ be linearly independent homogeneous 1-forms, then $\{df^i\}$ can be extended to linearly independent homogeneous 1-forms $\{df^\mu\}$ $(\mu=1,\ldots,m+n)$. If all $\{\tilde{f}^i\}\in C^\infty(U)$, then $\exists A$-CS $(f^\mu)\supset\{f^i\}$. We could have all extended parts be "Hermite".

For the proof, we note the extended part can be taken from the A-CS used.

Remark 51: Let $\{\alpha^\mu\}$ $(\mu=1,\ldots,m+n)$ be linearly independent homogeneous 1-forms, i.e., $(\alpha)=(d\xi)M$, $\mathscr{M}(m+n)\times(m+n)\ni M$ is nonsingular, $\exists M^{-1}$. Then there exist uniquely $\{X_\mu\}$ $(\mu=1,\ldots,m+n)$ such that

$$\langle X_\mu|\alpha^\mu\rangle=\delta_\mu{}^\nu 1_U. \tag{2D.18}$$

Further if $\{\alpha^\mu\}$ are all "Hermite", then $\{X_\mu\}$ are all "Hermite". Interchanging the roles of $\{\alpha^\mu\}$ and $\{X_\mu\}$, the similar statement is valid.

For the proof, one sees that $N=M^{-1}$, where $(X)=N(\partial)$. If $M\in\mathscr{H}(m+n)\times(m+n)$, then $M^{-1}\in\mathscr{H}(m+n)\times(m+n)$, which can be proved by putting $M=M_0+M_1$ where M_0 is Z_2 even part and M_1 odd, then $M^{-1}=(\Sigma_{k=0}^n(-M_0{}^{-1}M_1)^k)M_0{}^{-1}\in\mathscr{H}(m+n)\times(m+n)$ since $M_0{}^{-1}$, $M_0{}^{-1}M_1\in\mathscr{H}(m+n)\times(m+n)$. Thus one notes the nonsingular subset of $\mathscr{H}(m+n)\times(m+n)$ forms a group.

Let $\Omega(U,A)\ni\omega=\omega^h+\omega^a$ where ω^h, "Hermite" and ω^a, "anti-Hermite". If $d\omega=0$, then $d\omega^h=d\omega^a=0$ from Remark 44. Thus, for Poincaré's lemma, we may consider ω is "Hermite", without any loss of generality.

Poincaré's lemma: Let U be a contractible neighborhood, and "Hermitian" $\omega\in\Omega^{k+1}(U,A)$ $(k\in Z_+)$. If ω is closed, i.e., $d\omega=0$, then there exists a "Hermitian" ("anti-Hermitian") $\theta\in\Omega^k(U,A)$ when $k=$ even ($k=$ odd), such that $\omega=\mathrm{d}\theta$.

Aside from Kostant's algebraic proof, we will sketch an analytical proof, the cylinder construction, is possible as in the ordinary manifold cases, which are found in many textbooks.[12]

Proof: Let us imbed (U, A) of dim (m, n), as a graded submanifold, into a graded manifold $(I\times U, B)$ of dim $(m+1, n)$, where $I=R$, such that

$$\begin{pmatrix} j_0\colon & (U,A)\to & (I\times U,B)\\ j_1\colon & (U,A)\to & (I\times U,B)\end{pmatrix} \text{ by } \begin{pmatrix} j_0{}^*\colon & B(I\times U)\to & A(U)\\ j_1{}^*\colon & B(I\times U)\to & A(U)\end{pmatrix} \tag{2D.19}$$

Let $(t, \xi'^{\mu}; \mu = 1,...,m+n) = (t, r'^i, s'^j; i = 1,...,m, j = 1,...,n)$ be a HCS for $(I \times U, B)$, $t \in B(I \times U)_0$, and $(\xi^{\mu}; \mu = 1,...,m+n) = (r^i, s^j; i = 1,...m, j = 1,...,n)$ be a HCS for (U, A) such that

$$j_i^*(\xi'^{\mu}) = \xi^{\mu} \ (i = 0,1), j_1^*(t) = 1_U, j_0^*(t) = 0_U, \quad (2D.20)$$

where $j_i^*(i = 0,1)$ is an algebra homomorphism compatible with ${}^{\dagger\prime}$ conjugation, is assumed. $(\widetilde{t, r'^i})$ is a LCS for $I \times U$, and $(\widetilde{r^i})$ for U, and j_i^* $(i = 0,1)$ induces an algebra homomorphism $\tilde{j}_i^*$: $\mathscr{C}^{\infty}(I \times U) \to \mathscr{C}^{\infty}(U)$ such that

$$\tilde{j}_i^*(\widetilde{r'^{\mu}}) = \widetilde{r^{\mu}} \ (i = 0,1), \quad \tilde{j}_1^*(\tilde{t}) = 1, \ \tilde{j}_0^*(t) = 0. \quad (2D.21)$$

An algebra homomorphism compatible with d, j_i^*: $\Omega^p(I \times U, B) \to \Omega^p(U, A)$, $p \in Z_+$ is induced for $i = 0,1$.

Integration in terms of an even coordinate, say t, is well defined as follows: If $\partial_t F(t,\xi') = A(t,\xi')$ holds for given $A(t,\xi') \in B(I \times U)$, then $F(t,\xi')$ is well defined up to $C(\xi') \in B(I \times U)$. Thus $j_1^*F(t,\xi') - j_0^*F(t,\xi') = F(1,\xi) - F(0,\xi) \in A(U)$ is unambiguously defined, and we put symbolically

$$\int_0^1 dt A(t,\xi) \equiv j_1^*F(t,\xi') - j_0^*F(t,\xi'). \quad (2D.22)$$

Note the above integration preserves Z_2 grading.

Thus we can define a C-linear mapping,

$$K: \Omega^{p+1}(I \times U, B) \to \Omega^p(U, A) \ (p \in Z_+), \quad (2D.23)$$

such that

$$K \circ (d\xi')^{\mu} A_{\mu}(t,\xi') = 0_U \quad (2D.24)$$

and

$$K \circ dt A_{\nu}(t,\xi')(d\xi')^{\nu} = \left(\int_0^1 dt A_{\nu}(t,\xi')\right)(d\xi)^{\nu}, \quad (2D.25)$$

where note $|K| = 0$, and if $\alpha \in \Omega^p(I \times U, B)$ is "Hermite," then

$$(K \circ \alpha)^{\dagger\prime} = (-1)^p (K \circ \alpha). \quad (2D.26)$$

The basic property of K is valid: For all $\alpha \in \Omega(I \times U, B)$,

$$K \circ d\alpha + d(K \circ \alpha) = (j_1^* - j_0^*)\alpha. \quad (2D.27)$$

U is contractible: $\exists \phi'^i(t,\tilde{r}) \in C^{\infty}(I \times U)$ such that $\phi'^i(1,\tilde{r}) = \tilde{r}^i$, $\phi'^i(0,\tilde{r}) = \tilde{r}_0{}^i$ $(\tilde{r}_0{}^i \in R)$.

Let us define a mapping ϕ: $(I \times U, B) \to (U, A)$ by ϕ^*: $A(U) \to B(I \times U)$, such that

$$\phi^*(r^i) = \phi'^i(t, r'), \ \phi^*(s^j) = ts'^j. \quad (2D.28)$$

then $j_i^*\phi^*: A(U) \to A(U)$ $(i = 0,1)$: $j_1^*\phi^*$ is an identity mapping and $j_0^*\phi^*(r^i) = \tilde{r}_0{}^i 1_U$, $j_0^*\phi^*(s^j) = 0_U$, and then $j_1^*\phi^*: \Omega^{k+1}(U, A) \to \Omega^{k+1}(U, A)$ $(k \in Z_+)$:

$$j_1^*\phi^*\omega = \omega, \ j_0\phi^*\omega = 0 \ \text{for all} \ \omega \in \Omega^{k+1}(U, A). \quad (2D.29)$$

Put $\alpha = \phi^*\omega$ in (2D.27), then from $d\omega = 0$ and (2D.29),

$$\omega = d[K \circ (\phi^*\omega)], \quad (2D.30)$$

where $K \circ (\phi^*\omega) \equiv \theta$, the θ has suitable "Hermiticity" stated in the lemma, which follows from (2D.26), Remark 44 and that ϕ^* preserves "Hermiticity."Q.E.D.

Next we discuss Frobenius' theorem on our graded manifold with "Hermiticity", which is, as far as we know, new in incorporation of "Hermiticity".

We discuss Frobenius' theorem in such a way that it leads us to existence of 0-forms which can be extended to A-CS. Then some condition is needed on Z_2 even forms. Thus we prepare a concept, *"real" forms*: Let $\Omega(U)$ be an ordinary (real) exterior algebra on the real manifold U, and $\Omega_C(U)$ be complexification of $\Omega(U)$. Since $\sim$ is a homomorphism $\tilde{}: A(U) \to \mathscr{C}^{\infty}(U)$, $\tilde{}$ induces a homomorphism: $\Omega(U, A) \to \Omega_C(U)$, which we also denote $\tilde{}: \Omega(U, A) \to \Omega_C(U)$, where note $\tilde{s} = 0$, $\widetilde{ds} = 0$ for an odd $s \in A(U)$.[13] We call $\omega \in \Omega(U, A)$ is a "real" form if $\tilde{\omega} \in \Omega(U)$. All Z_2 odd elements $\in \Omega(U, A)$ are "real".

Remark 52: If Z_2 even "real" 1-forms $\{\omega^i\} \in \Omega^1(U, A)_0$ are linearly independent, then $\{\widetilde{\omega^i}\} \in \Omega(U)$ are also linearly independent.

If $\{\omega^i\}$ are "Hermitian" 1-forms, then they are "real".

Frobenius' Theorem: Let U be a sufficiently small contractible neighborhood of $P \in U$, and $\{\omega^i; i = 1,\cdots,k+p\} \in \Omega^1(U, A)$ be linearly independent homogeneous "real" 1-forms, and $\{\omega^i; i = 1,\cdots,k\}$ be Z_2 even, and $\{\omega^{i+k}; i = 1,\cdots,p\}$ be Z_2 odd. Then

$$d\omega^i = \sum_{j=1}^{k+p} \omega^j \theta_j{}^i \quad (i = 1,\cdots.k+p), \quad (2D.31)$$

where $\theta_j{}^i \in \Omega^1(U, A)$, if and only if

$$\omega^i = \sum_{j=1}^{k+p} (df^j) M_j{}^i \quad (i = 1,\cdots,k+p), \quad (2D.32)$$

where $\{df^i; i = 1,\cdots,k+p\}$ are linearly independent homogeneous "real" 1-forms, and in the notation as (2D.1), $M_{ji} = M_j{}^i$, M is a nonsingular $(k+p)\times(k+p)$ real matrix. In the above statement, further, if $\{\omega^i; i = 1,\cdots,k+p\}$ are "Hermitian" 1-forms, then in (2D.32), $\{f^j\}$ are "Hermitian" 0-forms, and M is a nonsingular $(k+p)\times(k+p)$ HCM.

Proof: "If" part is trivial, and therefore we show "only if" part. The proof is performed in the analytical method, which is found in many textbooks[12] in the ordinary manifold case. However, we should be careful about dealing not with functions but with elements of sheaf $A(U)$, but we fully make use of the fact that there exists an isomorphism $\tilde{}$: $C(U) \to C^{\infty}(U)$. As in the ordinary case, we consider a sufficiently small neighborhood, for even coordinates, but there are no such counterparts for odd coordinates.

We first note if we put $\omega^i = \Sigma_{j=1}^{k+p} \omega'^j N_j{}^i$ with a nonsingular $(k+p)\times(k+p)$ matrix N, then for $\{\omega'^j$: $j = 1,\cdots,k+p\}$, the condition (2D.31) holds. Note it is possible to take N such that ω' be in the form (2D.33) and "real." Thus it is sufficient to prove (2D.32) for ω', and we can assume,

$$\omega^i = d\eta^i - d\bar{\eta}^j \phi_{ji}(\eta,\bar{\eta}),$$

where $(\eta^i, \bar{\eta}^j; i = 1,\cdots,k+p; j = 1,\cdots,m+q)$ is a HCS for $A(U)$ and $\eta^i = r^i$ $(i = 1,\cdots,k)$, $\eta^{i+k} = s^i$ $(i = 1,\cdots,p)$, $\bar{\eta}^i = \bar{r}^i$ $(i = 1,\cdots,m)$, $\bar{\eta}^{i+m} = \bar{s}^i$ $(i = 1,\cdots,q)$. Thus $(\widetilde{r^i}, \widetilde{\bar{r}^j}; i = 1,...,k; j = 1,\cdots,m)$ is a LCS for U. We assume, without any loss of generality, $\widetilde{r^i}=\widetilde{\bar{r}^j}=0$ at $P\in U$, and U is sufficiently small.

In the first step, we will show that for a differential equation

$$\frac{\partial\xi^i(t)}{\partial t}=\bar\eta^j\phi_{ji}(\xi(t),t\bar\eta)\ \begin{pmatrix} i=1,\cdots,k+p\\ j=1,\cdots,m+q\end{pmatrix}, \qquad (2D.34)$$

or alternatively, for an integral equation

$$\xi^i(t)=\eta^i+\int_{1_{I\times U}}^{t}dt'\bar\eta^j\phi_{ji}(\xi(t'),t'\bar\eta), \qquad (2D.35)$$

there exists a unique solution $\xi^i(t)\in B(I\times U)$ with boundary condition $\xi^i(1_U)=\eta_i\in A(U)$ and $\xi^i(0_U)\equiv f^i\in A(U)$ such that $(\xi^i(t),\bar\eta^j)$ for fixed t, $t=t_0 1_U$ $(0\leqslant t_0\leqslant 1)$, be an A-CS for $A(U)$. Here one notes, in the statement above, strictly speaking, we should have considered, as we did in the proof of Poincaré's lemma, morphisms of graded manifolds many times, for everytime when we did substitutions in the arguments of function notation. For simplicity of notations, we neglect those processes and instead do with substitution in arguments of the function notation as such, and also ignore subscripts of 0_U or $1_{I\times U}$, with proper understanding. Here $(t,\eta^i,\bar\eta^j)$ is a A-CS for $B(I\times U)$, and $t\in B(I\times U)_0$. Now let us start with an integral equation

$$\xi^i(t)=\epsilon^i+\int_0^t dt'\bar\eta^j\phi_{ji}(\xi(t'),t'\bar\eta), \qquad (2D.36)$$

where $\epsilon^i\in B(I\times U)$, $\partial_t\epsilon^i=0$, $\widetilde{\epsilon^i}\in C^\infty(I\times U)$, $\xi^i(t)\in B(I\times U)$. If we expand both sides of (2D.36) in terms of odd coordinates $(s^i,\bar s^j;\ i=1,\cdots,p;\ j=1,\cdots,q)$,

$$\xi^i{}_{\mu\nu}(t)=\epsilon^i{}_{\mu\nu}+\int_0^t dt'G^i{}_{\mu\nu}(\xi_{\alpha\beta}(t'),\bar r,t'), \qquad (2D.37)$$

where

$$\xi^i(t)\equiv s^\mu\bar s^\nu\xi^i{}_{\mu\nu}(t),\quad \epsilon^i\equiv s^\mu\bar s^\nu\epsilon^i{}_{\mu\nu}$$

and

$$\bar\eta^j\phi_{ji}(\xi,t'\bar\eta)\equiv s^\mu\bar s^\nu G^i{}_{\mu\nu}(\xi^{i'}{}_{\alpha\beta}(t'),\bar r,t'), \qquad (2D.38)$$

and $\mu=(\mu_1,\cdots\mu_{(\mu)})$, $1\leqslant\mu_1<\cdots<\mu_{(\mu)}\leqslant p$, $s^\mu=s^{\mu_1}s^{\mu_2},\cdots,s^{\mu_{(\mu)}}$, for $(\mu)=0$, $s^\mu\equiv 1$, and $\bar s^\nu$ similarly understood and $G^i{}_{\mu\nu}(\cdot,\cdot,\cdot)\in\mathscr{C}^\infty(I\times U)$. An important remark is that

$$\frac{\partial}{\partial\xi^{i'}{}_{\alpha\beta}}G^i{}_{\mu\nu}=0\ \text{ for }\ (\alpha)>(\mu)\ \text{ or }\ (\beta)>(\nu). \qquad (2D.39)$$

Let us take $\sim$ on both sides of (2D.36) for $i=1,\cdots,k$, and then we have the integral equations in the same form as those appearing in the ordinary manifold case in Ref. 12, which are real integral equations following from ω^i being "real," and all discussions in the ordinary case apply, which say, for sufficiently small U, for fixed t $(0\leqslant t\leqslant 1)$, $(\widetilde{\xi^i(t)},\widetilde{\bar r^j};\ i=1,\cdots,k;\ j=1,\cdots,m)$ is a LCS for U, which coincides with $(\widetilde{r^i,\bar r^j})$ at $t=1$. Equations (2D.37) for $(\mu)+(\nu)>0$ are complex ones, but from general argument on integral equations of this type, from the original form (2D.36), and from (2D.39), which make iterative method applicable in (2D.37), starting from the solution for $(\mu)+(\nu)=0$, one finds there exists a solution $\xi^i{}_{\mu\nu}(t)$, if we use a A-CS $(r,\bar r,\lambda s,\lambda\bar s)$ where λ is a small positive number, such that $\xi^i{}_{\mu\nu}(1)=\lambda$ with $(\mu)=1$, $(\nu)=0$ and $\xi^i{}_{\mu\nu}(1)=0$, otherwise for $(\mu)+(\nu)>0$. General argument on (2D.37) for small t, can now be extended to $(0\leqslant t\leqslant 1)$, and it holds

$$\det\widetilde{\frac{\partial(\xi^i{}_{\mu\nu}(t))}{\partial(\xi^j{}_{\alpha\beta}(1))}}\neq 0,$$

which derives the followings from (2D.39);

$$\det\widetilde{\left(\frac{\partial\xi^i(t)}{\partial\eta^j}\right)}\neq 0,\quad (i,j=1,\cdots,k)$$

and

$$\det\widetilde{\left(\frac{\partial\xi^{i+k}(t)}{\partial(\lambda s)^j}\right)}\neq 0\quad (i,j=1,\cdots,p). \qquad (2D.40)$$

Summing up all of the above, we have a solution $\xi^i(t)$ $(i=1,\cdots,k+p)$ for $0\leqslant t\leqslant 1$, to (2D.36) in A-CS $(r,\lambda s,\bar r,\lambda\bar s)$ with sufficiently small U such that $(\xi^i(1);\ i=1,\cdots,k+p)=(r,\lambda s)$ and $(\xi^i(t),\bar r,\lambda\bar s;\ i=1,\cdots,k+p)$ is an A-CS for fixed t, $0\leqslant t\leqslant 1$. Here one notes that if one changes $s\rightarrow(1/\lambda)s$ in (2D.38) with the solution to (2D.37) unchanged, there results another solution $\xi^i(t)$ to (2D.36) in the A-CS $(r,s,\bar r,\bar s)$ such that $(\xi^i(t),\bar r,\bar s;\ i=1,\cdots,k+p)$ is an A-CS for fixed t, $0\leqslant t\leqslant 1$, which coincides with $(r,s,\bar r,\bar s)$ at $t=1$.

Following steps are exactly the same as those in the ordinary manifolds case. We substitute,

$$(\xi^i(t),t\bar\eta^j;\ i=1,\cdots,k+p;\ j=1,\cdots,m+q)\ \text{ for }\ (\eta^i,\bar\eta^j), \qquad (2D.41)$$

in (2D.33) and we put

$$\omega^i(t)\equiv d\xi^i(t)-d(t\bar\eta^j)\phi_{ji}(\xi(t),t\bar\eta)\in\Omega^1(I\times U,B), \qquad (2D.42)$$

where note that $\omega^i=\omega^i(1)\in\Omega^1(U,A)$. We work in an A-CS for $B(I\times U)$, $(t,f^i,\bar\eta^j;\ i=1,\cdots,k+p;\ j=1,\cdots,m+q)$, where $f^i\equiv\xi^i(0)$ $(i=1,\cdots,k+p)$. Then

$$\omega^i(t)=\sum_{j=1}^{k+p}df^j\frac{\partial\xi^i(t)}{\partial f^j}+\sum_{j=1}^{m+q}d\bar\eta^jF_{ji}(t), \qquad (2D.43)$$

where

$$F_{ji}(t)\equiv\frac{\partial\xi^i(t)}{\partial\bar\eta^j}-\phi_{ji}(\xi(t),t\bar\eta)t. \qquad (2D.44)$$

Under the substitution (2D.41), we put

$$\theta_j{}^i(t)\equiv dtP_{ji}(t)+\sum_{h=1}^{k+p}df^hQ_{hij}(t)+\sum_{h=1}^{m+q}d\bar\eta^hR_{hij}(t), \qquad (2D.45)$$

and then consider the condition (2D.31) and compare terms proportional to $dtd\bar\eta^j$ on both sides of (2D.31), we obtain

$$\partial_tF_{ji}(t)=-\sum_{h=1}^{k+p}F_{jh}(t)P_{hi}(t) \qquad (2D.46)$$

with

$$F_{ji}(0)=0\quad (i=1,\cdots,k+p;\ j=1,\cdots,m+q) \qquad (2D.47)$$

from (2D.44). If we consider expansion on both sides of (2D.46) like (2D.37), then given coefficient functions of $P_{hi}(t)$ are isomorphic to those $\in C^\infty(I\times U)$ and we have linear differential equations isomorphic to those which satisfy Lipschitz's condition and have unique solutions identically vanishing for $0\leqslant t\leqslant 1$. Thus we have $F_{ji}(t)=0$ and $\omega^i=\omega^i(1)=\Sigma_{j=1}^{k+p}df^j\ (\partial\eta^i/\partial f^j)$ where note the appearing matrix is nonsingular and "real" from $(\eta, \bar{\eta})$ being the original A-CS. Further, if $\{\omega^i\}$ are "Hermite", then $N \in \mathscr{H}(k+p) \times (k+p)$, and $\{\omega''\}$ are also "Hermite", and r.h.s. of (2D.34) are also, if $\xi^i(t)$ are, and we have "Hermitian" solutions $\xi^i(t)$, to (2D.35) and $(f, \bar{\eta})$ is a HCS. Q.E.D.

We do not know if Frobenius' theorem is valid or not when we relax conditions, "real" or "homogeneous." However it may be, we are not interested, at least, in the latter case, since nonhomogeneous 0-forms are not suitable as coordinates, chain rule of derivations invalidated.

As corollary of Frobenius' theorem, we have the complete integrability condition for the following equations.

Corollary 53: Let $\{X_i, i = 1,\cdots,k+p\}$ be linearly independent homogeneous "real" derivations, where "real" is similarly defined as in 1-forms, and the dimension of graded manifold be $(k+m, p+q)$ even dim, $k+m$, odd dim, $p+q$. Consider differential equations of $f \in A(U)$, where U is contractible,

$$X_i f = 0, \tag{2D.48}$$

where $|X_i| = 0$ $(i = 1,\cdots,k)$ $|X_{i+k}| = 1$ $(i = 1,\ldots, p)$. There exist homogeneous solutions $\{f^j; j = 1,\cdots,m+q\}$ which can be extended to an A-CS $\{f^\mu; \mu = 1,\cdots,(k+m) + (p+q)\} \supset \{f^j\}$, if and only if it holds that

$$[X_i, X_j] = \sum_{h=1}^{k+p} C_{ij}{}^h X_h, \quad (i,j = 1,\cdots,k+p), \tag{2D.49}$$

where $C_{ij}{}^h \in A(U)$. Further if all $\{X_i\}$ are "Hermite", then all $\{f^i: i = 1,\cdots,m+q\}$ can be "Hermite".

The proof goes in the same way as in the ordinary manifold case, since (2C.34), Remarks 49, 50, 51, and Frobenius' theorem are available.

3. GRADED SYMPLECTIC MANIFOLDS WITH "HERMITICITY" AND HAMILTONIAN FORMULAS

We call a graded manifold with "Hermiticity", (X, A), which is given a nonsingular "anti-Hermitian" closed 2-form ω of Z_2 degree zero, as a graded symplectic manifold with "Hermiticity", (X, A, ω). The 2-form, ω, is called as an "anti-Hermitian" symplectic form.

Now we can proceed as in the usual way, to define Poisson bracket bilinear operation on $A(U)$, which gives a graded Lie algebra structure on $A(U)$. We survey the process here emphasizing "Hermiticity" structures.

ω gives a $A(U)$ left linear map: $\mathrm{Der} A(U) \to \Omega^1(U, A)$, by $i_X\omega$ for $X \in \mathrm{Der} A(U)$. Since ω is nonsingular, and the map is an isomorphism, we can define a $A(U)$ left linear mapping, $\Omega^1(U, A) \ni \alpha \to X_{(\alpha)} \in \mathrm{Der} A(U)$ by

$$i_{X_{(\alpha)}}\omega = \alpha. \tag{3.1}$$

Remark 54: If $\alpha \in \Omega^1(U, A)$ is homogeneous, then $X_{(\alpha)}$ is homogeneous, $|X_{(\alpha)}| = |\alpha|$.

We write for $f \in A(U)$, $X_{(df)} \equiv X_f$.

Remark 55: For $f \in A(U)_{|f|}$,

$$X_{f\dagger} = (-1)^{|f|}(X_f)^{\dagger\prime}. \tag{3.2}$$

Thus for homogeneous "Hermitian" f, X_f is "Hermite" ("anti-Hermitian") if $|f| = 0$ $(|f| = 1)$.

For the proof, use Remark 42, in the definition (3.1).

Definition of Poisson bracket $\{\ ,\ \}$: C-bilinear mapping $\{\ ,\ \}: A(U) \otimes_C A(U) \to A(U)$ such that for $f, g \in A(U)$,

$$\{f, g\} \equiv \langle X_f | dg\rangle = X_f g. \tag{3.3}$$

Remark 56: For homogeneous $f, g \in A(U)$ and $h \in A(U)$,

$$\{f, gh\} = \{f, g\}h + (-1)^{|f||g|}g\{f, h\}. \tag{3.4}$$

For the proof, estimate $X_f(gh)$.

Theorem 57: $A(U)$ is a graded Lie algebra with respect to the bracket operation $\{f, g\}$; for homogeneous $f, g \in A(U)$

$$\{f, g\} = (-1)^{1+|f||g|}\{g, f\}, \tag{3.5}$$

$$\sum_{\substack{\text{cyclic}\\ \text{perm}}} (-1)^{|f||h|}\{f, \{g, h\}\} = 0 \quad \text{(Jacobi identity)}. \tag{3.6}$$

Furthermore $f \to X_f$ is a homomorphism of graded Lie algebras, that is

$$X_{\{f,g\}} = [X_f, X_g], \tag{3.7}$$

for $f, g \in A(U)$, where $i_{X_f}\omega = df$.

Proof: (3.5) follows from

$$\{f, g\} = i_{X_f} i_{X_g}\omega = -\langle X_f, X_g | \omega\rangle, \tag{3.8}$$

(3.7) from Remark 26,

$$\begin{aligned} i_{X_{\{f,g\}}}\omega &= d\{f, g\} = d i_{X_f} dg = \theta_{X_f} dg = \theta_{X_f} i_{X_g}\omega \\ &= [\theta_{X_f}, i_{X_g}]\omega = i_{[X_f, X_g]}\omega \end{aligned} \tag{3.9}$$

and (3.6) from (2C.32) with $n = 2$, $X_1 = X_f$, $X_2 = X_g$, $X_3 = X_h$. Q.E.D.

Remark 58: Let homogeneous $f, g \in A(U)$, then

$$\{f, g\}^\dagger = (-1)^{|f||g|}\{f^\dagger, g^\dagger\} \tag{3.10}$$

which implies if f, g are "Hermite", then $\{f, g\}$ is "anti-Hermite" ("Hermite") when $|f| = |g| = 1$ (otherwise).

For the proof, use Remark 42 and (3.2) on

$$\{f, g\} = i_{X_f} dg. \tag{3.11}$$

Now we give Darboux's theorem on our graded symplectic manifold with "Hermiticity", which makes so called canonical coordinates explicit. Besides "Hermite" p's and q's on the alternating part, "Hermite" odd coordinates on the quadratic part appear, and difference from Kostant's case consists in that the normalized coefficient of the quadratic part is $(+i/2)$ or $(-i/2)$ here instead of $(+1/2)$ or $(-1/2)$ in his case. Therefore in the sense above, there is also here, an invariant, the signature of the quadratic part as Kostant noted in his case. However we note later from physical point of view that the coefficients should be all $(+i/2)$ in connection with the conventional quantization procedure and the requirement of positive definite metric of state vectors. And one will find the "anti-Hermiteness" of our symplectic form ω fits perfectly with the conventional quantization.

Let dim (X, A, ω) be (m',n). Then from ω being nonsingular, one finds $m' = 2m$ even, considering $\sim$ of component matrix of ω whose antisymmetric part has nonzero determinant.

Definition of "Hermite" A-canonical coordinate system (HCCS): Let ω be as above. If there exists a HCS $(\xi^\mu;\ \mu = 1,\cdots,2m+n) \equiv (p^i; q^i; s^j: i = 1,\cdots,m; j = 1,\cdots,n)$, where $|\xi^\mu| = 0$ for $\mu = 1,\cdots,2m$, and $|\xi^{\mu+2m}| = 1$ for $\mu = 1,\cdots,n$, such that

$$\omega = \sum_{k=1}^{m} dp^k dq^k + \sum_{j=1}^{n} i(\epsilon_j/2) ds^j ds^j, \tag{3.12}$$

where $\epsilon_j = +1$ or -1, then we call the HCS $(p^i; q^i; s^j)$ as "Hermite" A-canonical coordinate system (HCCS).

Such a coordinate neighborhood U as above which admits a HCCS is called an *A-canonical coordinate neighborhood.*

Darboux's theorem: Any graded symplectic manifold (X, A, ω) with "Hermiticity" can be covered by A-canonical coordinate neighborhoods.

The proof is similar to the method adopted in Ref. 9 in the ordinary manifold case, and is accomplished in a few steps by making use of Corollary 53. The detail is given in the Appendix.

Remark 59: In HCCS $(p^i; q^i; s^j: i = 1,\cdots,m; j = 1,\cdots,n)$,

$$X_{p^i} = -\frac{\partial}{\partial q^i}, X_{q^i} = \frac{\partial}{\partial p^i}, X_{s^j} = -i\epsilon_j \frac{\partial}{\partial s^j}, \tag{3.13}$$

and as a consequence one has Poisson bracket relations,

$$\{q^i, p^j\} = \delta_{ij} 1_U,\ \{s^i, s^j\} = -i\delta_{ij}\epsilon_j 1_U, \tag{3.14a}$$

and

$$\{\ ,\ \} = 0 \text{ for other combinations.} \tag{3.14b}$$

Thus for $f, g \in A(U), f = f_0 + f_1$

$$\{f, g\} = \sum_{\alpha=0,1} \{f_\alpha, g\}, \tag{3.15}$$

$$X_{f_\alpha} = \sum_{k=1}^{m} \frac{\partial f_\alpha}{\partial q^k}\frac{\partial}{\partial p^k} - \frac{\partial f_\alpha}{\partial p^k}\frac{\partial}{\partial q^k} + \sum_{j=1}^{n} i(-1)^{|f_\alpha|}\epsilon_j \frac{\partial f_\alpha}{\partial s^j}\frac{\partial}{\partial s^j}, \tag{3.16}$$

and

$$\{f_\alpha, g\} = \sum_{k=1}^{m} \frac{\partial f_\alpha}{\partial q^k}\frac{\partial g}{\partial p^k} - \frac{\partial f_\alpha}{\partial p^k}\frac{\partial g}{\partial q^k} + \sum_{j=1}^{n} i(-1)^{|f_\alpha|}\epsilon_j \frac{\partial f_\alpha}{\partial s^j}\frac{\partial g}{\partial s^j}. \tag{3.17}$$

The signature of ω, a pair of numbers of odd coordinates $\epsilon_j = 1$ and of those $\epsilon_j = -1$, is invariant which is seen; let

$$\omega = (d\xi)(\Omega/2)(d\xi)^T \tag{3.18}$$

in the notations, (2D.1)–(2D.17) in HCS (ξ). Then

$$\Omega^\dagger = -\Omega \tag{3.19}$$

and Ω^{--} is symmetric and let $\Omega^{--} = i\Omega^{--\prime}$, then $\widetilde{\Omega^{--\prime}}$ is real symmetric, and now the result is known as Sylvester's law of inertia.

Finally we consider time evolution of elements $f(t) \in A(U)$, where t is considered as a real parameter. We assume there exists some element $H \in A(U)_0$ called Hamiltonian which is "Hermitian" and governs the time evolution by

$$\frac{\partial f}{\partial t} \equiv X_H f = \{H, f\} = -\{f, H\}. \tag{3.20}$$

Note the opposite sign found in (3.20) to that of classical mechanics appearing in the standard textbooks comes from that considered is, here, the time evolution of an element of sheaf $A(U)$, one may think, of functional form in a fixed coordinate neighborhood, while in the standard textbooks, the time evolution of values along the orbits in a fixed coordinate system.

In the time evolution (3.20), $\alpha \in \Omega(U, A)$ changes according to

$$\dot{\alpha} = \theta_{X_H}\alpha, \tag{3.21}$$

and especially one should note,

$$\dot{\omega} = \theta_{X_H}\omega = 0, \tag{3.22}$$

which gives invariants ω^k where $k = 0,1,\cdots,m$ in the ordinary symplectic manifold of dim $2m$, while here k is not limited by m, i.e., $k \in Z_+$, and those invariants bear statistical significance in the former, while the meaning of them is not quite clear yet in the latter.[3]

One may say here the physical meaning of classical Bose–Fermi systems, itself, is not clear. Truly we admit that in the direct sense, but, we believe it will play important roles in connection with quantized Bose–Fermi systems, especially in symmetry properties, so called sypersymmetry, of the quantized systems, such as those in supergravity theories and BRS symmetries in gauge theories.

Here we would like to comment on the relation between parameters and dynamical variables of Bose and Fermi properties appearing in the theories mentioned above, in connection with the graded symplectic manifold here. Bose parameters, say $\lambda^i \in C$, can be always incorporated in the graded symplectic manifold in the form $\lambda^i 1_U$, where $d(\lambda 1_U) = 0$. However, as far as the graded manifold we have discussed concerns, there is no room for constant odd elements v^i, $dv^i = 0$. Therefore we need to prepare two types of graded manifolds with "Hermiticity", B and $A(U)$, B for the parameters and $A(U)$ for the dynamical variables, where B of dim $(0, k)$ and $A(U)$ of dim $(2m,n)$ and we assume $A(U)$ to be B module and Z_2 property of $A(U)$ to be affected by B and $d(b 1_U) = 0$ for all $b \in B$. Then with an "anti-Hermitian" symplectic form ω, we obtain a graded symplectic manifold (X, A, ω) which allows coordinate transformations including parameters of Bose and/or Fermi type.

Finally in this section, we cite a lemma which is a byproduct in the proof of Darboux's theorem. We call a set of homogeneous "Hermitian" 0-forms $\{f^i;$ $i=1,\cdots,k<2m+n\}$ such that $\{df^i\}$ are linearly independent, as *sub*-HCCS (SHCCS) if $\{f^i\}$ satisfy Poisson bracket relations among them like a subset of HCCS.

Lemma 60: If $\{f^i; i=1,\cdots,k<2m+n, f^i\in A(U)\}$ is a SHCCS, then the SHCCS $\{f^i\}$ can be extended to a HCCS $\{f^\mu\}\supset$SHCCS $\{f^i\}$.

4. DISCUSSIONS AND CONCLUSIONS

Firstly we remark that canonical transformations, here, including those mixing Z_2 even and odd dynamical variables, are now treated on the same level as those of classical mechanics, which are found in many textbooks. From (3.12), two HCCS's, $(p^i; q^i; s^j)$ and $(p'^i; q'^i; s'^j)$ are related such that

$$\omega = d\alpha = d\alpha', \tag{4.1}$$

where α and α' are "Hermitian" 1-forms and suitably expressed by $(p^i; q^i; s^j)$ and $(p'^i; q'^i; s'^j)$, respectively, for example

$$\alpha = \sum_{k=1}^{m} p^k dq^k + \sum_{j=1}^{n} i\frac{\epsilon_j}{2} s^j ds^j. \tag{4.2}$$

For fixed choices of α and α', there exists "Hermite" $F\in A(U)_0$ such that

$$\alpha - \alpha' = dF, \tag{4.3}$$

since $d(\alpha-\alpha')=0$ from (4.1). Conversely by giving a suitable functional form of "Hermitian" F, various types of canonical transformations can be formed through (4.3), just as discussed in the standard textbooks[14] of classical mechanics, which we do not repeat here.

Secondly we extend the definition of canonical coordinate systems such that it admits some non-"Hermitian" coordinate systems, which are useful especially in quantum mechanics. If we put

$$a^k \equiv \frac{1}{\sqrt{2}}(p^k - iq^k),\quad a^{k\dagger} = \frac{1}{\sqrt{2}}(p^k + iq^k),\quad (k=1,\cdots,m) \tag{4.4}$$

then p^k, q^k are uniquely expressed by a^k, $a^{k\dagger}$, and thus all elements $f\in A(U)$ can be expressed by $(a^k; a^{k\dagger}; s^j$: $k=1,\cdots,m$, $j=1,\cdots,n)$. In this sense, $(a^k; a^{k\dagger}; s^j)$ may be included in A-CS's although the system does not fit the definition of A-CS given before since $\widetilde{a^k}, \widetilde{a^{k\dagger}}\notin C^\infty(U)$. In this A-CS $(a^k; a^{k\dagger}; s^j)$, ω has the form,

$$\omega = \sum_{k=1}^{m} i\, da^{k\dagger} da^k + \sum_{j=1}^{n} i\frac{\epsilon_j}{2} ds^j ds^j, \tag{4.5}$$

and Poisson bracket relations are

$$\{a^k, a^{j\dagger}\} = -i\delta_{kj} 1_U \quad (k,j=1,\cdots,m), \tag{4.6}$$

$$\{s^k, s^j\} = -i\epsilon_j \delta_{kj} 1_U \quad (k,j=1,\cdots,n), \tag{4.7}$$

and for other combinations

$$\{\ ,\ \} = 0. \tag{4.8}$$

Conversely if there exists a set $(a^k; a^{k\dagger}; s^j$: $k=1,\cdots,m$; $j=1,\cdots,n)$, such that a^k, $a^{k\dagger}$, Z_2 even and s^j, Z_2 odd 'Hermite," and (4.5) holds, then by the relation (4.4) we find a HCCS $(p^k; q^k; s^j)$. Thus we may include such a system in the definition of *A-canonical coordinate system* and we denote it as CCS. Further for a pair of odd element s^j, say s^1 and s^2, which have the same ϵ_j, $\epsilon_1=\epsilon_2$, we define new elements b, $b^\dagger$ by

$$b = \frac{1}{\sqrt{2}}(s_1 - is_2),\quad b^\dagger = \frac{1}{\sqrt{2}}(s_1 + is_2), \tag{4.9}$$

then

$$\sum_{j=1,2} i\frac{\epsilon_j}{2} ds^j ds^j = i\epsilon_1 db^\dagger db. \tag{4.10}$$

By the similar argument to that on a, $a^\dagger$, we may include a system (ξ^μ) such that b, $b^\dagger$ appear instead of s^1, s^2 in HCCS or CCS above, into the definition of CCS. In this case,

$$\{b, b^\dagger\} = -i\epsilon_1 1_U, \tag{4.11}$$

and for b, $b^\dagger$ with other elements,

$$\{b,\ \} = \{b^\dagger,\ \} = 0. \tag{4.12}$$

Now let us see how this classical system fits in with the conventional quantization procedure in simple-minded level, although we know that the procedure applied to manifolds having nontrivial global structure, is not justified and ignores the problem of operator orderings. The procedure reads: for a set of operators $(\hat{\xi}^\mu)$ corresponding to a CCS (ξ^μ), put

$$[\hat{\xi}^\mu, \hat{\xi}^\nu] = i\{\xi^\mu, \xi^\nu\}\hat{1}, \tag{4.13}$$

where

$$[\hat{\xi}^\mu, \hat{\xi}^\nu] \equiv \hat{\xi}^\mu\hat{\xi}^\nu - (-1)^{|\xi^\mu||\xi^\nu|}\hat{\xi}^\nu\hat{\xi}^\mu, \tag{4.14}$$

and we assume $\hat{\xi}^\mu$ is a Hermite operator if ξ^μ is "Hermite". From (4.13), it follows that

$$[\hat{p}^k, \hat{q}^j] = -i\delta_{kj}\hat{1}, \tag{4.15}$$

$$[\hat{s}^k, \hat{s}^j] = \epsilon_j\delta_{kj}\hat{1}, \tag{4.16}$$

and for all other combinations among $(\hat{p}^k; \hat{q}^k; \hat{s}^j)$

$$[\ ,\ ] = \hat{0}. \tag{4.17}$$

Here first note if we use Kostant's graded symplectic manifold which differs in reality concept from ours, and respect (4.13), then the r.h.s. of (4.16) becomes imaginary while the l.h.s. of (4.16) is Hermite from realness of s, which is a contradiction. Secondly note ϵ_j on the r.h.s. of (4.16) should be positive since the l.h.s. of (4.16) is a positive definite operator. Thus we put $\epsilon_j=1$. Next let us see the commutation relations in use of $\hat{a}^k$, $\hat{a}^{j\dagger}$ or $\hat{b}^k$, $\hat{b}^{j\dagger}$, where we assume n is even.

$$[\hat{a}^k, \hat{a}^{j\dagger}] = \delta_{kj}\hat{1}, \tag{4.18}$$

$$[\hat{b}^k, \hat{b}^{j\dagger}] = \delta_{kj}\hat{1}, \tag{4.19}$$

and for all other combinations among $(\hat{a}^k; \hat{a}^{k\dagger}; \hat{b}^j; \hat{b}^{j\dagger}$: $k=1,\cdots,m, j=1,\cdots,n/2)$,

$$[\ ,\ ] = \hat{0}. \tag{4.20}$$

(4.18) is seen in harmonic oscillator models, and (4.19) appears in treating fermions. We do not say $\hat{a}$ or $\hat{b}$ is an annihilation operator, which should follow from the structures of Hamiltonians. Finally note if we start from a Lagrangian including fermionic part then it leads to a constrained Hamiltonian system and requires special treatment, i.e., Dirac's method, which appeared in Ref. 3 and which will be discussed elsewhere in connection with the problem of existence of constraints in 'standard' form[9] in the graded symplectic manifold case.

As conclusions we sum up what is accomplished in this paper. A kind of complexification of graded manifold theory is given, which is different from Kostant's one in reality concept. As a conjugation operation to define reality concept in complex shief over $U, A(U)$, we adopt $\dagger$ "Hermitian" conjugation; for $a, b, \in A(U)$, $(ab)^\dagger = b^\dagger a^\dagger$ while Kostant adopted for $a, b \in A(U)$ $(ab)^* = a^* b^*$. We have shown the assumption of the existence of antiautomorphism of $A(U)$ reduces to the existence of "Hermite" coordinate systems. We also defined "Hermiticity" of derivations of $A(U)$, and of differential forms and of derivations of differential forms, all coordinate independent way. We have characterized "classical" Bose–Fermi systems by the graded symplectic forms ω, which are "anti-Hermitian" nonsingular closed 2-forms of Z_2 grading zero. Frobenius' theorem on the graded manifold, and also Darboux's theorem on the graded symplectic manifold are given. A comment on the relation of the dynamical variables and the parameters which appear in the theory in the superspace, is given in terms of the graded symplectic manifolds. A remark is given on the canonical transformations. We have seen how our system fits in with the naive conventional quantization procedures. Finally, we would like to note that it seems theoretical physicists use $\dagger$ for the reality concept while mathematicians, more or less, who talk about manifolds, use * and discuss a real subalgebra. In this situation, we would like to emphasize the contents presented here are new in that $\dagger$ is incorporated into the graded manifolds, as far as we know.

ACKNOWLEDGMENTS

This work was started at Utsunomiya University, and completed at Dalhousie University. It is supported in part by the National Sciences and Engineering Research Council Canada under Grant No. A3966. The author would like to express sincere thanks to Professor D. Kiang for his kind interest in the work and hospitality extended to him, by the Physics Department of Dalhousie University.

APPENDIX

Here we prove Darboux's Theorem by proving Lemma 60. We work on $A(U)$ over U, sufficiently small contractible neighborhood. We prepare Lemmas A1 and A2 for Lemma A3 which is nothing but Lemma 60.

Before going into the lemmas, we consider general properties of a component matrix of the "anti-Hermitian" symplectic form ω in some A-CS $(\xi^\mu; \mu = 1,\cdots,N)$. The ω is characterized by (i) $\omega^{\dagger t} = -\omega$, (ii) ω; nonsingular 2-form, i.e., the linear map: $\mathrm{Der}A(U) \to \Omega^1(U,A)$, by $i_X\omega$, is nonsingular, (iii) $d\omega = 0$, and (iv) $|\omega| = 0$. Let $(\xi^\mu; \mu = 1,\cdots,N)$ be an A-CS. Then ω can be written as

$$\omega = (d\xi)\frac{\Omega}{2}(d\xi)^T, \tag{A1}$$

where the notation is as those used from (2D.1) to (2D.17) and Ω is the component matrix. Only from Z_2 property, one may assume, by putting

$$\Omega = \begin{pmatrix} \Omega^{++} & \Omega^{+-} \\ \Omega^{-+} & \Omega^{--} \end{pmatrix}, \tag{A2}$$

that

$$\Omega^{++t} = -\Omega^{++}, \quad \Omega^{+-t} = \Omega^{-+}, \quad \Omega^{--t} = \Omega^{--}. \tag{A3}$$

Further if $(\xi^\mu; \mu = 1,\cdots,N)$ is a HCS, then from $\omega^{\dagger t} = -\omega$, one may assume Ω be "anti-Hermitian" matrix, i.e., $\Omega^\dagger = -\Omega$ where $\dagger$ is defined (2D.12), and together with (A3) one sees components of Ω^{++} are "Hermite" and those of Ω^{+-} Ω^{-+} and Ω^{--} are "anti-Hermite". From the definition of X_f, i.e.,

$$i_{X_f}\omega = df, \tag{A4}$$

one obtains

$$X_f = (\partial f)^T \Omega^{-1}(\partial), \tag{A5}$$

where $df = (d\xi)(\partial f)$ and (∂f) should be understood as

$$(\partial f) \in \mathscr{M}(\cdot + \cdot) \times (1 + 0) \quad \text{if} \quad |f| = 0, \tag{A6}$$

and

$$(\partial f) \in \mathscr{M}(\cdot + \cdot) \times (0 + 1) \ \text{if} \ |f| = 1. \tag{A7}$$

Thus one finds

$$X_{\xi^\mu}\xi^\nu = \{\xi^\mu, \xi^\nu\} = (\Omega^{-1})_{\mu\nu}, \tag{A8}$$

and also

$$(\Omega^{-1})^{++t} = -(\Omega^{-1})^{++}, \quad (\Omega^{-1})^{+-t} = -(\Omega^{-1})^{-+},$$
$$(\Omega^{-1})^{--t} = (\Omega^{-1})^{--}, \tag{A9}$$

and if $(\xi^\mu; \mu = 1,\cdots,N)$ is a HCS, then components of $(\Omega^{-1})^{++}$, $(\Omega^{-1})^{+-}$ and $(\Omega^{-1})^{-+}$ are "Hermite" and those of $(\Omega^{-1})^{--}$ "anti-Hermite". From (ii), we have

$$\det \tilde{\Omega} \neq 0. \tag{A10}$$

Now we give lemmas, where we suppose the total dimension of the graded symplectic manifold with "Hermiticity" be N.

Lemma A 1: Let $\{\xi^i; i = 1,\cdots,k < N\}$ be a set of homogeneous 0-forms such that $\{d\xi^i; i = 1,\cdots,k\}$ be linearly independent "real" 1-forms and

$$\{\xi^i, \xi^j\} = C^{ij}(\xi) \quad (i,j = 1,\cdots,k) \tag{A11}$$

be valid. Then there exists a A-CS $(\eta^\mu; \mu = 1,\cdots,N)$ such that

$$\{\xi^i, \eta^j\} = 0 \quad (i = 1,\cdots,k; \quad j = k+1,\cdots,N) \tag{A12}$$

and

$$X_{\xi^i} = \sum_{j=1}^{k} \{\xi^i, \eta^j\}\frac{\partial}{\partial \eta^j} \quad (i = 1,\cdots,k). \tag{A13}$$

Further if $\{\xi^i; i = 1,\cdots,k\}$ are all "Hermite", then A-CS $(\eta^\mu; \mu = 1,\cdots,N)$ could be HCS $(\eta^\mu; \mu = 1,\cdots,N)$.

Proof: We first note from (3.7) and the definition of X_f, that

$$[X_{\xi^i}, X_{\xi^j}]$$
$$= X_{C^{ij}(\xi)} = \sum_{m=1}^{k} (-1)^{|\xi^m|(1+|\xi^i|+|\xi^j|)} \frac{\partial C^{ij}(\xi)}{\partial \xi^m} X_{\xi^m}. \tag{A14}$$

Now Corollary 53 can be applied since $\{X_{\xi^j}\}$ are "real". In the case that $\{\xi^i\}$ be all "Hermite", $\{X_{\xi^i}\}$ are all "Hermite" or "anti-Hermite" from Remark 55, and then multiplying X_{ξ^i} suitably by i, we may apply the last part of Corollary 53. (A13) is trivial from the definition of Poisson bracket. Q.E.D.

We have defined a notion SHCCS $\{\xi^i; i = 1,\cdots,n < N\}$ just above Lemma 60. Here we introduce a notation SHCCS_m^k $\{\xi^i; i = 1,\cdots,k+m < N\}$ which implies that it is a SHCCS $\{\xi^i; i = 1,\cdots,k+m\}$ and that

$$\det \widetilde{\{\xi^i,\xi^j\}} \neq 0 \quad (i,j = 1,\cdots,k) \tag{A15}$$

and

$$\{\xi^i,\xi^{j+k}\} = 0 \quad (i = 1,\cdots,k+m;\ j = 1,\cdots,m), \tag{A16}$$

and sometimes SHCCS_m^k $\{\xi^i; i = 1,\cdots,k+m\}$ is written as SHCCS_m^k $\{\xi_2^i; \xi_1^j: i = 1,\cdots,k; j = 1,\cdots,m\}$ or simply SHCCS_m^k $\{\xi_2^i; \xi_1^j\}$.

Lemma A 2: For any SHCCS_m^k $\{\xi^i; i = 1,\cdots,k+m < N\} \equiv \mathrm{SHCCS}_m^k$ $\{\xi_2^i; \xi_1^j\}$, there exists a HCS $(\eta^\mu; \mu = 1,\cdots,N)$ such that

$$(\eta^1,\cdots,\eta^{k+m}) \equiv (\xi^1,\cdots,\xi^{k+m}), \tag{A17}$$

$$\{\xi_2^i,\eta^j\} = 0 \quad (i = 1,\cdots,k;\ \ j = k+m+1,\cdots,N), \tag{A18}$$

$$\{\xi^i,\eta^j\} = 0 \quad (i = 1,\cdots,k+m;\ \ j = k+2m+1,\cdots,N), \tag{A19}$$

and

$$\det \widetilde{\{\xi_1^i,\eta^j\}} \neq 0 \quad (i = 1,\cdots,m;\ \ j = k+m+1,\cdots,k+2m). \tag{A20}$$

Proof: From Lemma A1, there exists a HCS $(\bar\eta_0^i; \eta_0^j$ $i = 1,\cdots,k+m; j = k+m+1,\cdots N)$ such that

$$\{\xi^i,\eta_0^j\} = 0 \quad (i = 1,\cdots,k+m;\ \ j = k+m+1,\cdots,N) \tag{A21}$$

and

$$X_{\xi^i} = \sum_{\alpha=1}^{k+m} \{\xi^i,\bar\eta_0^\alpha\} \frac{\partial}{\partial \bar\eta_0^\alpha} \quad (i = 1,\cdots,k+m). \tag{A22}$$

From linear independence of $\{X_{\xi^i}; i = 1,\cdots,k+m\}$ we obtain

$$\det \widetilde{\{\xi^i,\bar\eta_0^j\}} \neq 0 \quad (i = 1,\cdots,k+m;\ \ j = 1,\cdots,k+m). \tag{A23}$$

From the definition of SHCCS_m^k $\{\xi^i\} \equiv \mathrm{SHCCS}_m^k$ $\{\xi_2^i, \xi_1^j\}$,

$$0 = \{\xi^i,\xi_1^j\}$$
$$= \sum_{\alpha=1}^{k+m} \{\xi^i,\bar\eta_0^\alpha\} \frac{\partial \xi_1^j}{\partial \bar\eta_0^\alpha} \quad (i = 1,\cdots,k+m;\ \ j = 1,\cdots,m). \tag{A24}$$

Since there exists an inverse of the matrix $\{\xi^i, \bar\eta_0^j\}$, $(i,j = 1,\cdots,k+m)$ from (A23), we have

$$\frac{\partial \xi_1^j}{\partial \bar\eta_0^\alpha} = 0 \quad (j = 1,\cdots,m;\quad \alpha = 1,\cdots,k+m), \tag{A25}$$

that is, ξ_1^j $(j = 1,\cdots,m)$ is expressed only in terms of η_0^α $(\alpha = k+m+1,\cdots,N)$ which is denoted as

$$\xi_1^j = \xi_1^j(\eta_0) \quad (j = 1,\cdots,m). \tag{A26}$$

From linear independence of $\{d\xi_1^j; j = 1,\ldots,m\}$, it follows

$$\mathrm{rank}\ \widetilde{\left(\frac{\partial \xi_1^j}{\partial \eta_0^\alpha}\right)} = m \quad (j = 1,\cdots,m;\quad \alpha = k+m+1,\cdots,N). \tag{A27}$$

From (A22), we have

$$\{\xi_2^i,\xi_2^j\} = \sum_{\alpha=1}^{k+m} \{\xi_2^i,\bar\eta_0^\alpha\} \frac{\partial \xi_2^j}{\partial \bar\eta_0^\alpha} \quad (i,j = 1,\cdots,k). \tag{A28}$$

From the definition of SHCCS_m^k $\{\xi_2^i; \xi_1^j\}$,

$$\mathrm{rank}\ \widetilde{\{\xi_2^i,\xi_2^j\}} = k \quad (i,j = 1,\cdots,k), \tag{A29}$$

and then from (A28) we obtain

$$\mathrm{rank}\ \widetilde{\left(\frac{\partial \xi_2^j}{\partial \bar\eta_0^\alpha}\right)} = k \quad (j = 1,\cdots,k;\quad \alpha = 1,\cdots,k+m). \tag{A30}$$

Now from (A27) and (A30), we may assume without loss of generality, that there exist a HCS $(\xi_2^1,\cdots,\xi_2^k; \xi_1^1,\cdots,\xi_1^m; \bar\eta_0^{k+1}, \cdots,\bar\eta_0^{k+m}; \eta_0^{k+2m+1},\cdots,\eta_0^N)$. Now again from Lemma A1, we obtain a HCS $(\xi_2^1,\cdots,\xi_2^k; \eta'^{k+1},\cdots,\eta'^N)$ such that

$$\{\xi_2^i,\eta'^j\} = 0 \quad (i = 1,\cdots,k;\quad j = k+1,\cdots N). \tag{A31}$$

Putting $\{\xi_1^1,\cdots,\xi_1^m; \eta_0^{k+2m+1},\cdots,\eta_0^N\} \equiv \{\zeta^1,\cdots,\zeta^{N-k-m}\}$, and using the above HCS $(\xi_2^i; \eta'^j: i = 1,\cdots,k; j = k+1,\cdots,N)$, we find

$$0 = \{\xi_2^i,\zeta^j\} = \sum_{\alpha=1}^{k} \{\xi_2^i,\xi_2^\alpha\} \frac{\partial \zeta^j}{\partial \xi_2^\alpha} \quad (i = 1,\cdots,k), \tag{A32}$$

and then

$$\frac{\partial \zeta^j}{\partial \xi_2^\alpha} = 0 \quad (\alpha = 1,\cdots,k;\quad j = 1,\cdots,N-k-m). \tag{A33}$$

Thus we have

$$\zeta^j = \zeta^j(\eta') \quad (j = 1,\cdots,N-k-m). \tag{A34}$$

From linear independence of $\{d\zeta^j; j = 1,\cdots,N-k-m\}$,

$$\mathrm{rank}\ \widetilde{\left(\frac{\partial \zeta^i}{\partial \eta'^j}\right)} = N-k-m$$
$$(i = 1,\cdots N-k-m;\quad j = k+1,\cdots N). \tag{A35}$$

Now from (A35) we may assume, without loss of generality, there exists a HCS $(\xi_2^1,\cdots,\xi_2^k; \xi_1^1,\cdots\xi_1^m; \eta'^{k+m+1},\cdots,\eta'^{k+2m}; \eta_0^{k+2m+1},\cdots,\eta_0^N)$ which is exactly the HCS $(\eta^\mu; \mu = 1,\cdots,N)$ in the Lemma A2 from (A8) and (A10). Q.E.D.

Here we quote Lemma 60 as Lemma A3.

Lemma A 3: For any SHCCS_m^k $\{\xi^i; i = 1,\cdots,k+m < N\}$, there exists a HCCS $(\xi^\mu; \mu = 1,\cdots,N)$ such that SHCCS_m^k $\{\xi^i\} \subset \mathrm{HCCS}(\xi^\mu)$.

Proof: For a given SHCCS_m^k $\{\xi_2^i; \xi_1^j\}$, we have a HCS $(\eta^\mu; \mu = 1,\cdots,N)$ in Lemma A2. If there exists an even coordinate η^j for some $j\in\{k+2m+1,\cdots,N\}$, satisfying (A19), then adding the η^j to the SHCCS^k_m $\{\xi^i; i=1,\cdots,k+m\}$, we obtain a SHCCS$^k_{m+1}$ $\{\xi^i; \eta^j: i=1,\cdots,k+m\}$ and thus we may assume, without loss of generality, that the above $\{\eta^j$; $j=k+2m+1,\cdots,N\}$ are all odd coordinates, which will be written as $\{\zeta^i; i=1,\cdots,r\}$ where $r=N-k-2m$, i.e., HCS $(\eta^\mu, \mu=1,\cdots,N)\equiv$HCS $(\xi^1_2,\cdots,\xi^k_2; \xi^1_1,\cdots,\xi^m_1; \eta^{k+m+1}, \cdots,\eta^{k+2m}; \zeta^1,\cdots,\zeta^r)$. Considering a matrix $\{\eta^\mu,\eta^\nu\}$ $(\mu,\nu=1,\cdots,N)$ in Lemma A2, and noting (A8) and (A10), we find

$$\det\widetilde{\{\zeta^\alpha,\zeta^\beta\}}\neq 0 \quad (\alpha,\beta=1,\cdots,r). \tag{A36}$$

We define a matrix $\{\zeta,\zeta\}$ symbolically as

$$\{\zeta,\zeta\}_{\alpha\beta}\equiv\{\zeta^\alpha,\zeta^\beta\} \tag{A37}$$

and then $\{\zeta,\zeta\}$ is a symmetric matrix whose components are all "anti-Hermitian" Z_2 even elements. Then multiplying each component by i, we define a matrix $i\{\zeta,\zeta\}$. Now $\widetilde{i\{\zeta,\zeta\}}$ is a real symmetric matrix, which can be diagonalized by some real orthogonal matrix written as $\tilde{\mathrm{O}}$ such that

$$\tilde{\mathrm{O}}\widetilde{i\{\zeta,\zeta\}}\tilde{\mathrm{O}}^t=d, \tag{A38}$$

and components of $\tilde{\mathrm{O}}$ are given by those of $\widetilde{i\{\zeta,\zeta\}}$, and we write symbolically

$$\tilde{\mathrm{O}}=\mathrm{O}(\widetilde{i\{\zeta\zeta\}}). \tag{A39}$$

Here we can define a matrix O unambiguously from $\tilde{\mathrm{O}}$ since components of $i\{\zeta,\zeta\}$ are all Z_2 even elements, as

$$\mathrm{O}\equiv\mathrm{O}(i\{\zeta,\zeta\}). \tag{A40}$$

Since $\widetilde{\mathrm{O}(i\{\zeta,\zeta\})}=\mathrm{O}(\widetilde{i\{\zeta,\zeta\}})$, the notations O and $\tilde{\mathrm{O}}$ are relevant. Let us define

$$\zeta'^\alpha\equiv\mathrm{O}_{\alpha\beta}\zeta^\beta \quad\text{or}\quad \zeta'=\mathrm{O}\zeta, \tag{A41}$$

and consider a new matrix $i\{\zeta',\zeta'\}$;

$$i\{\zeta',\zeta'\}=\mathrm{O}i\{\zeta,\zeta\}\mathrm{O}^t+(\text{proportional to }\zeta). \tag{A42}$$

Noting $\{\zeta^\alpha\}$ are all Z_2 odd, and $\widetilde{\zeta^\alpha}=0$, we have

$$\widetilde{i\{\zeta',\zeta'\}}=\tilde{\mathrm{O}}\widetilde{i\{\zeta,\zeta\}}\tilde{\mathrm{O}}^t=d, \tag{A43}$$

where

$$\det d\neq 0. \tag{A44}$$

From

$$\{\xi^i,\zeta^\alpha\}=0 \quad (i=1,\cdots,k+m;\quad \alpha=1,\cdots,r) \tag{A45}$$

and Jacobi identity (3.6), we find that

$$\{\xi^i,\{\zeta^\alpha,\zeta^\beta\}\}=0 \quad (i=1,\cdots,k+m;\quad \alpha,\beta=1,\cdots,r). \tag{A46}$$

Noting ζ'^1 being expressed only in terms of $\{\zeta,\zeta\}$ and ζ from (A40) and (A41), we have

$$\{\xi^i,\zeta'^1\}=\sum_{\alpha,\beta}\{\xi^i,\{\zeta^\alpha,\zeta^\beta\}\}\frac{\partial\zeta'^1}{\partial\{\zeta^\alpha,\zeta^\beta\}}+\sum_\alpha\{\xi^i,\zeta^\alpha\}\frac{\partial\zeta'^1}{\partial\zeta^\alpha}=0. \tag{A47}$$

From (A43) and (A44), it follows

$$\widetilde{i\{\zeta'^1,\zeta'^1\}}\gtrless 0, \tag{A48}$$

and it is well-defined that

$$\zeta''^1\equiv(\pm i\{\zeta'^1,\zeta'^1\})^{-1/2}\zeta'^1. \tag{A49}$$

Again from (A47) and Jacobi identity (3.6) we have

$$\{\xi^i,\{\zeta'^1,\zeta'^1\}\}=0 \quad (i=1,\cdots,k+m). \tag{A50}$$

From (A49) with (A47) and (A50), we see

$$\{\xi^i,\zeta''^1\}=0, \quad (i=1,\cdots,k+m). \tag{A51}$$

Noting, from Jacobi identity (3.6), that

$$\{\zeta'^1,\{\zeta'^1,\zeta'^1\}\}=0, \tag{A52}$$

and noting also $(\zeta'^1)^2=0$, we obtain

$$\{\zeta''^1,\zeta''^1\}=(\pm i\{\zeta'^1,\zeta'^1\})^{-(1/2)2}\{\zeta'^1,\zeta'^1\}=\mp i1_U. \tag{A53}$$

Here we note from (A51) and (A53) that we have obtained a SHCCS$^{k+1}_m$ $\{\xi^i_2; \zeta''^1; \xi^j_1; i=1,\cdots,k; j=1,\cdots,m\}$. Now by iterative process we may assume without loss of generality $r=N-k-2m=0$. Considering positions of zero components of the matrix $(\Omega^{-1})_{\mu\nu}=\{\eta^\mu,\eta^\nu\}$ $(\mu,\nu=1,\cdots,N)$ in Lemma A2 with $k+2m=N$, we can tell positions of zero components of the matrix Ω and obtain

$$\omega=\sum_\alpha dp^\alpha dq^\alpha+\sum_\beta i\frac{\epsilon_\beta}{2}ds^\beta ds^\beta+\sum_{j=1}^m d\xi^j_1\theta_j, \tag{A54}$$

where α and β summation give ξ_2-terms, and $\theta_j\in\Omega^1(U,A)$. From $d\omega=0$ we see $d(\Sigma^m_{j=1}d\xi^j_1\theta_j)=0$, and from Poincaré's Lemma, we put

$$\sum_{j=1}^m d\xi^j_1\theta_j\equiv d\omega'. \tag{A55}$$

From the form (A55), we derive the following form,

$$\omega'=\sum_{j=1}^m d\xi^j_1\left(A_j-\frac{\partial S}{\partial\xi^j_1}\right)+dS \tag{A56}$$

which gives the form as

$$\sum_{j=1}^m d\xi^j_1\theta_j=d\omega'=-\sum_{j=1}^m d\xi^j_1 d\left(A_j-\frac{\partial S}{\partial\xi^j_1}\right). \tag{A57}$$

Putting

$$\xi^j_1\equiv q'^j,\ A_j-\frac{\partial S}{\partial\xi^j_1}\equiv p'^j, \quad\text{if}\quad |\xi^j_1|=0, \tag{A58}$$

and

$$\xi^j_1=\sigma^j,\ A_j-\frac{\partial S}{\partial\xi^j_1}\equiv-\pi^j, \quad\text{if}\quad |\xi^j_1|=1, \tag{A59}$$

we find from $\omega^{\dagger t}=-\omega$ and from $|\omega|=0$ that without loss of generality, we may assume that p'^j be "Hermite" and π^j "anti-Hermite", and

$$|p'^j|=0 \quad\text{and}\quad |\pi^j|=1. \tag{A60}$$

$|\xi^j_1|=0$ term has an ordinary even coordinate form $dp'^j dq'^j$, and $|\xi^j_1|=1$ term can be rewritten as

$$d\sigma^j d\pi^j=\tfrac{1}{2}i(ds^j_+\,ds^j_+-ds^j_-\,ds^j_-), \tag{A61}$$

where

$$s^j_+=\frac{i}{\sqrt{2}}(\pi^j+i\sigma^j), \quad s^j_-=\frac{i}{\sqrt{2}}(\pi^j-i\sigma^j). \tag{A62}$$

All p'^j, q'^j, s^j_+, s^j_- are "Hermite". Q.E.D.

Finally let us note the above proof is valid for $k=m=0$, i.e., for SHCCS$^0_0$ $\{\xi^i\}$ which implies Lemma A3 is Darboux's theorem.

[1]R. Casalbuoni, Nuovo Cimento A **33**, 115, 389 (1976).
[2]Y. Ohnuki and T. Kashiwa, Prog. Theor. Phys. **60**, 548 (1978).
[3]Y. Nakano, Prog. Theor. Phys. **63**, 599 (1980).
[4]C. A. P. Calvao and C. Teitelboim, J. Math. Phys. **21**, 1863 (1980); and references cited therein.
[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*, Vol. 57 (Springer, Berlin, 1977).
[6]A consideration here on the relation between quantized systems and 'classical" systems does not logically exclude other possibilities.
[7]After completion of this work, the author has been informed that Frobenius' Theorem on graded manifolds has also been proved by R. Giachetti, R. Ragionieri, and R. Ricci (Univ. of Firenze, Italy).
[8]Prequantization and quantization of Kostant's geometric quantization program is beyond the scope of this paper, which will be discussed elsewhere.
[9]T. Maskawa and H. Nakajima, Prog. Theor. Phys. **56**, 1295 (1976).
[10]One may think of, for example, the Dirac Lagrangian which is real in terms of †.
[11]For more direct expression, $X\langle\omega|Y_n,\cdots,Y_1\rangle$ may be rewritten as $\langle\omega|Y_n,\cdots,Y_1\rangle\overleftarrow{X}$.
[12]H. Franders, *Differential Forms with Applications to the Physical Sciences* (Academic, New York, London, 1963).
[13]The notation $\widetilde{ds}$ is used for a different entity in Ref. 5.
[14]H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1950).

Cartan structures on Galilean manifolds: The chronoprojective geometry

G. Burdet,[a] C. Duval,[b] and M. Perrin[c]
Centre de Physique Théorique, CNRS, Marseille, France

(Received 20 July 1982; accepted for publication 23 September 1982)

A new geometry is constructed over Galilean manifolds expressing the compatibility requirement between the conformal equivalence notion of two Galilean structures and the projective equivalence notion of two affine connections. It is shown that it is the very geometry of the Newtonian cosmology (chronoprojective flatness is equivalent to isotropy of Newtonian cosmological models); moreover, it also explains various "accidental" symmetries in classical mechanics.

PACS numbers: 02.40.Dr, 02.40.Sf

INTRODUCTION

Let $P^2(V)$ denote the bundle of frames of second-order contact over a manifold V and let L be a Lie group with a closed subgroup L^0 such that $\dim(L/L^0) = \dim(V)$. An L/L^0 Cartan structure over the manifold V is a subbundle P of $P^2(V)$ with structure group L^0 and a canonically associated Cartan connection ω with respect to L; then it is in fact possible at each point of V to identify up to the first-order V with the homogeneous space L/L^0. Moreover, if we suppose that V is endowed with some (tensorial) structure α, it is then implicitly wished that L and L^0 have been artfully chosen to ensure that the homogeneous space L/L^0 is canonically endowed with a structure of the same kind as the original structure α.

Classical examples of Cartan structures are provided by the projective and conformal geometries.[1] In the projective geometry $L = \mathrm{PGl}(n,\mathbb{R})$ the projective general linear group and $L/L^0 = \mathbb{P}_n(\mathbb{R})$ is the projective space, a projective structure exists over any n-dimensional manifold V, i.e., V can be endowed with a class of symmetric connections which all determine the same set of geodesics. In the conformal geometry of Riemannian manifolds V_n, $L = O(n+1,1)$, and L/L^0 is a quadric diffeomorphic to the n-sphere S_n, i.e., V_n is endowed with a class of metric tensors which are proportional to each other. Note that because of the existence of a canonical connection (the Levi-Cività connection) over a Riemannian manifold there is a natural notion of conformal equivalence between torsionless symmetric linear connections corresponding to the above class of metrics.

A Newtonian space–time consists of a Galilean manifold (V_4, ψ, γ) and a chosen compatible symmetric linear connection Γ, a so-called Newtonian connection.[2] Obviously, a projective structure on $(V_4, \psi, \gamma, \Gamma)$ can be defined if space–time is endowed with a class of symmetric connections which give the same geodesics as Γ when parametrization is disregarded. On the other hand, it is also possible to introduce the notion of conformal equivalence on the Galilean manifold (V_4, ψ, γ). But, since a Newtonian connection is not canonically associated with the Galilean manifold this conformal equivalence notion does not yield a possible conformal structure over a Galilean manifold. Rather, one is naturally led to consider the compatibility between the projective structure and the conformal equivalence on the base manifold, and in this paper we show that the solution of this problem can be given by a particular Cartan structure we shall call the *chronoprojective geometry*.

It is worth noticing that it is possible to consider a manifold with compatible projective and conformal structure; this is known as a Weyl structure, and, for instance, a Weyl structure over a Lorentzian manifold is a model of Weyl space–time, which has been studied in Ref. 3. Although both approaches are similar, results are somewhat different because the Weyl geometry does not proceed from a Cartan structure.

This paper is organized as follows:

Section I consists of a brief recall of Galilean manifolds (V_4, ψ, γ) and the principal fiber bundle of Galilean frames $H(V_4)$.

In Sec. II, the conformal equivalence notion between two Galilean structures (V_4, ψ, γ) and (V_4, ψ', γ') is described, and the notion of conformal Galilean connection is introduced.

In Sec. III, the so-called chronoprojective group is defined and a chronoprojective Cartan structure P endowed with its natural torsionless Cartan connection is described.

In Sec. IV, the notion of admissible connection is defined, and the relations between two admissible conformal Galilean connections belonging to the same Galilean Cartan structure are studied and give rise to the chronoprojective equivalence notion. Finally, the chronoprojective transformations of a Galilean manifold are defined.

In Sec. V, the topology of the homogeneous space canonically associated with a chronoprojective Cartan structure, i.e., of the chronoprojective space–time, is described.

In Sec. VI, three equivalent definitions of infinitesimal chronoprojective Galilean transformations are given respectively on the base manifold V_4, on the bundle of first-order frames over V_4, and on the bundle of second-order frames.

In Sec. VII, the notion of chronoprojectively flat structure inherent to the above-developed geometry and the associated Weyl's curvature tensor are introduced. It is shown that the chronoprojective geometry is the very geometry of a Newtonian space–time and that any isotropic Newtonian space–time is chronoprojectively flat.

[a] Département de Mathématiques, Université de Dijon.
[b] Université d'Aix-Marseille II and Centre de Physique Théorique, CNRS, Marseille.
[c] On leave of absence from Département de Mathématiques, Université de Dijon.

Chronoprojective transformations are responsible for various "accidental" symmetries (Kepler problem, charged particle in a Dirac magnetic monopole field). As well those transformations which reproduce the direction of the presymplectic form of the evolution space of a massive test particle and project onto space–time transformations are in fact chronoprojective transformations. In particular, infinitesimal canonical transformations at most generate a 12-dimensional subalgebra of the chronoprojective algebra already known in the literature as the Schrödinger algebra. Originally, the Schrödinger group had been introduced as the largest group of space–time transformations which leave invariant the Schrödinger equation describing a free massive particle[4(a)] and the harmonic oscillator.[4(b)] An exhaustive classification of all compatible Hamiltonians has been given in Ref. 5(a) and the various types of mutual of two-, three-, and four-body interactions which are consistent with the Schrödinger invariance have been exhibited.[5(b)] The Schrödinger group is also a subgroup of the inhomogeneous two-sheeted symplectic group and has been studied from this point of view in Ref. 6, where an explicit expression of the Maslov index which characterizes metaplectic representations relevant in quantum mechanics is given.

This paper is a purely descriptive one; for conciseness proofs are generally omitted which are adapted from the ones used in projective and conformal geometries (see Ref. 1 and references therein), and italicized sentences take the place of propositions.

I. GALILEAN MANIFOLDS

A Galilean manifold is defined as a triple (V_4, ψ, γ), where V_4 is a four-dimensional C^∞-manifold, $\psi\in\mathscr{D}(V_4)$ is a differential 1-form of class one and γ is a positive semidefinite symmetric contravariant tensor field of degree 2 such that $\operatorname{Ker}\gamma$ is generated by ψ.

Let us now recall some properties we need for the following. Let us denote by S_x the associated space with ψ at $x\in V_4$:

$$S_x = \{w\in T_x(V_4) | w \lrcorner \psi = 0\} \quad \dim S_x = \operatorname{corank}\psi = 3. \quad (1.1)$$

Any element of S_x is called a *spacelike vector.*

The class of ψ being identified with the codimension of the intersection of S_x with the associated space with $d\psi$, the condition class $\psi = 1$ implies that $d\psi = 0$. Let S^*_x be a complementary subspace to the space generated by ψ in $T^*_x(V_4)$, for each x the inner product $\gamma_x : T^*_x(V_4)\times T^*_x(V_4)\to \mathbb{R}$ defines a linear isomorphism of S^*_x onto S_x as follows: with each $\alpha\in S^*_x(V_4)$ is associated $W^\alpha\in S_x$ defined by

$$W^\alpha \lrcorner \beta = \gamma_x(\alpha,\beta) \quad \forall\beta\in S^*_x. \quad (1.2)$$

One gets also an inner product, denoted by g_x, in S_x by setting

$$g_x(W^\alpha, W^\beta) = \gamma_x(\alpha,\beta). \quad (1.3)$$

If we denote by U the unique vector field such that

$$U\lrcorner\alpha = 0 \quad \forall\alpha\in S^*_x, \quad U\lrcorner\psi = 1. \quad (1.4)$$

U generates *timelike vectors*, and any vector $Y\in T_x(V_4)$ can be written as $Y = W_Y + \zeta_Y U$, where $W_Y\in S_x$ and ζ_Y is a function on V_4 linear with respect to Y. These properties are related to the fact that a Galilean manifold can be described as a bundle over a one-dimensional manifold, the projection being known as the universal time function. Obviously, g_x can be extended to an indefinite fiber metric $\tilde g_x$ over $T_x(V_4)$ by setting

$$\tilde g_x(Y,Y') = g_x(W_Y, W_{Y'}). \quad (1.5)$$

Let us introduce some notations which will be useful for the following: for $p\leqslant n$

$$\mathrm{O}^{n-p}(p) = \{g\in \mathrm{Gl}(n,\mathbb{R}) | gS^p(n)\,{}^t g = S^p(n),\ S^p(n) = \begin{pmatrix} \mathbb{1}_p & 0 \\ 0 & \mathrm{O}_{n-p}\end{pmatrix}\in M_n\Big\},$$

M_n being the $n\times n$ square matrices

$$\mathrm{O}_{n-p}(p) = \{g\in Gl(n,\mathbb{R}) | {}^t g S_p(n) g = S_p(n),\ S_p(n) = \begin{pmatrix} \mathrm{O}_{n-p} & 0 \\ 0 & \mathbb{1}_p\end{pmatrix}\in M_n\Big\},$$

$$\mathrm{CO}^{n-p}(p) = \{g\in \mathrm{Gl}(n,\mathbb{R}) | gS^p(n)\,{}^t g = \lambda^2 S^p(n),\quad \lambda\in\dot{\mathbb{R}} := \mathbb{R} - \{0\}\},$$

$$\mathrm{CO}_{n-p}(p) = \{g\in \mathrm{Gl}(n,\mathbb{R}) | {}^t g S_p(n) g = \lambda^2 S_p(n),\ \lambda\in\dot{\mathbb{R}}\}.$$

Note that $\mathrm{O}^0(p) = \mathrm{O}_0(p)$ is the usual orthogonal group denoted by $\mathrm{O}(n)$.

Let $P^1(V_4)$ be the bundle of linear frames over V_4; then $\mathbb{R}^4$ is the standard fibre of the tangent bundle $T(V_4)$ associated with $P^1(V_4)$. Since any element $r\in P^1(V_4)$ over $x\in V_4$ can be considered as a $1-1$ linear mapping of $\mathbb{R}^4$ onto $T_x(V_4)$: $y\mapsto ry = Y$, it is possible to associate with $\tilde g_x$ a bilinear form $(\ ,\ )_s$ on $\mathbb{R}^4$ defined by

$$(y, y')_s = (r^{-1}y, r^{-1}y') = \tilde g(Y,Y'). \quad (1.6)$$

This bilinear form can be written

$$(y,y')_s = {}^t y S^3(4) y', \quad (1.7)$$

where y is written as a column 4×1 matrix and ${}^t y$ as a row 1×4 matrix. Obviously, $(\ ,\ )_s$ is invariant by $\mathrm{O}^1(3)$, and it is easy to verify that this group is a semidirect product of the homogeneous Galilei group $\mathbb{R}^3 \circledS \mathrm{O}(3)$ by a dilatation, isomorphic to $\mathbb{R}^3\circledS(\mathrm{O}(3)\otimes\mathbb{R})$. The invariance of $(\ ,\ )_s$ by $\mathrm{O}^1(3)$ implies that relation (1.6) is independent of the choice of r modulo a right action of an element of $\mathrm{O}^1(3)$ as a subgroup of $\mathrm{Gl}(4),\mathbb{R}$ into $P^1(V_4)$, i.e., it leads to a reduction of $P^1(V_4)$ to a $\mathrm{O}^1(3)$-structure. Consequently, with any *Galilean manifold* (V_4, ψ, γ) *can be associated in a canonical way an* $\mathrm{O}^1(3)$*-structure* (of course, the converse is not true in general). The introduction of a connection in an $\mathrm{O}^1(3)$-structure over a Galilean manifold (V_4, ψ, γ) leads to the following properties on the base:

(i) $\nabla_Y\gamma = 0$,

(ii) $\nabla_Y\psi\wedge\psi = 0$

for all $Y\in T_x(V_4)$ and $x\in V_4$. But from ψ it is possible to construct the degenerate symmetric covariant tensor field of degree two $\Psi = \psi\otimes\psi$. Then, corresponding to the fiber metric Ψ_x, another degenerate bilinear form over $\mathbb{R}^4$ is induced, which can be written as

$$(y,y')_t = {}^t y S_1(4) y' \quad (1.8)$$

in compatibility with (1.7), due to the kernel condition linking γ and ψ. One checks that $(\ ,\ )_t$ is reproduced up to a positive scalar r_t^2 under the transposed action of $O^1(3)$, and from (ii) one gets in particular that the magnitude of timelike vectors is not preserved by parallel transfer.

Let us then consider the intersection of $O^1(3)$ with the group $O_3(1)$ which keeps $(\ ,\ )_t$ invariant, that is, the full homogeneous Galilei group H which is a double covering of $\mathbb{R}^3 \circledS O(3)$. Hence one is led to introduce a principal fiber bundle $H(V_4)$, which can be considered as an H-structure of degree 1, the *fiber bundle of Galilean frames*, corresponding to a cross section $\sigma_{\gamma,\psi}: V_4 \rightarrow P^1(V_4)/H$. On an H-structure it is possible to consider a so-called *Galilean connection* $\Gamma_{\gamma,\psi}$, which is the reduction of a linear connection with respect to which $\sigma_{\gamma,\psi}$ is parallel; then the parallel displacement of fibers of $P^1(V_4)/H$ preserves both fiber metrics $\tilde{g}$ and Ψ, so that, *with respect to a Galilean connection, γ and Ψ are parallel*:

$$\nabla\gamma = 0 \quad \text{and} \quad \nabla\Psi = 0. \tag{1.9}$$

Let $\theta = \{\theta^\mu \in \mathscr{D}(P^1(V_4)), \mu = 0,1,2,3\}$ be the canonical 1-form of $P^1(V_4)$ restricted to the bundle of Galilean frames and

$$\omega_\Gamma = \{\omega_\nu^\mu \in \mathscr{D}(H(V_4)),\ \mu,\nu \in [0,3] \,|\, \omega_\mu^0 = 0 \quad \forall\ \mu \in [0,3]$$

and

$$\omega_j^k + \omega_k^j = 0 \quad \text{for} \quad j,k \in [1,3]\}$$
$$= \{\varpi_0 = \{\omega_0^j\},\ \omega = \{\omega_j^k\}\}$$

be the connection 1-form on $H(V_4)$ defining a Galilean connection on V_4; the four 1-forms θ^μ and the six 1-forms ω_ν^μ define an absolute parallelism on $H(V_4)$. Then by using standard techniques one can show that it is not possible to define in a canonical way a unique Galilean connection on a given Galilean manifold. Hence it is clear that the geometry of a Galilean manifold is a less rigid structure than the geometry of a pseudo-Riemannian one, since, for instance, with a given Lorentzian manifold is associated a unique torsion-free connection (the Levi-Cività connection) on the fiber bundle of Lorentz frames while there does not exist a privileged (torsion-free) Galilean connection associated with a Galilean manifold.

II. CONFORMAL EQUIVALENCE ON A GALILEAN MANIFOLD

In Sec. I, the group $O^1(3) \simeq H \circledS \dot{\mathbb{R}}_t$, which keeps invariant the bilinear form $(\ ,\ )_s$ and reproduces up to a positive factor r_t^2 the bilinear form $(\ ,\ )_t$, has been introduced. On the other hand, we have seen that $(\ ,\ )_t$ is kept invariant by $O^1(3) \cap O_3(1)$. Moreover, the bilinear form $(\ ,\ )_t$ is also invariant under $CO^1(3) \cap O_3(1)$, which also reproduces up to a positive factor r_s^2 the bilinear form $(\ ,\ )_s$. So there are naturally introduced two dilatations which are gathered by defining the group $L^1 = CO^1(3) \cap CO_3(1) \simeq H \circledS (\dot{\mathbb{R}}_s \otimes \dot{\mathbb{R}}_t)$.

Now let us consider the cross section $\sigma_{\gamma,\psi}$ parallel with respect to a given Galilean connection and corresponding to a given embedding $H(V_4) \hookrightarrow P^1(V_4)$. Then $\sigma_{\gamma,\psi}$ composed with the natural mapping $\nu: P^1(V_4)/H \rightarrow P^1(V_4)/L^1$ defines a cross section of $P^1(V_4)/L^1$. Hence we are led to introduce a principal fiber bundle $L^1(V_4) \hookrightarrow P^1(V_4)$ with structure group L^1 the bundle of conformal Galilean frames. Moreover, the cross section $\nu \circ \sigma_{\gamma,\psi}$ is parallel with respect to the linear connection associated with the considered Galilean connection. Therefore, this linear connection is reducible to a connection in $L^1(V_4)$. Connections arising in this way will be called *conformal Galilean connections*. One can then consider $H(V_4)$ as a reduced bundle of $L^1(V_4)$ and $\Gamma_{\gamma,\psi}$ as a reduction of a conformal Galilean connection.

Now let us consider another triple (V_4, ψ', γ'). Corresponding to this geometrical structure on V_4 there is again a cross section $\sigma_{\gamma',\psi'}: V_4 \rightarrow P^1(V_4)/H$ and another embedding $H(V_4) \hookrightarrow P^1(V_4)$. In each fiber of $P^1(V_4)$ the two embeddings $H(V_4) \hookrightarrow P^1(V_4)$ define two H-orbits which are assumed to belong to an L^1-orbit. This entails that $\nu \circ \sigma_{\gamma,\psi}$ and $\nu \circ \sigma_{\gamma',\psi'}$ define the same $P^1(V_4)/L^1$ cross section, and, consequently, the *two Galilean structures are associated with two embeddings $H(V_4) \hookrightarrow P^1(V_4)$ corresponding to only one embedding of $L^1(V_4)$ into $P^1(V_4)$*. This situation occurs if the following notion of Galilean conformal equivalence is introduced: We say that the two triples (V_4, Ψ, γ) and (V_4, Ψ', γ') are *conformally equivalent* if $\gamma' = \rho_s \gamma$ and $\Psi' = \rho_t \Psi$, where ρ_s and ρ_t are positive suitably differentiable functions on V_4.

Let Π be the canonical projection $P^1(V_4) \rightarrow V_4$ $[L^1(V_4) \rightarrow V_4$, respectively$]$, then

$$\lambda(u) = \begin{pmatrix} \mathbb{1}_3 \otimes (\rho_s \circ \Pi)(u)^{1/2} & \\ & (\rho_t \circ \Pi)(u)^{-1/2} \end{pmatrix}$$

identified with an element of $\dot{\mathbb{R}}_s \otimes \dot{\mathbb{R}}_t$ characterizes the elements of L^1 which, on each fiber of $P^1(V_4)[L^1(V_4)$, respectively$]$, relate the two concerned H-orbits.

But due to the lack of a uniquely defined torsion-free Galilean connection, *there does not exist a one-to-one correspondence between the conformal equivalence classes of Galilean structures on V_4 and the embeddings $L^1(V_4) \hookrightarrow P^1(V_4)$ that is with the L^1-structures*, while such a one-to-one correspondence exists in the conformal geometry over pseudo-Riemannian manifolds. Merely a conformal equivalence class of Galilean structures corresponds to various embeddings into principal L^1-bundles on V_4, and we have to consider one L^1-bundle and its embedding into $P^1(V_4)$ only if a conformal Galilean connection is arbitrarily chosen. With respect to this conformal Galilean connection one has $\nabla\psi = \chi_t \otimes \psi$ and $\nabla\gamma = \chi_s \otimes \gamma$, where χ_s and χ_t are 1-forms on V_4; then, if the triple (V_4, ψ', γ') is conformally equivalent to (V_4, ψ, γ), one gets

$$\nabla\psi' = \chi_t' \otimes \psi \quad \text{with} \quad \chi_t' = \chi_t + \frac{1}{2}\frac{d\rho_t}{\rho_t}, \tag{2.1a}$$

$$\nabla\gamma' = \chi_s' \otimes \gamma \quad \text{with} \quad \chi_s' = \chi_s + \frac{d\rho_s}{\rho_s}. \tag{2.1b}$$

Now let φ be a diffeomorphism of a Galilean manifold (V_4, ψ, γ) onto itself, $\tilde{\varphi}$ the corresponding linear isomorphism of the tensor algebra $T(\varphi^{-1}(x))$ onto $T(x)$ and $\bar{\varphi}$ the induced automorphism of $P^1(V_4)$. The transformations φ of V_4 such that $\tilde{\varphi}\gamma = \rho_s \gamma$ and $\tilde{\varphi}\psi = \rho_t \psi$ are called the *conformal Galilean transformations* of V_4. In general such a transformation

$\bar{\varphi}$ is not an affine transformation; moreover, $\bar{\varphi}$ neither leaves invariant any linear connection nor maps any L^1-structure onto itself.

Several examples are known showing that the above considered set of transformation φ of V_4 is an infinite transformation group. Note that infinite transformation groups already appear when we look for *Galilean automorphisms* (accelerated frames) $\tilde{\varphi}\gamma = \gamma$ and $\tilde{\varphi}\Psi = \Psi$.

Hence it is pointed out once more that a Galilean structure is a less rigid geometrical object than a Lorentzian structure since conformal transformations of a Lorentzian manifold induce automorphisms of a CO(3,1)-structure and generate a transformation group of dimension at most equal to 15, while, according to the foregoing, the transformations of the basis which reproduce the geometrical Galilean structure generate in general an infinite-dimensional transformation group.

III. CHRONOPROJECTIVE CARTAN STRUCTURE ON A GALILEAN MANIFOLD

Let us consider the Lie group $O^2(3)$ which is from now on named the *chronoprojective group* (the reasons for this denomination will be clarified in Sec. VI). $O^2(3)$ is a 13-dimensional Lie group which can be written as $O^2(3) \approx (\mathbb{R}^3 \otimes \mathbb{R}^3) \circledS (O(3) \otimes Gl(2,\mathbb{R}))$.

Let us also consider the subgroup L^0 generated by the elements of $O^2(3)$, which admit ${}^t(0,0,0,0,1)$ as eigenvector; L^0 is nine-dimensional and can be written as $\mathbb{R}^3 \circledS (O(3) \otimes \mathbb{R} \otimes S_2)$ where S_2 is the two-dimensional solvable group. Let $\mathscr{O}^2(3)$ be the Lie algebra of $O^2(3)$ which as a vector space can be decomposed as $\mathscr{O}^2(3) = \ell^0 + a$, where ℓ^0 is the Lie algebra of L^0 and a is a four-dimensional Abelian Lie algebra. The subalgebra ℓ^0 is not reductive into $\mathscr{O}^2(3)$. $\{[\mathscr{O}^2(3),a]$ is not contained into a.$\}$ $\mathscr{O}^2(3)$ *is* but a *2-graded Lie algebra*, i.e., it can be written as

$$\mathscr{O}^2(3) = g_{-2} + g_{-1} + g_0 + g_1 + g_2 \tag{3.1}$$

such that $[g_p, g_q] \subset g_{p+q}$ with $g_p = 0$ when $|p| > 2$ and there exists a unique (up to a conjugation) element D in g_0 such that $[D, g_p] = p g_p$. In fact, $g_2 = g_{-2} = \mathbb{R}$ and $[g_2, g_{-2}]$ is proportional to D.

Moreover, $g_1 = g_{-1} = \mathbb{R}^3$ and $[g_1, g_{-1}] = 0$, $g_0 = O(3) \otimes \mathbb{R}^2$. Note that $\ell^0 = g_0 + g_1 + g_2$ and $a = g_{-2} + g_{-1}$. The class of the identity $e \in O^2(3)$ in the connected homogeneous space $O^2(3)/L^0$ will be called the *origin* of $O^2(3)/L^0$ and denoted by o. There is a natural representation ρ, usually called the *linear isotropy representation of* L^0 on the tangent space of $O^2(3)/L^0$ at the origin o. But $T_o(O^2(2)/L^0) = \mathscr{O}^2(3)/\ell^0 = a$ from the above decomposition; then the linear isotropy representation is defined by $\rho(g)x = \mathrm{Ad}(g)x (\mathrm{mod}\,\ell^0)$ for $g \in L^0$ and $x \in a$, where $\rho(g) \in Gl(4,\mathbb{R})$.

This representation is not faithful and possesses a one-dimensional kernel, $u_0 \in \mathbb{R}$ parametrizing an element of this kernel. ρ is faithful on a subgroup $\mathbb{R}^3 \circledS (CO(3) \otimes \mathbb{R})$ only, which is called the *linear isotropy group* of the homogeneous space. This group is isomorphic to the group L^1 introduced in Sec. II, and, in fact, L^0 can be written as a semidirect product $L^0 = \mathbb{R} \circledS L^1$ and there is a natural injective homomorphism $k: L^1 \to L^0$.

Let $G^1(4)$ and $G^2(4)$ be the structure groups of the fiber bundles of first- and second-order frames, respectively. There is a natural homomorphism $G^2(4) \to G^1(4)$, the kernel of which is denoted $N^2(4)$. One has the exact sequence

$$0 \to N^2(4) \to G^2(4) \to G^1(4) \to 1.$$

L^0 can be isomorphically embedded into $G^2(4)$ and there is a natural coordinate system in $G^2(4)$ given by $\{s^\lambda_\mu, s^\lambda_{\mu\nu} = s^\lambda_{\nu\mu}, \mu, \nu = 0,1,2,3\}$, where $\{s^\lambda_\mu\}$ is a natural coordinate system in $G^1(4) \approx Gl(4,\mathbb{R})$ so that the natural embedding $G^1(4) \hookrightarrow G^2(4)$ is given by $\{s^\lambda_\mu, 0\}$. In the present case the isomorphic embedding of L^0 into $G^2(4)$ is given by

$$s^\lambda_\mu = (\rho(g))^\lambda_\mu, \quad s^\lambda_{\mu\nu} = -\tfrac{1}{2}u_0[\delta^0_\mu(\rho(g))^\lambda_\nu + \delta^0_\nu(\rho(g))^\lambda_\mu] \tag{3.2}$$

and the injective homomorphism $k: L^1 \to L^0$ is realized by $\{s^\lambda_\mu, 0\}$ with the above-defined s^λ_μ.

If P denotes an L^0-structure of degree 2 on V_4, then $P/\ker\rho$ is a subbundle Q of $P^1(V_4)$ with $L^1 \subset G^1(4)$ as structure group. Conversely, let Q be an L^1-structure of degree 1 and k an injective homomorphism from L^1 into L^0. L^1 acts on the left on L^0 as follows $(a,n) \mapsto k(a)m$, $a \in L^1$, $m \in L^0$. So we can introduce the associated fiber bundle $(Q \times L^0)/L^1 = Q_k$, which is a principal fiber bundle, *the k-extension* of Q, with respect to the right action of L^0 over Q_k given by

$$((q\cdot m), m') \mapsto q\cdot(mm'), \quad q \in Q, \quad m, m' \in L^0,$$

$q\cdot m$ denoting the class of (q,m) into Q_k. If k is chosen as above then the k-extension of Q will be an L^0-structure of degree 2.

Now let us recall that a Cartan connection[1] with respect to $O^2(3)$ in an L^0-structure P is a 1-form w_c on P with values in the Lie algebra $\mathscr{O}^2(3)$ satisfying the following conditions:

(a) $w_c(A^*) = A$ for every $A \in \ell^0$, (3.3a)

where A^* is the fundamental vector field corresponding to A.

(b) $(Rg)^* w_c = \mathrm{ad}(g^{-1}) w_c$ for every element $g \in L^0$ (3.3b)

where ad is the adjoint action of L^0 on $\mathscr{O}^2(3)$.

(c) $w_c(X) \neq 0$ for every nonzero vector X of P. (3.3c)

Let (θ, Θ) be the canonical form on $P^2(V_4)$, where $\theta = \{\theta^\mu \in \mathscr{D}(P^2(V_4)), \mu = 0,1,2,3\}$ is an $\mathbb{R}^4$-valued one-form and Θ is a $gl(4,\mathbb{R})$-valued 1-form.

Then let $\{w^\lambda_{0'}, w^j_\lambda, w^0_0, w^{0'}_{0'}, j = 1,2,3, \lambda = 0,1,2,3\}$ be the $(g_{-2} + g_{-1} + \ell^1)$-valued 1-form obtained by restricting to P the canonical form (θ, Θ) on $P^2(V_4)$, it can be called *the canonical form of P*. [The above-used couples of indices are related to the graduation of $O^2(3)$ as follows: $w^0_{0'}$ is g_{-2}-valued, $\{w^j_{0'}\}$ is g_{-1}-valued, $\{w^k_j, w^0_0, w^{0'}_{0'}\}$ is g_0-valued and $\{w^j_{0'}\}$ is g_1-valued.] One can show that *this set of 12 1-forms* whose values in each point are linearly independent *can be supplemented in a unique way by a g_2-valued 1-form $w^{0'}_0$ to give rise to an $\mathscr{O}^2(3)$-valued 1-form defining a natural Cartan connection w_c for P. The natural Cartan connection is torsionless, i.e., the g_{-2} and g_{-1} components of its curvature 2-*

form Ω vanish, and its curvature satisfies the following conditions:

The component of Ω which takes its value on the dilatation D vanishes, i.e., $\Omega_D = 0$; (3.4a)

$$\sum_{j,k,l} \epsilon_{jkl} \Omega^j_0 \wedge w^k_{0'} \wedge w^l_{0'} = 0, \quad j,k,l \in [1,2,3]; \tag{3.4b}$$

$$\sum_{j,k} \epsilon_{jkl}\, \Omega^j_0 \wedge w^k_{0'} \wedge w^0_{0'} = 0, \quad \forall l, j,k,l \in [1,2,3]; \tag{3.4c}$$

$$\sum_{j,k} \epsilon_{jkl} \Omega^{0'}_{0'} \wedge w^j_{0'} \wedge w^k_{0'} = 0 \quad \forall l,\ j,k,l \in [1,2,3], \tag{3.4d}$$

where ϵ_{jkl} denotes the three-index permutation symbol.

This is proved by standard techniques which have been described for the projective and conformal geometries in Ref. 1 and for the contact geometry in Ref. 7; the appropriate construction for the present case will be given elsewhere in a more technical paper.[8]

Equipped with this unique Cartan connection, the L^0-structure P is parallelizable. Then the basic theorem (Ref. 1, p. 15) following which the group $\mathfrak{A}$ of automorphisms of a parallelizable manifold M is a Lie transformation group such that $\dim\mathfrak{A} \leqslant \dim M$, can be applied and ensures that *the group of automorphisms of P which preserves the Cartan connection is a Lie group with dimension at most equal to $\dim P = 13$.* Such an L^0-structure endowed with its natural Cartan connection will be called a *chronoprojective Cartan structure*.

IV. L^0-EQUIVALENCE OF TWO ADMISSIBLE CONFORMAL GALILEAN CONNECTIONS CHRONOPROJECTIVE TRANSFORMATIONS

Now let us consider the conformal Galilean connections over V_4 introduced in Sec. II. Among them the torsion-free connections are in one-to-one correspondence with the cross sections $\Gamma: V_4 \to P^2(V_4)/L^1$. Composed with the natural mapping $\mu: P^2(V_4)/L^1 \to P^2(V_4)/L^0$, these connections give rise to sections $\mu \circ \Gamma: V_4 \to P^2(V_4)/L^0$, i.e., to reductions of $P^2(V_4)$. In other words, *every torsion-free conformal Galilean connection Γ defines a reduction of the structure group of $P^2(V_4)$ to L^0* and induces an isomorphism $\mathscr{K}$ of $L^1(V_4)$ into $P^2(V_4)$. Then *Γ will be said to be admissible if it belongs to a Galilean Cartan structure P*, that is to say, if it induces P in the above-described manner, i.e., if the corresponding subbundle Q of relations (3.4) defining the unique natural Cartan connection. Let us denote by $\mathscr{W}$ the ℓ^1-valued curvature form of the conformal Galilean connection Γ and by θ the $\mathbb{R}^4$-valued canonical form of $P^1(V_4)$ restricted to Q and viewed as a $(g_{-2} + g_{-1})$-valued 1-form.
form of the conformal Galilean connection Γ and by θ the $\mathbb{R}^4$-valued canonical form of $P^1(V_4)$ restricted to Q and viewed as a $(g_{-2} + g_{-1})$-valued 1-form.

For convenience of notation, we shall set $\Omega_w = \{\Omega^j_0, \Omega^j_k - \delta^j_k\, \Omega^{0'}_{0'}, \Omega_D\}$ and we can write

$$\mathscr{K}^* \Omega_w = \mathscr{W} - \tfrac{1}{2}\mathscr{K}^*([w^{0'}_0, \theta]) + \theta^0 \mathbb{1}_4 \wedge \mathscr{K}^*(w^{0'}_0) \tag{4.1}$$

so that from relation (3.4), one gets

$$\mathscr{W}^0_0 = \mathscr{K}^*(w^{0'}_0) \wedge \theta^0, \tag{4.2a}$$

$$\sum_{j,k,l} \epsilon_{jkl} \mathscr{W}^j_0 \wedge \theta^k \wedge \theta^l = \tfrac{1}{2}\mathscr{K}^*(w^{0'}_0) \wedge \sum_{j,k,l} \epsilon_{jkl} \theta^j \wedge \theta^k \wedge \theta^l, \tag{4.2b}$$

$$\sum_{j,k} \epsilon_{jkl} \mathscr{W}^j_0 \wedge \theta^k \wedge \theta^0 = \tfrac{1}{2}\mathscr{K}^*(w^{0'}_0) \wedge \sum_{j,k} \epsilon_{jkl} \theta^j \wedge \theta^k \wedge \theta^0, \tag{4.2c}$$

$$\sum_{j,k} \epsilon_{jkl} \Big(\sum_{i=1}^{3} \mathscr{W}^i_i \Big) \wedge \theta^j \wedge \theta^k = \tfrac{1}{2}\mathscr{K}^*(w^{0'}_0) \wedge \sum_{j,k} \epsilon_{jkl} \theta^j \wedge \theta^k \wedge \theta^0. \tag{4.2d}$$

This first set of constraints has to be supplemented by another coming from the pullback to Q of relation (3.3b). It is expressed by a condition on the Ricci curvature tensor of Γ which provides a tenuous link between Γ and the structure of the base Galilean manifold (V_4, ψ, γ); explicitly one gets

$$\mathrm{Ric} = \varphi \otimes \psi - 4\psi \otimes \varphi, \tag{4.3}$$

where φ is an arbitrary covariant tensor field of degree one.

Finally a conformal Galilean connection Γ is admissible if its curvature form satisfies (4.2a, b, c, d) and its Ricci curvature tensor can be written as in (4.3).

Let us note that this notion of admissibility can be applied to connections of any subbundle of Q in particular to Galilean connections since any Galilean connection maps to a well-defined connection Γ by the homomorphism which defines the embedding of $H(V_4)$ into the L^1-structure Q. One verifies that admissibility conditions for a Galilean connection reduce to

$$\mathrm{Ric} = \rho \Psi = \rho \psi \otimes \psi, \tag{4.4}$$

where ρ is an arbitrary function on V_4.

In the following we shall speak of admissible connection without specifying the concerned subbundle of $L^1(V_4)$ if it is unnecessary, and by "*abus de langage*" we shall speak of *admissible Galilean manifold* to nominate a quadruple $(V_4, \psi, \gamma, \Gamma)$, where Γ is an admissible Galilean connection.

Two admissible torsion-free connections are said to be *L^0-equivalent* if they belong the the same Galilean Cartan structure P. It can be shown that *two admissible torsion-free connections defined by the $(g_0 + g_1)$-valued 1-forms ω* and ω' *are L^0-equivalent if there exists a g_2-valued function ξ on V_4 such that*

$$\omega' - \omega = \mathscr{K}^*([\theta, \xi \circ \Pi]), \tag{4.5}$$

where Π is the projection $\Pi: P \to V_4$.

Moreover, we have to ensure the compatibility between the L^0-equivalence of two connections Γ and Γ' and the conformal equivalence of (V_4, ψ, γ) and (V_4, ψ', γ'), i.e., to express that the covariant derivative with respect to the connection Γ' reproduces ψ' and γ'. Locally the L^0-equivalence is expressed by a projective equivalence of Christoffel's symbols which can be written

$$\Gamma'^{\lambda}_{\mu\nu} - \Gamma^{\lambda}_{\mu\nu} = (\delta^{\lambda}_{\mu} \psi_\nu + \delta^{\lambda}_{\nu} \psi_\mu)\eta, \tag{4.6}$$

where η is an $\mathbb{R}$-valued function on V_4, then *the compatibility is expressed by the following relations*:

$$\chi_t' = \chi_t + \frac{1}{2}\frac{d\rho_t}{\rho_t} - 2\eta\psi, \tag{4.7a}$$

$$\chi_s' = \chi_s + \frac{d\rho_s}{\rho_s} + 2\eta\psi, \tag{4.7b}$$

which lead to

$$d\,(\rho_s\cdot\rho_t^{1/2}) = \rho_s\cdot\rho_t^{1/2}\ (\chi_s' - \chi_s + \chi_t' - \chi_t). \tag{4.7c}$$

Therefore, we can say that the *two quadruples* (V_4,ψ,γ,Γ) *and* $(V_4,\psi',\gamma',\Gamma')$ *are chronoprojectively equivalent if* (V_4,ψ,γ) *and* (V_4,ψ',γ') *are conformally equivalent,* Γ *and* Γ' *are* L^0*-equivalent, and relations* (4.7*a*,*b*,*c*) *are satisfied.*

From the above relations (4.7) it is clear that, in general, the L^0-equivalence of two conformal Galilean connections does not lead to the situation described in Sec. II, that is, to consider only one embedding of $L^1(V_4)$ into $P^1(V_4)$ (which is expressed by (2.1a) and (2.1b). So it is significant to speak for example of the L^0-equivalence of two conformal Galilean connections. Nevertheless, it is worth noticing that relations (4.7) are greatly simplified in the case of the L^0-equivalence of two Galilean connections for which they reduce to the following relations

$$\frac{d\rho_t}{\rho_t} = 4\eta\psi, \tag{4.8a}$$

$$\frac{d\rho_s}{\rho_s} = -2\eta\psi, \tag{4.8b}$$

with

$$\rho_s\cdot\rho_t^{1/2} = \text{constant function on } V_4. \tag{4.8c}$$

This expresses the fact that two L^0-equivalent Galilean connections are mapped into the same conformal Galilean connection in $L^1(V_4)$ under the corresponding homomorphisms of $H(V_4)$ into $L^1(V_4)$.

A (local) diffeomorphism φ of V_4 induces a local isomorphism $\bar{\bar{\varphi}}$ of the bundle $P^2(V_4)$ if $\bar{\bar{\varphi}}$ sends a chronoprojective Cartan structure P into itself then φ will be called a (local) *chronoprojective Galilean transformation of* V_4. From this definition it is clear that a *transformation of* V_4 *is a chronoprojective Galilean transformation if and only if it is a conformal Galilean transformation which transforms an admissible conformal Galilean connection into an* L^0*-equivalent one such that* (4.7*a*, *b*, *c*) *are satisfied.*

V. THE HOMOGENEOUS SPACE $O^2(3)/L^0$: THE CHRONOPROJECTIVE SPACE–TIME

Let $M = O^2(3)/L^0$; it is easy to see that $M = (\mathbb{R}^3\times(\mathbb{R}^2 - \{0\}))/\mathbb{R}$. Taking into account that $\mathbb{R}^2 - \{0\}$ can be considered as a nontrivial principal $\dot{\mathbb{R}}$-bundle over the one-dimensional projective space over $\mathbb{R}$, i.e., the unit circle S^1, M can be described as a vector bundle of standard fiber $\mathbb{R}^3$ over S^1 associated with $\mathbb{R}^2 - \{0\}$, so that M is a good candidate to be a Galilean manifold with S^1 as time axis. Otherwise, $\mathbb{R} - \{0\}$ is also a trivial $\mathbb{R}^+$-principal bundle, so that M can be equivalently written as $(\mathbb{R}^3\times S^1)/\mathbb{Z}^2$ and appears as generalized Möbius space. Moreover, it can be shown that M is also an homogeneous coset space of the group $G = (\mathbb{R}^3\times\mathbb{R}^3)\circledS(O(3)\otimes SO(2))$ a ten-dimensional subgroup of $O^2(3)$. One gets $M = G/H$, where H denotes the full homogeneous Galilei group. Hence G can be considered as the fiber bundle of Galilean frames over M.

By considering the Maurer–Cartan form over G it is easy to verify that M is endowed with a Galilean structure and possesses a torsionless Newtonian connection Γ^0 with a null spatial component curvature.

In fact, (M,ψ,γ,Γ^0) is an exact solution of the Newton vacuum field equations with a unit reduced cosmological constant and is the model of a chronoprojectively flat Newtonian space–time (see Sec. VII) endowed with a globally nontrivial structure over the temporal projective line S^1; this justifies the denomination of this type of geometry over Galilean manifolds.

Let us consider $O^2(3)$ as a principal L^0-bundle over M. $O^2(3)$ can be identified with a chronoprojective structure in the following manner: on the one hand, each $f\in O^2(3)$ is a transformation of $O^2(3)/L^0$; on the other hand, any neighborhood of the origin o of $O^2(3)/L^0$ can be identified with a neighborhood of O in $\mathbb{R}^4$ in a natural way. Then any 2-jet of f can be considered as a 2-frame of $O^2(3)/L^0$ and the set $f(o)$ of all 2-frames obtained this way defines a chronoprojective structure which can be identified with $O^2(3)$. The Maurer–Cartan form of $O^2(3)$ becomes the natural Cartan connection of this chronoprojective structure over M, so it has no curvature (and no torsion); moreover, it is clear that $O^2(3)$ is the group of chronoprojective transformations of this chronoprojective structure over the chronoprojective space–time.

VI. INFINITESIMAL CHRONOPROJECTIVE GALILEAN TRANSFORMATIONS

In the previous sections chronoprojective Galilean transformations have been introduced as (local) diffeomorphisms φ of V_4 such that φ induces in a natural manner an automorphism $\bar{\bar{\varphi}}$ of the chronoprojective structure P. Equivalently (and by construction) they can also be defined as the conformal Galilean transformations of V_4, which prolonged to a mapping $\bar{\varphi}$ of $P^1(V_4)$ into $P^1(V_4)$ transform an admissible conformal Galilean connection into an L^0-equivalent one such that (4.7a,b,c) are satisfied.

Every vector field X on V_4 generates a one-parameter local group of (local) transformations.[9] This local group of transformations prolonged to $P^1(V_4)$ and to $P^2(V_4)$ induces a vector field X' on $P^1(V_4)$ and a vector field X'' on $P^2(V_4)$, then we call X an *infinitesimal chronoprojective Galilean transformation* if the local one-parameter group generated by X in a neighborhood of each point of V_4 consists of local chronoprojective Galilean transformations.

It can be shown that *the following conditions are mutually equivalent*:

(*A*) *On* $P^2(V_4)$:

(*i*) X'' *is tangent to P at every point of P*;

(*ii*) *the Lie derivative with respect to* X'' *of the natural Cartan connection is zero*: $L_{X''}\,w_c = 0$;

(*iii*) *the Lie derivative with respect to* X'' *of the standard horizontal vector field B corresponding to* ξ ($\xi\in\mathbb{R}^4$) *is zero for every* ξ.

The standard horizontal vector field associated with each element $\xi = (\xi^0,\xi^1,\xi^2,\xi^3)$ of $\mathbb{R}^4$ is the unique vector field

B on P with the following properties:

$$w_{0'}^{\lambda}(B) = \xi^{\lambda}, \quad \lambda = 0,1,2,3,$$
$$w_j^k(B) = w_0^0(B) = w_{0'}^{0'}(B) = w_{\lambda}^{o'}(B) = 0, \quad j,k = 1,2,3. \tag{6.1}$$

(B) *On* $P^1(V_4)$: *(iv) The Lie derivatives with respect to* X' *of the 1-form* $\omega_\Gamma = \{\overline{\omega}_0, \omega\}$ *of a Galilean connection which induces P and of the canonical 1-form* $\theta = \{\theta^0, \overline{\theta}\} = \{\theta^0, \theta^j, j = 1,2,3\}$ *of* $P^1(V_4)$ *reduced to the* L^1*-structure Q satisfy the following properties*:

$$L_{X'}\overline{\omega}_0 = -\tfrac{1}{2}\,((\varepsilon_t + \varepsilon_s)\circ\Pi)\overline{\omega}_0 + \zeta\overline{\theta}, \tag{6.2a}$$

$$L_{X'}\omega = 0, \tag{6.2b}$$

$$L_{X'}\overline{\theta} = -\tfrac{1}{2}\,(\varepsilon_s\circ\Pi)\overline{\theta}, \tag{6.2c}$$

$$L_{X'}\theta^0 = \tfrac{1}{2}\,(\varepsilon_t\circ\Pi)\theta^0, \tag{6.2d}$$

where ζ *is a function on Q and* ε_s *and* ε_t *are two functions on* V_4 *which are related by*

$$\Pi^*(d\varepsilon_s) = -\tfrac{1}{2}\Pi^*(d\varepsilon_t) = -\zeta\theta^0. \tag{6.3}$$

[In particular, note that X' is tangent to the L^1-structure at every point of $L^1(V_4) \hookrightarrow P^1(V_4)$.]

(C) *On* V_4: *(v) The Lie derivatives with respect to X of the tensors which define the Galilean manifold satisfy the following properties*:

$$L_X\Psi = \varepsilon_t\Psi \tag{6.4a}$$

(or equivalently $L_X\psi = \frac{1}{2}\varepsilon_t\psi$ which implies $d\varepsilon_t \wedge \psi = 0$),

$$L_X\gamma = \varepsilon_s\gamma, \tag{6.4b}$$

supplemented by

$$K(X,Y) = \eta\otimes Y + \eta(Y)\mathbb{1} \quad \text{for all vector fields } Y \text{ on } V_4, \tag{6.4c}$$

where $K(X,Y) = R(X,Y) - \nabla_Y A_X$, where A_X is the derivation defined by $L_X - \nabla_X$ (∇ denotes the covariant derivation) and R is the curvature tensor of an admissible connection which induces P; $\mathbb{1}$ denotes the identity tensor of type (1,1), while η is a covariant tensor field of type (0,1).

From (a) and (b) we deduce that

$$U(\varepsilon_s) = -\tfrac{1}{2}U(\varepsilon_t) \tag{6.5}$$

and

$$W_Y(\varepsilon_s) = 0,$$
$$W_Y(\varepsilon_t) = 0, \quad \text{for any vector field } Y, \tag{6.6}$$

where W_Y denotes the spatial component of Y and U is defined in (1.4). Moreover, one gets

$$U(\zeta_X) = \tfrac{1}{2}\varepsilon_t$$

where ζ is the timelike component of X.

From (c) we deduce that

$$\eta = U(\varepsilon_s)\psi. \tag{6.7}$$

Note that locally (c) is nothing else that the infinitesimal form of relation (4.6) which expresses the L^0-equivalence.

By determining all the vector fields X'' satisfying (i), (ii), and (iii), all the vector fields X' satisfying (iv) and all the vector fields satisfying (v) one gets three different realizations of the Lie algebra aut(P,w_c) of the automorphisms of the considered P structure. By the basic theorem recalled in Sec. III, one knows that the dimension of this Lie algebra is at most equal to 13. Moreover, it is easy to show that if the Lie algebra aut(P,w_c) is of dimension 13, the natural Cartan connection has vanishing curvature and torsion (the converse being in general not true).

VII. FLAT CHRONOPROJECTIVE GALILEAN STRUCTURES. APPLICATIONS TO COSMOLOGY AND MECHANICS

Let P and P' be chronoprojective Galilean structures on Galilean manifolds V and V', respectively. *A diffeomorphism* f: $V_4 \to V_4'$ *is called chronoprojective Galilean (with respect to P and P') iff f is prolonged to a mapping* $P^2(V_4)$ *onto* $P^2(V_4')$ *maps P onto P'*; hence it is a bundle isomorphism. By the very definition of P and P', the diffeomorphism f can also be called *chronoprojective Galilean with respect to the admissible connections* Γ *and* Γ', *which induce P and P', respectively.* A chronoprojective Galilean structure P is called *flat* if, for each point of V_4, there exists a neighborhood $\mathscr{U}$ and a chronoprojective Galilean diffeomorphism of $\mathscr{U}$ onto an open subset of the chronoprojective space–time M, M being taken as the standard model of chronoprojectively flat space–time by virtue of its properties given in Sec. V.

From usual argument one can show that *a chronoprojective structue P on a Galilean manifold* V_4 *is flat if and only if its natural Cartan connection has vanishing curvature.*

The Weyl's curvature tensor is defined by using the subset Ω_w (cf. Sec. IV) of the components of the natural Cartan connection curvature of P which generates a 2-form with values in the Lie algebra ℓ^1 and which can be lifted to the L^1-structure Q identified with the quotient $P/\ker\rho$ (ρ denoting the linear isotropy representation of L^0 defined in Sec. III). The Weyl's curvature tensor is of type (1,3), its components are expressed as functionals of the Christoffel's symbols $\Gamma^\lambda_{\mu\nu}$ of a connection Γ in Q which induces P, and are invariant under the L^0-equivalence. One gets

$$W^\lambda_{\ \mu\nu\rho} = R^\lambda_{\ \mu\nu\rho} + \tfrac{1}{3}(\delta^\lambda_\rho R_{(\mu\nu)} - \delta^\lambda_\nu R_{(\mu\rho)}) - \tfrac{1}{5}(\delta^\lambda_\rho R_{[\mu\nu]} - \delta^\lambda_\nu R_{[\mu\rho]} - 2\delta^\lambda_\mu R_{[\nu\rho]}), \tag{7.1}$$

where $R^\lambda_{\ \mu\nu\rho}$ are the components of the curvature of an admissible connection in a local coordinate system, $R_{(\mu\nu)} = \frac{1}{2}(R_{\mu\nu} + R_{\nu\mu})$ and $R_{[\mu\nu]} = \frac{1}{2}(R_{\mu\nu} - R_{\nu\mu})$, where $R_{\mu\nu}$ are the components of the Ricci tensor given by $R_{\mu\nu} = \Sigma^3_{\sigma=0} R^\sigma_{\ \mu\sigma\nu}$.

Then one can show that *a chronoprojective structure is flat if and only if its Weyl's curvature tensor vanishes.*

The expression (7.1) is formally identical to the one of the projective Weyl's curvature tensor. In fact this means that a chronoprojectively flat space–time is globally projectively flat and that the time-constant slices are conformally flat with respect to the induced metric.

Isotropic Newtonian space–time is chronoprojectively flat

By Newtonian space–time we mean a Galilean manifold (V_4,ψ,γ) equipped with a Newtonian connection, i.e., a Galilean connection whose curvature fulfills the nontrivial identity

$$\xi\cdot R(X,\gamma(\eta))\cdot Y = \eta\cdot R(Y,\gamma(\xi)):X \tag{7.2}$$

for any 1-form ξ and η on V_4 and vector fields X and Y and which satisfies the field equations given by [2,10]

$$\mathrm{Ric} = (4\pi G\rho - \Lambda)\Psi, \tag{7.3}$$

where G is the Newtonian gravitational constant, ρ is the matter density, and Λ the cosmological constant. Hence a Newtonian space–time is just what has been called an admissible Galilean manifold in Sec. IV [relation (4.4)]. So the chronoprojective geometry is the very geometry associated to a Newtonian space–time.

At each point of V_4 one can find a preferred coordinate system (a Galilean chart) such that the Newtonian laws of gravity correspond to $\Gamma^j_{00} = \Phi_{,j}$ and all other $\Gamma^{\alpha}_{\gamma\beta}$ vanish.[11] Φ denotes the Newtonian potential which satisfies Poisson's equation $\Delta\Phi = 4\pi G\rho - \Lambda$.
From relation (7.1) it is easy to see that the components of the Weyl's curvature tensor vanish except the W^j_{00k} components. Moreover, one checks that the components W^j_{00k} vanish if and only if $\partial_j\partial_k\Phi = \frac{1}{3}\delta_{jk}(4\pi G\rho - \Lambda)$. These conditions just express the cosmological isotropy hypothesis. Therefore, *the chronoprojective structure over an isotropic Newtonian space–time is flat*. However, this does not entail that the automorphism group of the chronoprojective structure over an isotropic space–time reaches its maximal dimension.

Let us exhibit two examples for which the maximal dimension is reached. This will be performed by looking at those vector fiels X on V_4 which satisfy the system (6.4a,b,c).

A. The isotropic empty space–time with a cosmological constant

It corresponds to $\rho = 0$. In a Galilean coordinate system, let ξ^0, ξ^j ($j = 1,2,3$) denote the components of a vector field X which verifies the system (6.4). By setting $\alpha = \sqrt{\Lambda/3}$ for $\Lambda \geqslant 0$ and $\alpha = i\sqrt{|\Lambda|/3}$ for $\Lambda < 0$, one gets

$$\begin{aligned}
\xi^0 &= c(1 - \cosh 2\alpha t)/2\alpha^2 + b\sinh(2\alpha t)/\alpha + a,\\
\xi^j &= L^j_k x^k + A^j\cosh\alpha t + V^j\sinh(\alpha t)/\alpha - \tfrac{1}{2}\varepsilon_s x^j,\\
\varepsilon_t &= 4b\cosh 2\alpha t - 2c\sinh(2\alpha t)/\alpha,\\
\varepsilon_s &= 2d - \tfrac{1}{2}\varepsilon_t,\\
\eta &= 2(c\cosh 2\alpha t - 2\alpha b\sinh 2\alpha t)\,dt,
\end{aligned} \tag{7.4}$$

where $L = \{L^j_k\}$ is a rotation matrix belonging to the Lie algebra of O(3) and $a, b, c, d, A^j, V^j \in \mathbb{R}$.

The family of vector fields we thus get depends upon 13 real parameters. The Lie bracket of two such vector fields is a vector field of the family and one verifies that the generated Lie algebra is $\mathrm{O}^2(3)$, the algebra of the chronoprojective group.

In terms of the above parametrization the commutation relations of $\mathrm{O}^2(3)$ are expressed by

$$\begin{aligned}
L'' &= [L', L],\\
\overline{V}'' &= L'\overline{V} - L\overline{V}' + \overline{V}'(b + a)\\
&\quad - \overline{V}(b' + d') + a\overline{A}' - a'\overline{A},\\
\overline{A}'' &= L'\overline{A} - L\overline{A}' + \overline{V}a' - \overline{V}'a\\
&\quad + \overline{A}(b' - d') - \overline{A}'(b - d),\\
a'' &= 2(b'a - ba'),\\
b'' &= ca' - c'a,\\
c'' &= 2(c'b - cb'),\\
d'' &= 0,
\end{aligned} \tag{7.5}$$

where $c \in g_{-2}$, $\overline{A} \in g_{-1}$, $(L, b, d) \in g_0$, $\overline{V} \in g_1$, and $a \in g_2$. These results still hold in the case $\alpha = 0$ which corresponds to $\Lambda = 0$.

From these results one can establish that the local diffeomorphisms which keep invariant the Galilean structure (i.e., which correspond to $\varepsilon_s = 0$, $\varepsilon_t = 0$, $\eta = 0$ implying $a = b = d = 0$) generate ten-dimensional Lie subalgebras of $\mathrm{O}^2(3)$, which are the Galilei Lie algebra if $\Lambda = 0$ and the two so-called Newton Lie algebras for $\Lambda > 0$ and $\Lambda < 0$. Note that these three Lie algebras can be obtained as contractions when the velocity of light tends to infinity of three ten-dimensional subalgebras of the algebra so(4,2), namely, the Poincaré algebra and the two De Sitter algebras so(4,1) and so(3,2), respectively.[12]

If one looks for the diffeomorphisms ($\xi^0 = 0$) which keep invariant the leaves of the foliation induced by ψ (slices of constant t), one gets another ten-dimensional subalgebra of the chronoprojective algebra which is the semidirect sum of the isochronous (derived) Galilei Lie algebra g' by a dilation $\mathbb{R}_d$ (parametrized by d). Then it is worth noticing that $\mathrm{O}^2(3)$ can also be interpreted as the derivation algebra of $g'\square\mathbb{R}_d$ (where $\square$ is the semidirect sum symbol).

B. The Newtonian cosmological model

We consider here the Newtonian cosmological model N_0 introduced in Ref. 13, which is globally strictly equivalent to the Friedmann model. There is no cosmological constant ($\Lambda = 0$) and the model is fully determined by the two constants B and K in the function $\varphi(\tau) = (B/K)(1 - \cosh\alpha\tau)$ with $\alpha = i\sqrt{K}$ for $K > 0$ and $\alpha = \sqrt{|K|}$ for $K < 0$, where it has been set $\tau(t) = \int(4\pi G\rho(t)/3B)^{1/3}\,dt$. B can be interpreted as the galactic matter density and $\varphi(\tau)$ is related to the Hubble coefficient H by $H = \varphi(\tau)^{-2}d\varphi(\tau)/d\tau$.

The resolution of the system (6.4) now leads to

$$\begin{aligned}
\xi^0 &= \varphi(\tau)\left\{-2c\,\frac{\cosh\alpha\tau - 1}{\alpha^2} + 2b\,\frac{\sinh\alpha\tau}{\alpha} + a\right\},\\
\xi^j &= L^j_k x^k + V^j\,\frac{\cosh\alpha\tau - 1}{\alpha^2}\\
&\quad + \frac{A^j}{2}\,\frac{\sinh\alpha\tau}{\alpha} - \frac{1}{2}\,\frac{\varepsilon_s}{2}\,x^j,\\
\varepsilon_t &= -8c\,\frac{\sinh\alpha\tau}{\alpha} + 4b\,\frac{\cosh 2\alpha\tau - \cosh\alpha\tau}{\cosh\alpha\tau - 1}\\
&\quad + 2a\,\frac{\alpha\sinh\alpha\tau}{\cosh\alpha\tau - 1},\\
\varepsilon_s &= 2d - \tfrac{1}{2}\varepsilon_t,\\
\eta &= \left(\frac{a\alpha^2}{\cosh\alpha\tau - 1} + 4(c\cosh\alpha\tau - \alpha b\sinh 2\tau)\right)dt,
\end{aligned} \tag{7.6}$$

where $L = \{L^j_k\}$ is a rotation matrix belonging to O(3), $a, b, c, d, A^j, V^j \in \mathbb{R}$. Again the family of vector fields satisfying (6.4) depends upon 13 real parameters and generates the Lie algebra of the chronoprojective group. This result remains valid for $\alpha = 0$ ($K = 0$).

By looking for the diffeomorphisms which keep invariant slices of constant t, once more one gets the semidirect sum of the isochronous Galilei Lie algebra by a dilation. But the invariance algebra of the the Galilean structure in the previously mentioned sense is only the nine-dimensional isochronous Galilei Lie algebra which is a subalgebra of the invariance algebras of the cosmological constant model.

A Newtonian model N_Λ with a cosmological constant Λ can also be considered, which appears as the Newtonian analogous of the Friedmann–Gamov–Lemaitre (FGL) model describing the large scale properties of the universe.[14] In fact, N_Λ is strictly equivalent to the FGL model by neglecting the radiation pressure, so the expansion of the universe can be described by N_Λ in the present matter-dominated era far from a hot period.

Then it is worth noticing that N_Λ also admits the maximal $O^2(3)$ invariance which strengthens the fact that the chronoprojective geometry is the very geometry of Newtonian cosmology.

Infinitesimal automorphisms of the evolution space of a massive test particle

It has been shown in Ref. 2 that $H(V_4)$ endowed with the closed 2-form $\sigma = m\,{}^t\theta \wedge \overline{\omega}_0$ can be considered as the evolution space of a freely falling massive ($m \neq 0$) particle into the Newtonian potential Φ, which induces the Galilean connection of the configuration space V_4. More explicitly, the leaves of the characteristic foliation E_σ of σ project upon nonisotropic geodesics of V_4 (the possible world lines of the particle), and the quotient $H(V_4)/E_\sigma$ is a six-dimensional manifold: the space of motions of the particle.

First it is interesting to note that *two L^0-equivalent Galilean connections Γ and Γ' without torsion, i.e., which belong to the same chronoprojective structure P, have the same geodesics, up to a change of parameter s vs. s', which, according to* (4.6), *must satisfy*

$$2\xi\psi_\mu \frac{dx^\mu}{ds'} + \frac{d^2s}{ds'^2} = 0\,. \tag{7.7}$$

Let us then define a *geodesic of a chronoprojective structure P* as a curve x_s in V_4 given by

$$x_s = \Pi'((\exp sB)u_0) \tag{7.8}$$

for some standard horizontal vector field B and for some point $u_0 \in P$, where $\Pi': P \to V_4$ is the projection. It can be shown that *if one disregards parametrization, every geodesic of P is a geodesic of an admissible Galilean connection Γ which induces P such that any tangent vector is an eigenvector of the Ricci tensor of Γ corresponding to a null eigenvalue.* But according to (4.4) every tangent vector to a geodesic lies in the kernel of ψ, so chronoprojective geodesics are spacelike and cannot correspond to world lines of freely falling massive particles.

Now let us study the behavior of the presymplectic form σ under chronoprojective transformations. From (6.2) it is easy to verify that

$$L_{X'}\sigma = -\tfrac{1}{2}[(\varepsilon_t + 2\varepsilon_s)\circ \Pi]\sigma,$$

and, according to (6.3), X' appears to be an inifintesimal canonical similitude [or a conformal (pre)symplectomorphism]. Thus it has been shown that the group of (local) chronoprojective transformations is a (local) group of automorphisms of the presymplectic form and, in particular, the Hamiltonian vector fields over $H(V_4)$ defined by $L_{X'}\sigma = 0$ generate a particular Lie subalgebra of the chronoprojective algebra which is described below. One can see that the two above-given realizations of the chronoprojective algebra obtained for cosmological models are characterized by $\varepsilon_t + 2\varepsilon_s = 4d$. If $d = 0$ is set into the chronoprojective Lie algebra, one gets a 12-dimensional subalgebra of $\mathscr{O}^2(3)$ known in the literature as the Schrödinger Lie algebra and denoted by sch. As a matter of fact, $\mathscr{O}^2(3) \simeq sch \,\square\, \mathbb{R}_d$. Therefore, the Hamiltonian vector fields over V_4 correspond to $d = 0$ and generate either the whole Schrödinger Lie algebra or only a subalgebra of it according to the considered model of space–time.

Finally, it can be shown that "accidental" symmetries in various problems such as the rising of the Lie algebra $\mathscr{O}^2(3) \oplus \mathscr{s}_2$ of canonical simplitudes of the Kepler problem (virial theorem) and of the Lie algebra $\mathscr{O}(3) \oplus \mathscr{O}(2,1)$ [15] of canonical transformations of the phase space of a charged particle moving in the field of a magnetic monopole can be introduced as chronoprojective transformations.[16]

[1] S. Kobayashi, *Transformation Groups in Differential Geometry* (Springer-Verlag, Berlin-Heidelberg-New York, 1972).

[2] H. Kunzle, Ann. Inst. H. Poincaré **17A**, 337–62 (1972); C. Duval and H. Kunzle, Rep. Math. Phys. **13**, 351–68 (1978); C. Duval and H. Kunzle, C.R. Acad. Sci. Paris **285**, 813–6 (1977).

[3] J. Ehlers, A. E. Pirani, and A. Schild, *General Relativity Papers in Honour of J. L. Synge*, edited by L. O.'Raifeartaigh (Oxford U.P., Oxford, 1972).

[4] (a) C. R. Hagen, Phys. Rev. D **5**, 337–88 (1972); U. Neiderer, Helv. Phys. Acta **45**, 802–10 (1972); (b) U. Niederer, Helv. Phys. Acta **46**, 191–200 (1973).

[5] (a) G. Burdet and M. Perrin, J. Math. Phys. **16**, 2172–6 (1975); (b) G. Burdet and M. Perrin, Lett. Nuovo Cimento **4**, 651–5 (1972).

[6] G. Burdet, M. Perrin, and M. Perroud, Commun. Math. Phys. **58**, 241–54 (1978).

[7] G. Burdet and M. Perrin, Trans. Amer. Math. Soc. **266**, 583–602 (1981).

[8] G. Burdet, C. Duval, and M. Perrin, to be published.

[9] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York-London, 1963), Vol. I.

[10] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), p. 294.

[11] W. G. Dixon, Commun. Math. Phys. **45**, 167–82 (1975).

[12] G. Burdet and M. Perrin, Nuovo Cimento A **25**, 181–90 (1975).

[13] J. M. Souriau, in *Géométrie symplectique et physique mathématique*, Souriau ed. (CNRS, Marseille, 1975), Vol. 237, pp. 59–113.

[14] J. M. Souriau, in *Proc. Colloque du Centenaire d'Einstein* (CNRS, Marseille, 1980).

[15] R. Jackiw, Ann. Phys. (N.Y.) **129**, 183–200 (1980).

[16] C. Duval, Thèse de Doctorat d'Etat, Marseille, 1982.

Time-dependent vector constants of motion, symmetries, and orbit equations for the dynamical system $\ddot{\mathbf{r}} = \hat{\imath}_r\{[\ddot{U}(t)/U(t)]r - [\mu_0/U(t)]r^{-2}\}$

Gerald H. Katzin
Department of Physics, North Carolina State University, Raleigh, North Carolina 27650-5367

Jack Levine
Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650-5548

(Received 2 March 1982; accepted for publication 27 January 1983)

The most general *time-dependent*, central force, classical particle dynamical systems (in n-dimensional Euclidean space, $n = 2$ or 3) of the form (a) $\ddot{\mathbf{r}} = \hat{\imath}_r F(r, t)$, ($r^2 \equiv \mathbf{r} \cdot \mathbf{r}$, $\mathbf{r} = \hat{\imath}_k x^k$, $k = 1,...,n$), which admit vector constants of motion of the form (b) $\mathbf{I} = U(r, t)(\mathbf{L}\times\mathbf{v}) + Z(r, t)(\mathbf{L}\times\mathbf{r}) + W(r, t)\mathbf{r}$ ($\mathbf{L}\equiv\mathbf{r}\times\mathbf{v}, \mathbf{v}\equiv\dot{\mathbf{r}}$) are obtained. It is found that the only class of such dynamical systems is (c) $\ddot{\mathbf{r}} = \hat{\imath}_r(\ddot{U}U^{-1}r - \mu_0 U^{-1}r^{-2})$, for which the concomitant vector constant of motion (b) takes the form (d) $\mathbf{I} = U(\mathbf{L}\times\mathbf{v}) - \dot{U}(\mathbf{L}\times\mathbf{r}) + \mu_0 r^{-1}\mathbf{r}$, where in (c) and (d) $U = U(t)$ is arbitrary ($\neq 0$). The dynamical system (c) includes both the time-dependent harmonic oscillator and a time-dependent Kepler system. Based upon infinitesimal *velocity-independent* mappings the complete symmetry group for the dynamical system (c) is obtained. This complete group of $[2 + n(n-1)/2]$ parameters contains a complete Noether symmetry subgroup of $[1 + n(n-1)/2]$ parameters. In addition to the $n(n-1)/2$ angular momenta, there is an energy-like constant of motion also associated with the Noether symmetries. By means of the vector constant of motion (d), the orbit equations of the dynamical system (c) are obtained. A one-dimensional procedure for obtaining constants of motion developed by Lewis and Leach is applied to the effective one-dimensional system concomitant to (c). Relations between constants of motion so obtained and those mentioned above are determined.

PACS numbers: 03.20. + i, 95.10.Ce, 41.70. + t

1. INTRODUCTION

It is well known that the classical time-independent Kepler dynamical system [1,2]

$$\ddot{\mathbf{r}} = -\hat{\imath}_r \frac{\Phi_0}{r^2}, \quad \Phi_0 \equiv \text{const},$$
$$r^2 \equiv \mathbf{r}\cdot\mathbf{r}, \quad \mathbf{r} \equiv \hat{\imath}_k x^k, \quad k = 1,...,n \tag{1.1}$$

($n = 2$ or 3) admits the Laplace–Runge–Lenz vector constant of motion

$$\mathbf{A}_0 \equiv \Phi_0^{-1}(\mathbf{L}\times\mathbf{v}) + \mathbf{r}/r, \quad \mathbf{v}\equiv\dot{\mathbf{r}}, \tag{1.2}$$

where

$$\mathbf{L}\equiv\mathbf{r}\times\mathbf{v} \tag{1.3}$$

is the angular momentum.[3]

In a recent paper[4,5] we determined all time-dependent dynamical systems of the form

$$\ddot{\mathbf{r}} = \hat{\imath}_r \Phi(t)/r^2, \tag{1.4}$$

which admitted quadratic constants of motion with explicit time dependence. In particular it was shown for the choice

$$\Phi(t) \equiv \lambda_0/(\alpha t + \beta), \quad \alpha, \beta, \lambda_0 \equiv \text{consts}, \tag{1.5}$$

that the time-dependent Kepler system (1.4) admits the time-dependent vector constant of motion[6]

$$\mathbf{A} \equiv \frac{\alpha t + \beta}{\lambda_0}(\mathbf{L}\times\mathbf{v}) - \frac{\alpha}{\lambda_0}(\mathbf{L}\times\mathbf{r}) + \frac{\mathbf{r}}{r}. \tag{1.6}$$

This *time-dependent* system (1.4), (1.5) with concomitant vector constant of motion (1.6) is a generalization of the *time-independent* system (1.1) with concomitant vector constant of motion (1.2). By use of (1.6) the orbit equations for the time-dependent dynamical system (1.4), (1.5) were easily integrated.

In this present paper we continue with the analysis of time-dependent central force dynamical systems which admit vector contants of motion. We determine the most general (central force) dynamical system of the form ($n = 2$ or 3)

$$\ddot{\mathbf{r}} = \hat{\imath}_r F(r, t), \tag{1.7}$$

which admits a vector constant of motion of the form

$$\mathbf{I} = U(r, t)(\mathbf{L}\times\mathbf{v}) + Z(r, t)(\mathbf{L}\times\mathbf{r}) + W(r, t)\mathbf{r}. \tag{1.8}$$

The class of dynamical systems (1.7) with concomitant constant of motion of the form (1.8) includes the dynamical system (1.4), (1.5) with constant of motion (1.6).

In Sec. 2 we determine the unknown functions $F(r, t)$, $U(r, t)$, $Z(r, t)$ and $W(r, t)$ appearing in (1.7) and (1.8) and thereby obtain the class of central force systems (1.7) with their associated time-dependent vector constants of motion (1.8). This result is summarized in Theorem 2.1 and Corollary 2.1.

In Sec. 3 we determine the complete groups of dynamical symmetries of the class of dynamical systems obtained in Sec. 2. These symmetries are based upon infinitesimal mappings which are functions of x^i and t (velocity-independent). It is shown that the above-mentioned complete group of symmetries includes a subgroup of Noether symmetries, and the concomitant Noether constants of motion are obtained. In addition to the anticipated angular momenta, these Noether constants of motion include a time-dependent "generalized energy" integral. The vector constants of motion of the form (1.8) are not among these Noether constants of mo-

 0022-2488/83/071761-11$02.50

tion. The results of Sec. 3 are summarized in Theorems 3.1 and 3.2.

In Sec. 4 a procedure is formulated for obtaining the orbit equations for the class of dynamical systems derived in Sec. 2. This procedure is based upon the concomitant vector constant of motion (1.8).

In Sec. 5 we determine the subclass of those dynamical systems obtained in Sec. 2 for which the force $F(r, t)$ is product-separable.

In Sec. 6 we give several examples which illustrate the procedure of Sec. 4 for determining the orbit equations.

In Sec. 7 a one-dimensional procedure for obtaining constants of motion developed by Lewis and Leach[7] is applied to the effective one-dimensional dynamical system associated with the dynamical system determined in Sec. 2. Relations between the constants of motion so obtained and those derived in Secs. 2 and 3 are determined.

2. DETERMINATION OF F, U, Z, W

For $\mathbf{I}$ (1.8) to be a constant of motion of the dynamical system (1.7) we must have[2]

$$\dot{\mathbf{I}} \doteq 0. \tag{2.1}$$

After forming the total time derivative $\dot{\mathbf{I}}$ of (1.8) we eliminate $\ddot{\mathbf{r}}$ by (1.7) and replace $\mathbf{L}\times\mathbf{r}$ and $\mathbf{L}\times\mathbf{v}$, respectively, by

$$\mathbf{L}\times\mathbf{r} \equiv r^2\mathbf{v} - (\mathbf{r}\cdot\mathbf{v})\mathbf{r}, \tag{2.2}$$

$$\mathbf{L}\times\mathbf{v} \equiv (\mathbf{r}\cdot\mathbf{v})\mathbf{v} - v^2\mathbf{r}, \quad v^2 = \mathbf{v}\cdot\mathbf{v}. \tag{2.3}$$

This leads to the equation

$$\begin{aligned}\dot{\mathbf{I}} \doteq \mathbf{v}\Big[&\frac{(\mathbf{r}\cdot\mathbf{v})^2}{r}U_{,r} \\ &+ (\mathbf{r}\cdot\mathbf{v})(U_{,t} + Z + rZ_{,r}) + rUF + r^2 Z_{,t} + W\Big] \\ +\mathbf{r}\Big[&-v^2\frac{(\mathbf{r}\cdot\mathbf{v})}{r}U_{,r} + v^2(U_{,t} + Z) - \frac{(\mathbf{r}\cdot\mathbf{v})^2}{r}Z_{,r} \\ &-(\mathbf{r}\cdot\mathbf{v})\Big(\frac{UF}{r} + Z_{,t} - \frac{1}{r}W_{,r}\Big) + W_{,t}\Big] \doteq 0.\end{aligned} \tag{2.4}$$

When (2.4) is expressed in rectangular coordinates $x^1, x^2, \ldots, x^n$ ($n = 2$ or 3), we obtain after rearrangement

$$\begin{aligned}&\dot{x}^i\dot{x}^j\dot{x}^m\big[(\delta^k_m x^i x^j - \delta^i_m x^j x^k)(1/r)U_{,r}\big] \\ &\quad + \dot{x}^i\dot{x}^j\big[\delta^k_j x^i(U_{,t} + Z + rZ_{,r}) \\ &\quad - \delta^i_j x^k(U_{,t} + Z) - x^i x^j x^k (1/r)Z_{,r}\big] \\ &\quad + \dot{x}^i\{\delta^k_i(rUF + r^2 Z_{,t} + W) \\ &\quad - x^i x^k\,[UF/r + Z_{,t} - (1/r)W_{,r}]\} + x^k W_{,t} \doteq 0.\end{aligned} \tag{2.5}$$

In order that (2.5) hold identically in the $\dot{x}$'s it follows (after appropriate symmetrization) that

$$\begin{aligned}(1/r)U_{,r}\,[&\delta^m_k x^i x^j - \delta^m_i x^j x^k + \delta^j_k x^i x^m - \delta^m_j x^m x^k \\ &- \delta^m_j x^i x^k + \delta^i_k x^j x^m] = 0,\end{aligned} \tag{2.6}$$

$$\begin{aligned}(U_{,t} + Z + rZ_{,r})(\delta^k_j x^i + \delta^k_i x^j) - 2(U_{,t} + Z)\delta^i_j x^k \\ - (2/r)Z_{,r} x^i x^j x^k = 0,\end{aligned} \tag{2.7}$$

$$(rUF + r^2 Z_{,t} + W)\delta^k_i - (1/r)(UF + rZ_{,t} - W_{,r})x^i x^k = 0, \tag{2.8}$$

$$W_{,t} = 0. \tag{2.9}$$

From (2.9) we have

$$W = W(r). \tag{2.10}$$

From (2.8) with $i \neq k$ it follows that

$$UF + rZ_{,t} - W_{,r} = 0. \tag{2.11}$$

Hence by use of (2.11) in (2.8) with $i = k$ it follows that

$$UF + rZ_{,t} + W/r = 0. \tag{2.12}$$

In (2.7) by contraction on i and j we obtain

$$U_{,t} + Z = 0. \tag{2.13}$$

Use of (2.13) in (2.7) followed by contraction on j and k gives

$$Z_{,r} = 0. \tag{2.14}$$

Hence

$$Z = Z(t). \tag{2.15}$$

By contraction on m and k in (2.6) we obtain

$$U_{,r}(nx^i x^j - \delta^i_j r^2) = 0. \tag{2.16}$$

If in (2.16) we choose $i = j$ (not summed), we find (since $n > 1$)

$$U_{,r} = 0. \tag{2.17}$$

Hence

$$U = U(t). \tag{2.18}$$

It now follows from (2.13), (2.15), and (2.18) that

$$Z(t) = -\dot{U}(t). \tag{2.19}$$

From (2.10), (2.11), and (2.12) we obtain

$$\frac{dW}{dr} + \frac{W}{r} = 0. \tag{2.20}$$

Integration of (2.20) gives

$$W = \mu_0/r, \quad \mu_0 = \text{const.} \tag{2.21}$$

By means of (2.19) and (2.21) we find that each of (2.11) and (2.12) reduces to

$$UF - r\ddot{U} + \mu_0/r^2 = 0. \tag{2.22}$$

We exclude the case $U = 0$, since (2.19), (2.21), and (2.22) then imply $Z = W = 0$, and hence by (1.8) $\mathbf{I} \equiv 0$.

From (2.22) we find

$$F(r, t) = (\ddot{U}/U)r - (1/U)\mu_0/r^2. \tag{2.23}$$

We have thus proved that (2.18), (2.19), (2.21), and (2.23) [where $U(t) \neq 0$, but otherwise arbitrary] are necessary for (2.5) to hold. By inspection it is easily seen that these conditions are also sufficient.

We summarize the above work in the following theorem.

Theorem 2.1: A necessary and sufficient condition that a time-dependent central force dynamical system (of two or three dimensions) of the form

$$\ddot{\mathbf{r}} = \hat{\imath}_r\, F(r, t), \tag{1.7$'$}$$

where $\mathbf{r} = \hat{\imath}_r r$ is the radius vector from the center of force, admits the time-dependent vector first integral of the form

$$\mathbf{I} = U(r, t)(\mathbf{L}\times\mathbf{v}) + Z(r, t)(\mathbf{L}\times\mathbf{r}) + W(r, t)\mathbf{r} \quad (\mathbf{v} \equiv \dot{\mathbf{r}}) \tag{1.8$'$}$$

is that

$$U = U(t) \neq 0 \quad \text{(otherwise arbitrary)},$$

$$Z = -\dot{U}(t),$$

$W=\mu_0/r,\quad \mu_0=\text{const},$
$F=(\ddot U/U)r-(1/U)\mu_0/r^2.$

The dynamical system (1.7′) and its concomitant vector first integral (1.8′) will then have the respective forms

$$\ddot{\mathbf r}=\hat\imath_r[(\ddot U/U)r-(1/U)\mu_0/r^2], \tag{2.24}$$

$$\mathbf I=U(t)(\mathbf L\times\mathbf v)-\dot U(t)(\mathbf L\times\mathbf r)+\mu_0\mathbf r/r.\quad\square \tag{2.25}$$

For the case $\mu_0=0$, the dynamical equation (2.24) reduces to the form of a time-dependent oscillator.[8,9] Hence we may state the following corollary.

Corollary 2.1: The time-dependent oscillator (in two or three dimensions)

$$\ddot{\mathbf r}=\hat\imath_r(\ddot U/U)r,\quad U(t)\neq 0,\quad \text{otherwise arbitrary}, \tag{2.26}$$

admits the time-dependent vector constant of motion

$$\mathbf I=U(\mathbf L\times\mathbf v)-\dot U(\mathbf L\times\mathbf r).\quad\square \tag{2.27}$$

Remark 1: In the rectangular coordinates the time-dependent oscillator (2.26) has the form

$$\ddot x^i-(\ddot U/U)x^i=0, \tag{2.28}$$

which has the general solution

$$x^i=a^iU+b^iUS,\quad a^i,b^i\equiv\text{consts}, \tag{2.29}$$

where

$$S\equiv\int U^{-2}\,dt. \tag{2.30}$$

Remark 2: When expressed in rectangular coordinates x^i the components of the vector constant of motion $\mathbf I$ (2.27) lead to the constants of motion F^i,

$$F^i\equiv U\dot x^i-\dot U x^i. \tag{2.31}$$

The F^i are Noether constants of motion of the time-dependent oscillator. This may be verified by taking $\omega\equiv-\frac12\ddot U/U$ and $C^i=U$ in Eqs. (6.21) and (6.25) of Ref. 9.

Remark 3: For the case $\mu_0=1$, $U=(\alpha t+\beta)/\lambda_0$ ($\alpha,\beta,\lambda_0=\text{const}$), the dynamical system (2.24) and constant of motion (2.25) reduce, respectively, to the time-dependent Kepler system (1.4), (1.5) and constant of motion (1.6).[4]

3. SYMMETRY MAPPINGS OF THE DYNAMICAL EQUATION (2.24)

In the preceding sections we determined which central force dynamical systems (1.7) admit vector constants of motion of the form (1.8). This led to the class of dynamical systems (2.24) with concomitant vector constant of motion (2.25).

In this present section we shall determine the symmetries of the dynamical equation (2.24), which in rectangular coordinates takes the form

$$E^i\equiv\ddot x^i-\left(\frac{\ddot U}{U}x^i-\frac{\mu_0}{U}\frac{x^i}{r^3}\right)=0,\quad r^2\equiv\delta_{ij}x^ix^j, \tag{3.1}$$

where $U=U(t)\neq0$ and $\mu_0\neq0$.[10] For generality in the symmetry analysis we shall consider (3.1) to be an n-dimensional dynamical system for arbitrary n (i.e., for $n\geqslant1$).

These symmetries will be based on infinitesimal mappings of the form

$$\begin{aligned}\bar x^i&=x^i+\delta x^i,\quad \delta x^i\equiv\xi^i(x,t)\delta a,\\ \bar t&=t+\delta t,\quad \delta t\equiv\xi^0(x,t)\delta a.\end{aligned} \tag{3.2}$$

The procedure for obtaining symmetries of (3.1) is determined by the requirement that $\delta E^i=0$ whenever $E^i=0$. For systems derivable from a Lagrangian this process will in general include Noether symmetries as a subcase. This procedure for obtaining symmetries is discussed in complete detail in previous papers.[11] In the paper referred to in Ref. 9 it is shown that for any dynamical system based upon a Lagrangian of the form

$$\mathscr L=\tfrac12\delta_{ij}\dot x^i\dot x^j-V(x,t) \tag{3.3}$$

(where x^i are rectangular coordinates in an n-dimensional Euclidean configuration space), the above-mentioned symmetry procedure leads to mapping functions ξ^i,ξ^0 of the form

$$\xi^i=\dot A_j(t)x^jx^i+B^i_j(t)x^j+C^i(t), \tag{3.4}$$

$$\xi^0=A_jx^j+B(t), \tag{3.5}$$

where the functions $A_j(t)$, $B(t)$, $B^i_j(t)$, and $C^i(t)$ appearing in (3.4) and (3.5) must satisfy the conditions[12]

$$\begin{aligned}A_kV_{,k}\delta^i_m&+2A_mV_{,i}-(\ddot A_kx^k+\ddot B)\delta^i_m\\&+2(\ddot A_mx^i+\ddot A_kx^k\delta^i_m+\dot B^i_m)=0,\end{aligned} \tag{3.6}$$

$$\begin{aligned}\dddot A_mx^mx^i&+\ddot B^i_mx^m+\ddot C^i\\&-(\dot A_kx^i+\dot A_mx^m\delta^i_k+B^i_k)V_{,k}\\&+2(\dot A_mx^m+\dot B)V_{,i}\\&+V_{,ik}(\dot A_mx^mx^k+B^k_mx^m+C^k)\\&+V_{,it}(A_mx^m+B)=0.\end{aligned} \tag{3.7}$$

For the dynamical system (2.24) it follows that the potential energy V appearing in (3.6) and (3.7) is given by

$$V(x,t)=-\frac12\frac{\ddot U}{U}r^2-\frac1U\frac{\mu_0}{r}. \tag{3.8}$$

We now proceed with the solution of (3.6) and (3.7) for the functions A_k, B^i_m, and C^i. By means of (3.8) we eliminate the V derivatives in (3.6) and obtain

$$\begin{aligned}A_k\left(-\frac{\ddot U}{U}x^k+\frac{\mu_0}{U}\frac{x^k}{r^3}\right)\delta^i_m+2A_m\left(-\frac{\ddot U}{U}x^i+\frac{\mu_0}{U}\frac{x^i}{r^3}\right)\\-(\ddot A_kx^k+\ddot B)\delta^i_m+2(\ddot A_mx^i+\ddot A_kx^k\delta^i_m+\dot B^i_m)=0.\end{aligned} \tag{3.9}$$

We consider the two cases $n>1$ and $n=1$.

Case 1, $n>1$ ($\mu_0\neq0$)[10]

In (3.9) consider $m\neq i$ and differentiate with respect to x^m to obtain

$$A_mx^i\left(\frac{\mu_0}{U}\frac{x^m}{r^5}\right)=0\quad(m\neq i,\ m\text{ not summed}). \tag{3.10}$$

Hence (3.10) implies

$$A_k=0\quad(k=1,\dots,n). \tag{3.11}$$

Use of (3.11) in (3.9) gives

$$2\dot B^i_m-\ddot B\delta^i_m=0. \tag{3.12}$$

Integrating (3.12), we obtain

$$B^i_j = \tfrac{1}{2}\dot{B}(t)\delta^i_j + b^i_j, \quad b^i_j = \text{const.} \tag{3.13}$$

It follows that (3.11) and (3.13) are necessary and sufficient to satisfy (3.9). This leaves (3.7) to be solved.

Use of (3.8), (3.11), and (3.13) in (3.7) followed by multiplication of the resulting equation by r^5 leads to

$$\begin{aligned} r^5\Big\{&\Big(\ddot{C}^i - \frac{\ddot{U}}{U}C^i\Big) \\ &+ x^i\Big[\frac{1}{2}\dddot{B} - 2\dot{B}\frac{\ddot{U}}{U} + B\Big(\frac{-\dddot{U}}{U} + \frac{\ddot{U}\dot{U}}{U^2}\Big)\Big]\Big\} \\ &+ \frac{\mu_0}{U}x^i x^k x^m\Big[\Big(\frac{1}{2}\dot{B} - \frac{B\dot{U}}{U}\Big)\delta^k_m - 3b^k_m\Big] \\ &+ \frac{\mu_0}{U}x^k x^m(C^i\delta^k_m - 3C^k\delta^i_m) = 0. \end{aligned} \tag{3.14}$$

Since $n > 1$, r^5 is irrational in the x's. It follows that the coefficient of r^5 in (3.14) must vanish because the remainder of the terms in (3.14) are polynomials in the x's. Since the coefficient of r^5 is a polynomial in the x's, we are thus led to the conditions

$$\ddot{C}^i - (\ddot{U}/U)C^i = 0, \tag{3.15}$$

$$\tfrac{1}{2}\dddot{B} - 2\dot{B}\ddot{U}/U + B(-\dddot{U}/U + \ddot{U}\dot{U}/U^2) = 0, \tag{3.16}$$

$$(\tfrac{1}{2}\dot{B} - B\dot{U}/U)\delta^k_m - \tfrac{3}{2}(b^k_m + b^m_k) = 0, \tag{3.17}$$

$$C^i\delta^k_m - \tfrac{3}{2}(C^k\delta^i_m + C^m\delta^i_k) = 0. \tag{3.18}$$

From (3.18) it follows that

$$C^i = 0. \tag{3.19}$$

From (3.17) we obtain

$$b^k_m + b^m_k = 0, \quad m \neq k, \tag{3.20}$$

$$\tfrac{1}{2}\dot{B} - B\dot{U}/U - 3b^m_m = 0 \quad (m \text{ not summed}). \tag{3.21}$$

From (3.21) it follows that

$$b^1_1 = b^2_2 = \cdots = b^n_n \equiv b_1 = \text{const} \tag{3.22}$$

and

$$\dot{B} - 2B\dot{U}/U = 6b_1. \tag{3.23}$$

Equation (3.23) can be solved for $B(t)$ to give

$$B(t) = b_2U^2 + 6b_1S, \quad S \equiv \int U^{-2}\,dt, \quad b_2 = \text{const.} \tag{3.24}$$

Equations (3.20) and (3.22) may be combined to give

$$b^i_j = \omega^i_j + b_1\delta^i_j, \quad \omega^i_j = -\omega^j_i = \text{const.} \tag{3.25}$$

By use of (3.25) we may express (3.13) in the form

$$B^i_j = (\tfrac{1}{2}\dot{B} + b_1)\delta^i_j + \omega^i_j. \tag{3.26}$$

Equation (3.16) remains to be satisfied. A straightforward calculation based on (3.23) shows that (3.16) is satisfied identically.

Equations (3.11), (3.19), (3.23), and (3.26) are necessary conditions for the two basic equations (3.9) and (3.14) to be satisfied. It is easily shown that these four conditions are also sufficient. By use of these four conditions the mapping functions ξ^i and ξ^0 of (3.4) and (3.5) are determined to have the form

$$\xi^i = b_1(6U\dot{U}S + 4)x^i + b_2(U\dot{U})x^i + \omega^i_j x^j, \tag{3.27}$$

$$\xi^0 = b_1(6U^2S) + b_2U^2. \tag{3.28}$$

Hence the symmetry mappings of the dynamical equations (2.24) are given by (3.2), (3.27), and (3.28). As is known these mappings will include Noether mappings as a special case.[13]

We now investigate which of the mappings determined by (3.2), (3.27), and (3.28) are Noether mappings. A mapping of the form (3.2) will be a Noether mapping if and only if there exists a function $\tau(x, t)$ which satisfies the equation[13]

$$\delta\mathscr{L} + \mathscr{L}\frac{d}{dt}\delta t = -\frac{d\tau}{dt}\delta a. \tag{3.29}$$

Expansion of (3.29) [with use of the relation[13] $\delta\dot{x}^i = (\dot{\xi}^i - \dot{x}^i\dot{\xi}^0)\delta a$] gives

$$\frac{\partial\mathscr{L}}{\partial\dot{x}^i}(\dot{\xi}^i - \dot{x}^i\dot{\xi}^0) + \frac{\partial\mathscr{L}}{\partial x^i}\xi^i + \frac{\partial\mathscr{L}}{\partial t}\xi^0 + \mathscr{L}\dot{\xi}^0 + \dot{\tau} = 0. \tag{3.30}$$

If (3.27), (3.28) and the Lagrangian $\mathscr{L}$ as determined by (3.3) and (3.8) are substituted in (3.30), the resulting equation may be written in the form

$$\begin{aligned} \dot{x}^i\dot{x}^i[b_1] \;&+ \dot{x}^i[6b_1(\dot{U}^2S + U\ddot{U}S + U^{-1}\dot{U})x^i \\ &+ b_2(\dot{U}^2 + U\ddot{U})x^i + \tau_{,i}] \\ &+ b_1[r^2(9\dot{U}\ddot{U}S + 7\dot{U}/U + 3U\dddot{U}S) + 2\mu_0/rU] \\ &+ b_2[\tfrac{1}{2}r^2(3\dot{U}\ddot{U} + U\dddot{U}) + \tau_{,t}] = 0. \end{aligned} \tag{3.31}$$

Equation (3.31) must hold identically in the x's. Since (3.31) is a polynomial in the $\dot{x}$'s, we obtain the conditions

$$b_1 = 0, \tag{3.32}$$

$$\tau_{,i} = -b_2(\dot{U}^2 + U\ddot{U})x^i, \tag{3.33}$$

$$\tau_{,t} = -\tfrac{1}{2}b_2r^2(3\dot{U}\ddot{U} + U\dddot{U}). \tag{3.34}$$

It is found that the integrability conditions $\tau_{,ij} = \tau_{,ji}$ and $\tau_{,it} = \tau_{,ti}$ are identically satisfied. The solution for τ is easily found from (3.33) and (3.34) to be

$$\tau = -\tfrac{1}{2}b_2r^2(\dot{U}^2 + U\ddot{U}). \tag{3.35}$$

(The constant of integration has been dropped.)

Hence in order for the mapping functions (3.27) and(3.28) to define a Noether mapping, it is necessary and sufficient that $b_1 = 0$ and that τ have the form (3.35). (The sufficiency is easily proven.)

We now return to the general symmetry solution defined by (3.2), (3.27), and (3.28) in order to obtain a set of generators which define the complete group of dynamical symmetries, and also to determine the subgroup of Noether symmetries. Since in these symmetry solutions b_1, b_2, and the $n(n-1)/2$ ω^i_j's are arbitrary constants, this leads to the following $2 + n(n-1)/2$ mapping vectors.

$$\xi^i(b_1) = (6U\dot{U}S + 4)x^i, \quad \xi^0(b_1) = 6U^2S, \tag{3.36}$$

$$\xi^i(b_2) = U\dot{U}x^i, \quad \xi^0(b_2) = U^2, \tag{3.37}$$

$$\xi^i(\omega^j_k) = \delta^i_k x^j - \delta^i_j x^k, \quad \xi^0(\omega^j_k) = 0. \tag{3.38}$$

From (3.36)–(3.38) we find the concomitant generators of the form $X \equiv \xi^i\partial_i + \xi^0\partial_t$ to be, respectively,

$$B_1 = (6U\dot{U}S + 4)x^i\partial_i + 6U^2S\partial_t, \tag{3.39}$$

$$B_2 = U\dot{U}x^i\partial_i + U^2\partial_t, \tag{3.40}$$

$$\Omega_{ij} = x^i\partial_j - x^j\partial_i. \tag{3.41}$$

These generators define the $[2 + n(n-1)/2]$-parameter complete group of dynamical symmetries with group algebra

$$[B_\alpha, \Omega_{ij}] = 0, \quad \alpha = 1, 2, \tag{3.42}$$

$$[B_1, B_2] = -6B_2, \tag{3.43}$$

$$[\Omega_{ij}, \Omega_{kl}] = \delta_{jk}\Omega_{il} - \delta_{jl}\Omega_{ik} - \delta_{ik}\Omega_{jl} + \delta_{il}\Omega_{jk}. \tag{3.44}$$

As determined above, in order for mapping functions ξ^i and ξ^0 of (3.27) and (3.28) to define a Noether symmetry, we must set $b_1 = 0$. Hence the generators defining the subgroup of Noether symmetries are B_2 and the Ω_{ij}. This subgroup defines the (complete) $[1 + n(n-1)/2]$-parameter group of Noether symmetries.

Case 2, $n = 1$ ($\mu_0 \neq 0$)[10]

It is easily shown in this case [that is, for the one-dimensional dynamical system (3.1)] that the symmetry mapping is again given by (3.27) and (3.28) (with $b_1 = 0$ for the Noether symmetry).

We are now prepared to state the following theorem.

Theorem 3.1: A necessary and sufficient condition that the n-dimensional ($n \geqslant 1$) dynamical equations (x^i rectangular coordinates)

$$\ddot{x}^i = (\ddot{U}/U)x^i - (\mu_0/U)x^i/r^3,$$
$$U = U(t), \quad \text{arbitrary } (\neq 0), \quad \mu_0 \neq 0, \tag{3.1'}$$

associated with the Lagrangian [refer to (3.3) and (3.8)]

$$\mathscr{L} = \tfrac{1}{2}\delta_{ij}\dot{x}^i\dot{x}^j + \tfrac{1}{2}(\ddot{U}/U)r^2 - (1/U)\mu_0/r,$$
$$r^2 \equiv \delta_{ij}x^ix^j, \tag{3.45}$$

admit infinitesimal symmetry mappings of the form

$$\bar{x}^i = x^i + \xi^i(x, t)\delta a,$$
$$\bar{t} = t + \xi^0(x, t)\delta a, \tag{3.46}$$

is that

$$\xi^i = b_1(6U\dot{U}S + 4)x^i + b_2U\dot{U}x^i + \omega^i_{\ j}x^j, \tag{3.27'}$$

$$\xi^0 = b_1(6U^2S) + b_2U^2, \tag{3.28'}$$

where $S \equiv \int U^{-2}\,dt$ and b_1, b_2, and $\omega^i_{\ j}$ $(= -\omega^j_{\ i})$ are arbitrary constants. The mapping vectors (3.27') and (3.28') define a complete group of dynamical symmetries of $2 + n(n-1)/2$ parameters with generators defined by

$$B_1 \equiv (6U\dot{U}S + 4)x^i\partial_i + 6U^2S\partial_t, \tag{3.39'}$$

$$B_2 \equiv U\dot{U}x^i\partial_i + U^2\partial_t, \tag{3.40'}$$

$$\Omega_{ij} \equiv x^i\partial_j - x^j\partial_i. \tag{3.41'}$$

These generators have the group algebra

$$[B_\alpha, \Omega_{ij}] = 0, \quad \alpha = 1, 2, \tag{3.42'}$$

$$[B_1, B_2] = -6B_2, \tag{3.43'}$$

$$[\Omega_{ij}, \Omega_{kl}] = \delta_{jk}\Omega_{il} - \delta_{jl}\Omega_{ik} - \delta_{ik}\Omega_{jl} + \delta_{il}\Omega_{jk}. \tag{3.44'}$$

The mappings (3.46) defined by (3.27') and (3.28') with $b_1 = 0$ are Noether mappings wherein

$$\tau(b_2) = -\tfrac{1}{2}b_2r^2(U\ddot{U} + \dot{U}^2), \tag{3.47}$$

$$\tau(\omega^i_{\ j}) = 0. \tag{3.48}$$

These mappings define the $[1 + n(n-1)/2]$-parameter complete group of Noether symmetries (with generators B_2 and Ω_{ij}) which is a subgroup of the above-mentioned group of $[2 + n(n-1)/2]$ parameters.□

Concomitant with each Noether symmetry mapping there will exist a Noether constant of motion of the form[13]

$$I_N = \frac{\partial\mathscr{L}}{\partial\dot{x}^i}\xi^i - \left(\frac{\partial\mathscr{L}}{\partial\dot{x}^i}\dot{x}^i - \mathscr{L}\right)\xi^0 + \tau. \tag{3.49}$$

Hence we may state the following theorem.

Theorem 3.2: The dynamical system described in Theorem 3.1 admits the $1 + n(n-1)/2$ Noether constants of motion:

$$\begin{aligned} I_N(b_2) &= U\dot{U}x^i\dot{x}^i - U^2\{\tfrac{1}{2}\delta_{ij}\dot{x}^i\dot{x}^j - [\tfrac{1}{2}(\ddot{U}/U)r^2 + (\mu_0/U)1/r]\} \\ &\quad - \tfrac{1}{2}(U\ddot{U} + \dot{U}^2)r^2 \doteq \tilde{k}, \end{aligned} \tag{3.50}$$

$$I_N(\omega^i_{\ j}) = x^i\dot{x}^j - x^j\dot{x}^i. \quad \square \tag{3.51}$$

Remark 1: Note that in (3.50) the term in the braces is the total energy [refer to (3.8)].

Remark 2: The Noether constants of motion (3.50) and (3.51) (based upon the complete group of Noether symmetries of Theorem 3.1) do not include the constants of motion which are the components (in rectangular coordinates) of the vector constant of motion **I** [(2.25)].[14]

Remark 3: It is of interest to note ($n = 2$ or 3) that the vector constant of motion **I** [(2.25)], the Noether constant of motion $I_N(b_2)$ [(3.50)], and the angular momentum **L** [(1.3)] associated with the time-dependent dynamical system (2.24) are functionally related in that

$$\mathbf{I}\cdot\mathbf{I} = -2I_N(b_2)\mathbf{L}\cdot\mathbf{L} + \mu_0^2. \tag{3.52}$$

A similar dependence between analogous constants of motion **A**, **L**, and E_3 was found for a time-dependent Kepler system.[4,6] This same functional dependence also exists between the Laplace–Runge-Lenz vector constant of motion $\mathbf{A}_0$, angular momentum **L**, and energy E for the time-independent Kepler system. (See, for example, Ref. 1, Chap. 3.) These similarities are not unexpected since the time-dependent dynamical system (2.24) includes both of the above-mentioned Kepler systems.

Remark 4: In obtaining the complete Noether symmetry group of the dynamical system (3.1), we assumed $\mu_0 \neq 0$ in order to exclude the case of the time-dependent oscillator for which the complete Noether symmetry group is known.[9,15,16] It turns out however that the Noether symmetries so obtained (with $\mu_0 \neq 0$) form a *subgroup* of the complete Noether symmetry group for the time-dependent oscillator. Hence the Noether constant of motion $I_N(b_2)$ with $\mu_0 = 0$ given by (3.50) will be a Noether constant of motion of the time-dependent oscillator (2.26). In fact, if we assume $\mu_0 = 0$ in $I_n(b_2)$, then it reduces to one of the Noether integrals $C_2(B)$ given by Eq. (6.24) of Ref. 9. This is easily verified by taking $\omega(t) \equiv -\tfrac{1}{2}\ddot{U}/U$, $B(t) \equiv U^2$, and $c_0 = 0$ in Eqs. (6.24) and (4.13) of Ref. 9.

Remark 5: For the case $U = (\alpha t + \beta)/\lambda_0$, $\mu_0 = 1$, the dynamical system (3.1) reduces to the time-dependent Kepler problem (1.4), (1.5) (discussed in Ref. 4), and the

Noether constant of motion $I_N(b_2)$ given by (3.50) reduces to the Noether constant of motion $I_N^*(\mu_2)$ of this time-dependent Kepler problem [Eq. (4.28) of Ref. 4].

4. ORBIT EQUATIONS

We now develop a procedure based on use of the vector constant of motion (2.25), which in general will lead to the orbit equation for the dynamical system (2.24) determined by a given function $U(t)$.

In two dimensions the vector constant of motion (2.25) may be expressed in terms of rectangular coordinates in the form

$$\mathbf{I} = \hat{\imath}_x I_x + \hat{\imath}_y I_y, \tag{4.1}$$

where

$$I_x \overset{\circ}{=} l_0(-U\dot{y} + \dot{U}y) + \mu_0 x/r \overset{\circ}{=} k_x, \quad k_x = \text{const}, \tag{4.2}$$

$$I_y \overset{\circ}{=} l_0(U\dot{x} - \dot{U}x) + \mu_0 y/r \overset{\circ}{=} k_y, \quad k_y = \text{const}, \tag{4.3}$$

and the angular momentum $\mathbf{L}$ (1.3) reduces to

$$L_z = (x\dot{y} - y\dot{x}) \overset{\circ}{=} l_0, \quad l_0 = \text{const}. \tag{4.4}$$

In terms of the plane polar coordinates r, ϕ, Eqs. (4.2), (4.3), and (4.4) take the respective forms

$$(\mu_0 - l_0 rU\dot{\phi})\cos\phi - l_0(U\dot{r} - \dot{U}r)\sin\phi \overset{\circ}{=} k_x, \tag{4.5}$$

$$(\mu_0 - l_0 rU\dot{\phi})\sin\phi + l_0(U\dot{r} - \dot{U}r)\cos\phi \overset{\circ}{=} k_y, \tag{4.6}$$

$$L(r, \dot{\phi}) \equiv r^2\dot{\phi} \overset{\circ}{=} l_0. \tag{4.7}$$

Equations (4.5) and (4.6) are solved for the coefficients of the $\sin\phi$ and $\cos\phi$ terms to obtain

$$\mu_0 - l_0^2 U/r \overset{\circ}{=} k_x \cos\phi + k_y \sin\phi, \tag{4.8}$$

$$l_0(U\dot{r} - \dot{U}r) \overset{\circ}{=} k_y \cos\phi - k_x \sin\phi, \tag{4.9}$$

where $\dot{\phi}$ was eliminated by use of (4.7).

Define the function $M(\phi)$ by

$$M(\phi) \equiv \mu_0 - k_x \cos\phi - k_y \sin\phi. \tag{4.10}$$

By means of (4.10), Eqs. (4.8) and (4.9) may be written in the respective forms

$$U(t) \overset{\circ}{=} rM(\phi)/l_0^2, \tag{4.11}$$

$$U\dot{r} - \dot{U}r \overset{\circ}{=} -M'/l_0, \tag{4.12}$$

where the prime indicates differentiation with respect to ϕ.

In order to obtain the orbit equation in the form $r = r(\phi)$, it is convenient to make the change in variable

$$r = 1/u, \tag{4.13}$$

where we regard $u = u(\phi)$. By means of (4.7) and (4.13) we find

$$\dot{\phi} \overset{\circ}{=} l_0 u^2. \tag{4.14}$$

By use of (4.13) and (4.14) it then follows that

$$\dot{r} \overset{\circ}{=} -l_0 u', \tag{4.15}$$

$$\ddot{r} \overset{\circ}{=} -l_0^2 u^2 u''. \tag{4.16}$$

Equations (4.11) and (4.12) transform into the respective forms

$$U(t) \overset{\circ}{=} M(\phi)/l_0^2 u, \tag{4.17}$$

$$u'M - uM' + l_0\dot{U} \overset{\circ}{=} 0. \tag{4.18}$$

When the dynamical equation (2.24) is expressed in the plane polar coordinates r, ϕ, we obtain

$$\ddot{r} - r\dot{\phi}^2 \overset{\circ}{=} (\ddot{U}/U)r - (1/U)\mu_0/r^2, \tag{4.19}$$

$$r\ddot{\phi} + 2\dot{r}\dot{\phi} \overset{\circ}{=} 0. \tag{4.20}$$

By means of (4.13), (4.14), (4.16), and (4.17) the dynamical equation (4.19) is transformed to

$$u'' + u \overset{\circ}{=} -(1/u^2 M)(\ddot{U} - \mu_0 u^3). \tag{4.21}$$

It is immediate that (4.20) integrates to (4.7) and hence to (4.14).

We shall now show how (4.17) and (4.18) can be used to derive the orbit equations.

In terms of the definition

$$w \equiv u/M \tag{4.22}$$

we express (4.17) in the form

$$U(t) \overset{\circ}{=} 1/l_0^2 w. \tag{4.23}$$

We assume (4.23) can be solved for t in terms of w in the form

$$t \overset{\circ}{=} \sigma(w). \tag{4.24}$$

Hence $\dot{U}(t)$ can be expressed in the form

$$\dot{U}(t) \overset{\circ}{=} \rho(w) \tag{4.25}$$

for some function $\rho(w)$.

We divide (4.18) by M^2 and make use of (4.22) and (4.25) to obtain

$$w' \overset{\circ}{=} -(l_0/M^2)\rho(w). \tag{4.26}$$

Equation (4.26) can be rewritten in the form

$$G(w) \overset{\circ}{=} \tau(\phi), \tag{4.27}$$

where

$$G(w) \equiv \int \frac{dw}{\rho(w)}, \tag{4.28}$$

$$\tau(\phi) \equiv -l_0 \int \frac{d\phi}{M^2} + k_0, \tag{4.29}$$

and k_0 is an arbitrary constant.[17]

From (4.27) we may write

$$w \overset{\circ}{=} Q(\tau) \tag{4.30}$$

for some function $Q(\tau)$.

By use of (4.13) and (4.22), we obtain from (4.30) the orbit equation in the $r = r(\phi)$ form

$$u = 1/r = M(\phi)Q[\tau(\phi)]. \tag{4.31}$$

It is not difficult to show that the orbit equation (4.31) satisfies the dynamical equation (4.21).

5. SEPARABLE FORCES

Preliminary to illustrating the procedure of Sec. 4 for obtaining orbits, we determine the functions $U(t)$ for which the time-dependent central force $F(r, t)$ ($\not\equiv 0$) [given in (1.7) and (2.24)] has the separable form

$$F(r, t) \equiv (\ddot{U}/U)r - (1/U)\mu_0/r^2 = T(t)R(r). \tag{5.1}$$

Note first that if $\mu_0 = 0$, then condition (5.1) is immediately satisfied. Hence in the analysis to follow we assume $\mu_0 \neq 0$.

Multiplication of (5.1) by $U(t)$ followed by differentiation of the result with respect to t leads to

$$\ddot{U}r = R(r)(\dot{U}T + U\dot{T}). \qquad (5.2)$$

In (5.2) consider the two cases

$$\text{(i)} \quad \ddot{U} = 0, \qquad (5.3)$$

$$\text{(ii)} \quad \ddot{U} \neq 0. \qquad (5.4)$$

For case (i)

$$U(t) = at^2 + bt + c, \quad a, b, c = \text{const}. \qquad (5.5)$$

It follows from (5.5) and (5.2) [since $R(r) \neq 0$] that $T(t)$ may be expressed in the form

$$T(t) = \lambda_0/U = \lambda_0/(at^2 + bt + c), \quad \lambda_0 = \text{const} \neq 0. \qquad (5.6)$$

Hence from (5.6) and (5.1) we find

$$R(r) = (1/\lambda_0)(2ar - \mu_0/r^2). \qquad (5.7)$$

It follows from (5.6), (5.7), and (5.1) that for case (i)

$$F(r, t) = [1/(at^2 + bt + c)](ar - \mu_0/r^2). \qquad (5.8)$$

For case (ii) (which assumes $\ddot{U} \neq 0$) we find from (5.2) that [note from (5.2) and (5.4) that $\dot{U}T + U\dot{T} \neq 0$]

$$\ddot{U}/(\dot{U}T + U\dot{T}) = R(r)/r = \alpha_0, \quad \alpha_0 = \text{const} \neq 0. \quad (5.9)$$

From (5.9) it is seen that

$$R(r) = \alpha_0 r. \qquad (5.10)$$

Use of (5.10) in (5.1) leads to

$$\ddot{U}r - \mu_0/r^2 = \alpha_0 UTr, \qquad (5.11)$$

which may be rewritten in the separable form

$$\ddot{U} - \alpha_0 UT = \mu_0/r^3. \qquad (5.12)$$

It follows from (5.12) that

$$r = r_0 = \text{const}. \qquad (5.13)$$

Hence from (5.10)

$$R(r) = \alpha_0 r_0. \qquad (5.14)$$

From (5.13) and (4.7) it follows that

$$\dot{\phi} = l_0/r_0^2 = \text{const}. \qquad (5.15)$$

Hence form (4.19), (5.1) (5.13), and (5.15) we find

$$\frac{\ddot{U}}{U} r_0 - \frac{1}{U}\frac{\mu_0}{r_0^2} = -\frac{l_0^2}{r_0^3}. \qquad (5.16)$$

From (5.16) we obtain

$$\ddot{U} + \omega_0^2 U = \mu_0/r_0^3, \quad \omega_0^2 \equiv l_0^2/r_0^4. \qquad (5.17)$$

Equation (5.17) gives

$$U(t) = \lambda_1 \cos \omega_0 t + \lambda_2 \sin \omega_0 t + r_0 \mu_0/l_0^2, \quad \lambda_1, \lambda_2 = \text{consts}. \qquad (5.18)$$

Comparison of (5.12), and (5.17) with use of (5.13) shows

$$T = -l_0^2/\alpha_0 r_0^4. \qquad (5.19)$$

It is easily verified that (5.14), (5.16), and (5.19) are consistent with (5.1) [with $r = r_0$].

We may now state the theorem to follow.

Theorem 5.1: A necessary and sufficient condition that the dynamical equation

$$\ddot{\mathbf{r}} = \hat{\imath}_r \left[\frac{\ddot{U}}{U} r - \frac{1}{U}\frac{\mu_0}{r^2} \right], \quad U = U(t), \qquad (2.24')$$

be expressible in the separable form

$$\ddot{r} = \hat{\imath}_r T(t) R(r) \qquad (5.20)$$

is either

(i) the function $U(t)$ has the form

$$U(t) = at^2 + bt + c, \quad a, b, c \equiv \text{consts}, \qquad (5.5')$$

in which case the dynamical equation (2.24′) takes the form

$$\ddot{\mathbf{r}} = \hat{\imath}_r [1/(at^2 + bt + c)](ar - \mu_0/r^2); \qquad (5.21)$$

or

(ii) the function $U(t)$ has the form

$$U(t) = \lambda_1 \cos \omega_0 t + \lambda_2 \sin \omega_0 t + r_0 \mu_0/l_0^2, \quad \omega_0^2 \equiv l_0^2/r_0^4, \quad \lambda_1, \lambda_2 \equiv \text{const}, \qquad (5.18')$$

and

$$r = r_0 = \text{const}, \qquad (5.13')$$

in which case the dynamical equation (2.24′) takes the form

$$\ddot{\mathbf{r}} = -\hat{\imath}_r l_0^2/r_0^3. \quad \square \qquad (5.22)$$

Remark: The necessity was proved above, and the sufficiency easily follows.

6. ILLUSTRATIVE ORBIT EXAMPLES

Example 1: $U(t) \equiv at^2$, $a > 0$

As our first example we choose

$$U(t) \equiv at^2, \; a > 0. \qquad (6.1)$$

This choice for $U(t)$ is a special case of (5.5) which leads to a separable force $F(r, t)$ [obtained from (5.8) with $b = c = 0$]. Use of (6.1) in (2.24) gives the dynamical equation

$$\ddot{\mathbf{r}} = \hat{\imath}_r (1/at^2)(2ar - \mu_0/r^2). \qquad (6.2)$$

We now obtain the orbit equation for the dynamical system (6.2). Following the discussion of Sec. 4, we have from (4.22), (4.23), and (6.1) that

$$at^2 = 1/l_0^2 w. \qquad (6.3)$$

Equation (4.24) now takes the form

$$t = 1/(al_0^2 w)^{1/2} \equiv \sigma(w). \qquad (6.4)$$

From (6.1) and (6.4) we find (4.25) becomes

$$\dot{U}(t) = 2a/(al_0^2 w)^{1/2} \equiv \rho(w). \qquad (6.5)$$

Equation (4.26) now takes the form

$$w' \overset{\circ}{=} -\frac{l_0}{M^2}\frac{2a}{(al_0^2 w)^{1/2}}, \quad \left(w' \equiv \frac{dw}{d\phi} \right). \qquad (6.6)$$

[Refer to (4.10) for the definition of $M = M(\phi)$.] By integration of (6.6) we have

$$\int \frac{(al_0^2 w)^{1/2}}{2a} dw = -l_0 \int \frac{d\phi}{M^2} + k_0. \qquad (6.7)$$

Hence, by inspection of (6.7), Eqs. (4.28) and (4.29) take the respective forms

$$G(w) = \int \frac{(al_0^2 w)^{1/2}}{2a} dw, \qquad (6.8)$$

$$\tau(\phi) = -l_0 \int \frac{d\phi}{M^2} + k_0. \tag{6.9}$$

By carrying out the integration of the left-hand side of (6.7) we obtain

$$w = -3a^{1/2} \int \frac{d\phi}{M^2} + c_0,$$
$$c_0 \equiv \frac{3k_0 a^{1/2}}{l_0}. \tag{6.10}$$

[Equation (6.10) should be compared with (4.30).] By use of (4.10), (4.13), and (4.22) in (6.10) we obtain the explicit orbit equation (4.31) in the form

$$\frac{1}{r} = (\mu_0 - k_x \cos\phi - k_y \sin\phi) \times \left[-3a^{1/2} \int \frac{d\phi}{(\mu_0 - k_x \cos\phi - k_y \sin\phi)^2} + c_0 \right]^{2/3}. \tag{6.11}$$

Example 2: $U(t) \equiv ae^{bt}$, $a > 0$

As a second example we choose

$$U(t) \equiv ae^{bt}, \quad a > 0. \tag{6.12}$$

In this case the dynamical equation (2.24) becomes

$$\ddot{\mathbf{r}} = \hat{\imath}_r [b^2 r - (e^{-bt}/a)\mu_0/r^2]. \tag{6.13}$$

The orbit analysis proceeds as follows. We find from (4.22), (4.23), and (6.12) that [corresponding to (4.24)]

$$t = -(1/b)\ln(awl_0^2) \equiv \sigma(w). \tag{6.14}$$

From (6.12) and (6.14) we obtain

$$\dot{U} = b/l_0^2 w \equiv \rho(w). \tag{6.15}$$

Hence (4.26) takes the form

$$\frac{dw}{d\phi} = -\frac{b}{l_0 w M^2}, \tag{6.16}$$

from which we find

$$w = \left[-\frac{2b}{l_0} \int \frac{d\phi}{M^2} + l_0 \right]^{1/2}, \quad k_0 = \text{const.} \tag{6.17}$$

From (4.22) and $u = 1/r$ we find from (6.17) and (4.10) the orbit equation

$$\frac{1}{r} = (\mu_0 - k_x \cos\phi - k_y \sin\phi) \times \left[-\frac{2b}{l_0} \int \frac{d\phi}{(\mu_0 - k_x \cos\phi - k_y \sin\phi)^2} + k_0 \right]^{1/2}. \tag{6.18}$$

7. LEWIS–LEACH TECHNIQUE APPLIED TO THE ONE-DIMENSIONAL EFFECTIVE SYSTEM ASSOCIATED WITH $\ddot{\mathbf{r}} = \hat{\imath}_r[(\ddot{U}/U)r - (1/U)\mu_0/r^2)]$

When expressed in plane polar coordinates the two-dimensional central force dynamical equation (2.24) leads to (4.19) and the ever-present angular momentum integral (4.7). By use of (4.7) $\dot{\phi}$ may be eliminated from (4.19) to obtain

$$\ddot{r} - \frac{l_0^2}{r^3} = \frac{\ddot{U}}{U} r - \frac{1}{U}\frac{\mu_0}{r^2}. \tag{7.1}$$

Equation (7.1) may be regarded as a one-dimensional dynamical system with effective potential[1]

$$V_E(r,t) \equiv \frac{l_0^2}{2r^2} - \frac{1}{2}\frac{\ddot{U}}{U} r^2 - \frac{\mu_0}{Ur}, \tag{7.2}$$

wherein we assume $l_0 \neq 0$.

From (7.2) it follows that the effective dynamical system (7.1) is obtainable from the Lagrangian $\mathscr{L}_E(\dot{r}, r, t)$

$$\mathscr{L}_E \equiv \frac{1}{2}\dot{r}^2 + \frac{l_0^2}{2r^2} + \frac{1}{2}\frac{\ddot{U}}{U} r^2 + \frac{1}{U}\frac{\mu_0}{r}. \tag{7.3}$$

From (7.3) we immediately find the Hamiltonian $H_E(q, p, t)$ of the effective dynamical system (7.1) to be

$$H_E = \frac{1}{2}p^2 + \frac{l_0^2}{2r^2} - \frac{1}{2}\frac{\ddot{U}}{U} r^2 - \frac{1}{U}\frac{\mu_0}{r}, \tag{7.4}$$

wherein $q \equiv r$ and $p = \dot{r}$.

Any constant of motion of the one-dimensional effective dynamical system characterized by the Lagrangian (7.3) or the Hamiltonian (7.4) must be a constant of motion of the two-dimensional system (2.24).

Over the past several years a series of papers (see Refs. 7, 18–21, and references contained therein) have appeared which deal with techniques of formulating constants of motion for certain types of one-dimensional dynamical systems with explicit time dependence. Of particular interest here is a constant of motion formula, recently published by Lewis and Leach,[7] for a class of one-dimensional dynamical systems which could possibly include the effective dynamical system characterized by the Hamiltonian (7.4). It is of interest therefore to determine if the Lewis–Leach formula is applicable to the effective one-dimensional dynamical system (7.4) and, if so, to compare the constants of motion so obtained with those obtained in the preceding sections for the associated two-dimensional system (2.24).

The Lewis and Leach formalism deals with one-dimensional dynamical systems characterized by Hamiltonians of the form

$$H = \tfrac{1}{2}p^2 + f[a(t),t]\alpha(t)(\tfrac{1}{2}c_1 q^2 - c_0 q) + \alpha^2(t) W(u), \tag{7.5}$$

where

$$u \equiv \alpha(t) q + \beta(t), \tag{7.6}$$

$$\alpha(t) \equiv [c_2 - c_1 a(t)]^{-1}, \tag{7.7}$$

$$\beta(t) \equiv -c_0[c_1 + c_2 a(t)]\alpha(t); \tag{7.8}$$

the function $a(t)$ satisfies the differential equation

$$\ddot{a}(t) = f[a(t), t] \tag{7.9}$$

and c_0, c_1, c_2 are constants such that

$$c_1^2 + c_2^2 = 1. \tag{7.10}$$

They have shown that such dynamical systems will admit a constant of motion P given by

$$P \equiv \tfrac{1}{2}[p/\alpha(t) + \dot{a}(t)(c_1 q - c_0)]^2 + W(u). \tag{7.11}$$

To apply the Lewis–Leach theory to our effective one-dimensional system (7.1), we must first determine if the effective Hamiltonian H_E (7.4) is a member of the class of Hamiltonians (7.5). Identifying

$$q \equiv r, \tag{7.12}$$

we find

$$p = \dot{r}. \tag{7.13}$$

Comparison of (7.4) with (7.5) leads to the requirement that

$$\frac{l_0^2}{2r^2} - \frac{r^2\ddot{U}(t)}{2U(t)} - \frac{\mu_0}{rU(t)} = (\tfrac{1}{2}c_1r^2 - c_0r)\alpha(t)f[a(t), t] + \alpha^2(t)W(u). \tag{7.14}$$

The solution to (7.14) may be divided into two cases: Case I, $\mu_0 \neq 0$, and Case II, $\mu_0 = 0$.

Case I, $\mu_0 \neq 0$

As shown in the Appendix for the case $\mu_0 \neq 0$ (7.14) leads to the necessary and sufficient conditions

$$a(t) = c_2/c_1 - U(t)/c_1\lambda_0, \quad c_1^2 + c_2^2 = 1, \quad c_1, \lambda_0 = \text{nonzero consts}, \tag{7.15}$$

$$f[a(t), t] = -\ddot{U}(t)/c_1\lambda_0, \tag{7.16}$$

$$\alpha(t) = \lambda_0/U(t), \tag{7.17}$$

$$c_0 = 0, \tag{7.18}$$

$$W(u) = l_0^2/2u^2 - \mu_0/\lambda_0 u, \tag{7.19}$$

$$u = \lambda_0 r/U(t). \tag{7.20}$$

By means of (7.13), (7.15), and (7.17)–(7.20) we evaluate the Lewis–Leach constant of motion P [(7.11)] for the effective one-dimensional dynamical system (7.1) to obtain for the case $\mu_0 \neq 0$

$$P_E(\dot{r}, r, t; l_0) = (1/\lambda_0^2)[\tfrac{1}{2}(U\dot{r} - \dot{U}r)^2 + l_0^2U^2/2r^2 - \mu_0U/r], \quad \mu_0 \neq 0. \tag{7.21}$$

Next we establish a relationship between the above-obtained $P_E(\dot{r}, r, t; l_0)$ [(7.21)] and the Noether constant of motion $I_N(b_2)$ [(3.50)] of the associated two-dimensional dynamical system (7.24). It is easily shown that in plane-polar coordinates the constant of motion $I_N(b_2)$ (3.50) is expressible in the form

$$I_N(b_2) \equiv I_N(\dot{r}, r, \dot{\phi}, t) = -[\tfrac{1}{2}(U\dot{r} - \dot{U}r)^2 + \tfrac{1}{2}r^2\dot{\phi}^2U^2 - \mu_0U/r]. \tag{7.22}$$

By use of the angular momentum integral (4.7) we find from (7.22) that

$$I_N(\dot{r}, r, \dot{\phi}, t) = I_N(\dot{r}, r, L/r^2, t) \equiv I_N^*(\dot{r}, r, t, L) \overset{\circ}{=} I_N^*(\dot{r}, r, t, l_0) = -\lambda_0 P_E(\dot{r}, r, t; l_0). \tag{7.23}$$

Hence the Noether constant of motion $I_N(b_2)$ (3.50) for the two-dimensional system (2.24) reduces on a dynamical path to the one-dimensional constant of motion P_E (7.21) (obtained by the Lewis–Leach procedure[7]) for the effective one-dimensional dynamical system (7.1). Conversely the Lewis–Leach one-dimensional constant of motion P_E can be converted by use of the angular momentum integral into the two-dimensional Noether constant of motion $I_N(b_2)$.

It is clear for this $\mu_0 \neq 0$ case that the Lewis–Leach procedure when applied to the effective one-dimensional system (7.1) did not lead to the components I_x (4.2) and I_y (4.3) of the vector constant of motion I [(2.25)]. By means of (7.23) it follows however that the Lewis–Leach constant of motion P_E (7.21) can be related to the sum of the squares of the components of the vector constant of motion **I** through the functional dependency relation (3.52).

Case II, $\mu_0 = 0$

With $\mu_0 = 0$, the dynamical system (2.24) reduces to the time-dependent harmonic oscillator (2.26). For this $\mu_0 = 0$ case (as shown in the Appendix) we find that, in order to satisfy (7.14), it is necessary and sufficient that

$$a(t) = c_2/c_1 - (1/c_1)\rho(t), \quad c_1^2 + c_2^2 = 1, \quad c_1 = \text{nonzero const}, \tag{7.24}$$

$$f[a(t), t] = -(1/c_1)\ddot{\rho}(t), \tag{7.25}$$

$$\alpha(t) = \rho^{-1}(t), \tag{7.26}$$

$$c_0 = 0, \tag{7.27}$$

$$W(u) = l_0^2/2u^2 + \tfrac{1}{4}\tau_0u^2, \quad \tau_0 = \text{const}, \tag{7.28}$$

$$u = r\rho^{-1}(t), \tag{7.29}$$

where $\rho(t)$ must satisfy[22]

$$\ddot{\rho} - (\ddot{U}/U)\rho - \tfrac{1}{2}\tau_0/\rho^3 = 0. \tag{7.30}$$

By use of (7.12), (7.13), (7.24), and (7.26)–(7.30), we evaluate the Lewis–Leach constant of motion P [(7.11)] for the effective one-dimensional system associated with the two-dimensional time-dependent oscillator (2.26) and obtain

$$P_E(\dot{r}, r, t; l_0, \tau_0) = \tfrac{1}{2}(\rho\dot{r} - \dot{\rho}r)^2 + (l_0^2/2r^2)\rho^2 + \tfrac{1}{4}\tau_0r^2/\rho^2, \quad \mu_0 = 0. \tag{7.31}$$

Since the function $\rho(t)$ appearing in (7.31) is determined by the second-order differential equation (7.30), it follows that there will be two $\rho(t)$ solutions and hence two concomitant constants of motion given by (7.31).

Note that for the case $\tau_0 = 0$ (7.30) has the two solutions

$$\rho_1 = U, \quad \rho_2 = US, \tag{7.32}$$

where S is defined by (2.30).

We next show how the Lewis–Leach constant of motion P_E (7.31) of the effective one-dimensional dynamical system associated with the two-dimensional time-dependent oscillator is related to a Noether constant of motion of the two-dimensional system.

The Noether constant of motion $C_2(B)$ [Ref. 9, Eq. (6.24)] referred to in Remark 4 of Sec. 3 may by a change of variable be expressed in an alternative form given by Ref. 9, Eq. (7.5). If this alternative form for the constant of motion $C_2(B)$ is then expressed in plane-polar coordinates and the angular variable $\dot{\phi}$ eliminated by use of the angular momentum integral (4.7), *the resulting form of the constant of motion* $C_2(B)$ *is identical to* P_E (7.31).

It is of interest to note that the condition [Ref. 9, Eq. (7.3)] for the existence of the Noether integral of the two-dimensional time-dependent oscillator is identical to (7.30), which was obtained as one of the conditions of compatibility of the effective Hamiltonian (7.4), for the case $\mu_0 = 0$, with the one-dimensional Lewis–Leach Hamiltonian (7.5).

APPENDIX: SOLUTION OF (7.14)

We form the third derivative of (7.14) with respect to r [where we take into account the r dependence of u by means of (7.6) and (7.12)] and obtain

$$\frac{-12l_0^4}{r^5}+\frac{6\mu_0}{r^4U}=\alpha^5W'''(u)\quad\left(W'\equiv\frac{dW}{du}\right).\qquad\text{(A1)}$$

Consistent with (7.6) we make the change in variables from (r,t) to $(u,\bar t)$ given by

$$r=[u-\beta(\bar t)]/\alpha(\bar t),\quad t=\bar t,\qquad\text{(A2)}$$

and express (A1) in the form

$$-12l_0^2+6\mu_0(u-\beta)/\alpha U=(u-\beta)^5W'''(u).\qquad\text{(A3)}$$

By forming the second derivative of (A3) with respect to u, we obtain

$$(u-\beta)^2W^{(5)}(u)+10(u-\beta)W^{(4)}(u)+20W'''(u)=0.\qquad\text{(A4)}$$

We differentiate (A4) with respect to $\bar t$ $(\equiv t)$ to obtain

$$[(u-\beta)W^{(5)}(u)+5W^{(4)}(u)]\dot\beta(t)=0.\qquad\text{(A5)}$$

First assume $\dot\beta\neq0$ in (A5). This implies

$$[u-\beta(t)]W^{(5)}(u)+5W^{(4)}(u)=0.\qquad\text{(A6)}$$

Differentiation of (A6) with respect to t implies $W^{(5)}(u)=0$, which with (A6) implies $W^{(4)}(u)=0$. Hence, by (A4), $W'''(u)=0$. It then follows from differentiation of (A3) with respect to u that $\mu_0=0$. With $\mu_0=0$ and $W'''(u)=0$ we find from (A1) that $l_0=0$, an excluded case. Hence $\dot\beta=0$ and $\beta=\beta_0=$ const.

By means of (7.8) the condition $\dot\beta=0$ implies

$$c_0\dot\alpha(t)=0.\qquad\text{(A7)}$$

If we assume $\dot\alpha=0$, then, by (7.7), $\alpha(t)=\alpha_0=$ const; by (7.9), $f[a(t),t]=0$; and by (7.6), (7.12) and the constancy of β, $u=\alpha_0r+\beta_0$. With this form of $u(r)$ and $\alpha=\alpha_0$, $\beta=\beta_0$ we find that differentiation of (A3) with respect to t gives $\mu_0\dot U=0$. We exclude the choice $\dot U(t)=0$ in order to retain the explicit time dependence of the dynamical system. Hence the above assumption of $\dot\alpha=0$ implies that we must have $\mu_0=0$. It therefore follows, if $\dot\alpha=0$, that (7.14) reduces to

$$l_0^2/2r^2-r^2\ddot U/2U=\alpha_0^2W(u).\qquad\text{(A8)}$$

From (A8) it follows that $\ddot U/U=$ const, which with the above requirement that $\mu_0=0$, would imply from (2.24) that the dynamical system is reduced to a *time-independent* harmonic oscillator. Hence we must exclude the case $\dot\alpha=0$. Therefore, from (A7) we must have $c_0=0$.

With $c_0=0$ we obtain form (7.8) that $\beta=0$ and hence (7.6) reduces to [with use of (7.12)]

$$u=\alpha(t)r.\qquad\text{(A9)}$$

With $\beta=0$ it follows that (A3) can be written in the form

$$6\mu_0/\alpha(t)U(t)=u^4W'''(u)+12l_0^2/u,\qquad\text{(A10)}$$

(where we still consider u as an independent variable).

At this point we consider the two possibilities: Case I, $\mu_0\neq0$, and Case II; $\mu_0=0$.

Case I, $\mu_0\neq0$

It follows from (A10) that

$$\alpha(t)U(t)=\lambda_0,\quad\lambda_0\equiv\text{const}\neq0.\qquad\text{(A11)}$$

From (7.7) and (A11) we have

$$c_2-c_1a(t)=U(t)/\lambda_0.\qquad\text{(A12)}$$

We must therefore exclude the case $c_1=0$, since otherwise by (A12) we obtain $U=$ const.

From (A11) and (7.9) it follows that

$$f=-\ddot U/c_1\lambda_0.\qquad\text{(A13)}$$

Keeping in mind that $c_0=0$ we use (A13) to reduce (7.14) to the form

$$\tfrac12 l_0^2-\mu_0r/U=(\lambda_0r/U)^2W(u).\qquad\text{(A14)}$$

By means of (A9) and (A11) we find $u=\lambda_0r/U$, which allows (A14) to be written in the form

$$W(u)=l_0^2/2u^2-\mu_0/\lambda_0u.\qquad\text{(A15)}$$

It may be easily verified that (A10) is satisfied by (A11) and (A15).

Hence for the case $\mu_0\neq0$ to tailor the Hamiltonian H [(7.5)], employed in the Lewis–Leach formalism, to the form of the Hamiltonian H_E [(7.4)], which characterizes the effective one-dimensional time-dependent system (7.1), it is necessary and sufficient that (7.15)–(7.20) hold.

Case II, $\mu_0=0$

With $\mu_0=0$ and (A9) we may rewrite (A1) in the form

$$W'''(u)=-12l_0^2/u^5.\qquad\text{(A16)}$$

From (A16) we obtain

$$W(u)=l_0^2/2u^2+\tfrac14\tau_0u^2+\tau_1u+\tau_2,\quad\tau_0,\tau_1,\tau_2\equiv\text{const.}\qquad\text{(A17)}$$

From the work prior to Case I, it was shown that $c_0=0$. Hence, if in (7.14) we use $c_0=0$, $\mu_0=0$, (7.9), (A17), and (A9), we obtain

$$(\tfrac12\ddot U/U+\tfrac12c_1\alpha\ddot a+\tfrac14\tau_0\alpha^4)r^2+(\tau_1\alpha^3)r+\tau_2\alpha^2=0.\qquad\text{(A18)}$$

It follows from (A18) [$\alpha(t)\neq0$; refer to (7.7)] that

$$\tau_1=0,\quad\tau_2=0,\qquad\text{(A19)}$$

and

$$\tfrac12\ddot U/U+\tfrac12c_1\alpha\ddot a+\tfrac14\tau_0\alpha^4=0.\qquad\text{(A20)}$$

By means of (7.7) and the definition

$$\rho(t)\equiv1/\alpha(t)=c_2-c_1a(t),\qquad\text{(A21)}$$

we find (A20) can be transformed into the form

$$\ddot\rho-\tfrac12\tau_0/\rho^3-(\ddot U/U)\rho=0.\qquad\text{(A22)}$$

The function $W(u)$ [(A17)] may be simplified by the use of (A19). From (A21) we may solve for $a(t)$ in terms of $\rho(t)$ and thereby calculate $f[a(t),t]$ by (7.9). By (A21) the function u (A9) may also be expressed in terms of $\rho(t)$.

The analysis of the $\mu_0=0$ case may be summarized by (7.24)–(7.30).

[1]H. Goldstein, *Classical Mechanics*, 2nd ed. (Addison-Wesley, Reading, MA, 1980).
[2]The coordinates x^i denote rectangular coordinates in Euclidean space. Unless indicated otherwise, the Einstein summation notation is employed. A dot over a symbol indicates total differentiation with respect to the time t. A comma indicates partial differentiation. [If $Z = Z(x^1,...,x^n)$, then $Z_{,i} = \partial Z/\partial x^i$; if $Z = Z(r, t)$, then $Z_{,r} \equiv \partial Z/\partial r, Z_{,t} \equiv \partial Z/\partial t$.] Unless indicated otherwise, a primed symbol denotes differentiation with respect to the angular variable ϕ. The symbol ($\doteq$) indicates equality on a dynamical path, i.e., for those $x^i = x^i(t)$ which are solutions of the dynamical equations (1.7).
[3]For simplicity we have assumed a unit mass.
[4]G. H. Katzin and J. Levine, J. Math. Phys. **23**, 552 (1982).
[5]While the paper of Ref. 4 was in press a paper [L. M. Berković, Cel. Mech. **24**, 407 (1981)] appeared on an analysis of the time-dependent Kepler system based upon transformations of the dynamical equations. This paper contains a good bibliography on the time-dependent Kepler problem.
[6]In addition to the vector constant of motion (1.6), and the angular momenta constants of motion, the system (1.4), (1.5) was shown to admit a "generalized energy integral" with explicit time dependence.
[7]H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 165 (1982).
[8]H. R. Lewis, Jr. [(a) Phys. Rev. Lett. **18**, 510 (1967); (b) J. Math Phys. **9**, 1976 (1968); (c) Phys. Rev. **172**, 1313 (1968)] showed that a one-dimensional time-dependent oscillator admits an exact time-dependent constant of motion.
[9]G. H. Katzin and J. Levine [J. Math. Phys. **18**, 1267 (1977)] extended this result by a method based upon the existence of dynamical symmetries admitted by the n-dimensional time-dependent oscillator and obtained a set of time-dependent constants of motion admitted by such a dynamical system.
[10]The case $\mu_0 = 0$ reduces (3.1) to a time-dependent oscillator. The symmetries for this case were previously obtained. See Ref. 9.
[11]G. H. Katzin and J. Levine, (a) J. Math. Phys. **17**, 1345 (1976); (b) **18**, 424 (1977), and the paper in Ref. 9.
[12]See Ref. 9, Eqs. (3.17) and (3.18).
[13]See Ref. 11(a).
[14]It should be noted that these Noether symmetry mappings are *velocity-independent*. However, it is known that any (scalar) constant of motion of a Lagrangian dynamical system can be expressed in the Noether form (3.49) for some *velocity-dependent* Noether mapping which satisfies (3.29). For a discussion of Noether theory based upon velocity dependent mappings see W. Sarlet and F. Cantrijn, SIAM Rev. **23**, 467 (1981).
[15]M. Lutzky, Phys. Lett. A **68**, 3 (1978).
[16]P. G. L. Leach, J. Math. Phys. **21**, 300 (1980).
[17]An integral similar to (4.29) also appears in the orbit analysis of a time-dependent Kepler problem. See Ref. 4.
[18]H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 2371 (1982).
[19]W. Sarlet and J. R. Ray, J. Math. Phys. **22**, 2504 (1981).
[20]P. G. L. Leach, Phys. Lett. A **84**, 161 (1981).
[21]J. R. Ray and J. L. Reid, J. Math. Phys. **21**, 2054 (1979).
[22]See Refs. 8 in which $\tau_0 = 2$ was used. See also Ref. 9 in which $c_0 \equiv \tau_0$ is arbitrary.

Deformation of Hamiltonian dynamics and constants of motion in dissipative systems

J. Taborda Duarte and R. Vilela Mendes
CFMC—Instituto Nacional de Investigacão Científica, Av. Gama Pinto, 2-1699 Lisboa Codex, Portugal

(Received 23 March 1982; accepted for publication 27 August 1982)

A necessary condition for the existence of arcs of vector fields with constants of motion is found. The result is applied to arcs obtained by deformation of Hamiltonian dynamics and illustrated in the Van der Pol and Lorenz models.

PACS numbers: 03.20. + i

I. PRELIMINARIES

For the vast majority of nonlinear dynamical systems it is usually extremely difficult to extract analytical information from the defining equations. A possile exception arises when the system can be shown to belong to some family and one of the members of the family is either exactly solvable or has nice properties. Then, in some cases, one can use the simpler system to obtain information on the properties of the family. This is the situation in the perturbation theory around the linear approximation, where the validity of extrapolation from the linear to the perturbed system hinges on the smallness of some physical parameter.

Another example arises when in $\dot{x} = X(x)$ the vector field can be decomposed into components, each having well-studied properties; for example, $X = X^{(S)} + X^{(V)}$, $X^{(S)}$ being a gradient and $X^{(V)}$ a volume-conserving or even a Hamiltonian field. In this case one could, for example, define a family (arc of vector fields) $X_\epsilon = 2\epsilon X^{(S)} + 2(1-\epsilon)X^{(V)}$ which for $\epsilon = \frac{1}{2}$ coincides with the original system and for $\epsilon = 0$ and $\epsilon = 1$ is purely (V) on purely (S). Properties of the original system can therefore be obtained from the deformation of its components. In this case, one cannot rely on the smallness of the deformation parameter ϵ and the validity of extrapolations must depend on the differentiability properties (in ϵ) of the arc.

The main purpose in the research that led to this paper was to identify families of dynamics which would supply nonperturbative information about its members once the analytical behavior of one of them is known. In this sense one might say that the main result is the notion of "arc of vector fields with constants of motion" (Definition 2). Once one realizes that the ingredients in the definition are what one needs to carry nonperturbative information along the family, the remaining results are a matter of computation using the differentiability properties of composite maps.

In Sec. 2 a necessary condition is obtained for the existence of the arcs (Theorem 1), which is then particularized for arcs obtained by deformation of Hamiltonian dynamics (Theorem 2).

As far as applications are concerned, we suggest that our results might be useful to help understand the transition between conservative and dissipative regimes.

When dissipation is added to a conservative system, there is typically a reduction of the phase space with the system tending at $t \to \infty$ to an attracting subset of lower dimension. Looking at dissipative systems as belonging to some family obtained by deformation of a conservative dynamics, one might hope to find which subsets in the conservative phase space do not change qualitatively when dissipation is turned on. In other words, although conservative dynamics is structurally unstable, it may happen that when restricted to some subsets of the phase space it remains stable for dissipative deformations. Finding the deformation stable subsets would supply analytical information on the nature and approximate location of the attractors.

Although the purpose of this paper is not to make a full exploration of the mathematical results of Sec. II, we have included as an illustration in Sec. III a study of the van der Pol and Lorenz models which were chosen for their simplicity and availability of numerical studies.

To conclude this section a quick review is made of some mathematical results needed in the sequel. All the material can be found with proofs in Ref. 1.

Let E and F be Banach spaces and U an open set in E. We denote by $C_b^k(U,F)$ the set of C^k maps $f: U \to F$ with the first k derivatives $D^k f: U \to L_s^k(E,F)$ bounded in U, that is, $\exists$ $k(f) \in \mathbb{R}^+$ such that $\sup_{x\in U} \|D^i f(x)\| < k(f)$ for all $0 \leqslant i \leqslant k$.

For $F = E$ one also uses the notation $\mathscr{X}_b^k(U)$ instead of $C_b^k(U,E)$. In this case one thinks of $f \in \mathscr{X}_b^k(U)$ as a C^k-vector field.

For $f \in C_b^k(U;F)$ define the norm $\|f\|_k = \Sigma_{i=0}^k \sup_{x\in U} \|D^i f(x)\|$. We will use the following:

Theorem: (1) $C_b^k(U,F)$ is a Banach space. (2) (i) Let M be a compact interval in R and U an open set in F; then $C^k(M,U) = \{ f \in C^k(M,F) | f(M) \subset U \}$ is open in $C^k(M,F)$.

(ii) The map comp: $C_b^{r+s}(U,E) \times C^r(M;U) \to C^r(M;E)$ is a C^s map, its derivative being given by

$$D\,\text{comp}(f,g)(\gamma,h)(x) = Df(g(x))(h(x)) + \gamma(g(x)),$$

$$\left.\begin{matrix} f \in C_b^{r+s}(U,E), & \gamma \in C_b^{r+s}(U,E), \\ g \in C^r(M;U), & h \in C^r(M;U) \end{matrix}\right\}, \quad x \in M.$$

II. ARCS OF VECTOR FIELDS AND CONSTANTS OF MOTION

Definition 1: Let (M,X) be a differentiable dynamical system. *A constant of motion* of (M,X) is any differentiable function $\Phi: M \to R$ such that for some solution γ of X we have $\Phi \circ \gamma = \text{const}$.

This notion is a generalization of the concept of first integral. Many systems which have no nontrivial first integrals have constants of motion. Below we will define a class of dynamical systems which is of special interest to us.

Definition 2: Let M be a differentiable manifold. A family $\epsilon \mapsto X_\epsilon$, defined by associating with each $\epsilon \in I = [-a,a]$ a vector field over M, is called an *arc of vector fields with constants of motion* if the following conditions are satisfied:

(i) Each X_ϵ has a constant of motion ϕ_ϵ over a periodic solution γ_ϵ.

(ii) The constant of motion ϕ_0 of X_0 is a first integral in a neighborhood of γ_0.

(iii) The maps, $\epsilon \mapsto X_\epsilon$, $\epsilon \mapsto \gamma_\epsilon$, and $\epsilon \mapsto \phi_\epsilon$, resp. $I \to \mathscr{X}^1_b(U)$, $I \to C^1_b(\mathbb{R};U)$, and $I \to C^1_b(U;\mathbb{R})$, are C^1-differentiable, U being an open set in M.

Then we prove the following:

Theorem 1: Let $\epsilon \mapsto X_\epsilon$ be an arc of vector fields with constants of motion defined on an open set U of a Banach space E. Then there is an (X_0-dependent) nontrivial 2-form β on U such that

$$\int_{\gamma_0} i\left(\frac{d}{d\epsilon} X_{\epsilon|_{\epsilon=0}}\right)\beta = 0. \tag{2.1}$$

Remark: Eq. (2.1) can also be written as

$$\int_0^{T_0} i\left(\frac{d}{d\epsilon} X_{\epsilon|_{\epsilon=0}}\right)(d\phi_0)\gamma_0(t)\, dt = 0. \tag{2.2}$$

Proof: The steps used in the proof are:

(1) Take the ϵ derivative at $\epsilon = 0$.

(2) Take the t derivative.

(3) Introduce the assumption that ϕ_0 is a first integral.

Let $T(\epsilon)$ be the period of the periodic solutions γ_ϵ, $T = \sup_{\epsilon\in I} T(\epsilon)$ and denote $\widetilde{M} = [0,T]$.

First step: Consider the diagram

$$I \xrightarrow{A} C^2_b(U,\mathbb{R}) \times C^1(\widetilde{M},U) \xrightarrow{\text{comp}} C^1(\widetilde{M},\mathbb{R})$$

defined by

$$\epsilon \mapsto (\phi_\epsilon,\gamma_\epsilon) \mapsto \phi_\epsilon \circ \gamma_\epsilon.$$

Therefore

$$\frac{d}{d\epsilon}(\phi_\epsilon \circ \gamma_\epsilon)_{|_{\epsilon=0}} = \frac{d}{d\epsilon}(\text{comp}\circ A)(\epsilon)_{|_{\epsilon=0}}.$$

Applying the chain rule and the theorem quoted in Sec. I, we conclude that $(d/d\epsilon)(\phi_\epsilon \circ \gamma_\epsilon)_{|_{\epsilon=0}}$ is a $C^1(\widetilde{M},R)$ map that associates with each $t \in \widetilde{M}$ the real number

$$D\phi_0(\gamma_0(t))\left(\frac{d}{d\epsilon}\gamma_{\epsilon|_{\epsilon=0}}(t)\right) + \frac{d}{d\epsilon}\phi_{\epsilon|_{\epsilon=0}}(\gamma_0(t))$$

or, denoting by primes the ϵ derivatives,

$$\frac{d}{d\epsilon}(\phi_\epsilon \circ \gamma_\epsilon)_{|_{\epsilon=0}} = D\phi_0(\gamma_0(t))(\gamma_0'(t)) + \phi_0'(\gamma_0(t)) = K_0', \tag{2.3}$$

where the last equality follows from the definition of constant of motion $(\phi_\epsilon \circ \gamma_\epsilon)(t) = K_\epsilon$.

Second step: As $K \in C^2(I \times \widetilde{M})$, it follows that $(d/dt)\, K'_{\epsilon|_{\epsilon=0}}$ is identically zero.

Taking the t derivative of Eq. (2.3),

$$D\phi_0'(\gamma_0(t))[X_0(\gamma_0(t))] + D\phi_0(\gamma_0(t))(\dot{\gamma}_0'(t)) + D^2\phi_0(\gamma_0(t))(\dot{\gamma}_0(t),\gamma_0'(t)) = 0$$

where the dot denotes the t derivative.

From $\dot{\gamma}_0 = X_0 \circ \gamma_0$, as γ_0 is a solution to X_0, we compute $\dot{\gamma}_0'(t)$:

$$\dot{\gamma}_0'(t) = DX_0(\gamma_0(t))(\gamma_0'(t)) + X_0'(\gamma_0(t))$$

and obtain

$$D\phi_0'(\gamma_0(t))(X_0(\gamma_0(t))) + D\phi_0(\gamma_0(t))(DX_0(\gamma_0(t))(\gamma_0'(t))) + D\phi_0(\gamma_0(t))(X_0'(\gamma_0(t))) + D^2\phi_0(\gamma_0(t))(\dot{\gamma}_0(t),\gamma_0'(t)) = 0. \tag{2.4}$$

Third step: Because ϕ_0 is a first integral of X_0 in an open nbd V of γ_0, we have $\forall\, x \in V\; D\phi_0(x)(X_0(x)) = 0$; hence, deriving the map $x \mapsto D\phi_0(x)(X_0(x))$, we obtain for $x \in V$ and $y \in E$:

$$D\phi_0(x)(DX_0(x)(y)) + D^2\phi_0(x)(X_0(x),y) = 0.$$

In particular, for $x = \gamma_0(t)$ and $y = \gamma_0'(t)$,

$$D\phi_0(\gamma_0(t))(DX_0(\gamma_0(t))(\gamma_0'(t))) + D^2\phi_0(\gamma_0(t))(X_0(\gamma_0(t)),\gamma_0'(t)) = 0 \quad \forall\; t \in \widetilde{M}. \tag{2.5}$$

Equation (2.4) reduces then to

$$D\phi_0'(\gamma_0(t))(X_0(\gamma_0(t)) + D\phi_0(\gamma_0(t))(X_0'(\gamma_0(t))) = 0 \quad \forall\; t \in \widetilde{M}. \tag{2.6}$$

Denoting by i the interior multiplication of a differential form by a vector field:

$$i(X_0)(d\phi_0')(\gamma_0(t)) + i(X_0')(d\phi_0)(\gamma_0(t)) = 0. \tag{2.7}$$

Let T_0 be the period of γ_0. As $[0,T_0] \subset [0,T]$, we restrict ourselves to the interval $[0,T_0]$ and integrate

$$\int_0^{T_0} i(X_0)(d\phi_0')(\gamma_0(t))\, dt + \int_0^{T_0} i(X_0')(d\phi_0)(\gamma_0(t))\, dt = 0;$$

γ_0 being a solution of X_0, this becomes

$$\int_{\gamma_0} d\phi_0' + \int_0^{T_0} i(X_0')(d\phi_0)(\gamma_0(t))\, dt = 0,$$

and by Stokes' theorem

$$\int_0^{T_0} i(X_0')(d\phi_0)(\gamma_0(t))\, dt = 0. \tag{2.2}$$

Because ϕ_0 is a first integral of X_0,

$$i(X_0)(d\phi_0) = 0,$$

in a neighborhood V of γ_0, we can find a differential 2-form β such that

$$d\phi_0 = i(X_0)\beta \quad \text{on} \quad V,$$

and from Eq. (2.2) one obtains

$$\int_0^{T_0} [\, i(X_0')i(X_0)\beta\,](\gamma_0(t))\, dt = 0,$$

i.e.,

$$\int_{\gamma_0} i(X_0')\beta = 0. \tag{2.1}$$

In the theorem we have just proved X_0' may be any vector field with a first integral. Of special interest for the applications is the case where X_0 is an Hamiltonian field.

From Ref. 2 we recall the main result in that paper:

"If M is a diff manifold $X \in \mathscr{X}(M)$ and $x \in M$, then it is possible to find, on an nbd Ω of x, a Riemannian metric g and

$N-1$ sympletic forms ω_i such that in Ω, $X = X_s + \Sigma_{i=1}^{N-1} X_{H_i}$, where X_s is gradient w.r.t. g and the X_{H_i} Hamiltonian w.r.t. ω_i.

We can then obtain the following:

Theorem 2: Let X be a vector field on $U \subset \mathbb{R}^N$ and $\{X_S, X_{H_1},\dots,X_{H_{N-1}}\}$ its gradient and Hamiltonian components. If the family $\epsilon \mapsto X_\epsilon$ defined by $X_\epsilon = X_{H_i} + \epsilon(X_S + \Sigma_{k\neq i} X_{H_k})$ is an arc of vector fields with constants of motion, then there is a solution γ_0 of X_{H_i} such that

$$\int_0^{T_0} \left[(\nabla H_i \cdot \nabla S) + \sum_{k\neq i} \omega_k(\nabla H_i, \nabla H_k)\right] \gamma_0(t))\, dt = 0. \tag{2.8}$$

Proof:

$$\frac{d}{d\epsilon} X_{\epsilon|_{\epsilon=0}} = X_S + \sum_{k\neq i} X_{H_k};$$

hence

$$\int_0^{T_0} i(X_0')(dH_i)(\gamma_0(t))\, dt = 0.$$

Let us make explicit the integrand function

$$\begin{aligned} i(X_0')(dH_i) &= i\Big(X_S + \sum_{k\neq i} X_{H_k}\Big)(dH_i) \\ &= i(X_S)(dH_i) + \sum_{k\neq i} i(X_{H_k})(dH_i) \\ &= (dH_i)(X_S) + \sum_{k\neq i} i(X_{H_k})(dH_i), \end{aligned}$$

and the result follows from the equalities

$$i(X_{H_k})(dH_i) = \omega_k(\nabla H_i, \nabla H_k), \quad dH_i(X_s) = \nabla H_i \cdot \nabla S.$$

III. APPLICATIONS

Combining the decomposition results[2] and the theorems in the previous section, we define the following strategy for searching for constants of motion (and attractors) in dynamical systems:

(a) Decompose the system into gradient and Hamiltonian components $\{X = X_S + \Sigma_{i=1}^{N-1} X_{H_i}\}$.

(b) Identify the constants of motion of each one of the Hamiltonian components X_{H_i}. These will be the Hamiltonian H_i itself plus a certain set $\{\phi_k\}$.

(c) Look for the closed orbits of the Hamiltonian components that satisfy the conditions of the theorem, Eq. (2.2), i.e., in coordinatewise notation

$$\int_0^{T_0} \left\{(\nabla\phi_k \cdot \nabla S) + \sum_{i=1}^{N-1} \omega_i(\nabla\phi, \nabla H_i)\right\} dt = 0. \tag{2.2'}$$

In general, for an N-dimensional system, the set of orbits that satisfy (2.2') for each ϕ_k might span an $(N-1)$-dimensional subspace Ω_k. The nonempty intersections of subsets of the Ω_k, i.e., the set of orbits that satisfy (2.2') simultaneously for certain subsets of $\{\phi_k\}$, would then supply information about the topological dimension and approximate location of the attractors for the arcs of vector fields associated to X_{H_i}.

(d) At this point, and before one gets the impression that a sure recipe has been obtained to find analytical approximations to the constants of motion of any dynamical system, one should remember that Eq. (2.2') is only a *necessary* not a sufficient condition for the existence of an arc (in the sense of the Definition 2). By applying (2.2'), all one obtains are analytical approximations to the constants of motion of the arcs of vector fields associated to the components X_{H_i} of the system. Left open is the question of whether the system actually belongs to an arc of vector fields of its components, i.e., whether it satisfies the necesssary differentiability conditions in the deformation parameter.

Therefore, one should complement this study by other methods, for example, using this analysis to complement and interpret numerical studies.

(e) If the closed orbits of the Hamiltonian components do not cover the whole phase space, one might try other Hamiltonians to explore the remaining regions. A natural choice is to use blown-up versions of the X_{H_i} for $H = \lambda^2 H_i(x/\lambda)$.

Whereas in (c) the arc of vector fields to be used is $X_\epsilon = X_{H_i} + \epsilon(X_S + \Sigma_{k\neq i} X_{H_k})$, in the case of an Hamiltonian H that is not a component, the arc is

$$X_\epsilon = (1-\epsilon)X_H + \epsilon\Big(X_S + \sum_i X_{H_i}\Big) = (1-\epsilon)X_H + \epsilon X.$$

The result has the same form as Eq. (2.2').

These techniques will now be illustrated in the Van der Pol and Lorenz models.

For the Van der Pol oscillator

$$\dot x = y = \frac{\partial S}{\partial x} + \frac{\partial H}{\partial y}, \quad S = \alpha\Big(\frac{x^2}{2} - \frac{x^4}{12}\Big),$$

$$\dot y = \alpha(1-x^2)y - x = \frac{\partial S}{\partial y} - \frac{\partial H}{\partial x}, \quad H = \frac{x^2+y^2}{2} - \alpha\Big(x - \frac{x^3}{3}\Big)y.$$

For values of the Hamiltonian H greater than

$$H > \tfrac{1}{2}\big[-\tfrac{1}{3}\alpha^2(2+\sqrt{1+3/\alpha^2})(1-\sqrt{1+3/\alpha^2})^2 + 2 + \sqrt{1+3/\alpha^2}\,\big],$$

the Hamiltonian orbits are not closed. In anticipation of the fact that one may need to explore wider regions of phase space than those covered by the closed orbits of H, we use a blown-up function

$$H_\lambda = \lambda^2 H(x/\lambda, y/\lambda) = \frac{x^2+y^2}{2} - \alpha\Big(x - \frac{x^3}{3\lambda^2}\Big)y.$$

In this two dimensional case, for the orbit $2H_\lambda = K$, Eq. (2.8) reduces to

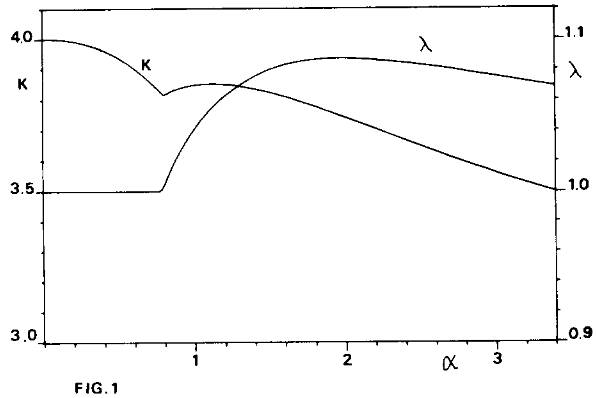


FIG. 1. Constant K for the Hamiltonian approximation to the limit cycle and blow-up parameter λ for the Van der Pol oscillator.

$$0=\int_A \Delta S=\alpha\int_A\int dxdy\,(1-x^2)$$
$$=4\alpha\int_0^{x_r} dx\,(1-x^2)\sqrt{\alpha^2x^2(1-x^2/3\lambda^2)^2-x^2+K}, \tag{3.1}$$

where the first equality follows from Stokes' theorem and x_r is the value for which the square root in the integrand vanishes.

The values of K and λ that satisfy Eq. (3.1) are plotted in the Fig. 1. One sees that for $\alpha > 0.75$ a certain amount of blowup is needed. Whenever $\lambda > 1$ is required, we have used the criterium of minimum blowup, i.e., we have chosen the smaller λ for which there is a closed orbit of $2H_\lambda$ satisfying Eq. (3.1). This means that for $\lambda > 1$ we use the separatrix between open and closed Hamiltonian orbits as an approximation to the limit cycle.

In the Figs. 2(a–c) we compare the exact limit cycle (dotted curve) obtained by numerical integration with our analytical Hamiltonian approximation

$$2H_\lambda = x^2+y^2-\alpha(x-x^3/3\lambda^2)y=K.$$

One sees that even for fairly large degrees of nonlinearity the overall shape and location of the limit cycle is reasonably approximated.

Our second example is the Lorenz model.[3] Extensive work on analytical approximations to the exact solutions of this model, in the limit of high Rayleigh number, has been done by Shimizu.[4] We will use the same change of variables as this author

$$x=\frac{X}{\sqrt{2\sigma(r-1)}}\quad m=\frac{Z}{r-1}-x^2,\quad \tau=t,$$

and write the Lorenz model $[\dot X=-\sigma X+\sigma Y$, $\dot Y=-Y+Xr-XZ, \dot Z=-bZ+XY]$ as

$$\dot x = p,$$
$$\dot p = -(\sigma+1)p-\sigma(r-1)x(x^2-1+m),$$
$$\dot m = -bm+(2\sigma-b)x^2.$$

Using these coordinates, our decomposition bear some formal resemblance to Shimizu's perturbation theory around the high Rayleigh number limit. This allowed some control and check on our results, which, however, are not restricted to any particular Rayleigh number range.

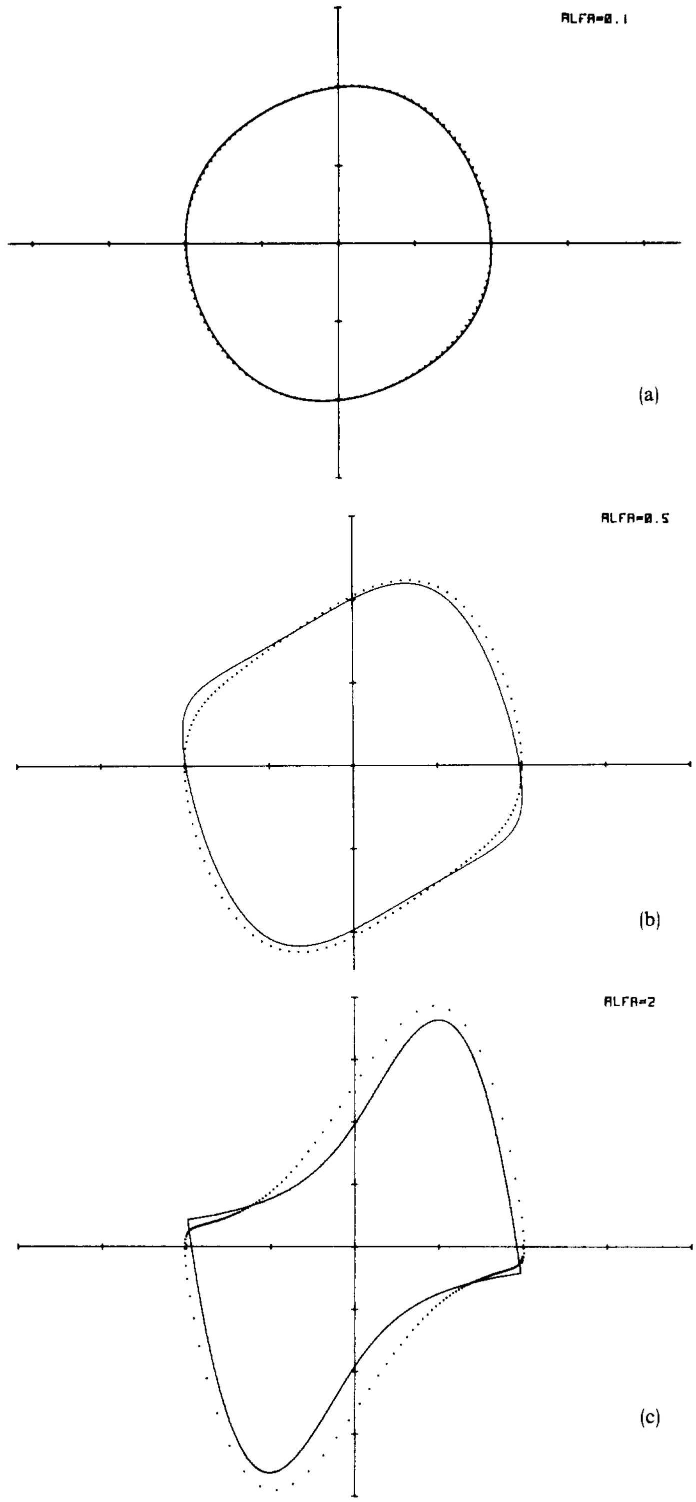
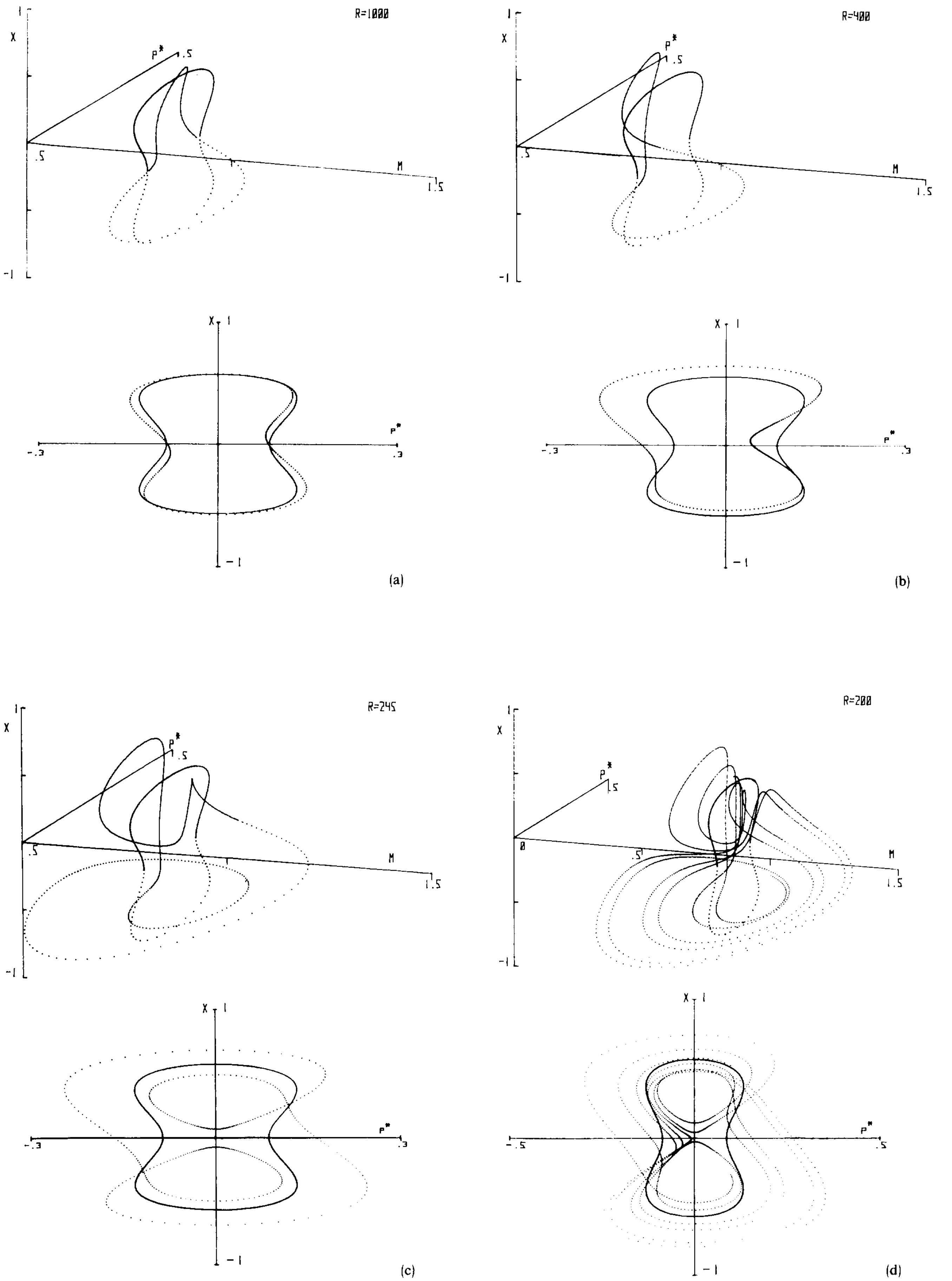


FIG. 2. Comparison of the limit cycle and the Hamiltonian orbit for three values of the nonlinearity parameter α in the Van der Pol.

Although odd-dimensional, the Lorenz model can be considered as imbedded in a space with one extra dimension and the same decomposition techniques applied. We will use



FIGS. 3(a)–3(d). Numerical solution and the Hamiltonian orbit for several Rayleigh values in the Lorenz model ($\sigma = 16$, $b = 4$).

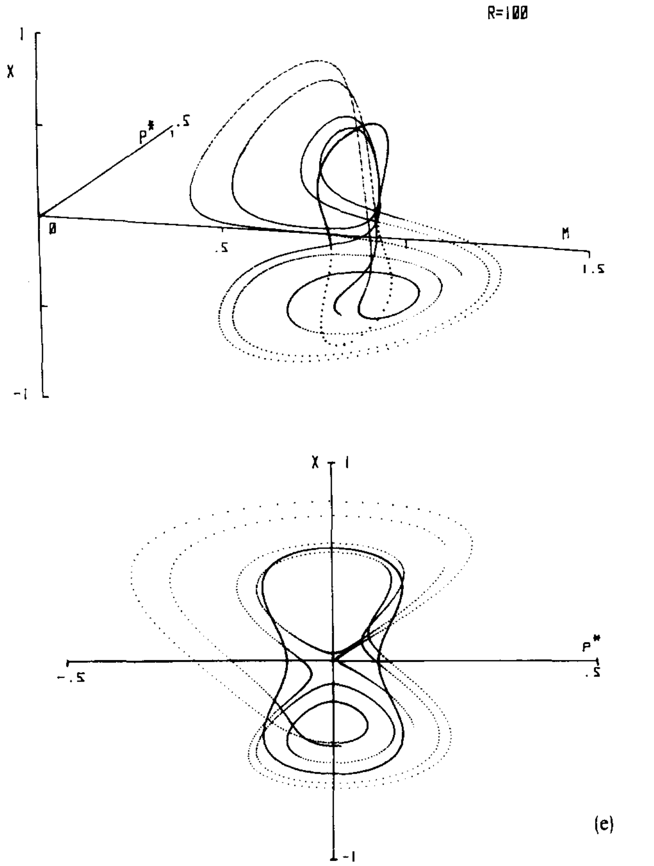


FIG. 3(e). Numerical solution and the Hamiltonian orbit for several Rayleigh values in the Lorentz model ($\sigma = 16$, $b = 4$).

the following decomposition:

$$\dot{x} = \frac{\partial S}{\partial x} + \frac{\partial H^1}{\partial p} + \frac{\partial H^2}{\partial m},$$
$$\dot{p} = \frac{\partial S}{\partial p} - \frac{\partial H^1}{\partial x}, \tag{3.2}$$
$$\dot{m} = \frac{\partial S}{\partial m} - \frac{\partial H^2}{\partial x},$$

with

$$S = -\tfrac{1}{2}(\sigma + 1)p^2 - \tfrac{1}{2}bm^2,$$
$$H^1 = \tfrac{1}{2}p^2 + \sigma(r-1)(\tfrac{1}{4}x^4 + \tfrac{1}{2}x^2(m-1)),$$
$$H^2 = -\tfrac{1}{3}(2\sigma - b)x^3.$$

The symplectic forms associated to the decomposition (3.2) are $\omega_1 = dx \wedge dp + dm \wedge dw$ and $\omega_2 = dx \wedge dm + dw \wedge dp$, w being the extra coordinate in the four-dimensional imbedding.

The Hamiltonian H^1 has closed orbits. Choosing H^1 to define the zero point in the arc, it follows from (3.2) that at $\epsilon = 0$ there are two constants of motion, namely

$$\phi_1 = H^1, \quad \phi_2 = m.$$

Application of the theorem leads to

$$0 = \int \{\nabla S \cdot \nabla H^1 + \omega_2(\nabla H^1, \nabla H^2)\}\, dt$$
$$= \int \{-(\sigma+1)p^2 + \tfrac{1}{2}\sigma(r-1)x^2[(2\sigma - b)x^2 - bm]\}\, dt. \tag{3.3a}$$

$$0 = \int [\nabla S \cdot \nabla m + \omega_1(\nabla m, \nabla H^1) + \omega_2(\nabla m, \nabla H^2)]\, dt$$
$$= \int [-bm + (2\sigma - b)x^2]\, dt. \tag{3.3b}$$

On the H^1 Hamiltonian orbits

$$p^2 + \sigma(r-1)[\tfrac{1}{2}x^4 + x^2(m-1)] = h,$$

where h is a constant. Making the replacement

$$dt = dx/2\sqrt{h - \sigma(r-1)[x^4/2 + x^2(m-1)]}$$

and the appropriate change of variables in Eqs. (3.3), they become

$$-(\sigma+1)hI_1 + \sigma(r-1)[(\sigma+1)(m-1) - \tfrac{1}{2}bm]I_2$$
$$+ \tfrac{1}{2}\sigma(r-1)(3\sigma - b + 1)I_4 = 0, \tag{3.4a}$$
$$-bmI_1 + (2\sigma - b)I_2 = 0, \tag{3.4b}$$

where

$$I_1 = \frac{2}{\sqrt{\alpha^2 + \beta^2}} K\left(\frac{\alpha^2}{\alpha^2 + \beta^2}\right), \tag{3.5a}$$
$$I_2 = -\beta^2 I_1 + 2\sqrt{\alpha^2 + \beta^2}\, E\left(\frac{\alpha^2}{\alpha^2 + \beta^2}\right), \tag{3.5b}$$
$$I_4 = \frac{2h}{3\sigma(r-1)} I_1 + \tfrac{4}{3}(1-m)I_2, \tag{3.5c}$$
$$\alpha^2 = (1-m) + \sqrt{(1-m)^2 + 2h/\sigma(r-1)} \tag{3.5d}$$
$$\beta^2 = -(1-m) + \sqrt{(1-m)^2 + 2h/\sigma(r-1)}. \tag{3.5e}$$

K and E are the complete elliptic integrals of first and second kind.[5]

Using Eqs. (3.5), Eqs. (3.4a)–(3.4b) are converted into the following equivalent set of equations:

$$\frac{2h}{\sigma(r-1)} = \frac{bm[(6\sigma - 2 - b)(1-m) - 3b]}{(2\sigma - b)(b+2)}, \tag{3.6a}$$
$$\frac{bm + (2\sigma - b)\beta^2}{\alpha^2 + \beta^2} K\left(\frac{\alpha^2}{\alpha^2 + \beta^2}\right) = (2\sigma - b)E\left(\frac{\alpha^2}{\alpha^2 + \beta^2}\right). \tag{3.6b}$$

Replacing (3.6a) into (3.5d) and (3.5e), one concludes that Eq. (3.6b) determines a solution for m independently of the Rayleigh number r. Once a solution for m is obtained numerically from Eq. (3.6b) for a given pair (σ,b), a solution for h is always obtained from (3.6a) for any r.

The existence of a *one-dimensional* Hamiltonian approximation to the constant of motion does *not* depend on the Rayleigh number. Referring back to our comments about topological dimension of attractors in (c) this sheds some doubt on the full turbulent nature of the r regions found in between the ranges of parameters for which limit cycles were found. It suggests that a (perhaps dense) set of periodic orbits might also exist in the turbulent regions. These remarks however, are only speculative, because of the limitations mentioned in (d).

For the popular values $\sigma = 16$ and $b = 4$, the numerical solution of Eq. (3.6b) leads to

$$m = 0.8586,$$
$$h = 0.11896(r-1). \tag{3.7}$$

In Figs. 3(a)–3(e) the Hamiltonian orbit

$$\frac{p^2}{\sigma(r-1)} + \left[\frac{x^4}{2} + x^2(m-1)\right] = \frac{h}{\sigma(r-1)}, \quad m = \text{const.}$$

for the values given in (3.7) is compared with the exact solution obtained by numerical integration for several Rayleigh number values. For the numerical integration, an integration step $\Delta t = 0.001$ is used in a Runge–Kutta algorithm, the solution being followed after 10^4 steps to remove the transient behavior. Plotted are points at 0.002 intervals and the initial conditions are chosen near the Hamiltonian orbit, namely $m(0) = 0.859$, $x(0) = 0$, $p(0) = \sqrt{h}$. In the figures $p^* = p/\sqrt{\sigma(r-1)}$.

From the inspection of the results, one notices that the Hamiltonian orbit gives a good estimate of the size and average position of the limit cycles. The perspective view and the projection in the p^*–x plane (where the H^1 dynamics takes place) allow an interpretation of these cycles as a distortion in the m direction of the Hamiltonian orbit obtained from the theorem.

Also in the "turbulent" regions ($r = 100, 200$) the exact solution winds around the Hamiltonian orbit, the projection in the p^*–x plane revealing the analytical orbit as an organizing center for the dynamics.

[1] J. Franks, "Manifolds of C^r-mappings and applications to dynamical systems," in *Studies in Analysis*, Advances in Mathematics, Supplementary Studies Vol. 4 (Academic, New York, 1979), pp. 271–90.

[2] R. V. Mendes and J. T. Duarte, J. Math. Phys. **22**, 1420 (1981).

[3] E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).

[4] T. Shimizu, Phys. Lett. A **69**, 379 (1979); **71**, 319 (1979); Physica A **97**, 383 (1979).

[5] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).

Irreversible quantum dynamics and the Hilbert space structure of quantum kinematics

N. Gisin[a]
Département de Physique Théorique, Université de Genève, 1211 Genève, 4, Switzerland

(Received 5 August 1982; accepted for publication 4 February 1983)

General dynamics compatible with the Hilbert space structure of quantum kinematics are considered. The general form of dynamics which preserve the set of closed linear submanifolds (i.e., properties) is deduced. Since the orthogonality relation is not necessarily preserved, the result generalizes Wigner's theorem and provides a model of some irreversible phenomena like spin relaxation, damped oscillator, etc. Connections with quantum logic and with statistical mechanics are presented.

PACS numbers: 03.65. – w

I. INTRODUCTION

The reversible dynamics for isolated quantum systems is given by the Schrödinger equation. Wigner's theorem[1,2] asserts that, starting from the Hilbert space structure of quantum kinematics, the unitary Schrödinger evolution can be deduced from the assumption that orthogonal states remain orthogonal throughout the evolution. In terms of quantum "logic," this means that the Schrödinger evolution is characterized by the fact that it induces symmetries of the property lattice $\mathscr{L}$, i.e., bijections of $\mathscr{L}$ into $\mathscr{L}$ which preserve the orthocomplements. A possible weakening of these assumptions is to suppose that the evolution induces only bijections of $\mathscr{L}$ into $\mathscr{L}$. That is to say, the evolution preserves the set of properties, but not all the relations between them. The interest for such generalization is obvious in the contexts of open quantum systems, relaxation phenomena, the measurement problem, etc. Such a generalization has been recently proposed by Daniel[3] (see also Ref. 4). The purpose of the present paper is to deduce the general form of such dynamics in the case of quantum mechanics on a Hilbert space (possibly with superselection variable; see the Appendix). It turns out that the corresponding evolution equations are not invariant under time reversal. Therefore, they describe irreversible phenomena, like a damped oscillator or spin relaxation, for example. In the latter, the angles between the spins are clearly not preserved; hence the evolution is not unitary.

Although our results concern quantum mechanics on a Hilbert space, our approach fits naturally into the framework of quantum "logic." Therefore, the paper is organized as follows. In the next section we justify our assumptions within an axiomatic approach to quantum physics. In Sec. III we return to quantum mechanics on a Hilbert space and present our mathematical results. In the last section we comment on them and establish the connection with the master equations of quantum statistical mechanics.

II. DETERMINISTIC EVOLUTIONS AND ELEMENTS OF REALITY

The aim of this section is to justify the assumptions of the theorems of the next section. For that purpose let us briefly sketch our—realistic—viewpoint about quantum physics.

Let us assume that the system exists by its own and always has some "elements of reality" (in Einstein's terminology[5]). Recall that an element of reality—or equivalently an "actual property" in Piron's terminology[6]—is a property of the physical system that can be tested and that is such that if we would actually perform a test, the positive result would always come out. Consequently, the elements of reality only depend on the system (and not on the measuring apparatus): They are in some way engraved in the system. The notion of state of a system can now be made precise: The state is the collection of all elements of reality. It is worth noting that in quantum mechanics the states, as defined above, uniquely determine the system's propensities[7] to actualize a property during an ideal measurement of the first kind (Gleason's theorem[6,8]).

To each evolution corresponds a semigroup of mapping of the state space into itself. Accordingly, during the evolution some elements of reality disappear and some other appear. Let us denote $\mathscr{A}$ the state space and μ_t the evolution, one has:

$$\mu_t\colon \mathscr{A}\to \mathscr{A}, \quad \mu_t\circ\mu_s=\mu_{t+s}, \quad \forall\, t, s\geqslant 0.$$

Moreover, a property b of the system is naturally represented by the set of states which make b actual. (For a discussion of the notion of property we refer to Refs. 6 and 9.) Accordingly, $b\subset \mathscr{A}$ and the set $\mathscr{L}$ of all properties is a subset of the power set of $\mathscr{A}$: $\mathscr{L}\subset P(\mathscr{A})$. The order relation given by the set theoretical inclusion provides $\mathscr{L}$ with the structure of a complete lattice (the lattice of quantum logic). Let us denote by $<$ this order relation, and by $\wedge$ the greatest lower bound and by $\vee$ the lowest upper bound. Henceforth, we shall identify any state $p\in\mathscr{A}$ with the corresponding minimal actual property $\wedge_{p\in b} b\in\mathscr{L}$. As usual, we assume that these properties are atoms of $\mathscr{L}$. Hence for $p\in\mathscr{A}$ we shall write $p\in\mathscr{L}$ instead of $\{p\}\in\mathscr{L}$.

At this point it may be useful to recall that in quantum mechanics the state space $\mathscr{A}$ is represented by the set of rays of a (complex separable) Hilbert space $\mathscr{H}$ and the properties are represented by the closed linear submanifolds of $\mathscr{H}$.

For any fixed t, μ_t induces a map γ_t from $\mathscr{L}$ into $P(\mathscr{A})$:

$$\gamma_t\colon\mathscr{L}\to P(\mathscr{A})$$

[a] Partially supported by the Swiss National Science Foundation.

 0022-2488/83/071779-04$02.50

$$b \mapsto \{ \mu_t(p) | p < b \}.$$

Our aim now is to characterize γ_t. The idea is that during the evolution the system remains the same, although the state changes. Hence we suppose that the evolution does not change the set of tests a physicist can carry out on the system. Now, recall that a property is determined by a test (more precisely by an equivalence class of tests). Consequently, following Daniel,[3] we shall say that *an evolution* μ_t *is deterministic if* the set of properties is preserved by the evolution:

$$\gamma_t(\mathscr{L}) = \mathscr{L} \quad \text{and} \quad \gamma_t \text{ is one-to-one.} \tag{1}$$

The terminology comes from the fact that to each element of reality at time t_0 there is a one-to-one correspondence with an element of reality at any time $t_1 < t_0$ or $t_1 > t_0$.

The following lemma displays some properties of deterministic evolutions.

Lemma 1: If μ_t is a deterministic evolution, then
(i) μ_t is bijective
(ii) $\forall\, p,q,r \in \mathscr{A}$,

$$p < q \vee r \Leftrightarrow \mu(p) < \mu(q) \vee \mu(r).$$

Proof: Per construction γ_t and γ_t^{-1} are order-preserving. Consequently, γ_t maps atoms onto atoms; hence μ_t is bijective, and γ_t and γ_t^{-1} preserve the upper and lower bounds. The conclusion follows.

The unitary Schrödinger evolution is a well-known example of deterministic evolution. It is characterized by the fact that it preserves orthogonal states. It is interesting to note that a deterministic evolution in a Hilbert space preserves biorthogonal linear submanifolds. Indeed a linear submanifold M is biorthogonal $M^{\perp\perp} = M$ (where $M^{\perp} = \{ \psi \in \mathscr{H} | \langle \psi | \varphi \rangle = 0 \ \forall \varphi \in M \}$) if and only if M is closed.[10] Clearly an evolution preserving orthogonal states *a fortiori* preserves biorthogonal linear submanifolds. But the converse is false. Hence the unitary evolution is not the only possible deterministic evolution.[3]

III. GENERAL FORM OF DETERMINISTIC EVOLUTIONS IN HILBERT SPACES

In this section we determine the general form of the deterministic evolutions defined in the preceeding section. Let us make the following definition:

Definition: A map $\mu: \mathscr{H} \rightarrow \mathscr{H}$ *preserves the superpositions* if $\forall\, \alpha, \beta \in \mathbb{C}^*$, $\varphi,\psi \in \mathscr{H}$, $\exists\, \delta, \zeta \in \mathbb{C}^*$ such that $\mu(\alpha\varphi + \beta\psi) = \delta\mu(\varphi) + \zeta\mu(\psi)$, where $\mathbb{C}^* = \mathbb{C} - \{0\}$.

The next two theorems are the main results of this section. We have already quoted them in Ref. 11.

Theorem 2: Let $\mathscr{H}$ be a complex Hilbert space, $\dim \mathscr{H} \geqslant 3$. Let $P(\mathscr{H})$ denote the set of closed linear manifolds of $\mathscr{H}$, and let $\mu_t: \mathscr{H} \rightarrow \mathscr{H}$ (t fixed). If μ_t is deterministic [see (1)], i.e., if μ_t induces a bijection $\mu_t: P(\mathscr{H}) \rightarrow P(\mathscr{H})$, then:

(i) μ_t preserves the superpositions;

(ii) $\exists$ a semilinear operator V_t acting on $\mathscr{H}$ (unique up to a multiplicative constant) and $\exists$ a map $g_t: \mathscr{H} \rightarrow \mathbb{C}^*$ such that

$$\mu_t(\psi) = g_t(\psi) \cdot V_t \psi, \quad \forall \psi \in \mathscr{H};$$

(iii) if μ_t is continuous, then the operator V_t is linear or antilinear.

Theorem 3: Let $\mathscr{H}$ be a complex separable Hilbert space, $\dim \mathscr{H} \geqslant 3$. Let $\mu_t: \mathscr{H} \rightarrow \mathscr{H}$ be a semigroup such that the maps $(t,\psi) \rightarrow \mu_t(\psi)$ are C^0 $\forall t \geqslant 0$, $\psi \in \mathscr{H}$, and $\exists$ a dense domain $\mathscr{D} \subset \mathscr{H}$ such that the maps $t \rightarrow \mu_t(\psi)$ are C^1 $\forall \psi \in \mathscr{D}$. If the μ_t are deterministic, then $\exists$ a C^0 contraction semigroup of linear operators W_t and $\exists$ a family of maps $h_t: \mathscr{H} \rightarrow \mathbb{C}^*$ such that

$$\mu_t(\psi) = h_t(\psi) \cdot W_t \psi \quad \forall\, \psi \in \mathscr{H}. \tag{2}$$

Proof 2: (i) Corollary of Lemma I, (ii).

(ii) μ_t induces a bijection between the raies of $\mathscr{H}$; hence μ_t induces a bijection σ from the projective space $\hat{\mathscr{H}}$ ($\simeq \mathscr{H}/\mathbb{C}$) into itself: $\sigma: \hat{\mathscr{H}} \rightarrow \hat{\mathscr{H}}$. Moreover, since μ_t preserves the superpositions, σ has the following property: If $\hat{\psi},\hat{\varphi},\hat{\chi}$ are three aligned points of $\hat{\mathscr{H}}$ (i.e., if the vectors ψ,φ,χ belong to a common plane of $\mathscr{H}$), then $\sigma(\hat{\psi})$, $\sigma(\hat{\varphi})$, $\sigma(\hat{\chi})$ are still aligned in $\hat{\mathscr{H}}$ [i.e., $\mu(\psi)$, $\mu(\varphi)$, $\mu(\chi)$ still belong to a common plane of $\mathscr{H}$]. Consequently, σ satisfies the assumptions of the fundamental theorem of projective geometry[12,13] (it is here that the assumption $\dim \mathscr{H} \geqslant 3$ is important). Accordingly, σ is induced by a semilinear transformation of $\mathscr{H}$.

(iii) Let $\psi,\varphi \in \mathscr{H}$ be such that $\mu_t(\psi) \perp \mu_t(\varphi)$. Consider the following map $f: \mathbb{C} \rightarrow \mathbb{R}$:

$$f(\alpha) = |\langle \mu_t(\psi) | \mu_t(\psi + \alpha\varphi) \rangle|^2 / \| \mu_t(\psi) \|^2 \cdot \| \mu_t(\psi + \alpha\varphi) \|^2.$$

The assumption implies that f is continuous. Let us note λ the automorphism of $\mathbb{C}$ corresponding to the semilinear operator V_t: $V_t(\alpha\psi) = \lambda(\alpha) \cdot V_t\psi$. One has

$$f(\alpha) = \| V_t\psi \|^2 / [\, \| V_t\psi \|^2 + |\lambda(\alpha)|^2 \cdot \| V_t\varphi \|^2] .$$

The automorphism λ is thus continuous, accordingly the operator V_t is linear or antilinear.

Proof 3: We divide this proof in three steps.

(a) $\forall \varphi$, $\psi \in \mathscr{D}$ the maps $t \rightarrow g_t(\varphi)\, V_t\psi$ and $t \rightarrow g_t(\varphi)^{-1} \cdot g_t(\psi)$ are C^1, where g_t and V_t are given by Theorem 2.

(b) $\exists$ a C^0 semigroup W_t of linear operators satisfying (2).

(c) The operators W_t are linear contractions.

(a) Let $\varphi \in \mathscr{D}$ be fixed, and put $\tilde{V}_t\psi \equiv g_t(\varphi) \cdot V_t\psi$ and $\tilde{g}_t(\psi) \equiv g_t(\varphi)^{-1} \cdot g_t(\psi)$. Hence $\mu_t(\psi) = \tilde{g}_t(\psi) \cdot \tilde{V}_t(\psi)$. Let $\psi \in \mathscr{D}$ be linearly independent of φ. One has

$$\mu_t(\varphi + \psi) = \tilde{g}_t(\varphi + \psi)\, \mu_t(\varphi) + \tilde{g}_t(\varphi + \psi)\, \tilde{g}_t(\psi)^{-1}\, \mu_t(\psi).$$

But $\mu_t(\varphi)$ and $\mu_t(\psi)$ are linearly independent; the differentiability properties of $t \rightarrow \mu_t(\psi)$ thus go over to the maps

$$t \mapsto \tilde{g}_t(\varphi + \psi) \quad \text{and} \quad t \mapsto \tilde{g}_t(\psi)^{-1}$$

Consequently, the maps $t \rightarrow \tilde{g}_t(\psi)$ and $t \rightarrow \tilde{V}_t(\psi)$ are C^1 for all $\psi \in \mathscr{D}$ linearly independent of φ. Finally, since $\tilde{V}_t(\varphi) = \mu_t(\varphi)$, this property is satisfied for all $\psi \in \mathscr{D}$.

(b) The semigroup law of μ_t implies

$$\tilde{V}_t \tilde{V}_s \psi = \omega_\psi(t,s)\, \tilde{V}_{t+s}\psi$$

where

$$\omega_\psi(t,s) = \tilde{g}_{t+s}(\psi) / \tilde{g}_t(\tilde{g}_s(\psi)\, \tilde{V}_s\psi) \cdot \tilde{g}_s(\psi)$$

Similar relations occur in the theory of projective unitary representations.[14] The only difference is that here the $\omega_\psi(t,s)$ are not necessarily of norm one (but they are $\neq 0$); consequently, one can generalize the usual arguments in order to take the ω's off. Let $\psi = \Sigma\, \alpha_i \psi_i$, $\alpha_i \in \mathbb{C}$, $\psi_i \in \mathscr{H}$. One has

$$\omega_\psi(t,s) \sum \alpha_i\, \tilde{V}_{t+s}\psi_i$$

$$= \sum \omega_{\psi_i}(t,s)\, \alpha_i\, \tilde{V}_{t+s}\psi_i.$$

Hence, the ω's are independent of ψ. Put $W_t\psi = \exp[\int_0^t f(t')\,dt']\, V_t\psi$, where $f(s) = \partial_t \omega(t,s)|_{t=0}$. One has $W_t \cdot W_s = \ell(t,s)\, W_{t+s}$, where

$$\ell(t,s) = \omega(t,s) \times \exp\left[\int_0^t f(t')\,dt' + \int_0^s f(s')\,ds' - \int_0^{t+s} f(u)\,du\right].$$

From the associativity law of the μ_t's one deduces that $\partial_t \ell(t,s) = \partial_s \ell(t,s) = 0$. But $\ell(0,s) = \ell(t,0) = 1$; hence $\ell(t,s) = 1$, $\forall t,s \geqslant 0$ and $W_t \cdot W_s = W_{t+s}$. Finally the operators W_t are linear because $W_t = (W_{t/2})^2$.

(c) We first prove that the W_t's are closed (for all fixed t); the boundness of the W_t's follows from the closed graph theorem.[15]

Let $\{\varphi_i\} \in \mathscr{H}$ be such that

$$\lim_{i\to\infty} \varphi_i = \varphi \quad \text{and} \quad \lim_{i\to\infty} W_t \varphi_i = \chi.$$

Since $\mu_t(\psi)$ is continuous in ψ, one has $\lim_{i\to\infty} \mu_t(\varphi_i) = \mu_t(\varphi)$. On the other hand,

$$h_t(\psi) = \langle W_t\psi | \mu_t(\psi)\rangle / \| W_t\psi\|^2.$$

Hence

$$\lim_{i\to\infty} h_t(\varphi_i) = \langle \chi | \mu_t(\varphi)\rangle / \| \chi \|^2$$

and

$$W_t\varphi = \frac{\mu_t(\varphi)}{h_t(\varphi)} = \lim_{i\to\infty} \frac{\mu_t(\varphi_i)}{h_t(\varphi)}$$

$$= \lim_{i\to\infty} \frac{h_t(\varphi_i)}{h_t(\varphi)} \cdot W_t\varphi_i = \frac{\langle \chi | \mu_t(\varphi)\rangle}{\| \chi \|^2 \cdot h_t(\varphi)} \cdot \chi.$$

$W_t\varphi$ is thus proportional to χ. Finally, let $\psi \in \mathscr{H}$ be linearly independent of φ. The same arguments show that $W_t(\varphi + \psi)$ is proportional to $\chi + W_t(\psi)$. Consequently, $W_t\varphi = \chi$. Finally one has[16]

$$\| W_t \| \leqslant M \cdot e^{at}$$

for some constants M and a. By redefining h_t and W_t one can thus include the bound $M \cdot e^{at}$ in the complex factor h_t and have $\| W_t \| \leqslant 1$. ∎

By looking carefully at the above proofs it may be seen that the assumption of a deterministic evolution μ_t is somewhat stronger than needed. Clearly, it is sufficient to assume that the evolution is injective on the rays and preserves the superpositions. In terms of lattices this amounts to assume that the evolution maps the lattice $\mathscr{L}$ into a not necessarily complete sublattice of $\mathscr{L}$.

IV. DISCUSSIONS

The results of the preceding section deserve some comments. First let us note that the complex number $h_t(\psi)$ in (2) does not affect the state (ray) defined by the vector $\mu_t(\psi)$. Consequently, $\forall\, \psi_0 \in \mathscr{H}$ the normalized vectors $\mu_t = W_t\psi_0 / \| W_t\psi_0\|$ are representatives of the evolution μ_t. Moreover, they satisfy the following evolution equation:

$$\dot{\psi}_t = -iH\psi_t + (\langle B\rangle_{\psi_t} - B)\,\psi_t, \tag{3}$$

where we have supposed that the generator Z of W_t is of the form $Z = -iH - B$ with $H^+ = H$ and $B^+ = B$. Let us notice that the evolution equation for the projectors is the following:

$$\dot{P}_t = -i[\mathrm{H}, \mathrm{P}_t] + [[P_t, B], P_t], \quad P_t = P_t^2 = P_t^+.$$

It is worthwhile to briefly consider the simple case $B = kH$ $(k > 0)$[17]:

$$\dot{\psi}_t = -iH\psi_t + k(\langle H\rangle_{\psi_t} - H)\,\psi_t. \tag{4}$$

Hence $(d/dt)\langle H\rangle_t = -2k(\langle H^2\rangle_t - \langle H\rangle_t^2) = -2k(\Delta H)^2 \leqslant 0$. Accordingly, the system dissipates energy, unless it is in an eigenstate of H. Furthermore, the stationary solutions (i.e., the solutions which correspond to states at rest) of (4) are exactly the same as those of the Schrödinger equation, and any solution of (4) tends asymptotically towards a stationary state. This asymptotical eigenstate corresponds to the lowest energy level on which the initial state has a nonvanishing component. Equation (4) has been successfully applied to spin relaxation[11,18] and to the damped oscillator.[17] In the latter the expected behavior was found for the coherent states. Notice that the nonlocal character of quantum physics[19,20] does not raise any problem.

Since Eq. (3) is nonconversative, one may expect some connections with the theory of open quantum systems,[21] and indeed one can deduce Eq. (3) in the framework of the master equations of quantum statistical mechanics. For that purpose, one applies precisely the same techniques as those which lead to the Pauli master equation, but instead of the partial trace projection one uses a pure state preserving projection.[22] Conversely, since W_t is a contraction semigroup, it can be dilated to a unitary evolution on a larger Hilbert space.[21]

Now let us stress the generality of our assumptions: The only nontechnical assumption is that the evolution preserves the set of properties, i.e., that the evolution is compatible with the Hilbert space structure and the usual interpretation of states and observables. We feel that this is a minimal assumption. Thus we answer the question of the connections between irreversible quantum dynamics and quantum "logic."[4] There is, however, an important implicit assumption: the Hilbert space. But this assumption is of a different nature, it concerns the kinematics and not the dynamics. Note that the Hilbert space is the only assumption about linearity. This leads us to our next remark.

It is well known that, in order to derive the linearity of the Schrödinger equation from the linearity of the Hilbert space, one has to assume that the evolution preserves orthogonal states (Wigner's theorem[1,2]). The latter assumption amounts to assuming that the evolution induces a symmetry of the lattice of properties, i.e., that not only the set of properties is preserved, but also all the relations between the propeties. Physically this means that nothing in the system changes, and indeed any Schrödinger evolution may be interpreted from the passive point of view, as is shown by the Heisenberg representation. On the other hand, in phenomena like spin relaxation, orthogonal states clearly do not remain orthogonal.

APPENDIX: QUANTUM DYNAMICS WITH SUPERSELECTION VARIABLES

The property lattice $\mathscr{L}$ of a quantum system with superselection (i.e., classical) variables is isomorphic to the direct union over a set Γ of Hilbert space lattices: $\mathscr{L} \approx V_{\gamma \in \Gamma} P(\mathscr{H}_\gamma)$.[5] Let us denote the atoms of $\mathscr{L}$ by $p_\alpha, q_\beta, r_\gamma, \cdots$, where $p, q, r, \cdots$ are atoms of $P(\mathscr{H}_\alpha)$, $P(\mathscr{H}_\beta)$, $P(\mathscr{H}_\gamma), \cdots$, respectively, and $\alpha, \beta, \gamma, \cdots \in \Gamma$. Let $\mathscr{A}_\gamma$ be the set of atoms of $P(\mathscr{H}_\gamma)$ and $\mathscr{A}$ the set of atoms of $\mathscr{L}$. We state without proof the following theorem which completes our study of deterministic evolutions.

Theorem 4: If $\mu: \mathscr{A} \rightarrow \mathscr{A}$ is deterministic, then $\exists\ \sigma: \Gamma \rightarrow \Gamma$ and $\nu_\alpha : \mathscr{A}_\alpha \rightarrow \mathscr{A}_{\sigma(\alpha)}$ such that σ is bijective and ν_α is deterministic $\forall\ \alpha \in \Gamma$, and $\mu(p_\alpha) = \nu_\alpha(p)_{\sigma(\alpha)}\ \forall\ p_\alpha \in \mathscr{A}$.

Consequently, a deterministic evolution for a classical system is nothing but a dynamical system.

[1] E. P. Wigner, *Group Theory* (Academic, New York, 1959).
[2] G. Emch and C. Piron, J. Math. Phys. **4**, 469 (1963).
[3] W. Daniel, Helv. Phys. Acta **55**, 330 (1982).
[4] V. Gorini and A. Zecca, J. Math. Phys. **16**, 667 (1975); see also A. Zecca, Int. J. Theor. Phys. **20**, 191 (1981).
[5] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).
[6] C. Piron, *Foundations of Quantum Physics* (Benjamin, New York, 1976).
[7] K. R. Popper, Br. J. Phil. Sci. **10**, 25 (1959).
[8] A. M. Gleason, J. Math. Mech. **6**, 885 (1957).
[9] D. Aerts "The classical and the nonclassical part of the description of an entity," Found. Phys. (to be published).
[10] T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, Berlin, 1976), Sec. V.1.
[11] N. Gisin, "A model of dissipative quantum dynamics," Ph.D. thesis, Department of Theoretical Physics, University of Geneva, 1982.
[12] R. Baer, *Linear Algebra and Projective Geometry* (Academic, New York, 1952).
[13] E. Artin, *Geometric Algebra* (Wiley-Interscience, New York, 1957).
[14] A. O. Barut and R. Rączka, *Theory of Group Representations and Applications* (Polish Scientific Publishers, Warsaw, 1977).
[15] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1972), Vol. 1.
[16] E. B. Davies, *One Parameter Semigroups* (Academic, London, 1980), Theorem 1.18.
[17] N. Gisin, J. Phys. A **14**, 2259 (1981).
[18] N. Gisin, Helv. Phys. Acta **54**, 457 (1981).
[19] H. Rauch *et al.*, Phys. Lett. A **54**, 425 (1975).
[20] A. Aspect, J. Dalibard, and G. Roger, Phys. Rev. Lett. **49**, 1804 (1982).
[21] E. B. Davies, *Quantum Theory of Open Systems* (Academic, Oxford, 1976).
[22] N. Gisin, Physica A **111**, 364 (1982).

The Fierz identities—A passage between spinors and tensors

Yasushi Takahashi
Theoretical Physics Institute, Department of Physics, University of Alberta, Edmonton, Alberta T6G 2J1, Canada

(Received 12 October 1982; accepted for publication 26 November 1982)

All possible Fierz identities among 16 elements in the Dirac algebra have been obtained. These relations impose various constraint conditions on Hermitian and non-Hermitian bilinear currents. Independent relations are sought out from the highly redundant system of constraint conditions. Relations between derivatives of a spinor and tensor currents are also obtained.

PACS numbers: 03.65. – w

I. INTRODUCTION

The purpose of this paper is to derive explicitly all the Fierz identities between various matrix elements of the Dirac gammas, and to apply them to obtain relations among 16 Hermitian currents J (scalar), J_5 (pseudoscalar), J_μ (vector), $J_{5\mu}$ (axial vector), $J_{\mu\nu}$ (skew tensor), and 20 non-Hermitian currents R_μ (complex vector) and $R_{\mu\nu}$ (complex skew tensor), in order to establish the passage from tensors to spinors and vice versa. These relations, which will be referred to as the Fierz constraint conditions, are not all independent. We have singled out 28 basic constraint relations from which all the other constraint relations can be derived, though the choice of the primary set is by no means unique.

The analysis of the Fierz constraint conditions provides us with the foundation for the spinor reconstruction theorem,[1,2] i.e., the spinor can be reconstructed *uniquely* from eight components of the currents, for example, J_5, J_i, J_{5i}, and the phase of R_0.

The Fierz identities can also be applied to obtaining the relations between the space–time derivative of a spinor and that of tensors. Hence, the physical quantities associated with the spinor field can be expressed in terms of the currents and their derivatives, without the knowledge of dynamics. We have illustrated such relations in Sec. V.

This investigation was motivated by a number of physical problems, such as the boson–fermion equivalence theory in (1 + 1)-dimensional space,[3] the supersymmetry theory,[4] the grand unification theory,[5] the proof of the positive definiteness of the gravitational energy with the spinor field,[6] and the gauge principle in particle physics.

From our viewpoint, there is no reason why the spinor is more fundamental than the tensors. A certain system of constrained tensors is equivalent to a spinor, although the introduction of a spinor may simplify the matter considerably in practice. For example, a tetrad in Minkowski space implies the existence of a spinor, and the orthogonality and completeness conditions are automatically satisfied, when the tetrad is expressed in terms of the spinor. Thus, the spinor monism is not the only possibility for unifying the constituents of the matter.

Throughout this paper, we adopt the notation used in the Appendix in Ref. 7.

II. THE FIERZ IDENTITIES

The basic Fierz identity[8] is written as

$$\delta_{ab}\delta_{cd} = \tfrac{1}{4}\sum_{C=1}^{16}(\gamma_C)_{ad}(\gamma_C)_{cb}, \tag{2.1}$$

where γ_C $(C = 1,2,...,16)$ are elements of the γ-algebra normalized as

$$\mathrm{Tr}[\gamma_A\gamma_B] = 4\delta_{AB}, \quad A,B = 1,2,...,16. \tag{2.2}$$

To derive further identities, multiply (2.1) by γ_A and γ_B to obtain

$$(\gamma_A)_{ab}(\gamma_B)_{cd} = \tfrac{1}{4}\sum_{C=1}^{16}(\gamma_A\gamma_C)_{ad}(\gamma_B\gamma_C)_{cb} \tag{2.3a}$$

$$= (\tfrac{1}{4})^2\sum_{C,D}\mathrm{Tr}[\gamma_C\gamma_A\gamma_D\gamma_B](\gamma_C)_{ad}(\gamma_D)_{cb}. \tag{2.3b}$$

Calculating the trace containing four gamms in (2.3b), we obtain all possible Fierz identities. However, the calculation of all the traces is extremely tedious if not impossible. We have used the form (2.3a) instead which can be calculated by hand much more directly than using the form (2.3b).

The lower case Latin letters, indicating spinor indices rearrangement, can be omitted without sacrificing clarity if we introduce the square bracket and write (2.3a) as

$$(\gamma_A)[\gamma_B] = \tfrac{1}{4}\sum_{C=1}^{16}(\gamma_A\gamma_C][\gamma_B\gamma_C). \tag{2.4}$$

Denoting as

$$\begin{aligned}&(I)\equiv(\ ),\quad (i\gamma_5)\equiv(5),\\ &(i\gamma_\mu)\equiv(\mu),\quad (i\gamma_5\gamma_\mu)\equiv(5\mu),\\ &(\sigma_{\mu\nu})\equiv(\mu\nu),\quad (i\gamma_5\sigma_{\mu\nu}) = -\tfrac{1}{2}i\epsilon_{\mu\nu\alpha\beta}(\sigma_{\alpha\beta})\equiv({}^*\mu\nu)\end{aligned} \tag{2.5}$$

and similarly for [ ], ( ], and [ ), we obtain, after straightforward but lengthy calculations, 15 relations in the Appendix (I), listed as (F-1)···(F-15).

To obtain relations between 16 Hermitian currents, we put

$$\begin{aligned}&(\ ) = [\ ] = (\ ] = [\ ) = \bar\psi\psi\equiv J,\\ &(5) = [5] = (5] = [5) = i\bar\psi\gamma_5\psi\equiv J_5,\\ &(\mu) = [\mu] = (\mu] = [\mu) = i\bar\psi\gamma_\mu\psi\equiv J_\mu,\\ &(5\mu) = [5\mu] = (5\mu] = [5\mu) = i\bar\psi\gamma_5\gamma_\mu\psi\equiv J_{5\mu},\\ &(\mu\nu) = [\mu\nu] = (\mu\nu] = [\mu\nu) = \bar\psi\sigma_{\mu\nu}\psi\equiv J_{\mu\nu}.\end{aligned} \tag{2.6}$$

Then, it follows immediately from (F-1)···(F-15) the relations in the Appendix (II), listed as (JJ-1)···(JJ-15).[9] There are altogether 187 relations among 16 quantities in (2.6). As we shall see later, only nine relations are independent, and the rest of the relations can be derived from them. Hence, seven quantities in (2.6) can be assigned arbitrarily. As has been shown, the spinor ψ can be reconstructed, except for the overall

 0022-2488/83/071783-08$02.50

phase, from seven quantities, for example, J_5, J_i, and J_{5i}, and the identities $(JJ\text{-}1)\cdots(JJ\text{-}15)$.

The overall phase can also be recovered, if we know, beside the Hermitian currents, the gauge variant quantity

$$R_\mu \equiv i\bar{\psi}^c \gamma_\mu \psi, \tag{2.7}$$

where $\bar{\psi}^c$ is the Pauli conjugate of the charge conjugation ψ^c of the spinor ψ. The quantity (2.7) satisfies the identities coming from (F-1)$\cdots$(F-15) together with the quantity

$$R_{\mu\nu} \equiv \bar{\psi}^c \sigma_{\mu\nu} \psi. \tag{2.8}$$

They are listed in the Appendix (III), (IV), (V), and (VI). In calculating $(RJ\text{-}1)\cdots(RJ\text{-}15)$, $(JR\text{-}1)\cdots(JR\text{-}15)$, $(RR\text{-}1)\cdots(RR\text{-}15)$, and $(R\bar{R}\text{-}1)\cdots(R\bar{R}\text{-}15)$, we assumed that ψ is the commuting quantity and, hence,

$$\bar{\psi}^c \psi = i\bar{\psi}^c \gamma_5 \psi = i\bar{\psi}^c \gamma_5 \gamma_\mu \psi = 0,$$

$$\bar{\psi}^c \psi^c = -\bar{\psi}\psi = -J, \tag{2.9}$$

$$i\bar{\psi}^c \gamma_5 \psi^c = -i\bar{\psi}\gamma_5\psi = -J_5,$$

$$i\bar{\psi}^c \gamma_\mu \psi^c = i\bar{\psi}\gamma_\mu \psi = J_\mu,$$

$$i\bar{\psi}^c \gamma_5\gamma_\mu \psi^c = -i\bar{\psi}\gamma_5\gamma_\mu\psi = -J_{5\mu}, \tag{2.10}$$

$$\bar{\psi}^c \sigma_{\mu\nu}\psi^c = \bar{\psi}\sigma_{\mu\nu}\psi = J_{\mu\nu},$$

which can be proved by the definition of the charge conjugation.

The relations $(RJ\text{-}1)\cdots(R\bar{R}\text{-}15)$ are again highly redundant and only 19 of them are independent, having only one component (namely the phase of R_0, say) arbitrary among 20 components of R_μ and $R_{\mu\nu}$.

III. INDEPENDENT RELATIONS AMONG BILINEAR QUANTITIES

As was seen in the preceding section, there are large numbers of relations among J, J_5, J_μ, $J_{5\mu}$, $J_{\mu\nu}$, R_μ, and $R_{\mu\nu}$ coming from the Fierz identities (F-1)$\cdots$(F-15). Under the condition

$$J^2 + J_5^2 \neq 0 \tag{3.1}$$

all the relations are summarized by the following 28 relations[10]:

$$J_{\mu\alpha}J_\alpha = -J_5 J_{5\mu}, \tag{3.2a}$$

$$*J_{\mu\alpha}J_\alpha = JJ_{5\mu}, \tag{3.2b}$$

$$J_\alpha J_\alpha = -J_{5\alpha}J_{5\alpha} = -(J^2 + J_5^2), \tag{3.2c}$$

$$J_{\mu\alpha}R_\alpha = iJR_\mu, \tag{3.3a}$$

$$R_{\mu\alpha}J_\alpha = -iJR_\mu \tag{3.3b}$$

$$*R_{\mu\alpha}J_\alpha = -iJ_5 R_\mu, \tag{3.3c}$$

$$R_\alpha \bar{R}_\alpha = 2(J^2 + J_5^2). \tag{3.3d}$$

[It appears that there are 35 relations in (3.2) and (3.3), but they are not all independent. For example, Eqs. (3.2a) and (3.2b) contain only seven independent relations.]

In order to show that all the Fierz constraint conditions $(JJ\text{-}1)\cdots(JJ\text{-}15)$ can be derived from (3.2), we note the identities, holding between two arbitrary skew-symmetric tensors $F_{\mu\nu}$ and $G_{\mu\nu}$,

$$F_{\mu\alpha}G_{\alpha\nu} - *F_{\nu\alpha}*G_{\alpha\mu} = -\tfrac{1}{2}\delta_{\mu\nu}F_{\alpha\beta}G_{\alpha\beta} = \tfrac{1}{2}\delta_{\mu\nu}*F_{\alpha\beta}*G_{\alpha\beta}, \tag{3.4a}$$

$$F_{\mu\alpha}*G_{\alpha\nu} + *F_{\nu\alpha}G_{\alpha\mu} = -\tfrac{1}{2}\delta_{\mu\nu}*F_{\alpha\beta}G_{\alpha\beta} = -\tfrac{1}{2}\delta_{\mu\nu}F_{\alpha\beta}*G_{\alpha\beta}, \tag{3.4b}$$

where $*F_{\mu\nu}$ is the dual of $F_{\mu\nu}$ defined by

$$*F_{\mu\nu} = -\tfrac{1}{2}i\,\epsilon_{\mu\nu\alpha\beta}F_{\alpha\beta}, \tag{3.5}$$

and similarly for $*G_{\mu\nu}$.

The identities

$$\epsilon_{\mu\nu\lambda\alpha}F_{\alpha\beta}A_\beta = -i(A_\mu^* F_{\nu\lambda} + A_\nu^* F_{\lambda\mu} + A_\lambda^* F_{\mu\nu}), \tag{3.6a}$$

$$\epsilon_{\mu\nu\lambda\alpha}*F_{\alpha\beta}A_\beta = i(A_\mu F_{\nu\lambda} + A_\nu F_{\lambda\mu} + A_\lambda F_{\mu\nu}) \tag{3.6b}$$

and also useful for an arbitrary skew-symmetric tensor $F_{\mu\nu}$ and an arbitrary vector A_μ. The relations such as

$$J_{5\alpha}J_\alpha = 0, \tag{3.7a}$$

$$J_\mu J_{\nu\lambda} + J_\nu J_{\lambda\mu} + J_\lambda J_{\mu\nu} + i\epsilon_{\mu\nu\lambda\alpha}JJ_{5\alpha} = 0, \tag{3.7b}$$

$$J_{\mu\nu} = \{J_5(J_\mu J_{5\nu} - J_\nu J_{5\mu}) - iJ\epsilon_{\mu\nu\alpha\beta}J_\alpha J_{5\beta}\}/J_\alpha J_\alpha, \tag{3.7c}$$

$$J_{\mu\alpha}J_{5\alpha} = -J_5 J_\mu, \tag{3.7d}$$

$$*J_{\mu\alpha}J_{5\alpha} = JJ_\mu, \tag{3.7e}$$

$$J_{\mu\alpha}J_{\nu\alpha} = \delta_{\mu\nu}J^2 + J_\mu J_\nu - J_{5\mu}J_{5\nu}, \tag{3.7f}$$

$$*J_{\mu\alpha}*J_{\nu\alpha} = \delta_{\mu\nu}J_5^2 + J_\mu J_\nu - J_{5\mu}J_{5\nu}, \tag{3.7g}$$

$$J_{\mu\alpha}*J_{\nu\alpha} = \tfrac{1}{4}\delta_{\mu\nu}J_{\alpha\beta}*J_{\alpha\beta} = \delta_{\mu\nu}JJ_5 \tag{3.7h}$$

follow from the relations (3.2), (3.4), and (3.6).

We now turn our attention to the relations (3.3). First, we note that having known all the J's from (3.2), Eq. (3.3a) determines R_i in terms of R_0 and all the J's. Then, Eqs. (3.3b) and (3.3c) determine $R_{\mu\nu}$ uniquely in terms of R_μ and all J's. The normalization of R_μ is fixed by (3.3d).

Obviously, we have

$$R_\alpha R_\alpha = 0 \tag{3.8}$$

from (3.3a), and

$$R_\alpha J_\alpha = 0 \tag{3.9}$$

from (3.3b). Using (3.3a), (3.7d), and (3.9), we obtain

$$R_\alpha J_{5\alpha} = 0. \tag{3.10}$$

Other relations,

$$R_{\mu\alpha}J_{5\alpha} = -J_5 R_\mu, \tag{3.11a}$$

$$*R_{\mu\alpha}J_{5\alpha} = JR_\mu, \tag{3.11b}$$

$$R_{\mu\alpha}J_{\nu\alpha} = *R_{\nu\alpha}*J_{\mu\alpha} = iJR_{\mu\nu} + R_\mu J_\nu, \tag{3.11c}$$

$$R_{\mu\alpha}*J_{\nu\alpha} = -*R_{\nu\alpha}J_{\mu\alpha} = iJ*R_{\mu\nu} + iJ_{5\mu}R_\nu, \tag{3.11d}$$

$$R_{\mu\alpha}R_\alpha = *R_{\mu\alpha}R_\alpha = R_{\mu\alpha}*R_{\nu\alpha} = 0, \tag{3.11e}$$

$$R_{\mu\nu} = \{R_\mu(iJJ_\nu - J_5J_{5\nu}) - R_\nu(iJJ_\mu - J_5J_{5\mu})\}/(J^2 + J_5^2) = \{iJ(R_\mu J_\nu - R_\nu J_\mu) - J_5\epsilon_{\mu\nu\alpha\beta}R_\alpha J_\beta\}/(J^2 + J_5^2), \tag{3.11f}$$

$$J_{\mu\nu} = \{J_5(J_{5\mu}J_\nu - J_{5\nu}J_\mu) - \tfrac{1}{2}iJ(\bar{R}_\mu R_\nu - \bar{R}_\nu R_\mu)\}/(J^2 + J_5^2), \tag{3.11g}$$

$$R_{\mu\alpha}R_{\nu\alpha} = *R_{\mu\alpha}*R_{\nu\alpha} = R_\mu R_\nu, \tag{3.11h}$$

$$R_{\mu\alpha}\bar{R}_\alpha = 2(J_5J_{5\mu} - iJJ_\mu), \tag{3.11i}$$

$$*R_{\mu\alpha}\bar{R}_\alpha = -2(JJ_{5\mu} + iJ_5J_\mu), \tag{3.11j}$$

$$R_\mu\bar{R}_\nu = \delta_{\mu\nu}(J^2 + J_5^2) + J_\mu J_\nu - J_{5\mu}J_{5\nu} - \epsilon_{\mu\nu\alpha\beta}J_\alpha J_{5\beta}, \tag{3.11k}$$

$$R_{\mu\alpha}\,{}^*\bar{R}_{\nu\alpha} = 2i(J_5 J_{\mu\nu} + J_\mu J_{5\nu}) - 2\delta_{\mu\nu} J J_5, \tag{3.11l}$$

$$i\epsilon_{\mu\nu\alpha\beta} R_\alpha \bar{R}_\beta = 2i(J_\mu J_{5\nu} - J_\nu J_{5\mu}), \tag{3.11m}$$

can be derived from (3.2) and (3.3) with the help of (3.4) and (3.6).

IV. ORTHOGONAL TETRAD AND PARAMETER REPRESENTATIONS

The relations (3.2c), (3.7a), (3.8), (3.9), and (3.10) imply that the four 4-vectors

$$\begin{aligned}
&\tfrac{1}{2}(R_\mu - \bar{R}_\mu)/(J^2 + J_5^2)^{1/2} \equiv h_\mu^{(1)},\\
&-\tfrac{1}{2}i(R_\mu + \bar{R}_\mu)/(J^2 + J_5^2)^{1/2} \equiv h_\mu^{(2)},\\
&J_{5\mu}/(J^2 + J_5^2)^{1/2} \equiv h_\mu^{(3)},\\
&J_\mu/(J^2 + J_5^2)^{1/2} \equiv h_\mu^{(0)}
\end{aligned} \tag{4.1}$$

form an orthonormal tetrad. Hence, any vector orthogonal to $h_\mu^{(0)},\ldots,h_\mu^{(3)}$ vanishes identically. If we use this property, the relations such as (3.11f) and (3.11g) can be proved without difficulty. The completeness of the tetrad (4.1)

$$\tfrac{1}{2}(R_\mu \bar{R}_\nu + R_\nu \bar{R}_\mu) - J_\mu J_\nu + J_{5\mu} J_{5\nu} = (J^2 + J_5^2)\delta_{\mu\nu} \tag{4.2}$$

is again the consequence of the Fierz identities (3.2) and (3.3) or (3.11k).

To the orthogonal tetrad formed by three spacelike $h_\mu^{(1)}, h_\mu^{(2)}, h_\mu^{(3)}$, and one timelike $h_\mu^{(0)}$, we may introduce the parameter representation in terms of the Euler angles and the pseudo-Euler angles.[2] Then, constructing the spinor-determining equation, we can go back to the original spinor, as was done in our previous papers.[1,2] We may also introduce the representation adopted by Takabayasi,[11] which shows that the Dirac field consists of rotators and their translational motion (and another parameter).

The alternative parameter representation

$$J_{5\mu} = 4|c_0|^2(S_1 C_2 a_i^{(1)} + S_2 a_i^{(2)}, iC_1 C_2), \tag{4.3a}$$

$$J_\mu = 4|c_0|^2(S_1 C_2 a_i^{(1)} + S_2 a_i^{(2)}, iC_1 C_2), \tag{4.3b}$$

$$R_\mu = -4c_0^2(S_1 S_2 a_i^{(1)} + C_2 a_i^{(2)} + i a_i^{(3)}, iC_1 S_2), \tag{4.3c}$$

$$\bar{R}_\mu = -4c_0^{*2}(S_1 S_2 a_i^{(1)} + C_2 a_i^{(2)} - i a_i^{(3)}, iC_1 S_2), \tag{4.3d}$$

$$\tfrac{1}{2}\epsilon_{ijk} J_{jk} = J(C_2 a_i^{(1)} - S_1 S_2 a_i^{(2)}) + J_5 C_1 S_2 a_i^{(3)}, \tag{4.3e}$$

$$iJ_{i4} = -J_5(C_2 a_i^{(1)} - S_1 S_2 a_i^{(2)}) + J C_1 S_2 a_i^{(3)}, \tag{4.3f}$$

$$\begin{aligned}\tfrac{1}{2}\epsilon_{ijk} R_{jk} = &-4c_0^2\{-J(S_2 a_i^{(1)} - S_1 C_2 a_i^{(2)} - iS_1 a_i^{(3)})\\ &+ iJ_5(C_1 a_i^{(2)} + iC_1 C_2 a_i^{(3)})\}/(J^2 + J_5^2)^{1/2},\end{aligned} \tag{4.3g}$$

$$\begin{aligned}iR_{i4} = &-4c_0^2\{J_5(S_2 a_i^{(1)} - S_1 C_2 a_i^{(2)} - iS_1 a_i^{(3)})\\ &+ iJ(C_1 a_i^{(2)} + iC_1 C_2 a_i^{(3)})\}/(J^2 + J_5^2)^{1/2}\end{aligned} \tag{4.3h}$$

has been adopted previously to show that from the Fierz identities (3.2) and (3.3),[2] the spinor can be recovered, where $a_i^{(r)}$ $(r = 1,2,3)$ are an orthogonal triad, represented by the three Euler angles, and also

$$|c_0|^2 = \tfrac{1}{4}(J^2 + J_5^2)^{1/2} \tag{4.4}$$

and

$$\begin{aligned} C_i &= \cosh \xi_i\\ S_i &= \sinh \xi_i. \end{aligned} \quad (i = 1,2) \tag{4.5}$$

The parameter representation (4.3) contains eight parameters J, J_5, ξ_1, ξ_2, three Euler angles and the phase of c_0. It is straightforward to verify that quantities (4.3) satisfy all the Fierz constraint conditions.

V. DERIVATIVES OF SPINORS AND TENSORS

Physical quantities associated with a spinor field, such as the energy–momentum tensor, contain derivatives of the spinor field. A certain combination of derivatives of the spinor field can be written as a combination of derivatives of tensor quantities with the help of the Fierz identities (F-1)…(F-15).

First, it is not difficult to prove the relation, from Eq. (2.4),

$$\begin{aligned}&(\gamma_A(\partial_\rho - \overleftarrow{\partial}_\rho))[\gamma_B] - (\gamma_A)[\gamma_B(\partial_\rho - \overleftarrow{\partial}_\rho)]\\ &\quad = \tfrac{1}{4}\sum_C (\gamma_A \gamma_C](\partial_\rho - \overleftarrow{\partial}_\rho)[\gamma_B \gamma_C).\end{aligned} \tag{5.1}$$

If we choose

$$\begin{aligned}&\gamma_A = \gamma_\mu, \quad \gamma_B = \gamma_\nu,\\ &(\ = \bar{\psi}(x),\ \ ) = \psi(x),\\ &[\ = \bar{\psi}^c(x),\ \ ] = \psi(x),\end{aligned} \tag{5.2}$$

we can read off from (F-10) the relation

$$\begin{aligned}&4R_\nu(\bar{\psi} i\gamma_\mu \nabla_\rho \psi)\\ &\quad = -\delta_{\mu\nu} J_\alpha \nabla_\rho R_\alpha + J_\mu \nabla_\rho R_\nu + J_\nu \nabla_\rho R_\mu + iJ\nabla_\rho R_{\mu\nu}\\ &\qquad + iJ_5 \nabla_\rho\, {}^*R_{\mu\nu} + J_{\mu\alpha} \nabla_\rho R_{\nu\alpha} + {}^*J_{\mu\alpha} \nabla_\rho\, {}^*R_{\nu\alpha}\\ &\qquad - \epsilon_{\mu\nu\alpha\beta} J_{5\beta} \nabla_\rho R_\alpha,\end{aligned} \tag{5.3}$$

and, from (F-10) with $(\ = \ [\ = \psi,\ ) \ = \ \psi,\ ] = \psi^c$, we obtain

$$\begin{aligned}&4(\bar{\psi} i\gamma_\mu \nabla_\rho \psi)\bar{R}_\nu\\ &\quad = -\delta_{\mu\nu} \bar{R}_\alpha \nabla_\rho J_\alpha + \bar{R}_\mu \nabla_\rho J_\nu + \bar{R}_\nu \nabla_\rho J_\mu - i\bar{R}_{\mu\nu} \nabla_\rho J\\ &\qquad - i\,{}^*\bar{R}_{\mu\nu} \nabla_\rho J_5 + \bar{R}_{\mu\alpha} \nabla_\rho J_{\nu\alpha} + {}^*\bar{R}_{\mu\alpha} \nabla_\rho\, {}^*J_{\nu\alpha}\\ &\qquad + \epsilon_{\mu\nu\alpha\beta} \bar{R}_\alpha \nabla_\rho J_{5\beta},\end{aligned} \tag{5.4}$$

where

$$\nabla_\rho \equiv \partial_\rho - \overleftarrow{\partial}_\rho. \tag{5.5}$$

Hence, we have

$$\begin{aligned}&8(\bar{\psi} i\gamma_\mu(\partial_\rho - \overleftarrow{\partial}_\rho)\psi)\bar{R}_\nu R_\nu\\ &\quad = -\bar{R}_\mu(J_\alpha \nabla_\rho R_\alpha) - (\bar{R}_\alpha \nabla_\rho J_\alpha)R_\mu + \bar{R}_\nu(J_\mu \nabla_\rho R_\nu) + (\bar{R}_\mu \nabla_\rho J_\nu)R_\nu\\ &\qquad + \bar{R}_\nu(J_\nu \nabla_\rho R_\mu) + (\bar{R}_\nu \nabla_\rho J_\mu)R_\nu + i\bar{R}_\nu(J\nabla_\rho R_{\mu\nu}) - i(\bar{R}_{\mu\nu} \nabla_\rho J)R_\nu\\ &\qquad + i\bar{R}_\nu(J_5 \nabla_\rho\, {}^*R_{\mu\nu}) - i({}^*\bar{R}_{\mu\nu} \nabla_\rho J_5)R_\nu + \bar{R}_\nu(J_{\mu\alpha} \nabla_\rho R_{\nu\alpha}) + (\bar{R}_{\mu\alpha} \nabla_\rho J_{\mu\alpha})R_\nu\\ &\qquad + \bar{R}_\nu({}^*J_{\mu\alpha} \nabla_\rho\, {}^*R_{\nu\alpha}) + ({}^*R_{\mu\alpha} \nabla_\rho\, {}^*J_{\nu\alpha})R_\nu - \epsilon_{\mu\nu\alpha\beta}\{\bar{R}_\nu(J_{5\beta} \nabla_\rho R_\alpha) - (\bar{R}_\alpha \nabla_\rho J_{5\beta})R_\nu\}.\end{aligned} \tag{5.6}$$

The right-hand side of this equation can be simplified, with the help of the Fierz identities in Sec. III, as

$$(\bar{\psi} i\gamma_\mu(\partial_\rho - \overleftarrow{\partial}_\rho)\psi)$$

$$= \tfrac{1}{4} i\{ - \partial_\rho \bar{R}_\nu (JR_{\mu\nu} + J_5 {}^*R_{\mu\nu}) - (J\bar{R}_{\mu\nu} + J_5 {}^*\bar{R}_{\mu\nu})\partial_\rho R_\nu + 2\partial_\rho J_\nu (JJ_{\mu\nu} + J_5 {}^*J_{\mu\nu}) + 4J_{5\mu}(J_5(\partial_\rho - \overleftarrow{\partial}_\rho)J)\}/(J^2 + J_5^2) \quad (5.7a)$$

$$= \tfrac{1}{4}\{J_\mu(\bar{R}_\nu \nabla_\rho R_\nu) - (\bar{R}_\nu \nabla_\rho J_\nu)R_\mu - \bar{R}_\mu (J_\nu \nabla_\rho R_\nu) + 4iJ_{5\mu}(J_5 \nabla_\rho J)\}/(J^2 + J_5^2), \quad (5.7b)$$

where, from (5.7a) to (5.7b), we have used the relations (3.11e) and (3.11f). We emphasize that the relations (5.7) involves only the Fierz identities and no dynamics has been considered at all. This expression agrees, when μ and ρ are contracted, with the Lagrangian obtained by Zhelnorovich (apart from the mass term).[12]

If we take

$$\gamma_A = i\gamma_5\gamma_\mu, \quad \gamma_B = \gamma_\nu, \quad (5.8)$$

we arrive, after a similar calculation, at

$$(\bar{\psi} i\gamma_5\gamma_\mu(\partial_\rho - \overleftarrow{\partial}_\rho)\psi) = \tfrac{1}{4}\{\partial_\rho \bar{R}_\nu (J {}^*R_{\mu\nu} - J_5 R_{\mu\nu}) - (J {}^*\bar{R}_{\mu\nu} - J_5 \bar{R}_{\mu\nu})\partial_\rho R_\nu + 2i(JJ_{\mu\nu} + J_5 {}^*J_{\mu\nu})\partial_\rho J_{5\nu} + 4i(J\nabla_\rho J_5)J_\mu\}/(J^2 + J_5^2) \quad (5.9a)$$

$$= -\tfrac{1}{4}\{J_{5\mu}(\bar{R}_\nu \nabla_\rho R_\nu) - (\bar{R}_\nu \nabla_\rho J_{5\nu})R_\mu - \bar{R}_\mu (J_{5\nu}\nabla_\rho R_\nu) + 4i(J\nabla_\rho J_5)J_\mu\}/(J^2 + J_5^2). \quad (5.9b)$$

This is the quantity used by Israel and Nester to prove the positive definiteness of the gravitational energy.[6]

VI. DISCUSSION

By now it becomes quite obvious that the investigation of the Fierz identities and constraint conditions opens up the passage between spinors and tensors, namely the tensor system satisfying the Fierz constraint conditions implies the existence of a spinor. Thus, the existence of the Fierz constraint conditions can be regarded as the evidence of the existence of the spinor. Moreover, when we start from the Hermitian current J, J_5, J_μ, $J_{5\mu}$, and $J_{\mu\nu}$ to recover the spinor, the gauge freedom of the first kind emerges.[1] This fact suggests that the Fierz constraint conditions among the Hermitian currents to establish the existence of the spinor play the same role as the Bianchi identity to the gauge variant potential in gauge field theory. We have assumed throughout this paper that the spinor is a commuting c-number and the problem of quantization has been untouched. Suppose that four 4-vectors form a tetrad, namely the four 4-vectors satisfy orthonormality and completeness. To quantize such a system, we may apply the well-known Dirac method.[13] Due to the complications of orthonormality and completeness, the Dirac method is expected to be rather cumbersome. But, as we have learned, the tetrad implies the existence of a spinor, and it can be expressed in terms of the spinor as in Eq. (4.1). Once the tetrad is written by the spinor, the orthogonality and the completeness conditions are automatically satisfied as a result of the Fierz identities. Hence, we may ignore them. Thus, the introduction of spinor may simplify the quantization of the original system. In fact, this is the method employed by Hara and Goto to deal with the particle with internal structure.[14]

The problem of quantization of a tensor system will, however, be deferred for future occasion.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada.

APPENDIX: ALL FIERZ IDENTITIES

(I)

$$(\gamma_A)_{ab}(\gamma_B)_{cd} = \tfrac{1}{4}\sum_C (\gamma_A\gamma_C)_{ad}(\gamma_B\gamma_C)_{cb} = (\tfrac{1}{4})^2 \sum_{C,D} \mathrm{Tr}[\gamma_C\gamma_A\gamma_D\gamma_B](\gamma_C)_{ad}(\gamma_D)_{cb}, \quad (A1)$$

$$(\gamma_A = I, \ \gamma_5, \ \gamma_\mu, \ i\gamma_5\gamma_\mu, \ \sigma_{\mu\nu} = (1/2i)[\gamma_\mu, \gamma_\nu], \quad (A2)$$

$$\mathrm{Tr}[\gamma_A\gamma_B] = 4\delta_{AB}, \quad (A3)$$

$$(\gamma_A)[\gamma_B] = \tfrac{1}{4}\sum_C (\gamma_A\gamma_C][\gamma_B\gamma_C) = (\tfrac{1}{4})^2 \sum_{C,D} \mathrm{Tr}[\gamma_C\gamma_A\gamma_D\gamma_B](\gamma_C][\gamma_D), \quad (A4)$$

$${}^*\sigma_{\mu\nu} \equiv -\tfrac{1}{2} i\epsilon_{\mu\nu\lambda\rho}\sigma_{\lambda\rho} = i\gamma_5\sigma_{\mu\nu}, \quad (A5)$$

$${}^*\sigma_{\alpha\beta}\, {}^*\sigma_{\alpha\beta} = -\sigma_{\alpha\beta}\sigma_{\alpha\beta}, \quad (A6)$$

$$(\ ) \equiv (I), \quad (5) \equiv (i\gamma_5), \quad (\mu) \equiv (i\gamma_\mu), \quad (5\mu) \equiv (i\gamma_5\gamma_\mu), \quad (\mu\nu) \equiv (\sigma_{\mu\nu}), \quad ({}^*\mu\nu) \equiv (i\gamma_5\sigma_{\mu\nu}) = ({}^*\sigma_{\mu\nu}), \ \text{etc.}, \quad (A7)$$

$$4(\ )[\ ] = (\ ][\ ) - (5][5) - (\alpha][\alpha) + (5\alpha][5\alpha) + \tfrac{1}{2}(\alpha\beta][\alpha\beta), \quad (F\text{-}1)$$

$$4(\ )[5] = (\ ][5) + (5][\ ) - i\{(\alpha][5\alpha) - (5\alpha][\alpha)\} + \tfrac{1}{2}(\alpha\beta][{}^*\alpha\beta), \quad (F\text{-}2)$$

$$4(\ )[\lambda] = (\ ][\lambda) + (\lambda][\ ) + i\{(5][5\lambda) - (5\lambda][5)\} + i\{(\alpha][\lambda\alpha) - (\lambda\alpha][\alpha)\} + \{(5\alpha][{}^*\lambda\alpha) + ({}^*\lambda\alpha][5\alpha)\}, \quad (F\text{-}3)$$

$$4(\ )[5\lambda] = (\ ][5\lambda) + (5\lambda][\ ) + i\{(5][\lambda) - (\lambda][5)\} + \{(\alpha][{}^*\lambda\alpha) + ({}^*\lambda\alpha][\alpha)\} + i\{(5\alpha][\lambda\alpha) - (\lambda\alpha][5\alpha)\}, \tag{F-4}$$

$$\begin{aligned} 4(\ )[\lambda\rho) = {} & (\ ][\lambda\rho) + (\lambda\rho][\ ) - \{(5][{}^*\lambda\rho) + ({}^*\lambda\rho][5)\} - i\{(\lambda][\rho) - (\rho][\lambda)\} + i\{(5\lambda][5\rho) - (5\rho][5\lambda)\} \\ & + i\{(\lambda\alpha][\rho\alpha) - (\rho\alpha][\lambda\alpha)\} + i\epsilon_{\lambda\rho\alpha\beta}\{(\alpha][5\beta) + (5\beta][\alpha)\}, \end{aligned} \tag{F-5}$$

$$4(5)[5] = -(\ ][\ ) + (5][5) - (\alpha][\alpha) + (5\alpha][5\alpha) - \tfrac{1}{2}(\alpha\beta][\alpha\beta), \tag{F-6}$$

$$4(5)[\lambda] = -i\{(\ ][5\lambda) - (5\lambda][\ )\} + \{(5][\lambda) + (\lambda][5)\} + i\{(\alpha][{}^*\lambda\alpha) - ({}^*\lambda\alpha][\alpha)\} - \{(5\alpha][\lambda\alpha) + (\lambda\alpha][5\alpha)\}, \tag{F-7}$$

$$4(5)[5\lambda] = -i\{(\ ][\lambda) - (\lambda][\ )\} + \{(5][5\lambda) + (5\lambda][5)\} - \{(\alpha][\lambda\alpha) + (\lambda\alpha][\alpha)\} + i\{(5\alpha][{}^*\lambda\alpha) - ({}^*\lambda\alpha][5\alpha)\}, \tag{F-8}$$

$$\begin{aligned} 4(5)[\lambda\rho] = {} & (5][\lambda\rho) + (\lambda\rho][5) + (\ ][{}^*\lambda\rho) + ({}^*\lambda\rho][\ ) - \{(\lambda][5\rho) + (5\rho][\lambda)\} \\ & + \{(5\lambda][\rho) + (\rho][5\lambda)\} + i\{({}^*\lambda\alpha][\rho\alpha) - ({}^*\rho\alpha][\lambda\alpha)\} + \epsilon_{\lambda\rho\alpha\beta}\{(\alpha][\beta) - (5\alpha][5\beta)\}, \end{aligned} \tag{F-9}$$

$$\begin{aligned} 4(\mu)[\lambda] = {} & -\delta_{\mu\lambda}\{(\ ][\ ) + (5][5) + (\alpha][\alpha) + (5\alpha][5\alpha)\} + (\mu][\lambda) + (\lambda][\mu) + (5\mu][5\lambda) \\ & + (5\lambda][5\mu) + i\{(\ ][\mu\lambda) - (\mu\lambda][\ )\} + i\{(5][{}^*\mu\lambda) - ({}^*\mu\lambda][5)\} + (\mu\alpha][\lambda\alpha) \\ & + ({}^*\mu\alpha][{}^*\lambda\alpha) + \epsilon_{\mu\lambda\alpha\beta}\{(\alpha][5\beta) - (5\beta][\alpha)\}, \end{aligned} \tag{F-10}$$

$$\begin{aligned} 4(\mu)[5\lambda) = {} & \delta_{\mu\lambda}\{i(\ ][5) - i(5][\ ) - (\alpha][5\alpha) - (5\alpha][\alpha)\} + \{(\mu][5\lambda) + (5\lambda][\mu) + (\lambda][5\mu) + (5\mu][\lambda)\} + \{(\ ][{}^*\mu\lambda) \\ & + ({}^*\mu\lambda][\ ) - (5][\mu\lambda) - (\mu\lambda][5)\} + i\{({}^*\lambda\alpha][\mu\alpha) - (\mu\alpha][{}^*\lambda\alpha)\} + \epsilon_{\mu\lambda\alpha\beta}\{(\alpha][\beta) + (5\alpha][5\beta)\}, \end{aligned} \tag{F-11}$$

$$\begin{aligned} 4(\mu)[\lambda\rho) = {} & i\delta_{\mu\lambda}\{(\ ][\rho) - (\rho][\ ) - i(5][5\rho) - i(5\rho][5) - i(\alpha][\rho\alpha) - i(\rho\alpha][\alpha) + (5\alpha][{}^*\rho\alpha) - ({}^*\rho\alpha][5\alpha)\} \\ & - i\delta_{\mu\rho}\{(\ ][\lambda) - (\lambda][\ ) - i(5][5\lambda) - i(5\lambda][5) - i(\alpha][\lambda\alpha) - i(\lambda\alpha][\alpha) + (5\alpha][{}^*\lambda\alpha) - ({}^*\lambda\alpha][5\alpha)\} \\ & + \{(\mu][\lambda\rho) + (\lambda\rho][\mu) + (\lambda][\mu\rho) + (\mu\rho][\lambda) - (\rho][\mu\lambda) - (\mu\lambda][\rho)\} + i\{(5\mu][{}^*\lambda\rho) - ({}^*\lambda\rho][5\mu) \\ & + (5\lambda][{}^*\mu\rho) - ({}^*\mu\rho][5\lambda) - (5\rho][{}^*\mu\lambda) + ({}^*\mu\lambda][5\rho)\} - i\epsilon_{\mu\lambda\rho\alpha}\{(\ ][5\alpha) + (5\alpha][\ ) - i(5][\alpha) + i(\alpha][5)\}, \end{aligned} \tag{F-12}$$

$$\begin{aligned} 4(5\mu)[5\lambda] = {} & \delta_{\mu\lambda}\{(\ ][\ ) + (5][5) - (\alpha][\alpha) - (5\alpha][5\alpha)\} + (\mu][\lambda) + (\lambda][\mu) + (5\mu][5\lambda) + (5\lambda][5\mu) - i\{(\ ][\mu\lambda) - (\mu\lambda][\ ) \\ & + (5][{}^*\mu\lambda) - ({}^*\mu\lambda][5)\} - (\mu\alpha][\lambda\alpha) - ({}^*\mu\alpha][{}^*\lambda\alpha) + \epsilon_{\mu\lambda\alpha\beta}\{(5\alpha][\beta) - (\beta][5\alpha)\}, \end{aligned} \tag{F-13}$$

$$\begin{aligned} 4(5\mu)[\lambda\rho) = {} & \delta_{\mu\lambda}\{(5][\rho) + (\rho][5) + i(\ ][5\rho) - i(5\rho][\ ) + (5\alpha][\rho\alpha) + (\rho\alpha][5\alpha) + i(\alpha][{}^*\rho\alpha) - i({}^*\rho\alpha][\alpha)\} \\ & - \delta_{\mu\rho}\{(5][\lambda) + (\lambda][5) + i(\ ](5\lambda) - i(5\lambda][\ ) + (5\alpha][\lambda\alpha) + (\lambda\alpha][5\alpha) + i(\alpha][{}^*\lambda\alpha) - i({}^*\lambda\alpha][\alpha)\} \\ & + \{(5\mu][\lambda\rho) + (\lambda\rho][5\mu) + (5\lambda][\mu\rho) + (\mu\rho][5\lambda) - (5\rho][\mu\lambda) - (\mu\lambda][5\rho)\} \\ & + i\{(\mu][{}^*\lambda\rho) - ({}^*\lambda\rho][\mu) + (\lambda][{}^*\mu\rho) - ({}^*\mu\rho][\lambda) - (\rho][{}^*\mu\lambda) + ({}^*\mu\lambda][\rho)\} \\ & - i\epsilon_{\mu\lambda\rho\alpha}\{(\ ][\alpha) + (\alpha][\ ) - i(5][5\alpha) + i(5\alpha][5)\}, \end{aligned} \tag{F-14}$$

$$\begin{aligned} 4(\mu\nu)[\lambda\rho] = {} & (\delta_{\mu\lambda}\delta_{\nu\rho} - \delta_{\mu\rho}\delta_{\nu\lambda})\{(\ ][\ ) - (5][5) - (\alpha][\alpha) + (5\alpha][5\alpha)\} + \delta_{\mu\lambda}\{(\nu][\rho) + (\rho][\nu) - (5\nu][5\rho) - (5\rho][5\nu) \\ & - i(\ ][\nu\rho) + i(\nu\rho][\ ) + i(5][{}^*\nu\rho) - i({}^*\nu\rho][5) - (\nu\alpha][\rho\alpha)\} - \delta_{\mu\rho}\{(\nu][\lambda) + (\lambda][\nu) - (5\nu][5\lambda) - (5\lambda][5\nu) \\ & - i(\ ][\nu\lambda) + i(\nu\lambda][\ ) + i(5][{}^*\nu\lambda) - i({}^*\nu\lambda][5) - (\nu\alpha][\lambda\alpha)\} + \delta_{\nu\rho}\{(\mu][\lambda) + (\lambda][\mu) - (5\mu][5\lambda) - (5\lambda][5\mu) \\ & - i(\ ][\mu\lambda) + i(\mu\lambda][\ ) + i(5][{}^*\mu\lambda) - i({}^*\mu\lambda][5) - (\mu\alpha][\lambda\alpha)\} - \delta_{\nu\lambda}\{(\mu][\rho) + (\rho][\mu) - (5\mu][5\rho) - (5\rho][5\mu) \\ & - i(\ ][\mu\rho) + i(\mu\rho][\ ) + i(5][{}^*\mu\rho) - i({}^*\mu\rho][5) - (\mu\alpha][\rho\alpha)\} + \epsilon_{\alpha\nu\lambda\rho}\{(5\mu][\alpha) - (\mu][5\alpha)\} \\ & + \epsilon_{\mu\alpha\lambda\rho}\{(5\nu][\alpha) - (\nu][5\alpha)\} + \epsilon_{\mu\nu\alpha\rho}\{(\alpha][5\lambda) - (5\alpha][\lambda)\} + \epsilon_{\mu\nu\lambda\alpha}\{(\alpha][5\rho) - (5\alpha][\rho)\} + \epsilon_{\mu\nu\lambda\rho}\{(\ ][5) \\ & + (5][\ )\} + (\mu\nu][\lambda\rho) - ({}^*\mu\nu][{}^*\lambda\rho) + (\lambda\nu][\mu\rho) + (\mu\rho][\lambda\nu) + (\mu\lambda][\nu\rho) + (\nu\rho][\mu\lambda). \end{aligned} \tag{F-15}$$

(II)

$(\ ) = \bar\psi\psi \equiv J$, $(5) = \bar\psi i\gamma_5\psi \equiv J_5$, $(\mu) = \bar\psi i\gamma_\mu\psi \equiv J_\mu$, $(5\mu) = \bar\psi i\gamma_5\gamma_\mu\psi \equiv J_{5\mu}$, $(\mu\nu) = \bar\psi\sigma_{\mu\nu}\psi \equiv J_{\mu\nu}$.

Putting $(\cdots) = [\cdots] = (\cdots] = [\cdots) \equiv J\cdots$, we have

$$3J^2 + J_5^2 + J_\alpha J_\alpha - J_{5\alpha}J_{5\alpha} - \tfrac{1}{2}J_{\alpha\beta}J_{\alpha\beta} = 0, \tag{JJ-1}$$

$$2JJ_5 - \tfrac{1}{2}J_{\alpha\beta}\,{}^*J_{\alpha\beta} = 0, \tag{JJ-2}$$

$$JJ_\mu - J_{5\alpha}\,{}^*J_{\mu\alpha} = 0, \tag{JJ-3}$$

$$JJ_{5\mu} - J_\alpha\,{}^*J_{\mu\alpha} = 0, \tag{JJ-4}$$

$$JJ_{\lambda\rho} + J_5\,{}^*J_{\lambda\rho} - i\epsilon_{\lambda\rho\alpha\beta}J_\alpha J_{5\beta} = 0, \tag{JJ-5}$$

$$J^2 + 3J_5^2 + J_\alpha J_\alpha - J_{5\alpha}J_{5\alpha} + \tfrac{1}{2}J_{\alpha\beta}J_{\alpha\beta} = 0, \tag{JJ-6}$$

$$J_5J_\lambda + J_{5\alpha}J_{\lambda\alpha} = 0, \tag{JJ-7}$$

$$J_5 J_{5\lambda} + J_\alpha J_{\lambda\alpha} = 0, \tag{JJ-8}$$
$$J_5 J_{\lambda\rho} - J\,{}^*J_{\lambda\rho} + J_\lambda J_{5\rho} - J_{5\lambda} J_\rho = 0, \tag{JJ-9}$$
$$2J_\mu J_\lambda + \delta_{\mu\lambda}\{J^2 + J_5^2 + J_\alpha J_\alpha + J_{5\alpha}J_{5\alpha}\} - 2J_{5\mu}J_{5\lambda} - J_{\mu\alpha}J_{\lambda\alpha} - {}^*J_{\mu\alpha}\,{}^*J_{\lambda\alpha} = 0, \tag{JJ-10}$$
$$J_\mu J_{5\lambda} + \delta_{\mu\lambda} J_\alpha J_{5\alpha} - J_{5\mu} J_\lambda - J\,{}^*J_{\mu\lambda} + J_5 J_{\mu\lambda} = 0, \tag{JJ-11}$$
$$J_\mu J_{\lambda\rho} + J_\lambda J_{\rho\mu} + J_\rho J_{\mu\lambda} + i\epsilon_{\mu\lambda\rho\alpha} J J_{5\alpha} - \delta_{\mu\lambda}(J_\alpha J_{\rho\alpha} + J_5 J_{5\rho}) + \delta_{\mu\rho}(J_\alpha J_{\lambda\alpha} + J_5 J_{5\lambda}) = 0, \tag{JJ-12}$$
$$2J_{5\mu}J_{5\lambda} - 2J_\mu J_\lambda - \delta_{\mu\lambda}\{J^2 + J_5^2 - J_\alpha J_\alpha - J_{5\alpha}J_{5\alpha}\} + J_{\mu\alpha}J_{\lambda\alpha} + {}^*J_{\mu\alpha}\,{}^*J_{\lambda\alpha} = 0, \tag{JJ-13}$$
$$J_{5\mu}J_{\lambda\rho} + J_{5\lambda}J_{\rho\mu} + J_{5\rho}J_{\mu\lambda} + i\epsilon_{\mu\lambda\rho\alpha} J J_\alpha - \delta_{\mu\lambda}(J_{5\alpha}J_{\rho\alpha} + J_5 J_\rho) + \delta_{\mu\rho}(J_{5\alpha}J_{\lambda\alpha} + J_5 J_\lambda) = 0, \tag{JJ-14}$$
$$\begin{aligned} &3J_{\mu\nu}J_{\lambda\rho} + {}^*J_{\mu\nu}\,{}^*J_{\lambda\rho} - 2J_{\lambda\nu}J_{\mu\rho} - 2J_{\mu\lambda}J_{\nu\rho} - (\delta_{\mu\lambda}\delta_{\nu\rho} - \delta_{\mu\rho}\delta_{\nu\lambda})\{J^2 - J_5^2 - J_\alpha J_\alpha + J_{5\alpha}J_{5\alpha}\} \\ &\quad - \delta_{\mu\lambda}\{2J_\nu J_\rho - J_{\nu\alpha}J_{\rho\alpha} - 2J_{5\nu}J_{5\rho}\} + \delta_{\mu\rho}\{2J_\nu J_\lambda - J_{\nu\alpha}J_{\lambda\alpha} - 2J_{5\nu}J_{5\lambda}\} - \delta_{\nu\rho}\{2J_\mu J_\lambda - J_{\mu\alpha}J_{\lambda\alpha} - 2J_{5\mu}J_{5\lambda}\} \\ &\quad + \delta_{\nu\lambda}\{2J_\mu J_\rho - J_{\mu\alpha}J_{\rho\alpha} - 2J_{5\mu}J_{5\rho}\} - 2\epsilon_{\mu\nu\lambda\rho} J J_5 = 0. \end{aligned} \tag{JJ-15}$$

(III)

Putting

$$(\ \equiv \bar\psi^c, \quad) = \ ] \equiv \psi, \quad [\ \equiv \bar\psi, \quad (\ ) = (\ ] = 0, \quad (5) = (5] = 0,$$
$$(\mu) = (\mu] = \bar\psi^c i\gamma_\mu \psi \equiv R_\mu, \quad (5\mu) = (5\mu] = 0, \quad (\mu\nu) = (\mu\nu] = \bar\psi^c \sigma_{\mu\nu}\psi \equiv R_{\mu\nu}, \quad [\cdots] = [\cdots) = J\cdots,$$

we have

$$R_\alpha J_\alpha - \tfrac{1}{2} R_{\alpha\beta} J_{\alpha\beta} = 0, \tag{RJ-1}$$
$$R_\alpha J_{5\alpha} + \tfrac{1}{2} i R_{\alpha\beta}\,{}^*J_{\alpha\beta} = 0, \tag{RJ-2}$$
$$R_\lambda J + i R_\alpha J_{\lambda\alpha} - i R_{\lambda\alpha} J_\alpha + {}^*R_{\lambda\alpha} J_{5\alpha} = 0, \tag{RJ-3}$$
$$R_\lambda J_5 + R_{\lambda\alpha} J_{5\alpha} + i R_\alpha\,{}^*J_{\lambda\alpha} + i\,{}^*R_{\lambda\alpha} J_\alpha = 0, \tag{RJ-4}$$
$$R_{\lambda\rho} J - {}^*R_{\lambda\rho} J_5 - i R_\lambda J_\rho + i R_\rho J_\lambda + i R_{\lambda\alpha} J_{\rho\alpha} - i R_{\rho\alpha} J_{\lambda\alpha} + i\epsilon_{\lambda\rho\alpha\beta} R_\alpha J_{5\beta} = 0, \tag{RJ-5}$$
$$R_\alpha J_\alpha + \tfrac{1}{2} R_{\alpha\beta} J_{\alpha\beta} = 0, \tag{RJ-6}$$
$$R_\lambda J_5 + i R_\alpha\,{}^*J_{\lambda\alpha} - i\,{}^*R_{\lambda\alpha} J_\alpha - R_{\lambda\alpha} J_{5\alpha} = 0, \tag{RJ-7}$$
$$i R_\lambda J - R_\alpha J_{\lambda\alpha} - R_{\lambda\alpha} J_\alpha - i\,{}^*R_{\lambda\alpha} J_{5\alpha} = 0, \tag{RJ-8}$$
$$R_{\lambda\rho} J_5 + {}^*R_{\lambda\rho} J - R_\lambda J_{5\rho} + R_\rho J_{5\lambda} + i\,{}^*R_{\lambda\alpha} J_{\rho\alpha} - i\,{}^*R_{\rho\alpha} J_{\lambda\alpha} + \epsilon_{\lambda\rho\alpha\beta} R_\alpha J_\beta = 0, \tag{RJ-9}$$
$$3R_\mu J_\lambda - R_\lambda J_\mu + \delta_{\mu\lambda} R_\alpha J_\alpha + i R_{\mu\lambda} J + i\,{}^*R_{\mu\lambda} J_5 - R_{\mu\alpha} J_{\lambda\alpha} - {}^*R_{\mu\alpha}\,{}^*J_{\lambda\alpha} - \epsilon_{\mu\lambda\alpha\beta} R_\alpha J_{5\beta} = 0, \tag{RJ-10}$$
$$3R_\mu J_{5\lambda} - R_\lambda J_{5\mu} + \delta_{\mu\lambda} R_\alpha J_{5\alpha} - {}^*R_{\mu\lambda} J + R_{\mu\lambda} J_5 - i\,{}^*R_{\mu\alpha} J_{\lambda\alpha} + i R_{\mu\alpha}\,{}^*J_{\lambda\alpha} - \epsilon_{\mu\lambda\alpha\beta} R_\alpha J_\beta = 0, \tag{RJ-11}$$
$$\begin{aligned} &3R_\mu J_{\lambda\rho} - R_{\lambda\rho} J_\mu - R_\lambda J_{\mu\rho} - R_{\mu\rho} J_\lambda + R_\rho J_{\mu\lambda} + R_{\mu\lambda} J_\rho - \delta_{\mu\lambda}\{R_\alpha J_{\rho\alpha} + R_{\rho\alpha} J_\alpha - i\,{}^*R_{\rho\alpha} J_{5\alpha} - i R_\rho J\} \\ &\quad + \delta_{\mu\rho}\{R_\alpha J_{\lambda\alpha} + R_{\lambda\alpha} J_\alpha - i\,{}^*R_{\lambda\alpha} J_{5\alpha} - i R_\lambda J\} + i\{{}^*R_{\lambda\rho} J_{5\mu} + {}^*R_{\mu\rho} J_{5\lambda} + {}^*R_{\lambda\mu} J_{5\rho}\} - \epsilon_{\mu\lambda\rho\alpha} R_\alpha J_5 = 0, \end{aligned} \tag{RJ-12}$$
$$- \delta_{\mu\lambda} R_\alpha J_\alpha + R_\mu J_\lambda + R_\lambda J_\mu + i R_{\mu\lambda} J + i\,{}^*R_{\mu\lambda} J_5 - R_{\mu\alpha} J_{\lambda\alpha} - {}^*R_{\mu\alpha}\,{}^*J_{\lambda\alpha} - \epsilon_{\mu\lambda\alpha\beta} R_\beta J_{5\alpha} = 0, \tag{RJ-13}$$
$$\begin{aligned} &\delta_{\mu\lambda}\{R_\rho J_5 + R_{\rho\alpha} J_{5\alpha} + i R_\alpha\,{}^*J_{\rho\alpha} - i\,{}^*R_{\rho\alpha} J_\alpha\} - \delta_{\mu\rho}\{R_\lambda J_5 + R_{\lambda\alpha} J_{5\alpha} + i R_\alpha\,{}^*J_{\lambda\alpha} - i\,{}^*R_{\lambda\alpha} J_\alpha\} \\ &\quad + R_{\lambda\rho} J_{5\mu} + R_{\mu\rho} J_{5\lambda} + R_{\lambda\mu} J_{5\rho} + i\{R_\mu\,{}^*J_{\lambda\rho} + R_\lambda\,{}^*J_{\mu\rho} + R_\rho\,{}^*J_{\lambda\mu} \\ &\quad - {}^*R_{\lambda\rho} J_\mu - {}^*R_{\mu\rho} J_\lambda - {}^*R_{\lambda\mu} J_\rho\} - i\epsilon_{\mu\lambda\rho\alpha} R_\alpha J = 0, \end{aligned} \tag{RJ-14}$$
$$\begin{aligned} &- 3R_{\mu\nu} J_{\lambda\rho} - {}^*R_{\mu\nu}\,{}^*J_{\lambda\rho} + R_{\lambda\nu} J_{\mu\rho} + R_{\mu\rho} J_{\lambda\nu} + R_{\mu\lambda} J_{\nu\rho} + R_{\nu\rho} J_{\mu\lambda} - (\delta_{\mu\lambda}\delta_{\nu\rho} - \delta_{\mu\rho}\delta_{\nu\lambda}) R_\alpha J_\alpha \\ &\quad + \delta_{\mu\lambda}\{R_\nu J_\rho + R_\rho J_\nu + i R_{\nu\rho} J - i\,{}^*R_{\nu\rho} J_5 - R_{\nu\alpha} J_{\rho\alpha}\} - \delta_{\mu\rho}\{R_\nu J_\lambda + R_\lambda J_\nu + i R_{\nu\lambda} J - i\,{}^*R_{\nu\lambda} J_5 - R_{\nu\alpha} J_{\lambda\alpha}\} \\ &\quad + \delta_{\nu\rho}\{R_\mu J_\lambda + R_\lambda J_\mu + i R_{\mu\lambda} J - i\,{}^*R_{\mu\lambda} J_5 - R_{\mu\alpha} J_{\lambda\alpha}\} - \delta_{\nu\lambda}\{R_\mu J_\rho + R_\rho J_\mu + i\,{}^*R_{\mu\rho} J_5 - R_{\mu\alpha} J_{\rho\alpha}\} \\ &\quad - \{\epsilon_{\alpha\nu\lambda\rho} R_\mu + \epsilon_{\mu\alpha\lambda\rho} R_\nu\} J_{5\alpha} + \{\epsilon_{\mu\nu\alpha\rho} J_{5\lambda} + \epsilon_{\mu\nu\lambda\alpha} J_{5\rho}\} R_\alpha = 0. \end{aligned} \tag{RJ-15}$$

(IV)

Put

$$(\ = \bar\psi, \quad) = \psi, \quad [\ = \bar\psi^c, \quad] = \psi,$$
$$[\mu) = \bar\psi^c i\gamma_\mu \psi = R_\mu = [\mu], \quad [\mu\nu) = \bar\psi^c \sigma_{\mu\nu}\psi = R_{\mu\nu} = [\mu\nu], \quad (\cdots) = (\cdots] = J\cdots.$$

$$J_\alpha R_\alpha - \tfrac{1}{2} J_{\alpha\beta} R_{\alpha\beta} = 0, \tag{JR-1}$$
$$J_{5\alpha} R_\alpha - \tfrac{1}{2} i J_{\alpha\beta}\,{}^*R_{\alpha\beta} = 0, \tag{JR-2}$$
$$3JR_\lambda - iJ_\alpha R_{\lambda\alpha} + iJ_{\lambda\alpha} R_\alpha - J_{5\alpha}\,{}^*R_{\lambda\alpha} = 0, \tag{JR-3}$$

$$J_5R_\lambda - iJ_\alpha\, {}^*R_{\lambda\alpha} - i\, {}^*J_{\lambda\alpha}R_\alpha + J_{5\alpha}R_{\lambda\alpha} = 0, \tag{JR-4}$$

$$3JR_{\lambda\rho} + J_5\, {}^*R_{\lambda\rho} + iJ_\lambda R_\rho - iJ_\rho R_\lambda - iJ_{\lambda\alpha}R_{\rho\alpha} + iJ_{\rho\alpha}R_{\lambda\alpha} + i\epsilon_{\lambda\rho\alpha\beta}J_{5\alpha}R_\beta = 0, \tag{JR-5}$$

$$J_\alpha R_\alpha + \tfrac{1}{2}J_{\alpha\beta}R_{\alpha\beta} = 0, \tag{JR-6}$$

$$3J_5R_\lambda - iJ_\alpha\, {}^*R_{\lambda\alpha} + i\, {}^*J_{\lambda\alpha}R_\alpha + J_{5\alpha}R_{\lambda\alpha} = 0, \tag{JR-7}$$

$$JR_\lambda - iJ_\alpha R_{\lambda\alpha} - iJ_{\lambda\alpha}R_\alpha - J_{5\alpha}\, {}^*R_{\lambda\alpha} = 0, \tag{JR-8}$$

$$3J_5R_{\lambda\rho} - J\, {}^*R_{\lambda\rho} + J_{5\rho}R_\lambda - J_{5\lambda}R_\rho - i\, {}^*J_{\lambda\alpha}R_{\rho\alpha} + i\, {}^*J_{\rho\alpha}R_{\lambda\alpha} - \epsilon_{\lambda\rho\alpha\beta}J_\alpha R_\beta = 0, \tag{JR-9}$$

$$3J_\mu R_\lambda - J_\lambda R_\mu + \delta_{\mu\lambda}J_\alpha R_\alpha - iJR_{\mu\lambda} - iJ_5\, {}^*R_{\mu\lambda} - J_{\mu\alpha}R_{\lambda\alpha} - {}^*J_{\mu\alpha}\, {}^*R_{\lambda\alpha} - \epsilon_{\mu\lambda\alpha\beta}J_{5\alpha}R_\beta = 0, \tag{JR-10}$$

$$\delta_{\mu\lambda}J_{5\alpha}R_\alpha - J_{5\lambda}R_\mu - J_{5\mu}R_\lambda - J\, {}^*R_{\mu\lambda} + J_5R_{\mu\lambda} - i\, {}^*J_{\mu\alpha}R_{\lambda\alpha} + iJ_{\mu\alpha}\, {}^*R_{\lambda\alpha} - \epsilon_{\mu\lambda\alpha\beta}J_\alpha R_\beta = 0, \tag{JR-11}$$

$$\begin{aligned}&3J_\mu R_{\lambda\rho} - J_{\lambda\rho}R_\mu - J_{\mu\rho}R_\lambda - J_{\lambda\mu}R_\rho - J_\lambda R_{\mu\rho} - J_\rho R_{\lambda\mu} - i\delta_{\mu\lambda}\{JR_\rho - iJ_\alpha R_{\rho\alpha} - iJ_{\rho\alpha}R_\alpha + J_{5\alpha}\, {}^*R_{\rho\alpha}\}\\ &\quad + i\delta_{\mu\rho}\{JR_\lambda - iJ_\alpha R_{\lambda\alpha} - iJ_{\lambda\alpha}R_\alpha + J_{5\alpha}\, {}^*R_{\lambda\alpha}\} - i\{J_{5\mu}\, {}^*R_{\lambda\rho} + J_{5\lambda}\, {}^*R_{\mu\rho} - J_{5\rho}\, {}^*R_{\mu\lambda}\} + \epsilon_{\mu\lambda\rho\alpha}J_5R_\alpha = 0,\end{aligned} \tag{JR-12}$$

$$\delta_{\mu\lambda}J_\alpha R_\alpha - J_\mu R_\lambda - J_\lambda R_\mu + iJR_{\mu\lambda} + iJ_5\, {}^*R_{\mu\lambda} + J_{\mu\alpha}R_{\lambda\alpha} + {}^*J_{\mu\alpha}\, {}^*R_{\lambda\alpha} - \epsilon_{\mu\lambda\alpha\beta}J_{5\alpha}R_\beta = 0, \tag{JR-13}$$

$$\begin{aligned}&3J_{5\mu}R_{\lambda\rho} - J_{5\lambda}R_{\mu\rho} + J_{5\rho}R_{\mu\lambda} - \delta_{\mu\lambda}\{J_5R_\rho + J_{5\alpha}R_{\rho\alpha} + iJ_\alpha\, {}^*R_{\rho\alpha} - i\, {}^*J_{\rho\alpha}R_\alpha\}\\ &\quad + \delta_{\mu\rho}\{J_5R_\lambda + J_{5\alpha}R_{\lambda\alpha} + iJ_\alpha\, {}^*R_{\lambda\alpha} - i\, {}^*J_{\lambda\alpha}R_\alpha\}\\ &\quad - i\{J_\mu\, {}^*R_{\lambda\rho} + J_\lambda\, {}^*R_{\mu\rho} + J_\rho\, {}^*R_{\lambda\mu} - {}^*J_{\lambda\rho}R_\mu - {}^*J_{\mu\rho}R_\lambda - {}^*J_{\lambda\mu}R_\rho\} + i\epsilon_{\mu\lambda\rho\alpha}JR_\alpha = 0.\end{aligned} \tag{JR-14}$$

(JR-15) is the same as (RJ-15).

(V)

($=$ [$\equiv\bar{\psi}^c$,) $=$] $\equiv\psi$.

Then

$$R_\alpha R_\alpha - \tfrac{1}{2}R_{\alpha\beta}R_{\alpha\beta} = 0, \tag{RR-1}$$

$$R_{\alpha\beta}\, {}^*R_{\alpha\beta} = 0, \tag{RR-2}$$

$$0 = 0, \tag{RR-3}$$

$$R_\alpha\, {}^*R_{\lambda\alpha} = 0, \tag{RR-4}$$

$$0 = 0, \tag{RR-5}$$

$$R_\alpha R_\alpha + \tfrac{1}{2}R_{\alpha\beta}R_{\alpha\beta} = 0, \tag{RR-6}$$

$$0 = 0, \tag{RR-7}$$

$$R_\alpha R_{\lambda\alpha} = 0, \tag{RR-8}$$

$$0 = 0, \tag{RR-9}$$

$$2R_\mu R_\lambda + \delta_{\mu\lambda}R_\alpha R_\alpha - R_{\mu\alpha}R_{\lambda\alpha} - {}^*R_{\mu\alpha}\, {}^*R_{\lambda\alpha} = 0, \tag{RR-10}$$

$$0 = 0, \tag{RR-11}$$

$$\begin{aligned}&R_\mu R_{\lambda\rho} + R_\lambda R_{\rho\mu} + R_\rho R_{\mu\lambda}\\ &\quad - \delta_{\mu\lambda}R_\alpha R_{\rho\alpha} + \delta_{\mu\rho}R_\alpha R_{\lambda\alpha} = 0,\end{aligned} \tag{RR-12}$$

$$\delta_{\mu\lambda}R_\alpha R_\alpha - 2R_\mu R_\lambda + R_{\mu\alpha}R_{\lambda\alpha} + {}^*R_{\mu\alpha}\, {}^*R_{\lambda\alpha} = 0, \tag{RR-13}$$

$$0 = 0, \tag{RR-14}$$

$$\begin{aligned}&-3R_{\mu\nu}R_{\lambda\rho} - {}^*R_{\mu\nu}\, {}^*R_{\lambda\rho} + 2R_{\lambda\nu}R_{\mu\rho}\\ &\quad + 2R_{\mu\lambda}R_{\nu\rho} - (\delta_{\mu\lambda}\delta_{\nu\rho} - \delta_{\mu\rho}\delta_{\nu\lambda})R_\alpha R_\alpha\\ &\quad + \delta_{\mu\lambda}\{2R_\nu R_\rho - R_{\nu\alpha}R_{\rho\alpha}\} - \delta_{\mu\rho}\{2R_\nu R_\lambda - R_{\nu\alpha}R_{\lambda\alpha}\}\\ &\quad + \delta_{\nu\rho}\{2R_\mu R_\lambda - R_{\mu\alpha}R_{\lambda\alpha}\} - \delta_{\nu\lambda}\{2R_\mu R_\rho - R_{\mu\alpha}R_{\rho\alpha}\}\\ &\quad = 0.\end{aligned} \tag{RR-15}$$

(VI)

Putting

($\equiv\bar{\psi}^c$, ($\equiv\bar{\psi}$,) $=\bar{\psi}^c$,) $\equiv\psi$,

[$\equiv\bar{\psi}$, [$\equiv\bar{\psi}^c$,] $\equiv\psi$,] $\equiv\bar{\psi}^c$,

$$\bar{\psi}^c\psi^c = -\bar{\psi}\psi = -J, \quad \bar{\psi}^c i\gamma_5\psi^c = -\bar{\psi}i\gamma_5\psi = -J_5,$$

$$\bar{\psi}^c i\gamma_\mu\psi^c = \bar{\psi}i\gamma_\mu\psi = J_\mu,$$

$$\bar{\psi}^c i\gamma_5\gamma_\mu\psi^c = -\bar{\psi}i\gamma_5\gamma_\mu\psi = -J_{5\mu},$$

$$\bar{\psi}^c\sigma_{\mu\nu}\psi^c = \bar{\psi}\sigma_{\mu\nu}\psi = J_{\mu\nu},$$

$$\bar{\psi}i\gamma_\mu\psi^c \equiv \bar{R}_\mu = (R_i^*, iR_0^*),$$

$$\bar{\psi}\sigma_{\mu\nu}\psi^c \equiv \bar{R}_{\mu\nu} = (R_{ij}^*, iR_{0j}^*)$$

we have

$$4J^2 = R_\alpha\bar{R}_\alpha - \tfrac{1}{2}R_{\alpha\beta}\bar{R}_{\alpha\beta}, \tag{R\bar{R}-1}$$

$$4JJ_5 = -\tfrac{1}{2}R_{\alpha\beta}\, {}^*\bar{R}_{\alpha\beta}, \tag{R\bar{R}-2}$$

$$4JJ_\lambda = -iR_\alpha\bar{R}_{\lambda\alpha} + iR_{\lambda\alpha}\bar{R}_\alpha, \tag{R\bar{R}-3}$$

$$4JJ_{5\lambda} = -R_\alpha\, {}^*\bar{R}_{\lambda\alpha} - {}^*R_{\lambda\alpha}\bar{R}_\alpha, \tag{R\bar{R}-4}$$

$$4JJ_{\lambda\rho} = iR_\lambda\bar{R}_\rho - iR_\rho\bar{R}_\lambda - iR_{\lambda\alpha}\bar{R}_{\rho\alpha} + iR_{\rho\alpha}\bar{R}_{\lambda\alpha}, \tag{R\bar{R}-5}$$

$$4J_5^2 = R_\alpha\bar{R}_\alpha + \tfrac{1}{2}R_{\alpha\beta}\bar{R}_{\alpha\beta}, \tag{R\bar{R}-6}$$

$$4J_5J_\lambda = -iR_\alpha\, {}^*\bar{R}_{\lambda\alpha} + i\, {}^*R_{\lambda\alpha}\bar{R}_\alpha, \tag{R\bar{R}-7}$$

$$4J_5J_{5\lambda} = R_\alpha\bar{R}_{\lambda\alpha} + R_{\lambda\alpha}\bar{R}_\alpha, \tag{R\bar{R}-8}$$

$$4J_5J_{\lambda\rho} = -i\, {}^*R_{\lambda\alpha}\bar{R}_{\rho\alpha} + i\, {}^*R_{\rho\alpha}\bar{R}_{\lambda\alpha} - \epsilon_{\lambda\rho\alpha\beta}R_\alpha\bar{R}_\beta, \tag{R\bar{R}-9}$$

$$\begin{aligned}4J_\mu J_\lambda &= -\delta_{\mu\lambda}R_\alpha\bar{R}_\alpha + R_\mu\bar{R}_\lambda\\ &\quad + R_\lambda\bar{R}_\mu + R_{\mu\alpha}\bar{R}_{\lambda\alpha} + {}^*R_{\mu\alpha}\, {}^*\bar{R}_{\lambda\alpha},\end{aligned} \tag{R\bar{R}-10}$$

$$4J_\mu J_{5\lambda} = i\, {}^*R_{\mu\alpha}\bar{R}_{\lambda\alpha} - iR_{\mu\alpha}\, {}^*\bar{R}_{\lambda\alpha} + \epsilon_{\mu\lambda\alpha\beta}R_\alpha\bar{R}_\beta, \tag{R\bar{R}-11}$$

$$\begin{aligned}4J_\mu J_{\lambda\rho} &= \delta_{\mu\lambda}\{R_\alpha\bar{R}_{\rho\alpha} + R_{\rho\alpha}\bar{R}_\alpha\}\\ &\quad - \delta_{\mu\rho}\{R_\alpha\bar{R}_{\lambda\alpha} + R_{\lambda\alpha}\bar{R}_\alpha\} + R_\mu\bar{R}_{\lambda\rho}\\ &\quad + R_\lambda\bar{R}_{\mu\rho} + R_\rho\bar{R}_{\lambda\mu}\\ &\quad + R_{\lambda\rho}\bar{R}_\mu + R_{\mu\rho}\bar{R}_\lambda + R_{\lambda\mu}\bar{R}_\rho,\end{aligned} \tag{R\bar{R}-12}$$

$$\begin{aligned}4J_{5\mu}J_{5\lambda} &= \delta_{\mu\lambda}R_\alpha\bar{R}_\alpha - R_\mu\bar{R}_\lambda - R_\lambda\bar{R}_\mu\\ &\quad + R_{\mu\alpha}\bar{R}_{\lambda\alpha} + {}^*R_{\mu\alpha}\, {}^*\bar{R}_{\lambda\alpha},\end{aligned} \tag{R\bar{R}-13}$$

$$
\begin{aligned}
4J_{5\mu}J_{\lambda\rho} = & \; i\delta_{\mu\lambda}\{{}^*R_{\rho\alpha}\bar{R}_\alpha - R_\alpha\,{}^*\bar{R}_{\rho\alpha}\} \\
& - i\delta_{\mu\rho}\{{}^*R_{\lambda\alpha}\bar{R}_\alpha - R_\alpha\,{}^*\bar{R}_{\lambda\alpha}\} \\
& - i\{R_\mu\,{}^*\bar{R}_{\lambda\rho} + R_\lambda\,{}^*\bar{R}_{\mu\rho} + R_\rho\,{}^*\bar{R}_{\lambda\mu} \\
& - {}^*R_{\lambda\rho}\bar{R}_\mu - {}^*R_{\mu\rho}\bar{R}_\lambda - {}^*R_{\lambda\mu}\bar{R}_\rho\},
\end{aligned}
\qquad (R\bar{R}\text{-}14)
$$

$$
\begin{aligned}
4J_{\mu\nu}J_{\lambda\rho} = & \; (\delta_{\mu\rho}\delta_{\nu\lambda} - \delta_{\mu\lambda}\delta_{\nu\rho})R_\alpha\bar{R}_\alpha \\
& + \delta_{\mu\lambda}\{R_\nu\bar{R}_\rho + R_\rho\bar{R}_\nu - R_{\nu\alpha}\bar{R}_{\rho\alpha}\} \\
& - \delta_{\mu\rho}\{R_\nu\bar{R}_\lambda + R_\lambda\bar{R}_\nu - R_{\mu\alpha}\bar{R}_{\nu\alpha}\} \\
& + \delta_{\nu\rho}\{R_\mu\bar{R}_\lambda + R_\lambda\bar{R}_\mu - R_{\mu\alpha}\bar{R}_{\lambda\alpha}\} \\
& - \delta_{\nu\lambda}\{R_\mu\bar{R}_\rho + R_\rho\bar{R}_\mu - R_{\mu\alpha}\bar{R}_{\rho\alpha}\} \\
& + R_{\mu\nu}\bar{R}_{\lambda\rho} - {}^*R_{\mu\nu}\,{}^*\bar{R}_{\lambda\rho} + R_{\lambda\nu}\bar{R}_{\mu\rho} \\
& + R_{\mu\rho}\bar{R}_{\lambda\nu} + R_{\mu\lambda}\bar{R}_{\nu\rho} + R_{\nu\rho}\bar{R}_{\mu\lambda}.
\end{aligned}
\qquad (R\bar{R}\text{-}15)
$$

[1]Y. Takahashi, Phys. Rev. D **26**, 2169 (1982).

[2]Y. Takahashi and K. Okuda, "A spinorization of a constrained vector system and a spinor reconstruction theorem," Univ. of Alberta Preprint Alberta Thy-12-1982, 1982.

[3]See E. H. Lieb and D. C. Mattis, *Mathematical Physics in One Dimension* (Academic, New York, 1966).

[4]See, for example, Freedman's review: D. Z. Freedman, in Proceedings of the 19th International Conference on High Energy Physics, Tokyo 1978, p. 535.

[5]See, H. Fritsch and P. Minkowski, Ann. Phys. (N.Y.) **93**, 193 (1975); H. Georgi and U. Nanopoulus, Nucl. Phys. **155**, 52 (1979).

[6]J. M. Nester, Phys. Lett. A **83**, 241 (1981); W. Israel and J. M. Nester, Phys. Lett. A **85**, 259 (1981).

[7]Y. Takahashi, *An Introduction to Field Quantization* (Pergamon, Oxford, 1968).

[8]M. Fierz, Z. Phys. **104**, 553 (1937).

[9]For the *JJ*-type Fierz conditions, see, for example, F. A. Kaempffer, Phys. Rev. D **23**, 918 (1981); J. D. Dumais, M.Sc. thesis, University of Alberta, 1975.

[10]For the independence of the first three relations, see Dumais' thesis in Ref. 9.

[11]T. Takabayasi, Nuovo Cimento **X7**, 118 (1958).

[12]V. A. Zhelnorovich, Dokl. Akad. Nauk SSSR **249**, 85 (1979) [Sov. Phys. Dokl. **24**, 899 (1979)].

[13]P. A. M. Dirac, lectures on quantum mechanics, Belfer Graduate School of Science, Yeshiva University, New York, 1964.

[14]O. Hara and T. Goto, Prog. Theor. Phys. **44**, 1383 (1970).

Superoperator perturbation theory for propagators

Brian Weiner
Quantum Theory Project, Department of Chemistry, University of Florida, Gainesville, Florida 32611

(Received 23 November 1982; accepted for publication 10 December 1982)

A well-defined superoperator perturbation theory for propagators is developed, based on equivalence classes of operators, which avoids the ambiguity of approaches based on a degenerate inner product. The Van Vleck formalism provides a natural tool for such a theory when self-consistent propagator approximations are chosen as zeroth-order approximations.

PACS numbers: 03.65.Bz

I. INTRODUCTION

Superoperators, as their name suggests, are glorified operators that map operators to operators. Intrinsically, however, they are not essentially different from operators as both are mappings from one linear space to another. The best known example of a superoperator is the Liouville operator $\mathscr{L}$ which is the infinitesimal generator of the time evolution of a quantum system, and thus acts on the algebra of observables of such a system. The action of $\mathscr{L}$ is defined in terms of the Hamiltonian H according to

$$\mathscr{L}(A) = [H, A] \tag{1.1}$$

and the super time evolution operator $\hat{U}(t)$ is given in terms of $\mathscr{L}$ by

$$\hat{U}(t)(A) = e^{i\mathscr{L}t}(A) = e^{iHt} A\, e^{-iHt}, \tag{1.2}$$

where A is an observable. The Heisenberg equation of motion is thus expressed as

$$-i\frac{\partial}{\partial t} A(t) = \mathscr{L}(A(t)) \tag{1.3}$$

and the Liouville–Von Neumann equation for the density operators $D(t)$ as

$$i\frac{\partial}{\partial t} D(t) = \mathscr{L}(D(t)). \tag{1.4}$$

The superoperator formalism has been used extensively in the field of nonequilibrium quantum statistics since the late 1950's[1] and the fairly recent review by Penrose[2] traces its development there. The tools of partitioning and inner projection, utilized to great effect by Löwdin in perturbation theory,[3] have played an essential role in the above topic. A comprehensive examination of perturbation theory for superoperators within this approach has just been completed[4] and is a good guide to the subject.

Superoperators were also introduced into a degenerate perturbation theory using the Van Vleck formalism[5] by Primas[6] (see references therein for earlier accounts of superoperator theory). The Van Vleck approach, as expounded by Redmon and Bartlett,[7] is an essential ingredient of the material presented in this paper.

It was not until 1970 that the superoperator formalism appeared in molecular quantum mechanics; then Goscinski and Lukman[8] applied it to propagator approximations. In the following years, many applications and extensions[9] further refined and improved this form of propagator perturbation theory.

The jump from operators to superoperators is, however, not completely free from mathematical snags, as normally operators are defined in a Hilbert space while superoperators are defined in the set of linear operators $\mathscr{L}(\mathscr{H})$, acting in $\mathscr{H}$.

The space $\mathscr{L}(\mathscr{H})$ is not a Hilbert space unless $\mathscr{H}$ is finite dimensional, though the bounded elements of $\mathscr{L}(\mathscr{H})$ form a Banach space $\mathscr{B}(\mathscr{H})$ with respect to the operator norm

$$\| A \| = \sup \|Av\|_{\mathscr{H}},\ \|v\|_{\mathscr{H}} = 1, \tag{1.5}$$

where $\| \ \|_{\mathscr{H}}$ is the norm in $\mathscr{H}$. Hence care must be exercised in the carrying over of Hilbert space operator properties to superoperators. In order to avoid such difficulties the subspace $\mathscr{B}_2(\mathscr{H})$ of Hilbert–Schmidt operators is often used as a core in which superoperators are defined; the space $\mathscr{B}_2(\mathscr{H})$ is a Hilbert space with inner product given by

$$(A \mid B)_{\mathrm{HS}} = \mathrm{Tr}\, A^\dagger B \tag{1.6}$$

[if $\dim(\mathscr{H}) = \infty$, then $\mathscr{B}_2(\mathscr{H})$ is properly contained in $\mathscr{L}(\mathscr{H})$]. The superoperators so defined may then be extended to larger domains.[10] A further practical difficulty associated with $\mathscr{L}(\mathscr{H})$ not being a Hilbert space is the lack of an inner product by which to construct matrix representations.

The propagator superoperator approach is based on an inner product space of operators; the Hermitian inner product involves the state density operator D, viz.,

$$(A \mid B)_\pm = \mathrm{Tr}\,\{[A^\dagger, B]_\pm\, D\}, \tag{1.7}$$

where

$$[A^\dagger, B]_\pm = A^\dagger B \pm BA^\dagger$$

and the sign used depends upon the type of operators A and B. A major drawback with this definition of inner product is that it is degenerate, i.e.,

$$(A \mid B)_\pm = 0 \quad \forall A \not\Rightarrow B = 0; \tag{1.8}$$

consequently, the superoperator adjoint operation and resolutions of the identity derived from this inner product are not well defined. These deficiencies are rather important as superoperator propagator perturbation theory is based on first forming an inner projection of a superoperator resolvent and then applying the partitioning technique (which both depend on a resolution of the identity) in conjunction with using the superoperator adjoint operation. A common way to surmount this problem in mathematical texts is to introduce

 0022-2488/83/071791-06$02.50

equivalence classes so that this degeneracy is removed when the inner product is defined on the space of equivalence classes. However, in the case of $(\ |\ )_-$ all operators are members of the zero equivalence class, clearly an unsatisfactory situation. In order to obtain a unified theory for both boson- and fermion-like operators it is possible to define the Hermitian inner product[11]

$$(A \mid B) = \mathrm{Tr}\,\{A^{\dagger}BD\}, \tag{1.9}$$

which in general is again degenerate. However, if one defines the equivalence classes

$$\Phi(X) = \{X + \eta;\ \mathrm{Tr}\{\eta^{\dagger}\eta D\} = 0\} \tag{1.10}$$

and considers $(\ |\ )$ to act in the space of such classes, it is then nondegenerate. The Hilbert of space $\mathscr{H}_D$ of equivalence classes so constructed is the one utilized in the Gel'fand–Naimark–Segal (GNS) construction to obtain representations of C^*-algebras.[12]

Hilbert spaces of equivalence classes induced by the inner product (1.9) have been used to develop a theory of self-consistent propagator approximation,[13] which will be briefly described in Sec. II of this article. In Sec. IV, it will be shown that such approximations form a natural starting point for a superoperator perturbation theory for propagators based on a Van Vleck approach (the salient features of which are quickly reviewed in Sec. III).

Every propagator can be canonically associated with a zeroth-order operator space—the space of operators that define it, i.e., the one-electron propagator with the space $\mathbf{f}_1$, the particle-hole/polarization type with $\mathbf{b}_1$, the two-electron one with $\mathbf{f}_2$, etc., where

$$\mathbf{f}_1 = \text{linear span}\ \{a_i,\ a_i^{\dagger};\ 1 \leqslant i \leqslant r\},$$

$$\mathbf{b}_1 = \text{linear span}\ \{a_i^{\dagger} a_j;\ 1 \leqslant i,j \leqslant r\},$$

$$\mathbf{f}_2 = \text{linear span}\ \{a_i a_j,\ a_j^{\dagger} a_i^{\dagger};\ 1 \leqslant i \leqslant j \leqslant r\}$$

(the operators, a_i, $a_i^{\dagger}$ are discrete field operators, based on a given basis of one-particle space $\mathscr{H}^1$). Each one of these spaces defines a self-consistent approximation to the associated propagator that corresponds to a model Hamiltonian and an approximate ground state $\tilde{\psi}$ that satisfies the vacuum condition for that manifold, viz.,

$$\{Q_k^{\dagger}\,\tilde{\psi},\ \tilde{\psi}, k = 1,...,\nu\}\ \text{is an orthonormal set of vectors} \tag{1.11}$$

and

$$Q_k \tilde{\psi} = 0, \quad k = 1,...,\nu, \tag{1.12}$$

where $\{Q_k, Q_k^{\dagger}, k = 1,...,\nu\}$ are linear combinations of operators from the appropriate space.

As higher order propagators determine lower order ones, higher order manifolds can either be considered as defining zeroth-order approximations (which we will take to be of the self-consistent type) to higher order propagators which implicitly define improved approximations to lower order propagators, or as manifolds that explicitly determine corrections to the lower order propagators, e.g., the manifold $\mathbf{f}_2$ determines as a self-consistent approximation to the two-electron propagators such that

$$\{Q_k, Q_k^{\dagger}, 1 \leqslant k \leqslant \tbinom{r}{2}\} \in \mathbf{f}_2 \tag{1.13}$$

and satisfy (1.11) and (1.12) for some approximate state $\tilde{\psi}$, which in turn determines an approximation to the one-electron propagator. Alternatively, one could correct the one-electron propagator perturbatively by the explicit inclusion of terms that involve elements of $\mathbf{f}_2$.

The type of perturbative corrections that result from the theory described in this manuscript ensure that at each order of correction one has a self-adjoint model Hamiltonian as well as ground and excited states. We can thus always construct representable propagators, i.e., propagators associated with a given ground state, that provide real excitation energies, in contrast, to polarization propagators calculated within the random phase approximation (RPA) that may predict complex energies.

II. SELF-CONSISTENT PROPAGATOR THEORY

If Ψ is the exact ground state of the Hamiltonian H, then we can construct a representation $\mathscr{L}_\Psi$ of the Liouville operator $\mathscr{L}$ acting in the Hilbert space of equivalence classes $\mathscr{H}_\Psi$, where these classes are defined by (1.10). When $D = |\Psi\rangle\langle\Psi|$ the inner product in this space is given by

$$(\Phi(A)|\Phi(B)) = \mathrm{Tr}\,\{A^{\dagger}B\,|\Psi\rangle\langle\Psi|\} = \langle\Psi|\,A^{\dagger}B\Psi\rangle. \tag{2.1}$$

The action of $\mathscr{L}_\Psi$ is defined by

$$\mathscr{L}_\Psi\cdot\Phi(A) = \Phi(\mathscr{L}(A)) = \Phi([H, A]). \tag{2.2}$$

It can be shown that (2.2) gives á well-defined symmetric operator if and only if

$$\langle\Psi|\,[H,A^{\dagger}B]\,\Psi\rangle = 0 \quad \forall\ A,B \tag{2.3}$$

and H is symmetric.[13]

The operator $\mathscr{L}_\Psi$ is in fact GNS representation of H-$\langle\Psi|\,H\Psi\rangle\cdot P_\Psi$, i.e.,

$$\Pi(H - EP_\Psi) = \mathscr{L}_\Psi, \tag{2.4}$$

where E is the ground state energy, P_Ψ is the projector onto Ψ, and the GNS representation for the state Ψ is defined by

$$\Pi(A)\cdot\Phi(B) = \Phi(AB) \quad \forall\ A,B. \tag{2.5}$$

The resolvent $(zI_\Psi - \mathscr{L}_\Psi)^{-1}$ is a well-defined Hilbert space operator and can thus be handled in standard ways.

The above resolvent is particularly useful in the construction of the Laplace transformations of various propagators. The causal propagator based on the operator manifold $\mathbf{M}$ can be expressed as

$$G^c(z) = (\Phi(\mathbf{t})|(zI_\Psi - \mathscr{L}_\Psi)^{-1}\Phi(\mathbf{t})) \pm \overline{\mathbf{\Delta}}(\Phi(\mathbf{t})|(\bar{z}I_\Psi + \mathscr{L}_\Psi)^{-1}\Phi(\mathbf{t}))^T\mathbf{\Delta}^T, \tag{2.6}$$

where

(i) $\Phi(\mathbf{t}) = \{\Phi(t_k);\ 1 \leqslant k \leqslant \nu\}$,

(ii) $\Phi(\mathbf{t}^{\dagger}) = \Phi(\mathbf{t}^{\dagger})\,\mathbf{\Delta}^{+}$,

(iii) $\mathbf{\Delta}_{kk'} = (\Phi(t_k^{\dagger})|\,\Phi(t_{k'})) = \langle\Psi\,|t_k t_{k'}\Psi\rangle, \quad 1 \leqslant k,k' \leqslant \nu$,

(iv) the image of the operator manifold $\mathbf{M}$ under Φ, i.e., $\Phi(\mathbf{M})$, is spanned by the vectors $\{\Phi(t_k);\ 1 \leqslant k \leqslant \nu\}$, and

(v) $\mathbf{\Delta}_T$ denotes the transpose of $\mathbf{\Delta}$, while $\overline{\mathbf{\Delta}}$ denotes the complex conjugate matrix. The retarded propagator is similarly given as

$$\mathbf{G}^R(Z) = (\Phi(\mathbf{t})|(zI_\Psi - \mathcal{L}_\Psi)^{-1}\Phi(\mathbf{t})) \pm \bar{\mathbf{\Delta}}(\Phi(\mathbf{t})|(zI_\Psi + \mathcal{L}_\Psi)^{-1}\Phi(\mathbf{t}))^T\mathbf{\Delta}^T. \tag{2.7}$$

The matrices of the resolvent operators appearing in (2.6) and (2.7) can be expressed in terms of the resolvents of the matrix $\mathcal{L}_\Psi$ of $\mathcal{L}$, by using the partitioning technique. Let

$$\mathcal{H}_\Psi = \Phi(\mathbf{M}) + \Phi(\mathbf{M})^\perp \tag{2.8}$$

so that

$$\Phi(\mathbf{M})^\perp = \Phi(\mathbf{M}^c), \tag{2.9}$$

where as a Banach space

$$\mathcal{B}(\mathcal{H}) = \mathbf{M} + \mathbf{M}^c. \tag{2.10}$$

Then,

$$\begin{aligned}(\Phi(\mathbf{t})|(zI_\Psi - \mathcal{L}_\Psi)^{-1}\Phi(\mathbf{t})) &= [(\Phi(\mathbf{t})|(zI_\Psi - \mathcal{L}_\Psi)\Phi(\mathbf{t})) \\ &\quad - (\Phi(\mathbf{t})|\mathcal{L}_\Psi\Phi(\mathbf{u}))(\Phi(\mathbf{u})|(zI_\Psi \\ &\quad - \mathcal{L}_\Psi)\Phi(\mathbf{u}))^{-1}(\Phi(\mathbf{u})|\mathcal{L}_\Psi\Phi(\mathbf{t}))]^{-1},\end{aligned} \tag{2.11}$$

where $\{\Phi(u_k), 1\leqslant k\leqslant \nu'\}$ is a complete orthonormal basis for $\Phi(\mathbf{M})^\perp$. Defining the "self-energy" term $\Sigma(z)$ by

$$\begin{aligned}\Sigma(z) &= (\Phi(\mathbf{t})|\mathcal{L}_\Psi\Phi(\mathbf{u}))(\Phi(\mathbf{u})|(zI_\Psi \\ &\quad - \mathcal{L}_\Psi)\Phi(\mathbf{u}))^{-1}(\Phi(\mathbf{u})|\mathcal{L}_\Psi\Phi(\mathbf{t})),\end{aligned} \tag{2.12}$$

one can write (2.11) in the succinct form

$$(\Phi(\mathbf{t})|(zI_\Psi - \mathcal{L}_\Psi)^{-1}\Phi(\mathbf{t})) = [(\Phi(\mathbf{t})|(zI_\Psi - \mathcal{L}_\Psi)\Phi(\mathbf{t})) - \Sigma(z)]^{-1}. \tag{2.13}$$

Expressions for the causal and retarded propagator can be obtained from (2.13).

The condition (2.3) for the existence of $\mathcal{L}_\Psi$ will not be satisfied unless Ψ is an exact stationary state of H, a situation that in practice is hardly ever realized. As the manifold $\mathbf{M}$ is of prime concern, the weaker condition

$$\langle\tilde{\Psi}|[H,A^+B]\,\tilde{\Psi}\rangle = 0 \quad \forall\ A,B\in\mathbf{M} \tag{2.14}$$

suggests itself as characterizing approximations to Ψ. If it is satisfied, it leads to the following definition for an approximation $K_{\tilde{\Psi}}$ to $\mathcal{L}_\Psi$

$$\begin{aligned}K_{\tilde{\Psi}}\,\tilde{\Phi}(A) &= P_{\tilde{\Psi}}(\mathbf{M})\tilde{\Phi}(\mathcal{L}(A)) \quad \forall\ A\in\mathbf{M}, \\ K_{\tilde{\Psi}}\tilde{\Phi}(A) &= P_{\tilde{\Psi}}(\mathbf{M})^\perp\tilde{\Phi}(\mathcal{L}(A)) \quad \forall\in\mathbf{M}^c,\end{aligned} \tag{2.15}$$

where $P_{\tilde{\Psi}}(\mathbf{M})$ is the orthogonal projector onto $\tilde{\Phi}(\mathbf{M})$, $P_{\tilde{\Psi}}(\mathbf{M})^\perp$ the projector onto the orthogonal compliment and $\tilde{\Phi}$ is the linear map to the Hilbert space of equivalence classes on $\tilde{\Psi}$. One obtains directly from (2.15) that

$$K_{\tilde{\Psi}} = P_{\tilde{\Psi}}(\mathbf{M})\,K_{\tilde{\Psi}}P_{\tilde{\Psi}}(\mathbf{M}) + P_{\tilde{\Psi}}(\mathbf{M})^\perp K_{\tilde{\Psi}}P_{\tilde{\Psi}}(\mathbf{M})^\perp, \tag{2.16}$$

i.e., that the spaces $\tilde{\Phi}(M)$ and $\tilde{\Phi}(M)^\perp$ are not coupled by $K_{\tilde{\Psi}}$. As $\tilde{\Psi}$ is a vector state (see, for example, Ref. 14), the GNS representation $\Pi_{\tilde{\Psi}}$ is a C^*-algebra isomorphism from $\mathcal{B}(\mathcal{H})\rightarrow\mathcal{B}(\mathcal{H}_{\tilde{\Psi}})$ and the image of $K_{\tilde{\Psi}}$ under $\pi_{\tilde{\Psi}}^{-1}$ is given by

$$K = \Pi_{\tilde{\Psi}}^{-1}(K_{\tilde{\Psi}}) = QHQ + P^\perp HP^\perp - \langle\tilde{\Psi}|H\tilde{\Psi}\rangle\,P_\Psi, \tag{2.17}$$

where the orthogonal projectors P and Q are given by

$$P = \sum_{k=1}^{\mu} q_k^\dagger|\tilde{\Psi}\rangle\langle\tilde{\Psi}|q_k, \quad P + Q = I - P_\Psi$$

and $\{\tilde{\Phi}q(_k^\dagger);\ 1\leqslant k\leqslant\mu\}$ is a complete orthonormal basis (conb) for $\tilde{\Phi}(\mathbf{M})$.

If Ψ is a vacuum for the manifold $\mathbf{M}$,[15] there must exist a basis $\{q_k^\dagger, q_k;\ 1\leqslant k\leqslant\mu'\}$ of $\mathbf{M}$ such that for $\mu\leqslant\mu'$

(i) $\{\tilde{\Psi}, q_k^\dagger\tilde{\Psi};\ 1\leqslant k\leqslant\mu\}$ is an orthonormal set of vectors and

(ii) $q_k\tilde{\Psi} = 0, \quad 1\leqslant k\leqslant\mu',$

$\quad\ \ q_k^\dagger\tilde{\Psi} = 0, \quad \mu+1\leqslant k\leqslant\mu'.$

Hence to obtain an approximate ground state $\tilde{\Psi}$ with the desired properties, we must solve the following problem:

$$\langle\tilde{\Psi}|H\tilde{\Psi}\rangle = \min_{\Psi\in\mathcal{C}}\langle\Psi|H\Psi\rangle, \tag{2.18}$$

where

$$\mathcal{C} = \mathrm{Vac}(\mathbf{M})\cap\mathrm{Sop}(\mathbf{M})$$

and

Vac($\mathbf{M}$) denotes the set of states that are vacuums for $\mathbf{M}$, while Sop($\mathbf{M}$), the set that satisfies (2.14) [actually one can replace Vac($\mathbf{M}$) by any set $W\subseteq$ Vac($\mathbf{M}$)]. If such a state is found one can then construct, according to the preceding discussion, a model Hamiltonian K, a model Liouville operator $\hat{K}$ defined as $[K,\ ]$, and a model superoperator $K_{\tilde{\Psi}}$ from which approximate resolvents and propagators can be built that decouple the spaces $\Phi(\mathbf{M})$ and $\Phi(\mathbf{M})^\perp$. However, it must be pointed out that the condition expressed in Eq. (2.14) cannot always be fulfilled for a given Hamiltonian H and manifold $\mathbf{M}$. A weakened form of this condition

$$|\langle\tilde{\Psi}|[H,A^+B]\,\tilde{\Psi}\rangle|^2\leqslant\epsilon \quad \forall\ A,B\in\mathbf{M} \tag{2.19}$$

for some small real positive number ϵ, leads to the set $\mathrm{Sop}_\epsilon(\mathbf{M})$ such that

$$\mathrm{Sop}_\epsilon(\mathbf{M})\subseteq\mathrm{Sop}(\mathbf{M}). \tag{2.20}$$

The model superoperator $K_{\tilde{\Psi}}$ based on $\tilde{\Psi}$ satisfying (2.19) has the property

$$\begin{aligned}K_{\tilde{\Psi}}\tilde{\Phi}(A) &= P_{\tilde{\Psi}}(\mathbf{M})\,\tilde{\Phi}(\hat{K}(A)) \quad \forall\ A\in\mathbf{M}, \\ K_{\tilde{\Psi}}\tilde{\Phi}(A) &= P_{\tilde{\Psi}}(\mathbf{M})^\perp\tilde{\Phi}(\hat{K}(A)) \quad \forall A\in\mathbf{M}^c,\end{aligned} \tag{2.21}$$

with $\hat{K}$ being replaced in (2.21) by $\hat{H}$ only when $\tilde{\Psi}\in$ Sop($\mathbf{M}$). Using the weakened condition of (2.14) is not a serious drawback for we have that as $\mathbf{M}\rightarrow\mathcal{B}(\mathcal{H})$, we still go towards the exact case, i.e., $\tilde{\Psi}\rightarrow\Psi$.

The term self-consistent approximation to a given propagator is thus to be interpreted as an approximate propagator such that:

(1) it is built on a model Hamiltonian whose eigenvalues (or eigenvalue differences) correspond to the poles of the propagator,

(2) the approximation represents a decoupling of the defining operator manifold of the propagator from all other operators,

(3) the state on which the propagator is defined is the ground state of the model Hamiltonian, and
(4) this state is also a vacuum for the defining manifold.

III. VAN VLECK PERTURBATION THEORY

There are two routes open to improve a self-consistent approximation to a given propagator. One can either
(a) go to a higher order propagator, i.e., enlarge the manifold **M**, and determine a self-consistent approximation and then use partitioning to obtain an improved approximation to the lower order propagator, or
(b) directly improve the lower order one by some form of perturbation theory.

We shall concentrate here on the second alternative and base a perturbation theory on the Van Vleck aproach,[5,7] which we modify to produce a self-adjoint operator at every level of approximation. Further, the form of the zeroth-order unperturbed Hamiltonians in this approach corresponds naturally to the form of the effective operators K produced by the self-consistent approximation procedure described in Sec. II.

Following Ref. 7, one considers a transformed Hamiltonian $\mathfrak{F}$ given by

$$\mathfrak{F} = U^{\dagger}HU, \tag{3.1}$$

an unperturbed reference state $\tilde{\Psi}$,

$$\tilde{\Psi} = U^{\dagger}\Psi, \tag{3.2}$$

and a reference orthogonal projector P such that

$$P\tilde{\Psi} = \tilde{\Psi} \tag{3.3}$$

and

$$\mathfrak{F} = P\mathfrak{F}P + P^{\perp}\mathfrak{F}P^{\perp}. \tag{3.4}$$

It follows from (3.1) and (3.2) that

$$\mathfrak{F}\tilde{\Psi} = E\tilde{\Psi}, \tag{3.5}$$

where

$$H\Psi = E\Psi. \tag{3.6}$$

The diagonal part A_D of an operator A with respect to P is defined to be

$$A_D = PAP + P^{\perp}AP^{\perp} \tag{3.7}$$

and the off-diagonal part by

$$A_X = PAP^{\perp} + P^{\perp}AP; \tag{3.8}$$

therefore

$$\mathfrak{F} = \mathfrak{F}_D \quad \text{and} \quad \mathfrak{F}_X = 0. \tag{3.9}$$

The transformed Hamiltonian $\mathfrak{F}$ can be developed in terms of an expansion of the infinitesimal generator G of U:

$$\mathfrak{F} = e^{-\{G_0 + G_1 + \cdots\}}H\,e^{\{G_0 + G_1 + \cdots\}} \tag{3.10}$$

such that

$$G_k = -G_k^{\dagger}, \quad k = 0,1,\cdots, \tag{3.11}$$

where approximations to $\mathfrak{F}$ are defined by

$$\mathfrak{F}_k = \exp\Big(-\sum_{l=0}^{k} G_l\Big)H\exp\Big(\sum_{l=0}^{k} G_l\Big) = U_k^{\dagger}\,HU_k. \tag{3.12}$$

In general $(\mathfrak{F}_k)_X \neq 0$, $(G_k)_D \neq 0$, although $(\mathfrak{F})_X = 0$ and $(G)_D = 0$. The zeroth-order approximation $\mathfrak{F}_0$ to $\mathfrak{F}$ is obtained by setting

$$\mathfrak{F}_0 = H_0 \quad \text{and} \quad U_0 = I, \tag{3.13}$$

where

$$H = H_0 + V.$$

(But note that higher order approximation $\mathfrak{F}_k$ of $\mathfrak{F}$ are given by $\mathfrak{F}_k = U_k^{\dagger}\,HU_k$.) In our case, the unperturbed part of the Hamiltonian will always be given by

$$H_0 = K = QHQ + P^{\perp}HP^{\perp} + \langle\tilde{\Psi}\,|\,H\tilde{\Psi}\rangle\,P_{\tilde{\Psi}}, \tag{3.14}$$

so that

$$V = H_X = QHP^{\perp} + P^{\perp}HQ + P_{\tilde{\Psi}}\,H(P+Q) + (P+Q)\,HP_{\tilde{\Psi}}. \tag{3.15}$$

The equations for $\mathfrak{F}_k$ and G_k in terms of $\{G_1,\ldots,G_{k-1}\}$, V, and R_γ are given in Eqs. (33) and (34) of Ref. 7, where R_γ is a resolvent acting in the space $P^{\perp}\mathscr{H}$, viz.,

$$R_\gamma = \sum_i \frac{|i\rangle\,\langle i|}{E_\gamma - E_i}, \tag{3.16}$$

where $\{|i\rangle\}$ is a conb for $P^{\perp}\mathscr{H}$ ($\{E_\gamma\}$ and $\{E_i\}$ are the eigenvalues of PHP and $P^{\perp}HP^{\perp}$, respectively.) For further details, the reader is referred to that paper. In practice, the whole space $\mathscr{H}$ cannot be used so a subspace is selected in which to apply the preceding perturbational construction, i.e., $P^{\perp}$ is replaced by the orthogonal projector P', where

$$PP' = P'P = 0 \tag{3.17}$$

and

$$P + P' \leqslant I. \tag{3.18}$$

IV. SUPEROPERATOR PERTURBATION THEORY

Consider the representation of the Liouville operator on the space of equivalence classes of operators based on the exact ground state Ψ of H and let Ψ_0 be a reference state so that as in Sec. III $\Psi = U\Psi_0$ and $\mathfrak{F} = U^{\dagger}HU$. Then

$$\begin{aligned}(\Phi(UA)\|\,\mathscr{L}_\Psi\,\Phi(UB))&= \langle\Psi\,|\{A^{\dagger}U^{\dagger}HUB - A^{\dagger}U^{\dagger}UBH\}\,\Psi\rangle\\&= \langle\Psi\,|UU^{\dagger}\{A^{\dagger}UHUB - A^{\dagger}BUU^{\dagger}H\}UU^{\dagger}\Psi\rangle\\&= \langle\Psi_0|\{(AU)^{\dagger}\,\mathfrak{F}(UB) - (AU)^{\dagger}(BU)\,\mathfrak{F}\}\,\Psi_0\rangle\\&= (\Phi_0(AU)\|\,\mathfrak{F}_{\Psi_0}\Phi_0(BU)) \quad \forall\; A,B,\end{aligned} \tag{4.1}$$

where Φ_0 is the equivalence class mapping and $\hat{\mathfrak{F}}_{\Psi_0} = \Pi_0(\hat{\mathfrak{F}})$ the GNS representation of $\hat{\mathfrak{F}}$ induced by Ψ_0, i.e.,

$$\hat{\mathfrak{F}}_{\Psi_0}\Phi_0(A) = \Phi_0(\hat{\mathfrak{F}}(A)) = \Phi_0([\mathfrak{F}, A]) \quad \forall\; A. \tag{4.2}$$

Defining the isometric isomorphism

$$\hat{U}\colon \mathscr{H}_\Psi \to \mathscr{H}_{\Psi_0} \tag{4.3}$$

by

$$\hat{U}\Phi(A) = \Phi_0(AU) \quad \forall\; A, \tag{4.4}$$

one then has that

$$(\Phi(UA)\|\,\mathscr{L}_\Psi\Phi(UB)) = (\Phi(A)\|\,\hat{U}^{\dagger}\,\hat{\mathfrak{F}}_{\Psi_0}\hat{U}\Phi(B)) \tag{4.5}$$

and finally that

$$(\Phi(A)|\,\mathscr{L}_\Psi\Phi(B)) = (\Phi_0(U^\dagger AU)|\,\hat{\mathfrak{F}}_{\Psi_0}\Phi_0(U^\dagger BU)). \tag{4.6}$$

[Note that $\Phi_0(A)\to\Phi_0(U^\dagger AU)$ does *not* define a linear map: $\mathscr{H}_{\Psi_0}\to\mathscr{H}_{\Psi_0}$!]

We shall now specialize Ψ_0 to be such that with $H_0 = QHQ + P^\perp HP^\perp + P_0HP_0$, it produces a self-consistent propagator for the manifold $\mathbf{M}$. Thus letting $\mathfrak{F}_0 = H_0$ we have that

$$(\Phi_0(\mathbf{t})|(zI_{\Psi_0} - \hat{\mathfrak{F}}_0)^{-1}\Phi_0(\mathbf{t})) = (z\mathbf{I} - \hat{\mathfrak{F}}_0)^{-1}, \tag{4.7}$$

where the matrix $\hat{\mathfrak{F}}_0$ is given by

$$\hat{\mathfrak{F}}_0\Phi_0(\mathbf{t}) = \Phi_0(\mathbf{t})\,\hat{\mathfrak{F}}_0 \tag{4.8}$$

and $\hat{\mathfrak{F}}_0$ is the GNS representation of $H_0 - P_0HP_0$, which is identical to that of $[\,\mathfrak{F}_0,\ ]$. Equation (4.7) can be re-expressed as

$$(\Phi_0(U_0^\dagger\,\mathbf{t}U_0)|(zI_{\Psi_0} - \hat{\mathfrak{F}}_0)^{-1}\Phi_0(U_0^\dagger\,\mathbf{t}U_0)) = (z\mathbf{I} - \hat{\mathfrak{F}}_0)^{-1}, \tag{4.9}$$

where $U_0 = I$, so that it serves to define a zeroth-order approximation to $\mathbf{R}_\mathbf{M}(z) = (\Phi(\mathbf{t})|(zI_\Psi - \mathscr{L}_\Psi)^{-1}\Phi(\mathbf{t}))$. In this case, the zeroth-order self-energy

$$\Sigma_0(z) = (\Phi_0(\mathbf{t})|\,\hat{\mathfrak{F}}_0\Phi_0(\mathbf{u}))(\Phi_0(\mathbf{u})|(zI_{\Psi_0} - \hat{\mathfrak{F}}_0)\times\Phi_0(\mathbf{u}))^{-1}(\Phi_0(\mathbf{u})|\,\hat{\mathfrak{F}}_0\Phi_0(\mathbf{t})) \tag{4.10}$$

is evidently zero as

$$\hat{\mathfrak{F}}_0 = \hat{Q}\,\hat{\mathfrak{F}}_0\hat{Q} + \hat{P}^\perp\,\hat{\mathfrak{F}}_0\,\hat{P}^\perp, \tag{4.11}$$

where

$$\hat{Q} = \sum_{i=1}^{\nu} |\Phi_0(t_i))(\Phi_0(t_i))|$$

and

$$\hat{Q} + \hat{P} = \hat{I} - |\Phi_0(I))(\Phi_0(I)|.$$

(One might note that the exact operator $\hat{\mathfrak{F}}_{\Psi_0}$ also has the property expressed by (4.11), i.e., it decouples $\hat{Q}\mathscr{H}_{\Psi_0}$ from $\hat{P}^\perp\mathscr{H}_{\Psi_0}$). The k th order approximation to $\mathbf{R}_\mathbf{M}(z)$ can be provided by

$$\mathbf{R}_\mathbf{M}^{(k)}(z) = \big[(\Phi_0(U_k^\dagger\mathbf{t}U_k)|(zI_{\Psi_0} - \hat{\mathfrak{F}}_k)\,\Phi_0(U_k^\dagger\mathbf{t}U_k)) - \Sigma_k\,(z)\big]^{-1}, \tag{4.12}$$

where

$$\Sigma_k(k) = (\Phi_0(U_k^\dagger\mathbf{t}U_k)|\,\hat{\mathfrak{F}}_k\Phi_0(U_k^\dagger\mathbf{u}\;U_k)) \times(\Phi_0(U_k^\dagger\mathbf{u}\;U_k)|(zI_{\Psi_0} - \hat{\mathfrak{F}}_k)\,\Phi_0(U_k^\dagger\mathbf{u}\;U_k))^{-1} \times(\Phi_0(U_k^\dagger\mathbf{u}\;U_k)|\,\hat{\mathfrak{F}}_k\;\Phi_0(U_k^\dagger\mathbf{t}\;U_k))$$

which is now nonzero as $U_k \neq I$. The operator $\hat{\mathfrak{F}}_k$ is defined to be the GNS representation of a modified k th order effective operator $\tilde{\mathfrak{F}}_k$ given by

$$\tilde{\mathfrak{F}}_k = Q\mathfrak{F}_kQ + P^\perp\mathfrak{F}_kP^\perp, \tag{4.13}$$

so that

$$\hat{\mathfrak{F}}_k = \Pi_0(\tilde{\mathfrak{F}}_k) \tag{4.14}$$

and also of the derivation operator defined by $\tilde{\mathfrak{F}}_k'$, where

$$\tilde{\mathfrak{F}}_k' = \tilde{\mathfrak{F}}_k + P_0\,\mathfrak{F}_k\,P_0, \tag{4.15}$$

i.e.,

$$\hat{\mathfrak{F}}_k = \Pi([\,\tilde{\mathfrak{F}}_k',\ ]). \tag{4.16}$$

As both the zeroth-order effective operator $\mathfrak{F}_0$ and the final effective operator $\mathfrak{F} = U^\dagger HU$ have the properties expressed by Eqs. (4.13) and (4.15), this modification of the Van Vleck procedure is well justified. Furthermore, our choice of H_0 ensures that the approximate infinitesimal generators G_k satisfy the Kemble condition for all k [i.e., $(G_k)_X = 0$]. The corrections to $\mathbf{R}_\mathbf{M}(z)$ arise from two sources:
(a) The superoperator matrix term

$$(\Phi_0(U_k^\dagger\mathbf{t}U_k)|(zI_{\Psi_0} - \hat{\mathfrak{F}}_k)\,\Phi_0(U_k^\dagger\mathbf{t}U_k))$$

and
(b) the self-energy $\Sigma_k(z)$.

The zeroth-order approximation to $\mathbf{R}_\mathbf{M}(z)$ provides a self-consistent propagator approximation; thus the manifold $\mathbf{M}$ is generated by $\{q_k^\dagger, q_k;\ 1\leqslant k\leqslant\mu'\}$ where $\{\Psi_0, q_k^\dagger\Psi_0;\ 1\leqslant k\leqslant\mu\}$ is an orthonormal set while $\{q_k\Psi_0 = 0;\ 1\leqslant k\leqslant\mu'\}$ and $\{q_k^\dagger\Psi_0 = 0;\ \mu+1\leqslant k\leqslant\mu'\}$ so that

$$\{\mathbf{t}\} = \{\mathbf{q}^\dagger,\mathbf{q}\}, \tag{4.17}$$

and we have zero blocks in the zeroth-order approximation, which of course can be excluded. The vacuum property of Ψ_0 wrt $\mathbf{M}$ is not inherited by Ψ_k or expressed equivalently, Ψ_0 is not a vacuum for $U_k^\dagger\mathbf{M}U_k$. Hence blocks that were zero in the initial approximation to $\mathbf{R}_\mathbf{M}(z)$ become nonzero in higher order ones.

As pointed out in Sec. III, it is not feasible to involve the whole Hilbert space $\mathscr{H}$ in practical calculations, and therefore only a given subspace is utilized. In the same way, only a subspace V of $\mathscr{B}(\mathscr{H})$ is considered in the superoperator approach. This necessitates an initial approximation of $\mathbf{R}_{\mathscr{B}(\mathscr{H})}(z)$ by the inner projection technique, viz.,

$$(\Phi(\mathscr{B}(\mathscr{H}))|(zI_\Psi - \mathscr{L}_\Psi)^{-1}\Phi(\mathscr{B}(\mathscr{H})) \simeq(\Phi(V)|(zI_\Psi - \mathscr{L}_\Psi)\,\Phi(V))^{-1}, \tag{4.18}$$

where the use of $\mathscr{B}(\mathscr{H})$ and V above is to be interpreted as bases for these spaces. This leads to

$$(\Phi(\mathbf{t})|(z\mathbf{I}_\Psi - \mathscr{L}_\Psi)^{-1}\Phi(\mathbf{t})) \simeq(\Phi(\mathbf{t})|\Phi(V))(\Phi(V)|(zI_\Psi\ \mathscr{L}_\Psi)\,\Phi(V))^{-1}(\Phi(V)|\,\Phi(\mathbf{t})). \tag{4.19}$$

The rhs of (4.19) can be further developed by the use of partitioning to give

$$\mathbf{R}_\mathbf{M}(z)\simeq\{(\Phi(\mathbf{t})|(z - \mathscr{L}_\Psi)\,\Phi(\mathbf{t})) - (\Phi(\mathbf{t})|\,\mathscr{L}_\Psi\Phi(\tilde{\mathbf{u}}))(\Phi(\tilde{\mathbf{u}})|(zI_\Psi - \mathscr{L}_\Psi)\,\Phi(\tilde{\mathbf{u}}))^{-1}(\Phi(\tilde{\mathbf{u}})|\Phi(\mathbf{t}))\}^{-1}, \tag{4.20}$$

where now

$$\Phi(V) = \text{linear span}\ \{\Phi(\mathbf{t}),\Phi(\tilde{\mathbf{u}})\}. \tag{4.21}$$

The approximation for $\mathbf{R}_\mathbf{M}(z)$ expressed in (4.20) has exactly the same form as (2.12), the exact expression for $\mathbf{R}_\mathbf{M}(z)$, except that $\mathbf{u}$ is replaced by $\tilde{\mathbf{u}}$; thus the perturbative approximation scheme outlined in this section can be directly applied.

V. DISCUSSION

The superoperator perturbation theory presented in this article has the following important properties:

(i) It is based on equivalence classes of operators that form a Hilbert space. This gives rigorous meaning to inner projections and the superoperator adjoint operation. It further allows us to use the machinery of Hilbert space operator theory in a valid fashion.

(ii) Superoperators acting in the space of equivalence classes based on the exact ground state of the Hamiltonian H can all be expressed in terms of superoperators acting in the space of equivalence classes based on a given reference state. Thus at every level of approximation we have a well-defined superoperator in a Hilbert space.

(iii) The approximate Liouville operators are all Hermitian.

(iv) The zeroth order of approximation provides an approximate propagator that is decoupled in a self-consistent manner.

(v) It inherits the advantages of the Van Vleck approach to multidimensional many-body theory as discussed in Ref. 7, and extends it to generate a perturbation theory for propagators.

Other forms of superoperator perturbation theory such as found in Ref. 9, although giving encouraging numerical results, use mathematical relationships that are not well-defined, i.e., resolutions of the identity, adjoint operations based on a degenerate inner product. Further, they do not relate explicitly to self-consistent propagator approximations nor do they provide model Liouville operators and stationary states. The desirability and importance of consistent approximations has been pointed out and discussed in particular for the particle-hole propagator within the random phase approximation.[16] A coupled-cluster inspired treatment of superoperator perturbation theory[17] also suffers from the defects just mentioned and, in particular, the explicit loss of Hermiticity in the propagator matrices.

In future publications we shall examine the relationship of the approach advocated in this article with other forms of propagator perturbation theory, and apply it to the electron and particle-hole propagators.

ACKNOWLEDGMENT

Osvaldo Goscinski originally suggested the use of the Van Vleck formalism for superoperator perturbation based on consistent propagators. I am also indebted to him for many stimulating discussions on the subject. A grant from the National Science Foundation is also gratefully acknowledged.

[1]S. Nakajima, Prog. Theor. Phys. **20**, 948 (1958); R. W. Zwanzig, J. Chem. Phys. **33**, 1338 (1960).

[2]O. Penrose, Rep. Prog. Phys. **42**, 1937 (1979).

[3]P. O. Löwdin, Int. J. Quantum Chem. **2**, 867 (1968).

[4]P. O. Löwdin, Quantum Theory Preprint TF 657, 1982.

[5]J. H. Van Vleck, Phys. Rev. **33**, 467 (1929).

[6]H. Primas, Rev. Mod. Phys. **35**, 710 (1963).

[7]L. T. Redmon and R. J. Bartlett, J. Chem. Phys. **76**, 1938 (1982).

[8]O. Goscinski and B. Lukman, Chem. Phys. Lett. **7**, 573 (1970).

[9]Y. Öhrn and G. Born, Adv. Quantum Chem. **13**, 1 (1981); P. Jorgensen, Ann. Rev. Phys. Chem. **26**, 359 (1975); J. Oddershede, Adv. Quantum Chem. **11**, 275 (1978); E. S. Nielsen, P. Jorgensen, and J. Oddershede, J. Chem. Phys. **73**, 6238 (1980).

[10]H. Sphon, J. Math. Phys. **17**, 57 (1976).

[11]B. Weiner and O. Goscinski, Int. J. Quantum Chem. Symp. 1 **21**, 299 (1977).

[12]I. Gel'fand, and M. A. Naimark, Mat. Sb. **12**, 197 (1943); I. E. Segal, Ann. Math. **48**, 930 (1947).

[13]O. Goscinski and B. Weiner, Phys. Scr. **21**, 385 (1980); B. Weiner and O. Goscinski, Int. J. Quantum Chem. **18**, 1109 (1980); **21**, 369 (1982).

[14]M. Takesaki, *Theory of Operator Algebras I* (Springer-Verlag, Berlin 1979).

[15]B. Weiner, Consistent Propagator Approximations" (unpublished). If Ψ is a vacuum the structure of the approximate propagator is greatly simplified.

[16]J. Linderberg and Y. Öhrn, Int. J. Quantum Chem. **12**, 161 (1977); Y. Öhrn and J. Linderberg, Int. J. Quantum Chem. **15**, 343 (1979).

[17]M. D. Prasad, S. Pal, and D. Mukherjee, Pramana **15**, 531 (1980).

Scattering theory for the dilation group. I. Simple quantum mechanical scattering

J. V. Corbett[a)]
Faculté des Sciences de Luminy and Centre de Physique Théorique, Section 2, CNRS Marseille, France

(Received 22 July 1982; accepted for publication 29 October 1982)

A theory of scattering based on the dilation group is developed for quantum mechanical one-particle systems. A scattering operator is defined that agrees with the usual scattering operator, whenever the usual wave operators exist and are asymptotically complete.

PACS numbers: 03.65.Nk

INTRODUCTION

The aim of this article is to lay the foundations for a scattering theory based on the dilation group along the lines of the Lax–Phillips scattering theory[1] that is based on the translation group.

The reason for choosing to base a scattering theory on the dilation group is that the time evolution of states in nonrelativistic quantum mechanics does not obey Huyghens' principle and thus does not easily fit into a theory based on the translation group. In fact the solutions of the Schrödinger equation, that are not bound states, asymptotically evolve as though they were being dilated. This can be most easily seen from the free Schrödinger equation. Let $f(x,t)$ be a solution of

$$-\frac{1}{i}\frac{\partial f}{\partial t} = -\frac{\nabla^2}{2m}f(x,t).$$

It is well known, and frequently called the evanescence of the wave packet that, for large positive times $t, f(x,t)$ looks like $(im/t)^{+3/2}\hat{f}(mx/t)$, where $\hat{f}$ denotes the Fourier transform of f. Let $\mathscr{U}(a)$ be a unitary representation of $\mathbf{R}_*^+$, the multiplicative group of positive real numbers, on $\mathbf{L}^2(\mathbf{R}^3)$ given by $(\mathscr{U}(a)f)(x) = a^{-3/2}f(a^{-1}x)$. Then $f(x,t)$ looks like $i^{3/2}(\mathscr{U}(t/m)\hat{f})(x)$ for large t. We now argue that for scattering processes it is the large time behavior of the wave functions that is most important, and hence if we develop a theory based on the dilation group we should be able to capture the large time behavior of the wave functions by dilating them to large spatial separations.

An outline of our results will be given now. As a theory of scattering processes we are still a long way from the completeness of the Lax–Phillips model, especially with regard to the analyticity properties of the scattering operator. To get such results seems to require that the Mellin transform be used in place of the Fourier transform of the Lax–Phillips theory, but we have not pursued this development.

We begin by defining dilating subspace representations of the group $\mathbf{R}_*^+$. These are subspaces of the Hilbert space, the elements of which are dilated out to infinity with respect to the spatial variables. These are shown to be equivalent to certain canonical representations of $\mathbf{R}_*^+$ on the Hilbert space $\mathbf{L}^2(\mathbf{R}_*^+, d\lambda/\lambda; K)$, where K is an auxiliary Hilbert space. These dilating representations $\mathscr{U}(a)$ are then related to time displacements $V(t)$ by the formula $\mathscr{U}(a)V(t)\mathscr{U}(a)^{-1} = V(at)$, for all $t\in\mathbf{R}$, $a\in\mathbf{R}_*^+$, and a canonical representation of these systems is obtained. They are essentially direct integrals of the irreducible representations of the affine group.

The quantum mechanical scattering systems are characterized by the fact that to the given representation $V(t)$ of time evolution we must associate two representations of $\mathbf{R}_*^+$. For each of these, called $\mathscr{U}^{\pm}(a)$, we obtain a map $W^{\pm}$ to the canonical model, and the dilation scattering operator S_d is defined $S_d = (W^{-1})^{-1}W^{+}$. It follows from the structure of representation theorem that the generator of $V(t)$ must be positive and have an absolutely continuous spectrum.

In the next section we compare S_d with the usual scattering operator S. This is done by introducing a free particle and defining dilation operators $\Omega_d^{\pm}$, first via the unitary maps to the canonical representations spaces, $\Omega_d^{\pm} = (W^{\mp})^{-1}W_0$, and secondly, in the special case when both $V(t)$ and $V_0(t)$ act on the same Hilbert space $\mathscr{H}$, as strong limits

$$\Omega_d^{\pm} = \operatorname*{s-lim}_{a\to\infty}(\mathscr{U}^{\mp}(a))^{-1}\mathscr{U}_0(a).$$

These two definitions are then shown to agree under certain addition conditions. Finally, under these same conditions, we show that the usual wave operators $\Omega^{\pm}(H,H_0)$ exist if and only if the dilation wave operators exist, and give a condition under which they are equal. This condition depends upon some additional structure. In the Appendix we discuss projective representations of SL$(2,R)$ and show how such representations exist on the space of scattering states when the usual wave operators $\Omega^{\pm}(H,H_0)$ exist and are asymptotically complete. In fact we demonstrate how the presence of appropriate projective representations of SL$(2,R)$ forces the time evolution $V(t)$ to have a dilation type behavior for large positive and negative times. Finally, in this section we show that if the dilation wave operators exist and if the dilating subspaces for $\mathscr{U}_0(a)$, $\mathscr{U}^{+}(a)$, and $\mathscr{U}^{-}(a)$ can be taken to be the same subsets of $\mathscr{H}$ then the $\Omega^{\pm}(H,H_0)$ are equal to the $\Omega_d^{\pm}$.

In the third section we briefly discuss the example of potential scattering. We outline a method of constructing the $\mathscr{U}^{\pm}(a)$, $a\in\mathbf{R}_*^+$, that should work for systems whose time delay may be infinite but increases more slowly than linearly. The idea is to construct the generators $A^{\pm}$ of the representations $\mathscr{U}^{\pm}(a)$ of $\mathbf{R}_*^+$.

[a)] Permanent and postal address: School of Mathematics and Physics, Macquarie University, North Ryde, New South Wales, Australia.

 0022-2488/83/071797-09$02.50

1. SCATTERING THEORY FOR THE DILATION GROUP

Let $\mathscr{U}(a)$, $a\in(0,\infty)$, be a continuous unitary representation of the multiplicative group $\mathbf{R}_*^+$ on the separable Hilbert space $\mathscr{H}$.

Definition 1: The continuous unitary representation $\mathscr{U}(a)$ of $\mathbf{R}_*^+$ is said to be dilating on $\mathscr{H}$ if there exists a closed subset D of $\mathscr{H}$ such that

(1) $\mathscr{U}(a)D\subset D$ for all a, $1\leqslant a<\infty$,

(2) $\bigcap_{a>1} \mathscr{U}(a)D=\{0\}$,

(3) $\bigcup_a \mathscr{U}(a)D$

is dense in $\mathscr{H}$, where $\cap$, $\cup$ are the set theoretical intersection and union. D is called the dilating subspace.

Example: Let $\mathscr{H}=\mathbf{L}^2(\mathbf{R}_*^+, d\lambda/\lambda; \mathscr{H}')$ be the space of functions on $(0,\infty)$ with values in the auxiliary Hilbert space $\mathscr{H}'$ that are square integrable with respect to the Haar measure $d\lambda/\lambda$ of R_*^+ on R_*^+. Henceforth we will denote this space by $\mathbf{L}^2(\mathbf{R}_*^+;\mathscr{H}')$. Let $\mathscr{U}_l(a)$, $a\in(0,\infty)$, be the left regular representation of $\mathbf{R}_*^+$, $(\mathscr{U}_l(a)f)(\lambda)=f(a^{-1}\lambda)$. Then $\mathscr{U}_l(a)$ is dilating on $\mathscr{H}$ with D given by

$$D=\mathbf{L}^2([1,\infty);\mathscr{H}').$$

There is another realization of the dilating subspace that we shall need. Let $\mathscr{U}_r(a)$, $a\in(0,\infty)$, be the representation of $\mathbf{R}_*^+$, $(U_r(a)f)(\lambda)=f(\lambda a)$. Then $\mathscr{U}_r(a)$ is dilating on $\mathscr{H}$ with D given by

$$D=\mathbf{L}^2((0,1];\mathscr{H}').$$

These examples are not really different, as the two representations are equivalent under the unitary map from $\mathbf{L}^2(\mathbf{R}_*^+;\mathscr{H}')$ to $\mathbf{L}^2(\mathbf{R}_*^+;\mathscr{H}')$ that corresponds to the change of variable $\lambda\to\lambda^{-1}$.

The reason for calling this representation dilating as opposed to contracting comes from the usual representation of dilations for the free Schrödinger particle. Take $\mathscr{H}_0=\mathbf{L}^2(\mathbf{R}^3)$, let $V(\theta)$ be the one-parameter group of dilations, whose generator is $A_0=(x\cdot p+p\cdot x)/2$ and whose action on $f(x)\in\mathbf{L}(\mathbf{R}^3)$ is $(V(\theta)f)(x)=(e^{-iA_0\theta}f)(x)$ $=e^{-3\theta/2}f(e^{-\theta}x)$. $\mathscr{H}_0$ carries, of course, the configuration space representation of the canonical commutation relations. $\mathscr{H}_0$ is unitarily equivalent to the Hilbert space $\mathbf{L}^2(\mathbf{R}_*^+, d\rho/\rho;K)$, where $\rho=x^2$, the square of the position coordinate, and $K=\mathbf{L}^2(S^2,d\Omega)$, S^2 being the unit sphere in $\mathbf{R}^3$ and $d\Omega$ its standard surface measure. The map from $\mathbf{L}^2(\mathbf{R}^3)$ to $\mathbf{L}^2(\mathbf{R}_*^+, d\rho/\rho;K)$ is given by

$$\tilde{f}(\rho;\omega)=(\mathscr{M}f)(\rho;\omega)=(\sqrt{2})^{-1}\rho^{3/4}f(\rho^{1/2}\omega),$$

where $\rho^{1/2}\omega=x\in\mathbf{R}^3$. Under this transformation $V(\theta)$ goes to $\tilde{V}(\theta)=\mathscr{M}V(\theta)\mathscr{M}^{-1}$, where $(\tilde{V}(\theta)\tilde{f})(\rho;\omega)=f(e^{-2\theta}\rho;\omega)$. When we put $s=e^{2\theta}$, $-\infty<\theta<\infty$, we get a unitary representation of $\mathbf{R}_*^+$, denoted by $\tilde{\mathscr{U}}(s)$, such that

$$(\tilde{\mathscr{U}}(s)\tilde{f})(\rho;\omega)=\tilde{f}(s^{-1}\rho;\omega).$$

This representation is dilating because if $\tilde{f}\in\mathbf{L}^2(\mathbf{R}_*^+;K)$ has support that does not contain the origin then support of $\tilde{\mathscr{U}}(s)\tilde{f}$ gets further from the origin as s increases. Dilating thus means spatially dilating. This same spatially dilating representation can be realized in momentum space. Again map $\mathbf{L}^2(\mathbf{R}^3)$ onto $\mathbf{L}^2(\mathbf{R}_*^+;K)$, where $k=\mathbf{L}^2(s^2,d\Omega)$, as above but the unitary map W_0 is given by

$$(W_0f)(\lambda;\omega)=(\sqrt{2})^{-1}\lambda^{3/4}\hat{f}(\lambda^{1/2}\omega),$$

where $\hat{f}(\lambda^{1/2}\omega)$ is the Fourier transform of $f\in\mathbf{L}^2(\mathbf{R}^3)$ evaluated at the momentum $p=\lambda^{1/2}\omega$. Under W_0, $V(\theta)$ goes to $\hat{V}(\theta)$, where $(\hat{V}(\theta)W_0f)(\lambda;\omega)=(W_0f)(e^{2\theta}\lambda;\omega)$. If we set $s=e^{2\theta}$, we again get a representation of the multiplicative group $\mathbf{R}_*^+$,

$$(\hat{\mathscr{U}}(s)g)(\lambda;\omega)=g(s\lambda;\omega).$$

This is the free energy realization of the spatially dilating representation of $\mathbf{R}_*^+$.

It should be remarked here that there is considerable freedom in the choice of the dilating subspace D. For example, for the free particle, with $\mathscr{U}(s)=\exp(-iA_0/2\ln s)$, $s\in(0,\infty)$, we may take D to be the set of elements in $\mathbf{L}^2(\mathbf{R}^3)$ whose support lies outside the ball $|x|=R<\infty$, where R may be chosen as large as we please, or we may take D to be the elements of $\mathbf{L}^2(\mathbf{R}^3)$, whose Fourier transforms have support inside the ball of radius R_0, no matter how small $R_0>0$ is.

This example is canonical in the sense of the following representation theorem.

Theorem 1: Let $\mathscr{U}(s)$ be a dilating continuous representation of $\mathbf{R}_*^+$ on the separable Hilbert space $\mathscr{H}$ and let D be a dilating subspace for $\mathscr{U}(s)$. Then there exists a Hilbert space K and a unitary map W from $\mathscr{H}$ onto $\mathbf{L}^2(\mathbf{R}_*^+, dx/x;K)$ such that

(i) $W[D]=\mathbf{L}^2([1,\infty), dx/x;K)$, and

(ii) $W\mathscr{U}(s)W^{-1}$ is the representation of $\mathbf{R}_*^+$ on $\mathbf{L}^2(\mathbf{R}_*^+,K)$ given by $(W\mathscr{U}(s)W^{-1}f)(x;\omega)=f(s^{-1}x;\omega)$. The representation is unique up to an isomorphism of K.

Proof: For all $s\in(0,\infty)$ define $D(s)=\mathscr{U}(s)[D]$. The conditions (2) and (3) of Definition 1 imply that

$$D(\infty)=\lim_{s\to 0} D(s)=\{0\}$$

and

$$D(0)=\lim_{s\to\infty} D(s)=\mathscr{H}.$$

For any real numbers a, b with $0<a\leqslant b<\infty$, define a closed subspace $\mathscr{M}(a,b)$,

$$\mathscr{M}(a,b)=D(a)\ominus D(b).$$

That is, $\mathscr{M}(a,b)$ is the orthogonal complement of $D(b)$ in $D(a)$. Let $P(a,b)$ be the orthogonal projection onto $\mathscr{M}(a,b)$. Then an easy calculation shows that $\mathscr{U}(s)P(a,b)\mathscr{U}(s)^{-1}=P(as,bs)$. On the other hand, the family $\{P(a,b)\}$, where (a,b) runs over all open intervals in $(0,\infty)$, generates a projection-valued measure $\{P(\Omega)\}$ on the Borel subsets Ω of $(0,\infty)$. Furthermore, for any $s\in(0,\infty)$ and any Borel subset Ω of $(0,\infty)$

$$\mathscr{U}(s)P(\Omega)\mathscr{U}(s)^{-1}=P(s\Omega). \quad (1)$$

Equation (1) defines a system of imprimitivity for the multiplicative group $\mathbf{R}_*^+$ on itself. By a version of Mackey's imprimitivity theorem (Ref. 2), Theorem 9.17) such a system is unitarily equivalent to a direct sum of irreducible systems of imprimitivity. The unique irreducible system of imprimitivity for $\mathbf{R}_*^+$ acting on itself by left translation is $\mathscr{H} = \mathbf{L}^2(\mathbf{R}_*^+, dx/x)$, with projection-valued measured χ_E, the characteristic function of the Borel subset E and representation given by the left regular representation $(\mathscr{U}_e(s)f)(x) = f(xs^{-1})$. Therefore, there is a separable Hilbert space K such that the system of imprimitivity $\{P(\Omega), \mathscr{U}(s)\}$ on $\mathbf{L}^2(\mathbf{R}_*^+, dx/x;K)$ given by $(\chi(E), \mathscr{U}_e(s))$, where $\chi(E)$ is the characteristic function on the Borel subset E and for any $F(x) \in \mathbf{L}^2(\mathbf{R}_*^+, dx/x;K)$, $(V(s)F)(x) = F(xs^{-1})$. To see this, let $F(x) = \Sigma_{i=1}^{\infty} f^i(x)e_i$, where e_i is an orthonormal basis in K and $(f^1(x), f^2(x),...)$ is a sequence of functions, each in $\mathbf{L}^2(\mathbf{R}_*^+, dx/x)$, in the direct sum of irreducible systems of imprimitivity that is given by the theorem.

Let W be the unitary map that realizes this equivalence; then by construction of the $\{P(\Omega)\}$ on $\mathscr{H}$, $D = P([1,\infty))\mathscr{H}$ and hence $W[D] = \mathbf{L}^2([1,\infty), dx/x;K)$. ■

The equivalence of the examples that follow Definition 1 implies the following corollary.

Corollary 1: Let $\mathscr{U}(s)$ be a dilating continuous unitary representation of $\mathbf{R}_*^+$ on the separable Hilbert space $\mathscr{H}$ and let D be a dilating subspace for $\mathscr{U}(s)$. Then there exists a Hilbert space K and a unitary map W from $\mathscr{H}$ on $\mathbf{L}^2(\mathbf{R}_*^+, d\lambda/\lambda;K)$ such that

(i) $W[D] = \mathbf{L}^2((0,1], d\lambda/\lambda;K)$

and

(ii) $(W\mathscr{U}(s)W^{-1}f)(\lambda;\omega) = f(s\lambda;\omega)$ for all $f \in \mathbf{L}^2(\mathbf{R}_*^+, d\lambda/\lambda;K)$.

We underline the fact that for different choices of dilating subspace D we will have different unitary maps W from $\mathscr{H}$ to $\mathbf{L}^2(\mathbf{R}_*^+, \lambda/\lambda;K)$.

We must now make a connection between these dilations and the time displacements of the system that singles out those dilations that are associated with scattering states. If we return to the example of the free Schrödinger particle then it is clear what the connection should be. If the free Hamiltonian H_0 is taken to be $-\Delta/2m$ then the dilations $V(\theta) = e^{-iA_0\theta}$ and time displacements $V(t) = e^{-iH_0t}$ satisfy $\mathscr{U}(s)V(t)\mathscr{U}(s)^{-1} = V(st)$, where we have taken $s = e^{2\theta}$ and written $\mathscr{U}(s)$ for $V(\frac{1}{2}\ln s)$; that is, $\mathscr{U}(s)$ is the unitary representation of $\mathbf{R}_*^+$ on $\mathbf{L}^2(\mathbf{R}^3)$.

Definition 2: Let $V(t)$, $t \in \mathbf{R}$ be the one-parameter unitary group of time displacements on the Hilbert space $\mathscr{H}$, and let $\mathscr{U}(s)$, $s \in \mathbf{R}_*^+$, be a dilating representation of $\mathbf{R}_*^+$ on $\mathscr{H}$, then we say that $\mathscr{U}(s)$ is a scattering representation of $\mathbf{R}_*^+$ for the time displacements $V(t)$ if, for all $s \in \mathbf{R}_*^+$ and all $t \in \mathbf{R}$,

$$\mathscr{U}(s)V(t)\mathscr{U}(s)^{-1} = V(st). \tag{2}$$

We immediately obtain a representation theorem for scattering representations of $\mathbf{R}_*^+$. The proof of this theorem follows the line of argument given in Reed and Simon (Ref. 3, Theorem XI.84) in their proof of von Neumann's theorem.

Theorem 2: Let $\mathscr{U}(s)$ be a continuous unitary representaiton of $\mathbf{R}_*^+$ and $V(t)$, $t \in \mathbf{R}$, a continuous unitary representation of $\mathbf{R}$ on the separable Hilbert space $\mathscr{H}$, such that $\mathscr{U}(s)$ is dilating and for $s \in \mathbf{R}_*^+$ and all $t \in \mathbf{R}$,

$$\mathscr{U}(s)V(t)\mathscr{U}(s)^{-1} = V(st).$$

That is, $\mathscr{U}(s)$ is a scattering representation of $\mathbf{R}_*^+$. Then there exists a Hilbert space K, and a unitary map W from $\mathscr{H}$ to $\mathbf{L}^2(\mathbf{R}_*^+, d\lambda/\lambda;K)$ such that $(W\mathscr{U}(s)W^{-1}f)(\lambda) = f(s\lambda)$ for all $f \in \mathbf{L}^2(\mathbf{R}_*^+;K)$ and $WV(t)W^{-1}$ is multiplication by $e^{-i\lambda t}$.

Proof: Let H_1 and A be the self-adjoint generators of $V(t)$ and $\mathscr{U}(s)$, $V(t) = e^{-iHt}$, $\mathscr{U}(s) = e^{-iA\ln s}$. Let $\mathscr{D}$ denote the Garding domain of $V(t)\mathscr{U}(s)$. $\mathscr{D}$ is the set of vectors in $\mathscr{H}$ of the form

$$\phi_F = \int_{-\infty}^{\infty}\int_0^{\infty} F(t,s)V(t)\mathscr{U}(s)\phi dt \frac{ds}{s},$$

where $\phi \in \mathscr{H}$ and $F(t,s) \in C_0^{\infty}(\mathbf{R}\times\mathbf{R}_*^+)$. It follows from the standard arguments that $\mathscr{D}$ has the following properties:

$\mathscr{D}$ is dense in $\mathscr{H}$, $\mathscr{D} \subset D(H_1)$, $\mathscr{D} \subset D(A)$, $V(t)\mathscr{D} \subset \mathscr{D}$, and $\mathscr{U}(s)\mathscr{D} \subset \mathscr{D}$. Let $\psi \in \mathscr{D}$, differentiating the equation

$$\mathscr{U}(s)V(t)\mathscr{U}(s)^{-1}\psi = V(st)\psi$$

with respect to t, and setting $t = 0$, we get

$$\mathscr{U}(s)H_1\mathscr{U}(s)^{-1}\psi = sH_1\psi. \tag{3}$$

Since $\mathscr{D}$ is a core for both H_1 and sH_1, Eq. (3) can be extended to hold for all $\psi \in D(H_1)$ and hence H_1 and sH_1 are unitarily equivalent. Let $\{E_1(\Omega)\}$ be the spectral family for H_1, then $\{\mathscr{U}(s)E_1(\Omega)\mathscr{U}(s)^{-1}\}$ is the spectral family for $sH_1 = \mathscr{U}(s)H_1\mathscr{U}(s)^{-1}$. Therefore, for any real λ,

$$\mathscr{U}(s)E_1(-\infty,\lambda)\mathscr{U}(s)^{-1} = E_1(-\infty,s^{-1}\lambda)$$

for all $s \in \mathbf{R}_*^+$. Let $D = \operatorname{Ran} E_1(-\infty,1]$. D is a dilating subspace for $\mathscr{U}(s)$ on $\mathscr{H}$, because $\mathscr{U}(s)D = \operatorname{Ran} E_1(-\infty,s^{-1}]$ for all $s \in \mathbf{R}_*^+$ and therefore by the properties of spectral projections we have that

(i) $\mathscr{U}(s)D \subset D, \quad s \geqslant 1$,

(ii) $\cap\mathscr{U}(s)D = \{0\}$,

(iii) $\overline{\cup\mathscr{U}(s)D} = \mathscr{H}$.

Now by Corollary 1, there exists an auxiliary Hilbert space K and a unitary map W of $\mathscr{H}$ onto $\mathbf{L}^2(\mathbf{R}_*^+, d\lambda/\lambda;K)$ such that $WD = \mathbf{L}^2((0,1], d\lambda/\lambda;K)$ and $W\mathscr{U}(s)W^{-1}$ is dilation by s. Furthermore, as $WE_1(-\infty,1]W^{-1} = \chi_{(0,1]}$ we have that $WE_1(-\infty,\lambda)W^{-1} = \chi_{(0,\lambda)}$ for all $\lambda \in \mathbf{R}$, and hence WH_1W^{-1} is multiplication by λ and $WV(t)W^{-1} = e^{-i\lambda t}$. ■

It may be recognized that this theorem is just a disguised version of the theorem, due to Gel'fand and Naimark[4] and Aslaksen and Klauder,[5] on the irreducible representations of the affine group. The second irreducible representation can be obtained by taking $V(t) = e^{-iHt}$.

The following consequences of this theorem are immediate.

Corollary 2: If $V(t) = e^{-iH_1t}$ and $\mathscr{U}(s)$, $s \in \mathbf{R}_*^+$, satisfies the conditions of Theorem 2, then

(1) H_1 is positive,

(2) the point spectrum of H_1 is $\{0\}$ or empty,

(3) Let $\sigma(H_1)$ be the spectrum of H_1, $\sigma(H_1) = [0,\infty)$, then $\sigma(H_1)\setminus\{0\}$ is absolutely continuous.

With regard to (3), it is inherent in this approach that we cannot tell whether 0 is an eigenvalue of H_1 or not. This point, that corresponds to the threshold of the absolutely continuous spectrum, must be considered separately in each case.

Corollary 3: If we have a continuous unitary representation $\mathscr{U}(s)$ of $\mathbf{R}_*^+$ and a continuous unitary representation $V(t)$ of $\mathbf{R}$ that satisfy

$$\mathscr{U}(s)V(t)\mathscr{U}(s)^{-1} = V(st)$$

for all $s\in\mathbf{R}_*^+$ and all $t\in\mathbf{R}$, then $\mathscr{U}(s)$ is a dilating representation of $\mathbf{R}_*^+$.

Proof: Just take $D = E_{H_1}(-\infty,1)\mathscr{H}$ as the dilating subspace for $\mathscr{U}(s)$, where $\{E_{H_1}(\lambda)\}$ is the spectral family for the self-adjoint generator H_1 of $V(t)$. This choice of dilating subspace for $\mathscr{U}(s)$ will be often used in the following.

We are now in a position to give a definition of a scattering system.

Definition 3: Let $\mathscr{H}$ be a separable Hilbert space and $V(t) = e^{-iH_1 t}$ a continuous unitary representation of $\mathbf{R}$ that describes the time evolution of the system described by $\mathscr{H}$. Then the pair $(\mathscr{H}, V(t))$ describes a scattering process for the system if there exists a pair of scattering representations $\mathscr{U}^{\pm}(s)$ of $\mathbf{R}_*^+$ on $\mathscr{H}$.

The triplet $(\mathscr{H}, V(t), \mathscr{U}^+(s))$ is called the outgoing dilation representation of $\mathbf{R}_*^+$ and the triplet $(\mathscr{H}, V(t), \mathscr{U}^{-1}(s))$ is called the incoming dilation representation of $\mathbf{R}_*^+$.

If $(\mathscr{H}, V(t))$ has both incoming and outgoing representations of $\mathbf{R}_*^+$ then by Theorem 2 there exist unitary operators $W^{\pm}$ from $\mathscr{H}$ to $\mathbf{L}^2(\mathbf{R}_*^+, d\lambda/\lambda; K^{\pm})$. The auxiliary Hilbert spaces $K^{\pm}$ can be chosen to be the same, because in both cases we have a spectral representation of the generator H_1 of $V(t)$.

Definition 4: If the pair $(\mathscr{H}, V(t))$ describes a scattering process then the scattering operators $\widetilde{S}_d$ and S_d are defined as follows. Let $f\in\mathscr{H}$, and put $f_- = W^- f$ and $f_+ = W^+ f$, then $\widetilde{S}_d$ is the map $\widetilde{S}_d : \mathbf{L}^2(\mathbf{R}_*^+; K) \to \mathbf{L}^2(\mathbf{R}_*^+; K)$,

$$\widetilde{S}_d f_- = f_+. \tag{4}$$

The scattering operator S_d is the map from $\mathscr{H}$ to itself that is obtained by pulling $\widetilde{S}_d$ back to $\mathscr{H}$.

The following properties of the scattering operators follow immediately from the definitions.

(1) $\widetilde{S}_d$ is a unitary map from $\mathbf{L}^2(\mathbf{R}_*^+; d\lambda/\lambda; K)$ to itself.

(2) $\widetilde{S}_d = W^+(W^-)^{-1}$. (5)

(3) $\widetilde{S}_d$ commutes with $e^{-i\lambda t}$ for all $t\in\mathbf{R}$ and hence is a decomposable operator, so that for all $f, g\in\mathbf{L}(\mathbf{R}_*^+; k)$

$$(f, \widetilde{S}_d g) = \int_0^\infty (f(\lambda), \widetilde{S}_d(\lambda) g(\lambda))_K \frac{d\lambda}{\lambda},$$

where each fiber $\widetilde{S}_d$ is unitary on K.

(4) S_d is a unitary map from $\mathscr{H}$ to itself.

(5) $S_d = (W^-)^{-1} W^+$. (6)

(6) S_d commutes with $e^{-iH_1 t}$ for all $\in\mathbf{R}$.

The dilation scattering operator S_d, or $\widetilde{S}_d$, has been defined without explicit reference to the free dynamics, even though condition (3) and the structure of the dilating representations have been taken from the properties of the free dynamics.

Therefore, S_d may exist for those interactions, such as the long-range potentials, when the usual free dynamics does not describe the asymptotic motion of the system.

2. DILATION WAVE OPERATORS

We have not shown that the scattering operator S_d gives the correct, physically observable, properties of a scattering system. One way to do this would be to show that the operators $W^{\pm}$ are integral operators whose kernels are generalized eigenfunctions of H_1. But we do not yet have a way to choose the appropriate generalized eigenfunctions. That is, we do not have a form of Lippmann–Schwinger equations for the eigenfunctions. To get such equations we need to define dilation wave operators.

In order to construct dilation wave operators we need the following free particle structure. Let $H_0 = \mathbf{L}^2(\mathbf{R}^3)$, $V_0(t) = e^{-iH_0 t}$ and $\mathscr{U}_0(s) = e^{-i(1/2)A_0 \ln s}$. The unitary map W_0 from $\mathscr{H}_0$ to $\mathbf{L}^2(\mathbf{R}_*^+, d\lambda/\lambda; K)$ takes $V_0(t)$ to $e^{-i\lambda t}$, and $\mathscr{U}_0(a)$ to $L(a)$, where $(L(a)f)(\lambda;\omega) = f(a\lambda;\omega)$.

With this structure it is clear that the dilation wave operators should be given by the expression

$$\Omega_d^{\pm} = (W^{\mp})^{-1} W_0. \tag{7}$$

These operators are unitary maps from $\mathscr{H}_0$ to $\mathscr{H}$ provided that the spectral multiplicities of the Hamiltonians H_0 and H_1 are equal. Furthermore, they satisfy the intertwining relations

$$\Omega_d^{\pm} V_0(t) = V(t)\Omega_d^{\pm} \tag{8}$$

and

$$\Omega_d^{\pm} \mathscr{U}_0(s) = \mathscr{U}^{\pm}(s)\Omega_d^{\pm}, \tag{9}$$

respectively, for all real t and for all real positive s.

If, in analogy with the definition of the usual scattering operator we define the dilation scattering operator to be $\Omega_d^+(\Omega_d^-)^{-1}$ then we observe that

$$\Omega_d^+(\Omega_d^-)^{-1} = (W^-)^{-1} W^+ = S_d, \tag{10}$$

where S_d is defined by Eq. (6).

The interchange of the signs in Eq. (7) has been made to get agreement with the usual definition of the wave operators $\Omega^{\pm}(H,H_1)$. Recall that if H and H_0 are the interacting and free Hamiltonians as the Hilbert space $\mathscr{H}_0$, then

$$\Omega^{\pm}(H,H_0) = \operatorname*{s-lim}_{t\to\mp\infty} e^{iHt} e^{-iH_0 t}.$$

The dilation wave operators $\Omega_d^{\pm}$ can be defined as strong limits of families of unitary operators once some connection between the Hilbert spaces $\mathscr{H}_0$ and $\mathscr{H}$ has been made. We will consider the simplest case in which $\mathscr{H}_0 = \mathscr{H}$. Assume that $(\mathscr{H}_0, V(t))$ is a scattering process in the sense of Definition 3 and that the free particle time evolution is given by $V_0(t)$ on $\mathscr{H}_0$.

Definition 5: The dilation wave operators $\Omega_d^{\pm}$ exist if the strong limits

$$\Omega_d^{\pm} = \operatorname*{s-lim}_{s\to\infty} (\mathscr{U}^{\mp}(s))^{-1} \mathscr{U}_0(s) \tag{11}$$

exist on $\mathscr{H}_0$.

It is an immediate consequence of this definition that, for all $s \in (0, \infty)$,

$$\mathscr{U}^{\mp}(s)\Omega_d^{\pm} = \Omega_d^{\pm}\mathscr{U}_0(s). \tag{12}$$

We must now check if this definition of $\Omega_d^{\pm}$ agrees with the formula of Eq. (7). We will show that this is so in the case when the dilating subspaces, D_0 for $\mathscr{U}_0(s)$, D_+ for $\mathscr{U}^+(s)$, and D_- for $\mathscr{U}^-(s)$ are equal.

Theorem 3: If the dilating subspaces D_0, D_+, and D_- are equal, then

$$\Omega_d^{\pm} = (W^{\mp})^{-1}W_0. \tag{13}$$

Proof: Let $f \in D_+$, the dilating subspace for the representation $\mathscr{U}^+(s)$, and define the representative f_+ of f on $\mathbf{L}^2(\mathbf{R}_*^+;K)$ as f_0, the representative of f as an element of D_0, the dilating subspace for the representation $\mathscr{U}_0(s)$. Then for any $s \in \mathbf{R}_*^+$ and any $f \in D$ we define the representative of $\mathscr{U}^+(s)f$ in $\mathbf{L}^2(\mathbf{R}_*^+;k)$ as the dilate of f_0 by s. In this way we have defined, for all $f \in D_0$,

$$f_+ = W_+f = W_0f = f_0$$

and

$$W_+\mathscr{U}^+(s)f = W_0\mathscr{U}_0(s)f = L(s)f$$

for all $s \in (0, \infty)$.
By the properties of the dilating subspace the dilates of D_+ are dense in $\mathscr{H}_0 = \mathscr{H}$ and hence we can define the representatives f_+ in $\mathbf{L}^2(\mathbf{R}_*^+;K)$ of all f in $\mathscr{H}_0$.

We first show that for all $f \in D_0$, $\Omega_d^- f = (W^+)^{-1}W_0f$. If $f \in D_0$, then $\mathscr{U}_0(s)f \in D_0$ for all $s \geqslant 1$. Thus we get in $\mathbf{L}^2(\mathbf{R}_*^+;K)$ the equality, for $s \geqslant 1$,

$$f_0 = W_+(\mathscr{U}^+(s))^{-1}\mathscr{U}_0(s)f = W_0f.$$

Pulling this back to $\mathscr{H}_0$ we have

$$(\mathscr{U}^+(s))^{-1}\mathscr{U}_0(s)f = (W_+)^{-1}W_0f$$

and thus $f = \Omega_d^- f = (W^+)^{-1}W_0f$, for all $f \in D_0$. Now let $g \in \mathscr{H}_0$ be such that $\mathscr{U}_0(s)g \in D_0$ for some $s \in (0, \infty)$. The set of such elements is dense in $\mathscr{H}_0$ by property 3 of Definition 1. It follows that $\mathscr{U}^+(s)\Omega_d^- g \in D_0$ because

$$\mathscr{U}^+(s)\Omega_d^- g = \Omega_d^- \mathscr{U}_0(s)g = \mathscr{U}_0(s)g.$$

Therefore,

$$W^+\mathscr{U}^+(s)\Omega_d^- g = W_0\mathscr{U}_0(s)g.$$

On the other hand, $W^+\mathscr{U}^+(s) = L(s)W^+$ and $W_0\mathscr{U}_0(s) = L(s)W_0$, where $L(s)$ acts on $\mathbf{L}^2(\mathbf{R}_*^+, d\lambda/\lambda;K)$ by dilation by s. Thus in $\mathbf{L}^2(\mathbf{R}_*^+, d\lambda/\lambda;K)$ we have

$$L(s)W_+\Omega_d^- g = L(s)W_0g$$

and, using the fact that $L(s)$ is an isometry,

$$W_+\Omega_d^- g = W_0g$$

or, on $\mathscr{H}_0$, $\Omega_d^- g = (W^+)^{-1}W_0g$.

This result extends to all of $\mathscr{H}_0$ by continuity and so $\Omega_d^- = (W_+)^{-1}W_0$.

An analogous argument shows that

$$\Omega_d^+ = (W^-)^{-1}W_0.$$

■

We have shown, in the special case when $\mathscr{H} = \mathscr{H}_0$ and the dilating subspaces D_0, D_+, and D_- can be chosen to be equal, that the dilation scattering operator S_d, given by $S_d = (W^{-1})^{-1}W^+$, is equal to the operator $\Omega_d^+(\Omega_d^-)^{-1}$, where the dilation wave operators are given by

$$\Omega_d^{(\pm)} = \operatorname*{s-lim}_{a\to\infty} (\mathscr{U}^{\mp}(a))^{-1}\mathscr{U}_0(a).$$

We will now show, at least in the special case described above, that $\Omega_d^{\pm} = \Omega^{\pm}(H,H_0)$, where the wave operators $\Omega^{\pm}(H,H_0) = \text{s-lim}_{t\to\mp\infty} V(t)^{-1}V_0(t)$, and that in this case the wave operators $\Omega^{\pm}(H,H_0)$ are asymptotically complete and that $S_d = S$, the usual scattering operator given by $S = \Omega^+(\Omega^-)^{-1}$.

To obtain this proof we have to construct some representations of the group $\mathrm{SL}(2,\mathbf{R})$ on the space of scattering states. These representations occur quite naturally whenever we have a quantum mechanical system whose time evolution is governed by a Schrödinger equation. The form of these representations is discussed in the Appendix. If

$$u(b) = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}, \quad s(a) = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix},$$

and

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

are the generators of $\mathrm{SL}(2,\mathbf{R})$ and $\pi(u(b))$, etc., the unitary representation of them on the Hilbert space $\mathscr{H}$, then we interpret $\pi(u(b))$ as the group of time translations and $\pi(s(a))$ as the dilation group. It is proven in the Appendix, for these representations,

$$\lim_{b\to\infty} \|\pi(u(b))f - \pi(s(b))(\pi(w))^3 f\| = 0. \tag{14}$$

This is just an abstract Hilbert space expression of the usual asymptotic behavior of the free time evolution $e^{-iH_0b}f$, $f \in \mathbf{L}^2(\mathbf{R}^3)$, and thus we are encouraged to conclude that the presence of a projective representation of $\mathrm{SL}(2,\mathbf{R})$ of the type described is the structure that forces this asymptotic behavior.

This claim is further enhanced by the fact that if the wave operators $\Omega^{\pm}(H,H_0)$ exist and are asymptotically complete with range $\mathscr{H}_{ac}(H)$, the spectral subspace of absolute continuity of H, and if we take $H_1 = H/\mathscr{H}_{ac}(H)$, then there exists on $\mathscr{H}_{ac}(H)$ two unitary representations of $\mathrm{SL}(2,\mathbf{R})$, $\pi^{\pm}(g)$ of the type discussed in the Appendix, where

$$\pi^{\pm}(g) = \Omega^{\pm}\pi_0(g)(\Omega^{\pm})^{-1}.$$

$\pi_0(g)$ is a unitary representation of $\mathrm{SL}(2,\mathbf{R})$ on $\mathbf{L}^2(\mathbf{R}^3)$ that is constructed in the Appendix. We thus have

Theorem 4: If the wave operators $\Omega^{\pm}(H,H_0)$ exist and are asymptotically complete then $(\mathscr{H}_{ac}(H), e^{-iH_1t})$ is a dilating scattering system with the representations of $\mathbf{R}_*^+$ given by $\mathscr{U}^{\mp}(a) = \Omega^{\pm}\mathscr{U}_0(a)(\Omega^{\pm})^{-1}$ and the dilation wave operators exist and equal the wave operators,

$$\Omega_d^{\pm} = \Omega^{\pm}(H,H_0).$$

This theorem is a direct consequence of the following, which is proved in the Appendix.

Theorem 5: Let $(\mathscr{H}_0, V_0(t))$ represent a free system and $(\mathscr{H}_0, V(t))$ a scattering system that carries true representations $\pi^{\pm}(g)$ of SL(2,**R**) so that

$$\pi^{\pm}(u(b)) = V(b),$$

$$\pi^{\pm}(s(a)) = \mathscr{U}^{\pm}(a^2),$$

and

$$\pi^{\pm}(w) = \Phi^{\pm},$$

then

(1) $\Omega_d^{\pm}$ exist on $\mathscr{H}_0$, if and only if $\Omega^{\pm}(H,H_0)$ exist on $\mathscr{H}_0$ and

(2) $\Omega_d^{\pm} = \Omega^{\pm}(H,H_0)$ if and only if either $\Omega^{\pm}(H,H_0)\Phi_0 = \Phi^{\pm}\Omega^{\pm}(H,H_0)$ or $\Omega_d^{\pm}\Phi_0 = \Phi^{\pm}\Omega_d^{\pm}$.

It sould be noted that the existence of the $\Omega^{\pm}(H,H_0)$, while sufficient to prove the existence of the $\Omega_d^{\pm}$, does not give the equality of the $\Omega^{\pm}(H,H_0)$ with the $\Omega_d^{\pm}$. The intertwining property of (2) of Theorem 5 is in some way equivalent to the completeness of the $\Omega^{\pm}(H,H_0)$, because by definition the $\Omega_d^{\pm}$ both have range equal to $\mathscr{H}_{ac}(H)$.

The converse, that the existence of the $\Omega_d^{\pm}$ implies the existence and completeness of the $\Omega^{\pm}(H,H_0)$ and that $\Omega^{\pm}(H,H_0) = \Omega_d^{\pm}$, follows from Theorem 3, in the special case that $\mathscr{H} = \mathscr{H}_0$ and $D_0 = D_+ = D_-$.

Theorem 6: Suppose that $V(t)$ and $V_0(t)$ act on the same Hilbert space $\mathscr{H}_0$, and that the dilating wave operators $\Omega_d^{\pm} = \text{s-lim}_{a\to\infty} (\mathscr{U}^{\mp}(a))^{-1}\mathscr{U}_0(a)$ exist as unitary operators on $\mathscr{H}_0$. If we can take the dilating subspaces D_0, D_+, and D_- to be the same subspace then the wave operators $\Omega^{\pm}(H,H_0)$ exist, are equal to the $\Omega_d^{\pm}$, and are therefore asymptotically complete.

Proof: Under these hypotheses, Theorem 3 implies that $\Omega_d^{\pm} = (W^{\mp})^{-1}W_0$ and hence that

$$\Omega_d^{\pm} V_0(t) = V(t)\Omega_d^{\pm} \quad \text{for all } t\in\mathbf{R},$$

and that

$$\Omega_d^{\pm}\mathscr{U}_0(a) = \mathscr{U}^{\mp}(a)\Omega_d^{\pm} \quad \text{for all } a\in\mathbf{R}_*^+.$$

Therefore if we set

$$\begin{aligned}\Phi^{\pm} &= (W^{\pm})^{-1}W_0\Phi_0W_0^{-1}W^{\pm}\\ &= \Omega_d^{\mp}\Phi_0(\Omega_d^{\mp})^{-1}\end{aligned}$$

then we have two unitary representations of SL(2,R) on $\mathscr{H}_0$, which are generated by $V(b)$, $b\in\mathbf{R}$, $\mathscr{U}^{\pm}(a^2)$, $a\in\mathbf{R}_*^+$, and $\Phi^{\pm}$. Moreover $\Omega_d^{\pm}\Phi_0 = \Phi^{\mp}\Omega_d^{\pm}$ and so, by Theorem 5, the wave operators $\Omega^{\pm}$ exist and equal the $\Omega_d^{\pm}$. ■

While it is imperative to have $D_+ = D_-$, the special assumptions needed to obtain the equivalence of the $\Omega_d^{\pm}$ with the asymptotically complete $\Omega^{\pm}(H,H_0)$ are, we believe, merely a technical problem and with careful work it should be possible to get rid of these restrictions. Nevertheless, we have that the dilation scattering operator S_d is equal to the usual scattering operator S for simple scattering processes where H has no bound states.

3. POTENTIAL SCATTERING

In this section, we give an example of how to construct the dilating representations $\mathscr{U}^{\pm}(a)$ of $\mathbf{R}_*^+$ without assuming, *a priori*, that the subspace on which these unitaries are defined is $\mathscr{H}_{ac}(H)$. The method used constructs the generators $A^{\pm}$ of the $\mathscr{U}^{\pm}(a)$.

The argument runs as follows. We define the subspace $\mathscr{H}$ on which the $\mathscr{U}^{\pm}(a)$ will ultimately act. Let

$$\mathscr{M} = \left\{ f\in\mathscr{H}_0 \,\middle|\, \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \|(K-2H)f_t\|dt = 0\right\}.$$

Here $K = i[H,A_0]$, where A_0 is the generator of the usual representation of the dilation group on $\mathbf{L}^2(\mathbf{R}^3)$. $\mathscr{H}$ is the closure of $\mathscr{M}$. It follows from the definition of $\mathscr{M}$ that if $f\in\mathscr{M}$ there exists a sequence of times $\{t_n\}$ tending to infinity with n such that

$$\lim_{n\to\infty} \|(K-2H)f_{t_n}\| = 0.$$

This means that $i[H,A(t_n)]f \to 2Hf$ as n tends to infinity. A^+ is then defined as the limit as n tends to infinity of $A(t_n) - F_n(H)$, where $F_n(H)$ is an operator valued function of H for each n, and we take the strong resolvent operator limit.

It should be noted that $\int_{-T}^{T} \|(K-2H)f_t\|dt$ is related to the classical time delay for the state f between the times $-T$ and T.[6] Therefore, $\mathscr{M}$ describes states whose time delay divergences as T tends to infinity slower than T^{-1}.

In order that this method should work, we must restrict the class of Hamiltonians $H = H_0 + V$ in the following way. V must be such that $K = i[H,A_0]$ exists and defines a self-adjoint operator with domain $D(K)$ such that $D(K)\cap D(H)$ is a core for both K and H.

Definition: Let

$$\mathscr{M} = \left\{ f\in\mathbf{L}^2(\mathbf{R}^3) \,\middle|\, \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \|(K-2H)f_t\|dt = 0\right\}. \quad (15)$$

$\mathscr{M}$ is clearly a linear subspace of $\mathbf{L}^2(\mathbf{R}^3)$ and is invariant under e^{-iHt} for all t.

Furthermore, we have the following property.

Theorem 7: If f_λ is an eigenvector of H with eigenvalue λ, $Hf_\lambda = \lambda f_\lambda$, and $\lambda \neq 0$, then f_λ does not belong to $\mathscr{M}$.

Proof: If $f\in\mathscr{M}$ then

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} (f,K(t)f)\,dt = 2(f,Hf). \quad (16)$$

Now assume that $f_\lambda\in\mathscr{M}$.

The left-hand side of (16) is 0 by the Virial theorem while the right-hand side is $2\lambda\|f_\lambda\|^2 \neq 0$. Hence $f_\lambda\notin\mathscr{M}$. ■

If we now assume that V is smooth enough that the following formal expansions are valid for t finite,

$$\rho^2(t)f = \frac{m}{2}r^2(t)f = \rho^2(0)f + tA_0f + \frac{t}{2}\int_0^t K(s)f\,ds$$

and

$$A_0(t)f = A_0f + \int_0^t K(s)f\,ds \quad \text{for all } f\in\mathscr{M},$$

then we have

Theorem 8: If $f\in\mathscr{M}$, then

$$\lim_{|t|\to\infty} \left\|\left(\frac{\rho^2(t)}{t^2} - H\right)f\right\| = 0 \quad (17)$$

and

$$\lim_{|t|\to\infty} \left\|\left(\frac{A_0(t)}{t} - 2H\right)f\right\| = 0. \quad (18)$$

Proof: We will only prove (17) as $t\to\infty$; the other limits can be proven in a similar way.
For $t>0$,

$$\left|\left|\left(\frac{\rho^2(t)}{t^2}-H\right)f\right|\right|\leqslant\frac{1}{t^2}\|\rho^2 f\|+\frac{1}{t}\|A_0 f\|+\frac{1}{2t}\int_0^t\|(K(s)-2H)f\|\,ds.$$

Now take the limit as $t\to\infty$, using the fact that $f\in\mathscr{M}$ to prove the result. ∎

Notice that as a result of (17), H must be positive on $\mathscr{M}$, for if $f\in\mathscr{M}$ then

$$\lim_{|t|\to\infty}\left(f,\frac{\rho^2(t)}{t^2}f\right)=(f,Hf). \tag{19}$$

The limit (19) has the obvious physical interpretation that if $f\in\mathscr{M}$ then as $|t|$ tends to infinity f_t is essentially free as its total energy is kinetic.

It also follows from (19) that if $(f,Hf)>0$ then as $|t|$ tends to infinity $\|r(t)f\|$ tends to infinity. We can now observe the following hierarchy of possibilities. If

$$\lim_{T\to\infty}\int_0^{\pm T}(K(s)-2H)f\,ds \tag{20}$$

exists for all $f\in\mathscr{M}$, then the generators $A^{\pm}$ exist and are given by

$$A^{+}=\lim_{t\to\infty}(A_0(t)-2Ht)=A_0+\int_0^{\infty}(K(s)-2H)\,ds, \tag{21a}$$

$$A^{-}=\lim_{t\to-\infty}(A_0(t)-2Ht)=A_0-\int_{-\infty}^{0}(K(s)-2H)\,ds. \tag{21b}$$

This is related to the property of H-smoothness (Ref. 3, XIII.7). If $(K-H)$ is positive then condition (20) implies that $(K-H)^{1/2}$ is H-smooth.

If the limit (20) does not exist, but there exists an operator-valued function $Q(t)$ that commutes with H such that

$$\lim_{T\to\infty}\int_0^{\pm T}(K(s)-2H-Q'(s))f\,ds<\infty, \tag{22}$$

then the generators $A^{\pm}$ exist and are given by

$$A^{\pm}=\lim_{t\to\infty}(A_0(t)-2Ht-Q(t))=A_0+\int_0^{\pm\infty}(K(s)-2H-Q'(s))\,ds. \tag{23}$$

This situation arises for long-range potentials. For example, if V is a function of $r=|x|$ only, say $V=c/r$, then by (17) we have that $\|(\rho^2(t)/t^2-H)f\|\to 0$ as $t\to\infty$ for all f in some dense subset of $\mathscr{M}$. If this subset is a core for H and r^2 and is invariant under e^{-iHt}, then the limit can be taken in the strong resolvent sense. This implies that for any smooth function g of a single variable

$$\|(g(\rho^2(t)/t^2)-g(H))f\|\to 0\quad\text{as}\quad t\to\infty.$$

Take $\delta V=K-2H$, then $\delta V=-(x\cdot\nabla V-2V)=-c/r$ as $H=H_0+V$, and hence $ds\,\delta V$ is C^{∞} except at the origin,

$$\|(\delta V(\rho^2(t)/t)-\delta V(\sqrt{H}))f\|\to 0.$$

Now as δV is homogeneous, this implies that

$$|t|\;\|(\delta V(r(t))-\delta V(\sqrt{H/mt}))f\to 0. \tag{24}$$

So if we choose $Q(s)$ to be an antiderivative of $\delta V(\sqrt{(H/m)}|s)$ for $|s|$ sufficiently large and zero otherwise, we get that

$$\lim_{T\to\infty}\int_0^{\pm T}(K(s)-2H-Q'(s))f\,ds<\infty$$

because if s_0 is large enough

$$\int_{s_0}^{T}\|(K(s)-2H-Q'(s))f\|ds=\int_{s_0}^{T}\left|\left|\left(dV(r(s))-\delta V\left(\sqrt{\frac{H}{m}}s\right)\right)f\right|\right|\,ds.$$

But $\|(\delta V(r(s))-\delta V(\sqrt{H/m}\,s))f\|$ converges to zero, as s goes to infinity, faster than $1/s$ so the integral is finite.

APPENDIX: UNITARY REPRESENTATIONS OF SL(2,R) AND WAVE OPERATORS

We are interested in the following continuous unitary representation of SL(2,R) on $\mathbf{L}^2(\mathbf{R}^3)$. Let

$$u(b)=\begin{pmatrix}1 & b\\ 0 & 1\end{pmatrix},\quad s(a)=\begin{pmatrix}a & 0\\ 0 & a^{-1}\end{pmatrix}$$

and

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

belong to SL(2,R) where $b\in\mathbf{R}$ and $a\in\mathbf{R}_*^+$. We have the following momentum space representation of SL(2,R).

$$(\pi_0(u(b))f)(p)=e^{-i(p^2/2)b}f(p), \tag{A1}$$

$$(\pi_0(s(a))f)(p)=a^{3/2}f(ap), \tag{A2}$$

and

$$(\pi_0(w)f)(p)=-i^{3/2}\hat{f}(p)=-i^{3/2}(\mathscr{F}f)(p), \tag{A3}$$

where

$$(\mathscr{F}f)(p)=\hat{f}(p)=\frac{1}{\sqrt{2\pi}}\int f(q)e^{-ip\cdot q}dq$$

is the Fourier transform of the function f considered as a unitary map from $\mathbf{L}^2(\mathbf{R}^3)$ to itself.

It follows from the general theory of generators and relations for SL(2,R) (Ref. 7, XI, Sec. 2) that to prove that this is a representation of SL(2,R) we must show that

$$\pi_0 s(a))=\pi_0(w)\pi_0(u(a^{-1}))\pi_0(w)\pi_0(u(a))\pi_0(w)\pi_0(u(a)^{-1}) \tag{A4}$$

and that

$$\begin{aligned}&(1)\ \pi_0(u(b))\text{ is a representation of the group }\mathbf{R},\\&(2)\ \pi_0(s(a))\text{ is a representation of the group }\mathbf{R}_*^+,\\&(3)\ \pi_0(w^2)=\pi_0(s(-1)),\\&(4)\ \pi_0(s(a))\pi_0(u(b))\pi_0(s(a^{-1}))=\pi_0(u(ba^2))\end{aligned} \tag{A5}$$

all hold.

We first check that

$$(\pi_0(w^2)f)(p) = -if(-p) = (\pi_0(s(-1))f(p).$$

Then we show that $\pi_0(s(a))$ defined by Eq. (A2) satisfies the relation (A4). This calculation is tedious but fortunately it has been done by Lang (Ref. 6, XI, Sec. 1) with a slightly difference choice of coordinates, namely, his variable x is related to p by $x = p/\sqrt{2\pi}$. Once the relation (A4) has been checked the test of the relations (A5) follow immediately. Except that $\Phi_0^4 = -I$, which reflects the fact that on $\mathbf{L}^2(\mathbf{R}^3)$ we really do not have unitary representation of SL$(2,R)$, but rather a unitary representation of its nontrivial twofold cover $\widetilde{\mathrm{SL}}(2,R)$, with the nontrivial element of the kernel of the covering map acting by minus the identity on $\mathbf{L}^2(\mathbf{R}^3)$. This unitary representation of the covering group $\widetilde{\mathrm{SL}}(2,R)$ of SL$(2,R)$ gives the projective representation of SL$(2,R)$, which is used in the following.

The relation (A5.4) is intimately related to the condition (2) that was imposed upon the dilating representation $\mathscr{U}(a)$ of $\mathbf{R}_*^+$ and the time evolution $V(t)$ to describe scattering processes. Recall that we demanded that

$$\mathscr{U}(a)V(t)\mathscr{U}(a)^{-1} = V(ta). \tag{A6}$$

If we take a new representation $\mathscr{U}'(a)$ of $\mathbf{R}_*^+$ given by $\mathscr{U}'(a) = \mathscr{U}(a^2)$, then $\mathscr{U}'(a)V(t)\mathscr{U}'(a)^{-1} = V(ta^2)$, which is relation (5.4) when we identify $V(t)$ with $\pi_0(u(t))$ and $\mathscr{U}'(a)$ with $\pi_0(s(a))$.(A5.4)

The relationship between this representation of SL$(2,R)$ and scattering theory is seen more readily in the configuration space realization of this representation. Let us denote by $\tilde{\pi}_0(g)$ the configuration space realization of this representation. The generators of SL$(2,R)$ take the following forms for $\tilde{f}(x)\in\mathbf{L}^2(\mathbf{R}^3)$.

$$\begin{aligned}(\tilde{\pi}_0(u(b))\tilde{f})(x) &= e^{i(b/2)\nabla^2}\tilde{f}(x),\\ (\tilde{\pi}_0(s(a))\tilde{f})(x) &= a^{-3/2}\tilde{f}(a^{-1}x),\end{aligned} \tag{A7}$$

and

$$(\tilde{\pi}_0(w)\tilde{f})(x) = -i^{-3/2}(\mathscr{F}\tilde{f})(x),$$

where again $\mathscr{F}$ is the Fourier transform operator on $\mathbf{L}^2(\mathbf{R}^3)$. The relation (4) holds for this realization, and it is easy to show that

$$\lim_{a\to\infty} \|\tilde{\pi}_0(s(a))\tilde{f} - \tilde{\pi}_0(u(a))(\tilde{\pi}_0(w))^3\tilde{f}\| = 0 \tag{A8}$$

for all $\tilde{f}\in\mathbf{L}^2(\mathbf{R}^3)$. To get this result we have used the fact that $(\tilde{\pi}_0(w))^2$ commutes with $\tilde{\pi}(u(a))$ for all a. Using the fact that $(\tilde{\pi}_0(w))^4 = -I$ we can rewrite (A8) as

$$\lim_{a\to\infty} \|\tilde{\pi}_0(u_0(a))\tilde{g} - i^{-3/2}\tilde{\pi}_0(s(a))\mathscr{F}\tilde{g}\| = 0. \tag{A9}$$

If we put $a = t/m$ and make the identifications $V_0(t) = \tilde{\pi}_0(u(t/m))$ and $\tilde{\pi}_0(s(t/m)) = \widetilde{\mathscr{U}}(t^2/m^2) = \widetilde{\mathscr{U}}'(t/m)$, we get the result, with $H_0 = -\nabla^2/2m$,

$$\lim_{t\to\infty}\int \left|(e^{-iH_0t}\tilde{g}(x) - i^{-3/2}(t/m)^{-3/2}\mathscr{F}\tilde{g}\left(\frac{mx}{t}\right)\right|^2 dx = 0. \tag{A9$'$}$$

A similar argument holds for the limit as t tends to minus infinity. Equation (4) holds for negative $a\neq 0$, but as we have only defined $\mathscr{U}(a)$ and $\mathscr{U}'(a)$ for positive a we must use the fact that $s(-a)W^2 = s(a)$ to get for negative a,

$$\begin{aligned}&\tilde{\pi}_0(s(-a))\\ &\quad = \tilde{\pi}_0(w)\tilde{\pi}_0(u(a^{-1}))\tilde{\pi}_0(w)\tilde{\pi}_0(u(a))\tilde{\pi}_0(w)\tilde{\pi}_0(u(a^{-1}))\tilde{\pi}_0(w^2).\end{aligned}$$

Then taking the limit as a tends to minus infinity and using $(\tilde{\pi}_0(w))^4 = -I$ we get, for $\tilde{f}\in\mathbf{L}^2(\mathbf{R}^3)$,

$$\lim_{a\to-\infty} \|\tilde{\pi}_0(s(-a))\tilde{f} + \tilde{\pi}_0(u(a))\tilde{\pi}_0(w)\tilde{f}\| = 0 \tag{A10}$$

or, using the fact that $\tilde{\pi}_0(w) = -i^{-3/2}\mathscr{F}$,

$$\lim_{a\to-\infty} \|\tilde{\pi}_0(u(a))\tilde{g} - i^{3/2}\pi_0(s(-a))\,\mathscr{F}^{-1}\tilde{g}\| = 0.$$

Now take $a = -t/m$ and use the equations

$$\tilde{\pi}_0(u(-t/m)) = V_0(-t) = e^{iH_0t} \quad \text{and}$$

$$\tilde{\pi}_0(s(t/m)) = \widetilde{\mathscr{U}}'(t/m)$$

to get

$$\lim_{t\to\infty}\int_{\mathbf{R}^3} \left|e^{iH_0t}\tilde{g}(x) - i^{3/2}(t/m)^{-3/2}(\mathscr{F}\tilde{g})\left(-\frac{mx}{t}\right)\right|^2 dx = 0. \tag{A11}$$

There are two observations that we wish to make concerning Eqs. (A10) and (A11). Firstly, these asymptotic relations between the time evolution and the dilating representations of $\mathbf{R}_*^+$ follow from the assumption that there is a unitary representation of SL$(2,R)$ in which the representative of $\begin{pmatrix}1 & b\\ 0 & 1\end{pmatrix}$ is the time evolution group of unitaries. Secondly, that Eqs. (A10) and (A11) are Hilbert space expressions for the asymptotic behavior of $\exp(-iH_0t)$, because if the integrands of these expressions are taken to tend to zero pointwise we have the usual pointwise asymptotic form for the free Hamiltonian H_0 (Ref. 8, Theorem IX.31).

In the light of these observations we can say that a quantum system $(\mathscr{H}, V(t))$ describes a simple scattering system if there exists a pair of unitary representations $\pi^\pm(g)$ of SL$(2,R)$ in each of which the representation of $u(b) = \begin{pmatrix}1 & b\\ 0 & 1\end{pmatrix}$ is e^{-iHb}, that is both $\pi^\pm(u(b)) = e^{-iHb}$ and the representatives of

$$s(a) = \begin{pmatrix}a & 0\\ 0 & a^{-1}\end{pmatrix}$$

are related to the dilating representations $\mathscr{U}^\pm(a)$ of $\mathbf{R}_*^+$,

$$\pi^\pm(s(a)) = \mathscr{U}^\pm(a^2) = \mathscr{U}'^\pm(a).$$

In this approach we have an element in addition to those in the dilation group approach, namely the representatives of the generator W of SL$(2,R)$. Suppose that $\Phi^\pm = \pi^\pm(w)$. Then it follows from the arguments used above that

$$\lim_{a\to\infty} \|V(a)g - \mathscr{U}^-(a^2)(\Phi^+)^{-3}g\| = 0 \tag{A12}$$

and

$$\lim_{a\to-\infty} \|V(a)g - \mathscr{U}^+(a^2)(\Phi^-)^3g\| = 0. \tag{A13}$$

If, as before, we define the dilation wave operators $\Omega_d^\pm$, in the case when $\mathscr{H}_0 = \mathscr{H}$, as $\Omega_d^\pm = \lim_{a\to\infty}\mathscr{U}^\mp(a)^{-1}\mathscr{U}_0(a)$, then we have the following theorem.

Theorem A1: Let $(\mathscr{H}_0, V_0(t))$ represent a free system and $(\mathscr{H}_0, V(t))$ a scattering system that carries two representations $\pi^\pm(g)$ of SL$(2,R)$ so that

$$\pi^\pm(u(b)) = V(b),\ \pi^\pm(s(a)) = \mathscr{U}^\pm(a^2)$$

and

$$\pi^\pm(w) = \Phi^\pm,$$

then

(1) $\Omega_d^\pm$ exists on $\mathscr{H}_0$ if and only if $\Omega^\pm(H,H_0)$ exists on $\mathscr{H}_0$, and

(2) $\Omega_d^\pm = \Omega^\pm(H,H_0)$ if and only if either $\Omega^\pm(H,H_0)\Phi_0 = \Phi^\pm\Omega^\pm(H,H_0)$ or $\Omega_d^\pm\Phi_0 = \Phi^\pm\Omega_d^\pm$.

Proof: The first conclusion follows from a simple $\epsilon/3$ argument using the asymptotic relations between the time evolution and the dilation representation.

We will only prove that if $\Omega^-(H,H_0)$ exists then Ω_d^- exists, the argument in the opposite direction is essentially the same. Suppose $\Omega^-(H,H_0) = \Omega^-$ exists, then for all $f \in \mathscr{H}_0$

$$\lim_{t\to\infty} \|V(t)\Omega^- f - V_0(t)f\| = 0.$$

Given $h \in \mathscr{H}_0$ we wish to find when there exists a unique $g \in \mathscr{H}_0$ such that

$$\lim_{a\to\infty} \|\mathscr{U}_0(a^2)h - \mathscr{U}^+(a^2)g\| = 0,$$

$$\begin{aligned}&\|\mathscr{U}_0(a^2)h - \mathscr{U}^+(a^2)g\|\\ &\leqslant \|\mathscr{U}_0(a^2)h - V_0(a)\Phi_0^3 h\| + \|V_0(a)\Phi_0^3 h - V(a)\Omega^-\Phi_0^3 h\|\\ &+ \|V(a)\Omega^-\Phi_0^3 h - \mathscr{U}^+(a^2)g\|.\end{aligned}$$

The first two terms converge to zero as $a\to\infty$ by the asymptotic relation between $\mathscr{U}_0(a^2)$ and $V_0(a)$ and by the assumption of the existence of Ω^- on all of $\mathscr{H}_0$. The third term vanishes in the limit as a tends to infinity if

$$g = (\Phi^+)^{-3}\Omega^-\Phi_0^3 h. \qquad \text{(A14)}$$

But for any h, g defined by (A14) belongs to $\mathscr{H}_0$ and so the dilation wave operator Ω_d^- exists and is given by

$$\Omega_d^- = (\Phi^+)^{-3}\Omega^-\Phi_0^3. \qquad \text{(A15)}$$

■

ACKNOWLEDGMENTS

This work was done while the author was on study leave from Macquarie University at the Centre de Physique Théorique, Luminy. I wish to thank Macquarie University for granting me my study leave, and to especially thank the C.P.T.–C.N.R.S. Luminy and the Faculté des Sciences de Luminy, Université d'Aix-Marseille II, for their generous hospitality, both personal and financial, during my stay. This work was benefitted from discussions with many people at the C.P.T., especially from Jean-Michel Combes, Alex Grossmann, and Eric Mourre. Finally, I wish to thank the Dean of the Faculté des Sciences de Luminy, Professor M. Mebkhout, for inviting me to Luminy and Madame Bourgeois and the secretariat staff for all their assistance.

[1] P. D. Lax and R. S. Phillips, *Scattering Theory* (Academic, New York, 1967).

[2] V. S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand Reinhold, New York, 1970), Vol. II.

[3] A. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. III, *Scattering Theory* (Academic, New York, 1979).

[4] I. M. Gel'fand and M. A. Naimark, Dokl. Akad. Nauk. SSSR **55**, 570 (1947).

[5] E. W. Aslaksen and J. R. Klauder, J. Math. Phys. **9**, 206 (1968).

[6] H. Narnhoffer, Phys. Rev. D **22**, 2387 (1980).

[7] S. Lang, $SL_2(R)$ (Addison-Wesley, Reading, MA, 1975).

[8] A. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. II, *Fourier Analysis, Self adjointness* (Academic, New York, 1977).

Non-Grassmann quantization of the massive Thirring model

Piotr Garbaczewski
Institute of Theoretical Physics, University of Wrocław, 50-205 Wrocław, Poland

(Received 22 April 1982; accepted for publication 29 July 1982)

A direct quantization of the c-number (semi) classical massive Thirring model in the inverse scattering formalism leads to the Bose massive Thirring model, which is equivalent to the conventional Fermi one, both having identical S-matrices and bound-state spectra.

PACS numbers: 03.70. + k, 03.65.Sq

A. There is a wide-spread belief among quantum field theorists that (semi) classical c-number spinor fields are unrelated to their quantized Fermi partners. A spectacular manifestation of this situation is the use of Grassmann algebra-valued spinor models as the would-be-the-only-reasonable pseudoclassical levels for Fermi systems. For example, the path integration methods if applied to spinor systems, by the very assumption do exclude the conventional path notion in the c-number function ring. An anticommuting, i.e., Grassmann algebra valued ring is then conventionally in use. From the practical point of view (perturbative calculations) this idea is quite justified, and it was consequently the main motivation for the studies of the Grassmann algebra valued massive Thirring model, which has been proved to be a completely integrable system.[1] There appeared, however, a problem of the quantization of this system via the quantum spectral transform method (which is successful for many other 1 + 1 dimensional models). This quantization route which we call a *Grassmann quantization* of the massive Thirring model still remains uncompleted.

Quite the contrary, in the series of papers, Refs. 2–4, we have investigated the relationships between the (semi) classical c-number spinor systems and the respective quantum Fermi models, following the idea of Ref. 5 that the c-number solutions of the classical spinor field equations should have some relevance for the construction of the appropriate quantum field theory. In Refs. 2 and 3 we have demonstrated that the relationship exists provided the Fermi models admit a "bosonization" in terms of free Bose fields. In the practical application of Ref. 4 it means that the Fermi massive Thirring model admits three different types of the asymptotic (Haag) expansions, depending on the choice of the state space, and provided one takes into account spaces generated by soliton coherent states, see, e.g., Ref. 4.

The underlying expansions appear either in terms of the massive vector boson without the Proca constraint, or in terms of the neutral massive scalar (then the relationship with the sine-Gordon model can be established), and under special circumstances only, in terms of the free (asymptotic) two-component fermion. The latter case fits into the conventional asymptotic completeness condition, otherwise the fermion being confined.

Because in the light of Ref. 4 there exists an indirect relationship of the c-number massive Thirring model (MT) to the Fermi MT, it is quite natural to state a problem of the direct quantization of the c-number massive Thirring model. This route we call a *non-Grassmann quantization* of the MT. We accomplish this quantization in the quantum inverse transform formalism of Ref. 6, by exploiting both the results of Refs. 7 and 8 concerning the complete integrability of the c-number MT and those on the quantization of the sine-Gordon model.[9]

We demonstrate that in the quantum inverse method, the Bose quantized MT has a lattice approximation, which is equivalent to that of the quantum sine-Gordon model. By repeating the arguments of Ref. 9, one is then capable of deriving a continuum limit in which both models have the bound-state spectrum (and the S-matrix) identical to this of the conventional Fermi MT.

B. The classical (semiclassical in fact) c-number massive Thirring model is known to be a completely integrable system.[7,8] The field equation

$$\begin{gathered}(-i\gamma_\nu\partial^\nu + m)\,\psi = g\gamma^\nu\psi(\bar\psi\gamma_\nu\psi), \quad \nu = 0,1,\ g>0\,,\\ \gamma^0 = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\quad \gamma^1 = \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix},\quad \gamma^5 = \gamma^0\gamma^1\,,\\ \bar\psi = \psi^*\gamma^0\,.\end{gathered}\tag{1}$$

can be rewritten as the system

$$\begin{aligned}&-i\psi_{1t} - i\psi_{1x} + m\psi_2 + 2g|\psi_2|^2\psi_1 = 0\,,\\ &-i\psi_{2t} + i\psi_{2x} + m\psi_1 + 2g|\psi_1|^2\psi_2 = 0\,,\end{aligned}\tag{2}$$

which is known to admit classical (c-number spinors) soliton solutions.[8] An equivalent description of Eq. (2) is known to be provided by the commutator $[X, T]_- = 0$ of the two objects:

$$\begin{aligned}X &= 2i\partial_x + g(\bar\psi\gamma_1\psi)\gamma^5 + (2mg)^{1/2}\\ &\quad\times\begin{pmatrix}0, \lambda\psi_2^* - \lambda^{-1}\psi_1^*\\ \lambda\psi_2 - \lambda^{-1}\psi_1, 0\end{pmatrix} - \frac{m}{2}(\lambda^2 - \lambda^{-2})\gamma^5\,,\\ T &= 2i\partial_t + g(\bar\psi\gamma_0\psi)\gamma^5 + (2mg)^{1/2}\\ &\quad\times\begin{pmatrix}0, \lambda\psi_2^* + \lambda^{-1}\psi_1^*\\ \lambda\psi_2 + \lambda^{-1}\psi_1, 0\end{pmatrix} - \frac{m}{2}(\lambda^2 + \lambda^{-2})\gamma^5\,,\end{aligned}\tag{3}$$

i.e., by the condition that all terms standing in the commutator at different powers of the spectral parameter λ do vanish. For $X = X(\lambda)$, we shall adopt the form

$$X' = \tfrac{1}{2} iX = -\partial_x + L(x,\lambda),$$

$$L(x,\lambda) = i\begin{pmatrix} \frac{1}{4}m(\lambda^{-2} - \lambda^{2}) + g(\rho_1 - \rho_2), & (\lambda\psi_2^* - \lambda^{-1}\psi_1^*)(mg/2)^{1/2} \\ (mg/2)^{1/2}(\lambda\psi_2 - \lambda^{-1}\psi_1), & \frac{1}{4}m(\lambda^{2} - \lambda^{-2}) - g(\rho_1 - \rho_2) \end{pmatrix}, \tag{4}$$

$$\rho_i = |\psi_i|^2, \quad i = 1,2$$

According to Ref. 6, a straightforward quantized version of the problem(1) appears if one replaces the classical fields $\psi_i(x)$ by the quantum operators $\hat{\psi}_i(x)$, satisfying the (equal $t = 0$ time) canonical commutation relations, not the canonical anticommutation ones (CAR) as demanded by convention:

$$[\hat{\psi}_i(x), \hat{\psi}_j^*(y)]_- = \alpha\delta_{ij}\delta(x-y), [\hat{\psi}_i(x), \hat{\psi}_j(y)]_- = 0, \tag{5}$$

provided we make a change in $L(x,\lambda)$; $L(x,\lambda) \to \hat{L}(x,\lambda)$,

$$\hat{L}(x,\lambda) = i\begin{pmatrix} \frac{1}{4}m(-\lambda^{2} - \lambda^{-2}) + g(-\xi\hat{\rho}_1 + \eta\hat{\rho}_2), & (mg/2)^{1/2}(\lambda\hat{\psi}_2^* - \lambda^{-1}\hat{\psi}_1^*) \\ (mg/2)^{1/2}(\lambda\hat{\psi}_2 - \lambda^{-1}\hat{\psi}_1), & \frac{1}{4}m(\lambda^{2} - \lambda^{-2}) + g(\eta\hat{\rho}_1 - \xi\hat{\rho}_2) \end{pmatrix}. \tag{6}$$

$$\hat{\rho}_i(x) = \hat{\psi}_i^*(x)\hat{\psi}_i(x), \quad i = 1,2, \quad \eta = \exp\gamma/\cosh\gamma,$$

$$\xi = \exp(-\gamma)/\cosh\gamma, \quad \gamma = \tfrac{1}{2}i \arcsin\alpha, \alpha\in R.$$

With the operator valued matrix $\hat{L}(x,\lambda)$ in hand, let us introduce the tensor product matrices

$$L' = \hat{L}\otimes I, \quad L'' = I\otimes\hat{L} \tag{7}$$

according to the rule

$$F\otimes G = \begin{pmatrix} f_{11}G, & f_{12}G \\ f_{21}G, & f_{22}G \end{pmatrix}. \tag{8}$$

Then the matrix equation

$$R(\lambda,\lambda')L'(x,\lambda)L''(x,\lambda') = L''(x,\lambda')L'(x,\lambda)R(\lambda,\lambda'), \tag{9}$$

can be solved by means of the 4×4 matrix $R = R(\lambda,\lambda')$, with the c-number matrix elements,[4]

$$R = \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & b & c & 0 \\ 0 & c & b & 0 \\ 0 & 0 & 0 & a \end{pmatrix}, \quad \begin{aligned} &a = 1, \\ &b = \sinh 2\gamma/\sinh(u + 2\gamma), \\ &c = \sinh u/\sinh(u + 2\gamma), \end{aligned} \tag{10}$$

where $\exp u = \lambda/\lambda' = \exp(v - v')$, $\exp v = \lambda$, $\exp v' = \lambda'$. Let us notice that the change of variables in (6),

$$v \to v - \gamma, \; v' \to v' - \gamma \tag{11}$$

does not affect the R-matrix (10) because $u = v - v' \to u$. Notice that (11) corresponds to the replacement $\lambda \to \lambda\exp(-\gamma)$. We shall adopt a bit more sophisticated version of (11), namely,

$$v \to v - (\gamma - i\pi/4), \quad v' \to v' - (\gamma - i\pi/4). \tag{12}$$

Recall that the parameter γ is purely imaginary: $\gamma = i\mu/2$, $\mu = \arcsin\alpha$. Consequently we arrive at

$$\hat{L}(x,\lambda) \to i\begin{pmatrix} -\frac{1}{2}m\sinh[2v - i(\mu - \pi/2)] + g(e^{i\mu/2}\hat{\rho}_2 - e^{-i\mu/2}\hat{\rho}_1), & (mg/2)^{1/2}(e^{v}\hat{\psi}_2^* - e^{-v}\hat{\psi}_1^*) \\ (mg/2)^{1/2}(e^{v}\hat{\psi}_2^* - e^{-v}\hat{\psi}_1^*), & \frac{1}{2}m\sinh[2v - i(\mu - \pi/2)] + g(e^{i\mu/2}\hat{\rho}_1 - e^{-i\mu/2}\hat{\rho}_2) \end{pmatrix}, \tag{13}$$

$$c = c(\lambda,\lambda') = \sinh(v - v')/\sinh(v - v' + i\mu),$$

$$b = (\lambda,\lambda') = i\sin\mu/\sinh(v - v' + i\mu),$$

$$\lambda = \exp v, \quad \lambda' = \exp v'.$$

C. In general how to apply the quantum spectral transform method on the continuum level is not straightforward. Usually one adopts some discretization scheme, like that in Ref. 9, where the quantum inverse scattering formalism as a basic ingredient includes a matrix equation:

$$X\Psi = \left(\frac{\partial}{\partial x} + iQ\right)\Psi = 0, \tag{14}$$

with $Q = Q(x) = i\hat{L}(x,\lambda)$. Its discretized version on a linear lattice of length L and spacing δ, $N = L/\delta$ reads

$$\Psi_{n+1} = L_n(x)\Psi_n,$$

$$L_n(\lambda) = I + i\int_{x_n}^{x_n+\delta} Q(z)\,dz = I - \int_{x_n}^{x_n+\delta} L(z,\lambda)\,dz, \tag{15}$$

$$x_m = -L/2 + n\delta, \quad n = 0,1,\ldots,N, \quad N = L/\delta.$$

In particular one finds

$$\Psi_{L/\delta} = \Psi_N := T(\lambda) = \overleftarrow{\prod_{n=0}^{N-1}} L_n(\lambda) = L_{N-1}(\lambda)\ldots L_0(\lambda). \tag{16}$$

The so defined transition operator for an interval $L = N\delta$,

$$T(\lambda) = \begin{pmatrix} A(\lambda) & B(\lambda) \\ C(\lambda) & D(\lambda) \end{pmatrix}, \tag{17}$$
$$R(\lambda,\lambda')(T(\lambda)\otimes T(\lambda')) = (T(\lambda')\otimes T(\lambda))R(\lambda,\lambda'),$$

is a fundamental object of the quantum inverse method. Upon discretization of (15), one represents $A(\lambda), B(\lambda), C(\lambda), D(\lambda)$ by operators in the $2N$ particle Hilbert space $\mathscr{H}_{2N} = \Pi_{i=1}^{N} \otimes (h \otimes \tilde{h})_i$ carrying a ($2N$ particle) Fock representation of the CCR algebra:

$$\psi_i(n)\Omega = 0,\ \forall i = 1,2,\ n = 1,2,\dots,N,\quad \Omega = \prod_{i=1}^{N} \otimes (\omega_0 \otimes \tilde{\omega}_0)_i\ ,$$
$$[\psi_i(n), \psi_j^*(m)]_- = \delta_{ij}\delta_{mn}\cdot\alpha,\quad [\psi_i(n), \psi_j(m)]_- = 0, \tag{18}$$
$$\psi_i(n) = \frac{1}{\sqrt{\delta}} \int_{R^1} \psi_i(x)\chi_n(x)\, dx,$$
$$\chi_n(x) = \begin{cases} 1, & x \in [x_n, x_n + \delta], \\ 0, & x \in [x_n, x_n + \delta]. \end{cases}$$

In particular we can consider the action of matrix elements of the operator $L_n(\lambda)$ on the Fock vacuum Ω. One immediately verifies that

$$\begin{aligned} L_{n21}\Omega &= 0, \\ L_{n11}\Omega &= \{1 - \tfrac{1}{2} im\delta \sinh[2v - i(\mu - \pi/2)]\}\Omega \\ &= \{1 + \tfrac{1}{2} m\delta \cosh(2v - i\mu)\}\Omega \\ &\cong \exp[\tfrac{1}{2} m\delta \cosh(2v - i\mu)]\cdot\Omega \\ &= \exp[a(\lambda)\cdot\delta]\cdot\Omega. \end{aligned} \tag{19}$$

In the above we use an identity $\sinh[(2v - i\mu) + i\pi/2] = i\cosh(2v - i\mu)$, Analogously,

$$L_{n22}\Omega \cong \exp[d(\lambda)\cdot\delta]\cdot\Omega = \exp\{-\tfrac{1}{2} m\delta \cosh(2v + i\mu)\}\Omega, \tag{20}$$

and consequently,

$$\exp[a(\lambda) + d(\lambda)]\delta = \exp(im\delta\cdot\sin\mu\cdot\sinh 2v), \tag{21}$$

i.e.,

$$\exp[a(\lambda) + d(\lambda)]N = \exp ikL = (\exp ik\delta)^{L/\delta}, \tag{22}$$
$$k = (m \sin\mu)\cdot\sinh 2v,$$

with [make a product of matrices L_n according to (16)]:

$$A(\lambda)\Omega = \exp[a(\lambda)N]\cdot\Omega,\quad D(\lambda)\Omega = \exp[d(\lambda)N]\cdot\Omega. \tag{23}$$

Hence in addition to the R-matrix (13), we have specified the reference (Fock) state Ω solving the eigenvalue problem for $A(\lambda), D(\lambda)$, Eqs. (19)–(23) and being annihilated by $C(\lambda)$. These data completely suffice to specify a representation of the algebra of A, B, C, D operators as defined by the commutation relation (17). Then we can construct the eigenvectors of the transfer operator

$$\mathrm{Tr}T(\lambda) = \mathscr{T}(\lambda) = A(\lambda) + D(\lambda), \tag{24}$$

as follows

$$|\lambda_1,\dots,\lambda_n) = \prod_{i=1}^{n} B(\lambda_i)\Omega, \tag{25}$$

provided we have satisfied the periodicity condition

$$\exp ik_i L = \prod_{j=n}^{n} \frac{\sinh(v_i - v_j + i\mu)}{\sinh(v_i - v_j - i\mu)} \tag{26}$$
$$k_i = (m \sin\mu)\cdot\sinh 2v_i,\quad v_i = \ln\lambda_i,\quad i = 1,\dots,n.$$

The respective eigenvalue reads

$$\mathscr{T}(\lambda)|\lambda_1,\dots,\lambda_n) = \Lambda(\lambda,\lambda_1,\lambda_2,\dots,\lambda_n)|\lambda_1,\dots,\lambda_n), \tag{27}$$
$$\Lambda(\lambda,\lambda_1,\dots,\lambda_n) = \exp[a(\lambda)L] \prod_{j=1}^{n} \frac{1}{c(\lambda_j,\lambda)} + \exp[d(\lambda)L] \prod_{j=1}^{n} \frac{1}{c(\lambda,\lambda_j)}$$

with $c(\lambda,\lambda')$ given by (13).

D. By recalling Ref. 9 we find that upon a mere identification [compare, e.g., (1.29) in Ref. 9],

$$m_{\mathrm{sG}}^2\cdot\delta/4 = m_{\mathrm{MT}}, \tag{28}$$

the above representation becomes isomorphic with this found for the quantum sine-Gordon model on a lattice. Obviously letting $\delta\to 0$ (continuum limit) must be accompanied by $m_{\mathrm{sG}}\to\infty$ to keep m_{MT} finite. The $m_{\mathrm{sG}}\to\infty$ demand is quite natural in the light of our previous analysis of the relationships between the sine-Gordon and xyz Heisenberg models.[10,11] These two models can be considered as equivalent in the continuum limit, upon the lattice identification analogous to that of (28):

$$l'^2/16\delta = m_{\mathrm{sG}}^2\cdot\delta/4 \tag{29}$$

of the xyz model parameter l' (an elliptic modulus of Jacobi theta functions), see, e.g. Refs. 11 and 12 and the sine-Gordon coupling constant m_{sG}, where $\delta\to 0$ means both $m_{\mathrm{sG}}\to\infty$ and $l'\to 0$ (the weak anisotropy limit of Ref. 13.).

In the above discussion one must, however, remember that the Bose MT algebra (17) is represented in the Hilbert space $\mathscr{H}_N = \Pi_{i=n}^{N} {}^{\otimes}(h\otimes\tilde{h})_i$, while this for the sine-Gordon system in $\mathscr{H}_N = \Pi_{i=1}^{N} {}^{\otimes} h_i$, and this for the xyz model can be represented in a proper subspace $P\mathscr{H}_N = \Pi_{i=1}^{N} {}^{\otimes}(ph)_i$ of $\mathscr{H}_N$, with p being a two-level projection of Ref. 10 in h. $P\mathscr{H}_N$ can be equivalently rewritten as $\Pi_{i=n}^{N} {}^{\otimes}(C_2)_i$, where C_2 is a two-dimensional vector space.

On the lattice level both the Bose MT and sine-Gordon representations of the algebra (17) are equivalent and both become equivalent to the representation of the xyz Heisenberg model algebra in the continuum limit. In this case the Coleman's equivalence with the Fermi MT is a straightforward consequence.

E. With respect to the mass spectrum or the S-matrix arising in the continuum limit of the above models, the (Coleman's) equivalence of the Bose MT and the Fermi MT is guaranteed by the lattice identification of the Bose MT with the sine-Gordon model in the quantum inverse method. The procedure of Ref. 9 allows then the recovery of a continuum limit for the spectrum of the lattice models.

A few words should be said about the related quantum fields. One knows that while passing from the *xyz* model to the Fermi MT, there is a natural way to recover Fermi fields from renormalized lattice spin 1/2 degrees.[10–14] However, if one starts from the lattice Bose models,[10] like the above sine-Gordon or Bose MT, the emergence of fermions is not apparent at all. The lesson of Refs. 4, 10, and 11 in this context is that these lattice models can be constrained via the so called spin 1/2 approximation to the *xyz* model. A continuum limit of such a projected lattice Bose model gives the *S*-matrix and the spectrum identical to that of the sine-Gordon/Fermi MT models. However, in contrast to the full Bose MT, the resulting state space is precisely the space of Fermi states of the quantum Bose field, see, e.g. Refs. 4 and 11.

On such a space the irreducible Fermi fields can be consistently defined. Certainly the Bose MT can be rewritten as the *reducible Fermi model*. For 1 + 1 dimensional models, a formal relationship with the spin (1/2) *xyz* Heisenberg model can be introduced by means of the previously defined projection *P*:

$$\begin{aligned} H_B &= PH_BP + PH_B(1-P) \\ &\quad + (1-P)H_BP + (1-P)H_B(1-P), \end{aligned} \tag{30}$$

where (this is a spin 1/2 approximation constraint)

$$PH_BP \equiv H_{xyz}. \tag{31}$$

For the sine-Gordon system in the continuum limit one arrives,[10,11] at the property rather rarely realized for lattice Bose systems:

$$\begin{aligned} &H_B \equiv H_{xyz} + (1-P)H_B(1-P), \\ &[H_B, P]_- = 0, \end{aligned} \tag{32}$$

which is in fact another version of the equivalence statement for the *xyz* and sine-Gordon models on the appropriate (a continuum limit of $P\mathscr{H}_N$) state space. The procedure of Ref. 11 with slight modifications can be repeated for the Bose MT, to prove that the formula (32) is valid in the continuum limit of the Bose MT. However, now the starting lattice Hilbert space of interest is $\mathscr{H}_{2N}$ and

$$\begin{aligned} P\mathscr{H}_{2N} &= P\prod_{i=1}^{N}{}^{\otimes}(h\otimes\tilde{h})_i := P\prod_{i=1}^{N}{}^{\otimes}(h_{2i-1}\otimes h_{2i}) \\ &= \prod_{i=1}^{N}{}^{\otimes}[(ph)_{2i-1}\otimes(ph)_{2i}] = \prod_{i=1}^{2N}{}^{\otimes}(ph)_i, \end{aligned} \tag{33}$$

where *p* is a two level projection of Ref. 10 in the single particle Hilbert space *h*.

If we start from the lattice CCR algebra generators associated with (18)

$$\begin{aligned} &[a_u, a_i^*] = \delta_{ui}, [\tilde{a}_u, \tilde{a}_i^*]_- = \delta_{ui}, a_u\Omega = 0 = \tilde{a}_u\Omega \ \forall k, \\ &[a_u^\#, \tilde{a}_i^\#]_- = 0 = [a_u, a_i]_- = [a_u^*, a_i^*]_-, \end{aligned} \tag{34}$$

the underlying projections are

$$\begin{aligned} P_u &= :\exp(-a_ua_u): + a_u^*:\exp(-a_u^*a_u):a_u \quad (\text{in } h), \\ \tilde{P}_u &= :\exp(-\tilde{a}_u\tilde{a}_u): + \tilde{a}_u^*:\exp(-\tilde{a}_u^*\tilde{a}_u):\tilde{a}_u \quad (\text{in } \tilde{h}), \\ P &= \prod_{u=1}^{N}(p_u\cdot\tilde{p}_u) \end{aligned} \tag{35}$$

and one easily checks that

$$\begin{aligned} &Pa_u^*P \equiv \sigma_u^+, \quad Pa_uP \equiv \sigma_u^-, \\ &P\tilde{a}_u^*P \equiv \tilde{\sigma}_u^+, \quad P_{\tilde{u}}P \equiv \tilde{\sigma}_u^-, \end{aligned} \tag{36}$$

determine the spin 1/2 SU(2) group generators for the linear chain of spins 1/2. Upon the change of labelling $\{\sigma_u^\pm, \tilde{\sigma}_u^\pm\}_{u=1,\ldots,N} \to \{\sigma_u^\pm\}_{u=1,\ldots,2N}, \{\sigma_{2i-1}^\pm, \tilde{\sigma}_u^\pm = \sigma_{2i}^\pm\}_{i=1,2,\ldots,N}$ being newly introduced, an application of the Jordan–Wigner transformation allows us to convert a $2N$ site spin 1/2 system into the $2N$ component Fermi system. This step was carefully investigated in Ref. 14.

ACKNOWLEDGMENT

I have received a copy of Ref. 6 prior to publication from Professor L. D. Faddeev to whom I am greatly indebted.

[1]A. G. Izergin and P. P. Kulish, Theor. Mat. Phys., 189 (1980).
[2]P. Garbaczewski, J. Math. Phys. **19**, 642 (1978).
[3]P. Garbaczewski, J. Math. Phys. **23**, 442 (1982).
[4]P. Garbaczewski, "Quantization of spinor fields. III. Fermions on coherent (Bose) domains," J. Math. Phys., (to be published) and Ann. Phys. N.Y. **139**, 293 (1982).
[5]R. Jackiw, Rev. Mod. Phys. **49**, 681 (1977).
[6]P. P. Kulish and E. K. Sklyanin, *Quantum Spectral Transform Method. Recent Developments*, Springer Lecture Notes in Physics, Vol. 151, edited by J. Hientarinta and J. Mickelsson (Springer-Verlag, Berlin, 1982).
[7]A. V. Mikhaĭlov, Pis'ma Zh. Eksp. Teor. Fiz. **23**, 356 (1976) [JETP Lett. **23**, 320 (1976)].
[8]E. A. Kuznetsov and A. V. Mikhaĭlov, Teor. Mat. Fiz. **30**, 323 (1977).
[9]E. K. Sklyanin, L. A. Tahtajan and L. D. Faddeev, Teor. Mat. Fiz. **40**, 194 (1979).
[10]P. Garbaczewski, J. Math. Phys. **22**, 574 (1981).
[11]P. Garbaczewski, "Mechanisms of the "fermion–boson reciprocity," J. Math. Phys, (to be published).
[12]H. Bergknoff and H. B. Thacker, Phys. Rev. D **19**, 3666 (1979).
[13]A. Luther, Phys. Rev. B **14**, 2153 (1976).
[14]M. Lüscher, Nucl. Phys. B **117**, 425 (1976).

An exactly solvable quantum field theory in three dimensions

E. C. G. Sudarshan and Xerxes Tata
Center for Particle Theory, Department of Physics, University of Texas at Austin, Austin, Texas 78712

(Received 18 May 1982; accepted for publication 30 July 1982)

We construct a quantum field theoretic model in three space dimensions and show that its spectrum can be exactly calculated. We also show that all the eigenvectors of the Hamiltonian can be obtained by a recursive procedure.

PACS numbers: 03.80. + r, 11.10.Ef, 11.20. − e

1. INTRODUCTION

Nonperturbative phenomena in quantum field theory are obviously of great importance. One such phenomenon, the formation of bound states, is an example of an unsolved problem almost as old as quantum field theory itself. The study of solvable quantum field theoretic models is of considerable interest since it may shed light on mechanisms responsible for some nonperturbative features of more realistic theories.

There exists considerable literature[1] on solvable quantum field theories in one space and one time dimension. To the best of our knowledge, there are no solvable models in three space and one time dimension. In this paper, we present a nonrelativistic (in the sense that negative frequencies are absent) field theory in three space dimensions for which the spectrum of the Hamiltonian can be exactly calculated. Moreover, we demonstrate that all the eigenstates of the Hamiltonian can be obtained by a recursive procedure. Thus, the S-matrix element for any process is, in principle, exactly calculable.

This paper is organized as follows. In the next section, we write down the Hamiltonian and the equations of motion for our model. It is noted that the vector space of states is a countable union of noncombining subspaces, each labeled by the values of two conserved quantum numbers, $\mathcal{N}_1$ and $\mathcal{N}_2$ (see Eqs. 2.3). In Sec. 3, it is shown that the spectrum of the Hamiltonian in any subspace is simply related to that in the subspaces with the same value of $\mathcal{N}_1$ but with $\mathcal{N}_2$ differing by one unit. In Sec. 4, it is shown that the spectrum of the Hamiltonian and the corresponding eigenstates in the family of subspaces with $\mathcal{N}_2 = 0$ can be readily obtained by explicitly solving the equations of motion. The complete spectrum can then be obtained from this by using the recursive procedure described in Sec. 3. In Sec. 5, we show that, for fixed $\mathcal{N}_1$, the eigenstates in the subspace labeled by $\mathcal{N}_2$ can be obtained from the corresponding states in the subspace labeled by $\mathcal{N}_2 - 1$. In doing so, we find that certain "eigenvalues" of the Hamiltonian (as obtained in Secs. 3 and 4) are spurious in that the corresponding eigenvector is null. We end with some concluding remarks in Sec. 6.

2. THE FERMIONIC LEE MODEL[2]

The model we consider consists of two fermion[3] fields N and θ interacting with a boson field V via a Yukawa-type interaction. It is assumed that the fields N and V are infinitely massive so that their energy is independent of the momentum. For simplicity, we assume the θ particle is massless. The Hamiltonian for the system is given by

$$H = m_0 V^\dagger V + \int d^3l\, l a^\dagger(l) a(l) + \int d^3l f(l)[V^\dagger N a(k) + a^\dagger(l) N^\dagger V]. \tag{2.1}$$

The quantization rules are,

$$\begin{aligned}
&\{N,N^\dagger\} = [V,V^\dagger] = 1,\\
&\{a(k),a^\dagger(l)\} = \delta(\vec{k}-\vec{l}),\\
&\{N,N\} = [V,V] = \{a(k),a(l)\} = 0,\\
&\{N,a(k)\} = \{N,a^\dagger(k)\} = [N,V] = [N,V^\dagger]\\
&\qquad = [a(k),V] = [a(k),V^\dagger] = 0,
\end{aligned} \tag{2.2}$$

together with their Hermitian conjugates.

The operators,

$$\mathcal{N}_1 = \mathcal{N}_V + \mathcal{N}_N, \tag{2.3a}$$

and

$$\mathcal{N}_2 = \mathcal{N}_\theta - \mathcal{N}_N, \tag{2.3b}$$

with

$$\mathcal{N}_V = V^\dagger V, \tag{2.4a}$$

$$\mathcal{N}_N = N^\dagger N, \tag{2.4b}$$

and

$$\mathcal{N}_\theta = d^3k\, a^\dagger(k) a(k), \tag{2.4c}$$

commute with the Hamiltonian. The vector space of states is, therefore, a countable union of disjoint subspaces labeled by the eigenvalues of $\mathcal{N}_1$ and $\mathcal{N}_2$.

We denote the eigenstates of the free Hamiltonian [the bilinear operator part of (2.1)] by $|\ \rangle$ whereas the eigenstates of the complete Hamiltonian are denoted by $|\ \rangle\rangle$. The phases of the free Hamiltonian eigenstates are defined by[4]

$$|V^a N\theta(k_1)\cdots\theta(k_b)\rangle \equiv (V^\dagger)^a a^\dagger(k_1)\cdots a^\dagger(k_b) N^\dagger|0\rangle, \tag{2.5a}$$

and

$$|V^a\theta(k_1)\cdots\theta(k_b)\rangle \equiv (V^\dagger)^a a^\dagger(k_1)\cdots a^\dagger(k_b)|0\rangle. \tag{2.5b}$$

For simplicity of notation, we shall label the exact eigenstates by their eigenvalues λ, suppressing all other labels that specify the state.

In the $\mathcal{N}_1 = a$, $\mathcal{N}_2 = b$ subspace (or, for short, the $V^a - \theta^b$ sector) there are just two[5] coupled amplitudes,

$$\psi_\lambda(k_1,\ldots,k_b) \equiv \langle\langle \lambda\, | V^a\theta(k_1),\ldots,\theta(k_b)\rangle \tag{2.6a}$$

and

$$\phi_\lambda(k_1,\ldots,k_{b+1}) \equiv \langle\langle \lambda\, | V^{a-1}N\theta(k_1),\ldots,\theta(k_{b+1})\rangle. \tag{2.6b}$$

The equations of motion satisfied by these can be readily obtained from the Hamiltonian. We find,

 0022-2488/83/071810-04$02.50

$$(\lambda - am_0 - k_1 \cdots - k_b)\psi_\lambda(k_1,k_2,\ldots,k_b)$$
$$= (-1)^b a \int d^3l f(l)\phi_\lambda(l,k_1,\ldots,k_b), \tag{2.7a}$$

and

$$[\lambda - (a-1)m_0 - k_1,\ldots,k_{b+1}]\phi_\lambda(k_1,\ldots,k_{b+1})$$
$$= (-1)^{b+1}[-f(k_1)\psi_\lambda(k_2,\ldots,k_{b+1})$$
$$+ f(k_2)\psi_\lambda(k_1,k_3,\ldots,k_{b+1})$$
$$+ \cdots(-)^i f(k_i)\psi_\lambda(k_1,k_2,\ldots,k_{i-1},k_{i+1},\ldots,k_{b+1}) + \cdots]. \tag{2.7b}$$

It is easily seen that for $a = b = 1$ these reduce to the equations of motion in the V-θ sector of the fermionic Lee model as given in Ref. 3.

Our purpose is to find those values of λ for which Eqs. (2.7) yield nontrivial solutions. To this end, we construct[6] the states $V^\dagger N|\lambda\rangle\rangle$ and $a^\dagger(k)|\lambda\rangle\rangle$ which reside in the subspace characterized by $(\mathcal{N}_1, \mathcal{N}_2 + 1)$ if the state $|\lambda\rangle\rangle$ was in the subspace labeled by $(\mathcal{N}_1, \mathcal{N}_2)$. By considering the action of the Hamiltonian on these new states, we are able to relate the spectral values in the $(\mathcal{N}_1, \mathcal{N}_2)$ subspace with those in the $(\mathcal{N}_1, \mathcal{N}_2 + 1)$ subspace. This forms the subject of the next section.

3. THE RELATIONSHIP BETWEEN THE SPECTRA IN THE $(\mathcal{N}_1,\mathcal{N}_2)$ AND $(\mathcal{N}_1,\mathcal{N}_2+1)$ SUBSPACES

As discussed at the conclusion of the previous section, we proceed by considering an eigenstate $|\lambda\rangle\rangle$ of the Hamiltonian in the particular subspace with $(\mathcal{N}_1, \mathcal{N}_2) = (a,b)$. It is clear that the states $a^\dagger(p)|\lambda\rangle\rangle$ and $V^\dagger N|\lambda\rangle\rangle$ belong to the subspace with $(\mathcal{N}_1, \mathcal{N}_2) = (a,b+1)$. The action of the Hamiltonian on these states can be readily calculated using Eqs. (2.1) and (2.2) to be

$$(H - \lambda - p)a^\dagger(p)|\lambda\rangle\rangle = f(p)V^\dagger N|\lambda\rangle\rangle \tag{3.1a}$$

and[7]

$$(H - \lambda - m_0)V^\dagger N|\lambda\rangle\rangle = \int d^3k f(k)a^\dagger(k)\mathcal{N}_1|\lambda\rangle\rangle$$
$$= a\int d^3k f(k)a^\dagger(k)|\lambda\rangle\rangle. \tag{3.2b}$$

If $|\mu\rangle\rangle$ is an eigenstate of the Hamiltonian in the $(a,b+1)$ sector, we easily obtain the equations of motion for the amplitudes

$$\sqrt{a}\langle\langle\mu|a^\dagger(p)|\lambda\rangle\rangle \equiv \eta_{\lambda\mu}(p) \tag{3.3a}$$

and

$$\langle\langle\mu|V^\dagger N|\lambda\rangle\rangle \equiv \sigma_{\lambda\mu}, \tag{3.3b}$$

to be

$$(\mu - \lambda - p)\eta_{\lambda\mu}(p) = \sqrt{a}f(p)\sigma_{\lambda\mu} \tag{3.4a}$$

and

$$(\mu - \lambda - m_0)\sigma_{\lambda\mu} = \sqrt{a}\int d^3k f(k)\eta_{\lambda\mu}(k). \tag{3.4b}$$

These are formally identical to the equations of motion in the lowest noninteracting sector of the Lee model,[2] with a coupling constant enhanced by a factor $\sqrt{a}$. It easily follows from Eqs. (3.4) that

$$\alpha_a(\mu - \lambda)\sigma_{\lambda\mu} = \sqrt{a}\int d^3k f(k)\delta(\mu - \lambda - k)$$
$$\equiv \sqrt{a}\tilde{f}(\mu - \lambda), \tag{3.5}$$

where

$$\alpha_a(z) \equiv z - m_0 - a\int d^3k \frac{f^2(k)}{z-k}. \tag{3.6}$$

It follows immediately that for μ to be in the spectrum of the Hamiltonian, σ_H, we must have either

(i) $\mu = \lambda + k$, $k > 0$, corresponding to the "scattering" solution[8] or

(ii) $\alpha_a(\mu - \lambda) = 0$, corresponding to the "discrete" solution.[8]

Before proceeding to analyze these further, we list some properties of the function α_a, regarded as a function of the complex variable z.

(i) $\alpha_a(z)$ is analytic in the z plane except for a cut along the positive real axis.

(ii) $\alpha_a(z)$ has no complex zeros.

(iii) For real, negative values of x, $\alpha_a'(x) > 0$ and $\alpha_a(-\infty) < 0$.

(iv) Again, for $x < 0$, $\alpha_a(x) - \alpha_1(x) \geqslant 0$.

In particular, if $\alpha_1(0^-) > 0$, $\alpha_a(0^-) > 0$ $\forall a$. This ensures a zero of $\alpha_a(x)$ if $\alpha_1(x)$ has a zero. The monotonicity of α_a ensures that the zero, if it exists, is unique. In this paper, we will assume that in the lowest nontrivial sector there is a discrete point in the spectrum at M_1, i.e., $\alpha_1(M_1) = 0$, $M_1 < 0$. It then immediately follows that for all a there exists M_a such that $\alpha_a(M_a) = 0$, with $M_a > M_b$ iff $a < b$.

We have thus shown that if $\lambda \in \sigma_H$, μ is not a solution to Eqs. (3.4) unless

$$\mu = \lambda + M_a, \tag{3.7a}$$

or

$$\mu = \lambda + k, \quad k > 0. \tag{3.7b}$$

In other words, $\mu \notin \sigma_H$ unless it is of the form (3.7). We will see in Sec. 5 that not all values of μ of the type (3.7) are in the spectrum. Before proceeding to do so, we first find the spectrum in the $(\mathcal{N}_1, \mathcal{N}_2) = (a,0)$ sector of the model. The remainder of the spectrum can be obtained using Eq. (3.7). We proceed to do so in the next section.

4. THE $(\mathcal{N}_1,\mathcal{N}_2) = (a,0)$ SUBSPACE

In the V^a sector of the model, the equations of motion (2.7) reduce to

$$(\lambda - am_0)\psi_\lambda = a\int d^3l f(l)\phi_\lambda(l), \tag{4.1a}$$

and

$$[\lambda - (a-1)m_0 - k]\phi_\lambda(k) = f(k)\psi_\lambda. \tag{4.1b}$$

Once again, carrying out the same manipulations that led to Eq. (3.5), we find that the spectrum consists of one discrete state with $\lambda = (a-1)m_0 + M_a$ and a continuum of scattering states with $\lambda = (a-1)m_0 + p$, $p \geqslant 0$. The corresponding (unnormalized) state vectors can easily be calculated. We find,

$$|B^{(a)}\rangle\rangle = |V^a\rangle + a\int \frac{d^3k f(k)}{M_a - k}|V^{a-1}N\theta(k)\rangle, \quad (4.2a)$$

for the discrete state and

$$|\lambda\rangle\rangle = \frac{\tilde{f}(\xi)}{\alpha_a^+(\xi)}|V^a\rangle + \int d^3k\left[\delta(\xi - k) + \frac{af(k)\tilde{f}(\xi)}{\xi - k + i\epsilon}\right]|V^{a-1}N\theta(k)\rangle, \quad (4.2b)$$

with $\xi = \lambda - (a-1)m_0$ for the continuum states. Here $\alpha_a^{\pm}(\xi) + \alpha_a(\xi \pm i\epsilon)$. The spectrum in this subspace, therefore, consists of

(i) a discrete point at $\lambda = (a-1)m_0 + M_a$

and

(ii) a continuum starting at $(a-1)m_0$ and extending to infinity.

It is interesting to notice that the discrete point in the spectrum occurs at $\lambda = (a-1)m_0 + M_a$. $(a-1)$ bare V particles act as mere spectators whereas only one of them is bound by the interaction, with an effective strength increased by a factor $\sqrt{a}$. This can also be seen from the wavefunction, since Eq. (4.2a) can be rewritten as[9]

$$|B^{(a)}\rangle\rangle = |V^{a-1}\rangle \otimes \left[\sqrt{a}\,|V\rangle + a\int d^3k \frac{f(k)}{M_a - k}|N\theta(k)\rangle\right]. \quad (4.3)$$

The presence of the other V particles merely enhances the effective coupling constant by the abovementioned factor.

This phenomenon of "limited interaction" can be naively understood if we recognize that the bare V particles can interact only in the presence of N particles. For more than one V particle to directly interact, more than one N particle would have to be present, which is not possible because of the fermionic nature of N coupled to the no-recoil structure of the Hamiltonian.

The spectrum in the subspace $(\mathcal{N}_1, \mathcal{N}_2) = (a,1)$ can be obtained from that in the $(a,0)$ subspace using Eqs. (3.7). In fact by repeating the procedure b times, we can obtain the spectrum in any arbitrary subspace. It is obvious that this procedure would lead to, among other things, eigenvalues of the form cM_a or "cM_a + continuum" with $c \neq 1$ in contradiction[10] with the considerations of the previous paragraph. The elimination of these spurious eigenvalues forms the subject of the next section.

5. THE ELIMINATION OF SPURIOUS SOLUTIONS

As we have pointed out in the last section, and also in our earlier papers,[3,6] it is the solution to the equations of motion for the Källén–Pauli amplitudes that yield nontrivial eigenvectors of the Hamiltonian. It is quite possible[3,6] to have nontrivial solutions to equations of motion for any auxiliary set of amplitudes which lead to trivial solutions for the Källén–Pauli system. It is, therefore, essential to check whether the solutions to Eqs. (3.4) really correspond to genuine eigenstates of the system. We proceed as follows.

For any eigenstate $|\lambda\rangle\rangle$ in the $\mathcal{N}_1 = a$ subspace, we define the state $|\mu\rangle\rangle$ by

$$|\mu\rangle\rangle = \int d^3p\, \eta_{\lambda\mu}(p)a^\dagger(p)|\lambda\rangle\rangle + \frac{\sigma_{\lambda\mu}}{\sqrt{a}}V^\dagger N|\lambda\rangle\rangle. \quad (5.1)$$

It is easily verified that

$$H|\mu\rangle\rangle = \mu|\mu\rangle\rangle, \quad (5.2)$$

when Eqs. (3.4) are satisfied. $|\mu\rangle\rangle$ is thus a new eigenstate. There are two types of eigenstates $|\mu\rangle\rangle$ that can be obtained from $|\lambda\rangle\rangle$ [See Eqs. (3.7)]. These are

(i) the "scattering" type with $\mu = \lambda + k$

and

(ii) the "discrete" type with $\mu = \lambda + M_a$.

The corresponding solutions are

$$\sigma_{\lambda\mu} = \sqrt{Z_a},$$

$$\eta_{\lambda\mu}(p) = \frac{\sqrt{a}f(p)\sqrt{Z_a}}{M_a - p}, \quad (5.3)$$

with

$$Z_a = \frac{1}{\alpha_a'(M_a)},$$

for the discrete type, and

$$\sigma_{\lambda\mu} = \frac{\sqrt{a}f(k)}{\alpha^+(k)},$$

$$\eta_{\lambda\mu}(p) = \delta(k-p) + \frac{\sqrt{a}f(p)\sqrt{Z_a}f(k)}{\alpha^+(k)[k - p + i\epsilon]}, \quad (5.4)$$

where

$$\mu = \lambda + k,$$

for the continuum type.

Following the same procedure as was used to obtain $|\mu\rangle\rangle$ from $|\lambda\rangle\rangle$, we now proceed to obtain the state $|\nu\rangle\rangle$ from $|\mu\rangle\rangle$. There are then three possibilities for ν, viz. $\nu = \lambda + 2M_a$, $\nu = \lambda + M_a + l$ and $\nu = \lambda + k + l$. We have calculated the state vector $|\nu\rangle\rangle$ in terms of the vector $|\lambda\rangle\rangle$ for the three cases. We find that for $\nu = \lambda + 2M_a$ the state vector vanishes on account of our choice statistics for the particles. This is completely in keeping with our earlier observation that "$2M_a$" could only arise due to a simultaneous interaction of two V^s which we had rule out earlier. For the state with $\nu = \lambda + k + M_a$, we find

$$
\begin{aligned}
|\nu\rangle\rangle = {} & \sqrt{a}\,\sqrt{Z_a}\int d^3q\,\frac{f(q)}{M_a-q}\,a^\dagger(q)a^\dagger(k)|\lambda\rangle\rangle \\
& + a^{3/2}\,\frac{f(k)}{\alpha^+(k)}\sqrt{Z_a}\int d^3p\,d^3q\,\frac{f(p)f(q)}{(M_a-q)(k-p+i\epsilon)}\,a^\dagger(q)a^\dagger(p)|\lambda\rangle\rangle \\
& + \sqrt{a}\,\sqrt{Z_a}\,\frac{f(k)}{\alpha^+(k)}\int d^3q\left[\frac{f(q)}{M_a-q}-\frac{f(q)}{k-q+i\epsilon}\right]a^\dagger(q)V^\dagger N|\lambda\rangle\rangle \\
& + \frac{\sqrt{Z_a}}{\sqrt{a}}\,V^\dagger N a^\dagger(k)|\lambda\rangle\rangle .
\end{aligned}
\tag{5.5}
$$

The same state is obtained whether we choose $\mu=\lambda+M_a$, $\nu=\mu+k$ or $\mu=\lambda+k$, $\nu=\mu+M_a$. Finally, for the state with $\nu=\lambda+k+l$, we have

$$
\begin{aligned}
|\nu\rangle\rangle = {} & a^\dagger(l)a^\dagger(k)|\lambda\rangle\rangle+\left[\frac{af(l)}{\alpha^+(l)}\int d^3q\,\frac{f(q)}{l-q+i\epsilon}\,a^\dagger(q)a^\dagger(k)|\lambda\rangle\rangle-(k\leftrightarrow l)\right] \\
& + \frac{a^2 f(k)f(l)}{\alpha^+(k)\alpha^+(l)}\int d^3p\,d^3q\,\frac{f(p)f(q)}{(l-q+i\epsilon)(k-p+i\epsilon)}\,a^\dagger(q)a^\dagger(p)|\lambda\rangle\rangle \\
& + V^\dagger\left[\frac{f(k)a^\dagger(l)}{\alpha^+(k)}-(k\leftrightarrow l)\right]N|\lambda\rangle\rangle \\
& + \frac{af(k)f(l)}{\alpha^+(k)\alpha^+(l)}\int d^3q\left[\frac{f(q)}{l-q+i\epsilon}-(k\leftrightarrow l)\right]a^\dagger(q)V^\dagger N|\lambda\rangle\rangle .
\end{aligned}
\tag{5.6}
$$

It is seen that $|\nu\rangle\rangle$ is antisymmetric under the change $(k\leftrightarrow l)$ as it should be.

We have thus seen that if $\lambda\in\sigma_H$, $\mu\in\sigma_H$ iff

$$\mu=\lambda+k,\quad k>0, \tag{5.7a}$$

or

$$\mu=\lambda+M_a, \tag{5.7b}$$

provided $\lambda\neq E+M_a$, $E\in\sigma_H$.

Since the spectrum in the $(\mathscr{N}_1,\mathscr{N}_2)=(a,0)$ sector is exactly known, and since any state in the $(\mathscr{N}_1,\mathscr{N}_2)=(a,b)$ sector can be reached by a sufficient number of applications of the operators $a^\dagger(p)$ and $V^\dagger(N)$. The spectrum in the $\mathscr{N}_1=a$ subspace thus consists of the points M_a, k, M_a+k, $M_a+k+l,\cdots$, where $k,l,\cdots,\geqslant 0$.

6. CONCLUDING REMARKS

In all renormalizable quantum field theories in three space dimensions, the scattering of fermions occurs via an exchange of a boson. In this paper, we have studied a considerably simplified version of one such theory and obtained the exact spectrum of the Hamiltonian. Furthermore, we have, in principle, obtained all the Hamiltonian eigenstates. The exact S matrix for this theory is, therefore, calculable. Although some features of our calculation were peculiar to the particular model (such as the "mass nonrenormalization" of a bare V particle in the presence of other V particles), all the results we obtained were in keeping with our intuitive expectations. Although all our considerations have been confined to a non-relativistic framework, it is hoped that some of the results obtained here may serve to elucidate some nonperturbative features of more realistic quantum field theories.

ACKNOWLEDGMENTS

This research was supported in part by a grant from the U. S. Department of Energy, Contract No. DE-AS05-76ER03992.

[1]A. Zamolodchikov and A. Zamolodchikov, Nucl. Phys. B **133**, 525 (1978) and references therein; R. Shankar and E. Witten, Nucl. Phys. B **141**, 349 (1978); for a review, see M. Karowski, Phys. Rep. **49**, 229 (1979). See also references therein.

[2]T. D. Lee, Phys. Rev. **95**, 1329 (1954); K. D. Friedrichs, Commun. Pure Appl. Math. **1**, 361 (1948); G. Källén and W. Pauli, K. Dan. Vidensk. Selsk. **30**, No. 7 (1955).

[3]X. Tata, E. C. G. Sudarshan, C. B. Chiu, and C. C. Chiang, Phys. Rev. D (in press).

[4]There can be at most one bare N particle since N is fermionic and does not carry momentum.

[5]This is so because the N content of the bare states is bounded by one. The corresponding situation for the Bosonic Lee model is much more complicated. For details of the solution in the two lowest sectors, see Ref. 6.

[6]M. Bolsterli, Phys. Rev. **166**, 1760 (1968); C. A. Nelson, J. Math. Phys. **13**, 1051 (1972); C. C. Chiang, C. B. Chiu, E. C. G. Sudarshan, and X. Tata, Phys. Rev. D (in press).

[7]For the case of the Lee model with all the fields quantized as bosons, the right-hand side of Eq. (3.2b) would read $\int d^2k f(k)a^\dagger(k)(N^\dagger N - V^\dagger V)|\lambda\rangle\rangle$. The reason why the spectrum of the fermionic Lee model is soluble is that an eigen-operator appears in place of $N^\dagger N - V^\dagger V$.

[8]The terms "scattering" solution and "discrete" solution are used in a loose sense. We mean the solutions corresponding to these eigenvalues would be the ones corresponding to the scattering and discrete solutions in the lowest nontrivial sector of the model.

[9]The factor $\sqrt{a}$ appears in front of $|V\rangle$ in the square bracket since $\langle V^a|V^a\rangle = a\|\,|V^{a-1}\rangle\otimes|V\rangle\|^2$.

[10]The fact that such spurious eigenvalues occur is not surprising in view of the considerations of Refs. 3 and 6. There, we had found solutions to equations of motion for the overcomplete auxiliary amplitudes for which there was no corresponding solution to the Källén–Pauli equations of motion. For the full dynamical content of the Lee model to be realized, the equations of motion for the overcomplete amplitudes had to be solved subject to the constraints discussed in the abovementioned papers. Equations (3.4) are akin to the overcomplete amplitude equations and hence it is not surprising to find solutions to these which do not correspond to any eigenstates of the system.

Dirac gravitational magnetic monopoles do not exist

D. K. Ross
Physics Department, Iowa State University, Ames, Iowa 50011

(Received 13 October 1982; accepted for publication 4 February 1983)

We show that the gravitational analog of the Dirac magnetic monopole does not exist for a gauge theory of gravitation with either a T_4 or an SL(2, C) $\otimes$ T_4 gauge group.

PACS numbers: 04.20.Cv, 14.80. – j, 04.20.Jb

I. INTRODUCTION

A number of papers have been written investigating the possibility of gravitational analogs of magnetic monopoles. In this paper, we will be interested in Dirac[1] type monopoles, not in the finite energy instantonlike 't Hooft[2]–Polyakov[3] monopoles also known as gravitational pseudoparticles. Dowker and Roche[4] were among the first to consider this possibility, and they exploited the analogy with electricity and magnetism. The idea that the NUT[5] solution to Einstein's field equations describes a gravitational dyon with both ordinary and "magnetic" mass has been considered by Newman and Demianski[6] and by Dowker.[7] The NUT parameter n_0' is often referred to as magnetic mass.[8] Misner[9] has analyzed NUT space in detail and has shown that it does not admit an interpretation without a periodic time coordinate. This acausal behavior precludes a single pole having a role classically. The NUT solution has a string of singularities in the metric, and this has more serious consequences, as we just mentioned, than the corresponding string of singularities in the vector potential in the electromagnetic case. Thus the NUT solution, while having many properties analogous to electromagnetic magnetic monopoles, still does not make a viable gravitational "magnetic monopole." The question then arises: Can we prove, in general, whether or not gravitational magnetic monopoles of the Dirac type can exist? We will show below that they cannot exist.

Ezawa and Tze[10] have written a very nice paper classifying Dirac monopoles, 't Hooft–Polyakov monopoles, and Nielsen–Olesen[11] vortices in terms of $\Pi_1(G)$, $\Pi_1(H)$, and $\Pi_1(G)$, respectively. G is the gauge group or structure group of the fiber bundle and H is an isotropy subgroup of G. Π_1 is the first homotopy group. We will rely heavily on Ezawa and Tze below, making appropriate modification so that their work can be applied in the context of a fiber bundle gauge theory of gravitation. To do this, we must first decide what gauge group is appropriate in the gravitational case. A certain variety exists in the literature. Utiyama[12] was the first to consider a gauge theory of gravitation. He used the Lorentz group as the gauge group, as did Sciama[13] later. A difficulty with Utiyama is that a manifold with curvature appears from the start. Kibble[14] extended this to the 10-parameter inhomogeneous Lorentz group. Other gauge groups include Gl(4) used by Yang[15] and SL(2,C) used by Carmeli.[16] Carmeli's theory yields the usual Eisntein field equations, but Yang's theory leads to field equations of higher order than those of general relativity, and hence to additional solutions which are apparently unphysical. Cho[17] has written down a very nice fiber bundle model using T_4 (the four-dimensional translation group) as the gauge group. This was also considered by earlier investigators.[18] This theory is consistent if spinor sources of the gravitational field are not included. Curvature is created naturally in this theory, not assumed *a priori*. If spinor sources are included, Cho[19] has also shown that the appropriate gauge group is SL(2, C) $\otimes$ T_4, which leads most naturally to the Einstein–Cartan theory, but can also give the Einstein theory although not as naturally. It is still an open question whether or not torsion plays a role in gravitation in the presence of spinor sources. In the following, for thoroughness, we shall consider the cases where T_4 and SL(2, C) $\otimes$ T_4 are the gauge groups.

One peculiarity of gravitation arises immediately if it is considered to be a gauge theory. In the T_4 case,[17] for example, the gauge group is not an internal symmetry group, but acts on space-time itself. In the fiber bundle formalism, this means that the gauge potential $B^{\,i}_{\ \mu}$, which is a connection in the fiber bundle, becomes identified as the nontrivial part of the vierbein fields

$$h_i^{\ \mu} = \delta_i^{\mu} + \mathscr{K} B_i^{\ \mu}, \tag{1}$$

which describe space-time itself. In order to apply the classification of Ezawa and Tze to gravity, we must generalize their work slightly (1) to apply to theories where G can act on the base space and (2) to apply to theories where G is not compact [T_4 and SL(2, C) $\otimes$ T_4 are not compact]. This latter provision is necessary since Ezawa and Tze assume and use compactness, which we do not have in the gravitational case.

We will show that the classification of Ezawa and Tze can be applied to T_4 and to SL(2, C) $\otimes$ T_4 gauge theories in Sec. II below. We will calculate the necessary homotopy groups in Sec. III, and we will finally complete our demonstration that gravitational "magnetic monopoles" of the Dirac type do not exist for a T_4 or SL(2, C) $\otimes$ T_4 gauge theory in Sec. IV.

II. APPLICATION OF THE WORK OF EZAWA AND TZE TO GRAVITATION

Let us look briefly at the proof of Ezawa and Tze[10] to see how to apply it in the case of a gravitational gauge theory. They take an element $U(x_0)$ in the field manifold and perform an equivalent transport of $U(x_0)$ along a loop l in the base space X to a point $x \in l$, getting

$$U_l(x) = T\,[\,g_l(x, x_0)\,]\;U(x_0), \tag{2}$$

 0022-2488/83/071814-03$02.50

where $T(G)$ is a certain holomorphic representation of the gauge group or symmetry group G, and

$$g_l(x, x_0) = P \exp\left(- ie \int_{x_0}^{x} dx_\mu \, A_\mu \right) \in G. \tag{3}$$

P is an ordering parameter along l. $g_l(x, x_0)$ draws a curve l^* in G, which is not necessarily a loop, as the loop in X is traversed. If l is swept over a 2-sphere S at a fixed time in X, l^* trace out a surface $g(s)$ in G with boundary $\partial g(s)$, a loop. R is defined as the subset of G, which is swept by all possible boundaries $\partial g(s)$. Thus with any sphere S, a member of the second relative homotopy group $\Pi_2(G,R,e)$ can be associated. Two such spheres enclose the same type of monopoles if they are mapped into the same homotopy class of $\Pi_2(G,R,e)$. Ezawa and Tze[10] then prove that monopoles are classified by the fundamental homotopy group $\Pi_1(R)$, using the exact homotopy sequence $\Pi_2(G) \rightarrowtail \Pi_2(G,R) \overset{\partial}{\rightarrowtail} \Pi_1(R)$. They go on to show that for Dirac monopoles, $R = G$ so that Dirac monopoles are classified by $\Pi_1(G)$.

Ezawa and Tze assume that G is compact, which is not true in the gravitational case. If we look at their proof that monopoles are classified by $\Pi_1(R)$, we see that compactness is not necessary, but only the weaker requirement that $\Pi_2(G) = 0$. Thus, for their work to apply to a gravitational gauge theory with either $G = T_4$ or $\mathrm{SL}(2, C) \circledS T_4$, we must show that

$$\Pi_2(T_4) = 0 \tag{4}$$

and

$$\Pi_2(\mathrm{SL}(2, C) \circledS T_4) = 0. \tag{5}$$

We calculate these in the next section.

A further stumbling block to the application of their proof to gravitation is the fact that in gravity, G also acts on the base space, space-time itself, through (1). We now want to show that this does not ruin their argument. We will look at the T_4 case. Similar arguments apply for $\mathrm{SL}(2, C) \circledS T_4$. Following Cho,[17] we shall use indices $\alpha, \beta = 1,2,3,4$ to refer to the structural group G with four commuting generators ξ_α. Indices $i,j,k = 1,2,3,4$ refer to four orthonormal vector fields e_i forming an orthonormal basis for space-time with commutation relations

$$[e_i, e_j] = T_{ij}{}^k e_k. \tag{6}$$

Finally $\mu, \nu = 1,2,3,4$ refer to a commuting coordinate basis for space-time. Using Cho's notation and explicitly including the gauge generators in (3) gives

$$g_l(x, x_0) = P \exp\left(- \int_{x_0}^{x} B_i^\alpha \xi_\alpha \, dx^i \right) \tag{7}$$

for the gravitational case. B_i^α is the gauge potential for the T_4 gauge theory, which is identified with the nontrivial part of the vierbein fields describing space-time as in (1). Using (1), we can write (7) as

$$g_l(x, x_0) = P \exp\left(- \frac{1}{\mathscr{K}} \int_{x_0}^{x} (h_i^\alpha - \delta_i^\alpha) \, \xi_\alpha \, dx^i \right). \tag{8}$$

Since a space-time vierbein now appears in $g_l(x, x_0)$, the question arises whether a closed loop in the base space can still lead to an open loop in l^* in G. If not, $\partial g(s) = 0$ and $R = \{0\}$ leading to a breakdown in the argument of Ezawa and Tze. We can now write (8) in terms of a coordinate basis, using the definition of a vierbein, as

$$g_l(x, x_0) = P \exp\left(- \frac{1}{\mathscr{K}} \int_{x_0}^{x} \xi_\alpha \, (dx^\alpha_{\text{curved}} - dx^\alpha_{\text{flat}} \right), \tag{9}$$

where $dx^\alpha_{\text{curved}}$ refers to curved space and dx^α_{flat} refers to flat space. The flat space term in (9), when integrated around a closed loop in the base space, always produces a closed loop in G leading to $\partial g(s) = 0$ and $R = \{0\}$. The curved space part of (9), when integrated around a closed loop in the base space, in general, will *not* produce a closed loop l^* in G, however. It is well known[20] that carrying a vector around a small parallelepiped in curved space gives

$$\Delta \xi^\alpha = R^\alpha{}_{\beta\mu\gamma} \, \xi^\beta \, dx^\eta \, d\hat{x}^\gamma \neq 0, \tag{10}$$

where $R^\alpha{}_{\beta\eta\gamma}$ is the curvature tensor. Thus the boundary $\partial g(s)$ is nonzero, in general, and the subset R of G is no longer trivial. The argument of Ezawa and Tze can thus be applied, even though G acts on space-time.

To summarize this section, gravitational "magnetic monopoles" are classified by $\Pi_1(T_4)$ [or $\Pi_1(\mathrm{SL}(2, C) \circledS T_4)$ in the more general case] *if* we can show that $\Pi_2(T_4) = \{0\}$ [or $\Pi_2(\mathrm{SL}(2, C) \circledS T_4) = \{0\}$]. We thus need to calculate these homotopy groups.

III. HOMOTOPY GROUPS

For a T_4 gauge theory, we need to calculate $\Pi_1(T_4)$ and $\Pi_2(T_4)$. For the more general $\mathrm{SL}(2, C) \circledS T_4$ gauge theory, we need $\Pi_1(\mathrm{SL}(2, C) \circledS T_4)$ and $\Pi_2(\mathrm{SL}(2, C) \circledS T_4)$. Let us look at the T_4 case first. This is easily handled using the following theorem[21]:

Theorem: If a space X is contractible by a homotopy that leaves x_0 fixed, then $\Pi_n(X, x_0) = \{0\}$ for each $n \geqslant 1$.

X is contractible if there is a point x_0 in X and a homotopy $H: X \times I \rightarrow X$ such that $H(x,0) = x$ and $H(x,1) = x_0$, where $x \in X$. In particular, the real line, Euclidean space of any dimension, an interval, and a convex figure in Euclidean space are all contractible spaces.[21] Now since T_4 is isomorphic to four-dimensional Euclidean space, we have that T_4 is contractible. Thus

$$\Pi_1(T_4) = \Pi_2(T_4) = \{0\}. \tag{11}$$

Turning to the $\mathrm{SL}(2, C) \circledS T_4$ gauge theory now, where $\circledS$ denotes the semidirect product, we need the following theorem[22]:

Theorem: Let X and Y be two given spaces and $x_0 \in X$, $y_0 \in Y$ be given points. Consider the product space $Z = X \times Y$ and the point $z_0 = (x_0, y_0)$ in Z. Then for every $n > 0$, we have $\Pi_n(Z, z_0) \approx \Pi_n(X, x_0) \otimes \Pi_n(Y, y_0)$, where $\otimes$ denotes the direct product and $\approx$ an isomorphism.

The proof of this theorem depends on using the obvious projections of loops in Z into loops in X and Y and, conversely, any pair of loops in X and Y determining a loop in Z. It is clear from the proof that $\mathrm{SL}(2, C) \circledS T_4$ is a product space in the sense of the theorem so that we can write

$$\Pi_n(\mathrm{SL}(2, C) \circledS T_4) \approx \Pi_n(\mathrm{SL}(2, C)) \otimes \Pi_n(T_4) \tag{12}$$

for $n > 0$.

We still require $\Pi_1(\mathrm{SL}(2, C))$ and $\Pi_2(\mathrm{SL}(2, C))$. Since SL(2, C) is simply connected, we have

$$\Pi_1(\mathrm{SL}(2, C)) = \{0\}. \tag{13}$$

Π_2 of all compact Lie groups is $\{0\}$,[23] and since SL(2, C) is compact, we further have

$$\Pi_2(\mathrm{SL}(2, C)) = \{0\}. \tag{14}$$

Putting (12) and (13) into (11) then gives

$$\Pi_1(\mathrm{SL}(2, C) \otimes T_4) = \{0\} \tag{15}$$

and

$$\Pi_2(\mathrm{SL}(2, C) \otimes T_4) = \{0\}. \tag{16}$$

IV. CONCLUSIONS

Using the results of the preceding section, we are now ready to put everything together. Since $\Pi_2(T_4) = \{0\}$ and $\Pi_2(\mathrm{SL}(2, C) \otimes T_4) = \{0\}$, the earlier arguments of Sec. II leads us to conclude that the classification of Ezawa and Tze can be applied to a gauge theory of gravitation based on a T_4 gauge group [or more generally, based on an SL(2, C) $\otimes$ T_4 gauge group]. Thus $\Pi_1(G)$ classifies the Dirac monopoles. However, we showed that $\Pi_1(T_4) = \{0\}$ and $\Pi_1(\mathrm{SL}(2, C) \otimes T_4) = \{0\}$. We thus conclude that Dirac-type gravitational "magnetic monopoles" do not exist. The only way out is if gravitation is based on a gauge group different from T_4 [or, more generally, SL(2, C) $\otimes$ T_4] with greatly different homotopy properties. This seems very unlikely.

[1] P. A. M. Dirac, Phys. Rev. **74**, 817 (1948); Proc. Roy. Soc. London Ser. A **133**, 60 (1931).
[2] G. 't Hooft, Nucl. Phys. B **79**, 276 (1974).
[3] A. M. Polyakov, Zh. Eksp. Teor. Fiz. Pisma Red. **20**, 430 (1974) [JETP Lett. **20**, 194 (1974)].
[4] J. S. Dowker and J. A. Roche, Proc. Phys. Soc. **92**, 1 (1967).
[5] E. Newman, L. Tamburino, and T. Unti, J. Math. Phys. **4**, 915 (1963).
[6] E. Newman and M. Demianski, Bull. Acad. Pol. Sci., Ser. Sci. Math. Astron. Phys. **14**, 653 (1966).
[7] J. Dowker, Gen. Relativity Gravitation **5**, 603 (1974).
[8] J. F. Plebański, Ann. Phys. (N.Y.) **90**, 196 (1975); J. F. Plebański and M. Demianski, Ann. Phys. (N.Y.) **98**, 98 (1976).
[9] C. W. Misner, J. Math. Phys. **4**, 924 (1963).
[10] Z. F. Ezawa and H. C. Tze, J. Math. Phys. **17**, 2228 (1976).
[11] H. B. Nielsen and P. Olesen, Nucl. Phys. B **61**, 45 (1973).
[12] R. Utiyama, Phys. Rev. **101**, 1597 (1956).
[13] D. W. Sciama, *Recent Developments in General Relativity* (Pergamon, New York, 1962).
[14] T. W. B. Kibble, J. Math. Phys. **2**, 212 (1961).
[15] C. N. Yang, Phys. Rev. Lett. **33**, 445 (1974).
[16] M. Carmeli, Nucl. Phys. B **38**, 621 (1972); Phys. Rev. Lett. **36**, 59 (1976).
[17] Y. M. Cho, Phys. Rev. D **14**, 2521 (1976).
[18] K. Hayashi and T. Nakano, Progr. Theoret. Phys. **38**, 491 (1967).
[19] Y. M. Cho, Phys. Rev. D **14**, 3335 (1976).
[20] R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1975), p. 160.
[21] F. H. Croom, *Basic Concepts of Algebraic Topology* (Springer-Verlag, New York, 1978), p. 112.
[22] S.-T. Hu, *Homotopy Theory* (Academic, New York, 1959), p. 144.
[23] M. E. Mayer, *Fiber Bundle Techniques in Gauge Theories* (Springer-Verlag, New York, 1977), p. 111.

Remarks on the Tomimatsu–Sato metrics[a)]

Cornelius Hoenselaers
Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 8046 Garching, Federal Republic of Germany
Frederick J. Ernst
Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616

(Received 8 April 1982; accepted for publication 11 March 1983)

After introducing a new way of writing the Tomimatsu–Sato solutions of Einstein's field equations, we consider the geometry in the neighborhood of the "poles." We also show that nonequatorial timelike and null geodesics can reach none of the ring singularities.

PACS numbers: 04.20.Jb

I. INTRODUCTION

Ever since the discovery of the Tomimatsu–Sato solutions[1,2] of Einstein's vacuum field equations, the structure of these solutions has puzzled those investigators who have dared to indulge in calculations with them.

The first Tomimatsu–Sato solution, which we shall designate by TS1, is the famous Kerr metric, which has been the subject of innumerable investigations. By comparison, very little has been done with any of the other TS solutions. However, a calculation[3] of the Weyl conform tensor invariants for TS2 led Ernst[4] and Economou[5] to consider exact vacuum solutions which approximate TS2 in the neighborhood of the "poles," the locations of apparent directional singularities in the metric. They found that these auxilliary solutions are of Petrov type D and that the poles are null surfaces rather than points. For TS3 the Weyl conform tensor invariants were evaluated by Hoenselaers,[6] but very little else is known about that solution.

The purpose of the present work is threefold: first, we shall cast the Tomimatsu–Sato solutions into a form which is more amenable to analysis than the original one; then, we shall address ourselves to the problem of the poles of TS3; and finally, we shall answer the long debated question concerning whether or not a timelike or null geodesic, not in the equatorial plane, can reach the ring singularities.

II. NEW FORM OF THE TS SOLUTIONS

In their original paper,[2] Tomimatsu and Sato wrote their series of solutions in the form

$$ds^2 = f^{-1}\left[\frac{A}{C}\left(\frac{dx^2}{x^2-1}+\frac{dy^2}{1-y^2}\right)+\rho^2\,d\phi^2\right] - f(dt-\omega\,d\phi)^2, \tag{2.1}$$

where $\rho^2 = (x^2-1)(1-y^2)$. The solutions correspond to Ernst potentials of the form

$$\epsilon = (\alpha-\beta)/(\alpha+\beta),$$

where α and β are polynomials of orders δ^2 and δ^2-1, respectively, in x and y and of order δ in p and q ($p^2+q^2=1$). Furthermore,

$$f = A/B,\quad A = \alpha\alpha^* - \beta\beta^*,$$
$$B = (\alpha+\beta)(\alpha^*+\beta^*),\quad C = p^{2\delta}(x^2-y^2)^{\delta^2-1}. \tag{2.2}$$

Some time ago we noticed that the even δ TS solutions can be written in the form

$$ds^2 = (B/C)[d\xi^2/(1+\xi^2)+d\eta^2/(1-\eta^2)] + B^{-1}[\xi^2\eta^2(\mu d\phi - \nu\,dt)^2 - (\sigma d\phi - \tau\,dt)^2], \tag{2.3}$$

while the odd δ TS solutions can be written in the form

$$ds^2 = (B/C)[d\xi^2/(1+\xi^2)+d\eta^2/(1-\eta^2)] + B^{-1}[\eta^2(\mu d\phi - \nu dt)^2 - \xi^2(\sigma d\phi - \tau dt)^2], \tag{2.4}$$

where

$$\xi^2 = x^2-1,\quad \eta^2 = 1-y^2,$$
$$B = \mu\tau - \nu\sigma,\quad C = p^{2\delta}(\xi^2+\eta^2)^{\delta^2-1}, \tag{2.5}$$

and where μ, ν, σ, and τ are polynomials in x and y. For TS2 these polynomials were listed in Ref. 4.

For TS3 one finds

$$\mu = ((px+1)/p)\{6(p^2\xi^8+q^2\eta^8)+32p^2\xi^4(\xi^2+1)\} + 4p\xi^6(3\xi^2+4)+\xi^2\tau, \tag{2.6a}$$

$$\nu = q\{p^2\xi^4(3\xi^4+8\xi^2\eta^2+6\eta^4)+q^2\eta^8\}, \tag{2.6b}$$

$$\sigma = 8q(px+1)\eta^4\{3\xi^2(\xi^2+\eta^2)+2(\xi^2+3\eta^2)-4\}\; 4q\xi^2\eta^4\{3(\xi^2+2\eta^2)-4\}+\eta^2\nu, \tag{2.6c}$$

$$\tau = p\{p^2\xi^8+q^2\eta^4(6\xi^4+8\xi^2\eta^2+3\eta^4)\}. \tag{2.6d}$$

From Eqs. (2.3) and (2.4), it is obvious that the degenerate metric induced on the "hypersurface" $\xi=0$ has Lorenzian signature for the even and Euclidean signature for the odd δ solutions.

III. THE "NORTH POLE" $x=1$, $y=1$ IN TS3

Considering the geometry of the TS3 solution near the "north pole" $x=1$, $y=1$, one is tempted to replace

$$\xi\to\lambda\xi,\quad \eta\to\lambda\eta,\quad t\to 32(p+1)t/(p^2\lambda^4), \tag{3.1}$$

and to rescale the metric by

$$ds^2\to[\lambda^2p^2/32(p+1)]ds^2, \tag{3.2}$$

a) This work supported in part by National Science Foundation grants PHY-75-08750 and PHY-79-08627.

and then to take the limit $\lambda \to 0$. In this way one arrives at the metric

$$ds^2 = (B/C)(d\xi^2 + d\eta^2) + B^{-1}\{\eta^2(\mu' d\phi - \nu dt)^2 - \xi^2(\sigma' d\phi - \tau dt)^2\}, \quad (3.3)$$

where

$$\mu' = p^3\xi^4, \quad \sigma' = -p^2 q\eta^4, \quad C = p^6(\xi^2 + \eta^2)^8, \quad (3.4)$$

and ν and τ are the same as in Eqs. (2.5b) and (2.5d). The auxiliary solution we obtain in this manner is actually the third in a series of nonasymptotically flat solutions given by Ernst.[7] The invariants of the Weyl tensor are given by

$$I_1 = 3(p^6(\xi^2 + \eta^2)^8/\beta^3)(p + iq)(p\xi^4 + iq\eta^4),$$
$$I_2 = 144(p^{12}(\xi^2 + \eta^2)^{15}/\beta^5)\xi^2\eta^2(p + iq)^2, \quad (3.5)$$

where

$$\beta = p^2\xi^6 - q^2\eta^6 + 3ipq\xi^2\eta^2(\xi^2 + \eta^2).$$

As in the case of the original TS3 metric, the invariants of the auxiliary metric tend to zero as ξ and η go to zero, regardless of the value of ξ/η.

On the other hand, the metric (3.3) cannot, in contrast to the $\delta = 2$ case, be regarded as a valid approximation to TS3, as here one has to perform the conformal rescaling (3.2).

Furthermore, the Hamilton–Jacobi equation for null geodesics with vanishing angular momentum becomes

$$p^2(\xi^2 + \eta^2)^8(S_{,\xi}{}^2 + S_{,\eta}{}^2) = p^2\xi^6 - q^2\eta^6,$$

which shows that no such geodesics can reach $\xi = 0$ for nonzero η. This is not what one would expect from our examination of the full TS3 metric.

To get an impression of how geodesics behave near the axis, we consider the equation of geodesic deviation:

$$\ddot{\zeta}^\alpha + \Gamma^\alpha_{\beta\gamma,\delta} u^\beta u^\gamma \zeta^\delta + 2\Gamma^\alpha_{\beta\gamma} u^\beta \dot{\zeta}^\gamma = 0 \quad (3.6)$$

for the deviation vector ζ^α, where the dot denotes the ordinary derivative with respect to the affine parameter of the geodesic, whose normalized tangent vector is u^α and along which the Γ's are evaluated.

For future use we want to keep the discussion fairly general. Therefore, we assume that the metric under consideration is of the form

$$ds^2 = a(dx^1)^2 + b\,(dx^2)^2 + dK^T g dK, \quad (3.7)$$

where

$$K = \begin{pmatrix} \phi \\ t \end{pmatrix}.$$

Furthermore, a, b, and the matrix g are functions of x^1 and x^2 only, symmetric under the reflection $x^2 \to -x^2$. Moreover, we take a geodesic with $x^2 = 0$ as reference geodesic. It is obvious from the symmetry of the metric that such geodesics exist.

Keeping in mind that under the above circumstances all first x^2 derivatives vanish in (3.6), the equation for $\zeta = \zeta^2$ becomes

$$\ddot{\zeta} - (1/2b)[a_{,22}v^2 + \dot{K}^T g_{,22}\dot{K}]\zeta + (\ln b)_{,1}\dot{\zeta}v = 0,$$

where $v = u^1$ and all expressions are evaluated at $x^2 = 0$. However, two of the conservation laws yield

$$g\dot{K} = k = \text{const}$$

and

$$av^2 + k^T g^{-1} k = \epsilon = 0, \pm 1. \quad (3.8)$$

We shall also use the resulting equation

$$\ddot{\zeta} - (1/2ab)[a\epsilon - ak^T g^{-1}k]_{,22}\zeta + (\ln b)_{,1}\dot{\zeta}v = 0 \quad (3.9)$$

in the following section.

We shall now consider null geodesics in the neighborhood of the axis by taking

$$k = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad x^1 = \xi, \quad x^2 = \eta, \quad \epsilon = 0.$$

Those geodesics have also been used by Szekeres and Morgan[8] for the analysis of the Curzon metric.

From the exact metric (2.3), one derives

$$ak^T g^{-1} k = -\frac{D_0\xi^6 - 9p^4q^2\xi^{16}\eta^2 + D_1}{p^6(\xi^2 + \eta^2)^8(1 + \xi^2)},$$

where $D_0 = D_0(\xi) = [32(1+p)p]^2 + O(\xi^2)$ and $D_1 = D_1(\xi, \eta) = O(\eta^4)$. Hence

$$\partial_{\eta\eta} ak^T g^{-1}k\,|_{\eta=0} \simeq 16[32(1+p)/(p^2\xi^6)]^2,$$

where the symbol $\simeq$ denotes the terms which diverge most rapidly as $\xi \to 0$. With

$$B \simeq 32p^4(1+p)\xi^{12},$$

and hence

$$b \simeq 32(1+p)/(p^2\xi^4),$$

Eq. (3.9) reduces to

$$\ddot{\zeta} - (4/\xi)\dot{\zeta}v + (8/\xi^4)\zeta = 0, \quad (3.10)$$

where we have kept only the most rapidly diverging terms. As

$$v^2 = \dot{\xi}^2 = -a^{-1}k^T g^{-1}k \simeq \xi^{-2},$$

we have

$$\ddot{\xi} \simeq -\xi^{-3}, \quad \dot{\zeta} = \zeta_{,\xi}\dot{\xi}, \quad \ddot{\zeta} = \zeta_{,\xi\xi}\dot{\xi}^2 + \zeta_{,\xi}\ddot{\xi},$$

and thus

$$\xi^2\zeta_{,\xi\xi} - 5\xi\zeta_{,\xi} + 8\zeta = 0. \quad (3.11)$$

The two independent solutions of Eq. (3.11) are

$$\zeta \propto \xi^2 \quad \text{and} \quad \zeta \propto \xi^4. \quad (3.12)$$

From $\dot{\xi} \simeq \xi^{-1}$ one concludes that $V = \xi^2$ is twice the affine parameter for small ξ. Now we use $\kappa = \eta/\xi^2$ and $V = \xi^2$ as coordinates and write the metric (2.4) in the form

$$ds^2 = P_1\,d\kappa^2 + (2P_2\kappa/V)d\kappa\,dV + (P_3/V^3)dV^2 + \kappa^2 N\,d\phi^2 - 2\kappa^2 V^2 L\,d\phi\,dt - V^3 M\,dt^2, \quad (3.13)$$

where

$$\pi = 32(1+p)/p^2, \quad \mu = q/p,$$
$$P_1 \simeq \pi,$$
$$P_2 \simeq \pi[1 + V(\tfrac{3}{2} - 8\kappa^2)],$$
$$P_3 \sim \pi\{\tfrac{1}{4} + V(\tfrac{1}{8} - \kappa^2) + V^2[\tfrac{1}{64} + \tfrac{1}{2}\kappa^2 + \kappa^4(1 + \tfrac{9}{4}\mu^2)] + V^3[(p-2)/128 + \tfrac{7}{16}\kappa^2 - \tfrac{13}{2}\kappa^4 + \kappa^6(6 - 5\mu^2)]\},$$
$$N \simeq \pi\{1 + \tfrac{3}{2}V + 9V^2(\tfrac{1}{16} - \kappa^4\mu^2)\}, \quad (3.14)$$
$$L \simeq 3\mu\{1 + 3\kappa^2 V - V^2[\tfrac{1}{2}\kappa^2 - \kappa^4(2 - 9\mu^2)]\},$$

$$M \simeq (1/\pi)\{1 - V(\tfrac{3}{2} + 9\mu^2\kappa^2) + 9V^2[\tfrac{3}{16} + \mu^2\kappa^2(\tfrac{3}{2} - 5\kappa^2)] - V^3[(p+54)/32 + \mu^2\kappa^2(\tfrac{243}{16} - 72\kappa^2 + 100\kappa^4 - 81\mu^2\kappa^4)]\}.$$

Following Ref. 4, we now perform a coordinate transformation

$$t = u + f(V, \kappa), \phi = \psi + g(V, \kappa), \tag{3.15}$$

where the functions f and g are to be chosen in such a way that the diverging terms with $d\kappa\, dV$, dV^2, $d\psi\, dV$, etc., are cancelled. In this way we obtain the metric

$$ds^2 = du\, dV + \pi(d\kappa^2 + \kappa^2\, d\psi^2) + h_{ij}\, dx^i\, dx^j. \tag{3.16}$$

h_{ij} is a tensor whose components go to zero at least as $o(V)$. Functions f and g which accomplish this are

$$f = \pi\{1/(4V^2) + (1/V)(\tfrac{1}{2} - \kappa^2) - \tfrac{3}{32}\ln V + 3\kappa^4 - 2\kappa^2 + V[(1-p)/64 - \tfrac{13}{16}\kappa^2 + 5\kappa^4 - 6\kappa^6 + \tfrac{1}{4}\kappa^6\mu^2]\},$$

$$g = -\tfrac{3}{2}\mu[\ln V + \tfrac{1}{2}V(2\kappa^2 - 1)].$$

An entirely analogous procedure is of course applicable to the "south pole" $x = 1, y = -1$. One can again, as in Ref. 4, identify the future of the south pole with the past of the north pole and vice versa. Thereby one arrives at a toruslike structure with the null geodesics threading through the opening of the doughnut, where $V = 0$. There the space-time becomes momentarily flat.

It is interesting to note that the metric near $V = 0$ is independent of the rotation parameter q. Furthermore, the part of the horizon $x = 1$ where $y \neq 1$ has been pushed out to $\kappa = \infty$ in our coordinates.

The whole structure of TS3 confirms earlier conclusions drawn from an investigation into the Voorhees metrics[9] and also agrees with the pictures we have of the Curzon metric, which the static limits of the TS solutions approach the large δ. We therefore believe that the poles of all Tomimatsu–Sato metrics are flat except for $\delta = 2$.

IV. GEODESICS NEAR THE SINGULARITY

As has been pointed out by Tomimatsu and Sato and confirmed by the calculation of the Weyl invariants, the TS metrics have curvature singularities where $\alpha + \beta$ vanishes. There are δ of those singularities, all lying in the equatorial plane $y = 0$. In the case of the even solutions, these are all naked singularties, while in the case of the odd δ solutions, at least the outermost ring singularity must lie outside of the possible event horizon at $x = 1$.

For the Kerr solution it is known[10] that geodesics off the equatorial plane $y = 0$ cannot reach the singularity. The derivation of this result was facilitated by the existence of a Killing tensor, which enabled Carter to solve the geodesic equations completely.

For the TS metrics with $\delta > 1$, for which no analogs of the Kerr–Killing tensor are known, it has been a long unanswered question whether off equatorial plane timelike or null geodesics can reach at least one of the ring singularities. Tomimatsu and Sato did show that geodesics confined to the equatorial plane reach the outermost ring singularity in a finite proper time.

We shall resolve the question concerning off equatorial plane geodesics by tackling the equation of geodesic deviation (3.9) with $x^1 = x$ and $x^2 = y$ in the neighborhood of one of the singularities, whose location we designate by $x = x_0, y = 0$.

From Eq. (2.2) it is clear that as the singularity is approached, A and B behave at least as $\xi = (x - x_0)$ and ξ^2, respectively. Then we may infer from Eq. (2.5) that the numerators of $g_{\phi\phi}$ and $g_{\phi t}$ also behave as ξ. From

$$\det g = -(x^2 - 1)(1 - y^2)$$

we find that $M = Bg$ becomes a singular matrix for small ξ. Furthermore, from Eq. (2.5) it follows that

$$(\ln b)_{,1} \simeq 2/\xi$$

and

$$(1/2ab)(a\epsilon - ak^T g^{-1} k)_{,22} \simeq c/\xi^4$$

with some constant c, while from Eq. (3.8) it follows that

$$v^2 = \dot{x}^2 = \dot{\xi}^2 = e^2/\xi^3$$

with another constant e, where, of course, k has to be chosen so that e^2 is positive. Now, again keeping only the dominant terms, we can write Eq. (3.9) in the form

$$\xi\zeta_{,\xi\xi} + \tfrac{1}{2}\zeta_{,\xi} + c'\zeta = 0.$$

With the *ansatz*

$$\zeta = \xi^m \sum_{n=0}^{\infty} a_n \xi^n,$$

the recursion relation for the a_n is found to be

$$a_n = -c' a_{n-1}[(n+m)(n+m-\tfrac{1}{2})]^{-1}$$

for $m = 0$ and $\tfrac{1}{2}$.

The general solution, a linear combination of the solutions belonging to the two m values, will miss the singularity unless the $m = 0$ part is absent.

Now let $y \simeq y_1\sqrt{\xi}$. Then we have from (3.7)

$$a\dot{\xi}^2 + b\dot{y}^2 + k^T g^{-1} k = \epsilon$$

with

$$a \simeq a_1\xi^2 + a_2 y^2,$$

$$b \simeq b_1\xi^2 + b_2 y^2,$$

$$k^T g^{-1} k \simeq (n_1\xi + n_2 y^2)/(d_1\xi^2 + d_2 y^2).$$

Consequently we find for the velocities

$$\dot{\xi} \simeq C + O(\xi), \quad \dot{y} \simeq (Cy_1/2\sqrt{\xi})[1 + O(\xi)],$$

if $\epsilon - k^T g^{-1} k = O(1)$ or

$$\dot{\xi} \simeq \hat{C}\sqrt{\xi} + O(\xi), \quad \dot{y} \simeq \hat{C}y_1/2 + O(\xi)$$

if $\epsilon - k^T g^{-1} k = O(\xi)$.

The y component of the geodesic equation reads

$$2b\ddot{y} + 2b_{,\xi}\dot{\xi}\dot{y} + b_{,y}\dot{y}^2 - a_{,y}\dot{\xi}^2 + (k^T g^{-1} k)_{,y} = 0.$$

By direct calculation of the leading order it can be shown that this equation cannot be satisfied for either of the above cases. Hence we conclude that no geodesic, save those confined to the equatorial plane, can reach the singularity. As our discussion was quite general and used only the dominant behavior of the various functions, the result holds for all Tomimatsu–Sato metrics.

V. ADDITIONAL REMARKS

As the complexity of the TS solutions increases tremendously with δ, we have found it necessary to perform some of the calculations with the help of a computer. To this end, one of us (CH) has developed the program POLYNOM, which is capable of performing algebraic operations including division and differentiation with polynomials. Even though the expressions (2.6) were first calculated by hand, a check by POLYNOM improved our confidence in them. The Weyl invariants of (3.3) were calculated entirely by computer.

[1]A. Tomimatsu and H. Sato, Phys. Rev. Lett. **29**, 1344 (1972).
[2]A. Tomimatsu and H. Sato, Prog. Theor. Phys. **50**, 95 (1973).
[3]J. E. Economou and F. J. Ernst, J. Math. Phys. **17**, 52 (1976).
[4]F. J. Ernst, J. Math. Phys. **17**, 1091 (1976).
[5]J. E. Economou, J. Math. Phys. **17**, 1095 (1976).
[6]C. Hoenselaers, Gen. Relativ. Gravit. **11**, 325 (1977).
[7]F. J. Ernst, J. Math. Phys. **18**, 233 (1977).
[8]P. Szekeres and F. H. Morgan, Commun. Math. Phys. **32**, 313 (1973).
[9]C. Hoenselaers, Prog. Theor. Phys. **59**, 1170 (1978).
[10]B. Carter, Phys. Rev. **174**, 1559 (1968).

Response to gravitational probes and induced Newton's constant

L. S. Brown and A. Zee
Department of Physics, University of Washington, Seattle, Washington 98195

(Received 7 December 1982; accepted for publication 4 February 1983)

We study the response of the action functional to external probes and derive a representation for Newton's constant without specializing to conformally flat space, as is normally done in standard discussions of induced gravity.

PACS numbers: 04.30. + x

Now that three of the four fundamental interactions have been unified, it is particularly urgent to understand how gravity is related to the other three interactions. The difficulty is, of course, that gravity looks quite different from the other three interactions, now known to be described by nonabelian gauge theories. In particular, while Yang–Mills theory is scale invariant in four-dimensional spacetime, the Hilbert–Einstein theory is not. An attempted forced marriage of gravity and the other interaction is attended by numerous unwanted guestinos, hitherto unobserved particles whose names end in "ino." An attractive alternative is that Einstein's theory of gravity actually represents a long distance effective phenomenological description.[1–4] In this case, Newton's constant G is determined by the other three gauge interactions. A formal representation for G was derived independently in Refs. 2 and 3:

$$\frac{1}{16\pi G} = -\frac{i}{96}\int (d^4x)\, x^2 \langle T^*[T(x)\, T(0)]\rangle. \tag{1}$$

Here the trace of the stress–energy tensor

$$T = T_\mu{}^\mu = \eta_{\mu\nu} T^{\mu\nu} \tag{2}$$

appears in the covariant time-ordered product $[T^*]$ taken in the flat-space vacuum expectation value. The flat-space metric is denoted by $\eta_{\mu\nu}$ which has the signature $(-\,+\,+\,+)$. In principle, this formula allows us to calculate Newton's constant.

The formula (1) holds in the approximation where gravity is treated as a classical field. Note that, as emphasized in Ref. 5, the appearance of the trace of the stress–energy tensor T rather than the stress tensor $T^{\mu\nu}$ itself is crucial since T is a soft operator. Calculations of Newton's constant G using formula (1) in various models have been attempted.[3,6] This formula has also been studied using considerations based on analyticity and positivity.[7]

One derives Eq. (1) by studying the response of the action functional to an external classical gravitational field

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + h_{\mu\nu}(x) \tag{3}$$

treating $h_{\mu\nu}$ as a small perturbation. In Refs. 2 and 3 $h_{\mu\nu}$ is chosen to be conformally flat, $h_{\mu\nu}(x) = \frac{1}{4}\eta_{\mu\nu} h(x)$, for arithmetical simplicity. The field $h(x)$ is coupled to $\eta_{\mu\nu}T^{\mu\nu}$, and this accounts for the appearance of T in Eq. (1) rather than $T^{\mu\nu}$. However, general coordinate invariance guarantees that physical results do not depend on what external probe one chooses to use. In this note, we shall derive formula (1) without restricting $g_{\mu\nu}$ to be conformally flat,[8] and we shall discuss the structure of the effective gravitational Lagrangian in some detail.

As a by-product of our discussion, we write down a representation for $K_{\mu\nu\lambda\rho}(x) = \langle 0|\, T^*[T_{\mu\nu}(x)\, T_{\lambda\rho}(0)]\,|0\rangle$ with the correct Schwinger terms included, which, as far as we know, has not been given in the literature. We thus add to the work of Boulware and Deser,[9] who, following Schwinger, analyzed years ago the commutator $[T_{\mu\nu}(x),\ T_{\lambda\rho}(0)]$.

As an interesting problem in applying the idea of induced gravity to cosmology one might wish to study the behavior of Newton's constant as a function of temperature. At finite temperature, Lorentz invariance is, of course, lost. The first step in the study would involve writing down the correct expression for $K_{\mu\nu\lambda\rho}$ at zero temperature and then generalizing to finite temperature.

The response of quantum fields to an arbitrary external field is described by the generating functional of connected Green's functions:

$$\begin{aligned} W[g] &= \frac{1}{2}\int (d^4x)\, h_{\mu\nu}(x)\, \langle T^{\mu\nu}(x)\rangle \\ &\quad + \frac{1}{2}\left(\frac{1}{2}\right)^2 \int (d^4x)(d^4y)\, h_{\mu\nu}(x)\, h_{\rho\sigma}(y) \\ &\quad \times \langle iT^*[T^{\mu\nu}(x)\, T^{\rho\sigma}(y)]\rangle + \cdots \\ &\equiv \frac{1}{2}\Lambda \int (d^4x)\, \eta^{\mu\nu} h_{\mu\nu}(x) \\ &\quad + \frac{1}{2}\left(\frac{1}{2}\right)^2 \int (d^4x)(d^4y)\, h_{\mu\nu}(x) \\ &\quad \times h_{\rho\sigma}(y)\, K^{\mu\nu\rho\sigma}(x-y) + \cdots . \end{aligned} \tag{4}$$

The generating functional is not altered by general coordinate transformations. This invariance is assured by requiring invariance under the infinitesimal transformations

$$\delta h_{\mu\nu} = \lambda^\sigma{}_{,\mu}(\eta_{\sigma\nu} + h_{\sigma\nu}) + \lambda^\sigma{}_{,\nu}(\eta_{\mu\sigma} + h_{\mu\sigma}) + \lambda^\sigma h_{\mu\nu,\sigma}. \tag{5}$$

Examining the terms which are linear in the gravitational field $h_{\mu\nu}$, we conclude that the two-point stress–tensor correlation function must obey the divergence condition

$$\partial_\nu K^{\mu\nu\rho\sigma}(x-y) = \Lambda\, [\eta^{\rho\sigma}\partial^\mu - \eta^{\mu\sigma}\partial^\rho - \eta^{\mu\rho}\partial^\sigma]\, \delta(x-y). \tag{6}$$

Hence we define

$$K^{\mu\nu\rho\sigma}(x-y)$$
$$=\bar{K}^{\mu\nu\rho\sigma}(x-y)$$
$$+\Lambda\,[\eta^{\mu\nu}\eta^{\rho\sigma}-\eta^{\mu\rho}\eta^{\nu\sigma}-\eta^{\mu\sigma}\eta^{\nu\rho}]\,\delta(x-y) \tag{7}$$

so that $\bar{K}^{\mu\nu\rho\sigma}$ is free of divergence,

$$\partial_\nu \bar{K}^{\mu\nu\rho\sigma}(x-y)=0. \tag{8}$$

Introducing the decomposition (7) into the generating functional expansion (4) gives

$$W[g]=\Lambda\int(d^4x)\left[\frac{1}{2}h_\mu{}^\mu-\frac{1}{4}h^{\mu\nu}h_{\mu\nu}+\frac{1}{8}h_\mu{}^\mu h_\nu{}^\nu\right]$$
$$+\frac{1}{2}\left(\frac{1}{2}\right)^2\int(d^4x)(d^4y)$$
$$\times h_{\mu\nu}(x)\,h_{\rho\sigma}(y)\,\bar{K}^{\mu\nu\rho\sigma}(x-y)+\cdots. \tag{9}$$

With the gravitational field $h_{\mu\nu}$ restricted to be slowly varying, we can identify this as the expansion of an effective local gravitational Lagrangian,

$$W[g]\to\int(d^4x)\sqrt{-g}\,\mathscr{L}_{\text{eff}}. \tag{10}$$

The terms in the square brackets in Eq. (9) are precisely the first few terms of the expansion of the determinantal factor $\sqrt{-g}$ in powers of $h_{\mu\nu}$. Hence

$$\mathscr{L}_{\text{eff}}=\Lambda+\cdots, \tag{11}$$

and we must identify Λ with the cosmological constant. We should note that, in view of Eq. (6), Λ can be computed from the stress-tensor correlation function as well as the simple single vacuum expectation value $\langle T^{\mu\nu}\rangle$.

To derive the Hilbert–Einstein piece of the effective Lagrangian, we need to first examine the kinematical structure of the correlation function or, equivalently, the structure of its Fourier transform

$$\bar{K}^{\mu\nu\rho\sigma}(k)=\int(d^4x)\,e^{-ikx}\,\bar{K}^{\mu\nu\rho\sigma}(x). \tag{12}$$

Taking account of the obvious symmetries of $\bar{K}^{\mu\nu\rho\sigma}(k)$, we may write its most general tensor structure as

$$\bar{K}^{\mu\nu\rho\sigma}(k)=\eta^{\mu\nu}\eta^{\rho\sigma}A(k^2)+(\eta^{\mu\rho}\eta^{\nu\sigma}+\eta^{\mu\sigma}\eta^{\nu\rho})\,B(k^2)$$
$$+(\eta^{\mu\nu}k^\rho k^\sigma+k^\mu k^\nu\eta^{\rho\sigma})\,C(k^2)$$
$$+(\eta^{\mu\rho}k^\nu k^\sigma+\eta^{\mu\sigma}k^\nu k^\rho$$
$$+\eta^{\nu\rho}k^\mu k^\sigma+\eta^{\nu\sigma}k^\mu k^\rho)\,D(k^2)$$
$$+k^\mu k^\nu k^\rho k^\sigma E(k^2). \tag{13}$$

The scalar functions $A,\ldots,E$ are devoid of kinematical singularities. The divergence condition

$$k_\mu\,\bar{K}^{\mu\nu\rho\sigma}(k)=0 \tag{14}$$

gives three constraints among the five scalar functions. Thus there are two independent combinations. The constraint equations are satisfied by $A=-k^2C$, $B=\frac{1}{2}k^2C$, $D=-\frac{1}{2}C$, and $E=0$. Hence there is a conserved tensor combination of order $k^\alpha k^\beta$ whose scalar coefficient is free of kinematical singularity. We introduce the notation

$$\Pi^{\mu\nu}=\eta^{\mu\nu}k^2-k^\mu k^\nu \tag{15}$$

and express this tensor as

$$L^{\mu\nu\rho\sigma}=(1/k^2)[\Pi^{\mu\rho}\Pi^{\nu\sigma}+\Pi^{\mu\sigma}\Pi^{\nu\rho}-2\Pi^{\mu\nu}\Pi^{\rho\sigma}]. \tag{16}$$

The combination of terms in the square brackets contain an overall factor of k^2, which is cancelled by the $1/k^2$ to produce a nonsingular tensor of order $k^\alpha k^\beta$. Thus $L^{\mu\nu\rho\sigma}$ is a tensor polynomial in k. There is no unique definition of the tensor which remains to complete the basis. It is, however, convenient to define this tensor by the requirement that it be traceless, and so we write

$$M^{\mu\nu\rho\sigma}=\Pi^{\mu\rho}\Pi^{\nu\sigma}+\Pi^{\mu\sigma}\Pi^{\nu\rho}-\tfrac{2}{3}\Pi^{\mu\nu}\Pi^{\rho\sigma}. \tag{17}$$

We now have

$$\bar{K}^{\mu\nu\rho\sigma}(k)=L^{\mu\nu\rho\sigma}K_1(k^2)+M^{\mu\nu\rho\sigma}K_2(k^2), \tag{18}$$

with $K_1(k^2)$ and $K_2(k^2)$ free of kinematical singularities. It is important to note that, with our choice of the tensor $M^{\mu\nu\rho\sigma}$, the scalar coefficient $K_1(k^2)$ is determined solely by the trace of the stress–energy tensor:

$$-12k^2K_1(k^2)+8\Lambda=\int(d^4x)\,e^{-ikx}\langle iT^*[T(x)\,T(0)]\rangle. \tag{19}$$

In particular, by setting $k_\mu=0$, one obtains a two-point representation of Λ.

In order to derive the effective low-energy gravitational Lagrangian, it is convenient to introduce the Fourier transform of the gravitational field and write

$$\int(d^4x)\int(d^4y)\,h_{\mu\nu}(x)\,h_{\rho\sigma}(y)\,\bar{K}^{\mu\nu\rho\sigma}(x-y)$$
$$=\int\frac{(d^4k)}{(2\pi)^4}h_{\mu\nu}(k)\,h_{\rho\sigma}(-k)\,\bar{K}^{\mu\nu\rho\sigma}(k). \tag{20}$$

Now to order h^2

$$\int\frac{(d^4k)}{(2\pi)^4}h_{\mu\nu}(k)\,L^{\mu\nu\rho\sigma}(k)\,h_{\rho\sigma}(k)$$
$$\simeq-8\int(d^4x)\sqrt{-g}\,R, \tag{21}$$

where R is the scalar curvature. Comparing with Eq. (9), we conclude that

$$\mathscr{L}_{\text{eff}}=\Lambda-K_1(0)\,R+\cdots, \tag{22}$$

and hence

$$1/16\pi G=-K_1(0). \tag{23}$$

Expanding the exponential in the integrand in Eq. (19) gives

$$12K_1(0)=\frac{1}{8}\int(d^4x)\,x^2\,\langle iT^*[T(x)\,T(0)]\rangle, \tag{24}$$

which, combined with Eq. (23), yields the formula (1) for Newton's constant.

Expanding Eq. (20) to higher powers of k yields higher order terms in the effective Lagrangian. Again examining the terms of order h^2, we find that

$$\int \frac{(d^4k)}{(2\pi)^4} h_{\mu\nu}(k)\, k^2 L^{\mu\nu\rho\sigma}(k)\, h_{\rho\sigma}(-k)$$

$$\simeq \int (d^4x)\sqrt{-g}\,(4C^2_{\mu\nu\rho\sigma} - \tfrac{4}{3}R^2) \tag{25}$$

while

$$\int \frac{(d^4k)}{(2\pi)^4} h_{\mu\nu}(k)\, M^{\mu\nu\rho\sigma}\, h_{\rho\sigma}(-k)$$

is proportional to

$$\int (d^4x)\sqrt{-g}\, C^2_{\mu\nu\rho\sigma},$$

where $C_{\mu\nu\rho\sigma}$ is the conformally invariant Weyl tensor. Thus we secure the Eddington–Weyl contribution to the effective Lagrangian

$$\mathscr{L}_{\text{eff}} = \cdots + \gamma C^2_{\mu\nu\rho\sigma} + \rho R^2 + \cdots, \tag{26}$$

where γ is proportional to $K_1'(0)$ and $K_2(0)$ and

$$\rho = -\tfrac{1}{6} K_1'(0), \tag{27}$$

with

$$K_1'(0) = \frac{\partial}{\partial k^2} K_1(k^2)\Big|_{k^2=0}. \tag{28}$$

In Ref. 5 it was argued that if ρ is set equal to zero in the classical Lagrangian, a calculable and finite value of ρ is induced since it is related to a matrix element of the soft operator $T(x)$. On the other hand, the coefficient γ will in general be divergent. The formal representation for γ involves the hard dimension four operator $T^{\mu\nu}$.

The structure of the stress-tensor correlation function can be exhibited in more detail if we write Källen–Lehmann representations for the scalar functions:

$$K_{1,2}(k^2) = \int_0^\infty ds\, \frac{\sigma_{1,2}(s)}{k^2+s}. \tag{29}$$

Here we shall proceed in a formal manner and disregard any possible convergence difficulties. Writing

$$K_1(k^2) = \int_0^\infty \frac{ds}{s}\sigma_1(s) - k^2\int_0^\infty \frac{ds}{s}\frac{\sigma_1(s)}{k^2+s}, \tag{30}$$

and, performing some rearrangement of terms, we obtain

$$\begin{aligned} K^{\mu\nu\rho\sigma}(k) &= M^{\mu\nu\rho\sigma}\int_0^\infty ds\,\frac{\sigma_2(s)-\sigma_1(s)/2}{k^2+s} \\ &\quad + \Pi^{\mu\nu}\Pi^{\rho\sigma}\int_0^\infty ds\,\frac{(4/3s)\,\sigma_1(s)}{k^2+s} \\ &\quad + L^{\mu\nu\rho\sigma}\int_0^\infty \frac{ds}{s}\sigma_1(s) \\ &\quad + \Lambda\,[\eta^{\mu\nu}\eta^{\rho\sigma} - \eta^{\mu\rho}\eta^{\nu\sigma} - \eta^{\mu\sigma}\eta^{\nu\rho}]. \end{aligned} \tag{31}$$

The traceless part corresponds to the spin-2 intermediate state contribution. Hence

$$\rho_2(s) = \sigma_2(s) - \sigma_1(s)/s \tag{32}$$

is the spectral weight for these states, while

$$\rho_0(s) = (4/3s)\,\sigma_1(s) \tag{33}$$

is the spectral density of the spin zero intermediate states.

One could clearly continue this analysis. For instance, one could examine the effective interaction of four gravitons. By extracting the appropriate coefficient, one could represent $1/G$ as an integral over the four-point function $\langle T^*[T_{\mu\nu}(z)\, T_{\lambda\rho}(y)\, T_{\sigma\tau}(x)\, T_{\eta\omega}(0)]\rangle$. General coordinate invariance guarantees an infinite hierarchy of identities.

[1] A. Sakharov, Dokl. Akad. Nauk SSSR **177**, 70 (1967) [Sov. Phys. Dokl. **12**, 1040 (1968)].

[2] S. L. Adler, Phys. Rev. Lett. **44**, 1569 (1980); Phys. Lett. B **95**, 241 (1980); *Proceedings of the 1980 Erice School*, edited by A. Zichichi (to be published); Rev. Mod. Phys. **54**, 729 (1982).

[3] A. Zee, Phys. Rev. Lett. **42**, 417 (1979); **44**, 703 (1980); Phys. Rev. D **23**, 858 (1981); Phys. Rev. Lett. **48**, 295 (1982); Phys. Lett. B **109**, 183 (1982); in *Grand Unified Theories and Related Topics*, Proceedings of the 1981 Kyoto School (World Scientific Publishing, 1981); in *Proceedings of the 1981 Erice School* (to be published).

[4] For earlier work and further references, see the reviews cited in Refs. 2 and 3.

[5] A. Zee, Phys. Lett. B **109**, 183 (1982).

[6] B. Hasslacher and E. Mottola, Phys. Lett. B **95**, 237 (1980).

[7] N. Khuri, Phys. Rev. Lett. **49**, 513 (1982); Phys. Rev. D **26**, 2664, 2671 (1982).

[8] This question was addressed but not resolved by S. Adler in his Erice lectures (Ref. 2). Our analysis was done some time ago and our results have already been summarized by A. Zee in his Erice lectures (Ref. 3), and cited by S. Adler in his review in Rev. Mod. Phys. (Ref. 2).

[9] D. Boulware and S. Deser, J. Math. Phys. **8**, 1468 (1967).

Null cone computation of gravitational radiation

R. A. Isaacson
Physics Division, National Science Foundation, Washington, D.C. 20550
J. S. Welling and J. Winicour[a)]
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 10 June 1982; accepted for publication 20 August 1982)

The production of gravitational waves is explored, both analytically and numerically, using a null cone formulation of axially symmetric gravitational and matter fields. The coupled field equations are written in an integral form, on a single conformally compactified patch, which is well suited for numerical computation. Some analytic and numerical solutions of the initial value problem are given. The total mass and radiation flux is studied in detail for a special class of collapsing dust configurations.

PACS numbers: 04.30. + x, 02.40. + m

I. INTRODUCTION

Recent experimental progress in the search for gravitational radiation has been enormously exciting. The beautiful results obtained from the binary pulsar PSR1913 + 16 firmly establish the existence of gravitational radiation reaction and quantitatively confirm the detailed predictions of general relativity.[1] During the past decade, the energy sensitivity of resonant Weber-bar gravitational wave receivers has improved by a factor of a million. While similar improvements in performance may still be needed for direct observation of radiation from natural astronomical sources, there is no shortage of promising new techniques and dedicated experimentalists willing to work toward achieving the goal.[2] Recently, prototypes of broad-band laser interferometric receivers have shown such promise of delivering detailed waveform information that serious consideration is being given to their use in creating a new astronomical window during the 1990s.[3]

Strong pulsed gravitational wave bursts, which are likely to be detected first, can only be emitted from astronomical objects with intense gravitational fields and large internal velocities. Prediction of the pulse structure from these sources demands solution to the coupled nonlinear equations of relativistic gravitation and of hydrodynamics and extraction of the elusive part representing gravitational radiation. Although theorists have been working on this formidable problem since 1918, it has now taken on renewed importance with the prospect of experimental data for comparison in the near future.

Most theoretical studies of this problem over the past 50 years can be grouped into three distinct approaches: perturbation methods,[4] asymptotic analysis of space–time,[5–9] and numerical solution via large-scale computers.[10] Perturbation methods have provided detailed formal estimates of radiation for a wide variety of processes, but questions of convergence obscure their application to describe realistic astronomical situations. Asymptotic techniques provide the only rigorous way to characterize radiation. However, they are generally unable to supply any link between sources and emerging waves. Computer techniques offer great promise for the future. So far, however, they have been applied to evolving Cauchy data for the motion of matter within a finite size box centered within a space–time. The numerical estimates of radiation made in this region are, at present, heuristic and without error bounds or basis for systematic improvement. Nevertheless, numerical and perturbative results are often in good qualitative and even quantitative agreement despite extrapolations beyond the expected realms of validity of the schemes used.

The purpose of this paper is to describe an approach to calculating radiation which is conceptually clear and well founded and which leads to surprisingly simple and numerically accurate techniques. The key ingredients of this program combine numerical and asymptotic methods. These include: (1) formulation of the problem in terms of the evolution of axisymmetric data specified on an initial null cone; (2) using conformal techniques to map the entire null hypersurface into a finite coordinate patch with one edge representing null infinity; (3) use of the field equations to find the rate of change of the data out of the initial hypersurface; and (4) direct computation of total energy and radiation flux from the fields found at null infinity. The difficult problem of combining the gravitational evolution with the matter hydrodynamics is left for the future and will not be attempted here. (This task is somewhat of a "black art" requiring highly specialized techniques of its own.[11])

In Sec. II, the mathematical formulation of the characteristic initial value problem is discussed. Previous results for vacuum[5] are extended to space–time with perfect fluids. In comparison to the usual Cauchy problem,[12] the procedure is enormously simplified. From the initial density, velocity, and one metric component (in the axisymmetric case), the full metric as well as the time derivatives of the matter and field variables are constructed sequentially from a miraculous hierarchy of explicit equations. At each stage, these formulae only involve differentiation and integration of known data (initial data or fields found at earlier steps), and are well suited to economical numerical evaluation.

Section III describes the asymptotic behavior of the gravitational field at large distance from matter. The powerful method of linkage integrals[8,13] is used to find the system's energy and radiation flux at null infinity.

a) Supported in part by the NSF Grant PHY-800823.

Section IV applies these tools to initial data corresponding to anisotropic dust models with vanishing gravitational shear. Exact analytic solutions to the field equations on the initial hypersurface are obtained. Radiation flux, total energy, and Newtonian comparisons are discussed.

Section V describes the numerical techniques used to solve the field equations. These are tested against the models of Sec. IV, and their accuracy is ascertained. Numerical models for dust distributions are explored in cases too complicated for analytic treatment.

II. NULL CONE EVOLUTION

Our prime assumption is the existence of a timelike geodesic whose points have future null cones which extend to null infinity without caustics or crossovers. For our purpose, the study of gravitational radiation, it is not essential that this timelike geodesic be complete. A geodesic segment will suffice. Thus the formation of an event horizon is not excluded, but our investigation would be restricted to the exterior asymptotically flat region. However, as later examples will illustrate, our treatment can extend to the interior of a particle horizon. Although our assumption encompasses a wide class of interesting astrophysical systems, it notably does not apply to a system such as two extremely separated stars, in a scattering state. In this case all null cones would undergo focussing. Thus, while we allow the existence of strong gravitational fields, we do rule out systems with "large quadrupole moments."

We use this timelike geodesic as the origin of a null coordinate system. In the presence of a hypersurface-orthogonal axisymmetry, the construction proceeds along the lines of Ref. 5, except that boundary conditions are imposed at the origin rather than null infinity. As a retarded time coordinate for the outgoing null cones, we take the proper time $x^0 = u$ along the timelike geodesic. A luminosity distance is chosen as the radial coordinate $x^1 = r$, so that surfaces of constant u and r have area $4\pi r^2$. Two ray labels x^A $(A = 2,3)$, which are constant along the outgoing null geodesics, complete the coordinate system. The line element then takes the form

$$\begin{aligned} ds^2 = (Vr^{-2}e^{2\beta} - U^2r^2e^{2\gamma})\,du^2 + 2e^{2\beta}\,du\,dr \\ + 2Ur^2e^{2\gamma}\,du\,d\theta - r^2(e^{2\gamma}\,d\theta^2 \\ + e^{-2\gamma}\sin^2\theta\,d\phi^2), \end{aligned} \tag{2.1}$$

where we choose $x^A = (\theta,\phi)$. For boundary conditions at the origin, we require that $t = u + r, x = r\sin\theta\cos\phi, y = r\sin\theta\sin\phi$, and $z = r\cos\theta$ define a smooth, local Fermi coordinate system, in which the metric takes the Minkowski form at $r = 0$. This requires $V = r + O(r^3), \beta = O(r^2)$, $U = O(r)$, and $\gamma = O(r^2)$. In addition, smoothness of the axis requires that $U/\sin\theta$ and $\gamma/\sin^2\theta$ be continuous at θ equal 0 and π.

As matter source, we consider an ideal fluid described by $T_{\mu\nu} = (\rho + p)w_\mu w_\nu - pg_{\mu\nu}$, with 4-velocity normalized by $w^\mu w_\mu = 1$. Axisymmetry implies $w_3 = 0$. The matter variables must also satisfy smoothness conditions at the origin, in the above Fermi coordinate system. In particular, in our null polar coordinates, this requires $w_1 = D - E\cos\theta + O(r)$ and $w_2 = Er\sin\theta + O(r^2)$, with $D^2 - E^2 = 1$. E vanishes for a matter flow with reflection symmetry about the equatorial plane. Also, for smoothness of the axis, $w_2/\sin\theta$ must be continuous.

In order to simplify the analysis of the characteristic initial value problem consider the tensor field $H_{\mu\nu} \equiv G_{\mu\nu} + 8\pi T_{\mu\nu}$. (We use unit $G = c = 1$.) Einstein's equation, $H_{\mu\nu} = 0$, implies the following:

hypersurface equations, $H_{1\nu} = 0$, (2.2)

gravitational evolution equations,

$$H_{AB} - \tfrac{1}{2}g_{AB}\,g^{CD}H_{CD} = 0, \tag{2.3}$$

matter evolution equations,

$$H_\mu{}^\nu{}_{;\nu} \equiv 8\pi T_\mu{}^\nu{}_{;\nu} = 0. \tag{2.4}$$

We now establish that the converse is also true. The proof follows from writing out the matter evolution equations in the form

$$H_\mu{}^\nu{}_{;\nu} = (-g)^{-1/2}[(-g)^{1/2}H_\mu{}^\nu]_{,\nu} + \tfrac{1}{2}g^{\alpha\beta}{}_{,\mu}H_{\alpha\beta}$$

and then dropping terms which vanish because of the hypersurface and gravitational evolution equations. For $\mu = 1$, this gives $g^{CD}H_{CD} = 0$. For $\mu = A$, it gives $(r^2H_{A0})_{,1} = 0$, from which $H_{A0} = 0$ follows by smoothness at the origin. In the same way, for $\mu = 0$, we obtain $H_{00} = 0$. Thus the hypersurface and evolution equations are equivalent to Einstein's equation. Note that this proof depends neither on the axisymmetry nor the ideal fluid model assumed in this paper.

The hypersurface equations are constraints which are intrinsic to each null hypersurface of constant u, i.e., no u derivatives appear. In terms of the Ricci tensor, they take the form

$$R_{11} = -8\pi T_{11} = -8\pi(\rho + p)w_1w_1, \tag{2.5a}$$

$$R_{12} = -8\pi T_{12} = -8\pi(\rho + p)w_1w_2, \tag{2.5b}$$

$$\begin{aligned} g^{AB}R_{AB} &= 8\pi(T - g^{AB}T_{AB}) \\ &= 8\pi[\rho - p + r^{-2}(\rho + p)e^{-2\gamma}(w_2)^2], \end{aligned} \tag{2.5c}$$

where, in terms of metric variables,

$$-\tfrac{1}{4}rR_{11} = \beta_{,1} - \tfrac{1}{2}r(\gamma_{,1})^2, \tag{2.6a}$$

$$\begin{aligned} -2r^2R_{12} = [r^4e^{2(\gamma-\beta)}U_{,1}]_{,1} - 2r^2[r^2(r^{-2}\beta)_{,12} \\ - \sin^{-2}\theta(\sin^2\theta\gamma)_{,12} + 2\gamma_{,1}\gamma_{,2}], \end{aligned} \tag{2.6b}$$

$$\begin{aligned} -r^2e^{2\beta}g^{AB}R_{AB} \\ = 2V_{,1} + \tfrac{1}{2}r^4e^{2(\gamma-\beta)}(U_{,1})^2 - (r^2\sin\theta)^{-1}(\sin\theta\,r^4U)_{,12} \\ + 2e^{2(\beta-\gamma)}[-1 + (\sin\theta)^{-1}(\sin\theta\beta_{,2})_{,2} - \gamma_{,22} \\ - 3\gamma_{,2}\cot\theta + (\beta_{,2})^2 + 2\gamma_{,2}(\gamma_{,2} - \beta_{,2})]. \end{aligned} \tag{2.6c}$$

Given an equation of state $p(\rho)$, the unconstrained initial data consists of the matter variables ρ, w_1, w_2, and the single metric variable γ, which determines the metric g_{AB} intrinsic to the surfaces of constant u and r. From this data, the remaining metric variables β, U, and V follow by solving the hypersurface equations. This task is enormously simplified since these potentially coupled partial differential equations actually form a hierarchy of purely radial ordinary differential equations, which may be solved by simple quadrature! The boundary conditions at the origin require that all integration constants vanish. Once the metric is found, the remaining matter variable w_0 is determined by normalization.

The gravitational evolution equations then determine the time derivative of the metric data γ, from information totally within the hypersurface. Because of axial symmetry, these equations contain only the one independent condition

$$R_{22} - e^{4\gamma} R_{33}/\sin^2\theta + 8\pi(\rho + p)(w_2)^2 = 0,$$

which gives

$$\begin{aligned}4r(r\gamma)_{,01} = {} & [2r\gamma_{,1}V - r^2(2\gamma_{,2}U + U_{,2} - U\cot\theta)]_{,1} \\ & - 2r^2(\sin\theta)^{-1}(\gamma_{,1}U\sin\theta)_{,2} + \tfrac{1}{2}r^4 e^{2(\gamma-\beta)}(U_{,1})^2 \\ & + 2e^{2(\beta-\gamma)}[(\beta_{,2})^2 + \beta_{,22} - \beta_{,2}\cot\theta \\ & + 4\pi(\rho + p)(w_2)^2]. \end{aligned} \tag{2.7}$$

The matter evolution equations determine the time derivatives of ρ, w_1, and w_2. By forming appropriate linear combinations they may also be cast into a simple hierarchy. To accomplish this, introduce null vectors $k_\alpha \equiv u_{,\alpha}$ and $l_\alpha \equiv 2w_\alpha - (k^\beta w_\beta)^{-1}k_\alpha$. To fix $\rho_{,0}$ we use

$$\begin{aligned}0 &= T_\alpha{}^\beta{}_{;\beta} l^\alpha \\ &= (\rho - p)_{;\alpha} w^\alpha + (k^\lambda w_\lambda)^{-1} p_{;\alpha} k^\alpha \\ &\quad + (\rho + p)[w^\alpha{}_{;\alpha} - (k^\lambda w_\lambda)^{-1} w_{\alpha;\beta} k^\alpha w^\beta].\end{aligned}$$

The only time derivatives which survive arise from the first term, giving

$$\begin{aligned}0 = {} & (1 - v_s{}^2)\rho_{,\alpha} w^\alpha - (w_1)^{-1} p_{,1} + (\rho + p)[w^1{}_{,1} \\ & + w^2{}_{,2} + 2(\beta_{,1} + r^{-1})w^1 + (2\beta_{,2} + \cot\theta)w^2 \\ & - (w_1)^{-1}(w_{1,1}w^1 + w_{1,2}w^2 + \tfrac{1}{2}g^{\alpha\beta}{}_{,1} w_\alpha w_\beta)],\end{aligned} \tag{2.8}$$

where $v_s \equiv (\partial p/\partial \rho)^{1/2}$ is the speed of sound. Next $w_{1,0}$ may be found by projecting along w^α

$$0 = T_\alpha{}^\beta{}_{;\beta} w^\alpha = (\rho + p)w^\alpha{}_{;\alpha} + \rho_{;\alpha} w^\alpha.$$

This may be put into the form

$$\begin{aligned}0 = {} & w_{1,0} + r^{-1}(r^2 e^{2\beta} w^1)_{,1} + (\sin\theta)^{-1} \\ & \times(\sin\theta\, e^{2\beta} w^2)_{,2} + (\rho + p)^{-1} e^{2\beta} \rho_{,\alpha} w^\alpha.\end{aligned} \tag{2.9}$$

Finally, $w_{2,0}$ may be found by using this condition to obtain

$$\begin{aligned}0 &= (\delta_\alpha{}^\tau + l^\tau w_\alpha) T_\tau{}^\beta{}_{;\beta} \\ &= (\rho + p) w^\beta [w_{\alpha;\beta} - (k^\lambda w_\lambda)^{-1} w_{\tau;\beta} k^\tau w_\alpha] \\ &\quad + (k^\lambda w_\lambda)^{-1} p_{;\beta} k^\beta w_\alpha - p_{;\alpha}.\end{aligned}$$

This gives $w_{2,0}$ in terms of $w_{1,0}$ from

$$0 = w_{[1} w_{2],\beta} w^\beta + \tfrac{1}{2} w_{[1} g^{\alpha\beta}{}_{,2]} w_\alpha w_\beta - (\rho + p)^{-1} w_{[1} p_{,2]}. \tag{2.10}$$

In summary, given γ, ρ, w_1, and w_2 on the initial hypersurface and an equation of state, we may determine β, U, V, and w_0 within this surface and then compute $\gamma_{,0}$, $\rho_{,0}$, $w_{1,0}$, and $w_{2,0}$, thus determining the evolution to the next hypersurface. This entire process only involves differentiation and radial integration of known functions within the hypersurface.

III. ASYMPTOTIC PROPERTIES

According to the scheme of Sec. II, the asymptotic behavior of the initial data determines all asymptotic properties. In order that the resulting space–time be asymptotically flat at future null infinity we require of this data that $\gamma = K + r^{-1}c + O(r^{-2})$ and the ρ be of compact support. The shear of a null hypersurface of constant u, which is proportional to $\gamma_{,1}$, then has the minimal r^{-2} falloff consistent with asymptotic flatness.

By integration of the hypersurface equations, the asymptotic behavior of the data leads to

$$\beta = H - c^2/(4r^2) + O(r^{-4}), \tag{3.1a}$$

$$\begin{aligned}U = {} & L + 2e^{2(H-K)}H_{,2}/r - [c\sin^2\theta\, e^{2(H-K)}]_{,2}/ \\ & (r^2\sin^2\theta) + O(r^{-3}),\end{aligned} \tag{3.1b}$$

$$\begin{aligned}V = {} & r^2(L\sin\theta)_{,2}/\sin\theta + re^{2(H-K)} \\ & \times[1 + 2(H_{,2}\sin\theta)_{,2}/\sin\theta + 4(H_{,2})^2 \\ & - 4H_{,2}K_{,2} - 2(K_{,2})^2 \\ & + K_{,22} + 3K_{,2}\cot\theta] - 2e^{2H}M + O(r^{-1}).\end{aligned} \tag{3.1c}$$

Furthermore, integration of the evolution equation for γ gives

$$\gamma_{,0} = K_{,0} + r^{-1}c_{,0} + O(r^{-2}), \tag{3.2}$$

where

$$K_{,0} = -\tfrac{1}{2}e^{-2K}\sin\theta(e^{2K}L/\sin\theta)_{,2}. \tag{3.3}$$

Thus the evolution equations are consistent with the initial asymptotic conditions.

The resulting asymptotic flatness is best exhibited by introducing a conformal geometry[7] $d\tilde{s}^2 = \Omega^2 ds^2$, with $\Omega = 1/r$. This leads to a smooth manifold with boundary, where the boundary I (future null infinity) consists of points $\Omega = 0$. Choosing $l = 1/r$ as a new radial coordinate, the line element takes the following form at I:

$$\begin{aligned}d\tilde{s}^2 = {} & -L^2 e^{2K} du^2 - 2e^{2H} du\, dl \\ & + 2Le^{2K} du\, d\theta - e^{2K} d\theta^2 - e^{-2K}\sin^2\theta\, d\phi^2.\end{aligned} \tag{3.4}$$

As required by asymptotic flatness, I is a null hypersurface, with null geodesics tangent to $n^{\mu'} = (1,0,L,0)$, in terms of the chart $x^{\mu'} = (u,l,\theta,\phi)$. This description of I differs, in two respects, from a standard conformal Bondi frame description,[8] in which the line element (3.4) would have vanishing L, H, and K. First, the angular coordinates are not constant along the null generators of I. This reflects the difference between specification of local inertial frames at the origin and at infinity. In the present treatment, θ and ϕ are pinned down by Fermi propagation at the origin and are not the natural Bondi coordinates (θ_B, ϕ_B) for I. For the same reason, the surfaces of constant u do not describe a Bondi slicing of I. Second, Ω is not a Bondi conformal factor, i.e., the intrinsic 2-geometry of cross sections of I is not that of a unit sphere and is not even cross-section-independent. These two features of the present formalism are unavoidable in a global treatment. They lead to no substantial problems as long as care is taken to describe asymptotic properties in terms of truly physical quantities, which are both coordinate-independent and Ω-independent. However, certain technical complications do arise, as described below.

One such set of physical quantities are the linkages, which describe the total energy–momentum, supermomentum, and angular momentum at a given retarded time.[8,13] They are linear representations of the BMS asymptotic symmetry group, whose generators $\xi^{\mu'}$ satisfy, at I,

$$\Omega^2 \mathcal{L}_\xi\, g_{\mu'\nu'} = 0.$$

We are particularly interested here in the Bondi energy–momentum associated with BMS translations. At I, these translation generators have the form $\xi^{\mu'} = An^{\mu'}$. In a standard conformal Bondi frame, A must be a linear combination of $l = 0$ and $l = 1$ spherical harmonics. In the present case, A can be determined by finding the transformation from the $x^{\mu'}$ chart with conformal factor $\Omega = l$ to a standard Bondi conformal frame. In this way, we find $A = \alpha e^{-2H}/\omega$, where α is a combination of $l = 0$ and $l = 1$ harmonics of the angular variables (θ_B, ϕ_B), with $\phi_B = \phi$ and

$$\frac{d\theta_B}{\sin\theta_B} = \frac{e^{2K}(d\theta - L\,du)}{\sin\theta}, \tag{3.5}$$

and where $\omega = e^K \sin\theta_B/\sin\theta$. Notice that the integrability condition for such a θ_B follows from the relationship (3.3) between $K_{,0}$ and L.

In order to evaluate the Bondi energy–momentum, the translation generators must first be extended to a neighborhood of I. Although this may be done quite arbitrarily,[13] the simplest extension in a null coordinate system is via the null hypersurface propagation law[8]

$$[\xi^{(\mu;\nu)} - \tfrac{1}{2}\xi^{\rho}{}_{;\rho}\, g^{\mu\nu}]u_{;\nu} = 0,$$

which in the x^μ chart reduces to

$$\xi^0{}_{,1} = 0, \tag{3.6a}$$

$$\xi^A{}_{,1} = -e^{2\beta} g^{AB}\xi^0{}_{,B}, \tag{3.6b}$$

$$\xi^1 = -\tfrac{1}{2} r[(\xi^A \sin\theta)_{,A}/\sin\theta - U\xi^0{}_{,2}]. \tag{3.6c}$$

These equations uniquely determine ξ^μ in terms of its value at I. The component of Bondi energy momentum associated with ξ^μ is then given by

$$P = (1/4\pi)\oint (\xi^{[\nu;\mu]} - \xi^{\rho}{}_{;\rho} B^{\mu\nu})\, dS_{\mu\nu}, \tag{3.7}$$

where $B^{\mu\nu} = u^{;[\mu} r^{;\nu]}/u^{;\alpha} r_{;\alpha}$ is the normalized bivector orthogonal to the integration surface. The integral is to be evaluated, in the limit $r \to \infty$, over a cross section of I with constant u. This limit may be calculated using the asymptotic expressions (3.1) for the metric variables and the formulae (3.6) for the derivatives of ξ^μ. In the final result, all θ derivatives of $\xi^0 = \alpha(\theta_B)e^{-2H}/\omega$ may be removed by a parts integration. After a straightforward but lengthy calculation, we find for the "α component" of energy–momentum:

$$\begin{aligned} P[\alpha] = (1/8\pi)\oint \alpha\omega^{-1} e^{-2K}\{ & 2e^{2K}M - 2c + 3c_{,2}\cot\theta \\ & + c_{,22} - 4c_{,2}(H+K)_{,2} - c[4(H_{,2})^2 \\ & - 8H_{,2}K_{,2} - 4(K_{,2})^2 + 2(H+K)_{,22} \\ & + 6(H+K)_{,2}\cot\theta]\}\sin\theta\, d\theta\, d\phi. \end{aligned} \tag{3.8}$$

In a standard Bondi frame, this reduces to Bondi's original expression[5] for the total energy E, for which we set $\alpha = 1$, as an integral of the mass aspect M [introduced in Eq. (3.1c)],

$$E = (4\pi)^{-1}\oint M \sin\theta\, d\theta\, d\phi.$$

Another asymptotic quantity of physical interest is the gravitational flux of energy–momentum to null infinity. This may be obtained by applying the divergence theorem to the linkage integrals (3.7). However, in order to put the resulting expression in a useful form, unwieldy manipulations using Einstein's equation would be necessary. It is preferable to begin with a conformal space flux expression in which these manipulations have already been carried out. By translating the expression given in Ref. 13 back into physical space, we obtain for the energy–momentum flux through a region dV of I

$$F[\alpha]\, dV = \frac{1}{32\pi}\alpha e^{2H}\omega[X_{\mu\nu}X^{\mu\nu} - \tfrac{1}{2}X^2]r^2 \sin\theta\, d\theta\, d\phi\, du, \tag{3.9}$$

where $X_{\mu\nu} = \mathscr{L}_\tau g_{\mu\nu}$, $X = g^{\mu\nu}X_{\mu\nu}$, and τ^μ is a BMS time translation with $\alpha = 1$. This flux expression is to be evaluated in the limit $r \to \infty$. As in the calculation of $P[\alpha]$, this limit may be evaluated using the asymptotic expressions for the metric variables (3.1) and (3.2) and Eqs. (3.6) for the BMS generator τ^μ, with $\tau^0 = e^{-2H}/\omega$. In this way, we obtain

$$\begin{aligned} F[\alpha]\, dV = (1/16\pi)(\alpha/\omega)e^{-2H}\{ & 2c_{,0} + 2c_{,2}L \\ & + cL_{,2} + cL\cot\theta + e^{-2K}\omega\sin\theta \\ & \times[(e^{2H}\omega)_{,2}(\omega^2\sin\theta)^{-1}]_{,2}\}^2 \\ & \times \sin\theta\, d\theta\, d\phi\, du. \end{aligned} \tag{3.10}$$

In a standard Bondi frame, the resulting energy flux leads to Bondi's original equation[5]

$$\frac{dE}{du} = -\frac{1}{4\pi}\oint (c_{,0})^2 \sin\theta\, d\theta\, d\phi.$$

An interesting feature, in the case of a highly compact matter distribution, is the existence of a particle horizon crossing the initial null cone and its concomitant white hole. (In the time-reversed version of our model, for which the initial cone would extend to past null infinity, this would correspond to an event horizon.) A sufficient, although not necessary, criterion for such a horizon is the existence, on the initial cone, of an $r =$ const surface which is antitrapped, i.e., whose two orthogonal sets of future null directions diverge. (By construction, the future directed set lying in the initial null cone automatically diverges.) In terms of our present formalism, this antitrapped surface condition takes the form

$$V < r^2(\sin\theta\, U)_{,2}/\sin\theta, \tag{3.11}$$

for some value of r. When this condition is satisfied, a singularity must exist somewhere in the past of the initial cone, although the geometry of the initial cone remains completely regular.

IV. RADIATIVE MODELS WITH SHEAR-FREE DATA

As a first approach to the general problem outlined in the previous sections, we now consider the mathematically simplest initial data which leads to the production of gravitational waves by matter. For the gravitational data, we set $\gamma = 0$ so that the initial null cone is shear-free. For the matter, we choose pressure-free dust which is initially "at rest" in the sense that its 4-velocity w_μ satisfies $\omega_1 = 1$, $w_A = 0$. In the Minkowski-space case, this corresponds exactly to a 4-velocity field, on the initial cone, which is comoving with the vertex world line. The initial data then consists entirely of the density ρ, with Minkowski space resulting from the choice $\rho = 0$.

The gravitational hypersurface and evolution equations, (2.5), and (2.7), now take the simplified integral form

$$\beta = 2\pi \int_0^r \rho r \, dr, \tag{4.1}$$

$$U = 2 \int_0^r dr \, r^{-4} e^{2\beta} \int_0^r s^4 (s^{-2} \beta_{,2})_{,s} \, ds, \tag{4.2}$$

$$\begin{aligned} V = \int_0^r dr \, \{ & -4\pi r^2 e^{2\beta} \rho - e^{2\beta} \\ & \times [-1 + \beta_{,2} \cot\theta + \beta_{,22} + (\beta_{,2})^2] \\ & - \tfrac{1}{4} r^4 e^{-2\beta} (U_{,1})^2 + (2 \sin\theta)^{-1} \\ & \times [r^{-2} \sin\theta \, (r^4 U)_{,1}]_{,2} \}, \end{aligned} \tag{4.3}$$

$$\begin{aligned} r\gamma_{,0} = \int_0^r dr \Big\{ & \frac{e^{2\beta}}{2r} [\beta_{,22} - \beta_{,2} \cot\theta + (\beta_{,2})^2] \\ & + \frac{r^3 e^{-2\beta}}{8} (U_{,1})^2 - \frac{\sin\theta}{4r} \left[\frac{(r^2 U)_{,1}}{\sin\theta} \right]_{,2} \Big\}. \end{aligned} \tag{4.4}$$

Furthermore, the asymptotic quantity K now vanishes so that the conformal factor ω, which determines the transformation to a standard conformal frame, equals 1. The Bondi energy and the energy flux then reduce to

$$E = (4\pi)^{-1} \oint M \sin\theta \, d\theta \, d\phi \tag{4.5}$$

and

$$F = (4\pi)^{-1} e^{-2H} [c_{,0} + \sin\theta \, (e^{2H} H_{,2} / \sin\theta)_{,2}]^2, \tag{4.6}$$

where we have set $F[\alpha] \, dV|_{\alpha=1} = F \sin\theta \, d\theta \, d\phi \, du$. It is convenient to introduce the relativistic internal energy $E_I = \int \rho w^\alpha \, dV_\alpha$, which is conserved for a dust model. In the present case, $E_I = \int \rho r^2 \sin\theta \, dr \, d\theta \, d\phi$ on the initial cone.

Gravitational radiation arises when the density is not spherically symmetric. Remarkably, there exists such a density distribution for which the gravitational equations (4.1)–(4.4) can be integrated analytically. This case, described below in model A, provides an important check on our numerical program for more complicated distributions. In addition, it provides valuable physical insight into the properties of Bondi energy and radiation. It turns out that the Newtonian binding energy, for the Newtonian analog of this special model, cannot be integrated analytically. For this reason, we also consider model B, consisting initially of a homogeneous spheroid, whose Newtonian binding energy is well known. We carry out an expansion, in powers of density, of the general relativistic model which shows, in the weak field case, that its binding energy differs from the Newtonian value, except for spherical distributions.

This feature stems from the physical meaning of shear-free data.[14] In the absence of matter, the shear comprises the entire data on the initial null cone, with vanishing shear determining Minkowski space. Thus, in the vacuum case, the simplest choice of data from a mathematical standpoint—zero shear—leads to the simplest interpretation from a physical standpoint—no incoming radiation. However, in the presence of matter, a shear free initial cone no longer corresponds to the absence of incoming waves, unless the matter distribution on the cone is also spherically symmetric. This can be understood by considering a collapsing sphere of dust with Schwarzschild exterior. Here gravitational radiation is clearly absent but the bending of light by matter introduces shear on a null cone whose vertex is displaced from the center of symmetry. Resetting this shear to zero is thus equivalent to introducing incoming radiation.

A. An exact model

We now present the analytic solution of the gravitational hypersurface and evolution equations (4.1)–(4.4) for an initial dust distribution described by the thick shell

$$\rho = \begin{cases} 0, & r < R, \\ k/(8\pi r), & R \leqslant r \leqslant \lambda(\theta) R, \\ 0, & r > \lambda(\theta) R. \end{cases}$$

Here k is a constant which determines the density scale, and $\lambda(\theta)$ describes the angular dependence of the outer boundary, which breaks spherical symmetry. The β integral (4.1) then becomes

$$\beta = \begin{cases} 0, & r \leqslant R, \\ k(r - R)/4, & R \leqslant r \leqslant \lambda R, \\ kR(\lambda - 1)/4, & r \geqslant \lambda R, \end{cases}$$

so that $H = kR(\lambda - 1)/4$. Note that $\beta_{,12}$ has δ-function behavior described by the integral

$$\int_{\lambda R^-}^{\lambda R^+} f\beta_{,12} \, dr = \tfrac{1}{4} k \lambda_{,2} R f|_{\lambda R},$$

for a test function f. When this is taken into account, the U integral (4.2) gives

$$U = \begin{cases} 0, & r \leqslant \lambda R, \\ -2e^{2H} H_{,2} \left(\dfrac{1}{3\lambda R} - \dfrac{1}{r} + \dfrac{2\lambda^2 R^2}{3r^3} \right), & r \geqslant \lambda R. \end{cases}$$

In the V integral (4.3), the δ-function contributions from $\beta_{,22}$ and $U_{,12}$ exactly cancel. We obtain

$$V = \begin{cases} r, & r \leqslant R, \\ 2R - 4k^{-1} - e^{k(r-R)/2}(r - 4k^{-1}), & R \leqslant r \leqslant \lambda R, \end{cases}$$

and, for $r \geqslant \lambda R$,

$$\begin{aligned} V = & \frac{2r^2 e^{2H}}{3\lambda R} \left[-H_{,22} - H_{,2} \cot\theta - 2(H_{,2})^2 + \frac{H_{,2}\lambda_{,2}}{\lambda} \right] \\ & + r e^{2H} [1 + 2H_{,22} + 2H_{,2} \cot\theta + 4(H_{,2})^2] - 2e^{2H} M \\ & + \frac{2e^{2H} \lambda^2 R^2}{3r} \Big[H_{,22} + H_{,2} \cot\theta - 4(H_{,2})^2 \\ & + \frac{2H_{,2}\lambda_{,2}}{\lambda} \Big] + \frac{4e^{2H} \lambda^4 R^4 (H_{,2})^2}{3r^3}, \end{aligned}$$

where the mass aspect is given by

$$\begin{aligned} M = \lambda R \, [& 1 + H_{,22} + H_{,2} \cot\theta + \tfrac{4}{3}(H_{,2})^2] \\ & + \lambda_{,2} H_{,2} R + (2/k)(e^{-2H} - 1) - Re^{-2H}. \end{aligned}$$

In the $\gamma_{,0}$ integral (4.4) the δ-function contributions from $\beta_{,22}$ and $U_{,12}$ again cancel. For $r \leqslant \lambda R$, we obtain $\gamma_{,0} = 0$, and, for $r \geqslant \lambda R$,

$$\begin{aligned} \gamma_{,0} = & \frac{e^{2H}}{3\lambda R} \left[H_{,22} - H_{,2} \cot\theta + 2(H_{,2})^2 - \frac{\lambda_{,2} H_{,2}}{\lambda} \right] \\ & + \frac{c_{,0}}{r} + \frac{\lambda^2 R^2 e^{2H}}{6r^3} \Big[H_{,22} - H_{,2} \cot\theta + 8(H_{,2})^2 \\ & + \frac{2\lambda_{,2} H_{,2}}{\lambda} \Big] - \frac{\lambda^4 R^4 e^{2H} (H_{,2})^2}{2r^5}, \end{aligned}$$

where

$$c_{,0} = -\tfrac{1}{2} e^{2H}[H_{,22} - H_{,2}\cot\theta + 3(H_{,2})^2].$$

From these results, we obtain for the Bondi energy (4.5) (after a parts integration)

$$E = (1/4\pi) \oint \{\lambda R\,[1 + \tfrac{2}{3}(H_{,2})^2] + e^{-2H}(2/k - R) - 2/k\} \sin\theta\, d\theta\, d\phi, \quad (4.7)$$

and, for the energy flux,

$$F = (1/16\pi)e^{2H}[H_{,22} - H_{,2}\cot\theta + (H_{,2})^2]^2. \quad (4.8)$$

In order to examine the content of these results, let us first consider the low density limit $k \to 0$. An expansion of the Bondi energy integral (4.7) gives

$$E = \tfrac{1}{8}kR^2 \int_0^\pi (\lambda^2 - 1)\sin\theta\, d\theta - \tfrac{1}{48}k^2R^3 \int_0^\pi [(\lambda-1)^2(\lambda+2) - \lambda(\lambda_{,2})^2] \times \sin\theta\, d\theta + O(k^3).$$

We may compare this with the corresponding Newtonian system consisting of dust, initially at rest, satisfying the Poisson equation $\nabla^2\Phi = 4\pi\rho$ (at an initial absolute time). For this model, the mass m of the Newtonian system equals the relativistic internal energy $E_{\rm I}$,

$$m = \oint \rho r^2 \sin\theta\, dr\, d\theta\, d\phi = \tfrac{1}{8}kR^2 \int_0^\pi (\lambda^2 - 1)\sin\theta\, d\theta,$$

and the Newtonian binding energy equals $B = \frac{1}{2}\int \rho\Phi r^2 \sin\theta\ dr\, d\theta\, d\phi$. We immediately see that $E = m + O(k^2)$. Furthermore, for spherically symmetric systems, the Newtonian binding energy is easy to compute, and we find that

$$E = m + B + O(k^3). \quad (4.9)$$

In the nonspherical case, the Poisson equation cannot be integrated to give an analytic expression for the initial Newtonian potential corresponding to our model. (This state-of-affairs is ironic since we have analytically integrated the general relativistic initial value equations.) In fact, Eq. (4.9) does not hold for nonspherical distributions, due to the incoming radiation discussed at the beginning of this section. In model B we furnish an explicit calculation and further discussion of this effect.

When λ does not have equatorial reflection symmetry, the system has nonvanishing Bondi momentum $P = (1/4\pi)\oint \cos\theta\, M \sin\theta\, d\theta\, d\phi$. In the low density limit, this Bondi momentum is $O(k^2)$, consistent with the "initially-at-rest" condition. At order k^2, this momentum arises from the incoming waves.

In the low density limit, the energy flux (4.8) has the dependence

$$F = (k^2R^2/256\pi)(\lambda_{,22} - \lambda_{,2}\cot\theta)^2 + O(k^3).$$

The existence of an $O(k^2)$ initial flux is quite distinct from what one would expect if there were a simple Newtonian correspondence. In that case, the Einstein radiation formula would lead only to an $O(k^4)$ flux (which, for our model, would initially vanish since, for a Newtonian system initially at rest, the quadrupole moment has vanishing third time derivative). The existence of this $O(k^2)$ outgoing flux is another effect arising from the incoming waves associated with the shear-free condition on the initial data. Also note that the flux vanishes to order k^2 when λ is a combination of $l = 0$ and $l = 1$ harmonics, consistent with the lack of monopole and dipole gravitational radiation. However, when λ contains $l = 1$ harmonics, an $O(k^4)$ flux arises from the $(H_{,2})^2$ term in (4.8); but this flux does then have a quadrupole ($\sin^4\theta$) angular distribution.

In the nonlinear regime, extreme effects arise from rapid angular behavior. If λ has the form $2 + cP_l(\cos\theta)$, then, for high l, the $(H_{,2})^2$ dominates the Bondi energy (4.7) so that $E \sim l^2$. Similarly, $F \sim l^4$ so that short-wavelength ripples are rapidly smoothed by gravitational radiation. In the high-density limit of any nonspherical distribution the result is more drastic: $E \sim k^2$ but now $e^{-2H}F \sim k^4$ so that, since $H \sim k$, the smoothing proceeds at a rate which increases exponentially with k.

Note that in the high-density limit of a spherical distribution the Bondi energy approaches the limiting value λR as $k \to \infty$. This can be understood in reference to the Schwarzschild radius at $r = 2\lambda R$, which is now exterior to the outer boundary of the shell at $r = \lambda R$. Thus, in this limit, there are antitrapped surfaces so that the spherically symmetric extension of the solution into the past must include part of the initial Schwarzschild singularity. The Bondi mass apparently saturates at the value λR because a particle, which satisfies our "at rest" condition, has zero energy (with respect to infinity) at half the Schwarzschild radius; i.e., $w_\mu \xi^\mu = 0$ at $r = M$ where ξ^μ is the static Killing vector.

The properties of this model, as described above, are manifest in Figs. 1 and 2. In addition, these figures reveal a linear regime in which the radiation power is proportional to k^2 (or k^4 in the exceptional $l = 1$ case), which extends to a density at which the Bondi mass differs by roughly 10% from the corresponding Newtonian mass. This gives an indication of the range of validity of linear perturbation calculations.

B. Homogeneous spheroids

We now consider the homogeneous density distribution $\rho = k/(8\pi)$, for $r \leqslant \lambda R$, and $\rho = 0$, for $r > \lambda R$. In particular, we take $\lambda = (1 - \epsilon^2\sin^2\theta)^{-1/2}$, which describes an oblate spheroid of eccentricity ϵ. In this case, the Newtonian mass m and binding energy B (for the spheroid initially at rest) are well known:

$$m = 4\pi\rho R^3/3(1 - \epsilon^2) \quad (4.10)$$

and

$$B = -(4\pi\rho m R^2 \sin^{-1}\epsilon)/5\epsilon(1 - \epsilon^2)^{1/2}. \quad (4.11)$$

The general relativistic hypersurface integrals (4.1)–(4.3) lead first to

$$\beta = \begin{cases} kr^2/8, & \text{for } r \leqslant \lambda R, \\ k\lambda^2R^2/8, & \text{for } r \geqslant \lambda R \end{cases}$$

so that $H = k\lambda^2R^2/8$. Here $\beta_{,12}$ has the δ-function behavior

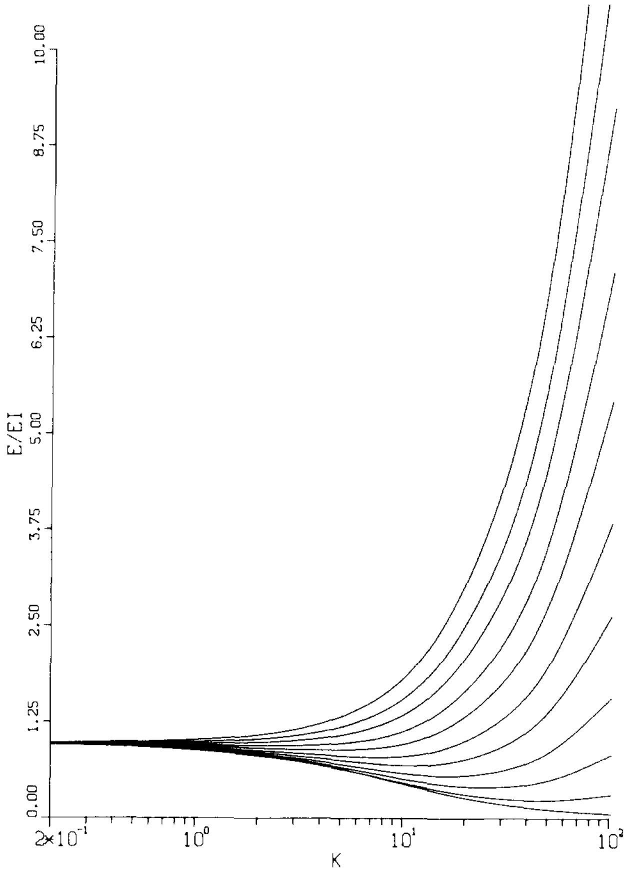


FIG. 1. Curves of E/E_1 vs k for thick dust shells. The surface of each shell has shape $\lambda_l = 2 + a_l P_l(\cos\theta)$, with curves corresponding to $l = 0,1,...,10$ drawn from bottom to top. (a_l is chosen to make E_1/k constant.) Note that E is positive for all shapes and densities. Large values of k correspond to models with particle horizons (white holes).

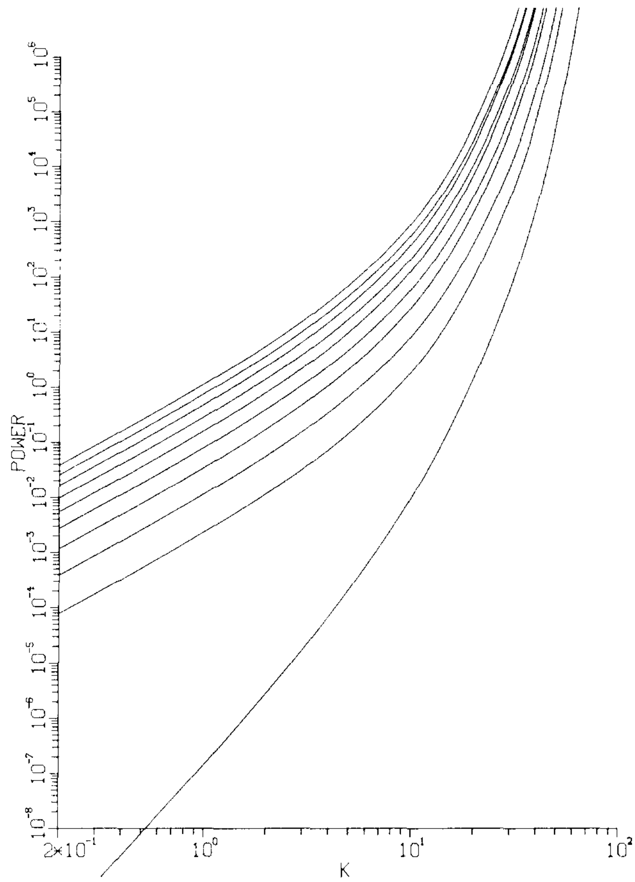


FIG. 2. Initial gravitational radiation power vs k for thick dust shells. Curves with $l = 1,2,...,10$ run from bottom to top. Radiation from $l = 1$ models only becomes significant for large k.

$$\int_{\lambda R}^{\lambda R_+} f\beta_{,12}\,dr = fkR^2\lambda\lambda_{,2}/4|_{\lambda R}.$$

Taking this into account, the U integral leads to $U = 0$, for $r \leqslant \lambda R$, and

$$U = 2e^{2H}H_{,2}\left(\frac{1}{r} - \frac{1}{3\lambda R} - \frac{2\lambda^2R^2}{3r^3}\right),$$

for $r \geqslant \lambda R$. Next, for $r \leqslant \lambda R$, the V integral gives

$$V = -re^{kr^2/4} + 2\int_0^r dr\, e^{kr^2/4},$$

where the right-hand side involves an integral.

As in the previous model, the δ-function contributions, at $r = \lambda R$, from $U_{,12}$ and $\beta_{,22}$ combine to cancel in the V integral. For $r \geqslant \lambda R$, we obtain

$$\begin{aligned} V|_{\alpha R}^{r} = \{ &re^{2H}[1 - H_{,22} - H_{,2}\cot\theta - 2(H_{,2})^2] \\ &- (4\lambda^2R^2/r)e^{2H}(H_{,2})^2(1 - \lambda^2R^2/3r^2) \\ &+ \tfrac{1}{2}(\sin\theta)^{-1}[\sin\theta\,(e^{2H})_{,2}(3r - 2r^2/3\lambda R \\ &+ 2\lambda^2R^2/3r)]_{,2}\}|_{\alpha R}^{r}. \end{aligned}$$

We now compare, at low density, the general relativistic and Newtonian models. From V, we find that the mass aspect has the expansion

$$\begin{aligned} M = &\frac{k\lambda^3R^3}{6} + \frac{kR^3(\lambda^2\lambda_{,2}\sin\theta)_{,2}}{4\sin\theta} \\ &- \frac{k^2\lambda^5R^5}{60} + \frac{k^2\lambda^3R^5(\lambda_{,2})^2}{24} + O(k^3). \end{aligned}$$

This leads to the Bondi energy

$$\begin{aligned} E = \frac{k}{12}\int_0^\pi d\theta\,\sin\theta\Big[&\lambda^3R^3 + \frac{k\lambda^3R^5(\lambda_{,2})^2}{4} \\ &- \frac{k\lambda^5R^5}{10}\Big] + O(k^3). \end{aligned}$$

After carrying out the angular integration, this may be put in the form

$$E = m - \{4\pi\rho mR^2/[5(1-\epsilon^2)^2]\}(1 - \tfrac{4}{3}\epsilon^2 + \tfrac{1}{7}\epsilon^6) + O(k^3). \tag{4.12}$$

The $O(k)$ first term is the mass of the Newtonian model. For spherical symmetry ($\epsilon = 0$), the $O(k^2)$ second term reduces to the Newtonian binding energy (4.11), as in our previous model.

In the nonspherical case, the Bondi energy has the small ϵ expansion

$$E = m - \tfrac{4}{5}\pi\rho mR^2[1 + \tfrac{2}{3}\epsilon^2 + \tfrac{1}{3}\epsilon^4 + O(\epsilon^6)] + O(k^3)$$

while the Newtonian binding energy has the expansion

$$B = -\tfrac{4}{3}\pi\rho m R^2\left[1 + \tfrac{2}{3}\epsilon^2 + \tfrac{8}{15}\epsilon^4 + O(\epsilon^6)\right].$$

We see that the $O(k^2)$ term in the Bondi energy equals the Newtonian binding energy up to $O(\epsilon^2)$ and *exceeds* the Newtonian binding energy at $O(\epsilon^4)$. This agrees with the physical picture presented at the beginning of this section: The shear-free data implies the presence of an incoming gravitational wave, with $O(k)$ amplitude at low density. The energy of this wave makes a *positive* $O(k^2)$ contribution which increases the Bondi energy above the Newtonian binding energy. At the same time, this also confirms, in the context of a radiation space–time, the close correspondence between Bondi energy and Newtonian energy, at low density and velocity. Referring to (4.12), in the $\epsilon \rightarrow 1$ pancake limit, the wave energy dominates and the total $O(k^2)$ Bondi energy is positive. The balance point, at which $E = m + O(k^3)$, occurs at $\epsilon \approx 0.898$.

To examine the prolate spheroidal case, the substitution $\epsilon^2 \rightarrow \epsilon^2/(\epsilon^2 - 1)$ gives the corresponding expressions for E, m, and B. The energy comparisons are completely analogous to the oblate results except that the condition for $E = m + O(k^3)$ is $\epsilon \approx 0.878$. In both the oblate and prolate case, a high eccentricity is necessary for the energy of incoming waves to be comparable to the Newtonian binding energy. The lower eccentricity, at the balance point, for the prolate spheroid is consistent with the interpretation that this incoming radiation arises from the shear-free condition. A cigar-shaped object has greater focussing power, and would therefore normally introduce greater shear, than a disc. Thus a stronger incoming wave is necessary, in the cigar case, to remove the shear.

Numerical results for this model are given in the next section.

V. NUMERICAL PROCEDURES AND RESULTS

In the formulation of a scheme for numerical solution to the field equations, we have been guided by Penrose's beautiful geometric characterization of radiation in asymptotically flat space–time. To discuss boundary conditions and fields at null infinity, we perform a coordinate transformation which assigns finite coordinate values to points at future null infinity. The physical space metric is necessarily singular at these points. However, all divergences may be collected into an overall conformal factor. The resulting conformal geometry is regular everywhere and completely characterizes the global structure of the asymptotically flat space–time. Therefore, it is ideally suited for numerical computation.

For the rest of this section, we will change the previous notation and denote the physical space coordinates, metric, and matter fields used in earlier sections by writing a tilde over the former symbols. We introduce new coordinates

$$u = \tilde{u}/a, \quad x = \tilde{r}/(a + \tilde{r}), \quad y = -\cos\tilde{\theta}, \quad \phi = \tilde{\phi},$$

where a is an arbitrary constant which may be chosen to set the scale of the geometry and $0 \leqslant x \leqslant 1$, $-1 \leqslant y \leqslant 1$. The coordinate x serves to compactify the radial direction and maps null infinity into the edge of the new coordinate patch at $x = 1$. New metric fields are introduced in the new frame by

$$\gamma(u,x,y) \equiv \tilde{\gamma}(\tilde{u},\tilde{r},\tilde{\theta})/\sin^2\tilde{\theta},$$
$$\beta(u,x,y) \equiv \tilde{\beta}(\tilde{u},\tilde{r},\tilde{\theta}),$$
$$U(u,x,y) \equiv a\tilde{U}(\tilde{u},\tilde{r},\tilde{\theta})/\sin\tilde{\theta},$$
$$S(u,x,y) \equiv a[\tilde{V}(\tilde{u},\tilde{r},\tilde{\theta}) - \tilde{r}]/\tilde{r}^2$$

and new matter fields are denoted by

$$\rho(u,x,y) \equiv a^2\tilde{\rho}(\tilde{u},\tilde{r},\tilde{\theta}),$$
$$p(u,x,y) \equiv a^2\tilde{p}(\tilde{u},\tilde{r},\tilde{\theta}),$$
$$w_1(u,x,y) \equiv \tilde{w}_1(\tilde{u},\tilde{r},\tilde{\theta}),$$
$$w_2(u,x,y) \equiv \tilde{w}_2(\tilde{u},\tilde{r},\tilde{\theta})/(a\sin\tilde{\theta}).$$

The physical and conformal metrics are related by

$$d\tilde{s}^2 = \Omega^{-2}\,ds^2$$

with conformal factor

$$\Omega = (\tilde{r} + a)^{-1} = (1 - x)/a.$$

Ω vanishes at null infinity as required. The conformal geometry has the metric

$$ds^2 = [(1-x)(1-x+xS)e^{2\beta} - x^2(1-y^2)U^2 f(\gamma)]\,du^2 + 2e^{2\beta}\,du\,dx + 2x^2 U f(\gamma)\,du\,dy - x^2[f(\gamma)dy^2/(1-y^2) + (1-y^2)\,d\phi^2/f(\gamma)],$$

where $f(\gamma) \equiv \exp[2\gamma(1-y^2)]$. From the discussion of asymptotic behavior in Sec. III, it follows that the new matter and metric fields are regular everywhere in the conformally compactified space–time.

In order to calculate the value of the metric fields resulting from the matter distribution, the field-equation hierarchy of Sec. II is transformed to the new coordinate frame. For fixed u, the compactified (x, y)-hypersurface is approximated by a lattice of discrete points (x_i, y_j), where $x_i = i/I$ for $i = 0,1,2,\ldots,I$ and $y_j = \pm j/J$ for $j = 0,1,2,\ldots,J$. At each lattice site, the values of γ, β, U, S, and $\gamma_{,u}$ are computed by numerical integration over x, using a simple rectangular approximation with the integrand evaluated at points $(x_{k+1/2}, y_j)$ for $k = 0,1,2,\ldots,i-1$. The sites are traversed in order of increasing x_i so that old values of the integrals may be incremented to obtain the new field values. This provides an efficient numerical procedure. The values of the total gravitational energy and radiation power are readily found by numerical integration over y, using the actual field values computed at null infinity ($x = 1$).

The exact thick shell solution of Sec. III A provides an independent check on the accuracy of the numerical calculations. The total energy and total radiation flux (power) are, in addition to their physical significance, sensitive indicators of computational error. For a typical model (with $k = 1.6$, $R = \frac{3}{7}$, $\lambda = 2 + 0.49P_4(-y)$ and a numerical grid with dimensions $I = 144$ and $J = 24$) the computed energy is 0.1996, compared to the exact value 0.2002. The computed power is 0.1022, while the exact value is 0.1024. Thus numerical computations are in excellent agreement with the exact results.

Figures 3 and 4 are graphs of the numerical calculations of energy and radiation power for the oblate spheroidal models of Sec. III B. The results for prolate spheroids are quite similar for the parameter ranges illustrated. Again, as in the shell model of Sec. III A, there is a linear regime, in which

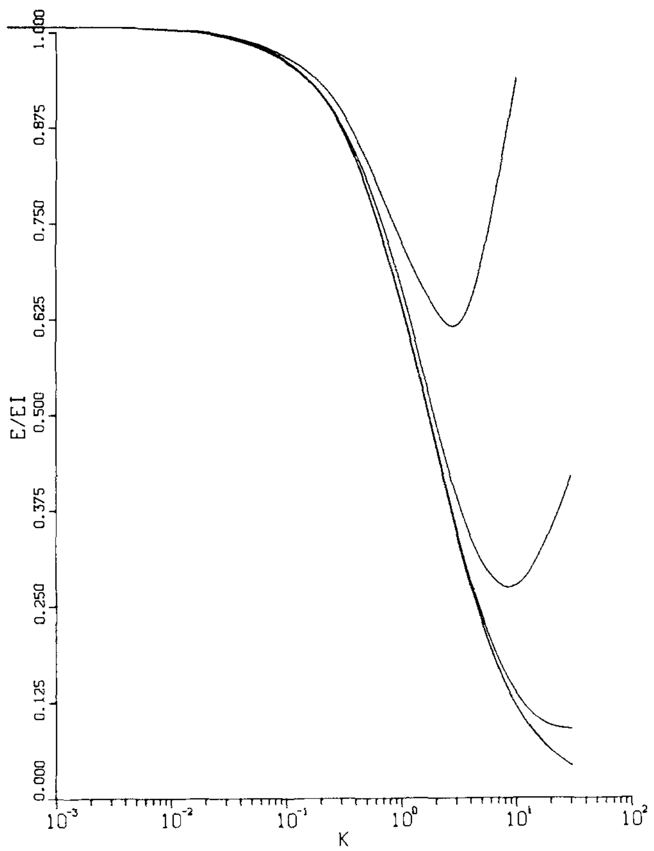


FIG. 3. E/E_1 vs k for dust spheroids. Curves of constant eccentricity $\epsilon = 0$, 0.3, 0.5, and 0.7 are plotted from bottom to top. (E_1/k is constant for all spheroids.) Here, curves for $\epsilon \leqslant 0.1$ would be indistinguishable from the $\epsilon = 0$ curve.

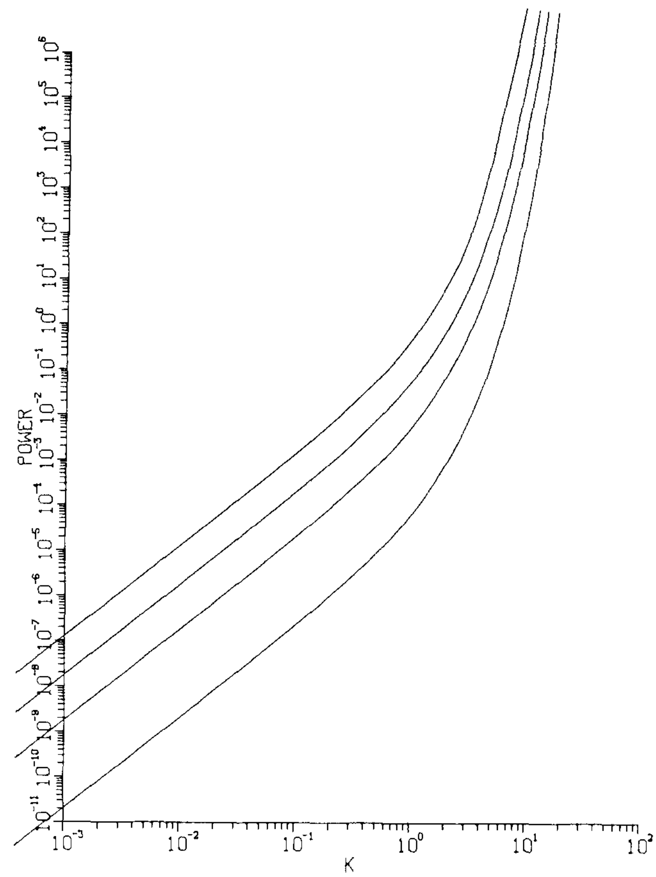


FIG. 4. Gravitation radiation power vs k from dust spheroids. Curves with $\epsilon = 0.1$, 0.3, 0.5, and 0.7 are drawn from bottom to top.

the power is proportional to k^2, which extends to a density at which $(E_1 - E)/E_1 \approx 0.1$.

Figure 5–11 illustrate the fields associated with a dumbell-shaped dust distribution. The dust configuration is given by

$$\rho = \frac{kr^2}{8\pi} \exp\left[-\left(\frac{r - r_0}{\sigma_r}\right)^2 - \left(\frac{\text{abs}(\cos\theta) - 1}{\sigma_y}\right)^2\right].$$

This gives two regions of dust, separated by a distance $2r_0$ in the axial direction. For the case in the illustrations, $r_0 = 1$, the density parameter $k = 2$, and the widths of the Gaussians σ_r and σ_y are 0.5 and 0.35, respectively. All of the figures are graphed in the compactified coordinates defined earlier in this section, and each shows a metric field in redefined (untilded) form. We again adopt the shear-free and "at rest" conditions.

Figure 5 shows the density distribution itself. Figures 6–8 show the metric components β through S, calculated by numerical integration of the hypersurface equations. Figure 9 shows the u derivative of γ, calculated from the evolution equation.

Figure 10 gives the flux F as a function of the y-coordinates. This is in its original (tilded) form, as given in Sec. IV. Note that the flux distribution is four-lobed. The inner, larger lobes are produced by the y derivative of H, while the outer lobes are contributed by the u derivative of γ.

The final figure shows the sign of the trapped surface condition given at the end of Sec. III. The upper plateaus represent regions where both of the two orthogonal (to an

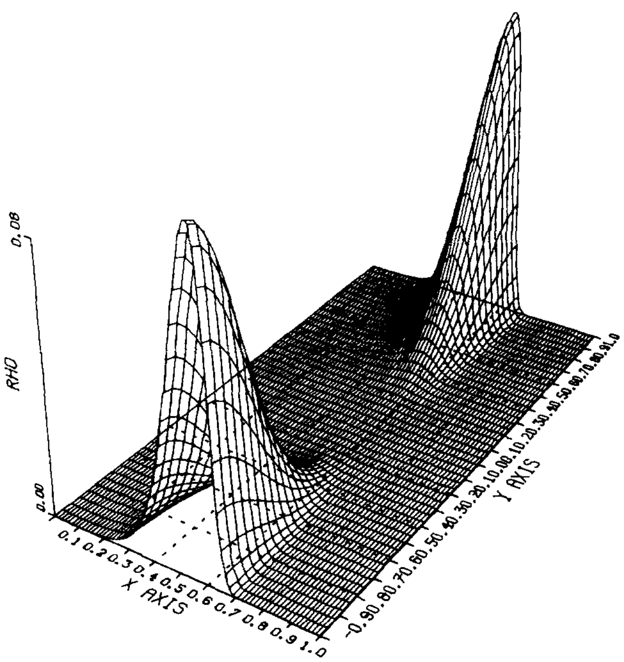


FIG. 5. The dust density distribution for a pair of blobs on the symmetry axis. The density formula is given in the text. For this and the following 3D plots, the origin lies along the $x = 0$ edge, and I along the $x = 1$ edge. The angles from 0 to π are spanned by y ranging from -1 to 1.

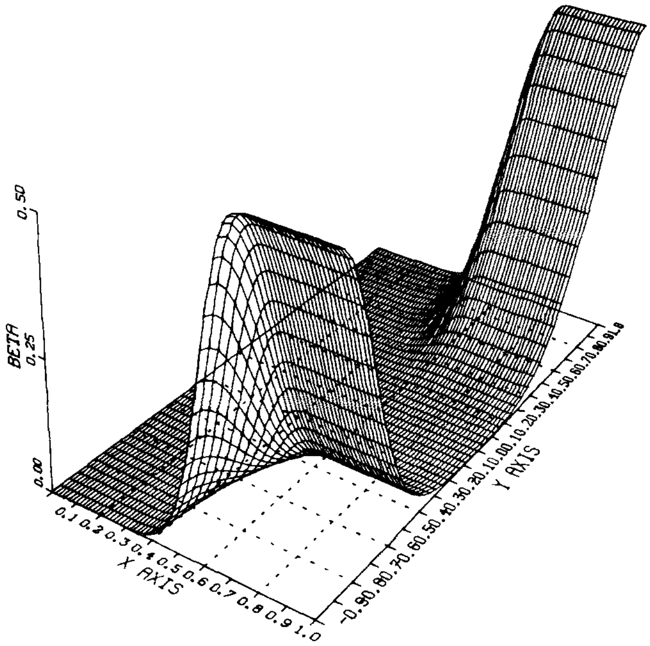


FIG. 6. Plot of the metric function β resulting from the density distribution of Fig. 5.

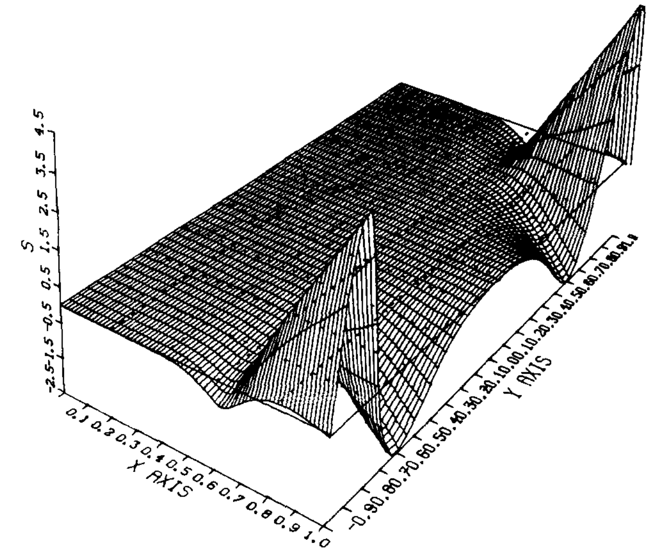


FIG. 8. Plot of the metric function S resulting from the density distribution of Fig. 5. This is the untilded S, which plays the role of V.

$r =$ const surface) sets of null directions diverge. Were these plateaus to form a complete band at some constant $x(r)$, an antitrapped surface would exist. In the case shown, this condition for an antitrapped surface is not fully satisfied. The internal energy cannot be computed analytically, but numerical integration gives $E/E_{\mathrm{I}} = 1.29$ for this case.

VI. OUTLOOK FOR THE FUTURE

The analytic and numerical results, obtained in our exploration of initial data corresponding to shear-free dust "at rest," constitute only the first step toward realistic dynamical models of radiating sources. We are currently modifying our computer codes to incorporate the effects of shear, pres-

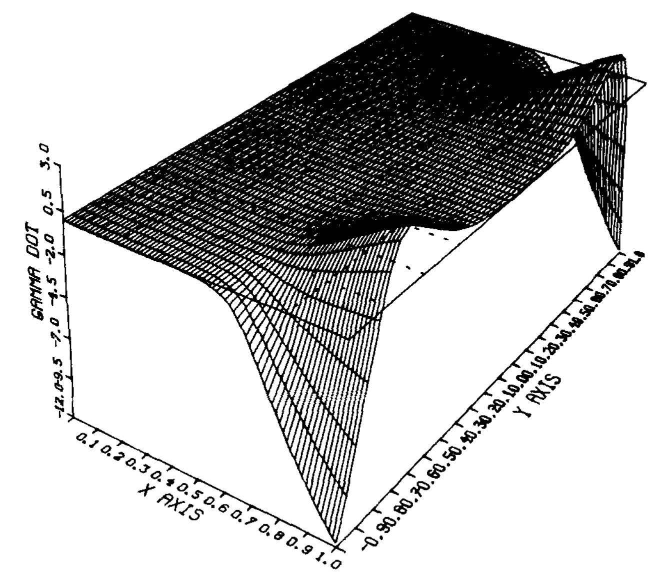


FIG. 9. Plot of the u derivative of γ resulting from the dust distribution of Fig. 5. γ itself is taken to be zero for this model. This is the untilded $\gamma_{,u}$, as defined in Sec. V.

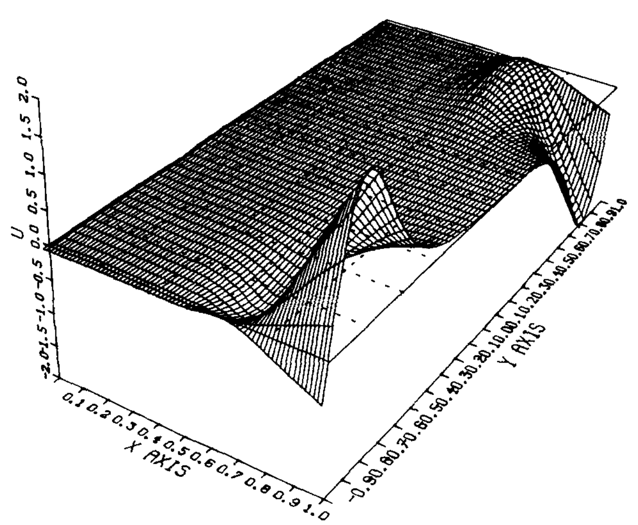


FIG. 7. Plot of the metric function U resulting from the density distribution of Fig. 5. This is the untilded U, as defined in Sec. V.

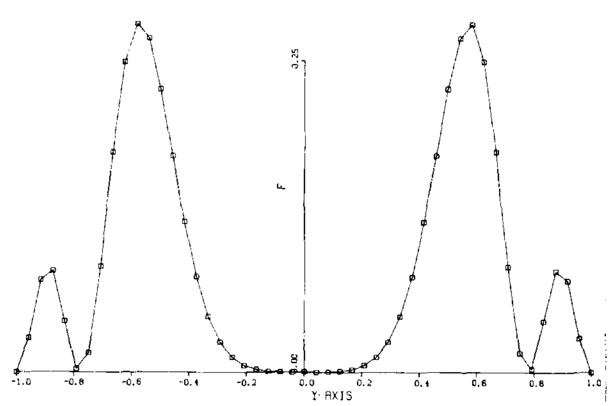


FIG. 10. Plot of the flux F, produced by the dust distribution of Fig. 5, vs the angular coordinate y. Note that the distribution is four-lobed.

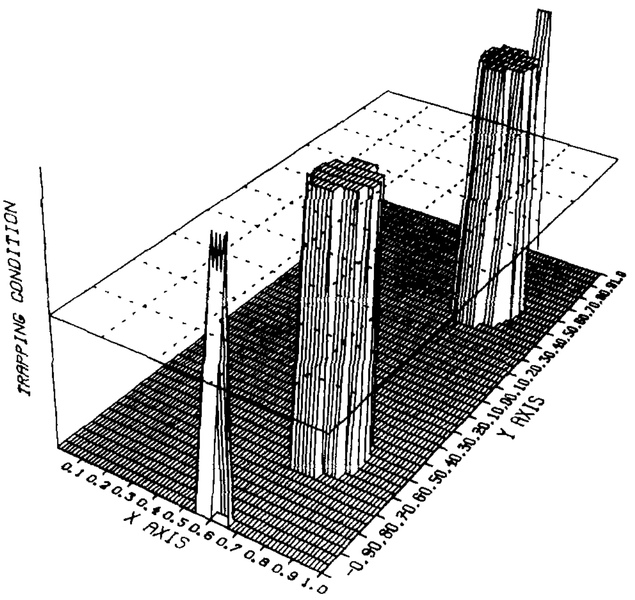


FIG. 11. This plot shows the sign of the trapped surface condition, as described in the text, for the dust distribution of Fig. 5.

sure, and velocity. An analytic integration of the initial value equations, for a model which includes any of these additional effects, would provide an extremely valuable check.

Some important conceptual problems remain in the physical interpretation of the data for the characteristic initial value problem. The low-density limit of our exact models has a consistent semi-Newtonian interpretation as an incoming gravitational wave superimposed upon a dust configuration. Work is proceeding on examining these results in the general context of a rigorous Newtonian limit,[15] in which $c \rightarrow \infty$. This should identify the appropriate conditions on the initial shear which rule out incoming radiation in the weak field case. The presence of some admixture of incoming radiation seems unavoidable in the strong field case. First, the characteristic initial value problem, set on an outgoing null hypersurface, has no *unique* evolution into the past unless the history of the news function, at future null infinity, is specified. Second, even if this history were known, it would be unfeasible, using existing techniques, to determine the radiation fields at past null infinity. Third, any local criterion to exclude incoming waves would be obscured, from a physical standpoint, by the backscattering of prior outgoing radiation, which normally occurs in the strong field case.

We feel that this subtle physical issue poses the most important problem in the theoretical analysis of gravitational radiation from a specific astrophysical source. It is a problem common to both the characteristic and spacelike approaches. As research progresses, "computer experiments," which evolve an initially weak field system into the strong field regime, should provide physical intuition to help formulate a theoretical resolution.

ACKNOWLEDGMENTS

One of us (R.A.I.) would like to thank John Stewart for a very stimulating discussion of numerical radiation calculations planned at Cambridge. He would also like to thank the Department of Physics and Astronomy of the University of Pittsburgh for its hospitality while part of this research was carried out. Finally, he is grateful to the MIT Mathlab Group for access to MACSYMA which was used to check several results. Another author (J.S.W.) would like to thank N. Zabusky for much advice on numerical methods.

[1]J. H. Taylor and L. A. Fowler, Sci. Am. **245**(4), 74 (1981); J. H. Taylor and J. M. Weisberg, Astrophys. J. **253**, 908 (1982).

[2]*Proceedings of the Second Marcel Grossman Conference on General Relativity*, edited by R. Ruffini (North-Holland, Amsterdam, 1982).

[3]K. Thorne, Rev. Mod. Phys. **52**, 285 (1980); R. Weiss, in *Sources of Gravitational Radiation*, edited by L. Smarr (Cambridge U.P., Cambridge, 1979); R. W. P. Drever, in *Experimental Gravity, Quantum Optics, and Measurement Theory*, edited by P. Meystre and M. O. Scully (Plenum, New York, 1982).

[4]*Isolated Gravitating Systems in General Relativity*, edited by J. Ehlers (North-Holland, Amsterdam, 1979); F. Pirani, in *Brandeis Summer Institute in Theoretical Physics* (Prentice-Hall, Englewood Cliffs, NJ, 1964), Vol. 1.

[5]H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner, Proc. Roy. Soc. London Ser. A **269**, 21 (1962).

[6]R. K. Sachs, Proc. Roy. Soc. London Ser. A **270**, 103 (1962).

[7]R. Penrose, in *Relativity, Groups and Topology*, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1964); R. Penrose, Proc. Roy. Soc. London Ser. A **284**, 159 (1965).

[8]L. Tamburino and J. Winicour, Phys. Rev. **150**, 1039 (1966).

[9]R. Geroch, in *Asymptotic Structure of Space–Time*, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977).

[10]*Sources of Gravitational Radiation*, edited by L. Smarr (Cambridge U.P., Cambridge, 1979); *Essays in General Relativity: A Festschrift for Abraham Taub*, edited by F. Tipler (Academic, New York, 1980).

[11]J. Wilson, in *Sources of Gravitational Radiation*, edited by L. Smarr (Cambridge U.P., Cambridge, 1979).

[12]J. York, Jr., in *Sources of Gravitational Radiation*, edited by L. Smarr (Cambridge U.P., Cambridge, 1979).

[13]R. Geroch and J. Winicour, J. Math. Phys. **22**, 803 (1981).

[14]An interesting discussion of the interplay between shear and gravitational energy is given by R. Penrose, in *Perspectives in Geometry and Relativity*, edited by B. Hoffman (Indiana U.P., Bloomington, 1966).

[15]H. P. Kunzle, Gen. Rel. Gravit. **7**, 445 (1976); G. Dautcourt, Acta Phys. Polon. **25**, 637 (1964); P. Havas, Rev. Mod. Phys. **36**, 938 (1964); J. Ehlers, in *Grundlagenprobleme der Physik* (*Festschrift for P. Mittelstaedt*), edited by J. Nitsch *et al.* (Mannheim, 1981).

The nonsymmetric Kaluza–Klein theory

M. W. Kalinowski[a)]
Department of Physics, University of Toronto, Toronto, Ontario, M5S 1A7, Canada

(Received 21 July 1982; accepted for publication 21 January 1983)

This paper is devoted to a five-dimensional unification of Moffat's theory of gravitation and electromagnetism. We found "interference effects" between gravitational and electromagnetic fields which appear to be due to the skew-symmetric part of the metric of Moffat's theory. Our unification called the nonsymmetric Kaluza–Klein theory becomes the classical Kaluza–Klein theory if the skew-symmetric part of the metric is zero. The possible generalization to an arbitrary gauge group is discussed.

PACS numbers: 04.50. + h

INTRODUCTION

The aim of this paper is to find the Kaluza–Klein analog for Moffat's theory of gravitation.[1–3] In other words, it will be a five-dimensional unification of Moffat's theory and classical Maxwell electromagnetism. Our unification, called nonsymmetric Kaluza–Klein theory, is analogous to the relation between Moffat's theory and general relativity. The diagram (Fig. 1) places our unification among the above-mentioned theories.

Roughly speaking, in general relativity, mass curves space–time. In Moffat's theory, mass and fermion charge (fermion number) curve and twist space–time. In the classical Kaluza–Klein theory, mass curves space–time, and electric charge curves the fifth dimension. In the nonsymmetric Kaluza–Klein theory, mass and fermion number curve and twist space–time, and electric charge curve and twists the fifth dimension.

Moffat's theory of gravitation is based on three fundamental geometrical quantities: two connections $\overline{\Gamma}^{\alpha}_{\beta\gamma}$ and $\overline{W}^{\alpha}_{\beta\gamma}$ and the nonsymmetric metric $g_{\alpha\beta}$. This nonsymmetric metric is equivalent to the existence of two geometrical objects defined on space–time: the symmetric metric tensor

$$\bar{g} = g_{(\alpha\beta)}\bar{\theta}^{\alpha} \otimes \bar{\theta}^{\beta}$$

and the two-form

$$\underline{g} = g_{[\mu\nu]}\bar{\theta}^{\mu} \wedge \bar{\theta}^{\nu}.$$

In the general theory of relativity we have only one connection with vanishing torsion and a symmetric metric on space–time. Thus we have only $\overline{\Gamma}$ and $\bar{g}$. Of course, in Moffat's theory connection $\overline{\Gamma}$ and $\overline{W}$ are interrelated and have nonvanishing torsion.

The classical Kaluza–Klein approach and its generalization to nonabelian gauge groups (see Ref. 4–8) was based on the following ideas.

On the space–time we have Riemannian geometry based on the metric tensor $\underline{g}$, and we have general relativity with the local coordinate invariance principle. Simultaneously, we have a principal fiber bundle over space–time

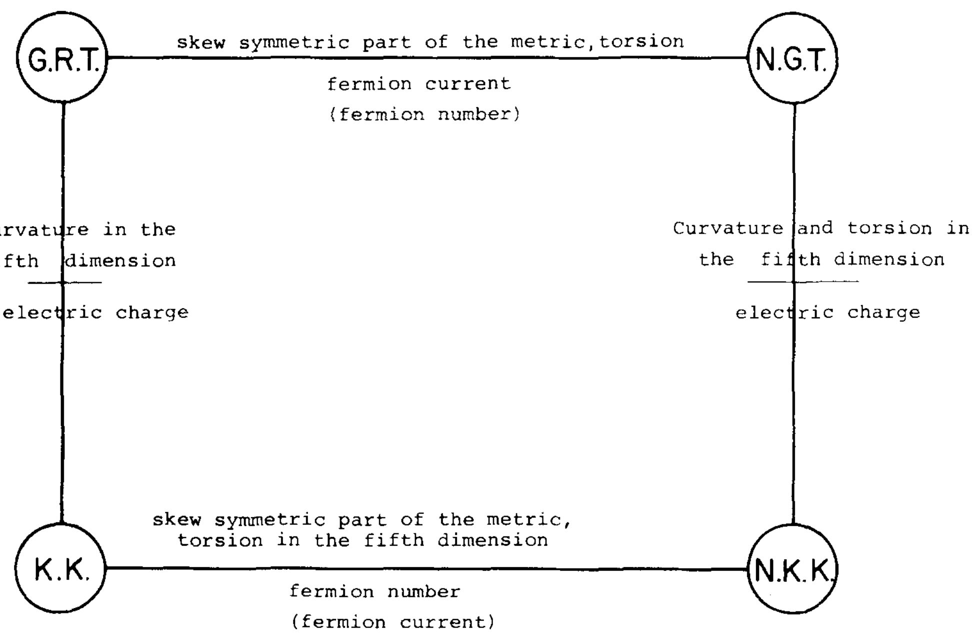


FIG. 1. The position of the nonsymmetric Kaluza–Klein theory among general relativity, nonsymmetric theory of gravitaiton, and the classical Kaluza–Klein theory: G.R.T. = general theory of relativity; N.G.T. = nonsymmetric theory of gravitation (Moffat's theory; real version); K.K. = Klein–Kaluza theory; N.K.K. = nonsymmetric Kaluza–Klein theory.

[a)] On leave of absence from the Institute of Philosophy and Sociology of Polish Academy of Sciences, 00-330 Warsaw Nowy Swiat 72, Poland.

 0022-2488/83/071835-11$02.50

with the structural group U(1) (in some generalization an arbitrary nonabelian group G). The connection on this bundle describes the electromagnetic field (or in the case of an arbitrary gauge group Yang–Mills field—gauge field). We have also the local gauge invariance principle for the electromagnetic field (or the Yang–Mills field).

The local coordinate invariance principle and the local gauge invariance principle seem to be two major concepts of physics. The Kaluza–Klein theory unifies these two concepts and reduces them to the first, the local coordinate invariance principle, but in a more than four-dimensional world. In the electromagnetic case we deal with a five-dimensional manifold [in general with $(n+4)$-dimensional for an arbitrary gauge group, where $n = \dim G$].

The basic idea is very simple. On the gauge group we have bi-invariant symmetric tensors (for example, the Cartan–Killing tensor). This tensor plays the role of a metric in the Lie algera of the gauge group G (normally it is supposed that G is semisimple). In the five-dimensional (electromagnetic) case we have as this tensor the number (-1).

On the fiber bundle we have the natural distribution of horizontal spaces induced by the connection.

On space–time acts the metric tensor $\bar{g}$.

We can divide every tangent vector to the fiber bundle in only one way (the connection is established) into two parts—horizontal and vertical. The horizontal part we can project onto space–time, and the vertical one, due to the connection, onto the Lie algebra of the gauge group. Thus we have natural (symmetric) metrization of the fiber bundle. We can "measure" independently the length of both parts by two (symmetric) metric tensors and after this add these two results. This construction was first introduced by Trautman.[9] Having the principal fiber bundle metrized in a natural way (the metric tensor is bi-invariant with respect to the gauge group action on the bundle), we introduce linear connections on the bundle which are compatible in some sense with the metric. The simplest solution is to suppose that this connection is the Levi-Civita connection. This was done in the five-dimensional Kaluza–Klein theory. If we calculate the Ricci curvature scalar for this connection, we get a sum of the Ricci curvature scalar on space–time and the electromagnetic Lagrangian. In the nonabelian case [$(n+4)$-dimensional] the result will be more complex; we get a sum of the Ricci curvature scalar on space–time, the Yang–Mills Lagrangian plus cosmological constant, which is 10^{127} times bigger than the upper limit from observational data. This makes us change geometry on the metrized fiber bundle, and abandon the Levi-Civita (Riemannian) connection. We must employ the connection with torsion. This was done in a natural geometrical way in Ref. 10. The cosmological constant vanishes (it is almost zero from observational data).

In the light of the new observational data concerning the quadrupole moment of mass for the sun (see Ref. 11), it seems that the general theory of relativity is unable to explain the perihelion movement of Mercury and Icarus.

Moffat's theory can explain the observational data (see Refs. 12 and 13). Moffat's theory due to using fermion current (fermion number $F = B - L$, where B is barion charge and L is lepton charge) as second gravitational charge (the first is the mass) seems to be closer to the elementary particle theory than general relativity.

Thus it would be natural and important because of further investigations to find the Kaluza–Klein analog for Moffat's theory.

This theory—the nonsymmetric Kaluza–Klein theory—unifies the coordinate invariance principle from Moffat's theory and the local gauge invariance principle.

Following ideas concerning the geometry of the Kaluza–Klein theory described above, it is necessary to find the natural nonsymmetric metrization of the fiber bundle over space–time. The existence of a nonsymmetric metric on the fiber bundle is equivalent to the existence of two bi-invariant geometrical objects: $\bar{\gamma}$ and $\underline{\gamma}$. The first $\bar{\gamma}$ is a symmetric bi-invariant tensor, and the second $\underline{\gamma}$ is a bi-invariant 2-form on the fiber bundle. The first is constructed and used in the classical Kaluza–Klein theory (natural symmetric metrization). It is necessary to construct the second one.

Following the basic idea of the previous construction, it is necessary to choose a bi-invariant skew-symmetric form on the gauge group G. We have a natural skew-symmetric form defined on the Lie algebra of G. It is the commutator. This form has values in the Lie algebra of the group. But the inner product of this form and the vector $C = h^{ab}\,\mathrm{Tr}[(X_a)^2]X_b$ (trace is with respect to the space of the generator representation and inner product is defined by the Killing–Cartan form) is a number. If the representation is real, we got what we were looking for. The form is bi-invariant with respect to the group action. This form is often zero. For example, it is zero for U(1) and all abelian groups.

Now following the idea of the symmetric metrization of the fiber bundle, we can build $\underline{\gamma}$ from $\underline{g}$ and this form. If the form is zero, $\underline{\gamma} = \pi^*(\underline{g})$, where π^* is a pullback of π (the natural projection on the fiber bundle).

In this paper we deal with the simplest case, the electromagnetic one. The general case will be done elsewhere.

But in this very simple case we got interesting results. The nonsymmetric Kaluza–Klein theory seems to be a real unified theory of electromagnetic and gravitational fields. It not only reduces two major principles of invariance to the local coordinate invariance principle, but it provides new effects, which are absent in the classical Kaluza–Klein theory. These effects are also absent in Moffat's theory of gravitation and in Maxwell's electromagnetism. Thus they are some "interference effects" between gravitational and electromagnetic fields. They are following:

(1) the new term in the electromagnetic Lagrangian,

$$(1/4\pi)(g^{[\mu\nu]}F_{\mu\nu})^2;$$

(2) the existence of an electromagnetic polarization of the vacuum $M_{\alpha\beta}$;

(3) the additional term for the Lorentz force term in the equation of motion for a test particle,

$$(g/m_0)g^{[\gamma\alpha]}H_{\gamma\beta}U^\beta,$$

where g is a charge of test particle and m_0 its rest mass;

(4) the new energy–momentum tensor ${}^{\mathrm{em}}T_{\mu\nu}$ for an electromagnetic field with zero trace;

(5) the source for the electromagnetic field—the con-

served current j_α.

All of these effects vanish if the metric of space–time becomes symmetric. In this case we get the classical Kaluza–Klein theory.

The paper is organized as follows. In the first section we introduce the notations and definitions of all geometrical quantities which we use throughout the paper. In the second section we define the natural nonsymmetric metrization of the principal fiber bundle. In the third section we formulate the nonsymmetric Kaluza–Klein theory. We calculate connections $\omega^A{}_B$ and $W^A{}_B$ on the five-dimensional manifold, which are analogous to connections $\overline{\omega}^\alpha_\beta$ and $\overline{W}^\alpha_\beta$ from Moffat's theory of gravitation. In Sec. 4 we write the geodetic equation on $\underline{P}$ (nonsymmetrically metrized electromagnetic bundle), and we find a new correction to the Lorentz force term. We calculate the 2-form of torsion and the 2-form of curvature for the connection $\omega^A{}_B$. After this we write the curvature tensor for $\omega^A{}_B$ and its contraction and the Moffat–Ricci tensor. Using obtained results, we calculate the Moffat–Ricci tensor and the Moffat–Ricci curvature scalar for the connection $W^A{}_B$.

In Sec. 5 we define the Palatini variational principle for the Moffat–Ricci curvature scalar $R\,(W)$. We get field equations for gravitational and electromagnetic fields. We discuss and interpret our results and point out all differences between the classical and the nonsymmetric Kaluza–Klein theory. We write down all "interference effects" between gravitational and electromagnetic fields which appear in our theory. In Sec. 6 we discuss some numerical predictions of the theory with a comparison to observational data. In Sec. 7 we deal in detail with an equation of motion for a test particle.

1. ELEMENTS OF GEOMETRY

In this section we introduce the notation and define geometric quantities used in the paper. We use a smooth principal fiber bundle $\underline{P}$, which includes in its definition the following list of differentiable manifolds and smooth maps: a total (bundle space $\underline{P}$; a base E (in our case it is a space–time); a projection π: $\underline{P}\to E$; a map $\Phi: p\times G\to \underline{P}$ defining the action of G on $\underline{P}$; if $a,b\in G$ and $\epsilon\in G$ is the unit element, then $\Phi\,(a)\circ\Phi\,(b) = \Phi\,(ba)$ and $\Phi\,(\epsilon) = id$ and $\Phi\,(a)p = \Phi\,(p,a)$; moreover, $\pi\circ\Phi\,(a) = \pi$. ω is a 1-form of a connection on $\underline{P}$ with values in the Lie algebra of the group G. Let $\Phi\,'(a)$ be the tangent map to $\Phi\,(a)$ where $\Phi\,^*(a)$ is contragredient to $\Phi\,(a)$ at the point a. The form ω is a form of ad type, i.e.,

$$\Phi\,^*(a)\omega = \mathrm{ad}'_{a^{-1}}\omega, \tag{1.1}$$

$\mathrm{ad}'_{a^{-1}}$ is the tangent map to the internal automorphism of the group G

$$\mathrm{ad}_a(b) = aba^{-1}.$$

Due to the form ω we can introduce the distribution field of linear elements H_r, $r\in\underline{P}$, where $H_r\subset T_r(\underline{P})$ is a subspace of the space tangent to $\underline{P}$ at a point r and

$$v\in H_r \Leftrightarrow \omega(v) = 0. \tag{1.2}$$

we have

$$T_r(\underline{P}) = V_r\otimes H_r \tag{1.3}$$

where H_r is called a subspace of horizontal vectors and V_r of vertical vectors. For vertical vectors $v\in V_r$ we have $\pi'(v) = 0$. This means that v is tangent to fibers. Let us define

$$v = \mathrm{hor}(v) + \mathrm{ver}(v), \quad \mathrm{hor}(v)\in H_r,\ \mathrm{ver}(v)\in V_r. \tag{1.4}$$

It is well known that the distribution H_r is equivalent to a choice of the connection ω. We use the operation "hor" for forms, i.e.,

$$(\mathrm{hor}\,\beta)(X,Y) = \beta\,(\mathrm{hor}\,X,\mathrm{hor}\,Y), \tag{1.5}$$

where $X,Y\in T_r(\underline{P})$. The 2-form of curvature of the connection ω is

$$\Omega = \mathrm{hor}\,d\omega. \tag{1.6}$$

It is also a form of ad type like ω. Ω obeys the structural Cartan's equation

$$\Omega = d\omega + \tfrac{1}{2}[\omega,\omega], \tag{1.7}$$

where

$$[\omega,\omega](X,Y) = [\omega(X),\omega(Y)].$$

Bianchi's identity for Ω is

$$\mathrm{hor}\,d\Omega = 0. \tag{1.8}$$

For the principal fiber bundle we use the following convenient scheme (Fig. 2). The map e: $U\to\underline{P}$, $U\subset E$, so that $e\circ\pi = \mathrm{id}$ is called a local section. From the physical point of view it means choosing the gauge. Thus

$$e^*\omega = e^*(\omega^a X_a) = A^a{}_\mu\,\bar\theta^\mu X_a, \tag{1.9}$$

$$e^*\Omega = e^*(\Omega^a X_a) = \tfrac{1}{2}F^a_{\mu\nu}\bar\theta^{\,\mu}\wedge\bar\theta^{\,\nu}X_a.$$

Further we introduce a notation

$$\Omega^a = \tfrac{1}{2}H^a{}_{\mu\nu}\theta^\mu\wedge\theta^\nu, \tag{1.10}$$

where $\theta^\mu = \pi^*(\bar\theta^\mu)$ and

$$F^a_{\mu\nu} = \partial_\mu A^a{}_\nu - \partial_\nu A^a_\mu + C^a_{bc}A^b_\mu A^c_\nu, X_a,$$

$$a = 1,2,\cdots\dim G = \mathrm{n},$$

are generators of the Lie algebra of the group G and

$$[X_a,X_b] = C^c{}_{ab}X_c.$$

In this paper we use also a linear connection on manifolds $\underline{P}$ and E using the formalism of differential forms. So the basic quantity is a 1-form of connection $\omega^A_{\cdot B}$. The 2-form of curvature is the following:

$$\Omega^A{}_B = d\omega^A{}_B + \omega^A{}_C\wedge\omega^C{}_B \tag{1.11}$$

and the 2-form of torsion

$$\Theta^A = D\theta^A, \tag{1.12}$$

where θ^A are basic forms and D means the exterior covariant derivative with respect to $\omega^A_{\cdot B}$. The following relations define the interrelation between our symbols and generally used

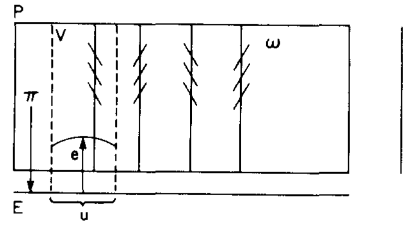


FIG. 2. Principal fiber bundle P.

ones:

$$\omega^A{}_B = \Gamma^A{}_{BC}\theta^C,$$
$$\Theta^A = \tfrac{1}{2} Q^A{}_{BC}\theta^B \wedge \theta^C, \qquad (1.13)$$
$$\Omega^A{}_B = \tfrac{1}{2} R^A{}_{BCD}\theta^C \wedge \theta^D,$$

where $\Gamma^A{}_{BC}$ are coefficients of the connection (they do not have to be symmetric in indices B and C), $R^{\cdot A}_{\cdot BCD}$ is a tensor of curvature and $Q^{\cdot A}_{\cdot BC}$ is a tensor of torsion. Covariant exterior differentiation with respect to $\omega^A{}_B$ is given by the formulas

$$D\Xi^A = d\Xi^A + \omega^A{}_C \wedge \Xi^C,$$
$$D\Sigma^A{}_B = d\Sigma^A{}_B + \omega^A{}_C \wedge \Sigma^C{}_B - \omega^C{}_B \wedge \Sigma^{\cdot A}_{\cdot C}. \qquad (1.14)$$

The forms of curvature $\Omega^{\cdot A}_{\cdot B}$ and torsion Θ^A obey Bianchi's identities

$$D\Omega^A{}_B = 0,$$
$$D\Theta^A = \Omega^{\cdot A}_{\cdot B} \wedge \theta^B. \qquad (1.15)$$

In the paper we use also Einstein's $+$ and $-$ differentiations for the nonsymmetric metric tensor g_{AB}:

$$Dg_{A+B-} = Dg_{AB} - g_{AD}Q^D{}_{BC}\theta^C, \qquad (1.16)$$

where D is the covariant exterior derivative with respect to $\omega^A{}_B$ and $Q^D{}_{BC}$ is the tensor of torsion for $\omega^A{}_B$. In a holonomic system of coordinates we easily get

$$Dg_{A+B-} = g_{A+B-;C}\theta^C$$
$$= (g_{AB,C} - g_{DB}\Gamma^D_{AC} - g_{AD}\Gamma^D_{CB})\theta^C. \qquad (1.17)$$

All quantities introduced in this section and their precise definitions can be found in Refs. 9, 14, 15, and 16.

2. THE NATURAL NONSYMMETRIC METRIZATION OF A BUNDLE $\underline{P}$

Let us introduce the principal fiber bundle P over the space–time E with the structural group G and with the projection π. Let us suppose that (E, g) is a manifold with nonsymmetric metric tensor

$$g_{\mu\nu} = g_{(\mu\nu)} + g_{[\mu\nu]}. \qquad (2.1)$$

Let us introduce a natural frame on $\underline{P}$

$$\theta^A = (\pi^*(\bar\theta^\alpha), \theta^a = \lambda\omega^a), \quad \lambda = \text{const}. \qquad (2.2)$$

It is convenient to introduce the following notations. Capital Latin indices A,B,C run $1,2,3,...,n+4$, $n = \dim G$. Lower Greek indices $\alpha,\beta,\gamma,\delta = 1,2,3,4$ and lower Latin cases $a,b,c,d = 5,6,...,n+4$. The overbar over θ^α and over other quantities indicates that these quantities are defined on E.

It is easy to see that the existence of the nonsymmetric metric on E is equivalent to introducing two independent geometrical quantities on E.

$$\bar g = g_{\alpha\beta}\bar\theta^\alpha \otimes \bar\theta^\beta = g_{(\alpha\beta)}\bar\theta^\alpha \otimes \bar\theta^\beta, \qquad (2.3)$$
$$\underline{g} = g_{\alpha\beta}\bar\theta^\alpha \wedge \bar\theta^\beta = g_{[\alpha\beta]}\bar\theta^\alpha \wedge \bar\theta^\beta, \qquad (2.4)$$

i.e., the symmetric metric tensor $\bar g$ on E and the 2-form $\underline{g}$. On the group G we can introduce a bi-invariant symmetric tensor called the Killing–Cartan tensor:

$$h(A,B) = \mathrm{Tr}(\mathrm{Ad}_A \circ \mathrm{Ad}_B), \qquad (2.5)$$

where $\mathrm{Ad}_A(C) = [A,C]$. It is easy to see that

$$h(A,B) = h_{ab}A^a \cdot B^b, \qquad (2.6)$$

where

$$h_{ab} = C^c{}_{ad}C^d{}_{bc}, \quad h_{ab} = h_{ba},$$
$$A = A^a X_a, \quad B = B^a X_a.$$

This tensor is distinguished by the group structure, but there are, of course, other bi-invariant tensors on G. Normally it is supposed that G is semisimple. It means $\det(h_{ab}) \neq 0$. What is a natural 2-form on G, or a natural skew-symmetric bi-invariant tensor? It is easy to see that

$$K(A,B) = h[[A,B],C], \quad C = h^{ab}\cdot\mathrm{Tr}[(X_a)^2]X_b \qquad (2.7)$$

has these properties and

$$K(A,B) = K_{bc}A^b \cdot B^c, \qquad (2.8)$$

where

$$K_{bc} = C^a{}_{bc}\cdot\mathrm{Tr}[(X_a)^2],$$
$$K_{bc} = -K_{cb}, \quad h^{ab}h_{ac} = \delta^b_c.$$

The trace Tr is here understood in the sense of the representation space of generators X_a. If the representation is a real, than K is a real too. The tensor K is zero in the two important cases:

(1) G is abelian;

(2) $\wedge_a \ \mathrm{Tr}[(X_a)^2] = 0$.

Thus K is zero for U(1). Let us turn to the nonsymmetric natural metrization of $\underline{P}$. Let us suppose that

$$\bar\gamma(X,Y) = g(\pi'X,\pi'Y) + \lambda^2 h(\omega(X),\omega(Y)), \qquad (2.9)$$
$$\bar\gamma(X,Y) = \bar g(\pi'X,\pi'Y) + \mu\lambda^2 K(\omega(X),\omega(Y)), \qquad (2.10)$$

$\mu = \text{const}$ and is dimensionless, $X,Y \in tan(\underline{P})$. The first formula (2.9) was introduced by Trautman (see Ref. 9) for the symmetric natural metrization of $\underline{P}$, and it was used to construct the Kaluza–Klein theory for U(1) and nonabelian generalizations of this theory.[7,8,10] It is easy to see that

$$\bar\gamma = \pi^* g + h_{ab}\theta^a \otimes \theta^b, \qquad (2.11)$$
$$\underline{\gamma} = \pi^* \underline{g} + \mu K_{ab}\theta^a \wedge \theta^b, \qquad (2.12)$$

or

$$\gamma_{(AB)} = \left(\begin{array}{c|c} g_{(\alpha\beta)} & 0 \\ \hline 0 & h_{ab} \end{array}\right), \qquad (2.13)$$

$$\gamma_{[AB]} = \left(\begin{array}{c|c} g_{[\alpha\beta]} & 0 \\ \hline 0 & \mu K_{ab} \end{array}\right). \qquad (2.14)$$

For

$$\gamma_{AB} = \gamma_{(AB)} + \gamma_{[AB]},$$

one easily gets

$$\gamma_{AB} = \left(\begin{array}{c|c} g_{\alpha\beta} & 0 \\ \hline 0 & l_{ab} \end{array}\right), \qquad (2.15)$$

where $l_{ab} = h_{ab} + \mu K_{ab}$. Tensor γ_{AB} has this simple form in the natural frame on P, θ^A. This frame is unholonomical, because

$$d\theta^a = \tfrac{1}{2}\lambda\,[H^a{}_{\mu\nu}\theta^\mu \wedge \theta^\nu - (1/\lambda^2)C^a{}_{bc}\theta^b \wedge \theta^c] \neq 0. \qquad (2.16)$$

We also introduce a dual frame

$$\bar\gamma(\xi_A) = \gamma_{(AB)}\theta^B. \qquad (2.17)$$

We have $\xi_A = (\xi_\alpha, \alpha_a)$ and

$$\underset{\xi_a}{\mathscr{L}}\, \bar{\gamma} = 0, \tag{2.18}$$

$$\underset{\xi_a}{\mathscr{L}}\, \underline{\gamma} = 0. \tag{2.19}$$

Thus γ is bi-invariant with respect to the group action on $\underline{P}$ (ξ_a are, of course, fundamental field on $\underline{P}$). In the case with $\mathrm{Tr}(X_a)^2 = 0$ for every a we have

$$\gamma_{AB} = \left(\begin{array}{c|c} g_{\alpha\beta} & 0 \\ \hline 0 & h_{ab} \end{array}\right). \tag{2.20}$$

For the electromagnetic case $[G = U(1)]$ one easily finds

$$\gamma_{AB} = \left(\begin{array}{c|c} g_{\alpha\beta} & 0 \\ \hline 0 & -1 \end{array}\right). \tag{2.21}$$

Now let us take a section $e{:}E \to \underline{P}$ and fit to it a frame v^a, $a = 5,6,\dots,n+4$, selecting $X^\mu \doteq$ const on a fiber in such a way that e is given by the condition

$$e^* v^a = 0.$$

Thus we have

$$\omega = (1/\lambda) v^a X_a + \pi^*(A^a_\mu \bar{\theta}^\mu) X_a,$$

where

$$e^*\omega = A = A^a{}_\mu \bar{\theta}^\mu X_a.$$

In this frame the tensor γ takes the form

$$\gamma_{AB} = \left(\begin{array}{c|c} g_{\alpha\beta} + \lambda^2 l_{ab} A^a_\alpha A^b_\beta & \lambda l_{cb} A^c_\alpha \\ \lambda l_{ac} A^c{}_\beta & l_{ab} \end{array}\right), \tag{2.22}$$

where

$$l_{ab} = h_{ab} + \mu k_{ab}.$$

The nonsymmetric theory of gravitation (see Refs. 1, 2, and 3) uses the nonsymmetric metric $g_{\mu\nu}$ such that

$$g_{\mu\nu} g^{\beta\nu} = g_{\nu\mu} g^{\nu\beta} = \delta^\beta_\mu, \tag{2.23}$$

where the order of indices is important. If G is semisimple and $\mathrm{Tr}(X_a)^2 = 0$ for every a,

$$l_{ab} = h_{ab}, \quad \det(h_{ab}) \neq 0$$

and

$$h_{ab} h^{bc} = \delta^c_a. \tag{2.24}$$

Thus one easily finds in this case:

$$\gamma_{AC}\gamma^{BC} = \gamma_{CA}\gamma^{CB} = \delta^B_A, \tag{2.25}$$

where the order of indices is important. We will have the same for the electromagnetic case $[G = U(1)]$. In general, if $\det(l_{ab}) \neq 0$, then

$$l_{ab} l^{ac} = l_{ba} l^{ca} = \theta^c_b, \tag{2.26}$$

where the order of indices is important. From (2.15) we have (2.25) for the general nonsymmetric metric γ.

3. FORMULATION OF THE NONSYMMETRIC KALUZA-KLEIN THEORY

Let P be the principal fiber bundle with the structural group $G = \mathrm{U}(1)$, over space–time E with a projection π, and let us define on this bundle a connection α. This bundle we call an electromagnetic bundle and α an electromagnetic connection. For the electromagnetic bundle $\underline{P}$ we can specify all quantities introduced in Sec. 1. We have

$$\Omega = d\alpha = \tfrac{1}{2}\pi^*(F_{\mu\nu}\bar{\theta}^\mu \wedge \bar{\theta}^\nu),$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\nu, \quad e^\alpha = A_\mu \bar{\theta}^\mu$$

and e is a local section of $\underline{P}$. A_μ is the 4-potential of the electromagnetic field and $\bar{F}_{\mu\nu}$ is its strength. Bianchi's identity is the following,

$$d\Omega = 0,$$

and due to this the 4-potential exists. It is, of course, the first Maxwell equation.

On space–time E we define a nonsymmetric metric tensor $g_{\alpha\beta}$ such that

$$g_{\alpha\beta} = g_{(\alpha\beta)} + g_{[\alpha\beta]},$$

$$g_{\alpha\beta} g^{\gamma\beta} = g_{\beta\alpha} g^{\beta\gamma} = \delta^\gamma_\alpha, \tag{3.1}$$

where the order of indices is important. We define also on E two connections $\bar{\omega}^\alpha{}_\beta$ and $\bar{W}^\alpha{}_\beta$:

$$\bar{\omega}^\alpha{}_\beta = \bar{\Gamma}^\alpha{}_{\beta\gamma}\bar{\theta}^\gamma \tag{3.2}$$

and

$$\bar{W}^\alpha{}_\beta = \bar{W}^\alpha{}_{\beta\gamma}\bar{\theta}^\gamma,$$

$$\bar{W}^\alpha{}_\beta = \bar{\omega}^\alpha{}_\beta - \tfrac{2}{3}\delta^\alpha{}_\beta \bar{W}, \tag{3.3}$$

where

$$\bar{W} = \bar{W}_\gamma \bar{\theta}^\gamma = \tfrac{1}{2}(\bar{W}^\sigma{}_{\cdot\gamma\sigma} - \bar{W}^\sigma{}_{\cdot\sigma\gamma})\bar{\theta}^\gamma.$$

For the connection $\bar{\omega}^\alpha{}_\beta$ we suppose the following conditions:

$$\bar{D}g_{\alpha+\beta-} = \bar{D}g_{\alpha\beta} - g_{\alpha\delta}\bar{Q}^\delta{}_{\beta\gamma}(\bar{\Gamma}) = 0, \tag{3.4}$$

$$\bar{Q}^\alpha{}_{\beta\alpha}(\bar{\Gamma}) = 0,$$

where $\bar{D}$ is the exterior covariant derivative with respect to $\bar{\omega}^\alpha{}_\beta$ and $\bar{Q}^\alpha{}_{\beta\gamma}(\bar{\Gamma})$ is the torsion of $\bar{\omega}^\alpha{}_\beta$.

Thus we have on space–time E all quantities from Moffat's theory of gravitation (see Ref. 1, 2, and 3). Now let us turn to the natural nonsymmetric metrization of the bundle $\underline{P}$. According to Sec. 2, we have

$$\bar{\gamma} = \pi^*\bar{g} - \theta^5 \otimes \theta^5 = \pi^*(g_{(\alpha\beta)}\bar{\theta}^\alpha \otimes \bar{\theta}^\beta) - \theta^5 \otimes \theta^5,$$

$$\underline{\gamma} = \pi^*\underline{g} = \pi^*(g_{[\alpha\beta]}\bar{\theta}^\alpha \wedge \bar{\theta}^\beta), \tag{3.5}$$

where $\theta^5 = \lambda\alpha$. From the classical Kaluza–Klein theory we know that $\lambda = 2\sqrt{G}/c^2$ (see Ref. 5). We work with such system of units that $G = c = 1$ and $\lambda = 2$

$$\gamma_{AB} = \left(\begin{array}{c|c} g_{\alpha\beta} & 0 \\ \hline 0 & -1 \end{array}\right). \tag{3.6}$$

We can introduce on $\underline{P}$ on a dual frame $\xi_A = (\xi_\alpha, \xi_5)$ similar as in Sec. 2 for general case.

Now we define, on P, a connection $\omega^A{}_B$ such that

$$D\gamma_{A+B-} = D\gamma_{AB} - \gamma_{AD}Q^D{}_{BC}(\Gamma)\theta^C = 0,$$

$$\underset{\xi_5}{\mathscr{L}}\, \omega^A_{\cdot B} = 0, \tag{3.7}$$

where

$$\omega^A{}_B = \Gamma^A{}_{BC}\theta^C,$$

D is the exterior covariant derivative with respect to the connection $\omega^A{}_B$ [see Eq. (1.16)] and $Q^D{}_{BC}(\Gamma)$ is the tensor of torsion for the connection $\omega^A{}_B$.

After some calculations on finds

$$\omega^A{}_B = \left(\begin{array}{c|c} \pi^*(\overline{\omega}^\alpha{}_\beta) + g^{\gamma\alpha}H_{\gamma\beta}\theta^5 & H_{\beta\gamma}\theta^\gamma \\ \hline g^{\alpha\beta}(h_{\gamma\beta} + 2F_{\beta\gamma})\theta^\gamma & 0 \end{array}\right), \tag{3.8}$$

where $H_{\beta\gamma}$ is a tensor on E such that

$$g_{\delta\beta}g^{\gamma\delta}H_{\gamma\alpha} + g_{\alpha\delta}\, g^{\delta\gamma}H_{\beta\gamma} = 2g_{\alpha\delta}\, g^{\delta\gamma}F_{\beta\gamma}. \tag{3.9}$$

we define on $\underline{P}$ a second connection

$$\omega^A{}_B = \left(\begin{array}{c|c} \pi^*(\overline{W}^\alpha{}_\beta) + g^{\gamma\alpha}H_{\gamma\beta}\theta^5 & H_{\beta\gamma}\theta^\gamma \\ \hline g^{\alpha\beta}(H_{\gamma\beta} + 2F_{\beta\gamma})\theta^\gamma & 0 \end{array}\right).$$

Thus we have on $\underline{P}$ all five-dimensional analog of quantities from Moffat's theory of gravitation, i.e.,

$$W^A{}_B, \quad \omega^A{}_B, \quad \text{and } \gamma_{AB}.$$

4. GEOMETRY OF THE MANIFOLD $\underline{P}$

Let us write an equation for geodesics with respect to the connection $\omega^A{}_B$ on $\underline{P}$:

$$U^B\nabla_B U^A = 0, \tag{4.1}$$

where $U^A(\tau)$ is a tangent vector to the geodesic line and ∇ means covariant derivative with respect to the connection $\omega^A{}_B$. Using (3.7), one easily finds

$$\begin{aligned} &\frac{\overline{D}U^\alpha}{d\tau} + U^5(2g^{\alpha\gamma}F_{\gamma\beta} + g^{\alpha\gamma}H_{\beta\gamma} + g^{\gamma\alpha}H_{\gamma\beta})U^\beta = 0, \\ &\frac{dU^5}{d\tau} + H_{\gamma\beta}U^\gamma U^\beta = 0, \end{aligned} \tag{4.2}$$

where $\overline{D}/d\tau$ means covariant derivative with respect to $\overline{\omega}^A{}_B$ along a curve to which $U^\alpha(\tau)$ is tangent.

In the classical Kaluza–Klein theory (see Ref. 5), $2U^5$ has the interpretation of (g/m_0) for a test particle and the system of equations (4.2) has first integral $U^5 = \text{const}$. In our case it is possible iff

$$H_{\gamma\beta} = -H_{\beta\gamma}. \tag{4.3}$$

And finally we get

$$\begin{aligned} &\frac{\overline{D}U^\alpha}{d\tau} + 2U^5(g^{\alpha\gamma}F_{\gamma\beta} - g^{[\alpha\gamma]}H_{\gamma\beta})U^\beta = 0, \\ &U^5 = \text{const} \quad \left(2U^5 = \frac{g}{m_0}\right). \end{aligned} \tag{4.4}$$

Thus we get the Lorentz-force term in the equation of motion for a test particle. This term really differs from the analogous term in the classical Kaluza–Klein theory. But, if the metric is symmetric, we get the classical Lorentz-force term. Let us turn to calculations of the torsion for $\omega^A{}_B$

$$\Theta^A(\Gamma) = D\theta^A. \tag{4.5}$$

One easily gets

$$\begin{aligned} &Q^{\alpha}_{\beta\gamma}(\Gamma) = \overline{Q}^{\alpha}_{\beta\gamma}(\overline{\Gamma}), \\ &Q^{5}_{\beta\gamma}(\Gamma) = 2(F_{\beta\gamma} - H_{\beta\gamma}), \\ &Q^{\alpha}_{5\gamma}(\Gamma) = -Q^{\alpha}_{\gamma 5}(\Gamma) = 2(g_{(\alpha\beta)}H_{\beta\gamma} - g^{\alpha\beta}F_{\beta\gamma}), \\ &Q^{5}_{5\gamma}(\Gamma) = Q^{5}_{\gamma 5} = 0. \end{aligned} \tag{4.6}$$

Let us define a tensor $K_{\beta\gamma}$ such that

$$H_{\beta\gamma} = F_{\beta\gamma} + K_{\beta\gamma}. \tag{4.7}$$

Now we have

$$Q^{5}_{\beta\gamma}(\Gamma) = -2K_{\beta\gamma}. \tag{4.8}$$

We will find a physical interpretation of this tensor.

Now we calculate the 2-form of curvature for the connection $\omega^A_{\cdot B}$:

$$\Omega^A_{\cdot B}(\Gamma) = d\omega^A_{\cdot B} + \omega^A_{\cdot C}\wedge\omega^C_{\cdot B}. \tag{4.9}$$

One easily gets, using (4.9), (3.7), and (4.3),

$$\begin{aligned} \Omega^\alpha{}_\beta = {} & \overline{\Omega}^\alpha{}_\beta(\overline{\Gamma}) + \overline{\nabla}_\mu(g^{\gamma\alpha}H_{\gamma\beta})\theta^\mu\wedge\theta^5 \\ & + [\, g^{\gamma\alpha}H_{\gamma\beta}F_{\mu\nu} - g^{\alpha\delta}(2F_{\delta[\mu} \\ & - H_{\delta[\mu}) - H_{\nu]\beta}]\,\theta^\mu\wedge\theta^\nu, \end{aligned} \tag{4.10a}$$

$$\begin{aligned} \Omega^\alpha_{\cdot 5} = {} & \{\nabla_{[\mu}[\, g^{\alpha\beta}(H_{\nu]\beta} - 2F_{\nu]\beta})] \\ & + \tfrac{1}{2}g^{\alpha\beta}(H_{\gamma\beta} - 2F_{\gamma\beta})\overline{Q}^{\gamma}_{\mu\nu}(\overline{\Gamma})\}\theta^\mu\wedge\theta^\nu \\ & + g^{\delta\alpha}g^{\gamma\beta}H_{\delta\gamma}(H_{\mu\beta} + 2F_{\beta\mu})\theta^5\wedge\theta^\mu, \end{aligned} \tag{4.10b}$$

$$\begin{aligned} \Omega^5{}_\beta = {} & [\overline{\nabla}_{[\mu}H_{|\beta|\nu} + \tfrac{1}{2}H_{\beta\gamma}\overline{Q}^{\gamma}_{\mu\nu}(\overline{\Gamma})]\theta^\mu\wedge\theta^\nu \\ & + g^{\delta\gamma}H_{\gamma\mu}H_{d\beta}\theta^\mu\wedge\theta^5, \end{aligned} \tag{4.10c}$$

$$\Omega^5_5 = g^{\gamma\beta}H_{\gamma[\mu}(H_{\nu]\beta} - 2F_{\nu]\beta})\theta^\mu\wedge\theta^\nu. \tag{4.10d}$$

$\overline{\Omega}^\alpha{}_\beta(\overline{\Gamma})$ is the 2-form of curvature of the connection $\overline{\omega}^\alpha{}_\beta$ and $\overline{\nabla}$ is the covariant derivative with respect to $\overline{\omega}^\alpha{}_\beta$. After some calculations, one gets the curvature tensor $R^A{}_{BCD}$ from Eqs. (4.10) and a contraction of this tensor:

$$A_{BC}(\Gamma) = R^{A}_{\cdot BCA}(\Gamma). \tag{4.11}$$

One obtains

$$\begin{aligned} A_{\beta\mu}(\Gamma) = {} & \overline{A}_{\beta\mu}(\overline{\Gamma}) + 4g^{\gamma\beta}H_{\gamma\beta}F_{\mu\alpha} + 2g^{\alpha\delta}H_{\delta\mu}H_{\alpha\beta} \\ & + (g^{\alpha\delta}F_{\alpha\delta})H_{\beta\mu}, \end{aligned} \tag{4.12}$$

$$A_{55}(\Gamma) = g^{\delta\alpha}g^{\gamma\beta}H_{\delta\gamma}H_{\alpha\beta} - 2g^{\delta\alpha}g_{\gamma\beta}H_{\delta\gamma}F_{\alpha\beta}, \tag{4.13}$$

where

$$\overline{A}_{\beta\mu}(\overline{\Gamma}) = \overline{R}^\alpha{}_{\beta\mu\alpha}(\overline{\Gamma})$$

is a contraction of the curvature tensor for the connection $\overline{\omega}^\alpha{}_\beta$. Now we pass to calculation of Moffat–Ricci tensor (see Refs. 1 and 3).

$$\begin{aligned} R_{\beta\mu}(\Gamma) &= A_{\beta\mu}(\Gamma) + \tfrac{1}{2}(R^\alpha{}_{\alpha\beta\mu} + R^{5}_{\cdot 5\beta\mu}) \\ &= \overline{R}_{\beta\mu}(\overline{\Gamma}) + 4g^{\gamma\delta}H_{\gamma\beta}F_{\mu\alpha} + 2g^{\alpha\delta}H_{\delta\mu}H_{\alpha\beta} \\ &\quad + (g^{\alpha\delta}F_{\alpha\delta})H_{\beta\mu} + (g^{\gamma\alpha}H_{\gamma\alpha}F_{\beta\mu}, \end{aligned} \tag{4.14}$$

where

$$\overline{R}_{\beta\mu}(\overline{\Gamma}) = A_{\beta\mu}(\overline{\Gamma}) + \tfrac{1}{2}\overline{R}^\alpha{}_{\alpha\beta\mu}$$

is the Moffat-Ricci tensor for the connection $\overline{\omega}^\alpha{}_\beta$:

$$\begin{aligned} R_{55}(\Gamma) = A_{55}(\Gamma) = {} & g^{\delta\alpha}g^{\gamma\beta}H_{\delta\gamma}H_{\alpha\beta} \\ & - 2g^{\delta\alpha}g^{\gamma\beta}H_{\delta\gamma}F_{\alpha\beta}. \end{aligned} \tag{4.15}$$

Using (4.14) and (4.15), one easily finds the Moffat-Ricci curvature scalar:

$$\begin{aligned} R(\Gamma) &= \gamma^{AB}R_{AB}(\Gamma) = g^{\beta\mu}R_{\beta\mu}(\Gamma) - R_{55}(\Gamma) \\ &= \overline{R}(\overline{\Gamma}) + 2(g^{[\mu\nu]}F_{\mu\nu})^2 - H^{\mu\alpha}F_{\mu\alpha}, \end{aligned} \tag{4.16}$$

where

$$\overline{R}(\overline{\Gamma}) = g^{\beta\mu}\overline{R}_{\beta\mu}(\overline{\Gamma})$$

is the Moffat-Ricci curvature scalar for the connection $\overline{\omega}^\alpha{}_\beta$ and

$$H^{\mu\alpha} = g^{\beta\mu} g^{\gamma\alpha} H_{\beta\gamma}.$$

Let us turn back to the connection $\overline{W}^{\alpha}{}_{\beta}$ and $W^{A}{}_{B}$ and calculate the Ricci–Moffat tensor and the Ricci–Moffat scalar curvature for $W^{A}{}_{B}$. It is easy to see that

$$R_{\beta\mu}(W) = R_{\beta\mu}(\Gamma) + \tfrac{2}{3} W_{[\beta,\mu]} \tag{4.17}$$

and

$$R_{55}(W) = R_{55}(\Gamma). \tag{4.18}$$

And finally we get

$$R(W) = \overline{R}(\overline{W}) + 2(g^{[\mu\nu]} F_{\mu\nu})^2 - H^{\mu\alpha} F_{\mu\alpha}, \tag{4.19}$$

where

$$\overline{R}(\overline{W}) = g^{\beta\mu} \overline{R}_{\beta\mu}(\overline{\Gamma}) + \tfrac{2}{3} g^{[\beta\mu]} \overline{W}_{[\beta,\mu]} \tag{4.20}$$

is the Moffat–Ricci curvature scalar for the connection $\overline{W}^{\alpha}{}_{\beta}$ (see Refs. 1 and 3).

5. THE VARIATIONAL PRINCIPLE AND FIELD EQUATIONS. INTERPRETATIONS AND CONCLUSION

Let us define the Palatini variational principle on the manifold $\underline{P}$ for $R(W)$

$$\delta \int_V R(W)\sqrt{\gamma}\, d^5x = 0, \quad V \subset \underline{P}, \tag{5.1}$$

where $\gamma = \det(\gamma_{AB}) = -\det(g_{\alpha\beta}) = -g$. We vary with respect to independent quantities $g_{\alpha\beta}$, $\overline{W}^{\alpha}{}_{\beta\gamma}$ and A_μ. After simple calculations one gets

$$\overline{R}_{\alpha\beta}(\overline{W}) - \tfrac{1}{2} g_{\alpha\beta} \overline{R}(\overline{W}) = 8\pi\, {}^{\text{em}}T_{\alpha\beta}, \tag{5.2}$$

$$\mathbf{g}^{[\mu\nu]}{}_{,\mu} = 0, \tag{5.3}$$

$$g_{\mu\nu,\sigma} - g_{\zeta\nu} \overline{\Gamma}^{\zeta}_{\mu\sigma} - g_{\mu\zeta} \overline{\Gamma}^{\zeta}_{\sigma\nu} = 0, \tag{5.4}$$

$$\partial_\mu (\mathbf{H}^{\alpha\mu}) = 4\mathbf{g}^{[\alpha\beta]} \partial_\beta (g^{[\mu\nu]} F_{\mu\nu}, \tag{5.5}$$

where

$${}^{\text{em}}T_{\alpha\beta} = (1/4\pi)[g^{\gamma\mu} H_{\gamma\alpha} F_{\mu\beta} - 2g^{[\mu\nu]} F_{\mu\nu} F_{\alpha\beta} - \tfrac{1}{4} g_{\alpha\beta} [H^{\mu\nu} F_{\mu\nu} - 2(g^{[\mu\nu]} F_{\mu\nu})^2]], \tag{5.6}$$

$$\mathbf{g}^{[\mu\nu]} = \sqrt{-g}\, g^{[\mu\nu]}, \quad \mathbf{H}^{\mu\alpha} = \sqrt{-g}\, g^{\beta\mu} g^{\gamma\alpha} H_{\beta\gamma}, \tag{5.7}$$

and

$$g_{\delta\beta} g^{\gamma\delta} H_{\gamma\alpha} + g_{\alpha\delta} g^{\delta\gamma} H_{\beta\gamma} = 2 g_{\alpha\delta} g^{\delta\gamma} F_{\beta\gamma}. \tag{5.8}$$

Equations (5.2) and (5.3) are equations for the gravitational field in the presence of electromagnetic sources. ${}^{\text{em}}T_{\alpha\beta}$ plays the role of an energy–momentum tensor for the electromagnetic field. Equation (5.4) is a compatibility condition for the metric on space–time [see Eq. (3.4)]. Equation (5.5) plays the role of the second Maxwell equation. It is easy to see that

$$g^{\alpha\beta}\, {}^{\text{em}}T_{\alpha\beta} = 0 \tag{5.9}$$

and

$$\partial_\alpha (4\mathbf{g}^{[\alpha\beta]} \partial_\beta (g^{[\mu\nu]} F_{\mu\nu})) = 0. \tag{5.10}$$

Now we are able to interpret all quantities in our theory. First of all, it is easy to see that $H_{\alpha\beta}$ plays the role of the second tensor of the electromagnetic strength, and Eq. (5.8) expresses the relationship between both tensors $F_{\alpha\beta}$ and $H_{\alpha\beta}$.

In the classical electrodynamics of continuous media[17] or in nonlinear electrodynamics[18] it is necessary to define both of these tensors. The first tensor $F_{\alpha\beta}$ is built from $(\vec{E}, \vec{B})$ and second $H_{\alpha\beta}$ from $(\vec{D}, \vec{H})$.

If the metric $g_{\alpha\beta}$ is symmetric then $F_{\alpha\beta} = H_{\alpha\beta}$. Thus it is interesting that the skew-symmetric part of the metric $g_{[\alpha\beta]}$ induces some kind of an electromagnetic polarization tensor of the vacuum.

In the classical electrodynamics of continuous media[11] and in nonlinear electrodynamics,[18] it is possible to define the electromagnetic polarization tensor of the continuous medium (classical electrodynamics) or the vacuum (nonlinear electrodynamics) called $M_{\alpha\beta}$:

$$H_{\alpha\beta} = F_{\alpha\beta} - 4\pi M_{\alpha\beta}. \tag{5.11}$$

It is easy to see that

$$4\pi M_{\alpha\beta} = -K_{\alpha\beta} \tag{5.12}$$

[see (4.7)]. Thus we got a geometrical interpretation of $M_{\alpha\beta}$

$$Q^5{}_{\alpha\beta}(\Gamma) = 8\pi M_{\alpha\beta}. \tag{5.13}$$

The electromagnetic polarization induced by the skew-symmetric part of the metric $g_{[\mu\nu]}$ is the torsion in the fifth dimension. This is in very good accordance with results from Ref. 19 and 20. The only difference is that in Refs. 19 and 20 the electromagnetic polarization has its origin from external sources and (5.13) plays the role of the Cartan equation in the Kaluza–Klein theory with torsion.

But this is not all. The skew-symmetric part of the metric $g_{[\mu\nu]}$ changes also the electromagnetic Lagrangian

$$\mathscr{L}_{\text{em}} = (1/8\pi)[2(g^{[\mu\nu]} F_{\mu\nu})^2 - H^{\mu\alpha} F_{\mu\alpha}]. \tag{5.14}$$

In (5.14) we have a new term $2(g^{[\mu\nu} F_{\mu\nu})^2$, which is an interaction between the skewon field and the electromagnetic one. This term vanishes if the metric is symmetric and is always nonnegative.

Thus classical electrodynamics in the nonsymmetric theory of gravitation (Moffat's theory) will be different from that in general relativity.

The skew-symmetric part of the metric induces also a source for the electromagnetic field. In Eq. (5.5) we get a current

$$\mathbf{j}^\alpha = (1/\pi)\mathbf{g}^{[\alpha\beta\nu]} \partial_\beta (g^{[\mu\nu]} F_{\mu\nu}), \tag{5.15}$$

which is conserved:

$$\mathbf{j}^\alpha{}_{,\alpha} = 0. \tag{5.16}$$

This current vanishes if the metric is symmetric. This is completely different than in classical Kaluza–Klein theory (see Refs. 4 and 5). In the classical approach based on a symmetric metric on space–time, one obtained the second Maxwell equation in the vacuum. Thus nonsymmetric Kaluza–Klein theory, combining Moffat's theory and the electromagnetic Maxwell theory, is much stronger than the classical Kaluza–Klein approach combining general relativity and electromagnetism.

In the nonsymmetric Kaluza–Klein theory there exist "interference effects" between gravitation and electromagnetism, which are absent in the classical approach. These new "interference effects" are the following:

(1) the new term in the electromagnetic Lagrangian

$$(1/4\pi)(g^{[\mu\nu]} F_{\mu\nu})^2;$$

(2) the existence of an electromagnetic polarization of

the vacuum $M_{\alpha\beta}$ which has geometrical interpretation as a torsion in the fifth dimension.

(3) the additional term for the Lorentz force term in the equation of motion for a test particle,

$$(g/m_0)g^{[\gamma\alpha]}H_{\gamma\beta}U^{\beta};$$

Due to these three fundamental "interference effects," we get other effects:

(1) the new energy–momentum tensor ${}^{\mathrm{em}}T_{\alpha\beta}$ for the electromagnetic field with zero trace;

(2) sources for the electromagnetic field—conserved current j^{α}.

All of these "interference effects" vanish if the metric of space–time becomes symmetric. In this case we get classical Kaluza–Klein theory.

6. NUMERICAL PREDICTIONS OF THE THEORY

Let us pass to Eq. (4.4). We get here an additional term for the Lorentz force

$$(q/m_0)g^{[\alpha\gamma]}H_{\gamma\beta}U^{\beta}. \tag{6.1}$$

In the Moffat theory of gravitation there is an exact nonsymmetric solution which is spherically symmetric and static (Schwarzschild-like solution). It has the following shape (see Ref. 3):

$$g_{\mu\nu} = \begin{bmatrix} -(1-2m/r)^{-1} & 0 & 0 & l^2/r^2 \\ 0 & -r^2 & 0 & 0 \\ 0 & 0 & -r^2\sin^2\theta & 0 \\ -l^2/r^2 & 0 & 0 & (1-2m/r)(1+l^4/r^4) \end{bmatrix}; \tag{6.2}$$

for inverse tensor $g^{\mu\nu}$ we similarly have:

$$g^{\mu\nu} = \begin{bmatrix} -(1-2m/r)(1+l^4/r^4) & 0 & 0 & -l^2/r^2 \\ 0 & -1/r^2 & 0 & 0 \\ 0 & 0 & -1/r^2\sin^2\theta & 0 \\ l^2/r^2 & 0 & 0 & (1-2m/r)^{-1} \end{bmatrix}, \tag{6.3}$$

where m is a mass and l^2 is a fermion charge. Let us estimate contribution of (6.1) to the Lorentz force term on a surface of the sun using (6.2) as a metric. In the Moffat theory (see Ref. 12)

$$l = l_{\odot} = (3.1 + 0.5)\times 10^3 \text{ km}, \tag{6.4}$$

and we have for the radius of the Sun

$$R_{\odot} = 0.7\times 10^6 \text{ km}. \tag{6.5}$$

Thus on the surface of the Sun we get

$$\omega_{\odot} = l_{\odot}^2/R_{\odot}^2 \simeq 10^{-6}. \tag{6.6}$$

If we consider Eq. (5.8) for (6.2) with (6.6), we get

$$H_{\beta\gamma} = F_{\beta\gamma} + \text{terms of higher order in } \omega_{\odot}. \tag{6.7}$$

Thus

$$H_{\beta\gamma} \cong F_{\beta\gamma} \tag{6.8}$$

and the electromagnetic polarization tensor induced by the skew-symmetric part of the metric $g_{[14]} = \omega_{\odot}$ is very small. One gets

$$(q/m_0)g^{[\alpha\gamma}H_{\gamma\beta}U^{\beta} \cong (q/m_0)g^{[\alpha\gamma]}F_{\gamma\beta}U^{\beta}. \tag{6.9}$$

But the only nonvanishing components of $G^{[\alpha\gamma]}$ are

$$G^{[14]} = -w_{\odot} \simeq 10^{-6}, \tag{6.10}$$

and the contribution (6.9) to the Lorentz force is 10^{-6} with comparison to the usual Lorentz force term. Thus it is negligible in the solar system. However for a neutron star we have (see Ref. 3)

$$l_N = 7 \text{ km}, \quad R_N = 6 \text{ km}, \quad \omega_N \approx 1, \tag{6.11}$$

and this new term should play a certain role. Unfortunately, only $g^{[14]} = \omega_N \neq 0$. Thus we have only a new term for the electric part of the electromagnetic field (for the second tensor of strength $H_{\alpha\beta}$). It is the same for the new term in the Lagrangian

$$2(g^{[\mu\nu]}F_{\mu\nu})^2 = 2\omega_N^2(F_{14})^2 = 2\omega_N^2\cdot E_z^2. \tag{6.12}$$

The electric field does not play any important role on a surface of neutron stars in contradiction to the magnetic field, and it does not contradict observational data.

It is also interesting to ask what will happen in the case of weak field approximation

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{(\mu\nu)} + h_{[\mu\nu]}, \quad |h_{\mu\nu}| < 1. \tag{6.13}$$

It is easy to see that coupling between $F_{\mu\nu}$ and $h_{[\mu\nu]}$ in the whole Lagrangian is of second order in $h_{[\mu\nu]}$ (now not only $h_{[14]} \neq 0$). Thus these interference effects are really very small and do not contradict experiments.

However, it would be possible to predict significant effects finding exact solutions of full field equations. It seems to be possible using the more general metric

$$g_{\mu\nu} = \begin{bmatrix} -\alpha & 0 & 0 & w \\ 0 & -r^2 & 0 & 0 \\ 0 & 0 & -r^2\sin^2\theta & 0 \\ -w & 0 & 0 & \gamma \end{bmatrix}, \tag{6.14}$$

where

$$\begin{aligned} \alpha &= [1 - 2m/r + \beta(r)]^{-1}, \\ w &= l^2/r^2, \\ \gamma &= [1 - 2m/r + \beta(r)](1 + l^4/r^4). \end{aligned} \tag{6.15}$$

But, unfortunately, up to now this solution is unknown.

7. EQUATION OF MOTION FOR A TEST PARTICLE. MORE CONCLUSIONS

Let us come back to Eq. (4.1). Due to the compatibility condition (3.7) we have (see Refs. 21 and 22) the first integral of motion for Eq. (4.1)

$$\gamma_{(AB)} U^A(t) U^B(t) = \text{const} \tag{7.1}$$

or

$$g_{(\alpha\beta)} U^\alpha(t) U^\beta(t) - (U^5)^2 = \text{const.} \tag{7.1'}$$

However, due to (4.3), we have

$$U^5 = \text{const.} \tag{7.2}$$

Thus we get

$$g_{(\alpha\beta)} U^\alpha(t) U^\beta(t) = \text{const.} \tag{7.3}$$

Let us rewrite Eq. (4.4) in the following form:

$$m_0 a^\alpha + q g^{\alpha\gamma} F_{\gamma\beta}\left(\frac{dx^\beta}{dt}\right) - q g^{[\alpha\gamma]} H_{\gamma\beta}\left(\frac{dx^\beta}{dt}\right) = 0, \tag{7.4}$$

where

$$(q/m_0) = 2U^5 \tag{7.4'}$$

and

$$a^\alpha = \frac{\overline{D} U^\alpha}{dt} = \frac{\overline{D}}{dt}\left(\frac{dx^\alpha}{dt}\right) \tag{7.5}$$

is a covariant 4-acceleration of a test particle. Let us consider an initial Cauchy problem for (7.4) such that

$$\begin{aligned} & x^\alpha(t_0) = x_0^\alpha, \\ & \frac{dx^\alpha}{dt}(t_0) = U_0^\alpha, \\ & g_{\alpha\beta} U_0^\alpha U_0^\beta = 1. \end{aligned} \tag{7.6}$$

Due to Eq. (7.3) we have for every $t \geqslant t_0$

$$g_{\alpha\beta}\frac{dx^\alpha}{dt}(t)\frac{dx^\beta}{dt}(t) = 1. \tag{7.7}$$

Now we will find an interpretation of the additional term for the Lorentz force in Eq. (7.4), i.e.,

$$-q g^{[\alpha\gamma]} H_{\gamma\beta}\frac{dx^\beta}{dt}. \tag{7.8}$$

To do this, let us consider Eq. (7.4) without this term, i.e.,

$$m_0 a^\alpha + q g^{\alpha\gamma} F_{\gamma\beta}\frac{dx^\beta}{dt} = 0. \tag{7.9}$$

This equation is a simple generalization of an equation for a charged point particle in general relativity to the nonsymmetric case. Now $g^{\alpha\gamma}$ is not symmetric and the covariant 4-acceleration is defined in terms of the connection $\overline{\omega}^\alpha{}_\beta$ on E. This connection is, of course, compatible with the nonsymmetric metric $g_{\alpha\beta}$. One easily checks that

$$\begin{aligned} & \frac{d}{dt}\left(g_{(\alpha\beta)}\frac{dx^\alpha}{dt}\frac{dx^\beta}{dt}\right) \\ & = 2U^5 g_{\alpha\delta}\, g^{\delta\gamma} F_{\gamma\beta}\left(\frac{dx^\beta}{dt}\right)\left(\frac{dx^\alpha}{dt}\right) \neq 0. \end{aligned} \tag{7.10}$$

Thus, in general, Eq. (7.9) does not have the first integral of motion (7.3). It means that we are unable in general to preserve the initial normalization for the 4-velocity of a test particle. If we want to have the normalization (7.7), we must add to Eq. (7.9) the auxiliary condition

$$\Phi(U^\alpha) = 0, \tag{7.11}$$

where

$$\Phi(U^\alpha) = g_{(\alpha\beta)} U^\alpha U^\beta - 1. \tag{7.12}$$

The auxiliary condition (7.11) is a nonhalonomic constraint. This constraint is nonintegrable and nonlinear (quadratic in velocities). According to the general theory of mechanical systems with constraints, we know that in such systems we have so-called reaction forces of constraints. Thus we should write (7.9) in a following form:

$$m_0 a^\alpha = -(2U^5 m_0) g^{\alpha\gamma} F_{\gamma\rho} U^\rho + Q^\alpha, \tag{7.13}$$

$$\Phi(U^\alpha) = g_{\alpha\beta} U^\alpha U^\beta - 1 = 0. \tag{7.14}$$

Q^α is a reaction force of the constraint (7.14). The force Q^α much be such that (7.14) is automatically satisfied during a motion. Let us find this force. In order to do this, let us multiply both sides of (7.13) by $g_{(\alpha\beta)} U^\beta$ and integrate from t_0 to t. One gets

$$\begin{aligned} & \tfrac{1}{2} m_0 \Phi(U^\alpha) = \tfrac{1}{2} m_0 (g_{(\alpha\beta)} U^\alpha U^\beta - 1) \\ & = \int_{t_0}^{t} (g_{(\alpha\beta)} U^\beta Q^\alpha - 2m_0 U^5 g_{(\alpha\beta)} g^{\alpha\gamma} F_{\gamma\rho} U^\beta U^\rho)\, dt. \end{aligned} \tag{7.15}$$

For (7.14) is satisfied we get

$$\int_{t_0}^{t} (g_{(\alpha\beta)} U^\beta Q^\alpha - 2m_0 U^5 g_{(\alpha\beta)} g^{\alpha\gamma} F_{\gamma\rho} U^\beta U^\rho)\, dt = 0. \tag{7.16}$$

However, (7.16) is satisfied for every t. Thus we get

$$g_{(\alpha\beta)} U^\beta Q^\alpha - 2m_0 U^5 g_{(\alpha\beta)} g^{\alpha\gamma} F_{\gamma\rho} U^\beta U^\rho = 0. \tag{7.17}$$

It is easy to see that Eq. (7.17) has a solution:

$$Q^\alpha = 2m_0 U^5 g^{\alpha\gamma} F_{\gamma\rho} U^\rho. \tag{7.18}$$

If we put (7.18) into (7.13), we get

$$m_0 a^\alpha = 0. \tag{7.19}$$

This solution has simple physical interpretation. Equation (7.19) is an equation of motion for an uncharged test particle. There is no Lorentz force. It corresponds to a choice $U^5 = 0$ or equivalently $q = 0$. Let us come back to Eq. (7.17) and transform it using condition (3.9). One gets

$$\begin{aligned} & [\, g_{\alpha\beta} U^\beta Q^\alpha + g_{\beta\alpha} U^\beta Q^\alpha + m_0 U^5 (g_{\delta\beta}\, g^{\gamma\delta} H_{\gamma\alpha} \\ & \quad + g_{\alpha\delta}\, g^{\delta\gamma} H_{\beta\gamma}) U^\alpha U^\beta \,] = 0. \end{aligned} \tag{7.20}$$

Equation (7.20) has a solution

$$\begin{aligned} Q^\alpha & = 2m_0 U^5 g^{[\alpha\gamma]} H_{\gamma\beta} U^\beta \\ & = q g^{[\alpha\gamma]} H_{\gamma\beta} U^\beta. \end{aligned} \tag{7.21}$$

Equation (7.21) gives us an interpretation for an additional term for Lorentz force in Eq. (7.4) or (4.4). This additional term is a reaction force of the nonintegrable, nonholonomic, nonlinear constraints (7.12).

Let us pass to the field $H_{\alpha\beta}$. This field plays the role of the second tensor of the electromagnetic strength. However, we have to do with only one electromagnetic field. Equation (5.8) expresses the relationship between $F_{\alpha\beta}$ and $H_{\alpha\beta}$. This equation is the linear equation for $H_{\alpha\beta}$. Difference between

$H_{\alpha\beta}$ and $F_{\alpha\beta}$ appears due to the skew-symmetric part of the metric $g_{\alpha\beta}$. If $g_{[\alpha\beta]}=0$, we have $H_{\alpha\beta}=F_{\alpha\beta}$. The second pair of Maxwell equation [Eq. (5.5)] is the same as in a nonlinear electrodynamics (see Ref. 18) or in a classical electrodynamics of continuous media (see Ref. 17). In Eq. (5.5) we have a sources, a conserved current j^α. This current is built from the skew-symmetric part of the metric $g_{[\alpha\beta]}$. Thus the real source for $H_{\alpha\beta}$ is the skew-symmetric part of metric. In the nonsymmetric theory of gravitation (see Ref. 3) a fermion current is a source for the differential equation of $g_{[\mu\nu]}$. In this way the fermion current becomes a source of a difference between $H_{\alpha\beta}$ and $F_{\alpha\beta}$. In the nonsymmetric theory of gravitation there is not a Lorentz-like force term connected with a fermion charge (see Refs. 1–3).

It is a very important property of this theory. Due to this the weak equivalence principle is satisfied, i.e., universal falling of all uncharged bodies [compare Eq. (7.19)]. This statement is not true for charged bodies. We have the Lorentz-force term. In the nonsymmetric Kaluza–Klein theory appears an additional term involving the tensor $H_{\alpha\beta}$ and the skew-symmetric part of metric $g^{[\alpha\gamma]}$. Due to this term the fermion charge has an influence on the motion of a charged test particle. It is, of course, an influence via a gravitational and an electromagnetic field (no additional Lorentz force with a fermion charge of a particle). But it is an influence. For example, the exact static, sterically symmetric solution of Moffat's theory has two sources: a mass-point m and a point fermion charge l^2 (see Refs. 1–3).

Let us pass to Eq. (5.8). We are able to solve this equation using iterative methods for the weak gravitational field. In order to do this, we write (5.8) in a form

$$H_{\beta\alpha}=(g_{\alpha\delta}\,g^{\delta\gamma}F_{\beta\gamma}-\tfrac{1}{2}g_{[\delta\beta]}\,g^{\gamma\delta}H_{\gamma\alpha}-\tfrac{1}{2}g_{[\alpha\delta]}g^{\delta\gamma}H_{\beta\gamma}), \tag{7.22}$$

and define following transformation,

$$\overset{(n+1)}{H}_{\beta\alpha}=M^{\mu\nu}{}_{\beta\alpha}\overset{(n)}{H}_{\mu\nu} \tag{7.23}$$

such that

$$\overset{(0)}{H}_{\beta\alpha}=F_{\beta\alpha} \tag{7.24}$$

$$\overset{(n+1)}{H}_{\beta\alpha}=\left(g_{\alpha\delta}\,g^{\delta\gamma}F_{\beta\gamma}-\tfrac{1}{2}g_{[\delta\alpha]}g^{\gamma\delta}\overset{(n)}{H}_{\gamma\alpha}-\tfrac{1}{2}g_{[\alpha\delta]}\,g^{\delta\gamma}\overset{(n)}{H}_{\beta\gamma}\right),\quad n=0,1,2,\cdots. \tag{7.24'}$$

One easily gets that

$$\overset{(n+1)}{H}_{\beta\alpha}=(M^{n+1})^{\mu\nu}{}_{\beta\alpha}\overset{(0)}{H}_{\mu\nu}=(M^{n+1})^{\mu\nu}{}_{\beta\alpha}F_{\mu\nu} \tag{7.25}$$

the power $(n+1)$ means $(n+1)$-iteration of the transformation (7.23). We get

$$\overset{(n+1)}{H}_{\beta\alpha}-\overset{(n)}{H}_{\beta\alpha}=-\tfrac{1}{2}\left[g_{[\delta\beta]}g^{\gamma\delta}\left(\overset{(n)}{H}_{\gamma\alpha}-\overset{(n-1)}{H}_{\gamma\alpha}\right)+g_{[\alpha\delta]}g^{\delta\gamma}\left(\overset{(n)}{H}_{\beta\gamma}-\overset{(n-1)}{H}_{\beta\gamma}\right)\right]. \tag{7.26}$$

Now let us suppose that the field $g_{\alpha\beta}$ is weak. It means that

$$g_{\alpha\beta}=\eta_{\alpha\beta}+h_{\alpha\beta}, \tag{7.27a}$$

$$g^{\alpha\beta}=\eta^{\alpha\beta}+\tilde{h}^{\alpha\beta}, \tag{7.27b}$$

$$|h_{\alpha\beta}|,\quad |\tilde{h}^{\alpha\beta}|<\alpha\ll 1, \tag{7.28}$$

where $\eta_{\alpha\beta}$ is the Minkowski tensor. In this case one gets

$$\tilde{g}^{\mu\delta}\cong\eta^{\mu\delta}-\eta^{\alpha\mu}\eta^{\gamma\delta}h_{\gamma\alpha}. \tag{7.29}$$

The skew-symmetric tensors

$$L_{\beta\nu}=-L_{\nu\beta} \tag{7.30}$$

form a natural linear six-dimensional vector space. Let us define the following norm in this space:

$$\|L\|=\max_{\beta,\nu=1,2,3,4}|L_{\beta\nu}|; \tag{7.31}$$

thus our space becomes a Banach space. For sufficiently small α one finds

$$\|\overset{(n+1)}{H}-\overset{(n)}{H}\|\leqslant\beta(\alpha)\|\overset{(n)}{H}-\overset{(n-1)}{H}\|, \tag{7.32}$$

where $0<\beta(\alpha)=96\alpha<1$, if $0<\alpha<1/96$. Equation (7.32) means that the transformation (7.23) is a contraction. According to the Banach theorem this transformation has a fix point

$$\overset{(\infty)}{H}_{\beta\alpha}=M^{\mu\nu}{}_{\beta\alpha}\overset{(\infty)}{H}_{\mu\nu} \tag{7.33}$$

such that

$$\overset{(\infty)}{H}_{\beta\alpha}=\lim_{n\to\infty}(M^n)^{\mu\nu}{}_{\beta\alpha}F_{\mu\nu}=\overset{(\infty)}{M}{}^{\mu\nu}_{\alpha\beta}F_{\mu\nu}. \tag{7.34}$$

The limit (7.34) is understood in a sense of the norm (7.31) and

$$\overset{(\infty)}{M}{}^{\mu\nu}{}_{\beta\alpha}=\lim_{n\to\infty}(M^n)^{\mu\nu}{}_{\beta\alpha}. \tag{7.35}$$

The limit (7.35) is understood in a sense of the usual linear operator topology generated by a topology of a Banach space. According to the Banach theorem there is one and only one fix point of the transformation (7.23) (in a weak field approximation). Thus we get that

$$H_{\beta\alpha}=\overset{(\infty)}{M}{}^{\mu\nu}{}_{\beta\alpha}F_{\mu\nu}. \tag{7.36}$$

Equation (7.36) is a solution of Eq. (5.8). In this case the additional term for the Lorentz force in Eq. (4.4) takes the form

$$-qg^{[\alpha\gamma]}\overset{(\infty)}{M}{}^{\mu\nu}{}_{\gamma\beta}U^\beta F_{\mu\nu}. \tag{7.37}$$

It is purely described by the tensor $F_{\mu\nu}$ and the metric tensor $g_{\alpha\beta}$. We have the same for the reaction force of constraints (7.14)

$$Q^\alpha=-qg^{[\alpha\gamma]}\overset{(\infty)}{M}{}^{\mu\nu}{}_{\gamma\beta}U^\beta F_{\mu\nu}. \tag{7.38}$$

For a nonholonomic (nonintegrable) constraint we have the following statement. A variational problem with differential (nonintegrable, nonholonomic) constraints cannot be reduced to a form where the variation of a certain quantity (an action) is put equal to zero. This is true in a much simpler case of linear nonholonomic constraints (see Ref. 23). Thus, unfortunately, we cannot formulate a principle of action for Eq. (4.4). However, we are still able to interpret the additional term in the Lorentz force as a reaction force of the nonholonomic constraints (7.11). From the geometrical point of view (the force Q^α is absorbed by a geometry) it seems that only metric geometry or Einstein geometry defined on the five-dimensional Kaluza–Klein manifold lead to the condition (7.1). The geometry defined by the metric $\bar{g} = g_{(\alpha\beta)}\bar{\theta}^\alpha \otimes \theta^\beta$, the 2-form $\underline{g} = g_{[\alpha\beta]}\bar{\theta}^\alpha \wedge \bar{\theta}^\beta$ and the connection $\overline{\omega}^\alpha{}_\beta$ satisfying the condition (3.4) we call Einstein geometry. If we want to get conditions (7.2) and (7.3) it seems that we have only three possibilities:

(1) Riemannian geometry (classical Kaluza–Klein theory);

(2) a generalization of the Einstein–Cartan theory and the Kaluza–Klein theory (see Refs. 19 and 20);

(3) Einstein geometry with the condition (4.3), i.e., the theory described in this paper.

The two first geometries are metric. The first is only a model of unification of electromagnetic and gravitational fields. This unification is to perfect. We do not get any "interference effects" between gravitational and electromagnetic fields. It seems that it is only five-dimensional notation of general relativity and Maxwell theory in Riemannian space–time. The second possibility due to Cartan equations in space–time and in the fifth dimension offers some inteference effects: additional current connected to spin sources, Israel energy–momentum tensor as a tensor of energy–momentum for the electromagnetic field, and contact interaction term of electromagnetic polarization in total energy–momentum tensor. Unfortunately, additional geometric degree of freedom, torsion is connected algebraically with external sources: spin and electromagnetic polarization of matter. Thus this torsion does not propagate. The third possibility seems to be more interesting. There are "interference effects" between gravitational and electromagnetic fields. Torsion propagates. It is interesting to notice that, despite completely different geometries in the second and third possibilities, we got the same equation connecting electromagnetic polarization existing in the theory to a torsion in the fifth dimension.

ACKNOWLEDGMENTS

I would like to thank Professor J. W. Moffat and Dr. G. Kunstatter for their kind hospitality and numerous extremely valuable discussions and advice during my stay at the Physics Department of the University of Toronto.

[1]J. W. Moffat, "New theory of gravitation," Phys. Rev. D **19**, 3557 (1979).
[2]J. W. Moffat, "Gauge invariance and string interactions in a generalized theory of gravitation," Phys. Rev. D **23**, 2870 (1981).
[3]J. W. Moffat, "Generalized theory of gravitation and its physical consequences," in *Proceedings of the VII International School of Gravitation and Cosmology, Erice, Sicily*, edited by V. de Sabatta (World Scientific Publishing, Singapore, 1982).
[4]T. Kaluza, Sitzungsber. Preuss. Akad. Wiss. 966 (1921).
[5]A. Lichnerowicz, *Théorie relativistes de la gravitation et de l'électromagnetisme* (Masson, Paris, 1955).
[6]Y. Rayski, "Unified theory and modern physics," Acta Phys. Polon. **XXVIII**, 89 (1965).
[7]R. Kerner, "Generalization of Kaluza–Klein theory for an arbitrary nonabelian gauge group," Ann. Inst. H. Poincaré A **IX**, 143 (1968).
[8]Cho-Y., "Higher dimensional unification of gravitation and gauge theories," J. Math. Phys. **16**, 2029 (1975).
[9]A. Trautman, "Fiber bundles associated with space–time," Rep. Math. Phys. **1**, 29 (1970).
[10]M. W. Kalinowski, "On vanishing of the cosmological constant in nonabelian Kaluza–Klein theories," Int. J. Theor. Phys. **16**, 5 (1983).
[11]H. A. Hill, R. J. Bos, and P. R. Goode, "Preliminary determination of the quadrupole moment of the sun from rotational splitting of global oscillations and its relevance to tests of general relativity," Phys. Rev. Lett. **33**, 1497 (1983).
[12]J. W. Moffat, "Consequences of a New Experimental Determination of the Quadrupole Moment of the Sun for Gravitational Theory," Phys. Rev. Lett. **50**, 709 (1983).
[13]J. W. Moffat, "The orbit of Icarus as a Test of a Theory of Gravitation," University of Toronto preprint, May 1982.
[14]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vols. I and II.
[15]A. Lichnerowicz, *Théorie globale des connexions et de groupe d'holonomie* (Ed. Cremonese, Rome, 1955).
[16]A. Einstein, *The Meaning of Relativity* (Princeton U.P., Princeton, N.J., 1953), pp. 133–65.
[17]S. R. de Groot and L. G. Suttorp, *Foundations of Electrodynamics* (North-Holland, Armsterdam, 1972).
[18]J. Plebański, *Nonlinear Electrodynamics* (Nordita, Copenhagen, 1970).
[19]M. W. Kalinowski, "Gauge fields with torsion," Int. J. Theor. Phys. **20**, 563 (1981).
[20]M. W. Kalinowski, "On a generalization of the Einstein–Cartan Theory and the Klein–Kaluza theory," Lett. Math. Phys. **5**, 489 (1981).
[21]M. A. Tonnelat, *Einstein's Unified Field Theory* (Gordon and Breach, New York, London, Paris, 1966).
[22]M. Wyman, "Unified field theory," Can. J. Math. **2**, 427 (1950).
[23]C. Lanczos, *The Variational Principles of Mechanics* (University of Toronto Press, Toronto, 1970).

The initial value formulation of higher derivative gravity

David R. Noakes
Physics Department, University of Chicago, Chicago, Illinois 60637

(Received 7 September 1982; accepted for publication 17 December 1982)

The initial value problem is considered for the conformally coupled scalar field and higher derivative gravity, by expressing the equations of each theory in harmonic coordinates. For each theory it is shown that the (vacuum) equations can take the form of a diagonal hyperbolic system with constraints on the initial data. Consequently these theories possess well-posed initial value formulations.

PACS numbers: 04.50. + h

I. INTRODUCTION

One major criterion of the physical acceptability of a gravitational theory (or a set of equations supposedly representing the time evolution of any physical quantities in a spacetime) is that it possesses a well-posed initial value formulation. There must be a set of quantities such that when the values of these quantities (the initial data) is specified on a spacelike 3-surface (the initial surface), there exists a unique solution of the equations of the theory in an open neighborhood of the initial surface which matches the initial data on the initial surface. Further, the value of all physical quantities at a point in the solution must depend only on the initial data within the past light cone of the point (only then is there relativistic causal development) and small perturbations of the initial data should produce only small perturbations of the solution (within a compact region). Since the initial value formulation reveals which variables are dynamically free and which are not, the details of a theory's initial value formulation must be known before the correct quantization can be established.

The initial value formulation of general relativity has been studied in detail. A summary of the application of the harmonic coordinate method of this paper to general relativity has been written by Bruhat.[1] As can be seen there, or in the exposition of the method in Sec. II below, in harmonic coordinates, Einstein's equation is a diagonal hyperbolic system with constraints on the initial data. Then a theorem proven by Leray[2] (see Sec. II) guarantees a well-posed initial value formulation.

The "conformally coupled scalar field"[3] has been used in calculations of quantum effects in the early universe,[4] but it is not obvious that it has a well-posed initial value formulation. The action (when no other matter is present) is[5]

$$S=\int \left((1/8\pi G)R+g^{cd}\phi_{;c}\phi_{;d}-\tfrac{1}{6}R\phi^{2}\right)\sqrt{-g}\,d^{4}x\,, \quad (1)$$

and the resulting field equations can be written

$$R_{ab}=\frac{-\frac{8}{3}\pi G}{1-\frac{4}{3}\pi G\phi^{2}}\left[2\phi_{;a}\phi_{;b}-\tfrac{1}{2}g_{ab}g^{cd}\phi_{;c}\phi_{;d}-\phi\phi_{;ab}\right], \quad (2)$$

$$g^{cd}\phi_{;cd}=0\,. \quad (3)$$

The second derivative of the ϕ term in (2) becomes a nondiagonal term in harmonic coordinates. Before Leray's powerful theorem applies, a diagonal hyperbolic form must be found.

Attempts to quantize general relativity[6] and attempts to regularize stress energy-momentum tensors of quantum fields propagating in curved spacetimes[7] have led investigators to consider gravitational actions involving curvature-squared terms. The most general action of this form is (for vacuum)

$$S=\int \left(\alpha R_{cd}R^{cd}+\beta R^{2}+(1/16\pi G)R\right)\sqrt{-g}\,d^{4}x\,, \quad (4)$$

where α and β are new universal constants (a Riemann-squared term can be eliminated using the Gauss–Bonet identity; the term linear in R is necessary for a proper Newtonian limit). The "higher derivative gravity" field equation which follows[8] from (4) is

$$\begin{aligned}(\alpha-2\beta)R_{;ab}&-\alpha\Box R_{ab}-(\tfrac{1}{2}\alpha-2\beta)g_{ab}\Box R\\&+2\alpha R^{cd}R_{acbd}-2\beta RR_{ab}-\tfrac{1}{2}g_{ab}(\alpha R^{cd}R_{cd}-\beta R^{2})\\&+(1/16\pi G)(R_{ab}-\tfrac{1}{2}g_{ab}R)=0\,. \quad (5)\end{aligned}$$

This equation obviously involves fourth derivatives of the metric, hence the label "higher derivative." If this is to be considered as a new, physically reasonable competitor for general relativity, its initial value formulation must be investigated.

In Sec. II, the method of harmonic coordinates as applied to general relativity is presented, because the procedures for the more complicated systems of interest here are simple extensions of the general relativity procedure. In Sec. III, this method is used to show that the conformally-coupled scalar field has a well-posed initial value problem. Section IV shows that the higher-derivative gravity equations, when treated as twenty second-order equations (not ten fourth-order equations), have a well-posed initial value problem.

In Secs. III and IV the equations of interest, when first written in harmonic coordinates, are nondiagonal systems, but equivalent diagonal hyperbolic systems are subsequently constructed. Not all quasilinear nondiagonal systems can be put in diagonal hyperbolic form. All of the nondiagonal systems considered here have a common form when naively expressed in harmonic coordinates: each of the variables in off-diagonal second-order terms also appears in second order (hyperbolically) in "its own" diagonal equation. There is a method for "turning off" such off-diagonal terms,[9] explained in Sec. III.

 0022-2488/83/071846-05$02.50

II. INITIAL VALUE FORMULATION OF GENERAL RELATIVITY

In this section, the initial value formulation of general relativity is outlined in a form that will later be extended to treat the conformally coupled scalar field and higher derivative gravity. A simple form (applying to C^∞ data and solutions) of Leray's theorem (which guarantees a well-posed initial value problem once the system of equations has been put in suitable ("diagonal hyperbolic" form) is stated. The theorem will eventually be brought to bear on each system of equations considered.

It is easily seen that four of the ten Einstein equations, when written out in a coordinate system, do not involve any second time derivatives of any metric component. These four, the (oa) components, do not specify time evolution, but are constraints that must be satisfied at all times. They are

$$G_{bc} h^b{}_a n^c = R_{bc} h^b{}_a n^c = K^b{}_{a;b} - K_{,a} , \tag{6}$$

$$G_{ab} n^a n^b = -\tfrac{1}{2}\{ \overset{3}{R} - K^2 + K_{ab} K^{ab} \} , \tag{7}$$

where n^a is the normal to the constant time surfaces, $h^a{}_b$ projects onto the surfaces, $\overset{3}{R}$ is the 3-curvature scalar on them and K_{ab} is the extrinsic curvature of the 3-surfaces, representing how they are imbedded in the spacetime. These are particular forms of the equations of Gauss and Codacci.[10] Equations (6) and (7) relate quantities determined by the local mass distribution (the left-hand sides with G_{ab} replaced by T_{ab}) to quantities determined by the metric (right-hand sides). In finding solutions from initial data, if these constraints are satisfied on the initial surface, the fully contracted Bianchi identity guarantees the constraints everywhere (once the evolution equations are solved).

Following Bruhat, define a "gauge potential"

$$F^a \equiv (1/\sqrt{-g})(\sqrt{-g} g^{ab})_{,b} = -g^{cd}\Gamma^a_{cd} . \tag{8}$$

The harmonic gauge (coordinate) condition is then

$$F^a = 0 . \tag{9}$$

The Ricci curvature can be expressed as

$$R_{ab} = -\tfrac{1}{2} g^{cd} g_{ab,cd} + \tfrac{1}{2}(g_{ac}F^c{}_{,b} + g_{bc}F^c{}_{,a}) + H_{ab} , \tag{10}$$

where H_{ab} represents terms strictly first order in the metric, whose detailed form is irrelevant to this discussion. Now, define

$$Q_{ab} = R_{ab} - \tfrac{1}{2}(g_{ac}F^c{}_{,b} + g_{bc}F^c{}_{,a})$$
$$= -\tfrac{1}{2} g^{cd} g_{ab,cd} + H_{ab} , \tag{11}$$

and choose as a field equation

$$Q_{ab} = 0 \tag{12}$$

(usually called "Einstein's equation in harmonic coordinates"). A solution of (12) is a solution of Einstein's equation only if the harmonic gauge condition (9) is also satisfied.

Regardless of whether Eq. (9) is satisfied or not, Eq. (12) [with Q_{ab} as in the second line of (11)] is a quasilinear diagonal system[9] of partial differential equations in ten unknowns g_{ab}. It is quasilinear because each equation is linear in its highest order term, and diagonal in that each equation invoves the highest order (two) derivative of only one of the unknowns. The system is "hyperbolic" in the sense that each of the second-order terms is a hyperbolic differential operator acting on one of the unknowns (as long as the metric is nonsingular and has hyperbolic signature). For such a system Leray[2] has proven a theorem, which, phrased for second-order equations and C^∞ solutions out of C^∞ data[11] is

Theorem: Given (a) a quasilinear, diagonal, second-order hyperbolic system,

$$h^{(k)ab}(x, u_{(j)}, u_{(j),c}) u_{(k),ab} + b^{(k)}(u_{(j)}, u_{(j),c}) = 0 , \tag{13}$$
$$k = 1,...,N,$$

in N unknowns $u_{(j)}$ (in n coordinates) where $h^{(k)ab}$ and $b^{(k)}$ are C^∞ in their arguments;
(b) a smooth $[(n-1)$-dimensional$]$ initial surface (region) S embedable in some n-manifold;
(c) a range of values of each of the unknowns (a subset $Y_{(j)}$ of R^n) for which all the $h^{(k)ab}$'s are Lorentz metrics (and the intersection of all the timelike cones is not empty) everywhere on S;
(d) initial data on S for the unknowns, lying in the allowed ranges $Y_{(j)}$,
a solution $u_{(j)}$ exists in an open neighborhood of S, and sufficiently small perturbations of the initial data produce small perturbations in the solution (for a precise statement, see Ref. 10, pp. 226–255), and initial data that differ only outside some region S' of S produce solutions that differ only outside of a causal future (defined by the $h^{(k)ab}$ metrics) of S'.

For any solution of the field equations generated from specified $g_{ab}, \dot{g}_{ab}$ (the dot indicating derivative in the timelike coordinate direction leading out of the initial surface), evaluating the covariant divergence of Q_{ab} [viewed as the first line of (11)] minus one half its trace implies (using the fact that the field equations are satisfied)

$$\tfrac{1}{2} g^{cd} F^a{}_{,cd} + p^a = 0 , \tag{14}$$

where p^a represents terms first order and homogeneous in F^b. This is a quasilinear diagonal hyperbolic system on F^a, so that if the initial data happens to satisfy

$$F^a = 0 , \tag{15}$$

$$\dot{F}^a = 0 \tag{16}$$

on the initial surface, then the solution necessarily has F^a zero everywhere. To get a solution in harmonic coordinates we must satisfy constraints (15) and (16) on the initial data. However, it can be seen[1] that on the initial surface, if (6), (7), (15), and the field equations in harmonic coordinates (12) are all satisfied, then (16) follows, and is not an independent constraint.

The last step in establishing that the initial value problem is well posed is the verification that the set of all constraints on the initial data is consistent (i.e., valid sets of initial data do exist). The usual procedure is to set g_{00} on the initial surface to (-1), $g_{0\alpha}$ initially to zero (these components are regarded as gauge, not geometric: they govern how the constant $\vec{x}$ lines leave the initial surface), and let (15) determine $\dot{g}_{0a}$ algebraically in terms of other components of the 4-metric and its dot. Then, Bruhat[1] details a solution of (6), (7) which fixes the conformal factor of the 3-metric and three components of $K_{\alpha\beta}$. The remaining information in the 3-metric (five "functions") and the remaining (three) components of $K_{\alpha\beta}$ are left free. $\dot{g}_{\alpha\beta}$ is determined algebraically from $K_{\alpha\beta}$ (for the choice of g_{0a} above, the components of the two are equal). After the (gauge) fixing of g_{0a}, there are eight "free functions," but specification of the coordinate system in the initial surface uses up three, and one more specifies the particular Cauchy surface in the 4-space foliation (by t) that the evolution starts from. Thus there are four functions in the initial data which specify the physical configuration, while the other eight are "gauge-fixing." Four free physical functions (divide by two to find two "degrees of freedom") are what are needed to specify the evolution of a spin-2 massless field, the graviton.

III. THE CONFORMALLY COUPLED SCALAR FIELD

The "conformally coupled scalar field equations in harmonic coordinates" [the analog of (11,12)] are

$$Q_{ab} = -\frac{1}{2} g^{cd} g_{ab,cd} - \frac{\frac{8}{3}\pi G\phi}{(1 - \frac{4}{3}\pi G\phi^2)} \phi_{,ab} + H'_{ab} = 0, \tag{17}$$

$$g^{cd}\phi_{,cd} = 0. \tag{18}$$

In this form these equations are not diagonal, as different second derivatives of the single variable ϕ appear in every equation. There is a change of variables, however, which can "turn off" some unwanted off-diagonal terms. Taking a derivative of the scalar wave equation (18) and formally treating $\phi_{,a}$ as an independent variable v_a, an equivalent set of equations to (17), (18) which are diagonal hyperbolic are

$$-\frac{1}{2} g^{cd} g_{ab,cd} - \frac{\frac{8}{3}\pi G\phi}{(1 - \frac{4}{3}\pi G\phi^2)} v_{a,b} + H'_{ab} = 0, \tag{19}$$

$$g^{cd}\phi_{,cd} = 0, \tag{20}$$

$$g^{cd} v_{a,cd} + g^{cd}{}_{,a} v_{c,d} = 0. \tag{21}$$

With this system, initial data for v_a is not free at all,

$$v_a|_s = (\dot{\phi}, \phi_{,\alpha}), \tag{22}$$

$$\dot{v}_a|_s = (\ddot{\phi}, \dot{\phi}_{,\alpha}) \tag{23}$$

($\ddot{\phi}$ is deduced by solving Eq. (20) on the surface). Leray's theorem guarantees that solutions (with the desired causal evolution and perturbation stability) exist for appropriately specified data. Then, given a solution, it is obvious that v_a and $\phi_{,a}$ satisfy the same hyperbolic differential equation (21), and have the same initial data, so

$$v_a = \phi_{,a} \tag{24}$$

everywhere in the solution. The conformally coupled scalar field has a well-posed initial value problem. The constraints on the initial data are very similar to those for general relativity, but the detailed form is changed by the $\phi_{;ab}$ contribution to R_{ab} [Eq. (2)] in the Gauss and Codacci equations (6), (7). This involves $\ddot{\phi}$ and first derivatives of the metric. $\ddot{\phi}$ can be elliminated by solving the scalar wave equation on the initial surface. Then, since the Gauss and Codacci equations are second order in the metric, their basic form and method of solution are unchanged. There is one more degree of freedom (which is ϕ).

The diagonal system was found using a simple test of whether the procedure of taking extra derivatives can turn off all off-diagonal terms in a particular nondiagonal system of equations.[9] If there exists a set of nonnegative integers $\{s_i\}$ [the (formal) index numbering the equations], and $\{t_j\}$ (the index numbering the unknowns) such that the principal part (terms of the highest order) in unknown u_i in equation j is of order $(s_i - t_j)$ when i equals j, and of lower order when i does not equal j, then the differentiation procedure can produce an equivalent diagonal system. Such a set of integers for the system (17), (18) is

$$s_{\text{grav eq},ab} = 3, \tag{25}$$

$$s_{\text{scalar eq}} = 2, \tag{26}$$

$$t_{g_{ab}} = 1, \tag{27}$$

$$t_\phi = 0. \tag{28}$$

Any equation with s less than the maximum value (here, the scalar wave equation) must be differentiated to produce a new equation to be added to the system, and any variable with t less than the maximum value must be differentiated to produce a (formally) new variable. This done, the offending off-diagonal terms can be (formally) replaced by terms lower order in the new variable(s). The old diagonal terms are unchanged.

This method is not highly specialized to the conformally coupled scalar field. For example, the coefficient 1/6 in Eq. (1) can be replaced by ξ, a nonminimal-coupling constant. The resulting field equations can be written in a form in which no new second-order terms appear, and the same method shows there is a well-posed initial value formulation, for any value of ξ. Further, Brans–Dicke theory[12] has equations of form similar, as originally posed, to those of the conformally-coupled scalar theory above. A well-posed initial value formulation of the original Brans–Dicke equations can be found directly using the method of this section, but it is well known that there is a transformation of variables that produces a diagonal hyperbolic set of equations for Brans–Dicke theory, except in the special case of ω (the Brans–Dicke constant) equal $-3/2$, when the transformation seems to break down. In this special case, the Brans–Dicke scalar field equation is conformally invariant (using a different change of variables), and the complete set of equations is equivalent to the set of equations derivable from Eq. (1) without the first term ($R/8\pi G$). The problem with this conformal-scalar version of Brans–Dicke theory is that the trace of the gravitational equation and the scalar field equation, which for all other values of ω are two distinct equations relating the scalar field to R, for this value are the same equation (if there is any other matter, appearing in the gravitational equations as T_{ab}, the difference of the two equations is T equal zero). With this reduction by one of the number of independent field equations, one function (e.g., ϕ) becomes totally unconstrained, and can be specified freely over a 4-volume before finding the solution (for the rest of the variables) in the 4-volume. Thus the ω equal $-3/2$ Brans–Dicke theory does not have a well-posed initial value formulation as defined by the Introduction.

IV. HIGHER DERIVATIVE GRAVITY

Stelle[8] has studied some classical aspects of higher derivative gravity (and reviewed earlier work). He has shown that linearized higher derivative gravity can be viewed as consisting of the (usual) massless graviton coupled to a massive (m_0) scalar field

$$m_0^2 = \frac{1}{32\pi G(3\beta - \alpha)} \tag{29}$$

and a massive (m_2) spin-2 field

$$m_2^2 = \frac{1}{16\pi G\alpha}. \tag{30}$$

It is not difficult to see that in the linearized theory the scalar and massive spin-2 field can be considered to be the linearized Ricci scalar and the linearized trace-free part of the Ricci tensor, respectively. Knowing this, a more natural, and more useful[13] form of the field equation is (for m_0 nonzero and α nonzero[14])

$$(\Box - m_0^2)R = 0, \tag{31}$$

$$\begin{aligned}(\Box - m_2^2)\widetilde{R}_{ab} &+ \tfrac{1}{6}(2m_2^2/m_0^2 + 1)R\widetilde{R}_{ab} + 2\widetilde{R}^{cd}C_{acbd} \\ &- 2\widetilde{R}_a{}^c\widetilde{R}_{bc} + \tfrac{1}{2}g_{ab}\widetilde{R}^{cd}\widetilde{R}_{cd} \\ &+ \tfrac{1}{3}(m_2^2/m_0^2 - 1)(R_{;ab} - \tfrac{1}{4}g_{ab}m_0^2R) = 0,\end{aligned} \tag{32}$$

where C_{acbd} is the Weyl curvature and $\widetilde{R}_{ab}$ is the trace-free part of the Ricci curvature

$$\widetilde{R}_{ab} = R_{ab} - \tfrac{1}{4}g_{ab}R \tag{33}$$

[Eq. (32) is trace-free].

In the linearized version of higher derivative gravity, R and $\widetilde{R}_{ab}$ look like a massive scalar and massive spin-2 field independent of the spin-2 null graviton only if the linearized field equations are treated as second-order equations with R and $\widetilde{R}_{ab}$ independent variables and the (linearized) Einstein equation is considered as an independent equation on g_{ab}, with R and $\widetilde{R}_{ab}$ terms as sources. The linear initial value formulation is then obvious. In the same style, the higher derivative gravity equations (31) and (32) (when both α and β are nonzero) can be considered as 20 second-order equations in independent variables g_{ab}, $\mathscr{R}$, $\widetilde{\mathscr{R}}_{ab}$. (The fully contracted Bianchi identity appears as a constraint below.) In harmonic coordinates,

$$-\tfrac{1}{2}g^{cd}g_{ab,cd} + H_{ab} = \widetilde{\mathscr{R}}_{ab} + \tfrac{1}{2}g_{ab}\mathscr{R}, \tag{34}$$

$$\begin{aligned}g^{cd}\widetilde{\mathscr{R}}_{ab,cd} &+ \tfrac{1}{3}(m_2^2/m_0^2 - 1)\mathscr{R}_{,ab} \\ &- \widetilde{\mathscr{R}}_{cd}g^{cf}g^{ed}(g_{ef,ab} + g_{ab,ef} - g_{ae,fb} - g_{fb,ae}) \\ &- \tfrac{1}{6}(\widetilde{\mathscr{R}}_{ab}g^{ef}g^{cd}g_{ef,cd}) + J_{ab} = 0,\end{aligned} \tag{35}$$

$$g^{cd}\mathscr{R}_{,cd} - m_0^2\mathscr{R} = 0 \tag{36}$$

[J_{ab} represents lower order terms in (35).] Equations (34)–(36) are analyzed by viewing $\mathscr{R}$ and $\widetilde{\mathscr{R}}_{ab}$ as unknowns, not necessarily equal to the curvature of g_{ab}. Unless constraints on initial data discussed below are in fact satisfied, $\widetilde{\mathscr{R}}_{ab}$ and $\mathscr{R}$ of a solution of (34)–(36) will not satisfy Bianchi's identities, and the true Ricci curvatures $\widetilde{R}_{ab}$ and R calculated from the metric g_{ab} of the solution will be entirely different. Hence the use of script symbols in (34)–(36) to distinguish them from the true curvature.

Again, the system is not diagonal, but the set of integers

$$(s_{(34)ab}, s_{(35)ab}, s_{(36)}) = (2,3,2), \tag{37}$$

$$(t_{(g)ab}, t_{(\widetilde{\mathscr{R}})ab}, t_{(\mathscr{R})}) = (0,1,0) \tag{38}$$

satisfies the test detailed in the previous section, implying that equations (34) and (36) and variables g_{ab} and $\mathscr{R}$ must be differentiated and added to the system. Thus a hyperbolic set equivalent to (34)–(36) is (34) and (36) together with

$$\begin{aligned}g^{cd}\widetilde{\mathscr{R}}_{ab,cd} &+ \tfrac{1}{3}(m_2^2/m_0^2 - 1)V_{a,b} \\ &- \widetilde{\mathscr{R}}_{cd}g^{cf}g^{de}(h_{efa,b} + h_{abe,f} - h_{aef,b} - h_{fba,e}) \\ &- \tfrac{1}{6}\widetilde{\mathscr{R}}_{ab}g^{ef}g^{cd}h_{efc,d} + J_{ab} = 0,\end{aligned} \tag{39}$$

$$g^{cd}V_{a,cd} + g^{cd}{}_{,a}V_{c,d} - m_0^2V_a = 0, \tag{40}$$

$$\begin{aligned}\tfrac{1}{2}g^{cd}h_{abe,cd} &+ \tfrac{1}{2}g^{cd}{}_{,e}h_{abc,d} + K_{abe} \\ &- \widetilde{\mathscr{R}}_{ab,e} - \tfrac{1}{4}g_{ab,e}\mathscr{R} - \tfrac{1}{4}g_{ab}\mathscr{R}_{,e} = 0,\end{aligned} \tag{41}$$

where K_{abe} is first order in g_{ab} and h_{abc}, and the initial data for V_a and h_{abc} is not specified freely at all, but so that, in the resulting solution (in analogy to the Sec. III discussion),

$$V_a = \mathscr{R}_{,a}, \tag{42}$$

$$h_{abc} = g_{ab,c}. \tag{43}$$

While appropriate (to the theorem) initial data does uniquely specify a solution of (34)–(36) in a neighborhood of S, new constraints on the new variables $\mathscr{R}$, $\widetilde{\mathscr{R}}_{ab}$ beyond those on g_{ab} for general relativity must be satisfied to obtain a solution of the higher derivative gravity equations with $\mathscr{R}$ and $\widetilde{\mathscr{R}}_{ab}$ equal to curvature of g_{ab}. Let

$$\Delta_{ab} = G_{ab} - \widetilde{\mathscr{R}}_{ab} + \tfrac{1}{4}g_{ab}\mathscr{R}, \tag{44}$$

where G_{ab} is the true Einstein curvature associated with g_{ab} of a solution of (34)–(36) (Δ_{ab} is the difference between G_{ab} and what should be G_{ab}), then similar manipulations to those that produced (14) on F^a here [out of (34) and (35)] imply instead

$$\tfrac{1}{2}g^{cd}F^a{}_{,cd} + p^a = g^{ab}\Delta_b{}^c{}_{;c} \tag{45}$$

and

$$\begin{aligned}g^{cd}(\Delta_a{}^b{}_{;b})_{,cd} &+ (\tfrac{2}{3} + m_2^2/3m_0^2)[-\Delta_a{}^cF^e{}_{,ec} - \tfrac{1}{2}F^e{}_{,e}F^d{}_{,da}] \\ &+ \tfrac{1}{2}G^{cd}[-g_{ae}F^e{}_{,dc} + g_{ac}F^e{}_{,ed} + g_{ce}F^e{}_{,da} - g_{cd}F^e{}_{,ea}] \\ &- \tfrac{1}{2}\Delta^{cd}[g_{ce}F^e{}_{,da} + g_{de}F^e{}_{,ca} - g_{cd}F^e{}_{,ea}] + L_a = 0,\end{aligned} \tag{46}$$

where L_a represents terms of first order or less, homogeneous in either F^a or $\Delta_a{}^b{}_{;b}$. The now familiar procedure produces a diagonal hyperbolic system out of (45), (46) by appending the derivative of (45) as a new equation and treating $F^a{}_{,e}$ as a new "independent" variable. Thus, if the initial data satisfies

$$F^a = 0, \tag{47}$$

$$\dot{F}^a = 0, \tag{48}$$

$$\Delta_a{}^b{}_{;b} = 0, \tag{49}$$

$$(\Delta_a{}^b{}_{;b})^{\cdot} = 0 \tag{50}$$

on S, then F^a and $\Delta_a{}^b{}_{;b}$ are zero everywhere in the solution, and the detailed form of the equation "$\Delta_a{}^b{}_{;b}$ equals zero" is the fully contracted Bianchi identity on $\mathscr{R}$ and $\widetilde{\mathscr{R}}_{ab}$. Of course, the Gauss and Codacci constraints (6), (7) [where the left-hand sides (G_{ab}) are calculated from $\mathscr{R}$ and $\tilde{\mathscr{R}}_{ab}$ data, while the right-hand sides are calculated from metric data] must still be satisfied on the initial surface. Equations (49) and (50) are the new constraints on the data. If all the constraints are satisfied, the solution of (34)–(36) is a solution of the higher derivative gravity equations, in harmonic coordinates, with

$$\tilde{R}_{ab} = \tilde{\mathscr{R}}_{ab} \,, \tag{51}$$

$$R = \mathscr{R} \,. \tag{52}$$

The last task is verification that the constraint system [the constraints for general relativity when R_{ab} not zero plus (49), (50)] on the initial surface is consistent. Again, the $\dot{F}^a$ equation (48) follows from the field equations "in harmonic coordinates," and (6), (7), (47) on the initial surface. Again, the F^a constraint (47) can be used to fix $\dot{g}_{0a}$ algebraically. Then, formally, (6), (7), (49), (50) are coupled together, requiring simultaneous solution. For use of the Cauchy–Kovalevskaya theorem,[15] considering analytic solutions generated by analytic data, the characteristic determinant of the system is a product of 2, one for the Gauss–Codacci equations (well known and well behaved), the other for the "Bianchi constraints." It is most convenient to regard $\tilde{\mathscr{R}}_{00}$ as the nondynamic component of $\tilde{\mathscr{R}}_{ab}$ (and eliminate it and $\dot{\tilde{\mathscr{R}}}_{00}$ from all expressions by enforcing trace-freeness) and then to decompose the space-space parts of $\tilde{\mathscr{R}}_{ab}$ and $\dot{\tilde{\mathscr{R}}}_{ab}$ on the initial surface into 3-trace and 3-trace-free parts:

$$\mathscr{A} = \overset{3}{g}{}^{\gamma\delta}\tilde{\mathscr{R}}_{\gamma\delta} \,, \tag{53}$$

$$\mathscr{A}_{\alpha\beta} = \tilde{\mathscr{R}}_{\alpha\beta} - \tfrac{1}{3}\overset{3}{g}_{\alpha\beta}\mathscr{A} \,, \tag{54}$$

$$\mathscr{B} = \overset{3}{g}{}^{\gamma\delta}\dot{\tilde{\mathscr{R}}}_{\gamma\delta} \,, \tag{55}$$

$$\mathscr{B}_{\alpha\beta} = \dot{\tilde{\mathscr{R}}}_{\alpha\beta} - \tfrac{1}{3}\overset{3}{g}_{\alpha\beta}\mathscr{B} \,. \tag{56}$$

Then the undotted Bianchi constraint (49) serves merely to specify $\dot{\tilde{\mathscr{R}}}_{0\alpha}$ and $\mathscr{B}$ algebraically in terms of $\mathscr{R}$, $\dot{\mathscr{R}}$, $\mathscr{A}$, $\mathscr{A}_{\alpha\beta}$, $\mathscr{B}_{\alpha\beta}$, $\tilde{\mathscr{R}}_{0\alpha}$. The "Bianchi-dot" constraint (50) becomes a system of four second-order equations, which can be regarded as fixing $\mathscr{A}$ and $\tilde{\mathscr{R}}_{0\alpha}$ on the initial surface (given data at a point, analytically) leaving $\mathscr{R}$, $\dot{\mathscr{R}}$, $\mathscr{A}_{\alpha\beta}$, $\mathscr{B}_{\alpha\beta}$ free. The characteristic determinant of these four equations (which is then multiplied by the other determinant) is nonsingular so the grand characteristic determinant is nonsingular, and the system of constraints is consistent. Higher derivative gravity does possess a well-posed initial value problem.

The satisfaction of all constraints leaves free all the same initial data functions seen in general relativity plus two scalar fields (corresponding to the "massive scalar field" R) and two trace-free symmetric 3-tensors (five free "functions" each) corresponding to $\tilde{\mathscr{R}}_{ab}$. Five degrees of freedom are associated with a massive spin-2 field, so $\tilde{\mathscr{R}}_{ab}$ retains some appearance of a massive spin-2 field in the full nonlinear theory, as it does in the linearized theory.

To recapitulate, higher derivative gravity does possess a well-posed initial value problem. The initial data is g_{ab}, $\mathscr{R}$, $\tilde{\mathscr{R}}_{ab}$ and their time (coordinate) derivatives on an initial surface, satisfying the following (surface) constraints: the Gauss and Codacci equations (6), (7) (with G_{ab} calculated from the data for $\mathscr{R}$ and $\tilde{\mathscr{R}}_{ab}$), the harmonic coordinate condition (47), and the "Bianchi constraints" (49), (50). Given the data, there exists a unique solution of the higher derivative gravity equations in a neighborhood of the initial surface. The solution evolves causally and is as stable to perturbations of the initial data as general relativity is.

ACKNOWLEDGMENTS

Thanks particularly to R. M. Wald, and also to Y. Choquet-Bruhat, R. Geroch, and P. Panagaden for help in discussion.

This research was supported, in part by NSF Grant PHY 80 26043 to The University of Chicago, and in part by a NSERC (Canada) scholarship.

[1]Y. Bruhat, in *Gravitation, an Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).

[2]J. Leray, *Hyperbolic Differential Equations* (Institute for Advanced Study, Princeton, NJ, 1953).

[3]R. Penrose, in *Relativity, Groups and Topology*, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1964).

[4]E.g., L. Parker, Phys. Rev. **183**, 1057 (1969); B. L. Hu and L. Parker, Phys. Rev. D **17**, 933 (1978), and references therein; M. V. Fischetti, J. B. Hartle, and B. L. Hu, Phys. Rev. D **20**, 1757 (1979).

[5]Convention: comma denotes coordinate derivative, semicolon denotes covariant derivative (Misner, Thorne, and Wheeler "$+\ -\ -$" convention) metric signature $(-\ +\ +\ +)$,

$$R^a{}_{bcd} = \Gamma^a{}_{bc,d} + \cdots, \quad R_{bd} = R^c{}_{bcd} \,,$$

$$\Box(Q^{\cdots}{}_{\cdots}) = g^{cd}(Q^{\cdots}{}_{\cdots})_{;cd} \,,$$

$$c = \hbar = 1 \,.$$

[6]S. Deser and P. van Nieuwenhuizen, Phys. Rev. D **10**, 401 (1974); **10**, 411 (1974); S. Deser, P. van Nieuwenhuizen, and H. S. Tsao, Phys. Rev. D **10**, 3337 (1974).

[7]G. T. Horowitz and R. M. Wald, Phys. Rev. D **17**, 414 (1978); R. M. Wald, Phys. Rev. D **17**, 1477 (1978).

[8]K. S. Stelle, Gen. Relativ. Gravit. **9**, 353 (1978).

[9]See, e.g., Y. Choquet-Bruhat, C. DeWitt-Moretee, and M. Dillard-Bleick, *Analysis, Manifolds and Physics* (North-Holland, Amsterdam, 1977), pp. 463–442.

[10]See, e.g., S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge University, Cambridge, 1972), pp. 44–47.

[11]A general statement (without proof) in modern terminology for arbitrary order is given in Ref. 9. A statement and proof for second-order systems but finite differentiability of initial data and solution is given in Ref. 10, pp. 226–255.

[12]C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961); R. H. Dicke, Phys. Rev. **125**, 2163 (1962).

[13]The form of Eq. (31) allows deduction of some black hole theorems. Using the arguments of J. D. Bekenstein [Phys. Rev. D **5**, 1239 (1972); **5**, 2403 (1972)], it is easy to see that the Ricci scalar R must be zero everywhere in the exterior of a black hole solution, either static or rotating, and $t\phi$-inversion symmetric, of the higher derivative gravity equations. This makes most black hole exteriors distinct from most stellar exteriors, since, in this theory, all realistic extended mass distributions must have nonzero R in their exteriors [see E. Pechlaner and R. Sexl, Comm. Math. Phys. **2**, 165 (1966)].

[14]Section IV's analysis does not apply to the case of α equal to zero (m_2^2 "diverges" and ceases to play a role). The equations are then simpler in that no derivatives of $\tilde{\mathscr{R}}_{ab}$ appear. The initial value problem can be shown to be qualitatively the same as that of the conformally coupled scalar field. In the case of α equal to 3β (m_0^2 "diverges") no derivatives of R appear but the analysis remains qualitatively that of Sec. IV.

[15]See, e.g., G. N. Polzhiy, *Equations of Mathematical Physics*, transl. Scripta Technica Inc. (Hayden, Rochelle Park, NJ, 1967), pp. 25–34.

The local nonsingularity theorem

Robert Geroch[a)]
Enrico Fermi Institute, 5630 Ellis Avenue, Chicago, Illinois 60637

(Received 30 November 1982; accepted for publication 10 February 1983)

It is proven that, for a certain class of hyperbolic systems (a class which includes Einstein's equation), sufficiently small initial data on a bounded patch of initial surface generates a solution nonsingular in the region determined by that initial data. This theorem is virtually a corollary of the boost theorem. Various consequences and possible generalizations are discussed.

PACS numbers: 04.90. + e, 02.30. + g

1. INTRODUCTION

Christodoulou and O'Murchadha[1] have recently proven what is called the boost theorem. The theorem has the following general structure. Consider, in Minkowski space–time, a system of second-order hyperbolic equations which falls within a certain, rather broad, class of such equations. Consider sufficiently smooth initial data for this system, specified at $t = 0$ in the Minkowski space–time. This initial data can be evolved to a solution of the hyperbolic system which remains nonsingular for at least a certain distance into the past and future from the initial surface. The boost theorem now asserts that, provided the initial data is sufficiently well-behaved asymptotically, this "certain distance" through which nonsingular evolution is possible is such that the region of smooth evolution includes boosts of the original initial-data surface. That is, the farther out on the initial surface one begins, the longer in time one must wait before encountering a singularity.

The class of hyperbolic systems to which this theorem applies includes, in particular, Einstein's equation (suitably formulated in a certain gauge), both in the source-zero case and in the presence of certain "hyperbolic" sources.

We here introduce a reformulation of the boost theorem. It asserts essentially that, given in a bounded region smooth initial data sufficiently close to the zero data, then this initial data must so evolve to be nonsingular in the entire domain of dependence of that region. That is, the solution must remain nonsingular insofar as it is determined by the given initial data. This "local nonsingularity theorem" is, by an elementary argument, a consequence of the original boost theorem.

The local theorem has a number of attractive features. First, it is somewhat simpler to state, for it avoids the necessity of maintaining control over global behavior. The statement in the case of Einsten's equation is particularly natural, for one avoids reference to boosts, which of necessity are with respect to a flat background metric related to the space–time metric by gauge conditions. Further, from the local nonsingularity theorem one can, by a simple geometrical argument, recover the original boost theorem. Thus, in effect, the boost theorem is separated into its "local, analytical" part (represented by the local version) and its "global, geometrical" part (represented by the argument which leads from the local to the global). This appears to be a natural separation. For example, there are, as it turns out, numerous ways to "globalize" the local nonsingularity theorem. There result variants of the boost theorem, in which there is modified both the required asymptotic behavior of the initial data and, correspondingly, the assertion as to how far into the future nonsingular evolution must be possible. Up to certain limits, any choice of the latter can be achieved by some choice of the former. The local nonsingularity theorem can be further simplified in the case of Einstein's equation, essentially as a consequence of the fact that the fields in this case have physical dimension. Finally, the local theorem suggests, as conjectures, a number of generalizations.

The nonsingularity theorem complements the well-known singularity theorems[2] for general relativity.

In Sec. 2, we summarize the statement of the boost theorem. In Sec. 3, we introduce the local nonsingularity theorem for the general hyperbolic system. We prove it, relate it back to the boost theorem, and discuss the resulting variants of the boost theorem. In Sec. 4, we consider the case of Einstein's equation. This differs from the general case in that the fields have physical dimension, which one may exploit, and in the critical role played by gauge. Several generalizations are conjectured in this section as well as in the conclusion.

2. THE BOOST THEOREM

We summarize the boost theorem. The details, including proofs and related results, may be found in Ref. 1.

Fix four-dimensional Minkowski space–time, with its flat metric η_{ab} and derivative operator ∇_a. Consider, on this space–time, a hyperbolic equation of the form

$$\gamma^{ab} = (u,\nabla u)\nabla_a \nabla_b u = \beta(u,\nabla u) \tag{1}$$

on the function u. Here, $\gamma^{ab}(u,\nabla u)$ and $\beta(u,\nabla u)$ are given functions smooth in their arguments. The former is required to be of Lorentz signature for all $(u,\nabla u)$, and such that $\gamma^{ab}(0,0) = \eta^{ab}$. Further, $\beta(u,\nabla u)$ is required to "vanish when its arguments vanish" in the following stronger sense: $\beta(u,\nabla u)$ must be bounded by a smooth multiple of $|u|^p|\nabla u|^q$, where p and q are nonnegative integers with $p + q \geqslant 2$. Note, e.g., that β at least quadratic in either argument will do. Note also that these conditions ensure that $u = 0$ is always a solution of (1).

Denote by S the hyperplane $t = 0$ in this space–time. Consider on S smooth initial data, $(u_0,\dot u_0)$, for Eq. (1). The

[a)] Supported by the NSF.

 0022-2488/83/071851-08$02.50

boost theorem asserts that small initial data can be evolved, via (1), far into the future. We must specify what "small" and "far" are to mean.

Measure the "size and asymptotic behavior" of data $(u_0, \dot{u}_0)$ by means of the number

$$\|(u_0,\dot{u}_0)\|^2 = \sum_{s=0}^{5}\int_S (1+r^2)^{s+\delta}(D_{a_1}\cdots D_{a_s}u_0)(D^{a_1}\cdots D^{a_s}u_0) + \sum_{s=0}^{4}\int_S (1+r^2)^{s+\delta+1}(D_{a_1}\cdots D_{a_s}\dot{u}_0)(D^{a_1}\cdots D^{a_s}\dot{u}_0). \tag{2}$$

Here, h_{ab} is the η-induced metric on S (used to raise and lower indices), D_a is the corresponding (flat) derivative operator on S, and r is h distance from some fixed origin in S. Further, δ is a certain number in the interval $(-\frac{3}{2},1)$ depending only on the form of the β in (1), and then only on its behavior near $u=0$ and $\nabla u = 0$. (In the most common case, with β at least quadratic in ∇u, δ can be chosen as close as one wishes to the lower end of this range.) Thus, $\|(u_0,\dot{u}_0)\|^2$ looks to the values of u_0 and its first five spatial derivatives, as well as $\dot{u}_0$ and its first four spatial derivatives. In order that this norm be finite, these values must vanish asymptotically sufficiently quickly with r that each of the 11 integrals in (2) converges. Note that the higher derivatives are required to vanish more quickly than the lower. In order that the norm be small, these values must be small "on the average." The "$1 + r^2$" in (2) is so chosen to approach r^2 for large r (and thus impose the correct asymptotic conditions on the data), and yet be bounded below for small r (and thus accommodate the terms in which the power to which it is raised is negative). We remark that, as a consequence of the Sobelev inequality,[3] the norm (2) bounds the numerical values, at points of S, of u_0 and its first three spatial derivatves and $\dot{u}_0$ and its first two.

Initial data $(u_0, \dot{u}_0)$ can, of course, be involved to obtain a solution of Eq. (1). But, since the equation can be nonlinear, the evolution will in general terminate at singularities. A solution u of (1) will be said to *admit boost* θ $[\in(0,1)]$ if it is defined and nonsingular at least in the region

$$t^2 < \theta(1+r^2) \tag{3}$$

of the Minkowski space–time. That is, the solution is required to be nonsingular in a region which includes the initial surface S, and also all boosts of S, up to speeds not exceeding θ, about its origin.

The boost theorem[1] now asserts: Given any $\theta\in(0,1)$, there exists a number $\epsilon > 0$ such that any data with $\|(u_0,\dot{u}_0)\|^2 < \epsilon$ evolves to a solution of (1) admitting boost θ. It is further asserted that one can, again by demanding that the norm (2) on the data be small, ensure that the values of the solution u and of its gradient $\nabla_a u$ will be small at all points of the region (3).

That some such result should hold—at least for *some* choice of the norm on the initial data—is perhaps not altogether surprising. The zero data evolves to the zero solution, which, of course, has no singularities. By a judicious choice of norm, one can require that data be close—in an intricate a sense as one wishes—to the zero data. One might expect thereby to be able to guarantee that the corresponding solution will remain small through a long time evolution—and in particular that it will remain nonsingular in a region including boosts of the initial surface. The key feature of the boost theorem is that it provides a certain explicit expression, (2), for the necessary "closeness" of the initial data to zero.

The theorem also applies with the single function u above replaced by m functions, $u^1,\ldots,u^m$, and with Eq. (1) replaced by m equations, one on each of the u^i. The γ^{ab}'s and β's, of course, can now be different in the different equations, and those in the ith equation can depend on the other u^j and their gradients. The initial data now consist of the values and first time derivatives, on S, of all the u^i. The norm (2) is modified to be the sum of the corresponding norms for each of the u^i. The theorem also applies to dimensions other than four [signature $(-,+,+,\ldots,+)$]. It is only necessary to modify the norm (2): For dimension n, increase the upper limits in the sums in (2) by $(n-4)/2$ (rounded downward), and decrease δ by $(n-4)/2$. The theorem also holds under somewhat weaker smootheness conditions on the initial data. Finally, the conclusion of the second part of the theorem (smallness of the resulting solution u) is actually much stronger than indicated above. We have only stated the conclusion we shall need.

To apply the theorem to the case of Einstein's equation, say with vanishing sources, one introduces a harmonic coordinate system, and chooses for the u's the components of the difference between the space–time metric and the flat metric adapted to that coordinate system. Then Einstein's equation splits into two sets of equations, one a hyperbolic system of the form (1), and the other a system of constraint equations on the initial data. Temporarily ignoring the constraints, one applies the boost theorem to the hyperbolic system. In this case, δ may be chosen arbitrarily close to $-\frac{3}{2}$, the lower end of its range. Further, the upper limits in the sums in (2) may in this case be reduced from 5 and 4 to 4 and 3, respectively. [This is a consequence of the particular form Einstein's equation takes, in more detail, of the fact that here $\gamma^{ab}(u,\nabla u)$ does not depend on ∇u.] One thus concludes that data for the hyperbolic system (whether or not the constraints are satisfied) can be evolved to admit large boosts—provided only that the norm of the data is sufficiently small. This conclusion therefore holds in particular when the constraints are satisfied, i.e., for the full Einstein equation. Note that the boosts through which evolution is guaranteed are those with respect to the flat background metric, and not the curved space–time metric (for which "boost" would anyway be difficult to define). This fact is not a serious defect, however, for the second part of the boost theorem allows one to ensure that the space–time and background metrics will be within any preassigned amount at each point, merely by requiring that the norm on the orignal data be sufficiently small. Boosts for the background metric will then be "approximately boosts" for the space-time metric. Finally, there can, of course, be included with Einstein's equation sources subject to suitable hyperbolic equations.

3. THE LOCAL NONSINGULARITY THEOREM

The boost theorem itself is certainly a type of "nonsingularity theorem." We here obtain a theorem which captures the "local" aspects of the boost theorem, omitting, e.g., its references to asymptotic structure and boosts.

Consider again the hyperbolic equation (1) in Minkowski space–time. Fix a compact, 3-submanifold with boundary C of the initial surface S $(t = 0)$. For example, C might be a closed 3-ball in S. Consider again smooth initial data $(u_0, \dot{u}_0)$ for Eq. (1), but now given *only* on the region C. Choose any extension of this data from C to all of S, and take the maximum evolution of the resulting data on S. Finally, take the domain of dependence[4] of C, using the metric $\gamma^{ab}(u,\nabla u)$, in this maximal evolution. We say that the data $(u_0,\dot{u}_0)$ given originally on C has *nonsingular evolution* if this domain of dependence is compact in the Minkowski space–time. The idea of this definition is the following. With respect to the flat metric η^{ab} in Minkowski space–time, the domain of dependence of the compact region C will, of course, be compact. With respect to metric $\gamma^{ab}(u,\nabla u)$ in the maximum evolution, the domain of dependence can fail to be compact only if it "terminates," in the past or the future from C, at a singularity of u. (Noncompactness would arise because such singular points must be omitted from the maximal evolution.) But, since these singular points are abutted by the γ-domain of dependence of C, their occurrence is already a consequence of the data, $(u_0,\dot{u}_0)$, given originally in C. Indeed, one checks that the definition is independent of the choice of extension of the data from C to all of S. In short, initial data on C has nonsingular evolution if there arises no singularity predicted entirely by that data.

Again fix region C and initial data $(u_0, \dot{u}_0)$ on C as above. We define the norm of the data by

$$\{(u_0,\dot{u}_0)\}^2 = \sum_{s=0}^{5} d^{\,2(s+\delta)}\int_C (D_{a_1}\cdots D_{a_s}u_0)(D^{a_1}\cdots D^{a_s}u_0) + \sum_{s=0}^{4} d^{\,2(s+\delta+1)}\int_C (D_{a_1}\cdots D_{a_s}\dot{u}_0)(D^{a_1}\cdots D^{a_s}\dot{u}_0), \quad (4)$$

where d is the diameter of C, the maximum distance between any two of its points. This will be recognized as the same norm as (2), except that the integrals are only over C (i.e., only where u_0 and $\dot{u}_0$ are defined), and the "$1 + r^2$" is replaced by "d^2." The expression (4), in contrast to (2), has a definite dimension.

We now state the local nonsingularity theorem:

Theorem 1: Fix compact C as above. Then there exists $\epsilon > 0$ such that any data $(u_0, \dot{u}_0)$ on C with norm (4) less than ϵ has nonsingular evolution.

Proof: Let, for contradiction, $(u_i, \dot{u}_i)$ $(i = 1, 2, ...)$ be data on C, each with singular evolution, and with $\{(u_i, \dot{u}_i)\}^2 \to 0$. Assume that each $(u_i, \dot{u}_i)$ has support in the interior of C. This is no loss of generality, for, fixing any compact region $\hat{C}$ containing C in its interior, there exists[5] a number λ (>1) such that any initial data on C can be extended to data on $\hat{C}$ with (i) the support of the extended data in the interior of $\hat{C}$ and (ii) the norm (4) of the extended data not exceeding λ times the norm of the original data. Further, choosing a subsequence if necessary, we may assume that $\{(u_i, \dot{u}_i)\}^2 \leqslant e^{-i}$ for all i.

Fix a point p of C, and also a sequence of points y_i of S whose distances from the origin are given by

$$|y_i| = d_i. \quad (5)$$

Denote by C_i that region of S obtained by translating C in S until its point p is located at point y_i. The C_i are thus copies of C. They will not overlap, for the y_i "go to infinity quickly enough," by (5). Finally, denote by $(u_0, \dot{u}_0)$ the smooth initial data on S determined as follows: At points outside of all the C_i, $u_0 = \dot{u}_0 = 0$, and, at points of C_i, $(u_0, \dot{u}_0)$ is the originally given data $(u_i, \dot{u}_i)$ (translated, similarly, so its support is in C_i).

We next bound the norm (2) of $(u_0, \dot{u}_0)$. First note that (2) can be represented as a sum of contributions from the C_i, for $u_0 = \dot{u}_0 = 0$ elsewhere. The distance, r, of any point of C_i from the origin must be between $|y_i| - d$ and $|y_i| + d$, and so from (5) we have

$$d^2[d^{-2} + (i-1)^2] \leqslant 1 + r^2 \leqslant d^2[d^{-2} + (i+1)^2]. \quad (6)$$

But (6) bounds the "$1 + r^2$" of (2) by the "d^2" of (4). We conclude that the contribution of region C_i to (2) is bounded by $f(i)\{(u_i, u_i)\}^2$, where f is some positive function whose values increase no faster than a power of i. Hence,

$$\|(u_0,\dot{u}_0)\|^2 \leqslant \sum_i f(i)\{(u_i,\dot{u}_i)\}^2 \leqslant \sum_i f(i)e^{-i} \quad (7)$$

is finite.

So, omitting a finite number of initial C_i in the construction of $(u_0, \dot{u}_0)$, we may cause $\|(u_0, \dot{u}_0)\|^2$ to be as small as we wish. Do so, and apply the boost theorem, to conclude: There exists a solution of u of (1) (i) having $(u_0, \dot{u}_0)$ as its initial data, (ii) admitting boost θ for some θ, and (iii) having u and ∇u less than a predetermined amount. Now the region (3) certainly includes, for all sufficiently large i, the η-domain of dependence of C_i (for these domains of dependence extend into the past and future by time not exceeding d). So, by (iii), this region also includes the $\gamma(u,\nabla u)$-domains of dependence. We conclude that, for all sufficiently large i, $(u_i, \dot{u}_i)$ on C has nonsingular evolution. This contradiction establishes the theorem.

Thus, all sufficiently small initial data on a fixed compact region C must have nonsingular evolution. Note that the proof is entirely elementary, and makes use of rather less than the full force of the boost theorem. Note also that the conclusion of Theorem 1 (although not the value of ϵ) is independent of the detailed form of the norm (4). This form, suggested by dimensional considerations, was selected for later convenience.

How does the "required degree of smallness of the initial data"—the ϵ given by the theorem—depend on the choice of the region C? We first remark that, for C with highly contorted boundary, the ϵ must in general be small. This effect arises because, for such a C, the λ in the first paragraph of the proof (the factor by which the norm of the data must be increased to extend it) becomes large.[5] But this is essentially an "edge effect." There is a simple way of avoiding it, and thus of dealing only with the dependence of ϵ on "size." Consider, on region C, not arbitrary initial data, but rather only data with support in the interior of C. Then the second sentence of the proof is unnecessary, and one immediately obtains an ϵ depending only on the diameter d of C, not on its shape. Call this function $\epsilon(d)$. Alternatively, instead of restricting the support of the data, one could restrict the region to be a member of a family of regions all dilations of each other (i.e., all of the same shape, but different sizes). Then, again, the required ϵ would depend only on the diameter d. However, one easily convinces oneself that the resulting $\hat{\epsilon}(d)$ is bounded above and below by multiples (which depend on the shape of the regions in the family) of the $\epsilon(d)$ above. Thus, the dependence of ϵ on size reduces to the study of one function of one variable, $\epsilon(d)$. The behavior of $\epsilon(d)$ for small d is of rather less interest, for, since the support of the data must lie in the interior of its region, one may always regard initial data as lying in a larger region, and thus us the norm and ϵ of that larger region. For large d, the behavior of $\epsilon(d)$ is quite simple: It must be bounded below.

Theorem 2: There exist positive numbers d_0 and ϵ_0 such that $\epsilon(d) \geqslant \epsilon_0$ for $d \geqslant d_0$,

The proof follows closely that of Theorem 1. Fix any d_0. Suppose, for contradiction, that we have data $(u_i, \dot{u}_i)$ $(i = 1, 2, \ldots)$ with support in the interior of region C_i of diameter $d_i \geqslant d_0$, all with singular evolution and with $\{(u_i, \dot{u}_i)\}^2 \to 0$. Again locate these regions on S, but choose for the locations y_i, instead of (5), $|y_i| = id_i$. (It may additionally be necessary to choose a subsequence to avoid overlapping of the regions.) There again results smooth data $(u_0, \dot{u}_0)$ on S. With (6) replaced by

$$d_i^2[(i-i)^2] \leqslant 1 + r^2 \leqslant d_i^2[d_0^{-2} + (i+1)^2], \tag{8}$$

we again conclude that $\|(u_0, \dot{u}_0)\|^2$ is finite. Apply the boost theorem. Again the η-domain of dependence of all but a finite number of the C_i is in the region (3), and so again we obtain a contradiction.

Theorems 1 and 2 taken together constitute what we wish to regard as the local contents of the boost theorem. Indeed, this view is supported by the observation that one can, by a simple geometrical argument from these two local theorems, essentially recover the original boost theorem. The argument, which basically reverses the proof of Theorem 1, is the following. Fix number $\theta \in (0,1)$. Consider on the hypersurface S data $(u_0, \dot{u}_0)$ with finite norm (2). Choose any point y of S, and denote by C the closed ball in S with center y and diameter d given by

$$d^2 = 4\theta^2(1 + |y|)^2. \tag{9}$$

Then the distance r of any point of C from the origin is between $|y| - d/2$ and $|y| + d/2$. So, we have

$$[(1-\theta)^2/4\theta^2]d^2 \leqslant 1 + r^2 \leqslant [(1+\theta^2)/\theta^2]d^2. \tag{10}$$

Again using (10) to relate the norms (2) and (4), we conclude that

$$\{(u_0, \dot{u}_0)\}^2 \leqslant c\|(u_0, \dot{u}_0)\|^2. \tag{11}$$

Here, the norm on the right is that of (2), the norm on the left is that of (4) using the data restricted to C, and c is a constant (depending on θ). Now invoke Theorems 1 and 2. It follows that, for $\|(u_0, \dot{u}_0)\|^2$ sufficiently small, the data restricted to C, for every C constructed as above, has nonsingular evolution. In Minkowski space–time, the union of the domains of dependence of the C's is, by (9), precisely the region (3). But small initial data on C evolves to a solution u in the domain of dependence of C with u and ∇u small, and hence with $\gamma^{ab}(u, \nabla u)$ near η^{ab}. Hence, for $\|(u_0, \dot{u}_0)\|^2$ sufficiently small, the corresponding solution admits boost θ.

Note that the argument above recovers only the "boost part" of the boost theorem. The fact that small data on C produces a solution in the domain of dependence of C with $\gamma^{ab}(u, \nabla u)$ near the η^{ab} is actually used in the argument. This feature is in a sense an unavoidabe consequence of the fact that the boost theorem refers to boosts with respect to the flat background metric η^{ab} rather than the physical metric $\gamma^{ab}(u, \nabla u)$, for which "boost" would anyway be difficult to define.

An interesting feature of the present local version is that there are numerous ways to "globalize" Theorems 1 and 2, yielding numerous variants of the boost theorem. By strengthening the norm (2) [increasing the functions $(1 + r^2)^{s+\delta}$ which appear], one can increase the size of the region (3) over which smooth evolution is guaranteed. In the other direction, weaker norms suffice for smaller regions. The simplest such variant is the following:

Theorem 3: Let $(u_0, \dot{u}_0)$ be initial data on S with norm

$$\sum_{s=0}^{5}\int_S (D_{a_1}\cdots D_{a_s}u_0)(D^{a_1}\cdots D^{a_s}u_0) + \sum_{s=0}^{4}\int_S (D_{a_1}\cdots D_{a_s}\dot{u}_0)(D^{a_1}\cdots D^{a_s}\dot{u}_0) \tag{12}$$

finite. Then the corresponding solution of (1) admits some time translation. To prove this, repeat the argument above, replacing (9) by $d = 2t_0$ and (10) by $d/2t_0 \leqslant 1 \leqslant d/2t_0$.

In fact, the class of theorems one obtains in this way is very large indeed. Let Σ be any open region in Minkowski space–time in which the initial hypersurface S is a Cauchy surface,[4] and whose closure, $\bar{\Sigma}$, in the conformal completion[6] of Minkowski space–time does not intersect null infinity $\mathscr{I}^{\pm}$. Then, we claim, there exists a norm of the form (2) such that any initial data $(u_0, \dot{u}_0)$, if sufficiently small in this norm, must evolve to a solution which is smooth and nonsingular everywhere in Σ. (The boost theorem and Theorem 3 are special cases for particlar choices of Σ). To see this, first replace (9), the expression for the diameter d of a ball in terms of its location y, by a new expression such that (i) the union of the domains of dependence of the balls includes $\bar{\Sigma}$ and (ii) $|y| - d/2 \to \infty$ as $|y| \to \infty$. (It is here that we need that $\bar{\Sigma}$ not intersect $\mathscr{I}^{\pm}$.) Let f be a function such that $f(r)$ exceeds $\max(d^{\delta}, d^{1+\delta}, \ldots, d^{5+\delta})$, where the maximum is over all the d's which occur for all those balls which include any point a distance r from the origin in S. This maximum exists, since, by (ii), it is over a collection of balls having their centers in a compact region of S. Finally, the appropriate norm is (2), but with the powers of $(1 + r^2)$ everywhere replaced by $f(r)$. For many Σ's, more delicate arguments can yield more delicate norms.

Of particular interest would be some result to the effect that, for sufficiently small initial data on S, the region of smooth evolution reaches null infinity, for such a result would allow us to study radiation fields. This case was explicitly excluded in the argument above, by the condition that Σ not meet $\mathscr{I}^{\pm}$. [Note, similarly, that the left side of (10) vanishes when $\theta = 1$.] Might there, however, be a more subtle argument involving Theorems 1 and 2, using a norm possibly radically different from that of (2), which does work in this case? It is perhaps instructive to see why this is unlikely. Fix initial data $(u_0, \dot{u}_0)$, small in some norm, on S, and let the region Σ above be such that there is some point p in $\overline{\Sigma}$ and $\mathscr{I}^{\pm}$. Let $\widehat{C}$ be any closed ball in S in the interior of $I^{-}(p) \cap S$. Now consider any collection of closed balls in S, the union of whose domains of dependence includes $\widehat{\Sigma}$. Then there must be within this collection balls which include $\widehat{C}$ and have arbitrarily large diameter d. Consider the norm (4) for the initial data restricted to these balls. The integrals over C will in particular include integrals over the fixed $\widehat{C}$, and so, as d becomes large, the norm will in general become infinite. Thus, we will in general be unable to conclude from Theorems 1 and 2 that the data, restricted to the balls in this collection, have nonsingular evolution. One checks that this argument will fail only if the data $(u_0, \dot{u}_0)$ vanish everywhere in $I^{-}(p) \cap S$. In short, one can show nonsingular evolution to null infinity only in the trivial case in which the zero data is responsible for that evolution.

Finally, we remark that this section is easily generalized to include systems of hyperbolic equations [rather than the single equation (1)] and to dimensions other than four.

4. EINSTEIN'S EQUATION

The previous section applies equally well, of course, when the hyperbolic system is Einstein's equation. However, as we shall see shortly, further simplifications are possible in this case. To make the discussion concrete, we deal only with the source-zero equation. The results may be generalized to include sources, themselves subject to suitable hyperbolic equations.

Let C be a compact 3-submanifold with boundary of the 3-manifold $\mathbb{R}^3$. An initial-data set on C consists of a pair $(q_{ab}, p_{ab}$ of symmetric tensor fields on C, the former positive-definite, satisfying the usual constraint equations[7] of general relativity. The key feature which simplifies the treatment of Einstein's equation is its scaling symmetry. Let Ω be any positive number. Then the initial-data set $(q'_{ab}, p'_{ab}) = (\Omega^2 q_{ab}, \Omega p_{ab})$ also satisfies the constraint equations. Further, the two sets of initial data evolve to solutions of Einstein's equation differing only by the constant factor Ω^2. In particular, either both have nonsingular evolution, or neither do. This scaling freedom is, of course, not exclusive to Einstein's equation. Indeed, it merely reflects the fact that q_{ab} and p_{ab}, as physical fields, have physical dimensions.

Now fix some flat, positive-definite metric h_{ab} on C. We define the norm of initial-data set (q_{ab}, p_{ab}) by

$$\begin{aligned}\{(q_{ab}, p_{ab})\}^2 &= \sum_{s=0}^{4} d^{2s-3} \int_C [D_{a_1} \cdots D_{a_s}(q_{ab} - h_{ab})] \\ &\quad \times [D_{a_1} \cdots D_{b_s}(q_{cd} - h_{cd})] h^{a_1 b_1} \cdots h^{a_s b_s} h^{ac} h^{bd} \qquad (13) \\ &\quad + \sum_{s=0}^{3} d^{2s-2} \int_C [D_{a_1} \cdots D_{a_s} p_{ab}] [D_{b_1} \cdots D_{b_s} p_{cd}] \\ &\quad \times h^{a_1 b_1} \cdots h^{a_s b_s} h^{ac} h^{bd}.\end{aligned}$$

Here, D_a is the h-derivative operator, and d the h-diameter of C. Comparison of (13) with (4) reveals a few minor changes. First, (13) has been expressed in more geometrical form. Second, the upper limits of the sums have been reduced from 5 and 4 to 4 and 3, respectively. As remarked in Sec. 2, this reduction is available because of the particular form taken by Einstein's evolution equations. Third, the number δ in (4) has now been set equal to $-\frac{3}{2}$. Application of the general considerations of Sec. 2 would yield $\delta > -\frac{3}{2}$, but as close as one wishes to this lower limit. The present choice, as we shall see, is made possible by the scaling symmetry. An important feature of this norm is that it is scale-invariant ("dimensionless"), provided we scale the flat metric by $h'_{ab} = \Omega^2 h_{ab}$. (The "$-3$" in the d exponent in the first sum scales the h-volume element in the integrals.)

We now state the local nonsingularity theorems for Einstein's equation.

Theorem 4: Given region C and flat h_{ab} as above, there exists $\epsilon > 0$ such that any initial data set (q_{ab}, p_{ab}) on C with norm (13) less than ϵ has nonsingular evolution.

Theorem 5: There exists $\epsilon_0 > 0$ such that, whenever C is an h-ball (of any diameter), $\epsilon = \epsilon_0$ suffices in Theorem 4.

Theorem 4 is just Theorem 1 restated in the present context, while Theorem 5 follows immediately from Theorem 4, using the scale symmetry and the fact that the norm (13) is scale-invariant. Of course, a result analogous to Theorem 5 holds for any fixed "shape" of C, varying only the "size." Note that we never used Theorem 2.

In this sense, then, a small initial-data set for Einstein's equation must have nonsingular evolution. It is curious that, in contrast to the general case, the dependence on size of the region is so simple for Einstein's equation. Indeed, Theorem 5 guarantees the existence of a "universal" pure number ϵ_0, the smallest such that any initial data on a ball must have nonsingular evolution provided its norm (13) is less than ϵ_0. What is the value of this number? It would seem to be difficult to obtain a good estimate. A lower limit may result from tracing through the original proof[1] of the boost theorem. Upper limits may be obtained by means of examples. Consider, for instance, the Kasner solutions.[8] The metric, in a suitable coordinate system, is

$$-dt^2 + t^{2p_1}dx^2 + t^{2p_2}dy^2 + t^{2p_3}dz^2, \qquad (14)$$

where p_1, p_2, and p_3 are numbers whose sum, as well as the sum of whose squares, is 1. Denote by C the spacelike 3-submanifold with boundary given by

$$\begin{aligned}&t = t_0, \\ &t_0^{2p_1}(1-p_1)^2 x^2 + t_0^{2p_2}(1-p_2)^2 y^2 + t_0^{2p_3}(1-p_3)^2 z^2 \leqslant 3t_0^2.\end{aligned} \qquad (15)$$

It is not difficult to check that the domain of dependence of C reaches "the singularity at $t = 0$." So, the induced initial data on C must have singular evolution. The data is given, in this coordinate system, as follows: The components of q_{ab} are $\mathrm{diag}(t_0^{2p_1}, t_0^{2p_2}, t_0^{2p_3})$, and those of p_{ab}, $2t_0^{-1}\,\mathrm{diag}(p_1 t_0^{2p_1}, p_2 t_0^{2p_2}, p_3 t_0^{2p_3})$. Let h_{ab} be the flat metric on C with components $\mathrm{diag}(t_0^{2p_1}(1-p_1)^2, t_0^{2p_2}(1-p_2)^2, t_0^{2p_3}(1-p_3)^2)$, so C is an h-ball of diameter $2\sqrt{3}t_0$. It is now easy to compute the norm (13), for only the first term in each sum contributes:

$$\{(q_{ab}, p_{ab})\}^2 = \frac{\pi}{6}\sum_{i=1}^{3}\frac{p_i^2(2-p_i)^2}{(1-p_i)^4} + 8\pi\sum_{i=1}^{3}\frac{p_i^2}{(1-p_i)^4}. \tag{16}$$

For the case $p_1 = p_2 = \frac{2}{3}$, $p_3 = -\frac{1}{3}$, for example, this norm has value $305\,995\pi/512 \approx 1878$. We conclude, therefore, that the number ϵ_0 cannot exceed this value. Clearly, one could substantially reduce this upper limit by more careful analyses of this and other examples.

There are at least three unnatural features of the local nonsingularity theorem for general relativity. We briefly discuss these features.

The norm (13) involves, in addition to the initial data, a fixed flat background metric h_{ab} on C—a metric that, of course, has no physical significance. It would be more natural to use a norm involving only physical fields. We indicate a possible line of attack on this issue. Fix C, a 3-manifold with boundary, topologically a 3-ball. On initial data (q_{ab}, p_{ab}) on C, introduce the following norm:

$$\begin{aligned}\{(q_{ab}, p_{ab})\}'^2 &= \max_{\partial C}(d^2\mathbb{R} - 8)^2 + \max_{\partial C}(d\pi_{ab} - 2n_{ab})(d\pi^{ab} - 2n^{ab}) \\ &+ \sum_{s=0}^{2} d^{2s+1}\int_C [D_{a_1}\cdots D_{a_s}R_{ab}]\,[D^{a_1}\cdots D^{a_s}R^{ab}] \\ &+ \sum_{s=0}^{3} d^{2s-2}\int_C [D_{a_1}\cdots D_{a_s}p_{ab}]\,[D^{a_1}\cdots D^{a_s}p^{ab}].\end{aligned} \tag{17}$$

Here, ∂C is the boundary of C, d is such that πd^2 is the q-area of ∂C, n_{ab} is the induced metric and π_{ab} the extrinsic curvature of ∂C, $\mathbb{R}$ is the scalar curvature of n_{ab}, D_a is the derivative operator defined by q_{ab}, R_{ab} is the Ricci tensor of q_{ab}, and indices are raised and lowered with q_{ab}. The key features of this norm are that it involves no flat background metric and that is is invariant under scaling.

The norm (17) vanishes if an only if $\mathbb{R} = 8/d^2$, $\pi_{ab} = 2n_{ab}/d$, $R_{ab} = 0$, and $p_{ab} = 0$. That is, the norm vanishes if and only if p_{ab} vanishes and q_{ab} is flat and such that C is a q-ball. We claim further that the norm (13) bounds (17): Given $\epsilon' > 0$ there exists $\epsilon > 0$ such that $\{(q_{ab}, p_{ab})\}^2 \leqslant \epsilon$ implies $\{(q_{ab}, p_{ab})\}'^2 \leqslant \epsilon'$. To see this, note[3] that, for the first sum in (13) small, $q_{ab} - h_{ab}$, as well as its first two derivatives, must be small at points of C, whence the first two terms in (17) are small. Then the first sum in (13) bounds the first in (17), and the second the second. But the more interesting question is the converse. It is true that, given by $\epsilon > 0$, there exists $\epsilon' > 0$ such that, whenever $\{(q_{ab}, p_{ab})\}'^2 \leqslant \epsilon_1$, then there exists flat metric h_{ab} on C such that C is an h-ball and $\{(q_{ab}, p_{ab})\}^2 \leqslant \epsilon$? A positive answer would mean that one may replace norm (13) by the "intrinsic" norm (17). It seems a reasonable conjecture that the answer is yes. Roughly speaking, the first term in (17) small implies that the q-induced metric on ∂C is close to that of a sphere, while the second term small then implies the existence of a flat h_{ab} near ∂C whose value and derivative on ∂C are close to those of q_{ab}. We thus achieve the required boundary conditions. Finally, the first sum in (17) small implies that the Ricci tensor of q_{ab} is small at points of C, i.e., that q_{ab} is "nearly flat." So, one might expect a flat h_{ab} on C close to q_{ab} in the sense of (13). It would be of interest to prove this conjecture, preferably for some less awkward replacement for (17).

The second feature, also involving the character of the norm (13), is closely related to the first. There exists of course initial data (q_{ab}, p_{ab}) on C which is for flat space–time, but is such that q_{ab} is far from flat and p_{ab} is far from zero. Such initial data may be obtained by taking a spacelike embedding—not in a hyperplane—of C in Minkowski space–time, and then evaluating the induced data. Now, for such data the norm (13) can be large, for it vanishes if and only if $q_{ab} = h_{ab}$ and $p_{ab} = 0$. Thus, we would be unable to conclude from Theorem 4 that such data has nonsingular evolution, despite the fact that its evolution (a flat space–time) is obviously nonsingular. In short, it would perhaps be more natural for the norm (13) to measure distance, not from the trivial data $(h_{ab}, 0)$, but rather from "the nearest data with flat evolution." We suggest, as one possible line for constructing such a norm, use of Witten's energy integral.[9] This integral is suggested by the fact that it vanishes for any data with flat evolution. But one would first have to modify the boundary conditions on the spinor potential, to accommodate the fact that our data, confined to a compact region C, are not asymptotically flat. Further, the integral itself would have to be modified, for as it stands it does not control the higher derivatives of the initial data which appear in (13). A norm along these lines, could it be found, might also serve as the "less awkward replacement" for (17).

The third unnatural feature involves the derivation from Theorem 4 of the original boost theorem and its generalizations. This derivation was carried out in Sec. 3 for the case of the general hyperbolic system. That argument made use of the fact that small initial data evolves to a small solution. Now this general framework is applied to Einstein's equation by working in a harmonic coordinate system. Thus, the corresponding fact in the Einstein case would read: By requiring that the norm (13) on initial data $(q_{ab}, p_{ab}$ on C be small, one can guarantee that, at every point, the components of the space-time metric g_{ab}, in a harmonic coordinate system, will be near diag$(-1, +1, +1, +1)$, and that the coordinate derivatives of these compounds will be small. This formulation is awkward, for it involves a coordinate system without physical significance. Of course, the final result—the generalized boost theorem—in the Einstein case manifests a similar awkwardness, for it makes reference to an unphysical flat background metric. But this seems to be unavoidable, for the region in which nonsingular evolution is to be guaranteed must be specified prior to the space–time metric. In short, while it seems difficult to improve the statement of the final result, one might hope to improve what goes into that result, i.e., to find a more physical version of "small initial data in C for Einstein's equation yields a nearly flat space–time metric."

Fix compact 3-manifold with boundary, C, and flat metric h_{ab} on C. Denote by $\mathscr{I}$ the set of all smooth initial data (q_{ab}, p_{ab}), satisfying the constraint equations, on C. The norm (13) defines neighborhoods of the point $(h_{ab}, 0)$ of $\mathscr{I}$. Consider the 4-manifold $M = C \times \mathbb{R}$, and fix once and for all positive-definite metric $\overset{+}{g}_{ab}$ on M. Denote by $\mathscr{S}$ the collection of all smooth source-zero solutions g_{ab} of Einstein's equation on M, with respect to which the domain of dependence of C is compact. Let the topology on $\mathscr{S}$ be that induced by the distance function

$$\begin{aligned} d(g_{ab}, g'_{ab}) = {} & \operatorname*{lub}_{M}(g_{ab} - g'_{ab}) \\ & \times (g_{cd} - g'_{cd})\overset{+}{g}{}^{ac}\overset{+}{g}{}^{bd} \\ & + \operatorname*{lub}_{M}(\overset{+}{\nabla}_a(g_{bc} - g'_{bc}))(\overset{+}{\nabla}_d(g_{ef} - g'_{ef})) \\ & \times \overset{+}{g}{}^{ad}\overset{+}{g}{}^{bc}\overset{+}{g}{}^{cf}, \end{aligned} \tag{18}$$

where $\overset{+}{\nabla}_a$ is the derivative defined by $\overset{+}{g}_{ab}$. Now, one might think of representing "the evolution of data" as a mapping from $\mathscr{I}$ to $\mathscr{S}$, whence "small data evolves to a nearly flat metric" would be expressed as continuity of this map near $(h_{ab}, 0)$ in $\mathscr{I}$. But this does not work. In the absence of gauge conditions, a given set of initial data does not even produce a unique space–time metric g_{ab} in M. Thus, we obtain no map, and so in particular cannot require its continuity.

An elegant way to say what one wants to say was suggested to me by Rafael Sorkin. Consider the mapping ζ from $\mathscr{S}$ to $\mathscr{I}$ obtained by merely evaluating the data on C. This mapping *is* well defined. We are interested, not in continuity of the map ζ, but rather in "continuity of its (nonexistent) inverse." We may express this by requiring that the mapping ζ be open in an appropriate region.

Theorem 6: Let $g_{ab} \in \mathscr{S}$ with $\zeta(g_{ab}) = (h_{ab}, 0) \in \mathscr{I}$, and let U be a neighborhood of g_{ab}. Then $\zeta[U]$ is a neighborhood of $(h_{ab}, 0)$.

This statement merely restates the second part of the boost theorem in the Einstein case. Thus, by choosing the mapping to be from the solutions to the data, rather than the other way, one avoids the problem of gauge.

5. CONCLUSION

To summarize, we have shown that, for certain hyperbolic systems, initial data given on a compact spacelike patch must evolve to a solution nonsingular insofar as determined by that data, provided the initial data is sufficiently small as measured by a certain norm. The class of hyperbolic systems to which this applies includes Einstein's equation.

One might seek to generalize these results in any of a number of directions. We suggest a few possibilities below.

The norm (4) on initial data $(u_0, \dot{u}_0)$ in the general case involves u_0 and its first five spatial derivatives, as well as $\dot{u}_0$ and its first four. Is control of such high derivatives really necessary to ensure nonsingular evolution? In the original boost theorem,[1] the data was not assumed to be smooth, and so a strong norm was needed in part merely to impose on the data sufficient smoothness that *some* evolution exist. But suppose one assumes, as here, that the data is already smooth (C^∞). Might it then be possible to use a weaker version of the norm (4), in which the last few terms in the sums are omitted?

An example will illustrate this remark. Consider in Minkowski space–time the equation

$$\eta^{ab}\nabla_a\nabla_b u = 2(1 + u)^{-1}\eta^{ab}\nabla_a u \nabla_b u. \tag{19}$$

Let C be the unit ball centered at the origin in the initial hypersurface S ($t = 0$). Then Theorem 1 guarantees that data on C with norm (4) sufficiently small must have nonsingular evolution. Will a weaker norm do? The answer in this example is yes. In fact, (19) is just the wave equation on the function $v = (1 + u)^{-1}$. Thus, singularities in u from evolution under (19) are just zeros in v from evolution under the wave equation. The solution v is given in terms of its initial data on S by

$$v(t,x,y,z) = \operatorname{avg}[v_0 + tn^a\nabla_a v_0 + t\dot{v}_0], \tag{20}$$

where avg is the average over the sphere given by the intersection of S and the light cone of (t, x, y, z) and n^a is the unit outward normal in S to this sphere. Thus, an acceptable norm on $(u_0, \dot{u}_0)$ must hold v_0 sufficiently close to one and ∇v_0 and $\dot{v}_0$ sufficiently close to zero that the v given by (20) is bounded away from zero. This is accomplished[10] by the norm (4) with the last three terms in each sum omitted, i.e., by the norm which involves only u_0 and its first two spatial derivatives, and $\dot{u}_0$ and its first. No further terms may be omitted.[11]

Thus, in at least one example, some higher-derivative terms in the norm (4) may be omitted. What would be of interest is, not weaker norms for special equations—and particularly not for such artificial equations—but rather a general result to the effect that a weaker norm suffices for all equations of the form (1), or else an example to show that no weaker norm works in general.

A similar question arises, of course, for the norm (13) for Einstein's equation. Now one has the added feature—perhaps making proofs easier, and certainly making examples more difficult—that the initial data must also satisfy the constraint equations.

The local nonsingularity theorem states that initial data must have nonsingular evolution if it is near the trivial data. A natural generalization is that initial data must have nonsingular evolution if it is near any initial data with nonsingular evolution. Such an assertion may be formulated as follows. Fix C, a compact 3-submanifold with boundary of R^3, with its flat metric h_{ab}. Denote by $\mathscr{I}$ the collection of all smooth initial data, $(u_0, \dot{u}_0)$, for (1) on C. Impose on $\mathscr{I}$ the norm topology from (4), i.e., that which arises from $\{(u_0 - \dot{u}'_0, u_0 - \dot{u}'_0)\}^2$ as the distance between points $(u_0, \dot{u}_0)$ and $(u'_0, \dot{u}'_0)$. In general some points of $\mathscr{I}$, i.e., some sets of initial data, will have nonsingular evolution; others singular. Theorem 1 in this language asserts that those with nonsingular evolution include a neighborhood of the point (0, 0). The natural generalization is:

Conjecture 7: The set of points of $\mathscr{I}$ with nonsingular evolution is open.

Note, for instance, that Conjecture 7 is true for the case of the hyperbolic equation (19).

One could in particular formulate this question for Einstein's equation. Then $\mathscr{I}$ is the collection of smooth initial data (q_{ab}, p_{ab}) on C satisfying the constraint equations, the topology on $\mathscr{I}$ is the norm topology from (13), and the assertion is again Conjecture 7. In the general case, Conjecture 7 replaces Theorem 1, but does not include the additional information contained in Theorem 2, i.e., that the degree of smallness required on initial data for nonsingular evolution is "uniform" in the size of the region C. For Einstein's equation, Conjecture 7 replaces Theorem 4. But in this case the corresponding "uniformity," given by Theorem 5, is an immediate consequence, by virtue of the scaling symmetry. Thus, in contrast to the general case, Conjecture 7 captures essentially the entire "local content" of the boost theorem for Einstein's equation.

There are several other interesting features of Conjecture 7 for Einstein's equation. First note that, whereas the actual distances, defined by the norm (13), between points of $\mathscr{I}$ depend on the flat background metric h_{ab}, the resulting topology on $\mathscr{I}$ does not. In fact, this topology is unchanged for h_{ab} any smooth positive–definite metric, flat or not. Thus, this topological version avoids entirely the artificial dependence on a flat background metric noted in Sec. 4. In addition, there now arises the possiblity of further generalizing Conjecture 7 by letting C be *any* compact 3-manifold possibly with boundary—not necessarily one embedded in $\mathbb{R}^3$. So, for example, C could be a 3-sphere. Conjecture 7 implies that initial dta (q_{ab}, p_{ab}) has nonsingular evolution if it is sufficiently close to *any* initial data having flat evolution. Thus, this topological version would circumvent the second unnatural feature, noted in Sec. 4, of the local nonsingularity theorem for general relativity.

Finally, we remark that one may generalize Theorem 6 and combine it with Conjecture 7.

Conjecture 8: The mapping $\zeta : \mathscr{S} \to \mathscr{I}$ is open.

(A mapping is open if the image under it of any open set is open.)

This generalizes Theorem 6 because it requires that the mapping ζ be open everywhere, not just near certain flat solutions. It also includes Conjecture 7 as a special case, for it requires in particular that $\zeta[\mathscr{S}]$ be open in $\mathscr{I}$, while $\zeta[\mathscr{S}]$ be precisely the points of $\mathscr{I}$ with nonsingular evolution. In short, Conjecture 8—an extremely simple statement—seems to carry out all the information about the local nonsingularity behavior of Einstein's equation. A proof would be of interest.

ACKNOWLEDGMENTS

I wish to thank Demetrios Christodoulou and Rafael Sorkin for helpful discussions.

[1]D. Christodoulou and N. O'Murchadha, Comm. Math. Phys. **80**, 271 (1981); D. Christodoulou, J. Math. Pures Appl. **60**, 99 (1981).

[2]See, for example, S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U.P., Cambridge 1973), Chap. 8.

[3]S. Sobelev, *Applications of Functional Analysis to Physics*, Translations of Mathematical Monographs (Am. Math. Soc., Providence, R.I., 1963).

[4]See, for example, R. Geroch, J. Math. Phys. **11**, 437 (1970).

[5]See Ref. 3, pp. 137.

[6]See, for example, Ref. 2, pp. 221; R. Geroch, in *Asymptotic Structure of Space-Time*, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977).

[7]See, for example, Ref. 2, Chap. 7.

[8]E. Kasner, Am. J. Math. **43**, 217 (1921).

[9]E. Witten, Comm. Math. Phys. **80**, 391 (1981).

[10]That is, this norm bounds the surface integrals in (20). See Ref. 3.

[11]This is seen by means of examples. Let u_0 vanish except near one sphere over which the average is taken in (20), there having a "bump" of amplitude κ and width λ. Then smallness of just the first two terms in the norm (4) bounds κ^2/λ; while the contribution of the second term in the average in (20) is of the order κ/λ. But, for $\lambda \to 0$ as $\kappa \to 0$, one can make κ/λ as large as one wishes, while holding κ^2/λ as small as one wishes.

The occupation statistics for indistinguishable dumbbells on a $2\times 2\times N$ lattice space[a)]

J. L. Hock
Department of Electrical Engineering and Computer Science, Marquette University, Milwaukee, Wisconsin 53233

R. B. McQuistan
Department of Physics and Laboratory for Surface Studies, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201

(Received 26 July 1982; accepted for publication 21 January 1983)

A 20-term recursion relation is derived that describes exactly the occupation statistics for indistinguishable dumbbells (which occupy two nearest neighbor sites) distributed on a $2\times 2\times N$ lattice space. On the basis of this recursion, the normalization, expectation, dispersion, and continuous representation of the statistics are also developed. When the lattice space is completely filled, the recursion relationship reduces to four terms, permitting a calculation of the orientational degeneracy of a completely filled $2\times 2\times N$ lattice space.

PACS numbers: 05.50. + q, 02.50. + s

I. INTRODUCTION

The problem of determining the occupational degeneracy for indistinguishable dumbbells distributed on a lattice space has its genesis in the statistical treatment of phenomena such as the adsorption and crystallization of diatomic molecules, the properties of binary alloys and the role of hydrogen bonding in elastomers and the action of muscle tissue.

The underlying difficulty in treating such systems is that there is statistical correlation in the sense that if a site is occupied then at least one of its nearest neighbor sites must be occupied. Thus, there is not a distribution of occupied sites but rather a distribution of pairs of occupied sites. Here, a dumbbell particle is one that occupies two linearly contiguous sites. As is generally true for statistical problems of this nature, exact solutions have been found for the one-dimensional case[1] only. Exact solutions for lattice spaces of higher dimensionality have been obtained for very special cases only, i.e., a completely filled two-dimensional rectangular lattice space using Pfaffians[2,3] and the transfer matrix method.[4] Consequently, either the spaces have been restricted[5,6] or approximation methods have been utilized[7,8] to solve this problem.

In the present paper we are concerned with the occupation statistics for indistinguishable, noninteracting dumbbells distributed on a nonsaturated, quasi-three-dimensional rectangular lattice space: a $2\times 2\times N$ array, in which most of the lattice sites have four nearest neighbors (see Fig. 1). The method presented here yields an exact recursion relation for the unsaturated, quasi-three-dimensional rectangular lattice space. To the best of our knowledge neither the Pfaffian nor the transfer matrix methods can be utilized to yields an exact solution in this case.[9]

II. RECURSION RELATIONS FOR $A[q,N]$

In the present section we derive a recursion relation for $A[q,N]$, the number of unique arrangements of q indistinguishable dumbbells on a $2\times 2\times N$ lattice space. We partition the $2\times 2\times N$ lattice into N blocks in which each block consists of four sites. Five arrays, $\alpha(N)$, $\beta(N)$, $\gamma(N)$, $\delta(N)$, and $\epsilon(N)$, are defined by the removal of zero through three sites from the Nth partition (see Fig. 2).

Thus, we call the quantities $A[q,N]$, $B[q,N]$, $C[q,N]$, $D[q,N]$, and $E[q,N]$, the numbers of all unique arrangements of q indistinguishable dumbbels on the lattices, $\alpha(N)$, $\beta(N)$, $\gamma(N)$, $\delta(N)$, $\epsilon(N)$, respectively.

We now determine certain subsets of which $A[q,N]$, $B[q,N]$, $C[q,N]$, $D[q,N]$, and $E[q,N]$ are composed, by examining the states of occupation of the compartments in the Nth partition of the arrays, $\alpha(N)$, ..., $\epsilon(N)$, respectively. There are a total of 34 configurations involving the occupation (or vacancy) of the Nth partition of the $\alpha(N)$ array. However, due to reflective and rotational symmetries there are only ten independent arrays (see Fig. 3). The numerical coefficient beneath each drawing indicates the degeneracy of the depicted arrangement. For example, the third drawing from the top on the left-hand side of Fig. 3 shows the situation when three of the compartments are vacant and one occupied by a dumbbell that extends into the adjacent compartment in the $(N-1)$th partition. The remaining $(q-1)$ dumbbells can then be distributed on a $\beta(N-1)$ array in $B[q-1,N-1]$ independent ways. Because the dumbbell that occupies compartments in the Nth and $(N-1)$th partitions can be placed in four equivalent positions, there are $4B[q-1,N-1]$ arrangements possible for this configuration.

By considering the ten arrangements shown in Fig. 3,

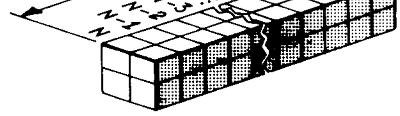


FIG. 1. A $2\times 2\times N$ rectangular lattice space.

[a)]The research reported here has been supported, in part, by the Air Force Office of Scientific Research under Grant AFOSR 81-0192.

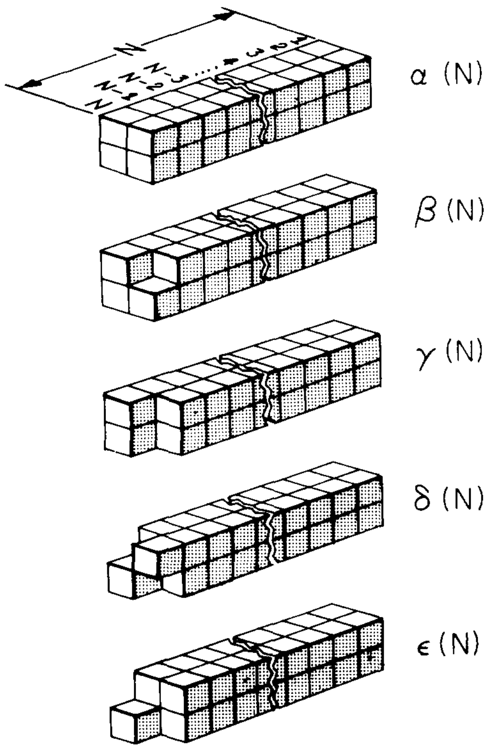


FIG. 2. The five lattice arrays required for the determination of the recursion relation satisfied by $A[q,N]$.

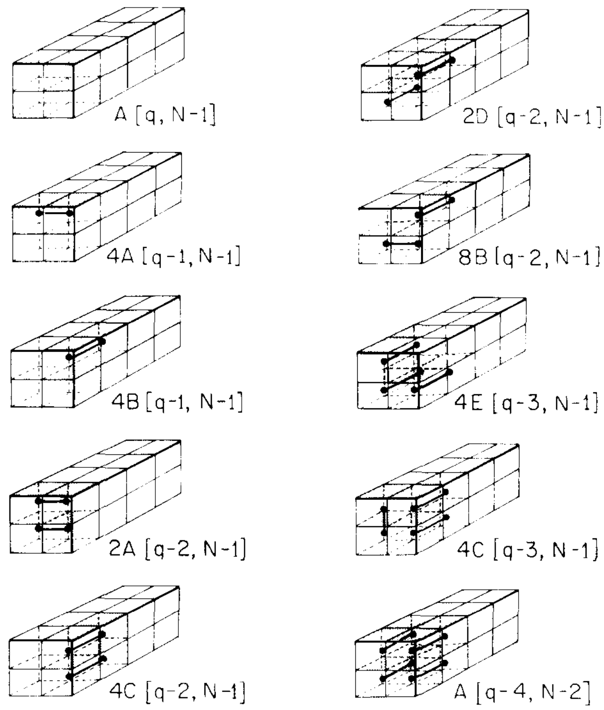


FIG. 3. This figure shows the ten independent arrays into which the set $A[q,N]$ can be decomposed.

we may write for the decomposition of $A[q,N]$,

$$\begin{aligned}A[q,N] = {} & A[q,N-1] + 4A[q-1,N-1] \\ & + 4B[q-1,N-1] + 2A[q-2,N-1] \\ & + 8B[q-2,N-1] + 4C[q-2,N-1] \\ & + 2D[q-2,N-1] + 4C[q-3,N-1] \\ & + 4E[q-3,N-1] + A[q-4,N-2]. \quad (1a)\end{aligned}$$

From the similar treatment of the arrays in Fig. 2, we write

$$\begin{aligned}B[q,N] = {} & A[q,N-1] + 2A[q-1,N-1] \\ & + 3B[q-1,N-1] + 2B[q-2,N-1] \\ & + 2C[q-2,N-1] + D[q-2,N-1] \\ & + E[q-3,N-1], \quad (1b)\end{aligned}$$

$$\begin{aligned}C[q,N] = {} & A[q,N-1] + A[q-1,N-1] \\ & + 2B[q-1,N-1] + C[q-2,N-1], \quad (1c)\end{aligned}$$

$$\begin{aligned}D[q,N] = {} & A[q,N-1] + 2B[q-1,N-1] \\ & + D[q-2,N-1], \quad (1d)\end{aligned}$$

$$E[q,N] = A[q,N-1] + B[q-1,N-1]. \quad (1e)$$

Equations (1a)–(1e) can be solved for $A[q,N]$ in terms of other A's by first eliminating E using Eq. (1e); next we use Eq. (1d) to eliminate B; C is then eliminated utilizing the equation that results from the foregoing operations on Eq. (1d). The two remaining equations contain A's and D's only and can be solved for the A's by substituting the reindexed D's into one of the equations from the other. These procedures yield

$$\begin{aligned}A[q,N] = {} & A[q,N-1] + 7A[q-1,N-1] \\ & + 6A[q-2,N-1] + A[q-1,N-2] \\ & + 6A[q-2,N-2] + 6A[q-3,N-2] \\ & - 7A[q-4,N-2] - 2A[q-3,N-3] \\ & - 10A[q-4,N-3] - 26A[q-5,N-3] \\ & - 8A[q-6,N-3] + A[q-5,N-4] \\ & + 2A[q-6,N-4] + 6A[q-7,N-4] \\ & + 9A[q-8,N-4] + A[q-8,N-5] \\ & - A[q-9,N-5] + 2A[q-10,N-5] \\ & - A[q-12,N-6]. \quad (2)\end{aligned}$$

Equation (2) is thus the recursion we seek for $A[q,N]$. We will use it to develop the occupation statistics for dumbbells on a $2\times2\times N$ lattice space.

Phares *et al.*[10–13] have solved recursions of the type given in Eq. (2) using a combinatorics function technique that yields solutions of multidimensional, multiterm difference equations.

III. GENERATING FUNCTIONS

We first form the polynominals

$$f_N(x) = \sum_{q=0}^{2N} A[q,N]x^q. \quad (3)$$

Utilizing the recursion derived in the previous section [Eq.

TABLE I. The computer-generated values of $A\,[q,N]$ for various values of q and N. The diagonal for $q = 2N$ is the saturation diagonal.

N \ q	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Sum over rows
0	1															1
1	1	4	2													7
2	1	12	42	44	9											108
3	1	20	142	440	588	288	32									1 511
4	1	28	306	1 672	4 863	7 416	5 470	1 620	121							21 497
5	1	36	534	4 248	19 774	55 200	91 200	84 984	40 553	8 204	450					305 184
6	1	44	826	8 680	56 333	235 132	637 914	1 112 668	1 208 714	771 436	261 500	39 080	1 681			4 334 009
7	1	52	1182	15 480	129 644	728 840	2 810 312	7 465 728	13 541 312	16 397 296	12 752 616	5 986 432	1 532 336	178 272	6272	61 545 775

(2)], Eq. (3) yields

$$\begin{aligned} f_{N+6}(x) = {} & [1 + 7x + 6x^2]\, f_{N+5}(x) \\ & + x[1 + 6x + 6x^2 - 7x^3]\, f_{N+4}(x) \\ & - 2x^3[1 + 5x + 13x^2 + 4x^3]\, f_{N+3}(x) \\ & + x^5[1 + 2x + 6x^2 + 9x^3]\, f_{N+2}(x) \\ & + x^8[1 - x + 2x^2]\, f_{N+1}(x) - x^{12} f_N(x). \end{aligned} \tag{4}$$

Equation (4), together with initial conditions for $f_0(x)$ through $f_5(x)$ gleaned from Table I will yield the numerical values for $A\,[q,N]$, $N > 5$ displayed in Table I.

To obtain $G(x,y)$, the so-called bivariant generating function, defined by

$$G(x,y) \equiv \sum_{N=0}^{\infty} f_N(x) y^N, \tag{5}$$

we impose the initial conditions on $f_N(x)$ (see Table I) and obtain (see Appendix A)

$$\begin{aligned} G(x,y) = H(x,y)/\{ & 1 - y[1 + 7x + 6x^2] \\ & - y^2 x[1 + 6x + 6x^2 - 7x^3] \\ & + 2y^3x^3[1 + 5x + 13x^2 + 4x^3] \\ & - y^4x^5[1 + 2x + 6x^2 + 9x^3] \\ & - y^5x^8[1 - x + 2x^2] + y^6x^{12}\}, \end{aligned} \tag{6}$$

where $H(x,y)$ is a polynominal of tenth degree in x and fourth degree in y.

IV. NORMALIZATION

In this section we determine for large values of N, the numerical values of

$$\Delta_N \equiv \sum_{q=0}^{2N} A\,[q,N], \tag{7}$$

the normalization of the statistics characterized by the recursion relationship given in Eq. (2). The recursion relationship for Δ_N is obtained from the generating function [Eqs. (3) and (4)with $x \equiv 1$]:

$$\begin{aligned} \Delta_{N+6} = {} & 14\Delta_{N+5} + 6\Delta_{N+4} - 46\Delta_{N+3} \\ & + 18\Delta_{N+2} + 2\Delta_{N+1} - \Delta_N. \end{aligned} \tag{8}$$

With the initial conditions (see Table I)

$$\begin{aligned} & \Delta_0 = 1, \quad \Delta_2 = 108, \quad \Delta_4 = 21\,497, \\ & \Delta_1 = 7, \quad \Delta_3 = 1511, \quad \Delta_5 = 305\,184, \end{aligned} \tag{9}$$

the generating function for Δ_N is given by Eq. (6) with $x \equiv 1$, i.e.,

$$\begin{aligned} \sum_{N=0}^{\infty} \Delta_N y^N & = \frac{H[1,y]}{1 - 14y - 6y^2 + 46y^3 - 18y^4 - 2y^5 + y^6} \\ & = \sum_{j=1}^{6} \frac{k_j}{1 - R_j y}, \end{aligned} \tag{10}$$

where the k_j's are constants and the R_j's are the reciprocals of the roots of the denominator; and where $H[1,y] \equiv 1 - 7y + 4y^2 + 3y^3 - y^4$ (see Appendix A), that is, the R_j's are the roots of $z^6 - 14z^5 - 6z^4 + 46z^3 - 18z^2 - 2z + 1$. From Descartes' rule of signs we see that there are at most four positive real roots and two negative real roots and no imaginary roots. The roots are determined to be

$$\begin{aligned} &R_1 = 14.200\,740\,57, \quad R_4 = 0.270\,655\,757, \\ &R_2 = 1.379\,071\,809, \quad R_5 = -0.225\,576\,045, \\ &R_3 = 0.410\,838\,846, \quad R_6 = -2.035\,730\,940. \end{aligned} \tag{11}$$

Consequently, an equivalent expression for the generating function for the Δ_N's may be written as

$$\sum_{j=1}^{6} \frac{k_j}{1 - R_j y} = \sum_{N=0}^{\infty} \sum_{j=0}^{6} (k_j R_j^N) y^N, \tag{12}$$

or

$$\Delta_N = \sum_{j=1}^{6} k_j R_j^N. \tag{13}$$

Because the values of R_2, ..., R_6 are less than R_1, we may write

$$\lim_{N\to\infty} \Delta_N = k_1 R_1^N. \tag{14}$$

As $y \to R_1^{-1}$ only the first term in Eq. (10) is significant so that

$$\lim_{y\to R_1^{-1}} \left\{ \frac{1 - 7y + 4y^2 + 3y^3 - y^4}{1 - 14y - 6y^2 + 46y^3 - 18y^4 - 2y^5 + y^6} - \frac{k_1}{1 - R_1 y} \right\} = 0. \tag{15}$$

Utilizing L'Hospital's rule we determine K_1,

$$k_1 = \frac{R_1^2(R_1^4 - 7R_1^3 + 4R_1^2 + 3R_1 - 1)}{14R_1^5 + 12R_1^4 - 138R_1^3 + 72R_1^2 + 10R_1 - 6} \cong 0.528\,471\,535. \tag{16}$$

Thus

$$\Delta_N \cong (0.528\,471\,535)(14.200\,740\,57)^N. \tag{17}$$

For $N = 5$, Eq. (17) yields 305 194.4. This is to be compared with a value of 305 184 (see Table I), giving an error of approximately 0.003%.

V. EXPECTATION, DISPERSION, AND CONTINUOUS REPRESENTATION

In this section we begin by calculating $\langle\theta\rangle_N$, the expectation value of the occupation of a $2\times2\times N$ lattice space,

$$\langle\theta\rangle_N \equiv 2\langle q\rangle_N/4N = \langle q\rangle_N/2N, \tag{18}$$

where

$$\langle q\rangle_N \equiv \sum_{q=0}^{2N} qA\,[q,N] \Big/ \sum_{q=0}^{2N} A\,[q,N],$$

or

$$\langle\theta\rangle_N = \frac{1}{2N\Delta_N} \sum_{q=0}^{2N} qA\,[q,N]. \tag{19}$$

We utilize Eq. (2) and assume the law of large numbers (see, e.g., Ref. 14), that for sufficiently large N

$$\langle\theta\rangle_N \cong \langle\theta\rangle_{N-1} \cdots \cong \langle\theta\rangle_\infty; \tag{20}$$

then Eq. (19) yields (see Appendix B)

$$\langle\theta\rangle_N = \frac{1}{2}\left[\frac{19R_1^5 + 3R_1^4 - 224R_1^3 + 131R_1^2 + 19R_1 - 12}{14R_1^5 + 12R_1^4 - 138R_1^3 + 72R_1^2 + 10R_1 - 6} \right] = 0.639\,594\,004. \tag{21}$$

Thus, if we assume the validity of the central limit theorem, the maximum number of arrangements occurs when the lattice space is approximately 64% filled. Figure 4 shows $A\,[q,5]$ as a function of q. In this case, the maximum occurs at $q = 6$ or $\langle\theta\rangle_5 = 0.60$.

$\langle\theta^2\rangle_N$, the dispersion in θ for a $2\times2\times N$ lattice space, defined by

$$\langle\theta^2\rangle_N = \langle q^2\rangle_N/4N^2, \tag{22}$$

where

$$\langle q^2\rangle_N \equiv \frac{1}{\Delta_N} \sum_{q=0}^{2N} q^2 A\,[q,N]$$

yields (see Appendix C)

$$\sigma_N^2 \equiv [\langle\theta^2\rangle_N - \langle\theta\rangle_N^2] = 0.076\,248\,755\,N^{-1}, \tag{23}$$

where again for large values of N we have assumed Eq. (20) to be valid.

We see that for large values of N, $A\,[q,N]$ can be represented as Gaussian distribution

$$A\,[\theta,N] = A_{\max} \exp\{-[\theta - \langle\theta\rangle_N]^2/2\sigma_N^2\}, \tag{24}$$

where

$$A_{\max} = \frac{k_1 R_1^N}{(8\pi N\sigma_N^2)^{1/2}} = 0.381\,755\,49\,\frac{(14.200\,740\,57)^N}{\sqrt{N}}$$

as determined by the normalization.

Figure 5 shows a comparison of $A\,[\theta,5]$, as calculated according to Eq. (2) (with appropriate initial conditions), with $A\,[\theta,5]$ as determined by Eq. (24).

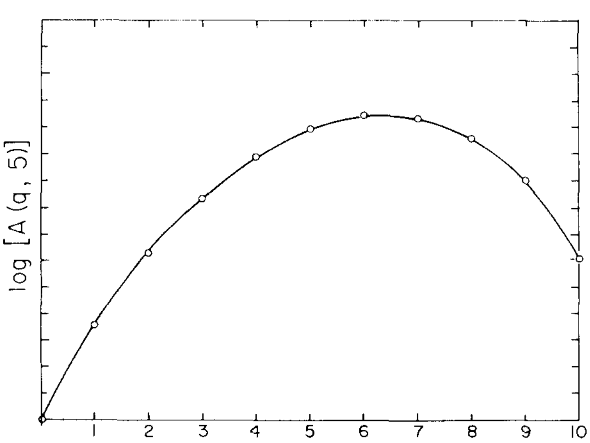


FIG. 4. $A\,[q,5]$ is shown as a function of q. The maximum value of A occurs at $q = 6$ or $\langle\theta\rangle_5 = 0.60$.

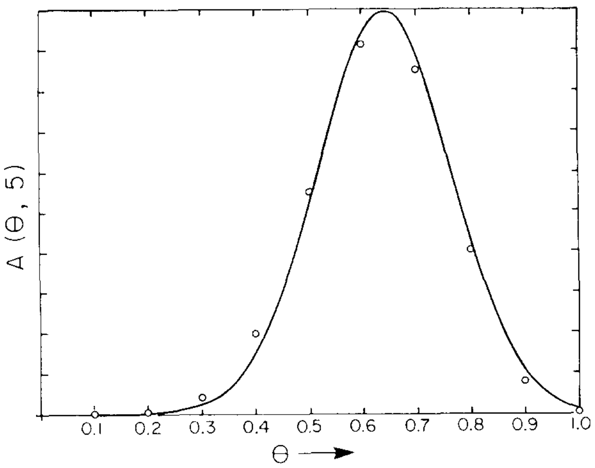


FIG. 5. The dots show $A[\theta,5]$ calculated according to the basic recursion relation, Eq. (2), with appropriate initial conditions. The continuous curve is $A[\theta,5]$, as given by Eq. (24).

VI. ORIENTATIONAL DEGENERACY FOR SATURATED LATTICE SPACES

We note that when $q > 2N$ no arrangements are possible. If, however, $q = 2N$, i.e., if the lattice space is completely filled, Eq. (2) reduces to

$$\begin{aligned} A_N = 6A_{N-1} - 7A_{N-2} - 8A_{N-3} \\ + 9A_{N-4} + 2A_{N-5} - A_{N-6}, \end{aligned} \tag{25}$$

where A_N has been written for $A[2N,N]$.

Equation (25) is not the simplest recursion that can be written, as can be seen from the following argument: If we require the Nth partition of the $\alpha(N)$-array (see Fig. 2) to be completely filled, a condition which is necessary if the entire space is to be occupied, then

$$\begin{aligned} A[q,N] = 2A[q-2,N-1] + 4C[q-3,N-1] \\ + A[q-4,N-2] \end{aligned} \tag{26}$$

and

$$C[q,N] = A[q-1,N-1] + C[q-2,N-1]; \tag{27}$$

that is, there can be no B, D, or E arrangements.

Equation (26) can be solved for $C[q,N]$ which, when reindexed, can be substituted for the C's in Eq. (27) to yield (after setting $q = 2N$, the condition for a saturated lattice) a recursion relationship for A_N, i.e.,

$$A_N = 3A_{N-1} + 3A_{N-2} - A_{N-3}. \tag{28}$$

By reindexing Eq. (28), we can construct Eq. (25): add

$$\begin{array}{lllllll} A_N & = 3A_{N-1} & + 3A_{N-2} & - A_{N-3} & & & \\ 0 & = 3A_{N-1} & - 9A_{N-2} & - 9A_{N-3} & + 3A_{N-4} & & \\ 0 & = & - A_{N-2} & + 3A_{N-3} & + 3A_{N-4} & - A_{N-5} & \\ 0 & = & & - A_{N-3} & + 3A_{N-4} & + 3A_{N-5} & - A_{N-6} \\ \hline A_N & = 6A_{N-1} & - 7A_{N-2} & - 8A_{N-3} & + 9A_{N-4} & + 2A_{N-5} & - A_{N-6} \end{array}$$

It is interesting to note that Eq. (28), itself, can be constructed from

$$A_N = 4A_{N-1} - A_{N-2} \tag{29}$$

by adding Eq. (29) to itself reindexed:

$$A_{N-1} = 4A_{N-2} - A_{N-3}.$$

But Eq. (29) does not yield the saturation degeneracy numbers given in the first diagonal of Table I unless an additional term of $2(-1)^N$ is appended to it, i.e.,

$$A_N = 4A_{N-1} - A_{N-2} + 2(-1)^N. \tag{30}$$

Utilizing Eq. (28), we may write[15]

$$A_N = cS^N \tag{31}$$

so that

$$S^3 - 3S^2 - 3S + 1 = 0 \tag{32}$$

or

$$(S+1)(S^2 - 4S + 1) = 0.$$

Then

$$A_N = c_1[2+\sqrt{3}]^N + c_2[2-\sqrt{3}]^N + c_3[-1]^N. \tag{33}$$

Employing the initial conditions (see the saturation diagonal in Table I)

$$A_0 = 1, \quad A_1 = 2, \quad A_2 = 9, \tag{34}$$

we obtain values for c_1, c_2, c_3 so that Eq. (33) becomes

$$\begin{aligned} A_N &= \tfrac{1}{6}\{(2+\sqrt{3})^{N+1} - 2(-1)^{N+1} + (2-\sqrt{3})^{N+1}\} \\ &= \tfrac{1}{6}\{(2+\sqrt{3})^{(N+1)/2} + (-1)^N(2-\sqrt{3})^{(N+1)/2}\}^2. \end{aligned} \tag{35}$$

As $N \to \infty$, we see from Eq. (35) that the so-called "molecular freedom" takes on the value $[2+3^{1/2}]^{1/2}$, in agreement with previously published numerical results.[16]

VII. CONCLUSION

A 19-term recursion relation has been derived that yields exactly the occupational degeneracy for indistinguishable dumbbells distributed on a $2\times2\times N$ lattice space. (Here we have assumed that the dumbbells have identical ends; if the ends are different then an additional factor of 2^q must be included in the foregoing results.)

On the basis of such a recursion we have calculated the generating functions, normalization, expectation, and variance of the associated statistics. A continuous representation for the occupational degeneracy is also presented, as is the orientational degeneracy for a completely filled lattice space.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to Professor D. Snider for several helpful discussions. We also thank Mr. R. Ristic for drawing the figures.

APPENDIX A

Equation (5) becomes

$$G(x,y)=\sum_{N=0}^{\infty} f_N(x)y^N=\sum_{N=0}^{5} f_N(x)y^N+\sum_{N=6}^{\infty} f_N(x)y^N$$
$$=\sum_{N=0}^{5} f_N(x)y^N+\sum_{N=0}^{\infty} f_{N+6}(x)y^{N+6}. \tag{A1}$$

Using the recursion for $f_{N+6}(x)$, Eq. (4), we may write

$$\begin{aligned}G(x,y)=&\sum_{N=0}^{5} f_N(x)y^N+y[1+7x+6x^2]\sum_{N=0}^{\infty} f_{N+5}(x)y^{N+5}\\&+y^2x[1+6x+6x^2-7x^3]\sum_{N=0}^{\infty} f_{N+4}(x)y^{N+4}\\&-2y^3x^3[1+5x+13x^2+4x^3]\sum_{N=0}^{\infty} f_{N+3}(x)y^{N+3}\\&+y^4x^5[1+2x+6x^2+9x^3]\sum_{N=0}^{\infty} f_{N+2}(x)y^{N+2}\\&+y^5x^8[1-x+2x^2]\sum_{N=0}^{\infty} f_{N+1}(x)y^{N+1}\\&-y^6x^{12}\sum_{N=0}^{\infty} f_N(x)y^N.\end{aligned} \tag{A2}$$

Thus

$$\begin{aligned}G(x,y)\{&1-y[1+7x+6x^2]\\&-y^2x[1+6x+6x^2-7x^3]\\&+2y^3x^3[1+5x+13x^2+4x^3]\\&-y^4x^5[1+2x+6x^2+9x^3]-y^5x^8[1-x+2x^2]\\&+y^6x^{12}\}=\sum_{N=0}^{5} f_N(x)y^N\\&-y[1+7x+6x^2]\sum_{N=0}^{4} f_N(x)y^N\\&-y^2x[1+6x+6x^2-7x^3]\sum_{N=0}^{3} f_N(x)y^N\\&+2y^3x^3[1+5x+13x^2+4x^3]\sum_{N=0}^{2} f_N(x)y^N\\&-y^4x^5[1+2x+6x^2+9x^3]\sum_{N=0}^{1} f_N(x)y^N\\&-y^5x^8[1-x+2x^2]f_0(x),\end{aligned} \tag{A3}$$

where (from Table I)

$$\begin{aligned}f_0(x)&=1,\\f_1(x)&=1+4x+2x^2,\\f_2(x)&=1+12x+42x^2+44x^3+9x^4,\\f_3(x)&=1+20x+142x^2+440x^3\\&\quad+588x^4+288x^5+32x^6,\\f_4(x)&=1+28x+306x^2+1672x^3+4863x^4\\&\quad+7416x^5+5470x^6+1620x^7+121x^8,\\f_5(x)&=1+36x+534x^2+4248x^3+19\,774x^4\\&\quad+55\,200x^5+91\,200x^6+84\,984x^7\\&\quad+40\,553x^8+8204x^9+450x^{10}.\end{aligned} \tag{A4}$$

$G(x,y)$ may then be written as Eq. (6), where $H(x,y)$ is the rhs of Eq. (A3).

To find $H(1,y)$ in Eq. (10), we see from the right-hand column of Table I, Eq. (A3), and Eq. (A4) that

$$H(1,y)=1-7y+4y^2+3y^3-y^4. \tag{A5}$$

APPENDIX B

From Eq. (19),

$$\langle q\rangle_N=2N\Delta_N\langle\theta\rangle_N=\sum_{q=0}^{2N} qA[q,N]. \tag{B1}$$

Utilizing the recursion for $A[q,N]$, Eq. (2), we may write

$$\begin{aligned}2N\Delta_N\langle\theta\rangle_N=&\sum_{q=0}^{2N} qA[q,N-1]\\&+7\sum_{q=0}^{2N} qA[q-1,N-1]\\&+6\sum_{q=0}^{2N} qA[q-2,N-1]\\&+\sum_{q=0}^{2N} qA[q-1,N-2]\\&+6\sum_{q=0}^{2N} qA[q-2,N-2]\\&+6\sum_{q=0}^{2N} qA[q-3,N-2]\\&-7\sum_{1=0}^{2N} qA[q-4,N-2]\\&-2\sum_{q=0}^{2N} qA[q-3,N-3]\\&-10\sum_{q=0}^{2N} qA[q-4,N-3]\\&-26\sum_{q=0}^{2N} qA[q-5,N-3]\\&-8\sum_{q=0}^{2N} qA[q-6,N-3]\\&+\sum_{q=0}^{2N} qA[q-5,N-4]\\&+2\sum_{q=0}^{2N} qA[q-6,N-4]\\&+6\sum_{q=0}^{2N} qA[q-7,N-4]\\&+9\sum_{q=0}^{2N} qA[q-8,N-4]\\&+\sum_{q=0}^{2N} qA[q-8,N-5]\\&-\sum_{q=0}^{2N} qA[q-9,N-5]\\&+2\sum_{q=0}^{2N} qA[q-10,N-5]\\&-\sum_{q=0}^{2N} qA[q-12,N-6].\end{aligned} \tag{B2}$$

However,

$$\sum_q qA\,[q-j,N-k\,]$$
$$=\sum_q (q+j)A\,[q,N-k\,]$$
$$=\sum_q qA\,[q,N-k\,]+j\sum_q A\,[q,N-k\,]$$
$$=2[N-k\,]\Delta_{N-k}\langle\theta\rangle_{N-k}+j\Delta_{N-k}, \tag{B3}$$

so that Eq. (B2) becomes

$$\begin{aligned}2N\Delta_N\langle\theta\rangle_N = {} & 2(N-1)\Delta_{N-1}\langle\theta\rangle_{N-1}\\ & + 14(N-1)\Delta_{N-1}\langle\theta\rangle_{N-1}+7\Delta_{N-1}\\ & + 12(N-1)\Delta_{N-1}\langle\theta\rangle_{N-1}+12\Delta_{N-1}\\ & + 2(N-2)\Delta_{N-2}\langle\theta\rangle_{N-2}+\Delta_{N-2}\\ & + 12(N-2)\Delta_{N-2}\langle\theta\rangle_{N-2}+12\Delta_{N-2}\\ & + 12(N-2)\Delta_{N-2}\langle\theta\rangle_{N-2}+18\Delta_{N-2}\\ & - 14(N-2)\Delta_{N-2}\langle\theta\rangle_{N-2}-28\Delta_{N-2}\\ & - 4(N-3)\Delta_{N-3}\langle\theta\rangle_{N-3}-6\Delta_{N-3}\\ & - 20(N-3)\Delta_{N-3}\langle\theta\rangle_{N-3}-40\Delta_{N-3}\\ & - 52(N-3)\Delta_{N-3}\langle\theta\rangle_{N-3}-130\Delta_{N-3}\\ & - 16(N-3)\Delta_{N-3}\langle\theta\rangle_{N-3}-48\Delta_{N-3}\\ & + 2(N-4)\Delta_{N-4}\langle\theta\rangle_{N-4}+5\Delta_{N-4}\\ & + 4(N-4)\Delta_{N-4}\langle\theta\rangle_{N-4}+12\Delta_{N-4}\\ & + 6(N-4)\Delta_{N-4}\langle\theta\rangle_{N-4}+42\Delta_{N-4}\\ & + 18(N-4)\Delta_{N-4}\langle\theta\rangle_{N-4}+72\Delta_{N-4}\\ & + 2(N-5)\Delta_{N-5}\langle\theta\rangle_{N-5}+8\Delta_{N-5}\\ & - 2(N-5)\Delta_{N-5}\langle\theta\rangle_{N-5}-9\Delta_{N-5}\\ & + 4(N-5)\Delta_{N-5}\langle\theta\rangle_{N-5}+20\Delta_{N-5}\\ & - 2(N-6)\Delta_{N-6}\langle\theta\rangle_{N-6}-12\Delta_{N-6}.\end{aligned} \tag{B4}$$

Utilizing Eq. (20), we see that

$$\begin{aligned}&2N\langle\theta\rangle_N\{-\Delta_N+14\Delta_{N-1}+6\Delta_{N-2}\\ &\quad -46\Delta_{N-3}+18\Delta_{N-4}+2\Delta_{N-5}-\Delta_{N-6}\}\\ &\quad +2\langle\theta\rangle_N\{-14\Delta_{N-1}-12\Delta_{N-2}\\ &\quad +138\Delta_{N-3}-72\Delta_{N-4}-10\Delta_{N-5}+6\Delta_{N-6}\}\\ &\quad +\{19\Delta_{N-1}+3\Delta_{N-2}-224\Delta_{N-3}\\ &\quad +131\Delta_{N-4}+19\Delta_{N-5}+12\Delta_{N-6}\}=0.\end{aligned} \tag{B5}$$

The coefficient of $2N\langle\theta\rangle_N$ vanishes because of Eq. (8). Thus, assuming $\Delta_N=k_1R_1^N$ [see Eq. (17)], we obtain Eq. (21).

APPENDIX C

From Eq. (22), we write

$$\langle q^2\rangle_N=4N^2\Delta_N\langle\theta^2\rangle_N=\sum_{q=0}^{2N}q^2A\,[q,N\,]. \tag{C1}$$

Substituting Eq. (2) for $A\,[q,N\,]$, letting

$$\langle\theta^2\rangle_N=\alpha/N+\langle\theta\rangle_N^2, \tag{C2}$$

and considering a typical term of the resulting expression,

$$\sum_q q^2A\,[q-j,N-k\,]$$
$$=\sum_q (q+j)^2A\,[q,N-k\,]$$
$$=\sum_q q^2A\,[q,N-k\,]+2j\sum_q qA\,[q,N-k\,]$$
$$+j^2\sum_q A\,[q,N-k\,]$$
$$=4(N-k\,)^2\Delta_{N-k}\left\{\frac{\alpha}{(N-k\,)}+\langle\theta\rangle_{N-k}^2\right\}$$
$$+4j(N-k\,)\Delta_{N-k}\langle\theta\rangle_{N-k}+j^2\Delta_{N-k}, \tag{C3}$$

we obtain

$$\begin{aligned}\sigma_N^2\equiv\langle\theta^2\rangle_N-\langle\theta\rangle_N^2=\alpha/N\\ =\frac{1}{N}\Bigg\{&\langle\theta\rangle_\infty^2\left[\frac{14R_1^5+24R_1^4-414R_1^3+288R_1^2+50R_1-36}{14R_1^5+12R_1^4-138R_1^3+72R_1^2+10R_1-6}\right]\\ &-\langle\theta\rangle_\infty\left[\frac{19R_1^5+6R_1^4-672R_1^3+524R_1^2+95R_1-72}{14R_1^5+12R_1^4-138R_1^3+72R_1^2+10R_1-6}\right]\\ &+\frac{1}{4}\left[\frac{31R_1^5-33R_1^4-1116R_1^3+967R_1^2+183R_1-144}{14R_1^5+12R_1^4-138R_1^3+72R_1^2+10R_1-6}\right]\Bigg\},\end{aligned} \tag{C4}$$

thus yielding Eq. (23).

[1]D. Lichtman and R. B. McQuistan, J.Math. Phys. **8**, 2441 (1967); R. B. McQuistan, Nuovo Cimento B **58**, 86 (1968).
[2]H. N. V. Temperley and M. E. Fisher, Philos. Mag. **6**, 1061 (1961).
[3]P. W. Kasteleyn, Physica **27**, 1209 (1961).
[4]E. H. Lieb, J. Math. Phys. **8**, 2339 (1967).
[5]R. B. McQuistan and S. J. Lichtman, J. Math. Phys. **11**, 3095 (1970).
[6]R. B. McQuistan, J. Math. Phys. **13**, 242 (1972).
[7]H. A. Bethe, Proc. R. Soc. London Ser. A **150**, 552 (1935).
[8]J. K. Roberts, Proc. R. Soc. London Ser. A **161**, 141 (1937).
[9]E.H. Lieb, Physica **73**, 237 (1974).
[10]A. F. Antippa and A. J. Phares, J. Math. Phys. **18**, 173 (1977).
[11]A. J. Phares, J. Math. Phys. **18**, 1838 (1977).
[12]A. J. Phares and R. J. Meier, J. Math. Phys. **22**, 1021 (1981).
[13]A. J. Phares, J. Math. Phys. **22** 1030 (1981).
[14]H. W. Gottinger, *Elements of Statistical Analysis* (De Gruyter, Berlin, 1980).
[15]A. J. Phares and K. H. Simmons, Am. J. Phys. **48**, 159 (1980).
[16]R. H. Fowler and G. S. Rushbrooke, Trans. Faraday Soc. **33**, 1272 (1937).

Field theory models in the scattering picture[a)]

M. O. Farrukh
Department of Physics, Faculty of Sciences, Lebanese University, Hadeth, Beirut, Lebanon

(Received 9 November 1982; accepted for publication 23 December 1982)

We propose general schemes for constructing quantum field theory models in the scattering picture. Our models include two types: those which are related to classical pure potential theories and those which are related to conventional quantum field theories. We also discuss the problem of symmetry breaking, and show how to avoid ultraviolet divergencies.

PACS numbers: 11.10.Cd, 11.30.Qc, 11.20.Dj, 11.30.Qc

I. INTRODUCTION

In a previous paper[1] we have shown that a field theory in the scattering picture can be formulated in terms of a potential density $\mathscr{H}(x)$, which satisfies the conditions of

(1) Hermiticity: $\mathscr{H}^{\dagger}(x) = \mathscr{H}(x)$,

(2) Poincaré covariance:
$U(\Lambda, a)\mathscr{H}(x)U^{-1}(\Lambda, a) = \mathscr{H}(\Lambda x + a)$.

The S matrix of the theory has the following expansion:

$$S = I + \sum_{n=1}^{\infty} (-i)^n \int dx_1 \int dx_2 \cdots \int dx_n \, J_n(x_1, x_2,\ldots, x_n) \times \mathscr{H}(x_1)\mathscr{H}(x_2)\cdots\mathscr{H}(x_n), \tag{1}$$

where $\{J_n(x_1, x_2,\ldots, x_n);\ n \in \mathscr{N}^*\}$ is a sequence of functions called the covariant causality functions. They are constructed as follows:

Let $n \in \mathscr{N}^*$ be a positive integer, and denote by Γ_n the set of all permutations on the numbers $(1,2,\ldots, n)$. Let $p \in \Gamma_n$ and denote by $\nu(p)$ the number of pairs of consecutive numbers which come in p in order. The characteristic function C_p is defined as follows:

$$C_p = (-1)^{n-1-\nu(p)}[n-1-\nu(p)]![\nu(p)]!, \tag{2}$$

using which we define the phase generating function

$$C_n(x_1, x_2,\ldots, x_n) = \sum_{p\in\Gamma_n} C_p \eta(x_{p_1} - x_{p_2})\eta(x_{p_2} - x_{p_3})\cdots\eta(x_{p_{n-1}} - x_{p_n}), \tag{3}$$

where $\eta(x) = \Theta(x^0)\Theta(x^2)$, and Θ is the Heaviside step function. If $j \leqslant n$, $j \in \mathscr{N}^*$, then the partitioning function $P_{n,j}$ is defined as follows:

$$P_{n,j}(x_1, x_2,\ldots, x_n) = \sum \frac{1}{n_1!\, n_2! \cdots n_j!} C_{n_1}(x_1, x_2,\ldots, x_{n_1}) \times C_{n_2}(x_{n_1+1},\ldots, x_{n_1+n_2})\cdots C_{n_j}(x_{n-n_j+1},\ldots, x_n), \tag{4}$$

where the summation is carried over all the numbers $n_1, n_2,\ldots, n_j \in \mathscr{N}^*$ such that $n_1 + n_2 + \cdots + n_j = n$. Finally, the covariant causality function J_n is defined as

$$J_n(x_1, x_2,\ldots, x_n) = \sum_{j=1}^{n} \frac{1}{j!} P_{n,j}(x_1, x_2,\ldots, x_n). \tag{5}$$

In particular, $J_1(x) = 1$ and $J_2(x, y) = \frac{1}{2}[\eta(x-y) - \eta(y - x) + 1]$. The S matrix thus constructed is unitary and Poincaré-invariant. Any other symmetry required to be possessed by the S matrix should be imposed on the potential density $\mathscr{H}(x)$.

In the previous paper[1] we have required $\mathscr{H}(x)$ to be local, in the sense that it should be a linear combination of densities of the form

$$\mathscr{D}(x) = \int dy_1 \int dy_2 \cdots \int dy_n \, R{:}\, \Phi(y_1)\Phi(y_2)\cdots\Phi(y_n){:} \times \Delta_n(x; y_1, y_2,\ldots, y_n) \tag{6}$$

in case of a neutral scalar field. The distribution Δ_n is of the form

$$\Delta_n(x; y_1, y_2,\ldots, y_n) = \int \bar{d}p_1 \int \bar{d}p_2 \cdots \int \bar{d}p_n \times e^{ip_1(y_1 - x)}\cdots e^{ip_n(y_n - x)}\widetilde{\Delta}_n(p_1, p_2,\ldots, p_n). \tag{7}$$

The vertex function $\widetilde{\Delta}_n$ may be expressed in terms of a function $F(z)$ which is analytic in a region containing the whole real line. There exists a number $m \in \mathscr{N}$, $m \leqslant n/2$ such that

$$\widetilde{\Delta}_n(p_1, p_2,\ldots, p_n) = F((p_1 + p_2 + \cdots + p_{n-m} - p_{n-m+1} - \cdots - p_n)^2) + \text{distinct permutations}. \tag{8}$$

The above concept of locality needs to be generalized to cases of other elementary fields. Also, there is some ambiguity in the form of the vertex function given in (8).

This paper is devoted to the solution of this problem. Our main results can be summarized as follows: Whatever the fields present in the model are, one has to construct a bilinear form $\mathscr{B}(y_1, y_2)$ and a multilinear form $\mathscr{M}(y_1, y_2,\ldots y_n)$, $n \geqslant 2$, such that:

(1) The two forms $\mathscr{B}$ and $\mathscr{M}$ are covariant scalars:

$$U(\Lambda, a)\mathscr{B}(y_1, y_2)U^{-1}(\Lambda, a) = \mathscr{B}(\Lambda y_1 + a, \Lambda y_2 + a),$$
$$U(\Lambda, a)\mathscr{M}(y_1, y_2,\ldots, y_n)U^{-1}(\Lambda, a) = \mathscr{M}(\Lambda y_1 + a,\ldots, \Lambda y_n + a). \tag{9}$$

(2) Any symmetry possessed by $\mathscr{H}(x)$ should be possessed by the forms $\mathscr{B}$ and $\mathscr{M}$ independently.

(3) The vertex function $\widetilde{\Delta}_{n+2}$ has the form

$$\widetilde{\Delta}_{n+2}(p_1, p_2,\ldots, p_{n+2}) = F\left(\left(\sum_{i=1}^{n} p_i - p_{n+1} - p_{n+2}\right)^2\right), \tag{10}$$

and the distribution Δ_{n+2} is deduced from (10) using (7).

[a)] Work supported by the Lebanese university council.

(4) The potential density $\mathscr{H}(x)$ is given by

$$\mathscr{H}(x) = \int dy_1 \int dy_2 \cdots \int dy_{n+2}\, R: \times \mathscr{M}(y_1, y_2, \ldots, y_n)\mathscr{B}(y_{n+1}, y_{n+2}): \times \Delta_{n+2}(x; y_1, y_2, \ldots, y_{n+2}). \tag{11}$$

The multilinear form $\mathscr{M}$ establishes the relation with conventional theories. If $n = 2$, then we have a pure potential theory—with no analog in conventional quantum field theory. If $n > 2$, then the form $\mathscr{M}$ is related to the interaction Hamiltonian of conventional quantum field theory as follows:

$$\mathscr{H}_{\text{int}}(x) = g\!: \mathscr{M}(x, x, \ldots, x):, \tag{12}$$

where g is the coupling constant.

II. CLASSICAL THEORIES

By a classical theory we mean a pure potential theory with a classical analogue.

Assume that the potential energy of two particles of charges q_1 and q_2 located at r_1 and r_2, respectively, is $q_1 q_2 f(|r_1 - r_2|)$, where $f: \mathscr{R}^*_+ \to \mathscr{R}$ is a function such that $\int_0^\infty r|f(r)|\,dr < +\infty$. Define now

$$F(z) = 8\pi \int_0^\infty f(r) M(\tfrac{1}{4} z r^2) r^2\, dr, \quad z \leqslant 0, \tag{13}$$

where

$$M(x) = \sum_{n=0}^{\infty} \frac{x^n}{(2n+1)!} \tag{14}$$

and assume that $F(z)$ defined by (13) for $z \leqslant 0$ has an analytic extension for all $z > 0$. If $\tilde{f}(k) = \int d^3 r f(r) e^{-ik\cdot r}$ is the Fourier transform of $f(r)$, then it follows from (13) that

$$\tilde{f}(k) = \tfrac{1}{2} F(-4k^2). \tag{15}$$

Define now the vertex function

$$\tilde{\Delta}(p_1, p_2, p_3, p_4) = F((p_1 + p_2 - p_3 - p_4)^2) \tag{16}$$

and assume that $\Phi(x)$ and $\Psi(x)$ are the free fields of particles of masses m_1 and m_2 and charges q_1 and q_2, respectively.

Consider now the bilinear form

$$\mathscr{B}(x, y) = m_1 q_1 \Phi^\dagger(x)\Phi(y) + m_2 q_2 \Psi^\dagger(x)\Psi(y) \tag{17}$$

and define the potential density

$$\mathscr{H}(x) = \int dy_1 dy_2 dy_3 dy_4\, R\!:\mathscr{B}(y_1, y_2)\mathscr{B}(y_3, y_4): \times \Delta(x; y_1, y_2, y_3, y_4), \tag{18}$$

where $\Delta(x; y_1, y_2, y_3, y_4)$ is formed from (16) by means of (7). One can easily verify that $\mathscr{H}(x)$ is a covariant self-adjoint density.

It is not difficult to verify that the operator $V = \int \mathscr{H}(0, x)\, d^3x$ has exactly the same behavior as the classical potential in the nonrelativistic limit as $c \to \infty$.

The two bilinear forms in (18) need not be identical, because, if we assume that like particles do not interact, then (18) should be replaced by

$$\mathscr{H}(x) = \int dy_1 dy_2 dy_3 dy_4\, R\!:\Phi^\dagger(y_1)\Phi(y_2)\Psi^\dagger(y_3)\Psi(y_4): \times g\Delta(x; y_1, y_2, y_3, y_4), \tag{19}$$

where $g = 2m_1 m_2 q_1 q_2$.

In view of the preceding discussion, we conclude that a relativistic pure potential theory has a vertex function given by (16) and a potential density given by

$$\mathscr{H}(x) = \int dy_1 dy_2 dy_3 dy_4\, R\!:\mathscr{B}_1(y_1, y_2)\mathscr{B}_2(y_3, y_4): \times \Delta(x; y_1, y_2, y_3, y_4), \tag{20}$$

where $\mathscr{B}_i$ $(i = 1,2)$ is a bilinear form satisfying

$$\mathscr{B}_i^\dagger(x, y) = \mathscr{B}_i(y, x),$$
$$U(\Lambda, a)\mathscr{B}_i(x, y)U^{-1}(\Lambda, a) = \mathscr{B}_i(\Lambda x + a, \Lambda y + a) \tag{21}$$

III. NONCLASSICAL THEORIES

By a nonclassical theory we mean a theory which allows for creation and annihilation of particles. Hence, it is not a pure potential theory.

In conventional quantum field theory, the interaction Hamiltonian which allows creation and annihilation of particles has the form (12). If the vertex function in (7) is the coupling constant g, then (12) can be put into the form

$$\mathscr{H}_{\text{int}}(x) = \int dy_1 dy_2 \cdots dy_n\!: \times \mathscr{M}(y_1, y_2, \ldots, y_n)\!:\Delta_n(x; y_1, y_2, \ldots, y_n),$$

$$\text{where} \quad \Delta_n = g \prod_{i=1}^{n} \delta^{(4)}(x - y_i). \tag{22}$$

One may expect that we can construct the corresponding scattering picture potential density by inserting the R product and modifying the Δ_n distribution in (22). A deeper look reveals the fact that this is the wrong way of dealing with the problem. In quantum electrodynamics[2] and pseudoscalar pion–nucleon coupling theory,[3] the interaction Hamiltonian is a trilinear form, which is set equal to zero by the R product. Other theories, such as the theory of weak interactions[4] and the famous Φ^4 theory of self-interaction of a neutral scalar field[3] are made by the R product pure potential theories. Whatever the number of the field variables in the interaction Hamiltonian is, the R product drops some of the term. Hence, some of the physical information is lost. Thus, to save all terms, we must multiply the interaction Hamiltonian by a factor before taking the R product. This factor should allow for the following requirements to be fulfilled:

(1) All terms in the original Hamiltonian should be preserved.

(2) The new factor should neither add any new interaction nor destroy any of the symmetries of the original Hamiltonian.

To satisfy the first requirement, we recall that the R product drops vacuum fluctation and self-energy terms. To preserve these terms, we need at most either two creation or two annihilation operators. Hence, if we multiply the interaction Hamiltonian by a bilinear form in all the field variables used, we can satsify the first requirement.

To show that the second requirement is satisfied, consider the following example of the Lagrangian of a scalar neutral field[5]:

$$\mathscr{L}(x) = \tfrac{1}{2}\,\partial_\mu \Phi(x)\partial^\mu \Phi(x) - \tfrac{1}{2}\, m^2\Phi^2(x) - P(\Phi(x)), \quad (23)$$

where $P(\Phi(x))$ is a polynomial of an even degree with a positive leading coefficient. The term $\lambda\Phi^2(x)$ is a mass shift term. Hence it does not add any new interaction. Also, since it is proportional to a term in the Lagrangian, it possesses all the symmetries of the Lagrangian.

The weak point in this deduction is that the term $\lambda\Phi^2(x)$ does not add any new interaction when it is added to and not multiplied by $P(\Phi(x))$. However, one may argue that apart from divergencies, $:P(\Phi(x)):$ and $R: \Phi^2(x)P(\Phi(x)):$ represent qualitatively the same theory. Thus, the second requirement is, in a sense, also satisfied.

Now, in general, the potential density should have the form (11). Comparing the vertex function (16) with its corresponding potential density (18), it seems to be very plausible that (10) is the vertex function of the potential density (11).

As an example, the interaction Hamiltonian of conventional quantum electrodynamics[2] is

$$\mathscr{H}_{\rm int}(x) = e\overline{\Psi}(x)\gamma^\mu\Psi(x)A_\mu(x), \quad (24)$$

where $\Psi(x)$ is a Dirac spinor and $A_\mu(x)$ is the electromagnetic field. The corresponding scattering picture potential density is

$$\begin{aligned}\mathscr{H}(x) = \int dy_1 dy_2 dy_3 dy_4 dy_5\, R{:}\overline{\Psi}(y_1)\gamma^\mu\Psi(y_2)A_\mu(y_3)\\ \times\big[a\overline{\Psi}(y_4)\Psi(y_5) + bA^\mu(y_4)A_\mu(y_5)\big]{:}\\ \times\Delta(x; y_1, y_2, y_3, y_4, y_5), \quad (25)\end{aligned}$$

and the vertex function is

$$\widetilde{\Delta}(p_1, p_2, p_3, p_4, p_5) = F((p_1+p_2+p_3-p_4-p_5)^2), \quad (26)$$

where a and b are real numbers and $F(z)$ is a real function of z, which is analytic in a region containing the whole real line. The numbers a and b, and the function F have still to be determined.

IV. SYMMETRY BREAKING

In the literature, the problem of approximate symmetries is treated in two ways: The first way is to break the symmetry by adding nonsymmetric terms to the Hamiltonian, such as, for example, the inclusion of medium strong, electromagnetic and weak interaction terms which violate SU(3) symmetry, in the Hamiltonian of hadrons,[6] supposed to contain a very strong SU(3)-invariant term from which the main contribution to the interaction comes. The mass differences among the hadrons forming an SU(3) multiplet can be accounted for from calculations of self-energies. This method cannot be applied within the framework of the scattering picture formalism, because of the presence of the R product which eliminates self energies.

The second way is to allow the symmetry to break spontaneously[7] by having a degenerate vacuum. This method, also, cannot be applied within the framework of the scattering picture formalism, because, the underlying Hilbert space has, by construction, a unique vacuum.

We shall treat the two aspects of the problem of approximate symmetries independently. On one hand, we allow for the presence of terms which violate the symmetry assumed to be approximately possessed by the potential density, such as, for example, the presence of an electromagnetic term which violates isospin invariance in a potential density containing an isospin-invariant strong-interaction term. On the other hand, we vary the masses of the particles forming a multiplet to make them identical with the physical values, which means that the assumed symmetry preserving term does not really preserve that symmetry. The effect of this mass variation on the potential density is treated as follows.

Suppose that $\{\Psi_i;\ i = 1,2,\ldots, n\}$ are different states of the same field in some internal space and that $F(\Psi_i)$ is an invariant under rotation in that internal space. Suppose also that the assumed mass of the Ψ field is m. When this symmetry is slightly broken, each component Ψ_i will have a different mass m_i. Thus, we have $\Psi_i \to \widetilde{\Psi}_i$ in the substitution $m \to m_i$. One may expect that we replace $F(\Psi_i)$ by $F(\widetilde{\Psi}_i)$, which violates the considered symmetry. However, in the limit as $m_i \to m$, $i = 1,2,\ldots, n$, we have

$$F((m_i/m)^\lambda \widetilde{\Psi}_i) \to F(\Psi_i), \quad \text{for any } \lambda \in \mathscr{R}.$$

We show in what follows that the choice of λ depends on the normalization of the field Ψ. So, if we normalize Ψ suitably, we can make $\lambda = 0$.

Going back to (17), we find that if the interaction is symmetric with respect to the interchange $\Phi \leftrightarrow \Psi$, then $q_1 = q_2 = q$, and (17) becomes

$$\mathscr{B}(x, y) = q\big[m_1\Phi^\dagger(x)\Phi(y) + m_2\Psi^\dagger(x)(y)\big] \quad (27)$$

Now, since Φ and Ψ are scalar fields, they have the expansions

$$\begin{aligned}\Phi(x) &= \int \frac{d^3p}{2E_p}\big[a(p)e^{-ipx} + b^\dagger(p)e^{ipx}\big],\\ \Psi(x) &= \int \frac{d^3p}{2E_p}\big[c(p)e^{-ipx} + d^\dagger(p)e^{ipx}\big]. \quad (28)\end{aligned}$$

If we define $u(p) = v(p) = \sqrt{m}$ and write

$$\begin{aligned}\Phi_0(x) &= \int \frac{d^3p}{2E_p}\big[a(p)u(p)e^{-ipx} + b^\dagger(p)v(p)e^{ipx}\big],\\ \Psi_0(x) &= \int \frac{d^3p}{2E_p}\big[c(p)u(p)e^{-ipx} + d^\dagger(p)v(p)e^{ipx}\big], \quad (29)\end{aligned}$$

then (27) becomes

$$\mathscr{B}(x, y) = q\big[\Phi_0^\dagger(x)\Phi_0(y) + \Psi_0^\dagger(x)\Psi_0(y)\big]. \quad (30)$$

Now, a general spinor field has the form

$$\begin{aligned}\Psi(x) = \int \frac{d^3p}{2E_p}\big[a_\lambda(p)u^\lambda(p)e^{-ipx}\\ + b^{\dagger\lambda}(p)v_\lambda(p)e^{ipx}\big] \quad (31)\end{aligned}$$

So, by normalizing the u and v functions to have the unit of $\sqrt{m}$, an expansion like (30) results upon breaking the symmetry, and hence $F(\Psi_i) \to F(\widetilde{\Psi}_i)$, with $\lambda = 0$.

As an example, the isospin covariant pion field is

$$\Phi_k(x) = \int \frac{d^3p}{2E_p}\big[a_k(p)e^{-ipx} - a^{\dagger j}(p)C_{jk}e^{ipx}\big], \quad k = -1,0,1, \quad (32)$$

where

$$C_{jk} = (-1)^{1-j}\delta_{j,-k}, \quad j,k = -1,0,1. \tag{33}$$

Also, the isospin covariant nucleon field is

$$\Psi_i(x) = \int \frac{d^3p}{2E_p}\,[n_{i,\lambda}(p)u^\lambda(p)e^{-ipx} + B_{ij}\bar{n}^{\dagger j,\lambda}(p)v_\lambda(p)e^{ipx}], \quad i = -\tfrac{1}{2},\tfrac{1}{2}, \tag{34}$$

where

$$B_{ij} = (-i)^{1/2-i}\delta_{ij}, \quad i,j = -\tfrac{1}{2},\tfrac{1}{2}. \tag{35}$$

The u and v functions are normalized such that

$$\bar{u}_\lambda(p)u^{\lambda'}(p) = -\bar{v}'^\lambda(p)v_\lambda(p) = 2m\delta_\lambda^{\lambda'} \tag{36}$$

where

$$\bar{u}_\lambda(p) = [u^\lambda(p)]^\dagger A, \quad \bar{v}^\lambda(p) = [v_\lambda(p)]^\dagger A, \quad A = \mathrm{diag}(1,1,-1,-1). \tag{37}$$

We define the conjugate nucleon field as

$$\bar{\Psi}^i(x) = [\Psi_i(x)]^\dagger A. \tag{38}$$

Define now the isospin invariant trilinear form

$$\mathscr{M}(x,y,z) = \bar{\Psi}^i(x)\gamma^5\Psi_j(y)\Phi_k(z)\langle\tfrac{1}{2}\,i|\tfrac{1}{2}ij1k\rangle, \tag{39}$$

where

$$\gamma^5 = -i\begin{pmatrix}0 & I_2\\ I_2 & 0\end{pmatrix}$$

and $I_2 = \mathrm{diag}(1,1)$, and $\langle\frac{1}{2}\,i|\frac{1}{2}\,jlk\rangle$ is the Clebsch–Gordan coefficient with values[8]

$$\langle\tfrac{1}{2}\tfrac{1}{2}|\tfrac{1}{2}\tfrac{1}{2}10\rangle = -\langle\tfrac{1}{2}-\tfrac{1}{2}|\tfrac{1}{2}-\tfrac{1}{2}10\rangle = 1/\sqrt{3}$$

$$\langle\tfrac{1}{2}-\tfrac{1}{2}|\tfrac{1}{2}\tfrac{1}{2}1-1\rangle = -\langle\tfrac{1}{2}\tfrac{1}{2}|\tfrac{1}{2}-\tfrac{1}{2}11\rangle = -\sqrt{2/3} \tag{40}$$

We now break the isospin symmetry and define:

(1) The proton field $\Psi_p(x) = \Psi_{1/2}(x)$ with $\bar{\Psi}_p(x) = \bar{\Psi}^{1/2}(x)$;

(2) The neutron field $\Psi_n(x) = \Psi_{-1/2}(x)$ with $\bar{\Psi}_p(x) = \bar{\Psi}^{-1/2}(x)$;

(3) The neutral pion field $\Phi(x) = \Phi_0(x)$;

(4) The charged pion field $\Theta(x) = \Phi_1(x)$ with $\Theta^\dagger(x) = -\Phi_{-1}(x)$.

Taking into account the normalization of the Φ_k and Ψ_i fields shown in (32) and (36), and assuming that the neutral pion mass is the mass of the isospin invariant pion field, we get from (39) and (40), upon breaking the isospin symmetry, the following trilinear form:

$$\begin{aligned}\tilde{\mathscr{M}}(x,y,z) = (1/\sqrt{3})&[\bar{\Psi}_p(x)\gamma^5\Psi_p(y)\Phi(z) - \bar{\Psi}_n(x)\gamma^5\Psi_n(y)\Phi(z)]\\ &- (2m_{\pi^\pm}/3m_{\pi^0})^{1/2}[\bar{\Psi}_n(x)\gamma^5\Psi_p(y)\Theta^\dagger(z) + \bar{\Psi}_p(x)\gamma^5\Psi_n(y)\Theta(z)].\end{aligned} \tag{41}$$

V. MEROMORPHIC POTENTIALS

By a meromorphic potential we mean a potential $f(r)$ such that its transform $F(z)$ given by (13) is a meromorphic function with no singularities other than poles of finite orders in the finite plane.

Let $f(r) = r^{n-1}e^{-\mu r}e^{-i\omega r}$, with $n \in \mathscr{N}$, $\mu \in \mathscr{R}_+^*$, and $\omega \in \mathscr{R}$. Applying (13) to this form, we get

$$F(z) = 32\pi\cdot 2^n\cdot n!\cdot\frac{\Sigma_k C_{n+1}^{2k+1}[2(\mu+i\omega)]^{n-2k}z^k}{[4(\mu+i\omega)^2 - z]^{n+1}}. \tag{42}$$

Hence, $F(z)$ is a meromorphic function with a pole of order $n+1$ at $z = 4(\mu+i\omega)^2$. These poles may lie anywhere on the complex plane except on the nonpositive part of the real line. This restriction can be eliminated by taking the limit $\mu \to 0^+$.

Now, if $\mu\omega = 0$, then $F(z)$ would have a real pole; thus $F(z)$ cannot be analytic on the whole real line as required by the theory. Hence, if we want to use a meromorphic potential, then we should have $\mu > 0$ and $\omega \neq 0$. Thus, neither the Coulomb potential ($n = 0, \mu = 0^+, \omega = 0$) nor the Yukawa potential ($n = 0, \mu > 0, \omega = 0$) can be used in a convergent theory.

VI. ULTRAVIOLET DIVERGENCIES

As proved in the previous paper,[1] if we could avoid divergencies in the second order of perturbation, then all the terms of the perturbation series would be finite. The second term in the series (1) has the form

$$S_2 = -\int dx\int dy\, J_2(x,y)\mathscr{H}(x)H(y), \tag{43}$$

where

$$J_2(x,y) = \tfrac{1}{2}[\eta(x-y) - \eta(y-x) + 1)]. \tag{44}$$

One can easily prove that

$$\eta(x) = \lim_{\epsilon\to 0^+} 8\pi\int\frac{e^{ipx}d^4p}{[(p^0 - i\epsilon)^2 - \mathbf{p}^2]^2}. \tag{45}$$

Now, setting

$$\mathscr{H}(x) = \int\tilde{\mathscr{H}}(p)e^{i\Sigma px}\,dp, \tag{46}$$

where the integration in (46) is carried over some phase space, we get

$$\begin{aligned}S_2 = -\int dp_1dp_2\delta^{(4)}(\Sigma p_1 - \Sigma p_2)\tilde{\mathscr{H}}(-p_1)\tilde{\mathscr{H}}(p_2)\\ \times\left(\frac{4\pi}{[(\Sigma p_2^0 - i\epsilon)^2 - (\Sigma\mathbf{p}_2)^2]^2} - \frac{4\pi}{[(\Sigma p_2^0 + i\epsilon)^2 - (\Sigma\mathbf{p}_2)^2]^2} + \frac{1}{2}\delta^{(4)}(\Sigma p_2)\right).\end{aligned} \tag{47}$$

The first two terms of (47) are similar, while the third one is always convergent. Hence, we study the convergence of the first term.

In Feynman graph terminology,[9] to an internal line representing the exchange of a massive particle of spin s we have a factor

$$\frac{d^3p}{2E_p}\Pi(p), \quad \Pi(p) = u^\lambda(p)\bar{u}_\lambda(p)\sim p^{2s}. \tag{48}$$

If we have n internal lines of total spin $s = s_1 + s_2 + \cdots + s_n$, and taking into consideration the factor $1/[(p^0 - i\epsilon)^2 - \mathbf{p}^2]^2$, we have the following estimate of the scattering amplitude:

$$\int \frac{d^3p_1}{2E_{p_1}} \cdots \frac{d^3p_n}{2E_{p_n}} \frac{\Pi_1(p_1)\otimes\cdots\otimes\Pi_n(p_n)}{[(p^0-i\epsilon)^2-\mathbf{p}^2]^2}$$

$$\times F(p''^2)F(p'^2) \sim \int p^{2n+2s-5}|F(p^2)|^2\,dp\,. \tag{49}$$

The convergence of (49) requires that, for large values of p, $F(p^2)\sim 1/p^{n+s-2+\epsilon}$, $\epsilon>0$. In other words, if for large values of z, $F(z)\sim 1/z^{\kappa}$, then the convergence of the theory implies that for any Feynman graph of n internal lines of massive particles of total spin s, we must have

$$\kappa > \tfrac{1}{2}(n+s)-1\,. \tag{50}$$

If we have a massless particle in the theory, then some restrictions are imposed on its projection matrix $\Pi(p)$. In the irreducible $2j+1$ and the reducible $2(2j+1)$ representations of fields, the limit $m\to 0$ for massless particles[10] has no effect on the high energy behavior of the projection matrices. Other representations need some care. For example,[11] for the spin-1 4-vector massive Proca field we have $\Pi(p)\sim p^2$ as expected, while for the spin-1 4-vector massless Maxwell field we have $\Pi(p)\sim 1$. Thus, the spin of the photon should be counted as zero in applying formula (50), if we use the 4-vector representation.

As an application, consider the potential density of quantum electrodynamics given in (25). Since we have five field variables, the maximum number of internal lines is $n=3$. If these internal lines are Dirac spinors, then $s=3/2$. This is the maximum value of s, because, as we have just said, a photon in this representation does not contribute to s in (50). Thus, $\kappa > \frac{1}{2}(n+s)-1=\frac{5}{4}$. If we require $F(z)$ to be meromorphic, then the minimum value of κ is $\kappa=2$.

VII. CONCLUSION AND DISCUSSION

The schemes outlined in the previous sections allow us to construct convergent quantum field theory models in a simple way. A scalar multilinear form is either written directly or deduced from classical Lagrangian field theory. This form is then multiplied by a bilinear form of all the fields used in the model. The R product is taken, and the result is smeared, using a distribution whose Fourier transform is a function analytic over the whole real line. The requirement that all terms of perturbation be finite imposes an asymptotic limit on this function. If we require this function to be meromorphic, then the whole problem reduces to determining a finite set of parameters. As an example, we have seen that if we use a meromorphic function for electrodynamics, then $F(z)\sim 1/z^2$ for large values of z. The simplest such form is

$$F(z)=1/[(z-\mu)^2+M^2]\,. \tag{51}$$

Hence, all what is left is to determine four parameters: a and b appearing in (25), and μ and M appearing in (51).

Also, we mention that our schemes allow for models of covariant potential theory, a thing which cannot be done in conventional quantum field theory.

[1]M. O. Farrukh, "Potential densities for quantum field theories in the scattering picture," J. Math. Phys. **24**, 1645 (1983).
[2]H. Muirhead, "Electromagnetic Interactions," in *The Physics of Elementary Particles* (Pergamon, Oxford, London, 1968).
[3]H. Muirhead, "Strong Interactions II," in Ref. 2.
[4]H. Muirhead, "The Weak Interactions," in Ref. 2.
[5]J. Glimm and A. Jaffe, "Boson Quantum Field Models," in *Mathematics of Contemporary Physics*, edited by R. F. Streater (Academic, London, 1972).
[6]W. R. Frazer, "The Eightfold Way," in *Elementary Particles* (Prentice-Hall, Englewood Cliffs, N. J., 1966).
[7]L. M. Sehgal, "Unified Theories of Weak and Electromagnetic Interactions, III, Physikalisches Institut Technische Hochschule Aachen, 1973.
[8]M. A. Preston, *Physics of the Nucleus* (Addison-Wesley, Reading, Mass., 1965).
[9]S. Weinberg, "Feynman Rules for Any Spin," Phys. Rev. **133**, B1318 (1964).
[10]S. Weinberg, "Feynman Rules for Any Spin. II. Massless Particles," Phys. Rev. **134**, B882 (1964).
[11]Y. Takahashi, *An Introduction to Field Quantization* (Pergamon, Oxford, London, 1969).

Splitting of the connection in gauge theories with broken symmetry

R. O. Fulp and L. K. Norris
Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650

(Received 29 December 1981; accepted for publication 17 September 1982)

We obtain a fully geometric analog of the Higgs mechanism whereby a symmetry-breaking Higgs field is used to impart mass to gauge fields. We do this by showing that under fairly general hypotheses a symmetry-breaking Higgs field ϕ on a "principal bundle with connection" (P,ω) allows the decomposition of the connection ω into a pair (ω',τ) where ω' is a connection on P that reduces to a ϕ-subbundle of P and where τ is a tensorial field on P. The gauge fields that remain massless are identified with the components of ω' while the gauge fields that acquire mass are identified with the components of τ. This decomposition of the connection is exploited in the case where the group of the bundle is the conformal group which scales some fixed metric of arbitrary signature. The geometry of such a bundle with connection generalizes Weyl geometry and provides a bundle setting for conformal gauge theories. We then show that the Weinberg–Salam electroweak theory can be recast as a conformal gauge theory. A primary feature of our conformal version of the Weinberg–Salam theory is that it provides a geometrical interpretation of the surviving component of the Higgs scalar field as an infinitesimal conformal factor.

PACS numbers: 11.15 – q, 02.40. – k

I. INTRODUCTION

It is now well known that gauge theories can be formulated as geometrical theories using the mathematical theory of connections on principal fiber bundles.[1–3] In the geometrization of gauge theory the two fundamental fields, the gauge fields (potentials) and the particle fields, are identified with geometrical objects defined on an appropriate bundle manifold. Specifically, gauge fields are identified with the 1-form components of a connection on the bundle and particle fields are identified with vector-valued functions on the bundle manifold that transform tensorially under different representations of the gauge group.

In this paper we investigate certain aspects of the geometrical interplay between particle and gauge fields in spontaneously broken gauge theories. Specifically this paper addresses the following two questions: (I) Can the symmetry-breaking Higgs field in spontaneously broken gauge theories be geometrized, that is be replaced by some geometrical object,[4] and (II) How does one distinguish, geometrically, the massless and massive vector bosons in spontaneously broken gauge theories?

We present an answer to (I) based on "generalized" *conformal geometry* that exploits the implicit metrical substructure of U(n)-type[5] gauge theories. In particular we will show that the component of the Higgs field that survives symmetry breakdown can be geometrized in terms of an infinitesimal conformal factor. Regarding question (II) we stress a fundamental theorem[6] on the decomposition of a connection into an induced connection and a tensorial 1-form on a subbundle. We are led to conjecture that massless and massive vector bosons be identified with the 1-form components of the induced connection and the tensorial form, respectively. As a specific application of these ideas we reformulate the Weinberg–Salam electroweak theory as a "generalized" conformal theory.

Let us begin by recalling a well known local construction of a symmetry-breaking Higgs field. For a local discussion there is no loss in generality in assuming that our principal bundle is trivial. Thus temporarily we let $P = M \times G$ where M denotes the four-dimensional spacetime base manifold and G denotes an n-dimensional gauge Lie group. We consider a representation of the gauge group on a vector space V and denote the action of G on V by $g\cdot v$ for all $g\in G$ and $v\in V$.

A (generalized) Higgs field on P is any map[7] $\phi: P\rightarrow V$ such that $\phi((a,g)) = g^{-1}\cdot\phi((a,e))$ where e is the identity in G. Such a Higgs field will be a "symmetry-breaking" Higgs field if it selects out a symmetry subgroup $H\subseteq G$ and a corresponding subbundle $Q = M\times H$. (See Sec. 2 for a more precise definition of Higgs field and symmetry-breaking.) In this local setting it is obvious how to construct a symmetry-breaking field. Choose a "gauge" in P, that is a section $x\rightarrow s(x) = (x,a(x))\in M\times G$, together with any vector $\xi_0\in V$. Then one may define a Higgs field $\phi: P\rightarrow V$ by

$$\phi(x,a(x)\cdot g) = g^{-1}\cdot\xi_0. \tag{1.1}$$

The Higgs field defined in this way is symmetry-breaking because over each $x\in M$ it selects out the subgroup G_{ξ_0}, the so-called isotropy subgroup, which is the set of all $g\in G$ such that $g\cdot\xi_0 = \xi_0$. The map ϕ thus reduces P to the subbundle $Q = M\times G_{\xi_0}\equiv\phi^{-1}(\xi_0)$.

Now although Higgs fields exist in general (any map $\phi: P\rightarrow V$ satisfying the transformation law stated above will do) symmetry-breaking Higgs fields do not. It is natural to investigate the extent to which a general Higgs field may be modified or decomposed in terms of symmetry-breaking fields and we turn now to this question, but restrict the discussion to Higgs fields that locally take the form[8]

$$\phi((x,a(x)\cdot g)) = g^{-1}\cdot\Omega(x)(f(x)\xi_0). \tag{1.2}$$

Here the notation is as in (1.1) above with the inclusion of an arbitrary smooth group-valued function Ω and an arbitrary smooth positive-valued function $f: M\rightarrow {}^{+}\mathbb{R}$.

Now it can happen that the gauge group G contains ${}^{+}\mathbb{R}$ as a subgroup that acts on V by scalar multiplication. In this

 0022-2488/83/071871-17$02.50

case we can "rechoose the gauge" so as to absorb the factor $f(x)$ in (1.2). By a further gauge transformation using $\Omega(x)\in G$ we can give ϕ the constant value ξ_0. In this case the Higgs field defined by (1.2) above is symmetry-breaking. (See Theorem 3.1 below.) On the other hand, if G does not contain ${}^+\mathbb{R}$ as a subgroup then it is clear from (1.2) that there is no possibility of rechoosing the gauge so that ϕ has everywhere the constant value ξ_0. The field ϕ will therefore not be a symmetry-breaking field in this case.

One approach to this problem follows from the observation that for each $r\in{}^+\mathbb{R}$ and suitable $f:M\to{}^+\mathbb{R}$, the level surface $f^{-1}(r)$ will be a submanifold N_r of M. Over each such submanifold the field defined by (1.2) will be a symmetry-breaking Higgs field. This "decomposition" approach is treated in detail in Sec. 3.

An alternative remedy for this situation would be to enlarge the principal bundle $P=M\times G$ to a new bundle $\overline{P}=M\times(G\times{}^+\mathbb{R})$ and to extend the action of G on V to an action of $G\times{}^+\mathbb{R}$ on V in such a way that the factor $f(x)$ in (1.2) above can be absorbed by an ${}^+\mathbb{R}$-valued gauge transformation. Then the extension of ϕ to $\overline{P}$ is a symmetry-breaking Higgs field. This method is clearly artificial unless the enlarged bundle and/or the extended field has a natural physical or perhaps geometrical meaning. Now if V is equipped with a metric $\rho: V\times V\to\mathbb{C}$ (as is the case in $U(n)$-type[5] gauge theories) then we may assume without loss of generality that ξ_0 in (1.2) above has unit length: $\rho(\xi_0,\xi_0)=1$. Thus in the gauge $\bar{s}$ in which ϕ takes the form $\phi(\bar{s}(x))=f(x)\cdot\xi_0$ we have that $\sqrt{f(x)}$ is the "length of $\phi(\bar{s}(x))$," that is $\rho(\phi(\bar{s}(x)),\phi(\bar{s}(x)))=f(x)$. Alternatively one may choose to interpret ϕ as a unit vector with respect to a conformally rescaled metric $\bar{\rho}$ defined by

$$\bar{\rho}_x(u,v)=(1/f(x))\rho(u,v).$$

In this way we are led to relate a Higgs field of the type (1.2) to a symmetry-breaking Higgs field and a conformal structure on an ${}^+\mathbb{R}$-enlarged bundle. In the case that $G=\mathrm{U}(n)$ then $G\times{}^+\mathbb{R}=\mathrm{U}(n)\times{}^+\mathbb{R}$ is the conformal group $\mathrm{CU}(n)$ and $\overline{P}$ will be referred to as a conformal bundle. In order to apply these ideas to gauge theories we review and develop in Sec. 2 the necessary background material on fiber metric and conformal structure Higgs fields on a principal bundle $P(M,G)$. The bundle P may be an arbitrary principal bundle provided only that the group G contain an appropriate copy of ${}^+\mathbb{R}$ as a subgroup.

Suppose now that on a principal bundle $P(M,G)$ one is given a symmetry-breaking Higgs field $\phi:P\to V$ that reduces P to a subbundle $Q(M,G_{\xi_0})$ (for example, ϕ may be a conformal structure Higgs field and Q a conformal subbundle). It is natural to ask if a connection ω on P also reduces to a connection on Q. It is known[6,9] that if ϕ is a symmetry-breaking Higgs field then ω reduces to a subbundle defined by ϕ if and only if $D\phi=0$, where $D\phi$ is the exterior covariant derivative of ϕ with respect to ω. More generally, suppose one is given a symmetry-breaking Higgs field ϕ along with a connection ω on P where the connection does not necessarily reduce to $Q=Q(M,H)=Q^{-1}(\xi_0)$. In this case we consider the question: In what sense and when is a connection naturally induced on Q by ω? Here by "naturally induced" we mean that at a minimum we require the "induced" connection to be the reduced connection when $D\phi=0$.

In language more closely related to physical theories this question can be rephrased as follows. Let e_a, $a=1,2,\dots,n$ be a basis of the n-dimensional Lie algebra $\mathfrak{g}$ of G. Let these vectors be arranged so that the first p of them $e_1,\dots,e_p$ form a basis for the subalgebra $\mathfrak{h}$ of $H\subseteq G$. In a local gauge the connection can hence be written as

$$A=\sum_{a=1}^{p}A_\mu^a e_a\,dx^\mu+\sum_{a=p+1}^{n}B_\mu^a e_a\,dx^\mu, \tag{1.3}$$

where the $A_\mu^a\,dx^\mu$ and $B_\mu^a\,dx^\mu$ are real-valued 1-forms. The question now is how to select from A a piece that will represent a connection on the subbundle $Q(M,H)$. Clearly the part $\Sigma_{a=1}^p A_\mu^a e_a\,dx^\mu$ is an obvious candidate inasmuch as a connection for $Q(M,H)$ must take its values in $\mathfrak{h}$. However, the splitting off of a $\mathfrak{h}$-valued part as in (1.3) is noninvariant in general because of the gauge transformation law

$$\overline{A}=\mathrm{Ad}(g^{-1})A+g^{-1}dg.$$

Note however that if the vectors $e_{p+1},\dots,e_n$ span a complement $\mathfrak{M}$ of $\mathfrak{h}$ in $\mathfrak{g}$ that is invariant by H, that is $\mathrm{Ad}(H)\mathfrak{M}=\mathfrak{M}$, then the splitting off of a $\mathfrak{h}$-valued piece of A will be invariant. There is such a 1-form $\Sigma_{a=1}^p A_\mu^a e_a\,dx^\mu$ at each point of M and the set of all 1-forms that arise in this way from a given connection will then piece together to define a connection on $Q(M,H)$. In general we have the theorem[6] that if there is a complement $\mathfrak{M}$ of $\mathfrak{h}$ in $\mathfrak{g}$ that is Ad-invariant by H, then the restriction of the $\mathfrak{h}$-component of every connection on $P(M,G)$ to $Q(M,H)$ is a connection on $Q(M,H)$.

This theorem and its consequences imply that if a symmetry-breaking Higgs field ϕ reduces $P(M,G)$ to $Q(M,H)$ then, under suitable hypotheses, there is a one-to-one correspondence between connections ω on P and pairs (ω_0,τ_0) where ω_0 is a connection on Q and τ_0 is a tensorial 1-form on Q. In case $D\phi=0$, ω_0 is just the reduction of ω to Q and $\tau_0\equiv0$. In general, however, the connection ω does not reduce to Q and, in this general case, a remnant ω_0 of ω still survives as a connection on Q, while the additional tensorial field τ_0 emerges representing the degrees of freedom lost when one passes from ω to ω_0 on Q. See Sec. 4 for the details of this decomposition of the connection.

To illustrate the role of ω_0 and τ_0 we present in Sec. 5 a bundle version of "generalized" conformal geometry. The conformal geometry is "generalized" in that it is defined on a subbundle of an arbitrary $\mathrm{Gl}(n,K)$ bundle over spacetime and is not restricted to subbundles of the frame bundle LM of the base manifold M. We will see that when the theory is specialized to a bundle version of Weyl geometry the tensorial field τ_0 turns out to be essentially the Weyl vector.[10]

We bring these ideas together in sec. 6 where we present as an application a conformal CU(2) geometrical model of the electroweak theory of Weinberg and Salam. In the model ω_0 is identified with the gauge field A representing the massless photon field while the different components of the tensorial field τ_0 are identified with the massive vector bosons $W^\pm$ and Z.

One might expect that a "$\mathrm{SU}(2)\times\mathrm{U}(1)$-bundle-with-connection" would provide a proper framework for the

Weinberg–Salam theory. This rather obvious description is of course possible but seems to give no new physical or mathematical insights. We have chosen instead to link the Weinberg–Salam theory to the geometry of connections on a principal bundle whose structure group is the conformal group $CU(2) = \mathbb{R}^+ \times U(2)$ associated with the usual metric on a two-dimensional complex vector space $\mathbb{C}^2$. We will see that the geometry of such a bundle-with-connection is entirely parallel to a fiber bundle version of Weyl geometry. A consequence of our geometric description of the Weinberg–Salam model is that it gives geometrical significance to the single surviving component η of the Higgs scalar field (see for example Ref. 11). Indeed we shall show that the Higgs scalar field may be geometrized in that η may be identified as an infinitesimal conformal factor in our model.

II. NOTATION AND GENERAL BACKGROUND

In this section we iterate some of the general facts necessary to the subsequent development and establish the notational conventions to be utilized throughout the paper. We also include a few theorems which are generalizations of theorems about metrical and conformal structures on subbundles of frame bundles. Some of the proofs of these generalizations are modifications of known results. We have relegated all the proofs of this section to Appendix A. The general arena for our results is that of a principal fiber bundle with connection. Generally $P(M,G)$ will denote a principal fiber bundle with bundle space P, base space M, and structure group G. We reserve the symbol π for the projection of P onto M and often do not explicitly remind the reader of its definition. Among other things we have as one our goals to study the interaction between Higgs fields and gauge fields. In the present context gauge fields are realized as the local gauge components of a connection 1-form on P and Higgs fields find expression as tensorial 0-forms on P (see precise definitions below).

Basically we utilize Kobayashi and Nomizu[6] as our standard source for basic facts and notational conventions regarding fiber bundles and connections.

Recall that a map s from an open subset U of M into P such that $\pi(s,(x)) = x$ for every $x \in U$ is called a *local section* of $P(M,G)$. We will refer to s as a *local gauge* in P. If ω is any connection 1-form on P we say that $s^*\omega$ is a *gauge field* on U. Moreover if $\{e_a\}$ is a basis of the Lie algebra $\mathfrak{g}$ of G and $s^*\omega = A^a e_a$ then $A^a = A^a_\mu dx^\mu$ are the *gauge potentials* of ω.

The following well-known theorem may be found in Kobayashi and Nomizu.[6] It will be utilized heavily in Sec. 4. We will refer to this theorem as the *Fundamental Theorem* and we will say that a subbundle $Q(M,H)$ of a principal bundle $P(M,G)$ satisfies the *Fundamental hypothesis* if

(1) there is a subspace $\mathfrak{M}$ of the Lie algebra $\mathfrak{g}$ of G such that $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{M}$ where $\mathfrak{h}$ is the Lie algebra of H and

(2) for every $h \in H$, $\mathrm{Ad}(h)(\mathfrak{M}) = \mathfrak{M}$.

Fundamental Theorem 2.1: Let $P(M,G)$ be a principal bundle and assume that $Q(M,H)$ is a subbundle of $P(M,G)$ which satisfies the Fundamental hypothesis. If ω is a connection on P then $\omega_\mathfrak{h}|TQ$ is a connection on Q and $\omega_\mathfrak{M}|TQ$ is a tensorial 1-form[7] on Q with values in $\mathfrak{M}$ (here, of course, $\omega_\mathfrak{h}$ and $\omega_\mathfrak{M}$ are the components of ω relative to the decomposition $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{M}$).

The subbundle Q to which ω is restricted may arise in many ways. One way is to use the result[6] that a bundle $P(M,G)$ has a reduction to a subgroup $H \subseteq G$ iff the bundle associated to the action of G on G/H has a global section.

Recall that if the structure group G of a principal bundle $P(M,G)$ acts on the left of a manifold F then a fiber bundle $E = E(M,F)$ can be constructed which depends on the action of G on F. This bundle is called the associated bundle. The members of E are equivalence classes of pairs $(u,f) \in P \times F$ where $(u_1,f_1) \sim (u_2,f_2)$ iff $u_2 = u_1 g$ and $f_2 = g^{-1} f$, for some $g \in G$. The equivalence class is denoted by $[u,f]$ for $(u,f) \in P \times F$. It is well known[2,6] that there is a one-to-one correspondence between sections of $E(M,G)$ and maps $\phi: P \to F$ having the property that for $u \in P$, $g \in G$

$$\phi(ug) = g^{-1} \cdot \phi(u).$$

Kobayshi and Nomizu[6] call such a map a tensorial 0-form[7] and Trautman[9] refers to ϕ a a *(generalized) Higgs field*. We will follow Trautman's convention but hasten to point out that to obtain a bundle version of the Higgs' mechanism one must restrict attention to *symmetry-breaking Higgs fields*. We say that a Higgs field ϕ is a *symmetry-breaking Higgs field* if it maps all of P onto a single orbit $G \cdot \xi$ of G on F (the *orbit* $G \cdot \xi$ of G through $\xi \in F$ is the set of all elements $g \cdot \xi$ for $g \in G$). We shall restrict our attention to Higgs fields ϕ having their values in a vector space, thus if we say $\phi: P \to F$ is a Higgs field it is understood that F is a vector space.

It was pointed out in the introduction that one way of modifying general Higgs fields to obtain symmetry-breaking Higgs fields is to introduce a generalized conformal structure. We now briefly lay the framework for a more thorough discussion of such structures.

Let $P(M,G)$ be an arbitrary principal bundle and let G act on the left of a vector space V. Let $\{r_1, r_2, \ldots, r_n\}$ denote a fixed but arbitrary basis of V. We utilize this basis throughout the paper and will refer to it as the *standard basis* of V. In Sec. 5 it will be assumed that G acts linearly and faithfully on V and thus that G may be identified with a subgroup of $\mathrm{Gl}(n,K)$ where $n = \dim V$ (here K is the scalar field of V and may be either $\mathbb{R}$ or $\mathbb{C}$). In particular $g \in G$ is identified with $(g_b{}^a) \in \mathrm{Gl}(n,K)$ where

$$g \cdot r_a = g_a^b r_b.$$

Even though G is a subgroup of $\mathrm{Gl}(n,K)$ it does not follow that P can be identified as a subbundle of the bundle LM of frames of M. Since any set of transition functions of P necessarily has its values in $G \subseteq \mathrm{Gl}(n,K)$ it follows that there is a $\mathrm{Gl}(n,K)$ bundle $\overline{P}$ which contains P as a subbundle. Generally $\overline{P}$ is not isomorphic to the bundle of frames LM, however, as it may not admit a soldering form.[9,12] On the other hand we can show that $\overline{P}$ is isomorphic to the bundle of all frames $(x, \{e_i\})$ where $x \in M$ and where $\{e_i\}$ is a basis of some appropriately chosen vector space E_x over x. More precisely, if E is any vector bundle over M then the set of pairs $(x, \{e_i\})$ with $x \in M$ and $\{e_i\}$ a basis of E_x is a principal fiber bundle over M. We call this bundle the bundle of frames of E and we denote it by $\mathscr{F}E$. In the special case when $E = TM$ we have $\mathscr{F}E = \mathscr{F}TM = LM$.

Theorem 2.2: Let $P(M,G)$ be a principal bundle whose group G acts linearly and faithfully on the left of a vector space V. Then $P(M,G)$ can be identified as a subbundle of the bundle of frames $\mathcal{F}E$ of the vector bundle $E = E(M,G)$ associated with the action of G on V.

Now assume that a fixed metric ρ is given on V and that it has arbitrary signature $m \leqslant n$. We say that a fiber metric γ on E is a *ρ-fiber metric on E* if at each point $x_0 \in M$ there exist local sections $X_1, X_2, \dots, X_n$ of E defined on some neighborhood U of x_0 such that for every $x \in U$

$$\gamma_x(X_i(x), X_j(x)) = \rho(r_i, r_j).$$

This is a minor modification of the definition of a fiber metric as found in Kobayashi and Nomizu[6] but has value when the bundle E is the bundle associated to the action of some relatively small subgroup of $\mathrm{Gl}(n,K)$. For example, if one has an SO(2) subbundle of the bundle of frames LM_0 of Minkowski space M_0 then there exits fiber metrics γ on the bundle E associated to an action of SO(2) on $\mathbb{R}^4$ such that it is impossible to gauge transform the metric γ so that its gauge components are the components of the Minkowski metric η. On the other hand this would not preclude the possibility that in an appropriate gauge the components of γ are precisely $\rho(r_i, r_j)$ for some constant metric ρ on $\mathbb{R}^4$.

If $\mathcal{B}(V)$ is the vector space of all bilinear maps from V into K we define an action of G on $\mathcal{B}(V)$ by

$$(g \cdot b)(v,w) = b(g^{-1}v, g^{-1}w),$$

for $g \in G$, $b \in \mathcal{B}(V)$, and $v,w \in V$. We say that τ is a *metric Higgs field* on P if τ is a Higgs field from P into $\mathcal{B}(V)$ such that $\tau(u)$ is a metric on V for each $u \in P$. We refer to τ as a *ρ-metric Higgs field* on P if τ maps P onto a single G orbit of $\rho \in \mathcal{B}(V)$. The following theorem is a minor modification of a similar well known theorem pertaining to metrics on G-subbundles of LM but is presented here (without proof) for use in Sec. 5.

Theorem 2.3: If $P(M,G)$ is a principal bundle and τ is a ρ-metric Higgs field on P then it induces a unique ρ-fiber metric on the bundle E associated with P and the action of G on $\mathcal{B}(V)$. Moreover there exist ρ-fiber metrics on E which do not arise from ρ-metric Higgs fields on P, but if $G = \mathrm{Gl}(n,K)$ then every ρ-fiber metric is induced by a unique ρ-metric Higgs field.

Remark 1: A proof of Theorem 2.3 utilizes the fact that if $\phi: P \to W$ is an arbitrary vector-valued Higgs field which maps onto a single orbit $G \cdot \xi$ of the action of G on W then

$$Q_\phi = \{u \in P \mid \phi(u) = \xi\}$$

is a subbundle of P with structure group the isotropy subgroup G_ξ of G ($g \in G_\xi$ iff $g \cdot \xi = \xi$). This fact is emphasized by Trautman[9] and will be utilized throughout this paper.

Remark 2: If $\overline{P}$ is the $\mathrm{Gl}(n,K)$ bundle (referred to above) which contains P as a subbundle then it is true that a given ρ-fiber metric on the bundle associated to P and H arises from a ρ-metric Higgs field on a subbundle of $\overline{P}$ which contains P. The group of this bundle is the smallest Lie subgroup of $\mathrm{Gl}(n,K)$ which contains G and the values of all those $\mathrm{Gl}(n,K)$-valued functions which carry ρ-gauges of $\overline{P}$ onto gauges of P.

We proceed to discuss conformal structures on P and E. We say that two bilinear (sesquilinear when $K = \mathbb{C}$) maps b_1, b_2 on V are (*conformally*) *equivalent* if there is a positive constant c such that $b_2 = cb_1$. The set of all maps equivalent to b is denoted by $[b]$ and the set of all $[b]$ for $b \in \mathcal{B}(V)$ is denoted $\mathcal{B}^*(V)$. Since $\mathcal{B}(V)$ is a vector space $\mathcal{B}^*(V)$ is a manifold since it is the set of all rays in a vector space.

Notice that G acts on $\mathcal{B}^*(V)$ via $g \cdot [b] = [g \cdot b]$ for $g \in G$, $b \in \mathcal{B}(V)$. To say that $\tilde{\tau}$ is a *conformal Higgs field* on P means $\tilde{\tau}$ is a smooth Higgs field from P onto $\mathcal{B}^*(V)$ such that for each $u \in P$, $\tilde{\tau}(u)$ is an equivalence class of *metrics* on V. Here when we say that $\tilde{\tau}$ is smooth we mean that each point of P is contained in an open set U on which there is defined a map $\tau: U \to \mathcal{B}(V)$ such that $\tilde{\tau}(u) = [\tau(u)]$ for each $u \in U$. A *ρ-conformal Higgs field* is a conformal Higgs field $\tilde{\tau}$ such that $\tilde{\tau}$ maps P onto a G-orbit of $[\rho]$. The proof of the following remark is easy and is left to the reader.

Remark 3: If $u_0 \in P$ and $\tilde{\tau}$ is a conformal Higgs field on P then there is a neighborhood U of $\pi(u_0)$ in M and a smooth $\mathcal{B}(V)$-valued Higgs field τ on $\pi^{-1}(U) = P \mid U$ such that $\tilde{\tau}(u) = [\tau(u)]$ for every $u \in P \mid U$. We refer to $\pi^{-1}(U)$ as a bundle neighborhood of u_0.

When we say that $\tilde{\gamma}$ is a *conformal structure on E* we mean that $\tilde{\gamma}$ is a map on M such that $\tilde{\gamma}_x$ is a conformal equivalence class of metrics on E_x. It is required that $\tilde{\gamma}$ be smooth. Thus each point of M is contained in *some* open subset U of M on which there is defined a fiber metric γ of $E \mid U$ such that $\tilde{\gamma}_x = [\gamma_x]$ for each $x \in U$.

The conformal structure $\tilde{\gamma}$ is a *ρ-conformal structure* if each point of M is contained in the domain U of a ρ-fiber metric γ on $E \mid U$ such that $\tilde{\gamma}_x = [\gamma_x]$ for each $x \in U$.

Theorem 2.4: If $P(M,G)$ is a principal bundle and $\tilde{\tau}: P \to \mathcal{B}^*(V)$ is a ρ-conformal Higgs field then $\tilde{\tau}$ induces a unique ρ-conformal structure on the bundle E associated to the usual action of G on V. In general there exist ρ-conformal structures on E which do not arise from a ρ-conformal Higgs field but in case $G = \mathrm{Gl}(n,K)$ every ρ-conformal structure on E is induced by a unique such field.

We now introduce the symmetry groups associated with the conformal structures discussed above. Let $\mathrm{U}(\rho)$ denote the set of all $g \in \mathrm{Gl}(n,K)$ such that $\rho(g \cdot v, g \cdot w) = \rho(v,w)$ for all $v,w \in V$ and let $\mathrm{CU}(\rho)$ denote the set of $g \in \mathrm{Gl}(n,K)$ such that for some $c > 0$ $\rho(g \cdot v, g \cdot w) = c\rho(v,w)$ for all $v,w \in V$. Observe that since ρ has arbitrary signature, $U(\rho)$ is not the usual unitary group although we sometimes refer to it in this way. Indeed, generally $U(\rho)$ is not compact. Also observe that $g \in \mathrm{CU}(\rho)$ iff $g = ch$ for some $c > 0$ and $h \in \mathrm{U}(\rho)$. Thus $\mathrm{CU}(\rho) \cong \mathrm{U}(\rho) \times {}^+\mathbb{R}$. Moreover there exists a "standard" basis $\{r_1, r_2, \dots, r_n\}$ of V such that the matrix of ρ, $J_{ij} \equiv \rho(r_i, r_j)$, is diagonal with only elements of $\{1, -1\}$ on its principal diagonal. We shall assume that the basis has been chosen so that this is so. Observe that $A \in \mathrm{U}(\rho)$ iff

$$AJA^\dagger = J.$$

Here $A^\dagger$ denotes the transpose of A when $K = \mathbb{R}$ and is the conjugate transpose of A when $K = \mathbb{C}$.

Remark 1: Observe that just as for the metric Higgs field, one has that a bundle $P(M,G)$ reduces to a $G \cap \mathrm{CU}(\rho)$ subbundle iff P admits a ρ-conformal Higgs field. On the other hand bundle reduction is generally not equivalent to the existence of a ρ-conformal structure on E. It is clear that the existence of a bundle reduction does always imply the existence of a ρ-conformal structure on E.

Remark 2: Generally one should not expect the existence of a ρ-conformal Higgs field on a principal bundle $P(M,G)$. For example it is well known[13] that in order for a Lorentzian metric to exist on a compact manifold M the Euler class of the tangent bundle of M must vanish. It follows from our next theorem that no η-conformal structure can then exist on a compact manifold whose tangent bundle has a nonvanishing Euler class.

Theorem 2.5: If $P(M,G)$ is a principal bundle and $G \subseteq \mathrm{Gl}(n,K)$ contains all positive multiples of the identity of $\mathrm{Gl}(n,K)$ then P admits a ρ-conformal Higgs field iff it admits a ρ-metric Higgs field.

Corollary 2.6: The bundle E associated to P has a ρ-conformal structure if and only if it has a ρ-fiber metric.

This concludes our effort to recast various known facts into the framework needed in the subsequent development.

III. SYMMETRY-BREAKING HIGGS FIELDS

As remarked in the introduction, symmetry-breaking Higgs fields do not exist in general. There are typically topological obstructions to their existence and we refer to the reader to the paper by Isham[14] for a detailed account of these obstructions. In this section we characterize those Higgs fields that are symmetry-breaking and we use this result to show that general Higgs fields are actually symmetry-breaking over submanifolds of the base space. In addition we prove a theorem which shows that under suitable hypotheses any two Higgs fields that map to the same orbit in a vector space are the same Higgs field up to a bundle automorphism.

As we stated in the introduction *local* symmetry-breaking Higgs fields are easily constructed. If $s:U\to P$ is any local gauge in P we define $\phi_U:(P\,|\,U)\to V$ as follows. Let $\xi\in V$ and choose any smooth function $\Omega: U\to G$. Let ϕ_U be defined by

$$\phi_U((x)g) = g^{-1}\Omega(x)\xi.$$

Then ϕ_U is a symmetry-breaking Higgs field on $P\,|\,U$. If we define $\bar{s}$ by $\bar{s}(x) = s(x)\Omega(x)$ we see that $\bar{s}$ is a new gauge in P and on this gauge $\phi_U(\bar{s}(x)) = \xi$ for all x. The question arises: how can one use local symmetry-breaking Higgs fields to obtain global ones? We can choose an open cover $\{U_\alpha\}$ of M such that for each α, U_α is the domain of a local gauge $s_\alpha: U_\alpha\to P$ of P. Also, as above we can define a symmetry-breaking Higgs field $\phi_\alpha:(P\,|\,U_\alpha)\to V$ which has some constant value $\xi_\alpha\in V$ on s_α. It is now trivial to show that there is a global Higgs field ϕ on P such that

$$\phi\,|(P\,|\,U_\alpha) = \phi_\alpha,$$

iff

$$\xi_\beta = g_{\alpha\beta}(x)^{-1}\xi_\alpha$$

for all $x\in U_\alpha\cap u_\beta$ where $g_{\alpha\beta}: U_\alpha\cap U_\beta\to G$ is the gauge transformation which carries the gauge s_α to the gauge s_β:

$$s_\beta(x) = s_\alpha(x)g_{\alpha\beta}(x).$$

Minor modification of this argument yields a proof of the following theorem.

Theorem 3.1: A Higgs field ϕ defined on a principal bundle $P(M,G)$ is a symmetry-breaking Higgs field if and only if there is a family $\{(U_\alpha,s_\alpha)\}$ of local gauges of P such that M is covered by the domains $\{U_\alpha\}$ of the gauges and ϕ is constant on each gauge s_α.

Since the existence of general Higgs fields poses no difficulty and since the existence of global symmetry-breaking Higgs fields imposes topological restrictions one could ask whether or not general Higgs fields admit a "decomposition" in terms of symmetry-breaking ones. We will see that under relatively general hypotheses it is possible to write M as

$$M = U_1\cup U_2\cup\cdots\cup U_s\cup\Sigma,$$

where Σ is a set of measure zero and where each U_i admits a foliation $\mathscr{U}_i$ such that for every leaf L of $\mathscr{U}_i$ the restriction of ϕ to $P\,|L$ is a symmetry-breaking Higgs field. We proceed to prove this theorem but we first establish some details necessary to the decomposition of M.

Generally the action of G on a vector space V does not foliate V. This is basically due to the singular nature of the mapping from $G\times V$ into V defined by $(g,x)\to g\cdot x$. This type of singularity is fairly well understood however,[15] and utilizing standard techniques we can decompose V as follows.

Let n_1 denote the dimension of an orbit of G in V of maximal dimension and define V_1 by

$$V_1 = \cup\{\mathscr{O}\,|\,\mathscr{O} \text{ is an orbit of } G \text{ such that } \dim\mathscr{O} = n_1\}.$$

One can now define a family of subsets $\{V_i\}$ of V and integers $\{n_i\}$ inductively so that the following conditions are satisfied:

(1) $n_1 > n_2 > \cdots > n_s$ and for each $1\leqslant i\leqslant s$, n_i is the largest integer which is the dimension of an orbit of G not lying in $\overline{V}_1\cup\overline{V}_2\cup\cdots\cup\overline{V}_{i-1}$ (for each i it can be shown that if an orbit intersects $\overline{V}_1\cup\cdots\cup\overline{V}_{i-1}$ then it is actually a subset of $\overline{V}_1\cup\cdots\cup\overline{V}_{i-1}$).

(2) for $1\leqslant i\leqslant s$, V_i is the union of the family of orbits of G in $V - \cup_{j=1}^{i-1}\overline{V}_j$ such that all the orbits of the family have dimension n_i.

(3) $V = \bigcup_{i=1}^{s} \overline{V}_i$.

One may then prove that each V_i is open and that $\Sigma = V - \cup_{i=1}^{s} V_i$ has measure zero in V. We call this decomposition the *singular decomposition* of V and we refer to Σ as the residual set of the decomposition. We delete the details of the proof that such a decomposition exists preferring to refer to Ref. 15 where similar decompositions are made.

It is perhaps worthwhile to compare this method of decomposing V to another method introduced by Michel and Radicati.[16] If $x\in V$ then the *stratum* of V through $x\in V$ is the set $S(x)$ of all $y\in V$ such that the isotropy subgroup of G determined by y is a conjugate of the isotropy subgroup G_x determined by x. Two points on the same orbit have conjugate isotropy subgroups, thus each stratum is the union of a family of orbits. Any two orbits in a stratum have the same dimension, thus for $x\in V_i$ one has $S(x)\subseteq V_i\cup\Sigma$. It follows that the set of interior points of a stratum is a subset of a single V_i but the boundary points may lie in the residual set Σ. Generally the decomposition of V into strata is a finer decomposition than the singular decomposition (see Appendix B for a typical example). We utilize the singular decomposition because we are interested in finding the largest bundles of the form $P|U$ on which a given Higgs field may be decomposed into symmetry-breaking Higgs fields over submanifolds of U.

Lemma 3.2: If G is a Lie group which acts on the left of a vector space V and if $V = \Sigma \cup V_1 \cup \cdots \cup V_s$ is the singular decomposition of V relative to this action then for each $1 \leqslant i \leqslant s$ there is a foliation $\mathscr{V}_i$ of V_i whose leaves are precisely the set of components of the orbits of G in V_i.

We relegate the proof to Appendix A.

Recall[17] that the action of a Lie group G on a manifold F is called a *regular action* if each point $x \in F$ is contained in an open subset $U_x \subseteq F$ such that for every orbit $\mathscr{O}$ of G in F, $\mathscr{O} \cap U_x$ is a connected (regular) submanifold of U_x. For example, if G is compact then every action of G is regular. Also the usual action of the Lorentz group on $\mathbb{R}^4$ is regular.

Theorem 3.3: Assume that $P(M,G)$ is a principal bundle, that G acts regularly on V and that $V = \Sigma \cup V_1 \cup \cdots \cup V_s$ is the singular decomposition of V relative to the action of G. Let $\phi: P \rightarrow V$ be a Higgs field and let $U_i = \pi(\phi^{-1}(V_i))$, $\Sigma_M = \pi(\phi^{-1}(\Sigma))$. Then

(1) $M = \Sigma_M \cup U_1 \cup \cdots \cup U_s$ where each U_i is open in M and Σ_M is residual in M,

(2) the Higgs field ϕ maps the bundle $P|U_i$ onto the open submanifold V_i of V and V_i is foliated by orbits of G,

(3) there exists a unique smooth mapping f_i from U_i into the possibly non-Hausdorff manifold V_i/G such that if $\eta_i : V_i \rightarrow V_i/G$ is the map which sends x to the oribt of x then

$$\eta_i \circ \phi = f_i \circ \pi.$$

Moreover, if the map f_i has constant rank then there is a foliation $\mathscr{U}_i$ of U_i such that if L is any leaf of $\mathscr{U}_i$ then ϕ maps $P|L$ onto a single orbit of G; consequently $\phi|(P|L)$ is a symmetry-breaking Higgs field.

Again we defer the proof to Appendix A.

Remark 1: Let $P(M,G)$ be a principal bundle whose structure group G is compact and assume that G acts on the left of a vector space V. If ϕ is a constant rank Higgs field which maps P into the interior of a single stratum S of the action of G on V then S admits a foliation $\mathscr{S}$ and by our theorem it follows that M admits a foliation $\mathscr{M}$ such that $\phi|(P|L)$ is symmetry-breaking for each leaf L of $\mathscr{M}$.

Remark 2: Theorem 3.3 above shows that it is possible to decompose M into the union of a finite number of open sets along with a residual set of measure zero such that each open set is foliated in such a way that ϕ is symmetry-breaking in the part of the bundle over each leaf. The basic assumption which allows us to do this is that each f_i has constant rank. This is only an assumption of convenience since if this were not true it is known that under fairly general hypotheses each U_i can be decomposed further so that on each piece of the decomposition f_i would have constant rank. Thus U_i would decompose into a union of a finite number of open sets along with a measure zero residual set in such a way that the rank of f_i is constant on each open set. Thus the theorem holds quite generally.

We have discussed the problems of existence of symmetry-breaking Higgs fields and their relationship to general Higgs fields. Our next theorem provides some insight regarding the uniqueness of such fields.

Theorem 3.4: Let $P(M,G)$ be a principal bundle and let G act on the left of a vector space V. Assume that ϕ and ψ are symmetry-breaking Higgs fields which map onto the same orbit $G \cdot \xi$ of G. If there is a normal subgroup N of G such that G is a semidirect product $G = N \circledS G_\xi$ of N and the isotropy subgroup G_ξ of ξ then there is an equivariant automorphism $\gamma: P \rightarrow P$ such that $\psi \circ \gamma = \phi$.

Again we defer the proof to Appendix A. This theorem will be utilized in our discussion of conformal geometry in Sec. 5. It also may be applied successfully to Higgs fields defined on the affine frame bundle AM which has been shown[12] to be the bundle relevant to a fiber bundle description of the metric affine theories of gravitation due to Hehl and others.[18]

IV. SPLITTING THE CONNECTION

It is well known that having a symmetry-breaking Higgs field on a principal bundle $P(M,G)$ is equivalent to the possibility of reducing the bundle P to a subbundle $Q(M,H)$ whose structure group H is a subgroup of G. In much of the physics literature on bundles any connection ω on the bundle P is either required to reduce to Q or the possibility of having a connection induced on Q by ω is ignored. On the other hand it is certainly well known that in spontaneously broken gauge theories certain components of the original gauge field survive as the components of a new gauge field and of course this new gauge field should be a connection on an appropriate subbundle of the original bundle. The Fundamental theorem gives us a method for dealing with this problem as it allows us to split a connection ω on P into a pair (ω_0, τ_0) where ω_0 is a new connection on Q and τ_0 is a tensorial form on Q provided that the Fundamental hypothesis is satisfied. In this section we are interested in some special implications of this result.

Let G be a Lie group and H a subgroup of G. We say that N is *a complement* of H in G provided N is a normal subgroup of G such that $G = N \circledS H$ is a semi-direct product of N and H.

Theorem 4.1: Assume that $P(M,G)$ is a principal bundle, that $Q(M,H)$ is a subbundle of $P(M,G)$ and that N is a complement of H in G. Then there is a bijective correspondence between the set of all connections on P and the set of all ordered pairs (ω_0, τ_0) where ω_0 is a connection on Q and τ_0 is an $\mathfrak{n}$-valued tensorial 1-form on Q.

Proof: The correspondence referred to in Theorem 4.1 is obtained as follows: given a connection ω on P the corresponding pair (ω_0, τ_0) is defined by

$$\omega_0 = \omega_{\mathfrak{h}} | TQ,$$

$$\tau_0 = \omega_{\mathfrak{n}} | TQ.$$

Conversely if (ω_0, τ_0) is given then the connection ω which gives (ω_0, τ_0) under this correspondence can be recovered via the formula

$$\omega(X) = \omega_0(d\bar{\pi}(X)) + \tau_0(d\bar{\pi}(X)) - d\mu(X - d\bar{\pi}(X)), \quad (4.1)$$

where X is any vector tangent to P at points of Q and where the property $R_g^* \omega = \mathrm{Ad}(g^{-1})\omega$ will recover its values at other points of P. Here μ is the map from P to N such that for each $u \in P$, $\mu(u)$ is the unique element of N such that $u\mu(u)$ is in Q. Moreover $\tilde{\pi}: P \to Q$ is defined by $\tilde{\pi}(u) = u\mu(u)$. The full details of the proof are similar to the corresponding theorem about connections on the affine frame bundle AM (see Ref. 6). ■

Generally if ω is a connection on a principal bundle P and Q is a subbundle of P then ω does not reduce to Q. Under the Fundamental hypothesis, however, we do know that there is a unique connection ω' on P such that

(1) ω' reduces to Q,

(2) ω' agrees with $\omega_{\mathfrak{h}} | TQ$ on Q.

It is clear that ω reduces to Q iff $\omega = \omega'$ and that this is true iff the tensorial form $\omega_{\mathfrak{n}} | TQ$ vanishes.

Theorem 4.2: Let $P(M,G)$ be a principal bundle and $Q(M,H)$ a subbundle which satisfies the Fundamental hypothesis. Let ω be a connection of P and let ω' be the unique connection on P which reduces to $\omega_{\mathfrak{h}} | TQ$ on Q.

(1) If $\tau = \omega - \omega'$ is the difference form then τ is uniquely determined by $\tau | TQ$. Moreover $\tau_{\mathfrak{h}} | TQ$ is a difference form on Q and ω reduces to Q iff $\tau_{\mathfrak{h}} | TQ = 0$.

(2) If G acts on a vector space V and $\phi: P \to V$ is a Higgs field which breaks the symmetry of P to Q then

$$D\phi = (\omega - \omega')\cdot\phi. \tag{4.2}$$

Moreover, $\tau_{\mathfrak{M}} | TQ$ is uniquely determined by $D\phi$.

Remark: In writing equation (4.2) we tacitly assumed the existence of an action of the Lie algebra $\mathfrak{g}$ on V. In fact if $A \in \mathfrak{g}$ and $v \in V$, then $A \cdot v$ is defined by

$$A\cdot v = \frac{d}{dt}\left[\exp(tA)\cdot v\right]\bigg|_{t=0},$$

where we identify $T_v V$ with V for each $v \in V$.

Proof of the Theorem: The proof of (1) is straightforward and is left to the reader. Actually (2) is also easy as generally we have $D\phi = d\phi + \omega\cdot\phi$ and $D'\phi = d\phi + \omega'\cdot\phi$. But ω' reduces to Q so that $D'\phi = 0$ and consequently $D\phi = (\omega - \omega')\cdot\phi$. This proves the first assertion of (2). Since $\phi: P \to V$ is a symmetry-breaking Higgs field we know that for some $\xi \in V$, $\phi^{-1}(\xi) = Q$ and $H = G_\xi$ is the isotropy subgroup of ξ. Thus if A is in the Lie algebra $\mathfrak{g}$ of G then A belongs to the subalgebra $\mathfrak{h}$ iff $A\cdot\xi = 0$. Thus $\tau_{\mathfrak{h}}\cdot\xi = 0$. But $D\phi = \tau\cdot\phi$ and consequently for $u \in Q$ and $X \in T_u Q$,

$$D_u\phi(X) = \tau_u(X)\cdot\phi(u) = \tau_u(X)\cdot\xi = (\tau_{\mathfrak{M}})_u\cdot\xi.$$

Note, however, that if $A,B \in \mathfrak{M}$ and $A\cdot\xi = B\cdot\xi$ then $A - B \in \mathfrak{h} \cap \mathfrak{M} = (O)$. Thus there is at most one $C \in \mathfrak{M}$ such that $D_u\phi(X) = C\cdot\xi$. Consequently $\tau_{\mathfrak{M}} | TQ$ is uniquely determined by $D\phi | TQ$. ■

Corollary: Assume that $P(M,G)$ is a principal bundle and that $Q(M,H)$ is a subbundle of P which satisfies the Fundamental hypothesis and which arises from a symmetry-breaking Higgs field $\phi: P \to V$. Then every connection ω on P arises as follows:

(1) choose an arbitrary connection ω_0 on Q,

(2) choose an arbitrary $\mathfrak{M}$-valued horizontal equivariant 1-form τ_0 on Q,

(3) require that $\omega_{\mathfrak{h}} | TQ = \omega_0$ and that $D\phi | TQ = \tau_0$.

Remark: We will see later in the paper that the components of τ play the role of physical fields. In the case that P is the conformal bundle ϕ may be taken to be real-valued and the form $(\tau_{\mathbf{R}} | TQ) = (\omega_{\mathbf{R}} | TQ)$ uniquely determines the Weyl vector usually associated with the electromagnetic field in Weyl geometry. We will also see later how to formulate the traditional Weinberg–Salam model of the electroweak interaction in geometrical language and upon doing so we will find that the gauge components of $(\tau_{\mathfrak{M}} | TQ) = (\omega_{\mathfrak{M}} | TQ)$ are precisely the fields $W^{\pm}, Z$. In both of these examples we have asserted that $(\tau_{\mathfrak{M}} | TQ) = (\omega_{\mathfrak{M}} | TQ)$. This is always true in the presence of our Fundamental hypothesis since on $Q = \phi^{-1}(\xi)$ we have $\mathfrak{h}\cdot\xi = 0$ and $\tau_{\mathfrak{M}}\cdot\xi = (\tau_{\mathfrak{h}} + \tau_{\mathfrak{M}})\cdot\xi = \tau\cdot\xi = (\omega - \omega')\cdot\xi = (\omega_{\mathfrak{h}} + \omega_{\mathfrak{M}})\cdot\xi - \omega'\cdot\xi = \omega_{\mathfrak{M}}\cdot\xi$ and on TQ this equation implies $\tau_{\mathfrak{M}} | TQ = \omega_{\mathfrak{M}} | TQ$.

We now develop some notation to be utilized in the final theorem of this section. Let $P(M,G)$ denote a principal fiber bundle and assume that G acts linearly on the left of a vector space V. Let ω be a connection on P and let ξ denote some fixed element of V. Assume that G_ξ has some fixed complement N in G and let $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{n}$ where $\mathfrak{h}$ is the Lie algebra of $H = G_\xi$. For every Higgs field $\phi: P \to V$ which maps onto the orbit $G\cdot\xi$ let ω_ϕ denote the connection on P which reduces to $\omega_{\mathfrak{h}} | TQ$ on Q and let τ_ϕ denote the $\mathfrak{n}$-valued tensorial difference form $\omega - \omega_\phi$ on P. It is of interest to know how a change in the Higgs field ϕ affects the pair (ω_ϕ, τ_ϕ) determined by ϕ and the connection ω. In particular assume that $\psi: P \to V$ is another Higgs field which maps onto the same orbit $G\cdot\xi$. Then ω_ϕ and ω_ψ agree with ω but on different subbundles of P. They are different connections but we shall see that in local gauges they transform as if their gauge components were the gauge components of a single connection.

Theorem 4.3: Assume that ϕ and ψ are Higgs fields from a principal bundle $P(M,G)$ into a vector space V such that both ψ and ϕ map into the same orbit $G\cdot\xi$ for $\xi \in V$. Assume that the isotropy subgroup G_ξ of ξ has a complement N in G and that ω is a connection on P. Then there exist local gauges s_ϕ in Q_ϕ and s_ψ in Q_ψ and a mapping n from dom $s_\phi \cap$ dom s_ψ into N such that

(1) $s_\psi{}^*\omega_\psi = s_\phi{}^*\omega_\phi = \mathrm{Ad}(n^{-1})(s_\psi^*(\omega_\phi)) + dL_n{}^{-1}\circ dn$

(2) $s_\psi{}^*\tau_\psi = s_\psi{}^*\tau_\phi - dR_n{}^{-1}\circ dn$.

Proof: It is easy to show that there exist gauges s_ϕ in Q_ϕ and s_ψ in Q_ψ such that $s_\psi(x) = s_\phi(x)n(x)^{-1}$ for some N-valued function on dom $s_\phi \cap$ dom s_ψ. If $H = G_\xi$ has Lie algebra $\mathfrak{h}$ and N has Lie algebra $\mathfrak{n}$ it follows from the definitions of ω_ϕ and ω_ψ that

$$s_\phi{}^*\omega_\phi = s_\phi{}^*\omega_{\mathfrak{h}},$$

$$s_\psi{}^*\omega_\psi = s_\psi{}^*\omega_{\mathfrak{h}}.$$

Moreover, since ω is a connection we have

$$s_\psi{}^*\omega = \mathrm{Ad}(n)(s_\phi{}^*\omega) + dL_n\circ dn^{-1}. \tag{4.3}$$

Let $\mu_\phi: P \to N$, $\mu_\psi: P \to N$ be the unique functions defined by requiring that $u\mu_\phi(u) \in Q_\phi$ and $u\mu_\psi(u) \in Q_\psi$ for each $u \in P$. Also let $\pi_\phi: P \to Q_\phi$, $\pi_\psi: P \to Q_\psi$ be defined by $\pi_\phi(u) = u\mu_\phi(u)$, $\pi_\psi(u) = u\mu_\psi(u)$. Since s_ϕ is a gauge in Q_ϕ we have that $d_x s_\phi(X)$ is tangent to Q_ϕ for each $x \in M$ and $X \in T_x M$. But $\pi_\phi | Q_\phi$ is the identity; thus $d\pi_\phi(d_x s_\phi(X)) = d_x s_\phi(X)$. It follows from these facts and Eq. (4.1) of the proof of Theorem 4.1 that

$$(s_\phi{}^*\omega)(X) = \omega_\phi(ds_\phi(X)) + \tau_\phi(ds_\phi(X))$$
$$= (s_\phi{}^*\omega_\phi + s_\phi{}^*\tau_\phi)(X)$$

with a similar equation involving ω_ψ and τ_ψ. Substituting these equations in (4.3) above we obtain

$$s_\psi{}^*\omega_\psi + s_\psi{}^*\tau_\psi = \mathrm{Ad}(n)(s_\phi{}^*\omega_\phi) + \mathrm{Ad}(n)(s_\phi{}^*\tau_\phi) + dL_n \circ dn^{-1}. \quad (4.4)$$

Each term in Eq. (4.4) can easily be identified as having all of its values either in $\mathfrak{n}$ or in $\mathfrak{h}$ except for $\mathrm{Ad}(n)\,(s_\phi{}^*\omega_\phi)$. We will show that $\mathrm{Ad}(n)\,(s_\phi{}^*\omega_\phi) = s_\phi{}^*\omega_\phi$. To see this recall that $\mathrm{Ad}(n) = d_e\sigma_n$ where $\sigma_n(x) = nxn^{-1}$ for each $x\in G$. Let $\pi: N\times H \to H$ be the map $\pi(n,h) = h$ and observe that if G is identified with $N\times H$ then $\pi(\sigma_n(n_1,h_1)) = h_1$ for arbitrary $n, n_1 \in N$ and $h_1\in H$. It follows that $d\pi\circ d\sigma_n = d\,(id_H)$ and thus that

$$(d_e\sigma_n)(v,w) = w$$

for arbitrary $v\in T_eN$, $w\in T_eH$. It follows that $\mathrm{Ad}(n)(b) = b$ for every $b\in\mathfrak{h}$ and that $\mathrm{Ad}(n)\,(s_\phi{}^*\omega_\phi) = s_\phi{}^*\omega_\phi$. By (4.4) we have

$$s_\psi{}^*\omega_\psi = s_\phi{}^*\omega_\phi = \mathrm{Ad}(n^{-1})(s_\psi{}^*\omega_\phi) + (dL_n{}^{-1}\circ dn)$$
$$s_\psi{}^*\tau_\psi = \mathrm{Ad}(n)(s_\phi{}^*\tau_\phi) + dL_n \circ dn^{-1}.$$

Now τ_ϕ is by definition the difference form $\omega - \omega_\phi$ and consequently is tensorial. Thus $\mathrm{Ad}(n)\,(s_\phi{}^*\tau_\phi) = s_\psi{}^*\tau_\psi$ and the last equation above becomes

$$s_\psi{}^*\tau_\psi = s_\psi{}^*\tau_\phi - dR_n{}^{-1}\circ dn.$$

V. GENERALIZED CONFORMAL GEOMETRIES

In this section we are concerned primarily with the "interaction" between ρ-conformal Higgs fields on a bundle and a connection on a bundle. In particular we formulate a version of Weyl theory on a very general class of principal fiber bundles. Indeed, throughout this section $P(M,G)$ will denote a principal bundle whose group G acts linearly and faithfully on an n-dimensional vector space V in such a way that the set of all positive multiples of the identity of $\mathrm{Gl}(n,K)$ is a subset of G (see Sec. 2 for notation). We also assume V has a metric ρ and since ${}^+\mathbb{R}\subseteq G$ acts on ρ via $s\cdot\rho = s^{-2}\rho$ it follows that the subalgebra $\mathbb{R}$ of the Lie algebra $\mathfrak{g}$ of G acts on ρ by $t\cdot\rho = (-2t)\rho$.

We now prove a preliminary result which does not depend on the interplay between connections and conformal structures but which gives us an easy method for dealing with conformal scaling techniques to be utilized in Sec. 6. We first develop further notation.

Observe that every element g of G, regarded as a matrix in $\mathrm{Gl}(n,K)$, has a unique decomposition $g = ca$ where $c>0$ and $|\det a| = 1$. Indeed the equation $g = ca$ implies $\det g = c^n(\det a)$ and thus $c = |\det g|^{1/n}$. On the other hand if we *define* c by $c = |\det g|^{1/n}$ and a by $a = c^{-1}g$ then it follows that $g = ca$ where $c>0$ and $|\det a| = 1$. Let H denote the set of all $a\in G$ such that $|\det a| = 1$. Clearly H is a subgroup of G and G is isomorphic to ${}^+\mathbb{R}\times H$ via the map $g\to(|\det(g)|^{1/n}, |\det(g)|^{-1/n}g)$. The natural action of G on its coset space ${}^+\mathbb{R} = G/H$ may be identified as

$$g\cdot r = r|\det(g)|^{1/n} \quad \text{for } r>0,\ g\in G. \quad (5.1)$$

Now if τ is a ρ-metric Higgs field on $P_{\tilde\tau}$ belonging to the fixed ρ-conformal Higgs field $\tilde\tau$ then, for each $u\in P$ there exists a unique $r(u)>0$ such that $\tau(u) = \phi(u)\rho$. Define $\phi(u) = r(u)^{-1/2}$ so that $\tau(u) = \phi(u)\cdot\rho$ for each u. For $a\in H\cap U(\rho)$ and $u\in P_{\tilde\tau}$, $\phi(ua)\cdot\rho = \tau(ua) = a^{-1}\cdot\tau(u) = \phi(u)\cdot(a^{-1}\cdot\rho) = \phi(u)\cdot\rho$ and $\phi(ua) = \phi(u)$. Thus ϕ is invariant under action of elements of $H\cap U(\rho)$. On the other hand, for $r>0$ and $u\in P_{\tilde\tau}$, $\phi(ur)\cdot\rho = \tau(ur) = r^{-1}\cdot\tau(u) = (r^{-1}\phi(u))\cdot\rho$ and $\phi(ur) = r^{-1}\phi(u) = r^{-1}\cdot\phi(u)$.

Thus ϕ is a Higgs field if ${}^+\mathbb{R}$ is acted on by elements of G as in (5.1) above. Conversely, if $\phi: G\to{}^+\mathbb{R}$ is a Higgs field and if we define $\tau(u) = \phi(u)^{-2}\rho$ it is easy to reverse the steps above and show that τ is a ρ-metric Higgs field on $P_{\tilde\tau}$. Thus we have:

Theorem 5.1: Let the structure group G of the principal bundle $P(M,G)$ contain all positive multiples of the identity. If $\tilde\tau$ is a ρ-conformal Higgs field on P then there is a bijection between the set of all ρ-metric Higgs field on $P_{\tilde\tau}$ which belong to $\tilde\tau$ and the set of all ${}^+\mathbb{R}$-valued Higgs fields on $P_{\tilde\tau}$. The ρ-metric Higgs field τ and the ${}^+\mathbb{R}$-valued Higgs field ϕ correspond to one another under the bijection if and only if

$$\tau(u) = \phi(u)\cdot\rho = \phi(u)^{-2}\rho.$$

At this point we introduce a connection ω on our bundle P. We have already seen in Sec. 2 that having a ρ-conformal structure $\tilde\tau: P\to\mathscr{B}^*(V)$ on P is equivalent to having a $\mathrm{CU}(\rho)\cap G$ subbundle $\tilde\tau^{-1}([\rho])$ of P. It is of interest to know when the connection ω reduces to this subbundle. It is known that ω reduces to $P_{\tilde\tau} := \tilde\tau^{-1}([\rho])$ iff $D\tilde\tau = 0$. Our next theorem shows that generally this provides us with a 1-form μ on P which will imply the existence of a "Weyl vector" for this theory.

Theorem 5.2: Let $\tilde\tau$ be a ρ-conformal Higgs field on P and let $P_{\tilde\tau}$ denote the $G\cap\mathrm{CU}(\rho)$ subbundle $\tilde\tau^{-1}([\rho])$ of P. If ω is a connection on P then ω reduces to $P_{\tilde\tau}$ iff $D\tilde\tau = 0$. Moreover if τ is a metric Higgs field such that $\tilde\tau(u) = [\tau(u)]$ for all $u\in P$ then ω reduces to $P_{\tilde\tau}$ iff

$$D\tau = \mu\tau$$

for some 1-form μ on P.

Proof: The proof of the first statement is well known[6,9] and consequently is omitted. To prove the second statement first let $\eta: \mathscr{B}(V)\to\mathscr{B}^*(V)$ be defined by $\eta(b) = [b]$. Recall that as a manifold $\mathscr{B}^*(V)$ is diffeomorphic with the unit sphere in $\mathscr{B}(V)$ with antipodal points identified. Thus if we wish to compute $d_b\eta$ at $b\in\mathscr{B}(V)$ then we may choose a coordinate patch at $\eta(b)\in\mathscr{B}^*(V)$ so that in these coordinates η is given by $\bar\eta(b) = (1/\|b\|)b$ where $\|b\|$ is the length of $b\in\mathscr{B}(V)$ relative to some fixed positive definite inner product on the finite-dimensional space $\mathscr{B}(V)$. But $(d_b\eta)(k) = (1/\|b\|)k - (b\cdot k/\|b\|^3)b$ and consequently $d_b\eta(k) = 0$ iff k is a multiple of b. Here $d_b\bar\eta$ denotes the usual Frêchet derivative of the map $\bar\eta$ from the finite dimensional vector space $\mathscr{B}(V)$ to itself. Note, however, that if $\tilde\tau(u) = [\tau(u)]$ then $\tilde\tau = \eta\circ\tau$ and

$$D\tilde\tau = d\eta\circ D\tau.$$

Thus if $X\in T_uP$ then $(D\tilde\tau_u)(X) = 0$ iff $(D\tau_u)(X)$ is a multiple of $\tau(u)$. But if $D\tau_u(X) = \mu(X)\tau(u)$ for some number $\mu(X)$ then

$$\mu(X) = \frac{D_u\tau(X)\cdot\tau(u)}{\tau(u)\cdot\tau(u)}$$

and conversely. Thus ω reduces to $P_{\tilde{\tau}}$ iff $D\tau = \mu\tau$ for some 1-form μ on P. ■

Remark: If there is a 1-form μ on P such that $D\tau = \mu\tau$ then it follows from the fact that both $D\tau$ and τ are tensorial forms[7] that μ is invariant under the action of G. Since μ is also necessarily horizontal it follows that $\mu = \pi^*\lambda$ for some 1-form λ on M.

Theorem 5.3: Let $\tau: P \rightarrow \mathscr{B}(V)$ be a ρ-metric Higgs field and ω a connection on P such that $D\tau = \mu\tau$ for some real 1-form μ. The following statements are equivalent:

(1) there is a smooth function $\mu: P \rightarrow \mathbb{R}$ such that $\mu = d\nu$ (thus μ is exact),

(2) there is a function $\nu: P \rightarrow \mathbb{R}$ such that $D(e^{-\nu}\tau) = 0$

(3) there is a function $\nu: P \rightarrow \mathbb{R}$ such that if $\hat{\tau}(u) = e^{-\nu(u)}\tau(u)$ then ω reduces to $\hat{\tau}^{-1}(r\cdot\rho)$ for each $r > 0$.

Proof: Note that for any smooth function $\nu: P \rightarrow \mathbb{R}$

$$\begin{aligned} D(e^{-\nu}\tau) &= d(e^{-\nu}\tau) + \omega\cdot(e^{-\nu}\tau) \\ &= e^{-\nu}[(D\tau - (d\nu)\tau]. \end{aligned}$$

Thus $D(e^{-\nu}\tau) = 0$ if and only if $D\tau = (d\nu)\tau$. Since there is at most one 1-form μ such that $D\tau = \mu\tau$ we see that μ is exact with $\mu = d\nu$ iff $D(e^{-\nu}\tau) = 0$. It is well known[6,9] that $D(e^{-\nu}\tau) = 0$ iff ω reduces to any "level surface" of $e^{-\nu}\tau$. ■

The following corollary is a trivial but important consequence of the theorem.

Corollary 5.4: Let $\tau: P \rightarrow \mathscr{B}(V)$ be a ρ-metric Higgs field and ω a connection on P such that $D\tau = (d\nu)\tau$ for some smooth function $\nu: P \rightarrow \mathbb{R}$. If $\Phi: P \rightarrow P$ is defined by $\Phi(u) = u\cdot e^{-\nu(u)}$ then

(1) Φ is an equivariant automorphism of P,

(2) Φ carries the $G \cap \mathrm{CU}(\rho)$ subbundle $P_{\tilde{\tau}} := \tilde{\tau}^{-1}([\rho])$ onto itself,

(3) Φ carries the $G \cap U(\rho)$ subbundles $P_r := \tau^{-1}(r\cdot\rho)\, r > 0$, onto other $G \cap U(\rho)$ subbundles $\Phi(P_r)$ of $P_{\tilde{\tau}}$,

(4) ω reduces to $P_{\tilde{\tau}}$ but generally does not reduce to the various P_r, $r > 0$. On the other hand, ω does reduce to the subbundles $\Phi(P_r)$, $r > 0$.

Remark: The automorphism Φ "changes scale" at each $x \in M$ in such a way that the P_r are "pushed into" subbundles on which ω reduces.

Generally, in case $\tau: P \rightarrow \mathscr{B}(V)$ is any metric Higgs field on P and ω is any connection such that $D\tau = \mu\tau$ for some 1-form μ on P it follows from the Remark preceding Theorem 5.3 that $\mu = \pi^*\lambda$ for some unique 1-form λ on M. It is customary to refer to λ as the Weyl vector of the pair (ω,τ) in spite of the obvious 1-form character of λ.

Theorem 5.5: Let P be a principal bundle whose structure group G satisfies: $^+\mathbb{R} \subseteq G \subseteq \mathrm{CU}(\rho)$. If τ is a ρ-metric Higgs field such that $\tau = \phi^{-2}\rho$ for some $^+\mathbb{R}$-valued Higgs field ϕ then for any connection ω on P the Weyl vector λ of (ω,τ) satisfies the following conditions:

(1) $D\phi = (-2)(\pi^*\lambda)\phi$,

(2) $\omega_{\mathbb{R}} = (-1/2)(\pi^*\lambda) - \dfrac{d\phi}{\phi}$,

(3) $D\tau = (-2)\left[\omega_{\mathbb{R}} + \dfrac{d\phi}{\phi}\right]\tau$.

Remark: Recall that $\omega_{\mathbb{R}}$ is the $\mathbb{R}$-component of the connection ω. Indeed, if $H = G \cap U(\rho)$ then $G = {}^+\mathbb{R} \times H$ and $\mathfrak{g} = \mathfrak{h} \oplus \mathbb{R}$. Thus $\omega = \omega_{\mathfrak{h}} + \omega_{\mathbb{R}}$. Moreover, since the action of H on $\mathscr{B}(V)$ leaves ρ invariant we know that the action of $\mathfrak{h}$ on ρ is zero. In particular,

$$\omega_{\mathfrak{h}}\cdot\tau = \omega_{\mathfrak{h}}\cdot(\phi^{-2}\rho) = \phi^{-2}(\omega_{\mathfrak{h}}\cdot\rho) = 0.$$

Proof of the Theorem: All three statements are easy consequences of the equation:

$$\begin{aligned} \mu\tau = D\tau &= (-2\phi^{-3})(d\phi)\rho + \omega\cdot(\phi^{-2}\rho) \\ &= (-2)\phi^{-3}(d\phi + \omega_{\mathbb{R}}\phi)\rho \\ &= (-2)\phi^{-1}(d\phi + \omega_{\mathbb{R}}\phi)\tau \\ &= (-2)\left(\frac{D\phi}{\phi}\right)\tau. \end{aligned}$$ ■

The final theorem of this section shows that our extension of the concept of a Weyl-vector to appropriate "conformal bundles" satisfies the usual transformation laws expected in a more classical treatment of the subject. In particular, how does the Weyl-vector change when one varies the ρ-metric Higgs field within a given ρ-conformal Higgs field?

Theorem 5.6: Let $P(M,G)$ be a principal bundle with $^+\mathbb{R} \subseteq G$, let ω be a connection on P, and let $\tilde{\tau}$ be a ρ-conformal Higgs field on P. If τ_1 and τ_2 are ρ-metric Higgs fields in $\tilde{\tau}$ with corresponding Weyl-vectors λ_1,λ_2 then there is a function $\xi: M \rightarrow {}^+\mathbb{R}$ such that

$$\lambda_1 = \lambda_2 - \tfrac{1}{2}d(\ln\xi).$$

Thus the two Weyl-vectors differ by a "gradient."

Proof: Choose $^+\mathbb{R}$-valued Higgs fields ϕ_1,ϕ_2 such that $\tau_1 = \phi_1^{-2}\rho$, $\tau_2 = \phi_2^{-2}\rho$ on $P_{\tilde{\tau}}$. Let $\theta = \phi_1\phi_2^{-1}$ on $P_{\tilde{\tau}}$ and recall from Theorem 5.5 that

$$D\phi_1 = (-2)(\pi^*\lambda_1)\phi_1,$$

$$D\phi_2 = (-2)(\pi^*\lambda_2)\phi_2.$$

Since θ is invariant under $G \cap \mathrm{CU}(\rho)$ it follows that $\theta = \pi^*\xi$ for some $\xi: M \rightarrow {}^+\mathbb{R}$. We claim that $\lambda_1 = \lambda_2 - \frac{1}{2}d(\ln\xi)$. To see this observe that $\phi_1 = \theta\phi_2$ and

$$D\phi_1 = \theta D\phi_2 + \phi_2 d\theta.$$

Thus $(-2)(\pi^*\lambda_1)\theta\phi_2 = (-2)(\pi^*\lambda_2)\theta\phi_2 + \phi_2 d\theta$ and $\pi^*\lambda_1 = \pi^*\lambda_2 - \frac{1}{2}d\theta/\theta$. ■

Weyl geometry is a consequence of Weyl's efforts to develop a unified field theory of gravitation and electromagnetism.[19] In Weyl's original theory the Weyl covariant vector λ on the base manifold was identified with the electromagnetic vector potential, and its "curl" $d\lambda$ was identified with the Maxwell field tensor. We shall see in Remark 2 below that in a bundle version of Weyl's theory it is $\omega_{\mathbb{R}}$ rather than $\pi^*(\lambda)$ that one should choose to "represent" the electromagnetic vector potential.

First we briefly clarify what is needed for a bundle formulation of Weyl theory. The appropriate principal bundle is the frame bundle LM of a four-dimensional manifold M. One should be given a "conformal structure" on M which in the present context is given by a conformal Higgs field $\tilde{\tau}$: $LM \rightarrow \mathscr{B}^*(\mathbb{R}^4)$. It is assumed, of course, that the Higgs field $\tilde{\tau}$ maps all of LM onto the orbit of $[\eta]$ where η is the usual (constant) Minkowski metric on $\mathbb{R}^4$. Weyl theory does not fix the metric within the given conformal structure thus the specific metric Higgs field $\tau: LM \rightarrow \mathscr{B}(\mathbb{R}^4)$ is left free to vary within the fixed class $\tilde{\tau}$. We see that $\tilde{\tau}$ breaks the symmetry of

LM to give us a subbundle COM with structure group $O(1,3)\times {}^+\mathbb{R}$. One also needs a connection ω on LM and the dynamics of ω must reduce to COM so that $D\tilde{\tau}=0$. Any such connection ω then decomposes as in the remark following Theorem 5.5:

$$\omega=\omega_0+\omega_{\mathbf{R}}.$$

Here ω_0 has its values in the Lie algebra $o(1,3)$ of $O(1,3)$ and $\omega_{\mathbf{R}}$ is real-valued.

Although the latter paragraph imposes significant restrictions on P, ω, and $\tilde{\tau}$ one does not yet have traditional Weyl theory. If it is required that ω be torsion-free then Weyl theory emerges. Thus we have[12]:

Remark 1: If $P(M,\mathrm{Gl}(4,\mathbb{R}))$ is a principal bundle over a four-dimensional manifold M then in order that P be a bundle model of Weyl geometry one must have:

(1) a connection ω on P,

(2) a symmetry-breaking conformal Higgs field $\tilde{\tau}$ on P such that $D\tilde{\tau}=0$,

(3) a soldering form θ on P such that $D\theta=0$.

The remainder of this section will be devoted to a proof of the following remark.

Remark 2: If $\omega=\omega_0+\omega_{\mathbf{R}}$ decomposes as above then the curvature Ω of ω also admits a decomposition as $\Omega=\Omega_0+\Omega_{\mathbf{R}}$ where Ω_0 is $o(1,3)$ valued, $\Omega_{\mathbf{R}}$ is real-valued and $\Omega_{\mathbf{R}}=d\omega_{\mathbf{R}}$. If λ is the Weyl vector of the pair (ω,τ) where τ is a metric Higgs field such that $\tilde{\tau}=[\tau]$ then it is true that $\Omega_{\mathbf{R}}=d(-2\pi^*\lambda)$ but λ clearly depends on the metrical substructure of $\tilde{\tau}$ whereas $\omega_{\mathbf{R}}$ does not. This suggests that $\omega_{\mathbf{R}}$ should play the role of the vector potential rather than $\pi^*(-2\lambda)$.

We now establish the assertions made in Remark 2.

Using the definition $\Omega=d\omega+1/2[\omega,\omega]$ of the curvature Ω of ω together with the commutation properties $[\omega_0,\omega_{\mathbf{R}}]=[\omega_{\mathbf{R}},\omega_{\mathbf{R}}]=0$ it is easy to show that Ω itself can be decomposed as

$$\Omega=\Omega_0+\Omega_{\mathbf{R}},$$

where

$$\Omega_0\equiv d\omega_0+1/2[\omega_0,\omega_0] \quad \text{and} \quad \Omega_{\mathbf{R}}=d\omega_{\mathbf{R}}.$$

Now let λ denote the Weyl vector of the pair (ω,τ). Then according to Theorem 5.5 we may write $\omega_{\mathbf{R}}=-2\pi^*(\lambda)-d\phi/\phi$ for some ${}^+\mathbb{R}$-valued Higgs field ϕ. Hence the ${}^+\mathbb{R}$-part of Ω may be expressed as

$$\Omega_{\mathbf{R}}=d\omega_{\mathbf{R}}=d\left(-2\pi^*\lambda-\frac{d\phi}{\phi}\right)=-2\pi^*(d\lambda).$$

$\Omega_{\mathbf{R}}$ is thus a horizontal invariant two-form on COM. If we do a "conformal change" from τ to $\bar{\tau}$, then according to Theorem 5.6 λ changes to $\bar{\lambda}=\lambda+\frac{1}{2}d(\ln\xi)$ for some ${}^+\mathbb{R}$-valued function ξ on M. Clearly $d\bar{\lambda}=d\lambda$ and thus $\Omega_{\mathbf{R}}=-2\pi^*(d\lambda)=-2\pi^*(d\bar{\lambda})$. ■

VI. WEINBERG–SALAM THEORY AS A CONFORMAL GAUGE THEORY

The purpose of this section is to give a geometric model of the Weinberg–Salam theory of the electroweak interaction. Normally the Weinberg–Salam model is presented[11] as a Lagrangian field theory whose Lagrangian is invariant under the Lie group of matrices $SU(2)\times U(1)$. Our presentation shall focus on a principal $\mathrm{Gl}(2,\mathbb{C})$ fiber-bundle-with-connection along with attendant Higgs fields. One of these Higgs fields will serve to define a conformal structure on the bundle as in the previous section while the other plays the same role as the usual scalar Higgs fields in the Lagrangian model.

A strict analogy with the Weinberg–Salam model would lead one to expect the geometrical arena to be a principal $SU(2)\times U(1)$ bundle with connection. We have chosen to utilize a $\mathrm{Gl}(2,\mathbb{C})$ bundle for two reasons.

First we wish to establish an analogy with the fiber bundle version of gravitational theories,[9,12] and these theories are generally modeled on the frame bundle of a spacetime M. By choosing a full $\mathrm{Gl}(2,\mathbb{C})$ bundle P we are assured by Theorem 2.2 that P may be identified as a bundle of frames of the bundle associated to P and the usual action of $\mathrm{Gl}(2,\mathbb{C})$ on $\mathbb{C}^2$. At this point we can, as in gravitational theories, let a metric choose a $U(2)$ subbundle. A suitably normalized scalar Higgs field may then be used to reduce this $U(2)$ subbundle to an electromagnetic $U(1)$ bundle as in the traditional Weinberg–Salam model.[20]

A second reason for choosing a bundle whose group contains $U(2)$ is that in doing so we allow the presence of a conformal structure which will enable us to give geometric meaning to the surviving component of the scalar Higgs field which in the usual model describes a scalar particle.

Throughout the remainder of this section we will assume given the following data:

(1) a spacetime manifold M along with a Lorentzian metric g,

(2) a $\mathrm{Gl}(2,\mathbb{C})$ principal fiber bundle P over M,

(3) a connection ω on P,

(4) a conformal Higgs field $\tilde{\tau}$ on P which reduces the connection ω ($D\tilde{\tau}=0$) and which maps onto the orbit of the class $[\rho]\in\mathscr{B}_*(\mathbb{C}^2)$ where ρ is the usual metric on $\mathbb{C}^2$: $\rho(z,w)=z_1\bar{w}_1+z_2\bar{w}_2$,

(5) a class of scalar Higgs fields $\psi:P\to\mathbb{C}^2$ which are assumed to be nontrivial symmetry-breaking Higgs fields.

The usual Weinberg–Salam model concerns itself with $u(2)=su(2)+u(1)$ valued gauge fields as well as scalar Higgs fields such as ψ above. Our first task is to show how to use the conformal structure to choose possible candidates for an appropriate $U(2)$-subbundle of P on which ω reduces. The components of the reduced connection ω will then play the role of the gauge fields which occur in the usual model.

Since we have assumed that $D\tilde{\tau}=0$ we know that if $\sigma:P\to\mathscr{B}(\mathbb{C}^2)$ is any metric Higgs field such that $\tilde{\tau}(u)=[\sigma(u)]$ for all $u\in P$, then

$$D\sigma=\mu_\sigma\sigma$$

for some 1-form μ_σ on P. The following observation is a consequence of Corollary 5.4.

Observation 1: If σ is a metric Higgs field on P such that $\mu_\sigma=d\nu_\sigma$ for some function $\nu_\sigma:P\to\mathbb{R}$ then the function

$$\Phi(u)=u\cdot e^{-\nu_\sigma(u)}$$

is an automorphism of P which carries $P_{\tilde{\tau}}:=\tilde{\tau}^{-1}([\rho])$ onto $P_{\tilde{\tau}}$ and which carries $P_r:\sigma^{-1}(r\cdot\rho)$ onto a $U(2)$-subbundle $\Phi(P_r)$ on which ω reduces. In particular, any one of the subbundles $\Phi(P_r)$, $r>0$ along with the reduced connection $\omega|\Phi(P_r)$ is an acceptable arena for Weinberg–Salam theory.

At this point if we fix $r>0$ then we see that for any nonvanishing Higgs field $\psi: P\to\mathbb{C}^2$ it follows that $(\psi/\|\psi\|)$, $\|\psi\|^2=\rho(\psi,\psi)$, is a Higgs field on $\Phi(P_r)$. Since $\psi/\|\psi\|$ maps $\Phi(P_r)$ onto a sphere of radius 1 in $\mathbb{C}^2$ and since the orbits of U(2) on $\mathbb{C}^2$ are precisely the set of all spheres in $\mathbb{C}^2$ it follows that $\chi := \psi/\|\psi\|$ is a symmetry-breaking Higgs field on $\Phi(P_r)$. It follows that $\chi^{-1}\binom{0}{1}$ is a U(1) subbundle of $\Phi(P_r)$ which we denote by Q_r. Its structure group is the isotropy subgroup of $\binom{0}{1}$ in U(2). If

$$A=e^{i\theta}\begin{pmatrix}\alpha & \beta\\ -\bar\beta & \bar\alpha\end{pmatrix},\quad |\alpha|^2+|\beta|^2=1,$$

is a typical element of U(2) we see that $A\binom{0}{1}=\binom{0}{1}$ if and only if

$$A=\begin{pmatrix}e^{2i\theta} & 0\\ 0 & 1\end{pmatrix}.$$

Thus we have:

Observation 2: If $\psi: P\to\mathbb{C}^2$ is a nontrivial Higgs field on P and $\chi := \psi/\|\psi\|$ then the subbundle of $\Phi(P_r)$ defined by $\chi^{-1}\binom{0}{1}\cap\Phi(P_r)$ is denoted by Q_r and is a U(1) subbundle of the U(2) bundle $\Phi(P_r)$. Its group is denoted by $U_c(1)$ and is called the charge conservation subgroup of U(2).[21] A matrix $A\in U(2)$ belongs to $U_c(1)$ if and only if

$$A=\begin{pmatrix}e^{2i\theta} & 0\\ 0 & 1\end{pmatrix}.$$

Observe that if τ^1, τ^2, τ^3 are the Pauli matrices and $T_1=\frac{1}{2}i\tau^1$, $T_2=\frac{1}{2}i\tau^2$, $T_3=\frac{1}{2}i\tau^3$, $T_4=iI$ where I is the 2×2 identity matrix, then $\{T_1,T_2,T_3,T_4\}$ is a basis of the Lie algebra $u(2)$ and, moreover, $u_c(1)$ is generated by $T_3+\frac{1}{2}T_4$. Observe also that if we consider the connection ω to be reduced to

$$P_{\tilde\tau}=\{u\,|\,\tilde\tau(u)=[\rho]\}$$

then ω is $cu(2)$ valued and we may write

$$\omega=\omega^1 gT_1+\omega^2 gT_2+\omega^3 gT_3+\tfrac{1}{2}\omega^4 g'T_4+\omega_{\mathbb{R}}g''I,$$

where g, g', g'' are coupling constants to be determined by empirical data. Since $\omega_{\mathbb{R}}$ will be assumed to be "flat" this term will not represent a particle in this model so we choose $g''=1$. Observe that if M is the subspace of $cu(2)$ generated by $\{T_1,T_2,\alpha T_3+\beta T_4\}$ for some constants α and β such that $\alpha T_3+\beta T_4\neq 0$ then $\mathrm{Ad}(h)(\mathfrak{M})\subseteq\mathfrak{M}$ for every $h\in U_c(1)\times{}^{+}\mathbb{R}\subseteq CU(2)$. Consequently by our Fundamental theorem of Sec. 2 we see that $\omega|TQ_r$ has five components, four of which are tensorial fields on Q_r. One of these tensorial fields is $\omega_{\mathbb{R}}|TQ_R$ which will be assumed to be exact and thus represents no particle in the model whereas the other three tensorial components of $\omega|TQ_r$ will represent massive particles (vector mesons). The only component of $\omega|TQ_r$ which fails to be tensorial is actually a connection on Q_r and will represent a massless particle (photon) in our model. It is desirable to compute the masses of the meson fields and with this in mind we observe that on any one of the U(2) bundles $\Phi(P_r)$, $\omega_{\mathbb{R}}$ vanishes [since ω reduces to $\Phi(P_r)$.] Also we regard the $\Phi(P_r)$, as being the natural candidates for the real "physical bundle" which is to be the arena for Weinberg–Salam theory. Thus we write $\omega=\omega_U+\omega_{\mathbb{R}}$ and will concentrate our efforts on ω_U, tacitly assuming it to be restricted to some specific $\Phi(P_r)$.

Consider the *real* bilinear form defined by

$$\mathscr{M}(\omega)=e^{\nu}\rho^{\mathbb{R}}(\omega_U\cdot v,\omega_U\cdot v),\tag{6.1}$$

where e^{ν} is a positive constant, $\rho^{\mathbb{R}}$ is the *real* matrix defined on $\mathbb{R}^4$ by $\rho^{\mathbb{R}}(x,y)=\Sigma_{i=1}^4 x_i y_i$, and where ω_U must be expanded in a basis of a *real* representation of $u(2)$. The vector v is a vacuum state vector represented in $\mathbb{R}^4$. It is known[11] that the masses of the particles defined by the components of ω_U on any one of the bundles $\Phi(P_r)$ are given by the eigenvalues of the matrix of the form (6.1). These may also be computed from the coefficients of the appropriate quadratic terms of our Lagrangian (6.7) below. In particular the paragraphs following (6.7) show the details of this computation.

Since the Fundamental theorem requires the use of a basis with special properties we emphasize the role that a change of basis plays in diagonalizing the mass matrix. Since the bilinear form $\mathscr{M}$ is basis-independent we present the details of the mass calculations in terms of $\mathscr{M}$.

To identify the components of ω_U to be assigned masses we must diagonalize the quadratic form $\mathscr{M}$ defined by (6.1). Since this form must be written in a real representation of $U(2)$ we express all vectors $X\in\mathbb{C}^2$ as four dimensional real vectors $X=(\mathrm{Re}\,X_1, \mathrm{Im}\,X_1, \mathrm{Re}\,X_2, \mathrm{Im}\,X_2)$ and we write the matrices T_1, T_2, T_3, T_4 as

$$M_1=\frac{1}{2}\begin{bmatrix}0 & J\\ J & 0\end{bmatrix},\quad M_2=\frac{1}{2}\begin{bmatrix}0 & I\\ I & 0\end{bmatrix},$$

$$M_3=\frac{1}{2}\begin{bmatrix}J & 0\\ 0 & -J\end{bmatrix},\quad M_4=\begin{bmatrix}J & 0\\ 0 & J\end{bmatrix},$$

where

$$J=\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}.$$

We identify T_i with M_i so that

$$\omega_U=g\omega^1 M_1+g\omega^2 M_2+g\omega^3 M_3+\tfrac{1}{2}g'\omega^4 M_4.$$

Since, on Q_r, our symmetry-breaking Higgs field χ has the value $(0,1)\in\mathbb{C}^2$ we see that the *real* vacuum state vector is $v=(0,0,1,0)$. Observe that $M_1\cdot v=(0,\frac{1}{2},0,0)$, $M_2\cdot v=(\frac{1}{2},0,0,0)$, $M_3\cdot v=(0,0,0,-\frac{1}{2})$, $M_4\cdot v=(0,0,0,\frac{1}{2})$ so that

$$\begin{aligned}\mathscr{M}(\omega)=e^{\nu}[&\tfrac{1}{4}g^2(\omega^1)^2+\tfrac{1}{4}g^2(\omega^2)^2\\ &+\tfrac{1}{4}g^2(\omega^3)^2+\tfrac{1}{4}(g')^2(\omega^4)^2-\tfrac{1}{2}gg'\omega^3\omega^4].\end{aligned}$$

We now wish to write $\mathscr{M}$ in terms of a new basis $\{\bar M_1,\bar M_2,\bar M_3,\bar M_4\}$ such that $\mathscr{M}$ is diagonalized in this new basis. It suffices to find an orthogonal matrix R such that $\bar M_i=\Sigma_{j=1}^4 R_{ji}(g_j M_j)$ where $g_1=g_2=g_3=g$ and $g_4=(g'/2)$. Since the first two terms of $\mathscr{M}(\omega)$ are already in diagonal form it suffices to choose R in the form

$$R=\begin{bmatrix}I & 0\\ 0 & A\end{bmatrix},$$

where A is a 2×2 orthogonal matrix. We write

$$A=\begin{bmatrix}\cos\alpha & \sin\alpha\\ -\sin\alpha & \cos\alpha\end{bmatrix}$$

for some α so that $\overline{M}_1 = gM_1, \overline{M}_2 = gM_2$,

$$\overline{M}_3 = (\cos\alpha)gM_3 + (\sin\alpha)(\tfrac{1}{2}g')M_4$$
$$\overline{M}_4 = (-\sin\alpha)gM_3 + (\cos\alpha)(\tfrac{1}{2}g')M_4.$$

In this new basis we have

$$\omega_U = \overline{\omega}^1\overline{M}_1 + \overline{\omega}^2\overline{M}_2 + \overline{\omega}^3\overline{M}_3 + \overline{\omega}^4\overline{M}_4$$

where $\overline{\omega}^1 = \omega^1, \overline{\omega}^2 = \omega^2, \overline{\omega}^3 = (\cos\alpha)\omega^3 + (\sin\alpha)\omega^4$, and $\overline{\omega}^4 = (-\sin\alpha)\omega^3 + (\cos\alpha)\omega^4$. Thus

$$\mathscr{M}(\omega) = \frac{e^\nu}{4}g^2(\overline{\omega}^1)^2 + \frac{e^\nu}{4}g^2(\overline{\omega}^2)^2 + \frac{e^\nu}{4}[(-g\cos\alpha + g'\sin\alpha)\overline{\omega}^3 + (g\sin\alpha + g'\cos\alpha)\overline{\omega}^4]^2.$$

We wish to choose α so that $\mathscr{M}(\omega)$ will be diagonalized. We also want one of the fields to be a $U_c(1)$ gauge field and hence a connection on the surviving $U_c(1)$ bundle. Thus it should have values in the Lie algebra generated by $M_3 + \frac{1}{2}M_4$ since this is the generator of $u(2)$ which annihilates the explicit vector v we used in computing $\mathscr{M}(\omega)$ above.[21] Clearly $\overline{\omega}^3$ and $\overline{\omega}^4$ are the only candidates for this field and in fact we may choose either to be the gauge field. We choose to let $\overline{\omega}^4$ be the component of the surviving connection on Q_r so we want $\overline{M}_4 = c(M_3 + \frac{1}{2}M_4)$ for some constant c. It follows that $(-\sin\alpha)g = c$. and $(\cos\alpha)g' = c$. Thus

$$\cos\alpha = g[g^2 + (g')^2]^{-1/2},$$
$$\sin\alpha = -g[g^2 + (g')^2]^{-1/2} \tag{6.2}$$

and

$$A = [g^2 + (g')^2]^{-1/2}\begin{bmatrix} g & -g' \\ g' & g \end{bmatrix}.$$

It follows that

$$\mathscr{M}(\omega) = \frac{e^\nu}{4}g^2(\overline{\omega}^1)^2 + \frac{e^\nu}{4}g^2(\overline{\omega}^2)^2 + \frac{e^\nu}{4}((g^2 + (g')^2)^{1/2})^2(\overline{\omega}^3)^2.$$

Thus $\overline{\omega}^1, \overline{\omega}^2, \overline{\omega}^3$ represent massive vector mesons with masses $\frac{1}{2}e^{\nu/2}g$, $\frac{1}{2}e^{\nu/2}g$, and $\frac{1}{2}e^{\nu/2}(g^2 + (g')^2)^{1/2}$, respectively. We see also that $(\overline{\omega}^4|TQ_r)\overline{M}_4$ represents a massless gauge field. Since $\overline{M}_4 = gg'[g^2 + (g')^2]^{-1/2}(M_3 + \frac{1}{2}M_4)$ it follows that its eigenvalues are a multiple of $gg'[g^2 + (g')^2]^{-1/2}$ and thus this gauge field interacts with charged particles whose fundamental unit of charge is $gg'[g^2 + (g')^2]^{-1/2}$. Moreover it follows from (6.2) that

$$\overline{\omega}^3 = \frac{g\omega^3 - g'\omega^4}{(g^2 + (g')^2)^{1/2}},$$
$$\overline{\omega}^4 = \frac{g'\omega^3 + g\omega^4}{(g^2 + (g')^2)^{1/2}}.$$

If we identify α as $-\theta_w$ where θ_w is the Weinberg angle then we may identify[11] $\overline{\omega}^3$ with $-Z$ and $\overline{\omega}^4$ with A where Z is the massive neutral field and A is the massless electromagnetic vector potential in the traditional form of Weinberg–Salam theory.

We would also like to emphasize the fact that the fields $\overline{\omega}^1, \overline{\omega}^2$ represent charged particles as they experience "mixing" under the action of the charge conservation group $U_c(1)$. Indeed, we have that $[\overline{M}_4,\overline{M}_1] = gg'[g^2 + (g')^2]^{-1/2}\overline{M}_2$ and $[\overline{M}_4,\overline{M}_2] = gg'[g^2 + (g')^2]^{-1/2}\overline{M}_1$ and consequently $U_c(1)$, acting via the adjoint representation on $u(2)$, rotates the plane spanned by $\{\overline{M}_1,\overline{M}_2\}$ onto itself. This suggests that $U_c(1)$ should be regarded as a "gauge group" for the pair $\overline{\omega}^1$, $\overline{\omega}^2$ and that these fields represent charged particles. As is customary we write them as complex fields:

$$W^{\pm} = \overline{\omega}^1 \mp i\overline{\omega}^2.$$

Finally we illustrate how to make this geometrical picture explicit in terms of a specific Lagrangian field theory. A typical Lagrangian that couples a $\mathbb{C}^2$-valued Higgs field ψ to a metric Higgs field σ on a $Gl(2,\mathbb{C})$ principal bundle with connection ω is

$$\mathscr{L} = 4\sigma_{ij}{}^\star D\psi^i \wedge D\overline{\psi}^j + 4\sigma^{ij}\sigma^{kl\star}D\sigma_{ik} \wedge D\overline{\sigma}_{jl} - V(\sigma(\psi,\psi)). \tag{6.3}$$

In this Lagrangian ${}^\star D\psi$ and ${}^\star D\sigma$ are the duals[22] of the covariant exterior derivatives $D\psi$ and $D\sigma$ of ψ and σ, respectively. Moreover,

$$V(\sigma(\psi,\psi)) = \widetilde{V}(\sigma(\psi,\psi))\pi^\star(d\Lambda),$$

where $\widetilde{V}(\sigma(\psi,\psi))$ is a typical symmetry-breaking potential[11] and $\pi^\star(d\Lambda)$ is the pullback under the projection $\pi:P\to M$ of the volume $d\Lambda$ on M defined by the spacetime metric. Observe that since both σ and ψ transform under $\mathbb{R}^+ \subseteq Gl(2,\mathbb{C})$, our Lagrangian is scale invariant.

We shall specialize the geometry to reduce this Lagrangian theory in first order in the $\mathbb{R}^+$-parameter to the Lagrangian of the Weinberg–Salam theory. In the process we will show that the single surviving component of the Higgs scalar field in the Weinberg–Salam theory is an infinitesimal conformal factor which scales the metric in the surviving $U_c(1)$ charge conservation subbundle of P.

Our first step is to impose a condition on the "bundle with connection and metric" which will insure that the geometry of the bundle is a suitable generalized Weyl geometry. This is reflected in the Lagrangian by requiring that $\mathscr{L}$ be restricted to only those metric Higgs fields σ which come out of a fixed conformal class defined by a ρ-conformal Higgs field $\tilde{\tau}:P\to\mathscr{B}^\star(\mathbb{C}^2)$ such that $D\tilde{\tau} = 0$. It follows from Theorem 5.2 that such σ satisfy $D\sigma = \mu_\sigma\sigma$ for some 1-form μ_σ on P. We also require the Weyl connection degenerate to a metric connection and consequently it follows from Theorem 5.5 that the $\mathbb{R}$-part of ω, $\omega_{\mathbb{R}}$ is exact. Thus there is a function $\xi:P\to\mathbb{R}$ such that $\omega_{\mathbb{R}} = d\xi$. It is not difficult to see[23] that $e^{-\xi}$ is actually an ${}^+\mathbb{R}$-valued Higgs field on P and thus that $e^{-\xi}\cdot\rho = e^{2\xi}\rho$ is a metric Higgs field on P in the class $\tilde{\tau}$. Since $D(e^{2\xi}\rho) = 2e^{2\xi}[d\xi - \omega_{\mathbb{R}}]\rho = 0$ we see that ω reduces to the U(2)-subbundle defined by $(e^{2\xi}\rho)^{-1}(r\cdot\rho)$. Observe that these subbundles may be defined more simply by the equation $\xi = -\ln r$; they are simply the "level surfaces" of ξ.

On the other hand, the fact that $\omega_{\mathbb{R}}$ is exact implies that μ_σ is exact where $D\sigma = \mu_\sigma\sigma$ as above. Thus $\mu_\sigma = d\nu_\sigma$ for some function $\nu_\sigma:P\to\mathbb{R}$. If, moreover, $\phi:P\to{}^+\mathbb{R}$ is the ${}^+\mathbb{R}$-valued Higgs field such that $\sigma = \phi\cdot\rho = \phi^{-2}\rho$ then we denote ν_σ by ν_ϕ and we have $d\nu_\phi = \mu_\sigma = -2[\omega_{\mathbb{R}} + d\phi/\phi] = d[-2\xi - 2\ln\phi]$. Thus, by modifying ξ if necessary, we have

$$\nu_\phi = -2\xi - 2\ln\phi \tag{6.4}$$

or

$$e^{\nu_\phi}e^{2\xi} = \phi^{-2}. \tag{6.5}$$

It follows from these remarks that the following statements are equivalent:

(1) $D\sigma = 0$,

(2) ν_ϕ is constant

(3) $\sigma = \phi^{-2}\rho = e^{\nu_\phi}(e^{-\xi}\cdot\rho)$ is a constant multiple of the metric Higgs field $(e^{-\xi}\cdot\rho)$.

Thus if $D\sigma = 0$ then the U(2)-subbundles defined by $\sigma^{-1}(r\cdot\rho)$, $r > 0$ are the same as those defined by the level surfaces of ξ.

If we restrict our Lagrangian to fields defined on these U(2)-subbundles then $D\sigma = 0$ and the Lagrangian becomes

$$\mathcal{L} = 4\phi^{-2}\rho_{ij}(\star D\psi^i \wedge \overline{D\psi^j}) - V(\phi^{-2}\|\psi\|^2). \tag{6.6}$$

If we assume ψ is a nontrivial symmetry-breaking Higgs field then $\|\psi\|$ is never zero and it defines a metric Higgs field τ by $\tau = \|\psi\|^{-2}\rho$. This field accounts for the "radial" dependence of ψ. Let $\chi = \psi/\|\psi\|$. Then χ accounts for the "spherical" dependence of ψ and is itself a symmetry-breaking Higgs field on any one of the U(2)-subbundles, $\xi =$ constant. We now show how to rewrite $\mathcal{L}$ in terms of $\|\psi\|$, ϕ, χ, and ν_τ. To do this simply observe that[24]

$$\begin{aligned} D\psi &= d(\|\psi\|\chi) + \omega\cdot(\|\psi\|\chi) \\ &= \|\psi\|d\chi + d(\|\psi\|)\chi + \omega_U\cdot(\|\psi\|\chi) + \omega_{\mathbf{R}}\cdot(\|\psi\|\chi) \\ &= \|\psi\|(d\chi + \omega_U\cdot\chi) + d(\|\psi\|)\chi + \omega_{\mathbf{R}}\cdot(\|\psi\|\chi) \\ &= \|\psi\|D\chi + \left[\frac{d(\|\psi\|)}{\|\psi\|} + \omega_{\mathbf{R}}\right]\|\psi\|\chi \end{aligned}$$

and thus

$$D\psi = \|\psi\|\left[D\chi - \tfrac{1}{2}(d\nu_\tau)\chi\right].$$

It follows that

$$\begin{aligned} \star D\psi^i \wedge \overline{D\psi^j} = \|\psi\|^2&[\star D\chi^i \wedge \overline{D\chi^j} \\ &+ \tfrac{1}{4}(\star d\nu_\tau \wedge d\nu_\tau)\chi^i\overline{\chi^j}] \\ &+ \|\psi\|^2[(-\tfrac{1}{2})(\star d\nu_\tau\chi^i) \wedge \overline{D\chi^j} \\ &- \star D\chi^i \wedge \tfrac{1}{2}(d\nu_\tau)\overline{\chi^j}]. \end{aligned}$$

But $\|\chi\|^2 = 1$ and $0 = D(\rho(\chi,\chi)) = \rho(D\chi,\chi) + \rho(\chi,D\chi)$ so that $\rho_{ij}\chi^i(\overline{D\chi^j}) = 0$. Also[22]

$$\begin{aligned} \rho_{ij}(\star D\chi^i \wedge (d\nu_\tau)\overline{\chi^j}) &= -\rho_{ij}(D\chi^i \wedge \star(d\nu_\tau\overline{\chi^j})) \\ &= -(\rho_{ij}(D\chi^i)\overline{\chi^j}) \wedge \star d\nu_\tau = 0. \end{aligned}$$

Thus

$$\rho_{ij}(\star D\psi^i \wedge \overline{D\psi^j}) = \|\psi\|^2[\rho_{ij}(\star D\chi^i \wedge \overline{D\chi^j}) + \tfrac{1}{4}(\star d\nu_\tau \wedge d\nu_\tau)$$

and

$$\begin{aligned} \mathcal{L} = 4\phi^{-2}\|\psi\|^2&[\rho_{ij}(\star D\chi^i \wedge \overline{D\chi^j}) \\ &+ \tfrac{1}{4}(\star d\nu_\tau \wedge d\nu_\tau)] - V(\phi^{-2}\|\psi\|^2). \end{aligned}$$

Since $\nu_\phi = -2\ln\phi - 2\xi$ and $\nu_\tau = -2\ln\|\psi\| - 2\xi$ we have that $e^{\nu_\phi - \nu_\tau} = \|\psi\|^2\phi^{-2}$. Thus our general Lagrangian is

$$\begin{aligned} \mathcal{L} = 4&[e^{\nu_\phi - \nu_\tau}\rho_{ij}(\star D\chi^i \wedge \overline{D\chi^j}) \\ &+ [\star de^{\frac{1}{2}(\nu_\phi - \nu_\tau)} \wedge de^{\frac{1}{2}(\nu_\phi - \nu_\tau)}]] - V(e^{\nu_\phi - \nu_\tau}). \end{aligned} \tag{6.7}$$

Here we have assumed that $\sigma = \phi^{-2}\rho$ satisfies $D\sigma = 0$ and thus that ν_ϕ is constant. We see that this new Lagrangian is a function only of χ, ν_τ and the connection ω. Moreover the components of the connection occur only in those terms involving $D\chi^i$. If we define a subbundle $Q_r \subseteq P_r = \{u\in P\,|\,\xi(u) = -\ln r\}$ by $Q_r = (\chi\,|P_r)^{-1}\binom{0}{1}$ then the Q_r are the charge conservation subbundles defined early in this section. On these bundles χ is constant and $D\chi = d\chi + \omega\cdot\chi = \omega\cdot\chi = \omega_U\cdot\chi$. It follows from the identity (C1) of Appendix C that the only terms of $\mathcal{L}$ having connection terms are those in

$$\begin{aligned} &4e^{\nu_\phi - \nu_\tau}\rho_{ij}(\star D\chi^i \wedge \overline{D\chi^j}) \\ &\quad = e^{\nu_\phi - \nu_\tau}\rho_{ij}g^{\alpha\beta}D_\alpha\chi^i\overline{D_\beta\chi^j}\pi^*(dA) \\ &\quad = e^{\nu_\phi - \nu_\tau}\rho_{ij}g^{\alpha\beta}[\omega_U(\bar\partial_\alpha)\cdot\chi]^i\overline{[\omega_U(\bar\partial_\beta)\cdot\chi]^j}(\pi^*dA) \\ &\quad = e^{\nu_\phi - \nu_\tau}g^{\alpha\beta}\rho(\omega_U(\bar\partial_\alpha)\chi, \omega_U(\bar\partial_\beta)\chi)\pi^*(dA). \end{aligned}$$

In this equation $\bar\partial_\alpha$ denotes $\partial/\partial(x^\alpha\circ\pi)$ where (x^α) are coordinates on M.

In *Minkowski space* $(g^{\alpha\beta}) = (\eta^{\alpha\beta})$ is the Minkowski metric and $\eta^{\alpha\beta} = 0$ for $\alpha \neq \beta$. One usually calculates the masses of the fields represented in the Lagrangian at the vacuum state. In our case we shall see that ν_τ may be identified as a multiple of the surviving component of the scalar Higgs field in the usual treatment of Weinberg–Salam theory. Thus ν_τ represents a perturbation away from the vacuum so that masses of particles associated with components of ω_U are obtained from the matrix of the quadratic form

$$\tilde{\mathcal{M}}(\omega) = e^{\nu_\phi}\eta^{\alpha\beta}\rho(\omega_U(\bar\partial_\alpha)\chi, \omega_U(\bar\partial_\beta)\chi) = e^{\nu_\phi}\rho(\omega_{U\alpha}\cdot\chi, \omega_U^\alpha\cdot\chi)$$

with $\nu_\tau = 0$. It is easy to see that the eigenvalues of the matrix associated with this form are the same as those of the matrix of

$$\mathcal{M}(\omega) = e^{\nu_\phi}\rho(\omega_U\cdot\chi, \omega_U\cdot\chi) \tag{6.8}$$

and this is the form utilized earlier to obtain the required masses of the components of ω_U.

At this point we show how our Lagrangian collapses to the one usually used in Weinberg–Salam theory. To get the usual Lagrangian one passes to first order. The Lagrangian (6.7) is defined on a family of fields $e^{(1/2)(\nu_\phi - \nu_\tau)}$ each having values in the group $^+\mathbb{R}$ while the usual Lagrangian is defined on fields with values in the Lie algebra $\mathbb{R}$ of $^+\mathbb{R}$. Thus if we approximate $e^{(1/2)(\nu_\phi - \nu_\tau)}$ by $1 + \frac{1}{2}(\nu_\phi - \nu_\tau)$ we should, up to minor modifications, obtain the usual Weinberg–Salam Lagrangian. Indeed if we take $\eta = -(1/\sqrt{2})\,\nu_\tau$ and $v = \sqrt{2} + (1/\sqrt{2})\nu_\phi$ then $(1/\sqrt{2})(\eta + v) = 1 + \frac{1}{2}(\nu_\phi - \nu_\tau)$ and, to first order

$$e^{(1/2)(\nu_\phi - \nu_\tau)} \approx (\eta + v)/\sqrt{2}.$$

With this approximation (6.7) becomes

$$\begin{aligned} \mathcal{L} = 4&\left(\frac{\eta + v}{\sqrt{2}}\right)^2\rho_{ij}(\star D\chi^i \wedge \overline{D\chi^j}) \\ &+ 2(\star d\eta \wedge d\eta) - V\left[\left(\frac{\eta + v}{\sqrt{2}}\right)^2\right]. \end{aligned}$$

If we use the identity (C1) of Appendix C once more and restrict our fields to Q_r we obtain ($\mathcal{L} = \tilde{\mathcal{L}}\pi^*(dA)$)

$$\mathscr{L} = \tfrac{1}{2}(\overline{\partial}^\mu \eta)(\overline{\partial}_\mu \eta) + \left[\frac{(\eta+v)^2}{2}\right] \times \rho_{ij}(\omega_U \cdot \chi)^i \, \overline{(\omega_U \cdot \chi)^j} - \widetilde{V}\left[\left(\frac{\eta+v}{\sqrt{2}}\right)^2\right] \tag{6.9}$$

which is clearly the usual Higgs Lagrangian for the Weinberg–Salam model.[11]

We also point out that in the usual model the field ψ may be written

$$\psi = \frac{v+\eta}{\sqrt{2}}\begin{pmatrix}0\\1\end{pmatrix}$$

in an appropriate U(2) gauge. In general we know that on Q_r $\psi = \|\psi\|\binom{0}{1}$ and also $e^{\nu_\phi - \nu_\tau} = \phi^{-2}\|\psi\|^2$. Thus, to first order we have

$$\|\psi\| \approx \phi\,[1 + \tfrac{1}{2}(\nu_\phi - \nu_\tau)] = \phi\left(\frac{\eta+v}{\sqrt{2}}\right)$$

and

$$\psi \approx \phi\left(\frac{\eta+v}{\sqrt{2}}\right)\begin{pmatrix}0\\1\end{pmatrix}.$$

It follows from (6.5) that on P_r

$$\phi = re^{-(1/2)\nu_\phi}. \tag{6.10}$$

Thus

$$\psi = re^{-(1/2)\nu_\phi}\left(\frac{\eta+v}{\sqrt{2}}\right)\begin{pmatrix}0\\1\end{pmatrix}$$

and if we like we may choose r such that $re^{-(1/2)\nu_\phi} = 1$ (recall[11] that the masses of $W^\pm$, Z and the charge e determine the coupling constants g, g' and an additional parameter which in our case is $\nu := \nu_\phi$).

Finally we wish to remark that this geometric model does indeed provide us with a geometric interpretation of the surviving Weinberg–Salam scalar field η. Using (6.10) and the Lagrangian (6.7) we obtain

$$\mathscr{L} = 4\big[e^{-\nu_\tau}(r\phi^{-2}\rho)_{ij}({}^*D\chi^i \wedge \overline{D\chi^j}) + ({}^*de^{\frac{1}{2}(\nu_\phi - \nu_\tau)} \wedge de^{\frac{1}{2}(\nu_\phi - \nu_\tau)}\big] - V(e^{\nu_\phi - \nu_\tau}).$$

Thus the term $e^{-\nu_\tau}$ serves only to scale the metric $(r\phi^{-2})\rho = r(\phi\cdot\rho)$. Once we choose r and the constant ν_ϕ it follows that $r\phi^{-2}$ is constant. Thus $e^{-\nu_\tau}$, as ν_τ varies, is a conformal scaling. It follows that the field equations serve to choose the conformal scale for the metric. Recall, however, that the Higgs scalar η was identified as $(-1/\sqrt{2})\,\nu_\tau$ in the first order version of $\mathscr{L}$. It follows then that η plays the role of an infinitesimal conformal factor in this model.

VII. CONCLUSIONS

A central objective of this paper has been to obtain a clearer understanding of the interaction between symmetry-breaking Higgs fields and gauge fields. To accomplish this purpose we have focused on the geometric aspects of both symmetry-breaking Higgs fields and gauge fields. Previous geometric work along these lines concentrated on gauge fields whose dynamics were restricted to those degrees of freedom represented by the symmetry group which survives after application of the Higgs mechanism. In mathematical terms this means that the connection reduced to the subbundle defined by the symmetry-breaking Higgs field. We feel that this restriction is not only unnecessary but often neglects some issues of central importance in the theory. To illustrate this point we remark that the Weinberg–Salam theory of the electroweak interaction may be represented within the theory of connections on a principal fiber bundle and that in this model not only is it the case that the dynamics of the gauge field are not restricted to the surviving symmetry group U(1) but in fact the extra dynamical degrees of freedom are needed to model the fields $W^\pm$ and Z. The first part of our paper underlines the distinctive role of symmetry-breaking fields among the set of all Higgs fields. It also shows that under reasonably general conditions a symmetry-breaking Higgs field ϕ allows a decomposition of a connection ω into two pieces, $\omega = \omega' + \tau$ where ω' is a new connection which reduces to a ϕ-subbundle and τ is a tensorial form. In the Weinberg–Salam theory ω' corresponds to the massless electromagnetic gauge field A and the components of τ correspond to the massive vector bosons $W^\pm$ and Z. We wish to point out a possible geometrical generalization of this process.

The Higgs mechanism is a technique which when applied to a specific type of Lagrangian eliminates massless scalar bosons and gives masses to certain vector bosons and possibly to other fields as well. Can we characterize, geometrically, which vector bosons are massive and which are massless? We conjecture that vector bosons which remain massless after symmetry breakdown should be represented by components of connections while vector bosons which acquire mass should be represented by tensorial fields. More generally we offer the following conjecture:

If a Lagrangian field theory is suitably geometrized with the gauge fields represented by the components of a connection ω then the symmetry-breaking mechanism should result in a decomposition of ω as $\omega = \omega' + \tau$ where the components of the connection ω' represent the massless vector bosons of the theory and the components of the tensorial field τ represent the massive vector bosons of the theory.

By applying the Fundamental Theorem of Sec. 2 to the connection developed in our version of the Weinberg–Salam model we have shown that the model is compatible with the conjecture.

A second main concern of this paper has been the question of whether or not the Higgs scalar field in spontaneously broken gauge theories can be geometrized.[25] After reviewing the structure of Higgs fields of the type that occur in spontaneously broken gauge theories we argued that such fields should be associated with symmetry breaking fields on an ${}^+\mathbb{R}$-enlarged bundle. This led to a study of generalized conformal geometry on an arbitrary principal bundle. From our results it was easily seen that Weyl geometry is a special case of this theory and also that the Weinberg–Salam theory may be modeled on a conformal bundle in such a way that it is completely parallel to a "trivial" Weyl geometry. Our conformal model of the Weinberg–Salam theory provided us with a geometrical interpretation of the scalar Higgs field which survives the symmetry-breaking mechanism in the usual model. We found that this field may be identified as an infinitesimal conformal factor and, in fact, that this factor may be obtained from the Weyl vector associated with the model. This identification was possible precisely because the Weyl "vector" was an exact differential form. The fact that it was sufficient to use a trivial Weyl geometry suggests the possibility of generalizing our model to the case where the Weyl vector is not exact. In such a generalization the total Higgs field could be eliminated from the theory in favor of bosons. In this theory one would obtain a final U(1) subbundle and would have in addition to the tensorial fields $W^{\pm},Z$, a new vector boson. In the same way as $W^{\pm},Z$ were defined by certain components of the original connection ω this new field would arise from the component $\omega_{\mathbb{R}}$ of ω which corresponds to the scale degree of freedom present in the conformal group.

APPENDIX A

In this appendix we prove those theorems of Sec. 2 which warrant some argument. We also include the proofs of the more technical results of Sec. 3.

Proof of Theorem 2.2: Let $\overline{P}$ be the bundle defined just prior to the statement of Theorem 2.2 in Sec. 2 and let $\{s_\alpha\}$ denote a family of local gauges in $\overline{P}$ whose domains cover M. For each α let $\bar{s}_\alpha(x) = (x\{[s_\alpha(x),r_i]\})$ for each x in the domain of s_α. It follows that $\{\bar{s}_\alpha\}$ is a family of local gauges in $\mathscr{F}E$ having the same transition functions as the $\{s_\alpha\}$. It follows easily that $\overline{P}$ is bundle isomorphic to $\mathscr{F}E$ and thus that P can be identified with a subbundle of $\mathscr{F}E$. ■

Proof of Theorem 2.4: Let $\tilde{\tau}:P\to\mathscr{B}^*(V)$ be a ρ-conformal Higgs field. It follows from the Remark preceding the statement of Theorem 2.4 that there exists an open cover $\{U_\alpha\}$ of M on which there is defined a family $\{s_\alpha\}$ of gauges of P and a family of metric Higgs fields $\tau_\alpha:(P\,|\,U_\alpha)\to\mathscr{B}(V)$ such that

$$\tilde{\tau}(s_\alpha(x)) = [\rho] \quad\text{and}\quad \tilde{\tau}(u) = [\tau_\alpha(u)].$$

Thus for α,β such that $U_\alpha\cap U_\beta\neq\emptyset$ we know that τ_β is a multiple of τ_α and since both τ_α and τ_β are Higgs fields there exist smooth functions $f_{\alpha\beta}:U_\alpha\cap U_\beta\to{}^+\mathbb{R}$ such that $\tau_\beta(u) = f_{\alpha\beta}(\pi(u))\tau_\alpha(u)$. If γ_α is the fiber metric on $E\,|\,U_\alpha$ induced by τ_α then $\gamma_\beta(x) = f_{\alpha\beta}(x)\gamma_\alpha(x)$ for all $x\in U_\alpha\cap U_\beta$. If we now define $\tilde{\gamma}$ on E by $\tilde{\gamma}(x) = [\gamma_\alpha(x)]$ where α is any index such that $x\in U_\alpha$ then it follows that $\tilde{\gamma}$ is a conformal structure on E. We also know that there exist smooth functions $c_\alpha:U_\alpha\to{}^+\mathbb{R}$ such that $\tau_\alpha(s_\alpha(x)) = c_\alpha(x)\rho$. If we define $X_{\alpha i}(x) = \sqrt{c_\alpha(x)}\,[s_\alpha(x),r_i]$ it follows that:

$$\gamma_\alpha(x)(X_{\alpha i}(x),X_{\alpha j}(x)) = \rho(r_i,r_j)$$

and thus $\tilde{\gamma}$ is a ρ-conformal structure on E.

Notice that if τ is a ρ-metric Higgs field and $f:M\to\mathrm{Gl}(n,K)$ is an arbitrary smooth map some of whose values fall outside of G then $\hat{\tau}(u) = f(\pi(u))\tau(u)$ is a metric Higgs field which is not a ρ-metric Higgs field. Its corresponding fiber metric $\hat{\gamma}$ is not a ρ-fiber metric. Thus if we define $\tilde{\gamma}$ by $\tilde{\gamma}(x) = [\hat{\gamma}(x)]$ then $\tilde{\gamma}$ is a ρ-conformal structure on E which is induced by no ρ-conformal Higgs field.

Finally if $G = \mathrm{Gl}(n,K)$ and $\tilde{\gamma}$ is a ρ-conformal structure on E then there is an open cover $\{U_\alpha\}$ of M and ρ-fiber metrics γ_α on $E\,|\,U_\alpha$ such that $\tilde{\gamma}(x) = [\gamma_\alpha(x)]$ for each $x\in U_\alpha$. Theorem 2.3 implies that there exists a ρ-metric Higgs field $\tau_\alpha:(P\,|\,U_\alpha)\to\mathscr{B}(V)$ which induces γ_α. If we define $\tilde{\tau}$ by $\tilde{\tau}(u) = [\tau_\alpha(u)]$ for $u\in(P\,|\,U_\alpha)$ then it follows that $\tilde{\tau}$ is a well-defined ρ-conformal Higgs field on P which induces $\tilde{\gamma}$. ■

Proof of Theorem 2.5: It is obvious that ρ-metric Higgs fields induce ρ-conformal Higgs fields. The converse follows from Remark 1 following Theorem 2.3 along with Theorems 5.6 and 5.7 of Ref. 6. ■

Proof of Corollary 2.6: If γ is a ρ-fiber metric on E then clearly $\tilde{\gamma}(x) := [\gamma(x)]$ is a ρ-conformal structure on E. Conversely assume that $\tilde{\gamma}$ is a ρ-conformal structure on E. Since E can be identified with the bundle associated to the $\mathrm{Gl}(n,K)$ bundle $\overline{P}$ it follows from Theorem 2.4 that there exists a unique ρ-conformal Higgs field $\tilde{\tau}$ on $\overline{P}$ which induces $\tilde{\gamma}$. It follows from Theorem 2.5 that there exists a ρ-metric Higgs field on $\overline{P}$ and thus that there exists a ρ-fiber metric on E.

Proof of Lemma 3.2: It suffices to show that if X and Y are vector fields on V which are everywhere tangent to orbits of G then $[X,Y]$ is everywhere tangent to orbits of G. To see this let $w\in V_i$ and let $A_1,A_2,\ldots,A_r$ generate a complement of the Lie algebra $\mathfrak{h}$ of $H = G_w$ in $\mathfrak{g}$. On a neighborhood of w, $X = \Sigma_{k=1}^r f_k A_k^*$ and $Y = \Sigma_{l=1}^r g_l A_l^*$. Since $[A_k^*,A_l^*] = [A_k,A_l]^*$ and A is everywhere tangent to orbits of G for every $A\in\mathfrak{g}$, it follows that $[X,Y](w)$ is tangent to $G\cdot w$ at w. The lemma now follows via standard arguments. ■

Proof of Theorem 3.3: The proofs of (1) and (2) are trivial consequences of the lemma and the definitions. We prove (3). Since G acts regularly on V we know that each orbit of G is regular and thus that the action of G foliates each V_i via regular leaves. By Palais, Ref. 26 page 19, V_i/G is a possibly non-Hausdorff manifold and by page 25 of the same article there is a smooth mapping $f_{\phi i}:U_i\to V_i/G$ such that $n_i\circ\phi = f_{\phi i}\circ\pi$. If $f_{\phi i}$ has constant rank then U_i is foliated by the components of the level surfaces of $f_{\phi i}$, i.e., if

$$\mathscr{U}_i \equiv \{L\,|\,L \text{ is a component of } f_{\phi i}^{-1}(p) \text{ for some } p\in V_i/G\},$$

then $\mathscr{U}_i$ is a foliation of U_i each leaf of which has dimension $\dim M = \mathrm{rank}(f_{\phi i})$. If L is any one of the leaves of $\mathscr{U}_i$ then $f_{\phi i}(L)$ is a single point of V_i/G and consequently $\phi\circ\pi$ maps $\pi^{-1}(L)$ onto a single orbit of G. Since $P\,|L = \pi^{-1}(L)$ it follows that ϕ maps $P\,|L$ onto a single orbit of G and thus $\phi\,|(P\,|L)$ is a symmetry-breaking Higgs field. ■

Proof of Theorem 3.4: Let $H = G_\xi$, $Q_\phi = \phi^{-1}(\xi)$, and $Q_\psi = \psi^{-1}(\xi)$. Recall that both Q_ϕ and Q_ψ are subbundles of P having G_ξ as structure group. We first construct an automorphism of P which carries Q_ϕ onto Q_ψ. We complete the proof via a sequence of remarks each of which has a trivial proof. We omit the details of (1) and (2).

(1) If $\psi:P\to G\cdot\xi\subseteq V$ is a Higgs field then there is a unique map $\mu_\psi:P\to N$ such that $\psi(u) = \mu_\psi(u)\xi$ for every $u\in P$. Moreover μ_ψ has the properties: $\mu_\psi(un) = n^{-1}\mu_\psi(u)$, $\mu_\psi(uh) = h^{-1}\mu_\psi(u)h$ for $u\in P$, $h\in H$, $n\in N$.

(2) If $\psi:P\to G\cdot\xi\subseteq V$ is a Higgs field then the map $\pi_\psi:P\to Q_\psi$ defined by $\pi_\psi(u) = u\cdot\mu_\psi(u)$ is a projection ($\pi_\psi^2 = \pi_\psi$) of P onto Q_ψ. Thus P is a trivial principal bundle over Q_ψ with structure group N. The inclusion $i:Q_\psi\to P$ is a global section of the bundle $\pi_\psi:P\to Q_\psi$ and π_ψ has the properties: $\pi_\psi(uh) = \pi_\psi(u)h$ and $\pi_\psi(un) = \pi_\psi(u)$ for $u\in P$, $n\in N$, $h\in H$.

(3) There is a unique equivariant automorphism $\gamma:P\to P$ such that $\psi\circ\gamma = \phi$.

To see this first observe that $\pi_\psi|Q$ carries Q_ϕ to Q_ψ since $\pi_\psi(P)\subseteq Q_\psi$. Moreover if $u\in Q_\phi$ and $h\in H$ then $\pi_\psi(uh) = \pi_\psi(u)h$ and $\pi(\pi_\psi(u)) = \pi(u\mu_\psi(u)) = \pi(u)$. Thus $\pi_\psi|Q_\phi:Q_\phi \to Q_\psi$ is a bundle isomorphism. We extend this isomorphism to an automorphism γ of P by requiring:

$$\gamma(u) = \pi_\psi(u\mu_\phi(u))\mu_\phi(u)^{-1}.$$

The remainder of the proof is an easy but tedious computation showing that $\psi\circ\gamma = \phi$ and that γ is indeed equivariant. We leave the details to the reader. ■

APPENDIX B

The singular decomposition of a vector space V introduced in Sec. 3 was tailored to the question of finding the largest bundles on which a general Higgs field is actually a symmetry-breaking Higgs field. The singular decomposition grouped together G-orbits of the same dimension. Another method of decomposing V is to write V as the union of strata, as introduced by Michel and Radicati.[16] A stratum in V is the union of all G-orbits that have, up to conjugation, the same isotropy subgroup. That these two decompositions are distinct and that the decomposition into strata is generally finer is illustrated by the following example.

Let $V = \mathbb{R}^3$ and let G be the identity component of SO(1,2). On V the inner product $x\cdot y = x^1y^1 - x^2y^2 - x^3y^3$ is G-invariant and defines a null cone.

It is easily seen that there are six classes of orbits and they may be characterized geometrically as follows: the upper and lower naps of the null cone, the upper and lower halves of hyperboloids of two sheets lying inside the null cone, hyperboloids of one sheet lying outside the null cone, and the origin. All these orbits except the orbit consisting of the origin alone are two dimensional. The singular decomposition thus takes the form $V = U\cup\Sigma$ where $\Sigma = \{0\}$ and U is the union of all the other two-dimensional orbits.

One can also easily show that there are four strata for this case: S_1 composed of the upper and lower naps of the null cone, S_2 composed of orbits that are upper and lower halves of hyperboloids of two sheets, S_3 composed of orbits that are hyperboloids of one sheet, and $S_4 = \{0\}$. Thus V decomposes into four strata as $V = S_1\cup S_2\cup S_3\cup S_4$. Hence the single nontrivial component U in the singular decomposition is the union of three strata $U = S_1\cup S_2\cup S_3$.

APPENDIX C

Let $P(M,G)$ be an arbitrary principal bundle over a manifold M with a metric g. Recall that if $u\in P$ then a k-form α on T_uP is said to be *horizontal* if $\alpha(X_1,\dots,X_k) = 0$ whenever at least one of the vectors $X_1,X_2,\dots,X_k$ is vertical. Let H^k_uP denote the set of all horizontal k-forms at u. Observe that if $\alpha\in H^k_uP$, $\beta\in H^l_uP$ then $\alpha\wedge\beta\in H^{k+l}_uP$. If $n = \dim M$ and $\alpha\in H^k_uP$, $k\leqslant n$, we define an n-k form $\star\alpha$ on T_uP by

$$\star\alpha(X_1,X_2,\dots,X_{n-k}) = \star(s^*\alpha)(d\pi(X_1),\dots,d\pi(X_{n-k})),$$

where $s:U\to P$ is any local section of P such that $s(\pi(u)) = u$ and $\star(s^*\alpha)$ is the usual Hodge dual of the k-form $(s^*\alpha)$ with respect to the metric g. Observe that if $\bar{s}$ is any other section of P through $x_0 = \pi(u)$ then there is a G-valued function h on a neighborhood U of x_0 in M such that $\bar{s}(x) = s(x)h(x)$ for all $x\in U$. Since $h(x_0) = e$ we have, for each vector X at $x_0 = \pi(u)$

$$d_{x_0}\bar{s}(X) = d_{x_0}s(X) + V$$

for some vertical vector V. Thus, for $X_1,X_2,\dots,X_k$ in $T_{x_0}M$

$$(\bar{s}^*\alpha)(X_1,\dots,X_k) = (s^*\alpha)(X_1,\dots,X_k)$$

and $(s^*\alpha)_{x_0} = (\bar{s}^*\alpha)_{x_0}$. It follows that $\star(s^*\alpha)_{x_0} = \star(\bar{s}^*\alpha)_{x_0}$ and that $\star\alpha$ is well-defined. It is clearly horizontal; thus $\star\alpha\in H^{n-k}_uP$. One consequence of these remarks is that if α,β belong to H^k_uP then $\star\alpha\wedge\beta\in H^n_uP$. Since $\dim H^n_uP = 1$ we have that $\star\alpha\wedge\beta$ is a multiple of $(\pi^*\eta)_u$ where η is the volume on M induced by the metric g on M. Define a map $\tilde{g}_u:H^1_uP\times H^1_uP\to\mathbb{R}$ by

$$\star\alpha\wedge\beta = \tfrac{1}{4}\tilde{g}_u(\alpha,\beta)(\pi^*\eta)_u.$$

It is clear that $\tilde{g}_u$ is bilinear. We claim that it is also nondegenerate and symmetric. To see this let s be any local section of P and let α and β be arbitrary 1-form fields defined on all of P. Then, for $x\in M$,

$$\begin{aligned} s^*(\star\alpha\wedge\beta)_x &= \tfrac{1}{4}\tilde{g}_{s(x)}(\alpha_{s(x)},\beta_{s(x)})s^*(\pi^*\eta)_x \\ &= \tfrac{1}{4}\tilde{g}_{s(x)}(\alpha_{s(x)},\beta_{s(x)})\eta_x. \end{aligned}$$

But it is known that if a and b are 1-forms on a spacetime M then

$$(\star a\wedge b)_x = \tfrac{1}{4}g_x(a_x,b_x)\eta_x$$

(this fact is not difficult to prove from first principles). It follows that

$$s^*(\star\alpha\wedge\beta)_x = \star(s^*\alpha)\wedge(s^*\beta) = \tfrac{1}{4}g_x((s^*\alpha)_x,(s^*\beta)_x)\eta_x$$

and consequently

$$\tilde{g}_{s(x)}(\alpha_{s(x)},\beta_{s(x)}) = g_x((s^*\alpha)_x,(s^*\beta)_x).$$

Since g_x is symmetric, so is $\tilde{g}_{s(x)}$. Since s is arbitrary we see that $\tilde{g}_u$ is symmetric for each $u\in P$. Also if $u\in P$ and (x^μ) are coordinates at $\pi(u)$ then $x^1\circ\pi$, $x^2\circ\pi,\dots,x^n\circ\pi$ are n-coordinates of a chart of P at u. Thus $d_u(x^\mu\circ\pi)$ is in H^1_uP for each μ. We have

$$\tilde{g}_u(d_u(x^\mu\circ\pi),d_u(x^\nu\circ\pi)) = g_{\pi(u)}(d_{\pi(u)}x^\mu,d_{\pi(u)}x^\nu) := g^{\mu\nu}$$

and consequently $\tilde{g}_u$ is nondegenerate. It follows that $\tilde{g}$ is a fiber metric on the subbundle

$$H^1P = \bigcup_{u\in P}(H^1_uP)$$

of T^*P.

If now $\phi:TP\to V$ is a vector-valued tensorial form on P and if $\{\phi^a\}$ are the components of ϕ in some basis $\{r_a\}$ of V then, for each a, ϕ^a is a horizontal 1-form on P so that

$$\star\phi^a\wedge\phi^b = \tfrac{1}{4}\tilde{g}(\phi^a,\phi^b)(\pi^*\eta) = \star\phi^b\wedge\phi^a. \tag{C1}$$

[1] D. Bleecker, *Gauge Theory and Variational Principles* (Addison-Wesley, Reading, MA, 1981).

[2] A. Trautman, Rep. Math. Phys. **1**, 29 (1970); M. E. Mayer and A. Trautman, Acta. Phys. Austriaca, Suppl. **XXIII**, 433–476 (1981).

[3] M. Daniel and C. M. Viallet, Rev. Mod. Phys. **52**, 175 (1980).

[4] This question has been raised in different forms by other authors. In particular Weisskopf [Physics Today **34** (11), 69 (1981)] has suggested the desirability of eliminating the Higgs field from present theory while retaining its effects.

[5]By a $U(n)$-type gauge theory we mean a gauge theory whose group is the subgroup $U(n)$ of $Gl(n,k)$ which leaves some metric ρ on K^n invariant. This metric ρ need not be positive definite. Consequently noncompact groups $O(m,n)$ as well as compact groups $O(n)$ are $U(n)$-type groups. Thus $U(n)$-type gauge theories implicitly define a metrical structure on their internal spaces. O'Raifeartaigh [Rep. Prog. Phys. **42**, 159 (1979)] has pointed out that, up to the present, gauge theories other than gravitational theories do not have a metrical substructure. By extending $U(n)$-type gauge theories to conformal $U(n)$ theories we are taking a step toward developing a metrical substructure on gauge theories of type $U(n)$.

[6]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry I* (Interscience, New York and London, 1963).

[7]If ρ is a representation of a Lie group G on a vector space V then a *pseudotensorial* p-form on P of type (ρ,V) is a V-valued p-form ψ on P satisfying $R_g{}^*\psi=\rho(g^{-1})\psi$. Here $R_g{}^*$ is the pull-back map on forms on P induced by the right "translation" map $R_g:P\to P$ defined by $R_g\xi=\xi g$. A p-form ψ that is pseudotensorial of type (ρ,V) and which in addition satisfies $\psi(\Delta_1,\ldots,\Delta_p)=0$ if any one of the vectors $\Delta_1,\ldots,\Delta_p$ is vertical is termed a tensorial p-form of type (ρ,V). For a more complete discussion of pseudotensorial forms see Ref. 6. Where no confusion would arise we will refer to a pseudotensorial p-form of type (ρ,V) as simply a pseudotensorial p-form. Moreover for brevity we will, following Trautman,[9] often refer to a tensorial zero-form as a Higgs field. The exterior covariant derivative of a pseudotensorial form ψ, denoted $D\psi$, is defined by $D\psi=\mathrm{hor}(d\psi)$. Note in particular that a connection 1-form ω is pseudotensorial while the curvature of ω, $\Omega\equiv D\omega$, is a tensorial 2-form.

[8]In spontaneously broken gauge theories one reparameterizes the Higgs field in terms of the vacuum state. If ξ_0 is chosen as the vacuum state then the first-order approximation $(1+g)\Omega\xi_0$ of the field $\phi=f\Omega\xi_0$, $f=e^g$, is the bundle version of the reparameterized Higgs field as it occurs in standard models (see Abers and Lee[11]).

[9]A. Trautman, Czech. J. Phys. B **29**, 107 (1979).

[10]J. A. Schouten, *Ricci Calculus* (Springer-Verlag, Berlin, Heidelberg, Gottingen, 1954).

[11]E. S. Abers and B. W. Lee, Phys. Rep. **9**, 1 (1973).

[12]L. K. Norris, R. O. Fulp, and W. R. Davis, Phys. Lett. A **79**, 278 (1980).

[13]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U.P., New York, 1973).

[14]C. J. Isham, Phys. Lett. B **102**, 251 (1981).

[15]M. Golubitsky and V. Guillemin, *Stable Mappings and their Singularities* (Springer-Verlag, New York, Heidelberg, Berlin, 1973).

[16]L. Michel and L. A. Radicati, Ann. Phys. **66**, 758 (1971).

[17]A foliation of a manifold N is *regular* if each point of N is contained in a neighborhood U of N such that for each leaf L of the foliation, $U\cap L$ is connected. If a Lie group G acts on a manifold F then we say that the action of G on F is *regular* if each point of F is contained in a neighborhood U of F such that for each orbit $\mathscr{O}$ of G in F it follows that $U\cap\mathscr{O}$ is connected.

[18]F. W. Hehl, P. von der Heyde, G. D. Kerlick, and J. M. Nester, Rev. Mod. Phys. **48**, 393 (1976).

[19]H. Weyl, Sitzungsber. Preuss. Akad. Wiss. **198**, p. 485; Ann. Phys. (Leipzig) **59**, 101 (1919).

[20]It should be noted that the Lie algebras of $U(2)$ and $SU(2)\times U(1)$ are identical and consequently few of the detailed mass calculations differ as one changes from $SU(2)\times U(1)$ to $U(2)$.

[21]We point out that the "vacuum vector' v was chosen, somewhat arbitrarily, as $\binom{0}{1}\in\mathbb{C}^2$. Clearly any vector on the unit sphere in $\mathbb{C}^2$ would do as well. The specific form of the subgroup $U_c(1)$ would then reflect our choice of v. As remarked earlier $U_c(1)=G_v$ where $G_v=\{g\in G\,|\,g\cdot v=v\}$ is the isotropy subgroup of v in G. If we chose $w=g\cdot v$ as the vacuum vector rather than v, then it is not difficult to show that the Lie algebra $\mathfrak{g}_v$ of G_v changes to $\mathfrak{g}_w=\mathrm{Ad}(g^{-1})\mathfrak{g}_v$. It follows from this equation that if $A\in\mathfrak{g}_v$ generates $\mathfrak{g}_v$ then $\mathfrak{g}_w$ is generated by $\mathrm{Ad}(g^{-1})A$.

[22]The details of the definition of the dual of a horizontal form on a principal fiber bundle are discussed in Appendix C.

[23]Recall from Theorem 5.5 that $0=-\frac{1}{2}\mu_\sigma=\omega_{\mathbf{R}}+d\phi/\phi$ for some ${}^+\mathbf{R}$-valued Higgs field ϕ. But this implies that $d\xi+(d\phi/\phi)=0$ or that $e^{-\xi}=k\phi$ for some constant k. Thus $e^{-\xi}$ is a Higgs field.

[24]Observe that the connection ω reduces to the $CU(2)$ subbundle $P_{\bar{\tau}}$ but that χ is not a Higgs field on $P_{\bar{\tau}}$ with respect to the usual action of $CU(2)\subseteq Gl(2,\mathbb{C})$ on $\mathbb{C}^2$. On the other hand χ is a Higgs field on $P_{\bar{\tau}}$ with respect to the action of $CU(2)$ on $\mathbb{C}^2$ defined by

$$(cA)\cdot w=Aw$$

for $c\in{}^+\mathbf{R}$, $A\in U(2)$, $w\in\mathbb{C}^2$. Thus

$$D\chi=d\chi+\omega\cdot\chi=d\chi+\omega_U\cdot\chi+\omega_{\mathbf{R}}\cdot\chi,$$

where $\omega_{\mathbf{R}}\cdot\chi=0$. It follows that

$$D\chi=d\chi+\omega_U\cdot\chi.$$

[25]For another approach to the problem of geometrizing the Weinberg–Salam model see Manton [Nucl. Phys. B **158**, 141 (1979)]. Unlike our approach Manton does not eliminate the Higgs particle from the theory and, in fact, he predicts its mass.

[26]R. S. Palais, *A Global Formulation of the Lie Theory of Transformation Groups* (Memoir of AMS, No. 23, 1957).

Higgs fields as Bargmann–Wigner fields and classical symmetry breaking[a)]

A. F. Furtado do Amaral, F. A. Doria, and M. Gleiser

Instituto de Física, Universidade Federal do Rio de Janeiro, C.P. 68528, Rio de Janeiro 21944 RJ, Brazil

(Received 3 August 1982; accepted for publication 22 October 1982)

We show that Higgs fields can be written as Bargmann–Wigner–Teitler spin-0 fields. When they interact with a gauge field, the usual inconsistency in the interaction equations is interpreted as Nambu's algebraic condition for nontrivial topological properties in the coupled system.

PACS numbers: 11.15.Ex, 11.30.Jw, 11.30.Qc

1. INTRODUCTION

Nambu[1] has recently discussed topologically nontrivial field configurations that arise out of solutions for the SU(2) [or SO(3)] field equations. In his paper Nambu notices that for a SU(2) field $\mathscr{F}_{\mu\nu}$ coupled in the usual manner to a Higgs field $\boldsymbol{\phi}$ in the adjoint representation together with the conditions

$$\boldsymbol{\phi}\cdot\boldsymbol{\phi} = 1, \tag{1.1a}$$

$$D_\mu \boldsymbol{\phi} = \partial_\mu \boldsymbol{\phi} + \mathbf{A}_\mu \times \boldsymbol{\phi} = 0, \tag{1.1b}$$

we can define a scalar field $f_{\mu\nu} = \mathscr{F}_{\mu\nu}\cdot\boldsymbol{\phi}$ that satisfies the Maxwell field equations. Moreover (1b) implies that one should have, because of consistency,

$$[D_\mu, D_\nu]\boldsymbol{\phi} = \mathbf{F}_{\mu\nu}\times\boldsymbol{\phi} = 0, \tag{1.2a}$$

together with a solution for (1b) which is

$$\mathbf{A}_\mu = \boldsymbol{\phi}\times\partial_\mu\boldsymbol{\phi} + a_\mu\boldsymbol{\phi}, \tag{1.2b}$$

where $a_\mu(x)$ is an arbitrary scalar spacetime-defined function. These conditions make it easier to analyze nontrivial topological aspects of the theory (which show up in local discontinuities in [(1.1b) and (1.2a)] and they lead to some well-known solutions such as the t'Hooft–Polyakov monopole[2]; thus they appear as physically meaningful.

We show here that Nambu conditions[1] arise very naturally when we consider the coupling between a SU(2) gauge field and a Bargmann–Wigner[3] spin-0 field which takes values in the SU(2) Lie algebra. Also as a consequence of our analysis we see that the Bargmann–Wigner SU(2) field—which at first sight could be seen only as an alternative way of introducing the Higgs field—breaks in a classical way the SU(2) symmetry down to U(1) without the need for a quartic self-interacting potential. After we have thus reduced the theory's symmetry, we can redefine the fields in such a way that while the U(1) gauge field is kept unchanged, the Bargmann–Wigner–Higgs field is (locally at least) gauged away. Obstructions to a global gauging away of the Higgs field are of a topological nature and they are identified to Nambu's vortex configurations. Algebraic pathologies also arise out of the Nambu conditions, since static solutions for Eq. (1) and (2) are quasiabelian in Solomon's sense[4] and as such have copied potentials.[5]

2. THE BARGMANN–WIGNER–TEITLER SPIN-0 FIELD

The Bargmann–Wigner (BW) equations have been long considered to be plagued by inconsistencies when minimal interaction is added. We will use here Teitler's Dirac algebra formulation for the spin-0 BW equations.[6] The Bargmann–Wigner–Teitler (BWT) equations are, in a local coordinate frame in Minkowski space,

$$(\gamma^\mu\partial_\mu + m)\Psi_{(m)} = 0, \tag{2.1a}$$

$$\Psi_{(m)} = -m\phi(x) + \phi_\mu(x)\gamma^\mu. \tag{2.1b}$$

As expected we derive from the above set the usual first-order massive spin-0 equations

$$\partial_\mu\phi^\mu = m^2\phi, \tag{2.2a}$$

$$\phi_\alpha = \partial_\alpha\phi, \tag{2.2b}$$

$$\partial_\alpha\phi_\beta - \partial_\beta\phi_\alpha = 0. \tag{2.2c}$$

In the massless case the BWT equations for a spin-0 field are

$$(\gamma^\alpha\partial_\alpha)\Psi_{(0)} = 0, \tag{2.3a}$$

$$\Psi_{(0)} = \phi_\mu(x)\gamma^\mu. \tag{2.3b}$$

These equations imply

$$\partial_\mu\phi^\mu = 0, \tag{2.4a}$$

$$\partial_\alpha\phi_\beta - \partial_\beta\phi_\alpha = 0. \tag{2.4b}$$

Because of the cocycle condition (2.4b) we have (locally at least, for a topologically nontrivial flat space) $\phi_\alpha = \partial_\alpha\phi$.

When we make the usual substitution $\partial_\mu \to D_\mu = \partial_\mu + A_\mu$, an inconsistency arises since (2.2c) and (2.4b) become $D_\alpha\phi_\beta - D_\beta\phi_\alpha = 0$, while (2.2b) (or in the massless case the covariant definition $\phi_\alpha = D_\alpha\phi$ lead to

$$D_\alpha\phi_\beta - D_\beta\phi_\alpha = [D_\alpha, D_\beta]\phi = \mathscr{F}_{\alpha\beta}\phi. \tag{2.5}$$

In the usual electromagnetic case, (2.5) immediately implies that either $\mathscr{F}_{\mu\nu} = 0$ or $\phi = 0$. However if ϕ is supposed to take values in the representation space for a nonabelian semisimple gauge group G, (2.5) has nontrivial solutions and opens up a highly nontrivial possibility for the coupled gauge and spin-0 Higgs system, as we are now dealing with that situation.

3. COUPLED SU(2) AND BARGMANN–WIGNER FIELDS

We will restrict our attention to the $G = \mathrm{SU}(2)$ [or SO(3)] case with ϕ in the adjoint representation. Equation (2.5) then becomes

$$\mathbf{F}_{\mu\nu}\times\boldsymbol{\phi} = 0. \tag{3.1}$$

This algebraic condition imples that the SU(2) symmetry is broken down to U(1) in a purely classical way. Equation (3.1) means that both $\mathbf{F}_{\mu\nu}$ and $\boldsymbol{\phi}$ are aligned in isospin space. An SU(2) gauge transformation then allows us to (locally at

[a)] Partially supported by CNPq and FINEP.

least) transform[7]

$$\mathbf{F}_{\mu\nu}(x)\to\mathbf{F}'_{\mu\nu}(x)=f_{\mu\nu}(\chi)\boldsymbol{\theta}, \tag{3.2a}$$

$$\boldsymbol{\phi}(x)\to\phi'(x)=\eta(x)\boldsymbol{\theta}, \tag{3.2b}$$

where both $f_{\mu\nu}(x)$ and $\eta(x)$ are scalar-valued functions and $\boldsymbol{\theta}$ is a constant Lie algebra element. Thus $\mathbf{F}$ is reduced to a U(1) gauge field. We should now consider here[7] two cases: either (i) $\det(f_{\mu\nu}(x))\neq 0$ everywhere in Minkowski space (but for a nowhere dense set with a void interior) or (ii) $\det(f_{\mu\nu}(x))=0$ over an open $U\subset$ Minkowski spacetime. In case (i) we have[7]

$$\mathbf{F}'_{\mu\nu}(x)=\partial_\mu\mathbf{A}_\nu-\partial_\nu\mathbf{A}_\mu \tag{3.3a}$$

with

$$\mathbf{A}_\mu(x)=(a_\mu(x)+\partial_\mu\lambda(x))\boldsymbol{\theta} \tag{3.3b}$$

in the particular gauge of Eqs. (3.2) where a_μ and λ are scalar-valued functions. Now Eqs. (3.1)–(3.3) together with the definition $\boldsymbol{\phi}_\mu=D_\mu\boldsymbol{\phi}$ lead to $\boldsymbol{\phi}_\mu=\partial_\mu\boldsymbol{\phi}$. We then notice that the combination

$$\mathbf{A}_\mu{}^{(\eta)}(x)=(a_\mu(x)+\partial_\mu\eta(x))\boldsymbol{\theta} \tag{3.4}$$

is also a potential for $\mathbf{F}_{\mu\nu}$. Consequently the (possibly local) gauge map $u(x)=\exp(-\eta(x)\boldsymbol{\theta})$ allows us to gauge away the Higgs field. We will see that in full detail when we examine the Lagrangian formulation for our spin-0 system. We only notice that for $\boldsymbol{\phi}$ as in (3.2b) both (1.1b) and (1.2b) are trivialized, since that case (1.1b) entails that $\partial_\mu\eta=0$ and (1.2b) reduce to (3.3b).

If (ii) $\det(f_{\mu\nu}(x))=0$ on an open U in Minkowski space, since $f=(1/2)f_{\mu\nu}dx^\mu\wedge dx^\nu$ is simplectic and closed (for $\mathbf{F}$ satisfies the abelian Bianchi condition, and thus $df=0$), we can find a coordinate system somewhere inside U such that the two-form $\mathbf{F}=(1/2)f_{\mu\nu}\boldsymbol{\theta}dx^\mu\wedge dx^\mu$ becomes

$$\mathbf{F}=\boldsymbol{\theta}dx^1\wedge dx^2 \tag{3.5}$$

in that coordinate system, as a consequence of Darboux's theorem in simplectic geometry—provided that $f_{\mu\nu}(x)$ be real valued.[7] (In the complex case we can split $f=\mathrm{Re}\,f+i\mathrm{Im}\,f$, and get the same result for either $\mathrm{Re}\,f$ or $\mathrm{Im}\,f$. We also notice that the transformation that leads to (3.5) is in general nonlinear coordinate transformation.) An obvious potential for (3.5) is given by the form

$$\mathbf{A}=(x^1)\boldsymbol{\theta}dx^2. \tag{3.6}$$

Another potential for (3.5) is

$$\mathbf{B}=((x^1)\boldsymbol{\theta}+h(x^2)\boldsymbol{\theta}')dx^2. \tag{3.7}$$

We notice that when $\boldsymbol{\theta}\times\boldsymbol{\theta}'\neq 0$, (3.6), and (3.7) are *not* gauge equivalent, and provide us with an example of the "gauge field copy" or "Wu–Yang ambiguity" phenomenon.[8] This makes our analysis more complicated to a certain extent, since the definitions $\boldsymbol{\phi}^A{}_\mu=D_\mu(A)\boldsymbol{\phi}=\partial_\mu\boldsymbol{\phi}$ and $\boldsymbol{\phi}^B=D_\mu(B)\phi\neq\partial_\mu\boldsymbol{\phi}$ do *not* coincide here. When we take $\boldsymbol{\phi}^B{}_\mu$ into (2.4) [with $\partial_\mu\to D_\mu(B)$] we check that

$$\partial_\mu\partial^\mu\boldsymbol{\phi}=\mathbf{j}\neq 0, \tag{3.8}$$

where $\mathbf{j}$ doesn't vanish because of the noncommuting $\boldsymbol{\theta}\times\boldsymbol{\theta}'$. However, we notice that the standard holonomy group generated by the affine connection form (3.7) will not be U(1), and thus the theory's symmetry group [for (3.7)] will still be an unbroken SU(2). This will always be the case when we have a copied potential, while when the two different potentials for the same field are (locally at least) gauge equivalent the local and infinitesimal holonomy groups always coincide with the reduced symmetry group.[9]

We thus exclude copies potentials from our analysis and restrict ourselves to potentials that are gauge equivalent to (3.6). These potentials will have as components

$$\mathbf{A}'_2=((x^1)\boldsymbol{\theta}+\partial_2\lambda(x)\boldsymbol{\theta}), \tag{3.9a}$$

$$\mathbf{A}'_k=\partial_k\lambda(x)\boldsymbol{\theta},\quad k\neq 2, \tag{3.9b}$$

and again we can define a new

$$\mathbf{A}_\mu{}^\eta=\mathbf{A}_\mu+\partial_\mu\eta(x)\boldsymbol{\theta} \tag{3.10}$$

[where $\mathbf{A}$ is given by (3.6)] so that the Higgs field ϕ is (at least locally) gauged away.

In cases (i) and (ii) the Higgs field has served its habitual purpose, that of breaking down the theory's symmetry group; however, we notice that the group's reduction has been performed in a purely classical way. Equations (3.4) and (3.10) also suggest that the Higgs field is a mere artifact in the present formulation: the moment we have reduced the symmetry group, it is (locally, at least) gauged away.

The Higgs field can be derived from a Dirac-like Lagrangian density,[10]

$$\mathscr{L}_{\mathscr{H}}=\overline{\boldsymbol{\Psi}}\cdot(\gamma^\mu\partial_\mu\boldsymbol{\Psi}) \tag{3.11}$$

which in the general case is a Clifford algebra valued object invariant under the action of the corresponding spinor representation of the Lorentz group. When we turn on the gauge interaction, (3.11) becomes

$$\mathscr{L}_{\mathscr{H},\mathrm{SU}(2)}=-(1/4)\mathbf{F}_{\mu\nu}\cdot\mathbf{F}^{\mu\nu}+\overline{\boldsymbol{\Psi}}\cdot(\gamma^\mu D_\mu\boldsymbol{\Psi}). \tag{3.12}$$

It is now also a (locally) SU(2)-invariant object. Besides the usual gauge field equations we get

$$D_\mu\boldsymbol{\phi}^\mu=0, \tag{3.13a}$$

$$D_\mu\boldsymbol{\phi}_\nu-D_\mu\boldsymbol{\phi}_\mu=0, \tag{3.13b}$$

which together with $\boldsymbol{\phi}_\mu=D_\mu\boldsymbol{\phi}$ lead to the condition $\mathbf{F}_{\mu\nu}\times\boldsymbol{\phi}=0$. We can now redefine our gauge potential according to the prescription given by (3.4) or (3.10). Both equations suggest in a very obvious way that we should look at the Higgs–BWT field as the nonhomogeneous part in the potential's transformation rule. After performing such a transformation (3.12) becomes the usual U(1) Lagrangian density, as a consequence of (3.13), since that condition leads to (3.1). The Higgs field has vanished in much the same way as in the case of Goldstone bosons.

4. EZAWA–TZE VORTICES

Such a gauge transformation can be always globally performed in Minkowski space. However, if we cut some holes in Minkowski space so that its first De Rham cohomology group $D^1(M)$ becomes nontrivial,[11] and if the theory's group is SO(3) (with a nonvanishing fundamental group π_1 [SO(3)], we can show that obstructions to the construction of global gauge transformations[12] are classified the same way as Nielsen–Olesen vortices. Actually the discontinuities that appear when one tries to globalize the local gauge mappings can be identified in a natural way to the vortices.

Ezawa and Tze[13] mathematically characterize the existence of a vortex by condition (1.1b),

$$D_\mu \boldsymbol{\phi} = 0,$$

or, in its integrated form,

$$g(x_0,x_0)\boldsymbol{\phi} = (P \exp \textstyle\int \mathbf{A}_\mu \, dx^\mu)\boldsymbol{\phi} = \boldsymbol{\phi}, \tag{4.1}$$

where P denotes the path-ordered nonabelian exponentiated integral, and Γ is a closed loop $\Gamma \subset M$ encircling the vortex, with $x_0 \in \Gamma$. Such maps from all loops around a vortex into the (nonreduced) gauge group $G = \mathrm{SO}(3)$ classify (mod gauge transformations) all possible vortices in the field; they exist provided that the De Rham group $D^1(M)$ be nontrivial (which is the case when we exclude the vortex region from our spacetime, thus punching a hole in Minkowski space); we also require the nontriviality of $\pi_1(G) = \pi_1(\mathrm{SO}(3))$, which is the case here.[14]

The existence of a nontrivial De Rham group D^1 entails a discontinuity that may block the extension of local gauge transformations that gauge away the Higgs field to the whole spacetime manifold. This has been proved elsewhere[15] and an example will make things clear.

Let ξ be a nontrivial 1-cocycle over M-{vortex} that is, if Γ encircles the vortex region,

$$\int_\Gamma \xi = n \neq 0. \tag{4.2}$$

The "Higgs field" $\boldsymbol{\phi} = \xi \boldsymbol{\theta}$ can be expressed as gradient of a scalar only locally, provided that

$$P \exp \int_\Gamma (\mathbf{A} + \boldsymbol{\phi}) \tag{4.3}$$

is mapped over a nontrivial element in $\pi_1((\mathrm{SO}(3)), C, 1)$, where C is the U(1) subgroup of SO(3) that stabilizes $\boldsymbol{\theta}$ and $\mathbf{A}$ is the (reduced) potential for $\mathbf{F}$.

5. CONCLUSION

We finally notice that since Eq. (3.5) is valid for any field in case (ii) above, there will be (noninertial) local frames where all static real U(1) fields look the same. Such frames are in general different for different fields, and different fields may wildly differ in their global properties, but they will always look locally the same. We do not have a clear interpretation for this phenomenon.

[1] Y. Nambu, Physica A **96**, 89–98 (1979).
[2] G. t'Hooft, Nucl. Phys. B **79**, 276–284 (1974); A. M. Polyakov, JETP Lett. **20**, 194–196 (1974).
[3] V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. **34**, 211–223 (1948).
[4] S. Solomon, Nucl. Phys. B **147**, 174–188 (1979); F. A. Doria, J. Math. Phys. **22**, 2943–2951 (1981).
[5] F. A. Doria, J. Math. Phys. **22**, 2943–2951 (1981). For a review of the literature on the problem see F. A. Doria, Commun. Math. Phys. **79**, 435–456 (1981).
[6] F. A. Doria, J. Math. Phys. **18**, 564–570 (1977); Let. Nuovo Cimento **18**, 37–40 (1977).
[7] Reference 4.
[8] Reference 5.
[9] F. A. Doria, "Gauge field copies and holonomy groups," Preprint IF-UFRJ (1982).
[10] Reference 6.
[11] On this see F. A. Doria and S. M. Abrahão, J. Math. Phys. **19**, 1650–1653 (1978).
[12] F. A. Doria, Commun. Math. Phys. **79**, 435–456 (1981).
[13] Z. F. Ezawa and H. C. Tze, J. Math. Phys. **17**, 2228–2232 (1976).
[14] See Refs. 12 and 13.
[15] Reference 12, Proposition 3.22.

Application of Benson's inequalities to the atomic electronic density

Frederick W. King
Department of Chemistry, University of Wisconsin-Eau Claire, Eau Claire, Wisconsin 54701

(Received 13 September 1982; accepted for publication 18 February 1983)

In this work, several upper bound estimates for the atomic electronic density are derived by making use of Benson's inequalities. In some cases, it has been possible to compare the results obtained using Benson's inequalities with some bounds recently derived by other workers.

PACS numbers: 31.20. – d, 31.15. + q

I. INTRODUCTION

Some time ago, Benson[1] proposed an elementary method which allowed a number of classical inequalities[2–4] to be derived. The basis of the method is as follows. If $P(u, x) > 0$ and $P(u,x)$ and $G(u,x)$ are continuously differentiable for x in $[a, b]$, then

$$P(u' - G_u P^{-1})^2 \geqslant 0, \tag{1}$$

$$P(u' - G_u P^{-1})^2 + 2u'G_u + 2G_x \geqslant \frac{d(2G)}{dx}, \tag{2}$$

$$\int_a^b [P(u')^2 + P^{-1}G_u^2 + 2G_x]dx \geqslant 2G(u(b),b) - 2G(u(a),a). \tag{3}$$

A subscript indicates the appropriate partial derivative. Despite the elementary nature of the above sequence of equations, Benson showed by judicious choice of the functions P and G that many interesting inequalities could be obtained.

A special case of Eq. (3) given by Benson is

$$P(u,x) = p(x); \quad G(u,x) = \tfrac{1}{2}u^2 g(x)p(x),$$

which leads to the result

$$\int_a^b [p(x)(u')^2 + \{p(x)g(x)^2 + (p(x)g(x))'\}u(x)^2]dx \geqslant u(b)^2 p(b)g(b) - u(a)^2 p(a)g(a). \tag{4}$$

In the remainder of this paper, a simplified form of Eq. (4) will be utilized, namely,

$$\int_a^b [(u(x)')^2 + \{g(x)^2 + g(x)'\}u(x)^2]dx \geqslant u(b)^2 g(b) - u(a)^2 g(a). \tag{5}$$

The main advantage of the above approach is that it provides a very straightforward approach to deriving bounds for the function u, given information on certain integrals involving u^2 and $(u')^2$. In some instances, however, the approach of Benson does not lead to the sharpest possible inequalities. This particularly appears to be the situation if additional information is known about the function $u(x)$. This point will be discussed further in the next section.

II. THEORY

In the present work, our interest is centered on the determination of bounds for the atomic electronic density. This topic has been the subject of recent interest,[5–12] particularly the determination of bounds for the asymptotic behavior. Considering the central role played by the electronic density in discussions of the static and dynamic behavior of matter, it is obviously very useful to know rigorous bounds for this fundamental quantity.

The following discussion will focus on the application of Eq. (5) to the electronic density for seven simple cases. The first couple of choices are selected in order to compare the resulting bounds with previous investigations. The last couple of cases examined are attempts to provide very sharp bounds for the electronic density.

The electronic density, which we will assume throughout to be radially symmetric, is defined by

$$\rho(\mathbf{r}) = N\int |\Psi(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)|^2 d\mathbf{r}_2 d\mathbf{r}_3 \ldots d\mathbf{r}_N,$$

where N is the number of electrons. Our results are restricted to atomic systems.

Case 1: $g = k$ (k is a constant).

If we set $g = k$, then the basic Benson inequality (5) becomes, on setting $a = 0$ and $u(a) = 0$,

$$ku(r_b)^2 \leqslant \int_0^{r_b} [u(x)']^2 dx + k^2 \int_0^{r_b} u(x)^2 dx. \tag{6}$$

If the optimum k is selected, then

$$u(r_b)^2 \leqslant 2\left[\int_0^{r_b} [u(x)']^2 dx \int_0^{r_b} u(x)^2 dx\right]^{1/2}. \tag{7}$$

which may be rewritten as

$$u(r)^2 < 2\left[\int_0^{\infty} [u(x)']^2 dx \int_0^{\infty} u(x)^2 dx\right]^{1/2}. \tag{8}$$

This special case of Benson's inequality has been known for some time.[13] The special case of Eq. (8) for $r = 0$ can be found in the book by Hardy *et al.*[2]

Employing the substitution

$$u(r) = r\rho(r)^{1/2} \tag{9}$$

and making use of the inequality, derived by Hoffmann-Ostenhof *et al.*,[5]

$$\int_0^{\infty} ([r\rho(r)^{1/2}]')^2 dr \leqslant \frac{T}{2\pi} \tag{10}$$

leads to the result

$$\rho(r) < (2NT)^{1/2}/2\pi r^2. \tag{11}$$

In Eq. (10) T is the total kinetic energy of the system. A slightly stronger result has been given by Hoffmann-Ostenhof *et al.*[5]:

$$\rho(r) \leqslant (2NT)^{1/2}/4\pi r. \tag{12}$$

For Eq. (8) we assume that $u(r)$ is continuous. No other as-

 0022-2488/83/071891-05$02.50

sumptions on the function $u(r)$ are employed, although the additional requirement that the integrals appearing in Eq. (8) converge will be employed; otherwise the result is rather trivial. The stronger result obtained by Hoffmann-Ostenhof, Eq. (12), follows from the inequality

$$|u(r)|^2 \leqslant \left[\int_0^\infty [u(x)']^2\,dx \int_0^\infty u(x)^2\,dx\right]^{1/2}, \tag{13}$$

which requires the hypothesis $u(0)=0$. This additional constraint allows the sharper inequality to be obtained, and can be observed to follow from a particular case of Block's inequalities[14,15]:

$$|u(r)|^2 \leqslant \frac{\tanh k(b-a)}{2k}\int_a^b \{[u(x)']^2 + k^2 u(x)^2\}dx, \tag{14}$$

which represents the uniform bound of the more general form of one of Block's inequalities:

$$|u(r)|^2 \leqslant \frac{\sinh k(r-a)\sinh k(b-r)}{k\sinh k(b-a)}\int_a^b \{[u(x)']^2 + k^2 u(x)^2\}dx. \tag{15}$$

Equations (14) and (15) are derived under the hypothesis

$$u(a) = u(b) = 0. \tag{16}$$

If, in place of Eq. (9), the substitution

$$u(r) = \rho(r)^{1/2} \tag{17}$$

is employed in Eq. (8), then bounds for the electronic density at the nucleus may be obtained, and these have been discussed elsewhere.[11] We note in passing that even for the case where $u(0)=0$ is not assumed, the inequality that follows from Benson's Eq. (5), i.e., Eq. (6), can be given in slightly sharper form:

$$ku(r)^2 \leqslant \tfrac{1}{2}(1+e^{-2kr})\int_0^\infty \{[u(x)']^2 + k^2u(x)^2\}dx, \tag{18}$$

which follows from an inequality of Block:

$$|u(r)|^2 \leqslant \frac{\cosh k(b-r)\cosh k(r-a)}{k\sinh k(b-a)}\int_a^b \{[u(x)']^2 + k^2u(x)^2\}dx. \tag{19}$$

Case 2: $g' + g^2 = k^2 e^{-2\Gamma x}$ (Γ is a positive constant). The differential equation to be solved is

$$g(x)' + g(x)^2 = k^2 e^{-2\Gamma x}. \tag{20}$$

The standard approach to handle a differential equation of this form is to employ the substitution

$$g(x) = v(x)'/v(x). \tag{21}$$

Using Eq. (21), Eq. (20) is converted to

$$v(x)'' - k^2 e^{-2\Gamma x} v(x) = 0. \tag{22}$$

The change of variable $y(x) = k^2 e^{-2\Gamma x}$ converts Eq. (22) into a modified Bessel differential equation. The solution of Eq. (22) is (in terms of constants c_1 and c_2)

$$v(x) = c_1 I_0(\gamma e^{-\Gamma x}) + c_2 K_0(\gamma e^{-\Gamma x}), \tag{23}$$

where $\gamma = k/\Gamma$ and I_0 and K_0 are modified Bessel functions of the first and second kind, respectively. The constant c_2 must be zero if $v(x)$ is finite at $x\to\infty$. From Eq. (23) we have that

$$g(x) = -ke^{-\Gamma x} I_1(\gamma e^{-\Gamma x})/I_0(\gamma e^{-\Gamma x}). \tag{24}$$

From Benson's inequality, Eq. (5),

$$\frac{u(r_a)^2 k e^{-\Gamma r_a} I_1(\gamma e^{-\Gamma r_a})}{I_0(\gamma e^{-\Gamma r_a})} < \int_{r_a}^\infty \{[u(x)']^2 + k^2 e^{-2\Gamma x}u(x)^2\}dx. \tag{25}$$

If we employ Eq. (9) and take advantage of the fact that the integrand in Eq. (25) is always positive, then

$$\rho(r) < \frac{I_0(\gamma e^{-\Gamma r})[2T + k^2 N\langle\Psi|e^{-2\Gamma r_1}|\Psi\rangle]}{4\pi r^2 k e^{-\Gamma r} I_1(\gamma e^{-\Gamma r})}. \tag{26}$$

From the asymptotic expansions for the modified Bessel functions of the first kind[16]:

$$I_0(z) \sim \frac{e^z}{(2\pi z)^{1/2}}\left[1 + \frac{1}{8z} + \frac{9}{128z^2} + \cdots\right] \quad z\to\infty, \tag{27}$$

$$I_1(z) \sim \frac{e^z}{(2\pi z)^{1/2}}\left[1 - \frac{3}{8z} - \frac{15}{128z^2} + \cdots\right] \quad z\to\infty, \tag{28}$$

it follows that Eq. (26) reduces to Eq. (11) in the limit $\Gamma\to 0$ when the optimum k is employed.

Hoffmann-Ostenhof *et al.*[5] have considered the problem of deriving bounds for expectation values involving exponential functions. Here we consider a different approach utilizing Sobolev's inequality.[17,18] Our bounds are restricted to expectation values of exponentially decreasing functions, and will allow us to express the expectation values in Eq. (26) in terms of the kinetic energy.

Using the Holder inequality, we have

$$\int e^{-kr}\rho(\mathbf{r})d\mathbf{r} \leqslant \left\{\int e^{-3/2kr}\,d\mathbf{r}\right\}^{2/3}\left\{\int \rho(\mathbf{r})^3\,d\mathbf{r}\right\}^{1/3}. \tag{29}$$

Sobolev's inequality takes the form (in R^3)

$$\left\{\int \phi(\mathbf{r})^6\,d\mathbf{r}\right\}^{1/2} \leqslant c\left\{\int |\nabla\phi(\mathbf{r})|^2\,d\mathbf{r}\right\}^{3/2}, \tag{30}$$

where the constant $c = 4/3^{3/2}\pi^2$. If we substitute

$$\phi(\mathbf{r}) = \rho(\mathbf{r})^{1/2} \tag{31}$$

in Eq. (30), then

$$\int \rho(\mathbf{r})^3\,d\mathbf{r} \leqslant c^2\left\{\int |\nabla\rho(\mathbf{r})^{1/2}|^2\,d\mathbf{r}\right\}^3. \tag{32}$$

Hence,

$$\int e^{-kr}\rho(\mathbf{r})d\mathbf{r} \leqslant c^{2/3}\left\{\int e^{-3/2kr}\,d\mathbf{r}\right\}^{2/3}\left\{\int |\nabla\rho(\mathbf{r})^{1/2}|^2\,d\mathbf{r}\right\}, \tag{33}$$

which simplifies on using

$$\int |\nabla\rho(\mathbf{r})^{1/2}|^2\,d\mathbf{r} \leqslant 2T \tag{34}$$

to give

$$\int e^{-2\Gamma r}\rho(\mathbf{r})d\mathbf{r} \leqslant \frac{16T}{27\Gamma^2}\left(\frac{2}{\pi^2}\right)^{1/3}. \tag{35}$$

We now make a simple evaluation of Eq. (35), using as a reference, the obvious result

$$\int e^{-2\Gamma r}\rho(\mathbf{r})d\mathbf{r} \leqslant N. \tag{36}$$

For the hydrogen atom, $\rho(r) = e^{-2r}/\pi$, and hence

$$\int e^{-2\Gamma r}\rho(\mathbf{r})d\mathbf{r} = \frac{1}{(\Gamma+1)^3}. \tag{37}$$

The right-hand side of Eq. (35) becomes

$$\frac{8}{27\Gamma^2}\left(\frac{2}{\pi^2}\right)^{1/3} \sim 0.174\,03\,\Gamma^{-2} \text{ (in atomic units)},$$

and hence the bound in Eq. (35) is a fairly satisfactory approximation. The inequality for the case of the hydrogen atom is sharpest for $\Gamma = 2$. Comparing Eqs. (35) and (36), we find that Eq. (35) is the better bound for $\Gamma > 0.417\,17$. For a general atom, Eq. (35) is superior to Eq. (36) when

$$0.348\,069 T^{1/2}/N < \Gamma. \tag{38}$$

For He Eq. (35) is better than Eq. (36) if $\Gamma > 0.7109$ and for the Be atom, if $\Gamma > 1.130$.

Returning to Eq. (26), we may rewrite this bound using Eqs. (35) and (36):

$$\rho(r) < \frac{I_0(\gamma e^{-\Gamma r})T}{2\pi r^2 k e^{-\Gamma r} I_1(\gamma e^{-\Gamma r})}\,[1 + 0.174\,035\gamma^2], \tag{39}$$

or the alternative form,

$$\rho(r) < \frac{I_0(\gamma e^{-\Gamma r})[2T + k^2 N]}{4\pi r^2 k e^{-\Gamma r} I_1(\gamma e^{-\Gamma r})}. \tag{40}$$

The optimum bound for Eq. (40) can be obtained by examining the limit $\Gamma \rightarrow 0$, which leads to

$$\rho(r) < (1/4\pi k r^2)[2T + k^2 N].$$

Case 3: $g' + g^2 = k^2 x^2$.

With the substitution $g(x) = v(x)'/v(x)$, the equation

$$g(x)' + g(x)^2 = k^2 x^2 \tag{41}$$

is transformed into

$$x^2 v(x)'' - k^2 x^4 v(x) = 0, \tag{42}$$

for which the solution is

$$v(x) = x^{1/2}[c_1 I_{1/4}(\tfrac{1}{2}kx^2) + c_2 K_{1/4}(\tfrac{1}{2}kx^2)]. \tag{43}$$

With the requirement that $v(x)$ remains finite as $x \rightarrow \infty$, we set $c_1 = 0$; hence

$$g(x) = \frac{1}{x} - \frac{kxK_{5/4}(\frac{1}{2}kx^2)}{K_{1/4}(\frac{1}{2}kx^2)}. \tag{44}$$

Since the integrand in Benson's inequality is positive for the present case, we obtain, using Eq. (9) and (10),

$$\rho(r) < \frac{K_{1/4}(\frac{1}{2}kr^2)[2T + Nk^2\langle\Psi|r_1^2|\Psi\rangle]}{4\pi k r^3 K_{3/4}(\frac{1}{2}kr^2)}. \tag{45}$$

Case 4: $g' + g^2 = k^2/x$.

On making the substitution $g(x) = v(x)'/v(x)$, the solution of the Riccati equation

$$g(x)' + g(x)^2 = k^2/x \tag{46}$$

is, with $\beta = 2kx^{1/2}$,

$$g(x) = kI_0(\beta)/x^{1/2}I_1(\beta), \tag{47}$$

which leads to the bound

$$\rho(r) < \frac{I_1(\beta)[2T + Nk^2\langle\Psi|r_1^{-1}|\Psi\rangle]}{4\pi r^{3/2} k I_0(\beta)}. \tag{48}$$

It can be shown,[11] that a slightly sharper bound can be obtained when the additional hypothesis that $u(r)$ vanishes at $r = 0$ is employed. The resulting bound is

$$\rho(r) < \frac{I_1(\beta)[2T + Nk^2\langle\Psi|r_1^{-1}|\Psi\rangle]}{4\pi r^{3/2} k\,[I_0(\beta) + \{I_1(\beta)K_0(\beta)/K_1(\beta)\}]} \tag{49}$$

$$= \frac{I_1(\beta)K_1(\beta)[2T + Nk^2\langle\Psi|r_1^{-1}|\Psi\rangle]}{2\pi r}. \tag{49a}$$

For large values of β, the denominator of Eq. (49) behaves like $\sim 2I_0(\beta)$. Therefore, in this limit, Eq. (49) is a sharper bound by a factor of 2. In the limit $\beta \rightarrow 0$, the additional factor in the denominator approaches zero, and hence Eqs. (48) and (49) become equivalent in this limit. The superior result, Eq. (49), is a direct consequence of the additional assumption on $u(r)$.

Case 5: $g' + g^2 = k/x^2$.

The solution of the Riccati equation in this case is elementary:

$$g = m_1/x, \quad m_1 = \tfrac{1}{2} + \tfrac{1}{2}(1 + 4k)^{1/2}, \tag{50}$$

$$g = m_2/x, \quad m_2 = \tfrac{1}{2} - \tfrac{1}{2}(1 + 4k)^{1/2}. \tag{51}$$

If we employ Benson's inequality and Eq. (9), then each of Eqs. (50) and (51) leads to bound for $\rho(r)$. The sharper of the two bounds is

$$\rho(r) < \frac{[2T + kN\langle\Psi|r_1^{-2}|\Psi\rangle]}{2\pi r[(1 + 4k)^{1/2} - 1]}. \tag{52}$$

The optimum k, restricted to positive values, for Eq. (52) is

$$k = \Omega + \Omega^{1/2}, \tag{53}$$

where

$$\Omega = \frac{2T}{N\langle\Psi|r_1^{-2}|\Psi\rangle}. \tag{54}$$

It follows from the well-known inequality[19]

$$N\langle\Psi|r_1^{-2}|\Psi\rangle \leqslant 8T \tag{55}$$

that $\Omega \geqslant \frac{1}{4}$ and hence $k \geqslant \frac{3}{4}$. Equation (52) reduces to

$$\rho(r) < \frac{N\langle\Psi|r_1^{-2}|\psi\rangle[2\Omega + \Omega^{1/2}]}{2\pi r[(1 + 4\Omega + \Omega^{1/2})^{1/2} - 1]}. \tag{56}$$

Case 6: $g' + g^2 = -k/x^2$.

The constant k is positive. From Eqs. (50) and (51), it is obvious that $-\frac{1}{4}$ is the most negative factor that is possible. Benson's inequality becomes, on using Eq. (50),

$$\tfrac{1}{2}r_b\rho(r_b) - \tfrac{1}{2}r_a\rho(r_a) \leqslant \int_{r_a}^{r_b} \{[(r\rho(r)^{1/2})']^2 - \tfrac{1}{4}\rho(r)\}dr. \tag{57}$$

If we employ Eq. (10), Eq. (57) may be rewritten as

$$\int_{r_a}^{r_b} \frac{\rho(r)}{r^2}(4\pi r^2\,dr) \leqslant 8T + 8\pi[r_a\rho(r_a) - r_b\rho(r_b)], \tag{58}$$

which is a generalization (for a radially symmetric density) of the well-known result

$$\int \frac{\rho(\mathbf{r})d\mathbf{r}}{r^2} \leqslant 8T. \tag{59}$$

Since the integrand in Eq. (57) is not necessarily positive for all r, it is clearly not possible to add the terms

$$\int_0^{r_a}\{([r\rho(r)^{1/2}]')^2-\tfrac{1}{4}\rho(r)\}dr+\int_{r_b}^{\infty}\{([r\rho(r)^{1/2}]')^2-\tfrac{1}{4}\rho(r)\}dr$$

to the right-hand side of the inequality. By way of example, consider the case of the hydrogen atom, for which this point can be resolved analytically. For the hydrogen atom, we have

$$([r\rho(r)^{1/2}]')^2-\tfrac{1}{4}\rho(r)=\rho(r)(r-3/2)(r-\tfrac{1}{2}). \tag{60}$$

The integrand is positive for all $r>3/2$ for the hydrogen atom, and hence for this case,

$$\rho(r)\leqslant\frac{1}{\pi r}\,[T-\tfrac{1}{8}N\langle\Psi|r_1^{-2}|\Psi\rangle]\quad\text{for } r>3/2. \tag{61}$$

Because of the importance of the region $r=0-0.5$ a.u., where the integrand is positive [see Eq. (60)], it is straightforward to show that Eq. (61) actually holds for all r.

From Eq. (58), we have

$$\rho(r)\leqslant\frac{1}{8\pi r}\left\{8T-N\langle\Psi|r_1^{-2}|\Psi\rangle+\int_r^{\infty}4\pi\rho(x)dx\right\}. \tag{62}$$

A bound for the last integral appearing in Eq. (62) can be obtained in the following manner. If we integrate the bound[5]

$$4\pi\rho(r)<\frac{1}{r^2}\left(\frac{T}{\alpha}+\frac{1}{2}\alpha N\right)(1-e^{-2\alpha r}), \tag{63}$$

we obtain

$$4\pi\int_r^{\infty}\rho(x)\,dx<\left(\frac{T}{\alpha}+\frac{1}{2}\alpha N\right)(1-E_2(2\alpha r))r^{-1}, \tag{64}$$

where α is an arbitrary positive parameter in Eq. (63) and $E_2(z)$ is an exponential integral.[16] Hence, Eq. (62) becomes

$$\begin{aligned}\rho(r)&<1/8\pi r\\ &\times\{8T-N\langle\Psi|r_1^{-2}|\Psi\rangle\\ &+(1/2\alpha r)(2T+\alpha^2N)(1-E_2(2\alpha r))\}.\end{aligned} \tag{65}$$

Case 7: $g'+g^2=-k^2/x$.

The Riccati equation to be solved is

$$g(x)'+g(x)^2=-k^2/x. \tag{66}$$

This may be converted into the following differential equation:

$$x^2v(x)''+k^2xv(x)=0. \tag{67}$$

The solution of Eq. (67) is, with $\beta=2kx^{1/2}$,

$$v(x)=x^{1/2}[c_1J_1(\beta)+c_2Y_1(\beta)], \tag{68}$$

where J_n and Y_n are Bessel functions of the first and second kind, respectively. If $v(x)'$ is finite as $x\to 0$, then $c_2=0$. The function g is

$$g(x)=\frac{k}{x^{1/2}}\,\frac{J_0(\beta)}{J_1(\beta)}. \tag{69}$$

Because of the oscillatory nature of the Bessel functions, we impose the restriction that $\beta<2.404\,825$ [the first zero of $J_0(\beta)$], i.e.,

$$r<1.445\,79\,k^{-2}. \tag{70}$$

Benson's inequality with u given by Eq. (9) gives

$$\rho(r)<\frac{J_1(\beta)}{kr^{3/2}J_0(\beta)}\int_0^r\{([x\rho(x)^{1/2}]')^2-k^2x\rho(x)\}dx. \tag{71}$$

Using Eq. (10), Eq. (71) can be rewritten as

$$\rho(r)<\frac{J_1(\beta)}{4\pi kr^{3/2}J_0(\beta)}\left\{2T+\frac{k^2}{Z}V_{\rm en}+k^2\int_r^{\infty}\rho(x)4\pi x\,dx\right\}, \tag{72}$$

where $V_{\rm en}$ is the electron-nuclear potential energy and Z is the nuclear charge. Both Eqs. (71) and (72) require the restriction given in Eq. (70).

A question of interest is whether or not a sharper form of Eq. (72) can be formulated. This can be answered in the affirmative, at least for one-electron systems. For the hydrogen atom, we have that

$$\begin{aligned}&([r\rho(r)^{1/2}]')^2-k^2r\rho(r)\\ &=\rho(r)[r-\{1+\tfrac{1}{2}k^2+\tfrac{1}{2}k[k^2+4]^{1/2}\}]\\ &\times[r-\{1+\tfrac{1}{2}k^2-\tfrac{1}{2}k[k^2+4]^{1/2}\}].\end{aligned} \tag{73}$$

In this case, the integrand in Eq. (71) is positive for

$$r>1+\tfrac{1}{2}k^2+\tfrac{1}{2}k[k^2+4]^{1/2} \tag{74}$$

and hence,

$$\rho(r)<\frac{J_1(\beta)}{4\pi kr^{3/2}J_0(\beta)}\left[2T+\frac{k^2}{Z}V_{\rm en}\right]. \tag{75}$$

The range of r for which Eq. (75) may be applied is governed by both Eqs. (70) and (74); that is,

TABLE I. Bounds for $\rho(r)$ for the hydrogen atom.

Radial distance	Bounds for $\rho(r)$ (in atomic units)					
(atomic units)	Eq. (11)	Eq. (48)	Eq. (49)	Eq. (56)	Eq. (65)	Exact
0.1	0.1592×10^2	0.7958	0.7958	0.5872×10^1	0.2171×10^1	0.2606
0.2	0.3979×10^1	0.3979	0.3973	0.2936×10^1	0.9270	0.2134
0.3	0.1768×10^1	0.2653	0.2631	0.1957×10^1	0.5564	0.1747
0.4	0.9947	0.1989	0.1948	0.1468×10^1	0.3852	0.1430
0.5	0.6366	0.1592	0.1530	0.1174×10^1	0.2887	0.1171
0.6	0.4421	0.1326	0.1246	0.9787	0.2279	0.9587×10^{-1}
0.7	0.3248	0.1137	0.1042	0.8389	0.1865	0.7849×10^{-1}
0.8	0.2487	0.9947×10^{-1}	0.8878×10^{-1}	0.7340	0.1569	0.6427×10^{-1}
0.9	0.1965	0.8842×10^{-1}	0.7678×10^{-1}	0.6524	0.1347	0.5261×10^{-1}
1.0	0.1592	0.7958×10^{-1}	0.6722×10^{-1}	0.5872	0.1176	0.4308×10^{-1}
1.5	0.7074×10^{-1}	0.5305×10^{-1}	0.3929×10^{-1}	0.3915	0.7053×10^{-1}	0.1585×10^{-1}
2.0	0.3979×10^{-1}	0.3979×10^{-1}	0.2633×10^{-1}	0.2936	0.4970×10^{-1}	0.5830×10^{-2}

$$1 + \tfrac{1}{2}k^2 + \tfrac{1}{2}k[k^2+4]^{1/2} < r < 1.445\,79\,k^{-2}. \qquad (76)$$

As a final remark on this case, we note that it is possible to derive an upper bound for the last integral appearing in Eq. (72) using an exponentially decreasing bound given by Hoffmann-Ostenhof *et al.*[5] The bound obtained requires information on the ionization potential.

Numerical results for some of the bounds discussed in this work are presented in Table I for the hydrogen atom. More detailed applications will be presented elsewhere. The value of k in each bound formula was optimized at each value of the radial coordinate r. The best bounds range from a factor of about 1.3 too high at medium range to about a factor of $3 \sim 4$ too high at both short and long range. At very long range, all the bounds give poor estimates because of the incorrect asymptotic behavior of the bounds as $r \to \infty$.

III. CONCLUSION

In this work, we have examined the application of Benson's inequalities to obtain upper bound estimates for the atomic electronic density. The bounds derived herein do not exhibit the correct long-range asymptotic behavior; that is, they do not decay exponentially as $r \to \infty$. Also, the bounds are not finite at $r = 0$. The problem of determining a reasonable bound which is both finite at the nucleus and decays exponentially for large r, is an unresolved problem. The few bounds for the electronic density which have been previously given in the literature, become infinite at $r = 0$. The exceptions are a recent bound derived by the author[11] and a bound derived specifically for $r = 0$ by Hoffmann-Ostenhof *et al.*[6]

The bounds derived in this work are satisfactory for values of r typically in the small to moderate range. Numerical applications will be discussed elsewhere.

For the situation were $g(x)' + g(x)^2$ is negative, only a limited number of functional forms have been examined. It is possible that a more judicious selection of the functional form of $g(x)' + g(x)^2$ would result in improved bounds.

ACKNOWLEDGMENT

Acknowledgment is made to the Donors of the petroleum Research Fund, administered by the American Chemical Society, for support of this research. The author would like to thank Ken Dykema for a helpful remark.

[1]D. C. Benson, J. Math. Anal. Appl. **17**, 292 (1967).
[2]G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, 2nd ed. (Cambridge U.P., Cambridge, England, 1952).
[3]E. F. Beckenbach and R. Bellman, *Inequalities* (Springer-Verlag, Berlin, 1961).
[4]D. S. Mitrinović, *Analytic Inequalities* (Springer-Verlag, Berlin, 1970).
[5]T. Hoffmann-Ostenhof and M. Hoffmann-Ostenhof, J. Phys. B **11**, 17 (1978).
[6]M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, and W. Thirring, J. Phys. B **11**, L571 (1978).
[7]T. Hoffmann-Ostenhof, M. Hoffmann-Ostenhof, and R. Ahlrichs, Phys. Rev. A **18**, 328 (1978).
[8]T. Hoffmann-Ostenhof, J. Phys. A **12**, 1181 (1979).
[9]R. Ahlrichs, M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, and J. D. Morgan III, Phys. Rev. A **23**, 2106 (1981).
[10]T. Hoffmann-Ostenhof and J. D. Morgan III, J. Chem. Phys. **75**, 843 (1981).
[11]F. W. King, J. Chem. Phys. **78**, 2459, 3091 (1983).
[12]F. W. King and K. J. Dykema, J. Phys. B (in press).
[13]B. Sz. Nagy, Acta Sci. Math. (Szeged) **10**, 64 (1941).
[14]H. D. Block, Proc. Amer. Math. Soc. **8**, 844 (1957).
[15]R. Redheffer, J. Math. Anal. Appl. **16**, 219 (1966).
[16]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1970).
[17]G. Rosen, SIAM J. Appl. Math. **21**, 30 (1971).
[18]G. Talenti, Ann. Mat. Pura Appl. **110**, 353 (1976).
[19]R. Blau, A. R. P. Rau, and L. Spruch, Phys. Rev. A **8**, 119 (1973).

Rigorous theory of spectra and radiation for a model in quantum electrodynamics[a]

Asao Arai

Department of Mathematics, Tokyo Institute of Technology, Oh-okayama, Meguro-ku, Tokyo 152, Japan

(Received 17 February 1982; accepted for publication 10 September 1982)

We study rigorously the problem of the Lamb shift and the spontaneous emission of light in a framework of nonrelativistic quantum electrodynamics by using an exactly soluble model of a harmonic oscillator atom interacting with a quantized electromagnetic field. We show that, under the perturbation of the electromagnetic field, all the point spectra corresponding to the excited states of the unperturbed atom disappear. This means that the "energy level shifts" (Lamb shifts) of the excited states of the atom cannot be described simply in terms of shifts of point spectra. Then, we give a rigorous mathematical meaning to both formal perturbation theories for the "energy level shifts" and for the transitions of the excited states due to the spontaneous emission of light, showing that the "energy level shifts" and the "decay probabilities" of the excited states of the atom are characterized in terms of the resonance pole of the S-matrix for the photon scattering by the atom. We also discuss broken symmetry aspects and infinite mass-renormalization of the model.

PACS numbers: 32.30.Jc, 32.70.Jz, 03.65.Db, 03.70. + k

I. INTRODUCTION: DESCRIPTION OF THE MODEL AND THE MAIN RESULTS

The purpose of this paper is to give a rigorous mathematical theory of spectra and radiation for a harmonic oscillator atom interacting with a quantized electromagnetic field. The main interest lies in giving a rigorous mathematical description to the "energy level shifts" and to the decay phenomena of the excited states of the atom due to the spontaneous emission of light, the former corresponding to the Lamb shifts in quantum electrodynamics[1–5] and being usually "defined" in terms of the formal perturbation theory. Such analysis is indeed necessary and is of theoretical importance, because, in view of the finite lifetime of the excited states of real atoms due to the spontaneous emission of light, one cannot expect that the "energy level shifts" are described in terms of shifts of point spectra and, if that is the case, then the meaning of the formal perturbation theory comes into question.

In the preceding works,[6,7] which are henceforth referred to as I and II, the author studied the same problem by using a similar, but rather simple model which describes a system of a one-dimensional harmonic oscillator coupled to a quantized, massless, scalar field in three space dimensions. He proved that all the point spectra but the lowest one of the unperturbed harmonic oscillator disappear due to the interaction with the field. This means that the "energy level shifts" of the oscillator due to the perturbation of the field cannot be described simply in terms of shifts of point spectra and hence that, without any reinterpretation, the formal perturbation theory makes no sense. Then, he gave a rigorous mathematical meaning to both formal perturbation theories for the "energy level shifts" and for the transitions of the excited states due to the spontaneous emission of the bosons, showing that the "energy level shifts" and the "decay probabilities" of the excited states are characterized in terms of the resonance pole of the S-matrix for the boson scattering, which is also a complex pole associated with the Wightman distributions or τ functions. In this paper, we shall show that results similar to those of (I,II) hold for the present model as well, and further, we shall consider broken symmetry aspects and infinite mass-renormalization of the model.

Similar models were studied by many authors from various physical points of view (see, e.g., Refs. 8 and 9 and references cited there). However, it seems that so far, the rigorous mathematical treatment based on the Hamiltonian formalism has not been given. Our model is not realistic, but we hope that the study of it serves as a step towards a rigorous construction of spectral and radiation theory for realistic models (see Refs. 10 and 11 for a general mathematical framework for the problem).

We now proceed to describe the model. We shall use the Coulomb gauge in quantizing the electromagnetic field. The underlying Hilbert space $\mathscr{H}$ of state vectors for the system is defined as the tensor product of $L^2(R^3)$ and $\mathscr{F}^{\mathrm{EM}}$, the Fock space for photons in the Coulomb gauge:

$$\mathscr{H} = L^2(R^3) \otimes \mathscr{F}^{\mathrm{EM}}. \tag{1.1}$$

Let $a^{(r)}(f)$ and $a^{(r)*}(f)$, $f \in L^2(R^3)$, $r = 1,2$, be the photon annihilation and creation operators in $\mathscr{F}^{\mathrm{EM}}$, respectively; they are densely defined on $\mathscr{F}_0^{\mathrm{EM}}$, the linear subspace of the finite particle vectors in $\mathscr{F}^{\mathrm{EM}}$, and leave it invariant, satisfying

$$(a^{(r)}(f)\Psi,\Phi) = (\Psi,a^{(r)*}(\bar{f})\Phi), \tag{1.2}$$

$$\begin{aligned} [a^{(r)}(f),a^{(s)*}(g)]\,\Psi &= \delta_{rs}(\bar{f},g)_{L^2(R^3)}\Psi, \\ [a^{(r)}(f),a^{(s)}(g)]\,\Psi &= 0, \end{aligned} \tag{1.3}$$

for all Ψ, Φ in $\mathscr{F}_0^{\mathrm{EM}}$ and all f, g in $L^2(R^3)$. Let $\mathbf{e}^{(r)}(\mathbf{k}) \in R^3$, $r = 1,2$, be the polarization vectors of photon with momentum $\mathbf{k}$, which satisfy

[a] Part of this work was submitted to University of Tokyo for the degree of Doctor of Sciences.

$$\mathbf{e}^{(r)}(\mathbf{k})\cdot\mathbf{k}=0, \quad \mathbf{e}^{(r)}(\mathbf{k})\cdot\mathbf{e}^{(s)}(\mathbf{k})=\delta_{rs}, \quad \mathbf{e}^{(1)}(\mathbf{k})\times\mathbf{e}^{(2)}(\mathbf{k})=\mathbf{k}/|\mathbf{k}|. \tag{1.4}$$

Then, in the Coulomb gauge, the time zero radiation field $\mathbf{A}(f)$ and its canonical conjugate $\boldsymbol{\pi}(f)$ are given by

$$A_\mu(f)=\frac{1}{\sqrt{2}}\sum_{r=1}^{2}\{a^{(r)*}(\hat{f}e_\mu^{(r)}/\sqrt{\omega}) + a^{(r)}(\widetilde{\hat{f}e_\mu^{(r)}}/\sqrt{\omega})\}, \quad \hat{f}/\sqrt{\omega}\in L^2(R^3), \tag{1.5}$$

$$\pi_\mu(f)=\frac{i}{\sqrt{2}}\sum_{r=1}^{2}\{a^{(r)*}(\sqrt{\omega}\hat{f}e_\mu^{(r)}) - a^{(r)}(\widetilde{\sqrt{\omega}\hat{f}e_\mu^{(r)}})\}, \quad \sqrt{\omega}\hat{f}\in L^2(R^3), \quad \mu=1,2,3, \tag{1.6}$$

where $\hat{f}$ denotes the Fourier transform of f and $\tilde{g}$ is defined by

$$\tilde{g}(\mathbf{k})=g(-\mathbf{k}), \tag{1.7}$$

and $\omega(\mathbf{k})$ is the energy of a free photon with momentum $\mathbf{k}$:

$$\omega(\mathbf{k})=|\mathbf{k}|. \tag{1.8}$$

We denote the free Hamiltonian of photons by H_0^{EM}, which is a nonnegative self-adjoint operator in $\mathscr{F}^{\mathrm{EM}}$ and is symbolically written as

$$H_0^{\mathrm{EM}}=\sum_{r=1}^{2}\int d^3\mathbf{k}\,\omega(\mathbf{k})a^{(r)*}(\mathbf{k})a^{(r)}(\mathbf{k}) \tag{1.9}$$

with $a^{(r)}(\mathbf{k})$ being the symbolic notation for $a^{(r)}(f)$ given by

$$a^{(r)}(f)=\int d^3\mathbf{k}\,a^{(r)}(\mathbf{k})f(\mathbf{k}). \tag{1.10}$$

All the operators in $\mathscr{F}^{\mathrm{EM}}$ [respectively $L^2(R^3)$] have natural extensions to $\mathscr{H}$; e.g., $a^{(r)}(f)$ in $\mathscr{F}^{\mathrm{EM}}$ [respectively $-i\nabla$ in $L^2(R^3)$] is extended as $I\otimes a^{(r)}(f)$ (respectively $-i\nabla\otimes I$). The extensions will be denoted by the same notations. We shall also denote the closure of a closable operator by the same notation.

The interaction of the harmonic oscillator atom with the quantized electromagnetic field is taken to be minimal. Let $m>0$, $e\in R^1$ and $\omega_0>0$ be parameters denoting the physical mass, charge of the "electron" and the spring constant of the oscillator, respectively. Then, the total Hamiltonian H in the dipole approximation is given formally by

$$H=\frac{1}{2(m-\delta m)}:(\mathbf{p}-e\mathbf{A}(\rho))^2:+\frac{1}{2}m\omega_0^2\mathbf{q}^2+H_0^{\mathrm{EM}}, \tag{1.11}$$

with

$$\mathbf{p}=-i\frac{\partial}{\partial\mathbf{q}}, \tag{1.12}$$

where : : denotes the Wick ordering and ρ is a real-valued function which denotes the charge distribution of the electron. The quantity

$$\delta m=\frac{2}{3}e^2\int d^3\mathbf{k}\frac{|\hat{\rho}(\mathbf{k})|^2}{|\mathbf{k}|^2} \tag{1.13}$$

will be found to give the mass-renormalization. Throughout the present paper we assume, for technical simplicity, that

$$\left.\begin{array}{c}\rho \text{ is a rotation invariant function, satisfying}\\ \rho\in\mathscr{S}(R^3),\ \hat{\rho}>0,\\ \int d^3\mathbf{x}\,\rho(\mathbf{x})=1.\end{array}\right\} \tag{AI}$$

We shall denote $\hat{\rho}(\mathbf{k})$ as $\hat{\rho}(|\mathbf{k}|)$.

Remarks:

(1) For the proof of the essential self-adjointness of H (see Theorem A.2), it is sufficient to assume that $\sqrt{\omega}\hat{\rho}$ and $\hat{\rho}/\omega$ are in $L^2(R^3)$.

(2) The total Hamiltonian without the dipole approximation is defined by replacing $\hat{\rho}(\mathbf{k})$ by $\hat{\rho}(\mathbf{k})\exp(-i\,\mathbf{k}\cdot\mathbf{q})$ in the right-hand side of (1.11). We can prove the existence of its self-adjoint extension for all e and the fundamental spectral property similar to that given in Ref. 11. The use of the dipole approximation permits us to solve the Heisenberg equations of motion exactly and hence to analyze the spectrum of the total Hamiltonian in detail. However, the total Hamiltonian without the dipole approximation leads to a complicated nonlinear Heisenberg equations of motion, which may not be solved exactly.

(3) Since we consider the one electron problem, (i.e., the charge one sector), the longitudinal (static) part of the electromagnetic field due to the Coulomb gauge does not appear in our Hamiltonian. It appears only in the Hamiltonian of the charge Z-sector with $Z\geqslant 2$ which is given formally in a general form by

$$H^{(Z)}=\sum_{j=1}^{Z}\left\{\frac{1}{2(m-\delta m)}:(\mathbf{p}_j-e\mathbf{A}_\rho(\mathbf{q}_j))^2:+V(\mathbf{q}_j)\right\} + \sum_{1\leqslant j<k\leqslant Z}\frac{e^2}{4\pi|\mathbf{q}_j-\mathbf{q}_k|}+H_0^{\mathrm{EM}}, \tag{1.14}$$

where $\mathbf{p}_j$ (respectively $\mathbf{q}_j$) denotes the momentum (respectively position) operator of the jth electron and $\mathbf{A}_\rho(\mathbf{q}_j)$ is defined by replacing $\hat{\rho}(\mathbf{k})$ by $\hat{\rho}(\mathbf{k})\exp(-i\,\mathbf{k}\cdot\mathbf{q}_j)$ in $\mathbf{A}(\rho)$. The operator $V(\mathbf{q})$ is the potential in which the electrons exist. [In our case, $V(\mathbf{q})=m\omega_0^2\mathbf{q}^2/2$.] The second term of (1.14) is the longitudinal (static) part of the electromagnetic field due to the Coulomb gauge. The Hamiltonian $H^{(Z)}$ is defined in the Hilbert space $(\otimes_{\mathrm{as}}^{Z}L^2(R^3))\otimes\mathscr{F}^{\mathrm{EM}}$, where $\otimes_{\mathrm{as}}^{Z}$ denotes the Z-fold antisymmetric tensor product. It would be of great interest to analyze the Hamiltonian $H^{(Z)}$.

Our first task is to establish the self-adjointness of the total Hamiltonian. We write

$$H=\widetilde{H}+\tfrac{1}{2}m\omega_0^2\mathbf{q}^2 \tag{1.15}$$

with

$$\widetilde{H}=\frac{1}{2(m-\delta m)}:(\mathbf{p}-e\mathbf{A}(\rho))^2:+H_0^{\mathrm{EM}}. \tag{1.16}$$

Let

$$\widetilde{H}_0=\frac{1}{2m}\mathbf{p}^2+H_0^{\mathrm{EM}}, \tag{1.17}$$

$$H_0=\widetilde{H}_0+\tfrac{1}{2}m\omega_0^2\mathbf{q}^2. \tag{1.18}$$

Then, we shall prove

Theorem A.1: $\widetilde{H}$ is essentially self-adjoint on any core for $\widetilde{H}_0$. In particular, if $m>\delta m$, then $\widetilde{H}$ is self-adjoint with $D(\widetilde{H})=D(\widetilde{H}_0)$. Furthermore, if $m>\delta m$ (respectively $m<\delta m$), then $\widetilde{H}$ is bounded below (respectively not bounded below):

$$\widetilde{H}\geqslant-\frac{e^2}{2(m-\delta m)}\left|\left|\frac{\hat{\rho}}{\sqrt{\omega}}\right|\right|_0^2, \tag{1.19}$$

where $\| \ \|_0^2$ denotes the norm of $L^2(R^3)$.

Theorem A.2: The total Hamiltonian H is essentially self-adjoint on any core for H_0. In particular, if $m > \delta m$, then H is self-adjoint with $D(H) = D(\widetilde{H}) \cap D(\mathbf{q}^2)$ and is bounded below with the same lower bound as that of $\widetilde{H}$ given in (1.19). If $m < \delta m$, then H is not bounded below.

For the proof of the first half of Theorem A.1, see Ref. 11. The second half can be easily proved. We shall prove Theorem A.2 in Sec. II.

The second half of Theorem A.2 shows that, if $m < \delta m$, then the ground state of H does not exist. Therefore, the case $m < \delta m$ may be unphysical. In fact, in this case, there exists an unphysical solution to the Heisenberg equations of motion which grows exponentially as time tends to a remote future or past (see Theorem 3.1).

Remark: If $m < \delta m$, then the Hamiltonian without the dipole approximation is not bounded below, either (see the proof of Theorem A.2 in Sec. II).

The next thing to analyze is the spectrum of H. We first want to remark on a mathematical feature of the problem. The total Hamiltonian can be rewritten as

$$H = H_0 + H_{\rm I}^{(1)} + H_{\rm I}^{(2)} + R, \tag{1.20}$$

where

$$H_{\rm I}^{(1)} = -\frac{e}{m}\,\mathbf{p}\cdot\mathbf{A}(\rho), \quad H_{\rm I}^{(2)} = \frac{e^2}{2m}\ :\ \mathbf{A}(\rho)^2\ :, \tag{1.21}$$

$$R = \frac{\delta m}{m - \delta m}\left(H_{\rm I}^{(1)} + H_{\rm I}^{(2)} + \frac{\mathbf{p}^2}{2m}\right). \tag{1.22}$$

The operator H_0 is taken to be the unperturbed Hamiltonian of the system and the operator $H_{\rm I}^{(1)} + H_{\rm I}^{(2)}$ is the minimal interaction. The operator R is the mass-renormalization counter term, which, in the lowest order in e, coincides with Bethe's[1] if $\hat{\rho}(\mathbf{k}) = \chi_{[0,m]}(|\mathbf{k}|)/(2\pi)^{3/2}$ with $\chi_{[0,m]}$ the characteristic function for the interval $[0,m]$. Concerning the unperturbed Hamiltonian H_0, we note that the point spectrum $\sigma_p(H_0)$ is embedded in the continuous spectrum. In fact,

$$\sigma_p(H_0) = \{E_n^{(0)}\}, \quad \sigma(H_0) = [\tfrac{3}{2}\omega_0, \infty), \tag{1.23}$$

where $\sigma(H_0)$ denotes the spectrum of H_0 and

$$E_n^{(0)} = (n + \tfrac{3}{2})\omega_0, \quad n = 0,1,2,..., \tag{1.24}$$

which is degenerate with multiplicity $M = (n+1)(n+2)/2$. Thus, from the perturbation theoretical point of view, the analysis of H gives a problem of perturbation of point spectra embedded in a continuous spectrum, which is hard to analyze in general (see, e.g., Refs. 12–19).

In order to analyze the spectrum of H, we first construct the exact solution to the Heisenberg equations of motion. We shall do this in Sec. III. Then, in Sec. IV, we shall prove the existence (respectively absence) of the asymptotic radiation fields in the case $m > \delta m$ (respectively $m < \delta m$), obtaining the explicit form of them, which permits us to analyze the spectrum of H in detail. In Sec. V we shall prove

Theorem B:

(1) Let $m > \delta m$ and

$$E_0 = \inf \sigma(H). \tag{1.25}$$

Then,

$$\sigma(H) = \sigma_{\rm ac}(H) = [E_0, \infty), \quad \sigma_p(H) = \{E_0\}, \quad \sigma_s(H) = \varnothing,$$

where $\sigma_{\rm ac}(H)$ [respectively $\sigma_s(H)$] denotes the absolutely (respectively singular) continuous spectrum of H. The eigenvalue E_0 is simple.

(2) Let $m < \delta m$. Then,

$$\sigma(H) = \sigma_{\rm ac}(H) = R^1, \quad \sigma_p(H) = \sigma_s(H) = \varnothing.$$

Remark: As is seen from the existence of the ground state of H *without infrared cutoff* [Theorem B-(1)], no infrared problem arises in our model in contrast to models in Refs. 10, 20, and 21. This may be due to that the electron is (harmonically) bound. In fact, in the case $\omega_0 = 0$, we can prove the absence of the dressed one electron states without infrared cutoff in the Fock space.[22]

Theorem B shows that the perturbation of the electromagnetic field makes all the point spectra (but the lowest one in the case $m > \delta m$) of H_0 completely disappear. This is natural in view of the spontaneous emission of light and leads us to expect the same for the Lamb shifts in real atoms. Then, however, a problem arises: What are the "energy level shifts" that the formal perturbation theory gives? Or, in other words, what does the formal perturbation theory approximate? In Theorem C below, we shall give a solution to this problem together with that of the spontaneous emission of light.

We must begin with describing some facts related to the formal perturbation theory. Let $\Psi_{n,j}^{(0)}$, $n \geqslant 0, j = 1,...,M$, be the eigenvectors of the unperturbed Hamiltonian H_0 with the eigenvalue $E_n^{(0)}$, and $\{P^{(0)}(E)\}$ be the spectral family associated with H_0. Put

$$\epsilon_{n,j}(z) = -(\Psi_{n,j}^{(0)}, H_{\rm I}^{(1)}(H_0 - z)^{-1}(1 - P^{(0)}(E_n^{(0)}))H_{\rm I}^{(1)}\Psi_{n,j}^{(0)}), \tag{1.26}$$

which is an analytic function of z in $\mathbb{C}\setminus[\tfrac{3}{2}\omega_0, \infty)$. We shall see in Lemma 6.2 that

$$\epsilon_{n,j} \equiv \lim_{\substack{z \to E_n^{(0)} \\ \operatorname{Im}(z) > 0}} \epsilon_{n,j}(z), \quad n \geqslant 0,\ j = 1,...,M, \tag{1.27}$$

exist and that $\epsilon_{n,j}$ does not depend on j. Let

$$E_{n,j}^{(2)}(e) = E_n^{(0)} + \operatorname{Re}(\epsilon_{n,j}) + (\Psi_{n,j}^{(0)}, H_{\rm I}^{(2)}\Psi_{n,j}^{(0)}) + (\Psi_{n,j}^{(0)}, (\delta m/2m^2)\,\mathbf{p}^2\Psi_{n,j}^{(0)}), \tag{1.28}$$

which are the formal perturbation expansions (see, e.g., Ref. 23, Chap. II), up to the second order in e, for the "perturbed energy levels." To this order the "energy level shift" is

$$\delta E_{n,j}^{(2)}(e) = E_{n,j}^{(2)}(e) - E_n^{(0)}. \tag{1.29}$$

The formal time dependent perturbation theory (the "Golden Rule") (see, e.g., Ref. 24) gives

$$\Gamma_n^{(2)}(e) \equiv -2\operatorname{Im}(\epsilon_{n,j}) > 0 \tag{1.30}$$

as the "decay probability," up to the second order in e, of the excited state $\Psi_{n,j}^{(0)}$. The decay is of course due to the spontaneous emission of light.

Now, the Heisenberg operators are given by

$$\mathbf{A}(f,t) = e^{itH}\mathbf{A}(f)e^{-itH}, \quad f \in \mathscr{S}(R^3), \ t \in R^1, \tag{1.31}$$

$$\mathbf{q}(t) = e^{itH}\mathbf{q}e^{-itH}, \tag{1.32}$$

$$\mathbf{p}(t) = e^{itH}\mathbf{p}e^{-itH}. \tag{1.33}$$

We consider the 2-point τ-functions in the case $m > \delta m$:

$$\tau_{\mu\nu}^{(h)}(t-s) = (\Omega, T[q_\mu(t) q_\nu(s)]\Omega), \tag{1.34}$$

$$\tau_{\mu\nu}^{(\mathrm{EM})}(f,g;t-s) = (\Omega, T[A_\mu(f,t) A_\nu(g,s)]\Omega), \quad f,g \in \mathscr{S}(R^3), \tag{1.35}$$

$$\tau_{\mu\nu}^{\mathrm{EM}}(f;t-s) = (\Omega, T[A_\mu(f,t) p_\nu(s)]\Omega), \quad f \in \mathscr{S}(R^3), \tag{1.36}$$

where Ω is the ground state of H and $T[\cdot]$ denotes the time ordered product. It will be shown by explicit construction in Sec. VI that $\tau_{\mu\nu}^{(h)}(t)$, $\tau_{\mu\nu}^{(\mathrm{EM})}(f,g;t)$ and $\tau_{\mu\nu}^{(\mathrm{EM},h)}(f;t)$ are all well-defined, and in particular that the latter two are continuous functionals on $\mathscr{S}(R^3) \times \mathscr{S}(R^3)$ and $\mathscr{S}(R^3)$, respectively, for each t. Hence we can write

$$\tau_{\mu\nu}^{(\mathrm{EM})}(f,g;t) \equiv \int d^3\mathbf{x}\, d^3\mathbf{y}\, f(\mathbf{x}) g(\mathbf{y}) \tau_{\mu\nu}^{(\mathrm{EM})}(\mathbf{x},\mathbf{y};t), \tag{1.37}$$

$$\tau_{\mu\nu}^{(\mathrm{EM},h)}(f;t) \equiv \int d^3\mathbf{x}\, f(\mathbf{x}) \tau_{\mu\nu}^{(\mathrm{EM},h)}(\mathbf{x};t). \tag{1.38}$$

Let

$$\hat{\tau}_{\mu\nu}^{(h)}(E) = \int_{-\infty}^{\infty} dt\, \tau_{\mu\nu}^{(h)}(t) e^{-itE}, \tag{1.39}$$

$$\hat{\tau}_{\mu\nu}^{(\mathrm{EM})}(\mathbf{x},\mathbf{y};E) = \int_{-\infty}^{\infty} dt\, \tau_{\mu\nu}^{(\mathrm{EM})}(\mathbf{x},\mathbf{y};t) e^{-itE}, \tag{1.40}$$

$$\hat{\tau}_{\mu\nu}^{(\mathrm{EM},h)}(\mathbf{x};E) = \int_{-\infty}^{\infty} dt\, \tau_{\mu\nu}^{(\mathrm{EM},h)}(\mathbf{x};t) e^{-itE}, \tag{1.41}$$

and

$$\mathbb{C}_+ = \{z \in \mathbb{C} \mid \mathrm{Re}(z) > 0\},$$
$$\Pi_\pm = \{z \in \mathbb{C} \mid \mathrm{Re}(z) > 0,\ \mathrm{Im}(z) \gtrless 0\}. \tag{1.42}$$

Then, we shall prove in Sec. VI

Theorem C: Let $m > \delta m$, and assume, in addition to (AI), that

$$\left.\begin{array}{l} \hat{\rho}(|\mathbf{k}|) \text{ has an analytic continuation } \hat{\rho}(z) \text{ onto } \Pi_- \text{ and} \\ \qquad \hat{\rho}(z) = O(|z|^{-3/2})(|z| \to \infty). \end{array}\right\} \quad (\mathrm{AII})$$

Then, all the three functions of E (with fixed $\mathbf{x},\mathbf{y}$) in (1.39)–(1.41) have meromorphic continuations from $(0,\infty)$ to $\mathbb{C}_+$ and there exists a constant $\lambda > 0$ such that, if $|e| < \lambda$, then the meromorphic continuations have in common a unique simple pole $\zeta(e)$ in Π_-, which is analytic in e, satisfying $\zeta(e) \to \omega_0$ as $e \to 0$. Let

$$\zeta(e) = \omega_0 + a_1 e + a_2 e^2 + \cdots, \quad |e| < \lambda.$$

Then:

(1) $a_1 = 0, \quad \mathrm{Im}(a_2) < 0.$

(2) $\Gamma_n^{(2)}(e) = -2ne^2\, \mathrm{Im}(a_2), \quad n \geqslant 0.$ (1.43)

(3) $\delta E_{n,j}^{(2)}(e) = \delta E_{0,1}^{(2)}(e) + ne^2\, \mathrm{Re}(a_2),$

$n \geqslant 0, \quad j = 1,\dots,M.$ (1.44)

This theorem shows that the formal perturbation expansions for the "energy level shifts" of the harmonic oscillator atom and the "decay probabilities" of the excited states are characterized in terms of a pole associated with the 2-point τ-functions, at least up to the second order in e. We shall also show that the pole $\zeta(e)$ is the resonance pole of the S-matrix for the photon scattering by the atom (see Sec. VI).

Remarks:

(1) The general n-point τ-functions are written as a sum of products of 2-point τ-functions.

(2) We can also consider the Hamiltonian

$$H' = \frac{1}{2m} : (\mathbf{p} - e\mathbf{A}(\rho))^2 : + \frac{1}{2} m\omega_0^2 \mathbf{q}^2 + H_0^{\mathrm{EM}} + \frac{\delta m}{2m^2}\mathbf{p}^2,$$

which is the "Taylor expansion" of H up to second order in e. It can be shown that results similar to those of Theorems B-(1) and C hold in this model as well. In particular, as is expected, the pole which characterizes the "energy level shifts" and the "decay probabilities" coincides with $\zeta(e)$ up to the second order in e.

In Sec. VII broken symmetry aspects of the model are considered: The model has a symmetry which is broken if $m < \delta m$. From this point of view, the unphysical solution in the case $m < \delta m$ may be regarded as a kind of "Goldstone boson." Then, we shall describe a mechanism ("Higgs Mechanism") by which the unphysical solution disappears and the physical theory can be obtained.

In the last section, we shall consider the point limit, $\rho(\mathbf{x}) \to \delta(\mathbf{x})$, of the interaction, which corresponds to the removal of the ultraviolet cutoff in momentum space and requires the infinite mass-renormalization. The point limit is taken in terms of the Wightman distributions constructed from the physical theory obtained in Sec. VII.

II. SELF-ADJOINTNESS OF THE TOTAL HAMILTONIAN

In this section we prove Theorem A.2. Let $L = H_0 + I$. Then, by the basic estimates

$$\|a^{(r)}(f)\Psi\| < \|f/\sqrt{\omega}\|_0 \|H_0^{\mathrm{EM}\,1/2}\Psi\|, \tag{2.1}$$

$$\|a^{(r)*}(f)\Psi\| < \|f/\sqrt{\omega}\|_0 \|H_0^{\mathrm{EM}\,1/2}\Psi\| + \|f\|_0 \|\Psi\|, \tag{2.2}$$

we can show that $H_1^{(j)}, j = 1,2$, and R given in (1.21) and (1.22) are H_0-bounded, so that

$$\|H\Psi\| \leqslant c\|L\Psi\|, \quad \Psi \in D(H_0),$$

for some constant $c > 0$. We can also show by commutation relations that

$$|(H\Phi, L\Psi) - (L\Phi, H\Psi)| \leqslant d\, \|L^{1/2}\Phi\|\, \|L^{1/2}\Psi\|,$$
$$\Phi, \Psi \in D(H_0)$$

for some constant $d > 0$. It is clear that H is symmetric on $D(H_0)$. Therefore, by the Nelson's commutator theorem (see, e.g., Ref. 25, §X.5) H is essentially self-adjoint on any core for H_0.

We next prove that, if $m > \delta m$, then H is actually self-adjoint with $D(H) = D(\widetilde{H}) \cap D(\mathbf{q}^2)$ and is bounded below. Let $m > \delta m$. Then, by Theorem A.1, we can take a constant $c > 0$ such that

$$\widetilde{H}_c \equiv \widetilde{H} + c \geqslant 0. \tag{2.3}$$

Therefore,

$$\mathrm{Re}(\widetilde{H}_c \Psi, \mathbf{q}^2 \Psi) \geqslant \mathrm{Re}([q_\mu, \widetilde{H}_c]\Psi, q_\mu \Psi)$$

for all Ψ in $D(H_0^2)$. But, the commutation relations give

$$[q_\mu, \widetilde{H}_c]\Psi = \frac{i}{m - \delta m}(p_\mu - eA_\mu(\rho))\Psi,$$

so that

$$\mathrm{Re}(\widetilde{H}_c\Psi,\mathbf{q}^2\Psi)\geqslant-\frac{1}{m-\delta m}\,\|\Psi\|^2.$$

Since $\widetilde{H}_c$ and $\mathbf{q}^2$ are H_0-bounded, this inequality extends to all Ψ in $D(H_0)$. Thus, we obtain

$$\|\widetilde{H}_c\Psi\|^2+\|\tfrac{1}{2}m\omega_0^2\mathbf{q}^2\Psi\|^2\leqslant\|(H+c)\Psi\|^2+\frac{m\omega_0^2}{m-\delta m}\,\|\Psi\|^2 \tag{2.4}$$

for all Ψ in $D(H_0)$. Since H is essentially self-adjoint on $D(H_0)$ as proved above, (2.4) implies that H is self-adjoint with $D(H)=D(\widetilde{H})\cap D(\mathbf{q}^2)$. The statement in regard to the bounded belowness of H follows from (1.19) and the positivity of $\mathbf{q}^2$.

Finally, we show that, if $m<\delta m$, then H is not bounded below. We can easily find a sequence $\{\Psi_n\}_{n=1}^{\infty}$ in $\mathscr{H}$ such that $\|\Psi_n\|=1$ and $(\Psi_n,H\Psi_n)\to-\infty$ as $n\to\infty$. Take, for example,

$$\Psi_n=\psi_n\otimes\Omega_F,$$

where

$$\psi_n(\mathbf{q})=\left(\frac{n}{\pi}\right)^{3/4}e^{-n\mathbf{q}^2/2}$$

and Ω_F is the Fock vacuum in $\mathscr{F}^{\mathrm{EM}}$. Thus, H is not bounded below. This concludes the proof of Theorem A.2.

III. CONSTRUCTION OF EXACT SOLUTION TO THE HEISENBERG EQUATIONS

In this section we shall construct explicitly the Heisenberg operators $\mathbf{A}(f,t)$ and $\mathbf{q}(t)$ as defined by (1.31) and (1.32). Formally the Heisenberg equations read (summation over repeated indices with respect to Greek letters is understood):

$$\left.\begin{aligned}&\left[(m-\delta m)\frac{d^2}{dt^2}+m\omega_0^2\right]q_\mu(t)=-e\int d^3\mathbf{x}\,\rho(\mathbf{x})\frac{\partial}{\partial t}A_\mu(\mathbf{x},t),\\&\left(\frac{\partial^2}{\partial t^2}-\Delta\right)A_\mu(\mathbf{x},t)=e\rho_{\mu\nu}(\mathbf{x})\frac{d}{dt}q_\nu(t),\end{aligned}\right\} \tag{3.1}$$

where

$$\rho_{\mu\nu}(\mathbf{x})=\frac{1}{((2\pi)^3)^{1/2}}\int d^3\mathbf{k}\,\hat{\rho}(\mathbf{k})d_{\mu\nu}(\mathbf{k})e^{i\mathbf{k}\mathbf{x}} \tag{3.2}$$

with

$$d_{\mu\nu}(\mathbf{k})=\delta_{\mu\nu}-(k_\mu k_\nu/\mathbf{k}^2), \tag{3.3}$$

and $\mathbf{A}(\mathbf{x},t)$ is the symbolic notation given by

$$\mathbf{A}(f,t)=\int d^3\mathbf{x}\mathbf{A}(\mathbf{x},t)f(\mathbf{x}). \tag{3.4}$$

Equations (3.1) can be exactly solved as an initial value problem. The following is a rigorous formulation of the solutions given in Refs. 8 and 9 by the theory of Fourier transform[26] (Ref. 8; Laplace transform is used in Ref. 9). We begin with preparing some technical lemmas.

A. Technical preliminaries

We formally define the operator G_ϵ, $\epsilon>0$, by

$$(G_\epsilon f)(\mathbf{k})=\int d^3\mathbf{k}'\,\frac{f(\mathbf{k}')}{\sqrt{|\mathbf{k}||\mathbf{k}'|}(\mathbf{k}^2-\mathbf{k}'^2+i\epsilon)}. \tag{3.5}$$

Let $M_\alpha(R^3)$ be the Hilbert space given by

$$M_\alpha(R^3)=\{f\,|\,\|f\|_\alpha\equiv\|\omega^\alpha f\|_{L^2(R^3)}<\infty\},\quad\alpha\in R^3. \tag{3.6}$$

Then, we have

Lemma 3.1: (1) G_ϵ is a bounded operator on $M_0(R^3)$ and has the strong limit

$$\operatorname*{s-lim}_{\epsilon\to+0}G_\epsilon\equiv G, \tag{3.7}$$

on $M_0(R^3)$. Furthermore, G is skew-symmetric on $M_0(R^3)$.

(2) G is also a bounded operator on $M_{-1/2}(R^3)$.

For the proof, see [I, Lemmas 4.1–4.3].

Let

$$D(z)=m\omega_0^2-z\left(m-\delta m-\tfrac{2}{3}e^2\int d^3\mathbf{k}\,\frac{\hat{\rho}(\mathbf{k})^2}{z-\mathbf{k}^2}\right), \tag{3.8}$$

which is analytic in the cut plane $\mathbb{C}\setminus[0,\infty)$.

Lemma 3.2:

(1) Let $m>\delta m$. Then, $D(z)$ has no zero in $\mathbb{C}\setminus[0,\infty)$.

(2) Let $m<\delta m$. Then $D(z)$ has a unique simple zero in $(-\infty,0)$. We denote it by $-E_B^2$ $(E_B>0)$.

The proof is elementary and is omitted.

Lemma 3.3:

$$D_\pm(s)\equiv\lim_{\epsilon\to+0}D(s\pm i\epsilon) \tag{3.9}$$

exist for each $s\in[0,\infty)$ and are continuous with respect to s. Furthermore,

$$\inf_{s\in[0,\infty)}|D_\pm(s)|>0. \tag{3.10}$$

The proof is quite similar to that of [I, Lemma 4.4].

Lemma 3.4: Let

$$Q(\mathbf{k})=ie\hat{\rho}(\mathbf{k})/D_+(\mathbf{k}^2). \tag{3.11}$$

Then,

$$(Q,d_{\mu\nu}Q)_0=\delta_{\mu\nu}\left\{\frac{1}{m\omega_0^2}-\left(\frac{\gamma_B}{E_B}\right)^2\theta(\delta m-m)\right\}, \tag{3.12}$$

where θ is the Heaviside function and γ_B is defined only in the case $\delta m>m$ by

$$\gamma_B=\frac{1}{(D'(-E_B^2))^{1/2}}>0. \tag{3.13}$$

Proof: Since Q is rotation invariant, we have

$$I_{\mu\nu}\equiv(Q,d_{\mu\nu}Q)_0=\delta_{\mu\nu}\tfrac{2}{3}\|Q\|_0^2.$$

But, since

$$D_-(\mathbf{k}^2)-D_+(\mathbf{k}^2)=\tfrac{8}{3}i\pi^2e^2|\mathbf{k}|^3\hat{\rho}(|\mathbf{k}|)^2, \tag{3.14}$$

we get

$$I_{\mu\nu}=\delta_{\mu\nu}\frac{1}{2\pi i}\int_0^\infty ds\left\{\frac{1}{D_+(s)}-\frac{1}{D_-(s)}\right\}\frac{1}{s}.$$

Then, the method of contour integration using Lemma 3.2 gives (3.12). ∎

Lemma 3.5: Let $T_{\mu\nu}$, $\mu,\nu=1,2,3$, be operators defined by

$$T_{\mu\nu}f=\delta_{\mu\nu}f+ie\omega^{5/2}QG\sqrt{\omega}\,\hat{\rho}d_{\mu\nu}f. \tag{3.15}$$

Then:

(1) $T_{\mu\nu}$ is a bounded operator on $M_0(R^3)$.

(2) $T_{\mu\nu}$ and $T^*_{\mu\nu}$ are also bounded operators on $M_\alpha(R^3)$ for $\alpha = \pm\frac{1}{2}, -1$, where $T^*_{\mu\nu}$ denotes the adjoint of $T_{\mu\nu}$ in $M_0(R^3)$.

Proof: (1) follows from Lemma 3.1-(1) and the fact that $\omega^{3/2}Q$ is in $L^\infty(R^3)$. By the skew-symmetry of G, we have

$$T^*_{\mu\nu} f = \delta_{\mu\nu} f + ie\sqrt{\omega}\,\hat{\rho}\, d_{\mu\nu} G\omega^{5/2}\overline{Q} f. \tag{3.16}$$

Hence (2) follows from Lemma 3.1-(2) and the fact that $\omega^2 Q$ is in $L^\infty(R^3)$. ∎

Remark: The operator $T_{\mu\nu}$ is symbolically given by a distribution kernel:

$$(T_{\mu\nu} f)(\mathbf{k}) = \int d^3\mathbf{k}' T_{\mu\nu}(\mathbf{k},\mathbf{k}') f(\mathbf{k}') \tag{3.17}$$

with

$$T_{\mu\nu}(\mathbf{k},\mathbf{k}') = \delta_{\mu\nu}\delta^3(\mathbf{k}-\mathbf{k}') - \frac{e^2\mathbf{k}^2\hat{\rho}(\mathbf{k})\hat{\rho}(\mathbf{k}')d_{\mu\nu}(\mathbf{k}')}{(\mathbf{k}^2 - \mathbf{k}'^2 + i0)D_+(\mathbf{k}^2)}. \tag{3.18}$$

Lemma 3.6:

(1) $$T^*_{\alpha\beta} d_{\beta\mu} T_{\mu\nu} + \theta(\delta m - m)(\gamma_\beta E_B)^2 (F_{\nu\beta},\ \cdot\ )_0 F_{\beta\alpha} = d_{\alpha\nu} I, \tag{3.19}$$

where

$$F_{\alpha\beta} = \frac{e d_{\alpha\beta}\hat{\rho}}{\omega^2 + E_B^2}. \tag{3.20}$$

(2) $$e^{(r)}_\alpha T_{\alpha\beta} d_{\beta\mu} T^*_{\mu\nu} e^{(s)}_\nu + m\omega_0^2 (e^{(s)}_\alpha Q,\ \cdot\ )_0 e^{(r)}_\alpha Q = \delta_{rs} I. \tag{3.21}$$

(3) $$T^*_{\alpha\beta} d_{\beta\mu} Q = i\gamma_B^2 F_{\alpha\mu}\theta(\delta m - m). \tag{3.22}$$

(4) $$\overline{T}_{\mu\nu} f = D^{\pm}_{-} T_{\mu\nu} f + (1 - D^{\pm}_{-}) \times \left(\delta_{\mu\nu} f - \frac{3}{8\pi}[d_{\mu\nu} f]\right),\quad f \in M_0(R^3), \tag{3.23}$$

where

$$D^{\pm}_{-}(\mathbf{k}) = D_+(\mathbf{k}^2)/D_-(\mathbf{k}^2), \tag{3.24}$$

$$[f](\mathbf{k}) = \int_{S^2} d\Omega(\mathbf{k}) f(|\mathbf{k}|\Omega(\mathbf{k})), \tag{3.25}$$

and the bar denotes the operator defined by

$$\overline{A} f = \overline{A\overline{f}}. \tag{3.26}$$

(5) If h is a rotation invariant function on R^3, then we have

$$T^*_{\alpha\beta} d_{\beta\mu} h T_{\mu\nu} = \overline{T}^*_{\alpha\beta} d_{\beta\mu} h \overline{T}_{\mu\nu} \tag{3.27}$$

as operator identities and

$$T^*_{\alpha\beta} d_{\beta\mu} h Q = -\overline{T}^*_{\alpha\beta} d_{\beta\mu} h\overline{Q},\quad \text{a.e.} \tag{3.28}$$

for h with $hQ \in M_0(R^3)$.

(6) $$e^{(r)}_\alpha T_{\alpha\beta} F_{\beta\mu} = -i\frac{m\omega_0^2 Q e^{(r)}_\mu \theta(\delta m - m)}{E_B^2}. \tag{3.29}$$

(7) $$T_{\mu\nu}\hat{\rho} = -\delta_{\mu\nu}\frac{i}{e}[m\omega_0^2 - (m - \delta m)\omega^2]Q. \tag{3.30}$$

(8) $$[\omega^2, T^*_{\mu\nu}] = d_{\mu\nu} ie\hat{\rho}(\omega^2 Q,\ \cdot\ )_0. \tag{3.31}$$

The proof of Lemma 3.6 is quite similar to that of [I, Lemma 4.9] and hence we omit it.

B. Exact solution to the Heisenberg equations of motion

Let $\hat{A}_\mu$ and $\hat{\pi}_\mu$ be the Fourier transforms of A_μ and π_μ respectively:

$$\hat{A}_\mu(f) = A_\mu(\hat{f}),\quad \hat{\pi}_\mu(f) = \pi_\mu(\hat{f}). \tag{3.32}$$

We define

$$b^{(r)}(f) = \frac{1}{\sqrt{2}}\left\{m\omega_0^2\left(\frac{Q}{\sqrt{\omega}}e^{(r)}_\mu, f\right)_0 q_\mu + i(\sqrt{\omega}Qe^{(r)}_\mu, f)_0 p_\mu + \hat{A}_\mu(T^*_{\mu\nu}e^{(r)}_\nu\sqrt{\omega}f) + i\hat{\pi}_\mu\left(T^*_{\mu\nu}e^{(r)}_\nu\frac{f}{\sqrt{\omega}}\right)\right\}, \tag{3.33}$$

$$b^{(r)*}(f) = \frac{1}{\sqrt{2}}\left\{m\omega_0^2\left(\frac{\overline{Q}}{\sqrt{\omega}}e^{(r)}_\mu, f\right)_0 q_\mu - i(\sqrt{\omega}\,\overline{Q}e^{(r)}_\mu, f)_0 p_\mu + \hat{A}_\mu(\overline{T}^*_{\mu\nu}\bar{e}^{(r)}_\nu\sqrt{\omega}\tilde{f}) - i\hat{\pi}_\mu\left(\overline{T}^*_{\mu\nu}\bar{e}^{(r)}_\nu\frac{\tilde{f}}{\sqrt{\omega}}\right)\right\},\quad f \in M_0(R^3). \tag{3.34}$$

$$C_\mu = \theta(\delta m - m)\left(\frac{E_B}{2}\right)^{1/2}\gamma_B\left\{\frac{m\omega_0^2}{E_B^2}q_\mu - \frac{1}{E_B}p_\mu + E_B\hat{A}_\alpha(F_{\alpha\mu}) + \hat{\pi}_\alpha(F_{\alpha\mu})\right\}, \tag{3.35}$$

$$D_\mu = \theta(\delta m - m)\left(\frac{E_B}{2}\right)^{1/2}\gamma_B\left\{\frac{m\omega_0^2}{E_B^2}q_\mu + \frac{1}{E_B}p_\mu - E_B\hat{A}_\alpha(F_{\alpha\mu}) + \hat{\pi}_\alpha(F_{\alpha\mu})\right\}. \tag{3.36}$$

Let

$$D_F = D_0 \otimes \mathscr{F}^{\mathrm{EM}}_0, \tag{3.37}$$

where

$$D_0 = \left\{\sum_{i,j,k}^{\text{finite}} \alpha_{ijk} h_i \otimes h_j \otimes h_k \,\middle|\, \alpha_{ijk} \in \mathbb{C}, h_i\text{: the } i\text{th Hermite function on } R^1,\quad i = 0,1,\ldots\right\}. \tag{3.38}$$

Then, by Lemma 3.5, $b^{(r)}(f), b^{(r)*}(f), f \in M_0(R^3)$, C_μ and D_μ are well-defined on D_F and leave it invariant, satisfying

$$(b^{(r)}(f)\Psi, \Phi) = (\Psi, b^{(r)*}(\bar{f})\Phi),\quad \Psi, \Phi \in D_F, \tag{3.39}$$

$$[b^{(r)}(f), b^{(s)*}(g)] = (\bar{f}, g)_0\delta_{rs},\quad [b^{(r)}(f), b^{(s)}(g)] = 0,$$
$$f, g \in M_0(R^3),\quad r,s = 1,2. \tag{3.40}$$

$$[b^{(r)}(f), C_\mu] = [b^{(r)}(f), D_\mu] = 0,$$
$$f \in M_0(R^3),\ r = 1,2,\ \mu = 1,2,3, \tag{3.41}$$

$$[C_\mu, D_\nu] = i\delta_{\mu\nu}\theta(\delta m - m),\quad \mu,\nu = 1,2,3. \tag{3.42}$$

The commutation relations (3.40) [respectively (3.41)] can be proved directly using Lemma 3.6-(2), 3.6-(4) [respectively 3.6-(6)]. The commutation relations (3.42) is proved by direct computations.

Lemma 3.7: D_F is a set of analytic vectors for C_μ and D_μ, $\mu = 1,2,3$. In particular, C_μ and D_μ are essentially self-adjoint on D_F.

Proof: The first half of the lemma follows from the standard number operator estimates for p_μ, q_μ, $a^{(r)}(f)$ and $a^{(r)*}(f)$ (see, e.g., Ref. 25, §X.6 and §X.7). It is clear that C_μ and D_μ are symmetric on D_F. Thus, by Nelson's analytic vector theorem (see, e.g., Ref. 25, §X.6), they are essentially self-adjoint on D_F. ■

Lemma 3.8: Let $m > \delta m$ and

$$\widehat{H} = H - E_0 \geqslant 0. \tag{3.43}$$

Then,

$$\|a^{(r)}(f)\Psi\| \leqslant c\|f\|_{-1/2}\|(\widehat{H}+1)^{1/2}\Psi\|, \tag{3.44}$$

$$\|a^{(r)*}(f)\Psi\| \leqslant c(\|f\|_{-1/2} + \|f\|_0)\|(\widehat{H}+1)^{1/2}\Psi\|, \quad r = 1,2, \tag{3.45}$$

$$\|q_\mu \Psi\| \leqslant c\|(\widehat{H}+1)^{1/2}\Psi\|, \tag{3.46}$$

$$\|p_\mu \Psi\| \leqslant c\|(\widehat{H}+1)^{1/2}\Psi\|, \quad \mu = 1,2,3, \tag{3.47}$$

for all $f \in M_{-1/2}(R^3) \cap M_0(R^3)$, $\Psi \in D(H^{1/2})$ and for some constant $c > 0$.

Proof: With the operator in (2.3), we have

$$\|H_0^{\mathrm{EM}1/2}\Psi\| \leqslant \|\widetilde{H}_c^{1/2}\Psi\| \quad \Psi \in D(\widetilde{H}_c^{1/2}).$$

On the other hand, it follows from the H-boundedness of $\widetilde{H}_c$ (2.4) that $\widetilde{H}_c$ is formbounded with respect to H (see, e.g., Ref. 25, §X.2). Therefore, together with (2.1) and (2.2), we obtain (3.44) and (3.45). Since q_μ^2 is also H-bounded by (2.4), we have (3.46). By (3.44) and (3.45), $A_\mu(f)$ is $\widehat{H}^{1/2}$-bounded for $\hat{f}$ in $M_{-1}(R^3) \cap M_{-1/2}(R^3)$. On the other hand, one can easily show that

$$\frac{1}{2(m-\delta m)}\|(p_\mu - eA_\mu(\rho))\Psi\|^2 \leqslant \|\widehat{H}^{1/2}\Psi\|^2 + d\|\Psi\|^2,$$
$$\Psi \in D(\widehat{H}^{1/2})$$

for some constant $d > 0$. Thus, we get

$$\|p_\mu \Psi\| \leqslant \|(p_\mu - eA_\mu(\rho))\Psi\| + |e|\,\|A_\mu(\rho)\Psi\| \leqslant c\|(\widehat{H}+1)^{1/2}\Psi\|.$$ ■

Lemma 3.8 gives estimates for $b^{(r)}(f)$ and $b^{(r)*}(f)$ in the case $m > \delta m$:

Lemma 3.9: Let $m > \delta m$, and let $b^{(r)\#}(f)$ denote either $b^{(r)}(f)$ or $b^{(r)*}(f)$. Then,

$$\|b^{(r)\#}(f)\Psi\| \leqslant c(\|f\|_{-1/2} + \|f\|_0)\|(\widehat{H}+1)^{1/2}\Psi\| \tag{3.48}$$

for all $f \in M_{-1/2}(R^3) \cap M_0(R^3)$, $\Psi \in D(\widehat{H}^{1/2})$ and some $c > 0$.

Proof: This follows from Lemmas 3.5 and 3.8. ■

Lemma 3.10:

(1) Let f be in $M_{-1/2}(R^3) \cap M_0(R^3) \cap M_1(R^3)$. Then, we have

$$[H, b^{(r)\#}(f)]\Psi = \pm b^{(r)\#}(\omega f)\Psi, \quad \Psi \in D_F \cap D(H_0), \tag{3.49}$$

where $+$ (respectively $-$) sign goes with $b^{(r)*}(\cdot)$ [respectively $b^{(r)}(\cdot)$].

(2)
$$[H, C_\mu]\Psi = -iE_B C_\mu \Psi, \quad [H, D_\mu]\Psi = iE_B D_\mu \Psi, \quad \Psi \in D_F \cap D(H_0). \tag{3.50}$$

$$e^{itH}C_\mu e^{-itH} = e^{tE_B}C_\mu, \quad e^{itH}D_\mu e^{-itH} = e^{-tE_B}D_\mu \quad \text{on } D_F. \tag{3.51}$$

Proof: The commutation relations (3.49) are proved by Lemma 3.6-(7), 3.6-(8). The part (2) can be proved by direct computations.. ■

Remark: If $M > \delta m$, then, by Lemma 3.9, (3.49) extends to all Ψ in $D(\widehat{H}^{3/2})$.

We now give the exact solution to the Heisenberg equations of motion:

Theorem 3.1: The Heisenberg operators defined by (1.31) and (1.32) have the following explicit form on the domain D_F:

$$A_\mu(f,t) = \frac{1}{\sqrt{2}}\sum_{r=1}^{2}\left\{b^{(r)*}\left(\frac{e_\nu^{(r)}}{\sqrt{\omega}}e^{it\omega}\overline{T}_{\mu\nu}\hat{f}\right) + b^{(r)}\left(\frac{e_\nu^{(r)}}{\sqrt{\omega}}e^{-it\omega}T_{\mu\nu}\tilde{\hat{f}}\right)\right\} + \left(\frac{E_B}{2}\right)^{1/2}\gamma_B(F_{\mu\nu},\hat{f})_0(C_\nu e^{tE_B} - D_\nu e^{-tE_B}), \tag{3.52}$$

$$q_\mu(t) = \frac{1}{\sqrt{2}}\sum_{r=1}^{2}\{b^{(r)*}(\sqrt{\omega}\,\overline{Q}e_\mu^{(r)}e^{it\omega}) + b^{(r)}(\sqrt{\omega}\,Qe_\mu^{(r)}e^{-it\omega})\} + \frac{\gamma_B}{\sqrt{2E_B}}(C_\mu e^{tE_B} + D_\mu e^{-tE_B}), \quad \mu = 1,2,3. \tag{3.53}$$

Proof: The outline is as follows: The initial conditions are checked by Lemma 3.4, Lemma 3.6-(1), 3.6-(3), and 3.6-(5). Using Lemma 3.10, one can get the expressions (3.52) and (3.53). ■

Remark: $\mathbf{A}(f,t)$ and $\mathbf{q}(t)$ satisfy the equations of motion (3.1) on D_F in the sharp-time operator-valued distribution sense with the time derivatives being taken in the strong topology.

Theorem 3.1 (the initial conditions) permits us to express $a^{(r)\#}(f)$, $\mathbf{p}$ and $\mathbf{q}$ in terms of $b^{(r)\#}(\ )$, $\mathbf{C}$ and $\mathbf{D}$:

Corollary 3.1: Let

$$W_+^{(r,s)}f = \frac{1}{2}\left(\frac{1}{\sqrt{\omega}}e_\mu^{(r)}T^*_{\mu\nu}e_\nu^{(s)}\sqrt{\omega} + \sqrt{\omega}\,e_\mu^{(r)}T^*_{\mu\nu}e_\nu^{(s)}\frac{1}{\sqrt{\omega}}\right)f, \tag{3.54}$$

$$W_-^{(r,s)}f = \frac{1}{2}\left(\frac{1}{\sqrt{\omega}}e_\mu^{(r)}T^*_{\mu\nu}\tilde{e}_\nu^{(s)}\sqrt{\omega} - \sqrt{\omega}\,e_\mu^{(r)}T^*_{\mu\nu}\tilde{e}_\nu^{(s)}\frac{1}{\sqrt{\omega}}\right)\tilde{f}, \tag{3.55}$$

$$F_{\pm,\mu}^{(r)} = \frac{1}{2}\sqrt{E_B}\,\gamma_B\left(\sqrt{\omega} \pm i\frac{E_B}{\sqrt{\omega}}\right)F_{\mu\nu}e_\nu^{(r)}. \tag{3.56}$$

Then,

$$a^{(r)}(f) = \sum_{s=1}^{2}\{-b^{(s)*}(\overline{W}_-^{(r,s)*}f) + b^{(s)}(W_+^{(r,s)*}f)\} + (F_{-,\mu}^{(r)},f)_0 C_\mu - (F_{+,\mu}^{(r)},f)_0 D_\mu, \tag{3.57}$$

$$a^{(r)*}(f) = \sum_{s=1}^{2}\{b^{(s)*}(\overline{W}_+^{(r,s)*}f) - b^{(s)}(W_-^{(r,s)*}f)\} + (F_{+,\mu}^{(r)},f)_0 C_\mu - (F_{-,\mu}^{(r)},f)_0 D_\mu, \tag{3.58}$$

$$q_\mu = \frac{1}{\sqrt{2}} \sum_{r=1}^{2} \{ b^{(r)*}(\sqrt{\omega}\, \overline{Q} e_\mu^{(r)}) + b^{(r)}(\sqrt{\omega}\, Q e_\mu^{(r)}) \} + \frac{\gamma_B}{\sqrt{2E_B}} (C_\mu + D_\mu), \tag{3.59}$$

$$p_\mu = \frac{im\omega_0^2}{\sqrt{2}} \sum_{r=1}^{2} \left\{ b^{(r)*}\left(\frac{\overline{Q}}{\sqrt{\omega}} e_\mu^{(r)} \right) - b^{(r)}\left(\frac{Q}{\sqrt{\omega}} e_\mu^{(r)} \right) \right\} - \frac{m\omega_0^2 \gamma_B}{E_B\sqrt{2E_B}} (C_\mu - D_\mu). \tag{3.60}$$

IV. ASYMPTOTIC FIELDS

In this section we consider the asymptotic limits, as $t \to \pm\infty$, of the operators given by

$$a_t^{(r)\#}(f) = e^{itH} e^{-itH_0} a^{(r)\#}(f) e^{itH_0} e^{-itH}, \tag{4.1}$$

with f in $M_{-1/2}(R^3) \cap M_0(R^3)$. If $m > \delta m$, then H_0^{EM} is form-bounded with respect to H (see the proof of Lemma 3.8) and hence $a_t^{(r)\#}(f)$ is well-defined on $D(\hat{H}^{1/2})$. On the other hand, if $m < \delta m$, then $a_t^{(r)\#}(f)$ is well-defined on D_F [see (3.57), (3.58) and Lemma 3.10].

Lemma 4.1: Each $W_-^{(r,s)}$ is a Hilbert–Schmidt operator on $M_0(R^3)$.

Proof: By direct computations, we have

$$(W_-^{(r,s)} f)(\mathbf{k}) = \int d^3\mathbf{k}'\, W_-^{(r,s)}(\mathbf{k},\mathbf{k}') f(\mathbf{k}'),$$

where

$$W_-^{(r,s)}(\mathbf{k},\mathbf{k}') = \frac{-ie\overline{Q}(\mathbf{k}')\hat{\rho}(\mathbf{k}) e_\mu^{(r)}(\mathbf{k}) e_\mu^{(s)}(\mathbf{k}') |\mathbf{k}'|^2}{2(|\mathbf{k}| + |\mathbf{k}'|)\sqrt{|\mathbf{k}|\,|\mathbf{k}'|}}.$$

But, it can be seen easily that $W_-^{(r,s)}(\cdot\,,\,\cdot)$ is in $L^2(R^6)$. Therefore, the lemma follows. ■

Theorem 4.1:

(1) Let $m > \delta m$. Then, for all Ψ in $D(\hat{H}^{1/2})$ and all f in $M_{-1/2}(R^3) \cap M_0(R^3)$, the strong limits

$$\operatorname*{s-lim}_{t \to \pm\infty} a_t^{(r)\#}(f)\Psi \equiv a_{\substack{\text{out}\\ \text{in}}}^{(r)\#}(f)\Psi \tag{4.2}$$

exist and are given explicitly by

$$a_{\text{in}}^{(r)\#}(f) = b^{(r)\#}(f), \tag{4.3}$$

$$a_{\text{out}}^{(r)}(f) = \sum_{s=1}^{2} b^{(s)}(L^{(r,s)} f), \tag{4.4}$$

$$a_{\text{out}}^{(r)*}(f) = \sum_{s=1}^{2} b^{(s)*}(\overline{L}^{(r,s)} f), \tag{4.5}$$

where

$$L^{(r,s)} f = \delta_{rs} f + \pi e\omega^3 \hat{\rho} Q e_\mu^{(s)} [f e_\mu^{(r)}]. \tag{4.6}$$

(2) Let $m < \delta m$ and ρ be in $C_0^\infty(R^3)$. Then, the strong limits (4.2) with Ψ in D_F do not exist in general.

Proof: (1) The existence of the strong limits (4.2) can be proved in the same way as in Refs. 27 and 28. We prove (4.3)–(4.5). By Lemma 3.10–(1) and (3.57), we have

$$a_t^{(r)}(f) = \sum_{s=1}^{2} \{ -b^{(s)*}(e^{it\omega} \overline{W}_-^{(r,s)*} e^{it\omega} f) + b^{(s)}(e^{-it\omega} W_+^{(r,s)*} e^{it\omega} f) \} \tag{4.7}$$

on D_F. Since $\overline{W}_-^{(r,s)*}$ is also a Hilbert–Schmidt operator on $M_0(R^3)$ by Lemma 4.1, it follows from the Proposition A1 in Appendix A that

$$\lim_{t \to \pm\infty} \| e^{it\omega} \overline{W}_-^{(r,s)*} e^{it\omega} f \|_0 = 0 \tag{4.8}$$

for all f in $M_0(R^3)$. On the other hand, let

$$X(t) = e^{-it\omega} W_+^{(r,s)*} e^{it\omega} f \tag{4.9}$$

with f in $\mathscr{S}(R^3)$. It is easy to see that $X(t)$ is strongly differentiable with respect to t in $M_0(R^3)$ and

$$\frac{d}{dt} X(t) = \frac{1}{2} e\omega^{3/2} Q e_\mu^{(s)} e^{-it\omega} \left(\frac{\hat{\rho} e_\mu^{(r)} e^{-it\omega}}{\sqrt{\omega}}, f \right)_0.$$

Therefore, we get

$$X(t) = W_+^{(r,s)*} f + \tfrac{1}{2} e\omega^{3/2} Q e_\mu^{(s)} \times \int_0^t d\tau\, e^{-i\tau\omega} \left(\frac{\hat{\rho} e_\mu^{(r)} e^{-i\tau\omega}}{\sqrt{\omega}}, f \right)_0.$$

By integration by parts we have

$$\left| \left(\frac{\hat{\rho} e_\mu^{(r)} e^{-it\omega}}{\sqrt{\omega}}, f \right)_0 \right| < \text{const} \times \frac{1}{\tau^2},$$

so that the strong limits $\text{s-lim}_{t \to \pm\infty} X(t) \equiv X_\pm$ in $M_0(R^3)$ exist. It is easy to check that they are given by

$$X_- = \delta_{rs} f, \quad X_+ = L^{(r,s)} f. \tag{4.10}$$

Since

$$\operatorname*{s-lim}_{n \to \infty} b^{(r)\#}(f_n) = b^{(r)\#}(f) \text{ on } D_F$$

as $f_n \to f$ in $M_0(R^3)$ (this can be proved by the number operator estimates for $a^{(r)\#}(\;)$, $\mathbf{p}$ and $\mathbf{q}$), the desired result with respect to $a_t^{(r)}(f)$ follows from (4.7)–(4.9) and (4.10). By a limiting argument we can extend the result to all f in $M_{-1/2}(R^3) \cap M_0(R^3)$ and all Ψ in $D(\hat{H}^{1/2})$. In the same way we can prove the result for $a_t^{(r)*}(f)$.

(2) In this case we have

$$a_t^{(r)}(f) = \sum_{s=1}^{2} \{ -b^{(s)*}(e^{it\omega} \overline{W}_-^{(r,s)*} e^{it\omega} f) + b^{(s)}(e^{-it\omega} W_+^{(r,s)*} e^{it\omega} f) \} + (F_{-,\mu}^{(r)}, e^{it\omega} f)_0 e^{tE_B} C_\mu - (F_{+,\mu}^{(r)}, e^{it\omega} f)_0 e^{-tE_B} D_\mu. \tag{4.11}$$

As proved above, the first term of (4.11) converges in the strong topology on D_F as $t \to \pm\infty$. But, we can see that the second (respectively the third) term diverges [respectively converges to zero for all f in $M_0(R^3)$] for some f's as $t \to \infty$ and converges to zero for all f in $M_0(R^3)$ (respectively diverges for some f's) as $t \to -\infty$. This proves the part (2). ■

V. ANALYSIS OF THE SPECTRUM OF THE TOTAL HAMILTONIAN

In this section we shall prove Theorem B.

A. The case $m > \delta m$

Lemma 5.1:

(1) Either $\sigma_p(H) = \varnothing$ or $\sigma_p(H) = \{E_0\}$ holds.

(2) the ground state of H, if it exists, is unique up to scalar multiples.

(3) Ψ is the ground state of H if and only if $b^{(r)}(f)\Psi = 0$ for all f in $M_{-1/2}(R^3)\cap M_0(R^3)$ and for $r = 1,2$.

Since we have established (3.57)–(3.60) and Theorem 4.1, the proof of this lemma is quite similar to that of [I, Lemma 6.2–6.4].

With Lemma 5.1 in mind, we shall prove the existence of the ground state of H by an explicit construction.

1. Construction of the ground state

Let

$$\mathbf{B} = \left(\frac{m\omega_0}{2}\right)^{1/2}\left(\mathbf{q} + i\frac{\mathbf{p}}{m\omega_0}\right), \tag{5.1}$$

$$Q^{(r)}_{\mu,\pm} = \tfrac{1}{2}\sqrt{m\omega_0}\,Qe^{(r)}_\mu\left(\frac{\omega_0}{\sqrt{\omega}} \pm \sqrt{\omega}\right). \tag{5.2}$$

Then, using (3.33), (3.34), (3.54), and (3.55), we can write

$$b^{(r)}(f) = \sum_{s=1}^{2}\{a^{(s)*}(W^{(s,r)}_{-}f) + a^{(s)}(W^{(s,r)}_{+}f)\} + (Q^{(r)}_{\mu,-},f)_0 B^*_\mu + (Q^{(r)}_{\mu,+},f)_0 B_\mu, \tag{5.3}$$

$$b^{(r)*}(f) = \sum_{s=1}^{2}\{a^{(s)*}(\overline{W}^{(s,r)}_{+}f) + a^{(s)}(\overline{W}^{(s,r)}_{-}f)\} + (\overline{Q}^{(r)}_{\mu,+},f)_0 B^*_\mu + (\overline{Q}^{(r)}_{\mu,-},f)_0 B_\mu. \tag{5.4}$$

The properties of $T_{\mu\nu}$ given in Lemma 3.6 read

$$W^*_+ W_+ - W^*_- W_- + P_+ - P_- = I, \tag{5.5}$$

$$\overline{W}^*_+ W_- - \overline{W}^*_- W_+ + P_{-+} - P_{+-} = 0, \tag{5.6}$$

$$W_+ W^*_+ - \overline{W}_-\overline{W}^*_- = I, \tag{5.7}$$

$$W_- W^*_+ - \overline{W}_+\overline{W}^*_- = 0, \tag{5.8}$$

where $W_\pm$, $P_\pm$, P_{+-} and P_{-+} are bounded operators on the Hilbert space

$$N_\alpha(R^3) \equiv (M_\alpha(R^3))^2, \tag{5.9}$$

with $\alpha = -1/2,0$, defined by

$$W_\pm = (W^{(r,s)}_\pm),\quad P_\pm = ((Q^{(s)}_{\mu,\pm},\cdot\,)_0 Q^{(r)}_{\mu,\pm}),$$
$$P_{\pm\mp} = ((Q^{(s)}_{\mu,\pm},\cdot\,)_0\overline{Q}^{(r)}_{\mu,\mp}). \tag{5.10}$$

Let

$$K^{(r,s)}f = \frac{1}{\sqrt{\omega}}e^{(r)}_\alpha T_{\alpha\beta}\,\omega d_{\beta\mu}T^*_{\mu\nu}e^{(s)}_\nu\frac{f}{\sqrt{\omega}}. \tag{5.11}$$

Then, by Lemma 3.5, $K^{(r,s)}$ is a bounded operator on $M_\alpha(R^3)$, $\alpha = -\tfrac{1}{2}, 0$. We can also see that

$$K = (K^{(r,s)}) \tag{5.12}$$

is a nonnegative bounded self-adjoint operator on $N_0(R^3)$.

Lemma 5.2: dim Ker $W_+ = 3$. The following serve as the three independent vectors in Ker W_+:

$$\omega_\mu \equiv (1+K)^{-1}\begin{pmatrix}\dfrac{Q}{\sqrt{\omega}}\,e^{(1)}_\mu\\[2mm] \dfrac{Q}{\sqrt{\omega}}\,e^{(2)}_\mu\end{pmatrix},\quad \mu = 1,2,3. \tag{5.13}$$

Proof: By the identity

$$(W_+f)^{(r)} = \frac{1}{2}\sum_{s=1}^{2}\frac{e^{(r)}_\alpha}{\sqrt{\omega}}T^*_{\alpha\beta}e^{(s)}_\beta\sqrt{\omega}\,((1+K)f)^{(s)}$$

and Lemma 3.6-(3), we have

$$(W_+\omega_\mu)^{(r)} = \frac{1}{2}\sum_{s=1}^{2}\frac{e^{(r)}_\alpha}{\sqrt{\omega}}T^*_{\alpha\beta}e^{(s)}_\beta e^{(s)}_\mu Q = 0.$$

Since ω_μ's are independent, it follows that dim Ker $W_+ \geqslant 3$. Let $f \in$ Ker W_+. Then, by Lemma 3.6-(2) and 3.6-(3), we get

$$(1+K)f = m\omega_0^2\sum_{r=1}^{2}(\sqrt{\omega}\,Qe^{(r)}_\mu,f^{(r)})_0\begin{pmatrix}\dfrac{Q}{\sqrt{\omega}}\,e^{(1)}_\mu\\[2mm] \dfrac{Q}{\sqrt{\omega}}\,e^{(2)}_\mu\end{pmatrix}, \tag{5.14}$$

i.e.,

$$f = m\omega_0^2\sum_{r=1}^{2}(\sqrt{\omega}\,Qe^{(r)}_\mu,f^{(r)})_0\,\omega_\mu,$$

which implies that dim Ker $W_+ = 3$. ■

By Lemma 5.1-(3) and (5.3), the ground state Ω, if it exists, is a vector in $\mathscr{H}$ satisfying

$$\sum_{r=1}^{2}\{a^{(r)*}((W_-f)^{(r)}) + a^{(r)}((W_+f)^{(r)}) + (Q^{(r)}_{\mu,-},f^{(r)})_0B^*_\mu + (Q^{(r)}_{\mu,+},f^{(r)})_0B_\mu\}\Omega = 0, \tag{5.15}$$

for all f in $N_0(R^3)\cap N_{-1/2}(R^3)$. This equation is equivalent to the following ones:

$$\left\{\sum_{r=1}^{2}a^{(r)*}((W_-\omega_\mu)^{(r)}) + (\Lambda_-)_{\mu\nu}B^*_\nu + (\Lambda_+)_{\mu\nu}B_\nu\right\}\Omega = 0,\quad \mu = 1,2,3, \tag{5.16}$$

$$\sum_{r=1}^{2}\{a^{(r)*}((W_-f)^{(r)}) + a^{(r)}((W_+f)^{(r)}) + (Q^{(r)}_{\nu,-},f^{(r)})_0B^*_\nu + (Q^{(r)}_{\nu,+},f^{(r)})_0B_\nu\}\Omega = 0, \tag{5.17}$$

where $\Lambda_\pm$ are 3×3 matrices on $\mathbb{C}^3$ given by

$$(\Lambda_\pm)_{\mu\nu} = \sum_{r=1}^{2}(Q^{(r)}_{\nu,\pm},\omega^{(r)}_\mu)_0. \tag{5.18}$$

In order to "solve" these equations, we need further some technical lemmas.

Lemma 5.3: Λ_+ is a strictly positive Hermitian matrix, and Λ_+^{-1} exists.

Proof: By (5.13), we have

$$Q^{(r)}_{\mu,\pm} = \tfrac{1}{2}\sqrt{m\omega_0}\,(\omega_0 \pm \omega)((1+K)\omega_\mu)^{(r)},$$

so that

$$\Lambda_\pm = \Lambda^{(1)} \pm \Lambda^{(2)}$$

where

$$\Lambda^{(1)}_{\mu\nu} = \tfrac{1}{2}\sqrt{m\omega_0}\,(\omega_0(1+K)\omega_\nu,\omega_\mu)_{N_0(R^3)},$$
$$\Lambda^{(2)}_{\mu\nu} = \tfrac{1}{2}\sqrt{m\omega_0}\,(\omega(1+K)\omega_\nu,\omega_\mu)_{N_0(R^3)}.$$

It is clear that $\Lambda^{(1)}$ is a strictly positive Hermitian matrix (note that $K \geqslant 0$). For $\Lambda^{(2)}$ we have by (5.14)

$$\Lambda^{(2)}_{\mu\nu} = \tfrac{1}{2}\sqrt{m\omega_0}\,m\omega_0^2\sum_{r=1}^{2}(\omega^{(r)}_\nu,\sqrt{\omega}\,Qe^{(r)}_\alpha)_0\sum_{s=1}^{2}(\sqrt{\omega}\,Qe^{(s)}_\alpha,\omega^{(s)}_\mu)_0,$$

so that $\Lambda^{(2)} \geqslant 0$. Thus, the lemma follows. ■

Lemma 5.4: Let

$$\Lambda = \Lambda_+^{-1}\Lambda_-. \tag{5.19}$$

Then,

$$\Lambda_{\mu\nu} = \Lambda_{\nu\mu}. \tag{5.20}$$

Proof: By taking the "matrix element" of (5.6) between $\bar{\omega}_\mu$ and ω_ν, we have

$$(\Lambda_+)_{\mu\alpha}(\Lambda_-)_{\nu\alpha} = (\Lambda_-)_{\mu\alpha}(\Lambda_+)_{\nu\alpha}.$$

Since $\Lambda_\pm$ are Hermitian matrices, this implies $\Lambda_+\bar{\Lambda}_- = \Lambda_-\bar{\Lambda}_+$, which, together with the Hermiteness of $\Lambda_\pm$, yields (5.20). ∎

Let $Y = (Y^{(r,s)})$ be an operator on $N_0(R^3)$ given by

$$Y^{(r,s)}f = W_-^{(r,s)} - (\Lambda_+^{-1})_{\mu\nu}(Q_{\mu,+}^{(s)}, f)_0 \times \sum_{p=1}^{2} W_-^{(r,p)}\omega_\nu^{(p)}, \quad f\in M_0(R^3). \tag{5.21}$$

By Lemma 4.1, each $Y^{(r,s)}$ is a Hilbert–Schmidt operator on $M_0(R^3)$. It follows from (5.7) that $(W_+W_+^*)^{-1}$ exists and hence that $W_+ \upharpoonright (\mathrm{Ker}\, W_+)^\perp$ is a one-to-one map from $(\mathrm{Ker}\, W_+)^\perp$ onto $N_0(R^3)$. Therefore

$$Z \equiv (W_+ \upharpoonright (\mathrm{Ker}\, W_+)^\perp)^{-1} \tag{5.22}$$

is a bounded operator on $N_0(R^3)$ with $\mathrm{Ran}\, Z = (\mathrm{Ker}\, W_+)^\perp$. Put

$$C = YZ. \tag{5.23}$$

Then, each $C^{(r,s)}$ is a Hilbert–Schmidt operator on $M_0(R^3)$.

Lemma 5.5: Let $C^{(r,s)}(\mathbf{k},\mathbf{k}')$ denote the Hilbert–Schmidt kernel of the operator $C^{(r,s)}$. Then,

$$C^{(r,s)}(\mathbf{k},\mathbf{k}') = C^{(s,r)}(\mathbf{k}',\mathbf{k}). \tag{5.24}$$

Proof: (5.24) is equivalent to $C^* = \bar{C}$, i.e.,

$$Z^*Y^* = \bar{Y}\bar{Z}. \tag{5.25}$$

By using (5.6), we can see that

$$W_+^*\bar{Y}f = W_-^*\bar{W}_+f - \sum_{r=1}^{2}\{(\bar{Q}_{\mu,-}^{(r)}, f^{(r)})_0 - \bar{\Lambda}_{\nu\mu}(\bar{Q}_{\nu,+}^{(r)}, f^{(r)})_0\}Q_{\mu,+}.$$

Also we have

$$Y^*\bar{W}_+f = W_-^*\bar{W}_+f - \sum_{r=1}^{2}\{(\bar{Q}_{\mu,-}^{(r)}, f^{(r)})_0 - \bar{\Lambda}_{\mu\nu}(\bar{Q}_{\nu,+}^{(r)}, f^{(r)})_0\}Q_{\mu,+}.$$

Therefore, by Lemma 5.4, we get

$$W_+^*\bar{Y} = Y^*\bar{W}_+,$$

which, combined with $W_+Z = I$, yields. (5.25). ∎

We now proceed to construct the ground state. Equations (5.16) and (5.17) are equivalent to

$$\left\{\sum_{r=1}^{2}(\Lambda_+^{-1})_{\mu\nu}a^{(r)*}((W_-\omega_\nu)^{(r)}) + \Lambda_{\mu\nu}B_\nu^* + B_\mu\right\}\Omega = 0, \quad \mu = 1,2,3, \tag{5.26}$$

$$\sum_{r=1}^{2}\{a^{(r)*}((Cf)^{(r)}) + a^{(r)}(f^{(r)}) + ((Z^*u_\mu)^{(r)}, f^{(r)})_0 B_\mu^*\}\Omega = 0, \quad f\in N_0(R^3), \tag{5.27}$$

where

$$u_\mu^{(r)} = Q_{\mu,-}^{(r)} - \bar{\Lambda}_{\mu\nu}Q_{\nu,+}^{(r)}, \quad \mu = 1,2,3. \tag{5.28}$$

Let

$$\Omega_0 = \psi_0 \otimes \Omega_F, \tag{5.29}$$

where ψ_0 is the ground state of the free harmonic oscillator:

$$\psi_0(\mathbf{q}) = \left(\frac{\pi}{m\omega_0}\right)^{-3/4} e^{-(m\omega_0/2)\mathbf{q}^2}. \tag{5.30}$$

Then, it follows that

$$a^{(r)}(f)\Omega_0 = 0, \quad B_\mu\Omega_0 = 0, \quad r = 1,2, \quad \mu = 1,2,3, \tag{5.31}$$

for all f in $M_0(R^3)$. Put

$$V = -\frac{1}{2}\sum_{r,s=1}^{2}\int d^3\mathbf{k}d^3\mathbf{k}'C^{(r,s)}(\mathbf{k},\mathbf{k}')a^{(r)*}(\mathbf{k})a^{(s)*}(\mathbf{k}') - \sum_{r=1}^{2}a^{(r)*}(\overline{(Z^*u_\mu)}^{(r)})B_\mu^* - \frac{1}{2}\Lambda_{\mu\nu}B_\mu^*B_\nu^*. \tag{5.32}$$

Since $C^{(r,s)}(\cdot\,,\,\cdot)\in L^2(R^6)$ and $(Z^*u_\mu)^{(r)}\in M_0(R^3)$, the operator V is well-defined on D_F and leaves it invariant. Furthermore, we can show that D_F is a set of analytic vectors for V. Thus, we can define the vector

$$\Omega = c\sum_{n=0}^{\infty}\frac{V^n\Omega_0}{n!}, \tag{5.33}$$

where $c>0$ is the normalization constant.

We can now prove

Lemma 5.6: The ground state of H exists and is equal to Ω given in (5.33) up to scalar multiples.

Proof: We need only to show that Ω satisfies (5.16) and (5.17). It is easy to see that

$$[V, a^{(r)}(f)]\Psi = \left\{\sum_{s=1}^{2}a^{(s)*}(C^{(s,r)}f) + ((Z^*u_\mu)^{(r)}, f)_0 B_\mu^*\right\}\Psi$$

for all f in $M_0(R^3)$ and all Ψ in D_F. Therefore, we have

$$\sum_{r=1}^{2}a^{(r)}(f^{(r)})V^n\Omega_0 = -n\left\{\sum_{r=1}^{2}a^{(r)*}((Cf)^{(r)}) + (Z^*u_\mu, f)_{N_0(R^3)}B_\mu^*\right\}V^{n-1}\Omega_0$$

for all f in $N_0(R^3)$ and $n\geqslant 1$, yielding (5.17). Similarly, we get

$$\left\{\sum_{r=1}^{2}a^{(r)*}(\overline{(Z^*u_\mu)}^{(r)}) + \Lambda_{\mu\nu}B_\nu^* + B_\mu\right\}\Omega = 0.$$

But, by (5.6) and $W_+Z = I$, we see that

$$\overline{Z^*u_\mu} = (\Lambda_+^{-1})_{\mu\nu}W_-\omega_\nu.$$

Therefore, (5.16) follows. ∎

2. Proof of Theorem B-(1)

By Lemmas 5.1 and 5.6, we have $\sigma_p(H) = \{E_0\}$, where E_0 is simple. Then, in the same method as in Refs. 27–29, we can show that $\hat{H}$ is unitarily equivalent to $H_0^{\rm EM}$ *acting in* $\mathscr{F}_0^{\rm EM}$. Thus, we get the theorem. ■

B. The case $m<\delta m$: Proof of Theorem B-(2)

Lemma 5.7: Let

$$\alpha_\mu = \frac{1}{\sqrt{2}}(C_\mu + iD_\mu), \quad \mu=1,2,3. \tag{5.34}$$

Then, there exists a normalized vector $\Omega_{\rm phys}$ in $\mathscr{H}$ such that

$$\alpha_\mu \Omega_{\rm phys} = b^{(r)}(f)\Omega_{\rm phys} = 0 \tag{5.35}$$

for all f in $M_0(R^3)$ and for $r=1,2$, $\mu=1,2,3$. The vector $\Omega_{\rm phys}$ is unique up to scalar multiples.

Proof: By (3.35) and (3.36), we have

$$\alpha_\mu = a_+ B_\mu - a_- B_\mu^* + b\hat{A}_\nu(F_{\mu\nu}) + c\hat{\pi}_\nu(F_{\mu\nu}), \tag{5.36}$$

where

$$a_\pm = \left(\frac{m\omega_0}{2E_B}\right)^{1/2} \frac{(1+i)\gamma_B}{2}\left(1 \pm \frac{\omega_0}{E_B}\right), \tag{5.37}$$

$$b = [(1-i)/2]E_B^{3/2}\gamma_B, \tag{5.38}$$

$$c = [(1+i)/2]E_B^{1/2}\gamma_B. \tag{5.39}$$

Let

$$U_-^{(s,r)} f = W_-^{(s,r)} f - \frac{1}{\sqrt{2}a}(Q_{\mu,+}^{(r)}, f)_0 \left(\frac{b}{\sqrt{\omega}} + ic\sqrt{\omega}\right) F_{\mu\nu} e_\nu^{(s)}, \tag{5.40}$$

$$U_+^{(s,r)} f = W_+^{(s,r)} f - \frac{1}{\sqrt{2}a}(Q_{\mu,+}^{(r)}, f)_0 \times\left(\frac{b}{\sqrt{\omega}} - ic\sqrt{\omega}\right) F_{\mu\nu} e_\nu^{(s)}, \quad f\in M_0(R^3). \tag{5.41}$$

and define

$$c^{(r)}(f) = \sum_{s=1}^{2} \{a^{(s)*}(U_-^{(s,r)} f) + a^{(s)}(U_+^{(s,r)} f)\}, \tag{5.42}$$

$$c^{(r)*}(f) = \sum_{s=1}^{2} \{a^{(s)*}(\overline{U}_+^{(s,r)} f) + a^{(s)}(\overline{U}_-^{(s,r)} f)\}. \tag{5.43}$$

Then, the properties of $T_{\mu\nu}$ given in Lemma 3.6 read

$$U_+^* U_+ - U_-^* U_- = I, \tag{5.44}$$

$$\overline{U}_+^* U_- - \overline{U}_-^* U_+ = 0, \tag{5.45}$$

$$U_+ U_+^* - \overline{U}_- \overline{U}_-^* = I, \tag{5.46}$$

$$U_- U_+^* - \overline{U}_+ \overline{U}_-^* = 0, \tag{5.47}$$

where $U_\pm$ are bounded operators on $N_0(R^3)$ given by

$$U_\pm = (U_\pm^{(r,s)}). \tag{5.48}$$

Further, by Lemma 4.1 and the fact that $Q_{\mu,+}^{(r)} \otimes (b/\sqrt{\omega} + ic\sqrt{\omega})F_{\mu\nu}e_\nu^{(s)}$ is in $L^2(R^6)$, U_- is a Hilbert–Schmidt operator on $N_0(R^3)$. Therefore, (5.42) and (5.43) give a proper linear canonical transformation (see Ref. 30, Chap II) and hence there exists a unique (up to scalar multiples) normalized vector Ψ_0 in $\mathscr{F}^{\rm EM}$ satisfying

$$c^{(r)}(f)\Psi_0 = 0, \quad r=1,2, \quad f\in M_0(R^3). \tag{5.49}$$

Let

$$T = \frac{1}{2a_+}\{a_- B_\mu^{*2} - 2[b\hat{A}_\nu(F_{\nu\mu}) + c\hat{\pi}_\nu(F_{\nu\mu})]B_\mu^*\}. \tag{5.50}$$

Then, we can see that the vector

$$\Omega_{\rm phys} \equiv N \sum_{n=0}^{\infty} \frac{T^n}{n!} \psi_0 \otimes \Psi_0 \tag{5.51}$$

is well-defined with N being the normalized constant and that it satisfies (5.35).

We next prove the uniqueness of $\Omega_{\rm phys}$. Suppose that there exists another normalized vector Ψ satisfying (5.35). Without loss of generality, we assume $(\Omega_{\rm phys}, \Psi) = 0$. Let $\tilde{\mathscr{H}}$ be the closure of the linear subspace

$$\{b^{(r_1)*}(f_1)\ldots b^{(r_n)*}(f_n)\alpha_{\mu_1}^* \ldots \alpha_{\mu_m}^* \Omega_{\rm phys}, \Omega_{\rm phys} \mid f_1,\ldots,f_n \in M_0(R^3), \; r_j = 1,2, \; \mu_i = 1,2,3, n,m\geqslant 1\}$$

Then, it is easy to see that Ψ is orthogonal to all vectors in $\tilde{\mathscr{H}}$. On the other hand, by (3.57)–(3.60), $\tilde{\mathscr{H}}$ is invariant by the algebra $A \equiv \{a^{(r)\#}(f), \mathbf{p},\mathbf{q} \mid f\in M_0(R^3), r=1,2,\}$, which is irreducible in $\mathscr{H}$ (see Ref. 30, Chap. I). Hence we have

$$\tilde{\mathscr{H}} = \mathscr{H}. \tag{5.52}$$

Therefore Ψ must be 0-vector. But this is a contradiction. ■

Proof of Theorem B-(2):

By (5.52) we can define a unitary operator U on $\mathscr{H}$ such that

$$U\Omega_{\rm phys} = \Omega_0, \tag{5.53}$$

$$Ub^{(r)\#}(f)U^{-1} = a^{(r)\#}(f), \quad f\in M_0(R^3), \quad r=1,2, \tag{5.54}$$

$$U\alpha_\mu U^{-1} = B_\mu, \quad \mu=1,2,3. \tag{5.55}$$

Using commutation relations (3.49) and (3.50) and the irreducibility of the algebra A, we can show that

$$UHU^{-1} = \frac{E_B}{2}(\mathbf{p}\cdot\mathbf{q} + \mathbf{q}\cdot\mathbf{p}) + H_0^{\rm EM} + \text{const.} \tag{5.56}$$

Thus, by Proposition A2 in Appendix B, we obtain the desired result. ■

VI. CHARACTERIZATION OF THE "ENERGY LEVEL SHIFTS" AND THE "DECAY PROBABILITIES"

A. 2-point τ-functions, proof of Theorem C

We first consider the 2-point functions in the case $m>\delta m$:

$$\mathscr{W}_{\mu\nu}^{(h)}(t-s) = (\Omega, q_\mu(t)q_\nu(s)\Omega), \tag{6.1}$$

$$\mathscr{W}_{\mu\nu}^{(\rm EM)}(f,g;t-s) = (\Omega, A_\mu(f,t)A_\nu(g,s)\Omega), \quad f,g\in\mathscr{S}(R^3), \tag{6.2}$$

$$\mathscr{W}_{\mu\nu}^{(\rm EM,h)}(f;t-s) = (\Omega, A_\mu(f,t)p_\nu(s)\Omega), \quad f\in\mathscr{S}(R^3), \tag{6.3}$$

By (3.33), (3.34), and (3.40) we can show that Ω is in $C^\infty(b^{(r)\#}(f))$ for all f in $M_{-1/2}(R^3)\cap M_0(R^3)$ and therefore that all the 2-point functions are well-defined.

Remark: The general n-point functions are also well-defined and can be written as a sum of products of the 2-point functions.

We can evaluate the 2-point functions by (3.40), Theorem 3.1 and the fact that $b^{(r)}(f)\Omega = 0$, $f\in M_{-1/2}(R^3)\cap M_0(R^3)$:

$$\mathscr{W}_{\mu\nu}^{(h)}(t) = \tfrac{1}{3}\delta_{\mu\nu}(Qe^{i\omega t}, Q)_{1/2}, \tag{6.4}$$

$$\mathscr{W}_{\mu\nu}^{(\rm EM)}(f,g;t) = \tfrac{1}{2}(e^{i\omega t}T_{\nu\beta}\hat{g}, d_{\beta\alpha}\widetilde{T_{\mu\alpha}\hat{f}})_{-1/2}, \tag{6.5}$$

$$\mathscr{W}_{\mu\nu}^{(\rm EM,h)}(f;t) = \frac{im\omega_0^2}{2}(Qe^{i\omega t}, d_{\nu\alpha}T_{\alpha\mu}\hat{f})_{-1/2}. \tag{6.6}$$

Employing the formula

$$\frac{1}{E}\{\theta(t)e^{-iEt}+\theta(-t)e^{iEt}\}$$
$$=\lim_{\epsilon\to+0}\frac{i}{\pi}\int_{-\infty}^{\infty}dE'\frac{e^{-iE't}}{E'^2-E^2+i\epsilon},$$

[$\theta(t)$: the Heaviside function]

we obtain

$$\tau_{\mu\nu}^{(h)}(t)=\frac{1}{2\pi i}\delta_{\mu\nu}\int_{-\infty}^{\infty}dE\frac{e^{-iEt}}{D_+(E^2)},\tag{6.7}$$

$$\tau_{\mu\nu}^{(\mathrm{EM})}(\mathbf{x},\mathbf{y};t)=\frac{i}{(2\pi)^4}\int_{-\infty}^{\infty}dE\left\{K_{\mu\nu}(E;\mathbf{x}-\mathbf{y})-\frac{e^2J_{\mu\alpha}(E;\mathbf{x})J_{\nu\alpha}(E;\mathbf{y})}{D_+(E^2)}\right\}e^{-iEt},\tag{6.8}$$

$$\tau_{\mu\nu}^{(\mathrm{EM},h)}(\mathbf{x};t)=\frac{im\omega_0^2e}{(2\pi)^{5/2}}\int_{-\infty}^{\infty}dE\frac{J_{\mu\nu}(E;\mathbf{x})}{D_+(E^2)}e^{-iEt},\tag{6.9}$$

where

$$K_{\mu\nu}(E;\mathbf{x})=\lim_{\epsilon\to+0}\int d^3\mathbf{k}\frac{d_{\mu\nu}(\mathbf{k})e^{-i\mathbf{kx}}}{E^2-\mathbf{k}^2+i\epsilon},\tag{6.10}$$

$$J_{\mu\nu}(E;\mathbf{x})=\lim_{\epsilon\to+0}\int d^3\mathbf{k}\frac{\hat\rho(\mathbf{k})d_{\mu\nu}(\mathbf{k})e^{-i\mathbf{kx}}}{E^2-\mathbf{k}^2+i\epsilon}.\tag{6.11}$$

To prove Theorem C we need

Lemma 6.1: Let $\hat\rho$ satisfy assumptions (AI) and (AII). Then:

(1) $K_{\mu\nu}(E;\mathbf{x})$ and $J_{\mu\nu}(E;\mathbf{x})$ have analytic continuations as functions of E from $(0,\infty)$ onto $\mathbb{C}_+$.

(2) $D_+(E^2)$ has an analytic continuation as a function of E from $(0,\infty)$ onto $\mathbb{C}_+$ and there exists a constant $\lambda>0$ such that, if $|e|<\lambda$, then the analytic continuation has a unique simple zero $\eta(e)$ in Π_-, which is analytic in e, satisfying $\eta(e)\to\omega_0$ as $e\to0$. Let

$$\eta(e)=\omega_0+b_1e+b_2e^2+\cdots,\quad |e|<\lambda.$$

Then,

$$b_1=0,\tag{6.12}$$

$$b_2=\frac{\omega_0\delta m}{2me^2}+\frac{\omega_0}{3m}P\int d^3\mathbf{k}\frac{\hat\rho(\mathbf{k})^2}{\omega_0^2-\mathbf{k}^2}-\frac{i2\pi^2\omega_0^2\hat\rho(\omega_0)^2}{3m}.\tag{6.13}$$

Proof: Since the proof of (1) is fairly easy, we prove only (2). Since (3.9) can be written as

$$D_\pm(E^2)=m\omega_0^2-E^2\left(m-\delta m-\frac{2}{3}e^2P\int d^3\mathbf{k}\frac{\hat\rho(\mathbf{k})^2}{E^2-\mathbf{k}^2}\right)\mp\tfrac{4}{3}i\pi^2e^2E^3\hat\rho(E)^2,\quad E>0,$$

the function

$$D_{\mathrm{II}}(z)=D(z^2)-\tfrac{8}{3}i\pi^2e^2z^3\hat\rho(z)^2$$

defines an analytic continuation of $D_+(E^2)$ from $(0,\infty)$ onto Π_-. Therefore, the function

$$F(z)=\begin{cases}D(z^2), & z\in\Pi_+,\\ D_+(z^2), & z\in(0,\infty),\\ D_{\mathrm{II}}(z), & z\in\Pi_-\end{cases}$$

gives an analytic continuation of $D_+(E^2)$ from $(0,\infty)$ onto $\mathbb{C}_+$. We can rewrite $F(z)$ as

$$F(z)=m\omega_0^2-mz^2+e^2(e^{-2}\delta mz^2+\phi(z))\equiv F(z,e),\tag{6.14}$$

where $\phi(z)$ is an analytic function in $\mathbb{C}_+$. Therefore, $F(z,e)$ is analytic in $\mathbb{C}_+\times\mathbb{C}$ as a function of two complex variables (z,e) and $F(z,0)$ has the unique simple zero at $z=\omega_0$. Thus, by the Weierstrass preparation theorem (see, e.g., Ref. 31, p. 188), there exists a complex neighborhood $\mathscr{N}$ of $(\omega_0,0)$ and a function $f(z,e)$ which is analytic and nonvanishing in $\mathscr{N}$ such that $F(z,e)$ can be expressed in the form

$$F(z,e)=(z-\eta(e))f(z,e)\quad\text{in }\mathscr{N},$$

where $\eta(e)$ is analytic in a neighborhood $\mathscr{U}$ of $e=0$ and $\eta(0)=\omega_0$. Taking Lemma 3.2 into account, we conclude that $\eta(e)\in\Pi_-$ for real $e\in\mathscr{U}$ and $D_{\mathrm{II}}(\eta(e))=0$.

We next prove the uniqueness of $\eta(e)$. Suppose

$$F(z_0(e))=0,\quad e\in R^1.$$

Then, it can be shown that, if $|e|<\delta$, then

$$\left|\left(1-\frac{\delta m}{m}\right)z_0(e)^2-\omega_0^2\right|\leqslant e^2c(\delta)\tag{6.15}$$

for some constant $c(\delta)$ depending on δ, where we have used the asymptotic behavior of $\hat\rho(z)$ at ∞ assumed in (AII). It follows from (6.15) that there exists a constant $\lambda>0$ such that, if $|e|<\lambda$, then $(z_0(e),e)$ is contained in $\mathscr{N}$ and $z_0(e)\to\omega_0$ as $e\to0$. Therefore, $z_0(e)$ must be equal to $\eta(e)$ for $|e|<\lambda$. Equations (6.12) and (6.13) follow from (6.14) by direct computations. ■

Lemma 6.2: The function $\epsilon_{n,j}(z)$ defined by (1.26) has an analytic continuation from Π_+ onto $\overline{\Pi}_-\cap\{z\in\mathbb{C}|\mathrm{Re}\,(z)\in(E_n^{(0)}-\omega_0,E_n^{(0)}+\omega_0)\}$. In particular, $\epsilon_{n,j}$ in (1.25) exists and is given explicitly as

$$\epsilon_{n,j}=\frac{e^2\omega_0}{3m}\left\{nP\int d^3\mathbf{k}\frac{\hat\rho(\mathbf{k})^2}{\omega_0^2-\mathbf{k}^2}-\frac{3}{2}\int d^3\mathbf{k}\frac{\hat\rho(\mathbf{k})^2}{|\mathbf{k}|(|\mathbf{k}|+\omega_0)}\right\}-\frac{i2\pi^2e^2\omega_0^2n\hat\rho(\omega_0)^2}{3m},\quad j=1,\dots,M.\tag{6.16}$$

Proof: By direct computations, we can show that

$$\epsilon_{n,j}(z)=-\frac{2\pi e^2\omega_0}{3m}\{(n+3)J_+(z)+nJ_-(z)\},\quad z\in\mathbb{C}\setminus[\tfrac{3}{2}\omega_0,\infty),\tag{6.17}$$

where

$$J_\pm(z)=\int_0^\infty dE\frac{E\hat\rho(E)^2}{E\pm\omega_0+E_n^{(0)}-z}.\tag{6.18}$$

It is easy to see that the function

$$J_-(z)_{\mathrm{II}}=J_-(z)+2\pi i(z+\omega_0-E_n^{(0)})\hat\rho(z+\omega_0-E_n^{(0)})$$

defines an analytic continuation of $J_-(z)$ from Π_+ onto $\overline{\Pi}_+\cap\{z\in\mathbb{C}|\mathrm{Re}\,(z)>E_n^{(0)}-\omega_0\}$. On the other hand, $J_+(z)$ is analytic in $\mathbb{C}\setminus[E_n^{(0)}+\omega_0,\infty)$. Thus, the first half of the lemma follows. (6.16) can be obtained from (6.17). ■

Proof of Theorem C: The first half of the theorem follows directly from (6.7)–(6.9) and Lemma 6.1, where $\zeta(e)$ is taken to be $\eta(e)$ itself. (1.43) follows from (6.13) and (6.16). To prove (1.44) we compute $E^{(2)}_{n,j}(e)$ explicitly:

$$E^{(2)}_{n,j}(e) = \left(1 + \frac{\delta m}{2m}\right)E^{(0)}_n + \mathrm{Re}\,(\epsilon_{n,j}),$$

so that

$$\delta E^{(2)}_{n,j}(e) - \delta E^{(2)}_{0,1}(e) = \frac{n\omega_0 \delta m}{2m} + \mathrm{Re}\,(\epsilon_{n,j} - \epsilon_{0,1}),$$

which, together with (6.13) and (6.16), yields (1.44). ■

B. Resonance pole of S-matrix

In this subsection we show that $\zeta(e)$ in Theorem C is also the resonance pole of the S-matrix for the photon scattering by the atom.

The Hilbert space $\mathscr{H}_{\substack{\text{in}\\ \text{out}}}$ of the photon scattering states are defined by

$$\mathscr{H}_{\substack{\text{in}\\ \text{out}}} = \mathscr{L}\{a^{(r_1)*}_{\substack{\text{in}\\ \text{out}}}(f_1)\cdots a^{(r_n)*}_{\substack{\text{in}\\ \text{out}}}(f_n)\Omega \mid f_j \in M_{-1/2}(R^3)\cap M_0(R^3),$$
$$r_j = 1,2,\ n = 0,1,\cdots\}, \tag{6.19}$$

where $\mathscr{L}\{\ \cdots\ \}$ denotes the closure of the linear span of $\{\ \cdots\ \}$. By relations (3.57)–(3.60) and the irreducibility of the algebra $\{a^{(r)\#}(f), \mathbf{p}, \mathbf{q} \mid r = 1,2, f\in M_{-1/2}(R^3)\cap M_0(R^3)\}$ in $\mathscr{H}$, we can prove the asymptotic completeness:

$$\mathscr{H} = \mathscr{H}_{\text{in}} = \mathscr{H}_{\text{out}}. \tag{6.20}$$

Then, the S-matrix S for the photon scattering is defined as an operator from $\mathscr{H}_{\text{in}}$ to $\mathscr{H}_{\text{out}}$ by

$$S a^{(r_1)*}_{\text{in}}(f_1)\cdots a^{(r_n)*}_{\text{in}}(f_n)\Omega = a^{(r_1)*}_{\text{out}}(f_1)\cdots a^{(r_n)*}_{\text{out}}(f_n)\Omega. \tag{6.21}$$

It follows from Theorem 4.1 and (6.20) that S is unitary. The n-photon S-matrix element is given by

$$S^{(n)}_{r_1\cdots r_n;s_1\cdots s_n}(f_1,\cdots,f_n;g_1,\cdots,g_n)$$
$$= (S a^{(r_1)*}_{\text{in}}(f_1)\cdots a^{(r_n)*}_{\text{in}}(f_n)\Omega, a^{(s_1)*}_{\text{in}}(g_1)\cdots a^{(s_n)*}_{\text{in}}(g_n)\Omega). \tag{6.22}$$

By commutation relations, we can show that the n-photon S-matrix element can be written as a sum of products of the one photon S-matrix element which is given explicitly by

$$S^{(1)}_{r,s}(f;g) = \int S^{(1)}_{r,s}(\mathbf{k};\mathbf{k}')\overline{f(\mathbf{k})}g(\mathbf{k}')\,d^3\mathbf{k}\,d^3\mathbf{k}', \tag{6.23}$$

with

$$S^{(1)}_{r,s}(\mathbf{k};\mathbf{k}') = \delta_{rs}\delta^3(\mathbf{k}-\mathbf{k}') - 2\pi i\delta(|\mathbf{k}| - |\mathbf{k}'|)\mathscr{T}(|\mathbf{k}|)e^{(r)}_\mu(\mathbf{k})e^{(s)}_\mu(\mathbf{k}'), \tag{6.24}$$

$$\mathscr{T}(E) = -\frac{e^2E\hat{\rho}(E)^2}{2D_+(E^2)}, \quad E > 0. \tag{6.25}$$

We can also show that all the off-diagonal matrix elements of S are zero.

Theorem 6.1: (1) The function $\mathscr{T}(E)$ originally defined in $(0,\infty)$ has a meromorphic continuation onto $\mathbb{C}_+$ which, for $|e| < \lambda$, has a unique pole $\zeta(e)$, where λ and $\zeta(e)$ are the same in Theorem C.

(2) Let $\zeta(e) = E(e) - i\Gamma(e)/2$, so that $\Gamma(e) > 0$, and let $\{I_e \mid e > 0\}$ be an arbitrary family of open intervals containing $E(e)$ and such that the Lebesgue measure of I_e is $O(\Gamma(e)^{1/2})$ as $e \to +0$. Then, for $E\in I_e$,

$$|\mathscr{T}(E)|^2 = \frac{3e^2\hat{\rho}(E)^2}{32\pi^2 E}\left\{\frac{\gamma(e)\Gamma(e)}{(E - E(e))^2 + \frac{1}{4}\Gamma(e)^2} + O(1)\right\}$$

as $e \to +0$, where

$$\gamma(e) = -\mathrm{Re}\,[D'_{\mathrm{II}}(\zeta(e))]^{-1}.$$

The proof is quite similar to that of (II, Theorem 3) and hence is omitted. (See also Ref. 14, Theorem 4).

The theorem shows that $\zeta(e)$ is the resonance pole of the S-matrix.

VII. BROKEN SYMMETRY ASPECTS

From now on we assume $m < \delta m$.

We have seen that, if $m < \delta m$, then the total Hamiltonian H is not bounded below [Theorems A.2 and B-(2)] and the unphysical mode $\{\alpha_\mu, \alpha^*_\mu\}$ (or $\{C_\mu, D_\mu\}$) appears. This phenomenon corresponds to the explicit symmetry breaking of H with respect to the linear canonical transformations of the dynamical variables which leave the quadratic form

$$\frac{1}{2|m - \delta m|}(\mathbf{p} - e\mathbf{A}(\rho))^2 + \frac{m\omega_0^2}{2}\mathbf{q}^2 + H^{\mathrm{EM}}_0$$

invariant. Therefore, the unphysical mode may be regarded as a kind of "Goldstone boson." The following is a procedure ("Higgs Mechanism") by which the unphysical mode disappears and the physical theory can be obtained.

Let $\mathscr{H}_{\text{phys}}$ be the closure of the linear subspace

$$\mathscr{H}^0_{\text{phys}} = \{b^{(r_1)*}(f_1)\cdots b^{(r_n)*}(f_n)\Omega_{\text{phys}}, \Omega_{\text{phys}} \mid f_1,\ldots,f_n \in M_0(R^3),$$
$$r_j = 1,2,\ n \geqslant 1\}, \tag{7.1}$$

and P be the orthogonal projection onto $\mathscr{H}_{\text{phys}}$. Then, by restricting the dynamical variables within $\mathscr{H}_{\text{phys}}$, we can see that the unphysical mode disappears. In fact, by (3.41), (3.52), and (3.53), we have

$$A^P_\mu(f,t) \equiv PA_\mu(f,t)P = \frac{1}{\sqrt{2}}\sum_{r=1}^{2}\left\{b^{(r)*}\left(\frac{e^{(r)}_\nu}{\sqrt{\omega}}e^{it\omega}\overline{T}_{\nu\mu}\hat{f}\right) + b^{(r)}\left(\frac{e^{(r)}_\nu}{\sqrt{\omega}}e^{-it\omega}T_{\nu\mu}\hat{f}\right)\right\}, \tag{7.2}$$

$$q^P_\mu(t) \equiv Pq_\mu(t)P = \frac{1}{\sqrt{2}}\sum_{r=1}^{2}\left\{b^{(r)*}(\sqrt{\omega}\,\overline{Q}e^{(r)}_\mu e^{it\omega}) + b^{(r)}(\sqrt{\omega}\,Qe^{(r)}_\mu e^{-it\omega})\right\}. \tag{7.3}$$

The projected fields $\mathbf{A}^P(f,t)$ and $\mathbf{q}^P(t)$ satisfy the field equations (3.1) in the operator-valued distribution sense. But the equal time commutation relations for the projected fields become noncanonical, which is a compensation for the disappearance of the unphysical mode.

We can also define the physical Hamiltonian H_{phys} in $\mathscr{H}_{\text{phys}}$ by

$$H_{\text{phys}} = PHP, \tag{7.4}$$

which is unitarily equivalent to H^{EM}_0 + const. in $\mathscr{F}^{\mathrm{EM}}$ [see Proof of Theorem B-(2)] and, in particular, satisfies

$$H_{\text{phys}}\Omega_{\text{phys}} = E\Omega_{\text{phys}} \tag{7.5}$$

for some real constant E.

Now we can construct the scattering theory in $\mathscr{H}_{\rm phys}$. Let

$$\tilde{a}_t^{(r)\#}(f) = e^{itH_{\rm phys}}Pe^{-itH_0}a^{(r)\#}(f)e^{itH_0}Pe^{-itH_{\rm phys}}, \quad f \in M_0(R^3),\ r=1,2, \tag{7.6}$$

which are well-defined on $\mathscr{H}^0_{\rm phys}$ [see (3.57) and (3.58)]. Then, we have

Theorem 7.1: For all Ψ in $\mathscr{H}^0_{\rm phys}$ and all f in $M_0(R^3)$, the strong limits

$$\underset{t\to\pm\infty}{\text{s-lim}}\ \tilde{a}_t^{(r)\#}(f)\Psi \equiv \tilde{a}^{(r)\#}_{\substack{\rm out\\ \rm in}}(f)\Psi \tag{7.7}$$

exist and are given explicitly by

$$\tilde{a}^{(r)\#}_{\rm in}(f) = b^{(r)\#}(f), \tag{7.8}$$

$$\tilde{a}^{(r)}_{\rm out}(f) = \sum_{s=1}^{2} b^{(s)}(L^{(r,s)}f), \tag{7.9}$$

$$\tilde{a}^{(r)*}_{\rm out}(f) = \sum_{s=1}^{2} b^{(s)*}(\overline{L}^{(r,s)}f), \tag{7.10}$$

where the operator $L^{(r,s)}$ is given by (4.6).

The proof is quite similar to that of Theorem 4.1

Theorem 7.1 permits us to develop the scattering theory in $\mathscr{H}_{\rm phys}$, but we omit the details (cf. Sec. VI-B).

VIII. POINT LIMIT—REMOVAL OF ULTRAVIOLET CUTOFF AND INFINITE MASS-RENORMALIZATION

In this section we consider the point limit, $\rho(\mathbf{x})\to\delta(\mathbf{x})$, of the interaction, which corresponds to the removal of the ultraviolet cutoff in momentum space. Since $\delta m\to\infty$ as $\rho(\mathbf{x})\to\delta(\mathbf{x})$, the point limit requires the infinite mass-renormalization.

We shall take the point limit in terms of the Wightman distributions $\{W^{(n,m)}_{\mu_1\cdots\mu_n;\nu_1\cdots\nu_m}\}_{n,m=1}^{\infty}$ for the projected fields in $\mathscr{H}_{\rm phys}$ defined by

$$W^{(n,m)}_{\mu_1\cdots\mu_n;\nu_1\cdots\nu_m}(f_1,t_1,\cdots,f_n,t_n;s_1,\ldots,s_m) = (\Omega_{\rm phys}, A^P_{\mu_1}(f_1,t_1)\cdots A^P_{\mu_n}(f_n,t_n)q^P_{\nu_1}(s_1)\cdots q^P_{\nu_m}(s_m)\Omega_{\rm phys}), \quad f_1,\ldots,f_n\in\mathscr{S}(R^3). \tag{8.1}$$

Using (7.2) and (7.3), one can easily see that, if $n+m$ is odd, then $W^{(n,m)}_{(\ldots)}$ is zero and, if $n+m$ is even, then it is written as a sum of products of the 2-point distributions given explicitly by

$$W^{(2,0)}_{\mu_1\mu_2}(f_1,t_1,f_2,t_2) = \tfrac{1}{2}(\overline{T}_{\mu_1\alpha}\hat{f}_1, d_{\alpha\beta}e^{-i(t_1-t_2)\omega}\overline{T}_{\beta\mu_2}\hat{f}_2)_{-1/2}, \tag{8.2}$$

$$W^{(1,1)}_{\mu;\nu}(f,t;s) = \tfrac{1}{2}(\overline{T}_{\mu\alpha}\hat{f}, d_{\alpha\nu}e^{-i(t-s)\omega}\overline{Q})_0, \tag{8.3}$$

$$W^{(0,2)}_{\nu_1\nu_2}(s_1,s_2) = \tfrac{1}{2}(Q, d_{\nu_1\nu_2}e^{-i(s_1-s_2)\omega}Q)_{1/2}. \tag{8.4}$$

We mean by $\rho\to\delta$ that $\hat{\rho}\to(2\pi)^{-3/2}$ with $\{\hat{\rho}\}$ being uniformly bounded.

Lemma 8.1:

(1) For each $\mathbf{k}\in R^3$,

$$\lim_{\rho\to\delta} D_+(\mathbf{k}^2) = m\omega_0^2 - \mathbf{k}^2\left(m + \frac{ie^2|\mathbf{k}|}{6\pi}\right) \equiv D^R_+(\mathbf{k}^2). \tag{8.5}$$

(2) Let

$$Q^R(\mathbf{k}) = \frac{ie}{(2\pi)^{3/2}D^R_+(\mathbf{k}^2)}. \tag{8.6}$$

Then,

$$\lim_{\rho\to\delta}\|Q - Q^R\|_{1/2} = 0. \tag{8.7}$$

(3) Let $T^R_{\mu\nu}$ be an operator given by

$$T^R_{\mu\nu}f = \delta_{\mu\nu}f - \frac{e^2\omega^{5/2}Gd_{\mu\nu}\sqrt{\omega}f}{(2\pi)^3D^R_+}. \tag{8.8}$$

Then, for all f in $M_0(R^3)$,

$$\lim_{\rho\to\delta}\|T_{\mu\nu}f - T^R_{\mu\nu}f\|_{-1/2} = 0. \tag{8.9}$$

Proof: The parts (1) and (2) are easily proved. By Lemma 3.1, $T^R_{\mu\nu}$ is a bounded operator from $M_0(R^3)$ to $M_{-1/2}(R^3)$. Let f be in $M_0(R^3)$. Then, we have

$$\|T_{\mu\nu}f - T^R_{\mu\nu}f\| \leqslant |e|\{\|\omega^2(Q-Q^R)Gd_{\mu\nu}\hat{\rho}\sqrt{\omega}f\|_0 + \|\omega^2Q^RGd_{\mu\nu}(\hat{\rho}-(2\pi)^{-3/2})\sqrt{\omega}f\|_0\}. \tag{8.10}$$

By the boundedness of G on $M_{-1/2}(R^3)$ and the dominated convergence theorem, the first term of (8.10) converges to zero as $\rho\to\delta$. On the other hand, we have [the second term of (8.10)] $\leqslant$const$\times\|G\|_{-1/2}\|(\hat{\rho}-(2\pi)^{-3/2})f\|_0\to 0$ as $\rho\to\delta$. ■

By Lemma 8.1 and (8.2)–(8.4), we obtain the following

Theorem 8.1: For all n, $m\geqslant 1$,

$$\lim_{\rho\to\delta} W^{(n,m)}_{\mu_1\cdots\mu_n;\nu_1\cdots\nu_m}(f_1,t_1,\ldots,f_n,t_n;s_1,\ldots,s_m) \equiv W^{(n,m)R}_{\mu_1\cdots\mu_n;\nu_1\cdots\nu_m}(f_1,t_1,\ldots,f_nt_n;s_1,\ldots,s_m), \quad f_j\in\mathscr{S}(R^3), t_j,s_j\in R^1, \tag{8.11}$$

exists; if $n+m$ is odd, then $W^{(n,m)R}_{(\ldots)}$ is zero and, if $n+m$ is even, then it is written as a sum of products of the renormalized 2-point distributions, $W^{(2,0)R}_{\mu\nu}$, $W^{(1,1)R}_{\mu;\nu}$ and $W^{(0,2)R}_{\mu\nu}$ which are obtained by replacing $T_{\mu\nu}$ (respectively Q) by $T^R_{\mu\nu}$ (respectively Q^R) in the right-hand side of (8.2), (8.3), and (8.4) respectively.

The sequence of the renormalized Wightman distributions $\{W^{(n,m)R}_{(\ldots)}\}_{n,m=1}^{\infty}$ is positive definite and has time translation invariance. Thus, by the Wightman reconstruction theorem, we can construct a Hilbert space, a Hamiltonian, fields and a unique vacuum in such a way that they give $\{W^{(n,m)R}_{(\ldots)}\}_{n,m=1}^{\infty}$ as the vacuum expectation values for the fields. By noting the unitary equivalence of $\mathscr{H}_{\rm phys}$ and $\mathscr{F}^{\rm EM}$ (cf. Sec. V-B and Sec. VII) or the explicit form of the fields (7.2) and (7.3), we can get a concrete realization for the renormalized theory and develop the scattering theory. Since it is an easy, but a tedious task, the details are omitted.

ACKNOWLEDGMENTS

The author would like to thank Professor H. Ezawa of Gakushuin University for helpful suggestions.

APPENDIX A: A PROPERTY OF HILBERT–SCHMIDT OPERATOR

Proposition A 1: Let S be a Hilbert–Schmidt operator on $L^2(R^n)$, $n\geqslant 1$. Then,

$$\lim_{t\to\pm\infty} \|Se^{iRt}f\|_{L^2(R^n)} = 0, \quad f\in L^2(R^n),$$

where

$$R(x) = |x|, \quad x\in R^n.$$

Proof: We can write as

$$(Se^{iRt}f)(x) = \int dy S(x,y) e^{i|y|t} f(y),$$

where $S(\cdot\ ,\ \cdot)\in L^2(R^{2n})$ is the Hilbert–Schmidt kernel of S. The square integrability of $S(\cdot\ ,\ \cdot)$ and the Riemann–Lebesgue lemma imply that

$$\lim_{t\to\pm\infty} (Se^{iRt}f)(x) = 0, \quad \text{a.e.}x.$$

Furthermore, we have by the Schwarz inequality

$$|(Se^{iRt}f)(x)|^2 \leqslant \|f\|^2_{L^2(R^n)} \int dy |S(x,y)|^2, \quad \text{a.e.}x.$$

The rhs does not depend on t and is integrable with respect to dx. Thus, by the dominated convergence theorem, we get the desired result.

APPENDIX B: SPECTRUM OF THE OPERATOR $\mathbf{p}\cdot\mathbf{q} + \mathbf{q}\cdot\mathbf{p}$

We shall consider the spectrum of a more general operator

$$A_n = \sum_{j=1}^{n} (p_j q_j + q_j p_j), \tag{B1}$$

which acts in $L^2(R^n)$, where q_j is the multiplication operator of j-th coordinate and $p_j = -i\partial/\partial q_j$. By applying the Nelson commutator theorem (see, e.g., Ref. 25, §X.5), we can prove that A_n is essentially self-adjoint on any core for the operator $\Sigma_{j=1}^n (p_j^2 + q_j^2)$.

Proposition A 2. For all $n\geqslant 1$, we have

$$\sigma(A_n) = \sigma_{ac}(A_n) = R^1, \quad \sigma_p(A_n) = \sigma_s(A_n) = \varnothing. \tag{B2}$$

Proof: We need only to prove (B2) for $n = 1$, because A_n $(n\geqslant 2)$ is identified with the operator $A_1 \otimes I \otimes \cdots \otimes I + I \otimes A_1 \otimes \cdots \otimes I + \cdots + I \otimes \cdots \otimes A_1$ acting in the n-fold tensor product of $L^2(R^1)$. We denote A_1 by A. The Hilbert space $L^2(R^1)$ is decomposed as a direct sum:

$$L^2(R^1) = L^2(R_+) \oplus L^2(R_-), \tag{B3}$$

where $R_+ = (0,\infty)$ and $R_- = (-\infty,0)$. Using the essential self-adjointness of A on $C_0^\infty(R^1\setminus\{0\})$, which follows from that of $p^2 + q^2$ (see Ref. 25, §X.4, Kalf–Walter–Schmincke–Simon theorem), A is reduced by $L^2(R_\pm)$. Thus, we need only to consider A in $L^2(R_\pm)$, which are denoted by $A_\pm$ respectively. For f in $L^2(R_+)$, we define a function uf on R^1 by

$$(uf)(\lambda) = \frac{1}{2\sqrt{\pi}} \int_{R_+} dq\, e^{-i/2(\lambda - i)\log q} f(q), \quad \lambda\in R^1. \tag{B4}$$

By change of variable we have

$$(uf)(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{R^1} dx\, e^{-i\lambda x} (vf)(x), \tag{B5}$$

where

$$(vf)(x) = \sqrt{2} e^x f(e^{2x}). \tag{B6}$$

It is easy to see that the map v is a unitary operator from $L^2(R_+)$ onto $L^2(R^1)$. Since uf is the Fourier transform of vf, it follows that u is a unitary operator from $L^2(R_+)$ onto $L^2(R^1)$. Now, let f be in $C_0^\infty(R_+)$. Then, by integration by parts, we have

$$(uA_+ f)(\lambda) = \lambda (uf)(\lambda). \tag{B7}$$

Since $C_0^\infty(R_+)$ is a core for A_+, (B7) extends to all f in $D(A_+)$, so that we have

$$uA_+ u^{-1} = \lambda. \tag{B8}$$

In the same way we can show that A_- is unitarily equivalent to the multiplication operator λ in $L^2(R^1)$. Thus, we obtain (B2) for $n = 1$.

[1]H. A. Bethe, Phys. Rev. **72**, 339 (1947).
[2]T. A. Welton, Phys. Rev. **74**, 1157 (1948).
[3]E. A. Power and S. Zienau, Trans. R. Philos. Soc. London A **251**, 427 (1959).
[4]J. M. Jauch and R. Rohrlich, *The Theory of Photons and Electrons*, 2nd ed. (Springer, Berlin, New York, 1976).
[5]*Foundations of Radiation Theory and Quantum Electrodynamics*, edited by A. O. Barut (Plenum, New York, 1980).
[6]A. Arai, J. Math. Phys. **22**, 2539 (1981).
[7]A. Arai, J. Math. Phys. **22**, 2549 (1981).
[8]R. E. Norton and W. K. R. Watson, Phys. Rev. **116**, 1597 (1959).
[9]K. Rzażewski and W. Żakowicz, J. Phys. A. Math. Gen. **9**, 1159 (1976).
[10]Ph. Blanchard, Commun. Math. Phys. **15**, 156 (1969).
[11]A. Arai, J. Math. Phys. **22**, 534 (1981).
[12]K. O. Friedrichs, Commun. Pure Appl. Math. **1**, 361 (1948).
[13]J. S. Howland, J. Math. Anal. Appl. **23**, 575 (1968).
[14]J. S. Howland, Pac. J. Math. **29**, 565 (1969).
[15]J. S. Howland, Bull. Am. Math. Soc. **78**, 280 (1972).
[16]J. S. Howland, Pac. J. Math. **55**, 157 (1974).
[17]W. Baumgärtel, Math. Nachr. **69**, 107 (1975).
[18]L. P. Horwitz and I. M. Sigal, Helv. Phys. Acta **51**, 685 (1978).
[19]B. Simon, Ann. Math. **97**, 247 (1973).
[20]J. Fröhlich, Fortsch. Phys. **22**, 159 (1974).
[21]J. Fröhlich, Ann. Inst. Henri Poincaré Sec. A **19**, 1 (1973).
[22]A. Arai, "A note on scattering theory in nonrelativistic quantum electrodynamics" (to appear in J. Phys. A. Math. Gen.).
[23]T. Kato, *Perturbation Theory for Linear Operators*, 2nd ed. (Springer, Berlin, New York, 1976).
[24]L. Landau and E. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, (Addison-Wesley, Reading, MA, 1958).
[25]M. Reed and B. Simon, *Methods of Modern Mathematical Physics Vol. II: Fourier Analysis, Self-adjointness* (Academic, New York, 1975).
[26]The equations of motion in Ref. 8 are slightly different from ours, because the gauge condition $\nabla\cdot\mathbf{A} = 0$ is ignored there.
[27]Y. Kato and N. Mugibayashi, Progr. Theor. Phys. **30**, 103 (1963).
[28]R. Hoeh-Krohn, Commun. Math. Phys. **21**, 256 (1971).
[29]S. Albeverio, "An introduction to some mathematical aspects of scattering theory in models of quantum fields," in *Scattering Theory in Mathematical Physics*, edited by J. A. Lavita and J. P. Marchand (Reidel, Dordrecht, Holland, 1974).
[30]F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[31]S. Bochner and W. T. Martin, *Several Complex Variables* (Princeton U.P., Princeton, 1948).

Exact dynamics of a model for a three-level quantum system interacting with a continuous spectrum

Russell Davidson
Department of Economics, Queen's University, Kingston, Ontario, K7L 3N6, Canada
John J. Kozak
Department of Chemistry and Radiation Laboratory, [a] *University of Notre Dame, Notre Dame, Indiana 46556*

(Received 24 August 1981; accepted for publication 14 January 1983)

Previous work of the authors on a three-level quantum system is extended to allow the radiation field in interaction with the system to have a continuous spectrum of possible frequencies. The limiting procedure involved in the passage from discrete to continuous spectra is complicated by the need to express sums over different discrete spectra as integrals with well-behaved limits. Exact expressions are found, for a spontaneous emission problem, for the time evolution of the probabilities that the system be in (each) one of its three states and representative calculations of these probabilities are presented. The emergence of irreversible behavior upon constructing the thermodynamic limit of the problem is plainly seen and this demonstration permits for the first time a discussion of the relative effectiveness of competing decay channels in three-level quantum systems without imposing any ad hoc assumptions (such as exponential decay). Rather, the actual form of the decay emerges as a consequence of the structure and parameters of the Hamiltonian defining the model, and hence one can examine the variety of circumstances in which the evolution can reasonably be described as exponential. The results obtained should be of great use in clarifying certain outstanding conceptual problems in radiation physics, particularly those which deal with the universality of exponential decay in three- (and two-) level quantum systems in interaction with a radiation field and the conditions under which nonexponential or even nonergodic behavior can emerge in such dissipative quantum systems.

PACS numbers: 32.80. – t, 03.40.Kf

I. INTRODUCTION

In this paper we continue work begun in an earlier one (Ref. 1). There we set up a model of a three-level quantum system interacting with a one-dimensional radiation field, and found exact expressions for the time evolution of the system in the sector describing the spontaneous emission of radiation by the system from its highest energy level. These expressions were obtained for a finite system; that is, one where the modes of the radiation field were discrete. Our task in this paper will be to extend the results to infinite systems, where there is a continuous spectrum of possible radiation frequencies. To this end, we start with a finite system for which the modes of the field are those characteristic of a one-dimensional box of length L, with periodic boundary conditions. Then L will tend to infinity. This seemingly simple procedure is in fact very complicated. In the eighth paper cited in Ref. 2 can be found the analogous calculations for a two-level atom, and while we will be able here to take advantage of many of the results of that paper, a good deal of additional work is necessary. The difficulty arises because the discrete-spectrum results are expressed as nested sums over three different discrete spectra, one simply that of the unperturbed radiation field, the second that of a two-level atom, and the third the spectrum for the three-level system itself. Before the $L \to \infty$ limit can be taken, these sums must be expressed as integrals for which limits exist in reasonably tractable form.

The (discrete-mode) Hamiltonian governing the system is

$$\begin{aligned} H = {} & \hbar\epsilon_3|3\rangle\langle 3| + \sum_\lambda \hbar(\epsilon_2 + \omega_\lambda)|2;\lambda\rangle\langle 2;\lambda| \\ & + \sum_{\lambda_1<\lambda_2}\sum \hbar(\omega_{\lambda_1} + \omega_{\lambda_2})|1;\lambda_1,\lambda_2\rangle\langle 1;\lambda_1,\lambda_2| \\ & + \sum_\lambda 2\hbar\omega_\lambda|1;2\lambda\rangle\langle 1;2\lambda| + \sum_\lambda \hbar\omega_\lambda|1;\lambda\rangle\langle 1;\lambda| \\ & + \sqrt{2}\Big\{\sum_\lambda \bar{h}_\lambda|2;\lambda\rangle\langle 3| + \sum_{\lambda_1>\lambda_2}\sum \bar{g}_{\lambda_1}|1;\lambda_2,\lambda_1\rangle\langle 2;\lambda_2| \\ & + \sum_{\lambda_1<\lambda_2}\sum \bar{g}_{\lambda_1}|1;\lambda_1,\lambda_2\rangle\langle 2;\lambda_2| + \sum_\lambda \bar{g}_\lambda\sqrt{2}|1;2\lambda\rangle\langle 2;\lambda| \\ & + \sum_\lambda \bar{f}_\lambda|1;\lambda\rangle\langle 3| + \text{Hermitian conjugate}\Big\}. \end{aligned} \qquad (1)$$

[see Eq. (2) of Ref. 1] The notation is as follows: $|3\rangle$ is the state with the quantum system in its highest energy level, with energy separation $\hbar\epsilon_3$ from the ground state. $|2;\lambda\rangle$ is a state with the system in its intermediate level, separated by $\hbar\epsilon_2$ from the ground state, and one photon in mode λ of the field. $|1;\lambda_1,\lambda_2\rangle$ has the system in its ground state, and one photon in each of modes λ_1 and λ_2, where $\lambda_1 < \lambda_2$ according

[a] The research described herein was supported by the Office of Basic Energy Sciences of the Department of Energy. This is Document No. NDRL-2275 from the Notre Dame Radiation Laboratory.

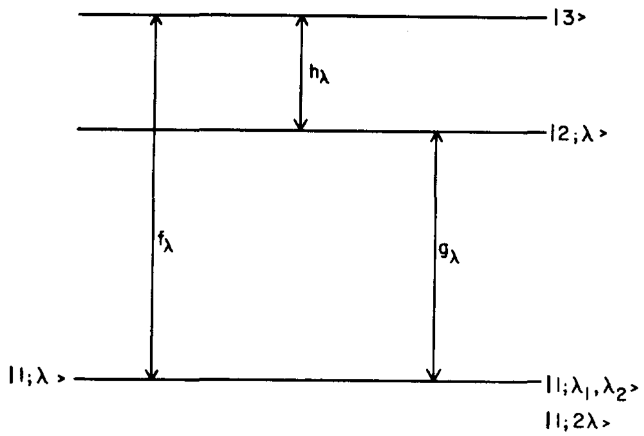


FIG. 1. The model for a three-level atom considered in this paper.

to some suitable ordering of the modes. $|1;2\lambda\rangle$ and $|1;\lambda\rangle$ are defined in the obvious way. The photon energies are $\hbar\omega_\lambda$, where the frequency ω_λ will be taken as $2\pi nc/L$, $n = 1, 2, 3, \ldots$, c = speed of light. The quantities $h_\lambda, g_\lambda, f_\lambda$ are transition matrix elements which measure the strength of the couplings of the system with the field. The bars in Eq. (1) and subsequently denote complex conjugates. The model is intended to describe the system shown in Fig. 1. The spontaneous emission problem starts with the system in state $|3\rangle$.

It is clear that if our quantum system is intended to model electronic excitations in an atom, then, at least for electric dipole radiation, one of the three transition matrix elements for each mode λ must vanish by the parity selection rule. However, this need not be the case for radiation of higher multipolarity. There exist other quantum systems, too, for which the three states will not have a well-defined parity and for which consequently the selection rule will not apply. In condensed phase spectroscopy, for example, in addition to finding matrix-induced shifts and splittings in the electronic spectra of matrix isolated solutes, one also finds that the matrix can allow a partial or complete relaxation of the usual free-molecular selection rules,[3] and it would seem useful to have rigorous results at hand to discuss aspects of this situation in detail.

Apart from this consideration, the calculations of this paper are further steps in the process of replacing Wigner–Weisskopf style approximations to the dynamics of the decay of excited quantum states by exact quantum-mechanical results. For many systems the improvement in accuracy achieved by this replacement will not be large (see Ref. 4 for bounds on the discrepancy between the exact and Wigner–Weisskopf exponential solutions) and may not even be experimentally detectable. However, the calculations presented here may have considerable theoretical interest. A good deal of attention has been given recently to various "semi-classical" models of the interaction between matter and radiation. See, for example, the works cited in Ref. 5. In judging the worth of these models it is most useful to have a collection of exact results against which the approximate results of semi-classical models can be compared. We intend in a future paper to attempt such comparisons.

In Sec. II, an infinite-system (continuous-spectrum) limit is found for $\rho_3(t)$, the probability that the system is in state $|3\rangle$ at time t. In Sec. III, the limit is found for $\rho_2(t)$, the probability of the system being in its intermediate state. The ground state probability is, of course, just $1-\rho_2(t)-\rho_3(t)$. A numerical investigation of these functions is presented in Sec. IV for a particular set of parameter values, and Sec. V contains concluding remarks.

II. CALCULATION OF $\rho_3(\tau)$

Our starting point is Eqs. (22) and (23) of the earlier paper, and our task is to replace all the discrete summations in these equations by integrals of which the infinite-system limit can readily be taken. Let us rewrite the expression for $\phi_3(t)$, the probability amplitude whose squared modulus is $\rho_3(t)$, the probability that, at time t, the system is in its state $|3\rangle$. We have

$$\phi_3(t) = \frac{1}{2\pi i}\int_C dz\, e^{-izt}\Bigg[2\pi i F(z) + (2\pi i)^2 \frac{1}{H_1(z)}\frac{2}{\hbar^2}\sum_\mu \frac{\overline{G}(\xi_\mu)}{H'(\xi_\mu)H(z-\xi_\mu)} \times\sum_\kappa \frac{G(\xi_\mu)-G(\xi_\kappa)+G(z-\xi_\mu)-G(z-\xi_\kappa)}{H'(\xi_\kappa)H(z-\xi_\kappa)}\Bigg]^{-1}, \tag{2}$$

where C is a Bromwich contour above and parallel to the real axis;

$$F(z) = \frac{1}{2\pi i}\left(\epsilon_3 - z - \sum_\mu \frac{2|f_\mu|^2}{\hbar^2(\omega_\mu - z)}\right); \tag{3}$$

$$H(z) = \frac{1}{2\pi i}\left(\epsilon_2 - z - \sum_\mu \frac{2|g_\mu|^2}{\hbar^2(\omega_\mu - z)}\right); \tag{4}$$

the zeros of H are denoted by the sequence ξ_κ;

$$H_1(z) = \sum_\kappa \frac{1}{H'(\xi_\kappa)H(z-\xi_\kappa)}; \tag{5}$$

the zeros of H_1 are denoted ζ_ν;

$$G(z) = -\frac{1}{(2\pi i)^2}\frac{\sqrt{2}}{\hbar}\sum_\lambda \frac{g_\lambda \overline{h}_\lambda}{\omega_\lambda - z}; \tag{6}$$

and $\overline{G}(\xi_\mu)$ denotes simply the complex conjugate of $G(\xi_\mu)$.

It is convenient to proceed directly to the dimensionless variables used for the purposes of numerical calculation in the earlier paper. Accordingly, we make the following set of definitions using a dimensionless coupling constant α:

$$\tau = \alpha\epsilon_3 t, \tag{7a}$$

$$\xi = z/\alpha\epsilon_3,\quad \beta_\lambda = \omega_\lambda/\alpha\epsilon_3,\quad \gamma_\mu = \xi_\mu/\alpha\epsilon_2,$$
$$\delta_\nu = \zeta_\nu/\alpha\epsilon_2, \tag{7b}$$

$$e = \epsilon_3/\epsilon_3,\quad \sigma^2 = \alpha\epsilon_3 L/c. \tag{7c}$$

Dimensionless coupling functions are introduced as follows:

$$g_\lambda = \hbar(\alpha\epsilon_2 c/L)^{1/2} g^{1/2}(\alpha e\beta_\lambda), \tag{8a}$$

$$f_\lambda = s\hbar(\alpha\epsilon_3 c/L)^{1/2} f^{1/2}(\alpha\beta_\lambda), \tag{8b}$$

$$h_\lambda = r\hbar(\alpha\epsilon_2 c/L)^{1/2} h^{1/2}(\alpha e\beta_\lambda), \tag{8c}$$

where the notations $g^{1/2}, f^{1/2}, h^{1/2}$ denote complex functions whose squared moduli will be written simply as g, f, h. The scaling parameters s and r are introduced in order to permit the normalization $g(1) = f(1) = h(1) = 1$. Functions corresponding to F, H, H_1, G are defined:

$$\hat{F}(\xi) = (2\pi i/\alpha\epsilon_3)F(\alpha\epsilon_3\xi), \tag{9a}$$

$$\hat{H}(\xi) = (2\pi i/\alpha\epsilon_2)H(\alpha\epsilon_2\xi), \tag{9b}$$

$$\hat{H}_1(\xi) = (\alpha\epsilon_2/(2\pi i)^2)H_1(\alpha\epsilon_2\xi), \tag{9c}$$

$$\hat{G}(\xi) = [(2\pi i)^2\sqrt{2}/h\alpha\epsilon_2]G(\alpha\epsilon_2\xi). \tag{9d}$$

From Eq. (2), then, we obtain

$$\begin{aligned}\hat{\phi}_3(\tau) &\equiv \phi_3(t)\\ &= \frac{1}{2\pi i}\int_C d\xi\, e^{-i\xi\tau}\\ &\quad\times\Bigg[\hat{F}(\xi) + \frac{1}{e\hat{H}_1(e\xi)}\sum_\mu \frac{\overline{G}(\hat{\gamma}_\mu)}{\hat{H}'(\gamma_\mu)\hat{H}(e\xi-\gamma_\mu)}\sum_\kappa\\ &\quad\times\frac{\hat{G}(\gamma_\mu)-\hat{G}(\gamma_\kappa)+\hat{G}(e\xi-\gamma_\mu)-\hat{G}(e\xi-\gamma_\kappa)}{\hat{H}'(\gamma_\kappa)\hat{H}(e\xi-\gamma_\kappa)}\Bigg]^{-1}.\end{aligned} \tag{10}$$

Let us collect here a few results on Laplace transforms and convolutions that will be of great use in our subsequent analysis. If two functions a, b, defined on the positive real axis have Laplace transforms A, B defined as follows:

$$\begin{aligned}A(\xi) &= i\int_0^\infty d\tau\,\exp(i\xi\tau)a(\tau),\\ B(\xi) &= i\int_0^\infty d\tau\,\exp(i\xi\tau)b(\tau),\end{aligned} \tag{11}$$

then the Laplace transform of the function ab is the convolution, $A*B$, of A and B, defined by the formula

$$(A*B)(\xi) \equiv \frac{1}{2\pi i}\int_D d\zeta\, A(\zeta)B(\xi-\zeta). \tag{12}$$

Either A or B, regarded as a function of the complex variable ξ, is defined and holomorphic in the region $\operatorname{Im}\xi > k$, for some k. The integrand in Eq. (12) will possess singularities with $\operatorname{Im}\zeta < k$ associated with the function A and singularities with $\operatorname{Im}\zeta > -k$ associated with B. In the representation (12), the contour D is defined as one which passes from $-\infty$ to $+\infty$, leaving the singularities associated with A on the left and those associated with B on the right, provided this is feasible. The inversion formula for Eq. (11) is then

$$a(\tau) = \frac{1}{2\pi i}\int_C d\xi\,\exp(-i\xi\tau)A(\xi). \tag{13}$$

Similarly, if one considers the inverse transform of the product AB, one finds that

$$\frac{1}{2\pi i}\int_C d\xi\, e^{-i\xi\tau}A(\xi)B(\xi) = i\int_0^\tau d\tau'\, a(\tau-\tau')b(\tau'). \tag{14}$$

We can now obtain an integral expression for the sum

$$\sum_\kappa \frac{\hat{G}(\gamma_\kappa)}{\hat{H}'(\gamma_\kappa)\hat{H}(e\xi-\gamma_\kappa)} \tag{15}$$

such that the $L\to\infty$ limit is easily taken. For this we define functions f_1 and f_2 as inverse Laplace transforms as follows:

$$f_1(\tau) = -\sum_\kappa \frac{e^{-i\gamma_\kappa\tau}}{\hat{H}'(\gamma_\kappa)} = \frac{1}{2\pi i}\int_C d\xi\,\frac{e^{-i\xi\tau}}{\hat{H}(\xi)}, \tag{16}$$

$$f_2(\tau) = \sum_\kappa \frac{e^{-i\gamma_\kappa\tau}\hat{G}(\gamma_\kappa)}{\hat{H}'(\gamma_\kappa)} = -\frac{1}{2\pi i}\int_C d\xi\,\frac{e^{-i\xi\tau}\hat{G}(\xi)}{\hat{H}(\xi)}. \tag{17}$$

Then the sum in Eq. (15) can be expressed as a contour integral for $\operatorname{Im}\xi > 0$.

$$-\frac{1}{2\pi i}\int_C d\zeta\,\frac{\hat{G}(\zeta)}{\hat{H}(\zeta)\hat{H}(e\xi-\zeta)},$$

and this is just the convolution

$$(-\hat{G}/\hat{H})*1/\hat{H}$$

evaluated at $e\xi$. It is easy to see that this convolution is also the value of the sum

$$\sum_\kappa \frac{\hat{G}(e\xi-\gamma_\kappa)}{\hat{H}'(\gamma_\kappa)\hat{H}(e\xi-\gamma_\kappa)}.$$

From (12) and (13) the convolution is just

$$i\int_0^\infty d\tau\, e^{ie\xi\tau}f_1(\tau)f_2(\tau), \tag{18}$$

and the $L\to\infty$ limit of this will be obtained by making use of the $L\to\infty$ limits of the functions f_1 and f_2. It has been seen in Ref. 2viii that these limits are well defined, and given by the following expressions:

$$f_1(\tau) = \frac{1}{2\pi i}\int_0^\infty d\zeta\, e^{-i\zeta\tau}\left(\frac{1}{\hat{H}^+(\zeta)} - \frac{1}{\hat{H}^-(\zeta)}\right) \tag{19}$$

and

$$f_2(\tau) = -\frac{1}{2\pi i}\int_0^\infty d\zeta\, e^{-i\zeta\tau}\left(\frac{\hat{G}^+(\zeta)}{\hat{H}^+(\zeta)} - \frac{\hat{G}^-(\zeta)}{\hat{H}^-(\zeta)}\right). \tag{20}$$

Here, the functions $\hat{H}^\pm$ and $\hat{G}^\pm$ are the limiting values of the functions $\hat{H}$ and $\hat{G}$, which become sectionally holomorphic[6] in the $L\to\infty$ limit:

$$\hat{H}^\pm(\zeta) = \frac{1}{\alpha} - \zeta + \frac{2}{\pi}\mathscr{P}\int_0^\infty d\lambda\,\frac{g(\alpha\lambda)}{\zeta-\lambda} \mp 2ig(\alpha\zeta), \tag{21}$$

$$\begin{aligned}\hat{G}^\pm(\zeta) &= \frac{2r}{\pi}\mathscr{P}\int_0^\infty d\lambda\,\frac{g^{1/2}(\alpha\lambda)\bar{h}^{1/2}(\alpha\lambda)}{\zeta-\lambda}\\ &\quad\mp 2irg^{1/2}(\alpha\zeta)\bar{h}^{1/2}(\alpha\zeta),\end{aligned} \tag{22}$$

($\mathscr{P}$ denotes Cauchy principal part).

In an exactly similar way, we obtain the result

$$\sum_\mu \frac{\overline{\hat{G}}(\gamma_\mu)}{\hat{H}'(\gamma_\mu)\hat{H}(e\xi-\gamma_\mu)} = i\int_0^\infty d\tau\, e^{ie\xi\tau}f_1(\tau)f_2^*(\tau), \tag{23}$$

where we define $f_2^*(\tau) = \overline{f_2(-\tau)}$.

We still need an integral expression for the summation:

$$\sum_\mu \frac{\overline{\hat{G}}(\gamma_\mu)[\hat{G}(\gamma_\mu)+\hat{G}(e\xi-\gamma_\mu)]}{\hat{H}'(\gamma_\mu)\hat{H}(e\xi-\gamma_\mu)}.$$

One part is easy:

$$\begin{aligned}\sum_\mu \frac{\overline{\hat{G}}(\gamma_\mu)\hat{G}(e\xi-\gamma_\mu)}{\hat{H}'(\gamma_\mu)\hat{H}(e\xi-\gamma_\mu)} &= -\left(\frac{\tilde{\hat{G}}}{\hat{H}}*\frac{\hat{G}}{\hat{H}}\right)(e\xi)\\ &= -i\int_0^\infty d\tau\, e^{ie\xi\tau}f_2(\tau)f_2^*(\tau),\end{aligned} \tag{24}$$

where

$$\tilde{G}(\xi) \equiv \overline{\hat{G}(\bar{\xi})}.$$

For the rest, define a function f_3:

$$f_3(\tau) = -\sum_\kappa \frac{|\hat{G}(\gamma_\kappa)|^2 e^{-i\gamma_\kappa\tau}}{\hat{H}'(\gamma_\kappa)}. \tag{25}$$

Then it is immediate that

$$\sum_\mu \frac{\hat{G}(\gamma_\mu)\tilde{G}(\gamma_\mu)}{\hat{H}'(\gamma_\mu)\hat{H}(e\xi-\gamma_\mu)} = -i\int_0^\infty d\tau\, e^{ie\xi\tau} f_1(\tau) f_3(\tau). \tag{26}$$

We still need an expression that permits us to take the infinite-system limit of f_3. By use of the residue theorem we obtain

$$\frac{1}{2\pi i}\int_C d\xi\, \frac{\hat{G}(\xi)\tilde{G}(\xi)e^{-i\xi t}}{H(\xi)}$$

$$= -\sum_\kappa \frac{|\hat{G}(\gamma_\kappa)|^2}{\hat{H}'(\gamma_\kappa)} e^{-i\gamma_\kappa\tau} - \frac{2r^2 e}{\sigma^2}\sum_\lambda h(\alpha e\beta_\lambda) e^{-ie\beta_\lambda\tau},$$

so that in the limit

$$f_3(\tau) = \frac{1}{2\pi i}\int_0^\infty d\zeta \left[\frac{\hat{G}^+(\zeta)\tilde{G}^+(\zeta)}{\hat{H}^+(\zeta)} - \frac{\hat{G}^-(\zeta)\tilde{G}^-(\zeta)}{\hat{H}^-(\zeta)}\right] e^{-i\zeta\tau}$$
$$+ \frac{2r^2}{\pi}\int_0^\infty d\lambda\, h(\alpha\lambda) e^{-i\lambda\tau}. \tag{27}$$

We may now put together the results (18), (23), (24), and (26) into Eq. (10) to obtain

$$\hat{\phi}_3(\tau) = \frac{1}{2\pi i}\int_C d\xi\, e^{-i\xi\tau}\Big[\hat{F}(\xi)$$
$$-\frac{i}{e}\int_0^\infty d\tau\, e^{ie\xi\tau}\big(f_2(\tau)f_2^*(\tau) + f_1(\tau)f_3(\tau)\big)$$
$$+\frac{2}{e\hat{H}(e\xi)}\int_0^\infty d\tau\, e^{ie\xi\tau} f_1(\tau) f_2^*(\tau)$$
$$\times \int_0^\infty d\tau'\, e^{ie\xi\tau'} f_1(\tau') f_2(\tau')\Big]^{-1}. \tag{28}$$

The integrand considered as a function of ξ is holomorphic for $\mathrm{Im}\,\xi > 0$, and so in particular along C, but has a branch cut along the real axis of ξ. We can deform C to a contour B (see Fig. 2) by Jordan's lemma, since the integrand in Eq. (28), apart from the factor $e^{-i\xi\tau}$, tends to zero as $|\xi| \to \infty$. Then the entire expression in Eq. (18) can be evaluated in terms of the limiting values of the integrand as ξ tends to positive real values either from above or below the real axis.

In order to obtain these limiting values, we are concerned with the analytic continuation of expressions like

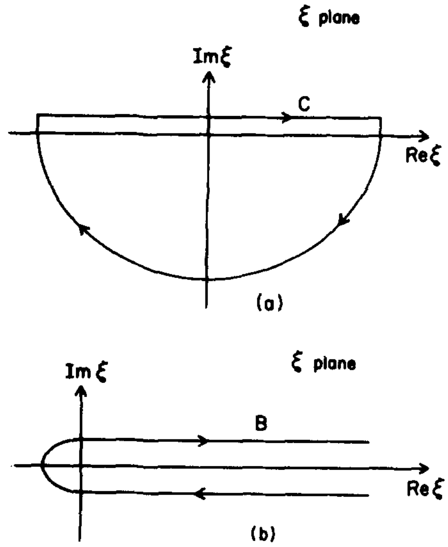


FIG. 2. (a) The Bromwich contour C chosen in the evaluation of the integral (28); (b) The deformed contour B.

$$i\int_0^\infty d\tau\, e^{i\xi\tau} p(\tau), \tag{29a}$$

defined for $\mathrm{Im}\,\xi > 0$. This expression defines a function of ξ which is analytic for $\mathrm{Im}\,\xi > 0$ and tends to a well-defined and finite limit as $\mathrm{Im}\,\xi \to 0^+$. Now the functions represented by $p(\tau)$ are all analytic for $\mathrm{Im}\,\tau < 0$, and in fact they all tend to zero as $|\tau| \to \infty$ in the lower half-plane. To see this, consider, for example, the definitions (16) and (17) of $f_1(\tau)$ and $f_2(\tau)$. Thus the function in (29a) can be continued through the negative real axis of ξ by the following formula:

$$i\int_0^\infty d\tau\, e^{i\xi\tau} p(\tau)$$
$$= i\int_0^{-\infty} d\tau\, e^{i\xi\tau} p(\tau) \quad \text{(Jordan's lemma)}$$
$$= -i\int_0^\infty d\tau\, e^{-i\xi\tau} p(-\tau). \tag{29b}$$

The last expression here is analytic for $\mathrm{Im}\,\xi < 0$ and thus when evaluated for real and positive ξ gives us the limiting value from below needed in order to evaluate Eq. (28).

Finally, then,

$$\hat{\phi}_3(\tau) = \frac{1}{2\pi i}\int_0^\infty d\xi\, e^{-i\xi\tau}\Bigg\{\Big[\hat{F}^+(\xi) - \frac{i}{e}\int_0^\infty d\theta\, e^{ie\xi\theta}\big(f_2(\theta)f_2^*(\theta) + f_1(\theta)f_3(\theta)\big)$$
$$+\frac{2}{e\hat{H}_1^+(e\xi)}\int_0^\infty d\theta\, e^{ie\xi\theta} f_1(\theta) f_2^*(\theta)\int_0^\infty d\theta'\, e^{ie\xi\theta'} f_1(\theta') f_2(\theta')\Big]^{-1}$$
$$-\Big[\hat{F}^-(\xi) + \frac{i}{e}\int_0^\infty d\theta\, e^{-ie\xi\theta}\big(f_2(-\theta)f_2^*(-\theta) + f_1(-\theta)f_3(-\theta)\big)$$
$$+\frac{2}{e\hat{H}_1^-(e\xi)}\int_0^\infty d\theta\, e^{-ie\xi\theta} f_1(-\theta) f_2^*(-\theta)\int_0^\infty d\theta'\, e^{-ie\xi\theta'} f_1(-\theta') f_2(-\theta')\Big]^{-1}\Bigg\}. \tag{30}$$

It is easy to see that

$$\hat{F}^{\pm}(\xi) = \frac{1}{\alpha} - \xi + \frac{2s^2}{\pi}\mathscr{P}\int_0^\infty d\lambda \frac{f(\alpha\lambda)}{\xi - \lambda} \mp 2is^2 f(\alpha\xi) \tag{31}$$

and

$$\hat{H}_1{}^{\pm}(\xi) = \mp i\int_0^\infty d\tau\, e^{\pm i\xi\tau}[f_1(\pm\tau)]^2, \tag{32}$$

since from the definition of H_1 it is clear that

$$\hat{H}_1 = -(1/\hat{H})*(1/\hat{H}). \tag{33}$$

Then the probability, $\rho_3(\tau)$, is given by

$$\rho_3(\tau) = |\hat{\phi}_3(\tau)|^2. \tag{34}$$

The qualitative features of $\hat{\phi}_3(\tau)$ are probably most easily inferred from Eq. (28), in which $\hat{\phi}_3(\tau)$ is expressed as an inverse Laplace transform. This expression may be compared with Eq. (16), for $f_1(\tau)$, which is the function corresponding to $\hat{\phi}_3(\tau)$ for the spontaneous emission of a two-level system. [See paper (iii) of Ref. 2, Eq. (12), and paper (viii), Eq. (15).] Both functions equal unity for $\tau = 0$ and tend to zero as $\tau \to \infty$. It is also instructive to consider the limiting forms of Eq. (28) if one or more of the coupling functions g, f, or h is zero. If $f = 0$, we still have a two-stage decay process, and the only simplification is that $\hat{F}(\xi) = 1/\alpha - \xi$. If $h = 0$, on the other hand, then $f_2 = f_3 = 0$, and so $\hat{\phi}_3(\tau) = (1/2\pi i)\int_C d\xi\, e^{-i\xi\tau}[\hat{F}(\xi)]^{-1}$, an expression exactly analogous to Eq. (16). It is clear that in this case the existence of state $|2\rangle$ does not influence the decay, because, as is clear from Fig. 1, the states $|2;\lambda\rangle$ are not coupled to either $|3\rangle$ or the final states $|1;\lambda\rangle$. If $g = 0$, the state $|3\rangle$ can decay either to $|2;\lambda\rangle$ or $|1;\lambda\rangle$. The resulting dynamics for $\hat{\phi}_3(\tau)$ are just like those of $f_1(\tau)$ but with coupling to two continua. This is clear when one notes that, for $g = 0$,

$$f_2 = 0,\quad f_1(\tau) = e^{-i\tau/\alpha},\quad f_3(\tau) = \frac{2r^2}{\pi}\int_0^\infty d\lambda\, e^{-i\lambda\tau}h(\alpha\lambda).$$

The integrand in Eq. (28) becomes accordingly

$$e^{-i\xi\tau}\left[\frac{1}{\alpha} - \zeta - \frac{2s^2}{\pi}\int_0^\infty \frac{d\lambda\, f(\alpha\lambda)}{\lambda - \zeta} - \frac{2r^2}{e\pi}\int_0^\infty \frac{d\lambda\, h(\alpha\lambda)}{(\lambda + 1/\alpha - e\zeta)}\right]^{-1}. \tag{35}$$

III. CALCULATION OF $\rho_2(\tau)$

For this calculation the starting point in Eqs. (18) and (24) of the earlier paper. These combine to yield the following expression for $\phi_{2,\lambda}(t)$, the probability amplitude for the state $|2,\lambda\rangle$ at the time t:

$$\phi_{2,\lambda}(t) = \frac{1}{2\pi i}\int_C dz\, e^{-izt}\frac{\tilde{\phi}_3(z)}{H_1(z)}\frac{2g_\lambda^*}{\hbar^2} \times\sum_\mu \frac{1}{(\xi_\mu - \omega_\lambda)H'(\xi_\mu)H(z-\xi_\mu)} \times\sum_\kappa \frac{G(\xi_\mu) - G(\xi_\kappa) + G(z-\xi_\mu) - G(z-\xi_\kappa)}{H'(\xi_\kappa)H(z-\xi_\kappa)}. \tag{36}$$

In this expression, $\tilde{\phi}_3(z)$ denotes the Laplace transform of $\phi_3(t)$ as given by Eq. (2):

$$\phi_3(t) = \frac{1}{2\pi i}\int_C dz\, e^{-izt}\tilde{\phi}_3(z). \tag{37}$$

As in the last section, we can move to dimensionless variables in Eq. (36). We obtain

$$\hat{\phi}_{2,\lambda}(\tau) \equiv \phi_{2,\lambda}(t) = \frac{\sqrt{2e}}{\sigma}\frac{1}{2\pi i}\bar{g}^{1/2}(\alpha e\beta_\lambda)\int_C d\xi\, e^{-i\xi\tau}\frac{\psi_3(\xi)}{\hat{H}_1(e\xi)} \times\sum_\mu \frac{1}{(\gamma_\mu - e\beta_\lambda)\hat{H}'(\gamma_\mu)\hat{H}(e\xi - \gamma_\mu)} \times\sum_\kappa \frac{\hat{G}(\gamma_\mu) - \hat{G}(\gamma_\kappa) + \hat{G}(e\xi - \gamma_\mu) - \hat{G}(e\xi - \gamma_\kappa)}{\hat{H}'(\gamma_\kappa)\hat{H}(e\xi - \gamma_\kappa)}. \tag{38}$$

Here the following definition has been used beyond those of Sec. II:

$$\psi_3(\xi) = \alpha\epsilon_3\tilde{\phi}_3(\alpha\epsilon_3\xi). \tag{39}$$

We wish now to make use of the structure of Eq. (38) as the inverse Laplace transform of a product of two functions of ξ. Accordingly, we define the function ψ_λ implicitly by the equation

$$\hat{\phi}_{2,\lambda}(\tau) = \frac{1}{2\pi i}\int_C d\xi\, e^{-i\xi\tau}\psi_3(\xi)\psi_\lambda(\xi), \tag{40}$$

so that if $\hat{\phi}_\lambda(\tau)$ denotes the inverse transform of $\psi_\lambda(\xi)$,

$$\hat{\phi}_\lambda(\tau) = \frac{1}{2\pi i}\int_C d\xi\, e^{-i\xi\tau}\psi_\lambda(\xi), \tag{41}$$

then from Eqs. (40) and (14) we obtain

$$\hat{\phi}_{2,\lambda}(\tau) = i\int_0^\tau d\tau'\,\hat{\phi}_3(\tau - \tau')\hat{\phi}_\lambda(\tau'), \tag{42}$$

since $\psi_3(\xi)$ is the Laplace transform of $\hat{\phi}_3(\tau)$. We shall be interested finally in $\rho_2(\tau)$, the probability that at time $t = \tau/\alpha\epsilon_3$ the atom is in the state $|2\rangle$. From Eq. (42), then,

$$\rho_2(\tau) = \sum_\lambda |\hat{\phi}_{2,\lambda}(\tau)|^2 = \int_0^\tau d\tau'\int_0^\tau d\tau''\,\hat{\phi}_3(\tau - \tau')\overline{\hat{\phi}}_3(\tau - \tau'') \times\sum_\lambda \hat{\phi}_\lambda(\tau')\overline{\hat{\phi}}_\lambda(\tau''). \tag{43}$$

The physical interpretation of the probability amplitude $\hat{\phi}_{2,\lambda}(\tau)$ is as follows. It is the convolution of $\hat{\phi}_3(\tau)$—the amplitude of the state $|3\rangle$ which can decay into $|2;\lambda\rangle$—and $\hat{\phi}_\lambda(\tau)$, which is exactly what one obtains as the amplitude of state $|2;\lambda\rangle$ in the decay of a two-level system (states $|2\rangle$ and $|1\rangle$) in the presence of a photon. [See paper (vii) of Ref. 2, Eq. (24).] This "photon" is here characterized by the amplitude $\bar{h}_\lambda\sqrt{2}/\hbar$, the transition matrix elements from $|3\rangle$ to $|2;\lambda\rangle$.

The next task, then, is to obtain the infinite-system limit of the expression

$$\sum_\lambda \hat{\phi}_\lambda(\tau')\overline{\hat{\phi}}_\lambda(\tau''),$$

since we already have the limit of the function $\hat{\phi}_3$. From Eq. (38) and the definition of $\hat{\phi}_\lambda$, we obtain

$$\hat{\phi}_\lambda(\tau)=\frac{\sqrt{2e}}{\sigma}\frac{1}{2\pi i}\,\overline{g^{1/2}}(\alpha e\beta_\lambda)$$
$$\times\int_C d\xi\, e^{-i\xi\tau}\frac{1}{\hat{H}_1(e\xi)}$$
$$\times\sum_\mu\frac{1}{(\gamma_\mu-e\beta_\lambda)\hat{H}'(\gamma_\mu)\hat{H}(e\xi-\gamma_\mu)}$$
$$\times\sum_\kappa\frac{\hat{G}(\gamma_\mu)-\hat{G}(\gamma_\kappa)+\hat{G}(e\xi-\gamma_\mu)-\hat{G}(e\xi-\gamma_\kappa)}{\hat{H}'(\gamma_\kappa)\hat{H}(e\xi-\gamma_\kappa)}. \tag{44}$$

Just as in the preceding section, we can conclude that

$$\frac{1}{\hat{H}_1(e\xi)}\sum_\kappa\frac{\hat{G}(\gamma_\mu)-\hat{G}(\gamma_\kappa)+\hat{G}(e\xi-\gamma_\mu)-\hat{G}(e\xi-\gamma_\kappa)}{\hat{H}'(\gamma_\kappa)\hat{H}(e\xi-\gamma_\kappa)}$$
$$=\hat{G}(\gamma_\mu)+\hat{G}(e\xi-\gamma_\mu)-\frac{2i}{\hat{H}_1(e\xi)}\int_0^\infty d\theta$$
$$\times e^{ie\xi\theta}f_1(\theta)f_2(\theta). \tag{45}$$

Let us denote the last term in this expression by $A(e\xi)$, and let $a(\tau)$ be the inverse Laplace transform of $A(\xi)$:

$$a(\tau)=\frac{1}{2\pi i}\int_C d\xi\, e^{-i\xi\tau}A(\xi). \tag{46}$$

In order to deal with the sum over μ in Eq. (44), consider the contour integral

$$-\frac{1}{2\pi i}\int_{C_2} d\zeta\,\frac{\hat{G}(\zeta)+\hat{G}(e\xi-\zeta)+A(e\xi)}{(\zeta-e\beta_\lambda)\hat{H}(\zeta)\hat{H}(e\xi-\zeta)}. \tag{47}$$

The contour C_2 is above and parallel to the real axis of ζ, and lies below the contour C. Thus, $\operatorname{Im}\zeta<\operatorname{Im}\xi$. The contour C_2 may be closed in the lower half-plane of ζ, and, within the closed contour so formed, the integrand of (47) has poles at the points $\zeta=e\beta_\lambda$, $\zeta=\gamma_\mu$ for all μ. Thus the integral (47) equals

$$\sum_\mu\frac{\hat{G}(\gamma_\mu)+\hat{G}(e\xi-\gamma_\mu)+A(e\xi)}{(\gamma_\mu-e\beta_\lambda)\hat{H}'(\gamma_\mu)\hat{H}(e\xi-\gamma_\mu)}$$
$$+\frac{r\,\overline{h^{1/2}}(\alpha e\beta_\lambda)}{\overline{g^{1/2}}(\alpha e\beta_\lambda)\hat{H}(e\xi-e\beta_\lambda)}. \tag{48}$$

Now from the definitions of f_1 and f_2, Eqs. (16) and (17), we see that

$$\frac{1}{2\pi i}\int_C d\xi\,\frac{e^{-i\xi t}}{\hat{H}(e\xi-\zeta)}=\frac{1}{e}e^{-i\zeta\tau/e}f_1\!\left(\frac{\tau}{e}\right), \tag{49}$$

$$\frac{1}{2\pi i}\int_C d\xi\,{}^{-i\xi t}\frac{\hat{G}(e\xi-\zeta)}{\hat{H}(e\xi-\zeta)}=-\frac{1}{e}e^{-i\zeta\tau/e}f_2\!\left(\frac{\tau}{e}\right). \tag{50}$$

If we make the definition

$$A(\tau;\zeta)\equiv\frac{1}{2\pi i}\int_C d\xi\, e^{-i\xi\tau}\frac{A(e\xi)}{\hat{H}(e\xi-\zeta)}, \tag{51}$$

then use of Eqs. (45), (48), (49), (50), and (51) in (44) allows us to write

$$\hat{\phi}_\lambda(\tau)=-\frac{\sqrt{2e}}{\sigma}\frac{1}{2\pi i}\,\overline{g^{1/2}}(\alpha\epsilon\beta_\lambda)\int_{C_2}d\zeta\,\frac{1}{(\zeta-e\beta_\lambda)}$$
$$\times\left\{\frac{1}{\hat{H}(\zeta)}\left[\frac{1}{e}e^{-i\zeta\tau/e}\left(\hat{G}(\zeta)f_1\right.\right.\right.$$
$$\left.\left.\times\left(\frac{\tau}{e}\right)-f_2\!\left(\frac{\tau}{e}\right)\right)+A(\tau;\zeta)\right]$$
$$\left.-\frac{r}{e}e^{-i\zeta\tau/e}f_1\!\left(\frac{\tau}{e}\right)\overline{h^{1/2}}(\alpha e\beta_\lambda)\frac{1}{\overline{g^{1/2}}(\alpha e\beta_\lambda)}\right\}. \tag{52}$$

This expression now contains no discrete summations. The complex conjugate, $\overline{\hat{\phi}_\lambda}(\tau)$, can easily be obtained from this expression by noting that the complex conjugate of an integral of the form

$$\int_C d\zeta\, P(\zeta)$$

for some function P is just

$$\int_{\bar{C}} d\zeta\,\bar{P}(\bar{\zeta}),$$

where $\bar{C}$ is a contour below and parallel to the real axis. Note that $\bar{P}(\bar{\zeta})$ is analytic in ζ if P is.

For the evaluation of $\Sigma_\lambda\hat{\phi}_\lambda(\tau')\overline{\hat{\phi}_\lambda}(\tau'')$, then we need the following sums over λ:

$$\frac{2e}{\sigma^2}\sum_\lambda\frac{g(\alpha e\beta_\lambda)}{(\zeta-e\beta_\lambda)(\zeta'-e\beta_\lambda)}=-1+\frac{\hat{H}(\zeta)-\hat{H}(\zeta')}{\zeta'-\zeta},$$
$$\frac{2er}{\sigma^2}\sum_\lambda\frac{\overline{h^{1/2}}(\alpha e\beta_\lambda)g^{1/2}(\alpha e\beta_\lambda)}{(\zeta-e\beta_\lambda)(\zeta'-e\beta_\lambda)}=\frac{\hat{G}(\zeta)-G(\zeta')}{\zeta'-\zeta},$$
$$\frac{2er}{\sigma^2}\sum_\lambda\frac{h^{1/2}(\alpha e\beta_\lambda)\overline{g^{1/2}}(\alpha e\beta_\lambda)}{(\zeta-e\beta_\lambda)(\zeta'-e\beta_\lambda)}=\frac{\tilde{G}(\zeta)-\tilde{G}(\zeta')}{\zeta'-\zeta},$$
$$\frac{2er^2}{\sigma^2}\sum_\lambda\frac{h(\alpha e\beta_\lambda)}{(\zeta-e\beta_\lambda)(\zeta'-e\beta_\lambda)}=\frac{\hat{E}(\zeta)-\hat{E}(\zeta')}{\zeta'-\zeta}. \tag{53}$$

These results follow directly from the definitions of the functions $\hat{H}$, $\hat{G}$, $\tilde{G}$, and

$$\hat{E}(\zeta)\equiv\frac{2er^2}{\sigma^2}\sum_\lambda\frac{h(\alpha e\beta_\lambda)}{\zeta-e\beta_\lambda}.$$

In the infinite-system limit, E becomes a sectionally holomorphic function:

$$\hat{E}(\zeta)\to\frac{2r^2}{\pi}\int_0^\infty d\lambda\,\frac{h(\alpha\lambda)}{\zeta-\lambda}. \tag{54}$$

The results (52) and (53) now yield

$$\sum_\lambda\hat{\phi}_\lambda(\tau')\overline{\hat{\phi}_\lambda}(\tau'')$$
$$=\frac{1}{(2\pi)^2}\int_C d\zeta\int_{\bar{C}}d\zeta'$$
$$\times\left\{\left[-1+\frac{\hat{H}(\zeta)-\hat{H}(\zeta')}{\zeta'-\zeta}\right]\frac{B(\tau';\zeta)\bar{B}(\tau'';\bar{\zeta}')}{\hat{H}(\zeta)\hat{H}(\zeta')}\right.$$
$$-\left[\frac{\hat{G}(\zeta)-\hat{G}(\zeta')}{\zeta'-\zeta}\right]\frac{1}{e}e^{-i\zeta\tau'/e}\frac{f_1(\tau'/e)\bar{B}(\tau'',\bar{\zeta}')}{\hat{H}(\zeta')}$$
$$-\left[\frac{\tilde{G}(\zeta)-\tilde{G}(\zeta')}{\zeta'-\zeta}\right]\frac{1}{e}e^{-i\zeta'\tau''/e}\bar{f}_1\!\left(\frac{\tau''}{e}\right)\frac{B(\tau';\zeta)}{\hat{H}(\zeta)}$$
$$+\left[\frac{\hat{E}(\zeta)-\hat{E}(\zeta')}{\zeta'-\zeta}\right]\frac{1}{e}e^{-i\zeta\tau'/e}f_1\!\left(\frac{\tau'}{e}\right)$$
$$\left.\times\frac{1}{e}e^{i\zeta'\tau''/e}\bar{f}_1\!\left(\frac{\tau''}{e}\right)\right\}, \tag{55}$$

where

$$B(\tau;\zeta)\equiv(1/e)e^{-i\zeta\tau/e}[\hat{G}(\zeta)f_1(\tau/e)-f_2(\tau/e)]+A(\tau;\zeta).$$

First we may observe that

$$\frac{1}{2\pi i}\int_C d\zeta\, \frac{B(\tau;\zeta)}{\hat{H}(\zeta)} = 0.$$

The proof is as follows. The integral equals

$$-\frac{2}{e} f_1\Big(\frac{\tau}{e}\Big) f_2\Big(\frac{\tau}{e}\Big) + \frac{1}{2\pi i}\int_C d\zeta\, \frac{A(\tau;\zeta)}{\hat{H}(\zeta)}.$$

Now from Eq. (51) we obtain

$$\frac{1}{2\pi i}\int_C d\zeta\, \frac{A(\tau;\zeta)}{\hat{H}(\zeta)} = \frac{1}{2\pi i}\int_C d\xi\, e^{-i\xi\tau} A(e\xi) \times \frac{1}{2\pi i}\int_{C_2} d\zeta\, \frac{1}{\hat{H}(\zeta)\hat{H}(e\xi-\zeta)}$$

$$= -\frac{1}{2\pi i}\int_C d\xi\, e^{-i\xi\tau} A(e\xi)\hat{H}_1(e\xi) \quad \text{[Eq. (33)]}$$

$$= \frac{2}{2\pi i}\int_C d\xi e^{-i\xi\tau} i\int_0^\infty d\theta\, e^{ie\xi\theta} f_1(\theta) f_2(\theta) \quad \text{[Eq. (45)]}$$

$$= \frac{2}{e} f_1\Big(\frac{\tau}{e}\Big) f_2\Big(\frac{\tau}{e}\Big).$$

Combining results we note the pairwise cancellation of the term $(2/e) f_1(\tau/e) f_2(\tau/e)$, thereby completing the proof.

Equation (55) can now be rearranged into the following form:

$$\begin{aligned}\sum_\lambda \hat{\phi}_\lambda(\tau')\bar{\hat{\phi}}_\lambda(\tau'') = \frac{1}{(2\pi)^2}\int_C d\zeta \int_{\bar{C}} d\zeta'\, \frac{1}{\zeta'-\zeta} &\times\Big\{\Big[\frac{\hat{G}(\zeta')\tilde{G}(\zeta')}{\hat{H}(\zeta')} - \frac{\hat{G}(\zeta)\tilde{G}(\zeta)}{\hat{H}(\zeta)} + \hat{E}(\zeta) - \hat{E}(\zeta')\Big]\frac{1}{e^2} e^{-i\zeta\tau'/e} e^{i\zeta'\tau''/e} f_1\Big(\frac{\tau'}{e}\Big)\bar{f}_1\Big(\frac{\tau''}{e}\Big) \\ &- \Big[\frac{\hat{G}(\zeta)}{\hat{H}(\zeta)} - \frac{\hat{G}(\zeta')}{\hat{H}(\zeta')}\Big]\frac{1}{e} e^{-i\zeta\tau'/e} f_1\Big(\frac{\tau'}{e}\Big)\Big[\bar{A}(\tau'';\bar{\zeta}') - \frac{1}{e} e^{i\zeta'\tau''/e}\bar{f}_2\Big(\frac{\tau''}{e}\Big)\Big] \\ &- \Big[\frac{\tilde{G}(\zeta)}{H(\zeta)} - \frac{\tilde{G}(\zeta')}{H(\zeta')}\Big]\frac{1}{e} e^{i\zeta'\tau''/e}\bar{f}_1\Big(\frac{\tau''}{e}\Big)\Big[A(\tau';\zeta) - \frac{1}{e} e^{-i\zeta\tau'/e} f_2\Big(\frac{\tau'}{e}\Big)\Big] \\ &+ \Big[\frac{1}{\hat{H}(\zeta')} - \frac{1}{\hat{H}(\zeta)}\Big]\Big[A(\tau';\zeta) - \frac{1}{e} e^{-i\zeta\tau'/e} f_2\Big(\frac{\tau'}{e}\Big)\Big]\Big[\bar{A}(\tau'';\bar{\zeta}') - \frac{1}{e} e^{i\zeta\tau''/e}\bar{f}_2\Big(\frac{\tau''}{e}\Big)\Big]\Big\}. \end{aligned} \tag{56}$$

Now if Q is a sectionally holomorphic function with a cut along the positive real axis, and with limiting values $Q^\pm(\zeta)$ on either side of the cut, then we have, for any two functions R_1 and R_2 holomorphic on and below the contour C, and such that their integrals along the closure of C in the lower half-plane tend to zero, that

$$\begin{aligned}\frac{1}{(2\pi i)^2}\int_C d\zeta\int_{\bar{C}} d\zeta'\, \frac{Q(\zeta)-Q(\zeta')}{\zeta'-\zeta} R_1(\zeta)\overline{R_2}(\bar{\zeta}') &= \frac{1}{2\pi i}\int_C d\zeta\, Q(\zeta) R_1(\zeta)\overline{R_2}(\zeta) - \frac{1}{2\pi i}\int_{\bar{C}} d\zeta'\, Q(\zeta') R_1(\zeta')\overline{R_2}(\bar{\zeta}') \\ &= \frac{1}{2\pi i}\int_B d\zeta\, Q(\zeta) R_1(\zeta)\overline{R_2}(\bar{\zeta}) = \frac{1}{2\pi i}\int_0^\infty d\zeta\, R_1(\zeta)\overline{R_2}(\zeta)[Q^+(\zeta) - Q^-(\zeta)]\end{aligned} \tag{57}$$

(where B is the contour of Fig. 2). In order to apply Eq. (57) to (56) in the infinite-system limit, notice from Eqs. (27) and (54) in this limit:

$$\frac{1}{2\pi i}\int_0^\infty d\zeta\, e^{-i\zeta\tau}\Big[\frac{\hat{G}^+(\zeta)\tilde{G}^+(\zeta)}{\hat{H}^+(\zeta)} - \hat{E}^+(\zeta) - \frac{\hat{G}^-(\zeta)\tilde{G}^-(\zeta)}{\hat{H}^-(\zeta)} + \hat{E}^-(\zeta)\Big] = f_3(\tau). \tag{58}$$

Then, the limit of Eq. (56) can be written as

$$\begin{aligned}&\frac{1}{e^2}\Big[f_1\Big(\frac{\tau'}{e}\Big)\bar{f}_1\Big(\frac{\tau''}{e}\Big) f_3\Big(\frac{\tau'-\tau''}{e}\Big) + f_1\Big(\frac{\tau'}{e}\Big)\bar{f}_2\Big(\frac{\tau''}{e}\Big) f_2\Big(\frac{\tau'-\tau''}{e}\Big) \\ &\quad + f_2\Big(\frac{\tau'}{e}\Big)\bar{f}_1\Big(\frac{\tau''}{e}\Big) f_2^*\Big(\frac{\tau'-\tau''}{e}\Big) + f_2\Big(\frac{\tau'}{e}\Big)\bar{f}_2\Big(\frac{\tau''}{e}\Big) f_1\Big(\frac{\tau'-\tau''}{e}\Big)\Big] \\ &\quad + \frac{1}{e} f_1\Big(\frac{\tau'}{e}\Big)\frac{1}{2\pi i}\int_0^\infty d\zeta\Big[\frac{\hat{G}^+(\zeta)}{\hat{H}^+(\zeta)} - \frac{\hat{G}^-(\zeta)}{\hat{H}^-(\zeta)}\Big] e^{-i\zeta\tau'/e}\bar{A}(\tau'';\zeta) \\ &\quad + \frac{1}{e}\bar{f}_1\Big(\frac{\tau''}{e}\Big)\frac{1}{2\pi i}\int_0^\infty d\zeta\Big[\frac{\tilde{G}^+(\zeta)}{\hat{H}^+(\zeta)} - \frac{\tilde{G}^-(\zeta)}{\hat{H}^-(\zeta)}\Big] e^{i\zeta\tau''/e} A(\tau';\zeta) \\ &\quad + \frac{1}{2\pi i}\int_0^\infty d\zeta\Big[\frac{1}{\hat{H}^+(\zeta)} - \frac{1}{\hat{H}^-(\zeta)}\Big]\Big\{A(\tau';\zeta)\bar{A}(\tau'';\zeta) \\ &\quad - \frac{1}{e} f_2\Big(\frac{\tau'}{e}\Big) e^{-i\zeta\tau'/e}\bar{A}(\tau'';\zeta) - \frac{1}{e}\bar{f}_2\Big(\frac{\tau''}{e}\Big) e^{i\zeta\tau''/e} A(\tau';\zeta)\Big\}.\end{aligned} \tag{59}$$

We have now succeeded in writing $\Sigma_\lambda \hat{\phi}_\lambda(\tau')\bar{\hat{\phi}}_\lambda(\tau'')$ in a form free of discrete summations, and thus ready for the infinite-system limit to be taken. Equation (59) is still very messy, however, and we can now embark on substantial simplifications designed to facilitate numerical computation. Since

$$A(\tau;\zeta) = \frac{1}{2\pi i e}\int_C d\xi\, \frac{e^{-i\xi\tau/e}A(\xi)}{\hat{H}(\xi-\zeta)} = \frac{i}{e}\int_0^{\tau/e} d\theta\, e^{-i\zeta\theta} f_1(\theta) a\left(\frac{\tau}{e}-\theta\right) \tag{60}$$

[Eqs. (14) and (49)], we obtain

$$\frac{1}{2\pi i}\int_0^\infty d\zeta \left[\frac{\hat{G}^+(\zeta)}{\hat{H}^+(\zeta)} - \frac{\hat{G}^-(\zeta)}{\hat{H}^-(\zeta)}\right] e^{-i\zeta\tau'/e}\bar{A}(\tau'';\zeta) = \frac{i}{e}\int_0^{\tau''/e} d\theta\, \bar{f}_1(\theta)\bar{a}\left(\frac{\tau''}{e}-\theta\right) f_2\left(\frac{\tau'}{e}-\theta\right). \tag{61}$$

Similar calculations permit us to evaluate all the terms in Eq. (59) which involve integrals over ζ. These terms become

$$\begin{aligned}\frac{1}{e^2}\Bigg[& if_1\left(\frac{\tau'}{e}\right)\int_0^{\tau''/e} d\theta\, \bar{f}_1(\theta)\bar{a}\left(\frac{\tau''}{e}-\theta\right) f_2\left(\frac{\tau'}{e}-\theta\right) - i\bar{f}_1\left(\frac{\tau''}{e}\right)\int_0^{\tau'/e} d\theta\, f_1(\theta) a\left(\frac{\tau'}{e}-\theta\right)\bar{f}_2\left(\frac{\tau''}{e}-\theta\right)\\ & + if_2\left(\frac{\tau'}{e}\right)\int_0^{\tau''/e} d\theta\, \bar{f}_1(\theta)\bar{a}\left(\frac{\tau''}{e}-\theta\right) f_1\left(\frac{\tau'}{e}-\theta\right) - i\bar{f}_2\left(\frac{\tau''}{e}\right)\int_0^{\tau'/e} d\theta\, f_1(\theta) a\left(\frac{\tau'}{e}-\theta\right)\bar{f}_1\left(\frac{\tau''}{e}-\theta\right)\\ & + \int_0^{\tau'/e} d\theta \int_0^{\tau''/e} d\theta'\, f_1(\theta)\bar{f}_1(\theta') a\left(\frac{\tau'}{e}-\theta\right)\bar{a}\left(\frac{\tau''}{e}-\theta'\right) f_1(\theta-\theta')\Bigg].\end{aligned} \tag{62}$$

The next step is to derive an expression for $a(\tau)$ in the infinite-system limit. This expression will turn out to be made up of yet more convolution integrals, with readily calculated limits. From the definition,

$$A(\xi) = -\frac{2i}{\hat{H}_1(\xi)}\int_0^\infty d\theta\, e^{i\xi\theta} f_1(\theta) f_2(\theta). \tag{63}$$

Now it is shown in the previous paper that as $z\to\infty$,

$$2\pi^2/H_1(z) = -\tfrac{1}{2}z + \epsilon_2 + O(z^{-1}).$$

In dimensionless variables, this becomes

$$1/\hat{H}_1(\xi) = \xi - 2/\alpha + O(\xi^{-1}).$$

Consequently,

$$1/\xi\hat{H}_1(\xi) - 1 = O(\xi^{-1})$$

for large ξ, and we can legitimately make the definition

$$h_1(\tau) = \frac{1}{2\pi i}\int_C d\xi\, e^{-i\xi\tau}\left[\frac{1}{\xi\hat{H}_1(\xi)} - 1\right]. \tag{64}$$

If, further, we define

$$f_4(\theta) = \frac{d}{d\theta}\,[f_1(\theta) f_2(\theta)], \tag{65}$$

then an integration by parts in Eq. (63) yields

$$A(\xi) = 2\left[\frac{1}{\xi\hat{H}_1(\xi)} - 1\right]\int_0^\infty d\theta\, e^{i\xi\theta} f_4(\theta) + 2\int_0^\infty d\theta\, e^{i\xi\theta} f_4(\theta). \tag{66}$$

Since $A(\xi)$ is the Laplace transform of $a(\tau)$, we can use Eq. (14) to obtain

$$a(\tau) = -2if_4(\theta) + 2\int_0^\tau d\tau'\, f_4(\tau') h_1(\tau-\tau'). \tag{67}$$

A limiting expression for f_4 is available from those for f_1 and f_2; for h_1, Eq. (64) gives, in the limit,

$$h_1(\tau) = -\frac{1}{\hat{H}_1(0)} + \frac{1}{2\pi i}\int_0^\infty d\zeta\, \frac{e^{-i\zeta\tau}}{\zeta}\left[\frac{1}{H_1^+(\zeta)} - \frac{1}{H_1^-(\zeta)}\right]. \tag{68}$$

Equation (32) allows for easy evaluation of this expression.

The final result of this section can now be written down. It is an expression for $\rho_2(\tau)$ in terms of the function $\hat{\phi}_3(\tau)$ considered in the previous section and five other functions of τ, namely, f_1, f_2, f_3, f_4, and h_1, combined as a sum of numerous products and convolutions. The expression, although long, now has a quite simple structure:

$$\rho_2(\tau) = \sum_\lambda |\hat{\phi}_{2,\lambda}(\tau)|^2 = \int_0^\tau d\tau' \int_0^\tau d\tau''\, \hat{\phi}_3(\tau-\tau')\bar{\hat{\phi}}_3(\tau-\tau'')\left[\sum_\lambda \hat{\phi}_\lambda(\tau')\bar{\hat{\phi}}_\lambda(\tau'')\right], \tag{69}$$

where

$$\sum_\lambda \hat{\phi}_\lambda(\tau')\overline{\hat{\phi}}_\lambda(\tau'') = \frac{1}{e^2}\Big\{f_1\Big(\frac{\tau'}{e}\Big)\overline{f}_1\Big(\frac{\tau''}{e}\Big)f_3\Big(\frac{\tau'-\tau''}{e}\Big) + \overline{f}_2\Big(\frac{\tau''-\tau'}{e}\Big)f_2\Big(\frac{\tau'}{e}\Big)\overline{f}_1\Big(\frac{\tau''}{e}\Big)$$
$$+f_2\Big(\frac{\tau'-\tau''}{e}\Big)f_1\Big(\frac{\tau'}{e}\Big)\overline{f}_2\Big(\frac{\tau''}{e}\Big) + f_1\Big(\frac{\tau'-\tau''}{e}\Big)f_2\Big(\frac{\tau'}{e}\Big)\overline{f}_2\Big(\frac{\tau''}{e}\Big)$$
$$-2f_1\Big(\frac{\tau'}{e}\Big)\int_0^{\tau''/e} d\theta f_2\Big(\frac{\tau'-\tau''}{e}+\theta\Big)\overline{f}_1\Big(\frac{\tau''}{e}-\theta\Big)\Big[\overline{f}_4(\theta) - i\int_0^\theta d\theta'\,\overline{f}_4(\theta')\,\overline{h}_1(\theta-\theta')\Big]$$
$$-2\overline{f}_1\Big(\frac{\tau''}{e}\Big)\int_0^{\tau'/e} d\theta\, \overline{f}_2\Big(\frac{\tau''-\tau'}{e}+\theta\Big)f_1\Big(\frac{\tau'}{e}-\theta\Big)\Big[f_4(\theta) + i\int_0^\theta d\theta'\,f_4(\theta')h_1(\theta-\theta')\Big]$$
$$-2\overline{f}_2\Big(\frac{\tau''}{e}\Big)\int_0^{\tau'/e} d\theta f_1\Big(\frac{\tau'-\tau''}{e}-\theta\Big)f_1\Big(\frac{\tau'}{e}-\theta\Big)\Big[f_4(\theta) + i\int_0^\theta d\theta'\,f_4(\theta')h_1(\theta-\theta')\Big]$$
$$-2f_2\Big(\frac{\tau'}{e}\Big)\int_0^{\tau''/e} d\theta f_1\Big(\frac{\tau'-\tau''}{e}+\theta\Big)\overline{f}_1\Big(\frac{\tau''}{e}-\theta\Big)\Big[\overline{f}_4(\theta) - i\int_0^\theta d\theta'\,\overline{f}_4(\theta')\,\overline{h}_1(\theta-\theta')\Big]$$
$$+4\int_0^{\tau'/e} d\theta\int_0^{\tau''/e} d\theta' f_1\Big(\frac{\tau'-\tau''}{e}-\theta+\theta'\Big)f_1\Big(\frac{\tau'}{e}-\theta\Big)\overline{f}_1\Big(\frac{\tau''}{e}-\theta'\Big)$$
$$\times\Big[f_4(\theta) + i\int_0^\theta d\eta\, f_4(\eta)h_1(\theta-\eta)\Big]\Big[\overline{f}_4(\theta') - i\int_0^{\theta'} d\eta'\,\overline{f}_4(\eta')\,\overline{h}_1(\theta'-\eta')\Big]\Big\}. \tag{70}$$

Again, the limiting forms of this expression for one or more zero coupling functions are interesting. If only $f = 0$, naturally enough there is no simplification in Eq. (70), since the details of the two-stage decay process must still be described. But, if $h = 0$, then $f_2 = f_3 = f_4 = 0$, so $\rho_2(\tau) = 0$. We saw before that the states involving $|2\rangle$ are inaccessible from state $|3\rangle$ in this case. The most interesting case is again that in which $g = 0$. Then

$$f_2 = f_4 = 0,\quad f_1(\tau) = e^{i\tau/\alpha},\quad f_3(\tau) = \frac{2r^2}{\pi}\int_0^\infty d\lambda\, h(\alpha\lambda)e^{-i\lambda\tau}$$

as before. Although we shall not need it, we also note the result that $h_1(\tau) = 2/\alpha$. [Compare Eq. (34) of paper (viii) of Ref. 2.] Then

$$\sum_\lambda \hat{\phi}_\lambda(\tau')\overline{\hat{\phi}}_\lambda(\tau'') = \frac{1}{e^2}f_1\Big(\frac{\tau'}{e}\Big)\overline{f}_1\Big(\frac{\tau''}{e}\Big)f_3\Big(\frac{\tau'-\tau''}{e}\Big)$$
$$= \frac{2r^2}{e^2\pi}\int_0^\infty d\lambda\, h(\alpha\lambda)\exp\Big[-\frac{i}{e}\Big(\frac{1}{\alpha}+\lambda\Big)(\tau'-\tau'')\Big].$$

Then, from Eq. (69),

$$\rho_2(\tau) = \frac{2r^2}{e^2\pi}\int_0^\infty d\lambda\, h(\alpha\lambda)\Big|i\int_0^\tau d\tau'\,\hat{\phi}_3(\tau')\exp\Big[-\frac{i}{e}\Big(\frac{1}{\alpha}+\lambda\Big)(\tau-\tau')\Big]\Big|^2.$$

A quantity of interest is the limit of this as $\tau\to\infty$, i.e., the probability that the system ends up in state $|2\rangle$ when that state cannot decay. The answer can be readily expressed in terms of ψ_3, the Laplace transform of $\hat{\phi}_3(\tau)$:

$$\rho_2(\infty) = \frac{2r^2}{e^2\pi}\int_0^\infty d\lambda\, h(\alpha\lambda)\Big|\psi_3^+\Big(\frac{1}{e}\Big(\lambda+\frac{1}{\alpha}\Big)\Big)\Big|^2. \tag{71}$$

We write ψ_3^+ here since ψ_3 is sectionally holomorphic like the other Laplace transforms of the paper. In fact, for the case $g = 0$, Eq. (35) gives

$$(\psi_3(\zeta))^{-1} = \frac{1}{\alpha} - \zeta - \frac{2s^2}{\pi}\int_0^\infty \frac{d\lambda\, f(\alpha\lambda)}{\lambda-\xi} - \frac{2r^2}{e\pi}\int_0^\infty \frac{d\lambda\, h(\alpha\lambda)}{\lambda+1/\alpha-e\zeta}.$$

Now, if in addition $f = 0$, one expects that $\rho_2(\infty) = 1$, since now state $|1\rangle$ is inaccessible from state $|3\rangle$. This is easily seen to be true, since Eq. (71) can then be written as

$$\rho_2(\infty) = \frac{1}{2\pi ie}\int_0^\infty d\lambda\Big[\psi_3^+\Big(\frac{1}{e}\Big(\lambda+\frac{1}{\alpha}\Big)\Big) - \psi_3^-\Big(\frac{1}{e}\Big(\lambda+\frac{1}{\alpha}\Big)\Big)\Big]$$
$$= \frac{1}{2\pi ie}\int_B d\zeta\,\psi_3\Big(\frac{1}{e}\Big(\zeta+\frac{1}{\alpha}\Big)\Big)$$
$$= 1,$$

since

$$\lim_{\zeta\to\infty}\zeta\psi_3\Big(\frac{1}{e}\Big(\zeta+\frac{1}{\alpha}\Big)\Big) = e.$$

IV. NUMERICAL RESULTS

In the previous sections, exact expressions were derived describing the emission of an excited three-level system in a (one-dimensional) field of radiation in the limit where the system size becomes of infinite extent and the mode spectrum becomes continuous. In the formulation of the problem, it was assumed that the state of the system excited initially was the state $|3\rangle$ and it is now of interest to follow the evolution of the system through the competing channels available for the decay of this quantum system. From Eqs. (30) and (34) derived in Sec. II, one can study the evolution from the excited state $|3\rangle$; then, from Eqs. (69) and (70) of Sec. III, one can determine the influence of the intermediate state $|2\rangle$. The functions f_λ, h_λ, and g_λ which specify the coupling between the states $|3\rangle$ and $|1\rangle$, $|3\rangle$ and $|2\rangle$, and $|2\rangle$ and $|1\rangle$, respectively, must, of course, be assigned before the numerical work can be performed. In our earlier work, computations were carried out for several different choices of coupling function; for example, $f(x)$ was taken as

$$f(x) = x^{-1/4},$$

$$f(x) = x^{-1/2}, \tag{72}$$

$$f(x) = 4x/(1+x)^2. \tag{73}$$

In the problem of the two-level atom, it was found[2v] that for sufficiently small values of the coupling parameter α, all of the above coupling functions led to strictly ergodic behavior in the time evolution of the system; however, our calculations also showed that the first two functions $f(x)$ allowed nonergodic behavior. Here we shall focus on results generated for three-level systems using the (always when $\alpha < \pi/8$) "ergodic" form factor, Eq. (73). Moreover, in this study, we have restricted our choice of the parameters of the model to the following set of values: $\alpha = 0.1$, $e = 2$, and $r = s = 1$. Specifying the parameter $e = \epsilon_3/\epsilon_2 = 2$ means that the three levels of the system are equally spaced with respect to the energy gap between the states $|3\rangle$ and $|2\rangle$, and $|2\rangle$ and $|1\rangle$. Setting $r = s = 1$ [see the definitions (8)] means that the functions f_λ, h_λ, and g_λ coupling the levels of the problem (see Fig. 1) are placed on an equal footing. Given these specifications, the results plotted in Fig. 3 were obtained, and we turn now to a discussion of the evolution profiles, $\rho_3(\tau)$ vs. τ and $\rho_2(\tau)$ vs. τ, displayed therein.

An acid test of the reliability of the present calculations in the accuracy with which the initial conditions of the problem, $\rho_3(\tau = 0) = 1.0$ and $\rho_2(\tau = 0) = 0.0$, are achieved. Although it is clear from the structure of Eqs. (69) and (70) that the probability $\rho_2(\tau)$ at $\tau = 0$ must vanish identically, the quantity $\rho_3(\tau = 0)$ can be computed. We determined that $\rho_3(\tau = 0) = 0.9892$, an accuracy comparable to that realized in the finite system calculation[1] when $\sigma^2 = 1.0$ [$\rho_3(\tau = 0) = 0.9928$] and somewhat better than that achieved when $\sigma^2 = 10.0$ [$\rho_3(\tau = 0) = 0.9405$]. A second, rather demanding test of the accuracy of the numerical work involves the function $\rho_2(\tau)$. Given the structure of the expression (70), it is seen that the real and imaginary parts of the wavefunction $\hat{\phi}_2(\tau)$ appearing therein are "mixed." Therefore, in the calculation of $\rho_2(\tau)$ not only should the real part of $\rho_2(\tau)$ be bounded between 0 and 1 (it is), but the imaginary part of $\rho_2(\tau)$ for all τ should vanish; the explicit calculation of Im $\rho_2(\tau)$ yielded a maximum value of 0.0089 at the maximum in the curve, $\rho_2(\tau)$ vs. τ, with the magnitude of this contribution much less elsewhere. For reasons of economy these computations were done in single precision; it is likely that the above checks can be improved in going to double precision.

Results corresponding to the above set of parameters of the model ($\alpha = 0.1$, $e = 2$, and $r = s = 1$) and various coupling functions were obtained previously for the discrete spectrum problem, and evolution profiles for $f(x) = x^{-1/4}$ and $f(x) = x^{-1/2}$ were displayed explicitly in Ref. 1. For small values of the coupling constant ($\alpha = 0.1$), it was found that the specific choice of coupling function had a relatively small (numerical) effect on the calculated evolution of $\rho_3(\tau)$, both for small systems ($\sigma^2 = 1.0$) and larger ones ($\sigma^2 = 10.0$); more noticeable differences in the evolution of the system were found in studying the probability $\rho_2(\tau)$. For both probabilities, the evolution determined for small α was much more sensitive to the other parameters of the model (e,r,s). To determine whether for given specification of e, r, and s these trends for small α persist in the limit $\sigma^2 \to \infty$, we focus here on a comparison of the discrete vs. continuous spectrum results for the same parameter specification ($\alpha = 0.1$, $e = 2$, $r = s = 1$), but for two different choices of coupling function, namely, Eqs. (72) and (73).

Both for finite systems (see Figs. 3 and 10 in Ref. 1 corresponding to $\sigma^2 = 1.0$ and $\sigma^2 = 10.0$, respectively) and the infinite one studied here (see Fig. 3), our calculations show that the *initial* decay of the system from the excited state $|3\rangle$, as monitored by the quantity $\rho_3(\tau)$, takes place on a time scale which seems to be relatively unaffected by the particular form factor $f(x)$ adopted in the calculations. In fact, the closeness with which the initial decays from the state $|3\rangle$ are in correspondence for finite versus infinite systems is quite remarkable, as can be seen in Fig. 4, where we

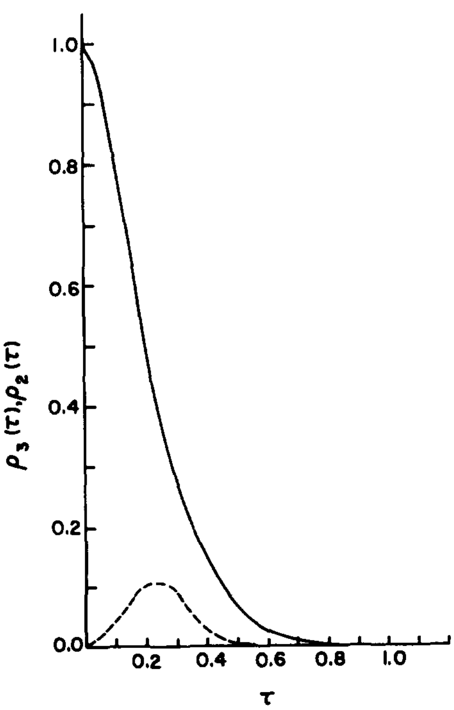


FIG. 3. The temporal behavior of the three-level atom studied in this paper. The parameters which characterize the system are $\alpha = 0.1$, $f(x) = 4x/(1+x)^2$, $e = 2$, $r = 1$, and $s = 1$. The solid curve describes the evolution of $\rho_3(\tau)$ while the dashed line describes the evolution of $\rho_2(\tau)$. We determine $\rho_3(0) = 0.9892$.

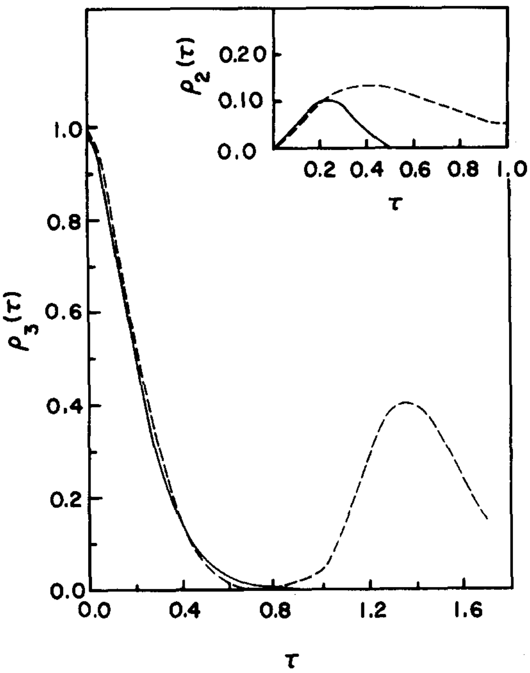


FIG. 4. A comparison of the effect of system size on the evolution from $|3\rangle$ and $|2\rangle$. The solid curves give the quantities $\rho_3(\tau)$ and $\rho_2(\tau)$ as determined from expressions (34) and (69), respectively, of this paper; here, $\alpha = 0.1$, $r = s = 1.0$, $e = 2.0$, and $f(x) = 4x/(1+x)^2$. The dotted curves give the quantities $\rho_3(\tau)$ and $\rho_2(\tau)$ as determined from expressions (40) and (41), respectively, of Ref. 1; in addition to the parameters specified above, we assign $\sigma^2 = 1.0$ and choose $f(x) = x^{-1/2}$.

display the evolution of $\rho_3(\tau)$ for $\sigma^2 = 1.0$ under the coupling function (72) and for $\sigma^2 \to \infty$ under the coupling function (73). For all intents and purposes, the decays are essentially coincident up to times $\tau \sim 1.0$; beyond this time there arise Poincaré recurrences in the evolution of the three-level quantum system in interaction with a *finite* number of modes of the field (the case $\sigma^2 = 1.0$), whereas the system remains de-excited for all times $\tau > 1.0$ in the continuous spectrum case. In both calculations, one finds a slight "shoulder" in the evolution curve in the vicinity of $\tau = 0$; the emergence of this behavior in exactly solvable, dynamical models characterized by a finite coupling constant α, and the collapse to strictly exponential decay in the weak-coupling limit ($\alpha \to 0$) have been discussed in our earlier work.[1,2viii]

Differences in the discrete vs. continuous spectrum problem seem to manifest themselves rather more dramatically in the evolution from the intermediate state $|2\rangle$ of the three-level system. Over the range of τ for which meaningful comparisons can be made ($\tau \leqslant 1.0$), the probability $\rho_2(\tau)$ of the system's being in the state $|2\rangle$ persists for a longer time in the discrete spectrum problem than in the continuous spectrum one; moreover, the maximum probability $\rho_2(\tau)$ realized in the former case is somewhat greater than that achieved in the latter one, over that same range of τ. It is worth noticing that the slightly faster decay of the function $\rho_3(\tau)$ for (very) short times for the case $\sigma^2 \to \infty$ vs. $\sigma^2 = 1.0$ is accompanied by a slightly greater probability of finding the system in the state $|2\rangle$ over that same time scale.

The conclusion which emerges from these calculations of $\rho_3(\tau)$ and $\rho_2(\tau)$ is that it is *not* the decay from the initial, excited state $|3\rangle$ that is most affected by the passage from a discrete to a continuous spectrum of radiation, but rather the decay from the intermediate state $|2\rangle$. We have already noted for finite systems[1] the apparently greater sensitivity of the probability $\rho_2(\tau)$ to the coupling function and to the parameters of the model. Since the probability $\rho_2(\tau)$ that the atom be in the state $|2\rangle$ at time τ reflects the importance of a competing decay channel available to a three-level quantum system, it is clearly of great interest to study the interplay between "direct" de-excitation events ($|3\rangle \to |1\rangle$) versus those that proceed through the intermediate state $|2\rangle$ as a function of the coupling function (with various choices of r and s), the coupling constant α, and the level splitting parameter e. This study, which is underway, will have a direct bearing on the assessment of earlier theories which have treated the problem of a two- or three-level quantum system in interaction with a radiation field in an approximate or phenomenological way, this owing to the absence of an exact solution to the time-dependent Schrödinger equation for the problem. However, with our earlier work on two level quantum systems (Ref. 2) and the work in Ref. 1 and in the present study on three-level ones, we now have available an *exact* quantum-statistical theory for such problems. It may be anticipated that the worth of these earlier theories will be decided on the basis of the success with which the detailed behavior of the probability $\rho_2(\tau)$ is reproduced as the parameters of the model are changed.

V. CONCLUDING REMARKS

In this paper, we have undertaken the study of a model for a three-level system in interaction with a continuous spectrum of radiation, and have succeeded in obtaining an exact solution to this important, quantum-statistical problem. The complexity of the derivations of this paper and the difficulty in evaluating numerically the expressions derived will serve, we hope, as good excuses for our not complicating the model at this stage with features such as three space dimensions, angular momentum, multipole expansions of the radiation field, and so on. Our main objective here was to display the essential mathematical issues that are involved and to show how the attendant analytical difficulties could be handled (exactly). It should be clear, however, that *no* difficulties of principle stand in the way of incorporating such features.

Exact solutions of models with an infinite number of modes or with continuous spectra are quite rare, and now that the present model has been solved it is worth commenting on the uses to which the theory and its generalizations can be put. Of perhaps greatest interest (to us) are the various theoretical and conceptual questions about the interaction of matter and radiation that are still outstanding. It is worth emphasizing that the $L \to \infty$ limit introduces inrreversible behavior and permits discussion of competing decay channels without any ad hoc assumptions such as exponential decay. Rather, the actual form of decay emerges as a consequence of the model, so that one can examine the circumstances in which it can be reasonably described as exponential. Situations in which *non*exponential decay can arise and the further problems of ergodicity and mixing in quantum dynamical systems[7] can also be studied without approximation. The explicit treatment here of a three-level quantum system allows, as well, some assessment of the extent to which these general notions pertain to concrete problems in spectroscopy, viz., fluorescence and phosphorescence. Taken together with the results obtained in our exact analysis of a two-level quantum system in interaction with a discrete [Ref. 2(iv)] or continuous [Ref. 2(v)] spectrum of radiation, we may, for example, calculate the line shape and line broadening in a variety of situations. As noted in the preceding section, the results obtained may be compared with those derived using approximate theories, e.g., the semiclassical theory of matter-radiation interaction, and a definitive statement can then be made concerning the regime of applicability of these approximate theories. At the very least, such a study should allow one to develop better approximations and thus enhance the range of applicability of these earlier, more intuitive theories.

[1]R. Davidson and J. J. Kozak, J. Math. Phys. **17**, 1692 (1976).

[2](i) R. Davidson and J. J. Kozak, J. Math. Phys. **11**, 189 (1970); (ii) R. Davidson and J. J. Kozak, J. Math. Phys. **11**, 1420 (1970); (iii) R. Davidson and J. J. Kozak, J. Math. Phys. **12**, 903 (1971); (iv) R. Davidson and J. J. Kozak, J. Math. Phys. **14**, 414 (1973); (v) R. Davidson and J. J. Kozak, J. Math. Phys. **14**, 425 (1973); (vi) J. J.-W. Yang, R. Davidson, and J. J. Kozak, J. Math. Phys. **15**, 491 (1974); (vii) R. Davidson and J. J. Kozak, J. Math. Phys. **16**, 1013 (1975); (viii) R. Davidson and J. J. Kozak, J. Math. Phys. **19**, 1074 (1978).

[3]See, for example, the discussions in *Matrix Isolation Spectroscopy* edited by A. J. Barnes, W. J. Orville-Thomas, A. Müller, and R. Gaufrès (Reidel, Dordrecht, 1981), especially the article by A. J. Rest, "Electronic spectroscopy of matrix isolated solutes," p. 27.

[4]M. Hillery, Phys. Rev. A **24**, 933 (1981).

[5](i) I. R. Senitzky, *Progress in Optics*, edited by E. Wolf (North-Holland, Amsterdam, 1978), Vol. XVI, p. 415; (ii) M. D. Crisp and E. T. Jaynes, Phys. Rev. **179**, 1253 (1969); **185** 2046 (1969); (iii) W. E. Lamb and M. O. Scully in *Polarisation: Matière et Rayonnement* (Société Francaise de Physïque, Paris, 1969); (iv) J. F. Clauser, Phys. Rev. A **6**, 49 (1972); (v) V. P. Maslov and M. V. Fedoriuk, *Semi-Classical Approximation in Quantum Mechanics* (Reidel, Dordrecht, 1981).

[6]For definitions and properties of sectionally holomorphic functions, see N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, 1953).

[7]See, for example, R. H. G. Helleman, "Self-Generated Chaotic Behavior in Nonlinear Mechanics," in *Fundamental Problems in Statistical Mechanics*, Vol. 5, edited by E. D. G. Cohen (North Holland, Amsterdam, 1980).

Stationary wave envelopes in nonlinear optics

Pranab Krishna Chanda and Dipankar Ray
Department of Applied Mathematics, Calcutta University, 92, A.P.C. Road, Calcutta 700 009, India

(Received 27 May 1982; accepted for publication 24 September 1982)

Small has given a set of equations for stationary electromagnetic wave envelopes in a cubically nonlinear two-dimensional medium and has obtained some particular solutions. These solutions are generalized. Some solutions of the Ginzburg–Pitaevski equation for superfluid are obtained as a by-product.

PACS numbers: 42.65. – k, 03.40.Kf, 67.90. + z

I. INTRODUCTION

In a recent paper[1] Small has obtained the following set of equations for stationary electromagnetic wave envelopes in a cubically nonlinear two-dimensional medium.

$$a_{xx} + a_{yy} - (\phi_x^2 + \phi_y^2 - k_0^2)a + \beta a^3 = 0,$$

$$a_x = \frac{\partial a}{\partial x}, \quad \text{etc.} \tag{1a}$$

$$(a^2\phi_x)_x + (a^2\phi_y)_y = 0, \quad \phi_x = \frac{\partial \phi}{\partial x}, \text{ etc.,} \tag{1b}$$

where the electric field is given by

$$\mathbf{E} = (a(x,y)\cos[\phi(x,y) - \omega t],0,0) \tag{2a}$$

for linearly polarized waves and by

$$\mathbf{E} = (a(x,y)\cos[\phi(x,y) - \omega t],a(x,y)\sin[\theta(x,y) - \omega t],0) \tag{2b}$$

for circularly polarized waves. Wave number vector

$$\mathbf{k}(x,y) \equiv \nabla\phi \equiv \left(\frac{\partial \phi}{\partial x},\frac{\partial \phi}{\partial y},0\right), \tag{3}$$

k_0 and β are constants, and

$$F(|\mathbf{P}|^2) = \beta a^2 \tag{4}$$

represents the nonlinearity of the medium where $\mathbf{P}$ is the polarization vector.

Equation (1) has been solved by Small when a and ϕ are functions of x only. In the present paper we seek solutions of (1) for (i) $a = a(x)$ and (ii) $a = a(r)$ where $r^2 = x^2 + y^2$ with no restriction on ϕ.

2. SOLUTIONS

Case (i): $a = a(x)$, $a_x \neq 0$

Here (1) is reduced to

$$\phi_x^2 + \phi_y^2 = \alpha(x), \tag{5a}$$

$$\alpha(x) = (a_{xx} + k_0^2 a + \beta a^3)/a, \tag{5b}$$

and

$$a(\phi_{xx} + \phi_{yy}) + 2a_x\phi_x = 0. \tag{5c}$$

It will be shown that (5) leads to

$$\phi_y = \text{const} = B \text{ (say)} \tag{6}$$

If $\phi_y = 0, \phi_x \neq 0$, then (6) is automatically satisfied with

$$B = 0.$$

If $\phi_y \neq 0$, $\phi_x = 0$, then the left-hand side of (5a) is a function of y only and the right-hand side is a function of x only. Hence both are constants. Hence, $\phi_y = $ const and (6) is satisfied.

If $\phi_x \neq 0$, $\phi_y \neq 0$, one can proceed as follows. Differentiating (5a) with respect to y,

$$\phi_x\phi_{xy} + \phi_y\phi_{yy} = 0.$$

Hence, eliminating ϕ_{yy} by use of (5c) and then dividing throughout by ϕ_y/a and integrating,

$$\frac{a^2\phi_x}{\phi_y} = \gamma(y),$$

where $\gamma(y)$ is an unspecified function of y. This readily gives

$$\phi = \phi(u),$$

where

$$u = X + Y, \quad X = \int \frac{dx}{a^2}, \quad Y = \int \frac{dy}{\gamma(y)}.$$

Using the above relations we get from (5a)

$$\phi_u^2(X_x^2 + Y_y^2) = \alpha(x)$$

or

$$\frac{X_x^2}{\alpha(x)} + \frac{Y_y^2}{\alpha(x)} = \frac{1}{\phi_u^2}.$$

Differentiating the above equation separately with respect to X and Y, respectively, and comparing the results,

$$\alpha(x)\left[\frac{X_x^2}{\alpha(x)}\right]_X + \alpha(x)\left[\frac{1}{\alpha(x)}\right]_X Y_y^2 = [Y_y^2]_Y.$$

Differentiating again successively with respect to X and Y,

$$\left\{\alpha(x)\left[\frac{1}{\alpha(x)}\right]_X\right\}_X [Y_y^2]_Y = 0.$$

Hence, either

$$\alpha(x)[1/\alpha(x)]_X = \text{const}$$

or

$$Y_y = \text{const.}$$

That $\alpha(x)[1/\alpha(x)]_X = $ const is not possible will be shown in Appendix A. From $Y_y = $ const and $Y = \int dy/\gamma(y)$ we have

$$\gamma(y) = \text{const}.$$

Differentiating $[X_x^2/\alpha(x) + Y_y^2/\alpha(x)] = 1/\phi_u^2$ with respect to Y and using $Y_y = $ const, we get

 0022-2488/83/071923-04$02.50

$$\left[\frac{1}{\phi_u^2}\right]_u = 0,$$

which gives $\phi_u = \text{const}$,

Hence, $\phi_y = \phi_u u_y = \phi_u Y_y = \text{const}$. Thus in this case of $\phi_x \neq 0$, $\phi_y \neq 0$ we also see that (6) is satisfied. Hence (6) is true for all cases. Now, using (6) and (5c) we get

$$\phi_x = A/a^2, \quad \text{where } A \text{ is a constant.} \tag{7}$$

Generalizing (6) and (7), we get

$$\phi = A\int \frac{dx}{a^2} + By. \tag{8}$$

Using (8) together with (5a) and (5b), we get

$$a_{xx} = A^2/a^3 + (B^2 - k_0^2)a - \beta a^3$$

or

$$\frac{d}{da}(a_x^2) = \frac{2A^2}{a^3} + 2(B^2 - k_0^2)a - 2\beta a^3,$$

or

$$\int \frac{\sqrt{2}a\, da}{[-\beta a^6 + 2(B^2 - k_0^2)a^4 + 2Ka^2 - 2A^2]^{1/2}} = x,$$

where K is a constant. The above equation can be rewritten in a more compact form as follows:

$$\int \frac{dZ}{[-2\beta Z^3 + 4(B^2 - k_0^2)Z^2 + 4KZ - 4A^2]^{1/2}} = x, \quad Z = a^2. \tag{9}$$

The integral in Eq. (9) is of the form of an elliptic integral and hence can be expressed in terms of standard elliptic integrals. Equation (9) then gives a implicity. ϕ is then given by (8).

Case (ii): $a = a(r)$, $a_r \neq 0$, **where** $r^2 = x^2 + y^2$, **tan** $\theta = y/x$

Here (1) is reduced to

$$r^2\phi_r^2 + \phi_\theta^2 = \delta(r)r^2, \tag{10a}$$

$$\delta(r) = \frac{(1/r)(ra_r)_r + k_0^2 a + \beta a^3}{a}, \tag{10b}$$

$$a\left[\frac{1}{r}(r\phi_r)_r + \frac{1}{r^2}\phi_{\theta\theta}\right] + 2a_r\phi_r = 0. \tag{10c}$$

It will be shown that (10) leads to

$$\phi_\theta = \text{const} = D \text{ (say)}. \tag{11}$$

If $\phi_\theta = 0$, $\phi_r \neq 0$ then (11) is automatically satisfied with $D = 0$. If $\phi_\theta \neq 0$, $\phi_r = 0$ then the left-hand side of (10a) is a function of θ only and the right-hand side is a function of r only. Hence both are constants. Hence, $\phi_\theta = \text{const}$ and thus (11) becomes satisfied.

If $\phi_\theta \neq 0$, $\phi_r \neq 0$ one can proceed as follows. Differentiating (10a) with respect to θ,

$$\phi_r\phi_{r\theta} + \frac{1}{r^2}\phi_\theta\phi_{\theta\theta} = 0.$$

Hence eliminating $\phi_{\theta\theta}$ by use of (10c) and then dividing throughout by ϕ_θ/ar and integrating,

$$\frac{a^2 r\phi_r}{\phi_\theta} = \chi(\theta),$$

where $\chi(\theta)$ is an unspecified function of θ. This readily gives,

$$\phi = \phi(v),$$

where

$$v = \rho + \Theta, \quad \rho = \int \frac{dr}{a^2 r}, \quad \Theta = \int \frac{d\theta}{\chi(\theta)}.$$

Using the above relations we get from (10a)

$$\phi_v^2\left(\rho_r^2 + \frac{\Theta_\theta^2}{r^2}\right) = \delta(r),$$

or

$$\frac{\rho_r^2}{\delta(r)} + \frac{\Theta_\theta^2}{r^2\delta(r)} = \frac{1}{\phi_v^2}.$$

Differentiating the above equation separately with respect to ρ and Θ, respectively, and comparing the results,

$$r^2\delta(r)\left[\frac{\rho_r^2}{\delta(r)}\right]_\rho + r^2\delta(r)\left[\frac{1}{r^2\delta(r)}\right]_\rho \Theta_\theta^2 = [\Theta_\theta^2]_\Theta.$$

Differentiating again successively with respect to ρ and Θ,

$$\{r^2\delta(r)[1/r^2\delta(r)]_\rho\}_\rho [\Theta_\theta^2]_\Theta = 0.$$

Hence, either

$$r^2\delta(r)[1/r^2\delta(r)]_\rho = \text{const}$$

or

$$\Theta_\theta = \text{const}.$$

That $r^2\delta(r)[1/r^2\delta(r)]_\rho = \text{const}$ is not possible will be shown in Appendix B.

From $\Theta_\theta = \text{const}$ and $\Theta = \int d\theta/\chi(\theta)$ we have $\chi(\theta) = \text{const}$. Differentiating

$$\rho_r^2/\delta(r) + \Theta_\theta^2/r^2\delta(r) = 1/\phi_v^2$$

with respect to Θ and using $\Theta_\theta = \text{const}$, $[1/\phi_v^2]_v = 0$, which gives $\phi_v = \text{const.}$ and hence, $\phi_\theta = \phi_v v_\theta = \theta_v \Theta_\theta = \text{const}$. Thus in this case of $\phi_r \neq 0$, $\phi_\theta \neq 0$ we also see that (11) is satisfied. Hence (11) is true for all cases. Now, using (11) and (10c), we get

$$\phi_r = C/ra^2, \tag{12}$$

where C is a constant. Generalizing (11) and (12) we get

$$\phi = C\int \frac{dr}{ra^2} + D\theta. \tag{13}$$

Using (13) together with (10a) and (10b) we get

$$r(ra_r)_r = C^2/a^3 + r^2[a(D^2/r^2 - k_0^2) - \beta a^3]. \tag{14}$$

a is implicitly given by (14). ϕ is then given by (13).

Case (iii): $a = $ **const**

Here (1) reduced to

$$\phi_x^2 + \phi_y^2 = \beta a^2 = \text{const}, \tag{15a}$$

$$\phi_{xx} + \phi_{yy} = 0. \tag{15b}$$

Differentiating (15a) separately with respect to x and y, respectively, adding and then using (15b), we get $\phi_{xy} = 0$, which gives

$$\phi = \xi(x) + \eta(y). \quad (16)$$

Eliminating ϕ from (16) and (15a) we get

$$\xi_x = \text{const} = L, \quad (17a)$$

$$\eta_y = \text{const} = M. \quad (17b)$$

Using (17a) and (17b) in (16) and then from (15) we get

$$\phi = Lx + My + N, \quad (18a)$$

where N is a constant, and

$$a = (1/\beta)(L^2 + M^2)^{1/2}. \quad (18b)$$

3. CONCLUSION

Summarily Eqs. (1) have been completely integrated for $a = a(x)$ and have been reduced to a single second-order differential equation for $a = a(r)$.

For $a = a(x)$ the solutions are given by (8) and (9) if $a \neq \text{const}$ and by (18) if $a = \text{const}$. For $a = a(r)$ Eqs. (1) have been reduced to (14) where ϕ is given by (13).

It is interesting to note that for $a = a(x)$, ϕ is a function of both x and y. Therefore, even when amplitude depends on x only, the phase depends on both x and y. Likewise we see for (13) and (14) that even when the amplitude depends on r only the phase depends on r and θ.

We further note that for $k_0 = 1$ and $\beta = -1$, Eqs. (1) reduce to Ginzburg–Pitaevski equations[2] for superfluids where the wavefunction of the superfluid is given by $a \exp(i\phi)$. Thus the present paper gives some solutions for the Ginzburg–Pitaevski equations for superfluids as well.

APPENDIX A

It is evident that when

$$\alpha(x)\left[\frac{1}{\alpha(x)}\right]_x = \text{const} = J \quad (A1)$$

there must be

$$\alpha(x)\left[\frac{X_x^2}{\alpha(x)}\right]_x = \text{const} = H. \quad (A2)$$

When $J \neq 0$ and $H \neq 0$, dividing (A2) by (A1) and using $X_x = 1/a^2$, we get

$$\alpha = (1 - ba^4)/da^4, \quad (A3)$$

where b and d are constants. From (A1) and using $X_x = 1/a^2$

$$\frac{\alpha_x}{\alpha} = -\frac{J}{a^2}. \quad (A4)$$

Eliminating α from (A3) and (A4) we get

$$a_x = J(1 - ba^4)/4a. \quad (A5)$$

Comparing (A3) and (5b), we get

$$\frac{(1 - ba^4)}{da^4} = \frac{a_{xx}}{a} + k_0^2 + \beta a^2. \quad (A6)$$

Expressing a_{xx} in (A6) in terms of a_x as in (A5),

$$\frac{1 - ba^4}{da^4} = -\frac{J^2}{16}\frac{(1 - ba^4)(1 + 3ba^4)}{a^4} + k_0^2 + \beta a^2. \quad (A7)$$

(A7) is an equation in a which can be satisfied only with discrete values of a. This means that a is a constant, which is not the case considered here.

When $J \neq 0$, $H = 0$, from (A2), and using $X_x = 1/a^2$, we get

$$\alpha = 1/a^4 g, \quad (A8)$$

where g is a constant. From (A1) and $X_x = 1/a^2$ we get

$$\frac{\alpha_x}{\alpha} = -\frac{J}{a^2}. \quad (A9)$$

Eliminating α from (A8) and (A9) we get

$$a_x = J/4a. \quad (A10)$$

Comparing (A8) and (5b) we get

$$\frac{1}{a^4 g} = \frac{a_{xx}}{a} + k_0^2 + \beta a^2. \quad (A11)$$

Expressing a_{xx} in (A11) in terms of a_x as in (A10),

$$\beta a^6 + k_0^2 a^4 = \left(\frac{1}{g} + \frac{J^2}{16}\right). \quad (A12)$$

(A12) is an equation in a which can be satisfied only with discrete values of a. This means that a is a constant which is not the case considered here.

When $J = 0$, $H \neq 0$, from (A1), we get

$$\alpha = 1/h, \quad (A13)$$

where h is a constant. Eliminating α from (A2) and (A13) and using $X_x = 1/a^2$,

$$1/a^4 = HX + s, \quad (A14)$$

where s is a constant. Differentiating (A14) with respect to x and using $X_x = 1/a^2$,

$$a_x = -(H/4)a^3. \quad (A15)$$

Comparing (A13) and (5b)

$$\frac{1}{h} = \frac{a_{xx}}{a} + k_0^2 + \beta a^2. \quad (A16)$$

Expressing a_{xx} in (A16) in terms of a_x as in (A15)

$$\frac{1}{h} = \frac{3H^2 a^4}{16} + \beta a^2 + k_0^2. \quad (A17)$$

(A17) is an equation in a which can be satisfied only with discrete values of a. This means that a is a constant, which is not the case considered here.

When $J = 0$, $H = 0$, from (A1), we get $\alpha = 1/h$, where h is a constant. From (A2), we get $a = 1/a^4 g$, where g is a constant. Eliminating α from above two relations we get, $a = (h/g)^{1/4} = \text{const}$, which is not the case considered here.

APPENDIX B

It is evident that when

$$r^2\delta(r)\left[\frac{1}{r^2\delta(r)}\right]_\rho = \text{const} = Q, \quad (B1)$$

there must be

$$r^2\delta(r)\left[\frac{\rho_r^2}{\delta(r)}\right]_\rho = \text{const} = R, \quad (B2)$$

When $Q \neq 0$, $R \neq 0$, dividing (B2) by (B1) and using $\rho_r = 1/a^2 r$ we get

$$\delta = \frac{(1-pa^4)}{qa^4}e^{-2n}, \tag{B3}$$

where p and q are constants and $\ln r = n$. From (B1) and using $\rho_r = 1/a^2 r$ we get

$$\frac{\delta_n}{\delta} = -\frac{Q}{a^2} - 2, \text{ where } \ln r = n. \tag{B4}$$

Eliminating δ from (B3) and (B4) we get

$$a_n = Q(1-pa^4)/4a. \tag{B5}$$

Comparing (B3) and (10b) and using $\ln r = n$,

$$\frac{(1-pa^4)}{qa^4}e^{-2n} = \frac{a_{nn}}{a}e^{-2n} + k_0^2 + \beta a^2. \tag{B6}$$

Expressing a_{nn} in (B6) in terms of a_n as in (B5),

$$\frac{(1-pa^4)}{qa^4}e^{-2n} = -\frac{Q}{16}\frac{(1-pa^4)(1+3pa^4)}{a^4}e^{-2n} + k_0^2 + \beta a^2. \tag{B7}$$

Differentiating with respect to n, and using (B5),

$$\frac{8a}{Q(1-pa^4)} = \frac{-4pa^3}{(1-pa^4)} + \frac{12Qpqa^3}{16+Qq+3Qpqa^4} - \frac{4}{a} - \frac{2\beta a}{k_0^2+\beta a^2}. \tag{B8}$$

(B8) is an equation in a which is satisfied only for discrete values of a. This means that a is a constant, which is not the case considered here.

When $Q \neq 0$, $R = 0$, from (B2) and using $\rho_r = 1/ra^2$,

$$\delta = \frac{1}{r^2a^4m} = \frac{e^{-2n}}{a^4m}, \tag{B9}$$

where m is a constant and $\ln r = n$. From (B1) and $\rho_r = 1/ra^2$,

$$\frac{\delta_n}{\delta} = -\frac{Q}{a^2} - 2, \quad \text{where } \ln r = n. \tag{B10}$$

Eliminating δ from (B9) and (B10),

$$a_n = Q/4a. \tag{B11}$$

Comparing (B9) and (10b) and using $\ln r = n$,

$$\frac{e^{-2n}}{a^4m} = \frac{a_{nn}}{a}e^{-2n} + k_0^2 + \beta a^2. \tag{B12}$$

Expressing a_{nn} in (B12) in terms of a_n as in (B11),

$$\frac{e^{-2n}}{a^4m} = -\frac{Q^2}{16a^4}e^{-2n} + k_0^2 + \beta a^2. \tag{B13}$$

Simplifying,

$$2n = \ln\left[\frac{(1/m + Q^2/16)}{(k_0^2+\beta a^2)a^4}\right].$$

Differentiating with respect to n and using (B11),

$$4\beta a^4 + (4k_0^2 + 3Q\beta)a^2 + 2Qk_0^2 = 0. \tag{B14}$$

(B14) is an equation in a which is satisfied only for discrete values of a. This means that a is a constant, which is not the case considered here.

When $Q = 0$, $R \neq 0$, from (B1) and using $\rho_r = 1/ra^2$, we get

$$\delta = \frac{1}{r^2 l} = \frac{e^{-2n}}{l} \tag{B15}$$

where l is a constant and $\ln r = n$. Eliminating δ from (B2) and (B15) and using

$$\rho_r = 1/ra^2,\ 1/a^4 = R\rho + T, \quad \text{where } T = \text{constant}. \tag{B16}$$

Differentiating (B16) with respect to n and using $\rho_r = 1/ra^2$,

$$a_n = -(R/4)a^3. \tag{B17}$$

Comparing (B15) and (10b), and using $\ln r = n$,

$$\frac{e^{-2n}}{l} = \frac{3R^2}{16}a^4e^{-2n} + k_0^2 + \beta a^2. \tag{B18}$$

Simplifying,

$$2n = \ln\left[\frac{16 - 3lR^2a^4}{(k_0^2+\beta a^2)(16l)}\right].$$

Differentiating with respect to n and using (B17),

$$6lR^3a^6(k_0^2+\beta a^2) + R\beta a^4(16 - 3lR^2a^4) = 4(16 - 3lR^2a^4)(k_0^2+\beta a^2). \tag{B19}$$

(B19) is an equation in a which is satisfied only for discrete values of a. This means that a is a constant, which is not the case considered here.

When $Q = 0$, $R = 0$, from (B1), we get $\delta = 1/r^2 l$, where l is a constant. From (B2), we get $\delta = 1/r^2a^4m$, where m is a constant. Eliminating δ from above two relations, $a = (l/m)^{1/4} =$ const, which is not the case considered here.

[1]R. D. Small, J. Math. Phys. **22**, 1497 (1981).

[2]V. L. Ginzburg and L. V. Pitaevski, Zh. Eksp. Teor. Fiz. **34**, 1240 (1958) [Sov. Phys. JETP 7, 858 (1958)].

Critical behavior of the spin van der Waals model[a)]

R. K. Pathria
Guelph-Waterloo Program for Graduate Work in Physics, Waterloo Campus, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

(Received 14 December 1982; accepted for publication 21 January 1983)

Thermodynamic behavior in the critical region of the spin van der Waals model is considered when the number of spins in the system is large but finite ($1 \ll N < \infty$). The specific heat curve is shown to possess a steep, yet smooth, maximum which turns into a singularity as the number of spins becomes infinite. In the case of the finite system, this maximum occurs at $\epsilon = \epsilon^* = O(N^{-1/3})$, where $\epsilon = (T - T_c)/T_c$, T_c being the critical temperature of the infinite system.

PACS numbers: 64.60.Cn, 05.70.Fh, 65.40. – f

I. INTRODUCTION

In a recent paper, Lee[1] has made an investigative study of the onset of long-range order in the spin van der Waals model when the number of spins in the system is finite ($N < \infty$) and also when it is infinite ($N \to \infty$). He has shown explicitly that, whereas an infinite system at low temperatures can support long-range order, a finite one cannot. Valid though it is, this result may leave one with the (erroneous) impression that the behavior of a finite system, in this case the spin van der Waals model, is completely devoid of features characteristic of a phase transition. This, of course, is true if one is looking, solely and strictly, for *spontaneous* magnetization, as indeed was the case with Lee's analysis, but is no longer true if one is observing properties such as the specific heat of the system. In that case, one would expect a "smoothed-out" behavior[2] characterized by a steep maximum in the specific-heat curve, whose precise height and location depend on the size of the system, such that, in the limit $N \to \infty$, it turns into a discontinuity at $T = T_c$. Thus, for any finite value of N, so long as it is large enough from a statistical point of view, one would expect to encounter a situation which, though nonsingular in principle, is indistinguishable in practice from the ones commonly associated with a phase transition.

To demonstrate this point, the partition function of the system has to be evaluated with a much greater degree of care than is ordinarily done because some of the approximations customarily made in such studies are valid only in the thermodynamic limit ($N \to \infty$), and may cause serious errors if they are admitted into the study of finite systems as well. Keeping this in mind, I have carried out a rigorous analysis of the specific heat of the finite-sized spin van der Waals model, which brings out very clearly the cooperative features alluded to in the foregoing.

II. BASIC FEATURES OF THE SPIN VAN DER WAALS MODEL

The spin van der Waals model consists of an aggregate of N $\frac{1}{2}$-spins situated on the sites of a regular lattice. The field-free partition function of the system is given by[1,3]

$$Z_N = \sum_{S=0}^{(1/2)N} \sum_{S_z=-S}^{S} g(S) e^{-\beta E(S, S_z)}, \tag{1}$$

where S and S_z refer to the "total spin" of the system, $E(S, S_z)$ is the eigenvalue of the Hamiltonian $\mathscr{H}$ associated with the state (S, S_z),

$$E(S, S_z) = -N^{-1}[JS(S+1) - \lambda S_z^2] \quad (J > 0), \tag{2}$$

while $g(S)$ is the corresponding degeneracy factor

$$g(S) = \binom{N}{\frac{1}{2}N - S} - \binom{N}{\frac{1}{2}N - S - 1}. \tag{3}$$

It will be noted that N here has been assumed to be an even number; the case of odd N can be treated likewise though, asymptotically, the results in the two cases should be identical.

For simplicity, we shall confine our analysis to the isotropic version of the model, viz., the one with $\lambda = 0$. The summation over S_z then yields a straightforward factor of $(2S + 1)$ and the partition function takes the form

$$Z_N = \sum_{S=0}^{(1/2)N} f(S), \tag{4}$$

where

$$f(S) = \binom{N}{\frac{1}{2}N - S} \frac{(2S+1)^2}{\frac{1}{2}N + S + 1} e^{\beta J N^{-1} S(S+1)}. \tag{5}$$

The function $f(S)$, which is a measure of the relative probabilities of S assuming different values, leads to the ratio

$$r(S) \equiv \frac{f(S)}{f(S-1)} = \frac{\frac{1}{2}N - S + 1}{\frac{1}{2}N + S + 1} \left(\frac{2S+1}{2S-1}\right)^2 e^{2\beta J N^{-1} S} \quad (S = 1, 2, \ldots, \tfrac{1}{2}N). \tag{6}$$

At the lower end of the summation, we have

$$r(1) = [9/(1 + 4N^{-1})] e^{2\beta J N^{-1}} \simeq 9 e^{2(\beta J - 2)N^{-1}}, \tag{7}$$

which is greater than unity for all values of the interaction parameter βJ, though it clearly indicates that βJ may have a critical value equal to 2. At the upper end of the summation, we have instead

$$r(\tfrac{1}{2}N) = \frac{N+1}{(N-1)^2} e^{\beta J} \simeq \frac{1}{N} e^{\beta J}, \tag{8}$$

which is considerably less than unity, except when the system is at so low a temperature that βJ is of order $\ln N$. For all

[a)] Work supported in part by the Natural Sciences and Engineering Research Council of Canada.

 0022-2488/83/071927-05$02.50

practical purposes, therefore, the summand in (4) is a nonmonotonic function of S, passing through a maximum at $S = S^*$, say; S^* is then the "most probable" value of S. Ordinarily one would expect S^* to be identical with the "mean value" $\bar{S}$; in the critical region, however, this may not be the case [see Eqs. (35)–(37)].

The precise value of S^* may be determined by equating the ratio (6) to unity, which leads to the implicit relation

$$\ln\left(\frac{\frac{1}{2}N + S^* + 1}{\frac{1}{2}N - S^* + 1}\right) - 2\ln\left(\frac{2S^* + 1}{2S^* - 1}\right) - 2\beta J\frac{S^*}{N} = 0. \tag{9}$$

If we are looking for an S^* of order N, then the middle term, being of order N^{-1}, drops out and we obtain the asymptotic relation

$$m^* = \tanh(\tfrac{1}{2}\beta J m^*), \quad (m = 2S/N), \tag{10}$$

which is formally identical with the well-known mean-field result of Weiss.[4] We thus find that the spin van der Waals model undergoes a phase transition at $\beta J = (\beta J)_c = 2$, i.e., $m^* = 0$ for $\beta J \leqslant 2$ and $0 < m^* \leqslant 1$ for $\beta J > 2$.

This simplistic picture changes dramatically when we recognize that between $S^* = 0$ on one hand and $0 < S^* = O(N)$ on the other, there exist regions in which $S^* = O(N^\alpha)$, with α lying between 0 and 1.[5] For instance, right at $\beta J = 2$, our basic relation (9) takes the asymptotic form

$$\left(\frac{4S^*}{N} + \frac{16S^{*3}}{3N^3} + \cdots\right) - \left(\frac{2}{S^*} + \cdots\right) - \frac{4S^*}{N} = 0, \tag{11}$$

which shows that the middle term is no longer negligible, with the result that S^* turns out to be nonzero and of order $N^{3/4}$:

$$(S^*)_{\beta J = 2} \simeq (3N^3/8)^{1/4}. \tag{12}$$

For βJ significantly less than 2, we have instead

$$\left(\frac{4S^*}{N} + \cdots\right) - \left(\frac{2}{S^*} + \cdots\right) - \frac{2\beta J S^*}{N} = 0, \tag{13}$$

whence it follows that

$$S^* \simeq \left(\frac{N}{2 - \beta J}\right)^{1/2} = O(N^{1/2}), \tag{14}$$

indicating the absence of cooperative behavior at higher temperatures. The region in the vicinity of $\beta J = 2$, therefore, requires a rather incisive, albeit cautious, investigation.

For this we introduce a parameter ϵ, defined by

$$\beta J = 2(1 - \epsilon), \tag{15}$$

such that ϵ increases with temperature. Our basic relation now becomes

$$\left(\frac{4S^*}{N} + \frac{16S^{*3}}{3N^3} + \frac{64S^{*5}}{5N^5} + \cdots\right) - \left(\frac{2}{S^*} + \cdots\right) - \frac{4(1-\epsilon)S^*}{N} = 0. \tag{16}$$

Now, so long as S^* is of order less than $N^{5/6}$, this relation reduces to

$$\frac{16S^{*3}}{3N^3} - \frac{2}{S^*} + \frac{4\epsilon S^*}{N} = 0, \tag{17}$$

with the result that

$$S^{*2} = \tfrac{1}{8}[(9\epsilon^2 N^4 + 24N^3)^{1/2} - 3\epsilon N^2]. \tag{18}$$

Note that, for $\epsilon = 0$, (18) reduces to our earlier result (12). In fact, so long as $|\epsilon|$ is of order $N^{-1/2}$, S^* is of order $N^{3/4}$; clearly, this may be regarded as the very bosom of the critical region.

If ϵ is positive and of order greater than $N^{-1/2}$, then (18) yields the result

$$S^{*2} \simeq (N/2\epsilon), \tag{19}$$

in agreement with (14). If, on the other hand, ϵ is negative and $|\epsilon|$ of order greater than $N^{-1/2}$, we obtain

$$S^{*2} \simeq \frac{3|\epsilon|N^2}{4}\left(1 + \frac{2}{3\epsilon^2 N}\right). \tag{20}$$

Finally, if $|\epsilon|$ in this case tends to be of order N^0 but stays much less than unity, then S^* becomes of order N and Eq. (16) reduces to

$$\frac{16S^{*3}}{3N^3} + \frac{64S^{*5}}{5N^5} + \frac{4\epsilon S^*}{N} = 0, \tag{21}$$

with the result that

$$S^{*2} = \frac{N^2}{24}[(25 + 180|\epsilon|)^{1/2} - 5] \simeq \frac{3|\epsilon|N^2}{4}\left[1 - \frac{9}{5}|\epsilon|\right]. \tag{22}$$

From the foregoing considerations, one might naively infer that the maximum of the specific heat curve lies in the region where $S^* = O(N^{3/4})$, i.e., where $|\epsilon| = O(N^{-1/2})$. Detailed analysis shows that this is not the case.

III. SPECIFIC HEAT IN THE CRITICAL REGION

The internal energy of the system under study is given by [see Eq. (2)]

$$U = -N^{-1}J(\overline{S^2} + \bar{S}). \tag{23}$$

Throughout the region of interest, the ratio $\bar{S}/\overline{S^2}$, being at most of order $N^{-1/2}$, is negligible; the foregoing expression may, therefore, be simplified to

$$U \simeq -N^{-1}J\,\overline{S^2} = -N^{-1}J(\bar{S}^2 + \Delta^2), \tag{24}$$

where Δ^2 denotes the variance of the variable S. In the region represented by (22), Δ is negligible in comparison with $\bar{S}$, and $\bar{S}$ is practically identical with S^*, with the result that the internal energy in this region is given by

$$U \simeq -N^{-1}JS^{*2} \simeq -\tfrac{3}{4}NJ|\epsilon|[1 - \tfrac{9}{5}|\epsilon|] \tag{25}$$

and the specific heat by

$$\frac{C}{Nk} = \frac{1}{Nk}\frac{\partial U}{\partial T} \simeq -\frac{\beta^2 J^2}{2N^2}\frac{\partial S^{*2}}{\partial \epsilon} \simeq \tfrac{3}{2}[1 - \tfrac{18}{5}|\epsilon|]. \tag{26}$$

We note that, as $|\epsilon|$ decreases, the specific heat does approach the standard value $\frac{3}{2}$ of C/Nk at the critical point $(\beta J = 2)$ of the infinite system. However, the slope of the specific heat curve remains *positive* throughout the region in which the foregoing approximation holds. We, on the other hand, wish to locate a smooth maximum, with vanishing slope, somewhere in the neighborhood of the erstwhile critical point $\epsilon = 0$. To achieve this goal, we shall now carry out a rigorous analysis of the specific heat of the system in this neighborhood.

For this, we first of all need to examine the "degree of steepness' of the maximum of the probability distribution function $f(S)$. This can be done by looking at the quantity

$$h(S) \equiv \frac{r(S+1)}{r(S)} - 1$$
$$= \left[\frac{(\frac{1}{2}N - S)(\frac{1}{2}N + S + 1)}{(\frac{1}{2}N - S + 1)(\frac{1}{2}N - S + 2)} \times \frac{(2S-1)^2(2S+3)^2}{(2S+1)^4} e^{2\beta J N^{-1}} \right] - 1 \quad (27)$$

at $S = S^*$, which serves as a useful and reliable measure of the variance of S; in fact, $\Delta^2 \propto |h(S^*)|^{-1}$. One readily finds that

$$h(S^*) \simeq -\frac{4N}{N^2 - 4S^{*2}} - \frac{2}{S^{*2}} + \frac{2\beta J}{N}, \quad (28)$$

where S^* is determined by the basic relation (16). It follows that, if S^* is of order N, the middle term of (28) drops out and the root-mean-square deviation Δ turns out to be of order $N^{1/2}$. Asymptotically, therefore, $\bar{S} = S^*$, thus justifying the passage from Eq. (24) to (25). This, however, will not be generally true in the critical region. For instance, at $\beta J = 2$, where S^* is given by (12), $h(S^*)$ turns out to be

$$h(S^*) \simeq -\frac{16S^{*2}}{N^3} - \frac{2}{S^{*2}} = -\frac{8}{S^{*2}}, \quad (29)$$

so that $\Delta = O(S^*)$ and hence $\bar{S}$ and S^*, while being of the same order in N, may differ from one another by a significant factor. A detailed study of $h(S^*)$, in conjunction with the basic relation (16), shows that whenever S is of order $N^{3/4}$ or less, the foregoing situation prevails. Accordingly, for such values of ϵ, the passage from (24) to (25) is not justified.

To deal with this situation, we have to make a more thorough use of the probability distribution function $f(S)$, which may now be approximated as

$$f(S) = \text{const} \times S^2 \exp\left[-\frac{2\epsilon S^2}{N} - \frac{4S^4}{3N^3} \right]. \quad (30)$$

The most probable value of S for the probability distribution (30) is indeed the same as the one given by Eq. (17). The mean and the mean-square values of S, however, are now given by

$$\bar{S} = \frac{2}{\sqrt{\pi}} \left(\frac{3N^3}{8} \right)^{1/4} D_{-2}(z)/D_{-3/2}(z) \quad (31)$$

and

$$\overline{S^2} = \frac{3}{2} \left(\frac{3N^3}{8} \right)^{1/2} D_{-5/2}(z)/D_{-3/2}(z), \quad (32)$$

where $D_p(z)$ are the parabolic cylinder functions[6] and z is the "scaled parameter" of the problem, defined by the relation

$$z = \left(\frac{3N}{2} \right)^{1/2} \epsilon \simeq \left(\frac{3N}{2} \right)^{1/2} \frac{T - T_c}{T_c}; \quad (33)$$

here T_c denotes the critical temperature of the infinite system. Noting that, for $\nu > 0$,

$$D_{-\nu}(0) = 2^{(1/2)\nu - 1} \Gamma(\tfrac{1}{2}\nu)/\Gamma(\nu), \quad (34)$$

we obtain at the erstwhile critical point ($z = 0$)

$$\bar{S}(0) = \frac{1}{\Gamma(\frac{3}{4})} \left(\frac{3N^3}{4} \right)^{1/4} = \frac{2^{1/4}}{\Gamma(\frac{3}{4})} S^*(0) \quad (35)$$

and

$$\overline{S^2}(0) = \frac{\Gamma(\frac{5}{4})}{\Gamma(\frac{3}{4})} \left(\frac{3N^3}{4} \right)^{1/2} = \frac{2^{1/2}\Gamma(\frac{5}{4})}{\Gamma(\frac{3}{4})} S^{*2}(0). \quad (36)$$

These results show very clearly that the variance of S is no longer negligible; in fact,

$$\frac{\overline{S^2}(0)}{\bar{S}^2(0)} = \Gamma(\tfrac{5}{4})\Gamma(\tfrac{3}{4}) = \frac{\pi}{\sqrt{8}} > 1. \quad (37)$$

For the specific heat, we now have

$$\frac{C}{Nk} = -\frac{2(1-\epsilon)^2}{N^2} \frac{\partial \overline{S^2}}{\partial \epsilon}; \quad (38)$$

cf. (26). Introducing the Whittaker functions $U(a,z)$, which are directly related to the parabolic cylinder functions through the relation[7]

$$U(a,z) = D_{-a-1/2}(z), \quad (39)$$

we obtain

$$\frac{C}{Nk} \simeq \frac{9}{4} \frac{U'(1,z)U(2,z) - U(1,z)U'(2,z)}{\{U(1,z)\}^2}$$
$$= \frac{9}{4} \left[1 - z\frac{U(2,z)}{U(1,z)} - \frac{3}{2} \left\{ \frac{U(2,z)}{U(1,z)} \right\}^2 \right]. \quad (40)$$

The temperature derivative of the specific heat is in turn given by

$$\frac{\partial}{\partial T}\left(\frac{C}{Nk} \right) = \frac{9}{4T_c} \left(\frac{3N}{2} \right)^{1/2} \left\{ -\frac{U(2,z)}{U(1,z)} + \left[3\frac{U(2,z)}{U(1,z)} + z \right] \times \left[1 - z\frac{U(2,z)}{U(1,z)} - \frac{3}{2} \left\{ \frac{U(2,z)}{U(1,z)} \right\}^2 \right] \right\}. \quad (41)$$

At $z = 0$, we obtain

$$\frac{C(0)}{Nk} = \frac{9}{4} \left[1 - \frac{3}{4} \left\{ \frac{\Gamma(\frac{5}{4})}{\Gamma(\frac{3}{4})} \right\}^2 \right] = 0.6087, \quad (42)$$

which is considerably less than the "standard" value $\frac{3}{2}$, as given by Eq. (26). The slope of the specific heat curve at $z = 0$ turns out to be $-0.3621N^{1/2}/T_c$. The fact that this is of order $N^{1/2}$ is not surprising because, in the vicinity of the point $z = 0$, the specific heat changes by a magnitude $O(1)$ over a temperature range $O(N^{-1/2})$. The fact that it is negative implies that the maximum of the specific heat curve lies in the region $z < 0$. The question then arises: is the corresponding value of $|z|$ of order unity?

To resolve this question, it seems worthwhile that we first examine the situation for large values of $|z|$. For this we make use of the asymptotic expansion of the parabolic cylinder functions appropriate to the domain where $z < 0$ and $|z| \gg 1$, viz.,[8]

$$D_{-\nu}(z) = -\frac{\sqrt{2\pi}}{\Gamma(\nu)} e^{-\nu\pi i} e^{(1/4)z^2} z^{\nu-1} \times \left\{1 + \frac{(\nu-1)(\nu-2)}{2z^2} + \frac{(\nu-1)(\nu-2)(\nu-3)(\nu-4)}{2.4z^4} + \cdots\right\}, \tag{43}$$

which leads to the much simpler result

$$\frac{D_{-\nu-1}(z)}{D_{-\nu}(z)} = -\frac{z}{\nu}\left\{1 + \frac{\nu-1}{z^2} - \frac{(\nu-1)(\nu-2)}{z^4} + \cdots\right\}. \tag{44}$$

Accordingly,

$$\frac{C}{Nk} \simeq -\frac{9}{4}\frac{\partial}{\partial z}\frac{D_{-5/2}(z)}{D_{-3/2}(z)} = \frac{3}{2}\left\{1 - \frac{1}{2z^2} - \frac{3}{4z^4} + \cdots\right\}, \tag{45}$$

which does approach the "standard" value $\frac{3}{2}$ but with a slope that remains throughout *negative*, though diminishing in magnitude as $|z|^{-3}$. Clearly, the specific heat maximum does not lie in the region where $|z| = O(1)$; it rather lies in a region where $|z|$ is infinitely large, i.e., where $|\epsilon|$ is of order greater than $N^{-1/2}$ [see Eq. (33)].

To proceed further, we require an improved version of the probability distribution function $f(S)$, which is readily found to be

$$f(S) = \text{const} \times S^2 \exp\left[-\frac{2\epsilon S^2}{N} - \frac{4S^4}{3N^3} - \frac{32S^6}{15N^5}\right]. \tag{30a}$$

Replacing the last exponential factor by its expansion in powers of (S^6/N^5), we obtain

$$\overline{S^2} = \left(\frac{3N^3}{8}\right)^{1/2} \times \frac{\Sigma_{n=0}^{\infty}(1/n!)\{-\frac{1}{5}\sqrt{6/N}\}^n \Gamma(3n+\frac{5}{2}) D_{-3n-5/2}(z)}{\Sigma_{n=0}^{\infty}(1/n!)\{-\frac{1}{5}\sqrt{6/N}\}^n \Gamma(3n+\frac{3}{2}) D_{-3n-3/2}(z)}, \tag{32a}$$

the parameter z being the same as defined earlier in Eq. (33). Utilizing the asymptotic expansion (43) for the function $D_{-\nu}(z)$, for $z<0$ and $|z| \gg 1$, we obtain

$$\overline{S^2} = \frac{3N^2(-\epsilon)\Sigma_{n=0}^{\infty}(1/n!)\eta^n\,[1 + (3n+\frac{3}{2})(3n+\frac{1}{2})/2z^2 + \cdots]}{4\,\Sigma_{n=0}^{\infty}(1/n!)\eta^n\,[1 + (3n+\frac{1}{2})(3n-\frac{1}{2})/2z^2 + \cdots]}, \tag{46}$$

where

$$\eta = \frac{\sqrt{6}}{5}\frac{z^3}{N^{1/2}} = \frac{9}{10}\epsilon^3 N < 0. \tag{47}$$

The leading terms of the series appearing in (46) yield the bulk result for $\overline{S^2}$, viz., $3N^2(-\epsilon)/4$. Other terms yield a multiplying factor of

$$\frac{e^\eta[1 + (9\eta^2 + 15\eta + \frac{3}{4})/2z^2 + \cdots]}{e^\eta[1 + (9\eta^2 + 9\eta - \frac{1}{4})/2z^2 + \cdots]} \simeq 1 + \frac{6\eta+1}{2z^2}, \tag{48}$$

whence it follows that

$$\overline{S^2} \simeq \frac{3N^2(-\epsilon)}{4}\left[1 + \left(\frac{9}{5}\epsilon + \frac{1}{3\epsilon^2 N}\right)\right]. \tag{49}$$

A comparison of (49) with Eqs. (20) and (22) shows that the difference between the quantities $\overline{S^2}$ and S^{*2}, which was quite glaring in the region $|z| = O(1)$, is present in the region under study as well, though here it is relatively subtle. For a further elucidation of this point see the Appendix.

A quick glance at the more accurate expression for $\overline{S^2}$ suggests that the reversal in the specific heat curve might take place in the region in which the two correction terms appearing in (49) become of the same order of magnitude, i.e., where $|\epsilon^3 N| = O(1)$ and hence $|z| = O(N^{1/6}) \gg 1$—in conformity with the conclusion arrived at following Eq. (45). For a precise location of the specific heat maximum, we substitute (49) into (38) to obtain

$$\frac{C}{Nk} \simeq \frac{3}{2}\left[1 + \left(\frac{8}{5}\epsilon - \frac{1}{3\epsilon^2 N}\right)\right]. \tag{50}$$

It is now straightforward to see that the specific heat of the system possesses a smooth maximum at $\epsilon^* = -(5/12N)^{1/3}$, the value of the maximum being

$$C^*/Nk \simeq \tfrac{3}{2}[1 + \tfrac{12}{5}\epsilon^*]. \tag{51}$$

This agrees with the standard value $\frac{3}{2}$ in the limit $N \to \infty$; for a finite system, however, we have a correction term of order $N^{-1/3}$. Needless to say, in the thermodynamic limit, the critical region, which we have examined here in rather minute detail, collapses into a single, critical point ($\epsilon = 0$) and, with it, several of the results reported in this paper get submerged into the singular behavior of the system at that point.

APPENDIX

To elucidate the slight, but significant, difference between the quantities $\overline{S^2}$ and S^{*2} in the relevant region of interest, we express the probability distribution function $f(S)$ in terms of the variable $y = (S - S^*)/S^*$, where S^* is determined by maximizing the expression (30a), i.e., by solving Eq. (16). In the present approximation, this gives

$$S^{*2} \simeq \frac{3N^2(-\epsilon)}{4}\left[1 + \left(\frac{2}{3\epsilon^2 N} + \frac{9}{5}\epsilon\right)\right]; \tag{A1}$$

cf. Eqs. (20) and (22), which hold in different domains of ϵ. The probability distribution function then takes the form

$$f(S) \equiv \exp[g(S)] = \exp[g(S^*) + \tfrac{1}{2}g''(S^*)S^{*2}y^2 + \tfrac{1}{6}g'''(S^*)S^{*3}y^3 + \cdots], \tag{A2}$$

where

$$g''S^{*2} \simeq -6\epsilon^2 N, \quad g'''S^{*3} \simeq -18\epsilon^2 N, \cdots. \tag{A3}$$

For $|\epsilon| = O(N^{-1/3})$, the root-mean-square deviation in y would be $O(N^{-1/6})$. As will be seen later, the mean value of y turns out to be considerably smaller—in fact $O(N^{-1/3})$. In view of this, the lower and upper limits on y, which are -1 and $(N-2S^*)/2S^*$ $[=O(N^{1/6})]$, may be replaced by $-\infty$ and $+\infty$, respectively. The mean-square value of S is then given by

$$\begin{aligned}\overline{S^2} &= S^{*2}\ \overline{(1+y)^2}\\ &= S^{*2}\left[1+\left(\frac{g'''S^{*3}}{g''^2S^{*4}}+\frac{1}{|g''|S^{*2}}\right)+\cdots\right]\\ &= S^{*2}\left[1-\frac{1}{3\epsilon^2N}+\cdots\right].\end{aligned} \tag{A4}$$

Substituting (A1) into (A4), we obtain the desired result (49). It appears worthwhile to point out here that although the relative difference between $\overline{S^2}$ and S^{*2} turns out to be $O(N^{-1/3})$, which for many purposes is unimportant, it does affect the numerical value of ϵ^*, which we are trying to determine accurately enough so as to establish the existence of a smooth maximum in the specific heat curve of the given system.

[1]M. H. Lee, J. Math. Phys. **23**, 464 (1982).

[2]R. K. Pathria, Phys. Rev. A **5**, 1451 (1972); S. Greenspoon and R. K. Pathria, Phys. Rev. A **8**, 2657 (1973); **11**, 1080 (1975).

[3]R. Dekeyser and M. H. Lee, Phys. Rev. B **19**, 265 (1979); D. K. Dacol, Physica (Utrecht) A **100**, 496 (1980); I. M. Kim and M. H. Lee, Phys. Rev. B **24**, 3961 (1981).

[4]R. K. Pathria, *Statistical Mechanics* (Pergamon, Oxford, 1972).

[5]S. Greenspoon and R. K. Pathria, Phys. Rev. A **9**, 2103 (1974); C. S. Zasada and R. K. Pathria, Phys. Rev. A **14**, 1269 (1976); **15**, 2439 (1977); H. R. Pajkowski and R. K. Pathria, J. Phys. A **10**, 561 (1977).

[6]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1980).

[7]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).

[8]E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U.P., London, 1927).

ERRATA

Erratum: Partial-range completeness and existence of solutions to two-way diffusion equations [J. Math. Phys. 22, 954 (1981)]

Richard Beals
Mathematics Department, Yale University, New Haven, Connecticut 06520

(Received 18 January 1983; accepted for publication 27 January 1983)

PACS numbers: 02.30.Jr, 99.10. + g

The stated results for two of the three cases considered are inaccurate due to lack of care in following Ref. 4. The indicated functions, though complete, are not independent. If $\mu_0 = 0$ and $\int h(\theta)d\theta < 0$ then $\{u_k; \lambda_k > 0\}$ are independent and complete where $h(\theta) > 0$, while if $\int h(\theta)d\theta > 0$ the constant function 1 must be included. Changing signs, one obtains the statement where $h(\theta) < 0$. Details will appear elsewhere.