On exomorphic types of phase transitions

D. B. Litvin
Department of Physics, The Pennsylvania State University, The Berks Campus, P.O. Box 2150, Reading, Pennsylvania 19608

J. Fuksa and V. Kopsky
Institute of Physics, Czechoslovak Academy of Science, Na Slovance 2, P.O. Box 24, 180 40 Praha 8, Czechoslovakia

(Received 21 May 1985; accepted for publication 22 October 1985)

An algorithmic method is presented to determine the irreducible representations that engender the irreducible representations associated with phase transitions involving a change of symmetry to a subgroup of index n. This method is based on the work of Ascher and Kobayashi [E. Ascher and J. Kobayashi, J. Phys. C **10**, 1349 (1977)] and the derivation of faithful irreducible representations contained in the permutation representation of transitive subgroups of permutation groups S_n. Character tables of all such irreducible representations, and their epikernels, associated with a change in symmetry to a subgroup of index $n = 2, 3, 4, 5$, and 6 are given explicitly. The relationship to exomorphic types of phase transitions is then discussed. The irreducible representations associated with the phase transitions O_h^1 to C_{4v}^1 in $BaTiO_3$ and D_{6h}^4 to D_{2h}^{16} in β-K_2SO_4 are derived and it is shown that these two phase transitions belong to the same exomorphic type.

I. INTRODUCTION

The use of group-theoretical methods to investigate structural phase transitions was introduced by Landau[1] over forty years ago. In the Landau method of determining the change of symmetry accompanying a phase transition, the lower symmetry phase is described by a density function, which is expanded in terms of basis functions of the irreducible representations of the higher symmetry phase. With the coefficients of the density function expansion as variational order parameters, a thermodynamic potential is constructed and minimized to determine the form of the density function and subsequently the symmetry of the lower symmetry phase.[2,3] The most extensive tabulations of changes in symmetry accompanying phase transitions derived using this method have been given by Toledano and Toledano.[4]

A number of necessary group-theoretical criteria have also been derived for use in determining the change in symmetry accompanying a phase transition.[3,5–9] These include the subduction criterion, chain subduction criteria, also called the chain criterion,[8] the Landau criterion for continuous phase transitions, and the Lifshitz homogeneity criterion. Using some or all of these criteria, tabulations of possible lower-phase symmetries have been derived for some phase transitions in crystals. For cases where the higher-phase symmetry group is a cubic space group, such tabulations have been given for O_h^1 by Goldrich and Birman[3] and Vinberg *et al.*,[10] for O_h^3 by Jaric,[9] and for O_h^5 by Sutton and Armstrong[11] and Ghozlen and Mlik.[12] Recently a computer program has been developed by Hatch and Stokes[13] and all the above mentioned criteria have been applied to all 230 space groups.

In parallel with the application of the Landau method with minimization, and the development and application of group-theoretical criteria, investigations into general theorems that apply to the change in symmetry accompanying a phase transition have also been developed. Such general theorems date back to the original papers of Landau.[1] It was shown by Landau that the irreducible representation associated with a phase transition, where the lower-phase symmetry group is a subgroup of index 2 of the higher-phase symmetry group, is a one-dimensional alternating irreducible representation. It was also conjectured that no phase transition between a higher-phase symmetry group and a lower-phase symmetry subgroup of index 3 is continuous. This so-called subgroup of index 3 theorem was shown to be valid for special cases by Anderson and Blout[14] and Boccara.[15] General proofs were subsequently given by Meisel, Gray, and Brown[16] and Brown and Meisel.[17] It has also been shown that the Landau subgroup of index 3 theorem cannot be extended to a subgroup of index n theorem with $n \neq 3$.[18]

Continuing the investigation into the group-theoretical aspects of phase transitions, Ascher and Kobayashi[19] have introduced the so-called "inverse Landau problem." This problem is to determine the irreducible representation associated with a phase transition between a given higher-phase symmetry group and a given lower-phase symmetry group. Following the work of Gufan and Sakhnenko[20] and Ascher and Kobayashi,[19] Kopsky has introduced the concept of "exomorphic" types of phase transitions.[21–24] For example, all phase transitions between a higher-phase symmetry group and lower-phase symmetry subgroup of index 2 belong to a single exomorphic type. Such a concept stresses the mathematical similarity among phase transitions and can be used in the study of the general properties of phase transitions. Two phase transitions belonging to the same exomorphic type have, for example, the same set of order parameters and the same mathematical form of the thermodynamic potential. The transitions can, however, differ in the physical interpretation of the order parameters and corresponding terms in the potential can be of different physical importance.[22] The concept of exomorphic types of phase transi-

 0022-2488/86/030661-07$02.50

tions can also be used as a basis of proofs of general theorems concerning phase transitions as, for example, in the alternate proof of the subgroup of index 3 theorem.[24]

In this paper we continue the study of exomorphic types of phase transitions. In Sec. II we briefly review the method of Ascher and Kobayashi and its connection to the subduction criterion.[3] We give an algorithmic method to determine the irreducible representations associated with a phase transition between a higher-phase symmetry group and a lower-phase symmetry subgroup of index n. We then determine and tabulate the irreducible representations that engender all irreducible representations associated with phase transitions where the subgroup index $n = 2, 3, 4, 5$, and 6. For each irreducible representation we also determine the epikernels, i.e., the isotropy groups, the subgroups that satisfy the subduction and chain-subduction criteria.

In Sec. III, we apply the results of Sec. II, to determine the irreducible representation associated with each of the two phase transitions O_h^1 to C_{4v}^1 and D_{6h}^4 to D_{2h}^{16}. We also determine the epikernels associated with each of these phase transitions. In Sec. IV we show that these two phase transitions belong to the same exomorphic type. We then derive additional phase transitions, which also belong to this exomorphic type.

II. IRREDUCIBLE REPRESENTATIONS ASSOCIATED WITH A PHASE TRANSITION

We consider a phase transition between a higher-phase symmetry group G and a lower-phase symmetry F, where F is a subgroup of G of index n. Let $D^\alpha(G)$ denote the irreducible representation of G associated with this phase transition. Given the groups G and F we consider the inverse Landau problem, to determine the possible irreducible representations associated with the phase transition.

We apply the subduction criterion

$$(D^\alpha(G)\downarrow F\,|\,D^1(F))\neq 0. \quad (1)$$

That is, the subduced representation $D^\alpha(G)\downarrow F$, the irreducible representation $D^\alpha(G)$ restricted to the elements of the subgroup F, must contain the identity representation $D^1(F)$ of F a nonzero number of times. Using the Frobenius Reciprocity Theorem,[25] Eq. (1) can be replaced by

$$(D^1(F)\uparrow G\,|\,D^\alpha(G))\neq 0. \quad (2a)$$

The irreducible representation $D^\alpha(G)$ must be contained a nonzero number of times in the induced representation $D^1(F)\uparrow G$.

We shall use the symbol $D_A^B(A)$ to denote the induced representation $D^1(B)\uparrow A$. Equation (2a) can then be rewritten as

$$(D_G^F(G)\,|\,D^\alpha(G))\neq 0. \quad (2b)$$

We shall also use the symbol $D_G = D_{G/H}\uparrow\uparrow G$ to denote the representation D_G of G "engendered by the representation $D_{G/H}$ of its factor group G/H. Engendering[26] is defined as follows: Let H be a normal subgroup of G. The cosets g_iH of the coset decomposition of G with respect to H are elements of the factor group G/H. If $D_{G/H}$ is a representation of G/H then to every coset g_iH of the factor group G/H corresponds a matrix $D_{G/H}(g_iH)$. To define the engendered representation $D_G = D_{G/H}\uparrow\uparrow G$, we set all matrices $D_G(g_kh)$, for all h of H, equal to the matrix $D_{G/H}(g_kH)$.

It has been shown [27,28] that

$$D_G^F = D_{G/H}^{F/H}(G/H)\uparrow\uparrow G. \quad (3)$$

The induced representation $D_G^F(G)$ is engendered by the induced representation $D_{G/H}^{F/H}(G/H)$ of the factor group G/H, where

$$H = \text{Core}\,F = \bigcap_{g\in G} gFg^{-1}. \quad (4)$$

From Eqs. (2b) and (3), it follows that an irreducible representation $D^\alpha(G)$ associated with a phase transition between the group G and subgroup F of G is such that

$$D^\alpha(G) = D^\alpha(G/H)\uparrow\uparrow G \quad (5)$$

and

$$(D_{G/H}^{F/H}(G/H)\,|\,D^\alpha(G/H))\neq 0. \quad (6)$$

That is, the irreducible representation $D^\alpha(G)$ is engendered by an irreducible representation $D^\alpha(G/H)$ of the factor group G/H, and $D^\alpha(G/H)$ must be contained in the induced representation $D_{G/H}^{F/H}(G/H)$ a nonzero number of times. In addition, since the kernel of $D^\alpha(G)$ is equal to the subgroup H (see Refs. 19 and 27), i.e.,

$$\ker D^\alpha(G) = H = \text{Core}\,F, \quad (7)$$

the irreducible representation $D^\alpha(G/H)$, which engenders $D^\alpha(G)$, is a faithful representation of G/H.

A matrix $D_A^B(a)$ of an induced representation $D_A^B(A)$ is also the matrix representing the permutation of the cosets of B in A under multiplication of the cosets by the element a of A (see Refs. 28 and 29). The group of matrices is called a "permutation representation" and represents a group of permutations that is transitive on the cosets of B in A. The dimension of this permutation representation is equal to the number of cosets of B in A. Consequently, the representation $D_{G/H}^{F/H}(G/H)$ is a permutation representation of a transitive subgroup T_n, isomorphic to G/H, of the symmetric group S_n, where n is the index of F in G.

A method to determine all possible irreducible representations $D^\alpha(G)$ associated with a phase transition between a group G and subgroup F of index n in G is based on Eqs. (5)–(7). Such irreducible representations satisfy the subduction criterion and, of course, are further restricted by the use of the chain subduction criterion, Landau criterion, and Lifshitz criterion. We have that an irreducible representation $D^\alpha(G)$ is engendered by a faithful irreducible representation $D^\alpha(G/H)$, which is contained in the permutation representation of a transitive subgroup T_n, isomorphic to G/H, of the symmetric group S_n. A method to determine the irreducible representations $D^\alpha(G)$ is as follows.

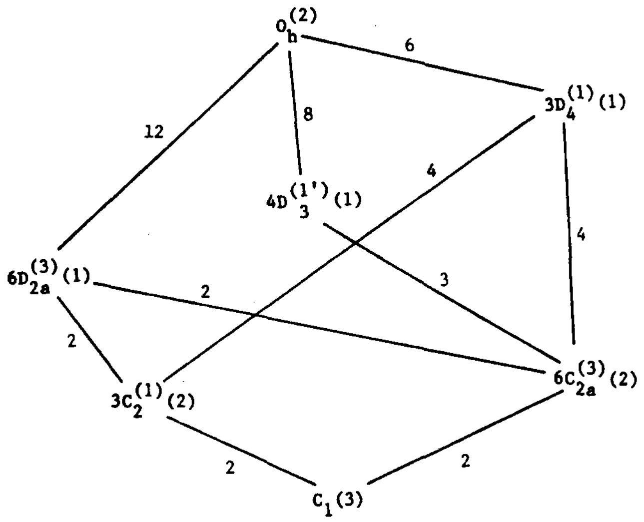
(1) Given the group G and subgroup F of index n, determine the subgroup H, Eq. (4), and the factor group G/H.

(2) Determine the transitive subgroup T_n, isomorphic to G/H, of the symmetric group S_n, and the faithful irreducible representations in the permutation representation of T_n.

(3) Each faithful irreducible representation of the permutation representation determines an irreducible represen-

TABLE I. Character table of the faithful irreducible representation contained in the permutation representation of the transitive subgroup 6/6 of S_6. Above each character is the number and cyclic notation of the elements in each class. The diagram shows the epikernels of the irreducible representation. The generators of each epikernel are listed below the diagram.

$6/6(48)C_2^2 \times O^2(O_h^{(2)})$										
1	3	8	6	6	1	3	8	6	6	
(1^6)	$(1^2,2^2)$	(3^2)	(2^3)	$(1^2,4)$	(2^3)	$(1^4,2)$	(6)	$(1^2,2^2)$	$(2,4)$	
3	−1	0	−1	1	−3	1	0	1	−1	(LA)



O_h^2:	(3456), (154236)
$3D_4^{(1)}$:	(3456), (36)(45); (1426), (16)(24); (1523), (15)(23).
$4D_3^{(1')}$:	(134)(256), (13)(25); (136)(254), (16)(24); (145)(263), (15)(23); (156)(234), (16)(24).
$6D_{2a}^{(3)}$:	(12), (36) (45); (12), (34)(56); (46), (15)(23); (46), (13)(25); (35), (16)(24); (35), (14)(26).
$6C_{2a}^{(3)}$:	(16)(24); (15)(23); (36)(45); (34)(56); (13)(25); (14)(26).
$3C_2^{(1)}$:	(12); (46); (35).

tation $D^{\alpha}(G/H)$, which in turn engenders, Eq. (5), a possible irreducible representation $D^{\alpha}(G)$ associated with the phase transition between G and subgroup F.

To implement this procedure requires the knowledge of all transitive subgroups T_n of the symmetric groups S_n, and all faithful irreducible representations contained in the permutation representation of each transitive subgroup. We have tabulated all transitive subgroups of the symmetric groups S_n for $n = 2, 3, 4, 5, 6$ and the faithful irreducible representations contained in the permutation representation of each transitive subgroup.[30] In Table I, we give an example from this tabulation. The table contains the following information.

(1) A symbol $n/m(p)$, where n is the degree of the symmetric group S_n, m is a serial number given to a transitive subgroup T_n, and p is the order of the transitive subgroup T_n. This is followed by a symbol or symbols, which denote the group T_n.

(2) The character table of the faithful irreducible representations contained in the permutation representation of T_n is given. The classes of elements are given in cycle length notation with the number of elements in each class given above the class symbol. The symbol "(LA)" is written to the right of the character table if the irreducible representation satisfies the Landau criterion.

(3) Using the lattices of the symmetric groups,[31] we have derived and tabulated the epikernels[24] for each faithful irreducible representation of the transitive subgroup T_n. The subgroup index of the epikernel is given along the line connected each pair of groups and the subduction frequency is given in parenthesis following the subgroup symbol. If there is more than one subgroup of a specific class, the number of such subgroups is given preceding the subgroup symbol.

(4) The generators of at least one epikernel of each class of epikernels is given. When the number of epikernels is not large, as in Table I, the generators of all epikernels in each class are given.

III. EXAMPLES

We shall consider two phase transitions: (1) the equitranslational transition from O_h^1 to C_{4v}^1 in $BaTiO_3$ and (2) the nonequitranslational transition from D^4_{6h} to D^{16}_{2h} in β-K_2SO_4. We shall determine the irreducible representations associated with these phase transitions and show that the respective irreducible representations are both engendered by the same faithful irreducible representation.

We first consider the phase transition from $G = O^1_h$ to $F = C^1_{4v}$, the equitranslational subgroup of O^1_h with the point group $C_{4v} = \{E, C_{4z}, C_{2z}, C^{-1}_{4z}, m_x, m_y, m_{xy}, m_{xy}\}$; C^1_{4v} is a subgroup of index $n = 6$ in O^1_h. The core of $F = C^1_{4v}$, see Eq. (4), is

$$H = \text{Core } C^1_{4v} = C^1_1,$$

where C^1_1 is the translational subgroup of O^1_h. It follows that $G/H = O^1_h/C^1_1$ and is isomorphic to the point group O_h of order 48. Then $D^{F/H}_{G/H}$ is a permutation representation of a transitive subgroup of order 48 of S_6. There is only one such transitive subgroup of S_6, the group denoted by 6/6(48) given in Table I. This permutation representation contains a single, Landau active, faithful irreducible representation whose character table is given in Table I. This character

TABLE II. Character table of the faithful irreducible representation contained in the permutation representation of the transitive subgroup 6/6 of S^6. In the first and second column are the number and cyclic notation of the elements of each class whose character is given in the third column. In the fourth column, we list in cyclic notation all elements of the transitive subgroup belonging to each class. Below each element we list the cosets of the factor groups O^1_h/C^1_1 and D^4_{6h}/C^2_2 isomorphic to this transitive subgroup of S_6.

1	(1^6)	3	(1)(2)(3)(4)(5)(6) $(E\vert 000)$ $\{(E\vert 000),(C_{2z}\vert 00\frac{1}{2})\}$			
3	$(1^1,2^2)$	-1	(35)(46) $(C_{2x}\vert 000)$ $\{(E\vert 010),(C_{2z}\vert 01\frac{1}{2})\}$	(12)(46) $(C_{2y}\vert 000)$ $\{(E\vert 110),(C_{2z}\vert 11\frac{1}{2})\}$	(12)(35) $(C_{2z}\vert 000)$ $\{(E\vert 100),(C_{2z}\vert 10\frac{1}{2})\}$	
8	(3^2)	0	(145)(263) $(C_{3xyz}\vert 000)$ $\{(C_3\vert 010),(C_6^{-1}\vert 01\frac{1}{2})\}$	(136)(254) $(C_{3\bar{x}yz}\vert 000)$ $\{(C_3^{-1}\vert 010),(C_6\vert 01\frac{1}{2})\}$	(134)(256) $(C_{3x\bar{y}z}\vert 000)$ $\{(C_3^{-1}\vert 000),(C_6\vert 00\frac{1}{2})\}$	(156)(234) $(C_{3xy\bar{z}}\vert 000)$ $\{(C_3^{-1}\vert 110),(C_6\vert 11\frac{1}{2})\}$
			(154)(236) $(C^{-1}_{3xyz}\vert 000)$ $\{(C_3^{-1}\vert 100),(C_6\vert 10\frac{1}{2})\}$	(163)(245) $(C^{-1}_{3\bar{x}yz}\vert 000)$ $\{(C_3\vert 110),(C_6^{-1}\vert 11\frac{1}{2})\}$	(143)(265) $(C^{-1}_{3x\bar{y}z}\vert 000)$ $\{(C_3\vert 000),(C_6^{-1}\vert 00\frac{1}{2})\}$	(165)(243) $(C^{-1}_{3xy\bar{z}}\vert 000)$ $\{(C_3\vert 100),(C_6^{-1}\vert 10\frac{1}{2})\}$
6	(2^3)	-1	(15)(23)(46) $(C_{2xy})\vert 000)$ $\{(C_{2x}\vert 000),(C_{22}\vert 00\frac{1}{2})\}$	(14)(26)(35) $(C_{2xz}\vert 000)$ $\{(C_{2xy}\vert 110),(C_{23}\vert 11\frac{1}{2})\}$	(12)(36)(45) $(C_{2yz}\vert 000)$ $\{(C_{2y}\vert 000),(C_{21}\vert 00\frac{1}{2})\}$	
			(13)(25)(46) $(C_{2\bar{x}y}\vert 000)$ $\{(C_{2x}\vert 100),(C_{22}\vert 10\frac{1}{2})\}$	(16)(24)(35) $(C_{2\bar{x}z}\vert 000)$ $\{(C_{2xy}\vert 000)(C_{23}\vert 00\frac{1}{2})\}$	(12)(34)(56) $(C_{2\bar{y}z}\vert 000)$ $\{(C_{2y}\vert 010),(C_{21}\vert 01\frac{1}{2})\}$	
6	$(1^2,4)$	1	(3456) $(C_{4x}\vert 000)$ $\{(C_{2y}\vert 100),(C_{21}\vert 10\frac{1}{2})\}$	(1426) $(C_{4y}\vert 000)$ $\{(C_{2xy}\vert 100),(C_{23}\vert 10\frac{1}{2})\}$	(1325) $(C_{4z}\vert 000)$ $\{(C_{2x}\vert 110),(C_{22}\vert 11\frac{1}{2})\}$	
			(3654) $(C_{4x}^{-1}\vert 000)$ $\{(C_{2y}\vert 110),(C_{21}\vert 11\frac{1}{2})\}$	(1624) $(C_{4y}^{-1}\vert 000)$ $\{(C_{2xy}\vert 010),(C_{23}\vert 01\frac{1}{2})\}$	(1523) $(C_{4z}^{-1}\vert 000)$ $\{(C_{2x}\vert 010),(C_{22}\vert 01\frac{1}{2})\}$	
1	(2^3)	-3	(12)(35)(46) $(\bar{1}\vert 000)$ $\{(\bar{1}\vert 000),(m_z\vert 00\frac{1}{2})\}$			
3	$(1^4,2)$	1	(12) $(m_x\vert 000)$ $\{(\bar{1}\vert 010),(m_z\vert 01\frac{1}{2})\}$	(35) $(m_y\vert 000)$ $\{(\bar{1}\vert 110),(m_z\vert 11\frac{1}{2})\}$	(46) $(m_z\vert 000)$ $\{(\bar{1}\vert 100),(m_z\vert 10\frac{1}{2})\}$	
8	(6)	0	(134256) $(S_{6xyz}\vert 000)$ $\{(S_3^{-1}\vert 10\frac{1}{2}),(S_6\vert 100)\}$	(143265) $(S_{6\bar{x}yz}\vert 000)$ $\{(S_3\vert 11\frac{1}{2}),(S_6^{-1}\vert 110)\}$	(163245) $(S_{6x\bar{y}z}\vert 000)$ $\{(S_3\vert 00\frac{1}{2}),(S_6^{-1}\vert 000)\}$	(145263) $(S_{6xy\bar{z}}\vert 000)$ $\{(S_3\vert 10\frac{1}{2}),(S_6^{-1}\vert 100)\}$
			(165243) $(S^{-1}_{6xyz}\vert 000)$ $\{(S_3\vert 01\frac{1}{2}),(S_6^{-1}\vert 010)\}$	(156234) $(S^{-1}_{6\bar{x}yz}\vert 000)$ $\{(S_3^{-1}\vert 01\frac{1}{2}),(S_6\vert 010)\}$	(154236) $(S^{-1}_{6x\bar{y}z}\vert 000)$ $\{(S_3^{-1}\vert 00\frac{1}{2}),(S_6\vert 000)\}$	(136254) $(S^{-1}_{6xy\bar{z}}\vert 000)$ $\{(S_3^{-1}\vert 11\frac{1}{2}),(S_6\vert 110)\}$
6	$(1^2,2^2)$	1	(13)(25) $(m_{xy}\vert 000)$ $\{(m_2\vert 00\frac{1}{2}),(m_x\vert 000)\}$	(16)(24) $(m_{xz}\vert 000)$ $\{(m_3\vert 11\frac{1}{2}),(m_{xy}\vert 110)\}$	(34)(56) $(m_{yz}\vert 000)$ $\{(m_1\vert 00\frac{1}{2}),(m_y\vert 000)\}$	
			(15)(23) $(m_{\bar{x}y}\vert 000)$ $\{(m_2\vert 10\frac{1}{2}),(m_x\vert 100)\}$	(14)(26) $(m_{\bar{x}z}\vert 000)$ $\{(m_3\vert 00\frac{1}{2}),(m_{xy}\vert 000)\}$	(36)(45) $(m_{\bar{y}z}\vert 000)$ $\{(m_1\vert 01\frac{1}{2}),(m_y\vert 010)\}$	
6	(2,4)	-1	(12)(3456) $(S_{4x}\vert 000)$ $\{(m_1\vert 11\frac{1}{2}),(m_y\vert 110)\}$	(1426)(35) $(S_{4y}\vert 000)$ $\{(m_3\vert 01\frac{1}{2}),(m_{xy}\vert 010)\}$	(1325)(46) $(S_{4z}\vert 000)$ $\{(m_2\vert 01\frac{1}{2}),(m_x\vert 010)\}$	
			(12)(3654) $(S_{4x}^{-1}\vert 000)$ $\{(m_1\vert 10\frac{1}{2}),(m_y\vert 100)\}$	(1624)(35) $(S_{4y}^{-1}\vert 000)$ $\{(m_3\vert 10\frac{1}{2}),(m_{xy}\vert 100)\}$	(1523)(46) $(S_{4z}^{-1}\vert 000)$ $\{(m_2\vert 11\frac{1}{2}),(m_x\vert 110)\}$	

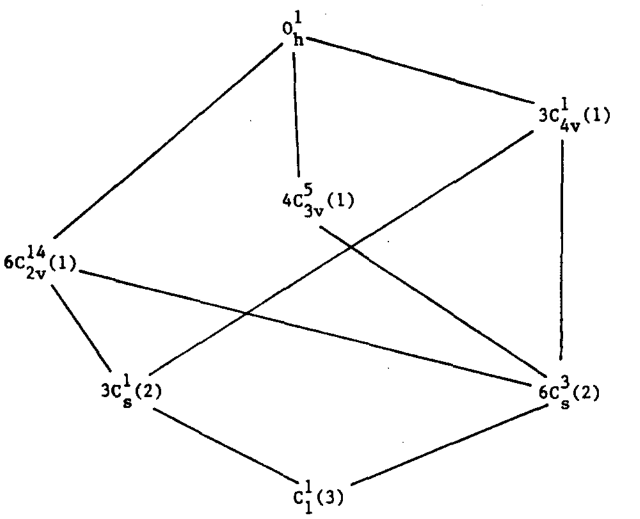


FIG. 1. Epikernels of the irreducible representation $D^{k=(0,0,0),4-}(O_h^1)$.

table is given in detail in Table II. In the first three columns we duplicate the first three rows of the character table given in 6/6(48) of Table I. To the right of each character we list explicitly in cyclic notation the elements of each class of this transitive subgroup of S_6.

The factor group $G/H = O_h^1/C_1^1$ is isomorphic to this transitive subgroup of S_6 denoted by 6/6(48). The isomorphism is between elements P_i of 6/6(48) and cosets $(R_i|000)C_1^1$ of G/H. In Table II we have denoted the coset $(R_i|000)C_1^1$ isomorphic to P_i by listing below the element P_i the coset representative $(R_i|000)$. This isomorphism and the faithful irreducible representation of the transitive subgroup 6/6(48) of S_6 determines the irreducible representation $D^\alpha(O_h^1/C_1^1)$, see Eq. (6), which in turn engenders, Eq. (5), the irreducible representation $D^\alpha(O_h^1)$ associated with the phase transition between O_h^1 and C_{4v}^1. This irreducible representation $D^\alpha(O_h^1)$ is denoted by $D^{(k=0,0,0),4-}(O_h^1)$ in the notation of Cracknell *et al.*[32]

Using the epikernels and generators of the epikernels given in Table I along with the isomorphism between the elements of 6/6(48) and cosets of O_h^1/C_1^1 given in Table II, we can derive the subgroups of O_h^1, which satisfy the chain-subduction criterion for phase transitions from O_h^1 associated with the irreducible representation $D^{(k=0,0,0),4-}(O_h^1)$. These epikernels are given in Fig. 1.

The second example is the phase transition from hexagonal $G' = D_{6h}^4$ to orthorhombic $F' = D_{2h}^{16}$. The subgroup D_{2h}^{16} has the translation subgroup generated by the hexagonal translations $(E|1,0,0)$, $(E|1,2,0)$, and $(E|0,0,1)$. The elements of D_{6h}^4, which are the coset representations of D_{2h}^{16} with respect to its translational subgroup, are

$$(E|0,0,0),\quad (\bar{1}|1,1,0),$$
$$(C_{2z}|0,0,\tfrac{1}{2}),\quad (m_z|1,1,\tfrac{1}{2}),$$
$$(C_{2x}|1,1,0),\quad (m_x|0,0,0),$$
$$(C_{22}|1,1,\tfrac{1}{2}),\quad (m_2|0,0,\tfrac{1}{2}).$$

Now D_{2h}^{16} is a subgroup of index $n = 6$ of D_{6h}^4. The core of $F' = D_{2h}^{16}$, see Eq. (4), is

$$H' = \text{Core}\, D_{2h}^{16} = C_2^2.$$

The group C_2^2 has the translational subgroup generated by the hexagonal translations $(E|2,0,0)$, $(E|0,2,0)$, and $(E|0,0,1)$. The elements of D_{6h}^4, which are the coset representatives of C_2^2 with respect to its translational subgroup, are $(E|0,0,0)$ and $(C_{2z}|0,0,\tfrac{1}{2})$. The factor group $G'/H' = D_{6h}^4/C_2^2$ is isomorphic to the point group O_h of order 48. It follows that $D_{G'/H'}^{F'/H'}$ is then a permutation representation of a transitive subgroup of order 48 of S_6. This is the same transitive group, 6/6(48) given in Table I, as that which arose in the first example given above.

The isomorphism between the elements P_i of 6/6(48) and the cosets $(R_i|\tau_i)C_2^2$ of G'/H' is given in Table II. Two lines below each element P_i of 6/6(48) given in Table II we have denoted the isomorphic coset $(R_i|\tau_i)C_2^2$ of $G'/H' = D_{6h}^4/C_2^2$. Since

$$(R_i|\tau_i)C_2^2 = (R_i|\tau_i)C_1^1 + (R_i|\tau_i)(C_{2z}|0,0,\tfrac{1}{2})C_1^1,$$

where C_1^1 is the translational subgroup of C_2^2, we list the two elements $(R_i|\tau_i)$ and $(R_i|\tau_i)(C_{2z}|0,0,\tfrac{1}{2})$. This isomorphism and the faithful irreducible representation of the transitive subgroup 6/6(48) of S_6 determines the irreducible representation $D^\alpha(D_{6h}^4/C_2^2)$, Eq. (6), which in turn engenders, Eq. (5), the irreducible representation $D^\alpha(D_{6h}^4)$ associated with the phase transition between D_{6h}^4 and D_{2h}^{16}. This irreducible representation $D^\alpha(D_{6h}^4)$ is denoted by $D^{k=(\frac{1}{2},0,0),2-}(D_{6h}^4)$ in the notation of Cracknell *et al.*[32]

Using the epikernels and generators of the epikernels given in Table I along with the isomorphism between elements of 6/6(48) and cosets of D_{6h}^4/C_2^2 given in Table II, we can derive the subgroups that satisfy the chain-subduction criterion for phase transitions from D_{6h}^4 associated with the irreducible representation $D^{k=(\frac{1}{2},0,0),2-}(D_{6h}^4)$. These epikernels are given in Fig. 2.

The above two examples are at first glance quite different, one being an equitranslational phase transition while the second is nonequitranslational. However, as we have seen, these two transitions are mathematically similar; the associated irreducible representations are engendered by the

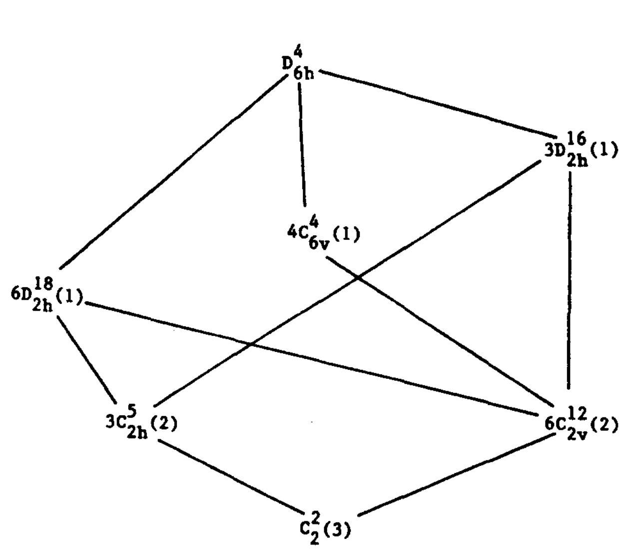


FIG. 2. Epikernels of the irreducible representation $D^{k=(\frac{1}{2},0,0),2-}(D_{6h}^4)$.

TABLE III. Phase transitions O_h^i to C_{4v}^j and O_h^i to D_{2d}^k that belong to the exomorphic type of phase transition characterized by the permutation representation of the transitive subgroup 6/6 of S_6. Here, $H = \text{Core}\, F = C_1^1$ for all cases.

$G = O_h^i$ i	$F = C_{4v}^j$ j	$F = D_{2d}^k$ k
1	1	5
2	6	8
3	7	5
4	4	8
5	9	11
6	10	11
7	11	12
8	12	12
9	9	9
10	12	10

same faithful irreducible representation. This mathematical similarity of different phase transitions has been codified by the concept of exomorphic types of phase transitions.[21–24]

IV. EXOMORPHIC TYPES OF PHASE TRANSITIONS

Two phase transitions, between a higher-phase symmetry group G and lower-phase symmetry F and between a higher-phase symmetry G' and lower-phase symmetry F', are said to be of the same exomorphic types if and only if[21] (1) the factor groups G/H, where $H = \text{Core}\, F$, and G'/H', where $H' = \text{Core}\, F'$, are isomorphic; and (2) there exists an isomorphism that maps the factor group F/H into F'/H'.

Alternatively,[24] we can state that two phase transitions are of the same exomorphic type if and only if a suitable labeling of the cosets g_iF and $g_i'F'$ in the coset decompositions G with respect to F, and G' with respect to F' exists such that the permutation representations $D_{G/H}^{F/H}(G/H)$ and $D_{G'/H'}^{F'/H'}(G'/H')$ are identical groups of permutations.

In the examples of Sec. III, both the transitions $G = O_h^1$ to $F = C_{4v}^1$ and $G' = D_{6h}^4$ to $F' = D_{2h}^{16}$ are of the same exomorphic type. The factor groups $G/H = O_h^1/C_1^1$ and $G'/H' = D_{6h}^4/C_2^2$ are isomorphic with the isomorphism given in Table II, where we find that $F/H = C_{4v}^1/C_1^1$ is isomorphic to $F'/H' = D_{2h}^{16}/C_2^2$. The permutation representations $D_{G/H}^{F/H}(G/H)$ and $D_{G'/H'}^{F'/H'}(G'/H')$ are identical groups of permutations isomorphic to the transitive subgroup 6/6(48) of S_6.

It follows from the above and Eqs. (1)–(6) that if the phase transitions from G to F and G' to F' are of the same exomorphic type, then the irreducible representations $D^\alpha(G)$ and $D^\alpha(G')$, which can be associated with the respective phase transitions, are each engendered by faithful irreducible representations contained in a single permutation representation. This is the permutation representation denoted by $D_{G/H}^{F/H}(G/H)$ and $D_{G'/H'}^{F'/H'}(G'/H')$, and is a permutation representation of a transitive subgroup, isomorphic to G/H and G'/H', of the symmetric group S_n.

If the permutation representation contains a single faithful irreducible representation then this faithful irreducible representation engenders the irreducible representations associated with all phase transitions belonging to the exo-

TABLE IV. Phase transitions D_{6h}^i to D_{2h}^j that belong to the exomorphic type of phase transition characterized by the permutation representation of the transitive subgroup 6/6 of S_6. Here Ch. 1 and Ch. 2 refer to the alternative choice of origins as given in the *International Tables for Crystallography*.[33] The shift in origin, with respect to the translational subgroup of D_{6h}^i is also given. Here, $H = \text{Core}\, F$ is given to the right on the same row as F.

G			F		H
D_{6h}^1	$D_{2h}^{13}\left(P\frac{2_1}{m}\frac{2_1}{m}\frac{2}{n}\right)$	Ch.1	$D_{2h}^{4}\left(P\frac{2}{b}\frac{2}{a}\frac{2}{n}\right)$	$(\frac{1}{2},\frac{1}{2},0)$Ch.1	C_2^1
	$D_{2h}^{7}\left(P\frac{2_1}{b}\frac{2}{m}\frac{2}{n}\right)$		$D_{2h}^{7}\left(P\frac{2}{m}\frac{2_1}{a}\frac{2}{n}\right)$		C_i^1
	$D_{2h}^{5}\left(P\frac{2_1}{c}\frac{2}{a}\frac{2}{m}\right)$	$(\frac{1}{2},\frac{1}{2},0)$	$D_{2h}^{5}\left(P\frac{2}{b}\frac{2_1}{m}\frac{2}{m}\right)$		C_s^1
D_{6h}^2	$D_{2h}^{10}\left(P\frac{2_1}{c}\frac{2_1}{c}\frac{2}{n}\right)$	$(\frac{1}{2},\frac{1}{2},0)$	$D_{2h}^{2}\left(P\frac{2}{n}\frac{2}{n}\frac{2}{n}\right)$	(0,0,1) Ch.1	C_2^1
	$D_{2h}^{6}\left(P\frac{2_1}{n}\frac{2}{c}\frac{2}{n}\right)$		$D_{2h}^{6}\left(P\frac{2}{c}\frac{2_1}{n}\frac{2}{n}\right)$		C_i^1
	$D_{2h}^{7}\left(P\frac{2_1}{c}\frac{2}{n}\frac{2}{m}\right)$	$(\frac{1}{2},\frac{1}{2},0)$	$D_{2h}^{7}\left(P\frac{2}{n}\frac{2_1}{c}\frac{2}{m}\right)$	$(\frac{1}{2},\frac{1}{2},0)$	C_s^1
D_{6h}^3	$D_{2h}^{16}\left(P\frac{2_1}{m}\frac{2_1}{n}\frac{2_1}{b}\right)$	$(\frac{1}{2},\frac{1}{2},0)$	$D_{2h}^{6}\left(P\frac{2}{b}\frac{2}{n}\frac{2_1}{n}\right)$	$(\frac{1}{2},\frac{1}{2},0)$	C_2^2
	$D_{2h}^{14}\left(P\frac{2_1}{b}\frac{2}{c}\frac{2_1}{n}\right)$		$D_{2h}^{12}\left(P\frac{2}{m}\frac{2_1}{n}\frac{2_1}{n}\right)$		C_i^1
	$D_{2h}^{13}\left(P\frac{2_1}{m}\frac{2}{n}\frac{2_1}{m}\right)$	$(\frac{1}{2},\frac{1}{2},0)$Ch. 2	$D_{2h}^{11}\left(P\frac{2}{b}\frac{2_1}{c}\frac{2_1}{m}\right)$		C_s^1
D_{6h}^4	$D_{2h}^{16}\left(P\frac{2_1}{c}\frac{2_1}{m}\frac{2_1}{n}\right)$	$(\frac{1}{2},\frac{1}{2},0)$	$D_{2h}^{6}\left(P\frac{2}{n}\frac{2}{a}\frac{2_1}{n}\right)$	$(\frac{1}{2},\frac{1}{2},0)$	C_2^2
	$D_{2h}^{12}\left(P\frac{2_1}{n}\frac{2}{m}\frac{2_1}{n}\right)$		$D_{2h}^{14}\left(P\frac{2}{c}\frac{2_1}{a}\frac{2_1}{n}\right)$		C_i^1
	$D_{2h}^{11}\left(P\frac{2_1}{c}\frac{2}{a}\frac{2_1}{m}\right)$		$D_{2h}^{13}\left(P\frac{2}{n}\frac{2_1}{m}\frac{2_1}{m}\right)$	$(0,\frac{1}{2},0)$ Ch. 2	C_s^1

morphic type. In the examples of the previous sections the irreducible representations $D^{k=(0,0,0),4-}(O_h^1)$ and $D^{k=(\frac{1}{2},0,0),2-}(D_{6h}^4)$ are associated with the phase transitions from $G=O_h^1$ to $F=C_{4v}^1$ and $G'=D_{6h}^4$ to $F'=D_{2h}^{16}$, respectively. These two phase transitions belong to the same exomorphic type, and both irreducible representations are engendered by the same faithful irreducible representation, denoted by $D^\alpha(O_h^1/C_1^1)$ and $D^\alpha(D_{6h}^4/C_2^2)$, the only faithful irreducible representation contained in the permutation representation of the transitive subgroup 6/6(48) of S_6.

The two phase transitions $G=O_h^1$ to $F=C_{4v}^1$ and $G'=D_{6h}^4$ to $F'=D_{2h}^{16}$ belong to the same exomorphic type whose permutation representation is the permutation representation of the transitive subgroup 6/6(48) of S_6. Additional equitranslational phase transitions belonging to this exomorphic type with $G=O_h^i$ and $F=C_{4v}^j$ and $F=D_{2d}^k$ as given in Table III. In Table IV we give the phase transitions between $G=D_{6h}^i$ and $F=D_{2h}^j$ that belong to this exomorphic type.

ACKNOWLEDGMENTS

One of us (DBL) gratefully acknowledges the support of the United States National Academy of Sciences–Czechoslovakian Academy of Science Exchange Program, and the financial support provided by The Berks Campus of The Pennsylvania State University.

We also thank Dr. V. Janovec of the Czechoslovak Academy of Science for discussions and for suggesting the examples used in this paper.

[1]L. D. Landau, Phys. Z. Sowjetunion **11**, 26 (1937); Zh. Eksp. Teor. Fiz. 7, 19 (1937). Reprinted in English, in *Collected Papers of L. D. Landau*, edited by D. Ter Haar (Pergamon, New York, 1965); L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Pergamon, New York, 1958), Chap. XIV.
[2]G. Ya. Lyubarskii, *Applications of Group Theory in Physics* (Pergamon, New York, 1960).
[3]J. L. Birman, in *Lecture Notes in Physics No. 79—Group Theory Methods in Physics*, edited by P. Kramer and A. Rieckers (Springer, New York, 1978), p. 203.
[4]P. Toledano and J. C. Toledano, Phys. Rev. B **14**, 3097 (1976); **16**, 386 (1977); **21**, 1139 (1980); **25**, 1946 (1982).
[5]J. L. Birman, Phys. Rev. Lett. **17**, 1216 (1966).
[6]F. E. Goldrich and J. L. Birman, Phys. Rev. **167**, 528 (1968).
[7]M. V. Jaric and J. L. Birman, Phys. Rev. B **16**, 2564 (1977).
[8]M. V. Jaric, Phys. Rev. B **23**, 3460 (1981).
[9]M. V. Jaric, Phys. Rev. B **25**, 2015 (1982).
[10]E. B. Vinborg, Y. M. Gufan, V. P. Sakhnenko, and Y. I. Sirotin, Sov. Phys. Crystallogr. **19**, 10 (1974).
[11]M. Sutton and R. L. Armstrong, Phys. Rev. B **25**, 1813 (1982); **27**, 1937 (1983).
[12]M. H. Ben Ghozlen and Y. Mlik, J. Phys. C **16**, 4365 (1983).
[13]D. M. Hatch and H. T. Stokes, in *Proceedings of the Thirteenth International Colloquium on Group Theoretical Methods in Physics*, edited by W. W. Zachary (World Scientific, Singapore, 1985).
[14]P. W. Anderson and E. I. Blount, Phys. Rev. Lett. **14**, 217 (1966).
[15]N. Boccara, Ann. Phys. NY **47**, 40 (1968).
[16]L. V. Meisel, D. M. Gray, and E. Brown, J. Math. Phys. **16**, 2520 (1975).
[17]E. Brown and L. V. Meisel, Phys. Rev. B **13**, 213 (1976).
[18]D. B. Litvin and J. L. Birman, Phys. Rev. B **26**, 5248 (1982).
[19]E. Ascher and J. Kobayashi, J. Phys. C **10**, 1349 (1977).
[20]M. Gufan and V. P. Sakhnenko, Zh. Eksp. Teor. Fiz. **63**, 1909 (1972).
[21]V. Kopsky, Phys. Lett. A **69**, 82 (1978).
[22]V. Kopsky, Ferroelectrics **24**, 3 (1980).
[23]V. Kopsky, Commun. Math. Chem. **11**, 35 (1981).
[24]V. Kopsky, Czech. J. Phys. B **33**, 485, 720, 845 (1983).
[25]L. Jansen and M. Boon, *Theory of Finite Groups; Applications in Physics* (North-Holland, Amsterdam, 1967), p. 139.
[26]Reference 25, p. 152.
[27]D. B. Litvin, J. N. Kotzev, and J. L. Birman, Phys. Rev. B **26**, 6947 (1982).
[28]E. Ascher, J. Phys. C **10**, 1365 (1977).
[29]W. Opechowski and T. Dreyfus, Acta Crystallogr. A **27**, 470 (1971).
[30]See AIP document JMAPA-27-0661-31 for 31 pages of transitive subgroups of S_n. Order by PAPS number and journal reference from American Institute of Physics, Physics Auxiliary Publication Service, 335 E. 45 St., New York, NY 10017. The price is $1.50 for microfiche or $5.00 for photocopies, airmail additional. Make check payable to American Institute of Physics. This material appears in the monthly Current Physics Microform edition of the Journal of Mathematical Physics on the frames following this article.
[31]J. Fuksa and V. Kopsky, in *"Proceedings of the 12th International Colloquium on Group Theoretical Methods in Physics,"* edited by G. Denardo, G. Ghirardi, and T. Weber (Springer, New York, 1984).
[32]A. P. Cracknell, B. L. Davies, S. C. Miller, and W. F. Love, *Kronecker Product Tables* (IFI/Plenum, New York, 1979), Vol. 1.
[33]*International Tables for Crystallography*, edited by T. Han (Reidel, Dordrecht, 1983), Vol. A.

Verma bases for representations of classical simple Lie algebras

Sai-Ping Li
Department of Physics, University of California—Irvine, Irvine, California 92717
R. V. Moody
Department of Mathematics, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, 57N OWO, and Concordia University, Montréal, Québec, Canada H3G IM8
M. Nicolescu, and J. Patera
Centre de recherches mathématiques, Université de Montréal, Montréal, Québec, Canada H3C 3J7

(Received 30 May 1985; accepted for publication 23 August 1985)

Complete bases are constructed for all finite-dimensional irreducible representations of the simple Lie algebras over C of the types A_n $(n>1)$, B_n and C_n $(2<n<6)$, D_n $(4<n<6)$, and G_2. Each basis vector is given as an explicit sequence of weight-lowering generators of the algebra acting on the highest weight vector of the representation space. A similar construction (due to D-N. Verma) for the highest weight representations of all Kac–Moody algebras of rank 2 is presented as well.

I. INTRODUCTION

The present article contains an exploitation of a new method due to D-N. Verma for construction of bases in irreducible representation spaces of simple Lie algebras. As far as we know no final account of the method has been prepared, not to mention published. In the first place, a proof for the construction requires the Demazure character formula that has only recently been proved to be correct for the semisimple Lie algebras and is still only conjectured in the Kac–Moody situation. In the second place, we have found that there is a further unsolved problem regarding a right choice of expressing the opposite involution of the Weyl group as a product of reflections. The problem occurs only for rank >3. The choice can be made in many different ways, some of which are fatal for the algorithm. Unfortunately there seems to be no proof at present that a successful choice can always be made. However, the numerous examples below attest that a suitable choice is often possible.

In this paper the method is applied to a series of particular cases, namely the simple Lie algebras of rank <6 and of all types except F_4 and E_6. For algebras A_n the results are given for all ranks. It is also clear that an extension of our results to algebras of types B_n C_n and D_n and ranks >6 is straightforward. Once the difficulties are avoided, the result of the method is a set of basis-defining inequalities which, in our opinion, is a striking development for the theory of semisimple Lie algebras and their finite-dimensional representations. Let us point out the following features of the inequalities.

(1) The inequalities define a set of linearly independent vectors that span the whole representation space.

(2) A given set of the inequalities refers to a Lie algebra of a specific type and applies to any irreducible finite-dimensional representation of the Lie algebra.

(3) The number of inequalities for a given algebra is equal to the number of positive roots of the algebra. For many particular representations, the number of inequalities required is often much smaller.

(4) Bases provided in this way are of a particularly convenient kind for applications: They consist of eigenvectors of the Cartan subalgebra and thus each basis vector is labeled by "additive quantum numbers" that are the components of a weight of the representation.

(5) The bases are not related to any fixed subalgebra(s) in general. This allows a relatively versatile further adaptation to a particular subalgebra of importance at any time.

(6) Matrix elements of suitably chosen generators of the algebra relative to a Verma basis can be easily calculated, but no general closed formulas can be given.

Once the inequalities were derived for a given algebra, they were checked by counting the number of basis vectors and comparing it with the dimension of the representation. In addition the dimensions of dominant weight subspaces were compared with an independent computation of the dominant weight multiplicities.[1]

Section II contains the two simplest examples: the Lie algebras A_1 and A_2. The first of them is elementary. However, the second one is quite nontrivial in spite of its simplicity. In Sec. III our results are presented. An account of the derivation is contained in Sec. IV together with some examples and an illustration of the difficulties of the procedure. In Sec. V we sketch some of the theory behind the construction. The purpose of this section is to bring to the readers' attention the beautiful ideas of D-N. Verma and to provide some insight into what otherwise seems like a magical prescription.The last section contains conclusions, comments, and the basis-defining inequalities for all rank 2 Lie and Kac–Moody algebras in a uniform form as found by Verma.[2]

The principle involved in the construction is heavily dependent on the use of Schubert submodules and the theory of SL_2-induced modules. Although we have had no luck in the cases of F_4 and E_6, there are many choices of opposite involution and it is far from clear that the method will not work for all simple Lie algebras.

 0022-2488/86/030668-10$02.50

II. TWO EXAMPLES

The simplest case is the Verma basis for the Lie algebra of rank 1. We choose the commutation relations as

$$[e,f] = h, \quad [h,e] = 2e, \quad [h,f] = -2f. \tag{1}$$

An irreducible representation is denoted by its highest weight $\Lambda = (L)$, where $L = 0,1,2,\ldots$ is an integer equal to twice the "angular momentum." In this case it is well known that the irreducible representation space $V(\Lambda)$ is of dimension $L+1$ and that it is spanned by the vectors

$$|LM\rangle = f^a|LL\rangle, \tag{2}$$

where

$$0 \leqslant a \leqslant L \tag{3}$$

and

$$M = L - 2a, \quad e|LL\rangle = 0. \tag{4}$$

The vectors $|LM\rangle$ are pairwise orthogonal. They can be normalized using the relations

$$e|LM\rangle = \tfrac{1}{2}\sqrt{(L-M)(L+M-2)}\,|LM+2\rangle, \tag{5}$$

$$f|LM\rangle = \tfrac{1}{2}\sqrt{(L+M)(L-M+2)}\,|LM-2\rangle, \tag{6}$$

$$h\,|LM\rangle = M\,|LM\rangle. \tag{7}$$

In this simple case Eq. (3) is the basis-defining inequality of Verma.

Our second example is the Lie algebra A_2. This time one has $e_i f_i, h_i, i = 1$ and 2, satisfying (1) for each simple root α_i. We make no use of the remaining generators of A_2. The Verma basis of an irreducible representation space $V(\Lambda)$, where $\Lambda = (m_1,m_2)$ is the highest weight, consists of all vectors

$$f_1^{a_3} f_2^{a_2} f_1^{a_1}|m_1,m_2\rangle, \tag{8}$$

such that

$$\begin{aligned} &0 \leqslant a_1 \leqslant m_1, \\ &0 \leqslant a_2 \leqslant m_2 + a_1, \\ &0 \leqslant a_3 \leqslant \min[m_2,a_2]. \end{aligned} \tag{9}$$

Let us consider some particular representations. Thus for the representation $\Lambda = (10)$, which is of dimension 3, the inequalities (9) allow exactly three sets of values for the exponents a_1, a_2, and a_3 in (8):

$$\Lambda = (1,0): \quad |1,0\rangle, \quad f_1|1,0\rangle, \quad f_2 f_1|1,0\rangle. \tag{10}$$

Similarly one finds the bases for the representations (0 1), (2 0), and (1 1). Namely,

$$\Lambda = (0,1): \quad |0,1\rangle, \quad f_1|0,1\rangle, \quad f_2 f_1|0,1\rangle; \tag{11}$$

$$\begin{aligned} \Lambda = (2,0): \quad &|2,0\rangle, \quad f_1|2,0\rangle, \quad f_1^2|2,0\rangle, \\ &f_1^2|2,0\rangle, \quad f_2 f_1^2|2,0\rangle, \\ & \quad f_2^2 f_1^2|2,0\rangle; \end{aligned} \tag{12}$$

$$\begin{array}{lccccc} \Lambda = (1,1): & & & |1,1\rangle & & \\ & & & (000) & & \\ & f_1|1,1\rangle & & & & f_2|1,1\rangle \\ & (001) & & & & (010) \\ & & f_2 f_1|1,1\rangle & & f_1 f_2|1,1\rangle & \\ & & (011) & & (110) & \\ & f_2^2 f_1|1,1\rangle & & & & f_1 f_2 f_1|1,1\rangle \\ & (021) & & & & (111) \\ & & & f_1 f_2^2 f_1|1,1\rangle & & \\ & & & (121) & & \end{array} \tag{13}$$

In the last example the basis vectors are arranged into levels of successive applications of f_i's, and under each one we show the values of a_1, a_2, and a_3 as $(a_1 a_2 a_3)$. Clearly no other values of the parameters a_1, a_2, a_3 are alowed by (9) for $m_1 = m_2 = 1$. The two vectors $f_2 f_1|1,1\rangle$ and $f_2 f_2|1,1\rangle$ of zero weight are linearly independent since $f_1 f_2 \neq f_2 f_1$.

III. BASIS-DEFINING INEQUALITIES

A. The Lie algebras A_n

In this case the basis-defining inequalities can be written in a form that applies to any rank $\geqslant 1$. Any basis vector for $V(\Lambda)$, $\Lambda = (m_1, m_2, \ldots, m_n)$, is then

$$(f_1^{a_N} f_2^{a_{N-1}} \cdots f_n^{a_{N-n+1}})(f_1^{a_{N-n}} \cdots f_{n-1}^{a_{N-2n+2}}) \cdots (f_1^{a_3} f_2^{a_2}) f_1^{a_1}|\Lambda\rangle,$$

$$N = n(n+1)/2. \tag{14}$$

Brackets are used in (14) only to indicate the regularities of the sequence of f's. The basis-defining inequalities are found in Table I.

B. The Lie algebras B_n and C_n, $2 \leqslant n \leqslant 6$

The basis-defining inequalities for B_n and C_n are closely related due to the duality of their simple roots. Thus for each rank they have to be calculated for only one of the two algebras. For the other one they are obtained by renumbering of the roots and by a substitution of a_i's. Nevertheless, we write them out for both types.

TABLE I. Basis vectors and defining inequalities of an irreducible A_n representation $(m_1, m_2,..., m_n)$. A dotted line with A_r indicates the last inequality for A_r.

$(f_1^{a_N} f_2^{a_{N-1}} \cdots f_n^{a_{N-n+1}})(f_1^{a_{N-n}} \cdots f_{n-1}^{a_{N-2n+2}}) \cdots (f_1^{a_3} f_2^{a_2}) f_1^{a_1} \vert\Lambda\rangle, \quad N = n(n+1)/2$	
$0 \leqslant a_1 \leqslant m_1$	A_1
$0 \leqslant a_2 \leqslant m_2 + a_1$	
$0 \leqslant a_3 \leqslant \min[m_2, a_2]$	A_2
$0 \leqslant a_4 \leqslant m_3 + a_2$	
$0 \leqslant a_5 \leqslant \min[m_3 + a_3, a_4]$	
$0 \leqslant a_6 \leqslant \min[m_3, a_5]$	A_3
$0 \leqslant a_7 \leqslant m_4 + a_4$	
$0 \leqslant a_8 \leqslant \min[m_4 + a_5, a_7]$	
$0 \leqslant a_9 \leqslant \min[m_4 + a_6, a_8]$	
$0 \leqslant a_{10} \leqslant \min[m_4, a_9]$	A_4
$0 \leqslant a_{11} \leqslant m_5 + a_7$	
$0 \leqslant a_{12} \leqslant \min[m_5 + a_8, a_{11}]$	
$0 \leqslant a_{13} \leqslant \min[m_5 + a_9, a_{12}]$	
$0 \leqslant a_{14} \leqslant \min[m_5 + a_{10}, a_{13}]$	
$0 \leqslant a_{15} \leqslant \min[m_5, a_{14}]$	A_5
$0 \leqslant a_{16} \leqslant m_6 + a_{11}$	
$0 \leqslant a_{17} \leqslant \min[m_6 + a_{12}, a_{16}]$	
$0 \leqslant a_{18} \leqslant \min[m_6 + a_{13}, a_{17}]$	
$0 \leqslant a_{19} \leqslant \min[m_6 + a_{14}, a_{18}]$	
$0 \leqslant a_{20} \leqslant \min[m_6 + a_{15}, a_{19}]$	
$0 \leqslant a_{21} \leqslant \min[m_6, a_{20}]$	A_6
$\vdots$	
	A_{n-1}
$0 \leqslant a_{N-n+1} \leqslant m_n + a_{N-2n+2}$	
$0 \leqslant a_{N-n-2} \leqslant \min[m_n + a_{N-2n+3}, a_{N-n+1}]$	
$0 \leqslant a_{N-n+3} \leqslant \min[m_n + a_{N-2n+4}, a_{N-n+2}]$	
$\vdots$	
$0 \leqslant a_{N-1} \leqslant \min[m_n + a_{N-n+1}, a_{N-2}]$	
$0 \leqslant a_N \leqslant \min[m_n, a_{N-1}]$	A_n

Somewhat special is the lowest case of B_2 or C_2. Although the two algebras are isomorphic, it is sometimes convenient to distinguish two forms of the algebra by using opposite numbering of the simple roots. Our results are shown in Table II. The two sets of inequalities shown in the table correspond to different numberings of the simple roots and both refer to the basis vectors of the same form as given there. Consequently, Table II defines two quite different bases for each irreducible space.

For B_n (C_n) the root α_n is the short (long) one among the simple roots. A generic basis vector is of the form

$$(f_1^{a_{n^2}} \cdots f_n^{a_{n(n-1)+1}} \cdots (f_1^{a_{2n}} \cdots f_n^{a_{n+1}}) f_1^{a_n} \cdots f_n^{a_1} \times \vert m_1, m_2,..., m_n\rangle. \qquad (15)$$

TABLE II. Basis vectors and defining inequalities of an irreducible representation (m_1,m_2) of B_2 or C_2. The two cases differ by numbering of the simple roots.

$f_1^{a_4} f_2^{a_3} f_1^{a_2} f_2^{a_1} \vert m_1, m_2\rangle$	
 1 2	 1 2
$0 \leqslant a_1 \leqslant m_2$	$0 \leqslant a_1 \leqslant m_2$
$0 \leqslant a_2 \leqslant m_1 + a_1$	$0 \leqslant a_2 \leqslant m_1 + 2a_1$
$0 \leqslant a_3 \leqslant \min[m_1 + a_2, 2a_2]$	$0 \leqslant a_3 \leqslant \min[\frac{1}{2}(m_1 + a_2), a_2]$
$0 \leqslant a_4 \leqslant \min[m_1, \frac{1}{2}a_3]$	$0 \leqslant a_4 \leqslant \min[\mathrm{m}_1, \mathrm{a}_3]$

The basis-defining inequalities are given in Tables III, V, VII, and IX for B_n ($3 \leqslant n \leqslant 6$) and in Tables IV, VI, VIII, and X for C_n ($3 \leqslant n \leqslant 6$). The transition $B_n \to C_n$ is done by the following transformation:

$$m_n \leftrightarrow 2m_n, \qquad a_r \leftrightarrow 2a_r, \quad \text{for all } r = 1 \mod n. \qquad (16)$$

C. The Lie algebras D_n, $4 \leqslant n \leqslant 6$

The basis vectors are taken of the form

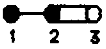
$$(f_1^{a_{n(n-1)}} \cdots f_n^{a_{n(n-2)+1}}) \cdots (f_1^{a_{2n}} \cdots (f_n^{a_{n+1}})(f_1^{a_n} \cdots f_n^{a_1}) \times \vert m_1, m_2,..., m_n\rangle \qquad (17)$$

TABLE III. Basis-defining inequalities for any irreducible representation (m_1,m_2,m_3) of B_3.


1 2 3

$0 \leqslant a_1 \leqslant m_3$
$0 \leqslant a_2 \leqslant m_2 + a_1$
$0 \leqslant a_3 \leqslant m_1 + a_2$
$0 \leqslant a_4 \leqslant \min[m_2 + a_2, 2a_2]$
$0 \leqslant a_5 \leqslant \min[m_2 + a_3, m_1 + a_4, a_3 + \frac{1}{2}a_4]$
$0 \leqslant a_6 \leqslant \min[m_2, \frac{1}{2}a_4, a_5]$
$0 \leqslant a_7 \leqslant \min[m_1 + a_5, 2a_3, 2a_5 - 2a_6]$
$0 \leqslant \mathrm{a}_7 \leqslant \min[m_1 + a_6, \frac{1}{2}a_7]$
$0 \leqslant a_9 \leqslant \min[m_1, a_8]$

TABLE IV. Basis-defining inequalities for any irreducible representation (m_1,m_2,m_3) of C_3.



$0\leqslant a_1\leqslant m_3$
$0\leqslant a_2\leqslant m_2+2a_1$
$0\leqslant a_3\leqslant m_1+a_2$
$0\leqslant a_4\leqslant \min[a_2,\frac{1}{2}(m_2+a_2)]$
$0\leqslant a_5\leqslant \min[a_3+a_4,m_1+2a_4,m_2+a_3]$
$0\leqslant a_6\leqslant \min[m_2,a_4,a_5]$
$0\leqslant a_7\leqslant \min[a_3,a_5-a_6,\frac{1}{2}(m_1+a_5)]$
$0\leqslant a_8\leqslant \min[a_7,m_1+a_6]$
$0\leqslant a_9\leqslant \min[a_8,m_1]$

for all ranks. The basis defining inequalities are given in Tables XI–XIII.

D. The Lie algebra G_2

Here α_2 is the short root. A generic vector is written as

$$f_2^{a_6}f_1^{a_5}f_2^{a_4}f_1^{a_3}f_2^{a_2}f_1^{a_1}|m_1,m_2\rangle \tag{18}$$

and the basis-defining inequalities are found in Table XIV.

IV. DERIVATION OF BASIS-DEFINING INEQUALITIES

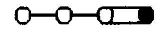
In order to set up a generic form of basis vectors for an Lie algebra $\mathfrak{g}$ relative to a Cartan subalgebra $\mathfrak{h}$ one has to choose a form of the opposite involution, which we denote by inv. That is, the element of the Weyl group of $\mathfrak{g}$ that transforms every positive root into a negative root or, equivalently, every highest weight of a representation into the lowest one. The involution can be written as a sequence of N reflections r_i, $1\leqslant i\leqslant n$, in planes orthogonal to simple roots α_i of $\mathfrak{g}$, where N is the number of positive roots of $\mathfrak{g}$. The actual expression of inv in terms of reflections r_i is far from unique. Thus, for instance, we took inv $= r_1r_2r_1$ for A_2, but the choice $\text{inv}_1 = r_2r_1r_2$ would have been equally admissible. Once inv is fixed, say

$$\text{inv} = r_{i_N}r_{i_{N-1}}\cdots r_{i_2}r_{i_1}, \tag{19}$$

TABLE V. Basis-defining inequalities for any irreducible representation (m_1,m_2,m_3,m_4) of B_4.

$0\leqslant a_1\leqslant m_4$
$0\leqslant a_2\leqslant m_3+a_1$
$0\leqslant a_3\leqslant m_2+a_2$
$0\leqslant a_4\leqslant m_1+a_3$
$0\leqslant a_5\leqslant \min[m_3+a_2,2a_2]$
$0\leqslant a_6\leqslant \min[m_3+a_3,m_2+a_5,a_3+\frac{1}{2}a_5]$
$0\leqslant a_7\leqslant \min[m_3+a_4,m_1+a_6,a_4+a_6,a_4+\frac{1}{2}a_5]$
$0\leqslant a_8\leqslant \min[m_3,\frac{1}{2}a_5,a_6,a_7]$
$0\leqslant a_9\leqslant \min[m_2+a_6,2a_3,2a_4-2a_7+2a_6,2a_6-2a_8]$
$0\leqslant a_{10}\leqslant \min[m_2+a_7,m_1+a_9,a_4+\frac{1}{2}a_9,a_7-a_8+\frac{1}{2}a_9]$
$0\leqslant a_{11}\leqslant \min[m_2+a_8,\frac{1}{2}a_9,a_{10}]$
$0\leqslant a_{12}\leqslant \min[m_2,a_{11}]$
$0\leqslant a_{13}\leqslant \min[m_1+a_{10},2a_4,2a_7-2a_8,2a_{10}-2a_{11}]$
$0\leqslant a_{14} v \min[m_1+a_{11},\frac{1}{2}a_{13}]$
$0\leqslant a_{15}\leqslant \min[m_1+a_{12},a_{14}]$
$0\leqslant a_{16}\leqslant \min[m_1,a_{15}]$

TABLE VI. Basis-defining inequalities for any irreducible representation (m_1,m_2,m_3,m_4) of C_4.



$0\leqslant a_1\leqslant m_4$
$0\leqslant a_2\leqslant m_3+2a_1$
$0\leqslant a_3\leqslant m_2+a_2$
$0\leqslant a_4\leqslant m_1+a_3$
$0\leqslant a_5\leqslant \min[\frac{1}{2}(m_3+a_2),a_2]$
$0\leqslant a_6\leqslant \min[m_2+2a_5,m_3+a_3,a_3+a_5]$
$0\leqslant a_7\leqslant \min[m_1+a_6,m_3+a_4,a_4+a_5,a_4+a_6]$
$0\leqslant a_8\leqslant \min[m_3,a_5,a_6,a_7]$
$0\leqslant a_9\leqslant \min[\frac{1}{2}(m_2+a_6),a_3,a_6-a_8,a_4+a_6-a_7]$
$0\leqslant a_{10}\leqslant \min[m_1+2a_9,m_2+a_7,a_4+a_9,a_7-a_8+a_9]$
$0\leqslant a_{11}\leqslant \min[m_2+a_8,a_9,a_{10}]$
$0\leqslant a_{12}\leqslant \min[m_2,a_{11}]$
$0\leqslant a_{13}\leqslant \min[\frac{1}{2}(m_1+a_{10}),a_4,a_7-a_8,a_{10}-a_{11}]$
$0\leqslant a_{14}\leqslant \min[m_1+a_{11},a_{13}]$
$0\leqslant a_{15}\leqslant \min[m_1+a_{12},a_{14}]$
$0\leqslant a_{16}\leqslant \min[m_1,a_{15}]$

then the general form of a basis vector is chosen to be

$$f_{i_N}^{a_N}\cdots f_{i_2}^{a_2}f_{i_1}^{a_1}|\Lambda_2\rangle, \tag{20}$$

where f_q is the generator corresponding to the simple root α_q. For definiteness we assume that f_q, e_q, and h_q satisfy the commutation relations (1) for every q, $1\leqslant q\leqslant n$.

The derivation of the basis defining inequalities proceeds recursively. Suppose that the first $k-1$ inequalities are already known and that $i_k = p$ in (20), $1\leqslant k\leqslant N$. The upper limit $L(k)$ of a_k is then found from the product of $h_p = 2\alpha_p/(\alpha_p,\alpha_p)$ with the weight μ of the vector (20) with $a_k = a_{k+1} = \cdots = a_N = 0$:

$$\langle\mu,h_p\rangle = \left\langle \Lambda - \sum_{r=1}^{k-1} a_r\alpha_{i_r},h_p\right\rangle = m_p - \sum_{r=1}^{k-1} a_r\langle\alpha_{i_r},h_p\rangle, \tag{21}$$

where $\langle\alpha_i,h_p\rangle$ is the element A_{ip} of the Cartan matrix A of the Lie algebra $\mathfrak{g}$. Next we split out of (21) the terms with a's corresponding to the same generator f_p:

$$\langle\mu,h_p\rangle = m_p\left\langle \sum_{r=1}^{k-1}(1-\delta_{pi_r})a_r\alpha_{i_r} + \sum_{r=1}^{k-1}\delta_{pi_r}a_r\alpha_{i_r},h_p\right\rangle. \tag{22}$$

Here δ_{pq} is the Kronecker symbol. Then

$$L(k) = m_p - \sum_{r=1}^{k-1}(1-\delta_{pi_r})a_r\langle\alpha_{i_r},h_p\rangle - \sum_{r=1}^{k-1}\delta_{pi_r}b_r\langle\alpha_{i_r},h_p\rangle, \tag{23}$$

where b_r is the average value of a_r,

$$b_r = \tfrac{1}{2}(\max a_r + \min a_r). \tag{24}$$

It is calculated from the first $k-1$ inequalities assuming that all the parameters a_r that occur in the first sum of (23) are fixed, i.e., those corresponding to f_r in (20) with $r\neq p$.

Consider an example of the Lie algebra C_3 with the standard numbering of its three simple roots, where α_3, is the long root, and let us illustrate a derivation of the inequalities of Table IV. An irreducible representation space of the highest weight $\Lambda = (m_1,m_2,m_3)$ decomposes into the direct sum $V(\Lambda) = \oplus_\mu V(\Lambda;\mu)$ of subspaces $V(\Lambda;\mu)$ labeled by the weights μ of the weight system $\Omega(\Lambda)$ of Λ. We choose the opposite involution of C_3 as follows:

TABLE VII. Basis-defining inequalities for an irreducible representation (m_1,m_2,m_3,m_4) of B_5.

$0<a_1<m_5$
$0<a_2<m_4+a_1$
$0<a_3<m_3+a_2$
$0<a_4<m_2+a_3$
$0<a_5<m_1+a_4$
$0<a_6<\min[2a_2,m_4+a_2]$
$0<a_7<\min[a_3+\frac{1}{2}a_6,m_3+a_6,m_4+a_3]$
$0<a_8<\min[m_2+a_7,m_4+a_4,a_4+a_7,a_4+\frac{1}{2}a_6]$
$0<a_9<\min[m_1+a_8,m_4+a_5,a_5+a_7,a_5+a_8,a_5+\frac{1}{2}a_6]$
$0<a_{10}<\min[\frac{1}{2}a_6,a_7,a_8,a_9,m_4]$
$0<a_{11}<\min[2a_3,m_3+a_7,2a_4+2a_7-2a_8,2a_5+2a_7-2a_9,2a_7-2a_{10}]$
$0<a_{12}<\min[m_2+a_{11},m_3+a_8,a_4+\frac{1}{2}a_{11},a_5+a_8-a_9+\frac{1}{2}a_{11},a_8-a_{10}+\frac{1}{2}a_{11}]$
$0<a_{13}<\min[m_1+a_{12},m_3+a_9,a_5+\frac{1}{2}a_{11},a_5+a_{12},a_9-a_{10}+\frac{1}{2}a_{11},a_9-a_{10}+a_{12}]$
$0<a_{14}<\min[a_{12},a_{13},m_3+a_{10},\frac{1}{2}a_{11}]$
$0<a_{15}<\min[a_{14},m_3]$
$0<a_{16}<\min[2a_4,m_2+a_{12},2a_5+2a_8-2a_9,2a_8-2a_{10},2a_5+2a_{12}-2a_{13},2a_9+2a_{12}-2a_{10}-2a_{13},2a_{12}-2a_{14}]$
$0<a_{17}<\min[m_1+a_{16},m_2+a_{13},a_5+\frac{1}{2}a_{16},a_9-a_{10}+\frac{1}{2}a_{16},a_{13}-a_{14}+\frac{1}{2}a_{16}]$
$0<a_{18}<\min[m_2+a_{14},a_{17},\frac{1}{2}a_{16}]$
$0<a_{19}<\min[m_2+a_{15},a_{18}]$
$0<a_{20}<\min[m_2,a_{19}]$
$0<a_{21}<\min[m_1+a_{17},2a_5,2a_9-2a_{10},2a_{13}-2a_{14},2a_{17},-2a_{18}]$
$0<a_{22}<\min[m_1+a_{18},\frac{1}{2}a_{21}]$
$0<a_{23}<\min[m_1+a_{19},a_{22}]$
$0<a_{24}<\min[m_1+a_{20},a_{23}]$
$0<a_{25}<\min[m_1,a_{24}]$

$$\mathrm{inv}=r_1r_2r_3r_1r_2r_3r_1r_2r_3. \tag{25}$$

Successive application of individual reflections from inv transform $\Lambda=(m_1,m_2,m_3)$ into the lowest weight, $\mathrm{inv}\,\Lambda=-\Lambda=(-m_1,-m_2,-m_3)$ of $\Omega(\Lambda)$. For details of the action of the Weyl group on weights see, for instance, Ref. 1 or 3. A Verma basis of $V(\Lambda)$ consists then of the vectors $|\mu\rangle$ given by (15) with $n=3$. The weight μ is then

$$\mu=\Lambda-(a_1+a_4+a_7)\alpha_3-(a_2+a_5+a_8)\alpha_2-(a_3+a_6+a_9)\alpha_1. \tag{26}$$

Expressing Λ, α_1, α_2, and α_3 in the basis of fundamental weights ("omega basis" of Table 2 of Ref. 1), one has

$$\mu=(m_1-(a_3+a_6+a_9)-a_2-a_5-a_9,\; m_2-2(a_2+a_3+a_5+a_6+a_8+a_9)-a_1-a_5-a_7,\; m_3-2(a_1-a_4+a_7)-a_2-a_5-a_8). \tag{27}$$

The first three inequalities of Table IV are a direct consequence of the weight algorithm. The upper limit L (4) of a_4 is then calculated according to (23):

TABLE VIII. Basis-defining inequalities for an irreducible representation (m_1,m_2,m_3,m_4,m_5) of C_5.

$0<a_1<m_5$
$0<a_2,m_4+2a_1$
$0<a_3<m_3+a_2$
$0<a_4<m_2+a_3$
$0<a_5<m_1+a_4$
$0<a_6<\min[a_2,\frac{1}{2}(m_4+a_2]$
$0<a_7<\min[m_3+2a_6,m_4+a_3,a_3+a_6]$
$0<a_8<\min[m_2+a_7,m_4+a_4,a_4+a_6,a_4+a_7]$
$0<a_9<\min[m_1+a_8,m_4+a_5,a_5+a_6,a_5+a_7,a_5+a_8]$
$0<a_{10}<\min[m_4,a_6,a_7,a_8,a_9]$
$0<a_{11}<\min[a_3,\frac{1}{2}(m_3+a_7),a_7-a_{10},a_4+a_7-a_8,a_5+a_7-a_9]$
$0<a_{12}<\min[m_2+2a_{11},m_3+a_8,a_4+a_{11},a_5+a_8-a_9+a_{11},a_8-a_{10}+a_{11}]$
$0<a_{13}<\min[m_1+a_{12},m_3+a_9,a_5+a_{11},a_5+a_{12},a_9-a_{10}+a_{11},a_9-a_{10}+a_{12}]$
$0<a_{14}<\min[m_3+a_{10},a_{11},a_{12},a_{13}]$
$0<a_{15}<\min[m_3,a_{14}]$
$0<a_{16}>\min[a_4,\frac{1}{2}(m_2+a_{12}),a_8-a_{10},a_5+a_8-a_9,a_5+a_{12}-a_{13},a_{12}-a_{14},a_9-a_{10}+a_{12}-a_{13}]$
$0<a_{17}<\min[m_1+2a_{16},m_2+a_{13},a_5+a_{16},a_9-a_{10}+a_{16},a_{13}-a_{14}+a_{16}]$
$0<a_{18}<\min[m_2+a_{14},a_{16},a_{17}]$
$0<a_{19}<\min[m_2+a_{15},a_{18}]$
$0<a_{20}<\min[m_2,a_{19}]$
$0<a_{21}<\min[a_5,\frac{1}{2}(m_1+a_{17}),a_9-a_{10},a_{13}-a_{14},a_{17}-a_{18}]$
$0<a_{22}<\min[m_1+a_{18},a_{21}]$
$0<a_{23}<\min[m_1+a_{19},a_{22}]$
$0<a_{24}<\min[m_1+a_{20},a_{23}]$
$0<a_{25}<\min[m_1,a_{24}]$

TABLE IX. Basis-defining inequalities for an irreducible representation $(m_1,m_2,m_3,m_4,m_5,m_6)$ of B_6.

$0 < a_1 < m_6$

$0 < a_2 < m_5 + a_1$

$0 < a_3 < m_4 + a_2$

$0 < a_4 < m_3 + a_3$

$0 < a_5 < m_2 + a_4$

$0 < a_6 < m_1 + a_5$

$0 < a_7 < \min[2a_2, m_5 + a_2]$

$0 < a_8 < \min[m_4 + a_7, m_5 + a_3, a_3 + \frac{1}{2}a_7]$

$0 < a_9 < \min[m_3 + a_8, m_5 + a_4, a_4 + a_8, a_4 + \frac{1}{2}a_7]$

$0 < a_{10} < \min[m_2 + a_9, m_5 + a_5, a_5 + a_9, a_5 + a_8, a_5 + \frac{1}{2}a_7]$

$0 < a_{11} < \min[m_1 + a_{10}, m_5 + a_6, a_6 + a_{10}, a_6 + a_9, a_6 + a_8, a_6 + \frac{1}{2}a_7]$

$0 < a_{12} < \min[m_5, a_{11}, a_{10}, a_9, a_8, \frac{1}{2}a_7]$

$0 < a_{13} < \min[2a_3, m_4 + a_8, 2a_4 + 2a_8 - 2a_9, 2a_5 + 2a_8 - 2a_{10}, 2a_6 + 2a_8 - 2a_{11}, 2a_8 - 2a_{12}]$

$0 < a_{14} < \min[m_3 + a_{13}, m_4 + a_9, a_4 + \frac{1}{2}a_{13}, a_9 + a_5 - a_{10} + \frac{1}{2}a_{13}, a_9 + a_6 - a_{11} + \frac{1}{2}a_{13}, a_9 - a_{12} + \frac{1}{2}a_{13}]$

$0 < a_{15} < \min[m_2 + a_{14}, m_4 + a_{10}, a_5 + \frac{1}{2}a_{13}, a_5 + a_{14}, a_6 + a_{10} - a_{11} + \frac{1}{2}a_{13}, a_6 + a_{10} - a_{11} + a_{14}, a_{10} - a_{12} + \frac{1}{2}a_{13}, a_{10} - a_{12} + a_{14}]$

$0 < a_{16} < \min[m_1 + a_{15}, m_4 + a_{11}, a_6 + \frac{1}{2}a_{13}, a_6 + a_{14}, a_6 + a_{15}, a_{11} - a_{12} + \frac{1}{2}a_{13}, a_{11} - a_{12} + a_{14}, a_{11} - a_{12} + a_{15}]$

$0 < a_{17} < \min[m_4 + a_{12}, \frac{1}{2}a_{13}, a_{14}, a_{15}, a_{16}]$

$0 < a_{18} < \min[m_4, a_{17}]$

$0 < a_{19} < \min[m_3 + a_{14}, 2a_4, 2a_5 + 2a_9 - 2a_{10}, 2a_6 + 2a_9 - 2a_{11}, 2a_5 + 2a_{14} - 2a_{15}, 2a_6 + 2a_{14} - 2a_1,$
$\quad 2a_6 + 2a_{10} - 2a_{11} + 2a_{14} - 2a_{15}, 2a_{10} - 2a_{12} + 2a_{14} - 2a_{15}, 2a_{11} - 2a_{12} + 2a_{14} - 2a_{16}, 2a_9 - 2a_{12}, 2a_{14} - 2a_{17}]$

$0 < a_{20} < \min[m_2 + a_{19}, m_3 + a_{15}, a_5 + \frac{1}{2}a_{19}, a_6 + a_{10} - a_{11} + \frac{1}{2}a_{19}, a_6 + a_{15} - a_{16} + \frac{1}{2}a_{19}, a_{10} - a_{12} + \frac{1}{2}a_{19},$
$\quad a_{11} - a_{12} + a_{15} - a_{16} + \frac{1}{2}a_{19}, a_{15} - a_{17} + \frac{1}{2}a_{19}]$

$0 < a_{21} < \min[m_1 + a_{20}, m_3 + a_{16}, a_6 + \frac{1}{2}a_{19}, a_6 + a_{20}, a_{11} - a_{12} + \frac{1}{2}a_{19}, a_{11} - a_{12} + a_{20}, a_{16} - a_{17} + \frac{1}{2}a_{19}, a_{16} - a_{17} + a_{20}]$

$0 < a_{22} < \min[m_3 + a_{17}, \frac{1}{2}a_{19}, a_{20}, a_{21}]$

$0 < a_{23} < \min[m_3 + a_{18}, a_{22}]$

$0 < a_{24} < \min[m_3, a_{23}]$

$0 < a_{25} < \min[m_2 + a_{20}, 2a_5, 2a_6 + 2a_{10} - 2a_{11}, 2a_6 + 2a_{15} - 2a_{16}, 2a_6 + 2a_{20} - 2a_{21},$
$\quad 2a_{11} - 2a_{12} + 2a_{15} - 2a_{16}, 2a_{11} - 2a_{12} + 2a_{20} - 2a_{21}, 2a_{16} - 2a_{17} + 2a_{20} - 2a_{21}, 2a_{10} - 2a_{12}, 2a_{15} - 2a_{17}, 2a_{20} - 2a_{22}]$

$0 < a_{26} < \min[m_1 + a_{25}, m_2 + a_{21}, a_6 + \frac{1}{2}a_{25}, a_{11} - a_{12} + \frac{1}{2}a_{25}, a_{16} - a_{17} + \frac{1}{2}a_{25}, a_{21} - a_{22} + \frac{1}{2}a_{25}]$

$0 < a_{27} < \min[m_2 + a_{22}, a_{26}, \frac{1}{2}a_{25}]$

$0 < a_{28} < \min[m_2 + a_{23}, a_{27}]$

$0 < a_{29} < \min[m_2 + a_{24}, a_{28}]$

$0 < a_{30} < \min[m_2, a_{29}]$

$0 < a_{31} < \min[m_1 + a_{26}, 2a_6, 2a_{11} - 2a_{12}, 2a_{16} - 2a_{17}, 2a_{21} - 2a_{22}, 2a_{26} - 2a_{27}]$

$0 < a_{32} < \min[m_1 + a_{27}, \frac{1}{2}a_{31}]$

$0 < a_{33} < \min[m_1 + a_{28}, a_{32}]$

$0 < a_{34} < \min[m_1 + a_{29}, a_{33}]$

$0 < a_{35} < \min[m_1 + a_{30}, a_{34}]$

$0 < a_{36} < \min[m_1, a_{35}]$

$$L(4) = m_3 - a_3\langle\alpha_1,h_3\rangle - a_2\langle\alpha_2,h_3\rangle - b_1\langle\alpha_3,h_3\rangle, \tag{28}$$

where b_1 is found from (24). Namely, using the first three inequalities one has

$$\max a_1 = m_3,$$
$$\min a_1 = \max[0,\frac{1}{2}(a_2 - m_2)] = \min[0,\frac{1}{2}(m_2 - a_2)]$$
$$= a_2 + \min[a_2,\frac{1}{2}(m_2 + a_2)]. \tag{29}$$

Substituting (29) into (24) and using that in (28) together with the matrix elements

$$\langle\alpha_1,h_3\rangle = 0, \quad \langle\alpha_2,h_3\rangle = -1, \quad \langle\alpha_3,h_3\rangle = 2 \tag{30}$$

of the C_3-Cartan matrix, one gets $L(4) = \min[a_2,\frac{1}{2}(m_2 + a_2)]$. The upper limits $L(k)$ of Table IV with $5 < k < 9$ are found in exactly the same way.

Finally let us illustrate a failure of the method. Let us choose again the Lie algebra C_3 but this time put

$$\text{inv} = r_1r_2r_3r_2r_1r_3r_2r_3r_2 \tag{31}$$

instead of (25). It is a perfectly valid expression of the opposite involution, because inv $\Lambda = -\Lambda$. Furthermore it implies that the basis vectors are of the form

$$f_1^{a_9}f_2^{a_8}f_3^{a_7}f_2^{a_6}f_1^{a_5}f_3^{a_4}f_2^{a_3}f_3^{a_2}f_2^{a_1}|m_1,m_2,m_3\rangle \tag{32}$$

rather than (15). Derivation of the first five inequalities proceeds without problem according to the procedure described above. One gets

$$\begin{aligned} &0 < a_1 < m_2, \\ &0 < a_2 < m_3 + a_1, \\ &0 < a_3 < \min[2a_2, m_3 + a_2], \\ &0 < a_4 < \min[m_3, \frac{1}{2}a_3], \\ &0 < a_5 < m_1 + a_1 + a_3. \end{aligned} \tag{33}$$

At the next step when $L(6) = \langle\Lambda - (b_1 + b_3)\alpha_2 - a_5\alpha_1 - (a_2 + a_4)\alpha_3, h_2\rangle$, one needs to find the average $b_1 + b_3$ of $a_1 + a_3$. However, it is not clear how to proceed because the variables a_1 and a_3 are not independent in (33) so that (24) does not apply. A numerical evaluation is undoubtedly all but useless.

V. COMPLETENESS OF THE VERMA BASES

The choice of positive roots for $\mathfrak{g}$ (relative to a Cartan subalgebra $\mathfrak{h}$) determines a decomposition

TABLE X. Basis-defining inequalities for an irreducible representation $(m_1,m_2,m_3,m_4,m_5,m_6)$ of C_6.

$0<a_1<m_6$
$0<a_2<m_5+2a_1$
$0<a_3<m_4+a_2$
$0<a_4<m_3+a_3$
$0<a_5<m_2+a_4$
$0<a_6<m_1+a_5$
$0<a_7<\min[a_2,\frac{1}{2}(m_5+a_2]$
$0<a_8<\min[m_4+2a_7,m_5+a_3,a_3+a_7]$
$0<a_9<\min[m_3+a_8,m_5+a_4,a_4+a_8,a_4+a_7]$
$0<a_{10}<\min[m_2+a_9,m_5+a_5,a_5+a_9,a_5+a_8,a_5+a_7]$
$0<a_{11}<\min[m_1+a_{10},m_5+a_6,a_6+a_{10},a_6+a_9,a_6+a_8,a_6+a_7]$
$0<a_{12}<\min[m_5,a_7,a_8,a_9,a_{10},a_{11}]$
$0<a_{13}<\min[a_3,\frac{1}{2}(m_4+a_8),a_8-a_{12},a_4+a_8-a_9,a_5+a_8-a_{10},a_6+a_8-a_{11}]$
$0<a_{14}<\min[m_3+2a_{13},m_4+a_9,a_4+a_{13},a_5+a_9-a_{10}+a_{13},a_6+a_9-a_{11}+a_{13},a_9-a_{12}+a_{13}]$
$0<a_{15}<\min[m_2+a_{14},m_4+a_{10},a_5+a_{13},a_5+a_{14},a_6+a_{10}-a_{11}+a_{13},a_6+a_{10}-a_{11}+a_{14},a_{10}-a_{12}+a_{13},$
$a_{10}-a_{12}+a_{14}]$
$0<a_{16}<\min[m_1+a_{15},m_4+a_{11},a_6+a_{13},a_6+a_{14},a_6+a_{15},a_{11}-a_{12}+a_{13},a_{11}-a_{12}+a_{14},a_{11}-a_{12}+a_{15}]$
$0<a_{17}<\min[m_4+a_{12},a_{13},a_{14},a_{15},a_{16}]$
$0<a_{18}<\min[m_4,a_{17}]$
$0<a_{19}<\min[a_4,\frac{1}{2}(m_3+a_{14}),a_5+a_9-a_{10},a_6+a_9-a_{11},a_5+a_{14}-a_{15},a_6+a_{14}-a_{16},$
$a_6+a_{10}-a_{11}+a_{14}-a_{15},a_{10}-a_{12}+a_{14}-a_{15},a_{11}-a_{12}+a_{14}-a_{16},a_9-a_{12},a_{14}-a_{17}]$
$0<a_{20}<\min[m_2+2a_{19},m_3+a_{15},a_5+a_{19},a_6+a_{10}-a_{11}+a_{19},a_6+a_{15}-a_{16}+a_{19},a_{10}-a_{12}+a_{19},$
$a_{11}-a_{12}+a_{15}-a_{16}+a_{19},a_{15}-a_{17}+a_{19}]$
$0<a_{21}<\min[m_1+a_{20},m_3+a_{16},a_6+a_{19},a_6+a_{20},a_{11}-a_{12}+a_{19},a_{11}-a_{12}+a_{20},a_{16}-a_{17}+a_{19},$
$a_{16}-a_{17}+a_{20}]$
$0<a_{22}<\min[m_3+a_{17},a_{19},a_{20},a_{21}]$
$0<a_{23}<\min[m_3+a_{18},a_{22}]$
$0<a_{24}<\min[m_3,a_{23}]$
$0<a_{25}<\min[a_5,\frac{1}{2}(m_2+a_{20}),a_6+a_{10}-a_{11},a_6+a_{15}-a_{16},a_6+a_{20}-a_{21},a_{11}-a_{12}+a_{15}-a_{16},$
$a_{11}-a_{12}+a_{20}-a_{21},a_{16}-a_{17}+a_{20}-a_{21},a_{10}-a_{12},a_{15}-a_{17},a_{20}-a_{22}]$
$0<a_{26}<\min[m_1+2a_{25},m_2+a_{21},a_6+a_{25},a_{11}-a_{12}+a_{25},a_{16}-a_{17}+a_{25},a_{21}-a_{22}+a_{25}]$
$0<a_{27}<\min[m_2+a_{22},a_{25},a_{26}]$
$0<a_{28}<\min[m_2+a_{23},a_{27}]$
$0<a_{29}<\min[m_2+a_{24},a_{28}]$
$0<a_{30}<\min[m_2,a_{29}]$
$0<a_{31}<\min[a_6,\frac{1}{2}(m_1+a_{26}),a_{11}-a_{12},a_{16}-a_{17},a_{21}-a_{22},a_{26}-a_{27}]$
$0<a_{32}<\min[m_1+a_{27},a_{31}]$
$0<a_{33}<\min[m_1+a_{28},a_{32}]$
$0<a_{34}<\min[m_1+a_{29},a_{33}]$
$0<a_{35}<\min[m_1+a_{30},a_{34}]$
$0<a_{36}<\min[m_1+a_{31},a_{35}]$

TABLE XI. Basis-defining inequalities for an irreducible representation (m_1,m_2,m_3,m_4) of D_4.

$0<a_1<m_4$
$0<a_2<m_3$
$0<a_3<m_2+a_1+a_2$
$0<a_4<m_1+a_3$
$0<a_5<\min[a_3,m_2+a_2]$
$0<a_6<\min[a_3,m_2+a_1,m_2+a_3-a_5]$
$0<a_7<\min[a_4+a_6,a_4+a_5,m_1+a_5+a_6,m_2+a_4]$
$0<a_8<\min[a_5,a_6,a_7,m_2]$
$0<a_9<\min[a_4,m_1+a_6,a_7-a_8]$
$0<a_{10}<\min[a_7-a_8,a_4,m_1+a_5,m_1+a_7-a_9]$
$0<a_{11}<\min[a_9,a_{10},m_1+a_8]$
$0<a_{12}<\min[a_{11},m_1]$

TABLE XII. Basis-defining inequalities for an irreducible representation (m_1,m_2,m_3,m_4,m_5) of D_5.

$0<a_1<m_5$
$0<a_2<m_4$
$0<a_3<m_3+a_1+a_2$
$0<a_4<m_2+a_3$
$0<a_5<m_1+a_4$
$0<a_6<\min[a_3,m_3+a_2]$
$0<a_7<\min[a_3,m_3+a_1,m_3+a_3-a_6]$
$0<a_8<\min[a_4+a_6,a_4+a_7,m_2+a_6+a_7,m_3+a_4]$
$0<a_9<\min[m_1+a_8,m_3+a_5,a_5+a_6,a_5+a_7,a_5+a_8]$
$0<a_{10}<\min[m_3,a_6,a_7,a_8,a_9]$
$0<a_{11}<\min[a_4,m_2+a_7,a_8-a_{10},a_5+a_8-a_9]$
$0<a_{12}<\min[a_4,m_2+a_6,m_2+a_8-a_{11},a_5+a_8-a_9,a_8-a_{10}]$
$0<a_{13}<[\min m_1+a_{11}+a_{12},a_9+a_{11}-a_{10},a_9+a_{12}-a_{10},m_2+a_9,$
$a_5+a_{11},a_5+a_{12}]$
$0<a_{14}<\min[a_{11},a_{12},a_{13},m_2+a_{10}]$
$0<a_{15}<\min[a_{14},m_2]$
$0<a_{16}<\min[a_5,m_1+a_{12},a_{13}-a_{14},a_9-a_{10}]$
$0<a_{17}<\min[a_5,m_1+a_{11},m_1+a_{13}-a_{16},a_9-a_{10},a_{13}-a_{14}]$
$0<a_{18}<\min[a_{16},a_{17},m_1+a_{14}]$
$0<a_{19}<\min[a_{18},m_1+a_{15}]$
$0<a_{20}<\min[a_{19},m_1]$

TABLE XIII. Basis-defining inequalities for an irreducible representation $(m_1,m_2,m_3,m_4,m_5,m_6)$ of D_6.

$0<a_1<m_6$
$0<a_2<m_5$
$0<a_3<m_4+a_1+a_2$
$0<a_4<m_3+a_3$
$0<a_5<m_2+a_4$
$0<a_6<m_1+a_5$
$0<a_7<\min[m_4+a_2,a_3]$
$0<a_8<\min[a_3,m_4+a_1,m_4+a_3-a_7]$
$0<a_9<\min[m_4+a_4,m_3+a_7+a_8,a_4+a_7,a_4+a_8]$
$0<a_{10}<\min[m_4+a_5,m_2+a_9,a_5+a_7,a_5+a_8,a_5+a_9]$
$0<a_{11}<\min[m_4+a_6,m_1+a_{10},a_6+a_7,a_6+a_8,a_6+a_9,a_6+a_{10}]$
$0<a_{12}<\min[m_4,a_7,a_8,a_9,a_{10},a_{11}]$
$0<a_{13}<\min[a_4,m_3+a_8,a_9+a_5-a_{10},a_9+a_6-a_{11},a_9-a_{12}]$
$0<a_{14}<\min[a_4,m_3+a_7,m_3+a_9-a_{13},a_9+a_5-a_{10},a_9+a_6-a_{11},a_9-a_{12}]$
$0<a_{15}<\min[m_3+a_{10},m_2+a_{13}+a_{14},a_5+a_{13},a_5+a_{14},a_{10}-a_{12}+a_{13},a_{10}-a_{12}+a_{14},$
$\quad a_6+a_{10}-a_{11}+a_{13},a_6+a_{10}-a_{11}+a_{14}]$
$0<a_{16}<\min[m_3+a_{11},m_1+a_{15},a_6+a_{13},a_6+a_{14},a_6+a_{15},a_{11}-a_{12}+a_{13},a_{11}-a_{12}+a_{14},a_{11}-a_{12}+a_{15}]$
$0<a_{17}<\min[a_{13},a_{14},a_{15},a_{16},m_3+a_{12}]$
$0<a_{18}<\min[m_3,a_{17}]$
$0<a_{19}<\min[a_5,m_2+a_{14},a_{10}-a_{12},a_{15}-a_{17},a_6+a_{10}-a_{11},a_6+a_{15}-a_{16},a_{11}-a_{12}+a_{15}-a_{16}]$
$0<a_{20}<\min[a_5,m_2+a_{13},m_2+a_{15}-a_{19},a_{10}-a_{12},a_{15}-a_{17},a_6+a_{10}-a_{11},a_6+a_{15}-a_{16},$
$\quad a_{11}-a_{12}+a_{15}-a_{16}]$
$0<a_{21}<\min[m_2+a_{16},m_1+a_{19}+a_{20},a_{11}-a_{12}+a_{19},a_{11}-a_{12}+a_{20},a_{16}-a_{17}+a_{19},$
$\quad a_{16}-a_{17}+a_{20},a_6+a_{19},a_6+a_{20}]$
$0<a_{22}<\min[m_2+a_{17},a_{19},a_{20},a_{21}]$
$0<a_{23}<\min[a_{22},m_2+a_{18}]$
$0<a_{24}<\min[m_2,a_{23}]$
$0<a_{25}<\min[a_6,m_1+a_{20},a_{11}-a_{12},a_{16}-a_{17},a_{21}-a_{22}]$
$0<a_{26}<\min[a_6,m_1+a_{19},m_1+a_{21}-a_{25},a_{11}-a_{12},a_{16}-a_{17},a_{21}-a_{22}]$
$0<a_{27}<\min[a_{25},a_{26},m_1+a_{22}]$
$0<a_{28}<\min[a_{27},m_1+a_{23}]$
$0<a_{29}<\min[a_{28},m_1+a_{24}]$
$0<a_{30}<\min[m_1,a_{29}]$

$$\mathfrak{g}=\mathfrak{g}_-\oplus\mathfrak{h}\oplus\mathfrak{g}_+.$$

Then $\mathfrak{b}=\mathfrak{h}\oplus\mathfrak{g}_+$ is a maximal solvable subalgebra (Borel subalgebra) of $\mathfrak{g}$. Let V_Λ be the representation space for the representation of the highest weight Λ and let $|\Lambda\rangle$ be a highest weight vector. For each $w\in W$, we denote by $w|\Lambda\rangle$ some nonzero vector in the (one-dimensional) weight space $V^{w\Lambda}$ of V_Λ and define the Schubert submodule $V_\Lambda(w)$ as the $\mathfrak{b}$-module generated by $w|\Lambda\rangle$. Thus $V_\Lambda(w)$ is obtained by applying to $w|\Lambda\rangle$ the raising operators e_i and operators from the Cartan subalgebra $\mathfrak{h}$. Evidently,

$$V_\Lambda(1)=\mathbb{C}(|\Lambda\rangle)\quad\text{and}\quad V_\Lambda(\text{inv})=V(\Lambda).$$

Thus we have the filtration

$$V_\Lambda(1)\subset V_\Lambda(r_{i_1})\subset V_\Lambda(r_{i_2}r_{i_1})\subset\cdots\subset V(\text{inv})=V_\Lambda.$$

TABLE XIV. Basis vectors and defining inequalities for an irreducible representation (m_1,m_2) of G_2.

$0<a_1<m_1$
$0<a_2<m_2+3a_1$
$0<a_3<\min[a_2,\tfrac{1}{3}(m_2+2a_2)]$
$0<a_4<\min[2a_3,\tfrac{1}{2}(m_2+3a_3)]$
$0<a_5<\min[\tfrac{1}{2}a_4,\tfrac{1}{3}(m_2+a_4)]$
$0<a_6<\min[a_5,m_2]$

Each Schubert submodule $V_\Lambda(w)$ is a sum of weight spaces and hence has a character $\text{Char}_\Lambda(w)$. An important fact is the Demazure character recursion formula: if α is a simple root, then

$$\text{Char}_\Lambda(r_\alpha w)=\Delta_\alpha(\text{Char}_\Lambda(w)):=\frac{\text{Char}_\Lambda(w)-\{r_\alpha\,\text{Char}_\Lambda(w)\}e^{-\alpha}}{1-e^{-\alpha}},$$

provided that $r_\alpha w$ is a longer word in the Weyl group than w.

The relevance of this to the Verma basis construction is that after the k th step we have a basis of $V_\Lambda(r_{i_k}\cdots r_{i_1})$. To understand what is happening as we pass from $V_\Lambda(w)$ to $V_\Lambda(r_\alpha w)$ it is necessary to look at what Verma calls SL_2 induction.

Let e,f,h be the basis (1) of $\mathfrak{sl}_2(\mathbb{C})$ over $\mathbb{C}$ and let $\mathfrak{b}=\mathbb{C}h+\mathbb{C}e$. Let M be a $\mathfrak{b}$-module of $\dim M<\infty$, which is a sum of $\mathfrak{h}$-weight spaces. Let $U(\mathfrak{sl}_2(\mathbb{C}))$ and $U(\mathfrak{b})$ be the universal enveloping algebras of $\mathfrak{sl}_2(\mathbb{C})$ and $\mathfrak{h}$, respectively. If we induce M to an $\mathfrak{sl}_2(\mathbb{C})$ module, we obtain

$$M'=U(\mathfrak{sl}_2(\mathbb{C}))\otimes_{U(\mathfrak{b})}M.$$

Here, M' has a unique minimal submodule J of finite codimension. We define

$$\text{Ind}_{\mathfrak{sl}_2(\mathbb{C})}(M)=M'/J.$$

Thus $\text{Ind}_{\mathfrak{sl}_2(\mathbb{C})}(M)$ is the maximal finite-dimensional quotient of M'. Given $D,U\in\mathbb{C}$ such that $(U-D)/\in\mathbb{N}$, there is a unique $\mathfrak{b}$-module

$$V^{D,U} = \mathbb{C}x + \mathbb{C}ex + \cdots + \mathbb{C}e^{1/2(U-D)}x,$$

such that

(i) $he^jx = (D+2j)e^jx, \quad j = 0,1,...,\frac{1}{2}(U-D)$;

(ii) $e^jx = 0, \quad \text{if } j > \frac{1}{2}(U-D)$;

(iii) $\dim V^{D,U} = \frac{1}{2}(U-D) + 1$.

Every indecomposable finite dimensional $\mathfrak{b}$-module is of this form. We let 1_A denote the one-dimensional module $V^{A,A}$ and V^L, $L\in N$, the familiar $V^{-L,L}$ which is even an $\mathfrak{sl}_2(\mathbb{C})$-module. Note that

$$V^{D,U}\otimes_{\mathbb{C}} 1_R \cong V^{D+R,U+R}.$$

In particular,

$$V^{D,U} \cong V^L \otimes 1_A, \quad \text{where } L = \tfrac{1}{2}(U-D), \quad A = \tfrac{1}{2}(U+D).$$

One shows that

$$\mathrm{Ind}_{\mathfrak{sl}_2(\mathbb{C})}(V^{D,U}) = \begin{cases} V^L\otimes V^A, & \text{where } L = \frac{1}{2}(U-D),\ A = \frac{1}{2}(U+D), \\ & \text{if } \frac{1}{2}(U+D)\in\mathbb{N} \text{ and } \frac{1}{2}(U-D)\in\mathbb{N}; \\ 0, & \text{otherwise.} \end{cases} \tag{34}$$

In the case that this induced module is not 0, it has the character

$$\begin{aligned} &\mathrm{Char}(\mathrm{ind}(V^{D,U})) \\ &= \{\mathrm{Char}(V^{D,U}) - \{r_\alpha\, \mathrm{Char}(V^{D,U})\}e^{-\alpha}\}/(1-e^{-\alpha}) \\ &= \Delta_\alpha(\mathrm{Char}(V^{D,U})). \end{aligned}$$

Here α is the simple root of $\mathfrak{sl}_2(\mathbb{C})$ corresponding to e. It follows that $\mathrm{Char}(\mathrm{Ind}_{\mathfrak{sl}(\mathbb{C})}(M)) = \Delta_\alpha(\mathrm{Char}(M))$ for any finite dimensional $\mathfrak{b}$-module in which all the indecomposable submodules satisfy the condition (34).

If we have a tensor product $M = V^{D_1U_1}\otimes\cdots\otimes V^{D_k,U_k}$ of indecomposable modules, each satisfying the nonvanishing condition (34), then

$$M \cong V^{D_1}\otimes\cdots\otimes V^{D_k}\otimes 1^A,$$

where $D_j = \frac{1}{2}(U_j - D_j)$ and $A = \frac{1}{2}\Sigma(U_j + D_j)$. Thus

$$\mathrm{Ind}_{\mathfrak{sl}_2(\mathbb{C})}(M) = V^{D_1}\otimes\cdots\otimes V^{D_k}\otimes V^A.$$

Notice that if $u_1,...,u_N$ is some basis of M then $\{f^ku_j|0<k<A,\ 1<j<N\}$ is a basis for $\mathrm{Ind}_{\mathfrak{sl}_2(\mathbb{C})}(M)$.

Now we return to the ascending sequence of Schubert modules. When we pass from $V_\Lambda(w)$ to $V_\Lambda(r_\alpha w)$ it is not hard to see that $V_\Lambda(r_\alpha w)$ is the $\mathfrak{sl}_2^{(\alpha)}(=\mathbb{C}f_\alpha + \mathbb{C}[e_\alpha f_\alpha] + \mathbb{C}e_\alpha)$-module in V_Λ generated by $V_\Lambda(w)$. Comparing the induction character formula and the Demazure formula shows that

$$V_\Lambda(r_\alpha w) = \mathrm{Ind}_{\mathfrak{sl}_2(\mathbb{C})}(V_\Lambda(w)).$$

Suppose that $V_\Lambda(w)$ has a basis of the form (20), where $0<a_j<U_j$ $(j=1,...,N)$ and each U_j is a function of $a_1,...,a_{j-1}$ (precisely as we have obtained in the Verma construction). Choose some *fixed* set of values for all a_j's for which $f_j\neq f_\alpha$. The remaining a_j's—say, $a_{j_1},...,a_{j_r}$—may still vary but they now have some lower constraints, just as in the examples of Sec. IV. Suppose that we find

$$\begin{aligned} &D_1<a_{j_1}<U_1 \\ &\quad\vdots \\ &D_r<a_{j_r}<U_r, \end{aligned} \tag{35}$$

where $D_1,...,D_r$, $U_1,...,U_r$ are *independent* of $a_{j_1},...,a_{j_r}$. As $a_{j_1},...,a_{j_r}$ vary, the corresponding basis vectors determine a basis for a subspace which apparently is a $\mathfrak{b}^\alpha(=\mathbb{C}h_\alpha + \mathbb{C}e_\alpha)$-module, M^α, at least modulo some submodule K. Furthermore M^α/K has the structure $V^{D_1,U_1}\otimes\cdots\otimes V^{D_r,U_r}\otimes 1_R$, where R is set to get the h_α-weight correct. When we induce M^α, we arrive at

$$\mathrm{Ind}(M^\alpha) = V^{1/2(U_1-D_1)} \otimes\cdots\otimes V^{1/2(U_r-D_r)}\otimes V^{R+(1/2\Sigma(U_i+D_i)},$$

with basis

$$f_\alpha^a f_{i_N}^{a_N}\cdots f_{i_2}^{a_2} f_{i_1}^{a_1}|\Lambda\rangle, \quad 0<a<R+\frac{1}{2}\sum(U_i+D_i).$$

This gives some credence to the Verma construction. The conditions (35) are exactly what we have seen are necessary to continue the construction.

VI. CONCLUSIONS AND COMMENTS

The inequalities calculated in this work allow one to write down a complete basis in any finite-dimensional space $V(\Lambda)$, irreducible with respect to a representation Λ of any of the Lie algebra/group of types A_n, $1<n$, B_n and C_n, $2<n<6$, D_n, $4<n<6$, and G_2. There apparently is no difficulty in computing the inequalities for ranks $n>6$ of any of the series of Lie algebras. However, we have failed so far in our attempts to derive the inequalities for F_4 and E_6.

A truly efficient construction of bases in large representation spaces irreducible with respect to a Lie algebra of high rank cannot follow the prescription above as it is. Indeed it would be impractical to write down thousands of basis vectors. Fortunately, it is hardly ever necessary as we have pointed out elsewhere.[3] It is advantageous to build the Verma basis in subspaces $V(\Lambda;\mu)$ with μ dominant and then to transform it if necessary to other subspaces with weights on the same Weyl group orbit using the "charge conjugation operators" of Ref. 3. The result is a major economy of efforts.

Finally let us compare the bases of Verma with the orthonormal bases of Gelfand and Zeitlin.[4,5] In the case of $\mathfrak{g} = A_n$, a basis vector (pattern) of Gelfand–Zeitlin coincides (up to a normalization) with that of Verma in the same space $V(\Lambda)$ only as long as the corresponding weight subspace $V(\Lambda;\mu)$ is one dimensional. When $\dim V(\Lambda;\mu)>1$, the two bases are in general different. The Gelfand–Zeitlin bases[5] for representations of the algebras B_n and D_n are different from those described here. There is no correspondence between any single basis vectors. The basis consists of vector patterns that are not eigenvectors of any Cartan subalgebra while those of this article are in the same way as in the case of A_n. That also is the reason why Gelfand–Zeitlin bases for representations of the Lie algebras of the orthogonal group have so far found only limited use in applications.

For completeness we present also the following beautiful (unpublished) result of Verma;[2] namely a uniform way of writing the basis-defining inequalities for a rank 2 Lie and Kac–Moody algebra of any type. Such an algebra is specified by its Cartan matrix

$$\begin{bmatrix} 2 & -A \\ -B & 2 \end{bmatrix}, \quad A,B \in \{1,2,...\}. \tag{36}$$

We ignore the elementary case $A = B = 0$ of the nonsimple algebra $A_1 \times A_1$. Cartan matrices of the algebras of rank 2 of this article are the particular cases of (34): A_2 ($A = B = 1$), B_2 or C_2 ($A = 2$, $B = 1$), and G_2 ($A = 3$, $B = 1$). The cases $A = B = 2$ and $A = 1$, $B = 4$ correspond to the affine Kac–Moody algebras. The rest are the Kac–Moody algebras of hyperbolic types; there are infinitely many of them of rank 2.

Given an irreducible representation (m_1,m_2) of $\mathfrak{g}$, a generic basis vector of the representation space is of the form

$$\cdots f_2^{a_k} f_1^{a_{k-1}} \cdots f_2^{a_4} f_1^{a_3} f_2^{a_2} f_1^{a_1} |m_1,m_2\rangle, \tag{37}$$

where the exponents a_i take all the values within the following inequalities:

$$\begin{aligned}
&0 \leqslant a_1 \leqslant m_1,\\
&0 \leqslant a_2 \leqslant m_2 + A a_1,\\
&\quad\vdots\\
&0 \leqslant a_{2i-1} \leqslant \min\left[a_{2i-2}\frac{\sqrt{B/A}\,\rho_{i-3/2}}{\rho_{i-2}}, \frac{a_{2i-2}\rho_{i-1} + m_2}{\sqrt{A/B}\,\rho_{i-3/2}} \right],\\
&\quad i \geqslant 2, \qquad\qquad (38)\\
&0 \leqslant a_{2i}\\
&\quad \leqslant \min\left[a_{2i-1}\frac{\rho_{i-1}}{\sqrt{B/A}\,\rho_{i-3/2}}, \frac{a_{2i-1}\sqrt{A/B}\,\rho_{i-1/2} + m_2}{\rho_{i-1}} \right],\\
&\quad\vdots
\end{aligned}$$

The coefficients ρ_i depend on the off-diagonal elements of the Cartan matrix (34). Putting $C = \sqrt{AB}$, one has

$$\begin{aligned}
\rho_0 &= 1\\
\rho_{1/2} &= c\\
\rho_1 &= -1 + c^2\\
\rho_{3/2} &= -2c + c^3\\
\rho_2 &= 1 + 3c^2 + c^4\\
\rho_{5/2} &= 3c - 4c^3 + c^5\\
\rho_3 &= -1 + 6c^2 - 5c^4 + c^6.\\
&\text{etc.}
\end{aligned} \tag{39}$$

ACKNOWLEDGMENTS

The authors wish to thank Professor D-N. Verma who both explained his method to us and provided us with an early outline of his ideas. They would also like to thank the Aspen Center for Physics for its hospitality.

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

[1] M. R. Bremner, R. V. Moody, and J. Patera, *Tables of Dominant Weight Multiplicities for Representation of Simple Lie Groups* (Marcel Dekker, New York, 1984).

[2] D-N. Verma, presented at the CMS Summer Seminar on Lie Theory and Related Topics, University of Windsor, Windsor, Ontario, Canada, 1984 (unpublished).

[3] R. V. Moody and J. Patera, "General charge conjugation operators in simple Lie groups," J. Math. Phys. **25**, 2838 (1984).

[4] I. M. Gelfand and M. L. Zeitlin, "Finite dimensional representations of the group of unimodular matrices," Dokl. Akad. Nauk SSSR **71**, 825 (1950); G. E. Baird and L. C. Biedenharn, J. Math Phys. **4**, 1449 (1963).

[5] I. M. Gelfand and M. L. Zeitlin, "Finite dimensional representations of the group of orthogonal matrices," Dokl. Akad. Nauk SSSR **71**, 1017 (1950).

Pure Lie algebraic approach to the modified Korteweg–de Vries equation

G. Post
Department of Applied Mathematics, Twente University of Technology, P. O. Box 217, 7500 AE Enschede, The Netherlands

(Received 30 May 1985; accepted for publication 31 October 1985)

It is explained, using only Lie algebraic means, how the modified-KdV-like equations arise. As an example, the modified-KdV equation is treated.

I. INTRODUCTION

Some years ago, the Kyoto school (Date, Jimbo, Kashiwara, Miwa, and others) explained how KdV-like equations arise, in the context of infinite-dimensional Lie algebras (see, for example, Refs. 1 and 2).

However, some aspects are not very transparent, in particular the construction of the equations. Kac[3] explained how these arise by considering the $\mathscr{g}'$-module $L(\Lambda_0) \otimes L(\Lambda_0)$. In the case of the KdV, $\mathscr{g} = A_1^{(1)}$. In this paper is explained, extending Kac's method, how one finds the modified equations by taking a $\mathscr{g}'$-module $L(\Lambda) \otimes L(\tilde{\Lambda})$, where, in general, $\Lambda \neq \tilde{\Lambda}$.

II. THE τ-FUNCTIONS AND EQUATIONS

Let us start with two integrable irreducible highest weight modules $L(\Lambda)$ and $L(\tilde{\Lambda})$ over a Kac–Moody algebra $\mathscr{g}'(A)$. We assume that A is a symmetrizable $(n+1)\times(n+1)$-matrix and $\mathscr{g}'(A)$ is generated by $e_0,...,e_n$ and $f_0,...,f_n$. We define $\alpha_i^\vee = [e_i, f_i]$.

The modules $L(\Lambda)$ and $L(\tilde{\Lambda})$ are completely determined by their labels $s_i = \langle \Lambda, \alpha_i^\vee \rangle$ and $\tilde{s}_i = \langle \tilde{\Lambda}, \alpha_i^\vee \rangle$.

Moreover, integrability requires that $s_i, \tilde{s}_i \in \mathbf{Z}_+$ (see Kac,[3] Lemma 10.1). The highest weight vector of $L(\Lambda)$ [resp. $L(\tilde{\Lambda})$] is denoted by v_1 (resp. v_2).

The principal gradation δ of $\mathscr{g}'(A)$ is defined by $\delta(e_i) = -\delta(f_i) = 1$, and the gradation δ_L of $L(\Lambda)$ or $L(\tilde{\Lambda})$ by $\delta_L(g\cdot v_i) = -\delta(g)$ [$i = 1,2, g \in \mathscr{g}'(A)$].

From these two modules we form the $\mathscr{g}'(A)$-module $L(\Lambda) \otimes L(\tilde{\Lambda})$. The $\mathscr{g}'(A)$-action is defined by

$$x(v\otimes w) = (xv)\otimes w + v \otimes (xw), \quad x\in\mathscr{g}'(A), \quad v\in L(\Lambda), \quad w \in L(\tilde{\Lambda}). \tag{2.1}$$

In general this module is not irreducible any more. But we know that it is completely reducible (Kac,[3] Corollary 10.7[b]). Because of this, the submodule generated by $v_1 \otimes v_2$ is irreducible. We denote this module by L_{high} and we see that $L_{\text{high}} \cong L(\Lambda + \tilde{\Lambda})$.

Further we introduce the following contravariant nondegenerate Hermitian form H:

$$H(v\otimes w, v'\otimes w') = H_1(v,v')\cdot H_2(w,w'), \tag{2.2}$$

where H_1 and H_2 are the unique contravariant nondegenerate Hermitian forms in $L(\Lambda)$ and $L(\tilde{\Lambda})$, satisfying $H_i(v_i,v_i) = 1$. Here H is (taken) linear in the second argument, antilinear in the first argument, and satisfies the contravariance condition

$$H(x\cdot u, u') = -H(u, \omega_0(x)\cdot u'), \tag{2.3}$$

where $x\in\mathscr{g}'(A)$, $u,u' \in L(\Lambda)\otimes L(\tilde{\Lambda})$, and ω_0 is the antilinear Cartan involution of $\mathscr{g}'(A)$ (cf. Kac,[3] § 11.5).

We define $L_{\text{low}} = L_{\text{high}}^{\perp}$, that is, L_{low} is the orthocomplement of L_{high} w.r.t. H. The contravariance (2.3) implies that L_{low} is a submodule of $L(\Lambda)\otimes L(\tilde{\Lambda})$ under the action of $\mathscr{g}'(A)$. We clearly have the direct sum of submodules:

$$L(\Lambda)\otimes L(\tilde{\Lambda}) = L_{\text{high}} \oplus L_{\text{low}}. \tag{2.4}$$

Next we introduce the τ-functions. Denote by G the group of automorphisms of $L(\Lambda)$ [or $L(\tilde{\Lambda})$] generated by $\exp(tf_i)$ and $\exp(te_i)$ ($i = 0,...,n$, $t\in\mathbf{C}$). The G-action defined by $g(v\otimes w) = (gv)\otimes(gw)$ is well defined in $L(\Lambda)\otimes L(\tilde{\Lambda})$. Of great importance is the fact that $g(v_1 \otimes v_2) \in L_{\text{high}}$ ($g\in G$). We denote $\tau_i(g) = g\cdot v_i$ ($i = 1,2$). Thus we find

$$\tau_1(g)\otimes\tau_2(g) \in L_{\text{high}}. \tag{2.5}$$

The following consequence is our central equation.

Theorem 1:

$$H(u, \tau_1(g)\otimes\tau_2(g)) = 0 \quad (\text{for all } u\in L_{\text{low}} \text{ and } g\in G). \tag{2.6}$$

III. HIROTA POLYNOMIALS

We assume that we have a realization of $L(\Lambda)$ and $L(\tilde{\Lambda})$ as polynomials, where the conditions

$$L(\Lambda) \cong \mathbf{C}[x_j^{(0)}, j\in E_+], \quad L(\tilde{\Lambda}) \cong \mathbf{C}[x_j^{(1)}, j \in E_+] \tag{3.1}$$

are satisfied, such that there exist p_i, q_j, $c \in \mathscr{g}'(A)$ ($i,j\in E_+$) with $[p_i, q_j] = \delta_{ij}\cdot c$ (all others zero) and

$$p_j \cdot v^{(i)} = \frac{\partial}{\partial x_j^{(i)}}\cdot v^{(i)}, \tag{3.2a}$$

$$q_j\cdot v^{(i)} = \alpha x_j^{(i)}\cdot v^{(i)}, \tag{3.2b}$$

$$c\cdot v^{(i)} = \alpha v^{(i)} \quad \text{and} \quad v_i = 1 (\in \mathbf{C}[x^{(i)}]). \tag{3.2c}$$

Here $v^{(i)} \in \mathbf{C}[x^{(i)}]$ (arbitrary), $\alpha \in \mathbf{Z}_+ \setminus \{0\}$, $E_+ \subset \mathbf{Z}_+\setminus\{0\}$. We denote the subalgebra $\langle p_i, q_j, c\rangle$ ($i,j\in E_+$) by $\mathscr{S}$, and we see that the $\mathscr{S}$-modules $L(\Lambda)$ and $L(\tilde{\Lambda})$ remain irreducible; applying q's to 1 generates any element of $\mathbf{C}[x^{(i)}]$, and applying p's brings any polynomial back to $1 \in \mathbf{C}[x^{(i)}]$.

In practice, we proceed from the other side.[4] Picking an $\mathscr{S}$, we investigate when the highest weight module $L(\Lambda)$ is irreducible as an $\mathscr{S}$-module. Considering $\mathbf{C}[x^{(i)}]$ as a $\mathscr{g}'(A)$-module, the elements of $\mathscr{g}'(A)$ are represented by differential operators of infinite order.

We make some more assumptions on $\mathscr{S}$. We require

$$\delta(p_i) = -\delta(q_i) = i \quad \text{and} \quad \omega_0(q_j) = -(1/j)p_j. \tag{3.3}$$

 0022-2488/86/030678-04$02.50

Clearly H_1 and H_2 are then given by

$$H_i(P(x_j^{(i)}),Q(x_j^{(i)})) = \overline{P}\left(\frac{1}{j}\frac{\partial}{\partial x_j^{(i)}}\right) Q(x_j^{(i)})\Big|_{x_j^{(i)}=0} \quad (i=0,1) . \tag{3.4}$$

So the monomials form an orthogonal basis of $\mathbb{C}[x^{(i)}]$ with square length given by

$$H(x_1^{k_1}\cdots x_s^{k_s}, x_1^{k_1}\cdots x_s^{k_s}) = \prod_{j=1}^{s} j^{-k_j}(k_j!) .$$

In order to derive the Hirota polynomials, we introduce a new set of variables by

$$2x_j = x_j^{(0)} + x_j^{(1)} , \tag{3.5a}$$

$$2y_j = x_j^{(0)} - x_j^{(1)} . \tag{3.5b}$$

One easily computes the action of $\mathscr{S}$ on $L(\Lambda) \otimes L(\widetilde{\Lambda})$:

$$p_j = \frac{\partial}{\partial x_j} , \tag{3.6a}$$

$$q_j = 2\alpha x_j , \tag{3.6b}$$

$$c = 2\alpha \cdot 1 . \tag{3.6c}$$

The space of Hirota polynomials is now defined by

$$\text{Hir} := L_{\text{low}} \cap \mathbb{C}[y] . \tag{3.7}$$

Now (3.6b) implies $L_{\text{low}} \supset \text{Hir} \otimes C[x]$ and (3.6a) implies $L_{\text{low}} \subset \text{Hir} \otimes \mathbb{C}[x]$, so $L_{\text{low}} = \text{Hir} \otimes \mathbb{C}[x]$.

Hence we deduce for Q free and $P \in \text{Hir}$:

$$H(Q(x_j)P(y_j), \tau_1(x^{(0)};g) \cdot \tau_2(x^{(1)};g)) = 0.$$

Using (3.4) and (3.5) this can be rewritten as

$$\overline{Q}\left(\frac{1}{j}\frac{\partial}{\partial x_j}\right)\overline{P}\left(\frac{1}{j}\frac{\partial}{\partial y_j}\right) \times (\tau_1(x_j+y_j;g)\cdot\tau_2(x_j-y_j;g))\Big|_{\substack{x=0\\y=0}} = 0 .$$

As Q is arbitrary we find the following theorem.

Theorem 2:

$$\overline{P}\left(\frac{1}{j}\frac{\partial}{\partial y_j}\right)(\tau_1(x_j+y_j;g)\cdot\tau_2(x_j-y_j;g))\Big|_{y=0} = 0 , \tag{3.8}$$

for all $P \in \text{Hir}$ and $g \in G$. We remark that τ_i also might be an element of some completion $\overline{G}$ of G; from an analytic point of view, these τ_i turn out to be more interesting.

IV. SOME EXAMPLES

One can wonder if there is anything left after the constraints (3.1)–(3.3). The answer is yes. Kac[3] describes the following classes.

Class 1: Let $\mathscr{S}$ be the principal Heisenberg subalgebra of an affine matrix $X_N^{(k)}$, which is symmetric if $k=1$. Then $L(\Lambda_0)$ satisfies our requirements, so we can look at $L(\Lambda_0) \otimes L(\Lambda_0)$. (See Kac,[3] Chap. 14.) (In general, Λ_i is given by $\langle \Lambda_i, \alpha_j^\vee \rangle = \delta_{ij}$.)

Class 2: Because of the symmetry of $A_l^{(1)}$, $D_l^{(1)}$, $E_6^{(1)}$, $E_7^{(1)}$, and $D_{l+1}^{(2)}$, we expect to remain irreducible as principal Heisenberg algebra module:

$A_l^{(1)}$ - $L(\Lambda_i)$ $(i=0,\dots,l)$,

$D_l^{(1)}$ - $L(\Lambda_0)$, $L(\Lambda_1)$, $L(\Lambda_{l-1})$, $L(\Lambda_l)$,

$E_6^{(1)}$ - $L(\Lambda_0)$, $L(\Lambda_1)$, $L(\Lambda_5)$,

$E_7^{(1)}$ - $L(\Lambda_0)$, $L(\Lambda_6)$,

$D_{l+1}^{(2)}$ - $L(\Lambda_0)$, $L(\Lambda_l)$.

(Here we have enumeration like Kac,[3] Chap. 4.) Equations corresponding to $L(\Lambda) \otimes L(\widetilde{\Lambda})$, with $\Lambda \neq \widetilde{\Lambda}$ are called modified equations. In the next sections we investigate $A_1^{(1)}$.

V. DESCRIPTION OF $A_1^{(1)}$

In fact we describe a realization of the derived algebra

$$A_1^{(1)\prime} \cong \text{sl}(2,\mathbb{C}) \otimes \mathbb{C}[t,t^{-1}] \oplus \mathbb{C}\cdot c ,$$

with generators $e_0 = f \otimes t$, $e_1 = e \otimes t$, $f_0 = e \otimes t^{-1}$, $f_1 = f \otimes t^{-1}$. Here e,f,h form a Chevalley basis for $\text{sl}(2,\mathbb{C})$.

We fix the bilinear nondegenerate invariant symmetric form $(\cdot|\cdot)$ on $\text{sl}(2,\mathbb{C})$ by $(h|h) = 2$, and define the bracket on $A_1^{(1)\prime}$ by

$$[g_1 \otimes P_1(t), g_2 \otimes P_2(t)] = [g_1,g_2] \otimes P_1(t)P_2(t) \oplus \frac{1}{2}\,\text{Res}\left(\frac{dP_1}{dt}\cdot P_2\right)(g_1|g_2)c \tag{5.1}$$

and

$$[c, g_1 \otimes P_1(t)] = 0 ,$$

$$\text{for } g_1,g_2 \in \text{sl}(2,\mathbb{C}), \quad P_1(t),P_2(t) \in \mathbb{C}[t,t^{-1}] .$$

Note that the principal gradation is given by $\delta(g \otimes t^k) = k$. The principal Heisenberg subalgebra $\mathscr{S}$ is defined by

$$p_j = (e+f) \otimes t^j ,$$

$$q_j = (1/j)(e+f) \otimes t^{-j}; \ c \quad (j \in \mathbb{Z}_+, j \text{ odd}) . \tag{5.2}$$

One immediately checks $[p_i,q_j] = \delta_{ij}\cdot c$ and $\omega_0(q_j) = -(1/j)p_j$. In this case $E_+ = \{j \in \mathbb{Z}_+ \mid j \text{ odd}\}$.

For future use we define

$$A(z) = \sum_{i\in\mathbb{Z}} z^{-i}(\overline{A}_i \otimes t^i) = \sum_{i\in\mathbb{Z}} A_i z^i ,$$

$$\text{with } \overline{A}_i = \begin{cases} h, & i \text{ even} , \\ f-e , & i \text{ odd} . \end{cases} \tag{5.3}$$

The homogenous components A_i are elements of $A_1^{(1)\prime}$ with $\delta(A_i) = -i$. Moreover A_i $(i \in \mathbb{Z})$ and p_i,q_j, and c form a basis for the vector space $A_1^{(1)\prime}$. A short calculation shows

$$[p_j, A(z)] = 2z^j A(z) , \quad [q_j, A(z)] = 2(z^{-j}/j)A(z) . \tag{5.4}$$

VI. IRREDUCIBLE $\mathscr{S}$-MODULES

We investigate for which Λ $L(\Lambda)$ remains irreducible considered as an $\mathscr{S}$-module. This can be determined by counting dimensions. Let $V_j(\mathbf{1})$ denote all elements of $L(\Lambda)$ of the form $g \cdot v_\Lambda$ with $\delta(g) = -j$ and v_Λ the highest weight vector, and let $\mathscr{g}_j'(r)$ denote all elements of $\mathscr{g}'(A)$ of degree j in the gradation $\hat{\delta}$, given by $\hat{\delta}(e_i) = r_i$, $r = (r_0,\dots,r_n)$. By $\mathbf{1}$ we denote the vector $(1,\dots,1)$. The following formula is valid:

$$\dim_q L(\Lambda) = \sum_{j\geqslant 0} \dim V_j(1) q^j$$

$$= \prod_{j\geqslant 1} (1-q^j)^{\dim {}^t\mathfrak{g}_j(s+1) - \dim \mathfrak{g}_j(1)} \quad (6.1)$$

Here ${}^t\mathfrak{g}(A)$ is the Kac–Moody algebra belonging to ${}^tA$: ${}^t\mathfrak{g}(A) = \mathfrak{g}({}^tA)$. In the case of $A_1^{(1)}$ we have ${}^t\mathfrak{g}(A_1^{(1)}) = \mathfrak{g}(A_1^{(1)})$, as $A_1^{(1)}$ is symmetric. Further $s = (s_0,...,s_n)$, where $s_i = \langle \Lambda, \alpha_i^\vee \rangle$ as before. Formula (6.1) is Proposition 10.10 of Kac.[3]

The irreducible $\mathscr{S}$-module $\hat{L}(\Lambda)$, isomorphic to polynomials in the variables x_j ($j\in E_+$) has q-dimension:

$$\sum_{j\geqslant 0} \dim \hat{V}_j(1) \cdot q^j = \prod_{j\geqslant 1} (1-q^j)^{-\dim \mathscr{S}_j(1)} \quad (6.2)$$

(cf. Kac,[3] Proposition 14.5). Requiring irreducibility we need $V_j(1) = \hat{V}_j(1)$:

$$\dim {}^t\mathfrak{g}_j(s+1) = \dim \mathfrak{g}_j(1) - \dim \mathscr{S}_j(1) \quad (j\geqslant 1). \quad (6.3)$$

In our case

$${}^tA_1^{(1)} = A_1^{(1)},$$

$$\dim \mathfrak{g}_j(1) = \begin{cases} 2, & j \text{ odd}, \\ 1, & j \text{ even}, \end{cases} \quad \dim \mathscr{S}_j(1) = \begin{cases} 1, & j \text{ odd}, \\ 0, & j \text{ even}. \end{cases}$$

That is, we need $\dim \mathfrak{g}_j(s+1) = 1$ ($j\geqslant 1$). In particular $\dim \mathfrak{g}_1(s+1) = 1$. We have two cases:

$$1: \begin{cases} s_0 > 0, \\ s_1 = 0, \end{cases} \quad 2: \begin{cases} s_0 = 0, \\ s_1 > 0. \end{cases}$$

Applying to $\dim \mathfrak{g}_2(s+1) = 1$, we find $s_0 = 1$ (resp. $s_1 = 1$). These values satisfy (6.3). They correspond to Λ_0 and Λ_1, respectively. We can describe these representations completely. First by (5.4) we find

$$\sigma_i(A(z)) = a_i \exp\Big(2\sum_{j\geqslant 1} z^j x_j^{(i)}\Big) \exp\Big(-2\sum_{j\geqslant 1} \frac{1}{j} z^{-j} \frac{\partial}{\partial x_j^{(i)}}\Big), \quad (6.4)$$

where σ_i is the representation on $L(\Lambda_i)$ and a_i is determined by Λ_i ($i = 0,1$). We have

$$\alpha_0^\vee = -h\otimes 1 + \tfrac{1}{2}c, \quad \alpha_1^\vee = h\otimes 1 + \tfrac{1}{2}c \quad (6.5)$$

or

$$c = \alpha_0^\vee + \alpha_1^\vee, \quad A_0 = h\otimes 1 = \tfrac{1}{2}(\alpha_1^\vee - \alpha_0^\vee). \quad (6.6)$$

So we find $\sigma_i(c)\cdot 1 = \langle \Lambda_i, c\rangle \cdot 1 = 1$ ($i = 0$ or 1) and $\sigma_i(A_0)\cdot 1 = \langle \Lambda_i, \tfrac{1}{2}\alpha_1^\vee - \tfrac{1}{2}\alpha_0^\vee\rangle = a_i \cdot 1$. Therefore,

$$\sigma_i(c) = \mathrm{Id} \quad (\alpha = 1) \quad \text{and} \quad a_0 = -\tfrac{1}{2}, \quad a_1 = \tfrac{1}{2}. \quad (6.7)$$

VII. THE KdV HIERARCHIES

Kac[3] shows that the Hirota polynomials belonging to $L(\Lambda_0)\otimes L(\Lambda_0)$ lead to the KdV hierarchy. It is clear (for example, by taking $\bar{A}_{i,\text{new}} = -\bar{A}_i$, which simply leads to $a_1 = -\tfrac{1}{2}$), that $L(\Lambda_1)\otimes L(\Lambda_1)$ has the same Hirota polynomials. There are two cases left, $L(\Lambda_0)\otimes L(\Lambda_1)$ and $L(\Lambda_1)\otimes L(\Lambda_0)$, which are also essentially the same. We pick $L(\Lambda_0)\otimes L(\Lambda_1)$, and start to count the q-dimension of Hir:

$$\dim_q \mathrm{Hir} = \sum_{j\geqslant 0} \dim H_j \cdot q^j \quad (7.1)$$

(where H_j consist of the elements of Hir of degree j). Moreover we have

$$\dim_q L(\Lambda_0)\otimes L(\Lambda_1) = \dim_q L(\Lambda_0)\cdot \dim_q L(\Lambda_1)$$

$$= \prod_{j\geqslant 1} (1-q^{2j-1})^{-2}, \quad (7.2)$$

$$\dim_q L_{\text{high}} = \dim_q L(\Lambda_0 + \Lambda_1)$$

$$= \prod_{j\geqslant 1} (1-q^{2j-1})^{-2} \cdot \prod_{j\geqslant 1} (1-q^{4j-2}), \quad (7.3)$$

so

$$\dim_q L_{\text{low}} = \prod_{j\geqslant 1} (1-q^{2j-1})^{-2} \cdot \Big\{1 - \prod_{j\geqslant 1} (1-q^{4j-2})\Big\}. \quad (7.4)$$

By $L_{\text{low}} = \mathrm{Hir}\otimes \mathbb{C}[x]$, we have

$$\dim_q L_{\text{low}} = \dim_q \mathrm{Hir}\cdot \prod_{j\geqslant 1} (1-q^{2j-1})^{-1}, \quad (7.5)$$

so

$$\dim_q \mathrm{Hir} = \prod_{j\geqslant 1} (1-q^{2j-1})^{-1} \Big\{1 - \prod_{j\geqslant 1} (1-q^{4j-2})\Big\}. \quad (7.6)$$

The q-dimension of the modified-KdV hierarchy has been calculated also by Sato and Mori.[5]

From (7.3) one calculates the q-dimension of

$$\text{Constraint:} = L_{\text{high}} \cap \mathbb{C}[y],$$

$$\dim_q \text{Constraint} = \prod_{j\geqslant 1} (1-q^{2j-1})^{-1} \cdot \prod_{j\geqslant 1} (1-q^{4j-2}), \quad (7.7)$$

and one finds Table I.

Of course Hir and Constraint are for each degree complementary orthogonal subspaces of $\mathbb{C}[y]$.

The easiest way to calculate Hir seems to be to calculate Constraint first and then to take the orthocomplement. This is explained, and done in the Appendix for degree $\leqslant 5$.

We consider Hir_2 and Hir_3, the subspaces of Hir of degree 2 and 3. They give rise to the following equations:

$$D_1^2(\tau_1(x+y)\tau_2(x-y))|_{y=0} = 0, \quad (7.8)$$

$$(D_1^3 - 4D_3)(\tau_1(x+y)\tau_2(x-y))|_{y=0} = 0. \quad (7.9)$$

Remember that τ_1 and τ_2 both satisfy the KdV hierarchy. We put

$$u = 2\frac{\partial^2}{\partial x_1^2}\log \tau_1 \quad \text{and} \quad v = \frac{\partial}{\partial x_1}\log \tau_2 - \frac{\partial}{\partial x_1}\log \tau_1. \quad (7.10)$$

TABLE I. Dimensions of $\mathbb{C}[y]$, Hir, and Constraint.

Degree	0	1	2	3	4	5	6	7
$\mathbb{C}[y]$	1	1	1	2	2	3	4	5
Hir	0	0	1	1	1	2	3	4
Constraint	1	1	0	1	1	1	1	1

Writing out (7.8) and dividing by $\tau_1 \cdot \tau_2$ yields

$$\frac{\tau_1''}{\tau} - 2\frac{\tau_1'}{\tau_1}\cdot\frac{\tau_2'}{\tau_2} + \frac{\tau_2''}{\tau_2} = 0 \quad \left(\text{where } ' = \frac{\partial}{\partial x_1} = \frac{\partial}{\partial x}\right).$$

Using

$$v^2 = \left(\frac{\tau_2'}{\tau_2}\right)^2 - 2\frac{\tau_2'}{\tau_2}\cdot\frac{\tau_2'}{\tau_1} + \left(\frac{\tau_1'}{\tau_1}\right)^2$$

and

$$v_x = \frac{\partial^2}{\partial x_1^2}\log\tau_2 - \frac{1}{2}u = \frac{\tau_2''}{\tau_2} - \left(\frac{\tau_2'}{\tau_2}\right)^2 - \frac{1}{2}u,$$

we find

$$\tfrac{1}{2}u + v^2 + (v_x + \tfrac{1}{2}u) = u + v_x + v^2 = 0, \qquad (7.11)$$

the Miura transformation.

With the same kind of manipulations, and with the use of (7.11), Eq. (7.9) leads to the modified KdV:

$$4v_t = v_{xxx} - 6v_x v^2, \qquad (7.12)$$

where $t = x_3$ and $x = x_1$.

APPENDIX: CALCULATION OF HIROTA POLYNOMIALS

Calculations can be done by using the definition of L_{low} and Hir:

$$\text{Hir} = \{P(y)\,|\,H(g\cdot 1, P(y)) = 0, \quad \text{for all } g \in U(\mathscr{g}'(A))\}. \qquad (\text{A1})$$

We choose the following basis for $U(\mathscr{g}'(A))$:

$$q_{l_4}^{\beta_{l_4}}\cdots q_1^{\beta_1}\cdot A_{l_3}^{\alpha_{l_3}}\cdots A_2^{\alpha_2}\cdot A_1^{\alpha_1}\cdot p_{l_2}^{\gamma_{l_2}}$$
$$\cdots p_1^{\gamma_1}\cdot A_{-l_1}^{\alpha_{-l_1}}\cdots A_{-1}^{\alpha_{-1}}\cdot A_0^{\alpha_0}\cdot c^{\gamma}. \qquad (\text{A2})$$

Then we find that

$$q_l^{\beta_l}\cdots q_1^{\beta_1} A_m^{\alpha_m}\cdots A_2^{\alpha_2} A_1^{\alpha_1}\cdot 1 \quad \text{span } L(\Lambda_0)\otimes L(\Lambda_1). \qquad (\text{A3})$$

We carelessly do not write the representation.

Working in this "basis," (A1) reads

$$\text{Hir} = \{P(y)\,|\,H(q_l^{\beta_l}\cdots q_1^{\beta_1}\cdot A_m^{\alpha_m}\cdots A_1^{\alpha_1}\cdot 1, P(y)) = 0\}. \qquad (\text{A4})$$

Moreover we know $\omega_0(q_i) = -(1/i)p_i$, but $p_i(P(y)) = 0$, so in (A4) we can take $\beta_1 = \cdots = \beta_l = 0$, and there remains

$$\text{Hir} = \{P(y)\,|\,H(A_m^{\alpha_m}\cdots A_1^{\alpha_1}\cdot 1, P(y)) = 0\}. \qquad (\text{A5})$$

From now on we take homogenous polynomials $P(y)$ of principal degree $\|\alpha\| = \alpha_1 + 2\alpha_2 + \cdots + m\cdot\alpha_m$, all other constraints in (A5) already being satisfied.

Using (6.4) and (6.7) the action of A_i on $L(\Lambda_0)\otimes L(\Lambda_1)$ is given by

$$A_i = \frac{1}{2}\sum_l \left\{p_{l+i}(2x^{(1)})p_l\left(-2\frac{\tilde\partial}{\partial x^{(1)}}\right)\right.$$
$$\left. - p_{l+i}(2x^{(0)})p_l\left(-2\frac{\tilde\partial}{\partial x^{(0)}}\right)\right\}, \qquad (\text{A6})$$

where $p_k(x)$ denotes the Schur polynomial fixed by

$$\sum p_k(x)z^k = \exp\left(\sum_{j>1} x_{2j-1}z^{2j-1}\right) \qquad (\text{A7})$$

and

$$\frac{\tilde\partial}{\partial x^{(i)}} = \left(\frac{\partial}{\partial x_1^{(i)}}, \frac{1}{2}\frac{\partial}{\partial x_2^{(i)}}, \frac{1}{3}\frac{\partial}{\partial x_3^{(i)}}, \ldots\right).$$

We are only interested in the components in $\mathbb{C}[y]$ of L_{high} and L_{low}, which we called Constraint and Hir. But, in general,

$$A_m^{\alpha_m}\cdots A_1^{\alpha_1}\cdot 1 = \sum_\beta x^\beta Q_\beta(y), \quad x^\beta = x_m^{\beta_m}\cdots x_1^{\beta_1} \qquad (\text{A8})$$

generates components outside Constraint, namely these parts of (A8) with $\beta \neq 0$.

But

$$L_{\text{high}} = \text{Constraint}\otimes\mathbb{C}[x]$$

so that all polynomials $Q_\beta(y)$ are in Constraint. We are only interested in the $Q_\beta(y)$ with degree $\|\alpha\|$, this is $Q_0(y)$. We write

$$A_m^{\alpha_m}\cdots A_1^{\alpha_1}\cdot 1 \sim Q_0(y),$$

where Q_0 is found by the following substitutions in (A6):

$$2x^{(1)} \to -2y, \quad 2x^{(0)} \to 2y,$$
$$-2\frac{\tilde\partial}{\partial x^{(1)}} \to \frac{\tilde\partial}{\partial y}, \quad -2\frac{\tilde\partial}{\partial x^{(0)}} \to -\frac{\tilde\partial}{\partial y},$$

which exactly cancel all terms containing x.

We calculate the Hirota polynomials of degree $\leqslant 5$: degree 0 and degree 1,

$$\text{Hir}_0 = \text{Hir}_1 = \{0\};$$

degree 2,

$$\text{Hir}_2 = \mathbb{C}[y]_2 = \langle y_1^2\rangle, \quad D_1^2(\tau_1\cdot\tau_2) = 0;$$

degree 3,

$$2A_3\cdot 1 \sim p_3(-2y) - p_3(2y) = -\tfrac{8}{3}y_1^3 - 4y_3,$$
$$\text{Hir}_3 = \langle y_1^3 - 12y_3\rangle, \quad (D_1^3 - 4D_3)(\tau_1\cdot\tau_2) = 0;$$

degree 4,

$$A_3\cdot A_1\cdot 1 = -2A_3\cdot y_1 \sim -p_3(-2y)y_1 - p_4(-2y)$$
$$+ p_3(2y)y_1 + p_4(2y) = \tfrac{4}{3}y_1^4 - 4y_1y_3,$$
$$\text{Hir}_4 = \langle y_1^4 + 24y_1y_3\rangle\cdot(D_1^4 + 8D_1D_3)(\tau_1\cdot\tau_2) = 0;$$

degree 5,

$$2A_5\cdot 1 \sim p_5(-2y) - p_5(2y) = -\tfrac{8}{15}y_1^5 - 8y_1^2y_3 - 4y_5,$$
$$\text{Hir}_5 = \langle y_1^5 + 24y_1^2y_3, y_1^5 + 80y_5\rangle,$$
$$(D_1^5 + 8D_1^2D_3)(\tau_1\cdot\tau_2) = 0,$$
$$(D_1^5 + 16D_5)(\tau_1\cdot\tau_2) = 0.$$

[1]E. Date, M. Kashiwara, M. Jimbo, and T. Miwa, "Transformation groups for soliton equations," in *Proceedings of RIMS Symposium on Non-Linear Integrable Systems—Classical Theory and Quantum Theory* (World Scientific, Singapore, 1983).
[2]M. Jimbo and T. Miwa, "Solitons and infinite dimensional Lie algebras," RIMS, Kyoto University **439** (1983).
[3]V. G. Kac, *Infinite Dimensional Lie Algebras* (Birkhäuser, Boston, 1983).
[4]J. Lepowsky and R. L. Wilson, "Construction of the affine Lie algebra $A_1^{(1)}$," Comm. Math. Phys. **62**, 43 (1978).
[5]M. Sato and Y. Mori, RIMS, Kôkyûroku **388**, 183 (1980).

Some integrals involving three modified Bessel functions. I

A. Gervois[a)] and H. Navelet
Service de Physique Théorique, CEN Saclay, 91191 Gif-sur-Yvette Cedex, France

(Received 30 May 1985; accepted for publication 31 October 1985)

The integrals $\int_0^\infty t^{1+\mu} I_\mu(at)K_\nu(bt)K_\nu(ct)dt$ and $\int_0^\infty t^{1-\nu} I_\nu(at)K_\nu(bt)K_\nu(ct)dt$ are calculated with the help of the factorization properties of the Appell function F_4. Results are given for real parameters a, b, c, both when they are and are not in a triangle configuration.

I. INTRODUCTION

Few results exist for integrals of products of three Bessel functions. A formal expression is the well-known formula[1]

$$\int_0^\infty t^{\lambda-1} J_\mu(at)J_\nu(bt)K_\rho(ct)dt$$
$$= 2^{\lambda-2}\frac{a^\mu b^\nu}{c^{\lambda+\mu+\nu}} \times \frac{\Gamma((\lambda+\mu+\nu+\rho)/2)\Gamma((\lambda+\mu+\nu-\rho)/2)}{\Gamma(\mu+1)\Gamma(\nu+1)} \times F_4\Big(\frac{\lambda+\mu+\nu+\rho}{2}, \frac{\lambda+\mu+\nu-\rho}{2},$$
$$\mu+1, \nu+1; -\frac{a^2}{c^2}, -\frac{b^2}{c^2}\Big), \tag{1.1}$$

for a complex a, b, c (parameters) and λ, μ, ν, ρ (indices) provided that

$$\mathrm{Re}(\lambda+\mu+\nu\pm\rho)>0, \tag{1.2}$$
$$\mathrm{Re}(c\pm ia\pm ib)>0. \tag{1.3}$$

Here F_4 is the Appell function[2] that is defined as a double series inside the domain $|a|+|b|<|c|$. Other integrals may be obtained using linear combinations of Bessel functions.

The practical interest of (1.1) is small as the double series is difficult to handle. Moreover, condition (1.3) is often too restrictive; for example, integrals like

$$\int_0^\infty t^{\lambda-1} J_\mu(at)J_\nu(bt)J_\rho(ct)dt \tag{1.4}$$

or

$$\int_0^\infty t^{\lambda-1} K_\mu(at)K_\nu(bt)K_\rho(ct)dt \tag{1.5}$$

exist for any *real* (*positive*) a, b, c, (and real λ, μ, ν, ρ) but the result cannot be reached easily by (1.1) as very little is known about the analytical continuation of F_4 (see Ref. 3), in particular when a, b, c may be considered as the sides of a triangle.

In two recent papers,[4,5] we showed that *for real (positive) a, b, c, and real λ, μ, ν, ρ*, integrals of the form (1.4) may be calculated by analytical continuation of (1.1) *when the function F_4 factorizes into products of hypergeometric functions ${}_2F_1$ of one variable only*; the nonanalyticity of ${}_2F_1$ is easily controlled as it reduces at most to a cut along the real axis. In this article, we extend the method to some integrals involving two K functions and an I function (and consequently to some integrals with three K functions). *Again $a, b, c, \lambda, \mu, \nu, \rho$ are real.*

Factorization possibilities for F_4 are listed in Refs. 1 and 4. Two factorizations are required for each integral; the integrals that can be calculated reduce to four.

(i) $$\int_0^\infty t^{1+\mu} I_\mu(at)K_\nu(bt)K_\nu(ct)dt \tag{1.6}$$

$(a<b+c,\quad 1+\mu-|\nu|>0)$.

(ii) $$\int_0^\infty t^{1-\nu} I_\nu(at)K_\nu(bt)K_\nu(ct)dt \tag{1.7}$$

$(a<b+c,\quad |\nu|<1)$.

(iii) Setting $\mu=-\nu$ in (1.6), we deduce

$$\int_0^\infty t^{1-\nu} K_\nu(at)K_\nu(bt)K_\nu(ct)dt, \tag{1.8}$$

any real $a, b, c>0,\quad -1<\nu<\frac{1}{2}$,

an integral that appears in the calculation of universal numbers in polymer theory.[6] Its calculation was our first motivation for this work.

(iv) $$\int_0^\infty I_\mu(at)K_\nu(bt)K_\rho(ct)dt, \tag{1.9}$$

$a<b+c$,

for any real μ, ν, ρ (here $\lambda=1$).

Integrals (1.6) and (1.9) were obtained by Bailey[7] for $c>a+b$. We give here the complete results for (1.6)–(1.8) both in nontriangle and triangle configurations. As to integral (1.9), it is also possible to reduce it to a sum of products of hypergeometric functions ${}_2F_1$, but the expressions are simple only when $\mu=\pm\nu$ and $a=b$. Calculations differ in detail but are similar to those of Ref. 5 and will be given in a forthcoming paper.[8]

This article is organized as follows. In the main section (Sec. II), we derive (1.6) in the nontriangle and triangle configurations. Integrals (1.7) and (1.8) are more briefly calculated in Sec. III. Results are collected in Tables I and II.

II. CALCULATION OF INTEGRAL (1.6)

We use the definitions

$$K_\nu = (\pi/2\sin\pi\nu)(I_{-\nu}-I_\nu),\quad \nu \text{ is not an integer}, \tag{2.1}$$
$$I_\nu(x) = e^{-i\nu\pi}J_\nu(e^{i\pi/2}x),$$

to rewrite the integral

$$\mathscr{L}_\mu = \int_0^\infty t^{1+\mu} I_\mu(at)K_\nu(bt)K_\nu(ct)dt$$

[a)] Chercheur Centre National de la Recherche Scientifique.

as

$$\mathcal{L}_\mu = \frac{\pi}{2\sin\pi\nu}\int_0^\infty t^{1+\mu}e^{-i\mu\pi/2}J_\mu(e^{i\pi/2}at) \times[e^{i\nu\pi/2}J_{-\nu}(e^{i\pi/2}bt) - e^{-i\nu\pi/2}J_\nu(e^{i\pi/2}bt)] \times K_\nu(ct)dt,$$

or, from (1.1) and provided that separately each integral has a meaning (i.e., $a+b<c$),

$$\begin{aligned}\mathcal{L}_\mu &= \frac{\pi}{2\sin\pi\nu}\frac{2^\mu a^\mu}{c^{2+2\mu}}\Big\{\Big(\frac{c}{b}\Big)^\nu\frac{\Gamma(1+\mu-\nu)}{\Gamma(1-\nu)} \\ &\quad\times F_4\Big(1+\mu,1+\mu-\nu;1+\mu,1-\nu;\frac{a^2}{c^2},\frac{b^2}{c^2}\Big);-\Big(\frac{b}{c}\Big)^\nu\frac{\Gamma(1+\mu+\nu)}{\Gamma(1+\nu)} \\ &\quad\times F_4\Big(1+\mu,1+\mu+\nu;1+\mu,1+\nu;\frac{a^2}{c^2},\frac{b^2}{c^2}\Big)\Big\} \\ &= \frac{\pi}{2\sin\pi\nu}\frac{2^\mu a^\mu}{c^{2+2\mu}}\Big\{\frac{\Gamma(1+\mu-\nu)}{\Gamma(1-\nu)} \times\Big(\frac{c}{b}\Big)^\nu(1-x)^{1+\mu-\nu} \\ &\quad\times {}_2F_1\Big(1+\mu-\nu,1+\mu;1-\nu;-\frac{y(1-x)}{1-y}\Big) -\Big(\frac{b}{c}\Big)^\nu\frac{\Gamma(1+\mu+\nu)}{\Gamma(1+\nu)}(1-x)^{1+\mu+\nu} \\ &\quad\times {}_2F_1\Big(1+\mu+\nu,1+\mu;1+\nu;-\frac{y(1-x)}{1-y}\Big)\Big\},\end{aligned} \tag{2.2}$$

where we took into account the factorization of F_4 (see Ref. 1)

$$F_4\Big(\alpha,\beta;1+\alpha-\beta,\beta;\frac{-x}{(1-x)(1-y)},\frac{-y}{(1-x)(1-y)}\Big) = (1-y)^\alpha\,{}_2F_1\Big(\alpha,\beta;1+\alpha-\beta;-\frac{x(1-y)}{1-x}\Big). \tag{2.3}$$

Parameters x, y are given by the transformation

$$-\frac{x}{(1-y)(1-x)} = \frac{a^2}{c^2},\quad -\frac{y}{(1-x)(1-y)} = \frac{b^2}{c^2} \tag{2.4a}$$

or

$$x = \frac{b^2+a^2-c^2-4\tilde\Delta}{2b^2},\quad y = \frac{b^2+a^2-c^2-4\tilde\Delta}{2c^2}, \tag{2.4b}$$

where

$$4\tilde\Delta = \delta^{1/2} \tag{2.4c}$$

$$\delta = [(b+a)^2-c^2][(b-a)^2-c^2], \tag{2.4d}$$

and $\delta^{1/2}$ is the positive determination of the square root so that x, y go to zero for large c.

A. NONTRIANGLE CONFIGURATION: $c>a+b$

We introduce the hyperbolic "angles" u_a, u_b, u_c and the "area" $\tilde\Delta$ such that

$$\begin{aligned}c^2 &= a^2+b^2+2ab\cosh u_c,\\ a^2 &= b^2+c^2-2bc\cosh u_a,\\ b^2 &= c^2+a^2-2ac\cosh u_b,\end{aligned} \tag{2.5a}$$

$$u_c = u_a + u_b, \tag{2.5b}$$

$$\tilde\Delta = \tfrac12 ab\sinh u_c = \tfrac12 bc\sinh u_a = \tfrac12 ca\sinh u_b. \tag{2.5c}$$

Then,

$$\begin{aligned}1-x &= (c/b)e^{-u_a},\\ 1-y &= (c/a)e^{-u_b},\\ -x(1-y)/(1-x) &= e^{-2u_b},\\ -y(1-x)/(1-y) &= e^{-2u_a},\end{aligned} \tag{2.5d}$$

and expression (2.2) is rewritten as

$$\begin{aligned}\mathcal{L}_\mu &= \frac{\pi}{2\sin\pi\nu}\frac{2^\mu a^\mu}{(bc)^{\mu+1}}\Big\{\frac{\Gamma(1+\mu-\nu)}{\Gamma(1-\nu)}e^{-u_a(1+\mu-\nu)} \times{}_2F_1(1+\mu-\nu,1+\mu;1-\nu;e^{-2u_a}) \\ &\quad -\frac{\Gamma(1+\mu+\nu)}{\Gamma(1+\nu)}e^{-u_a(1+\mu+\nu)} \times{}_2F_1\big((1+\mu+\nu,1+\mu;1+\nu;e^{-2u_a}\big)\Big\}.\end{aligned} \tag{2.6}$$

The main point is now to reduce the hypergeometric ${}_2F_1$ to Legendre functions outside the cut. Formulas may be found in Ref. 9. We use successively

$$e^{-u(1+\mu\mp\nu)}\frac{\Gamma(1+\mu\mp\nu)}{\Gamma(1\mp\nu)}\,{}_2F_1(1+\mu\mp\nu,1+\mu;1\mp\nu;e^{-2u}) = (e^{-i\pi(\mu+1/2)}/\sqrt{\pi})2^{-1/2-\mu}(\sinh u)^{-1/2-\mu} \times Q^{\mu+1/2}_{-1/2\mp\nu}(\cosh u), \tag{2.7a}$$

$$Q^{\mu+1/2}_{-1/2\mp\nu}(\cosh u) = \frac{\pi}{2}\frac{e^{i\pi(\mu+1/2)}}{\sin\pi(\mu+1/2)}\Big[P^{\mu+1/2}_{\nu-1/2}(\cosh u) - \frac{\Gamma(\mu\mp\nu+1)}{\Gamma(-\mu\mp\nu)}P^{-\mu-1/2}_{\nu-1/2}(\cosh u)\Big], \tag{2.7b}$$

and, after some manipulations,

$$\mathcal{L}_\mu = \frac12\sqrt{\frac{\pi}{2}}\Big(\frac{a}{bc}\Big)^\mu\frac{\Gamma(1+\mu+\nu)\Gamma(1+\mu-\nu)}{bc} \times(\sinh u_a)^{-\mu-1/2}P^{-\mu-1/2}_{\nu-1/2}(\cosh u_a),$$

to be compared to a similar result for $\int t^{1+\mu}J_\mu K_\nu K_\nu$ in Ref. 10. In a more symmetric way, using (2.5c), we get

$$\mathcal{L}_\mu = 2^{-2\mu-2}\sqrt{\frac{\pi}{2}}\frac{(abc)^\mu}{\tilde\Delta^{2\mu+1}} \times\Gamma(1+\mu+\nu)\Gamma(1+\mu-\nu)(\sinh u_a)^{\mu+1/2} \times P^{-\mu-1/2}_{\nu-1/2}(\cosh u_a) \tag{2.8}$$

$[\tilde\Delta = \tfrac12 bc\sinh u_a,\quad$ see Eq. (2.5c)$]$.

For $\mu=-\nu$, Eq. (2.8) becomes

$$\mathscr{L}_{-\nu} = 2^{2\nu-2}\left(\frac{\pi}{2}\right)^{1/2}\frac{\widetilde{\Delta}^{2\nu-1}}{(abc)^{\nu}}$$
$$\times\Gamma(1-2\nu)(\sinh u_a)^{1/2-\nu}P^{\nu-1/2}_{\nu-1/2}(\cosh u_a). \tag{2.9}$$

Both results (2.8) and (2.9) are written in Table I, lines 1 and 2.

Result (2.8) seems very far from that of Bailey,[7] which is written with the help of artificial angles φ, ϕ, and ψ. Setting

$$a = c'\sinh\varphi, \quad b = c'\sinh\phi, \quad c = c'\cosh\varphi\cosh\phi,$$

we get

$$\cosh u_a = 1/\tanh\psi, \quad \sinh u_a = 1/\sinh\psi.$$

The P Legendre function is turned into a Q function, with the help of the Whipple formula,[9]

$$P^{-\mu-1/2}_{-\nu-1/2}\left(\frac{1}{\tanh\psi}\right) = \frac{i}{\Gamma(1+\mu+\nu)}e^{-i\pi(\nu+1/2)}$$
$$\times\sqrt{\frac{2}{\pi}}\sinh\psi\, Q^{\nu}_{\mu}(\cosh\psi),$$

and we get the final expression

$$\mathscr{L}_{\mu} = 2^{\mu}(\sinh\varphi)^{\mu}\left(\frac{\cosh^2\psi/2}{\cosh^2\phi}\right)^{\mu+1}\frac{\Gamma(1+\mu-\nu)}{c'^{2+\mu}}$$
$$\times e^{-i\pi\nu}Q^{\nu}_{\mu}(\cosh\psi). \tag{2.10}$$

It differs from result (3.5) of Ref. 7, where $e^{-i\pi\nu}$ is replaced by $\sin\mu\pi/\sin(\mu+\nu)\pi$. The result of Ref. 7 corresponds to a nonstandard definition of the Q_μ (Barnes notation, See Ref. 11).

Notice that results (2.8)–(2.10) hold when going to the limit where ν is an integer.

B. TRIANGLE CONFIGURATION: $|a-b|<c<a+b$

Now, the integrals $\int t^{1+\mu}I_\mu I_{\pm\mu}K_\nu$ do not separately converge, but their difference does. As the result is an analytical function of parameters a, b, c, the expression (2.2) has an analytical continuation in the region $|a-b|<c<a+b$, which can be obtained from that of the hypergeometric series. In other words, the two hypergeometric

$${}_2F_1\left(1+\mu\mp\nu, 1+\mu; 1\mp\nu; -y\frac{(1-x)}{(1-y)}\right)$$

still exist although they no longer correspond to any integral $\int t^{1+\mu}I_\mu I_{\pm\nu}K_\nu$. They are now complex quantities but their difference, which remains proportional to integral (1.6), is real.

Parameters x, y defined by (2.4a) are now complex numbers

$$x = (b^2+a^2-c^2-4i\Delta)/2b^2,$$
$$y = (b^2+a^2-c^2-4i\Delta)/2c^2, \tag{2.11a}$$

where

$$4\Delta = \sqrt{-\delta} \tag{2.11b}$$

is the positive determination of the square root.

Introducing the angles φ_a, φ_b, φ_c and the area Δ of the triangle,

$$a^2 = b^2+c^2-2bc\cos\varphi_a,$$
$$b^2 = c^2+a^2-2ac\cos\varphi_b, \tag{2.12a}$$
$$c^2 = a^2+b^2-2ab\cos\varphi_c,$$

$$\pi = \varphi_a+\varphi_b+\varphi_c, \tag{2.12b}$$
$$\Delta = \tfrac{1}{2}ab\sin\varphi_c = \tfrac{1}{2}bc\sin\varphi_a = \tfrac{1}{2}ca\sin\varphi_b, \tag{2.12c}$$

we have now

$$1-x = (c/b)e^{-i\varphi_a}, \quad 1-y = (c/a)e^{-i\varphi_b},$$
$$\frac{-x(1-y)}{1-x} = e^{-2i\varphi_b}, \quad \frac{-y(1-x)}{1-y} = e^{-2i\varphi_a}, \tag{2.12d}$$

and the analytical continuation of (2.2) is

$$\mathscr{L}_{\mu} = \frac{\pi}{2\sin\pi\nu}\frac{2^{\mu}a^{\mu}}{c^{2+2\mu}}\left(\frac{c}{b}\right)^{1+\mu}\left\{\frac{\Gamma(1+\mu-\nu)}{\Gamma(1-\nu)}\right.$$
$$\times e^{-i\varphi_a(1+\mu-\nu)}{}_2F_1(1+\mu-\nu,1+\mu;1-\nu;\ e^{-2i\varphi_a})$$
$$-\frac{\Gamma(1+\mu+\nu)}{\Gamma(1+\nu)}e^{-i\varphi_a(1+\mu+\nu)}$$
$$\left.+{}_2F_1\left(1+\mu+\nu, 1+\mu; 1+\nu; e^{-2i\varphi_a}\right)\right\}. \tag{2.13}$$

TABLE I. List of the integrals calculated in the paper for $c>a+b$. The $P^{-\mu-1/2}_{\nu-1/2}(\cosh u)$ are the Legendre functions outside the cut; a, b, c, λ, μ, ν, ρ are all real.

$$\int_0^\infty t^{1+\mu}I_\mu(at)K_\nu(bt)K_\nu(ct)dt = 2^{-2\mu-2}\sqrt{\frac{\pi}{2}}\frac{\Gamma(1+\mu+\nu)\Gamma(1+\mu-\nu)}{\widetilde{\Delta}^{2\mu+1}}(abc)^{\mu}(\sinh u_a)^{\mu+1/2}P^{-\mu-1/2}_{\nu-1/2}(\cosh u_a), \quad 1+\mu-|\nu|>0;$$

$$\int_0^\infty t^{1-\nu}I_{-\nu}(at)K_\nu(bt)K_\nu(ct)dt = 2^{2\nu-2}\sqrt{\frac{\pi}{2}}\Gamma(1-2\nu)\frac{\widetilde{\Delta}^{2\nu-1}}{(abc)^{\nu}}(\sinh u_a)^{1/2-\nu}P^{\nu-1/2}_{\nu-1/2}(\cosh u_a), \quad \nu<\tfrac{1}{2};$$

$$\int_0^\infty t^{1-\nu}I_\nu(at)K_\nu(bt)K_\nu(ct)dt = 2^{2\nu-2}\sqrt{\frac{\pi}{2}}\Gamma(1-2\nu)\frac{\widetilde{\Delta}^{2\nu-1}}{(abc)^{\nu}}\{(\sinh u_c)^{1/2-\nu}P^{\nu-1/2}_{\nu-1/2}(\cosh u_c)-(\sinh u_b)^{1/2-\nu}P^{\nu-1/2}_{\nu-1/2}(\cosh u_b)\}, \quad |\nu|<1;$$

$$\int_0^\infty t^{1-\nu}K_\nu(at)K_\nu(bt)K_\nu(ct)dt$$
$$= \frac{\pi}{\sin\pi\nu}2^{2\nu-3}\sqrt{\frac{\pi}{2}}\Gamma(1-2\nu)\frac{\widetilde{\Delta}^{2\nu-1}}{(abc)^{\nu}}$$
$$\times\{(\sinh u_a)^{1/2-\nu}P^{\nu-1/2}_{\nu-1/2}(\cosh u_a)+(\sinh u_b)^{1/2-\nu}P^{\nu-1/2}_{\nu-1/2}(\cosh u_b)-(\sinh u_c)^{1/2-\nu}P^{\nu-1/2}_{\nu-1/2}(\cosh u_c)\},$$
$$\nu\neq 0, \quad -1<\nu<\tfrac{1}{2};$$

$$c^2 = a^2+b^2+2ab\cosh u_c, \quad b^2 = a^2+c^2-2ac\cosh u_b, \quad a^2 = b^2+c^2-2bc\cosh u_a, \quad u_c = u_a+u_b, \quad \widetilde{\Delta} = \tfrac{1}{2}ab\sinh u_c = \tfrac{1}{2}b_c\sinh u_a = \tfrac{1}{2}ca\sinh u_b.$$

The following proceeds as in Sec. II A, with some supplementary care as we deal with complex numbers. We indicate some intermediate steps for $e^{-i\varphi(1+\mu-\nu)}{}_2F_1(1+\mu-\nu;1-\nu;e^{-2i\varphi})$, which we intend to express in terms of the Legendre function *on* the cut. Again, all formulas are in Ref. 9. We have successively

$$e^{-i\varphi(1+\mu-\nu)}\,{}_2F_1(1+\mu-\nu,1+\mu;1-\nu;e^{-2i\varphi}) = [e^{i\varphi(1+\mu-\nu)}\,{}_2F_1(1+\mu-\nu,1+\mu;1-\nu;e^{2i\varphi})]^*, \tag{2.14}$$

where * denotes the complex conjugate, then[9]

$${}_2F_1(1+\mu-\nu,1+\mu;1-\nu;e^{2i\varphi}) = \Gamma(1-\nu)e^{i\nu\varphi}(1-e^{2i\varphi})^{-1-\mu}P^{\nu}_{-1-\mu}(i\cot\varphi),$$

where the argument of $(1-e^{2i\varphi})$ is $\varphi-\pi/2$, then[12]

$$P^{-\nu}_{-1-\mu}(i\cot\varphi) = \sqrt{\frac{2}{\pi}\sin\varphi}\;\frac{e^{i\pi\mu}e^{i\pi/4}}{\Gamma(-\mu-\nu)}Q^{-\mu-1/2}_{-\nu-1/2}(\cos\varphi-i0)$$

and[9]

$$Q^{-\mu-1/2}_{-\nu-1/2}(\cos\varphi-i0) = e^{-(i\pi/2)\mu}e^{-i\pi/4}[Q^{-\mu-1/2}_{-\nu-1/2}(\cos\varphi) + (i\pi/2)P^{-\mu-1/2}_{-\nu-1/2}(\cos\varphi)],$$

where the last two functions are Legendre functions on the cut. We have the intermediate result

$$\frac{\Gamma(1+\mu\mp\nu)}{\Gamma(1\mp\nu)}e^{-i\varphi(1+\mu\mp\nu)}\times{}_2F_1(1+\mu\mp\nu,1+\mu;1\mp\nu;e^{-2i\varphi}) = \frac{\Gamma(1+\mu\mp\nu)}{\Gamma(-\mu\mp\nu)}(2\sin\varphi)^{-1-\mu}\sqrt{\frac{2}{\pi}\sin\varphi}\;e^{-i\pi(\mu+1/2)}\times[Q^{-\mu-1/2}_{\mp\nu-1/2}(\cos\varphi)-(i\pi/2)P^{-\mu-1/2}_{\mp\nu-1/2}(\cos\varphi)]. \tag{2.15}$$

The difference (2.13) may be rewritten only in terms of $P^{-\mu-1/2}_{\nu-1/2}$ by using all the relations[9] between $P^{-\mu-1/2}_{\mp\nu-1/2}$ and $Q^{-\mu-1/2}_{\mp\nu-1/2}$. As expected, the imaginary part disappears and we finally get

$$\mathscr{L}_\mu = \frac{1}{2}\sqrt{\frac{\pi}{2}}\,\frac{\Gamma(1+\mu-\nu)\Gamma(1+\mu+\nu)}{bc}\Big(\frac{a}{bc}\Big)^\mu\times(\sin\varphi_a)^{-1/2-\mu}P^{-\mu-1/2}_{\nu-1/2}(\cos\varphi_a)$$

$$\mathscr{L}_\mu = 2^{-2\mu-2}\sqrt{\frac{\pi}{2}}\Gamma(1+\mu-\nu)\Gamma(1+\mu+\nu)\frac{(abc)^\mu}{\Delta^{2\mu+1}}\times P^{-\mu-1/2}_{\nu-1/2}(\cos\varphi_a)\times(\sin\varphi_a)^{1/2+\mu}, \tag{2.16}$$

and, for $\mu=-\nu$,

$$\mathscr{L}_{-\nu} = 2^{2\nu-2}\Big(\frac{\pi}{2}\Big)^{1/2}\Gamma(1-2\nu)\frac{\Delta^{2\nu-1}}{(abc)^\nu}\times P^{\nu-1/2}_{\nu-1/2}(\cos\varphi_a)\times(\sin\varphi_a)^{1/2-\nu}. \tag{2.17}$$

These formulas are very similar to the ones obtained for the nontriangle configuration [Eqs. (2.8) and (2.9)] and are reported on lines 1 and 2 in Table II.

Result (2.17) can be checked in a different way. Expanding $I_{-\nu}(at)$, we have

$$\mathscr{L}_\mu = \sum_{m=0}^{\infty}\frac{1}{m!}\frac{1}{\Gamma(1+m-\nu)}\times\int_0^\infty\Big(\frac{a}{2}t\Big)^{2m-\nu}t^{1-\nu}K_\nu(bt)K_\nu(ct)dt$$

$$= \frac{1}{2}\sum_{m=0}^{\infty}\frac{1}{m!}\frac{1}{\Gamma(1+m-\nu)}\Big(\frac{bc}{2a}\Big)^\nu\Big(\frac{a}{c}\Big)^{2m}\times\frac{\Gamma(1+m)\Gamma^2(1+m-\nu)\Gamma(1+m-2\nu)}{\Gamma(2+2m-2\nu)}\times{}_2F_1\Big(1+m,1+m-\nu;2+2m-2\nu;1-\frac{b^2}{c^2}\Big)$$

(see Ref. 9, p. 101). Now, with duplication formulas[13]

$$\Gamma(2m-2\nu+2) = (1/\sqrt{\pi})2^{2m-2\nu+1}\Gamma(m-\nu+1)\Gamma(m+\tfrac{3}{2}-\nu)$$

and

$${}_2F_1\Big(1+m,1+m-\nu;2+2m-2\nu;1-\frac{b^2}{c^2}\Big)$$

TABLE II. The same as in Table I for a triangle configuration ($|a-b|<c<a+b$). The $P^{-\mu-1/2}_{\nu-1/2}(\cos u)$ are the Legendre functions on the cut.

$$\int_0^\infty t^{1+\mu}I_\mu(at)K_\nu(bt)K_\nu(ct)dt = 2^{-2\mu-2}\sqrt{\frac{\pi}{2}}\Gamma(1+\mu+\nu)\Gamma(1+\mu-\nu)\frac{(abc)^\mu}{\Delta^{2\mu+1}}(\sin\varphi_a)^{\mu+1/2}P^{-\mu-1/2}_{\nu-1/2}(\cos\varphi_a),\quad 1+\mu-|\nu|>0;$$

$$\int_0^\infty t^{1-\nu}I_{-\nu}(at)K_\nu(bt)K_\nu(ct)dt = 2^{2\nu-2}\sqrt{\frac{\pi}{2}}\Gamma(1-2\nu)\frac{\Delta^{2\nu-1}}{(abc)^\nu}(\sin\varphi_a)^{1/2-\nu}P^{\nu-1/2}_{\nu-1/2}(\cos\varphi_a),\quad \nu<\tfrac{1}{2};$$

$$\int_0^\infty t^{1-\nu}I_\nu(at)K_\nu(bt)K_\nu(ct)dt = \Big(\frac{\pi}{2}\Big)^{3/2}2^{\nu-1/2}\frac{\Gamma(1/2-\nu)}{\pi}\frac{\Delta^{2\nu-1}}{(abc)^\nu} - 2^{2\nu-2}\sqrt{\frac{\pi}{2}}\Gamma(1-2\nu)\frac{\Delta^{2\nu-1}}{(abc)^\nu}\times\{(\sin\varphi_b)^{1/2-\nu}P^{\nu-1/2}_{\nu-1/2}(\cos\varphi_b)+(\sin\varphi_c)^{1/2-\nu}P^{\nu-1/2}_{\nu-1/2}(\cos\varphi_c)\},\quad |\nu|<1;$$

$$\int_0^\infty t^{1-\nu}K_\nu(at)K_\nu(bt)K_\nu(ct)dt = -\Big(\frac{\pi}{2}\Big)^{5/2}2^{\nu-1/2}\frac{\Gamma(1/2-\nu)}{\pi\sin\pi\nu}\frac{\Delta^{2\nu-1}}{(abc)^\nu} + \frac{\pi}{\sin\pi\nu}2^{2\nu-3}\sqrt{\frac{\pi}{2}}\Gamma(1-2\nu)\frac{\Delta^{2\nu-1}}{(abc)^\nu}\{(\sin\varphi_a)^{1/2-\nu}P^{\nu-1/2}_{\nu-1/2}(\cos\varphi_a) + (\sin\varphi_b)^{1/2-\nu}P^{\nu-1/2}_{\nu-1/2}(\cos\varphi_b)+(\sin\varphi_c)^{1/2-\nu}P^{\nu-1/2}_{\nu-1/2}(\cos\varphi_c)\},\quad \nu\neq0,\quad -1<\nu<\tfrac{1}{2};$$

$a^2=b^2+c^2-2bc\cos\varphi_a$, $b^2=c^2+a^2-2ac\cos\varphi_b$, $c^2=a^2+b^2-2ab\cos\varphi_c$,

$\pi=\varphi_a+\varphi_b+\varphi_c$, $\Delta=\frac{1}{2}ab\sin\varphi_c=\frac{1}{2}bc\sin\varphi_a=\frac{1}{2}ca\sin\varphi_b$.

$$= \left(\frac{b+c}{2c}\right)^{-2-2m} {}_2F_1\Big(m+1, \nu+\frac{1}{2}; \frac{3}{2}+m-\nu; \left(\frac{b-c}{b+c}\right)^2\Big),$$

we get

$$\mathscr{L}_\mu = \frac{\sqrt{\pi}2^\nu}{(b+c)^2}\left(\frac{bc}{a}\right)^\nu \times \sum_{m=0}^{\infty} \frac{\Gamma(1+m)\Gamma(1+2m-2\nu)}{m!\Gamma(m+\frac{3}{2}-\nu)}\left(\frac{a}{b+c}\right)^{2m} \times {}_2F_1\Big(1+m, \nu+\frac{1}{2}; \frac{3}{2}+m-\nu; \left(\frac{b-c}{b+c}\right)^2\Big)$$

$$= \frac{\sqrt{\pi}2^\nu}{(b+c)^2}\left(\frac{bc}{a}\right)^\nu \frac{\Gamma(1-2\nu)}{\Gamma(3/2-\nu)}$$

$$\times F_1\Big(1; 1-2\nu, \nu+\frac{1}{2}; \frac{3}{2}-\nu; \left(\frac{a}{b+c}\right)^2, \left(\frac{b-c}{b+c}\right)^2\Big),$$

where F_1 is the first Appell function.[2] For these peculiar values of indices, F_1 reduces to a hypergeometric[2–7,9–14]

$$F_1\Big(1; 1-2\nu, \nu+\frac{1}{2}; \frac{3}{2}-\nu; \left(\frac{a}{b+c}\right)^2, \left(\frac{b-c}{b+c}\right)^2\Big) = \frac{(b+c)^2}{4bc}\,{}_2F_1\Big(1, 1-2\nu; \frac{3}{2}-\nu; \frac{a^2-(b-c)^2}{4bc}\Big) = \frac{(b+c)^2}{4bc}\,{}_2F_1\Big(1, 1-2\nu; \frac{3}{2}-\nu; \frac{1-\cos\varphi_a}{2}\Big),$$

and the link with $P_{\nu-1/2}^{\nu-1/2}(\cos\varphi_a)$ is now straightforward.

III. CALCULATION OF INTEGRAL (1.7)

The proof follows the same scheme as above. Splitting $K_\nu(bt)$ into two terms $I_{\pm\nu}(bt)$ for $c>a+b$, we get

$$\mathscr{M}_\nu = \int_0^\infty t^{1-\nu} I_\nu(at)K_\nu(bt)K_\nu(ct)dt = \frac{\pi}{2\sin\pi\nu}\frac{2^{-\nu}}{c^{2-\nu}}\Big[\left(\frac{a}{b}\right)^\nu \frac{1}{\Gamma(\nu+1)} F_4\Big(1, 1-\nu; 1-\nu, 1+\nu; \frac{a^2}{c^2}, \frac{b^2}{c^2}\Big) - \left(\frac{ab}{c^2}\right)^\nu \frac{1}{\Gamma(\nu+1)} F_4\Big(1, 1+\nu; 1+\nu, 1+\nu; \frac{a^2}{c^2}, \frac{b^2}{c^2}\Big)\Big]$$

$$= \frac{\pi}{2\sin\pi\nu}\frac{2^{-\nu}}{c^{2-\nu}}\frac{1}{\Gamma(\nu+1)}\Big[\left(\frac{a}{b}\right)^\nu {}_2F_1\Big(1,1-\nu; 1+\nu; -\frac{x(1-y)}{1-x}\Big) - \left(\frac{ba}{c^2}\right)^\nu (1-x)(1-y)\,{}_2F_1(1,1-\nu; 1+\nu; xy)\Big],$$

where x, y are again given by (2.4)–(2.11).

where x, y are again given by (2.4)–(2.11).

A. When $c>a+b$

We have

$$\mathscr{M}_\nu = \frac{\pi}{2\sin\pi\nu}\frac{2^{-\nu}}{\Gamma(\nu+1)}\Big\{\frac{1}{ac}\left(\frac{ac}{b}\right)^\nu e^{-u_b} \times {}_2F_1(1, 1-\nu; 1+\nu; e^{-2u_b}) - \frac{1}{ab}\left(\frac{ab}{c}\right)^\nu e^{-u_c}\,{}_2F_1\Big(1, 1-\nu; 1+\nu; e^{-2u_c}\Big)\Big\}, \quad (3.1)$$

where the u_a, u_b, u_c are again the hyperbolic angles defined in Eqs. (2.5). The hypergeometric are rewritten by means of relations (2.7) for $\mu=-\nu$ (and the lower sign). Whence, after some easy calculations

$$\mathscr{M}_\nu = \left(\frac{\pi}{2}\right)^{1/2} 2^{2\nu-2}\Gamma(1-2\nu)\frac{\tilde{\Delta}^{2\nu-1}}{(abc)^\nu} \times\{(\sinh u_c)^{1/2-\nu}P_{\nu-1/2}^{\nu-1/2}(\cosh u_c) - (\sinh u_b)^{1/2-\nu}P_{\nu-1/2}^{\nu-1/2}(\cosh u_b)\}, \quad (3.2)$$

which is reported in Table I, line 3.

B. When $|a-b|<c<a+b$

We have

$$\mathscr{M}_\nu = \frac{\pi}{2\sin\pi\nu}\frac{2^{-\nu}}{\Gamma(\nu+1)} \times\Big\{\frac{1}{ac}\left(\frac{ac}{b}\right)^\nu e^{-i\varphi_b}\,{}_2F_1(1, 1-\nu; 1+\nu; e^{-2i\varphi_b}) + \frac{1}{ab}\left(\frac{ab}{c}\right)^\nu [e^{-i\varphi_c} \times {}_2F_1(1, 1-\nu; 1+\nu; e^{-2i\varphi_c})]^*\Big\}, \quad (3.3)$$

where [ ]* means the complex conjugate and $\varphi_a, \varphi_b, \varphi_c$ are the triangle angles, Eqs. (2.12). Setting $\mu=-\nu$ in Eq. (2.15) (with lower sign) and expressing $Q_{\nu-1/2}^{\nu-1/2}$ in terms of $P_{\nu-1/2}^{\pm(\nu-1/2)}$ only, we get

$$\frac{e^{-i\varphi}}{\Gamma(\nu+1)}\,{}_2F_1(1, 1-\nu; 1+\nu; e^{-2i\varphi}) = -\frac{2^{\nu-1}}{\cos\nu\pi}(\sin\varphi)^{\nu-1}\sqrt{\frac{\pi}{2}\sin\varphi}\; P_{\nu-1/2}^{\nu-1/2}(\cos\varphi) - \frac{i}{2}\frac{e^{i\nu\pi}\sqrt{\pi}}{\cos\nu\pi}\frac{(\sin\varphi)^{2\nu-1}}{\Gamma(\nu+1/2)}$$

and

$$\mathscr{M}_\nu = \left(\frac{\pi}{2}\right)^{3/2} 2^{-1/2-\nu} \frac{\Gamma(1/2-\nu)}{\pi} \frac{\Delta^{2\nu-1}}{(abc)^\nu} - \left(\frac{\pi}{2}\right)^{1/2} 2^{2\nu-2}\Gamma(1-2\nu) \frac{\Delta^{2\nu-1}}{(abc)^\nu} \times\{(\sin\varphi_b)^{1/2-\nu}P^{\nu-1/2}_{\nu-1/2}(\cos\varphi_b) + (\sin\varphi_c)^{1/2-\nu}P^{\nu-1/2}_{\nu-1/2}(\cos\varphi_c)\} \tag{3.4}$$

(see Table II, line 3).

C. We end with some remarks

(i) Results (3.2)–(3.4) hold too when ν goes to an integer value (actually $\nu = 0$).

(ii) For $\nu = 0$, $\int_0^\infty tI_0(at)K_0(bt)K_0(ct)dt$ may be calculated either with $\mathscr{L}_0$ or $\mathscr{M}_0$, which gives a consistency test. Setting $\nu = 0$ in (2.9) and (3.2), we have to verify that

$$\sqrt{\sinh u_a}P^{-1/2}_{-1/2}(\cosh u_a) = \sqrt{\sinh u_c}P^{-1/2}_{-1/2}(\cosh u_c) - \sqrt{\sinh u_b}P^{-1/2}_{-1/2}(\cosh u_b).$$

As in Ref. 9 $P^{-1/2}_{-1/2}(\cosh u) = \sqrt{2/\pi}(u/\sqrt{\sinh u})$, we get $u_a = u_c - u_b$, which is precisely Eq. (2.5b). Similarly, setting $\nu = 0$ in (2.17) and (3.4), we have to check the relation

$$\sqrt{\sin\varphi_a}\, P^{-1/2}_{-1/2}(\cos\varphi_a) = -\sqrt{\sin\varphi_b}P^{-1/2}_{-1/2}(\cos\varphi_b) - \sqrt{\sin\varphi_c}P^{-1/2}_{-1/2}(\cos\varphi_c) + \sqrt{2\pi}.$$

As

$$P^{-1/2}_{-1/2}(\cos\varphi) = \sqrt{\frac{2}{\pi}}\frac{\varphi}{\sqrt{\sin\varphi}},$$

we get condition (2.12), i.e., $\varphi_a + \varphi_b + \varphi_c = \pi$.

(iii) Collecting results of equations (2.9), (2.17), (3.2), and (3.4), we get the last integral (1.8), which is reported in Tables I and II, lines 4. Formulas do not hold for $\nu = 0$ as we introduce a singular factor $(\sin\nu\pi)^{-1}$, but it is easy to verify that the numerator vanishes too (see above), so the result is finite. We should get the linear term in powers of ν and with this method, we need the derivative of P^τ_σ relative to indices for $\sigma = 0, -\frac{1}{2}$. The calculation is possible but is complicated. It will be given elsewhere.[8]

ACKNOWLEDGMENTS

We are very grateful to Dr. B. Duplantier who suggested this study to us and corrected the manuscript.

[1]W. N. Bailey, Proc. Lond. Math. Soc. **40**, 37 (1936).

[2]P. Appell and J. Kampé de Feriet, *Fonctions hypergeometriques et hypersphériques* (Gauthier-Villars, Paris, 1926).

[3]See, for example, A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions,* Vol. I (McGraw–Hill, New York, 1953), p. 240.

[4]A. Gervois and H. Navelet, J. Math. Phys. **26**, 633 (1985).

[5]A. Gervois and H. Navelet, J. Math. Phys. **26**, 645 (1985).

[6]B. Duplantier (private communication).

[7]W. N. Bailey, J. London Math. Soc. **11**, 16 (1936). Our integrals (1.9) and (1.6) correspond to (3.3) and (3.5), respectively, but this latter is derived with another definition of the Q-function.

[8]A. Gervois and H. Navelet, J. Math. Phys. **27**, 688 (1986).

[9]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966), pp. 151ff.

[10]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965), p. 696, formula (10).

[11]See Ref. 3, p. 179.

[12]Reference 10, p. 1006, formula 8.738.2.1, which is a consequence of Whipple's formula, but there is still an error in the revised version: exp $i\pi(\nu + \frac{1}{2})$ must be replaced by exp $i\pi(\nu + \frac{1}{4})$.

[13]Reference 9, pp. 3 and 51.

[14]Reference 10, p. 1054, formula 9.182.1.

Some integrals involving three modified Bessel functions. II

A. Gervois[a)] and H. Navelet
Service de Physique Théorique, CEN Saclay, 91191 Gif-sur-Yvette Cedex, France

(Received 30 May 1985; accepted for publication 31 October 1985)

The integrals $\int_0^\infty Z_\mu(at)K_\nu(bt)K_\rho(ct)dt$, where $Z_\mu = I_\mu, K_\mu$, are calculated, with the help of the factorization properties of the function F_4. Results are given for real parameters a, b, c both when they are and are not in a triangle configuration. Some generalizations using derivation with respect to the parameters are considered.

I. INTRODUCTION

In a companion paper,[1] using the formal result[2]

$$\int_0^\infty t^{\lambda-1} I_\mu(at) I_\nu(bt) K_\rho(ct)dt = 2^{\lambda-2}\frac{a^\mu b^\nu}{c^{\lambda+\mu+\nu}}\frac{\Gamma((\lambda+\mu+\nu+\rho)/2)}{\Gamma(\mu+1)} \times\frac{\Gamma((\lambda+\mu+\nu-\rho)/2)}{\Gamma(\nu+1)} \times F_4\left(\frac{\lambda+\mu+\nu+\rho}{2},\frac{\lambda+\mu+\nu-\rho}{2}, \mu+1,\nu+1;\frac{a^2}{c^2},\frac{b^2}{c^2}\right), \quad (1.1)$$

with

$$\mathrm{Re}(\lambda+\mu+\nu\pm\rho)>0, \quad (1.2)$$

$$\mathrm{Re}(c-a-b)>0, \quad (1.3)$$

we showed that some integrals of the form

$$\int_0^\infty t^{\lambda-1} Z_\mu(at)K_\nu(bt)K_\rho(ct)dt, \quad Z_\mu = I_\mu, K_\mu, \quad (1.4)$$

can be explicitly calculated when the Appell function F_4 (see Ref. 3) factorizes into functions of one variable only, actually hypergeometrics ${}_2F_1$. The purpose of this paper is to complete this study.

The main remark is the following: when F_4 factorizes into hypergeometrics ${}_2F_1$, it is possible to perform its analytical continuation outside region (1.3)—expect at most for a cut along the real axis—to calculate integrals derived from (1.1), even if (1.1) itself does not converge. Such considerations were used in the calculation of integrals of the form $\int_0^\infty dt\, t^{\lambda-1}J_\mu J_\nu J_\rho\, dt$, when $\mathrm{Im}(\pm a\pm b\pm c)>0$ (see Ref. 4) and were used in Ref. 1 to get $\int_0^\infty dt\, t^{1+\mu}I_\mu K_\nu K_\nu$ in the region $c>a+b$ and $(a-b)<c<a+b$, though separately each integral $\int_0^\infty t^{1+\mu}I_\mu I_{\pm\nu}K_\nu\, dt$ does not exist. The most interesting situation for applications corresponds to the case when a, b, c are real and may be considered as the sides of a triangle, i.e., $|a-b|<c<a+b$. Notice too, that, up to a proportionality factor $(2/i\pi)\exp((i\pi/2)(\lambda+\mu+\nu-\rho))$, integrals $\int_0^\infty t^{\lambda-1}I_\mu I_\nu K_\rho\, dt$ and $\int_0^\infty t^{\lambda-1}J_\mu J_\nu H_\rho^{(1)}\, dt$ coincide when they exist.

The factorization cases for F_4 (see Refs. 2–4) can be divided into two classes: (i) $\lambda = 2\pm\mu$, $\nu=\pm\rho$, which leads to the calculation of

$$\int_0^\infty dt\, t^{1-\mu}I_\mu K_\nu K_\nu, \quad \int_0^\infty t^{1-\nu}I_{-\nu}K_\nu K_\nu\, dt, \quad \int_0^\infty t^{1-\nu}K_\nu K_\nu K_\nu\, dt$$

(see Ref. 1); and (ii) $\lambda=1$, any μ, ν, ρ, as

$$F_4(\alpha,\beta;\gamma,\gamma';X(1-Y),Y(1-X)) = {}_2F_1(\alpha,\beta,\gamma;X)\,{}_2F_1(\alpha,\beta,\gamma',Y) \quad (1.5a)$$

whenever

$$\alpha+\beta+1=\gamma+\gamma'. \quad (1.5b)$$

This formula derives from the finite summation by Burchnall and Chaundy[5],

$$F_4(\alpha,\beta;\gamma,\gamma';X(1-Y),Y(1-X)) = \sum_{r=0}^\infty \frac{(\alpha)_r(\beta)_r}{(\gamma)_r(\gamma')_r}\frac{(\alpha+\beta+1-\gamma-\gamma')_r}{r!} \times Y^rX^r\,{}_2F_1(\alpha+r,\beta+r,\gamma+r,X) \times {}_2F_1(\alpha+r,\beta+r,\gamma'+r,Y), \quad (1.6a)$$

when

$$\alpha+\beta+1=\gamma+\gamma'-n, \quad n \text{ integer}, \quad (1.6b)$$

and the same considerations may apply.

In the present paper, we study factorization of class (ii) [Eqs. (1.5)]. We assume that the parameters a, b, c and indices μ, ν, ρ are real though this restriction is probably not necessary. In Sec. II, we study the transformation $(a^2/c^2, b^2/c^2)\to(X, Y)$ on the parameters and give the expression for the general integral (1.4).

In Sec. III, we derive formulas for special values of indices ($\mu=\pm\nu$ or $\nu=\rho$) following Ref. 4, with complementary results when $a=b$ or $b=c$ ("isosceles" case). In Sec. IV, we use derivation with respect to the parameters to get integrals where $\alpha+\beta+1-\gamma-\gamma'$ is increased by 1, 2, ..., n units, generalizing formula (1.6b) to positive and negative integers. We indicate some other results coming from derivation with the integrals of Ref. 1.

Results are collected in Tables I–III. The cases when two lengths are equal ($a=b$ or $b=c$), which are of some interest in physical situations, are given in the Appendix [formulas (A1)–(A3)].

[a)] Chercheur Centre National de la Recherche Scientifique.

 0022-2488/86/030688-08$02.50

II. GENERAL FORMULAS

From $K_\nu = (\pi/2\sin\pi\nu)(I_{-\nu} - I_\nu)$ and using (1.1), we get

$$\int_0^\infty I_\mu(at)K_\nu(bt)K_\rho(ct)dt = \frac{\pi}{4\sin\pi\nu}\frac{a^\mu}{c^{1+\mu}}\frac{1}{\Gamma(1+\mu)}\left\{\left(\frac{c}{b}\right)^\nu \frac{\Gamma((1+\mu-\nu+\rho)/2)\Gamma((1+\mu-\nu-\rho)/2)}{\Gamma(1-\nu)}\right.$$
$$\times F_4\left(\frac{1+\mu-\nu+\rho}{2}, \frac{1+\mu-\nu-\rho}{2}, 1+\mu, 1-\nu; \frac{a^2}{c^2}, \frac{b^2}{c^2}\right)$$
$$-\left(\frac{c}{b}\right)^{-\nu}\frac{\Gamma((1+\mu+\nu+\rho)/2)\Gamma((1+\mu+\nu-\rho)/2)}{\Gamma(1+\nu)}$$
$$\left.\times F_4\left(\frac{1+\mu+\nu+\rho}{2}, \frac{1+\mu+\nu-\rho}{2}, 1+\mu, 1+\nu; \frac{a^2}{c^2}, \frac{b^2}{c^2}\right)\right\},$$

provided $c > a+b$, $1+\mu-|\nu|-\rho| > 0$

The factorization property (1.5a) reads

$$F_4\left(\frac{1+\mu\mp\nu+\rho}{2}, \frac{1+\mu\mp\nu-\rho}{2}, 1+\mu, 1\mp\nu; \frac{a^2}{c^2}, \frac{b^2}{c^2}\right)$$
$$= {}_2F_1\left(\frac{1+\mu\mp\nu+\rho}{2}, \frac{1+\mu\mp\nu-\rho}{2}, 1+\mu; X\right){}_2F_1\left(\frac{1+\mu\mp\nu+\rho}{2}, \frac{1+\mu\mp\nu-\rho}{2}, 1\mp\nu; Y\right),$$

where X, Y are such that

$$X(1-Y) = a^2/c^2, \quad Y(1-X) = b^2/c^2, \tag{2.1a}$$

with the asymptotic condition

$$X, Y \to 0, \quad \text{when } c \to +\infty. \tag{2.1b}$$

Rewriting

$${}_2F_1\left(\frac{1+\mu\mp\nu+\rho}{2}, \frac{1+\mu\mp\nu-\rho}{2}, 1+\mu; X\right)$$

as

$$(1-X)^{\pm\nu}{}_2F_1\left(\frac{1+\mu\pm\nu+\rho}{2}, \frac{1+\mu\pm\nu-\rho}{2}, \quad 1+\mu, X\right),$$

and after some transformations, we get the final expression

$$\int_0^\infty I_\mu(at)K_\nu(bt)K_\rho(ct)dt = \frac{a^\mu b^\nu}{c^{\mu+\nu+1}}\frac{\Gamma((1+\mu+\nu+\rho)/2)\Gamma((1+\mu+\nu-\rho)/2)}{4\Gamma^2(\mu+1)}$$
$$\times\Gamma\left(\frac{1+\mu-\nu+\rho}{2}\right)\Gamma\left(\frac{1+\mu-\nu-\rho}{2}\right)$$
$$\times {}_2F_1\left(\frac{1+\mu+\nu+\rho}{2}, \frac{1+\mu+\nu-\rho}{2}, 1+\mu; X\right)$$
$$\times {}_2F_1\left(\frac{1+\mu+\nu+\rho}{2}, \frac{1+\mu+\nu-\rho}{2}, 1+\mu; 1-Y\right). \tag{2.2}$$

This result was already obtained by Bailey,[6] for $c > a+b$ again, with the correspondence $a = c\sin\varphi\sin\phi$, $b = c\cos\varphi\cos\phi$ (and $X = \sin^2\phi$, $1-Y=\sin^2\varphi$).

The main point is that *this result is still true when* $c < a+b$, actually when

$$|a-b| < c < a+b, \tag{2.3}$$

i.e., when a, b, c may be considered as the sides of a triangle ("triangle configuration") as $\int_0^\infty I_\mu K_\nu K_\rho\,dt$ is an analytical function of variables a, b, c and exists when (2.3) is fulfilled, although separately $\int_0^\infty I_\mu I_{\pm\nu}K_\rho\,dt$ does not converge. The integral is obtained by analytical continuation of the function ${}_2F_1$, which exists everywhere but (at most) on a cut. In the same way we can get $\int_0^\infty K_\mu K_\nu K_\rho\,dt$ for any real positive a, b, c, though $\int_0^\infty I_{\pm\mu}K_\nu K_\rho\,dt$ does not exist when $c < |a-b|$.

It remains to give explicit expressions for X, Y, both for $c > a+b$ and $|a-b| < c < a+b$. The case $c < |a-b|$, which appears for the calculation of $\int_0^\infty K_\mu K_\nu K_\rho\,dt$, is not necessary because of the symmetry of the roles of a, b, c and μ, ν, ρ.

Equations (2.1) are rewritten as

$$X=(a^2+c^2-b^2-\sqrt{\delta})/2c^2,$$
$$Y=(b^2+c^2-a^2-\sqrt{\delta})/2c^2, \tag{2.4a}$$
$$\delta=a^4+b^4+c^4-a^2b^2-b^2c^2-c^2a^2, \tag{2.4b}$$

where $\sqrt{\delta}$ denotes the positive square root when $c>a+b$ and $\sqrt{\delta}=i\sqrt{-\delta}$ in the triangle configuration. We rewrite (2.4) using angle variables.

If $c>a+b$, we set

$$c^2=a^2+b^2+2ab\cosh u_c,\quad a^2=b^2+c^2-2bc\cosh u_a,$$
$$b^2=a^2+c^2-2ac\cosh u_b, \tag{2.5a}$$

where u_a, u_b, u_c are hyperbolic angles

$$u_c=u_a+u_b, \tag{2.5b}$$

and $\sqrt{\delta}/4=\tilde{\Delta}$ may be thought as a measure of a pseudoarea

$$\tilde{\Delta}=\tfrac{1}{2}ab\sinh u_c=\tfrac{1}{2}bc\sinh u_a=\tfrac{1}{2}ca\sinh u_b. \tag{2.5c}$$

Then,

$$X=(a/c)e^{-u_b},\quad Y=(b/c)e^{-u_a},$$
$$1-X=(b/c)e^{u_a},\quad 1-Y=(a/c)e^{u_b}, \tag{2.5d}$$

and

$$0<X,\ Y<1,\quad -1<1-2X,\quad 1-2Y<1. \tag{2.5e}$$

If $|a-b|<c<a+b$, we define true angles φ_a, φ_b, φ_c,

$$a^2=b^2+c^2-2bc\cos\varphi_a,$$
$$b^2=c^2+a^2-2ac\cos\varphi_b, \tag{2.6a}$$
$$c^2=a^2+b^2-2ba\cos\varphi_c,$$

with

$$\varphi_a+\varphi_b+\varphi_c=\pi, \tag{2.6b}$$

and $\sqrt{-\delta}/4=\Delta$ is the true area of the triangle

$$\Delta=\tfrac{1}{2}ab\sin\varphi_c=\tfrac{1}{2}bc\sin\varphi_a=\tfrac{1}{2}ca\sin\varphi_b. \tag{2.6c}$$

The corresponding expressions for the variables are

$$X=(a/c)e^{-i\varphi_b},\quad Y=(b/c)e^{-i\varphi_a},$$
$$1-X=(b/c)e^{i\varphi_a},\ 1-Y=(a/c)e^{i\varphi_b}. \tag{2.6d}$$

Notice that

$$1-Y=X^*,\quad 1-X=Y^*, \tag{2.6e}$$

where X^* denotes the complex conjugate of X, which insures the reality of integral (2.2) in that case.

All formulas (2.2)–(2.6) are collected in Table I.

III. SPECIAL CASES

Expression (2.2) simplifies when the indices take peculiar values, as the hypergeometric ${}_2F_1$ reduces to Legendre functions. We examine successively cases $\mu=\pm\nu$, $\mu=\pm\rho$, and get the limits when μ, ν, ρ go to zero. The results are reported in Tables I and II. All formulas concerning Legendre functions may be found in Ref. 7.

A. Case $\mu=\pm\nu$

We start with the relations

$${}_2F_1\left(\frac{1+\rho}{2}+\nu,\frac{1-\rho}{2}+\nu,1+\nu;x\right)$$
$$=(1-x)^{-\nu/2}x^{-\nu/2}P^{\nu}_{\rho-1}(1-2x), \tag{3.1a}$$

TABLE I. Some formulas with general (real) a, b, c. Cases $\mu=\pm\nu$, $\rho=\nu$ are explicitly written in terms of the Legendre functions on the cut (real argument with modulus less than 1) or outside the cut.

$$\int_0^\infty I_\mu(at)K_\nu(bt)K_\rho(ct)dt=\frac{a^\mu b^\nu}{4c^{1+\mu+\nu}}\frac{\Gamma((\mu+1+\nu+\rho)/2)\Gamma((\mu+1+\nu-\rho)/2)\Gamma((\mu+1-\nu+\rho)/2)\Gamma((\mu+1-\nu-\rho)/2)}{\Gamma^2(\mu+1)}$$
$$\times{}_2F_1\left(\frac{1+\mu+\nu+\rho}{2},\frac{1+\mu+\nu-\rho}{2},1+\mu;X\right){}_2F_1\left(\frac{1+\mu+\nu+\rho}{2},\frac{1+\mu+\nu-\rho}{2},1+\mu;1-Y\right),$$

$1+\mu>|\nu|+|\rho|,\quad a<b+c;$

$$\int_0^\infty I_\nu(at)K_\nu(bt)K_\rho(ct)dt=\frac{1}{4c}\Gamma\left(\frac{1+\rho}{2}+\nu\right)\Gamma\left(\frac{1-\rho}{2}+\nu\right)\Gamma\left(\frac{1+\rho}{2}\right)\Gamma\left(\frac{1-\rho}{2}\right)P^{-\nu}_{(\rho-1)/2}(1-2X)P^{-\nu}_{(\rho-1)/2}(2Y-1),$$

$|\rho|<1+2\inf(0,\nu),\quad a<b+c;$

$$\int_0^\infty I_\mu(at)K_\rho(bt)K_\rho(ct)dt=\frac{1}{a}Q^{\rho}_{(\mu-1)/2}\left(\frac{2}{X}-1\right)Q^{-\rho}_{(\mu-1)/2}\left(\frac{2}{1-Y}-1\right);$$

$$\int_0^\infty K_\nu(at)K_\nu(bt)K_\rho(ct)dt=\frac{\pi}{8c\sin\pi\nu}\Gamma\left(\frac{1+\rho}{2}\right)\Gamma\left(\frac{1-\rho}{2}\right)\left\{\Gamma\left(\frac{1+\rho}{2}-\nu\right)\Gamma\left(\frac{1-\rho}{2}-\nu\right)P^{\nu}_{(\rho-1)/2}(1-2X)P^{\nu}_{(\rho-1)/2}(2Y-1)\right.$$
$$\left.-\Gamma\left(\frac{1+\rho}{2}+\nu\right)\Gamma\left(\frac{1-\rho}{2}+\nu\right)P^{-\nu}_{(\rho-1)/2}(1-2X)P^{-\nu}_{(\rho-1)/2}(2Y-1)\right\}$$
$$=\frac{\pi}{2a\sin\pi\rho}\{Q^{\rho}_{-(\mu+1)/2}(1-2Y)Q^{-\rho}_{-(\mu+1)/2}(2X-1)-Q^{-\rho}_{(\mu-1)/2}(1-2Y)Q^{\rho}_{(\mu-1)/2}(2X-1)\}.$$

$c>a+b,\quad X=(a/c)e^{-u_b},\quad Y=(b/c)e^{-u_a},\quad 1-X=(b/c)e^{u_a},\quad 1-Y=(a/c)e^{u_b}.$

$a^2=b^2+c^2-2bc\cosh u_a,\quad b^2=c^2+a^2-2ac\cosh u_b,\quad c^2=a^2+b^2+2ab\cosh u_c,\quad u_c=u_a+u_b.$

Correspondence with Bailey's result $a=c\sin\varphi\sin\phi$, $b=c\cos\varphi\cos\phi$,

$X=\sin^2\phi,\quad 1-Y=\sin^2\varphi,\quad 1-2X=\cos 2\phi,\quad 2Y-1=\cos 2\varphi.$

$|a-b|<a<a+b,\quad X=(a/c)e^{-i\varphi_b},\quad Y=(b/c)e^{-i\varphi_a},\quad 1-X=Y^*=(b/c)e^{i\varphi_a},\quad 1-Y=X^*=(a/c)e^{i\varphi_b},\quad \varphi_a+\varphi_b+\varphi_c=\pi,$

$a^2=b^2+c^2-2bc\cos\varphi_a,\quad b^2=c^2+a^2-2ac\cos\varphi_b,\quad c^2=a^2+b^2-2ab\cos\varphi_c.$

TABLE II. The same when one of the indices is zero. K denotes the elliptic function.

$$\int_0^\infty K_0(at)K_0(bt)K_\rho(ct)dt = \frac{\pi}{2a\sin\pi\rho}\{Q^{\rho}_{-1/2}(1-2Y)Q^{-\rho}_{-1/2}(2X-1) - Q^{-\rho}_{-1/2}(1-2Y)Q^{\rho}_{-1/2}(2X-1)\}, \quad |\rho|<1;$$

$$\int_0^\infty I_0(at)K_0(bt)K_0(ct)dt = \frac{1}{c}K\left(\sqrt{\frac{a}{c}}e^{-u_b/2}\right)K\left(\sqrt{\frac{a}{b}}e^{u_b/2}\right), \quad c>a+b,$$

$$= \frac{1}{b}K\left(\sqrt{\frac{a}{b}}e^{i\varphi_c/2}\right)K\left(\sqrt{\frac{a}{b}}e^{-i\varphi_c/2}\right), \quad |a-b|<c<a+b;$$

$$\int_0^\infty K_0(at)K_0(bt)K_0(ct)dt = \frac{\pi}{2c}\left\{K\left(\sqrt{\frac{a}{c}}e^{-u_b/2}\right)K\left(\sqrt{\frac{b}{c}}e^{-u_a/2}\right) + K\left(\sqrt{\frac{a}{c}}e^{u_b/2}\right)K\left(\sqrt{\frac{b}{c}}e^{u_a/2}\right)\right\}, \quad c>a+b;$$

$$= \frac{\pi}{2c}\left\{K\left(\sqrt{\frac{a}{c}}e^{i\varphi_b/2}\right)K\left(\sqrt{\frac{b}{c}}e^{i\varphi_a/2}\right) + K\left(\sqrt{\frac{a}{c}}e^{-i\varphi_b/2}\right)K\left(\sqrt{\frac{b}{c}}e^{-i\varphi_a/2}\right)\right\}, \quad |a-b|<c<a+b.$$

$$ {}_2F_1\left(\frac{1+\rho}{2}+\nu, \frac{1-\rho}{2}+\nu, 1+\nu; z\right) = (1-z)^{-\nu/2}(-z)^{-\nu/2}P^{\nu}_{(\rho-1)/2}(1-2z), \tag{3.1b}$$

which hold on and outside the cut, respectively.

For $c>a+b$, as $0<X$, $Y<1$, we take expression (3.1a) and, with the help of definitions (2.5), Eq. (2.2) becomes

$$\int_0^\infty I_\nu(at)K_\nu(bt)K_\rho(ct)dt = \frac{1}{4c}\Gamma\left(\frac{1+\rho}{2}+\nu\right)\Gamma\left(\frac{1-\rho}{2}+\nu\right) \times\Gamma\left(\frac{1+\rho}{2}\right)\Gamma\left(\frac{1-\rho}{2}\right) \times P^{-\nu}_{(\rho-1)/2}(1-2X)P^{-\nu}_{(\rho-1)/2}(2Y-1), \tag{3.2}$$

provided $|\rho|<1+2\inf(0,\nu)$.

For $|a-b|<c<a+b$, we start with expression (3.1b) and definitions (2.6). Taking some care in the determination of the powers, we get finally the same expression (3.2). The integral $\int_0^\infty I_{-\nu}K_\nu K_\rho\, dt$ is derived by replacing ν by $-\nu$ (as $K_\nu = K_{-\nu}$) and we finally get the result for the product of three K_α functions:

$$\int_0^\infty K_\nu(at)K_\nu(bt)K_\rho(ct)dt = \frac{\pi}{2\sin\pi\nu}\frac{1}{4c}\Gamma\left(\frac{1+\rho}{2}\right)\Gamma\left(\frac{1-\rho}{2}\right) \times\left\{\Gamma\left(\frac{1+\rho}{2}-\nu\right)\Gamma\left(\frac{1-\rho}{2}-\nu\right) \times P^{\nu}_{(\rho-1)/2}(1-2X)P^{\nu}_{(\rho-1)/2}(2Y-1) - \Gamma\left(\frac{1+\rho}{2}+\nu\right)\Gamma\left(\frac{1-\rho}{2}+\nu\right) \times P^{-\nu}_{(\rho-1)/2}(1-2X)P^{-\nu}_{(\rho-1)/2}(2Y-1)\right\}. \tag{3.3}$$

Results are reported in Table I, lines 2 and 4.

B. Case $\nu = \pm\rho$

In both triangle and nontriangle configurations, we use the identity[7]

$$2^{(\mu-1)/2}\frac{\Gamma((1+\mu)/2)\Gamma((1+\mu)/2-\rho)}{\Gamma(\mu+1)} \times(z+1)^{-(1+\mu+\rho)/2}(z-1)^{\rho/2} \times{}_2F_1\left(\frac{1+\mu}{2}+\rho, \frac{1+\mu}{2}, 1+\mu; Z\right) = e^{i\pi\rho}Q^{\rho}_{(\mu-1)/2}(z),$$

where Q is the Legendre function outside the cut and $z = 2/Z - 1$. We get the unique formula (see Table I, line 3)

$$\int_0^\infty I_\mu(at)\kappa_\rho(bt)K_\rho(ct)dt = \frac{1}{a}Q^{\rho}_{(\mu-1)/2}\left(\frac{2}{X}-1\right)Q^{-\rho}_{(\mu-1)/2}\left(\frac{2}{1-Y}-1\right) \tag{3.4}$$

and, for $\int_0^\infty K_\mu K_\rho K_\rho\, dt$, another expression

$$\int_0^\infty K_\mu(at)K_\rho(bt)K_\rho(ct) = \frac{\pi}{2\sin\pi\mu}\frac{1}{a} \times\left\{Q^{\rho}_{-(\mu+1)/2}\left(\frac{2}{X}-1\right)Q^{-\rho}_{-(\mu+1)/2}\left(\frac{2}{1-Y}-1\right) - Q^{\rho}_{(\mu-1)/2}\left(\frac{2}{X}-1\right)Q^{-\rho}_{(\mu-1)/2}\left(\frac{2}{1-Y}-1\right)\right\}. \tag{3.5}$$

This expression looks different from (3.3). To check that they are actually the same, we make in Eq. (3.5) the substitution $\rho\to\nu$, $\mu\to\rho$, $c\leftrightarrow a$, $2/x-1\Rightarrow 1-2Y$, $2/(1-Y)-1\Rightarrow 2X-1$, and then rewrite the Q^{σ}_{τ} in terms of the $P^{\pm\sigma}_{\tau}$ (Table I, line 5).

C. Case $\mu=\nu=\rho=0$

Expressions for $\int_0^\infty I_0K_0K_0\, dt$ are easy to get, when using (3.2) or (3.5) and going to the limit. We have

$$\int_0^\infty I_0(at)K_0(bt)K_0(ct)\, dt = (\pi^2/4c)P^0_{-1/2}(1-2X)P^0_{-1/2}(2Y-1).$$

The function $P^0_{-1/2}$ can be expressed in terms of the elliptic function K but the definition differs depending on whether the argument is on or outside the cut.

For $-1<x<1$,

$$P^0_{-1/2}(x) = (2/\pi)K(\sqrt{(1-x)/2}),$$

and for any other z,

$$P^0_{-1/2}(z) = \frac{2}{\pi}\sqrt{\frac{2}{z+1}}K\left(\frac{\sqrt{z-1}}{\sqrt{z+1}}\right).$$

We shall need, too,

$$Q^0_{-1/2}(z) = \sqrt{\frac{2}{z+1}}K\left(\frac{\sqrt{2}}{\sqrt{z+1}}\right).$$

We get then

$$\int_0^\infty I_0(at)K_0(bt)K_0(ct)\,dt$$

$$= \frac{1}{c}K\left(\sqrt{\frac{a}{c}}e^{-u_b/2}\right)K\left(\sqrt{\frac{a}{c}}e^{u_b/2}\right), \quad c>a+b, \tag{3.6a}$$

$$= \frac{1}{b}K\left(\sqrt{\frac{a}{b}}e^{i\varphi_c/2}\right)K\left(\sqrt{\frac{a}{b}}e^{-i\varphi_c/2}\right),$$

$$|a-b|<c<a+b, \tag{3.6b}$$

the second formula being symmetrical by exchange of b and c.

The derivation of $\int_0^\infty K_0K_0K_0\,dt$ is more subtle, as we need the behavior of $P^\sigma_{\tau-1/2}$ for $\sigma\approx 0$ or $\tau\approx 0$, depending on whether we start from (3.3) or (3.5). For instance, we set $\rho=0$ in Eq. (3.5). We have

$$\int_0^\infty K_\mu(at)K_0(bt)K_0(ct)\,dt$$

$$= \frac{\pi}{2\sin\pi\mu}\frac{1}{a}\times\left\{Q^0_{-(\mu+1)/2}\left(\frac{2}{X}-1\right)Q^0_{-(\mu+1)/2}\left(\frac{2}{1-Y}-1\right) - Q^0_{(\mu-1)/2}\left(\frac{2}{X}-1\right)Q^0_{(\mu-1)/2}\left(\frac{2}{1-Y}-1\right)\right\}.$$

As[7]

$$\frac{\partial}{\partial\sigma}Q^0_{\sigma-1/2}(z)\bigg|_{\sigma=0} = -\frac{\pi^2}{2}P^0_{-1/2}(z),$$

the indetermination is removed and

$$\int_0^\infty K_0(at)K_0(bt)K_0(ct)\,dt$$

$$= \frac{\pi^2}{4a}\left\{P^0_{-1/2}\left(\frac{2}{X}-1\right)Q^0_{-1/2}\left(\frac{2}{1-Y}-1\right) + P^0_{-1/2}\left(\frac{2}{1-Y}-1\right)Q^0_{-1/2}\left(\frac{2}{X}-1\right)\right\},$$

where the Legendre functions are outside the cut. In terms of the elliptic function, we have finally

$$\int_0^\infty K_0(at)K_0(bt)K_0(ct)\,dt$$

$$= (\pi/2a)\sqrt{X}\sqrt{1-Y}\{K(\sqrt{1-X})K(\sqrt{1-Y}) + K(\sqrt{X})K(\sqrt{Y})\}$$

$$= \frac{\pi}{2c}\left\{K\left(\sqrt{\frac{b}{c}}e^{u_a/2}\right)K\left(\sqrt{\frac{a}{c}}e^{u_b/2}\right) + K\left(\sqrt{\frac{a}{c}}e^{-u_b/2}\right)K\left(\sqrt{\frac{b}{c}}e^{-u_a/2}\right)\right\}, \quad c>a+b, \tag{3.7a}$$

$$= \frac{\pi}{2c}\left\{K\left(\sqrt{\frac{b}{c}}e^{i\varphi_a/2}\right)K\left(\sqrt{\frac{a}{c}}e^{i\varphi_b/2}\right) + K\left(\sqrt{\frac{a}{c}}e^{-i\varphi_b/2}\right)K\left(\sqrt{\frac{b}{c}}e^{-i\varphi_a/2}\right)\right\},$$

$$|a-b|<c<a+b. \tag{3.7b}$$

All these results are reported in Table II.

IV. DERIVATION

New integrals may be calculated using the derivation with respect to the parameters. From

$$\int_0^\infty t^\lambda Z_{\mu\pm1}K_\nu K_\rho\,dt$$

$$= \epsilon_Z\left(\frac{\partial}{\partial a}\mp\frac{\mu}{a}\right)\int_0^\infty t^{\lambda-1}Z_\mu K_\nu K_\rho\,dt \tag{4.1}$$

$$(\epsilon_Z = 1 \text{ if } Z_\mu = I_\mu = I_\mu,\ \epsilon_Z = -1 \text{ if } Z_\mu = K_\mu),$$

or

$$\int_0^\infty t^\lambda Z_\mu K_\nu K_{\rho\pm1}\,dt$$

$$= -\left(\frac{\partial}{\partial c}\mp\frac{\rho}{c}\right)\int_0^\infty t^{\lambda-1}Z_\mu K_\nu K_\rho\,dt, \tag{4.2}$$

it is easy to reach any integral where the indices may be lowered or raised by one or more units (provided it converges); the power term is always raised. This was already used in Ref. 4 for getting new integrals of three J functions.

Each of the operations above corresponds to increasing by one unit the quantity

$$\alpha+\beta+1-\gamma-\gamma',$$

where α, β, γ, γ' are the indices of the F_4 function [Eq. (1.1)]. In the frame of the present paper, starting with $\alpha+\beta+1=\gamma+\gamma'$, we get integrals such that

$$\alpha+\beta+1=\gamma+\gamma'+n, \quad n \text{ positive integer}, \tag{4.3}$$

which is the opposite situation of that proposed by Burchnall and Chaundy,[5] reported here in Eqs. (1.6).

As an illustration, we calculate

$$\int_0^\infty tZ_\mu(at)K_\nu(bt)K_{\mu\pm\nu}(ct)\,dt, \quad Z_\mu = I_\mu, K_\mu.$$

A similar integral, but with three J functions, was calculated differently elsewhere,[8] but using a basic formula,[9] which is not adequate here. We consider only the + sign (the − sign amounts to changing ν into $-\nu$). We have

$$\int_0^\infty tI_\mu(dt)K_\nu(bt)K_{\mu+\nu}(ct)\,dt$$

$$= \left(\frac{\mu+\nu-1}{c}-\frac{\partial}{\partial c}\right)\int_0^\infty I_\mu(at)K_\nu(bt)K_\nu(ct)\,dt.$$

This calculation was already performed in Ref. 8. We get the functional relation

TABLE III. Some integrals obtained by derivation with respect to the parameters.

$$\int_0^\infty tI_\mu(at)K_\nu(bt)K_{\mu\pm\nu}(ct)dt = \frac{a^\mu b^{\pm\nu}}{c^{\mu\pm\nu}}\frac{\Gamma(\mu\pm\nu)\Gamma(1\mp\nu)}{4\Gamma(1+\mu)}\mathscr{F}, \quad 2+\mu>|\nu|+|\mu\pm\nu|,$$

$$\mathscr{F} = \frac{1}{2\tilde{\Delta}}\left\{{}_2F_1\left(\mu\pm\nu, 1;\mu+1;\frac{a}{c}e^{u_b}\right) - {}_2F_1\left(\mu\pm\nu, 1;\mu+1;\frac{a}{c}e^{-u_b}\right)\right\}, \quad c>a+b$$

$$= \frac{1}{\Delta}\operatorname{Im}{}_2F_1\left(\mu\pm\nu, 1;\mu+1;\frac{a}{c}e^{i\varphi_b}\right), \quad |a-b|<c<a+b;$$

$$\int_0^\infty tI_0(at)K_\nu(bt)K_\nu(ct)dt = \frac{\Gamma(\nu)\Gamma(1-\nu)}{4\tilde{\Delta}}\sinh(\nu u_a), \quad c>a+b\,|\nu|<1$$

$$= \frac{\Gamma(\nu)\Gamma(1-\nu)}{4\Delta}\sin(\nu\varphi_a), \quad |a-b|<c<a+b;$$

$$\int_0^\infty tI_0(at)K_0(bt)K_0(ct)dt = \frac{u_a}{4\tilde{\Delta}}, \quad c>a+b,$$

$$= \frac{\varphi_a}{4\Delta}, \quad |a-b|<c<a+b;$$

$$\int_0^\infty tI_\mu(at)K_0(bt)K_\mu(ct)dt = \left(\frac{a}{c}\right)^\mu \frac{1}{4\tilde{\Delta}}\sum_{n=0}^\infty \left(\frac{a}{c}\right)^{n+1}\frac{\sinh[(n+1)u_b]}{\mu+n+1}, \quad c>a+b,$$

$$= \left(\frac{a}{c}\right)^\mu \frac{1}{4\Delta}\sum_{n=0}^\infty \left(\frac{a}{c}\right)^{n+1}\frac{\sin[(n+1)\varphi_b]}{\mu+n+1}, \quad |a-b|<c<a+b;$$

$$\int_0^\infty tK_0(at)K_0(bt)K_0(ct)dt = -\frac{1}{4\tilde{\Delta}}\left\{u_a\ln\left(\frac{a}{c}\right) - \sum_{n=0}\left(\frac{a}{c}\right)^{n+1}\frac{\sinh[(n+1)u_b]}{(n+1)^2}\right\} \quad c>a+b$$

$$= \frac{1}{4\Delta}\left\{-\varphi_a\ln\left(\frac{a}{c}\right) + \sum_{n=0}\left(\frac{a}{c}\right)^{n+1}\frac{\sin[(n+1)\varphi_b]}{(n+1)^2}\right\}.$$

(i) $c>a+b$, $c^2=a^2+b^2+2ab\cosh u_c$, $a^2=b^2+c^2-2bc\cosh u_a$, $b^2=a^2+c^2-2ac\cosh u_b$,
$u_c=u_a+u_b$, $\tilde{\Delta}=\frac{1}{2}ab\sinh u_c=\frac{1}{2}bc\sinh u_a=\frac{1}{2}ca\sinh u_b$.

(ii) $|a-b|<a<a+b$, $a^2=b^2+c^2-2bc\cos\varphi_a$, $b^2=c^2+a^2-2ac\cos\varphi_b$, $c^2=a^2+b^2-2ab\cos\varphi_c$,
$\varphi_a+\varphi_b+\varphi_c=\pi$, $\Delta=\frac{1}{2}ab\sin\varphi_c=\frac{1}{2}bc\sin\varphi_a=\frac{1}{2}ac\sin\varphi_b$.

$$\begin{aligned}(\mu+\nu)F_4(\mu+\nu,1;\mu+1,\nu+1;X(1-Y),Y(1-X)) &= (c^2/4\tilde{\Delta})[\mu\,{}_2F_1(\mu+\nu,1,\nu+1;Y) \\ &\quad +\nu\,{}_2F_1(\mu+\nu,1,1+\mu;X)], \quad c>a+b \\ &= (c^2/4\Delta)\operatorname{Im}[\mu\,{}_2F_1(\mu+\nu,1,\nu+1;Y) \\ &\quad +\nu\,{}_2F_1(\mu+\nu,1,1+\mu;X)], \quad |a-b|<c<a+b,\end{aligned}$$

whence

$$\begin{aligned}\int_0^\infty tI_\mu(at)K_\nu(bt)K_{\mu+\nu}(ct)\,dt &= \frac{a^\mu b^\nu}{8\tilde{\Delta}c^{\mu+\nu}}\frac{\Gamma(\mu+\nu)\Gamma(1-\nu)}{\Gamma(1+\mu)} \\ &\quad\times\{{}_2F_1(\mu+\nu,1,1+\mu;1-Y) \\ &\quad -{}_2F_1(\mu+\nu,1,1+\mu;X)\}, \quad \text{if } c>a+b, \\ &= -\frac{a^\mu b^\nu}{8\Delta c^{\mu+\nu}}\frac{\Gamma(\mu+\nu)\Gamma(1-\nu)}{\Gamma(1+\mu)} \\ &\quad\times\{[{}_2F_1(\mu+\nu,1,1+\mu;X) \\ &\quad -{}_2F_1(\mu+\nu,1,1+\mu;X^*)]/i\}, \\ &\quad \text{if } |a-b|<c<a+b,\end{aligned} \tag{4.4}$$

where $\tilde{\Delta}$, Δ are the areas (2.5c) and (2.6c).

The associate integral $\int_0^\infty t\,I_{-\mu}K_\nu K_{\mu+\nu}\,dt$ is obtained by replacing both μ and ν by the opposite $-\mu$, $-\nu$ in Eq. (4.4). We get then the integral

$$\int_0^\infty tK_\mu K_\nu K_{\mu+\nu}\,dt.$$

Again, we may be interested by some limit cases when one of the indices μ or ν goes to zero.

For example,

$$\begin{aligned}\int_0^\infty tI_0(at)K_\nu(bt)K_\nu(ct)\,dt &= \frac{\Gamma(\nu)\Gamma(1-\nu)}{4\tilde{\Delta}}\sinh\nu u_a \quad (c>a+b) \\ &= \frac{\Gamma(\nu)\Gamma(1-\nu)}{4\Delta}\sin\nu\varphi_a \quad (|a-b|<c<a+b)\end{aligned} \tag{4.5}$$

and

$$\begin{aligned}\int_0^\infty tI_0(at)K_0(bt)K_0(ct)\,dt &= u_a/4\tilde{\Delta} \quad (c>a+b) \\ &= \varphi_a/4\Delta \quad (|a-b|<c<a+b),\end{aligned} \tag{4.6}$$

results formally very similar to that of Ref. 8. For $\nu=0$, we have

$$\int_0^\infty t I_\mu(at) K_0(bt) K_\mu(ct)\,dt$$

$$= \left(\frac{a}{c}\right)^\mu \frac{1}{8\tilde{\Delta}} \left\{ \sum_{n>0} \frac{(1-Y)^{n+1}}{\mu+n+1} - \sum_{n>0} \frac{X^{n+1}}{\mu+n+1} \right\}, \quad c>a+b,$$

$$= -i\left(\frac{a}{c}\right)^\mu \frac{1}{8\Delta} \left\{ \sum_{n>0} \frac{(1-Y)^{n+1}}{\mu+n+1} - \sum_{n>0} \frac{X^{n+1}}{\mu+n+1} \right\}, \tag{4.7}$$

$$|a-b|<c<a+b,$$

which reduces to (4.6) when $\mu = 0$.

Starting with (4.7), we derive

$$\int_0^\infty t K_0(at) K_0(bt) K_0(ct)$$

$$= -\frac{1}{4\tilde{\Delta}} \left\{ u_a \ln\left(\frac{a}{c}\right) - \sum_{n=0}^{\infty} \left(\frac{a}{c}\right)^{n+1} \frac{\sinh(n+1)u_b}{(n+1)^2} \right\}, \quad c>a+b,$$

$$= \frac{1}{4\Delta} \left\{ -\varphi_a \ln\left(\frac{a}{c}\right) + \sum_{n=0}^{\infty} \left(\frac{a}{c}\right)^{n+1} \frac{\sin(n+1)\varphi_b}{(n+1)^2} \right\},$$

$$|a-b|<c<a+b. \tag{4.8}$$

We notice that the summation on the right-hand side may be written as the integral

$$\int_{(a/c)e^{-u_b}}^{(a/c)e^{u_b}} \ln(1-z)\frac{dz}{z}$$

or

$$\int_{(a/c)e^{i\varphi_b}}^{(a/c)e^{-i\varphi_b}} \ln\frac{(1-z)}{z}dz,$$

respectively, as $|z|<1$ (c is the largest length). These results are summarized in Table III.

In case $a=c$ ($b<2a$), the last formula (4.8) reduces to

$$\int_0^\infty t\,[K_0(at)]^2 K_0(bt)\,dt = \frac{1}{4\Delta}\sum_{n>0} \frac{\sin(n+1)\varphi_b}{(n+1)^2}, \tag{4.9}$$

and the series is Lobachevski's function.[10]

As a last remark, we want to emphasize that the derivation method explained in Sec. IV increases considerably the number of integrals that can be calculated and consequently the number of F_4 functions that are the sum of products of functions in one variable. It may be used to get other integrals that those corresponding to $\alpha+\beta+1=\gamma+\gamma'\pm n$. As an example, we list some integrals related to the first class of factorization of the Appell function and the integral $\int_0^\infty t^{1-\nu} Z_\nu K_\nu K_\nu\, dt$, which was studied in Ref. 1:

$$\int_0^\infty dt\, t^{3-\nu} Z_\nu(at) K_\nu(bt) K_\nu(ct)$$

$$= \left(\frac{\partial}{\partial a} - \frac{\nu-1}{a}\right)\left(\frac{\partial}{\partial a} + \frac{\nu}{a}\right) \times \int_0^\infty t^{1-\nu} Z_\nu(at) K_\nu(bt) K_\nu(ct)\,dt$$

and more generally

$$\int_0^\infty t^{1+2n-\nu} Z_\nu(at) K_\nu(bt) K_\nu(ct)\,dt$$

$$= \left[\left(\frac{\partial}{\partial a} - \frac{\nu-1}{a}\right)\left(\frac{\partial}{\partial a} + \frac{\nu}{a}\right)\right]^n \times \int_0^\infty t^{1-\nu} Z_\nu(at) K_\nu(bt) K_\nu(ct)\,dt.$$

Other possibilities are

$$\int_0^\infty t^{3-\nu} Z_{\nu\mp 1}(at) K_{\nu\pm 1}(bt) K_\nu(ct)\,dt$$

$$= \epsilon_Z \left(\frac{\partial}{\partial b} \mp \frac{\nu}{b}\right)\left(\frac{\partial}{\partial a} \pm \frac{\nu}{a}\right) \times \int_0^\infty t^{1-\nu} Z_\nu(at) K_\nu(bt) K_\nu(ct)\,dt,$$

$$\int_0^\infty t^{3-\nu} Z_{\nu\mp 1}(at) K_{\nu\mp 1}(bt) K_\nu(ct)\,dt$$

$$= \epsilon_Z \left(\frac{\partial}{\partial a} \pm \frac{\nu}{a}\right)\left(\frac{\partial}{\partial b} \pm \frac{\nu}{b}\right) \times \int_0^\infty t^{1-\nu} Z_\nu(at) K_\nu(bt) K_\nu(ct)\,dt,$$

where $Z_\nu = I_\nu$, K_ν and $\epsilon_Z = \pm 1$ depending on whether Z_μ is I_μ or K_μ; we do not intend to carry out a more complete investigation.

ACKNOWLEDGMENTS

The authors are greatly indebted to Professor M. L. Mehta for his suggestions and comments for the redaction of the manuscript.

APPENDIX: CASE $\nu=\mu, \rho$; $b=a, c$

We derive briefly here some formulas where two of the three parameters a, b, c are equal.

1. Case $\mu=\nu$ and $a=b$

When $c>2a$, $1-2X=1-2Y=\sqrt{1-4a^2/c^2}$ and Eq. (3.3) simplifies into

$$\int_0^\infty I_\nu(at) K_\nu(at) K_\rho(ct)\,dt$$

$$= \frac{\Gamma((1+\rho)/2)\Gamma((1-\rho)/2)}{4c} \times \Gamma\left(\frac{1+\rho}{2}+\nu\right)\Gamma\left(\frac{1-\rho}{2}+\nu\right) \times P^{-\nu}_{(\rho-1)/2}\left(\sqrt{\frac{1-4a^2}{c^2}}\right) \times P^{-\nu}_{(\rho-1)/2}\left(-\sqrt{1-\frac{4a^2}{c^2}}\right). \tag{A1}$$

When $c<2a$ (triangle configuration) $1-2X=1-2Y=i\sqrt{4a^2/c^2-1}$ is pure imaginary. The Legendre functions $P^{-\nu}_{(\rho-1)/2}(\pm i\sqrt{4a^2/c^2-1})$ may be rewritten in terms of real quantities. Let define $\varphi=\varphi_c/2$, with $\sin\varphi=c/2a$, $\cos\varphi=\sqrt{1-c^2/4a^2}$. Then, from the Whipple formula,[7]

$$P^{\nu}_{(\rho-1)/2}(\pm i\cot\varphi) = \sqrt{\frac{2}{\pi}\frac{e^{i\pi}(\rho/2\mp\frac{1}{4})}{\Gamma(-\nu-(\rho-1)/2)}} \times\sqrt{\sin\varphi}\, Q^{-\rho/2}_{-\nu-1/2}(\cos\varphi\mp i0),$$

or, in terms of the Legendre function on the cut,

$$Q^{-\rho/2}_{-\nu-1/2}(\cos\varphi\mp i0) = [Q^{-\rho/2}_{-\nu-1/2}(\cos\varphi)\pm i(\pi/2)P^{-\rho/2}_{-\nu-1/2}(\cos\varphi)] \times e^{-i(\pi/4)\rho}e^{-i(\pi/4)\rho(1\pm 1)},$$

which finally leads to

$$\int_0^\infty I_\nu(at)K_\nu(at)K_\rho(ct)\,dt = \frac{1}{4\pi a}\frac{\Gamma((1+\rho)/2)\Gamma((1-\rho)/2)\Gamma((1-\rho)/2-\nu)}{\Gamma(1+\rho)/2-\nu)} \times\left|Q^{-\rho/2}_{\nu-1/2}\left(\sqrt{1-\frac{c^2}{4a^2}}\right) + i\frac{\pi}{2}P^{-\rho/2}_{\nu-1/2}\left(\sqrt{1-\frac{c^2}{4a^2}}\right)\right|^2. \quad \text{(A2)}$$

Corresponding expressions for $\int_0^\infty I_{-\nu}K_\nu K_\rho\,dt$ and $\int_0^\infty [K_\nu]^2K_\rho\,dt$ are straightforward and we do not indicate them here.

2. Case $\nu=\rho$ and $b=c$

For the sake of convergence, we are in a triangle configuration and $a<2c$. From (3.5), we get

$$\int_0^\infty I_\mu(at)[K_\rho(ct)]^2\,dt = \frac{1}{a}Q^{\rho}_{(\mu-1)/2}\left(i\sqrt{\frac{4c^2}{a^2}-1}\right) \times Q^{-\rho}_{(\mu-1)/2}\left(-i\sqrt{\frac{4c^2}{a^2}-1}\right).$$

Using again Whipple's formula,[7] we transform $Q^{\pm\rho}_{(\mu-1)/2}$ into Legendre functions $P^{-\mu/2}_{\rho-1/2}$ on the cut. We have

$$Q^{\pm\rho}_{(\mu-1)/2}(\pm i\cot\varphi) = \sqrt{\frac{\pi}{2}}\Gamma\left(\frac{\mu+1}{2}\pm\rho\right)e^{\pm i\pi(\rho-1/4)} \times P^{-\mu/2}_{\rho-1/2}(\cos\varphi\mp i0)\sqrt{\sin\varphi} = \sqrt{\frac{\pi}{2}}\Gamma\left(\frac{\mu+1}{2}\pm\rho\right)e^{\pm i\pi(\rho-1/4)}e^{\pm i(\pi/4)\mu} \times P^{-\mu/2}_{\rho-1/2}(\cos\varphi)\sqrt{\sin\varphi},$$

where $\varphi=\varphi_a/2$, $\sin\varphi=a/2c$. Whence

$$\int_0^\infty I_\mu(at)[K_\rho(ct)]^2dt = \frac{\pi}{4c}\Gamma\left(\frac{\mu+1}{2}+\rho\right)\Gamma\left(\frac{\mu+1}{2}-\rho\right) \times\left[P^{-\mu/2}_{\rho-1/2}\left(\sqrt{1-\frac{a^2}{4c^2}}\right)\right]^2. \quad \text{(A3)}$$

[1]A. Gervois and H. Navelet, J. Math. Phys. **27**, 682 (1986).
[2]W. N. Bailey, Proc. Lond. Math. Soc. **40**, 37 (1936).
[3]P. Appell and J. Kampe de Feriet, *Fonctions hypergéométriques et hypersphériques* (Gauthier-Villars, Paris, 1926).
[4]A. Gervois and H. Navelet, J. Math. Phys. **26**, 633, 645 (1985).
[5]J. L. Burchnall and T. W. Chaundy, Quart. J. Math. **11**, 249 (1940) Eq. (54); A. W. Niukkanen, J. Phys. A: Math. Gen. **18**, 1399 (1985).
[6]W. N. Bailey, J. London Math. Soc. **11**, 16 (1936), expression (3.3).
[7]See, for example, W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966), pp. 151ff.
[8]A. Gervois and H. Navelet, J. Math. Phys. **25**, 3350 (1984).
[9]A. L. Dixon and W. L. Ferrar, Quart. J. Math. **4**, 193 (1933).
[10]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965), p. 933, formula 8.261.

On the linearization problem for ultraspherical polynomials

B. Bassetti, E. Montaldi, and M. Raciti
Dipartimento di Fisica dell'Università degli Studi di Milano, I - 20133 Milano, Italy and Istituto Nazionale di Fisica Nucleare, I - 20133 Milano, Italy

(Received 25 July 1985; accepted for publication 31 October 1985)

A direct proof of a formula established by Bressoud in 1981 [D. M. Bressoud, SIAM J. Math. Anal. 12, 161 (1981)], equivalent to the linearization formula for the ultraspherical polynomials, is given. Some related results are briefly discussed.

I. INTRODUCTION

The linearization problem for a family of polynomials $\{A_n(x)\}$, orthogonal with respect to the weight $W(x)$ on the interval $(-1,1)$, amounts, as is well known, to the evaluation of the integrals

$$\int_{-1}^{1} dx\, W(x)\, A_m(x)\, A_n(x)\, A_r(x).$$

Apart from their mathematical interest, integrals of this type are very useful in physical applications. The most common example occurs in the quantum theory of angular momentum,[1] where $A_n(x)$ is the nth Legendre polynomial $P_n(x)$. More generally one can consider the case in which the $\{A_n(x)\}$ are ultraspherical or Gegenbauer polynomials.[2,3]

In recent years many investigations have been devoted to the properties of the so-called q-hypergeometric functions; a discussion on the subject, including physical applications, can be found in a survey paper by Andrews.[4]

In the present note we want to give a direct proof of the following result obtained by Bressoud[5] as the limiting case $(q\to 1)$ of the linearization formula for q-ultraspherical polynomials:

$$\begin{aligned}(1-2sx+s^2)^{-\lambda}(1-2tx+t^2)^{-\lambda} &= \sum_{m,n=0}^{\infty}\binom{m+n}{n}\frac{\Gamma(\lambda+m)\Gamma(\lambda+n)}{\Gamma(\lambda)\Gamma(\lambda+m+n)}\\ &\quad\times s^m t^n\, {}_2F_1(\lambda,2\lambda+m+n;\\ &\quad\lambda+m+n+1;st)\, C^{\lambda}_{m+n}(x),\end{aligned} \tag{1.1}$$

whence one easily obtains

$$\begin{aligned}\int_{-1}^{1} dx\,(1-x^2)^{\lambda-1/2}(1-2sx+s^2)^{-\lambda}&\\ \times(1-2tx+t^2)^{-\lambda}C^{\lambda}_r(x)&\\ =\frac{\pi 2^{1-2\lambda}\Gamma(2\lambda+r)}{r!(\lambda+r)\,[\Gamma(\lambda)]^2}\frac{(2\lambda)_r}{(\lambda)_r}t^r&\\ \times{}_2F_1(\lambda,2\lambda+r;\lambda+r+1;st)&\\ \times{}_2F_1(-r,\lambda;2\lambda;1-s/t).&\end{aligned} \tag{1.2}$$

Equation (1.2) is a rather unusual version of the linearization formula for the ultraspherical polynomials:

$$\begin{aligned}C^{\lambda}_m(x)\,C^{\lambda}_n(x) &= \sum_{r=0}^{\min(m,n)}\frac{m+n+\lambda-2r}{m+n+\lambda-r}\frac{(m+n-2r)!}{(2\lambda)_{m+n-2r}}\\ &\quad\times\frac{(\lambda)_r(\lambda)_{m-r}(\lambda)_{n-r}(2\lambda)_{m+n-r}}{r!(m-r)!(n-r)!(\lambda)_{m+n-r}}C^{\lambda}_{m+n-2r}(x),\end{aligned} \tag{1.3}$$

as can be checked by expanding both sides as a double series in s and t.

Although Eq. (1.3) is a standard result, whose history is presented, for instance, in a book by Askey,[6] we believe that a simple derivation of the more compact equation [(1.2)] is of some interest. This is done in Sec. II. Section III is devoted to a brief discussion of some related results.

II. THE PROOF OF EQ. (1.2)

The crucial remark is that the two functions ${}_2F_1$ can be rewritten in terms of suitable Legendre functions, whose product is essentially the Laplace transform of a product of two Bessel functions. These, in turn, arise in a quite natural way when one expands the exponential in Gegenbauer polynomials.

Having this in mind we first transform the left-hand side of Eq. (1.2) in the following way:

$$\begin{aligned}G^{\lambda}_r(s,t)&\equiv\int_{-1}^{1}dx\,(1-x^2)^{\lambda-1/2}(1-2sx+s^2)^{-\lambda}\\ &\quad\times(1-2tx+t^2)^{-\lambda}\,C^{\lambda}_r(x)\\ &=[\Gamma(\lambda)]^{-2}\int_0^{\infty}d\xi\,\xi^{\lambda-1}e^{-(1+s^2)\xi}\\ &\quad\times\int_0^{\infty}d\eta\,\eta^{\lambda-1}e^{-(1+t^2)\eta}\\ &\quad\times\int_{-1}^{1}dx(1-x^2)^{\lambda-1/2}e^{2(s\xi+t\eta)x}\,C^{\lambda}_r(x),\end{aligned}$$

whence, by integrating[7] over x, we arrive at

$$\begin{aligned}G^{\lambda}_r(s,t)&=\frac{\pi 2^{1-2\lambda}\Gamma(2\lambda+r)}{r![\Gamma(\lambda)]^3}\\ &\quad\times\int_0^{\infty}d\xi\int_0^{\infty}d\eta\,(s\xi+t\eta)^{-\lambda}(\xi\eta)^{\lambda-1}\\ &\quad\times e^{-[(1+s^2)\xi+(1+t^2)\eta]}I_{\lambda+r}(2(s\xi+t\eta)).\end{aligned} \tag{2.1}$$

Next we perform the change of variables

$$\xi=\Delta^{-1}[(1+t^2)\alpha-t\beta],\quad \eta=\Delta^{-1}[s\beta-(1+s^2)\alpha],$$

where

$$\Delta=s\,(1+t^2)-t\,(1+s^2)=(s-t)\,(1-st).$$

 0022-2488/86/030696-03$02.50

Without loss of generality we assume $0<t<s<1$ so that $\Delta>0$. Equation (2.1) now reads

$$G_r^\lambda(s,t)=\frac{\pi 2^{1-2\lambda}\Gamma(2\lambda+r)}{r![\Gamma(\lambda)]^3}(st)^{\lambda-1}\Delta^{1-2\lambda}\times\int_0^\infty d\alpha\,\alpha^{-\lambda}I_{\lambda+r}(2\alpha)\int_{[(1+s^2)/s]\alpha}^{[(1+t^2)/t]\alpha}d\beta\times\left(\frac{1+t^2}{t}\alpha-\beta\right)^{\lambda-1}\left(\beta-\frac{1+s^2}{s}\alpha\right)^{\lambda-1}e^{-\beta}. \tag{2.2}$$

The inner integral is simply Poisson's integral representation for Bessel functions,[8] and we get

$$G_r^\lambda(s,t)=\frac{\pi^{3/2}2^{1-2\lambda}\Gamma(2\lambda+r)}{r![\Gamma(\lambda)]^2}(st)^{-1/2}\Delta^{1/2-\lambda}\times\int_0^\infty d\alpha\,\alpha^{-1/2}I_{\lambda+r}(2\alpha)\,I_{\lambda-1/2}\left(\frac{\Delta}{2st}\alpha\right)e^{-A\alpha}, \tag{2.3}$$

where

$A=(s+t)\,(1+st)/2st.$

By evaluating the Laplace transform[9] we have

$$G_r^\lambda(s,t)=\frac{\pi^{3/2}2^{1-2\lambda}[\Gamma(2\lambda+r)]^2}{r![\Gamma(\lambda)]^2}\Delta^{1/2-\lambda}(st)^{-1/2}c^{1/2}\times P_{\lambda-1}^{-(\lambda+r)}(\cosh\tau)\,P_{\lambda+r-1/2}^{-(\lambda-1/2)}(\cosh\tau'), \tag{2.4}$$

where

$\sinh\tau=2c,\quad \sinh\tau'=(\Delta/2st)\,c,$

$\cosh\tau\cosh\tau'=Ac,$

or, equivalently,

$$c=\sqrt{st}/(1-st),\quad \cosh\tau=\frac{1+st}{1-st},\quad \cosh\tau'=\frac{s+t}{2\sqrt{st}}.$$

Finally, recalling that[10]

$$P_{\lambda-1}^{-(\lambda+r)}\left(\frac{1+st}{1-st}\right)=\frac{(st)^{(\lambda+r)/2}}{\Gamma(\lambda+r+1)}\,{}_2F_1\left(1-\lambda,\lambda;\lambda+r+1;-\frac{st}{1-st}\right)=\frac{(st)^{(\lambda+r)/2}}{\Gamma(\lambda+r+1)}(1-st)^\lambda\times{}_2F_1(\lambda,2\lambda+r;\lambda+r+1;st)$$

and

$$P_{\lambda+r-1/2}^{-(\lambda-1/2)}\left(\frac{s+t}{2\sqrt{st}}\right)=2^{-(\lambda-1/2)}\left(\frac{s-t}{2\sqrt{st}}\right)^{\lambda-1/2}\left(\frac{t}{s}\right)^{r/2}\times\frac{1}{\Gamma(\lambda+\frac12)}\,{}_2F_1\left(-r,\lambda;2\lambda;1-\frac{s}{t}\right),$$

Eq. (1.2) follows at once.

III. SOME OTHER RESULTS

We first observe that the rhs of Eq. (2.1) can be handled in a different way. By performing the change of variable $\eta=\xi\omega$ in the inner integral, and then integrating over ξ, we have

$$\int_0^\infty d\omega\,\omega^{\lambda-1}\frac{(s+t\omega)^r}{[(1+s)^2+(1+t)^2\omega]^{2\lambda+r}}\,{}_2F_1\left(2\lambda+r,\lambda+r+\tfrac12;2\lambda+2r+1;\frac{4(s+t\omega)}{(1+s)^2+(1+t)^2\omega}\right)=B(\lambda,\lambda)\,t^r\,{}_2F_1(\lambda,2\lambda+r;\lambda+r+1;st)\,{}_2F_1(-r,\lambda;2\lambda;1-s/t). \tag{3.1}$$

Next let us consider the case $s=t$ of Eq. (1.2):

$$\int_{-1}^1 dx\,(1-x^2)^{\lambda-1/2}(1-2tx+t^2)^{-2\lambda}C_r^\lambda(x)=\frac{\pi 2^{1-2\lambda}\Gamma(2\lambda+r)\,(2\lambda)_r}{r!(\lambda+r)\,[\Gamma(\lambda)]^2(\lambda)_r}t^r\times{}_2F_1(\lambda,2\lambda+r;\lambda+r+1;t^2), \tag{3.2}$$

or, equivalently

$$(1-2tx+t^2)^{-2\lambda}=\sum_{r=0}^\infty\frac{(2\lambda)_r}{(\lambda)_r}t^r\,{}_2F_1(\lambda,2\lambda+r;\lambda+r+1;t^2)\,C_r^\lambda(x). \tag{3.3}$$

It is interesting to derive this formula in a direct way.

To this aim we note that

$$\int_{-1}^1 dx\,(1-x^2)^{\lambda+r-1/2}C_{2n}^{2\lambda+r}(x)=\pi^{1/2}\frac{(\lambda)_n(2\lambda+r)_n}{n!}\frac{\Gamma(\lambda+r+\frac12)}{\Gamma(\lambda+r+n+1)}, \tag{3.4}$$

as one can verify by expanding $C_{2n}^{2\lambda+r}$ in powers of x^2, integrating term by term, and then summing the resulting ${}_3F_2$ by Watson's theorem.[11] The integral

$$\int_{-1}^1 dx\,(1-x^2)^{\lambda-1/2}C_{r+2n}^{2\lambda}(x)\,C_r^\lambda(x)=\frac{\pi 2^{1-2\lambda}\Gamma(2\lambda+r)\,(2\lambda)_r\,(\lambda)_n\,(2\lambda+r)_n}{r!(\lambda+r)\,[\Gamma(\lambda)]^2(\lambda)_r\,(\lambda+r+1)_n n!}, \tag{3.5}$$

needed to arrive at Eq. (3.2), follows by expressing C_r^λ through Rodrigues' formula, then integrating r times by parts, and finally using Eq. (3.4).

As a final remark we point out that Eq. (3.2), multiplying both sides by s^r and then summing over r, yields

$$\int_{-1}^1 dx\,(1-x^2)^{\lambda-1/2}(1-2tx+t^2)^{-2\lambda}(1-2sx+s^2)^{-\lambda}=2^{1-2\lambda}\sin(\pi\lambda)\int_0^1 du$$

$$\times u^{2\lambda-1}(1-u)^{\lambda}(1-t^2u)^{-\lambda}(1-stu)^{-2\lambda}$$
$$= B\left(\tfrac{1}{2}\lambda+\tfrac{1}{2}\right) F_1(2\lambda,\lambda,2\lambda;\lambda+1;t^2,st). \tag{3.6}$$

The second line of this equation has been obtained by employing, for the $_2F_1$ appearing in Eq. (3.2), Euler's integral representation; in the last line, use has been made of Picard's[12] single integral representation for the first Appell function F_1.

[1] E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge U. P., Cambridge, 1963).

[2] R. Askey, Lett. Math. Phys. **6**, 299 (1982).

[3] M. Rahman and M. J. Shah, J. Math. Phys. **26**, 627 (1985).

[4] G. E. Andrews, SIAM Rev. **16**, 441 (1974).

[5] D. M. Bressoud, SIAM J. Math. Anal. **12**, 161 (1981).

[6] R. Askey, *Orthogonal Polynomials and Special Functions* (SIAM, Philadelphia, 1975), Lecture 5.

[7] Bateman Manuscript Project, *Tables of Integral Transforms* (McGraw–Hill, New York, 1954), Vol. 2, p. 281, Eq. (7).

[8] Bateman Manuscript Project, *Higher Transcendental Functions* (McGraw–Hill, New York, 1953), Vol. 2, p. 81, Eq. (10).

[9] Bateman Manuscript Project, *Tables of Integral Transforms* (McGraw–Hill, New York, 1954), Vol. 1, p. 196, Eqs. (12) and (13). Formula (13) is readily obtained by expanding both the Bessel functions and integrating term by term. Equation (12) follows from the previous one as a particular case of a well-known factorization formula; see, for instance, W. N. Bailey, *Generalized Hypergeometric Functions* (Cambridge U. P., Cambridge, 1935), p. 81.

[10] Bateman Manuscript Project, *Higher Transcendental Functions* (McGraw–Hill, New York, 1953), Vol. 1, p. 122, Eq. (3), p. 130, Eq. (29).

[11] See Ref. 10, p. 189, Eq. (6).

[12] See Ref. 10, p. 231, Eq. (5).

Fractional approximation to the vacuum–vacuum amplitude of a ϕ^4-potential theory in zero dimensions

Eduardo Chalbaud and Pablo Martín
Universidad Simón Bolívar, Departamento de Física, Apdo. 80659, Caracas 1080A, Venezuela

(Received 3 April 1985; accepted for publication 22 October 1985)

Here, the vacuum–vacuum amplitude with a ϕ^4-potential in terms of the fractional approximation to the partition function of a zero-dimensional field theory is presented. This fractional approximation has been obtained from both the power series and the asymptotic expansion. The power series diverges, nonetheless the fractional approximations are excellent. All the approximations from first to seventh degree are presented, with maximum errors from 0.6% to 1.6×10^{-5}%, respectively.

I. INTRODUCTION

In a recent paper,[1] it has been shown that the Green's function for $U(N)$ invariant-matrix ϕ^4-theories in zero space-time dimensions can be expressed in terms of a class of orthogonal polynomials $P_n(\phi)$, which are orthogonal on the interval $(-\infty,\infty)$ with respect to the weight function $\exp(-m^2\phi^2-\lambda\phi^4)$. These polynomials are obtained in terms of the function $I_\beta(\alpha)$ and derivatives [see Eq. (3.7), Ref. 1]

$$I_\beta(\alpha)=\int_{-\infty}^{+\infty}\exp(-\alpha\phi^2-\beta\phi^4)d\phi. \tag{1}$$

This function is the partition function of a zero-dimensional field theory, and it is used in instanton techniques.[2] It is related to the modified Bessel function $K_{1/4}$; its calculation, as well as that of its derivatives, is cumbersome.[3] For this reason, in field theory calculations we are often restricted to its asymptotic form.[4] Recently a method of obtaining fractional approximations to Coulomb and Bessel functions has been published,[5–7] which allows an easy calculation of the functions with great accuracy. We have applied this method to the function $I_\beta(\alpha)$, and we have arranged the solution in such a way that the computation of its derivatives is also easily obtained. This is the first time that this method of obtaining fractional approximations has been applied to a function whose power series has a zero radius of convergence.

The procedure and determination of the coefficients of the approximation is described in Sec. II, where we also show explicitly how to obtain the derivatives. The results are discussed in Sec. III, and we conclude this paper with a short summary in Sec. IV.

II. DETERMINATION OF THE COEFFICIENTS OF THE APPROXIMATION

Here for convenience we change the notation in Ref. 1, and, making a trivial change of variables, we define

$$I(x)=\int_{-\infty}^{+\infty}\exp(-s^2-xs^4)ds. \tag{2}$$

The power series for this function is given by

$$I_p(x)=\sum_{n=0}^{\infty}a_nx^n,\quad a_n=\frac{(-1)^n}{n!}\Gamma\left(\frac{4n+1}{2}\right). \tag{3}$$

The radius of convergence of this series is zero, and for that reason the following asymptotic expansion is more often used:

$$I_a(x)=\frac{1}{x^{1/4}}\sum_{i=0}^{\infty}\frac{b_i}{x^i}+\frac{1}{x^{3/4}}\sum_{j=0}^{\infty}\frac{B_j}{x^j}, \tag{4a}$$

$$b_n=[(4n-3)/(16n^2-8n)]b_{n-1},\quad b_0=\tfrac{1}{2}\Gamma(\tfrac{1}{4}), \tag{4b}$$

$$B_n=[(4n-1)/(16n^2+8n)]B_{n-1},$$
$$B_0=-\tfrac{1}{2}\Gamma(\tfrac{3}{4}). \tag{4c}$$

However, from the point of view of the fractional approximation, the coefficients of the power series are as important as those of the asymptotic expansion. In spite of the zero radius of convergence of the power series, if a fractional approximation is determined with the asymptotic expansion only, its accuracy is very poor.

Because of the form of Eqs. (3) and (4), we will consider approximations of the form

$$Y(x)=\left[\frac{b_0}{(1+x)^{1/4}}\sum_{j=0}^{n}p_jx^j+\frac{B_0}{(1+x)^{3/4}}\sum_{i=0}^{n}p_ix^i\right]\times\left(\sum_{k=0}^{n}q_kx^k\right)^{-1}, \tag{5}$$

where all the polynomials are monic. This fractional approximation presents the following important features.

(i) For large x we obtain the factors $x^{-1/4}$ and $x^{-3/4}$ as in the asymptotic expansion.

(ii) For small x, because of the unity in the radical, the approximation is regular at $x=0$, and an increasing number of terms of the power series, Eq. (3), can be obtained using polynomials of higher degree.

(iii) Since the parameters of the denominator of both terms of our approximation are the same, the parameters of our fractional approximation are given in a unique way by simple linear algebraic equations.

This third feature is an improvement on our previous work on the Coulomb function,[5] where a free parameter had to be obtained by trial and error. To obtain the parameters of our approximation, we proceed as in Refs. 5 and 6. By using the Taylor expansion of our approximating form, Eq. (5), and equating the coefficients of equal powers of x with the power series, Eq. (3), we get a number of equations relating

TABLE I. P_i and p_i parameters of the polynomial numerators, and q_i parameters of the polynomial denominator of $Y(x)$.

$n=1$		
$q_0 = 0.251$	$p_0 = 0.626$	$P_0 = 1.126$
$n=2$		
$q_0 = 0.076\,4$	$p_0 = 0.298\,0$	$P_0 = 0.660\,7$
$q_1 = 0.722\,9$	$p_1 = 1.097\,9$	$P_1 = 1.597\,9$
$n=3$		
$q_0 = 0.009\,76$	$p_0 = 0.085\,84$	$P_0 = 0.225\,74$
$q_1 = 0.198\,32$	$p_1 = 0.493\,60$	$P_1 = 1.013\,70$
$q_2 = 0.919\,37$	$p_2 = 1.294\,37$	$P_2 = 1.794\,37$
$n=4$		
$q_0 = 0.002\,6757$	$p_0 = 0.040\,4574$	$P_0 = 0.111\,9601$
$q_1 = 0.079\,6896$	$p_1 = 0.285\,7118$	$P_1 = 0.674\,6508$
$q_2 = 0.619\,9074$	$p_2 = 1.128\,5700$	$P_2 = 1.933\,1757$
$q_3 = 1.488\,3780$	$p_3 = 1.863\,3780$	$P_3 = 2.363\,3780$

p_j, P_i, and q_k with the known a_n. Similarly, by using the asymptotic expansion of Eq. (5) and equating the coefficients of equal powers of $1/x$ with the asymptotic expansion, Eq. (4), we get two sets of equations: one set relating p_j and q_k to the known b_i, and another set relating P_i and q_k to the known B_j. As usual, we have to choose the same number of equations as the number of parameters to be determined. There are several possibilities, so we have to develop criteria for choosing the best set of equations. In order to obtain optimum accuracy, we have found that the following criteria gives the best results for approximations of order n $(n>1)$: We take $n+1$ equations from the power set and $2n-1$ equations from the asymptotic expansion, n of these from the b_i set and $n-1$ from the B_j set.

For the first-order approximation we have three parameters to determine, q_0, p_0, and P_0, so the best results are obtained if we take one equation from the power set, and two from the asymtotic set, one from the b_i set and one from the B_i set. In this case the set of linear equations can be simply solved symbolically, and we obtain

$$q_0 = (3b_0 + 7B_0)/8(a_0 - b_0 - B_0), \tag{6a}$$

$$p_0 = (3a_0 + 4B_0)/8(a_0 - b_0 - B_0), \tag{6b}$$

$$P_0 = (7a_0 - 4b_0)/8(a_0 - b_0 - B_0). \tag{6c}$$

For higher orders, the expressions become too cumbersome to be solved symbolically, and it is best to solve numerically. The results are presented in Sec. III. Also, since the b_i expansion dominates over the B_j expansion, for orders higher than 1, we take one more equation from the b_i set than from the B_j set, as stated above.

In order to obtain the derivatives of our function to any order, it is best to cast it in a pole-residue form

$$Y(x) = \frac{b_0}{(1+x)^{1/4}}\left(1 + \sum_{i=1}^{n} \frac{c_i}{(x+d_i)}\right) + \frac{B_0}{(1+x)^{3/4}}\left(1 + \sum_{i=1}^{n} \frac{C_i}{(x+d_i)}\right), \tag{7}$$

$$Y(x) = b_0 Y_1(x) + B_0 Y_2(x). \tag{7'}$$

The Leibnitz formula for the mth-order derivative of $Y_1(x)$ [and a corresponding expression for $Y_2(x)$],

$$D_m Y_1(x) = \sum_{l=0}^{n} \binom{m}{l} D_{m-l}\left(\frac{1}{(1+x)^{1/4}}\right) \times D_l\left(1 + \sum_{i=1}^{n} \frac{c_i}{(x+d_i)}\right), \tag{8}$$

can then be used straightforwardly to obtain the derivatives of any order in terms of the pole-residue parameters c_i, C_i, and d_i.

III. RESULTS

We have computed all approximations from first to seventh degree. The fractional parameters for the approximations of order 1 to 4 are listed in Table I, and the fractional parameters for the approximations of order 5 to 7 in Table II. The pole-residue parameters for these approximations are

TABLE II. P_i and p_i parameters of the polynomial numerators, and q_i parameters of the polynomial denominator of $Y(x)$.

$n=5$		
$q_0 = 2.277\,141\,9\times10^{-4}$	$p_0 = 8.318\,063\,39\times10^{-3}$	$P_0 = 0.023\,951\,711\,5$
$q_1 = 0.010\,695\,668$	$p_1 = 0.071\,046\,612\,7$	$P_1 = 0.191\,568\,447$
$q_2 = 0.151\,313\,739$	$p_2 = 0.418\,315\,358$	$P_2 = 0.900\,232\,545$
$q_3 = 0.794\,704\,645$	$p_3 = 1.337\,998\,77$	$P_3 = 2.188\,779\,83$
$q_4 = 1.580\,728\,79$	$p_4 = 1.955\,728\,78$	$P_4 = 2.455\,728\,78$
$n=6$		
$q_0 = 5.408\,356\,6\times10^{-5}$	$p_0 = 3.536\,767\,43\times10^{-3}$	$P_0 = 0.010\,307\,6919$
$q_1 = 3.270\,007\,0\times10^{-3}$	$p_1 = 0.033\,499\,909\,5$	$P_1 = 0.094\,887\,8859$
$q_2 = 0.062\,857\,400\,8$	$p_2 = 0.235\,159\,751$	$P_2 = 0.561\,765\,239$
$q_3 = 0.482\,123\,701$	$p_3 = 1.011\,824\,33$	$P_3 = 1.914\,110\,34$
$q_4 = 1.568\,420\,29$	$p_4 = 2.319\,713\,77$	$P_4 = 3.447\,827\,31$
$q_5 = 2.135\,393\,74$	$p_5 = 2.510\,393\,74$	$P_5 = 3.010\,393\,74$
$n=7$		
$q_0 = 3.855\,027\,53\times10^{-6}$	$p_0 = 6.184\,076\,76\times10^{-4}$	$P_0 = 1.818\,515\,51\times10^{-3}$
$q_1 = 3.211\,630\,63\times10^{-4}$	$p_1 = 6.695\,523\,71\times10^{-3}$	$P_1 = 0.019\,795\,6468$
$q_2 = 9.132\,658\,99\times10^{-3}$	$p_2 = 0.057\,358\,635\,2$	$P_2 = 0.152\,933\,401$
$q_3 = 0.112\,933\,141$	$p_3 = 0.342\,488\,726$	$P_3 = 0.766\,943\,155$
$q_4 = 0.652\,895\,540$	$p_4 = 1.261\,599\,34$	$P_4 = 2.280\,527\,57$
$q_5 = 1.790\,897\,02$	$p_5 = 2.575\,731\,66$	$P_5 = 3.748\,566\,74$
$q_6 = 2.224\,836\,83$	$p_6 = 2.599\,836\,83$	$P_6 = 3.099\,836\,82$

TABLE III. d_i poles and c_i,C_i residues of $Y(x)$.

$n=1$		
$d_1=0.251$	$c_1=$ 0.375	$c_1=$ 0.875
$n=2$		
$d_1=0.128\,5$	$c_1=$ 0.372 2	$C_1=$ 1.012 8
$d_2=0.594\,5$	$c_2=$ 0.002 8	$C_2=-0.137\,8$
$n=3$		
$d_1=0.070\,51$	$c_1=$ 0.683 42	$C_1=$ 1.948 06
$d_2=0.220\,29$	$c_2=-0.478\,00$	$C_2=-1.288\,84$
$d_3=0.628\,57$	$c_3=$ 0.169 58	$C_3=$ 0.215 78
$n=4$		
$d_1=0.052\,229\,55$	$c_1=$ 1.068 190 4	$C_1=$ 3.076 536 7
$d_2=0.141\,744\,9$	$c_2=-1.000\,195\,3$	$C_2=-2.755\,335\,6$
$d_3=0.407\,520\,4$	$c_3=$ 0.285 414 0	$C_3=$ 0.567 886 8
$d_4=0.886\,883\,2$	$c_4=$ 0.021 590 93	$C_4=-0.014\,087\,88$

given in Table III and Table IV, respectively.

For the first-degree approximation we took one equation from the power series and two from the asymptotic expansion (one each from the b_i and B_i expansions). This approximation was compared to the exact function, and this choice of parameters was found to be the best. The maximum error occurs at $x=0.5$ and is 0.6%. The error decreases rapidly with increasing x, and is already less than 0.03% at $x=10$. For the $n=2$ approximation we took, according to our prescription, three equations from the power set and three from the asymptotic expansion: two from the b_i set and one from the B_j set. The accuracy improves, as compared to the first degree approximation, and the maximum error occurs also at $x=0.5$, but is only 0.16%. The error also rapidly decreases with increasing x, and at $x=10$ it is less than 0.003%. The third-degree approximation has nine fractional parameters and here, as in all approximations of degree greater than 1, we followed our prescription to choose the equations, described in Sec. II. The maximum error now occurs at $x=0.2$, and it is only 0.01%. As above, the error also decreases rapidly with increasing x, and at $x=10$ is only 0.0001%. The fourth-degree approximation has a maximum error of 0.005% at $x=0.3$, and by $x=5$, it reproduces exactly six decimals.

The maximum errors of the approximations of fifth, sixth, and seventh order occur at $x=0.2$, $x=0.25$, and $x=0.2$, respectively, and are $3.3\times10^{-4}\%$, $2\times10^{-4}\%$, and $1.6\times10^{-5}\%$, respectively. The error in all these approximations decreases from the maximum error by more than an order of magnitude for $x<0.05$ and for $x>1$. We should point out that as we increase the order of our approximation from n to $n+1$, the maximum error always decreases, but it decreases by roughly a factor of 2 if n is odd, and by roughly a factor of 10 if n is even.

It is important to point out that though the power series diverges, it contains very valuable information for the fractional approximation. We have found that if we take fewer equations from the set of equations obtained from the power series than the optimum number described above ($n+1$ for $n>1$), the accuracy of the approximation markedly decreases.

IV. SUMMARY

We have presented a fractional approximation to the function $I_\beta(\alpha)$ of Ref. 1, which permits the computation of the vacuum–vacuum amplitude with a ϕ^4-potential. Since for this computation the function and its derivatives are employed, we have cast the approximation in a pole-residue form, from which derivatives of any order can be easily obtained in terms of these parameters. The parameters of all approximations from first to seventh degree have been presented, and their maximum error of 0.6%, 0.16%, 0.01%, 0.005%, $3.3\times10^{-4}\%$, $2\times10^{-4}\%$, and $1.6\times10^{-5}\%$, respectively, indicate that the amplitude can be obtained with great precision. We have also indicated the procedure to follow if we desire higher-degree approximations and have in-

TABLE IV. d_i poles and c_i,C_i residues of $Y(x)$.

$n=5$		
$d_1=0.037\,735\,354\,6$	$c_1=$ 2.633 187 71	$C_1=$ 7.642 316 00
$d_2=0.083\,684\,422\,0$	$c_2=-$ 3.307 527 74	$C_2=-9.369\,519\,76$
$d_3=0.197\,455\,000$	$c_3=$ 1.054 729 98	$C_3=$ 2.771 254 94
$d_4=0.449\,624\,886$	$c_4=-$ 0.063 484 111 7	$C_4=-0.206\,478\,512$
$d_5=0.812\,229\,122$	$c_5=$ 0.058 094 162 7	$C_5=$ 0.037 427 337 1
$n=6$		
$d_1=0.031\,249\,708\,0$	$c_1=$ 4.700 493 83	$C_1=$ 13.688 206 3
$d_2=0.064\,090\,291\,9$	$c_2=-$ 6.562 131 30	$C_2=-18.783\,326\,4$
$d_3=0.140\,927\,593$	$c_3=$ 2.405 312 69	$C_3=$ 6.585 435 12
$d_4=0.319\,739\,026$	$c_4=-$ 0.276 355 019	$C_4=-$ 0.723 359 755
$d_5=0.633\,718\,755$	$c_5=$ 0.093 873 509 0	$C_5=$ 0.108 410 934
$d_6=0.945\,668\,365$	$c_6=$ 0.013 806 289 0	$C_6=-$ $3.661\,936\,90\times10^{-4}$
$n=7$		
$d_1=0.025\,192\,377\,2$	$d_1=$ 12.743 721 7	$C_1=$ 37.226 569 1
$d_2=0.046\,583\,931\,6$	$d_2=-20.941\,238\,3$	$C_2=-60.498\,062\,9$
$d_3=0.089\,752\,292\,4$	$d_3=$ 10.067 112 0	$C_3=$ 28.415 060 8
$d_4=0.181\,416\,130$	$d_4=-$ 1.768 204 12	$C_4=-$ 4.749 908 58
$d_5=0.359\,216\,912$	$d_5=$ 0.214 222 566	$C_5=$ 0.461 293 036
$d_6=0.627\,124\,695$	$d_6=$ 0.333 109 499	$C_6=$ 0.010 398 363 9
$d_7=0.895\,550\,485$	$d_7=$ 0.026 075 157 8	$C_7=$ $9.650\,161\,20\times10^{-3}$

dicated roughly how the accuracy increases with n. In this paper we have extended the method of fractional approximations to the case where the power series has zero radius of convergence.

[1]P. A. Amundsen and P. H. Damgaard, J. Math. Phys. **25**, 1633 (1984).
[2]D. J. Wallace, *Proceedings of the Symposium on Nonlinear (Soliton) Structure and Dynamics in Condensed Matter*, edited by A. R. Bishop and T. Schneider (Springer, New York, 1978), pp. 104–115.
[3]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972), pp. 374–379.
[4]W. B. Yeung, J. Math. Phys. **21**, 477 (1980).
[5]E. Chalbaud and P. Martín, J. Math. Phys. **25**, 1268 (1984).
[6]P. Martín and A. Guerrero, J. Math. Phys. **26**, 705 (1985).
[7]P. Martín, G. Donoso, and J. Zamudio-Cristi, J. Math. Phys. **21**, 280 (1980).

On a particular transcendent solution of the Ernst system generalized on n fields

B. Leaute and G. Marcilhacy
Institut Henri Poincaré, Laboratoire de Physique Théorique, C.N.R.S., U.A. n° 769, 11, rue Pierre et Marie Curie, 75231 Paris Cedex 05, France

(Received 30 July 1985; accepted for publication 31 October 1985)

A particular solution, a function of a particular form of the fifth Painlevé transcendent, of the Ernst system generalized to n fields is determined, which characterizes both the stationary axially symmetric fields, the solution of the Einstein $(n-1)$ Maxwell equations, and one class of axially symmetric static self-dual $SU(n+1)$ Yang–Mills fields.

I. INTRODUCTION

In a recent paper, Gürses has shown that the system of the Einstein–$(n-1)$-Maxwell equations, valid for the case of stationary axially symmetric metrics, can also be understood as describing a particular class of static axially symmetric $SU(n+1)$ self-dual Yang–Mills fields.[1] The system of nonlinear partial differential equations, which plays a prominent part here, may be reformulated by means of some changes of unknown functions under a particularly compact form that constitutes a generalization to n complex fields ζ^a of the well-known Ernst system.[2,3]

We have investigated this generalized system and obtained a first form of solutions: $\zeta^a = \zeta^a(v)$, the n fields being determined as functions of one arbitrary harmonic function v. This solution covers, in particular, the vacuum gravitational case, when the metric is that of Papapetrou.[4]

The purpose of this article is to present another form of the solution that will be found by means of the method of separation of variables ρ and z. This procedure, which we have already used in previous papers on the Ernst equation,[5] will be fully developed here.

To start with we would like to briefly recall some aspects of Gürses' article, in particular the initial Einstein–$(n-1)$-Maxwell system and the Ernst n fields formulation, which is the object of our study. The intermediate stages will be omitted here because they have already been explained.[3] We will then develop the arguments and computations that lead to our new solution. The latter can be written under the form of a hyperspherical representation of the n complex fields ζ^a parametrized by a particular form of the fifth Painlevé transcendental function. Finally some particular solutions corresponding to the specific choices of the integration constants will be given.

II. RECALL OF THE FORMULATION OF THE BASIC SYSTEM

The coupled Einstein–Abelian gauge fields equations are given by

$$S_{\mu\nu} = \gamma_{ab}\left(F^a_{\mu\alpha}F^{b\alpha}_{\nu} - \tfrac{1}{4}g_{\mu\nu}F^a_{\alpha\beta}F^{b\alpha\beta}\right), \tag{1}$$

$$F^{a\mu\nu}_{;\nu} = 0, \tag{2}$$

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu, \tag{3}$$

where γ_{ab} is a diagonal matrix, which, by a suitable choice of the basis to the potential A^a_μ, can be taken as the Kronecker symbol δ_{ab}; $a, b,... = 1, 2,... (n-1)$ $(n>0)$. A semicolon denotes the covariant differentiation with respect to the Riemannian connection. The considered space-time is stationary and axially symmetric; its metric can therefore be written under the form

$$ds^2 = f(dt + \omega\, d\varphi)^2 - f^{-1}\left[e^{\lambda\gamma}(d\rho^2 + dz^2) + \rho^2\, dz^2\right], \tag{4}$$

where the functions f, ω, γ depend only on the ρ and z coordinates. It is again assumed the gauge potential one-form A^a has the two components

$$A^a \equiv A^a_\mu\, dx^\mu = A^a_t\, dt + A^a_\varphi\, d\varphi, \tag{5}$$

which also depend only on ρ and z.

The system of field equations (1)–(3) is explicitly written by taking into account the assumptions made about the metric (4) and the gauge fields (5). In our work[3] we have described how, following Ref. 1, the system obtained could be analyzed and reformulated in a manner similar to that of Ernst in the Einstein–Maxell case.[2] We shall not repeat all the details of the analysis here, but limit ourselves to relating the result, namely, that it is possible to introduce a set of n complex functions ζ^a (a: from 1 to n) of variables ρ and z governed by the equations

$$\Lambda\nabla^2\zeta^a = 2\zeta^{b*}\nabla\zeta^b\cdot\nabla\zeta^a, \tag{6}$$

with

$$\Lambda \equiv \zeta^b\zeta^{b*} - 1,$$

where ∇, ∇^2 denote, respectively, the gradient and Laplacian operators in cylindrical coordinates (ρ, z) and relatively to the flat tridimensional metric; the symbol * denotes the complex conjugation. The summation on the repeated indices is applied to numeration indices $a, b,...$ of the fields.

This system is basic in researching the solutions of the field equations (1)–(3); a simple examination shows that it is a generalization to n fields ζ^a of the well-known Ernst system.

III. PARTICULAR TRANSCENDENTAL SOLUTION

We intend to research a particular solution of the system (6) by using the method of separation of variables. Consequently, we put

$$\zeta^a(\rho,z) = \chi^a(\rho)\lambda^a(z). \tag{7}$$

By inserting this expression in (6) we see that we have necessarily

 0022-2488/86/030703-04$02.50

$$\lambda^a(z) = e^{i\alpha_a z}, \tag{8}$$

where $\{\alpha_a\}$ denotes a set of n real constants. Taking into account of this result, the system (6) can be explicitly written as

$$\Lambda(\chi^{a''} + (1/\rho)\chi^{a'} - \alpha_a^2\chi) = 2(\chi^{b*}\chi^{b'}\chi^{a'} - \alpha_a\alpha_b\chi^{b*}\chi^b\chi^a), \tag{9}$$

with $\Lambda \equiv \chi^a\chi^{a*} - 1$ and $' \equiv \frac{d}{d\rho}$.

By then multiplying the two members, in a contracted manner, by ζ^{a*} and by taking the imaginary part of the result, we find a differential equation that can be integrated and we thus obtain the first integral

$$\chi^{a'}\chi^{a*} - \chi^a\chi^{a*'} = 2ia\Lambda^2, \tag{10}$$

a being a first real constant of integration. By again multiplying the two members of (9) by χ^{a*} but *without summation on the repeated index a* and by taking the imaginary part of the result and with the help of the relation (10), we arrive again at a differential equation that can be integrated to obtain a set of n first integrals:

$$\chi^{a'}\chi^{a*} - \chi^a\chi^{a*'} = 2i\Lambda(a\chi^{\underline{a}}\chi^{\underline{a}*} + b_{\underline{a}}). \tag{11}$$

To indicate the nonsummation we underline the repeated index a. We have a set n real constants of integration $\{b_a\}$ that are satisfactory, given (10):

$$a + \sum_{a=1}^{n} b_a = 0. \tag{12}$$

To pursue the calculation it is advisable to take into account the relation (11) in the system (9). But it is possible to progress only by adding the assumption that *all constants* α_a, introduced in (8), *are equal*: $\alpha_a = \alpha$, with a from 1 to n.

In order to find a system where each equation governs only, in fact, one of the unknown functions, we are led to make the following change of functions:

$$\begin{aligned}
\chi^1 &= \Omega\cos\omega_1 e^{iu_1},\\
\chi^2 &= \Omega\sin\omega_1\cos\omega_2 e^{iu_1},\\
\chi^{n-1} &= \Omega\sin\omega_1\sin\omega_2\cdots\cos\omega_{n-1}e^{iu_{n-1}},\\
\chi^n &= \Omega\sin\omega_1\sin\omega_2\cdots\sin\omega_{n-1}e^{iu_n}.
\end{aligned} \tag{13}$$

We thus define $2n$ real functions $\{\Omega;\omega_1,\ldots,\omega_{n-1};u_1,\ldots,u_n\}$ of the variable ρ. The first integrals (11) now may be written

$$\begin{aligned}
u_1' &= \frac{\Lambda}{\rho}\left(a + \frac{b_1}{(\Omega\cos\omega_1)^2}\right),\\
u_2' &= \frac{\Lambda}{\rho}\left(a + \frac{b_2}{(\Omega\sin\omega_1\cos\omega_2)^2}\right),\\
u_n' &= \frac{\Lambda}{\rho}\left(a + \frac{b_n}{(\Omega\sin\omega_1\cdots\sin\omega_{n-1})^2}\right),
\end{aligned} \tag{14}$$

with $\Lambda \equiv \Omega^2 - 1$.

By taking these expressions into account, Eq. (9) can be reformulated in terms of functions $(\Omega;\omega_i)$ only. In so doing, we obtain a system of n differential equations that then can be combined linearly so as to obtain another equivalent of equations where each of these contains only one function twice derived. Hence by straightforward calculation it is possible to determine successively $n-1$ first integrals. To observe this fact let us consider, for example, $n=3$. We then find the following equations:

$$\Lambda\Omega\left(\omega_2'' + 2\frac{\cos\omega_1}{\sin\omega_1}\omega_1'\omega_2' + \frac{1}{\rho}\omega_2'\right) - 2\Omega'\omega_2' + \frac{\Lambda^3}{\rho^2\Omega^3\sin^4\omega_1}\left(b_2^2\frac{\sin\omega_2}{\cos^3\omega_2} - b_3^2\frac{\cos\omega_2}{\sin^3\omega}\right) = 0, \tag{15}$$

$$\Lambda\Omega\left(\omega_1'' - \sin\omega_1\cos\omega_1\omega_2'^2 + \frac{1}{\rho}\omega_1'\right) - 2\Omega'\omega_1' + \frac{\Lambda^3}{\rho^2\Omega^3}\times\left[b_1^2\frac{\sin\omega_1}{\cos^3\omega_1} - \frac{\cos\omega_1}{\sin^3\omega_1}\left(\frac{b_2^2}{\cos^2\omega_2} + \frac{b_3^2}{\sin^2\omega_2}\right)\right] = 0, \tag{16}$$

$$\Lambda[\Omega'' + (1/\rho)\Omega' - \Omega(\omega_1'^2 + \sin^2\omega_1\omega_2'^2)] - 2\Omega\Omega'^2 + \alpha^2(1+\Omega^2)\Omega + \frac{\Lambda^3}{\rho^2\Omega^3}\left[a^2\Omega^4 - \frac{b_1^2}{\cos^2\omega_1} - \frac{1}{\sin^2\omega_1}\left(\frac{b_2^2}{\cos^2\omega_2} + \frac{b_3^2}{\sin^2\omega_2}\right)\right] = 0. \tag{17}$$

By investigating Eq. (15) we verify that it admits the first integral

$$\omega_2'^2 = \frac{\Lambda^2}{\rho^2(\Omega\sin\omega_1)^4}\left(k_2^2 - \frac{b_2^2}{\cos^2\omega_2} - \frac{b_3^2}{\sin^2\omega_2}\right). \tag{18}$$

Inserting this quantity in (16) we show also that the resulting equation admits the first integral

$$\omega_1'^2 = \frac{\Lambda^2}{\rho^2\Omega^4}\left(k_1^2 - \frac{b_1^2}{\cos^2\omega_1} - \frac{k_2^2}{\sin^2\omega_1}\right). \tag{19}$$

It is clear that, without further demonstration, this process can be easily generalized from 3 to n fields; the sequence of $n-1$ first integrals then being

$$\begin{aligned}
\omega_{n-1}'^2 &= \frac{\Lambda^2}{\rho^2(\Omega\sin\omega_1\cdots\sin\omega_{n-1})^4}\times\left(k_{n-1}^2 - \frac{b_{n-1}^2}{\cos^2\omega_{n-1}} - \frac{b_n^2}{\sin^2\omega_{n-1}}\right),\\
\omega_{n-2}'^2 &= \frac{\Lambda^2}{\rho^2(\Omega\sin\omega_1\cdots\sin\omega_{n-2})^4}\times\left(k_{n-2}^2 - \frac{b_{n-2}^2}{\cos^2\omega_{n-2}} - \frac{k_{n-1}^2}{\sin^2\omega_{n-2}}\right),\\
\omega_2'^2 &= \frac{\Lambda^2}{\rho^2(\Omega\sin\omega_1)^4}\left(k_2^2 - \frac{b_2^2}{\cos^2\omega_2} - \frac{k_3^2}{\sin^2\omega_2}\right),\\
\omega_1'^2 &= \frac{\Lambda^2}{\rho^2\Omega^4}\left(k_1^2 - \frac{b_1^2}{\cos^2\omega_1} - \frac{k_2^2}{\sin^2\omega_1}\right).
\end{aligned} \tag{20}$$

In these expressions there are $n-1$ constants of integration that must all be chosen positive, from which we derive the notation $(k_1^2,\ldots,k_{n-1}^2)$. The insertion of results (18) and (19) in (17) gives

$$\Omega'' + \frac{1}{\rho}\Omega' - 2\frac{\Omega\Omega'^2}{\Omega^2-1} + \frac{1}{\rho^2}(\Omega^2-1)^2\left(a^2\Omega - \frac{k_1^2}{\Omega^3}\right) + \alpha^2\frac{\Omega^2+1}{\Omega^2-1}\Omega = 0 \tag{21}$$

as the equation governing the function $\Omega(\rho)$. We can easily establish that this equation, proved formally for three fields, is, in fact, independent of the number n of fields. The $n-1$ first integrals (20) arise only from the last integration constant k_1^2 in (21). By the change

$$Y=\Omega^2, \tag{22}$$

we finally obtain the equation

$$Y''+\frac{1}{\rho}Y'-\left(\frac{1}{2Y}+\frac{1}{Y-1}\right)Y'^2 +\frac{2}{\rho^2}(Y-1)^2\left(a^2Y-\frac{k_1^2}{Y}\right)+2\alpha^2\frac{Y+1}{Y-1}Y=0, \tag{23}$$

in which we identify a *particular form of the differential equation that defines the fifth Painlevé transcendent* with the parameters $\underline{\alpha}=-2a^2$, $\underline{\beta}=2k_1^2$, $\underline{\gamma}\equiv 0$, $\underline{\delta}=-2\alpha^2$ nonidentically zero[6,7] (see commentary in the last paragraph).

The functions $(\omega_1,...,\omega_{n-1})$ and $(u_1,...,u_n)$, characterized by the first integrals (20) and (14), are then connected functionally to the Painlevé transcendent $Y(\rho)$. For example, we thus have, for $\omega_1(\rho)$,

$$\arcsin U_1=\frac{\epsilon}{k_1}\int_{\rho_0}^{\rho}\frac{Y-1}{Y}\frac{d\rho}{\rho},\quad \text{with } \epsilon\equiv\pm 1, \tag{24}$$

and

$$U_1=(1/\nu)(\cos^2\omega_1-\mu),$$

$$\mu\equiv\frac{k_1^2+b_1^2-k_2^2}{2k_1^2},\quad \nu\equiv\left(\mu^2-\frac{b_1^2}{k_1^2}\right)^{1/2}.$$

Thus at the end of this development it appears that the n complex fields ζ^a governed by the generalized Ernst system (6), by assumption of the form $\zeta^a(\rho,z)=\chi^a(\rho)e^{i\alpha z}$, are determined as functionals of $Y=\zeta^a\zeta^{a*}$ defined, by Eq. (23), as the fifth transcendental function of Painlevé.

IV. ON SOME PARTICULAR CASES—COMMENTARY

The consideration of first integrals (14) and (20) and Eq. (21) leads us to examine the possibility of various particular cases corresponding to specific choices of various constants (b_i), (k_i^2), and α. Let us look at some eventualities.

(i) When a particular constant b_i is made equal to zero it is without significant effect on the nature of solutions. But if all the constants b_i are made equal to zero and a also, in virtue of (12), the functions (u_i) are then constants; as a consequence of (7) all the fields ζ^a have the same phase u, which is a linear function of ρ.

Another notable eventuality: a choice of (b_i) such as $a=-\Sigma_{i=1}^n b_i=0$. In this situation, as in the previous one, the Painlevé transcendent defined by (22), has only two nonzero parameters.

(ii) The equations governing the functions (ω_i) admit particular constant solutions; we can see that in the example of (15) and (16). Equation (15) admits the particular solution $\omega_2=$ const defined by

$$\tan^2\omega_2=b_3/b_2. \tag{25}$$

This imposes, of course, the same sign for b_2 and b_3. If this particular solution is used, we also have the possibility of $\omega_1=$ const, with

$$\tan^2\omega_1=(b_2+b_3)/b_1, \tag{26}$$

where b_2+b_3 and b_1 have the same sign. This process can naturally be extended to the general case with n fields, where, beginning by ω_{n-1}, it is possible to find successively a sequence of constant solutions (ω_i) characterized by relations of the same type as the previous ones [(25) and (26)].

(iii) We finally envisage the situation when the constant α is taken equal to zero. In this case all the fields ζ^a depend only on the radial coordinate ρ. Instead of this variable ρ we introduce v, defined by $v=\log\rho$, which is a harmonic function; consequently we have to refer back to our previous study.[3] Let us point out, however, some elements of the integration process in the formalism used here. Equation (23), with $\alpha=0$, can be written in terms of the variable v:

$$\frac{d^2Y}{dv^2}-\left(\frac{1}{2Y}+\frac{1}{Y-1}\right)\left(\frac{dY}{dv}\right)^2 +2(Y-1)^2\left(a^2Y-\frac{k_1^2}{Y}\right)=0. \tag{27}$$

We recognize here a particular form of Eq. 38 of the classification of Painlevé and Gambier.[6,7] Thus the following first integral can immediately be written as

$$\left(\frac{dY}{dv}\right)^2=4(Y-1)^2(-a^2Y^2+2KY-k_1^2), \tag{28}$$

where K is an integration constant. The calculation is pursued using several ways based on the sign of trinom: $a^2+2K+k_1^2$.

If $a^2+2K+k_1^2>0$, we have

$$2\lambda(v-v_0)=\arcsin\left\{\frac{(a^2+K)(Y-1)-\lambda^2}{(K^2-a^2k_1^2)^{1/2}|Y-1|}\right\}, \tag{29}$$

EL4 with $\lambda\equiv(a^2+2K+k_1^2)^{1/2}$.

If $a^2+2K+k_1^2<0$, we have

$$2\lambda_1(v-v_0)=\log\left\{\frac{\gamma Y+\delta+(-a^2Y+2KY-k_1^2)^{1/2}}{|Y-1|}\right\}, \tag{30}$$

with $\lambda_1\equiv[-(a^2+2K+k_1^2)]^{1/2}$,

$\gamma\equiv(a^2+K)/\lambda_1,\quad \delta\equiv\lambda_1+\gamma,$

and v_0 an integration constant.

There are again other more particular cases that we omit. Taking into account (27), we can also give a more explicit form for the expression (24):

$$\arcsin U_1=\frac{\epsilon}{2k_1}\int_{Y_0}^{Y}\frac{dY}{Y(-a^2Y+2KY-k_1^2)^{1/2}} \tag{31}$$

or

$$\arcsin U_1=\frac{\epsilon}{2k_1}\left\{\arcsin\left(\frac{k_1^2-KY}{(K^2-k_1^2a^2)^{1/2}Y}\right)+\text{const}\right\}. \tag{32}$$

The calculation of the phases $u_1,...,u_n$ also could be performed but we think the above statement will suffice.

The values $n=1$ and $n=2$ correspond, respectively, to the vacuum Einstein and Einstein–Maxwell cases. The separable solutions of Ernst' equations (6) connected with these situations are known thanks to the investigations of

Morris and Dodd ($n = 1$) and Lakshmanan and Kaliappan ($n = 2$), which have proved the functional dependence of these solutions on the fifth Painlevé transcendent.[8,9]

[1]M. Gürses, Phys. Rev. D **30**, 486 (1984).
[2]F. J. Ernst, Phys. Rev. **167**, 1175 (1968); **168**, 1415 (1968).
[3]B. Léauté and G. Marcilhacy, J. Math. Phys. **26**, 1938 (1985).
[4]A. Papapetrou, Ann. Phys. Leipzig **12**, 309 (1953).
[5]B. Léauté and G. Marcilhacy, Phys. Lett. A **87**, 159 (1982); Lett. Nuovo Cimento **40**, 102 (1984).
[6]P. Painlevé, Acta Math. **25**, 1 (1902), B. Gambier, Acta Math. **33**, 1 (1909).
[7]B. E. Ince, *Ordinary Differential Equations* (Dover, New York, 1956), Chap. XIV.
[8]H. Morris and R. Dodd, Phys. Lett. A **75**, 249 (1980).
[9]M. Lakshmanan and P. Kaliappan, J. Math. Phys. **24**, 795 (1983).

To the complete integrability of long-wave–short-wave interaction equations

A. Roy Chowdhury[a)]
International Center for Theoretical Physics, Trieste, Italy
Pranab K. Chanda
High Energy Physics Division, Department of Physics, Jadavpur University, Calcutta, India

(Received 14 June 1985; accepted for publication 28 October 1985)

It is shown that the nonlinear partial differential equations governing the interaction of long and short waves are completely integrable. The methodology used is that of Ablowitz *et al.* [M. J. Ablowitz, A. Ramani, and A. Segur, Lett. Nuovo Cimento **23**, 333 (1980); M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715, 1006 (1980)], though in the last section of our paper the problem also has been discussed in the light of the procedure due to Weiss *et al.* [J. Weiss, M. Tabor, and G. Carnevale, J. Math. Phys. **24**, 522 (1983)] and a Backlünd transformation has been obtained.

I. INTRODUCTION

In recent years there have been extensive studies for the understanding of the complete integrability of nonlinear partial differential equations. Of late two methods have been advocated—one is the technique of Ablowitz *et al.*[12] and the other is that of Weiss *et al.*[3] Although these methods differ in the actual mode of calculation, and sometimes in the finer details of the results, in principle, both of these methods are the same. So here we initially apply the former approach and show how the "resonances" are formed and the arbitrariness of the expansion coefficients together with that of the wave front are deduced. We then discuss how our results fit in the formalism of Weiss *et al.*

II. THEORY

The nonlinear equations under consideration read (Newell[4,5])

$$A_t = 2S(BC)_x,$$

$$B_t = 2iB_{xx} = K_3 A_x B + \overline{K}_3 AB_x + iK_4 A^2 B - 2iSB^2 C, \qquad (1)$$

$$C_t = 2iC_{xx} = \overline{K}_3 A_x C - K_4 AC_x - iK_4 A^2 C + 2iSBC^2.$$

The third of this set is really the complex conjugate of the second; that is $B^* = C$, and here we have followed the notation of Ablowitz and Segur.[6]

Following the procedure of Ablowitz *et al.*[1,2]

$$A = \phi^\alpha \sum a_j \phi^j(x,t), \qquad (2a)$$

$$B = \phi^\beta \sum b_j \phi^j(x,t), \qquad (2b)$$

$$C = \phi^\gamma \sum c_j \phi^j(x,t). \qquad (2c)$$

To determine the dominant behavior we initially assume

$$A \sim \phi^\alpha a_0, \quad B \sim \phi^\beta b_0, \quad C \sim \phi^\gamma c_0.$$

So matching the most singular terms in (1) for $\phi(x,t) = 0$ we get $\alpha = -2, \beta = \gamma = -1$. Then corresponding coefficients are related as:

$$\phi_t a_0 = 4\phi_x^3, \qquad (3)$$
$$Sb_0 c_0 = 2\phi_x^2.$$

Now to determine the next-to-leading order terms we set

$$A \sim a_0 \phi^{-2} + a_r \phi^{r-2},$$
$$B \sim b_0 \phi^{-1} + b_r \phi^{r-1}, \qquad (4)$$
$$C \sim c_0 \phi^{-1} + c_r \phi^{r-1}$$

in the "reduced" set of equations and obtain

$$[(r-2)\phi_t]a_r + [-2S(r-2)c_0\phi_x]b_r - 2S(r-2)b_0\phi_x c_r = 0,$$
$$2[(r-1)(r-2)\phi_x^2 - 2Sb_0c_0]b_r - 2Sb_0^2 c_r = 0, \qquad (5)$$
$$+2[(r-1)(r-2)\phi_x^2 - 2Sb_0c_0]c_r - 2Sc_0^2 b_r = 0.$$

This set of homogeneous equation can have a nonvanishing solution only if the determinant is zero, that is,

$$\begin{vmatrix} (r-2)\phi_t & -2S(r-2)c_0\phi_x & -2S(r-2)b_0\phi_x \\ 0 & 2i[(r-1)(r-2) - 2Sb_0c_0] & -2iSb_0 \\ 0 & -2iSc_0^2 & 2i[(r-1)(r-2)\phi_x^2 - 2Sb_0c_0] \end{vmatrix} = 0.$$

Using Eqs. (3) we reduce this to the two equations

$$(r-1)(r-2) - 4 = \pm 2,$$

[a)] Permanent address: High Energy Physics Division, Department of Physics, Jadavpur University, Calcutta, India.

 0022-2488/86/030707-03$02.50

so either $r(r-3)=0$ or $(r-4)(r+1)=0$. That is, $r=0, 3, 4$, and -1. As has been elaborately discussed in the papers by Ablowitz *et al.*[1,2] the resonance at $r=-1$ corresponds to the arbitrariness of the wave front.

We now proceed to determine the coefficients at the resonance positions. With no loss of generality we assume that $\phi(x,t)=x-f(t)$ and all the coefficients a_j, b_j, and c_j are functions of t only. We then have

$$a_0=-(4/f_t),\quad b_0 \ \text{arbitrary},\quad c_0=2/Sb_0. \tag{6}$$

For $j>1$ we now consider the recurrence relation obtained by linearization with respect to the nonleading terms

$$\begin{pmatrix} (j-2)f_t & 2S(j-2)c_0 & 2S(j-2)b_0 \\ 0 & 2i[(j-1)(j-2)-2Sb_0c_0] & -2iSb_0^2 \\ 0 & -2iSc_0^2 & 2i[(j-1)(j-2)-2Sb_0c_0] \end{pmatrix}\begin{pmatrix} a_j \\ b_j \\ c_i \end{pmatrix}=\begin{pmatrix} 0 \\ -(j-2)b_{j-1}f_t \\ -(j-2)c_{j-1}f_t \end{pmatrix}, \tag{7}$$

when $j=1$ we have

$$\begin{pmatrix} -f_t & -2Sc_0 & -2Sb_0 \\ 0 & -4iSb_0c_0 & -2iSb_0^2 \\ 0 & -2iSc_0^2 & -4iSb_0c_0 \end{pmatrix}\begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix}=\begin{pmatrix} 0 \\ b_0f_t \\ c_0f_t \end{pmatrix}, \tag{8}$$

which easily yields

$$a_1=-2i/3(\text{const});\quad b_1=if_tb_0/12;\quad c_1=if_t/6Sb_0. \tag{9}$$

Now for the first resonance $j=2$, we have

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & -8i & -2iSb_0^2 \\ 0 & 8i/Sb_0^2 & 8i \end{pmatrix}\begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{10}$$

This leads to

$$a_2=\text{arbitrary function of } t,\quad b_2=c_2=0. \tag{11}$$

At the second resonance for $j=3$,

$$\begin{pmatrix} f_t & 4/b_0 & 2Sib_0 \\ 0 & -4i & -2iSb_0^2 \\ 0 & 8i/Sb_0^2 & 4i \end{pmatrix}\begin{pmatrix} a_3 \\ b_3 \\ c_3 \end{pmatrix}=\begin{pmatrix} 0 \\ -b_2f_t \\ -c_2f_t \end{pmatrix}, \tag{12}$$

which yields

$$a_3=0,\quad b_3=\text{arbitrary function of } t,\quad c_3=2b_3/Sb_0^2. \tag{13}$$

For the third resonance at $j=4$ we get

$$\begin{pmatrix} 2f_t & 8/b_0 & 4Sb_0 \\ 0 & 4i & -2iSb_0^2 \\ 0 & 8i/Sb_0^2 & -4i \end{pmatrix}\begin{pmatrix} a_4 \\ b_4 \\ c_4 \end{pmatrix}=\begin{pmatrix} 0 \\ -2b_3f_t \\ -2c_3f_t \end{pmatrix}. \tag{14}$$

An interesting result that follows from (14) is that we again get

$$c_3=2b_3/Sb_0^2$$

as from the previous set.

We also deduce that

$$c_4=(1/2iSb_0^2)(4ib_4+2b_3f_t),$$
$$a_4=8b_4/f_tb_0+2ib_3/b_0, \tag{15}$$
$$b_4=\text{arbitrary function of time.}$$

At this point it is perhaps not out of place to discuss the formation of the "resonance coefficients," if the form of the wave front $\phi(x,t)$ had been different. Actually we also repeated the whole calculation with $\phi(x,t)=t+g(x)$ and observed the following results:

$$c_1=(b_0+12ib_{0x}g_x-18ib_0g_{xx})/6iSb_0^2,$$
$$b_1=(b_0-12ig_xb_{0x}+18ib_0g_{xx})/6iSb_0^2, \tag{16}$$
$$a_1=(2i/3)g_x-4g_xg_{xx},$$
$$a_2=\text{arbitrary function of } x,$$
$$b_2=(Sb_0^2c_{0xx}+4b_{0xx}g_x^2)/12g_x^4, \tag{17}$$
$$c_2=(Sb_0^2c_{0xx}-b_{0xx}g_x^2)/3Sb_0^2g_x^2,$$

together with $c_0=2g_x^2/Sb_0$. We will not quote further results about other a_j, b_j, and c_j's because they are quite cumbersome.

In our above calculations we have clearly demonstrated that the set of coupled nonlinear partial differential equations describing the interaction of long and short waves are actually completely integrable and conform to the Painlevé test.

Before concluding we show that it is also possible to consider these equations following the procedure suggested by Weiss *et al.*[3] Substituting the full series (2a)–(2c) in (1) we get

$$\sum A_{jt}\phi^{j-2}+\sum(j-2)A_j\phi^{j-3}\phi_t=2S\sum_j\sum_m B_jC_{mx}\phi^{j+m-2}$$
$$+2S\sum_j\sum_m B_{jx}C_m\phi^{j+m-2}+2S\sum\sum(m+j-2)B_jC_m\phi^{j+m-3}\phi_x, \tag{18}$$

$$\sum B_{jt}\,\phi^{j-1}+\sum (j-1)B_j\,\phi^{j-2}\phi_t-2i\sum B_{jxx}\,\phi^{j-1}-4i\sum (j-1)B_{jx}\,\phi^{j-2}\phi_x-2i\sum (j-1)(j-2)B_j\,\phi^{j-3}\phi_x^2$$

$$-2i\sum (j-1)B_j\,\phi^{j-2}\phi_{xx}=K_3\sum A_{jx}B_x\,\phi^{j+m-3}+K_3\sum\sum (j-2)A_jB_m\,\phi^{j+m-4}\phi_x-\overline{K}_3\sum\sum A_j\,B_{mx}\,\phi^{j+m-3}$$

$$-\overline{K}_3\sum\sum (m-1)A_j\,B_m\,\phi^{j+m-4}\phi_x+iK_4\sum\sum\sum A_j\,A_k\,B_m\,\phi^{j+K+m-5}$$

$$-2iS\sum\sum\sum B_j\,B_kC_m\,\phi^{j+K+m-3}, \quad (19)$$

and a similar equation for C. Assuming $\phi(x,t)=f(x)+t$, we obtain from the leading terms $\alpha=-2$, $\beta=\gamma=-1$, $A_0=4f_x^3$, and $C_0=2\phi_x^2/SB_0$; B_0 is arbitrary. Then coefficient ϕ^{-2} yields

$$A_{0t}-A_1\,\phi_t=2SB_0C_{0x}+2SB_{0x}C_0-2SB_0C_1\,\phi_x-2SB_1C_0\,\phi_x,$$

together with

$$\begin{aligned}&-B_0\,\phi_t+4iB_{0x}\,\phi_x+2iB_0\,\phi_{xx}=-4iSB_0C_0B_1-2iSB_0^2C_1,\\&-C_0\,\phi_t-4iC_{0x}\,\phi_x-2iC_0\,\phi_{xx}=-4iSB_0C_0C_1-2iSB_1C_0^2.\end{aligned}\quad (20)$$

So the first Eq. of (20) yields that A_1 cannot be determined, that is, it is arbitrary. But if we assume that $A_1=A_2=\cdots=0$ for $i>0$, then it reduces to

$$A_{0t}=2S(B_0C_0)_x,$$

the original nonlinear equation if we also have $B_0C_1=-B_1C_0$.

Equating now coefficients of ϕ^{-1} we get

$$\begin{aligned}A_{1t}&=2SB_1C_{0x}+2SB_0C_{1x}+2SB_{0x}C_1+2SB_{1x}C_0,\\B_{0t}&=2iB_{0xx}-2iSB_i^2C_0-4iSB_0B_1C_1-2iSB_0^2C_2-4iSB_2B_0C_0,\\C_{0t}&=-2iC_{0xx}+2iSC_1^2B_0+4iSC_0C_1B_1+2iSC_0^2B_2,\end{aligned}\quad (21)$$

which clearly indicate the indeterminacy of the coefficients B_2, C_2, etc. So if together with the assumption that $A_i=0$ for $i>1$ we also consider $B_j=0$, $C_j=0$ for $j\gg 2$ then from the second and third equations of (20) we get

$$-B_0\,\phi_t+4iB_{0x}\,\phi_x+2iB_0\,\phi_{xx}=-2iSB_0C_0B_1, \quad (22a)$$

$$-C_0\,\phi_t-4iC_{0x}\,\phi_x-2iC_0\,\phi_{xx}=2iSB_0C_0C_1. \quad (22b)$$

From (22a) and (22b) we deduce that ϕ satisfies

$$-\phi_t+2i\phi_{xx}=0. \quad (23)$$

Up till now the form of ϕ has been quite general or arbitrary. The second and third equations of those obtained by equating ϕ^0 in Eq. (19) yields

$$\begin{aligned}B_{1t}&=2iB_{1xx}=-2iSB_1^2C_1,\\C_{1t}&=2iC_{1xx}=2iSC_1^2B_1,\end{aligned}$$

the original nonlinear equations. So we have actually demonstrated that it is possible to truncate the series (2a)–(2c) only with a few number of terms, and can really have[7]

$$\begin{aligned}A&=\phi^{-2}A_0,\\B&=\phi^{-1}[B_0+B_1\,\phi],\\C&=\phi^{-1}[C_0+C_1\,\phi],\end{aligned}$$

which is nothing but equations connecting two sets of solutions (A,B,C) and (A_0,B_1,C_1) of the same equation and so can be thought of as a Backlünd transformation.

ACKNOWLEDGMENTS

One of the authors (A.R.C.) would like to thank Professor Abdus Salam, the International Atomic Energy Agency and UNESCO for hospitality at the International Center for Theoretical Physics, Trieste.

[1]M. J. Ablowitz, A. Ramani, and A. Segur, Lett. Nuovo Cimento **23**, 333 (1980).
[2]M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715, 1006 (1980).
[3]J. Weiss, M. Tabor, and G. Carnevale, J. Math. Phys. **24**, 522 (1983).
[4]A. C. Newell, SIAM J. Appl. Math. **35**, 650 (1978).
[5]A. C. Newell, Proc. R. Soc. London Ser. A **365**, 283 (1979).
[6]M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981).
[7]J. Weiss, J. Math. Phys. **24**, 1405 (1983).

Realizing the Berezin integral as a superspace contour integral

Alice Rogers
Department of Mathematics, King's College, Strand, London WC2R 2LS, England

(Received 8 April 1985; accepted for publication 31 October 1985)

Integration on supermanifolds, using contours and the covariant differential forms of Kostant [B. Kostant, "Graded manifolds, graded Lie theory and prequantisation," in *Lecture Notes in Mathematics*, Vol. 570 (Springer, Berlin, 1977)], is described; the good properties that these integrals naturally acquire are considered. It is then shown that the formal process of integration over even and odd variables introduced by Berezin [F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966)] using partly covariant and partly contravariant forms can be regarded as a special case of these contour integrals over covariant forms.

I. INTRODUCTION

The standard rule for integration over anticommuting variables

$$\int \theta \, \bar{d}\theta = 1, \quad \int \bar{d}\theta = 0, \tag{1.1}$$

was introduced by Berezin[1] as a formal process in his work on the quantization of Fermi fields, and is widely and successfully used in the path integral approach to field quantization. With the advent of supersymmetry, much greater demands have been put on this form of integration than originally envisaged, and the formalism stretched beyond its limits. The aim of this paper is to show that, by realizing the abstract Berezin integral as a contour integral in superspace, one extends the possibilities of "odd" integration because the good properties of conventional integration can be used. Some steps in this direction have been taken by Rabin[2]; these are described in detail in Sec. II. A very early work by Fairlie and Martin[3] describes the Berezin integral as an integral around a closed loop with "measure" $(1/\theta^2)\, d\theta$. Although $1/\theta^2$ is undefined, the approach taken in this paper owes much to the spirit of Ref. 3.

The symbol "$\bar{d}\theta$" in (1.1) does not represent a coordinate differential; indeed in order to obtain the rule (1.1) in all coordinate systems, $\bar{d}\theta$ must transform contravariantly (like a vector). The formal integration process can be extended to the case of several anticommuting variables, and even combined with conventional integration over real variables in cases where both odd and even variables are present. The key definition and theorem, due to Berezin and his coworkers, and clearly presented by Leĩtes,[4] is the following. Suppose U is an open set in $\mathbf{R}^m$ and $\theta^1,\ldots,\theta^n$ are n anticommuting "variables." Also, given $x \in U$,

$$f(x,\theta) := f_{(0)}(x) + \sum f_{(j)}(x)\theta^{(j)} + \cdots + f_{(1\cdots n)}(x)\theta^1\cdots\theta^n . \tag{1.2}$$

Here a "function" f of the real x's and anticommuting θ's is a formal Taylor series in the θ's with coefficients that are functions of the x's. Then

$$\int_U d^m x\, \bar{d}^n\theta f(x,\theta) := \int_U f_{(1\cdots n)}(x)\,dx , \tag{1.3}$$

where the second integral is simply the usual Riemann integral. Now, if one allows formal change of variables

$$y = h(x,\theta), \quad \phi = \eta(x,\theta) ,$$

with

$$y^i = h^i_0(x) + \sum_{\substack{k,l=1\\k<l}}^{n} h^i_{k,l}(x)\theta^k\theta^l + \cdots, \quad i = 1,\ldots,m , \tag{1.4}$$

$$\phi^j = \sum_{k=1}^{n} \eta^j_k(x)\theta^k + \cdots, \quad j = 1,\ldots,n , \tag{1.5}$$

then the key result, establishing good behavior under change of variables,[3] is

$$\int_{h_0(U)} d^m y\, \bar{d}^n\phi f(y,\phi) = \int_U d^m x\, \bar{d}^n\theta f[h(x,\theta),\eta(x,\theta)]\, J(x,\theta) , \tag{1.6}$$

where $J(x,\theta)$ is the superdeterminant of the matrix $M^{ij}(x,\theta)$ with

$$\begin{aligned} M^{ik} &= \frac{\partial h^i}{\partial x^k}, \quad 1<i,k<m ,\\ M^{ij+m} &= \frac{\partial h^i}{\partial \theta^j} \quad 1<i<m, \quad 1<j<n ,\\ M^{lk} &= \frac{\partial \eta^l}{\partial x^k} \quad 1<l<n, \quad 1<k<m ,\\ M^{lj+m} &= \frac{\partial \eta^l}{\partial \theta^j}, \quad 1<l,j<m . \end{aligned} \tag{1.7}$$

The result (1.6) is only true if each of the coefficient functions in the Taylor expansion (1.2) of f is C^∞ with compact support. A valid transformation rule is of course quite essential to any coordinate-independent definition of superintegration.

In the application of superspace integration to supersymmetric quantum field theories, one rapidly finds the formalism described above too limited. In the first place supersymmetry involves changes of variable of the form

$$\theta^{j\prime} = \theta^j + \epsilon^j , \tag{1.8}$$

where ϵ^j is an odd Grassmann element independent of the θ's. Thus a more general form of transformation than (1.4)

 0022-2488/86/030710-08$02.50

and (1.5) is required, and instead of the abstract θ variables a specific Grassmann algebra is needed. Thus one uses the type of superspace proposed by Volkov and Akulov[5] and Salam and Strathdee,[6] where the x's are even Grassmann variables, the θ's are odd Grassmann variables, and $f(x, \theta)$ denotes a function mapping superspace into The Grassmann algebra. Such a superspace has been extensively discussed in the literature,[7–9] together with supermanifolds made by patching bits of superspace together. The necessary adjustment to the proof of the transformation result (1.6) has been worked out by Fung. His proof is quoted in Ref. 10. But other difficulties remain, in particular the following.

(a) A much wider class of function is considered in quantum physics than the C^∞ functions with compact support that the transformation rule (1.6) requires for its validity.

(b) Although the formalism allows one to integrate over superspace, there is no consistent rule for integration over subspaces when the odd dimension of the subspace is lower than that of the full superspace. By analogy with the method for integrating a p-form over a p-dimensional submanifold of a conventional real manifold, one might expect to be able to integrate a "(p, q)-form" over a (p, q)-dimensional subspace of superspace. However, this will not work for the $d^m x\, \bar{d}^n\theta$ type of form because the contravariantly transforming $d\theta$ part will not "pull back." Now, being forced to integrate over the full superspace of the theory puts severe restrictions on the dimensions of invariants one can obtain by superspace integration; particularly in N-extended supersymmetry this has prevented the construction of actions as superspace integrals because the high dimension of the superspace volume element $d^4x\, \bar{d}^{4N}\theta$ means that very negative-dimensional pre-pre-pre ··· potentials are required. A coherent theory of subspace integration should open up an enormous number of new possibilities.

(c) The mathematical theory of integration on conventional manifolds is very elegant; from the fundamental definition

$$\int_\sigma \omega = \int_{I^m} \sigma^*(\omega) \tag{1.9}$$

for integrating an m-form ω over an m-simplex on some manifold, nice functorial behavior (that is, good behavior under change of variable) and Stokes' theorem follow almost automatically.[11] Also integration provides a link between the topological and differentiable structure of a manifold, and is the key to the "topological" aspects of quantum field theory. It is thus clearly highly desirable to extend the Berezin formalism to a full theory or integration of "superforms" over "superchains" with similarly nice and useful properties.

Many of the difficulties with Berezin integration stem from the fact that a "funny" differential from $\bar{d}\theta$ is used. Now there do exist ordinary differential forms on superspace (and more general supermanifolds), that is to say covariant graded-antisymmetric forms. They are described very fully by Kostant[12] and have the useful pullback properties; further details are given in Sec. II. (Other authors have considered these objects, but Kostant's treatment is the most detailed, and the most far reaching.)

The aim of this paper is to express the Berezin integral as a contour integral over this kind of form, because such integrals automatically have nice transformation properties; thus difficulty (a) can be tackled. A further paper is planned where the contour approach will be extended to overcome difficulty (b) and to develop a formalism of the kind envisaged in (c).

The outline of this paper is that in Sec. II the general formalism for contour integrals in superspace is introduced; here the point of view that superspace is a Banach space (with additional structure)[9] is essential. In Sec. III it is shown how the rule (1.1) for integration over a single odd variable, and its higher-dimensional analogs, can be obtained by suitable choice of contours. Section IV extends the approach to mixed odd and even superspace.

II. CONTOUR INTEGRALS IN SUPERSPACE

Contour integrals are a means of "pulling back" an integral in a space that is *algebraically* (as well as possibly geometrically) more complicated than $\mathbb{R}^m$. A familiar example, of course, is complex contour integration; if γ: $[0, 1] \to \mathbb{C}$ is piecewise C_1 and f: $\mathbb{C} \to \mathbb{C}$, one has the one-dimensional contour integral

$$\int_\gamma f(z)dz = \int_0^1 f[\gamma(t)]\gamma'(t)dt\,. \tag{2.1}$$

This involves the algebraic structure of $\mathbb{C}$ because the right-hand side of (2.1) includes multiplication of complex numbers.

The key point of this section is that a similar approach to contour integration—which leads automatically to nice properties—can be applied to superspace and supermanifolds, provided an approach is taken where these spaces are Banach spaces with additional structure. Such an approach is provided by the "G^∞ supermanifolds" of Ref. 9. (A review of various approaches to supermanifolds is in Ref. 7.) The Banach space structure is essential for the notion of a C_1 map from I^m (the unit cube in $\mathbb{R}^m$) into superspace. Additionally the Banach space and Grassmann algebra structure must interplay in such a way that the chain rules (2.7) and (2.8) are obeyed, which enables differential forms on superspace to be pulled back to $\mathbb{R}^m$. The important ideas and notation for supermanifolds needed in this paper are the following.

(i) The set[12] M_∞. This is the set of finite sequences of positive integers $\mu = (\mu_1,...,\mu_k)$ with $1 \leqslant \mu_1 < \mu_2 < \cdots < \mu_k$, $|\mu| := k$. The empty sequence Ω is included in M_∞.

(ii) The "Grassmann" algebra[9] B_∞. This is the Banach space l_1 of infinite sequences of real numbers (for convenience labeled by elements of M_∞ rather than straightforwardly with Z)

$$x = (x_\Omega, x_{(1)}, x_{(2)}, x_{(1,2)},...)\,,$$

such that

$$\|x\| := \sum_{\mu\in M_\infty} |x_\mu| < \infty\,.$$

If, for given μ in M_∞, β_μ denotes the sequence in l_1 with $x_\mu = 1$ and all other terms zero, then one can extend the multiplication rules

$$\beta_\Omega\beta_\mu = \beta_\mu\beta_\Omega = \beta_\mu, \quad \forall\ \mu\in M_\infty\ , \tag{2.2}$$

$$\beta_{(i)}\beta_{(j)} = -\beta_{(j)}\beta_{(i)}, \quad \forall \text{ integers } i,j\ , \tag{2.3}$$

$$\beta_\mu = \beta_{(\mu_1)}\beta_{(\mu_2)}\cdots\beta_{(\mu_k)}, \quad \forall\ \mu\in M_\infty\ , \tag{2.4}$$

to give l_1 the structure of a Banach algebra, denoted B_∞. (Full details are in Ref. 9.)

(iii) The (m, n)-dimensional superspace $B_\infty^{m,n}$ is the Cartesian product of m copies of the even part of B_∞ and n copies of the odd part. A typical element is denoted $(x, \theta) = (x^1,\dots, x^m; \theta^1,\dots, \theta^n)$.

(iv) ϵ: $B_\infty \to \mathbf{R}$ denotes the unique algebra homomorphism that sets the generators $\beta_1, \beta_2,\dots$ to zero; $\epsilon_{m,n}$: $B_\infty^{m,n} \to \mathbf{R}^m$ is defined by

$$\epsilon_{m,n}(x^1,\dots, x^m; \theta^1,\dots, \theta^n) = [\epsilon(x^1),\dots, \epsilon(x^m)]\ . \tag{2.5}$$

(v) A notion of differentiation of B_∞-valued functions of $B_\infty^{m,n}$ can be defined. [For instance, if $m = 0$, $n = 1$, one has $\partial f/\partial\theta$ defined by

$$f(\theta + \epsilon) = f(\theta) + \epsilon\frac{\partial f}{\partial\theta}(\theta) + O(\|\epsilon\|^2)\ . \tag{2.6}$$

Again, full details are in Ref. 9.] Infinitely differentiable functions are called G^∞ and have terminating Taylor series in the odd variables. Two important chain rules are (a) given U open in $\mathbf{R}^m$, V open in $B_\infty^{m,n}$, and γ: $U \to B_\infty^{m,n}$ with $\gamma(U) \subset V$, $f\in G^\infty(V)$, then if γ is piecewise C^1, $f\circ\gamma$ is piecewise C^1 and

$$\partial_i(f\circ\gamma) = \sum_{k=1}^{m+n}\partial_i(\gamma^k)G_k f\circ\gamma, \quad i = 1,\dots, p \tag{2.7}$$

(here G_k denotes differentiation with respect to the kth Grassmann variable); and (b) given $V\subset B_\infty^{p,q}$, h: $V\to B_\infty^{m,n}$, W open in $B_\infty^{m,n}$ with $h(V) \subset W$, and $f\in G^\infty(W)$, then

$$G_l(f\circ h) = \sum_{k=1}^{m+n} G_l(h)\circ G_k f\circ h, \quad l = 1,\dots, p+q\ . \tag{2.8}$$

[One can define a topology on $B_\infty^{m,n}$ and then define supermanifolds by a suitable patching of open sets in $B_\infty^{m,n}$ (see Refs. 7–9).]

(vi) Following Kostant,[12] if $G^\infty(V)$ denotes the B_∞ module of G^∞ functions on an open set V in superspace, a vector field is an endomorphism X of $G^\infty(V)$ such that

$$Xfg = (Xf)g + (-1)^{|X||f|}fXg\ , \tag{2.9}$$

for all f, g in $G^\infty(V)$ and

$$X af = (-1)^{|X||a|}a\,Xf, \tag{2.10}$$

for all a in B_∞, f in $G^\infty(V)$. (Here $|f|$ denotes the Grassmann degree of f, and so on.) In the following der (V) denotes the set of vector fields on V.

(vii) Of great importance to the theory of integration are the differential forms on V. Again following Kostant,[12] a p-form on V is a p-linear mapping ω on der (V) such that

$$\langle\xi_1,\dots,f\xi_i,\dots,\xi_p|\omega\rangle = (-1)^{|f|\sum_{l=1}^{i-1}|\xi_l|}f\langle\xi_1,\dots,\xi_p|\omega\rangle\ , \tag{2.11}$$

and

$$\langle\xi_1,\dots,\xi_i,\xi_{i+1},\dots,\xi_p|\omega\rangle = (-1)^{1+|\xi_i||\xi_{i+1}|}\langle\xi_1,\dots,\xi_{i+1},\xi_i,\dots,\xi_p|\omega\rangle\ . \tag{2.12}$$

In the following, $\Omega^p(V)$ denotes the set of p-forms on V,

$$\Omega^0(V) := G^\infty(V), \quad\text{and}\quad \Omega(V) := \bigoplus_{p=0}^{\infty}\Omega^p(V)\ . \tag{2.13}$$

An exterior product $\wedge$ and exterior derivative d can be defined on $\Omega(V)$ (see Ref. 12). The exterior derivative satisfies $dd = 0$.

A p-form β can be expanded in local coordinates $z^k = (x^i, \theta^j)$ as

$$\beta = \sum_{1<i_1<\cdots<i_p<m+n} dz^{i_1}\wedge\cdots\wedge dz^{i_p}f_{i_1\cdots i_p} \tag{2.14}$$

The exterior product and derivative take the expected form in local coordinates. The forms on superspace (and more generally on supermanifolds) form a bigraded algebra (there is the "super" Z_2 grading, as well as the usual Z grading of differential forms), which is bigraded commutative. In particular

$$\begin{aligned} dx^i\wedge dx^j &= -dx^j\wedge dx^i,\\ d\theta^k\wedge dx^j &= -dx^j\wedge d\theta^k,\\ d\theta^k\wedge d\theta^l &= +d\theta^l\wedge d\theta^k. \end{aligned} \tag{2.15}$$

[Among other things, this means that $(d\theta^i)^2$ may be nonzero, and there is no upper limit on the degree of a differential form on a supermanifold.] Under G^∞ maps between supermanifolds these forms have a well behaved pullback. Also, if one has a C^∞ map from a real manifold into a G^∞ supermanifold (a possibility, since a G^∞ supermanifold is *a fortiori* a C^∞ manifold), then a p-form on the G^∞ manifold will pull back to a p-form on the C^∞ manifold. Thus one may define contour integration of p-forms on an (m, n)-dimensional supermanifold in the following manner: Let γ: $I^p \to Y$ be piecewise C^1, and ω be a p-form on Y. Then

$$\int_\gamma \omega := \int_{I^p}\gamma^*(\omega)\ . \tag{2.16}$$

In particular, for a one-dimensional contour γ: $I\to B_\infty^{0,1}$ and f: $B_\infty^{0,1}\to B_\infty$,

$$\int_\gamma d\theta f(\theta) = \int_0^1 \gamma'(t)dt f[\gamma(t)]\ . \tag{2.17}$$

The Banach space property of B_∞ is crucial in giving (2.16) and (2.17) well-defined meaning. As well as reparametrization invariance of such integrals, one has the transformation rule

$$\int_{h\circ\gamma} = \int_\gamma h^*\omega\ , \tag{2.18}$$

if h: $Y\to Z$ is a G^∞ map of supermanifolds, and ω is a p-form on Z. Also one has Stokes' theorem

$$\int_{\partial\gamma}\omega = \int_\gamma d\beta\ , \tag{2.19}$$

if ω is a $(p-1)$-form on Y. These are proved exactly as the analogous classical theorems are proved, using the fact that

$$(h\circ\gamma)^*\omega = \gamma^*h^*\omega\ , \tag{2.20}$$

and

$$d(\gamma^*\beta) = \gamma^* d\beta. \quad (2.21)$$

In a previous paper,[13] contour integrals in even superspace (first considered by De Witt[8]) have been described, and a Cauchy theorem proved showing that the integral of the function round a closed curve (bounding a region where the function is G^∞) is zero. The following example of a contour integral in odd superspace shows that the Cauchy theorem is not true for odd superspace. (In fact it is easy to see where the proof in Ref. 13 breaks down for anticommuting variables.)

Example 2.1: Let $\gamma: I \to B_\infty^{0,1}$ be the curve

$$\gamma(t) = (\sin 2\pi t)\beta_1 + (\cos 2\pi t)\beta_2. \quad (2.22)$$

(Recall that β_1 and β_2 are anticommuting generators in the algebra B_∞.) Then

$$\int_\gamma d\theta = 0, \quad (2.23)$$

but

$$\int_\gamma \theta\, d\theta = 2\pi \int_0^1 -\beta_1\beta_2\, dt = -2\pi\beta_1\beta_2. \quad (2.24)$$

This breakdown of the Cauchy theorem means that one cannot express the integral of an arbitrary function along a curve simply in terms of functions evaluated at its end points; this of course is related to the fact that one cannot antidifferentiate the function f with $f(\theta) = \theta$. As already mentioned, Rabin[2] made the very appealing observation that the results

$$\int_\gamma d\theta = 0 \quad \text{(by Stokes' theorem)}, \quad (2.25)$$

$$\int_\gamma d\theta\, \theta \neq 0, \quad \text{in general}, \quad (2.26)$$

for any closed curve γ is very suggestive of the Berezin rule (1.11). There are three apparent difficulties with this approach.

(i) Although $\int_\gamma d\theta\, \theta$ is not in general zero, one is bound to have $\int_\gamma d\theta\, \theta = c$, where c is a noninvertible element of B_∞ $[\epsilon(c) = 0]$. Thus one cannot simply normalize by dividing by c. Rabin proposed to avoid this difficulty by defining, for a given closed contour γ,

$$A(\gamma) = \int_\gamma d\theta\, d\theta, \quad (2.27)$$

and then defining $\int_{\text{Berezin}} f(\theta) d\theta$ to be the quantity uniquely defined by

$$\int_\gamma d\theta f(\theta) = A(\gamma) \int_{\text{Berezin}} f(\theta) d\theta, \quad (2.28)$$

for all γ.

(ii) A second difficulty with Rabin's approach is that it does not immediately extend to cases where there is more than one odd variable. For instance, $\int_\gamma \theta^1 d\theta^1 d\theta^2$ is not zero for an arbitrary closed contour γ, since $\theta^1 d\theta^1 d\theta^2$ is not an exact form.

(iii) Also the one-form $d\theta$ transforms covariantly, and thus does not correspond to the $\bar{d}\theta$ of the Berezin integral.

In Sec. III it is shown how Rabin's approach can be adapted and applied to integration of several odd variables. Difficulty (i) is overcome by using the existence of elements c of B_∞, which, although having $\epsilon(c) = 0$, do not annihilate any nonzero elements of B_∞, so that division by c, where possible, is unique. Difficulty (ii) is overcome by a careful choice of contour and (iii) is overcome by brute force.

III. A CONTOUR REPRESENTATION OF THE BEREZIN INTEGRAL

The work in Sec. II describes a natural way to integrate on supermanifolds in a consistent manner; the good properties of the pullback map allow one to harness the good properties of standard integration on $\mathbb{R}^m$. It is highly desirable to include the Berezin integral in this general formalism, so that results can be proved by appeal to general properties of integrals, and the difficulties (described in the Introduction), which are encountered when treating the Berezin integral as a purely formal process, can be circumvented. A preliminary lemma is required.

Lemma 3.1: For each integer $n = 1,2,\ldots$ let the element c_n of the infinite-dimensional algebra B^∞ be defined by

$$c_n = \sum_{s=1}^{\infty} \alpha_{s,n},$$

where

$$\alpha_{s,n} = (1/2^{2s})\beta_{(2s-1)n+1}\beta_{(2s-1)n+2}\cdots\beta_{2sn} \quad (3.1)$$

(recall that β_i, $i = 1, 2,\ldots$ denote generators of B_∞). Then, for each $n = 1, 2,\ldots$ and for all a in B_∞,

$$ac_n = 0 \quad \text{if and only if} \quad a = 0. \quad (3.2)$$

Outline of proof: The full proof is too long to include here. The first step is to suppose that $a = \Sigma_{\mu\in M_\infty} a^\mu \beta_\mu$ satisfies

$$ac_n = 0. \quad (3.3)$$

It is easily shown that, if the sequence μ does not contain any subsequence of the form $[(2s-1)n+1, (2s-1)n+2,\ldots,2sn]$ (where s is some integer) then $a^\mu = 0$. Also, by induction over the number of subsequences in μ, one can prove that all a^μ must be zero.

Corollary 3.2: If $x = ac_n$, then one can set

$$a := (1/c_n)x \quad (3.4)$$

unambiguously, and thus give a partial meaning to division by c_n.

Now the first example of a contour representation of a Berezin integral, that of Berezin integration on $B_\infty^{0,1}$, will be described. To do this one first specifies the contour (or more accurately, a sum of contours, that is, a chain) γ, and then, for an arbitrary coordinate system θ defines the one-form $\bar{d}\theta$, (Definition 3.3). Theorem 3.4 then establishes that the integral

$$\frac{1}{c_1}\int_\gamma \bar{d}\theta(\gamma) f(\theta)$$

has all the properties of the Berezin integral.

Definition 3.3: Let $\rho : B_\infty^{0,1} \to B_\infty^{0,1}$ be a coordinate system on $B_\infty^{0,1}$. For each $s = 1,2,\ldots$, let $\gamma_s : I \to B_\infty^{0,1}$ be the contour such that

$$\rho \circ \gamma_s(t) = (1/2^s)(\beta_{2s-1} \sin t + \beta_{2s} \cos t) . \qquad (3.5)$$

Also, let

$$\gamma = \sum_{s=1}^{\infty} \gamma_s . \qquad (3.6)$$

Now suppose θ is some other, quite arbitrary, coordinate system on $B_\infty^{0,1}$. That is,

$$\theta = a\rho + b ,$$

with a an invertible even element of B_∞ and b an odd element. (a) Set $\bar{d}\theta(\gamma) := (1/a) d\rho$. (b) Let

$$\int_{\text{Berezin}} \bar{d}\theta f(\theta) = \frac{1}{c_1} \bar{d}\theta(\gamma) f(\theta) . \qquad (3.7)$$

The following theorem establishes that the differential $\bar{d}\theta$ transforms in the required manner, that is, evaluation of the Berezin integral gives the required result (and thus is independent of the choice of coordinate ρ, relative to which γ is defined). And also that the integral (3.7) is invariant under change of the coordinate θ.

Theorem 3.4:

(a) $$\frac{1}{c_1} \int_\gamma \bar{d}\theta(\gamma)(\theta p + q) = p \qquad (3.8)$$

(if p, q are arbitrary elements of B_∞).

(b) Suppose $\phi: B_\infty^{0,1} \to B_\infty^{0,1}$ is a further coordinate on $B_\infty^{0,1}$. Then

$$\bar{d}\phi = \frac{\partial\theta}{\partial\phi} \bar{d}\theta . \qquad (3.9)$$

(c) If $\theta = h \circ \phi$,

$$\int_\gamma \bar{d}\theta(\gamma) f(\theta) = \int_\gamma \bar{d}\phi(\gamma) \frac{1}{\partial\theta/\partial\phi} f \circ h(\phi) . \qquad (3.10)$$

Proof: (a) One has

$$\frac{1}{c_1} \int \bar{d}\theta(\gamma)(\theta p + q)$$
$$= \frac{1}{c_1} \sum_{s=1}^{\infty} \int_{\gamma_s} \bar{d}\theta(\gamma)(\theta p + q)$$
$$= \frac{1}{c_1} \sum_{s=1}^{\infty} \int_0^1 dt \Big\{ (\beta_{2s-1} \cos t - \beta_{2s} \sin t) \times \frac{1}{a}$$
$$\times [a(\beta_{2s-1} \sin t + \beta_{2s} \cos t) p + q] \Big\}$$
$$= p . \qquad (3.11)$$

(b) There exist c, d in B_∞ such that $\phi = c\theta + d$. Thus

$$\bar{d}\phi = \frac{1}{ca} d\rho = \frac{\partial\theta}{\partial\phi} \bar{d}\theta . \qquad (3.12)$$

(c) This follows immediately from (b).

This completes the description of the contour approach to Berezin integration over one variable. For several variables the approach is similar but inevitably more complicated. Again the first step is to specify a chain $\gamma = \Sigma_{s=1}^{\infty} \gamma_s$ on $B_\infty^{0,n}$, and then to define $\bar{d}^n\theta$ with reference to the coordinate system in which γ has a simple form.

Definition 3.5: Let $\rho = (\rho^1, \ldots, \rho^n)$, with $\rho: B_\infty^{0,n} \to B_\infty^{0,n}$, be a coordinate system on $B_\infty^{0,n}$. For each $s = 1, 2, \ldots$, let $\gamma_s: I^n \to B_\infty^{0,1}$ be the map such that, for $k = 1,\ldots,n$,

$$\rho^k \circ \gamma_s(t^1, \ldots, t^n) = \beta_{(2s-1)n+2k-1} \sin t^k + \beta_{(2s-1)n+2k} \cos t^k . \qquad (3.13)$$

Then, if $\theta = (\theta^1, \ldots, \theta^n)$ is another (arbitrary) coordinate system on $B_\infty^{0,n}$

(a) $d^n\theta(\gamma) = 1/J d\rho^n \cdots d\rho^1$,

where

$$J = \det\left(\frac{\partial\theta^i}{\partial\rho^j}\right) ; \qquad (3.14)$$

(b) $$\int_{\text{Berezin}} \bar{d}^n\theta f(\theta^1, \ldots, \theta^n)$$
$$:= \frac{1}{c_n} \int_\gamma \bar{d}\theta^n(\gamma) f(\theta^1, \ldots, \theta^n) . \qquad (3.15)$$

Again, a theorem is proved to establish that evaluation of the integral defined in (3.15) gives the required result (independently of the choice of coordinate system ρ in which γ takes a simple form), and also that the "volume form" $d^n\theta \cdots d\theta^1$ transforms correctly and that the integral (3.15) is invariant under a change of coordinate.

Theorem 3.6: (a) If $f: B_\infty^{0,n} \to B_\infty$ with

$$f(\theta^1, \ldots, \theta^n) = \theta^1\theta^2 \cdots \theta^n p + \text{lower-order terms}, \qquad (3.16)$$

then

$$\frac{1}{c_n} \int_\gamma \bar{d}^n\theta(\gamma) f(\theta^1, \ldots, \theta^n) = p . \qquad (3.17)$$

(b) If $\phi = (\phi^1, \ldots, \phi^n)$ is a further coordinate system on $B_\infty^{0,n}$ then

$$\bar{d}^n\phi(\gamma) = \bar{d}^n\theta(\gamma) \times \det\left(\frac{\partial\theta^i}{\partial\phi^j}\right) . \qquad (3.18)$$

(c) If $\theta = h \circ \phi$, and $g: B_\infty^{0,n} \to B_\infty$,

$$\int_{\text{Berezin}} \bar{d}^n\theta(\gamma) g(\theta)$$
$$= \int_{\text{Berezin}} \bar{d}^n\phi(\gamma) \frac{1}{\det(\partial\theta^i/\partial\phi^j)} [g \circ h(\phi)] . \qquad (3.19)$$

Proof: (a) In the particular coordinate system ρ

$$\frac{1}{c_n} \int_\gamma \bar{d}^n \rho(\gamma) f(\rho^1, \ldots, \rho^n) = p , \qquad (3.20)$$

by direct calculation. Thus to establish (3.17) in an arbitrary coordinate system, it is sufficient to prove that

$$\theta^1 \cdots \theta^n \times \frac{1}{\det(\partial\theta^i/\partial\rho^j)} = \rho^1 \cdots \rho^n + \text{lower-order terms} , \qquad (3.21)$$

and that when

$$\theta^{i_1}\theta^{i_2} \cdots \theta^{i_k} \times \frac{1}{\det(\partial\theta^i/\partial\rho^j)} \quad (\text{with } k < n)$$

is expanded in terms of the ρ^i, there is no term containing $\rho^1 \cdots \rho^n$. The facts are essentially proved most elegantly by Fung (quoted by Van Niewenhuizen[10]) in the course of his proof of the transformation rule of the Berezin integral.

(Other proofs in the literature do not include all necessary cases, because they omit the possibility of translations of the ρ^i.)

(b) $\bar{d}^n\phi(\gamma)$

$$= d^n\rho \cdots d\rho^1 \times \frac{1}{\det(\partial\phi^i/\partial\rho^j)}$$
$$= d\rho^n \cdots d\rho^1 \times \frac{1}{\det(\partial\phi^i/\partial\theta^k)} \times \frac{1}{\det(\partial\theta^k/\partial\rho^j)}$$
$$= \bar{d}^n\theta(\gamma) \times \det\left(\frac{\partial\theta^k}{\partial\phi^i}\right). \quad (3.22)$$

(c) The result follows immediately from (b).

IV. COMBINING THE BEREZIN INTEGRAL WITH INTEGRATION OVER EVEN VARIABLES

As mentioned in the Introduction, the Berezin transformation rule (1.6) is not always valid if the coefficient functions in the θ-expansion of the integrand do not have compact support. In a previous paper,[13] the author has shown how some of the difficulty can be alleviated by treating the even part of the integral as a contour integral in even superspace (as suggested by De Witt[8]) rather than simply ignoring transformations in nilpotent even directions. The theory presented in Ref. 13 is an uncomfortable hybrid of contour integration over even variables and formal integration over odd variables, and the hope was expressed that the odd integration could be incorporated into the contour approach and thus odd and even integration more happily married. Clearly with the approach outlined in Sec. III this should now be possible. Vladimirov and Volovich[14] have also considered integrals on even superspace; they express the Berezin integral in odd superspace as an integral on $\mathbb{R}^n$, but not explicitly as a contour integral.

First, two examples are presented showing how the Berezin transformation rule breaks down when the integrand does not have compact support.

Example 4.1. Integration on $B_L^{1,1}$:

$$\text{(a)} \quad \int_{0\ \text{Berezin}}^{1} dy\, \bar{d}\phi\, y = 0. \quad (4.1)$$

Set

$$y = x + \theta\alpha \quad (\alpha \neq 0), \quad \theta = \phi. \quad (4.2)$$

Then, [according to the rule (1.6)], the integral (4.1) becomes

$$\int_{0\ \text{Berezin}}^{1} dx\, \bar{d}\theta (x + \theta\alpha) = \alpha \neq 0. \quad (4.3)$$

$$\text{(b)} \quad \int_{0\ \text{Berezin}}^{1} dy\, \bar{d}\phi\, y^2\phi = \frac{1}{3}. \quad (4.4)$$

Set

$$y = x + \theta\alpha \quad (\alpha \neq 0), \quad \phi = \beta x + \theta (\beta \neq 0). \quad (4.5)$$

Then, [according to the rule (1.7)], the integral (4.4) becomes

$$\int_{0\ \text{Berezin}}^{1} dx\, \bar{d}\theta (1 + \alpha\beta)(x + \theta\alpha)^2(\beta x + \theta) = \frac{1}{3} + \frac{4}{3}\alpha\beta. \quad (4.6)$$

These examples show that the contour approach (which certainly will provide a good transformation rule) must alter either the rule for calculating $dx\, d\theta$ in terms of $dy\, d\phi$ or the rule (1.3) for evaluating the integral. Really the difficulty stems from the fact that there are already ambiguities present in purely even superspace integration[13]; this is because different contours in $B_\infty^{m,0}$ may project down onto the same region in $\mathbb{R}^m$. This ambiguity also shows up in the following definition of integration over a "region" in superspace.

Definition 4.2: (a) A mapping $\sigma_s: I^{m+n} \to B_\infty^{m,n}$ is called a Berezin s-contour with fiducial coordinate system $\chi = (r, \rho)$ if

$$\text{(i)} \quad \chi \circ \sigma_s = \tau \times \gamma_s, \quad (4.7)$$

where $\tau: I^m \to B_\infty^{m,0}$ with

$$\epsilon \circ \tau(t^1,\ldots,t^m) = \epsilon \circ \tau(t'^1,\ldots,t'^m), \quad (4.8)$$

only if

$$(t^1,\ldots,t^m) = (t'^1,\ldots,t'^m) \quad (4.9)$$

[a typical element of I^{m+n} is denoted $(t^1,\ldots,t^m;\ u^1,\ldots,u^n)$]; and

$$\text{(ii)} \quad \gamma_s: I^n \to B_\infty^{0,n}$$

is as defined in Definition 3.5.

(b) The formal sum

$$\sigma = \sum_s \sigma_s \quad (4.10)$$

is called a Berezin chain with fiducial coordinate system χ.

(c) Given any other coordinate system $\psi = (x, \theta)$ on $B_\infty^{m,n}$,

$$[d^m x\, \bar{d}^n\theta](\sigma) : d^m r\, d^n\rho \times J, \quad (4.11)$$

where $J =$ superdeterminant (M^{ij}) with

$$M^{ij} = \frac{\partial\psi^j}{\partial\chi^i}. \quad (4.12)$$

[In the future J will be denoted superdet $(\partial\psi^j/\partial\chi^i)$. Equation (4.12) makes it clear which is the row index and which is the column index.]

(d) Given $U \in \mathbb{R}^m$ and $f: \epsilon_{m,n}^{-1}(U) \to B_\infty$,

$$\int_{U\ \text{Berezin}} d^m x\, d^n\theta f(x, \theta)$$
$$:= \int_\sigma [d^m x\, \bar{d}^n\theta](\sigma) f(x, \theta), \quad (4.13)$$

where σ is as above, and additionally

$$\epsilon_{m,n} \circ \psi \circ \sigma_s(I^{m+n}) = U, \quad (4.14)$$

for each $s = 1, 2,\ldots$. [The notation $f(x, \theta)$ really means $f \circ \psi$; it seems preferable to use the accepted notation.] Note that by setting $n = 0$, one has a definition of integration in the purely even superspace $B_\infty^{m,0}$, coinciding with that of Ref. 13.

The following theorem proves the transformation rule for the Berezin integral. (Once again, in this formalism, this is almost trivial.) It also shows how near the value of an integral defined by (4.13) comes to the Berezin value (1.3);

and to what extent the integral depends on the choice of coordinates systems $\chi = (r, \rho)$, in which the contours σ_s take the simple form (4.7). (The lack of complete coordinate independence might seem fatal, making it impossible to define an integral on a general supermanifold. In fact an equivalent difficulty applies to the standard Berezin integral[15,16]; it is not such a difficulty as it might appear, for reasons that are explained below.)

Theorem 4.3: The notation of Definition 4.2 is used.

(a) Suppose that $\psi' = (y, \phi)$ are a further set of coordinates on $B_\infty^{m,n}$. Then

$$[d^m y\, \bar{d}^n \phi](\sigma) = [d^m x\, \bar{d}^n \theta](\sigma) \times \text{superdet}\, \frac{\partial \psi'^k}{\partial \psi^l}. \tag{4.15}$$

(b) If h: $B_\infty^{m,n} \to B_\infty^{m,n}$,

$$h = \psi \circ \psi'^{-1}, \tag{4.16}$$

$$\int_{\epsilon_{m,n} h[\epsilon_{m,n}^{-1}(U)]\ \text{Berezin}} d^m x\, \bar{d}^n \theta f(x, \theta) = \int_{U\ \text{Berezin}} d^m y\, \bar{d}^n \phi \ \text{superdet}\left(\frac{\partial \psi'^k}{\partial \psi^l}\right) f \circ h(y, \phi). \tag{4.17}$$

(c) If the coordinates $\psi = (x, \theta)$ satisfy

$$\frac{\partial x^i}{\partial \rho^j} = 0, \quad i = 1,...,m; \quad j = 1,...,n, \tag{4.18}$$

and

$$f(x, \theta) = f^{(n)}(x)\theta^1 \cdots \theta^n + \text{lower-order terms}, \tag{4.19}$$

then

$$\int_{U\ \text{Berezin}} d^m x\, \bar{d}^n \theta f(x, \theta) = \int_U d^m x f^{(n)}(x), \tag{4.20}$$

where the second integral is to be regarded as an integral on $B_\infty^{m,0}$.

(d) If f has a θ expansion (with each coefficient function of compact support contained in U),

$$\int_{U\ \text{Berezin}} d^m x\, \bar{d}^n \theta f(x, \theta) = \int_U d^m x f^{(n)}(x). \tag{4.21}$$

Proof: (a) and (b) are immediate.

(c) Choosing σ as in Definition 4.2,

$$\int_{U\ \text{Berezin}} d^m \bar{d}^n \theta f(x, \theta) = \int_\sigma d^m r\, d^n \rho \ \text{superdet}\left(\frac{\partial \psi^i}{\partial \chi^j}\right) f(x, \theta) \tag{4.22}$$

$$= \int_{I^{m+n}} \sigma^*(d^m r)\sigma^*(d^n \rho)\sigma^* \times \left(\frac{\det(\partial x^i/\partial r^j)}{\det(\partial \theta^k/\partial \rho^l)} f(x, \theta)\right) \tag{4.23}$$

$$= \int_{I^m} dt^1 \cdots dt^m \det\left(\frac{\partial r^i \circ \sigma}{\partial t^j}\right)(t^1,...,t^m) \times \det\left(\frac{\partial x^i}{\partial r^j}\right)(r \circ \sigma(t^1,...,t^m)) f^{(n)}[x(t^1,...,t^m)] \tag{4.24}$$

$$= \int_{\sigma'} d^m x f^n(x), \tag{4.25}$$

where this last integral is an integral on $B_\infty^{m,0}$ with $x \circ \sigma'(t^1,...,t^m) = k \circ \tau(t^1,...,t^m)$, where τ is a factor of $\chi \circ \sigma$ [Eq. (4.7)] and k: $B_\infty^{m,0} \to B_\infty^{m,0}$ is defined by

$$x = k \circ r. \tag{4.26}$$

The result follows.

(d) Choose intermediate coordinates $\psi' = (y, \phi)$ such that

$$x = y \quad \text{and} \quad \phi = \rho. \tag{4.27}$$

Then

$$\int_{U\ \text{Berezin}} d^m x\, \bar{d}^n \theta f(x, \theta) = \int_{U\ \text{Berezin}} d^m y\, \bar{d}^n \phi f^{(n)}(y)\phi^n \cdots \phi^1 + \text{lower-order terms}$$

$$= \int_\sigma d^m y\, d^n \phi [f^{(n)}(y)\phi^n \cdots \phi^1 + \text{lower-order terms}]$$

$$= \sum_{s=1}^{\infty} \int_{I^{m+n}} dt^1 \cdots dt^m\, du^n \cdots du^1 \times [f^{(n)}(y(\sigma_s(t, u))) \times \rho^1 \circ \sigma_s(u) \cdots \rho^n \circ \sigma_s(u) + \text{lower-order terms}] \times \det\left(\frac{\partial y^i}{\partial t^i}\right) \times \frac{\partial \rho^n \circ \sigma_s}{\partial u^n} \cdots \frac{\partial \rho^1 \circ \sigma_s}{\partial u^1}. \tag{4.28}$$

Now for each t in I^n it is possible to replace this integral by

$$\int_{I^{m+n}} dt^1 \cdots dt^m\, du^1 \cdots du^n [f^{(n)}[h(t)] \times \rho^n(u) + \text{lower-order terms}] \times \det\left(\frac{\partial h^i}{\partial t^s}(t)\right) \times \frac{\partial \rho^n \circ \sigma_s}{\partial u^n} \cdots \frac{\partial \rho^1 \circ \sigma_s}{\partial u^1}, \tag{4.29}$$

where

$$y[\sigma(t, u)] = y \circ \chi^{-1}[r \circ \sigma(t), \rho \circ \sigma(u)] = h(t) + \text{terms involving } u.$$

This is because the change from (4.28) to (4.29) involves changing the even part of the contour. Because the Cauchy theorem holds for even contour integrals,[13] this change is equivalent to integration over an interpolating contour. Since f has compact support contained in U, the value of this interpolating integral will be zero. □

V. CONCLUSIONS

What has been achieved? The main result is the expression of the Berezin integration as a particular case of a much wider type of integration, the incorporation of the formal Berezin integration process into the most natural type of integration on superspace. By putting even and odd integration on an equal footing, results such as Theorem 4.3 are much more easily proved than formerly.

But, as explained in the Introduction, the aim of this paper was to lay down foundations for a much broader theory of integration on supermanifolds. Future papers will consider applying the contour integration methods of this paper to general supermanifolds and subsupermanifolds, and investigating the relation between integration and topology. The lack of full coordinate-independence of the definition of integration (Definition 4.2) does not preclude integration on as wide a class of supermanifolds as one might have thought, because in practice many supermanifolds admit subatlases where the transition functions have the new even coordinates independent of the old odd coordinates.[16] A full investigation of this matter is underway. Obviously the contour approach here can handle integration over subsupermanifolds as easily as full supermanifolds; a p-form on a supermanifold can always be integrated over a mapping from I^p into the supermanifold. The interesting question is whether any Berezin-type theory of integration on subsupermanifolds can be developed as a special case. Finally, the question of the topological aspect is most intriguing. The lack of a Cauchy theorem for odd superspace integration certainly destroys any hope of the usual homotopy-based results, but this should not matter since many supermanifolds are homotopic to the underlying manifold, being effectively vector bundles over this manifold; it remains to be seen if integration can probe the vector bundle structure, and also the topology of supermanifolds that are not simply vector bundles over some conventional manifold, but have patching in the odd directions.[9]

ACKNOWLEDGMENT

This research was supported by the Science and Engineering Research Council of Great Britain (SERC) under advanced research fellowship No. B/AF/687.

[1]F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[2]J. M. Rabin, "The Berezin integral as a contour integral," in *Supersymmetry in Physics*, edited by V. A. Kosteleckeý and D. K. Campbell (North-Holland, Amsterdam, 1985). Also published as Physica D **15**, 65 (1985).
[3]D. B. Fairlie and D. Martin, Nuovo Cimento A **18**, 373 (1973).
[4]D. A. Leĩtes, Usp. Mat. Nauk **35**, 3 (1980); Russian Math. Surveys **35**, 1 (1980).
[5]D. V. Volkov and V. P. Akulov, Phys. Lett. B **46**, 109 (1973).
[6]Abdus Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974).
[7]M. Batchelor, "Graded manifolds and supermanifolds," in *Mathematical Aspects of Superspace, Proceedings of the Nato Workshop*, Hamburg, July 1983 (Reidel, Dordrecht, 1984).
[8]B. S. De Witt, *Supermanifolds* (Cambridge U. P., Cambridge, England, 1984).
[9]A. Rogers, J. Math. Phys. **21**, 1352 (1980).
[10]P. Van Nieuwenhuizen, Phys. Rep. **68**, 189 (1981).
[11]N. Hicks, *Notes on Differential Geometry* (Van Nostrand, New York, 1965).
[12]B. Kostant, "Graded manifolds, graded Lie theory and pre-quantisation," in *Lecture Notes in Mathematics*, Vol. 570 (Springer, Berlin, 1977).
[13]A. Rogers, J. Math. Phys. **26**, 385 (1985).
[14]V. S. Vladimirov and I. V. Volovich, Teor. Mat. Fiz. **60**, 169 (1984).
[15]F. A. Berezin, Sov. J. Nucl. Phys. **30**, 605 (1979).
[16]A. Rogers, J. Math. Phys. **26**, 2749 (1985).

Resolving the singularities in the space of Riemannian geometries

Arthur E. Fischer
Department of Mathematics, University of California, Santa Cruz, California 95064

(Received 9 July 1985; accepted for publication 28 October 1985)

A method is described for unfolding the singularities of superspace, $\mathcal{S} = \mathcal{M}/\mathcal{D}$, the space of Riemannian geometries of a manifold M. This extended, or unfolded superspace, is described by the projection $\mathcal{S}_{F(M)} = (\mathcal{M} \times F(M))/\mathcal{D} \to \mathcal{M}/\mathcal{D} = \mathcal{S}$, where $F(M)$ is the frame bundle of M. The unfolded space $\mathcal{S}_{F(M)}$ is an infinite-dimensional manifold without singularities. Moreover, as expected, the unfolding of $\mathcal{S}_{F(M)}$ at each geometry $[g_0] \in \mathcal{S}$ is parametrized by the isometry group $I_{g_0}(M)$ of g_0. The construction is completely natural, gives complete control and knowledge of the unfolding at each geometry necessary to make $\mathcal{S}_{F(M)}$ a manifold, and is generally covariant with respect to all coordinate transformations. A similar program is outlined, based on the methods of this paper, of desingularizing the moduli space of connections on a principal fiber bundle.

I. INTRODUCTION AND BACKGROUND

Let M be a C^∞ compact connected n-dimensional manifold, $\mathcal{D} = \mathrm{Diff}(M)$, the ILH (inverse limit Hilbert) Lie group of C^∞ diffeomorphisms of M, and $\mathcal{M} = \mathrm{Riem}(M)$, the ILH manifold of C^∞ Riemannian metrics on M. Then $\mathcal{D}$ acts naturally on the right on $\mathcal{M}$ by pullback

$$\Psi: \mathcal{M} \times \mathcal{D} \to \mathcal{M}, \quad (g,f) \mapsto f^*g .$$

The resulting orbit space

$$\mathcal{S} = \mathcal{M}/\mathcal{D}$$

of isometry classes of Riemannian metrics is the space of *Riemannian geometries of M*.

The space $\mathcal{S}$ has been of considerable interest both to physicists and mathematicians for some time. For $n = 3$, $\mathcal{S}$ is Wheeler's superspace,[1] the natural configuration space for a possible quantum theory of gravity. On the mathematical side, Palais[2] and Ebin[3] gave a detailed analysis of the action of Ψ, culminating in the Ebin–Palais Slice Theorem. These results were used in investigations by Fischer,[4] who showed that $\mathcal{S}$ is stratified by manifolds with the strata being labeled by the conjugacy classes in $\mathcal{D}$ of the isometry groups $I_g(M)$. Later, Bourguignon[5] gave further mathematical details of these results.

Unfortunately, the space $\mathcal{S}$ is not in general a manifold, inasmuch as Riemannian metrics may have isometry groups with different dimensions or different numbers of connected components. Since the isometry groups are the isotropy groups of the action Ψ, these differences cause the orbit space $\mathcal{S}$ to have singularities, complicating its structure.

In this paper, we resolve the singularities in $\mathcal{S}$ by constructing a *manifold* $\mathcal{S}_{FM}$ that covers $\mathcal{S}$ by a projection

$$\pi_1: \mathcal{S}_{FM} \to \mathcal{S}$$

that is continuous and open, and such that for each $[g] \in \mathcal{S}$, $\pi_1^{-1}([g])$ is a finite-dimensional closed submanifold of $\mathcal{S}_{FM}$. The pair $(\mathcal{S}_{F(M)}, \pi_1)$ is a *resolution of the singularities of* $\mathcal{S}$, and $\mathcal{S}_{FM}$ is the *resolution space*. In this resolution, the "fibers" $\pi_1^{-1}([g])$ are a measure of the singularities in $\mathcal{S}$. Our construction is completely natural, as we now briefly describe.

Let $F(M)$ denote the frame bundle of M. Then $\mathcal{S}_{FM}$ is constructed by enlarging $\mathcal{M}$ to the product space $\mathcal{M} \times F(M)$ and considering the right action of $\mathcal{D}$ on the enlarged space

$$(\mathcal{M} \times F(M)) \times \mathcal{D} \to \mathcal{M} \times F(M),$$
$$((g,u),f) \mapsto (f^*g, f^*u),$$

where $f^*u = \hat{f}^{-1}(u)$ and $\hat{f}: F(M) \to F(M)$ is the natural lift of a diffeomorphism $f \in \mathcal{D}$ to the frame bundle $F(M)$. The advantage of introducing the action of $\mathcal{D}$ on $\mathcal{M} \times F(M)$ is that this action is now free, for if $(f^*g, f^*u) = (g,u)$, then $f \in I_g(M)$ is an isometry of g and f fixes the frame u. But by a classical theorem of Riemannian geometry, an isometry of a connected manifold that fixes a frame must be the identity (see, e.g., Helgason,[6] p. 62). Thus $f = \mathrm{id}_M$ and the action is free.

In Sec. III we show that the resulting orbit space

$$\mathcal{S}_{FM} = (\mathcal{M} \times F(M))/\mathcal{D}$$

is an ILH manifold which naturally projects onto $\mathcal{S}$,

$$\pi_1: \mathcal{S}_{FM} \to \mathcal{S}, \quad [(g,u)] \mapsto [g] .$$

Moreover, for $[g_0] \in \mathcal{S}$, the "fiber" $\pi_1^{-1}([g_0]) \subseteq \mathcal{S}_{FM}$ is diffeomorphic to the $(n^2 + n - k)$-dimensional orbit manifold

$$I_{g_0}(M) \backslash F(M),$$

where $k = \dim I_{g_0}(M)$ and $I_{g_0}(M)$ acts on $F(M)$ on the left by push-foward of frames

$$I_{g_0}(M) \times F(M) \to F(M), \quad (u,f) \mapsto \hat{f}(u) .$$

Note that if $\mathcal{S} = \mathcal{M}/\mathcal{D}$ were a manifold, then $\mathcal{M} \to \mathcal{S}$ would be a $\mathcal{D}$-principal fiber bundle (PFB) and

$$\mathcal{S}_{FM} = (\mathcal{M} \times F(M))/\mathcal{D} \to \mathcal{S} = \mathcal{M}/\mathcal{D}$$

would be the associated fiber bundle with standard fiber $F(M)$ over the base space $\mathcal{S}$. Although $\mathcal{M} \to \mathcal{S}$ is not in general a principal fiber bundle, this point of view is useful in motivating our construction (see discussion preceding Proposition 3.6 and the discussion following Theorem 6.1).

Our construction of $\mathcal{S}_{FM}$ is related to the following

 0022-2488/86/030718-21$02.50

construction (see Sec. IV and also Ebin,[3] Fischer,[4] and Bourguignon[5]). For a point $x_0 \in M$, let

$$\mathscr{D}'_{x_0} = \{f \in \mathscr{D} \mid f(x_0) = x_0 \text{ and } T_{x_0} f = I_{x_0}\}$$

denote the subgroup of $\mathscr{D}$ that fixes the point x_0 and whose derivative at x_0, $T_{x_0} f: T_{x_0}M \to T_{x_0}M$, is the identity I_{x_0} isomorphism of the tangent space $T_{x_0}M$. Then $\mathscr{D}'_{x_0}$ is a closed ILH Lie subgroup of $\mathscr{D}$ and acts on $\mathscr{M}$ by pullback,

$$\mathscr{M} \times \mathscr{D}'_{x_0} \to \mathscr{M}, \quad (g,f) \mapsto f^*g .$$

This action is free (for the reasons stated above) and the orbit space

$$\mathscr{G}_{x_0} = \mathscr{M}/\mathscr{D}'_{x_0}$$

is a manifold which naturally projects onto $\mathscr{G}$,

$$\pi: \mathscr{G}_{x_0} \to \mathscr{G}, \quad (g) \mapsto [g] .$$

If $[g_0] \in \mathscr{G}$, then the "fiber" $\pi^{-1}([g_0])$ can be canonically identified with the double coset manifold

$$I_{g_0}(M) \backslash \mathscr{D} / \mathscr{D}'_{x_0} = \{I_{g_0}(M) \circ f \circ \mathscr{D}'_{x_0} \mid f \in \mathscr{D}\} .$$

Moreover, if a frame $u_0 \in F(M)$ is chosen, then there is a diffeomorphism

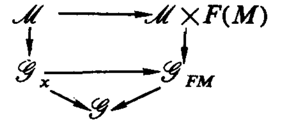
$$d_{u_0}^{-1}: \mathscr{G}_{FM} \to \mathscr{G}_{x_0} ,$$

which maps fibers of $\mathscr{G}_{FM}$ to fibers of $\mathscr{G}_{x_0}$ diffeomorphically [Proposition 6.2; if M is nonreversible, $F(M)$ must be replaced by $F^+_{u_0}(M)$; see Sec. II].

Thus the resolution $\mathscr{G}_{FM} \to \mathscr{G}$ may be thought of as a canonical resolution, and for each $x \in M$, there is a particular resolution $\mathscr{G}_x \to \mathscr{G}$. If a frame u at x is chosen, then u induces a "representation" $d_u^{-1}: \mathscr{G}_{FM} \to \mathscr{G}_x$ of the canonical resolution space $\mathscr{G}_{FM}$ onto the particular resolution space $\mathscr{G}_x$.

Although $\mathscr{G}_{FM}$ and $\mathscr{G}_x$ are diffeomorphic (not canonically), the construction of $\mathscr{G}_{FM}$ is more geometrical and natural than the construction of $\mathscr{G}_x$ inasmuch as we do not need to fix a point, thereby giving such a point a preferred status, and we do not need to restrict to diffeomorphisms in $\mathscr{D}'_x$. Thus the construction of $\mathscr{G}_{FM}$ is generally covariant, i.e., is covariant with respect to all diffeomorphisms, whereas the construction of $\mathscr{G}_x$ is covariant only with respect to the subgroup $\mathscr{D}'_x$. These considerations are of importance in applications in general relativity.

We also consider other relationships between $\mathscr{G}_{FM}$ and $\mathscr{G}_x$. In particular, we show that $\mathscr{G}_{FM}$ is the base space of a $\mathscr{D}$-PFB (principal fiber bundle) $\mathscr{M} \times F(M) \to \mathscr{G}_{FM}$ (Sec. III) and that $\mathscr{G}_x$ is the base space of a $\mathscr{D}'_x$-PFB $\mathscr{M} \to \mathscr{G}_x$ (Sec. V). Moreover, for each frame u at x there is a reduction of the $\mathscr{D}$-PFB to the $\mathscr{D}'_x$-PFB. These relationships are summarized by the commuting pentagon (see Sec. VI)



Lastly, we construct a fiber bundle (see Sec. VI)

$$E = E(M, \mathscr{G}_{FM}, \mathrm{GL}(n), F(M))$$
$$= (F(M) \times \mathscr{G}_{FM})/\mathrm{GL}(n)$$

over M with standard fiber $\mathscr{G}_{FM}$, with projection

$$\pi_E: E \to M ,$$

and which is associated to the frame bundle $F(M)$. Each fiber $\pi_E^{-1}(x)$ in this bundle can be canonically identified with $\mathscr{G}_x$ so that

$$E = \bigcup_{x \in M} \mathscr{G}_x .$$

Moreover, for a frame u at x, the usual identification in the construction of an associated fiber bundle of the standard fiber with the fiber at x

$$\mathscr{G}_{FM} \to \mathscr{G}_x \approx \pi_E^{-1}(x)$$

is given by d_u^{-1}. Thus we have constructed a bundle $E \to M$ whose standard fiber is the canonical resolution space $\mathscr{G}_{FM}$, and whose fiber at x is the particular resolution space $\mathscr{G}_x$. Thus E provides a bundle point of view for tying together the canonical resolution space $\mathscr{G}_{FM}$ with all of the particular resolution spaces $\mathscr{G}_x$, $x \in M$. Inasmuch as E is the totality of all the particular spaces, E may properly be deemed to be the *grand resolution space of* $\mathscr{G}$.

Throughout this paper we shall be dealing with infinite-dimensional manifolds of C^∞ maps such as $\mathscr{M}$, $\mathscr{D}$, and $\mathscr{D}'_x$, submanifolds of these spaces, and maps and diffeomorphisms between these spaces. When considering such manifolds and maps between such manifolds, we shall always take "manifold," "submanifold," "map," and "diffeomorphism" in the ILH sense (see Omori,[7] Ebin,[3] and Ebin and Marsden[8] for further details regarding these spaces).

II. SOME GEOMETRY OF THE FRAME BUNDLE

Throughout this section, M will denote a C^∞ connected n-dimensional manifold, but not necessarily compact. Let

$$\pi_{FM}: F(M) \to M$$

denote its frame bundle, a $\mathrm{GL}(n) = \mathrm{GL}(n, R)$ principal fiber bundle (PFB) over M. By differentiation, a diffeomorphism $f: M \to M$ has a *natural lift* to an automorphism $\hat{f}: F(M) \to F(M)$, where $\hat{f}$ maps the frame $u = (v_1,..., v_n)$ at x, $v_i \in T_xM$, $1 \leqslant i \leqslant n$, to the frame $\hat{f}(u) = T_xf \cdot v_1,..., T_xf \cdot v_n)$ at $f(x)$. The lift $\hat{f}$ covers f, i.e., $\pi_{FM} \circ \hat{f} = f \circ \pi_{FM}$, and if $A \in \mathrm{GL}(n)$, $\hat{f}(u \cdot A) = \hat{f}(u) \cdot A$, or $\hat{f} \circ R_A = R_A \circ \hat{f}$, where $R_A: F(M) \to F(M)$ is the diffeomorphism of the frame bundle corresponding to A. We let $\mathrm{AUT}(FM)$ denote the group of automorphisms of the frame bundle, so that $\hat{f} \in \mathrm{AUT}(FM)$. Then the natural lift induces a group monomorphism

$$\mathscr{D} \to \mathrm{AUT}(FM), \quad f \mapsto \hat{f},$$

since if $f_1, f_2 \in \mathscr{D}$, $\widehat{f_1 \circ f_2} = \hat{f}_1 \circ \hat{f}_2$.

Similarly, if $X \in \mathscr{X}(M)$ is a vector field on M, the *natural lift of* X is a vector field $\hat{X}: F(M) \to T(F(M))$ on $F(M)$, defined by

$$\hat{X}(u) = \frac{d}{d\lambda} \hat{f}_\lambda(u)|_{\lambda=0} \in T_u F(M) ,$$

where $u \in F(M)$, $x = \pi_{F(M)}(u) \in M$, and f_λ is the local one-parameter flow of X of local diffeomorphisms in a neighborhood of x. The natural lift satisfies $(R_A)_* \hat{X} = \hat{X}$, i.e., $\hat{X} \in \mathscr{X}_{\mathrm{GL}(n)}(FM)$, the Lie algebra of $\mathrm{GL}(n)$-invariant vector fields on $F(M)$. Moreover, with respect to the Lie bracket of vector fields on M and on $F(M)$, the natural lift induces a Lie algebra monomorphism

$$\mathscr{X}(M)\to\mathscr{X}_{\mathrm{GL}(n)}(FM),\quad X\mapsto\hat{X}: F(M)\to T(F(M)).$$

The group monomorphism $f\mapsto\hat{f}$ induces a left action of $\mathscr{D}$ on $F(M)$,

$$\mathscr{D}\times F(M)\to F(M),\quad (f,u)\mapsto\hat{f}(u),$$

the action of *push-forward of frames*. If we pretend that $\mathscr{D}$ is a Lie group with Lie algebra $\mathscr{X}(M)$, then the Lie algebra monomorphism $X\mapsto\hat{X}$ can be interpreted as the infinitesimal generator of this action. Note that the Lie algebra structure on $\mathscr{X}(M)$ given by the usual bracket of vector fields corresponds to the *right* Lie algebra structure of $\mathscr{D}$.

For a frame $u\in F(M)$, let

$$\mathscr{D}_u=\{f\in\mathscr{D}\,|\,\hat{f}(u)=u\}$$

denote the *isotropy group at u* of the action of $\mathscr{D}$ on $F(M)$, the subgroup of diffeomorphisms that fix the frame u. If $f\in\mathscr{D}_u$, then f must fix the base point $x=\pi_{FM}(u)$, and since $u=(v_1,\ldots,v_n)$ are linearly independent, T_xf: $T_xM\to T_xM$ must be the identity I_x of T_xM. Thus if $x=\pi_{FM}(u)$,

$$\mathscr{D}_u=\mathscr{D}'_x=\{f\in\mathscr{D}\,|\,f(x)=x\ \text{and}\ T_xf=I_x\}.$$

If M is orientable, then the frame bundle consists of two $\mathrm{GL}^0(n)$-principal fiber bundles, say $F^+(M)$ and $F^-(M)$, where $\mathrm{GL}^0(n)$ is the connected component of the identity of $\mathrm{GL}(n)$. Let $\mathscr{D}^+\subseteq\mathscr{D}$ denote the group of orientation-preserving diffeomorphisms of M. Then $\mathscr{D}^+$ acts transitively on both $F^+(M)$ and $F^-(M)$. Thus if there exists an orientation-reversing diffeomorphism of M, then $\mathscr{D}$ acts transitively on $F(M)$, and thus if $u\in F(M)$ is a frame at x, then

$$\mathscr{D}/\mathscr{D}_u=\mathscr{D}/\mathscr{D}'_x\to F(M),\quad f\circ\mathscr{D}'_x\to\hat{f}(u)$$

is a bijection. If there does not exist an orientation-reversing diffeomorphism, then $\mathscr{D}=\mathscr{D}^+$ and $\mathscr{D}$ is transitive on $F^+(M)$ and $F^-(M)$ separately, but is not transitive on $F(M)$. In this case let $F^+_u(M)$ denote the bundle of oriented frames that have the same orientation as u [i.e., the bundle $F^+(M)$ such that $u\in F^+(M)$]. Then

$$\mathscr{D}/\mathscr{D}_u=\mathscr{D}/\mathscr{D}'_x\to F^+_u(M),\quad f\circ\mathscr{D}'_x\to\hat{f}(u)$$

is a bijection. We shall say that M is *reversible* if M is orientable and there exists an orientation-reversing diffeomorphism, and M is *nonreversible* if M is orientable and there does not exist an orientation-reversing diffeomorphism (see Fried[9] for examples of nonreversible manifolds).

If M is not orientable, then $F(M)$ is connected, and so $\mathscr{D}$ is transitive on $F(M)$. In this case,

$$\mathscr{D}/\mathscr{D}_u=\mathscr{D}/\mathscr{D}'_x\to F(M),\quad f\circ\mathscr{D}'_x\to\hat{f}(u)$$

is a bijection.

The above bijections are interesting inasmuch as they represent the frame bundle as a homogeneous space. In Sec. IV, we shall show that if M is compact, then the coset spaces $\mathscr{D}/\mathscr{D}'_x$ are manifolds, and that they are diffeomorphic to $F(M)$ [or $F^+_u(M)$].

Now let g be a Riemannian metric on M, let Γ denote its associated Levi-Civita connection on $F(M)$, and let ω denote the corresponding $\mathfrak{gl}(n)=\mathfrak{gl}(n;R)$-valued connection one-form on $F(M)$. Let

$$\gamma: \mathfrak{gl}(n)\times\mathfrak{gl}(n)\to\mathbb{R},\quad (C,D)\to\mathrm{tr}(C^t\circ D)$$

denote the "Euclidean" inner product on $\mathfrak{gl}(n)$. Here C^t denotes the transpose of C with respect to the Euclidean metric on $\mathbb{R}^n$. In coordinates,

$$\gamma(C,D)=\sum_{i,j=1}^{n}C^i_jD^i_j.$$

The metric g, its Levi-Civita connection one-form ω, and the inner product γ induce a natural Riemannian metric $\hat{g}$ on $F(M)$, defined by

$$\hat{g}=(\pi_{FM})^*g+\gamma\cdot(\omega\otimes\omega),$$

so that if $u\in F(M)$ and $Z_1, Z_2\in T_u\,F(M)$, $\hat{g}$ is given by

$$\hat{g}(u)\cdot(Z_1,Z_2)=g(T_u\pi_{FM}\cdot Z_1, T_u\pi_{FM}\cdot Z_2)+\gamma\cdot(\omega(Z_1),\omega(Z_2)).$$

Note that if $A\in O(n)$, $R^*_A\hat{g}=\hat{g}$, but that $\hat{g}$ is not invariant by $\mathrm{GL}(n)$. Also, the projection π_{FM}: $F(M)\to M$ is a Riemannian submersion with respect to the Riemannian metrics $\hat{g}$ on $F(M)$ and g on M.

Let

$$I_g(M)=\{f\in\mathscr{D}\,|\,f^*g=g\}$$

denote the Lie group of *isometries of* (M,g), and let

$$\mathscr{I}_g(M)=T_e(I_g(M))=\{X\in\mathscr{X}(M)\,|\,L_Xg=0\ \text{and}\ X\ \text{is a complete vector field}\}$$

denote its Lie algebra of *complete Killing vector fields* (or *complete infinitesimal isometries*), taken with the Lie algebra structure given by the usual bracket of vector fields. This Lie algebra structure correponds to the Lie algebra of *right* invariant vector fields on $I_g(M)$.

$I_g(M)$ acts on M on the left as a Lie transformation group,

$$I_g(M)\times M\to M,\quad (f,x)\mapsto f(x),$$

and the action can be lifted to a left action of $I_g(M)$ on $F(M)$,

$$I_g(M)\times F(M)\to F(M),\quad (f,u)\mapsto\hat{f}(u)$$

(see Fig. 1). The infinitesimal generator of this lifted action is given by

$$\mathscr{I}_g(M)\to\mathscr{X}_{\mathrm{GL}(n)}(FM),\quad X\mapsto\hat{X}.$$

For $u\in F(M)$, let

$$\chi_u: I_g(M)\to F(M),\quad f\to\hat{f}(u)$$

denote the orbit map through u, and let

$$\hat{I}_g(u)=\{\hat{f}(u)\,|\,f\in I_g(M)\}\subseteq F(M)$$

denote the orbit through u. Then χ_u is a smooth map with derivative at the identity $e=\mathrm{id}_M\in I_g(M)$ given by

$$T_e\chi_u: \mathscr{I}_g(M)\to T_uF(M),\quad X\to\hat{X}(u).$$

Let

$$\hat{\mathscr{I}}_g(u)=\text{range}\ T_e\chi_u=\{X(u)\in T_uF(M)\,|\,X\in\mathscr{I}_g(M)\}\subseteq T_uF(M).$$

The following computation will be of use.

Lemma 2.1: Let $f\in\mathscr{D}$. Then

$$\hat{\mathscr{I}}_{(f^{-1})^*g}(\hat{f}(u))=T_u\hat{f}(\hat{\mathscr{I}}_g(u)).$$

Proof: The isometry group of $(f^{-1})^*g$ is $I_{(f^{-1})^*g}(M) = f \circ I_g(M) \circ f^{-1}$, and so $\mathscr{I}_{(f^{-1})^*g}(M) = f_*(\mathscr{I}_g(M))$. Thus if $X \in \mathscr{X}(M)$,

$$f_* X(\hat{f}(u)) = (\hat{f}_* \hat{X})(\hat{f}(u)) = T_u \hat{f} \circ \hat{X} \circ \hat{f}^{-1}(\hat{f}(u)) = T_u \hat{f} \cdot \hat{X}(u).$$

Thus

$$\hat{\mathscr{I}}_{(f^{-1})^*g}(\hat{f}(u)) = f_*(\mathscr{I}_g(M)) = \{\widehat{f_* X}(\hat{f}(u)) = T_u \hat{f} \cdot \hat{X}(u) \mid X \in \mathscr{I}_g(M)\} = T_u \hat{f} \cdot (\mathscr{I}_g(M)).$$

□

Concerning the action of $I_g(M)$ on $F(M)$, we have the following theorem.

Theorem 2.2: The action

$$I_g(M) \times F(M) \to F(M), \quad (f,u) \to \hat{f}(u)$$

of push-forward of frames is smooth, free, and proper. For $u \in F(M)$, the orbit

$$\hat{I}_g(u) = \{\hat{f}(u) \mid f \in I_g(M)\} \subseteq F(M)$$

is a closed submanifold of $F(M)$ with tangent space at $u_1 = \hat{f}(u) \in \hat{I}_g(u)$ given by

$$\hat{\mathscr{I}}_g(u_1) = \{\hat{X}(u_1) \mid X \in \mathscr{I}_g(M)\} = T_u \hat{f} \cdot (\hat{\mathscr{I}}_g(u)).$$

The orbit map

$$\chi_u : I_g(M) \to \hat{I}_g(u) \subseteq F(M)$$

is a diffeomorphism onto its image.

With respect to the metric $\hat{g} = \pi_{FM}^* g + \gamma \cdot (\omega \otimes \omega)$, $I_g(M)$ acts on $F(M)$ as a group of isometries.

Proof: The smoothness of the action is a consequence of the smoothness of the action of $I_g(M)$ on M. The action is free, since if $\hat{f}(u) = u$, then $f(x_0) = x_0$, where $x_0 = \pi_F(u)$, and $T_{x_0} f = I_{x_0}$. Since f is an isometry (and M is connected), f is the identity.

To show properness, we must show that the map

$$I_g(M) \times F(M) \to F(M) \times F(M), \quad (f,u) \mapsto (u, \hat{f}(u))$$

is a proper mapping, i.e., that the inverse image of a compact set is compact. Equivalently, if $(u_n, \hat{f}_n(u_n)) \to (u_0, u_1)$ converges, then f_n has a convergent subsequence in $I_g(M)$. But $u_n \to u_0$ implies $x_n = \pi_{FM}(u_n) \to x_0 = \pi_{FM}(u_0)$, and $\hat{f}_n(u_n) \to u_1$ implies $f_n(x_n) \to x_1 = \pi_{FM}(u_1)$. Then, by the properness of the action of $I_g(M)$ on M, f_n has a convergent subsequence in $I_g(M)$.

Since the action is free, the orbit map χ_u: $I_g(M) \to F(M)$ is an injective immersion, and so the orbit $\hat{I}_g(u)$ is an immersed submanifold. Since the action is proper, the orbit mapping is a closed mapping, and hence is a homeomorphism onto its image. Hence the orbit $\hat{I}_g(u)$ is a closed submanifold and χ_u is a diffeomorphism onto $\hat{I}_g(u)$. At $u_1 = \hat{f}(u) \in \hat{I}_g(u)$, the tangent space $T_{u_1}(\hat{I}_g(u)) = \text{range-}T_e\chi_{u_1} = \hat{\mathscr{I}}_g(u_1) = T_u \hat{f} \cdot (\hat{\mathscr{I}}_g(u))$, where the last equality follows from Lemma 2.1 (since $f^*g = g$).

Lastly, for any diffeomorphism $f \in \mathscr{D}$,

$$\hat{f}^*\hat{g} = \hat{f}^*(\pi_{FM}^* g + \gamma \cdot (\omega \otimes \omega)) = \pi_{FM}^*(f^*g) + \gamma \cdot (\hat{f}^*\omega \otimes \hat{f}^*\omega).$$

Thus if $f \in I_g(M)$, $f^*g = g$ and $\hat{f}^*\omega = \omega$, so $\hat{f}^*\hat{g} = \hat{g}$. Thus $I_g(M)$ acts as a group of isometries on $F(M)$ with respect to the Riemannian metric $\hat{g}$. □

Since the action of $I_g(M)$ on $F(M)$ is smooth, free, and proper, and is a group of isometries with respect to the Riemannian metric $\hat{g}$, a standard construction asserts the existence of local cross sections that are orthogonal to the orbits, and that are equivariant with respect to the action (see, e.g., Palais,[10] p. 108). These local cross sections are constructed as follows. Let $\exp_{\hat{g}}(u)$ denote the exponential map of $\hat{g}$ on $F(M)$ at $u \in F(M)$, and let $N_u \subseteq T_u F(M)$ denote a normal neighborhood of the origin of $T_u F(M)$. Thus

$$\exp_{\hat{g}}(u) : N_u \to F(M)$$

is a diffeomorphism onto a neighborhood of u. Now $\hat{\mathscr{I}}_g(u) \subseteq T_u F(M)$ is the tangent space at u to the orbit $\hat{I}_g(u)$. Let $\hat{\mathscr{I}}_g^{\perp}(u) \subseteq T_u F(M)$ denote the orthogonal complement to $\hat{\mathscr{I}}_g(u)$ with respect to the inner product $\hat{g}(u)$ on $T_u F(M)$. Thus we have the orthogonal direct sum

$$T_u F(M) = \hat{\mathscr{I}}_g(u) \otimes \hat{\mathscr{I}}_g^{\perp}(u)$$

(see Fig. 1). Then a local cross-section at u for the action of $I_g(M)$ on $F(M)$ is given by

$$C_u = \exp_{\hat{g}}(u)\,(N_u \cap \hat{\mathscr{I}}_g^{\perp}(u)) \subseteq F(M).$$

The following properties of these local cross sections are standard.

Proposition 2.3: The subspace C_u given above is a closed submanifold of $F(M)$ containing u, and C_u has the following properties: (1) C_u is orthogonal to the orbit $\hat{I}_g(u)$, i.e.,

$$T_u C_u = \hat{\mathscr{I}}_g^{\perp}(u);$$

(2) C_u is equivariant with respect to the action of $I_g(M)$ on $F(M)$, i.e., if $f \in I_g(M)$,

$$C_{\hat{f}(u)} = \hat{f}(C_u);$$

and (3) C_u is a local cross section for the action, i.e., (3a) if $f \in I_g(M)$ and $\hat{f}(C_u) \cap C_u \neq \varnothing$, then $f = \mathrm{id}_M$, and (3b) when restricted to C_u, the action

$$I_g(M) \times C_u \to F(M), \quad (f,u') \mapsto \hat{f}(u')$$

is a diffeomorphism onto an open invariant neighborhood of the orbit $\hat{I}_{g(u)}$. □

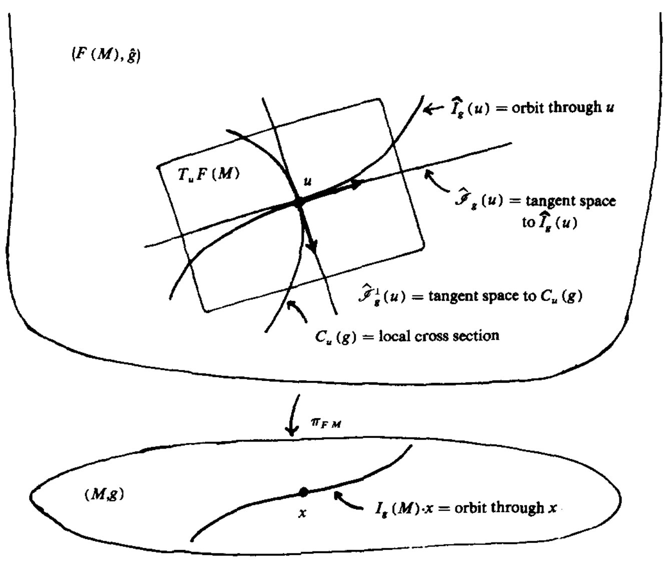


FIG. 1. The action of $I_g(M)$ on $F(M)$.

Remarks: (1) The usual properties of a slice for a group action (see, e.g., Palais[10]) are restricted to properties (3) and (4) above when the action is free, i.e., when the slice is a cross section.

(2) We shall also need to know the behavior of the cross section C_u under the action of $\mathscr{D}$. For this purpose, let $C_u(g) = C_u$ denote the cross section with its dependence on g explicitly displayed. For $f \in \mathscr{D}$, the exponential map $\exp_g$ satisfies

$$\exp_{(\hat{f}^{-1})^* g}(\hat{f}(u)) = \hat{f} \exp_g(u) \cdot (T_u \hat{f})^{-1},$$

so that the cross section satisfies

$$C_{\hat{f}(u)}((f^{-1})^* g) = \hat{f}(C_u(g)).$$

Note that property (2) above follows when $f \in I_g(M)$. □

The following assemblage of results is a corollary of the existence of the local equivariant cross sections constructed above (see, e.g., Abraham and Marsden,[11] pp. 266 and 276).

Theorem 2.4: For the action of $I_g(M)$ on $F(M)$,

$$I_g(M) \times F(M) \to F(M), \quad (f,u) \mapsto \hat{f}(u),$$

the orbit space

$$I_g(M) \backslash F(M)$$

has a smooth manifold structure such that the orbit projection map

$$\pi: F(M) \to I_g(M) \backslash F(M), \quad u \mapsto [u]$$

is a submersion. Moreover, for $u \in F(M)$,

$$\ker T_u \pi = T_u(\hat{I}_g(u)) = \hat{\mathscr{I}}_g(u)$$

and

$$\text{range } T_u \pi = T_{[u]}(I_g(M) \backslash F(M)) \approx \hat{\mathscr{I}}_g(u) \backslash T_u F(M) \approx \hat{\mathscr{I}}_g^{\perp}(u).$$

The submersion π: $F(M) \to I_g(M) \backslash F(M)$ has the structure of a smooth (left) principal fiber bundle with total space $F(M)$, base space $I_g(M) \backslash F(M)$, and structure group $I_g(M)$. □

III. RESOLVING THE SINGULARITIES IN $\mathscr{S}$

Throughout the rest of this paper, M will denote a C^∞ *compact* connected n-dimensional manifold. We now combine the action of $\mathscr{D}$ on $\mathscr{M}$ with the action of $\mathscr{D}$ on $F(M)$. The resulting orbit space will resolve the singularities in $\mathscr{M}/\mathscr{D}$.

Dual to the left action of $\mathscr{D}$ on $F(M)$, we have the right action of *pull-back of frames by diffeomorphisms*,

$$F(M) \times \mathscr{D} \to F(M), \quad (u,f) \mapsto f^* u \equiv (\hat{f})^{-1}(u).$$

On the product manifold $\mathscr{M} \times F(M)$, $\mathscr{D}$ then acts on the right by pull-back,

$$\Phi: (\mathscr{M} \times F(M)) \times \mathscr{D} \to \mathscr{M} \times F(M),$$

$$((g,u),f) \mapsto (f^* g, f^* u).$$

With respect to the ILH manifold structures of $\mathscr{M}$ and $\mathscr{D}$, this action is ILH smooth (see Ebin[3]).

For $(g,u) \in \mathscr{M} \times F(M)$, let

$$\Phi_{(g,u)}: \mathscr{D} \to \mathscr{M} \times F(M), \quad f \mapsto (f^* g, f^* u)$$

denote the *orbit map at* (g,u), and let

$$\mathscr{O}_{(g,u)} = \Phi_{(g,u)}(\mathscr{D}) = \{(f^* g, f^* u) \mid f \in \mathscr{D}\} \subseteq \mathscr{M} \times F(M)$$

denote the *orbit through* (g,u). For $f \in \mathscr{D}$, let

$$\Phi_f = f^*: \mathscr{M} \times F(M) \to \mathscr{M} \times F(M), \quad (g,u) \mapsto (f^* g, f^* u)$$

denote the diffeomorphism of $\mathscr{M} \times F(M)$ corresponding to f.

Let

$$S_2(M) = C^\infty (T^*M \otimes_{\text{sym}} T^*M)$$

denote the space of smooth two-covariant symmetric tensor fields of M. Since $\mathscr{M}$ is open in $S_2(M)$, the tangent space of $\mathscr{M}$ at g is

$$T_g \mathscr{M} = \{g\} \times S_2(M) \approx S_2(M),$$

which we identify with $S_2(M)$. Thus the tangent space of $\mathscr{M} \times F(M)$ at $(g,u) \in \mathscr{M} \times F(M)$ is taken to be

$$\begin{aligned} T_{(g,u)}(\mathscr{M} \times F(M)) &= T_g \mathscr{M} \times T_u F(M) \\ &= S_2(M) \times T_u F(M) \\ &= S_2(M) \oplus T_u F(M), \end{aligned}$$

where we take the tangent space with its direct sum structure.

The orbit map $\Phi_{(g,u)}$ is smooth with derivative at the identity $e \in \mathscr{D}$ given by

$$\begin{aligned} T_e \Phi_{(g,u)}: \mathscr{X}(M) \to T_{(g,u)}&(\mathscr{M} \times F(M)) \\ &= S_2(M) \times T_u F(M), \end{aligned}$$

$$X \mapsto L_X g - \hat{X}(u),$$

where $\hat{X}$ is the natural lift of the vector field X. Thus the infinitesimal generator of the action is given by

$$\mathscr{X}(M) \to \mathscr{X}(\mathscr{M} \times F(M)),$$

$$X \mapsto X^*: \mathscr{M} \times F(M) \to T(\mathscr{M} \times F(M)),$$

where

$$\begin{aligned} X^*(g,u) &= T_e \Phi_{(g,u)} \cdot X \\ &= L_X g - \hat{X}(u) \in S_2(M) \oplus T_u F(M). \end{aligned}$$

On $\mathscr{M} \times F(M)$, we introduce the following weak L_2-Riemannian metric Λ. For $(g,u) \in \mathscr{M} \times F(M)$, $h_1, h_2 \in S_2(M)$, and $Z_1, Z_2 \in T_u F(M)$, let

$$\Lambda(g,u) \cdot ((h_1, Z_1), (h_2, Z_2)) = \left(\int_M \langle h_1, h_2 \rangle_g \, du_g \right) + \hat{g}(u) \cdot (Z_1, Z_2),$$

where $\langle h_1, h_2 \rangle_g$ denotes the pointwise metric on $T^*M \otimes_{\text{sym}} T^*M$ induced by g, du_g is the volume element associated with g (a measure, not an n form, unless M is orientable), and $\hat{g} = \pi_{FM}^* g + \gamma \cdot (\omega \otimes \omega)$ is the metric on $F(M)$ introduced in Sec. II.

For $s > n/2$, we let $\mathscr{M}^s$ denote the Hilbert manifold of H^s Riemannian metrics, $S_s^2(M)$ the space of H^s two-covariant symmetric tensor fields, and $\mathscr{D}^{s+1}$ the group of H^{s+1} diffeomorphisms of M (see Omori,[7] Ebin,[3] and Ebin and Marsden[8] for more information about these spaces). Then $\mathscr{D}^{s+1}$ is a topological group and acts continuously on $\mathscr{M}^s$. Let

$$\Phi^s: (\mathscr{M}^s \times F(M)) \times \mathscr{D}^{s+1} \to \mathscr{M}^s \times F(M),$$

$$((g,u),f)\mapsto(f^*g,f^*u)$$

denote the corresponding C^0 action on $\mathscr{M}^s\times F(M)$.

The action Φ has the following properties.

Theorem 3.1. Let

$$\Phi: (\mathscr{M}\times F(M))\times\mathscr{D}\to\mathscr{M}\times F(M),$$
$$((g,u),f)\mapsto(f^*g,f^*u)$$

be the action described above. Then Φ is a smooth, free, and proper action, and $\mathscr{D}$ acts as a group of isometries with respect to the weak Riemannian metric Λ on $\mathscr{M}\times F(M)$.

Proof: As in Ebin,[3] the C^0 action

$$\Phi^{s+1}: (\mathscr{M}^s\times F(M))\times\mathscr{D}^{s+1}\to\mathscr{M}^s\times F(M)$$

induces an ILH-smooth action of $\mathscr{D}$ on $\mathscr{M}\times F(M)$. The action is free, since if $(f^*g,f^*u)=(g,u)$, $f^*g=g$ and $\hat{f}(u)=u$, so f is an isometry fixing a frame and so must be the identity. To show the action is proper, we must show that the map

$$(\mathscr{M}\times F(M))\times\mathscr{D}\to(\mathscr{M}\times F(M))\times(\mathscr{M}\times F(M)),$$
$$((g,u),f)\mapsto((g,u),\ (f^*g,f^*u))$$

is proper. Thus let $(g_n,u_n)\to(g_0,u_0)$ and $(f_n^*g_n, f_n^*u_n)\to(g_1,u_1)$ converge in $(\mathscr{M}\times F(M))\times(\mathscr{M}\times F(M))$. Then $g_n\to g_0$ and $f_n^*g_n\to g_1$, so by the properness of the action of $\mathscr{D}$ on $\mathscr{M}$ [see remark (3) below], f_n has a convergent subsequence f_{n_i} in $\mathscr{D}$. Thus the action of $\mathscr{D}$ on $\mathscr{M}\times F(M)$ is proper.

To show that Λ is invariant by $\mathscr{D}$, let $f\in\mathscr{D}$, $(g,u)\in\mathscr{M}\times F(M)$, $h_1, h_2\in S_2(M)$, and $Z_1, Z_2\in T_uF(M)$. Then

$$\begin{aligned}((f^*)^*\Lambda(g,u))\cdot((h_1,Z_1),(h_2,Z_2)) &= \Lambda(f^*g,f^*u)\cdot((f^*h_1,T_uf^*\cdot Z_1),(f^*h_2,T_uf^*\cdot Z_2))\\ &= \left(\int\langle f^*h_1,f^*h_2\rangle_{f^*g}\,du_{f^*g}\right)\\ &\quad+f^*g(f^*u)\cdot(T_uf^*h_2\cdot Z_1,T_uf^*\cdot Z_2)\\ &= \left(\int(\langle h_1,h_2\rangle_g\circ f)\,f^*(du_g)\right)\\ &\quad+(\hat{f})^*\hat{g}(\hat{f}^{-1}(u))\cdot(T_u\hat{f}^{-1}\cdot Z_1,T_u\hat{f}^{-1}\cdot Z_2)\\ &= \left(\int f^*(\langle h_1,h_2\rangle_g\,du_g)\right)+\hat{g}(u)\cdot(Z_1,Z_2)\\ &= \left(\int\langle h_1,h_2\rangle_g du_g\right)+\hat{g}(u)\cdot(Z_1,Z_2)\\ &= \Lambda(g,u)\cdot((h_1,Z_1),(h_2,Z_2)).\end{aligned}$$

□

Remarks: (1) It is interesting that the action of $\mathscr{D}$ on $\mathscr{M}\times F(M)$ is free although it is not free on either of the factors.

(2) Generally, if $M_1\times G\to M$ is a proper action, and $M_2\times G\to M_2$ is any other action, then the action on the product space $(M_1\times M_2)\times G\to M_1\times M_2$ is proper.

(3) Ebin (Ref. 3, Proposition 6.13) shows that if $g_1, g_2\in\mathscr{M}^s$, $s>n/2$, and $f_n\in\mathscr{D}^{s+1}$ is a sequence such that $f_n^*g_1\to g_2$, then f_n has a convergent subseqence in $\mathscr{D}^{s+1}$. Using the strong H^s metric ρ^s on $\mathscr{D}^s$, it then follows that the action of $\mathscr{D}^{s+1}$ on $\mathscr{M}^s$ is proper (see Palais[12]). □

Since we are working with infinite-dimensional manifolds, we cannot immediately conclude (as we can for finite-dimensional manifolds) that the orbits are embedded submanifolds, or that the quotient space $(\mathscr{M}\times F(M))/\mathscr{D}$ is a manifold. To get the first of these results, we must show that the orbit map is an immersion, and to get the second, we must construct local equivariant cross sections for the action (see Palais[12]). For these results, we shall need a direct sum decomposition of $S_2(M)\oplus T_uF(M)$.

First we recall some results of Ebin.[3] Let

$$\Psi: \mathscr{M}\times\mathscr{D}\to\mathscr{M},\quad (g,f)\mapsto f^*g$$

denote the usual pull-back action of $\mathscr{D}$ on $\mathscr{M}$. For $g\in\mathscr{M}$, let

$$\Psi_g: \mathscr{D}\to\mathscr{M},\quad f\mapsto f^*g$$

denote the orbit map at g. Then Ψ_g is a smooth map with derivative at the identity $e\in\mathscr{D}$ given by

$$\alpha_g=T_e\Psi_g: \mathscr{X}(M)\to S_2(M),\quad X\mapsto L_Xg,$$

where $L_Xg\in S_2(M)$ is the Lie derivative of g with respect to the vector field X. Then range α_g is closed in $S_2(M)$ and has closed L_2-orthogonal complement

$$\mathring{S}_2(g)=\{h\in S_2(M)\,|\,\delta_gh=0\},$$

the space of C^∞ divergence free two-covariant symmetric tensor fields on M. In local coordinates, the divergence is given by $(\delta_gh)_i=-g^{jk}h_{ij|k}$, where the vertical bar denotes covariant differentiation. Thus $S_2(M)$ splits L_2-orthogonally as

$$S_2(M)=\mathring{S}_2(g)\oplus\text{range }\alpha_g,\quad h=\mathring{h}+L_Xg.$$

The pieces $\mathring{h}$ and L_Xg are uniquely determined, but the vector field X is determined only up to a Killing vector field. We shall refer to this splitting as the *canonical splitting of* $S_2(M)$.

Let

$$\mathscr{O}_g=\{f^*g\,|\,f\in\mathscr{D}\}\subseteq\Psi_g(\mathscr{D})\subseteq\mathscr{M}$$

denote the orbit through g. Then $\mathscr{O}_g$ is a smooth closed submanifold of $\mathscr{M}$ with tangent space at g given by

$$T_g\mathscr{O}_g=\text{range }\alpha_g.$$

Orthogonal to $\mathscr{O}_g$ there exists a slice $S_g\subseteq\mathscr{M}$, also a smooth closed manifold of $\mathscr{M}$, with tangent space at g given by

$$T_gS_g=\mathring{S}_2(g).$$

Thus the canonical splitting can be written as

$$T_g\mathscr{M}=T_gS_g\oplus T_g\mathscr{O}_g.$$

We now construct a similar splitting for $S_2(M)\oplus T_uF(M)$. Recall that for $g\in\mathscr{M}$ and $u\in F(M)$, $\hat{I}_g(u)\subseteq F(M)$ is the orbit of $I_g(M)$ through u with tangent space at u given by $T_u(\hat{I}_g(u))\subseteq\hat{\mathscr{I}}_g(u)\subseteq T_uF(M)$. Let $\hat{\mathscr{I}}_g^\perp(u)$ denote the orthogonal complement of $\hat{\mathscr{I}}_g(u)$ with respect to $\hat{g}(u)$, so that we have the $\hat{g}(u)$-orthogonal splitting of $T_uF(M)$,

$$T_uF(M)=\hat{\mathscr{I}}_g(u)\oplus\hat{\mathscr{I}}_g^\perp(u)$$

(see Fig. 2).

Theorem 3.2: For $(g,u)\in\mathscr{M}\times F(M)$, let

$$\alpha_{(g,u)}=T_e\Phi(g,u): \mathscr{X}(M)\to S_2(M)\oplus T_uF(M),$$
$$X\mapsto X^*(g,u)=L_Xg-\hat{X}(u)$$

be the derivative of the orbit map

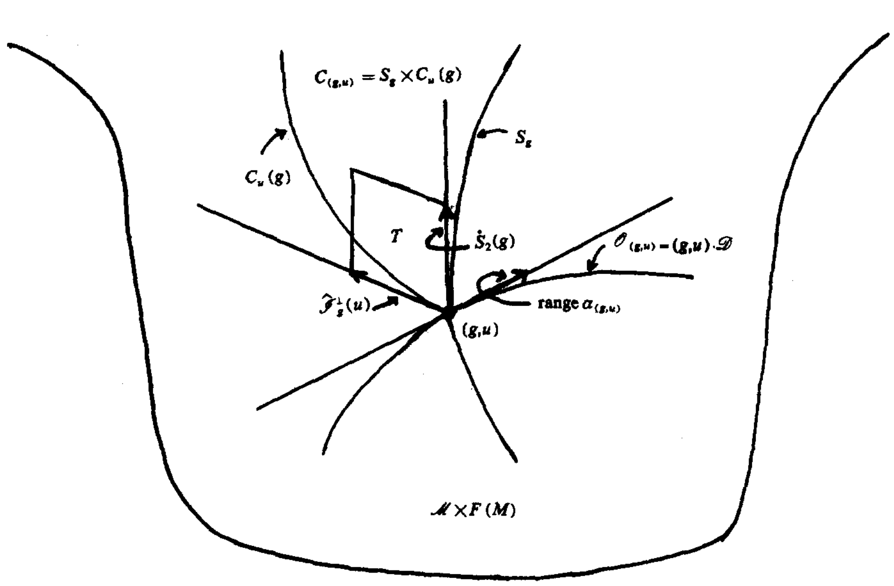


FIG. 2. The direct sum decomposition

$$T_{(g,u)}(\mathscr{M}\times F(M)) = S_2(M) \oplus T_u F(M) = \mathring{S}_2(g) \oplus \mathscr{I}^{\perp}_g(u) \oplus \text{range}\, \alpha_{(g,u)}$$

with

$$T = T_{(g,u)} C_{(g,u)} = \mathring{S}_2(g) \oplus \hat{\mathscr{I}}_g(u).$$

$$\Phi_{(g,u)} : \mathscr{D} \to \mathscr{M}\times F(M), \quad f \mapsto (f^*g, f^*u)$$

at the identity $e\in\mathscr{D}$. Then $\alpha_{(g,u)}$ has closed range and $\mathring{S}_2(g) \oplus \hat{\mathscr{I}}^{\perp}_g(u)$ is a closed complement of range $\alpha_{(g,u)}$ in $S_2(M) \oplus T_u F(M)$. Thus there is a direct sum decomposition

$$S_2(M) \oplus T_u F(M) = \mathring{S}_2(g) \oplus \hat{\mathscr{I}}^{\perp}_g(u) \oplus \text{range } \alpha_{(g,u)}$$

If $h + Z_u \in S_2(M) \oplus T_u F(M)$, then

$$\begin{aligned} h + Z_u &= \mathring{h} + W^{\perp}_u + \alpha_{(g,u)}(X+Y) \\ &= \mathring{h} + W^{\perp}_u + (L_{(X+Y)} g - (\hat{X}(u) + \hat{Y}(u))), \end{aligned}$$

where $h = \mathring{h} + L_X g$ is the canonical splitting of h, $W_u = Z_u + \hat{X}(u)$, and $W_u = W^{\perp}_u + W^{\parallel}_u \in \hat{\mathscr{I}}^{\perp}_g(u)$,is the splitting of W_u according to the $\hat{g}(u)$-orthogonal splitting of $T_u F(M)$, and $Y \in \mathscr{I}_g(M)$ is the unique Killing vector field on M such that $\hat{Y}(u) = -W^{\parallel}_u$.

Each of the above summands is equivalent with respect to $\mathscr{D}$, i.e., if $f\in\mathscr{D}$, then

$$\mathring{S}_2(f^*g) = f^*(\mathring{S}_2(g)),$$
$$\hat{\mathscr{I}}^{\perp}_{f^*g}(f^*u) = T_u f^{-1}(\hat{\mathscr{I}}^{\perp}_g(u)),$$

and

$$\begin{aligned} \text{range } \alpha_{(f^*g, f^*u} &= T_{(g,u)} f^*(\text{range } \alpha_{(g,u)}) \\ &= \{T_{(g,u)} f^*\cdot(h + Z_u) \\ &\quad = f^*h + T_u \hat{f}^{-1}\cdot Z_u \,|\, h + Z_u \in \text{range } \alpha_{(g,u)}\}, \end{aligned}$$

where the "$\perp$" on the left-hand side of the second equation is the $(\hat{f}^*\hat{g})(f^*u)$-orthogonal complement of $\hat{\mathscr{I}}_{f^*g}(f^*u)$.

Proof: $\alpha_{(g,u)}$ is injective, since if $\alpha_{(g,u)}(X) = L_X g - \hat{X}(u) = 0$, then $L_X g = 0$ and $\hat{X}(u) = 0$. Thus $X(x_0) = 0$, $x_0 = \pi_{FM}(u)$, and $T_{x_0} X = 0$. But it is classical that a Killing vector field on a connected manifold that vanishes at a point and whose derivative vanishes at that point must be identically zero. Note that if $X(x_0) = 0$, $T_{x_0} X = 0$ is equivalent to $\nabla X(x_0) = 0$.

Since $\ker \alpha_{(g,u)} = 0$, it follows by an elementary argument that a necessary and sufficient condition for $\alpha_{(g,u)}$ to have closed range is that if $X_n \in \mathscr{X}(M)$ is a sequence of vector fields such that $\alpha_{(g,u)}(X_n) \to 0$, then there exists a subsequence $X_{n_i} \to 0$. So assume $\alpha_{(g,u)}(X_n) = L_{X_n} g - \hat{X}_n(u) \to 0$. Then $\alpha_g(X_n) = L_{X_n} g \to 0$ and $\hat{X}_n(u) \to 0$. Since α_g has closed range (Ebin,[3] Proposition 6.10), X_n has a convergent subsequence $X_{n_i} \to X$ that converges to an element $X \in \mathscr{I}_g(M)$. Thus $\hat{X}_{n_i}(u) \to \hat{X}(u) = 0$. Thus from the above $X = 0$. Thus range $\alpha_{(g,u)}$ is closed in $S_2(M) \oplus T_u F(M)$, and $\alpha_{(g,u)}$ is an isomorphism onto its range.

Now $\mathring{S}_2(g) = \ker \delta_g$ is closed in $S_2(M)$, and since $\hat{\mathscr{I}}^{\parallel}_g(u)$ is finite dimensional, $\mathring{S}_2(g) \oplus \hat{\mathscr{I}}^{\perp}_g(u)$ is closed in $S_2(M) \oplus T_u F(M)$. Thus both range $\alpha_{(g,u)}$ and $\mathring{S}_2(g) \oplus T_u F(M)$ are closed in $S_2(M) \oplus T_u F(M)$. Since algebraically complementary *closed* subspaces of a Fréchet space are topologically complementary (Wilansky,[13] p. 62), it is sufficient to show that $\mathring{S}_2(g) \oplus \hat{\mathscr{I}}^{\perp}_g(u)$ is an algebraic complement to range $\alpha_{(g,u)}$.

Thus let $h + Z_u \in (\mathring{S}_2(g) \oplus \hat{\mathscr{I}}^{\perp}_g(u)) \cap$ range $\alpha_{(g,u)}$. Thus $\delta_g h = 0$, $Z_u \in \hat{\mathscr{I}}^{\perp}_g(u)$, and there exists an $X \in \mathscr{X}(M)$ such that $h + Z_u = L_X g - X(u)$. Thus $h = L_X g = 0$ by the L_2-orthogonality of the canonical splitting, and so X is a killing vector field. Thus $\hat{X}(u) \in \hat{\mathscr{I}}_g(u)$. But $\hat{X}(u) = -Z_u \in \hat{\mathscr{I}}^{\perp}_g(u)$, so $\hat{X}(u) = 0$. Hence, as above, $X = 0$, and so $h + Z_u = 0$.

Now let $h + Z_u \in S_2(M) \oplus T_u F(M)$. Let $h = \mathring{h} + L_X g$ be the canonical splitting of h, so X is determined only up to a Killing vector field. Let $W_u = Z_u + \hat{X}(u) \in T_u F(M)$, and let $W_u = W^{\parallel}_u + W^{\perp}_u \in \hat{\mathscr{I}}_g(u) \oplus \hat{\mathscr{I}}^{\perp}_g(u)$ denote the splitting of W_u according to the splitting of $T_u F(M)$.

Let $Y \in \mathscr{I}_g(M)$ be the unique Killing vector field on M such that $\hat{Y}(u) = -W^{\parallel}_u$. Such a Y exists by the definition of $\hat{\mathscr{I}}_g(u)$, and it is unique since a Killing vector field is determined by its value and the values of its derivative at a single point. Then

$$\begin{aligned} h + Z_u &= \mathring{h} + L_X g - \hat{X}(u) + (Z_u + \hat{X}(u)) \\ &= \mathring{h} + W^{\perp}_u + W^{\parallel}_u + L_X g - \hat{X}(u) \\ &= (\mathring{h} + W^{\perp}_u) + L_X g - \hat{X}(u) - \hat{Y}(u) \end{aligned}$$

$$= \mathring{h} + W_u^\perp + \alpha_{(g,u)}(X+Y),$$

where

$$\alpha_{(g,u)}(X+Y) = L_{(X+Y)}\, g - \hat{X}(u) - \hat{Y}(u)$$
$$= L_X\, g - \hat{X}(u) - \hat{Y}(u)$$

since $L_Y g = 0$. Thus $S_2(M) \oplus T_u F(M)$ is a sum of $\mathring{S}_2(g) \oplus \hat{\mathscr{I}}_g^\perp(u)$ and range $\alpha_{(g,u)}$, and thus from the above is a direct sum.

Now we consider the equivariance of each summand. First, if $h \in \mathring{S}_2(f^*g) = \{h \in S_2(M) \,|\, \delta_{f^*g} h = 0\}$, then $(f^{-1})^*(\delta_{f^*g} h) = \delta_g((f^{-1})^* h) = 0$ so $(f^{-1})^* h \in \mathring{S}_2(g)$ and $h \in f^*(\mathring{S}_2(g))$. Hence $\mathring{S}_2(f^*g) \subseteq f^*(\mathring{S}_2(g))$. Since this inclusion is true for all f, $\mathring{S}_2(g) = \mathring{S}_2((f^{-1})^*(f^*g)) \subseteq (f^{-1})^* \mathring{S}_2(f^*g)$, which gives the reverse inclusion $f^*(\mathring{S}_2(g)) \subseteq \mathring{S}_2(f^*g)$ and hence equality.

For the second summand, recall from Lemma 2.1 (with f^{-1} replacing f) that $\hat{\mathscr{I}}_{f^*g}(f^*u) = T_u \hat{f}^{-1} \cdot (\hat{\mathscr{I}}_g(u))$. Thus for $Z_1, Z_2 \in T_{f^*u} F(M)$,

$$(\hat{f}^*\hat{g})(f^*u) \cdot (Z_1, Z_2)$$
$$= \hat{g} \circ \hat{f}(\hat{f}^{-1}(u)) \cdot (T_{f^*u} \hat{f} \cdot Z_1, T_{f^*u} \hat{f} \cdot Z_2)$$
$$= \hat{g}(u) \cdot (T_{f^*u} \hat{f} \cdot Z_1, T_{f^*u} \hat{f} \cdot Z_2).$$

Thus

$$T_{f^*u} \hat{f}\colon T_{f^*u} F(M) \to T_u F(M)$$

is an isometry of the inner product spaces $(T_{f^*u} F(M), \hat{f}^*\hat{g}(f^*u))$ and $(T_u F(M), \hat{g}(u))$, and so maps orthogonal subspaces to orthogonal subspaces.

The equivariance of the third summand follows from

$$T_{(g,u)} f^*(S_2(M) \oplus T_u F(M))$$
$$= S_2(M) \oplus T_u \hat{f}^{-1} \cdot (T_u(F(M)))$$

and the equivariance of the first two summands, where

$$f^*\colon \mathscr{M} \times F(M) \to \mathscr{M} \times F(M), \quad (g,u) \mapsto (f^*g, f^*u)$$

is the diffeomorphism of $\mathscr{M} \times F(M)$ corresponding to $f \in \mathscr{D}$.

Alternately, by direct computation, for $f \in \mathscr{D}$ and $X \in \mathscr{X}(M)$,

$$\alpha_{(f^*g, f^*u)}((f^{-1})_* X)$$
$$= L_{(f^{-1})_* X} f^*g - (f^{-1})_* X(f^*u)$$
$$= f^*(L_X g) - T_u \hat{f}^{-1} \cdot \hat{X}(u)$$

(see the proof of Lemma 2.1)

$$= T_{(g,u)} f^*(L_X g - \hat{X}(u))$$
$$= T_{(g,u)} f^*(\alpha_{(g,u)}(X)).$$

Thus

$$\alpha_{(f^*g, f^*u)}((f^{-1})_* \mathscr{X}(M)) = T_{(g,u)} f^*(\alpha_{(g,u)}(\mathscr{X}(M))),$$

and since $(f^{-1})_* \mathscr{X}(M) = \mathscr{X}(M)$,

$$\text{range}\, \alpha_{(f^*g, f^*u)} = T_{(g,u)} f^*(\text{range}\, \alpha_{(g,u)}). \qquad \square$$

Remarks: (1) Note that the summands are direct sum complements but not L_2-orthogonal complements with respect to Λ. Indeed

$$\Lambda(g,u) \cdot (\mathring{h} + w_u^\perp, L_X g - \hat{X}(u)) = -\hat{g}(u)(w_u^\perp, \hat{X}(u)),$$

which is not zero in general. In fact the L_2-adjoint of $\alpha_{(g,u)}$ is

$$\alpha^*_{(g,u)}\colon S_2(M) \oplus T_u F(M) \to \mathscr{X}(M), \quad h + Z_u \mapsto 2\delta_g h,$$

so the L_2-orthogonal complement of range $\alpha_{(g,u)}$ is $\ker \alpha^*_{(g,u)} = \mathring{S}_2(g) \oplus \{0\}$. Thus the usual Fredholm alternative approach to splittings does not work in this case, so we must construct closed linear complements directly.

(2) In the canonical splitting of $h = \mathring{h} + L_X g$, X is determined only up to a Killing vector field. In the splitting of $S_2(M) \oplus T_u F(M)$, however, $W_u^\perp$ and the sum $X + Y$ are uniquely determined. Indeed, if $X_1 \in \mathscr{I}_g(M)$ and $X_{\text{new}} = X + X_1$ replaces X, then $(W_u)_{\text{new}} = Z_u + \hat{X}(u) + \hat{X}_1(u)$. But since $\hat{X}_1(u) \in \hat{\mathscr{I}}_g(u)$, $\hat{X}_1^\perp(u) = 0$, and so $(W_u)^\perp_{\text{new}} = (Z_u + \hat{X}(u))^\perp = w_u^\perp$. Similarly, Y_{new} is now chosen so that

$$\hat{Y}_{\text{new}}(u) = -(W_u)^\parallel_{\text{new}} = -(Z_u + \hat{X}(u))^\parallel - \hat{X}_1(u),$$

since $\hat{X}_1^\parallel(u) = \hat{X}_1(u)$. Thus $Y_{\text{new}} = Y - X_1$, so that $X_{\text{new}} + Y_{\text{new}} = (X + X_1) + Y - X_1 = X + Y$. In short, the ambiguity in the choice of X in the canonical splitting of h is reflected solely in an ambiguity in the choice of Y. The sum $X + Y$, however, is uniquely determined. $\square$

Using the above splitting, we can now show that the orbits are closed submanifolds of $\mathscr{M} \times F(M)$ (see Fig. 2).

Proposition 3.3: For $(g,u) \in \mathscr{M} \times F(M)$, the orbit

$$\mathscr{O}_{(g,u)} = \{(f^*g, f^*u) \,|\, f \in \mathscr{D}\}$$

through (g,u) is a closed submanifold of $\mathscr{M} \times F(M)$, with tangent space at (f^*g, f^*u) given by

$$T_{(f^*g, f^*u)} \mathscr{O}_{(g,u)} = \text{range}\, \alpha_{(f^*g, f^*u)}$$
$$= T_{(g,u)} f^* \cdot (\text{range}\, \alpha_{(g,u)}),$$

and where

$$\text{range}\, \alpha_{(g,u)}$$
$$= \{L_X g - \hat{X}(u) \in S_2(M) + T_u F(M) \,|\, X \in \mathscr{X}(M)\}.$$

The orbit map

$$\Phi_{(g,u)}\colon \mathscr{D} \to \mathscr{O}_{(g,u)} \subseteq \mathscr{M} \times F(M)$$

is a diffeomorphism onto $\mathscr{O}_{(g,u)}$.

Proof: For $s > n/2$, consider the action

$$\Phi^{s+1}\colon (\mathscr{M}^s \times F(M)) \times \mathscr{D}^{s+1} \to \mathscr{M}^s \times F(M),$$
$$((g,u), f) \mapsto (f^*g, f^*u).$$

By standard composition properties of Sobolev spaces, if $(g,u) \in \mathscr{M}^{s+k} \times F(M)$, $k \geqslant 0$, the orbit map

$$\Phi^{s+1}_{(g,u)}\colon \mathscr{D}^{s+1} \to \mathscr{M}^s \times F(M)$$

is of class C^k (see Fischer and Marsden[14]). Also $\Phi^{s+1}_{(g,u)}$ is injective. For $k \geqslant 1$, its derivative at the identity is

$$\alpha^{s+1}_{(g,u)} = T_e \Phi^{s+1}_{(g,u)}\colon \mathscr{X}^{s+1}(M) \to S_2^s(g) \oplus T_u F(M),$$
$$X \mapsto L_X g - \hat{X}(u).$$

Since $\ker \alpha^{s+1}_{(g,u)} = \{0\}$, and since $\alpha^{s+1}_{(g,u)}$ has a closed range with a closed direct sum complement $\mathring{S}_2^s(g) \oplus \hat{\mathscr{I}}_g(u)$, $\Phi^{s+1}_{(g,u)}$ is an injective immersion at the identity.

For $f \in \mathscr{D}^{s+1}$, let

$$f^*\colon \mathscr{M}^s \times F(M) \to \mathscr{M}^s \times F(M), \quad (g,u) \mapsto (f^*g, f^*u)$$

denote the diffeomorphism of $\mathscr{M}^s \times F(M)$ corresponding to $f \in \mathscr{D}^{s+1}$, and let

$$R_f\colon \mathscr{D}^{s+1}\to\mathscr{D}^{s+1},\quad h\mapsto h\circ f$$

denote right translation by f. Then

$$T_eR_f\colon T_e\mathscr{D}^{s+1}=\mathscr{X}^{s+1}(M)\to T_f\mathscr{D}^{s+1},\quad X\mapsto X\circ f,$$

so that

$$T_fR_{f^{-1}}(T_f\mathscr{D}^{s+1})=T_f\mathscr{D}^{s+1}\circ f^{-1}=\mathscr{X}^{s+1}(M).$$

Since Φ^{s+1} is an action, we have the identity

$$\Phi^{s+1}_{(g,u)}\circ R_f=f^*\circ\Phi^{s+1}_{(g,u)}$$

or

$$\Phi^{s+1}_{(g,u)}=f^*\circ\Phi^{s+1}_{(g,u)}\circ R_{f^{-1}},$$

and so

$$T_f\Phi^{s+1}_{(g,u)}=T_{(g,u)}f^*\circ T_e\Phi^{s+1}_{(g,u)}\circ T_fR_{f^{-1}}$$
$$=T_{(g,u)}f^*\circ\alpha^{s+1}_{(g,u)}\circ T_fR_{f^{-1}}.$$

Thus

$$\begin{aligned}\text{range }T_f\,\Phi^{s+1}_{(g,u)}&=T_f\Phi^{s+1}_{(g,u)}(T_f\mathscr{D}^{s+1})\\&=T_{(g,u)}f^*\circ\alpha^{s+1}_{(g,u)}(\mathscr{X}^{s+1}(M))\\&=T_{(g,u)}f^*(\text{range }\alpha^{s+1}_{(g,u)})=\text{range }\alpha^{s+1}_{(f^*g,f^*u)},\end{aligned}$$

so that

$$\mathring{S}^s_2(f^*g)\oplus\hat{\mathscr{I}}^{\perp}_{f^*g}(f^*u)=T_{(g,u)}f^*(\mathring{S}_2(g)\oplus\hat{\mathscr{I}}^{\perp}_g(u))$$

is a closed linear complement of range $T_f\Phi^{s+1}_{(g,u)}$ in $S_2(M)\oplus T_{f^*u}F(M)$. Thus $\Phi^{s+1}_{(g,u)}$ is an injective immersion and so the orbit is an immersed submanifold.

Moreover, since the action is proper, the orbit map $\Phi^{s+1}_{(g,u)}$ is a proper map. Since a proper map between metrizable spaces is closed, the orbit $\mathscr{O}^s_{(g,u)}=\Phi^{s+1}_{(g,u)}(\mathscr{D}^{s+1})$ is closed in $\mathscr{M}^s\times F(M)$, and since $\Phi^{s+1}_{(g,u)}$ is a continuous injection, $\Phi^{s+1}_{(g,u)}$ is a homeomorphism onto its image. By the implicit function theorem, $\mathscr{O}^s_{(g,u)}$ is then a closed C^k submanifold of $\mathscr{M}^s\times F(M)$ and $\Phi^{s+1}_{(g,u)}$ is a C^k diffeomorphism onto $\mathscr{O}^s_{(g,u)}$ (see Abraham and Marsden,[11] Example 1.6.F.a, which is easily generalized to Hilbert manifolds). Moreover, the tangent space of $\mathscr{O}^s_{(g,u)}$ at (f^*g,f^*u) is given by range $T_e\Phi^{s+1}_{(f^*g,f^*u)}=T_{(g,u)}f^*(\text{range }T_e\Phi^{s+1}_{(g,u)})$, the equality following by Theorem 3.2 That $\mathscr{O}_{(g,u)}$ is an ILH submanifold and that $\Phi_{(g,u)}$ is an ILH diffeomorphism onto $\mathscr{O}_{(g,u)}$ is now checked in a standard manner. □

We now construct local equivariant cross sections for the action Φ. For $g\in\mathscr{M}$, we let $S_g\subseteq\mathscr{M}$ denote Ebin's slice at g for the action of Ψ of $\mathscr{D}$ on $\mathscr{M}$, and for $(g,u)\in\mathscr{M}\times F(M)$, we let $C_u(g)\subseteq F(M)$ denote the local cross section at $u\in F(M)$ for the action of $I_g(M)$ on $F(M)$.

Theorem 3.4: Let

$$\Phi\colon(\mathscr{M}\times F(M))\times\mathscr{D}\to\mathscr{M}\times F(M),$$
$$((g,u),f)\mapsto(f^*u,f^*u)$$

be the action of pullback on $\mathscr{M}\times F(M)$. For $(g,u)\in\mathscr{M}\times F(M)$, there exists a submanifold

$$C_{(g,u)}=S_g\times C_u(g)\subseteq\mathscr{M}\times F(M)\quad\text{(see Fig. 2)}$$

containing (g,u) and which satisfies the following conditions.

(1) At (g,u),

$$T_{(g,u)}C_{(g,u)}=\mathring{S}_2(g)\oplus\hat{\mathscr{I}}^{\perp}_g(u),$$

so that $C_{(g,u)}$ is transversal to the orbit $\mathscr{O}_{(g,u)}$ through (g,u).

(2) $C_{(g,u)}$ is equivariant with respect to $\mathscr{D}$, i.e., if $f\in\mathscr{D}$, $C_{(f^*g,f^*u)}=f^*(C_{(g,u)})$.

(3) If $f\in D$ and $f^*(C_{(g,u)})\cap C_{(g,u)}\neq\varnothing$, then $f=\mathrm{id}_M$.

(4) The action Φ, restricted to $C_{(g,u)}$,

$$\Phi_{\restriction}\colon C_{(g,u)}\times\mathscr{D}\to\mathscr{M}\times F(M),$$
$$((g_1,u_1),f)\mapsto(f^*g_1,f^*u_1),$$

is a diffeomorphism onto an open invariant neighborhood $U\subseteq\mathscr{M}\times F(M)$ of the orbit $\mathscr{O}_{(g,u)}$.

Proof: From Ebin's Slice Theorem, S_g is a submanifold containing g. Thus $S_g\times C_u(g)$ is a submanifold of $\mathscr{M}\times F(M)$ containing (g,u).

(1) $$T_{(g,u)}(S_g\times C_u(g))=T_gS_g\oplus T_uC_u(g)=\mathring{S}_2(g)\oplus\hat{\mathscr{I}}^{\perp}_g(u).$$

(2) In Ebin's construction of the slice S_g, he exponentiates an open neighborhood of $\mathring{S}_2(g)$ using the exponential map of the $\mathscr{D}$-invariant L_2-Riemannian metric on $\mathscr{M}$. It is easily deduced from his construction [using the equivariance of $\mathring{S}_2(g)$ and of the exponential map] that the slice so constructed is equivariant with respect to $\mathscr{D}$, i.e., for $f\in\mathscr{D}$, $S_{f^*g}=f^*(S_g)$. This equivariance, together with the equivariance of the local cross sections $C_u(g)$ [see remark 2 following Proposition 2.2, taking f^{-1} instead of f] implies

$$C_{(f^*g,f^*u)}=S_{f^*g}\times C_{f^*u}(f^*g)=f^*(S_g)\times f^*(C_u(g))$$
$$=f^*(S_g\times C_u(g))=f^*(C_{(g,u)}).$$

(3) If for $f\in\mathscr{D}$, $(g_1,u_1)\in S_g\times C_u(g)$ and $(f^*g_1,f^*u_1)\in S_g\times C_u(g)$, then g_1, $f^*g_1\in S_g$ and u_1, $f^*u_1\in C_u(g)$. By property 2 of Ebin's Slice Theorem, $f^*S_g\cap S_g\neq\varnothing$ implies $f\in I_g(M)$. Thus $f\in I_g(M)$, and since $C_u(g)$ is a local cross section for the action of $I_g(M)$ on $F(M)$, $u_1,f^*u_1\in C_u(g)$ with $f\in I_g(M)$ implies $f=\mathrm{id}_M$.

(4) Since the action Φ is smooth and proper, Φ is smooth and proper, and hence a closed map. By property (3), Φ is an injection and hence is a homeomorphism onto its image.

The derivative of Φ at $((g_1,u_1),f)\in C_{(g,u)}\times\mathscr{D}$ is given by

$$T_{((g_1,u_1),f)}\Phi\colon$$
$$T_{(g_1,u_1)}C_{(g,u)}\times T_f\mathscr{D}\to S_2(M)\oplus T_{f^*u_1}F(M),$$
$$(h+Z_u,X_f)\mapsto T_{(g_1,u_1)}f^*\cdot(h+Z_u+\alpha_{(g_1,u_1)}(X)),$$

where $X=X_f\circ f^{-1}\in\mathscr{X}(M)$. Since $C_{(g,u)}$ is transversal to $\mathscr{O}_{(g,u)}$, by the openness of transversality, if $C_{(g,u)}$ is chosen small enough, $C_{(g,u)}$ will be transversal to all the orbits it meets. The transversality of $C_{(g,u)}$ together with the injectivity of $\alpha_{(g_1,u_1)}=T_e\Phi_{(g_1,u_1)}$ implies that $T_{((g_1,u_1),f)}\Phi$ is an isomorphism at every $((g_1,u_1),f)\in C_{(g,u)}\times\mathscr{D}$. From this it follows that Φ is a diffeomorphism onto an open invariant neighborhood of $\mathscr{O}_{(g,u)}$ (see Ebin,[3] pp. 32–34). □

Remarks: (1) Again we remark that the usual properties of a slice restrict to properties (3) and (4) when the action is free.

(2) Since

$$T_{(g,u)}S_{(g,u)} = \mathring{S}_2(g) \oplus \hat{\mathscr{I}}_g^{\perp}(u)$$

and

$$T_{(g,u)}\mathscr{O}_{(g,u)} = \text{range}\,\alpha_{(g,u)}$$

(Proposition 3.3), the splitting $S_2(M) = \mathring{S}_2(g) \oplus \hat{\mathscr{I}}_g^{\perp}(u) \oplus \text{range}\,\alpha_{(g,u)}$ can be written as

$$T_{(g,u)}(\mathscr{M}\times F(M)) = T_{(g,u)}C_{(g,u)} \oplus T_{(g,u)}\mathscr{O}_{(g,u)},$$

which gives a geometric interpretation of the first two summands as the tangent space to the local cross section $C_{(g,u)}$ at (g,u) and the third summand as the tangent space to the orbit $\mathscr{O}_{(g,u)}$ through (g,u). □

The existence of equivariant local cross sections now implies the following.

Theorem 3.5: For the action

$$\Phi\colon (\mathscr{M}\times F(M))\times\mathscr{D}\to\mathscr{M}\times F(M),$$

$$((g,u),f)\mapsto(f^*g,f^*u),$$

let

$$\mathscr{G}_{FM} = ((\mathscr{M}\times F(M))/\mathscr{D}$$

denote the orbit space, and let

$$\pi\colon \mathscr{M}\times F(M)\to\mathscr{G}_{FM}, \quad (g,u)\mapsto[(g,u)]$$

denote the orbit projection map. Then $\mathscr{G}_{FM}$ has a smooth manifold structure such that π is a submersion with

$$\ker T_{(g,u)}\pi = T_{(g,u)}\mathscr{O}_{(g,u)} = \text{range}\,\alpha_{(g,u)}$$

and

$$\text{range}\, T_{(g,u)}\pi = T_{[(g,u)]}\mathscr{G}_{FM} \approx \frac{S_2(M)\oplus T_uF(M)}{\text{range}\,\alpha_{(g,u)}} \approx \mathring{S}_2(g)\oplus\hat{\mathscr{I}}_g^{\perp}(u).$$

Moreover, $\pi\colon\mathscr{M}\times F(M)\to\mathscr{G}_{FM}$ has the structure of a principal fiber bundle with total space $\mathscr{M}\times F(M)$, base space $\mathscr{G}_{FM}$, and structure group $\mathscr{D}$. This principal fiber bundle has a natural connection given by the direct sum decomposition

$$S_2(M)\oplus T_uF(M) = H_{(g,u)}\oplus V_{(g,u)},$$

where

$$V_{(g,u)} = \ker T_{(g,u)}\pi = \text{range}\,\alpha_{(g,u)} = T_{(g,u)}\mathscr{O}_{(g,u)}$$

is the vertical subspace at (g,u) and

$$H_{(g,u)} = \mathring{S}_2(g)\oplus\hat{\mathscr{I}}_g^{\perp}(u)$$

is the horizontal subspace at (g,u).

Proof: The proof proceeds as in the finite-dimensional case (see, e.g., Abraham and Marsden,[11] p. 262). The main idea is to use the local equivariant cross sections $C_{(g,u)}$ as charts for the orbit space $\mathscr{G}_{FM} = (\mathscr{M}\times F(M))/\mathscr{D}$. Thus a chart for $\mathscr{G}_{FM}$ at $[(g,u)]$ is constructed as follows. For $(g,u)\in\mathscr{M}\times F(M)$, let

$$\pi_{\restriction}\colon C_{(g,u)}\to\mathscr{G}_{FM}$$

denote the restriction of $\pi\colon\mathscr{M}\times F(M)\to\mathscr{G}_{FM}$ to the local cross section $C_{(g,u)}$ at (g,u). Then $\pi_{\restriction}$ is a bijection onto its image. Since $C_{(g,u)}$ is a submanifold, there exists a chart (U,φ) at (g,u), $U\subseteq\mathscr{M}\times F(M)$, that has the submanifold property, i.e.,

$$\varphi\colon U\to E\times F,$$

$$\varphi(U\cap C_{(g,u)}) = \varphi(U)\cap(E\times\{0\})\subseteq E\times\{0\}.$$

Let

$$\pi_{\restriction}^{-1}\colon \pi(U\cap C_{(g,u)})\to U\cap C_{(g,u)}$$

denote the inverse of $\pi_{\restriction}$ restricted to $U\cap C_{(g,u)}$. If $[(g_1,u_1)]\in\pi(U\cap C_{(g,u)})$, then

$$\pi^{-1}([(g_1,u_1)]) = C_{(g,u)}\cap\mathscr{O}_{(g_1,u_1)} = \{(g_2,u_2)\}\subseteq U\cap C_{(g,u)}.$$

Thus

$$\varphi\circ\pi_{\restriction}^{-1}\colon \pi(U\cap C_{(g,u)})\to E\times\{0\},$$

$$[(g_1,u_1)]\mapsto\varphi(g_2,u_2)$$

is a bijection onto its image. Now take $(\pi(U\cap C_{(g,u)}), \varphi\circ\pi_{\restriction}^{-1})$ as a chart for $\mathscr{G}_{FM}$ at $[(g,u)]$. From the smoothness of the action and the equivariance of the local cross sections, a routine check then shows that the overlap maps between any two charts of $[(g,u)]$ are ILH smooth, and thus $\mathscr{G}_{FM}$ is a smooth ILH manifold.

Now let (U,φ) be a chart at $(g,u)\in\mathscr{M}\times F(M)$ and let $(\pi(U\cap C_{(g,u)}), \varphi\circ\pi^{-1})$ be a chart at $\pi(g,u) = [(g,u)]\in\mathscr{G}_{FM}$. Shrink U if necessary so that $\pi(U) = \pi(U\cap C_{(g,u)})$. In these charts

$$(\varphi\circ\pi_{\restriction}^{-1})\circ\pi\circ\varphi^{-1}\colon \varphi(U)\subseteq E\times F\to E\times\{0\},$$

$$(e,f)\mapsto(e,0)$$

is a smooth submersion, which implies that $\pi\colon \mathscr{M}\times F(M)\to\mathscr{G}_{FM}$ is a smooth submersion. As any submersion admits smooth local cross-sections, π is a $\mathscr{D}$-PFB over $\mathscr{G}_{FM}$.

That the assignment $(g,u)\mapsto H_{(g,u)} = \mathring{S}_2(g)\oplus\hat{\mathscr{I}}_g^{\perp}(u)$ defines a connection on the PFB $\pi\colon\mathscr{M}\times F(M)\to\mathscr{G}_{FM}$ follows from the equivariance of the splitting

$$S_2(M)\oplus T_uF(M) = H_{(g,u)}\oplus V_{(g,u)}$$

(Theorem 3.2), since if $f\in\mathscr{D}$

$$H_{(f^*g,f^*u)} = \mathring{S}_2(f^*g)\oplus\hat{\mathscr{I}}_{f^*g}^{\perp}(f^*u) = T_{(g,u)}f^*(\mathring{S}_2(g)\oplus\hat{\mathscr{I}}_g^{\perp}(u)).$$

That $H_{(g,u)}$ depends smoothly on (g,u) follows by arguments as in Ebin and Marsden,[8] Appendix A, and Fischer *et al.*[15] □

We have now constructed a manifold $\mathscr{G}_{FM}$ which is the base space of a $\mathscr{D}$-PFB $\pi\colon\mathscr{M}\times F(M)\to\mathscr{G}_{FM}$. In turn, $\mathscr{G}_{FM}$ covers the space $\mathscr{G} = \mathscr{M}/\mathscr{D}$ of Riemannian geometries by the projection

$$\pi_1\colon (\mathscr{M}\times F(M))/\mathscr{D}\to\mathscr{M}/\mathscr{D}, \quad [(g,u)]\mapsto[g]$$

so that we have the sequence of mappings

$$\mathscr{M}\times F(M)\xrightarrow{\pi}(\mathscr{M}\times F(M))/\mathscr{D} = \mathscr{G}_{FM}\xrightarrow{\pi_1}\mathscr{M}/\mathscr{D} = \mathscr{G},$$

$$(g,u)\mapsto[(g,u)]\mapsto[g].$$

This sequence is analogous to the sequence of maps

$$P\times F\xrightarrow{\pi}(P\times F)/G = E\xrightarrow{\pi_E}P/G = M$$

used in the construction of a fiber bundle

$$E = E(M,F,G,P) = (P\times F)/G$$

over M, with standard fiber F, and which is associated with a G-PFB $\pi_P: P\to M$. In this construction, the first map $\pi: P\times F\to E$ is also a G-PFB with total space $P\times F$ and base space E, and the second map $\pi_E: E\to M$ is the fiber bundle associated with $P\to M$. The $\mathscr{D}$-PFB $\pi: \mathscr{M}\times F(M)\to \mathscr{G}_{FM}$ is analogous to this first map, and the covering $\pi_1: \mathscr{G}_{FM}\to\mathscr{G}$ is analogous to the second map, with $\mathscr{D}$ playing the role of G, $\mathscr{M}$ the role of P, and $F(M)$ the role of F. Thus if $\mathscr{M}\to\mathscr{M}/\mathscr{D}$ *were* a $\mathscr{D}$-PFB, then $\pi_1: \mathscr{G}_{FM}\to\mathscr{G}$ would be the associated fiber bundle with standard fiber $F(M)$.

Interestingly, even though $\mathscr{M}\to\mathscr{M}/\mathscr{D}$ is not in general a PFB, so that the second map of the construction $\pi_1: \mathscr{G}_{FM}\to\mathscr{G}$ is not a fiber bundle, the first step of the construction $\pi: \mathscr{M}\times F(M)\to\mathscr{G}_{FM}$ is a $\mathscr{D}$-PFB, completely analogous to the G-PFB $\pi: P\times F\to E$.

We will return to this analogy in Sec. VI (see also the remark following Proposition 3.6 below).

Since $\pi_1: \mathscr{G}_{FM}\to\mathscr{G}$ is not a fiber bundle, the fibers $\pi_1^{-1}([g])\subseteq\mathscr{G}_{FM}$ are not diffeomorphic to each other. Indeed, if we think of $\mathscr{G}$ as a singular manifold covered by the manifold $\mathscr{G}_{FM}$, then the deviation of the fibers $\pi_1^{-1}([g])$ from the standard fiber $F(M)$ is a measure of the degree of these singularities. We now compute these fibers.

Proposition 3.6: Let

$$\mathscr{M}\times F(M)\overset{\pi}{\to}\mathscr{G}_{FM}\overset{\pi_1}{\to}\mathscr{G},\quad (g,u)\mapsto[(g,u)]\mapsto[g]$$

be the sequence of mappings described above. Then for $[g_0]\in\mathscr{G}$, the fiber $\pi_1^{-1}([g_0])$ is a closed submanifold of $\mathscr{G}_{FM}$ and is given by

$$\pi_1^{-1}([g_0])=\pi(\mathscr{O}_{g_0}\times F(M))$$
$$=(\mathscr{O}_{g_0}\times F(M))/\mathscr{D}\approx I_{g_0}(M)/F(M),$$

where $\mathscr{O}_{g_0}=\{f^*g_0|f\in\mathscr{D}\}\subseteq\mathscr{M}$ is the $\mathscr{D}$-orbit of $g_0\in\mathscr{M}$, and where

$$\hat{\pi}_{g_0}: I_{g_0}(M)\backslash F(M)\to\pi_1^{-1}([g_0]),\quad [u]\mapsto[(g_0,u)]$$

is a diffeomorphism. At $[(g_0,u)]\in\pi_1^{-1}([g_0])$, the tangent space to the fiber $\pi_1^{-1}([g_0])$ is

$$T_{[(g_0,u)]}(\pi_1^{-1}([g_0]))\approx\frac{\mathring{S}_2(g_0)\oplus\hat{\mathscr{I}}^{\perp}_{g_0}(u)}{\mathring{S}_2(g_0)}\approx\hat{\mathscr{I}}^{\perp}_{g_0}(u).$$

Proof: $\pi_1^{-1}([g_0])=\{[(g,u)]\in\mathscr{G}_{FM}|[g]=[g_0]$ and $u\in F(M)\}$, so that $(g_1,u_1)\in[(g,u)]\in\pi_1^{-1}([g_0])$ iff $[g_1]=[g_0]$ iff $g_1=f^*g_0$ for some $f\in\mathscr{D}$. Thus

$$\pi_2^{-1}([g_0])$$
$$=\{[(f^*g_0,u)]\in\mathscr{G}_{FM}|f\in\mathscr{D}\text{ and }u\in F(M)\}$$
$$=\pi(\mathscr{O}_{g_0}\times F(M)).$$

Alternatively, $\pi_1\circ\pi(g_0,u)=[g_0]$, and so

$$(\pi_1\circ\pi)^{-1}([g_0])=\pi^{-1}\circ\pi_1^{-1}([g_0])]=\mathscr{O}_{g_0}\times F(M).$$

Thus $\pi_1^{-1}([g_0])=\pi(\mathscr{O}_{g_0}\times F(M))$.

Since $\mathscr{O}_{g_0}\times F(M)$ is a closed $\mathscr{D}$-invariant submanifold of $\mathscr{M}\times F(M)$, its orbit space

$$(\mathscr{O}_{g_0}\times F(M))/\mathscr{D}=\pi(\mathscr{O}_{g_0}\times F(M))=\pi_2^{-1}([g_0])$$

is a closed submanifold of $\mathscr{G}_{FM}$.

If $f\in\mathscr{D}$ and $u\in F(M)$, $[(f^*g_0,u)]=[(g_0,(f^{-1})^*u)]$, and so

$$\pi_2^{-1}([g_0])=\{[(g_0,u)]\in\mathscr{G}_{FM}|u\in F(M)\}$$
$$=\pi(\{g_0\}\times F(M)).$$

For $g_0\in[g_0]$, let

$$\pi_{g_0}: F(M)\to\mathscr{G}_{FM},\quad u\mapsto\pi(g_0,u)=[(g_0,u)],$$

so that π_{g_0} is surjective onto the fiber $\pi_2^{-1}([g_0])$. If $\pi_{g_0}(u_1)=\pi_{g_0}(u_2)$, then $[(g_0,u_1)]=[(g_0,u_2)]$, and so there exists an $f\in\mathscr{D}$ such that $f^*g_0=g_0$ and $f^*u_1=u_2$, so that $f\in I_{g_0}(M)$. Conversely, if $f\in I_{g_0}(M)$, $\pi_{g_0}(f^*u)=\pi_{g_0}(u)$ so that π_{g_0} is invariant by the action of $I_{g_0}(M)$ on $F(M)$, and hence passes to the quotient manifold $I_{g_0}(M)\backslash F(M)$, where, by the previous statement, it is injective. Hence

$$\tilde{\pi}_{g_0}: I_{g_0}(M)\backslash F(M)\to\pi_2^{-1}([g_0])\subseteq\mathscr{G}_{FM},$$
$$[u]\mapsto[(g_0,u)],$$

is a bijection.

Now $\pi_{g_0}: F(M)\to\mathscr{G}_{FM}$ is a smooth map with derivative at $u\in F(M)$ given by

$$T_u\pi_{g_0}: T_uF(M)\to T_{[(g_0,u)]}\mathscr{G}_{FM}\approx\mathring{S}_2(g_0)\oplus\hat{\mathscr{I}}^{\perp}_{g_0}(u),$$
$$Z_u\mapsto Z_u^{\perp}.$$

Similarly, $\tilde{\pi}_{g_0}$ is smooth with derivative at $[u]\in I_{g_0}(M)\backslash F(M)$ given by

$$T_u\tilde{\pi}_{g_0}: T_uF(M)/\hat{\mathscr{I}}_{g_0}(u)\approx\hat{\mathscr{I}}^{\perp}_{g_0}(u)\to\mathring{S}_2(g_0)\oplus\hat{\mathscr{I}}^{\perp}_{g_0}(u),$$
$$Z_u^{\perp}\mapsto Z_u^{\perp}.$$

Hence $\tilde{\pi}_{g_0}$ is an injective immersion onto the closed submanifold $\pi_1^{-1}([g_0])$, and is thus a diffeomorphism of $I_{g_0}(M)\backslash F(M)$ and $\pi_1^{-1}([g_0])$. Also, the tangent space of $\pi_1^{-1}([g_0])$ at $[(u,g_0)]$ is then given by range $T_u\tilde{\pi}_{g_0}\approx\hat{\mathscr{I}}^{\perp}_{g_0}(u)$. □

Remarks: The above proposition continues our previous analogy with the construction of the associated fiber bundle $E=(P\times F)/G$ and the sequence of maps

$$P\times F\overset{\pi}{\to}E\overset{\pi_E}{\to}M.$$

Thus if $x\in M$ and $p_0\in\pi_p^{-1}(x)$, then

$$\pi_E^{-1}(x)=\{[(p,f)]\in E\,|\pi_p(p)=x\text{ and }f\in F\}$$
$$=\pi(\mathscr{O}_{p_0}\times F)$$
$$=(\mathscr{O}_{p_0}\times F)/G=\{[(p_0,f)]\in E\,|f\in F\}$$
$$=\pi(\{p_0\}\times F)$$

where $\mathscr{O}_{p_0}=p_0\cdot G\subseteq P$ is the G-orbit through p_0. The fibers $\pi_1^{-1}([g_0])$ of Proposition 3.6 are given by the same formulas.

Moreover, each $p_0\in P$ induces a diffeomorphism of the standard fiber F onto the fiber $\pi_E^{-1}(x_0)$

$$\pi_{p_0}: F\to\pi_E^{-1}(x_0)\subseteq E,\quad f\mapsto\pi(p_0,f)=[(p_0,f)],$$

where $x_0=\pi_p(p_0)$. The map π_{g_0} of the proposition plays the role of the map π_{p_0} in the standard construction. However, since $\mathscr{M}\to\mathscr{M}/\mathscr{D}$ is not a *bona fide* PFB, π_{g_0} is only injective when we pass to the quotient manifold $I_{g_0}(M)\backslash F(M)$, and thus the fibers $\pi_1^{-1}([g_0])$ are not in general diffeomorphic to the "standard fiber" $F(M)$. □

In summary, $\pi_1: \mathscr{G}_{FM} \to \mathscr{G}$ is an infinite-dimensional manifold that covers $\mathscr{G}$, and whose "fiber" $\pi_1^{-1}([g])$ at the geometry $[g] \in \mathscr{G}$ is diffeomorphic to the finite-dimensional manifold $I_g(M) \backslash F(M)$. Thus $\mathscr{G}_{FM}$ unfolds the singularities of $\mathscr{G}$, and the degree of this unfolding at $[g]$ (or the degree of the singularity of $\mathscr{G}$ at $[g]$) is measured by the deviation of the fiber $\pi_1^{-1}([g])$ from the "standard fiber" $F(M)$. Thus as expected, $I_g(M)$ parametrizes the degree of the singularity of $\mathscr{G}$ at $[g]$.

IV. THE FRAME BUNDLE AS A HOMOGENEOUS MANIFOLD

In this section, we construct a principal fiber bundle

$$\pi: \mathscr{D} \to F(M)$$

with total space $\mathscr{D}$, structure group $\mathscr{D}'_{x_0}$, and base space $F(M)$ [if M is nonreversible, replace $F(M)$ with $F^+_{u_0}(M)$, where u_0 is a frame at x_0]. We also consider an interesting double coset manifold

$$I_{g_0(M)} \backslash \mathscr{D}/\mathscr{D}'_{x_0} = \{I_{g_0}(M) \circ f \circ \mathscr{D}'_{x_0} \mid f \in \mathscr{D}\}.$$

We shall use these results in Secs. V and VI.

For $x_0 \in M$, let

$$\mathscr{X}'_{x_0}(M) = \{X \in \mathscr{X}(M) \mid X(x_0) = 0 \text{ and } TX(x_0) = 0\}.$$

Then $\mathscr{X}'_{x_0}$ is a closed subspace of $\mathscr{X}(M)$. A finite-dimensional complement can be constructed as follows.

Let $\{Y_i, Y_\alpha\} \subseteq \mathscr{X}(M)$, $1 \leqslant i \leqslant n$, $1 \leqslant \alpha \leqslant n^2$, be $n^2 + n$ vector fields on M such that $\{Y_i(x_0)\}$, $1 \leqslant i \leqslant n$, span $T_{x_0}M$, and such that for $1 \leqslant \alpha \leqslant n^2$, $\{Y_\alpha(x_0) = 0\}$ and $\{TY_\alpha(x_0)\}$ span $T_0(TM)$. Note that

$$TY_\alpha(x_0): T_{x_0}M \to T_{Y_\alpha(x_0)}(TM) = T_0(TM),$$

so that it makes sense to require that the set $\{TY_\alpha(x_0)\}$ span $T_0(TM)$.

Let $\mathscr{Y}(M)$ denote the $(n^2 + n)$-dimensional subspace of $\mathscr{X}(M)$ spanned by $\{Y_i, Y_\alpha\}$. Since $\mathscr{Y}(M)$ is finite dimensional, it is closed in $\mathscr{X}(M)$. Moreover, we have the following.

Proposition 4.1: The subspaces $\mathscr{Y}(M)$ and $\mathscr{X}'_{x_0}(M)$ are closed complementary subspaces of $\mathscr{X}(M)$, and so $\mathscr{X}(M)$ has the direct sum decomposition

$$\mathscr{X}(M) = \mathscr{X}'_{x_0}(M) \oplus \mathscr{Y}(M), \quad X = X' + Y.$$

Proof: Since $\mathscr{Y}(M)$ and $\mathscr{X}'_{x_0}(M)$ are closed in $\mathscr{X}(M)$, it is sufficient to show that the above direct sum is true in the algebraic sense.

Since $\{Y_i(x), TY_\alpha(x)\}$ is a basis for $T_xM \oplus T_0(TM)$, for each $X \in \mathscr{X}(M)$, there is a unique solution $\{\lambda^i_0, \lambda^\alpha_0\}$ to the $n^2 + n$ equations

$$\lambda^i Y_i(x_0) = X(x_0),$$

$$\lambda^i TY_i(x_0) + \lambda^\alpha TY_\alpha(x_0) = TX(x_0).$$

Let $Y = \lambda^i_0 Y_i + \lambda^\alpha_0 Y_\alpha \in \mathscr{X}(M)$. Then $X' = X - Y \in \mathscr{X}'_{x_0}(M)$ and $X = X' + Y$. Thus $\mathscr{X}(M) = \mathscr{X}'_{x_0}(M) + \mathscr{Y}(M)$.

Conversely, since $\{Y_i, Y^\alpha\}$ is a basis for $\mathscr{Y}(M)$, if $X \in \mathscr{X}'_{x_0}(M) \cap \mathscr{Y}(M)$, then $X = \lambda^i_0 Y_i + \lambda^\alpha_0 Y_\alpha$ for a unique set of coefficients $\{\lambda^i_0, \lambda^\alpha_0\}$. But since $X \in \mathscr{X}'_{x_0}(M)$, these coefficients all vanish, and so $X = 0$. Thus the above sum is an algebraic direct sum, and hence a topological one also. □

Remarks: (1) Let $\mathscr{X}(M)/\mathscr{X}'_{x_0}(M)$ denote the quotient space of $\mathscr{X}(M)$ by the closed subspace $\mathscr{X}'_{x_0}(M)$. Then

$$\mathscr{Y}(M) \to \mathscr{X}(M)/\mathscr{X}'_{x_0}(M), \quad Y \mapsto Y + \mathscr{X}'_{x_0}(M)$$

is an isomorphism of $\mathscr{Y}(M)$ and $\mathscr{X}(M)/\mathscr{X}'_{x_0}(M)$.

(2) The space $\mathscr{X}'_{x_0}(M)$ is invariant by $\mathscr{D}'_{x_0}$, i.e., if $f \in \mathscr{D}'_{x_0}$, then $f_*(\mathscr{X}'_{x_0}(M)) = \mathscr{X}'_{x_0}(M)$. However, $f_*(\mathscr{Y}(M))$ is not in general equal to $\mathscr{Y}(M)$, although $f_*(\mathscr{Y}(M))$ will be another closed complement of $\mathscr{X}'_{x_0}(M)$. □

Theorem 4.2: Let $x_0 \in M$. Then $\mathscr{D}'_{x_0}$ is a closed ILH Lie subgroup of $\mathscr{D}$ with Lie algebra $\mathscr{X}'_{x_0}(M)$. The quotient space $\mathscr{D}/\mathscr{D}'_{x_0}$ has a smooth manifold structure such that the projection

$$\pi: \mathscr{D} \to \mathscr{D}/\mathscr{D}'_{x_0}, \quad f \mapsto f \circ \mathscr{D}'_{x_0}$$

is a submersion with

$$\ker T_f\pi = \mathscr{X}'_{x_0}(M) \circ f$$

and

$$\text{range } T_f\pi = T_{f \circ \mathscr{D}'_{x_0}}(\mathscr{D}/\mathscr{D}'_{x_0}) \approx \mathscr{X}(M)/\mathscr{X}'_{x_0}(M).$$

The submersion π has the structure of a principal fiber bundle with total space $\mathscr{D}$, base space $\mathscr{D}/\mathscr{D}'_{x_0}$ and structure group $\mathscr{D}'_{x_0}$.

If $u_0 \in F(M)$ is a frame at $x_0 = \pi_{FM}(u_0)$ and M is reversible or nonorientable, then

$$\tilde{\psi}_{u_0}: \mathscr{D}/\mathscr{D}'_{x_0} \to F(M), \quad f \circ \mathscr{D}'_{x_0} \mapsto \hat{f}(u_0)$$

is a diffeomorphism. If M is nonreversible then $\tilde{\psi}_{u_0}$ is a diffeomorphism onto $F^+_{u_0}(M) \subseteq F(M)$.

If $\mathscr{Y}(M) \subseteq \mathscr{X}(M)$ is a closed complement of $\mathscr{X}'_{x_0}(M)$, then a smooth connection on the bundle $\pi: \mathscr{D} \to \mathscr{D}/\mathscr{D}'_{x_0}$ is given by the direct sum decomposition

$$T_f\mathscr{D} = \mathscr{X}(M) \circ f = \mathscr{X}'_{x_0}(M) \circ f \oplus \mathscr{Y}(M) \circ f$$

$$= V_f \oplus H_f,$$

where $V_f = \ker T_f\pi = \mathscr{X}'_{x_0}(M) \circ f$ is the vertical subspace at f and $H_f = \mathscr{Y}(M) \circ f \approx \mathscr{X}(M)/\mathscr{X}'_{x_0}(M)$ is the horizontal subspace at f.

Proof: For $s > n/2 + 1$ and $u_0 \in F(M)$, let

$$\psi^s_{u_0}: \mathscr{D}^s \to F(M), \quad f \mapsto \hat{f}(u_0)$$

denote the orbit map at u_0. Then $\psi^s_{u_0}$ is smooth with derivative at $f \in \mathscr{D}^s$ given by

$$T_f\psi^s_{u_0}: T_f\mathscr{D}^s \to T_{\hat{f}(u_0)}(F(M)), \quad X_f \mapsto \hat{X}(\hat{f}(u_0)),$$

where $X = X_f \circ f^{-1} \in \mathscr{X}(M)$. Since $T_f\psi^s_{u_0}$ is clearly surjective, $\psi^s_{u_0}$ is a smooth submersion and so by the implicit function theorem

$$(\psi^s_u)^{-1}(u_0) = \mathscr{D}^s_{u_0} = (\mathscr{D}^s)'_{x_0}$$

$$= \{f \in \mathscr{D}^s \mid f(x_0) = x_0 \text{ and } T_{x_0}f = I_{x_0}\}$$

is a smooth closed submanifold of $\mathscr{D}^s$ with tangent space at $f \in \mathscr{D}^s_{u_0}$ given by

$$\ker T_f\psi^s_u = \mathscr{X}'_{x_0}(M)^s \circ f.$$

It is also a subgroup, and so it is a topological group whose group operations have the same smoothness properties as those of $\mathscr{D}^s$. Therefore

$$\mathscr{D}'_{x_0} = \bigcap_{s>n/2+1} (\mathscr{D}^s)'_{x_0}$$

is a closed ILH Lie subgroup of D with Lie algebra

$$\mathscr{X}(M) = \bigcap_{s>n/2+1} \mathscr{X}'_x(M)^s.$$

Now $\mathscr{D}^s_{u_0}$ acts on $\mathscr{D}^s$ on the right by composition

$$\mathscr{D}^s \times \mathscr{D}^s_{u_0} \to \mathscr{D}^s, \quad (f,h) \mapsto f \circ h,$$

and this action is a free C^0 action. The orbit map $\psi^s_{u_0}$: $\mathscr{D}^s \to F(M)$ is invariant by this action, since if $f \in \mathscr{D}^s$ and $h \in \mathscr{D}^s_{u'}$

$$\psi^s_{u_0}(f\circ h) = \widehat{f\circ h}(u_0) = \hat{f}\circ\hat{h}(u_0) = \hat{f}(u_0) = \psi^s_{u_0}(f),$$

and so passes to the quotient space

$$\tilde{\psi}^s_{u_0}: \mathscr{D}^s/\mathscr{D}^s_{u_0} \to F(M), \quad f\circ\mathscr{D}^s_{u_0} \mapsto \hat{f}(u_0),$$

where it is an injective map. Since $\psi^s_{u_0}$ is a submersion, $\tilde{\psi}^s_{u_0}$ is an open map, and hence a homeomorphism onto its image. If M is reversible or nonorientable, this image is all of $F(M)$. If M is nonreversible, this image is $F^+_{u_0}(M)$. In either case we induce a smooth structure on $\mathscr{D}^s/\mathscr{D}^s_{u_0}$ by declaring $\tilde{\psi}_{u_0}$ to be a diffeomorphism onto $F(M)$ [or $F^+_{u_0}(M)$]. With this differential structure, π^s: $\mathscr{D}^s \to \mathscr{D}^s/\mathscr{D}^s_{u_0}$ is a smooth submersion because $\psi^s_{u_0}$: $\mathscr{D}^s_u \to F(M)$ is. Since any submersion admits smooth local sections, π^s becomes a PFB with structure group $\mathscr{D}^s_{u_0}$. The bundle structure is only C^0 because $\mathscr{D}^s_{u_0}$ acts only continuously on $\mathscr{D}^s$. However, it does give a smooth ILH bundle structure to π: $\mathscr{D} \to \mathscr{D}/\mathscr{D}_{u_0}$.

If $\mathscr{Y}(M) \subseteq \mathscr{X}(M)$ is a closed complement of $\mathscr{X}'_{x_0}(M)$, then the direct sum decomposition

$$\mathscr{X}(M) = \mathscr{X}'_{x_0}(M) \oplus \mathscr{Y}(M)$$

can be right translated to $T_f\mathscr{D} = \mathscr{X}(M)\circ f$ by

$$T_eR_f: T_e\mathscr{D} = \mathscr{X}(M) \to T_f\mathscr{D}, \quad X \mapsto X\circ f.$$

Thus

$$T_f\mathscr{D} = T_eR_f\cdot\mathscr{X}(M) = T_eR_f\cdot\mathscr{X}'_{x_0}(M) \oplus T_eR_f\cdot\mathscr{Y}(M),$$

or

$$\mathscr{X}(M)\circ f = (\mathscr{X}'_{x_0}(M)\circ f) \oplus (\mathscr{Y}(M)\circ f),$$

is a right equivariant direct sum decomposition of $T_f\mathscr{D}$ by closed complementary subspaces. In particular, if $h \in \mathscr{D}$, then

$$T_fR_h(\mathscr{Y}(M)\circ f) = \mathscr{Y}(M)\circ f\circ h,$$

so that the horizontal spaces $H_f = \mathscr{Y}(M)\circ f = T_eR_f\cdot\mathscr{Y}(M)$ satisfy

$$T_fR_h\cdot H_f = H_{f\circ h}.$$

Moreover, from the smoothness of the group operations in $\mathscr{D}$, and by using arguments as in Ebin and Marsden,[8] Appendix A, it follows that the distribution $f \mapsto H_f = \mathscr{Y}(M)\circ f$ is smooth in f, thereby defining a smooth connection on the $\mathscr{D}'_{x_0}$-PEB π: $\mathscr{D} \to \mathscr{D}/\mathscr{D}'_{x_0}$. □

Remark: Thus choosing a direct sum complement $\mathscr{Y}(M) \subseteq \mathscr{X}(M)$ to $\mathscr{X}'_{x_0}(M)$ is equivalent to giving a connection on the PFB π: $\mathscr{D} \to \mathscr{D}/\mathscr{D}'_{x_0}$. □

For $g_0 \in \mathscr{M}$, the isometry group $I_{g_0}(M)$ acts on $\mathscr{D}$ on the left,

$$I_{g_0}(M) \times \mathscr{D} \to \mathscr{D}, \quad (k,f) \mapsto k\circ f.$$

Since $I_{g_0}(M)$ is compact and acts freely, the quotient space

$$I_{g_0}(M)\backslash\mathscr{D} = \{I_{g_0}(M)\circ f \mid f \in \mathscr{D}\}$$

is a manifold (see Ebin,[3] p. 22). Also, $I_{g_0}(M)$ acts naturally on the left on the quotient manifold, $\mathscr{D}/\mathscr{D}'_{x_0}$,

$$I_{g_0}(M) \times (\mathscr{D}/\mathscr{D}'_{x_0}) \to \mathscr{D}/\mathscr{D}'_{x_0},$$
$$(k, f\circ\mathscr{D}'_{x_0}) \mapsto k\circ f\circ\mathscr{D}'_{x_0}.$$

This action is free, for if $k\circ f\circ\mathscr{D}'_{x_0} = f\circ\mathscr{D}'_{x_0}$, then $f^{-1}\circ k\circ f \in \mathscr{D}'_{x_0}$, and since $(f^{-1}\circ k\circ f)^*(f^*g_0) = f^*g_0$, $(f^{-1}\circ k\circ f) \in I_{f^*g_0}(M) \cap \mathscr{D}'_{x_0}$. Hence $f^{-1}\circ k\circ f = \mathrm{id}_M$ and so $k = \mathrm{id}_M$. Since $I_{g_0}(M)$ is compact, the resulting quotient space

$$I_{g_0}(M)\backslash(\mathscr{D}/\mathscr{D}'_{x_0}) = \{I_{g_0}(M)\circ(f\circ\mathscr{D}'_{x_0}) \mid f \in \mathscr{D}\}$$

is a manifold, since $\mathscr{D}/\mathscr{D}'_{x_0}$ is a finite-dimensional manifold.

Similarly, $\mathscr{D}'_{x_0}$ acts naturally on the right on the quotient manifold $I_{g_0}(M)\backslash\mathscr{D}$,

$$(I_{g_0}(M)\backslash\mathscr{D}) \times \mathscr{D}'_{x_0} \to I_{g_0}(M)\backslash\mathscr{D},$$
$$(I_{g_0}(M)\circ f, h) \mapsto I_{g_0}(M)\circ f\circ h.$$

This action is also free, for if $I_{g_0}(M)\circ f\circ h = I_{g_0}(M)\circ f$, then $I_{g_0}(M)\circ f\circ h\circ f^{-1} = I_{g_0}(M)$, and so $f\circ h\circ f^{-1} \in I_{g_0}(M)$. Thus $h \in I_{(f^{-1})^*g_0}(M) \cap \mathscr{D}'_{x_0} = \{\mathrm{id}_M\}$. A slight modification of Theorem 4.2 then shows that the quotient space

$$(I_{g_0}(M)\backslash\mathscr{D})/\mathscr{D}'_{x_0} = \{(I_{g_0}(M)\circ f)\circ\mathscr{D}'_{x_0} \mid f \in \mathscr{D}\}$$

is a manifold. Clearly the quotient manifolds $I_g(M)\backslash(\mathscr{D}/\mathscr{D}'_{x_0})$ and $(I_{g_0}(M)\backslash\mathscr{D})/\mathscr{D}'_{x_0}$ are canonically identifiable with each other, and we can remove the parentheses

$$I_{g_0}(M)\backslash(\mathscr{D}/\mathscr{D}'_{x_0}) \approx I_{g_0}(M)\backslash\mathscr{D}/\mathscr{D}'_{x_0} \approx (I_{g_0}(M)\backslash\mathscr{D})/\mathscr{D}'_{x_0}.$$

The double coset manifold $I_{g_0}(M)\backslash\mathscr{D}/\mathscr{D}'_{x_0}$ has two interesting interpretations. For $g_0 \in \mathscr{M}$, the $\mathscr{D}$ orbit of g_0 in $\mathscr{M}$, $\mathscr{O}_{g_0} = \{f^*g_0 \mid f \in \mathscr{D}\}$, is a closed submanifold of $\mathscr{M}$ and is $\mathscr{D}'_{x_0}$ invariant, and so $\mathscr{D}'_{x_0}$ acts on $\mathscr{O}_{g_0}$ on the right,

$$\mathscr{O}_{g_0} \times \mathscr{D}'_{x_0} \to \mathscr{O}_{g_0}, \quad (g,h) \mapsto h^*g.$$

This action is free, and since $\mathscr{O}_{g_0}$ is a closed invariant submanifold of $\mathscr{M}$, the resulting quotient space

$$\mathscr{O}_{g_0}/\mathscr{D}'_{x_0} \subseteq \mathscr{M}/\mathscr{D}'_{x_0}$$

is a closed submanifold of $\mathscr{M}/\mathscr{D}'_{x_0}$ (see Theorem 5.2). Let $(g) = g\cdot\mathscr{D}'_{x_0} \in \mathscr{O}_{g_0}/\mathscr{D}'_{x_0}$ denote elements in the orbit space.

Proposition 4.3: Let $g_0 \in \mathscr{M}$, $u_0 \in F(M)$, and $x_0 = \pi_{FM}(u_0) \in M$. First assume that M is reversible or nonorientable. Then the diagram

$$\begin{array}{ccc} & I_{g_0}(M)\backslash\mathscr{D}/\mathscr{D}'_{x_0} & \\ \eta_1 \swarrow & & \searrow \eta_2 \\ \mathscr{O}_{g_0}/\mathscr{D}'_{x_0} & \xrightarrow{\quad\eta_3\quad} & I_{g_0}(M)/F(M) \end{array},$$

where

$$\eta_1: I_{g_0}(M)\circ f\circ\mathscr{D}'_{x_0} \mapsto (f^*g_0) = f^*g_0\cdot\mathscr{D}'_{x_0} \in \mathscr{O}_{g_0}/\mathscr{D}'_{x_0},$$
$$\eta_2: I_{g_0}(M)\circ f\circ\mathscr{D}'_{x_0} \mapsto [\hat{f}(u)] = \hat{I}_{g_0}(u) \in I_{g_0}(M)\backslash F(M),$$
$$\eta_3: (f^*g_0) \mapsto [\hat{f}(u_0)],$$

is a commuting diagram of diffeomorphisms.

If M is nonreversible, the above remains true if $F(M)$ is replaced by $F^+_{u_0}(M)$.

Proof: The proof follows by noting that the quotient manifold $I_{g_0}(M)\backslash\mathscr{D}$ is naturally diffeomorphic to the $\mathscr{D}$-orbit $\mathscr{O}_{g_0}$ by the diffeomorphism

$$I_{g_0}(M)\backslash\mathscr{D}\to\mathscr{O}_{g_0},\quad I_{g_0}(M)\circ f\mapsto f^*g_0.$$

Similarly, for $u_0\in F(M)$, $x_0=\pi_{FM}(u_0)$, the quotient manifold $\mathscr{D}/\mathscr{D}'_{x_0}$ is naturally diffeomorphic to $F(M)$ [or $F^+_{u_0}(M)$ if M is nonreversible] by the map

$$\tilde{\psi}_{u_0}:\ \mathscr{D}/\mathscr{D}'_{x_0}\to F(M),\quad f\circ\mathscr{D}'_{x_0}\mapsto\hat{f}(u_0).$$

Then the identifications $I_{g_0}(M)\backslash\mathscr{D}\approx\mathscr{O}_{g_0}$ and $\mathscr{D}/\mathscr{D}'_{x_0}\approx F(M)$ imply the chain of identifications

$$\mathscr{O}_{g_0}/\mathscr{D}'_{x_0}\approx(I_{g_0}(M)\backslash\mathscr{D})/\mathscr{D}'_{x_0}\approx I_{g_0}(M)\backslash\mathscr{D}/\mathscr{D}'_{x_0},$$
$$\approx I_{g_0}(M)\backslash(\mathscr{D}/\mathscr{D}'_{x_0})\approx I_{g_0}(M)\backslash F(M)$$

[or in the last step, $I_{g_0}(M)\backslash F^+_{u_0}(M)$ if M is nonreversible], which we diagram as

$$(f^*g_0)=f^*g_0\cdot\mathscr{D}'_{x_0}\leftarrow I_{g_0}(M)\circ f\circ\mathscr{D}'_{x_0}$$
$$\mapsto I_{g_0}(M)\cdot(\hat{f}(u))=[\hat{f}(u)].$$

□

V. ANOTHER RESOLUTION OF $\mathscr{G}$

As mentioned in the Introduction, for $x_0\in M$, the group $\mathscr{D}'_{x_0}$ acts freely on $\mathscr{M}$. In this section we show that the orbit space $\mathscr{M}/\mathscr{D}'_{x_0}$ is a manifold that resolves the singularities of $\mathscr{M}/\mathscr{D}$. This result follows easily from the methods used in Sec. III, together with a modification of the canonical splitting of $S_2(M)$. We begin with this modified splitting.

Proposition 5.1: Let $x_0\in M$, and let

$$\mathscr{X}(M)=\mathscr{X}'_{x_0}(M)\oplus\mathscr{Y}(M)$$

be the direct sum decomposition of $\mathscr{X}(M)$ according to Proposition 4.1. Let $g\in\mathscr{M}$. Then $\mathscr{Y}(M)$ can always be chosen so that

$$\mathscr{I}_g(M)\subseteq\mathscr{Y}(M).$$

If $\mathscr{Y}(M)$ is so chosen, then we have the direct sum decomposition

$$S_2(M)=\mathring{S}(g)\oplus\alpha_g(\mathscr{Y}(M))\oplus\alpha_g(\mathscr{X}'_{x_0}(M)),$$
$$h=\mathring{h}+L_Y g+L_{X'}g,$$

where each of the summands is a closed subspace of $S_2(M)$.

Proof. Since $\dim\mathscr{I}_g(M)\leq\frac{1}{2}n(n+1)<\dim\mathscr{Y}(M)=n^2+n$, and since a Killing vector field X is determined by $X(x_0)$ and $TX(x_0)$, it follows that $\mathscr{Y}(M)$ can be chosen so that $\mathscr{I}_g(M)\subseteq\mathscr{Y}(M)$.

From the canonical splitting, $S_2(M)$ splits as

$$S_2(M)=\mathring{S}_2(g)\oplus\text{range}\,\alpha_g,$$

where $\mathring{S}_2(g)$ and $\alpha_g(\mathscr{X}(M))$ are closed subspaces of $S_2(M)$. Thus it is sufficient to show the direct sum decomposition

$$\alpha_g(\mathscr{X}(M))=\alpha_g(\mathscr{Y}(M))\oplus\alpha_g(\mathscr{X}'_{x_0}(M))$$

for the closed subspace range $\alpha_g=\alpha_g(\mathscr{X}(M))$. Now $\alpha_g(\mathscr{Y}(M))$ is finite dimensional, and hence closed in $\alpha_g(\mathscr{X}(M))$. Also, $\alpha_g(\mathscr{X}'_{x_0}(M))$ is closed in $\alpha_g(\mathscr{X}(M))$, since $\alpha_g(\mathscr{X}(M))$ is closed in $S_2(M)$, and $\mathscr{X}'_{x_0}(M)$ is closed in $\mathscr{X}(M)$. Thus if $X_n\in\mathscr{X}'_{x_0}(M)$ is a sequence of vector fields such that $L_{X_n}g\to 0$,

$$X_n\to X\in\mathscr{I}_g(M)\cap\mathscr{X}'_{x_0}(M)=\ker\alpha_g\cap\mathscr{X}'_{x_0}(M)=\{0\},$$

and so $X=0$. Thus $\alpha_g(\mathscr{X}'_{x_0}(M))$ is closed in $\alpha_g(\mathscr{X}(M))$. Thus we must only show that the above direct sum is an algebraic direct sum.

Since $\mathscr{X}(M)=\mathscr{X}'_{x_0}(M)\oplus\mathscr{Y}(M)$, clearly

$$\alpha_g(\mathscr{X}(M))=\alpha_g(\mathscr{X}'_{x_0}(M))\oplus\alpha_g(\mathscr{Y}(M)).$$

So let $h\in\alpha_g(\mathscr{X}'_{x_0}(M))\cap\alpha_g(\mathscr{Y}(M))$. Then $h=L_Xg=L_Yg$ for some $X\in\mathscr{X}'_{x_0}(M)$ and $Y\in\mathscr{Y}(M)$, and so $L_{X-Y}g=0$. Thus $X-Y\in\mathscr{I}_g(M)\subseteq\mathscr{Y}(M)$ [since we are assuming $\mathscr{I}_g(M)\subseteq\mathscr{Y}(M)$], and so $X\in Y+\mathscr{Y}(M)=\mathscr{Y}(M)$. Thus $X\in\mathscr{X}'_{x_0}(M)\cap\mathscr{Y}(M)=\{0\}$, and so $X=0$. □

Remarks: (1) We shall refer to the above splitting of $S_2(M)$ as the *modified canonical splitting*. Thus if $h\in S_2(M)$ and $h=\mathring{h}+L_Xg$ is its canonical splitting, and

$$X=X'+Y\in\mathscr{X}'_{x_0}(M)\oplus\mathscr{Y}(M),\quad\mathscr{I}_g(M)\subseteq\mathscr{Y}(M),$$

is the splitting of X, then $h=\mathring{h}+L_Yg+L_{X'}g$ is the modified canonical splitting of h, where each of the pieces is uniquely determined. Moreover, even though X is determined only up to a Killing vector field, X' is uniquely determined, for if X is replaced by $X+X_1$ with $X_1\in\mathscr{I}_g(M)\subseteq\mathscr{Y}(M)$, then $(X+X_1)'=X'$ since $X'_1=0$. Thus $X+X_1$ splits as

$$X+X_1=X'+(Y+X_1),\quad X'\in\mathscr{X}'_{x_0}(M),$$
$$(Y+X_1)\in\mathscr{Y}(M),$$

where $Y+X_1$ is determined only up to a Killing vector field. Of course, the piece $L_Yg=L_{(Y+X_1)}g$ is uniquely determined.

(2) The summands in the modified canonical splitting are not L_2-orthogonal. Indeed, if $L_Yg\in\alpha_g(\mathscr{Y}(M))$ and $L_{X'}g\in\alpha_g(\mathscr{X}'_{x_0}(M))$,

$$\int\langle L_Yg,L_{X'}g\rangle_g\ d\mu_g$$

is not zero in general.

(3) When α_g is restricted to $\mathscr{X}'_{x_0}(M)$ it is injective and hence is an isomorphism onto its image. Thus $\alpha_g(\mathscr{X}'_{x_0}(M))\approx\mathscr{X}'_{x_0}(M)$. When α_g is restricted to $\mathscr{Y}(M)$, then $\ker\alpha_g=\mathscr{I}_g(M)\subseteq\mathscr{Y}(M)$, and so

$$\alpha_g(\mathscr{Y}(M))\approx\mathscr{Y}(M)/\mathscr{I}_g(M).$$

Thus

$$\text{range}\,\alpha_g=\alpha_g(\mathscr{X}'_{x_0}(M))\oplus\alpha_g(\mathscr{Y}(M))$$
$$\approx\mathscr{X}'_{x_0}(M)\oplus(\mathscr{Y}(M)/\mathscr{I}_g(M)).$$

(4) If $u_0\in F(M)$ is a frame at x_0, then $\mathscr{Y}(M)$ can be identified with $T_{u_0}F(M)$ by the map

$$\mathscr{Y}(M)\to T_uF(M),\quad Y\mapsto\hat{Y}(u_0),$$

and $\mathscr{I}_g(M)\subseteq\mathscr{Y}(M)$ gets mapped to $\hat{\mathscr{I}}_g(u_0)$. Thus the following spaces are isomorphic:

$$\alpha_g(\mathscr{Y}(M))\approx\mathscr{Y}(M)/\mathscr{I}_g(M)\approx T_uF(M)/\hat{\mathscr{I}}_g(u_0)$$
$$\approx\hat{\mathscr{I}}^{\perp}_g(u_0).$$

In particular,

$$\alpha_g(\mathscr{Y}(M)) \to \hat{\mathscr{I}}_g^\perp(u), \quad L_Y g \mapsto \hat{Y}^\perp(u)$$

is an isomorphism. □

For fixed $x_0 \in M$, let

$$\psi': \mathscr{M} \times \mathscr{D}'_{x_0} \to \mathscr{M}, \quad (g,f) \mapsto f^*g$$

be the right action of $\mathscr{D}'_{x_0}$ on $\mathscr{M}$. For $g \in \mathscr{M}$, let

$$\psi'_g: \mathscr{D}'_{x_0} \to \mathscr{M}, \quad f \mapsto f^*g$$

denote the orbit map through g, and let

$$\mathscr{O}'_g = \psi'_g(\mathscr{D}'_{x_0}) = \{f^*g | f \in \mathscr{D}'_{x_0}\} \subseteq \mathscr{M}$$

denote the orbit through g.

From the modified canonical splitting, together with the methods of Sec. III, we now have all of the ingredients necessary for the following theorem.

Theorem 5.2: The action of $\mathscr{D}'_{x_0}$ on $\mathscr{M}$ described above is smooth, free, and proper. The L_2 metric on $\mathscr{M}$ is invariant under this action.

For $g \in \mathscr{M}$, the orbit $\mathscr{O}'_g$ is a closed submanifold of $\mathscr{M}$ and the orbit map

$$\psi'_g: \mathscr{D}'_{x_0} \to \mathscr{O}'_g \subseteq \mathscr{M}$$

is a diffeomorphism. The tangent space to $\mathscr{O}'_g$ at $g_1 = f^*g \in \mathscr{O}'_g$ is

$$T_{g_1}\mathscr{O}'_g = \alpha_{g_1}(\mathscr{X}'_{x_0}(M)) = f^*(\alpha_g(\mathscr{X}'_{x_0}(M))).$$

The orbit space $\mathscr{M}/\mathscr{D}'_{x_0}$ has the structure of a smooth manifold such that the orbit projection map

$$\pi: \mathscr{M} \to \mathscr{M}/\mathscr{D}'_{x_0}, \quad g \mapsto (g)$$

is a smooth submersion. Moreover the projection π has the structure of a principal fiber bundle with total space $\mathscr{M}$, base space $\mathscr{M}/\mathscr{D}'_{x_0}$, and structure group $\mathscr{D}'_{x_0}$. For $g \in \mathscr{M}$,

$$\ker T_g\pi = T_g\mathscr{O}'_g = \alpha_g(\mathscr{X}'_{x_0}(M)),$$

and

$$\text{range } T_g\pi = T_{(g)}(\mathscr{M}/\mathscr{D}'_{x_0}) \approx S_2(M)/\alpha_g(\mathscr{X}'_{x_0}(M)).$$

Proof: Since $\mathscr{D}'_{x_0}$ is a closed ILH Lie subgroup of $\mathscr{D}$, smoothness and properness of the action follow from smoothness and properness of the action of $\mathscr{D}$ on $\mathscr{M}$. The action is free, since if $f^*g = g, f \in \mathscr{D}'_{x_0}$, then $f = \mathrm{id}_M$. The invariance of the L_2 metric is inherited for the action of any subgroup of $\mathscr{D}$ on $\mathscr{M}$.

For $s > n/2$, $k > 1$, and $g \in \mathscr{M}^{s+k}$, the orbit map

$$\psi'_g: (\mathscr{D}_{s+1})'_{x_0} \to \mathscr{M}^s, \quad f \mapsto f^*g$$

is a C^k map with derivative at $f \in (\mathscr{D}^{s+1})'_{x_0}$ given by

$$T_f\psi'_g: T_f(\mathscr{D}^{s+1})'_{x_0} \to T_{f^*g}\mathscr{M}^s \approx S_2^s(g), \quad X_f \mapsto f^*(L_X g),$$

where $X = X_f \circ f^{-1} \in \mathscr{X}'_{x_0}(M)^{s+1}$. Thus $T_f\psi'_g$ is an injection with closed range $T_f\psi'_g = f^*(\alpha_g(\mathscr{X}'_{x_0}(M)^{s+1}))$. Now let $\mathscr{Y}^{s+1}(M) \subseteq \mathscr{X}^{s+1}(M)$ be chosen so that

$$\mathscr{I}_g^{s+1}(M) = \{X \in \mathscr{X}^{s+1}(M) | L_X g = 0\} \subseteq \mathscr{Y}^{s+1}(M).$$

Then $\mathring{S}_2^s(g) \oplus \alpha_g(\mathscr{Y}^{s+1}(M))$ is a closed complement of $\alpha_g(\mathscr{X}'_{x_0}(M)^{s+1})$ and $f^*(\mathring{S}_2^s(g) \oplus \alpha_g(\mathscr{Y}^{s+1}(M)))$ is a closed complement of range $T_f\psi'_g$. Thus ψ'_g is an injective immersion and so $(\mathscr{O}'_g)^s$ is an immersed C^k submanifold. Since the action is proper, $(\mathscr{O}'_g)^s$ is a closed C^k submanifold and ψ'_g is a C^k diffeomorphism onto $(\mathscr{O}'_g)^s$. The C^∞ case then follows.

Using the exponential map of the L_2 metric on $\mathscr{M}^s$, a local cross section $C_g^s \subseteq \mathscr{M}^s$ at $g \in \mathscr{M}^{s+k}$, $k > 1$, is constructed by exponentiating a sufficiently small ball in $\mathring{S}_2^s(g) \oplus \alpha_g(\mathscr{Y}^{s+1}(M))$ onto $\mathscr{M}^s$. The resulting cross section contains the Ebin slice S_g^2 as well as small pieces of the orbits corresponding to $\alpha_g(\mathscr{Y}^{s+1}(M)) \approx \mathscr{Y}(M)/\mathscr{I}_g(M)$.

At f^*g, $f \in (\mathscr{D}^{s+1})'_{x_0}$, $f^*(\mathring{S}_2^s(g) \oplus \alpha_g(\mathscr{Y}^{s+1}(M)))$ is a closed complement of range $T_f\psi'_g = f^*(\alpha_g(\mathscr{X}'_{x_0}(M)^{s+1})) = \alpha_{f^*g}(\mathscr{X}'_{x_0}(M)^{s+1})$, the last equality following since $f_*(\mathscr{X}'_{x_0}(M)^{s+1}) = \mathscr{X}'_{x_0}(M)^{s+1}$. Thus if we exponentiate $f^*(\mathring{S}_2^s(g) \oplus \alpha_g(\mathscr{Y}^{s+1}(M)))$ along the orbit $(\mathscr{O}'_g)^s$, then by equivariance of the exponential map, the local cross sections are equivariant,

$$C_{f^*g}^s = f^*(C_g^s)$$

along the orbit. The remainder of the theorem then follows from the existence of these equivariant local cross sections as in Theorem 3.5. □

Remarks: (1) The modified canonical splitting of $S_2(M)$,

$$S_2(M) = \mathring{S}_2(g) \oplus \alpha_g(\mathscr{Y}(M)) \oplus \alpha_g(\mathscr{X}'_{x_0}(M)),$$

can be written as

$$T_g\mathscr{M} = T_{(g)}(\mathscr{M}/\mathscr{D}'_{x_0}) \oplus T_g(\mathscr{O}'_g).$$

(2) The modified canonical splitting does not in and of itself define a connection on the PFB $\pi: \mathscr{M} \to \mathscr{M}/\mathscr{D}'_{x_0}$. Indeed, if $g_2 = f^*g_1$, $f \in \mathscr{D}'_{x_0}$, and $\mathscr{Y}_1(M)$ and $\mathscr{Y}_2(M)$ are closed complements of $\mathscr{X}'_{x_0}(M)$ such that $\mathscr{I}_{g_1}(M) \subseteq \mathscr{Y}_1(M)$ and $\mathscr{I}_{g_2}(M) \subseteq \mathscr{Y}_2(M)$, then

$$\mathring{S}_2(g_2) \oplus \alpha_{g_2}(\mathscr{Y}_2(M)) \neq f^*(S_2(g_1) \oplus \alpha_{g_1}(\mathscr{Y}_1(M)))$$

unless $\mathscr{Y}_2(M) = (f^{-1})_*(\mathscr{Y}_1(M))$. However, a connection on π is defined by a smooth distribution $g \mapsto \mathscr{Y}_g(M)$ such that for each g,

(1) $\mathscr{Y}_g(M)$ is a closed complement of $\mathscr{X}'_{x_0}(M)$,

(2) $\mathscr{I}_g(M) \subseteq \mathscr{Y}_g(M)$,

(3) $\mathscr{Y}_{f^*g}(M) = (f^{-1})_*(\mathscr{Y}_g(M))$, for $f \in \mathscr{D}'_{x_0}$.

Note that if condition (2) holds at g, then, from (3),

$$\mathscr{I}_{f^*g}(M) = (f^{-1})_*(\mathscr{I}_g(M)) \subseteq (f^{-1})_*(\mathscr{Y}_g(M)) = \mathscr{Y}_{f^*g}(M),$$

so that (2) is then automatically true along the orbit $\mathscr{O}'_g$. □

The manifold $\mathscr{M}/\mathscr{D}'_{x_0}$ covers $\mathscr{M}/\mathscr{D}$ by the projection

$$\pi_1: \mathscr{M}/\mathscr{D}'_{x_0} \to \mathscr{M}/\mathscr{D}, \quad (g) \mapsto [g],$$

i.e., π_1 takes the $\mathscr{D}'_{x_0}$-orbits of $\mathscr{M}$ to the $\mathscr{D}$-orbits of $\mathscr{M}$. As in Sec. III the "fiber" $\pi_1^{-1}([g])$ over $[g]$ is a measure of the degree of the singularity of $\mathscr{M}/\mathscr{D}$ at $[g]$.

Proposition 5.3: Let $x_0 \in M$ and consider the sequence of projections

$$\mathscr{M} \xrightarrow{\pi} \mathscr{M}/\mathscr{D}'_{x_0} \xrightarrow{\pi_1} \mathscr{M}/\mathscr{D}, \quad g \mapsto (g) \mapsto [g].$$

Then for $[g_0] \in \mathscr{M}/\mathscr{D}$, $\pi_1^{-1}([g_0])$ is an embedded submanifold of $\mathscr{M}/\mathscr{D}'_{x_0}$ and

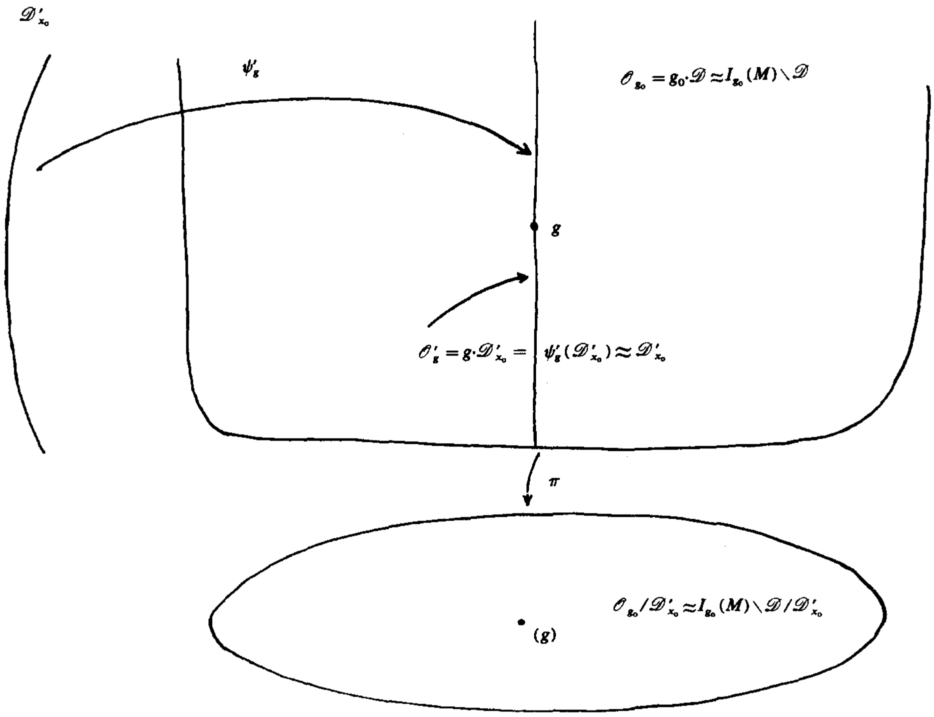


FIG. 3. The $\mathscr{D}'_{x_0}$-principal fiber bundle over $\mathscr{O}_{g_0}/\mathscr{D}'_{x_0} \approx I_{g_0}(M) \backslash \mathscr{D}/\mathscr{D}'_{x_0}$, with fiber at $(g) \in \mathscr{O}_{g_0}/\mathscr{D}'_{x_0}$ given by $\mathscr{O}'_g = g \cdot \mathscr{D}'_{x_0}$.

$$\pi_1^{-1}([g_0]) = (\mathscr{O}_{g_0}) = \mathscr{O}_{g_0}/\mathscr{D}'_{x_0} \approx I_{g_0}(M) \backslash \mathscr{D}/\mathscr{D}'_{x_0},$$

where the last identification is given by the diffeomorphism $I_{g_0}(M) \circ f \circ \mathscr{D}'_{x_0} \mapsto (f^* g_0)$.

Proof: $\pi_1^{-1}([g_0]) = \{(g) \in \mathscr{M}/\mathscr{D}'_{x_0} \mid [g] = [g_0]\}$. Thus if $(g) \in \pi_1^{-1}([g_0])$, and $g_1 \in (g)$, then $[g_1] = [g_0]$ and so $g_1 = f^* g_0$ for some $f \in \mathscr{D}$. Hence

$$\pi_1^{-1}([g_0]) = \{(f^* g_0) \in \mathscr{M}/\mathscr{D}'_{x_0} \mid f \in \mathscr{D}\} = \pi(\mathscr{O}_{g_0}).$$

Alternately, $\pi_1 \circ \pi: g \to [g]$, and so

$$(\pi_1 \circ \pi)^{-1}([g_0]) = \pi^{-1}(\pi_1^{-1}([g_0])) = \mathscr{O}_{g_0}.$$

Thus $\pi_1^{-1}([g_0]) = \pi(\mathscr{O}_{g_0})$. The identification $\mathscr{O}_{g_0}/\mathscr{D}'_{x_0} \approx I_{g_0}(M) \backslash \mathscr{D}/\mathscr{D}'_{x_0}$ has been shown in Proposition 4.3.

Since $\mathscr{D}'_{x_0}$ acts freely on the closed submanifold $\mathscr{O}_{g_0}$, it follows, as in Theorem 5.2, that $\mathscr{O}_{g_0}/\mathscr{D}'_{x_0} = \pi(\mathscr{O}_{g_0})$ is a closed submanifold of $\mathscr{M}/\mathscr{D}'_{x_0}$. □

Thus $\mathscr{M}/\mathscr{D}'_{x_0} \to \mathscr{M}/\mathscr{D}$ is an unfolding of the singularities of $\mathscr{M}/\mathscr{D}$.

As in Theorem 5.2, the projection

$$\pi: \mathscr{O}_{g_0} \to \mathscr{O}_{g_0}/\mathscr{D}'_{x_0}, \quad g \mapsto (g)$$

is a $\mathscr{D}'_{x_0}$-PFB with total space $\mathscr{O}_{g_0}$, base space $\mathscr{O}_{g_0}/\mathscr{D}'_{x_0} \approx I_{g_0}(M) \backslash \mathscr{D}/\mathscr{D}'_{x_0}$, and with fiber at $(g) \in \mathscr{O}_{g_0}/\mathscr{D}'_{x_0}$ given by $\pi^{-1}((g)) = \mathscr{O}'_g$ (see Fig. 3). Thus each $\mathscr{D}$-orbit $\mathscr{O}_{g_0}$ is the total space of a PFB over $I_{g_0}(M) \backslash \mathscr{D}/\mathscr{D}'_{x_0}$ and whose fibers are the $\mathscr{D}'_{x_0}$-orbits $\mathscr{O}'_g$ of $\mathscr{O}_{g_0}$. Thus, roughly speaking, the orbit $\mathscr{O}_{g_0}$ is a $I_{g_0}(M) \backslash \mathscr{D}/\mathscr{D}'_{x_0}$-thickening of the orbit $\mathscr{O}'_{g_0}$.

VI. REDUCTION OF THE BUNDLE $\mathscr{M} \times F(M) \to \mathscr{G}_{FM}$

We have now constructed two principal fiber bundles, which we diagram in Fig. 4, where $u_0 \in F(M)$ is a frame at $x_0 \in M$, and where the four projection maps have been described previously, but are here relabeled. In Fig. 4,

$$i_{u_0}: \mathscr{M} \to \mathscr{M} \times F(M), \quad g \mapsto (g,u),$$

and

$$d_{u_0}: \mathscr{M}/\mathscr{D}'_{x_0} \to (\mathscr{M} \times F(M))/\mathscr{D}, \quad (g) \mapsto [(g, u_0)].$$

Here d_{u_0} is well defined, since if $g_1 \in (g)$, $g_1 = h^* g$ for some $h \in \mathscr{D}'_{x_0}$, so $[(g_1, u_0)] = [(h^* g, u_0)] = [(g, u_0)]$ since $h^* u_0 = u_0$.

The base spaces $\mathscr{M}/\mathscr{D}'_{x_0}$ and $(\mathscr{M} \times F(M))/\mathscr{D}$ in each of the above principal fiber bundles are resolutions of the singularities in $\mathscr{M}/\mathscr{D}$. Here we complete our analysis by showing that these base spaces are diffeomorphic [with the possible replacing of $F(M)$ by $(F^+_{u_0}(M))$], and that the pair (i_{u_0}, d_{u_0}) is a reduction of the $\mathscr{D}$-PFB $\pi_1: \mathscr{M} \times F(M) \to \mathscr{G}_{FM}$ to the $\mathscr{D}'_{x_0}$-PFB $\pi_3: \mathscr{M} \to \mathscr{G}_{x_0}$. Our definition of a reduced subbundle is slightly more general than that of Kobayashi and Nomizu,[16] p. 53, inasmuch as we only require the base spaces to be diffeomorphic and not identical.

We also construct a fiber bundle over M associated with

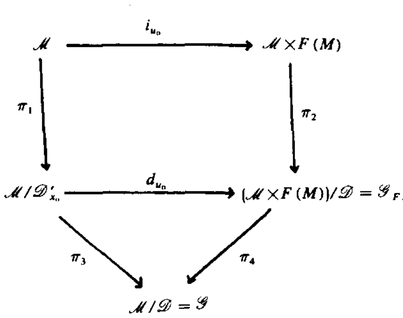


FIG. 4. A commuting pentagon showing the relationships between the two principal fiber bundles $\mathscr{M} \to \mathscr{M}/\mathscr{D}'_{x_0}$, $\mathscr{M} \times F(M) \to \mathscr{G}_{FM}$, and the space of Riemannian geometries $\mathscr{G}$.

the frame bundle, with standard fiber $\mathscr{G}_{FM}$, and whose fiber at x is $\mathscr{G}_x$. This bundle will represent the *grand resolution* of the singularities of $\mathscr{G}$.

Theorem 6.1: Let $x_0 \in M$, and let π_1: $\mathscr{M} \to \mathscr{M}/\mathscr{D}'_{x_0}$ and π_2: $\mathscr{M} \times F(M) \to (\mathscr{M} \times F(M))/\mathscr{D}$ be the two principal fiber bundles described previously. For a frame $u_0 \in F(M)$ over x_0, $\pi_{FM}(u_0) = x_0$, let

$$i_{u_0}: \mathscr{M} \to \mathscr{M} \times F(M), \quad g \mapsto (g,u_0).$$

Then with respect to the inclusion mapping i: $\mathscr{D}'_{x_0} \to \mathscr{D}$, i_{u_0} is an embedding of the $\mathscr{D}'_{x_0}$-PFB π_1 into the $\mathscr{D}$-PFB π_2, so that π_1 is a subbundle of π_2. The induced map on the base spaces is

$$d_{u_0}: \mathscr{M}/\mathscr{D}'_{x_0} \to (\mathscr{M} \times F(M))/\mathscr{D}, \quad (g) \mapsto [(g,u_0)],$$

so that the upper square of Fig. 4 commutes. Moreover, d_{u_0} is a diffeomorphism onto its image which is either $(\mathscr{M} \times F(M))/\mathscr{D}$ if M is reversible or nonorientable, or $\mathscr{M} \times F^+_{u_0}(M))/\mathscr{D}$ if M is nonreversible. Thus in the former case the pair (i_{u_0}, d_{u_0}) is a reduction of the structure group $\mathscr{D}$ of π_2 to the structure group $\mathscr{D}'_{x_0}$ of π_1, and the subbundle π_1 is a reduced subbundle of π_2. In the latter case, π_1 is a reduced subbundle of

$$\pi_2^+: \mathscr{M} \times F^+_{u_0}(M) \to (\mathscr{M} \times F^+_{u_0}(M))/\mathscr{D}.$$

The lower triangle of Fig. 4 commutes.

Proof: Let $h \in \mathscr{D}'_{x_0}$, and let u_0 be a frame at x_0. Then $h^* u_0 = u_0$, so that

$$i_{u_0}(h^* g,u) = h^*(g,u) = h^*(i_{u_0}(g)).$$

Thus with respect to the inclusion i: $\mathscr{D}'_{x_0} \to \mathscr{D}$, i_{u_0} is a PFB homomorphism. It is clearly an embedding, so that π_3 is a subbundle of π_1. Let d_{u_0} denote the induced mapping of the base spaces, defined so the square in Fig. 4 commutes. Thus $d_{u_0}(\pi_3(g)) = d_{u_0}((g)) = \pi_1(i_{u_0}(g)) = [(g,u_0)]$, as given previously. Since i_{u_0} is an embedding, the induced map d_{u_0} is also an embedding. Thus we must show that d_{u_0} is a submersion.

The derivative of d_{u_0}: $\mathscr{M}/\mathscr{D}'_{x_0} \to (\mathscr{M} \times F(M))/\mathscr{D}$ at (g) is easily computed to be

$$\begin{aligned} T_{(g)} d_{u_0}: \quad & T_g(\mathscr{M}/\mathscr{D}'_{x_0}) \approx S_2(M)/\alpha_g(\mathscr{X}'_{x_0}) \\ & \to T_{[(g,u_0)]}((\mathscr{M} \times F(M))/\mathscr{D}) \\ & \qquad \approx (S_2(M) \oplus T_{u_0}F(M))/\alpha_{(g,u_0)}(\mathscr{X}(M)), \\ & h + \alpha_g(\mathscr{X}'_{x_0}(M)) \mapsto h + \alpha_{(g,u_0)}(\mathscr{X}(M)). \end{aligned}$$

Now $T_{(g)} d_{u_0}$ is well defined, since d_{u_0} is well defined, or since $\alpha_g(\mathscr{X}'_{x_0}(M)) \subseteq \alpha_{(g,u_0)}(\mathscr{X}(M))$, and $T_{(g)} d_{u_0}$ is injective, since d_{u_0} is an embedding. Alternately, if

$$h_1 + \alpha_{(g,u_0)}(\mathscr{X}(M)) = h_2 + \alpha_{(g,u_0)}(\mathscr{X}(M)),$$

then $h_1 - h_2 \in \alpha_{(g,u_0)}(\mathscr{X}(M))$, and so $h_1 - h_2 = L_X g - X(u_0)$ for some $X \in \mathscr{X}(M)$. Thus $h_1 - h_2 = L_X g$ and $X(u_0) = 0$, and so $h_1 - h_2 \in \alpha_g(\mathscr{X}'_{x_0}(M))$, or $h_1 = h_2 + \alpha_g(\mathscr{X}'_{x_0}(M))$.

To show that $T_{(g)} d_{u_0}$ is surjective, let $h + Z_{u_0} + \alpha_{(g,u_0)}(\mathscr{X}(M)) \in T_{[(g,u_0)]}(\mathscr{M} \times F(M))/\mathscr{D}$, and let $Y \in \mathscr{X}(M)$ be such that $\hat{Y}(u_0) = Z_{u_0}$. Then

$$\begin{aligned} T_{(g)} d_{u_0}(h + L_Y g + \alpha_g(\mathscr{X}'_{x_0}(M))) \\ &= h + L_Y g + \alpha_{(g,u_0)}(\mathscr{X}(M)) \\ &= h + Z_{u_0} + (-Z_{u_0} + L_Y g) + \alpha_{(g,u_0)}(\mathscr{X}(M)) \\ &= h + Z_{u_0} + \alpha_{(g,u_0)}(Y) + \alpha_{(g,u_0)}(\mathscr{X}(M)) \\ &= h + Z_{u_0} + \alpha_{(g,u_0)}(\mathscr{X}(M)), \end{aligned}$$

since $\alpha_{(g,u_0)}(Y) = L_Y g - \hat{Y}(u_0) = L_Y g - Z_{u_0}$. Thus $T_{(g)} d_{u_0}$ is an isomorphism. The same arguments then show that

$$d^s_{u_0}: \mathscr{M}^s/(\mathscr{D}^{s+1})'_{x_0} \to (\mathscr{M}^s \times F(M))/\mathscr{D}^{s+1}$$

is an embedding and a submersion, and hence an isomorphism onto its image. That d_{u_0} is then an ILH diffeomorphism onto its image then follows in the usual manner.

If M is reversible or nonorientable, $\mathscr{D}$ acts transitively on $F(M)$. Hence if $[(g,u_1)] \in (\mathscr{M} \times F(M))/\mathscr{D}$ and $f \in \mathscr{D}$ is such that $u_0 = f^* u_1$, then

$$i_{u_0}(f^* g) = [(f^* g,u_0)] = [(f^* g,f^* u_1)] = [(g,u_1)],$$

so that $d_{u_0}((f^* g)) = [(g,u_1)]$. Hence d_{u_0} is surjective. If M is nonreversible, a similar argument shows that d_{u_0} is onto $(\mathscr{M} \times F^+_{u_0}(M))/\mathscr{D}$.

Now $\pi_4 \circ d_{u_0}((g)) = \pi_4([(g,u)]) = [g] = \pi_3(g)$ so the bottom triangle commutes. □

Remark: The inverse of d_{u_0} is given by

$$d^{-1}_{u_0}: \mathscr{G}_{FM} \to \mathscr{G}_{x_0}, \quad [(g,u)] \mapsto (f^* g),$$

where f is a solution to the equation $u = \hat{f}(u_0)$ [assuming M is reversible or nonorientable; otherwise replace $\mathscr{G}_{FM}$ with $\mathscr{G}_{F^+_{u_0}M} = (\mathscr{M} \times F^+_{u_0}(M))/\mathscr{D}$]. Thus

$$\begin{aligned} d_{u_0} \cdot (f^* g) &= [(f^* g,u_0)] = [(g,(f^{-1})^* u_0)] \\ &= [(g,\hat{f}(u_0))] = [(g,u)]. \end{aligned}$$ □

There is an interesting way to interpret the pentagon diagram associated with this theorem which shows the reasonability of the diffeomorphism d_{u_0}: $\mathscr{G}_{x_0} \to \mathscr{G}_{FM}$. This interpretation continues our interpretation of $\mathscr{G}_{FM} \to \mathscr{G}$ as being the pseudofiber bundle associated with the pseudo-$\mathscr{D}$-PFB $\mathscr{M} \to \mathscr{M}/\mathscr{D}$ (see the discussion preceding Proposition 3.6).

For a G-PFB π_P: $P \to M$, let H be a closed subgroup of G, and let

$$E = E(M, G/H, G, P) = (P \times (G/H))/G$$

denote the associated fiber bundle over M with the homogeneous fiber G/H, and with projection map

$$\pi_E: E \to M, \quad (P, aH) \cdot G \mapsto \pi_P(p).$$

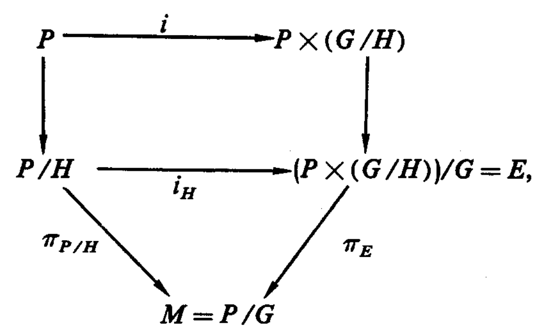
Then E can be canonically identified with P/H, namely,

$$\begin{aligned} i_H: \ & P/H \to (P \times (G/H))/G, \\ & (p) = pH \mapsto (p,H) \cdot G = [(p,H)]. \end{aligned}$$

Moreover, from the inclusion

$$i: P \to P \times (G/H), \quad p \mapsto (p,H),$$

we get the following pentagon diagram associated with the construction of the associated bundle E:



where $P \to P/H$ is an H-PFB, $P \times (G/H) \to E$ is a G-PFB, $\pi_{P/H}$: $P/H \to P/G$, $pH \mapsto pG$ and π_E: $E \to M$, $(p, aH)\cdot G \mapsto pG$ are fiber bundles over M, i_H is a diffeomorphism, and (i, i_H) is a reduction of the G-PFB to the H-PFB.

Now if $x_0 \in M$, $\mathscr{D}'_{x_0}$ is a closed ILH Lie subgroup of $\mathscr{D}$. Thus if we pretend that $\mathscr{M} \to \mathscr{M}/\mathscr{D}$ is a $\mathscr{D}$-PFB, then the associated fiber bundle

$$E(\mathscr{M}/\mathscr{D}, \mathscr{D}/\mathscr{D}'_{x_0}, \mathscr{D}, \mathscr{M}) = (\mathscr{M} \times (\mathscr{D}/\mathscr{D}'_{x_0}))/\mathscr{D}$$

with standard fiber $\mathscr{D}/\mathscr{D}'_{x_0}$ is canonically diffeomorphic to $\mathscr{M}/\mathscr{D}'_{x_0}$ by the map

$$\mathscr{M}/\mathscr{D}'_{x_0} \to (\mathscr{M} \times (\mathscr{D}/\mathscr{D}'_{x_0}))/\mathscr{D}, \quad (g) \mapsto [(g, \mathscr{D}'_{x_0})].$$

If $u_0 \in F(M)$ is a frame at x_0, then

$$\tilde{\psi}_{u_0}: \mathscr{D}/\mathscr{D}'_{x_0} \to F(M), \quad f \circ \mathscr{D}'_{x_0}(M) \mapsto \hat{f}(u_0)$$

is a diffeomorphism [if M is nonreversible, replace $F(M)$ with $F^+_{u_0}(M)$]. If we identify $\mathscr{D}/\mathscr{D}'_{x_0}$ with $F(M)$ [or $F^+_{u_0}(M)$], then

$$d_{u_0}: \mathscr{M}/\mathscr{D}'_{x_0} \to (\mathscr{M} \times (\mathscr{D}/\mathscr{D}'_{x_0}))/\mathscr{D}$$
$$\approx (\mathscr{M} \times F(M))/\mathscr{D}$$
$$[\text{or } (\mathscr{M} \times F^+_{u_0}(M))/\mathscr{D}]$$

is the diffeomorphism from $\mathscr{G}_{x_0} = \mathscr{M}/\mathscr{D}'_{x_0}$ to the total space $\mathscr{G}_{FM} = (\mathscr{M} \times F(M))/\mathscr{D}$ of the fiber bundle with standard fiber $F(M)$ associated with $\mathscr{M} \to \mathscr{M}/\mathscr{D}$. Thus d_{u_0} is analogous to the diffeomorphism i_H: $P/H \to (P \times (G/H))/H$, even though $\mathscr{M} \to \mathscr{M}/\mathscr{D}$ is not a *bona fide* fiber bundle. Moreover, the pentagon diagram of Theorem 6.1 is then analogous to the general pentagon diagram for the construction of the associated fiber bundle $E(M, G/H, G, P)$. Of course the projections $\mathscr{G}_{x_0} \to \mathscr{G}$ and $\mathscr{G}_{FM} \to \mathscr{G}$ are not fiber bundles since $\mathscr{G}_{FM} \to \mathscr{G}$ is not a principal fiber bundle.

We now make three other remarks regarding Theorem 6.1. First, recall that the structure group G of a PFB $P(M, G)$ is reducible to a closed subgroup H if and only if the associated fiber bundle $E = E(M, G/H, G, P)$ admits a cross section σ: $M \to P/H = E$ (see, e.g., Kobayashi–Nomizu,[16] p. 57). Thus since the structure group $\mathscr{D}$ of the PFB π_1: $\mathscr{M} \times F(M) \to \mathscr{G}_{FM}$ is reducible to the closed subgroup $\mathscr{D}'_{x_0}$ (assuming M is reversible or nonorientable), the associated fiber bundle

$$\pi_5: \frac{(\mathscr{M} \times F(M)) \times (\mathscr{D}/\mathscr{D}'_{x_0})}{\mathscr{D}} \approx \frac{\mathscr{M} \times F(M)}{\mathscr{D}'_{x_0}}$$
$$\to \frac{\mathscr{M} \times F(M)}{\mathscr{D}}$$

must admit a cross section

$$\sigma: \frac{\mathscr{M} \times F(M)}{\mathscr{D}} \to \frac{\mathscr{M} \times F(M)}{\mathscr{D}'_{x_0}}.$$

Here the identification of $((\mathscr{M} \times F(M)) \times (\mathscr{D}/\mathscr{D}'_{x_0}))/\mathscr{D}$ with $(\mathscr{M} \times F(M))/\mathscr{D}'_{x_0}$ is the canonical identification of $(P \times G/H)/G$ with P/H discussed above, where $P \to M$ now corresponds to the *bona fide* PFB π_1: $\mathscr{M} \times F(M) \to \mathscr{G}_{FM}$, and not the pseudo-PFB $\mathscr{M} \to \mathscr{G}$, since π_1 is the PFB whose structure group $\mathscr{D}$ is reduced to $\mathscr{D}'_{x_0}$.

For a frame $u_0 \in F(M)$ at x_0, an explicit cross section is given by

$$\sigma_{u_0}: \frac{\mathscr{M} \times F(M)}{\mathscr{D}} \to \frac{\mathscr{M} \times F(M)}{\mathscr{D}'_{x_0}},$$
$$[(g,u)] \mapsto ((f^{-1})^* g, u_0)\cdot\mathscr{D}'_{x_0},$$

where $f \in \mathscr{D}$ satisfies $f^* u_0 = u$ [or $\hat{f}(u) = u_0$], and where we denote elements of $(\mathscr{M} \times F(M))/\mathscr{D}'_{x_0}$ by $(g,u)\cdot\mathscr{D}'_{x_0}$. To show σ_{u_0} is well defined, we must show that it is independent of the choice of $f \in \mathscr{D}$ that satisfies $f^* u_0 = u$, and that it is independent of the choice of representative of $[(g,u)]$. Thus, first, if $f_1, f_2 \in \mathscr{D}$ are such that $f_1^* u_0 = u$ and $f_2^* u_0 = u$, then $(f_2^{-1})^* f_1^* u_0 = (f_1 \circ f_2^{-1})^* u_0 = u_0$, and so $f_1 \circ f_2^{-1} \in \mathscr{D}'_{x_0}$. Thus

$$\sigma_{u_0}([(g,u)]) = ((f_1^{-1})^* g, u_0)\cdot\mathscr{D}'_{x_0}$$
$$= ((f_2^{-1})^* f_1^* (f_1^{-1})^* g, (f_2^{-1})^* f_1^* u_0)\cdot\mathscr{D}'_{x_0}$$
$$= ((f_2^{-1})^* g, u_0)\cdot\mathscr{D}'_{x_0},$$

and so σ_{u_0} is independent of the choice of solution $f^* u_0 = u$.

Second, if $[(g,u_2)] = [(g_1,u_1)]$, then $(g_2,u_2) = (f_1^* g_1, f_1^* u_1)$ for some $f_1 \in \mathscr{D}$. Thus if $f^* u_0 = u_1$ then $f_1^* f^* u_0 = (f \circ f_1)^* u_0 = f_1^* u_1$ and so

$$\sigma_{u_0}([f_1^* g_1, f_1^* u_1])$$
$$= (((f \circ f_1)^{-1})^* f_1^* g_1, u_0)\cdot\mathscr{D}'_{x_0}$$
$$= ((f^{-1})^* g_1, u_0)\cdot\mathscr{D}'_{x_0} = \sigma_{u_0}([(g_1,u_1)]),$$

since $f^* u_0 = u_1$. Thus σ_{u_0} is well defined and if $f^* u_0 = u$, then

$$\pi_5(\sigma_{u_0}([(g,u]))$$
$$= \pi_5(((f^{-1})^* g, u_0)\cdot\mathscr{D}'_{x_0}) = [((f^{-1})^* g, u_0)]$$
$$= [(g, f^* u_0)] = [(g,u)],$$

so that σ_{u_0} is a cross section.

As our second remark, we consider the question of whether or not the natural connection on the $\mathscr{D}$-PFB π_2: $\mathscr{M} \times F(M) \to \mathscr{G}_{FM}$ is reducible to a connection on the reduced $\mathscr{D}'_{x_0}$-PFB π_1: $\mathscr{M} \to \mathscr{G}_{x_0}$. Since the Lie algebra $\mathscr{X}(M)$ of $\mathscr{D}$ admits a direct sum decomposition $\mathscr{X}(M) = \mathscr{X}'_{x_0}(M) \oplus \mathscr{Y}(M)$, where $\mathscr{X}'_{x_0}(M)$ is the Lie algebra of $\mathscr{D}'_{x_0}$, the natural connection on π_2 is reducible to a connection on π_1 if

$$\text{Ad}(\mathscr{D}'_{x_0})\cdot(\mathscr{Y}(M)) = \mathscr{Y}(M)$$

(Kobayashi and Nomizu,[16] p. 83). But if $h \in \mathscr{D}'_{x_0}$ and $Y \in \mathscr{Y}(M)$ with flow f_λ, then

$$\text{Ad}(h)\cdot Y = \frac{d}{d\lambda}(h \circ f_\lambda \circ h^{-1})|_{\lambda=0}$$

$= Th \circ Y \circ h^{-1} = h^*Y.$

Since h^*Y is not in $\mathscr{Y}(M)$ for all $h \in \mathscr{D}'_{x_0}$,

$$\mathrm{Ad}(\mathscr{D}'_{x_0}) \cdot \mathscr{Y}(M) \neq \mathscr{Y}(M).$$

Thus the $\mathscr{X}'_{x_0}(M)$-component of the connection one-form of the natural connection on π_2: $\mathscr{M} \times F(M) \rightarrow \mathscr{G}_{FM}$ when pulled back to π_1: $\mathscr{M} \rightarrow \mathscr{G}_{x_0}$ is *not* a connection one-form on π_1. This is another way of seeing that there is apparently no natural connection on π_1: $\mathscr{M} \rightarrow \mathscr{G}_{x_0}$ [see also remark (2) following Proposition 5.2].

As a last remark regarding Theorem 6.1, we note that for $[g_0] \in \mathscr{G}$, the fibers

$$\pi_3^{-1}([g_0]) = \pi_1(\mathscr{O}_{g_0})$$
$$= \{(f^*g_0) \in \mathscr{G}_{x_0} \mid f \in \mathscr{D}\} \subseteq \mathscr{G}_{x_0}$$

and

$$\pi_4^{-1}([g_0]) = \pi_2(\mathscr{O}_{g_0} \times F(M))$$
$$= \{[(f^*g_0, u)] \in \mathscr{G}_{FM} | f \in \mathscr{D},\ u \in F(M)\}$$
$$= \{[(g_0, u)] \in \mathscr{G}_{FM} | u \in F(M)\} \subseteq \mathscr{G}_{FM}$$

are embedded submanifolds. Since the bottom triangle in Fig. 4 commutes, the diffeomorphism d_{u_0}: $\mathscr{G}_{x_0} \rightarrow \mathscr{G}_{FM}$ restricts to these submanifolds to give a diffeomorphism (if M is reversible or nonorientable)

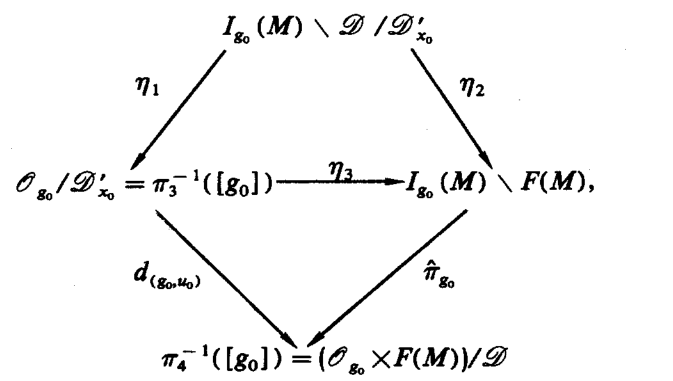
$$d_{(g_0,u_0)} = d_{u_0} \restriction \pi_3^{-1}([g_0]):$$
$$\pi_3^{-1}(([g_0])) \rightarrow \pi_4^{-1}([g_0]),$$
$$(f^*g_0) \mapsto [(f^*g_0, u_0)] = [(g_0, (f^{-1})^*u_0)].$$

From Proposition 3.6, there is also a diffeomorphism

$$\tilde{\pi}_{g_0}: I_{g_0}(M) \setminus F(M) \rightarrow \pi_4^{-1}([g_0]), \quad [u] \mapsto [(g_0, u)].$$

Thus another layer can be added to the commuting triangle of Proposition 4.3.

Proposition 6.2: Let $g_0 \in \mathscr{M}$ and let $u_0 \in F(M)$ be a frame at x_0. Assume M is reversible or nonorientable [otherwise replace $F(M)$ with $F^+_{u_0}(M)$]. Then the following rhombic diagram is a commuting diagram of diffeomorphisms:



where η_1, η_2, η_3 are given in Proposition 4.3, $\hat{\pi}_{g_0}$ is given in Proposition 3.6, and $d_{(g_0,u_0)}$: $(f^*g_0) \mapsto [(f^*g_0, u_0)]$.

Proof: That the upper triangle is a commuting triangle of diffeomorphisms is Proposition 4.3. The lower triangle commutes since

$$\hat{\pi}_{g_0}(\eta_3(F^*g_0)) = \hat{\pi}_{g_0}([\tilde{f}(u_0)]) = [(g_0, \hat{f}(u_0))]$$
$$= [(g_0, (f^{-1})^*u_0)] = [(f^*g_0, u_0)]$$
$$= d_{(g_0,u_0)} \cdot (f^*g_0).$$

Since η_3 and $\hat{\pi}_{g_0}$ are diffeomorphisms, so is $d_{(g_0,u_0)}$. □

Finally, we construct a fiber bundle π_E: $E \rightarrow M$ associated with the frame bundle π_{FM}: $F(M) \rightarrow M$, with standard fiber $\mathscr{G}_{FM}$, and whose fiber at each $x \in M$, $\pi_E^{-1}(x)$ is naturally diffeomorphic with $\mathscr{G}_x$ [assuming M is reversible or nonorientable; otherwise replace $\mathscr{G}_{FM}$ with $\mathscr{G}_{F^+_{u_0}(M)} = \mathscr{M} \times F^+_{u_0}(M))/\mathscr{D}$, where u_0 is a frame at x_0].

Consider the right action of $\mathrm{GL}(n)$ on $\mathscr{M} \times F(M)$,

$$(\mathscr{M} \times F(M)) \times \mathrm{GL}(n) \rightarrow \mathscr{M} \times F(M),$$
$$((g,u), A) \mapsto (g,u) \cdot A \equiv (g, u \cdot A).$$

Since $\hat{f} \circ R_A = R_A \circ \hat{f}$ for all $A \in \mathrm{GL}(n)$ and $f \in \mathscr{D}$, this action commutes with the action of $\mathscr{D}$ on $\mathscr{M} \times F(M)$,

$$(f^*g, f^*u) \cdot A = (f^*g, f^*u \cdot A)$$
$$= (f^*g, f^*(u \cdot A)) = f^*(g, u \cdot A),$$

and so passes to a right $\mathrm{GL}(n)$-action on the quotient manifold $\mathscr{G}_{FM} = (\mathscr{M} \times F(M))/\mathscr{D}$,

$$\mathscr{G}_{FM} \times GL(n) \rightarrow \mathscr{G}_{FM}, \ ([(g,u)], A)$$
$$\mapsto [(g,u)] \cdot A = [(g,u) \cdot A] = [(g, u \cdot A)].$$

Let

$$E = E(M, \mathscr{G}_{FM}, \mathrm{GL}(n), F(M))$$
$$= (F(M) \times \mathscr{G}_{FM})/\mathrm{GL}(n)$$

denote the resulting fiber bundle associated with the frame bundle π_{FM}: $F(M) \rightarrow M$ and with standard fiber $\mathscr{G}_{FM}$, and let

$$\pi_E: E \rightarrow M, \quad (u, [(g, u_1)]) \cdot \mathrm{GL}(n) \mapsto \pi_{FM}(u)$$

denote the projection map, where we denote elements of E by

$$(u, [(g, u_1)]) \cdot \mathrm{GL}(n)$$
$$= \{(u \cdot A, [(g, u_1 \cdot A)]) | A \in \mathrm{GL}(n)\} \in E.$$

Then for $x_0 \in M$,

$$\pi_E^{-1}(x_0) = \{(u_0, [(g,u)]) \cdot \mathrm{GL}(n) |$$
$$[(g,u)] \in \mathscr{G}_{FM} \text{ and } \pi_{FM}(u_0) = x_0\}$$

(see Fig. 5).

Now let

$$\pi: F(M) \times \mathscr{G}_{FM} \rightarrow E = (F(M) \times \mathscr{G}_{FM})/\mathrm{GL}(n)$$

denote the natural orbit projection map, so as previously discussed (see discussion preceding Proposition 3.6), π is a $\mathrm{GL}(n)$-PFB over E. Moreover, if $u_0 \in F(M)$ is a frame at $x_0 = \pi_{FM}(u_0)$, there is a diffeomorphism

$$\pi_{u_0}: \mathscr{G}_{FM} \rightarrow \pi_E^{-1}(x_0) \subseteq E,$$
$$[(g,u)] \mapsto \pi(u_0, [(g,u)]) = (u_0, [(g,u)]) \cdot \mathrm{GL}(n)$$

of the standard fiber $\mathscr{G}_{FM}$ onto $\pi_E^{-1}(x_0)$. Remarkably, the fibers $\pi_E^{-1}(x)$ can be naturally identified with the $\mathscr{G}_x$'s.

Proposition 6.3: Let $x_0 \in M$ and let $u_0 \in F(M)$ be a frame at x_0. Then the map

$$\pi_{u_0} \circ d_{u_0}: \mathscr{G}_{x_0} \rightarrow \pi_E^{-1}(x_0),$$
$$(g) \mapsto (u_0, [(g, u_0)]) \cdot \mathrm{GL}(n)$$

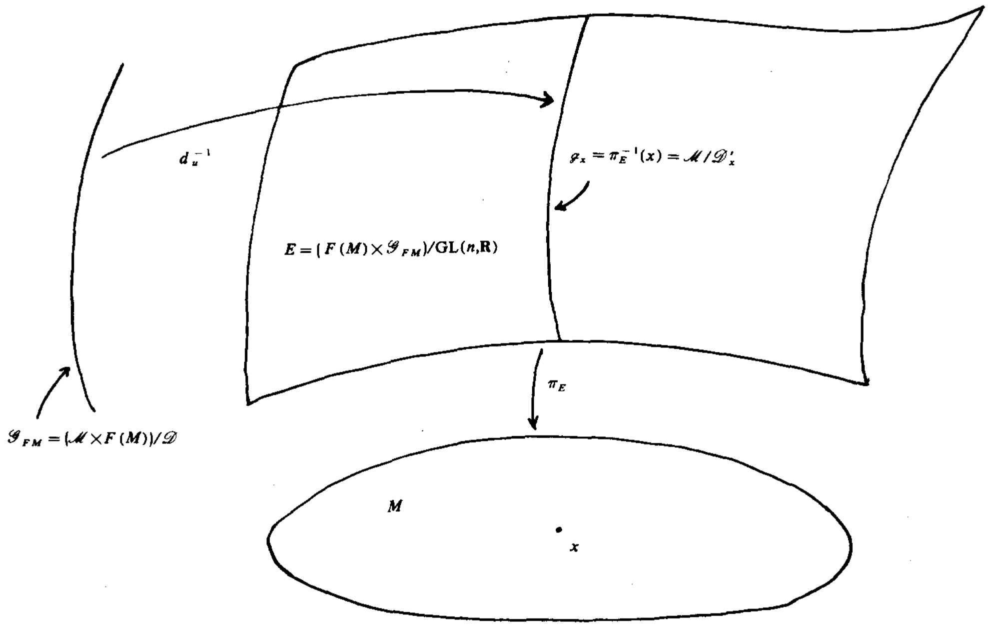


FIG. 5. The grand resolution $E = E(M, \mathscr{G}_{FM}, \mathrm{GL}(n), F(M))$ of $\mathscr{G}$. The fiber bundle E has standard fiber the canonical resolution space $\mathscr{G}_{FM}$, and has fiber at $x \in M$ the particular resolution space $\mathscr{G}_x$. Here E incorporates all the particular resolutions in its total space, parametrized by $x \in M$.

is a diffeomorphism independent of the choice of frame u_0 at x_0, where

$$d_{u_0}: \mathscr{G}_{x_0} \to \mathscr{G}_{FM}, \quad (g) \mapsto [(g,u_0)]$$

is the diffeomorphism of Theorem 6.1.

Proof: For $A \in \mathrm{GL}(n)$,

$$\begin{aligned}\pi_{u_0 \cdot A}([(g,u)]) &= (u_0 \cdot A, [(g,u)]) \cdot \mathrm{GL}(n) \\ &= (u_0, [(g,u)] \cdot A^{-1}) \cdot \mathrm{GL}(n) \\ &= \pi_{u_0}([(g,u)] \cdot A^{-1}) \\ &= \pi_{u_0} \circ \widetilde{R}_{A^{-1}}([(g,u)]),\end{aligned}$$

where

$$\widetilde{R}_A: \mathscr{G}_{FM} \to \mathscr{G}_{FM}, \quad [(g,u)] \mapsto [(g,u)] \cdot A = [(g,u \cdot A)]$$

denotes the diffeomorphism of $\mathscr{G}_{FM}$ induced by A and the action of $\mathrm{GL}(n)$ on $\mathscr{G}_{FM}$. Thus $\pi_{u_0 \cdot A} = \pi_{u_0} \circ \widetilde{R}_{A^{-1}}$. Similarly, the diffeomorphism $d_{u_0}: (g) \mapsto [(g,u_0)]$ satisfies

$$\begin{aligned}d_{u_0 \cdot A}((g)) &= [(g,u_0 \cdot A)] = [(g,u_0)] \cdot A \\ &= \widetilde{R}_A([g,u_0]) = \widetilde{R}_A(d_{u_0}((g))),\end{aligned}$$

so that $d_{u_0 \cdot A} = \widetilde{R}_A \circ d_{u_0}$. Thus the composition

$$\begin{aligned}\pi_{u_0} \circ d_{u_0}: \mathscr{G}_{x_0} &\to \mathscr{G}_{FM} \to \pi_E^{-1}(x_0), \\ (g) &\mapsto [(g,u_0)] \mapsto (u_0,[g,u_0]) \cdot \mathrm{GL}(n)\end{aligned}$$

satisfies

$$\pi_{u_0 \cdot A} \circ d_{u_0 \cdot A} = (\pi_{u_0} \circ \widetilde{R}_{A^{-1}}) \circ (\widetilde{R}_A \circ d_{u_0}) = \pi_{u_0} \circ d_{u_0}$$

and so is independent of the frame u_0 at x_0. □

Thus we have proven the following (see Fig. 5).

Theorem 6.4: Let

$$\begin{aligned}E &= E(M, \mathscr{G}_{FM}, \mathrm{GL}(n), F(M)) \\ &= (F(M) \times \mathscr{G}_{FM})/\mathrm{GL}(n)\end{aligned}$$

be the fiber bundle over M with the standard fiber $\mathscr{G}_{FM}$, and which is associated with the frame bundle $F(M)$. Let

$$\pi_E: E \to M, \quad (u,[(g,u_1)]) \cdot \mathrm{GL}(n) \mapsto \pi_{FM}(u)$$

denote the projection map for E. Then if $x \in M$, the fiber

$$\pi_E^{-1}(x) = \{(u,[g,u_1]) \cdot \mathrm{GL}(n) \mid [(g,u_1)] \in \mathscr{G}_{FM} \text{ and } \pi_{FM}(u) = x\}$$

is naturally diffeomorphic with $\mathscr{G}_x = \mathscr{M}/\mathscr{D}'_x$ by the map

$$\pi_u \circ d_u: (g) \mapsto (u,[(g,u)]) \cdot \mathrm{GL}(n),$$

which is independent of the choice of frame u at x.

If u_0 is a frame at x_0, then u_0 induces a diffeomorphism

$$\begin{aligned}\pi_{u_0}: \mathscr{G}_{FM} &\to \pi_E^{-1}(x_0), \\ [(g,u)] &\mapsto (u_0, [(g,u)]) \cdot \mathrm{GL}(n),\end{aligned}$$

and a diffeomorphism

$$d_{u_0}^{-1}: \mathscr{G}_{FM} \to \mathscr{G}_{x_0}, \quad [(g,u)] \mapsto (f^* u),$$

where $f \in \mathscr{D}$ satisfies $\hat{f}(u) = u_0$. □

The above construction provides a "grand" bundle viewpoint of Theorem 6.1 in the following sense. The canonical resolution of $\mathscr{G}$ is $\mathscr{G}_{FM} \to \mathscr{G}$, and for each $x \in M$, there is a particular resolution $\mathscr{G}_x \to \mathscr{G}$. Moreover, a frame u at x induces a diffeomorphism $d_u^{-1}: \mathscr{G}_{FM} \to \mathscr{G}_x$ that gives a representation of the canonical resolution $\mathscr{G}_{FM}$ onto a particular resolution $\mathscr{G}_x$. If each fiber $\pi_E^{-1}(x)$ in the bundle $E \to M$ is identified with the particular resolution $\mathscr{G}_x$, then the total space E contains all the particular resolutions $\mathscr{G}_x$,

$$E = \bigcup_{x \in M} \mathscr{G}_x.$$

Thus in this picture, the base space M parametrizes the particular resolutions, and if a frame u at x is given, the diffeomorphism

$$d_u^{-1}: \mathscr{G}_{FM} \to \mathscr{G}_x$$

is the usual identification of the standard fiber with the fiber at x that is induced in an associated fiber bundle when a point $u \in F(M)$ in the principal fiber bundle is chosen. This is the bundle viewpoint of Theorem 6.1.

Moreover, inasmuch as the bundle $E \to M$ contains the totality of all the particular resolutions $\mathscr{G}_x$ as its fibers, and each frame $u \in F(M)$ gives a representation of the canonical resolution $\mathscr{G}_{FM}$, or standard fiber of E, onto the particular resolution $\mathscr{G}_x$, $x = \pi_{FM}(u)$, or fiber of E at x, $E = (F(M) \times \mathscr{G}_{FM})/\mathrm{GL}(n)$ may properly be deemed to be the *grand resolution of* $\mathscr{G}$.

VII. FURTHER WORK

The techniques used in this paper can be applied to desingularize the moduli space of connections on a principal fiber bundle. In outline, this program would proceed as follows. Let $P = P(M, G)$ denote a principal fiber bundle over a compact connected manifold M with structure Lie group G. Let $\mathscr{C}(P)$ denote the space of all connections on P, and let $\mathrm{Aut}(P)$ denote the group of automorphisms of P that cover the identity of M. Then $\mathrm{Aut}(P)$ acts on $\mathscr{C}(P)$ on the right by pullback,

$$C(P) \times \mathrm{Aut}(P) \to C(P); \quad (\omega, F) \to F^*\omega .$$

The resulting orbit space

$$C(P)/\mathrm{Aut}(P)$$

is the space of moduli of connections on P. Because of the presence of nonisomorphic isotropy groups at different connections, this action is not free, and thus the resulting orbit space in general is not a manifold. Here the isotropy group at a connection is just the symmetry group of that connection. Note that this situation is entirely analogous to the nonmanifold nature of the space of Riemannian geometries (Fischer[4]). Another remark is that the action of $\mathrm{Aut}(P)$ in general is not even effective, inasmuch as the center of G can be identified with a subgroup of $\mathrm{Aut}(P)$ that fixes every connection. This subgroup, however, is normal in $\mathrm{Aut}(P)$ and can be factored out to produce an effective action.

Our idea to produce a free action is completely analogous to extending $\mathscr{M}$ to $\mathscr{M} \times F(M)$ in order to get a free action of $\mathscr{D}$. Thus we extend $\mathscr{C}(P)$ to $\mathscr{C}(P) \times P$. Then $\mathrm{Aut}(P)$ acts on the right on this product.

$$(\mathscr{C}(P) \times P) \times \mathrm{Aut}(P) \to \mathscr{C}(P) \times P,$$

$$((\omega, p), F) \mapsto (F^*\omega, F^{-1}(p)) .$$

This action is free, since if $F \in \mathrm{Aut}(P)$ satisfies $F^*\omega = \omega$ and $F(p) = p$ for some $p \in P$, then $F = \mathrm{id}_P$. A further argument shows that $\mathrm{Aut}(P)$ acts smoothly and properly. It then follows that the orbit space $(\mathscr{C}(P) \times P)/\mathrm{Aut}(P)$ is a smooth infinite-dimensional manifold, and that the natural projection

$$\pi\colon (\mathscr{C}(P) \times P) \mapsto (\mathscr{C}(P) \times P)/\mathrm{Aut}(P)$$

has the structure of a principal fiber bundle with structure group $\mathrm{Aut}(P)$. The base space $(\mathscr{C}(P) \times P)/\mathrm{Aut}(P)$ in turn covers $\mathscr{C}(P)/\mathrm{Aut}(P)$ by the further projection

$$[(\omega, p)] \to [\omega] .$$

In this sense, $(\mathscr{C}(P) \times P)/\mathrm{Aut}(P)$ is a desingularization of $\mathscr{C}(P)/\mathrm{Aut}(P)$. Note that $\mathrm{Aut}(P)$ acts naturally on P on the left, so that if $\mathscr{C}(P) \to \mathscr{C}(P)/\mathrm{Aut}(P)$ were a principal fiber bundle over the manifold $\mathscr{C}(P)/\mathrm{Aut}(P)$, then $(\mathscr{C}(P) \times P)/\mathrm{Aut}(P)$ would be the associated fiber bundle over the base manifold $\mathscr{C}(P)/\mathrm{Aut}(P)$ with standard fiber P. This situation is exactly analogous to the desingularization of the space of Riemannian geometries (compare Sec. I).

A similar desingularization can be accomplished by considering the subgroup $\mathrm{Aut}_p(P)$ of $\mathrm{Aut}(P)$ that fixes an arbitrary point p of P, and then taking the orbit space of $\mathscr{C}(P)$ by this subgroup. The resulting orbit space

$$\mathscr{C}(P)/\mathrm{Aut}_p(P)$$

is then a manifold, since the action of this restricted group is free, smooth, and proper. Moreover, this desingularization is isomorphic to the one above. Also, these particular desingularizations can be tied together by an infinite-dimensional fiber bundle associated with $P(M, G)$, analogous to Theorem 6.4, where the canonical desingularization $(\mathscr{C}(P) \times P)/\mathrm{Aut}(P)$ acts as a model for all the particular desingularizations $\mathscr{C}(P)/\mathrm{Aut}_p(P)$. The details of these considerations will be published elsewhere.

Finally, we remark that an announcement of the results contained in this paper is contained in Ref. 17. In that work, and occasionally in the current work, because of the geometric suggestiveness of the term unfolding as a smooth covering manifold of a singular manifold, we use it interchangeably with the term desingularization.

[1] J. A. Wheeler, "Geometrics and the issue of the final state," in *Relativity, Groups, and Topology*, edited by B. DeWitt and C. DeWitt (Gordon and Breach, New York, 1974).

[2] R. S. Palais, "Letter of Palais to Lang," 1961 (unpublished).

[3] D. Ebin, "On the space of Riemannian metrics," Symp. Pure Math., Am. Math. Soc. **15**, 11 (1970).

[4] A. Fischer, "The Theory of superspace," in *Relativity*, edited by M. Carmeli, S. Fickler, and L. Witten (Plenum, New York, 1970).

[5] J-P. Bourguignon, "Une stratification de l'espace des structures Riemanniennes," Comp. Math. **30**, 1 (1975).

[6] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).

[7] H. Omori, "On the group of diffeomorphisms on a compact manifold," Symp. Pure Math., Am. Math. Soc. **15**, 167 (1970).

[8] D. Ebin and J. Marsden, "Groups of diffeomorphisms and the motion of an incompressible fluid," Ann. Math. **92**, 102 (1970).

[9] D. Fried, "Some irreversible three-manifolds," preprint, University of California, Santa Cruz, 1980.

[10] R. S. Palais, "Slices and equivariant imbeddings," in *Seminar on Transformation Groups*, edited by A. Borel (Princeton U.P., Princeton, NJ, 1960).

[11] R. Abraham and J. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978), 2nd ed.

[12] R. S. Palais, "Slices for the action of Diff(M) on the space of Riemannian metrics," University of California, preprint, 1977.

[13] A. Wilansky, *Modern Methods in Topological Vector Spaces* (McGraw-Hill, New York, 1978).

[14] A. Fischer and J. Marsden, "The manifold of conformally equivalent metrics," Can. J. Math. **29**, 193 (1977).

[15] A. Fischer, J. Marsden, and V. Moncrief, "The structure of the space of solutions of Einstein's equations. I. One Killing field," Ann. Inst. H. Poincaré **33**, 147 (1980).

[16] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1963), Vol. 1.

[17] A. Fischer, "Unfolding the singularities in superspace," Gen. Relativ. Gravit. **15**, 1191 (1983).

Effective determinism in a classical field theory with spacelike characteristics

J. Isenberg[a)]
Department of Mathematics, University of Oregon, Eugene, Oregon, 97403

G. Horndeski
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario N2L-3G1, Canada

(Received 29 August 1985; accepted for publication 28 October 1985)

For certain classical field theories, it appears as if spacelike characteristic hypersurfaces can occur in the solutions. In principle, this means that acausal propagation is allowed, and the Cauchy problem breaks down. To investigate if indeed this happens, a class of Kasner-like solutions of a generalized Einstein–Maxwell theory is examined. It is found that the spacelike characteristics almost never form, and when they do, they cause no trouble.

I. INTRODUCTION

One of the most useful ways in which to study a given classical field theory[1] is by formulating it as a Cauchy, or initial value, problem. In such a formulation, initial data (possibly satisfying constraint equations) prescribed on a spacelike hypersurface in space-time is evolved (via evolution equations) into a space-time solution of the field equations. This is often a very practical way to obtain physically interesting solutions;[2] and the form of the Cauchy problem for a particular theory (number and type of constraints, number of evolution equations, number of free variables) often reveals much about the nature of that theory.[3]

How does one tell if a given classical field theory admits a well-posed Cauchy problem? While unfortunately there is no canonical "yes–no" test that settles the question for every theory, there are two criteria (one positive and one negative) that have proven to be very valuable. The first criterion, the positive one, is to attempt to show that the field equations are hyperbolic (perhaps in a slightly modified sense if the field theory involves gauge freedom).[4] This method has been used to show that most of the important established theories in physics—e.g., the Maxwell, the Einstein, the Yang–Mills–Higgs (any group), the Dirac, and (in a certain sense) the supergravity[5] theories—do admit Cauchy formulations. But it does not work[6] for some of the theories appearing recently in the physics literature—e.g., some of the $R + R^2$ theories of gravity,[7] certain versions of the skyrmion theory,[8] and possibly some string theories.[9,10]

The other critierion, the negative one, concerns the existence of spacelike characteristic hypersurfaces. For a given field theory, a *characteristic hypersurface* is a three-dimensional submanifold embedded in space-time across which discontinuities in the second derivatives of the solutions can (in principle) develop. The existence of *spacelike* characteristics prohibits a Cauchy formulation in the usual sense because initial data specified on such a hypersurface do not generally determine all the necessary second derivatives of the fields on the hypersurface, and hence evolution cannot be carried out.

Spacelike characteristic hypersurfaces can cause other problems as well. Characteristics, being the loci of second-derivative discontinuities, govern the propagation of wave fronts and shocks (the direction of propagation must be *tangent* to the characteristic hypersurface). Thus if a particular field theory permits spacelike characteristic hypersurfaces, there is (in principle) nothing to stop the acausal propagation of signals.

In view of these consequences, classical field theories that allow spacelike characteristics seem to be very unappealing. We might feel justified in throwing out theories that, like some of the $R + R^2$ theories and like the generalized Einstein–Maxwell (GEM) theory developed by one of us,[11] have this feature. Yet, it is important that we recall that a characteristic hypersurface is one across which second-derivative discontinuities can occur *in principle*. Nothing says that there necessarily *are* solutions that have such discontinuities.

To illustrate this point in a much simplified context, let us consider a particle mechanics problem with the equation of motion

$$M(q, \dot{q})\, \ddot{q}(t) = F(q, \dot{q}) , \tag{1.1}$$

where M and F are a pair of real-valued functions. Here, there are no characteristic hypersurfaces in the usual sense; however, the zeros of M can play a similar role, as we shall see. Now if M is bounded away from zero, then Eq. (1.1) together with a set of initial data $\{q(0), \dot{q}(0)\}$ is sufficient to determine the solution $q(t)$ both into the future and into the past, regardless of the choice of F. If M is identically zero for all values of q and $\dot{q}$, then Eq. (1.1) becomes a constraint. The tricky situation occurs if $M(q, \dot{q})$ is not *identically* zero, yet is zero for certain values of q and $\dot{q}$ (the set of which we label by Q). Then as $q(t)$ evolves, whenever $M(q(t), \dot{q}(t)) \neq 0$, Eq. (1.1) determines the continued evolution. Yet if $M(q,\dot{q}) = 0$, then Eq. (1.1) fails to specify $\ddot{q}$, and there can, in principle, be a jump in $\ddot{q}$ (the analogy to characteristics is thus evident).

Whether, in fact, such discontinuities can occur in otherwise smooth solutions $q(t)$, depends critically upon the properties of $F(q, \dot{q})$ as well as of $M(q, \dot{q})$. For some choices of these functions, the jumps do occur in certain solutions.

[a)] Portions of this research were done while working at the University of Waterloo, Waterloo, Ontario N2L-3G1, Canada.

 0022-2488/86/030739-07$02.50

For other choices of F and M, all solutions with initial data $(q(0), \dot{q}(0)) \notin Q$ are forced by Eq. (1.1) to avoid the set Q while solutions with $(q(0), \dot{q}(0)) \in Q$ stay in Q for all time. Finally, there are choices of the functions F and M for which *most* trajectories avoid Q, but certain ones do not and these evolve smoothly in and out of Q. In none of these three cases can one claim that the initial value problem works in the ordinary sense. Yet in both the second and the third cases, all solutions evolve in an effectively deterministic way.

It would be nice if one could demonstrate that this same phenomenon of "effective determinism" characterizes all of the solutions of some of the field theories that permit spacelike characteristic hypersurfaces. While we have not been able to do this, we *have* been able to show that in the case of the GEM theory,[12] this behavior does occur if one restricts attention to a certain class of cosmological models.

We describe these results in this paper. First, however, we briefly discuss characteristics in general field theories and note how the spacelike ones occur (Sec. II). Then in Sec. III, we describe the GEM theory and recall that it allows spacelike characteristics[13] (timelike ones as well). In the first part of Sec. IV, we impose a number of symmetry conditions on the fields of the GEM theory and thereby vastly simplify the system of field equations (and also vastly restrict the class of solutions). Finally, in the latter part of Sec. IV, we use phase plane analysis to show that these simplified (Kasner-like) models are all effectively deterministic.

II. CHARACTERISTIC HYPERSURFACES

Characteristic hypersurfaces play an important role in the study of systems of partial differential equations, and so accordingly much has been written about them.[14] Here we only make a few remarks, focusing on what allows spacelike characteristics to occur.

We start by considering a scalar field theory on Minkowski space-time background, with the scalar field $\varphi: \mathbf{R}^4 \to \mathbf{R}$ to satisfy the field equation

$$M^{\mu\nu}[\varphi, \nabla\varphi, \eta]\, \nabla_\mu \nabla_\nu\, \varphi = F[\varphi, \nabla\varphi, \eta]\,. \qquad (2.1)$$

Here $M^{\mu\nu}$ and F are both specified functions of φ, of $\nabla\varphi$, and of the Minkowski metric η. (We assume the summation convention.) Now consider a properly embedded submanifold Σ^3 in Minkowski space-time. We wish to rewrite (2.1) in a form that isolates from the rest of the equation the term containing second derivatives normal to Σ^3. In this form, we can identify whether or not Σ^3 can be a characteristic hypersurface, and we also have the appropriate setup for turning (2.1) into an evolution equation. So let us choose local coordinates (x^0, x^1, x^2, x^3) compatible with Σ^3 in the sense that $\{\partial/\partial x^1, \partial/\partial x^2, \partial/\partial x^3\}$ are tangent to Σ^3. We may rewrite (2.1) in the form

$$M^{00}\, \ddot{\varphi} = -2M^{0i}\, \nabla_i\, \dot{\varphi} - M^{ij}\, \nabla_i \nabla_j\, \varphi + F\,, \qquad (2.2)$$

where we use the "·" to denote ∇_0 and where the Latin indices run from 1 to 3 (Greek go from 0 to 3).

Now let us presume that φ and $\dot{\varphi}$ are known on Σ^3 (the derivatives of φ and $\dot{\varphi}$ tangent to Σ^3 are then automatically known as well). If M^{00} is nonvanishing everywhere on Σ^3, then (2.2) determines $\ddot{\varphi}$ everywhere on Σ^3. It follows that there are no discontinuities in $\dot{\varphi}$ across Σ^3, and one has a chance of evolving φ from Σ^3 into a space-time neighborhood of Σ^3. If, however, M^{00} has zeros on Σ^3, then $\ddot{\varphi}$ is not fully determined, discontinuities may occur, and the Cauchy problem (relative to Σ^3) fails. If M^{00} vanishes *everywhere* on Σ^3, then Σ^3 is defined to be a characteristic hypersurface.

The Klein–Gordon field equation, of course, takes the form (2.1). It has $M^{\mu\nu}[\varphi, \nabla\varphi, \eta] = \eta^{\mu\nu}$, and $F[\varphi, \nabla\varphi, \eta] = m^2\varphi$, so we have $M^{00} = \eta^{-1}(dx^0, dx^0)$. This vanishes iff dx^0 is null, and so Σ^3 is characteristic iff it is a null hypersurface. Thus in the Klein–Gordon theory, spacelike characteristics do not occur.

As an example of theory in which they do occur, consider the choice

$$M^{\mu\nu} = \eta^{\mu\nu} - \nabla^\mu \varphi \nabla^\nu \varphi \qquad (2.3)$$

and $F = m^2\varphi$. In this case, $M^{00} = \eta^{-1}(dx^0, dx^0) - \dot{\varphi}\dot{\varphi}$. Thus for *any* choice of Σ^3 (spacelike, null, timelike, or mixed) there exist sets of data $(\varphi, \dot{\varphi})|_{\Sigma^3}$ such that Σ^3 is characteristic.

Most field theories (including the one we shall be focusing on) have more than one field component, so we need a generalization of the notion of characteristic hypersurface for a field theory whose field equations take the form

$$M^{A\mu\nu}_{B}\, [\psi^C, \nabla\psi^C, \eta]\, \nabla_\mu \nabla_\nu \psi^B = E^A\, [\psi^C, \nabla\psi^C, \eta] \qquad (2.4)$$

(where the capital Latin indices range from 1 to n, the total number of field components in the theory). The generalization is straightforward and obvious for those theories with no gauge freedom. In such theories, if we again choose a hypersurface Σ^3 embedded in space-time and work with hypersurface-compatible coordinates (x^0, x^i), we may rewrite (2.4) as

$$M^{A00}_{B}\, \ddot{\psi}^B = -2M^{A0i}_{B}\, \nabla_i \dot{\psi}^B - M^{Aij}_{B}\, \nabla_i \nabla_j \psi^B + E^A\,. \qquad (2.5)$$

Here, if M^{A00}_{B} is an invertible matrix (at each point in Σ^3) then we may solve (2.5) for $\ddot{\psi}^B$ in terms of the data $(\psi^B, \dot{\psi}^B)|_{\Sigma^3}$, and no discontinuities are allowed. If, however, $\det(M^{A00}_{B})$ vanishes anywhere, then (2.5) fails to determine $\ddot{\psi}^B$ and the problems discussed above follow. So in analogy with the definition above, the hypersurface Σ^3 is a characteristic hypersurface iff $\det(M^{A00}_{B}) = 0$ everywhere on Σ^3 [for the given data $(\psi^B, \dot{\psi}^C)|_{\Sigma^3}$].

As in the one-component case, there are simple examples of multicomponent field theories that have no spacelike characteristics—e.g., the vector Klein–Gordon field theory with field equation $\nabla^2\psi^\alpha = m^2\psi^\alpha$—and there are also simple examples that permit spacelike characteristics—e.g., the theory with field equation

$$(\eta^{\mu\nu} - \psi^\mu \psi^\nu)\nabla_\mu \nabla_\nu\, \psi^\alpha = m^2\psi^\alpha\,.$$

We need two more generalizations: We need to consider field theories with gauge freedom, and we need to allow the metric to be a nonfixed, dynamic field. Neither generalization causes much trouble. In the case of theories with gauge freedom (e.g., the Maxwell field) one finds that, regardless of the choice of Σ^3 and regardless of the values of the fields on Σ^3, there are some components of the fields (e.g., A_0) whose second derivatives are not determined by the field equations. This seems at first to be a disaster; however, one also finds that the strange components can invariably be isolated, and one can then base one's analysis of characteristic hypersurfaces and of the initial value problem on the behavior of the other components (freedom of the strange components from determination just reflects the freedom of gauge). As for allowing the metric to be a dynamic field, it *does* follow that the determination of whether Σ^3 is spacelike, timelike, or null becomes dependent on the field data. However, the definition of characteristic hypersurface does not change, and this new twist is harmless.

To illustrate this, and also to provide a familiar standard with which to compare the GEM theory when we discuss it below, we shall now describe the Einstein–Maxwell theory and how its characteristic hypersurfaces behave. The Einstein–Maxwell action principle is given by

$$\mathscr{S}_{\rm EM}(g,A)=\int_{M^4}\left[\sqrt{-g}\left(R-\frac{1}{4}F^{\mu\nu}F_{\mu\nu}\right)\right], \qquad (2.6)$$

where the metric $g_{\mu\nu}$ and the vector potential A_μ are the basic fields, R is the scalar curvature based on the Riemannian connection associated to $g_{\mu\nu}$, and $F_{\mu\nu}$ is the electromagnetic field associated to A_μ. This action is invariant under gauge transformations $A_\mu \to A_\mu + \nabla_\mu \lambda$ and also under space-time diffeomorphisms. Now varying (2.6) with respect to g and A, one obtains the Einstein–Maxwell field equations:

$$G_{\mu\nu}=1/2\,(F_{\mu\alpha}F^{\alpha}_{\ \nu}-1/4\,g_{\mu\nu}F^{\alpha\beta}F_{\alpha\beta}), \qquad (2.7a)$$

$$\nabla_\mu F^{\mu}_{\ \nu}=0. \qquad (2.7b)$$

These may be cast into the form (2.4). Then if we choose a hypersurface Σ^3 along with the appropriate coordinates, we get the following (analogous to 2.5):

$$\begin{pmatrix} Q^{AB} & O \\ 0 & P^{ij}\end{pmatrix}\begin{pmatrix}\ddot g_B \\ \ddot A_j\end{pmatrix}=\begin{pmatrix}C^A\\ D^i\end{pmatrix}. \qquad (2.8)$$

Here, we use the convenient "paired index convention" exemplified by the identity

$$g_B \leftrightarrow g_{ab}:\begin{pmatrix}g_1\\g_2\\g_3\\g_4\\g_5\\g_6\end{pmatrix}\leftrightarrow\begin{pmatrix}g_{11}\\g_{12}\\g_{13}\\g_{22}\\g_{23}\\g_{33}\end{pmatrix}. \qquad (2.9)$$

The matrices Q^{AB} and P^{ij} are defined by

$$Q^{AB}=\begin{pmatrix} q^{11} & 2q^{12} & 2q^{13} & q^{14} & 2q^{15} & q^{16}\\ q^{21} & 2q^{22} & 2q^{23} & q^{24} & 2q^{25} & q^{26}\\ \cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\ \cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\ \cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\ q^{61} & 2q^{62} & 2q^{63} & q^{64} & 2q^{65} & q^{66}\end{pmatrix} \qquad (2.10a)$$

with

$$q^{AB}\leftrightarrow q^{abcd}:=\tfrac{1}{2}\,(g^{a0}g^{b\,(c}g^{d)0}+g^{b0}g^{a\,(c}g^{d)0} - g^{a0}g^{b0}g^{cd}-g^{a\,(c}g^{d)b}g^{00} - g^{ab}g^{0c}g^{0d}+g^{ab}g^{00}g^{cd}) \qquad (2.10b)$$

and

$$p^{ij}:=g^{00}g^{ij}-g^{i0}g^{j0}, \qquad (2.11)$$

while the vectors C^A and D^i are certain functionals of the field data $(A_\mu, g_{\mu\nu}, \dot A_\mu\, \dot g_{\mu\nu})|_{\Sigma^3}$ that we do not need and therefore will not write out explicitly.

Note that in (2.8), the $\ddot g_{00}$ and $\ddot g_{00}$ and the A_0 are missing from the list of second "time" derivatives. These are the components of the gravitional and electromagnetic fields that are left out of the dynamics for *any* Σ^3 and for *any* initial data. They are in a certain sense "pure gauge" (elsewhere[15] we have labeled them "atlas fields") and can be disregarded in the analysis of the characteristics.

For a given set of Einstein–Maxwell field data, Σ^3 is a characteristic hypersurface iff either $\det(Q^{AB})$ or $\det(P^{ij})$ vanishes everywhere on Σ^3. One easily verifies that if Σ^3 is a null hypersurface (which is true iff $g^{00}=0$ and $g^{0i}\neq 0$) then Q^{AB} and P^{ij} both are degenerate. A bit more work shows that, in fact, Q or P is degenerate *only* if Σ^3 is a null hypersurface. It follows that for the Einstein–Maxwell field theory, the characteristic hypersurfaces are necessarily tangent to the local null cones and therefore are never spacelike (with the attendent problems). We note again that in the Einstein–Maxwell theory the notion of spacelike, etc. depends upon the fields. Regardless, we obtain well-defined causal propagation, and the Einstein–Maxwell theory is straightforwardly shown to have a well-posed Cauchy problem.[4]

III. THE GENERALIZED EINSTEIN–MAXWELL FIELD THEORY

The GEM theory[11] may be characterized as the most general classical theory which (1) involves (only) the two fields $g_{\mu\nu}$ (gravitational metric) and A_μ (electromagnetic potential); (2) is derived from an action principle $\mathscr{S}(g,A)$, which is invariant under space-time diffeomorphisms; (3) conserves both charge and energy-momentum (gauge invariance is a consequence of this assumption); (4) has second-order field equations; and (5) becomes the standard Einstein theory if $F_{\mu\nu}=0$ and becomes the standard Maxwell theory if $R^{\mu}{}_{\nu\alpha\beta}=0$. These conditions together imply that the action must take the form[11]

$$\begin{aligned}\mathscr{S}_{\rm GEM}(g,A) &= \int_{M^4}\sqrt{-g}\Big[R-\frac{1}{4})^{\mu\nu}F_{\mu\nu}+\frac{\mu}{2}F_{\alpha\beta}F^{\gamma\partial}R^{\ \alpha\beta *}_{*\ \gamma\partial}\Big]\\ &= \mathscr{S}_{\rm EM}(g,A)+\frac{\mu}{2}\int\sqrt{-g}\,F_{\alpha\beta}F^{\gamma\lambda}R^{\ \alpha\beta *}_{*\ \gamma\lambda}, \end{aligned} \qquad (3.1)$$

where $R^{\ \alpha\beta *}_{*\ \gamma\lambda}=\epsilon^{\alpha\beta\mu\nu}R_{\mu\nu}{}^{\lambda\sigma}\epsilon_{\lambda\sigma\gamma\lambda}$ (a double dual) and μ is a coupling constant.

Varying $\mathscr{S}_{\rm GEM}(g, A)$, we obtain the field equations for this theory:

$$G_{\mu\nu} = \tfrac{1}{2}\,[F_{\mu\alpha}F_{\nu}{}^{\alpha} - \tfrac{1}{4}g_{\mu\nu}F^{\alpha\beta}F_{\alpha\beta}] - \mu[F^{\alpha}{}_{\lambda}F^{\beta\lambda}R^{\ *\ *}_{\mu\alpha\nu\beta} + \nabla^{\beta}F^{*}_{\mu\alpha}\nabla^{\alpha}F^{*}_{\nu\beta}] \tag{3.2a}$$

and

$$\nabla_{\mu}F^{\nu\mu} + (\mu/2)\nabla_{\mu}F_{\beta\gamma}R^{\ \nu\mu\beta\gamma}_{**} = 0\,. \tag{3.2b}$$

The physical consequences of the GEM theory, and the degree to which it is consistent with experiment and observation, are discussed elsewhere.[11] Here, our main concern is with the characteristics and their consequences. So we proceed to carry out the sort of analysis discussed in Sec. II. Since the fields and the gauge freedom are the same as in the Einstein–Maxwell theory, this analysis leads to an equation similar to Eq. (2.8). In particular, we get[13]

$$\begin{pmatrix} \hat{Q}^{AB} & \mu S^{Aj} \\ \mu V^{iB} & \hat{P}^{ij} \end{pmatrix} \begin{pmatrix} \ddot{g}_{B} \\ \ddot{A}_{j} \end{pmatrix} = \begin{pmatrix} \hat{C}^{A} \\ \hat{D}^{i} \end{pmatrix}, \tag{3.3}$$

where

$$\hat{P}^{ij} := P^{ij} + \mu R^{\,0i0j}_{**} \quad [\text{for } P^{ij} \text{ from (2.11)}]\,, \tag{3.4}$$

$$S^{Aj} \leftrightarrow S^{abj} := +\frac{1}{2(\det g)}\,(\epsilon^{0mda}\epsilon^{0jeb} + \epsilon^{0mdb}\epsilon^{0jea})\nabla_{e}F_{md}\,, \tag{3.5}$$

$$\hat{Q}^{AB} = \begin{pmatrix} \hat{q}^{11} & 2\hat{q}^{12} \\ \text{same pattern as (2.10a)} & \end{pmatrix}, \tag{3.6a}$$

with

$$\hat{q}^{AB} \leftrightarrow \hat{q}^{abcd} := q^{abcd} + [\mu/2(\det g)]F_{k\lambda}F_{j}{}^{\lambda}\epsilon^{0ak[c}\epsilon^{d]jb0}\,,$$

$$\text{for } q^{abcd} \text{ from (2.12b)}\,, \tag{3.6b}$$

and

$$V^{iB} = \begin{pmatrix} V^{11} & 2V^{12} & 2V^{13} & V^{14} & 2V^{15} & V^{16} \\ V^{21} & 2V^{22} & 2V^{23} & V^{24} & 2V^{25} & V^{26} \\ V^{31} & 2V^{32} & 2V^{33} & V^{34} & 2V^{35} & V^{36} \end{pmatrix}, \tag{3.7a}$$

with

$$V^{iB} := \tfrac{1}{2}S^{Bi}\,. \tag{3.7b}$$

Again, the expressions for $\hat{C}^{A}$ and $\hat{D}^{i}$ are not needed.

To see whether Σ^3 is a characteristic hypersurface for a given set of field data $(A_i, \dot{A}_i, g_{ij}, \dot{g}_{ij})$, we must calculate

$$\det\begin{pmatrix} \hat{Q} & \mu S \\ \mu V & \hat{P} \end{pmatrix}.$$

This is, unfortunately, a real mess in general. One can show, however, that spacelike (as well as timelike and null) characteristic hypersurfaces can exist. As a simple example, let $M^4 = S^3 \times R$, let $\Sigma^3 = S^3 \times \{0\}$, and pick data

$$g_{ij} = \sigma_{ij}\,, \quad \dot{g}_{ij} = 0\,, \quad A_{\mu} = 0\,, \quad \dot{A}_{m} = (1/x)\varphi_{m}\,, \tag{3.8}$$

where σ_{ij} is the spherical metric of constant positive curvature with $R = -1/(3\mu)$, φ_m is a divergence-free vector harmonic on round S^3, and x is a normalizing constant set to guarantee that $\dot{A}_m\dot{A}_n\sigma^{mn} = (-1/\mu)$. If we take the coupling constant μ to be negative, then it is straightforward to check that, for this data, the matrices S^{Aj} and V^{iB} vanish, and $\det(\hat{Q}^{AB}) = 0$ everywhere on Σ^3. In addition, the fields (3.8) satisfy the constraint equations of Horndeski's theory [obtained by setting $\nu = 0$ in Eq. (3.2)]. Thus we have valid data for which Σ^3 is a spacelike characteristic hypersurface.

Does this mean that the Cauchy problem fails for the GEM theory? We address this question in the next section.

IV. STUDY OF SPACELIKE CHARACTERISTICS IN A SMALL CLASS OF SOLUTIONS

While the analysis of the entire set of spacetime solutions of the generalized Einstein–Maxwell theory is beyond the scope of this (or any other) paper, we can obtain a fairly complete picture of a very small subset of such solutions. This subset—or "minisuperspace" in the parlance of geometrodynamics[16]—is defined by the following conditions:

$$\begin{aligned} &\text{(a) } M^4 = \mathbf{R}^4\,, \\ &\text{(b) isometry group} = \mathbf{R}^3 \times S^1\,, \\ &\text{(c) } A_\mu \text{ is pure electric.} \end{aligned} \tag{4.1}$$

The second of these conditions is the key one that makes this class manageable. It reduces us to the set of locally rotationally symmetric (LRS) Bianchi type I homogeneous space-times. In appropriate coordinates (t, x, y, z) the fields in such space-times depend only upon t and so the field equations are reduced from a system of partial differential equations to a system of ordinary differential equations.

The fields may be written in the form

$$\begin{aligned} g &= -dt^2 + \alpha(t)[dx^2 + dy^2] + \beta(t)dz^2\,, \\ A &= a(t)dz\,. \end{aligned} \tag{4.2}$$

Then if we use a convenient set of conjugate variables $L(t)$, $K(t)$ and $\mathscr{E}(t)$ (these are essentially the canonical momentum variables conjugate to α, β, and a, as defined via the Legendre transformation), the field equations are as follows:

$$L^2 + 2KL = \mathscr{E}^2(1 - 3\mu L^2)\,, \tag{4.3}$$

$$\dot{\alpha} = -2\alpha L\,, \tag{4.4a}$$

$$\dot{\beta} = -2\beta K\,, \tag{4.4b}$$

$$\dot{a} = (1/\sqrt{\alpha})\mathscr{E}\,, \tag{4.4c}$$

$$\dot{L} = \tfrac{3}{2}L^2 - \tfrac{1}{2}(1 - \mu L^2)\mathscr{E}^2\,, \tag{4.5a}$$

$$\dot{K} = K(K + L) + \mathscr{E}^2[\tfrac{3}{2} + \mu KL - L^2(1 + 2\mu) + \mu\mathscr{E}^2] - \mu L^2\mathscr{E}^2(1 + 2\mu L^2)/(1 - \mu L^2)\,, \tag{4.5b}$$

$$(1 - \mu L^2)\dot{\mathscr{E}} = \mathscr{E}L\,[(2 - \mu\mathscr{E}^2) + \mu L^2(1 + \mu\mathscr{E}^2)]\,. \tag{4.5c}$$

Equation (4.3) is the one constraint on the choice of initial data for these space-times. Equations (4.4) and (4.5) specify the evolution.

Where are the spacelike characteristics? If the coupling constant μ is negative, then, in fact, there are none. However, for positive μ, a spacelike hypersurface with specified data $(\alpha, \beta, a, L, K, \mathscr{E})$ is characteristic iff $1 - \mu L^2 = 0$. This is evident in (4.5c), and also in (4.5b).

Before proceeding to analyze the behavior of solutions of the system (4.3)–(4.5), we note that conditions (4.1)—which we have used to reduce the full system of Eq. (3.3) to the simplified (4.3)–(4.5)—are consistent with the full system. That is, if one chooses initial data of the sort that conditions (4.1) demand, and if one then evolves using (3.3), the resulting space-time will satisfy conditions (4.1) for all time. The verification of this is straightforward.[17]

The most useful method, for our purposes, of studying how the solutions of (4.3)–(4.5) behave is via a qualitative trajectory (or "phase") portrait. For a six-dimensional system, which (4.3)–(4.5) appears to be, trajectory portraits are rather unwieldy. However, we can, in fact, reduce our system to one with only two essential dimensions as follows: First, we break the six-dimensional system into a pair of three-dimensional systems, one of which is a slave to the other. The primary system consists of Eq. (4.3) and (4.5); it is independent of α, β, and a and hence may be solved by itself for $L(t)$, $K(t)$, and $\mathscr{E}(t)$. The slave system involves Eq. (4.4); using the fields $L(t)$, $K(t)$, and $\mathscr{E}(t)$, obtained as solutions to the primary system, one solves it for $\alpha(t)$, $\beta(t)$, and $a(t)$. Focusing on the primary system, we may now use the constraint equation (4.3) to reduce it to two dimensions. Specifically, we solve (4.3) for K,

$$K = (1/2L)[\mathscr{E}^2(1 - 3\mu L^2) - L^2]; \tag{4.6}$$

and then as long as we are careful with $L = 0$, we are left with just a two-dimensional unconstrained system with the evolution equations

$$\frac{d}{dt} L = \frac{3}{2} L^2 - \frac{1}{2} (1 - \mu L^2)\mathscr{E}^2 \tag{4.7a}$$

and

$$(1 - \mu L^2) \frac{d}{dt} \mathscr{E} = \mathscr{E} L [(2 - \mu \mathscr{E}^2) + \mu L^2(1 + \mu \mathscr{E}^2)], \tag{4.7b}$$

for the variables L and $\mathscr{E}$.

We have studied (4.7) both qualitatively and numerically, with particular attention given to the behavior of solutions that have (at some time in their history) data that approach the characteristic values—namely $L = \pm 1/\sqrt{\mu}$, $\mathscr{E}$ = anything.[18] The trajectory portrait for this system is sketched in Fig. 1. The most striking feature of this portrait is the role of the locus of characteristics (by which we mean the points in phase plane with $L = \pm 1/\sqrt{\mu}$) as an almost impenetrable fence. Only the "pure Kasner" space-times—those with $\mathscr{E} = 0$—have trajectories that intersect this locus. Other solutions approach arbitrarily close but never reach it. Indeed, one finds that for $L = \pm 1/\sqrt{\mu}$, Eq. (4.7b) becomes a constraint that is *only* satisfied by $\mathscr{E} = 0$.

What this implies is that the only space-times in the class under consideration, which contain spacelike characteristic hypersurfaces, are the pure Kasner space-times; and in these, there are no discontinuities of any sort across those spacelike characteristics.

More can be said about the behavior of the full class of these Bianchi type I, LRS, pure electric space-time solutions of the GEM field equations, as portrayed by the trajectory portrait (Fig. 1). We shall briefly discuss some of this further in the Appendix. But regarding the main question—do spacelike characteristic hypersurfaces develop in solutions, and do they cause trouble—the issue is settled for this small class of space-times. Spacelike characteristics rarely develop; and when they do, they cause no trouble. The Cauchy problem essentially survives.

V. CONCLUSION

The set of space-time solutions of the generalized Einstein–Maxwell theory that we have examined is a very restricted one. And there is no sense in which the GEM theory is "generic" among all those classical field theories that seem

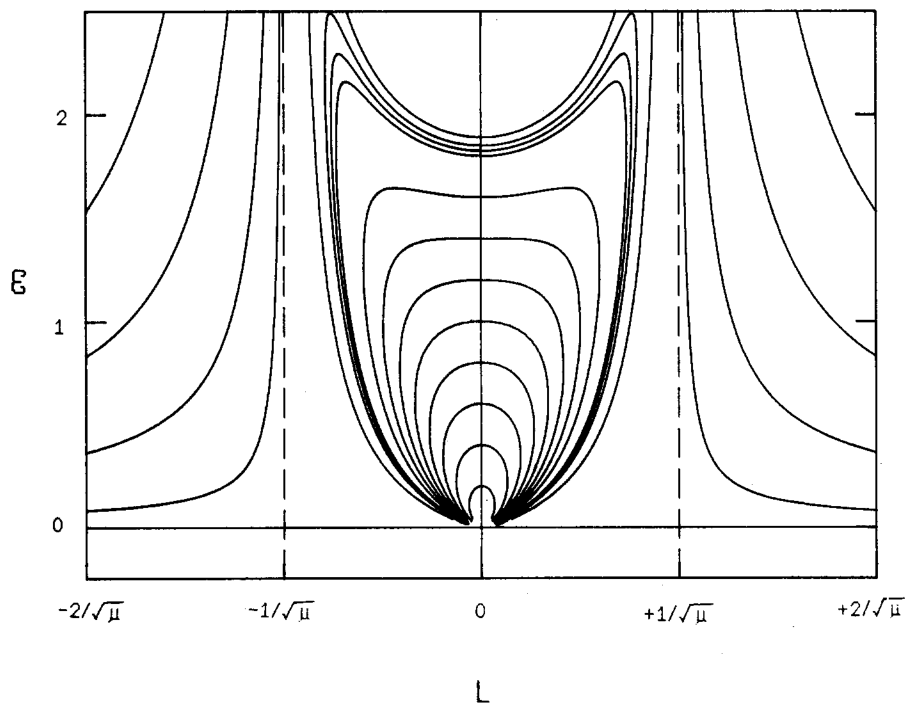


FIG. 1. L-$\mathscr{E}$ trajectory portrait for Bianchi I LRS electric model space-time solutions of the GEM theory. These are representative orbits of the primary variables L and $\mathscr{E}$ for the model solutions of the GEM theory discussed in Sec. IV and in the Appendix. Note the barrierlike behavior of the loci of characteristic data: $L = \pm 1/\sqrt{\mu}$. We have not included any trajectories with $\mathscr{E} < 0$; these are mirror images of the $\mathscr{E} > 0$ trajectories.

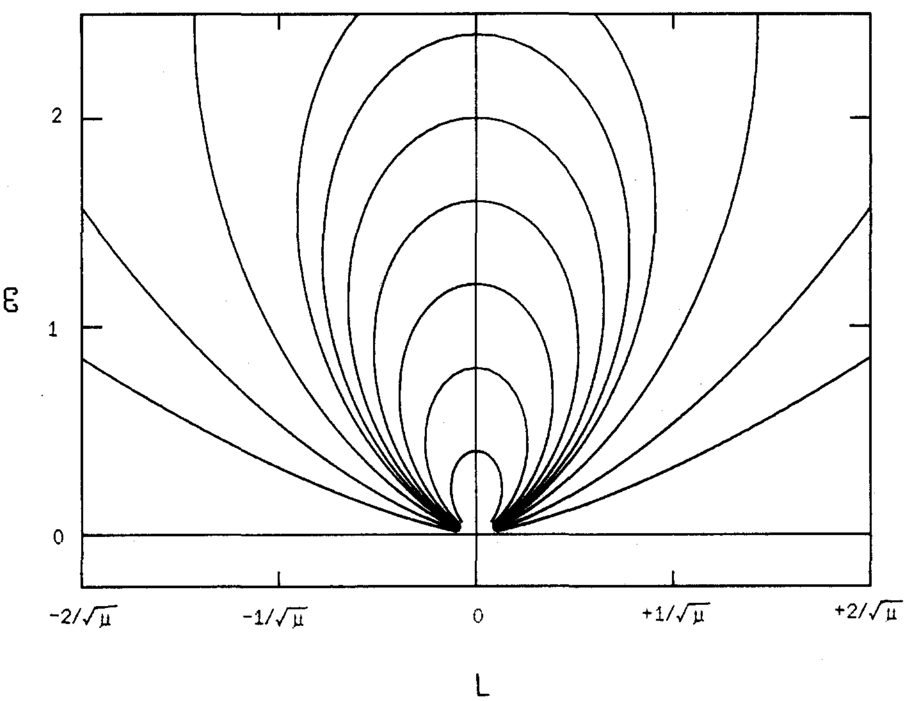


FIG. 2. L-$\mathscr{E}$ trajectory portrait for Bianchi I LRS electric model space-time solutions of the Einstein–Maxwell theory. These are representative orbits of the primary variables L and $\mathscr{E}$ for model solutions of the Einstein–Maxwell theory ($\mu \to 0$ limit of GEM). For small $|L|$ and small $|\mathscr{E}|$, the trajectories are much like those of Fig. 1. For larger $|L|$ and $|\mathscr{E}|$, the character of the trajectories is quite different for the two theories.

to allow spacelike characteristic hypersurfaces. Yet, our results do give some measure of support to the contention that, even though certain field theories admit spacelike characteristics *in principle*, these almost never actually occur in solutions, and when they do occur there are rarely any discontinuities across them. The solutions are then essentially deterministic.

ACKNOWLEDGMENTS

We wish to thank B. Marshman and T. Wagner for help with the numerical work and the figures.

One of us (J.I.) was partly supported by National Science Foundation Grant No. MCS83-03998.

APPENDIX: SOME FEATURES OF THE BIANCHI I LRS SOLUTIONS OF THE GEM FIELD EQUATIONS

Here we mention some of the features of the solutions of (4.7) that, though somewhat peripheral to the main point regarding spacelike characteristics, are still interesting. In this discussion, it is useful to keep in mind how the behavior of K is related to that of L and $\mathscr{E}$ [via Eq. (4.6)], and how the behavior of the metric components α and β is determined by integrating over L and K: from (4.4), we get

$$\alpha(t) = \exp\left(-\int_0^t L\right), \tag{A1a}$$

$$\beta(t) = \exp\left(-\int_0^t K\right). \tag{A1b}$$

First, we recall that the pure Kasner (LRS) space-times (which solve the Einstein equations as the $\mathscr{E} = 0$ limit of the GEM equations) are represented by the trajectories along the $\mathscr{E} = 0$ axis in Fig. 1. The big bang in these cosmological models occurs at large $|L|$; as time proceeds and $|L|$ decreases to zero, the space-time approaches (but never reaches) a large maximal static state.

Other than these Kasner models, there are basically two types of solutions here (ignoring the direction of time). The first kind, those with $|L| > (1\sqrt{\mu})$, are much like the Kasner models in the remote past: They start with a big bang, and then gradually decelerate their expansion (with $|L|$ decreasing). As time proceeds however, $|\mathscr{E}|$ begins to grow exponentially regardless of how small its initial value was. Within a finite time, $\mathscr{E}$ blows up, carrying K along with it, and one reaches a singularity. (Note that for $K \to \infty$, $\beta \to 0$, and so the metric becomes singular.) Hence, space-time solutions of this kind, unlike the Kasner models, have finite lives with singularities on either end.

The other kind of space-time—those with $|L| < (1/\sqrt{\mu})$—are more like Kasner regarding their ultimate fate. They have one big bang type singularity in the past, and approach asymptotically a maximal static state. This is not apparent from Fig. 1, since L is finite throughout the trajectories of this type. Recall, however, that if L approaches zero with $\mathscr{E}$ finite, then K blows up, thereby forcing β to vanish (hence a singularity). So a space-time represented by one of these trajectories must terminate (in the past) at the intersection of its trajectory with the $L = 0$ axis (for $\mathscr{E} \neq 0$)

Note that the Kasner solutions in this system are unstable to perturbations in $\mathscr{E}$. That is, given any set of Kasner data with $L \neq 0$, and $L \neq \pm 1/\sqrt{\mu}$, if one perturbs $\mathscr{E}$ then the resulting space-time will have large $\mathscr{E}$ either in the past or in the future.

It is interesting to compare the solutions just discussed with the analogous space-time solutions of the standard Einstein–Maxwell equations. The trajectory portrait for these is sketched in Fig. 2. We see that in many respects, these space-times are much like the $|L| < (1/\sqrt{\mu})$ solutions of the GEM theory.

[1]For our purposes here, we shall adopt the loose definition that a classical field theory is simply a system of partial differential equations (or "field equations") for some functions (or "fields") on space-time.

[2]Some discussion of the practical application of this method of constructing space-times appears in L. Smarr, *Sources of Gravitational Radiation* (Cambridge U.P., London, 1979).

[3]See, e.g., J. Isenberg and J. Nester, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 1.

[4]Y. Choquet-Bruhat, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).

[5]D. Bao, Y. Choquet-Burhat, J. Isenberg, and P. Yasskin, J. Math. Phys. **26**, 329 (1985).

[6]One's failure to demonstrate hyperbolicity does *not* imply the nonexistence of a well-posed Cauchy problem.

[7]Theories of the sort discussed by E. Fairchild, Phys. Rev. D **16**, 2438 (1977) have this problem, as is shown in unpublished work of J. Isenberg and P. Yasskin.

[8]T. Skyrme, J. Math. Phys. **12**, 1735 (1971).

[9]See, e.g., L. Brink, "Superstrings," preprint, Institute of Theoretical Physics, Göteberg, Sweden, 1984.

[10]Since most of these new theories have been motivated by quantum considerations, some may treat the existence or nonexistence of a well-posed classical Cauchy problem as an irrelevant issue. However, we believe that one should be a bit suspicious of a quantum field theory whose classical version admits no Cauchy formulation.

[11]G. Horndeski, J. Math. Phys. **17**, 1980 (1976); G. Horndeski, Phys. Rev. D **16**, 1691 (1977).

[12]We have studied this theory rather than some of the other more popular ones because it is considerably simpler.

[13]In J. Math. Phys. **20**, 1745 (1979), Horndeski first showed that spacelike characteristics can, in principle, occur in solutions of his theory.

[14]See, e.g., R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Wiley, New York, 1962), Vol. II.

[15]J. Isenberg, Ann. Phys. **129**, 223 (1980).

[16]C. Misner, in *Magic Without Magic*, edited by J. Klauder (Freeman, San Francisco, 1973).

[17]J. Isenberg and G. Horndeski (unpublished).

[18]By "characteristic data" we mean data such that the hypersurface Σ^3 with that data is a characteristic hypersurface.

The structure of the space of homogeneous Yang–Mills fields

Moshe Molelekoa
National University of Lesotho, P.O. Roma 180, Lesotho, Africa

(Received 26 July 1985; accepted for publication 28 October 1985)

The structure of the space of solutions of Yang–Mills equations is examined for solutions that are required to have a specified set of infinitesimal space-time symmetries. It is shown that when the set consists of Killing vector fields, which are tangent to a compact spacelike Cauchy surface, the space is a smooth ILH manifold near each solution that has only trivial gauge symmetries.

I. INTRODUCTION

Arms[1] and Moncrief[2] have shown that when a space-time manifold has a compact spacelike Cauchy surface M, the Yang–Mills (YM) equations can be split up into evolution and constraint equations on M. By considering infinitesimal isometries, we give necessary and sufficient conditions for a YM field to be homogeneous (see Sec. IV for definition) relative to a given Lie algebra of Killing vector fields, in terms of the Cauchy data on M.

Furthermore, Arms[3] examined the structure of the space of solutions of the YM equations for such space-times. She has shown that this space is a smooth (ILH) manifold near each field that has only trivial gauge symmetries. In this paper, it is shown that this result also holds when we restrict ourselves to those fields that are required to be homogeneous relative to a set of Killing vector fields, which are tangent to the subspace M.

The constructions and proofs use standard results about elliptic operators. These operators are defined on spaces of tensor fields, which are defined on the compact manifold M. The fact that the space of smooth gauge fields on M is an affine ILH space (an inverse limit of Hilbert spaces, as defined by Omori[4]) easily follows from the work of Kondracki and Rogulski.[5] The rest of the material uses this fact to establish the main result.

Notations and general definitions are given in Sec. II. Further notation is introduced in Sec. III, where the work of Arms on the structure of the space of YM fields is reviewed. In Sec. IV, symmetry of gauge fields[6,7] is reviewed and some results of Sec. III are extended to the space of homogeneous gauge fields. Only the infinitesimial space-time symmetries are considered.

In Sec. V, a symmetry of a YM field is expressed in terms of the Cauchy data on a compact spacelike Cauchy surface of a four-dimensional space-time. This is followed in Sec. VI by the extension of the main results of Arms to the space of homogeneous YM fields. Homogeneity here is with respect to Killing vector fields, which are tangent to the Cauchy surface. The concluding remarks in Sec. VII discuss the corresponding results when the vector fields are not all tangent to the Cauchy surface. A primary example in this respect is a space-time that is static.

II. NOTATIONS AND DEFINITIONS

Let G be an n-dimensional Lie group with Lie algebra $\mathfrak{g}$. Let M be a (real, paracompact) smooth manifold of dimension m; $\pi: P\to M$, a principal G-bundle; ω, a connection form on P; and Ω, the corresponding curvature form.

By an automorphism of P, we mean one that is also a bundle map. By a tensorial form on P, we shall mean one that is of type Ad G (cf. Kobayashi and Nomizu[8]). A tensorial form on M is a pullback of one on P by a (local) cross section.

The connection form ω represents a gauge field on M, of type P, with gauge group G (the internal symmetry group of the field).

A gauge (on M, of type P) is simply a trivialization of the principal bundle. Relative to a gauge, the connection form ω on P can be expressed uniquely by a family of forms $\{A_\alpha\}$ each defined in an open subset U_α of M (see Ref. 8, p. 66). A choice of gauge implies the use of the special cross sections σ_α: $U_\alpha\to P$, which result from the trivialization. The local field A_α is related to ω by $A_\alpha = \sigma_\alpha^*\omega$. In the sequel we drop the subscript α and write σ, A, and U.

The $\mathfrak{g}$-valued one-form A is usually referred to as a (local) gauge potential of the field, and $F=\sigma^*\Omega$ is the corresponding field strength.

Let $\mathscr{A}$ denote the collection of all gauge fields on M, of type P. Each element of $\mathscr{A}$ will be considered (relative to some fixed gauge) as being defined on M as described above and will be expressed by a representative A of the collection $\{A_\alpha\}$.

Similarly, each tensorial form on P, relative to a gauge, may be represented on M by a collection of locally defined $\mathfrak{g}$-valued forms. The set of tensorial r-forms $\mathscr{D}^r$, will therefore consist of those $\mathfrak{g}$-valued r-forms on M that are pullbacks of tensorial r-forms on P. For example, the field strength F is an element of $\mathscr{D}^2$. The set $\mathscr{D}^r$ is a real vector space of infinite dimensions.

Suppose that $\{X_1,X_2,...,X_n\}$ is a basis for $\mathfrak{g}$. Then, any $\mathfrak{g}$-valued r-form ψ, say, may be written uniquely as the sum $\psi^a X_a$ (summed over the a's from 1 to n), where each ψ^a is a real-valued r-form.

Note: The summation convention will be used throughout.

If ϕ is another $\mathfrak{g}$-valued s-form defined on the same space as ψ, then the $(r+s)$-form denoted by $b_\phi(\psi)$ and defined by (the bracket)

$$b_\phi(\psi) = [\phi,\psi] = (\phi^a\wedge\psi^b)[X_a,X_b] \qquad (2.1)$$

is independent of the basis chosen.

 0022-2488/86/030746-09$02.50

As examples, of the use of this bracket, we have the curvature form

$$\Omega = d\omega + \tfrac{1}{2}b_\omega(\omega)\,,$$

and

$$D\phi = d\phi + b_\omega(\phi)$$

is the covariant derivative of any tensorial form ϕ relative to the connection form ω.

In Secs. V and VI we use the notation in Ref. 1 for a space-time manifold as follows.

Let ${}^4S$ denote a four-dimensional space-time manifold that has a compact spacelike Cauchy surface M; ${}^4g$, the metric on ${}^4S$; and g, the restriction of ${}^4g$ on M.

The metric ${}^4g$ has signature $(-\ +\ +\ +)$ and so g is positive definite.

In general, tensors on ${}^4S$ have a preceding superscript 4, while tensors on M do not.

For points of ${}^4S$ that are on M, we shall often use Gaussian normal coordinates (GNC) (x^0,x^1,x^2,x^3), where $x^0 = t$ and $t = 0$ on M.

Lowercase Greek indices range from 0 to 3 and are lowered or raised by ${}^4g$; Latin indices from the middle of the alphabet range from 1 to 3 and are manipulated by g.

Let $\mathfrak{z}$ denote the future pointing unit normal to M. When we use the GNC, the resulting frame of vector fields $\partial_\alpha = \partial/\partial x^\alpha$ is such that

$$\partial_0 = \partial_t = \mathfrak{z}\,.$$

III. THE SPACE OF GAUGE FIELDS

Several results on the space $\mathscr{A}$ of connections on a principal fiber bundle over a compact Riemannian manifold have been announced since Singer[9] published his paper in 1978. In particular, the action of the group $\mathscr{G}$ of gauge transformations on $\mathscr{A}$ has been shown to have a slice (see, e.g., Refs. 5 and 10). More recently, Arms has pointed out a corresponding slice for the action of $\mathscr{G}$ on the cotangent bundle over $\mathscr{A}$ from a more general result concerning momentum mappings.[11]

The latter result will be stated in this section in terms of the action of $\mathscr{G}$ on the tangent bundle $\mathscr{T}\mathscr{A}$ over $\mathscr{A}$. It will then be shown that with an appropriate definition of space-time symmetry for elements of $\mathscr{T}\mathscr{A}$, the action of the more general group of automorphisms of the principal bundle covering a given group of isometries, has a slice. In fact this slice may be chosen to coincide with the slice used in Refs. 3 and 11 above. This will be useful in proving the main results.

Let M be a compact oriented Riemannian manifold of dimension m, and G a Lie group of dimension n. Suppose that the Lie algebra $\mathfrak{g}$ of G has a positive definite real inner product γ, which is invariant under the adjoint action of G on $\mathfrak{g}$. Let P be a principal G-bundle over M. We introduce the usual (see, e.g., Refs. 3 and 12) inner product $\langle , \rangle$ on the vector space $\mathscr{D}^r$ (and extend it componentwise to $\mathscr{D}^r \times \mathscr{D}^s$) as follows.

If B and E are tensorial r-forms on M, then

$$\langle B,E\rangle = \int_M \text{trace}\,(B\wedge *E)$$

where $*$ is the Hodge star operator and the trace is relative to γ. In particular, relative to any basis $\{X_1,X_2,\ldots,X_n\}$ of $\mathfrak{g}$,

$$\text{trace}\,(B\wedge *E) = \gamma_{cd}B^c\wedge(*E^d)\,,$$

where $\gamma_{cd} = \gamma(X_c,X_d)$.

It is easy to show that $\langle , \rangle$ is well defined if we recall that the values of tensorial forms on M at each point are uniquely determined up to the action of Ad G, and γ is invariant under this action.

Denote by $\mathscr{A}$, the collection of all smooth gauge fields over M of type P, as described in Sec. II. Since any two connection forms differ by a tensorial one-form, the space $\mathscr{A}$ is an affine space modeled on the vector space $\mathscr{D}^1$.

Let $\mathscr{A}_k$ denote the collection of all gauge fields on M of Sobolev class H^k, i.e., the components of each element of $\mathscr{A}_k$, in local coordinates, are locally H^k-functions. Then $\mathscr{A}_k$ carries the natural structure of an affine space modeled on the Hilbert space $\mathscr{D}^1_k$ of tensorial one-forms of Sobolev class H^k (see Ref. 5).

It follows from the continuous inclusion $\mathscr{D}^1_{k+1} \subseteq \mathscr{D}^1_k$ and the Sobolev lemma, that $\mathscr{D}^1$ can be identified with the intersection of all the Hilbert spaces $\mathscr{D}^1_k$ for $k > m/2 + 1$. Then $\mathscr{D}^1$, as an inverse limit of these Hilbert spaces (an ILH space), is a smooth ILH manifold, as defined by Omori.[4] The affine space $\mathscr{A}$ is therefore also a smooth ILH manifold.

Note: All smooth ILH manifolds will be referred to simply as smooth manifolds.

Another example of a smooth manifold is $\mathscr{D}^r$ for $0 \leqslant r \leqslant m$. This follows from considerations similar to those for $r = 1$.

We now review part of the work of Arms[3] concerning the space of gauge fields, in order to establish notation and the slice theorem.

Since the smooth manifold $\mathscr{A}$ is an affine space, we can identify its tangent bundle $\mathscr{T}\mathscr{A}$ with $\mathscr{A}\times\mathscr{D}^1$; so

$$\mathscr{T}\mathscr{A} = \mathscr{A}\times\mathscr{D}^1\,.$$

Now, $\langle\ ,\ \rangle$ is a (weak) Riemannian metric on $\mathscr{A}$ and hence also on $\mathscr{T}\mathscr{A}$.

Let K: $\mathscr{T}\mathscr{A}\to\mathscr{D}^0$ be defined by

$$K(A,E) = -D^*E\,,$$

where D^* is the adjoint of the operator D relative to the metric $\langle\ ,\ \rangle$, and D is covariant differentiation relative to A.

Note: All adjoints of operators will be relative to the inner product $\langle\ ,\ \rangle$. Formulas for these adjoints are calculated in the usual way of integrating by parts. They will differ from those of Arms because she uses the cotangent bundle $\mathscr{T}^*\mathscr{A}$ (the natural phase-space) and tensor densities instead of our tangent bundle and tensor fields.

It is easy to show that K is a smooth map whose derivative at (A,E) is given by

$$K'(a,e) = -D^*e - b^*_a E\,, \tag{3.1}$$

where b_a is defined by Eq. (2.1) and the tangent space is identified with $\mathscr{D}^1\times\mathscr{D}^1$. Moreover, its adjoint is given by

$$K'^*(V) = (b_E V, -DV) \tag{3.2}$$

where V is a tensorial function.

The operator K'^* is elliptic because the highest order derivatives in $K'^*(\psi)$ are given by $-d\psi$ of the second component, and d is an elliptic operator on smooth functions.

Note: Ellipticity here is in the sense of Douglis and Nirenberg as extended by Hormander.[13]

Since K'^* is elliptic, we have the orthogonal (direct sum) splittings (relative to $\langle\, ,\,\rangle$)

$$\mathscr{D}^1\times\mathscr{D}^1 = \operatorname{Im} K'^* \oplus \operatorname{Ker} K'$$

and

$$\mathscr{D}^1\times\mathscr{D}^1 = \operatorname{Im} J\circ K'^* \oplus \operatorname{Ker} K'\circ J, \tag{3.3}$$

where $J(a,e) = (e, -a)$.

Remark: Equation (3.3) is true if we replace $\mathscr{D}^1$ by the Hilbert space $\mathscr{D}^1_k$ for suitably large k. Then, by applying the regularity lemmas (available because of the ellipticity of the operator), the result extends to $\mathscr{D}^1$.

Let $\mathscr{G}$ denote the group of all gauge transformations, i.e., all automorphisms of P that cover the identity map of M. Then $\mathscr{G}$ acts on $\mathscr{A}$ and $\mathscr{D}^1$ (and hence on their Cartesian product) in the usual way, viz. by pullbacks. From the results in Arms, we deduce that for each (A,E) in $\mathscr{T}\mathscr{A}$, the orbit $\mathscr{G}(A,E)$ of (A,E) under $\mathscr{G}$ is a smooth manifold. Its tangent space at (A,E) is precisely $\operatorname{Im} J\circ K'^*$, the first factor in Eq. (3.3). This space is a closed subspace of $\mathscr{D}^1\times\mathscr{D}^1$ because of the ellipticity of $J\circ K'^*$.

The action of $\mathscr{G}$ on $\mathscr{T}\mathscr{A}$ has a (smooth) slice $\mathscr{S}$ at (A,E), which may be given by an open ball

$$\mathscr{S} = \{(A+a, E+e)\colon\ (a,e)\in \operatorname{Ker} K'\circ J,\ \rho((a,e),(0,0))<\delta\} \tag{3.4}$$

for some $\delta>0$ and (strong) metric ρ, which may be chosen to be invariant under gauge transformations and the isometries of M by using the inner product $\langle\, ,\,\rangle$. This result implies the following: Let f be any gauge transformation. (1) If $f(A,E) = (A,E)$, then $f(\mathscr{S})\subseteq\mathscr{S}$. (2) If $f(\mathscr{S})\cap\mathscr{S}$ is not empty, then $f(A,E) = (A,E)$.

Moreover, $\mathscr{T}\mathscr{A}$ is locally diffeomorphic to $\mathscr{G}(A,E)\times\mathscr{S}$ near (A,E).

IV. SPACE OF HOMOGENEOUS GAUGE FIELDS

We now consider homogeneous gauge fields. Although our main results concern symmetries relative to Killing vector fields, we also use isometries in establishing the results.

Let $\pi\colon P\to M$, G and ω be as in Sec. II. By an automorphism of ω is meant any automorphism of P under which ω is invariant.

Let h be any transformation of M. The gauge field on M, defined by the connection form ω on P, is said to have h as a symmetry if there is an automorphism $L(h)$ of ω that covers h.

Suppose that H is a Lie group of transformations of M. The gauge field is said to be homogeneous[6,7] relative to H if each h in H is a symmetry of ω. In such a case there is a map

$$L\colon\ H\times P\to P$$

such that

$$L(h,p) = L(h)(p)\,,$$

where $L(h)$ is an automorphism of ω covering h.

Remarks: In general, homogeneity does not imply that the connection is invariant under some action of H on P covering the given action on M. Moreover, the above definition (of homogeneity) for symmetry may be more appropriate in certain important physical examples (see Refs. 6 and 7). However, if the map L is smooth and H is a one-parameter group of transformations, then homogeneity is equivalent to the invariance of the connection under an action of H on P covering the given action (cf. Proposition 3.2 of Ref. 7). This result will be useful later.

Now let H be a Lie group of isometries of M. Suppose that $L(H)$ is the group of all automorphisms of P that cover the elements of H. Then the action of $L(H)$ on P induces an action on the set of connection and tensorial forms. This action on $\mathscr{A}$ and $\mathscr{D}^1$ is given (locally, for a fixed gauge) by

$$L(h)(A) = \sigma^*(L(h)^*\omega) \tag{4.1}$$

and

$$L(h)(E) = \sigma^*(L(h)^*\phi)\,, \tag{4.2}$$

where $L(h)$ covers h, σ is an appropriate (local) cross section of P (see Sec. II), and A and E are elements of $\mathscr{A}$ and $\mathscr{D}^1$ corresponding to ω and ϕ, respectively. The action extends to Cartesian products componentwise. For example, on $\mathscr{T}\mathscr{A}$ the action is then given by

$$L(h)(A,E) = (\sigma^*L(h)^*\omega, \sigma^*L(h)^*\phi)\,. \tag{4.3}$$

Proposition 4.1: Any automorphism of P that induces an isometry on M is also an isometry of $\mathscr{D}^r$.

Proof: Suppose $L(h)$ is an automorphism of P that covers the isometry h of M.

Let

$$L(h)(\sigma_1(x)) = \sigma_2(h(x))\cdot[T_{L(h)}(x)]^{-1}\,, \tag{4.4}$$

for some local cross sections σ_1 and σ_2, with $T_{L(h)}(x)$ an element of G. Then, for any tensorial r-form ϕ on P,

$$\sigma_1^*L(h)^*\phi = \operatorname{Ad} T_{L(h)}h^*(\sigma_2^*\phi)\,. \tag{4.5}$$

Note that, by our convention, $\sigma_1^*\phi$ and $\sigma_2^*\phi$ are represented by the same symbol on M because they represent the same tensorial form on P. Suppose then that a and e are tensorial r-forms on M. We calculate the inner product

$$\begin{aligned}\langle L(h)a, L(h)e\rangle &= \langle \operatorname{Ad} T_{L(h)}h^*a, \operatorname{Ad} T_{L(h)}h^*e\rangle\\ &= \langle h^*a, h^*e\rangle \quad \text{(from invariance of } \gamma)\\ &= \langle a,e\rangle\,.\end{aligned}$$

The last equation holds because h is an isometry (cf. definition of $\langle\, ,\,\rangle$). Hence, $L(h)$ is an isometry of $\mathscr{D}^r$.

Definition: Let h be any diffeomorphism of M. We shall say h is a symmetry of $(A,E)\in\mathscr{T}\mathscr{A}$ if

$$L(h)(A,E) = (A,E)$$

[cf. Eq. (4.3)] for some automorphism $L(h)$ of the connection, which covers h.

Proposition 4.2: Suppose that h is an isometry of M and a symmetry of $(A,E)\in\mathscr{T}\mathscr{A}$. Let $\mathscr{S}$ be the set defined in Eq. (3.4). Then, for any automorphism $L(h)$ of P, which covers h, the following hold.

(1) If $L(h)(A,E) = (A,E)$, then $L(h)(\mathscr{S})\subseteq\mathscr{S}$.

(2) If $L(h)(\mathscr{S})\cap\mathscr{S}$ is not empty, then $L(h)(A,E) = (A,E)$.

To prove this proposition, we need the following lemma.

Lemma 4.1: Suppose that $(A,E) \in \mathscr{T}\mathscr{A}$ and $L(h)$ is an automorphism of P that covers h and satisfies $L(h)(A,E) = (A,E)$. Then $L(h)$ is an automorphism of the vector subspace Ker $K' \circ J$ of $\mathscr{D}^1 \times \mathscr{D}^1$.

Proof: Let f be any gauge transformation. Then

$$L(h)(f(A,E)) = L(h)(f(L(h)^{-1}(A,E)))$$
$$= (L(h)^{-1} f L(h))(A,E) .$$

Since $L(h)^{-1} f L(h)$ is a gauge transformation, we conclude that $L(h)$ maps $\mathscr{G}(A,E)$ into itself.

Now, $L(h)(A,E) = (A,E)$ implies that $L(h)$ maps the tangent space Im $J \circ K'^*$ of $\mathscr{G}(A,E)$ at (A,E) into itself.

Since $L(h)$ is an isometry of $\mathscr{D}^1 \times \mathscr{D}^1$, we conclude from Eq. (3.3) that $L(h)$ maps Ker $K' \circ J$ into itself.

Proof of Proposition 4.2: Suppose $L(h)(A,E) = (A,E)$ and $(A + a, E + e)$ lies in $\mathscr{S}$. From the definition it follows that

$$L(h)\ (A + a, E + e) = L(h)(A,E) + L(h)(a,e)$$
$$= (A,E) + L(h)(a,e)$$
$$= (A,E) + (a_1,e_1) ,$$

where (a_1,e_1) lies in Ker $K' \circ J$, by Lemma 4.1.

The metric ρ used to define the ball $\mathscr{S}$, is invariant under $\mathscr{G}$ and the isometries of M. So ρ is invariant under $L(h)$ (cf. proof of Proposition 4.1). This implies that $L(h)(\mathscr{S}) \subseteq \mathscr{S}$ and so (1) is proved.

To prove (2), suppose that $L(h)(\mathscr{S}) \wedge \mathscr{S}$ is nonempty. Since (A,E) has symmetry h, there is a lift, $L_1(h)$ say, of h such that

$$L_1(h)(A,E) = (A,E) .$$

Since $L_1(h)$ covers h, there is a gauge transformation f such that $L(h) = L_1(h) \circ f$.

Since $L(h)(\mathscr{S}) \cap \mathscr{S}$ is nonempty, there exists (A_1,E_1) and (A_2,E_2) in $\mathscr{S}$ such that $L(h)(A_1,E_1) = (A_2,E_2)$. But,

$$L(h)(A_1,E_1) = f(L_1(h)(A_1,E_1)) = f(A_3,E_3)$$

for some (A_3,E_3) in $\mathscr{S}$. The latter follows from (1) with L_1 instead of L.

Hence $f(\mathscr{S}) \cap \mathscr{S}$ is nonempty. Since $\mathscr{S}$ is a slice for the action of $\mathscr{G}$ on $\mathscr{T}\mathscr{A}$ at (A,E), this implies that $f(A,E) = (A,E)$. So,

$$L(h)(A,E) = L_1(h)(A,E) = (A,E) . \qquad \blacksquare$$

Let us now examine the question of infinitesimal isometries. These are the kinds of symmetries that will be considered in the main results.

We recall from Refs. 6 and 7, that a vector field X on M is a symmetry of the connection form ω if it lifts to a G-invariant vector field $\bar{X}$ on P such that

$$L_{\bar{X}}\omega = 0 . \tag{4.6}$$

We also recall that X is a symmetry of ω if and only if ω is invariant relative to the flow of X, i.e., if and only if the flow of X lifts to the flow of $\bar{X}$ which fixes the connection. Furthermore, X is a symmetry of ω if and only if

$$L_X A = DW = dW + [A,W] \tag{4.7}$$

holds (locally) on M for some $\mathfrak{g}$-valued function

$$W = A(X) + V , \tag{4.8}$$

where V is a tensorial function and A represents ω on M.

The map W is called a symmetry function of A relative to X. It may be found as follows.

Suppose that $\{L(h_t)\}$ is the flow of the G-invariant vector field $\bar{X}$ on P [see Eq. (4.6)] covering the flow $\{h_t\}$ of X.

For values of t near 0, we may use the same cross section σ in Eq. (4.4) and write

$$L(h_t)(\sigma(x)) = \sigma(h_t(x))T_t(x)^{-1} . \tag{4.9}$$

The function T_t is referred to as a transformation function of A relative to the symmetry h_t.

Differentiation at $t = 0$ yields the required function

$$W(x) = \frac{d}{dt}\, T_t(x) \quad \text{(evaluated at } t = 0\text{)} . \tag{4.10}$$

An equivalent form of Eq. (4.7) is

$$i_X F = DV , \tag{4.11}$$

where V is defined in Eq. (4.8) and F is the field strength. Thus X is a symmetry of A if and only if Eq. (4.11) holds for some tensorial function V.

Let $\mathscr{h}$ be a Lie algebra of vector fields on M. A gauge field with connection form ω, is said to be homogeneous relative to $\mathscr{h}$ if each element of the Lie algebra is a symmetry of ω. In this case, each element of the connected Lie group H generated by $\mathscr{h}$ is a symmetry of ω. Hence ω is homogeneous relative to H.

V. SPACE-TIME SYMMETRIES OF YANG–MILLS FIELDS

In this section we discuss the symmetric properties of Yang–Mills (YM) fields relative to infinitesimal isometries. We express a symmetry in terms of the initial data on a Cauchy surface for the YM fields and the isometries.

Some useful results from Moncrief[2] and Arms[3] will be stated in terms of our notation. Specifically, the YM equation splits into evolution and constraint equations of the initial data on a compact spacelike Cauchy surface. A necessary and sufficient condition for a Killing vector field to be a symmetry of a YM field is given by Proposition 5.1 in terms of the initial state.

We use the notation as described in Sec. II for the four-dimensional space-time manifold ${}^4S$, which has a compact spacelike Cauchy surface M. The Lie algebra $\mathfrak{g}$ of the gauge group G will be assumed to have a positive definite real inner product γ, which is invariant under the adjoint action of G on $\mathfrak{g}$.

Let ${}^4A$ be a gauge field on ${}^4S$ and ${}^4F$, the corresponding field strength. The gauge field defines a YM field if the YM equation

$${}^4D^*({}^4F) = 0$$

is satisfied. Here ${}^4D^*$ is formally $*{}^4D*$, where the latter stars are the Hodge star operators, and ${}^4D$ is covariant differentiation with respect to ${}^4A$.

The linearized YM equation at a solution ${}^4A$ (of the YM equation) is

$${}^4D^*\,{}^4D({}^4a) + b^*_{({}^4a)}({}^4F) = 0 . \tag{5.1}$$

We shall often work with the Gaussian normal coordinates (GNC) as described in Sec. II. Relative to the GNC, the metric ${}^4g$ of ${}^4S$ has

$$ {}^4g_{00} = -1 \text{ and } {}^4g_{0i} = 0, \quad \text{for } i = 1,2, \text{ and } 3 . $$

We define a generalized electric field E on M to be the restriction to M of the one-form $-i_{\not{z}}({}^4F)$.

Note: This electric field differs from that defined in Arms where it is a vector density. The latter results from defining E as the momentum conjugate to the restriction A of ${}^4A$ to M, in the Hamiltonian formalism. The resulting Hamiltonian equations then give the evolution and constraint equations.

In our notation, the resulting equations are given as follows.

The constraint equation is given by

$$ K(A,E) = 0 , \tag{5.2} $$

where K is defined in Sec. III. The evolution equation for E is given by

$$ L_{\not{z}} E = D^*F + b_C(E) + \Gamma E , \tag{5.3} $$

where C is the restriction of $-i_{\not{z}}({}^4A)$ to M, and

$$ \Gamma E = 2i_{E^*}(\not{k}) - \text{trace}(\not{k})E , $$

where E^* is the $\mathfrak{g}$-valued vector field associated with E via the metric g on M, and $\not{k}$ is the second fundamental form on M. In GNC, ΓE has

$$ (\Gamma E)_i = 2\Gamma^k_{0i}E_k - \Gamma^k_{0k}E_i , $$

where Γ^k_{0i} are the Christoffel symbols for ${}^4g$.

The result of the Cauchy problem for YM fields then is that for each tensorial function C, the evolution equation defines a unique YM ${}^4A$ on ${}^4S$ such that $K(A,E) = 0$ on M. A choice of C defines a choice of gauge. So, a solution of $K(A,E) = 0$ defines the YM field ${}^4A$ uniquely up to this choice of gauge.

Let ${}^4X$ be a Killing vector field on ${}^4S$ and let $N\not{z} + X$ be its restriction on M, with X tangent to M. Define a map Φ: $\mathscr{T}\mathscr{A} \to \mathscr{D}^1 \times \mathscr{D}^1$ by

$$ \Phi(A,E) = (NE - i_XF, -L_XE - N(\Gamma E) - [A(X),E] - D^*(NF)) , $$

where ΓE is as defined before.

Proposition 5.1: The Killing vector field ${}^4X$ is a symmetry of the YM field ${}^4A$ if and only if the corresponding (A,E) in Ker K satisfies $\Phi(A,E) \in \text{Im } J \circ K'^*$, i.e.,

$$ (1) \quad -NE + i_XF = DV $$

and

$$ (2) \quad L_XE + N(\Gamma E) + [A(X),E] + D^*(NF) = [E,V] $$

for some tensorial function V on M.

Note: Proposition 5.1 could also be expressed by saying that ${}^4X$ is a symmetry of ${}^4A$ if and only if the vector field Φ on $\mathscr{T}\mathscr{A}$ at (A,E) is tangent to the orbit of (A,E) under $\mathscr{G}$.

To prove the proposition, we need the following results.

Lemma 5.1: Suppose ${}^4X$ is given by ${}^4X^\alpha \partial_\alpha$ in GNC, then ${}^4X^0$ is independent of t, whereas

$$ \partial_t({}^4X^i) = {}^4g^{ik}(\partial_kN) . $$

Lemma 5.2: When restricted to M, the tensorial one-form $i_{\not{z}}\, {}^4D(i_{({}^4X)}{}^4F)$ becomes $-D^*(NF) - L_XE - N(\Gamma E) - [A(X),E]$, where ΓE is as defined before.

Proof of Lemma 5.1: Since ${}^4X$ is a Killing vector field, Killing's equations

$$ {}^4X_{\mu;\nu} + {}^4X_{\nu;\mu} = 0 $$

hold.

Let $\mu = \nu = 0$ in this equation and obtain $2\,\partial_0\, {}^4X_0 = 0$, using the fact that $\Gamma^\alpha_{00} = 0$ in GNC. This proves the first result.

Now let $\mu = 0$ and $\nu = k$. Then the corresponding Killing's equations are

$$ \partial_k({}^4X_0) + \partial_0({}^4X_k) = 2\Gamma^\alpha_{0k}\, {}^4X_\alpha . $$

This implies that

$$ \partial_t({}^4X_k) = 2\Gamma^i_{0k}\, {}^4X_i - \partial_k({}^4X_0) . \tag{5.4} $$

Also,

$$ \partial_t({}^4X^i) = \partial_t({}^4g^{ik}\, {}^4X_k) = (\partial_t\, {}^4g^{ik})\, {}^4X_k + {}^4g^{ik}(\partial_t\, {}^4X_k) . $$

If we use Eq. (5.4), the fact that ${}^4X_0 = -{}^4X^0 = -N$, and also that $\Gamma^j_{i0} = \frac{1}{2}\, {}^4g^{kj}(\partial_t\, {}^4g_{ik})$, we obtain the second result.

Proof of Lemma 5.2: For any tensorial form ${}^4B$ and vector field ${}^4Y$ on ${}^4S$,

$$ {}^4D(i_{({}^4Y)}{}^4B) + i_{({}^4Y)}\, {}^4D({}^4B) = L_{({}^4Y)}\, {}^4B + [{}^4A({}^4Y), {}^4B] . $$

So,

$$ \begin{aligned} i_{\not{z}}\, {}^4D(i_{({}^4X)}{}^4F) &= -{}^4D(i_{\not{z}}\, i_{({}^4X)}\, {}^4F) + L_{\not{z}} i_{({}^4X)}\, {}^4F + [{}^4A(\not{z}), i_{({}^4X)}\, {}^4F] \\ &= {}^4Di_{{}^4X}\, i_{\not{z}}\, {}^4F + i_{{}^4X}L_{\not{z}}\, {}^4F + i_{[\not{z},{}^4X]}\, {}^4F + [{}^4A(\not{z}), i_{{}^4X}\, {}^4F] \\ &= {}^4Di_{{}^4X} i_{\not{z}}\, {}^4F + i_{{}^4X}(L_{\not{z}}\, {}^4F + [{}^4A(\not{z}), {}^4F]) + i_{[\not{z},{}^4X]}\, {}^4F \\ &= {}^4Di_{{}^4X} i_{\not{z}}\, {}^4F + i_{{}^4X}\, {}^4Di_{\not{z}}\, {}^4F + i_{[\not{z},{}^4X]}\, {}^4F = L_{{}^4X} i_{\not{z}}\, {}^4F + [{}^4A({}^4X), i_{\not{z}}\, {}^4F] + i_{[\not{z},{}^4X]}\, {}^4F . \end{aligned} $$

On M, ${}^4X = N\not{z} + X$, ${}^4A(\not{z}) = -C$, and, by Lemma 5.1, $[\not{z}, {}^4X] = (\partial^iN)\partial_i$, where the latter expression is grad(N). So, when restricted to M, $i_{\not{z}}\, {}^4D(i_{{}^4X})\, {}^4F$ becomes $N(-L_{\not{z}}E) - L_XE + N[C,E] - [A(X),E] + i_{\text{grad}(N)}F$. This last expression, by Eq. (5.3), is $-ND^*F + i_{\text{grad}\,N}F - N(\Gamma E) - L_XE - [A(X),E]$. The result then follows from the fact that the first two terms add up to $-D^*(NF)$.

Proof of Proposition 5.1: Suppose that ${}^4X$ is a symmetry of ${}^4A$. Then, By Eq. (4.11),

$$ i_{{}^4X}\, {}^4F = {}^4D({}^4V) $$

for some tensorial function ${}^4V$.

Since ${}^4X = N\not{z} + X$ on M, the equation on M becomes

$$ -NE + i_XF = DV \tag{5.5} $$

where V is the restriction of ${}^4V$ to M.

Also,

$$[i_{\mathfrak{z}}\, {}^4F,{}^4V] = i_{\mathfrak{z}}[{}^4F,{}^4V] = i_{\mathfrak{z}}({}^4D)^2({}^4V) = i_{\mathfrak{z}}\, {}^4D(i_{{}^4X}\, {}^4F)\,.$$

By using Lemma 5.2, we see that the last result implies that

$$-[E,V] = -D^*(NF) - L_X E - N(\Gamma E) - [A(X),E]\,. \tag{5.6}$$

The two Eqs. 5.5 and 5.6 imply that $\Phi(A,E)$ is an element of Im $J\circ K'^*$.

Conversely, given that $\Phi(A,E) = J\circ K'^*(V)$ on M, we extend V to a tensorial function ${}^4V$ on ${}^4S$ such that

$$i_{\mathfrak{z}}\, {}^4D({}^4V) = i_{\mathfrak{z}}(i_{{}^4X}\, {}^4F)\,. \tag{5.7}$$

This result, together with what is given, implies that

$$i_{{}^4X}({}^4F) = {}^4D({}^4V) \quad \text{on } M\,.$$

Now, $i_{{}^4X}({}^4F) - {}^4D({}^4V)$ is a solution of the linearized YM equation (5.1) (see, e.g., Ref. 1). This solution vanishes on M and hence vanishes on ${}^4S$ provided

$$L_{\mathfrak{z}}(i_{{}^4X}\, {}^4F - {}^4D({}^4V)) = 0, \quad \text{on } M\,.$$

To prove the latter equation, note that on M,

$$\begin{aligned} &L_{\mathfrak{z}}(i_{{}^4X}\, {}^4F - {}^4D({}^4V)) \\ &\quad = i_{\mathfrak{z}} d(i_{{}^4X}\, {}^4F - {}^4D\, {}^4V) + d i_{\mathfrak{z}} i_{{}^4X}\, {}^4F - {}^4D\, {}^4V) \\ &\quad = i_{\mathfrak{z}} d(i_{{}^4X}\, {}^4F - {}^4D\, {}^4V) \quad [\text{by Eq (5.7)}]\,. \end{aligned}$$

But, on M,

$$i_{\mathfrak{z}} d(i_{{}^4X}\, {}^4F - {}^4D\, {}^4V) = i_{\mathfrak{z}}\, {}^4D(i_{{}^4X}\, {}^4F - {}^4D\, {}^4V)\,,$$

because

$$\begin{aligned} i_{\mathfrak{z}}[{}^4A, i_{{}^4X}\, {}^4F - {}^4D\, {}^4V] &= [{}^4A(\mathfrak{z}), i_{{}^4X})\, {}^4F - {}^4D\, {}^4V] \\ &\quad + [{}^4A, i_{\mathfrak{z}}(i_{{}^4X}\, {}^4F - {}^4D\, {}^4V)] \end{aligned}$$

vanishes on M.

On the other hand, we have on M (using Lemma 5.2) that

$$\begin{aligned} i_{\mathfrak{z}}\, {}^4D(i_{{}^4X}\, {}^4F - {}^4D\, {}^4V) &= i_{\mathfrak{z}}\, {}^4D(i_{{}^4X}\, {}^4F) - i_{\mathfrak{z}}[{}^4F,{}^4V] \\ &= -D^*(NF) - L_X E - N(\Gamma E) \\ &\quad - [A(X),E] + [E,V]\,. \end{aligned}$$

The latter expression vanishes, by hypotheses.

Thus we have, on ${}^4S$, that

$$i_{({}^4X)}\, {}^4F = {}^4D({}^4V)$$

where ${}^4V$ is a tensorial form.

By Eq. (4.11), we see that this implies that ${}^4X$ is a symmetry of ${}^4A$ ■

VI. THE SPACE OF HOMOGENEOUS YANG–MILLS FIELDS

In this section we examine the structure of the space of YM fields that are homogeneous relative to a Lie algebra of Killing vector fields that are tangent to some compact spacelike Cauchy surface in the four-dimensional space-time manifold. Our main result is Proposition 6.1, which shows that the space is a smooth manifold near each field with only trivial gauge symmetries. An outline of the proof of this theorem is as follows.

Suppose that ${}^4S$ is a space-time manifold with a compact spacelike Cauchy surface M, then the YM fields on ${}^4S$ are completely determined by the elements of $\mathscr{C} = \operatorname{Ker} K$ (Sec. V). If h is an isometry of ${}^4S$ that fixes M, it follows that an element of $\mathscr{C}$ has symmetry h if and only if the corresponding YM field has the same symmetry. A similar result therefore holds also for Killing vector fields that are tangent to M.

Let (A,E) be an element of $\mathscr{C}$ and let $\mathscr{S}$ be the slice at (A,E) described in Sec. III. There exists[3] a local diffeomorphism $\mathscr{B}$, defined near (A,E) in $\mathscr{T}\mathscr{A}$, which satisfies the property that if (A,E) has only trivial gauge symmetries, then $\mathscr{B}$ maps a neighborhood of (A,E) in $\mathscr{C}\cap\mathscr{S}$ to an affine ILH space. Thus $\mathscr{C}\cap\mathscr{S}$ is a smooth manifold near (A,E).

Suppose $\mathscr{k}$ is a Lie algebra of Killing vector fields that are tangent to M. Let $\mathscr{H}$ be the set of all elements of $\mathscr{C}$ that are homogeneous relative to $\mathscr{k}$. If (A,E) lies in $\mathscr{H}$ and has trivial gauge symmetries, then $\mathscr{B}$ maps a neighborhood of (A,E) in $\mathscr{H}\cap\mathscr{S}$ to an affine ILH space. Finally we show that near (A,E), $\mathscr{H}$ equals $\mathscr{G}(A,E)\times(\mathscr{H}\cap\mathscr{S})$ and so $\mathscr{H}$ is a smooth manifold near (A,E).

Since each element of $\mathscr{k}$ is tangent to M, it follows that each element is independent of the t coordinate of the GNC (cf. Lemma 5.1). Thus ${}^4X$ in $\mathscr{k}$ may be identified with its restriction X on M.

We shall need several lemmas.

Lemma 6.1: Suppose that $(A,E)\in\mathscr{C}$ and $(A_1,E_1)\in\mathscr{C}\cap\mathscr{S}$ have the same symmetry X of $\mathscr{k}$. Furthermore, suppose that the dimensions of their gauge symmetry groups are the same. Then the two gauge fields have the same set of symmetry functions relative to X.

Proof: Since (A_1,E_1) lies in the slice $\mathscr{S}$, its gauge symmetry group $\mathscr{G}(A_1,E_1)$ lies in $\mathscr{G}(A,E)$ [the isotropy group of (A,E) under $\mathscr{G}$]. This follows from the second property of a slice (see the end of Sec. III). So, the Lie algebra of $\mathscr{G}_{(A_1,E_1)}$ lies in the Lie algebra of $\mathscr{G}_{(A,E)}$. Since their dimensions are the same, they are equal. Hence (A,E) and (A_1,E_1) have the same set of infinitesimal gauge symmetries.

Since X is a symmetry of (A_1,E_1), there is a lift $\{L(h_t)\}$ of the flow $\{h_t\}$ of X to the bundle space such that

$$L(h_t)(A_1,E_1) = (A_1,E_1)\,.$$

Since X is a symmetry of (A_1,E_1), there is a lift $\{L(h_t)\}$ of the flow $\{h_t\}$ of X to the bundle space such that

$$L(h_t)(A_1,E_1) = (A_1,E_1)\,.$$

So, the function defined by Eq. (4.9) is a transformation function of both (A,E) and (A_1,E_1). Therefore, there is a common symmetry function W_X.

Now, the lifts of h_t, which fix a connection, differ by a gauge symmetry of the connection. This implies, by Eq. (4.9), that the corresponding transformation functions differ by a gauge symmetry. Hence the symmetry functions, which are given by Eq. (4.10), differ by infinitesimal gauge symmetries. The result then follows from the fact that the latter symmetries for (A,E) and (A_1,E_1) are identical. ■

Let (A,E) be an element of $\mathscr{C}$ such that (A,E) has only trivial gauge symmetries (i.e., those common to all fields) and is homogeneous relative to $\mathscr{k}$. Then, for each X in $\mathscr{k}$, Proposition 5.1 implies that

$$i_X F = D(V_X)$$

and

$$L_X E + [A(X),E] = [E,V_X] ,$$

for some tensorial function V_X (since $N=0$).

This result can be written as

$$L_X A = DW_X \equiv dW_X + [A,W_X]$$

and

$$L_X E = [E,W_X] ,$$

where $W_X = A(X) + V_X$ is the symmetry function [cf. Eq. (4.8)].

If we extend L_X to ordered pairs componentwise, then the two equations may be combined to give

$$L_X(E,-A) = K'^*(W_X) . \tag{6.1}$$

Let $\{L(h_t)\}$ again be a lift of the flow $\{h_t\}$ of X that fixes (A,E). Define the linear operator α_X by

$$\alpha_X(B) = \frac{d}{dt}(L(h_t)B) \quad \text{(evaluated at } t=0) , \tag{6.2}$$

where B is any tensorial form on M. Extend α_X to ordered pairs componentwise. The operator α_X is also given by

$$\alpha_X(B) = L_X(B) + [W_X,B] . \tag{6.3}$$

The latter follows from the fact that Eqs. (4.2) and (4.9) lead to

$$L(h_t)B = \mathrm{Ad}\, T_t(h_t^* B) \quad \text{(for small values of } t) . \tag{6.4}$$

Now, differentiation of the terms in Eqs. (6.4) at $t=0$, gives the required result.

Let us define Ker α as follows:

$$\mathrm{Ker}\,\alpha = \{(a,e)\in\mathscr{D}^1\times\mathscr{D}^1: \alpha_X(a) = \alpha_X(e) = 0, \text{ for all } X \text{ in } \mathscr{k}\} .$$

Lemma 6.2: $\mathscr{S}\cap\mathscr{H} = \{(A+a,E+e)\in\mathscr{S}\cap\mathscr{C}: (a,e)\in\mathrm{Ker}\,\alpha\}$, where Ker α is as defined above.

Proof: Let $(\bar{A},\bar{E}) = (A+a,E+e)\in\mathscr{S}\cap\mathscr{C}$. Then Eq. (6.1) yields

$$L_X(\bar{E},-\bar{A}) = \bar{K}'^*(W_X) + \alpha_X(e,-a) , \tag{6.5}$$

where $\bar{K}'^*$ is the adjoint of the derivative of K at $(\bar{A},\bar{E})$.

Suppose that $(a,e)\in\mathrm{Ker}\,\alpha$. Then $\alpha_X(e,-a) = (0,0)$ and so Eq. (6.5) implies that X is a symmetry of $(\bar{A},\bar{E})$. This is true for all X in $\mathscr{k}$ and so $(\bar{A},\bar{E})\in\mathscr{S}\cap\mathscr{H}$.

Conversely, suppose that $(\bar{A},\bar{E})\in\mathscr{S}\cap\mathscr{H}$ and $X\in\mathscr{k}$. Let $\bar{W}_X$ be a symmetry function of $(\bar{A},\bar{E})$ relative to X. Then $L_X(\bar{E},-\bar{A}) = \bar{K}'^*(\bar{W}_X)$ and so

$$\alpha_X(e,-a) = \bar{K}'^*(\bar{W}_X - W_X) .$$

There is a one-to-one correspondence between the infinitesimal gauge symmetries of the YM field corresponding to (A,E) and the elements of Ker K'^* (see, e.g., Ref. 1). Since (A,E) has only trivial gauge symmetries, Ker $K'^* \subseteq$ Ker $\bar{K}'^*$. But Ker $\bar{K}'^* \subseteq$ Ker K'^* follows from an argument similar to that in the proof of Lemma 6.1 (the first part). Hence,

$$\mathrm{Ker}\,K'^* = \mathrm{Ker}\,\bar{K}'^* . \tag{6.6}$$

By Lemma 6.1, the two fields have the same set of symmetry functions. Again, by the argument at the end of the proof of Lemma 6.1, this means that $\bar{W}_X - W_X$ lies in Ker $\bar{K}'^*$ and so $\alpha_X(e,-a) = (0,0)$.

Since the latter is true for each X in $\mathscr{k}$, $(a,e)\in\mathrm{Ker}\,\alpha$. This completes the proof.

We now define the operator $\mathscr{R}$ as follows: Since K'^* is elliptic, there is an orthogonal splitting

$$\mathscr{D}^0 = \mathrm{Im}\,K' \oplus \mathrm{Ker}\,K'^* \tag{6.7}$$

[cf. Remark after Eq. (3.3)]. The linear map $K'K'^*$ is invertible as a map from Im K' into itself. Let $\mathscr{R}$ be its inverse on Im K', and extend $\mathscr{R}$ to be zero on Ker K'^*.

Lemma 6.3: The operator α_X [Eqs. (6.2) and (6.3)] commutes with K' and with $\mathscr{R}$. Moreover,

$$\alpha_X b_e^* a = b_e^* \alpha_X a - b_a^* \alpha_X e .$$

Proof: Let h be an isometry of M. Suppose h lifts to an automorphism $L(h)$ of the principal bundle P. A direct computation shows that

$$D_h(B) = L(h)D(L(h)^{-1}B) ,$$

where D_h is covariant differentiation relative to $L(h)A$ and B is any tensorial form.

Let V be a tensorial function. Then we have the following result:

$$\begin{aligned}\langle D_h^*(L(h)E),V\rangle &= \langle L(h)E,D_h V\rangle \\ &= \langle L(h)E,L(h)D(L(h)^{-1}V)\rangle \\ &= \langle E,D(L(h)^{-1}V)\rangle \quad \text{(by Prop. 4.1)} \\ &= \langle D^*E,L(h)^{-1}V\rangle \\ &= \langle L(h)D^*E,V\rangle .\end{aligned}$$

Hence,

$$\begin{aligned}K(L(h)A,L(h)E) &= -D_h^*(L(h)E) \\ &= -L(h)D^*E = L(h)K(A,E) ,\end{aligned}$$

i.e., $L(h)$ commutes with K. This implies that

$$K'(L(h)a,L(h)e) = L(h)K'(a,e)$$

whenever $L(h)(A,E) = (A,E)$ and (a,e) lies in $\mathscr{D}^1\times\mathscr{D}^1$.

In particular, let $\{L(h_t)\}$ be a lift of the flow $\{h_t\}$ of the vector field X, which fixes (A,E). Then

$$K'(L(h_t)a,L(h_t)e) = L(h_t)K'(a,e) .$$

Differentiation relative to t at 0 yields the result that

$$K'\alpha_X(a,e) = \alpha_X K'(a,e) ,$$

i.e., K' commutes with α_X. This implies also that α_X maps Im K' into itself.

Now, by using the inner product $\langle\ ,\ \rangle$, it is easy to show that K'^* also commutes with α_X. Therefore α_X commutes with $\Delta = K'K'^*$.

Let $\Delta\mathscr{R} = \mathscr{R}\Delta = \mathscr{p}$ be the orthogonal projection of $\mathscr{D}^0$ onto Im K' [see Eq. (6.7)]. Then

$$\mathscr{R}\alpha_X(a) = \mathscr{R}\not{h}\alpha_X(a) \quad (\text{since } \mathscr{R} \text{ vanishes on } \operatorname{Ker} K'^*)$$
$$= \mathscr{R}\alpha_X\not{h}(a) \quad (\text{since } \alpha_X \text{ commutes with } K' \text{ and } K'^*)$$
$$= \mathscr{R}\alpha_X\Delta\mathscr{R}(a)$$
$$= \mathscr{R}\Delta\alpha_X\mathscr{R}(a) = \not{h}\alpha_X\mathscr{R}(a) = \alpha_X\not{h}\mathscr{R}(a)$$
$$= \alpha_X\mathscr{R}(a) \quad (\text{since } \operatorname{Im}\mathscr{R} \subseteq \operatorname{Im} K')\,.$$

So, α_X commutes with $\mathscr{R}$.

Lastly, if a and e are tensorial one-forms on M, then, in local coordinates, b_e^*a can be shown to equal $[a_k,e^k]$. So, in local coordinates,

$$\alpha_X b_e^* a = L_X[a_k,e^k] + [W_X[a_k,e^k]]$$
$$= [(L_Xa)_k,e^k] + [a_k,(L_Xe)^k]$$
$$+ [[W_X,a_k],e^k] + [a_k,[W_X,e^k]].$$

Now observe that

$$(L_Xa)_k = L_X(a_k) + (\partial_k X^i)a_i \quad \text{and}$$
$$= (L_Xe)^k L_X(e^k) - (\partial_i X^k)e_i\,.$$

From this we deduce that

$$\alpha_X b_e^* a = [(\alpha_X a)_k,e^k] + [a_k,(\alpha_X e)^k]$$

because the other terms cancel out. ■

Proposition 6.1: Suppose that (A,E) lies in $\mathscr{C}$ and has only trivial gauge symmetries. Furthermore, suppose that (A,E) is homogeneous relative to $\not{h}$. Then the space $\mathscr{H}$, of YM fields that are homogeneous relative to $\not{h}$ is a smooth manifold near (A,E).

Proof: Since $\mathscr{S}$ is a slice for the action of $\mathscr{G}$ on $\mathscr{T}\mathscr{A}$,

$$\mathscr{T}\mathscr{A} = \mathscr{G}(A,E)\times\mathscr{S}$$

near (A,E), where $\mathscr{G}(A,E)$ is the orbit of (A,E) under $\mathscr{G}$. Hence,

$$\mathscr{H} = \mathscr{G}(A,E)\times(\mathscr{S}\cap\mathscr{H})$$

near (A,E).

Since $\mathscr{G}(A,E)$ is a smooth manifold, it suffices to show that $\mathscr{S}\cap\mathscr{H}$ is a smooth manifold near (A,E).

From Lemma 6.2 follows that

$$\mathscr{S}\cap\mathscr{H} = \{(A+a,E+e)\in\mathscr{S}\cap\mathscr{C}:\ (a,e)\in\operatorname{Ker}\alpha\}\,.$$

Let $\mathscr{B}(A+a,E+e) = (A+a,E+e) + K'^*\mathscr{R}b_e^*(a)$, where $\mathscr{R}$ is as defined before Lemma 6.3.

Since $b_e^*(a) = [a_k,e^k]$ in local coordinates, it follows that the derivative of $\mathscr{B}$ at (A,E) is the identity map. This implies that $\mathscr{B}$ is a local diffeomorphism at (A,E) in $\mathscr{T}\mathscr{A}$.

From Lemma 6.3, it follows that $\mathscr{B}$ maps $(A,E) + \operatorname{Ker}\alpha$ into itself. Also, from Ref. 3, we note that $\mathscr{B}$ maps $\mathscr{S}\cap\mathscr{C}$ into $(A,E) + \operatorname{Ker} K'\cap\operatorname{Ker} K'\circ J$ and also maps a neighborhood of (A,E) in $\mathscr{S}\cap\mathscr{C}$ diffeomorphically onto an open subset of $(A,E) + \operatorname{Ker} K'\cap\operatorname{Ker} K'\circ J$. So,

$$\mathscr{B}(\mathscr{H}\cap\mathscr{S})$$
$$= \mathscr{B}(\mathscr{H}\cap\mathscr{S}\cap\mathscr{C}) = \mathscr{B}(\mathscr{C}\cap\mathscr{S}\cap\{(A,E)+\operatorname{Ker}\alpha\})$$
$$\subseteq (A,E) + \operatorname{Ker} K'\cap\operatorname{Ker} K'\circ J\cap\operatorname{Ker}\alpha\,.$$

Since $\mathscr{B}$ is a local diffeomorphism, it follows that $\mathscr{B}$ maps a neighborhood of (A,E) in $\mathscr{H}\cap\mathscr{S}$ diffeomorphically onto an open subset of the affine ILH space $(A,E) + \operatorname{Ker} K'\cap\operatorname{Ker} K'\circ J\cap\operatorname{Ker}\alpha$. Hence the result follows. ■

Note: From the proof of this proposition, it is clear that the same result holds if the condition about trivial gauge symmetries is dropped and $\mathscr{H}$ is replaced by homogeneous YM fields that have the same dimension for their gauge symmetry groups, so that Eq. (6.6) may hold.

VII. GENERALIZATIONS

Let ${}^4S$ be a nonflat space-time manifold that contains a compact spacelike Cauchy surface of constant mean curvature. Marsden and Tipler[14] have shown that, under certain generic conditions, ${}^4S$ has a foliation by a family of Cauchy surfaces of distinct constant mean curvatures. This result implies that the Killing vector fields, if any, are tangent to the Cauchy surface. If ${}^4S$ is an Einstein space-time, the latter result also holds. Thus our results apply to a large class of space-time manifolds. A major class of examples, which do not satisfy the generic conditions stated above, are the static space-times. The latter contain Killing vector fields that are timelike, and hence are not of the type considered here.

The results in this paper form part of a Ph.D. dissertation.[6] Some generalization to the case of static space-times is included in the dissertation. Briefly, the Lie algebra $\not{h}$ is allowed to include the unit normal $\mathfrak{z}$. This algebra is shown to be spanned by $\mathfrak{z}$ and vector fields that are tangent to the Cauchy surface. The problem then reduces to that of describing the space of YM fields on M, which are homogeneous relative to the Killing vector fields of M. Proposition 6.1 then holds in this case provided the space of YM fields itself has a tangent space at A. The latter may be expressed by saying that the space is linearization stable at A (a concept that is used by Marsden and others in the study of the structure of the space-time manifold).

This extension does not exhaust all space-times that are usually considered in physics, and therefore the general case remains to be examined. Also, the structure of the space of homogeneous YM fields has been determined only near those fields that have only the trivial gauge symmetries.

ACKNOWLEDGMENTS

The author is grateful to Professor J. M. Arms of the University of Washington, Seattle, Washington for suggesting this investigation and for the many regular discussions held during the preparation of the Ph.D. dissertation from which this work has been extracted for publication.

[1]J. M. Arms, J. Math. Phys. **20**, 443 (1979).
[2]V. Moncrief, Ann. Phys. NY **108**, 87 (1977).
[3]J. M. Arms, Math. Proc. Cambridge Philos. Soc. **90**, 361 (1981).
[4]H. Omori, Proc. Sympos. Pure Math. XV, 11 (1970).
[5]W. Kondracki and J. Rogulski, "The stratification structure in the space of gauge equivalent Yang–Mills connections," preprint, Warsaw (1980).

[6]M. Molelekoa, "Space-time symmetries of Yang–Mills fields," Ph.D. thesis, University of Washington 1982.
[7]M. Molelekoa, J. Math. Phys. **26**, 192 (1985).
[8]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, Vol. I (Interscience, New York, 1963).
[9]I. M. Singer, Commun. Math. Phys. **60**, 7 (1978).
[10]P. K. Mitter and C. M. Viallet, Commun. Math. Phys. **79**, 457 (1981).
[11]J. M. Arms, J. E. Marsden, and V. Moncrief, Commun. Math. Phys. **78**, 455 (1981).
[12]T. P. Branson, "Gauge-invariant analytic structures on the Yang–Mills phase-space," preprint, Purdue University, West Lafayette, IN.
[13]L. Hörmander, *Linear Partial Differential Operators* (Springer, Berlin, 1963), pp. 267 and 268; Ann. Math. **83**, 129 (1966).
[14]J. E. Marsden and F. J. Tipler, Phys. Rep. **66**, 109 (1980).

Ermakov and non-Ermakov systems in quantum dissipative models

Antônio B. Nassar
Department of Physics, University of California at Los Angeles, Los Angeles, California 90024

(Received 20 June 1985; accepted for publication 31 October 1985)

Via the hydrodynamical formulation of quantum mechanics, a unified protocol to treat the quantum time-dependent harmonic oscillator with friction is presented, described by two different models: an explicitly time-dependent, linear Schrödinger equation (Caldirola–Kanai model) and a logarithmic nonlinear Schrödinger equation (Kostin model). For the former model, an Ermakov system that makes it possible to obtain an invariant of Ermakov–Lewis-type is derived. For the latter model, a non-Ermakov system is derived instead and it is shown that neither an exact nor an approximate invariant of Ermakov–Lewis-type exists.

I. INTRODUCTION

Since Lewis' rediscovery of an exact invariant of the time-dependent harmonic oscillator,[1] the theory of invariants (constants of motion or first integrals) has become a center of intense research with diversified applications in classical and quantum physics.[2–12] Essentially, Lewis showed that a conserved quantity for the time-dependent harmonic oscillator is given by

$$I = \tfrac{1}{2}[(\dot{q}\alpha - \dot{\alpha}q)^2 + (q/\alpha)^2], \tag{1.1}$$

where q and α satisfy, respectively,

$$\ddot{q} + \omega^2(t)q = 0 \tag{1.2}$$

and

$$\ddot{\alpha} + \omega^2(t)\alpha = 1/\alpha^3. \tag{1.3}$$

In fact, this problem traces back to Ermakov[13] who derived (1.1) (the Ermakov–Lewis invariant) by eliminating $\omega^2(t)$ between (1.2) and (1.3) (the Ermakov system).

Thus far, the solution of the so-called Ermakov–Lewis problem and its generalizations has been found, mainly, by the following four methods[14–16]: (1) Kruskal's method of (exact) adiabatic invariants,[17] (2) Leach's method of time-dependent canonical transformations,[16,18] (3) Noether's theorem as developed by Katzin and Levine,[19] Lutzky,[20] and Ray and Reid,[21] and (4) the Lie theory of extended groups as presented by Leach[22] and Gauthier.[9] These methods have also been applied in the search of invariants for dissipative systems with an underlying explicitly time-dependent Lagrangian[23–26]: this type of Lagrangian model can be interpreted as a special example for a scalar in a Riemannian curvated configuration space where the metric describes a friction force[27–39] or, alternatively, as an example of a system with a time-varying mass[40–42] (the Caldirola–Kanai model).

Recently, there has been an increasing number of papers trying to remedy the conceptual difficulties of this model. Their main argument is that a physically reasonable description of quantum dissipative systems by the Caldirola–Kanai model is attainable by an inclusion of a stochastic external force, representing the interaction of the particle with a chaotic bath[27–43] or by considering times shorter than the inverse friction constant $1/\nu$ (see Ref. 39). Furthermore, it has been shown that by a proper rescaling transformation one can reduce the Caldirola–Kanai Lagrangian/Hamiltonian into another without dissipation.[11,25,38] This, then, may justify *a priori* the existence of an exact or approximate invariant for classical or quantum dissipative systems described by the Caldirola–Kanai model.[24,25]

In order to avoid some of the above-mentioned ambiguities, quantum mechanical treatment of dissipative processes also has been introduced through nonlinear Schrödinger equations. Among them, the Kostin nonlinear Schrödinger equation was the first to be discovered,[44] and subsequently derived, within the realm of stochastic mechanics by Skagerstam,[45] Yasue,[46] and the author.[47] Much work has been built upon this nonlinear Schrödinger equation: We point out the works of Weiner and Forman,[48] Brüll and Lange,[49] Yasue,[50] Griffin and Kan[51] for the compelling physical reasons that motivate the study of this model. In fact, Caldeira and Leggett[52] have given a possible justification for the use of nonlinear wave equations (such as the Kostin nonlinear Schrödinger equation) for the description of nonconservative systems, based on their finding that damping tends to destroy interference effects of two Gaussian wave packets in a harmonic potential.

In light of the above discussion, one important question motivates our work here: how could one treat phenomenologically different quantum dissipative models (such as the quantized Caldirola–Kanai and the Kostin models), and investigate the possibility of finding (or not) Ermakov–Lewis-type invariants *all in the same scheme*?

In this paper we answer this question from a new perspective by studying the one-degree-of-freedom quantum dissipative time-dependent harmonic oscillator using the hydrodynamical formulation of quantum mechanics, which has proven to be enormously advantageous *vis á vis* other formulations.[53–66]

In Sec. II, we study the damped time-dependent harmonic oscillator described by the Caldirola–Kanai model and derive an Ermakov pair of equations (Ermakov system), which generates a corresponding Ermakov–Lewis-type invariant. In Sec. III, by proceeding in the same scheme, we show that this feature is not shared by the Kostin model: it comprises, conversely, a non-Ermakov pair of equations (non-Ermakov system), therefore not producing either an exact or an approximate Ermakov–Lewis-type invariant.

 0022-2488/86/030755-04$02.50

II. AN ERMAKOV SYSTEM (IN THE CALDIROLA–KANAI MODEL)

We begin with the quantum time-dependent harmonic oscillator with friction described by the explicitly time-dependent Schrödinger equation (Caldirola–Kanai model)[28–39]

$$i\hbar\frac{\partial\psi}{\partial t}(x,t) = -\frac{\hbar^2}{2m}e^{-\nu t}\frac{\partial^2\psi}{\partial x^2} + \frac{1}{2}e^{\nu t}m\omega^2(t)x^2\psi(x,t), \tag{2.1}$$

where $\psi(x,t)$, ν, and $\omega(t)$ are the wave function, constant friction coefficient, and time-dependent harmonic oscillator frequency, respectively.

To obtain the quantum fluid dynamics description of (2.1), we write the wave function $\psi(x,t)$ in the form

$$\psi(x,t) = \phi(x,t)\exp[iS(x,t)]. \tag{2.2}$$

After substitution of (2.2) into (2.1) we obtain from its real and imaginary parts

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} + \nu v + \omega^2(t)x = -\frac{1}{m}\frac{\partial V_{\text{qu}}}{\partial x} \tag{2.3}$$

and

$$\frac{\partial\rho}{\partial t} + \frac{\partial}{\partial x}(\rho v) = 0, \tag{2.4}$$

where $\rho\equiv\phi^2$ is the quantum fluid density, $v\equiv(\hbar/m)e^{-\nu t}\,\partial S/\partial x$ is the quantum fluid velocity, and $V_{\text{qu}} \equiv -(\hbar^2/2m)e^{-2\nu t}\rho^{-1/2}(\partial^2/\partial x^2)\rho^{1/2}$ is the quantum potential.

An essential unique feature of the quantum potential is that the force arising from it is unlike a mechanical force of a wave pushing on a particle with a pressure proportional to the wave intensity. So it follows that the expectation value of the quantum force vanishes for all times, i.e., $\langle\partial V_{\text{qu}}/\partial x\rangle = 0$ (see Ref. 43). Further if we prepare the fluid particle initially in a Gaussian wave packet centered at $x = 0$, $\rho(x,0) = [\pi\sigma(0)]^{-1/2}\exp[-x^2/\sigma(0)]$, and any initial velocity v_0, we may split (2.3) into

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} + \nu v + \omega^2(t)x = k(t)(x-q) \tag{2.5}$$

and

$$\frac{\partial}{\partial x}\left[\frac{\hbar^2 e^{-2\nu t}}{2m^2}\rho^{-1/2}\frac{\partial^2\rho^{1/2}}{\partial x^2}\right] = k(t)(x-q), \tag{2.6}$$

where $q(t)$ is the expectation (classical) value of x $[\langle x\rangle = q(t)]$ that will be determined in concomitance with $k(t)$.

Integrating (2.6) (assuming that ρ vanishes for $|x|\to\infty$ at any time), one obtains

$$\rho(x,t) = [\pi\sigma(t)]^{-1/2}\exp[-(x-q)^2/\sigma(t)], \tag{2.7a}$$

where

$$\sigma^2(t)\equiv\hbar^2 e^{-2\nu t}/m^2 k(t). \tag{2.7b}$$

Next, substituting (2.7) into (2.4) and integrating, we find

$$v(x,t) = (\dot\sigma/2\sigma)(x-q) + \dot q, \tag{2.8}$$

where the constant of integration must be zero since ρ vanishes for $|x|\to\infty$.

Inserting (2.8) into (2.5) we have our main result

$$\left(\frac{\ddot\sigma}{2\sigma} - \frac{\dot\sigma^2}{4\sigma^2} + \frac{\nu\dot\sigma}{2\sigma} + \omega^2(t) - \frac{\hbar^2 e^{-2\nu t}}{m^2\sigma^2}\right)(x-q) + [\ddot q + \nu\dot q + \omega^2(t)q] = 0. \tag{2.9}$$

This equation is identically satisfied if

$$\ddot\alpha + \nu\dot\alpha + \omega^2(t)\alpha = e^{-2\nu t}/\alpha^3 \tag{2.10a}$$

and

$$\ddot q + \nu\dot q + \omega^2(t)q = 0, \tag{2.10b}$$

where we have made $\sigma\equiv(\hbar/m)\alpha^2$.

By eliminating $\omega^2(t)$ between (2.10a) and (2.10b) and after some manipulations we find[23–25]

$$\dot I = 0, \tag{2.11a}$$

where

$$I\equiv\tfrac{1}{2}\{e^{2\nu t}[(\dot q\alpha - \dot\alpha q)^2] + (q/\alpha)^2\} \tag{2.11b}$$

is an Ermakov–Lewis-type invariant. So, (2.10a) and (2.10b) constitute an *Ermakov system*.

III. A NON-ERMAKOV SYSTEM (IN THE KOSTIN MODEL)

Now, we consider the quantum time-dependent harmonic oscillator with friction described by the nonlinear Schrödinger equation (Kostin model)[44–51]

$$i\hbar\frac{\partial\psi}{\partial t}(x,t) = -\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} + \left[\frac{1}{2}m\omega^2(t)x^2 + \frac{\hbar\nu}{2i}\ln\frac{\psi(x,t)}{\psi^*(x,t)}\right]\psi(x,t), \tag{3.1}$$

where the nonlinear term $(\hbar\nu/2i)\ln(\psi/\psi^*)$ accounts for the dissipation.

To obtain the quantum fluid dynamics description of (3.1), we proceed as in Sec. II, we express the wave function $\psi(x,t)$ as in (2.2) and obtain

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} + \nu v + \omega^2(t)x = -\frac{1}{m}\frac{\partial V_{\text{qu}}}{\partial x} \tag{3.2}$$

and

$$\frac{\partial\rho}{\partial t} + \frac{\partial}{\partial x}(\rho v) = 0, \tag{3.3}$$

where now $\rho\equiv\phi^2$, $v\equiv(\hbar/m)(\partial S/\partial x)$, and $V_{\text{qu}}\equiv-(\hbar^2/2m)\rho^{-1/2}(\partial^2/\partial x^2)\rho^{1/2}$.

By following closely the same scheme developed in Sec. II, we split (3.2) into

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} + \nu v + \omega^2(t)x = k(t)(x-q) \tag{3.4}$$

and

$$\frac{\partial}{\partial x}\left[\frac{\hbar^2}{2m^2}\rho^{-1/2}\frac{\partial^2}{\partial x^2}\rho^{1/2}\right] = k(t)(x-q). \tag{3.5}$$

Equation (3.5) yields

$$\rho(x,t) = [\pi\sigma(t)]^{-1/2}\exp[-(x-q)^2/\sigma(t)], \tag{3.6a}$$

where

$$\sigma^2(t) \equiv \hbar^2/m^2k(t) . \tag{3.6b}$$

The fluid-particle velocity $v(x, t)$ is obtained by substituting (3.6) into (3.3) and integrating: we obtain (2.8) with $\sigma(t)$ given by (3.6b).

Analogously, we have the main result

$$\left(\frac{\ddot{\sigma}}{2\sigma} - \frac{\dot{\sigma}^2}{4\sigma^2} + \frac{\nu\dot{\sigma}}{2\sigma} + \omega^2(t) - \frac{\hbar^2}{m^2\sigma^2}\right)(x - q) + [\ddot{q} + \nu\dot{q} + \omega^2(t)q] = 0 . \tag{3.7}$$

This equation is identically satisfied if

$$\ddot{\alpha} + \nu\dot{\alpha} + \omega^2(t)\alpha = 1/\alpha^3 \tag{3.8a}$$

and

$$\ddot{q} + \nu\dot{q} + \omega^2(t)q = 0 , \tag{3.8b}$$

where $\sigma \equiv (\hbar/m)\alpha^2$.

By eliminating $\omega^2(t)$ between (3.8a) and (3.8b) and after some manipulations we find

$$\dot{I} = \nu e^{2\nu t}(q/\alpha)^2 , \tag{3.9a}$$

where

$$I = \tfrac{1}{2} e^{2\nu t}[(\dot{q}\alpha - \dot{\alpha}q)^2 + (q/\alpha)^2] . \tag{3.9b}$$

So, (3.8a) and (3.8b) constitute a *non-Ermakov system*, since no Ermakov–Lewis-type invariant can be found (except in the trivial case as $\nu = 0$).[71]

IV. CONCLUSIONS

In summary, the relevant advantage of our method *vis à vis* the previously mentioned methods is that we can deal with linear quantum systems as well as nonlinear quantum systems within the same scheme (the hydrodynamical formulation of quantum mechanics). This makes it possible to compare and distinguish, in a clear fashion, what we called Ermakov and non-Ermakov systems, based on whether or not one can find a general invariant in the form of (1.1). Moreover, the protocol developed here can be used toward a deeper understanding of the role of the theory of invariants in conjunction with other types of nonlinear wave mechanical theories.[67–70] It poses some new perspectives and an alternative route that suggests further research and generalizations. Work in this direction is in progess and will be published in a forthcoming paper.

ACKNOWLEDGMENTS

I would like to thank Professor J. R. Ray, Dr. H. R. Lewis, Professor P. L. L. Leach, Professor R. Hasse, and Dr. D. Schuch for very stimulating correspondence and for sharing reprints of their recent works. I am also indebted to Professor S. Putterman for many fruitful discussions, and R. Berg and F. L. A. Machado for useful suggestions and careful reading of this manuscript. Special thanks to Professor J. M. Bassalo and Professor P. T. Alencar for encouragement and interest in this work.

[1] H. R. Lewis, Phys. Rev. Lett. **18**, 510 (1967); **18**, 636 (E) (1967).
[2] H. R. Lewis and W. B. Riesenfeld, J. Math. Phys. **10**, 1458 (1969).
[3] J. R. Ray, Lett. Nuovo Cimento **30**, 372 (1981).
[4] J. G. Hartley and J. R. Ray, Phys. Rev. A **24**, 2873 (1981); D **25**, 382 (1982).
[5] J. R. Ray, Phys. Rev. A **26**, 729 (1982); **28**, 2603 (1983); D **25**, 3417 (1982).
[6] H. Kohl and R. M. Dreizler, Phys. Lett. A **98**, 95 (1983).
[7] A. K. Dhara and S. V. Lawande, Phys. Rev. A **30**, 560 (1984); J. Phys. A **17**, 2423 (1984).
[8] H. R. Lewis and K. R. Symon, Phys. Fluids **27**, 192 (1984).
[9] S. Gauthier, J. Phys. A **17**, 2633 (1984).
[10] B. Abraham-Schrauner, Phys. Fluids **27**, 197 (1984).
[11] M. R. Feix and H. R. Lewis, J. Math. Phys. **26**, 68 (1985).
[12] A. B. Nassar, Phys. Rev. A **32**, 1862 (1985).
[13] V. P. Ermakov, Univ. Izv. Kiev **20** (9), 1 (1980).
[14] J. R. Ray, Lett. Nuovo Cimento **27**, 424 (1980).
[15] W. Sarlet and J. R. Ray, J. Math. Phys. **22**, 2504 (1981).
[16] H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 165 (1982).
[17] W. Sarlet. Ann. Phys. (N.Y.) **92**, 248 (1975); J. Phys. A **11**, 843 (1978).
[18] P. G. L. Leach, SIAM J. Appl. Math. **34**, 496 (1978).
[19] G. H. Katzin and J. Levine, J. Math. Phys. **18**, 1267 (1977).
[20] M. Lutzky, Phys. Lett. A **68**, 3 (1978).
[21] J. R. Ray and J. L. Reid, J. Math. Phys. **22**, 91 (1981).
[22] P. G. L. Leach, J. Math. Phys. **22**, 465 (1981).
[23] M. Lutzky, J. Math. Phys. **21**, 1370 (1980).
[24] L. S. Wollenburg, Phys. Lett. A **79**, 269 (1980).
[25] J. R. Reid and J. R. Ray, J. Math. Phys. **21**, 1583 (1980); **23**, 503 (1982).
[26] P. G. L. Leach, Phys. Lett. A **98**, 89 (1983).
[27] J. Messer, Acta Phys. Austriaca **50**, 75 (1979); Lett. Math. Phys. **2**, 281 (1978).
[28] W. H. Steeb and A. Kunick, Phys. Rev. A **25**, 2889 (1982).
[29] J. L. Reid and J. J. Cullen, Prog. Theor. Phys. **68**, 989 (1982).
[30] J. Heinrichs, Z. Phys. B **50**, 269 (1983).
[31] H. Gzyl, Phys. Rev. A **27**, 2297 (1983).
[32] G. Braunss, J. Math. Phys. **25**, 266 (1984).
[33] J. M. Cervero and J. Villaroel, J. Phys. A **17**, 1777 (1984).
[34] Y. R. Yang, Phys. Rev. A **29**, 18 (1984).
[35] B. J. McCoy and R. G. Carbonell, Phys. Rev. A **29**, 399 (1984).
[36] A. B. Nassar, Lett. Nuovo Cimento **41**, 476 (1984).
[37] L. F. Urrutia and E. Hernandez, Int. J. Theor. Phys. **23**, 1105 (1984).
[38] C. C. Gerry, J. Math. Phys. **25**, 1820 (1984).
[39] A. M. Jayannavar, Lett. Nuovo Cimento **42**, 141 (1985).
[40] J. R. Ray, Lett. Nuovo Cimento **25**, 47 (1979).
[41] R. K. Colegrave and M. S. Abdalla, J. Phys. A **15**, 1549 (1982).
[42] P. G. L. Leach, J. Phys. A **16**, 3261 (1983).
[43] A. B. Nassar, Phys. Lett. A **106**, 43 (1984).
[44] M. D. Kostin, J. Chem. Phys. **57**, 3589 (1972).
[45] B. K. Skagerstam, Phys. Lett. B **58**, 21 (1975); J. Math. Phys. **18**, 308 (1977).
[46] K. Yasue, Phys. Lett. B **64**, 239 (1976).
[47] A. B. Nassar, Phys. Lett. A **109**, 1 (1985).
[48] T. H. Weiner and R. E. Forman, Phys. Rev. B **10**, 315, 325 (1974).
[49] L. Brüll and H. Lange, J. Math. Phys. **25**, 786 (1984).
[50] K. Yasue, Ann. Phys. (N.Y.) **114**, 479 (1978).
[51] J. J. Griffin and K. K. Kan, Rev. Mod. Phys. **48**, 467 (1976).
[52] A. O. Caldeira and A. J. Leggett, Phys. Rev. A **31**, 1059 (1985).
[53] G. Casati and I. Guarneri, Phys. Rev. Lett. **42**, 1579 (1979).
[54] H. E. Wilhelm, Phys. Rev. D **1**, 2278 (1970).
[55] I. Bialynicki-Birula and Z. Bialynicka, Phys. Rev. D **3**, 2410 (1971); I. Bialynicki-Birula, Phys. Rev. D **3**, 2413 (1971).
[56] K. K. Kan and J. J. Griffin, Phys. Rev. C **15**, 1126 (1977).
[57] P. Frobrich, R. Lipperheide, and G. Thoma, Ann. Phys. (N.Y.) **104**, 478 (1977).
[58] F. Guerra, Phys. Rep. **77**, 263 (1981).
[59] S. K. Ghosh and B. M. Deb, Phys. Rep. **92**, 1 (1982); J. Phys. A **17**, 2463 (1984).
[60] S. J. Putterman and P. H. Roberts, Physica A **117**, 369 (1983).
[61] D. Bohm and B. J. Hilley, Found. Phys. **14**, 255 (1984).
[62] K. Wódkiewicz, Phys. Rev. A **29**, 1527 (1984).
[63] V. M. Strutinzky, A. G. Magner, and M. Brack, Z. Phys. A **318**, 205 (1984).
[64] G. Terlecki, N. Grün, and W. Scheid, J. Phys. B **17**, 3719 (1984).
[65] S. Olarin and J. I. Popescu, Rev. Mod. Phys. **57**, 339 (1985).
[66] A. B. Nassar, J. Phys. A **18**, L509 (1985).

[67]R. W. Hasse, Phys. Lett. B **85**, 197 (1979); Rep. Prog. Phys. **41**, 1027 (1978); J. Math. Phys. **16**, 2005 (1975). See also R. W. Hasse and K. Albrecht, Hadronic J. **5**, 1479 (1982).

[68]K. Albrecht, Phys. Lett. B **56**, 127 (1975).

[69]G. Süssmann, Los Alamos Seminar Talk, July 1973. See also P. Szilas and G. Süssmann, Z. Phys. B **50**, 209 (1983).

[70]D. Schuch, K-M. Chung, and H. Hartmann, J. Math. Phys. **25**, 3086 (1984).

[71]We can also verify that Eqs. (3.8a) and (3.8b) do not form an Ermakov pair of equations by applying the nonlinear superposition law to Eq. (3.8b) and proceeding along the line of Ref. 25. In this case, one cannot construct the corresponding physical (Ermakov–Lewis) invariant in the strict sense of Ermakov, although a trivial mathematical (non-Ermakov–Lewis) invariant quantity $I^* = I - \nu\int^t e^{2\nu s}(q(s)/\alpha(s))^2\,ds$ can be formed [with I given by Eq. (3.9b)]. We may substantiate further on this remark, bearing in mind the very importance of a physical (Ermakov–Lewis) invariant: its use as an artifact to construct an exact solution for the underlying Schrödinger equation. In other words, the presence of a physical (Ermakov–Lewis) invariant is a one-way link with a sure exact solution (the reciprocal may not necessarily be true). Thus, this comes about to corroborate, *a fortiori*, the already known facts in the literature: (1) that the Caldirola–Kanai model is exactly solvable (see Refs. 27–42), while (2) the Kostin model does not yield an exact solution (see Refs. 27, 48, 51, and 67).

On information gain by quantum measurements of continuous observables

Masanao Ozawa
Department of Mathematics, College of General Education, Nagoya University, Chikusa-ku, Nagoya 464, Japan

(Received 23 July 1985; accepted for publication 31 October 1985)

A generalization of Shannon's amount of information into quantum measurements of continuous observables is introduced. A necessary and sufficient condition for measuring processes to have a non-negative amount of information is obtained. This resolves Groenewold's conjecture completely including the case of measurements of continuous observables. As an application the approximate position measuring process considered by von Neumann and later by Davies is shown to have a non-negative amount of information.

I. INTRODUCTION

A pertinent approach to quantum measurements should have at least two aspects. One is the statistical aspect that describes statistics of the results of measurements. The other is the dynamical aspect that describes the dynamics of the processes of measurements. The conventional approach due to von Neumann[1] harmonizes these two aspects in a very simple way. The statistical formula (E) (see Ref. 1, p. 295) and the repeatability hypothesis (M) (see Ref. 1, p. 335) describe the statistics of the measurements completely. On the other hand, it is proved that measurements with these statistical properties can be described by a quantum-mechanical interaction between the observed system and the apparatus (see Ref. 1, Chap. VI). However, this beautiful theory can be applied only to measurements of discrete observables. For measurements of continuous observables, von Neumann proposes the approximation with step-function operators having a discrete spectrum.

There are several reasons for going further beyond von Neumann's theory; some of them other than the discontent of von Neumann himself (see Ref. 1, p. 223) are as follows.

(1) As pointed out by Wigner,[2] if there is an additive conservation law throughout the process of a measurement then the measurement of the quantity that does not commute with the conserved quantity cannot satisfy the repeatability hypothesis (see Refs. 3 and 4 for general proofs).

(2) Some results on approximate simultaneous measurements of noncommuting observables strongly demand new statistics for measurements of continuous observables (see Refs. 5 and 6).

(3) von Neumann's device of approximating continuous observables with discrete observables destroys the symmetry that continuous observables have naturally (see Ref. 7, p. 66).

In order to provide a basis of general considerations for the above nonidealized measurements, simultaneous measurements and covariant measurements, we have started[4] with an axiomatic approach. In our previous work,[4,8–10] we have discussed the following problems.

(1) When is the statistical description consistent with the dynamical description?

(2) How does the dynamical description determine the state change of the observed system caused by the measurement?

(3) How is the conditional expectation of the statistics of quantum measurements related to the state change?

In order to resolve the above problems, we have introduced the mathematical notion of a measuring process, which is a mathematical generalization of the dynamical description of quantum measurement. It is proved that this notion contains enough data to determine the state change caused by the measurement. On the other hand, a mathematical generalization of the statistical description of quantum measurement was previously introduced by Davies–Lewis,[11] which is referred to as an instrument. Then our solutions are based on the following result (see Ref. 4, Theorem 5.1): The statistical description is consistent with the dynamical description if and only if it is described by a completely positive instrument.

In the present paper, we shall introduce an information-theoretical aspect to our general theory of quantum measurements besides the above two aspects. Such a consideration was first done by Groenewold[12] for discrete idealized measurements.

In information theory, the information obtained by observation of a system is measured by the change of entropy in the observed system and it is proved that the average of *a posteriori* entropy is not larger than *a priori* entropy.[13] For the conventional description of quantum measurement due to von Neumann[1] and Lüders,[14] the corresponding quantum mechanical analog of the above inequality is conjectured by Groenewold[12] and proved by Lindblad.[15] Suppose that a discrete observable $X = \Sigma_i x_i P_i$ is measured by a conventional repeatable measuring process at the initial state ρ. Then we get the *a posteriori* state $\rho_i = (1/\mathrm{Tr}[P_i \rho])P_i \rho P_i$ with probability $p_i = \mathrm{Tr}\,[P_i \rho]$ for any measured value x_i. In this case, the *a priori* entropy is $\mathrm{S}[\rho] = -\mathrm{Tr}[\rho \log \rho]$ and the *a posteriori* entropy given the result x_i is $\mathrm{S}[\rho_i] = -\mathrm{Tr}[\rho_i \log \rho_i]$. Then the Groenewold–Lindblad inequality is as follows:

$$\mathrm{S}[\rho] - \sum_i p_i \mathrm{S}[\rho_i] \geqslant 0.$$

The left-hand side of the above inequality is just a quantum-mechanical analog of the amount of information introduced by Shannon in information theory. Thus conventional repeatable quantum measurements can be well interpreted as an

 0022-2488/86/030759-05$02.50

information transmission from the observed system to the apparatus.

In the present paper, we shall consider the generalization of the above inequality to our general quantum measurements. It is natural to expect that every physically relevant description of quantum measurement admits the above information-theoretical interpretation, that is, a generalized Groenewold–Lindblad inequality holds. However, since our most general description of quantum measurement is obtained axiomatically from the sole requirement of consistency of statistics and dynamics, it is not *a priori* true to hold the required inequality. Thus our problem is the following: What condition characterizes the general description of quantum measuring processes for which the average amount of information is always non-negative? In the following sections, we shall discuss and resolve the above problem.

In Sec. II, we provide necessary preliminaries for measuring processes and state changes. In Sec. III, we introduce a quantum-mechanical generalization of Shannon's amount of information. In Sec. IV, we obtain a necessary and sufficient condition for the generalized Groenewold–Lindblad inequality. Our main result can be stated using the terminology introduced in our previous work[4,9] as follows: We say that a measuring process is quasicomplete if when an *a priori* state is pure then almost all *a posteriori* states are pure. Then the generalized Groenewold–Lindblad inequality holds for every *a priori* state if and only if the measuring process is quasicomplete. In Sec. V this result will be applied to show that von Neumann's model of approximate position measurement (see Ref. 1, pp. 442–445) always satisfies the generalized Groenewold–Lindblad inequality.

II. MEASURING PROCESSES AND STATE CHANGES

A quantum system is described by a Hilbert space $\mathscr{H}$. Denote the algebra of all bounded operators on $\mathscr{H}$ by $\mathscr{L}(\mathscr{H})$ and the algebra of all trace class operators on $\mathscr{H}$ by $\mathscr{T}(\mathscr{H})$. A *state* is described by a density operator, i.e., a positive trace one operator on $\mathscr{H}$. A *semiobservable* X with value space $(\Lambda,\mathscr{B}(\Lambda))$ is a positive operator valued measure X: $\mathscr{B}(\Lambda)\to\mathscr{L}(\mathscr{H})$ on a Borel space $(\Lambda,\mathscr{B}(\Lambda))$ such that $X(\Lambda) = 1$. A semiobservable is called an *observable* if it is a spectral measure, i.e., projection valued.

Consider the following description of a measuring process of a quantum system. The observed system S and the apparatus M are described by Hilbert spaces $\mathscr{H}$ and $\mathscr{K}$, respectively. A measurement is carried out by an interaction during a finite time interval from time 0 to t, whose time evolution is given by a unitary operator $U = \exp(-itH)$ on $\mathscr{H}\otimes\mathscr{K}$, where H is the Hamiltonian of the system $S+M$. Before the interaction, S is in the (unknown) state ρ and M is in the (known) state σ, so that the system $S+M$ is in $\rho\otimes\sigma$. Thus by the interaction the state of $S+M$ changes into $U(\rho\otimes\sigma)U^*$. Let X be the semiobservable in S to be measured and $\widetilde{X}$ be the observable in M to show the value of X. Then the probability distribution Prob $[X\in dx|\rho]$ of the value of X at time 0 should coincide with the probability distribution Prob $[\widetilde{X}\in dx|U(\rho\otimes\sigma)U^*]$ of $\widetilde{X}$ at time t, that is, we have

$$\mathrm{Tr}[X(dx)\rho] = \mathrm{Tr}[(1\otimes\widetilde{X}(dx))U(\rho\otimes\sigma)U^*]. \quad (2.1)$$

Equation (2.1) is our sole requirement for the above interaction to be a measurement of X in the initial state ρ. We shall refer to any four-tuple $\langle\mathscr{K},\widetilde{X},\sigma,U\rangle$ consisting of a Hilbert space $\mathscr{K}$, an observable $\widetilde{X}$ in $\mathscr{K}$ with value space $(\Lambda,\mathscr{B}(\Lambda))$, a density operator σ on $\mathscr{K}$, and a unitary operator U on $\mathscr{H}\otimes\mathscr{K}$ satisfying Eq. (2.1) for any density operator ρ on $\mathscr{H}$ as a *measuring process* of a semiobservable X in the observed system S (cf. Ref. 4, Definition 3.1).

Denote by $\mathrm{CP}(\mathscr{T}(\mathscr{H}))$ the space of all completely positive maps on $\mathscr{T}(\mathscr{H})$. A $\mathrm{CP}(\mathscr{T}(\mathscr{H}))$-valued map $\mathscr{I}$ on $\mathscr{B}(\Lambda)$ is called a *CP instrument* if it satisfies the following conditions (I1) and (I2).

(I1) For each disjoint sequence $\{B_i\}$ in $\mathscr{B}(\Lambda)$,

$$\mathscr{I}(\cup_i B_i) = \sum_i \mathscr{I}(B_i),$$

where the sum is convergent in the strong operator topology of CP $(\mathscr{T}(\mathscr{H}))$.

(I2) For each density operator on $\mathscr{H}$, Tr $[\mathscr{I}(\Lambda)\rho] = \mathrm{Tr}[\rho]$.

The dual $\mathscr{I}(B)^*$ of $\mathscr{I}(B)$ is defined by the relation $\mathrm{Tr}[(\mathscr{I}(B)^*a)\rho] = \mathrm{Tr}[a(\mathscr{I}(B)\rho)]$, for all $B\in\mathscr{B}(\Lambda)$, $a\in\mathscr{L}(\mathscr{H})$, and $\rho\in\mathscr{T}(\mathscr{H})$.

As shown in Ref. 4, every measuring process $\langle\mathscr{K},\widetilde{X},\sigma,U\rangle$ determines a unique CP instrument $\mathscr{I}$ by the relation

$$\mathscr{I}(B)\rho = E_{\mathscr{K}}[(1\otimes\widetilde{X}(B))U(\rho\otimes\sigma)U^*], \quad (2.2)$$

for all $B\in\mathscr{B}(\Lambda)$ and $\rho\in\mathscr{T}(\mathscr{H})$, where $E_{\mathscr{K}}$ stands for the partial trace over $\mathscr{K}$. In this case, we have

$$\mathscr{I}(B)^*1 = X(B), \quad (2.3)$$

for all $B\in\mathscr{B}(\Lambda)$. The CP instrument $\mathscr{I}$ determines the state change caused by this measuring process, as follows. For any $B\in\mathscr{B}(\Lambda)$, let S_B be the subensemble of the measured system S in which the outcome of the measurement is in B and let ρ_B be the state of S_B at the instant after this measurement. Then, we have

$$\rho_B = (1/\mathrm{Tr}[\mathscr{I}(B)\rho])\mathscr{I}(B)\rho, \quad (2.4)$$

for all $B\in\mathscr{B}(\Lambda)$ with Tr $[\mathscr{I}(B)\rho]\neq 0$. For any $x\in\Lambda$, let S_x be the subensemble of the measured system S in which the outcome of the measurement is x and let ρ_x be the state of S_x at the instant after this measurement. Then by our statistical interpretation, we should impose the following requirements (A1) and (A2) on the family $\{\rho_x;\ x\in\Lambda\}$.

(A1) The function $x\to\rho_x$ is strongly $\mathscr{B}(\Lambda)$-measurable from Λ into the space of all density operators on $\mathscr{H}$.

$$\text{(A2)} \quad \int_B \rho_x\, \mathrm{Tr}[\mathscr{I}(dx)\rho] = \mathscr{I}(B)\rho,$$

where the integral is a Bochner integral.

We call any family $\{\rho_x; x\in\Lambda\}$ satisfying conditions (A1) and (A2) a family of *a posteriori states* with respect to the *a priori state* ρ. By Ref. 9, Theorem 4.5, a family of *a posteriori* states always exists and is unique in the following sense: If $\{\rho'_x; x\in\Lambda\}$ is another family of *a posteriori* states with respect to the *a priori* state ρ, then $\rho'_x = \rho_x$ for almost all $x\in\Lambda$ with respect to $\mathrm{Tr}[\mathscr{I}(dx)\rho]$.

For the detailed discussion, we shall refer the reader to Refs. 4 and 8–10.

III. AMOUNT OF INFORMATION OBTAINED BY MEASUREMENTS

From an information-theoretical point of view, a measurement gives us some information about the observed system. Since the state of the system is affected by our knowledge about the system, the information obtained by the measurement will be measured by the entropy change corresponding to the state change caused by the measurement.

Let X be a semiobservable on a Hilbert space $\mathscr{H}$ with value space $(\Lambda,\mathscr{B}(\Lambda))$. Now, assume that a measurement of X is carried out in the initial state ρ of the observed system by a measuring process $\mathbf{M} = \langle \mathscr{K},\tilde{X},\sigma,U\rangle$. Let $\mathscr{I}$ be the corresponding CP instrument. We call ρ the *a priori* state. Let $\{\rho_x;x\in\Lambda\}$ be a family of *a posteriori* states with respect to the *a priori* state ρ. Then the measurement changes the state from the *a priori* state ρ to the *a posteriori* states ρ_x, when the result $\tilde{X}=x$ is obtained. The entropy of the *a priori* state ρ is

$$S[\rho] = -\mathrm{Tr}[\rho\log\rho], \tag{3.1}$$

which we call the *a priori entropy*. The entropy of the *a posteriori* state ρ_x is

$$S[\rho_x] = -\mathrm{Tr}[\rho_x\log\rho_x], \tag{3.2}$$

which we call the *a posteriori entropy*. Thus in this case the information of this measurement is given by the following entropy change $I[\rho,\mathbf{M}|x]$:

$$I[\rho,\mathbf{M}|x] = S[\rho] - S[\rho_x], \tag{3.3}$$

which we call the *conditional amount of information* given $\tilde{X}=x$. Since the result $\tilde{X}=x$ of measurement is probabilistic event distributed by the probability $\mathrm{Prob}\,[X\in dx|\rho]$, our expected amount $I[\rho,\mathbf{M}]$ of information of this measurement is the average of the conditional amount of information, i.e.,

$$\begin{aligned} I[\rho,\mathbf{M}] &= \int_\Lambda I[\rho,\mathbf{M}|x]\ \mathrm{Prob}\ [X\in dx|\rho] \\ &= S[\rho] - \int_\Lambda S[\rho_x]\ \mathrm{Tr}\ [\mathscr{I}(dx)\rho], \end{aligned} \tag{3.4}$$

which we call the *amount of information* of the measuring process $\mathbf{M}$ with *a priori* state ρ. Since the family $\{\rho_x;x\in\Lambda\}$ of *a posteriori* states is determined uniquely up to μ-almost everywhere, where $\mu(dx) = \mathrm{Tr}[\mathscr{I}(dx)\rho]$, the amount of information $I[\rho,\mathbf{M}]$ does not depend on particular choice of the family of *a posteriori* states with respect to ρ.

Now we can state our generalization of the Groenewold–Lindblad inequality as follows.

(GL) $I[\rho,\mathbf{M}]\geqslant 0$, for any *a priori* state ρ with $S[\rho]<\infty$.

If the measuring process $\mathbf{M}$ is the conventional repeatable measurement of a discrete observable $X=\Sigma_i x_i P_i$. Then $\mathscr{I}(B)\rho = \Sigma\{P_i\,\rho P_i;x_i\in B\}$. In this case, statement (GL) is proved in Ref. 15, Theorem 2. In the next section we shall discuss when a given measuring process satisfies condition (GL).

In the rest of this section, we shall introduce another information theoretical quantity motivated by more classical interpretation. Let $\rho=\Sigma_i w_i\,\rho_i$ be the orthogonal decomposition of ρ into pure states ρ_i. Let $\nu(dx|i)$ be a transition probability defined by

$$\nu(dx|i) = \mathrm{Tr}[X(dx)\rho_i]. \tag{3.5}$$

Then the measurement of X can be interpreted as an information channel with input space $\{i=1,2,\cdots\}$ and output space $(\Lambda,\mathscr{B}(\Lambda))$ in the *a priori* distribution $\{w_i\}$ such that if the input parameter is i then the output distribution is $\nu(dx|i)$. As well as every information channel, we can define Shannon's information $I[\nu,\{w_i\}]$ for this information channel as follows:

$$I[\nu,\{w_i\}] = \sum_i\int_\Lambda w_i\nu(dx|i)\log\left\{\frac{\nu(dx|i)}{\Sigma_i w_i\nu(dx|i)}\right\}. \tag{3.6}$$

Let $\mu(dx)=\mathrm{Tr}[X(dx)\rho]$ and $\nu_i(dx)=\nu(dx|i)$. Then we have

$$I[\nu,\{w_i\}] = \sum_i w_i \text{c-}S[\nu_i|\mu], \tag{3.7}$$

where $\text{c-}S[\nu_i|\mu]$ stands for the classical relative entropy of ν_i with respect to μ, i.e.,

$$\text{c-}S[\nu_i|\mu] = \int_\Lambda \nu_i(dx)\log\left(\frac{d\nu_i}{d\mu}\right)(x). \tag{3.8}$$

Since the quantity $I[\nu,\{w_i\}]$ depends only on X and ρ, we shall refer to it as the *classical amount of information* of the measurement of X with *a priori* state ρ and write $\text{c-}I[\rho,X]=I[\nu,\{w_i\}]$, i.e.,

$$\text{c-}I[\rho,X] = \sum_i w_i\ \text{c-}S[\mathrm{Tr}[X(\cdot)\rho_i]\,|\,\mathrm{Tr}[X(\cdot)\rho]], \tag{3.9}$$

where $\rho=\Sigma_i w_i\,\rho_i$ is the orthogonal decomposition into pure states. By the classical theorem (see Ref. 16, p. 11), we have

$$\text{c-}I[\rho,X]\geqslant 0, \tag{3.10}$$

for all semiobservable X and states ρ.

IV. GENERALIZED GROENEWOLD–LINDBLAD INEQUALITY

In what follows, we shall fix a measuring process $\mathbf{M}=\langle\mathscr{K},\tilde{X},\sigma,U\rangle$ of a semiobservable X on a Hilbert space $\mathscr{H}$ with value space $(\Lambda,\mathscr{B}(\Lambda))$ and the corresponding CP instrument $\mathscr{I}$. A family $\{\rho_x;x\in\Lambda\}$ of *a posteriori* states with respect to an *a priori* state ρ is called *pure* if ρ_x is a pure state for almost all $x\in\Lambda$ with respect to $\mathrm{Tr}[\mathscr{I}(dx)\rho]$. A measuring process $\mathbf{M}$ is called *quasicomplete* if it satisfies the following condition.

(QC) For any pure state ρ, a family $\{\rho_x;x\in\Lambda\}$ of *a posteriori* states with respect to ρ is pure.

A measuring process $\mathbf{M}$ is called *complete* if a family of *a posteriori* states is pure for any *a priori* states. To justify the terminology, consider the conventional repeatable measuring process of a discrete observable $X=\Sigma_i x_i P_i$. In this case, the corresponding CP instrument $\mathscr{I}$ is of the form $\mathscr{I}(B)\rho = \Sigma\{P_i\,\rho P_i;x_i\in B\}$. Thus the *a posteriori* state ρ_x for $X=x_i$, written as $\rho(x_i)$, is of the form $\rho(x_i) = (1/\mathrm{Tr}[P_i\,\rho])P_i\,\rho P_i$. Then it is easy to see that this measuring process is quasicomplete and it is complete if and only if the spectrum of X is simple, i.e., every P_i is one dimensional.

Theorem 1: If the measuring process **M** is quasicomplete then the amount of information is not less than the classical amount of information, i.e.,

$$\mathrm{I}[\rho,\mathbf{M}] \geqslant \text{c-I}[\rho,X], \tag{4.1}$$

for any *a priori* state ρ with $\mathrm{S}[\rho] < \infty$.

Proof: Let $\mu(dx) = \mathrm{Tr}[\mathscr{I}(dx)\rho]$. Let $\rho = \Sigma_i w_i \rho_i$ be the orthogonal decomposition of ρ into pure states. Then we have

$$\mathrm{S}[\rho] = -\sum_i w_i \log w_i.$$

Let $\nu_i(dx)$ be the probability measure defined by $\nu_i(dx) = \mathrm{Tr}[\mathscr{I}(dx)\rho_i]$. Then $\nu_i(dx) = \mathrm{Tr}[X(dx)\rho_i]$ and $\nu_i \ll \mu$ for all i. Let $\{\rho_i(x); x \in \Lambda\}$ be a family of *a posteriori* states with respect to ρ_i. By the assumption $\rho_i(x)$ is pure for μ-almost all x and hence we can assume that $\rho_i(x)$ is pure for all $x \in \Lambda$ without any loss of generality. For any $B \in \mathscr{B}(\Lambda)$, we have

$$\begin{aligned}
\mathscr{I}(B)\rho &= \sum_i w_i \mathscr{I}(B)\rho_i \\
&= \sum_i w_i \int_B \rho_i(x)\mathrm{Tr}[\mathscr{I}(dx)\rho_i] \\
&= \sum_i w_i \int_B \rho_i(x)\nu_i(dx) \\
&= \sum_i w_i \int_B \rho_i(x)\Big(\frac{d\nu_i}{d\mu}\Big)(x)\mu(dx) \\
&= \int_B \Big\{\sum_i w_i \Big(\frac{d\nu_i}{d\mu}\Big)(x)\rho_i(x)\Big\}\mu(dx).
\end{aligned}$$

Thus $\rho_x = \Sigma_i w_i(d\nu_i/d\mu)(x)\rho_i(x)$ defines a family $\{\rho_x; x \in \Lambda\}$ of *a posteriori* states with respect to ρ. Let $f_i(x) = w_i(d\nu_i/d\mu)(x)$ for all i. By Ref. 15, Corollary, p. 247, we have, for all $x \in \Lambda$,

$$\begin{aligned}
\mathrm{S}[\rho_x] &= \mathrm{S}\Big[\sum_i f_i(x)\rho_i(x)\Big] \\
&\leqslant \sum_i \mathrm{S}[f_i(x)\rho_i(x)] \\
&= -\sum_i \mathrm{Tr}[f_i(x)\rho_i(x) \log(f_i(x)\rho_i(x))] \\
&= -\sum_i f_i(x) \log f_i(x) \\
&= -\sum_i w_i\Big(\frac{d\nu_i}{d\mu}\Big)(x) \log w_i\Big(\frac{d\nu_i}{d\mu}\Big)(x) \\
&= -\sum_i \Big(\frac{d\nu_i}{d\mu}\Big)(x) w_i \log w_i \\
&\quad -\sum_i w_i\Big(\frac{d\nu_i}{d\mu}\Big)(x) \log\Big(\frac{d\nu_i}{d\mu}\Big)(x).
\end{aligned}$$

Thus we have

$$\begin{aligned}
&\int_\Lambda \mathrm{S}[\rho_x]\mu(dx) \\
&\leqslant -\sum_i w_i \log w_i \\
&\quad -\sum_i w_i \int_\Lambda \nu_i(dx) \log\Big(\frac{d\nu_i}{d\mu}\Big)(x) \\
&= \mathrm{S}[\rho] - \sum_i w_i \text{c-S}[\nu_i|\mu].
\end{aligned}$$

Consequently, if $\mathrm{S}[\rho] < \infty$ then

$$\mathrm{S}[\rho] - \int_\Lambda \mathrm{S}[\rho_x]\mathrm{Tr}[\mathscr{I}(dx)\rho] \geqslant \sum_i w_i \text{c-S}[\nu_i|\mu],$$

whence,

$\mathrm{I}[\rho,\mathbf{M}] \geqslant \text{c-I}[\rho,X]$. Q.E.D.

Theorem 2: A measuring process **M** is quasicomplete if and only if the amount of information $\mathrm{I}[\rho,\mathbf{M}]$ is non-negative for any *a priori* state ρ with $\mathrm{S}[\rho] < \infty$.

Proof: Suppose that a given measuring process **M** satisfies $\mathrm{I}[\rho,\mathbf{M}] \geqslant 0$ for any *a priori* state ρ with finite entropy. Suppose that ρ is a pure state. Then $\mathrm{S}[\rho] = 0$ and hence

$$-\int_\Lambda \mathrm{S}[\rho_x]\mathrm{Tr}[\mathscr{I}(dx)\rho] = \mathrm{I}[\rho,\mathbf{M}] \geqslant 0.$$

Since $\mathrm{S}[\rho_x] \geqslant 0$, we have $\mathrm{S}[\rho_x] = 0$ for almost all $x \in \Lambda$ with respect to $\mathrm{Tr}[\mathscr{I}(dx)\rho]$. Thus the family of *a posteriori* state $\{\rho_x; x \in \Lambda\}$ with respect to ρ is pure, so that **M** is quasicomplete. The converse part of the assertion follows from Theorem 1 and inequality (3.10). Q.E.D.

Therefore, we have proved that a necessary and sufficient condition for the generalized Groenewold–Lindblad inequality (GL) is the quasicompleteness (QC) of the measuring process.

V. VON NEUMANN'S APPROXIMATE POSITION MEASUREMENTS

In Ref. 1, pp. 442–445, von Neumann considers the following measuring processes that measure the position observable approximately. The measured system and the apparatus system are one-dimensional systems described by Hilbert spaces $\mathscr{H} = \mathrm{L}^2(\mathbf{R})$ and $\mathscr{K} = \mathrm{L}^2(\mathbf{R})$, respectively; their wave function will be denoted by $\psi(x)$ and $\eta(y)$. The interaction is described by the Hamiltonian H of the form $H = -ix(\partial/\partial y)$. The measurement is carried out by the interaction from time 0 to 1. The pointer position of the apparatus system is the position observable Y on the Hilbert space $\mathscr{K}$. Assume that the prepared state of the apparatus system is a pure state $\sigma = |\xi\rangle\langle\xi|$ such that $\xi(y)$ is bounded. Thus we have a measuring process $\mathbf{M} = \langle \mathscr{K}, Y, |\xi\rangle\langle\xi|, U\rangle$, where $U = \exp(-iH)$.

By computations in Ref. 1, p. 443, we have

$$U|\psi(x)\eta(y)\rangle = |\psi(x)\eta(y-x)\rangle. \tag{5.1}$$

The corresponding CP instrument $\mathscr{I}$ is obtained by Eq. (2.2) as follows:

$$\mathscr{I}(dy)\rho = \mathrm{E}_{\mathscr{K}}[(1\otimes Y(dy))U(\rho\otimes|\xi\rangle\langle\xi|)U^*], \tag{5.2}$$

for all $\rho \in \mathscr{T}(\mathscr{H})$. Let A_y be the multiplication operator on $\mathscr{H} = \mathrm{L}^2(\mathbf{R})$ such that $(A_y\psi)(x) = \xi(y-x)\psi(x)$ for all

$\psi \in \mathscr{H}$. Then, we have, for all $\psi \in \mathscr{H}$, $a \in \mathscr{L}(\mathscr{H})$, and $B \in \mathscr{B}(\mathbf{R})$,

$$\begin{aligned}
\langle \psi | \mathscr{I}(B)^* a | \psi \rangle
&= \mathrm{Tr}[a \mathscr{I}(B)(|\psi\rangle\langle\psi|)] \\
&= \mathrm{Tr}[(a \otimes Y(B)) U(|\psi\rangle\langle\psi| \otimes |\xi\rangle\langle\xi|) U^*] \\
&= \int_{\mathbf{R}} \int_{\mathbf{R}} \langle \psi(x) \xi(y) | \\
&\quad \times U^*(a \otimes Y(B)) U | \psi(x) \xi(y) \rangle dx\, dy \\
&= \int_{B} \int_{\mathbf{R}} \langle \xi(y-x) \psi(x) | \\
&\quad \times a \otimes 1 | \xi(y-x) \psi(x) \rangle dx\, dy \\
&= \int_B \langle A_y \psi | a | A_y \psi \rangle dy = \langle \psi | \int_B A_y^* a A_y \, dy | \psi \rangle. \qquad (5.3)
\end{aligned}$$

Thus,

$$\mathscr{I}(B)\rho = \int_B A_y \rho A_y^* \, dy, \qquad (5.4)$$

for all $\rho \in \mathscr{T}(\mathscr{H})$ and $B \in \mathscr{B}(\mathbf{R})$. Now, we have shown that the CP-instrument $\mathscr{I}$ corresponding to the measuring process **M** considered by von Neumann[1] is just the covariant instrument considered by Davies (see Ref. 7, Theorem 4.6.1; and see also Ref. 17, Theorem 4). By Ref. 7, Theorem 4.6.1, the corresponding semiobservable X is the *approximate position observable* such that

$$X(B)\psi(x) = \left(\int_{\mathbf{R}} \chi_B(y) |\xi(x-y)|^2 \, dy \right) \psi(x). \qquad (5.5)$$

It is easy to see that a family $\{\rho_y; y \in \mathbf{R}\}$ of *a posteriori* states with respect to an *a priori* state ρ is given by

$$\rho_y = (1/\mathrm{Tr}[A_y \rho A_y^*]) A_y \rho A_y^*. \qquad (5.6)$$

Thus, this measurement is, obviously, quasicomplete. By Theorem 2, we can conclude that von Neumann's approximate position measurement satisfies (GL), i.e., the generalized Groenewold–Lindblad inequality holds for every *a priori* state with finite entropy.

[1] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, translated by R. T. Beyer (Princeton U.P., Princeton, NJ, 1955).
[2] E. P. Wigner, "Die Messung Quantenmechanischer Operatoren," Z. Phys. **143**, 101 (1952).
[3] H. Araki and M. M. Yanase, "Measurement of quantum mechanical operators," Phys. Rev. **130**, 622 (1960).
[4] M. Ozawa, "Quantum measuring processes of continuous observables," J. Math. Phys. **25**, 79 (1984).
[5] E. Arthurs and J. L. Kelly Jr., "On the simultaneous measurement of a pair of conjugate observables," Bell Syst. Tech. J. **44**, 725 (1965).
[6] C. Y. She and H. Heffner, "Simultaneous measurement of noncummuting observables," Phys. Rev. **152**, 1103 (1966).
[7] E. B. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976).
[8] M. Ozawa, "Conditional expectation and repeated measurements of continuous quantum observables," Lect. Notes Math. **1021**, 518 (1983).
[9] M. Ozawa, "Conditional probability and *a posteriori* states in quantum mechanics," Publ. RIMS, Kyoto Univ. **21**, 279 (1985).
[10] M. Ozawa, "Concepts of conditional expectations in quantum theory," J. Math. Phys. **26**, 1948 (1985).
[11] E. B. Davies and J. T. Lewis, "An operational approach to quantum probability," Commun. Math. Phys. **17**, 239 (1970).
[12] H. J. Groenewold, "A problem of information gain by quantum measurements," Int. J. Theor. Phys. **4**, 327 (1971).
[13] N. Abramson, *Information Theory and Coding* (McGraw–Hill, New York, 1963).
[14] V. G. Lüders, "Über die Zustandsänderung durch den Messprozess," Ann. Phys. Leipzig **6**, 322 (1951).
[15] G. Lindblad, "An entropy inequality for quantum measurements," Commun. Math. Phys. **28**, 245 (1972).
[16] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*, translated by A. Feinstein (Holden–Day, San Francisco, 1964).
[17] E. B. Davies, "On the repeated measurement of continuous observables in quantum mechanics," J. Funct. Anal. **6**, 318 (1970).

Evaluation of a class of integrals that arise in certain path integrals

Patrick L. Nash
Division of Earth and Physical Sciences, The University of Texas at San Antonio, San Antonio, Texas 78285-0663

(Received 3 June 1985; accepted for publication 6 November 1985)

A commonly encountered n-dimensional integral associated with a relativistic quadratic Lagrangian is explicitly evaluated for arbitrary n. In the limit $n\rightarrow\infty$, this integral is given by the usual Van Vleck–Morette determinant. The main advantage of the present approach is that it is simple and direct.

I. INTRODUCTION

Let $y_k \leftrightarrow y_k^\alpha \in M_4$, where $\alpha, \beta,... = 1,2,3,4$ and $k = 0,1,...,N$. We set $y_0 = 0 = y_N$ and put

$$v_k^\alpha = y_k^\alpha - y_{k-1}^\alpha . \tag{1}$$

This paper is devoted to the explicit evaluation of the integral

$$F_N = N^2 \int_{-\infty}^{\infty} \prod_{n=1}^{N-1} (-i\pi^{-2}\, d^4 y_n) \times \exp\left\{ i \sum_{k=1}^{N} [\tilde{v}_k g v_k + N^{-1}\kappa \tilde{y}_k A v_k] \right\}, \tag{2}$$

where $\sim$ denotes the transpose, $g\leftrightarrow g_{\alpha\beta}$ is the metric tensor on M_4,

$$g_{\alpha\beta} = \mathrm{diag}(1,1,1,-1), \tag{3}$$

$A\leftrightarrow A_{\alpha\beta} = -A_{\beta\alpha}$ is a covariant rank-2, constant antisymmetric space-time tensor, and κ is a constant. This type of integral arises in the phase space path integral evaluation of the Green's function associated to the Dirac equation for an electron moving in a constant background electromagnetic field.[1,2] Heretofore this expression has only been evaluated in the limit $N\rightarrow\infty$ and by altogether different techniques[3–5] than employed here.

In the next section we explicitly evaluate F_N for arbitrary N. The technique that we utilize is quite elementary, and it is gratifying to find a simple solution to a simple problem, viz., a Feynman path integral over a quadratic Lagrangian. The last section is devoted to an application of our results.

II. EVALUATION OF F_N

For each $N = 1,2,...$, we consider $y_k \in M_4$, $k = 0,1,...,N+1$ with $y_0 = 0 = y_{N+1}$. The y_k, $k = 1,...,N$, are our integration variables. We define, for each N, a real $N\times N$ matrix $W_N \leftrightarrow W_{Nkh}$ according to

$$v_k^\alpha = W_{Nkh}\, y_h^\alpha , \tag{4}$$

where the Einstein summation convention over repeated indices is operative. The W_N has matrix elements equal to $+1$ along the main diagonal, -1 directly below the main diagonal, and zero elsewhere. We put (the tilde denotes transpose)

$$T_N = W_N + \tilde{W}_N , \tag{5}$$

and note that

$$\sum_{k=1}^{N+1} \tilde{v}_k g v_k = y_h^\alpha g_{\alpha\beta} T_{Nhk}\, y_k^\beta \equiv \tilde{y}(g\otimes T_N) y . \tag{6}$$

In addition we write

$$\sum_{k=1}^{N+1} \tilde{y}_k A v_k = \tilde{y}(A\otimes W_N) y . \tag{7}$$

In this notation we find that F_N of Eq. (2) is given by

$$F_N = N^2(-i\pi^{-2})^{N-1}\pi^{2(N-1)} \times [\det\{-i(g\otimes T_{N-1} + N^{-1}\kappa A\otimes W_{N-1})\}]^{-1/2},$$

where we have used the well-known relation $\int d^n x\, e^{-\tilde{x}Ax} = \pi^{n/2}[\det A]^{-1/2}$. We factor out $-ig\otimes I$, where I denotes the $(N-1)\times(N-1)$ unit matrix, and use $\det(-ig\otimes I) = (-i)^{2(N-1)}$ to obtain

$$F_N = N^2[\det(\delta\otimes T_{N-1} + N^{-1}\kappa g^{-1}A\otimes W_{N-1})]^{-1/2}, \tag{8}$$

where $\delta\leftrightarrow\delta_\beta^\alpha$. The determinant in Eq. (8) may be evaluated as follows: We put $\alpha = (N+1)^{-1}\kappa$ and consider

$$\begin{aligned}\det(\delta\otimes T_N + \alpha g^{-1}A\otimes W_N) &= \det[\delta\otimes(W_N + \tilde{W}_N) + \alpha g^{-1}A\otimes W_N]\\ &= \det(\delta\otimes W_N)\det[\delta\otimes W_N^{-1}\tilde{W}_N + (\delta + \alpha g^{-1}A)\otimes I]\\ &= \det[\delta\otimes W_N^{-1}\tilde{W}_N + (\delta+\alpha g^{-1}A)\otimes I],\end{aligned}$$

where I denotes the $N\times N$ unit matrix, and we have used $\det(W_N) = 1$. Clearly, once the characteristic polynomial of $W_N^{-1}\tilde{W}_N$ is determined, this determinant may be explicitly evaluated. We shall proceed with a straightforward calculation of this polynomial.

In order to calculate W_N^{-1}, we set $W_N = I - V_N$. From the definition of W_N, we see that V_N is nilpotent, verifying $(V_N)^N = 0$. In terms of V_N,

$$W_N^{-1} = \sum_{n=0}^{\infty} (V_N)^n = \sum_{n=0}^{N-1} (V_N)^n .$$

Therefore, as the reader may readily verify, W_N^{-1} is given by the lower triangular matrix with matrix elements equal to $+1$ on and below the principal diagonal, and zero elsewhere. We find that $W_N^{-1}\tilde{W}_N$ is the matrix whose elements are equal to $+1$ in the first column, -1 directly above the principal diagonal, and equal to zero elsewhere. We denote the characteristic polynomial of $W_N^{-1}\tilde{W}_N$ by $D_N(\lambda)$, which is defined according to

 0022-2488/86/030764-03$02.50

$$D_N(\lambda) = \det(W_N^{-1}\widetilde{W}_N - \lambda I) . \quad (9)$$

We expand this determinant in terms of the elements in the last column and deduce the simple recurrence relation

$$D_N = -\lambda D_{N-1} + 1 . \quad (10)$$

Hence,

$$D_N(\lambda) = \sum_{n=0}^{N} (-\lambda)^n = [1-(-\lambda)^{N+1}](1+\lambda)^{-1} . \quad (11)$$

Let λ_n, $n = 1,...,N$, denote the eigenvalues of $W_N^{-1}\widetilde{W}_N$, $D_N(\lambda_n) = 0$. In terms of these eigenvalues we may express D_N as a product

$$D_N(\lambda) = D_N(0) \prod_{n=1}^{N} (\lambda - \lambda_n)(0-\lambda_n)^{-1} = \prod_{n=1}^{N} (\lambda_n - \lambda)\lambda_n^{-1} ,$$

or

$$D_N(\lambda) = \prod_{n=1}^{N} (\lambda_n - \lambda) , \quad (12)$$

since

$$\prod_{n=1}^{N} \lambda_n = \det W_N^{-1}\widetilde{W}_N = 1 = D_N(0) .$$

We may now explicitly evaluate F_N of Eq. (8), which we rewrite as ($\alpha = \kappa/N$)

$$F_N = N^2[\det\{\delta \otimes W_{N-1}^{-1}\widetilde{W}_{N-1} + (\delta + \alpha g^{-1}A)\otimes I\}]^{-1/2} . \quad (13)$$

We denote the eigenvalues of $g^{-1}A$ by $\pm E$, $\pm iB$, and find that

$$\begin{aligned}\det[\delta \otimes W_{N-1}^{-1}\widetilde{W}_{N-1} &+ (\delta + \alpha g^{-1}A)\otimes I]\\ &= \prod_{n=1}^{N-1} (\lambda_n + 1 + \alpha E)(\lambda_n + 1 - \alpha E)\\ &\quad\times(\lambda_n + 1 + i\alpha B)(\lambda_n + 1 - i\alpha B)\\ &= D_{N-1}(-1-\alpha E)D_{N-1}(-1+\alpha E)\\ &\quad\times D_{N-1}(-1-i\alpha B)D_{N-1}(-1+i\alpha B)\\ &= -(\alpha^4E^2B^2)^{-1}[1-(1+\alpha E)^N]\\ &\quad\times[1-(1-\alpha E)^N][1-(1+i\alpha B)^N]\\ &\quad\times[1-(1-i\alpha B)^N] .\end{aligned}$$

In this expression the last equality follows from Eq. (11), while the third equality is a consequence of Eq. (12). Therefore we arrive at our main result

$$\begin{aligned}F_N = (EB\kappa^2)[&-\{1-(1+\kappa E/N)^N\}\\ &\times\{1-(1-\kappa E/N)^N\}\{1-(1+i\kappa B/N)^N\}\\ &\times\{1-(1-i\kappa B/N)^N\}]^{-1/2} . \quad (14)\end{aligned}$$

Passing to the limit $N \to \infty$, we find that

$$F = \lim_{n\to\infty} F_N = \kappa^2 EB\,[4\sinh(\kappa E/2)\sin(\kappa B/2)]^{-1} . \quad (15)$$

III. APPLICATION

Let $F_{\alpha\beta} = A_{\beta,\alpha} - A_{\alpha,\beta}$, where $A_\alpha = 1/2x^\beta F_{\beta\alpha}$, describe a constant electromagnetic field. In his classic realization of the Green's function for a Dirac electron a constant electromagnetic field, Schwinger[6] employed a so-called proper time method to evaluate the propagator $\langle x'',s|x',0\rangle = \langle x''|e^{-is\Pi^2}|x'\rangle$. Here $\Pi^2 = \Pi g^{-1}\widetilde{\Pi} = \Pi_\alpha \Pi^\alpha$, where $\Pi_\alpha = p_\alpha - eA_\alpha$ and $e = -|e|$ is the charge of the electron. As is very well known, this propagator may also be evaluated using the phase space path integral technique in which one writes

$$\langle x'',s|x',0\rangle = \lim_{N\to\infty} \langle x''|(1-isN^{-1}\Pi^2)^N|x'\rangle ,$$

and inserts complete set of states to obtain

$$\begin{aligned}\langle x'',s|x',0\rangle &= \lim_{N\to\infty} (2\pi)^{-4N}\int d^4p_N \prod_{n=1}^{N-1} d^4p_n\, d^4x_n\\ &\quad\times\exp\Big\{i\sum_{k=1}^{N}[p_k(x_k - x_{k-1}) - sN^{-1}\Pi_k g^{-1}\widetilde{\Pi}_k]\Big\}\\ &= \lim_{N\to\infty} \frac{1}{i^N}\Big(\frac{N}{4\pi s}\Big)^{2N}\int \prod_{n=1}^{N-1} d^4x_n\\ &\quad\times\exp\Big\{\frac{is}{4N}\sum_{k=1}^{N}\Big[\tilde{v}'_k g v'_k + \frac{2es}{N}\tilde{x}_k F v'_k\Big]\Big\} ,\end{aligned}$$

where in the last step we have integrated over momenta and set $x_k - x_{k-1} = v'_k$. Here it is understood that $x_N = x''$ and $x_0 = x'$. We denote the argument of the exponential as $iS[x]/2$, where S is the action associated with the Lagrangian $L = \tilde{\dot{x}}g\dot{x}/2 + e\tilde{x}F\dot{x}$ and the path $\{x_k\}$. To further simplify this expression for $\langle x'',s|x',0\rangle$, we change integration variables to the Feynman variables $y = x - \bar{x}$, where $\bar{x}$ is the solution of the Euler–Lagrange equations verifying $\bar{x}(0) = x'$ and $\bar{x}(s) = x''$. Using $S[x] = S[\bar{x}] + S[y]$ (the y_k vanish at the boundaries of the interval, that is, $y_0 = 0 = y_N$) we find that the propagator is given by

$$\begin{aligned}\langle x'',s|x',0\rangle &= \exp\frac{\{iS[\bar{x}]\}}{2} \lim_{N\to\infty} \frac{N^2}{i(4\pi s)^2}\int \prod_{n=1}^{N-1}\frac{(-id^4y_n)}{\pi^2}\\ &\quad\times\exp\Big\{i\sum_{k=1}^{N}\Big[\tilde{v}_k g v_k + \frac{2es}{N}\tilde{x}_k F v_k\Big]\Big\} .\end{aligned}$$

This is of the form of Eq. (2) with $\kappa = 2es$. Hence

$$\langle x'',s|x',0\rangle = \frac{e^{iS[\bar{x}]/2}}{i(4\pi s)^2}\lim_{N\to\infty} F_N = e^{iS[\bar{x}]/2}\frac{F}{i(4\pi s)^2} ,$$

which, using Eq. (15), we record as

$$\langle x'',s|x',0\rangle = \frac{e^{iS[\bar{x}]/2}}{i(4\pi s)^2}\frac{esE}{\sinh(esE)}\frac{esB}{\sin(esB)} . \quad (16)$$

For completeness we remark that $S[\bar{x}]$ can easily be computed, and is given by

$$S[\bar{x}] = e\tilde{x}'Fx'' + \tfrac{1}{2}\tilde{r}\{(eF)\coth(eFs)\}r , \quad (17)$$

where $r = x'' - x'$. An expression equivalent to Eqs. (16) and (17) was first given by Schwinger [see Eq. (3.20) of Ref. 6] using a proper time formalism.

[1]A. M. Chebotarev and V. P. Maslov, "Feynman path integrals," in *Lecture Notes in Physics*, Vol. 106, edited by S. Albeverio, Ph. Combe, R. Høegh-Krohn, G. Rideau, M. Sirugue-Collin, M. Sirugue, and R. Stora (Springer, New York, 1979), pp. 58–72.
[2]L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley–Interscience, New York, 1981).
[3]B. S. DeWitt, Rev. Mod. Phys. **29**, 377 (1957).
[4]G. J. Papadopoulos and J. T. Devreese, Phys. Rev. D **13**, 227 (1976).
[5]G. J. Papadopoulos, *Path Integrals*, edited by G. J. Papadopoulos and J. T. Devreese (Plenum, New York, 1978), pp. 85–169.
[6]J. Schwinger, Phys. Rev. **82**, 664 (1951).

Path integration of the time-dependent forced oscillator with a two-time quadratic action

Tian Rong Zhang and Bin Kang Cheng[a)]
Department of Physics and Center for Relativity, The University of Texas at Austin, Austin, Texas 78712

(Received 25 July 1985; accepted for publication 31 October 1985)

Using the prodistribution theory proposed by DeWitt-Morette [C. DeWitt-Morette, Commun. Math. Phys. **28**, 47 (1972); C. DeWitt-Morette, A. Maheshwari, and B. Nelson, Phys. Rep. **50**, 257 (1979)], the path integration of a time-dependent forced harmonic oscillator with a two-time quadratic action has been given in terms of the solutions of some integrodifferential equations. We then evaluate explicitly both the classical path and the propagator for the specific kernel introduced by Feynman in the polaron problem. Our results include the previous known results as special cases.

I. INTRODUCTION

The propagator of a particle from position to position in nonrelativistic quantum mechanics can be expressed symbolically as

$$K(x,T;x_0,0) = \int_{x(0)=x_a}^{x(T)=x_b} \exp\left\{\left(\frac{i}{\hbar}\right) S[x]\right\} D[x], \qquad (1)$$

where the symbol $D[x]$ implies that integrations are performed over all possible paths x from $x(0)=x_a$ to $x(T)=x_b$. In this path-integral theory of polarons,[1] Feynman introduced for the first time a two-time quadratic action functional. The problem is to evaluate the path integral (1) for an action of the form

$$S[x] = \int_0^T \left[\frac{1}{2} m(\dot{x}^2(t) - \omega_0^2 x^2(t)) + f(t)x(t)\right] dt - \int_0^T dt \int_0^T G(t,s)[x(t)-x(s)]^2 ds \qquad (2)$$

with the symmetric kernel $G(t,s)$ and the time-dependent force $f(t)$. Following Feynman's polygonal approach,[2] the path integral (1) has been evaluated exactly for the two-time quadratic action[3] with $\omega_0 = f(t) = 0$, for the harmonic oscillator with a two-time quadratic action[4] and for the constant forced harmonic oscillator with a two-time quadratic action.[5]

Using the prodistribution theory proposed by DeWitt-Morette,[6,7] Maheshwari[8] and Bosco[9] calculated the path integration of the harmonic oscillator with different memory terms. Recently, Khandekar *et al.*[10] have derived a general formula of a two-time quadratic action with generalized memory by the same technique. In this paper, we use the Cameron–Martin transformation[7] to derive the general formula for the propagator of a time-dependent forced harmonic oscillator with a two-time quadratic action. For the specific kernel $G(t,s)$ used by Feynman in the polaron problem,[1] we are able to calculate explicitly the classical path and the propagator.

a) On leave from Departamento de Física, Universidade Federal do Parana, Caixa Postal 19.081, 80.000 Curitiba, Brazil.

II. GENERAL SOLUTION

The propagator of the action (2) can be written

$$K(x_b,T;x_a,0) = \int_{Z_+} d\omega_+^w(z) \exp\left\{\frac{i}{\hbar} V\left[x_b + \sqrt{\frac{\hbar}{m}} z\right]\right\} \times \delta[x_b + (\sqrt{\hbar/m}) z(0) - x_a], \qquad (3)$$

with

$$V[x] = \int_0^T \left[-\frac{1}{2} m\omega_0^2 x^2(t) + f(t)x(t)\right] dt - \int_0^T dt \int_0^T G(t,s)[x(t)-x(s)]^2 ds. \qquad (4)$$

Here $\omega_+^w(z)$ is the Wiener prodistribution on Z_+ with covariance

$$A_+^w(t,s) = \theta(t-s)(T-t) + \theta(s-t)(T-s), \qquad (5)$$

and Z_+ is the space of continuous paths $\{z(t)\}$ such that $z(T)=0$. The step function θ is equal to one for positive arguments, and zero otherwise.

Let Y_+ be a copy of Z_+ and make change of variable from $z \equiv Z_+$ to $y \equiv Y_+$ defined by

$$x_b + (\sqrt{\hbar/m})\, z(t) = u(t) + (\sqrt{\hbar/m})\, y(t),$$

where $u(t)$ is the classical path from $(0,x_a)$ to (T,x_b). Using the classical equation of motion, we obtain

$$K(x_b,T;x_a,0) = \exp\left[\frac{i}{\hbar} S_{\rm cl}\right] \int_{Y_+} d\omega_+^w(y) \times \exp\left\{-\frac{1}{2} \mathrm{im}\, \omega_0^2 \int_0^T y^2(t)\, dt - i\int_0^T dt \int_0^T G(t,s)[y(t)-y(s)]^2 ds\right\} \times \delta[(\sqrt{\hbar/m})\, y(0)], \qquad (6)$$

where

$$S_{\rm cl} = \int_0^T \left[\frac{1}{2} m(\dot{u}^2(t) - \omega_0^2 u^2(t)) + f(t)u(t)\right] dt - \int_0^T dt \int_0^T G(t,s)[u(t)-u(s)]^2 ds \qquad (7)$$

 0022-2488/86/030767-05$02.50

is the classical action function along the classical path u.

Now by using the Cameron–Martin formula the integrand in (6) can be absorbed into a new prodistribution $\omega_+(y)$ on Ω_+ (also a copy of Z_+). Finally, by combining the δ-function with the integrator $d\omega_+(y)$, the propagator can be reexpressed as an integral over the space Ω of paths vanishing both at 0 and T. The Gaussian prodistribution $\omega(y)$ on the space Ω, which is Leray associated (Ref. 7, p. 283) to the Gaussian prodistribution $\omega_+(y)$, is normalized to

$$\omega(\Omega) = \int_{\Omega_+} d\omega_+(y)\delta[y(0)] = (2\pi i)^{-1/2}|\det \mathbf{A}_+(t,s)|^{-1/2}. \quad (8)$$

Its covariance $\mathbf{A}(t,s)$ is the kernel of Jacobi operator, which, for the case of two-time quadratic action (see the Appendix for derivations), is

$$\frac{1}{4}m\left[\left(\frac{d}{dt}\right)^2 + \omega_0^2\right]\mathbf{A}(t,s) - \left[\int_0^T G(\tau,t)\,d\tau\right]\mathbf{A}(t,s) + \int_0^T G(t,\tau)\mathbf{A}(\tau,s)\,d\tau = \delta(t-s),$$
$$\text{with } \mathbf{A}(0,s) = \mathbf{A}(T,s) = 0. \quad (9)$$

Furthermore, the classical path $u(t)$ satisfies the following integrodifferential equation (see the Appendix)

$$\frac{1}{4}m(\ddot{u} + \omega_0^2 u) + \int_0^T G(t,s)[u(t) - u(s)]\,ds\,\frac{1}{4}(t),$$
$$u(0) = x_0, \quad u(T) = x. \quad (10)$$

With the help of (8)–(10), the propagator has the form

$$K(x_b,T;x_a,0) = (m/2\pi i\hbar)^{1/2}|\det \mathbf{A}(0,T)|^{-1/2} \times \exp\left\{\frac{i}{\hbar}\left[mu(t)\dot{u}(t)|_0^T + \frac{1}{2}\int_0^T f(t)u(t)\,dt\right]\right\}. \quad (11)$$

In order to evaluate the determinent of $\mathbf{A}(0,T)$, we consider the function $\mathbf{A}_\mu(0,T)$ satisfying the equation

$$\frac{1}{4}m\left[\left(\frac{d}{dt}\right)^2 + \Omega_0^2\right]\mathbf{A}_\mu(t,s) + \mu\int_0^T G(t,\tau)\,\mathbf{A}_\mu(\tau,s)\,d\tau = \delta(t-s) \quad (12)$$

and boundary conditions

$$\mathbf{A}_\mu(0,s) = \mathbf{A}_\mu(T,s) = 0, \quad (13)$$

with

$$\Omega_0^2 = \omega_0^2 + \int_0^T G(\tau,t)\,dt,$$

and μ a parameter. By using the well-known formula

$$\delta(\ln\det \mathbf{A}_\mu) = \mathrm{Tr}(\mathbf{A}_\mu^{-1}\,\delta\mathbf{A}_\mu), \quad (14)$$

we have

$$|\det \mathbf{A}_1| = |\det \mathbf{A}_0|\exp\left\{\int_{\mu=0}^1 \mathrm{Tr}(\mathbf{A}_\mu^{-1}\,\delta\mathbf{A}_\mu)\right\}. \quad (15)$$

Taking the variation of (12) gives

$$\mathbf{A}_\mu^{-1}\,\delta\mathbf{A}_\mu = -\int_0^T G(t,\tau)\left\{\int_0^1 \mathbf{A}_\mu(\tau,s)\,\delta\mu\right\}d\tau. \quad (16)$$

Finally, we obtain

$$|\det \mathbf{A}(0,T)| = |\det \mathbf{A}_0(0,T)| \times\exp\left\{-\int_0^T dt\int_0^1 d\mu\int_0^T d\tau[G(t,\tau)\mathbf{A}_\mu(\tau,t)]\right\}. \quad (17)$$

Here, $|\det \mathbf{A}_0(0,T)| = \sin\Omega_0 T$ is the well-known result for the harmonic oscillator.

Therefore, the propagator has been given in terms of the classical path $u(t)$ and the kernel $\mathbf{A}_\mu(t,s)$ of a generalized μ-parametrized Jacobi operator. But the integrodifferential equations (10) and (12) are usually very difficult to solve for the general $G(t,s)$.

III. A SPECIAL CASE

For the kernel $G(t,s)$ of the polaron problem,[1] we have

$$G(t,s) = \tfrac{1}{4}m\Omega^2\omega^2\phi(t,s),$$

with (18)

$$\phi(t,s) = \cos[\omega(\tfrac{1}{2}T - |t-s|)]/2\omega\sin(\tfrac{1}{2}\omega T).$$

Let $\Omega_0^2 = \Omega^2 + \omega_0^2$, then Eqs. (12) and (10) become

$$\left[\left(\frac{d}{dt}\right)^2 + \Omega_0^2\right]\mathbf{A}_\mu(t,s) - \mu\Omega^2\omega^2\int_0^T \phi(t,\tau)\mathbf{A}_\mu(\tau,s)d\tau = \delta(t,s), \quad \mathbf{A}_\mu(0,s) = \mathbf{A}_\mu(T,s) = 0, \quad (19)$$

and

$$\ddot{u} + \Omega_0^2 u = \Omega^2\omega^2\int_0^T \phi(t,s)u(s)ds + \frac{f(t)}{m},$$
$$u(0) = x_0, \quad u(T) = x. \quad (20)$$

Since $\phi(t,s)$ is the kernel of operator $(d/dt)^2 + \omega^2$, after acting the operator on both sides, Eqs. (19) and (20) become

$$\left[\left(\frac{d}{dt}\right)^2 + \Omega_+^2\right]\left[\left(\frac{d}{dt}\right)^2 + \Omega_-^2\right]\mathbf{A}_\mu(t,s) = -\left[\left(\frac{d}{dt}\right)^2 + \omega^2\right]\delta(t,s), \quad (21)$$

and

$$\left[\left(\frac{d}{dt}\right)^2 + \omega_+^2\right]\left[\left(\frac{d}{dt}\right)^2 + \omega_-^2\right]u(t) = \left[\left(\frac{d}{dt}\right)^2 + \omega^2\right]\frac{f(t)}{m}, \quad (22)$$

where

$$\Omega_\pm^2 = \tfrac{1}{2}(\Omega_0^2 + \omega^2) \pm \{[\tfrac{1}{2}(\Omega_0^2 - \omega^2)]^2 + \mu\Omega^2\omega^2\}^{1/2} \quad (23)$$

and

$$\omega_\pm^2 = \tfrac{1}{2}(\Omega_0^2 + \omega^2) \pm \{[\tfrac{1}{2}(\Omega_0^2 - \omega^2)]^2 + \Omega^2\omega^2\}^{1/2} \quad (24)$$

The solutions of Eqs. (21) and (22) can be expressed as

$$A_\mu(t,s) = [1/(\Omega_-^2 - \Omega_+^2)] \times [(\omega^2 - \Omega_+^2)A_+(t,s) - (\omega^2 - \Omega_-^2)A_-(t,s)], \tag{25}$$

and

$$u(t) = [1/(\omega_-^2 - \omega_+^2)] \times [(\omega^2 - \omega_+^2)u_+(t) - (\omega^2 - \omega_-^2)u_-(t)], \tag{26}$$

where $A_\pm(t,s)$ and $u_\pm(t)$ satisfy the equations

$$\left[\left(\frac{d}{dt}\right)^2 + \Omega_\pm^2\right] A_\pm(t,s) = -\delta(t,s) \tag{27}$$

and

$$\left[\left(\frac{d}{dt}\right)^2 + \omega_\pm^2\right] u_\pm(t) = \frac{f(t)}{m}. \tag{28}$$

In order to get the boundary conditions for $A_\pm(t,s)$, we need to substitute the solutions (25) into the original equation (19). Then we obtain

$$A_\pm(T,s) = A_\pm(0,s),$$

$$(\omega^2 - \Omega_+^2)A_+(0,s) = (\omega^2 - \Omega_-^2)A_-(0,s), \tag{29}$$

$$\left(\frac{d}{dt}A_+(t,s) - \frac{d}{dt}A_-(t,s)\right)\Big|_{t=0} = \left(\frac{d}{dt}A_+(t,s) - \frac{d}{dt}A_-(t,s)\right)\Big|_{t=T}.$$

Then after lengthy but straightforward calculations, (17) becomes

$$|\det A(0,T)| = |\det A_0(0,T)| \times \exp\left\{-\int_0^T dt \int_0^1 d\mu \int_0^T d\tau [G(t,\tau)A_\mu(\tau,t)]\right\}$$
$$= |\det A_0(0,T)| \exp\left\{-\int_0^T dt \int_0^1 d\mu \times \left[\frac{\Omega^2\omega^2}{\Omega_+^2 - \Omega_-^2}(A_-(t,t) - A_+(t,t))\right]\right\}$$
$$= \left(\frac{(\omega_+^2 - \omega_-^2)\sin^2(\omega T/2)}{2D\sin(\omega_+ T/2)\sin(\omega_- T/2)}\right)^{1/2}, \tag{30}$$

where

$$D = \frac{(\omega_+^2 - \omega^2)}{\omega_+}\sin\left(\frac{\omega_- T}{2}\right)\cos\left(\frac{\omega_+ T}{2}\right) - \frac{(\omega_-^2 - \omega^2)}{\omega_-}\sin\left(\frac{\omega_+ T}{2}\right)\cos\left(\frac{\omega_- T}{2}\right). \tag{31}$$

The general solution of (22) can be written as

$$u(t) = a_+\sin(\omega_+ t) + b_+\cos(\omega_+ t) + a_-\sin(\omega_- t) + b_-\cos(\omega_- t) + u_p(t), \tag{32}$$

with the particular solution

$$u_p(t) = \frac{1}{m(\omega_-^2 - \omega_+^2)}\int_0^t f(s)\left[\frac{\omega^2 - \omega_+^2}{\omega_+}\sin\omega_+(t-s) - \frac{\omega^2 - \omega_-^2}{\omega_-}\sin\omega_-(t-s)\right] ds. \tag{33}$$

Substituting (32) into (20) and using the boundary conditions of $u(t)$ give

$$\begin{aligned}
&b_+ + b_- = x_0,\\
&a_+\sin\omega_+T + b_+\cos\omega_+T + a_-\sin\omega_-T + b_-\cos\omega_-T = x - u_p(T),\\
&\frac{a_+\sin\omega_+T - b_+(1-\cos\omega_+T)}{\omega_+^2 - \omega^2} + \frac{a_-\sin\omega_-T - b_-(1-\cos\omega_-T)}{\omega_-^2 - \omega^2} = -\frac{u_p^+(T)}{\omega_+^2 - \omega^2} - \frac{u_p^-(T)}{\omega_-^2 - \omega^2},\\
&\frac{\omega_+[a_+(1-\cos\omega_+T) + b_+\sin\omega_+T]}{\omega_+^2 - \omega^2} + \frac{\omega_-[a_-(1-\cos\omega_-T) + b_-\sin\omega_-T]}{\omega_-^2 - \omega} = \frac{\dot u_p^+(T)}{\omega_+^2 - \omega^2} + \frac{\dot u_p^-(T)}{\omega_-^2 - \omega^2},
\end{aligned} \tag{34}$$

with

$$u_p^\pm(t) = \frac{\pm 1}{m(\omega_-^2 - \omega_+^2)}\int_0^t f(s)\left(\frac{\omega^2 - \omega_\pm^2}{\omega_\pm}\right)\sin\omega_\pm(t-s)\, ds. \tag{35}$$

Solving the system of equations (34) gives

$$\begin{aligned}
b_\pm = &\pm\frac{1}{d}\left\{(x - x_0)\left(\omega_-\sin\frac{1}{2}\omega_-T\cos\frac{1}{2}\omega_+T - \omega_+\sin\frac{1}{2}\omega_+T\cos\frac{1}{2}\omega_-T\right)\right.\\
&\left.+\int_0^T\frac{f(s)}{m}\left[\cos\frac{1}{2}\omega_-T\cos\omega_+\left(\frac{1}{2}T - s\right) - \cos\frac{1}{2}\omega_+T\cos\omega_-\left(\frac{1}{2}T - s\right)\right]ds\right\}\\
&\pm\frac{2x_0\omega_\mp(\omega_-^2 - \omega_+^2)}{d(\omega_\mp^2 - \omega^2)}\cos\frac{1}{2}\omega_\pm T\sin\frac{1}{2}\omega_\mp T,
\end{aligned} \tag{36}$$

$$a_\pm = \pm(1/\sin\omega_\pm T)\{[(\omega_\pm^2 - \omega^2)/(\omega_+^2 - \omega_-^2)](x - x_0) - u_p^\pm(T) \pm (1 - \cos\omega_\pm T)b_\pm\}, \tag{37}$$

where

$$d = 2(\omega_+^2 - \omega_-^2)\left\{\frac{\omega_+\sin\frac{1}{2}\omega_+T\cos\frac{1}{2}\omega_-T}{\omega_+^2 - \omega^2} - \frac{\omega_-\sin\frac{1}{2}\omega_-T\cos\frac{1}{2}\omega_+T}{\omega_-^2 - \omega^2}\right\}. \tag{38}$$

Integrating (10), we have

$$\dot{u}(T)=\dot{u}(0)-\omega_0^2\int_0^T u(t)\,dt+\frac{1}{m}\int_0^T f(t)\,dt, \tag{39}$$

since $G(t,s)=G(s,t)$. Combining (11), (30), and (39), we obtain our principal result, namely the propagator, for the system (2),

$$K(x_b,T;x_a,0)=\left(\frac{m(\omega_+^2-\omega_-^2)\sin^2(\omega T/2)}{4\pi i\hbar D\sin(\omega_+T/2)\sin(\omega_-T/2)}\right)^{1/2}$$

$$\times\exp\left\{(i/2\hbar)\left[m(x-x_0)\,\dot{u}(0)-x\int_0^T[m\omega_0^2u(t)-f(t)]\,dt+\frac{1}{m}\int_0^T u(t)f(t)\,dt\right]\right\}. \tag{40}$$

Here we should mention that (40) is invalid when $\sin(\omega_+T/2)\sin(\omega_-T/2)=0$, which will be considered in the following paragraph. After a lengthy but straightforward calculation, we can show that (40) reduces to (36) of Ref. 4 for $f(t)=0$ and to (38) of Ref. 5 for $f(t)=f$.

IV. DISCUSSION (WHEN $\omega_0=0$)

For completeness, we consider the special case $\omega_0=0$, which implies $\omega_-=0$, and the classical path is of the form

$$u_0(t)=s_+\sin(\omega_+t)+c_+\cos(\omega_+t)+c_0+c_1t+u_p^0(t), \tag{41}$$

with

$$u_p^0(t)=\frac{1}{m\omega_+^2}\int_0^t f(s)\left[\frac{\Omega^2}{\omega_+}\sin\omega_+(t-s)+\omega^2(t-s)\right]ds, \tag{42}$$

and

$$s_+=\frac{\Omega^2}{2\omega_+^2}(x-x_0)\cot\frac{\omega_+T}{2}-\frac{\Omega^2}{2m\omega_+^3}\int_0^T f(t)\left[\cos\omega_+t-\cot\frac{\omega_+T}{2}\sin\omega_+t+1\right]dt, \tag{43}$$

$$c_+=\frac{-\Omega^2}{2\omega_+^2}(x-x_0)+\frac{\Omega^2}{2m\omega_+^3}\int_0^T f(t)\left[\cot\frac{\omega_+T}{2}(\cos\omega_+t-1)+\sin\omega_+t\right]dt, \tag{44}$$

$$c_0=\frac{\Omega^2}{2\omega_+^2}(x-x_0)-\frac{\Omega^2}{2m\omega_+^3}\int_0^T f(t)\left[\cot\frac{\omega_+T}{2}(\cos\omega_+t-1)+\sin\omega_+t\right]dt+x_0, \tag{45}$$

$$c_1=\frac{\omega^2}{\omega_+^2T}(x-x_0)-\frac{\omega^2}{m\omega_+^2}\int_0^T f(t)(T-t)\,dt. \tag{46}$$

Taking the limit value of (30) as $\omega_-\to0$, we obtain

$$K(x,T;x_0,0)=\left(\frac{m}{2\pi i\hbar T}\right)^{1/2}\cdot\frac{\omega_+\sin\frac12\omega T}{\omega\sin\frac12\omega_+T}\cdot\exp[(i/\hbar)S_{\rm cl}], \tag{47}$$

with

$$\begin{aligned}S_{\rm cl}&=\tfrac12 m(x-x_0)(s_+\omega_++c_1)\\&+\frac12\frac{x\Omega^2}{\omega_+^2}\int_0^T f(t)\,dt+\frac12 s_+\int_0^T f(t)(\sin\omega_+t)\,dt\\&+\frac12 c_+\int_0^T f(t)(\cos\omega_+t)\,dt+\frac12 c_0\int_0^T f(t)\,dt\\&+\frac12 c_1\int_0^T tf(t)\,dt\\&-\frac{1}{2m\omega_+^2}\int_0^T dt\int_0^t ds\,f(t)f(s)\\&\times[(\Omega^2/\omega_+)\sin\omega_+(t-s)-\omega^2(t-s)].\end{aligned} \tag{48}$$

When a memory term is present, the exact solution of a quadratic system given by (47) is not of the form

$$K(x,T;x_0,0)=\left|\det\frac{\partial^2S_{\rm cl}}{\partial x\partial x_0}\right|^{1/2}\cdot\exp\left[\frac{i}{\hbar}S_{\rm cl}\right].$$

Indeed here the Van Vleck–Morette determinant

$$\frac{\partial^2S_{\rm cl}}{\partial x\partial x_0}=-\frac{m\Omega^2}{2\omega_+}\cdot\frac{\cos\frac12\omega_+T}{\sin\frac12\omega_+T}-\frac{m\omega^2}{\omega_+^2T} \tag{49}$$

is different from the normalization in (47). For the case of $\omega_0=f(t)=0$, Eq. (40) reduces to (3.45) in Ref. 3 as we expect.

ACKNOWLEDGMENTS

We thank C. DeWitt-Morette for valuable discussions. One of us (B.K.C.) would like to thank C. DeWitt-Morette for kind hospitality at the Center for Relativity, University of Texas at Austin.

We would like to thank The Conselho Nacional de Desenvolviment Cientifico e Tecnologico, Cansello Nacional de Pesquisas (CNPq), Brazil, for financial support. This work was supported in part by the National Science Foundation, Grant No. PHY 84-04931.

APPENDIX: LAGRANGE OPERATOR AND JACOBI OPERATOR FOR THE SYSTEM WITH A TWO-TIME QUADRATIC ACTION

We derive the Lagrange equation and the Jacobi equation for the system with a two-time quadratic action by using the one-parameter variation method.[7,11] Let $\Lambda(\Omega)$ be the space of C^1 mapping x on $\Omega = (0,T) \subset R$ into R. Introducing a one parameter family of functions $\alpha(u,t) \in \Lambda(\Omega)$, where $u \in [0,1]$ such that $\alpha(0,t)$ is the classical path and $\alpha(1,t) = x(t)$, we can rewrite the action functional (2) as

$$S[\alpha] = \int_0^T \left\{ \frac{1}{2} m[\dot{\alpha}^2(u,t) - \omega_0^2 \alpha^2(u,t)] + f(t)\alpha(u,t) - 2\alpha^2(u,t) \int_0^T G(t,s)\, ds + 2\alpha(u,t) \times \int_0^T G(t,s)\, \alpha(u,s)\, ds \right\} dt, \quad \text{(A1)}$$

since $G(t,s) = G(s,t)$. Furthermore, we have

$$(S \circ \alpha)'(u) = m\dot{\alpha}(u,t) \frac{\partial \alpha(u,t)}{\partial u} \Big|_0^T + \int_0^T \Big[4 \int_0^T G(t,s)\alpha(u,s)\, ds - m\ddot{\alpha}(u,t) - m\omega_0^2 \alpha(u,t) + f(t) - 4\alpha(u,t) \int_0^T G(t,s) ds \Big] \times \frac{\partial \alpha(u,t)}{\partial u}\, dt, \quad \text{(A2)}$$

and

$$(S \circ \alpha)''(u) = \int_0^T \Big[4 \int_0^T G(t,s) \frac{\partial \alpha(u,t)}{\partial u}\, ds - m \frac{\partial \ddot{\alpha}(u,t)}{\partial u} - m\omega_0^2 \frac{\partial \alpha(u,t)}{\partial u} - 4 \frac{\partial \alpha(u,t)}{\partial u} \int_0^T G(t,s)\, ds \Big] \frac{\partial \alpha(u,t)}{\partial u}\, dt + \int_0^T \Big[4 \int_0^T G(t,s)\alpha(u,s)\, ds - m\dot{\alpha}(u,t) - m\omega_0^2 \alpha(u,t) + f(t) - 4\alpha(u,t) \times \int_0^T G(t,s)\, ds \Big] + \frac{\partial^2 \alpha(u,t)}{\partial u^2}\, dt. \quad \text{(A3)}$$

For variation, keeping the end point fixed,

$$\frac{\partial \alpha(u,0)}{\partial u} = \frac{\partial \alpha(u,T)}{\partial u} = 0,$$

(A2) reduces to

$$(S \circ \alpha)'(0) = \int_\Omega \Big[\hat{L}(t)\alpha(u,t) \frac{\partial \alpha(u,t)}{\partial u} \Big]_{u=0} dt,$$

where

$$\hat{L}(t) = 4 \int_0^T G(t,s)(\cdot)\, ds - m \left(\frac{d}{dt}\right)^2 - m\omega^2 + f(t) - 4 \int_0^T G(t,s)\, ds \quad \text{(A4)}$$

is an integrodifferential operator. By definition we obtain the Lagrange equation

$$\hat{L}(t)\, \alpha(0,t) = 0, \quad \text{(A5)}$$

which is exactly equivalent to (10) as needed. With the help of (A5), we obtain

$$(S \circ \alpha)''(0) \equiv \int_0^T \Big[\hat{J}(t) \frac{\partial \alpha(0,t)}{\partial u} \Big] \frac{\partial \alpha(0,t)}{\partial u}\, dt, \quad \text{(A6)}$$

where the Jacobi operator

$$\hat{J}(t) = -m \left(\frac{d}{dt}\right)^2 - m\omega_0 - 4 \int_0^T G(t,s)\, ds + 4 \int_0^T G(t,s)(\cdot)\, ds \quad \text{(A7)}$$

is also an integrodifferential operator.

[1]R. P. Feynman, Phys. Rev. **97**, 660 (1955).
[2]R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw–Hill, New York, 1965).
[3]D. C. Khandekar, S. V. Lawande, and K. V. Bhagwat, J. Phys. A **16**, 4209 (1983).
[4]B. K. Cheng, Phys. Lett. A **103**, 357 (1984).
[5]B. K. Cheng, Rev. Bras. Fis. **14**, 221 (1984).
[6]C. DeWitt-Morette, Commun. Math. Phys. **28**, 47 (1972).
[7]C. DeWitt-Morette, A. Maheshwari, and B. Nelson, Phys. Rep. **50**, 257 (1979).
[8]A. Maheshwari, J. Phys. A **8**, 1019 (1975).
[9]E. Bosco, Phys. Lett. A **95**, 277 (1983).
[10]D. C. Khandekar, S. Datta, and S. V. Lawande, Phys. Rev. A **31**, 3574 (1985).
[11]Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds and Physics* (North-Holland, New York, 1982).

Time-ordering techniques and solution of differential difference equation appearing in quantum optics

G. Dattoli, J. Gallardo,[a] and A. Torre[b]
ENEA, Dip. TIB, Divisione Fisica Applicata, C.R.E. Frascati, C.P. 65, 00044 Frascati, Rome, Italy

(Received 5 August 1985; accepted for publication 3 October 1985)

Time-ordering techniques based on the Magnus expansion and the Wei–Norman algebraic procedure are discussed and their relevance and usefulness to quantum optics are stressed.

I. INTRODUCTION

This paper has a twofold motivation: (a) to discuss relatively unknown time-ordering techniques, and (b) to show that these techniques are a useful tool to solve a large class of differential finite difference equations, too.

The problem of time-ordering expansion is as old as quantum mechanics and the most common treatment of it is the Feynman–Dyson[1] diagrammatic technique. However, alternative rigorous procedures, apparently not widely known, have been developed through the years by Magnus[2] and Wei and Norman.[3] These techniques offer definite advantages with respect to the well-known Feynman–Dyson[1] expansion and are tailored to be suited for a class of Hamiltonian operators appearing in many problems of quantum optics.

The considerations we develop here are general enough to be applied to diverse physical problems such as two-level molecular dynamics, stimulated Compton scattering, and the acousto-optic effect.

Let us briefly review the problems underlying the operator time evolution and time-ordering expansion. From elementary quantum mechanics[4] the evolution of the wave function of a physical system driven by a time-dependent Hamiltonian operator can be found formally by writing the solution of the Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = \hat{H}(t)\psi \tag{1.1}$$

as

$$\psi(t) = \hat{U}(t)\psi(0), \tag{1.2}$$

where $\hat{U}(t)$ is the time-evolution operator obeying the equation

$$i\hbar\frac{\partial\hat{U}(t)}{\partial t} = \hat{H}(t)\hat{U}(t), \quad \hat{U}(0) = \hat{1}. \tag{1.3}$$

If the operator $\hat{H}$ is time independent or if it commutes with itself at different times ($[\hat{H}(t),\hat{H}(t')] = 0$), the solution of (1.3) is straightforward, i.e.,

$$\hat{U}(t) = \exp\left[-\frac{i}{\hbar}\int_0^t \hat{H}(t')dt'\right]\hat{U}(0). \tag{1.4}$$

If the operator $\hat{H}(t)$ does not commute at different times, time-ordering problems arise and the solution of (1.3) cannot be expressed in the simple form (1.4).

The technique most frequently adopted to deal with the evolution of $\hat{U}(t)$ is the use of the Feynman–Dyson[1] expansion

$$\left\{\exp\left[\frac{i}{\hbar}\int_0^t \hat{H}(t')dt'\right]\right\}_+ = 1 - \frac{i}{\hbar}\int_0^t dt_1\,\hat{H}(t_1) + \left(\frac{i}{\hbar}\right)^2\int_0^t dt_1\int_0^{t_1} dt_2\,\hat{H}(t_1)\hat{H}(t_2) + \cdots. \tag{1.5}$$

Where $\{\cdot\}_+$ denotes the time ordering and plays the role of the "chronological" operator. The expansion (1.5) is a perturbation series with all the practical disadvantages of the perturbative expansion. Indeed as noticed elsewhere,[5] the operator $\hat{U}^{(n)}(t)$, obtained by truncating the series, is no more a unitary operator, furthermore it is expected to be accurate for a small time interval or when $\hat{H}(t)$ can be treated as a perturbation. In many problems $H(t)$ cannot be considered as a perturbation or an accurate evaluation requires an excessively large number of infinitesimal orders. However, it must be stressed that in many cases (1.5) can be easily handled and each term can be usefully understood in terms of the symbolic Feynman diagrams.[1]

To go beyond the expansion (1.5) we require at least (a) a functional form of $\hat{U}(t)$ which preserves the unitary nature of the evolution operator; and (b) an exact form of the operator without any recourse to perturbation, or if perturbation is needed, a method that allows the expansion at any higher order.

Two methods, essentially complementary, have been proposed that satisfy the above requirements.

The first due to Magnus[2] consists in writing

$$\hat{U}(t) = \exp\{\hat{A}(t)\}, \quad \hat{A}(0) = 0, \tag{1.6}$$

where $\hat{A}(t)$ is a functional of $\hat{H}(t)$, more precisely an infinite series whose nth term is a sum of integrals of n-fold multiple commutators of $\hat{H}(t)$.

This method is now briefly reviewed, we follow a simpler but rigorous version due to Pechukas and Light.[5] The search for the time-displacement operator expressed in the form (1.6) meets both the requirements (a) and (b) and is an immediate generalization to a more complicated case of the corresponding expression (1.4).

According to Ref. 5, for the time derivative of $\hat{U}(t)$,

[a] Permanent address: University of California at Santa Barbara, Quantum Institute, Santa Barbara, California 93106.
[b] Also at Istituto Nazionale di Fisica Nucleare, Sezione Napoli, Napoli, Italy.

 0022-2488/86/030772-09$02.50

$$\frac{d}{dt}\hat{U}(t) = \left[\frac{e^{\mathrm{ad}\hat{A}}-1}{\mathrm{ad}\,A}\frac{d}{dt}\hat{A}\right]\hat{U}(t), \tag{1.7}$$

and the evolution equation (1.3), one immediately obtains

$$\frac{d\hat{A}(t)}{dt} = \left[\frac{\mathrm{ad}\,\hat{A}}{e^{\mathrm{ad}\hat{A}}-1}\right]\frac{\hat{H}}{i\hbar}. \tag{1.8}$$

(The operator $\mathrm{ad}\,\hat{A}$ is a linear operator defined as $(\mathrm{ad}\,\hat{A})\hat{X} = [\hat{A},\hat{X}]$.) Expanding in series of the operator on the right-hand side of (1.8) we get

$$\frac{d\hat{A}(t)}{dt} = \left(1 - \frac{1}{2}\mathrm{ad}\,\hat{A} + \sum_{n=1}^{\infty}\frac{(-)^{n+1}}{(2n)!}B_n(\mathrm{ad}\,\hat{A})^{2n}\right)\frac{\hat{H}}{i\hbar}, \tag{1.9}$$

where the B_n are the Bernoulli numbers $B_1 = \frac{1}{6}$, $B_2 = \frac{1}{30}, \ldots$. Solving the above equation by iteration one ends up with the expression

$$\hat{A} = \sum_{n=1}^{\infty}\hat{A}_n, \tag{1.10}$$

where the $(n+1)$th term reads

$$\hat{A}_{n+1} = \int_0^t dt'\left[-\frac{1}{2}\mathrm{ad}\,\hat{A}_n + \frac{1}{12}\sum_{m=1}^{n-1}\mathrm{ad}\,\hat{A}_m\,\mathrm{ad}\,\hat{A}_{n-m}\right]\frac{\hat{H}}{i\hbar}. \tag{1.11}$$

The first four terms are

$$\begin{aligned}
\hat{A}_1(t) &= -\frac{i}{\hbar}\int_0^t \hat{H}(t')dt', \\
\hat{A}_2(t) &= -\frac{1}{2}\left(\frac{i}{\hbar}\right)^2\int_0^t dt'\int_0^{t'} dt''[\hat{H}(t''),\hat{H}(t')], \\
\hat{A}_3(t) &= -\frac{1}{6}\left(\frac{i}{\hbar}\right)^3\int_0^t dt'\int_0^{t'} dt''\int_0^{t''} dt''' \\
&\quad\times\{[\hat{H}(t'''),[\hat{H}(t''),\hat{H}(t')]] \\
&\quad+[[\hat{H}(t'''),H(t'')],H(t')]\}, \\
\hat{A}_4(t) &= -\frac{1}{12}\left(\frac{i}{\hbar}\right)^4\int_0^t dt'\int_0^{t'} dt''\int_0^{t''} dt'''\int_0^{t'''} dt'''' \\
&\quad\times\{[\hat{H}(t''''),[[\hat{H}(t'''),\hat{H}(t'')],\hat{H}(t')]] \\
&\quad+[[\hat{H}(t''''),[\hat{H}(t'''),\hat{H}(t'')]],\hat{H}(t')] \\
&\quad+[[\hat{H}(t''''),\hat{H}(t''')],[\hat{H}(t''),\hat{H}(t')]]\}.
\end{aligned} \tag{1.12}$$

We note that the structure of the Magnus expansion corresponds to the continuous version of the Baker–Hausdorff disentangling theorem.[6]

The second method we describe is the Wei–Norman algebraic procedure.[3] This technique is complementary to the previous one, in the sense that it works when the Hamiltonian operator $\hat{H}(t)$ can be expressed in terms of the generators of an n-dimensional finite Lie algebra. The Magnus expansion, on the other side, applies when the multiple commutators in (1.11) converge to a c number.

According to Ref. 3 we write the Hamiltonian as

$$\hat{H}(t) = \sum_{j=1}^{m\leqslant n} a_j(t)\hat{L}_j, \tag{1.13}$$

where the $\hat{L}_j$ are the generators of the Lie algebra, the $a_j(t)$ are linearly independent functions of t, and the index j runs from 1 to $m \leqslant n$, where n is the dimensionality of the algebra.

The form of the solution (1.4), being valid, within this framework in the case of a one-dimensional algebra, suggests the following generalization to the case (1.13)

$$\hat{U}(t) = \prod_{j=1}^{n}\exp[g_j(t)\hat{L}_j], \quad g_j(0) = 0. \tag{1.14}$$

The functions $g_j(t)$, entering the above expression, can be obtained from a set of nonlinear differential equations whose specific form depends on the $a_j(t)$, and the algebraic structure constants involved in replacing (1.14) in (1.3) immediately yields

$$\sum_{i=1}^{n}\dot{g}_i(t)\left[\prod_{j=1}^{i-1}\exp(g_j(t)\hat{L}_j)\right]\hat{L}_i\cdot\left[\prod_{j=i}^{n}\exp(g_j(t)\hat{L}_j)\right] = \sum_{i=1}^{n}a_i(t)\hat{L}_i\hat{U}(t). \tag{1.15}$$

After a postmultiplication by the inverse operator $\hat{U}^{-1}$ and the direct computation of the expression

$$\left[\prod_{j=1}^{i-1}\exp(g_j(t)\hat{L}_j)\right]\hat{L}_i\left[\prod_{j=i-1}^{1}\exp(-g_j(t)\hat{L}_j)\right] = \sum_{k=1}^{n}\xi_{ki}\hat{L}_k, \tag{1.16}$$

we find

$$\sum_{i=1}^{n}a_i(t)\hat{L}_i = \sum_{j=1}^{n}\sum_{i=j}^{n}\dot{g}_j(t)\xi_{ij}\hat{L}_i, \tag{1.17}$$

where the matrix elements $\xi_{i,j}$ depend on the algebraic structure constants and on the g functions.

The linear independence of the generators reduces (1.17) into the nth-order system of differential equations

$$\begin{pmatrix}a_1\\ a_n\end{pmatrix} = \begin{bmatrix}\xi_{1,1}\cdots\xi_{1,n}\\ \xi_{n,2}\cdots\xi_{n,n}\end{bmatrix}\begin{pmatrix}\dot{g}_1\\ \dot{g}_n\end{pmatrix}. \tag{1.18}$$

It is therefore clear that once the explicit form of $a_i(t)$ and $\xi_{i,k}$ are known one can determine the functions g_i solving a set of nonlinear differential equations. For the proof of invertibility of (1.18) see Ref. 3. Let us finally point out that it has been shown[3] that uncoupling theorem holds for all solvable Lie algebras and for the real "split three-dimensional" simple Lie algebra (see Sec. II).

The paper is organized as follows: In Sec. II we will discuss systems that allow exact solutions, in Sec. III we will discuss perturbation methods, and finally in Sec. IV we present some conclusive remarks.

II. EXACT SOLUTIONS

In this section we will apply the above-discussed techniques to specific cases of physical interest in quantum optics.

The first we consider is a Hamiltonian operator that is a generalization to the two-level case of the so-called Kano Hamiltonian,[7] namely ($\hbar = 1$)

$$\hat{H} = \omega(t)\hat{J}_3 + \Omega^*(t)\hat{J}_+ + \Omega(t)\hat{J}_- + \beta(t), \tag{2.1}$$

where $\omega(\tau)$, $\Omega^*(\tau)$, $\Omega(\tau)$, and $\beta(\tau)$ are time-dependent, complex nonsingular functions, furthermore the $\hat{J}$ operators obey the well-known angular momentum commutation relations

$$[\hat{J}_+,\hat{J}_-] = 2\hat{J}_3, \quad [\hat{J}_\pm,\hat{J}_3] = \mp\hat{J}_\pm \ . \tag{2.2}$$

Assuming, for simplicity, that the Hamiltonian (2.1) drives a system of coupled harmonic oscillators with n_+, n_- initial quanta in the upper and lower level, the more general time evolution of the state can be described by the wave function

$$|\psi(t)\rangle = \sum_{l=-n_+}^{n_-} C_l(t)|n_+ + l, n_- - l\rangle \ , \tag{2.3}$$

where l is an integer accounting for the number of exchanged photons and the $C_l(t)$ are time-dependent coefficients denoting the amplitude probabilities of l emissions at time t. The Schrödinger equation gives for the coefficients $C_l(t)$ the following motion equation:

$$i\frac{dC_l}{dt} = \omega(t)\left[\frac{n_+ - n_-}{2} + l\right]C_l(t) + \beta(t)C_l(t)$$
$$+ \Omega(t)\sqrt{(n_- - l)(n_+ + l + 1)}C_{l+1}(t) \tag{2.4}$$
$$+ \Omega^*(t)\sqrt{(n_+ + l)(n_- - l + 1)}C_{l-1}(t) \ ,$$
$$C_l(0) = c_l \ .$$

(To deduce (2.4) we have used the Schwinger realization of the angular momentum algebra, namely (see Ref. 4) $\hat{J}_+ = \hat{a}_+^\dagger\hat{a}_-$, $\hat{J}_- = \hat{a}_-^\dagger\hat{a}_+$, $\hat{J}_3 = \frac{1}{2}(\hat{a}_+^\dagger\hat{a}_+ - \hat{a}_-^\dagger\hat{a}_-)$, where $\hat{a}_\pm$, $\hat{a}_\pm^\dagger$ are creation annhilation operators $([\hat{a}_\pm^\dagger,\hat{a}_\pm] = -1[\hat{a}_+^\dagger,\hat{a}_-] = 0)$. The initial condition should be $C_l(0) = \delta_{l,0}$, we have assumed a generic discrete function to discuss slightly a more general problem.)

This differential difference equation already has been studied in Ref. 8 where it was pointed out that it belongs to the family of Raman–Nath (RN) equations[9] (i.e., spherical or SU_2 RN). We must stress that, according to the discussion of the previous section, the introduction of Eq. (2.4) is not a necessary step. The analytical expression of the evolution operator indeed can be found by means of the Hamiltonian operator (2.1). We have introduced this rather artificial step to remark that the technique we discuss here is also a powerful tool to solve equations of the RN type.

Adopting the same procedure of Ref. 8, we use the transformation

$$C_l(t) = (-i)^l \exp\left\{-i\int_0^t \omega(t')dt'\left[\frac{n_+ - n_-}{2} + l\right]\right\}$$
$$\times\exp\left[-i\int_0^t \beta(t')dt'\right]\cdot M_l \ , \tag{2.5}$$

which, once inserted in (2.4), yields

$$\frac{d}{dt}M_l = -\Omega(t)\exp\left\{-i\int_0^t dt'\,\omega(t')\right\}$$
$$\cdot\sqrt{(n_- - l)(n_+ + l + 1)}M_{l+1}$$
$$+ \Omega^*(t)\exp\left\{+i\int_0^t dt'\,\omega(t')\right\}$$
$$\cdot\sqrt{(n_+ + l)(n_- - l + 1)}M_{l-1} \ , \tag{2.6}$$

$$M_l(0) = \sum_k i^k c_k \delta_{l,k} \ .$$

We now can solve the problem of finding the explicit solution of $M_l(t)$ exploiting the Wei–Norman technique discussed in the previous section.

The structure of (2.5) suggests the following equation for the evolution operator:

$$\frac{d\hat{U}(t)}{dt} = \hat{T}(t)\hat{U}(t), \quad \hat{U}(0) = \hat{1} \ ,$$
$$\hat{T}(t) = -\Omega(t)\exp\left\{-i\int_0^t dt'\,\omega(t')\right\}\hat{J}_- \tag{2.7}$$
$$+ \Omega^*(t)\exp\left\{+i\int_0^t dt'\,\omega(t')\right\}\hat{J}_+ \ .$$

According to (14) the explicit solution of (2.6) can be written as

$$\hat{U}(t) = \exp\{2h(t)\hat{J}_3\}\exp\{g(t)\hat{J}_+\}\exp\{-f(t)\hat{J}_-\}\hat{1} \ . \tag{2.8}$$

Before giving the differential equations from which one can derive the functions f, g, and h, we notice that, to calculate the functional form of (2.6), it will be sufficient to evaluate the following matrix element:

$$\langle l|\hat{U}(t)|k\rangle = \sum_{m=0}^{\infty}\sum_{r=0}^{\infty}\frac{(-1)^r[g(t)]^m\cdot[f(t)]^r}{m!r!}$$
$$\times\langle l|\exp[2h(t)\hat{J}_3]\hat{J}_+^m\hat{J}_-^r|k\rangle \ . \tag{2.9}$$

After some algebra and exploiting the properties of the $\hat{J}_-$ operators we find

$$\langle l|\hat{U}(t)|k\rangle$$
$$= \exp\{2h(t)[\tfrac{1}{2}(n_+ - n_-) + l]\}[g(t)]^{l-k}$$
$$\times\left[\binom{n_- - k}{l-k}\binom{n_+ + l}{l-k}\right]^{1/2}$$
$$\times {}_2F_1(-n_+ + k, n_- - k + 1; l - k + 1; f(t)\cdot g(t)) \ . \tag{2.10}$$

Combining (2.5), (2.6), and (2.10) we finally easily find

$$C_l(t)$$
$$= \exp\left\{-i\int_0^t \beta(t')dt'\right\}\exp\left\{\left[\frac{1}{2}(n_+ - n_-) + l\right]\mathscr{H}(t)\right\}$$
$$\times\sum_k(-i)^{l-k}c_k[g(t)]^{l-k}$$
$$\times\left[\binom{n_- - k}{l-k}\binom{n_+ + l}{l-k}\right]^{1/2}$$
$$\times {}_2F_1(-n_+ + k, n_- - k + 1; l - k + 1; g(t)\cdot f(t)) \ , \tag{2.11}$$

where

$$\mathscr{H}(t) = 2h(t) - i\int_0^t \omega(t')dt' \tag{2.12}$$

and ${}_2F_1(\cdots)$ is the hypergeometric function.[10] The result (2.12) is a more general expression of that obtained in Ref. 8(b).

Let us now consider the problem of writing the differential equations satisfied by (f,g,h).

It is easy to derive from (1.18) and from the algebraic structure of the angular momentum operators the following equations:

$$\dot{h}(t) = \Omega(t)\exp\{\mathscr{H}(t)\}g(t),$$
$$\dot{g}(t) = \Omega^*(t)\exp\{-\mathscr{H}(t)\} - g(t)\cdot\dot{h}(t), \quad (2.13)$$
$$\dot{f}(t) = \Omega(t)\exp\{\mathscr{H}(t)\}, \quad h(0) = f(0) = g(0) = 0.$$

It can be shown that the solution of (2.13) depends on the single Riccati equation

$$\dot{u} - u^2 + r(t)u + q(t) = 0, \quad \dot{h}(t) = u(t), \quad u(0) = 0, \quad (2.14)$$

$$r(t) = -\frac{d}{dt}\ln\Omega(t) + i\omega(t), \quad q(t) = -|\Omega(t)|^2.$$

The solution of (2.14) can be explicitly written in a restricted number of cases. In the less general case $\omega(t) = \omega_0 = \text{const}$; $\Omega(t) = \Omega = \text{const}$ and real $\beta(t) = 0$, we easily get

$$h(t) = (i/2\omega_0 t - \ln(1 - p(t))^{1/2} - i\arctan\left(\left(\frac{\omega_0}{\delta}\right)\tan\left(\frac{\delta t}{2}\right)\right),$$
$$g(t) = [p(t)]^{1/2}[1 - p(t)]^{1/2} \times \exp\left\{i\arctan\left(\left(\frac{\omega_0}{\delta}\right)\tan\left(\frac{\delta t}{2}\right)\right)\right\}, \quad (2.15)$$
$$f(t) = \left[\frac{p(t)}{1 - p(t)}\right]^{1/2}\exp\left\{-i\arctan\left(\left(\frac{\omega_0}{\delta}\right)\tan\left(\frac{\delta t}{2}\right)\right)\right\},$$
$$p(t) = \left(\Omega\frac{\sin\delta t/2}{\delta/2}\right)^2, \quad \delta = \sqrt{\omega_0^2 + 4\Omega^2}.$$

In the case of $n_+ = 0$ we get [see Ref. 8(a)]

$$C_l(t) = \binom{n_-}{l}^{1/2}\exp\left\{in_-\arctan\left(\frac{\omega_0}{\delta}\tan\frac{\delta t}{2}\right)\right\}(\alpha(t))^l \times (1 - |\alpha(t)|^2)^{(n_- - l)/2}, \quad (2.16)$$
$$\alpha(t) = (-i)\left(\Omega\frac{\sin\delta t/2}{\delta/2}\right)\exp\left[-i\arctan\left(\frac{\omega_0}{\delta}\tan\left(\frac{\delta t}{2}\right)\right)\right].$$

The results obtained so far are very general. We can now discuss some interesting limiting cases, when the number of "excitation quanta" $n_\pm$ are very large.

A somewhat crude approach to the problem could be that of taking the asymptotic limits $n_\pm \to \infty$ in (2.11). This procedure gives the right functional form of the coefficient C_l but raises doubts on the correct expression of the functions f, g, and h. The appropriate and rigorous procedure requires the so-called group contraction method,[11] which will allow us to understand the intimate connection between the SU_2 algebra and its contraction to the "harmonic oscillator" and "shift" algebras.

We introduce a three-dimensional Lie algebra with generators $\hat{H}_1$, $\hat{H}_2$, and $\hat{H}_3$ with commutation relations[3]

$$[\hat{H}_1, \hat{H}_2] = 2\lambda\hat{H}_2,$$
$$[\hat{H}_1, \hat{H}_3] = -2\lambda\hat{H}_3, \quad (2.17)$$
$$[\hat{H}_2, \hat{H}_3] = -\delta\hat{H}_1,$$

where λ and δ are numbers which define the explicit form of the $\hat{H}$ operators. We leave, for the moment, the operators in (2.17) undefined and notice that an evolution operator driven by the $\hat{H}_s$ (we mean by this a Hamiltonian of the type $\hat{H} = [\omega(t)/2]\hat{H}_1 + \Omega^*(t)\hat{H}_2 - \Omega(t)\hat{H}_3)$ [see (2.8)] exhibits f, g, and h functions defined by the differential equations

$$\dot{h}(t) = \delta g(t)\dot{f}(t),$$
$$\dot{g}(t) = \Omega^*(t)\exp\{-\mathscr{H}_\lambda^{(t)}\} - \lambda g(t)\dot{h}(t), \quad (2.18)$$
$$\dot{f}(t) = \Omega(t)\exp\{\mathscr{H}_\lambda^{(t)}\},$$
$$\mathscr{H}_\lambda^{(t)} = 2\lambda h(t) - i\int_0^t \omega(t')dt'.$$

It is easy to understand that when $\lambda = \delta = 1$ the $\hat{H}$ operators can be identified with

$$\hat{H}_1 = 2\hat{J}_3, \quad \hat{H}_2 = \hat{J}_+, \quad \hat{H}_3 = -\hat{J}_-. \quad (2.19)$$

Furthermore if $\lambda = 0$ and $\delta = 1$ the SU_2 algebra contracts to the three-dimensional non-Abelian algebra with generators $\{\hat{a}^+, \hat{a}, \hat{I}\}$, where $\hat{a}^+$, $\hat{a}$ are creation annihilation operators. Finally when $\lambda = \delta = 0$ all our algebra collapses in a "shift algebra" with generators $\{\hat{E}^+, \hat{E}^-, \hat{I}\}$, where the $\hat{E}^\pm$ are shift operators.

Let us now discuss the cases $n_- \to \infty$ and $n_\pm \to \infty$.

A. Large number of lower level excitation quanta ($n_- \to \infty$)

In this case the SU_2 RN equation (2.4) reduces to the so-called harmonic RN equation[12]

$$i\frac{dC_l}{dt} = \omega(t)(n_+ + l)C_l + \bar{\Omega}^*(t)\sqrt{n_+ + l}\,C_{l-1} + \bar{\Omega}(t)\sqrt{n_+ + l + 1}\,C_{l+1} + \bar{\beta}(t)C_l,$$
$$\bar{\Omega}(t) = \Omega(t)\sqrt{n_-}, \quad \bar{\beta}(t) = \beta(t) - \frac{n_- + n_+}{2}\omega(t). \quad (2.20)$$

This expression suggests the following identification of the $\hat{H}$ operators:

$$\hat{H}_1 = -n_-\hat{1}, \quad \hat{H}_2 = \sqrt{n_-}\,\hat{a}^+, \quad \hat{H}_3 = -\sqrt{n_-}\,\hat{a}. \quad (2.21)$$

Therefore, setting $\lambda = 0$, $\delta = 1$ in (2.18) we immediately find the solutions

$$g(t) = \int_0^t \Omega^*(t')\exp\left[i\int_0^{t'}\omega(t'')dt''\right]dt',$$
$$f(t) = \int_0^t \Omega(t')\exp\left[-i\int_0^{t'}\omega(t'')dt''\right]dt', \quad (2.22)$$
$$h(t) = \frac{1}{2}\int_0^t g(t')\dot{f}(t')dt' - \frac{1}{2}\int_0^t [\dot{g}(t')f(t') - g(t')\dot{f}(t')]dt'.$$

Furthermore, using the well-known asymptotic properties of the hypergeometric function,[12] we find for C_l the expression

$$C_l(t)=\exp\left\{-i\int_0^t \bar{\beta}(t')dt'\right\}$$
$$\times\exp\{-i(n_+ + l)\int_0^t \omega(t')dt'\}\exp\left\{-\frac{1}{2}\alpha(t)\cdot\gamma(t)\right\}$$
$$\times\exp\left\{-\frac{1}{2i}\int_0^t [\bar{\Omega}^*(t')\gamma(t')+\bar{\Omega}(t')\alpha(t')]dt'\right\}$$

$$\times\sum_k c_k(-i)^{l-k}\sqrt{(n_+ + k)!(n_+ + l)!}$$
$$\times\mathscr{F}^{l-k}(t)L^{l-k}_{n_+ + k}(\alpha(t)\cdot\gamma(t)), \quad (2.23)$$

where the L^l_n are the generalized Laguerre polynomials and

$$\mathscr{F}(t)=\int_0^t \bar{\Omega}^*(t')\exp\left\{+i\int_0^{t'}\omega(t'')dt''\right\}dt',$$
$$\alpha(t)=-i\mathscr{F}(t)\exp\left\{-i\int_0^{t'}\omega(t'')dt''\right\},$$
$$\gamma(t)=i\left[\int_0^t \bar{\Omega}(t')\exp\left\{-i\int_0^{t'}\omega(t'')dt''\right\}dt'\right]$$
$$\times\exp\left\{i\int_0^t \omega(t')dt'\right\}. \quad (2.24)$$

It is easy to derive from (2.23) the solution already found in Ref. 12 when $C_l(0)=\delta_{l,0}$, $\beta(t)=0$, $\omega_0=\text{const}$, and $\bar{\Omega}=\bar{\Omega}^*$ const, namely

$$C_l(t)=\sqrt{\frac{n_+!}{(n_+ + l)!}}\alpha(t)^l\cdot e^{-in_+\omega_0 t}$$
$$\times\exp\left\{i\frac{\omega_0}{2}\int_0^t |\alpha(t')|^2dt'\right\}$$
$$\times\exp(-|\alpha(t)|^2/2)L^l_n[|\alpha(t)|^2], \quad (2.25)$$
$$\alpha(t)=-i\left(\bar{\Omega}\frac{\sin\omega_0 t/2}{\omega_0/2}\right)e^{-i\omega_0 t/2}.$$

B. Large number of upper and lower number of excitation quanta ($n_\pm\to\infty$)

Equation (2.4) reduces, in this hypothesis, to the so-called shift RN equation, namely

$$i\frac{dC_l}{dt}=\omega(t)lC_l+\bar{\Omega}(t)C_{l+1}+\bar{\Omega}^*(t)C_{l-1}+\bar{\beta}(t)C_l,$$
$$\bar{\Omega}=\sqrt{n_+n_-}\,\Omega(t),\quad \bar{\beta}=\beta(t)+(n_+ - n_-)/2\omega(t). \quad (2.26)$$

The identification of the $\hat{H}$ operators is straightforward:

$$\hat{H}_1=0,\quad \hat{H}_2=\sqrt{n_+n_-}\hat{E}^+,\quad \hat{H}_3=-\sqrt{n_+n_-}\hat{E}^-. \quad (2.27)$$

Therefore setting $\lambda=\delta=0$ in (36) and exploiting again the asymptotic properties of the hypergeometric function for large $n_\pm$ (see Ref. 10), we finally find

$$C_l=\exp\left(-i\int_0^t\bar{\beta}(t')dt'\right)$$
$$\times\sum_k c_k\exp\left(-ik\int_0^t\omega(t')dt'\right)\left[\frac{\alpha(t)}{\gamma(t)}\right]^{(l-k)/2}$$
$$\times J_{l-k}[2\sqrt{\alpha(t)\cdot\gamma(t)}], \quad (2.28)$$

where $J_l(\cdot)$ is the cylindrical Bessel function of the first kind and integer order, furthermore

$$\alpha(t)=-i\left[\int_0^t\bar{\Omega}^*(t')\cdot\exp\left(-i\int_0^{t'}\omega(t'')dt''\right)\right]$$
$$\times\exp\left(-i\int_0^t\omega(t')dt'\right),$$
$$\gamma(t)=i\left[\int_0^t\bar{\Omega}(t')\exp\left(+i\int_0^{t'}\omega(t'')dt''\right)\right]$$
$$\times\exp\left(+i\int_0^t\omega(t')dt'\right). \quad (2.29)$$

It is easy to see that when $\beta(t)=0$, $\Omega(t)=\Omega^*(t)=\text{const}$ $C_l(0)=\delta_{l,0}$ (2.29) reduces to the well-known solution[13]

$$C_l=(-i)^l\exp\left(-il\frac{\omega_0 t}{2}\right)J_l\left[2\bar{\Omega}\frac{\sin\omega_0 t/2}{\omega_0/2}\right]. \quad (2.30)$$

Before concluding this section we stress that the analysis we have presented is very general and based on the Wei–Norman technique. However, while this procedure is strictly necessary for the SU_2 algebra in the case of the "harmonic oscillator" algebra, the Magnus expansion is equally useful (see the Appendix).

III. PERTURBED SOLUTIONS

In the previous section we considered particularly significant cases that admit exact solutions. In this section we will discuss different situations where exact solutions are not available but nontrivial perturbed solutions may be obtained.

The analysis we develop in this section is relevant, e.g., to the evolution of quantum systems driven by Hamiltonians of the type

$$\hat{H}=\omega(t)\hat{J}_3+\epsilon(t)\hat{J}_3^2+[\Omega^*(t)\hat{J}_+ +\Omega(t)\hat{J}_-]+\beta(t), \quad (3.1)$$

where $\epsilon(t)$ is a nonsingular time-dependent function that can be treated as a perturbation.

To illustrate the method we shall restrict ourselves to the algebraically simpler case of ϵ, ω, and $\Omega=\Omega^*$ time-independent constants. However, we stress that the identical procedure applies to the Hamiltonian (3.2).

We now will consider the specific problem of the stimulated Thomson scattering of two counterpropagating electromagnetic waves.[14]

According to Ref. 14 this process can be described by a spherical RN equation of the type

$$i\frac{dC_l}{dt}=\omega_0 lC_l+\epsilon l^2C_l+\Omega[\sqrt{(n_- - l)(n_+ l+1)}C_{l+1}(t)$$
$$+\sqrt{(n_- - l+1)(n_+ + l)}C_{l-1}(t)],\ C_l(0)=\delta_{l,0}. \quad (3.2)$$

The main difference between the above equation and the one considered in the previous section is the presence of the quadratic term in l. However, if this term can be treated as a perturbation (as happens in many cases of physical interest), one can find a perturbed solution to first order in ϵ [(3.2)]. We proceed as in the previous section. Namely we introduce the function

$$C_l(t) = (-i)^l \exp\{-i(\omega_0 + \epsilon l)lt\} M_l(t), \tag{3.3}$$

which, once inserted in (3.2), gives the new expression

$$\begin{aligned}\frac{dM_l(t)}{dt} &= \Omega\{\sqrt{(n_+ + l)(n_- - l + 1)} \\ &\quad \times \exp[i(\omega_0 + \epsilon(2l-1))(t)] M_{l-1}(t) \\ &\quad - \sqrt{(n_- - l)(n_+ l + 1)} \\ &\quad \times \exp[-i(\omega_0 + \epsilon(2l+1))(t)] M_{l+1}(t)\}, \\ M_l(0) &= i^l \delta_{l,0}.\end{aligned} \tag{3.4}$$

The above expression, even if more complicated than those discussed in Sec. II, suggests the following structure for the motion equation of the evolution operator:

$$\begin{aligned}\frac{d\hat{U}}{dt} &= \hat{T}(t)\hat{U}, \\ \hat{T}(t) &= \Omega e^{i\omega_0 t} \exp\{i\epsilon[2\hat{J}_3 - (n_+ - n_-) - 1]t\}\hat{J}_+ \\ &\quad - \Omega e^{-i\omega_0 t}\exp\{-i\epsilon[2\hat{J}_3 - (n_+ - n_-) + 1]t\}\hat{J}_-.\end{aligned} \tag{3.5}$$

The presence of the $\hat{J}_3$ operator at the exponent, does not allow any exact solution of (3.5) in closed form.

However, expanding the exponents up to the first order in ϵ the $\hat{T}$ operator can be written as

$$\begin{aligned}\hat{T}(t) &\simeq b^0(t)\hat{H}_2 + C^0(t)\hat{H}_3 + \epsilon[b^1(t)\hat{H}_2 \\ &\quad + C^1(t)\hat{H}_3 + p(t)\hat{H}_1\hat{H}_3 + q(t)\hat{H}_1\hat{H}_2],\end{aligned} \tag{3.6}$$

where the $\hat{H}$ are the operators introduced in (2.19):

$$\begin{aligned}&b^0(t) = \Omega e^{i\omega_0 t}, \quad b^1(t) = -t\Omega i(\Delta n + 1)e^{i\omega_0 t}, \\ &C^0(t) = \Omega e^{-i\omega_0 t}, \quad C^1(t) = -ti\Omega(\Delta n - 1)e^{-i\omega_0 t}, \\ &p(t) = \frac{\partial}{\partial\omega_0}C^0(t), \quad q(t) = \frac{\partial}{\partial\omega_0}b^0(t), \\ &\Delta n \equiv n_+ - n_-.\end{aligned} \tag{3.7}$$

A convenient form to find the time evloution of $\hat{U}(t)$ is

$$\begin{aligned}\hat{U} &= \exp\{\epsilon\zeta(t)\hat{H}_1^2\}\exp\{\epsilon\eta(t)\hat{H}_2^2\}\exp\{\epsilon\gamma(t)\hat{H}_3^2\} \\ &\quad \times\exp\{\epsilon\delta(t)\hat{H}_1\cdot\hat{H}_2\} \\ &\quad \times\exp\{\epsilon\theta(t)\hat{H}_1\hat{H}_3\}\exp\{\epsilon\lambda(t)\hat{H}_2\hat{H}_3\} \\ &\quad \times\exp\{g_1(t)\hat{H}_1\}\exp\{g_2(t)\hat{H}_2\}\exp\{g_3(t)\hat{H}_3\}\hat{I},\end{aligned} \tag{3.8}$$

where the functions in the exponents are specified by the following system of differential equations:

$$\begin{aligned}g_1(t) &= h(t) + \epsilon h^1(t), \\ g_2(t) &= g(t) + \epsilon g^1(t), \\ g_3(t) &= f(t) + \epsilon f^1(t).\end{aligned} \tag{3.9}$$

Here, $f(t)$, $g(t)$, and $h(t)$ are the functions defined in the previous section and furthermore

$$\begin{aligned}b^1(t) &= e^{2h(t)}[g^1(t)^2 + g(t)^2\dot{f}^1(t)] + 2b^0(t)h^1(t) \\ &\quad + 2C^0(t)g^1(t) - 2b^0(t)(2\zeta(t) \\ &\quad + \lambda(t)) + 2\eta(t)C^0(t), \\ C^1(t) &= 2b^0(t)\gamma(t) + e^{-2h(t)}[\dot{f}^1(t) - 2\dot{f}(t)h^1(t)] \\ &\quad - 4\zeta(t)C^0(t), \\ 0 &= \dot{h}^1(t) - \dot{f}^1(t)g(t) - \dot{f}(t)g^1(t) - 2\delta(t)C^0(t), \\ q(t) &= \dot{\delta}(t) + (4\zeta(t) + \lambda(t))b^0(t) - 2\eta(t)C^0(t), \\ p(t) &= \dot{\theta}(t) + 2\gamma(t)b^0(t) - C^0(t)(4\zeta(t) + \lambda(t)), \\ 0 &= \dot{\lambda}(t) + 2\theta(t)b^0(t) - 2\delta(t)C^0(t), \\ 0 &= \dot{\zeta}(t) + \theta(t)b^0(t) - \delta(t)C^0(t), \\ 0 &= \dot{\eta}(t) + 2\delta(t)b^0(t), \\ 0 &= \dot{\gamma}(t) - 2\theta(t)C^0(t).\end{aligned} \tag{3.10}$$

Finally, using the properties of the $\hat{H}$ operators we find the following expression for the $C_l(t)$ coefficients:

$$\begin{aligned}C_l(t) &= \left[\binom{n_-}{l}\binom{n_+ + l}{l}\right]^{1/2} \exp\left\{i(n_- - n_+)\left[\arctan\left(\frac{\omega_0}{\delta}\tan\frac{\delta t}{2}\right) - \frac{\omega_0 t}{2}\right]\right\}\alpha(t)^l(1 - |\alpha(t)|^2)^{(n_- - n_+ - l)/2} \\ &\quad \times\Big\{[1 - i\epsilon l^2 t + \epsilon(n_+ - n_- + 2l)(h^1(t) - i\alpha(t)(1 - |\alpha(t)|^2)^{1/2}\exp[2i\arctan(\omega_0/\delta\tan\delta t/2)]f^1(t)) \\ &\quad + \epsilon\lambda(t)(\tfrac{1}{2}(n_+ - n_- + 2l)^2 - (n_+ + l)(n_- - l + 1)]\,{}_2F_1(-n_+, n_- + 1; l + 1; |\alpha(t)|^2) \\ &\quad - i\epsilon\frac{\exp[-2i\arctan(\omega_0/\delta\tan\delta t/2)]}{\alpha(t)(1 - |\alpha(t)|^2)^{1/2}}\Big[(g^1(t) + \alpha(t)^2(1 - |\alpha(t)|^2) \\ &\quad \times\exp[4i\arctan(\omega_0/\delta\tan\delta t/2)]f^1(t))\cdot l\,{}_2F_1(-n_+, n_- + 1; l; |\alpha(t)|^2) \\ &\quad + \frac{(n_- - l)(n_+ + l + 1)}{(l+1)}\alpha(t)^2(1 - |\alpha(t)|^2)\exp[4i\arctan(\omega_0/\delta\tan\delta t/2)] \\ &\quad \times f^1(t)\,{}_2F_1(-n_+, n_- + 1; l + 2; |\alpha(t)|^2)\Big] - i\epsilon\frac{(n_+ - n_- + 2l)}{\alpha(t)(1 - |\alpha(t)|^2)^{1/2}} \\ &\quad \times\Big[l\delta(t)e^{-i\omega_0 t}(1 - |\alpha(t)|^2)\,{}_2F_1(-n_+, n_- + 1; l; |\alpha(t)|^2) \\ &\quad + \theta(t)e^{i\omega_0 t}\frac{n_- - l}{l+1}(n_+ + l + 1)\alpha(t)^2\,{}_2F_1(-n_+, n_- + 1; l + 2; |\alpha(t)|^2)\Big]\end{aligned}$$

$$-\epsilon\Big[\eta(t)e^{-2i\omega_0 t}\frac{(1-|\alpha(t)|^2)}{\alpha(t)^2}l(l+1)\,_2F_1(-n_+,n_-+1;l-1;|\alpha(t)|^2)\Big]$$
$$-\epsilon\gamma(t)e^{2i\omega_0 t}\frac{\alpha(t)^2}{(1-|\alpha(t)|^2)}\frac{(n_--l)(n_--l-1)(n_++l+1)(n_++l+2)}{(l+1)(l+2)}$$
$$\times\,_2F_1(-n_+,n_-+1;l+3;|\alpha(t)|^2)\Big\}. \quad (3.11)$$

The results we have derived show that a perturbed analysis of quantum systems driven by Hamiltonians like (3.1) in principle can be carried out analytically.

However, we must point out that (a) the system of differential equations (3.10) cannot be solved straight forwardly; and (b) the expression (3.11) is rather complicated and in the present framework further simplifications cannot be made.

In any case these results can be usefully exploited when physical assumptions allow some simplifications or as a test of numerical analysis.

Let us now discuss the "contraction" of Eq. (50) to (namely for large n_-)

$$i\frac{dC_l}{dt}=\omega_0 lC_l+\epsilon l^2C_l+\bar\Omega\big[\sqrt{n+l+1}\,C_{l+1}+\sqrt{n+l}\,C_{l-1}\big],\quad C_l(0)=c_l. \quad (3.12)$$

(This case is relevant to Hamiltonians of the type $\hat H=\omega_0\hat a^+\hat a+\epsilon(\hat a^+\hat a)^2+\Omega[\hat a^++\hat a]$.) The perturbed solution of (3.12) can be found using the same technique of group contraction discussed in the previous section. We omit the details of the calculations and write directly the solution

$$C_l(t)=\exp\Big(\frac{i\omega_0}{2}\int_0^t|\alpha(\tau')|^2\,d\tau'\Big)\exp\Big(\frac{-1}{2}|\alpha(t)|^2\Big)\sum_k\sqrt{\frac{(n+k)!}{(n+l)!}}\exp(-ik\omega_0 t)\alpha(t)^{l-k}c_k\{A_{l-k}(t)+iD_{l-k}(t)\}, \quad (3.13)$$

where

$$A_{l-k}(t)=L^{l-k}_{n+k}(\cdot)[1-\epsilon kC(t)]+\frac{\epsilon}{|\alpha(t)|}\frac{\partial}{\partial\omega_0}|\alpha(t)|\big[|\alpha(t)|^2L^{l-k+1}_{n+k-1}(\cdot)+(n+k+1)L^{l-k-1}_{n+k+1}(\cdot)\big]$$
$$+\frac{\epsilon}{|\alpha(t)|}\Big[R(t)-\frac{2}{9}\frac{\partial}{\partial\omega_0}|\alpha(t)|^3\Big]\big[|\alpha(t)|^2L^{l-k+1}_{n+k-1}(\cdot)-(n+k+1)L^{l-k-1}_{n+k+1}(\cdot)\big]$$
$$+\frac{2\epsilon}{|\alpha(t)|}k\frac{\partial}{\partial\omega_0}|\alpha(t)|\big[n+k+1)L^{l-k-1}_{n+k+1}(\cdot)$$
$$-|\alpha(t)|^2L^{l-k-1}_{n+k-1}(\cdot)\big]-\frac{\epsilon}{|\alpha(t)|}\frac{\partial}{\partial\omega_0}|\alpha(t)|\big[|\alpha(t)|^4L^{l-k+2}_{n+k-2}(\cdot)-(n+k+1)(n+k+2)L^{l-k-2}_{n+k+2}(\cdot)\big],$$
$$D_{l-k}(t)=-\epsilon l^2t+\epsilon[2C(t)|\alpha(t)|^2+G(t)-7/6t\,|\alpha(t)|^4+(2(n+k)+1)(2C(t)-t\,|\alpha(t)|^2)]L^{l-k}_{n+k}(\cdot) \quad (3.14)$$
$$-\epsilon(t/2)\big[|\alpha(t)|^2L^{l-k+1}_{n+k-1}(\cdot)-(n+k+1)L^{l-k-1}_{n+k+1}(\cdot)\big]+\epsilon[2t\,|\alpha(t)|^2-3C(t)]$$
$$\times\big[|\alpha(t)|^2L^{l-k+1}_{n+k-1}(\cdot)+(n+k+1)L^{l-k-1}_{n+k+1}(\cdot)\big]$$
$$+\epsilon tk\big[(n+k+1)L^{l-k-1}_{n+k+1}(\cdot)-L^{l-k+1}_{n+k-1}(\cdot)|\alpha(t)|^2\big]$$
$$-\frac{\epsilon}{|\alpha(t)|^2}\Big[C(t)+\frac{t}{2}|\alpha(t)|^2\Big]\big[|\alpha(t)|^4L^{l-k+2}_{n+k-2}(\cdot)$$
$$+(n+k+1)(n+k+2)L^{l-k-2}_{n+k+2}(\cdot)\big]$$

and

$$C(t)=\Big(\frac{\bar\Omega}{\omega_0}\Big)^2\Big[t-\frac{\sin\omega_0 t}{\omega_0}\Big],$$
$$G(t)=-\frac{2}{3}\Big(\frac{\bar\Omega}{\omega_0}\Big)^4\Big[-\frac{t}{4}\cos 2\omega_0 t-\frac{13}{4}t-2t\cos\omega_0 t+\frac{1}{6}\omega_0^2t^3+\frac{3}{4\omega_0}\sin 2\omega_0 t+\frac{4}{\omega_0}\sin\omega_0 t\Big], \quad (3.15)$$
$$R(t)=-\frac{2}{3}\Big(\frac{\bar\Omega}{\omega_0}\Big)^3\Big[7t\cos\frac{\omega_0 t}{2}+2t\cos\omega_0 t\cos\frac{\omega_0 t}{2}-\frac{7}{2\omega_0}\sin\frac{3}{2}\omega_0 t-\frac{15}{2\omega_0}\sin\frac{\omega_0 t}{2}\Big].$$

[The $L^l_n(\cdot)$ are the generalized Laguerre polynomials of argument $|\alpha(\tau)|^2$.] As final example we consider the further "contraction" of (3.2) when both $n_\pm$ are large quantities. In this case the spherical RN equation reduces to

$$i\frac{dC_l}{dt}=\omega_0 lC_l+\epsilon l^2C_l+\bar\Omega[C_{l+1}+C_{l-1}],\quad C_l(0)=c_l. \quad (3.16)$$

Using the same procedure leading to (3.12) we find

$$C_l=\sum_k(-i)^{l-k}e^{-i[\omega_0(l-k)t/2]}c_k\{A_{l-k}+iD_{l-k}\}, \quad (3.17)$$

where

$$A_{l-k} = J_{l-k}(\cdot) + \epsilon\frac{\partial}{\partial\omega_0}\left(\bar{\Omega}\frac{\sin\omega_0 t/2}{\omega_0/2}\right)[(2l-1)J_{l-k-1}(\cdot) - (2l+1)J_{l-k+1}(\cdot)] + \frac{1}{2}\epsilon\bar{\Omega}^2\frac{\partial}{\partial\omega_0}\left(\bar{\Omega}\frac{\sin\omega_0 t/2}{\omega_0/2}\right)^2[J_{l-k+2}(\cdot) - J_{l-k-2}(\cdot)],$$

$$D_{l-k} = \left[-\epsilon l^2 t + 4\epsilon\frac{\bar{\Omega}^2}{\omega_0^2}\left(\frac{\sin\omega_0 t}{\omega_0} - t\right)\right]J_{l-k}(\cdot) + \frac{\epsilon t}{2}\left(\bar{\Omega}\frac{\sin\omega_0 t/2}{\omega_0/2}\right)[(2l-1)J_{l-k-1}(\cdot) + (2l+1)J_{l-k+1}(\cdot)] + \epsilon\frac{\bar{\Omega}^2}{\omega_0^2}\left(t\cos\omega_0 t - \frac{\sin\omega_0 t}{\omega_0}\right)[J_{l-k-2}(\cdot) + J_{l-k+2}(\cdot)]. \quad (3.18)$$

The argument of the Bessel functions left indicated by $(\cdot)$ is $2\bar{\Omega}(\sin\omega_0 t/2)/(\omega_0/2)$.

IV. CONCLUSIONS

In this work we have discussed a rather general technique that can be usefully applied to a number of physical problems. Furthermore the techniques we have developed may be of interest even for mathematicians, since they amount to a useful tool to construct solutions for a large class of differential finite difference equations.

Concerning this last point we wish to add the following comment.

In some cases when the structure of the RN equation is particularly simple, namely when it can be derived from Hamiltonians that do not involve noncommuting operators [as, e.g., the $(\hat{E}^+, \hat{E}^-, \hat{I})$ generators] more direct methods can be used.

To give an example we reconsider the shift RN equation with constant coefficients (see Sec. II):

$$i\frac{dC_l}{dt} = \omega_0 l C_l + \Omega[C_{l+1} + C_{l-1}], \quad C_l(0) = \delta_{l,0}. \quad (4.1)$$

Using the transformation (2.5) we get

$$\frac{dM_l}{dx} = -e^{-i\beta x}M_{l+1} + e^{i\beta x}M_{l-1},$$

$$M_l(0) = i^l\delta_{l,0} \quad (x = \Omega t, \quad \beta = \omega_0/\Omega). \quad (4.2)$$

Multiplying both sides of (4.2) by s^l and summing over l we find

$$\frac{d\Gamma(x,s)}{dx} = \left(se^{+i\beta x} - \frac{1}{s}e^{-i\beta x}\right)\Gamma(x,s),$$

$$\Gamma(0,s) = 1, \quad \Gamma(x,s) = \sum_{l=-\infty}^{\infty} s^l M_l(x), \quad (4.3)$$

$$0 < |s| < \infty.$$

Equation (4.3) can be solved straightforwardly, namely

$$\Gamma(x,s) = \exp\left\{\frac{\sin\beta x/2}{\beta/2}\left[se^{i\beta x/2} - \frac{1}{s}e^{-i\beta x/2}\right]\right\}. \quad (4.4)$$

Therefore, using the Bessel generating function $[\sum_{l=-\infty}^{+\infty} t_l J_l(x) = e^{x/2(t-1/t)}]$ we easily find [see also Ref. (10)]

$$C_l(t) = (-i)^l e^{-il\omega_0 t/2} J_l[2\Omega\sin\omega_0 t/2/\omega_0/2]. \quad (4.5)$$

This is only a particularly simple example that shows that in a few selected cases simpler and more direct techniques are available. In any case when noncommuting operators are involved with time-dependent coefficients, time-ordering techniques, of the type discussed here, are a necessary step.

A further point, relevant to the differential finite difference equation, that we want to touch upon is the fact that we have discussed only homogeneous equations. We have not mentioned inhomogeneous cases which may arise treating perturbed solutions of nonlinear differential finite difference systems [to give an example, $i\,dC_l/dt = \Omega(C_{l+1} + C_{l-1}) + \xi|C_l|^2$, with ξ an expansion parameter]. We want to give a simple example that shows that even in this hypothesis exact solutions can be found.

The equation we consider is the following:

$$i\frac{dC_l}{dt} = \Omega(C_{l+1} + C_{l-1}) + f_l(t), \quad C_l(0) = \delta_{l,0}, \quad (4.6)$$

where $f_l(t)$ is a generic function depending both on the time and on the discrete index l.

It is easy to verify that the solution of (4.6) can be written as

$$C_l(t) = C_l^h(t) + \sum_{m=-\infty}^{+\infty}(-i)\int_0^t C_{l-m}^h(t-t')f_m(t')dt', \quad (4.7)$$

where $C_l^h(t)$ is the solution of the homogeneous case $[C_l^h(t) = (-i)^l J_l(2\Omega t)]$.

In a forthcoming paper we will apply all the previously found results to particular physical problems.

ACKNOWLEDGMENTS

The authors recognize the contributions of M. Richetta for checking the calculations and A. Dipace for critical remarks.

It is a pleasure to thank A. Renieri, S. Solimeno, and C. D. Cantrell for kind interest and comment. One of the authors (G. Dattoli) acknowledges the warm hospitality of the Department of Physics of the Napoli University where part of this work was carried out.

J. Gallardo wishes to acknowledge the support of the Office of Naval Research under Contract No. N00014-80-C-0308.

APPENDIX: A NOTE ON THE MAGNUS EXPANSION

The Magnus method has been discussed in the Introduction but, in the following sections, we have been mainly concerned with the Wei–Norman method, which has the advantage of being more general.

We have exploited that method even for cases in which it is not strictly necessary. In this way, however, we have shown the intimate connection between apparently disconnected problems.

In this Appendix we will discuss some examples where the Magnus expansion directly applies.

We consider a Hamiltonian of the type

$$\hat{H} = \omega_0 \hat{a}^+ \hat{a} + \Omega(\hat{a}^+ + \hat{a}) \quad (\hbar = 1), \tag{A1}$$

from which we can write down an interaction Hamiltonian of the type

$$\begin{aligned}\hat{H}_{\text{int}} &= \Omega e^{+i\omega_0 t \hat{a}^+ \hat{a}} (\hat{a}^+ + \hat{a}) e^{-i\omega_0 t \hat{a}^+ \hat{a}} \\ &= \Omega(\hat{a}^+ e^{+i\omega_0 t} + \hat{a} e^{-i\omega_0 t}).\end{aligned} \tag{A2}$$

The equation of the evolution operator writes

$$i\frac{d\hat{U}}{dt} = \hat{H}_{\text{int}} \hat{U}, \quad \hat{U}(0) = \hat{I}, \tag{A3}$$

whose explicit solution, using the expressions (1.12), reads

$$\begin{aligned}\hat{U}(t) &= \exp\left\{ i\frac{\omega_0}{2} \int_0^t dt' \, \Omega^2 \left(\frac{\sin \omega_0 t'/2}{\omega_0/2} \right)^2 \right\} \\ &\times \exp\left\{ -\frac{1}{2} \Omega^2 \left(\frac{\sin \omega_0 t/2}{\omega_0/2} \right)^2 \right\} \\ &\times \exp\left[-i\Omega \left(\frac{\sin(\omega_0 t/2)}{\omega_0/2} \right) \hat{a}^+ e^{i(\omega_0 t/2)} \right] \\ &\times \exp\left[-i\Omega \left(\frac{\sin(\omega_0 t/2)}{\omega_0/2} \right) \hat{a} e^{-i(\omega_0 t/2)} \right].\end{aligned} \tag{A4}$$

Using the above expression and assuming that we start from the vacuum, we easily get the following expression[12] for the evolution of $|\psi\rangle$:

$$\begin{aligned}|\psi\rangle &= \sum_{l=0}^{\infty} \exp\left\{ i\frac{\omega_0}{2} \int_0^t |\alpha(\tau')^2 d\tau' \right\} \\ &\times [(\alpha(t))^l / \sqrt{l!}] e^{-(|\alpha(t)|^2)/2} |l\rangle.\end{aligned} \tag{A5}$$

If we consider the Hamiltonian

$$\hat{H} = \omega_0 \hat{a}^+ \hat{a} + \epsilon(\hat{a}^+ \hat{a})^2 + \Omega(\hat{a}^+ + \hat{a}), \tag{A6}$$

the situation is considerably more complicated than before. However, following the same steps as before one finds

$$\begin{aligned}\hat{H}_{\text{int}} &= \Omega e^{+i[\omega_0 \hat{a}^+ \hat{a} + \epsilon(\hat{a}^+ \hat{a})^2]t} (\hat{a}^+ + \hat{a}) \\ &\quad \times e^{-i[\omega_0 \hat{a}^+ \hat{a} + \epsilon(\hat{a}^+ \hat{a})^2]t} \\ &= \Omega\{\exp[i(-\omega_0 - \epsilon(2\hat{a}^+ \hat{a} + 1))t]\hat{a} \\ &\quad + \exp[i(\omega_0 + \epsilon(2\hat{a}^+ \hat{a} - 1))t]\hat{a}^+\}.\end{aligned} \tag{A7}$$

If one is interested in a perturbed solution in ϵ, the Magnus procedure can be applied. The calculations are quite cumbersome, the Magnus series[11] must be calculated up to the fourth term and the results for $C_l(t)$ coincide, as they must, with (3.13) (for further comments and details see Ref. 12).

[1] R. P. Feyman, *Quantum Electrodynamics* (Benjamin, New York, 1961).
[2] N. Magnus, Commun. Pure Appl. Math. 7, 649 (1954).
[3] J. Wei and E. Norman, J. Math. Phys. A 4, 575 (1963).
[4] G. Baym, *Lectures on Quantum Mechanics* (Benjamin, New York, 1969).
[5] P. Pechukas and J. C. Light, J. Chem. Phys. 44, 3897 (1966).
[6] See, e.g., D. Finkelstein, Commun. Pure Appl. Math. 8, 245 (1955).
[7] Y. Kamo, Phys. Lett. A 56, 7 (1976).
[8] (a) P. Bosco, G. Dattoli, and M. Richetta, J. Phys. A 17, L395 (1984); (b) P. Bosco, J. Gallardo, and G. Dattoli, J. Phys. A 17, 2739 (1984); (c) G. Dattoli, J. Gallardo, and A. Torre, Lett. Nuovo Cimento 42, 163 (1985).
[9] C. W. Raman and N. S. Nath, Proc. Ind. Acad. Sci. 2, 406 (1936).
[10] N. N. Lebedev, *Special Functions and Their Applications* (Dover, New York, 1972).
[11] R. Gilmore, *Lie Group, Lie Algebras and Some of Their Applications* (Wiley, New York, 1972).
[12] F. Ciocci, G. Dattoli, and M. Richetta, J. Phys. A 17, 1333 (1984); G. Dattoli, M. Richetta, and I. Pinto, Nuovo Cimento D 4, 293 (1984).
[13] B. Mache, Opt. Commun. 28, 131 (1979); M. V. Fedorov, Sov. Phys. JETP 46, 69 (1977); P. Bosco and G. Dattoli, J. Phys. A 16, 4409 (1984).
[14] G. Dattoli, J. Gallardo, and A. Torre, submitted for publication.

A unified treatment of Wigner $\mathscr{D}$ functions, spin-weighted spherical harmonics, and monopole harmonics

Tevian Dray[a)]
School of Mathematics, Institute for Advanced Study, Princeton, New Jersey 08540

(Received 18 April 1985; accepted for publication 30 September 1985)

A unified, self-contained treatment of Wigner $\mathscr{D}$ functions, spin-weighted spherical harmonics, and monopole harmonics is given, both in coordinate-free language and for a particular choice of coordinates.

I. INTRODUCTION

We show in this paper that three independent generalizations of the usual spherical harmonics on S^2, namely Wigner $\mathscr{D}$ functions,[1–4] spin-weighted spherical harmonics,[5–7] and monopole harmonics[8–9] are completely equivalent.

It is well known that the Wigner $\mathscr{D}$ functions form an orthogonal basis for $L^2(S^3)$, the square-integrable functions on S^3. Monopole harmonics, on the other hand, form an orthonormal basis for (square integrable) sections of (all of the) complex line bundles over S^2. A standard result from the theory of fiber bundles (see, e.g., Ref. 10), however, asserts that these two concepts are entirely equivalent; this will be discussed in more detail in Sec. II below. Thus, monopole harmonics are equivalent to Wigner $\mathscr{D}$ functions. Finally, spin-weighted spherical harmonics can also be interpreted as sections of complex line bundles over S^2 and are therefore the same as monopole harmonics. (This was checked in coordinates by Dray.[11]) Goldberg *et al.*[6] showed directly that the spin-weighted spherical harmonics are equivalent to the Wigner $\mathscr{D}$ functions. Thus, all three of these concepts are equivalent; this paper is devoted to making this equivalence precise.

The paper is divided into two parts. In Part I (Secs. II–V) we give precise mathematical definitions in coordinate-free language of all three kinds of harmonics and establish their equivalence. In Part II (Secs. VI–VIII) we repeat the results of Part I in a particular choice of coordinates, thus establishing a direct connection between the precise mathematical definitions of Part I and the standard literature, which is mostly in the coordinate language of Part II. Parts I and II are written so as to be independent of each other; part of the purpose of this paper is to serve as a dictionary between the coordinate and coordinate-free versions of these results. Some readers may prefer to skip Part I on first reading. However, we feel that it is only in the coordinate-free language of Part I that the fundamental nature of the equivalence of the three kinds of harmonics becomes apparent.

The equivalence of the monopole harmonics to the Wigner $\mathscr{D}$ functions is at least implicitly contained in Refs. 9 and 12 while the interpretation of the spin-weighted spherical harmonics as sections of complex line bundles, and thus their equivalence to monopole harmonics, is also known. However, several features of our presentation are new. Foremost among these is the fact that the standard definition of spin-weighted spherical harmonics does *not* make explicit the fact that they are sections of a fiber bundle.[13] We interpret the standard definition as defining spin-weighted spherical harmonics to be functions on the (unit) tangent bundle to S^2; we make this precise in Sec. IV below and show the equivalence of our definition to the standard definition in Sec. VIII.

For monopole harmonics the situation is somewhat better in that the fiber bundle structure has been given explicitly.[10] However, the monopole harmonics themselves have only been given explicitly with respect to one particular trivializing cover of the complex line bundles.[8,14] We introduce the monopole harmonics in Sec. V as sections of the complex line bundles *irrespective* of local trivializations. The *explicit* coordinate version of both the monopole harmonics and the spin-weighted spherical harmonics in an *arbitrary* local trivialization ("spin-gauge"), given in Sec. VIII [(175) with (166)], is new.

The definition of the Wigner $\mathscr{D}$ functions in Sec. III as the matrix representation of SU(2) acting on irreducible representations of SU(2) in $L^2(S^3)$ is also new. This is usually done only for *integer* spin.[15] Our approach has the advantage that it can be done in a coordinate-free way, i.e., *without* introducing a parametrization in terms of Euler angles.

Finally, one motivation for this work was the desire to provide a self-contained, consistent presentation of these three types of harmonics in order to eliminate the necessity of worrying about which conventions have been used in the three different sets of literature.[16]

The paper is organized as follows. In Sec. II we introduce the mathematical concepts and notation that we will use throughout Part I. Sections III, IV, and V define, respectively, Wigner $\mathscr{D}$ functions, spin-weighted spherical harmonics, and monopole harmonics in abstract, coordinate-free language. Each definition is compared to the previous definition(s) as it arises. Finally, in Part II the results of Part I are rewritten in coordinate language and related to previous work. Section VI reproduces the notation of Sec. II in coordinate language, while Sec. VII does the same for the Wigner $\mathscr{D}$ functions of Sec. III. Section VIII then discusses both spin-weighted spherical harmonics (Sec. IV) and monopole harmonics (Sec. V) in coordinate language.

II. NOTATION

In this section we define angular momentum operators and give the basic properties of complex line bundles in order to fix our notation. The results are standard; our presentation is largely based on Kuwabara[17] and Greub and Petry.[10]

a) Present address: Department of Mathematics, University of York, York YO1 5DD, England.

 0022-2488/86/030781-12$02.50

Explicit coordinate versions of most of the results appear in Sec. VI. The generalization of many of the concepts presented here to higher-dimensional vector bundles over higher-dimensional spaces is discussed by Guillemin and Uribe.[18]

Let the isomorphism of SU(2): = SU(2,C) and S^3 be given by

$$T\colon \mathrm{SU}(2) \xrightarrow{\approx} S^3 \tag{1}$$

and let $B \in \mathrm{SU}(2)$ act on S^3 on the left via

$$p \mapsto Bp := T[B(T^{-1}p)]. \tag{2}$$

The corresponding action of SU(2) on $L^2(S^3)$ is (L^2 denotes the set of square integrable functions)

$$f \mapsto D(B)f, \tag{3a}$$

$$D(B)f|_p := f|_{B^{-1}p}, \tag{3b}$$

$$D(B'B)f \equiv D(B')D(B)f. \tag{3c}$$

Similarly, let $A \in \mathrm{SO}(3) := \mathrm{SO}(3,\mathbf{R})$ act on S^2 on the left via

$$x \mapsto Ax; \tag{4}$$

the corresponding action of SO(3) on $L^2(S^2)$ is

$$g \mapsto D(A)g, \tag{5a}$$

$$D(a)g|_x := g|_{A^{-1}x}. \tag{5b}$$

Consider

$$\mathrm{U}(1) = \{H(\lambda) : \lambda \in [0,2\pi)\} \subset \mathrm{SU}(2), \tag{6a}$$

$$H(a)H(b) = H(a+b), \tag{6b}$$

$$H(\lambda) = 1 \Leftrightarrow \lambda = 0 \pmod{2\pi}.$$

The Hopf bundle is defined to be the principal bundle

$$\begin{array}{l} \mathrm{U}(1) \to S^3 \\ \qquad\quad \downarrow \pi \\ \qquad\quad S^2 \end{array}, \tag{7}$$

where the (right) action of U(1) on S^3 is given by

$$p \mapsto pe^{i\lambda} := T[(T^{-1}p)H(\lambda)]. \tag{8}$$

Thus, $\pi(pe^{i\lambda}) = \pi(p)$ and $\pi(Bpe^{i\lambda}) = \pi(Bp)$. We therefore get an induced map

$$\hat{\pi} : \mathrm{SU}(2) \to \mathrm{SO}(3), \qquad \hat{\pi}(B)\pi(p) \equiv \pi(Bp). \tag{9}$$

We will *assume* that the U(1) subgroup of SU(2) in (6) has been chosen so that the Chern class of the Hopf bundle is $[R]$, where $R = -(i/2)\Omega$ and Ω is the volume form on S^2. Thus the Hopf bundle has Chern number

$$\oint_{S^2} \frac{iR}{2\pi} = +1 \tag{10}$$

instead of -1 (the only other possibility). [This can always be achieved by replacing $H(\lambda)$ by $H(-\lambda)$ if necessary.]

We wish to introduce a basis $\Lambda_a \in \mathrm{su}(2)$, the Lie algebra of SU(2), which satisfies

$$[\Lambda_a, \Lambda_b] = \epsilon_{abc}\Lambda_c, \tag{11}$$

where the indices run from 1 to 3 and ϵ_{abc} is the totally antisymmetric tensor defined by $\epsilon_{123} = +1$. However, if we define

$$\beta_a(\tau) := \exp(\tau\Lambda_a) \in \mathrm{SU}(2), \tag{12}$$

for Λ_a satisfying (11), then β_a is periodic in τ *with period* 4π.

We define

$$\Lambda_3 := \frac{d}{d\tau}\bigg|_{\tau=0} H\left(-\frac{\tau}{2}\right) \tag{13}$$

(the minus sign is conventional) and choose Λ_1 and Λ_2 so that (11) is satisfied; note that we now have

$$H(\lambda) \equiv \beta_3(-2\lambda). \tag{14}$$

Introduce angular momentum operators on S^3 via

$$J_a : L^2(S^3) \to L^2(S^3), \qquad f \mapsto i\frac{d}{d\tau}\bigg|_{\tau=0} D(\beta_a(\tau))f, \tag{15a}$$

i.e.,

$$J_a f|_p = -i\frac{d}{d\tau}\bigg|_{\tau=0} f|_{\beta_a(\tau)p}. \tag{15b}$$

Then we have

$$[J_a, J_b] = i\epsilon_{abc}J_c. \tag{16}$$

Define

$$J^2 := \sum_a J_a^2, \qquad J_\pm := J_1 \pm iJ_2. \tag{17}$$

Then [compare (131) below] $J^2 \equiv -\frac{1}{4}\Box_3$, where $\Box_3$ is the standard Laplacian on S^3.

If we now define angular momentum operators on S^2 by

$$\hat{J}_a : L^2(S^2) \to L^2(S^2), \qquad g \mapsto i\frac{d}{d\tau}\bigg|_{\tau=0} D(\alpha_a(\tau))g, \tag{18}$$

where $\alpha_a(\tau) := \hat{\pi}(\beta_a(\tau))$, then we have

$$\hat{J}_a g|_{\pi(p)} \equiv J_a(g \circ \pi)|_p \quad [g \in L^2(S^2)]. \tag{19}$$

Note that J_a and $\hat{J}_a$ are Hermitian operators.

The complex line bundles E_n associated with the Hopf bundle can be defined as follows. Let U(1) act on $\mathbb{C}$ via multiplication, i.e.,

$$\mathbb{C} \to \mathbb{C}, \qquad z \mapsto e^{i\lambda}z. \tag{20}$$

Define

$$E_n := S^3 \times_n \mathbb{C} = \{[(p,z)]\}, \tag{21}$$

where the square brackets denote equivalence classes under the relation

$$(p,z) \sim (pe^{i\lambda}, e^{in\lambda}z), \tag{22}$$

E_n is a fiber bundle over S^3 with fiber $\mathbb{C}$,

$$\begin{array}{l} E_n \\ \downarrow \pi_n \\ S^2 \end{array}, \tag{23}$$

and there is a U(1) action on E_n given by

$$[(p,z)] \mapsto [(pe^{i\lambda}, z)]. \tag{24}$$

The projection π_n is given by

$$\pi_n([(p,z)]) := \pi(p). \tag{25}$$

There is a natural connection on E_n (Ref. 19) so that the curvature of E_n is [see (155)]

$$R_n = -(in/2)\Omega, \tag{26}$$

where Ω is the volume form on S^2, so that the Chern number of E_n is [compare (10)]

$$\oint_{S^2} \frac{iR_n}{2\pi} \equiv +n. \tag{27}$$

We will assume throughout the remainder of the paper that this connection has been chosen.[20]

There is an existence and uniqueness theorem which says that a line bundle over S^2 with curvature R exists if and only if

$$\oint \frac{iR}{2\pi} \in \mathbf{Z} \tag{28}$$

and that this bundle is unique up to strong bundle isomorphism.

Let

$$\widehat{F}_n := \{f \in C^\infty(S^3) : f(pe^{i\lambda}) = e^{in\lambda} f(p)\}. \tag{29}$$

Given any $f \in \widehat{F}_n$ we can obtain a C^∞ section σ_f of E_n by

$$\begin{aligned} \sigma_f &: S^2 \to E_n, \\ x &\mapsto [(p, f(p))], \end{aligned} \tag{30}$$

for any p such that $\pi(p) = x$. Denote by Q the map

$$Q(f) := \sigma_f. \tag{31}$$

Note that Q is one-to-one: A C^∞ section σ_f uniquely determines a function f on S^3 via (30) which much be in $\widehat{F}_n$ in order to be well-defined. We write the inverse mapping as

$$f_\sigma := Q^{-1}(\sigma). \tag{32}$$

Given any smooth (C^∞) local section

$$\begin{aligned} &U_A \subset S^2, \\ &\hat{\gamma}_A : U_A \to S^3, \\ &\pi \circ \hat{\gamma}_A \equiv \mathbf{1} \end{aligned} \tag{33}$$

of the Hopf bundle we can interpret a section $\sigma \in \Gamma_n$ of E_n as a *function*

$$g_A^\sigma := f_\sigma \circ \hat{\gamma}_A \in L^2(U_A). \tag{34}$$

For $x \in U_A$ and sections σ, τ of E_n we define the scalar product

$$\begin{aligned} \langle \sigma, \tau \rangle_x &:= \overline{f_\sigma(\hat{\gamma}_A(x))} \cdot f_\tau(\hat{\gamma}_a(x)), \\ &\equiv \overline{g_A^\sigma} g_A^\tau |_x. \end{aligned} \tag{35}$$

Note that since f_σ and f_τ are *both* elements of $\widehat{F}_n$ the norm is independent of the choice of local section γ_A so long as x is in the domain of definition of γ_A. We can now define

$$\langle \sigma, \tau \rangle := \oint_{S^2} \langle \sigma, \tau \rangle_x \, dx. \tag{36}$$

Note that Q is *not* an isometry but satisfies [compare (144)]

$$\langle f, f \rangle \equiv (\pi/2) \langle Q(f), Q(f) \rangle. \tag{37}$$

Thus, if we define $F_n := \widehat{F}_n \cap L^2(S^3)$ and let $\Gamma_n := L^2(S^2 \to E_n)$ denote the set of square integrable sections of E_n, then Q clearly gives a one-to-one correspondence

$$Q : F_n \to \Gamma_n. \tag{38}$$

We can now introduce angular momentum operators $\hat{L}_a$ on Γ_n via

$$\hat{L}_a \sigma := Q(J_a(Q^{-1}\sigma)). \tag{39}$$

For the connection chosen above [compare (26)] the Laplacian on Γ_n is [see (157) below][17,19]

$$\Delta_n \equiv -\hat{L}^2 + n^2/4. \tag{40}$$

III. WIGNER $\mathscr{D}$ FUNCTIONS

Wigner[1] introduced the functions $\mathscr{D}^l_{qm}$ as the matrix elements of finite rotations acting on irreducible representations of the rotation group [SO(3)]. Our presentation is largely based on that of Edmonds.[2] Other standard references are Rose[3] and the more modern treatment given by Biedenharn and Louck.[4]

We can define an irreducible representation of SU(2) on S^3 for each l as follows.

Choose $\phi_{ll} \in L^2(S^3)$ with

$$J^2 \phi_{ll} = l(l+1)\phi_{ll}, \quad J_3 \phi_{ll} = l\phi_{ll}, \tag{41}$$

and define

$$\phi_{l,l-n} = [n(2l-n+1)]^{-1/2} J_- \phi_{l,l-n+1} \quad (n = 1,2,\ldots,2l). \tag{42}$$

Then

$$J_3 \phi_{lm} = m\phi_{lm}, \tag{43}$$

$$J_\pm \phi_{lm} = [(l \mp m)(l \pm m + 1)]^{1/2} \phi_{l,m\pm 1}, \tag{44}$$

and

$$\langle \phi_{lm}, \phi_{l'm'} \rangle = \langle \phi_{ll}, \phi_{ll} \rangle \delta_{ll'} \delta_{mm'}, \tag{45}$$

where $\langle \cdot, \cdot \rangle$ denotes the L^2 norm on S^3. Note that since $[J^2, J_-] = 0$, (42) implies

$$J^2 \phi_{lm} \equiv l(l+1)\phi_{lm}; \tag{46}$$

this also follows directly from (43) and (44). From (43)–(46) we see that there is a (matrix) representation of SU(2) on the vector spaces

$$W^l := \text{Span}\{\phi_{lm} : 0 \leqslant l - |m| \in \mathbf{Z}\}, \tag{47}$$

for each l. The representation is irreducible because W^l is generated by the action of J_- on ϕ_{ll} [see (42)].

We can now define the Wigner $\mathscr{D}$ functions to be the matrix representation of SU(2) acting on W^l:

$$\begin{aligned} D(B) &: W^l \to W^l, \\ \phi_{lm} &\mapsto \sum_q \phi_{lq} \mathscr{D}^l_{qm}(B) \end{aligned} \tag{48}$$

with $D(B)$ as in (3). Note that this construction is independent of the choice of ϕ_{ll} satisfying (41).

Before deriving the properties of the Wigner $\mathscr{D}$ functions we first show that, for *integer* spin ($l \in \mathbf{Z}$), our definition is the same as the usual one in terms of spherical harmonics on S^2. We can introduce the usual spherical harmonics on S^2 via

$$\begin{aligned}&\hat{J}_3 Y_{lm} = mY_{lm},\\&\hat{J}_{\pm} Y_{lm} = [(l \mp m)(l \pm m+1)]^{1/2} Y_{l,m\pm 1},\\&\langle Y_{lm}, Y_{l'm'}\rangle = \delta_{ll'}\delta_{mm'}\\&\quad (l\in \mathbf{Z},\quad 0 < l-|m| \in \mathbf{Z})\end{aligned} \tag{49}$$

and a choice of phase for each l, where $\hat{J}_{\pm} := \hat{J}_1 \pm i\hat{J}_2$ and the norm is now the $\mathbf{L}^2$ norm on $\mathbf{S}^2$. Defining $Y^l_{0m} := Y_{lm} \circ \pi \in \mathbf{L}^2(\mathbf{S}^3)$ we see that the Y^l_{0m} satisfy (43)–(46) and thus

$$D(B)Y^l_{0m} \equiv \sum_q Y^l_{0q}\mathscr{D}^l_{qm}(B). \tag{50}$$

But

$$\begin{aligned}D(B)Y^l_{0m}|_p &= Y_{lm} \circ \pi|_{B^{-1}p}\\&= Y_{lm}|_{\hat{\pi}(B^{-1})\pi(p)}\\&= D(\hat{\pi}(B))Y_{lm} \circ \pi,\end{aligned}$$

and therefore

$$D(\hat{\pi}(B))Y_{lm} \equiv \sum_q Y_{lq}\mathscr{D}^l_{qm}(B), \tag{51}$$

which is equivalent to the standard definition of the functions $\mathscr{D}^l_{qm}$ in terms of SO(3).

We now establish the various properties of the $\mathscr{D}^l_{\rm qm}$. From (45) we have

$$\begin{aligned}&\langle\phi_{ll},\phi_{ll}\rangle\delta_{ll'}\delta_{mm'}\\&\quad= \oint_{\mathbf{S}^3} \overline{\phi_{lm}(p)}\phi_{l'm'}(p)dS\\&\quad\equiv \oint_{\mathbf{S}^3} \overline{\phi_{lm}(B^{-1}p)}\phi_{l'm'}(B^{-1}p)dS\\&\quad\equiv \sum_{q,q'}\oint_{\mathbf{S}^3} \overline{\phi_{lq}(p)\mathscr{D}^l_{qm}(B)}\phi_{l'q'}(p)\mathscr{D}^{l'}_{q'm'}(B)dS\\&\quad\equiv \sum_q \overline{\mathscr{D}^l_{qm}(B)}\mathscr{D}^{l'}_{qm'}(B)\langle\phi_{ll},\phi_{ll}\rangle\end{aligned}$$

and therefore

$$\sum_q \overline{\mathscr{D}^l_{qm}(B)}\mathscr{D}^l_{qm'}(B) \equiv \delta_{mm'}. \tag{52}$$

From (3c) we also have

$$\mathscr{D}^l_{qm}(B'B) = \sum_n \mathscr{D}^l_{qn}(B')\mathscr{D}^l_{nm}(B). \tag{53}$$

Setting $B' = B^{-1}$ and relabeling indices we get

$$\delta_{mm'} = \sum_q \mathscr{D}^l_{mq}(B^{-1})\mathscr{D}^l_{qm'}(B), \tag{54}$$

so that we finally obtain

$$\overline{\mathscr{D}^l_{qm}(B)} \equiv \mathscr{D}^l_{mq}(B^{-1}). \tag{55}$$

In other words, the matrix $(\mathscr{D}^l)_{qm} = \mathscr{D}^l_{qm}$ is unitary; $(\mathscr{D}^l)^{-1} = \overline{(\mathscr{D}^l)}^t$.

Define the operators L_a and K_a on $\mathbf{L}^2(\mathrm{SU}(2))$ via

$$L_a h|_B := -i\frac{d}{d\tau}\bigg|_{\tau=0} h|_{\beta_a(\tau)B}, \tag{56a}$$

$$K_a h|_B := +i\frac{d}{d\tau}\bigg|_{\tau=0} h|_{B\beta_a(\tau)}, \tag{56b}$$

where $h \in \mathbf{L}^2(\mathrm{SU}(2))$. We then have

$$\begin{aligned}\phi_{lq}(L_a\mathscr{D}^l_{qm}(B)) &= -i\frac{d}{d\tau}\bigg|_{\tau=0}\phi_{lq}\mathscr{D}^l_{qm}(\beta_a(\tau)B)\\&= -i\frac{d}{d\tau}\bigg|_{\tau=0} D(\beta_a(\tau)B)\phi_{lq}\\&\equiv -J_a[D(B)\phi_{lq}],\end{aligned} \tag{57}$$

whereas

$$\begin{aligned}\phi_{lq}(K_a\mathscr{D}^l_{qm}(B)) &= i\frac{d}{d\tau}\bigg|_{\tau=0}\phi_{lq}\mathscr{D}^l_{qm}(B\beta_a(\tau))\\&= i\frac{d}{d\tau}\bigg|_{\tau=0} D(B\beta_a(\tau))\phi_{lq}\\&\equiv D(B)[J_a\phi_{lq}].\end{aligned} \tag{58}$$

Using (43) and (44) it is now easy to compute

$$\begin{aligned}L_3\mathscr{D}^l_{qm} &= -q\mathscr{D}^l_{qm},\\L_{\pm}\mathscr{D}^l_{qm} &= -[(l\pm q)(l\mp q+1)]^{1/2}\mathscr{D}^l_{q\mp 1,m};\end{aligned} \tag{59}$$

$$\begin{aligned}K_3\mathscr{D}^l_{qm} &= m\mathscr{D}^l_{qm},\\K_{\pm}\mathscr{D}^l_{qm} &= [(l\mp m)(l\pm m+1)]^{1/2}\mathscr{D}^l_{q,m\pm 1};\end{aligned} \tag{60}$$

where $L_{\pm} := L_1 \pm iL_2$ and $K_{\pm} := K_1 \pm iK_2$. Note that *both* L_a and K_a satisfy the usual commutation relations, namely those satisfied by J_a, and that $L^2 \equiv K^2$.

From (59) and (60) we see that the $\mathscr{D}^l_{qm}$ are orthogonal, i.e.,

$$\langle\mathscr{D}^l_{qm},\mathscr{D}^{l'}_{q'm'}\rangle = \delta_{ll'}\delta_{qq'}\delta_{mm'}C_l, \tag{61}$$

where the norm is the $\mathbf{L}^2$ norm on SU(2) and where $C_l := \langle\mathscr{D}^l_{ll},\mathscr{D}^l_{ll}\rangle$. [Note that by (59) and (60) the normalization depends only on l.] But from (52) we have

$$\sum_q \langle\mathscr{D}^l_{qm},\mathscr{D}^l_{qm}\rangle \equiv 2\pi^2, \tag{62}$$

so that

$$C_l = 2\pi^2/(2l+1) \tag{63}$$

[which shows that $\mathscr{D}^l_{qm} \in \mathbf{L}^2(\mathrm{SU}(2))$].

Note that $\mathbf{L}^2(\mathbf{S}^3)$ is of course isomorphic to $\mathbf{L}^2(\mathrm{SU}(2))$ via

$$\begin{aligned}&T^*: \mathbf{L}^2(\mathbf{S}^3) \to \mathbf{L}^2(\mathrm{SU}(2)),\\&\quad f \mapsto f\circ T.\end{aligned} \tag{64}$$

We thus define

$$\begin{aligned}\hat{\mathscr{D}}^l_{qm} &:= T_*\mathscr{D}^l_{qm},\\&\equiv \mathscr{D}^l_{qm}\circ T^{-1} \in \mathbf{L}^2(\mathbf{S}^3).\end{aligned} \tag{65}$$

Under this isomorphism we have

$$L_a(f\circ T) \equiv J_a f\circ T, \tag{66}$$

and we see that the matrix representation of SU(2) on the space spanned by the $\hat{\mathscr{D}}^l_{qm}$ for fixed l and m [given by (59)] is *not* the same as the matrix representation of SU(2) on W^l [given by (43) and (44)]. We can fix this by defining

$$\mathscr{Y}^l_{qm}(B) := \sqrt{(2l+1)/4\pi}\,\mathscr{D}^l_{-q,m}(B^{-1}), \\ \equiv \sqrt{(2l+1)/4\pi}\,\overline{\mathscr{D}^l_{m,-q}(B)}, \tag{67}$$

and

$$\hat{\mathscr{Y}}^l_{qm} := \mathscr{Y}^l_{qm} \circ T^{-1} \in \mathrm{L}^2(\mathrm{S}^3), \tag{68}$$

where the factor $\sqrt{(2l+1)/4\pi}$ has been added for convenience.

Note that

$$L_a \mathscr{Y}^l_{qm}|_B = -i \frac{d}{d\tau}\Big|_{\tau=0} \mathscr{Y}^l_{qm}|_{\beta_a(\tau)B}$$
$$= -i\sqrt{\frac{2l+1}{4\pi}} \frac{d}{d\tau}\Big|_{\tau=0} \mathscr{D}^l_{-qm}|_{[\beta_a(\tau)B]^{-1}}$$
$$= +i\sqrt{\frac{2l+1}{4\pi}} \frac{d}{d\tau}\Big|_{\tau=0} \mathscr{D}^l_{-qm}|_{B^{-1}\beta_a(\tau)},$$

so that

$$L_a \mathscr{Y}^l_{qm}|_B \equiv \sqrt{(2l+1)/4\pi}\, K_a \mathscr{D}^l_{-qm}|_{B^{-1}} \tag{69}$$

and similarly

$$K_a \mathscr{Y}^l_{qm}|_B \equiv \sqrt{(2l+1)/4\pi}\, L_a \mathscr{D}^l_{-qm}|_{B^{-1}}. \tag{70}$$

Thus

$$L_3 \mathscr{Y}^l_{qm} = m\mathscr{Y}^l_{qm}, \\ L_\pm \mathscr{Y}^l_{qm} = [(l\mp m)(l\pm m+1)]^{1/2}\mathscr{Y}^l_{q,m\pm 1}; \tag{71}$$

$$K_3 \mathscr{Y}^l_{qm} = +q\mathscr{Y}^l_{qm}, \\ K_\pm \mathscr{Y}^l_{qm} = -[(l\mp q)(l\pm q+1)]^{1/2}\mathscr{Y}^l_{q\pm 1,m}, \tag{72}$$

and

$$\langle \mathscr{Y}^l_{qm}, \mathscr{Y}^{l'}_{q'm'}\rangle = (\pi/2)\delta_{ll'}\delta_{qq'}\delta_{mm'}. \tag{73}$$

The matrix representation of SU(2) on the spaces

$$W^l_q := \mathrm{span}\{\hat{\mathscr{Y}}_{qm} : 0 < l - |m| \in \mathbf{Z}\}, \tag{74}$$

for each q ($0<l-|q|\in\mathbf{Z}$) is now precisely the same as on W^l and there are $2l+1$ spaces W^l_q for each l. Using the Peter–Weyl theorem[21] we conclude that

$$\mathrm{L}^2(\mathrm{S}^3) \equiv \bigoplus_{l,q} W^l_q, \tag{75}$$

and that $\{\sqrt{2/\pi}\,\hat{\mathscr{Y}}^l_{qm}\}$ therefore forms an *orthonormal* basis for $\mathrm{L}^2(\mathrm{S}^3)$. The Wigner $\mathscr{D}$ functions $\{\hat{\mathscr{D}}^l_{qm}\}$ thus form an *orthogonal* basis for $\mathrm{L}^2(\mathrm{S}^3)$.

We now derive a property of the $\mathscr{Y}^l_{qm}$ that will be crucial in what follows. Note that

$$L_a \mathscr{Y}^l_{qm}|_{BH(\lambda)} \equiv L_a(\mathscr{Y}^l_{qm}(BH(\lambda))), \tag{76}$$

but that

$$K_a \mathscr{Y}^l_{qm}|_{BH(\lambda)} = i\frac{d}{d\tau}\Big|_{\tau=0} \mathscr{Y}^l_{qm}|_{BH(\lambda)\beta_a(\tau)}, \tag{77}$$

which, in general, is *not* equal to

$$K_a(\mathscr{Y}^l_{qm}(BH(\lambda))) = i\frac{d}{d\tau}\Big|_{\tau=0} \mathscr{Y}^l_{qm}|_{B\beta_a(\lambda)H(\tau)}. \tag{78}$$

However, since $[\beta_3(\tau), H(\lambda)] = 0$ we see that we *do* have

$$K_3 \mathscr{Y}^l_{qm}|_{BH(\lambda)} \equiv K_3(\mathscr{Y}^l_{qm}(BH(\lambda))). \tag{79}$$

But since the $\mathscr{Y}^l_{qm}$ are fully determined by their eigenvalues with respect to L^2, L_3, and K_3 and since they form a basis for $\mathrm{L}^2(\mathrm{SU}(2))$, we conclude that

$$\mathscr{Y}^l_{qm}(BH(\lambda)) \equiv c(\lambda)\mathscr{Y}^l_{qm}(B), \tag{80}$$

where c may depend on (l,q,m). Finally, using (14) we have

$$q\mathscr{Y}^l_{qm}|_B = K_3\mathscr{Y}^l_{qm}|_B$$
$$= i\frac{d}{d\tau}\Big|_{\tau=0} \mathscr{Y}^l_{qm}|_{B\beta_3(\lambda)}$$
$$= -\frac{i}{2}\frac{d}{d\lambda}\Big|_{\lambda=0} \mathscr{Y}^l_{qm}|_{BH(\lambda)}.$$

Thus

$$\frac{d}{d\lambda}\Big|_{\lambda=0} c(\lambda) = 2iq, \tag{81}$$

which, together with $c(0) = 1$ and $c(a+b) = c(a)c(b)$ [which follows from (6b)], implies $c(\lambda) = e^{2iq\lambda}$, so that

$$\mathscr{Y}^l_{qm}(BH(\lambda)) \equiv e^{2iq\lambda}\mathscr{Y}^l_{qm}(B) \tag{82}$$

and therefore

$$\hat{\mathscr{Y}}^l_{qm}(pe^{i\lambda}) \equiv e^{2iq\lambda}\hat{\mathscr{Y}}^l_{qm}(p), \tag{83}$$

i.e., $\hat{\mathscr{Y}}^l_{qm} \in F_{2q}$. In fact $\{\sqrt{2/\pi}\,\mathscr{Y}^l_{qm}\}$ for fixed q clearly forms an orthonormal basis for F_{2q}. We can now write (77) as (79) together with

$$K_\pm \mathscr{Y}^l_{qm}|_{BH(\lambda)} \equiv e^{\pm 2i\lambda}K_\pm(\mathscr{Y}^l_{qm}(BH(\lambda))). \tag{84}$$

IV. SPIN-WEIGHTED SPHERICAL HARMONICS

Newman and Penrose[5] introduced spin weighted spherical harmonics in a particular choice of spin gauge. The (trivial) generalization to an arbitrary spin gauge (for a particular choice of coordinates) can be found in Dray.[11] (A spinorial definition has been given by Penrose and Rindler.[7] See also Refs. 22 and 23.) Consider the complexified tangent bundle

$$\begin{array}{l} T_{\mathbf{C}}\mathrm{S}^2 := T\mathrm{S}^2 \otimes \mathbf{C} \\ \downarrow \\ \mathrm{S}^2 \end{array}, \tag{85}$$

and let m be a (complex) vector field on S^2, i.e., a *local* section of the tangent bundle

$$m : U \to T_{\mathbf{C}}\mathrm{S}^2 \quad (U \subset \mathrm{S}^2), \tag{86}$$

which satisfies

$$\langle m,m\rangle = 0, \quad \langle m,\bar{m}\rangle = 2, \tag{87}$$

at each point of U.

A quantity Q is said to have[5] *spin weight* s [we write $\mathrm{sw}(Q) = s$] if under the transformation

$$m \to e^{i\kappa}m, \tag{88a}$$

Q transforms according to

$$Q \mapsto e^{is\kappa}Q. \tag{88b}$$

What does this mean? We interpret this imprecise definition as follows.

Consider the space V consisting of all elements v of $T_{\mathbf{C}}\mathrm{S}^2$ satisfying (87). There is a natural decomposition

$$V = V_0 \cup \overline{V}_0, \tag{89a}$$

where for *any* $v_0 \in V_0$ we have

$$V_0 = \{e^{i\chi} v_0;\quad \chi\colon S^2 \to [0,2\pi)\}. \tag{89b}$$

Note that V_0 is a subbundle of the tangent bundle (85), i.e.,

$$\begin{array}{c} V_0 \\ \downarrow \tilde{\pi} \\ S^2 \end{array} \tag{90}$$

Furthermore, since we have a natural U(1) action on V_0, namely

$$V_0 \to V_0,\quad v \mapsto e^{i\lambda} v, \tag{91a}$$

and since

$$\tilde{\pi}(e^{i\lambda} v) \equiv \tilde{\pi}(v), \tag{91b}$$

V_0 is clearly a circle bundle over S^2. But since $V_0 \otimes \mathbb{R}$ is equivalent to the *real* tangent bundle TS^2 and since TS^2 is bundle isomorphic to E_1, there is a fiber-preserving isomorphism[24]

$$\eta\colon V_0 \overset{\approx}{\to} S^3,\quad \eta(e^{i\lambda} v) \equiv \eta(v) e^{i\lambda}. \tag{92}$$

We therefore interpret "quantities of spin weight s" to be functions on V_0, i.e., elements of $L^2(S^3)$, with a particular behavior under the circle action. We must, however, be extremely careful here: The vector field m in the usual definition of spin-weighted spherical harmonics has a definite behavior under the circle action, so we are *not* free to specify this independently. We claim that the correct choice is to require m to behave in the same way as K_+ under the circle action, namely [compare (84)]

$$m \mapsto e^{+2i\lambda} m \tag{93}$$

[so that $\kappa = 2\lambda$ in (88)]. We will see below [compare (177c)] that this correctly reproduces the standard definition in coordinates.

We thus define a "quantity of spin weight s" to be a function

$$\tilde{f}\colon V \to \mathbb{C} \tag{94a}$$

satisfying

$$\tilde{f}(e^{i\lambda} v) \equiv e^{+2is\lambda} \tilde{f}(v). \tag{94b}$$

We can turn $\tilde{f}$ into a function $f := \tilde{f} \circ \eta^{-1}$ on S^3, and we therefore define *spin-weighted functions on* S^3 to be elements f of $L^2(S^3)$ satisfying

$$f(pe^{i\lambda}) \equiv e^{+2is\lambda} f(p), \tag{95}$$

i.e., $f \in F_{2s}$ for some s, and define s to be the *spin weight* of f ($\mathrm{sw}(f) = s$). But note that from (83) we have

$$\mathrm{sw}(\hat{\mathscr{Y}}^l_{qm}) \equiv q, \tag{96}$$

so that $\{\hat{\mathscr{Y}}^l_{qm}\}$ for fixed q forms a basis for the functions of spin weight q for $2q \in \mathbb{Z}$. We call the $\hat{\mathscr{Y}}^l_{qm}$ *spin-weighted spherical harmonics on* S^3.

Note further that, for *integer* spin ($l \in \mathbb{Z}$), (72) implies

$$\mathscr{Y}^l_{qm} = \begin{cases} \left[\dfrac{(l-q)!}{(l+q)!}\right]^{1/2} (-1)^q (K_+)^q \mathscr{Y}^l_{0m} & (0<q<l), \\ \left[\dfrac{(l+q)!}{(l-q)!}\right]^{1/2} (-1)^q (K_-)^{|q|} \mathscr{Y}^l_{0m} & (-l<q<0), \\ 0 & (l<|q|), \end{cases} \tag{97}$$

which is equivalent to the standard definition of spin-weighted spherical harmonics for integer spin.

V. MONOPOLE HARMONICS

Greub and Petry[10] were the first to introduce the idea of using a Hilbert space of sections of complex line bundles to obtain a description of the Dirac magnetic monopole which is free of string singularities. Wu and Yang[8] independently discovered the same idea and gave an orthonormal basis for this Hilbert space (with respect to a particular trivialization of the bundles) which they called *monopole harmonics*. (A combined treatment of these two approaches can be found in Biedenharn and Louck.[9])

Consider the electromagnetic field $\hat{F}$ of a Dirac magnetic monopole of strength g located at the origin. The field $\hat{F}$ is a spherically symmetric, time-independent two-form over $\mathbb{R}^4$ so it is sufficient to consider the pullback F of $\hat{F}$ to S^2. Then we have

$$F = g\Omega, \tag{98}$$

where Ω is the volume form on S^2, i.e.,

$$\oint_{S^2} \Omega = 4\pi. \tag{99}$$

Maxwell's equations imply that F is closed, i.e., $dF = 0$, but we do *not* assume that F is exact, i.e., we do not assume that there exists a *globally* defined vector potential A satisfying $dA = F$.

The Schrödinger equation for a particle with electric charge e and mass m moving in this field can be written

$$i\,\partial_t \psi = -\frac{1}{2m}\left(\partial_r^2 + \frac{2}{r}\partial_r + \frac{\Delta}{r^2}\right)\psi, \tag{100}$$

which we interpret as follows. Make the ansatz

$$\psi = e^{-iEt} \rho(r) \sigma,$$

where σ is a section of the line bundle over S^2 with curvature $R = -ieF$ (see Ref. 25). Then Δ represents the natural Laplacian [compare (26) and (40)] on this bundle. The Schrödinger equation now becomes

$$r^2(2mE\rho + \rho'' + 2\rho'/r)\sigma = -\rho(\Delta\sigma), \tag{101}$$

so that the angular part of the wave function, σ, must be an eigensection of Δ.

For the line bundle with curvature $R = -ieF$ to exist we must have [see (28)]

$$\oint \frac{eF}{2\pi} \in \mathbb{Z}, \tag{102}$$

which is just the Dirac quantization condition

$$2eg \in \mathbb{Z}; \tag{103}$$

σ is thus a section of E_{2q} for $q = eg$. We thus have [see (40)]

$$\Delta \equiv \Delta_{2q} \equiv -\hat{L}^2 + q^2 \tag{104}$$

so that eigensections of Δ are also eigensections of $\hat{L}^2$. We are only interested in sections σ that are square integrable, so that $\sigma \in \Gamma_{2q}$, and thus $Q^{-1}\sigma \in F_{2q}$. But we have seen that $\{\hat{\mathscr{Y}}^l_{qm}\}$ forms a basis for F_{2q}. We are thus led to define the *monopole harmonics*

$$\mathscr{Y}_{qlm} := Q(\hat{\mathscr{Y}}^l_{qm}) \in \Gamma_{2q}. \tag{105}$$

The monopole harmonics can also be defined intrinsically by the conditions

$$\begin{aligned} &\mathscr{Y}_{qlm} \in \Gamma_{2q},\\ &\hat{L}^2\mathscr{Y}_{qlm} = l(l+1)\mathscr{Y}_{qlm},\\ &\hat{L}_3\mathscr{Y}_{qlm} = m\mathscr{Y}_{qlm},\\ &\langle \mathscr{Y}_{qlm}, \mathscr{Y}_{q'l'm'}\rangle = \delta_{ll'}\delta_{qq'}\delta_{mm'}; \end{aligned} \tag{106}$$

these follow from the definition (105) together with (71), (73), and (37).

We thus see that the monopole harmonics are completely equivalent to the spin-weighted spherical harmonics on S^3, where the equivalence is given by the mapping Q.

VI. COORDINATE NOTATION

In this section we introduce a particular coordinatization of SU(2) [and thus also of SO(3) and S^3] in terms of Euler angles. We then introduce the complex line bundles over S^2 and angular momentum operators in terms of these coordinates.

The group $SU(2) := SU(2,\mathbb{C})$ can be defined as the set of 2×2 complex matrices satisfying

$$B^{-1} = \overline{B^t}, \quad \det B = 1, \tag{107a}$$

or equivalently

$$B = \begin{pmatrix} a & b\\ -\bar{b} & \bar{a}\end{pmatrix} \quad (a\bar{a} + b\bar{b} = 1). \tag{107b}$$

Similarly, $SO(3) := SO(3,\mathbb{R})$ can be defined as the set of 3×3 real matrices satisfying

$$A^{-1} = A^t, \quad \det A = 1. \tag{108}$$

We choose the parametrization for SU(2) given by

$$\begin{aligned} a &= \cos(\beta/2)\, e^{-i(\gamma+\alpha)/2},\\ b &= \sin(\beta/2)\, e^{+i(\gamma-\alpha)/2}, \end{aligned} \tag{109a}$$

where

$$\beta \in [0,\pi], \quad \alpha \in [0,2\pi), \quad \gamma \in [0,4\pi); \tag{109b}$$

we write $B(\alpha,\beta,\gamma)$ for the matrix so determined.[26]

We consider S^3 to be the subspace of $\mathbb{C}^2$ defined by

$$S^3 := \left\{\begin{pmatrix} u\\ v\end{pmatrix} \in \mathbb{C}^2 : u\bar{u} + v\bar{v} = 1\right\}. \tag{110}$$

We choose the parametrization

$$\begin{aligned} u &= \cos(\theta/2)e^{-i(\psi+\phi)/2},\\ v &= -\sin(\theta/2)e^{-i(\psi-\phi)/2}, \end{aligned} \tag{111a}$$

where

$$\theta \in [0,\pi], \quad \phi \in [0,2\pi), \quad \psi \in [0,4\pi); \tag{111b}$$

we write (θ,ϕ,ψ) for the point so determined.[26] The isomorphism between SU(2) and S^3 can be given as

$$\begin{aligned} T:\ & SU(2) \xrightarrow{\approx} S^3,\\ & B(\alpha,\beta,\gamma) \mapsto (\beta,\alpha,\gamma), \end{aligned} \tag{112a}$$

or equivalently

$$\begin{pmatrix} a & b\\ -\bar{b} & \bar{a}\end{pmatrix} \mapsto \begin{pmatrix} a\\ -\bar{b}\end{pmatrix}. \tag{112b}$$

The metric on S^3 is now

$$\begin{aligned} ds^2 &:= du\, d\bar{u} + dv\, d\bar{v}\\ &\equiv \tfrac{1}{4}(d\theta^2 + d\phi^2 + d\psi^2 + 2\cos\theta\, d\phi\, d\psi), \end{aligned} \tag{113}$$

so that the Laplacian on S^3 is given by

$$\Box_3 = 4\left(\partial_\theta^2 + \cot\theta\,\partial_\theta + \frac{1}{\sin^2\theta}\partial_\phi^2 + \frac{1}{\sin^2\theta}\partial_\psi^2 - \frac{2\cos\theta}{\sin^2\theta}\partial_\psi\,\partial_\phi\right). \tag{114}$$

We consider S^2 to be the subspace of $\mathbb{R}^3$ defined by

$$S^2 := \left\{\begin{pmatrix} x\\ y\\ z\end{pmatrix} \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\right\} \tag{115}$$

and we choose the usual parametrization in terms of spherical coordinates

$$x + iy = e^{i\phi}\sin\theta, \quad z = \cos\theta, \tag{116}$$

where θ and ϕ have the same ranges as in (111b). Consider the elements $H(\lambda)$ of SU(2) defined by

$$H(\lambda) := \begin{pmatrix} e^{+i\lambda} & 0\\ 0 & e^{-i\lambda}\end{pmatrix} \equiv B(0,0,-2\lambda). \tag{117}$$

Then $\{H(\lambda) : \lambda \in [0,2\pi)\}$ is isomorphic to U(1) so we can define the Hopf bundle via

$$\begin{array}{c} U(1) \rightarrow S^3\\ \downarrow \pi\ ,\\ S^2 \end{array} \tag{118a}$$

where the projection is the obvious map

$$\pi(\theta,\phi,\psi) := (\theta,\phi) \tag{118b}$$

or equivalently

$$\begin{pmatrix} u\\ v\end{pmatrix} \mapsto \begin{pmatrix} \mathrm{Re}(-2\bar{u}v)\\ \mathrm{Im}(-2\bar{u}v)\\ u\bar{u} - v\bar{v}\end{pmatrix} \tag{118c}$$

and the circle action is defined by

$$\begin{aligned} (\theta,\phi,\psi)e^{i\lambda} &:= T(B(\theta,\phi,\psi)H(\lambda))\\ &\equiv (\theta,\phi,\psi - 2\lambda). \end{aligned} \tag{118d}$$

From (118c) we see that we get an induced map

$$\hat{\pi} : SU(2) \rightarrow SO(3) \tag{119a}$$

which can be defined by

$$\hat{\pi}(B)\pi\left(\begin{pmatrix} p\\ q\end{pmatrix}\right) \equiv \pi\left(B\begin{pmatrix} p\\ q\end{pmatrix}\right). \tag{119b}$$

Direct calculation shows that under this map we have

$$\hat{\pi}:\begin{pmatrix} a & b\\ -\bar{b} & \bar{a}\end{pmatrix}\mapsto\begin{pmatrix}\mathrm{Re}(a^2-b^2) & \mathrm{Im}(a^2+b^2) & \mathrm{Re}(2ab)\\ -\mathrm{Im}(a^2-b^2) & \mathrm{Re}(a^2+b^2) & -\mathrm{Im}(2ab)\\ -\mathrm{Re}(2\bar{a}b) & \mathrm{Im}(2\bar{a}b) & a\bar{a}-b\bar{b}\end{pmatrix}, \tag{119c}$$

so that

$$A(\alpha,\beta,\gamma):=\hat{\pi}(B(\alpha,\beta,\gamma))$$

$$\equiv\begin{pmatrix}\cos\alpha\cos\beta\cos\gamma-\sin\alpha\sin\gamma & -\cos\alpha\cos\beta\sin\gamma-\sin\alpha\cos\gamma & \cos\alpha\sin\beta\\ \sin\alpha\cos\beta\cos\gamma+\cos\alpha\sin\gamma & -\sin\alpha\cos\beta\sin\gamma+\cos\alpha\cos\gamma & \sin\alpha\sin\beta\\ -\sin\beta\cos\gamma & \sin\beta\sin\gamma & \cos\beta\end{pmatrix}. \tag{119d}$$

Noting that $A(\alpha,\beta,\gamma)\equiv A(\alpha,\beta,\gamma+2\pi)$ we can choose the parametrization of SO(3) to be given by (119d), where[26]

$$\beta\in[0,\pi],\quad \alpha\in[0,2\pi),\quad \gamma\in[0,2\pi). \tag{119e}$$

Define $\alpha_a(\tau)\in SO(3)$, $a=1,2,3$, to be the matrix which rotates S^2 about the ath axis counterclockwise through an angle τ. Thus

$$\alpha_1(\tau)=\begin{pmatrix}1&0&0\\0&\cos\tau&-\sin\tau\\0&\sin\tau&\cos\tau\end{pmatrix}\equiv A\left(\frac{3\pi}{2},\tau,\frac{\tau}{2}\right),$$

$$\alpha_2(\tau)=\begin{pmatrix}\cos\tau&0&\sin\tau\\0&1&0\\-\sin\tau&0&\cos\tau\end{pmatrix}\equiv A(0,\tau,0), \tag{120}$$

$$\alpha_3(\tau)=\begin{pmatrix}\cos\tau&-\sin\tau&0\\ \sin\tau&\cos\tau&0\\0&0&1\end{pmatrix}\equiv A(0,0,\tau).$$

Note that

$$A(\alpha,\beta,\gamma)\equiv\alpha_3(\alpha)\alpha_2(\beta)\alpha_3(\gamma), \tag{121}$$

so that the parametrization (119d) of SO(3) is just the usual one in terms of Euler angles: $A(\alpha,\beta,\gamma)$ is the element of SO(3) that rotates the sphere S^2 first by γ around the z axis, then by β around the (original) y axis, and finally by α around the (original) z axis. We wish to find matrices $\beta_a\in SU(2)$, $a=1,2,3$, satisfying

$$\hat{\pi}(\beta_a(\tau))\equiv\alpha_a(\tau). \tag{122a}$$

Although $\hat{\pi}$ is a *two*-to-one mapping, the additional requirement that

$$\beta_a(0)\equiv I \tag{122b}$$

determines the $\beta_a(\tau)$ uniquely as

$$\beta_1(\tau)=B\left(\frac{3\pi}{2},\tau,\frac{5\pi}{2}\right)\equiv\begin{pmatrix}\cos(\tau/2)& i\sin(\tau/2)\\ i\sin(\tau/2)&\cos(\tau/2)\end{pmatrix},$$

$$\beta_2(\tau)=B(0,\tau,0)\equiv\begin{pmatrix}\cos(\tau/2)&\sin(\tau/2)\\-\sin(\tau/2)&\cos(\tau/2)\end{pmatrix}, \tag{123}$$

$$\beta_3(\tau)=B(0,0,\tau)\equiv\begin{pmatrix}e^{-i\tau/2}&0\\0&e^{+i\tau/2}\end{pmatrix}.$$

From (109) we see that the general element of SU(2) can be written

$$B(\alpha,\beta,\gamma)\equiv\beta_3(\alpha)\,\beta_2(\beta)\,\beta_3(\gamma). \tag{124}$$

Note that (121) and (124) imply

$$A(\alpha,\beta,\gamma)^{-1}\equiv A(-\gamma,-\beta,-\alpha),$$
$$B(\alpha,\beta,\gamma)^{-1}\equiv B(-\gamma,-\beta,-\alpha). \tag{125}$$

Define

$$\Lambda_a:=\left.\frac{d}{d\tau}\right|_{\tau=0}\beta_a(\tau), \tag{126a}$$

$$\hat{\Lambda}_a:=\left.\frac{d}{d\tau}\right|_{\tau=0}\alpha_a(\tau) \tag{126b}$$

and notice that both Λ_a and $\hat{\Lambda}_a$ satisfy the commutation relations (11).[27] Note that (13) and (14) are also satisfied.

We now intoduce angular mometum operators. Using the chain rule the definition (18) of angular momentum on S^2 is equivalent to

$$\hat{J}_a g=+i(x\quad y\quad z)\hat{\Lambda}_a\begin{pmatrix}\partial_x\\ \partial_y\\ \partial_z\end{pmatrix}g, \tag{127}$$

which yields the familiar result

$$\hat{J}_3=-i\,\partial_\phi,$$
$$\hat{J}_\pm=e^{\pm i\phi}(\pm\partial_\theta+i\cot\theta\,\partial_\phi). \tag{128}$$

Similarly, the definition (15) of angular momentum on S^3 is equivalent to

$$J_a f=+i(u\quad v)\bar{\Lambda}_a\begin{pmatrix}\partial_u\\ \partial_v\end{pmatrix}f+i(\bar{u}\quad\bar{v})\Lambda_a\begin{pmatrix}\partial_{\bar{u}}\\ \partial_{\bar{v}}\end{pmatrix}f, \tag{129}$$

which yields

$$J_3=-i\,\partial_\phi,$$
$$J_\pm=e^{\pm i\phi}(\pm\partial_\theta+i\cot\theta\,\partial_\phi-(i/\sin\theta)\partial_\psi) \tag{130}$$

so that

$$J^2:=J_3^2-J_3+J_+J_-\equiv-\tfrac{1}{4}\Box_3. \tag{131}$$

Both of these operators satisfy the standard commutation relations (16), e.g.,

$$[J_3,J_\pm]=\pm J_\pm,\quad [J_+,J_-]=2J_3 \tag{132}$$

(all others zero).

We now introduce the complex line bundles E_n over S^2 associated with the Hopf bundle (118). The points of E_n are equivalence classes

$$[(\theta,\phi,\psi;z)]\in S^3\times_n\mathbf{C} \tag{133a}$$

under the equivalence relation

$$(\theta,\phi,\psi;z) \sim (\theta,\phi,\psi - 2\lambda;e^{in\lambda}z). \tag{133b}$$

We thus obtain the line bundle

$$\begin{array}{c} E_n \\ \downarrow \pi_n , \\ S^2 \end{array} \tag{134a}$$

where the projection π_n is given by

$$\pi_n([(\theta,\phi,\psi;z)]) := (\theta,\phi). \tag{134b}$$

There is a one-to-one correspondence Q between (smooth) functions on S^3 satisfying

$$f(\theta,\phi,\psi) \equiv e^{-in\psi/2}F(\theta,\phi) \tag{135}$$

and (smooth) sections

$$\sigma: S^2 \to E_n \tag{136}$$

of the line bundles E_n, which is given by

$$Q(f)(\theta,\phi) := [(\theta,\gamma,\psi;f(\theta,\phi,\psi))]. \tag{137}$$

We will use the notation

$$Q(f) =: \sigma_f, \quad Q^{-1}(\sigma) =: f_\sigma. \tag{138}$$

Given any (smooth) local section[28]

$$\begin{aligned} &U_A \subset S^2, \\ &\hat{\gamma}_A : U_A \to S^3, \\ &\quad (\theta,\phi) \mapsto (\theta,\phi,\gamma_A(\theta,\phi)) \end{aligned} \tag{139}$$

of the Hopf bundle we can interpret any section (136) of E_n as a *function* on U_A via

$$\begin{aligned} g_A^\sigma &:= f_\sigma \circ \hat{\gamma}_A \\ &\equiv e^{-i(n/2)\gamma_A}F_\sigma, \end{aligned} \tag{140}$$

where F_σ is defined from f_σ as in (135). There is thus a one-to-one correspondence Q_A between (smooth) functions on U_A and (smooth) sections of E_n (restricted to U_A), which is given by

$$Q_A^{-1}(\sigma) := Q^{-1}(\sigma) \circ \hat{\gamma}_A \equiv g_A^\sigma, \tag{141}$$

i.e.,

$$Q_A(F) = [(\theta,\phi,\psi;e^{-i(n/2)(\psi-\gamma_A)}F)]. \tag{142}$$

There is a natural norm on the space of sections (136) of E_n given by

$$\langle \sigma,\tau \rangle := \oint_{S^2} \bar{F}_\sigma F_\tau \, dS. \tag{143}$$

Note that for σ,τ both sections of E_n we have [compare (37)]

$$\begin{aligned} \langle f_\sigma, f_\tau \rangle &= \int_{S^3} \bar{f}_\sigma f_\tau \, dS \\ &\equiv \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \int_{\psi=0}^{4\pi} \bar{F}_\sigma F_\tau \frac{1}{8} \sin\theta \, d\theta \, d\phi \, d\psi \\ &\equiv (\pi/2)\langle \sigma,\tau \rangle. \end{aligned} \tag{144}$$

Given any operator Z on the space Γ_n of square-integrable sections of E_n we can obtain an operator on $L^2(U_A)$ defined by

$$Z^A := Q_A^{-1} \circ Z \circ Q_A \tag{145}$$

and this is clearly a one-to-one mapping of the corresponding operator spaces. Here Z^A will be referred to as the operator a *with respect to* (the section of the Hopf bundle) $\hat{\gamma}_A$.

We can introduce angular momentum operators $\hat{L}_a$ on Γ_n via

$$\hat{L}_a \sigma := Q(J_a(Q^{-1}\sigma)) \tag{146}$$

and thus obtain the operators[29]

$$\begin{aligned} L_3^A &= -i\partial_\phi + (n/2)\,\partial_\phi\gamma_A, \\ L_\pm^A &= e^{\pm i\phi}(\pm\partial_\theta + i\cot\theta\,\partial_\phi \\ &\quad - (n/2)((1/\sin\theta) \mp i\,\partial_\theta\gamma_A + \cot\theta\,\partial_\phi\gamma_A)), \end{aligned} \tag{147}$$

$$(L^2)^A = -\Box_2^A + \frac{n\cos\theta}{\sin^2\theta} L_3^A + \frac{n^2}{4\sin^2\theta},$$

on $L^2(U_A)$, where

$$\begin{aligned} \Box_2^A &:= \Box_2 + i(n/2)(\Box_2\gamma_A) \\ &\quad + in[(\partial_\theta\gamma_A)\,\partial_\theta + (\partial_\phi\gamma_A)\partial_\phi/\sin^2\theta] \\ &\quad - (n^2/4)[(\partial_\theta\gamma_A)^2 + (\partial_\phi\gamma_A)^2/\sin^2\theta] \end{aligned} \tag{148}$$

is the operator obtained from $\Box_2$ by the substitutions

$$\partial_\theta \mapsto \partial_\theta + i(n/2)(\partial_\theta\gamma_A), \tag{149}$$

$$\partial_\phi \mapsto \partial_\phi + i(n/2)(\partial_\phi\gamma_A), \tag{150}$$

where $\Box_2$ denotes the Laplacian on S^2.

The natural connection on E_n is given by[10,17]

$$\tilde{d}\sigma \equiv Q(e^{-i(n/2)\psi}\,dF_\sigma + i(n/2)\,f_\sigma \cos\theta\,d\phi), \tag{151}$$

where d denotes the exterior derivative on S^2. The connection one-form ω_n^A of the bundle E_n with respect to $\hat{\gamma}_A$ is defined by

$$\tilde{d}[(\theta,\phi,\gamma_A;1)] =: [(\theta,\phi,\gamma_A;\omega_n^A]. \tag{152}$$

But

$$Q^{-1}([(\theta,\phi,\gamma_A;1)]) \equiv e^{-i(n/2)(\gamma_A-\psi)} \tag{153}$$

so that

$$\omega_n^A \equiv i(n/2)(\cos\theta\,d\phi + d\gamma_A). \tag{154}$$

The curvature E_n is thus

$$\begin{aligned} R_n &= d\omega_n^A \\ &\equiv -i(n/2)\sin\theta\,d\theta \wedge d\phi \\ &\equiv -i(n/2)\Omega \end{aligned} \tag{155}$$

as desired [compare (26)].

The Laplacian Δ_n on E_n associated with the connection (151) can now be defined as follows:

$$\Delta_n := Q_A \circ \Delta_n^A \circ Q_A^{-1}, \tag{156a}$$

where[17]

$$\begin{aligned} \Delta_n^A &:= \Box_2 + 2g^{ab}(\omega_n^A)_a\nabla_b + g^{ab}(\nabla_a(\omega_n^A)_b \\ &\quad + (\omega_n^A)_a(\omega_n^A)_b), \end{aligned} \tag{156b}$$

where g_{ab} is the standard metric on S^2 and ∇_a denotes covariant differentiation on S^2 (so that $\Box_2 \equiv g^{ab}\nabla_a\nabla_b$). Inserting our choice (154) for the connection ω_n^A in (156) we obtain

$$\Delta_n^A = +\Box_2^A - \frac{n\cos\theta}{\sin^2\theta} L_3^A - \frac{n^2}{4}\frac{\cos^2\theta}{\sin^2\theta}, \tag{157a}$$

so that

$$\Delta_n \equiv -\hat{L}^2 + n^2/4. \tag{157b}$$

VII. WIGNER $\mathscr{D}$ FUNCTIONS

We now introduce the Wigner $\mathscr{D}$ functions[1–4] as the matrix elements of finite rotations acting on irreducible representations of the group SU(2).

We introduce irreducible representations of SU(2) on S^3 as follows. For $\binom{u}{v} \in S^3$ define

$$\phi_{lm} := \frac{u^{l-m} v^{l+m}}{\sqrt{(l-m)!(l+m)!}} \quad (0 \leqslant 2l \in \mathbb{Z},\ 0 \leqslant l - |m| \in \mathbb{Z}). \tag{158}$$

It is easy to check that $\phi_{lm} \in L^2(S^3)$. The operators J_a [(129) and (130)] when acting on ϕ_{lm} take the form

$$J_3 = \tfrac{1}{2}(v\,\partial_v - u\,\partial_u), \quad J_+ = v\,\partial_u, \quad J_- = u\,\partial_v, \tag{159}$$

and direct calculation shows that Eqs. (43)–(46) are satisfied so that there is an irreducible matrix representation of SU(2) on the vector spaces W^l defined in (47). We define the Wigner $\mathscr{D}$ functions by (3) and (48) and write

$$\mathscr{D}^l_{qm}(\alpha,\beta,\gamma) := \mathscr{D}^l_{qm}(B(\alpha,\beta,\gamma)). \tag{160}$$

Then the $\mathscr{D}^l_{qm}(\alpha,\beta,\gamma)$ of course satisfy properties (52) through (55), in particular,

$$\overline{\mathscr{D}^l_{qm}(\alpha,\beta,\gamma)} \equiv \mathscr{D}^l_{mq}(-\gamma,-\beta,-\alpha), \tag{161}$$

where we have used (125).

We now turn to the angular momentum operators L_a and K_a defined in (56). From (66) and the expressions (130) for J_a it is clear that

$$L_3 = -i\,\partial_\alpha,$$
$$L_\pm = e^{\pm i\alpha}(\pm\partial_\beta + i\cot\beta\,\partial_\alpha - (i/\sin\beta)\,\partial_\gamma). \tag{162}$$

We derive expressions for the K_a as differential operators by noticing that, using the chain rule, definition (56b) is equivalent to

$$K_a h \equiv +i(a\quad b)\Lambda_a \begin{pmatrix}\partial_a\\ \partial_b\end{pmatrix} h + i(\bar a\quad \bar b)\overline{\Lambda}_a \begin{pmatrix}\partial_{\bar a}\\ \partial_{\bar b}\end{pmatrix} h, \tag{163}$$

which yields

$$K_3 = +i\,\partial_\gamma,$$
$$K_\pm = -e^{\mp i\gamma}(\pm\partial_\beta + (i/\sin\beta)\,\partial_\alpha - i\cot\beta\,\partial_\gamma). \tag{164}$$

Using (59)–(61) and (63) one can show that[30]

$$\mathscr{D}^l_{qm}(\alpha,\beta,\gamma) = \left[\frac{(l+m)!(l-m)!}{(l+q)!(l-q)!}\right]^{1/2} \left(\sin\frac{\beta}{2}\right)^{2l} \times \sum_{n=n_{\min}}^{n_{\max}} \binom{l+q}{n}\binom{l-q}{n-q-m}(-1)^{l+m-n} \times e^{-iq\alpha}(\cot(\beta/2))^{2n-m-q} e^{-im\gamma}, \tag{165a}$$

where

$$n_{\min} = \max(0, m+q), \quad n_{\max} = \min(l+q, l+m). \tag{165b}$$

Defining $\hat{\mathscr{Y}}^l_{qm}$ via (67) and (68) we thus obtain

$$\hat{\mathscr{Y}}^l_{qm}(\theta,\phi,\psi) = \left[\frac{(l+q)!(l-q)!(2l+1)}{(l+m)!(l-m)!(4\pi)}\right]^{1/2} \left(\sin\frac{\theta}{2}\right)^{2l} \times \sum_{k=k_{\min}}^{k_{\max}} \binom{l+m}{k}\binom{l-m}{k+q-m}(-1)^{l-q-k} \times e^{+im\phi}(\cot(\theta/2))^{2k+q-m} e^{-iq\psi}, \tag{166a}$$

where

$$k_{\min} = \max(0, m-q), \quad k_{\max} = \min(l+m, l-q). \tag{166b}$$

The properties of the $\mathscr{Y}^l_{qm}$ are easily obtained from (71)–(73) using the isomorphism (64). Define the operator ð ("edth") on $L^2(S^3)$ by

$$\eth f := -K_+(f\circ T)\circ T^{-1} \tag{167a}$$

[compare (66); the minus sign is conventional] so that

$$\bar\eth f \equiv K_-(f\circ T)\circ T^{-1}. \tag{167b}$$

Explicitly, we have

$$\eth = e^{-i\psi}(\partial_\theta + (i/\sin\theta)\partial_\phi - i\cot\theta\,\partial_\psi), \tag{168}$$
$$\bar\eth = e^{+i\psi}(\partial_\theta - (i/\sin\theta)\partial_\phi + i\cot\theta\,\partial_\psi)$$

and

$$[\eth,\bar\eth] = -2i\,\partial_\psi. \tag{169}$$

Then we have

$$J_3\hat{\mathscr{Y}}^l_{qm} = m\hat{\mathscr{Y}}^l_{qm},$$
$$J_\pm\hat{\mathscr{Y}}^l_{qm} = [(l\mp m)(l\pm m+1)]^{1/2}\hat{\mathscr{Y}}^l_{q,m\pm1}; \tag{170}$$

$$i\,\partial_\psi\hat{\mathscr{Y}}^l_{qm} = q\hat{\mathscr{Y}}^l_{qm},$$
$$\eth\hat{\mathscr{Y}}^l_{qm} = [(l-q)(l+q+1)]^{1/2}\hat{\mathscr{Y}}^l_{q+1,m}, \tag{171}$$
$$\bar\eth\hat{\mathscr{Y}}^l_{qm} = -[(l+q)(l-q+1)]^{1/2}\hat{\mathscr{Y}}^l_{q-1,m},$$

and

$$\langle\hat{\mathscr{Y}}^l_{qm}, \hat{\mathscr{Y}}^{l'}_{q'm'}\rangle = (\pi/2)\delta_{ll'}\delta_{qq'}\delta_{mm'}, \tag{172}$$

so that $\{\sqrt{2/\pi}\,\hat{\mathscr{Y}}^l_{qm}\}$ forms an orthonormal basis for $L^2(S^3)$.

Comparing (128), (130), (170), and (172) we see that we can *define* the usual spherical harmonics on S^2 to be [compare (49)]

$$Y_{lm}(\theta,\phi) := \hat{\mathscr{Y}}^l_{0m}(\theta,\phi,0) \quad (l \in \mathbb{Z},\ 0 \leqslant l - |m| \in \mathbb{Z}). \tag{173}$$

VIII. SPIN-WEIGHTED SPHERICAL HARMONICS AND MONOPOLE HARMONICS

Having defined the functions $\hat{\mathscr{Y}}^l_{qm} \in L^2(S^3)$ in terms of Wigner $\mathscr{D}$ functions we now show how to obtain the usual coordinate definitions of both spin-weighted spherical harmonics and monopole harmonics. We first note that by comparing (166) and (135) we see that we can define a section $\mathscr{Y}_{qlm}$ of the line bundle E_{2q} by (105) so that [compare (137)]

$$\mathscr{Y}_{qlm}(\theta,\phi) := [(\theta,\phi,\psi; \hat{\mathscr{Y}}^l_{qm}(\theta,\phi,\psi))]. \tag{174}$$

The $\mathscr{Y}_{qlm}$ of course satisfy (106).

Given a local section of the Hopf bundle defined by the function $\gamma_A \in L^2(U_A)$ and (139), we thus obtain the functions [compare (141)]

$$\begin{aligned}\mathcal{Y}^A_{qlm}(\theta,\phi) &:= Q_A^{-1}(\mathcal{Y}_{qlm})\\ &\equiv \hat{\mathcal{Y}}^l_{qm}(\theta,\phi,\gamma_A(\theta,\phi))\\ &\equiv e^{-iq\gamma_A(\theta,\phi)}\hat{\mathcal{Y}}^l_{qm}(\theta,\phi,0)\in \mathbf{L}^2(U_A).\end{aligned} \tag{175}$$

The properties of the $\mathcal{Y}^A_{qlm}$ are completely analogous to those of the $\hat{\mathcal{Y}}^A_{qlm}$, i.e., (170)–(172). Before giving them explicitly, however, we need to introduce an operator on sections of E_n analogous to ð. We do this by defining [compare (146)]

$$\begin{aligned}\hat{ð}:\ &\Gamma_n \to \Gamma_{n+2},\\ &\hat{ð}\sigma := -Q(K_+(Q^{-1}\sigma))\end{aligned} \tag{176a}$$

and

$$\begin{aligned}\hat{\bar{ð}}:\ &\Gamma_n \to \Gamma_{n-2},\\ &\hat{\bar{ð}}\sigma := +Q(K_-(Q^{-1}\sigma)),\end{aligned} \tag{176b}$$

so that [see (145)][31]

$$ð^A = e^{-i\gamma_A}\Big[\partial_\theta + \frac{i}{\sin\theta}\partial_\phi + \frac{n}{2}\Big(-\cot\theta + i\,\partial_\theta\gamma_A - \frac{1}{\sin\theta}\partial_\phi\gamma_A\Big)\Big], \tag{177a}$$

$$\bar{ð}^A = e^{+i\gamma_A}\Big[\partial_\theta - \frac{i}{\sin\theta}\partial_\phi - \frac{n}{2}\Big(-\cot\theta - i\,\partial_\theta\gamma_A - \frac{1}{\sin\theta}\partial_\phi\gamma_A\Big)\Big]. \tag{177b}$$

The vector m of Sec. IV is given by[11]

$$m = e^{-i\gamma_A}(\partial_\theta + (i/\sin\theta)\partial_\phi), \tag{177c}$$

where the choice of the function $\gamma_A(\theta,\phi)$ is referred to as the choice of a *spin gauge*. Note that since

$$\sigma\in\Gamma_n \Rightarrow \bar{\sigma}\in\Gamma_{-n}, \tag{178a}$$

we have

$$\overline{\hat{ð}\sigma} \equiv \hat{\bar{ð}}\bar{\sigma}. \tag{178b}$$

The properties of the $\mathcal{Y}^A_{qlm}$ are thus [compare (169)–(171)]

$$\begin{aligned}&L_3^A\mathcal{Y}^A_{qlm} = m\mathcal{Y}^A_{qlm},\\ &L_\pm^A\,\mathcal{Y}^A_{qlm} = [(l\mp m)(l\pm m+1)]^{1/2}\mathcal{Y}_{q,l,m\pm 1};\end{aligned} \tag{179}$$

$$\begin{aligned}&ð^A\mathcal{Y}^A_{qlm} = [(l-q)(l+q+1)]^{1/2}\mathcal{Y}^A_{q+1,l,m},\\ &\bar{ð}^A\mathcal{Y}^A_{qlm} = -[(l+q)(l-q+1)]^{1/2}\mathcal{Y}^A_{q-1,l,m},\\ &[ð^A,\bar{ð}^A]\mathcal{Y}^A_{qlm} = -2q\mathcal{Y}^A_{qlm}.\end{aligned} \tag{180}$$

Furthermore, if γ_A is chosen so that its domain of definition U_A is dense in $\mathbf{S}^2$, then (172) becomes

$$\oint_{\mathbf{S}^2}\overline{\mathcal{Y}^A_{qlm}(\theta,\phi)}\mathcal{Y}^A_{q'l'm'}(\theta,\phi) = \delta_{ll'}\delta_{qq'}\delta_{mm'}. \tag{181}$$

Finally, note that for *integer* spin ($l\in\mathbf{Z}$), (173) and (175) imply

$$\mathcal{Y}^A_{0lm}(\theta,\phi) \equiv Y_{lm}(\theta,\phi) \tag{182}$$

and that from (180) we now obtain [compare (97)]

$$\mathcal{Y}^A_{qlm}(\theta,\phi) = \begin{cases}\Big[\dfrac{(l-q)!}{(l+q)!}\Big]^{1/2}(ð^A)^q Y_{lm}(\theta,\phi) & (0<q<l),\\ \Big[\dfrac{(l+q)!}{(l-q)!}\Big]^{1/2}(-1)^q(\bar{ð}^A)^{|q|}Y_{lm}(\theta,\phi) & (-l<q<0),\\ 0 & (l<|q|),\end{cases} \tag{183}$$

where defined [i.e., for $(\theta,\phi)\in U_A$]. But the standard coordinate definition of the spin-weighted spherical harmonics for *integer* spin is just[6,11] (183) in the "standard" spin gauge[32]

$$\gamma_0(\theta,\phi) := 0, \tag{184a}$$

so that[33]

$$\begin{aligned}{}_qY_{lm} &:= \mathcal{Y}^0_{qlm}(\theta,\phi) = \hat{\mathcal{Y}}^l_{qm}(\theta,\phi,0)\\ &\equiv \sqrt{\frac{2l+1}{4\pi}}\ \overline{\mathcal{D}^l_{m,-q}(\theta,\phi,0)}.\end{aligned} \tag{184b}$$

In an arbitrary "spin gauge," (175) implies[34]

$$\begin{aligned}\mathcal{Y}^A_{qlm}(\theta,\phi) &\equiv \sqrt{\frac{2l+1}{4\pi}}\ \overline{\mathcal{D}^l_{m,-q}(\theta,\phi,\gamma_A(\theta,\phi))}\\ &\equiv e^{-iq\gamma_A(\theta,\phi)}\mathcal{Y}^0_{qlm}(\theta,\phi).\end{aligned} \tag{185}$$

One normally *defines* the spin-weighted spherical harmonics in standard gauge for half-integer spin by (184). Thus, the standard spin-weighted spherical harmonics are just the $\mathcal{Y}^A_{qlm}$ in a particular (dense) trivialization of the line bundles E_{2q} [namely the one induced by (184a)].

The monopole harmonics Y_{qlm} are even easier. They are defined by[8,11] (179) and (181) (and a choice of phase for each l) in the gauges[32]

$$\gamma_a(\theta,\phi) := -\phi,\quad \gamma_b(\theta,\phi) := +\phi. \tag{186a}$$

With an appropriate choice of phase we have[35]

$$Y^a_{qlm} \equiv \mathcal{Y}^a_{qlm},\quad Y^b_{qlm} \equiv \mathcal{Y}^b_{qlm}, \tag{186b}$$

so that the monopole harmonics Y_{qlm} of Ref. 8 are just the $\mathcal{Y}_{qlm}$ in a particular trivializing cover of E_{2q} [namely the one defined by (186) (Ref. 32)].

ACKNOWLEDGMENT

I am deeply indebted to Malcolm Adams for numerous discussions and several critical readings of the manuscript (especially Part I), all of which served to remind me that physics is not necessarily mathematics.

[1]E. P. Wigner, *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren* (Vieweg, Braunschweig, 1931). A revised version of this book was later published in English: E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).

[2]A. R. Edmonds, *Angular Momentum in Quantum Mechanics*, (Princeton U. P., Princeton, NJ, 1957).

[3]M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957).

[4]L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics* (Addison–Wesley, Reading, MA, 1981). (See also Ref. 9.)

[5]E. T. Newman and R. Penrose, J. Math. Phys. **7**, 863 (1966).

[6]J. N. Goldberg, A. J. MacFarlane, E. T. Newman, F. Rohrlich, and E. C. G. Sudarshan, J. Math. Phys. **8**, 2155 (1967).

[7]R. Penrose and W. Rindler, *Spinors and Space-Time* (Cambridge U. P., Cambridge, 1984).

[8]T. T. Wu and C. N. Yang, Nucl. Phys. B **107**, 365 (1976).

[9]L. C. Biedenharn and J. D. Louck, *The Racah–Wigner Algebra in Quantum Theory* (Addison–Wesley, Reading, MA, 1981). (See also Ref. 4.)

[10]W. Greub and H.-R. Petry, J. Math. Phys. **16**, 1347 (1975).

[11]T. Dray, J. Math. Phys. **26**, 1030 (1985).

[12]Peter Batenburg, Doctoraalscriptie (in English), University of Utrecht, 1984 (unpublished).

[13]The interpretation of spin-weighted *functions* (95) as sections of complex line bundles was given in Ref. 22. The standard definition of spin-weighted spherical harmonics (for integrer spin) is (183), which is given in terms of the differential operator ð. The interpretation of ð as an operator on sections of line bundles was given in Ref. 23. However, Ref. 22 does not discuss spin-weighted spherical *harmonics* at all, and although Ref. 23 does give a precise definition of them it does not discuss them in any detail. The author wishes to thank Ted Newman for providing these two references.

[14]This is to be contrasted with the "standard spin gauge" for spin-weighted spherical harmonics [see Ref. 11 and (184a) below], which amounts to giving only a *local* trivialization of the complex line bundles (which of course does *not* cover S^2 but only a dense subspace of S^2).

[15]The standard procedure (see, e.g., Ref. 9) for *half-integer* spin is to exponentiate the angular momentum operators.

[16]A previous effort along these lines, namely Ref. 6, unfortunately uses an internally inconsistent choice of conventions.

[17]R. Kuwabara, J. Math. Tokushima Univ. **16**, 1 (1982).

[18]V. Guillemin and A. Uribe, "Clustering theorems with twisted spectra," Princeton Univ. preprint, 1985.

[19]See, e.g. Ref. 10. Equation (40) can be thought of as the *definition* of this preferred connection, which will be given in coordinates in Sec. VI.

[20]As *Riemannian manifolds* the spaces E_n $(n \neq 0)$ can be thought of as the lens spaces $S^3/Z_{|n|}$. In particular, $E_{\pm n}$ are isomorphic and the Hopf bundle thought of in this way is isomorphic to S^3.

[21]F. Peter and H. Weyl, Math. Ann. **97**, 737 (1927); J. F. Adams, *Lectures on Lie Groups* (Benjamin, New York, 1969).

[22]W. D. Curtis and D. E. Lerner, J. Math. Phys. **19**, 874 (1978).

[23]M. Eastwood and P. Tod, Math. Proc. Cambridge Philos. Soc. **92**, 317 (1982).

[24]We could just as well have used $\overline{V}_0$ in constructing the bundle (90). Only one of these bundles is *strong* bundle isomorphic to the Hopf bundle (7) but this does not affect the argument leading up to (92).

[25]The minus sign comes about because one usually writes the momentum as $p - eA$, i.e., as the operator $-i(\nabla - ieA)$, so that the connection is $-ieA$ and the curvature is $-ie\,dA \equiv -ieF$.

[26]Note that the coordinates (α,β,γ) and (θ,ϕ,ψ) are not well defined at the poles $\beta = 0, \pi$ and $\theta = 0,\pi$, respectively.

[27]This follows immediately from the definition of $\alpha_a(\tau)$ as the usual rotation matrices.

[28]Note that $\hat{\gamma}_A$ is a *section*, whereas γ_A is a *function*.

[29]These agree with (24) of Ref. 11 with $\gamma_A \equiv -\gamma$ and $n \equiv 2s$. We have omitted the hats for simplicity.

[30]This agrees with (4.1.12) and (4.1.15) of Ref. 2 if we note that Edmonds defines

$$
\begin{aligned}
{}_E\mathscr{D}^l_{qm}(\alpha,\beta,\gamma) &:= \mathscr{D}^l_{qm}(\beta_3^{-1}(\alpha)\,\beta_2^{-1}(\beta)\,\beta_3^{-1}(\gamma)) \\
&\equiv \mathscr{D}^l_{qm}(B(\gamma,\beta,\alpha)^{-1}) \\
&\equiv \mathscr{D}^l_{qm}(-\alpha,-\beta,-\gamma) \\
&\equiv \overline{\mathscr{D}^l_{qm}(\alpha,-\beta,\gamma)}.
\end{aligned}
$$

However, if we interpret (3.4) of Ref. 6 (see also our Ref. 16) as defining

$$
\begin{aligned}
{}_G\mathscr{D}^l_{qm}(\alpha,\beta,\gamma) &:= \mathscr{D}^l_{qm}(B(\alpha,\beta,\gamma)^{-1}) \\
&\equiv \mathscr{D}^l_{qm}(-\gamma,-\beta,-\alpha) \\
&\equiv \overline{\mathscr{D}^l_{mq}(\alpha,\beta,\gamma)} \\
&\equiv {}_E\mathscr{D}_{mq}(\alpha,-\beta,\gamma),
\end{aligned}
$$

then we are forced to conclude that (3.9) of Ref. 6 is missing a factor $(-1)^{m+m'}$. Finally, note that although no explicit expression analogous to (165) is given in Ref. 4 the functions $D_{lqm}(\alpha,\beta,\gamma)$ defined there are identical to our functions $\mathscr{D}^l_{qm}(\alpha,\beta,\gamma)$.

[31]This agrees with (19) of Ref. 11, where $\gamma_A \equiv -\gamma$ and $n \equiv +2s$. We have omitted the hats for simplicity.

[32]Note that the section of the Hopf bundle (139) induced by γ_0 is defined everywhere on S^2 except at the poles $\theta = 0,\pi$, whereas the sections induced by γ_a,γ_b are each defined everywhere on S^2 except at *one* pole (namely $\theta = 0$ and $\theta = \pi$, respectively).

[33]This agrees with the ${}_qY_{lm}$ of Ref. 11 in the standard spin gauge but differs from Ref. 6 by a factor $(-1)^q$.

[34]The factor $(-1)^s$ in (28) of Ref. 11 is thus incorrect.

[35]This agrees with both Ref. 8 and Ref. 11.

Screen observables in relativistic and nonrelativistic quantum mechanics

R. Werner
Fachbereich Physik, Universität Osnabrück, Postfach 4469, D-4500 Osnabrück, Federal Republic of Germany

(Received 3 April 1985; accepted for publication 22 October 1985)

Screen observables, which measure the arrival coordinates of free particles at a hyperplane containing a timelike direction, are defined by a covariance property with respect to an irreducible representation of the Poincaré or Galilei group. For each representation with $m > 0$ the set of screen observables is constructed explicitly and a unique ideal screen observable of greatest intrinsic accuracy is singled out.

I. INTRODUCTION

It is well known that the classical concept of particle trajectories has no proper analog in quantum mechanics. Only the position of a particle at a single time is usually considered as a quantum mechanical observable. This limitation seems strange in view of the standard laboratory techniques of particle detection; whereas the measurement of "position at time t" would somehow require the instantaneous construction of a trap for the particle, measuring devices like counters, photographic plates, or scintillating screens are usually made sensitive for a long time interval during which the arrival of the particle is expected. The arrival time itself, or the time at which the detector responds, is often considered an important part of the information. The purpose of the present paper is to show how such measurements may be given an idealized description within the quantum mechanical formalism.

Observables describing the joint measurement of arrival time and arrival location of a particle at a screen will be characterized by a covariance condition with respect to a given irreducible representation of the group of (relativistic or nonrelativistic) space-time symmetries. Further conditions concerning the intrinsic "accuracy" of the measurement are then used to single out an essentially unique "ideal screen observable" for each representation. These conditions are in close analogy to Wightman's characterization of the Newton–Wigner position observable.[1] However, Wightman's assumption that the observable should be projection valued (rather than a general positive operator-valued measure[2]) turns out to be too restrictive in the case of screen observables and many similar problems, e.g., observables for the orientation of a rotator, phase-space variables, or the position of a photon. Therefore, this assumption was replaced by the minimality of a certain quadratic form, describing an "uncertainty" intrinsic to the measurement. This "variance form" (defined for general observables over $\mathbf{R}^n$, see Sec. III) vanishes for projection-valued observables, but not only for these. Screen observables with vanishing variance form exist in the nonrelativistic and spinless relativistic cases. For higher spin the variance form is necessarily positive, since in this case the unbounded operators describing the expectation values of arrival time and location fail to commute.

The general notion of covariant observables has been studied by several authors.[2,3] The basic structural result, a covariant version of the Naimark dilation theorem,[4,5] is presented in Sec. II in a slightly extended version and leads to a general construction procedure for covariant observables. To the author's knowledge this procedure has not been applied systematically to the case of screen observables. However, some special cases have been obtained by Kijowski[6] and Ludwig.[7]

II. COVARIANT OBSERVABLES

An observable is the theoretical description of a measuring device. The possible outcomes of individual measurements form a set X and to each measurable subset $\sigma \subset X$ an observable F over X for systems described in the Hilbert space $\mathscr{H}$ associates an operator $F(\sigma) \in \mathscr{B}(\mathscr{H})$ with $0 \leqslant F(\sigma) \leqslant 1$. The quantity tr $WF(\sigma)$ is to be interpreted as the probability that measurement on systems prepared according to W ($W \geqslant 0$, tr $W = 1$) yields an outcome $x \in \sigma$. Obviously F must be a positive operator-valued measure on X. For technical reasons it is useful to consider not only the measure F but also its integrals over a suitable class of functions. More precisely we define for a Hilbert space $\mathscr{H}$ and a locally compact space X, with $\mathscr{C}_0(X)$ the space of continuous complex functions on X, vanishing at infinity: An *observable* F over X in $\mathscr{H}$ is a linear map $F: \mathscr{C}_0(X) \to \mathscr{B}(\mathscr{H})$ such that $f \geqslant 0 \Rightarrow F(f) \geqslant 0$ and $\|F(f)\| \leqslant \|f\|$.

An observable F uniquely defines an operator-valued Baire measure, which we shall denote by the same letter, so that $F(f) = \int_X f(x) F(dx)$. The operator $F(X) := \sup\{F(f) \mid f \leqslant 1\}$ determines the probability that the apparatus measuring F responds at all. An observable is called *normalized* if $F(X) = 1$. [Every observable F can be considered to be normalized over the one-point compactification $X \cup \{\infty\}$, when the measure of the "no event" result "∞" is defined by $F(\{\infty\}) = 1 - F(X)$.]

The set of observables over X is convex, and compact in the initial topology induced by the functionals $F \to \langle \varphi, F(f) \psi \rangle$ for $\varphi, \psi \in \mathscr{H}$ and $f \in \mathscr{C}_0(X)$. The subset of normalized observables is in general not closed in this topology. A *decision observable*[8] is a normalized observable F such that $F: \mathscr{C}_0(X) \to \mathscr{B}(\mathscr{H})$ is a *-algebraic homomophism or, equivalently, the associated measure is projection valued with $F(X) = 1$. In many textbooks, beginning with von Neumann's, only this restricted class is used. However, since the aim of this paper cannot be achieved within this class, we have to make use of the more general concept above, which

 0022-2488/86/030793-11$02.50

was introduced by Ludwig, Davies, and others.

Now let G be some physical symmetry group acting on the parameter space X by homeomorphisms and denote by T_g: $\mathscr{C}_0(X)\to\mathscr{C}_0(X)$ the action lifted to $\mathscr{C}_0(X)$ [i.e., $(T_g f)(x) = f(g^{-1}x)$]. Suppose furthermore that G acts by symmetries on a quantum system described in a Hilbert space $\mathscr{H}$, i.e., there is a projective representation U of G by unitary (and possibly antiunitary) operators. Then we shall call an observable F over X *U-covariant*, if $U_g F(f) U_g^* = F(T_g f)$ for $g\in G$ and $f\in\mathscr{C}_0(X)$. Equivalently, the measure F satisfies $U_g F(\sigma) U_g^* = F(g\sigma)$ for $g\in G$ and $\sigma\subset X$ measurable. The compact convex set of U-covariant observables over X will be denoted by $\mathscr{M}(U,X)$. An $F\in\mathscr{M}(U,X)$ will be called *pure*, if it cannot be decomposed into a sum of nonproportional elements of $\mathscr{M}(U,X)$. [Not every extreme point of $\mathscr{M}(U,X)$ is pure.] In typical applications (e.g., arrival time, position, phase-space, or screen observables), G is a group characterizing the kinematic properties of a particle.

One cornerstone of the constructions below is Mackey's theory of decision observables in $\mathscr{M}(U,X)$, called "systems of imprimity" by him. We can only indicate some basic results of this theory and must refer the reader to the literature.[9] Suppose G acts transitively on X and $F\in\mathscr{M}(U,X)$ is a descision observable for some unitary representation U in $\mathscr{H}$. Then, heuristically, we can diagonalize all operators $F(f)$, so that $\mathscr{H} = \mathscr{L}^2(X,dx;\mathscr{K})$, the space of square integrable functions over X with values in a Hilbert space $\mathscr{K}$, and $F(f)$ is multiplication by $f(x)$. By covariance, U must act as $(U_g\psi)(x) = \mathscr{D}(g,x)\psi(g^{-1}x)$, where $\mathscr{D}(g,x)$ is a transformation of $\mathscr{K}$, which must be unitary if the measure dx is invariant. If $\mathscr{D}(g_1,x)\,\mathscr{D}(g_2,g_1^{-1}x) = \mathscr{D}(g_1g_2,x)$, U is a representation. In particular $\mathscr{D}(\cdot,x_0)$ restricts to a representation of the subgroup $H = \{g\in G / gx_0 = x_0\}$. The representations of the conjugate subgroups belonging to different x_0 are conjugate via a suitable transformation $\mathscr{D}$, hence only one representation of the abstract group H is involved, which in turn characterizes the pair F,U up to unitary equivalence: Let G be separable and locally compact, $H\subset G$ a closed subgroup, and $X = G/H$. Identify X with a Borel subset of G, so that every $g\in G$ has a unique decomposition $g = \tilde{x}[g]\tilde{h}[g]$, with $\tilde{x}\in X$, $\tilde{h}\in H$, and suppose for simplicity that X has a G-invariant measure dx. Now let $\mathscr{D}$: $H\to\mathscr{U}(\mathscr{K})$ be a continuous unitary representation. Define $\mathscr{H} = \mathscr{L}^2(X, dx;\mathscr{K})$ and $F(f)$, $U_g\in\mathscr{B}(\mathscr{H})$ by $(F(f)\psi)(x) = f(x)\psi(x)$ and $(U_g\psi)(x) = \mathscr{D}(\tilde{h}[g^{-1}x]^{-1})\psi(g^{-1}x)$. Then U is a continuous unitary representation of G and $F\in\mathscr{M}(U,X)$, where U is called the representation induced from $\mathscr{D}$, and $\{F,U\}$ the canonical system of imprimitivity induced from $\mathscr{D}$. The results used below are (1) every decision observable $F\in\mathscr{M}(U,X)$ is induced from a representation $\mathscr{D}$ uniquely determined by F and U, and (2) $\mathscr{D}$ is irreducible iff the set of operators $\{F(f), U_g \mid f\in\mathscr{C}_0(X), g\in G\}$ is irreducible.

The second cornerstone of the constructions below is the following "dilation theorem," which reduces the construction of general covariant observables to the construction of imprimitivity systems.

Proposition 1: Let G be a group acting on the locally compact space X and U a projective representation of G by unitary and antiunitary operators on a Hilbert space $\mathscr{H}$. Let $F\in\mathscr{M}(U,X)$. Then there is a Hilbert space $\hat{\mathscr{H}}$ and a projective representation $\hat{U}$ of G with the same factor as U and such that $\hat{U}_g$ is unitary (resp. antiunitary) iff U_g is. Moreover there is an $\hat{F}\in\mathscr{M}(\hat{U},X)$ and a contraction V: $\mathscr{H}\to\hat{\mathscr{H}}$ such that

(1) $\hat{F}$ is a decision observable,

(2) $V^*\hat{F}(f)V = F(f)$, for $f\in\mathscr{C}_0(X)$,

(3) $VU_g = \hat{U}_g V$, for $g\in G$,

and

(4) $\{\hat{F}(f)V\psi \mid f\in\mathscr{C}_0(X),\ \psi\in\mathscr{H}\}$ is total in $\hat{\mathscr{H}}$.

Moreover $\hat{H}$, $\hat{U}$, V, and $\hat{F}$ are determined up to unitary equivalence and are called the *dilation* of F.

If G is a topological group, the action $G\times X\to X$ is continuous, and U is a continuous ray representation, then $\hat{U}$ is also continuous.

Sketch of proof: (For details see Refs. 4 and 5.) As a positive map on an Abelian C^*-algebra, F: $\mathscr{C}_0(X)\to\mathscr{B}(\mathscr{H})$, is completely positive. Hence we may apply the Stinespring dilation theorem and only have to check that this construction is naturally covariant.[10] Explicitly, we construct $\hat{\mathscr{H}}$ as the Hausdorff completion[11] of the algebraic tensor product $\tilde{\mathscr{H}} := \mathscr{C}_0(X)\otimes\mathscr{H}$ with respect to the sesquilinear form $\langle f_1\otimes\psi_1, f_2\otimes\psi_2\rangle := \langle\psi_1, F(\bar{f}_1 f_2)\psi_2\rangle$, which is positive by the complete positivity of F. We then define $\hat{F}$, $\hat{U}$, and V^* as the extensions by continuity of the linear operators

$$\tilde{F}(f')f\otimes\psi = (f'f)\otimes\psi,\quad \tilde{V}^* f\otimes\psi = F(f)\psi,$$

and

$$\tilde{U}_g f\otimes\psi = T_g f\otimes U_g\psi,$$

resp. the antilinear operator $\tilde{U}_g f\otimes\psi = \overline{T_g f}\otimes U_g\psi$ if U_g is antiunitary. If $\hat{\mathscr{H}}'$, $\hat{U}'$, $\hat{F}'$, V' is another dilation, define $\tilde{S}$: $\tilde{\mathscr{H}}\to\hat{\mathscr{H}}'$ by $\tilde{S}f\otimes\psi = \hat{F}'(f)V'\psi$. The extension S: $\hat{\mathscr{H}}\to\hat{\mathscr{H}}'$ is then a unitary equivalence with $V' = SV$.

Q.E.D.

Similarly, as for the GNS case of the Stinespring dilation, we have the following "Radon–Nikodym" result.

Proposition 2: Let G, X, $\mathscr{H}$, and U be as above and F, $F_1\in\mathscr{M}(U,X)$ be covariant observables with $f\geqslant 0\Rightarrow F_1(f)\leqslant\lambda F(f)$ with $\lambda\geqslant 0$. Let $\hat{\mathscr{H}}$, $\hat{U}$, V, and $\hat{F}$ be the dilation of F. Then there is a unique $R\in\mathscr{B}(\hat{\mathscr{H}})$ such that

(1) $0\leqslant R\leqslant\lambda\cdot 1$,

(2) $[R,\hat{F}(f)] = 0$, $[R,\hat{U}_g] = 0$, for $f\in\mathscr{C}_0(X)$, $g\in G$,

(3) $V^*R\hat{F}(f)V = F_1(f)$.

Consequently F is pure iff $\hat{F}(\mathscr{C}_0(X))\cup\hat{U}(G)$ is an irreducible set of operators.

Proof: Since R commutes with $\hat{F}$, Eq. (3) determines the matrix elements of R for a dense set of vectors. Define on $\tilde{\mathscr{H}}$ as in the proof of Proposition 1 the sesquilinear form $\tilde{R}(f_1\otimes\psi_1, f_2\otimes\psi_2) := \langle\psi_1, F_1(\bar{f}_1 f_2)\psi_2\rangle$. This is positive by the (complete) positivity of F_1 and bounded by λ 1 by the positivity of $\lambda F - F_1$. Hence $\tilde{R}$ extends to a bounded form on $\hat{\mathscr{H}}$, given by an operator R with the above properties.

Q.E.D.

We may now combine the imprimitivity theorem and the dilation theorem to a general construction procedure for covariant observables on homogeneous spaces $X = G/H$. The dilating operator V identifies U as a subrepresentation of the representation $\hat{U}$, which by imprimitivity must be induced from some representation $\hat{\mathscr{D}}$ of H.

Construction procedure: Let G be a separable locally compact group, H a closed subgroup, and $X = G/H$. Let U: $G \rightarrow \mathscr{U}(\mathscr{H})$ be a strongly continuous unitary representation. In order to construct $F \in \mathscr{M}(U,X)$, proceed as follows.

Step 1: Select a continuous unitary representation $\hat{\mathscr{D}}$: $H \rightarrow \mathscr{U}(\hat{\mathscr{K}})$.

Step 2: Induce from $\hat{\mathscr{D}}$ the canonical system of imprimitivity consisting of a unitary representation $\hat{U}: G \rightarrow \mathscr{U}(\hat{\mathscr{H}})$ and a decision observable $\hat{F} \in \mathscr{M}(\hat{U},X)$.

Step 3: Find an intertwining operator V: $\mathscr{H} \rightarrow \hat{\mathscr{H}}$ between the representations U and $\hat{U}$, such that $\|V\| \leqslant 1$ and $\hat{F}(\mathscr{C}_0(X))V\mathscr{H} = \hat{\mathscr{H}}$.

Assertion: $F(f) := V^* \hat{F}(f) V \, [f \in C_0(X)]$ defines a U-covariant observable. All $F \in \mathscr{M}(U,X)$ can be constructed in this way; $\hat{\mathscr{D}}$ and V are uniquely determined by F up to unitary equivalence, and F is pure in $\mathscr{M}(U,X)$ iff $\hat{\mathscr{D}}$ is irreducible.

Of course, $\hat{\mathscr{D}}$ is not completely arbitrary if nontrivial intertwining operators V are to exist in step 3.

III. $\mathbb{R}^n$-COVARIANT OBSERVABLES

Screen observables, will be defined as observables, which are based on a hyperplane X, and which are covariant with respect to a group containing the translations along X. Since some basic properties of screens depend only on this restricted covariance condition, it is useful to study covariant observables with $G = X = \mathbb{R}^n$ separately. Throughout this section a representation U: $\mathbb{R}^n \rightarrow \mathscr{U}(\mathscr{H})$ will be assumed to be given.

In order to construct a pure $F \in \mathscr{M}(U,X)$ we have to pick first an irreducible representation of the group $H = \{e\}$, which is trivial. The induction process then yields the regular representation $(\hat{U}_y \psi)(x) := \psi(x - y)$ in $\mathscr{L}^2(\mathbb{R}^n, d^n x)$ and $(\hat{F}(f)\psi)(x) = f(x) \cdot \psi(x)$, i.e., the usual representation of translations and position over $X = \mathbb{R}^n$. (This instance of the imprimitivity theorem is known as von Neumann's uniqueness theorem.) Starting instead with a reducible representation of $\{e\}$ we obtain a direct multiple of the system $\{\hat{U},\hat{F}\}$ and hence the following result: Every $F \in \mathscr{M}(U,X)$ is of the form $F(f) = \Sigma_i V_i^* \hat{F}(f) V_i$, where for all i, V_i: $\mathscr{H} \rightarrow \mathscr{L}^2(X, dx)$ intertwines U and $\hat{U}$, and $F(X) := \Sigma_i V_i^* V_i \leqslant 1$.

With each state $W \in \mathscr{T}(\mathscr{H})$ we may associate the positive measures $\mu_W(dx) = \operatorname{tr} WF(dx)$ over X and $\nu_W(d\xi) = \operatorname{tr}(W F(X)^{1/2} E(d\xi) F(X)^{1/2})$ over the dual Ξ of X, where E is the spectral measure of U. [If $\operatorname{tr} WF(X) = 1$, μ_W and ν_W are both normalized.] The relationship between these measures can be studied very easily by the above representation for F: $\mu_W(dx) = \operatorname{tr} \hat{W}\hat{F}(dx)$ and ν_W, with $\int \nu_W(d\xi) e^{i\xi x} = \operatorname{tr} \hat{W}\hat{U}_x$, are the distributions of "pseudoposition" and "pseudomomentum" for the Schrödinger system $\{\hat{F},\hat{U}\}$ in the state $\hat{W} := \Sigma_i V_i W V_i^* \in \mathscr{T}(\hat{\mathscr{H}})$. In particular if $\operatorname{tr} WF(X) = 1$, the product of the variances of the probability measures μ_W and ν_W is bounded away from zero. For screen observables these are the *uncertainty relations* between energy and arrival time and between transverse momentum and arrival location.[6]

Both μ_W and ν_W are absolutely continuous. If E_{ac} denotes the projector onto the absolutely continuous spectral subspace of U, then for all $f \in \mathscr{C}_0(X)$: $F(f)(1 - E_{ac}) = 0$. For example, if U describes the time evolution of the system ($\dim X = 1$), this implies that a covariant arrival time observable is insensitive to bound states of the Hamiltonian. In the case of the harmonic oscillator there are no covariant time observables at all. This is not surprising, since quantities like "the time t at which the particle reaches the origin" make sense only modulo periods of U. In the classical case this difficulty is reflected in the nonexistence of variables globally canonically conjugate to the energy. Of course, in the case of the oscillator (but not in more complex cases) we may consider instead covariance with respect to $G = X = \mathbb{R}/\mathbb{Z}$.

The support of ν_W is contained in the absolutely continuous spectrum $\Sigma := \operatorname{supp}\{E_{ac} E(\cdot)\}$ of U. This imposes constraints on the support of μ_W. For example, if there is a state W such that μ_W is nonzero and supported by a proper cone in X, then by the "edge of the wedge" theorem[12] we must have $\Sigma = \mathbb{R}^n$. In particular, unless $\Sigma = \mathbb{R}^n$, we cannot find a projection-valued $F \in \mathscr{M}(U,X)$, since for such F we could choose μ_W to be supported by an arbitrary set of positive measure. We may also use the edge of the wedge theorem for the reverse Fourier transform to conclude that if Σ is contained in a proper cone the support of μ_W is equal to X (or $\mu_W = 0$). Since this spectral condition is satisfied for screen observables, we conclude that there is no apparatus preparing particles in such a way that they avoid with certainty some nontrivial patch of some translation covariant screen. Of course this does not mean that μ_W cannot be highly concentrated. For example, some of the ideal screen observables constructed below (namely those for spinless and/or nonrelativistic particles) are easily seen to be concentratable[13] in the sense that we can make $\mu_W(\sigma) > 1 - \epsilon$ for any open set σ.

Another characteristic difference between covariant decision observables and the more general type considered here is the following: For a decision observable, the weak closure of the range $F(\mathscr{C}_0(X))$ is an Abelian von Neumann algebra. In the general case, however, this space will not be an algebra and the von Neumann algebra generated by the range of an observable may be considerably larger than the range itself. To be specific, suppose that $F \in \mathscr{M}(U,X)$ is pure $\mathbb{R}^n$-covariant. Since $F(f) = V^* \hat{F}(f) V$, the von Neumann algebra generated by $F(\mathscr{C}_0(X))$ is the closure of $V^* \mathscr{M} V$, where $\mathscr{M} \subset \mathscr{B}(\hat{\mathscr{H}})$ is the von Neumann algebra generated by $\hat{F}(\mathscr{C}_0(X))$ and VV^*. It is useful to think of $\hat{F}(\mathscr{C}_0(X))''$ as the space of functions of "pseudoposition" $\hat{Q}$ and of VV^* as a specific function $\lambda(\hat{P})$ of "pseudomomentum." Here, $\mathscr{M}$ is clearly invariant under "position" translations and since it contains exponential functions of "position" it is also invariant under "momentum" translations. By the general theory of phase-space translation invariant operator spaces[14] the commutant of $\mathscr{M}$ is generated by the operators $\exp(ip \cdot \hat{Q})$

with $p\in\mathbf{R}^n$ a period of λ. A nontrivial period of λ is impossible if Σ (and hence supp λ) is contained in a proper cone. In this case the range $F(\mathscr{C}_0(X))$ of every pure U-covariant observable is irreducible on its support $F(X)\mathscr{H}$.

By the spectral theorem each decision observable F over $\mathbf{R}$ is uniquely related to a self-adjoint operator $Z=\int xF(dx)$, where Z is interpreted as an expectation value functional in the sense that tr WZ is the expectation value of the probability measure tr $WF(\cdot)$. This definition and interpretation of Z immediately generalizes to the case of general (i.e., not projection-valued) observables. However, it is then no longer true that Z uniquely determines F. In particular, $\mathrm{tr}(WZ^2)$ need not be equal to the second moment of the probability measure tr $WF(\cdot)$. The failure of this equality is measured by the variance form of F, introduced in the following proposition. It is formulated for the multidimensional case and does not depend on a covariance condition.

Proposition 3: Let F: $\mathscr{C}_0(\mathbf{R}^n)\to\mathscr{B}(\mathscr{H})$ be an observable.

(1) Then the integrals $Z_\nu\varphi:=\int x_\nu F(dx)\varphi$ $(\nu=1,...,n)$ converge strongly for φ in the domain

$$\mathscr{D}(Z):=\left\{\varphi\in\mathscr{H}\,\middle|\,\int\left(\sum_{\nu=1}^{n}x_\nu^2\right)\langle\varphi,F(dx)\varphi\rangle<\infty\right\}.$$

The symmetric operators Z_ν defined by this formula are called the *expectation operators* of F.

(2) For φ, $\psi\in\mathscr{D}(Z)$ and $\nu,\mu=1,...,n$, define

$$\Delta_{\nu\mu}(\varphi,\psi):=\int x_\nu x_\mu\langle\varphi,F(dx)\psi\rangle-\langle Z_\nu\varphi,Z_\mu\psi\rangle,$$

where Δ is called the *variance form* of F and is positive in the sense that for $\varphi_\nu\in\mathscr{D}(Z)$: $\Sigma_{\nu\mu}\Delta_{\nu\mu}(\varphi_\nu,\varphi_\mu)\geqslant0$; F is called *variance free*, if $\mathscr{D}(Z)$ is dense and $\Delta=0$.

Proof:

(1) (The validity of this statement is erroneously denied in Ref. 15, p. 339.) Consider the set of cutoff functions $h\in\mathscr{C}_0(\mathbf{R}^n)$, with $0\leqslant h\leqslant1$ and compact support, directed by pointwise ordering. The claim is that for $\varphi\in\mathscr{D}(Z)$ and $\nu=1,...,n$, $h\to F(x_\nu\cdot h)\varphi$ is a norm Cauchy net. Consider the dilation $\hat{\mathscr{H}}$, $\hat{\mathscr{F}}$, V of F (see Proposition 1 with trivial group). Then for two cutoffs h, h':

$$\begin{aligned}
&\|F(x_\nu h)\varphi-F(x_\nu h')\varphi\|^2\\
&\quad=\|V^*\hat{F}(x_\nu(h-h'))V\varphi\|^2\\
&\quad\leqslant\|\hat{F}(x_\nu(h-h'))V\varphi\|^2\\
&\quad=\langle V\varphi,\hat{F}(x_\nu^2(h-h')^2)V\varphi\rangle\\
&\quad=\int x_\nu^2(h(x)-h'(x))^2\langle\varphi,F(dx)\varphi\rangle.
\end{aligned}$$

As h and h' increase to 1, this goes to 0 by dominated convergence.

(2) For the dilation $\hat{F}$, define $\hat{Z}_\nu$ and $\mathscr{D}(\hat{Z})$ analogously. Then $\mathscr{D}(Z)=\{\varphi\mid V\varphi\in\mathscr{D}(\hat{Z})\}$ and the above argument shows that for

$$\varphi\in\mathscr{D}(Z),\quad Z_\nu\varphi=V^*\hat{Z}_\nu V\varphi,$$

then

$$\Delta_{\nu\mu}(\varphi,\psi)=\langle\hat{Z}_\nu V\varphi,\hat{Z}_\mu V\psi\rangle-\langle V^*Z_\nu V\varphi,V^*Z_\mu V\psi\rangle.$$

Hence if

$$\varphi_\nu\in\mathscr{D}(Z):\sum_{\nu\mu}\Delta_{\nu\mu}(\varphi_\nu,\varphi_\mu)=\|\phi\|^2-\|V^*\phi\|^2\geqslant0,$$

with

$$\phi=\sum_\nu\hat{Z}_\nu V\varphi_\nu\text{ and }\|V\|\leqslant1.\qquad\text{Q.E.D.}$$

The variance form Δ describes an "uncertainty," which is intrinsic to the measurement of F. Decision observables are variance-free. The converse is false, in general, but holds for normalized observables of compact support. Thus variance-free observables are a natural extension of the class of decision observables. Even for variance-free observables, the symmetric operators Z_ν need not have any self-adjoint extensions. They also need not commute, although they are by definition jointly measured by F; $(\Delta_{\nu\mu}-\Delta_{\mu\nu})$ is just the commutator of Z_ν and Z_μ, written as a quadratic form on $\mathscr{D}(Z)$. The following example shows that this form may indeed be nonzero.

Example: Covariant phase-space observables (cf. Refs. 14). Let $X=\mathbf{R}^{2n}$ be a phase-space, equipped with a symplectic form $\sigma(x,y)=\sigma^{\nu\mu}x_\nu y_\mu$, and let U: $X\to\mathscr{U}(\mathscr{H})$ be an irreducible representation of the Weyl relations $U_xU_y=\exp[(i/2)\,\sigma(x,y)\cdot U_{x+y}]$. Then the self-adjoint generators R^ν of U defined by $U_x=\exp(ix_\nu R^\nu)$ satisfy the canonical commutation relations $i[R^\nu,R^\mu]=\sigma^{\nu\mu}\cdot1$. The U-covariant observables F over X are all of the form $F(f)=\int dx\,f(x)U_x\dot{F}U_x^*$, where $\dot{F}\geqslant0$ is a trace class operator. We shall assume a suitable normalization of Lebesgue measure dx and tr $\dot{F}=1$, so that F becomes normalized. The measure $\mu_W=\mathrm{tr}\,WF(\cdot)$ has the Radon–Nikodym density $x\to\mathrm{tr}(WU_x\dot{F}U_x^*)$, which should be thought of as a convolution of W and $\dot{F}$. The moments of μ_W can be calculated quite simply from the "moments" of W and $\dot{F}$ (we set $x^\nu=\sigma^{\nu\mu}x_\mu$ and tr $W=1$)

$$\int x^\nu\mu_W(dx)=\mathrm{tr}\,WR^\nu-\mathrm{tr}\,\dot{F}R^\nu,$$

$$\begin{aligned}\int x^\nu x^\mu\mu_W(dx)&=\mathrm{tr}(WR^\nu R^\nu)+\mathrm{tr}\,\dot{F}R^\nu R^\mu+i\sigma^{\nu\mu}\\
&\quad-\mathrm{tr}\,WR^\nu\,\mathrm{tr}\,\dot{F}R^\mu-\mathrm{tr}\,WR^\mu\,\mathrm{tr}\,\dot{F}R^\nu.\end{aligned}$$

(Note that this is real symmetric by virtue of the commutation relations.)

Since $U_x\mathscr{D}(Z)=\mathscr{D}(Z)$ and U_x is irreducible, $\mathscr{D}(Z)=\{0\}$ or $\mathscr{D}(Z)$ is dense. The above formula shows that the first case occurs iff one of the moments $\mathrm{tr}(\dot{F}R^\nu R^\mu)$ diverges. Otherwise

$$\mathscr{D}(Z)=\bigcap_\nu\mathscr{D}(R^\nu)$$

is independent of F. Then $Z_\nu=R_\nu-(\mathrm{tr}\,\dot{F}R_\nu)\cdot1$ with $\sigma^{\nu\mu}R_\nu=R^\mu$. Since F is normalized, Δ is invariant under U and hence must be the restriction to $\mathscr{D}(Z)$ of a multiple of the identity

$$\Delta_{\mu\nu}=\{\mathrm{tr}(\dot{F}R_\nu R_\mu)-\mathrm{tr}\,\dot{F}R_\nu\cdot\mathrm{tr}\,\dot{F}R_\mu+i\sigma_{\nu\mu}\}\cdot1.$$

For calculating the second moments of μ_W, it is sufficient to know Z_ν and the symmetric part of $\Delta_{\nu\mu}$. However, $\Delta_{\nu\mu}+\Delta_{\mu\nu}$ is constrained by the condition that the complex matrix Δ, with its prescribed antisymmetric part $i\sigma_{\nu\mu}$, is positive. This implies inequalities also for the symmetric part of Δ, equivalent to the usual uncertainty relations for the trace operator $\dot{F}$. Thus there is a lower bound to the intrinsic variance Δ of any joint measurement of position and momentum (even independently of the covariance condition). This supplements the usual uncertainty relations, which refer only to the impossibility of preparing certain states. The observables F for which the variance Δ is minimal are given by coherent states $\dot{F}$ and are indeed the most widely used covariant phase-space observables. A similar condition of minimal variance form will be used below to single out ideal screen observables.

IV. DEFINITION OF SCREEN OBSERVABLES

In this section the only difference between the relativistic and the nonrelativistic case will be the choice of the kinematic group K, which will be either the Galilei or the Poincaré group (including reflections). The group K acts by affine transformations on the four-dimensional space-time manifold, in which we shall consider a fixed three-dimensional hyperplane X containing a timelike direction. Thus in a suitable coordinate system, which will also be held fixed in the sequel, $X = \{(x_0,x_1,x_2,x_3)\in\mathbb{R}^4 | x_3 = 0\}$. The intersection of the worldline of a classical particle with X can be interpreted as the event of the particle hitting a screen that is stationary in the 1-2 plane. An ensemble of classical free particles will thus produce a probability distribution over X. Shifting this ensemble by a transformation $g\in K$ with $gX\subset X$ will shift this probability distribution by the same transformation g. This aspect of the classical trajectory concept can be transcribed into quantum mechanics as a covariance condition for an observable as follows. Let G denote the connected component of the subgroup of $g\in K$ with $gX\subset X$. The kinematic properties of the kind of quantum particles considered are characterized by an irreducible projective representation U of K in a Hilbert space $\mathscr{H}$ with mass m and spin s. Then a *screen observable* is an observable F over X covariant for the restriction of U to G, i.e., $F\in\mathscr{M}(U\upharpoonright G,X)$.

The set of screen observables is very large, which is not unreasonable, since there are many screenlike measuring devices, supposedly associated with different observables over X. The detection of a particle by such a device generally depends on some scattering or ionization process in the screen. The cross section of this process (as a function of the particle's momentum) enters the expression for the probability that the particle is detected at all, given by the operator $F(X)$. On the other hand, for the observable associated to the classical kinematic concept of particles meeting a timelike plane, (almost) all particles will be "detected" so that $F(X) = 1$, independently of momentum, i.e., the kinematic concept abstracts from any material realization of the screen. In this sense the aim of constructing "ideal" quantum screen observables mimicking the classical kinematic concept is opposed to the construction of faithful models for real screens. Nevertheless the ideal screen observables described below can be useful in the analysis of concrete experiments, e.g., when a more detailed analysis is too difficult and can reasonably be expected to have little influence on the results.

The nonuniqueness of screen observables arises largely because the representation $U\upharpoonright G$ is reducible. (The commutant of $U\upharpoonright G$ is generated by functions of the components P_3 and S_3 of momentum and spin normal to the screen.) Hence if $A\in\mathscr{B}(\mathscr{H})$ commutes with U_G and F is a screen observable, then so is $\tilde{F}(f) = A^*F(f)A$. In order to define unique ideal screen observables, we therefore have to introduce conditions excluding such transformations. Specifically we shall define an *ideal screen observable* as a screen observable, which can be decomposed into pure components F satisfying the four conditions below.

Condition 1: $F(X)$ is a projection to a subspace of the spectral subspace $[P_3>0]$ of P_3.

This condition introduces the restriction that the screen should be "one-sided." We could also have restricted ourselves to $P_3<0$ or a symmetric/antisymmetric subspace with respect to $P_3\to -P_3$. Together with condition 3 this condition implies that the total detection probability $F(X)$ is equal to one for $P_3>0$. It also forces the decomposition of an ideal screen observable into pure components to be as simple as possible, namely a direct sum.

Condition 2: F is also covariant for the transformation $\theta\in K$ with $\theta(x_0,x_1,x_2,x_3) = (-x_0, -x_1,x_2, -x_3)$.

Without this condition a transformation of F with $A = \exp\{iP_3\cdot a\}$ would still be admissible, so that there would be no justification to associate F with the hyperplane $x_3 = 0$ rather than $x_3 = a$. The reflection θ chosen here leaves P_3 and S_3 invariant.

Condition 3: If $\varphi\in\mathscr{H}$ is a differentiable vector for the whole group K and $\varphi(p)$ has compact support in the half space $p_3>0$, then the second moments of the measure $\langle\varphi,F(\cdot)\varphi\rangle$ exist.

This "regularity condition" rules out transformation by self-adjoint unitary functions of p_3.

Condition 4: The variance form $\Delta_{\nu\mu}$ of F is minimal.

The variance form Δ (see Sec. III) describes the intrinsic "uncertainties" in the measurement of F. The first three conditions single out one parameter families of pure screen observables. The variance form [or more precisely, all expressions $\Sigma_{\nu\mu}\Delta_{\nu\mu}(\varphi_\nu,\varphi_\mu)$ with $\varphi_\nu\in\mathscr{D}(Z)$] depends monotonically on this parameter Γ. Thus condition 4 demands the choice of the minimal value of Γ. A pure screen observable is characterized by conditions 1–4 up to the choice of the subset $\gamma\subset\{-s,..., +s\}$ of the spectrum of S_3 by which it is supported. By forming a direct sum of such observables with $\gamma = \{n\}$, we obtain the following result.

Main result: Given an irreducible representation of K with mass m and spin s, there is a unique ideal screen observable such that $F(X)$ is the projection onto $[P_3>0]$ and $F(f)$ commutes with the spin component S_3. The latter condition is redundant in the nonrelativistic case and for $s = 0$, $\frac{1}{2}$, 1.

With the necessary changes for a spacelike hyperplane X these conditions define an "ideal position observable," which turns out to be the standard position observable in the nonrelativistic case and the Newton–Wigner–Wightman observable in the (massive) relativistic case. In this sense the above conditions characterize the exact analogs of the standard position observables in the screen case.[16]

V. CONSTRUCTION OF SCREEN OBSERVABLES

We shall begin by introducing some notations and the explicit form of the representation U describing the particle.

It will be convenient to take the symbol K now for the covering group of the Poincaré/Galilei group. (The action of K on space-time is defined in the canonical way.) Then $G\subset K$ will be the connected component of the subgroup leaving X invariant, $H\subset G$ the subgroup leaving the origin invariant (i.e., boosts along the screen and rotations of the screen) and $T\subset H$ will denote the subgroup of rotations (i.e., $T\simeq\mathbb{R} \bmod 4\pi$). Here T is also a subgroup of the rotation subgroup $SU_2\subset K$ serving as "little group" in the construction of representations of K.

In order to define U in the relativistic case let $\beta(p)\in K$, for $p\in\mathbb{R}^4$, $p^0>|\mathbf{p}|$, denote the pure boost taking $\bar{p}:=(\sqrt{p^\mu p_\mu},0,0,0)$ to p. Later we shall also need the pure boost $b(p)\in H$ taking $\hat{p}:=\left(\sqrt{p_0^2-p_1^2-p_2^2},0,0,p_3\right)$ to p. This $b(p)$ is independent of p_3. We will denote by $\rho(p)\in SU_2$ the rotation, which is the product of the pure boosts from $\bar{p}$ to p to $\hat{p}$ and back to $\bar{p}$. Now let $m>0$ and s be an integer or half integer. (We shall not consider $m=0$, for the sole reason of saving space.)

Let $\mathscr{D}:SU_2\to\mathscr{U}(\mathscr{K})$ be the irreducible representation with $\dim\mathscr{K}=2s+1$, and let $\mathscr{H}:=\mathscr{L}^2(\mathbb{R}^3,d^3p/p_0,\mathscr{K})$, where $p_0=(m^2+p^2)^{1/2}$. Then for $(y,\Lambda)\in K$,

$$[U(y,\Lambda)]\psi(\mathbf{p})=e^{iy\cdot p}\,\mathscr{D}(\beta(p)^{-1}\Lambda\beta(\Lambda^{-1}p))\psi(\Lambda^{-1}p).$$

In the nonrelativistic case the group elements will be parametrized as $(y_0,y;\,u,R)\in K$, where $y_0\in\mathbb{R}$, $y,u\in\mathbb{R}^3$, and $R\in SU_2$. Here K acts on $\mathbb{R}^{1+3}$ via $(y_0,y;R)(x_0,x)=(x_0+y_0,Rx+ux_0+y)$.

Now let $m>0$ and s be an integer or half integer and $\mathscr{D}$: $SU_2\to\mathscr{U}(\mathscr{K})$ as in the relativistic case. Let $\mathscr{H}=\mathscr{L}^2(\mathbb{R}^3,d^3p,\mathscr{K})$. Then

$$[U(x_0,y,u,R)\psi](p)=\exp i\left\{\frac{m}{2}u\cdot y+\frac{p^2}{2m}y_0-p\cdot y\right\}\times\mathscr{D}(R)(R^{-1}\psi(p-mu))$$

defines a projective representation of K. For later use we introduce the notation $b(p):=(0,0;p_1/m,p_2/m;\,1)\in H$.

The main results of this section are collected in the following theorem.

Proposition 4: Let $\hat{\mathscr{D}}:H\to\mathscr{U}(\hat{\mathscr{K}})$ be a continuous unitary representation and $\mathscr{C}$: $\mathbb{R}\to\mathscr{B}(\mathscr{K},\hat{\mathscr{K}})$ a measurable function such that for each $p_3\in\mathbb{R}$, $\mathscr{C}(p_3)$: $\mathscr{K}\to\hat{\mathscr{K}}$ intertwines $\mathscr{D}\restriction T$ and $\hat{\mathscr{D}}\restriction T$ and

$$\mathscr{C}(p_3)\mathscr{C}(p_3)^*+\mathscr{C}(-p_3)\mathscr{C}(-p_3)^*\leqslant 1.$$

Let $\hat{\mathscr{H}}=\mathscr{L}^2(X,d^3x;\hat{\mathscr{K}})$, $\hat{F}(f)\in\mathscr{B}(\hat{\mathscr{H}})$ the operator of multiplication by $f\in\mathscr{C}_0(X)$, and define V: $\mathscr{H}\to\hat{\mathscr{H}}$ by

$$(V\psi)(x_0,\mathbf{x})=(2\pi)^{-3/2}\int\frac{d^3p}{p_0}|p_3|^{1/2}\times\exp(-ip_0x_0+i\mathbf{p}\cdot\mathbf{x})\hat{\mathscr{D}}(b(p))\mathscr{C}(p_3)\times\mathscr{D}(\rho(p))\psi(p)$$

in the relativistic case, and

$$(V\psi)(x_0,\mathbf{x})=(2\pi)^{-3/2}\int d^3p\left|\frac{p_3}{m}\right|^{1/2}\times\exp\left(-i\frac{p^2}{2m}x_0+i\mathbf{p}\cdot\mathbf{x}\right)\hat{\mathscr{D}}(b(p))\times\mathscr{C}(p_3)\psi(p)$$

in the nonrelativistic case. Then $F(f):=V^*\hat{F}(f)V$ defines a screen observable, and all screen observables are constructed in this way. Moreover, F is pure and satisfies conditions 1, 2, and 3 of Sec. IV iff $\hat{\mathscr{D}}$ is irreducible, $\mathscr{C}(p_3)=0$ for $p_3<0$, $\mathscr{C}(p_3)\equiv\mathscr{C}$ is a constant isometry for $p_3>0$, and $\hat{\mathscr{D}}(\theta)\mathscr{C}=\mathscr{C}\mathscr{D}(\theta)$ for some antiunitary operators $\hat{\mathscr{D}}(\theta)$ and $\mathscr{D}(\theta)$ in $\hat{\mathscr{K}}$ and $\mathscr{K}$, representing the reflection $p_2\to-p_2$.

Proof: Both relativistic and nonrelativistic construction follow the scheme outlined in Sec. II. In the following proof these two cases will be distinguished by labels (R i) and (NR i), $i=1,2,3$. In step (1) the representation $\hat{U}$: $G\to\mathscr{U}(\hat{\mathscr{H}})$ will be induced from the given representation $\hat{\mathscr{D}}$. According to the scheme, we then have to find all intertwining operators between the reducible representations $\hat{U}$ and $U\restriction G$. To this end $\hat{U}$ is transformed in step (2) by an isometry $\hat{V}^*$: $\hat{\mathscr{H}}\to\tilde{\mathscr{H}}:=\mathscr{L}^2(\Xi,d^3\xi,\hat{\mathscr{K}})$ to a representation $\tilde{U}$ whose irreducible components are more readily analyzed. In step (3) the intertwining operators $\tilde{V}^*$: $\tilde{\mathscr{H}}\to\mathscr{H}$ between $\tilde{U}$ and U are constructed and parametrized by $\mathscr{C}$. Then $V=\hat{V}\tilde{V}$ takes the form given in the theorem. By the results of Sec. II we thus obtain the most general screen observable. The remaining properties are checked in step (4).

(R 1) Identify X with the translation subgroup of G. Then the decomposition $g=\tilde{x}[g]\tilde{h}[g]\in XH$ is simply $(y,\Lambda)=(y,1)(0,\Lambda)$. Hence $\tilde{h}[g^{-1}x]^{-1}=\tilde{h}[g]$, and hence $[\hat{U}(y,\Lambda)\psi](x)=\hat{\mathscr{D}}(\Lambda)\psi(\Lambda^{-1}(x-y))$, for $\psi\in\hat{\mathscr{H}}:=\mathscr{L}^2(X,dx;\hat{\mathscr{K}})$.

(R 2) Set $\tilde{\mathscr{H}}:=\mathscr{L}^2(\Xi_+,d\xi_0\,d\xi_1\,d\xi_2;\hat{\mathscr{K}})$, where $\Xi_+=\{\xi\in\mathbb{R}^3|\xi_0>(\xi_1^2+\xi_2^2)^{1/2}\}$ is considered as a subset of the dual space of X. Then

$$(\tilde{U}(y,\Lambda)\psi)(\xi)=e^{iy\cdot\xi}\hat{\mathscr{D}}(b(\xi)^{-1}\Lambda b(\Lambda^{-1}\xi))\psi(\Lambda^{-1}\xi),$$

for $(y,\Lambda)\in G$, $\psi\in\tilde{\mathscr{H}}$ defines a representation; $\tilde{U}$ depends only on $\hat{\mathscr{D}}\restriction T$. The commutant of $\hat{\mathscr{D}}(T)$, together with the function $\xi_0^2-\xi_1^2-\xi_2^2$, generates the commutant of $\tilde{U}$. Now $\hat{V}$: $\tilde{\mathscr{H}}\to\hat{\mathscr{H}}$, defined by

$$(\hat{V}\psi)(x)=(2\pi)^{-3/2}\int d^3\xi e^{-i\xi\cdot x}\hat{\mathscr{D}}(b(\xi))\psi(\xi)$$

is an isometry intertwining $\tilde{U}$ and $\hat{U}$.

(R 3) Suppose $\tilde{V}^*$: $\tilde{\mathscr{H}}\to\mathscr{H}$ intertwines $\tilde{U}$ and $U\restriction G$. Then since $\tilde{V}^*$ intertwines the translation subgroups, $(\tilde{V}^*\psi)$ $(p_1,p_2,p_3)\in\mathscr{K}$ can only depend on $\psi(p_0,p_1,p_2)\in\hat{\mathscr{K}}$. Assume this correspondence to be given by $\tilde{\mathscr{C}}^*(p)$: $\hat{\mathscr{K}}\to\mathscr{K}$. Then $\tilde{V}^*$ intertwines iff

$$\tilde{\mathscr{C}}(p)^*\hat{\mathscr{D}}(b(p)^{-1}\Lambda b(\Lambda^{-1}p))=\mathscr{D}(\beta(p)^{-1}\Lambda\beta(\Lambda^{-1}p))\tilde{\mathscr{C}}^*(\Lambda^{-1}p)$$

for all p and $\Lambda\in H$. Using the identity

$$\rho(p)\beta(p)^{-1}\Lambda\beta(\Lambda^{-1}p)=b(p)^{-1}\Lambda b(\Lambda^{-1}p)\rho(\Lambda^{-1}p),$$

with

$$\rho(p)=\beta(\hat{p})^{-1}b(p)^{-1}\beta(p),$$

and the definition

$$\tilde{\mathscr{C}}(p)=|p_3|^{1/2}\mathscr{C}(p)\mathscr{D}(\rho(p)),$$

this condition becomes equivalent to

$$\mathscr{C}(p)=\mathscr{C}(\Lambda p)=\hat{\mathscr{D}}(r)\mathscr{C}(p)\mathscr{D}(r)^*,$$

for $\Lambda \in H, r \in T$, i.e., $\mathscr{C}(p) = \mathscr{C}(p_3)$ intertwines rotations. The bound on $\mathscr{C}$ results from $\|\widetilde{V}^*\psi\|^2 \leqslant \|\psi\|^2$, where the factors $|p_3|^{1/2}$ cancel in the substitution $(p_1, p_2, p_3) \to (\xi_0, \xi_1, \xi_2)$.

(NR 1) In this case U is a projective representation of K. The associated central extension $\widetilde{K} \supset K$ has typical elements (y_0, y, u, R, ζ) with $\zeta \in \mathbb{C}$, $|\zeta| = 1$, and a multiplication law $g'' = gg'$ extended by

$$\zeta'' = \zeta\zeta' \exp[(im/2)(u \cdot Ry' - y \cdot Ru' - y_0 u \cdot Ru')].$$

Now, U becomes a true representation of $\widetilde{K}$ with the definition $U(0,0,0,1,\zeta) = \zeta \cdot 1$. The subgroup of $\widetilde{K}$ generated by G (resp. H) and the center of $\widetilde{K}$ will be denoted by $\widetilde{G}$ (resp. $\widetilde{H}$). Then $X = G/H = \widetilde{G}/\widetilde{H}$. Identifying X with the translation subgroup in $\widetilde{G}$ we obtain the Mackey decomposition $g = \tilde{x}[g] \cdot \tilde{h}[g] \in \widetilde{G}$ given by $(y_0, y, u, R, \zeta) = (y_0, y, 0, 1, 1)$ $(0,0,u,R,\zeta \cdot \exp[(im/2) \times u \cdot y])$. We may now apply the construction procedure to $\widetilde{G}$ and $\widetilde{H}$. Proposition 1 yields the additional information that $\widehat{U}$ can be considered as a projective representation of K with the same factor as U or, equivalently, that $\widehat{U}(0,0,0,1,\zeta) = \zeta \cdot 1$. This imposes on the representation $\widehat{\mathscr{D}}$: $\widetilde{H} \to \mathscr{U}(\widehat{\mathscr{K}})$, from which $\{\widehat{U}, \widehat{F}\}$ is induced, the constraint $\widehat{\mathscr{D}}(0,0,0,1,\zeta) = \zeta \cdot 1$, and, since the extension $\widetilde{H} = H \otimes U(1)$ is trivial, $\widehat{\mathscr{D}}(0,0,u,R,\zeta) = \zeta \cdot \widehat{\mathscr{D}}(u, R)$, where $\widehat{\mathscr{D}}$ is a representation of H in $\widehat{\mathscr{K}}$. Then the induction procedure yields $\widehat{\mathscr{H}} = \mathscr{L}^2(X, dx, \widehat{\mathscr{K}})$ and

$$\begin{aligned}[\widehat{U}(y_0, y, u, R, \zeta)\psi](x_0, x) &= \zeta \exp[(im/2)u \cdot \{2x - y - (x_0 - y_0) \cdot u\}] \\ &\quad \cdot \widehat{\mathscr{D}}(u,R)\psi(x_0 - y_0; R^{-1}(x - y - (x_0 - y_0)u)).\end{aligned}$$

(NR 2) Choose $\widehat{\mathscr{H}} = \mathscr{L}^2(\mathbb{R}^3, d\xi_0\, d\xi_1\, d\xi_2; \widehat{\mathscr{K}})$ and $\widehat{V}$: $\widehat{\mathscr{H}} \to \widehat{\mathscr{H}}$ as

$$(\widehat{V}\psi)(x) := (2\pi)^{-3/2} \int d^3\xi \times \exp(-i\xi_0 x_0 + i\boldsymbol{\xi} \cdot \mathbf{x}) \widehat{\mathscr{D}}(b(\xi))\psi(\xi).$$

This operator is clearly unitary and $\widetilde{U}_g := \widehat{V}^* \widehat{U}_g \widehat{V}$ takes the form

$$\begin{aligned}(\widetilde{U}(y_0, y; u, R, \zeta)\psi)(\xi_0, \xi) &= \zeta \exp i\{(m/2)\, y \cdot u - y \cdot \xi + y_0 \xi_0\} \widehat{\mathscr{D}}(0,R) \\ &\quad \times \psi(\xi_0 - u \cdot \xi + (m/2)u^2, R^{-1}(\xi - mu)).\end{aligned}$$

(NR 3) If $\widehat{V}^*$: $\widehat{\mathscr{H}} \to \mathscr{H}$ intertwines $\widetilde{U}$ and $U \restriction G$, we conclude as before that

$$(\widetilde{V}^*\psi)(p_1, p_2, p_3) = \left|\frac{p_3}{m}\right|^{1/2} \mathscr{C}(p)^* \psi\left(\frac{p^2}{2m}, p_1, p_2\right),$$

with

$$\mathscr{C}(p)^* \widehat{\mathscr{D}}(0,R) = \mathscr{D}(R) \mathscr{C}(R^{-1}(p - mu))^*.$$

Thus, $\mathscr{C}(p)$ depends only on p_3 and intertwines $\widehat{\mathscr{D}} \restriction T$ and $\mathscr{D} \restriction T$. We check easily that $\widetilde{V}$ is contractive iff $\mathscr{C}(p_3)\mathscr{C}(p_2)^* + \mathscr{C}(-p_3)\mathscr{C}(-p_3)^* \leqslant 1$ for each p_3 and that $V = \widetilde{V}\widehat{V}$: $\widehat{\mathscr{H}} \to \mathscr{H}$ is given by the formula in the theorem.

(4) Purity of F and irreducibility of $\widehat{\mathscr{D}}$ are equivalent by the general results of Sec. II. Here $F(X)$ is supported by $[p_3 \geqslant 0]$ iff V vanishes off this subspace, i.e., $\mathscr{C}(p_3) = 0$ for $p_3 < 0$, and $F(X)$ is a projection iff V is an isometry iff in addition almost every $\mathscr{C}(p_3)$ is an isometry for $p_3 > 0$. Consider now the transformation θ of condition 2: Since it contains a time inversion and the Hamiltonian is positive, it can only be represented by an antiunitary operator $U(\theta)$, which is unique up to a phase by the irreducibility of U: $K \to \mathscr{U}(\mathscr{H})$. In both theories $(U(\theta)\psi)(p) = \mathscr{D}(\theta)\psi(\hat{\theta}p)$, where $\mathscr{D}(\theta)$ is an antiunitary operator in $\mathscr{K}$ and $\hat{\theta}(p_1, p_2, p_3) = (p_1, -p_2, p_3)$. Now suppose F is also covariant for θ. Then by proposition 1 there must be an antiunitary operator $\widehat{U}(\theta)$ on $\widehat{\mathscr{H}}$ extending the representation $\widehat{U}$ to θ, satisfying $\widehat{U}(\theta)\widehat{F}(f)\widehat{U}(\theta)^* = \widehat{F}(T_\theta f)$. These conditions imply $(\widehat{U}(\theta)\psi)(x) = \widehat{\mathscr{D}}(\theta)\psi(\theta x)$ with $\widehat{\mathscr{D}}(\theta)$ antiunitary in $\widehat{\mathscr{K}}$. Moreover V intertwines $\widehat{U}(\theta)$ and $U(\theta)$ iff $\widehat{\mathscr{D}}(\theta)\mathscr{C}(p_3) = \mathscr{C}(p_3)\mathscr{D}(\theta)$. Note that if $\widehat{\mathscr{D}}$ is irreducible, $\widehat{\mathscr{D}}(\theta)$ is also unique up to a phase. We may then pick eigenbases $\{\zeta_m\} \subset \mathscr{K}$, $\{\hat{\zeta}_m\} \subset \widehat{\mathscr{K}}$ for the generators of $\mathscr{D} \restriction T$, $\widehat{\mathscr{D}} \restriction T$ with $\mathscr{D}(\theta)\zeta_m = \zeta_m$, $\widehat{\mathscr{D}}(\theta)\hat{\zeta}_m = \hat{\zeta}_m$. Then $\mathscr{C}(p_3)\zeta_m = \mathscr{C}_m(p_3)\hat{\zeta}_m$ with $2s + 1$ functions $\mathscr{C}_m$: $\mathbb{R}^+ \to \{-1,0,1\}$.

Finally suppose that the variance of $\langle\psi, F(\cdot)\psi\rangle$ is finite. Then certain combinations of first derivatives of $\widehat{\mathscr{D}}(b(p))\mathscr{C}(p_3)\mathscr{D}(\rho(p))\psi(p)$ [resp. $\widehat{\mathscr{D}}(b(p))\mathscr{C}(p_3) \times \psi(p)$] are square integrable. In particular, this function is absolutely continuous. Condition 3 requires that this is the case for a large class of differentiable functions ψ. Since $\widehat{\mathscr{D}}(b(p))$ and $\mathscr{D}(\rho(p))$ are clearly continuous, this implies that $\mathscr{C}$ is continuous on $\mathbb{R}^+$. Since $\mathscr{C}(p_3)$ is contained in a discrete set, this function must be constant. It will be seen in Sec. VI that the properties stated in the theorem indeed imply condition 3. Q.E.D.

This theorem characterizes the direct summands of ideal screen observables up to the choice of two elements. One of these is the isometry $\mathscr{C}$: $\mathscr{K} \to \widehat{\mathscr{K}}$. Since $\mathscr{C}$ intertwines the representations of rotations and the reflection θ, it is characterized completely by the subset $\gamma \subset \{-s, -s+1, \ldots, +s\}$ of the spectrum of the generator of $\mathscr{D} \restriction T$ by which it is supported. The second object to be chosen is an irreducible representation $\widehat{\mathscr{D}}$: $H \to \mathscr{U}(\widehat{\mathscr{K}})$, which is only constrained by the condition that γ must also be contained in the spectrum $\sigma(L)$ of the generator L of $\widehat{\mathscr{D}} \restriction T$. We now briefly describe the relevant representations $\widehat{\mathscr{D}}$ of H.

In the relativistic case $H \simeq \mathrm{SL}(2,\mathbb{R})$, and $\widehat{\mathscr{D}}$ has three generators M_1, M_2, and L. With $M_\pm := \pm iM_1 + M_2$, the Lie algebra is defined by $[L, M_\pm] = \pm M_\pm$ and $[M_+, M_-] = -2L$. Both $M_\pm$ and L commute with antiunitary involution $\widehat{\mathscr{D}}(\theta)$. We may choose a basis $|n\rangle$ with $L|n\rangle = n|n\rangle$ and $\widehat{\mathscr{D}}(\theta)|n\rangle = |n\rangle$. Then

$$M_+|n\rangle = \lambda(n)|n+1\rangle, \quad M_-(n)|n\rangle = \lambda(n-1)|n-1\rangle$$

with $\lambda(n) \in \mathbb{R}$ and

$$|\lambda(n)|^2 - |\lambda(n-1)|^2 = 2n.$$

Hence $|\lambda(n)|^2 = n(n+1) + \Gamma$, where $\Gamma = M_1^2 + M_2^2 - L^2$ is the Casimir invariant. For $n \in \sigma(L)$, both $|\lambda(n)|^2$ and $|\lambda(n-1)|^2$ must be non-negative.

If the spin s and hence $\sigma(L)$ is integer, we obtain a continuous series with $\sigma(L) = \mathbb{Z}$ for $\Gamma > 0$, the trivial representation for $\Gamma = 0$, and two series with $\Gamma = -n_0(n_0 - 1)$, $n_0 \geqslant 0$, in which $\sigma(L)$ is bounded above and below, respectively. If $s \in \mathbf{Z}+\frac{1}{2}$, the discrete series is given similarly and the continuous series is obtained for $\Gamma > \frac{1}{4}$.

In the nonrelativistic case, H is the twofold covered Euclidean group, i.e., a semidirect product of $\mathbf{R}^2$ with $T \simeq \mathbf{R}/4\pi\mathbf{Z}$. We shall write the elements of H as $(\mathbf{p},\alpha)$ with $(\mathbf{p},\alpha)\,(\mathbf{p}',\alpha') = (\mathbf{p}+R_\alpha\,\mathbf{p}',\alpha+\alpha')$. Then for each $m \in \frac{1}{2}\mathbf{Z}$, $\widehat{\mathscr{D}}(\mathbf{p},\alpha) = e^{im\alpha}$ is a one-dimensional representation. For the other irreducible representations we take $\widehat{\mathscr{K}}$ to be the subspace of $\mathscr{L}^2(T,d\alpha)$ with $\psi(\alpha+2\pi) = \pm\psi(\alpha)$, where the sign distinguishes the integer and half-integer spin cases. Then we set

$$(\widehat{\mathscr{D}}(0,\alpha)\psi)(\beta) = \psi(\beta-\alpha),$$
$$(\widehat{\mathscr{D}}(\theta)\psi)(\alpha) = \overline{\psi(-\alpha)},$$

and

$$(\widehat{\mathscr{D}}(p,0)\psi)(\alpha) = \exp(im\sqrt{\Gamma}\mathbf{e}\cdot\mathbf{p})\psi(\alpha),$$

with

$$\mathbf{e}_\alpha = \begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix}.$$

The quantity $\sqrt{\Gamma}$ parametrizes the orbital radii in Mackey's construction of representations of semidirect products and is a characteristic length for the screen observable in question. This is seen in the following example.

Example: For nonrelativistic, spinless particles we have $\gamma = \{0\}$, so that we obtain a one parameter family of pure screen observables satisfying conditions 1, 2, and 3. When $\widehat{\mathscr{K}} \subset \mathscr{L}^2(T,d\alpha)$, $\mathscr{C}: \mathscr{K} \to \widehat{\mathscr{K}}$ maps to the constant function and, for $\Gamma > 0$,

$$\begin{aligned}(V_\Gamma\psi)(x,\alpha) &= (2\pi)^{-3/2}\int_{p_3>0} d^3p\,|p_3|^{1/2} \\ &\quad\times \exp i\Big\{-\frac{p^2}{2m}x_0 + i\mathbf{p}\cdot(\mathbf{x}+\sqrt{\Gamma}\mathbf{e}_\alpha)\Big\}\psi(p).\end{aligned}$$

Thus $F_\Gamma(f) = V_\Gamma^*\widehat{F}(f)V_\Gamma = F_0(M_\Gamma(f))$, where

$$M_\Gamma(f)(x_0,\mathbf{x}) = \frac{1}{2\pi}\int_0^{2\pi} d\alpha\, f(x_0,\mathbf{x}+\sqrt{\Gamma}\mathbf{e}_\alpha)$$

denotes the average of f over circles of radius $\sqrt{\Gamma}$.

Evidently, this smearing out of the observable F_0 will not change the expectation operators Z_0, Z_1, and Z_3, while strictly increasing the variance form Δ. Thus condition 4 of Sec. IV singles out F_0 as the unique ideal screen observable for nonrelativistic spinless particles. For higher spin and relativistic particles, the calculation of Δ and hence the evaluation of condition 4, is more involved and will be carried out in Sec. VI.

VI. EXPECTATION AND VARIANCE OF SCREEN OBSERVABLES

In this section we shall compute the expectation operators Z_ν and variance form $\Delta_{\nu\mu}$ $(\nu,\mu = 0,1,2)$ introduced in Proposition 3 for the screen observables characterized by Proposition 4. We have $Z_\nu = V^*\widehat{Z}_\nu V$, where $\widehat{Z}_\nu$ is the operator of multiplication by x_ν in $\widehat{\mathscr{H}} = \mathscr{L}^2(X,dx_0\,dx_1\,dx_2;\widehat{\mathscr{K}})$. It is useful to consider also the operators $\widetilde{Z}_\nu = \widehat{V}^*\widehat{Z}_\nu\widehat{V}$ in the space $\widetilde{\mathscr{H}} = \mathscr{L}^2(\Xi,d^3\xi,\widehat{\mathscr{K}})$ from the proof of Proposition 4. Here,

$$\begin{aligned}(\widehat{V}\psi)(x_0,\mathbf{x}) &= (2\pi)^{-3/2}\int d^3\xi\,\exp(-i\xi_0x_0+i\boldsymbol{\xi}\cdot\mathbf{x}) \\ &\quad\times\widehat{\mathscr{D}}(b(\xi))\psi(\xi)\end{aligned}$$

is a Fourier transform with an additional twist depending on the representation $\widehat{\mathscr{D}}$ of H. It is then easy to check that

$$\widetilde{Z}_0 = -i\frac{\partial}{\partial\xi_0} + \widetilde{Y}_0(\xi),$$

and

$$\widetilde{Z}_k = i\frac{\partial}{\partial\xi_k} + \widetilde{Y}_k(\xi)\quad (k=1,2),$$

where

$$\widetilde{Y}_k(\xi) = i\widehat{\mathscr{D}}(b(\xi))^*\frac{\partial}{\partial\xi_k}\widehat{\mathscr{D}}(b(\xi)).$$

These operators are ξ-dependent linear combinations of the generators of $\widehat{\mathscr{D}}$ in $\widehat{\mathscr{K}}$ and will be computed below. Then $Z_\nu = \widetilde{V}^*\widetilde{Z}\widetilde{V}$, where $\widetilde{V}$ is the transformation from the proof of Proposition 4. In calculating this expression, we use that the isometry $\mathscr{C}(p_3)$ determining $\widetilde{V}$ is constant, by omitting terms like $\mathscr{C}(p_3)^*(\partial/\partial p_3)\mathscr{C}(p_3)$. This is also important for calculating the variance form $\Delta_{\nu\mu}(\phi,\psi) = \langle\widetilde{Z}_\nu\widetilde{V}\phi,(1-\widetilde{V}\widetilde{V}^*)\widetilde{Z}_\mu\widetilde{V}\psi\rangle$: since $\mathscr{C}$ is constant, $i(\partial/\partial\xi_\mu)\widetilde{V}\psi$ is again in the range of $\widetilde{V}$. Therefore the contributions from the differential operators in $\widetilde{Z}_\mu$ to the variance form vanish, and

$$\Delta_{\nu\mu}(\phi,\psi) = \int d^3\xi\,\langle(\widetilde{V}\phi)(\xi),\widetilde{\Delta}_{\nu\mu}(\xi)(\widetilde{V}\phi)(\xi)\rangle_{\widehat{\mathscr{K}}},$$

where

$$\widetilde{\Delta}_{\nu\mu}(\xi) = P_\gamma\widetilde{Y}_\nu(\xi)(1-P_\gamma)\widetilde{Y}_\mu(\xi)P_\gamma,$$

and $P_\gamma = \mathscr{C}\mathscr{C}^* =$ the spectral projection of the generator of $\widehat{\mathscr{D}} \restriction T$ for the set $\gamma \subset \{-s,\ldots,+s\}$.

Consider first the slightly simpler nonrelativistic case. Then $Y_0(\xi) \equiv 0$ and Y_1 and Y_2 are independent of ξ, and equal to the boost generators for $\widehat{\mathscr{D}}$. Thus in an eigenbasis $\{|n\rangle\}$ of L, $(\pm iY_1+Y_2)|n\rangle = \sqrt{\Gamma}|n\pm 1\rangle$. Then $\widetilde{\Delta}$ is proportional to Γ. For example, if $\gamma = \{-s,\ldots,+s\}$ the commutator form is

$$\begin{aligned}(1/i)(\Delta_{12}-\Delta_{21}) &= (1/i)[Z_1,Z_2] \\ &= \Gamma\cdot(P_{+s}-P_{-s}),\end{aligned}$$

which vanishes only for $s=0$ or $\Gamma=0$. The choice $\Gamma=0$ is the only case in which the screen observable becomes variance free, and is thus demanded by condition 4 for ideal screen observables. This means that $\widehat{\mathscr{D}}$ represents boosts trivially, so that the factor $\widehat{\mathscr{D}}(b(p))$ can be omitted from the definition of V in Proposition 4. For $Z_\nu = \widetilde{V}^*\widetilde{Z}_\nu\widetilde{V}$ we then obtain

$$Z_0 = \frac{m}{i}p_3^{-1/2}\frac{\partial}{\partial p_3}p_3^{-1/2},$$
$$Z_k = i\frac{\partial}{\partial p_k} + \frac{p_k}{m}Z_0\quad (k=1,2).$$

These operators are defined on the domain $\mathscr{D}(Z) \subset \mathscr{L}^2(\mathbb{R}^2\times\mathbb{R}^+, dp^3, \mathscr{K})$ of functions ϕ such that $Z_\nu\phi \in \mathscr{H}$ and $\lim_{p_3\to 0} p_3^{-1/2}\phi(p) = 0$. Since Z_0 has defect indices $(\infty, 0)$, it is maximally symmetric. [The functions $\phi(p) = \sqrt{p_3}\exp(-p_3^2/2m)\cdot\tilde{\phi}(p_1,p_2)$ are in the defect space.] Also, Z_1 has defect indices (∞,∞) [with $\phi_\pm(p) = \sqrt{p_3}e^{\pm p_1}\tilde{\phi}(p^2/2m, p_2)$] and thus admits self-adjoint extensions. However, there is no proper extension A of Z_1 that still satisfies the covariance condition $U_x A U_x^* = A - x_1\cdot 1$ for translations (x_0, x_1, x_2) along the screen. Formally, the operators Z_ν can be obtained by substituting operators for position and momentum in the formulas $z_0 = -m(q_3/p_3)$, $z_k = q_k + (p_k/m)z_0$ for the arrival coordinates of a classical free particle at $x_3 = 0$.

In the relativistic case the calculation of $Y_\nu(\xi)$, or, what is the same thing, the calculation of the Lie algebra valued form $b(\xi)^{-1}db(\xi)$ requires a little more work and yields the following result:

$$Y_0(\xi) = \frac{-1}{\lambda^2}(\xi_1 M_1 + \xi_2 M_2),$$

$$Y_1(\xi) = \frac{-1}{\lambda}M_1 + \frac{\xi_1}{\lambda+\xi_0}Y_0(\xi) - \frac{\xi_2}{\lambda(\lambda+\xi_0)}L,$$

$$Y_2(\xi) = \frac{-1}{\lambda}M_2 + \frac{\xi_2}{\lambda+\xi_0}Y_0(\xi) + \frac{\xi_1}{\lambda(\lambda+\xi_0)}L,$$

where $\lambda = (\xi_0^2 - \xi_1^2 - \xi_2^2)^{1/2}$ and M_1, M_2, and L are the self-adjoint generators of the representation $\hat{\mathscr{D}}$ of H. Then with $M_\pm = \pm iM_1 + M_2$ and

$$\Delta_\pm = \tfrac{1}{4}\{P_\gamma M_+(1-P_\gamma)M_-P_\gamma \pm P_\gamma M_-(1-P_\gamma)M_+P_\gamma\},$$

and assuming that γ has no one-element gaps $[P_\gamma M_+(1-P_\gamma)M_+P_\gamma = 0]$:

$$\tilde{\Delta}_{\nu\mu}(\xi) = \lambda^{-4}(\xi_\nu\xi_\mu - \lambda^2 g_{\nu\mu})\cdot\Delta_+ - i\lambda^{-3}\epsilon_{\nu\mu\tau}\xi^\tau\cdot\Delta_-.$$

Independently of assumptions about γ, we have $\tilde{\Delta}_{\nu\mu}(\xi)g^{\nu\mu} = -1/2\lambda^2$, so that the variance form is predominantly spacelike. The commutators $i[Z_\nu, Z_\mu]$ are all proportional to Δ_-. The variance form depends on the characteristic parameter Γ of $\hat{\mathscr{D}}$ via the matrix elements of M_1 and M_-. The following lemma asserts that choosing Γ smaller decreases Δ in a very strong sense.

Lemma 5: Let $\gamma \subset \{-s,\ldots,+s\}$ and Δ^Γ the variance form of the relativistic screen observable characterized by γ and Γ. Then if ϕ_ν $(\nu = 0,1,2)$ are in the domain of Δ^Γ, for Γ in some interval,

$$\Gamma \mapsto \sum_{\nu,\mu=0}^{2}\Delta^\Gamma_{\nu\mu}(\phi_\nu,\phi_\mu)$$

is an increasing function of Γ.

Proof: By the above formula for Δ in terms of $\Delta(\xi)$, the assertion is equivalent to the monotonicity of

$$\sum_{\nu\mu}\langle\phi_\nu, P_\gamma Y_\nu(\xi)(1-P_\gamma)Y_\mu(\xi)P_\gamma\phi_\mu\rangle = \left|\left|\sum_\mu(1-P_\gamma)Y_\mu(\xi)P_\gamma\phi_\mu\right|\right|^2 =: \|\phi\|^2,$$

for any choice of ξ and $\phi_\nu \in \hat{\mathscr{K}}$. Since $(1-P_\gamma)LP_\gamma = 0$, we may write $\phi = (1-P_\gamma)(M_+\phi_+ + M_-\phi_-)$ with $\phi_\pm$ independent of Γ. Then

$$\|\phi\|^2 = \sum_{n\notin\gamma}|\langle n, M_+\phi_+\rangle + \langle n|M_-\phi_-\rangle|^2$$

$$= \sum_{n\notin\gamma}|(n(n-1)+\Gamma)^{1/2}\langle n-1,\phi_+\rangle + (n(n+1)+\Gamma)^{1/2}\langle n+1,\phi_-\rangle|^2.$$

We can check by differentiation that each term in this sum increases with Γ for any choice of $\langle n+1,\phi_\mp\rangle$. Q.E.D.

By definition, an ideal screen observable F is a direct sum of screen observables F_i, determined by the parameters Γ_i and γ_i. The sets γ_i must be disjoint and we may as well assume $\cup\gamma_i = \{-s,\ldots,+s\}$ or, equivalently, $F(X) =$ projection onto $[p_3>0]$. By condition 4, we have to choose each Γ_i as small as possible consistent with γ_i. For any choice of the partition $\{\gamma_i\}$ there is hence a unique ideal screen observable. It may happen that different choices of $\{\gamma_i\}$ yield the same observable: If Γ is taken to be the infimum of the values admissable for γ, the representation $\hat{\mathscr{D}}$ may become reducible. For example if $\gamma = \{-s,\ldots,+s\}$ and s is an integer, the minimal choice of Γ is $\Gamma = 0$, in which case the representation is decomposed into three parts with $L>0$, $L=0$, and $L<0$. (The critical value for half-integer spin is $\Gamma = \frac{1}{4}$.) In particular, for spin $s = 0, \frac{1}{2}$, or 1, there is only one normalized ideal screen observable.

For higher spin, condition 4 does not single out a unique partition $\{\gamma_i\}$. For example, for the partition of $\{-s,\ldots,+s\}$ into one-element sets, the variance determining operators are $\Delta_+^{(1)} = \frac{1}{2}|L|$ and $\Delta_-^{(1)} = \frac{1}{2}L$. On the other hand, for $\gamma = \{-s,\ldots,+s\}$, $\Delta_\pm^{(2)} = \frac{1}{4}(s(s+1)+\Gamma)\{|-s\rangle\langle -s| \pm |s\rangle\langle s|\}$, i.e., contributions to the variance come only from the largest and smallest eigenvalue of S_3. For $s \geqslant \frac{3}{2}$ the resulting variance forms are clearly not comparable in operator ordering. In order to characterize a unique ideal screen observable in these cases we have to impose an additional condition. The choice of the one-element partition is equivalent to the condition that each $F(f)$ commutes with S_3.

Explicit expressions for Z_ν and $\Delta_{\nu\mu}$ in the Hilbert space $\mathscr{L}^2(\mathbb{R}^3, d^3p/p_0; \mathscr{K})$ of the given representation can be assembled from the above formulas for $\hat{Y}(\xi)$ and an expression for $\tilde{V}^*(\partial/\partial\xi_\nu)$ in terms of $\partial/\partial p_\nu$ and the form $\rho(p)^{-1}d\rho(p)$. Since we did not find the result very illuminating, we shall only note the resulting differential operators in the case $s = 0$:

$$Z_0 = -\frac{ip_0}{p_3}\frac{\partial}{\partial p_3} + \frac{ip_0}{2p_3^2} = -\frac{p_0}{p_3}\circ Q_3,$$

$$Z_k = i\frac{\partial}{\partial p_k} + \frac{p_k}{p_0}Z_0 = Q_k + \frac{p_k}{p_0}\circ Z_0 \quad (k=1,2),$$

where Q_1, Q_2, Q_3 denote the Newton–Wigner position observable and $A\circ B = \frac{1}{2}(AB+BA)$, the Jordan product. Once again these expressions are formally the same as the arrival coordinates of a relativistic free classical particle.

VII. DISCUSSION

A basic interpretation rule of quantum mechanics states that to each measuring device we may associate an operator-valued measure in Hilbert space, called an "observable." However, there is no theory of measurement that would—at least in principle—allow the computation of this measure from a blueprint of the measuring apparatus. Thus, if quantum mechanics is not to be left empirically vacuous, we have to make a preliminary choice of observables at least for some basic methods of measurement. By combining a few basic observables with the quantum mechanical descriptions of motion in external fields and scattering with test particles, a fairly detailed theoretical description of many measuring devices can be developed. At a later stage in the development of the theory the initially chosen basic observables may be replaced by more realistic descriptions of actual measuring devices. The screen observables constructed above may serve as basic observables in this program.

The characterization of screen observables in Sec. IV is formulated entirely in terms of an axiomatically postulated representation of the kinematic group K. In this sense a screen observable measures a "kinematic property" of quantum particles. Other choices of covariance conditions correspond to different aspects of "quantum kinematics", e.g., to observables for position and momentum (and phase-space variables in the nonrelativistic case). In some of the sets $\mathscr{M}(U\restriction G, G/H)$ $(H\subset G\subset K)$ of covariant observables "ideal" elements may be singled out that measure the given property as sharply as quantum mechanics allows. Ideal observables satisfying different covariance conditions are usually incommensurable (e.g., positions at two different times), but as the example of phase-space observables shows, it is possible to have a joint covariant measurement of nonideal covariant observables (smeared out position and momentum). There are, however, limits to this joint measurability: If we choose $G\subset K$ and $X=G/H$ too large, $\mathscr{M}(U\restriction G,X)$ may be empty. For example, there is no phase-space observable that is covariant under the Galilei group including time translation.

In the nonrelativistic case, Wigner–Weyl quantization is a map $f\mapsto F_w(f)$ from functions on phase space to operators on Hilbert space, which is covariant under the entire affine symplectic group (including the Galilei group). Thus, by the preceding remark F_w cannot be an observable and "probabilities" calculated via F_w may indeed be negative or infinite. Since classically the coordinates of arrival at a screen can be expressed as functions on phase space, Wigner–Weyl quantization induces a screen "observable" which is not positive (hence not an observable in the sense of Sec. II) but automatically possesses the correct covariance properties. It turns out that the operator $F_w(\chi_\sigma)$ thus associated to a subset $\sigma\subset x$ of the screen is just the integral of the so-called probability current over σ. The kernel for $F_w(\chi_\sigma)$ in momentum representation contains a factor $\frac{1}{2}(p_3+p_3')$, which is clearly not positive-definite. Replacing this arithmetic mean by the geometric mean $\sqrt{|p_3p_3'|}$ and hence by a positive-definite kernel, we obtain precisely the ideal nonrelativistic screen observable constructed above.[7]

For the Wigner–Weyl quantization it is inessential whether the hypersurface describing the screen is flat. But we may also construct a positive-operator-valued measure for any hypersurface having everywhere a timelike tangent vector: The hypersurface is approximated by flat pieces on each of which an observable is defined as the Galilei/Poincaré translate of a corresponding piece of "flat" screen observable. (The covariance condition for the flat screen observable makes the result unique.) The problem with this approach is that the resulting measure need not satisfy the condition $\|F(f)\|\leqslant\|f\|$, even if the trajectories of classical free particles meet the given hypersurface at most once. This shows that even for a flat screen the measurement of an effect $F(\sigma)$, $\sigma\in X$, is not to be considered as localized near the space-time set σ but depends on the whole hyperplane X. This phenomenon has a well-known analog in the case of Newton–Wigner position observables: Since the Poincaré translates of its projections do not commute, reassembling such translates to form a measure on a curved, spacelike hypersurface necessarily leads to probabilities >1 [i.e., $\|F(f)\|\leqslant\|f\|$ fails].

Experimentally, a curved screen can be realized as a curved piece of photographic film. In order to obtain a reasonable theoretical description of such measuring devices, it is necessary to take into account that upon "first contact" with the screen the particle is absorbed or at least perturbed in its free motion. One framework in which this influence of measuring devices on the dynamics can be expressed is Davies' theory of quantum stochastic processes.[2] It would be interesting to work out a theory of screens in this context and to see its relationship to the covariant observable approach presented above.

In the relativistic case, the question of locality that arose in the discussion of curved screens suggests the following line of research: There is a natural "second quantization" procedure also for non-projection-valued observables.[17] Applying this construction to the screen observables, we obtain observables for the free quantum fields measuring the number of particles arriving at any part $\sigma\in X$ of the screen. It is easy to see that such observables are not strictly localized near σ in the sense of quantum field theory. Are these observables in some sense approximately localized? Is it possible to describe the counting of particles at a screen by an observable that is strictly localized? These problems lead back to the fundamental interpretation problem indicated at the beginning of this section: the strategy of postponing the detailed analysis of measuring devices in quantum mechanics by studying at first only their covariance properties has its analog in quantum field theory in the program of developing an interpretation of the theory in terms of localization properties alone (i.e., in terms of a net of local algebras[18]) and postponing the working out of a detailed theory of measurement. Thus quantum field theory and relativistic quantum mechanics both intrinsically contain a description of the spatio-temporal properties of physical systems. It would be interesting to see whether these descriptions can be united in a coherent view.

ACKNOWLEDGMENTS

This paper grew out of a collaboration with Heinz-Jürgen Schmidt on measurement theory in general and time observables in particular. It is a pleasure for me also to thank Alex Grossman for an invitation to the CNRS Marseille and many stimulating discussions.

[1]A. S. Wightman, "On the localizability of quantum mechanical systems," Rev. Mod. Phys. **34**, 845 (1962).

[2]E. B. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976).

[3]A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).

[4]D. P. L. Castrigiano and R. W. Henrichs, "Systems of covariance and subrepresentations of induced representations," Lett. Math. Phys. **4**, 169 (1980).

[5]U. Cattaneo, "On Mackey's imprimitivity theorem," Comment. Math. Helv. **54**, 629 (1979).

[6]J. Kijowski, "On the time operator in quantum mechanics and the Heisenberg uncertainty relations for energy and time," Rep. Math. Phys. **6**, 361 (1974).

[7]G. Ludwig, *Foundations of Quantum Mechanics II* (Springer, Berlin, 1985).

[8]G. Ludwig, *Foundations of Quantum Mechanics I* (Springer, Berlin, 1983).

[9]A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (PWN–Polish Scientific, Warsaw, 1980).

[10]H. Scutaru, "Some remarks on covariant completely positive linear maps on C^*-algebras," Rep. Math. Phys. **16**, 79 (1979).

[11]N. Bourbaki, *Topologie Générale* (Herrmann, Paris, 1971).

[12]R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964).

[13]D. P. L. Castrigiano, "On Euclidean systems of covariance for massless particles," Lett. Math. Phys. **5**, 303 (1981).

[14]R. Werner, "Quantum harmonic analysis on phase space," J. Math. Phys. **25**, 1404 (1984); S. T. Ali and E. Prugovečki, "Systems of imprimitivity and representations of quantum mechanics of fuzzy phase spaces," J. Math. Phys. **18**, 219 (1977); E. Prugovečki, *Stochastic Quantum Mechanics and Quantum Spacetime* (Reidel, Dordrecht, 1984); P. Busch, "Indeterminacy relations and simultaneous measurements in quantum theory," Int. J. Theor. Phys. **25**, 63 (1985).

[15]N. I. Akhiezer and I. U. Glazman, *Theory of Linear Operators in Hilbert Space II* (Pitman, London, 1981).

[16]R. Werner and D. P. L. Castrigiano, "On the uniqueness of the position observable for photons," Osnabrück preprint, 1985.

[17]R. Werner, "The macroscopic observable of kinetic theories," in preparation.

[18]R. Haag and D. Kastler, "An algebraic approach to quantum field theory," J. Math. Phys. **5**, 848 (1964).

The one-dimensional inverse scattering problem for an increasing potential

Gerhard Kristensson
Division of Electromagnetic Theory, Royal Institute of Technology, S-100 44 Stockholm, Sweden

(Received 20 March 1985; accepted for publication 28 October 1985)

The one-dimensional inverse scattering problem is considered for potentials that grow without limit for large values of x. The Marchenko method is established for this class of potentials, and several properties of the solution to the Schrödinger equation are developed. In the derivation of the Marchenko equation an extension of the triangularity condition is used. Some brief remarks on the relation to the inverse radial problem and the generalization to hard core potentials are made.

I. INTRODUCTION

The inverse scattering problem on the whole x axis has been analyzed extensively by several authors; for a review see, e.g., Refs. 1 and 2. The starting point is the Schrödinger equation,

$$-y'' + qy = k^2 y, \quad -\infty < x < \infty,$$

where the potential $q(x)$ is assumed real and satisfies

$$\int_{-\infty}^{\infty} (1+x^2)|q(x)|dx < \infty.$$

For such potentials the Marchenko equation can be constructed and the potential q successfully recovered from the Fourier transform of the scattering data.[1]

The solution of the inverse radial scattering problem is also well known, see Ref. 1 for a review. For s-waves this half-line problem can, at least formally, be extended to a scattering problem on the whole x axis, where the potential on the complementary half axis is defined as infinite (hard core). In this paper we consider potentials defined on the whole axis that eventually go to infinity for large values of x. Potentials of this kind have similarities with both the problems discussed above. Being defined on the whole x axis, they are, of course, connected to the line problem. However, many properties of our inverse problem have a direct counterpart in the radial problem, since we have a perfectly reflecting potential (generalization of hard core). Thus, in this sense the potentials treated here can be considered as intermediate between, and extensions of both, the full- and half-line problems, but it should be emphasized that the analogy is partly formal.

The class of potentials discussed in this paper has been studied by Kulish[3] in a short mathematical note. In this paper we develop his results further and also extend the results in several directions. We also give the proofs in detail, something that is missing in Ref. 3.

Problems and applications relevant to this class of potentials can be found in, e.g., Refs. 4–6, and references given there. Applications to the Stark effect are discussed in Ref. 4. Interesting applications are also found in solitary-wave behavior in solutions to the nonlinear Korteweg–de Vries equation.[1,4–6] We refer to these papers for more details.

We now introduce some definition and notation that are useful later on. Let L denote the Schrödinger operator

$$L = -\frac{d^2}{dx^2} + q(x),$$

where the potential $q(x)$ is a real-valued, locally integrable function and is defined on the whole axis. The potentials that are of interest in this paper belong to the class Q, which in addition to the assumption above has the following property:

$$Q = \{q \mid q \in L_0^1 \cap L_1^1, \text{ and } q(x) \to \infty, \text{ as } x \to \infty\}. \quad (1.1)$$

The space L_p^1 is

$$L_p^1 = \left\{ q \,\middle|\, \int_{-\infty}^{b} |x^p q(x)|dx < \infty, \text{ for all finite } b \right\}. \quad (1.2)$$

In some applications below we strengthen the assumptions on the potentials somewhat so that in addition $q \in L_2^1$. At any rate, these assumptions include most potentials of physical interest, e.g., piecewise continuous potentials with finite jump discontinuities.

Notice that we do not specify how the potential grows as $x \to \infty$. It can grow arbitrarily fast or slow as long as it eventually goes to infinity as $x \to \infty$. Nor do we assume the potential to be differentiable.

This class of potentials is, roughly speaking, well behaved as $x \to -\infty$, where q is "small." However, for large positive values of x, q grows beyond all limits, and serves as an impenetrable barrier. It is in fact possible to extend the class Q somewhat, so that the divergence point of q can be finite, i.e., $q(x) \to \infty$, as $x \to b < \infty$, and $q(x) \equiv \infty$ for $x > b$, provided the potential q gives a differential operator of the limit-point case (cf. Sec. II) at the singular point b. The definition of L_p^1 spaces then also has to be modified accordingly, but we do not pursue this extension any further, except for some short remarks below.

Potentials of class Q are, as we discussed above, intermediate between the full-line problem and the radial problem. The analysis presented in this paper follows the standard treatment of the Marchenko formalism quite closely. However, it is important to take a fresh look at the proofs since this class of potentials is not included in the classical inverse problems. Thus, we go through the various steps in the inverse Marchenko formalism, to see what modifications and extensions have to be made for a potential in class Q.

We define some more notations that are convenient for

 0022-2488/86/030804-12$02.50

the analysis below. Let $y(x,\lambda)$ be a solution of the Schrödinger equation

$$Ly = -y''(x,\lambda) + q(x)y(x,\lambda) = \lambda y(x,\lambda), \quad -\infty \leqslant x \leqslant \infty,$$

where λ is an arbitrary complex number. Throughout this paper a prime denotes differentiation with respect to x, and the real and imaginary parts of a complex number are indicated by indices 1 and 2, respectively, i.e., $\lambda = \lambda_1 + i\lambda_2$. This parameter λ (also denoted E in the literature) is, in suitable units, the energy in the Schrödinger equation. It is also convenient to introduce the wave number k, defined by $\lambda = k^2$, and $k = k_1 + ik_2$. For arbitrary real a and b we obtain by partial integration (a bar denotes the complex conjugate)

$$\lambda \int_a^b |y|^2\, dx = -y'\bar{y}|_{x=a}^{x=b} + \int_a^b (|y'|^2 + q|y|^2)dx. \tag{1.3}$$

The real and imaginary parts are

$$\lambda_1 \int_a^b |y|^2\, dx = -\frac{1}{2}\frac{d}{dx}|y|^2|_{x=a}^{x=b} + \int_a^b (|y'|^2 + q|y|^2)dx, \tag{1.4}$$

$$\lambda_2 \int_a^b |y|^2\, dx = -\frac{i}{2} W_x(y,\bar{y})|_{x=a}^{x=b}, \tag{1.5}$$

where the Wronskian $W_x(f,g) = f(x)g'(x) - f'(x)g(x)$ is used.

This paper is organized in sections, and each section provides an important step in the derivation of the Marchenko equation. In Sec. II we introduce a solution, well behaved at infinity, called the regular solution, which has similarities with the regular solution in the radial problem. Several properties of this solution are developed, some briefly discussed in Ref. 3, some new. We continue in Secs. III and IV by defining the Jost solution and the Jost function, and derive some of the specific properties they have for a potential in class Q. The Marchenko equation is derived in Sec. V by taking the Fourier transform of the relation between the regular solution and the Jost solution, and using the properties of the support of the Fourier transform. This can be considered as a generalization of the triangularity condition used in the standard treatment. In this context we also introduce the theory of Hardy spaces, and for the convenience of the reader we have collected some important and useful results on Hardy spaces in the Appendix. Section V also contains a uniqueness theorem. Some simple examples of the theory are given in Sec. VI.

II. THE REGULAR FUNCTION

At the beginning of this century Weyl[7] developed the theory of singular boundary value problems. The results were further extended by several authors, and for this paper the results of Hille[8,9] are the most interesting. Consider the differential equation

$$Ly = -y''(x,\lambda) + q(x)y(x,\lambda) = \lambda y(x,\lambda), \tag{2.1}$$

where $q(x)$ is real and locally integrable [not necessarily belonging to the class Q defined in Eq. (1.1)]. [More general differential equations of the type

$$-w''(x,\lambda) + P(x)w'(x,\lambda) + Q(x)w(x,\lambda) = \lambda w(x,\lambda)$$

can always be reduced to this form by the transform

$$y(x,\lambda) = w(x,\lambda)\exp\left\{-\frac{1}{2}\int^x P(s)ds\right\},$$

$$q(x) = Q(x) - P'(x)/2 + P^2(x)/4.]$$

Weyl[7] showed that the following properties hold for the solution of Eq. (2.1) restricted to $x \geqslant 0$.

(i) For every nonreal value of λ, Eq. (2.1) has at least one nontrivial solution of $L^2(0,\infty)$.

(ii) If for a particular value of λ, Eq. (2.1) has two linearly independent solutions (and hence all solutions) in $L^2(0,\infty)$, then this property holds for all values of λ, real or complex.

If the second property holds, L is said to be of the limit-circle case at infinity, otherwise L is said to be of the limit-point case at infinity. Thus in the limit-point case there exists, for every nonreal value of λ, exactly one solution to Eq. (2.1) that belongs to $L^2(0,\infty)$.

For a special kind of potential $q(x)$, the theory of Weyl can be extended somewhat. The following theorem shows the existence of a $L^2(a,\infty)$ solution for every a (see also Refs. 8 and 9), and in Theorem 3 below we collect the main result of this section.

Theorem 1: Let $q(x)$ be real and locally integrable in $[a,\infty)$, and let $q(x)\to\infty$ as $x\to\infty$. Then for all complex λ there exists one and only one linearly independent, nontrivial solution $y \in L^2(a,\infty)$ to $Ly = \lambda y$, $x \geqslant a$. Furthermore, y' and $\sqrt{|q-\lambda|}\, y \in L^2(a,\infty)$.

Notice that no boundary conditions are imposed on y at $x = a$. The theorem can be further extended so that the singular point where $q(x)\to\infty$ can be finite, as commented upon in the Introduction. The proof of this theorem has similarities to the one given by Weyl[7] and Hille,[8,9] but several extensions occur; intermediate results in the proof will be used later on in this paper, so we prefer to give the proof in detail.

Proof: The potential $q(x)$ defines an operator L of the limit-point case at infinity, see, e.g., Coddington and Levinson.[10] The uniqueness of the solution is therefore already clear by the results of Weyl [(i) and (ii) above], and to complete the proof we have to prove the existence of such a solution (in fact, only real values of λ are necessary, but for later use we treat also complex λ).

Let $y_1(x,\lambda)$ and $y_2(x,\lambda)$ be solutions of $Ly = \lambda y$ for $x \geqslant a$, satisfying the boundary conditions

$$y_1(a,\lambda) = 0, \quad y_1'(a,\lambda) = -1, \tag{2.2}$$

$$y_2(a,\lambda) = 1, \quad y_2'(a,\lambda) = 0. \tag{2.3}$$

The solutions $y_i(x,\lambda)$, $i = 1,2$, are linearly independent, since

$$W_x(y_1,y_2) = W_a(y_1,y_2) = 1 \neq 0,$$

and any solution (up to a multiplicative constant) of $Ly = \lambda y$ can be written as

$$y(x,\lambda) = y_1(x,\lambda) + my_2(x,\lambda). \tag{2.4}$$

We will now show that it is possible to choose the constant m so that $y(x,\lambda)\in L^2(a,\infty)$.

Let $\lambda=\lambda_1+i\lambda_2$ be a fixed complex number, b and c real numbers such that $a<b<c<\infty$, and choose $b=b(\lambda)$ such that

$$q(x)>\lambda_1 \text{ for all } x>b. \tag{2.5}$$

Equation (1.4) gives

$$\operatorname{Re}\{\overline{y(c,\lambda)}\,y'(c,\lambda)\}-\operatorname{Re}\{\overline{y(b,\lambda)}\,y'(b,\lambda)\} =\int_b^c [|y'|^2+(q-\lambda_1)|y||^2]dx. \tag{2.6}$$

Define

$$F(m,c)=\operatorname{Re}\{\overline{y(c,\lambda)}\,y'(c,\lambda)\}, \tag{2.7}$$

which depends on m according to Eq. (2.4), and defines a quadratic form in m. In the complex m plane, $F(m,c)=0$ defines a circle C_c:

$$|m-m_c|=r_c.$$

The center m_c and the radius r_c of the circle C_c are, after some algebra, found to be

$$m_c=-(y_1'\,\overline{y_2}+y_1\,\overline{y_2'})_{x=c}/(2\operatorname{Re}\{\overline{y_2}y_2'\}_{x=c}),$$

$$r_c=|2\operatorname{Re}\{\overline{y_2}y_2'\}_{x=c}|^{-1}.$$

Two possibilities can now occur.

(i) $\operatorname{Re}\{\overline{y_2}y_2'\}_{x=c}<0$ for all $c>b$. In this case we obtain, by means of Eq. (2.6) applied to the solution $y_2(x,\lambda)$,

$$\int_b^c [|y_2'|^2+(q-\lambda_1)|y_2|^2]dx<-\operatorname{Re}\{\overline{y_2}y_2'\}_{x=b},$$

for all $c>b$,

and, letting $c\to\infty$, we obtain that y_2' and $\sqrt{|q-\lambda_1|}\,y_2\in L^2(b,\infty)$, and they obviously also belong to $L^2(a,\infty)$. In this case the theorem is proven since $y=y_2$ is the solution satisfying the theorem.

(ii) $\operatorname{Re}\{\overline{y_2}y_2'\}_{x=c'}>0$ for some $c'>b$. Now $\operatorname{Re}\{\overline{y_2}y_2'\}$ is a monotonically increasing function for $x>b$, since $q(x)>\lambda_1$ for $x>b$, and we conclude that $\operatorname{Re}\{\overline{y_2}y_2'\}_{x=c}>0$ for $c>c'>b$. From the definition of $F(m,c)$ and $\operatorname{Re}\{\overline{y_2}y_2'\}_{x=c}>0$, for $c>c'$ we have that

$$F(m,c)\lessgtr 0\begin{Bmatrix}\text{inside}\\ \text{outside}\end{Bmatrix}C_c, \quad \textit{for } c>c'.$$

For all m values inside C_c we thus have

$$\int_b^c [|y'|^2+(q-\lambda_1)|y|^2]dx<-\operatorname{Re}\{\bar{y}y'\}_{x=b}, \quad c>c'. \tag{2.8}$$

Furthermore, we see that $C_{c_1}\subset C_{c_2}$ for $c_1>c_2>c'$, i.e., the circles C_c are nesting for increasing values of c. The family of circles C_c then converge to a single point $m(\lambda)$, as $c\to\infty$. In fact, $\operatorname{Re}\{\overline{y_2}y_2'\}_{x=c}$ is monotonically increasing for large values of c and its limit must be $+\infty$, so $r_c\to 0$, otherwise all m values inside the circle $\lim_{c\to\infty} C_c$ would give us a solution $y=y_1+my_2\in L^2(a,\infty)$, which contradicts the fact that L is of the limit-point case at infinity. For this value of $m=m(\lambda)$, both y' and $\sqrt{|q-\lambda_1|}y\in L^2(b,\infty)$, by Eq. (2.8), and they obviously also belong to $L^2(a,\infty)$, and the theorem is proved.

From the proof we see that the solution $y(x,\lambda)\in L^2(a,\infty)$ has the following property

$$\frac{d}{dx}|y(x,\lambda)|^2<0, \quad \text{for } x>b(\lambda), \tag{2.9}$$

and no other linearly independent solutions have this property.

The function $m(\lambda)$ plays a fundamental role in the analysis presented below. We now prove some important properties of this function. Equation (2.9) shows that the positive function $|y|^2$ is monotonically decreasing for sufficiently large x. Thus, $|y|^2$ must have a limit as $x\to\infty$, which necessarily has to be zero, since $y\in L^2(a,\infty)$. The function $m(\lambda)$ is thus

$$m(\lambda)=-\lim_{x\to\infty}\{y_1(x,\lambda)/y_2(x,\lambda)\}. \tag{2.10}$$

On the other hand we have

$$m(\lambda)=W_x(y_1,y)=W_a(y_1,y)=y(a,\lambda). \tag{2.11}$$

From Eq. (1.5) and Eqs. (2.2)–(2.4) we see that

$$\lambda_2\int_a^x |y|^2dx=-\frac{i}{2}W_x(y,\bar{y})+\operatorname{Im}\{m(\lambda)\}. \tag{2.12}$$

As $x\to\infty$, the left-hand side converges and thus $W_x(y,\bar{y})$ must converge as $x\to\infty$. However, this limit has to be zero, since $y\to 0$ as $x\to\infty$, as proved above, and $|y'|$ is bounded for all sufficiently large x. To see this, we use the differential equation $Ly=\lambda y$ to prove the following equation for a general $\lambda+\lambda_1+i\lambda_2$:

$$\frac{d}{dx}|y'|^2=(q-\lambda_1)\frac{d}{dx}|y|^2+2\lambda_2\operatorname{Im}\{y\overline{y'}\}.$$

Equation (2.9) gives, for $x>b(\lambda)$,

$$|y'(x,\lambda)|^2<|y'(b,\lambda)|^2+2\lambda_2\int_b^x \operatorname{Im}\{y\overline{y'}\}dt$$

$$<|y'(b,\lambda)|^2+2|\lambda_2|\int_b^x |y||y'|dt.$$

Both y and y' are in $L^2(a,\infty)$, the integral converges, and $|y'|$ must be bounded for sufficiently large x (we will actually prove below that $\lim_{x\to\infty}|y'|=0$). As a corollary to Theorem 1 we thus obtain, by letting $x\to\infty$ in Eq. (2.12),

$$\operatorname{Im}\{m(\lambda)\}=\lambda_2\int_a^\infty |y|^2dx. \tag{2.13}$$

We see that the imaginary part of $m(\lambda)$ and λ have the same sign. Furthermore, it is shown in Hille[8] that $m(\lambda)$ is a holomorphic function in the upper and lower half planes of λ and there satisfies

$$m(\bar{\lambda})=\overline{m(\lambda)}, \quad \operatorname{Im}\{\lambda\}\neq 0. \tag{2.14}$$

We can in our case say even more. The function $m(\lambda)$ is closely related to the spectral function of the following self-adjoint operator A on $[a,\infty)$:

$$Ay=Ly, \quad y'(a,\lambda)=0. \tag{2.15}$$

Hille[8] shows that the spectrum of A consists only of a point spectrum of isolated points $\{\lambda_n\}$, bounded from below, and with no limit point in the finite λ plane. All λ_n's are, of course, real. Furthermore, Hille[8] shows that the solutions $y_i(x,\lambda)$ [and $y_i'(x,\lambda)$], $i = 1,2$, for any fixed value of x, are entire functions of λ of order $\frac{1}{2}$ and of finite type (normal type of this order).

We now state an important property of $m(\lambda)$. For the proof we refer to Hille.[8]

Theorem 2: The function $m(\lambda)$ is a meromorphic function with simple poles at $\{\lambda_n\}$. It satisfies Eq. (2.14) for $\lambda \notin \{\lambda_n\}$. The function $m(\lambda)$ has the representation

$$m(\lambda) = \sum_n \frac{\sigma_n}{\lambda - \lambda_n}, \tag{2.16}$$

where the sum converges absolutely for all $\lambda \notin \{\lambda_n\}$, and where the residues σ_n are real and nonpositive for all n. Here σ_n are the jumps in the spectral function corresponding to the self-adjoint operator A in Eq. (2.15).

From Eqs. (2.1)–(2.4) it is easy to obtain

$$m(\mu) - m(\lambda) = (\mu - \lambda) \int_a^\infty y(x,\mu)\, y(x,\lambda) dx,$$

and we get as $\mu \to \lambda$

$$\frac{d}{d\lambda} m(\lambda) = \int_a^\infty y^2(x,\lambda) dx.$$

The solution $y(x,\lambda) \in L^2(a,\infty)$ constructed in Theorem 1 and given by Eq. (2.4) with $m = m(\lambda)$ has a unique continuation to the whole real axis $(-\infty, \infty)$. Furthermore, it is a meromorphic function of λ and its only singularities are simple poles at $\{\lambda_n\}$, which have no limit point in the finite λ plane. These poles are, of course, not of any importance in our scattering problem, but have their origin from the construction of the solution $y(x,\lambda)$. However, we can remove these poles by multiplying the solution $y(x,\lambda)$ with an entire function with simple zeros at $\{\lambda_n\}$, and nowhere else. The existence of such an entire function is given by Weierstrass's factorization theorem.[11] In the following theorem we collect the main result of this section.

Theorem 3: For every real a and every complex λ, there exists a solution $\phi \in L^2(a,\infty)$ to $L\phi = \lambda\phi$, where q satisfies the assumptions in Theorem 1. It is possible to choose this solution such that $\phi(x,\lambda)$ and $\phi'(x,\lambda)$ for each fixed x are entire functions of λ; $\phi(x,\lambda)$ is unique up to a multiplication with an entire function of λ without zeros, and ϕ can be chosen real for real λ. The $\phi(x,\lambda)$ is called the regular solution of the scattering problem.

The regular solution has a number of important properties, some of which will be useful later on in this paper. First we prove the following lemma.

Lemma 1: Let $\phi(x,\lambda)$ be the regular solution found in Theorem 3 and assume that a is chosen such that $q(t) \geqslant 0$ for $t \geqslant a$. Then $\phi'(a,\lambda) + ik\phi(a,\lambda)$ has no zeros as a function of k in the closed upper half plane of k, i.e., $k_2 \geqslant 0$ $(\lambda = k^2)$.

Proof: Assume that k is a zero. Then $\phi'(a,k^2) = -ik\phi(a,k^2)$, and by use of Eqs. (1.4) and (1.5) we obtain

$$k_2|\phi(a,k^2)|^2 = -\int_a^\infty [|\phi'|^2 + (q - \lambda_1)|\phi|^2] dx,$$

$$k_1|\phi(a,k^2)|^2 = -\lambda_2 \int_a^\infty |\phi|^2 dx,$$

since $\lim_{x\to\infty} \phi'\bar{\phi} = 0$ as proved above. The second equation implies that $k_1 = 0$, since $k_2 \geqslant 0$. The first equation is then simplified to

$$k_2|\phi(a,k^2)|^2 = -\int_a^\infty [|\phi'|^2 + (q + (k_2)^2)|\phi|^2] dx.$$

However, this equation cannot be satisfied for any nontrivial ϕ since $q(x) \geqslant 0$ for $x \geqslant a$; we have a contradiction, and the lemma is proved.

The next lemma is a simple consequence of Theorem 2.

Lemma 2: Let A be the operator defined in Eq. (2.15) and let the spectrum of A, $\sigma(A) = \{\lambda_n\}_{n=1}^\infty$, be ordered such that

$$\lambda_1 \leqslant \lambda_2 \leqslant \cdots.$$

Then $\mathrm{Im}\{km(\lambda)\} \geqslant 0$ for all k, provided $k_2 \geqslant 0$, and $|\lambda| \geqslant -\lambda_1$. (Do not confuse the first two points of the spectrum λ_1 and λ_2 with the real and imaginary parts of the complex number λ.)

Proof: Theorem 2 provides us with a representation of $m(\lambda)$ in terms of the spectral function

$$m(\lambda) = \sum_{n=1}^\infty \frac{\sigma_n}{\lambda - \lambda_n},$$

where $\{\lambda_n\}$ are the values of the point spectrum of A and $\{\sigma_n\}$ are all nonpositive numbers. Simple calculations give $(\lambda = k^2)$

$$\mathrm{Im}\{km(\lambda)\} = -k_2 \sum_{n=1}^\infty \sigma_n \frac{\lambda_n + |\lambda|}{|\lambda_n - \lambda|^2} \geqslant 0,$$

$$\text{for all } k_2 \geqslant 0,\ |\lambda| \geqslant -\lambda_1,$$

and the lemma is proved.

Note: The result of the lemma is in fact valid for a larger class of self-adjoint operators A for which $y'(a,\lambda) = 0$, namely those operators that have a spectrum that is bounded from below. In such cases the representation of $m(\lambda)$ is

$$m(\lambda) = \int \frac{d\rho(t)}{t - \lambda},$$

and the lemma can be proved in analogy with the proof above.

In the rest of the paper we assume for simplicity that $\lambda_1 \geqslant 0$. There is in fact no loss of generality in this assumption, and an analogous treatment can be made for $\lambda_1 < 0$, but the details become more complicated.

Lemma 3: The regular solution $\phi(x,\lambda)$ of Theorem 3 satisfies

$$|ik\phi(a,\lambda) - \phi'(a,\lambda)|/|ik\phi(a,\lambda) + \phi'(a,\lambda)| \leqslant 1, \tag{2.17}$$

for all $k_2 \geqslant 0$. Here a is any number such that $q(t) \geqslant 0$, for $t \geqslant a$.

Proof: From Lemma 1 we see that the denominator of Eq. (2.17) is never zero, so the quotient is well defined. Furthermore, we see from Eqs. (2.2)–(2.4), and the construction of the regular solution $\phi(x,\lambda)$, that

$$m(\lambda) = -\phi(a,\lambda)/\phi'(a,\lambda). \tag{2.18}$$

From Lemma 2 we have $\mathrm{Im}\{km(\lambda)\} \geqslant 0$, or equivalently

$$|-ikm(\lambda) - 1|/|-ikm(\lambda) + 1| \leqslant 1.$$

However, this inequality is equivalent to Eq. (2.17), and the proof of the lemma is completed.

Lemma 4: The regular solution $\phi(x,\lambda)$ of Theorem 3 satisfies

$$|ik\phi(x,\lambda)e^{-ikx}|/|[ik\phi(a,\lambda)+\phi'(a,\lambda)]e^{-ika}|<e^{K\theta(x)},$$

for all $x<a$, $k_2>0$. Here K is a numerical constant,

$$\theta(x)=\int_x^a (t-x)|q(t)|dt,$$

and a is any number such that $q(t)>0$ for $t>a$.

Proof: The regular solution $\phi(x,\lambda)$ satisfies the Volterra equation

$$\phi(x,\lambda)=\phi(a,\lambda)\cos k(a-x) - \phi'(a,\lambda)[\sin k(a-x)]/k + \int_x^a \frac{\sin k(t-x)}{k} q(t)\phi(t,\lambda)dt. \quad (2.19)$$

Define

$$h(x,\lambda)=ik\phi(x,\lambda)e^{ik(a-x)}/[ik\phi(a,\lambda)+\phi'(a,\lambda)]. \quad (2.20)$$

Then $h(x,\lambda)$ satisfies

$$h(x,\lambda)=\frac{1}{2}\left(1+e^{2ik(a-x)}\frac{ik\phi(a,\lambda)-\phi'(a,\lambda)}{ik\phi(a,\lambda)+\phi'(a,\lambda)}\right) + \int_x^a D_k(t-x)q(t)h(t,\lambda)dt,$$

where

$$D_k(x)=(1/2ik)\,(e^{2ikx}-1). \quad (2.21)$$

For $x<a$ and $k_2>0$, we have by Lemma 3

$$|h(x,\lambda)|<1+\int_x^a |D_k(t-x)|\,|q(t)|\,|h(t,\lambda)|dt. \quad (2.22)$$

Using the estimate

$$|D_k(x)|<Kx/(1+|k|x)<Kx, \quad \text{for } k_2>0, x>0, \quad (2.23)$$

where K is an appropriate constant independent of x and k, we obtain by iteration

$$|h(x,\lambda)|<\exp\left\{K\int_x^a (t-x)|q(t)|dt\right\},$$

and the lemma is proved.

We close this section by proving an additional property of the regular solution.

Theorem 4: Let $\phi(x,\lambda)$ be the solution of Theorem 3 and assume that the potential q belongs to class Q. Then the integral

$$\int_{-\infty}^{\infty} q(x)\phi(x,\lambda)e^{-ikx}dx,$$

is convergent for all fixed $\lambda=k^2$, such that $k_2>0$, and holomorphic in $k_2>0$.

Proof: We define a real positive function

$$g(x,\lambda)=|u'(x,\lambda)+2iku(x,\lambda)|^2 = e^{2k_2x}|\phi'(x,\lambda)+ik\phi(x,\lambda)|^2,$$

where

$$u(x,\lambda)=\phi(x,\lambda)e^{-ikx}.$$

Here $u(x,t)$ satisfies

$$-u''-2iku'+qu=0. \quad (2.24)$$

Simple calculations give

$$g'(x,\lambda)=q(x)e^{2k_2x}\left(\frac{d}{dx}|\phi(x,\lambda)|^2-2k_2|\phi(x,\lambda)|^2\right).$$

The solution ϕ satisfies Eq. (2.9) and we have

$$g'(x,\lambda)<0, \text{ for } x>a, \text{ and } k_2>0,$$

where a is defined such that $a>b(\lambda)$ in Eq. (2.5) and $q(x)>0, x>a$. The positive function $g(x,\lambda)$ is thus monotonically decreasing, the limit $\lim_{x\to\infty} g(x,\lambda)$ must exist, and we show that this limit has to be zero. For real values of k this is easy to see. Then $|u|=|\phi|$ and $|\phi|\to 0$ as $x\to\infty$, since both ϕ and $\phi'\in L^2(a,\infty)$. Thus $|\phi'|$ converges and this limit must be zero, otherwise $\phi'\notin L^2(a,\infty)$.

From the differential equation in Eq. (2.24) we obtain the following integral expression by partial integration:

$$\int_a^b (|u'|^2-2ik\bar{u}u'+q|u|^2)dx=\bar{u}u'|_{x=a}^{x=b},$$

and we choose a as above. The real and imaginary parts are

$$\int_a^b (|u'|^2+q|u|^2+2k_1\,\mathrm{Im}\{\bar{u}u'\})dx = \left\{\frac{1}{2}\frac{d}{dx}|u|^2-k_2|u|^2\right\}\Big|_{x=a}^{x=b}, \quad (2.25)$$

$$2k_2\int_a^b \mathrm{Im}\{\bar{u}u'\}dx=\{k_1|u|^2+\mathrm{Im}\{\bar{u}u'\}\}|_{x=a}^{x=b}. \quad (2.26)$$

For real k we have already proven that $\lim_{x\to\infty} g(x,\lambda)=0$, so assume $k_2>0$, and eliminate the integral over $\mathrm{Im}\{\bar{u}u'\}$, to obtain

$$k_2\int_a^b (|u'|^2+q|u|^2)dx = e^{2k_2x}\left(\frac{1}{2}k_2\frac{d}{dx}|\phi|^2-k_1\,\mathrm{Im}\{\bar{\phi}\phi'\}\right)\Big|_{x=a}^{x=b}.$$

This equation can be simplified by using [cf. the derivation of Eq. (2.13)]

$$\lambda_2\int_b^\infty |\phi|^2dx=\mathrm{Im}\{\bar{\phi}\phi'\}_{x=b}.$$

We get

$$k_2\int_a^b (|u'|^2+q|u|^2)dx = \frac{1}{2}k_2e^{2k_2x}\frac{d}{dx}|\phi|^2|_{x=a}^{x=b}+k_1\,\mathrm{Im}\{\bar{\phi}(a)\phi'(a)\}e^{2k_2a} - 2k_1^2k_2e^{2k_2b}\int_b^\infty |\phi|^2dx.$$

The integral on the left-hand side either converges to a finite limit or diverges to $+\infty$, as $b\to\infty$, for every potential q in class Q. The right-hand side, however, cannot diverge to $+\infty$ for $k_2>0$ as $b\to\infty$, due to Eq. (2.9), and we conclude that both u' and $u\sqrt{|q|}\in L^2(a,\infty)$, and, hence also u. Since both u and $u'\in L^2(a,\infty)$, $\lim_{x\to\infty}|u|=0$, and again, since $g(x,\lambda)$ has a limit, $|u'|$ must converge as $x\to\infty$, and this limit must be zero, since $u'\in L^2(a,\infty)$.

We collect the conclusions made above and find that

$$\int_a^b q(x)\phi(x,\lambda)e^{-ikx}\,dx$$
$$= [u'(x,\lambda) + 2iku(x,\lambda)]\big|_{x=a}^{x=b}$$

converges as $b\to\infty$, and the first part of the theorem is proven.

The remaining part of the proof is to show that the integral over the interval $(-\infty,a]$ is convergent. With the same notation as in the proof of Lemma 4 and the estimate Eq. (2.23) we get

$$|h(x,\lambda)|\leqslant 1 + K\int_x^a (t-x)|q(t)|\,|h(t,\lambda)|dt.$$

There is no loss of generality assuming $a>0$ and by use of Lemma 4 we get

$$\begin{aligned}|h(x,\lambda)| &\leqslant 1 + K\int_0^a t\,|q(t)|\,|h(t,\lambda)|dt \\ &\quad + K|x|\int_x^a |q(t)|\,|h(t,\lambda)|dt \\ &\leqslant 1 + Ke^{K\theta(0)}\int_0^a t\,|q(t)|dt \\ &\quad + K|x|\int_x^a |q(t)|\,|h(t,\lambda)|dt \\ &\leqslant C + K|x|\int_x^a |q(t)|\,|h(t,\lambda)|dt,\end{aligned}$$

where the constant C is independent of x and λ, but depends on a and q. Define $H(x,\lambda) = |h(x,\lambda)|/[C(1+|x|)]$ and $Q(x) = K(1+|x|)|q(x)|$. We obtain

$$H(x,\lambda)\leqslant 1 + \int_x^a Q(t)\,H(t,\lambda)dt.$$

Iteration gives

$$H(x,\lambda)\leqslant\exp\left\{\int_x^a Q(t)dt\right\}\leqslant\exp\left\{\int_{-\infty}^a Q(t)dt\right\}.$$

For a potential in class Q we thus get

$$|\phi(x,\lambda)e^{-ikx}|\leqslant K'(1+|x|),\quad \text{for } k_2\geqslant 0,\quad x\leqslant a,$$

where the constant K' is independent of x, but depends on ϕ, a, and k. Since $q\in L_0^1\cap L_1^1$ we can conclude that $\int_{-\infty}^a q\phi e^{-ikx}\,dx$ is finite and the integral $\int_{-\infty}^{\infty} q\phi e^{-ikx}\,dx$ is convergent. The holomorphic properties now easily follow from above and the theorem is proved for all potentials in class Q.

We see that in the integrand both the potential and the exponential function are increasing functions of x. However, the regular solution ϕ compensates this increase by a decrease, as $x\to\infty$, so that the integral in Theorem 4 converges.

III. THE JOST SOLUTION

In the preceding section the regular solution of $Ly=\lambda y$ was investigated. This solution has the property of being well behaved at large positive values of x. The Jost solution is instead well behaved at large negative values of x. For the Jost solution it is more convenient to let the dependence of the parameter λ be in terms of the wave number k $(\lambda = k^2)$.

For convenience we introduce three positive, monotonically increasing functions $\alpha(x)$, $\beta(x)$, and $\gamma(x)$ defined as

$$\alpha(x) = \int_{-\infty}^x |q(t)|dt, \tag{3.1}$$

$$\beta(x) = \int_{-\infty}^x (x-t)|q(t)|dt, \tag{3.2}$$

$$\gamma(x) = \int_{-\infty}^x (1+|t|)|q(t)|dt. \tag{3.3}$$

The main result of this section is collected in the following theorem.

Theorem 5: For each k, $k_2\geqslant 0$, there exists a unique solution $f(x,k)$, the Jost solution, to the differential equation $Lf = k^2 f$, where q is in class Q, such that f satisfies the boundary condition

$$\lim_{x\to-\infty} f(x,k)e^{ikx} = 1. \tag{3.4}$$

$f(x,k)$ satisfies the integral equation

$$f(x,k) = e^{-ikx} + \int_{-\infty}^x \frac{\sin k(x-t)}{k}\,q(t)f(t,k)dt. \tag{3.5}$$

For each fixed x, $f(x,k)$ and $f'(x,k)$ are holomorphic in $k_2>0$ and continuous in $k_2\geqslant 0$, and $f(x,k)$ satisfies

$$\overline{f(x,k)} = f(x,-\bar{k}),\quad \text{for } k_2\geqslant 0. \tag{3.6}$$

Furthermore, $f(x,k)$ satisfies

$$|f(x,k)e^{ikx} - 1|\leqslant[\alpha(x)/|k|]e^{\alpha(x)/|k|},\quad k_2\geqslant 0,\quad k\neq 0, \tag{3.7}$$

$$|f(x,k)e^{ikx} - 1|\leqslant K'[1+\max(0,x)] \times e^{K\beta(x+1)}\gamma(x)/(1+|k|),\quad k_2\geqslant 0, \tag{3.8}$$

$$|f'(x,k)e^{ikx} + ik|\leqslant K''[1+\max(0,x)] \times e^{K\beta(x+1)}\gamma^2(x),\quad k_2\geqslant 0, \tag{3.9}$$

where K, K', and K'' are appropriate constants independent of x and k. We also have that

$$\frac{d}{dk}f(x,k) \equiv \dot{f}(x,k)$$

exists for all $k_2\geqslant 0$, $k\neq 0$, and that $k\dot{f}(x,k)$ is continuous in $k_2\geqslant 0$ with $\lim_{k\to 0} \dot{f}(x,k) = 0$. If furthermore, $q\in L_2^1$ then $\dot{f}(x,k)$ exists and is continuous at $k=0$.

Proof: The proof of this theorem follows closely the proof of the corresponding results for the full-line problem with more well-behaved potentials given by Deift and Trubowitz.[2] We refer to their paper for details, and give here only details of the proof where alterations from Ref. 2 are needed. We thus refer to Ref. 2 concerning the existence and uniqueness properties of f, as well as its holomorphic properties.

By standard iteration of the Volterra equation in Eq. (3.5) we easily get Eq. (3.7) by means of Eq. (2.21) and the estimate

$$|D_k(x)|\leqslant 1/|k|,\quad k_2\geqslant 0,\quad k\neq 0,\quad x\geqslant 0.$$

To derive Eq. (3.8) we start by iterating the Volterra equation using the estimate Eq. (2.23). We get

$$|f(x,k)e^{ikx}|<\exp\{K\beta(x)\}. \quad (3.10)$$

By means of the Volterra equation, Eq. (2.23), and Eq. (3.10), we now obtain for all $x>0$

$$\begin{aligned}|f(x,k)e^{ikx}-1| &< K\int_{-\infty}^{x}(x-t)|q(t)|\,|f(t,k)e^{ikt}|dt\\ &< Kx\int_{-\infty}^{x}|q(t)|\,|f(t,k)e^{ikt}|dt\\ &\quad +K\int_{-\infty}^{0}(-t)|q(t)|\,|f(t,k)e^{ikt}|dt\\ &< Kx\,e^{K\beta(x)}\alpha(x)+C_1,\end{aligned}$$

for some appropriate choice of the constant C_1, which depends on q, but not on x and k. For $x<0$ we similarly have some constant C_2 such that

$$|f(x,k)e^{ikx}-1|<C_2\int_{-\infty}^{x}(-t)|q(t)|dt.$$

Equation (3.8) follows from these inequalities and Eq. (3.7), and $\alpha(x)<\beta(x+1)$.

The estimate Eq. (3.9) can be obtained from the identity

$$f'(x,k)e^{ikx}+ikf(x,k)e^{ikx}=\int_{-\infty}^{x}q(t)\,f(t,k)e^{ik(2x-t)}\,dt,$$

which follows from Eq. (3.5). By use of Eq. (3.8) we get, for $k_2>0$,

$$\begin{aligned}&|f'(x,k)e^{ikx}+ik|\\ &\quad<|k|\,|f(x,k)e^{ikx}-1|\\ &\qquad+\int_{-\infty}^{x}|q(t)|\,|f(t,k)e^{ikt}-1|dt+\alpha(x)\\ &\quad<K''[1+\max(0,x)]e^{K\beta(x+1)}\gamma(x)\\ &\qquad+K''\int_{-\infty}^{x}\frac{(1+|t|)|q(t)|e^{K\beta(t+1)}\gamma(t)dt}{1+|k|},\end{aligned}$$

from which we obtain Eq. (3.9).

What remains to be proved are the properties of $f(x,k)$. However, they are quite similar to Ref. 2, and we do not repeat the details since the generalization is obvious, and the theorem is proved.

From Eq. (3.8) we see that $f(x,k)e^{ikx}-1\in L^2(-\infty,\infty)$ as a function of k_1, for each fixed value of x, and $k_2>0$. Furthermore, $f(x,k)e^{ikx}-1\in H^2$ for each fixed x. For the convenience of the reader we have collected some important results on the Hardy space H^2 in the Appendix. From Theorem A.1 we find that there exists an $L^2(-\infty,\infty)$-function $A(x,t)$ of t, with support in $(-\infty,x]$, for each x, and we have the representation

$$f(x,k)=e^{-ikx}+\int_{-\infty}^{x}A(x,t)e^{-ikt}\,dt. \quad (3.11)$$

For real k the Wronskian between the two solutions $f(x,\pm k)$ is

$$W_x(f(x,k),f(x,-k))=2ik,\quad k \text{ real}, \quad (3.12)$$

so for real $k\neq 0$, we see that $f(x,\pm k)$ are two linearly independent solutions of $Lf=k^2f$.

IV. THE JOST FUNCTION

We define the Jost function $F(k)$ as usual:

$$F(k)=W_x(\phi(x,\lambda),f(x,k)). \quad (4.1)$$

This function is holomorphic in $k_2>0$ and continuous in $k_2\geq 0$ as seen from the analysis in Secs. II and III.

Theorem 6: The Jost function $F(k)$ defined in Eq. (4.1) has the following properties for a potential q in class Q.

(i) $\overline{F(k)}=F(-\bar{k}),\quad k_2\geq 0. \quad (4.2)$

(ii) $F(k)=\int_{-\infty}^{\infty}q(t)\phi(x,\lambda)e^{-ikt}\,dt,\quad k_2\geq 0. \quad (4.3)$

(iii) The only roots of $F(k)$ in $k_2\geq 0$ are simple and purely imaginary (except possibly $k=0$).

(iv) $F(k)$ has a root in $k_2>0$ if and only if $\phi(x,\lambda)$ is in $L^2(-\infty,\infty)$.

(v) For every a,

$$F(k)=-e^{-ika}[ik\phi(a,\lambda)+\phi'(a,\lambda)][1+O(1/|k|)],\quad k_2\geq 0. \quad (4.4)$$

Proof: Property (i) is a simple consequence of the fact that ϕ is even in k (remember $\lambda=k^2$) and real for real λ, and Eq. (3.6).

To prove Eq. (4.3) we assume for a moment that k is real $\neq 0$. Equations (3.12) and (4.1) and the fact that ϕ is even in k and real give

$$2ik\phi(x,\lambda)=F(-k)f(x,k)-F(k)f(x,-k),\quad k \text{ real}. \quad (4.5)$$

Insert Eq. (3.5) and use Theorem 4

$$\begin{aligned}2ik\phi(x,\lambda) &=F(-k)e^{-ikx}-F(k)e^{ikx}\\ &\quad+2ik\int_{-\infty}^{x}\frac{\sin k(x-t)}{k}q(t)\phi(t,\lambda)dt\\ &=e^{-ikx}\Big\{F(-k)-\int_{-\infty}^{\infty}q(t)\phi(t,\lambda)e^{ikt}dt+o(1)\Big\}\\ &\quad-e^{ikx}\Big\{F(k)-\int_{-\infty}^{\infty}q(t)\phi(t,\lambda)e^{-ikt}dt+o(1)\Big\},\\ &\quad\text{as } x\to\infty.\end{aligned}$$

Since both ϕ and ϕ' go to zero as $x\to\infty$, Eq. (4.3) follows for real $k\neq 0$. However, the integral $\int_{-\infty}^{\infty}q(t)\phi(t,\lambda)e^{-ikt}dt$ also has a unique analytic continuation in the upper half plane of k as shown in Theorem 4, and $F(k)$ is continuous on the real axis; thus we have proved (ii).

To prove (iii) we assume that $F(k)=0$. First assume k is real $\neq 0$. Then by Eqs. (4.2) and (4.5), $\phi(x,k^2)\equiv 0$, which is the trivial case. Thus the root cannot be real $\neq 0$. Now assume $k_2>0$, and use Eq. (1.5) with $y(x,\lambda)=f(x,k)$. We get

$$2k_1k_2\int_{-\infty}^{a}|f(x,k)|^2\,dx=-(i/2)\{f(a,k)\,\overline{f'(a,k)}-f'(a,k)\,\overline{f(a,k)}\},$$

since the Wronskian vanishes for the lower limit for $k_2>0$. We also have from Eqs. (4.1) and (2.18) that

$$F(k) = -\phi'(a,k^2)\{m(k^2)f'(a,k)+f(a,k)\}=0,$$

and since $\phi'(a,k^2)\neq 0$, for $k_2>0$ [see the discussion below Eq. (2.15)], we have

$$2k_1k_2\int_{-\infty}^{a}|f(x,k)|^2dx = -|f'(a,k)|^2\,\mathrm{Im}\{m(k^2)\}. \tag{4.6}$$

Equation (2.13) shows that $\mathrm{Im}\{m(k^2)\}$ has the same sign as λ_2, i.e., $\mathrm{Im}\{m(k^2)\}\leq 0$, for $k_1\leq 0$, $k_2>0$. Equation (4.6) then gives that $k_1=0$, and all roots are purely imaginary. We now prove that all roots of $F(k)$ in $k_2>0$ must be simple. Assume that k is a root of F. Then according to Eq. (4.1), ϕ and f are linearly dependent and there exists a constant $A\neq 0$, such that

$$f(x,k)=A\,\phi(x,k^2). \tag{4.7}$$

We also use the following identity:

$$\frac{d}{dx}W_x(\dot\phi,f)=\frac{d}{dx}W_x(\dot f,\phi)=2kf(x,k)\phi(x,k^2).$$

We can now compute $(d/dk)\,F(k)=\dot F(k)$ at the root k:

$$\dot F(k)=W_x(\dot\phi,f)+W_x(\phi,\dot f)$$
$$=2kA\int_{-\infty}^{\infty}\phi^2(x,k^2)dx\neq 0, \tag{4.8}$$

where we have used Eq. (4.7), the derived properties of f as $x\to-\infty$ and ϕ as $x\to\infty$, and we conclude that the roots of F are simple; thus (iii) is proved.

Assume that k is a root of F such that $k_2>0$. For this value of k Eq. (4.7) holds. On the left-hand side of Eq. (4.7) there is a function $\in L^2(-\infty,a)$, on the right-hand side a function $\in L^2(a,\infty)$; thus $\phi(x,k^2)\in L^2(-\infty,\infty)$. We prove the converse by assuming $F(k)\neq 0$, $k_2>0$. Then the kernel of the resolvent

$$G(x,x')=f(x_<,k)\phi(x_>,k^2)/F(k),$$

where $x_<\ (x_>)=\min(\max)(x,x')$ is well defined. For any $g\in L^2(-\infty,\infty)$ we have

$$[(L-\lambda)^{-1}g](x)$$
$$=\int_{-\infty}^{\infty}G(x,x')g(x')dx'$$
$$=F^{-1}(k)\left\{\phi(x,\lambda)\int_{-\infty}^{x}f(x',k)g(x')dx'\right.$$
$$\left.+f(x,k)\int_{x}^{\infty}\phi(x',\lambda)g(x')dx'\right\},$$

and no nontrivial $L^2(-\infty,\infty)$ function exists satisfying $Ly=\lambda y$.

We also show that for $k=0$ there does not exist any eigenfunction, i.e., $k=0$ is never in the point spectrum of the self-adjoint operator L on $(-\infty,\infty)$ without boundary conditions. This is easy to see, since we have $f(x,0)\to 1$, as $x\to-\infty$; a second solution is given by

$$f(x,0)\int^{x}[f(t,0)]^{-2}dt\to x,\quad \text{as } x\to-\infty;$$

and $Ly=0$ has no nontrivial solution in $L^2(-\infty,\infty)$.

Finally, we prove (v) simply by using Theorem 5:

$$F(k)=\phi(a,\lambda)e^{-ika}[f'(a,k)e^{ika}+ik]$$
$$-\phi'(a,\lambda)e^{-ika}[f(a,k)e^{ika}-1]$$
$$-e^{-ika}\{ik\phi(a,\lambda)+\phi'(a,\lambda)\}$$
$$=-e^{-ika}\{ik\phi(a,\lambda)+\phi'(a,\lambda)\}$$
$$\times\{1+O(1/|k|)\},$$

and (v) and the theorem are proved.

Note that by Lemma 1, for large enough $|k|$ (and a proper choice of a), $F(k)$ cannot be zero, and there are at most a finite number of roots to $F(k)$ in the upper half plane of k. We can rephrase this by saying that the self-adjoint operator L on $L^2(-\infty,\infty)$ (no boundary conditions) has only a finite number of eigenvalues.

The fundamental relation between the solutions of the scattering problem given by Eq. (4.5) is more conveniently written as

$$\psi(x,k)=f(x,-k)-R(k)f(x,k),\quad k \text{ real}, \tag{4.9}$$

where we have defined the scattering solution ψ and the reflection coefficient R for real k as

$$\psi(x,k)=-2ik\phi(x,k^2)/F(k), \tag{4.10}$$

$$R(k)=F(-k)/F(k). \tag{4.11}$$

Equation (4.9) has, of course, the physical interpretation that the solution $\psi(x,k)$ consists of two parts, one incoming wave, represented by $f(x,-k)$, and one reflected wave, given by $-R(k)f(x,k)$.

From the definition of ψ in Eq. (4.10) we immediately see that ψ can be continued analytically into the upper half plane of k and thus ψ is meromorphic in $k_2>0$ with simple poles at the roots of F in $k_\nu=i\beta_\nu$, $\beta_\nu>0$, $i=1,\dots,n$ (n bound states), and the residues are [use Eqs. (3.11), (4.7), and (4.8)]

$$\mathrm{Res}\{\psi(x,k)\}_{k=i\beta_\nu}=\frac{-if(x,i\beta_\nu)}{\int_{-\infty}^{\infty}f^2(t,i\beta_\nu)dt}$$
$$=-if(x,i\beta_\nu)\,M_\nu$$
$$=-i\left(e^{\beta_\nu x}+\int_{-\infty}^{x}A(x,t)e^{\beta_\nu t}dt\right)M_\nu. \tag{4.12}$$

Since the left-hand side of Eq. (4.9) is a meromorphic function in $k_2>0$, there must be a cancellation of the singularities in the right-hand side of the equation, since the right-hand side, in general, is not defined for these complex k.

The reflection coefficient $R(k)$ satisfies

$$|R(k)|=1,\quad k \text{ real}, \tag{4.13}$$

due to Eq. (4.2), which is nothing but energy conservation in our problem, and the fact that we have an impenetrable barrier. From the definition of $R(k)$, we see that $R(k)$ is in general only defined for real k and for such k we have, due to Theorem 6,

$$R(k)=e^{2ika}\frac{\phi'(a,\lambda)-ik\phi(a,\lambda)}{\phi'(a,\lambda)+ik\phi(a,\lambda)}\left\{1+O\left(\frac{1}{|k|}\right)\right\}$$
$$=R_\infty(k)+R_2(k),\quad k \text{ real}, \tag{4.14}$$

where R_∞ is defined as

$$R_\infty = e^{2ika}\frac{\phi'(a,\lambda) - ik\phi(a,\lambda)}{\phi'(a,\lambda) + ik\phi(a,\lambda)}. \tag{4.15}$$

In this expression a acts as a parameter.

However, from what has been said above, $R_\infty(k)$ can be continued analytically into the upper half plane of k, and there it satisfies, due to Lemma 3 (a is assumed to be chosen according to the assumptions in Lemma 3),

$$|R_\infty(k)e^{-2ika}|\leqslant 1, \quad \text{for } k_2 \geqslant 0. \tag{4.16}$$

The other part of $R(k)$, here called $R_2(k)$, belongs to $L^2(-\infty,\infty)$. Do not confuse $R_2(k)$ with the imaginary part of a complex number. The index 2 here is used to indicate that the function $R_2(k)$ belongs to $L^2(-\infty,\infty)$. Similarly, the index ∞ on $R_\infty(k)$ is to indicate that $R_\infty(k) \times \exp\{-2ika\}$ is bounded in the upper half plane of k. Notice that, since a is arbitrarily large, R_∞ is decaying, as a function of k, faster than $\exp\{-2k_2a\}$ for any real a.

V. THE MARCHENKO EQUATION

We start this section by proving an important theorem, which has certain analogs to Theorem A.1 in the Appendix, cf. also Ref. 12 on analyticity in tubes.

Theorem 7: Let $u(z)$ be holomorphic in $\mathrm{Im}\{z\}>0$ and

$$|u(z)|\leqslant K, \text{ in } \mathrm{Im}\{z\}>0,$$

i.e., $u\in H^\infty$. Then the Fourier transform

$$\hat{u}(k) = \int_{-\infty}^{\infty} u(x)e^{-ikx}\,dx,$$

is a tempered distribution with support in $[0,\infty)$.

Proof: It is clear that $\hat{u}\in\mathscr{S}'$ [$\mathscr{S}=\mathscr{S}(-\infty,\infty)$ is the Schwartz space, and $\mathscr{S}'$ is the space of tempered distributions], since the function $u\in\mathscr{S}'$ (Theorem A.2 in the Appendix), and the Fourier transform maps $\mathscr{S}'$ into $\mathscr{S}'$. We show that

$$\int_{-\infty}^{\infty}\hat{u}(k)\phi(k)dk = \int_{-\infty}^{\infty}u(x)\hat{\phi}(x)dx = 0,$$

for all functions $\phi(k)\in\mathscr{S}$ with $\mathrm{supp}\{\phi\}\subset(-\infty,0)$. For every such ϕ there exists an $\epsilon>0$, such that $\phi=0$, for $x\geqslant-\epsilon$. We also have that as a function of the complex variable $z=x+iy$,

$$\hat{\phi}(z) = \int_{-\infty}^{\infty}\phi(k)e^{-ikz}\,dk = \int_{-\infty}^{-\epsilon}\phi(k)e^{-ikz}\,dk,$$

is well defined for all $z, y\geqslant 0$, and, furthermore, $\hat{\phi}(z)$ is holomorphic in $y>0$. For $y\geqslant 0$, we have

$$|\hat{\phi}(x+iy)|\leqslant\int_{-\infty}^{-\epsilon}|\phi(k)|e^{ky}\,dk\leqslant Ce^{-\epsilon y},$$

and similarly for all derivatives of ϕ. Thus we get for $y\geqslant 0$

$$|\hat{\phi}(x+iy)|\leqslant C_N e^{-\epsilon y}(1+|z|^N)^{-1},$$
$$\text{for all integer } N\geqslant 0.$$

Here C_N is a constant dependent on N and ϕ. For every $y>0$, we have

$$\int_{-\infty}^{\infty}u(x)\hat{\phi}(x)dx = \int_{-\infty}^{\infty}u(x+iy)\hat{\phi}(x+iy)dx,$$

by the Cauchy theorem, the assumption $|u(z)|\leqslant K$ in $y>0$ (and by Theorem A.2 in the Appendix a.e. on the real axis), and the estimate on $\hat{\phi}(x+iy)$ above. Now choose N large enough so that

$$\left|\int_{-\infty}^{\infty}u(x)\hat{\phi}(x)dx\right|\leqslant Ce^{-\epsilon y}, \quad \text{for all } y\geqslant 0,$$

and we conclude that

$$\int_{-\infty}^{\infty}\hat{u}(k)\phi(k)dk = 0, \quad \text{for all } \phi(k)\in\mathscr{S},$$
$$\mathrm{supp}\{\phi\}\subset(-\infty,0).$$

Therefore, supp $\{\hat{u}\}\subset[0,\infty)$, and the theorem is proved.

Theorem 7 is not stated in its weakest form. The theorem also holds for functions u with certain polynomial growth along the real axis,[12] but the present formulation of the theorem is sufficient for our purposes.

Note that a holomorphic function $f(z)$ in $\mathrm{Im}\{z\}>0$, defined by $f(z)=u(z)e^{itz}$, where $u(z)$ satisfies the assumptions in Theorem 6, has $\mathrm{supp}\{\hat{f}\}\subset[t,\infty)$.

The relation between the scattering solution ψ and the Jost solution f for real k is given by Eq. (4.9)

$$\psi(x,k) = f(x,-k) - R(k)f(x,k). \tag{5.1}$$

By Theorem 6 and Lemma 4 we find that ψ as a function of k, for fixed $x\leqslant a$ and in the absence of bound states, satisfies

$$|\psi(x,k)e^{-ikx}|\leqslant C, \quad \text{in } k_2\geqslant 0, \tag{5.2}$$

where the constant C is independent of k (but depends, of course, on x, a, and q).

We are now ready to tie everything together and derive the Marchenko equation for our scattering problem. This is done simply by taking the Fourier transform of Eq. (5.1), after the following rearrangements and use of Eq. (3.11):

$$\psi(x,k) + R_\infty(k)\left(e^{-ikx} + \int_{-\infty}^{x}A(x,t)e^{-ikt}\,dt\right) - e^{ikx}$$
$$= -R_2(k)\left(e^{-ikx} + \int_{-\infty}^{x}A(x,t)e^{-ikt}\,dt\right)$$
$$+\int_{-\infty}^{x}A(x,t)e^{ikt}\,dt.$$

If no bound states are present we take the Fourier transform

$$\hat{f}(y) = \int_{-\infty}^{\infty}f(k)e^{-iky}\,dk,$$

for $y<x$ and apply Theorem 7. Since the parameter a that appears in $R_\infty(k)$ can be arbitrarily large, it suffices to consider the case $x\leqslant a$. We get

$$0 = -\hat{R}_2(x+y) - \int_{-\infty}^{x}A(x,t)\hat{R}_2(t+y)dt + 2\pi A(x,y),$$

or

$$A(x,y) = A_0(x+y) + \int_{-\infty}^{x}A(x,t)\,A_0(t+y)dt, \; y<x, \tag{5.3}$$

where

$$A_0(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R(k)e^{-iky}\,dk = \frac{1}{2\pi}\hat{R}_2(y), \qquad (5.4)$$

and where the last equality holds for $y < 2a$ due to the support of the tempered distribution $\hat{R}_\infty(y)$. When bound states are present at $i\beta_\nu$, $i = 1,\ldots,n$, we replace $A_0(y)$ by

$$\tilde{A}_0(y) = A_0(y) + \sum_{\nu=1}^{n} M_\nu e^{\beta_\nu y},$$

where the normalization constant M_ν is given by Eq. (4.12)

$$M_\nu = \left\{\int_{-\infty}^{\infty} f^2(t,i\beta_\nu)\,dt\right\}^{-1}.$$

To see the correspondence and the similarities between the formalism presented here and the Marchenko equation for the radial scattering problem for s-waves, we write Eq. (5.4) as

$$A_0(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} [R(k) - R_\infty(k)]e^{-iky}\,dk, \quad y < 2a, \qquad (5.5)$$

where $R_\infty(k)$ is given by Eq. (4.15). Compare this expression with the corresponding expression for the radial problem, with S-matrix $S(k)$ (see Ref. 1)

$$A_0(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} [S(k) - 1]e^{-iky}\,dk.$$

This expression can be obtained formally from Eq. (5.5) by introducing the boundary conditions for the radial problem in Eq. (4.15) at $a = 0$, i.e., $\phi(0,k) = 0$, and $\phi'(0,k) = 1$.

The reconstruction of the potential from the solution of the Marchenko equation is now identical to the standard case.[1] The potential is obtained from

$$q(x) = 2\frac{d}{dx}[A(x,x)], \qquad (5.6)$$

and $A(x,t)$ satisfies the hyperbolic equation

$$\left(\frac{d}{dx^2} - \frac{d}{dt^2}\right)A(x,t) = q(x)\,A(x,t), \quad t < x. \qquad (5.7)$$

Finally, we prove a uniqueness theorem for the potential q in this class Q.

Theorem 8: A potential q (in class Q) without bound states is uniquely defined by its reflection coefficient $R(k)$.

Proof: Let ψ_i and f_i, $i = 1,2$, be two scattering and Jost solutions, respectively, giving the same reflection coefficient $R(k)$, where $|R(k)| = 1$. Introduce

$$u(x,k) = [\psi_1(x,k) - \psi_2(x,k)]e^{-ikx},$$
$$h(x,k) = [f_1(x,k) - f_2(x,k)]e^{ikx}.$$

In the absence of bound states, $u(x,k)\in H^\infty$, see Eq. (5.2), and $h(x,k)\in H^2$. Furthermore, we have for real k,

$$u(x,k) = \overline{h(x,k)} - R(k)e^{-2ikx}h(x,k). \qquad (5.8)$$

The left-hand side is a boundary function of an H^∞ function, which also is in $L^2(-\infty,\infty)$, since the right-hand side is in $L^2(-\infty,\infty)$. The rest of the proof is an application of the Plancherel theorem (see the Appendix) and the results on the support of the Fourier transformation. We get

$$0 = \int_{-\infty}^{\infty} \hat{u}(x,t)\hat{h}(x,-t)\,dt = 2\pi\int_{-\infty}^{\infty} u(x,k)h(x,k)\,dk$$
$$= 2\pi\int_{-\infty}^{\infty} [|h(x,k)|^2 - R(k)e^{-2ikx}h^2(x,k)]\,dk.$$

On the other hand we have

$$\int_{-\infty}^{\infty} |u(x,k)|^2 dk = \int_{-\infty}^{\infty} [2|h(x,k)|^2 - 2\,\mathrm{Re}\{R(k)e^{-2ikx}h^2(x,k)\}]\,dk = 0,$$

so $u(x,k)\equiv 0$ and $\psi_1(x,k)\equiv\psi_2(x,k)$. Thus $q_1 = (\psi_1'' + k^2\psi_1)/\psi_1 = q_2$ and the theorem is proved.

Note: If bound states are present the potential is uniquely determined by its reflection coefficient $R(k)$, its bound state energies $-\beta_\nu^2$, $\nu = 1,\ldots,n$, and its normalization constants M_ν, $\nu = 1,\ldots,n$.

VI. EXAMPLES

A simple example that illustrates some of the results above is the linear potential

$$q(x) = xH(x),$$

where $H(x)$ is the step function [$H(x) = 1, x\geqslant 0$, and zero otherwise]. The scattering solution is easily calculated

$$\psi(x,k) = \begin{cases} e^{ikx} - R(k)e^{-ikx}, & x\leqslant 0, \\ C(k)\mathrm{Ai}(x - k^2), & x\geqslant 0, \end{cases}$$

where Ai is the Airy function[13] and

$$R(k) = \frac{\mathrm{Ai}'(-k^2) - ik\,\mathrm{Ai}(-k^2)}{\mathrm{Ai}'(-k^2) + ik\,\mathrm{Ai}(-k^2)}, \qquad (6.1)$$
$$C(k) = 2ik/[\mathrm{Ai}'(-k^2) + ik\,\mathrm{Ai}(-k^2)].$$

It is easy to see that

$$\int_0^\infty t\,\mathrm{Ai}(t - k^2)e^{-ikt}\,dt = -\mathrm{Ai}'(-k^2) - ik\,\mathrm{Ai}(-k^2),$$

which is equivalent to Eq. (4.3) in Theorem 6. It is also straightforward to evaluate the asymptotic behavior of the reflection coefficient $R(k)$ in Eq. (6.1). We get

$$R(k) = e^{i\{(4/3)/k^3 + \pi/2\}}\{1 + O(k^{-3})\}, \quad \text{as } k\to\infty.$$

Potentials of a general power of x can be analyzed by asymptotic methods given by Brander.[14]

An example of an exponential potential is ($\beta > 0$)

$$q(x) = (e^{\beta x} - 1)\,H(x).$$

The solution to this problem is

$$\psi(x,k) = \begin{cases} e^{ikx} - R(k)e^{-ikx}, & x\leqslant 0, \\ C(k)K_\nu((2/\beta)e^{\beta x/2}), & x\geqslant 0, \end{cases}$$

where $\nu = (2i/\beta)(k^2 + 1)^{1/2}$, K is the modified Bessel function of second kind,[13] and

$$R(k) = \frac{K_\nu'(2/\beta) - ikK_\nu(2/\beta)}{K_\nu'(2/\beta) + ikK_\nu(2/\beta)}, \qquad (6.2)$$
$$C(k) = 2ik/[K_\nu'(2/\beta) + ikK_\nu(2/\beta)].$$

Now $K_{i\alpha_1}(\alpha_2)$ is real for real α_1 and α_2 (see Ref. 13). The asymptotic behavior of $R(k)$ in Eq. (6.2) is given by

$$R(k) = \beta^{2\nu} \frac{\Gamma(1+\nu)}{\Gamma(1-\nu)} \{1 + O(k^{-1})\}$$
$$= \exp\{\nu \ln(1-\nu^2) + 2\nu \ln(\beta/e) + i\pi/2\}$$
$$\times \{1 + O(k^{-1})\}, \quad \text{as } k \to \infty.$$

Similarly, for the pure exponential potential $(\alpha,\beta > 0)$

$$q(x) = \alpha\, e^{\beta x}.$$

We get $(\nu = 2ik/\beta)$

$$R(k) = \frac{\Gamma(1+\nu)}{\Gamma(1-\nu)} \left(\frac{\beta^2}{\alpha}\right)^{\nu}. \tag{6.3}$$

A more complicated potential is the Morse potential $(\alpha,\beta > 0)$

$$q(x) = \alpha^2 (e^{2x/\beta} - 2e^{x/\beta}).$$

The scattering solution $\psi(x,k)$ is

$$\psi(x,k) = \exp(\alpha\beta e^{x/\beta})$$
$$\times \{ e^{ikx}\, {}_1F_1(ik\beta + \tfrac{1}{2} + \alpha\beta; 2ik\beta + 1; -2\alpha\beta e^{x/\beta})$$
$$- R(k) e^{-ikx}\, {}_1F_1(-ik\beta + \tfrac{1}{2}\alpha\beta; -2ik\beta + 1;$$
$$-2\alpha\beta e^{x/\beta})\},$$

where

$$R(k) = e^{-2i\pi k\beta}(-2\alpha\beta)^{-2ik\beta}$$
$$\times \frac{\Gamma(\frac{1}{2} - \alpha\beta - ik\beta)\Gamma(1+2ik\beta)}{\Gamma(\frac{1}{2} - \alpha\beta + ik\beta)\Gamma(1-2ik\beta)}, \tag{6.4}$$

and ${}_1F_1$ is the Kummer's function.[13]

ACKNOWLEDGMENT

The work reported in this paper is sponsored by the National Swedish Board for Technical Development (STU) and their support is gratefully acknowledged.

APPENDIX: HARDY SPACES H^p

For the convenience of the reader we collect in this appendix some useful results on the Hardy space H^p. A general introduction to H^2 is given by Dym and McKean,[15] p. 160. For general H^p, $0 < p \leqslant \infty$ we refer to Ref. 16.

We introduce the notation D for the open upper half plane, i.e.,

$$D = \{z = x + iy | y > 0\},$$

and define the translation of the argument of a function f defined in D by

$$f_y(x) \equiv f(x+iy).$$

The Hardy space H^p, $0 < p < \infty$, is defined as

$$H^p = \{f(z) | f \text{ holomorphic in } D \text{ and } \sup_{y>0} \|f_y\|_p < \infty\}, \tag{A1}$$

where

$$\|f\|_p = \left\{ \int_{-\infty}^{\infty} |f(x)|^p dx \right\}^{1/p}.$$

H^∞ is defined as

$$H^\infty = \{f(z) | f \text{ holomorphic in } D \text{ and } |f| \text{ bounded in } D\}. \tag{A2}$$

The Fourier transform is defined as

$$\hat{f}(k) = \int_{-\infty}^{\infty} f(x) e^{-ikx} dx, \tag{A3}$$

$$\check{f}(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) e^{ikx} dx. \tag{A4}$$

For L^2 functions we have (in the L^2 sense)

$$f(x) = \check{\hat{f}}(x),$$

and the Plancherel theorem

$$\|\hat{f}\|_2 = \sqrt{2\pi} \|f\|_2,$$

or more generally

$$\int_{-\infty}^{\infty} \hat{f}(-y)\hat{g}(y) dy = 2\pi \int_{-\infty}^{\infty} f(x) g(x) dx.$$

For a function f in H^2 we have the following important theorem.[16]

Theorem A. 1: A function f is in H^2 if and only if there exists a function $\hat{f} \in L^2(0,\infty)$ with support in $[0,\infty)$ such that

$$f(z) = \int_0^{\infty} \hat{f}(k) e^{ikz} dk, \quad z \in D, \tag{A5}$$

and

$$\sup_{y>0} \|f_y\|_2 = (1/\sqrt{2\pi}) \|\hat{f}\|_2. \tag{A6}$$

This theorem shows that we can identify H^2 with the Fourier transform of functions in $L^2(0,\infty)$, more precisely

$$H^2 = \{f \in L^2(-\infty,\infty) | \hat{f}(k) = 0, \text{ for almost all } k < 0\}.$$

Similarly, we define $\tilde{H}^2$

$$\tilde{H}^2 = \{f \in L^2(-\infty,\infty) | \hat{f}(k) = 0, \text{ for almost all } k > 0\},$$

which has corresponding holomorphic properties in the lower complex half-plane. The isometry

$$L^2(-\infty,\infty) \simeq H^2 \oplus \tilde{H}^2 \tag{A7}$$

holds and the following orthogonal projections on L^2 onto $\tilde{H}^2$ and H^2 show that the decomposition of Eq. (A7) is orthogonal:

$$(P^- f)(x) = \frac{1}{2\pi} \int_{-\infty}^{0} \left\{ \int_{-\infty}^{\infty} f(t) e^{-ikt} dt \right\} e^{ikx} dk$$
$$= (1_{(-\infty,0)} \hat{f})^{\vee}(x), \tag{A8}$$

$$(P^+ f)(x) = \frac{1}{2\pi} \int_{0}^{\infty} \left\{ \int_{-\infty}^{\infty} f(t) e^{-ikt} dt \right\} e^{ikx} dk$$
$$= (1_{(0,\infty)} \hat{f})^{\vee}(x). \tag{A9}$$

We close this Appendix by giving the following important theorems for functions in H^p (see Ref. 16).

Theorem A.2: If $f \in H^p$, $1 < p < \infty$, then the boundary function

$$f(x) = \lim_{y \to 0} f(x+iy)$$

exists pointwise almost everywhere and $f \in L^p$.

Theorem A.3: Let $0<p,r<\infty$. If $f\in H^p$ and if its boundary function is in L^r, then f is also in H^r.

Theorem A.4: If $f\in H^p$, $1\leqslant p\leqslant\infty$, then

$$f(z)=\frac{y}{\pi}\int_{-\infty}^{\infty}\frac{f(t)dt}{(x-t)^2+y^2}.$$

Furthermore,

$$f(z)=\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{f(t)}{t-z}dt,\ y>0,$$

and $f(z)=0$, for all $y<0$. Conversely, if $f(t)\in L^p(-\infty,\infty)$, $1\leqslant p\leqslant\infty$, and

$$f^*(z)=\frac{y}{\pi}\int_{-\infty}^{\infty}\frac{f(t)dt}{(x-t)^2+y^2}.$$

is holomorphic in D or alternatively $(p\neq\infty)$

$$f^*(z)=\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{f(t)}{t-z}dt\equiv 0,\quad y<0,$$

then $f^*\in H^p$ and its boundary function $f(x)=f^*(x)$ in the sense of mean convergence (L^p norm).

Theorem A.5: If $f\in H^p$, $1\leqslant p<\infty$ then

$$\lim_{y\to 0}\|f_y\|_p^p=\|f\|_p^p,$$

and

$$\lim_{y\to 0}\|f_y-f\|_p^p=0.$$

Furthermore, if $0<y_1<y_2$, then

$$\|f_{y_2}\|_p\leqslant\|f_{y_1}\|_p\leqslant\|f\|_p.$$

[1]K. Chadan and P. C. Sabatier, *Inverse Problems in Quantum Scattering Theory* (Springer, New York, 1977).
[2]P. Deift and E. Trubowitz, "Inverse Scattering on the Line," Commun. Pure Appl. Math. **XXXII**, 121 (1979).
[3]P. P. Kulish, "Inverse scattering problem for the Schroedinger equation on the axis," Math. Zametki **4**, 677 (1968) [Math. Notes **4**, 895 (1968)].
[4]S. Graffi and E. Harrell, "Inverse scattering for the one-dimensional Stark effect and application to the cylindrical KdV equation," Ann. Inst. H. Poincaré **XXXVI**, 41 (1982).
[5]F. Calogero and A. Degasperis, "Inverse spectral problem for the one-dimensional Schrödinger equation with an additional linear potential," Lett. Nuovo Cimento **23**, 143 (1978).
[6]O. M. Lomonosova, "The inverse scattering problem for the Schrödinger equation on the line and a new class of solutions of the Korteweg–de Vries equation," Sov. Math. Dokl. **25**, 751 (1982).
[7]H. Weyl, "Über gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlichen Funktionen," Math. Ann. **68**, 220 (1910).
[8]E. Hille, *Lectures on Ordinary Differential Equations* (Addison–Wesley, Reading, MA, 1969).
[9]E. Hille, "Green's transforms and singular boundary value problems," J. Math. Pures Appl. **42**, 332 (1963).
[10]E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw–Hill, New York, 1955).
[11]E. Hille, *Analytic Function Theory* (Ginn, Boston, 1959), Vol. I.
[12]M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vols. I and II (Academic, New York, 1972 and 1975).
[13]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Natl. Bur. Stand., Washington, D. C., 1964).
[14]O. Brander, "A peculiar property of the wavefunctions for confining potentials," Phys. Lett. A **75**, 471 (1980).
[15]H. Dym and H. P. McKean, *Fourier Series and Integrals* (Academic, New York, 1972).
[16]P. L. Duren, *Theory of H^p Spaces* (Academic, New York, 1970).

Sturmian eigenvalue equations with a Bessel function basis

G. H. Rawitscher
Department of Physics, University of Connecticut, Storrs, Connecticut 06268
G. Delic
Nuclear Theory Research Group, Department of Physics and Schonland Research Center for Nuclear Sciences, University of the Witwatersrand, 1 Jan Smuts Avenue, Johannesburg 2001, South Africa

(Received 9 July 1985; accepted for publication 28 October 1985)

A non-self-adjoint Sturmian eigenvalue equation of the form $Av = f$, encountered in quantum scattering theory, is solved as a complex general matrix eigenvalue problem. The matrix form is obtained on expansion of the solution in a discrete set of spherical Sturmian–Bessel functions of complex argument. This set of basis functions gives better convergence behavior for both the eigenvalues and eigenfunctions when compared to the results of a Chebyshev polynomial method reported previously.

I. INTRODUCTION

The theory of potential scattering can be formulated by expansion of the scattering waves in terms of a set of basis states. For low incident energies, when compound nuclear resonances dominate,[1] a convenient set of basis states are the Gamow functions,[2] as was shown by Kapur and Peierls[3] in 1938. A great deal of effort has already been devoted to the numerical evaluation of Gamow states, either in configuration space[4] or in momentum space.[5,6]

For high energies, where a smooth optical potential gives an adequate description of the nucleon–nucleus interaction, it is usual to solve the Schrödinger equation numerically, and avoid unnecessary expansions. However, when the optical potential is replaced by a set of coupled equations, as is done in the case of a microscopic description of the nucleon–nucleus interaction, and if the number N of channels becomes large (larger than ~ 40), then an expansion in a set of Sturmian basis functions[7,8] can become preferable. The conventional numerical method of solving the equations on a mesh of radial steps involves a large amount of computer time, which increases as N^3, and at the same time becomes unreliable, since it is not easy to numerically satisfy the outgoing wave boundary conditions in all N channels. On the other hand, the computing time for an expansion in terms of a basis of Sturmian states, increases as N^2, and the validity of the boundary conditions is automatically assured.

One such basis is the set of Sturmian eigenfunctions for a square well potential. They are proportional to the product of the radial distance r times a spherical Bessel function of complex argument $K_j r$, where the K_j's are discrete complex wave numbers. A general Sturmian function is defined in a radial interval from 0 to a. At the upper limit the logarithmic derivative of this function is required to be the same as that of the outgoing Hankel (or Coulomb) function for the real physical energy of the channel in question. At the origin this Sturmian function is required to vanish, and between 0 and a it obeys the Schrödinger equation for the given optical potential, which, however, is multiplied by a complex scalar (called the eigenvalue) such that the boundary condition is satisfied. This differs from the Gamow states case, where the potential is kept fixed and the energy is made complex instead. For the Sturmian–Bessel case the potential is a square well of radius a, which is added to the centripetal potential for angular momentum l. For this case the Sturmian–Bessel functions and the Gamow functions are identical.

The Sturmian–Bessel basis, has already shown its usefulness in that it gives rise to expressions for the scattering T-matrix in complex momentum space, which can be solved numerically without difficulty.[9] Furthermore, the required integrals of products of two such functions times a potential, over the finite radial interval $[0,a]$ can be computed rapidly and with great accuracy for the case that the potential is a Gaussian function, by employing semianalytical expressions for the error function.[10] A fast algorithm for obtaining the complex Bessel wave numbers K_j is also available.[11]

A basis of Sturmian–Bessel functions is also useful for calculating the Sturmian eigenfunctions for a general potential, be it for the case of a single channel or for a set of coupled channels. These general Sturmian eigenfunctions in turn are useful for providing succinct separable representations for the multichannel Green's functions that occur in a set of coupled equations[12] for the corresponding T-matrix operator, and for the nonlocality of the corresponding optical potential. These general Sturmian states are also useful in assessing the strength of the potentials through the size of the eigenvalues, and for producing corrections to the distorted wave Born approximation.[12]

In view of the usefulness of Sturmian functions in scattering theory it is of interest to assess the accuracy with which such functions v can be calculated for a general potential $\bar{V}$ by expanding them into a set of basis functions.

The eigenvalue equation to be solved is

$$Av = \alpha \bar{V} v, \tag{1}$$

where A is a linear second-order differential operator that is not necessarily self-adjoint because it can contain a complex, diffuse, local interaction potential V_0. For the choice of boundary conditions discussed above and in Ref. 13, α and v are the corresponding complex eigenvalues and eigenfunctions. In this case the v's are not Gamow states because the energy, contained in A, is still real while the potential $\bar{V}$ is made complex (with a positive, or emissive, imaginary part) through the multiplication by the scalar α.

In a previous paper[13] we solved Eq. (1) for an uncou-

 0022-2488/86/030816-08$02.50

pled general complex potential by expanding v in terms of a set of Chebyshev polynomials, and we have examined the feasibility and accuracy of this method for several examples that are commonly used in nuclear physics applications. It is the purpose of the present paper to make a similar type of analysis with a basis set of Sturmian–Bessel functions, and compare the accuracy achieved in this case with that for the Chebyshev basis.

An advantage of the Bessel function basis (BFB) over the Chebyshev polynomial basis (CPB) is that the quantities α and v converge more rapidly with the size of the basis. A disadvantage of the BFB is that it may not have the stable numerical properties characteristic of the CPB. Furthermore, the standard operations of differentiation and integration on expansions in the basis, while trivial in the Chebyshev case, often need to be evaluated numerically in the Bessel function basis.

For certain pathological potentials it could happen that the corresponding Sturmian functions do not form a complete set, for example, when the integral of the square of such a function from 0 to a vanishes. We have not encountered this situation in our various applications. This possibility, however, does not invalidate the general usefulness of these functions, and the need to understand their convergence behavior.

This study is divided into six sections and an appendix. Sections II and III describe the Sturmian–Bessel functions and explain how they are obtained. Sections IV and V discuss application of the method to case five of Ref. 13 and compare the results for the eigenvalues and eigenfunctions with the Chebyshev polynomial basis of Ref. 13. Section VI summarizes the conclusions, and the Appendix describes a method of finding the complex wave numbers of the spherical Bessel functions.

II. THE STURMIAN–BESSEL FUNCTION BASIS

The Sturmian–Bessel function basis set is related to the regular solution of the equation

$$\left[\frac{d^2}{dr^2}+k^2-\frac{l(l+1)}{r^2}\right]f_{ln}(r)=\bar{U}_{ln}(r)f_{ln}(r)\,, \qquad (2)$$

where $k^2=2\mu E/\hbar^2$,

$$\bar{U}_{ln}(r)=(2\mu/\hbar^2)(\bar{V}_{ln}+i\,\bar{W}_{ln})\,,\quad r<a\,,$$
$$=0\,,\quad r>a\,, \qquad (3)$$

with $\bar{V}_{ln}$ and $\bar{W}_{ln}$ real constants determined by the boundary condition on $f_{ln}(r)$ at $r=a$ (see below); E is the center-of-mass energy in MeV of a particle mass μ, k^2 is a real constant, and l refers to orbital angular momentum and has only integer values. The interior ($r<a$) solution of Eq. (2) with the potential (3) is

$$f_{ln}(r)=K_{ln}r\,j_l(K_{ln}r)\,, \qquad (4)$$

where j_l is the spherical Bessel function of complex argument[14] with

$$K_{ln}^2=k^2-\bar{U}_{ln}\,, \qquad (5)$$

and the subscript n in Eqs. (4) and (5) indicates that Eq. (2) has solutions only for a discrete set of complex potentials $\bar{U}_{ln}$ for fixed l. This set is discrete because the boundary condition on $f_{ln}(r)$ at $r=a$ is

$$\left[\frac{1}{f_{ln}(r)}\frac{df_{ln}(r)}{dr}\right]_{r=a}=\left[\frac{1}{h_l^{(+)}(kr)}\frac{dh_l^{(+)}(kr)}{dr}\right]_{r=a}, \qquad (6)$$

where

$$h_l^{(+)}(kr)=i^{l+1}kr\,h_l^{(1)}(kr)\,, \qquad (7)$$

and $h_l^{(1)}(kr)$ is a spherical Bessel function of the first kind[14] of real argument. If $l=0$, then the right-hand side of (6) is ik, while if $l\neq 0$, but $ka<l$, it is not too different from ik.

The boundary condition (6) is a complex transcendental equation for the roots K_{ln}, and has been solved[15] for $l=0$ by Nussenzveig and Joffily with a complex square well potential and by Kaus and Pearson for $l=0,1$ with the real case. A numerical method of obtaining the solution, using a rapidly convergent Newton iteration technique is described in the Appendix.

The solutions (4) form a discrete set corresponding to the discrete set of complex numbers

$$K_{ln}a=A_{ln}+iB_{ln}\,,\quad n=1,2,\ldots, \qquad (8)$$

which solve the boundary condition (6) for each l. For different values of the asymptotic wave number k (or energy E), or different l, another set of numbers (8) is obtained. As n increases by unity, the corresponding function $f_{ln}(r)$ acquires an additional node inside $r=a$ and the real part of the square well potential (3) becomes more negative. The corresponding values of the imaginary part are all positive since the square well has to be "emissive" in order that the $f_{ln}(r)$ have an asymptotic boundary condition of only "outgoing waves" (see Appendix A of Ref. 13). As n increases, the real part of $K_{ln}a$ in (8) increases by approximately π. The corresponding magnitude of B_{ln} (which is always negative) initially increases slowly as a function of n and then decreases (see the Appendix).

The Sturmian BFB, $\phi_{ln}(r)$, $n=1,2,\ldots$, is defined by

$$\phi_{ln}(r)=N_{ln}K_{ln}r\,j_l(K_{ln}r),\quad r<a\,, \qquad (9a)$$
$$=N'_{ln}\,h_l^{(+)}(kr)\,,\quad r>a\,. \qquad (9b)$$

The basis is orthonormal with the normalization constants N_{ln} chosen such that

$$(\bar{V}_{ln}+i\bar{W}_{ln})\int_0^a\phi_{ln}(r)\phi_{ln'}(r)dr=\delta_{nn'}\,, \qquad (10)$$

which follows uniquely from the boundary condition of Eq. (6).

The Sturmian BFB set defined by Eqs. (9) and (10) is used to expand solutions of Eq. (1) corresponding to diffuse complex interactions as described in the next section.

III. SOLUTION OF EIGENVALUE EQUATIONS WITH THE BASIS

The solutions of Eq. (1) are also denoted as positive energy channel Weinberg states in Ref. 16. They are regular at the origin, asymptotically "outgoing waves" (with real wave number k) and obey the eigenvalue equation

$$(T_l-E)v_{lj}(r)=-(\alpha_{lj}+1)\bar{V}(r)v_{lj}(r)\,. \qquad (11)$$

Here

$$T_l = \frac{\hbar^2}{2\mu}\left[-\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2}\right]. \quad (12)$$

The potential $\bar{V}$ is complex, and $\alpha_{lj}, j = 1,2,...$, is the discrete eigenvalue. Equation (11) is a special case of Eq. (3) of Ref. 13 with U_0 set equal to $\bar{U}$.

In the present study an approximation to v_{lj} is obtained by expansion on the BFB $\phi_{lj}(r)$ described in Sec. II,

$$v_{lj}^{(N)}(r) = \sum_{n=1}^{N} \phi_{ln}(r)\, e_{nj}^{(N)}. \quad (13)$$

The coefficients in the expansion of Eq. (13) are the eigenvectors of the matrix equation[16]

$$\sum_{n'=1}^{N} V_{nn'}\, e_{n'j}^{(N)} = (\alpha_{lj}^{(N)} + 1)^{-1} e_{nj}^{(N)}, \quad n = 1,...,N, \quad (14)$$

obtained by inserting (13) into (11), multiplying both sides by $\phi_{ln'}(r)$, integrating over r from 0 to a, and using the normalization property (10) written in the form

$$\int_0^a \phi_{ln}(r)(T_l - E)\phi_{ln'}(r)\,dr = -\delta_{nn'}. \quad (15)$$

The matrix elements $V_{nn'}$ are given by

$$V_{nn'} = \int_0^a \phi_{ln}(r)\bar{V}(r)\phi_{ln'}(r)dr. \quad (16)$$

The accuracy of the expansion as a function of the basis size N is studied for eigenvalues α_{lj} in Sec. IV and eigenfunctions v_{lj} in Sec. V. A comparison is made with the results of Ref. 13 for the expansion of $v_{lj}(r)$ on a basis of Chebyshev polynomials.

IV. CONVERGENCE FOR EIGENVALUES

Case five, as discussed in Ref. 13, corresponds to a neutron at 15 MeV (Lab) scattering from ^{16}O and therefore the mass of the projectile and target are those of a neutron and the nucleus ^{16}O. The value of $2\mu/\hbar^2$ is $0.047\,832\times 16/17$ fm^{-2} MeV^{-1} and μ is the reduced mass. The center-of-mass energy is 14.11 MeV, the wave number k is $0.797\,215\,2\ fm^{-1}$ and the matching radius a, beyond which the potential $\bar{V}$ is set to zero, has the value of 7.39 fm. The values of the angular momentum quantum number l range from 0 to 10 in the calculations reported here.

The functions $\phi_{ln}(r)$, defined in Eq. (9) were calculated by solving Eq. (2) in single precision by means of the Numerov method[16] with a step size $\Delta r = 7.2168\times 10^{-3}$ fm. The required depth $\bar{V}_{ln} + i\,\bar{W}_{ln}$ of the square well potential (3) is obtained by solving the transcendental equation (6) by the iterative procedure, described in the Appendix. The integrals required for the normalization condition (10) and the matrix elements $V_{nn'}$, Eq. (16), were calculated by quadrature using Simpson's rule. The diagonalization of the matrix equation (14) was performed with the EISPACK routines using IBM double precision arithmetic[17] and the resulting values of $\alpha_{lj}^{(N)}$ and $v_{lj}^{(N)}(r)$ were studied as a function of N. Eigenvalues obtained by the Bessel function method of the present work for $l = 0$ to 10 and $j \leqslant 10$ were compared with the previous results[13] obtained for case five with the Chebyshev polynomial method. Differences were typically of the order of one digit in the fourth significant figure and exceeded this by several digits only in a few instances.

Figure 1 shows the rate of convergence of $\alpha_{lj}^{(N)}$ to a prescribed error of one digit in the Sth significant figure as a function of N the number of BFB elements used in (13) for $l = 0$. This figure, when compared to Fig. (2d) of Ref. 13, shows a substantial improvement over the CPB. The BFB requires approximately half the number of basis states used by the CPB and the rate of convergence from $S = 1$ to $S = 4$ is twice that of the CPB. Figure 1 shows a result which is typical for the BFB, namely, a rapid convergence beyond $S = 1$, whereas convergence in the CPB is slower for case five. Thus the *difference* in basis size $(N_4^{j=8} - N_4^{j=1})$ required to produce an accuracy of $S = 4$ for all eigenvalues upto $j = 8$ is smaller for the BFB case (~ 16) than for the CPB (~ 24).

Another measure of convergence for the eigenvalues is given in Fig. 2. This figure shows all eigenvalues of magnitude $\leqslant 110$ (for $l = 0$ to 10), which have converged to four

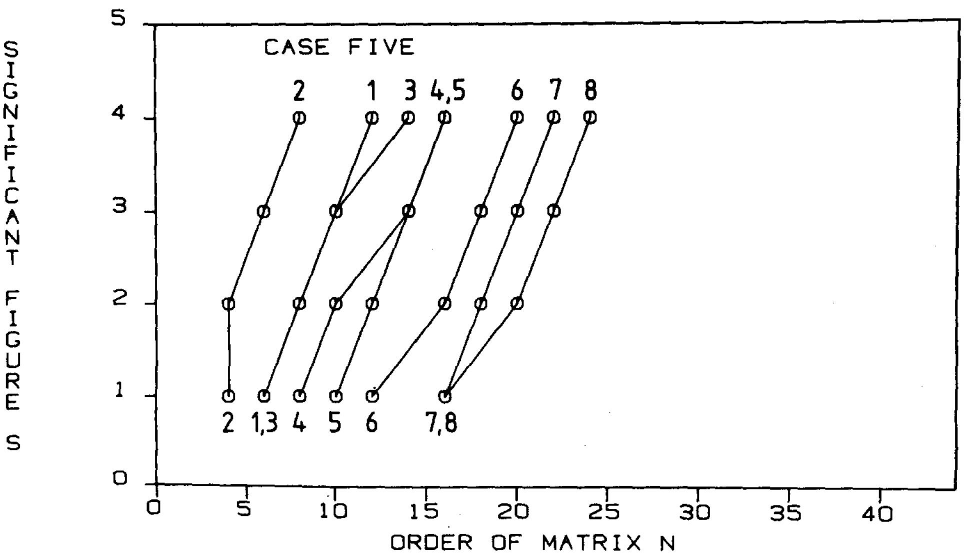


FIG. 1. Rate of convergence to one digit in the Sth decimal as a function of matrix order for $j = 1$ to 8 and $l = 0$. The numbers on the curves give the j of the eigenvalue.

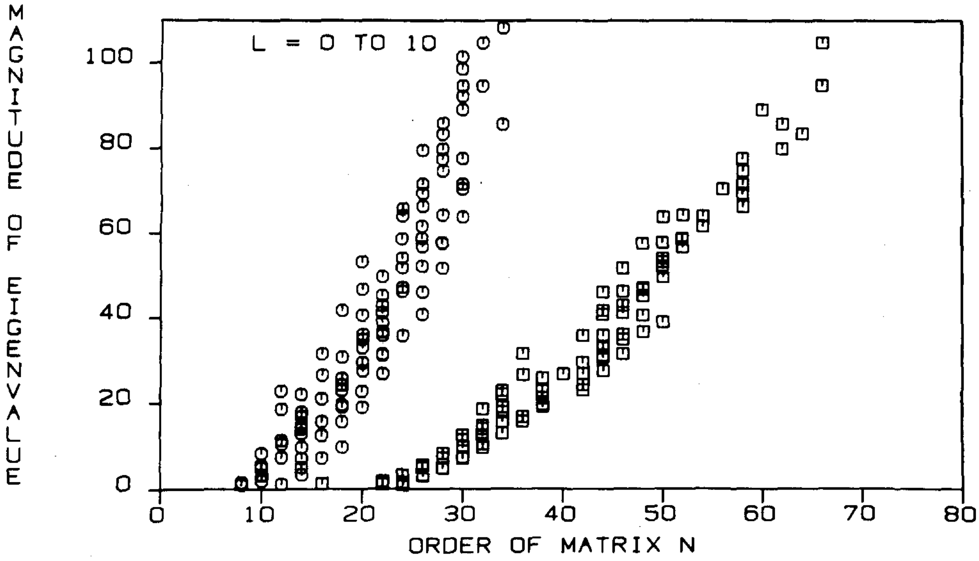


FIG. 2. The matrix order N required to produce convergence of one digit in the fourth figure for $l = 0$ to 10, as a function of the magnitude of the eigenvalues. All eigenvalues that have reached this prescribed error are included. The circles and squares correspond, respectively, to the Bessel function and Chebyshev polynomial basis.[13]

significant figures as a function of the matrix order N required to achieve this limit. The squares correspond to the results of Fig. 4 of Ref. 13 for the CPB and the circles are the results of the BFB for the present calculation. This type of plot is important in assessing how large a matrix is required to ensure convergence of all complex eigenvalues lying inside a circle of prescribed radius. For both the CPB and the BFB, the points are clustered along a line, an estimate of the efficiency of either basis is given by the slope of such a line. A comparison shows that in the BFB case such a line has a slope ~ 4.6 compared to ~ 2.7 for the CPB. This comparison confirms the result of Fig. 1, namely, that for eigenvalues of similar magnitude the BFB requires approximately half the number of basis states when compared to the CPB results. However, the members of the BFB are complex while those of the CPB are real. Thus a direct comparison of the two bases should take account of the fact that one complex function is equivalent to two real functions.

V. CONVERGENCE FOR EIGENFUNCTIONS

The expansion coefficients $e_j^{(N)}$ on the BFB are the eigenvectors corresponding to the eigenvalues $\alpha_{lj}^{(N)}$ for the $N \times N$ matrix eigenvalue equation (14). The eigenvectors may be normalized by the same condition as used in Sec. III of Ref. 13. In the comparison of different basis sizes N_S, error curves

$$A_S\left[v_{lj}^{(N_4)}(r) - v_{lj}^{(N_S)}(r)\right], \quad S = 1,2,3, \qquad (17)$$

were produced, with A_S as a scale factor and (as in Sec. IV) $S = 1,2,3$, and 4 correspond, respectively, to convergence to within one digit in the first to fourth significant figure for

TABLE I. Comparison of maximum error on $0 < r < 7.39$ fm.[a]

l	j	N_S	Chebyshev polynomial basis[b] Real	Imaginary	N_S	Bessel function basis Real	Imaginary[c]
0	1	12	0.46(− 2)	0.41(− 2)	6	0.16(− 1)	0.21(− 1)
		16	0.68(− 3)	0.65(− 3)	8	0.42(− 2)	0.26(− 2)
		18	0.31(− 3)	0.29(− 3)	10	0.84(− 3)	0.73(− 3)
		22	...	...	12	...	...
	10	30	0.37	0.38	22	0.29	0.26
		36	0.76(− 1)	0.76(− 1)	24	0.29(− 1)	0.32(− 1)
		40	0.12(− 1)	0.11(− 1)	26	0.53(− 2)	0.68(− 2)
		48	...	...	28	...	...
4	1	8	2.7	1.4	4	0.37	0.37
		12	1.3	0.26	6	0.55(− 1)	0.38(− 1)
		18	0.28(− 1)	0.13	8	0.11(− 1)	0.91(− 2)
		26	...	...	10	...	...
	8	30	1.6	1.2	20	0.34	0.36
		40	0.34	0.31	22	0.28(− 1)	0.41(− 1)
		44	0.18	0.20	24	0.76(− 2)	0.96(− 2)
		52	...	...	26	...	...

[a] The maximum error is the maximum value of $|v_{lj}^{(N_4)} - v_{lj}^{(N_S)}|$ for $S = 1,2,3$.
[b] From Ref. 13.
[c] The number in parenthesis is the exponent of 10.

real and imaginary parts of the eigenvalue α_{lj}. Thus the area under the modulus of the error curve is a measure of the magnitude of the error remaining on truncation after N_S terms. Error curves (17) were produced for $l = 0$ ($j = 1$ and 10), $l = 4$ ($j = 1$, and 8), and $l = 6$ ($j = 1$).

Table I shows a comparison of the maximum error produced on the interval of approximation [0,7.39] by the BFB and the CPB for $l = 0$ and 4. For the case of $l = 0$ and $j = 1$, with approximately twice the number of basis states, the CPB has a maximum error that is as much as five times smaller than that of the BFB. In this example, the two methods give a comparable (maximum) error if the CPB has approximately four terms more than the BFB. Thus, for $l = 0, j = 1$, the rapid convergence of the eigenvalue in the BFB when compared to the CPB does not necessarily imply a smaller error for the eigenfunction. However, for $l = 0$, $j = 10$ and $l = 4, j = 1$, similar maximum errors are obtained only when the CPB has approximately 14 basis elements more than the BFB. For $l = 4, j = 8$ the difference in basis size is of the order of 22 and in the case of $l = 6, j = 1$ (not shown) this becomes a difference of 30 terms more for the CPB.

The error curves (17) for the real part of the eigenfunctions are shown in Figs. 3 and 4 with parts (a) and (b) corresponding, respectively, to BFB and CPB. For $l = 0$, $j = 1$, it is seen in Fig. 3 that the CPB with $N_3 = 18$ provides a much better approximation than does the BFB with $N_3 = 10$ over the whole interval in that it differs from zero less than the BFB does. However, in Fig. 4, the converse is the case, with the BFB providing the better approximation.

For $l > 6$, neither the BFB nor the CPB performs well, and large basis sizes are required to reduce errors in the eigenfunctions even though eigenvalues have converged. As noted in Ref. 13, this is due to the eigenfunctions of case five inside the potential having oscillations of amplitude orders of magnitude greater than the eigenfunction outside.

VI. CONCLUSIONS

A novel method of solution for a non-self-adjoint Sturmian eigenvalue equation has been proposed. The method

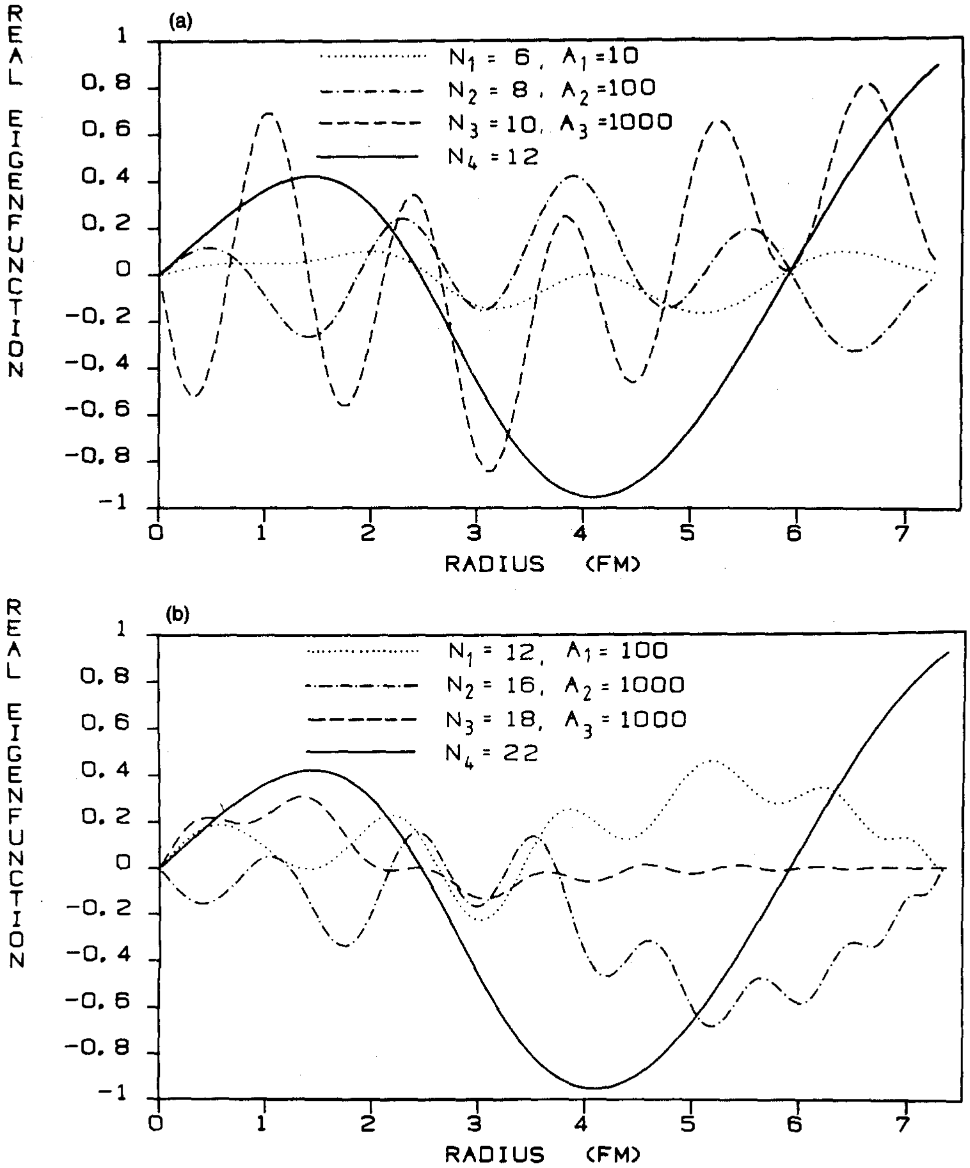


FIG. 3. Real part of the eigenfunction (unbroken line) and error curves (broken lines) for $l = 0, j = 1$ corresponding to the (a) Bessel function and (b) Chebyshev polynomial basis,[13] respectively. The values of N_S and A_S used in Eq. (17) are given in the symbol key; N_4 is the value used to generate the eigenfunction.

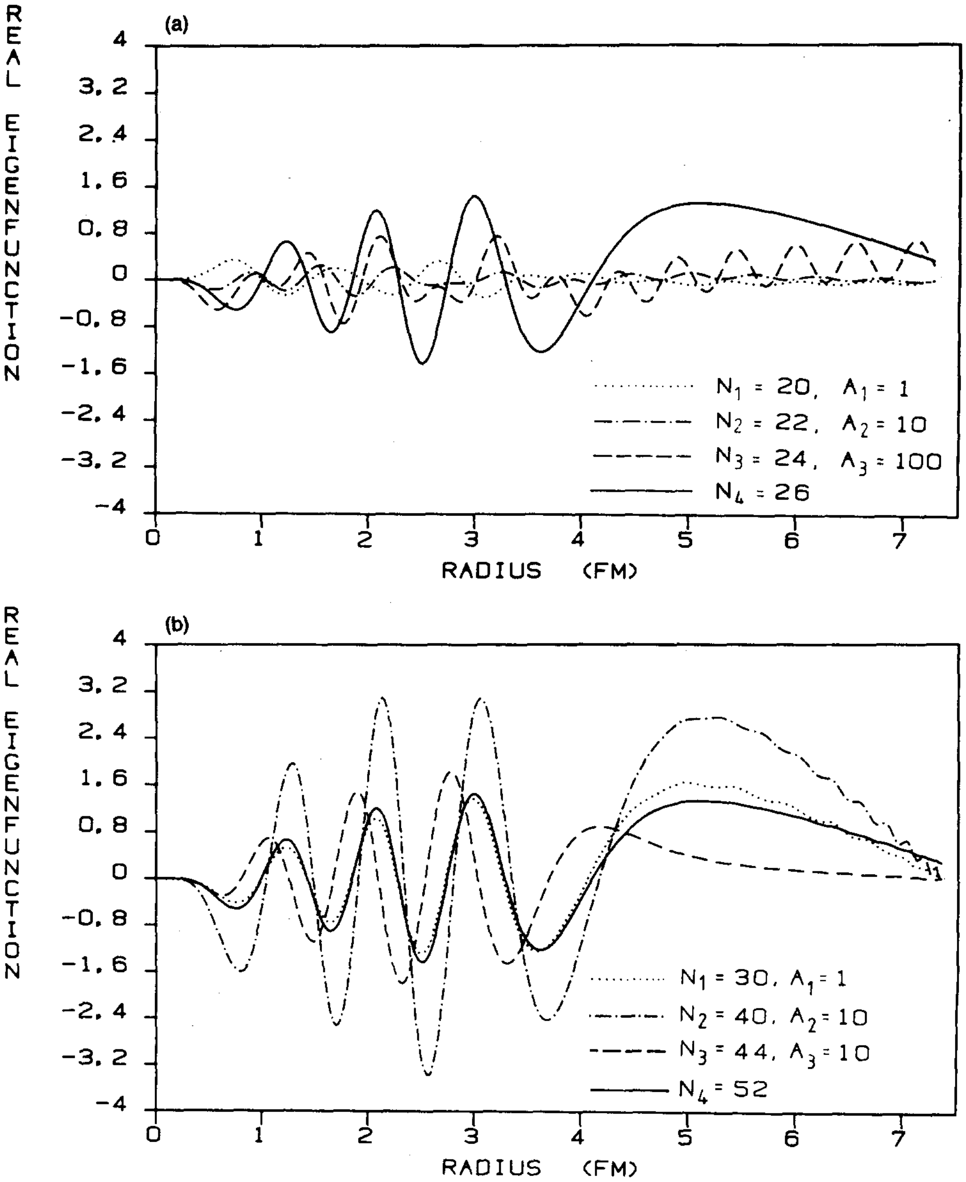


FIG. 4. Real part of the eigenfunction (unbroken line) and error curves (broken lines) for $l = 4$, and $j = 8$ corresponding to the (a) Bessel function and (b) Chebyshev polynomial basis,[13] respectively. The values of N_S and A_S used in Eq. (17) are given in the symbol key; N_4 is the value used to generate the eigenfunction.

consists of expansion of the eigensolution on a discrete basis of suitably normalized spherical Bessel functions of complex argument. The basis functions are members of a discrete set of functions corresponding to distinct roots of a complex transcendental equation obtained from an asymptotic boundary condition. Substitution of the expansion into the second-order differential equation of interest leads to a complex matrix eigenvalue problem that is solved by conventional techniques. The complex eigenvalues of the matrix are those of the required complex two-point boundary value problem. The corresponding eigenvectors are the expansion coefficients on the Bessel function basis.

The method has been compared in detail with one using a Chebyshev polynomial basis reported in another study[13] for a realistic case. Comparison of the rate of convergence for eigenvalues showed that the Bessel function basis method was approximately 70% more efficient in terms of the number of basis states when compared with the Chebyshev polynomial method. Convergence for eigenfunctions was investigated by comparison of error curves, on the interval of approximation, for different truncations of the Bessel function and Chebyshev polynomial bases. For small orbital angular momentum l and eigenvalues of small magnitude, the two types of bases produced similar errors for similar basis size. However, for larger l and eigenvalues of larger magnitude the Bessel function basis proved superior in that a substantially smaller number of basis functions was required for a prescribed error compared to the Chebyshev polynomial basis.

In conclusion, this and the previous report[13] have established the stability of two independent methods of generating numerically, in realistic cases, the elements v_{lj} of a set of basis functions used in the construction of finite rank approximations to non-self-adjoint integral operators. Investigation of the applicability of the Sturmian expansion method to scattering theory[9,12] is in progress.[8]

ACKNOWLEDGMENTS

G. Delic wishes to express his appreciation to Professor G. H. Rawitscher for his kind hospitality during a visit to the University of Connecticut in January 1983.

This research was supported in part by the National Science Foundation Grant No. PHY 8214571 and the Council for Scientific and Industrial Research (R.S.A.).

APPENDIX: ROOTS OF THE SQUARE WELL TRANSCENDENTAL EQUATION

In the case that the potential $\bar{U}$ in Eq. (2) is a square well, the solutions are spherical Bessel functions as given by Eq. (4). The argument of these functions is the complex variable $z = Kr$. At the matching point $r = a$, z is denoted as $Z = A + iB$.

The boundary condition (6) on the solution, Eq. (4), at $r = a$ for the square well case reduces to the requirement

$$D_l(Z, z_0) = 0 \tag{A1}$$

with

$$D_l(Z, z_0) = Z\,[j_{l-1}(Z)/j_l(Z)] - (l + z_0)\,, \tag{A2}$$

where

$$z_0 = \frac{a}{h_l^{(+)}(kr)} \left.\frac{dh_l^{(+)}(kr)}{dr}\right|_{r=a} = x_0 + iy_0\,. \tag{A3}$$

The roots $Z_{ln} = A_{ln} + i B_{ln}$ of Eq. (A1) are found by Newton's iteration method[14] in the complex plane, using as starting values $Z_{ln}^{(1)}$, which are arrived at by a method described below. Figure 5 illustrates the resulting values of A_{ln} and B_{ln} for the kinematic conditions of case five, with $a = 7.39$ fm and $k = 0.797\,215\,2\ \text{fm}^{-1}$. For large values of n and even l, A_{ln} is close to a half-integral multiple of π, while for odd l, A_{ln} lies close to an integral multiple of π. As n decreases, a bend in the trajectory of points (A_{ln}, B_{ln}) occurs (close to $A_{ln} \sim 2\pi$). This behavior can be understood by examining the behavior of $D_l(Z, z_0)$ For large values of Z the asymptotic form of the Bessel functions can be utilized, and we find[11] two branches for the solutions, denoted as (+) and (−). They lie close to the points $a_{ln}^{(+)}$ and $a_{ln}^{(-)}$ given by

$$a_{ln}^{(-)} = (2n + l - 1)(\pi/2)\,, \tag{A4}$$

$$a_{ln}^{(+)} = (2n + l)(\pi/2)\,. \tag{A5}$$

For the (−) branch the asymptotic solutions are

$$A_{ln}^{(-)} = a_{ln}^{(-)} + y_0^2/(a_{ln}^{(-)})^3 + l(l+1)/(2a_{ln}^{(-)})\,, \tag{A6}$$

$$B_{ln}^{(-)} = -y_0/a_{ln}^{(-)}\,. \tag{A7}$$

For the (+) branch these solutions are

$$A_{ln}^{(+)} = a_{ln}^{(+)}(1 - 1/y_0^2)\,, \tag{A8}$$

$$B_{ln}^{(+)} = -a_{ln}^{(+)}/y_0\,. \tag{A9}$$

Here x_0 and y_0 are the real and imaginary parts of z_0, defined in Eq. (A3).

The behavior of the solutions described above arises from the fact that asymptotically the Bessel functions in (A2) contain sines or cosines (of complex argument). The solutions are multivalued because the real part of the circular functions are oscillatory functions of the real part of the variable. A detailed discussion of the properties of $D_l(Z, z_0)$ is given elsewhere.[11]

The numerical procedure consists in starting the iterative solutions at large values of n on the (−) branch. The starting values $Z_{ln}^{(1)}$ are taken from Eqs. (A6) and (A7). The result is illustrated in Fig. 5 by the points lying to the right of the minimum, i.e., $A_{ln} > 3\pi$. The value of n is decreased successively by one unit, the previously found values of Z_{ln} serving to construct the guess for the next value of n.

This procedure is continued until we arrive near the bend of the curve, where the (+) branch of the solution is reached. Even though Z_{ln} is not so large that the asymptotic expressions for the Bessel functions are valid, it is nevertheless found that Eqs. (A8) and (A9) provide adequate guesses for the starting values of the iterative procedure.

The procedure developed usually requires three or four iterations in order to achieve an accuracy of 1 part in 10^8 for each n. For the parameter values used, i.e., $0<l<10$ and $10<ka<60$, no roots were found to be missing. The valleys of convergence around each root appear to be sufficiently broad to make the method reliable. The results have been checked for $l = 0$ and $l = 1$ by comparison with a graphic method.[11]

Another check lies in the comparison with the Chebyshev basis expansion for the eigenvalues and eigenfunctions of Eq. (1) for the square well case of Ref. 13.

The CPU time required on an IBM 360 model 168 com-

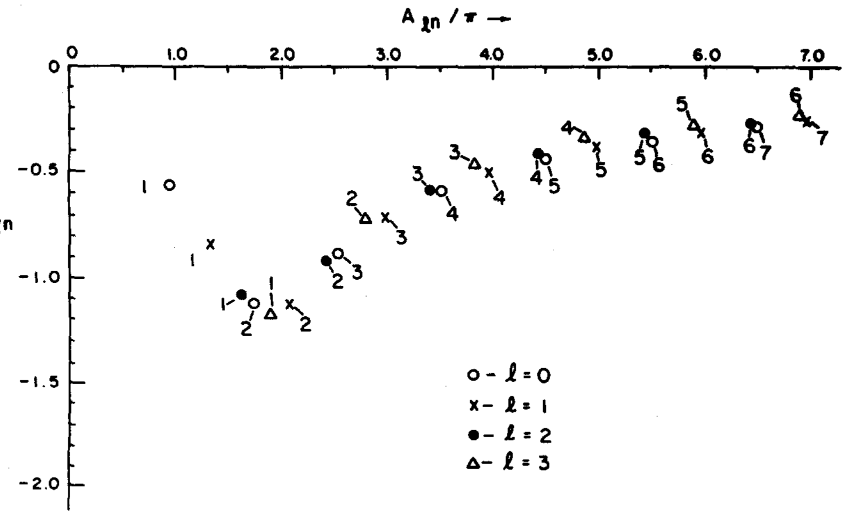


FIG. 5. Argand plot of the complex roots $A_{ln} + iB_{ln}$ of Eq. (A1) for l values 0 to 3 as indicated in the symbol key. The integers on the points are the values of n and the value of ka for this case was 5.891 420.

puter with double precision arithmetic is typically less than three seconds in order to obtain 1200 different complex roots of the transcendental equation.

[1]A. M. Lane and R. G. Thomas, Rev. Mod. Phys. **30**, 257 (1958); C. Mahaux and H. A. Weidenmüller, *Shell Model Approach to Nuclear Reactions* (North-Holland, Amsterdam, 1969), Chap. 8.

[2]G. Gamow, Z. Phys. **51**, 204 (1928).

[3]P. L. Kapur and R. Peierls, Proc. R. Soc. London Ser. A **166**, 277 (1938).

[4]T. Vertse, K. F. Pal, and Z. Balogh, Comput. Phys. Comm. **27**, 309 (1982).

[5]R. Landau, Phys. Rev. C **27**, 2191 (1983); E. Hernandez and A. Mondragon, Phys. Rev. C **29**, 722 (1984).

[6]B. Gyarmati, A. T. Kruppa, Z. Papp, and G. Wolf, Nucl. Phys. A **417**, 393 (1984).

[7]S. Weinberg, Phys. Rev. B **133**, 232 (1963); S. Weinberg, in *Lectures on Particles and Field Theory, Brandeis Summer Institute in Theoretical Physics*, edited by S. Deser and K. W. Ford (Prentice-Hall, Engelwood Cliffs, 1964), Vol. II, Chap. 5; R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966), pp. 198 and 282; G. H. Rawitscher, Phys. Rev. C **25**, 2196 (1982).

[8]G. H. Rawitscher, "Accuracy analysis of a Sturmian expansion method for the solution of scattering equations," to be submitted for publication.

[9]G. H. Rawitscher and G. Delic, Phys. Rev. C **29**, 747 (1984).

[10]G. H. Rawitscher and E. S. Hirschorn, "Accurate evaluation of an integral involving the product of two Bessel functions and a Gaussian," submitted to J. Comput. Phys.

[11]G. H. Rawitscher and R. Hadley, unpublished.

[12]G. H. Rawitscher and G. Delic, Phys. Rev. C **29**, 1153 (1984).

[13]G. Delic and G. H. Rawitscher, J. Comput. Phys. **57**, 188 (1985).

[14]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).

[15]H. M. Nussenzveig, Nucl. Phys. **11**, 499 (1959); S. Joffily, Nucl. Phys. A **215**, 301 (1973); P. Kaus and C. J. Pearson, Nuovo Cimento **28**, 500 (1963).

[16]G. H. Rawitscher, Phys. Rev. C **25**, 2196 (1982).

[17]B. T. Smith, J. M. Boyle, J. J. Dongarra, B. S. Garbow, Y. Ikebe, V. C. Klema, and C. B. Moler, *Matrix Eigensystem Routines—EISPACK Guide* (Springer, New York, 1976), 2nd ed.

A note on helicity and self-duality[a)]

Abhay Ashtekar[b)]
Physics Department, Syracuse University, Syracuse, New York 13210 and Physique Théorique, IHP, Université de Paris VI, 75231 Paris, France

(Received 19 August 1985; accepted for publication 28 October 1985)

It is pointed out that, even if restricted to only self-dual (or anti-self-dual) fields, photon and linearized graviton states of *both* helicities can be constructed by dropping the restriction to positive-frequency fields. Consequently, contrary to the usual belief, it may not be necessary to work with both self-dual and anti-self-dual fields to obtain the Hilbert space of all quantum states in full quantum gravity.

I. INTRODUCTION

For free spin-1 and spin-2 fields, the helicity operator is generally defined in textbooks[1] in the momentum representation. This representation is not available for self-interacting fields of Yang–Mills theory and general relativity. Therefore, recently, Birula *et al.*[2] translated the usual definition in terms of space-time fields themselves, without any reference to Fourier transforms, and obtained the relation

$$\hat{S} = -i|s|D \tag{1}$$

between the helicity operator $\hat{S}$ and the duality operator D, where s is the spin of the field. This relation immediately implies that self-dual fields (eigenvectors of D with eigenvalue $+i$) have positive helicity while the anti-self-dual fields (eigenvectors of D with eigenvalue $-i$) have negative helicity, in agreement with usage implicit in literature. Birula *et al.* then used Eq. (1) to *define* the helicity operator for nonlinear Yang–Mills and Einstein fields.

The purpose of this note is to point out an oversight: Already for linear fields, Eq. (1) holds *only* if one restricts oneself to positive frequencies. We shall see that, for negative-frequency fields, the correct helicity operator requires an extra minus sign on the right side of Eq. (1), so that negative-frequency, self-dual fields have *negative* helicity. *For a field that has both positive- and negative-frequency parts, therefore, there is no simple relation between duality and helicity.* Thus, although as pointed out in Ref. 2, Eq. (1) itself makes no reference to decomposition into momentum states, its domain of validity cannot be specified without recourse to the positive- and negative-frequency decomposition. Since the operation of taking positive-frequency parts of fields is nonlocal, contrary to appearances, the helicity operator is also nonlocal. Consequently, *a priori* it is not clear how to extend the definition of this operator to Yang–Mills and Einstein fields except in the asymptotic and weak field limits.

Section II discusses the main result. While the correction of the oversight in Ref. 2 serves only to clarify a technical issue, the final picture that emerges from this correction does have significant conceptual implications to certain programs that are being pursued in the general context of quantum gravity. These are briefly discussed in Sec. III.

II. HELICITY AND DUALITY

Let us begin with Maxwell fields in Minkowski space-time. Since we do not wish to tie ourselves to positive- or negative-frequency fields from the beginning, it is convenient to introduce the one-photon Hilbert space using real solutions of Maxwell's equations. Since details of this construction have appeared in the literature,[3] we will only recall the main steps without entering into issues of rigor involving functional analysis.

Denote by V the vector space of real solutions F_{ab} to Maxwell's equation in Minkowski space, which induce, on any Cauchy surface, C^∞ initial data of compact support. The vector space V is equipped with a natural symplectic structure Ω; Ω: $V \times V \to R$:

$$\Omega(F,\tilde{F}) := \int_\Sigma (\tilde{F}_{ab}A^b - F_{ab}\tilde{A}^b)\,ds^a, \tag{2}$$

where A_a is any smooth vector potential of F_{ab}. This symplectic structure governs the classical Poisson brackets as well as quantum commutators between Maxwell fields. To construct the Hilbert space of one-photon states, one needs to make V into a complex vector space. This is achieved by introducing[4] a linear operator J:

$$J\cdot F = iF^+ + (-i)F^-, \tag{3}$$

where $F^{\pm}_{ab}$ are the positive- and negative-frequency parts of F_{ab}. Note that, since F_{ab} is a real tensor field, so is $J\cdot F_{ab}$. However, since $J^2 = -1$, J can be thought of as the operation of multiplication by i. Thus, one can simply define

$$(a+ib)\cdot F := aF + bJ\cdot F, \tag{4}$$

for all real numbers a and b; Eq. (4) endows the space V of real Maxwell fields with the structure of a complex vector space. Next, it is easy to verify that the complex structure J so introduced is compatible with the symplectic structure Ω:

$$\langle F,\tilde{F}\rangle := (1/4\hbar)[\Omega(F,J\tilde{F}) + i\Omega(F,\tilde{F})] \tag{5}$$

is the Hermitian inner product on the complex vector space V. Denote the Cauchy completion of $(V,\langle\,,\,\rangle)$ by H. This H is the Hilbert space of one-photon quantum states. Thus, to obtain the one-photon Hilbert space, one has to introduce a complex structure J such that (V, J, Ω) is a Kahler space.

[a)] Dedicated to the memory of Dr. N. R. Gordon.
[b)] Alfred P. Sloan Research Fellow.

 0022-2488/86/030824-04$02.50

The J defined above is the unique such complex structure for which the natural action of the Poincaré group on the resulting H is unitary.

Let us now examine the relation between the Hilbert space H constructed above and the more familiar Hilbert space H^+ of positive frequency solutions F_{ab}^+ to Maxwell's equation. Consider the mapping Λ^+: $H \to H^+$ defined by $\Lambda^+ \cdot F = F^+$. The definition of J gives

$$\Lambda^+ \cdot J \cdot F = iF^+ = i\Lambda^+ \cdot F. \tag{6}$$

Thus, the operation of J on a real solution just corresponds to multiplying its positive frequency part by i. Furthermore, it is easy to show that

$$\begin{aligned}\langle F, \tilde{F} \rangle &\equiv (i/2\hbar)\Omega(F^-, \tilde{F}^+) \\ &= \frac{1}{(2\pi)^{3/2}} \cdot \frac{1}{\hbar} \int \hat{A}_a^*(\mathbf{K}, |\mathbf{K}|) \tilde{A}^a(\mathbf{K}, |\mathbf{K}|) \frac{d^3\mathbf{K}}{2|\mathbf{K}|} \\ &= \langle \Lambda^+ \cdot F, \Lambda^+ \cdot \tilde{F} \rangle_+ . \end{aligned} \tag{7}$$

Here, $\hat{A}_a(\mathbf{K}, K_0)$ is the Fourier transform of any potential A_a (of F_{ab}) satisfying the Lorentz gauge condition $\partial_a A^a = 0$,

$$\begin{aligned} A_a(X) = \frac{1}{2(\pi)^{3/2}} \int [&\hat{A}_a(\mathbf{K}, |\mathbf{K}|) e^{i\mathbf{K}\cdot\mathbf{x} - i|\mathbf{K}|t} \\ &+ \hat{A}_a(\mathbf{K}, -|\mathbf{K}|) e^{i\mathbf{K}\cdot\mathbf{x} + i|\mathbf{K}|t}] \frac{d^3\mathbf{K}}{2|\mathbf{K}|}, \end{aligned} \tag{8}$$

and $\langle\, , \rangle_+$ is the usual, textbook[1] inner product on the H^+. Thus, $(H, J, \langle\, , \rangle)$ is, via Λ^+, naturally isomorphic to $(H^+, i, \langle\, , \rangle_+)$.

Next, let us consider the mapping Λ^- that sends any F_{ab} to its negative frequency part F_{ab}^-. Now, we have

$$\Lambda^- J \cdot F = (-i)F^- = (-i)\Lambda^- \cdot F. \tag{9}$$

It is easy to verify that $\langle F, \tilde{F} \rangle = \langle \Lambda^- \cdot F, \Lambda^- \cdot \tilde{F} \rangle$, so that $(H, J, \langle\, , \rangle)$ is naturally isomorphic to $(H^-, (-i), \langle\, , \rangle_-)$. The extra minus sign in Eq. (9) relative to Eq. (6) will turn out to be crucial.

We are now ready to examine the unitary representation of the Poincaré group on H. Given by Killing vector field ξ^a on Minkowski space, we have a densely defined self-adjoint operator $\hat{\xi}$ and H:

$$\hat{\xi} \cdot F := -\hbar J \cdot L_\xi F, \tag{10}$$

where $L_\xi F \equiv L_\xi F_{ab}$ is the Lie derivative of F_{ab} with respect to (w.r.t.) ξ^a. [Note that Eq. (10) is a direct generalization of the expression $P_x = -\hbar i\, \partial/\partial x$ of the momentum operator, which generates space translations in the x direction in nonrelativistic quantum mechanics. The presence of the complex structure J in the right side of (10) ensures that ξ is a self-adjoint—rather than anti-self-adjoint—operator on H.] The four-momentum operator P_a, is given, in terms of its component $P_a t^a$, along any vector t^a by

$$P_a t^a = \hat{\mathbf{t}}, \tag{11}$$

where $\hat{\mathbf{t}}$ is the self-adjoint operator on H corresponding to the translational Killing vector defined by t^a. [The choice of sign in the definition (10) of ξ is dictated by the requirement that P_a be future—rather than past—pointing. Note that, in our convention, η_{ab} has signature $- + + +$.] The angular momentum operator M_{ab} is defined (w.r.t. an origin 0 in Minkowski space) in terms of its contraction $f^{ab}M_{ab}$ with any skew tensor f^{ab} by

$$f^{ab}M_{ab} := \hat{\xi}_f, \tag{12}$$

where ξ_f^a is the Lorentz Killing field $\xi_f^a = 2f^{ab}X_b$, X_b being the position vector (w.r.t. the origin 0) of the point at which ξ^a is evaluated. Finally, the Pauli–Lubanski spin vector operator S^a is given by

$$S^a := \tfrac{1}{2}\epsilon^{abcd} P_b M_{cd}, \tag{13}$$

where ϵ^{abcd} is the alternating tensor field defined by the Minkowskian metric η_{ab}. For classical zero rest mass particles, the spin vector is parallel to the four-momentum and the proportionality factor gives the helicity. We wish to find the corresponding helicity operator $\hat{S}$ on the Hilbert space H.

To find $\hat{S}$, we proceed as follows. Fix a constant vector field V^a on Minkowski space. Then the component $V_a S^a$ of the spin operator S^a on H is given by

$$\begin{aligned}(V_a S^a) \cdot F_{mn} &= \tfrac{1}{2}\epsilon^{abcd} V_a P_b \cdot M_{cd} \cdot F_{mn} \\ &= \epsilon^{abcd} V_a P_b \cdot (-\hbar J) \\ &\quad \cdot [X_d \nabla_c F_{mn} + F_{mc} \nabla_n X_d + F_{cn} \nabla_m X_d] \\ &= \epsilon^{abcd} V_a (-\hbar J)^2 \\ &\quad \cdot \nabla_b [X_d \nabla_c F_{mn} + F_{mc}\eta_{nd} + F_{cn}\eta_{md}] \\ &= \epsilon^{abcd} V_a (-\hbar J)^2 \\ &\quad \cdot [(\nabla_b F_{mc})\eta_{nd} + (\nabla_b F_{cn})\eta_{md}] \\ &= -\tfrac{1}{2}\epsilon^{abcd} V_a (-\hbar J)^2 \\ &\quad \cdot [(\nabla_m F_{cb})\eta_{nd} + (\nabla_n F_{bc})\eta_{md}] \\ &= V^a (-\hbar J)^2 \cdot [\nabla_m^* F_{an} + \nabla_n^* F_{ma}] \\ &= V^a (-\hbar J)^2 \cdot (\nabla_a^* F_{mn}) \\ &= V^a P_a (-\hbar J) \cdot {}^*F_{mn}, \end{aligned} \tag{14}$$

where ∇ is the derivative operator compatible with η_{ab}. (Here, we have used the source-free Maxwell equations; $\nabla_{[a}F_{bc]} = 0$ is used in the fifth step and $\nabla_{[a}^* F_{bc]} = 0$ in the seventh.) Thus, on the Hilbert space H of one-photon states, we are led to the defined $\hat{S}$:

$$\hat{S} \cdot F = -\hbar J \cdot D \cdot F, \tag{15}$$

where D is the duality operator, $D \cdot F_{ab} = {}^*F_{ab}$. Note that, although neither J nor D admits *real* Maxwell fields F_{ab} as eigenvectors, $\hat{S}$ does! If F_{ab} is such that F_{ab}^+ is self-dual (i.e., ${}^*F^+ = iF^+$), or, equivalently, F_{ab}^- is anti-self-dual, F_{ab} is an eigenvector of $\hat{S}$ with eigenvalue $+\hbar$, while if F_{ab}^+ is anti-self-dual, F_{ab} is an eigenvector of $\hat{S}$ with eigenvalue $-\hbar$. (It is interesting to note that $\hat{S}$ is, modulo $\hbar$, just the product of two complex structures J and D.)

One can repeat the above analysis for linearized gravity. The helicity operator on the Hilbert space of (linearized) gravitons turns out to be

$$\hat{S} = -2\hbar J \cdot D, \tag{16}$$

and, in general, we have

$$\hat{S} = -|s|\hbar J \cdot D, \tag{17}$$

where s is the spin of the field.

TABLE I. Relation between duality, frequency, and helicity.

Duality/Frequency	Positive-freq. fields	Negative-freq. fields
Self-dual fields	+ *ve* helicity	− *ve* helicity
Anti-self-dual fields	− *ve* helicity	+ *ve* helicity

The above framework based on real solutions is well suited to compare and contrast results emerging from the use of positive and negative frequencies. Using the isomorphism Λ^+ between H and H^+ we can transport various operators on H, associated with the Poincaré action, to H^+. Equations (6) and (10) imply that the Poincaré generator $\hat{\xi}$ on H^+, associated with a Killing field ξ^a, is given by

$$\hat{\xi} \cdot F^+ := \Lambda^+ \cdot \hat{\xi} \cdot (\Lambda^+)^{-1} \cdot F^+ = (\hbar/i) L_\xi F^+, \quad (18)$$

whence it follows that the helicity operator on H^+ is

$$\hat{S} \cdot F^+ = |s| (\hbar/i) D \cdot F^+. \quad (19)$$

Hence if a positive-frequency field F^+ is self-dual, it is an eigenvector of $\hat{S}$ with *positive* eigenvalue, $|s|\hbar$. On the negative-frequency Hilbert space H^-, on the other hand, we have

$$\hat{\xi} \cdot F^- := \Lambda^- \cdot \hat{\xi} \cdot (\Lambda^-)^{-1} \cdot F^- = i\hbar L_\xi F^-, \quad (20)$$

and, consequently, the helicity operator is given by

$$\hat{S} \cdot F^- = |s| i\hbar D \cdot F^-. \quad (21)$$

Thus, negative-frequency, self-dual fields are eigenvectors of the helicity operator with the negative eigenvalue $-|s|\hbar$.

Our results can be summarized in Table I.

Remarks: (i) In our presentation, we purposely avoided tying ourselves to positive- or negative-frequency fields from the beginning. Instead, we began with real Maxwell fields, constructed the one-particle Hilbert space, introduced the correct Poincaré generators, and *then* translated our results to the positive- and negative-frequency Hilbert spaces $H^\pm$. However, one could have also just begun with H^+ and H^-. How would one then know that, given a Killing field ξ^a on Minkowski space, the corresponding self-adjoint generator ξ on $H^\pm$ [Eqs. (18) and (20)] should differ by a relative sign? The answers lies in the fact that, on H^+ as well as H^-, the choice of sign is forced upon us by the requirement that the four-momentum operator P_a should be future pointing (i.e., that $t^a P_a$ should be a negative definite operator for all future pointing time translations t^a; recall that η_{ab} has signature $- + + +$).

(ii) If one works with spaces of solutions to nonlinear equations such as Yang–Mills or Einstein, one cannot repeat the above analysis because the space of solutions to these equations does not have a natural vector-space structure. However, one can then work with asymptotic states. This can be done, *without linearizing the equations*, by using the structure at future *or* past null infinity $I^\pm$ of space-time.[5] On $I^\pm$, one can isolate the radiative modes of the exact, nonlinear theories. The space of radiative modes has a natural affine space structure. One can use this, together with the action of the symmetry group at $I^\pm$ to obtain a Hilbert space of one-particle states. These are asymptotic states of the exact theories. One can then introduce the four-momentum and helicity operators. Although in technical details this construction differs from the one given in this paper, the final results are the same. The asymptotic Yang–Mills particles and gravitons have zero mass, and spin-1 and -2, respectively. The relation between positive and negative frequencies, duality, and helicity is the same as in Table I.[6]

III. DISCUSSION

The belief that self-dual fields always correspond to positive helicity and anti-self-dual ones to negative helicity was implicit in literature on self-dual and anti-self-dual solutions to (Lorentzian) Yang–Mills and Einstein equations for several years before the publication of Ref. 2, and has had significant impact on the general way of thinking on many problems. In particular, it had led one to believe that it is *essential* to have both self-dual and anti-self-dual fields to incorporate gravitons of both helicities in quantum gravity. This scenario has been a major motivation behind attempts to understand the "interaction" between the H-spaces and the dual H-spaces as well as efforts aimed at "combining" self-dual and anti-self-dual solutions to Einstein's equation using twistor methods.[7] As we have seen, the correct picture in linear theories is that one can incorporate states of both helicities in either of two ways: one can work with positive frequency fields, both self-dual *and* anti-self-dual ones; or, one can work just with self-dual fields without any restriction on frequency. Consequently, in quantum gravity, *a priori*, there are two possible types of avenues. The first would involve the introduction of a generalization of the notion of positive-frequency fields and a study of the interaction between *positive-frequency*—rather than arbitrary—self-dual and anti-self-dual configurations. The second possibility is to work with *all* self-dual configurations (or, "self-dual parts" of real configurations[8]) without worrying about frequencies. This attractive strategy would not have been viable if one were forced to use both self-dual and anti-self-dual fields to incorporate both helicities.

ACKNOWLEDGMENTS

Discussions with Kumar Narain and Phil Yasskin motivated this investigation.

The work was supported in part by the NSF Grant No. PHY8310041 to Syracuse University and by the Alfred P. Sloan Foundation.

[1]See, e.g., W. W. Frazer, *Elementary Particles* (Prentice–Hall, Englewood Cliffs, NJ, 1968); P. Roman, *Introduction to Quantum Field Theory* (Wiley, New York, 1969); I. Bialynicki-Birula and Z. Bialynicka-Birula, *Quantum Electrodynamics* (Pergamon, Oxford, 1975).

[2]I. Birula, E. T. Newman, J. Porter, J. Winicour, B. Lukacs, Z. Perjes, and A. Sebestyen, J. Math. Phys. **22**, 2530 (1981).

[3]I. E. Segal, *Mathematical Problems of Relativistic Physics* (Am. Math. Soc., Providence, RI, 1963), Chap. III, and references therein; A. Magnon, "*Sur la quantification des champs en espace-temps courbes*," these d'état, Université de Marseille, 1978.

[4]Actually, as defined, J does not leave V invariant. One must first Cauchy-complete V with respect to the inner product of Eq. (7) and then introduce J on the completed space. For details, see Ref. 3.

[5]See, e.g., S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1974), pp. 118–124.

[6]A. Ashtekar, Phys. Rev. Lett. **46**, 573 (1981). Details are reported in A. Ashtekar, *Notes on Quantum Gravity* (Bibliopolis, Naples, in press), Chap. II.C.

[7]Another major motivation came purely from classical general relativity: to find "the general solution" to Einstein's equation.

[8]More precisely, one would work with the usual phase space of general relativity and introduce a (complex) polarization on it that contains precisely the (appropriately defined) anti-self-dual directions. Then, the quantum wave functions would depend only on the self-dual part of the curvature. It is conceivable that one can introduce an appropriate Hermitian inner product on the space of these states without having to introduce the analog of the complex structure J, or, of the positive- and negative-frequency decomposition. If this turns out to be possible, one would obtain the analog of the full Fock space of both helicity gravitons. However, one would be able to introduce the notion of particle number, spin, mass, and helicity only in the asymptotic or the weak field limit. For details, see A. Ashtekar, *Notes on Quantum Gravity* (Bibliopolis, Naples, in press), Chap. III.E.

On the two-point functions of interacting Wightman fields[a)]

Klaus Baumann

Institut für Theoretische Physik der Universität Göttingen, Bunsenstrasse 9, D-3400 Göttingen, Federal Republic of Germany

(Received 23 July 1985; accepted for publication 28 October 1985)

Let A be a relativistic local field. If its two-point function in momentum space $\widetilde{W}_2(p)$ has a falloff such that for some $\alpha > 0$, $\widetilde{W}_2(p)e^{\alpha\sqrt{p^2}}$ is still a tempered distribution, then A is necessarily a generalized free field.

I. INTRODUCTION

Let A be a relativistic local field in n space-time dimensions. We assume that A fulfills all Wightman axioms.[1]

It is well known, since the early days of axiomatic quantum field theory, that if the two-point function of a scalar field A in momentum space $\widetilde{W}_2(p)$ has the form $\widetilde{W}_2(p) = \theta(p^0)\delta(p^2 - m^2)$, then A is a free field of mass m (Jost–Schroer theorem[2]). Of course this is true for integer spins, too. Some years later Greenberg[3] and Borchers (unpublished) proved independently that if $\widetilde{W}_2(p) \equiv 0$ for $p^2 > M^2 > 0$, then A is a generalized free field. In 1966 Vasilev[4] got the same result if $\widetilde{A}(p)$ decreased like $e^{-\mu|p^2|}$, $\mu > 0$. In this paper we extend these results further: If there exists an $\alpha > 0$ such that $\widetilde{W}_2(p)e^{\alpha\sqrt{p^2}} \in S'$, then A has to be a generalized free field. We want to emphasize that for this result we do not have to assume that A transforms finite covariantly under Lorentz transformations. So this result remains true even if A has infinitely many components.

The paper is organized as follows: In Sec. II we give the precise formulation and the proof of our result. This will be applied in Sec. III to solutions of the wave equation. We show how restrictions to timelike planes of such solutions cannot look alike. Therefore this part may be of independent interest.

II. MAIN THEOREM

Theorem 1: Let $A(x)$ be a relativistic quantum field in n space-time dimensions, obeying all Wightman axioms. If there exists an $\alpha > 0$ such that we have the bound

$$\widetilde{W}_{A^+A}(p)e^{8\alpha\sqrt{p^2}} \in S'(\mathbb{R}^n)$$

for the Fourier transformed two-point function, then (i) $A(x)$ is a generalized free field, if $n \geqslant 3$, and (ii) for $n = 2$, $A(x)$ is a generalized free field, if there are no zero mass states in the energy–momentum spectrum.

Proof:

(1) By the very assumption $\widetilde{A}(p)\Omega e^{4\alpha\sqrt{p^2}}$ defines a vector-valued tempered distribution.

(2) In the following we shall consider the matrix element

$$\widetilde{F}(p,q) = \underbrace{\left(\psi,\left\{\widetilde{A}\left(\frac{p-q}{2}\right)\widetilde{A}\left(\frac{p+q}{2}\right)\right.\right.}_{\widetilde{F}_+(p,q)} \mp \underbrace{\left.\left.\widetilde{A}\left(\frac{p+q}{2}\right)\widetilde{A}\left(\frac{p-q}{2}\right)\right\}\Omega\right)}_{\widetilde{F}_-(p,q)},$$

$$= \widetilde{F}_+(p,q) \mp \widetilde{F}_-(p,q),$$

where the minus, resp. plus, sign has to be chosen for bosons, resp. fermions, and ψ is any vector with compact momentum in the domain of $A^+(f)$. By the spectrum condition we get

(i) $p \in \overline{V}^+ \cap \operatorname{supp}\{E(p)\psi\}$,

(ii) $\operatorname{supp} \widetilde{F}_+(p,\cdot) \subseteq -p + \overline{V}^+$,

(iii) $\operatorname{supp} \widetilde{F}_-(p,\cdot) \subseteq p - \overline{V}^+$.

We have to show $\widetilde{F}(p,q) \equiv 0$ for $p \in \overline{V}^+ \setminus \{0\}$ and all ψ.

(3) Using the Cauchy–Schwartz inequality we see that

(iv) $\widetilde{F}_+(p,q)e^{2\alpha\sqrt{(p+q)^2}}$

and

(v) $\widetilde{F}_-(p,q)e^{2\alpha\sqrt{(p-q)^2}}$

exist as tempered distributions. This implies immediately the existence of

$$\widetilde{F}(p,q)\cosh\sigma\sqrt{q^2}, \quad \sigma \in \mathbb{R}, \quad |\sigma| < 2\alpha$$

as a tempered distribution. For example, for $q \in -p + \overline{V}^+$ we have

$$\widetilde{F}_+(p,q)\cosh\sigma\sqrt{q^2} = \underbrace{\widetilde{F}_+(p,q)\cosh 2\alpha\sqrt{(p+q)^2}}_{\text{exists as a distribution}} \; \underbrace{\frac{\cosh\sigma\sqrt{q^2}}{\cosh 2\alpha\sqrt{(p+q)^2}}}_{C^\infty \text{ and all derivatives are bounded by polynomials on supp } \widetilde{F}_+}.$$

[a)] This work contains parts of the author's "Habilitationsschrift," accepted by the Physics Department, University of Göttingen.

(4) For our convenience we define

$$\tilde{G}_p(\sigma,q) := \tilde{F}(p,q)\cosh\sigma\sqrt{q^2}$$

and

$$G_p(\sigma,x) := \left(\frac{1}{2\pi}\right)^{n/2}\int e^{iqx}\tilde{G}_p(\sigma,q)d^nq.$$

Lemma 1: (i) $G_p(\sigma,x)$ is, for $|\sigma|\leqslant 2\alpha$, a weak solution of the ultrahyperbolic equation

$$(\partial_\sigma^2 + \partial_{x_0}^2 - \partial_{x_1}^2 - \cdots - \partial_{x_{n-1}}^2)G_p(\sigma,x) = 0.$$

(ii) $G_p(\sigma,x) = 0$, if $x^2 < 0$, $|\sigma| < 2\alpha$.

Proof: (a) $\partial_\sigma^2\tilde{G}_p(\sigma,q) = q^2\tilde{G}_p(\sigma,q)$, which is equivalent to (i).

(b) $G_p(0,x)$

$$= \left(\frac{1}{2\pi}\right)^{n/2}\int e^{iqx} \times\left(\psi,\left[\tilde{A}\left(\frac{p-q}{2}\right),\tilde{A}\left(\frac{p+q}{2}\right)\right]_\mp\Omega\right)d^nq$$

$$= \left(\frac{1}{2\pi}\right)^{n/2}\int e^{ipy} \times(\psi,[A(y+x),A(y-x)]_\mp\Omega)d^ny$$

$$= 0, \quad \text{if } x^2 < 0$$

by the locality of $[A(x),A(y)]_\mp$;

$$(\partial_\sigma\tilde{G}_p)(\sigma,q) = \tilde{F}(p,q)\sqrt{q^2}\sinh\sigma\sqrt{q^2}$$

and therefore

$$(\partial_\sigma G_p)(0,x) \equiv 0.$$

(c) By repeated use of the mean value theorem of Asgeirson[5] for ultrahyperbolic equations, we get for $|\sigma| < 2\alpha$,

$$G_p(\sigma,x) = 0, \quad \text{if } \sigma^2 + x_0^2 - \mathbf{x}^2 < 0.$$

(d) For fixed values of σ we have $G_p(\sigma,x) = 0$, if $x^2 < -\sigma^2$. But $G_p(\sigma,x) = G_p^+(\sigma,x) \mp G_p^-(\sigma,x)$, where $G_p^+(\sigma,x)$, resp. $G_p^-(\sigma,x)$, are boundary values of functions analytic in $\mathbb{R}^n + iV^+$, resp. $\mathbb{R}^n - iV^+$. By the double cone theorem we get $G_p(\sigma,x) = 0$, if $x^2 < 0$.

(5) For smooth functions $f\in\mathscr{D}([-\alpha,\alpha])$ we define

$$G_p^f(\sigma,x) := \int f(\tau)G_p(\sigma-\tau,x)d\tau.$$

The following properties are obvious for $|\sigma| < \alpha$:

(i) $(\partial_\sigma^2 + \partial_{x_0}^2 - \Delta_{\mathbf{x}})G_p^f(\sigma,x) = 0$,

(ii) $G_p^f(\sigma,x) = 0$, if $x^2 < 0$,

(iii) $$\tilde{G}_p^f(\sigma,q) = \int f(\tau)\tilde{G}_p(\sigma-\tau,q)d\tau = \tilde{F}(p,q)\{\tilde{f}_s(\sqrt{-q^2})\cosh\sigma\sqrt{q^2} + \tilde{f}_A(\sqrt{-q^2})\sinh\sigma\sqrt{q^2}\}.$$

By this convolution we have achieved that $\tilde{G}_p^f(\sigma,q)$ decreases quite fast in all directions of q space!

(6) *Lemma 2:* $G_p^f(\sigma,x) = 0$, for $|\sigma| < \alpha$ and all x.

Outline of the proof: We shall show first that $G_p^f(\sigma,x)$ is a C^∞ function. Together with the locality (ii), this implies

$$(\partial_\sigma^k\partial_{x_0}^l G_p^f)(0,0,\mathbf{x}) \equiv 0, \text{ for all } k, l \text{ and all } \mathbf{x}.$$

Then we use a theorem by Strichartz[6] on solutions of the ultrahyperbolic equation (i), which states that the vanishing of all these derivatives rules out all nontrivial solutions.

Proof: (a) Let us treat first the case where p varies only over V^+, e.g., by using testing functions $\tilde{h}(p)$ with $\operatorname{supp}\tilde{h}\subset V^+$ and defining $G_h^f(\sigma,x) = \int\tilde{h}(p)G_p^f(\sigma,x)d^np$.

(b) We claim that $G_h^f(\sigma,x)$ is a C^∞ function:

$$G_p^f(\sigma,x) = \int\tilde{F}(p,q)\{\tilde{f}_S(\sqrt{-q^2})\cosh\sigma\sqrt{q^2} + \tilde{f}_A(\sqrt{-q^2})\sinh\sigma\sqrt{q^2}\}e^{iqx}d^nq,$$

and $\tilde{F}(p,q) = \tilde{F}_+(p,q) \mp \tilde{F}_-(p,q)$. We decompose $\tilde{F}_+(p,q)$ further into

$$\tilde{F}_+(p,q) = \tilde{F}_+^{\mathrm{I}}(p,q) + \tilde{F}_+^{\mathrm{II}}(p,q),$$

with

(i) $\operatorname{supp}\tilde{F}_+^{\mathrm{I}}(p,\cdot) \subseteq -\frac{2}{3}p + \bar{V}^+$,

and

(ii) $\operatorname{supp}\tilde{F}_+^{\mathrm{II}}(p,\cdot) \subseteq \{-p+\bar{V}^+\}\setminus\{-\frac{1}{3}p + V^+\}$,

which is obviously possible. Now we have

$$\int\underbrace{\tilde{F}_+^{\mathrm{I}}(p,q)\cosh 2\alpha\sqrt{(p+q)^2}}_{\text{distribution}}\;\underbrace{\frac{\tilde{f}_S(\sqrt{-q^2})\cosh\sigma\sqrt{q^2} + \tilde{f}_A(\sqrt{-q^2})\sinh\sigma\sqrt{q^2}}{\cosh 2\alpha\sqrt{(p+q)^2}}}_{\substack{\text{restricted to } \operatorname{supp}\tilde{F}_+^{\mathrm{I}}(p,\cdot),\\ \text{this is a testing function}}}e^{iqx}d^nq$$

and

$$\int\tilde{F}_+^{\mathrm{II}}(p,q)\;\underbrace{\{\tilde{f}_S(\sqrt{-q^2})\cosh\sigma\sqrt{q^2} + \tilde{f}_A(\sqrt{-q^2})\sinh\sigma\sqrt{q^2}\}e^{iqx}d^nq.}_{\substack{\text{restricted to } \operatorname{supp}\tilde{F}_+^{\mathrm{II}}(p,\cdot),\text{ this is a testing}\\ \text{function because of the rapid decrease of}\\ \tilde{f}\text{ for real arguments}}}$$

Therefore after integrating with $\tilde{h}(p)$ both integrals define C^∞ functions. For $\tilde{F}_-(p,q)$ we use an analogous decomposition. In this way we have shown that $G_h^f(\sigma,x)$ is a C^∞ function in σ and x.

(c) Now we make use of the following theorem given by Strichartz.[6]

Theorem: Let $u(y)\in\mathscr{D}'(\mathbf{R}^3)$ be a weak solution of

$$(\partial^2_{y_0}-\partial^2_{y_1}-\partial^2_{y_2})u(y)=Mu(y),\quad M\in\mathbf{R}.$$

If $(\partial^k_{y_1}\partial^l_{y_2}u)(y_0,0,0)\equiv 0$, for all k,l, then $u(y)\equiv 0$.

If the space-time dimension is 2 ($n=2$), we have

$$(\partial^2_\sigma+\partial^2_{x_0}-\partial^2_{x_1})G^f_h(\sigma,x)=0$$

and because $G^f_h(\sigma,x)$ is a C^∞ function that vanishes if $x^2<0$, we have $(\partial^k_\sigma\partial^l_{x_0}G^f_h)(0,0,x_1)\equiv 0$, for all k,l. By identifying $y_0=x_1$, $y_1=x_0$, $y_2=\sigma$, and putting $M=0$ we get from a local version of the above theorem $G^f_h(\sigma,x)\equiv 0$. If the space-time dimension is greater than 2 ($n\geqslant 3$), we do a Fourier transformation with respect to $x_2,\dots,x_{n-1}$:

$$H^f_h(\sigma,x_0,x_1|q_2,\dots,q_{n-1}) := \int G^f_h(\sigma,x)\exp -i(x_2q_2+\cdots+x_{n-1}q_{n-1}) \times dx_2\cdots dx_{n-1}$$

and H^f_h is analytic with respect to $q_2,\dots,q_{n-1}$ because the support of $G^f_h(\sigma,\cdot)$ is contained in $\overline{V}^+\cup\overline{V}^-$. But $H^f_h(\sigma,x_0,x_1|q_2,\dots,q_{n-1})=0$, if $x_0^2-x_1^2<0$ and is still a C^∞ function. Therefore we can apply the above theorem with $M=\Sigma^{n-1}_{k=2}q^2_k$ and we get $H^f_h\equiv 0$.

(d) Up to now we have shown that under the assumptions of Theorem 1, $\widetilde{F}(p,q)$ vanishes for $p\in V^+$. Therefore the support of $\widetilde{F}(\cdot,q)$ is contained in $\{p^2=0,p_0>0\}$.

Let us treat the case $n\geqslant 3$ first: Assume supp $\tilde{h}$ is contained in a small neighborhood of $p=p_0(1,1,0,\dots,0)$. Then $\widetilde{G}^f_h(\sigma,q)$ will not decrease for large q's if q_0 is approximately equal to q_1 because $(\cosh 2\alpha\sqrt{(p+q)^2})^{-1}$ does not decrease if p and q are parallel and lightlike. To circumvent this difficulty we take $g\in\mathscr{D}(\mathbf{R})$ and consider instead of $\widetilde{G}^f_h(\sigma,q)$, the product $\widetilde{G}^f_h(\sigma,q)\tilde{g}(q_1)$, which will decrease for large q_1. In all other directions we argue as earlier. Its Fourier transform

$$\int G^f_h(\sigma,x_0,x_1-y,x_2,\dots,x_{n-1})g(y)dy$$

is therefore a C^∞ function, fulfills the ultrahyperbolic equation, and vanishes if $x_0^2-x_2^2-\cdots-x^2_{n-1}<0$. By continuity we get

$$\int(\partial^k_\sigma\partial^l_{x_0}G^f_h)(0,0,x_1-y,x_2,\dots,x_{n-1})g(y)dy\equiv 0,$$

for all values of $x_1,x_2,\dots,x_{n-1}$. Then we proceed as in part (c). The above trick is obviously only possible if $n>2$.

(e) What happens in two-dimensional space-time? The following example shows that certain peculiarities show up in two-dimensional models.

Take a free vector field $j^\mu(x)$ with canonical dimension 1 obeying the conservation laws $\partial_\mu j^\mu=0=\partial_\mu\epsilon^\mu{}_\nu j^\nu$. Therefore

$$j^0=j_+(\xi)+j_-(\eta),\quad j^1=j_+(\xi)-j_-(\eta),$$

where ξ and η denote the light cone coordinates. The commutation relations are

$$(1/i)[j_+(\xi),j_+(\xi')]=-\delta'(\xi'-\xi),$$

$$(1/i)[j_-(\eta),j_-(\eta')]=-\delta'(\eta'-\eta),$$

$$[j_+(\xi),j_-(\eta)]=0.$$

Now take the Wick products $\Theta_+(\xi)=\frac{1}{2}{:}j^2_+{:}(\xi)$ and $\Theta_-(\eta)=\frac{1}{2}{:}j^2_-{:}(\eta)$. These local fields still obey the massless Klein–Gordon equation $\Box\Theta_+=0=\Box\Theta_-$ because $\Box=4\partial_\xi\partial_\eta$ but both are genuine Lie fields and we have

$$(1/i)[\Theta_+(\xi),\Theta_+(\xi')] = (1/24\pi)\delta'''(\xi'-\xi) - \delta'(\xi'-\xi)[\Theta_+(\xi)+\Theta_+(\xi')]$$

and a similar relation for $\Theta_-(\eta)$.

Therefore, we have to exclude zero mass states in two space-time dimensions in the sense that

$$\left[\widetilde{A}\left(\frac{p-q}{2}\right),\widetilde{A}\left(\frac{p+q}{2}\right)\right]_\mp\Omega=0,$$

for all $p\in V^+$ implies

$$\left[\widetilde{A}\left(\frac{p-q}{2}\right),\widetilde{A}\left(\frac{p+q}{2}\right)\right]_\mp\Omega =\left(\Omega,\left[\widetilde{A}\left(\frac{p-q}{2}\right),\widetilde{A}\left(\frac{p+q}{2}\right)\right]_\mp\Omega\right)\cdot\Omega.$$

III. REMARK ON THE WAVE EQUATION

There is a remarkable connection between local distributions in n dimensions and weak solutions of the wave equation in $n+1$ dimensions given by Gårding (see Ref. 5).

Let F_1 and F_2 be local distributions, i.e., $F_i(\xi)=0$ if $\xi^2<0$, then

$$\widetilde{G}(q,q_n):=\int\left\{F_1(\xi)\cos(q_n\sqrt{\xi^2}) + F_2(\xi)\frac{\sin(q_n\sqrt{\xi^2})}{\sqrt{\xi^2}}\right\}e^{iq\xi}d\xi$$

is a solution of the $(n+1)$-dimensional wave equation

$$\left(\frac{\partial^2}{\partial q_0^2}-\sum_{i=1}^n\frac{\partial^2}{\partial q_i^2}\right)\widetilde{G}(q,q_n)=0,$$

with

$$\widetilde{G}(q,0)=\widetilde{F}_1(q)\text{ and }\frac{\partial\widetilde{G}}{\partial q_n}(q,0)=\widetilde{F}_2(q).$$

(This looks like an "initial value problem" on the timelike plane $q_n=0$!) The converse is true, too. The restriction of a tempered solution of the wave equation to a timelike plane is the Fourier transform of a local distribution.

Therefore, by similar methods as used in the above proof, we get the following theorem.

Theorem 2: Let $\widetilde{G}\in S'(\mathbf{R}^{n+1})$ fulfill the wave equation

$$\left(\frac{\partial^2}{\partial q_0^2}-\sum_{i=1}^n\frac{\partial^2}{\partial q_i^2}\right)\widetilde{G}(q,q_n)=0.$$

If the restrictions $\widetilde{F}_1(q)$ and $\widetilde{F}_2(q)$ of $\widetilde{G}$ and of $\partial\widetilde{G}/\partial q_n$ to the plane $q_n=0$ fulfill the conditions (i) there is a $p\in\overline{V}^+$ such that $\widetilde{F}_i=\widetilde{F}_{i+}-\widetilde{F}_{i-}$, $i=1,2$, with supp $\widetilde{F}_{i+}\subseteq -p+V^+$ and supp $\widetilde{F}_{i-}\subseteq p-\overline{V}^+$; and (ii) there is an $\alpha>0$ and a $\beta>0$ such that

$$\widetilde{F}_{i+}(q)e^{\alpha\sqrt{(q+\beta p)^2}}\in S'(\mathbf{R}^n),$$

$$\tilde{F}_{i-}(q)e^{\alpha\sqrt{(q-\beta p)^2}} \in S'(\mathbf{R}^n);$$

then $\tilde{G} = 0$.

ACKNOWLEDGMENT

I want to thank very much Dr. W. Lücke for his encouraging help.

[1]R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964); R. Jost, *The General Theory of Quantized Fields* (Am. Math. Soc., Providence, RI, 1965).

[2]R. Jost, "Properties of Wightman functions," in *Lectures on Field Theory and the Many-Body Problem*, edited by E. R. Caianello (Academic, New York, 1961).

[3]O. W. Greenberg, "Heisenberg fields which vanish on domains of momentum space," J. Math. Phys. 3, 859 (1962).

[4]A. N. Vasilev, "A characteristic of the generalized free field," Sov. Phys. JETP 22, 543 (1966).

[5]A. S. Wightman, "Analytic functions of several complex variables, part VII," in *Dispersion Relations and Elementary Particles*, edited by C. De Witt and R. Omnès (Hermann, Paris, 1960).

[6]R. S. Strichartz, "The stationary observer problem for $\Box u = Mu$ and related equations," J. Differential Equations 9, 205 (1971).

Hyperfunctions and renormalization

Shigeaki Nagamachi
Technical College, Tokushima University, Tokushima 770, Japan
Nobumichi Mugibayashi
Department of Electrical Engineering, Kobe University, Kobe 657, Japan

(Received 10 July 1985; accepted for publication 28 October 1985)

A condition is found so that the Wick-ordered power series of scalar fields in four-dimensional space-time is defined as a Fourier-hyperfunction field, and the derivative coupling model is investigated in the framework of hyperfunction quantum field theory.

I. INTRODUCTION

Since Wightman and Gårding[1] formulated the quantum field theory in an axiomatic way by regarding fields as operator-valued tempered distributions, many authors have attempted to generalize the theory to take in more fields represented by more singular generalized functions by restricting the class of test functions. Among others, we only mention the strictly localizable field theory of Jaffe,[2] and the hyperfunction quantum field theory of the present authors.[3] For the choice of test functions in quantum field theory, the reader should refer to a short review of Wightman.[4] One of our motivations to generalize the theory in this way is the wish to manage so-called nonrenormalizable fields in some appropriate framework, since it has been said that the exclusion of nonrenormalizable fields from the theory is due to the axiom of temperedness. For example, the model of neutral scalar field with derivative coupling in four-dimensional space-time is concerned with the interaction of the second kind[5] and usually classified as nonrenormalizable by a simple power-counting argument. In the present paper we revisit this model in the new light of the hyperfunction quantum field theory and it will be shown that its Wightman functions are inevitably not tempered distributions. This problem was once treated quite formally by Okubo.[6]

The paper is organized as follows. In Sec. II we examine the singularity of the two-point Wightman function of neutral scalar field. The main result of Sec. II is that entire functions of the two-point function are well defined as Fourier hyperfunctions. In Sec. III, Wick-ordered entire functions of the free field operator are studied. A condition is found so that the Wick-ordered formal power series of the free field has well-defined Wightman functions in the sense of the Fourier hyperfunction (Theorems 3.3 and 3.4), and it is shown that, thanks to the equivalence between Euclidean field theory and Minkowski (hyperfunction) quantum field theory (see Nagamachi and Mugibayashi[7]), these Wightman functions satisfy the axioms of Ref. 3. In Sec. IV, the derivative coupling model is investigated. To accomplish the infinite renormalization rigorously, we use the Euclidean lattice formulation with an infinitesimal lattice spacing and represent it in the language of nonstandard analysis. It is proved that this model is renormalizable by the infinite field strength renormalization and its Wightman functions satisfy the axioms of Ref. 3, but they are never tempered distributions.

II. SINGULARITY OF THE TWO-POINT WIGHTMAN FUNCTION OF THE FREE NEUTRAL SCALAR FIELD IN FOUR-DIMENSIONAL SPACE-TIME

It is well known that the two-point Wightman function of the free neutral scalar field in four-dimensional space-time,

$$D_m^{(-)}(x) = (2\pi)^{-3}\int e^{-ik\cdot x}\delta(k^2-m^2)\theta(k_0)dk$$
$$= (2\pi)^{-3}\int [2\omega(\mathbf{k})]^{-1}e^{-i\omega(\mathbf{k})x_0}e^{i\mathbf{kx}}\,d\mathbf{k}$$
$$[k\cdot x = k_0x_0-\mathbf{kx},\quad k^2=(k_0)^2-\mathbf{k}^2,\quad \omega(\mathbf{k}) = (\mathbf{k}^2+m^2)^{1/2}], \tag{2.1}$$

is a boundary value of the analytic continuation of the two-point Schwinger function of the free field,

$$S_m(y) = (2\pi)^{-4}\int [p^2+m^2]^{-1}e^{ip\cdot y}\,dp$$
$$= (2\pi)^{-3}\int [2\omega(\mathbf{p})]^{-1}e^{-\omega(\mathbf{p})|y_0|}e^{i\mathbf{py}}\,d\mathbf{p}$$
$$[p\cdot y = p_0y_0+\mathbf{py},\quad p^2=(p_0)^2+\mathbf{p}^2], \tag{2.2}$$

which is analytic in $y_0\neq 0$. That is, we have

$$\lim_{\epsilon\to+0} S_m(ix_0+\epsilon,\mathbf{x}) = \lim_{\epsilon\to+0} D_m^{(-)}(x_0-i\epsilon,\mathbf{x}) = D_m^{(-)}(x).$$

It is also well known that the bilinear form $C(f,g)$ on $\mathscr{S}(\mathbf{R}^4)\times\mathscr{S}(\mathbf{R}^4)$, defined by

$$C(f,g) = \int S_m(x-y)f(x)g(y)dx\,dy,$$

determines a unique Gaussian measure $d\Phi_C$ on $\mathscr{S}'(\mathbf{R}^4)$ satisfying

$$e^{-C(f,f)/2} = \int e^{i\Phi(f)}\,d\Phi_C,\quad f\in\mathscr{S}(\mathbf{R}^4). \tag{2.3}$$

If $m=0$, we can evaluate $D_0^{(-)}(x)$ explicitly:

 0022-2488/86/030832-08$02.50

$$D_0^{(-)}(x) = [2(2\pi)^3]^{-1}\int |\mathbf{k}|^{-1}e^{-i|\mathbf{k}|x_0}e^{i\mathbf{kx}}\,d\mathbf{k}$$

$$= \lim_{\epsilon\to+0}[2(2\pi)^3]^{-1}\int |\mathbf{k}|^{-1} \times\exp[-i|\mathbf{k}|(x_0-i\epsilon)]\exp(i|\mathbf{k}|\,|\mathbf{x}|\cos\theta) \times|\mathbf{k}|^2\sin\theta\, d|\mathbf{k}|\,d\theta\,d\phi$$

$$= \lim_{\epsilon\to+0}[2(2\pi)^2]^{-1}\int_0^\infty [i|\mathbf{k}|]^{-1} \times\exp[-i|\mathbf{k}|(x_0-i\epsilon)] \times[e^{i|\mathbf{k}|\,|\mathbf{x}|}-e^{-|\mathbf{k}|\,|\mathbf{x}|}]d|\mathbf{k}|$$

$$= \lim_{\epsilon\to+0}(2\pi)^{-2}[(x_0-i\epsilon)^2-\mathbf{x}^2]^{-1}.$$

Thus $D_0^{(-)}(x)$ has singularity on the light cone with order ϵ^{-2} at the origin of the space-time.

In the massive case $D_m^{(-)}(x)$ is shown to have a similar singularity as follows (see Glimm and Jaffe[8]). Let

$$g(\epsilon) = [2(2\pi)^3]^{-1}\int\omega(\mathbf{k})^{-1}e^{-\omega(\mathbf{k})\epsilon}\,d\mathbf{k} = (2\pi)^{-2}\int_0^\infty\omega(\mathbf{k})^{-1}e^{-\omega(\mathbf{k})\epsilon}|\mathbf{k}|^2\,d|\mathbf{k}|,$$

then we have

$$|D_m^{(-)}(x_0-i\epsilon,\mathbf{x})|\leqslant g(\epsilon) = D_m^{(-)}(-i\epsilon,0) \tag{2.4}$$

and

$$\epsilon^2 g(\epsilon) = (2\pi)^{-2}\int_0^\infty\exp\left[-s\left\{1+\left(\frac{\epsilon m}{s}\right)^2\right\}^{1/2}\right] \times s/\{1+(\epsilon m/s)^2\}^{1/2}\,ds,$$

where we have let $|\mathbf{k}| = s/\epsilon$. Thus we have

$$\epsilon^2 g(\epsilon)\to(2\pi)^{-2}\int_0^\infty e^{-s}s\,ds = (2\pi)^{-2},\quad \text{as } \epsilon\to+0,$$

and $D_m^{(-)}(x)$ has the singularity of order ϵ^{-2} at the origin.

The expression (2.1) of $D_m^{(-)}(x)$ shows that the support of its Fourier transform is contained in the forward light cone. Therefore $D_m^{(-)}(x)$ is the boundary value of the function $D_m^{(-)}(z)$, which is holomorphic in the backward light cone (see Theorem 3.3.1 of Kawai[9]).

Here we give an estimate for $S_m(y)$ of (2.2). Since $S_m(y)$ is rotationally invariant, we may assume $y = (y_0,0,0,0)$. Then

$$|S_m(y)| = S_m(y) = (2\pi)^{-3}\int[2\omega(\mathbf{p})]^{-1}e^{-\omega(\mathbf{p})|y_0|}\,d\mathbf{p} = g(|y_0|).$$

For sufficiently small $\epsilon>0$, we have

$$\omega(\mathbf{p})\geqslant\begin{cases} m+\epsilon|\mathbf{p}|^2, & \text{for } |\mathbf{p}|\leqslant 1,\\ m+\epsilon|\mathbf{p}|, & \text{for } |\mathbf{p}|\geqslant 1.\end{cases}$$

Therefore

$$g(t)\leqslant\text{const}\,e^{-mt}\left[\int_0^1 e^{-t\epsilon r^2}r^2\,dr+\int_1^\infty e^{-t\epsilon r}r^2\,dr\right] \leqslant\text{Const}\,e^{-mt}[t^{-3/2}+t^{-3}],$$

and we have for large y

$$|S_m(y)|\leqslant\text{const } e^{-m|y|}|y|^{-3/2}. \tag{2.5}$$

The n-fold product $[D_m^{(-)}(x)]^n$ of $D_m^{(-)}(x)$ is also a boundary value of the holomorphic function $[D_m^{(-)}(z)]^n$, and defines a hyperfunction; moreover it is a distribution since the order of growth is ϵ^{-2n} when it approaches the real axis. But

$$\sum_{n=0}^\infty b_n[D_m^{(-)}(x)]^n \tag{2.6}$$

is a distribution if and only if all except finite b_n's are vanishing (see Vogt[10]). On the other hand, if

$$\sum_{n=0}^\infty b_n z^n \tag{2.7}$$

is an entire function, then (2.6) is a hyperfunction. Moreover it is a Fourier hyperfunction because $D_m^{(-)}(x_0-i\epsilon,\mathbf{x})$, for ϵ fixed, is a bounded function by (2.4). The power series (2.7) determines an entire function if and only if its coefficients b_n satisfy the condition

$$\lim_{n\to\infty}|b_n|^{1/n} = 0. \tag{2.8}$$

Finally we look into the relation between the condition (2.8) and the infraexponential condition for the Fourier transformation of (2.6). We start with a series of propositions.

Proposition 2.1: The Fourier transformation of $[D_m^{(-)}(x)]^2$ is

$$[4(2\pi)^3]^{-1}\theta(p_0)\theta(p^2-(2m)^2)[p^2-(2m)^2]^{1/2}/(p^2)^{1/2}. \tag{2.9}$$

Proof: See (B.1) of Bogolubov, Logunov, and Todorov.[11] See also the Appendix.

Proposition 2.2: The Fourier transformation $F_n(p)$ of $[D_0^{(-)}(x)]^n$ for $n\geqslant 2$ is

$$F_n(p) = (2\pi)^{1-2n}4^{1-n}(n-1)^{-1}[(n-2)!]^{-2} \times(p^2)^{n-2}\theta(p_0)\theta(p^2). \tag{2.10}$$

From (2.9) and (2.10) we have

$$[D_0^{(-)}(x)]^n = (2\pi)^{4-2n}4^{2-n}(n-1)^{-1} \times[(n-2)!]^{-2}(-\Box)^{n-2}D_0^{(-)}(x)^2$$

for $n\geqslant 2$.

Proposition 2.3: The Fourier transformation of $[D_m^{(-)}(x)]^n$ is dominated by the function $F_n(p)$, (2.10).

The proof of the latter two propositions will be given in the Appendix.

It is well known that the entire function

$$h(p) = \sum_{n=2}^\infty c_n(p^2)^{n-2} \tag{2.11}$$

is a Fourier hyperfunction if and only if it is infraexponential; this is equivalent to saying that $\Sigma_{m=2}^\infty c_n(p^2)^n$ is infraexponential, i.e., $\Sigma_{n=2}^\infty c_n z^n$ is an entire function of order $\frac{1}{2}$ and type 0. More explicitly, (2.11) is a Fourier hyperfunction if and only if the coefficients satisfy the condition

$$\lim_{n\to\infty}[(2n)!|c_n|]^{1/2n} = 0 \tag{2.12}$$

(see Theorem 2.2.10 of Boas[12]).

The condition (2.12) coincides with the requirement that the operator

$$\sum_{n=2}^{\infty} c_n(-\Box)^{n-2} \tag{2.13}$$

be a local operator (see the Appendix). In order that the function

$$\sum_{n=2}^{\infty} b_n F_n(p)$$

be a Fourier hyperfunction it is necessary and sufficient that

$$c_n = b_n/[(2\pi)^{2n-1}4^{n-1}(n-1)\{(n-2)!\}^2]$$

satisfy (2.12), that is,

$$\lim_{n\to\infty} |b_n|^{1/2n} = 0,$$

and this is equivalent to the condition (2.8). Since $\exp z$ is an entire function, we have the following theorem.

Theorem 2.4: $\exp\{D_m^{(-)}(x)\}$ is a Fourier hyperfunction, but not a distribution.

III. WICK PRODUCTS AND ENTIRE FUNCTIONS OF THE FREE NEUTRAL SCALAR FIELD

For the free field operator $\phi(x)$, the Wick product is defined by

$$:\phi(x_1)\cdots\phi(x_l): = \sum_{r=0}^{[l/2]} (-1)^r \sum_{C_r} [x_{j_1}\cdots x_{j_{2r}}] \times\phi(x_{k_1})\cdots\phi(x_{k_{l-2r}}). \tag{3.1}$$

Here $[l/2]$ is the greatest integer less than or equal to $l/2$. The sum Σ_{C_r} is over all partitions of the integers $1,\ldots,l$ into two subsets $\{j_1,\ldots,j_{2r}\}$ and $\{k_1,\ldots,k_{l-2r}\}$ for which $j_1<j_2<\cdots<j_{2r}$, $k_1<k_2<\cdots<k_{l-2r}$. The hafnian $[x_{j_1},\ldots,x_{j_{2r}}]$ is defined by the vacuum expectation value

$$[x_{j_1},\ldots x_{j_{2r}}] = (\Omega,\phi(x_{j_1})\cdots\phi(x_{j_{2r}})\Omega) = \sum_{\text{pairing}} \prod_{s=1}^{r} D_m^{(-)}(x_{i_s}-x_{k_s}),$$

where the summation is over all pairings $(i_1,k_1),\ldots,(i_r,k_r)$ of $\{j_1,\ldots,j_{2r}\}$ such that $i_s<k_s$ for $s=1,\ldots,r$. Then $:\phi(x)^l:$ is defined (formally) as

$$:\phi(x)^l: = \lim_{x_1,\ldots,x_l\to x} :\phi(x_1)\cdots\phi(x_l):.$$

It is a well-defined field operator (see Wightman and Gårding[1]).

Let be defined

$$\rho^{(i)}(x) = \sum_{n=0}^{\infty} a_n^{(i)} \frac{:\phi(x)^n:}{n!}, \tag{3.2}$$

then Theorem A.1 of Jaffe[13] reads as follows.

Theorem 3.1: As a formal power series we have

$$(\Omega,\rho^{(1)}(x_1)\cdots\rho^{(n)}(x_n)\Omega) = \sum_{r_{ij}=0;1<i<j<n}^{\infty} \frac{A(R)T^R}{R!}, \tag{3.3}$$

where

$$r_{ij}=r_{ji},\quad r_{ii}=0,$$
$$R_i = \sum_{j=1}^{n} r_{ij},\quad t_{ij} = D_m^{(-)}(x_i-x_j),$$
$$R! = \prod_{1<i<j<n} (r_{ij})!,\quad T^R = \prod_{1<i<j<n} (t_{ij})^{r_{ij}},$$
$$A(R) = \prod_{j=1}^{n} a_{R_j}^{(j)}. \tag{3.4}$$

Corollary 3.2: In the case of

$$\rho^{(i)}(x) = :e^{g_i\phi(x)}: = \sum_{n=0}^{\infty} g_i^n \frac{:\phi(x)^n:}{n!},$$

(3.3) becomes

$$(\Omega,\rho^{(1)}(x_1)\cdots\rho^{(n)}(x_n)\Omega) = \exp\left\{\sum_{1<i<j<n} g_i g_j t_{ij}\right\}. \tag{3.5}$$

For the random field $\Phi(f)$ defined by (2.3), we also define the Wick product by

$$:\Phi(f_1)\cdots\Phi(f_l): = \sum_{r=0}^{[l/2]} (-1)^r \times\sum_{C_r}\int \Phi(f_{j_1})\cdots\Phi(f_{j_{2r}})d\Phi_C \times\Phi(f_{k_1})\cdots\Phi(f_{k_{l-2r}}), \tag{3.6}$$

where Σ_{C_r} is the same as (3.1). Then we have

$$:e^{\Phi(f)}: = \sum_{n=0}^{\infty} \frac{:\Phi(f)^n:}{n!} = e^{-C(f,f)/2}e^{\Phi(f)},$$

and

$$\int :e^{\Phi(f_1)}:\cdots:e^{\Phi(f_n)}: d\Phi_C = \exp\left[-\sum_{i=1}^{n}\frac{C(f_i,f_i)}{2}\right]\int \exp\left[\phi\left(\sum_{j=1}^{n} f_j\right)\right] d\Phi_C = \exp\left\{\sum_{1<i<j<n} C(f_i,f_j)\right\}. \tag{3.7}$$

If f_i converges to the point measure $g_i\delta_{x_i}$ in such a way that $x_i\neq x_j$ for $i\neq j$, the right-hand side of (3.7) converges to

$$\exp\left\{\sum_{1<i<j<n} g_i g_j S_m(x_i-x_j)\right\},$$

which is the Schwinger function corresponding to (3.5).

Theorem 3.3: If

$$\lim_{n\to\infty} [|a_n^{(i)}|^2/n!]^{1/n} = 0, \tag{3.8}$$

then the right-hand side of (3.3) is an entire function of t_{ij}.

Proof: Let

$$\|R\| = \sum_{1<i<j<n} r_{ij}, \tag{3.9}$$

then it follows from Lemaire's theorem that (3.3) is an entire function of t_{ij} if

$$\lim_{\|R\|\to\infty} [|A(R)|/R!]^{1/\|R\|} = 0. \tag{3.10}$$

Since the multinomial theorem implies

$$R_i! = \left(\sum_{j=1}^{n} r_{ij}\right)! < n^{R_i}\prod_{j=1}^{n}(r_{ij})!,$$

and it follows from (3.4) and (3.9) that

$$\sum_{i=1}^{n} R_i = 2\|R\|, \quad \prod_{i=1}^{n}\prod_{j=1}^{n}(r_{ij})! = (R!)^2,$$

we have

$$\left[\frac{|A(R)|}{R!}\right]^2 = \frac{\prod_{i=1}^{n}|a_{R_i}^{(i)}|^2}{(R!)^2} < \prod_{i=1}^{n}\left(|a_{R_i}^{(i)}|^2\frac{n^{R_i}}{R_i!}\right)$$

and

$$\left[\frac{|A(R)|}{R!}\right]^{1/\|R\|} < \prod_{i=1}^{n}\left[|a_{R_i}^{(i)}|^2\frac{n^{R_i}}{R_i!}\right]^{1/2\|R\|} = n\prod_{i=1}^{n}\left[\frac{|a_{R_i}^{(i)}|^2}{R_i!}\right]^{1/2\|R\|}.$$

Since $[|a_{R_i}^{(i)}|^2/R_i]^{1/\|R\|}$ $(i=1,...,n)$ remain bounded as $\|R\|\to\infty$, and at least one R_i satisfies $nR_i \geqslant 2\|R\|$, the condition (3.10) follows from (3.8).

Theorem 3.4: The condition (3.8) is necessary so that

$$(\Omega,\rho^{(i)}(x_1)\rho^{(i)}(x_2)\Omega) = \sum_{n=0}^{\infty} a_n^{(i)2}\frac{t_{12}^n}{n!}$$

is an entire function of t_{12}.

Proof: Obvious.

By virtue of Theorem 3.3, if the coefficients $a_n^{(i)}$ of $\rho^{(i)}(x)$ in (3.2) satisfy the condition (3.8), then the right-hand side of (3.3) defines a Fourier hyperfunction (of type I). Next we have to show that these Fourier hyperfunctions satisfy the (modified) Wightman axioms, formulated in Nagamachi and Mugibayashi[3]: (R0) Fourier hyperfunction property, (R1) relativistic covariance, (R2) positivity, (R3) local commutativity, (R4) spectral condition, and (R5) cluster property. Of these, (R0) and (R1) are obviously satisfied.

In order to show other properties, we define $\rho_N^{(i)}(x)$ as a truncation of $\rho^{(i)}(x)$, that is, as a Wick polynomial of the form

$$\rho_N^{(i)}(x) = \sum_{n=0}^{N} a_N^{(i)}\frac{:\phi(x)^n:}{n!},$$

so that the Wightman functions for $\rho_N^{(i)}(x)$ satisfy all the (unmodified) Wightman axioms (see Wightman and Gårding[1]). Since the right-hand side of (3.3) is an absolutely convergent series, $(\Omega,\rho_N^{(1)}(x_1)\cdots\rho_N^{(n)}(x_n)\Omega)$ converges to $(\Omega,\rho^{(1)}(x_1)\cdots\rho^{(n)}(x_n)\Omega)$ as $N\to\infty$ in the sense of Fourier hyperfunctions. Thus we easily see the positivity (R4). However, it is difficult to verify (R3), (R4), and (R5) in this manner, so we use the Euclidean theory.

The theory of Osterwalder and Schrader[14] states that the Schwinger functions for $\rho_N^{(i)}(x)$ satisfy the Osterwalder–Schrader (OS) axioms, namely, (E0) distribution property, (E1) Euclidean covariance, (E2) positivity, (E3) symmetry, and (E4) cluster property. The Schwinger function corresponding to the Wightman function $(\Omega,\rho^{(1)}(x_1)\cdots\rho^{(n)}(x_n)\Omega)$ is just the right-hand side of (3.3) with t_{ij} replaced by $S_m(y_i - y_j)$.

Now we show that they satisfy the modified OS axioms (E0′) – (E4′) of Nagamachi and Mugibayashi.[7] Axiom (E0′) follows from the estimate (2.5); (E1′) and (E3′) are obvious from (2.2) and (3.3); and (E2′) follows from the positivity of Schwinger functions of $\rho_N^{(i)}(x)$. Finally we show (E4′). Let $S_n(y_1,...,y_n)$ be the Schwinger function (3.3) with $t_{ij} = S_m(y_i - y_j)$. From (2.5), t_{ij} converges to zero as $y_i - y_j$ goes to infinity. If $y_i - y_j$, $1\leqslant i\leqslant k<j\leqslant n$, goes to infinity,

$$T^R = \prod_{1\leqslant i<j\leqslant n}(t_{ij})^{r_{ij}}$$

converges to zero unless all r_{ij} for $1\leqslant i\leqslant k<j\leqslant n$ are zero. Therefore we have

$$T^R = \left[\prod_{1\leqslant i<j\leqslant k}(t_{ij})^{r_{ij}}\right]\prod_{k<i<j\leqslant n}(t_{ij})^{r_{ij}}.$$

Thus, $S_n(y_1,...,y_k,y_{k+1}+\lambda a,...,y_n+\lambda a)$ for $a(\neq 0)\in\mathbf{R}^4$ converges to $S_k(y_1,...,y_k)S_{n-k}(y_{k+1},...,y_n)$ as λ goes to infinity. This proves (E4′).

The theory of Nagamachi and Mugibayashi[7] states that the Schwinger functions satisfying (E0′)–(E4′) uniquely define Wightman–Fourier hyperfunctions (of mixed type) satisfying (R0′)–(R5′) of Ref. 7. Since (3.3) is a Fourier hyperfunction of type I, the resulting Wightman–Fourier hyperfunctions satisfy all the (modified) Wightman axioms (R0)–(R5) of Ref. 3.

Let $\rho^{(i)}(x)$ of (3.2) be either $\phi(x)$ or $\rho_\pm(x) = :e^{\pm ig\phi(x)}:$, then by the reconstruction theorem (Theorem 6.1 of Ref. 3), the Wightman functions (3.3) define the system of fields $\phi(x)$ and $\rho_\pm(x)$, that is, there exists a system of Hilbert space H with a unique vacuum Ω, unitary representation $U(a,\Lambda)$ of the Poincaré group, and field operators $\phi(x)$ and $\rho_\pm(x)$ defined on a dense subset D of H satisfying the modified Wightman axioms (W0)–(W5) of Ref. 3.

IV. DERIVATIVE COUPLING MODEL

The Lagrangian density of the derivative coupling theory of neutral scalar meson in four-dimensional space-time is

$$L(x) = L_F(x) + L_I(x), \tag{4.1}$$

with

$$L_F(x) = -\bar\psi(x)(\gamma^\mu\partial_\mu + M)\psi(x) - \tfrac{1}{2}\{(\partial_\mu\phi(x))^2 + m^2\phi(x)^2\},$$
$$L_I(x) = ig(\bar\psi(x)\gamma^\mu\psi(x))\partial_\mu\phi(x).$$

Since the coupling constant g has the dimension of length in natural units, the power-counting argument suggests that this theory belongs to the class of nonrenormalizable quantum field theories.

We quantize it by path integral; more precisely, we calculate the Schwinger function in Euclidean lattice theory (of infinitesimal lattice spacing). The reason the Schwinger function can be calculated explicitly is that a set of transformations

$$\psi(x) = e^{ig\phi(x)}\psi'(x), \quad \bar\psi(x) = e^{-ig\phi(x)}\bar\psi'(x)$$

converts $L(x)$ of $(\psi(x), \bar\psi(x), \phi(x))$ into $L_F(x)$ of $(\psi'(x), \bar\psi'(x), \phi(x))$.

From now on we use the nonstandard analysis (see Davis[15]). Let N be an infinitely large hyperreal number and define $L = N^2$ and an infinitesimal $\Delta = \sqrt{\pi}/N$. Let $^*\mathbf{Z}$ be the set of all hyperintegers and $\Gamma = \Delta^*\mathbf{Z}/2L\Delta^*\mathbf{Z}$ be a lattice with an infinitesimal lattice spacing Δ and an infinitely large length $2\sqrt{\pi}N$. The number of lattice sites of Γ is equal to L. Let e_μ be the vector of length Δ parallel to the μth coordinate axis ($\mu = 0,...,3$). Let $^*\mathbf{R}$ be the set of all hyperreal numbers. We define a measure $G(\Phi)$ on $^{**}\mathbf{R}^L$ by

$$G(\Phi) = C\exp\left\{\frac{1}{2}\sum_{y\in\Gamma^4}\Phi(y) \times\left[\sum_{\mu=0}^{3}\frac{\Phi(y+e_n)+\Phi(y-e_\mu)-2\Phi(y)}{\Delta^2} - m^2\Phi(y)\right]\Delta^4\right\}\prod_{y\in\Gamma^4}d\Phi(y), \qquad (4.2)$$

where C is a normalization constant. We also define a measure $D(\Psi^1,\Psi^2)$ on the hyper-Grassmann algebra Λ generated by $\{\Psi^1_\alpha(y), \Psi^2_\alpha(y);\ \alpha = 1,...,4,\ y\in\Gamma^4\}$ by

$$D(\Psi^1,\Psi^2) = C'\exp\left\{\sum_{y\in\Gamma^4}\Psi^{2T}(y)\left[\sum_{\mu=0}^{3}\gamma^E_\mu\nabla_\mu + M\right]\Psi^1(y)\Delta^4\right\} \times\prod_{y\in\Gamma^4}\prod_{\alpha=1}^{4}d\Psi^1_\alpha(y)d\Psi^2_\alpha(y), \qquad (4.3)$$

where C' is another normalization constant,

$$\Psi^2(y) = (\Psi^2_1(y),...,\Psi^2_4(y))^T,$$

$$\Psi^1(y) = (\Psi^1_1(y),...,\Psi^1_4(y))^T,$$

$$\gamma^E_0 = \begin{pmatrix}\sigma_0 & 0\\ 0 & -\sigma_0\end{pmatrix}, \quad \gamma^E_j = \begin{pmatrix}0 & -i\sigma_j\\ i\sigma_j & 0\end{pmatrix},$$

$$j = 1,2,3,$$

$$\sigma_0 = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \quad \sigma_1 = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},$$

$$\sigma_2 = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}, \quad \sigma_3 = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},$$

and

$$\nabla_\mu\Psi_k(y) = \begin{cases}\nabla^+_\mu\Psi_k(y) = (\Psi_k(y+e_\mu) - \Psi_k(y))/\Delta, & \text{if } k = 1,2,\\ \nabla^-_\mu\Psi_k(y) = (\Psi_k(y) - \Psi_k(y-e_\mu))/\Delta, & \text{if } k = 3,4.\end{cases}$$

Namely $\nabla_\mu = \frac{1}{2}\{(\nabla^+_\mu + \nabla^-_\mu) + \gamma^E_0(\nabla^+_\mu - \nabla^-_\mu)\}$. The integration of the Grassmann algebra is given by the rule

$$\int d\Psi^\mu_\alpha(y) = 0, \quad \int\Psi^\mu_\alpha(y)d\Psi^\mu_\alpha(y) = 1$$

(see Berezin[16]).

Now, $G(\Phi)$ is a Gaussian measure whose covariance

$$\int\Phi(y_1)\Phi(y_2)G(\Phi)$$

has the standard part that coincides with the Schwinger function $S_m(y_1 - y_2)$ of free neutral scalar field of mass m.

The covariance

$$\int\Psi^1_\alpha(y_1)\Psi^2_\beta(y_2)D(\Psi^1,\Psi^2)$$

of the measure $D(\Psi^1,\Psi^2)$ has the standard part that coincides with the Schwinger function $R_{M;\alpha,\beta}$ of free Dirac field of mass M without suffering from the doubling problem of lattice fermions (see Nagamachi and Mugibayashi[17]).

Let us define the Euclidean lattice interaction Lagrangian density $L_I(y)$, which corresponds to $L_I(x)$ of (4.1), by

$$L_I(y) = \frac{1}{2}\Psi^{2T}(y)e^{ig\Phi(y)}\sum_{\mu=0}^{3}\gamma^E_\mu[P_+\Psi^1(y+e_\mu) \times(\exp[-ig\Phi(y+e_\mu)] - e^{-ig\Phi(y)})/\Delta + P_-\Psi^1(y-e_\mu)(e^{-ig\Phi(y)} - \exp[-ig\Phi(y-e_\mu)])/\Delta], \qquad (4.4)$$

where $P_\pm$ is the projection operator defined by

$$P_\pm = (1\pm\gamma^E_0)/2.$$

If we replace infinitesimal differences by derivatives, $L_I(y)$ becomes (the Euclidean version of) $L_I(x)$.

Now we calculate Schwinger functions. The two-point Schwinger function is calculated as

$$\int\Psi^1_\alpha(y_1)\Psi^2_\beta(y_2)\exp\left(\sum_{y\in\Gamma}4L_I(y)\Delta^4\right)D(\Psi^1,\Psi^2)G(\Phi) \times\left\{\int\exp\left(\sum_{y\in\Gamma}4L_I(y)\Delta^4\right)D(\Psi^1,\Psi^2)G(\Phi)\right\}^{-1}. \qquad (4.5)$$

If we change the variables of integration by

$$\Psi^1(y) = e^{ig\Phi(y)}\Psi'^1(y), \quad \Psi^2(y) = e^{-ig\Phi(y)}\Psi'^2(y),$$

then (4.5) becomes

$$\int e^{ig\Phi(y_1)}\Psi'^1_\alpha(y_1)e^{-ig\Phi(y_2)}\Psi'^2_\beta(y_2)D(\Psi'^1,\Psi'^2)G(\Phi) = \int\Psi'^1_\alpha(y_1)\Psi'^2_\beta(y_2)D(\Psi'^1,\Psi'^2) \times\int e^{ig\Phi(y_1)}e^{-ig\Phi(y_2)}G(\Phi).$$

The standard part of $\int\Psi'^1_\alpha(y_1)\Psi'^2_\beta(y_2)D(\Psi'^1,\Psi'^2)$ is the two-point Schwinger function

$$R_{M;\alpha,\beta}(y) = \left\{\sum_{\mu=0}^{3}\gamma^E_\mu\left(\frac{\partial}{\partial y_\mu}\right) + M\right\}_{\alpha,\beta}S_m(y) \qquad (4.6)$$

of the free Dirac field, where $y = y_1 - y_2$ (see Nagamachi and Mugibayashi[17]). But $\int e^{ig\Phi(y_1)}e^{-ig\Phi(y_2)}G(\Phi)$ is infinitely large and has no standard part. This infinity, however, can be removed by Wick ordering with respect to the Gaussian measure $G(\Phi)$, that is, $\int :e^{ig\Phi(y_1)}: :e^{-ig\Phi(y_2)}: G(\Phi)$ has the standard part $\exp\{g^2S_m(y)\}$, where $y = y_1 - y_2$ with $y_1 \neq y_2$. Thus, after the infinity is removed, the two-point Schwinger function of the derivative coupling theory defined by (4.1) turns out

$$R_{M;\alpha,\beta}(y)e^{g^2S_m(y)} \qquad (4.7)$$

and the corresponding Wightman function is

$$G_{M;\alpha,\beta}(x)e^{g^2D^{(-)}_m(x)} = W^{1,2}_{\alpha,\beta}(x).$$

Here $x = x_1 - x_2$, and

$$G_{M;\alpha,\beta}(x) = (i\gamma^\mu\partial_\mu + M)_{\alpha\beta} D_m^{(-)}(x) = W_{0,\alpha,\beta}^{1,2}(x)$$

is the two-point Wightman function of the free Dirac field

$$W_{0,\alpha,\beta}^{1,2}(x) = \mathscr{W}_{0,\alpha,\beta}^{1,2}(x_1,x_2) = (\Omega,\psi_\alpha^1(x_1)\psi_\beta^2(x_2)\Omega),$$

where $\psi^1(x_1) = \psi(x_1)$ and $\psi^2(x_2) = \bar{\psi}(x_2)$.

Let

$$Z^{-1/2} = \exp\left\{g^2\int\Phi(0)^2 G(\Phi)\right\} \tag{4.8}$$

be an infinitely large number corresponding to $\exp\{g^2 S_m(0)\}$, then

$$:e^{\pm ig\Phi(y)}: = Z^{-1/2} e^{\pm ig\Phi(y)}.$$

If we define renormalized fields $\Psi_R^i(y)$, $i = 1,2$, by

$$\Psi_R^i(y) = Z^{-1/2}\Psi^i(y) \tag{4.9}$$

and the renormalized Lagrangian density $L_{R,I}(y)$ and a measure $D_R(\Psi_R^1,\Psi_R^2)$ by

$$L_{R,I}(\Psi_R^1(y),\Psi_R^2(y),\Phi(y)) = L_I(\Psi^1(y),\Psi^2(y),\Phi(y)),$$
$$D_R(\Psi_R^1,\Psi_R^2) = D(\Psi^1,\Psi^2),$$

then the quantity

$$\int \Psi_{R,\alpha}^1(y_1)\Psi_{R,\beta}^2(y_2)\exp\left(\sum_{y\in\Gamma^4} L_{R,I}(y)\Delta^4\right)$$
$$\times D_R(\Psi_R^1,\Psi_R^2)G(\Phi)$$
$$\times\left\{\int\exp\left(\sum_{y\in\Gamma^4} L_{R,I}(y)\Delta^4\right)D_R(\Psi_R^1,\Psi_R^2)G(\Phi)\right\}^{-1}$$

has the standard part (4.7), since this is equal to the expression (4.5) multiplied by Z^{-1}.

Let

$$\mathscr{W}_{0,\alpha}^r(x_1,\ldots,x_n) = (\Omega,\psi_{\alpha_1}^{r_1}(x_1)\cdots\psi_{\alpha_n}^{r_n}(x_n)\Omega)$$

be the n-point Wightman distribution of free Dirac field, where $r = (r_1,\ldots,r_n)$, $\alpha = (\alpha_1,\ldots,\alpha_n)$, $r_j = 1,2$, $\alpha_j = 1,\ldots,4$. Then the corresponding Wightman function of ψ, $\bar{\psi}$ in (4.1) is

$$\mathscr{W}_\alpha^r(x_1,\ldots,x_n)$$
$$= \exp\left\{\sum_{1<i<j<n}(-1)^{r_i+r_j-1}D_m^{(-)}(x_i - x_j)\right\}$$
$$\times\mathscr{W}_{0,\alpha}^r(x_1,\ldots,x_n). \tag{4.10}$$

Thus the theory defined by (4.1) is renormalized by the field strength renormalization (4.9), but its Wightman functions (4.10) are not tempered distributions.

Let $\{H, \Omega; \phi(x), \rho_\pm(x); U(a,\Lambda)\}$ be the system defined at the end of Sec. III, and let $\{K, \Phi; \psi(x), \bar{\psi}(x); V(a,\Lambda)\}$ be the system of the free Dirac field, then each of these systems separately satisfies the (modified) Wightman axioms. Therefore the system

$$\{H\otimes K,\Omega\otimes\Phi; \Phi(x)\otimes I_K,\rho_\pm(x)\otimes I_K,$$
$$I_H\otimes\psi(x),I_H\otimes\bar{\psi}(x); U(a,\Lambda)\otimes V(a,\Lambda)\}$$

also satisfies the axioms. Let $\rho^{(s)}(x)$ stand for either $\phi(x)$ or $\rho_\pm(x)$. If the product $\rho^{(s)}(x)\otimes\psi_\alpha^r(x)[=(\rho^{(s)}(x)\otimes I_K)\cdot(I_H\otimes\psi_\alpha^r(x))]$ at the same point x is defined, we can say that the original model (4.1) satisfies the modified Wightman axioms.

The vector-valued Fourier hyperfunctions $\Pi_{i=1}^n\rho^{(s_i)}(x_i)\Omega$ [resp. $\Pi_{i=1}^n\psi_{\alpha_i}^{r_i}(x_i)\Phi$] are the boundary values of vector-valued holomorphic functions $P^{(s)}(\zeta_0,\zeta_1,\ldots,\zeta_{n-1})$ [resp. $\Psi_\alpha^r(\zeta_0,\zeta_1,\ldots,\zeta_{n-1})$] defined in $\mathscr{T}_+^n = \mathbf{R}^{4n}\times iV_+^n$ where V_+ is the forward light cone and $\zeta_0 = z_1$, $\zeta_j = z_{j+1} - z_j$, $z_j = x_j + iy_j$ (see Theorem 4.3 of Ref. 3). The function $P^{(s)}(\zeta_0,\ldots,\zeta_{n-1})\otimes\Psi_\alpha^r(\zeta_0,\ldots,\zeta_{n-1})$ is holomorphic in $\mathscr{T}_+^n$ and defines the Fourier hyperfunction

$$\Pi_{i=1}^n(\rho^{(s_i)}(x_i)\otimes\psi_{\alpha_i}^{r_i}(x_i))(\Omega\otimes\Phi),$$

which takes on its values in $H\otimes K$. Thus the product $\rho^{(s)}(x)\otimes\psi_\alpha^r(x)$ at the same point is defined.

APPENDIX: PROOFS OF PROPOSITOINS. A CONDITION FOR THE LOCAL OPERATOR

Proof of Proposition 2.1: By (2.1), the calculation of the Fourier transform of $[D_m^{(-)}(x)]^2$ reduces to the convolution of $\delta(k^2 - m^2)\theta(k_0)$;

$$(2\pi)^{-4}\int\delta(k^2 - m^2)\delta((p-k)^2 - m^2)$$
$$\times\theta(k_0)\theta(p_0 - k_0)dk. \tag{A1}$$

Since the distribution (A1) is Lorentz invariant, if $m\neq 0$ we may evaluate it in the rest frame in which $p = (E,0,0,0)$.

Introducing independent variables k_0, k^2, θ, and ϕ by

$$k_1 = r\sin\theta\cos\phi,\quad k_2 = r\sin\theta\sin\phi,$$
$$k_3 = r\cos\theta,\quad r = [(k_0)^2 - k^2]^{1/2},$$
$$0<\theta<\pi,\quad 0<\phi<2\pi,$$

(A1) becomes

$$(2\pi)^{-4}\int\delta(k^2 - m^2)\delta(E^2 - 2Ek_0 + k^2 - m^2)\theta(k_0)$$
$$\times\theta(E - k_0)(r/2)\sin\theta\, dk^2\, d\theta\, d\phi\, dk_0$$
$$= (2\pi)^{-4}\int\delta(E_2 - 2Ek_0)\theta(k_0 - m)\theta(E - k_0)$$
$$\times[(k_0)^2 - m^2]^{1/2}(1/2)\sin\theta\, d\theta\, d\phi\, dk_0$$
$$= (2\pi)^{-3}(2E)^{-1}[(E/2)^2 - m^2]^{1/2}\theta(E - 2m)$$
$$= (4(2\pi)^3)^{-1}[E^2 - (2m)^2]^{1/2}\theta(E - 2m)/E. \tag{A2}$$

Since this function is Lorentz invariant, we may write this result in the following invariant form:

$$(4(2\pi)^3)^{-1}\theta(p_0)\theta(p^2 - (2m)^2)$$
$$\times[p^2 - (2m)^2]^{1/2}/(p^2)^{1/2}. \tag{A3}$$

Since

$$\mathscr{S}'\times\mathscr{S}'\ni(f,g)\rightarrow f\otimes g\in\mathscr{S}'\otimes\mathscr{S}'$$

is continuous (see Treves[18]), and

$$(2\pi)^{-4}\langle\theta(k_0)\delta(k^2 - m^2)$$
$$\otimes\theta(q_0)\delta(q^2 - m^2),\ \phi(k+q)\rangle$$
$$= 4^{-1}(2\pi)^{-3}\langle\theta(p_0)\theta(p^2 - (2m)^2)$$
$$\times[p^2 - (2m)^2]^{1/2}/(p^2)^{1/2},\ \phi(p)\rangle,$$

the Fourier transformation of $[D_0^{(-)}(x)]^2$ turns out to be

$$4^{-1}(2\pi)^{-3}\theta(p_0)\theta(p^2). \tag{A4}$$

Proof of Proposition 2.2: For $n=2$, (2.10) coincides with (A4). For general n we can prove it by induction as follows:

$$\lim_{m\to 0}(2\pi)^{-1}\int (k^2)^{n-2}\theta(k^2)\theta(k_0)$$
$$\times\delta((p-k)^2-m^2)\theta(p_0-k_0)dk$$
$$=\lim_{m\to 0}(2\pi)^{-1}\int (k^2)^{n-2}\theta(k^2)$$
$$\times\delta(E^2-2Ek_0+k^2-m^2)\theta(k_0)\theta(E-k_0)$$
$$\times(r/2)\sin\theta\, dk^2\, d\theta\, d\phi\, dk_0$$
$$=\lim_{m\to 0}\int (2Ek_0-E^2+m^2)^{n-2}\theta(2Ek_0-E^2+m^2)$$
$$\times\theta(k_0)\theta(E-k_0)[(E-k_0)^2-m^2]^{1/2}dk_0$$
$$=\int_{E/2}^{E}E^{n-1}(2k_0-E)^{n-2}(E-k_0)\theta(E)dk_0$$
$$=4^{-1}[n(n-1)]^{-1}E^{2(n-1)}\theta(E)$$
$$=[4n(n-1)]^{-1}(p^2)^{n-1}\theta(p_0)\theta(p^2)\,.$$

The last equality follows from the same reasoning as that leading from (A2) to (A3).

Proof of Proposition 2.3: The following inequality holds.

$$(2\pi)^{-1}\int (k^2)^{n-2}\theta(k^2-(nm)^2)\theta(k_0)$$
$$\times\delta((p-k)^2-m^2)\theta(p_0-k_0)dk$$
$$=(2\pi)^{-1}\int (k^2)^{n-2}\theta(k^2-(nm)^2)\theta(k_0)$$
$$\times\delta(E^2-2Ek_0+k^2-m^2)$$
$$\times\theta(E-k_0)(r/2)\sin\theta\, dk^2\, d\theta\, d\phi\, dk_0$$
$$=\int(-E^2+2Ek_0+m^2)^{n-2}$$
$$\times\theta(-E^2+2Ek_0+m^2-(nm)^2)$$
$$\times\theta(k_0)\theta(E-k_0)[(k_0)^2+E^2-2Ek_0-m^2]^{1/2}$$
$$\times\theta((k_0)^2+E^2-2Ek_0-m^2)dk_0$$
$$=\int_{E/2+(n^2-1)m^2/2E}^{E-m}(-E^2+2Ek_0+m^2)^{n-2}$$
$$\times[(E-k_0)^2-m^2]^{1/2}\theta(E-(n+1)m)dk_0$$
$$<\int_{E/2+n^2m^2/2E}^{E-m+m^2/2E}(-E^2+2Ek_0)^{n-2}$$
$$\times[(E-k_0+m^2/2E)^2-m^2]^{1/2}$$
$$\times\theta(E-(n+1)m)dk_0$$
$$=[4n(n-1)]^{-1}E^{2(n-1)}\theta(E-(n+1)m)$$
$$=[4n(n-1)]^{-1}(p^2)^{n-1}$$
$$\times\theta(p_0)\theta(p^2-\{(n+1)m\}^2)\,. \qquad (A5)$$

The last equality follows in the same way as (A3) coming from (A2). Using (A3) and (A5) we can prove by induction that the Fourier transform of $[D_m^{(-)}(x)]^n$ is dominated by

$$[(2\pi)^{2n-1}4^{n-1}(n-1)\{(n-2)!\}^2]^{-1}$$
$$\times(p^2)^{n-2}\theta(p_0)\theta(p^2-(nm)^2)$$

for $n>2$, which is turn is dominated by (2.10).

The condition that (2.13) is a local operator: Equation (2.13) is a local operator if and only if the support of

$$\sum_{n=0}^{\infty}c_n(-\Box)^n\delta(x) \qquad (A6)$$

is concentrated at the origin. Since $\delta(x)$ is represented as a hyperfunction by

$$\delta(x)=\prod_{j=0}^{3}\frac{i}{(2\pi z_j)}$$

and

$$(-\Box)^n=\sum_{n_0+\cdots+n_3=n}(-\partial_0^2)^{n_0}\prod_{i=1}^{3}\partial_i^{2n_i}\frac{n!}{n_0!\cdots n_3!},$$

$(-\Box)^n\delta(x)$ is represented by

$$(-\Box)^n\delta(x)=\sum_{n_0+\cdots+n_3=n}\frac{d_{n_0,\ldots,n_3}}{z_0^{2n_0+1}\cdots z_3^{2n_3+1}},$$

where

$$d_{n_0,\ldots,n_3}=(n_0+\cdots+n_3)!(-1)^{n_0}\prod_{j=0}^{3}\frac{(2n_j)!}{2\pi(n_j)!}.$$

The support of (A6) is concentrated at the origin if and only if

$$\sum_{n=0}^{\infty}c_n\sum_{n_0+\cdots+n_3=n}\frac{d_{n_0,\ldots,n_3}}{z_0^{2n_0+1}\cdots z_3^{2n_3+1}}$$

is analytic for $z_j\neq 0, j=0,\ldots,3$, in other words,

$$\sum_{n=0}^{\infty}c_n\sum_{n_0+\cdots+n_3=n}d_{n_0,\ldots,n_3}w_0^{2n_0+1}\cdots w_3^{2n_3+1}$$

is entire, that is

$$\lim_{n_0+\cdots+n_3=n\to\infty}[|c_n|\,|d_{n_0,\ldots,n_3}|]^{1/(2n+4)}=0\,. \qquad (A7)$$

Since the inequality

$$|d_{n_0,\ldots,n_3}|<|d_{n,0,0,0}|=(2n)!/(2\pi)^4$$

holds for $n=n_0+\cdots+n_3$, (A7) is equivalent to

$$\lim_{n\to\infty}[(2n)!|c_n|]^{1/2n}=0\,,$$

which is (2.12).

[1]A. S. Wightman and L. Gårding, Ark. Fys. 28, 129 (1964).
[2]A. M. Jaffe, Phys. Rev. 158, 1454 (1965).
[3]S. Nagamachi and N. Mugibayashi, Commun. Math. Phys. 46, 119 (1976).
[4]A. S. Wightman, Adv. Math. Suppl. Stud. B 7, 769 (1981).
[5]N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Wiley, New York, 1959).
[6]S. Okubo, Nuovo Cimento 19, 574 (1961).
[7]S. Nagamachi and N. Mugibayashi, Commun. Math. Phys. 49, 257 (1976).
[8]J. Glimm and A. Jaffe, *Quantum Physics, A Functional Integral Point of View* (Springer, New York, 1981).
[9]T. Kawai, J. Fac. Sci. Univ. Tokyo Sect. 1 17, 467 (1970).
[10]D. Vogt, J. Reine Angew, Math. 261, 134 (1973).
[11]N. N. Bogolubov, A. A. Logunov, and I. T. Todorov, *Introduction to Axi-*

omatic *Quantum Field Theory* (Benjamin, London, 1975).
[12]R. P. Boas, Jr., *Entire Functions* (Academic, New York, 1959).
[13]A. M. Jaffe, Ann. Phys. N.Y. **32**, 127 (1965).
[14]K. Osterwalder and R. Schrader, Commun. Math. Phys. **31**, 83 (1973).
[15]M. Davis, *Applied Nonstandard Analysis* (Wiley, New York, 1977).
[16]F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[17]S. Nagamachi and N. Mugibayashi, "Nonstandard analysis of Euclidean Fermi fields," preprint.
[18]F. Treves, *Topological Vector Spaces, Distributions and Kernels* (Academic, New York, 1967).

Volterra algebra and the Bethe–Salpeter equation

J. Faraut
Départment de Mathématiques, Université Louis Pasteur, Strasbourg, France
G. A. Viano
Dipartimento di Fisica dell'Università, Genova, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Genova, Genova, Italy

(Received 21 January 1985; accepted for publication 22 October 1985)

The diagonalization of the Bethe–Salpeter equation for the absorptive part of the amplitude is reconsidered. In particular the mathematical tools required by the diagonalization, like the Volterra algebra and the related spherical Laplace transform, are investigated in detail.

I. INTRODUCTION

In the past some attention has been paid to the problem of a partial diagonalization of the Bethe–Salpeter equation at fixed momentum transfer. The interested reader is referred to Refs. 1–5. In this paper we return to these problems with particular attention to the mathematical aspect of the questions.

The main points are the following: (a) to utilize to its full extent the symmetry of the problem; and (b) to diagonalize the equation with a mathematical tool that allows dealing with a class of amplitudes sufficiently large that their growth properties have physical interest.

Concerning the point (a), it has been recognized long ago[1,2] that, taking the momentum transfer Q to be fixed, the equation is invariant under the action of the $SO_0(1,2)$ group in the case of nonforward scattering ($Q^2<0$), and it is invariant under the action of the $SO_0(1,3)$ group in the case of forward scattering ($Q=0$). Next it was observed that the diagonalization of the Bethe–Salpeter equation for the absorptive part of the amplitude is considerably simpler than that involving the whole amplitude.[3,4] Indeed the support properties of the absorptive part of the amplitude are such that the integral in the Bethe–Salpeter equation is carried by a bounded region. Furthermore if we suppose, following Refs. 2–4, that the external momenta are spacelike (as we shall indeed assume hereafter), then it follows that the internal momenta are spacelike, too.

Concerning the point (b), the diagonalization was tried by means of the Laplace transform in order to deal with amplitudes, which show a power growth, in agreement with theoretical asymptotic bounds and with experiments.[3–5] In spite of these efforts, a safe mathematical formulation of the spherical Laplace transform, as well as of the Volterra algebra, has not been given up to now, as it has been explicitly remarked in Ref. 3. The purpose of this paper is precisely that of covering this gap.

The paper is organized as follows: in Sec. II we consider the generalized Volterra equations in $\mathbf{R}^n$; in Sec. III we introduce the so-called Volterra algebra on the hyperboloid with one sheet, and the associated spherical Laplace transform. This algebra may be regarded as the analog of the algebra of zonal functions on the upper sheet of the hyperboloid with two sheets, associated with the spherical Fourier transform. In Sec. IV we show that the analytical tools developed in the previous sections can be applied to the Bethe–Salpeter equation for the absorptive part of the amplitude in the case of spinless particles.

II. VOLTERRA KERNEL AND VOLTERRA INTEGRAL EQUATION

Let $X=\mathbf{R}^n$ and Ω be the forward light cone in $\mathbf{R}^n$ defined by

$$x_0^2 - x_1^2 - \cdots - x_{n-1}^2 > 0, \quad x_0 > 0. \tag{2.1}$$

We consider on X the ordering associated with Ω: for x and y in X we note $x>y$ if $x-y$ belongs to Ω, and $x \geqslant y$ if $x-y$ belongs to the closure $\overline{\Omega}$ of Ω. For this ordering, the set

$$D(y,x) = \{z \in X \mid y \leqslant z \leqslant x\} \tag{2.2}$$

is bounded. It is empty if $(x-y)$ does not belong to $\overline{\Omega}$.

A kernel $K(x,y)$ is said to be a Volterra kernel if $K(x,y)$ is continuous on $\Gamma = \{(x,y) \mid (x-y) \in \overline{\Omega}\}$ and vanishes out of Γ. The product of two Volterra kernels K_1 and K_2 is given by

$$K_1 \# K_2(x,y) = \int_{D(y,x)} K_1(x,z) K_2(z,y)\, dz, \tag{2.3}$$

where $K_1 \# K_2$ is again a Volterra kernel. Hence, the space $V(X)$ of Volterra kernels is an algebra: the Volterra algebra of the ordered space X.

If K is a Volterra kernel we define $K^{\#k}$ by

$$K^{\#1} = K, \tag{2.4a}$$

$$K^{\#k} = K^{\#(k-1)} \# K. \tag{2.4b}$$

Problem: For K and B given in $V(X)$ find a kernel A in $V(X)$ such that

$$A(x,y) - \int_{D(y,x)} K(x,z)\, A(z,y)\, dz = B(x,y). \tag{2.5}$$

This is a Volterra integral equation of the second kind. It can be written as follows:

$$A - K \# A = B. \tag{2.6}$$

Theorem 2.1: Equation (2.5) has a solution that is unique. It is given by

$$A(x,y) = B(x,y) + \int_{D(y,x)} R(x,z) B(z,y)\, dz, \tag{2.7}$$

where

 0022-2488/86/030840-09$02.50

$$R(x,y) = \sum_{k=1}^{\infty} K^{\#k}(x,y) \,. \tag{2.8}$$

The series (2.8) converges uniformly on bounded sets and R is a Volterra kernel.

Proof: This result was proved by Riesz[6] in a more general setting; the following proof is essentially due to him.

For $(x,y)\in\overline{\Omega}$ we let

$$r(x,y) = [(x_0-y_0)^2 - (x_1-y_1)^2 - \cdots - (x_{n-1}-y_{n-1})^2]^{1/2} \,, \tag{2.9}$$

and define

$$H_n(\alpha) = \pi^{(n-2)/2} 2^{(\alpha-1)} \Gamma(\alpha/2)\Gamma([\alpha+2-n]/2) \,, \tag{2.10}$$

where Γ denotes the Euler gamma function. Riesz defines the kernels I_α (generalized Riemann–Liouville kernels):

$$I_\alpha(x,y) = \begin{cases} 1/H_n(\alpha)\, r(x,y)^{(\alpha-n)} \,, \\ \quad \text{if } (x-y)\in\overline{\Omega} \,; & (2.11a) \\ 0 \,, \quad \text{if not} \,. & (2.11b) \end{cases}$$

The kernel I_α is locally integrable if $\alpha > n-2$, and satisfies the following composition relationship:

$$I_\alpha \# I_\beta = I_{\alpha+\beta} \,. \tag{2.12}$$

(a) Let K be a Volterra kernel. For x_0, y_0 fixed, there exists a constant $M>0$, such that for x and y in $D(y_0,x_0)$, $|K(x,y)|<M$. Therefore

$$|K(x,y)| < M H_n(n) I_n(x,y) \,, \tag{2.13}$$

and

$$|K^{\#k}(x,y)| \leqslant [M H_n(n)]^k I_n^{\#k}(x,y) \,. \tag{2.14}$$

The composition relationship (2.12) gives

$$I_n^{\#k} = I_{nk} \,. \tag{2.15}$$

Finally we obtain, for x and y in $D(y_0,x_0)$

$$|K^{\#k}(x,y)| \leqslant \{[M H_n(n)]^k / H_n(kn)\} \times [r(x,y)]^{n(k-1)} \,. \tag{2.16}$$

(b) We define

$$R(x,y) = \sum_{k=1}^{\infty} K^{\#k}(x,y) \,. \tag{2.17}$$

By the inequality (2.16) it follows that the series (2.17) converges uniformly on bounded sets and its sum is a Volterra kernel.

We verify easily that

$$A = B + R\#B \tag{2.18}$$

is a solution of Eq. (2.5). For proving uniqueness, let us assume that

$$K\#A = A \,, \tag{2.19}$$

then

$$K^{\#k}\#A = A \,,$$

and as k goes to infinity, we obtain $A=0$. ■

III. VOLTERRA ALGEBRA ON THE HYPERBOLOID WITH ONE SHEET

A. The hyperboloid with one sheet

Let X be the hyperboloid with one sheet in $\mathbb{R}^3$, defined by

$$-x_0^2 + x_1^2 + x_2^2 = 1 \,. \tag{3.1}$$

The Lorentz group $G = SO_0(1,2)$ acts transitively on X. The pseudo-Riemannian metric induced on X by the Minkowski metric

$$ds^2 = dx_0^2 - dx_1^2 - dx_2^2 \tag{3.2}$$

is invariant under G. We shall denote by $d\sigma$ the corresponding surface element. The isotropy subgroup of the point $e_2 = (0,0,1)$ is $H = SO_0(1,1)$; i.e., the one parameter subgroup of the following matrices:

$$h_\vartheta = \begin{pmatrix} \cosh\vartheta & \sinh\vartheta & 0 \\ \sinh\vartheta & \cosh\vartheta & 0 \\ 0 & 0 & 1 \end{pmatrix} . \tag{3.3}$$

We introduce also the one-parameter subgroup A of the following matrices:

$$a_\xi = \begin{pmatrix} \cosh\xi & 0 & \sinh\xi \\ 0 & 1 & 0 \\ \sinh\xi & 0 & \cosh\xi \end{pmatrix} , \tag{3.4}$$

and we define

$$A_+ = \{a_\xi \in A \,|\, \xi > 0\} \,. \tag{3.5}$$

Let Ω be the forward light cone in $\mathbb{R}^3$, defined by $x_0^2 - x_1^2 - x_2^2 > 0$, $x_0 > 0$. As illustrated in Sec. II, we can associate with Ω an ordering in $\mathbb{R}^3$: for $x = (x_0,x_1,x_2)$, $y = (y_0,y_1,y_2)$, we note $x>y$ if $x-y$ belongs to Ω, and $x \geqslant y$ if $x-y$ belongs to the closure $\overline{\Omega}$ of Ω. For this ordering, the set

$$D(y,x) = \{z\in X \,|\, y \leqslant z \leqslant x\} \tag{3.6}$$

is bounded. This ordering is invariant under G. Let G_+ be the set of g in G such that $g\,e_2 > e_2$, $e_2 = (0,0,1)$. The set G_+ is a semigroup and $A_+ = A \cap G_+$.

Proposition 3.1: The semigroup G_+ has the following decomposition:

$$G_+ = H A_+ H \,. \tag{3.7}$$

Proof: It is enough to prove that, if x belongs to X and satisfies $x > e_2$, then there exist $\xi > 0$, and ϑ real such that $x = h_\vartheta a_\xi e_2$, i.e.,

$$x_0 = \sinh\xi \cosh\vartheta \,,$$
$$x_1 = \sinh\xi \sinh\vartheta \,,$$
$$x_2 = \cosh\xi \,.$$

The point x belongs to X,

$$-x_0^2 + x_1^2 + x_2^2 = 1 \,,$$

and $x - e_2$ belongs to Ω,

$$x_0^2 - x_1^2 - (x_2-1)^2 > 0 \,, \quad x_0 > 0 \,.$$

Therefore $x_2 > 1$, and there exists $\xi > 0$ such that $x_2 = \cosh\xi$. We have

$$x_0^2 - x_1^2 = (\sinh\xi)^2 \,, \quad x_0 > 0 \,,$$

therefore there exists ϑ real such that

$$x_0 = \sinh \xi \cosh \vartheta ; \quad x_1 = \sinh \xi \sinh \vartheta . \qquad \blacksquare$$

The numbers (ξ,ϑ) will be called the polar coordinates of X. In terms of them, the surface element is given by $d\sigma = \sinh \xi \, d\xi \, d\vartheta$.

B. The Volterra algebra

Let Γ be the graph of the ordering, defined in Sec. III A, restricted to X: $\Gamma = \{(x,y) \in X \times X \mid y < x\}$. A function $K(x,y)$ on $X \times X$ is called a Volterra kernel if K is continuous on Γ and vanishes out of Γ. In order to make clear our exposition, let us restate the composition product of two Volterra kernels K_1 and K_2 in the case of the hyperboloid with one sheet; this product is defined by

$$K_1 \# K_2(x,y) = \int_{D(y,x)} K_1(x,z) K_2(z,y) \, d\sigma(z) , \tag{3.8}$$

where the set $D(y,x)$ is defined by Eq. (3.6). The integral makes sense without boundedness assumption on the kernels K_1 and K_2, since the set $D(y,x)$ is bounded. Furthermore the kernel $K_1 \# K_2$ is again a Volterra kernel, hence the set of Volterra kernels is an algebra which is called the Volterra algebra $V(X)$ of the ordered space X.

The kernel K is said to be invariant under G if, for any g in G

$$K(gx, gy) = K(x,y) . \tag{3.9}$$

The set $V(X)^\natural$ of the invariant Volterra kernels is a subalgebra of $V(X)$.

We can identify an invariant Volterra kernel K with a function f on G, which is continuous on the closure $\bar{G}_+$ of G_+, vanishes out of $\bar{G}_+$, and it is bi-invariant under H. The identification is given by

$$K(ge_2, e_2) = f(g) . \tag{3.10}$$

Hence we shall consider the elements of the algebra $V(X)^\natural$ as functions on G as well.

Theorem 3.1: The algebra $V(X)^\natural$ of invariant Volterra kernels is commutative.

Proof: We define

$$J = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} , \tag{3.11}$$

and we consider the automorphism σ of G defined by

$$\sigma(g) = JgJ . \tag{3.12}$$

The set of elements of G fixed under σ is the subgroup H. This automorphism is associated with the symmetry of X given by $x \to Jx$, in the following sense:

$$J(ge_2) = \sigma(g) e_2 . \tag{3.13}$$

The fixed points of this symmetry are e_2 and $-e_2$. With respect to the ordering associated with the forward light cone this symmetry is decreasing, i.e.,

$$y < x \Leftrightarrow Jx < Jy . \tag{3.14}$$

For a function f on G we let

$$f^\sigma(g) = f(\sigma(g)) , \tag{3.15}$$

$$f^\vee(g) = f(g^{-1}) . \tag{3.16}$$

Next we have the following lemma.

Lemma: For a function f in $V(X)^\natural$ we have

$$f^\sigma = f^\vee . \tag{3.17}$$

Proof: If f belongs to G_+, then by Proposition 3.1 we have

$$g = h_1 a_\xi h_2 ,$$

with h_1 and h_2 in H, $\xi > 0$; then

$$\sigma(g) = h_1 \sigma(a_\xi) h_2 = h_1 a_{-\xi} h_2 ,$$

$$g^{-1} = h_2^{-1} a_{-\xi} h_1^{-1} .$$

Therefore, if f is bi-invariant under H

$$f(\sigma(g)) = f(g^{-1}) ,$$

and the lemma is proved. Theorem 3.1 follows from the lemma because, if f_1 and f_2 belong to $V(X)^\natural$, we have

$$(f_1 \# f_2)^\sigma = f_1^\sigma \# f_2^\sigma , \tag{3.18}$$

$$(f_1 \# f_2)^\vee = f_2^\vee \# f_1^\vee . \tag{3.19}$$

Indeed for the corresponding Volterra kernels we have

$$K^\sigma(x,y) = K(\sigma(x), \sigma(y)) , \tag{3.20}$$

$$K^\vee(x,y) = K(y,x) , \tag{3.21}$$

and the proof follows easily. $\blacksquare$

Remark: For a function f belonging to $V(X)^\natural$ neither f^σ nor $f^\vee$ belongs to $V(X)$, but to the Volterra algebra related to the ordering associated with the backward light cone.

A function belonging to $V(X)^\natural$ depends only on one variable; for such a function f and for $\xi > 0$, we will use the following notation:

$$f(h_1 a_\xi h_2) = f[\cosh \xi] . \tag{3.22}$$

The following proposition provides an explicit formula for the composition product in $V(X)^\natural$.

Proposition 3.2: For two functions f_1 and f_2 in $V(X)^\natural$ we have

$$f_1 \# f_2 [\cosh \xi] = 2 \int_0^\xi \left\{ \int_0^{\alpha(\xi,\tau)} f_1 [\cosh \xi \cosh \tau - \sinh \xi \sinh \tau \cosh \vartheta] \, d\vartheta \right\} \times f_2 [\cosh \tau] \sinh \tau \, d\tau , \tag{3.23}$$

where $\alpha = \alpha(\xi,\tau)$ is the positive root of the equation

$$\cosh \xi \cosh \tau - \sinh \xi \sinh \tau \cosh \alpha = 1 . \tag{3.24}$$

Proof: Let K_1 and K_2 be the corresponding Volterra kernels. We have

$$K_1 \# K_2(x, e_2) = \int_X K_1(x,z) \, K_2(z, e_2) \, d\sigma(z) . \tag{3.25}$$

We let

$$x = a_\xi e_2 , \quad y = h_\vartheta a_\tau e_2 .$$

Then the integral becomes

$$K_1 \# K_2(a_\xi e_2, e_2) = \iint K_1(a_\xi e_2, h_\vartheta a_\tau e_2) K_2(a_\tau e_2, e_2) \times \sinh \tau \, d\tau \, d\vartheta . \tag{3.26}$$

Because of the invariance of the kernel K_1 we have

$$K_1(a_\xi e_2, h_\vartheta a_\tau e_2) = K_1(a_{-\tau} h_{-\vartheta} a_\xi e_2, e_2) .$$

If $g = a_{-\tau} h_{-\vartheta} a_\xi$ belongs to G_+, then for Proposition (3.1) we have

$$a_{-\tau} h_{-\vartheta} a_\xi = h_1 a_s h_2 ,$$

with h_1 and h_2 belonging to H; furthermore

$$\cosh s = \cosh \xi \cosh \tau - \sinh \xi \sinh \tau \cosh \vartheta$$
$$\geqslant 1 . \tag{3.27}$$

■

C. The Poisson kernel and the spherical functions

We introduce now the one parameter subgroup N of G consisting in the following matrices:

$$n(z) = \begin{pmatrix} 1 + \frac{1}{2} z^2 & z & \frac{1}{2} z^2 \\ z & 1 & z \\ -\frac{1}{2} z^2 & -z & 1 - \frac{1}{2} z^2 \end{pmatrix}. \tag{3.28}$$

The map

$$N \times A \rightarrow X , \quad (n(z), a_\xi) \mapsto n(z) a_\xi e_2$$

is a diffeomorphism of $N \times A$ on the open set $\{x \in X \mid x_0 + x_2 > 0\}$, which contains the set $\{x \in X \mid x > e_2\}$. If $x = n(z) a_\xi e_2$, we have

$$x_0 = \sinh \xi + \tfrac{1}{2} z^2 e^\xi , \quad x_1 = z e^\xi , \quad x_2 = \cosh \xi - \tfrac{1}{2} z^2 e^\xi ,$$

and $n(z) a_\xi$ belongs to G_+, i.e., $g e_2 > e_2$, if and only if $\xi > 0$ and $|z| < 1 - e^{-\xi}$. It follows that G_+ is contained in NA_+H.

The numbers (z, ξ) are called horicyclic coordinates. In terms of them, the surface element is given by $d\sigma = e^\xi \, dz \, d\xi$.

For a complex number λ we define the function P^λ by

$$P^\lambda(x) = e^{-\lambda\xi} , \quad \text{if } x = n(z) a_\xi e_2 , \tag{3.29}$$

and the Poisson kernel $P^\lambda(x, \vartheta)$ by

$$P^\lambda(x, \vartheta) = P^\lambda(h_{-\vartheta} x) . \tag{3.30}$$

We have

$$P^\lambda(x) = (x_0 + x_2)^{-\lambda} \tag{3.31}$$

$$P^\lambda(a_\xi e_2, \vartheta) = (\cosh \xi + \sinh \xi \cosh \vartheta)^{-\lambda} . \tag{3.32}$$

Proposition 3.3: (a) The function P^λ satisfies the relationship

$$P^\lambda(n(z) a_\xi x) = e^{-\lambda\xi} P^\lambda(x) . \tag{3.33}$$

(b) For $x > e_2$, $\operatorname{Re} \lambda > 0$,

$$\int_H | P^\lambda(hx) | dh = \int_{-\infty}^{+\infty} | P^\lambda(x, \vartheta) | d\vartheta < \infty . \tag{3.34}$$

(c) Furthermore, if g belongs to G_+,

$$\int_H P^\lambda(ghx) \, dh = P^\lambda(g e_2) \int_H P^\lambda(hx) \, dh . \tag{3.35}$$

Proof: To prove (a) we let

$$x = n(z') a_\tau e_2 ,$$

then

$$n(z) a_\xi x = n(z'') a_{\xi + \tau} e_2 ,$$

with

$$z'' = z + e^{-\xi} z'$$

because

$$n(z'') = n(z) a_\xi n(z') a_{-\xi} .$$

(The point is that the subgroup A normalizes the subgroup N.) Hence

$$P^\lambda(n(z) a_\xi x) = e^{-\lambda(\xi + \tau)} = e^{-\lambda\xi} P^\lambda(x) .$$

To prove (b) we may assume that $x = a_\xi e_2$, with $\xi > 0$, then

$$\int_H | P^\lambda(hx) | dh$$
$$= \int_{-\infty}^{+\infty} (\cosh \xi + \sinh \xi \cosh \vartheta)^{-\operatorname{Re} \lambda} d\vartheta < \infty .$$

To prove (c) we let

$$F(g) = \int_H P^\lambda(ghx) \, dh .$$

The function F is right invariant under H. If $g = n(z) a_\xi$ we have

$$F(n(z) a_\xi) = \int_H P^\lambda(n(z) a_\xi hx) \, dh ,$$

and using (a) we obtain

$$F(n(z) a_\xi) = e^{-\lambda\xi} \int_H P^\lambda(hx) \, dh .$$

Since G_+ is contained in NA_+H, (c) is proved. ■

For $\operatorname{Re} \lambda > 0$ we define the spherical function Φ_λ on G_+ by

$$\Phi_\lambda(g) = \int_H P^\lambda(hge_2) \, dh . \tag{3.36}$$

The function Φ_λ is biinvariant under H and

$$\Phi_\lambda [\cosh \xi] = \int_{-\infty}^{+\infty} (\cosh \xi + \sinh \xi \cosh \vartheta)^{-\lambda} d\vartheta . \tag{3.37}$$

This function is essentially the Legendre function of the second kind. With the classical notation we have

$$\Phi_\lambda [\cosh \xi] = 2 Q_{\lambda - 1} (\cosh \xi)$$

[see Ref. 7, Vol. 1, formula 3.7(3), p. 155].

Next we have the following theorem.

Theorem 3.2: The spherical function Φ_λ satisfies a product formula: for g_1 and g_2 in G_+

$$\int_H \Phi_\lambda(g_1 h g_2) \, dh = \Phi_\lambda(g_1) \Phi_\lambda(g_2) . \tag{3.38}$$

This formula can also be written, when $g_1 = a_\xi$ and $g_2 = a_\tau$, as follows:

$$\int_{-\infty}^{+\infty} \Phi_\lambda [\cosh \xi \cosh \tau + \sinh \xi \sinh \tau \cosh \vartheta] \, d\vartheta$$
$$= \Phi_\lambda [\cosh \xi] \, \Phi_\lambda [\cosh \tau] . \tag{3.39}$$

Proof: Using formula (3.35) [see Proposition (3.3)], we have for h and h' in H:

$$\int_H P^\lambda(h' g_1 h g_2 e_2) \, dh = P^\lambda(h' g_1 e_2) \Phi_\lambda(g_2) .$$

The product formula is obtained by integrating with respect to h'. ■

A more general product formula has been proved by Durand for the Legendre functions of the second kind [Ref. 8 formula (13), p. 359, or Ref. 9, formula (3.5), p. 80].

D. The spherical Laplace transform

We define the spherical Laplace transform $\tilde{f}$ of a function f belonging to the Volterra algebra $V(X)^\natural$ by

$$\tilde{f}(\lambda) = \int_X f(x)P^\lambda(x)\, d\sigma(x), \tag{3.40}$$

whenever the integral converges. Integrating in polar coordinates and using the previous notations, we obtain

$$\tilde{f}(\lambda) = \int_0^{+\infty} f[\cosh\xi]\Phi_\lambda[\cosh\xi]\sinh\xi\, d\xi. \tag{3.41}$$

Theorem 3.3: Let $\alpha > 0$, and let $V(X)^\natural_\alpha$ be the space of functions f in $V(X)$ such that

$$\|f\|_\alpha = \int_X |f(x)|P^\alpha(x)\, d\sigma(x) < \infty. \tag{3.42}$$

The space $V(x)^\natural_\alpha$ is a subalgebra of $V(X)^\natural$, and for two functions f_1 and f_2 in $V(x)^\natural_\alpha$ we have

$$\|f_1 \# f_2\|_\alpha \leqslant \|f_1\|_\alpha \|f_2\|_\alpha. \tag{3.43}$$

The spherical Laplace transform $\tilde{f}$ of a function f in $V(x)^\natural_\alpha$ is defined for $\operatorname{Re}\lambda \geqslant \alpha$, analytic for $\operatorname{Re}\lambda > \alpha$, and, for f_1 and f_2 in $V(x)^\natural_\alpha$, we have

$$\widetilde{f_1 \# f_2}(\lambda) = \tilde{f}_1(\lambda)\tilde{f}_2(\lambda). \tag{3.44}$$

Remark: Since

$$\Phi_\lambda[\cosh\xi] \sim c(\lambda)e^{-\lambda\xi} \quad (\xi\to\infty),$$

with

$$c(\lambda) = 2^{(\lambda+1)}\int_0^{+\infty}(1+\cosh\vartheta)^{-\lambda}\, d\vartheta$$
$$= 2\sqrt{\pi}\,\Gamma(\lambda)/\Gamma(\lambda+\tfrac{1}{2}),$$

a function f of $V(X)^\natural$ belongs to $V(x)^\natural_\alpha$ if and only if

$$\int_0^{+\infty} |f[\cosh\xi]|\, e^{(1-\alpha)\xi}\, d\xi < \infty. \tag{3.45}$$

Proof: (a) Let f belong to $V(x)^\natural_\alpha$. For $\operatorname{Re}\lambda \geqslant \alpha$, and for g in G we have

$$\int_X f(g^{-1}x)P^\lambda(x)\, d\sigma(x) = \int_X f(x)P^\lambda(gx)\, d\sigma(x).$$

Using polar coordinates we obtain

$$\int_{H\times A_+} f(a_\xi)P^\lambda(gha_\xi)\, dh \sinh\xi\, d\xi,$$

and using formula (3.25) [Proposition (3.3)] we get

$$\int_{A_+} f(a_\xi)P^\lambda(g)\left(\int_H P^\lambda(ha_\xi)\, dh\right)\sinh\xi\, d\xi$$
$$= P^\lambda(g)\tilde{f}(\lambda).$$

If K is the corresponding Volterra kernel, the previous relationship can be written

$$\int_X K(x,y)P^\lambda(x)\, d\sigma(x) = P^\lambda(y)\int_X K(x,e_2)\, P^\lambda(x)\, d\sigma(x).$$

(b) Let f_1 and f_2 belong to $V(x)^\natural_\alpha$; K_1 and K_2 be the corresponding Volterra kernels. We have

$$\int_X\int_X K_1(x,y)K_2(y,e_2)P^\lambda(x)\, d\sigma(x)\, d\sigma(y)$$
$$= \int_X K_1(x,e_2)P^\lambda(x)\, d\sigma(x)\int_X K_2(y,e_2)P^\lambda(y)\, d\sigma(y),$$

and therefore

$$\widetilde{f_1 \# f_2}(\lambda) = \tilde{f}_1(\lambda)\tilde{f}_2(\lambda).$$ ■

Let us compute the spherical Laplace transform of a function f, using horicyclic coordinates:

$$\tilde{f}(\lambda) = \int_0^{+\infty}\left[\int_{|z|<1-e^{-\xi}} f(n(z)a_\xi)\, dz\right]e^{-(\lambda-1)\xi}\, d\xi. \tag{3.46}$$

We define the Abel transform of f as

$$\mathfrak{A}f(\xi) = e^{\xi/2}\int_{|z|<1-e^{-\xi}} f(n(z)a_\xi)\, dz, \tag{3.47}$$

so that the spherical Laplace transform is the composition of the Abel transform and the usual Laplace transform:

$$\tilde{f}(\lambda) = \int_0^{+\infty}\mathfrak{A}f(\xi)e^{-(\lambda-1/2)\xi}\, d\xi. \tag{3.48}$$

It follows that, under the Abel transform, the composition product of the Volterra algebra $V(X)^\natural$ is transformed in the usual convolution.

We have

$$\mathfrak{A}f(\xi) = e^{\xi/2}\int_{|z|<1-e^{-\xi}} f\left[\cosh\xi - \frac{1}{2}z^2e^\xi\right]dz$$
$$= \int_{|z|<2\sinh(\xi/2)} f\left[\cosh\xi - \frac{1}{2}z^2\right]dz$$
$$= \sqrt{2}\int_0^\xi f[\cosh r]\frac{\sinh r}{\sqrt{\cosh\xi-\cosh r}}\, dr. \tag{3.49}$$

Let us recall the classical inversion of the Abel integral transform: if

$$g(v) = \int_0^v \frac{f(u)}{\sqrt{v-u}}\, du, \tag{3.50}$$

then

$$f(v) = \frac{1}{\pi}\int_0^v \frac{g'(u)}{\sqrt{v-u}}\, du = \frac{1}{\pi}\frac{d}{dv}\int_0^v \frac{g(u)}{\sqrt{v-u}}\, du. \tag{3.51}$$

Therefore the Abel transform $\mathfrak{A}$ is inverted in the following way: if $F(\xi) = \mathfrak{A}f(\xi)$, then

$$f[\cosh\xi] = \frac{1}{\sqrt{2}\pi}\int_0^\xi \frac{F'(r)}{\sqrt{\cosh\xi-\cosh r}}\, \mathrm{dr}$$
$$= \frac{1}{\sqrt{2}\pi}\frac{1}{\sinh\xi}\frac{d}{d\xi}\int_0^\xi \frac{F(r)\sinh r}{\sqrt{\cosh\xi-\cosh r}}\, dr. \tag{3.52}$$

Finally, using the inversion formula for the usual Laplace transform, we obtain the following inversion formula for the spherical Laplace transform.

Proposition 3.4: Let f be a function in $V(x)^{\natural}_{\alpha}$ such that, for $\sigma \geqslant \alpha$,

$$\int_{-\infty}^{+\infty} |\tilde{f}(\sigma+i\nu)|\, |\nu| d\nu < \infty , \quad (3.53)$$

then, for $\xi > 0$,

$$f[\cosh \xi] = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \psi(\xi,\sigma+i\nu) \tilde{f}(\sigma+i\nu)\, d\nu , \quad (3.54)$$

with

$$\psi(\xi,\lambda) = \frac{\lambda - \frac{1}{2}}{\pi\sqrt{2}} \int_0^{\xi} \frac{e^{(\lambda-1/2)r}}{\sqrt{\cosh \xi - \cosh r}}\, dr . \quad (3.55)$$

Proof: As a function of ν, $\tilde{f}(\sigma+i\nu)$ is the usual Fourier transform of

$$e^{-(\sigma-1/2)\xi}F(\xi) = e^{-(\sigma-1/2)\xi}\mathfrak{A}f(\xi)$$

[see formula (3.48)]. Therefore

$$F(\xi) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{(\sigma-1/2+i\nu)\xi}\tilde{f}(\sigma+i\nu)\, d\nu ,$$

and, since

$$\int_{-\infty}^{+\infty} |\tilde{f}(\sigma+i\nu)|\, |\nu|\, d\nu < \infty ,$$

we have

$$F'(\xi) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (\sigma - 1/2 + i\nu) \times e^{(\sigma-1/2+i\nu)\xi}\tilde{f}(\sigma+i\nu)\, d\nu .$$

Therefore, using the inversion formula for the Abel transform and interchanging the order of integration we obtain the result. ■

Remark: Since for $\xi < 0$ we have $F(\xi) = 0$, and therefore $F'(\xi) = 0$, then, for $\xi > 0$

$$\int_{-\infty}^{+\infty} \left(\sigma - \frac{1}{2} + i\nu\right) e^{-(\sigma-1/2+i\nu)\xi}\tilde{f}(\sigma+i\nu)\, d\nu = 0 .$$

Therefore

$$F'(\xi) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \left(\sigma - \frac{1}{2} + i\nu\right) \times \cosh\left[\left(\sigma - \frac{1}{2} + i\nu\right)\xi\right] \tilde{f}(\sigma+i\nu)\, d\nu ,$$

and

$$f[\cosh \xi] = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \vartheta(\xi,\sigma+i\nu) \tilde{f}(\sigma+i\nu)\, d\nu , \quad (3.56)$$

with

$$\vartheta(\xi,\lambda) = \frac{\sqrt{2}(\lambda - \frac{1}{2})}{\pi} \int_0^{\xi} \frac{\cosh(\lambda - \frac{1}{2})\, r}{\sqrt{\cosh \xi - \cosh r}}\, dr .$$

This function can be expressed in terms of the Legendre function of the first kind as follows:

$$\vartheta(\xi,\lambda) = (\lambda - \tfrac{1}{2})\, P_{\lambda-1}(\cosh \xi) . \quad (3.57)$$

The inversion formula of Proposition (3.4) and the last one [Eq. (3.56)], have been obtained by Crönstrom and Klink.[10] The spherical Laplace transform has been studied in a more general setting by Mizony.[11]

E. Volterra algebra on the hyperboloid with one sheet in $\mathbb{R}^4$

A Volterra algebra can be associated to the hyperboloid with one sheet in $\mathbb{R}^n$ defined by

$$-x_0^2 + x_1^2 + x_2^2 + \cdots + x_{n-1}^2 = 1 \quad (3.58)$$

in a way similar to that exposed above; indeed the results are essentially the same.

We will write below, without proof, the explicit form of the results concerning $V(X)^{\natural}$ for X being the hyperboloid with one sheet in $\mathbb{R}^4$: $-x_0^2 + x_1^2 + x_2^2 + x_3^2 = 1$. Indeed, in the latter, we shall need these results.

A function f of $V(X)^{\natural}$ depends only on one variable. For $\xi \geqslant 0$, we write

$$f(h_1 a_{\xi}\, h_2) = f[\cosh \xi] ; \quad h_1,h_2 \in H . \quad (3.59)$$

For two functions f_1 and f_2 in $V(X)^{\natural}$ we have

$$f_1 \# f_2[\cosh \xi] = 2\pi \int_0^{\xi} \left\{ \int_0^{\alpha(\xi,\tau)} f_1[\cosh \xi \cosh \tau - \sinh \xi \sinh \tau \cosh \vartheta] \sinh \vartheta\, d\vartheta \right\} \times f_2\, [\cosh \tau] (\sinh \tau)^2\, d\tau , \quad (3.60)$$

where $\alpha(\xi,\tau)$ is the positive root of the equation

$$\cosh \xi \cosh \tau - \sinh \xi \sinh \tau \cosh \alpha = 1 .$$

This product is related to the ordinary convolution product in the following way: if we let

$$F_1[\cosh \xi] = \int_0^{\xi} f_1[\cosh \tau] \sinh \tau\, d\tau ,$$

we obtain

$$f_1 \# f_2[\cosh \xi] \sinh \xi = 2\pi \int_0^{\xi} F_1[\cosh (\xi - \tau)]\, f_2[\cosh \tau] \sinh \tau\, d\tau . \quad (3.61)$$

The spherical functions are given for $\mathrm{Re}\, \lambda > 1$, by

$$\Phi_{\lambda}[\cosh \xi] = \int_0^{+\infty} (\cosh \xi + \sinh \xi \cosh \vartheta)^{-\lambda} \sinh \vartheta\, d\vartheta = \frac{1}{\lambda - 1} \frac{1}{\sinh \xi} e^{-(\lambda-1)\xi} . \quad (3.62)$$

The spherical Laplace transform reduces to the ordinary Laplace transform as follows. If we assume that, for $\alpha > 2$,

$$\int_0^{+\infty} |f[\cosh \xi]|\, e^{-(\alpha-2)\xi}\, d\xi < \infty , \quad (3.63)$$

then, for $\mathrm{Re}\, \lambda \geqslant \alpha$,

$$\tilde{f}(\lambda) = 2\pi \int_0^{+\infty} \Phi_{\lambda}[\cosh \xi]\, f[\cosh \xi]\, (\sinh \xi)^2\, d\xi = \frac{2\pi}{(\lambda - 1)} \int_0^{+\infty} e^{-(\lambda-1)\xi} f[\cosh \xi] \sinh \xi\, d\xi . \quad (3.64)$$

If we let

$$F[\cosh \xi] = \int_0^{\xi} f[\cosh \tau] \sinh \tau\, d\tau , \quad (3.65)$$

then we obtain

$$\tilde{f}(\lambda) = 2\pi \int_0^{+\infty} e^{-(\lambda-1)\xi} F[\cosh \xi]\, d\xi . \tag{3.66}$$

The spherical Laplace transform carries a convolution product into an ordinary product:

$$\widetilde{f_1 \# f_2}(\lambda) = \tilde{f}_1(\lambda)\tilde{f}_2(\lambda) . \tag{3.67}$$

If we assume furthermore that, for $\sigma > \alpha$,

$$\int_{-\infty}^{+\infty} |\tilde{f}(\sigma + i\nu)|\, |\nu|\, d\nu < \infty , \tag{3.68}$$

then

$$\begin{aligned} f[\cosh \xi] &= \frac{1}{4\pi^2} \frac{1}{\sinh \xi} \frac{d}{d\xi} \int_{-\infty}^{+\infty} e^{(\sigma-1+i\nu)\xi} \tilde{f}(\sigma+i\nu)\, d\nu \\ &= \frac{1}{4\pi^2} \frac{1}{\sinh \xi} \int_{-\infty}^{+\infty} (\sigma - 1 + i\nu)\, e^{(\sigma-1+i\nu)\xi} \tilde{f}(\sigma+i\nu)\, d\nu \\ &= \frac{1}{2\pi^2} \int_{-\infty}^{+\infty} (\sigma - 1 + i\nu) \frac{\sinh[(\sigma-1+i\nu)\xi]}{\sinh \xi} \tilde{f}(\sigma+i\nu)\, d\nu . \end{aligned} \tag{3.69}$$

The proof of the last equality uses the same argument as for proving (3.56).

IV. BETHE–SALPETER EQUATION

A. The Bethe–Salpeter equation as a Volterra equation

We consider the following integral equation for the scattering amplitude A (see Refs. 1–5):

$$A(P,K,Q) = B(P,K,Q) + \int_{\mathbb{R}^4} N(P,P',Q)\, A(P',K,Q)\, d^4P' . \tag{4.1}$$

In this equation, A is an unknown function and represents the amplitude; B and N are supposed to be known and represent the potential and the interaction kernel, respectively. The four momenta P,P',K,Q are graphically represented in Fig. 1 [see also Ref. (3)].

In some instances it may be useful to introduce the so-called Mandelstam variables: $s = (P+K)^2$ = squared energy in the center of mass system; $t = Q^2$ = squared momentum transfer in the center of mass system.

We restrict our attention to the absorptive part of the amplitude; indeed it satisfies an integral equation with the same kinematic structure as the whole amplitude.[3,4] But the support conditions of the absorptive part of the amplitude A, of the potential B and of the interaction kernel N imply that the integral equation (3.1) is a Volterra equation.

Indeed the absorptive part of the amplitude vanishes if $(P+K)^2 < 0$ or if $(P_0+K_0) < 0$ (see, for instance, Refs. 12 and 13). It follows that, as a function of P and K, the support of A is contained in the set

$$\{(P,K) \mid (P+K)^2 > 0 \text{ and } (P_0+K_0) > 0\} . \tag{4.2}$$

The function B has the same property, and N satisfies a similar one: the support of N is contained in the set

$$\{(P,P') \mid (P-P')^2 > 0 \text{ and } (P_0 - P_0') > 0\} . \tag{4.3}$$

Moreover we suppose that the amplitude A, the potential B, and the kernel N are continuous functions on the sets containing the support; this latter assumption is quite restrictive, but at this stage of our research we do not take care of specific and more realistic models for the kernel and the potential.

Therefore by letting $P = x$, $K = -y$, $P' = z$, and forgetting about Q, which is regarded as a fixed parameter (recall that we are working at fixed momentum transfer), the Bethe–Salpeter equation becomes a Volterra equation of the type we considered in Sec. II and we can apply to it the results of that section.

B. Partial diagonalization of the Bethe–Salpeter equation

The functions A, B, and N involved in the Bethe–Salpeter equation are functions depending on three four-vectors, invariant under the Lorentz group $SO_0(1,3)$ acting simultaneously on the three vectors. For fixed Q, the functions A, B, and N are functions depending on two four-vectors, invariant under the subgroup G of the Lorentz transformation fixing the vector Q. We will look at the two following

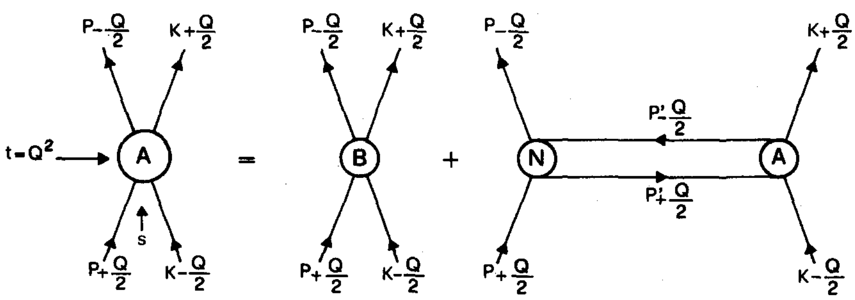


FIG. 1. Graphic representation of the Bethe–Salpeter equation.

cases: (a) forward scattering, in this case $Q=0$ and the group G is $SO_0(1,3)$ itself, and (b) nonforward scattering, in this case $t=Q^2$ is negative. We may choose a coordinate system such that $Q=(0,0,0,\sqrt{-t})$; in this case the group G is then $SO_0(1,2)$.

1. Forward scattering

Before analyzing the Bethe–Salpeter equation, let us consider a Volterra kernel $A(x,y)$ in $\mathbb{R}^4$ invariant under the Lorentz group $SO_0(1,3)$, acting simultaneously on the two vectors x and y. If x and y are spacelike we let

$$x=\rho u, \quad \rho>0; \quad -u_0^2+u_1^2+u_2^2+u_3^2=1, \tag{4.4a}$$

$$y=rv, \quad r>0; \quad -v_0^2+v_1^2+v_2^2+v_3^2=1. \tag{4.4b}$$

The invariance property of $A(x,y)$ implies that $A(x,y)$ depends only on $\rho,r,$ and the inner product $(u,v)=u_0v_0-u_1v_1-u_2v_2-u_3v_3$:

$$A(x,y)=A[\rho,r;(u,v)]. \tag{4.5}$$

The support condition of the kernel A implies that

$$-(u,v)\geqslant(r^2+\rho^2)/2r\rho. \tag{4.6}$$

Let us analyze explicitly the product of two Volterra kernels N and A, invariant under $SO_0(1,3)$:

$$N\#A(x,y)=\int_{D(y,x)}N(x,z)\,A(z,y)\,dz. \tag{4.7}$$

To this purpose we can choose the following coordinates (see also Ref. 4):

$$x=(\rho\sinh\xi,0,0,\rho\cosh\xi), \quad \rho>0, \tag{4.8a}$$

$$y=(0,0,0,r), \quad r>0, \tag{4.8b}$$

$$z=(\rho'\sinh\tau\cosh\vartheta,\ \rho'\sinh\tau\sinh\vartheta\cos\varphi,$$
$$\rho'\sinh\tau\sinh\vartheta\sin\varphi,\ \rho'\cosh\tau),\ \rho'>0. \tag{4.8c}$$

From $(z-y)\in\bar{\Omega}$, we get

$$\tau\geqslant0, \quad \cosh\tau\geqslant(r^2+\rho'^2)/2r\rho', \tag{4.9}$$

which implies

$$\tau\geqslant|\log(\rho'/r)|. \tag{4.10}$$

From the condition $(x-z)\in\bar{\Omega}$ we get

$$\rho\sinh\xi\geqslant\rho'\sinh\tau\cosh\vartheta, \tag{4.11a}$$

$$\cosh\xi\cosh\tau-\sinh\xi\sinh\tau\cosh\vartheta$$
$$\geqslant(\rho^2+\rho'^2)/2\rho\rho'\geqslant1, \tag{4.11b}$$

which imply

$$\xi\geqslant0, \quad \xi-\tau\geqslant|\log(\rho'/\rho)|, \tag{4.12}$$

and

$$|\vartheta|\leqslant\alpha(\xi,\tau), \tag{4.13}$$

where $\alpha=\alpha(\xi,\tau)$ is the positive root of the equation

$$\cosh\xi\cosh\tau-\sinh\xi\sinh\tau\cosh\alpha=1. \tag{4.14}$$

Then

$$0\leqslant|\log(\rho'/\rho)|\leqslant\tau\leqslant\xi-|\log(\rho'/\rho)|\leqslant\xi, \tag{4.15}$$

and it follows that

$$\sqrt{r\rho}\,e^{-\xi/2}\leqslant\rho'\leqslant\sqrt{r\rho}\,e^{\xi/2}. \tag{4.16}$$

We obtain finally

$$N\#A[\rho,r;\cosh\xi]$$
$$=2\pi\int_{\sqrt{r\rho}\,e^{-\xi/2}}^{\sqrt{r\rho}\,e^{\xi/2}}\rho'^3\,d\rho'\int_0^{\xi}d\tau(\sinh\tau)^2A[\rho',r;\cosh\tau]$$
$$\times\left\{\int_0^{\alpha(\xi,\tau)}N[\rho',r;\cosh\xi\cosh\tau\right.$$
$$\left.-\sinh\xi\sinh\tau\cosh\vartheta]\sinh\vartheta\,d\vartheta\right\}, \tag{4.17}$$

which is, precisely, for the integration with respect to τ and ϑ, a convolution product of the type considered in Sec. III E.

Under suitable conditions we can consider the partial spherical Laplace transform of the amplitude A with respect to the angular variables

$$\tilde{A}(\rho,r;\lambda)=2\pi\int_0^{+\infty}\Phi_\lambda[\cosh\xi]\,A[\rho,r;\cosh\xi]$$
$$\times(\sinh\xi)^2\,d\xi$$
$$=\frac{2\pi}{\lambda-1}\int_0^{+\infty}e^{-(\lambda-1)\xi}A[\rho,r;\cosh\xi]$$
$$\times\sinh\xi\,d\xi. \tag{4.18}$$

Since the spherical Laplace transform carries out the convolution product of Volterra kernels on the hyperboloid into the usual product, it is possible to analyze the Bethe–Salpeter equation. We obtain a partial diagonalization of the equation as follows:

$$\tilde{A}(\rho,r;\lambda)$$
$$=\tilde{B}(\rho,r;\lambda)+\int_0^{+\infty}\tilde{N}(\rho,\rho';\lambda)\,\tilde{A}(\rho',r;\lambda)\,\rho'^3\,d\rho'. \tag{4.19}$$

2. Nonforward scattering

We consider now Volterra kernels $A(x,y)$ in $\mathbb{R}^4$, invariant under the Lorentz group $SO_0(1,2)$ acting simultaneously on the two four-vectors x and y. We write $x=(x',x_3)$ with $x'=(x_0,x_1,x_2)$, and also $y=(y',y_3)$ with $y'=(y_0,y_1,y_2)$. The vectors x' and y' are the components of x and y orthogonal to the fixed momentum transfer Q. We will further assume that x' and y' are spacelike.

Next we let

$$x'=\rho u, \quad \rho>0; \quad -u_0^2+u_1^2+u_2^2=1, \tag{4.20a}$$

$$y'=rv, \quad r>0, \quad -v_0^2+v_1^2+v_2^2=1. \tag{4.20b}$$

The invariance property of $A(x,y)$ implies that $A(x,y)$ depends only on ρ, r, (u,v), x_3, and y_3:

$$A(x,y)=A[\rho,r;(u,v);x_3,y_3]. \tag{4.21}$$

For computing the explicit form of the product of two Volterra kernels N and A, invariant under $SO_0(1,2)$, we use the following coordinates:

$$x=(\rho\sinh\xi,0,\rho\cosh\xi,x_3), \quad \rho>0, \tag{4.22a}$$

$$y=(0,0,r,y_3), \quad r>0, \tag{4.22b}$$

$$z=(\rho'\sinh\tau\cosh\vartheta,\ \rho'\sinh\tau\sinh\vartheta,$$
$$\rho'\cosh\tau,z_3), \quad \rho'>0. \tag{4.22c}$$

Now let us determine the limits of integration. Since

$x-z$ belongs to $\bar{\Omega}$, we have $x_0 - z_0 > |x_3 - z_3|$, and therefore

$$x_3 - \rho \sinh \xi + \rho' \sinh \tau \cosh \vartheta < z_3 < x_3 + \rho \sinh \xi - \rho' \sinh \tau \cosh \vartheta . \quad (4.23)$$

Similarly, from $(z-y) \in \bar{\Omega}$ we get

$$y_3 - \rho' \sinh \tau \cosh \vartheta < z_3 < y_3 + \rho' \sinh \tau \cosh \vartheta , \quad (4.24)$$

and adding the inequalities (4.23) and (4.24) we obtain

$$(x_3 + y_3)/2 - \tfrac{1}{2}\rho \sinh \xi < z_3 < (x_3 + y_3)/2 + \tfrac{1}{2}\rho \sinh \xi . \quad (4.25)$$

The determination of the limits of integration for ρ', τ, and ϑ is similar to the forward scattering case. Indeed we have

$$N \# A[\rho, r; \cosh \xi; x_3, y_3] = 2 \int_{\beta_1}^{\beta_2} dz_3 \int_{\sqrt{r\rho}\, e^{-\xi/2}}^{\sqrt{r\rho}\, e^{\xi/2}} \rho'^2 \, d\rho' \times \int_0^{\xi} d\tau \sinh \tau \, A[\rho', r; \cosh \tau; x_3, y_3] \times \left\{ \int_0^{\alpha(\xi,\tau)} N[\rho, \rho'; \cosh \xi \cosh \tau - \sinh \xi \sinh \tau \cosh \vartheta; z_3, y_3] \, d\vartheta \right\} , \quad (4.26)$$

β_1 and β_2 being the lower and upper bounds for z_3 given by formula (4.25).

As in the forward scattering case we consider the partial spherical Laplace transform of a Volterra kernel A with respect to the angular variables

$$\widetilde{A}(\rho, r; \lambda; x_3, y_3) = \int_0^{+\infty} A[\rho, r; \cosh \xi; x_3, y_3] \Phi_\lambda [\cosh \xi] \sinh \xi \, d\xi , \quad (4.27)$$

with

$$\Phi_\lambda [\cosh \xi] = 2 Q_{\lambda - 1} (\cosh \xi) ,$$

where Q_ν denotes the Legendre function of the second kind. Finally we obtain a partial diagonalization of the Bethe–Salpeter equation as follows:

$$\widetilde{A}(\rho, r; \lambda; x_3, y_3) = \widetilde{B}(\rho, r; \lambda; x_3, y_3) + \int_{-\infty}^{+\infty} dz_3 \int_0^{+\infty} \rho'^2 \, d\rho' \, \widetilde{N}(\rho, \rho'; \lambda; x_3, z_3) \times \widetilde{A}(\rho', r; \lambda; z_3, y_3) . \quad (4.28)$$

Remark: As observed by Banarjee *et al.*[14] and by Nussinov and Rosner,[2] the Bethe–Salpeter equation can be further diagonalized, if the equation remains invariant under dilatation. To this purpose we perform a Mellin transform involving the radial variables ρ, ρ', and r. Recall that this dilatation invariance holds true in the limit of zero mass for the internal particles.[2]

[1] L. Sertorio and M. Toller, "Complex angular momentum and Lorentz group," Nuovo Cimento 33, 413 (1964).
[2] S. Nussinov and J. Rosner, "Absorptive parts and the Bethe–Salpeter equation for forward scattering," J. Math. Phys. 7, 1670 (1966).
[3] H. D. I. Abarbanel and L. M. Saunders, "Laplace transforms and the diagonalization of Bethe–Salpeter equations for absorptive parts," Phys. Rev. D 2, 711 (1970).
[4] H. D. I. Abarbanel and L. M. Saunders, "Partial diagonalization of absorptive part equations by Laplace transforms," Ann. Phys. NY 64, 254 (1971).
[5] C. Crönstrom, "Partial diagonalization of Bethe–Salpeter type equations," Ann. Phys. NY 92, 262 (1975).
[6] M. Riesz, "L'intégrale de Riemann–Liouville et le problème de Cauchy," Acta Math. 81, 1 (1949) (in particular, see p. 201).
[7] *Higher Transcendental Functions*, edited by R. Erdelyi (McGraw–Hill, New York, 1953).
[8] L. Durand, "Nicholson-type integrals for products of Gegenbauer functions and related topics," in *Theory and Applications of Special Functions*, edited by R. Askey (Academic, New York, 1975), pp. 353–374.
[9] L. Durand, "Product formulas and Nicholson-type integrals for Jacobi functions I: Summary of results," SIAM J. Math. Anal. 9, 76 (1978).
[10] C. Crönstrom and W. Klink, "Generalized O(1,2) expansions of multiparticle amplitudes," Ann. Phys. NY 69, 218 (1972).
[11] M. Mizony, "Une transformation de Laplace–Jacobi," SIAM J. Math. Anal. 14, 987 (1983).
[12] S. Ferrara, G. Mattioli, G. Rossi, and M. Toller, "Semi-group approach to multiperipheral dynamics," Nucl. Phys. B 53, 366 (1973).
[13] A. Bassetto and M. Toller, "Harmonic analysis on the one-sheet hyperboloid and multiperipheral inclusive distributions," Ann. Inst. H. Poincaré, A 18, 1 (1973).
[14] M. K. Banerjee, M. Kugler, C. A. Levison, and I. J. Muzinich, "High-energy behaviour in field theory and dilatational symmetry," Phys. Rev. B 137, 1280 (1965).

Variational principles and orthogonal polynomials

K. M. Case
The Rockefeller University, New York, New York 10021

(Received 9 July 1985; accepted for publication 22 October 1985)

It is known that there is a close relationship between the theories of scattering and orthogonal polynomials. Here variational principles analogous to those of scattering theory are shown to hold for problems involving difference equations such as those for orthogonal polynomials. Some applications are indicated.

I. INTRODUCTION

Previously,[1] we have seen an important relation between scattering theory and the theory of orthogonal polynomials. A theorem about one implies a theorem about the other.

For some time variational principles have been known for scattering theory.[2] Unfortunately, these tend to be neither minima or maxima. Hence, their usefulness in approximation schemes is rather limited. However, we have recently seen that such variational principles can yield exact results for some quantities of interest. One suspects then that there are variational principles for orthogonal polynomials. This we demonstrate here. As application, we obtain an exact result relating small changes in the spectral function to small changes in the coefficients of the difference equations satisfied by the polynomials.

Our program is (a) we summarize the basic facts of the relation between orthogonal polynomials and scattering theory; (b) a variational principle is obtained; and (c) an application is made.

Closely related to the difference equations describing orthogonal polynomials are the difference equations describing discrete scattering in one dimension. These are of importance in applying the inverse scattering transform to discrete problems—such as the Toda lattice. The variational principles for discrete one-dimensional scattering can readily be found paralleling the approach in this article.

II. SUMMARY

Here we collect known results[1] which will be needed later.

In the theory of orthogonal polynomials, we are given some nondecreasing function $\rho(\lambda)$ defined on the real axis. We are to find polynomials $\psi(\lambda,n)$ such that (i) $\psi(\lambda,n)$ is a polynomial of exact degree n and its leading coefficient is positive; and (ii) the orthonormality relations hold, i.e.,

$$\int_{-\infty}^{\infty} \psi(\lambda,n)\psi(\lambda,m)d\rho(\lambda) = \delta(m,n)\ . \qquad (1)$$

It is then shown that the polynomials satisfy the three-term recursion relation

$$a(n+1)\psi(\lambda,n+1) + b(n)\psi(\lambda,n) + a(n)\psi(\lambda,n-1) = \lambda\psi(\lambda,n)\ , \quad n=0,1,2,\dots\ . \qquad (2)$$

[Here we have defined $a(0)\psi(\lambda,-1)$ to be zero.] From the orthonormality relations, we readily obtain the explicit formulas

$$b(n) = \int_{-\infty}^{\infty} \lambda\psi^2(\lambda,n)d\rho(\lambda)\ , \qquad (3)$$

$$a(n+1) = \int_{-\infty}^{\infty} \lambda\psi(\lambda,n)\psi(\lambda,n+1)d\rho(\lambda)\ . \qquad (4)$$

The question we (almost) will address is what are the relations between small changes in a, b, and ρ. (This will be made more precise later.)

From the viewpoint of scattering theory we take Eq. (2) for $n \geqslant 0$ with the boundary conditions $a(0)\psi(\lambda,-1) = 0$, $\psi(\lambda,0) = C = 1/\sqrt{\int_{-\infty}^{\infty} d\rho(\lambda)}$ as fundamental for the discussion of orthogonal polynomials. Further, we restrict attention to the case when $a(\infty)$ and $b(\infty)$ exist and the limits are approached at least as fast as $1/n^2$. [This is the situation when the support of $d\rho(\lambda)$ is compact.]

Some simplifications are possible (without loss of generality). Thus, we will take $b(\infty) = 0$ and $a(\infty) = \frac{1}{2}$. If $b(\infty) \neq 0$, the spectrum obtained below is merely shifted by $b(\infty)$. If $a(\infty) \neq \frac{1}{2}$, the continuous spectrum is merely stretched by a factor $2a(\infty)$.

Denote as "regular" those solutions of Eq. (2) with the given initial conditions which for a fixed λ are bounded as $n \to \infty$. With the assumed conditions it is readily shown that such solutions exist for all λ such that

$$-1 \leqslant \lambda \leqslant 1\ . \qquad (5)$$

These solutions are conveniently described by z such that

$$\lambda = \tfrac{1}{2}[z + z^{-1}]\ . \qquad (6)$$

The statement then is that the Jacobi matrix formed from the $a(n)$, $b(n)$ has a continuous spectrum in the region defined by Eq. (5), or alternatively for z lying on the unit circle ($z = e^{i\theta}$). In addition, there may be some discrete eigenvalues λ_i corresponding to square summable solutions of Eq. (1). It has been shown that these eigenvalues are (i) real, (ii) simple, (iii) finite in number, and (iv) lie outside or at the edge of the continuum. (In z, they are real and within the unit circle or at $z = \pm 1$.) These results imply that the $\rho(\lambda)$ used to form our orthogonal polynomials has only a finite number of jumps outside the interval $-1 \leqslant \lambda \leqslant 1$ plus a continuous part in the interval.

Useful auxiliary solutions of Eq. (2) are defined for $|z| \geqslant 1$ by

$\lim_{n\to\infty} |\psi_+ - z^n| \to 0$

and for $|z|<1$ by

$\lim_{n\to\infty} |\psi_- - z^{-n}| \to 0.$

Further, we use Eq. (2) to define $f_\pm(z)$ as $f_\pm(z) = a(0)\psi_\pm(z,-1)$. [$f_+(z)$ is called the Jost function since it plays the same role as the function of that name in scattering theory.]

Some needed properties are (i) on the unit circle

$$\psi_+(z,n) = \psi_-^*(z,n) = \psi_-(z^{-1},n);$$

(ii) $\psi(\lambda,n) = (C/i\sin\theta) \times [f_-(z)\psi_+(z,n) - f_+(z)\psi_-(z,n)]$, for $z=e^{i\theta}$;

(iii) in particular, the asymptotic behavior as $n\to\infty$ is $\psi(\lambda,n)\to[2C|f_+(z)|/\sin\theta]\sin(n\theta+\eta)$, where

$$\eta(\theta) = -\arg f_+(z); \tag{7}$$

and (iv) the zeros of $f_+(z)$ in the unit circle determine the discrete eigenvalues.

It is clear that f_+ plays a fundamental role in the theory of orthogonal polynomials. From its definition, f_+ depends only on the coefficients $a(n),b(n)$. The principal application of the variational principle we will develop will be to elucidate this.

First, we will need a representation for f_+.

III. A REPRESENTATION FOR f_+

This follows from the following properties of $f_+(z)$.

(i) On the unit circle $S = e^{2i\eta} = f_+(1/z)/f_+(z)$.

(ii) $f_+(z)$ is analytic within the unit circle except for a simple pole at $z=0$ where the residue is

$$R = \frac{1}{2}\prod_{i=1}^{\infty}\frac{1}{2a(n)}. \tag{8}$$

(iii) There are at most a finite number of simple zeros of f_+ within the unit circle. These are at real points z_i.

These then imply that

$$f_+(z) = \frac{Re^{\Gamma(z)}}{z}\prod_{i=1}^{N}\left(1-\frac{z}{z_i}\right), \tag{9}$$

where

$$\Gamma(z) = -\frac{1}{\pi}\oint_c \frac{\eta'(z')dz'}{z'-z}, \tag{10}$$

$$\eta' = \bar{\eta} + \frac{1}{2}\arg\prod_{i=1}^{N}\frac{(z'-z_i)}{(1/z'-z_i)}, \quad \bar{\eta} = \eta(z') - \theta', \tag{11}$$

and c is the unit circle. [The argument leading to Eq. (9) is essentially one previously given.[3] The only difference is that here f_+ has a simple pole at $z=0$ with known residue. Previously, the function $P(z)$ had no pole but the value of $P(0)$ was known.]

IV. A QUESTION

We now can ask what are the first-order changes in f_+ when we consider small changes $\delta a(n)$, $\delta b(n)$ in the coefficients.

From the integral representation of Eq. (9) we see we need $\delta R/\delta a$, $\delta R/\delta b$, $\delta z_i/\delta a$, $\delta z_i/\delta b$, and $\delta\bar{\eta}/\delta a$, $\delta\bar{\eta}/\delta b$. The first two of these variations follow trivially from the explicit expression of Eq. (8). Indeed,

$$\frac{\delta R}{\delta a(n)} = -\frac{R}{a(n)}, \tag{12}$$

and

$$\frac{\delta R}{\delta b(n)} = 0. \tag{13}$$

The functional derivatives $\delta z_i/\delta a$ and $\delta z_i/\delta b$ can be obtained by a simple argument which then suggests how to find $\delta\bar{\eta}/\delta a$ and $\delta\bar{\eta}/\delta b$. Note that

$$\lambda_i = (z_i + z_i^{-1})/2 \tag{14}$$

is a discrete eigenvalue. Therefore,

$$\lambda_i\,\psi(\lambda_i,n) = a(n+1)\psi(\lambda_i,n+1) + b(n)\psi(\lambda_i,n) + a(n)\psi(\lambda_i,n-1). \tag{15}$$

Multiplying by $\psi(\lambda_i,n)$ and summing over all n, gives the expression

$$\lambda_i = \left[2\sum_n a(n+1)\psi(\lambda_i,n)\psi(\lambda_i,n+1) + \sum_n b(n)\psi^2(\lambda_i,n)\right]\left[\sum_n \psi^2(\lambda_i,n)\right]^{-1}. \tag{16}$$

Now regard this as a functional $\lambda(\psi)$. Consider variations of this λ in the vicinity of $\psi\equiv\psi(\lambda_i,n)$. (Here we imagine the a,b held fixed.) Then, in virtue of Eq. (15),

$$\left.\frac{\delta\lambda}{\delta\psi}\right|_{\psi=\psi(\lambda_i,n)} = 0. \tag{17}$$

This, of course, is merely a discrete form of the Rayleigh–Ritz principle. In the usual quantum mechanical application—to obtain approximations to the eigenvalues—we most often consider the lowest eigenvalue and make use of the minimum property to improve accuracy. However, while not necessarily a maximum or a minimum, the functional is stationary at each eigenvalue.

Let us now use Eq. (16) to compute the change in λ_i due to small changes in $a(n)$, $b(n)$. Since we have demonstrated the stationary property, we need only calculate the change in λ_i due to the changes where a and b occur explicitly. Thus,

$$\frac{\delta\lambda_i}{\delta a(n)} = \frac{2\psi(\lambda_i,n)\psi(\lambda_i,n-1)}{\Sigma_m\,\psi^2(\lambda_i,m)} \tag{18a}$$

and

$$\frac{\delta\lambda_i}{\delta b(n)} = \frac{\psi(\lambda_i^2,n)}{\Sigma_m\,\psi^2(\lambda_i,m)}. \tag{18b}$$

From Eq. (14) we see $\delta\lambda_i = \frac{1}{2}(1-1/z_i^2)\delta z_i$,

$$\therefore\ \frac{\delta z_i}{\delta a(n)} = \left(\frac{2}{1-z_i^{-2}}\right)\frac{2\psi(\lambda_i,n)\psi(\lambda_i,n-1)}{\Sigma_m\,\psi^2(\lambda_i,m)} \tag{19a}$$

and

$$\frac{\delta z_i}{\delta b(n)} = \left(\frac{2}{1-z_i^{-2}}\right)\frac{\psi^2(\lambda_i,n)}{\Sigma_m\,\psi^2(\lambda_i,m)}. \tag{19b}$$

This derivation suggests how we may calculate the functional derivatives $\delta\bar{\eta}/\delta a(n)$ and $\delta\bar{\eta}/\delta b(n)$. If we can construct a functional for $\bar{\eta}$, which is a stationary problem, then the functional derivatives can be computed using only the explicit changes in $a(n)$ and $b(n)$. We turn to the construction of such a variational principle.

V. THE VARIATIONAL PRINCIPLE

Let us construct an integral form of the difference equation. From this several expressions for the phase shift will be given. Combining these will give the variational principle.

Rewrite Eq. (2) in the form

$$\tfrac{1}{2}\{\psi(\lambda,n+1)+\psi(\lambda,n-1)\}-\lambda\psi(\lambda,n)=\gamma(n)\,, \tag{20}$$

where

$$\begin{aligned}\gamma(n) = &-[a(n+1)-\tfrac{1}{2}]\psi(\lambda,n+1)\\ &-[a(n)-\tfrac{1}{2}]\psi(\lambda,n-1)\\ &-b(n)\psi(\lambda,n)\,.\end{aligned} \tag{21}$$

The homogeneous form of Eq. (20),

$$\tfrac{1}{2}\{\psi_0(\lambda,n+1)+\psi_0(\lambda,n-1)\}-\lambda\psi_0(\lambda,n)=0\,, \tag{22}$$

has as a solution vanishing at $n=-1$

$$\psi_0^{(1)}(\lambda,n)=z^{n+1}-z^{-(n+1)}\equiv 2i\sin(n+1)\theta\,. \tag{23}$$

A second solution of Eq. (22) is

$$\psi_0^{(2)}(\lambda,n)=z^{n+1}+z^{-(n+1)}\,. \tag{24}$$

Then, $\psi_0^{(1)}$ and $\psi_0^{(2)}$ are linearly independent since

$$\begin{aligned}W[\psi_0^{(1)},\psi_0^{(2)}] &= \psi_0^{(1)}(\lambda,n+1)\psi_0^{(2)}(\lambda,n)\\ &\quad-\psi_0^{(2)}(\lambda,n+1)\psi_0^{(1)}\\ &=2(z-z^{-1})\neq 0\,.\end{aligned} \tag{25}$$

A Green's function satisfying

$$\tfrac{1}{2}\{G(\lambda,m+1;n)+G(\lambda,m-1;n)\}-\lambda G(\lambda,m;n)=\delta(m,n)$$

is then readily constructed as

$$\begin{aligned}G(\lambda,m;n) &= -\frac{\psi_0^{(1)}(\lambda,n)\psi_0^{(2)}(\lambda,m)}{z-z^{-1}}\,,\quad m\geqslant n\,,\\ &= -\frac{\psi_0^{(2)}(\lambda,n)\psi_0^{(1)}(\lambda,m)}{z-z^{-1}}\,,\quad m\leqslant n\,,\end{aligned} \tag{26}$$

An "integral equation" for the scattering problem incorporating both the boundary condition and the difference equation is then

$$\psi(\lambda,n)=\psi_0^{(1)}(\lambda,n)+\sum_m G(\lambda,m;n)\gamma(m)\,. \tag{27}$$

As $n\to\infty$, we then see that

$$\psi(\lambda,n)\to z^{n+1}-\bar{z}^{(n+1)}+E\,[z^{n+1}+\bar{z}^{(n+1)}]\,, \tag{28}$$

with

$$E=-\frac{1}{z-z^{-1}}\sum_m \psi_0^{(1)}(\lambda,m)\gamma(m)\,. \tag{29}$$

The phase shift η is defined so that

$$\psi(\lambda,n)\underset{n\to\infty}{\sim} D\sin(n\theta+\eta)=D\sin[(n+1)\theta+\bar{\eta}]\,. \tag{30}$$

Since $z^{n+1}-z^{-(n+1)}=2i\sin(n+1)\theta$ and $z^{n+1}+z^{-(n+1)}=2\cos(n+1)\theta$, we see on comparing Eqs. (28) and (30) that $D\cos\bar{\eta}=2i$, $D\sin\bar{\eta}=2E$, and, therefore,

$$\tan\bar{\eta}=\frac{2E}{2i}=\sum_m\frac{\psi_0^{(1)}(\lambda,m)\gamma(m)}{2\sin\theta}\equiv\frac{K}{2\sin\theta}\,. \tag{31}$$

An alternative expression is readily obtained. Let

$$J=\sum_n \gamma(n)\psi(\lambda,n)-\sum_{mn}G(\lambda,m;n)\gamma(m)\gamma(n)\,. \tag{32}$$

But for the solution of our problem

$$\sum_m G(\lambda,m;n)\gamma(m)=\psi(\lambda,n)-\psi_0^{(1)}(\lambda,n)\,,$$

$$\therefore J=\sum_n\gamma(n)\psi_0^{(1)}(\lambda,n)=2\sin\theta\tan\bar{\eta}\,,$$

i.e.,

$$\tan\bar{\eta}=J/(2\sin\theta) \tag{33}$$

is another expression.

Combining the two expression of Eq. (31) and (33), we also have[4]

$$\tan\bar{\eta}=[1/(2\sin\theta)]K^2/J\,. \tag{34}$$

It is now maintained that this expression as a functional of ψ is stationary when the ψ satisfying the integral equation is used. Thus, consider variations in ψ keeping $a(n)$, $b(n)$ fixed. We have

$$\delta\tan\bar{\eta}=[1/(2\sin\theta)]\{2\delta K-\delta K\}\,,$$

since when ψ is the correct solution $K=J$. One has $\delta K=\Sigma_m\,\psi_0^{(1)}(\lambda,m)\delta\gamma(m)$ with

$$\begin{aligned}\delta\gamma(m) = -\sum_m\{&[a(m+1)-\tfrac{1}{2}]\delta\psi(\lambda,m+1)\\ &+[a(m)-\tfrac{1}{2}]\delta\psi(\lambda,m-1)\\ &+b(m)\delta\psi(\lambda,m)\}\,,\end{aligned} \tag{35}$$

and

$$\begin{aligned}\delta J=&\sum_n\delta\gamma(n)\psi(\lambda,n)+\sum_n\gamma(n)\delta\psi(\lambda,n)\\ &-\sum_{mn}\gamma(n)G(\lambda,m;n)\delta\gamma(m)\\ &-\sum_{mn}\delta\gamma(n)G(\lambda,m;n)\gamma(m)\,.\end{aligned}$$

But

$$\begin{aligned}&\sum_{mn}\gamma(n)G(\lambda,m;n)\delta\gamma(m)\\ &\quad=\sum_{mn}\gamma(m)G(\lambda,n;m)\delta\gamma(n)\\ &\quad=\sum_n[\psi(\lambda,n)-\psi_0^{(1)}(\lambda,n)]\delta\gamma(n)\,,\end{aligned}$$

$$\begin{aligned}\therefore\ \delta J=&\,2\sum_n\delta\gamma(n)\psi_0^{(1)}(\lambda,n)\\ &+\sum_n\gamma(n)\delta\psi(\lambda,n)-\sum_n\psi(\lambda,n)\delta\gamma(n)\,.\end{aligned} \tag{36}$$

Inserting Eq. (35) for $\delta\gamma(n)$ and doing some relabeling, we find the last two terms on the right just cancel,

$$\therefore \delta J = 2\sum_n \delta\gamma(n)\psi_0^{(1)}(\lambda,n) = 2\delta K, \tag{37}$$

and so $\delta \tan \bar{\eta} = 0$ as claimed.

We have shown $\tan \bar{\eta}$ is stationary with respect to small variations in the wave functions. Hence, to calculate changes due to small changes in $a(n)$, $b(n)$, we merely need to look at the explicit changes in our variational expression.

As before

$$\delta \tan \bar{\eta} = \frac{1}{2\sin\theta}\{2\delta K - \delta J\},$$

$$\delta K = \sum_n \psi_0^{(1)}(\lambda,n)\delta\gamma(n),$$

where now

$$\delta\gamma(m) = -\delta a(n+1)\psi(\lambda,n+1) - \delta a(n)\psi(\lambda,n-1) - \delta b(n)\psi(\lambda,n),$$

$$\delta J = \sum_n \delta\gamma(n)\psi(\lambda,n) - \sum_{mn} G(\lambda,m;n)\delta\gamma(m)\gamma(n) - \sum_{mn} G(\lambda,m;n)\gamma(m)\delta\gamma(n).$$

Since

$$\sum_m G(\lambda,m;n)\gamma(m) = \psi(\lambda,n) = \psi_0^{(1)}(\lambda,n),$$

$$\delta J = \sum_n \psi_0^{(1)}(\lambda,n)\delta\gamma(n) - \sum_{mn} G(\lambda,m;n)\delta\gamma(m)\gamma(n),$$

we have

$$\sum_{mn} G(\lambda,m;n)\delta\gamma(m)\gamma(n) = \sum_{mn} G(\lambda,n;m)\gamma(m)\delta\gamma(n) = \sum_n \psi(\lambda,n) - \psi_0^{(1)}(\lambda,n),$$

$$\therefore \delta J = 2\sum_n \psi_0^{(1)}(\lambda,n)\delta\gamma(n) - \sum_n \psi(\lambda,n)\delta\gamma(n).$$

Hence,

$$\delta \tan \bar{\eta} = -\frac{1}{2\sin\theta}\sum_n \psi(\lambda,n)\delta\gamma(n), \tag{38}$$

and then

$$\frac{\delta \tan \bar{\eta}}{\delta a(n)} = -\frac{1}{2\sin\theta}\psi(\lambda,n)\psi(\lambda,n-1), \tag{39}$$

$$\frac{\delta \tan \bar{\eta}}{\delta b(n)} = -\frac{1}{2\sin\theta}\psi^2(\lambda,n). \tag{40}$$

VI. CONCLUSION

It has been demonstrated that the difference equations describing orthogonal polynomials have variational principles completely analogous to those occurring in the theory of scattering by potentials. These are principles for the eigenvalues and phase shifts.

As an application, it has been seen how the functional derivatives of the Jost function for discrete problems can be found via the variational principles.

It may be mentioned that there are other problems involving difference equations which are closely related to those discussed here. For example, in applying the inverse scattering methods to discrete problems—such as the Toda lattice—one is interested in difference equations like Eq. (2), but for $-\infty < n < \infty$. Then the primary quantities of interest are the reflection and transmission coefficients. Variational principles for these are readily obtained. These are straightforwardly obtained discrete forms of the principles discussed in Ref. 5. From these the functional derivatives of the reflection and transmission functions with respect to the coefficients in the difference equations are easily found.

ACKNOWLEDGMENT

This work was supported in part by the Department of Energy under Contract No. DE-AC02-81ER10807.006.

[1]K. M. Case, J. Math. Phys. **15**, 2166 (1974).

[2]See, for example, P. M. Morse and H. Feshbach, in *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Chap. 9.

[3]K. M. Case and M. Kac, J. Math. Phys. **14**, 594 (1973).

[4]See Ref. 2, p. 1123.

[5]M. D. Arthur and K. M. Case, J. Math. Phys. **23**, 1771 (1982).

Spinors and space-times

Krystyna Bugajska
Institut für Theoretische Physik der Technischen, Universität Clausthal, 3392 Clausthal-Zellerfeld, West Germany

(Received 5 December 1984; accepted for publication 27 September 1985)

Dirac spinors for vector space-times $R^{s,t}$ are considered as real spinors of appropriate extended vector spaces $R^{s',t'}$. This extension is determined by the condition $R_{s',t'} \cong R^{\mathbb{C}}_{s,t}$. The catalog of reality, chirality conditions allowed for Dirac spinors by the analysis of corresponding Clifford algebras and their left minimal ideals are completed.

I. INTRODUCTION

In supersymmetric field theories, as well as in some unified theories of gravitation and gauge interactions, we are specially interested in various space-time dimensions. The fundamental problem for all such theories is an analysis of the kind of spinors that can be defined in the n-dimensional space-time of the signature (s,t); $s + t = n$.

In this article we will investigate algebraic properties of spinors and their consequences on the possible physical theories. In some sense we shall follow in the spirit of the articles in Refs. 1–3, however, some of our conclusions will be different.

In supersymmetric field theories we have to deal with different kinds of division algebras and for some space-time, appropriate spinor spaces are also modules over some division algebras. Hence the idea arises to link the properties of supersymmetric field theories in various space-time dimensions with the properties of spinors.

Although this idea has been used previously in the literature[1] the conclusion arrived at is doubtful. First, the association of the given space-time with one of the division algebras $\mathbb{R}$,$\mathbb{C}$, or $\mathbb{H}$ does not lead us to the conclusion that there is some relationship between the dimension n (or "transverse" dimension $n-2$) and division algebra. Also we cannot anticipate the further association $n = 10$ with the algebra of octonions $\mathbb{O}$. The known relation between $\mathrm{Spin}(7) \subseteq \mathrm{SO}(8)$ and the subgroup of $\mathrm{GL}_R(\mathbb{O})$ generated by right multiplication by unit octonions, which are purely imaginary, has a completely different origin. It comes from the fact that every linear transformation of R^8 can be written[4] as a sum of eight transformations of the form $x \to (ax)b$, where the product is taken from the algebra of octonions. The similar situation is met in the four-dimensional case. Then every linear transformation of R^4 can be written as a sum of four transformations of the form $x \to axb$ with the product from the quaternion algebra $\mathbb{H}$. It implies that the Spin (4) group is given by $x \to pxq^{-1}$, where p and q are pure imaginary unit quaternions.

In this paper we shall consider spinors as elements of the left minimal ideals of the corresponding Clifford algebra. However, if we start with some vector space-time $R^{s,t}$ then the left minimal ideals $\mathscr{S}(s,t)$ of its Clifford algebra $R_{s,t}$ are modules over $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$, respectively, depending on the concrete signature (s,t). Unfortunately, for even $(s + t = 2r)$-dimensional vector space-time $R^{s,t}$, spinor spaces are over $\mathbb{R}$ or $\mathbb{H}$. For these reasons we construct Dirac spinors as elements of a left minimal ideal of the complexified Clifford algebra $R^{\mathbb{C}}_{s,t}$. However $R^{\mathbb{C}}_{s,t}$ is isomorphic to the universal Clifford algebra C_n of the complexified vector space-time $(R^{s,t})^{\mathbb{C}} = \mathbb{C}^n$. Thus we see that C_n does not reflect the algebraic properties of the appropriate Clifford algebra prior to the complexification. In other words we lose the algebraic properties of $R_{s,t}$ related to concrete signature (s,t). Sometimes we may rediscover the relation between the starting vector space-time and one of the division algebras $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$ but only when using isomorphism between Spin (s,t) group and one of the classical groups.

In this paper we proceed in a slightly different way. Namely we embed the vector space $R^{s,t}$ into the corresponding vector space $R^{s',t'}$ (where $2r = s + t = s' + t' - 1$) determined by the condition $R^{\mathbb{C}}_{s,t} \cong R_{s',t'}$. The advantages of such an approach are the following.

(a) So defined Dirac spinor space inherits the symmetries determimed by scalar products induced by anti-involutions of real Clifford algebras.

(b) The above symmetries are closely related to the symmetries of maximal supergravities in higher-dimensional space-times. It suggests that we should work with real Clifford algebra approach.[2]

(c) We obtain a precise interdependence between a dimension and signature of any space-time $R^{s,t}$ on the one side and the division algebras related to pinor and spinor modules of $R_{s,t}$ on the other.[5]

However, if we take into account the Ogievetsky and Sokatchev construction of the supergravity potential by means of complex space-time coordinates then we should consider also the complexified Chevalley approach.

II. GENERAL CONSIDERATION

Let $R^{s,t}$ be an $(m = s + t)$-dimensional vector space-time with the space dimension s and the time dimension t. Let us denote by $R_{s,t}$ its corresponding universal Clifford algebra (CA). It is known[5] that any CA $R_{s,t}$ can be realized by its matrix representation. It exhibits the character of $R_{s,t}$ as the real algebra of endomorphisms of a right F-linear space $\mathscr{S}(s,t) \equiv \mathscr{S}$, i.e.,

$$R_{s,t} \leftrightsquigarrow \mathrm{End}_F \mathscr{S}. \qquad (2.1)$$

Here F is an appropriate ring uniquely related with a given CA $R_{s,t}$. A space $\mathscr{S}$ is called the "real" spinor space of the orthogonal vector space $R^{s,t}$. Thus we treat the spinors as elements of the underlying vector space of the faithful matrix

 0022-2488/86/030853-06$02.50

representation of the CA $R_{s,t}$. This underlying vector space can be given by an minimal left ideal of $R_{s,t}$:

$$\mathcal{S}(s,t) = R_{s,t} f. \quad (2.2)$$

Here f is a primitive idempotent $f^2 = f$ of $R_{s,t}$, which uniquely determines $\mathcal{S}(s,t)$. Sometimes, the left minimal ideal $\mathcal{S}$ as a right F-module is called a pinor module,[6] whereas one introduces also the notion of spinor module $\Sigma \equiv \Sigma(s,t)$ as a left minimal ideal of the even subalgebra $R^{(0)}_{s,t}$ of the CA $R_{s,t}$. The first to introduce spinors as elements of a left minimal ideal of corresponding CA was Sauter.[7]

Although for a fixed dimension $n = s + t$ the geometric properties of a given CA $R_{s,t}$ depend on the concrete signature, we have only a few possibilities for their algebraic types. Namely, for $n = s + t = 2r$, CA $R_{s,t}$ can be realized only as the algebra of $2^r \times 2^r$ real matrices $\mathbf{R}(2^r)$ or as the algebra $\mathbf{H}(2^{r-1})$ of $2^{r-1} \times 2^{r-1}$ matrices with quaternionic entries. For n odd, i.e., $n = 2r + 1$, CA $R_{s,t}$ can be only in one of the following algebraic types:

$$^2\mathbf{R}(2^r), \quad \mathbb{C}(2^r), \quad ^2\mathbf{H}(2^{r-1}). \quad (2.3)$$

Moreover, because we have algebraic isomorphisms between

$$R^{(0)}_{s,t} \cong R_{s,t-1}, \quad \text{for } t > 1,$$

and (2.4)

$$R^{(0)}_{s,t} \cong R_{t,s-1}, \quad \text{for } s > 1,$$

we see that we can obtain a link between a given space-time $R^{s,t}$, but only one of the division algebras $\mathbb{R}$, $\mathbb{C}$, or $\mathbf{H}$. Thus the suggestion[1] that octonion algebra could come naturally to our spinor analysis when we increase the dimension of space-time is not true.

Any space $\mathcal{S}(s,t)$ is the underlying space not only for the faithful representation of the corresponding CA $R_{s,t}$, but among others also for Spin(s,t) group and for the group $G_\pm$ determined by scalar products on $\mathcal{S}(s,t)$. The former group is related with the symmetry of the whole physical theory based on a given space-time $R^{s,t}$ whereas the latter groups are related[2] with symmetries of maximal supergravities in dimension n.

In the physical theories one of the most fundamental equations is the Dirac equation

$$(i\partial\!\!\!/ + m)\psi = 0, \quad \partial\!\!\!/ = \gamma^\mu \partial_\mu. \quad (2.5)$$

Here γ^μ are the complex matrices that satisfy the relations

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}, \quad \eta^{\mu\nu} = \text{diag}(\underbrace{+ + \cdots +}_{s} \underbrace{- \cdots -}_{t}),$$

and ψ is a sufficiently differentiable function from $R^{s,t}$ to a 2^r-dimensional complex vector space equipped with Spin(s,t) symmetry.

We see that only in the case when $R_{s,t}$ has $\mathbb{C}(2^r)$ realization, matrices γ_μ generate CA $R_{s,t}$ and ψ can be taken as a function from $R^{s,t}$ to $\mathcal{S}(s,t)$.

Although in the case when CA has its matrix representation with the quaternionic entries, we can still believe that (by means of the well-known realization of the quaternions in terms of Pauli matrices) we are able to preserve the validity of (2.5); it is obvious that this is completely impossible when the algebraic type of $R_{s,t}$ is given by $\mathbf{R}(2^r)$.

The main obstruction against the validity of Dirac equation (2.5) in terms of the geometric objects related to $R^{s,t}$ is the presence of the imaginary unit i. For this reason in the relativistic quantum mechanics we have to pass either to the complexified space-time $(R^{3,1})^{\mathbb{C}}$ or to a one-dimensional extension $R^{4,1}$ of our space-time $R^{3,1}$. In both cases we introduce the imaginary unit i to appropriate Clifford algebras as well as to related spinor spaces. Of course it is in contradiction with Dirac's original intension of factoring the wave operator in four-dimensional real space-time.

We can consider the just mentioned two possibilities because the algebraic types of $R^{\mathbb{C}}_{3,1}$ and $R_{4,1}$ are the same and are given by the full matrix algebras $\mathbb{C}(4)$. Let us return to the Dirac mass equation (2.5). By construction it can be satisfied only in the case when γ_μ matrices have complex entries as well as when the field ψ has complex components. Such a complex matrix representation of the generators γ_μ of an appropriate Clifford algebra will be called the Dirac algebra and related with its spinor space will be called the Dirac space. However, for even-dimensional vector space $R^{s,t}$ we can never obtain the dirac algebra as its Clifford algebra. The simplest way is to pass to the complexified picture. But in this case we lost the dependence on the concrete signature (s,t) and we lost the natural possibility of aditional symmetries of Dirac spinor spaces related with scalar products; besides, we have to introduce more than one additional real dimension. For this reason it seems to be more natural and convenient to pass to the only one-dimensional extended vector space R^{n+1}. We can do this by considering our starting vector space $R^{s,t}$ as a subspace of an appropriate $(n+1)$-dimensional vector space formed by deleting from the canonical basis $(e_0, e_1, \ldots, e^n)$ for R^{n+1} a vector e_0.

The additional dimension e_0 is timelike or spacelike depending on the signature (s,t).

Our aim is to pass to such a bigger vector space whose corresponding Clifford algebra is algebraicly isomorphic to the complexified CA of the starting orthogonal space $R^{s,t}$,

$$\begin{array}{ccl} R^{s,t} & \hookrightarrow & R^{s',t'} \\ \downarrow & & \downarrow \\ R_{s,t} & \hookrightarrow & R_{s',t'} =: D_{s,t}. \\ & \searrow & \wr\| \\ & & R^{\mathbb{C}}_{s,t} \end{array} \quad (2.6)$$

Here $s + t = s' + t' - 1$.

We shall denote a left minimal ideal of $R_{s,t}$ (i.e., a "real" spinor space) by $\mathcal{S} = \mathcal{S}(s,t)$; a left minimal ideal of $R^{(0)}_{s,t}$ (i.e., a space of "even" spinors) by $\Sigma = \Sigma(s,t)$ and a left minimal ideal of $D_{s,t} := R_{s',t'}$ (i.e., a space of Dirac spinors) by $\Psi = \Psi(s,t)$.

As we have told for even-dimensional vector space-time $R^{s,t}$, $s + t = 2r$, we have two algebraic types of universal Clifford algebras:

$$\text{(I) } R_{s,t} \cong \mathbf{R}(2^r) \quad \text{and} \quad \text{(II) } R_{s,t} \cong \mathbf{H}(2^{r-1}). \quad (2.7)$$

It appears that for each of these cases we have to consider two possibilities: with additional spacelike dimension, and with additional timelike dimension. This space- or timelike character of the extension of $R^{s,t}$ to $R^{s',t'}$ depends on the concrete signature (s,t).

For odd-dimensional vector space-time $R^{s,t}$, $s + t$

$=2r+1$, we have that either $R_{s,t}$ is realized as a matrix algebra with complex entries, i.e., $\mathbb{C}(2^r)$, or as a matrix algebra over some, the so-called double field. Thus in the former case we have just what we need, whereas in the latter cases our Clifford algebras do not possess "one generator's" extension having the complex matrix realization. Hence in these cases all we can do is to pass to the complexification $\mathbb{C}^n$ of a starting vector space $R^{s,t}$ itself.

III. DIRAC, WEYL, AND MAJORANA SPINORS

Let $R^{s,t}$ be an even-dimensional orthogonal vector space. Let $R^{s',t'}$, $s'+t'=s+t+1$ be a vector space whose corresponding universal Clifford algebra $R_{s',t'}$ is algebraically isomorphic to $R^{\mathbb{C}}_{s,t}$. As we have told, a Dirac spinor space $\Psi(s,t)$ is (by definition) formed by a minimal left ideal of $R_{s',t'}$, i.e., $\Psi(s,t)=\mathscr{S}(s't')$.

Now the additional dimension e_0 allows us to construct the projective operators $W_\pm$ according to the formula

$$W_\pm=\tfrac{1}{2}(1+\eta e_0). \tag{3.1}$$

It is easy to see that $W^2_\pm=W_\pm$ if $\eta=1$ for $e_0^2=+1$ and $\eta=i$ when $e_0^2=-1$.

We shall call operators $W_\pm$ Weyl operators. They decompose the Dirac 2^r complex-dimensional space Ψ onto two Weyl subspaces Ψ_+ and Ψ_-, respectively. Although the space Ψ is equipped with higher symmetries G_+ and G_- the operators $W_\pm$ break them and subspaces Ψ_+ and Ψ_- inherit only appropriate Spin(s,t) symmetry.

Let $\{e_1,\ldots,e_{s+t}\}$ form an orthogonal basis for the corresponding vector space $R^{s,t}$. Let us denote their product by e_J, i.e.,

$$e_J=e_1e_2\cdots e_{s+t}. \tag{3.2}$$

Because the product e_0e_J plays the role of a pure imaginary unit in CA $R_{s',t'}=D_{s,t}$ for any (s,t), the Weyl operators can be written

$$W_\pm=\tfrac{1}{2}(1\mp\eta e_J) \tag{3.3}$$

($ie_J=e_0$ for $e_0^2=-1$ and $ie_J=-e_0$ for $e_0^2=+1$). It can be checked that for $s-t=2, 6 \bmod 8$ the additional dimension e_0 has to be a space-line, i.e., $e_0^2=+1$. We shall denote these cases by A (see Table I). Although the algebraic type of $R_{s,t}$ for $s-t=2 \bmod 8$ is given by $\mathbb{R}(2^r)$ algebra, and for $s-t=6 \bmod 8$ is given by $\mathbb{H}(2^{r-1})$ algebra, their even Clifford subalgebras are isomorphic. They are realized as the matrix algebra $\mathbb{C}(2^{r-1})$. Namely for $s-t=6 \bmod 8$ we have $t-s=2 \bmod 8$. Hence the known isomorphism

$$R^{(0)}_{s,t}\cong R^{(0)}_{t,s}$$

tells us that the Clifford algebras to type A have algebraically equivalent even subalgebras.

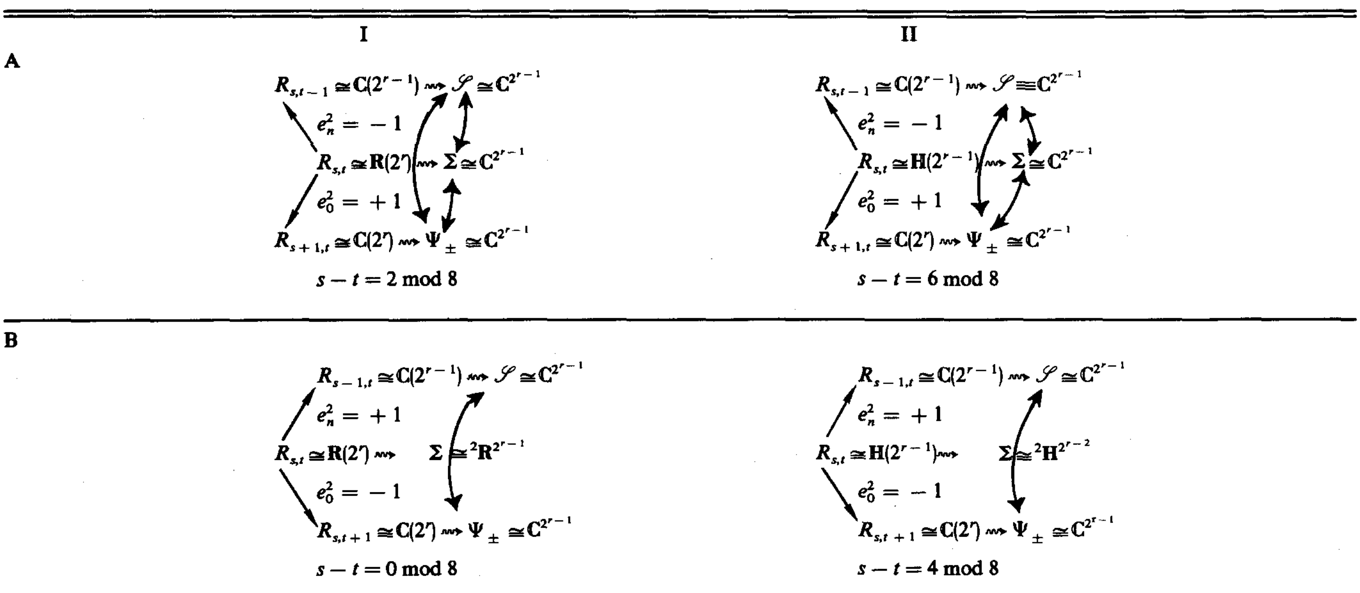
Moreover, the product e_J given by (3.2) plays the role of the pure imaginary unit in these even Clifford algebras. Thus we obtain the vector space isomorphism between the Weyl subspaces $\Psi_\pm$ and even spinor space Σ, respectively. Moreover we always have a vector space isomorphism between the Weyl subspaces and the Dirac space of corresponding lower-dimensional vector space $R^{s-1,t-1}$. Hence for $s-t=2,6 \bmod 8$ we have

$$\Psi_\pm \Leftrightarrow \Sigma(s,t) \Leftrightarrow \Psi(s-1-t-1)=\mathscr{S}_{s,t-1} \tag{3.4}$$

as vector spaces.

The situation will change when we have to enlarge our starting vector space $R^{s,t}$ by a timelike dimension e_0, $e_0^2=-1$ to get $R_{s,t+1}\cong R^{\mathbb{C}}_{s,t}$. Then for both possible algebraic types of universal Clifford algebras, the even spinor spaces are not isomorphic to the Weyl subspaces (see Table I). For these cases the product e_J of the basis elements of corresponding generating vector space $R^{s,t}$ has the property $e_J^2=+1$. It implies that the even Clifford subalgebras, and hence their left modules, have to be over double fields. We obtain that $\Sigma_{s,t}\cong{}^2\mathbb{R}(2^{r-1})$, for $s-t=0 \bmod 8$ and $\Sigma_{s,t}\cong{}^2\mathbb{H}(2^{r-2})$, for $s-t=4 \bmod 8$. Now let us return to the general case of even-dimensional vector space-times. Let γ_μ denote the matrix representation of e_μ in the Dirac algebra $D_{s,t}\cong\mathbb{C}(2^r)$. It is known (by the Noether–Skolem theorem)

TABLE I. Relations between Clifford algebras.



that any two matrix representations of $R_{s,t}$ are related by the similarity transformations. Thus we see immediately that if $R_{s,t}$ is algebraically equivalent to $\mathbb{R}(2^r)$, i.e., if it is of type I, then we can find the representation of $e_\mu \in R_{s',t'}$, $\mu = 1,...,s+t$ by the real matrices γ^0_μ. It implies that any other representation γ_μ can be obtained as

$$\gamma_\mu = \epsilon_0 \mathscr{A}^{-1}\gamma^0_\mu \mathscr{A}, \quad \epsilon_0 = +1. \tag{3.5}$$

Hence

$$\gamma^*_\mu = \mathscr{A}^{-1*}\mathscr{A}\gamma_\mu \mathscr{A}^{-1}\mathscr{A}^*. \tag{3.6}$$

In other words for any matrix representation of the Dirac algebra of type I we have

$$\gamma^*_\mu = B\gamma_\mu B^{-1}, \tag{3.7}$$

with $B = \mathscr{A}^{-1}\mathscr{A}$. It implies that

$$BB^* = 1. \tag{3.8}$$

However, the relation (3.7) means that when we transform the spinor space Ψ by means of an operator C,

$$\psi \rightsquigarrow C\psi := B^{-1}\psi^*, \tag{3.9}$$

then elements e_μ are represented by the same matrices γ_μ.

We shall call C the charge conjugation operator, and $\psi^c = C\psi$ the charge conjugated spinor.

It is easy to see that $C^2 = 1$. The property allows us to construct the projective Majorana operators

$$M_\pm = \tfrac{1}{2}(1 \pm C), \tag{3.10}$$

which decompose the Dirac spinor space into two so-called Majorana subspaces

$$\Psi_R = \tfrac{1}{2}(1+C)\Psi, \qquad \Psi_I = \tfrac{1}{2}(1-C)\Psi. \tag{3.11}$$

Now we see that if a primitive idempotent f, which determines our Dirac spinor space Ψ, is also a primitive one of the Clifford algebra $R_{s,t}$ then $B = \text{id}$ and the Majorana decomposition is given as the decomposition onto real and pure imaginary real subspaces of the Dirac space Ψ. Namely, let

$$\Psi = R_{s',t'} \cdot f. \tag{3.12}$$

Now any primitive idempotent f can be written as

$$f = (1/2^\chi)(1+\omega_1)...(1+\omega_\chi), \tag{3.13}$$

where $\{\omega_i\}$ is a set of square one pairwise commuting elements of $R_{s',t'}$. Then number χ is uniquely determined by the signature (s',t') (see Ref. 8). It can be checked that

$$\chi(s,t) = \chi(s',t'), \tag{3.14}$$

for CA of type I. Hence we can fix a primitive idempotent of corresponding Dirac algebra $D_{s,t} = R_{s',t'}$ as exactly the same as of CA $R_{s,t}$. Thus we obtain immediately the vector space isomorphism between the Dirac spinor space $\Psi(s,t)$ and the complexification $\mathscr{S}^{\mathbb{C}}(s,t)$ of the "real" spinor space $\mathscr{S}(s,t)$, or equivalently we obtain that

$$\Psi_R = \mathscr{S}(s,t). \tag{3.15}$$

However, we should notice that although the algebraic properties of the pure imaginary units in $\Psi(s,t)$ and $\mathscr{S}(s,t)$ are the same, their geometric features are different. They behave differently under the anti-involutions induced by identity and reflection transformation of the starting vector space. This implies different symmetry properties of $\Psi(s,t)$ and $\mathscr{S}(s,t)^{\mathbb{C}}$, respectively.

Now we can ask the question of when Weyl spinors are Majorana ones, i.e., when is the subspace $M_+\Psi$ equal to the subspace $W_+\Psi$? This is the case when

$$C\bar{\gamma} = \bar{\gamma}C, \quad \text{where } \bar{\gamma} = \eta\gamma_1...\gamma_{s+t}, \tag{3.16}$$

but

$$C\bar{\gamma} = \eta^*\gamma_1\bar{\gamma}_{s+t}C = \begin{cases} -\bar{\gamma}C, & \text{for } s-t = 2 \bmod 8, \\ \bar{\gamma}C, & \text{for } s-t = 0 \bmod 8. \end{cases} \tag{3.17}$$

Thus for $s-t = 0 \bmod 8$ Weyl operators are equivalent to their complex conjugates.

Now let us consider the case when the universal Clifford algebra $R_{s,t}$ is realized by the matrix algebra with quaternionic entries, i.e., let us consider the algebraic type II. We can check that for these Clifford algebras we have

$$\chi(s',t') = \chi(s,t) + 1. \tag{3.18}$$

Thus we see that $\Psi(s,t)$ cannot be considered as a vector space isomorphic to the complexification $\mathscr{S}^{\mathbb{C}}(s,t)$ of $\mathscr{S}(s,t)$. The elements $e_\mu \in R^{s,t}$ are represented by the complex matrices $\gamma_\mu \in D_{s,t}$, which cannot be made real. However, γ^*_μ matrices also generate the same Clifford algebra $R_{s,t}$. Again by the Noether–Skolem theorem we have

$$\gamma_\mu = \epsilon\widetilde{B}^{-1}\gamma^*_\mu\widetilde{B}, \tag{3.19}$$

with $\widetilde{B}^*\widetilde{B} = \pm 1$ and $\epsilon = \pm 1$.

Now let $\widetilde{B}^*\widetilde{B} = 1$. Then we can factor $\widetilde{B}$ in the following way:

$$\widetilde{B} = \widetilde{A}^*\widetilde{A}^{-1}. \tag{3.20}$$

This implies that

$$\widetilde{A}^{-1}\gamma_\mu\widetilde{A} = \epsilon\widetilde{A}^{*-1}\gamma^*_\mu\widetilde{A}^*. \tag{3.21}$$

Thus if $\epsilon = +1$, the γ_μ's have their matrix representation given by real matrices, i.e., we have the case of CA of type I. If $\epsilon = -1$, γ_μ's have their matrix representation given by pure imaginary matrices. We have the following lemma.

Lemma 1: Let γ_i be a matrix representation of the universal Clifford algebra $R_{s,t}$ with pure imaginary entries. Then matrices $i\gamma_\mu$ generate the universal CA $R_{t,s}$.

Proof: Obvious.

We see that in the considered case matrices, $i\gamma_\mu$ have to be real. Because for $s-t = 2 \bmod 8$ we have $t-s = 6 \bmod 8$, we obtain that in the case II A we have $\widetilde{B}\widetilde{B}^* = 1$ and $\epsilon = -1$. In other words for $s-t = 6 \bmod 8$ we can fix such a basis of the Dirac space Ψ in which the elements e_μ are represented by pure imaginary matrices γ_μ. In this basis

$$\gamma^*_\mu = -\gamma_\mu \tag{3.22}$$

and the charge conjugation $\psi^c = C\psi$ is given by

$$\psi^c = \psi^*. \tag{3.23}$$

In the general case we have that similar to (3.9)

$$C\psi = \psi^c = \widetilde{B}^{-1}\psi^* \tag{3.24}$$

and Majorana operators

$$M_{\pm} = \tfrac{1}{2}(1 \pm C). \tag{3.25}$$

Example: Let the γ_μ be real 4×4 matrices that generate $R_{3,1} \subset R_{4,1}$. Then pure imaginary matrices $i\gamma_\mu \in R_{2,3}$ generate a subalgebra of $R_{2,3}$ isomorphic to $R_{1,3}$. In other words because we can construct Majorana spinors for the vector space $R^{3,1}$ $(3-1=2)$ this is also possible for vector space-time $R^{1,3}$ $(1-3=6-8)$. Nevertheless in both cases (i.e., $s-t=2,6 \bmod 8$) Majorana spinors $\psi=\psi^c$ cannot be given as Weyl spinors because the necessary conditions (3.16) cannot be satisfied.

Now let us consider the last possibility $s-t=4 \bmod 8$, i.e., the type II B. We see that in this case we are not able to find a representation in which matrices γ_μ generating $R_{s,t}$ are pure imaginary ones. If they were, then the CA $R_{t,s}$ would be isomorphic to the algebra of real matrices. However, this is impossible because also $t-s=4 \bmod 8$. Thus in this case $\widetilde{B}^*\widetilde{B} = -1$ and we cannot factor $\widetilde{B}$ as previously in Eq. (3.20). Of course we can introduce a charge conjugation operator C by (3.24), but it does not allow us to construct the Majorana operators. Although in this case we also can take the decomposition of the Dirac spinor space Ψ onto its real and pure imaginary part, but these subspaces will never be invariant with respect to the Spin (s,t) group. Besides they are not underlying spaces for a faithful representation of endomorphisms generated by γ_μ. As a matter of fact the case of signature $s-t=4 \bmod 8$ can be considered as having the true quaternionic nature. Already we have seen that then both algebraic (s,t) spinors (i.e., left minimal ideals of $R_{s,t}$) and even spinors (i.e., left minimal ideals of even subalgebra $R^{(0)}_{s,t}$) are right modules over the quaternionic ring. Moreover we can see that owing to the property $\widetilde{B}^*\widetilde{B} = -1$, the Dirac spinor space Ψ possesses also a quaternionic structure. Let us take the Weyl decompostion of Ψ onto $\Psi_\pm$ subspaces. We can define an operator $\widetilde{C}$ by

$$\widetilde{C}\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} = \begin{pmatrix}-\widetilde{B}^{-1}\psi_2^*\\ \widetilde{B}^{-1}\psi_1^*\end{pmatrix}, \tag{3.26}$$

where $\psi_1 \in \Psi_+$ and $\psi_2 \in \Psi_-$. Now we have $\widetilde{C}^2 = 1$ and we can look for spinors that satisfy the following generalized Majorana condition

$$\widetilde{C}\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} = \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}. \tag{3.27}$$

We can write (3.27) as

$$\psi_i = \epsilon_i^k \widetilde{B}^{-1}\psi_k^*, \tag{3.27'}$$

with

$$\epsilon_i^k = \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}.$$

Because the relation (3.27') is preserved under the transformation of ψ_i by the group SU(2), we have defined the multiplication of the Dirac space by quaternions [SU(2) is equivalent to the group of unit quaternions].

IV. CONCLUSIONS

In this paper we treat spinors as elements of the universal Clifford algebra $R_{s,t}$ generated by a vector space-time $R^{s,t}$. We have introduced Dirac spinor $\Psi(s,t)$ as "real" spinors of an appropriate extended starting vector space. As a matter of fact we have considered only the cases of even-dimensional vector space-time $s+t=2r$ because for odd-dimensional vector spaces we can do nothing except complexification.

An additional dimension e_0 allows us to construct the product $e_0e_J = e_0e_1 \ldots e_{s+t}$, which plays the role of a pure imaginary unit i. Also, by means of e_0 we construct the Weyl operators $W_\pm = \tfrac{1}{2}(1+\eta e_0)$. Thus for any signature (s,t) we can decompose the Dirac spinor space onto two subspaces of Weyl spinors. For some signature (s,t) we can introduce a charge conjugation operator C; $c^2 = +1$. In these cases we can decompose the Dirac spinor space $\Psi(s,t)$ also onto the so-called Majorana spinor subspaces.

For $s-t=0,2 \bmod 8$, any primitive idempotent of the Clifford algebra $R_{s,t}$ is also a primitive one of the Dirac algebra $D_{s,t} = R_{s',t'}$. Hence we can fix such a base of Dirac spinor space Ψ in which charge conjugation means only complex conjugation, i.e., $\psi^c = \psi^*$. In other words in these cases just $S(s,t)$ forms the Majorana spinor space $\Psi_R = \tfrac{1}{2}(1+C)\Psi$. Moreover in this base the elements $e_\mu \in R^{s,t}$, $\mu = 1,\ldots,s+t$ are represented by real matrices $\gamma_\mu \in D_{s,t}$. Of couse matrices $\{i\gamma_\mu\}$ also belong to the Dirac algebra $D_{s,t}$. However in the case of $s-t=0 \bmod 8$, they generate a subalgebra of $D_{s,t}$, which is isomorphic to the starting Clifford algebra itself. For $s-t=2 \bmod 8$, $\{i\gamma_\mu\}$ generate a subalgebra that is isomorphic to $R_{t,s}$; $t-s=6 \bmod 8$. Hence for $s-t=0 \bmod 8$, Majorana spinors can be taken also as Weyl spinors, whereas for $s-t=2 \bmod 8$ it is impossible. Nevertheless this fact implies that elements $e_\mu \in R^{s,t}$, $s-t=6 \bmod 8$ can be represented by pure imaginary matrices γ_μ in the Dirac algebra $D_{s,t}$. Hence we can introduce a charge conjugation operator C, $C^2=1$, as well as Majorana spinors also for a vector space-time of signature $s-t=6 \bmod 8$. For $s-t=4 \bmod 8$, matrices $\gamma_\mu \in D_{s,t}$ as well as matrices $i\gamma_\mu \in D_{s,t}$ create isomorphic subalgebras of $D_{s,t}$, which are algebraically equivalent to the Clifford algebra $R_{s,t}$. Thus in this case either γ_μ or $i\gamma_\mu$ matrices cannot be given as real matrices. Moreover although γ_μ^* matrices generate also the $R_{s,t}$ subalgebra of $D_{s,t}$, we have in this case $\gamma_\mu^* = \epsilon\widetilde{B}^{-1}\gamma_\mu\widetilde{B}$ with $\widetilde{B}^*\widetilde{B} = -1$. Hence in this case we can not construct an operation of charge conjugation with property $C^2=1$. It means that we cannot fix any subspace of the Dirac space Ψ as the Majorana spinor space. Nevertheless we can use the Weyl decomposition (which always exists) and introduce the so-called SU(2) charge conjugation. This fact together with the quaternionic of the "real" spinor module $\mathscr{S}(s,t)$ and of the "even" spinor module reflects the "true" quaternionic nature of spinors for signature $s-t=4 \bmod 8$.

ACKNOWLEDGMENTS

I wish to thank Professor H. Doebner for his kind hospitality at Clausthal University as well as for his help.

This research was supported by the Humboldt Foundation.

[1]T. Kugo and P. Townsend, "Supersymmetry and the division algebras," CERN preprint TH. 3459-CERN, 1982.

[2]R. Coquereaux, "Modulo 8 periodicity of real Clifford algebras and particle physics," CERN preprint TH. 3279-CERN, 1982.
[3]K. Wetterich, "Massless spinors in more than 4 dimensions," CERN preprint TR 3310-CERN, 1982.
[4]W. Greub, C. R. Math. Rep. Acad. Sci. Canada, I, 293 1979; II, 135 (1980).
[5]I. R. Porteous, *Topological Geometry* (van Nostrand, London, 1969).
[6]P. Lounesto, "Spinor valued regular functions in hypercomplex analysis," 1979 Report No. HTKK-MAT-A 154.
[7]F. Sauter, Z. Phys. **63**, 803 (1930).
[8] $-\chi(s,t)=\chi:=t-H(t-s)$, where H is a Radon–Hurwitz number; see Ref. 5.

Explicit reduction of spl(1,2) to osp(1,2) as a simplified model for applications of supersymmetry to nuclear physics

M. Baake and P. Reinicke
Physikalisches Institut, Universität Bonn, Nussallee 12, D-5300 Bonn 1, West Germany

(Received 28 May 1985; accepted for publication 28 October 1985)

The reduction of irreducible representations of the superalgebra spl(1,2) to the subsuperalgebra osp(1,2) is given explicitly. The oscillator representations are discussed in detail and their relevance for nuclear physics is outlined by a simple model containing three ground-state nuclei. The results have important implications on a recently suggested mechanism for the symmetry breaking of spl(6,2m) via osp(6,2m).

I. INTRODUCTION

A. Foreword

While considering a well-defined problem of nuclear physics we realized the necessity of a simple model in the framework of which all calculations can be carried out explicitly. Such a model is presented in this paper where we separate two aspects, namely the mathematical one and the physical one.

In the mathematical part (Secs. II and III) we describe the reduction of the superalgebra spl(1,2) to its subalgebra osp(1,2) explicitly in terms of generators, irreducible representations, and ordinary oscillator representations. This part can be read clearly without the rest and its scope is, of course, more general than needed for the physical application.

In the physical part we discuss the implications for nuclear physics. Here the reader always should keep in mind that our simple model has been constructed in order to facilitate a well-founded understanding of the analogous, but much more complex, structure in realistic models.

B. Physical motivation

Since its introduction by Arima and Iachello[1] in the mid-1970's the interacting boson approximation (IBA) for the collective excitations of even–even nuclei has been well established from the theoretical as well as from the experimental point of view.[2,3]

The generalization to odd–even nuclei consequently lead to the so-called interacting boson fermion approximation[4] (IBFA) and, in a further step, to boson fermion supersymmetries[5–7] (SUSY).

In the last case, one starts with the supergroup U(6/2m), where the 6 stems from the bosonic degrees of freedom (one s-boson and one d-boson) and 2m reflects the fermionic degrees of freedom. For realistic applications, 2m is at least of the order of 10 which obviously leads to a huge number of possible group chains for the desired dynamic symmetry breaking. Since presently a microscopic theory of nuclear SUSY is missing, the relevance of the several chains can be tested only phenomenologically.

In what follows, we use the terminology of Lie superalgebras, and write spl(n,m) for the algebra that belongs to U(n/m).

Apart from the standard route, breaking SUSY in the first step by going to the maximal Lie subalgebra, one can keep SUSY for one or more further steps by first going to a Lie subsuperalgebra.

For several reasons—two of them being the importance of the o(6) limit in the IBA and of the concept of seniority in both even–even and odd–even nuclei—it looks very promising to introduce the Lie superalgebra osp(6,2m) with its subalgebra o(6) × sp(2m), i.e., to consider the chain

$$\mathrm{spl}(6,2m) \supset \mathrm{osp}(6,2m) \supset \mathrm{o}(6)\times\mathrm{sp}(2m) \supset \cdots . \qquad (1.1)$$

This has been suggested recently by Morrison and Jarvis.[8]

As is well known,[9–11] the rep theory of Lie superalgebras shows some pathological properties, unknown from the theory of ordinary Lie algebras, which may cause problems for physical applications. Especially for the chain (1.1), the Hamiltonian turns out to be non-Hermitian and the reduction spl(6,2m)↓osp(6,2m) produces a mixture of different boson and fermion numbers equivalent to a mixture of different nuclei.

In order to make the whole structure transparent, one needs a simple model, in the framework of which one can calculate each step explicitly. Such a model is presented in this paper while a discussion of (1.1) and other realistic chains is deferred to a forthcoming publication.

The necessity to simplify the physical structure is obvious since reps of, let us say, spl(6,12) are too large to be dealt with explicitly. Omitting the bosonic degrees of freedom normally attached to the d-boson, i.e., keeping only the s-boson, one arrives at spl(1,2m) with 2m fermionic degrees of freedom. In this paper, we will consider the case $m = 1$, i.e., spl(1,2), since the pathologic properties already arise in this simple example. To see this one has to consider the two chains

$$\begin{aligned}&\text{(i)}\quad \mathrm{spl}(1,2) \supset \mathrm{osp}(1,2) \supset \mathrm{sl}(2),\\ &\text{(ii)}\quad \mathrm{spl}(1,2) \supset \mathrm{gl}(1)\times\mathrm{sl}(2).\end{aligned} \qquad (1.2)$$

In this model, a nucleus is defined as the set of all irreducible sl(2) multiplets belonging to the same numbers N_B and N_F of bosons and fermions, respectively. Since we have solely an s-boson we can only describe ground states, i.e., a nucleus is a single sl(2) multiplet.

 0022-2488/86/030859-10$02.50

Clearly, from what has been said above, we need an oscillator rep of spl(1,2) containing one boson s and two fermions a and b, the reduction of which then yields the whole structure of our simple model.

C. Survey of contents

After this long discussion of motivation we can now describe how the paper is organized and give some of the main results.

Sections II and III are devoted to the mathematical structure. As the explicit reduction of spl(1,2) reps to osp(1,2) is an interesting problem in itself and a direct pursuit of Ref. 9, we do this job first in Sec. II.

Then, in Sec. III, we present the oscillator reps of spl(1,2) and its decomposition. Although for the physical application only the rep with one boson and two fermions (Sec. III B) is needed, we shortly present, for the sake of completeness, the oscillator rep built of one fermion and two bosons (Sec. III A). In both cases the reduction to osp(1,2) is given. Through this reduction, a non-Hermiticity occurs in one case for the quadratic Casimir operator of osp(1,2), which turns out to be not even a normal operator with all the consequences like state mixing and change of metric. This is discussed in detail in Secs. III C and III D.

Last but not least, in Sec. IV, we consider the physical implications of the preceding sections. First, we discuss the direct consequences of the non-Hermiticity that arose in Sec. III C. Then, in Sec. IV B, we consider the relevance of dynamic symmetry breaking via osp(1,2) by inverse analysis. There we obtain the remarkable result that with the ansatz of a Hermitian sl(2)-invariant Hamiltonian—from which one surely would have started if one had not known anything about SUSY—it is impossible to establish a true osp(1,2) supersymmetry.

Normally, a simplified ansatz for the Hamiltonian consists of a sum of linear and quadratic Casimir operators that is motivated—but not justified—by a special property of IBA, namely that this ansatz yields the most general Hamiltonian quartic in the creation and annihilation operators, except for one term that is the product of two linear Casimir operators.[4,12] For this type of ansatz, dynamic symmetry then simply means that one only has to take the Casimir operators of a suitably chosen chain of subalgebras.

In general, one cannot expect the same property and, indeed, in our simple model already several terms exist that are linearly independent of the Casimir operators. Perhaps the most striking example is a transition operator between two nuclei belonging to the same supermultiplet.

II. REDUCTION OF spl(1,2) TO osp(1,2)

To give an explicit reduction of the irreps of spl(1,2) to those of osp(1,2) we will use the results of Scheunert *et al.*[9] The even generators of spl(1,2) are Q_m ($m = 1,2,3$) and B, the odd ones are $V_\pm$ and $W_\pm$. The commutation relations of the algebra read

$$\begin{aligned}
&[Q_3,Q_\pm] = \pm Q_\pm, \quad [Q_+,Q_-] = 2Q_3,\\
&[B,Q_\pm] = [B,Q_3] = 0, \quad [B,V_\pm] = \tfrac{1}{2}V_\pm,\\
&[B,W_\pm] = -\tfrac{1}{2}W_\pm, \quad [Q_\pm,V_\pm] = [Q_\pm,W_\pm] = 0,\\
&[Q_3,V_\pm] = \pm\tfrac{1}{2}V_\pm, \quad [Q_3,W_\pm] = \pm\tfrac{1}{2}W_\pm,\\
&[Q_\pm,V_\mp] = V_\pm, \quad [Q_\pm,W_\mp] = W_\pm,\\
&\{V_\pm,V_\pm\} = \{V_\pm,V_\mp\} = \{W_\pm,W_\pm\}\\
&\qquad = \{W_\pm,W_\mp\} = 0,\\
&\{V_\pm,W_\pm\} = \pm Q_\pm, \quad \{V_\pm,W_\mp\} = -Q_3 \pm B,
\end{aligned} \tag{2.1}$$

where

$$Q_\pm = Q_1 \pm iQ_2.$$

The generators of the osp(1,2) algebra are Q_m ($m = 1,2,3$) and $V_{\pm 1/2}$, satisfying the commutation relations

$$\begin{aligned}
&[Q_3,Q_\pm] = \pm Q_\pm, \quad [Q_+,Q_-] = 2Q_3,\\
&[Q_\pm,V_{\pm 1/2}] = 0, \quad [Q_3,V_{\pm 1/2}] = \pm\tfrac{1}{2}V_{\pm 1/2},\\
&[Q_\pm,V_{\mp 1/2}] = V_{\pm 1/2},\\
&\{V_{\pm 1/2},V_{\pm 1/2}\} = \pm\tfrac{1}{2}Q_\pm, \quad \{V_{\pm 1/2},V_{\mp 1/2}\} = -\tfrac{1}{2}Q_3.
\end{aligned} \tag{2.2}$$

We will use the notation of Ref. 9 with the following modifications: In order to distinguish between the $(\pm q,q)$ rep and the $(b,q)|_{b=\pm q}$ rep of spl(1,2) we will label the first one by $[\pm q,q]$ and the second one by $(\pm q,q)$. We label the odd generators of osp(1,2) by $V_{\pm 1/2}$ instead of $V_\pm$ and we will choose $V^\ddagger_{\pm 1/2} = \mp V_{\mp 1/2}$ instead of $V^\ddagger_{\pm 1/2} = \pm V_{\mp 1/2}$, where $\ddagger$ denotes the grade adjoint operation.

The irreps of the algebra osp(1,2) are labeled by a half-integer q. For $q \geqslant \frac{1}{2}$, such a rep contains two sl(2) multiplets with isospin q and $q - \frac{1}{2}$, respectively (for the notation, cf. Ref. 9 or the Appendix). The dimension therefore is $4q + 1$. The generators of the osp(1,2) algebra in the q rep can be written in matrix notation,

$$\begin{aligned}
Q_k &= \left(\begin{array}{c|c} D^{(q)}(Q_k) & 0 \\ \hline 0 & D^{(q-1/2)}(Q_k) \end{array}\right),\\
V_{\pm 1/2} &= \frac{1}{2}\left(\begin{array}{c|c} 0 & A^{(q)}_\pm \\ \hline \pm {}^tA^{(q)}_\mp & 0 \end{array}\right),
\end{aligned} \tag{2.3}$$

where

$$\begin{aligned}
A^{(q)}_+ &= \left(\begin{array}{ccccc} \sqrt{2q} & & & & \\ & \sqrt{2q-1} & & 0 & \\ & & \ddots & & \\ & 0 & & \sqrt{2} & \\ & & & & \sqrt{1} \\ \hline 0 & & \cdots & & 0 \end{array}\right),\\
A^{(q)}_- &= \left(\begin{array}{ccccc} 0 & & \cdots & & 0 \\ \hline \sqrt{1} & & & & \\ & \sqrt{2} & & 0 & \\ & & \ddots & & \\ & 0 & & \sqrt{2q-1} & \\ & & & & \sqrt{2q} \end{array}\right),
\end{aligned} \tag{2.4}$$

and $D^{(q)}(Q_k)$ is the usual matrix rep of sl(2) of dimension $2q+1$.

The irreps of the algebra spl(1,2) have been systematically constructed.[9] The different reps as well as the different possibilities (one has to normalize the states relative to each other) are characterized by constants $\alpha, \beta, \gamma, \delta, \epsilon, \zeta, \tau$, and ω. In order really to define a rep these constants have to fulfill a system of nonlinear equations. For the sake of completeness, we list the most important relations in the Appendix.

To come now to an explicit reduction of spl(1,2) reps to osp(1,2) reps we first consider the $[\pm q,q]$ reps of spl(1,2). Here and in what follows we exclude the case $q=0$ since this gives the trivial representation.

Before we can directly compare the rep matrices we have to localize the subalgebra osp(1,2) in terms of spl(1,2) generators. Since the sl(2) part, given by the generators Q_k, $1<k<3$, is the same for both algebras, we only have to express $V_{\pm 1/2}$ by the odd generators of spl(1,2). One finds

$$V_{\pm 1/2}=\tfrac{1}{2}(V_\pm + W_\pm), \tag{2.5}$$

which can directly be verified by the commutation relations.

Choosing $\alpha=1$ ($\beta=1$) for the $[q,q]$ ($[-q,q]$) representation, one obtains

$$Q_k=\begin{pmatrix} D^{(q)}(Q_k) & 0 \\ 0 & D^{(q-1/2)}(Q_k)\end{pmatrix},$$
$$\frac{1}{2}(V_\pm + W_\pm)=\frac{1}{2}\begin{pmatrix} 0 & A_\pm^{(q)} \\ \pm {}^tA_\mp^{(q)} & 0\end{pmatrix}, \tag{2.6}$$

and a comparison with (2.3) immediately gives the following theorem.

Theorem 1:

$$[\pm q,q]_{\rm spl}\downarrow_{\rm osp}\simeq(q)_{\rm osp}, \tag{2.7}$$

i.e., the $[\pm q,q]$ reps of spl(1,2) stay irred on osp(1,2).

Now consider the (b,q) representation of spl(1,2). In what follows, $b\neq \pm q$ is assumed and b is called the baryon number according to Ref. 9. We shall prove the following theorem.

Theorem 2:

$$(b,q)_{\rm spl}\downarrow_{\rm osp}\simeq(q)_{\rm osp}+(q-\tfrac{1}{2})_{\rm osp}. \tag{2.8}$$

Notice that we are not writing the direct sum. To show the equivalence we will have to use a nonunitary matrix, which results in a change of the metric.

Proof: Consider first the case $q=\frac{1}{2}$. This representation contains only three multiplets, namely $|b,q,q_3\rangle$, $|b+\frac{1}{2},q-\frac{1}{2},q_3\rangle$, and $|b-\frac{1}{2},q-\frac{1}{2},q_3\rangle$ (hence $\delta=\zeta=\tau=\omega=0$). In this case one has

$$Q_k=\begin{pmatrix} D^{(1/2)}(Q_k) & & 0 \\ & D^{(0)}(Q_k) & \\ 0 & & D^{(0)}(Q_k)\end{pmatrix}. \tag{2.9}$$

In general we order the blocks with decreasing isospin q and if the isospin is the same with decreasing baryon number b. Notice that in the $(\frac{1}{2})+(0)$ rep of osp(1,2) Q_k reads as (2.9). Hence, in order to show (2.8), we have to find a matrix M that commutes with Q_k and satisfies

$$M\tfrac{1}{2}(V_\pm + W_\pm)M^{-1}=V_{\pm 1/2}. \tag{2.10}$$

From $[M,Q_k]=0$ we obtain

$$M=\begin{pmatrix} \mathbf{1}_2 & 0 & \\ 0 & \lambda_2 & \mu_1 \\ & \mu_2 & \lambda_3\end{pmatrix},\quad \mathbf{1}_2=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \tag{2.11}$$

where an overall constant has already been fixed.

Now, since $b\neq\pm q$ we can choose the two independent constants α and β as follows:

$$\alpha=\sqrt{b+\tfrac{1}{2}},\quad \beta=\sqrt{-b+\tfrac{1}{2}}, \tag{2.12}$$

so that

$$\tfrac{1}{2}(V_\pm + W_\pm)=\frac{1}{2}\begin{pmatrix} 0 & \alpha\cdot A_\pm^{(1/2)} & \beta\cdot A_\pm^{(1/2)} \\ \pm\alpha\cdot {}^tA_\mp^{(1/2)} & 0 & 0 \\ \pm\beta\cdot {}^tA_\mp^{(1/2)} & 0 & 0\end{pmatrix}. \tag{2.13}$$

Choosing

$$M=\begin{pmatrix} \mathbf{1}_2 & 0 & \\ 0 & \alpha & \beta \\ & \beta & -\alpha\end{pmatrix}, \tag{2.14}$$

we obtain

$$M\tfrac{1}{2}(V_\pm + W_\pm)M^{-1}=\frac{1}{2}\begin{pmatrix} 0 & A_\pm^{(1/2)} & 0 \\ \pm {}^tA_\mp^{(1/2)} & 0 & 0 \\ 0 & 0 & 0\end{pmatrix}=V_{\pm 1/2}, \tag{2.15}$$

where $V_{\pm 1/2}$ is given in the $(\frac{1}{2})+(0)$ rep of osp(1,2). Note that there are other possible choices of M that satisfy (2.10), but that it is not possible to choose M unitary.

Let now $q\geq 1$. In the (b,q) rep of spl(1,2) and in the $(q)+(q-\frac{1}{2})$ rep of osp(1,2) Q_k reads

$$Q_k=\begin{pmatrix} D^{(q)}(Q_k) & & & 0 \\ & D^{(q-1/2)}(Q_k) & & \\ & & D^{(q-1/2)}(Q_k) & \\ 0 & & & D^{(q-1)}(Q_k)\end{pmatrix}. \tag{2.16}$$

From Schur's lemma the most general nonsingular matrix M (up to an overall factor) that commutes with Q_k is

$$M=\begin{pmatrix} \lambda_1\cdot \mathbf{1}_{2q+1} & 0 & 0 & 0 \\ 0 & \lambda_2\cdot \mathbf{1}_{2q} & \mu_1\cdot \mathbf{1}_{2q} & 0 \\ 0 & \mu_2\cdot \mathbf{1}_{2q} & \lambda_3\cdot \mathbf{1}_{2q} & 0 \\ 0 & 0 & 0 & \mathbf{1}_{2q-1} \end{pmatrix}. \tag{2.17}$$

Since $b \neq \pm q$ the constants α, β, and δ are nonzero, so that without loss of generality we choose

$$\alpha=\sqrt{(q+b)/2q},\quad \beta=\delta=\sqrt{(q-b)/2q}. \tag{2.18}$$

Thus, in the (b,q) rep we have

$$\frac{1}{2}(V_\pm + W_\pm)=\frac{1}{2}\begin{pmatrix} 0 & \alpha A_\pm^{(q)} & \beta A_\pm^{(q)} & 0 \\ \pm\alpha^t A_\mp^{(q)} & 0 & 0 & \beta A_\pm^{(q-1/2)} \\ \pm\beta^t A_\mp^{(q)} & 0 & 0 & -\alpha A_\pm^{(q-1/2)} \\ 0 & \pm\beta^t A_\mp^{(q-1/2)} & \mp\alpha^t A_\mp^{(q-1/2)} & 0 \end{pmatrix}. \tag{2.19}$$

Now choosing $\lambda_2=\alpha\lambda_1$, $\lambda_3=-\alpha$, $\mu_1=\lambda_1\beta$, and $\mu_2=\beta$ one obtains

$$M\tfrac{1}{2}(V_\pm + W_\pm)M^{-1}=\frac{1}{2}\left(\begin{array}{cc|cc} 0 & A_\pm^{(q)} & & \\ \pm{}^tA_\mp^{(q)} & 0 & \multicolumn{2}{c}{0} \\ \hline & & 0 & A_\pm^{(q-1/2)} \\ \multicolumn{2}{c|}{0} & \pm{}^tA_\mp^{(q-1/2)} & 0 \end{array}\right)=V_{\pm 1/2}, \tag{2.20}$$

where $V_{\pm 1/2}$ is given in the $(q)+(q-\frac{1}{2})$ rep of osp(1,2). This completes the proof of (2.8).

If $b=\pm q$, $q>\frac{1}{2}$, one gets, for each sign, two inequivalent not completely reducible representations. After restriction to osp(1,2) one obtains for each case separately (2.8). Here again it is not possible to choose the corresponding matrix M unitary.

III. OSCILLATOR REPRESENTATIONS

Let us now proceed to an explicit construction of oscillator reps of spl(1,2) and osp(1,2), which, in principle, have been discussed by Palev[13,14] in a more general context without regarding the connection between spl(1,2) and osp(1,2).

A. One fermion and two bosons

In this section we shortly present the infinite-dimensional representation S containing one fermion a and two bosons r and s, with

$$\{a,a^+\}=[r,r^+]=[s,s^+]=1,$$
$$[r,s]=[r^+,s]=[r,a]=[s,a]=[r^+,a]=[s^+,a]=0. \tag{3.1}$$

Concentrating first on spl(1,2) we obtain

$$V^+=a^+r,\quad V_-=-a^+s,\quad W_+=as^+,\quad W_-=ar^+,$$
$$Q_+=s^+r,\quad Q_-=r^+s,\quad Q_3=\tfrac{1}{2}(s^+s-r^+r), \tag{3.2}$$
$$B=\widehat{N}_F+\tfrac{1}{2}\widehat{N}_B,$$

where $\widehat{N}_F=a^+a$ and $\widehat{N}_B=r^+r+s^+s$. Since $V_+^+=W_-$ and $V_-^+=-W_+$, this defines a star representation of spl(1,2).

It is easy to calculate

$$\vec{Q}^2=(\widehat{N}_B/2)((\widehat{N}_B/2)+1),\quad K_2^{spl}=\widehat{N}_F(1-\widehat{N}_F)\equiv 0. \tag{3.3}$$

Hence, the spin of a single sl(2) multiplet contained in the oscillator representation is given by $q=\frac{1}{2}\cdot N_B$, $N_B\in \mathbb{N}_0$. For every value of N_B we can have $N_F=0$ or $N_F=1$. Now, an irreducible spl(1,2) constituent consists of one sl(2) multiplet with $q=\frac{1}{2}\cdot N_B$, $N_F=0$ (hence $b=q$) and another one with $q=\frac{1}{2}\cdot(N_B-1)$, $N_F=1$ (hence $b=q+\frac{1}{2}$), where $N_B=0$ gives the trivial representation.

Consequently, the infinite-dimensional oscillator representation S contains exactly all *nontypical* representations of type $[q,q]$, i.e.,

$$S=\bigoplus_{2q=0}^{\infty}[q,q]. \tag{3.4}$$

For example, taking $q=1$, one gets

$$|1;1,1\rangle=(1/\sqrt{2})(s^+)^2|0\rangle,\quad |1;1,0\rangle=r^+s^+|0\rangle,$$
$$|1;1,-1\rangle=(1/\sqrt{2})(r^+)^2|0\rangle, \tag{3.5}$$
$$|\tfrac{1}{2};\tfrac{1}{2},\tfrac{1}{2}\rangle=a^+s^+|0\rangle,\quad |\tfrac{1}{2};\tfrac{1}{2},-\tfrac{1}{2}\rangle=a^+r^+|0\rangle,$$

where the vacuum $|0\rangle$ is defined by

$$a|0\rangle=r|0\rangle=s|0\rangle=0,\quad \langle 0|0\rangle=1. \tag{3.6}$$

Simultaneously, the vacuum spans the rep space of the trivial [0,0] constituent.

Going now to the subalgebra osp(1,2) by $V_{\pm 1/2}:=\frac{1}{2}(V_\pm + W_\pm)$, we obtain

$$V_{+1/2}=\tfrac{1}{2}(a^+r+as^+),\quad V_{-1/2}=\tfrac{1}{2}(-a^+s+ar^+),$$
$$Q_+=s^+r,\quad Q_-=r^+s,\quad Q_3=\tfrac{1}{2}(s^+s-r^+r). \tag{3.7}$$

Consequently,

$$[V_{+1/2},V_{-1/2}]=\tfrac{1}{4}(2\hat{N}_F-\hat{N}_B+2\hat{N}_F\hat{N}_B),$$
$$K_2^{osp}=(\hat{N}_F+\hat{N}_B/2)(1/2+\hat{N}_B/2). \tag{3.8}$$

By comparison to spl(1,2) one can see that K_2^{osp} takes the eigenvalues $q(q+\frac{1}{2})$ with $q=\frac{1}{2}\cdot N_B$, since $N_F+\frac{1}{2}\cdot N_B=b=q$. This, of course, agrees with

$$[q,q]_{spl}\downarrow_{osp}\simeq(q)_{osp}. \tag{3.9}$$

In this case no splitting occurs, which means that no dynamic symmetry breaking can be created by osp(1,2). Therefore we turn to the next case.

B. One boson and two fermions

We now construct the oscillator representation T containing one boson s and two fermions a and b, with

$$[s,s^+]=\{a,a^+\}=\{b,b^+\}=1,$$
$$[s,a]=[s,b]=[s^+,a]=[s^+,b]=0, \tag{3.10}$$
$$\{a,a\}=\{b,b\}=\{a,b\}=\{a,b^+\}=0.$$

Then we have for the generators of spl(1,2),

$$V_+=s^+a,\quad V_-=-s^+b,\quad W_+=sb^+,\quad W_-=sa^+,$$
$$Q_+=b^+a,\quad Q_-=a^+b,\quad Q_3=\tfrac{1}{2}(b^+b-a^+a), \tag{3.11}$$
$$B=\hat{N}_B+\tfrac{1}{2}\hat{N}_F,$$

where

$$\hat{N}_B=s^+s\quad\text{and}\quad\hat{N}_F=a^+a+b^+b. \tag{3.12}$$

As in the previous case, these relations define a star representation with $V_+^\dagger=W_-$ and $V_-^\dagger=-W_+$. Calculating the quadratic Casimir operators of spl(1,2) and its sl(2) constituent one gets

$$K_2^{spl}=\hat{N}-\hat{N}^2,\quad \hat{N}=\hat{N}_B+\hat{N}_F,$$
$$K_2^{sl}=\vec{Q}^2=\tfrac{3}{4}\cdot\hat{N}_F(2-\hat{N}_F). \tag{3.13}$$

The possible values of N_F are 0, 1, and 2. For $N_B\in\mathbb{N}_0$, $\vec{Q}^2$ takes the values 0 and $\frac{3}{4}$, and B the values N_B, $N_B+\frac{1}{2}$, and N_B+1. Thus, for $N_B\geqslant 1$, the irreducible constituents of this oscillator representation are star representations of type $(N_B+\frac{1}{2},\frac{1}{2})$, all being four dimensional. The case $N_B=0$ gives the sum of two reps, namely $[0,0]\oplus[\frac{1}{2},\frac{1}{2}]$, which will be treated separately. Altogether we have

$$T=[0,0]\oplus[\tfrac{1}{2},\tfrac{1}{2}]\oplus\bigoplus_{n=1}^{\infty}(n+\tfrac{1}{2},\tfrac{1}{2}). \tag{3.14}$$

With the vacuum $|0\rangle$, defined by the relations $a|0\rangle=b|0\rangle=s|0\rangle=0$ and $\langle 0|0\rangle=1$, one can construct an explicit basis of the $(N_B+\frac{1}{2},\frac{1}{2})$ constituent for $N_B=n\geqslant 1$, namely

$$|n+\tfrac{1}{2};\tfrac{1}{2},\tfrac{1}{2}\rangle=(1/\sqrt{n!})b^+(s^+)^n|0\rangle,$$
$$|n+\tfrac{1}{2};\tfrac{1}{2},-\tfrac{1}{2}\rangle=(1/\sqrt{n!})a^+(s^+)^n|0\rangle,$$
$$|n+1;0,0\rangle=(1/\sqrt{(n+1)!})\cdot(s^+)^{n+1}|0\rangle, \tag{3.15}$$
$$|n;0,0\rangle=-(1/\sqrt{(n-1)!})a^+b^+(s^+)^{n-1}|0\rangle.$$

This corresponds to the following choice of parameters, which causes the representation matrices of the generators to be real:

$$\alpha=\gamma=\sqrt{n+1},\quad\beta=-\epsilon=\sqrt{n},\quad\delta=\zeta=\tau=\omega=0. \tag{3.16}$$

The matrices become

$$Q_+=\left(\begin{array}{cc|c}0&1&\\0&0&0\\\hline &0&0\end{array}\right),\quad Q_-=\left(\begin{array}{cc|c}0&0&\\1&0&0\\\hline &0&0\end{array}\right),$$
$$Q_3=\left(\begin{array}{cc|c}\frac{1}{2}&0&\\0&-\frac{1}{2}&0\\\hline &0&0\end{array}\right), \tag{3.17}$$

$$B=\left(\begin{array}{cc|cc}n+\frac{1}{2}&0&&\\0&n+\frac{1}{2}&\multicolumn{2}{c}{0}\\\hline &&n+1&0\\\multicolumn{2}{c|}{0}&0&n\end{array}\right), \tag{3.18}$$

$$V_+=\left(\begin{array}{cc|cc}&&0&-\sqrt{n}\\\multicolumn{2}{c|}{0}&0&0\\\hline 0&\sqrt{n+1}&&\\0&0&\multicolumn{2}{c}{0}\end{array}\right),$$
$$V_-=\left(\begin{array}{cc|cc}&&0&0\\\multicolumn{2}{c|}{0}&0&-\sqrt{n}\\\hline -\sqrt{n+1}&0&&\\0&0&\multicolumn{2}{c}{0}\end{array}\right), \tag{3.19}$$

$$W_+=\left(\begin{array}{cc|cc}&&\sqrt{n+1}&0\\\multicolumn{2}{c|}{0}&0&0\\\hline 0&0&&\\0&\sqrt{n}&\multicolumn{2}{c}{0}\end{array}\right),$$
$$W_-=\left(\begin{array}{cc|cc}&&0&0\\\multicolumn{2}{c|}{0}&\sqrt{n+1}&0\\\hline 0&0&&\\-\sqrt{n}&0&\multicolumn{2}{c}{0}\end{array}\right). \tag{3.20}$$

For the case $N_B=0$ we obtain

$$|\tfrac{1}{2};\tfrac{1}{2},\tfrac{1}{2}\rangle=b^+|0\rangle,\quad|\tfrac{1}{2};\tfrac{1}{2},-\tfrac{1}{2}\rangle=a^+|0\rangle,$$
$$|1;0,0\rangle=s^+|0\rangle,\quad|0;0,0\rangle=|0\rangle. \tag{3.21}$$

The matrices read as above but with $n=0$. Since now $\beta=\epsilon=0$, we have a four-dimensional, fully reducible rep that is block diagonal from the beginning, namely $[0,0]\oplus[\frac{1}{2},\frac{1}{2}]$.

Hence, Eq. (3.14) is obvious. If $\mathscr{V}$ denotes the infinite-dimensional rep space of T and $\mathscr{V}_n:=\text{span}(|n+\frac{1}{2};\frac{1}{2},\pm\frac{1}{2}\rangle, |n+1;0,0\rangle,|n;0,0\rangle)$, we can also write

$$\mathscr{V}=\bigoplus_{n=0}^{\infty}\mathscr{V}_n, \tag{3.22}$$

where only $\mathscr{V}_0$ is further decomposable with respect to T.

C. Reduction of T to osp(1,2)

We will now investigate the reduction of T to the subalgebra osp(1,2). Clearly, from Sec. II, the net result will be

$$T|_{\rm osp} \simeq \bigoplus_{n=0}^{\infty} ((q=0)_{\rm osp} + (q=\tfrac{1}{2})_{\rm osp}), \tag{3.23}$$

where the problems concerned with the reduction will be discussed explicitly. For $N_B = n \geqslant 1$ we reduce the representation $(n+\frac{1}{2},\frac{1}{2})$ in detail. The generators of osp(1,2) are, in this context,

$$Q_+ = b^+a, \quad Q_- = a^+b, \quad Q_3 = \tfrac{1}{2}(b^+b - a^+a), \tag{3.24}$$
$$V_{+1/2} = \tfrac{1}{2}(s^+a + sb^+), \quad V_{-1/2} = \tfrac{1}{2}(-s^+b + sa^+).$$

In the spl(1,2) basis $|n+\frac{1}{2};\frac{1}{2},\frac{1}{2}\rangle$, $|n+\frac{1}{2};\frac{1}{2},-\frac{1}{2}\rangle$, $|n+1;0,0\rangle$, $|n;0,0\rangle$, the last two operators have the matrices

$$V_{+1/2} = \frac{1}{2}\left(\begin{array}{cc|cc} \multicolumn{2}{c|}{0} & \sqrt{n+1} & -\sqrt{n} \\ & & 0 & 0 \\ \hline 0 & \sqrt{n+1} & \multicolumn{2}{c}{0} \\ 0 & \sqrt{n} & & \end{array}\right), \tag{3.25}$$

$$V_{-1/2} = \frac{1}{2}\left(\begin{array}{c|ccc} 0 & & 0 & 0 \\ & & \sqrt{n+1} & -\sqrt{n} \\ \hline -\sqrt{n+1} & 0 & & \\ -\sqrt{n} & 0 & \multicolumn{2}{c}{0} \end{array}\right).$$

In order to block-diagonalize them without changing the matrices of Q_+, Q_-, and Q_3 we will use the transformation matrix M,

$$M = \left(\begin{array}{c|cc} \mathbb{1}_2 & \multicolumn{2}{c}{0} \\ \hline 0 & \sqrt{n+1} & -\sqrt{n} \\ & -\sqrt{n} & \sqrt{n+1} \end{array}\right), \tag{3.26}$$

$$M^{-1} = \left(\begin{array}{c|cc} \mathbb{1}_2 & \multicolumn{2}{c}{0} \\ \hline 0 & \sqrt{n+1} & \sqrt{n} \\ & \sqrt{n} & \sqrt{n+1} \end{array}\right), \quad \det(M) = 1.$$

This gives $\tilde{V}_{\pm 1/2} = MV_{\pm 1/2}M^{-1}$ with

$$\tilde{V}_{+1/2} = \frac{1}{2}\left(\begin{array}{ccc|c} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right), \tag{3.27}$$

$$\tilde{V}_{-1/2} = \frac{1}{2}\left(\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right).$$

Clearly, the transformation defines a new basis on the representation space $\mathscr{V}_n$, namely

$$\begin{aligned} &|(n)\tfrac{1}{2};\tfrac{1}{2},\tfrac{1}{2}\rangle = (1/\sqrt{n!})b^+(s^+)^n|0\rangle, \quad |(n)\tfrac{1}{2};\tfrac{1}{2},-\tfrac{1}{2}\rangle = (1/\sqrt{n!})a^+(s^+)^n|0\rangle, \\ &|(n)\tfrac{1}{2};0,0\rangle = \left[\frac{1}{\sqrt{n!}}(s^+)^{n+1} - \frac{\sqrt{n}}{\sqrt{(n-1)!}}a^+b^+(s^+)^{n-1}\right]|0\rangle, \\ &|(n)0;0,0\rangle = \left[\frac{\sqrt{n}}{\sqrt{(n+1)!}}(s^+)^{n+1} - \frac{\sqrt{n+1}}{\sqrt{(n-1)!}}a^+b^+(s^+)^{n-1}\right]|0\rangle. \end{aligned} \tag{3.28}$$

The matrix M is not unitary and, furthermore, it is impossible to block-diagonalize $V_{+1/2}$ and $V_{-1/2}$ with a unitary matrix. Hence, the Hermitian inner product must be destroyed by this transformation. Indeed one obtains the dual basis

$$\begin{aligned} &((n)\tfrac{1}{2};\tfrac{1}{2},\tfrac{1}{2}| = \langle 0|s^n\cdot b\cdot(1/\sqrt{n!}), \quad ((n)\tfrac{1}{2};\tfrac{1}{2},-\tfrac{1}{2}| = \langle 0|s^n\cdot a\cdot(1/\sqrt{n!}), \\ &((n)\tfrac{1}{2};0,0| = \langle 0|\left[s^{n+1}\cdot\frac{1}{\sqrt{n!}} + s^{n-1}\cdot ba\cdot\frac{\sqrt{n}}{\sqrt{(n-1)!}}\right], \\ &((n)0;0,0| = \langle 0|\left[-s^{n+1}\cdot\frac{\sqrt{n}}{\sqrt{(n+1)!}} - s^{n-1}\cdot ba\cdot\frac{\sqrt{n+1}}{\sqrt{(n-1)!}}\right], \end{aligned} \tag{3.29}$$

with the result that

$$((n)\tfrac{1}{2};0,0| \neq |(n)\tfrac{1}{2};0,0)^+, \quad ((n)0;0,0| \neq |(n)0;0,0)^+. \tag{3.30}$$

Hence, we no longer have the canonical dual basis. This forces us to distinguish between the bilinear form $(\cdot,\cdot)$ defined on $\mathscr{V}_n^* \times \mathscr{V}_n$, where $\mathscr{V}_n^*$ is the dual space, and the inner product of $\mathscr{V}_n$. The latter results from the old bilinear form $\langle\cdot,\cdot\rangle$ by the mapping

$$\iota: \mathscr{V}_n \to \mathscr{V}_n^*,$$
$$|\varphi\rangle \mapsto \langle\varphi| = \iota|\varphi\rangle := \sum_{j=1}^{4}\langle j|\varphi\rangle^*\langle j| \tag{3.31}$$

via

$$\langle\varphi|\psi\rangle := \langle\iota|\varphi\rangle,\psi\rangle. \tag{3.32}$$

Here, $\langle j|$, $1 \leqslant j \leqslant 4$, denotes the canonical dual basis to the Fock space basis constructed above.

For the new bilinear form $(\cdot,\cdot)$, the map ι induces

$$\begin{aligned} &\iota(c_1|(n)\tfrac{1}{2};\tfrac{1}{2},\tfrac{1}{2}) + c_2|(n)\tfrac{1}{2};\tfrac{1}{2},-\tfrac{1}{2}) \\ &\quad + c_3|(n)\tfrac{1}{2};0,0) + c_4|(n)0;0,0)) \\ &\quad = ((n)\tfrac{1}{2};\tfrac{1}{2},\tfrac{1}{2}|c_1^* + ((n)\tfrac{1}{2};\tfrac{1}{2},-\tfrac{1}{2}|c_2^* \\ &\qquad + ((n)\tfrac{1}{2};0,0|[(2n+1)c_3^* + 2\sqrt{n(n+1)}c_4^*] \\ &\qquad + ((n)0;0,0|[2\sqrt{n(n+1)}c_3^* + (2n+1)c_4^*]. \end{aligned} \tag{3.33}$$

As a consequence, we get, for the new basis vectors,

$$\| \,|(n)\tfrac{1}{2};0,0)\|^2 = \| \,|(n)0;0,0)\|^2 = 2n+1 \qquad (3.34)$$

and

$$\langle \iota |(n)\tfrac{1}{2};0,0),|(n)0;0,0)\rangle = 2\sqrt{n(n+1)}. \qquad (3.35)$$

This means that the reduction of the spl(1,2) representation to the subalgebra osp(1,2) results in a basis that is neither normalized nor even orthogonal in the old metric (3.32), which cannot be changed for physical reasons.

In the previously discussed oscillator rep S with one fermion and two bosons (Sec. III A) we avoid the problem since only nontypical spl reps of type $[q,q]$ occur that stay irred after restriction to osp(1,2).

D. Quadratic Casimir operator of osp(1,2) and its non-Hermiticity

To get a deeper insight into the problems mentioned above, let us calculate the quadratic Casimir operator $K_2^{\rm osp}$ in the oscillator rep T. We obtain

$$K_2^{\rm osp} = \Sigma + \overline{\Sigma}, \qquad (3.36)$$

where

$$\Sigma = (\hat{N}_{\rm F} + \tfrac{1}{2}\hat{N}_{\rm B}) + \tfrac{1}{4}\hat{N}_{\rm F}(1-\hat{N}_{\rm F}) - \tfrac{1}{2}\hat{N}_{\rm F}(\hat{N}_{\rm B}+\hat{N}_{\rm F}),$$
$$\overline{\Sigma} = -\tfrac{1}{2}(s^+as^+b + a^+sb^+s), \qquad (3.37)$$
$$\Sigma^+ = \Sigma, \quad \overline{\Sigma}^+ = -\overline{\Sigma}.$$

in the *old* spl(1,2) basis, defined by Eq. (3.15), one has, for the several irreducible constituents, the block matrices

$$K_2^{\rm osp} = \frac{1}{2}\begin{pmatrix} \mathbf{1}_2 & 0 & \\ 0 & n+1 & -\sqrt{n(n+1)} \\ & \sqrt{n(n+1)} & -n \end{pmatrix}, \qquad (3.38)$$

while, after the similarity transformation by M, one has

$$K_2^{\rm osp} = \begin{pmatrix} \tfrac{1}{2} & 0 & & \\ 0 & \tfrac{1}{2} & & 0 \\ & 0 & \tfrac{1}{2} & 0 \\ & & 0 & 0 \end{pmatrix}. \qquad (3.39)$$

Clearly, this means that the *new* basis vectors $|(n)\tfrac{1}{2};\tfrac{1}{2},\tfrac{1}{2}),\ldots,$ $|(n)0;0,0)$ form an eigenbasis of $K_2^{\rm osp}$ to *real* eigenvalues. This is only possible, if $[\Sigma,\overline{\Sigma}]\neq 0$. Indeed,

$$[\Sigma,\overline{\Sigma}] = \tfrac{1}{4}(s^+as^+b - a^+sb^+s)$$
$$+\tfrac{1}{2}(s^+as^+b\cdot\hat{N}_{\rm B} - \hat{N}_{\rm B}\cdot a^+sb^+s), \qquad (3.40)$$

and, consequently,

$$[K_2^{\rm osp},(K_2^{\rm osp})^+] = 2[\overline{\Sigma},\Sigma]\neq 0. \qquad (3.41)$$

But the last relation shows that $K_2^{\rm osp}$ in this oscillator rep is not even a normal operator, which explains the appearance of nonorthogonal eigenstates.

IV. PHYSICAL IMPLICATIONS

In this section we try to translate the mathematical structure of our simplified model to the language of nuclear physics.

A. Consequences of the reduction spl(1,2)↓osp(1,2)

The rep theory of graded Lie algebras possesses some properties unknown from ordinary Lie algebras. Concepts like Hermitian reps, complete reduciblity, and the Schur lemma have to be generalized.[10] Consequently, one should expect some problems for application to physics; we deal with one in our simple model.

In the oscillator rep T of Sec. III B the quadratic Casimir operator of osp(1,2) reads

$$K_2^{\rm osp} = \Sigma + \overline{\Sigma}, \qquad (4.1)$$

with

$$\Sigma^+ = \Sigma, \quad \overline{\Sigma}^+ = -\overline{\Sigma}, \quad \text{and } [\Sigma,\overline{\Sigma}]\neq 0.$$

Hence, $K_2^{\rm osp}$ is neither a Hermitian nor even a normal operator. If $K_2^{\rm osp}$ is now part of a Hamiltonian, as is suggested by the usual scheme of dynamic symmetry breaking, we get in general a mixture of states correlated with a change of the metric. This means that the new state vectors are neither normalized nor even orthogonal with respect to the original metric (cf. Sec. III C).

In the context of our model the reduction spl(1,2)↓osp(1,2) causes a mixture of states that correspond to nuclei with *different* nucleon numbers since the definition of what we call nucleus has been given in terms of spl(1,2) state vectors. Recalling Eq. (3.15) from Sec. III B, we obviously have

$$N_{\rm B} = n+1, \quad N_{\rm F} = 0, \quad \text{for the state } |n+1;0,0\rangle,$$
$$N_{\rm B} = n-1, \quad N_{\rm F} = 2, \quad \text{for the state } |n;0,0\rangle, \qquad (4.2)$$
$$N_{\rm B} = n, \quad N_{\rm F} = 1, \quad \text{for the states } |n+\tfrac{1}{2};\tfrac{1}{2},\pm\tfrac{1}{2}\rangle.$$

Hence, the total number of quasiparticles is $N = N_{\rm B} + N_{\rm F} = n+1$, while the numbers of real particles are $2n+2$, $2n$, and $2n+1$, respectively, defining ground-state nuclei with spin-0 or -$\tfrac{1}{2}$. As the reduction mixes the nuclei $|n+1;0,0\rangle$ and $|n;0,0\rangle$, one cannot hope to find any meaningful expressions for transition elements or selection rules.

Let us add a further note to the reduction discussed in Sec. III C. As the reader may have noticed, no problem is present in the case $N_{\rm B}=0$ [Eq. (3.21)] since then the osp(1,2) part is block diagonal from the beginning. But the states then contain at most either a fermion or a boson. As, additionally, one of the states is the vacuum, this case is irrelevant for nuclear physics. That is why we only investigated the case $N_{\rm B}\geqslant 1$.

B. An inverse analysis for dynamic symmetry breaking

We will now discuss the relevance of the dynamic symmetry breaking in the context of our simple model with one boson s and two fermions a,b. To this end, we will proceed backwards, i.e., we will start with the most general sl(2)-invariant Hamiltonian H that is compatible with the underlying physics. Then H should be a function of the creation and annihilation operators, i.e.,

$$H = H(s,s^+,a,a^+,b,b^+), \qquad (4.3)$$

and should consist only of a sum of terms that are bilinear or quartic in the operators, where we disregard an overall constant in H. Furthermore, we demand that in every term the number of creation operators equals the number of annihilation operators. Then, the most general expression for H turns out to be

$$H = \alpha \cdot \widehat{N}_{\mathrm{F}} + \beta \cdot \widehat{N}_{\mathrm{B}} + \gamma \cdot \widehat{N}_{\mathrm{F}}^2 + \delta \cdot \widehat{N}_{\mathrm{B}}^2 + \epsilon \cdot \widehat{N}_{\mathrm{F}} \cdot \widehat{N}_{\mathrm{B}} + \zeta \cdot s^+ a s^+ b + \eta \cdot a^+ s b^+ s. \tag{4.4}$$

The non-Hermiticity problem occurring with the last two terms will be discussed later on. If we now rearrange the Hamiltonian for terms that are spl(1,2) and osp(1,2) invariant we obtain

$$\begin{aligned} H = {} & \tilde{\alpha} \cdot \widehat{N} + \tilde{\beta} \cdot K_2^{\mathrm{spl}} + \tilde{\gamma} \cdot K_2^{\mathrm{osp}} + \tilde{\delta} \cdot K_2^{\mathrm{sl}} \\ & + \tilde{\epsilon} \cdot (\widehat{N}_{\mathrm{B}} - \widehat{N}_{\mathrm{F}}) + \tilde{\zeta} \cdot (N_{\mathrm{B}} - N_{\mathrm{F}})^2 \\ & + \tilde{\eta}(s^+ a s^+ b - a^+ s b^+ s), \end{aligned} \tag{4.5}$$

with

$$\begin{aligned} K_2^{\mathrm{spl}} &= \widehat{N} - \widehat{N}^2, \\ K_2^{\mathrm{osp}} &= \tfrac{5}{4}\widehat{N}_{\mathrm{F}} + \tfrac{1}{2}\widehat{N}_{\mathrm{B}} - \tfrac{3}{4}\widehat{N}_{\mathrm{F}}^2 - \tfrac{1}{2}\widehat{N}_{\mathrm{F}}\widehat{N}_{\mathrm{B}} \\ &\quad - \tfrac{1}{2}(s^+ a s^+ b + a^+ s b^+ s), \\ K_2^{\mathrm{sl}} &= \tfrac{3}{4}\widehat{N}_{\mathrm{F}}(2 - \widehat{N}_{\mathrm{F}}). \end{aligned} \tag{4.6}$$

The coefficients are related via

$$\begin{pmatrix} \alpha \\ \vdots \\ \eta \end{pmatrix} = {}^t\widetilde{M} \begin{pmatrix} \tilde{\alpha} \\ \vdots \\ \tilde{\eta} \end{pmatrix},$$

$$\widetilde{M} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & -2 & 0 & 0 \\ \frac{5}{4} & \frac{1}{2} & -\frac{3}{4} & 0 & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{3}{2} & 0 & -\frac{3}{4} & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix},$$

$$\det \widetilde{M} = -6 \neq 0. \tag{4.7}$$

Let us pause for two remarks. First, we have two linearly independent terms that are spl(1,2) invariant, namely K_2^{spl} and $\widehat{N}$. Second, by the appearance of seven linearly independent sl(2) invariants, we truly cannot write the Hamiltonian as a pure sum of Casimir operators without suppressing some terms. The most striking one is

$$\Omega := s^+ a s^+ b - a^+ s b^+ s, \tag{4.8}$$

which is a consequence of the imposed combination of a boson with two fermions. Furthermore, Ω can serve as transition operator, which will be discussed below.

The only non-Hermitian expression in Eq. (4.5) is now K_2^{osp}, which can be written as (cf. Sec. III D)

$$K_2^{\mathrm{osp}} = \Sigma + \overline{\Sigma},$$

with

$$\overline{\Sigma} = -\tfrac{1}{2}(s^+ a s^+ b + a^+ s b^+ s), \quad \Sigma^+ = \Sigma, \quad \overline{\Sigma}^+ = -\overline{\Sigma}. \tag{4.9}$$

If we arrange the Hamiltonian to be Hermitian from the beginning, i.e., if we take $\eta = -\zeta$, it will be impossible to establish a true osp(1,2) symmetry! This is because we will get

$$\begin{aligned} H = {} & \alpha' \cdot \widehat{N} + \beta' \cdot K_2^{\mathrm{spl}} + \gamma' \cdot K_2^{\mathrm{sl}} + \delta' \cdot (\widehat{N}_{\mathrm{B}} + \tfrac{1}{2}\widehat{N}_{\mathrm{F}}) \\ & + \epsilon' \cdot (\widehat{N}_{\mathrm{B}} - \widehat{N}_{\mathrm{F}})^2 + \zeta' \cdot (s^+ a s^+ b - a^+ s b^+ s), \end{aligned} \tag{4.10}$$

where the primed coefficients are related to the unprimed ones by

$$\begin{pmatrix} \alpha \\ \vdots \\ \zeta \end{pmatrix} = {}^t M' \begin{pmatrix} \alpha' \\ \vdots \\ \zeta' \end{pmatrix},$$

$$M' = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & -2 & 0 \\ \frac{3}{2} & 0 & -\frac{3}{4} & 0 & 0 & 0 \\ \frac{1}{2} & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\det M' = \tfrac{3}{2} \neq 0. \tag{4.11}$$

One has to conclude that the appearance of an osp(1,2)-invariant term that is not automatically spl(1,2)-invariant is equivalent to the appearance of a term proportional to $(s^+ a s^+ b + a^+ s b^+ s)$ in the original Hamiltonian.

Clearly, the two possibilities to rearrange the Hamiltonian shown above are correlated to the chains $\mathrm{spl}(1,2) \supset \mathrm{osp}(1,2) \supset \mathrm{sl}(2)$ and $\mathrm{spl}(1,2) \supset \mathrm{gl}(1) \times \mathrm{sl}(2)$, respectively. The first one produces a non-Hermitian Hamiltonian with all its problems discussed above while the second one avoids the whole trouble by throwing out the non-Hermitian contribution.

Furthermore, for the second chain, the operator Ω defined by Eq. (4.8) can be interpreted as a transition operator. In the spl(1,2) basis we find

$$\Omega = \left(\begin{array}{c|cc} 0 & \multicolumn{2}{c}{0} \\ \hline 0 & 0 & \sqrt{n(n+1)} \\ & \sqrt{n(n+1)} & 0 \end{array} \right), \tag{4.12}$$

which means that Ω describes an interaction only between the states $|n+1;0,0\rangle$ and $|n;0,0\rangle$, i.e., replacing two bosons by two fermions or vice versa. In terms of selection rules this reads as

$$\Delta b = 0, \pm 1, \quad \Delta q = 0, \quad \Delta q_3 = 0. \tag{4.13}$$

This may serve as a verification of the impossibility of writing the Hamiltonian as a sum of Casimir operators, since this at least means dropping the operator Ω, which is a direct consequence of supersymmetry.

V. CONCLUSION

In the mathematical part of this paper, the reduction of the simple Lie superalgebra spl(1,2) to its subalgebra osp(1,2) has been presented. The decomposition of representations was outlined explicitly in terms of similarity transformations, whereby the necessity emerged to change the metric.

The same program was carried out with the ordinary oscillator representations after separating out their irreducible constituents. In one case, the quadratic Casimir operator of osp(1,2) was not normal. This reflects the fact that the osp superalgebras do not have star reps but at most grade star reps.[10,11]

In the physical part of this paper, the reduction spl(1,2)↓osp(1,2) served as a simplified model for the investigation of dynamic supersymmetry breaking in nuclear physics. We found that the chain spl(1,2) ⊃ osp(1,2) ⊃ sl(2) produces a Hamiltonian that is neither Hermitian nor even normal. Its diagonalization causes mixtures of different nuclei and, simultaneously, a change of the metric. Hence, selection rules and expressions for transition elements will have no satisfactory physical interpretation. For these reasons, we claim this breaking mechanism to be unphysical.

One can avoid all these problems by taking the chain spl(1,2) ⊃ gl(1)×sl(2), which, of course, breaks supersymmetry already in the first step.

Although the restriction of our model is obvious, the consequences are quite far reaching. The whole structure, like non-Hermiticity, state mixing, and so on, is also present in a realistic model involving the chain spl(6,2m) ⊃ osp(6,2m) ⊃ ..., which was recently proposed by Morrison and Jarvis.[8] A more detailed discussion of this model as well as an alternative approach without the problems described above will be given in a forthcoming publication.

ACKNOWLEDGMENT

We wish to thank A. Gelberg and V. Rittenberg for helpful discussions and for reading the manuscript.

APPENDIX: IRREDUCIBLE REPRESENTATIONS OF spl(1,2)

In what follows we briefly present the main results on the classification of spl(1,2) taken from Scheunert *et al.*[9]

The states of an irrep are labeled by the baryon number b and the isospin q with its projection q_3, corresponding to the operators B, $\vec{Q}^2$, and Q_3, respectively. For convenience, this terminology has been taken over from high energy physics (cf. Ref. 9) although the numbers b and q may have a different meaning in the context of nuclear physics.

In a single irrep we have at most the states

$$|b;q,q_3\rangle, \quad |b+\tfrac{1}{2};q-\tfrac{1}{2},q_3\rangle, \qquad |b-\tfrac{1}{2};q-\tfrac{1}{2},q_3\rangle, \quad |b;q-1,q_3\rangle. \tag{A1}$$

The generators act on these states as follows:

$$Q_3|b;q,q_3\rangle = q_3|b;q,q_3\rangle,$$
$$Q_\pm|b;q,q_3\rangle = \sqrt{(q\mp q_3)(q\pm q_3+1)}\,|b;q,q_3\pm 1\rangle, \tag{A2}$$

$$V_\pm|b;q,q_3\rangle = \pm\alpha\sqrt{q\mp q_3}\,|b+\tfrac{1}{2};q-\tfrac{1}{2},q_3\pm\tfrac{1}{2}\rangle,$$
$$W_\pm|b;q,q_3\rangle = \pm\beta\sqrt{q\mp q_3}\,|b-\tfrac{1}{2};q-\tfrac{1}{2},q_3\pm\tfrac{1}{2}\rangle, \tag{A3}$$

$$V_\pm|b+\tfrac{1}{2};q-\tfrac{1}{2},q_3\rangle = 0,$$
$$W_\pm|b+\tfrac{1}{2};q-\tfrac{1}{2},q_3\rangle = \gamma\sqrt{q\pm q_3+\tfrac{1}{2}}\,|b;q,q_3\pm\tfrac{1}{2}\rangle \pm\delta\sqrt{q\mp q_3-\tfrac{1}{2}}\,|b;q-1,q_3\pm\tfrac{1}{2}\rangle, \tag{A4}$$

$$V_\pm|b-\tfrac{1}{2};q-\tfrac{1}{2},q_3\rangle = \epsilon\sqrt{q\pm q_3+\tfrac{1}{2}}\,|b;q,q_3\pm\tfrac{1}{2}\rangle \pm\zeta\sqrt{q\mp q_3-\tfrac{1}{2}}\,|b;q-1,q_3\pm\tfrac{1}{2}\rangle,$$
$$W_\pm|b-\tfrac{1}{2};q-\tfrac{1}{2},q_3\rangle = 0, \tag{A5}$$

$$V_\pm|b;q-1,q_3\rangle = \tau\sqrt{q\pm q_3}\,|b+\tfrac{1}{2};q-\tfrac{1}{2},q_3\pm\tfrac{1}{2}\rangle,$$
$$W_\pm|b;q-1,q_3\rangle = \omega\sqrt{q\pm q_3}\,|b-\tfrac{1}{2};q-\tfrac{1}{2},q_3\pm\tfrac{1}{2}\rangle, \tag{A6}$$

where the constants $\alpha,\beta,...,\tau,\omega$ are independent of q_3, but may depend on q and b.

We can now classify the irreps of spl(1,2).

(a) $q=0$ gives the trivial rep of spl(1,2).

(b) Setting $q\geqslant\tfrac{1}{2}$, $\beta=\delta=\epsilon=\zeta=\tau=\omega=0$, $b=q$, $\alpha\cdot\gamma=1$, where $\alpha\neq 1$ is arbitrary, yields the rep of type $[q,q]$ that contains the multiplets $|b;q,q_3\rangle$ and $|b+\tfrac{1}{2};q-\tfrac{1}{2},q_3\rangle$ but not the multiplets $|b-\tfrac{1}{2};q-\tfrac{1}{2},q_3\rangle$ and $|b;q-1,q_3\rangle$ The dimension is $4q+1$.

(c) Setting $q\geqslant\tfrac{1}{2}$, $\alpha=\gamma=\delta=\zeta=\tau=\omega=0$, $b=-q$, $\beta\cdot\epsilon=1$, where $\beta\neq 0$ is arbitrary, yields the rep of type $[-q,q]$ that contains the multiplets $|b;q,q_3\rangle$ and $|b-\tfrac{1}{2};q-\tfrac{1}{2},q_3\rangle$ but not the multiplets $|b+\tfrac{1}{2};q-\tfrac{1}{2},q_3\rangle$ and $|b;q-1,q_3\rangle$. The dimension again is $4q+1$.

(d) Setting $q=\tfrac{1}{2}$, $\delta=\zeta=\tau=\omega=0$, $\alpha\cdot\gamma=\tfrac{1}{2}+b$, $\beta\cdot\epsilon=\tfrac{1}{2}-b$, defines the four-dimensional rep of type $(b,\tfrac{1}{2})$, where only the multiplet $|b;q-1,q_3\rangle$ is missing. Solutions with nonzero α and β are equivalent. For $b=\mp\tfrac{1}{2}$ we will exclude the solutions with vanishing α or β, since they are irrelevant for our purpose.

(e) If an irrep ($q\geqslant 1$) contains all four multiplets, the constants $\alpha,\beta,...,\tau,\omega$ must solve the following nonlinear system of equations:

$$\begin{aligned}
&\alpha\epsilon+\zeta\tau=0, \quad \beta\gamma+\delta\omega=0,\\
&\alpha\gamma+\beta\epsilon=1, \quad \beta\epsilon+\zeta\omega=1,\\
&\alpha\gamma+\delta\tau=1, \quad \delta\tau+\zeta\omega=1,\\
&\alpha\delta+\beta\zeta=0, \quad \gamma\tau+\epsilon\omega=0,\\
&\alpha\gamma-\beta\epsilon=b/q, \quad \delta\tau-\zeta\omega=-b/q,\\
&\alpha\gamma(q+\tfrac{1}{2})-\delta\tau(q-\tfrac{1}{2})=b+\tfrac{1}{2},\\
&-\beta\epsilon(q+\tfrac{1}{2})+\zeta\omega(q-\tfrac{1}{2})=b-\tfrac{1}{2}.
\end{aligned} \tag{A7}$$

For $b\neq\pm q$ the solution of these equations may be given in terms of the constants α, β, and δ. The only possible solutions require these constants to be nonzero. Representations with different (nonvanishing) values for α,β, and δ are equivalent. If $b=\pm q$, there are some additional solutions, where some of the free constants vanish. Since for our purpose these additional solutions are irrelevant, we will not go into details.

[1]A. Arima and F. Iachello, Ann. Rev. Nucl. Part. Sci. **31**, 75 (1981) and references therein.
[2]*Interacting Bosons in Nuclear Physics*, edited by F. Iachello (Plenum, New York, 1979).
[3]R. F. Casten and P. v. Brentano, Phys. Lett. **152**, 22 (1985).
[4]*Interacting Bose–Fermi Systems in Nuclei*, edited by F. Iachello (Plenum, New York, 1981).
[5]A. B. Balantekin, I. Bars, and F. Iachello, Nucl. Phys. A **370**, 284 (1981).
[6]F. Iachello, Nucl. Phys. A **374**, 636c (1982).
[7]P. van Isacker, A. Frank, and H-Z. Sun, Ann. Phys. NY **157**, 183 (1984).
[8]I. Morrison and P. D. Jarvis, Nucl. Phys. A **435**, 461 (1985).
[9]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977).
[10]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146 (1977).
[11]M. Scheunert, "The theory of Lie superalgebras," in *Lecture Notes in Mathematics*, Vol. 716 (Springer, Berlin, 1979).
[12]P. Castaños, E. Chacón, A. Frank, and M. Moshinsky, J. Math. Phys. **20**, 35 (1979).
[13]T. D. Palev, J. Math. Phys. **21**, 1293 (1980).
[14]T. D. Palev, J. Math. Phys. **23**, 1100 (1982).

Coherent states of the real symplectic group in a complex analytic parametrization. II. Annihilation-operator coherent states

C. Quesne[a)]

Physique Théorique et Mathématique CP 229, Université Libre de Bruxelles, Boulevard du Triomphe, B 1050 Brussels, Belgium

(Received 27 June 1985; accepted for publication 31 October 1985)

In the present series of papers, the coherent states of $\mathrm{Sp}(2d,R)$, corresponding to the positive discrete series irreducible representations $\langle\lambda_d + n/2,...,\lambda_1 + n/2\rangle$, encountered in physical applications, are analyzed in detail with special emphasis on those of $\mathrm{Sp}(4,R)$ and $\mathrm{Sp}(6,R)$. The present paper discusses the annihilation-operator coherent states, i.e., the eigenstates of the noncompact lowering generators corresponding to complex eigenvalues. These states generalize the coherent states introduced by Barut and Girardello for $\mathrm{Sp}(2,R)$, and later on extended by Deenen and Quesne to the $\mathrm{Sp}(2d,R)$ irreducible representations of the type $\langle(\lambda + n/2)^d\rangle$. When $\lambda_1,...,\lambda_d$ are not all equal, it was shown by Deenen and Quesne that the eigenvalues do not completely specify the eigenstates of the noncompact lowering generators. In the present work, their characterization is completed by a set of continuous labels parametrizing the (unitary-operator) coherent states of the maximal compact subgroup $\mathrm{U}(d)$. The resulting coherent states are therefore of mixed type, being annihilation-operator coherent states only as regards the noncompact generators. A realization in a subspace of a Bargmann space of analytic functions shows that such coherent states satisfy a unity resolution relation in the representation space of $\langle\lambda_d + n/2,...,\lambda_1 + n/2\rangle$, and therefore may be used as a continuous basis in such space. The analytic functions and the differential operators representing the representation space discrete bases and the $\mathrm{Sp}(2d,R)$ generators, respectively, are found in explicit form. It is concluded that the annihilation-operator coherent state representation provides the mathematical foundation for the use of differentiation operators with respect to the noncompact raising generators in symbolic expressions of the $\mathrm{Sp}(2d,R)$ generators. This is to be compared with the habit of replacing a boson annihilation operator by a symbolic differentiation with respect to the corresponding creation operator, accounted for by the Bargmann representation of such operators.

I. INTRODUCTION

The purpose of the present series of papers is to study the generalized coherent states (CS) of the real symplectic group $\mathrm{Sp}(2d,R)$, corresponding to the positive discrete series irreducible representations (irreps) $\langle\lambda\rangle \equiv \langle\lambda_d + n/2,...,\lambda_1 + n/2\rangle$ (see Refs. 1–3). Special emphasis is laid on the CS of $\mathrm{Sp}(4,R)$ and $\mathrm{Sp}(6,R)$, amenable to some interesting applications in physical problems.

The first paper in this series[4] (henceforth referred to as I and whose equations will be subsequently quoted by their number preceded by I) was devoted to the unitary-operator CS, as defined by Klauder,[5] Perelomov,[6] and Gilmore.[7] As is well known, such CS exist in one-to-one correspondence with the points of the coset space $\mathrm{Sp}(2d,R)/H$, where H is the stability group of the irrep $\langle\lambda\rangle$ lowest weight state $|(\lambda)_{\min}\rangle$, chosen as reference state. It was shown in I that a convenient parametrization of the CS is provided by a complex symmetric $d\times d$ matrix $\mathbf{u}$, subject to the condition $\mathbf{I} - \mathbf{u}^*\mathbf{u} > 0$, and by a set of parameters $\mathbf{z}$ characterizing the CS of the maximal compact subgroup $\mathrm{U}(d)$, corresponding to the irrep $[\lambda] \equiv [\lambda_1 + n/2,...,\lambda_d + n/2]$. Such a parametrization is based upon a factorization of the coset space $\mathrm{Sp}(2d,R)/H$ into the product of coset spaces $\mathrm{Sp}(2d,R)/\mathrm{U}(d)$ and $\mathrm{U}(d)/H$.

The present paper deals with a generalization to $\mathrm{Sp}(2d,R)$ of the $\mathrm{Sp}(2,R)$ CS introduced by Barut and Girardello.[8] The latter are defined as the eigenstates $|w)$ of the noncompact lowering generator $D = D_{11}$,

$$D|w) = w^*|w), \tag{1.1}$$

corresponding to a complex eigenvalue w^*. They can be termed annihilation-operator CS because the lowering generator D annihilates the lowest weight state of the $\mathrm{Sp}(2,R)$ irrep $\langle\lambda\rangle$, and therefore plays the same role with respect to that reference state as the oscillator annihilation operator with respect to the oscillator ground state.

When we consider $\mathrm{Sp}(2d,R)$ instead of $\mathrm{Sp}(2,R)$, as shown in Eq. (I 2.5a), the lowest weight state $|(\lambda)_{\min}\rangle$ is annihilated by the set of noncompact lowering generators $D_{ij} = D_{ji}$, $i,j = 1,...,d$. Since the latter commute with one another, we may search for their common eigenstates $|\mathbf{w})$,

$$D_{ij}|\mathbf{w}) = w_{ij}^*|\mathbf{w}), \quad i,j = 1,...,d, \tag{1.2}$$

corresponding to some complex eigenvalues $w_{ij}^* = w_{ji}^*$. Here, $\mathbf{w}$ denotes the complex symmetric $d\times d$ matrix whose elements are w_{ij}. In the case where $\lambda_1 = \cdots = \lambda_d = \lambda$, Deenen and Quesne[9,10] proved that for any complex symmetric matrix $\mathbf{w}$, Eq. (1.2) does have a uniquely defined solution, and they exhibited the latter in explicit form for $\mathrm{Sp}(4,R)$ and $\mathrm{Sp}(6,R)$.

In contrast, when all the λ_i's are not equal, Deenen and Quesne[11] demonstrated that for any complex symmetric ma-

[a] Maître de recherches F.N.R.S.

trix $\mathbf{w}$, Eq. (1.2) has more than one independent solution (actually the number of independent solutions is equal to the dimension Λ of the $U(d)$ irrep $[\lambda]$), hence the CS are not completely specified by $\mathbf{w}$. The missing labels cannot be defined as the eigenvalues of some extra lowering generators, since no such operator commutes with the whole set of generators D_{ij}. This means that a straightforward generalization of the Barut–Girardello Sp(2,R) CS does not exist for Sp($2d,R$) whenever $\lambda_1,...,\lambda_d$ are not all equal. To face this difficulty, Deenen and Quesne[11] used as missing labels the Gel'fand patterns (λ) characterizing the rows of the irrep[12–14] $[\lambda]$, thereby introducing partially coherent states (PCS) $|\mathbf{w};(\lambda))$, specified by the continuous parameters $\mathbf{w}$ and the discrete labels (λ).

The present paper proposes an alternative to this procedure, wherein instead of the discrete labels (λ), continuous parameters $\mathbf{z}$ are used to completely specify the eigenstates of D_{ij}, $i,j=1,...,d$, which are therefore fully CS in contrast to the PCS of Ref. 11. As in I, the parameters $\mathbf{z}$ characterize the $U(d)$ CS corresponding to the irrep $[\lambda]$.

The CS $|\mathbf{w},\mathbf{z})$ are defined in Sec. II, and some of their properties are reviewed in Sec. III. In Sec. IV, the $U(d)$ CS representation is realized in a subspace of a Bargmann space of analytic functions,[15] wherein the complex parameters $\mathbf{z}$ are given a well-defined meaning. In Sec. V, a similar procedure is applied to both the parameters $\mathbf{w}$ and $\mathbf{z}$, thereby showing that the CS $|\mathbf{w},\mathbf{z})$ satisfy a unity resolution relation in a subspace of a Bargmann space. Finally, Sec. VI contains some concluding remarks.

Before proceeding, a few words about notations are in order. In I, we used an angular bracket for the unitary-operator CS $|\mathbf{u},\mathbf{z}\rangle$, and a caret above the symbols denoting the corresponding quantities, such as the measure $d\hat{\sigma}(\mathbf{u},\mathbf{z})$, the functional representation $\hat{\psi}(\mathbf{u},\mathbf{z})$, etc. In the present paper, a round bracket is used as the notation for the annihilation-operator CS $|\mathbf{w},\mathbf{z})$. The measure $d\sigma(\mathbf{w},\mathbf{z})$, the functional representation $\psi(\mathbf{w},\mathbf{z})$, etc., corresponding to such states are denoted by the same symbols as in I, but without the caret. For the quantities not associated with CS, such as boson operators, generators, etc., the definitions and notations of I are used without change.

II. DEFINITION OF THE ANNIHILATION-OPERATOR COHERENT STATES

In the case where $\lambda_1=\cdots=\lambda_d=\lambda$, (see Refs. 9 and 10) it has proved convenient to rewrite Eq. (1.2) in the unitary-operator CS representation, corresponding to the states $|\mathbf{u}\rangle$ defined in Eq. (I 3.3),

$$\langle\mathbf{u}|D_{ij}|\mathbf{w})=w_{ij}^*\langle\mathbf{u}|\mathbf{w}),\quad i,j=1,...,d. \tag{2.1}$$

Since the corresponding representation of D_{ij} is the partial differential operator

$$\hat{\mathscr{D}}_{ij}=\Delta_{u_{ij}}=(1+\delta_{ij})\frac{\partial}{\partial u_{ij}}, \tag{2.2}$$

Eq. (2.1) is indeed equivalent to the following system of first-order partial differential equations for the overlap $\langle\mathbf{u}|\mathbf{w})$,

$$\Delta_{u_{ij}}\langle\mathbf{u}|\mathbf{w})=w_{ij}^*\langle\mathbf{u}|\mathbf{w}),\quad i,j=1,...,d. \tag{2.3}$$

Its solution is given by

$$\langle\mathbf{u}|\mathbf{w})=G(\mathbf{w}^*)\exp(\tfrac{1}{2}\operatorname{tr}\mathbf{u}\mathbf{w}^*), \tag{2.4}$$

where $G(\mathbf{w}^*)$ is an arbitrary function of $\mathbf{w}^*$. If we fix the normalization of $|\mathbf{w})$ in such a way that

$$\langle(\lambda)_{\min}|\mathbf{w})=1, \tag{2.5}$$

then

$$G(\mathbf{w}^*)=1, \tag{2.6}$$

and Eq. (2.4) becomes

$$\langle\mathbf{u}|\mathbf{w})=\exp(\tfrac{1}{2}\operatorname{tr}\mathbf{u}\mathbf{w}^*). \tag{2.7}$$

Equation (2.7) uniquely determines the states $|\mathbf{w})$, which can be expanded as follows:

$$|\mathbf{w})=\int d\hat{\sigma}(\mathbf{u})\exp(\tfrac{1}{2}\operatorname{tr}\mathbf{u}\mathbf{w}^*)|\mathbf{u}\rangle, \tag{2.8}$$

where $d\hat{\sigma}(\mathbf{u})$ is the measure corresponding to the CS $|\mathbf{u}\rangle$ given in Eqs. (I 6.2), (I 6.10), and (I 6.13). Therefore, for any complex symmetric matrix $\mathbf{w}$, Eq. (1.2) has a unique solution $|\mathbf{w})$, subject to the normalization condition (2.5).

For those cases where the λ_i's are not all equal, the above procedure converts Eq. (1.2) into the following relation:

$$\langle\mathbf{u},\mathbf{z}'|D_{ij}|\mathbf{w})=w_{ij}^*\langle\mathbf{u},\mathbf{z}'|\mathbf{w}),\quad i,j=1,...,d, \tag{2.9}$$

where $|\mathbf{u},\mathbf{z}'\rangle$ is now the unitary-operator CS defined in Eq. (I 3.6). Since from Eq. (I 5.3a), Eq. (2.2) remains valid, we obtain the system of partial differential equations

$$\Delta_{u_{ij}}\langle\mathbf{u},\mathbf{z}'|\mathbf{w})=w_{ij}^*\langle\mathbf{u},\mathbf{z}'|\mathbf{w}),\quad i,j=1,...,d, \tag{2.10}$$

which only differs from Eq. (2.3) by an extra $\mathbf{z}'$ dependence. The solution of Eq. (2.10) is given by

$$\langle\mathbf{u},\mathbf{z}'|\mathbf{w})=G(\mathbf{z}',\mathbf{w}^*)\exp(\tfrac{1}{2}\operatorname{tr}\mathbf{u}\mathbf{w}^*), \tag{2.11}$$

where $G(\mathbf{z}',\mathbf{w}^*)$ is an arbitrary function of $\mathbf{z}'$ and $\mathbf{w}^*$. If we retain the normalization condition (2.5), then

$$G(\mathbf{0},\mathbf{w}^*)=1. \tag{2.12}$$

Hence the $\mathbf{z}'$ dependence of the right-hand side of Eq. (2.11) remains arbitrary. Therefore, for any given complex symmetric matrix $\mathbf{w}$, Eq. (1.2) admits more than one independent solution.

Let us consider the states $|\mathbf{w},\mathbf{z})$ defined by the relation

$$\langle\mathbf{u},\mathbf{z}'|\mathbf{w},\mathbf{z})=\hat{K}(\mathbf{z}';\mathbf{z}^*)\exp(\tfrac{1}{2}\operatorname{tr}\mathbf{u}\mathbf{w}^*)=\langle\mathbf{z}'|\mathbf{z}\rangle\langle\mathbf{u}|\mathbf{w}), \tag{2.13}$$

where $\mathbf{z}$ has the same meaning as $\mathbf{z}'$, i.e., parametrizes the $U(d)$ CS corresponding to the irrep $[\lambda]$, and $\hat{K}(\mathbf{z}';\mathbf{z}^*)$ is the $U(d)$ CS overlap, whose explicit expression is given in Eq. (I 4.22). As in I, we may choose for $\mathbf{z}$ either the $\mathbf{x}$ or $\mathbf{y}$ parameters. From Eqs. (2.11) and (2.12), it is obvious that the states $|\mathbf{w},\mathbf{z})$ are solutions of Eq. (1.2) and satisfy the normalization condition (2.5), i.e.,

$$D_{ij}|\mathbf{w},\mathbf{z})=w_{ij}^*|\mathbf{w},\mathbf{z}), \tag{2.14}$$

and

$$\langle(\lambda)_{\min}|\mathbf{w},\mathbf{z})=1. \tag{2.15}$$

Moreover, they are uniquely determined by Eq. (2.13), since they can be expanded as

$$|\mathbf{w},\mathbf{z}) = \int d\hat{\sigma}(\mathbf{u},\mathbf{z}')\hat{K}(\mathbf{z}';\mathbf{z}^*)\exp(\tfrac{1}{2}\,\mathrm{tr}\,\mathbf{u}\mathbf{w}^*)|\mathbf{u},\mathbf{z}'\rangle, \tag{2.16}$$

by making use of Eq. (I 6.1).

For any complex symmetric matrix $\mathbf{w}$, the states $|\mathbf{w},\mathbf{z})$ form an uncountably infinite set of linearly dependent solutions of Eq. (1.2). In Ref. 11, it has indeed been shown that Eq. (1.2) has exactly Λ independent solutions $|\mathbf{w};(\lambda))$, which may be labeled by the Gel'fand patterns (λ) associated with the irrep[12–14] $[\lambda]$, and may be defined by the relation

$$\langle \mathbf{u};(\lambda')|\mathbf{w};(\lambda)) = \delta_{(\lambda'),(\lambda)}\langle \mathbf{u}|\mathbf{w}), \tag{2.17}$$

where $|\mathbf{u};(\lambda')\rangle$ is the PCS given in Eq. (I 3.4).

We now proceed to prove that the known expansion of the unitary-operator CS into PCS, given in Eq. (I 8.5),

$$|\mathbf{u},\mathbf{z}\rangle = \sum_{(\lambda)} [\hat{\phi}_{(\lambda)}(\mathbf{z})]^*|\mathbf{u};(\lambda)\rangle, \tag{2.18}$$

leads to a similar expansion of the annihilation-operator CS in terms of the Λ independent solutions of Eq. (1.2),

$$|\mathbf{w},\mathbf{z}) = \sum_{(\lambda)} [\hat{\phi}_{(\lambda)}(\mathbf{z})]^*|\mathbf{w};(\lambda)). \tag{2.19}$$

For such purpose, all we have to show is that the overlap of the right-hand side of Eq. (2.19) with the bra $\langle \mathbf{u},\mathbf{z}'|$ satisfies Eq. (2.13). From Eqs. (2.17) and (2.18), we immediately obtain the result

$$\langle \mathbf{u},\mathbf{z}'|\sum_{(\lambda)} [\hat{\phi}_{(\lambda)}(\mathbf{z})]^*|\,\mathbf{w};(\lambda)) = \left\{\sum_{(\lambda)} \hat{\phi}_{(\lambda)}(\mathbf{z}')[\hat{\phi}_{(\lambda)}(\mathbf{z})]^*\right\}\langle \mathbf{u}|\mathbf{w}), \tag{2.20}$$

where the factor between curly braces may be rewritten as

$$\sum_{(\lambda)} \hat{\phi}_{(\lambda)}(\mathbf{z}')[\hat{\phi}_{(\lambda)}(\mathbf{z})]^* = \sum_{(\lambda)} \langle \mathbf{z}'|(\lambda)\rangle\langle(\lambda)|\mathbf{z}\rangle = \langle \mathbf{z}'|\mathbf{z}\rangle, \tag{2.21}$$

thus completing the proof of Eq. (2.19).

In the next section, we shall review some properties of the states $|\mathbf{w},\mathbf{z})$. Their demonstration could be based on Eq. (2.19) and the corresponding properties of the PCS $|\mathbf{w};(\lambda))$ established in Ref. 11. As an alternative, for the sake of easiness we shall start from the definition (2.13) of $|\mathbf{w},\mathbf{z})$.

III. SOME PROPERTIES OF THE ANNIHILATION-OPERATOR COHERENT STATES

To begin with, let us determine the overlap of the states $|\mathbf{w},\mathbf{z})$ with the discrete basis states of $\mathscr{F}_{(\lambda)}$, introduced in Eq. (I 2.6),

$$|\mathbf{N};(\lambda)\rangle = F_{\mathbf{N}}(\mathbf{D}^\dagger)|(\lambda)\rangle, \tag{3.1}$$

where $F_{\mathbf{N}}(\mathbf{D}^+)$ is defined in Eq. (I 2.7). From the Hermitian conjugate of Eq. (2.14), it results that

$$\phi_{\mathbf{N}(\lambda)}(\mathbf{w},\mathbf{z}) \equiv (\mathbf{w},\mathbf{z}|\mathbf{N};(\lambda)\rangle = F_{\mathbf{N}}(\mathbf{w})(\mathbf{w},\mathbf{z}|(\lambda)\rangle. \tag{3.2}$$

To calculate $(\mathbf{w},\mathbf{z}|(\lambda)\rangle$, it is convenient to expand $|(\lambda)\rangle$ on the continuous basis of $U(d)$ CS associated with the irrep $[\lambda]$. For such purpose, we need their unity resolution relation, Eq. (I 6.20), rewritten in the following form:

$$\int d\hat{\rho}(\mathbf{z})|\mathbf{z}\rangle\langle\mathbf{z}| = I_{[\lambda]}; \tag{3.3}$$

we obtain

$$(\mathbf{w},\mathbf{z}|(\lambda)\rangle = \int d\hat{\rho}(\mathbf{z}')(\mathbf{w},\mathbf{z}|\mathbf{z}'\rangle\langle\mathbf{z}'|(\lambda)\rangle, \tag{3.4}$$

where from Eq. (2.13)

$$(\mathbf{w},\mathbf{z}|\mathbf{z}'\rangle = (\mathbf{w},\mathbf{z}|\mathbf{0},\mathbf{z}'\rangle = \langle\mathbf{z}|\mathbf{z}'\rangle. \tag{3.5}$$

By using Eq. (3.3) again, Eq. (3.4) becomes

$$(\mathbf{w},\mathbf{z}|(\lambda)\rangle = \langle\mathbf{z}|(\lambda)\rangle = \hat{\phi}_{(\lambda)}(\mathbf{z}). \tag{3.6}$$

Finally, by combining Eqs. (3.2) and (3.6), we get the desired result

$$\phi_{\mathbf{N}(\lambda)}(\mathbf{w},\mathbf{z}) = F_{\mathbf{N}}(\mathbf{w})\hat{\phi}_{(\lambda)}(\mathbf{z}). \tag{3.7}$$

Similarly, the overlap of $|\mathbf{w},\mathbf{z})$ with the discrete basis states of Eq. (I 2.9),

$$|([l][\lambda])\alpha[h]\beta(k)q\rangle = [P_{[l]}(\mathbf{D}^\dagger)\times|(\ )\rangle]^{\alpha[h]}_{\beta(k)q}, \tag{3.8}$$

classified according to the group chain $Sp(2d,R) \supset U(d) \supset SO(d)$, is given by

$$\phi_{[l]\alpha[h]\beta(k)q}(\mathbf{w},\mathbf{z}) \equiv (\mathbf{w},\mathbf{z}|([l][\lambda])\alpha[h]\beta(k)q\rangle = [P_{[l]}(\mathbf{w})\times\hat{\phi}_{(\ )}(\mathbf{z})]^{\alpha[h]}_{\beta(k)q}. \tag{3.9}$$

Anticipating the results of Sec. V, where the set of states $|\mathbf{w},\mathbf{z})$ will be shown to form a continuous basis of $\mathscr{F}_{(\lambda)}$, we may interpret the functions (3.7) and (3.9) as the respective annihilation-operator CS representation of the states (3.1) and (3.8). By contrast to the unitary-operator CS representation, the present representation is very simple, and actually can be written in explicit form for $Sp(4,R)$ and $Sp(6,R)$. For such purpose, we only have to introduce into Eqs. (3.7) and (3.9) the explicit form of $\hat{\phi}_{(\lambda)}(\mathbf{z})$ given by

$$\hat{\phi}_\mu(z) = [(\lambda_1-\lambda_2)!]^{1/2}[(\lambda_1-\mu)!(\mu-\lambda_2)!]^{-1/2}z^{\mu-\lambda_2}, \tag{3.10a}$$

or

$$\hat{\phi}_m(z) = [(2j)!]^{1/2}[(j-m)!(j+m)!]^{-1/2}z^{j+m}, \tag{3.10b}$$

in the $U(2)$ case, and by

$$\hat{\phi}_{\mu_1\mu_2\nu}(\mathbf{x}) = N^{\lambda_1\lambda_2\lambda_3}_{\mu_1\mu_2\nu}\sum_k \binom{\mu_1-\mu_2-k}{\mu_1-\nu}\binom{\mu_1-\lambda_2}{k} \times x_1^{\mu_1-\lambda_2-k}x_3^{\nu-\mu_2-k}(x_2-x_1x_3)^{\mu_2-\lambda_3+k}, \tag{3.11}$$

in the $U(3)$ case. Here the Gel'fand patterns (λ) are denoted by $(\mu+n/2)$ and $\left(\begin{smallmatrix}\mu_1+n/2 & \mu_2+n/2\\ \nu+n/2\end{smallmatrix}\right)$ for $U(2)$ and $U(3)$, respectively, $j=(\lambda_1-\lambda_2)/2$, $m=\mu-(\lambda_1+\lambda_2)/2$, and $N^{\lambda_1\lambda_2\lambda_3}_{\mu_1\mu_2\nu}$ is the following normalization coefficient:

$$N^{\lambda_1\lambda_2\lambda_3}_{\mu_1\mu_2\nu} = (-1)^{\mu_2-\lambda_3}[(\mu_1-\mu_2+1)(\lambda_1-\lambda_2)!(\lambda_1-\lambda_3+1)!(\lambda_2-\lambda_3)!(\mu_1-\nu)!(\nu-\mu_2)!]^{1/2} \times[(\lambda_1-\mu_1)!(\mu_1-\lambda_2)!(\mu_1-\lambda_3+1)!(\lambda_1-\mu_2+1)!(\lambda_2-\mu_2)!(\mu_2-\lambda_3)!]^{-1/2}. \tag{3.12}$$

Equations (3.10) to (3.12) are valid for generic irreps, and are demonstrated in Appendix A. The results for U(3) irreps, for which $\lambda_1 > \lambda_2 = \lambda_3$, or $\lambda_1 = \lambda_2 > \lambda_3$, can be easily obtained by specializing Eqs. (3.11) and (3.12), and by using Eq. (I 3.16) relating the **x** and **y** parametrizations. They are given by

$$\hat{\phi}_{\mu_1\lambda_2\nu}(\mathbf{x}) = [(\lambda_1 - \lambda_2)!]^{1/2} \times [(\lambda_1 - \mu_1)!(\mu_1 - \nu)!(\nu - \lambda_2)!]^{-1/2} \times x_1^{\mu_1 - \nu} x_2^{\nu - \lambda_2}, \quad (3.13)$$

and

$$\hat{\phi}_{\lambda_1\mu_2\nu}(\mathbf{y}) = (-1)^{\mu_2 - \lambda_3}[(\lambda_1 - \lambda_3)!]^{1/2} \times [(\mu_2 - \lambda_3)!(\lambda_1 - \nu)!(\nu - \mu_2)!]^{-1/2} \times y_2^{\mu_2 - \lambda_3} y_3^{\nu - \mu_2}, \quad (3.14)$$

respectively.

Let us now consider the action of the Sp($2d,R$) generators on the annihilation-operator CS $|\mathbf{w},\mathbf{z})$. We shall proceed to prove that any generator X is equivalent to some partial differential operator $\mathscr{X}$ with respect to **w** and **z** when applied to a bra $(\mathbf{w},\mathbf{z}|$,

$$(\mathbf{w},\mathbf{z}|X = \mathscr{X}(\mathbf{w},\mathbf{z}|, \quad (3.15)$$

and find the explicit expression of $\mathscr{X}$. Then by taking the Hermitian conjugate of Eq. (3.15), it will result that

$$X^\dagger|\mathbf{w},\mathbf{z}) = \mathscr{X}^*|\mathbf{w},\mathbf{z}), \quad (3.16)$$

where $\mathscr{X}^*$ is the complex conjugate of $\mathscr{X}$, i.e., a partial differential operator with respect to $\mathbf{w}^*$ and $\mathbf{z}^*$. Once again anticipating the results of Sec. V, we may interpret $\mathscr{X}$ as the annihilation-operator CS representation of X.

From the Hermitian conjugate of Eq. (2.14), it is obvious that the representation of $\mathbf{D}^\dagger$ is simply given by

$$\mathscr{D}^\dagger = \mathbf{w}. \quad (3.17)$$

To find the representation of $\mathscr{E}$ and $\mathscr{D}$, it is convenient to take the scalar product of both sides of Eq. (3.15) with a continuous basis state $|\mathbf{u},\mathbf{z}'\rangle$, and to use the Hermiticity properties and the unitary-operator CS representation of the Sp($2d,R$) generators, respectively, given in Eqs. (I 2.2) and (I 5.3). In this way, we obtain the following relations:

$$(\mathbf{w},\mathbf{z}|E_{ij}|\mathbf{u},\mathbf{z}'\rangle = \langle\mathbf{u},\mathbf{z}'|E_{ji}|\mathbf{w},\mathbf{z})^* = [\mathbf{u}^*\Delta_{u^*} + \mathring{\mathscr{E}}'^*]_{ji}(\mathbf{w},\mathbf{z}|\mathbf{u},\mathbf{z}'\rangle, \quad (3.18)$$

and

$$(\mathbf{w},\mathbf{z}|D_{ij}|\mathbf{u},\mathbf{z}'\rangle = \langle\mathbf{u},\mathbf{z}'|D^\dagger_{ij}|\mathbf{w},\mathbf{z})^* = \{\mathbf{u}^*\tilde{\mathring{\mathscr{E}}}'^* + \mathring{\mathscr{E}}'^*\mathbf{u}^* + [\mathbf{u}^*\Delta_{u^*} - (d+1)\mathrm{I}]\mathbf{u}^*\}_{ij} \times (\mathbf{w},\mathbf{z}|\mathbf{u},\mathbf{z}'\rangle, \quad (3.19)$$

where $\mathring{\mathscr{E}}'^*_{ij}$ is a differential operator with respect to $\mathbf{z}'^*$. It is now straightforward to convert the differential operators with respect to $\mathbf{u}^*$ and $\mathbf{z}'^*$ into differential operators with respect to **w** and **z** in Eqs. (3.18) and (3.19), by applying Eq. (2.13) and the following relations:

$$[\Delta_{u_{ij}}, u_{kl}] = [\Delta_{w_{ij}}, w_{kl}] = \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}, \quad (3.20)$$

$$u^*_{ij}\exp(\tfrac{1}{2}\,\mathrm{tr}\,\mathbf{u}^*\mathbf{w}) = \Delta_{w_{ij}}\exp(\tfrac{1}{2}\,\mathrm{tr}\,\mathbf{u}^*\mathbf{w}), \quad (3.21)$$

$$(\mathbf{u}^*\Delta_{u^*})_{ji}\exp(\tfrac{1}{2}\,\mathrm{tr}\,\mathbf{u}^*\mathbf{w}) = (\mathbf{w}\Delta_w)_{ij}\exp(\tfrac{1}{2}\,\mathrm{tr}\,\mathbf{u}^*\mathbf{w}), \quad (3.22)$$

$$\mathring{\mathscr{E}}'^*_{ji}\langle\mathbf{z}|\mathbf{z}'\rangle = \langle\mathbf{z}'|E_{ji}|\mathbf{z}\rangle^* = \langle\mathbf{z}|E_{ij}|\mathbf{z}'\rangle = \mathring{\mathscr{E}}_{ij}\langle\mathbf{z}|\mathbf{z}'\rangle. \quad (3.23)$$

The results for $\mathscr{E}$ and $\mathscr{D}$ read

$$\mathscr{E} = \mathbf{w}\Delta_w + \mathring{\mathscr{E}}, \quad (3.17')$$

and

$$\mathscr{D} = \Delta_w\mathring{\mathscr{E}} + \tilde{\mathring{\mathscr{E}}}\Delta_w + \Delta_w[\mathbf{w}\Delta_w - (d+1)\mathrm{I}], \quad (3.17'')$$

where $\mathring{\mathscr{E}}$ is given by Eqs. (I 5.9) and (I 5.10) for U(2) and U(3) respectively. We note that Eqs. (3.17) have the same structure as the corresponding equations in the PCS representation corresponding to the states $|\mathbf{w};(\lambda))$, Eq. (6.15) of Ref. 11.

In Sec. V, we shall realize both the functions $\phi_{\mathrm{N}(\lambda)}(\mathbf{w},\mathbf{z})$ and the differential operators $\mathscr{X}$ in a subspace of a Bargmann Hilbert space of analytic functions, thereby showing that the annihilation-operator CS satisfy a unity resolution relation in $\mathscr{F}_{(\lambda)}$. For such purpose, it is useful to first establish similar results for the functions $\hat{\phi}_{(\lambda)}(\mathbf{z})$ and the differential operators $\mathring{\mathscr{E}}_{ij}$, corresponding to the U(d) CS representation. This is the topic of the next section.

IV. REALIZATION OF THE U(d) COHERENT STATE REPRESENTATION IN A BARGMANN SPACE OF ANALYTIC FUNCTIONS

Let us consider the U(d) group generated by the operators

$$E_{ij} = \sum_{s=1}^{n}\eta_{is}\,\xi_{js} + \frac{n}{2}\delta_{ij}, \quad i,j = 1,\dots,d, \quad (4.1)$$

where η_{is}, ξ_{is}, $i = 1,\dots,d$, $s = 1,\dots,n$, are the boson creation and annihilation operators introduced in I. This U(d) group can be embedded into a larger group U(dn) in the following way[13]:

$$\begin{matrix} U(dn) & \supset & U(d) & \times & U(n) \\ [N+\tfrac{1}{2},(\tfrac{1}{2})^{dn-1}] & & [\lambda] \equiv [\lambda_1 + n/2,\dots,\lambda_d + n/2] & & [\lambda]' \equiv [\lambda_1 + d/2,\dots,\lambda_d + d/2,(d/2)^{n-d}]. \end{matrix} \quad (4.2)$$

Here U(dn) and U(n) are generated by the operators

$$E_{is,jt} = \tfrac{1}{2}(\eta_{is}\xi_{jt} + \xi_{jt}\eta_{is}) = \eta_{is}\xi_{jt} + \tfrac{1}{2}\delta_{ij}\delta_{st}, \qquad (4.3)$$

and

$$E_{st} = \sum_{i=1}^{d} \eta_{is}\xi_{it} + \frac{d}{2}\delta_{st}, \qquad (4.4)$$

respectively. In Eq. (4.2), underneath each group we have indicated the labels characterizing its irreps. All N-boson states belong to a single U(dn) irrep characterized by $[N+\frac{1}{2},(\frac{1}{2})^{dn-1}]$. Within any such irrep, the U(d) and U(n) groups are complementary.[13,16] This means that the reduction of the U(dn) irrep into irreps of U$(d)\times$U(n) is multiplicity free and that any U(d) irrep $[\lambda]$, for which $\Sigma_i\lambda_i = N$, occurs with multiplicity equal to the dimension of the U(n) irrep $[\lambda]'$, and conversely. The representation space of $[\lambda]$ may therefore be realized by selecting all N-boson states transforming under a U(n) irrep characterized by $[\lambda]'$, and belonging to a definite row of the latter, e.g., the highest weight state.

In the Bargmann representation,[15] η_{is} and ξ_{is} are, respectively, represented by some complex variables $\overline{w}_{is}$, and the corresponding differential operators $\partial/\partial\overline{w}_{is}$. Any boson state $|\psi\rangle$ is represented by an analytic function $\psi(\overline{\mathbf{w}})$ in the dn complex variables $\overline{w}_{is}$, $i=1,\dots,d$, $s=1,\dots,n$. The space spanned by the analytic functions $\psi(\overline{\mathbf{w}})$ is a Hilbert space, whose scalar product is defined by

$$(\chi,\psi) = \int d\mu(\overline{\mathbf{w}})[\chi(\overline{\mathbf{w}})]^*\psi(\overline{\mathbf{w}}), \qquad (4.5)$$

where

$$d\mu(\overline{\mathbf{w}}) = \prod_{i=1}^{d}\prod_{s=1}^{n} d\mu(\overline{w}_{is}), \qquad (4.6)$$

and

$$d\mu(\overline{w}_{is}) = \pi^{-1}\exp(-\overline{w}_{is}^*\overline{w}_{is})d\overline{w}_{is}\, d\overline{w}_{is}^*. \qquad (4.7)$$

In Bargmann space, the operators $E_{is,jt}$, E_{ij}, and E_{st} are represented by partial differential operators, denoted for simplicity's sake by the same symbols, for instance,

$$E_{ij} = \sum_{s=1}^{n}\overline{w}_{is}\frac{\partial}{\partial\overline{w}_{js}} + \frac{n}{2}\delta_{ij}. \qquad (4.8)$$

The functions $\psi_{(\lambda)}(\overline{\mathbf{w}})$ of the d^2 variables $\overline{w}_{ij}$, $i,j=1,\dots,d$, which are the simultaneous solutions of the system of partial differential equations

$$\begin{aligned}&\sum_{k=1}^{d}\overline{w}_{ki}\frac{\partial}{\partial\overline{w}_{ki}}\psi_{(\lambda)}(\overline{\mathbf{w}}) = \lambda_i\psi_{(\lambda)}(\overline{\mathbf{w}}),\\ &\sum_{k=1}^{d}\overline{w}_{ki}\frac{\partial}{\partial\overline{w}_{kj}}\psi_{(\lambda)}(\overline{\mathbf{w}}) = 0, \quad i<j,\end{aligned} \qquad (4.9)$$

span a subspace of Bargmann space, which provides us with a realization of the U(d) irrep $[\lambda]$ representation space. The functions $\psi_{(\lambda)}(\overline{\mathbf{w}})$ can be written as[13,17]

$$\psi_{(\lambda)}(\overline{\mathbf{w}}) = \psi_{(\lambda)_{\min}}(\overline{\mathbf{w}}) \times Z_{(\lambda)}\left(\frac{\overline{w}_{i1}}{\overline{w}_{d1}}, \frac{\overline{w}_{id,12}}{\overline{w}_{d-1d,12}}, \dots, \frac{\overline{w}_{13\cdots d,12\cdots d-1}}{\overline{w}_{2\cdots d,1\cdots d-1}}\right), \qquad (4.10)$$

in terms of the lowest weight state function

$$\psi_{(\lambda)_{\min}}(\overline{\mathbf{w}}) = M^{[\lambda]}(\overline{w}_{d1})^{\lambda_1-\lambda_2}(\overline{w}_{d-1d,12})^{\lambda_2-\lambda_3} \times\cdots\times(\overline{w}_{2\cdots d,1\cdots d-1})^{\lambda_{d-1}-\lambda_d}(\overline{w}_{1\cdots d,1\cdots d})^{\lambda_d}, \qquad (4.11)$$

and of some polynomials $Z_{(\lambda)}$ in the indicated ratios, subject to the condition that $\psi_{(\lambda)}(\overline{\mathbf{w}})$ should be a polynomial in $\overline{w}_{ij}$. Here $\overline{w}_{i_1\cdots i_r j_1\cdots j_r}$ denotes the minor of order r of det $\overline{\mathbf{w}}$ corresponding to rows $i_1,\dots,i_r$, and columns $j_1,\dots,j_r$, and $M^{[\lambda]}$ is a normalization coefficient given by

$$M^{[\lambda]} = \left[\prod_{i<j}(\lambda_i-\lambda_j+j-i)\right]^{1/2} \times\left[\prod_i(\lambda_i+d-i)!\right]^{-1/2}. \qquad (4.12)$$

From Eqs. (4.10) and (4.11), both $\psi_{(\lambda)_{\min}}$ and $Z_{(\lambda)}$ are only depending upon some special combinations of the variables $\overline{w}_{ij}$. If we make the following change of variables

$$x_{ji} = \begin{cases}\dfrac{\overline{w}_{ij+1\cdots d,1\cdots d-j+1}}{\overline{w}_{j\cdots d,1\cdots d-j+1}}, & \text{if } i=1,\dots,j-1,\\ \overline{w}_{j\cdots d,1\cdots d-j+1}, & \text{if } i=j,\\ \overline{w}_{ji}, & \text{if } i=j+1,\dots,d,\end{cases} \qquad (4.13)$$

for any $j=1,\dots,d$, then $\psi_{(\lambda)_{\min}}$ becomes a polynomial in x_{ii}, $i=1,\dots,d$,

$$\psi_{(\lambda)_{\min}}(x_{11},\dots,x_{dd}) = M^{[\lambda]}(x_{dd})^{\lambda_1-\lambda_2}(x_{d-1,d-1})^{\lambda_2-\lambda_3} \times\cdots\times(x_{22})^{\lambda_{d-1}-\lambda_d}(x_{11})^{\lambda_d}, \qquad (4.14)$$

whereas $Z_{(\lambda)}$ is converted into a polynomial in x_{ji}, $j>i$, which coincides with the U(d) CS representation $\hat{\phi}_{(\lambda)}(\mathbf{x})$ of the Gel'fand state $|(\lambda)\rangle$ in the $\mathbf{x}$ parametrization,

$$Z_{(\lambda)}(x_{d1},\dots,x_{d,d-1},x_{d-1,1},\dots,x_{d-1,d-2},\dots,x_{21}) = \hat{\phi}_{(\lambda)}(\mathbf{x}). \qquad (4.15)$$

The latter assertion can be easily checked for U(2) and U(3) by explicit construction of $Z_{(\lambda)}$ using the raising operator technique of Appendix A. By way of illustration, it is proved for U(2) in Appendix B.

Up to some fixed, irrelevant dependence on $x_{11},\dots,x_{dd}$, the bases $\psi_{(\lambda)}(\mathbf{x})$ of the subspace of Bargmann space, characterized by a given U(d) irrep $[\lambda]$ and of highest weight with respect to U(n), therefore provide us with a realization of the U(d) CS functions $\hat{\phi}_{(\lambda)}(\mathbf{x})$ corresponding to the same U(d) irrep. This is corroborated by the fact that the CS representation $\mathring{\mathscr{E}}_{ij}$ of the U(d) generators, given in I, coincides with the differential operators with respect to $\mathbf{x}$, defined by the following relation:

$$E_{ij}\psi_{(\lambda)}(\mathbf{x}) = \psi_{(\lambda)_{\min}}(x_{11},\dots,x_{dd})\mathring{\mathscr{E}}_{ij}\hat{\phi}_{(\lambda)}(\mathbf{x}), \qquad (4.16)$$

where E_{ij} is given by Eq. (4.8). This statement is again easily checked for U(2) and U(3). It is proved for U(2) in Appendix B.

Finally, the U(d) CS measure $d\hat{\rho}(\mathbf{x})$, defined in Eqs. (3.3), (I 6.17), and (I 6.20), can be derived from Bargmann measure $d\mu(\overline{\mathbf{w}})$, given in Eqs. (4.6) and (4.7). For such purpose, we start from the orthogonality properties of the U(d) irrep bases $\psi_{(\lambda)}(\overline{\mathbf{w}})$ in Bargmann space,

$$\int d\mu(\overline{\mathbf{w}})[\psi_{(\lambda')}(\overline{\mathbf{w}})]^*\psi_{(\lambda)}(\overline{\mathbf{w}}) = \delta_{(\lambda'),(\lambda)}, \quad (4.17)$$

and of the U(d) CS functions $\hat{\phi}_{(\lambda)}(\mathbf{x})$,

$$\int d\hat{\rho}(\mathbf{x})[\hat{\phi}_{(\lambda')}(\mathbf{x})]^*\phi_{(\lambda)}(\mathbf{x}) = \delta_{(\lambda'),(\lambda)}. \quad (4.18)$$

By comparing Eqs. (4.17) and (4.18), we get the following equation:

$$\int d\mu(\overline{\mathbf{w}})[\psi_{(\lambda')}(\overline{\mathbf{w}})]^*\psi_{(\lambda)}(\overline{\mathbf{w}}) = \int d\hat{\rho}(\mathbf{x})[\hat{\phi}_{(\lambda')}(\mathbf{x})]^*\hat{\phi}_{(\lambda)}(\mathbf{x}). \quad (4.19)$$

To obtain the U(d) CS measure $d\hat{\rho}(\mathbf{x})$, we therefore only have to make the change of variables (4.13) and successively integrate over x_{ji}, $j<i$, and x_{ii}, in the left-hand side of Eq. (4.19). This procedure, although straightforward, becomes quite tedious for d values greater than 3. As an example, we detail the derivation for $d=2$ in Appendix B.

The above results are valid for generic U(d) irreps, i.e., irreps for which $\lambda_1>\lambda_2>\cdots>\lambda_d$, and for the $\mathbf{x}$ parametrization of CS. To conclude the present section, let us briefly indicate how the $\mathbf{y}$ parametrization of the CS representation can be realized in Bargmann space, and the results extended to nongeneric irreps.

As for the former point, it is straightforward to check that the polynomial $Z_{(\lambda)}$ in the $d(d-1)/2$ variables x_{ji}, $j>i$, can alternatively be written in terms of the $d(d-1)/2$ variables

$$y_{ji} = \overline{w}_{i+1\cdots j-1\, ij+1\cdots d,1\cdots d-i}/\overline{w}_{i+1\cdots d,1\cdots d-i}, \quad j>i, \quad (4.20)$$

related to the previous ones by the determinantal relation

$$y_{ji} = x_{ji} - \sum_{i<k<j} x_{jk}x_{ki} + \sum_{i<k<l<j} x_{jl}x_{lk}x_{ki} - \cdots + (-1)^{j-i-1}x_{j,j-1}x_{j-1,j-2}\cdots x_{i+1,i}. \quad (4.21)$$

Since Eq. (4.21) is nothing but the relation between the $\mathbf{x}$ and $\mathbf{y}$ parameters, Eq. (I 3.23b), the transformed polynomial $Z_{(\lambda)}(\mathbf{y})$ coincides with the U(d) CS representation $\hat{\phi}_{(\lambda)}(\mathbf{y})$ of the Gel'fand state $|(\lambda)\rangle$ in the $\mathbf{y}$ parametrization.

As for the latter point, we note that the transition from the generic irreps to nongeneric ones can be carried out by setting some of the $\mathbf{x}$ or $\mathbf{y}$ parameters equal to zero. In Bargmann space, this arises quite naturally from the fact that for nongeneric irreps $Z_{(\lambda)}$ does not depend on the whole set of variables x_{ji} or y_{ji}, $j>i$. In deriving the U(d) CS measure from Bargmann measure, extra integrations over the missing variables can then be performed. For instance, for cases b or c U(3) irreps (corresponding to $\lambda_1>\lambda_2=\lambda_3$ or $\lambda_1=\lambda_2>\lambda_3$, respectively), $Z_{(\lambda)}$ is independent of the variable $x_{21}=\overline{w}_{13,12}/\overline{w}_{23,12}$ or $y_{32}=\overline{w}_{21}/\overline{w}_{31}$, which is set equal to zero in the corresponding CS representation. Integration over x_{21} or y_{32}, in addition to x_{ji}, $j<i$, leads to the results contained in Eqs. (I 6.12), (I 6.17), and (I 6.22).

After these preliminaries on a realization of the U(d) CS representation in a Bargmann space of analytic functions, we can now proceed to derive similar results for the Sp($2d,R$) annihilation-operator CS representation.

V. REALIZATION OF THE Sp($2d,R$) ANNIHILATION-OPERATOR COHERENT STATE REPRESENTATION IN A BARGMANN SPACE OF ANALYTIC FUNCTIONS

As is well known,[1] the Sp($2d,R$) group can be embedded into the larger group Sp($2dn,R$), whose generators are the operators $E_{is,jt}$, defined in Eq. (4.3), and

$$D^\dagger_{is,jt} = D^\dagger_{jt,is} = \eta_{is}\eta_{jt}, \quad (is)\leqslant(jt),$$
$$D_{is,jt} = D_{jt,is} = \xi_{is}\xi_{jt}, \quad (is)\leqslant(jt). \quad (5.1)$$

As a matter of fact, we have the following group chain:

$$\begin{matrix} \mathrm{Sp}(2dn,R) & \supset & \mathrm{Sp}(2d,R) & \times & \mathrm{O}(n) \\ \langle(\tfrac12)^{dn}\rangle \text{ or } \langle(\tfrac12)^{dn-1}\tfrac32\rangle & & \langle\lambda\rangle \equiv \langle\lambda_d+n/2,\ldots,\lambda_1+n/2\rangle & & (\lambda_1\cdots\lambda_d), \end{matrix} \quad (5.2)$$

where O(n) is generated by the operators

$$\Lambda_{st} = -\Lambda_{ts} = -i(E_{st}-E_{ts}), \quad 1\leqslant s<t\leqslant n. \quad (5.3)$$

In Eq. (5.2), underneath each group we have indicated the labels characterizing its irreps. All the boson states belong to one of two irreps of Sp($2dn,R$), $\langle(\frac12)^{dn}\rangle$ or $\langle(\frac12)^{dn-1}\frac32\rangle$, according to whether the total number N of bosons is even or odd. Within each one of them, the Sp($2d,R$) and O(n) groups are complementary.[1,16] The representation space $\mathscr{F}_{\langle\lambda\rangle}$ of $\langle\lambda\rangle$ may therefore be realized by selecting all the boson states transforming under an O(n) irrep characterized by $(\lambda_1\cdots\lambda_d)$, and belonging to a definite row of the latter, e.g., the highest weight state.

For practical purposes made clear later on, we now denote by g_{is} instead of $\overline{w}_{is}$ the complex variables representing the boson creation operators η_{is} in Bargmann representation. Let $\mathscr{H}$ be the corresponding Bargmann space, i.e., the Hilbert space spanned by the analytic functions $\psi(\mathbf{g})$, and endowed with the scalar product defined in Eqs. (4.5)–(4.7) with $\mathbf{g}$ substituted for $\overline{\mathbf{w}}$. In such a space, the O(n) and Sp($2d,R$) generators are represented by the differential operators

$$\Lambda_{st} = -i\sum_{i=1}^{d}\left(g_{is}\frac{\partial}{\partial g_{it}} - g_{it}\frac{\partial}{\partial g_{is}}\right), \quad (5.4)$$

and

$$D^\dagger_{ij} = \sum_{s=1}^n g_{is}\, g_{js},$$
$$D_{ij} = \sum_{s=1}^n \frac{\partial^2}{\partial g_{is}\, \partial g_{js}}, \qquad (5.5)$$
$$E_{ij} = \sum_{s=1}^n g_{is} \frac{\partial}{\partial g_{js}} + \frac{n}{2}\delta_{ij},$$

respectively.

The functional images of the boson states belonging to $\mathscr{F}_{\langle\lambda\rangle}$ span a subspace of Bargmann space, denoted by $\mathscr{H}_{\langle\lambda\rangle}$. By definition, the functions in $\mathscr{H}_{\langle\lambda\rangle}$ have definite transformation properties under the $O(n)$ group generated by the operators (5.4), that is, they are the highest weight states of all the $O(n)$ irreps characterized by the same labels $(\lambda_1 \cdots \lambda_d)$.

To construct such functions, an explicit procedure was devised in Ref. 18. To begin with, we introduce the new variables

$$a_{i\alpha} = 2^{-1/2}(g_{i,2\alpha-1} - ig_{i,2\alpha}),$$
$$b_{i\alpha} = 2^{-1/2}(g_{i,2\alpha-1} + ig_{i,2\alpha}),$$
$$c_i = g_{in} \quad (\text{only when } n = 2\nu + 1), \qquad (5.6)$$
$$i = 1,\ldots,d, \quad \alpha = 1,\ldots,\nu = [n/2].$$

Under transformation (5.6), Bargmann measure remains invariant. Next, we define

$$\begin{aligned} w_{ij} = w_{ji} &= \sum_{s=1}^n g_{is}\, g_{js} \\ &= \sum_{\alpha=1}^{\nu} (a_{i\alpha} b_{j\alpha} + a_{j\alpha} b_{i\alpha}), \quad \text{when } n = 2\nu, \\ &= \sum_{\alpha=1}^{\nu} (a_{i\alpha} b_{j\alpha} + a_{j\alpha} b_{i\alpha}) + c_i c_j, \quad \text{when } n = 2\nu+1, \\ \overline{w}_{i\alpha} &= a_{i\alpha}, \quad \alpha = 1,\ldots,\nu, \\ \overline{w}_{i\rho} &= b_{i\rho}, \quad \rho = 1,\ldots,\nu - i, \\ &= c_i, \quad \rho = \nu - i + 1 \quad (\text{only when } n = 2\nu+1). \\ & i,j = 1,\ldots,d. \end{aligned} \qquad (5.7)$$

Then, when using Wong's definition of weight and raising generators,[19] it turns out that the highest weight states of all the $O(n)$ irreps characterized by $(\lambda_1 \cdots \lambda_d)$ only depend upon the variables $w_{ij} = w_{ji}$, and $\overline{w}_{ij}$, $i,j = 1,\ldots,d$, and are the simultaneous solutions of the following system of partial differential equations:

$$\sum_{k=1}^d \overline{w}_{ki} \frac{\partial}{\partial \overline{w}_{ki}} \psi(\mathbf{w},\overline{\mathbf{w}}) = \lambda_i \psi(\mathbf{w},\overline{\mathbf{w}}),$$
$$\qquad (5.8)$$
$$\sum_{k=1}^d \overline{w}_{ki} \frac{\partial}{\partial \overline{w}_{kj}} \psi(\mathbf{w},\overline{\mathbf{w}}) = 0, \quad i < j.$$

The restriction of the $Sp(2d,R)$ generators to $\mathscr{F}_{\langle\lambda\rangle}$ is easily determined from the relation

$$\frac{\partial}{\partial g_{is}} \psi(\mathbf{w},\overline{\mathbf{w}}) = \left[\sum_j g_{js} \Delta_{w_{ij}} + \gamma_s \frac{\partial}{\partial \overline{w}_{i\alpha}} \right] \psi(\mathbf{w},\overline{\mathbf{w}}),$$
$$s = 2\alpha - 1 \text{ or } 2\alpha, \qquad (5.9)$$

where

$$\gamma_s = \begin{cases} 2^{-1/2}, & \text{if } s = 2\alpha - 1 < 2d, \\ -2^{-1/2} i, & \text{if } s = 2\alpha \leqslant 2d, \\ 0, & \text{if } s > 2d. \end{cases} \qquad (5.10)$$

The result reads

$$D^\dagger_{ij}\, \psi(\mathbf{w},\overline{\mathbf{w}}) = w_{ij}\, \psi(\mathbf{w},\overline{\mathbf{w}}), \qquad (5.11a)$$
$$\begin{aligned} D_{ij}\, \psi(\mathbf{w},\overline{\mathbf{w}}) = \Bigg(\sum_k \bigg[\Delta_{w_{ik}} \Big(\sum_l \overline{w}_{kl} \frac{\partial}{\partial \overline{w}_{jl}} + \frac{n}{2}\delta_{kj} \Big) \\ + \Big(\sum_l \overline{w}_{kl} \frac{\partial}{\partial \overline{w}_{il}} + \frac{n}{2}\delta_{ki} \Big) \Delta_{w_{kj}} \bigg] \\ + \{ \Delta_w [\mathbf{w}\Delta_w - (d+1)\mathbf{I}] \}_{ij} \Bigg) \psi(\mathbf{w},\overline{\mathbf{w}}), \end{aligned} \qquad (5.11b)$$
$$E_{ij}\, \psi(\mathbf{w},\overline{\mathbf{w}}) = \left[(\mathbf{w}\Delta_w)_{ij} + \sum_k \overline{w}_{ik} \frac{\partial}{\partial \overline{w}_{jk}} + \frac{n}{2}\delta_{ij} \right] \psi(\mathbf{w},\overline{\mathbf{w}}). \qquad (5.11c)$$

Since the Gel'fand states $|(\lambda)\rangle$ are annihilated by the generators D_{ij}, it results from Eq. (5.11b) that their functional images $\psi_{(\lambda)}(\mathbf{w},\overline{\mathbf{w}})$ do not depend upon the variables w_{ij}. For such functions, Eq. (5.8) reduces to Eq. (4.9). Taking into account that $\overline{w}_{ij} = a_{ij}$ are Bargmann variables, we conclude that all the results of Sec. IV are applicable to $\psi_{(\lambda)}(\overline{\mathbf{w}})$. In particular, the transformation (4.13) converts it into the following product:

$$\psi_{(\lambda)}(\overline{\mathbf{w}}) = \psi_{(\lambda)_{\min}}(x_{11},\ldots,x_{dd})\hat{\phi}_{(\lambda)}(\mathbf{x}). \qquad (5.12)$$

From Eq. (5.11a), it now results that the functional image of $|\mathbf{N};(\lambda)\rangle$ is given by

$$\psi_{\mathbf{N}(\lambda)}(\mathbf{w},\overline{\mathbf{w}}) = F_{\mathbf{N}}(\mathbf{w})\psi_{(\lambda)}(\overline{\mathbf{w}}). \qquad (5.13)$$

Comparison with Eqs. (3.7) and (5.12) shows that

$$\psi_{\mathbf{N}(\lambda)}(\mathbf{w},\overline{\mathbf{w}}) = \psi_{(\lambda)_{\min}}(x_{11},\ldots,x_{dd})\phi_{\mathbf{N}(\lambda)}(\mathbf{w},\mathbf{x}). \qquad (5.14)$$

Apart from a fixed, irrelevant dependence on $x_{11},\ldots,x_{dd}$, contained in $\psi_{(\lambda)_{\min}}(x_{11},\ldots,x_{dd})$, the Bargmann representation $\psi_{\mathbf{N}(\lambda)}(\mathbf{w},\overline{\mathbf{w}})$ of the $\mathscr{F}_{\langle\lambda\rangle}$ discrete bases $|\mathbf{N};(\lambda)\rangle$ therefore coincides with their annihilation-operator CS representation $\phi_{\mathbf{N}(\lambda)}(\mathbf{w},\mathbf{x})$. Hence the subspace $\mathscr{H}_{\langle\lambda\rangle}$ of Bargmann space carries a realization of the annihilation-operator CS representation, wherein the variables $w_{ij} = w_{ji}$ and x_{ji} $(j > i)$ are those combinations of g_{is} defined in Eqs. (4.13), (5.6), and (5.7).

The corresponding realization of the $Sp(2d,R)$ generators X results from Eq. (5.11) where Eqs. (4.8) and (4.16) are considered. It is straightforward to check the coincidence between the differential operators $\mathscr{X}$ with respect to $\mathbf{w}$ and $\mathbf{x}$, defined by the relation

$$X\psi_{\mathbf{N}(\lambda)}(\mathbf{w},\overline{\mathbf{w}}) = \psi_{(\lambda)_{\min}}(x_{11},\ldots,x_{dd})\mathscr{X}\phi_{\mathbf{N}(\lambda)}(\mathbf{w},\mathbf{x}), \qquad (5.15)$$

and the annihilation-operator CS representation of the $Sp(2d,R)$ generators, given in Eq. (3.17), as expected. Note that a preliminary account of the derivation of $\mathscr{X}$ in Bargmann space was already given in Ref. 20.

Finally, by the same way as the $U(d)$ CS measure was deduced from Bargmann measure in Sec. IV, it can be proved that the annihilation-operator CS satisfy a unity resolution relation with a measure $d\sigma(\mathbf{w},\mathbf{z})$, which, at least in principle, can be obtained from Bargmann measure. For such purpose, let us calculate in $\mathscr{H}_{\langle\lambda\rangle}$ the overlap of two discrete bases of $\mathscr{F}_{\langle\lambda\rangle}$,

$$\langle N';(\lambda')|\mathbf{N};(\lambda)\rangle = \int d\mu(\mathbf{g})[\psi_{\mathbf{N}'(\lambda')}(\mathbf{w},\overline{\mathbf{w}})]^*\psi_{\mathbf{N}(\lambda)}(\mathbf{w},\overline{\mathbf{w}}). \quad (5.16)$$

After performing the transformations (5.6), (5.7), and (4.13), then introducing Eq. (5.14) into the right-hand side of Eq. (5.16), the integrations over $\overline{w}_{i\alpha}$, $i=1,...,d$, $\alpha=d+1,...,\nu$, $\overline{w}_{i\rho}$, $i=1,...,d$, $\rho=1,...,\nu-i$ or $\nu-i+1$ (according as $n=2\nu$ or $2\nu+1$), and x_{ji}, $1<j<i<d$, can, at least in principle, be carried out. It remains an integral over w_{ij}, $1<i<j<d$, and x_{ji}, $1<i<j<d$, containing some measure $d\sigma(\mathbf{w},\mathbf{x})$,

$$\langle \mathbf{N}';(\lambda')|\mathbf{N};(\lambda)\rangle = \int d\sigma(\mathbf{w},\mathbf{x})[\phi_{\mathbf{N}'(\lambda')}(\mathbf{w},\mathbf{x})]^*\phi_{\mathbf{N}(\lambda)}(\mathbf{w},\mathbf{x}). \quad (5.17)$$

By taking the definition (3.2) of $\phi_{\mathbf{N}(\lambda)}(\mathbf{w},\mathbf{x})$ into account, Eq. (5.17) can be transformed into the following relation:

$$\langle \mathbf{N}';(\lambda')|\mathbf{N};(\lambda)\rangle = \int d\sigma(\mathbf{w},\mathbf{x})\langle \mathbf{N}';(\lambda')|\mathbf{w},\mathbf{x})(\mathbf{w},\mathbf{x}|\mathbf{N};(\lambda)\rangle. \quad (5.18)$$

As a result, we obtain the searched for unity resolution relation

$$\int d\sigma(\mathbf{w},\mathbf{x})|\mathbf{w},\mathbf{x})(\mathbf{w},\mathbf{x}| = I_{\langle\lambda\rangle}, \quad (5.19)$$

showing that the annihilation-operator CS form a continuous basis of $\mathscr{F}_{\langle\lambda\rangle}$. Although the procedure described above enables us to prove Eq. (5.19), it is not suitable for deriving the explicit form of $d\sigma(\mathbf{w},\mathbf{x})$. A more convenient alternative method for such purpose has been described elsewhere.[9]

VI. CONCLUSION

In the present paper, we have extended to all positive-discrete series irreps of Sp$(2d,R)$ the annihilation-operator CS introduced by Barut and Girardello for Sp$(2,R)$ (see Ref. 8) and later on generalized by Deenen and Quesne to the Sp$(2d,R)$ positive-discrete series irreps of the type $\langle(\lambda+n/2)^d\rangle$. The CS we have introduced for such purpose are of mixed type, in the sense that they are annihilation-operator CS as regards the noncompact generators only, while presenting some features of the unitary-operator CS (namely their parametrization) as concerns the compact generators.

We did show that such CS form a continuous basis of the irrep representation space $\mathscr{F}_{\langle\lambda\rangle}$, and we did obtain in compact form the corresponding representation of both the discrete bases of $\mathscr{F}_{\langle\lambda\rangle}$ and the Sp$(2d,R)$ generators. In addition, quite detailed formulas were written down for Sp$(4,R)$ and Sp$(6,R)$.

In contrast to what was done in I for the unitary-operator CS, we did not address ourselves to the determination of the explicit form, the reproducing kernel, nor the measure of the annihilation-operator CS. Even for the simplest case of the irreps $\langle(\lambda+n/2)^d\rangle$ treated in Ref. 9, such a determination is indeed quite tedious. As a matter of fact, the lacking quantities are irrelevant to the most important application of the annihilation-operator CS, that is of conceptual nature and will now be reviewed.

One of the many interesting properties of the standard CS consists in the resulting Bargmann representation, wherein the boson creation and annihilation operators η and ξ are, respectively, realized by the operator of multiplication by the complex variable g, and by the corresponding differential operator $\partial/\partial g$. Bargmann representation therefore provides the mathematical foundation for the widespread habit of replacing ξ by a symbolic differentiation with respect to η. In the same way, in the Sp$(2d,R)$ annihilation-operator CS representation, each noncompact raising generator $D^\dagger_{ij}$ is realized by an operator of multiplication by a complex variable w_{ij}, while the remaining generators become differential operators with respect to the variables w_{ij} [and some extra variables z_{ji} parametrizing U(d) CS]. The annihilation-operator CS representation, therefore, provides the mathematical foundation for the recently introduced use of differentiation operators with respect to $D^\dagger_{ij}$ in symbolic expressions of the Sp$(2d,R)$ generators.[21,22] In conclusion, it presents some of the simplifying features of Bargmann representation that are lacking in the unitary-operator CS representation.

APPENDIX A: U(2) AND U(3) BASIS FUNCTIONS IN THE COHERENT STATE REPRESENTATION

The purpose of this appendix is to derive the explicit form of the U(2) and U(3) basis functions $\hat{\phi}_\mu(z)$ and $\hat{\phi}_{\mu_1\mu_2\nu}(\mathbf{x})$, respectively, given in Eq. (3.10a) and in Eqs. (3.11) and (3.12).

As explained in Sec. 7 of I, the functional image $\hat{\phi}_{(\lambda)}(\mathbf{z})$ of a U(d) Gel'fand state $|(\lambda)\rangle$ in the CS representation can be obtained from that of the U(d) irrep lowest weight state

$$\hat{\phi}_{(\lambda)_{min}}(\mathbf{z}) = 1, \quad (A1)$$

by applying an appropriate polynomial $P_{(\lambda)}$ in the differential operators $\mathring{\mathscr{E}}_{ij}$,

$$\hat{\phi}_{(\lambda)}(\mathbf{z}) = P_{(\lambda)}(\mathring{\mathscr{E}})1. \quad (A2)$$

This polynomial can be expressed in terms of U(d), U$(d-1)$,...,U(2) raising operators. A U(n) raising operator R^n_m, $m=1,...,n-1$, is herein defined as a polynomial in the U(n) generators, transforming any Gel'fand state characterized by a U$(n-1)$ irrep $[h_1\cdots h_{n-1}]$, and of lowest weight in U$(n-1)$, into a Gel'fand state specified by the U$(n-1)$ irrep $[h_1\cdots h_m+1\cdots h_{n-1}]$, and still of lowest weight in U$(n-1)$. It differs from a Nagel–Moshinsky raising operator,[23] in the substitution of U$(n-1)$ lowest-weight states for their highest-weight ones, and can be constructed by using similar techniques.

For U(2) and U(3), our raising operators read

$$R^2_1 = \mathring{\mathscr{E}}_{12}, \quad (A3)$$

$$R^3_1 = \mathring{\mathscr{E}}_{23}, \quad R^3_2 = \mathring{\mathscr{E}}_{13}(\mathring{\mathscr{E}}_{11}-\mathring{\mathscr{E}}_{22}-1)+\mathring{\mathscr{E}}_{12}\mathring{\mathscr{E}}_{23}, \quad (A4)$$

and the normalized polynomials $P_{(\lambda)}(\mathring{\mathscr{E}})$ are given by

$$P_{\mu}(\mathring{\mathscr{E}}) = \mathscr{N}_{\mu}^{\lambda_1\lambda_2}(R_1^2)^{\mu-\lambda_2}, \tag{A5}$$

$$P_{\mu_1\mu_2\nu}(\mathring{\mathscr{E}}) = \mathscr{N}_{\mu_1\mu_2\nu}^{\lambda_1\lambda_2\lambda_3}(R_1^2)^{\nu-\mu_2}(R_1^3)^{\mu_1-\lambda_2}(R_2^3)^{\mu_2-\lambda_3}, \tag{A6}$$

where

$$\mathscr{N}_{\mu}^{\lambda_1\lambda_2} = [(\lambda_1-\mu)!]^{1/2}[(\lambda_1-\lambda_2)!(\mu-\lambda_2)!]^{-1/2}, \tag{A7}$$

$$\mathscr{N}_{\mu_1\mu_2\nu}^{\lambda_1\lambda_2\lambda_3} = [(\mu_1-\mu_2+1)(\lambda_1-\mu_1)!(\lambda_1-\mu_2+1)!(\lambda_2-\mu_2)!(\mu_1-\nu)!]^{1/2}$$
$$\times[(\lambda_1-\lambda_2)!(\lambda_1-\lambda_3+1)!(\lambda_2-\lambda_3)!(\mu_1-\lambda_2)!(\mu_1-\lambda_3+1)!(\mu_2-\lambda_3)!(\nu-\mu_2)!]^{-1/2}. \tag{A8}$$

For U(2), from Eqs. (A3) and (I 5.9) we obtain the relation

$$(R_1^2)^\alpha 1 = (\lambda_1-\lambda_2)![(\lambda_1-\lambda_2-\alpha)!]^{-1}z^\alpha, \tag{A9}$$

which can be proved by induction over α. Then Eq. (3.10a) results from the combination of Eq. (A9) with Eqs. (A5) and (A7).

Proceeding through the same way for U(3), from Eqs. (A3), (A4), and (I 5.10) we get the following results

$$(R_2^3)^\alpha 1 = (-1)^\alpha(\lambda_1-\lambda_3+1)!(\lambda_2-\lambda_3)![(\lambda_1-\lambda_3-\alpha+1)!(\lambda_2-\lambda_3-\alpha)!]^{-1}(x_2-x_1x_3)^\alpha, \tag{A10}$$

$$(R_1^3)^\beta(x_2-x_1x_3)^\alpha = (\lambda_1-\lambda_2)![(\lambda_1-\lambda_2-\beta)!]^{-1}x_1^\beta(x_2-x_1x_3)^\alpha, \tag{A11}$$

and

$$(R_1^2)^\gamma x_1^\beta(x_2-x_1x_3)^\alpha = \gamma!\sum_k \binom{\lambda_2-\lambda_3+\beta-\alpha-k}{\lambda_2-\lambda_3+\beta-\alpha-\gamma}\binom{\beta}{k}x_1^{\beta-k}x_3^{\gamma-k}(x_2-x_1x_3)^{\alpha+k}. \tag{A12}$$

Their combination with Eqs. (A6) and (A8) finally leads to Eqs. (3.11) and (3.12).

APPENDIX B: REALIZATION OF THE U(2) COHERENT STATE REPRESENTATION IN A BARGMANN SPACE OF ANALYTIC FUNCTIONS

The purpose of this appendix is to detail, for the $d=2$ case, the realization of the U(d) CS representation in a Bargmann space of analytic functions as introduced in Sec. IV.

For U(2), the normalized lowest weight state function $\psi_{\lambda_2}(\overline{\mathbf{w}})$ is given by

$$\psi_{\lambda_2}(\overline{\mathbf{w}}) = (\lambda_1-\lambda_2+1)^{1/2}[(\lambda_1+1)!\lambda_2!]^{-1/2}\times(\overline{w}_{21})^{\lambda_1-\lambda_2}(\overline{w}_{12,12})^{\lambda_2}. \tag{B1}$$

From the latter, any function

$$\psi_\mu(\overline{\mathbf{w}}) = \psi_{\lambda_2}(\overline{\mathbf{w}})Z_\mu\left(\frac{\overline{w}_{11}}{\overline{w}_{21}}\right) \tag{B2}$$

can be obtained by applying the relation

$$\psi_\mu(\overline{\mathbf{w}}) = \mathscr{N}_{\mu}^{\lambda_1\lambda_2}(E_{12})^{\mu-\lambda_2}\psi_{\lambda_2}(\overline{\mathbf{w}}), \tag{B3}$$

where E_{12} and $\mathscr{N}_{\mu}^{\lambda_1\lambda_2}$ are defined in Eqs. (4.8) and (A7), respectively. The result reads

$$\psi_\mu(\overline{\mathbf{w}}) = [(\lambda_1-\lambda_2+1)!]^{1/2}\times[(\lambda_1+1)!\lambda_2!(\lambda_1-\mu)!(\mu-\lambda_2)!]^{-1/2}\times(\overline{w}_{11})^{\mu-\lambda_2}(\overline{w}_{21})^{\lambda_1-\mu}(\overline{w}_{12,12})^{\lambda_2}. \tag{B4}$$

By making the transformation

$$x_{11} = \overline{w}_{12,12},\quad x_{12} = \overline{w}_{12},\quad x_{21} = \frac{\overline{w}_{11}}{\overline{w}_{21}},\quad x_{22} = \overline{w}_{21}, \tag{B5}$$

we obtain

$$\psi_{\lambda_2}(x_{11},x_{22}) = (\lambda_1-\lambda_2+1)^{1/2}[(\lambda_1+1)!\lambda_2!]^{-1/2}\times(x_{22})^{\lambda_1-\lambda_2}(x_{11})^{\lambda_2}, \tag{B6}$$

and

$$Z_\mu(x_{21}) = [(\lambda_1-\lambda_2)!]^{1/2}\times[(\lambda_1-\mu)!(\mu-\lambda_2)!]^{-1/2}(x_{21})^{\mu-\lambda_2}. \tag{B7}$$

The comparison of Eq. (3.10a) with Eq. (B7), where we set $z=x_{21}$, leads to the conclusion that

$$Z_\mu(z) = \hat{\phi}_\mu(z). \tag{B8}$$

From Eq. (B5) and the fact that $\psi_\mu(\mathbf{x})$ does not depend on x_{12}, it results that

$$\begin{aligned}
\frac{\partial}{\partial\overline{w}_{11}}\psi_\mu(\mathbf{x}) &= \left(\overline{w}_{22}\frac{\partial}{\partial x_{11}}+\frac{1}{\overline{w}_{21}}\frac{\partial}{\partial x_{21}}\right)\psi_\mu(\mathbf{x}),\\
\frac{\partial}{\partial\overline{w}_{12}}\psi_\mu(\mathbf{x}) &= -\overline{w}_{21}\frac{\partial}{\partial x_{11}}\psi_\mu(\mathbf{x}),\\
\frac{\partial}{\partial\overline{w}_{21}}\psi_\mu(\mathbf{x}) &= \left(-\overline{w}_{12}\frac{\partial}{\partial x_{11}}-\frac{\overline{w}_{11}}{\overline{w}_{21}^2}\frac{\partial}{\partial x_{21}}+\frac{\partial}{\partial x_{22}}\right)\psi_\mu(\mathbf{x}),\\
\frac{\partial}{\partial\overline{w}_{22}}\psi_\mu(\mathbf{x}) &= \overline{w}_{11}\frac{\partial}{\partial x_{11}}\psi_\mu(\mathbf{x}).
\end{aligned} \tag{B9}$$

Let us now introduce Eq. (B9) into the left-hand side of Eq. (4.16), and take into account that

$$\begin{aligned}
x_{11}\frac{\partial}{\partial x_{11}}\psi_\mu(\mathbf{x}) &= \lambda_2\,\psi_\mu(\mathbf{x}),\\
x_{22}\frac{\partial}{\partial x_{22}}\psi_\mu(\mathbf{x}) &= (\lambda_1-\lambda_2)\psi_\mu(\mathbf{x}).
\end{aligned} \tag{B10}$$

For E_{11} for instance, we obtain

$$E_{11}\psi_\mu(\mathbf{x}) = \left(\overline{w}_{11}\frac{\partial}{\partial\overline{w}_{11}} + \overline{w}_{12}\frac{\partial}{\partial\overline{w}_{12}} + \frac{n}{2}\right)\psi_\mu(\mathbf{x})$$
$$= \left(x_{11}\frac{\partial}{\partial x_{11}} + x_{21}\frac{\partial}{\partial x_{21}} + \frac{n}{2}\right)\psi_\mu(\mathbf{x})$$
$$= \psi_{\lambda_2}(\mathbf{x})(z\,\partial + \lambda_2 + n/2)\hat{\phi}_\mu(z), \tag{B11}$$

where we set again $z = x_{21}$ and $\partial = \partial/\partial x_{21}$. Hence, by comparing with the right-hand side of Eq. (4.16), we get

$$\mathring{\mathscr{E}}_{11} = z\partial + \lambda_2 + n/2. \tag{B12}$$

For the remaining generators, the results read

$$\mathring{\mathscr{E}}_{22} = -z\,\partial + \lambda_1 + n/2, \quad \mathring{\mathscr{E}}_{12} = z(\lambda_1 - \lambda_2 - z\partial),$$
$$\mathring{\mathscr{E}}_{21} = \partial, \tag{B13}$$

in accordance with Eq. (I 5.9).

For the determination of the U(2) CS measure from Bargmann one, we first note that the inverse of transformation (B5) and the corresponding Jacobian are

$$\overline{w}_{11} = x_{21}x_{22}, \quad \overline{w}_{12} = x_{12},$$
$$\overline{w}_{21} = x_{22}, \quad \overline{w}_{22} = (x_{21}x_{22})^{-1}(x_{11} + x_{12}x_{22}), \tag{B14}$$

and

$$\frac{\partial(\overline{\mathbf{w}},\overline{\mathbf{w}}^*)}{\partial(\mathbf{x},\mathbf{x}^*)} = |x_{21}|^{-2}, \tag{B15}$$

respectively. Then the left-hand side of Eq. (4.19) becomes

$$\int d\mu(\overline{\mathbf{w}})\psi_{\mu'}(\overline{\mathbf{w}}^*)\psi_\mu(\overline{\mathbf{w}})$$
$$= \pi^{-4}(M^{[\lambda]})^2\int dx_{21}\,dx_{21}^*\,|x_{21}|^{-2}Z_{\mu'}(x_{21}^*)Z_\mu(x_{21})$$
$$\times\int dx_{22}\,dx_{22}^*$$
$$\times\exp[-(1+|x_{21}|^2)|x_{22}|^2]\,|x_{22}|^{2(\lambda_1-\lambda_2)}$$
$$\times\int dx_{11}\,dx_{11}^*\,\exp(-|x_{21}x_{22}|^{-2}|x_{11}|^2)\,|x_{11}|^{2\lambda_2}$$
$$\times\int dx_{12}\,dx_{12}^*\,\exp\Bigg\{-|x_{21}|^{-2}$$
$$\times\left[(1+|x_{21}|^2)|x_{12}|^2 + \frac{x_{11}^*}{x_{22}^*}x_{12} + \frac{x_{11}}{x_{22}}x_{12}^*\right]\Bigg\}. \tag{B16}$$

It is straightforward to perform the integrations over x_{12}, x_{11}, and x_{22}. The result reads

$$\int d\mu(\overline{\mathbf{w}})\psi_{\mu'}(\overline{\mathbf{w}}^*)\psi_\mu(\overline{\mathbf{w}})$$
$$= \pi^{-1}(\lambda_1 - \lambda_2 + 1)\int dz\,dz^*$$
$$\times(1+|z|^2)^{-(\lambda_1-\lambda_2+2)}\hat{\phi}_{\mu'}(\mathbf{z}^*)\hat{\phi}_\mu(\mathbf{z}), \tag{B17}$$

in accordance with Eq. (4.19), when Eqs. (I 4.24), (I 6.11), (I 6.17), and (I 6.21) are taken into account.

[1]M. Moshinsky and C. Quesne, J. Math. Phys. 12, 1772 (1971).
[2]G. Rosensteel and D. J. Rowe, Int. J. Theor. Phys. 16, 63 (1977).
[3]A. U. Klimyk, J. Math. Phys. 24, 224 (1983).
[4]C. Quesne, J. Math. Phys. 27, 428 (1986).
[5]J. R.Klauder, J. Math. Phys. 4, 1055, 1058 (1963); 5, 177 (1964).
[6]A. M. Perelomov, Commun. Math. Phys. 26, 222 (1972); Usp. Fiz. Nauk 123, 23 (1977) [Sov. Phys. Usp. 20, 703 (1977)].
[7]R. Gilmore, Ann. Phys. NY 74, 391 (1972); Rev. Mex. Fis. 23, 143 (1974).
[8]A. O. Barut and L. Girardello, Commun. Math. Phys. 21, 41 (1971).
[9]J. Deenen and C. Quesne, J. Math. Phys. 25, 1638 (1984).
[10]Reference 9 actually deals with the case where $\lambda_1 = \cdots = \lambda_d = 0$, but it is straightforward to extend the results to the case where $\lambda_1 = \cdots = \lambda_d = \lambda \neq 0$ by replacing n by $n + 2\lambda$ in all formulas.
[11]J. Deenen and C. Quesne, J. Math. Phys. 25, 2354 (1984).
[12]I. M. Gel'fand and M. L. Tseitlin, Dokl. Akad. Nauk SSSR 71, 825 (1950).
[13]M. Moshinsky, J. Math. Phys. 4, 1128 (1963).
[14]G. E. Baird and L. C. Biedenharn, J. Math. Phys. 4, 1449 (1963).
[15]V. Bargmann, Commun. Pure Appl. Math. 14, 187 (1961).
[16]M. Moshinsky and C. Quesne, J. Math. Phys. 11, 1631 (1970).
[17]In Refs. 13 and 18, the basis states of U(d) irreps are constructed by starting from the highest weight state. Here we use instead the lowest weight one because it has been chosen as the reference state to define CS.
[18]J. Deenen and C. Quesne, J. Phys. A: Math. Gen. 16, 2095 (1983). Here the variables g_{is}, $\overline{w}_{i\alpha}$, and $\overline{w}_{i\rho}$ are denoted by z_{is}, $u_{i\alpha}$, and $v_{i\rho}$, respectively.
[19]M. K. F. Wong, J. Math. Phys. 8, 1899 (1967).
[20]J. Deenen and C. Quesne, "Group Theoretical Methods in Physics," in *Lecture Notes in Physics*, Vol. 180, edited by M. Serdaroğlu and E. Inönü (Springer, Berlin, 1983), p. 444.
[21]O. Castaños, E. Chacón, M. Moshinsky, and C. Quesne, J. Math. Phys. 26, 2107 (1985).
[22]O. Castaños, P. Kramer, and M. Moshinsky, J. Phys. A: Math. Gen. 18, 493 (1985).
[23]J. G. Nagel and M. Moshinsky, J. Math. Phys. 6, 682 (1965).