Determination of basis for the irreducible representations of the unitary group for $U(p+q) \downarrow U(p) \otimes U(q)$

C. R. Sarma
Department of Physics, Indian Institute of Technology, Bombay, India

(Received 16 July 1980; accepted for publication 24 April 1981)

A direct procedure is outlined for determining the basis spanning finite dimensional irreducible representations of $U(p+q)$ adapted to the subgroup $U(p) \otimes U(q)$. Using a tableau based analysis, it is shown that the realization of the semimaximal states follows readily from a knowledge of the matrix elements of the generators E_{ii+1} of $U(p)$, $U(q) \subset U(p+q)$.

PACS numbers: 02.20. – a

I. INTRODUCTION

The use of the canonical basis spanning finite dimensional irreducible representations (irreps) of the unitary group has been well established in the study of many-body problems. Based on extensive studies by Moshinsky,[1] Baird and Biedenharn,[2] and others,[3,4] considerable notational and other simplifications have been developed[5,6] leading to large scale studies of many-particle systems using computers.[7,8]

In many applications, however, it is necessary to choose a noncanonical basis for the unitary group $U(p+q)$ adapted to the subgroup $U(p) \otimes U(q)$. The usefulness of such a basis has been discussed in recent years by a number of workers.[9–11] Details of the realization of such a basis using lowering operators which are polynomials of the generators of the unitary group was considered recently by Mickelsson.[9] Though rules have been given for the construction of these operators, their explicit realization and actual construction of the required basis using them is not easy.

Alternatively, the S-function method of Littlewood,[12] as elaborated by Wybourne,[13] could also be used to generate the basis adapted to the restriction $U(p+q) \downarrow U(p) \otimes U(q)$. In this approach there exist simple tableau based rules for obtaining the irreps of $U(p) \otimes U(q)$, which occur in the above restriction. A determination of the required basis states spanning a given product representation of the subgroup is, however, not quite straightforward.

In the present paper we have used the tableau method to obtain the subduction series occurring in the restriction of the group to the subgroup $U(p) \otimes U(q)$. The series is obtained as was done by Wybourne[13] and Robinson,[14] by reducing the skew representation $[\lambda] - [\mu]$, where $[\lambda]$ is an irrep of $U(p+q)$ and $[\mu]$ is an irrep of $U(q)$ contained entirely within $[\lambda]$ and realized using a tensor product of the first p orbitals of the fundamental representation of the group. For each of the product irreps occuring in the subduction series the semimaximal states have been determined. The semimaximality condition[9] has been used to determine the subgroup adapted basis in terms of the set spanning $[\lambda]$. Knowing the matrix elements of the generators of $U(p)$ and $U(q)$, this basis set has been determined to within a normalization factor and ambiguity due to the multiplicity of occurrence of $[\nu]$ in $[\lambda] - [\mu]$.

The procedure has been outlined in Sec. 2 and a brief discussion presented in Sec. 3.

2. THE RESTRICTION $U(p+q) \downarrow U(p) \otimes U(q)$

Consider an ordered set of orthonormal single particle basis orbitals $\{\phi_i | i = 1,2,..., p, p+1,..., p+q\}$ spanning the fundamental representation space V_{p+q} of the unitary group $U(p+q)$. Let

$$[\lambda] \equiv [\lambda_1, \lambda_2,..., \lambda_{p+q}], \lambda_1 \geqslant \lambda_2 \geqslant ... \geqslant \lambda_{p+q} \geqslant 0, \quad (1)$$
$$\sum_{i=1}^{p+q} \lambda_i = N,$$

be an irrep occurring in the reduction of the tensor basis spanning $V_{p+q} \otimes^N$, where N is the number of particles. The basis spanning $V_{p+q} \otimes^N$ may be represented as a tensor of rank N in terms of single particle orbital occupancies, as

$$(N_1, N_2,..., N_{p+q}) \equiv \phi_1^{N_1} \phi_2^{N_2} ... \phi_{p+q}^{N_{p+q}}, \quad (2)$$

where

$$\sum_{i+1}^{p+q} N_i = N, \quad N_1, N_2,..., N_{p+q} \geqslant 0. \quad (3)$$

Using nonstandard Wigner operators of the permutation group S_N it is quite straightforward to generate a canonical basis spanning the irrep $[\lambda]$ using the reducible basis of Eq. (2) and the detailed procedures outlined in an earlier note.[15] As was shown in that note[15] such a basis transforms as the canonical Gel'fand or Weyl basis under the action of the generators $\{E_{ij} | i,j = 1,2,...,p,p+1,..., p+q\}$ of $U(p+q)$. Using the Weyl tableau to represent the basis, we observe that the first p indices define a standard subtableau structure of λ corresponding to a $[\mu]$ of $U(p)$ obtained in the reduction of $V_p \otimes^{N'}$, where

$$N' + N'' = N, \quad \sum_{i+1}^{p} N_i = N', \quad \sum_{i=p+1}^{p+q} N_i = N''. \quad (4)$$

In terms of the branching rules for $U(p+q)$,[16] this essentially means that if we delete N'' boxes successively from the given Weyl tableau for $[\lambda]$ such that no two of the N_{p+1} or N_{p+2} or ... N_{p+q} boxes share a column, the residue still defines a standard Weyl tableau for the irrep $[\mu]$ of $U(p)$. The deleted portion $[\lambda] - [\mu]$, which is a "skew tableau" in the sense defined by Robinson,[14] forms a reducible representation for the subgroup $U(q)$, whose fundamental representation space is spanned by thge last q orbitals. Given a $[\lambda]$ of $U(p+q)$ and $[\mu]$ of $U(p)$, contained entirely within it, the rules obtaining the possible irreps of $U(p) \otimes U(q)$ occurring in the reduction of $[\mu] \otimes ([\lambda] - [\mu])$ can be stated as[13]

 0022-2488/82/071235-04$02.50

(i) replace the Young tableau box structure for $[\lambda]$ by corresponding dots; (ii) leaving the subtableau structure corresponding to a given $[\mu]$ unaltered, fill the residual portion of (i) with a set of a's in all possible ways so that no two of them share a column in the tableau; (iii) assign sets of b's, c's, etc., as in (ii), ensuring at each stage that the letters a, b, c, etc., define a lattice permutation read from left to right along each row taken from top to bottom.

Exhausting the skew portion of the tableau $[\lambda] - [\mu]$ as above, we obtain a set of $[\lambda]\downarrow[\mu] \otimes [\nu]$ by arranging the set of a's, b's, etc., as defining the rows of tableaus for irreps of U(q). As an illustration, consider $[4,3,2]\downarrow[2,1] \otimes ([4,3,2] - [2,1])$ of $U(3) \otimes U(3) \subset U(6)$. Using the rules above, we readily obtain

$$[4,3,2]\downarrow \quad \otimes \left(\begin{array}{cccccc} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & - & \cdot & \\ \cdot & \cdot & & & & \end{array} \right)$$

$$\begin{array}{llllllllllllllll} & \cdot & \cdot & a & a & & \cdot & \cdot & a & a & & \cdot & \cdot & a & a & \\ = & \cdot & a & b & & + & \cdot & a & b & & + & \cdot & a & b & & \\ & a & b & & & & a & c & & & & b & b & & & \\ & & & & & & & & & & & & & & & \\ & \cdot & \cdot & a & a & & \cdot & \cdot & a & a & & \cdot & \cdot & a & a & \\ + & \cdot & a & b & & + & \cdot & b & b & & + & \cdot & b & b & , & \\ & b & c & & & & a & c & & & & c & c & & & \end{array}$$

so that we obtain the possible representations subduced from [4,3,2] as

$$\begin{aligned} &[4,3,2]\downarrow[2,1] \otimes ([4,3,2] - [2,1]) \\ &\quad = [2,1] \otimes ([4,2] + [4,1^2] + [3^2] \\ &\qquad + 2\,[3,2,1] + [2^3]). \end{aligned} \tag{5}$$

Thus, given a fixed $[\lambda]$ of U($p+q$) and a fixed $[\mu]$ of U(p) contained entirely in $[\lambda]$, it is easy to obtain the subduction series

$$[\lambda]\downarrow[\mu] \otimes ([\lambda] - [\mu]) = \sum_{[\nu]} \Gamma_{\lambda\mu\nu}\,[\mu] \times [\nu], \tag{6}$$

where $\Gamma_{\lambda\mu\nu}$ is the multiplicity of occurrence of $[\nu]$ in $[\lambda] - [\mu]$.

For convenience, we will assume that the irreps $[\nu]$ on the right side of Eq. (6) are ordered such that for $i < j$, the irrep $[\nu^{(i)}]$ has a greater row symmetry in the corresponding tableau than $[\nu^{(j)}]$. In terms of the partitions defining the irreps this implies that the first nonzero difference $[\nu_1^{(i)} - \nu_1^{(j)}, \nu_2^{(i)} - \nu_2^{(j)}, \ldots]$ is positive where the suffixes define the partitions of the irreps in the usual sense [cf., Eq. (1)].

Let us now consider the procedure for generating the basis states spanning the product representations subduced as in Eq. (6). Let $|\mu;l\rangle$ and $|\nu;k\rangle$ be the canonical Weyl basis spanning the irreps of U(p) and U(q) occurring in the restriction $[\lambda]\downarrow[\mu] \otimes [\nu]$ of U($p+q$). Corresponding to a particular $[\lambda]\downarrow[\mu] \otimes [\nu]$ occurring in the reduction of $[\mu] \otimes ([\lambda] - [\mu])$ defined by Eq. (6), we observe that the basis states for the subduction can be expressed as

$$[\lambda]\downarrow|\mu;l\rangle|\nu;k\tau\rangle = \sum_j \langle \lambda j|\,\mu l;\nu k\tau\rangle\,|\lambda;j\rangle, \tag{7}$$

where $|\lambda;j\rangle$ are the canonical Weyl bases spanning the irrep $[\lambda]$ of U($p+q$) and the subduction coefficients can be chosen to be elements of a real orthogonal matrix satisfying[16]

$$\sum_j \langle \lambda j|\,\mu l;\nu k\tau\rangle\,\langle\,\mu l';\nu k'\tau'|\lambda j\rangle = \delta_{ll'}\cdot\delta_{kk'}\cdot\delta_{\tau\tau'} \tag{8}$$

and

$$\sum_{k,\tau} \langle \lambda' j'|\,\mu l;\nu k\tau\rangle\,\langle\,\mu l;\nu k\tau|\lambda j\rangle = \delta_{\lambda\lambda'}\cdot\delta_{jj'}. \tag{9}$$

In Eqs. (7)–(9), the index τ has been introduced to distinguish between the states spanning multiply occurring $[\nu] \in [\lambda] - [\mu]$ [i.e., $\Gamma_{\lambda\mu\nu} > 1$ on the right of Eq. (6)].

As a first step in determining the linear combination occurring on the right side of Eq. (7), consider the semimaximal states for the irreps defined by Eq. (6). The semimaximal states (s.m.) can be defined using the weight raising generators E_{ij} $(i<j = 1,2,\ldots,p,\ p+1,\ldots,p+q)$ of the unitary group U($p+q$) as[9]

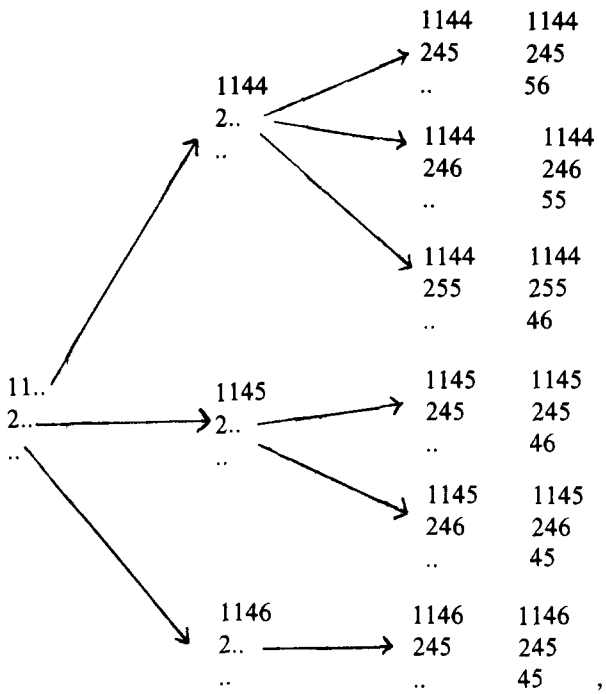
$$E_{ij}|(\text{s.m.})\rangle = 0 \tag{10}$$

for $i<j = 1,2,\ldots,p$ and $i<j = p+1,\ldots,p+q$.

The semimaximality condition defined by Eq. (10) can also be restated as requiring that the component states spanning $[\mu]$ and $[\nu]$ of Eq. (7) be of maximal weight. In terms of the partitions $[\mu_1, \mu_2, \ldots, \mu_p]$ and $[\nu_1, \nu_2, \ldots, \nu_q]$ of $[\mu]$ and $[\nu]$, respectively, defined as in Eq. (1) for the subgroups U(p) and U(q), we observe that the orbital occupancy indices of Eq. (2) take the values

$$\{N_k^{(\mu)} = \mu_k \,|\, k = 1,\ldots,p\} \text{ and } \{N_{p+l}^{(\nu)} = \nu_l \,|\, l = 1,\ldots,q\}. \tag{11}$$

Let $|\mu;1\rangle$ and $|\nu;1\rangle$ denote, respectively, the maximal weight states of $[\mu]$ and $[\nu]$ with occupancy indices as in Eq. (11). For nonzero subduction coefficients to exist on the right side of Eq. (7), we require that the same numbers of indices $1,\ldots,p,\ldots,p+q$ be assignable in all possible ways to the tableau structure of $[\lambda]$, so that the result is a standard Weyl tableau. The assignment of indices $1,\ldots,p$ in numbers equal to that given by Eq. (11) to the $[\mu]$ substructure of $[\lambda]$ is straightforward, since the former defines a standard subtableau of the latter. The assignment of $p+1,\ldots,p+q$ in numbers, as given by Eq. (11) to the skew portion $[\lambda] - [\mu]$, is also straightforward and follows from rules (i), (ii), and a modified (iii), which led to Eq. (6). Rule (iii) given there is modified by the statement that the indices $p+1,\ldots,p+q$ be assigned to the skew portion $[\lambda] - [\mu]$ of $[\lambda]$ in all possible ways consistent with Eq. (11), so that the resulting structure defines a standard Weyl tableau for $[\lambda]$. This can be readily illustrated for the restriction $[4,3,2]\downarrow[2,1] \otimes [3,2,1]$ of $U(6)\downarrow U(3) \otimes U(3)$. The self-explanatory branching diagram leading to allowed standard Weyl tableaus having occupancies $N_1 = 2$, $N_2 = 1$, $N_3 = 0$, $N_4 = 3$, $N_5 = 2$, $N_6 = 1$ is



where the branches leading to nonstandard tableaus have been omitted. This result indicates that only a subset of $|\lambda;j\rangle$ leads to nonvanishing coefficients $\langle\lambda j|\,\mu 1;\nu 1\tau\rangle$ on the right side of Eq. (7).

Once a subset has been obtained as above, the semimaximality condition of Eq. (11) can be used on the right side of Eq. (7) to eliminate contributions from lower weight states of higher row symmetry irreps occurring on the right of Eq. (6). In doing this, it is only necessary to apply E_{ii+1} $(i=p+1, p+2,...,p+q-1)$ to both sides of Eq. (7) and equate the result on the right side to zero, since the left side already gives zero. This need to use only E_{ii+1} instead of all $E_{ij}(i<j)$ follows from the commutation relations satisfied by E_{ij},

$$E_{ij}=[E_{ij-1},E_{j-1j}]=E_{ij-1}-E_{j-1j}E_{ij-1}, \tag{12}$$

which lead to

$$\begin{aligned}E_{ij}|(\text{s.m.})\rangle &= -E_{j-1j}E_{ij-1}|(\text{s.m.})|\\ &=...\\ &=(-)^{j-i}E_{j-1j}E_{j-2j-1}\cdots E_{ii+1}|(\text{s.m.})\rangle=0.\end{aligned} \tag{13}$$

Equations (12) and (13) imply that only $q-1$ independent equations,

$$E_{ii+1}\sum_j\langle\lambda j|\,\mu 1;\nu 1\tau\rangle|\lambda;j\rangle=0$$
$$\text{for } i=p+1,...,+p+q, \tag{14}$$

can be used to eliminate the contributions from lower weight states of higher symmetry irreps of Eq. (6). The final combination occurring on the right of Eq. (7) after this elimination can only define a semimaximal state for the given restriction $[\lambda]\downarrow[\mu]\otimes[\nu]$ and does not have any contribution from the lower symmetry irreps $[\nu]\in[\lambda]-[\mu]$ occurring in Eq. (6). This again follows since, by definition, at least one row of the tableau corresponding to $[\nu']$ is of less length than a corresponding one for $[\nu]$. This forces the corresponding set of indices with occupancies as in Eq. (11) to share a column in the skew portion $[\nu']\in[\lambda]-[\mu]$. The final set of essential unknowns occurring on the right side of Eq. (7) is equal to the number of times $[\nu]$ occurs in $[\lambda]-[\mu]$.

As an illustration consider again $[4,3,2]\downarrow[2,1]\otimes[3,2,1]$ of $U(6)\downarrow U(3)\otimes U(3)$. For notational convenience we label the states occurring in the final column of the branching diagram for this example discussed earlier as ψ_i $(i=1,...,6)$, read from top to bottom. We also replace the corresponding subduction coefficients on the right side of Eq. (7) by a_i $(i=1,...,6)$ and require the determination of these unknowns for

$$[4,3,2]\downarrow\left|\begin{matrix}1 & 1\\ 2\end{matrix}\right\rangle\times\left|\begin{matrix}4 & 4 & 4\\ 5 & 5\\ 6\end{matrix}\right\rangle=\sum_{i=1}^{6}a_i\psi_i.$$

Applying the generators E_{45} and E_{56} to both sides of the above expression, noting that the left side yields zero in either case and using the simple expressions for the matrix elements of these generators for the basis on the right of the above equation, we obtain the results

$$\left[\frac{\sqrt{2}}{\sqrt{5}}a_1+a_3+\frac{2\sqrt{2}}{\sqrt{5}}a_4\right]\left|\begin{matrix}1 & 1 & 4 & 4\\ 2 & 4 & 5\\ 4 & 6\end{matrix}\right\rangle$$
$$+\left[\frac{\sqrt{3}}{\sqrt{5}}a_2+\frac{2\sqrt{3}}{\sqrt{5}}a_5\right]\left|\begin{matrix}1 & 1 & 4 & 4\\ 2 & 4 & 6\\ 4 & 5\end{matrix}\right\rangle=0$$

and

$$[a_1+\sqrt{2}a_2]\left|\begin{matrix}1 & 1 & 4 & 4\\ 2 & 4 & 5\\ 5 & 5\end{matrix}\right\rangle$$
$$+\left[\frac{\sqrt{3}}{2\sqrt{2}}a_4+\frac{\sqrt{3}}{2\sqrt{2}}a_4+\frac{\sqrt{3}}{2}a_5\right.$$
$$\left.+\frac{\sqrt{15}}{2\sqrt{2}}a_6\right]\left|\begin{matrix}1 & 1 & 4 & 5\\ 2 & 4 & 5\\ 4 & 5\end{matrix}\right\rangle=0.$$

Since the basis states of [4,3,2] occurring on the left side of the above equations form an orthonormal set, we obtain four defining equations for the six unknowns $a_1,...,a_6$. Using these equations to eliminate four of them, we get

$$[4,3,2]\downarrow\left|\begin{matrix}1 & 1\\ 2\end{matrix}\right\rangle\times\left|\begin{matrix}4 & 4 & 4\\ 5 & 5\\ 6\end{matrix}\right\rangle$$
$$=a_4\left[-\frac{2\sqrt{2}}{\sqrt{5}}\psi_3+\psi_4-\frac{1}{\sqrt{5}}\psi_6\right]$$
$$+a_5\left[(2\sqrt{2})\psi_1-2\psi_2-\frac{4}{\sqrt{5}}\psi_3+\psi_5-\frac{\sqrt{2}}{\sqrt{5}}\psi_6\right].$$

Since the semimaximality condition holds for all arbitrary values a_4, a_5, we can choose linearly independent semimaximal states for this restriction, as

$$[4,3,2]\downarrow\left|\begin{matrix}1 & 1\\ 2\end{matrix}\right\rangle\times\left|\begin{matrix}4 & 4 & 4\\ 5 & 5 & ;1\\ 6\end{matrix}\right\rangle$$
$$=\frac{1}{\sqrt{14}}[(2\sqrt{2})\psi_3-(\sqrt{5})\psi_4+\psi_6]$$

and

$$[4,3,2]\downarrow \left|\begin{matrix}1 & 1\\ 2\end{matrix}\right\rangle \times \left|\begin{matrix}4 & 4 & 4\\ 5 & 5 & ;2\\ 6\end{matrix}\right\rangle$$

$$= \frac{1}{\sqrt{83}}[(2\sqrt{10})\psi_1 - (2\sqrt{5})\psi_2 - 4\psi_3 + (\sqrt{5})\psi_5 - (\sqrt{2})\psi_6].$$

An orthonormal pair of states may be generated from these using the Schmidt orthogonalization procedure. We observe that the occurrence of the two basis states for the restriction coincides with the twofold multiplicity of occurrence of [3,2,1] in [4,3,2] − [2,1], as illustrated in Eq. (5).

3. SUMMARY AND DISCUSSION

The procedure outlined in Sec. 2 for generating the subduced representations of the unitary group consists basically of the following. Given $[\lambda]$ of $U(p+q)$ with definite substructure $[\mu]$ of $U(p)$ we first obtain the subduction series as in Eq. (6) using rules (i)–(iii) presented there. For any $[\mu]\otimes[\nu]$ occurring in the reduction of $[\mu]\otimes([\lambda]-[\mu])$ we generate the maximal weight states for the component representations having the occupancies as in Eq. (11). Assigning these orbital indices to the tableau corresponding to $[\lambda]$ the required Weyl tableaus are generated. Using these to define a linear combination as in Eq. (7), the generators $E_{i\,i+1}$ $(i=p+1,...,p+q)$ are applied to it and the result equated to zero. This leads to defining relations among the subduction coefficients and makes it possible to eliminate some of them. The number of remaining coefficients is equal to the multiplicity index $\Gamma_{\lambda\mu\nu}$. If $\Gamma_{\lambda\mu\nu}=1$, only one unknown is left which can be fixed by normalization. If, on the other hand, $\Gamma_{\lambda\mu\nu}>1$ we find that a set of unique semimaximal states cannot be obtained. In this case a method similar to the one used by Patera and Sharp[17] and Harter and Patterson[18] in their studies of angular momentum projected states has been employed to obtain linearly independent semimaximal states. These can then be Schmidt orthogonalized. Once the semimaximal states have been realized the others follow on using the lowering generators $E_{ij}(i>j=1,...,p$ and $i>j=p+1,...,p+q)$ of the subgroup $U(p)\otimes U(q)\subset U(p+q)$.

Since a program has been developed recently for generating the canonical basis states for the finite dimensional irreps of $U(n)$[19] corresponding to a specified occupancy index set and for determining the matrix elements of the generators[8] of the group between any two of these, the task of using the present procedure is considerably simplified.

[1] M. Moshinsky, *Group Theory and the Many-Body Problem* (Gordon and Breach, London, 1968).
[2] E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 463 (1963).
[3] J. D. Louck, Am. J. Phys. **38**, 3 (1970).
[4] M. Ciftan, J. Math. Phys. **10**, 1635 (1969).
[5] J. Paldus, Phys. Rev. A **14**, 1620 (1976); Int. J. Quantum Chem. Symp. **9**, 165 (1976).
[6] C. R. Sarma and S. Rettrup, Theoret. Chim. Acta **46**, 63, 73 (1977).
[7] B. R. Brooks and H. F. Schaefer III, J. Chem. Phys. **70**, 5092 (1979).
[8] C. R. Sarma and S. Rettrup, J. Phys. A (London) **13**, 2267 (1980).
[9] J. Mickelsson, J. Math. Phys. **12**, 2378 (1971).
[10] I. G. Kaplan, Int. J. Quantum Chem. **10**, 699 (1976).
[11] A. U. Klimyk and B. Gruber, J. Math. Phys. **20**, 1995, 2011 (1979).
[12] D. E. Littlewood, *Theory of Group Characters and Matrix Representations of Groups* (Oxford University, Oxford, 1958).
[13] B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).
[14] G. de B. Robinson, *Representation Theory of the Symmetric Group* (University of Toronto, Toronto, Ontario, 1961).
[15] C. R. Sarma and G. G. Sahasrabudhe, J. Math. Phys. **21**, 638 (1980).
[16] C. W. Patterson and W. G. Harter, Phys. Rev. A **15**, 2372 (1977).
[17] J. Patera and R. T. Sharp, Nuovo Cimento A **12**, 365 (1972).
[18] W. G. Harter and C. W. Patterson, *A Unitary Calculus for Electronic Orbitals* (Springer, Heidelberg, 1976).
[19] S. Rettrup and C. R. Sarma, Phys. Lett. A **75**, 184 (1980).

Branching rules for the supergroup SU(N/M) from those of SU($N+M$) [a)]

A. Baha Balantekin
Wright Nuclear Structure Laboratory, Physics Department, Yale University, New Haven, Connecticut 06520

Itzhak Bars
J. W. Gibbs Laboratory, Physics Department, Yale University, New Haven, Connecticut 06520

(Received 29 October 1981; accepted for publication 26 February 1982)

The decomposition of representations of supergroups into representations of subgroups is needed in practical applications. In this paper we set up and exploit a fruitful one-to-one correspondence between the Lie group branching SU $(N+M)\supset$SU$(N)\otimes$SU$(M)\otimes$U(1) and the supergroup branchings SU$(N/M)\supset$SU$(N)\otimes$SU$(M)\otimes$U(1) and SU$(N_1+N_2/M_1+M_2)\supset$SU$(N_1/M_1)\otimes$SU$(N_2/M_2)\otimes$U(1). A simple and useful prescription is discovered for obtaining the SU(N/M) branching rules from those of SU$(N+M)$ for any representation. A large class of examples, sufficient for many physical applications we can foresee, are explicitly worked out and tabulated.

PACS numbers: 02.20 + b, 11.30.Pb, 21.60.Fw

I. INTRODUCTION

Superalgebras and supergroups are relevant to mixed systems of bosons and fermions.[1] They were introduced into physics in the context of dual models[2] and were subsequently used in supersymmetric field theories[3] and supergravity.[4] The first experimental evidence of a supersymmetric phenomenon in Nature has recently come from nuclear physics: A scheme based on the supergroup U(6/4) has been suggested to describe many properties of nuclei in the Os–Pt region.[5] Also, lately the representations of the supergroup SU(N/M) with a vanishing eigenvalue of the cubic Casimir operator have been used to construct models of composite quarks and leptons.[6]

Classification aspects of superalgebras have been studied extensively.[7] However, with the new physical applications[5,6] one needed especially an explicit construction of the representations of superalgebras. The most extensive and practical results were obtained by introducing a symmetrization–antisymmetrization procedure (called supersymmetrization),[8] of direct products of fundamental (defining) representations. Part of these results have also been obtained by other methods.[9] In particular, a labelling of irreducible representations in terms of Dynkin superdiagrams was given by Kac,[9] who also computed the mathematical properties (dimensions, etc.) of "typical" representations. These are the representations containing an equal number of bosons and fermions. However, "nontypical" representations are infinitely more numerous than "typical" ones. In the applications of Refs. 5 and 6 it is nontypical representations that were needed. The properties of both typical and nontypical representations were computed using Young supertableaux methods and given in the form of practical formulas in Ref. 8. The relation between supertableaux and Kac–Dynkin diagrams is obtained in Ref. 10. In addition to these finite dimensional representations, there are the infinite dimensional unitary representations of noncompact supergroups, which are realized on coset spaces in the form of induced representations. In supersymmetric field theoretical applications[3,4] the supercoset space plays the role of superspace, which includes the Minkowski position x_μ and the Majorana Grassmann variables θ^{α}_{α}. Geometrical properties of such representations are presently under study in the context of supersymmetric field theory and supergravity and will be reported elsewhere.

The rules of supersymmetrization are summarized by supertableaux,[8] which are similar in appearance to ordinary Young tableaux but indicate opposite symmetrization properties for the bosonic and fermionic components of the fundamental representations. The generalized supertableau contains both dotted ⊠ and undotted ⊡ boxes with a slash through each box to distinguish them from ordinary Young tableaux. An undotted box ⊡ corresponds to the covariant basis of the fundamental representation, which may be written in terms of a tensor with a single *lower* index ϕ_A. Similarly, a dotted box ⊠ corresponds to the conjugate of the fundamental representation whose contravariant basis may be written as ϕ^A with an *upper* index.

In previous papers[8] we constructed the representations of SU(N/M), Osp$(N/2M)$, and P(N) type supergroups in terms of direct products of (covariant and contravariant) fundamental representations. This was done by using analogies with the known methods[11] for constructing representations of ordinary Lie groups, which are summarized in the next paragraph. We formally associated certain supergroups with ordinary Lie groups by observing similar properties in their fundamental (defining) representations:

$$\begin{aligned}&\text{SU}(N/M)\text{ with SU}(N+M),\\&\text{Osp}(N/2M)\text{ with O}(N+2M),\\&\text{P}(N)\text{ with Sp}(2N).\end{aligned}\tag{1.1}$$

We then established a formal one-to-one correspondence between their higher irreducible representations whenever the shape of the supertableau for the supergroup was the same as the Young tableau for the corresponding ordinary Lie group. It was then possible to construct explicit irreducible *matrix representations* of the supergroup element, in direct product space. The super *character* was calculated and given in the form of a compact and explicit formula involving the supertrace of powers of the fundamental representation, that is Str(U^n), where U is the group element in the fundamental representation. These expressions were com-

a)Research supported in part by the U.S. Department of Energy, Contract Nos. DE-AC02-76ER03074 and DE-AC02-76ER03075.

 0022-2488/82/071239-09$02.50

pared to those of the ordinary groups and were seen to differ only in the replacement of supertrace by trace, provided the tableaux were in one-to-one correspondence. The correspondence guaranteed the equality of the various combinatoric factors indicated by the tableaux, except for the supertrace $\leftrightarrow$ trace distinction. This basic observation allowed us to calculate characters of ordinary Lie groups and thus derive our new formula for supercharacters. From supercharacters we easily wrote down formulas for the *dimensions* of these representations and *eigenvalues of all Casimir operators*.[8] Our formulas, when specialized to ordinary Lie groups, give the known results, but in some cases in a new form, in particular when dotted boxes are involved.

The rules of tensor products for obtaining irreducible representations of *ordinary Lie groups* can be summarized as follows. For the group SU(N) the basis for higher irreducible representations can be constructed in terms of direct products of only the fundamental representation $\square \sim \phi_a$, $a = 1,2,...,N$, according to the rules of Young tableaux.[11] However, as is often convenient and more transparent in physical applications, physicists use tensors containing both lower (covariant) and upper (contravariant) indices, where an upper index may be represented by a dotted box $\phi^a \sim \dot{\square}$.[12] An irreducible SU($N$) tensor is obtained by symmetrizing lower and upper indices *independently*, according to independent Young tableaux, and then subtracting all possible traces between upper and lower indices, by using the Kronecker delta $\delta_a{}^b$, which is an invariant of SU(N). The *traceless* tensor thus obtained forms the basis of an irreducible representation of SU(N). All these operations can be summarized by a generalized Young tableau containing both dotted and undotted boxes, as in Fig. 1.

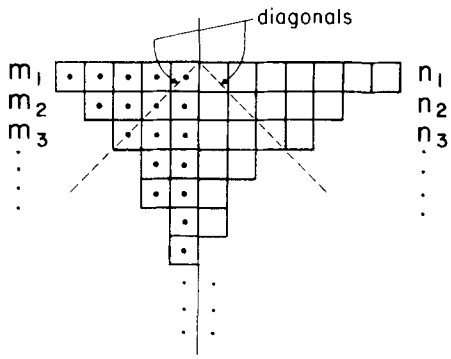
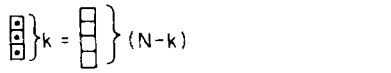


FIG. 1. A generalized Young tableau.

Consider such a tableau with r undotted and $\dot{r}$ dotted rows. The Young tableau is legal provided $n_1 \geqslant n_2 \geqslant \cdots \geqslant n_r \geqslant 0$ and $m_1 \geqslant m_2 \geqslant \cdots \geqslant m_{\dot{r}} \geqslant 0$. For SU($N$) the tensor will vanish and the tableau is meaningless unless

$$r \leqslant N, \quad \dot{r} \leqslant N - r. \tag{1.2}$$

These inequalities, in particular the last one, follow from the remarks below and the fact that if more than N indices are antisymmetrized in any column, the tensor vanishes. Any column with k dotted boxes may be replaced by a column with $N - k$ undotted boxes, that is



(1.3)

This is seen by using the SU(N) invariance properties of the completely antisymmetric Levi–Civita tensor $\epsilon_{a_1 a_2 \cdots a_N}$ or $\epsilon^{a_1 a_2 \cdots a_N}$ in N dimensions. Thus, any tableau containing dotted boxes can be converted to a conventional tableau containing only undotted boxes. Because of Eq. (1.3), any column containing exactly N dotted or N undotted boxes may be ignored as far as the SU(N) transformation is concerned, since that part of the tensor represents a singlet of SU(N) (but does not vanish).

For groups O($2N$), O($2N + 1$), and Sp($2N$) the complex conjugate representation is equivalent to the original representation $\dot{\square} \sim \square$. Therefore, there are no dotted boxes. An irreducible basis may be constructed just as in SU(N) by using Young tableaux with undotted boxes satisfying $n_1 \geqslant n_2 \geqslant \cdots \geqslant 0$. Because of Eq. (1.3), it is sufficient to consider N rows to obtain all irreducible representations (except spinor representations), but the tensor would not vanish unless the number of rows is more than the dimension of the fundamental representation ($2N$ or $2N + 1$). Unlike SU(N), for these groups there are invariant tensors with two lower indices. They can be taken as the Kronecker delta δ_{ab} for SO($2N$) and SO($2N + 1$) and as a standard $2N \times 2N$ antisymmetric matrix

$$C_{ab} = \left(\begin{array}{c|c} 0 & I_N \\ \hline -I_N & 0 \end{array}\right)$$

for Sp($2N$). After doing the appropriate Young tableau symmetrizations all possible "traces" must be taken away by contracting with these tensors. This means that traces are taken away when SO($2N$) or SO($2N + 1$) indices are symmetrized and when Sp($2N$) indices are antisymmetrized. The "traceless" tensor thus obtained forms the basis of an irreducible representation of the above groups.

For the supergroups in Eq. (1.1), the forms of the supertableaux are the same as the corresponding groups except for $\square$ or $\dot{\square}$ replacing $\square$ or $\dot{\square}$ and "supertraces" taken away instead of traces by contracting with the appropriate superinvariant tensor.[8] The "supertraceless" tensors so obtained form irreducible bases for the supergroups. For supergroups the restrictions of Eq. (1.2) are removed and Eq. (1.3) no longer holds. This is because there are no completely antisymmetric Levi–Civita tensors for supergroups. The invariant superdeterminant which plays a similar role to the determinant cannot be written in terms of a tensor with a finite number of indices since it is a *ratio* of two ordinary determinants. As a consequence, *supertableaux can have any number of rows* with dotted as well as undotted boxes. Furthermore, supertableaux with dotted boxes can no longer be converted to supertableaux with only undotted boxes since Eq. (1.3) does not hold [*except for special cases as in Ref. 10*]. Therefore, unlike SU(N), representations described by dotted supertableaux are independent ones and *the notation with dotted boxes is necessary*.

The SU(N/M) $\supset$ SU(N) $\otimes$ SU(M) $\otimes$ U(1) branching rules of the superrepresentations is of primary importance in practical physical applications.[5,6] This decomposition is also the key for establishing the relation[10] between supertableaux and the Kac–Dynkin diagrams. The irreducibility of our representations, which was discussed to a limited extent in our

previous work,[8] becomes evident after this connection. In this article we will explicitly study these branching rules. We will also indicate the branching rules with respect to other subgroups, such as $SU(N_1+N_2/M_1+M_2)\supset SU(N_1/M_1)\otimes SU(N_2/M_2)\otimes U(1)$, which follow from the same methods. The same techniques apply to $Osp(N/2M)$ and $P(N)$.

II. SU(N/M) BRANCHING RULES FROM THOSE OF SU(N + M)

As in our previous work, we continue to make progress by further exploiting the relationship between $SU(N/M)$ and $SU(N+M)$. Here we will compare the branching rules for

$$SU(N+M)\supset SU(N)\otimes SU(M)\otimes U(1)$$

and

$$SU(N/M)\supset SU(N)\otimes SU(M)\otimes U(1). \tag{2.1}$$

In the $N+M$ dimensional fundamental representation, the U(1) generator is a traceless matrix for $SU(N+M)$ and a supertraceless one for $SU(N/M)$. Up to an overall constant, it is given as

$$U(1):\ \left(\begin{array}{c|c} 1/N & 0 \\ \hline 0 & -1/M \end{array}\right) \text{ for } SU(N+M),$$

$$U(1):\ \left(\begin{array}{c|c} 1/N & 0 \\ \hline 0 & 1/M \end{array}\right) \text{ for } SU(N/M). \tag{2.2}$$

In the following we will work our way up by starting with a few simple examples and eventually arrive at some general observations that hold for any representation.

The fundamental representation $\phi_A \sim \square$ (or $\boxslash$) can be split into the direct sum $\phi_A = \phi_a \oplus \psi_\alpha$. Here the N dimensional piece ϕ_a, $a = 1,2,...,N$, transforms like the fundamental representation of $SU(N)$, is singlet under $SU(M)$, and carries the U(1) charge $1/N$. We denote this part by $\phi_a \sim (\square,1)_{1/N}$. Similarly, the M dimensional piece ψ_α, $\alpha = 1,2,...,M$, belongs to the fundamental representation of $SU(M)$, is a singlet under $SU(N)$, and carries the U(1) charge $(-1/M)$ for $SU(N+M)$ and $1/M$ for $SU(N/M)$. We denote it by $\psi_\alpha \sim (1,\square)_{-1/M}$ or $(1,\square)_{1/M}$, respectively. For $SU(N+M)$ ϕ_a and ψ_α are both bosons (or both fermions). For $SU(N/M)$ one of them is a boson and the other is a fermion. In representations of Class I[8] we chose ϕ = boson and ψ = fermion. It is sufficient to restrict ourselves to only class I representations since all other representations (Class II and mixed cases) can be obtained from those of pure Class I representations by simply switching bosons and fermions in the final *basis* without changing the matrix representation of the *group element*.[8]

From the above explanation, the branching equation (2.1) for the fundamental representation $\phi_A = \phi_a \oplus \psi_\alpha$ can be expressed in terms of tableaux as

$$\square = (\square,1)_{1/N} \oplus (1,\square)_{-1/M} \quad \text{for } SU(N+M),$$

$$\boxslash = (\square,1)_{1/N} \oplus (1,\square)_{1/M} \quad \text{for } SU(N/M). \tag{2.3}$$

Next consider the *completely symmetric* (completely supersymmetric) tensor with n indices $\phi_{(A_1A_2\cdots A_n)}$. By specializing each index $A_i = a_i \oplus \alpha_i$, where $a_i = 1,2,...,N$ and $\alpha_i = 1,2,...,M$, we can find the various $SU(N)\otimes SU(M)\otimes U(1)$ components of this tensor as

$$\phi_{(A_1A_2\cdots A_n)} = \phi_{(a_1a_2\cdots a_n)} \oplus \phi_{(a_1a_2\cdots a_{n-1})(\alpha_n)} \oplus \cdots$$

$$\oplus \phi_{(a_1a_2)(\alpha_3\alpha_4\cdots\alpha_n)} \oplus \phi_{(a_1)(\alpha_2\alpha_3\cdots\alpha_n)} \oplus \phi_{(\alpha_1\alpha_2\cdots\alpha_n)}. \tag{2.4}$$

The U(1) charges can be computed by assigning $1/N$ to each a_i and $-1/M$ ($+1/M$) to each α_i. For $SU(N+M)$ $\phi_{(a_1a_2\cdots a_{n-k})(\alpha_{n-k+1}\cdots\alpha_N)}$ must be completely symmetrized in both sets of indices a_i or α_i since the original indices $(A_1\cdots A_N)$ were completely symmetrized. Thus, it transforms as the direct product representation

$$\left(\overbrace{\square\square\square\square\square}^{n-k},\ \overbrace{\square\square\square\square\square}^{k}\right)$$

of $SU(N)\times SU(M)$. But for $SU(N/M)$, since supersymmetrization[8] of $(A_1A_2\cdots A_n)$ implies symmetrization of the bosons ϕ_{a_i} and antisymmetrization of the fermions ψ_{a_i}, the bosonic indices $(a_1a_2\cdots a_{n-k})$ are symmetric but the fermionic indices $(\alpha_{n-k+1}\cdots\alpha_n)$ are antisymmetric. Thus, from (2.4) we can write the branching rule[8]

$$\text{Su(N+M):}\ \overbrace{\square\square\square\square\square\square\square}^{n} = \sum_{k=0}^{n} \left(\overbrace{\square\square\square}^{(n-k)},\ \overbrace{\square\square}^{k}\right)_{(n-k)/N-k/M}$$

$$\text{Su(N/M):}\ \overbrace{\square\square\square\square\square\square\square}^{n} = \sum_{k=0}^{n} \left(\overbrace{\square\square\square}^{(n-k)},\ \left\{\begin{array}{c}\square\\\square\end{array}\right\}k\right)_{(n-k)/N+k/M} \tag{2.5}$$

The number of terms appearing in the first of these equations is $(n+1)$. However, in the second equation some of the Young tableaux

$$\left\{\begin{array}{c}\square\\\square\\\square\end{array}\right\}k$$

for $SU(M)$ vanish if $k > M$. Thus, if $n \leqslant M$ there will be $n+1$ terms, but if $n > M$ there will be fewer terms.

Note that the *pictures* in Eq. (2.5) are independent of the value of N and M. They are completely determined by the permutation symmetry of the original one row tableau. Therefore, the comparison of the branching rule for Lie groups and Lie supergroups need not be restricted to groups that have identical subgroups. For example, instead of comparing the branching rules of SU(6/4) to those of SU(10) we may just as well compare them to those of SU(75), since the *pictures* of SU(10) and SU(75) are identical except that Young tableaux for SU(10) with more than 10 rows vanish. Similarly, for a given shape of the supertableau the pictures of SU(6/4) are identical with those of any other $SU(N/M)$ except for the illegal $SU(N)$ or $SU(M)$ tableaux that vanish. Thus, in branching rule calculations, we will always consider N and M to be as large as necessary so that none of the $SU(N)$ or $SU(M)$ tableaux vanish for both $SU(N+M)$ and $SU(N/M)$. Thus, the branching rule *pictures* that are obtained will be generic to the tableau and independent of the values of N and M. They will depend only on the numbers m_i and n_i of the dotted and undotted boxes in the original tableau as, e.g., in Fig. 1. After obtaining the branching rule for a *given tableau* specified by m_i and n_i (not for given M and N), we can specialize to any desired values of N and M and

eliminate, if necessary, any illegal SU(N) or SU(M) tableau that does not satisfy Eq. (1.2) for given values of r, $\dot{r}$, and N (or M). In this way we obtain branching rules for the whole series of SU(N/M) supergroups rather than specific N,M. Equation (2.5), which is the first such example, was obtained in Ref. 8 and through other methods in Ref. 9.

Next we consider the SU($N+M$) tableau

and SU(N/M) supertableau

which correspond to a tensor $\phi_{(A_1A_2A_3);B_1}$. By specializing the indices $A_i = a_i \oplus \alpha_i$, $B_1 = b_1 \oplus \beta_1$, we obtain the various components. The *independent* components are specified by considering the indices a_i to be *lower* than the indices b_i within SU(N) and both a_i and b_i to be lower than the α_i or β_i within SU($N+M$), when they are allowed to take values $A = 1,2,\ldots,N+M$. Then we obtain

$$\begin{aligned}\phi_{(A_1A_2A_3);B_1} &= \phi_{(a_1a_2a_3);b_1} \oplus \phi_{(a_1a_2a_3);\beta_1} \oplus \phi_{(a_1a_2)(\alpha_3);b_1}, \\ &\oplus \phi_{(a_1a_2)(\alpha_3);\beta_1} \oplus \phi_{(a_1)(\alpha_2\alpha_3);b_1} \\ &\oplus \phi_{(a_1)(\alpha_2\alpha_3);\beta_1} \oplus \phi_{(\alpha_1\alpha_2\alpha_3);\beta_1}. \end{aligned} \tag{2.6}$$

Note that we did not include $\phi_{(\alpha_1\alpha_2\alpha_3);b_1}$, and some others, even though they could appear as possible components of the tensor. This is because of the ordering rule $a_i < b_i < \alpha_i < \beta_i$, which allows us to select the independent components of the tensor *only once*. According to this rule, we cannot allow b_1 to appear in the second row when the first row contains only α's. The component with the symmetries of $\phi_{(\alpha_1\alpha_2\alpha_3);b_1}$ is already counted, as seen below.

For SU($N+M$) Eq. (2.6) can be written in terms of SU(N) $\otimes$ SU(M) Young tableaux as

$\phi_{(a_1a_2a_3);\,b_1} = ($  $,1)$

$\phi_{(a_1a_2)(a_3);\,b_1} = ($  $,\square)$

$\phi_{(a_1)(a_2a_3);\,b_1} = ($ $)$

$\phi_{(a_1a_2a_3);\,\beta_1} = ($  $)$ (2.7)

$\phi_{(a_1a_2)(a_3);\,\beta_1} = ($  $) + ($ $)$

$\phi_{(a_1)(a_2a_3);\,\beta_1} = ($ $) + ($ $)$

$\phi_{(a_1a_2a_3);\,\beta_1} = ($  $)$

Note that $\phi_{(a_1a_2)(\alpha_3);\beta_1}$ has two irreducible SU(M) components since the α_3 index in the first row and the β_1 index in the second row are not forced to be in any symmetry relation relative to each other. Thus, we obtain the two irreducible pieces, because for SU(M)

$$\square \otimes \square = \square\square \oplus \begin{smallmatrix}\square\\\square\end{smallmatrix} \tag{2.8}$$

Similarly, $\phi_{(a_1)(\alpha_2\alpha_3);\beta_1}$ has two irreducible components since

$$\square\square \otimes \square = \square\!\square\!\square \oplus \square\square\square \tag{2.9}$$

The second piece in Eq. (2.9) corresponds to the component $\phi_{(\alpha_1\alpha_2\alpha_3);b_1}$ mentioned above.

Thus, for SU($N+M$) $\supset$ SU(N) $\otimes$ SU(M) $\otimes$ U(1) the branching rule for this tableau is

$$\begin{aligned} \begin{smallmatrix}\square\\\square\end{smallmatrix} = &(\begin{smallmatrix}\square\\\square\end{smallmatrix},1)_{4/N} \oplus (1,\begin{smallmatrix}\square\\\square\end{smallmatrix})_{-4/M} \oplus (\square\square,\square)_{3/N-1/M} \\ &\oplus (\square,\square\square)_{1/N-3/M} \oplus (\begin{smallmatrix}\square\\\square\end{smallmatrix},\square)_{3/N-1/M} \\ &\oplus (\square,\begin{smallmatrix}\square\\\square\end{smallmatrix})_{1/N-3/M} \oplus (\square\square,\begin{smallmatrix}\square\\\square\end{smallmatrix})_{2/N-2/M} \\ &\oplus (\begin{smallmatrix}\square\\\square\end{smallmatrix},\square\square)_{2/N-2/M} \oplus (\square\square,\square\square)_{2/N-2/M} \end{aligned} \tag{2.10a}$$

Note that in the final result the *pictures* are symmetric under the interchange of SU(M) and SU(N), as they should be, since the permutation symmetry of the original tableau does not distinguish between SU(N) and SU(M) indices. Identical results would be obtained by considering the a_i,b_i indices to be higher compared to the α_i indices. The SU(N) $\leftrightarrow$SU(M) interchangeability of the branching rule reflects this fact. Thus, for every irreducible component (X,Y) there exists another irreducible component (Y,X), where X and Y represent the *pictures* of Young tableaux.

The same reasoning can be applied step by step to the SU(N/M) group. The only difference is that whenever the α_i's were symmetrized within SU($N+M$) they should be antisymmetrized within SU(N/M) and vice versa. This is required by the supersymmetrization indicated by the supertableau[8]. The a_i's have the same permutation properties in both SU($N+M$) and SU(N/M) as in the previous example. Thus, for SU(N/M), Eq. (2.7) will be modified by changing every SU(M) row into a column and vice versa. This means that every irreducible component (X,Y) that appeared for SU($N+M$) will have a counterpart $(X,\tilde{Y})$ for SU(N/M), where $\tilde{Y}$ is an SU(M) Young tableau reflected along the diagonal relative to Y. Thus, the analog of Eq. (2.10a) for SU(N/M) $\supset$ SU(N) $\otimes$ SU(M) $\otimes$ U(1) becomes

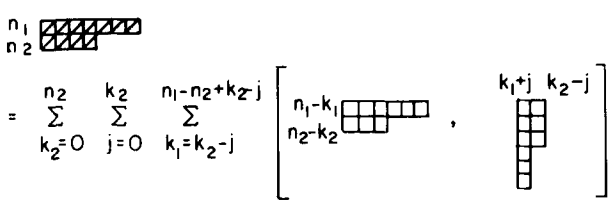
$$\begin{aligned} \begin{smallmatrix}\square\\\square\end{smallmatrix} = &(\begin{smallmatrix}\square\\\square\end{smallmatrix},1)_{4/N} \oplus (1,\begin{smallmatrix}\square\\\square\end{smallmatrix})_{4/M} \oplus (\square\square,\square)_{3/N+1/M} \\ &\oplus (\square,\begin{smallmatrix}\square\\\square\end{smallmatrix})_{1/N+3/M} \oplus (\begin{smallmatrix}\square\\\square\end{smallmatrix},\square)_{3/N+1/M} \\ &\oplus (\square,\begin{smallmatrix}\square\\\square\end{smallmatrix})_{1/N+3/M} \oplus (\square\square,\square\square)_{2/N+2/M} \\ &\oplus (\begin{smallmatrix}\square\\\square\end{smallmatrix},\begin{smallmatrix}\square\\\square\end{smallmatrix})_{2/N+2/M} \oplus (\square\square,\begin{smallmatrix}\square\\\square\end{smallmatrix})_{2/N+2/M} \end{aligned} \tag{2.10b}$$

Note that the $1/M$ pieces in the U(1)'s have switched signs. Furthermore, the pictures are now symmetric under the SU(N)$\leftrightarrow$SU(M) interchange only after being reflected along the diagonal. That is, for every irreducible component $(X,\tilde{Y})$ there exists also $(Y,\tilde{X})$.

We use the same methods as above for tableaux containing n_1 boxes in the first row and n_2 boxes in the second row, where $n_1 \geqslant n_2 \geqslant 0$. The result for SU($N+M$) is

$$\begin{smallmatrix}n_1\\n_2\end{smallmatrix}\overbrace{\square\cdots\square}^{\text{}} = \sum_{k_2=0}^{n_2} \sum_{j=0}^{k_2} \sum_{k_1=k_2-j}^{n_1-n_2+k_2-j} \left[\begin{smallmatrix}n_1-k_1\\n_2-k_2\end{smallmatrix}\overbrace{\square\cdots\square}^{\text{}}, \begin{smallmatrix}k_1+j\\k_2-j\end{smallmatrix}\overbrace{\square\cdots\square}^{\text{}} \right] \tag{2.11a}$$

while for SU(N/M) we only need to reflect the SU(M) tableau and obtain



(2.11b)

The U(1) charges were not indicated for lack of space but they are computed by counting the number of boxes for SU(N) and SU(M) tableaux and given as $(n_1 + n_2 - k_1 - k_2)/N \mp (k_1 + k_2)/M$, with the upper sign for SU($N + M$) and the lower one for SU(N/M). The number of independent irreducible terms appearing in the sum of Eq. (2.11) is easily computed to be

$$\tfrac{1}{2}(n_1 - n_2 + 1)(n_2 + 1)(n_2 + 2). \tag{2.12}$$

Some of these terms may vanish if N or M are too small and the tableau becomes illegal according to Eq. (1.2). Equations (2.11) reduce to (2.5) for the special case $n_2 = 0$.

After obtaining the result for 1 and 2 rows, it is immediate to arrive at the branching rule for 1 and 2 columns, simply by reflecting each tableau along the diagonal. These and a few other cases not containing dotted boxes are summarized in Table I.

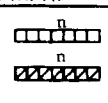
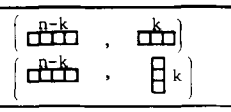
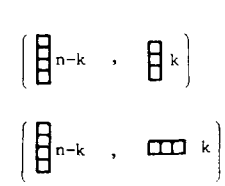
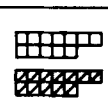
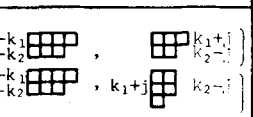
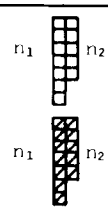
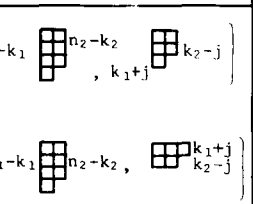
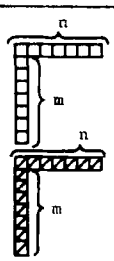
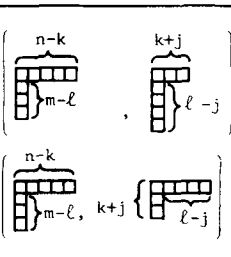
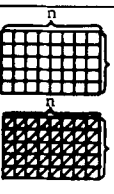
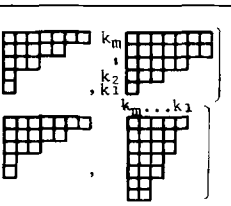
Returning to tableaux containing dotted boxes, we begin with the fundamental contravariant representation, $\phi^A = \phi^a \oplus \psi^a$, which may be written as

$$\square = (\square,1)_{-1/N} \oplus (1,\square)_{1/M} \quad \text{for SU}(N+M),$$

$$\square\!\!\square = (\square,1)_{-1/N} \oplus (1,\square)_{-1/M} \quad \text{for SU}(N/M). \tag{2.13}$$

The next simplest case is the adjoint representation $\phi_A{}^B$, which is a traceless matrix for SU($N + M$) and supertraceless for SU(N/M). Specializing the indices $A = a \oplus \alpha$, $B = b \oplus \beta$, we obtain the components

TABLE I. Examples of branching rules for SU(N/M) as compared with those of SU($N + M$). Note the limits on the number of boxes in the pictures. Only covariant representations are considered.

SU(N+M) / SU(N/M) IRREP	SU(N) ⊗ SU(M) IRREP	U(1) CHARGE	LIMITS ON NUMBERS OF BOXES	TOTAL NUMBER OF TERMS IF ALL TABLEAUX ARE LEGAL [SEE EQ. (1.2)]
		$\frac{n-k}{N} \mp \frac{k}{M}$	$0 \le k \le n$	$n + 1$
		$\frac{n-k}{N} \mp \frac{k}{M}$	$0 \le k \le n$	$n + 1$
		$\frac{n_1+n_2-k_1-k_2}{N} \mp \frac{k_1+k_2}{M}$	$0 \le k_2 \le n_2$ $0 \le j \le k_2$ $k_2-j \le k_1 \le n_1-n_2+k_2-j$	$\frac{1}{2}(n_1-n_2+1)(n_2+1)(n_2+2)$
		$\frac{n_1+n_2-k_1-k_2}{N} \mp \frac{k_1+k_2}{M}$	$0 \le k_2 \le n_2$ $0 \le j \le k_2$ $k_2-j \le k_1 \le n_1-n_2+k_2-j$	$\frac{1}{2}(n_1-n_2+1)(n_2+1)(n_2+2)$
		$\frac{n+m-k-\ell}{N} \mp \frac{k+\ell}{M}$	$j = 0,1$ $j \le \ell \le m$ $1-j-\delta_{\ell 0} \le k \le n-1+(1-j)\delta_{\ell m}$ $\delta_{\ell i}$ = Kronecker delta	$2nm+n-m+1$
		$\frac{nm}{N} - \left(\frac{1}{N} \pm \frac{1}{M}\right)\left(\sum_{i=1}^{m} k_i\right)$	$0 \le k_1 \le k_2 \ldots \le k_m \le n$	$\frac{(n+m)!}{n!m!}$

$$\phi_A{}^B = \phi_a{}^b \oplus \phi_a{}^\beta \oplus \phi_\alpha{}^b \oplus \phi_\alpha{}^\beta. \tag{2.14}$$

The pieces $\phi_a{}^b$ and $\phi_\alpha{}^\beta$ are not irreducible with respect to SU(N) and SU(M), respectively, since we have not yet insured that they are traceless. Thus

$$\phi_a{}^b = \tilde{\phi}_a{}^b + \frac{1}{N}\delta_a{}^b\phi = (\Box\Box,1) \oplus (1,1),$$

$$\phi_\alpha{}^\beta = \tilde{\phi}_\alpha{}^\beta \mp \frac{1}{M}\delta_\alpha{}^\beta\phi = (1,\Box\Box) \oplus (1,1), \tag{2.15}$$

where the singlet part $\phi = (1,1)$ is identical in both pieces so that the tracelessness (supertracelessness) condition on $\phi_A{}^B$ is satisfied. The result for SU($N+M$) is

$$\Box\boxdot = (\Box\boxdot,1)_0 \oplus (1,\Box\boxdot)_0 \oplus (1,1)_0$$
$$\oplus (\Box,\boxdot)_{1/N+1/M} \oplus (\boxdot,\Box)_{-1/N-1/M}, \tag{2.16a}$$

while for SU(N/M) we have

$$\boxslash\boxslash = (\Box\boxdot,1)_0 \oplus (1,\Box\boxdot)_0 \oplus (1,1)_0$$
$$\oplus (\Box,\boxdot)_{1/N-1/M} \oplus (\boxdot,\Box)_{-1/N+1/M}. \tag{2.16b}$$

Next we consider the tensor $\phi_{A_1A_2\cdots A_n}^{B_1B_2\cdots B_m}$ with both lower and upper indices symmetrized (or supersymmetrized) and satisfying the trace (or supertrace) condition. Specializing the indices, we have

$$\phi_{A_1A_2\cdots A_n}^{B_1B_2\cdots B_m} = \sum_{k=0}^{n}\sum_{l=0}^{m} \phi_{(a_1a_2\cdots a_{n-k})(\alpha_{n-k+1}\cdots\alpha_n)}^{(b_1b_2\cdots b_{m-l})(\beta_{m-l+1}\cdots\beta_m)}. \tag{2.17}$$

The U(1) quantum numbers can be calculated by assigning $1/N$ for each a_i, $-1/N$ for each b_i, $(\mp 1/M)$ for each α_i, and $(\mp 1/M)$ for each β_i, where the upper sign is for SU($N+M$) and the lower signs for SU(N/M). The various terms in the sum are, in general, reducible with respect to SU(N) $\otimes$ SU(M). For example, for $n=2$ and $m=2$,

$$\phi_{(a_1a_2)}{}^{(b_1b_2)} = (\boxdot\boxdot\Box\Box,1) \oplus (\boxdot\Box,1) \oplus (1,1),$$

$$\phi_{(a_1\alpha_2)}{}^{(b_1b_2)} = (\boxdot\boxdot\Box,\Box) \oplus (\boxdot,\Box), \quad \text{etc.} \tag{2.18}$$

Note that, for each reduction which is achieved by using a Kronecker delta $\delta_a{}^b$ (or $\delta_\alpha{}^\beta$), we get to eliminate one dotted and one undotted box from the picture of an SU(N) [or an SU(M)] tableau. Just as the ϕ in Eq. (2.15), we must be aware that the tracelessness of the original tensor $\phi_{A_1\cdots A_n}{}^{B_1\cdots B_m}$ imposes that some of the pieces in the various traces are identical and should not be counted more than once. To insure this property we count only the traces calculated by contracting with the SU(M) $\delta_\alpha{}^\beta$ and ignore those obtained with $\delta_a{}^b$, since they are the same ones. With these conditions, we arrive at the SU($N+M$) branching rule

$$\overbrace{\boxdot\boxdot\boxdot\boxdot\boxdot}^{m}\overbrace{\Box\Box\Box\Box\Box\Box\Box}^{n}$$

$$= \sum_{k=0}^{n}\sum_{\ell=0}^{m}\sum_{i=0}^{\min(k,\ell)} \left[\overbrace{\boxdot\boxdot\boxdot\boxdot}^{m-\ell}\overbrace{\Box\Box\Box}^{n-k},\ \overbrace{\boxdot\boxdot\boxdot}^{\ell-i}\overbrace{\Box\Box\Box\Box}^{k-i}\right] \tag{2.19a}$$

The sum over i takes care of the traces and produces the pieces similar to the $\phi \sim (1,1)$ of Eq. (2.15). The U(1) charge, whic depends only on the number of boxes, is given as $(n-k-m+l)/N - (k-l)/M$.

For SU(N/M) the reasoning is identical; however, the α_i's and β_i's should now be antisymmetrized as opposed to being symmetrized in the previous case. Therefore, the SU(M) tableaux should be changed relative to the previous case by substituting columns instead of rows. Otherwise, every step can be repeated to obtain

$$\boxslash\boxslash\boxslash\boxslash\boxslash\boxslash\boxslash\boxslash\boxslash$$

$$= \sum_{k=0}^{n}\sum_{\ell=0}^{m}\sum_{i=0}^{\min(k,\ell)} \left[\overbrace{\boxdot\boxdot\boxdot\boxdot}^{m-\ell}\overbrace{\Box\Box\Box}^{n-k},\ \ell-i \left\{\begin{matrix}\boxdot\\\boxdot\\\Box\\\Box\end{matrix}\right\} k-i\right] \tag{2.19b}$$

We see that in going from SU($N+M$) to SU(N/M), the SU(M) tableaux get reflected independently for the dotted and undotted boxes along their respective diagonals. The diagonals are shown in Fig. 1.

Our method should be quite clear to the reader by now. Without giving any more details, we list our results for a few tableaux containing dotted boxes in Table II. We emphasize that these correspond to arbitrarily large representations of arbitrarily large groups. Together with Table I we expect that these concrete results should be quite sufficient for a variety of physical applications that we can now foresee. More complicated cases can be worked out, if necessary, with the same methods.

III. BRANCHING RULE FOR SU(N_1+N_2/M_1+M_2)$\supset$SU(N_1/M_1)$\otimes$SU(N_2/M_2)$\otimes$U(1)

This branching is again obtained from that of SU($N+M$)$\supset$SU(N)$\times$SU(M)$\times$U(1) by a reinterpretation of the boxes in the tableaux. Let us first identify the fundamental representation $\phi_A = \phi_a \oplus \phi_\alpha$ as follows: ϕ_a contains N_1 bosons and M_1 fermions and belongs to the fundamental representation $\phi_a \sim \boxslash$ of SU(N_1/M_1); similarly, ϕ_α contains N_2 bosons and M_2 fermions and belongs to the fundamental representation $\phi_\alpha \sim \boxslash$ of SU(N_2/M_2).

The U(1) generator, which is a $N_1+M_1+N_2+M_2$ dimensional diagonal matrix in the fundamental representation, is identified as

$$\begin{pmatrix} 1/(N_1-M_1) & 0 \\ 0 & -1/(N_2-M_2)\end{pmatrix} \begin{matrix} N_1+M_1 \\ N_2+M_2 \end{matrix}. \tag{3.1}$$

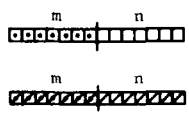
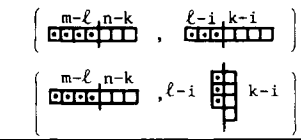
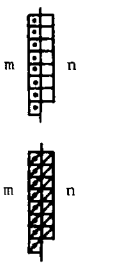
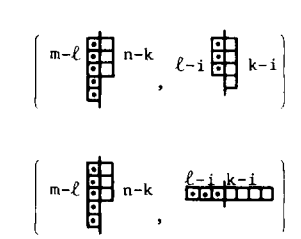
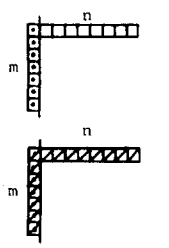
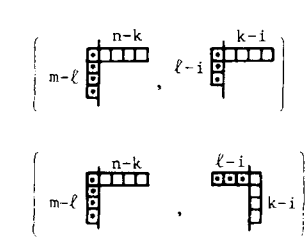
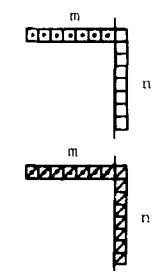
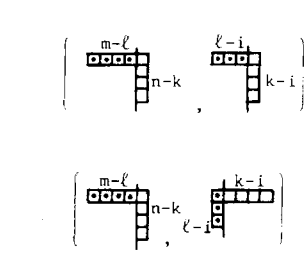
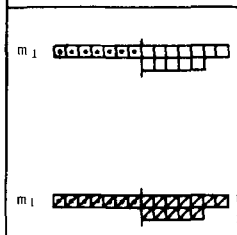
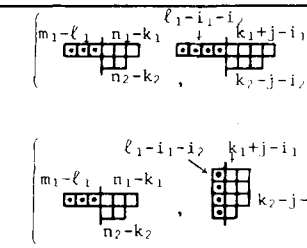
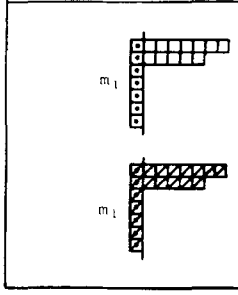
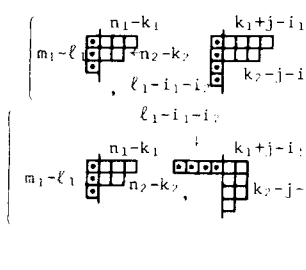
Thus, for each index a or α we obtain the U(1) charges $1/(N_1-M_1)$ or $-1/(N_2-M_2)$, respectively. Similarly, the U(1) is computed by replacing every $1/N$ or $-1/M$ in the old expressions by $1/(N_1-M_1)$ or $-1/(N_2-M_2)$, respectively. Therefore, $\phi_A = \phi_a \oplus \phi_\alpha$ may be written in terms of tableaux as

$$\boxslash = (\boxslash,1)_{1/(N_1-M_1)} \oplus (1,\boxslash)_{-1/(N_2-M_2)}. \tag{3.2}$$

Note the formal similarity to Eq. (2.3) for SU($N+M$), except that *every box* is replaced by a slashed box, and the values of the U(1) charges are computed by different assignments to a and α, as explained above.

The completely supersymmetric tensor $\phi_{(A_1A_2\cdots A_n)}$ can be decomposed by specializing each index $A_i = a_i \oplus \alpha_i$, just as in Eq. (2.4). However, the meaning of $\phi_{(a_1,a_2,\cdots,a_n)}$, etc., now differs from the SU($N+M$) case in that the indices $(a_1,a_2,\cdots)$ or $(\alpha_1,\alpha_2,\cdots)$ are supersymmetrized rather than simply symmetrized. Therefore, Eq. (2.5) now gets replaced by

TABLE II. Same as Table I, except for mixed covariant–contravariant representations.

SU(N+M) IRREP SU(N/M)	SU(N) ⊗ SU(M) IRREP	U(1) CHARGE	LIMITS ON NUMBERS OF BOXES	TOTAL NUMBER OF TERMS IF ALL TABLEAUX ARE LEGAL [SEE EQ. (1.2)]
		$\frac{n-k-m+\ell}{N} \mp \frac{k-\ell}{M}$	$0 \le k \le n$ $0 \le \ell \le m$ $0 \le i \le \min(k,\ell)$	$\frac{1}{6}(m+1)(m+2)(3n-m+3)$ if $n \ge m$ $\frac{1}{6}(n+1)(n+2)(3m-n+3)$ if $n \le m$
		$\frac{n-k-m+\ell}{N} \mp \frac{k-\ell}{M}$	$0 \le k \le n$ $0 \le \ell \le m$ $0 \le i \le \min(k,\ell)$	$\frac{1}{6}(m+1)(m+2)(3n-m+3)$ if $n \ge m$ $\frac{1}{6}(n+1)(n+2)(3m-n+3)$ if $n \le m$
		$\frac{n-k-m+\ell}{N} \mp \frac{k-\ell}{M}$	$i = 0,1$ $i \le k \le n$ $i \le \ell \le m$	$2mn + m + n + 1$
		$\frac{n-k-m+\ell}{N} \mp \frac{k-\ell}{M}$	$i = 0,1$ $i \le k \le n$ $i \le \ell \le m$	$2mn + m + n + 1$
		$\frac{n_1+n_2-m_1}{N} - \left(\frac{1}{N} \mp \frac{1}{M}\right)\left(k_1+k_2-\ell_1\right)$	$0 \le \ell_1 \le m_1$ $0 \le k_2 \le n_2$ $0 \le j \le k_2$ $k_2-j \le k_1 \le n_1-n_2+k_2-j$ $0 \le i_2 \le \min(k_2-j,\ell_1)$ $0 \le i_1 \le \min(k_1-k_2+2j,\ell_1-i_2)$	$\frac{1}{2}(n_2+1) \times \{3m_1n_2+2)(n_1-n_2) + 2m_1(2n_1+1)+2\}$, $m_1 \ge 1$
		$\frac{n_1+n_2-m_1}{N} - \left(\frac{1}{N} \mp \frac{1}{M}\right)\left(k_1+k_2-\ell_1\right)$	$i_1 = 0,1$; $i_2 = 0,1$ $i_1 + i_2 \le \ell_1 \le m_1$ $i_2 \le k_2 \le n_2$ $0 \le j \le k_2-i_2$ $i_1-i_2+k_2-j \le k_1 \le n_1-n_2+k_2-j$	

$$\overset{n}{\square\square\square\square\square\square\square} = \sum_{k=0}^{n} (\overset{n-k}{\square\square\square\square}, \overset{k}{\square\square\square}), \tag{3.3}$$

with the U(1) charges given by $[(n-k)/(N_1-M_1) - k/(N_2-M_2)]$.

The analysis is the same for any other supertableau and the result for the new branching rule is obtained from the known cases of SU($N+M$) by simply replacing every SU(N) or SU(M) box $\square$ by slashed boxes $\boxslash$ belonging to SU(N_1/M_1) or SU(N_2/M_2), respectively. Therefore, all the results listed

in Tables I and II are directly generalized to the new branching.

IV. DISCUSSION AND CONCLUSIONS

We have established a one-to-one correspondence between the branching rules $SU(N+M)\supset SU(N)\otimes SU(M)\otimes U(1)$ and $SU(N/M)\supset SU(N)\otimes SU(M)\otimes U(1)$ as well as $SU(N_1+N_2/M_1+M_2)\supset SU(N_1/M_1)\otimes SU(N_2/M_2)\otimes U(1)$. Some concrete examples, which we expect to be sufficient for most physical applications, have been explicitly worked out and listed in Tables I and II. More complicated cases can be analyzed with the methods given here.

For $SU(N+M)$ branching rules, there are useful lists available in the literature.[13] Our $SU(N+M)$ results are in complete agreement with these known cases. One virtue of our approach is that we are not limited by large dimensions of representations or groups. Thus, in our Tables I and II one finds large dimensions not covered in the extensive lists of Ref. 13.

In making comparisons with these lists one must be aware that some Young tableaux for $SU(N)$ or $SU(M)$ become illegal [see Eq. (1.2)] and vanish if either N or M are too small. One may use the available $SU(N+M)$ lists to derive additional practical $SU(N/M)$ results not covered in this paper explicitly, provided N and M are large enough to insure no tableau vanishes. Useful branching rules can be obtained from Ref. 13 or similar lists by noting the following general observations which follow from our analysis above.

Consider an arbitrary Young tableau T for $SU(N+M)$ and the $SU(N)\otimes SU(M)\otimes U(1)$ branching rule

$$T=\sum\oplus(X,Y), \tag{4.1}$$

where X and Y denote Young tableaux for $SU(N)$ and $SU(M)$, respectively. In general, T,X,Y contain both dotted and undotted boxes. In tables such as Ref. 13, Dynkin indices are used to label a representation. They must be converted to Young tableau notation in order to apply our method below.

For every $SU(N+M)$ Young tableau T, as in Fig. 1, we can define an $SU(N/M)$ supertableau $\mathbb{T}$, with identical numbers n_i, m_i for its rows, except that every box $\square$ or $\boxdot$ is replaced by a slashed box $\boxslash$ or $\boxbslash$. We may then consider the branching rule for $SU(N/M)\to SU(N)\otimes SU(M)\otimes U(1)$ as

$$\mathbb{T}=\sum\oplus(X,\tilde{Y}), \tag{4.2}$$

where $\tilde{Y}$ is the reflection of Y along its diagonals. The diagonals are shown in Fig. 1. Similarly, the branching rule for $SU(N_1+N_2/M_1+M_2)\supset SU(N_1/M_1)\otimes SU(N_2/M_2)\otimes U(1)$ can be written as

$$\mathbb{T}=\sum\oplus(\mathbb{X},\mathbb{Y}), \tag{4.3}$$

where $\mathbb{X}$ and $\mathbb{Y}$ are the supertableaux analogous to X and Y. Equations (4.2) and (4.3) follow from (4.1) if one uses tensor language $\phi_{B_1B_2\cdots}{}^{A_1A_2\cdots}$ and specializes each index $A_i=a_i\oplus\alpha_i$, etc. It is necessary to consider the meaning of supersymmetrization[8] in relation to ordinary symmetrization, and then extracting irreducible $SU(N)$ and $SU(M)$ components. These statements can be understood by following the examples in Tables I and II.

Equation (4.1) contains the terms $(1,T)$ and $(T,1)$, where T is the $SU(N)$ or $SU(M)$ representation with the most boxes that could appear in the branching rule. The *pictures* that represent the decomposition $T=\Sigma\oplus(X,Y)$ are independent of N or M. Therefore, we will assume that N and M are large enough so that the Young tableau T or the reflection from its diagonals $\tilde{T}$ does not vanish for $SU(N)$ or $SU(M)$. This insures that every (X,Y) or $(X,\tilde{Y})$ that could appear in the sum does not vanish for $SU(N+M)$ or $SU(N/M)$. After obtaining the branching rules for such large N,M we can apply the result to smaller N,M, as it may be necessary in some practical application. Then, we only need to eliminate the illegal $SU(N)$ or $SU(M)$ tableaux according to Eq. (1.2)

The following statements hold for any decomposition noted in Eqs. (4.1), (4.2), and (4.3), and can be used as a check in any calculation.

(1) The total number of undotted *minus* dotted boxes is identical in every term of the sum and equal to the same quantity for T or $\mathbb{T}$.

(2) For a term (X,Y) that appears in Eq. (4.1), there should be another term (Y,X) provided it does not vanish according to Eq. (1.2). For Eq. (4.2) this implies that for every term $(X,\tilde{Y})$ there should be a $(Y,\tilde{X})$, while in Eq. (4.3) for every $(\mathbb{X},\mathbb{Y})$ there should exist a $(\mathbb{Y},\mathbb{X})$.

(3) The maximum number of rows and columns that can appear in any (X,Y) or $(X,\tilde{Y})$ or $(\mathbb{X},\mathbb{Y})$, as well as the general shape of these Young tableaux, are predetermined by the numbe of rows and columns and general shape of T or $\mathbb{T}$. See the examples in Tables I and II.

(4) The dimension of $\mathbb{T}$ or T on the left-hand side of these equations should match with the sum of the dimensions on the right-hand side. For this purpose one can use the practical dimension formulas for the numbers of bosons and fermions developed in Ref. 8. On the right-hand side of Eq. (4.2) a fermion is obtained when the $SU(M)$ representation $\tilde{Y}$ contains an odd total number of dotted *plus* undotted boxes (independent of X). (This is a fermion in a Class I representation. For Class II we demand an odd number of boxes in X rather than in Y. While in mixed Class I–II representations the roles of bosons and fermions may be interchanged.[8] The representation of the group element is independent of the class.)

We have not discussed branching rules for $Osp(N/2M)$ and $P(N)$. However, from the remarks in the Introduction and Ref. 8, and the methods of this paper, it is clear that Eqs. (4.1), (4.2), and (4.3) must also apply to these groups and that detailed branching rules can be obtained with the same approach as the present paper.

The results of this paper are used in Ref. 10 to establish the relation between Kac–Dynkin diagrams,[9] which give a unique labeling of representations of supergroups, and Young supertableaux, which are practical for obtaining concrete results, as illustrated here.

[1] L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975).
[2] A. Neveu and J. H. Schwarz, Nucl. Phys. B **31**, 86 (1971); P. Ramond, Phys. Rev. D **3**, 2415 (1971).

[3]J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974).
[4]S. Deser and B. Zumino, Phys. Lett. B **62**, 335 (1976); D. Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara, Phys. Rev. D **14**, 912 (1976). For a recent review see P. van Nieuwenhuizen, Phys. Rep. C **68**, 189 (1981).
[5]A. B. Balantekin, I. Bars, and F. Iachello, Phys. Rev. Lett. **47**, 19 (1981); Nucl. Phys. A **370**, 284 (1981); for a preliminary version of this work, see I. Bars, "Dynamical Supersymmetry SU(6/4) in Nuclei," in *Group Theoretical Methods in Physics, Lecture Notes in Physics*, Vol. 135, edited by K. B. Wolf (Springer, Berlin, 1980); *Interacting Bose–Fermi Systems in Nuclei*, edited by F. Iachello (Plenum, New York 1981).
[6]T. Banks, A. Schwimmer, and S. Yankielowicz, Phys. Lett. B **96**, 67 (1980); I. Bars and S. Yankielowicz, *ibid.* **101**, 159 (1981); I. Bars, *ibid.*, **106**, 105 (1981), Yale Preprint YTP82-04.
[7]P. G. O. Freund and I. Kaplansky, J. Math. Phys. **17**, 228 (1976); V. G. Kac, Commun. Math. Phys. **53**, 31 (1977).
[8]A. B. Balantekin and I. Bars, J. Math. Phys. **22**, 1149 (1981); **22**, 1810 (1981); A. B. Balantekin, J. Math. Phys. **23**, 486 (1982).
[9]V. G. Kac, in *Differential Geometrical Methods in Mathematical Physics*, edited by K. Bleuler, H. R. Petry, and A. Reetz (Springer, Berlin, 1978); Y. Ne'eman and S. Sternberg, Proc. Natl. Acad. Sci. USA **77**, 3127 (1980); P. H. Dondi and P. D. Jarvis, Z. Phys. **C4**, 201 (1980); J. Phys. A **14**, 547 (1981); M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977); M. Marcu, *ibid.* **21**, 1277 (1980); J. Thierry-Mieg and B. Morel, Harvard Preprint HUTMP 80/8100; J. P. Hurni and B. Morel, University of Geneva Preprint UGVA-DPT 1981/03-280. [There are some incorrect comments in this paper about the irreducibility of the representations in Ref. 5 which coincide with Eq. (2.5).]
[10]I. Bars, Yale Preprint YTP 81-25.
[11]M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1962); H. Weyl, *Classical Groups* (Princeton U.P. Princeton, 1946).
[12]R. C. King, J. Math Phys. **11**, 280 (1970); J. Abramsky and R. C. King, Nuovo Cimento A **67**, 153 (1970); N. El Samra and R. C. King, J. Phys. A **12**, 2305 (1979).
[13]See, e.g., W. G. McKay and J. Patera, *Tables of Dimensions, Indices and Branching Rules for Representations of Simple Lie Algebras* (Marcel Dekker, New York, 1981); B. G. Wybourne, *Symmetry Principles and Atomic Spectroscopy*, Appendix II, by P. H. Butler, on Tables of Group Properties (Interscience, New York, 1970); E. B. Dynkin, Am. Math. Soc. Tr., Ser. 2, **6**, 353 (1957); H. Freudental, Indag. Math. **16**,490 (1954).

Spectral theory for the periodic sine-Gordon equation: A concrete viewpoint

M. Gregory Forest [a)]
Department of Mathematics, Ohio State University, Columbus, Ohio 43210
and David W. McLaughlin [b),c)]
Department of Mathematics, University of Arizona, Tucson, Arizona 85721

(Received 2 January 1980; accepted for publication 8 May 1981)

A summary of the spectral theory for quasiperiodic sine- and sinh-Gordon equations is given. Analogies with whole-line solitons and scattering theory motivates the discussion. The relation between the ingredients in the inverse spectral solution of the periodic sine-Gordon equation and physical characteristics of sine-Gordon waves is emphasized. The explicit topics covered are summarized in the table of contents in the Introduction.

PACS numbers: 02.30. + g

I. INTRODUCTION

The inverse scattering transform has been an extremely useful tool in the study of solitons. Its utility results, to a large extent, from (i) the practical importance of these isolated, localized nonlinear pulses and, at the same time, (ii) the precise yet simple description of these isolated pulses by the scattering transform. As a result, sparse configurations of solitons in isolation from each other are now very well understood.

Frequently, however, physical situations arise which involve a high density of solitons. For example, finite-length nonlinear oscillators such as Josephson oscillators[1] or oscillators in condensed matter physics[2] often contain many densely packed solitons. Another example is the dispersive smoothing of shock waves. Molecular dynamics simulations of shocks in conservative lattices[3] show the shock is smoothed by a high frequency nonlinear wave train which can be interpreted as a dense configuration of many solitons. When the density of solitons is high, their tails interact; they are certainly not in isolation.

For systems with a high density of solitons, one desires to compute either their deterministic or statistical behavior. The only mathematical transform for such calculations which currently exists is the "*inverse spectral transform* under periodic boundary conditions." Although it is just as precise as the scattering transform, the periodic spectral transform has not been used in many applications, primarily because of its complicated description. In this series of papers, we use the inverse spectral solutions of the periodic sine-Gordon equation to study several concrete problems. Our intent in this series is to describe and use the spectral transform in a style which should make it easier for scientists to apply the transform to their own studies of dense collections of solitons.

This first paper describes the present status of the particular periodic spectral transform which is appropriate for the sine-Gordon equation. In subsequent papers, we use this background to develop a Hamiltonian description of slowly modulating N-phase wave trains, verify the validity of this Hamiltonian description of wave trains using a formulation of the transform in terms of differentials on a Riemann surface (see Ref. 4), and use the spectral transform to investigate finite-length nonlinear oscillators.

This first paper summarizes the present status of the spectral solution of the periodic sine-Gordon equation. Although we do present some new material [in particular, evidence of spines in the spectrum (Sec. III), a representation of single phase radiation as a sequential translation of kink–antikink components (Sec. V), an explicit numerical calculation of the paths of the μ coordinates in a degenerate situation (Sec. VI), and a spectral characterization of "separable solutions" (Sec. VII)] our main goal is to present a concrete description of the periodic spectral transform for the sine-Gordon equation. In each section we summarize those theoretical facts which we use throughout the series; and, in considerable detail, we illustrate these facts and their consequences with concrete examples. We emphasize the physical information which is carried by the spectral data such as the manner in which this data classifies the elementary excitations (kink trains, breather trains, radiation) which are present in the wave and the manner in which it specifies their physical characteristics (frequencies, wave numbers, etc.). Sections II, III, and IV provide general information about waves which contain arbitrary numbers of basic excitations, while the last sections describe specific, detailed information about waves which contain one and two basic excitations.

We emphasize that our viewpoint in this paper is considerably different from the viewpoints presented in the theoretical papers on the periodic spectral transform. We argue formally, using scattering theory and solitons in isolation to motivate and interpret the periodic theory. We assume the reader is familiar with sine-Gordon solitons and their scattering descriptions, as developed by Takhatajian and Faddeev,[5] at least to the extent summarized in the Introduction and Appendix of Ref. 6. Some knowledge of the band-gap structure of the spectrum of the Schrödinger equation with periodic potential (Floquet theory of Hill's equa-

[a)]Supported in part by N.S.F. Grant No. MCS-8002969.
[b)]Supported in part by N.S.F. Grant No. MSC-7903533 and by D.O.E. (Los Alamos National Laboratory, Group T-7).
[c)] Address for the academic year 1981–1982: New York University, Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, New York 10012.

 0022-2488/82/071248-30$02.50

tion) will also be helpful.

A detailed outline is provided by the section headings.

II. PRELIMINARIES

1. Sine-Gordon spectral theory: analogies between periodic and whole-line problems

Consider the sine-Gordon equation in Laboratory coordinates,

$$u_{tt}-u_{xx}+\sin u=0, \quad -\infty<x<\infty. \tag{II.1}$$

To integrate all soliton equations, one adjoins to the nonlinear partial differential equation a linear eigenvalue problem whose "potential" is given by a solution of the soliton equation. [The most familiar example is the use of the linear Schrödinger equation $(-\psi_{xx}+u(x)\psi=E\psi)$ to integrate the nonlinear Korteweg–deVries equation[7] $(u_t-6uu_x+u_{xxx}=0)$.] The integration of the sine-Gordon equation is accomplished through the following linear eigenvalue problem:

$$\left[\begin{pmatrix}0&-1\\1&0\end{pmatrix}\frac{d}{dx}+\frac{i}{4}w\begin{pmatrix}0&1\\1&0\end{pmatrix}+\frac{1}{16\sqrt{E}}\begin{pmatrix}e^{iu}&0\\0&e^{-iu}\end{pmatrix}-\sqrt{E}\right]\psi=0, \tag{II.2}$$

where $w\equiv u_x+u_t, \psi=(\psi_1,\psi_2)^T$. Indeed, the linear system

$$\left\{\left[\begin{pmatrix}0&-1\\1&0\end{pmatrix}\frac{d}{dx}+\frac{i}{4}w\begin{pmatrix}0&1\\1&0\end{pmatrix}+\frac{1}{16\sqrt{E}}\begin{pmatrix}e^{iu}&0\\0&e^{-iu}\end{pmatrix}-\sqrt{E}\right]\psi=0,\right\} \tag{II.3a}$$

$$\left\{\left[\begin{pmatrix}0&-1\\1&0\end{pmatrix}\frac{d}{dt}+\frac{i}{4}w\begin{pmatrix}0&1\\1&0\end{pmatrix}-\frac{1}{16\sqrt{E}}\begin{pmatrix}e^{iu}&0\\0&e^{-iu}\end{pmatrix}-\sqrt{E}\right]\psi=0,\right\} \tag{II.3b}$$

which consists of the linear problem (II.2) augmented by a time flow for ψ, is compatible ($\psi_{xt}=\psi_{tx}$) if and only if the potentials u and w satisfy a sine-Gordon system,

$$u_t+u_x=w,$$
$$w_t-w_x=-\sin u.$$

In this manner, the linear system (II.3) implicitly carries the content of the sine-Gordon field as a compatibility condition. This method for the integration of the sine-Gordon equation is due to Lamb[8] and has been clarified and developed by Ablowitz, Kaup, and Newell[9] and Takhatajian and Faddeev.[5] Since we use the exposition of the latter authors, we refer to Eq. (II.2) as the "Takhatajian–Faddeev eigenvalue problem."

Notice that if the potentials (u,w) are purely imaginary,

$$u\equiv i\mathcal{U}, \quad w\equiv i\mathcal{W},$$

this sine-Gordon system becomes the sinh-Gordon system

$$\mathcal{U}_t+\mathcal{U}_x=\mathcal{W},$$
$$\mathcal{W}_t-\mathcal{W}_x=-\sinh\mathcal{U}.$$

Linear problem (II.3), with $u=i\mathcal{U}, w=i\mathcal{W}$, will integrate this sinh-Gordon system. As we will describe in the text, the mathematical properties of the sinh-Gordon problem are far simpler and better understood than those of the sine-Gordon problem.

The origin of the sine-Gordon equation (II.1) as a compatibility condition for the system (II.3) is a fact which is local in (x,t), independent of boundary conditions. Once boundary conditions are imposed upon the sine-Gordon equation, the potentials u and w of the Takhatajian–Faddeev eigenvalue problem (II.2) inherit the same boundary conditions. This boundary behavior of the potentials fixes the type

of spectrum which the eigenvalue problem (II.2) possesses. The nature of this spectrum, in turn, is related to the classes of fundamental excitations of the sine-Gordon field under the particular boundary conditions.

For example, consider the sine-Gordon equation for real fields which vanish at $|x| = \infty$:

$$u_{tt} - u_{xx} + \sin u = 0, \quad -\infty < x < \infty, \tag{II.4a}$$

$$u(x,t) \to 0 (\mathrm{mod} 2\pi) \text{ as } |x| \to \infty, \tag{II.4b}$$

$$u_t(x,t) \to 0 \text{ as } |x| \to \infty. \tag{II.4c}$$

It is well known that the fundamental excitations in this case consist of solitons (kinks), antisolitons (antikinks), breathers, and radiation. (See Fig. 12 and Ref. 6 for a detailed discussion.) When the potentials $u(x)$ and $w(x)$ of the Takhatajian–Faddeev eigenvalue problem satisfy these vanishing boundary conditions, this eigenvalue problem, viewed over the entire x axis, has continuous spectrum filling the positive E axis, and discrete bound states (eigenvalues) which occur either at points on the negative real E axis or in complex conjugate pairs [see Fig. 2(b)]. The continuous spectrum for $E > 0$ is related to radiation degrees of freedom for the sine-Gordon field, the negative eigenvalues $E_j < 0$ label kink or antikink solutions, and the conjugate bound state pairs E_j, E_j^* label breather excitations. *As the potentials u,w flow in time according to the sine-Gordon equation, the locations of the bound state eigenvalues remain fixed and determine the speeds and widths of the solitons and the frequencies of the breathers.* This temporal invariance of the spectrum is central to the integration of the sine-Gordon equation, which is said to "generate an *isospectral flow*" for the Takhatajian–Faddeev eigenvalue problem.

In this paper, we are concerned with the sine-Gordon equation under *periodic boundary conditions*, with period L [10]:

$$u_{tt} - u_{xx} + \sin u = 0, \quad -\infty < x < \infty, \tag{II.5a}$$

$$e^{iu(x+L,t)} = e^{iu(x,t)}, \tag{II.5b}$$

$$u_t(x+L,t) = u_t(x,t). \tag{II.5c}$$

When the potentials u and w are real and satisfy these periodic boundary conditions, the spectrum of (II.2), still defined over the whole x axis, is continuous spectrum. This spectrum consists of curves (or bands) which we will interpret as follows.

(i) The discrete bound states for potentials with vanishing boundary conditions have spread into narrow bands of continuous spectrum, and (ii) short "spines" of spectrum have grown off the positive E axis [see Fig. 4(b)].

The basic excitations of the real periodic sine-Gordon field turn out to consist of periodic trains of kinks or antikinks (associated with the narrow bands of spectrum on the negative E axis), trains of breathers (associated with the narrow bands in conjugate pairs), and radiation which behaves much as Fourier modes for the linear Klein–Gordon field (and is presumably associated with the spines off the positive real axis). The isospectrality is now manifested by the invariance of these bands of spectrum under the periodic sine-Gordon flow. This spectral classification of the fundamental excitations in the periodic sine-Gordon field will be established in the following sections. We emphasize that, to be useful in applications, the correspondence between the location of the spectrum and physical characteristics of the potentials must be concretely understood; the manner by which some of these detailed facts can be obtained is described in the body of the paper and in paper III.

With these remarks in mind, we view the Takhatajian–Faddeev eigenvalue problem (II.2) as a family of eigenvalue problems, defined over the entire x axis. This family, indexed by a time parameter t, is generated by the periodic sine-Gordon equation (II.5). Since the spectrum σ is invariant in t, σ is determined for all time by the periodic initial data $\mathring{u},\mathring{v}$,

$$u(x,t=0) \equiv \mathring{u}, \quad u_t(x,t=0) \equiv \mathring{v}(x),$$

$$e^{i\mathring{u}(x+L)} = e^{i\mathring{u}(x)},$$

$$\mathring{v}(x+L) = \mathring{v}(x),$$

via

$$\left[\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \frac{d}{dx} + \frac{i}{4}(\mathring{v} + \mathring{u}_x) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \frac{1}{16\sqrt{E}} \begin{pmatrix} e^{i\mathring{u}} & 0 \\ 0 & e^{-i\mathring{u}} \end{pmatrix} - \sqrt{E} \right] \psi = 0. \tag{II.6}$$

We begin with a discussion of the general properties of the spectrum σ, and later consider several important special cases of $\mathring{u},\mathring{v}$.

2. Transfer matrix, Floquet discriminant, and Floquet solutions

First, we define the tools which are used to display the spectrum σ, namely, the *transfer matrix*, *Floquet discriminant*, and *Floquet solutions*. Fix a point x_0, and a basis of solutions of (II.6), $\{\phi_+(x,x_0,E), \phi_-(x,x_0,E)\}$, by the following initial conditions at $x = x_0$:

$$\phi_+(x=x_0,x_0,E) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \phi_-(x=x_0,x_0,E) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Notice that $\phi_\pm(x+L,x_0,E)$ are also solutions of (II.6); this follows quickly from the periodicity of $\mathring{w}$, $\exp(\pm i\mathring{u})$. Therefore, we can expand these "new" solutions on the basis $\phi_\pm(x,x_0,E)$:

$$\phi_+(x+L,x_0,E) = t_{11}(E)\phi_+(x,x_0,E) + t_{12}(E)\phi_-(x,x_0,E),$$

$$\phi_-(x+L,x_0,E) = t_{21}(E)\phi_+(x,x_0,E) + t_{22}(E)\phi_-(x,x_0,E),$$

or more concisely,

$$\begin{pmatrix} \phi_+(x+L,x_0,E) \\ \phi_-(x+L,x_0,E) \end{pmatrix} = T(E) \begin{pmatrix} \phi_+(x,x_0,E) \\ \phi_-(x,x_0,E) \end{pmatrix}. \tag{II.7}$$

The (2×2) "transfer matrix" $T(E)$, defined by (II.7), transfers the basis $\phi_\pm(x,x_0,E)$ across one period L of the potentials $\mathring{u},\mathring{w}$. Iterating this formula (II.7) across N periods yields

$$\begin{pmatrix} \phi_+(x+NL,x_0,E) \\ \phi_-(x+NL,x_0,E) \end{pmatrix} = T^N(E) \begin{pmatrix} \phi_+(x,x_0,E) \\ \phi_-(x,x_0,E) \end{pmatrix}. \tag{II.8}$$

By definition, a complex number E belongs to the spectrum σ of the Takhatajian–Faddeev eigenvalue problem (II.6) if and only if the solutions $\phi_\pm(x,x_0,E)$ are bounded for all x,[11] or equivalently, if and only if $\phi_\pm(x+NL,x_0,E)$ are bounded for all integers N. From (II.8), we see that such bounded behavior for large N is possible if and only if both eigenval-

ues, $\rho_\pm(E)$, of the transfer matrix $T(E)$ have unit modulus. Thus we obtain

Fact 1: $E\in\sigma$ if and only if $|\rho_\pm(E)| = 1$, where $\rho_\pm(E)$ are the eigenvalues of the transfer matrix defined as the two roots of

$$\det[T(E) - \rho(E)I] = 0.$$

Computing this determinant yields

$$\rho^2(E) - \Delta(E)\rho(E) + 1 = 0, \tag{II.9}$$

where $\Delta(E) \equiv$ trace of $T(E)$. The roots are given explicitly by

$$\rho_\pm(E) = \{\Delta(E) \pm [\Delta^2(E) - 4]^{1/2}\}/2. \tag{II.10}$$

The function $\Delta(E)$ is known as the *Floquet discriminant*, and is central to the entire theory. From (II.9) and (II.10) it follows that

$$\rho_-(E)\rho_+(E) = 1,$$
$$\Delta(E) = \rho_-(E) + \rho_+(E),$$

so that $|\rho(E)| = 1$ if and only if $\Delta(E)$ is real, with $|\Delta(E)|\leqslant 2$. We therefore have

Fact 2: $E\in\sigma$ iff $\Delta(E)$ is real, with $|\Delta(E)|\leqslant 2$.

Remark. Note that the eigenvalue problem (II.6) with purely imaginary $u = i\mathscr{U}$, $w = i\mathscr{W}$ is self-adjoint; hence, its spectrum σ is real. On the other hand, when the potentials u and w are real, the eigenvalue problem is *not* self-adjoint. In this case, the spectrum σ need not be real; nevertheless, σ does lie on curves of real $\Delta(E)$.

Next, we change basis from $\{\phi_\pm(x,x_0,E)\}$ to one in which the transfer matrix $T(E)$ becomes diagonal. This new basis is called the *Floquet basis* and consists of the *Floquet solutions*, $\{\psi_\pm(x,x_0,E)\}$ defined by

$$\psi_\pm(x + L,x_0,E) = \rho_\pm(E)\psi_\pm(x,x_0,E). \tag{II.11}$$

The eigenvalues $\rho_\pm(E)$ are called the *Floquet multipliers*; the representation (II.11) clearly indicates that Floquet theory is a natural generalization of strictly periodic eigenfunctions ($\rho = 1$) and can be used to develop generic properties of the spectrum σ.

3. Generic properties of the spectrum σ

Using these ingredients [transfer matrix $T(E)$; Floquet discriminant $\Delta(E)$; Floquet multipliers $\rho_\pm(E)$; the basis $\phi_\pm$; Floquet basis $\psi_\pm$], we can summarize the generic properties of the spectrum σ in

Theorem II.1: *Useful properties of the spectrum* σ:

(i) $E\in\sigma$ iff $|\rho(E)| = 1$

iff $\Delta(E)$ is real and $|\Delta(E)|\leqslant 2$.

(ii) The spectrum σ is continuous spectrum. It consists of a countable number of smooth curves, called *bands of spectrum*; the endpoints of these curves (except $E = 0$ and $E = \infty$) are simple periodic and antiperiodic [$\psi(x + L) = -\psi(x)$] eigenvalues, and comprise the *simple spectrum*.

(iii) E is a *periodic* (*antiperiodic*) *eigenvalue*

iff $\Delta(E) = +2(-2)$.

(iv) For *real* potentials $\mathring{u}$ and $\mathring{w}$, the following symmetries hold: (a) If $E\in\sigma$, then $E^*\in\sigma$.[12]

(b) If E is a periodic (antiperiodic) eigenvalue, then E^* is a periodic (antiperiodic) eigenvalue.

(v) For purely *imaginary* potentials, $E\in\sigma \Rightarrow E = E^* > 0$.

Proof: Part (i) is clear from Sec. II.2. However, it is instructive to iterate formula (II.11) N periods to the left and right:

$$\psi_\pm(x + NL,x_0,E) = [\rho_\pm(E)]^N\psi_\pm(x,x_0,E),$$
$$\psi_\pm(x - NL,x_0,E) = [\rho_\pm(E)]^{-N}\psi_\pm(x,x_0,E). \tag{T1}$$

It then becomes obvious that for $|\rho_\pm| \neq 1$, the corresponding eigenfunctions must blow up either at $x = +\infty$ or $x = -\infty$. For $|\rho_\pm| = 1$, (T1) shows that the eigenfunctions remain bounded for all x. Since

$$\Delta(E) = \rho_+(E) + \rho_-(E) = \rho_+(E) + [\rho_+(E)]^{-1},$$

it immediately follows from $|\rho_+| = 1$ that Δ is real and $|\Delta(E)|\leqslant 2$. Part (i) is thereby proved.

To prove Part (iii), we use (T1) with $N = 1$. Clearly $\rho = +1(-1)$ yields periodic (antiperiodic) Floquet solutions; since $\rho = \{\Delta(E) + [\Delta^2(E) - 4]^{1/2}\}/2$, $\Delta(E) = +2(-2)$ determines these periodic (antiperiodic) eigenvalues.

For Part (ii), since σ is characterized by $|\rho| = 1$, from (T1) we see that $|\psi_\pm|$ repeats the same values over each period L. It follows that while bounded for all x, $\psi_\pm$ cannot vanish as $|x|\to\infty$; thus $\psi_\pm$ are generalized eigenfunctions and σ is continuous spectrum. To see that σ lies in *bands*, we use the basic fact, without proof, that $\Delta(E)$ is an analytic function of E except for essential sigularities at $E = 0, \infty$. [See the *Remark* at the end of this section.] By Part (i), σ is characterized by $\Delta(E)$ real, $|\Delta(E)|\leqslant 2$; $\Delta(E)$ analytic then yields that σ lies on smooth curves of $\mathrm{Im}(\Delta(E)) = 0$, terminating only at those points where $\Delta(E) \pm 2$, $\Delta'(E) = 0$, that is, the *simple spectrum*. Thus Part (ii) is established.

To verify Part (iv), let

$$\psi(x,x_0,E) = \begin{pmatrix}\psi_1(x,x_0,E)\\ \psi_2(x,x_0,E)\end{pmatrix}$$

denote a Floquet solution of the eigenvalue problem (II.6) at $E\in\sigma$. Simply inserting into (II.6) shows that $(\psi_2^*(x,x_0,E), -\psi_1^*(x,x_0,E))^T$ is a Floquet solution at E^* if we assume the reality of $\mathring{u}(x)$ and $\mathring{w}(x)$. From the hypothesis $E\in\sigma,\psi_1,\psi_2$ and likewise ψ_2^*, $-\psi_1^*$ are bounded: $E^*\in\sigma$. The periodicity (anti) follows immediately upon inspection. Part (v) was noted earlier.

We close this section with a technical representation of the discriminant $\Delta(E)$, which is extremely useful for displaying qualitative features of the spectrum σ. In the basis $\{\phi_\pm(x,x_0,E)\}$, we have the following

Theorem II.2: *Eigenfunction representation of* $\Delta(E)$: The Floquet discriminant $\Delta(E)$ can be represented as

$$\Delta(E) = \phi_{+,1}(x_0 + L,x_0,E) + \phi_{-,2}(x_0 + L,x_0,E). \tag{II.12}$$

Proof: In this basis $\{\phi_\pm(x,x_0,E)\}$, we have

$$\begin{pmatrix}\phi_+(x + L,x_0,E)\\ \phi_-(x + L,x_0,E)\end{pmatrix} = T(E)\begin{pmatrix}\phi_+(x,x_0,E)\\ \phi_-(x,x_0,E)\end{pmatrix}.$$

Setting $x = x_0$ and using the initial conditions for $\phi_{\pm}$ at $x = x_0$ yields

$$\phi_+(x_0 + L, x_0, E) = t_{11}(E)\begin{pmatrix}1\\0\end{pmatrix} + t_{12}(E)\begin{pmatrix}0\\1\end{pmatrix},$$

$$\phi_-(x_0 + L, x_0, E) = t_{21}(E)\begin{pmatrix}1\\0\end{pmatrix} + t_{22}(E)\begin{pmatrix}0\\1\end{pmatrix},$$

from which we find

$$T(E) = \begin{bmatrix}\phi_{+,1}(x_0 + L, x_0, E) & \phi_{+,2}(x_0 + L, x_0, E)\\ \phi_{-,1}(x_0 + L, x_0, E) & \phi_{-,2}(x_0 + L, x_0, E)\end{bmatrix}.$$

Since $\Delta(E) =$ trace of $T(E)$, we have (II.12).

Remark: Since $\phi_{\pm}(x, x_0, E)$ satisfies an initial-value problem with $(1,0)^T$, $(0,1)^T$ initial data, it can be established by a Picard iteration scheme that $\phi_{+,1}(x, x_0, E)$ and $\phi_{-,2}(x, x_0, E)$ (and $E^{1/2}\phi_{+,2}$, $E^{1/2}\phi_{-,1}$) are analytic functions of E except for essential singularities at $E = 0, \infty$ (check, for example, the special cases $\mathring{u} = \mathring{w} \equiv 0$). Hence Eq. (II.12) shows that the Floquet discriminant $\Delta(E)$ enjoys the same analyticity properties.

III. QUALITATIVE INSIGHT INTO THE SPECTRUM σ

The purpose of Secs. II and III of this paper is to connect properties of the spectrum σ with the solution of the periodic initial-value problem for the sine-Gordon equation. Thus far, in Sec. II we have discussed the general nature of the spectrum σ; here we seek more detailed, qualitative information about the structure of the bands of σ. To obtain this information, we imbed a whole-line scattering problem into each period. Thereby, we provide simple derivations of properties of the Floquet spectrum σ based upon better-known properties from whole-line scattering theory; we then interpret the relation of this information about the band structure of σ to solutions of the sine-Gordon equation.

With these remarks in mind, we specialize to potentials $\mathring{u}(x), \mathring{w}(x) = \mathring{u}_x(x) + \mathring{v}(x)$ such that $\mathring{u}(x)$ has compact support (mod 2π if $\mathring{u}$ is real) within each period, and $\mathring{w}(x)$ has compact support within each period. We call such potentials *truncated potentials*. (See Fig. 1.) Although our derivations are restricted to this class of truncated potentials, the specific properties which we obtain are in fact more general. We emphasize that this class contains potentials which are far from the vacuum ($\mathring{u} = \mathring{w} \equiv 0$); thus, our representation gives exact spectral information about sine-Gordon field configurations which contain arbitrary numbers of the basic excitations (periodic trains of kinks and antikinks, breathers and radiation). For convenience, we summarize useful whole-line scattering properties in Appendix A.

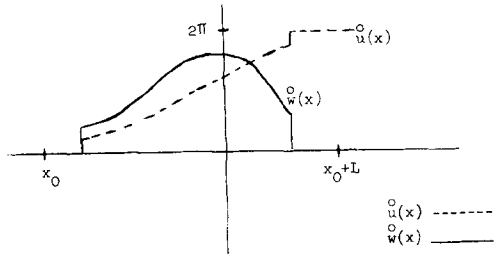


FIG. 1. Truncated potentials.

The ordering for the remainder of Sec. III is to first use infinite-line scattering theory to arrive at a "scattering representation" of the discriminant $\Delta(E)$, from which we deduce the general qualitative structure of the spectrum; we relate this structure to the fundamental excitations of the sine-Gordon field; finally, we summarize the spectrum appropriate to the sinh-Gordon field.

Theorem III.1: *Scattering representation of the Floquet discriminant* $\Delta(E)$:

Let $\mathring{u}(x)$ and $\mathring{w}(x)$ denote truncated potentials (Fig. 1). In terms of the scattering parameter[13] $\lambda, \lambda^2 \equiv E$,

$$\Delta(\lambda) = a(\lambda)e^{-i\alpha(\lambda)L} + \bar{a}(\lambda)e^{i\alpha(\lambda)L}, \quad \mathrm{Im}\lambda \geq 0. \qquad \text{(III.1)}$$

Here $\alpha(\lambda) \equiv \lambda - 1/16\lambda$ and $[a(\lambda)]^{-1}$ denotes the transmission coefficient across the basic period of the potentials, with the potentials continued beyond this basic period by

$$e^{i\mathring{u}(x)} \equiv 1, \quad \mathring{w}(x) \equiv 0 \quad \text{outside} \quad [x_0, x_0 + L].$$

For λ real, representation (III.1) reduces to

$$\Delta(\lambda) = 2|a(\lambda)|\cos\{\alpha(\lambda)L - \mathrm{ph}[a(\lambda)]\}, \quad \lambda \text{ real}, \qquad \text{(III.2)}$$

where $\mathrm{ph}[a(\lambda)]$ denotes the phase of $a(\lambda)$.

We present the proof of Theorem III.1 in Appendix A so as not to delay the exposition. Here, we merely summarize those properties of $a(\lambda)$ which are needed to use the scattering representation of $\Delta(\lambda)$.

Theorem III.2: *Facts from whole-line scattering theory*:

(i) $[a(\lambda)]^{-1} \equiv$ transmission coefficient.

(ii)

$$a(\lambda) \to \begin{cases} 1 & \text{as } \lambda \to \infty, \mathrm{Im}\lambda \geq 0 \\ \exp[(i/2)(\mathring{u}(+\infty) - \mathring{u}(-\infty))] & \text{as } \lambda \to 0. \end{cases}$$

(iii) $|a(\lambda)|^2 \pm |b(\lambda)|^2 = 1, \lambda$ real. Here $+$ is for real u, w and $-$ is for purely imaginary u, w.

(iv) $a(\lambda_j) = 0$ if and only if λ_j is a bound state eigenvalue, $j = 1, 2, \ldots, N$. For purely imaginary potentials $u = i\mathscr{U}$, $w = i\mathscr{W}$, no bound states exist. For real potentials, these bound states occur either on the positive imaginary λ axis (in which case they are associated with kinks), or in pairs $(\lambda_j, -\lambda_j^*)$ (in which case they are associated with breathers).

(v) With real potentials, $a(\lambda)$ can possess multiple zeros. In this case, small perturbations of the potentials will remove this degeneracy.

We summarize these properties in Fig. 2 in both the λ and E planes.

We now deduce from the scattering representation of Δ (Theorem III.1) a series of facts about the spectrum σ. These facts will be phrased in the E plane. We begin with

Fact 1: For real potentials the positive real E axis ($E > 0$) is continuous spectrum, with no gaps in σ on the positive real E axis. (This part of the spectrum is associated with radiation in the sine-Gordon field.)

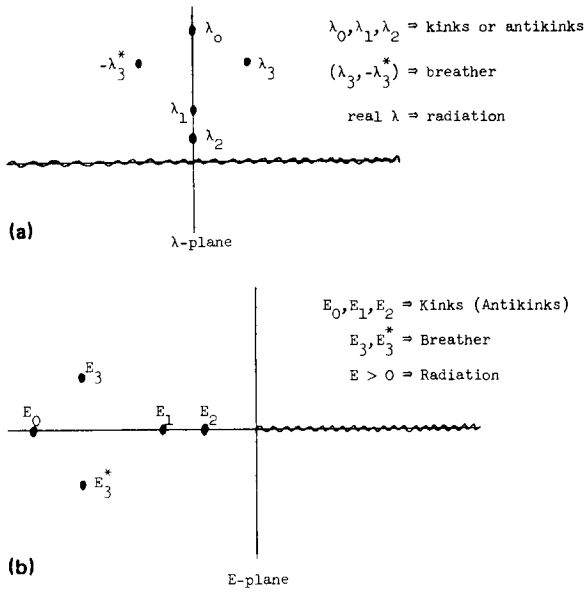


FIG. 2. Spectrum σ for whole-line problem with vanishing boundary conditions.

This fact is obvious from the scattering representation of Δ. For $E>0$, or equivalently for real λ,

$$|a(\lambda)|^2 + |b(\lambda)|^2 = 1 \Rightarrow |a(\lambda)| \leqslant 1 \Rightarrow |\Delta| \leqslant 2.$$

Since Δ is real and always bounded in magnitude by 2, Theorem III.1 establishes Fact 1.

Next, consider the special case $\mathring{u}(x) = \mathring{w}(x) \equiv 0$. In this case of zero potentials, the scattering representation of Δ, together with the total transmission $[a(\lambda) = \bar{a}(\lambda) \equiv 1]$ property of zero potentials, gives

$$\Delta(\lambda) = e^{-i\alpha(\lambda)L} + e^{i\alpha(\lambda)L},$$

where $\mathrm{Im}\lambda \geqslant 0$, $\alpha(\lambda) \equiv \lambda - 1/16\lambda$. When $\lambda = -\lambda^*$,

$$\Delta(\lambda) = 2\cosh\left[\left(|\lambda| + \frac{1}{16|\lambda|}\right)L\right];$$

when λ is real,

$$\Delta(\lambda) = 2\cos[\alpha(\lambda)L].$$

From these formulas we deduce (see Fig. 3)

Fact 2: For $\mathring{u}(x) = \mathring{w}(x) \equiv 0$,

(a) $|\Delta(E)| \to \infty$ exponentially as $|E| \to \infty, E \notin (0, \infty)$.

(b) For $E<0$, $\Delta(E)$ is real with $\Delta(E) \to \infty$ exponentially as $E \to 0, \infty$.

(c) For $E>0$, the oscillations of $\Delta(E)$ always reach ± 2; all periodic and antiperiodic eigenvalues are double; $\Delta(E) \sim 2\cos(E^{1/2}L)$ as $E \to +\infty$, and its graph appears very regular near $E = +\infty$.
$\Delta(E) \sim 2\cos(L/16\sqrt{E})$ as $E \to 0^+$, and its graph appears very dense near $E = 0, E>0$.

For arbitrary real potentials $(\mathring{u}, \mathring{w} \neq 0)$, $\Delta(E)$ asymptotes to these formulas as $E \to 0, \infty$, with the following exception near $E = 0$. First, there is a possible phase shift of π as $E \to 0, E>0$. With the "charge" M defined $\mathring{u}(x_0 + L) = \mathring{u}(x) + 2M\pi$, the phase shift is zero for M even and π for M odd. That is, for general potentials $\mathring{u}, \mathring{w}$,

$$\Delta(E) \sim (-1)^M 2\cos(L/16\sqrt{E}) \quad \text{near} \quad E = 0, E>0,$$

$$\Delta(E) \sim 2\cos(E^{1/2}L) \quad \text{near} \quad E = +\infty, E \quad \text{real}.$$

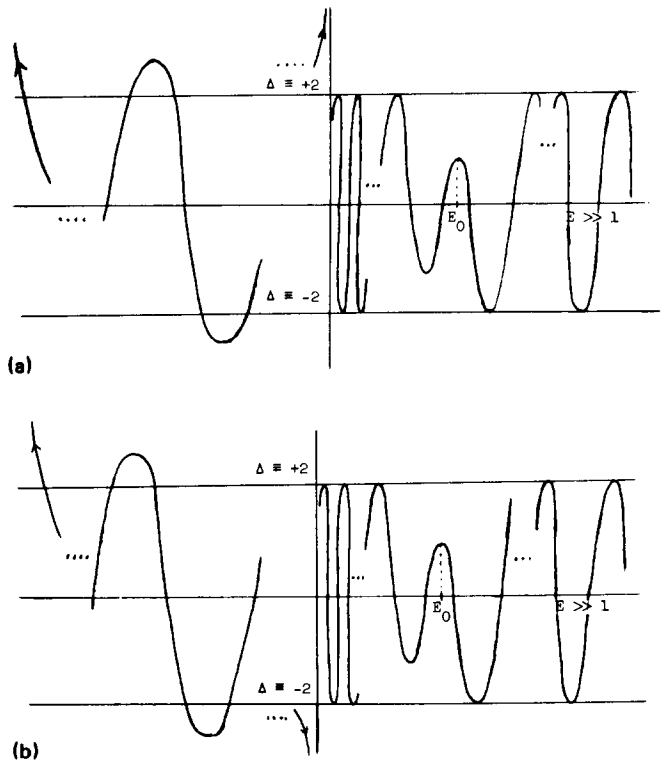


FIG. 3. Graph of $\Delta(E)$ vs E real. $\mathring{u}(x_0 + L) = \mathring{u}(x_0) + 2M\pi$, (a) M even; (b) M odd.

Second is the effect of the "charge" of $\mathring{u}(x)$ on the graph near $E = 0, E<0$:

$$\lim_{E \to 0^-} \Delta(E) = \begin{cases} +\infty, & M \text{ even} \\ -\infty, & M \text{ odd} \end{cases}.$$

With these remarks, we can sketch $\Delta(E)$ vs E real (Fig. 3).

From the scattering representation for real potentials, we have found the entire positive real E axis is continuous spectrum. There is an interesting ramification of this fact.

Fact 3: For real potentials, the spectrum σ has *spines* coming off the positive real E axis (see Fig. 4).

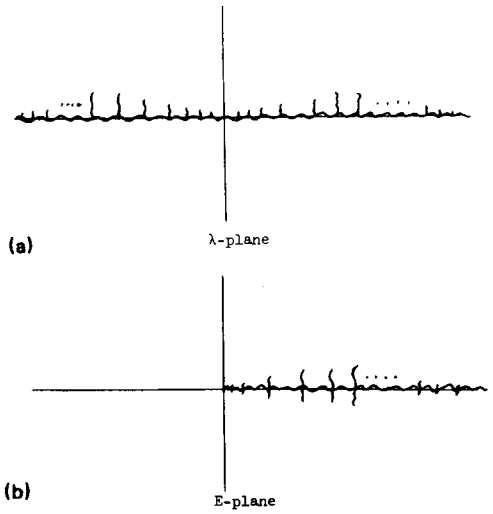


FIG. 4. Spines in the Floquet spectrum σ.

The proof of Fact 3, a rather surprising result in sine-Gordon spectral theory, is an elementary exercise in complex function theory.[14,15] The gist of the argument is as follows. Refer to the graph of $\Delta(E)$ vs E real, $E>0$ (Fig. 3), and consider a local maximum $(E_0,\Delta(E_0))$, where $|\Delta(E_0)|<2$ and $\Delta'(E_0)=0$. [Remember $\Delta(E)$ is real for E real and the spectrum σ lies on curves of $\mathrm{Im}\Delta(E)\equiv 0$, with $|\Delta|<2$.] From a more general perspective, though, $\Delta(E)\equiv\Delta_R+i\Delta_I$ is an analytic function of the complex variable $E=E_R+iE_I$. Thus, $\Delta_R(E_R,E_I)$ is harmonic and E_0 must be a saddle point on the surface $z\equiv\Delta_R(E_R,E_I)$. Thus, leaving E_0 along the real E axis one descends the saddle along a curve of steepest descent. But since E_0 is a saddle point, there is another direction off the real axis into the complex E plane which is a steepest ascent path. (In fact, the curve is determined by $\Delta_I\equiv 0$.) Then by continuity (since $|\Delta(E_0)|<2$), this curve gives a short band of spectrum, a *spine*. Clearly the same holds true at a local minimum E_0, where $|\Delta(E_0)|<2$.[16]

As $E\to 0,+\infty$, the behavior of $\Delta(E)$ shows the length of these spines must asymptote to zero. They tend to be very short anyway; this seems intuitively clear from the above arguments since a spine is a steepest descent (ascent) path of Δ_R and so should exceed 2 in magnitude very quickly (for moderate period L). In fact, Eq. (III.1) for $\Delta(\lambda)$, $\mathrm{Im}\lambda\geqslant 0$, shows $|\Delta(\lambda)|$ grows exponentially as a function of $\lambda_I\geqslant 0$:

$$\Delta(\lambda)=a(\lambda)e^{-i\alpha(\lambda)L}+\bar{a}(\lambda)e^{i\alpha(\lambda)L},$$

from which we have

$$|\Delta(\lambda)|\sim|a(\lambda)|e^{\lambda_I L(1+1/16|\lambda|^2)},\quad \lambda_I>0. \tag{III.3}$$

This completes our discussion of the part of the spectrum σ which is connected to the positive real E axis.

For real potentials, there also exists parts of the spectrum which are not connected to the positive real E axis. These are the periodic generalizations of kinks (antikinks) and breathers. From the exponential growth of $\Delta(\lambda)$ for $\lambda_I>0$, (III.3), we realize any such bands of spectrum in the upper-half λ plane, or off the positive real E axis, are very short and sparse. They originate from a different type of oscillation in $\Delta(\lambda)$, coming from zeros of $a(\lambda)$, rather than from sinusoidal oscillations of $\exp[i\alpha(\lambda)L]$.

Assume we have some curve $\Delta_I\equiv 0$ in the upper-half λ plane. If there is to be any spectrum along such a curve, we must have $-2\leqslant\Delta_R\leqslant 2$. However, from the exponential behavior of $\Delta(\lambda)$ for $\lambda_I>0$, Eq. (III.3), $|\Delta|=|\Delta_R|$ will usually exceed 2 unless $a(\lambda)$ has a zero very near the curve. The zeros of $a(\lambda)$ are finite in number and occur only at the bound states of the truncated potentials $\mathring{u}(x),\mathring{w}(x)$. Usually zeros of $a(\lambda)$ are isolated and, $|\Delta(\lambda)|$ will exceed 2 very quickly. Thus, there will be a short band of spectrum near the zeros of $a(\lambda)$.

These facts, together with the symmetries of Theorem (II.1), yield the rather generic spectral profile of Figs. 5. (Compare Figs. 2.)

As the period L becomes infinite, the spines go away and the entire positive real E axis remains continuous spectrum (radiation in the sine-Gordon field), the conjugate pair of bands in the E plane shrinks to conjugate poles (a breather), and the three bands on the negative real E axis shrink to poles and describe three kinks (or antikinks). [Refer to Theorem III.2, Part (iv).] Thus, Figs. 5 collapse to Figs. 2 in the infinite-period limit. Based on these whole-line analogies for the limiting configurations, we see the conjugate pair of bands must be related to trains of breathers while the bands on the negative real E axis correspond to trains of kinks or antikinks.

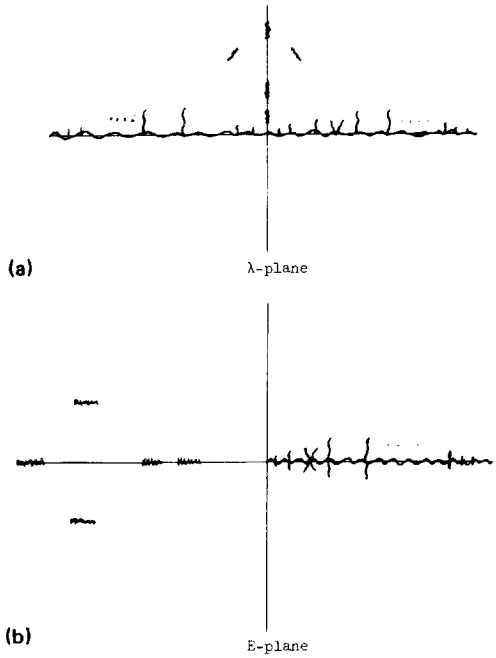


FIG. 5. Generic profile of the Floquet spectrum σ.

Fact 4: For real potentials, the eigenvalue problem (II.2) is *not* self-adjoint. This can lead to rather eccentric spectral properties; for example, $a(\lambda)$ can have multiple zeros. However, such properties do not appear to be generic; arbitrarily small changes in the potentials remove them.

This completes our use of the "scattering representation" to obtain properties of the spectrum σ for real potentials. We now turn to purely imaginary potentials, which are appropriate for the sinh-Gordon field. In this case, the spectrum σ is much simpler because the eigenvalue problem (II.2) is self-adjoint. We summarize its properties in

Fact 5: For purely imaginary potentials, the spectrum σ consists of bands on the positive E axis, separated by gaps. These gaps on the real axis become very narrow as $(\sqrt{E}-1/16\sqrt{E})\to\infty$; that is, as $E\to 0^+$ and as $E\to+\infty$.

These properties of the spectrum for purely imaginary potentials are established from the scattering representation of $\Delta(E)$, with arguments which exactly parallel the case of real potentials. The only difference is the sign in $|a^2|=1\mp|b^2|$ (Theorem III.2). The lower sign results in gaps on the positive real axis; the upper sign leads to spines in the spectrum.

Notice that the spectrum for purely imaginary potentials is directly analogous to the well-known spectrum of Hill's equation. This simplicity of the spectrum for self-adjoint linear problems should be contrasted with the complicated spectral structure of the non-self-adjoint linear problem for the real sine-Gordon equation.

IV. FINITE DEGREES OF FREEDOM

1. $2N$ invariant simple spectra $\{E_j\}$; determination of a Riemann surface

Material in previous sections indicates that the elementary excitations of the periodic sine-Gordon field consist of trains of kinks, trains of breathers, and trains of radiation. (Of course, the distinctions may not be as well defined as in the whole-line case.) For the remainder of the paper, we seek detailed qualitative and quantitative connections between the spectrum σ and the physical characteristics of these basic excitations, such as frequencies, dispersion relations, and amplitudes. First, we use the invariant spectrum σ to fix a Riemann surface; then we use this Riemann surface to construct special solutions which consist of a finite number of elementary excitations. These special finite degree of freedom solutions admit several equivalent representations, each of which has its own utility. In this section we construct two of these [the "μ-representation" of Eqs. (IV.3) and (IV.4) and the "theta function representation" of Eq. (IV.20)] and discuss their properties. The Riemann surface enters these constructions as follows.

We begin by seeking all solutions of the sine-Gordon equation with a prescribed spectrum σ. Formulas such as

$$\rho(E) = \tfrac{1}{2}\{\Delta(E) + [\Delta^2(E) - 4]^{1/2}\}$$

for the Floquet multiplier $\rho(E)$ indicate that such potentials will be fundamentally related to the analyticity structure of the function

$$[\Delta^2(E) - 4]^{1/2} = [(\Delta(E) - 2)(\Delta(E) + 2)]^{1/2}.$$

This function has branch points at $E = 0, \infty$,[18] and at the simple periodic and antiperiodic eigenvalues $\{E_j\}$. Since the spectrum σ consists of bands that terminate precisely at the branch points of $(\Delta^2 - 4)^{1/2}$, the branch cut structure of $(\Delta^2 - 4)^{1/2}$ can be chosen to coincide with the spectrum σ.[19] Thus, the appropriate function theory of the sine-Gordon equation stems from the two-sheeted Riemann surface of $[\Delta^2(E) - 4]^{1/2}$: two copies of the E plane connected along cuts between branch points $\{E_j\}$ (including $0, \infty$) consistent with the constraints of Theorem II.1. For example, for real potentials,

$$E_j \text{ a branch point} \Rightarrow E_j^* \text{ a branch point.} \quad \text{(IV.1)}$$

Since there are, in general, infinitely many simple periodic and antiperiodic eigenvalues, $[\Delta^2(E) - 4]^{1/2}$ will have infinitely many branch points, leading to a Riemann surface of infinite genus.

However, the special case of exactly $2N$ *simple* zeros of $\Delta^2(E) - 4$ provides sine-Gordon fields which contain a finite number (N) of basic excitations. In that case an infinite product expansion yields (a derivation in Paper II)

$$[\Delta^2(E) - 4]^{1/2} = C \prod_{j > 2N+1} \left(1 - \frac{E}{E_j}\right) \prod_{j<0} \left(1 - \frac{E_j}{E}\right) \times \left[\prod_{k=1}^{2N+1} (E - E_k)\right]^{1/2};$$

the number of branch points is finite, leading to a finite genus Riemann surface. (This case of N degrees of freedom is potentially useful in applications. For example, consider a study of the sine-Gordon equation with arbitrary periodic initial data. First, spectral analyze the data to determine the structure and location of the spectrum σ. Next, retain only the "most dominant" bands[17] to obtain a finite-band approximation of σ. This finite-band spectrum then yields an approximation to the full wave in terms of a finite number of basic sine-Gordon excitations.)

With this motivation, we pose the *finite band inverse problem*: the strategy is to prescribe $(2N + 1)$ branch points, $\{E_k, k = 1,...,2N + 1\}$, consistent with the symmetries, and then construct the most general solution of the sine-Gordon equation which yields this simple spectrum. For convenience, we summarize these symmetries:

$$E_{2N+1} = 0,$$

$$\text{Real potentials} \Rightarrow \quad \text{(i)} \quad E_k \notin [0, \infty] \text{ and}$$

$$\text{(ii)} \quad E_k = E_k^* < 0 \text{ or } E_k \text{ occur in conjugate pairs}$$

$$\text{Imaginary potentials} \Rightarrow E_R = E_k^* > 0. \quad \text{(IV.2)}$$

With the simple spectrum $\{E_k\}$ fixed, consider

$$R^2(E) = \prod_{k=1}^{2N+1} (E - E_k)$$

and its two sheeted Riemann surface.

(For the case $N = 3$, this Riemann surface may be realized from the cut structure depicted in Fig. 6. Notice the branch cuts need not, in fact often do not,[20] coincide with the spectrum σ; however, they could be so chosen.) In the next sections we show that the *general* sine-Gordon solution *with this prescribed N band spectrum* is an "N-phase wave train"; these wave trains can be represented in terms of N variables, $\{\mu_1,...,\mu_N\}$, each of which moves on the Riemann surface just specified.

2. μ-representation of the N-phase wave train

We now construct a representation of the N-phase wave train. We follow an approach of Daté,[21] which is very straighforward and far easier than alternatives in the literature. First, we state the representation and then detail its construction.

Theorem IV.1: *μ-representation of the N-phase wave train*:

With the simple spectrum fixed as $\Sigma^{(s)} = \{E_k, k = 1,2,...,2N\}$, the general N-phase sine-Gordon wave train admits the representation

$$u(x,t) = i \ln\left[\left(\prod_{l=1}^{N} \mu_l(x,t)\right) \Big/ P^{1/2}\right], \quad \text{(IV.3)}$$

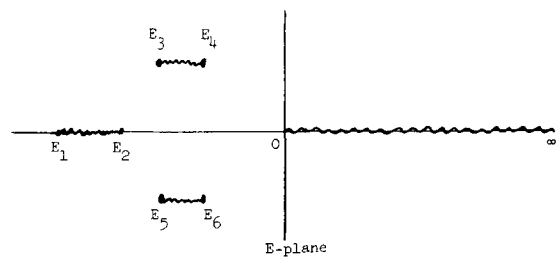


FIG. 6. Sample cut structures for genus $N = 3$.

where $\{\mu_l(x,t)\}$ satisfy the ordinary differential equations

$$(\mu_l)_x = \frac{2i\left(1 \mp \frac{1}{16P^{1/2}}\prod_{j\neq l}\mu_j\right)\left[\prod_{j=1}^{2N+1}(\mu_l - E_j)\right]^{1/2}}{\prod_{j\neq l}(\mu_l - \mu_j)}. \quad \text{(IV.4)}$$

Here $P = \Pi E_k$ and $E_{2N+1} \equiv 0$.

We will refer to this representation as the "μ-representation of $u(x,t)$." It constructs the general N-phase sine-Gordon wave from the following data: (i) the simple spectrum $\Sigma^{(s)} = \{E_k\}$ [which, along with $0, \infty$, are the branch points of $R^2(E) = (\Pi(E - E_k))$], (ii) N degrees of freedom $\{\mu_j(x,t)\}$ [which reside on the Riemann surface of $R(E)$], (iii) N branch determinations $\{R(\mu_l) = [\Pi(\mu_l - E_j)]^{1/2}$ for $l = 1,2,...,N\}$ [which fix the Riemann sheet on which μ_l initially resides], and (iv) the branch determination of $P^{1/2} = (\Pi E_k)^{1/2}$. We remark here, and will discuss at the end of this section, that problems exist in imposing the reality constraints on this representation. The constraint of purely imaginary potentials (sinh-Gordon) is easy; that of real potentials (sine-Gordon) is difficult.

Next, we derive this representation through a collection of lemmas, which serves to outline the construction. Instead of using the Takhatajian–Faddeev linear system to construct the sine-Gordon wave train, we employ an equivalent linear system for quadratic eigenfunctions $\{f,g,h\}$,

$$f \equiv -i(\psi_1\phi_2 + \phi_1\psi_2)/2, \quad g \equiv \psi_1\phi_1, \quad h \equiv -\psi_2\phi_2, \quad \text{(IV.5)}$$

where both ψ and ϕ denote solutions of the Takhatajian–Faddeev linear system (II.3). We have

Lemma (Squared eigenfunction system): The vector function $\mathbf{F} \equiv (f,g,h)^T$ satisfies the linear system

$$\mathbf{F}_x = \begin{pmatrix} 0 & i(\sqrt{E} \mp e^{iu}/16\sqrt{E}) & i(\sqrt{E} \mp e^{-iu}/16\sqrt{E}) \\ 2i(\sqrt{E} \mp e^{-iu}/16\sqrt{E}) & -iw/2 & 0 \\ 2i(\sqrt{E} \mp e^{iu}/16\sqrt{E}) & 0 & iw/2 \end{pmatrix} F \quad \text{(IV.6)}$$

or, in component form,

$$f_x = i(\sqrt{E} \mp e^{iu}/16\sqrt{E})g + i(\sqrt{E} \mp e^{-iu}/16\sqrt{E})h,$$
$$g_x = -i(w/2)g + 2i(\sqrt{E} \mp e^{-iu}/16\sqrt{E})f, \quad \text{(IV.6')}$$
$$h_x = i(w/2)h + 2i(\sqrt{E} \mp e^{iu}/16\sqrt{E})f.$$

The verification of this lemma follows by direct calculation. We emphasize that system (IV.6) is compatible if and only if the potentials u and w satisfy the sine-Gordon system

$$u_x + u_t = w,$$
$$w_x - w_t = \sin u.$$

This squared eigenfunction system is fundamental to the theory of the sine-Gordon equation (see Paper III in this series). As a system of ordinary differential equations, it possesses a "first integral."

Lemma (Basic constant of motion): Let (f,g,h) solve the squared eigenfunction system (IV.6). Then

$$P(E) \equiv f^2(x,t;E) - g(x,t;E)h(x,t;E)$$

is independent of both x and t.

The proof that P is a constant of motion is immediate from the component form of the system, (IV.6'). Recall that we are working on an inverse problem, the construction of an N-phase wave from spectral data. The simple spectrum $\Sigma^{(s)} = \{E_k, k = 1,...,2N\}$ enters into the construction procedure through the basic constant of motion $P(E)$. Fix $2N$ complex constants $\{E_k\}$ consistent with the constraints (IV.2), and consider a polynomial form of the constant $P(E)$,

$$P(E) = f^2(x,t;E) - g(x,t;E)h(x,t;E) = \prod_{k=1}^{2N}(E - E_k). \quad \text{(IV.7)}$$

To achieve this polynomial $P(E)$ one seeks squared eigenfunction solutions (f,g,h) of (IV.6) which themselves are polynomial in the eigenvalue parameter $\sqrt{E}$,

$$f = \frac{1}{\sqrt{E}}\sum_{j=1}^{N} f_j E^j, \quad g = \sum_{j=0}^{N} g_j E^j, \quad h = \sum_{j=0}^{N} h_j E^j. \quad \text{(IV.8)}$$

We emphasize the existence of a solution (f,g,h) to the squared eigenfunction system forces the potentials (u,w) to satisfy the sine-Gordon equation. The additional requirement (beyond existence) that f,g,h are polynomial in $\sqrt{E}$ further constrains the sine-Gordon solution. It turns out that this polynomial constraint selects wave train solutions.

Lemma (Existence of squared eigenfunctions which are polynomial in $\sqrt{E}$): Polynomial solutions of linear system (IV.6) which have the form (IV.8) exist if and only if the potentials u and w satisfy

$$e^{-2iu} = -g_0/h_0, \quad \text{(IV.9a)}$$
$$w = 4f_N/g_N. \quad \text{(IV.9b)}$$

In this case, the coefficients satisfy

$$(f_j)_x = ig_{j-1} \mp \frac{i}{16}e^{iu}g_j + ih_{j-1} \mp \frac{i}{16}e^{-iu}h_j,$$
$$(g_j)_x = 2if_j \mp \frac{i}{8}e^{-iu}f_{j+1} - \frac{i}{2}wg_j, \quad \text{(IV.10a)}$$
$$(h_j)_x = 2if_j \mp \frac{i}{8}e^{iu}f_{j+1} + \frac{i}{2}wh_j,$$

$(j = 0,1,...,N)$ together with the constraints

$$f_0 \equiv f_{N+1} \equiv 0, \quad \text{(IV.10b)}$$
$$g_N + h_N \equiv 0.$$

To establish this existence lemma, we insert the polynomial ansatz (IV.8) into linear system (IV.6). System (IV.10)

for the coefficients $\{f_j,g_j,h_j\}$ results, from which the necessity of formulas (IV.9) follows by minor manipulations. As for the sufficiency, one replaces u and w in (IV.10) by formulas (IV.9). In this manner, (IV.10) becomes a closed system of nonlinear ordinary differential equations. The existence of a solution for this nonlinear system is elementary.

Several remarks are appropriate at this stage. First, Eqs. (IV.10) yield

$$g_N = -h_N = \text{const}, \tag{IV.11a}$$

$$g_0 h_0 = \text{const}. \tag{IV.11b}$$

Without loss of generality, we choose $g_N = +1$. Then the polynomial constant $P(E)$ takes the form

$$P(E) = f^2 - gh = (-g_N h_N)E^{2N} + \cdots + (-g_0 h_0)E^0 = \prod_{k=1}^{2N} (E - E_k),$$

from which we obtain

$$g_N = 1, \quad h_N = -1, \quad g_0 h_0 = -\prod_1^{2N} E_k. \tag{IV.12}$$

These formulas are useful to keep in mind during the construction process.

Constraint (IV.9) is crucial. It gives a formula for the potentials u and w in terms of the coefficients g_0,h_0,g_N,h_N. Next we show these potentials have *N degrees of freedom*, where N is the degree of the polynomials g and h.

Lemma (*u has N degress of freedom*).

If (f,g,h) admit polynomial representations (IV.8), then

$$u(x,t) = i \ln\left[\prod_{j=1}^{N} \mu_j(x,t) \Big/ \left(\prod_{j=1}^{2N} E_j\right)^{1/2}\right], \tag{IV.13}$$

where the N degrees of freedom $\{\mu_1(x,t),\ldots,\mu_N(x,t)\}$ are the zeros of the squared eigenfunction $g(x,t;E)$,

$$g(x,t;E) = \prod_{j=1}^{N} [E - \mu_j(x,t)], \tag{IV.14}$$

and the $\{E_j\}$ are the zeros of the constant $P(E) = f^2 - gh = \Pi(E - E_k)$.

To prove this lemma, one uses (IV.9a), (IV.12), and (IV.14):

$$\left.\begin{aligned} e^{-2iu} &= -g_0/h_0 \\ h_0 g_0 &= -\prod E_k \end{aligned}\right\} \Rightarrow e^{-iu} = \prod \mu_j \Big/ \left(\prod E_k\right)^{1/2}.$$

Next, we derive a system of differential equations for the variables $(\mu_1,\ldots,\mu_N)$.

Lemma (*Dynamical system for the* μ *variables*):

Let (f,g,h) admit polynomial representation (IV.8), with $g(x,t;E) = \Pi[E - \mu_j(x,t)]$. Then

$$\mu_{l,x} = \frac{2i\left(1 \mp \dfrac{1}{16P^{1/2}} \prod_{j\neq l} \mu_j\right)\left[\prod_{k=1}^{2N+1} (\mu_l - E_k)\right]^{1/2}}{\prod_{j\neq l} (\mu_l - \mu_j)}. \tag{IV.15}$$

To verify that the N zeros of $g,\{\mu_j\}$, satisfy dynamical system (IV.15), one begins with Eq. (IV.6b') for g,

$$g_x = \frac{-iw}{2} g + 2i\left(\sqrt{E} \mp \frac{e^{-iu}}{16\sqrt{E}}\right) f.$$

Inserting the product representation $g = \Pi(E - \mu_j)$ into this equation yields

$$\left[\prod (E - \mu_j)\right]_x \Big|_{E=\mu_l} = 2i\left[(\mu_l)^{1/2} \mp e^{-iu}/16(\mu_l)^{1/2}\right] f(x,t;\mu_l).$$

The potential e^{-iu} is known in terms of $\{\mu_j\}$ by (IV.13). To find $f(x,t;\mu_l)$, one uses

$$P(\mu_l) = \prod_{k=1}^{2N} (\mu_l - E_k) = f^2 - gh\,|_{E=\mu_l} = f^2(x,t;\mu_l).$$

These expressions yield (IV.15).

This completes the proof of the "μ-representation of u." We emphasize that the wave train u is constructed from the following data: (i) the simple spectrum $\{E_k\}$, (ii) N branch determinations $\{R(\mu_l) = [\Pi(\mu_l - E_k)]^{1/2}, l = 1,\ldots,N\}$, (iii) the branch determination of $P^{1/2} = (\Pi E_k)^{1/2}$, and (iv) the N dynamical variables $\{\mu_j\}$ which satisfy dynamical system (IV.15). Actually, this same data fixes the potentials u and w, as well as the polynomial squared eigenfunction (f,g,h). For example, the coefficients f_j in the polynomial squared eigenfunction f are easily shown to satisfy the system

$$\begin{pmatrix} 1 & \mu_1 & \mu_1^2 & \cdots & \mu_1^{N-1} \\ 1 & \mu_2 & \mu_2^2 & \cdots & \mu_2^{N-1} \\ \vdots & & & & \\ 1 & \mu_N & \mu_N^2 & \cdots & \mu_N^{N-1} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix} = \begin{pmatrix} f(\mu_1)/(\mu_1)^{1/2} \\ f(\mu_2)/(\mu_2)^{1/2} \\ \vdots \\ f(\mu_N)/(\mu_N)^{1/2} \end{pmatrix}, \tag{IV.16}$$

where

$$f(\mu_j) = \left[\prod_{k=1}^{2N} (\mu_j - E_k)\right]^{1/2} = [P(\mu_j)]^{1/2}.$$

By inverting this linear system, the data fixes the squared eigenfunction f.

The preceding verification of the μ-representation of the N-phase wave consists entirely of easy manipulations. There remains a very difficult question concerning the global nature of the coordinates $\{\mu_1,\ldots,\mu_N\}$. As it stands, the μ-representation generates complex potentials u and w; however, we are only interested in potentials which are either real or purely imaginary. Under such reality constraints, what is the topological nature of the manifold of N-phase waves? Is this manifold coordinatized globally by the $\{\mu_j\}$ variables?

For the case of purely imaginary potentials (sinh-Gordon), the answers are known. In this case, constraints (IV.2)

force the constants $\{E_k, k = 1,...,2N\}$ to be real and positive. We order them as $0 < E_1 < E_2 < \cdots < E_{2N} < \infty$. Then choose μ_j *real, positive,* and

$$\mu_j \in [E_{2j-1}, E_{2j}], \quad j = 1,2,...,N. \tag{IV.17}$$

Under the flow (IV.4), each μ_j travels on one sheet of the Riemann surface from E_{2j-1} to E_{2j}, changes sheets, then travels back to E_{2j-1} again. Topologically, this path is a circle. If we let $\mathcal{M}_N$ denote the manifold of purely imaginary N-phase waves with fixed simple spectrum,

$$\mathcal{M}_N \equiv \left\{ \begin{matrix} u:\mathbb{R}\to C \\ w:\mathbb{R}\to C \end{matrix} \middle| u^* = -u,\ w^* = -w;\ \Sigma^{(s)} \right\},$$

then $\mathcal{M}_N$ is an N-torus parametrized by $\{\mu_1,...,\mu_N\}$ which lie in the "gaps" (IV.17). This case is exactly analogous to that of Hill's equation. (See the preliminary section of Ref. 4 and recall that for purely imaginary potentials, the Takhatajian–Faddeev linear eigenvalue problem is self-adjoint.)

For real potentials (sine-Gordon), the situation is not well understood. Consider the manifold of real N-phase waves with fixed simple spectrum,

$$\mathcal{M}_N \equiv \left\{ u:\mathbb{R}\to\mathbb{R}, w:\mathbb{R}\to\mathbb{R} \,|\, \Sigma^{(s)} = \{E_1,...,E_{2N}\} \right\}.$$

Here, the constants E_j must satisfy the reality constraints of (IV.2); that is, either E_j are real and negative or they occur in conjugate pairs. Locally the $\{\mu_j\}$ (now complex) coordinatize the manifold $\mathcal{M}_N$; however, except in the single-phase case ($N = 1$), their global properties are unknown. For example μ_1, although complex valued, has only one real dimension, yet it does not live on a fixed complex curve which is independent of $(\mu_2,...,\mu_N)$. Moreover, the μ_j variables need not be distinct, but can collide in pairs. In summary, in the case of real potentials, it is not certain that the manifold $\mathcal{M}_N$ is globally a torus; nor is it clear that the manifold $\mathcal{M}_N$ can be globally coordinatized by $\{\mu_1,...,\mu_N\}$. In later sections, we investigate the μ coordinates for $N = 1$ and $N = 2$ in some detail.

3. Θ-function representation of the N-phase wave train

In the last section, we developed the "μ-representation of the N-phase wave train." This representation clearly shows that the wave train has N degrees of freedom $\{\mu_1,...,\mu_N\}$, but the μ variables satisfy a complicated dynamical system (IV.4). Fortunately, this flow can be transformed to straight-line motion by exploiting the fact that each variable μ_j resides on the Riemann surface of

$$R^2(E) = \prod_{j=1}^{2N+1} (E - E_j). \tag{IV.18}$$

In this section we use calculus on this Riemann surface to replace the variables $\{\mu_1,...,\mu_N\}$ by N "phases," each of which is *linear* in space (x) and time (t). The explicit formula for the wave train u in terms of these N phases involves theta functions of N variables. Since this "Θ-function representation of u" is not central to our discussion, we refer the reader to Ref. 20 for most details. To state the results, we first describe the appropriate cycles and differentials on the underlying Riemann surface which are used to both integrate the μ equations (IV.15) and define the Θ-functions of N variables. (The integration of the μ equations introduces the phase variables, and the Θ functions are used to express the μ variables, hence the potentials, in terms of these phases.)

On the Riemann surface of $R^2(E) = \Pi(E - E_k)$, $E_{2N+1} \equiv 0$, we introduce two families of closed curves (which form a basis for contour integration and are typically referred to as "canonical cycles"), $\{a_1,a_2,...,a_N\}$ and $\{b_1,b_2,...,b_N\}$. We illustrate these a–b cycles for the special case $N = 3$; for the cut structure of Fig. 6, the precise paths are depicted below in Fig. 7. (See Ref. 22 for such matters.)

Next, introduce N differentials (abelian differentials of the first kind)

$$dU_\nu = \frac{C_{\nu 1} E^{N-1} + \cdots + C_{\nu N}}{R(E)}\, dE, \quad \nu = 1,2,...,N, \tag{IV.19a}$$

where the matrix of constants $C = (C_{\nu\mu})$ is fixed in terms of $\{E_j\}$ by normalization conditions

$$\oint_{a_\mu} dU_\nu = \delta_{\nu\mu}. \tag{IV.19b}$$

From these differentials, define the "period matrix" $B = (B_{\mu\nu})$ by

$$B_{\mu\nu} = \oint_{b_\mu} dU_\nu. \tag{IV.19c}$$

First we use the differentials dU_ν to linearize the μ equations. Introduce the change of variables

$$\tau_\pm \equiv t \pm x, \tag{IV.20a}$$

in terms of which Eqs. (IV.15) simplify to

$$(\mu_l)_{\tau_j} = 2i \left[\frac{\prod_{j\neq l} \mu_j}{16\left(\prod E_k\right)^{1/2}} \right]^{(\frac{1}{2})} \left[\frac{\left(\prod_{k=1}^{2N+1} (\mu_l - E_k)\right)^{1/2}}{\prod_{j\neq l} (\mu_l - \mu_j)} \right]. \tag{IV.20b}$$

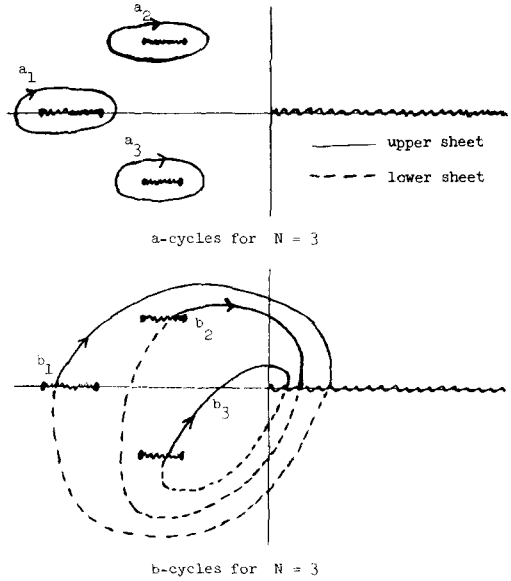


FIG. 7. Cycles for Genus $N = 3$.

Then, in terms of the differentials dU_j, (IV.19a), define abelian integrals of the first kind, $l_j \equiv l_j(\mu_1,...,\mu_N)$, by

$$l_j(\boldsymbol{\mu}) \equiv -\sum_{k=1}^{N} \int_{\mathring{\mu}_k}^{\mu_k} dU_j = -\sum_{l=1}^{N} C_{jl} \sum_{k=1}^{N} \int_{\mathring{\mu}_k}^{\mu_k} \frac{E^{N-l}}{R(E)}\, dE. \tag{IV.21a}$$

Next, compute derivatives of $l_j(\boldsymbol{\mu})$ with respect to $\tau_\pm$, and make use of (IV.20b):

$$\frac{\partial l_j(\boldsymbol{\mu})}{\partial \tau_\pm} = -\sum_{l=1}^{N} C_{jl} \sum_{k=1}^{N} \frac{\mu_k^{N-l}}{R(\mu_k)} \frac{\partial \mu_k}{\partial \tau_\pm}$$

$$= -\sum_{l=1}^{N} (2iC_{jl}) \sum_{k=1}^{N} \left(\frac{\prod_{m\neq k} \mu_m}{16P^{1/2}} \right)^{(9)_{11}} \frac{\mu_k^{N-l}}{\prod_{n\neq k} (\mu_k - \mu_n)}.$$

The right-hand side is independent of $\boldsymbol{\mu}$; thus, $l_j(\boldsymbol{\mu})$ flows linearly with $\tau_\pm$. This observation is immediate for $N = 1$. We illustrate the $N = 2$ case:

$$\frac{\partial l_j(\mu_1,\mu_2)}{\partial \tau_+} = -2iC_{j1} \left(\sum_{k=1}^{2} \frac{\mu_k}{\prod_{n\neq k} (\mu_k - \mu_n)} \right) - 2iC_{j2} \left(\sum_{k=1}^{2} \frac{1}{\prod_{n\neq k} (\mu_k - \mu_n)} \right)$$

$$= -2iC_{j1}(1) - 2iC_{j2}(0)$$

$$= -2iC_{j1}.$$

Similarly,

$$\frac{\partial l_j(\mu_1,\mu_2)}{\partial \tau_-} = \frac{2iC_{j2}}{16P^{1/2}}.$$

These two expressions are trivially integrated to yield

$$l_j(\mu_1,\mu_2) = -2i\,[\,C_{j1}\tau_+ - (C_{j2}/16P^{1/2})\tau_-\,] + \mathring{l}_j$$

or, in terms of x and t,

$$l_j = -2i\left[\left(C_{j1} + \frac{C_{j2}}{16P^{1/2}}\right)x + \left(C_{j1} - \frac{C_{j2}}{16P^{1/2}}\right)t\right] + \mathring{l}_j.$$

Thus, we have explicitly reduced the x,t dependence to two phases l_1 and l_2; the space and time flows have been linearized. The above computation carries through in the same manner for general N with the use of Lagrange interpolation. The result of that analysis is

$$l_j(\mu_1,...,\mu_N) = -2i\left[\left(C_{j1} + \frac{(-1)^N C_{jN}}{16P^{1/2}}\right)x + \left(C_{j1} + \frac{(-1)^{N+1} C_{jN}}{16P^{1/2}}\right)t\right] + \mathring{l}_j. \tag{IV.21b}$$

There are N phases in the potentials u,w. Mathematically, Eqs. (IV.21) for $j = 1,...,N$ represent the "Jacobi inversion problem": given the phases l_j, find the unknowns $\mu_1,...,\mu_N$. This classical problem is solved in terms of the zeros of the Riemann theta function. This explains the role of Θ functions in the exact, closed-form, periodic solutions of sine-Gordon and other nonlinear evolution equations integrated by the inverse spectral transform. (We suggest the reader consider the inversion problem for $N = 1$.) We turn next to display these exact N-phase solutions in terms of Θ functions.

From the ingredients in (IV.19), the N-dimensional Θ function is constructed:

$$\Theta(\mathbf{P};B) = \sum \exp\{\pi i(B\,\mathbf{k},\mathbf{k}) + 2\pi i(\mathbf{P},\mathbf{k})\},$$

$$\mathbf{P} = (P_1,P_2,...,P_N).$$

(We remark that as defined, ImB is positive definite, which yields very rapid convergence for these series—an advantage of the representation to follow.) The solution of the Jacobi inversion problem in terms of $\Theta(\mathbf{P};B)$ then yields

$$u(x,t) = 2i \ln\left[\frac{\Theta(\mathbf{l}(x,t) + \frac{1}{2};B)}{\Theta(\mathbf{l}(x,t);B)}\right], \tag{IV.22}$$

where

$$\mathbf{l}(x,t) = (l_1,...,l_N),$$

$$l_j(x,t) = -2i\left\{\left[\left(C_{j1} + \frac{1}{16P^{1/2}} C_{jN}\right)x + \left(C_{j1} - \frac{1}{16P^{1/2}} C_{jN}\right)t\right]\right\} + l_j(0,0),$$

$$E_{2N+1} \equiv 0, \quad P = \prod_{k=1}^{2N} E_k,$$

$$\mathbf{l} + \tfrac{1}{2} \equiv (l_1 + \tfrac{1}{2}, l_2 + \tfrac{1}{2},..., l_N + \tfrac{1}{2}).$$

We refer to Eq. (IV.22) as the "Θ-function representation of $u(x,t)$"; notice that it explicitly displays the sine-Gordon solution $u(x,t)$ as a *multiphase wave train*. That is, it has N phases, $\{l_\nu(x,t), \nu = 1,...,N\}$, *each of which depends linearly on x and t*. The wave form is parametrized by the fixed simple spectrum $\Sigma^{(S)} = \{E_1,...,E_{2N}\}$ and the N constants $l_1(0,0),...,l_N(0,0)$.

Any choice of these parameters will yield potentials u and w which (i) solve the complex sine-Gordon equation, (ii) contain N degrees of freedom, and (iii) possess a Takhatajian–Faddeev spectrum which lies in $N + 1$ bands and terminates at $\{E_1,E_2,...,E_{2N},0,\infty\}$. If, in addition, the points $\{E_j\}$ satisfy the constraints specified in Eqs. (IV.2), and the real part of the constant $\mathbf{l}(0,0)$ is chosen properly, then the potentials will either be real or pure imaginary. In general, these potentials will be quasiperiodic in space and time with N spatial periods and N temporal periods. To obtain periodic potentials, the $2N$ points $\{E_j\}$ must satisfy $(N - 1)$ constraints, which ensure the N spatial periods are commensurate. An additional constraint is needed to ensure fixed period L. The inverse problem as posed (given the simple spectrum $\{E_j\}$, construct the potentials) naturally yields quasiperiodic potentials. In some applications, such as modulating wave trains, this quasiperiodic class of waves is most natural; in other instances, such as finite-length Josephson transmission lines, the periodic constraints must be imposed.

These representations of the wave $u(x,t)$ will be of little use until the connection between the input parameters $\{E_1,...,E_{2N}\}$ and the physical characteristics (such as wavenumbers, frequencies, amplitudes, and periods) is clearly understood. In the next section, we discuss this connection for sine-Gordon waves in great detail for the $N = 1$ case; in the final section, we consider a special class of $N = 2$-phase sine-Gordon waves.

V. TRAVELING WAVES (SINGLE-PHASE WAVES)

In this section we consider single-phase traveling waves which may be constructed by elementary means. The goal of this section is to relate these well known families of traveling waves to the less familiar single-phase $(N=1)\mu$ and Θ representations. We restrict ourselves to real (sine-Gordon) waves. The sinh-Gordon case follows similarly.

1. Reduction to $N=1$

The single-phase wave takes the form

$$u(x,t) = 2i \ln\left[\frac{\Theta(l(x,t)+\frac{1}{2}B)}{\Theta(l(x,t);B)}\right], \tag{V.1a}$$

where the single phase $l(x,t)$ is given, with $P = E_1E_2$, by

$$l(x,t) = -2iC\left[(1+1/16P^{1/2})x + (1-1/16P^{1/2})t\right] + l(0,0). \tag{V.1b}$$

From this representation we see that the single-phase sine-Gordon solution is a traveling wave; that is, a function of x and t through only the one linear combination $(\kappa x+\omega t)$, with "phase velocity" U given by

$$U \equiv \frac{\omega}{\kappa} = \frac{16P^{1/2}-1}{16P^{1/2}+1}. \tag{V.1c}$$

2. Direct ansatz method to display traveling waves

Once we realize the $N=1$ solution $u(x,t)$ is a traveling wave, there is a direct and elementary approach which catalogues all single-phase traveling wave solutions of the sine-Gordon equation. One merely seeks a solution in the form of a traveling wave:

$$u(x,t) = u_T(\kappa x+\omega t).$$

Inserting this ansatz into the sine-Gordon equation yields an equation for u_T as a function of the "phase" $\theta = \kappa x+\omega t$,

$$(\omega^2-\kappa^2)u_T'' = -\sin u_T,$$

where $'$ denotes $d/d\theta$. This "effective oscillator" equation may be integrated once to obtain the energy equation

$$\tfrac{1}{2}(\kappa^2-\omega^2)(u_T')^2 + V_{\text{eff}} = E, \quad \text{if} \quad U^2=\omega^2/\kappa^2<1, \tag{V.2a}$$

$$\tfrac{1}{2}(\omega^2-\kappa^2)(u_T')^2 + v_{\text{eff}} = \epsilon = -E, \quad \text{if} \quad U^2=\omega^2/\kappa^2>1, \tag{V.2b}$$

where V_{eff} and v_{eff} are the effective potentials given by

$$V_{\text{eff}}(u) = \cos u,$$

$$v_{\text{eff}}(u) = -\cos u.$$

The potential energy diagrams are sketched in Fig. 8.

As we catalog all traveling waves, we first classify waves as having phase speed $|U| = |\omega/\kappa|$, which (i) exceeds, or (ii) is

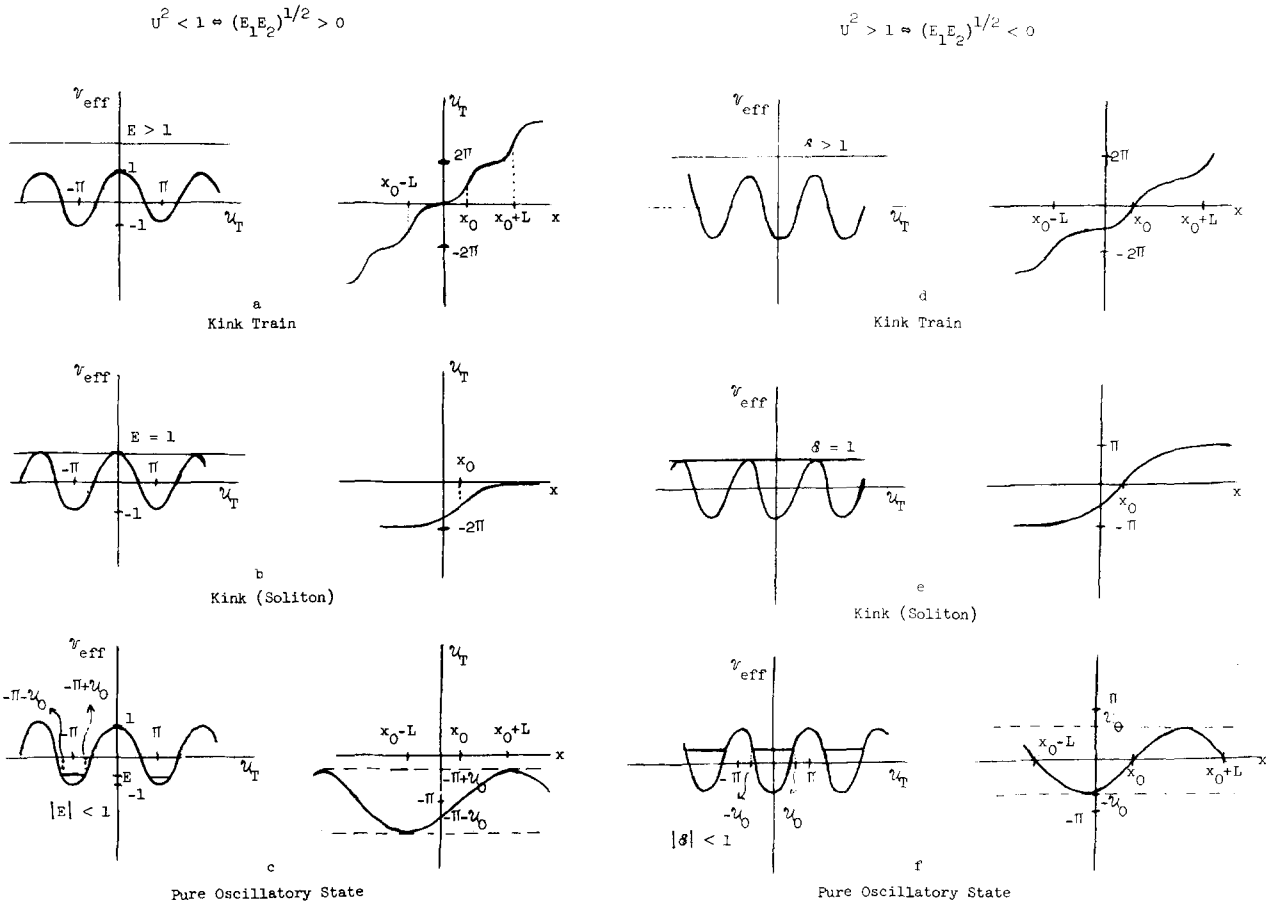


FIG. 8. Potential energy diagrams and corresponding solutions $(N=1)$.

exceeded by, the characteristic speed $c = 1$. For phase velocities satisfying $U^2 < 1$, there are three classes of solutions which are depicted in Figs. 8(a), 8(b), and 8(c). Notice when the energy parameter $E = 1$, the traveling wave is a single kink which rises from -2π to 0 [this corresponds to picking the positive square root when solving for u_T from (V.2a); the negative determination yields a single antikink which falls steadily from 0 to -2π]. For $E > 1, u_T$ is a monotonic sequence of kinks (antikinks), and should be thought of as a "kink train." (We amplify this interpretation in Sec. V.4.) This kink train is periodic in θ (mod2π), with period $\mathscr{P}$ as computed directly from the energy equation (V.2a):

$$\left(\frac{\kappa^2-\omega^2}{2}\right)^{1/2}\int_0^{2\pi}\frac{du}{(E-\cos u)^{1/2}} = \mathscr{P}(\kappa,\omega,E), \qquad E>1. \quad \text{(V.3)}$$

When $-1 < E < 1$, the traveling wave becomes oscillatory about $u_T = -\pi$. The solutions are strictly periodic for this range of E with θ-period 2,

$$2\left(\frac{\kappa^2-\omega^2}{2}\right)^{1/2}\int_{u_-}^{u_+}\frac{du}{(E-\cos u)^{1/2}} = 2\mathscr{P}(\kappa,\omega,E), \qquad -1<E<1. \quad \text{(V.4)}$$

In this formula, the "turning points," $u_\mp = -\pi \mp u_0(E)$, denote consecutive zeros of $E - V_{\text{eff}}(u) = E - \cos u$, as depicted in Fig. 8(c); these turning points fix the amplitude of oscillation by

$$\text{Amplitude of Oscillation} = u_0. \quad \text{(V.5)}$$

[Notice that the spatial period L and temporal period T can also be computed from the energy equation (V.2a) simply by changing variables from θ to x and t, respectively. From these, the wavenumber $\kappa = 2\pi/L$ and frequency $\omega = 2\pi/T$ immediately follow.]

When $U^2 > 1$, the Figs. 8(d), 8(e), and 8(f) apply with similar conclusions.

The main point of this subsection is that, using this "effective oscillator" approach, one can quickly catalog all traveling waves as belonging to one of *six distinct classes*. Once the phase velocity U is fixed by either $|U| < 1$ or $|U| > 1$, the traveling wave is either a train of kinks ($E > 1$), a single kink ($E = 1$), or an oscillatory state ($|E| < 1$). Moreover, the input parameters ($U = \omega/\kappa, E$) admit clear physical interpretation; they characterize the wavenumber, frequency, and amplitude of the wave; the phase velocity is given by ω/κ and the θ-period, x-period, and t-period are all defined by loop integrals.

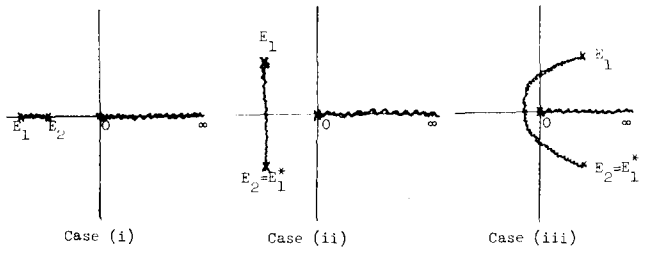


FIG. 9. Various locations of E_1, E_2 with appropriate branch cuts ($N = 1$).

The μ-representation of $u(x,t)$ for single-phase waves has input parameters E_1, E_2. The location of E_1, E_2 in the complex plane (subject to the constraints $E_j \in \Sigma^{(S)} \Rightarrow E_j^* \in \Sigma^{(S)}$ and $E_j \notin (0,\infty)$) must yield an equivalent catalog for the traveling-wave solutions. We discuss these topics next.

3. μ-representation for $N = 1$; catalog and physical characteristics of the traveling waves in terms of $\Sigma^{(S)}$

For the $N = 1$ case, the simple spectrum $\Sigma^{(S)} = \{E_1, E_2\}$ can be located in the complex E plane in only two distinct ways:

Case 1. $E_1 < E_2 < 0$, so both lie on the negative real E axis.

Case 2. $E_2 = E_1^*, E_1 \neq E_2$, so they occur as a conjugate pair.

These possible locations E_1, E_2 lead to the three configurations depicted in Fig. 9, which also displays the canonical branch cuts. We now use the analogy with whole-line sine-Gordon scattering theory (refer to Theorem III.2 and Figs. 2 and 5) to guess the excitation classification based on the location of E_1 and E_2.

Case 1, $E_1 < E_2 < 0$, appears to arise from an isolated pole $E_j < 0$, which spreads along the negative real axis into a band of spectrum; this should be the periodic analog of one kink, namely, a train of kinks.

Case 2, $E_2 = E_1^*, E_1 \neq E_2$, can be viewed (with E_1, E_1^* located near the positive real axis) as related to radiation degrees of freedom. It then seems that E_1, E_1^* should give a pure oscillatory radiation excitation.

We will verify these guesses next using the single-phase μ-representation of $u(x,t)$, which we now recall:

$$u(x,t) = i\ln[-\mu(x,t)/(E_1E_2)^{1/2}], \quad \text{(V.6a)}$$

where $\mu(x,t)$ satisfies

$$(\mu)_x = 2i(1 \pm 1/16(E_1E_2)^{1/2})[\mu(\mu-E_1)(\mu-E_2)]^{1/2} \quad \text{(V.6b)}$$

We aim to deduce many facts directly from (V.6a) and (V.6b), without need of the exact integration of these μ equations in terms of Θ functions. First, the relation between $u(x,t)$ and $\mu(x,t)$, (V.6a), together with the μ equations, quickly yields

$$\frac{u_t}{u_x} = \frac{\mu_t}{\mu_x} = \frac{16(E_1E_2)^{1/2}-1}{16(E_1E_2)^{1/2}+1}. \quad \text{(V.7)}$$

Thus, the μ-representation confirms the $N = 1$ sine-Gordon solution is a traveling wave; that is,

$$u(x,t) = u_T(\kappa x + \omega t),$$

with the phase velocity $U = \omega/\kappa$, from (V.7), given by

$$U = \frac{u_t}{u_x} = \frac{16(E_1E_2)^{1/2}-1}{16(E_1E_2)^{1/2}+1}. \quad \text{(V.8)}$$

We therefore find that the phase velocity is a function of E_1 and E_2 only through the product $(E_1E_2)^{1/2}$. We sketch U as a function of $(E_1E_2)^{1/2}$ in Fig. 10; note the following facts ($c\equiv 1$ is the characteristic speed):

(i) $(E_1E_2)^{1/2}\gtrless 0 \Leftrightarrow |U|\lesseqgtr 1$; the sign determination of $(E_1E_2)^{1/2}$ fixes the phase speed relative to the characteristic speed.

(ii) $|(E_1E_2)^{1/2}|\lesseqgtr \frac{1}{16} \Leftrightarrow U \lesseqgtr 0$; the wave is at rest when $|(E_1E_2)^{1/2}|$ lies on the circle of radius $\frac{1}{16}$ and travels to the right (left) when $|(E_1E_2)^{1/2}|$ lies inside (outside) that circle.

Moreover, now that the μ-representation has implied traveling waves, we can make contact between the μ-representation and the "effective oscillator" approach. The first step is provided by formula (V.8) for U in terms of E_1,E_2; we now need to relate the energy parameter E of the effective oscillator to E_1 and E_2. To do so, recall the energy equation (V.2a), which can be written in terms of $x(\partial/\partial x = \kappa\, d/d\theta)$ as

$$\tfrac{1}{2}(U^2-1)(u_T)_x^2 - \cos u_T = -E. \tag{V.9a}$$

This equation (V.9a) can also be obtained from the μ-representation. First,

$$u_x^2 = \frac{-\mu_x^2}{\mu^2} = \frac{4[1+16(E_1E_2)^{1/2}]^2}{[16(E_1E_2)^{1/2}]^2}\,\frac{(\mu-E_1)(\mu-E_2)}{\mu}$$

which, upon using (V.8), may be reduced to

$$\tfrac{1}{2}(U^2-1)u_x^2 = \cos u + \tfrac{1}{2}E_1/(E_1E_2)^{1/2} + E_2/(E_1E_2)^{1/2}. \tag{V.9b}$$

Fixing $U^2<1$, and comparing (V.9a) and (V.9b), we find the energy E is given in terms of E_1 and E_2:

Case 1: $E_1<E_2<0$

$$E = \tfrac{1}{2}((E_1/E_2)^{1/2} + (E_2/E_1)^{1/2}) > 1. \tag{V.10a}$$

Case 2: $E_1 = E_2^*, E_1 \neq E_2$

$$E = -\cos[\mathrm{ph}(E_1)] \in (-1,1). \tag{V.10b}$$

For the choice $U^2>1$, the same formulas hold with the total energy E replaced with $-E$.

We now see directly from the μ-representation that the location of E_1,E_2 catalogs the traveling-wave solutions in an equivalent way to that of the effective oscillator parameters (U,E). In summary,

Case 1. $E_1<E_2<0 \Rightarrow$ kink trains, Figs. 8(a) and 8(d).

Case 2. $E_1 = E_2^*, E_1 \neq E_2 \Rightarrow$ oscillatory solutions, Figs. 8(c) and 8(f). Moreover, in the limiting case $E_1 = E_2 < 0$ (the transition state between Case 1 and Case 2), the total energy $E\equiv 1$ and the resulting wave form is a single-kink—the sine-Gordon soliton. Thus,

Case 3. $E_1 = E_2 < 0 \Rightarrow$ "soliton limit"—a single kink, Figs. 8(b) and 8(e).

We now turn to the physical interpretations of the parameters E_1 and E_2. This information is contained in formulas (V.8) for the phase velocity U and (V.10) for the energy E. We elaborate by showing how the amplitude of the oscillatory states is fixed by the phase of E_1, and is independent of the amplitude $|E_1|$.

Fix $U^2<1$ (the other case follows in the same manner), and consider the μ-representation for this cut structure. From

$$u(x,t) = i\ln(-\mu(x,t)/(E_1E_2)^{1/2}),$$

reality of the solution implies $|\mu(x,t)| = (E_1E_2)^{1/2} = |E_1|$: *$\mu(x,t)$ resides on the circle of radius $|E_1|$* (see Fig. 11) with $\mathrm{ph}(\mu) = -(\pi + u(x,t))$,

$$\mu(x,t) = |E_1| e^{-(\pi + u(x,t))}.$$

Since $u(x,t)$ oscillates between $|u(x,t)+\pi| \leqslant u_0$, we find the path of $\mu(x,t)$ as shown in Fig. 11; thus,

$$u_0 \equiv \text{amplitude of oscillation} \equiv \mathrm{ph}E_1. \tag{V.11}$$

At this stage in the presentation, we have used the $N=1$ μ-representation to show the sine-Gordon solutions are traveling waves; the input parameters (E_1,E_2) catalog all the traveling waves and describe their physical characteristics (phase velocity and amplitude). Next, we use the $N=1$ Θ-function representation to arrive at an exact decomposition formula for each class of solutions (kink trains and oscillatory states). This decomposition formula firmly establishes that kinks and antikinks comprise the fundamental building blocks for each $N=1$ solution.

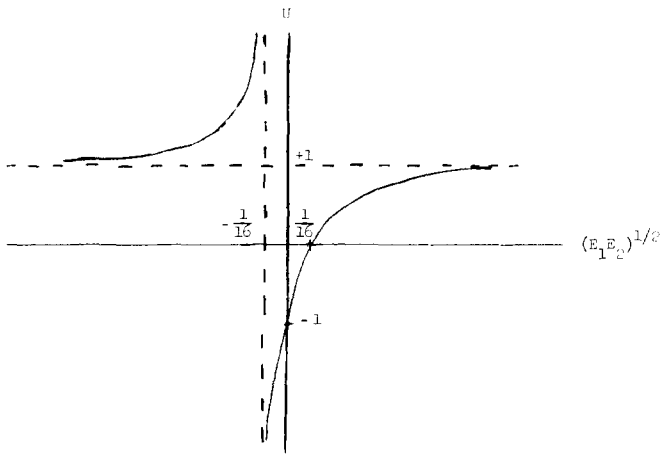


FIG. 10. Graph of U vs $(E_1E_2)^{1/2}$.

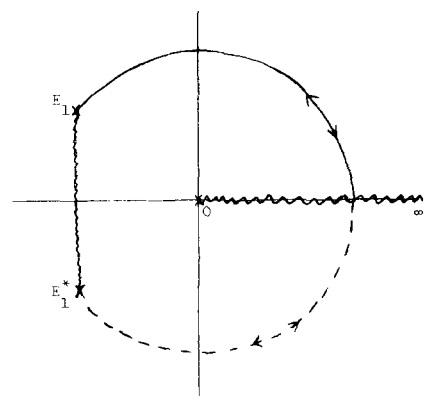


FIG. 11. μ-cycle; $N=1, E_1 = E_2^*, E_1 \neq E_2$.

the traveling waves and describe their physical characteristics (phase velocity and amplitude). Next, we use the $N=1$ Θ-function representation to arrive at an exact decomposition formula for each class of solutions (kink trains and oscillatory states). This decomposition formula firmly establishes that kinks and antikinks comprise the fundamental building blocks for each $N=1$ solution.

4. Infinite-product representation of Θ functions and summation representation of the traveling waves

Toda[23] used an infinite-product representation of the theta function to interpret the periodic traveling-wave solutions of the Korteweg–deVries equation and the Toda lattice as a sum of soliton shapes successively shifted by one period. Generically, if we let $u_T(\xi)$ denote the periodic traveling wave with period L, and $u_S(\xi)$ denote a soliton shape centered at $\xi=0$, Toda's formula is of the form

$$u_T(\xi)=\sum_{n=-\infty}^{\infty} u_S(\xi+nL).$$

This formula allows the beautiful interpretation of the traveling wave as a soliton on a ring. It seems one of the remarkable properties of these completely integrable nonlinear equations that such a formula is exact. In this section, we adapt Toda's arguments to the sine-Gordon equation; we show the "helical wave" (Fig. 14) literally is a kink train, while the oscillatory state admits a very interesting interpretation.

We begin by recalling the well-known single-soliton solutions of the whole-line sine-Gordon equation. They are classified as "kinks" or "antikinks," and are depicted in Fig. 12. The familiar formulas for these solitons, centered at x_0, are

$$u_{\substack{K\\AK}}(x-x_0,t)=4\arctan(e^{\pm\phi}), \tag{V.12}$$

where

$$\phi=(x-x_0-vt)/(1-v^2)^{1/2}; \tag{V.13}$$

a little manipulation places these in the equivalent form

$$u_{\substack{K\\AK}}(x-x_0,t)=2i\ln\left(\frac{1\mp ie^{\phi}}{1\pm ie^{\phi}}\right), \tag{V.14}$$

where the choice of branch for ln is taken to correspond with Fig. 12.

Now, the Θ-function representation for single-phase traveling waves is [Eq. (V.1)]

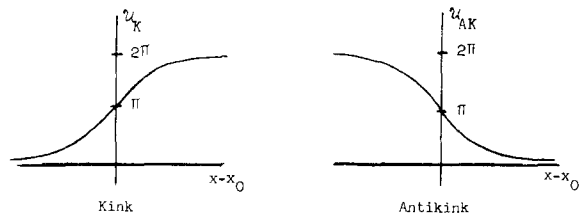


FIG. 12. Kink, antikink shapes.

$$u(x,t)=2i\ln\left(\frac{\Theta(l(x,t)+\tfrac{1}{2};B)}{\Theta(l(x,t);B)}\right), \tag{V.15a}$$

where the phase $l(x,t)$ takes the explicit form[24]

$$l(x,t)=-2iC\left[\left(1+\frac{1}{16(E_1E_2)^{1/2}}\right)(x-x_0)+\left(1-\frac{1}{16(E_1E_2)^{1/2}}\right)t\right]-\tfrac{1}{2}B\mp\tfrac{1}{4}. \tag{V.15b}$$

With this representation, the wavenumber κ and frequency ω are given by

$$\kappa=\frac{-4\pi C}{F}\left(1+\frac{1}{16(E_1E_2)^{1/2}}\right), \tag{V.16}$$

$$\omega=\frac{-4\pi C}{F}\left(1-\frac{1}{16(E_1E_2)^{1/2}}\right),$$

where the factor $F=\operatorname{Im}B$ for the kink train and $F=\operatorname{Im}(2B)$ for the oscillatory state. The normalization constant $(C_{11})\equiv C$ and period "matrix" $(B_{11})\equiv B$ [see Eq. (IV.19)] are functions only of the eigenvalues E_1,E_2; explicit formulas for C and B in terms of elliptic integrals are given in an appendix, but for now we note only their general form:

$$\begin{aligned}&\text{For } E_1<E_2<0,\quad B=i\operatorname{Im}(B),\quad C<0.\\&\text{For } E_1=E_2^{*},E_1\neq E_2,\quad B=-\tfrac{1}{2}+i\operatorname{Im}(B),\quad C<0.\end{aligned} \tag{V.17}$$

Although the one-dimensional Θ function is defined by the infinite series

$$\Theta(l;B)=\sum_{n=-\infty}^{\infty}\exp(i\pi Bn^2+i2\pi ln),$$

it also admits an infinite-product representation.[25] In Appendix B, we use infinite products to show the Θ-function representation (V.15) of the single-phase sine-Gordon wave has the series representation[26]

$$u(x,t)=\sum_{n=1}^{\infty}2i\ln\left(\frac{1-ie^{\alpha_n}}{1+ie^{\alpha_n}}\right)+\sum_{n=0}^{-\infty}2i\ln\left((-1)\frac{1-ie^{\alpha_n}}{1+ie^{\alpha_n}}\right), \tag{V.18a}$$

where

$$\alpha_n=\kappa(x-x_0)+\omega t+2n\pi iB. \tag{V.18b}$$

Now, by specializing to the two possible locations of E_1,E_2, using formulas (V.14) for the kink, antikink wave forms, and noting the facts (V.17), we find

Theorem V.1. (Decomposition formulas for single-phase sine-Gordon waves)
Fix $|U|<1$ and the choice $-\frac{1}{4}$ in $l(x,t)$. Then
(i) for $E_1<E_2<0$,

$$u(x-x_0,t)=\sum_{n=-\infty}^{\infty}\{u_K(x-x_0-nL,t)+\pi(\operatorname{sgn}(n)-1)\} \tag{V.19}$$

$$=\sum_{n=-\infty}^{\infty}\left\{2i\ln\left(\frac{1-ie^{\theta_n}}{1+ie^{\theta_n}}\right)+\pi(\operatorname{sgn}(n)-1)\right\} \tag{V.19'}$$

(ii) for $E_1 = E_2^*, E_1 \neq E_2$,

$$u(x - x_0, t) = \sum_{n=-\infty}^{\infty} \{u_K(x - x_0 - 2nL, t) + u_{AK}(x - x_0 - (2n-1)L, t) + 2\pi(\mathrm{sgn}(n) - 1)\} \tag{V.20}$$

$$= \sum_{n=-\infty}^{\infty} \left\{2i \ln\left(\frac{1 - ie^{\theta_{2n}}}{1 + ie^{\theta_{2n}}}\right) + 2i \ln\left(\frac{1 + ie^{\theta_{2n}'}}{1 - ie^{\theta_{2n}'}}\right) + 2\pi(\mathrm{sgn}(n) - 1)\right\}, \tag{V.20'}$$

where

$$\theta_n \equiv \kappa(x - x_0 - nL) + \omega t,$$

$\kappa, \omega, L = 2\pi/\kappa$ are given by Eq. (V.16), and

$$\mathrm{sgn}(n) = \begin{cases} +1 & \text{for } n > 0, \\ -1 & \text{for } n \leqslant 0. \end{cases}$$

We display these decomposition formulas in Figs. 13 and 14.

We now interpret these results: For $E_1 < E_2 < 0$ (the "helical wave" solution), we find the wave is literally a kink train: a sum of translated kink shapes successfully shifted by one period L. [We note the choice of $(+\frac{1}{4})$ in $l(x,t)$ yields an antikink train.] Moreover, the wave train is explicitly periodic (mod2π), with spatial period

$$L = \frac{\mathrm{Im}(B)}{2|C|(1 + 1/16(E_1E_2)^{1/2})}. \tag{V.21}$$

As E_1 and E_2 coalesce, the period L becomes infinite (use the explicit formulas for B,C in Appendix D), the kinks in the sum move infinitely far apart, and only the $n = 0$ term (a single kink centered at x_0) remains; in this manner, the wave train reduces to a single kink as E_1, E_2 coalesce along the negative real axis:

$$\lim_{E_1 \to E_2} L = +\infty,$$

$$\lim_{E_1 \to E_2} \kappa(x - x_0) + \omega t = \phi \equiv \frac{x - x_0 - vt}{(1 - v^2)^{1/2}},$$

$$v \equiv \frac{1 - 16|E_1|}{1 + 16|E_1|},$$

$$\lim_{E_1 \to E_2} u(x,t) = 4 \arctan\left[\exp\left(\frac{x - x_0 - vt}{(1 - v^2)^{1/2}}\right)\right]. \tag{V.22}$$

In this soliton limit, the analyticity structure is depicted in Fig. 15. Thus, the collapse of the kink train to a single kink as E_1, E_2 coalesce is consistent with our intuition from scattering theory. Before this limit is taken, the series representation (V.19) shows, for $E_1 < E_2 < 0$, that the sine-Gordon solution is a train of distorted kinks, or kink "shapes;" the speed and width of each kink in the train is not that of the soliton except in the lone survivor of the infinite-period limit. This distortion in the speed and width of each kink from that of a kink in isolation is the only effect of the interaction of the "tails" of the individual soliton components as they form a kink train.

For $E_1 = E_2^*, E_1 \neq E_2$, the oscillatory wave solution, the series decomposition (V.20) is even more interesting. It shows that the fundamental building block of the oscillatory state is a kink–antikink pair, bound together to form the "bumps" in the wave. The full wave is then shown by this formula to be a train of kink–antikink pairs, successively shifted over each period $2L$. The building blocks and full oscillatory state are depicted in Fig. 14. Moreover, this representation shows the solution is truly periodic, with spatial period $2L$ given by (V.21). Once again, as E_1, E_1^* collide on the negative real axis, $E_1 = E_1^* < 0$, the branch cut collapses to a single pole on the negative real axis (Fig. 15); the period L becomes infinite, and only the single kink centered at $x = x_0$ survives. Before this limit is taken, the individual kink, antikink components are distorted kinks and antikinks. The interaction of the tails alters the speeds and widths from that of the solitons in isolation to that of the wave train; the soliton speed-width relationship is recovered only for the single soliton that survives the infinite-period limit.

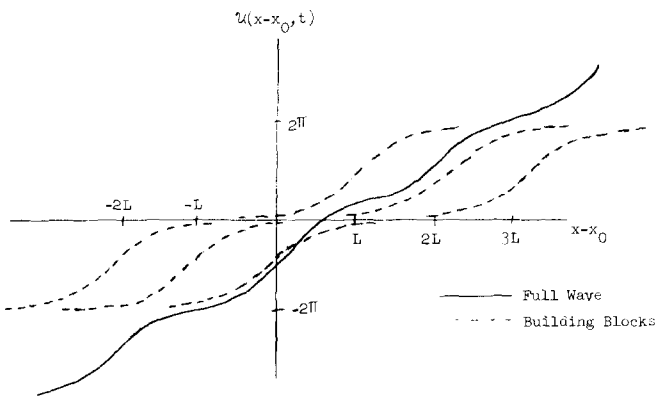


FIG. 13. Graphic representation of the decomposition formula for the kink train, $E_1 < E_2 < 0$.

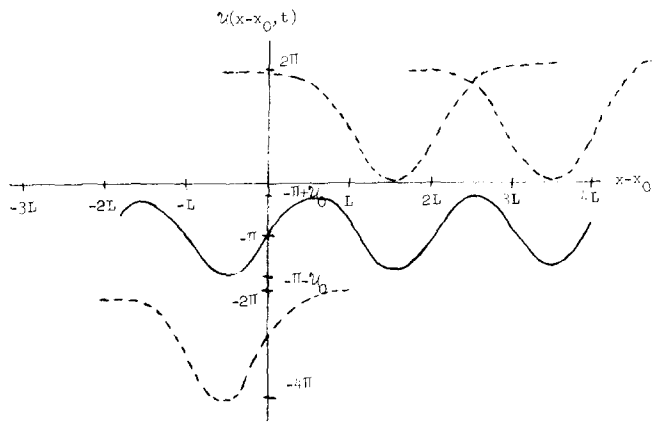


FIG. 14. Graphic representation of the decomposition formula for the oscillatory state, $E_1 = E_2^*, E_1 \neq E_2$.

Thus, these decomposition formulas provide insight into the nature of the single-phase periodic wave trains. We emphasize that these representations are exact, and suggest a fundamental role for the whole-line soliton components in the periodic theory. Appendix D completes this section with detailed formulas for the constants C and B in the Θ-function representation, in terms of E_1 and E_2. The wavenumber κ, frequency ω, and other physical characteristics are then expressed in terms of $C(E_1,E_2), B(E_1,E_2)$.

VI. THE DEGENERATE CASE OF PURE SOLITONS: $N = 1,2$

In this section, we analyze the μ-representation of one and two phase sine-Gordon waves *in the degenerate case of pure solitons*. From this analysis, we gain: (i) very explicit information about the μ paths for the specialized soliton configurations, which in turn suggests the nature of the μ paths for the periodic configurations, (ii) an understanding of the "Jacobi inversion problem" on the sphere (genus zero Riemann surface), an instructive exercise since in this case the inversion problem is solved explicitly in terms of elementary functions (rather than Θ functions), (iii) the recovery of the well-known $N = 1$ (kink) and $N = 2$ (kink–kink pair, breather) soliton formulas, with the physical characteristics explicitly labeled by the spectral parameters E_j. Incidentally, this analysis also illustrates how whole-line soliton constructions are imbedded in the periodic framework at the level of the μ equations. More importantly, it displays concretely the type and origin of the difficulty with the μ variables.

For $N = 1$, the degeneracy takes the form $E_1 = E_2 \equiv -K$, and we denote $\mu \equiv \mu_1$, $P^{1/2} = (E_1E_2)^{1/2} \equiv K$, and $R(\mu) = [\mu(\mu - E_1)(\mu - E_2)]^{1/2} \equiv (\mu)^{1/2}(\mu + K)$. Thus, the μ-representation for this degenerate $N = 1$ case becomes

$$u(x,t) = i \ln(-\mu/K), \tag{VI.1}$$

$$(\mu)_x = 2i(1 \pm 1/16K)(\mu + K)\sqrt{\mu}. \tag{VI.2}$$
t

This pair (VI.2) is easily integrated to yield

$$\ln A(x,t) = \theta + \ln A^0, \tag{VI.3a}$$

where

$$\theta(x,t) \equiv 2[(\sqrt{K} + 1/16\sqrt{K})(x - x_0) + (\sqrt{K} - 1/16\sqrt{K})(t - t_0)], \tag{VI.3b}$$

$$A(x,t) \equiv \frac{[1 - i(\mu(x,t)/K)^{1/2}]}{[1 + i(\mu(x,t)/K)^{1/2}]}, \tag{VI.3c}$$

and $A^0 \equiv A(x_0,t_0)$.

In general, $\mu(x,t)$ as calculated from (VI.3) leads to a complex potential $u(x,t)$ through (VI.1). One must choose the initial condition $\mu(x_0,t_0)$, or equivalently the integration constant A^0, to ensure reality of u. The correct choice, as seen from (VI.1), is that $\mu(x_0,t_0)$ [and therefore $\mu(x,t)$] is constrained to a circle of radius K,

$$|\mu(x,t)| = K = -E_1. \tag{VI.4}$$

This reality constraint yields

$$\mu(x,t) = \left(\frac{1 - ie^{\theta + \ln|A^0|}}{1 + ie^{\theta + \ln|A^0|}}\right)^2,$$

which leads to the familiar $N = 1$ soliton formula:

$$u(x,t) = 4 \arctan(e^{\phi}), \tag{VI.5a}$$

$$\phi \equiv \frac{(x - x_1) - v(t - t_0)}{(1 - v^2)^{1/2}}, \quad v = \frac{8\sqrt{K}}{1 + 16\sqrt{K}}. \tag{VI.5b}$$

Thus, the solution is the single soliton—a kink. More importantly, though, $\mu(x,t)$ is seen to live on the circle of radius K in the cut E plane, and takes an infinite amount of "time" to travel from the pole at $-K$ around the circle back to $-K$. If we spread the pole at $-K$ into either periodic configuration, this soliton μ path indicates the *periodic* μ paths shown in Fig. 16. The point is that the degenerate case of a single soliton suggests the correct μ path for the $N = 1$ periodic configurations, (as computed by the techniques of Sec. V).

We now turn to the degenerate case of two solitons, with bound state eigenvalues at

$$E_1 = E_2 = -K_1, \quad E_3 = E_4 = -K_2; \tag{VI.7}$$

there are two possible cases:

(i) kink–kink pair $K_1 > K_2 > 0$,

(ii) breather $K_1 = K_2^*$.

For now we label the spectrum with K_1, K_2 via Eq. (VI.7), which yields

$$P = \prod_{j=1}^{4} E_j = K_1^2 K_2^2,$$

$$R(\mu) = \left[\mu \prod_{j=1}^{4} (\mu - E_j)\right]^{1/2} \equiv (\mu)^{1/2}(\mu + K_1)(\mu + K_2), \tag{VI.8a}$$

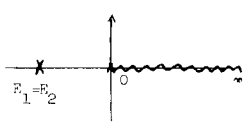


FIG. 15. Spectrum in E-plane for single soliton.

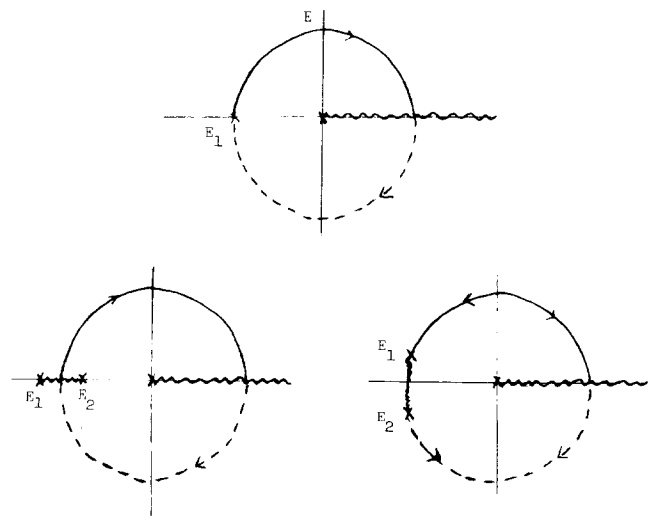


FIG. 16. Single phase μ-paths.

and we fix $P^{1/2} = +K_1K_2$. With these ingredients, the $N=2$ μ-representation of $u(x,t)$ becomes

$$u(x,t) = i\ln(\mu_1\mu_2/K_1K_2), \tag{VI.8b}$$

where

$$(\mu_k)_x = 2i\left[\left(1 \mp \frac{\mu_{j\neq k}}{16K_1K_2}\right)\frac{R(\mu_k)}{(\mu_k - \mu_{j\neq k})}\right]. \tag{VI.8c}$$

The same algebraic manipulations which were used to linearize the general μ equations in Sec. IV will place these μ equations (VI.8c) in the form

$$\frac{d\mu_1}{R(\mu_1)} + \frac{d\mu_2}{R(\mu_2)} = \frac{\pm 2i}{16K_1K_2}\begin{pmatrix} dx \\ dt \end{pmatrix},$$
$$\frac{\mu_1 d\mu_1}{R(\mu_1)} + \frac{\mu_2 d\mu_2}{R(\mu_2)} = 2i\begin{pmatrix} dx \\ dt \end{pmatrix}. \tag{VI.9}$$

In this degenerate case, the integration of these equations is elementary. [In the general nondegenerate case of Sec. (IV), their integration requires hyperelliptic integrals]. After this elementary integration and algebraic simplification, Eq. (VI.9) become

$$A_j(x,t) = e^{\theta_j(x,t)}A_j^0, \tag{IV.10a}$$

with A_j defined in terms of μ_j by

$$A_j \equiv \left(\frac{1 - i(\mu_1/K_j)^{1/2}}{1 + i(\mu_1/K_j)^{1/2}}\right)\left(\frac{1 - i(\mu_2/K_j)^{1/2}}{1 + i(\mu_2/K_j)^{1/2}}\right), \tag{IV.10b}$$

and with the phases θ_j given by

$$\theta_j \equiv 2(\sqrt{K_j} + 1/16\sqrt{K_j})(x - x_0) + (\sqrt{K_j} - 1/16\sqrt{K_j})(t - t_0). \tag{VI.11}$$

Alternatively, consideration of combinations of $\sqrt{\mu_1}, \sqrt{\mu_2}$ (rather than μ_1, μ_2 individually) leads to the equivalent representation:

$$\left(\frac{\mu_1\mu_2}{K_1K_2}\right)^{1/2} = \frac{\sqrt{K_1}(A_1^0e^{\theta_1} - 1)(A_2^0e^{\theta_2} + 1) - \sqrt{K_2}(A_1^0e^{\theta_1} + 1)(A_2^0e^{\theta_2} - 1)}{\sqrt{K_2}(A_1^0e^{\theta_1} - 1)(A_2^0e^{\theta_2} + 1) - \sqrt{K_1}(A_1^0e^{\theta_1} + 1)(A_2^0e^{\theta_2} - 1)}, \tag{VI.12a}$$

$$\sqrt{\mu_1} + \sqrt{\mu_2} = \frac{i(K_1 - K_2)(A_1^0e^{\theta_1} - 1)(A_2^0e^{\theta_2} - 1)}{\sqrt{K_1}(A_1^0e^{\theta_1} + 1)(A_2^0e^{\theta_2} - 1) - \sqrt{K_2}(A_1^0e^{\theta_1} - 1)(A_2^0e^{\theta_2} + 1)}, \tag{VI.12b}$$

where

$$\exp(iu/2) = \mu_1\mu_2/K_1K_2. \tag{VI.12c}$$

As in the $N=1$ case, these representations will in general yield complex u and w. We must select the integration constants A_j^0 to ensure reality of u and w. First, we make this selection for the kink–kink case, then for the breather.

Case i (Kink-kink pair: $K_1 > K_2 > 0$.) In this case, Eqs. (VI.11) show the phases θ_1 and θ_2 are real. The constraint A_j^0 purely imaginary will ensure real u and w. To verify this constraint, allow $\theta_1 \to \infty$ in (VI.12). This forces $\mu_1 \to -K_1$. If A_2^0 is chosen purely imaginary, μ_2 will lie on a circle of radius K_2 and the potentials u and w will be real. Similarly, allowing $\theta_2 \to \infty$ in (VI.12) forces $\mu_2 \to -K_2$. The choice A_1^0 purely imaginary then places μ_1 on a circle of radius K_1 and yields real u and w. Since the A_j^0 are integration constants, they are independent of θ_1 and θ_2; these choices are valid for all θ_1 and θ_2.

Finally, consider the μ paths for this kink–kink pair. When μ_1 is held at $-K_1$, the μ_2 path is a circle of radius K_2 centered at the origin. When μ_2 is held at $-K_2$, the μ_1 path is a circle of radius K_1. These configurations place one soliton at ∞; the second soliton does not interact at all with the one at ∞. As μ_1 moves away from $-K_1$ (and the first soliton moves in from ∞), the μ_2 path is no longer a perfect circle. This distortion accounts for the effect of the first soliton on the second. Indeed, its path depends upon the value μ_1 through (VI.12a) and (VI.12b). Consideration of these μ paths leads one to guess that in the periodic generalization, a *kink–kink-train*, the μ_i paths will qualitatively appear as shown below in Fig. 17.

Case ii (Breather: $K_1 = K_2^*$). Since K_1, K_2 are complex, two *real* phases must be deduced from the complex phases θ_1, θ_2 of Eq. (VI.11). These real phases will label physically the two degrees of freedom in the breather waveform: (1) the translation of the breather envelope at a fixed speed, and (2) the beating or "breathing" of the envelope at a fixed frequency. $K_1 = K_2^*$ shows $\theta_1 = \theta_2^*$ and two real phases, η_1 and η_2, are identified simply as the real and imaginary parts of θ_1, θ_2. We therefore define

$$\eta_1 \equiv -2\sin(\phi/2)\left[\left(\sqrt{K} + \frac{1}{16\sqrt{K}}\right)(x - x_0) + \left(\sqrt{K} - \frac{1}{16\sqrt{K}}\right)(t - t_0)\right], \tag{VI.13}$$

$$\eta_2 \equiv 2\cos(\phi/2)\left[\left(\sqrt{K} - \frac{1}{16\sqrt{K}}\right)(x - x_0) + \left(\sqrt{K} + \frac{1}{16\sqrt{K}}\right)(t - t_0)\right], \tag{VI.14}$$

where

$$\theta_1 = \eta_1 + i\eta_2, \quad \theta_2 = \eta_1 - i\eta_2, \tag{VI.15}$$

and

$$E_1 = E_2^* = Ke^{i\phi}$$
$$(\sqrt{K_1} = i(K)^{1/2}e^{i\phi/2}, \sqrt{K_2} = -i(K)^{1/2}e^{-i\phi/2}).$$

Next we pick the integration constants A_j^0 to ensure reality. The asymptotic trick which we employed for the kink–kink pair is not available for the breather—the two degrees of freedom do not separate. Instead, a tedious exercise is required to determine the constraints. We sketch the details next.

From Eq. (VI.12c), reality of $u(x,t)$ clearly implies the constraint

$$|\mu_1(x,t)\mu_2(x,t)| = K_1K_2 = K^2. \tag{VI.16}$$

This gives one piece of information in determining the

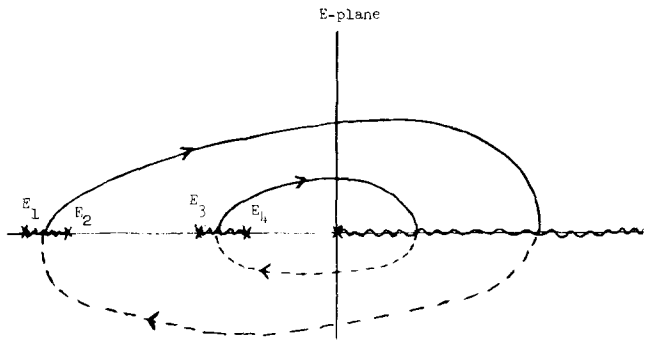


FIG. 17. μ-paths for kink-kink train.

constraints:

$$|\mu_1^0\mu_2^0| = K^2. \tag{VI.17}$$

Next we work with Eq. (VI.12a) for $(\mu_1\mu_2)^{1/2}/K$, which by (VI.16) must have modulus one. This condition implies

$$A_1^0 = (A_2^0)^*. \tag{VI.18a}$$

Moreover, using the particular form of A_1^0, A_2^0, Eq. (VI.10b), together with the constraint $|\mu_1^0\mu_2^0| = K^2$, the condition $\mathrm{Re}A_1^0 = \mathrm{Re}A_2^0$ actually forces both quantities to vanish:

$$\mathrm{Re}A_1^0 = \mathrm{Re}A_2^0 \equiv 0. \tag{VI.18b}$$

Thus, reality of $u(x,t)$ leads to

$$A_1^0 \equiv i|A_1^0|, \quad A_2^0 \equiv -i|A_1^0| = (A_1^0)^*. \tag{VI.18c}$$

Equation (VI.12a) then simplifies to

$$\frac{-(\mu_1\mu_2)^{1/2}}{K} = \frac{\left[1 + i\tan\frac{\phi}{2}\left(\frac{\sin(\eta_2 - P)}{\cosh(\eta_1 + \ln|A_1^0|)}\right)\right]}{\left[1 - i\tan\frac{\phi}{2}\left(\frac{\sin(\eta_2 - P)}{\cosh(\eta_1 + \ln|A_1^0|)}\right)\right]}, \tag{VI.19}$$

where $P = \frac{\pi}{2} - \mathrm{ph}(A_1^0) = \begin{cases} 0 & \text{if } \mathrm{Im}A_1^0 > 0, \\ -\pi & \text{if } \mathrm{Im}A_1^0 < 0. \end{cases}$

Since $u(x,t) = 2i\ln((\mu_1\mu_2)^{1/2}/K)$, Eq. (VI.12c), we arrive at the usual breather formula

$$u(x,t) = 4\arctan\left(\frac{\tan(\phi/2)\sin\eta_2}{\cosh(\eta_1 + \ln|A_1^0|)}\right). \tag{VI.20}$$

The physical characteristics of the breather are clearly labeled in terms of the simple spectrum $\{E_1, E_2 = E_1^*\}$; in particular, (i) the phase $\eta_1(x,t)$ labels the translation of the envelope, while $\eta_2(x,t)$ labels the "breathing" of the envelope, (ii) from $\eta_1(x,t)$, we find the envelope velocity (U) depends only on $K = |E_1|$, given by (for $P^{1/2} > 0$)

$$U = \frac{16K - 1}{16K + 1},$$

(iii) from $\eta_2(x,t)$, the beating frequency (ν) of the envelope in its own frame depends only on ϕ = phase (E_1),

$$\nu = \cos(\phi/2).$$

These formulas clearly show how the two breather degrees of freedom vary with the location of $E_1, E_2 = E_1^*$.

We return now to the determination of the paths of $\mu_1(x,t), \mu_2(x,t)$. Notice $(\mu_1\mu_2)^{1/2}$ lives on a circle [see (VI.17)]. Using this constraint, together with (VI.8c), we deduce from (VI.12b) that $(\sqrt{\mu_1} + \sqrt{\mu_2})$ is purely imaginary. To obtain more detailed information about these μ paths, we use a computer to evaluate and plot μ_1 and μ_2 from (VI.12a) and (VI.12b) for a variety of input parameters $(E_1, E_2 = E_1^*, \mu_1^0, \mu_2^0)$. The results of these computer plots can be summarized as follows.

(1) As expected, for fixed values of E_1 and $E_2 = E_1^*$, the paths of $\mu_1(x,t), \mu_2(x,t)$ do not lie on any *fixed* curve in the complex plane. Indeed, the path of μ_1 depends continuously on the value of μ_2. However, one might expect that, for fixed E_1 and $E_2 = E_1^*$, all cycles of $\mu_1(x,t)$ [respectively $\mu_2(x,t)$] will be equivalent as paths of integration on the Riemann surface of $R(\mu)$. (This means all paths of μ_1, respectively μ_2, are in the same homology class on the Riemann surface.) Even this is *not* the case.

(2) Fix $E_1, E_2 = E_1^*$, and choose μ_1^0, μ_2^0 consistent with reality of $u(x,t)$. To observe *closed* cycles of μ_1, μ_2 (i.e., *periodic orbits*) one must fix a value of the translation phase $\eta_1(x,t)$ and then flow according to the breathing phase $\eta_2(x,t)$. Physically, this amounts to riding with the breather along the straight line motion in x,t space determined by $\eta_1(x,t) \equiv$ const, thereby experiencing only the periodic beating of the envelope. The following surprising result appears generic for every fixed location of $E_1, E_2 = E_1^*$.

(3) There is a critical value $\eta_1{}^c$ of the translation phase $\eta_1(x,t)$ corresponding to a critical line $[\eta_1(x,t) \equiv \eta_1{}^c]$ in x,t space. In the "half-space" $\eta_1(x,t) < \eta_1{}^c$, all paths of $\mu_1(x,t)$ [respectively $\mu_2(x,t)$] under the flow of η_2 are equivalent, as paths of integration, to a closed loop Γ_1 (respectively Γ_2), such that: $E_i \in \mathrm{Int}\Gamma_i; 0, E_i^* \in \mathrm{Ext}\Gamma_i$. (See Fig. 18.) In the other half-space $\eta_1(x,t) > \eta_1{}^c$, the μ cycles have interchanged roles, that is, $\mu_1(x,t)$ cycle $= \Gamma_2$ and $\mu_2(x,t)$ cycle $= \Gamma_1$.

By considering a sequence of plots of $\mu_1(x,t), \mu_2(x,t)$ along $\eta_1(x,t) \equiv C$, with C ranging through values centered at $\eta_1{}^c$, we observe the smooth evolution of Fig. 18(a) into Fig. 18(b). The critical value $\eta_1{}^c$ can only be determined approximately. Along the lines $\eta_1(x,t) \equiv a, a \cong \eta_1{}^c, a < \eta_1{}^c$, Fig. 19(a) depicts the μ_1 path. Then along the neighboring line $\eta_1(x,t) \equiv b, b \cong \eta_1{}^c, b > \eta_1{}^c$, Fig. 19(b) depicts the μ_1 path. (The paths are switched for μ_2.)

(5) There are two unusual events occuring here: (a) the $\mu_i(x,t)$ appear to pass right through the stationary points E_1, E_2 of the ODE's (VI.8c), and (b) $\mu_1(x,t)$ and $\mu_2(x,t)$ evidently collide, corresponding to a singularity of the ODE's (VI.8c). Both events violate the smooth evolution of these μ paths described above in (4), *unless they occur simultaneous-*

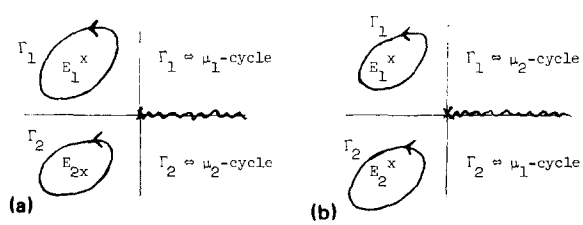


FIG. 18. (a) Half-space $\eta_1(s,t) < \eta_1{}^c$, (b) half-space $\eta_1(x,t) > \eta_1{}^c$.

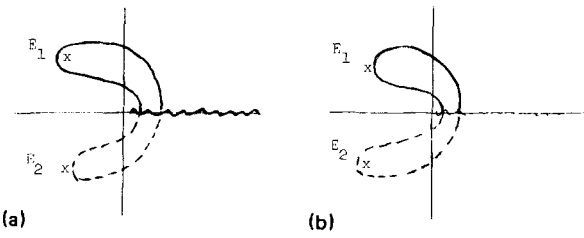


FIG. 19. (a) μ_1 path for $\eta_1(x,t)\equiv a \lesssim \eta_1{}^c$, μ_2 path for $\eta_1(x,t)\equiv b \gtrsim \eta_1{}^c$. (b) μ_1 path for $\eta_1(x,t)\equiv b \gtrsim \eta_1{}^c$, μ_2 path for $\eta_1(x,t)\equiv a \lesssim \eta_1{}^c$.

ly. That is, for example, $\mu_1 \to \mu_2$ in such a way that the limit of $\mu_1 - E_1/\mu_1 - \mu_2$ is finite and nonzero [refer to (VI.8c)]. This behavior occurs along $\eta_1(x,t)\equiv\eta_1{}^c$, and only at one instant in (x,t).

Remarks: (1) We have therefore shown that even for this simplified case of pure solitons, the spectral variables $\mu_j(x,t)$ do not lie in any fixed homology class (class of equivalent paths of integration). This is consistent with the findings of McKean[27] and Trubowitz.[28] This is somewhat disheartening; the μ variables are clearly the natural variables for all the manipulations displayed thus far, yet appear to be unnatural topologically. One hopes (at least we do) that these apparent pathologies can be overcome, perhaps by a suitable change of coordinates.

(2) The inferences about the periodic problem based on this breather analysis are that *in a certain confined region of x,t space*, the μ_j cycle contains the canonical a_j cycle, $j = 1,2$ (Fig. 20). In the complement of that region, the μ_1,μ_2 cycles are interchanged. Apparently μ_1 can collide with μ_2, in which case the coordinates break down.

VII. SEPARABLE SOLUTIONS ($N = 2$)

1. Definitions and motivation

Separable solutions of the sine-Gordon equation

$$u_{tt} - u_{xx} + \sin u = 0, \tag{VII.1}$$

are defined by the ansatz that the x,t dependence separates, for example,

$$u(x,t) = 4\arctan(f(x)g(t)). \tag{VII.2}$$

This ansatz was initially suggested by Lamb[29] in the context of optical pulse propagation. Since then, Costabile *et al.*[1] and Fulton[30] have applied separable solutions to study the oscillatory behavior of finite one-dimensional Josephson transmission lines, and Trullinger[31] has used such solutions to analyze boundary effects on charge density waves. The most detailed mathematical study of the "separability" of the sine-Gordon equation is given by Osborne and Stuart.[32] The salient feature of the above references is the fact that the separable ansatz, (VII.2), leads to elliptic function solutions for $f(x)$ and $g(t)$; moreover, from Refs. 1 and 30 it is clear on physical grounds that these oscillatory sine-Gordon solutions have at least two degrees of freedom, and therefore represent the nonlinear interaction of two or more periodic traveling waves.

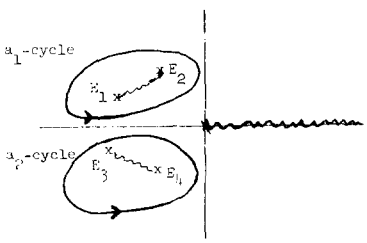


FIG. 20. Canonical a cycles. Genus 2.

We know (refer to Sec. IV) that the general representation of periodic $N = 2$-phase waves leads to *hyperelliptic functions* on the genus 2 Riemann surface of

$$R^2(E) = E\prod_{k=1}^{4}(E - E_k), \quad \sum^{(S)} \equiv \{E_1,\ldots E_4\},$$

not to *elliptic functions* on a genus 1 Riemann surface. The question naturally arises as to the source of this degeneracy in the function theory. In this section, we identify and describe this degeneracy. We will analyze the special case studied in Ref. 1, the "open circuit" boundary conditions,

$$u_x(0,t) = u_x(L,t) = 0. \tag{VII.3}$$

The authors give explicit 2 degrees of freedom separable solutions satisfying these boundary conditions. There are many other types (as discussed in Ref. 32) of separable solutions, which can be analyzed similarly. Our approach is outlined below.

We first show that the "open circuit" boundary conditions, (VII.3), result from spatial symmetry and periodicity of the initial data $\overset{\circ}{u}(x),\overset{\circ}{v}(x)$; therefore, N-phase solutions with one additional constraint (initial data symmetric about $x_0 = 0$) will satisfy open circuit boundary conditions. We then show how this spatial symmetry implies a spectral symmetry $E_j \Rightarrow 1/16^2 E_j$. With this symmetry in the simple spectrum $\Sigma^{(S)}$, the μ-representation and Θ-function representation are used to show these special solutions are *standing waves* whose x,t flows *separate* and can be explicitly integrated in terms of *elliptic functions*.

2. Open circuit ⇔ spatial symmetry

We first show the open circuit boundary conditions (VII.3) result from periodic initial data which are even functions about $x_0 = 0$. The converse is much easier and does not require proof. We state the result in

Theorem VII.1: Let $u(x,t)$ be a solution of the sine-Gordon equation, (VI.1), with smooth, periodic initial data $u(x,t=0)\equiv\overset{\circ}{u}(x), u_t(x,t=0)\equiv\overset{\circ}{v}(x)$, even about $x = 0$:

$$\begin{aligned}&\overset{\circ}{u}(x+L) = \overset{\circ}{u}(x)(\mathrm{mod}2\pi),\\ &\overset{\circ}{u}(-x) = \overset{\circ}{u}(x),\\ &\overset{\circ}{v}(x+L) = \overset{\circ}{v}(x),\\ &\overset{\circ}{v}(-x) = \overset{\circ}{v}(x).\end{aligned} \tag{VII.4}$$

Then: $u(x,t)$ is even and periodic in x for all time t, and satisfies the open circuit boundary conditions at $x = 0, \pm 2L,\ldots$. That is, $u(x,t)$ satisfies

(i) $u(x+L,t) = u(x,t)(\mathrm{mod}2\pi)$,

(ii) $u(-x,t) = u(x,t)$,

(iii) $u_x(0,t) = u_x(nL,t) = 0$.

Proof of Theorem: Part (i) is established by McKean.[27] To prove Part (ii), we begin with $u(x,t)$ satisfying the hypotheses $(\mathring{u}(-x),\mathring{v}(-x)) = (\mathring{u}(x),\mathring{v}(x))$. Now define $u_{\pm}(x,t)$ by

$$u_{\pm}(x,t) = \tfrac{1}{2}(u(x,t) \pm u(-x,t)).$$

We aim to show $u_-(x,t) \equiv 0$. It follows from the sine-Gordon equation that $u_{\pm}(x,t)$ satisfy the system

$$(\partial_{tt} - \partial_{xx})u_+ + \sin u_+ \cos u_- = 0,$$
$$(\partial_{tt} - \partial_{xx})u_- + \sin u_- \cos u_+ = 0,$$
$$u_{\pm}(x,0) = \mathring{u}_{\pm}(x),$$
$$\partial_t u_{\pm}(x,0) = \mathring{v}_{\pm}(x).$$

But $\begin{cases} \mathring{u}(-x) = \mathring{u}(x) \\ \mathring{v}(-x) = \mathring{v}(x) \end{cases}$ implies $\begin{cases} \mathring{u}_-(x) = 0, \\ \mathring{v}_-(x) = 0, \end{cases}$

so that $u_-(x,t) \equiv 0$, proving Part (ii). Part (iii) follows rather easily, since $(\mathring{u}(-x),\mathring{v}(-x)) = (\mathring{u}(x),\mathring{v}(x))$ together with smoothness implies $\partial_x u(x,t)|_{x=0} = 0$. Periodicity then gives $u_x(x,t)|_{x=nL} = 0$ for any integer n.

3. Spatial symmetry ⇔ spectral symmetry

We now use the Floquet theory of Sec. II to characterize even potentials by a symmetry in the spectrum. We begin with

Theorem VII.2: Consider the Takhatajian–Faddeev eigenvalue problem (II.6), with periodic, even initial data

$$(\mathring{u}(-x),\mathring{v}(-x)) = (\mathring{u}(x),\mathring{v}(x)),$$
$$\mathring{u}(x+L) = \mathring{u}(x) + 2\pi M, M \equiv \text{"charge" of } \mathring{u}(x),$$
$$\mathring{v}(x+L) = \mathring{v}(x).$$

Then the Floquet discriminant $\Delta(E)$ satisfies the symmetry relation

$$\Delta(1/16^2 E) = (-1)^M \Delta(E).$$

The proof is given in Appendix C. (We also note the converse is true). It is now an easy exercise to deduce symmetries in the simple spectrum $\Sigma^{(S)}$. In particular, we find

Corollary VII.2. Under spatial symmetry as in Theorem VII.2,

$$E_j \in \sum^{(S)} \Rightarrow 1/16^2 E_j \in \sum^{(S)}. \tag{VII.5}$$

Consider the implications of this corollary for the single-phase case. From Sec. V, we know the $N=1$ simple spectrum $\Sigma^{(S)}_{N=1}$ occurs in only two forms $\Sigma^{(S)}_{N=1} = \{E_1 < E_2 < 0\}$ or $\{E_1 = E_2^*, E_1 \neq E_2\}$. For the *kink train case*, $E_1 < E_2 < 0$, the symmetry (VIII.5) implies $E_2 = 1/16^2 E_1$. But the phase velocity U satisfies (Sec. V)

$$U = \frac{16(E_1E_2)^{1/2} - 1}{16(E_1E_2)^{1/2} + 1};$$

since U cannot be infinite,[33] we find that the only open circuit kink trains are at rest ($U=0$). The oscillatory case, $E_1 = E_2^*, E_1 \neq E_2$, yields the same result. Physically, this time-independent nature of open circuit single-phase solutions is quite obvious. For example, in the Josephson transmission line, u represents the magnetic flux. A single-phase *traveling* wave u will either transfer flux through the left ($U<0$) or right ($U>0$) boundary, violating the open circuit boundary conditions; thus, stationary waves.

The physical considerations of Refs. 1 and 30 imply nonstationary separable solutions must be standing waves. This forces $N \geqslant 2$; next, we investigate $N=2$-phase solutions (Sec. IV) in the light of this spectral symmetry.

4. Consequences of the spectral symmetry ($N=2$): separability, standing waves, and elliptic functions

We now use the spectral symmetry, which characterizes the sine-Gordon solutions satisfying open circuit boundary conditions, to show first that these solutions are, in fact, separable, and second, to describe the degeneracy in the function theory. Our analysis uses both Θ function and μ-representations of $u(x,t)$ (Sec. IV). The μ-representation is summarized by

$$u(x,t) = i \ln(\mu_1(x,t)\mu_2(x,t)/P^{1/2}), \tag{VII.6a}$$

where $P = \Pi E_j, \Sigma^{(S)} \equiv \{E_1,\ldots,E_4\}$, and $\mu_{1,2}$ satisfy

$$(\mu_l)_t = 2i\left(1 \mp \frac{\mu_{j\neq l}}{16P^{1/2}}\right)\left[\mu_l \prod_{j=1}^{4}(\mu_l - E_j)\right]^{1/2} \Big/ \prod_{j\neq l}(\mu_l - \mu_j). \tag{VII.6b}$$

$$u(x,t) = 2i \ln\left(\frac{\Theta(\mathbf{l}(x,t) + \frac{1}{2};B)}{\Theta(\mathbf{l}(x,t);B)}\right), \tag{VII.7a}$$

where $\mathbf{l}(x,t) = (l_1(x,t), l_2(x,t))$,

$$l_\nu(x,t) = -2i\left[\left(C_{\nu 1} + \frac{1}{16P^{1/2}}C_{\nu 2}\right)x + \left(C_{\nu 1} - \frac{1}{16P^{1/2}}C_{\nu 2}\right)t\right] + l_\nu(0,0), \tag{VII.7b}$$

$$E_5 \equiv 0, \quad P = \prod_{k=1}^{4} E_k.$$

First note that the spectral symmetry $E_j \in \Sigma^{(S)} \Rightarrow 1/16^2 E_j \in \Sigma^{(S)}$ forces

$$\sum^{(S)} \equiv \{E_1, E_2, 1/16^2 E_1, 1/16^2 E_2\}, \tag{VII.8a}$$

$$R^2(E) = E\prod_{j=1}^{4}(E - E_j)(E - 1/16^2 E_j), \tag{VII.8b}$$

$$P = \prod_{j=1}^{4} E_j \equiv \frac{1}{16^4}. \tag{VII.8c}$$

With $\Sigma^{(S)}$ given by (III.8a), we consider two configurations of the spectrum; for convenience, we list and graph these below, together with the branch cut structure and canonical a_i cycles. We note that $\Delta^2(E)$ (and therefore $\Sigma^{(S)}$) is invariant under the map $E \to 1/16^2 E$, and moreover the circle $|E| = \frac{1}{16}$ is mapped onto itself by this transformation. This fact is manifested in Figs. 21(a) and 21(b); in each case, there is a clear symmetry about the circle $|E| = \frac{1}{16}$. (The entire analysis which follows rests on this observation.)

This spectral symmetry leads to analogous relations among the holomorphic differentials on the Riemann surface of $R(E)$, Eq. (VII.8b). We describe these facts in

Case 1. Two trains of kinks, antikinks:

$$E_1 < E_2 < 1/16^2 E_2 < 1/16^2 E_1.$$

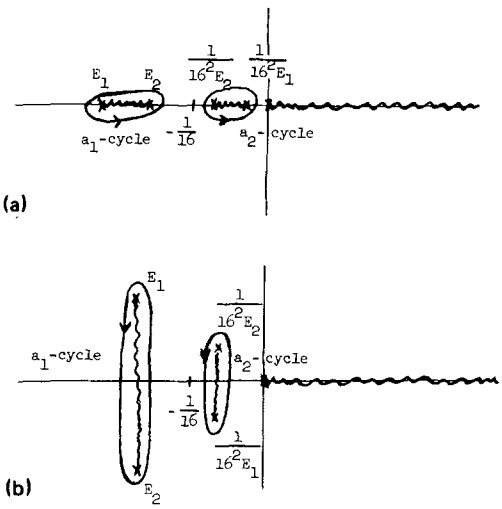


FIG. 21. Branch cut structures for two-phase solutions with open circuit boundary conditions. (a) Branch cuts and a_i cycles for 2-phase kink–kink trains under open circuit B.C. (b) Branch cuts and a_i cycles for 2-phase breather or plasma oscillations satisfying open circuit B.C.

Case 2. *Breather trains, plasma oscillations*:

$E_1, E_1^*, 1/16^2E_1, 1/16^2E_1^*$, with $|E_1| > \frac{1}{16}$.

*Lemma VII.*1. Define a basis $\{dI, dJ\}$ of holomorphic differentials on the Riemann surface of $R^2(E) = E\,\Pi_{j=1}^2(E - E_j)(E - 1/16^2E_j)$ by

$$dI \equiv \frac{dE}{R(E)}, \quad dJ \equiv \frac{EdE}{R(E)}. \tag{VII.9}$$

In terms of these differentials, denote the a_i periods (with the a_i cycles depicted in Figs. 21) by

$$I(a_i) \equiv \oint_{a_i} dI, \quad J(a_i) \equiv \oint_{a_i} dJ. \tag{VII.10}$$

Then:

(i) $dI(E) = -16dJ(\epsilon)$,

$-16dJ(E) = dI(\epsilon)$, where $\epsilon = 1/16^2E$,

(ii) $I(a_1) = 16J(a_2)$,

$I(a_2) = 16J(a_1)$.

The proof is an immediate consequence of the change of variables $\epsilon = 1/16^2E$.

With this lemma, we can immediately deduce

Fact 1 (*Standing waves*): $N = 2$-phase solutions with the spectral symmetry $E_j \in \Sigma^{(S)} \Rightarrow 1/16^2E_j \in \Sigma^{(S)}$ are *standing waves*. The proof of this fact comes by analyzing the two phases $l_{1,2}(x,t)$ in the θ-function representation (VII.7). The normalization constants $C_{\nu\mu}$, as defined by (IV.19b), are given in terms of $I(a_i), J(a_i)$ by [with $W \equiv J(a_1)I(a_2) - I(a_1)J(a_2)$]

$$C_{11} = \frac{I(a_2)}{W}, \quad C_{12} = \frac{-J(a_2)}{W},$$

$$C_{21} = \frac{-I(a_1)}{W}, \quad C_{22} = \frac{J(a_1)}{W}.$$

Now using Lemma VII.1, Part (ii), the phases $l_{1,2}(x,t)$ become [with the choice $P^{1/2} = (1/16^4)^{1/2} = +1/16^2$]

$$l_1(x,t) = \frac{-32i}{[I(a_2)]^2 - [I(a_1)]^2} \times [(I(a_2) - I(a_1))x + (I(a_2) + I(a_1))t] + l_1(0,0),$$

$$l_2(x,t) = \frac{-32i}{[I(a_2)]^2 - [I(a_1)]^2} \times [(I(a_2) - I(a_1))x - (I(a_2) + I(a_1))t] + l_2(0,0).$$

Thus, *the phase velocities are equal but opposite: standing waves.*

The next result we state as

Fact 2 (*Separability*): The $\binom{x}{t}$-flows explicitly *separate*. That is, there exists one combination of the variables μ_1 and μ_2 which depends on time t and is independent of x, and a second combination which depends on x and is independent of t.

We use the μ equations (VII.6b):

$$(\mu_1)_x = 2i(1 \pm 16\mu_2)R(\mu_1)/(\mu_1 - \mu_2),$$
$$(\mu_2)_x = 2i(1 \pm 16\mu_1)R(\mu_2)/(\mu_2 - \mu_1). \tag{VII.11}$$

which can be written in the equivalent differential form:

$$\frac{d\mu_1}{R(\mu_1)}(\mu_1 - \mu_2) = 2i(1 \mp 16\mu_2)\binom{dx}{dt}, \tag{VII.12a}$$

$$\frac{d\mu_2}{R(\mu_2)}(\mu_1 - \mu_2) = 2i(-1 \pm 16\mu_1)\binom{dx}{dt}. \tag{VII.12b}$$

Algebraic manipulation then yields quite simple expressions for the $\binom{x}{t}$-flows:

$$\frac{d\mu_1}{R(\mu_1)} + \frac{d\mu_2}{R(\mu_2)} = \pm 32i\binom{dx}{dt}, \tag{VII.13a}$$

$$\frac{16\mu_1 d\mu_1}{R(\mu_1)} + \frac{16\mu_2 d\mu_2}{R(\mu_2)} = 32i\binom{dx}{dt}. \tag{VII.13b}$$

Then, simply adding the dx equations and subtracting the dt equations in (VII.13a) and (VII.13b), we find the $\binom{x}{t}$-flows explicitly *separate*. The x dependence is given by

$$\left[\frac{1 + 16\mu_1}{R(\mu_1)}\right](\mu_1)_x + \left[\frac{1 + 16\mu_2}{R(\mu_2)}\right](\mu_2)_x = 64i,$$
$$\left[\frac{1 + 16\mu_1}{R(\mu_1)}\right](\mu_1)_t + \left[\frac{1 + 16\mu_2}{R(\mu_2)}\right](\mu_2)_t = 0, \tag{VII.14a}$$

while the t dependence is given by

$$\left[\frac{1 - 16\mu_1}{R(\mu_1)}\right](\mu_1)_t + \left[\frac{1 - 16\mu_2}{R(\mu_2)}\right](\mu_2)_t = -64i,$$
$$\left[\frac{1 - 16\mu_1}{R(\mu_1)}\right](\mu_1)_x + \left[\frac{1 - 16\mu_2}{R(\mu_2)}\right](\mu_2)_x = 0. \tag{VII.14b}$$

This means the integral

$$\int^{\mu_1}\left[\frac{1 + 16\mu}{R(\mu)}\right]d\mu + \int^{\mu_2}\left[\frac{1 + 16\mu}{R(\mu)}\right]d\mu$$

is independent of t and linear in x, while the integral

$$\int^{\mu_1}\left[\frac{1-16\mu}{R(\mu)}\right]d\mu+\int^{\mu_2}\left[\frac{1-16\mu}{R(\mu)}\right]d\mu$$

is independent of x and linear in t.

We can also conclude from these relations

Fact 3 (*Elliptic functions*): The $\binom{x}{t}$-flow of the 2-phase solutions with the spectral symmetry $E_j \Rightarrow 1/16^2E_j$ in $\Sigma^{(S)}$ depends on *elliptic* functions which live on one of two genus 1 Riemann surfaces:

$$w(z)=\left|(z\pm\tfrac{1}{16})(z-z_1)(z-z_2)\right|^{1/2}, \tag{VII.15a}$$

where

$$z_i\equiv\tfrac{1}{2}(E_i+1/16^2E_i),\quad i=1,2, \tag{VII.15b}$$

$$\sum^{(S)}=\{E_1,1/16^2E_1,E_2,1/16^2E_2\}.$$

To see this fact, we first note from Fact 2 and Eq. (VII.14) that the x flow is characterized in terms of one particular differential, $(1+16E)dE/R(E)$, as opposed to the usual case which requires both differentials in the holomorphic basis (see the θ-function representation). Similarly, the t flow depends only on the one differential $[(1-16E)/R(E)]dE$. The symmetry in $R(E)$ suggests a transformation which reduces the $\binom{x}{t}$-differentials to elliptic. The map that accomplishes this reduction is

$$z=\tfrac{1}{2}(E+1/16^2E); \tag{VII.16}$$

the resulting differentials for the $\binom{x}{t}$-flows are

$$x\text{-flow:}\quad\frac{(1+16E)dE}{R(E)}\to\frac{-16dz}{\left|2(z-\frac{1}{16})(z-z_1)(z-z_2)\right|^{1/2}},$$

$$t\text{-flow:}\quad\frac{(1-16E)dE}{R(E)}\to\frac{-16dz}{\left|2(z+\frac{1}{16})(z-z_1)(z-z_2)\right|^{1/2}},$$

with z_i defined in Fact 3. This proves Fact 3.

We emphasize these differentials involve square roots of cubics in z; they are elliptic. In general, they yield elliptic functions with distinct moduli.

We remark that the map $z=\frac{1}{2}(E+1/16^2E)$ is invariant to the spectral symmetry $E_j\Rightarrow 1/16^2E_j$ (which, of course, motivated its use) and therefore does not distinguish between the points $E_j, 1/16^2E_j$. The branch points become one in the image space, effecting the reduction to elliptic functions. The appearance of $\pm\frac{1}{16}$ as branch points for the $\binom{x}{t}$-flow is related to two facts. First, $\pm\frac{1}{16}$ are the fixed points of the mapping $z(E)$. Second, the circle of symmetry in the E plane $|E|=\frac{1}{16}$, gets mapped onto the slit $-\frac{1}{16}<z<+\frac{1}{16}$.

In this section, we have described a class of "open-circuit separable solutions" in terms of inverse spectral theory. The generality of these separable solutions, as well as the origin of their elliptic nature, becomes clear from this inverse spectral analysis. Other classes of separable solutions could certainly be analyzed by similar symmetries in the spectral transform. These two-phase separable solutions could be used (i) to clarify theoretical studies, since they are the most concrete examples of 2-phase periodic waves, and (ii) to study, by the methods of inverse spectral theory, the effects of boundary conditions on the sine-Gordon equation.

Note added in proof: Since this work was completed, some progress has been made in the understanding of real, N-phase sine-Gordon waves by E. Date of Kyote University and by N. Ercolani and G. Forest, Ohio State University.

ACKNOWLEDGMENTS

We have enjoyed many discussions with our colleague H. Flaschka. All sections of this survey were directly influenced by these interactions. We also acknowledge conversations with H. McKean and E. Trubowitz, who clarified for us difficulties with the μ coordinates for sine-Gordon waves. Professor McKean also made his manuscript available to us before its publication.

APPENDIX A: SCATTERING MOTIVATION

1. Infinite-line scattering properties

Consider the eigenvalue problem (II.2) under vanishing boundary conditions at $|x|=\infty$, (II.4b) and (II.4c). Here u and w are taken real.[5,6,8,34]

The "Jost" solutions are defined in terms of their asymptotic behavior at $x=\pm\infty$: the vanishing boundary conditions on $u(x)$, $w(x)$ reduce (II.2) to ($\lambda=\sqrt{E}$)

$$\left[\begin{pmatrix}0&-1\\1&0\end{pmatrix}\frac{d}{dx}+\left(\frac{1}{16\lambda}-\lambda\right)\right]\psi=\mathbf{0}. \tag{A1}$$

A basis of solutions for (A1) is

$$\psi_1=\begin{pmatrix}1\\-i\end{pmatrix}e^{-i(\lambda-1/16\lambda)x},\quad\psi_2=\begin{pmatrix}1\\i\end{pmatrix}e^{i(\lambda-1/16\lambda)x}.$$

The *Jost solutions* $\mathbf{f}_i,\mathbf{g}_i, i=1,2$ are then defined as solutions of the full problem (II.2) satisfying the following boundary conditions:

$$\mathbf{f}_1\sim\begin{pmatrix}1\\i\end{pmatrix}e^{i(\lambda-1/16\lambda)x}\quad\text{as}\quad x\to+\infty,$$

$$\mathbf{f}_2\sim\begin{pmatrix}1\\-i\end{pmatrix}e^{-i(\lambda-1/16\lambda)x}\quad\text{as}\quad x\to+\infty,$$

$$\mathbf{g}_1\sim\begin{pmatrix}1\\i\end{pmatrix}e^{i(\lambda-1/16\lambda)x}\quad\text{as}\quad x\to-\infty,$$

$$\mathbf{g}_2\sim\begin{pmatrix}1\\-i\end{pmatrix}e^{-i(\lambda-1/16\lambda)x}\quad\text{as}\quad x\to-\infty.$$

If we now define fundamental matrices for (II.2) in terms of these Jost solutions,

$$F(x,\lambda)\equiv(\mathbf{f}_1\mathbf{f}_2),\quad G(x,\lambda)\equiv(\mathbf{g}_1\mathbf{g}_2); \tag{A2}$$

then the transfer matrix $T(\lambda)$ maps the basis elements $\mathbf{g}_1,\mathbf{g}_2$, which behave like the "free" eigenfunctions at $x=-\infty$, across the influence of the potentials, into the solutions $\mathbf{f}_1,\mathbf{f}_2$, which behave like the "free" eigenfunctions at $x=+\infty$. That is,

$$F(x,\lambda)=G(x,\lambda)T(\lambda), \tag{A3}$$

$$T(\lambda)=\begin{pmatrix}t_{11}(\lambda)&t_{12}(\lambda)\\t_{21}(\lambda)&t_{22}(\lambda)\end{pmatrix}.$$

It follows that for λ *real*, $T(\lambda)$ can be expressed in terms of two complex-valued functions:

$$T(\lambda) = \begin{pmatrix} a(\lambda) & -b^*(\lambda) \\ b(\lambda) & a^*(\lambda) \end{pmatrix}, \quad \lambda \quad \text{real.} \tag{A4}$$

The realization of $a(\lambda), b(\lambda)$ in terms of reflection and transmission coefficients is illustrated by the following example. Consider an incoming plane wave from $x = +\infty$ with unit amplitude, which will impinge upon the potential, resulting in transmission through to $-\infty$ of some parts of the wave while the remainder is reflected back to $+\infty$.

From (A3), we have the following [after dividing by $a(\lambda)$]:

$$\frac{1}{a(\lambda)}\mathbf{g}_2(x,\lambda) = \frac{b^*(\lambda)}{a(\lambda)}\mathbf{f}_1(x,\lambda) + \mathbf{f}_2(x,\lambda).$$

Referring to the asymptotic nature of $\mathbf{f}_i, \mathbf{g}_i$, we write this equation in the asymptotic form:

$$\frac{1}{a(\lambda)}\begin{pmatrix} 1 \\ -i \end{pmatrix} e^{-i(\lambda - 1/16\lambda)x} \qquad x \to -\infty$$

$$\frac{b^*(\lambda)}{a(\lambda)}\begin{pmatrix} 1 \\ i \end{pmatrix} e^{i(\lambda - 1/16\lambda)x} + \begin{pmatrix} 1 \\ -i \end{pmatrix} e^{-i(\lambda - 1/16\lambda)x}. \tag{A5}$$

$$x \to +\infty$$

Since $e^{-i(\lambda - 1/16\lambda)x}$ refers to a left-running plane wave, while $e^{i(\lambda - 1/16\lambda)x}$ is right-running, we interpret (A5) as the free plane-wave solution of unit amplitude, $\binom{1}{-i} e^{-i(\lambda - 1/16\lambda)x}$, launched from $x = +\infty$ and traveling toward the potential; the term

$$\frac{b^*(\lambda)}{a(\lambda)}\begin{pmatrix} 1 \\ i \end{pmatrix} e^{i(\lambda - 1/16\lambda)x}$$

as $x \to +\infty$ is the reflected part of the wave, traveling to the right, back to $x = +\infty$, while the left-hand side,

$$\frac{1}{a(\lambda)}\begin{pmatrix} 1 \\ -i \end{pmatrix} e^{-i(\lambda - 1/16\lambda)x},$$

is the transmitted part of the wave which has made it through the potential and is traveling to the left out to $x = -\infty$.

Due to these interpretations, $1/a(\lambda) \equiv T_R(\lambda)$ is referred to as the *right transmission coefficient*, and $b^*(\lambda)/a(\lambda) \equiv R_R(\lambda)$ is the *right reflection coefficient*. In the same manner, we can launch the free plane-wave eigenfunction from $-\infty$,

$$\begin{pmatrix} 1 \\ i \end{pmatrix} e^{i(\lambda - 1/16\lambda)x},$$

which is right-running, and find the *left transmission coefficient* $T_L(\lambda) \equiv 1/a(\lambda)$ and the *left reflection coefficient* $R_L(\lambda) \equiv b(\lambda)/a(\lambda)$. (Use $\mathbf{g}_1 + [b(\lambda)/a(\lambda)]\mathbf{g}_2 = [1/a(\lambda)]\mathbf{f}_1$.) The following facts then follow about *whole-line scattering theory*.

(i) $T_R(\lambda) = T_L(\lambda) = 1/a(\lambda)$ is the transmission coefficient,

(ii) $\mathbf{f}_1, \mathbf{g}_2$ and thereby $a(\lambda)$ can be analytically continued into the upper-half λ plane, where $(\mathrm{Im}\lambda \geqslant 0)$

(iii)

$$a(\lambda) \to \begin{cases} 1 & \text{as} \quad \lambda \to \infty, \\ \exp[(i/2)(u(\infty) - u(-\infty))] & \text{as} \quad \lambda \to 0, \end{cases}$$

(iv) $|a(\lambda)|^2 + |b(\lambda)|^2 = 1$, λ real,

(v) the real λ axis is continuous spectrum and is associated with radiation in the sine-Gordon field,

(vi) $a(\lambda_j) = 0$ iff λ_j is a bound state eigenvalue; these occur on the positive imaginary axis or in pairs $(\lambda_j, -\lambda_j^*)$.

In the general case of vanishing boundary conditions at $|x| = \infty$, (II.4b) and (II.4c), the transfer matrix $T(\lambda)$, as defined by Eq. (A3), does not extend to the upper-half λ plane. Rather, $T(\lambda)$ is represented by four complex functions:

$$T(\lambda) = \begin{pmatrix} a(\lambda) & \overline{b(\lambda)} \\ b(\lambda) & \overline{a(\lambda)} \end{pmatrix}, \quad \lambda \in C, \tag{A6}$$

where $a(\lambda)$ is analytic in the upper-half λ plane, $\mathrm{Im}\lambda \geqslant 0$, and $\overline{a(\lambda)}$ is analytic in the lower-half λ plane, $\mathrm{Im}\lambda \leqslant 0$. In the overlapping region, $\mathrm{Im}\lambda = 0$, [see (A4)]

$$\overline{a(\lambda)} = a^*(\lambda), \quad \overline{b(\lambda)} = -b^*(\lambda), \quad \lambda \quad \text{real.} \tag{A7}$$

However, for the restricted class of compact support potentials, all of these functions $(a, b, \bar{a}, \bar{b})$ can be analytically continued into the whole λ plane.

2. Proof of Theorem III.1

From Theorem II.2,

$$\Delta(\lambda) = \phi_{+,1}(x_0 + L, x_0, \lambda) + \phi_{-,2}(x_0 + L, x_0, \lambda).$$

Using scattering theory of Sec. A1, we have the following "asymptotic" behavior near $x_0, x_0 + L$ [with $\alpha(\lambda) \equiv \lambda - 1/16\lambda$]:

$$\phi_+(x,x_0,\lambda) \sim \frac{1}{2}\begin{pmatrix} 1 \\ -i \end{pmatrix} e^{-i\alpha(\lambda)(x - x_0)} + \frac{1}{2}\begin{pmatrix} 1 \\ i \end{pmatrix} e^{i\alpha(\lambda)(x - x_0)} \quad \text{near} \quad x = x_0,$$

$$\phi_-(x,x_0,\lambda) \sim \frac{i}{2}\begin{pmatrix} 1 \\ -i \end{pmatrix} e^{-i\alpha(\lambda)(x - x_0)} + \frac{-i}{2}\begin{pmatrix} 1 \\ i \end{pmatrix} e^{i\alpha(\lambda)(x - x_0)} \quad \text{near} \quad x = x_0,$$

$$\phi_+(x,x_0,\lambda) \sim \frac{1}{2}\begin{pmatrix} 1 \\ i \end{pmatrix} \overline{a(\lambda)} e^{i\alpha(\lambda)x - x_0)} - \frac{1}{2}\begin{pmatrix} 1 \\ -i \end{pmatrix} b(\lambda) e^{-i\alpha(\lambda)(x + x_0)}$$

$$- \frac{1}{2}\begin{pmatrix} 1 \\ i \end{pmatrix} \overline{b(\lambda)} e^{i\alpha(\lambda)(x + x_0)} + \frac{1}{2}\begin{pmatrix} 1 \\ -i \end{pmatrix} a(\lambda) e^{-i\alpha(\lambda)(x - x_0)} \quad \text{near} \quad x = x_0 + L,$$

$$\phi_-(x,x_0,\lambda) \sim \frac{-i}{2}\begin{pmatrix} 1 \\ i \end{pmatrix} \overline{a(\lambda)} e^{i\alpha(\lambda)(x - x_0)} + \frac{i}{2}\begin{pmatrix} 1 \\ -i \end{pmatrix} b(\lambda) e^{-i\alpha(\lambda)(x + x_0)}$$

$$+ \frac{-i}{2}\begin{pmatrix} 1 \\ i \end{pmatrix} \overline{b(\lambda)} e^{i\alpha(\lambda)(x + x_0)} + \frac{i}{2}\begin{pmatrix} 1 \\ -i \end{pmatrix} a(\lambda) e^{-i\alpha(\lambda)(x - x_0)} \quad \text{near} \quad x = x_0 + L,$$

$$\phi_-(x,x_0,\lambda)\sim \frac{-i}{2}\begin{pmatrix}1\\ i\end{pmatrix}\overline{a(\lambda)}e^{i\alpha(\lambda)(x-x_0)}+\frac{i}{2}\begin{pmatrix}1\\ -i\end{pmatrix}b(\lambda)e^{-i\alpha(\lambda)(x+x_0)}$$
$$+\frac{-i}{2}\begin{pmatrix}1\\ i\end{pmatrix}\overline{b(\lambda)}e^{i\alpha(\lambda)(x+x_0)}+\frac{i}{2}\begin{pmatrix}1\\ -i\end{pmatrix}a(\lambda)e^{-i\alpha(\lambda)(x-x_0)} \qquad \text{near } x=x_0+L,$$

from which we find

$$\Delta(\lambda)=\phi_{+,1}(x_0+L,x_0,\lambda)+\phi_{-,2}(x_0+L,x_0,\lambda)=a(\lambda)e^{-i\alpha(\lambda)L}+\overline{a(\lambda)}e^{i\alpha(\lambda)L}. \tag{A8}$$

Now, *for λ real*, Eq. (A7) yields $\overline{a(\lambda)}=a^*(\lambda)$. Thus we write $a(\lambda)=|a(\lambda)|e^{i\mathrm{pha}(\lambda)}$, $a^*(\lambda)=|a(\lambda)|e^{-i\mathrm{pha}(\lambda)}$, and (A8) becomes

$$\Delta(\lambda)=2|a(\lambda)|\cos[\alpha(\lambda)L-\mathrm{pha}(\lambda)],\quad \lambda \text{ real}. \tag{A9}$$

APPENDIX B: DERIVATION OF THE DECOMPOSITION FORMULAS

In Ref. 25 the Θ function we have defined in the paper is denoted θ_4, and there are similarly defined Θ functions $\theta_1,\theta_2,\theta_3$. In fact, in terms of θ_i, the $N=1$ sine-Gordon Θ-function representation becomes

$$u(x,t)=2i\ln\left(\frac{\Theta(l(x,t)+\frac{1}{2};B)}{\Theta(l(x,t);B)}\right)$$
$$=2i\ln\left(\frac{\theta_4(l(x,t);B)}{\theta_3(l(x,t);B)}\right). \tag{B1}$$

We then show the ratio $\theta_4(l;B)/\theta_3(l;B)$ is given by

$$\frac{\theta_4(l;B)}{\theta_3(l;B)}=\frac{i\theta_2(l+B/2+\frac{1}{2};B)}{\theta_2(l+B/2;B)};$$

then we use the infinite-product representation of θ_2 in the form

$$\theta_2(l;B)=(\text{constant})e^{i\pi l}\prod_{n=1}^{\infty}(1+e^{2\pi i(nB-l)})$$
$$\times\prod_{-\infty}^{0}(1+e^{-2\pi i(nB-l)}).$$

Combining these facts with formula (V.15b) for $l(x,t)$ yields, after some manipulation,

$$\frac{\theta_4(l;B)}{\theta_3(l;B)}=\prod_{n=1}^{\infty}\frac{1-ie^{\alpha_n}}{1+ie^{\alpha_n}}\prod_{n=-\infty}^{0}(-1)\frac{1-ie^{\alpha_n}}{1+ie^{\alpha_n}},$$

where

$$\alpha_n=\kappa(x-x_0)+\omega t+2n\pi iB.$$

Using this infinite product in the formula (B1) for $u(x,t)$ yields Eq. (V.18) in the paper. □

APPENDIX C: PROOF: SPATIAL SYMMETRY ⇔ SPECTRAL SYMMETRY

The following proof is based on a whole-line argument by E. Overman. Consider any solution

$$\phi=\begin{pmatrix}\phi_1(x,E)\\ \phi_2(x,E)\end{pmatrix}$$

of the Takhatajian–Faddeev eigenvalue problem, (VI.4); then define the function $\psi(x,E)$ by

$$\psi(x,E)\equiv\begin{pmatrix}e^{(i/4)u(x)}\phi_1(x,E)\\ e^{(-i/4)u(x)}\phi_2(x,E)\end{pmatrix}. \tag{C1}$$

With $(\ )'=d/dx(\ )$, it easily follows that ψ satisfies the associated eigenvalue problem

$$\left(-\psi_2+\frac{i}{4}\mathring{v}(x)\psi_2\right)$$
$$+\left(\frac{1}{16\sqrt{E}}e^{(i/2)u(x)}-E^{1/2}e^{(-i/2)u(x)}\right)\psi_1=0, \tag{C2}$$

$$\left(\psi_1+\frac{i}{4}\mathring{v}(x)\psi_1\right)+\left(\frac{1}{16\sqrt{E}}e^{(-i/2)u(x)}-E^{1/2}e^{(i/2)u(x)}\right)\psi_2$$
$$=0.$$

Looking ahead to the spatial symmetry, we also define

$$\phi(x,E)\equiv\psi(-x,E). \tag{C3}$$

With this foundation, we now assume, as in Theorem VI.2, symmetry of the potentials:

$$(\mathring{u}(-x),\mathring{v}(-x))=(\mathring{u}(x),\mathring{v}(x)).$$

Inserting into (C2), we find

Lemma C 1: The following statements are equivalent.

(1) $\psi(x,E)$ solves (C2) at E.

(2) $\begin{pmatrix}\psi_2(-x,1/16^2E)\\ \psi_1(-x,1/16^2E)\end{pmatrix}=\begin{pmatrix}\phi_2(x,1/16^2E)\\ \phi_1(x,1/16^2E)\end{pmatrix}$

solves (C2) at E.

The symmetries in the spectrum are displayed with the Floquet discriminant $\Delta(E)$; recall the "Eigenfunction Representation of $\Delta(E)$," Eq. (II.12) (with $x_0\equiv 0$),

$$\Delta(E)=\phi_{+,1}(L,E)+\phi_{-,2}(L,E), \tag{C4}$$

where $\phi_\pm(x,E)$ are the basis for the Takhatajian–Faddeev eigenvalue problem (VI.4) normalized by

$$\phi_+(x=0,E)=\begin{pmatrix}1\\ 0\end{pmatrix},\quad \phi_-(x=0,E)=\begin{pmatrix}0\\ 1\end{pmatrix}. \tag{C5}$$

Thus, $\phi_\pm(x,E)$ are two particular cases of the function ϕ above, and following the relations (C1) and (C3) we define

$$\psi_\pm(x,E)\equiv\begin{pmatrix}e^{(i/4)u(x)} & \phi_{\pm,1}(x,E)\\ e^{(-i/4)u(x)} & \phi_{\pm,2}(x,E)\end{pmatrix}, \tag{C6}$$

$$\phi_\pm(x,E)=\psi_\pm(-x,E). \tag{C7}$$

We also note the initial conditions, which follow from (C5),

$$\psi_{\pm}(0,E)=\begin{pmatrix} e^{(i/4)\,\mathring{u}(x)} \\ 0 \end{pmatrix},\quad \psi_{-}(0,E)=\begin{pmatrix} 0 \\ e^{(-i/4)\,\mathring{u}(0)} \end{pmatrix}, \tag{C8a}$$

and the boundary condition

$$\psi_{\pm}(L,E)=\begin{pmatrix} e^{(i/4)\,\mathring{u}(L)} & \phi_{\pm,1}(L,E) \\ e^{(-i/4)\,\mathring{u}(L)} & \phi_{\pm,2}(L,E) \end{pmatrix}. \tag{C8b}$$

We aim to relate $\Delta(E)$ with $\Delta(1/16^2E)$, where

$$\Delta(E)=\phi_{+,1}(L,E)+\phi_{-,2}(L,E)$$
$$=\phi_{+,1}(-L,E)+\phi_{-,2}(-L,E),$$
$$\Delta(1/16^2E)=\phi_{+,1}(L,1/16^2E)+\phi_{-,2}(L,1/16^2E).$$

The approach is to map from $\boldsymbol{\phi}_{\pm}$ to $\psi_{\pm}$ using (C6), and then connect $\psi_{\pm}$ at E and $1/16^2E$ using Lemma C1.

From Lemma C1, we know

$$\psi_{\pm}(x,E)=\begin{pmatrix} \psi_{\pm,1}(x,E) \\ \psi_{\pm,2}(x,E) \end{pmatrix}$$

and

$$\begin{pmatrix} \psi_{\pm,2}(-x,1/16^2E) \\ \psi_{\pm,1}(-x,1/16^2E) \end{pmatrix}$$

are solutions of the *same* Eq. (C2); using the initial conditions (C8) at $x=0$,

$$\begin{pmatrix} \psi_{\pm,2}(0,1/16^2E) \\ \psi_{\pm,1}(0,1/16^2E) \end{pmatrix}=\begin{cases} e^{(i/4)u(0)}\begin{pmatrix}0\\1\end{pmatrix} \\ e^{(-i/4)u(0)}\begin{pmatrix}1\\0\end{pmatrix} \end{cases}$$

we find that the eigenfunctions are proportional:

$$\begin{pmatrix} \psi_{\pm,2}(-x,1/16^2E) \\ \psi_{\pm,1}(-x,1/16^2E) \end{pmatrix}=e^{\pm(i/2)u(0)}\psi_{\mp}(x,E). \tag{C9}$$

Thus, we compute

$$\Delta(1/16^2E)=\phi_{+,1}(L,1/16^2E)+\phi_{-,2}(L,1/16^2E)$$
$$=e^{(-i/4)u(L)}\psi_{+,1}(L,1/16^2E)+e^{(i/4)u(L)}\psi_{-,2}(L,1/16^2E)$$
$$=e^{(-i/4)u(L)}e^{(i/2)u(0)}\psi_{-,2}(-L,E)+e^{(i/4)u(L)}e^{(-i/2)u(0)}\psi_{+,1}(-L,E)$$
$$=e^{(i/2)[u(0)-u(L)]}\phi_{-,2}(-L,E)+e^{(i/2)[u(L)-u(0)]}\phi_{+,1}(-L,E).$$

From the periodicity of $\mathring{u}(x)$,

$$\mathring{u}(L)-\mathring{u}(0)=2\pi M,\quad M\equiv\text{"charge of }\mathring{u}(x)\text{,"}$$

we find

$$\Delta(1/16^2E)=(-1)^M\Delta(E).$$

The argument can be reversed in the following way. Note that Lemma C1 holds if and only if $(\mathring{u}(-x),\mathring{v}(-x))=(\mathring{u}(x),\mathring{v}(x))$. Equating $\Delta(E)$ and $(-1)^M\Delta(1/16^2E)$ implies $(\mathring{u}(-L),\mathring{v}(-L))=(\mathring{u}(L),\mathring{v}(L))$. The result then follows using periodicity of $\mathring{u}(x),\mathring{v}(x)$. □

APPENDIX D: ELLIPTIC INTEGRAL FORMULAS

The normalization constant C and period "matrix" B are both expressed in terms of two fundamental loop integrals, which we denote $I(a),I(b)$. With the branch cuts of Fig. 9, we denote canonical paths by "a" and "b" cycles as shown in Fig. 22, where we also depict useful contours of integra-

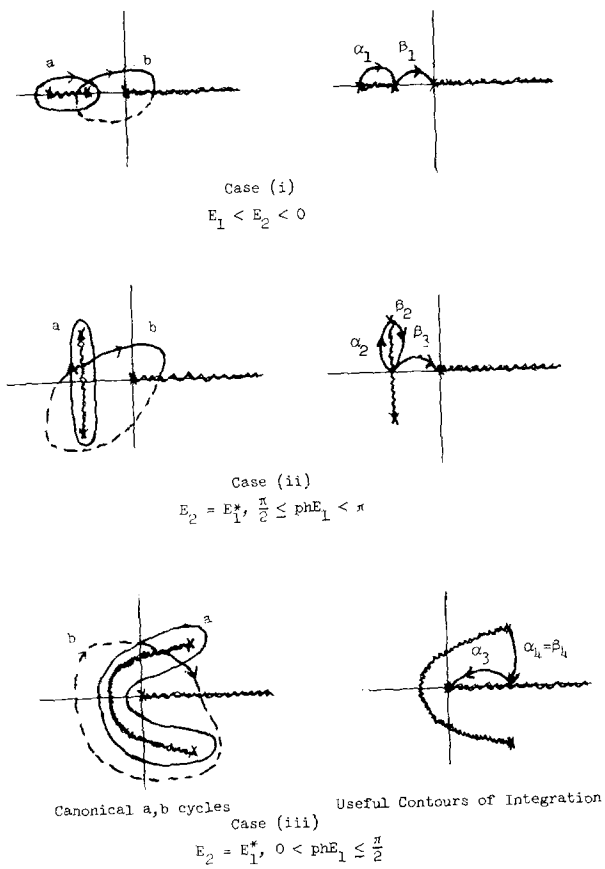


FIG. 22. Canonical a,b cycles and useful contours of integration for all $N=1$ cut structures.

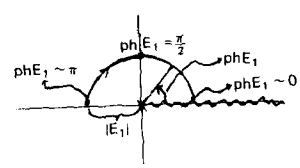


FIG. 23. Graphic limit as $\mathrm{ph}E_1:\pi\to 0$, $|E_1|$ constant.

tion for each structure. In terms of the holomorphic differential $dI \equiv dE/R(E), R^2(E) = E(E-E_1)(E-E_2)$, we define

$$I(a) = \oint_a dI, \quad I(b) = \oint_b dI.$$

Then the constants C and B are explicitly given by

$$C = 1/I(a), \quad B = I(b)/I(a).$$

We now display the detailed information in Tables I–IV.

TABLE I. General structure of the loop integrals $I(a), I(b)$, normalization constant C, and period matrix B.

In the notation of Fig. 22, $I(a), I(b)$ have the following representations in terms of the contours α_i, β_i; the explicit facts then follow by routine complex integration, with $C = 1/I(a), B = I(b)/I(a)$.

Case (i):
$$I(a) = 2\int_{\alpha_1} dI \;;\; I(b) = \int_{\beta_1} dI$$
$$\Rightarrow I(a) < 0, \quad \mathrm{Re}\, I(b) = 0, \ \mathrm{Im}\, I(b) < 0.$$
$$\Rightarrow C < 0, \quad B = i\,\mathrm{Im}\, B, \quad \mathrm{Im}\, B > 0.$$

Case (ii):
$$I(a) = 4\,\mathrm{Re}\int_{\alpha_2} dI \;;\; I(b) = -\tfrac{1}{2}I(a) + 2\int_{\beta_2} dI + 2i\,\mathrm{Im}\int_{\beta_2} dI$$
$$\Rightarrow I(a) < 0, \quad \mathrm{Re}\, I(b) = -\tfrac{1}{2}I(a), \ \mathrm{Im}\, I(b) < 0.$$
$$\Rightarrow C < 0, \quad B = -\tfrac{1}{2} + i\,\mathrm{Im}\, B, \ \mathrm{Im}\, B > 0.$$

Case (iii):
$$I(a) = 4\int_{\alpha_3} dI + 4\,\mathrm{Re}\int_{\alpha_4} dI \;;\; I(b) = -\tfrac{1}{2}I(a) + 2i\,\mathrm{Im}\int_{\beta_4} dI$$
$$\Rightarrow I(a) < 0, \quad \mathrm{Re}\, I(b) = -\tfrac{1}{2}I(a), \ \mathrm{Im}\, I(b) < 0.$$
$$\Rightarrow C < 0, \quad B = -\tfrac{1}{2} + i\,\mathrm{Im}\, B, \ \mathrm{Im}\, B > 0.$$

TABLE II. Elliptic integral formulas for the normalization constant C and period matrix B.

The loop integrals $I(a), I(b)$, and thus the normalization constant C and period matrix B, can be expressed in terms of familiar elliptic integrals.[35] In detail (refer to Fig. 22 and Table I).

Case (i):
$E_1 < E_2 < 0$,
$C = -(-E_1)^{1/2}/4K(s)$,
$B = iK'(s)/K(s)$,
where the modulus $s^2 = (E_2 - E_1)/-E_1$.

Case (ii):
$E_1 = E_2^*, E_1 \neq E_2, \pi/2 < \mathrm{ph}E_1 < \pi$,
$C = -|E_1|^{1/2}/4K(s)$,
$B = -\tfrac{1}{2} + [i/2K(s)][F(\phi,s') + F(\psi,s')]$,
where $s^2 = \tfrac{1}{2}(1 + \cos(\mathrm{ph}E_1)), s'^2 = \tfrac{1}{2}(1 - \cos(\mathrm{ph}E_1))$,
$\phi = \arccos(s^2/s'^2), \psi = \arcsin(s/s')$.

Case (iii):
$E_1 = E_2^*, E_1 \neq E_2, 0 < \mathrm{ph}E_1 < \pi/2$,
$C = -|E_1|^{1/2}/4[F(\phi,s) + F(\psi,s)]$,
$B = -\tfrac{1}{2} + iK(s')/[2(F(\phi,s) + F(\psi,s))]$,
where s, s', ϕ, ψ are as defined in Case (ii).

TABLE III. Soliton limit ($E_1 = E_2 < 0$) formulas.

From the explicit elliptic integral formulas in Table II, the "soliton limit" as E_1, E_2 collide on the negative real axis is computed. The results are the same for each cut structure, $E_1 < E_2 < 0$ and $E_1 = E_2^*, E_1 \neq E_2$. (Refer to Figs. 15 and 22.)

$$\lim_{E_1 \to E_2} I(b) = 2\int_{E_1}^{0} \frac{dE}{(E-E_1)\sqrt{E}} = -i\infty,$$

$$\lim_{E_1 \to E_2} I(a) = \int_{|E-E_1|=\epsilon} \frac{dE}{(E-E_1)\sqrt{E}} = -2\pi i \quad [\text{residue at } E_1] = \frac{-2\pi}{\sqrt{|E_1|}},$$

$$\lim_{E_1 \to E_2} L = +\infty,$$

$$\lim_{E_1 \to E_2} \kappa(x - x_0) + \omega t = \frac{1}{(1-v^2)^{1/2}}[x - x_0 - vt\,], \quad v = \frac{1 - 16|E_1|}{1 + 16|E_1|}$$

(for the choice $|U| < 1$).

TABLE IV. Contrasts in physical characteristics of the oscillatory states in terms of E_1, E_1^*.

To display the difference in the (subcharacteristic speed, $U^2 < 1$) *oscillatory states* based on the relative location of $E_1 = E_2^*, E_1 \neq E_2$, we consider two limits that essentially "cover" the E-plane. First, we display the "angular dependence" of the physical characteristics, and second the "radial dependence."

1. Fix $|E_1| \equiv$const, and consider $\mathrm{ph}E_1$: $\pi \to 0$ (see Fig. 23)

E = Energy: $+1 \to -1$.

u_0 = Amplitude of oscillation = $\mathrm{ph}E_1$: $\pi \to 0$.

$$\kappa = \text{Wavenumber} = \frac{2\pi|C|}{\mathrm{Im}(B)}\left(1 + \frac{1}{16(E_1E_2)^{1/2}}\right):\left(\frac{\sqrt{|E_1|}}{4} + \frac{4}{\sqrt{|E_1|}}\right)\to 0.$$

$$\omega = \text{Frequency} = \frac{2\pi|C|}{\mathrm{Im}(B)}\left(1 - \frac{1}{16(E_1E_2)^{1/2}}\right):\left(\frac{-\sqrt{|E_1|}}{4} + \frac{4}{\sqrt{|E_1|}}\right)\to 0.$$

$\mathrm{Im}B \propto$ separation distance between crests: $+\infty \to 0$.

$|C| \propto$ width of each kink (antikink) component: $\sqrt{|E_1|} \to 0$.

We note that each of the above limits is *monotone decreasing*.

2. Fix $\mathrm{ph}E_1 \equiv$constant, and consider $|E_1|$: $0 \to \infty$

E = Energy = $-\cos(\mathrm{ph}E_1)$ remains constant, independent of $|E_1|$.

u_0 = Amplitude of oscillation = $\mathrm{ph}E_1$ remains constant, independent of $|E_1|$.

κ = Wavenumber: $+\infty$ *decreasing to* $\kappa_{\min}|_{|E_1|=16} = 1$, *increasing to* $+\infty$.

ω = Frequency: $+\infty$ *decreasing to* $-\infty, \omega|_{|E_1|=16} = 0$.

$\mathrm{Im}B \propto$ Separation distance between crests *remains constant*.

$|C| \propto$ Width of each kink (antikink) component: $0 \to +\infty$.

[1] G. Costabile, R. D. Parmentier, B. Savo, D. W. McLaughlin, and A. C. Scott, Appl. Phys. Lett. **32**, 587 (1978).

[2] A. Bishop, private communication.

[3] B. L. Holian and G. K. Straub, "Molecular Dynamics of Shock Waves in One-Dimensional Chains," Phys. Rev. B **18**, 1593 (1978).

[4] H. Flaschka, M. G. Forest, and D. W. McLaughlin, "Multiphase Averaging and the Inverse Spectral Solution of K.dV.," Commun. Pure Appl. Math. **33**, 739 (1980).

[5] L. A. Takhatajian and L. D. Faddeev, "Essentially Nonlinear One-Dimensional Model of Classical Field Theory," Theor. Math. Phys. **21**, 1046 (1974).

[6] D. W. McLaughlin and A. C. Scott, "Perturbation Analysis of Fluxon Dynamics," Phys. Rev. A **18**, 1652 (1978).

[7] C. S. Gardner, J. M. Greene, M. P. Kruskal, and R. M. Miura, "Korteweg-de Vries Equation and Generalizations, 6. Methods for Exact Solution," Commun. Pure Appl. Math. **27**, 97 (1974).

[8] G. L. Lamb, Jr., *Elements of Soliton Theory* (Wiley, New York, 1979).

[9] M. J. Ablowitz, D. Kaup, and A. Newell, J. Math. Phys. **15**, 1852 (1974).

[10] Notice (II.5b) means $u(x+L,t) = u(x,t) (\mathrm{mod} 2\pi)$ if u is real, and $u(x+L,t) = u(x,t)$ if u is purely imaginary.

[11] Actually, the spectrum is the closure of this set.

[12] $(\,)^*$ refers to complex conjugation here and in the remainder of the paper.

[13] As will become evident in the next section, the appropriate mathematical function theory takes place on a Riemann surface built from two copies of the E plane, with branch cuts along the continuous spectrum of the Takhatajian–Faddeev eigenvalue problem. In whole-line scattering theory, this spectrum consists of the positive real E axis, and thus the branch cut structure coincides with that of the function $\lambda^2 = E$. However, for periodic problems the continuous spectrum is more complicated; it consists of bands of spectrum with endpoints $\{E_j\}$. Thus the branch cuts for the periodic problem coincide with that of the function $\lambda^2 = E\,\Pi_j(E - E_j)$. Therefore, for whole-line scattering theory, $\lambda = \sqrt{E}$ is a natural coordinate (except near ∞); however, for periodic problems it is not a natural local coordinate (except near 0). Nevertheless, at times we will work with the parameter λ since we aim to use facts from whole-line scattering theory.

[14] The proof follows almost verbatim the arguments for the steepest descent method of asymptotics [Ref. 15].

[15] E. T. Copson, *Asymptotic Expansions* (Cambridge University, London, 1965), pp. 65–68.

[16] We emphasize that this proof is an analyticity argument, independent of the compact support restrictions inherent in the scattering representation of Δ. The absence of gaps in the spectrum on the positive real E axis and the existence of spines are *not* restricted to compact support potentials.

[17] At present, a precise definition of "most dominant" bands is unavailable. Nonetheless, this is the nonlinear analog of approximating a linear wave by the N most dominant Fourier modes.

[18] The branch points at $E = 0, \infty$ arise not as zeros of $\Delta^2 - 4$, but due to the essential singularities of $\Delta(E)$ at $E = 0, \infty$.

[19] This is completely analogous to whole-line scattering theory: The $L^2(\mathbb{R})$ Green's function G is an analytic function of the energy parameter E except for essential singularities at $E = 0, \infty$ and poles at the bound state eigenvalues; $G(E)$ has a branch cut which can be chosen along the positive real axis to coincide with the continuous spectrum. For the periodic problem the whole-line Green's function is an analytic function of E, except for essential singularities at $E = 0, \infty$ and branch points at the periodic and antiperiodic eigenvalues; the branch cut structure can be chosen to coincide with the spectrum σ.

[20] V. B. Matveev, "Abelian Functions and Solitons," Preprint No. 373, University of Wroclaw (1973).

[21] E. Date and S. Tanaka, "Periodic Multi-Soliton Solutions of the Korteweg-de Vries Equation and Toda Lattice," Prog. Theor. Phys. Supp. **59**, 107 (1976).

[22] G. Springer, *Introduction to Riemann Surfaces* (Addison–Wesley, Reading, Mass., 1957), Chap. 10.

[23] M. Toda, "Studies of an Exponential Lattice," Phys. Rep. **18**, 1 (1975).

[24] For this single-phase case, $\mathrm{Re}[I(0,0)]$ is chosen to yield $u(x,t)$ real, while $\mathrm{Im}[I(0,0)]$ centers the wave train. The choices $(-\frac{1}{4}$ or $\frac{1}{4})$ are shown to yield kinks or antikinks, respectively.

[25] E. T. Whittaker G. N. Watson, *A Course of Modern Analysis* (Cambridge University, London, 1902).

[26] From these representations (V.18), $\mathrm{Im}(\alpha_n) \equiv k, k \in Z$, implies $|1 \mp ie^{\alpha_n}| = |1 \pm ie^{\alpha_n}|$, and the wave is indeed *real*.

[27] H. P. McKean, "The sine-Gordon and sinh-Gordon Equations on the Circle," Com. Pure Appl. Math. **34**, 197 (1981).

[28] Private communication.

[29]G. L. Lamb, Jr., "Analytical Descriptions of Ultrashort Optical Pulse Propagation in a Resonant Medium," Rev. Mod. Phys. **43**, 99 (1971).
[30]T. A. Fulton, in *Superconductor Applications: SQUIDS and Machines*, edited by B. B. Schwartz and S. Foner (Plenum, New York, 1977), p. 125.
[31]S. Trullinger, private communication.
[32]A. Osborne and A. E. G. Stuart, "On the Separability of the sine-Gordon Equation and Similar Quasilinear Partial Differential Equations," J. Math. Phys. **19**, 1573 (1978).
[33]$U = \infty \Rightarrow u(x,t)$ is independent of $x \Rightarrow \hat{u}(x) \equiv \text{const} \Rightarrow \Sigma^{(s)}$ is empty.
[34]V. E. Zakharov, L. A. Takhatajian, and L. D. Faddeev, "Complete Description of Solutions of the 'sine-Gordon' Equations," Sov. Phys. Dokl. **19**, 824 (1975).
[35]P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Scientists* (Springer, New York, 1971).

New sum rule for products of Bessel functions with application to plasma physics

Barry S. Newberger
Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 2 February 1982; accepted for publication 5 March 1982)

In our investigations of the linear theory of the stability of relativistic beam-plasma systems immersed in a magnetic field we have been led to consider sum rules for an infinite series of products of Bessel functions of the form $\Sigma_{n=-\infty}^{\infty}(n^j J_n^2)/(n+\mu)$. In this work we report on the sum of this series treated as a special case of a more general infinite series. We also mention the extension of the results beyond the range of the parameters for which formulae are explicitly given and indicate how intermediate results obtained may be useful in their own right. Finally, an additional application of our result is indicated.

PACS numbers: 02.30.Gp

I. INTRODUCTION

In considering the linear theory of the stability of a charged particle beam penetrating a magnetized plasma, one is confronted with an infinite series of products of Bessel functions of the form

$$\sum_{n=-\infty}^{\infty}\frac{n^j J_n^2(z)}{n+\mu},$$

or similar series which can be obtained from it by differentiation with respect to z. In the context of the physical problem, μ is essentially the ratio of the Doppler shifted wave frequency as seen by an individual particle in its orbit to its Larmor frequency. The variable z is essentially the product of the wavenumber of the mode perpendicular to the magnetic field and the Larmor radius of an individual particle. The dispersion relation is obtained by an integration over a distribution of particles and, hence, a numerical investigation of the dispersion relation is complicated considerably by the presence of the infinite sums. This is especially true when one observes that the rates of convergence may vary widely as the parameters vary over the physically interesting domain. Furthermore, one would like to obtain analytic results in limiting cases, particularly those of large and small magnetic field. The latter case corresponds to the asymptotic regime in μ and progress can only be made if a sum rule can be found. Results from the application of our sum rules to the plasma physics problem have been presented.[1]

As will be seen in the work to follow, a sum rule can be obtained for the more general case

$$\sum_{n=-\infty}^{\infty}\frac{(-1)^n n^j J_{\alpha-\gamma n}J_{\beta+\gamma n}}{n+\mu},$$

from which our required result follows as a special case ($\alpha=\beta=0, \gamma=1$). Here, as above, j is a nonnegative integer and, in our specific application, runs from 0 to 2 but will not be so restricted in the results to follow. A special case ($\alpha=\beta$, $j=0$) of our result has been given.[2]

In the following section, we will sketch the derivation of the desired sum rule. Our method is direct calculation. In Sec. III, we will discuss extension of the results to parameter ranges beyond those for which formulae have been given. Furthermore, given the result for $j=0$, it is easily shown that it follows from a contour integration. This is also briefly discussed in this section. On the other hand, when $j\neq 0$, the contour integration method has difficulties in that one cannot ensure that the integrand dies off sufficiently rapidly as the contour is enlarged to ∞. The direct method illustrates the difficulty explicitly and shows how it can be circumvented. The applicability of the method to other series is also discussed here and one example is given. In the last section, applications will be briefly addressed. Included will be explicit results for our own special case, an infinite series which has appeared in the solution of a particular differential equation arising in an electrical engineering problem, which has come to be called "Turkin's" function,[3] and a brief discussion of the application of certain "intermediate" results.

II. DERIVATION OF THE SUM RULE

In this section we present the derivation of the sum rule for the infinite series of the form

$$S=\sum_{n=-\infty}^{\infty}\frac{(-1)^n n^j J_{\alpha-\gamma n}(z)J_{\beta+\gamma n}(z)}{n+\mu}, \tag{2.1}$$

where j is a nonnegative integer, μ is nonintegral but otherwise complex and unrestricted, and z is a complex parameter as are α and β. We restrict them, for the moment, to satisfy $\mathrm{Re}(\alpha+\beta)>-1$. The parameter γ is real and restricted to the interval (0,1].

It is useful to rewrite S in the following way:

$$S=\sum_{n=-\infty}^{\infty}\frac{(-1)^n[n^j-(-\mu)^j]}{n+\mu}J_{\alpha-\gamma n}J_{\beta+\gamma n}+(-\mu)^j\sum_{n=-\infty}^{\infty}\frac{(-1)^n J_{\alpha-\gamma n}J_{\beta+\gamma n}}{n+\mu}. \tag{2.2}$$

We will suppress the variable z throughout this section. We now consider the two sums separately:

$$S_1\equiv\sum_{n=-\infty}^{\infty}\frac{(-1)^n J_{\alpha-\gamma n}J_{\beta+\gamma n}}{n+\mu} \tag{2.3a}$$

and

$$S_2\equiv\sum_{n=-\infty}^{\infty}\frac{(-1)^n[n^j-(-\mu)^j]J_{\alpha-\gamma n}J_{\beta+\gamma n}}{n+\mu}=\sum_{n=-\infty}^{\infty}(-1)^n\sum_{k=0}^{j-1}(-1)^k\mu^k n^{j-1-k}J_{\alpha-\gamma n}J_{\beta+\gamma n}. \tag{2.3b}$$

 0022-2488/82/071278-04$02.50

We will consider S_1 first.

$$S_1 = \sum_{n=-\infty}^{\infty} \frac{(-1)^n J_{\alpha-\gamma n} J_{\beta+\gamma n}}{n+\mu} = \frac{2}{\pi} \sum_{n=-\infty}^{\infty} \frac{(-1)^n}{n+\mu} \int_0^{\pi/2} J_{\alpha+\beta}(2z\cos\theta) \times \cos[(2\gamma n - \eta)\theta]\, d\theta,$$

where

$$\eta \equiv \alpha - \beta, \text{ and } \mathrm{Re}(\alpha+\beta) > -1. \tag{2.4}$$

The integral representation of the product of Bessel functions is well known.[4]

Expanding $\cos[(2\gamma n - \eta)\theta]$ and separating the summation into a pair of sums over odd and even values of n, respectively, gives

$$S_1 = \frac{2}{\pi} \int_0^{\pi/2} d\theta\, J_{\alpha+\beta}(2z\cos\theta) \times \left\{ \frac{\cos\eta\theta}{2} \sum_{n=-\infty}^{\infty} \left[-\frac{\cos[(4n+2)\gamma\theta}{n+\mu/2+1/2} + \frac{\cos 4n\gamma\theta}{n+\mu/2} \right] + \frac{\sin\eta\theta}{2} \sum_{n=-\infty}^{\infty} \left[-\frac{\sin[(4n+2)\gamma\theta]}{n+\mu/2+1/2} + \frac{\sin 4n\gamma\theta}{n+\mu/2} \right] \right\}, \tag{2.5}$$

there being no difficulty with the interchange of orders of summation and integration.

Further expanding the trigonometric functions, we find we must evaluate sums of the form

$$\sum_{n=-\infty}^{\infty} \frac{\begin{Bmatrix}\cos 4n\gamma\theta \\ \sin 4n\gamma\theta\end{Bmatrix}}{n+\mu/2+1/2} \text{ and } \sum_{n=-\infty}^{\infty} \frac{\begin{Bmatrix}\cos 4n\gamma\theta \\ \sin 4n\gamma\theta\end{Bmatrix}}{n+\mu/2}.$$

Using the formulae[5]

$$\sum_{n=-\infty}^{\infty} \frac{\cos n\phi}{n-a} = -\frac{\pi\cos a(\pi-\phi)}{\sin\pi a}$$

and

$$\sum_{n=-\infty}^{\infty} \frac{\sin n\phi}{n-a} = \frac{\pi\sin a(\pi-\phi)}{\sin\pi a}, \quad 0<\phi<2\pi,$$

where a is nonintegral but otherwise unrestricted, we have for S_1,

$$S_1 = \int_0^{\pi/2} d\theta\, J_{\alpha+\beta}(2z\cos\theta) \times \left\{ \cos\eta\theta \left[-\frac{\cos[\pi/2+\mu/2(\pi-4\gamma\theta)]}{\sin[(\mu/2+1/2)\pi]} + \frac{\cos[\mu/2(\pi-4\gamma\theta)]}{\sin\mu\pi/2} \right] + \sin\eta\theta \left[-\frac{\sin[\pi/2+\mu/2(\pi-4\gamma\theta)]}{\sin[(\mu/2+1/2)\pi]} + \frac{\sin[\mu/2(\pi-4\gamma\theta)]}{\sin\mu\pi/2} \right] \right\} \tag{2.6a}$$

or

$$S_1 = \frac{1}{\sin(\mu\pi/2)\cos(\mu\pi/2)} \int_0^{\pi/2} d\theta\, J_{\alpha+\beta}(2z\cos\theta) \times (\cos\eta\theta\,\{\sin(\mu\pi/2)\sin[\mu/2(\pi-4\gamma\theta)] + \cos(\mu\pi/2)\cos[\mu/2(\pi-4\gamma\theta)]\} + \sin\eta\theta\,\{-\sin(\mu\pi/2)\cos[\mu/2(\pi-4\gamma\theta)] + \cos(\mu\pi/2)\sin[\mu/2(\pi-4\gamma\theta)]\}). \tag{2.6b}$$

Simplifying the trigonometric functions reduces S_1 to

$$\frac{1}{\sin(\mu\pi/2)\cos(\mu\pi/2)} \int_0^{\pi/2} d\theta\, J_{\alpha+\beta}(2z\cos\theta) \times \cos[(\eta+2\gamma\mu)\theta], \tag{2.7}$$

or, upon employing the integral representation for the product of Bessel functions once again,

$$S_1 = \frac{\pi}{\sin\mu\pi} J_{\alpha+\gamma\mu}(z) J_{\beta-\gamma\mu}(z), \tag{2.8}$$

which is the desired result.

We now proceed to consider S_2, Eq. (2.3b). Interchanging orders of summation, it is sufficient to consider a sum of the form

$$\hat{S}_2 = \sum_{n=-\infty}^{\infty} (-1)^n n^p J_{\alpha-\gamma n} J_{\beta+\gamma n}, \quad p \text{ integer}, = \frac{2}{\pi} \sum_{n=-\infty}^{\infty} (-1)^n \int_0^{\pi/2} d\theta\, J_{\alpha+\beta}(2z\cos\theta) n^p \times \cos[(2\gamma n - \eta)\theta]. \tag{2.9a}$$

Again expanding $\cos[(2\gamma n - \eta)\theta]$ and observing that we can absorb the factor n^p by differentiation with respect to $2\gamma\theta$ p times, we have

$$\hat{S}_2 = \frac{2(-1)^{p^*}}{\pi} \sum_{n=-\infty}^{\infty} (-1)^n \int_0^{\pi/2} d\theta\, J_{\alpha+\beta}(2z\cos\theta) \times \left[\cos\eta\theta \frac{d^p}{d(2\gamma\theta)^p} \begin{Bmatrix}\cos 2n\gamma\theta \\ \sin 2n\gamma\theta\end{Bmatrix} + \sin\eta\theta \frac{d^p}{d(2\gamma\theta)^p} \begin{Bmatrix}\sin 2n\gamma\theta \\ \cos 2n\gamma\theta\end{Bmatrix} \right], \tag{2.10}$$

where we define

$$p^* = p/2, \quad p \text{ even} = (p-1)/2, \quad p \text{ odd},$$

and the upper term in braces pertains if p is even and conversely if p is odd.

One is now tempted to interchange orders of summation and integration but is then confronted with infinite series which do not converge in the ordinary sense. Progress can be made by appealing to the theory of generalized functions.[6,7] This allows us to proceed formally.

Interchanging orders of summation, integration, and differentiation, we first observe that the sine series vanishes. Thus in Eq. (2.10), the first term in square brackets pertains if p is even and the second pertains if p is odd. We will denote this in what follows by a subscript e or o as appropriate in the pth order derivative. Separating sums over even and odd values of n as before, we have

$$\hat{S}_2 = \frac{2(-1)^{p^*}}{\pi}\int_0^{\pi/2} d\theta\, J_{\alpha+\beta}(2z\cos\theta)$$
$$\times\left\{\left[\cos\eta\theta\frac{d^p}{d(2\gamma\theta)_e^p} + \sin\eta\theta\frac{d^p}{d(2\gamma\theta)_o^p}\right]\right.$$
$$\left.\times\left[-\sum_{n=-\infty}^{\infty}\cos[(2n+1)(2\gamma\theta)] + \sum_{n=-\infty}^{\infty}\cos 4\eta n\theta\right]\right\}. \tag{2.11}$$

Expanding the trigonometric functions and again observing that the sine series vanishes, we are left with

$$\hat{S}_2 = \frac{2^p(-1)^{p^*}}{2\pi\gamma}\int_0^{2\pi\gamma} dx\, J_{\alpha+\beta}\left(2z\cos\left(\frac{x}{4\gamma}\right)\right)$$
$$\times\left\{\left[\cos\left(\frac{\eta x}{4\gamma}\right)\frac{d^p}{dx_e^p} + \sin\left(\frac{\eta x}{4\gamma}\right)\frac{d^p}{dx_o^p}\right]\right.$$
$$\left.\times\left[\left(1-\cos\frac{x}{2}\right)\sum_{n=-\infty}^{\infty}\cos nx\right]\right\}, \tag{2.12}$$

where $x \equiv 4\gamma\theta$.

We now assume that $\mathrm{Re}(\alpha+\beta) \geqslant p$. This certainly ensures that the pth derivative of $J_{\alpha+\beta}$ is integrable. We will return to this assumption in the next section.

Since the integrand is an even function of x, we will integrate over the full interval $[-2\pi\gamma, 2\pi\gamma]$ and take half the value obtained. Integrating by parts p times and observing that, under the assumption that $\mathrm{Re}(\alpha+\beta) \geqslant p$, the integrated terms vanish, we have

$$\hat{S}_2 = \frac{2^p(-1)^{p^*}(-1)^p}{4\pi\gamma}$$
$$\times\int_{-2\pi\gamma}^{2\pi\gamma}\frac{d^p}{dx^p}\left\{J_{\alpha+\beta}\left(2z\cos\left(\frac{x}{4\gamma}\right)\right)\begin{bmatrix}\cos(\eta x/4\gamma)\\ \sin(\eta x/4\gamma)\end{bmatrix}\right\}$$
$$\times\left(1-\cos\left(\frac{x}{2}\right)\right)\sum_{n=-\infty}^{\infty}\cos nx\, dx, \tag{2.13}$$

where the upper term in [ ] pertains if p is even and the lower if p is odd. Now using a result in the theory of generalized functions,[6,7] we put

$$\sum_{n=-\infty}^{\infty}\cos nx = 2\pi\sum_{n=-\infty}^{\infty}\delta(x-2n\pi),$$

and $\hat{S}_2$ becomes

$$\hat{S}_2 = \frac{2^p(-1)^{p^*}(-1)^p}{2\gamma}$$
$$\times\int_{-2\pi\gamma}^{2\pi\gamma} dx\,\frac{d^p}{dx^p}\left\{J_{\alpha+\beta}\left(2z\cos\left(\frac{x}{4\gamma}\right)\right)\begin{bmatrix}\cos(\eta x/4\gamma)\\ \sin(\eta x/4\gamma)\end{bmatrix}\right\}$$
$$\times\left(1-\cos\left(\frac{x}{2}\right)\right)\sum_{n=-\infty}^{\infty}\delta(x-2n\pi)\, dx. \tag{2.14}$$

The result for $\hat{S}_2$ now follows immediately. If $\gamma < 1$, only the $n=0$ term in the sum contributes. If $\gamma = 1$, both the $n=0$ and $n=1$ terms contribute. Furthermore, in the case $\gamma = 1$, the δ-functions act at the end points of the integration and, hence, only contribute one half the value they otherwise would. We now define

$$H(x) \equiv 0, \quad x < 0,$$
$$\equiv \tfrac{1}{2}, \quad x = 0,$$
$$\equiv 1, \quad x > 0. \tag{2.15}$$

We also observe that by virtue of the $[1-\cos(x/2)]$ term, the contribution from $n=0$ vanishes. The extension to γ on the interval (1,2) is also immediate. We will discuss this again briefly in the following section. Thus we have, evaluating $[1-\cos(x/2)]$ at $x = 2\pi$,

$$\hat{S}_2 = H(\gamma-1)\frac{2^p(-1)^{p^*}(-1)^p}{\gamma}$$
$$\times\frac{d^p}{dx^p}\left\{J_{\alpha+\beta}\left(2z\cos\left(\frac{x}{4\gamma}\right)\right)\begin{bmatrix}\cos(\eta x/4\gamma)\\ \sin(\eta x/4\gamma)\end{bmatrix}\right\}\Bigg|_{x=2\pi}. \tag{2.16}$$

Substituting into (2.2), we arrive at our final expression for S:

$$S = \frac{\pi(-\mu)^j J_{\alpha+\gamma\mu}(z)J_{\beta-\gamma\mu}(z)}{\sin\pi\mu}$$
$$-\frac{H(\gamma-1)}{\gamma}(-2)^j\sum_{k=0}^{j-1}{}_{\mathrm{o}}^{\mathrm{e}}\left(\frac{\mu}{2}\right)^k\begin{bmatrix}(-1)^{(j-1-k)/2}\\ (-1)^{(j-2-k)/2}\end{bmatrix}$$
$$\times\frac{d^{j-1-k}}{dx^{j-1-k}}\left\{J_{\alpha+\beta}\left(2z\cos\left(\frac{x}{4\gamma}\right)\right)\begin{bmatrix}\cos(\eta x/4\gamma)\\ \sin(\eta x/4\gamma)\end{bmatrix}\right\}\Bigg|_{x=2\pi}, \tag{2.17}$$

where again the upper term in [ ] pertains if $j-1-k$ is even and the lower if $j-1-k$ is odd. We have appended the superscript e and subscript o to the summation symbol as a reminder.

III. EXTENSION OF THE RESULTS

In this section we will discuss the validity of our results for values of the parameters outside their domains of restriction imposed in Sec. II. It is convenient to consider S_1 and S_2 separately.

We have shown that S_1 as defined in Eq. (2.3a) is, as in Eq. (2.8), equal to

$$\frac{\pi J_{\alpha-\gamma\mu}J_{\beta+\gamma\mu}}{\sin\pi\mu} \quad \text{for } 0 \leqslant \gamma \leqslant 1,\ \mathrm{Re}(\alpha+\beta) > -1,$$

and μ complex but nonintegral. Thus, S_1 considered as a function of α, β, or γ is analytic since the Bessel functions are entire functions of their order. It follows by analytic continuation that the restriction on the real part of α and β is lifted and that, as far as S_1 is concerned, γ can be any complex number. It also follows from Eq. (2.8) that S_1 is a meromorphic function of μ, with poles at the integers. It is interesting to observe that if we take Eq. (2.8) to be the definition of S_1 then the infinite series Eq. (2.3a) follows immediately from the expansion formula for meromorphic functions,[8,9] provided $\mathrm{Re}(\alpha+\beta) \geqslant -1$, since there is a closed contour C_N on which $1/\sin\pi\mu$ is bounded as the length of the contour $L \to \infty$, and one can easily show that $|J_{\alpha-\gamma\mu}J_{\beta+\gamma\mu}| = O(|\mu|^{-N-1})$ for fixed N such that $\mathrm{Re}(\alpha+\beta) \geqslant N$.

The situation with S_2 is slightly more complicated. Our results have been derived under the assumption $\mathrm{Re}(\alpha+\beta) \geqslant p$. We observe, however, that S_2 from its definition Eq. (2.9a) is an entire function of α and β. This is also

true of the expression (2.15). Thus, by analytic continuation the result holds provided only that the required derivatives of the Bessel functions exist at zero. Thus, the condition $\mathrm{Re}(\alpha+\beta)\geqslant p$ is sufficient but not necessary. An example is $\alpha+\beta=0$. Our application to a problem in plasma physics is a special case of this with $\alpha=\beta=0$ and results are illustrated in the following section.

To extend the range of γ beyond 1, care must be taken to include the additional terms in the sum over δ-functions. That is, S_2 is discontinuous at the integers; for example on the interval (1,2), the term in Eq. (2.16) which appears for $\gamma=1$ is multiplied by 2 and the function H we introduced earlier achieves this extension automatically. In principle one could continue along the γ axis in this way and thereby extend the results to arbitrary real γ. We note too that since γ appears symmetrically, our results embrace negative real γ and we need not concern ourselves with such values of γ specifically.

Finally, we remark that the results we have derived can be immediately taken over to the case where the infinite sums involve the modified Bessel function of the first kind.

We wish to remark here that it was not our intent to be exhaustive. We confined our attention to the specific series arising in the physical problem of immediate interest. Nevertheless, the method appears to be easily applicable to other similar series. Furthermore, since it is straightforward, we would expect such generalizations to be immediately transparent. As a case in point, consider a result of Kendall[10] which established the following:

$$\gamma\sum_{n=-\infty}^{\infty}J_{\alpha-n\gamma}(z)J_{\beta+n\gamma}(z)=J_{\alpha+\beta}(2z), \tag{3.1}$$

for $0<\gamma<2$. Our method used to sum $\hat{S}_2$ immediately generalizes this result to all finite γ:

$$\gamma\sum_{n=-\infty}^{\infty}J_{\alpha-n\gamma}(z)J_{\beta+n\gamma}(z)=\sum_{k=0}^{\infty}H(\gamma-k)J_{\alpha+\beta}\left(2z\cos\frac{k\pi}{\gamma}\right)\cos\left(\frac{\eta k\pi}{\gamma}\right), \tag{3.2}$$

where η and H are as defined in Sec. II. Indeed, this result is simpler to obtain than that for $\hat{S}_2$ in that the sum need not be broken into summation over even and odd n.

IV. APPLICATIONS

The immediate application is to the physical problem described in the introduction. It is a straightforward application of Eq. (2.16), with $\alpha=\beta=0$, $\gamma=1$, and j running from 0 to 3, to show

$$\sum_{n=-\infty}^{\infty}\frac{J_n^2(z)}{n+\mu}=\frac{\pi J_\mu(z)J_{-\mu}(z)}{\sin\pi\mu}, \tag{4.1}$$

$$\sum_{n=-\infty}^{\infty}\frac{nJ_n^2(z)}{n+\mu}=-\frac{\mu\pi J_\mu(z)J_{-\mu}(z)}{\sin\pi\mu}+1, \tag{4.2}$$

$$\sum_{n=-\infty}^{\infty}\frac{n^2J_n^2(z)}{n+\mu}=\frac{\mu^2\pi J_\mu(z)J_{-\mu}(z)}{\sin\pi\mu}-\mu, \tag{4.3}$$

and

$$\sum_{n=-\infty}^{\infty}\frac{n^3J_n^2(z)}{n+\mu}=-\frac{\mu^3\pi J_\mu(z)J_{-\mu}(z)}{\sin\pi\mu}+\mu^2+\frac{z^2}{2}. \tag{4.4}$$

These sums are simple enough that they can be confirmed directly since explicit expressions for series of the form

$$\sum_{n=1}^{\infty}n^jJ_n^2$$

are available for j small.[2] This, however, suggests a further application of our results. The expression for $\hat{S}_2$, Eq. (2.15) is essentially an "inversion" formula for Neumann's expansion of a polynomial in an infinite series of Bessel functions and a generalization of Schlömilch's expansion of the type Σn^pJ_n to products of Bessel functions.[4]

Finally, we mention an application to an infinite series of products of Bessel functions which appears as the Fourier coefficients of the solution of a certain differential equation arising in a problem in electric circuits. Numerical values of this series, which has been called "Turkin's" function, have been given in a publication of the Ministry of Communications of the USSR.[3] Turkin's function is defined by

$$T_m(z,\alpha)=\sum_{n=-\infty}^{\infty}\frac{J_n(z)J_{n-m}(z)}{n-\alpha},\quad m\text{ integer} \tag{4.5a}$$

$$=(-1)^m\sum_{n=-\infty}^{\infty}\frac{(-1)^nJ_n(z)J_{m-n}(z)}{n-\alpha}. \tag{4.5b}$$

This is clearly a special case of our result for $j=0,\mu=-\alpha$, $\gamma=1,\alpha=m,\beta=0$. Thus we have

$$T_m(z,\alpha)=-\frac{(-1)^m\pi}{\sin\alpha\pi}J_{m-\alpha}(z)J_\alpha(z). \tag{4.6}$$

This result simplifies considerably the numerical evaluation of Turkin's function.

ACKNOWLEDGMENTS

The physical problem stimulating the calculation presented in this work was related to me by Don S. Lemons and Carlos R. Handy. I have enjoyed many conversations with them. I have expecially profited from many discussions with Huan Lee. This work was supported by the U.S. Department of Energy.

[1]Barry Newberger and Don Lemons, Bull. Am. Phys. Soc. **26**, 1001 (1981).
[2]Eldon R. Hansen, *A Table of Series and Products* (Prentice-Hall, Englewood Cliffs, NJ, 1975), p. 399.
[3]John Todd, Math. Rev. **32**, 1006 (1966).
[4]G. N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd ed. (Cambridge U. P., London, 1952), p. 150.
[5]T. J. Bromwich, *An Introduction to the Theory of Infinite Series*, (Macmillan, London, 1949), pp. 369–370.
[6]M. J. Lighthill, *Introduction to Fourier Analysis and Generalized Functions* (Cambridge U. P., London, 1958), p. 68.
[7]J. Mikusinski and R. Sikorski, "The Elementary Theory of Distributions," *Rozprawy Matematyczne*, No. XII, **1957**, pp. 1–54.
[8]E. C. Titchmarsh, *The Theory of Functions* (Oxford U. P., London, 1939), p. 110.
[9]Phillip M. Morse and Herman Feshback, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. I, Chap. 4, p. 413.
[10]D. G. Kendall, Quart. J. Math. **13**, 172 (1942).

Some algebraic, geometric, and system-theoretic properties of the Special Functions of mathematical physics [a),b)]

Robert Hermann
The Association for Physical and Systems Mathematics, 53 Jordan Road, Brookline, Massachusetts 02146

(Received 28 January 1981; accepted for publication 26 February 1982)

It is known that many of the Special Functions of mathematical physics appear as matrix elements of Lie group representations. This paper is concerned with a beginning attack on the converse problem, i.e., finding conditions that a given function be a matrix element. The methods used are based on a combination of ideas from system theory, functional analysis, Lie theory, differential algebra, and linear ordinary differential equation theory. A key idea is to attach a *symbol* as an element of a commutative algebra. In favorable cases, this symbol defines a Riemann surface, and a meromorphic differential form on that surface. The topological and analytical invariants attached to this form play a key role in system theory. The Lie algebras of the groups appear as linear differential operators on this Riemann surface. Finally, it is shown how the Picard–Vessiot–Infeld–Hull theory of factorization of linear differential operators leads to realization of many Special Functions as matrix representations of group representations.

PACS numbers: 02.30.Gp, 02.40. – k, 02.10.Sp, 02.30.Jr

1. INTRODUCTION

One classical way to study the Special Functions is to describe them as solutions of linear ordinary differential equations with analytic meromorphic coefficients.[1,2] More recently, it has been recognized that they also arise as matrix elements[3] of Lie group representations, and that many of their properties are more naturally related to this genesis. However, the relation between these two descriptions has not been studied in detail. The purpose of this paper is to present work aimed toward linking the two basic approaches, using hybrid methods coming from linear system theory,[4–6] classical analysis, the Picard–Vessiot theory[7–10] of differential algebra, and Riemann surface theory.[11]

A topic in mathematical system theory[4–6] called the *realization problem* deals with the relation between the two approaches. There one asks how a continuous complex valued function $t \to f(t)$ of a real variable, $0 \leqslant t < \infty$, may be represented in the form

$$f(t) = \langle v^d / e^{At} v \rangle, \tag{1.1}$$

where v is an element of a topological vector space V, v^d an element of its dual, and $t \to e^{At}$ is a one parameter continuous semigroup of operators on V. There is also a more general question of realizing a matrix function of t,

$$f(t) = \begin{pmatrix} f_{11}(t),\ldots,f_{1n}(t) \\ \vdots \\ f_{m1}(t),\ldots,f_{mn}(t) \end{pmatrix} \tag{1.2}$$

as follows:

$$f(t) = \begin{pmatrix} \langle v_1^d / e^{At} v_1 \rangle \cdots \langle v_1^d / e^{At} v_n \rangle \\ \vdots \\ \langle v_m^d / e^{At} v_n \rangle \cdots \langle v_m^d / e^{At} v_n \rangle \end{pmatrix}, \tag{1.3}$$

where $(v_1^d,\ldots,v_m^d)$ and $(v_1,\ldots v_n)$ are elements of v^d and v, respectively. The first is called the *scalar*, the second the *multivariable* case. Similar realization questions also appear frequently in probability and statistics.

In the system theory literature, this is presented in a slightly different form, emphasizing the *input–output* properties, and utilizing the basis-free methods of modern *linear algebra*. Namely, one is given three vector spaces (V,U,Y) called the *state*, *input*, and *output* spaces, respectively, a triple (A,B,C) of linear maps

$$\begin{aligned} &A\colon V \to U, \\ &B\colon U \to V, \\ &C\colon V \to Y, \end{aligned} \tag{1.4}$$

and a system of linear, time-invariant differential equations:

$$\begin{aligned} &\frac{dv}{dt} = Av + Bu, \\ &y = Cv. \end{aligned} \tag{1.5}$$

The solution of (1.5) with zero initial conditions

$$v(0) = 0 \tag{1.6}$$

is then

$$y(t) = \int_0^t Ce^{A(t-\tau)} Bu(\tau)\, d\tau. \tag{1.7}$$

The map

input curves → output curves

is a Volterra integral operator whose kernel is the function (which is a matrix-valued function when bases are chosen for U and Y)

$$t \to C \exp(tA)B. \tag{1.8}$$

In the system theory literature,[4–6] there is a complete answer to the question of existence and classification of such state-space realizations under some sort of natural "equivalence" only in case the elements of the matrix f are functions of t of a special type, namely, they belong to the vector space

[a)] A revision of "Infinite Dimensional Linear Systems, Riemann Surfaces, and Lie Group Harmonic Analysis, Part I" (unpublished).

[b)] Supported by a grant from the Ames Research Center (NASA), # NSG2402, from the Army Research Office, # ILIG1102RHN7-05MATH, and from the National Science Foundation.

 0022-2488/82/071282-13$02.50

(that we will call $\mathscr{R}$ of functions which are a finite sum of those of the form

$$t^j e^{\lambda t}, \quad j \text{ integer} \geqslant 0, \quad \lambda \in \mathbb{C}. \tag{1.9}$$

[These are the functions that are precisely the inverse Laplace transform of rational functions $D(s): s \to F(s)$ of the complex variables, which vanish at $s = \infty$. In the harmonic analysis literature, these functions are called *exponential polynomials*.] In fact, these results (proved in definitive form only in the early 1960's) have played an important role in the algebrazation of system theory in the last 20 years. The key fact here is that V can be taken to be a finite-dimensional vector space, so that the methods of linear algebra and finitely generated module theory suffice.

There have been many generalizations of these basic results to wider classes of functions. These mainly are based on various ideas of functional analysis. However, in certain of these papers,[12–14] there are indications that the Special Functions, and their genesis in Lie group harmonic analysis, play a role.

The purpose of this paper is to build stronger links between this dual genesis of the Special Functions. It will also give us the opportunity to adapt certain aspects of the general machinery of functional analysis, linear ordinary differential equation theory,[15] and differential algebra[9,10] to this problem. The treatise by Yosida[16] will be the standard reference for the functional analysis material.

2. THE TITCHMARSH ALGEBRA

Let R_+ be the non-negative real numbers. We will also denote it as $[0, \infty)$. Let $C(R_+)$ be the vector space of continuous, complex-valued functions on R_+. An element $f \in C(R_+)$ is a continuous map

$$f: R_+ \to \mathbb{C}, \tag{2.1}$$
$$t \to f(t), \quad t \geqslant 0.$$

Introduce the operation

$$(f_1, f_2) \to f_1 * f_2, \tag{2.2}$$

$$(f_1 * f_2)(t) = \int_0^t f_1(\tau) f_2(t - \tau)\, d\tau. \tag{2.3}$$

It is called the *causal convolution*. It makes $C(R_+)$ into a commutative associative algebra. By a theory of Titchmarch[16,17] this algebra has *no divisors of zero*. We will call this the *Titchmarsh algebra*.

A unit element can be added to the Titchmarsh algebra (in this case it is just the Dirac delta function) to make an *integral domain*. The quotient field, denoted by $\mathscr{M}$ is the basic object in the Mikusinski theory of generalized functions.[17]

$C(R_+)$ has another important algebraic structure,[18] namely, it is a *differential algebra*, i.e., there is a linear map $d: C(R_+) \to C(R_+)$ such that

$$d(f_1 * f_2) = df_1 * f_2 + f_1 * df_2, \tag{2.4}$$

for $f_1, f_2 \in C(R_+)$,

$$d(f)(t) = tf(t). \tag{2.5}$$

This differential algebra structure on $C(R_+)$ (and on the quotient field $\mathscr{M}$ considered by Mikusinski) plays a basic role in certain system-theoretic questions.

3. PICARD–VESSIOT INTEGRAL OPERATORS AND THE GALOIS GROUP

We have just seen that associated with any linear integral operator of the form

$$y(t) = \int_0^t f(t - \tau) u(\tau)\, d\tau \tag{3.1}$$

there is assigned an element f (which is its *symbol*) in a commutative differential algebra

$$(C(R_+), *, d).$$

We can now make use of ideas of *differential algebra* to study the algebraic properties of such operators.

Consider linear operators

$$D: C[0, \infty) \to C[0, \infty)$$

of the form

$$D = r_n * d^n + \cdots + r_0 * \tag{3.2}$$

with $r_0, \ldots, r_n \in \mathscr{R}$. Let us say that the operator (3.1) is of *Picard–Vessiot type* if there is a D of the form (3.2) such that

$$D(f) = r, \tag{3.3}$$

for some $r \in \mathscr{R}$.

Given an $f \in C(R_+)$, which satisfies an equation of the form (3.3), we assign to f a subfield $\mathscr{R}(f)$ of $\mathscr{M}$ as follows:

$\mathscr{R}(f)$ is the smallest differential subfield of $\mathscr{M}$ containing $\mathscr{R}, f$, and all solutions of the homogeneous equation

$$D(h) = 0,$$
$$h \in \mathscr{M}.$$

Definition: The Galois group of the integral operator (3.1) is the group of automorphisms of the differential field $\mathscr{R}(f)$ which leave fixed each element of $\mathscr{R}$.

In this paper, we will not pursue in detail this purely algebraic approach. Assuming $\mathscr{R}(f)$ can be generated by elements that are Laplace transformable,[9] $\mathscr{R}(f)$ is isomorphic to a Picard–Vessiot field of analytic functions, in the classical sense.[10] It is known[2] that, at least for certain types of such fields, the Picard–Vessiot group is related to the *algebraic closure* of the *monodromy group*.[2] This gives us a way of approaching the subject that is much closer to traditional mathematical physics. We will now review some relevant material.

4. LAPLACE TRANSFORM

It is well known[9] that the Titchmarsh algebra structure can be studied by the Laplace transform

$$\mathscr{L}(f)(s) = \int_0^\infty e^{-st} f(t)\, dt. \tag{4.1}$$

On the submanifold of $C(R_+)$, where the Laplace transform exists in some appropriate sense, it transforms the product (2.3) into the ordinary product of functions

$$\mathscr{L}(f_1 * f_2) = \mathscr{L}(f_1)\mathscr{L}(f_2). \tag{4.2}$$

We also have

$$\mathscr{L}(df) = \frac{d}{ds}\mathscr{L}(f). \tag{4.3}$$

Thus, to the extent that the Laplace transform is defined, the "differential algebra" $\{C(R_+),*,d\}$ is translated over to "differential algebra" in the classical Picard–Vessiot sense, involving linear, ordinary differential operators with analytic coefficients.

Let us now be more precise about these ideas, and define subspaces of $C(R_+)$ in the following way: For $s\in\mathbb{C}$, set

$$C(R_+,s) = \{f\in C(R_+): |f(t)| \leqslant e^{st} | \quad \text{for all } t\geqslant 0\}, \tag{4.4}$$

$L^1(R_+,|e^{st}|dt) \equiv$ set of Lebesgue measurable maps: $R_+\to\mathbb{C}$ which are absolutely integrable over $0\leqslant t<\infty$ with respect to the measure $|e^{-st}|dt$. (4.5)

Then

$$C(R_+,s)\subset L^1(R_+,s') \tag{4.6}$$

if the real part of $(s'-s)>0$.

Notice that these are precise classes of functions to which the classical theory of Laplace transform theory[9] applies. One of the areas to be treated in this paper is the description of certain sufficient conditions, deriving from functional analysis,[16] which imply that elements of $C(R_+)$, defined as matrix elements of semigroups, belong to these classes. We will also compare conditions of this nature arising from linear differential equation[15] and Volterra path integral theory.[20] These growth conditions will also play a role in the system-theoretic realization questions.

5. HILLE–YOSIDA THEORY AND THE GROWTH CONDITIONS

Let us now suppose that V is a topological vector space (Ref. 16, p. 25) with the complex numbers $\mathbb{C}$ as field of scalars. Let $L(V,V)$ be the space of linear continuous maps: $V\to V$. Let V^d be the dual vector space, i.e., the vector space of *continuous* linear maps: $V\to\mathbb{C}$. (At least for the moment, we do not impose any topological structure on V^d.)

A *semigroup* in V is map $R_+\to L(V,V)$, $t\to g(t)$, such that the following conditions are satisfied:

$$g(0) = \text{identity}, \tag{5.1}$$

$$g(t_1+t_2) = g(t_1)g(t_2), \quad \text{for } t_1,t_2\in R_+. \tag{5.2}$$

For each $v\in V$, the map $t\to g(t)v$ of $R_+\to V$ is continuous. (5.3)

Given $v\in V$, $v^d\to V^d$, we can form the function

$$f: R_+\to\mathbb{C}$$

by the following formula:

$$f(t) = \langle v^d, g(t)v\rangle. \tag{5.4}$$

The function f, defined by formula (5.4), is then an element of $C(R_+)$. We are looking for conditions that f belongs to the subsets $C(R_+,s)$, $L^1(R_+,e^{-st}dt)$, defined in Sec. 2, for some $s\in R_+$. Conditions of this sort are proved in Yosida's treatise,[16] resulting from now-classical work by Hille and Yosida.

Theorem 5.1: Suppose that the topology on V is defined by a Banach space norm. In addition, suppose that conditions (5.1)–(5.4) are satisfied. Then,

$$f\in C(R_+,s), \quad \text{for some } s\in R_+.$$

Proof: For the proof, see Sec. 1, Chap. 9 of Ref. 16.

Theorem 5.2: Suppose that V is a sequentially complete, locally convex topological vector space. Suppose again, that conditions (5.1)–(5.4) are satisfied. Then,

$$f\in L^1(R_+, e^{-st}dt), \quad \text{for some } s\in R_+.$$

Proof: See Sec. 3, Chap. 9 of Ref. 16.

6. EXISTENCE OF THE LAPLACE TRANSFORM FOR FUNCTIONS DEFINED VIA LINEAR, TIME-DEPENDENT SYSTEMS

In this section, we will use the theory of linear ordinary differential equations[15] to prove that the Laplace transform of certain elements of $C(R_+)$ exists.

Again, let t be a real variable, $t\geqslant 0$. Let $M(n,C)$ be the Lie algebra (under commutator) of $n\times n$ complex matrices. Let $\alpha: R_+\to M(n,C)$ be a continuous curve in $M(n,\mathbb{C})$. Consider the linear, ordinary differential equation

$$\frac{dx}{dt} = \alpha(t)x + z(t) \tag{6.1}$$

for the curve

$$t\to\begin{pmatrix} x(t)\\ \vdots\\ x_n(t)\end{pmatrix} = x(t) \tag{6.2}$$

in $\mathbb{C}^n$. $t\to f(t)$ is a given continuous curve in $\mathbb{C}^n$. Let $t\to f(t)$ be the element of $C(R_+)$ defined as follows:

$$f(t) = \gamma^T x(t), \tag{6.3}$$

$$x(0) = x_0, \tag{6.4}$$

with

$$\gamma\in\mathbb{C}^n, \quad \gamma^T = \text{matrix transpose}.$$

Our concern is to find sufficient conditions (involving the data α, z, x_0) that the Laplace transform f exist. The standard asymptotic theory of linear ordinary differential equations (Ref. 15, Chap. 4) provides one type of condition of this sort. The Volterra product integral formalism[20] provides another technique for deriving such estimates. The results from the two techniques seem similar, hence we will only consider the method of Ref. 15 here.

Let $|\ |$ denote the standard quadratic norm on $\mathbb{C}^n$ and $M(n,\mathbb{C})$.

$$|x| = (|x_1|^2+\cdots+|x_n|^2)^{1/2}, \tag{6.5}$$

$$|\alpha| = \sup_{x\in\mathbb{C}^n}\frac{|\alpha x|}{|x|}, \quad \text{for } \alpha\in M(n,\mathbb{C}). \tag{6.6}$$

Then, using formula (6.2) of Chap. 4, p. 56 of Ref. 15, we

have

$$|x(t)| \leqslant |x_0| \exp \int_0^t \lambda(\alpha(\tau))d\tau + \int_0^t |z(\tau)| \exp \int_\tau^t \lambda((\alpha(s))ds\, d\tau, \tag{6.7}$$

where $\lambda(\alpha)$ is the greatest eigenvalue of $\alpha + \alpha^*$. From (6.3), we also have

$$|f(t)| \leqslant |\gamma||x(t)|. \tag{6.8}$$

Equations (6.7) and (6.8) give estimates that are useful for finding sufficient conditions that $t \to f(t)$ have at most exponential growth, hence that its Laplace transform exists.

Example: f satisfies a Sturm–Liouville equation

$$\frac{d^2 f}{dt^2} = -a^2 \lambda f + uf \quad (\lambda \text{ constant}). \tag{6.9}$$

Convert this into a system in the usual way.

$$f = x_1, \tag{6.10}$$

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \tag{6.11}$$

$$\frac{dx}{dt} = \left[\begin{pmatrix} 0, & a \\ -a, & 0 \end{pmatrix} + \begin{pmatrix} 0, & 0 \\ u/a, & 0 \end{pmatrix} \right] x. \tag{6.12}$$

Hence,

$$\alpha = \begin{pmatrix} 0, & \lambda \\ -\lambda, & 0 \end{pmatrix} + \begin{pmatrix} 0, & 0 \\ u(t)a^{-1}, & 0 \end{pmatrix},$$

$$\alpha + \alpha^* = \begin{pmatrix} 0, & u^* a^{-1^*} \\ u, a^{-1}, & 0 \end{pmatrix}.$$

Then,

$$\lambda(\alpha + \alpha^*) = |u(t)a^{-1}|. \tag{6.13}$$

Equation (6.13), substituted into (6.7), gives sufficient conditions for exponential growth.

Of course, this is but the simplest of a whole series of sufficient conditions that the Laplace transform exists for solutions of linear, ordinary differential equations.

Another series of sufficient conditions for

$$\int_0^\infty e^{-st} f(t)dt$$

to exist is derivable from the hypothesis that $t \to f(t)$ is a matrix element of a Lie group representation in a Banach space.

7. LAPLACE TRANSFORM AND THE WEYL ALGEBRA OF LINEAR DIFFERENTIAL OPERATORS

We now review another classical topic that is relevant to the system-theoretic problem, namely, the linear differential equations with polynomial coefficients satisfied by a function and its Laplace transform.

Let "z" denote a complex variable. Let $\mathbb{C}[z]$ denote the polynomial in z with complex coefficients. A linear differential operator of the form

$$D = p_n(z) \frac{d^n}{dz^n} + \cdots + p_0(z), \quad p_0, \ldots, p_n \in \mathbb{C}[z] \tag{7.1}$$

is called a *Weyl operator*. The operators form an associative algebra under composition called the *Weyl algebra* and denoted by $\mathscr{W}$.

Consider the Laplace tranform

$$(f)(w) = \int e^{-zw} f(z)dz. \tag{7.2}$$

The integral in (7.2) is over a curve in the complex t plane, ranging from $z = a$ to $z = b$. If D is a Weyl operator, then $\mathscr{L}(Df)$ is obtained by applying a differential operator in z, which we call $\mathscr{L}(D)$, to $\mathscr{L}(f)$ plus some boundary terms. Let us formulate this algebraically in the following way.

$$\mathscr{L}(Df) = \mathscr{L}(D)(\mathscr{L}f) + \int \frac{d}{dt} [(\theta(D))(\mathbf{e}_s, f)]dt, \tag{7.3}$$

where $\mathbf{e}_s$ is the function $t \to e^{-st}$ and

$$\theta(D): (f_1, f_2) \to \theta(f_1, f_2)$$

is a bilinear differential operator in the functions f_1, f_2.

In particular, then, we have

$$\mathscr{L}(Df) = \mathscr{L}(D)(\mathscr{L}f) \tag{7.4}$$

for all functions on γ which are C^∞, but vanish to the infinite order at the boundaries. This shows that

$$\mathscr{L}(D_1 D_2) = \mathscr{L}(D_1)\mathscr{L}(D_2). \tag{7.5}$$

$\mathscr{L}$ is a homomorphism of the Weyl algebra. We will show how to construct $\mathscr{L}$ in terms of Lie algebra theory.

Let $\mathscr{H}$ be the three-dimensional Lie algebra generated by three elements $(p, q, 1)$ satisfying the following commutation relations:

$$[p, q] = 1,$$

$$[1, p] = 0 = [1, q].$$

$\mathscr{H}$ is called the *Heisenberg Lie algebra*.

Let $U(\mathscr{H})$ be the universal associative enveloping algebra associated with $\mathscr{H}$. Let $\mathscr{I}$ be the associative algebra ideal generated by elements of the form

$$\Delta - \Delta 1,$$

$$\Delta \in U(\mathscr{H}).$$

We will now show that the associative algebra

$$U(\mathscr{H})/\mathscr{I}$$

is isomorphic to the Weyl algebra $\mathscr{W}$.

Define a Lie algebra homomorphism of $\mathscr{H}$ into $\mathscr{W}$ as follows:

$$p \to \frac{d}{dt},$$

$$q \to \text{multiplication by } t,$$

$$1 \to \text{multiplication by } 1.$$

This representation, call it $\rho: \mathscr{H} \to \mathscr{W}$, extends as an associative algebra homomorphism to $U(\mathscr{H})$:

$$\rho(h_1 \cdots h_n) = \rho(h_1) \cdots \rho(h_n), \quad \text{for } h_1, \ldots, h_n \in \mathscr{H}.$$

The ideal $\mathscr{I}$ is mapped into zero, hence ρ passes to the quotient to define an associative algebra homomorphism

$$\rho: U(\mathscr{H})/\mathscr{I} \to \mathscr{W}.$$

It is now readily seen that ρ is an isomorphism.

Now, for the Laplace transform, we have

$$\mathscr{L}(t) = -\frac{d}{ds}, \tag{7.6}$$

$$\mathscr{L}\left(\frac{d}{dt}\right) = s.$$

Define a Lie algebra isomorphism

$$\alpha: \mathscr{H} \to \mathscr{H}$$

as follows:

$$\begin{aligned} &\alpha(q) = -p, \\ &\alpha(p) = q, \\ &\alpha(1) = 1. \end{aligned} \tag{7.7}$$

α extends to an associative algebra isomorphism

$$\alpha: U(\mathscr{H}) \to U(\mathscr{H}).$$

α maps $\mathscr{I}$ into $\mathscr{I}$, hence defines an isomorphism

$$\mathscr{W} \to \mathscr{W}.$$

It is obvious that this is just the map $\mathscr{L}$ defined by relation (7.4).

8. MONODROMY PROPERTIES OF LINEAR ORDINARY DIFFERENTIAL EQUATIONS

Now, we review certain classical material about the monodromy group of linear ordinary differential equations.[2,21] Consider an inhomogeneous linear ordinary differential equation

$$a_n(z)d_z^n(y) + \cdots + a_0 y = f. \tag{8.1}$$

Here,

$$d_z = \frac{d}{dz}.$$

The coefficients $a_0,\ldots,a_n$ are polynomials in z. We can have this make sense at $z = \infty$ also, by substituting $u = z^{-1}$, in the usual way. Let Z be the "Riemann sphere," i.e., the complex plane with variable z, with $z = \infty$ added on (conformally) in this way. We can solve (8.1) for y as an analytic function of z in a neighborhood about any point in which $a_n(z) \neq 0$, and analytically continue it throughout Z. Let Z' be Z with the singular points of (8.1), i.e., the points where $a_n(z) = 0$ omitted. Let $\mathbf{Z}'$ be the simply connected covering space to Z'. Let

$$\phi: \mathbf{Z}' \to Z'$$

be the projection map. π is a principal fiber bundle with structure group

$$\pi_1(Z'),$$

i.e., the fundamental group of Z'. Construct $\mathbf{Z}$ in the following way. Pick $z_0 \in Z'$ arbitrarily. Let us then say that two curves which begin at z_0 are *equivalent* is they satisfy the following relations:

they have the same endpoint,
they are homotopic, with endpoints fixed.

$\mathbf{Z}'$ is then the quotient of the space of paths under this equivalence relation. $\pi_1(Z')$ is the quotient of the space of loops based at z_0. Path composition defines the action

$$\pi_1(Z') \times \mathbf{Z}' \to \mathbf{Z}',$$

which defines the principal fiber bundle action of $\pi_1(Z')$.

Now, z_0 is a point chosen so that the *homogeneous* differential equation

$$a_n d_z^n + \cdots + a_0 = 0 \tag{8.2}$$

has n linearly independent solutions which are analytic in a neighborhood of z_0. Fix such a neighborhood and let V be the complex vector space of functions spanned by the solutions of (8.1). Since the difference of two solutions of (8.1) is a solution of (8.2), V includes the space of solutions of (8.2). It is equal to it if and only if $f \neq 0$. Hence, if $f \neq 0$,

$$\dim V = n + 1.$$

One can start with a solution at z_0 and analytically continue it along paths in Z'. The result is (by the monodromy principle of function theory) independent of the homotopy class of the path. In particular, analytic continuation along closed loops defines a homomorphism

$$\pi_1(Z') \to \mathrm{GL}(V),$$

i.e., a linear representation of $\pi_1(Z')$, which is called the *monodromy group* of the differential equation (8.1). It is the basic object of study in the classical work on linear, ordinary differential equations, e.g., by Fuchs, Schwarz, Poincaré, Klein, Picard, and Vessiot. It also plays a basic role in the theory of *automorphic functions*. The monodromy group can also be defined as the holomony group of a connection on a vector bundle.

Now, if $f \neq 0$, V is an $(n+1)$-dimensional vector space. Let V_0 be the n-dimensional linear subspace spanned by the solutions (in a neighborhood of z_0) of the solutions of the homogeneous equation (8.2).

Theorem 8.1: The monodromy group leaves V_0 invariant and acts as the identity in V/V_0.

Proof: That it leaves V_0 invariant just means that analytical continuation of a solution of (8.2) along a closed loop returns to a solution of (8.2), which is obvious. The trivial action in V/V_0 follows since analytical continuation of a solution y of (8.1) returns to another solution y, and $y - y_1$ lies in V_0. Q.E.D.

Let K be the kernel of the monodromy representation. It is an invariant subgroup of $\pi_1(Z')$. Let S' be the orbit space of K acting on $\mathbf{K}'$. The quotient group

$$\pi_1(z')/K$$

acts on S'. The map

$$\phi: \mathbf{Z}' \to Z'$$

passes to the quotient to define a covering map

$$\phi_K: S' \to Z'.$$

One can similarly define the *homogeneous monodromy group* associated with the differential equation (8.2). It is an action of $\pi_1(Z')$ on V_0. Let K_0 be its kernel. There is a homomorphism

$$\pi_1(Z')/K \to \pi_1(Z')/K_0,$$

which defines the inhomogeneous monodromy group as an

extension of the *homogeneous one*. This extension is *central*, i.e., its kernel is in the center. Thus, the inhomogeneous monodromy group is determined by the second group cohomology of the homogeneous one.

9. MONODROMY CONDITIONS THAT A SOLUTION OF A WEYL EQUATION BE ALGEBRAIC

A function y defined in a region of the complex z plane is said to be *algebraic* if it satisfies a relation of the form

$$p_n(z)y(z)^n + \cdots + p_0(z) = 0, \quad \text{for all } z \in Z, \tag{9.1}$$

where the $p_0,\ldots,p_n$ are polynomials in z (in other words, if the field of functions containing y and the rational functions is an algebraic extension of the rationals). We will be concerned with conditions that specified differential equations have algebraic solutions. Suppose that

$$D(y) = f \tag{9.2}$$

is such an equation, where D is a Weyl operator.

As in Sec. 8, we can study this question by choosing a nonsingular point z_0, letting V be the vector space of analytic functions in a neighborhood of z_0 generated by the solutions to (9.2), and continuing along all loops in the Riemann sphere minus the singular points of (9.2). Here is the classical answer.[2] This work has recently been given a modern setting by Baldassari and Dwork.[22]

Theorem 9.1: An element $y \in V$ is an algebraic function if and only if the orbit of y under the monodromy group is finite. In particular, all elements of V are algebraic functions if and only if the monodromy group is finite.

Proof: Suppose first that y is algebraic over $\mathbb{C}[z]$, i.e., that it satisfies an algebraic relation of the form (9.1). Now, analytically continue y over curves in Z'. The coefficients $p_0,\ldots,p_n$ remain unchanged, but y transforms into the orbits of; y under the monodromy group. Hence, the elements in this orbit also satisfy such an algebraic relation. But, there are only a finite number of such functions, i.e., the orbit is finite.

Conversely, suppose

$$y = y_1,\ldots,y_n$$

are the orbits of y under the monodromy group. Let

$$p_1,\ldots,p_n$$

be the elementary symmetric function in $y_1,\ldots,y_n$ in the function of z such that

$$\prod_{j=1}^{n}(y_j(z) - \lambda) = \sum_{j=0}^{n} p_j(z)\lambda^j, \tag{9.3}$$

for all $\lambda \in \mathbb{C}$.

As this relation is analytically contained under the monodromy group (with λ held fixed) the $y_1,\ldots,y_n$ are permuted, i.e., the left-hand side remains invariant under the monodromy group. Hence, they are meromorphic functions over the whole Riemann sphere, i.e., are rational functions on z. Now, substitute $\lambda = y_1(z)$ to obtain a relation of the form (9.1).

Remark: A word about the relevance of this to the theory of Special Functions might be useful. Most of them satisfy an equation of type (9.2). Some of them (e.g., the Bessel functions) have the property that their *Laplace transform* is an *algebraic function*.[23] Similarly, certain of the solutions of the hypergeometric equation are algebraic, and the Laplace transform of a hypergeometric equation is another hypergeometric equation.

10. THE GALOIS GROUP OF A PICARD–VESSIOT EXTENSION OF THE RATIONALS AND THE MONODROMY GROUP

Let us return to the Titchmarsh differential algebra $C(R_+)$ with the rational subfield $\mathscr{R}$. Let $t \to f(t)$ be an element of $C(R_+)$ which is a Picard–Vessiot element relative to the subalgebra $\mathscr{R}$. We can associate algebraic invariants with f (and *a fortiori* the input–output systems associated with f) by using the notion of the *Galois group*[2,9,10] of a *differential field*.

Let $\mathscr{M}$ be the quotient field of the integral domain obtained from $C(R_+)$ by adjoining a unit element. The differential extends to $\mathscr{M}$, to define it as a *differential field*.[10] Suppose f is an element of $C(R_+)$, i.e., $t \to f(t)$ is a continuous map, satisfying the following conditions:

(a) f lies in a Picard–Vessiot differential field $\mathscr{M}(f)$ such that

$$\mathscr{R} \subset \mathscr{M}(f) \subset \mathscr{M}, \tag{10.1}$$

(b) f is Laplace transformable, i.e.,

$$\int_0^\infty |f(t)e^{-st}|dt < \infty, \quad \text{for some } s \in \mathbb{C}. \tag{10.2}$$

This Laplace transform sends f into a function

$$s \to \mathscr{L}(f)(s),$$

which is analytic in a half-plane, and satisfies a linear ordinary differential equation with polynomial coefficients. The Galois group of this differential equation is clearly the "abstract" Galois of $\mathscr{M}(f)$.

The monodromy group of this differential equation is clearly contained in the Galois group.[2] Since the Galois group is algebraic, it contains the algebraic closure of the monodromy. In certain cases (e.g., if the differential equation is Fuchsian) the algebraic closure of the monodromy group is the Galois group.[2] It also follows on general algebraic principles that the finiteness of the Galois group of the differential field $\mathscr{M}(f)$ over the rationals $\mathscr{R}$ implies that $\mathscr{M}(f)$ is an algebraic extension of $\mathscr{M}(f)$.

In the 19th century considerable work (e.g., by Schwarz, Klein, Fuchs, and Jordan) went into finding explicit conditions that the Galois group is finite. These results and generalizations are most readily accessible in recent work by Baldassarri and Dwork.[22]

11. REALIZATION THEORY FOR INFINITELY DIFFERENTIABLE FUNCTIONS AND THE SPECIAL FUNCTIONS

The Special Functions of mathematical physics have many quasialgebraic properties (linked, say, to differential algebra) that play an important role both in their applica-

tions to applied problems and in the study of their general properties. The recursion relations are particularly important, linking algebraically each whole family. Finally, the integral representations link the Special Functions to Lie group theory. In Ref. 24, a realization method was presented for functions $t \to f(t)$ on R_+ that were infinitely often differentiable. I will review it here, then in several examples study its relation to recursion relations for the Special Functions.

Let the state space X be the space of C^∞, complex-valued functions on the half-line $0 \leqslant \tau < \infty$. Denote a typical element of X by

$$x: \tau \to x(\tau).$$

Let

$$A: C^\infty[0, \infty) \to C^\infty[0, \infty)$$

be the following linear operator:

$$A(x) = \frac{dx}{d\tau}. \tag{11.1}$$

Let the input and output vector space be $\mathbb{C}$, i.e., "scalar input–output". Let $B: \mathbb{C} \to X$, $C: X \to \mathbb{C}$ be defined as follows:

$$B(u) = uf, \quad \text{for } u \in \mathbb{C}, \tag{11.2}$$

$$C(x) = x(0), \quad \text{for } x \in X \equiv C^\infty[0, \infty). \tag{11.3}$$

The corresponding input–output system in state space form,

$$\frac{dx}{dt} = Ax + Bu, \tag{11.4}$$

$$y = Cx,$$

has the convolution solution (for zero initial condition)

$$y(t) = \int_0^t C \exp(t - \tau) Bu(\tau) dt. \tag{11.5}$$

Now, with A given by (11.1),

$$\exp(tA)(x)(\tau) = x(t + \tau). \tag{11.6}$$

Thus, using (11.2), (11.3), and (11.6), (11.5) takes the following form:

$$\begin{aligned} y(t) &= \int_0^t C[\exp(t - \tau) Bu(\tau)] d\tau \\ &= \int_0^t C[\exp(t - \tau)(f) u(\tau)] d\tau \\ &= \int_0^t C([\exp(t - \tau)(f)](\tau) u(\tau)) d\tau \\ &= \int_0^t f(t - \tau) u(\tau) d\tau. \end{aligned} \tag{11.7}$$

Let us sum up as follows.

Theorem 11.1: With the choice of data indicated above, the zero initial condition input–output relations corresponding to the scalar input–output system (11.4) take the form (11.7), so that it is a causal convolution operator with the function f as kernel (i.e., a multiplication in the Titchmarsh algebra).

What seems to be done to make this a "realization" in the sense of Sec. 5 is to impose topologies correctly. This can readily be done with the standard[16] locally vector space topologies for C^∞ functions.

Of course, another state space realization is obtained by taking X to be the closure of the elements $f, Af, A^2 f, \ldots$ in $C^\infty[0, \infty)$. This can be computed readily, and linked to many areas of classical analysis and mathematical physics, by choosing "f" to be one of the Special Functions.

Example:

$$f(t) = J_m(t),$$

the mth order Bessel function, m an integer.

The basic recursion relation we need is the following:

$$2 \frac{d}{dt} J_m(t) = J_{m-1}(t) - J_{m+1}. \tag{11.8}$$

Also,

$$J_{-n}(t) = (-1)^n J_n(t). \tag{11.9}$$

Thus, relations (11.8) and (11.9) provide, for each n, $A^n(f) \equiv A^n(B1)$,

$$A = \frac{d}{dt},$$

a relation of the following form:

$$A^n(f) = \sum_{\substack{j = m - n \\ j \geqslant 0}}^{m+1} a_j J_j, \quad a_j \in R. \tag{11.10}$$

This realizes A as an *infinite Jacobi matrix*. How to use this to embed A in a Lie algebra of operators isomorphic to the group of rigid motion of R^2 has been described in work by Baras, Brockett, and Fuhrmann.[12–14]

However, a more geometrically natural way to embed it in a Lie algebra of operators is to use the integral representations of the Bessel functions to construct integral intertwining operators between this representation on functions and another. Let $C^\infty(S^1)$ be the C^∞, complex-valued functions on the unit circle S^1 in R^2. These can be exhibited as the C^∞ functions $\theta \to h(\theta)$ of the real variables θ such that

$$h(\theta + 2\pi) = h(\theta).$$

Let

$$\phi: C^\infty(S^1) \to C^\infty(R_+)$$

be the following linear map:

$$\phi(h)(t) = \int_0^\pi e^{it \cos\theta} h(\theta) \, d\theta. \tag{11.11}$$

Set

$$\alpha(h)(\theta) = i \cos\theta. \tag{11.12}$$

Then,

$$\begin{aligned} \phi(\alpha(h))(t) &= \int_0^\pi e^{it \cos\theta} (i \cos\theta) h(\theta) d\theta \\ &= \frac{d}{dt} \int_0^\pi e^{it \cos\theta} h(\theta) d\theta \end{aligned}$$

(differentiation under the integral sign justified by standard rules)

$$\begin{aligned} &= \frac{d}{dt} (\phi(h)(t)), \\ &= \phi\alpha = A\phi, \end{aligned} \tag{11.13}$$

i.e., ϕ intertwines α and A.

h is now one element of a Lie algebra of first order linear differential operators on $C^\infty(S^1)$, which is isomorphic to the Lie algebra of E(2), the group of rigid motions in R^2.

$$\alpha = i\sin\theta, \quad \beta = i\cos\theta,$$

$$\gamma = \frac{d}{d\theta}. \tag{11.14}$$

This Lie algebra and related ones acting on $C^\infty(S^1)$ [e.g., that of SL $(2, R)$] has been extensively investigated in Ref. 25 from the point of view of "deformation" of Lie group representations. It is instructive to see how the other operators in the Lie algebra (11.14) behave relative to the intertwining operator ϕ. Note that

$$\beta^2 = 1 - \alpha^2; \tag{11.15}$$

hence,

$$\begin{aligned}\phi\beta &= \phi((1-\alpha^2)^{1/2}) \\ &= (\phi - A^2\phi)^{1/2} \\ &= (1-A^2)^{1/2}\phi,\end{aligned} \tag{11.16}$$

$$\begin{aligned}\phi(\gamma(h))(t) &= \int_0^\pi e^{it\cos\theta}\frac{d}{d\theta}(h(\theta))d\theta \\ &= \int_0^\pi e^{it\cos\theta} it\cos\alpha h(\theta)d\theta \\ &= t\int_0^\pi e^{it\cos\theta}\alpha(h)(\theta)d\theta \\ &= t\phi(\alpha(h))(\theta) \\ &= tA\phi(h)(\theta).\end{aligned} \tag{11.17}$$

Thus, we have

$$\phi\gamma = BA\phi, \tag{11.18}$$

where

$$B = \text{multiplication by } t. \tag{11.19}$$

B and A define the Lie algebra of the Heisenberg group

$$[A,B] = 1. \tag{11.20}$$

We can now sum up as follows.

Theorem 11.2: The scalar input–output realization of the Bessel functions leads to an embedding of A in the Lie algebra of the group of rigid motions of R^2, and realization of the Lie algebra in terms of an algebraic extension of the universal enveloping algebra of the Heisenberg Lie algebra, namely,

$$A, \quad (1-A^2)^{1/2}, \tag{11.21}$$

$$BA$$

form the Lie algebra of E(2). B is just the differential of the Titchmarsh algebra constructed and utilized in previous sections.

Remark: Such realizations of one Lie algebra in terms of an algebraic extension of another have appeared[25,26] in relation to the theory of deformations of Lie algebras and their representations, particularly in terms of what one calls the *Gell-Mann formula*.

12. REALIZATION THEORY IN TERMS OF THE LAPLACE TRANSFORM AND RIEMANN SURFACE THEORY

In Sec. 11 we discussed various aspects of the realization theory in case a given element f of the Titchmarsh algebra satisfied a special condition, namely, it was infinitely often differentiable. Now we discuss the realization theory under another assumption, namely,

$$\int_0^\infty |e^{-st}f(t)|\,dt < \infty, \quad \text{for } s\in R. \tag{12.1}$$

Then, the Laplace transform

$$\mathcal{L}(f)(s) = \int_0^\infty e^{-st}f(t)dt \tag{12.2}$$

will exist as an analytic function of the complex s plane in some region of the complex s plane. It will be convenient to define the Laplace transform as a complex analytic one-differential form

$$\omega(f) = \mathcal{L}(f)ds. \tag{12.3}$$

Thus, we have

$$\omega(f) = \int_0^\infty e^{-st}f(t)dt\wedge ds. \tag{12.4}$$

This can be interpreted in the following integral-geometric way:

$$e^{-st}f(t)dt\wedge ds \tag{12.5}$$

is a two-differential form on the manifold $R_+\times\mathbb{C}$. $\omega(f)$ is the form on $\mathbb{C}$ (or a region) that results from integrating this form over the fiber of the Cartesian projection map

$$(t,s)\to s \quad \text{of } R_+\times\mathbb{C}\to\mathbb{C}.$$

Once defined as an analytic one-differential form in some region of the complex s plane, $\omega(f)$ can be analytically continued. We will make the following assumption about the conditions $\omega(f)$ should satisfy:

$\omega(f)$ can be analytically continued to be a meromorphic form in a fixed neighborhood U about the point $s=\infty$ of the Riemann sphere. $\omega(f)$ has at most a pole of order one at $s=\infty$. (12.6)

Now, set

$$z = s^{-1}$$

and let

$X =$ vector space of one-forms which are meromorphic in U with a pole only at $s=\infty$, i.e., those which are of the form

$$\omega = (a_{-1}z^{-1} + a_0 + a_1 z + \cdots)dz. \tag{12.7}$$

Then,

$$a_{-1} = \text{res}\,(\omega, s=\infty), \tag{12.8}$$

the *residue* of the meromorphic form ω at the point at infinity.

To construct the system, proceed as follows. Set

$X' =$ vector space of one-forms ω of type (12.7), i.e., that are meromorphic in U with a single pole of type (12.7) at $s=\infty$; (12.9)

$C(\omega) =$ residue of ω at $s = \infty$, i.e., the left-hand side of (12.7); (12.10)

C is a linear map $X' \to \mathbb{C}$.

Define a linear map

$$B: \mathbb{C} \to X'$$

as follows:

$$B(c) = c\omega(f), \quad \text{for } c \in \mathbb{C}, \tag{12.11}$$

where $\omega(f)$ is the one-form which results from Laplace transforming f. Finally, define a linear map

$$A: X' \to X'$$

as follows:

$$A(\omega) = s\omega - C(\omega)ds'. \tag{12.12}$$

Set

$X =$ vector subspace of X' spanned by

$$\omega(f), A\omega(f), A^2\omega(f), \ldots .$$

Now, form the scalar input–output system with state space X:

$$\frac{dx}{dt} = Ax + Bu,$$
$$y = Cx. \tag{12.13}$$

Now one can explain the relation between the input–output system (12.2) and the one constructed in the "time domain" in Sec. 11.

Theorem 12.1: With the notation explained above, suppose that $f \in C^\infty[0, \infty)$ is a function such that the Laplace transform of f and all its derivatives exists in the region U of the complex s plane. Then, the input–output relations of the system (12.13), with initial conditions $x(0) = 0$, are of the form

$$y(t) = \int_0^t f(t - \tau)u(\tau)d\tau. \tag{12.14}$$

ω is related to the system as follows:

$$\omega(f) = (C(s - A)^{-1}B)ds. \tag{12.15}$$

We can make this realization more explicit. With ω of form (12.7) and $z = s^{-1}$, we have

$$\omega = (a_{-1}s + a_0 + a_1 s^{-1} \cdots)\left(-\frac{1}{s^2}ds\right)$$
$$= -(a_{-1}s^{-1} + a_0 s^{-2} + a_1 s^{-3} + \cdots)ds.$$

Hence

$$A(\omega) = s\omega + a_{-1}ds$$
$$= -(a_0 s^{-1} + a_1 s^{-3} + \cdots)ds. \tag{12.16}$$

Notice that A is what functional analysts call the *shift* operator:

$$(a_{-1}, a_0, a_1, \ldots) \to (a_0, a_1, a_2, \ldots). \tag{12.17}$$

It is also closely related to the "annihilation" operators of quantum mechanics, and thus to the Heisenberg group. These formulas also appear in the functional analysis approach[12–14] to the realization problem. However, the advantage of this interpretation of the formula is that it suggests various "geometrizations."

For example, one might postulate that the open set U be identified with an open set of another Riemann surface (not necessarily the Riemann sphere constructed from the complex variable s and the point at infinity) such that the differential forms ω which lie in the state space of the realization can be extended to be meromorphic one-forms on M. For example, f of $f(t)$ is one of the Bessel functions; its Laplace transform[19] is a meromorphic form on the Riemann surface M associated with the algebraic curves

$$w^2 = s^2 + 1. \tag{12.18}$$

One can similarly go through the tables of the Laplace transform of the Special Functions and reel off the associated Riemann surfaces.

Thus, one who is familiar with the theory of Riemann surfaces[11] will suspect that the *Riemann–Roch theorem* will play a role in the study of systems associated with the elements of the Titchmarsh algebra which lie in Picard–Vessiot extensions of $\mathscr{R}$. As preparation for such a study in a later publication, I will now show that it does indeed play a role in the simplest case, where f belongs to $\mathscr{R}$ itself, where the corresponding systems have finite dimensional state spaces.

13. THE RIEMANN–ROCH THEOREM APPLIED TO FINITE DIMENSIONAL SYSTEMS

The realizations constructed in Sec. 12 involve meromorphic differential forms on Riemann surfaces. The classical Riemann–Roch theorem[11] is the basic structural result about such geometric objects.

I will now show how the case of a finite dimensional state space (where the symbol is a rational function) can be treated in this spirit.

Theorem 13.1: Let $f: t \to f(t)$ be a C^∞ function of the real variable t that satisfies a linear, ordinary differential equation with constant coefficients. Then

(a) The Laplace transform

$$\mathscr{L}(f) = \int_0^\infty e^{-st} f(t)dt$$

exists and is a rational function of s, i.e., of the form

$$\mathscr{L}(f) = \frac{N(s)}{D(s)}, \tag{13.1}$$

where N and D are polynomials such that

N and D have no factors in common. (13.2)
(They are then said to be *coprime*.)

(b) The realizations constructed in Secs. 11 and 12 in terms of functions of t and meromorphic one-forms, respectively, are isomorphic under Laplace transform.

(c) The state space X for the realization constructed by means of the Laplace transform in Sec. 12 consists of all rational one-forms of the form

$$\omega = \frac{P(s)}{D(s)}ds, \tag{13.3}$$

where $s \to P(s)$ is a polynomial on s such that

$$\text{degree } P < \text{degree } D. \tag{13.4}$$

(d) The dimension of X, i.e., the Macmillan degree of the system, is equal to the degree of the polynomial $D(s)$.

Proof: These facts are well-known properties of the Laplace transform.[19]

My goal now is to identify the state space X with one of the vector spaces that occur in the Riemann–Roch theorem. Let S be the compact Riemann surface obtained from adding the point at infinity to the complex plane with complex variable s.

Now, let us recall certain concepts from classical Riemann surface theory. Let S be an arbitrary Riemann surface, i.e., a connected, one-dimensional complex manifold. Let F be a meromorphic function on S. At a point $p \in S$, with analytic coordinate z valid in a neighborhood of p with $z(p) = 0$, F is said to be of *order n at p* if

$$F = z^n F',$$

where F' is analytic about p, $F'(p) \neq 0$.

We then define

$$\mu(F,p) = n. \tag{13.5}$$

With p held fixed, the map $p \to \mu(F, p)$ is a *valuation* of the field of meromorphic functions.

Similarly, if ω is a one-form which is meromorphic in a neighborhood of p, and $\omega = F\,dz$, define

$$\mu(\omega, p) = \mu(F, p).$$

Now, let us return to the case where S is a Riemann surface of genus zero, its complex numbers parametrized by "s", with the point at infinity added on. Let ω be meromorphic one-form such that

$$\omega(\mu, \infty) \geqslant -1. \tag{13.6}$$

If ω is of the form

$$= (N/D)\,ds \tag{13.7}$$

with coprime polynomials N, D, then (13.6) says that ω is either analytic at ∞ or has a simple pole. Since the one-form ds has a pole of order two at ∞,

$$ds = -s^2 d(s^{-1}).$$

N/D must vanish to at least the first order, with

$$\text{degree } N < \text{degree } D. \tag{13.8}$$

Thus,

$$T(s) = N/D \tag{13.9}$$

is the transfer function of a finite-dimensional linear system.

Theorem 13.2: Let X be the set of meromorphic one-forms which satisfy the following conditions:

$$\mu(\omega', \infty) \geqslant -1, \tag{13.10}$$

$$\mu(\omega', p) \geqslant \mu(\omega, p) \tag{13.11}$$

for all $p \in S$. Then, X consists of all one-forms of the form

$$\omega'(P/D)\,ds, \tag{13.12}$$

where P is a polynomial such that

$$\text{degree } P \leqslant \text{degree } D. \tag{13.13}$$

Proof: Suppose

$$\omega' = F\,ds,$$

with F a rational function. Condition (13.7) implies that DF has no pole at finite r, i.e., is a polynomial P. Condition (13.10) implies that this polynomial is of order $< P$. Q.E.D.

Let S continue as the compact Riemann surface of genus zero. Let $\omega = Tds = (N/D)\,ds$ be a meromorphic one-form on S such that degree $N <$ degree D. Then,

$$\nu(\omega, \infty) \geqslant -1.$$

We assume familiarity with the classical Riemann–Roch theorem as described by Weyl.[11] Let $s_1,\ldots,s_m$ be the poles of ω at finite points of S. Let δ be the following divisor:

$$\delta = (\infty)^{-1}(s_1)^{\nu(\omega, s_1)}\cdots(s_m)^{\nu(\omega, s_m)}. \tag{13.14}$$

Then, δ agrees with the divisor of ω at finite points, but may not agree at infinity. We tie it down at ∞. Let $X(\delta)$ be the meromorphic differential one-forms which are "multiples" of δ, in the sense that

$$\nu(\omega, p) \geqslant \delta(p) \tag{13.15}$$

at all $p \in S$. [Thus, a divisor is a mapping of $S \to$(integers), which is equal to zero for all but a finite number of p's. The classical notation (13.14) for δ indicates the points, in this case $p = \infty, s_1,\ldots,s_m$, at which δ is nonzero.]

δ^{-1} is the divisor defined as follows:

$$\delta^{-1}(p) = -\delta(p), \quad \text{for all } p \in S,$$

or

$$\delta^{-1} = (\infty)^1(s_1)^{-\nu(\omega, s)}\cdots(s_m)^{-\nu(\omega, s_m)}. \tag{13.16}$$

Let

$$X(\delta^{-1}) = \text{ set of meromorphic functions } f \text{ such that } \nu(F, p) \geqslant \delta^{-1}(p) \quad \text{for all } p \in S.$$

General Riemann–Roch formula

$$\dim(X)(\delta^{-1}) - \dim(X(\delta)) = (\text{degree } \delta) + 1 - (\text{genus } S). \tag{13.17}$$

[Degree δ = sum of $\delta(p)$, $p \in S$.]

Formula (13.17) holds for an arbitrary divisor δ on an arbitrary, compact Riemann surface S. If we take δ as defined by (13.13), notice that δ^{-1} is "positive," in the sense that $\delta^{-1}(p) \geqslant 0$ for all $p \in S$.

Thus,

$$X(\delta^{-1}) = 0,$$

since any meromorphic function which has no poles is a constant. The condition $\delta^{-1}(\infty) = 1$ forces this constant to be zero. Also, degree $\delta = -$ degree $D - 1$, genus $S = 0$. Then, the Riemann–Roch formula takes the special form

$$\dim X(\delta) = \text{degree } D. \tag{13.18}$$

We can sum up as follows.

Theorem 13.3: Let E be the complex line-bundle on the $P_1(\mathbb{C})$ whose dimension is the degree of D. Then, the state space of the scalar input–output system whose transfer function is $T(s) = N(s)/D(s)$ is the space $\Gamma(E)$ of holomorphic cross sections of E. The dimension of this realization, i.e., the minimal dimension of the state space, is thus "naturally" computed using the Riemann–Roch formula.

14. EMBEDDING OF THE "DYNAMIC" OPERATOR OF THE STANDARD TIME-DOMAIN STATE-SPACE REALIZATION IN A LIE ALGEBRA VIA THE PICARD–VESSIOT–INFELD–HULL FACTORIZATION

Recall that we have defined a state-space realization of a (scalar input–output) linear system

$$y(t) = \int_0^t f(t-\tau)\mu(\tau)\,d\tau \tag{14.1}$$

as a system of linear differential equations of the form

$$\frac{dx}{dt} = Ax + Bu, \tag{14.2}$$

such that the solution of (14.2) with initial conditions $x(0) = 0$ is given by the integral operator (14.1).

Now, A is a linear operator $X \to X$ on the state vector space. It is of obvious interest from the Lie theory–Special Functions point of view[2,24] to know when A can be embedded in a larger Lie algebra $\mathcal{G}$ of operators in the vector space X which acts *irreducibly*. One might also require certain conditions relating $\mathcal{G}$, B, and C, which I will not attempt to formulate here.

In previous work[24] I have followed up suggestions by Infeld and Hull[27] relating Special Functions as solutions of second order, ordinary linear differential equations and factorizations of these operators into first order operators. The Lie algebra generated by these first order operators seems to be important in the study of the mathematical structure of these special functions. I will now investigate, in a preliminary way, how the Infeld–Hull structure affects the standard time-domain realizations constructed in Sec. 11.

In Sec. 11 we have chosen the state space X to be a space of C^∞ functions on the interval $0 \leqslant t < \infty$. Now we choose it more algebraically as a subspace of a Picard–Vessiot differential field extension of the field of rational functions of the complex variable t.

Let

$$\mathbb{C}[t]$$

be the integral domain of polynomials in the variable t. (We now use standard notation of algebra.)

Let $\mathbb{C}(t)$ be the quotient field, i.e., the field of rational functions.

Let

$$d_t : \mathbb{C}(t) \to \mathbb{C}(t)$$

be the usual derivative operator. It defines

$$(\mathbb{C}(t), d_t)$$

as a differential field.[9-10] $\mathbb{C}[t]$ is, of course, a subdifferential algebra.

Let

$$D : \mathbb{C}(t) \to \mathbb{C}(t)$$

be a linear differential operator with coefficients in $\mathbb{C}(t)$, i.e., D is of the form

$$D = a_n d_t^n + \cdots + a_0, \tag{14.3}$$

with $a_n, \ldots, a_0 \in \mathbb{C}(t)$.

Let $\mathcal{P}(D)$ be the Picard–Vessiot extension of $\mathbb{C}(t)$ obtained by adjoining to $\mathbb{C}(t)$ all the solutions of

$$D(f) = 0. \tag{14.4}$$

Thus, $\mathcal{P}(D)$ is the smallest *differential field* containing $\mathbb{C}(t)$ and all solutions of (14.4). [More classically, $\mathcal{P}(D)$ can be realized as a space of locally analytic functions defined on certain regions of the complex t plane.]

Now, let

$$X = \text{ set of all } f \in \mathcal{P}(D) \text{ such that } f \text{ is analytic in a neighborhood of } t = 0. \tag{14.5}$$

X is then a differential subalgebra of $\mathcal{P}(D)$. Let U and Y be the complex numbers considered as *input* and *output* spaces. As in Sec. 11, let

$$C : X \to Y,$$

$$C(f) = f(0), \quad \text{for } f \in X, \tag{14.6}$$

$$A : X \to X,$$

$$A = d_t. \tag{14.7}$$

Choosing one $f \in X$ defines a system of the form (14.5), with

$$B(u) = uf, \quad \text{for } u \in U \equiv \mathbb{C}. \tag{14.8}$$

Thus, we obtain an input–output system of the form (14.5).

Now, the operator D factors into a product of first order operators in $\mathcal{P}(D)$. Of course, finding such a factorization explicitly enables one to solve $D = 0$ by integrations, so it is in practice rarely possible to find them. However, there may be subdifferential fields $\mathcal{F}$ of $P(D)$ such that D takes some simple form in $\mathcal{F}$. Here is one such possibility.

Definition: The subdifferential field $\mathcal{F}$ of $\mathcal{P}(D)$ is said to be an *Infeld–Hull extension of* $(\mathbb{C}(z), D)$ if there are first order differential operators

$$D_1, \ldots, D_n$$

and a zeroth order operator D_0, with coefficients in $\mathcal{F}$, such that

$$D = D_1 \cdots D_n + D_0. \tag{14.9}$$

The Lie algebra $\mathcal{L}$ of operators on $\mathcal{F}$ generated by the operators $D_0, D_1, \ldots, D_n$ is called the *Lie algebra associated with the Infeld–Hull factorization (14.9).*

In the paper by Infeld and Hull[27] and in Vilenkin's treatise[3] one finds many examples of such factorizations. Some of the associated Lie algebras can be readily calculated. They seem to be infinite-dimensional Lie algebras of a relatively simple algebraic structure.

The operator which defines the "dynamics" of the system A is now realized as a differential operator, in fact just d_t. Thus, given a $u \in X$ such that

$$Du = 0, \tag{14.10}$$

we can associate with it the input–output system (14.6)–(14.8), and the Lie algebra of differential operators generated by A and $\mathcal{L}(D)$. In certain favorable cases, A will in fact belong to $\mathcal{L}(D)$.

Let us now compute certain examples.

Example 1: The harmonic oscillator

$$D = d_t^2 - t^2 + \lambda, \tag{14.11}$$

$\lambda \in \mathbb{C}$.

Set

$$D_1 = d_t + t,$$

$$D_2 = d_t - t, \tag{14.12}$$

$$D_1 D_2 = (d_t + t)(d_t - t),$$
$$= d_t^2 + t d_t - d_t t - t^2$$
$$= d_t^2 - t^2 - 1.$$

Hence,

$$D = D_1 D_2 + D_0, \tag{14.13}$$

with

$$D_0 = 1 + \lambda, \tag{14.14}$$

$$[D_1, D_2] = -1$$
$$= -(1 + \lambda)^{-1} D_0, \tag{14.15}$$

$$A = \tfrac{1}{2}(D_1 - D_2). \tag{14.16}$$

These formulas describe the situation. $\mathscr{L}(D)$ is the Heisenberg group, and the dynamic operator A belongs to it.

Of course, the linear subspace of X generated by $u, Au, A^2u, \ldots$, which determines the *minimal* (or *controllable*) realization, and is calculable by the familiar annihilation–creation operator formalism associated with the harmonic oscillators in quantum mechanics.

Example 2: The Bessel equation.

Suppose

$$D_1 = (d_t + nt^{-1}),$$

$$D_2 = d_t - (n-1)t^{-1} \quad (n \text{ is a complex number}), \tag{14.17}$$

$$D_1 D_2 = d_t^2 + t^{-1} d_t - n^2 t^{-2} \tag{14.18}$$
$$\equiv \text{the Bessel operator.}$$

The Lie algebra $\mathscr{L}$ generated by D_1 and D_2 is computed in Ref. 28. It is an infinite-dimensional, but graded, Lie algebra which is in some sense an infinite-dimensional generalization of the Heisenberg algebra. It can be expressed[28] in terms of the universal enveloping algebra of the Lie algebra $\mathscr{G}$ of the group G of rigid motions of the two-dimensional Euclidean plane, which is a three-dimensional solvable Lie group. As explained in Ref. 28, the element A of $\mathscr{L}$ is identified with an element of $\mathscr{G}$, in which is the underlying value (in terms of this formalism) for the appearance of the Bessel functions as matrix elements[3] of representations of G.

In Refs. 3 and 28, the Legendre and Whittaker functions are treated in parallel with these two examples. The only qualitative difference is that the field $\mathscr{F}$ in which the second order operator D factors in the "Infeld–Hull" form is a quadratic extension of the rational field $\mathbb{C}(t)$, whereas in the two simpler examples treated above it is $\mathbb{C}(t)$ itself.

These examples suggest the introduction of some further concepts, which we now briefly present in preparation for further research.

15. WEYL AND SCHLESINGER SYSTEMS

Return to the case where the kernel f of the input–output relations is an element of the Titchmarsh algebra, where the latter is extended to include all Lebesgue measurable functions

$$f: [0, \infty) \to \mathbb{C},$$

which are *locally integrable* with respect to Lebesgue measure. Such functions determine distributions, in the sense of Schwartz, in the usual way. Let us say that the input–output relation

$$u \to \int_0^t f(t - \tau) u(\tau) d\tau$$

are of *Weyl type* if f, as a distribution, satisfies a linear, ordinary differential equation with *polynomial* coefficients. Let us say that it forces a *Schlesinger system* if the distribution satisfies a linear, ordinary differential equation whose coefficients belong to a field of algebraic functions in the complex variable t of *finite* genus.

The further study of these systems, and their relations with Lie groups, will appear in later publications.

[1] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, 1927).
[2] L. Schlesinger, *Handbuch der linearen Differentialglechungen* (Leipzig, 1895–1898).
[3] N. J. Vilenkin, *Special Functions and the Theory of Group Representations* (American Mathematical Society, Providence, RI, 1968).
[4] B. D. O. Anderson and J. Moore, *Linear Optimal Control* (Prentice-Hall, Englewood Cliffs, N.J., 1971).
[5] R. Brockett, *Finite Dimensional Linear Systems* (Wiley, New York, 1970).
[6] M. Wonham, *Linear Multivariable Control* (Springer, Berlin, 1974).
[7] E. Picard, *Traité d'Analyse* (Gauthier-Villars, Paris, 1928), Vol. 3, Chap. 7.
[8] E. Vessiot, "Sur les integrations des équations différentielles lineaires," Am. Sci. Ecole Norm. **131**, 192–280 (1892).
[9] F. Kolchin, *Differential Algebra and Algebraic Groups* (Academic, New York, 1973).
[10] I. Kaplansky, *Introduction to Differential Algebra* (Herman, Paris, 1957).
[11] H. Weyl, *The Concept of a Riemann Surface* (Addison-Wesley, Reading, MA, 1964).
[12] J. Baras, R. Brockett, and P. Fuhrmann, "State space models for infinite dimensional linear systems," IEEE Trans. Auto. Control **19**, 693–700 (1974).
[13] J. Baras and R. Brockett, "H^2-functions and infinite dimensional realization theory," SIAM J. Control **13**, 221–241 (1975).
[14] R. Brockett and R. Fuhrmann, SIAM J. Control **14**, 107–119 (1976).
[15] P. Hartman, *Ordinary Differential Equations* (Wiley, New York, 1964).
[16] K. Yosida, *Functional Analysis*, 6th ed. (Springer, Berlin, 1980).
[17] J. Mikusinski, *Operational Calculus* (Pergamon, London, 1959).
[18] R. Hermann, "Linear systems and the differential field of Heaviside operators," Syst. Control Lett. **1**, 211–215 (1981).
[19] G. Doetsch, *Introduction to the Theory and Application of the Laplace Transformation* (Springer, Berlin, 1974).
[20] J. Dollard and D. C. Friedman, *Product Integration* (Addison-Wesley, Reading, MA, 1979).
[21] E. G. C. Poole, *Linear Differential Equations* (Oxford U. P., London, 1936).
[22] F. Baldassarri and B. Dwork, "On second order linear differential equations with algebraic solutions," Am. J. Math. **101**, 42–76 (1979).
[23] F. Oberhettinger and L. Bodii, *Tables of Laplace Transforms* (Springer, Berlin, 1973).
[24] R. Hermann and C. Martin, "Applications of algebraic geometry to system theory, Part VI; Infinite dimensional linear systems and properties of analytic functions," *Proceedings of the 1976 Ames (NASA) Conference on*

Geometric Control (Math. Sci. Press, Brookline, MA, 1977).
[25]R. Hermann, "Analytic continuation of group representations, III," Commun. Math. Phys. **3**, 75–97 (1966).
[26]R. Hermann, "The Gell-Mann formula for representations of semisimple groups," Commun. Math. Phys. **2**, 155–164 (1966).
[27]L. Infeld and T. E. Hull, Rev. Mod. Phys. **123**, 21 (1951).
[28]R. Hermann, "Infeld-Hull factorization, Galois-Picard-Vessiot theory for differential operators," J. Math. Phys. **22**, 1163 (1981).

New similarity solutions for the Ernst equations with electromagnetic fields

Elliot Fischer
Institute for Defense Analyses,[a] *Alexandria, Virginia 22311*

(Received 2 February 1982; accepted for publication 12 March 1982)

A new class of exact similarity solutions is found for the Ernst equations with electromagnetic fields. The original coupled nonlinear partial differential equations are reduced to a system of coupled nonlinear ordinary differential equations. The reduced system is solvable in a manner identical to previous similarity solutions found by Kaliappan and Lakshmanan. These solutions may be considered the extension of the Curzon solution (static, uncharged) to the stationary, charged solution.

PACS numbers: 02.30.Jr

I. INTRODUCTION

Similarity variables are often used to reduce partial differential equations in two independent variables to ordinary differential equations. The similarity form can be found systematically using either the differential geometric techniques of Harrison and Estabrook[1] or the classical techniques of Bluman and Cole.[2]

Recently, similarity variables have been employed to find new exact solutions of the Einstein and Einstein–Maxwell equations. Fischer[3] found two different similarity forms for the Ernst equation

$$(\mathrm{Re}\,\epsilon)\vec{\nabla}^2\epsilon = \vec{\nabla}\epsilon\cdot\vec{\nabla}\epsilon, \tag{1}$$

where ϵ is a single complex function depending only on the cylindrical variables ρ and z and $\vec{\nabla}^2$ is the axisymmetric Laplacian. The two similarity forms

$$\epsilon = \epsilon(\rho/z) \tag{2a}$$

and

$$\epsilon = \epsilon(\rho^2 + z^2) \tag{2b}$$

both reduce (1) to ordinary differential equations which are solvable in terms of elementary functions and quadratures. More recently, Kaliappan and Lakshmanan[4] (referred to hereafter as K and L) used the similarity variable ρ/z in (2a) to find exact solutions to the Ernst form of the coupled Einstein–Maxwell equations:

$$(\xi\xi^* + \eta\eta^* - 1)\vec{\nabla}^2\xi = 2\vec{\nabla}\xi\cdot(\xi^*\vec{\nabla}\xi + \eta^*\vec{\nabla}\eta), \tag{3a}$$

$$(\xi\xi^* + \eta\eta^* - 1)\vec{\nabla}^2\eta = 2\vec{\nabla}\eta\cdot(\xi^*\vec{\nabla}\xi + \eta^*\vec{\nabla}\eta), \tag{3b}$$

where ξ and η are complex functions related to the complex Ernst function ϵ and the complex Maxwell function Ψ for the electromagnetic field by

$$\epsilon = (\xi - 1)/(\xi + 1),\ \Psi = \eta/(\xi + 1). \tag{4}$$

It is the purpose of this paper to show that the system (3) is exactly solvable in terms of the similarity variable of (2b),

$$\zeta = \rho^2 + z^2. \tag{5}$$

Remarkably, these solutions have the same functional form as those found by K and L, differing only in the similarity variable used. They may be considered as the extension of the Curzon solution[5] to a more general physical situation.

[a] Present Address: Bell Laboratories, Whippany, New Jersey 07981.

2. SIMILARITY FORM

For future convenience, we use the similarity variable

$$r = \ln[(1 + \zeta^2)^{1/2} - \zeta] = \ln[(1 + (\rho^2 + z^2)^2)^{1/2} - (\rho^2 + z^2)] \tag{6a}$$

or

$$\zeta = -\sinh r. \tag{6b}$$

The solution procedure is analogous to that of K and L. Equation (3) reduces to a set of coupled ordinary differential equations:

$$(\xi\xi^* + \eta\eta^* - 1)(4r\xi'' + 6\xi') = 8r(\xi^*\xi'^2 + \eta^*\xi'\eta'), \tag{7a}$$

$$(\xi\xi^* + \eta\eta^* - 1)(4r\eta'' + 6\eta') = 8r(\xi^*\xi'\eta' + \eta^*\eta'^2), \tag{7b}$$

where the prime denotes differentiation with respect to r. We now let

$$\xi = \kappa \exp\left[i\int\tau dr\right], \tag{8a}$$

$$n = p \exp\left[i\int q dr\right], \tag{8b}$$

where κ, τ, p and q are real functions of r. Substituting (8) into (7) and equating real and imaginary parts results in four coupled ordinary differential equations:

$$(\kappa^2 + p^2 - 1)[4r(\kappa'' - \kappa\tau^2) + 6\kappa'] = 8r[p(\kappa'p' - \kappa\tau pq) + \kappa(\kappa'^2 - \kappa^2\tau^2)], \tag{9a}$$

$$(\kappa^2 + p^2 - 1)[4r(p'' - pq^2) + 6p'] = 8r[\kappa(\kappa'p' - \kappa\tau pq) + p(p'^2 - p^2q^2)], \tag{9b}$$

$$(\kappa^2 - p^2 - 1)[4r(2\kappa'\tau + \kappa\tau') + 6\kappa\tau] = 8r[\kappa(2\kappa\kappa'\tau) + p(\kappa\tau p' + \kappa' pq)], \tag{9c}$$

$$(\kappa^2 + p^2 - 1)[4r(2p'q + pq') + 6pq] = 8r[p(2pp'q) + \kappa(\kappa\tau p' + \kappa' pq)]. \tag{9d}$$

Multiplying (9c) by κ and (9d) by p and adding yields

$$\frac{d(\kappa^2\tau + p^2q)}{\kappa^2\tau + p^2q} + \frac{3}{2}\frac{dr}{r} = \frac{2(2\kappa\kappa' + 2pp')}{\kappa^2 + p^2 - 1}, \tag{10}$$

which integrates to

$$\kappa^2\tau + p^2q = Cr^{-3/2}(\kappa^2 + p^2 - 1)^2, \tag{11}$$

where C is a constant of integration.

Similarly, (9c)$\times\kappa$ − (9d)$\times p$ yields

 0022-2488/82/071295-02$02.50

$$d\left[\frac{r^{3/2}(\kappa^2\tau - p^2 q)}{\kappa^2 + p^2 - 1}\right] = Cd(\kappa^2 - p^2), \tag{12}$$

which integrates to

$$\kappa^2\tau - p^2 q = r^{-3/2}[C(\kappa^2 - p^2) + D](\kappa^2 + p^2 - 1), \tag{13}$$

where D is a second integration constant. From (11) and (13) we obtain

$$\tau = (1/2\kappa^2)r^{-3/2}(\kappa^2 + p^2 - 1)[C(2\kappa^2 - 1) + D], \tag{14a}$$

$$q = (1/2p^2)r^{-3/2}(\kappa^2 + p^2 - 1)[C(2p^2 - 1) - D], \tag{14b}$$

which are identical in form to K and L equations (16) and (17) except for the additional factor of $r^{-3/2}$.

Using (14), (9a) may be rewritten as

$$(\kappa^2 + p^2 - 1)(4r\kappa'' + 6\kappa') - 8r(pp'\kappa' + \kappa\kappa'^2)$$
$$= r^{-2}(\kappa^2 + p^2 - 1)^3\kappa^{-3}C^2\left[\left(1 - \frac{D}{C}\right)^2 - 4\kappa^4\right], \tag{15}$$

which may be rewritten as

$$\frac{r^{3/2}\kappa'}{(\kappa^2 + p^2 - 1)}\frac{d}{dr}\left[\frac{r^{3/2}\kappa'}{(\kappa^2 + p^2 - 1)}\right]$$
$$= -\frac{1}{8}(C - D)^2\frac{d}{dr}\kappa^{-2} - C^2\frac{d}{dr}\frac{\kappa^2}{2}. \tag{16}$$

Equation (16) may be integrated to obtain

$$\frac{1}{2}\left(\frac{r^{3/2}\kappa'}{\kappa^2 + p^2 - 1}\right)^2 = \left[-\frac{\kappa^{-2}}{8}(C - D)^2 - \frac{C^2\kappa^2}{2} + E\right], \tag{17}$$

where E is the third integration constant.

Analogously, (9b) may be integrated to

$$\frac{1}{2}\left(\frac{r^{3/2}p'}{\kappa^2 + p^2 - 1}\right)^2 = \left[-\frac{p^{-2}(C + D)^2}{8} - \frac{C^2p^2}{2} + F\right], \tag{18}$$

where F is the fourth integration constant. Equations (16) and (17) are identical to K and L equations (21) and (22), again with the exception of the factor of $r^{3/2}$.

At this point, K and L solve their equations (21) and (22) for $(\kappa^2 + p^2 - 1)$ and equate; the same can be done for Eqs. (17) and (18) here, as follows. Rewrite (17) as

$$\kappa'^2\Big/\left[\frac{(C - D)^2}{8\kappa^2} + \frac{C^2\kappa^2}{2} - E\right] = -2r^{-3}(\kappa^2 + p^2 - 1)^2, \tag{19a}$$

$$p'^2\Big/\left[\frac{(C + D)^2}{8p^2} - \frac{C^2p^2}{2} - F\right] = -2r^{-3}(\kappa^2 + p^2 - 1)^2. \tag{19b}$$

Equating the left-hand sides of (19a) and (19b) and making the change of variable

$$\kappa^2 = x, \quad p^2 = y \tag{20}$$

yields K and L equation (24). Thus, the functional forms for κ and p here and in K and L are identical and their results may now be used directly. The equation resulting from the procedure described above may be integrated to obtain y as a function of x with a fifth constant G [K and L (25a)], and substituted into (17) to obtain a first order equation for x, which may be solved to finally obtain κ as

$$\kappa = [(x_1t^2 + x_2)/(t^2 + 1)]^{1/2}, \tag{21}$$

where

$$t = -\alpha + A\coth(A\nu r + \tfrac{1}{2}\ln B) \tag{22}$$

and x_1, x_2, ν, α, and A are new constants that are combinations of the constants $C - G$ given by K and L equations (27) and (30).

Finally, to obtain the Ernst potentials ξ and η, the expression $\exp[i\int\tau dr]$ is required. To find this, use (14a) and the fact that

$$(x + y - 1) = -\frac{dt}{dr}\frac{r^{3/2}}{C(t^2 + 1)}, \tag{23}$$

so that the factor of $r^{-3/2}$ in (14a) cancels with the factor of $r^{3/2}$ in (23) to give the same expression as (K and L) for $\int\tau dr$, and

$$\exp\left[i\int\tau dr\right] = \exp i\arctan\left[\frac{(\sqrt{x_2} - \sqrt{x_1})t}{(\sqrt{x_2} + \sqrt{x_1})t^2}\right]e^{iH}, \tag{24}$$

where the integration constant H could be identified with a NUT parameter. A second NUT parameter is produced by the term $\exp i\int qdr$.

Thus, amazingly, the functional expressions for ξ and η here are identical with those found in K and L, the difference in the solutions being the different similarity variables. These new solutions, like those of K and L, contain eight constants and are finite everywhere, therefore representing the external gravitational and electromagnetic fields of an axisymmetric, stationary, charged, rotating body. Furthermore, the similarity variable $\rho^2 + z^2$ is the same as that found in the Curzon solution,[5] and these new solutions may be thought of as the extension of the Curzon solution (static, uncharged) to the more general situation.

[1] B. K. Harrison and F. B. Estabrook, J. Math. Phys. **12**, 653 (1971).
[2] G. W. Bluman and J. D. Cole, *Similarity Methods for Differential Equations* (Springer, New York, 1974).
[3] E. Fischer, J. Phys. A **13**, L81 (1980).
[4] P. Kaliappan and M. Lakshmanan, J. Math. Phys. **22**, 2447 (1981).
[5] H. E. J. Curzon, Proc. London Math. Soc. **23**, (1924).

Quantization as a consequence of the symmetry group: An approach to geometric quantization

Víctor Aldaya and José A. de Azcárraga
Departamento de Física Teórica, Universidad de Valencia, Facultad de C. Físicas, Burjasot (Valencia), Spain

(Received 4 November 1981; accepted for publication 26 February 1982)

A method is proposed to obtain the dynamics of a system which only makes use of the group law. It incorporates many features of the traditional geometric quantization program as well as the possibility of obtaining the classical dynamics: The classical or quantum character of the theory is related to the choice of the group, avoiding thus the need of quantizing preexisting classical systems and providing a group connection between the quantum and classical systems, i.e., the classical limit. The method is applied to the free-particle dynamics and the harmonic oscillator.

PACS numbers: 02.40. + m, 02.20. + b, 03.65. − w

I. INTRODUCTION

In a previous paper[1] we have briefly outlined a method of quantization which follows closely the traditional method of geometric quantization of Souriau, Kostant, and others.[2–4] The underlying rationale, however, is rather different: Instead of searching for quantizations of previously defined classical systems, the new approach tends to build directly, and without any ingredient other than a group law, the dynamical quantum systems. Thus, their quantum character is already determined by the symmetry group. The method is based on the close relation which exists among the spatial and dynamical properties of a system and its symmetry group, as well as of the slightly different character which the symmetry groups of classical and quantum systems present. Clearly, the difficulty of the procedure is the determination of the group for the case of interacting systems; in this respect, it fares no better than the conventional Kostant–Souriau (KS) theory. Nevertheless, our method may be applied directly in configuration space, thus avoiding the problem of characterizing the manifold of solutions[5] of the classical system in order to quantize it.

From a general point of view, our procedure is similar to that followed in classical mechanics when building dynamical systems associated with a Lie group. There, the symplectic manifolds appear contained in the group coalgebra, where the group operates in a natural way. In the present case, the manifold sought is a contact manifold which appears as a submanifold of a Lie group $\tilde{G}$ (of trivial cohomology). $\tilde{G}$ will be obtained as a central extension of a Lie group G—the classical group in the usual meaning of the word—by a one-dimensional Lie group $T = \mathrm{U}(1)$, and the contact form will be defined by the (vertical) component along T of the canonical 1-form defined on $\tilde{G}$. The space–time variables need not to be initially known; the group structure itself will determine the corresponding identification.

This method of quantization also provides a way of defining the classical limit of the quantum theory. The classical limit—in its Hamilton–Jacobi version—is recuperated using $T = \mathbb{R}$ instead of $T = \mathrm{U}(1)$ in the general theory, i.e., opening the compact group U(1). In this way, the classical limit does not involve a limiting procedure, but is associated to a different invariance group.

To facilitate the comparison with the Kostant–Souriau theory, a short summary of results is given in Sec. II. Section III is devoted to the development of the above-mentioned program, and in Secs. IV and VI two applications are presented, the case of the free particle and that of the harmonic oscillator; the classical limit is described in Sec. V. Finally, some comments about the extension of the theory to the relativistic case are made in Sec. VII.

II. THE FORMALISM OF GEOMETRIC QUANTIZATION OF SOURIAU AND KOSTANT: AN OUTLINE

In this paragraph we review briefly the formalism of geometric quantization of Souriau and Kostant[2–4] in the form which is more useful for our purposes (see also Ref. 6). The first part (till Definition 2.7) deals with the formalism of prequantization. The rest will give a brief resumé of the quantization procedure which still might be completed with the introduction of the bundle of half-forms, which allow the definition of an integration volume for the polarized manifolds. Proofs will be omitted, the reader being referred to the quoted literature and specially to Refs. 2 and 3.

Definition 2.1: *A quantum manifold* P is a $(2n + 1)$-dimensional differentiable manifold such that

(QM a) There exists on P a 1-form Θ such that the pair (P,Θ) defines a contact structure, i.e., the (contact) form is of constant class $2n + 1$,

$$\dim(\mathrm{rad}\Theta \cap \mathrm{rad}d\Theta) = 0. \tag{2.1a}$$

(QM b) The group U(1) acts on P effectively, i.e.,

$$z_P\xi = \xi \Rightarrow z = 1 \quad \forall\ \xi \in P \quad [z \in U(1)] \tag{2.1b}$$

and *the* dynamical vector field associated with Θ is the vector field Ξ_P, which generates the action of U(1) on P.

In fact, since Θ is a contact form [(QM a)] it follows that there is a unique vector field X such that

$$i_X\Theta = 1, \quad i_X\, d\Theta = 0; \tag{2.2}$$

(QM b) then asserts that $X \equiv \Xi_P$. Ξ_P thus plays the role of the Hamiltonian vector field for P and the quotient $(P/\Xi_P \equiv S,\ d\Theta = \omega)$ is then a symplectic manifold.

An alternative way of defining P is established by means of the following

 0022-2488/82/071297-09$02.50

Proposition 2.1: *A quantum manifold* is a principal bundle $(P \xrightarrow{\pi} S, \mathrm{U}(1))$ and a connection form $\Theta \in \Lambda^1(P)$ such that the curvature 2-form $\Omega = \mathrm{curv}\,\Theta$ defines on $S \approx P/U(1)$ a symplectic structure.

Definition 2.2: Let (P,Θ) and (P',Θ') be two quantum manifolds, and let $F: P \to P'$ be a diffeomorphism. F is called a *quantomorphism* if

$$F^*\Theta' = \Theta, \tag{2.3a}$$

$$F(z_P \xi) = z_{P'} F(\xi) \quad \forall z \in \mathrm{U}(1), \quad \forall \xi \in P. \tag{2.3b}$$

As a function, F is thus a U(1) function [(2.3b)] so that the diagram

$$\begin{array}{ccc} P & \xrightarrow{F} & P' \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi'} \\ S & \xrightarrow{f} & S \end{array},$$

in which f is the associated symplectomorphism, is commutative. When $S = S'$ and $f = I_S$, the *quantizations* are said to be *equivalent*.

An alternative way of defining F is given by the following:

Proposition 2.2: A *quantomorphism* is an isomorphism between the principal bundles $(P \xrightarrow{\pi} S, \mathrm{U}(1); \Theta)$ and $(P' \xrightarrow{\pi'} S', \mathrm{U}(1); \Theta')$ (compatible with the connections Θ, Θ').

Theorem 2.1 (on the existence of a quantization): Let (S,ω) be a symplectic manifold. Then, for a quantization (P,Θ) of (S,ω) to exist, it is necessary and sufficient that ω be of integer cohomology class, $[\omega] \in \mathrm{Im}\epsilon$, where ϵ is the canonical map $\epsilon: H^2(S,\mathbb{Z}) \to H^2(S,\mathbb{R})$.[7]

The possible quantizations of S are classified according to the following:

Proposition 2.3: The set of inequivalent quantizations of a quantizable manifold (S,ω) are parametrized by the group $H^1(S,U(1))$ or, equivalently, by the number of different characters (or of irreducible unitary representations) of the first homotopy group $\pi^1(S)$ of S.

Definition 2.3: Let $(P \xrightarrow{\pi} S, \Theta)$ be a quantum manifold, and F a quantomorphism over the base S. An *infinitesimal quantomorphism* $\tilde{X}$ is a vector field on P, $\tilde{X} \in \mathscr{X}(P)$, such that

$$L_{\tilde{X}}\Theta = 0. \tag{2.4}$$

This expression is the infinitesimal counterpart of $F^*\Theta = \Theta$ [cf. (2.3a)], $\tilde{X}$ being the vector field associated with F.

Proposition 2.4: The necessary and sufficient condition for $\tilde{X}$ to be an infinitesimal quantomorphism is the existence of a real function f on S, $f \in \mathscr{F}(S)$, such that

$$i_{\tilde{X}}\Theta = f, \quad i_{\tilde{X}}\, d\Theta = -df. \tag{2.5}$$

Indeed, to prove this proposition, it is sufficient to realize that $i_{\tilde{X}}\Theta \equiv f$ defines a function on S, for which it is sufficient to check that

$$L_{\Xi_P}(i_{\tilde{X}}\Theta) = 0. \quad \blacksquare \tag{2.6}$$

Theorem 2.2 (definition of *quantum lifting*): Given a globally Hamiltonian vector field X on S, there is a unique vector field $\tilde{X}$ on P, π-projectable on X, such that $\tilde{X}$ is an infinitesimal quantomorphism.

Proof: since X is globally Hamiltonian, there exists a function $f \in \mathscr{F}(S)$ such that $i_X\omega = -df$. The theorem now follows from Proposition 2.4 ■

We shall denote henceforth by $\tilde{X}_f$ the infinitesimal quantomorphism determined by f. A useful expression for $\tilde{X}_f$ is provided by

$$\tilde{X}_f = X_f + [f - \Theta(X_f)]\Xi_P, \tag{2.7}$$

where $i_{X_f}\omega = -df$; for $f = \text{const}$, $\tilde{X}_f = (\text{const})\cdot\Xi_P$. Using local coordinates $\{q^i, p_i\}$ for S, $\{q^i, p_i, z\}$ for P, we have

$$\Theta = p_i dq^i + \frac{dz}{iz}, \quad \omega = d p_i \wedge dq^i, \quad \Xi_P = iz\frac{\partial}{\partial z} \tag{2.8}$$

and

$$\tilde{X}_f = \frac{\partial f}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial f}{\partial q^i}\frac{\partial}{\partial p_i} + \left(f - p_i\frac{\partial f}{\partial p_i}\right)\Xi_P. \tag{2.9}$$

Theorem 2.3: The mapping $f \to \tilde{X}_f$ is a Lie algebra isomorphism in the sense that

$$[\tilde{X}_f, \tilde{X}_g] = -\tilde{X}_{\{f,g\}}, \tag{2.10}$$

where the Poisson bracket is defined by

$$\{f,g\} = -\omega(X_f,X_g) = L_{X_g}\cdot f = \left(\frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i}\right).$$

Definition 2.4: Let $(P \xrightarrow{\pi} S, \Theta)$ be a quantum manifold. The *space of wavefunctions* $\mathscr{H}(P)$ is the space of the complex U(1) functions on P, i.e.,

$$\psi: P \to \mathbb{C} / \psi(z_P\xi) = z_{\mathbb{C}}\psi(\xi) \qquad \forall \xi \in P, \ \forall z \in U(1). \tag{2.11}$$

From the expression of Ξ [(2.8)] and the definition (2.11) it follows immediately that

$$\Xi.\psi = i\psi, \tag{2.12}$$

where the dot means derivation.

Definition 2.5: Given a function $f \in \mathscr{F}(S)$, the operator which acts linearly on $\mathscr{H}(P)$ as defined by

$$\hat{f}\psi \equiv -i\tilde{X}_f.\psi = -iX_f.\psi + [f - \Theta(X_f)]\psi \tag{2.13}$$

may be called the *prequantized operator* associated with the magnitude f. The definition of the usual quantum operator requires a *polarization*; see Definitions 2.7 and 2.9.

Taking advantage of the canonical Liouville measure which exists on any simplectic manifold and of the fact that, because of (2.11), $\psi^*(\xi)\Phi(\xi)$ is defined on S, a Hermitian scalar product may be defined on $\mathscr{H}(P)$ by means of the following

Definition 2.6: Given $\psi, \Phi \in \mathscr{H}(P)$, the scalar product is given by

$$\langle \psi, \Phi \rangle \equiv \int_S \omega^n \psi^*(\xi)\Phi(\xi). \tag{2.14}$$

Theorem 2.4: The *quantization map* $f \to \hat{f}$ satisfies:

(a) $\hat{f}$ is a Hermitian operator;

(b) $\hat{}$ is a regular linear map of the space $\mathscr{F}(S)$ of real

functions on S on the space of Hermitian operators on $\mathscr{H}(P)$;

(c) $\hat{1} = 1|_{\mathscr{H}(P)}$;

(d) $[\hat{f},\hat{g}] = i(\{f,g\})\hat{}$.

The above ends what is usually called *prequantization*. To obtain an irreducible representation of the Lie algebra of quantum operators ($\hat{p} = -i\partial/\partial q$ and $\hat{q} = q$, for instance) and thus complete the quantization procedure, the notion of polarization is introduced according to the following definition:

Definition 2.7: Let $(P \xrightarrow{\pi} S, \Theta)$ be a quantum manifold. A (closed) submanifold $Q \subset P$ on which Θ is zero, $\Theta|_Q = 0$, is called (Souriau) a *Planck manifold*. The pair (Q,i), where i is the injection of Q into P ($i^*\Theta \equiv \Theta|_Q$), is thus a (nongeneral) integral manifold of the Pfaff form Θ.

Taking into account that $\Theta(\Xi_P) = 1 \neq 0$, the following is clear:

Proposition 2.5: The restriction $\pi|_Q$ of π to Q is a local diffeomorphism.

When $\pi|_Q$ is a global diffeomorphism, $\pi(Q) \equiv N$ is a submanifold of S diffeomorphic to Q on which ω is zero, $\omega|_N = 0$. Hence, N is an isotropic $[\omega(X,X') = 0 \ \forall \ X,X' \in T(N)]$ submanifold of S. It is frequent to define a *polarization* as an isotropic differential system on S; since the system is integrable, this is equivalent to giving a foliation by the Planck (integral) manifolds.

Once N has been defined, the (horizontal) vector fields of $\mathscr{X}(N)$ may be lifted to $\mathscr{X}(P)$ by means of the connection Θ.

Definition 2.8: Given $X \in \mathscr{X}(N) \subset \mathscr{X}(S)$, these exists a unique $\widehat{X} \in \mathscr{X}(P)$, the *isotropic lifting*, such that $\Theta(\widehat{X}) = 0$ (horizontality) and $\pi^T \circ \widehat{X} = X \circ \pi$. Explicitly, $\widehat{X}$ is given by

$$\widehat{X} = X - \Theta(X)\Xi_P. \tag{2.15}$$

The fields $\widehat{X}$ constitute an integrable differential system (they close into an algebra) which reproduces the Planck manifold $Q \subset P$.

The reduction of the algebra of the prequantized operators (*quantization*) is now achieved by imposing on the $\psi \in \mathscr{H}(P)$ the so-called *Planck condition*:

Definition 2.9: The polarized Hilbert space $\mathscr{H}_Q(P)$ is given by the $\psi \in \mathscr{H}(P)$ such that

$$\widehat{X}.\psi = 0 \quad \forall \ \widehat{X} \in \mathscr{X}(Q). \tag{2.16}$$

When the Planck manifold is a maximal isotropic manifold, (2.16) may be rewritten in the form $[\widehat{X}_f = \tilde{X}_f - fiz\partial/\partial z]$

$$\hat{f}_j \psi = f_j \psi, \quad j = 1,\ldots,\tfrac{1}{2}(\dim S), \tag{2.17}$$

where $i_{X_{f_j}}\omega = -df_j$ and $X_{f_j} \in \mathscr{X}(N)$; physically $\{\hat{f}_j\}$ corresponds to a *maximal set of commuting observables*.

Note: The Planck constant $\hbar \equiv h/2\pi$ has been put equal to 1 throughout, but it is simple to include it explicitly in the corresponding formulae. In that case, $\Theta = p_i dq^i + \hbar dz/iz$, $\Xi = (i/\hbar)z\partial/\partial z$, $\Xi.\psi = (i/\hbar)\psi$ [(2.8),(2.12)]; $\hat{f} = -i\hbar\widehat{X}_f$ [(2.13)]; $[\hat{f},\hat{g}] = i\hbar(\{f,g\})\hat{}$, $\widehat{X}_f = X_f - (1/\hbar)fiz\partial/\partial z$, and so on.

III. QUANTIZATION AND SYMMETRY: A GROUP THEORETICAL APPROACH

In the symmetry group approach to geometric quantization, two group entities play a special role: the canonical 1-form on a Lie group and the symplectic cohomology of a Lie group. We briefly review here the relevant concepts.

A. Canonical 1-form on a Lie group

Let G be a Lie group and $\mathscr{G}$ its associated Lie algebra, considered as the tangent space $T_e(G)$ at the unity $e \in G$. Let L_g (R_g) be the left (right) translations, $L_g : s \to gs$ ($R_g : s \to sg$). $X \in \mathscr{X}(G)$ is called a *left (right)-invariant vector field* if $L_g^T \circ X \circ L_g^{-1} = X$, i.e., if $L_g^T X_s = X_{gs}$ ($R_g^T \circ X \circ R_g^{-1} = X$, i.e., $R_g^T X_s = X_{sg}$) $\forall g,s \in G$. Calling $\mathscr{X}^L(G)$ ($\mathscr{X}^R(G)$) the vector space of the left (right)-invariant vector fields X_G^L (X_G^R), the above is symbolically written $(L_g)_* X_G^L = X_G^L$ ($(R_g)_* X_G^R = X_G^R$). It is simple to see that $\mathscr{X}^L(G) \approx \mathscr{X}^R(G) \approx T_e(G)$ and that $[X_G^L, X_G^R] = 0$.

The *canonical left (right) 1-form* $\theta^{L,R}$ on G [8] is the left (right)-invariant $\mathscr{G}$-valued 1-form uniquely determined by the condition $\theta^{L,R}(Z_G^{L,R}) = Z$ ($\forall \ Z_G^{L,R} \in \mathscr{X}^{L,R}(G)$). If $\{Z_i\}$ is a basis of $\mathscr{G}$, then

$$\theta^{L,R} = \theta_G^{L,R\,(i)} \circ Z_i, \tag{3.1}$$

where $\{\theta_G^{L,R\,(i)}\}$ is the dual basis for the ordinary (i.e., $\mathbb{R}$-valued) left (right)-invariant 1-form on G, i.e., such that $(L_g)^* \theta_s^{L\,(i)} = \theta_{g^{-1}s}^{L\,(i)}$ ($(R_g)^* \theta_s^{R\,(i)} = \theta_{sg^{-1}}^{R\,(i)}$), and $\theta_G^{L,R\,(i)}(Z_{G(j)}^{L,R}) = \delta_j^i$. This last expression,

$$\theta_g^{L,R\,(i)}(Z_{g(j)}^{L,R}) = \delta_j^i, \tag{3.2}$$

will be often used in Sec. IV.

For the case of θ^L, one finds that the canonical left-invariant 1-form transforms as $\mathrm{ad}(g^{-1})$ under right translations R_g and has zero Lie derivative with respect to right-invariant vector fields, $L_{X_G^R}\theta^L = 0$. Thus, the left-invariant $\theta^{L\,(i)}$ are left-translated in a way which preserves the notion of incidence on left-invariant vector fields, and are kept invariant in the customary "physical" meaning (zero Lie derivative) under right-invariant vector fields. Analogous considerations can be made for θ^R through substituting right for left.

B. Symplectic cohomology on a Lie group[2]

Let G be a Lie group and $g \mapsto g_{\mathscr{G}^*} \equiv D_{\mathscr{G}^*}(g)$ the (coadjoint) representation on the coalgebra $\mathscr{G}^*$ of G.

A *symplectic cocycle* of G is a $\mathscr{G}^*$-valued cocycle γ such that its derivative $D(\gamma)(e)$ at the unity $e \in G$ is antisymmetric, i.e., γ is a differentiable mapping $\gamma : G \to \mathscr{G}^*$ such that[9]

$$\gamma(g'g) = \gamma(g') + g'_{\mathscr{G}^*}\gamma(g) \quad \forall \ g',g \in G, \tag{3.3a}$$

where $g_{\mathscr{G}^*}$ is the representation of g on $\mathscr{G}^*$, and $D(\gamma)(e)\,Z$ satisfies[10]

$$[D(\gamma)(e).Z](Z') \equiv \Omega(Z',Z) = -\Omega(Z',Z) \quad \forall \ Z,Z' \in \mathscr{G}. \tag{3.3b}$$

Two 1-cocycles γ,γ' are cohomologous if their difference is a coboundary,

$$\Delta_{\mu_0} : g \mapsto g_{\mathscr{G}^*}\mu_0 - \mu_0, \tag{3.4}$$

where $\mu_0 \in \mathscr{G}^*$. Equation (3.4) is an equivalence relation between symplectic cocycles which defines the (vector) *space of symplectic cohomology* $H_s(G,\mathscr{G}^*)$ of G. $H_s(G,\mathscr{G}^*)$ is clear-

ly a vector subspace of $H(G,\mathscr{G}^*)$, the cohomology space of the coadjoint representation. Two cohomologous symplectic 1-cocycles γ and γ' define two cohomologous Ω and Ω'; indeed, from $\gamma'(g)=\gamma(g)+\Delta_{\mu_0}(g)$, one gets

$$\Omega'=\Omega+\mu_0\cdot\mathrm{ad},\quad \Omega'(Z,Z')=\Omega(Z,Z')+\mu_0\cdot[Z,Z']. \tag{3.5}$$

C. Symplectic cohomology and central extensions: $H^2(G,\mathrm{U}(1))$

From the definition of Ω as a mapping $\Omega:\mathscr{G}\to\mathscr{G}^*$, it may be derived that

$$\Omega([Z,Z'])-Z_{\mathscr{G}}\cdot\Omega(Z')-Z'_{\mathscr{G}}\cdot\Omega(Z)=0, \tag{3.6}$$

where $Z_{\mathscr{G}}\cdot$ indicates the (coadjoint) representation of $\mathscr{G}$ on $\mathscr{G}^*$.[11] Equation (3.6) written as $\partial\Omega(Z,Z')=0$ indicates that Ω may be identified with a 1-$\mathscr{G}^*$-cocycle of the Lie algebra $\mathscr{G}$ valued on the $\mathscr{G}$-module $\mathscr{G}^*$ ($\mathscr{G}$ acts on $\mathscr{G}^*$ through $Z_{\mathscr{G}}\cdot$) in accordance with the standard definition[12] of cohomology in Lie algebras. It is interesting now to relate $H_s(G,\mathscr{G}^*)$ and the group $H^2(G,\mathrm{U}(1))$ [or $H^2_0(G,\mathbb{R})$], which characterizes the possible central extensions of G by U(1) (or $\mathbb{R}$.[13,14] A two-cocycle $\xi(g',g)$ of G (*called* an exponent of G since in the case of U(1) it appears as such) is a function $\xi:G\times G\to\mathbb{R}$ which satisfies

$$\xi(g_1,g_2)+\xi(g_1g_2,g_3)=\xi(g_1,g_2g_3)+\xi(g_2,g_3),$$
$$\xi(e,e)=0. \tag{3.7a}$$

A 2-coboundary (*a trivial exponent*) is a 2-cocycle of the form

$$\xi(g_1,g_2)=\zeta(g_1,g_2)-\zeta(g_1)-\zeta(g_2), \tag{3.7b}$$

where $\zeta:G\to\mathbb{R}$ satisfies $\zeta(e)=0$; Eqs. (3.7) determine $H^2_0(G,\mathrm{U}(1))$.

Taking now a canonical coordinate system (G may always be supposed simply connected by going to the covering group) or making use of the formal group structure,[15] we may represent by the same letter the corresponding elements of G and $\mathscr{G}$. Then, the 2-cocycle in $H^2_0(\mathscr{G},\mathbb{R})$ associated with (3.7a) may be written

$$\begin{aligned}\Sigma(g_1,g_2)&=\lim_{t\to0}(1/t^2)[\xi(tg_1,tg_2)+\xi([tg_1]^{-1},[tg_2]^{-1})\\&\quad+\xi([tg_1][tg_2],[tg_1]^{-1}[tg_2]^{-1})]\\&=-\Sigma(g_2,g_1),\end{aligned} \tag{3.8a}$$

and is called the *infinitesimal exponent*.[16] The Σ associated with a coboundary is determined[14] by the real-valued linear function Λ on $\mathscr{G}$,

$$\Lambda(g)=\lim_{t\to0}(1/t)\zeta(tg),\quad \Sigma(g_1,g_2)=\Lambda([g_1,g_2]). \tag{3.8b}$$

Because of the linearity of Λ, the trivial $\Sigma(g_1,g_2)$ in (3.8b) may be written as $\mu_0\cdot[g_1,g_2]$. Then, cohomologous infinitesimal exponents Σ' and Σ are related by

$$\Sigma'(g_1,g_2)=\Sigma(g_1,g_2)+\mu_0\cdot[g_1,g_2] \tag{3.8c}$$

[compare with (3.5)]. There is thus a one-to-one correspondence between the classes of cohomologous symplectic 1-cocycles of G, of $\mathscr{G}^*$-1-cocycles of $\mathscr{G}$, of real valued 2-forms Ω on $\mathscr{G}\times\mathscr{G}$ and of real valued 2-cocycles of $\mathscr{G}$: $H_s(G,\mathscr{G}^*)\simeq H_s(\mathscr{G},\mathscr{G}^*)\simeq H^2_0(\mathscr{G},\mathbb{R})$. Since G is simply connected, $H_s(G,\mathscr{G}^*)\simeq H^2_0(G,\mathbb{R})\simeq H^2_0(G,\mathrm{U}(1))$.

D. Two examples

Let E be a vector space of dimension $2n$ considered as an additive group, $X''=X'+X$. The central extensions of E by U(1) are characterized by $H^2_0(E,\mathrm{U}(1))$, i.e., by the different symplectic forms which may be defined on E. The possible extensions $\tilde{E}_{(m)}$ are given by the group law

$$\begin{pmatrix}X'\\z'\end{pmatrix}*\begin{pmatrix}X\\z\end{pmatrix}=\begin{pmatrix}X'+X\\z'\cdot ze^{i\omega(X,X')/2}\end{pmatrix}. \tag{3.9}$$

The symplectic forms ω are given by $m\xi(X',X)$ or $m\Omega(X',X)\equiv m\Sigma(X',X)$, $m\in\mathbb{R}^+$ (the fact that here $\xi=\Sigma$ is accidental, and is not true in general). The group defined by (3.9) is called the Weyl-Heisenberg group.

The extensions of E by $\mathbb{R}$ are similarly obtained by substituting $\theta'+\theta+\frac{1}{2}m\Omega(X',X)$ for the lower line of (3.9).

E. Definition of a quantum manifold

Let $\tilde{G}$ be a Lie group of trivial symplectic cohomology obtained as a central extension of a group G (of nontrivial symplectic cohomology) by a Lie group T of dimension one [T will be taken as U(1); the case of $\mathbb{R}$ will be relevant for the classical limit of the theory and will be discussed in Sec. V). $\tilde{G}(T,\tilde{G}/T=G)$ is a principal bundle over the base G.

Let θ^L be the canonical left 1-form on $\tilde{G}$ and θ^L_v its vertical component as defined by the bundle projection $\pi:\tilde{G}\to G$. The characteristic module of θ^L_v, $\mathscr{C}_{\theta^L_v}\equiv\mathrm{rad}\,\theta^L_v\cap\mathrm{rad}\,d\theta^L_v$ defines an integrable differential system on $\tilde{G}$ (this is immediately seen by using the identity $i_{[X,Y]}=L_Xi_Y-i_XL_Y$).

Proposition 3.1: Given $\tilde{G}$ as before, the quotient $P=\tilde{G}/\mathscr{C}_{\theta^L_v}$ together with $\theta^L_v/\mathscr{C}_{\theta^L_v}$ as connection form has the structure of a quantum manifold.

Proof: Because θ^L_v is the vertical component of θ and Ξ the generator of T, we have $i_\Xi\theta^L_v=1$ since θ^L is the canonical form on $\tilde{G}$. From the properties of the canonical form we obtain $L_{X^L_{(i)}}\theta^{L(0)}=-C^0_{ij}\theta^{L(j)}$ and in particular (since $X^L_{(0)}$ is central), $L_{X^L_{(0)}}\theta^{L(0)}=-C^0_{0j}\theta^{L(j)}$, i.e., $L_\Xi\theta^L_v=0$. From this expression we get $i_\Xi d\theta^L_v=0$. Thus, $\tilde{G}/\mathscr{C}_{\theta^L_v}$ will be a quantum manifold (Definition 2.1) if θ_v satisfies the additional condition of being of class $2s+1$ on $\tilde{G}$ since then it will define a contact structure on the quotient $\tilde{G}/\mathscr{C}_{\theta^L_v}$. θ^L_v will be of class $2s+1$ if $\theta^L_v\wedge(d\theta^L_v)^s\neq0$ and $(d\theta^L)^{2s+1}=0$. The last condition is obtained by noticing that on a canonical chart $d\theta_v$ may be identified with $d\gamma$, interpreting γ as an ordinary 1-form on G. The rank of $d\theta_v$ is thus equal to the dimension $2s$ of one of the symplectic orbits of G on $\mathscr{G}^*$ defined by γ, which is even.[2] The condition $\theta^L_v\wedge(d\theta^L_v)^s\neq0$ now follows from $(d\theta^L_v)^s\neq0$ and $\theta^L_v(\Xi)\neq0$, $\Xi\in\mathrm{rad}(d\theta_v)$. ∎

Definition 3.1: The vector fields belonging to the quotient of $\mathscr{X}^R(\tilde{G})$ by $\mathscr{C}_{\theta^L_v}$ are called (*pre*) *quantum operators*.

Definition 3.2: A *polarization* is a subspace of $\mathscr{H}^L(\tilde{G})$ containing $\mathscr{C}_{\theta^L_v}$ which on the base of the quantum manifold P generates an isotropic submanifold for the sympletic form $d(\theta^L_v/\mathscr{C}_{\theta^L_v})$.

Definition 3.3: The space of the wavefunctions is the space of all the T-equivariant functions defined on P.

As we shall see, the above simple definitions—all based

on the group of the system—are directed towards recovering the more traditional formalism of geometric quantization, where P is a contact manifold. Nevertheless, since both the quantum operators and the polarizations appear here as derivations canonically defined on $\tilde{G}$, the proposed quantization process may be directly realized on the group $\tilde{G}$ itself despite the fact that $\tilde{G}$ may not to be a manifold of dimension $2n+1$.[17] This will be illustrated in Secs. IV and VI.

It will become apparent from the examples that not only the vertical component θ^L_v and θ^L will have a physical meaning but that the other components will reproduce, on the trajectories of $\mathscr{C}_{\theta^L_v}$, the Newton equations of motion.

IV. THE FREE QUANTUM NONRELATIVISTIC PARTICLE

In this case $\tilde{G}$ is $\tilde{G}_{(m)}$, the central extension of the Galilean group by $T \equiv U(1)$; the parameter m (the mass) characterizing the group extension[14] is an element of $H^2(G,U(1))$ [or $H_s(G,\mathscr{S}^*)$]. The group law of $\tilde{G}_{(m)}$ is obtained from that of G, $g'' = g' * g$,

$$\begin{aligned} &B'' = B' + B, \quad B \in \mathbb{R}, \\ &\mathbf{A}'' = \mathbf{A}' + \mathbf{A} + \mathbf{V}'B, \quad \mathbf{A},\mathbf{V} \in \mathbb{R}^3, \\ &\mathbf{V}'' = \mathbf{V}' + \mathbf{V}, \end{aligned} \tag{4.1}$$

and is given by $\tilde{g}'' = \tilde{g}' * \tilde{g}$, where $\tilde{g} \equiv (g,\zeta)$ and

$$(g'',\zeta'') = (g' * g, \zeta'\zeta \exp[i\xi_{(m)}(g',g)]). \tag{4.2}$$

The rotations have been omitted in (4.1) and (4.2) for the sake of simplicity. In any case, for spin zero there is no contribution from $SU(2,C)$ since this group is semisimple and has trivial cohomology.

Let us take in (4.2) the cocycle

$$\xi_{(m)}(g',g) = m[\mathbf{A}'\cdot\mathbf{V} + B(\mathbf{V}'\cdot\mathbf{V} + \tfrac{1}{2}\mathbf{V}'^2)]. \tag{4.3}$$

The right-invariant vector fields ($X^R = \partial g''/\partial g'|_{g'=e}$) are given by

$$\begin{aligned} &X^R_B = \frac{\partial}{\partial B}, \quad X^R_{\mathbf{A}} = \frac{\partial}{\partial \mathbf{A}} + m\mathbf{V}\Xi, \\ &X^R_{\mathbf{V}} = B\frac{\partial}{\partial \mathbf{A}} + \frac{\partial}{\partial \mathbf{V}} + mB\mathbf{V}\Xi, \\ &X^R_\zeta \equiv \Xi = i\zeta\frac{\partial}{\partial \zeta}, \end{aligned} \tag{4.4R}$$

and the left-invariant vector fields ($X^L = \partial g''/\partial g|_{g=e}$) are given by

$$\begin{aligned} &X^L_B = \frac{\partial}{\partial B} + \mathbf{V}\frac{\partial}{\partial \mathbf{A}} + \tfrac{1}{2}m\mathbf{V}^2\Xi, \\ &X^L_{\mathbf{A}} = \frac{\partial}{\partial \mathbf{A}}, \quad X^L_{\mathbf{V}} = \frac{\partial}{\partial \mathbf{V}} + m\mathbf{A}\Xi, \\ &X^L_\zeta \equiv \Xi = i\zeta\frac{\partial}{\partial \zeta}; \end{aligned} \tag{4.4L}$$

it is immediately verified that $[X^L, X^R] = 0$. The canonical left 1-form on $\tilde{G}_{(m)}$, $\theta^L = \theta^{L(i)}_{\tilde{G}} \circ X_{(i)}$, is now easily computed from (3.2), with the result

$$\begin{aligned} &\theta^{L(B)} = dB, \quad \theta^{L(\mathbf{A})} = d\mathbf{A} - \mathbf{V}dB, \\ &\theta^{L(\mathbf{V})} = d\mathbf{V}, \end{aligned} \tag{4.5a}$$

$$\theta^{L(\zeta)} \equiv \theta^L_{\mathbf{v}} \equiv \Theta^L = -m\mathbf{A}d\mathbf{V} - \tfrac{1}{2}m\mathbf{V}^2 dB + \frac{d\zeta}{i\zeta}; \tag{4.5b}$$

and again one may check that the defining properties (Sec. IIIA)

$$\begin{aligned} &L_{X^R}\theta^L = 0, \quad L_{X^{L(i)}}\theta^{L(j)} = -C^j_{ik}\theta^{L(k)} \\ &\qquad \forall\ i,j = B,\mathbf{A},\mathbf{V},\zeta. \end{aligned} \tag{4.6}$$

[which correspond to $(R_g)^*\theta^L = \mathrm{ad}(g^{-1})\theta^L$ and $(L_g)^*\theta^L = \theta^L$] are fulfilled. We shall simply denote $\Theta^L = \Theta$ in what follows.

The characteristic module $\mathscr{C}_\Theta$ is determined by the fields satisfying $i_X\Theta = 0$, $i_X d\Theta = 0$. This characterizes uniquely a differential system composed by one vector field,

$$X = \frac{\partial}{\partial B} + \mathbf{V}\frac{\partial}{\partial \mathbf{A}} + \tfrac{1}{2}m\mathbf{V}^2\Xi \equiv X^L_B. \tag{4.7}$$

Its integral manifold is determined by the equations (of motion)

$$\frac{dB}{ds} = 1, \quad \frac{d\mathbf{A}}{ds} = \mathbf{V}, \quad \frac{d\mathbf{V}}{ds} = 0, \quad \frac{d\zeta}{ds} = \tfrac{1}{2}m\mathbf{V}^2(i\zeta), \tag{4.8}$$

where s is the integration parameter and is given by

$$B = s, \quad \mathbf{A} = \mathbf{V}s + \mathbf{K}, \quad \mathbf{V} = \mathbf{V}_0 \equiv \mathbf{P}/m, \quad \zeta = ze^{i(\mathbf{P}^2/2m)s}, \tag{4.9}$$

where $\mathbf{K}$, $\mathbf{V}_0$, and z are the integration constants (and the one corresponding to B has been set equal to zero). Equations (4.9) suggest the *identifications*

$$B = t, \quad \mathbf{A} = \mathbf{x} \tag{4.10}$$

which, together with $\mathbf{V} = \mathbf{v} \equiv \mathbf{p}/m$ *define* the evolution space variables in terms of the group parameters.[18] The manifold of solutions is obtained by means of the change of variables (the generalized "Hamilton–Jacobi" transformation):

$$\mathbf{p} = \mathbf{P}, \quad \mathbf{x} = \frac{\mathbf{P}}{m}t + \mathbf{K}, \quad \zeta = ze^{i(\mathbf{P}^2/2m)t}, \tag{4.11}$$

where $\mathbf{K}$ ("center of mass"), $\mathbf{P}$ (momentum), and z are the integration constants of the equations of motion which parametrize the *quantum manifold* $P = \tilde{G}/\mathscr{C}_\Theta$ (Proposition 3.1). In terms of the variables $\mathbf{x}$, t, and $\mathbf{p}$, Θ is written

$$\Theta = -\mathbf{x}d\mathbf{p} - \frac{\mathbf{p}^2}{2m}dt + \frac{d\zeta}{i\zeta}. \tag{4.12}$$

Passing Θ to the quotient P, (4.12) takes the form

$$\Theta_P = -\mathbf{K}d\mathbf{P} + dz/iz \tag{4.13}$$

in terms of the integration constants $\mathbf{K}$, $\mathbf{P}$, and z; clearly, (P,Θ_P) is a contact manifold.

The *pre*quantum operators associated with $\mathbf{K}$, $\mathbf{P}$ are given (Definition 3.1) by $-(i/m)X^R_{\mathbf{V}}$, $-iX^R_{\mathbf{A}}$ defined on P (in fact, $\hat{\mathbf{K}} = (i/m)X^R_{\mathbf{V}}$, $\hat{P} = -iX^R_{\mathbf{A}}$, but the difference of sign may be incorporated into the definition of the fields). From (4.4R) and (4.10), (4.11), we obtain, for the vector fields on P,

$$X^R_{\mathbf{A}} = \frac{\partial}{\partial \mathbf{K}} + \mathbf{P}iz\frac{\partial}{\partial z}, \quad X^R_{\mathbf{V}} = m\frac{\partial}{\partial \mathbf{P}}. \tag{4.14}$$

Thus, the operators act on the wave functions (Definition

3.3) through the expressions[19]

$$\hat{\mathbf{P}}\psi = \left(-i\frac{\partial}{\partial \mathbf{K}} + \mathbf{P} \right)\psi, \quad \hat{\mathbf{K}}\psi = i\frac{\partial \psi}{\partial \mathbf{P}}, \tag{4.15}$$

where the T-equivariance property (2.12) has been used. It is easy to check that (4.15) are the same expressions one would have obtained from (2.13) and (2.9) by taking $(\mathbf{K},\mathbf{P})$ as coordinates for S and $f = \mathbf{P},\mathbf{K}$.

The reduction of the representation is achieved by taking a polarization. Using the one defined by the left (Definition 3.2) vector fields X^L_B and $X^L_{\mathbf{A}}$,

$$X^L_{\mathbf{A}} = \frac{\partial}{\partial \mathbf{K}}, \tag{4.16}$$

the condition (2.16), $X^L_A \cdot \psi = 0$, gives $\psi \neq \psi(\mathbf{K})$ and thus (4.15) reads

$$\hat{\mathbf{P}}\cdot\psi = \mathbf{P}\psi, \quad \hat{\mathbf{K}}\cdot\psi = i\frac{\partial \psi}{\partial \mathbf{P}}; \tag{4.17}$$

(4.16) is also the polarization which is obtained from (2.15) by putting $X = \partial/\partial \mathbf{K}$.

These polarization conditions may be imposed directly on the evolution space (on $\tilde{G}_{(m)}$), without resorting to the manifold of solutions S or to the quotient P. In the same way, the quantum operators can be defined directly on the evolution space; thus the whole quantization process can be performed directly on $\tilde{G}_{(m)}$. To do this, we start with C-valued functions on $\tilde{G}_{(m)}$, $\psi = \psi(\mathbf{x},\mathbf{p},t,\zeta)$. The condition of U(1) function for ψ, $\Xi\cdot\psi = i\psi$ implies that ψ is of the form

$$\psi = \zeta \Phi(\mathbf{x},\mathbf{p},t). \tag{4.18}$$

The polarizations are now given directly by the left vector fields (4.4L). $X^L_{\mathbf{A}}\cdot\psi = 0$ now reads

$$\frac{\partial \psi}{\partial \mathbf{x}} = 0, \quad \psi = \zeta\varphi(\mathbf{p},t); \tag{4.19}$$

$\varphi(\mathbf{p},t)$ is the momentum space wavefunction. In this scheme, the condition $X^L_B\cdot\psi = 0$ imposed by the vector field of the time displacements of (4.4L) takes its full meaning:

$$X^L_B\cdot\psi = 0 \rightarrow i\frac{\partial \varphi}{\partial t} = \frac{\mathbf{p}^2}{2m}\varphi, \quad \psi = \zeta F(p) e^{-i(\mathbf{p}^2/2m)t}. \tag{4.20}$$

Equation (4.20) is the Schrödinger equation which selects the functions which are physical wavefunctions. It appears here as one more polarization (condition) on ψ compatible with $X^L_{\mathbf{A}}$.

The quantum operators are similarly obtained from (4.4R), and one gets

$$\hat{\mathbf{P}} \equiv -iX^R_{\mathbf{A}} \rightarrow \hat{\mathbf{P}}\varphi(\mathbf{p},t) = \mathbf{p}\varphi(p,t), \tag{4.21a}$$

$$\hat{\mathbf{K}} \equiv \frac{i}{m}X^R_{\mathbf{V}} \rightarrow \hat{K}\varphi(\mathbf{p},t) = \left(i\frac{\partial}{\partial \mathbf{p}} - \frac{\mathbf{p}}{m}t \right)\varphi(\mathbf{p},t) \tag{4.21b}$$

after factorization of the terms in ζ. Equation (4.21b) allows the identifications of $\hat{\mathbf{x}}$ with $i\partial/\partial \mathbf{p}$, since $\mathbf{K} = \mathbf{x} - (\mathbf{p}/m)t$.

A final comment on polarizations may be in order here. The "transverse" polarization $X^L_{\mathbf{V}}$, which also reduces the quantum representation, is not compatible with the Hamiltonian vector field ($[X_B, X_{\mathbf{V}}] = X_{\mathbf{A}}$) and thus does not lead to solutions of the Schrödinger equation. From the physical point of view, this means that the quantization theory based on the quantum group $\tilde{G}$ already dictates the appropriate polarization by specifying the dependence of the Hamiltonian function on $\mathbf{x}$ and $\mathbf{p}$ (here, $H = \frac{1}{2}m\mathbf{V}^2$).

Different cocycles $\xi(g',g)$ differing in a coboundary, on the other hand, may be used to define the *same* quantum group $\tilde{G}$, the central extension of G by U(1) [as is clear from the group extension theory; see (3.7)]. Clearly, the theory must be insensitive to the choice of a particular cocycle (with the same m), and this is indeed the case, although intermediate expressions may vary. We collect here some useful equivalent cocycles and the corresponding left and right canonical forms $\theta^{L,R}_{\mathbf{v}}$ on $\tilde{G}_{(m)}$, starting with the cocycle used in the text [Formulae (4.22) give, in this order, $\xi_{(m)}(g',g)$, Θ^R, and Θ^L],

$$m[\mathbf{A}'\mathbf{V} + B(\mathbf{V}\mathbf{V}' + \tfrac{1}{2}\mathbf{V}'^2)]; \quad -m\mathbf{V}d\mathbf{A} + \frac{d\zeta}{i\zeta};$$

$$-m\mathbf{A}d\mathbf{V} - \frac{m}{2}\mathbf{V}^2 dB + \frac{d\zeta}{i\zeta}, \tag{4.22a}$$

$$\frac{m}{2}[\mathbf{V}.\mathbf{A}' - \mathbf{A}\mathbf{V}' + \mathbf{V}\mathbf{V}'B];$$

$$-\frac{m}{2}[\mathbf{V}d\mathbf{A} - \mathbf{A}d\mathbf{V}] + \frac{d\zeta}{i\zeta};$$

$$-\frac{m}{2}[\mathbf{A}d\mathbf{V} - \mathbf{V}d\mathbf{A}] - \frac{m}{2}\mathbf{V}^2 dB + \frac{d\zeta}{i\zeta}, \tag{4.22b}$$

$$-m[\tfrac{1}{2}B\mathbf{V}'^2 + \mathbf{V}'\mathbf{A}]; \quad m\mathbf{A}d\mathbf{V} + \frac{d\zeta}{i\zeta};$$

$$m\mathbf{V}d\mathbf{A} - \tfrac{1}{2}m\mathbf{V}^2 dB + \frac{d\zeta}{i\zeta}, \tag{4.22c}$$

$$m[\mathbf{A}'\mathbf{V} - B'(\tfrac{1}{2}\mathbf{V}^2 + \mathbf{V}\mathbf{V}')];$$

$$-m\mathbf{V}d\mathbf{A} + d(\tfrac{1}{2}m\mathbf{V}^2 B) + \frac{d\zeta}{i\zeta};$$

$$-m\mathbf{A}d\mathbf{V} + mB\mathbf{V}d\mathbf{V} + \frac{d\zeta}{i\zeta}. \tag{4.22d}$$

V. THE CLASSICAL LIMIT

Let us now take $T = \mathbb{R}$, the additive group of real numbers, and let us now call $\overline{G}_{(m)}$ the central extension of the Galilei group by $\mathbb{R}$. The group law for $\overline{G}_{(m)}$ is given by (4.1) and

$$(g'',\theta'') = (g'*g,\ \theta' + \theta + \xi_{(m)}(g',g)), \tag{5.1}$$

which determines the composition law for arbitrary elements $(g,\theta) \in \overline{G}_{(m)}$ ($g \in G$, $\theta \in \mathbb{R}$) and replaces (4.2).

The procedure of Secs. III and IV may be followed now to determine a manifold P' which is a local chart (a vector bundle neighborhood) of the previously found manifold. The $\mathbb{R}$ functions $\psi(\mathbf{K},\mathbf{P},\theta)$ satisfy now the condition

$$\Xi\cdot\psi = 1, \quad \Xi \equiv \frac{\partial}{\partial \theta}, \tag{5.2}$$

so that, $\psi(\mathbf{K},\mathbf{P},\theta) = \Phi(\mathbf{K},\mathbf{P}) + \theta$. The polarization $X^L_{\mathbf{A}}\cdot\psi = 0$ gives $\Phi = \Phi(\mathbf{P})$, and for the operators $\hat{\mathbf{P}}$ and $\hat{\mathbf{K}}$ one finds

$$\hat{\mathbf{P}}\cdot\psi = \mathbf{P}, \quad \hat{\mathbf{K}}\cdot\psi = \frac{\partial \psi}{\partial \mathbf{P}}. \tag{5.3}$$

Again, the results may be directly obtained in configuration space, where $\psi = \psi(\mathbf{x},\mathbf{p},t,\theta)$. The constraints on ψ imposed by Ξ, $X^L_{\mathbf{A}} = \partial/\partial\mathbf{x}$ [(4.4L)] require that $\Phi = \Phi(\mathbf{p},t)$, and now $X^L_B \cdot \psi = 0$ gives

$$\frac{\partial \Phi}{\partial t} + \frac{\mathbf{p}^2}{2m} = \frac{\partial \Phi}{\partial t} + H = 0. \tag{5.4}$$

Identifying Φ with the Hamilton–Jacobi function S the theory obtained is thus the classical limit[20] of the quantum theory defined by $\tilde{G}_{(m)}$. Its group-theoretical interpretation is clear: the classical (Hamilton–Jacobi) theory is defined by $\bar{G}_{(m)}$ obtained from $\tilde{G}_{(m)}$ substituting $\mathbb{R}$ for U(1). Physically, one could say that the opening of the compact U(1) to $\mathbb{R}$ (taking a local chart) eliminates the quantum prescription and gives the associated classical theory.

As for operators in configuration space, they are obtained from (4.4R) with Ξ given by (5.2);

$$\begin{aligned}
\hat{\mathbf{P}} &\equiv X^R_{\mathbf{A}} = \frac{\partial}{\partial \mathbf{x}} + \mathbf{p}\frac{\partial}{\partial \theta}, \\
\hat{\mathbf{K}} &\equiv X^R_{\mathbf{V}} = t\frac{\partial}{\partial \mathbf{x}} + m\frac{\partial}{\partial \mathbf{p}} + t\mathbf{p}\frac{\partial}{\partial \theta},
\end{aligned} \tag{5.5}$$

and the acting on ψ reproduce (5.3).

VI. THE QUANTUM HARMONIC OSCILLATOR

We consider now the nontrivial example of the quantum oscillator. We shall not assume, as in the case of the free particle, anything about the space, time, or forces; all these quantities will be naturally defined by the quantum group once it is properly determined.

Let $\tilde{G}_{(m,\omega)}$ be the group defined by the following composition law ($\tilde{g}'' = \tilde{g}' * \tilde{g}$):

$$\begin{aligned}
&\lambda'' = \lambda' + \lambda, \quad \lambda \equiv \omega B \in \mathbb{R}, \\
&\mathbf{C}'' = \mathbf{C}' e^{-i\lambda} + \mathbf{C}, \quad \mathbf{C} \in \mathbb{C}^3, \ \mathbf{C}^+ = (\mathbf{C})^*, \\
&\mathbf{C}''^+ = \mathbf{C}'^+ e^{i\lambda} + \mathbf{C}^+, \\
&\zeta'' = \zeta'\zeta \exp[\tfrac{1}{2}i(i\mathbf{C}'\mathbf{C}^+ e^{-i\lambda} - i\mathbf{C}'^+ \mathbf{C} e^{i\lambda})].
\end{aligned} \tag{6.1}$$

This group is the central extension by U(1) of a group $G_{(\omega)}$ which contracts to the Galilei group[21] G when the frequency ω of the oscillator tends to zero (the rotations of G are again ignored). Indeed, the change of variables defined by

$$\begin{aligned}
\mathbf{C} &\equiv (m/\omega)^{1/2}(\omega \mathbf{A} + i\mathbf{V})/\sqrt{2}, \\
\mathbf{C}^+ &\equiv (m/\omega)^{1/2}(\omega \mathbf{A} - i\mathbf{V})/\sqrt{2},
\end{aligned} \tag{6.2}$$

where $\omega^2 m$ is the Hooke constant of the spring, allows us to rewrite (6.1) in the form

$$\begin{aligned}
B'' &= B' + B, \\
\mathbf{A}'' &= \mathbf{A}' \cos\omega B + \mathbf{A} + (\mathbf{V}'/m)\sin\omega B, \\
\mathbf{V}'' &= \mathbf{V}' \cos\omega B + \mathbf{V} - \omega \mathbf{A}' \sin\omega B, \\
\zeta'' &= \zeta'\zeta \exp\{i\tfrac{1}{2}m[\mathbf{A}'\mathbf{V}\cos\omega B - \mathbf{V}'\mathbf{A}\cos\omega B \\
&\quad + (\mathbf{V}\mathbf{V}'/\omega + \omega \mathbf{A}'\mathbf{A})\sin\omega B]\}.
\end{aligned} \tag{6.3}$$

In the zero force limit $\omega \to 0$, the first three equations of (6.3) reproduce (4.1), and the last one gives

$$\zeta'' = \zeta'\zeta \exp\{i\tfrac{1}{2}m[(\mathbf{V}\mathbf{A}' - \mathbf{V}'\mathbf{A}) + \mathbf{V}\mathbf{V}'B]\}, \tag{6.4}$$

i.e., the group law of $\tilde{G}_{(m)}$ where $\xi_{(m)}(g',g)$ is given in terms of the (Bargmann) cocycle (4.22b).

We now proceed with the quantization procedure which again will be based on the left canonical form. As mentioned before, this means that the left- and right-invariant vector fields define polarizations and operators, respectively. The left vector fields are given by

$$\begin{aligned}
X_\lambda &= \frac{\partial}{\partial \lambda} - i\mathbf{C}\frac{\partial}{\partial \mathbf{C}} + i\mathbf{C}^+\frac{\partial}{\partial \mathbf{C}^+}, \\
X_{\mathbf{C}} &= \frac{\partial}{\partial \mathbf{C}} - \frac{i}{2}\mathbf{C}i\zeta\frac{\partial}{\partial \zeta}, \\
X_{\mathbf{C}^+} &= \frac{\partial}{\partial \mathbf{C}^+} + \frac{i}{2}\mathbf{C}i\zeta\frac{\partial}{\partial \zeta}, \\
X_\zeta &= i\zeta\frac{\partial}{\partial \zeta},
\end{aligned} \tag{6.5L}$$

and the right ones by

$$\begin{aligned}
X_\lambda &= \frac{\partial}{\partial \lambda}, \\
X_{\mathbf{C}} &= e^{-i\lambda}\frac{\partial}{\partial \mathbf{C}} + \frac{i}{2}e^{-i\lambda}\mathbf{C}^+ i\zeta\frac{\partial}{\partial \zeta}, \\
X_{\mathbf{C}^+} &= e^{i\lambda}\frac{\partial}{\partial \mathbf{C}^+} - \frac{i}{2}e^{i\lambda}\mathbf{C}i\zeta\frac{\partial}{\partial \zeta}, \\
X_\zeta &= i\zeta\frac{\partial}{\partial \zeta}.
\end{aligned} \tag{6.5R}$$

The canonical left 1-form on $\tilde{G}_{(m,\omega)}$ is given by

$$\theta^{L(\zeta)} \equiv \Theta = \frac{i}{2}(\mathbf{C}^+ d\mathbf{C} - \mathbf{C}d\mathbf{C}^+) - \mathbf{C}\mathbf{C}^+ d\lambda + \frac{d\zeta}{i\zeta}, \tag{6.6a}$$

$$\theta^{L(\mathbf{C})} = d\mathbf{C} + i\mathbf{C}d\lambda, \tag{6.6b}$$

$$\theta^{L(\mathbf{C}^+)} = d\mathbf{C}^+ - i\mathbf{C}^+ d\lambda, \tag{6.6c}$$

$$\theta^{L(\lambda)} = d\lambda, \tag{6.6d}$$

and it may be seen that $L_{X^R}\theta^L = 0$.

The characteristic module $\mathscr{C}_\Theta$ is determined by the vector field

$$X = \frac{\partial}{\partial \lambda} - i\mathbf{C}\frac{\partial}{\partial \mathbf{C}} + i\mathbf{C}^+\frac{\partial}{\partial \mathbf{C}^+}; \tag{6.7}$$

the equations of motion are given by

$$\frac{d\mathbf{C}}{ds} = -i\mathbf{C}, \quad \frac{d\mathbf{C}^+}{ds} = i\mathbf{C}^+, \tag{6.8a}$$

$$\frac{d\lambda}{ds} = 1, \quad \frac{d\zeta}{ds} = 0, \tag{6.8b}$$

and thus the trajectories determined by the generator (6.7) of $\mathscr{C}_\Theta$ are

$$C = C_0 e^{-is}, \quad C^+ = C_0^+ e^{is}, \quad B = s, \quad \zeta = \zeta_0 \equiv z. \tag{6.9}$$

[Again, one finds that (6.8a) may be derived from the other components (6.6b) and (6.6c) of θ.]

On $P \equiv \tilde{G}_{(m,\omega)}/\mathscr{C}_\Theta$, the form Θ reduces to

$$\Theta_P = \frac{i}{2}(\mathbf{C}_0^+ d\mathbf{C}_0 - \mathbf{C}_0 d\mathbf{C}_0^+) + \frac{dz}{iz} \tag{6.10}$$

in terms of the constants of the motion $\mathbf{C}_0$, $\mathbf{C}_0^+$, and z, as it should. Nevertheless, we shall continue to work on the group $\tilde{G}_{(m,\omega)}$ to take advantage of keeping the time dependence as given by the identification

$$B = \lambda/\omega = t. \tag{6.11}$$

To determine the wavefunctions, we adopt as polarization the one generated by the subgroup $X_{\mathbf{C}}^L, X_\lambda^L$ [(6.5L)]. The Planck condition, together with the condition of the U(1) function for the ψ defined on $\tilde{G}_{(m,\omega)}$, $\psi = \psi(\mathbf{C},\mathbf{C}^+,\lambda,\zeta)$, give

$$\begin{aligned} &\Xi.\psi = i\psi \Rightarrow \psi = \zeta\Phi(\mathbf{C},\mathbf{C}^+,\lambda), \\ &X_{\mathbf{C}}^L.\psi = 0 \Rightarrow \frac{\partial\psi}{\partial\mathbf{C}} + \frac{\mathbf{C}^+}{2}\psi = 0 \\ &\qquad \Rightarrow \psi = \zeta\varphi(\mathbf{C}^+,\lambda)e^{-\mathbf{C}\cdot\mathbf{C}/2}, \\ &X_\lambda^L.\psi = 0 \Rightarrow \frac{\partial\varphi}{\partial\lambda} = -i\mathbf{C}^+\frac{\partial\varphi}{\partial\mathbf{C}^+}. \end{aligned} \tag{6.12}$$

This last equation {which in the definition of the group corresponds to the evolution of ψ [(6.11); cf. (4.20)]} is nothing but the evolution equation in the Bargmann–Fock–Segal picture[22] from which the usual description of the quantum harmonic oscillator may be recovered.[23] In particular, when the scalar product (Definition 2.6) is applied to the polarized functions,

$$\psi = \zeta\varphi(\mathbf{C}^+,\lambda)e^{-\mathbf{C}\cdot\mathbf{C}/2}, \tag{6.13}$$

it takes the Bargmann form

$$\langle\psi',\psi\rangle = \int_{\mathbb{R}^n} d^3C^+\,\varphi'^*(\mathbf{C}^+,\lambda)\varphi(\mathbf{C}^+,\lambda)e^{-|\mathbf{C}|^2}, \tag{6.14}$$

which corresponds to a scalar product for holomorphic functions φ with weight $\rho = \exp(-|\mathbf{C}|^2)$.

We now proceed to identify physically all the group parameters and describe the full contents of the canonical form on $\tilde{G}_{(m,\omega)}$. In terms of the parameters of $\tilde{G}_{(m)}$, the fields in (6.5R) and (6.5L) are written[24] as

$$\begin{aligned} X_B^R &= \frac{\partial}{\partial B}, \\ X_{\mathbf{A}}^R &= \cos\omega B\frac{\partial}{\partial\mathbf{A}} - \omega\sin\omega B\frac{\partial}{\partial\mathbf{V}} \\ &\quad + \frac{m}{2}[\omega\mathbf{A}\sin\omega B + \mathbf{V}\cos\omega B]\Xi, \\ X_{\mathbf{V}}^R &= \frac{\sin\omega B}{\omega}\frac{\partial}{\partial\mathbf{A}} + \cos\omega B\frac{\partial}{\partial\mathbf{V}} \\ &\quad + \frac{m}{2}\left[\mathbf{V}\frac{\sin\omega B}{\omega} - \mathbf{A}\cos\omega B\right]\Xi, \\ X_\zeta^R &= i\zeta\frac{\partial}{\partial\zeta} \equiv \Xi \end{aligned} \tag{6.15R}$$

and

$$\begin{aligned} X_B^L &= \frac{\partial}{\partial B} + \mathbf{V}\frac{\partial}{\partial\mathbf{A}} - \omega^2\mathbf{A}\frac{\partial}{\partial\mathbf{V}}, \\ X_{\mathbf{A}}^L &= \frac{\partial}{\partial\mathbf{A}} - \frac{m}{2}\mathbf{V}\Xi, \\ X_{\mathbf{V}}^L &= \frac{\partial}{\partial\mathbf{V}} + \frac{m}{2}\mathbf{A}\Xi, \\ X_\zeta^L &= \Xi. \end{aligned} \tag{6.15L}$$

Despite the similarity of notation with the $\tilde{G}_{(m)}$ case, the above vector fields correspond to $\tilde{G}_{(m,\omega)}$; the two situations are distinguished by commutators such as $[X_B, X_{\mathbf{A}}] = \omega^2 X_{\mathbf{V}}$. This commutator—we note in passing—explains why the polarization defined by $X_{\mathbf{A}}^L$ is now unsuitable as a polarization for the harmonic oscillator and another one had to be introduced, leading naturally to the Bargmann–Fock–Segal picture.

To compare with the free case, we now return to the left canonical form on $\tilde{G}_{(m,\omega)}$ which is given by [cf. (4.5a) and (4.22b)]

$$\theta^{L(B)} = dB, \quad \theta^{L(\mathbf{A})} = d\mathbf{A} - \mathbf{V}dB, \quad \theta^{L(\mathbf{V})} = d\mathbf{V} + \omega^2\mathbf{A}dB, \tag{6.16a}$$

$$\begin{aligned} \theta^{L(\zeta)} \equiv \Theta &= \frac{m}{2}[\mathbf{V}d\mathbf{A} - \mathbf{A}d\mathbf{V}] \\ &\quad - (\tfrac{1}{2}m\mathbf{V}^2 + \tfrac{1}{2}m\omega^2\mathbf{A}^2)dB + \frac{d\zeta}{i\zeta}. \end{aligned} \tag{6.16b}$$

The vector field of $\mathscr{C}_\Theta$ is given by

$$X = \frac{\partial}{\partial B} + \mathbf{V}\frac{\partial}{\partial\mathbf{A}} - \omega^2\mathbf{A}\frac{\partial}{\partial\mathbf{V}} \equiv X_B^L \tag{6.17}$$

and the corresponding equations $dB/ds = 1$, $d\mathbf{A}/ds = \mathbf{V}$, $d\mathbf{V}/ds = -\omega^2\mathbf{A}$ and $d\zeta/ds = 0$ again dictate the identifications

$$B = t, \quad \mathbf{A} = \mathbf{x}, \quad m\mathbf{V} = \mathbf{p}, \quad \textit{and} \quad \mathbf{F} = -\omega^2 m\mathbf{x}. \tag{6.18}$$

The group thus includes the dynamics; in fact, $\theta^{L(A)} = 0$ defines the velocity, and $\theta^{L(v)} = 0$ is Newton's law. As for Θ^L, it identifies the Hamiltonian $\mathbf{p}^2/2m + \frac{1}{2}m\omega^2\mathbf{x}^2$ as the factor accompanying dB. We give now, for completeness, the action of $G_{(\omega)}$ on the evolution space variables $\mathbf{x},p,t$:

$$\begin{aligned} \mathbf{x}' &= \mathbf{x} + \mathbf{A}\cos\omega t + \frac{\mathbf{V}}{\omega}\sin\omega t, \\ \mathbf{p}' &= \mathbf{p} - m\omega\mathbf{A}\sin\omega t + m\mathbf{V}\cos\omega t, \\ t' &= t + B. \end{aligned} \tag{6.19}$$

These transformations leave invariant, as they should, the oscillator equation $d^2\mathbf{x}/dt^2 = -m\omega^2\mathbf{x}$ and, for $\omega \to 0$, become the familiar Galilei group transformations.

VII. CONCLUDING REMARKS

All developments in this paper, being based on the Galilei group, are obviously nonrelativistic. The problem of ex-

tending the formalism to the relativistic situation appears already in the free case because the Poincaré group does not admit nontrivial extensions by U(1) (nor by other unitary groups) as a consequence of its trivial symplectic cohomology. Indeed, it is not specially illustrative to enlarge an invariance by means of a direct product.[25] Even the requirement of strict invariance of the (initially) classical theory (applied to the Lagrangian or to the Poincaré–Cartan form by imposing *zero* derivative) does not provide a reason to generalize the (classical) relativistic theory since the Poincaré group already admits strict invariance. Besides these difficulties in the use of $U(1) \otimes \mathscr{P}$ as a quantum group, one might add that the elementary quantizing relations, $[x^i, p_j] = i\delta^i_j$, are not compatible with the Poincaré group since, as is well known, the operator $\mathbf{x}$ above does not belong to algebra of $\mathscr{P}$. This constitutes the root of the difficulty of defining a relativistic position operator (Ref. 27; see also 28). Indeed, the above shortcomings merely reflect the fact that the axiomatization of the quantization process of a relativistic theory trails behind that of its nonrelativistic (and hence only approximate) counterpart. A way out to this situation, which is still based on the Poincaré group, would be to look for infinite-dimensional extensions containing both $\mathscr{P}$ and U(1) (and bypassing in this way the no go theorems). This would lead in a natural way to relativistic quantum field theory, whose quantization process may be considered as based on that of an infinity of oscillators. Another possibility would be to consider supersymmetric extensions. This line of research seems worth pursuing, and work is in progress in this direction.

Finally, we conclude by mentioning that possibility exists of using, in the general theory (Sec. III, esp. IIIE), a group T larger than U(1).

ACKNOWLEDGMENTS

This work was supported in part by the Comisión Asesora de Investigación Científica y Técnica under Contract No. 4604-79.

[1]V. Aldaya and J. A. de Azcárraga, Lett. Nuovo Cimento **33**, 341 (1982).
[2]J. M. Souriau, *Structure des systèmes dynamiques* (Dunod, Paris, 1970).
[3]B. Kostant, *Quantization and Unitary Representations*, in *Lecture Notes in Mathematics* **170** (Springer-Verlag, Berlin, 1970), p. 87.
[4]Recent general presentations of the theory are given in D. J. Simms and N. M. Woodhouse, *Lectures in Geometric Quantization* (Springer-Verlag, Berlin, 1976); R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin-Cummins, New York, 1978), Sec. 5.4; J. Śniatycki, *Geometric Quantization and Quantum Mechanics* (Springer-Verlag, New York, 1980); N. Woodhouse, *Geometric Quantization* (Clarendon, Oxford, 1980).
[5]See, e.g., V. Aldaya and J. A. de Azcárraga, La Rivista del Nuovo Cimento **3** (10)(1980).
[6]V. Aldaya and J. A. de Azcárraga, Int. J. Theor. Phys. (1982).
[7]In fact, this condition should be applied to $\omega/\hbar$, where $\hbar$ is the Planck constant.
[8]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.
[9]Given a Lie group G and a Linear representation $D_E(G)$ of G on a vector space E, and *E-1 cocycle* of G is defined as a differentiable mapping $\theta: G \to E / \theta(gg') = \theta(g) + g_E \theta(g')$. Given an arbitrary (but fixed) $\mu \in E$, the mapping $\Delta_\mu: G \to E / \Delta_\mu: g \to g_E \mu - \mu$ is an E-1 cocycle [since $\Delta_\mu(gg') = \Delta_\mu(g) + g_E \Delta_\mu(g')$] called a *1-coboundary*. This may be understood by defining F-valued *cochains* on G and realizing that the *coboundary mapping* of a zero *cochain* (i.e., of a constant function $\mu: g \to \mu(g) \equiv \mu \in E \ \forall\ g$) is a 1-coboundary (i.e., $\partial\mu \equiv \Delta_\mu$).
[10]Since $\mathscr{G}^*$ is a vector space ($\approx \mathbb{R}^n$), $T(\mathscr{G}^*) = \mathscr{G}^*$. Thus $D(\gamma)(e) \equiv \Omega: \mathscr{G} \to \mathscr{G}^*$ and (3.3b) shows that Ω is a 2-form $\Omega: \mathscr{G} \times \mathscr{G} \to \mathbb{R}$. Obviously, $\Omega(Z).Z' = -\Omega(Z').Z = \Omega(Z, Z')$.
[11]This relation holds for any Ω derived from an E-cocycle θ of G, not only for symplectic ones γ. In this last case (3.6) leads (by applying it to Z'') to the Jacobi identity $\Omega(Z).([Z', Z'']) + \Omega(Z').([Z'', Z]) + \Omega(Z'').([Z, Z']) = 0$.
[12]N. Jacobson, *Lie Algebras* (Interscience, New York, 1962).
[13]A. G. Kurosh, *The Theory of Groups* (Chelsea, New York, 1960), Vol. II; see also L. Michel, in *Group Theoretical Methods in Elementary Particles* (Gordon and Breach, New York, 1964).
[14]V. Bargmann, Ann. Math. **59**, 1 (1954).
[15]P. J. Serre, *Lie Algebras and Lie Groups* (Benjamin, New York, 1965).
[16]Infinitesimal exponents solve the corresponding Lie algebra extension problem. The ξ's appearing in (3.8) correspond to the commutator $q(g_1 g_2) = g_1 g_2 g_1^{-1} g_2^{-1}$, and they are associated with the infinitesimal exponents since $[g_1, g_2] = \lim_{t \to 0} (1/t^2) q(t g_1, t g_2)$.
[17]It is in this context that the above definition of polarization [as a subspace of $\mathscr{H}^L(\tilde{G})$ containing the characteristic module $\mathscr{C}_{\theta^L}$] acquires its full meaning.
[18]The observation of (4.9) and (4.5a) exemplifies the comment made at the end of Sec. III.
[19]Apart from a possible sign which depends on the election of the group parameters, operators are defined by $X^R_O \to \hat{O} = -iX^R_O$ (or $-i\hbar X^R_O$ when $\hbar$ is explicitly included).
[20]L. D. Landau and E. M. Lifschitz, *Quantum Mechanics* (Pergamon, New York, 1959), Sec. 15.
[21]This justifies the notation $\tilde{G}_{(m,\omega)}$.
[22]V. Bargmann, Comm. Pure Appl. Math. **XIV**, 187–214 (1961); Proc. Natl. Acad. Sci. USA **48**, 199–204 (1962). See also Simms *et al.* and Śniatycki (Ref. 4) and A. Galindo and P. Pascual, *Mecánica Cuántica* (Alhambra, Madrid, 1978), Sec. 4.4.
[23]The zero-point energy (the 3/2 term) may be introduced through the analysis of the metaplectic structure.[4]
[24]The difference between (6.15R), (6.15L) and (4.4R), (4.4L) in the limit $\omega \to 0$ is due to the different choices of the cocycle.
[25]One might say in this respect that $\tilde{G}_{(m)}$ is more a quantum group than $\mathscr{P} \otimes U(1)$ is; see Ref. 26 in this respect, where the necessity of introducing $\tilde{G}_{(m)}$ is justified in a different context.
[26]V. Aldaya and J. A. de Azcárraga, J. Math. Phys. **22**, 1425 (1981).
[27]T. D. Newton and E. P. Wigner, Rev. Mod. Phys. **21**, 400 (1949).
[28]J. A. de Azcárraga, L. Oliver, and J. Pascual, Phys. Rev. D **8**, 4375 (1973).

Joint distributions, quantum correlations, and commuting observables

Arthur Fine
University of Illinois at Chicago Circle, Chicago, Illinois 60680

(Received 25 August 1981; accepted for publication 26 February 1982)

We provide necessary and sufficient conditions for several observables to have a joint distribution. When applied to the bivalent observables of a quantum correlation experiment, we show that these conditions are equivalent to the Bell inequalities, and also to the existence of deterministic hidden variables. We connect the no-hidden-variables theorem of Kochen and Specker to these conditions for joint distributions. We conclude with a new theorem linking joint distributions and commuting observables, and show how violations of the Bell inequalities correspond to violations of commutativity, as in the theorem.

PACS numbers: 02.50. – r, 03.65.Bz

1. INTRODUCTION

The question of when joint probabilities exist in quantum mechanics is not entirely settled, although several results in the literature suggest that joint probabilities exist only for commuting observables.[1] We show here that the special question of whether observables can have a joint distribution in a given, fixed state lies at the center of recent investigations into hidden variables; in particular, it is the key to the Bell theorems[2] and the no-hidden-variables result of Kochen and Specker.[3] We conclude with a new theorem linking joint distributions and commuting observables.

2. STATISTICAL OBSERVABLES

In this section we establish a framework, and results on joint probabilities, to be applied below to quantum mechanics. We begin by defining a *statistical observable* (or, *observable*, for short) as a pair $\langle A, P_A \rangle$, where A is a real-valued function and P_A is a probability measure on the Borel subsets of the reals ($\mathbf{R}$). Intuitively, $P_A(S)$ gives the probability that A takes a value in S. Thus every random variable paired with its distribution function is a statistical observable. In quantum mechanics every self-adjoint operator $\tilde{A}$ gives rise to statistical observables $\langle A, P_A^{\Psi} \rangle$, where A maps a sequence of unit rays ("states") ϕ_n to a real number λ iff $\mathrm{Inf}||\tilde{A}\phi_n - \lambda\phi_n|| = 0$, and where $P_A^{\Psi}(S) = \langle \chi_S(A) \rangle_{\Psi}$, for χ_S, the characteristic function of the set S and Ψ any state (i.e., unit ray). It is convenient to refer to the function A alone, in the pair $\langle A, P_A \rangle$, as the (statistical) observable, suppressing reference to the measure P_A. Using this convention, we define a *joint distribution* of statistical observables A_1, A_2,...,A_n as a probability measure $P_{A_1,\ldots,A_n}$ on the Borel subsets of $\mathbf{R}^n$ returning each measure associated with each observable as marginals; i.e., satisfying

$$P_{A_1,\ldots,A_i,\ldots,A_n}(\mathbf{R}\times\ldots\times S\times\ldots\times\mathbf{R}) = P_{A_i}(S),$$

where Borel set S occurs in the i^{th} place in the Cartesian product. It is trivial to show that observables always have a joint distribution, since the product measure

$$P_{A_1,\ldots,A_n} = P_{A_1}\cdots P_{A_n}$$

always suffices. If, however, one is given a set of observables with certain fixed joint distributions already defined for various tuples of observables in the set, then it is a nontrivial question whether there exists a joint distribution for all the observables that returns the fixed joints as marginals. If so, we shall say that there is a joint distribution *compatible with* the fixed joints. We now establish some results of this sort, having in mind an application to quantum correlation experiments. (Intuitively, below, think of the fixed joints as the ones quantum mechanics gives—in some state—for pairs of commuting observables.)

Theorem 1: Let observables $A_1, A_2, \ldots, A_n; B_1, \ldots, B_m$ be given together with joint distributions P_{A_i,B_j} for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. There exists a joint distribution for all $n + m$ observables compatible with the given joints if and only if there exists a joint distribution $P_{B_1,\ldots,B_m}$ and corresponding joint distributions $P_{A_i,B_1,\ldots,B_m}$, each of which is compatible with $P_{B_1,\ldots,B_m}$ and P_{A_i,B_j} for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$.

Proof: Clearly, if there is a joint distribution for all the $n + m$ observables, compatible with the joints for the AB pairs, then the stated conditions hold. To establish the converse, notice that these conditions enable one to define density functions $\rho_i = dP_{A_i,B_1,\ldots,B_m}$ on $\mathbf{R}^{m+1}$ and a density $\beta = dP_{B_1,\ldots,B_m}$ on $\mathbf{R}^m$ such that for $\mathbf{y} = \langle y_1, \ldots, y_m \rangle$, $\int_{\mathbf{R}}\rho_i(x_i, \mathbf{y})dx_i = \beta(\mathbf{y})$ for $i = 1, \ldots, n$. Then for $\mathbf{x} = \langle x_1, \ldots, x_n \rangle$ we can define a probability density ρ on $\mathbf{R}^{n+m}$ by

$$\rho(\mathbf{x}, \mathbf{y}) = [\rho_1(x_1,\mathbf{y})\ldots\rho_n(x_n,\mathbf{y})]/\beta^{n-1}(\mathbf{y}). \tag{1}$$

(For $\beta = 0$, we can set the left-hand side to zero as well.) This is a proper density, for

$$\int_{\mathbf{R}^{n+m}}\rho(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y} = \int_{R^m}\beta(\mathbf{y})d\mathbf{y} = 1.$$

Moreover, we get the given distributions $P_{A_i,B_1,\ldots,B_m}$ back as marginals because

$$\int_{\mathbf{R}^{n-1}}\rho(\mathbf{x},\mathbf{y})dx_1\ldots dx_{i-1}dx_{i+1}\cdots dx_n = \rho_i(x_i,\mathbf{y}).$$

for $i = 1, 2, \ldots, n$. Finally, since each ρ_i returns $P_{A_i,B_j}(j = 1, \ldots, m)$ as marginals, the probability measure on the Borel subsets of $\mathbf{R}^{n+m}$ corresponding to the density ρ is the required joint distribution. To apply the theorem it is

 0022-2488/82/071306-05$02.50

useful to state an immediate corollary.

Corollary: Necessary and sufficient for the existence of a joint distribution for observables $A_1, ..., A_n; B_1, B_2$, compatible with given joints $P_{A_i,B_j} (1 \leqslant i \leqslant k < n$ and $j = 1,2)$, is the existence of a joint distribution P_{B_1,B_2} and of distributions P_{A_i,B_1,B_2}, each of the latter compatible with P_{B_1,B_2} and with the given P_{A_i,B_j}.

The corollary enables one to reduce the general problem to conditions on triples of obervables, which we now study in a special case.

Theorem 2: If A, B, B' are bivalent observables (each mapping into $\{x,y\}$) with given joint distributions $P_{A,B}, P_{A,B'}$ and $P_{B,B'}$, then necessary and sufficient for the existence of a joint distribution $P_{A,B,B'}$, compatible with the given joints for the pairs, is the satisfaction of the following system of inequalities:

$$P(A) + P(B) + P(B') \leqslant 1 + P(AB) + P(AB') + P(BB'), \tag{2a}$$

$$P(AB) + P(AB') \leqslant P(A) + P(BB'), \tag{2b}$$

$$P(AB) + P(BB') \leqslant P(B) + P(AB'), \tag{2c}$$

and

$$P(AB') + P(BB') \leqslant P(B') + P(AB), \tag{2d}$$

where we write $P(\ )$ for the probability that each enclosed observable takes the value x.[4]

Proof: Write $\bar{S}$ for the observable taking value y iff S takes value x, and let $\alpha = P(ABB')$. Then the terms in a distribution $P_{A,B,B'}$, if there were one compatible with the given joint distributions for pairs, would have to satisfy

$$P(AB\bar{B}') = P(AB) - \alpha, \tag{3a}$$

$$P(A\bar{B}B') = P(AB') - \alpha, \tag{3b}$$

$$P(A\bar{B}\bar{B}') = P(A) - P(AB) - P(AB') + \alpha, \tag{3c}$$

$$P(\bar{A}BB') = P(BB') - \alpha, \tag{3d}$$

$$P(\bar{A}B\bar{B}') = P(B) - P(AB) - P(BB') + \alpha, \tag{3e}$$

$$P(\bar{A}\bar{B}B') = P(B') - P(AB') - P(BB') + \alpha, \tag{3f}$$

$$P(\bar{A}\bar{B}\bar{B}') = 1 - P(A) - P(B) - P(B') + P(AB) + P(AB') + P(BB') - \alpha. \tag{3g}$$

Using $0 \leqslant \alpha \leqslant \min\{P(AB), P(AB'), P(BB')\}$, the condition that each term in (3) be non-negative produces the system (2). For example, requiring (3c) to be non-negative yields (2b). Conversely, if the system (2) is satisfied then choosing α as above insures that Eqs. (3) define the required distribution $P_{A,B,B'}$.

If we combine Theorem 2 with the corollary to Theorem 1, we get a good working condition for when bivalent observables $A_1, ..., A_n, B_1, B_2$ with preassigned joints P_{A_i,B_j} for $1 \leqslant i \leqslant k < n$ and $j = 1,2$, have a compatible joint distribution; namely, when there exist joint distributions P_{A_i,B_j} (for $k < l \leqslant n$ and $j = 1,2$) such that the system (2) of inequalities is simultaneously satisfiable for $A = A_i$, $B = B_1$, and $B' = B_2$, $1 \leqslant i \leqslant n$. In special cases these inequalities form an especially tractable system.

Theorem 3: If A_1, A_2, B_1, B_2 are bivalent observables with joint distributions P_{A_i,B_j} (for $i = 1,2$ and $j = 1,2$), then necessary and sufficient for there to exist a joint distribution P_{A_1,A_2,B_1,B_2} compatible with the given joints is that the following system of inequalities is satisfied:

$$-1 \leqslant P(A_iB_j) + P(A_i,B_{j'}) + P(A_{i'}B_{j'}) - P(A_{i'}B_j) - P(A_i) - P(B_{j'}) \leqslant 0, \tag{4}$$

for $i \neq i' = 1,2$ and $j \neq j' = 1,2$.

Proof: To show necessity note that, assuming the distribution P_{A_1,A_2,B_1,B_2}, for $i \neq i' = 1,2$, and $j \neq j' = 1,2$,

$$P(A_iB_jB_{j'}) = P(A_iA_{i'}B_jB_{j'}) + P(A_i\bar{A}_{i'},B_jB_{j'}) \leqslant P(A_{i'}B_j) + P(B_{j'}) - P(A_{i'}B_{j'}) \tag{5}$$

and

$$P(\bar{A}_iB_jB_{j'}) = P(\bar{A}_iA_{i'}B_jB_{j'}) + P(\bar{A}_i\bar{A}_{i'}B_jB_{j'}) \leqslant P(A_{i'}B_{j'}) + P(B_j) - P(A_{i'}B_j). \tag{6}$$

Also

$$0 \leqslant P(A_i\bar{B}_j\bar{B}_{j'}) = P(A_i) - P(A_iB_j) - P(A_iB_{j'}) + P(A_iB_jB_{j'}), \tag{7}$$

and

$$0 \leqslant P(\bar{A}_i\bar{B}_j\bar{B}_{j'}) = 1 - P(A_i) - P(B_j) - P(B_{j'}) + P(A_iB_j) + P(A_iB_{j'}) + P(\bar{A}_iB_jB_{j'}). \tag{8}$$

Then (5) with (7) yields the right-hand side of (4), and (6) with (8) yields the left-hand side of (4). In order to show sufficiency, consider inequalities (2), first for $B = B_1$, $B' = B_2$, and $A = A_1$ and then, similarly, for $A = A_2$. If these eight inequalities hold simultaneously for one and the same $P(B_1B_2)$ then, by Theorem 2 and the corollary to Theorem 1, we have the required P_{A_1,A_2,B_1,B_2}. To show that inequalities (4) guarantee all this, let $n = 1,2$ and $m \neq k = 1, 2$; set

$$\gamma = \min(P(A_nB_m) + P(B_k) - P(A_nB_k), P(B_m), P(B_k)) \tag{9}$$

and define $P(B_1B_2) = \gamma$. We can fill out the rest of the distribution P_{B_1,B_2} by letting $P(\bar{B}_1B_2) = P(B_1) - \gamma$, $P(B_1\bar{B}_2) = P(B_2) - \gamma$ and $P(\bar{B}_1\bar{B}_2) = 1 - P(B_1) - P(B_2) + \gamma$. Then (9) and the left-hand side of (4) imply that $P(A_i) + P(B_1) + P(B_2) \leqslant 1 + P(A_iB_1) + P(A_iB_2) + P(B_1B_2)$ for $i = 1, 2$. Similarly, (9) and the right-hand side of (4) imply the remaining six inequalities corresponding to (2b), (2c), and (2d) for the successive $A = A_1, A_2$; $B = B_1$, and $B' = B_2$.

3. CORRELATION EXPERIMENTS AND HIDDEN VARIABLES

We apply the preceding results to quantum correlation experiments. These involve distinct measurements of two noncommuting, bivalent obvservables (with values ± 1) A_1, A_2 in spacetime region R_1 and of two noncommuting, bivalent observables B_1, B_2 (values ± 1) in region R_2. Ideally, R_1 and R_2 would be spacelike separated. In any case, we assume that each A_i commutes with each B_j. Each measurement is performed on one of a correlated pair of particles, for example, on one of pairs of photons emitted in the singlet state from an atomic cascade (see Ref. 2). Various sets of assumptions about the workings of the experiment have been shown to lead to the probabilities of the experiment (i.e., the observed distributions for A_i, B_j and for the commuting pairs A_i, B_j) being constrained by the system of inequalities (4).

Let us refer to these, collectively, as *the Bell /CH inequalities*. It follows from Theorem 3, that the Bell/CH inequalities hold for the probabilities of a quantum correlation experiment if and only if there exists a joint distribution P_{A_1,A_2,B_1,B_2} for the observables of the experiment that is compatible with the observed distributions for the singles A_i and B_j and the commuting pairs A_i, B_j. We now show, in turn, that the existence of such a joint distribution function is equivalent to the existence of a deterministic hidden variables theory for the experiment. Such a theory is defined as follows. Let $A_1,\ldots,A_n,\ldots$ be observables of a quantum system, in a given state Ψ. A *deterministic hidden variables theory* for these observables (in that state) consists of a classical probability space $\Omega = \langle \Lambda, \sigma(\Lambda), P \rangle$, where Λ is a nonempty set (the "hidden variables" = "complete states" of the system), $\sigma(\Lambda)$ is a σ-algebra of subsets of Λ and P is a probability measure on $\sigma(\Lambda)$. We require that there is a mapping $A \to A(\,)$ from the observables $A = A_i$ to random variables on Ω, where the range of $A(\,)$ is the spectrum of A and satisfying

$$P^{\Psi}_{A} = P_{A(\,)}, \tag{D_1}$$

for each given observable $A = A_i$, and

$$P^{\Psi}_{A,B} = P_{A(\,),B(\,)} \tag{D_2}$$

for all commuting pairs, A, B among the given observables. {In (D_2) the left-hand side is the quantum joint distribution, determined by

$$P^{\Psi}_{A,B}(S \times T) = \langle \chi_S(A)\chi_T(B) \rangle_{\Psi}. \tag{10}$$

On the right-hand side of (D_2),

$$P_{A(\,),B(\,)}(S \times T) = P[A^{-1}(S) \cap B^{-1}(T)] \tag{11}$$

is the joint distribution of the random variables $A(\,)$, $B(\,)$.}

It is straightforward to see that there exists such a deterministic hidden variables theory for $A_1, A_2, \ldots$ if and only if there is a joint distribution for $A_1, A_2, \ldots$ compatible with the quantum mechanical distributions $P^{\Psi}_{A_i}$ and $P^{\Psi}_{A_i,B_j}$. For given such a hidden variables theory we can define the distribution for $A_1, A_2, \ldots$ by

$$P_{A_1,\ldots,A_n,\ldots} = P_{A_1(\,)} \cdots P_{A_n(\,)} \cdots; \text{ i.e.,}$$

as the usual product measure. Conversely, suppose we have a joint distribution $P_{A_1,\ldots,A_n,\ldots}$ compatible with the quantum single and joint probabilities (for commuting pairs), then let Λ consist of all sequences $\langle a_1, a_2, \ldots \rangle$, where $a_i \in$ *spectrum of* A_i. Let $\sigma(\Lambda)$ consist of all the infinite-dimensional Borel subsets of Λ^{ω}, and define P by

$$P(S_1 \times \cdots \times S_n \times \cdots) = P_{A_1,\ldots,A_n,\ldots}(S_1 \times \cdots \times S_n \times \cdots).$$

Then (D_1) and (D_2) follow from the compatibility requirements on $P_{A_1,\ldots,A_n,\ldots}$ if we associate with observable A_i the random variable $A_i(\,)$ defined by

$$A_i(\lambda) = a_i \quad \text{for } \lambda = \langle a_1,\ldots,a_i,\ldots \rangle \in \Lambda.$$

Clearly, this same equivalence between hidden variables and joint distributions obtains if we replace the left-hand side of (D_1) and (D_2) by any given distributions. In the case of the quantum correlation experiments, the weight of evidence suggest that the observed distributions are those of quantum mechanics (see Ref. 2). But even if this were not so, we could ask about the possibility for a hidden variables theory returning the experimentally observed probabilities, whatever they are, on the left-hand side of (D_1) and (D_2). We summarize the bearing of our results on this question in the following theorem.

Theorem 4: For a correlation experiment with observables A_1 A_2, B_1, B_2 and with exactly the four pairs A_i, B_j $(i = 1,2; j = 1,2)$ commuting, the following statements are mutually equivalent: (1) The Bell/CH inequalities hold for the single and double probabilities of the experiment; (2) there is a joint distribution P_{A_1,A_2,B_1,B_2} compatible with the observed single and double distributions; (3) there is a deterministic hidden variables theory for A_1, A_2, B_1, B_2 returning the observed single and double distributions; and (4) there is a well-defined joint distribution (for the noncommuting pair) P_{B_1,B_2} and joint distributions P_{A_1,B_1,B_2} and P_{A_2,B_1,B_2}, each of the latter compatible with P_{B_1,B_2} and with the observed single and double distributions.[5]

4. OTHER HIDDEN VARIABLES

There are observables whose quantum mechanical probabilities for certain states of correlated quantum systems violate the Bell/CH inequalities. Likewise, in most of the correlation experiments the observed probabilities also violate these inequalities. Thus both theoretically and experimentally we have a refutation of the possibility of deterministic hidden variables. Before the investigations initiated by Bell on correlated systems, however, there were other no-hidden-variables results. The strongest recent one is due to Kochen and Specker (Ref. 3). We show here the connection between their work and our investigation of joint probabilities and deterministic hidden variables.

Kochen and Specker begin by defining a hidden variables theory, for a set $\mathbf{0}$ of observables of a quantum system in state Ψ, exactly as in our definition in the preceding section for such a deterministic hidden variables theory, including (D_1) for every $A \in \mathbf{0}$, but not requiring (D_2) for commuting pairs. Let us refer to this as a *weak hidden variables theory*. They then suggest that a reasonable-looking formal requirement, in addition, would be to have the algebra of operators mirrored by the algebra of random variables. Thus they add the requirement

$$f(A)(\lambda) = f[A(\lambda)] \tag{KS}$$

for all $\lambda \in \Lambda$ and for every Borel function f (and for all $A \in \mathbf{0}$).

Our first result here is to show that if the set $\mathbf{0}$ is large enough, then (KS) is equivalent to (D_2). Specifically, define a set of observables $\mathbf{0}$ to be *large enough* if (1) whenever $A \in \mathbf{0}$ and $B \in \mathbf{0}$ and $AB = BA$, then $AB \in \mathbf{0}$, and also there is some observable $C \in \mathbf{0}$ such that $A = f(C)$ and $B = g(C)$ for Borel functions f and g; and (2) whenever $A \in \mathbf{0}$ and S is a Borel set, then $\chi_S(A) \in \mathbf{0}$.

Lemma: If $\mathbf{0}$ is large enough, then for $A \in \mathbf{0}$, $B \in \mathbf{0}$ and $AB = BA$, (KS) implies

$$AB(\lambda) = A(\lambda)B(\lambda). \tag{PR}$$

Proof: We have that $A = f(C)$ and $B = g(C)$ for $C \in \mathbf{0}$. By (KS), $A(\lambda) = f[C(\lambda)]$ and $B(\lambda) = g[C(\lambda)]$. But $AB = fg(C)$. So by (KS), $AB(\lambda) = fg(C)(\lambda) = fg[C(\lambda)]$ $= f[C(\lambda)]g[C(\lambda)] = A(\lambda)B(\lambda)$.

Theorem 5: If **0** is large enough, then a weak hidden variables theory for **0** satisfies (KS) only if it satisfies (D_2) for all commuting pairs A, B in **0**.

Proof: It follows from (D_1) and the lemma, that

$$P^{\Psi}_{A,B}(S\times T) = \langle \chi_S(A)\chi_T(B)\rangle_{\Psi} = P^{\Psi}_{\chi_S(A)\chi_T(B)}(1)$$
$$= P_{\chi_S[A(\ )]\chi_T[B(\ )]}(1).$$

By (KS), this yields

$$P^{\Psi}_{A,B}(S\times T) = P[\{\lambda \mid \chi_S(A(\lambda)) = \chi_T(B(\lambda)) = 1\}]$$
$$= P[A^{-1}(S)\cap B^{-1}(T)] = P_{A,B}(S\times T).$$

There is nearly a converse to this theorem, as follows.

Theorem 6: If **0** is large enough, then the following are equivalent. (1) There is a deterministic hidden variables theory for **0**; (2) there is a weak hidden variables theory for **0** satisfying (KS) almost everywhere; (3) there is a weak hidden variable theory for **0** satisfying (PR) almost everywhere.

Proof: We show that (1) implies (2), that (2) implies (3), and that (3) implies (1). To show that (1) implies (2), suppose we have (D_2) for all commuting pairs A, B in **0**. We want to show that $f(A)(\lambda)^{a.e.} = f[A(\lambda)]$; i.e., that $P[\{\lambda \mid f(A)(\lambda)\neq f[A(\lambda)]\}] = 0$. Let y be any number in the spectrum of $f(A)$, and let $S = \{\lambda \mid f[A(\lambda)] = y\}$ and $T = \{\lambda \mid f[A(\lambda)] = y\}$. We want $P(\bar{S}\cap T) = P(S\cap\bar{T}) = 0$. This will follow if we have $P(S) = P(T) = P(S\cap T)$. From (D_1) and the usual rules for functions of observables, we have $P(T) = P[\{\lambda \mid A(\lambda)\in f^{-1}(y)\}] = P^{\Psi}_A(f^{-1}(y)) = P^{\Psi}_{f(A)}(y) = P(S)$. $= P(S)$.] Using the spectral representation of A, it follows that $\chi_D(A)\chi_{f(D)}(f(A)) = \chi_D(A)$ for any set D, where $f(D) = \{f(x) \mid x\in D\}$. Hence, $P^{\Psi}_{A,f(A)}(D\times f(D)) = \langle \chi_D(A)\rangle_{\Psi} = P^{\Psi}_A(D)$. In particular, $P^{\Psi}_{A,f(A)}(f^{-1}(y)\times\{y\}) = P^{\Psi}_A(f^{-1}(y)) = P(S) = P(T)$. But, $P_{A,f(A)}(f^{-1}(y)\times\{y\}) = P(S\cap T)$. The conclusion now follows from (D_2). That (2) implies (3) is a consequence of the lemma. Finally, the derivation of (1) from (3) has already been carried out elsewhere[6] and, since it involves no new principles, need not be repeated here.

This theorem has an immediate corollary that applies to the correlation experiments.

Corollary: If **0** is large enough, then a necessary condition for there to exist a weak hidden variables theory for **0** that satisfies (KS) [or (PR)] is that there exists a joint distribution for every finite subset of **0**, one compatible with all the well-defined quantum mechanical single and joint probabilities in that subset.

If we consider the observables A_1, A_2, B_1, and B_2 for a correlation experiment, then clearly there is a finite, large enough set **0** containing them all. According to the corollary above, and Theorem 4, the failure of the Bell/CH inequalities for particular correlated systems implies that there is no weak hidden variables theory satisfying (KS) for any finite large enough set of observables of such a system. It was just the tying down of the no-hidden-variables results to such finite systems of observables that was the central concern of the Kochen and Specker results. Our work in this section and the previous one shows that the Bell/CH inequalities for the correlation experiments achieve the same end.

5. COMMUTING OBSERVABLES

Our investigations suggest that what the different hidden variables programs have in common, and the common source of their difficulties, is the provision of joint distributions in those cases where quantum mechanics denies them. In this section, we formulate an intuitive criterion for a joint distribution, and show that its satisfaction in quantum mechanics leads to the usual connection between joint distributions and commuting operators.

If A and B are random variables over a common probability space with measure P, then for any two-place Borel function f and any Borel set S, the joint distribution $P_{A,B}$ is well defined, as is the random variable $f(A,B)$, and they satisfy the condition that $P_{A,B}(f^{-1}(S)) = P_{f(A,B)}(S)$. We now propose, essentially, the same condition as a criterion for when several observables of a quantum system have a joint distribution, as follows.

We shall say that observables $A_1, \ldots, A_n$ of a quantum system satisfy *the joint distribution condition* [briefly, (**jd**)] just in case, corresponding to every n-place Borel function f, there is an observable of the system with operator $f(A_1, \ldots, A_n)$, and corresponding to every state Ψ of the system there is probability measure $\mu_{\Psi,A_1,\ldots,A_n}$ on the Borel sets of $\mathbf{R}^n$ that returns the quantum single distributions $P^{\Psi}_{A_i}$ as marginals, such that

$$\mu_{\Psi,A_1,\ldots,A_n}(f^{-1}(S)) = P^{\Psi}_{f(A_1,\ldots,A_n)}(S) \tag{12}$$

for every state Ψ and Borel set S of reals.

Theorem 7: Observables $A_1, \ldots, A_n$ satisfy (**jd**) if and only if all pairs commute.[7]

Proof: If $A_1, \ldots, A_n$ form a commuting set then $f(A_1, \ldots, A_n)$ is well defined for every n-place Borel function f, and the usual joint distribution determined by $\mu_{\Psi,A_1,\ldots,A_n}(S_1\times\cdots\times S_n) = \langle \chi_{S_1}(A_1)\cdots\chi_{S_n}(A_n)\rangle_{\Psi}$ satisfies (**jd**) for all states Ψ. To show the converse we will show that if (**jd**) holds and $A = A_i$, $B = A_j$ then the spectral projections $\chi_S(A)$, $\chi_T(B)$ commute for any Borel sets S, T of reals. So suppose that i,j are fixed and S, T are given Borel sets. Then there are n-place Borel functions f and Borel sets of reals S', T' such that

$$\mathbf{R}\times\cdots\times S\times\cdots\times\mathbf{R} = f^{-1}(S') \tag{13}$$

and

$$\mathbf{R}\times\cdots\times T\times\cdots\times\mathbf{R} = f^{-1}(T'), \tag{14}$$

where S occurs in the ith place in (13), and T in the j th place in (14). For example, we can define a Borel function f by $f(x_1, \ldots, x_n) = 0$ for $x_i\notin S$ and $x_j\notin T$, $f(x_1, \ldots, x_n) = 1$ for $x_i\in S$ and $x_j\in T$, $f(x_1, \ldots, x_n) = 2$ for $x_i\in S$ and $x_j\notin T$, and $f(x_1, \ldots, x_n) = 3$ for $x_i\notin S$ and $x_j\in T$. If $S' = \{1, 2\}$ and $T' = \{1, 3\}$ then (13) and (14) hold. For such an f, we have from (**jd**) that

$$P^{\Psi}_A(S) = \mu_{\Psi,A_1,\ldots,A_n}(\mathbf{R}\times\cdots\times S\times\cdots\times\mathbf{R})$$
$$= \mu_{\Psi,A_1,\ldots,A_n}(f^{-1}(S'))$$
$$= P^{\Psi}_{f(A_1,\ldots,A_n)}(S'). \tag{15}$$

Since (15) holds for all states Ψ, it follows that $\chi_S(A) = \chi_{S'}(f(A_1, \ldots, A_n))$. Similarly,

$\chi_T(B) = \chi_{T'}(f(A_1, ..., A_n))$. Hence, $\chi_S(A)$ commutes with $\chi_T(B)$.

The criterion (**jd**) and Theorem 7 help us to understand the significance of the violations of the Bell inequalities for the correlation experiments, for the observables A_1, A_2, B_1, B_2 of the experiments (with values ± 1) do not form a commuting set. Hence, by Theorem 7, if $f(x_1, x_2, y_1, y_2) = x_1y_1 + x_1y_2 + x_2y_2 - x_2y_1$ and we try the correspondence rule $f(A_1, A_2, B_1, B_2) = A_1B_1 + A_1B_2 + A_2B_2 - A_2B_1$ then Eq. (12) will fail for some set S and state Ψ. In particular, if S is the closed interval from -2 to $+2$, then $f^{-1}(S) \supseteq \{-1, 1\}^4$ and the left side of (12) must be 1 for any measure. But in certain singlet states Ψ (namely, those for which the Bell/CH inequalities fail) the quantum mechanical probability on the right side of (12) will differ from 1. Thus violations of the Bell/CH inequalities are particular cases where (**jd**) fails, as Theorem 7 tells us it somewhere must, for observables not all pairs of which commute. [Of course, it is Bell's important and lasting contribution to have found cases especially simple, and also experimentally tractable, where (**jd**) does fail.]

It seems natural to take (**jd**) as a criterion for when observables have a joint distribution. It is a coarse-grained criterion, not sensitive to the particular state of a system. As we have seen in the preceding sections, more finely grained criteria (and hidden variables are among them) are equivalent to constraints (like the Bell/CH inequalities) that some quantum systems violate in certain states. These violations have been experimentally confirmed. Perhaps, then, we ought to accept the straight-line induction; that where (**jd**) fails, and quantum mechanics does not give a well-defined joint distribution, neither would experiments. After all, if we hold that probabilities (including joint probabilities) are real properties, then some observables may simply not have them.

ACKNOWLEDGMENTS

Work for this study was supported by a grant from the National Science Foundation. My research assistant for this project, Dana Fine, provided the insight for the general formulation of Theorem 1 and, more especially, of its corollary. Paul Teller suggested the term "statistical observable" to replace the term "statistical variable" that I had used in previous publications. Patrick Suppes and Isaac Levi have been helpful correspondents.

[1] See K. Urbanik, Studia Math. **21**, 117 (1961); L. Cohen, J. Math. Phys. 7, 781 (1966); E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton, U.P., Princeton, 1967), pp. 117–119; S. Gudder, J. Math. Mech. **18**, 325 (1968); V. V. Kuryshkin, in *The Uncertainly Principle and Foundations of Quantum Mechanics*, edited by W. Price and S. Chissick (Wiley, New York, 1977), pp. 61–83; S. Bugajski, Z. Naturforsch. Teil A **33**, 1383 (1978).

[2] See the survey of this literature by J. Clauser and A. Shimony, Rep. Prog. Phys. **41**, 1881 (1978).

[3] S. Kochen and E. Specker, J. Math. Mech. **17**, 59 (1967).

[4] For the special case where the range of the observables is ± 1 and where $\frac{1}{2} = P(A) = P(B) = P(B')$, a system of inequalities equivalent to (2) was first discovered by P. Suppes and M. Zanotti, Synthese (to be published). I want to thank Suppes and Zanotti for sharing their work with me, and for some very stimulating exchanges of ideas.

[5] We can add a fifth equivalent statement to this list; namely, (5) there exists a factorizable (so-called "local") stochastic hidden variables theory for A_1, A_2, B_1, B_2 returning the observed single and double distributions. See J. Clauser and M. Horne, Phys. Rev. D **10**, 526 (1974), for the definitions here. It is well known that (3) implies (5), and Clauser and Horne show that (5) implies (1). Thus the equivalence follows from our proof that (1) implies (3). It is easier, however, indeed trivial, to show that (5) implies (2), and to get the equivalence from that of (2) to (3). See my "Hidden variables, joint probability and the Bell inequalities," Phys. Rev. Lett. **48**, 291 (1982), which also contains another derivation of (3) from (1).

[6] A. Fine, Synthese **29**, 257 (1974). This is reprinted, with a relevant correction to the proof, in *Logic and Probability in Quantum Mechanics*, edited by P. Suppes (Reidel, Dordrecht, 1976), pp. 249–281.

[7] For pairs of discrete observables a computational proof is contained in A. Fine, Brit. J. Philos. Sci. **24**, 1 (1973). I want to thank Robert Latzer for correspondence that helped me find the simple, general derivation below.

Theory of nonbijective canonical transformations in mechanics: Application to the Coulomb problem [a)]

M. Boiteux

Laboratoire d'Ultrasons, [b)] Université Pierre et Marie Curie, Tour 13, 4 Place Jussieu, 75230 Paris Cedex 05, France

(Received 11 November 1980; accepted for publication 12 May 1981)

It is shown that the study of nonbijective transformation requires a fiber bundle formulation of mechanics. The conditions upon which nonbijective canonical point transformations can be defined are given. Then, as an example, we apply that theory to the study of the Coulomb problem in two and three dimensions. The Hopf fibration leads to an inverse harmonic oscillator problem.

PACS numbers: 03.20. + i

I. INTRODUCTION

For a decade, there had been an upsurge of interest in canonical transformations from both classical and quantal view points. From the beginning, it has been clear from examples (Stiefel *et al.*,[1] Cisneros *et al.*,[2] Boiteux[3,4]) that nonbijective point transformations might be useful. In quantum mechanics, those transformations connect operators with different spectra which as such cannot be deduced from one another by unitary transformations. If we wish to restore a one-to-one relation between states and matrix elements, in order to get the spectrum of an operator from that of another, we must kill the superfluous states by auxiliary conditions. These constraints result in the existence of a Lie group of transformations. Subsequently, Moshinsky *et al.*[5-8] systematically studied that group on examples and named it the ambiguity group.

However, all the previous papers lack mathematical rigor, due to the nonregularity of nonbijective transformations; also a clear definition of the ambiguity group is missing. This paper intends to solve these problems in the case of point transformations. The general situation can be studied in the same way, but with a little more involved mathematics.

II. OUTLINE

It turns out that the suitable mathematical framework for our purposes is fiber bundle theory.

Therefore in Sec. III we give some definitions and mathematical results in the differential geometry of fiber bundles. Then we state our main theorem about the conditions under which a nonbijective canonical point transformation can be defined.

Section IV is devoted to some applications. In particular, we study trivial nonbijective point transformations and discuss the Coulomb problem in two and three dimensions.

Finally in the conclusion we outline the extension to general nonbijective canonical transformations.

III. SOME DEFINITIONS AND RESULTS[9]

Definition 1: The configuration space of a dynamical system S with n degrees of freedom is a manifold X^n with local coordinates $\{x^i\}$.

Definition 2: The manifold $X^n \times R$ with local coordinates (x^i,t) is the spatiotemporal configuration space of S.

Definition 3: Let M be a manifold equipped with a metric g and let $\bar{M}$ be a fibration over M equipped with a metric $\bar{g}$. Let Π: $\bar{M} \to M$ be a fiber bundle. Let $\bar{M}_m$ be the fiber at a point m of the base space M.

Since $\bar{M}$ is equipped with $\bar{g}$, we can define the orthogonal space to the kernel of the projection Π (this kernel is the so-called vertical space of the fiber at the point $\bar{M}$). We consider the restriction $\bar{g}^{-1}$ of $\bar{g}$ to this space which it is natural to call the horizontal space.

Definition 4: We said that Π is a Riemannian submersion if Π is an isometry of $\bar{g}^{-1}$ to g.

Theorem 1: Hermann[10] has given a sufficient condition that a mapping of Riemannian manifolds be a fiber bundle.

If $\phi: X \to B$ is a C^∞ map of manifolds, $\phi_*: X_x \to B_{\phi(x)}$ is the linear map on tangent vectors induced by ϕ such that (i) ϕ has maximal rank on X, i.e., $\phi_*(X_x) = B_{\phi(x)}$ for all $x \in X$; (ii) X and B are Riemannian manifolds, and the isomorphism $\phi_*: X_x/\phi_*^{-1}(0) \sim B_{\phi(x)}$ preserves the inner products defined by the metrics on these spaces for all x in X. If X is complete as a Riemannian space, so is B. ϕ is then a locally trivial fiber space. If in addition the fibers of ϕ are totally geodesic submanifolds of X, ϕ is a fiber bundle with structure group the Lie group of isometries of the fiber.

Remark: If the typical fiber F is compact, G is compact. A finite group is a Lie group with zero dimension.

Definition 6: Let A_g be an operator on M which is naturally defined in terms of g (e.g., the Laplace–Beltrami operator). Let $A_{\bar{g}}, A_{\bar{g}_m}$ be the corresponding operators on $\bar{M}, \bar{M}_m$ respectively.

Define $A_{\bar{g}}^F$ on $\bar{M}$ by $A_{\bar{g}}^F(\bar{f})(\bar{m}) = A_{\bar{g}_m}(\bar{f}^{-1}\bar{M}_m)(\bar{m})$, where $m = \Pi\bar{m}$. $A_{\bar{g}}^F$ is in a sense the "vertical part" of $A_{\bar{g}}$. Write $A_{\bar{g}}^h = A_{\bar{g}} - A_{\bar{g}}^F$ and call it the "horizontal part" of $A_{\bar{g}}$.

Theorem 2: (Berard Bergery and Bourguignon, Ref. 11). $[A_F, A_{\bar{g}}] = 0$ if the fibration $\Pi: (M,\bar{g}) \to (M,g)$ is a Riemannian submersion with totally geodesic fibers.

Then, if A kills the constant functions, $A_{\bar{g}}^h(f \circ \Pi) = A_g(f) \circ \Pi$ and $A_{\bar{g}}^h$ acting on functions lifted from M can be identified with a natural operator A on M. This is strictly shown for the Laplacian by Bérard Bergery and Bourguignon and indeed we need only their result in our applications. But the result is true also for A from its expres-

[a)] This paper covers a part of a Ph.D. thesis by the author, Paris University (1979).

[b)] Associated with the Centre National de la Recherche Scientifique.

 0022-2488/82/071311-04$02.50

sion from vector fields.

Remarks: There exists an expression of the vertical operator A_F in terms of vertical vector fields; on another hand, the Lie algebra of the holonomy group is generated by some vertical operators.

We now specialize Theorem 2 in order to study quantum problems.

Let $\bar{f}'$ be defined on $\bar{M}$; then if $L_F \bar{f}' = 0, \bar{f}'$ admits a unique representative $\bar{f}'$ on M ($\bar{f}'$ is deduced from f' by composition).

If $f \in L^2(M,g)$, then $f \circ \Pi \in L^2(\bar{M},\bar{g})$.

If the fiber is compact in order that the functions constant along the fibers will be in L^2, then we have the main theorem 3.

Theorem 3: Let A_g be an operator on a manifold (M,g) naturally defined in terms of the metric g. Then if $\Pi:(\bar{M},\bar{g}) \to (M,g)$ is a Riemannian submersion with totally geodesic fibers with compact structural group G, then the A_F-invariant part of the spectrum of the operator $A_{\bar{g}}$ on $\bar{M}$ is equivalent to that of A_g. Moreover, if G acts transitively on the fiber, A_F can be defined in terms of vector fields defined by G.

Under these conditions, Π is called a diastrophic canonical point transformation and G the ambiguity group of that transformation.

Remarks: When $\bar{M} = M$, we recover ordinary canonical point transformations. When G is finite, it is better to directly use the invariance under the generators of G than Theorems 2 and 3.

IV. EXAMPLES AND APPLICATIONS

IV.1 Examples

IV.1.1 Trivial diastrophic transformation

This corresponds to the case of trivial fiber bundles $\bar{M} = M \times F$.

IV.1.2 Quadratic transformations

One can give a classification of quadratic transformations preserving rotational invariance based on geometric properties of numbers.[12] A theorem by Hurwitz says that such transformation can exist only when the dimension d is 1, 2, 4, or 8. We will not discuss here the case $d = 8$ connected with octonions.

1) $d = 1$: There exists only one[13]: $x = \bar{x}^2$ and $M = \mathbb{R}^+$, $\bar{M} = \mathbb{R} - \{0\}$. The ambiguity group is $\mathbb{Z}_2 = C_2$, the cyclic group with two elements.

2) $d = 2$: $\mathbb{R}^2 \sim \mathbb{C} \ni z = x_1 + ix_2$. There are only two kinds of quadratic transformations:

a) $z = \bar{z}\,\bar{z}^*$

and $M = \mathbb{R}^+, \bar{M} = \mathbb{R}^2 - \{0\}$. The fibers are circles and $G = \mathrm{U}(1)$ (notice that the fibers can be identified with the group).

b) $z = \bar{z}^2$

and $M = \mathbb{R}^2 - (0), \bar{M}$ is a two-sheeted covering of M and $G = Z_2$.

3) $d = 4$: $\mathbb{R}^4 \sim \mathbb{H} \ni q = x_1 + ix_2 + jx_3 + kx_4$ the field of quaternions. Also, there are only two kinds of quadratic transformations:

a) $q = \bar{q}\bar{q}^*$,

$M = \mathbb{R}^+, \bar{M} = \mathbb{R}^4 - \{0\}$. $G = \mathrm{SU}(2)$ and the fibers are three-dimensional spheres in $\mathbb{R}^4$ centered at the origin. (Notice here again that the fibers can be identified with the group itself.)

b) $x_1 = \bar{x}_1^2 - \bar{x}_2^2 - \bar{x}_3^2 + \bar{x}_4^2$,
$x_2 = 2(\bar{x}_1\bar{x}_2 + \bar{x}_3\bar{x}_4)$,
$x_3 = 2(\bar{x}_1\bar{x}_3 - \bar{x}_2\bar{x}_4)$,

$M = \mathbb{R}^3 - \{0\} = S^2 \times \mathbb{R}^+; M = \mathbb{R}^4 - \{0\} = S^3 \times \mathbb{R}^+$. This is the Hopf fibration when $|q| = 1: S^3 \to S^2$. $G = \mathrm{U}(1)$ and the fibers are circles in $\mathbb{R}^4$ centered at the origin.

4) $d = 8$: Then the Hopf fibration $S^7 \times \mathbb{R}^+ \to S^4 \times \mathbb{R}^+$ may be interesting for further study.

IV.2 Applications

IV.2.1 The Coulomb problem in two dimensions

In this case we shall see that G is finite. The Coulomb potential $V(r) = -e^2/r$ is singular at the origin. Then $M = \mathbb{R}^2 - \{0\}$. Therefore, the corresponding Hamiltonian $H = p^2/2m - e^2/r$ or, equivalently, the Schrödinger operator $S_E = H - E$ is not self-adjoint but only essentially self-adjoint. S_E has a unique self-adjoint extension anyway. But rS_E is not symmetric on $L^2(R^2 \setminus 0, dx_1 dx_2)$, because of the factor r. Using the transformation $z = \bar{z}^2$ (IV.1.2.2b). $(Vf)(z) = f(\bar{z}^2)$, $V: L^2(R^2 \setminus 0, dx_1\, dx_2) \to L^2(R^2 \setminus 0, d\bar{x}_1\, d\bar{x}_2)$; then $V(rS_E)V^{-1}$ is symmetric on the subspace of $L^2(R^2 \setminus 0, dx_1\, dx_2/4r = d\bar{x}_1\, d\bar{x}_2)$ for which V^{-1} exists because the factors r and $1/r$ cancel.

Now $S'_E = V(rS_E)\, V^{-1} \equiv (-(1/8m)\bar{\Delta} - E\bar{r}^2 - e^2)$ is a harmonic oscillator Schrödinger operator.

As already mentioned, it is preferable to directly use the invariance of the wavefunction under the discrete symmetry $I \in z_2$.

Then

$$S'_E \bar{\psi} = (-(1/8m)\bar{\Delta} - E\bar{r}^2 - e^2)\bar{\psi},$$
$$I\bar{\psi} = \bar{\psi}, \quad \bar{\psi} \in \mathcal{L}^2(M, d\bar{x}_1\, d\bar{x}_2).$$

Thus the bound-state Coulomb problem in two dimensions is equivalent to an inverse two-dimensional harmonic oscillator inverse problem the energy of which is defined and equal to $4e^2$, the angular frequency $\omega = (-8E/m)^{1/2}$ of which is unknown, and the eigensolution of which are univariant under symmetry $(x_1, x_2) \to (-x_1, -x_2)$.

IV.2.2 The Coulomb problem in three dimensions

Now, G is a continuous group. The first step is identical to the $d = 2$ case. But now, we use the quadratic transformation (IV.1.3b). Here more details on the Hopf fibration[14] are needed in order to get a complete answer.

$$M = S^2 \times R^+, \quad \bar{M} = S^3 \times R^+.$$

This transformation reads in coordinates x_1, x_2, x_3 and $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4$ (see Kustaanheimo and Stiefel[15]).

$$x_1 = \bar{x}_1^2 - \bar{x}_2^2 - \bar{x}_3^2 + \bar{x}_4^2,$$
$$x_2 = 2(\bar{x}_1\bar{x}_2 + \bar{x}_3\bar{x}_4),$$
$$x_3 = 2(\bar{x}_1\bar{x}_3 - \bar{x}_2\bar{x}_4),$$

and satisfy

$$x_1^2 + x_2^2 + x_3^2 = r^2 = (\bar{x}_1^2 + \bar{x}_2^2 + \bar{x}_3^2 + \bar{x}_4^2)^2 = \bar{r}^4.$$

Let us introduce the notation

$$\bar{q} = \bar{z}_0 + j\bar{z}_1 \quad \bar{z}_0, \bar{z}_1 \in \mathbb{C}$$

induced by the relation $\mathbb{H} \sim \mathbb{C} \oplus \mathbb{C}$. S^3 is the underlying space of the group SU(2) which can be parametrized by the Euler angles ψ, θ, ϕ. If $q\bar{q}^* = 1$,

$$\begin{pmatrix} z_0 & iz_1^* \\ iz_1 & z_0^* \end{pmatrix} = e^{i(\phi/2)\sigma_3} e^{i(\theta/2)\sigma_1} e^{i(\psi/2)\sigma_3}$$

$$= \begin{pmatrix} \cos\frac{\theta}{2} e^{i(\psi+\phi)/2} & i\sin\frac{\theta}{2} e^{i(\phi-\psi)/2} \\ i\sin\frac{\theta}{2} e^{i(\psi-\phi)/2} & \cos\frac{\theta}{2} e^{-i(\psi+\phi)/2} \end{pmatrix},$$

from which we get

$$\bar{x}_1 = \bar{r}\cos\frac{\phi+\psi}{2}\cos\frac{\theta}{2},$$

$$\bar{x}_2 = \bar{r}\sin\frac{\phi+\psi}{2}\cos\frac{\theta}{2},$$

$$\bar{x}_3 = \bar{r}\cos\frac{\phi-\psi}{2}\sin\frac{\theta}{2},$$

$$\bar{x}_4 = \bar{r}\sin\frac{\phi-\psi}{2}\sin\frac{\theta}{2},$$

and

$$x_1 = r\sin\phi\cos\theta,$$

$$x_2 = r\sin\phi\sin\theta,$$

$$x_3 = r\cos\theta.$$

So (ϕ, θ) are the polar angles on the sphere S^2. Then ψ parametrizes the fibers which are great circles on S^3. It follows that the Hopf fibration is totally geodesic and the conditions of Theorem 3 are satisfied since the group $S^1 = U_1$ is compact.

The unit vectors tangent to the fiber at $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4)$ are $\pm(1/\bar{r})(-\bar{x}_4, -\bar{x}_3, -\bar{x}_2, \bar{x}_1)$. The length element is

$$d\bar{s}^2 = d\bar{r}^2 + \bar{r}^2[d\theta^2 + \sin^2\theta\,(d\varphi^2 + d\psi^2) + 2\cos\theta\, d\varphi\, d\psi],$$

from which we get the Laplacian

$$\bar{\Delta}_4 = \frac{\partial^2}{\partial\bar{r}^2} + \frac{3}{\bar{r}}\frac{\partial}{\partial\bar{r}} + \frac{4}{\bar{r}^2}\left[\frac{\partial^2}{\partial\theta^2} + \cot\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\left(\frac{\partial^2}{\partial\phi^2} + \frac{\partial^2}{\partial\psi^2}\right) - 2\cos\theta\frac{\partial^2}{\partial\phi\partial\psi}\right].$$

But

$$\Delta_3 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right],$$

where $r = \bar{r}^2$. Here, the symmetry of $V(rS_E)V^{-1}$ arises again from an $(1/r)\times r$ cancellation arising from $\Pi^*(d\bar{x}_1, d\bar{x}_2, d\bar{x}_3, d\bar{x}_4) = (dx_1, dx_2, dx_3)/4r$.

Then, the auxiliary condition which ensures the wave function $\bar{\psi}_E$ is univalued, is $\Delta_F\bar{\psi}_E = 0$, or equivalently $(\partial/\partial\psi)\bar{\psi}_E = 0$, where $\partial/\partial\psi$ is the infinitesimal generator along the great circles.

Then the Schrödinger equation reads

$$\left(-\frac{1}{8m}\bar{\Delta}_4 - E\bar{r}^2 - e^2\right)\bar{\psi}_E = 0, \quad \bar{\psi}\in\mathscr{L}^2(\bar{M}),$$

$$\bar{r}\bar{L}_F\bar{\psi}_E = \bar{r}\frac{\partial}{\partial\psi}\bar{\psi}_E = \left(-\bar{x}_4\frac{\partial}{\partial\bar{x}_1} + \bar{x}_3\frac{\partial}{\partial\bar{x}_2} - \bar{x}_2\frac{\partial}{\partial\bar{x}_3} + \bar{x}_1\frac{\partial}{\partial\bar{x}_4}\right)\bar{\psi}_E = 0.$$

The reader is referred to Boiteux[3] for the resolution of that system. (See also Miyachi.[16]) In other words, the last condition exactly kills the irrelevant states among those of a degenerated level of the 4-d harmonic oscillator in order to get the correct accidental degeneracy of the corresponding level of hydrogen atom.

Remarks: The classical (Kepler) problem is solved in the same way. Thus the canonical nonbijective transformation works for the classical and the corresponding quantum problem as well. In particular, the ambiguity group is the same.

The case of the hydrogen atom with a planar barrier through the origin also is easily solved because it transforms into a planar barrier in $R^4 - \{0\}$ by means of the Hopf fibration.

Indeed, the condition $A\frac{F}{\bar{g}}\bar{f} = 0$ can be replaced by $A\frac{F}{\bar{g}}\bar{f} = a\bar{f}$ and still preserves a one-to-one correspondence. However, in this case various sections of the fibration are inequivalent. In particular $A\frac{h}{\bar{g}}$ now depends on a.

V. CONCLUSION

We have seen that nonbijective point transformations can be used in both classical and quantum mechanics. They require the introduction of particular fiber bundles which satisfy the conditions of Theorem 3. However, it is not too difficult to extend these results to general nonbijective transformations. We will now outline that extension.

We consider the cotangent bundle $T^*(X^n)$ to the configuration manifold (the phase space) equipped with the natural metric $\bar{g}$ inherited from the natural Sasaki metric[17] of the tangent bundle $T(X^n)$ as the base of a suitable fibration.

Then all the aspects of Theorem 3 can be extended. In particular, A may be the Liouville operator (or Lie derivative). That is the classical situation.

As for the quantal case, that structure is transported by the Weyl quantization which associates to Poisson brackets, Frechet–Poisson brackets.[18] An elementary example has been discussed recently by Newton.[19]

Note added in proof: Since the completion of this work, a paper by G. H. Ringwood and J. T. Devreese has been published in J. Math. Phys. **21**, 1390 (1980), dealing with the same problem. Their work is based on the construction of propagators in quotient spaces. The identity between the propagator prescriptions and nonlinear canonical transformations is not automatically fulfilled. Therefore it seems that the reliability of their results is not due to their method,

which in general is not correct, but to an underlying property of the transformation used, namely the Kustaanheimo–Stiefel map (see our results).

ACKNOWLEDGMENTS

The author wishes to thank Professor H. Bacry for suggesting to him the possible relevance of the fiber bundle structure to his problem and for his useful encouragement, and Doctor J. -P. Bourguignon for communicating his results prior to publication, for his very clear comments and discussions on Riemannian submersions, and for a critical reading and corrections of the manuscript. The author also wishes to take this opportunity to acknowledge the kind hospitality of Professor Moshinsky and his group at Mexico University during the fall of 1974.

[1]E. Stiefel and G. Scheifeld, C. R. Acad. Sci. Paris **267**, 950 (1968).
[2]A. Cisneros and H. V. McIntosh, J. Math. Phys. **10**, 277 (1969).
[3]M. Boiteux, Physica **65**, 581 (1973).
[4]M. Boiteux, Physica **75**, 603 (1974).
[5]M. Moshinsky and M. A. Mello, J. Math. Phys. **16**, 2017 (1975).
[6]M. Moshinsky, P. Kramer, and T. H. Seligman, J. Math. Phys. **19**, 683 (1978).
[7]M. Moshinsky and T. H. Seligman, Ann. Phys. **114**, 243 (1978).
[8]M. Moshinsky and T. H. Seligman, Ann. Phys. **120**, 402 (1979).
[9]For general information on fiber bundles: N. Steenrod, *The Topology of Fibre Bundles* (Princeton U. P., Princeton, N. J., 1951).
[10]R. Hermann, Proc. Am. Math. Soc. **11**, 236 (1980).
[11]L. Berard Bergery and J. -P. Bourguignon, "Laplacian and Riemannian Submersion with Totally Geodesics Fibers," in *Proc. Janos Bolay Math. Soc. Meeting* (Budapest, Sept., 1979); *Proc. Colloquium on Global Differential Geometry and Global Analysis* (Berlin, Nov., 1979) (to appear).
[12]The author is indebted to H. Bacry for pointing out this classification.
[13]In the following, the bar refers to variables of $\bar{M}$. The stars mean conjugation.
[14]H. Hopf, Math. Ann. **104**, 657 (1931); Selecta Heinz Hopf **38** (Springer, Berlin, 1969). For related works, see L. H. Ryder, J. Phys. A **13**, 437 (1980).
[15]P. Kustaanheimo and E. Stiefel, J. Reine Angew. Math. **218**, 204 (1965).
[16]M. Miyachi and M. Ikeda, Math. Japan. **15**, 127 (1970); the author acknowledges Doctor Miyachi for pointing out his work for him.
[17]A. Besse, *Manifolds all of whose geodesics are closed, Ergebrisse der Mathmatik und Grenzbebiete, Vol. 93* (Springer, Berlin, 1978), pp. 46–47.
[18]F. J. Bloore, J. Phys. A **6**, L-7 (1973).
[19]R. G. Newton, Ann. Phys. **125**, 327 (1980).

The quartic anharmonic oscillator in stochastic electrodynamics [a)]

L. Pesquera
Departamento de Fisica Teorica, Universidad de Santander, Santander, Spain

P. Claverie
Laboratoire de Chimie Quantique, Institut de Biologie Physico-Chimique, 13, rue Pierre et Marie Curie, 75005 Paris, France

(Received 24 February 1981; accepted for publication 10 April 1981)

The case of a slightly anharmonic oscillator (with a βx^4 perturbing potential) is examined in the framework of stochastic electrodynamics (SED) in full detail. We obtain the stationary probability density and the mean energy, which differs from the quantum result at order β^2. Using Kubo's linear response theory we obtain the absorption curve: the maximum absorption frequencies do not coincide with the quantum transition frequencies. From the calculation of the emission energy we show that the "radiation balance" is *not* exactly satisfied as soon as $\beta \neq 0$, a property which disagrees with the quantum results. Finally, we discuss the consequences of this lack of radiation balance concerning Kirchhoff's law.

PACS numbers: 03.65.Bz, 41.70. + t, 02.50. + s

I. INTRODUCTION

Stochastic electrodynamics (SED) is a classical theory which has been proposed as a possible alternative to quantum theory. Basically SED is classical electrodynamics (including Lorentz–Dirac radiation damping) supplemented by the assumption that there exists a (classical) electromagnetic field in the whole space, which is so complex that it is appropriately treated as a stochastic field.[1–4] Assuming that its stochastic properties are Lorentz-invariant, it may be found that this background field must have zero mean value and a spectral density (see, e.g., Ref. 4a):

$$\int_{-\infty}^{\infty} e^{i\omega\theta} \mathscr{B}_{\mathscr{E}}(\theta)\, d\theta = \mathscr{S}_{\mathscr{E}}(\omega) = \frac{2\hbar}{3c^3}|\omega|^3, \tag{1}$$

where $\mathscr{B}_{\mathscr{E}}(\theta)$ denotes the correlation function of the stochastic process $\mathscr{E}(t)$ (this relationship between $\mathscr{B}$ and $\mathscr{S}$ is the same as in Ref. 3).

Consequently, the equation of motion in SED for a nonrelativistic charged particle (mass m, charge e) is a Lorentz–Dirac type equation (called a Braffort–Marshall equation)

$$m\ddot{\mathbf{r}} = \mathbf{F}(\mathbf{r}) + m\tau\dddot{\mathbf{r}} + e\mathscr{E}(t) \tag{2}$$

where, in addition to the external known force $\mathbf{F}(\mathbf{r})$ and the usual radiation damping force $m\tau\dddot{\mathbf{r}}$ (where $\tau = 2e^2/3mc^3$), we have a stochastic electromagnetic force $e\mathscr{E}(t)$. This last term is written in the electric-dipole approximation, which neglects the magnetic force of the background field and the spatial dependence of the electric field $\mathscr{E}$.[4b]

Equation (2) has so-called runaway (i.e., self-accelerating) solutions. A procedure to eliminate these undesirable solutions is to use, instead of Eq. (2), an integro-differential equation (see, e.g., Ref. 5). Applying this procedure to Eq. (2) and keeping the terms up to order τ ($\tau = 0.26 \times 10^{-6}$ a.u.) in the Taylor expansion of the deterministic force (Ref. 1, Sec. 3B), we obtain the following equation:

$$m\ddot{\mathbf{r}} = \mathbf{F}(\mathbf{r}) + \tau\overline{\nabla}\mathbf{F}\cdot\dot{\mathbf{r}} + e\mathscr{E}^{(\tau)}(t), \tag{3}$$

where $\mathscr{E}^{(\tau)}(t)$ is a modified stochastic field with zero mean value and a spectral density (for each component)

$$\mathscr{S}^{(\tau)}_{\mathscr{E}}(\omega) = \frac{\mathscr{S}_{\mathscr{E}}(\omega)}{1+\tau^2\omega^2} = \frac{2\hbar}{3c^3}\frac{|\omega|^3}{1+\tau^2\omega^2}. \tag{4}$$

The theory constitutes, therefore, a well-defined problem of mathematical physics, but its solution is rather difficult, due to the nonwhite character of the stochastic field. It is only recently that the nonstandard techniques required to treat nonlinear problems in SED have been worked out (Refs. 3, 6, and 21b).

In the case of linear systems, Eq. (3) is easily solved by applying the Fourier transform method. Although the results are not fully identical with those of quantum theory, they are rather satisfactory.[1–4,7]

However, for nonlinear systems, such as the anharmonic oscillator[8,9] and the Kepler problem,[10,11] the results obtained until now are not in agreement with quantum theory.

The aim of this paper is to study in full detail a slightly anharmonic oscillator (with a βx^4 perturbing potential) in SED. Some preliminary results for this problem were reported in Refs. 8 and 9.

The outline of the paper is as follows. In Sec. II, using the techniques worked out to solve non-Markovian stochastic differential equations such as Eq. (3), we obtain the stationary probability density W_0 and the average energy, which is not the same as in quantum theory. Then, using W_0 and the Kubo linear response theory, we calculate in Sec. III the absorption coefficient at each frequency. It is found that the first and second maxima of the absorption curve do not coincide with the quantum and experimental results.

In Sec. IV we obtain the emitted power at each frequency, and we study the balance between absorbed and emitted power at each frequency ("radiative balance").

This balance is not satisfied in the background field of SED. Finally, we show that the anharmonic oscillator would fulfill this radiative balance condition only with a background field having the (Rayleigh–Jeans) ω^2-spectrum, and we discuss the consequences of this lack of radiation balance in SED concerning Kirchhoff's law.

 0022-2488/82/071315-08$02.50

II. THE STATIONARY STATE OF THE (QUARTIC) ANHARMONIC OSCILLATOR IN SED

A. The stationary density in the Markovian approximation

We would like first to recall the basic ideas of the methods used to solve non-Markovian stochastic differential equations[6] [such as the Braffort–Marshall equation (3)]. Indeed, this topic is not yet very well known except by specialists. In order to solve equations of this kind we use the fact that the damping and stochastic forces are much smaller than the deterministic force, due to the presence of the (small) parameter τ.

Therefore the relaxation time (namely, the interval of time required for the perturbation to have a finite effect with respect to the deterministic motion) is of order $1/\tau \simeq 10^6$ a.u., whereas the correlation time of the stochastic force is of order of 1 a.u. (Ref. 3, Sec. 4.A).

In this case, namely a small stochastic force with a correlation time short with respect to the relaxation time, two main lines of approximations to the solution of Eq. (3) are possible. They will be denoted by Lax-5 and Lax-6, respectively, according to Secs. 5 and 6 of Lax's important paper.[12]

(a) In the Lax-5 method, the exact (non-Markovian) process in phase space $\{\mathbf{r}(t), \mathbf{p}(t)\}$ solution of (3) is approximated by a Markov process; the corresponding Fokker–Planck equation (FPE) is used as an approximate evolution equation for the probability density of the exact process.

This method is based on the existence of a Markovian limit (Khas'minskii limit[13]) of the process when τ goes to zero (for a review see Ref. 14). It should be mentioned that this approach was initiated by Stratonovich,[15] independently of Lax.[12]

(b) In the Lax-6 method, using a stochastic Liouville equation, an exact generalized Fokker–Planck equation for the probability density in phase space is obtained.[16] This equation was also obtained by de la Peña-Auerbach and Cetto[17] by using the projection operator method or "smoothing method" of Frisch.[18]

From this exact generalized FPE we obtain an approximate second-order partial derivative equation of the "Fokker–Planck *type*" (see Ref. 16, Sec. III, and Ref. 19, Ref. 6c, Sec. II-B). The lack of uniqueness of equations obtained by the two methods (a) and (b) is due to the non-Markovian character mentioned above.

These equations may be reduced to a unique FPE for the stationary probability density in terms of some relevant constants of the deterministic motion corresponding to $\mathbf{F}(\mathbf{r})$ (Ref. 6c, Sec. III), using a method devised by H. Haken (Ref. 19, Sec. XI-C-2). For an isotropic multiperiodic system, due to the isotropic character of the damping and stochastic forces, such relevant constants of motion are those two action variables which are invariant under rotation (Ref. 20, Chap. II-2-a).

This reduced FPE can also be obtained directly by calculating, through perturbation methods, the variation of these "constants of motion" under the effect of the damping and stochastic forces, and by averaging these variations in order to get the drift and diffusion coefficients. The perturbation method may be applied directly to the equations of motion[21] or, following the main idea of Kubo's linear response theory,[22] it may be applied to the Liouville equation.[23]

For one-dimensional systems the reduced stationary FPE involves a single constant of motion, e.g., the energy E:

$$\frac{\partial}{\partial E}\left[G^{E}W_0 + G^{EE}\frac{\partial W_0}{\partial E}\right] = 0, \tag{5}$$

where G^{E} (G^{EE}) is the reduced drift (diffusion) coefficient. It should be noted that $W_0(E)$ is to be normalized in the whole phase space, i.e., $\int\int W_0(E)dq\,dp = 1$. Since the pair of variables $\{t,E\}$ is canonical, we have $dq\,dp = dt\,dE$; hence this normalization condition becomes

$$1 = \int\int W_0(E)dt\,dE = \int dE\, W_0(E)\oint_E dt = \int dE\, W_0(E)T(E),$$

where $T(E)$ denotes the period of the orbit with energy E. In other words, $W_0(E)$ is normalized with respect to the volume element $T(E)dE$ instead of dE alone (note that a similar situation is also encountered in statistical mechanics).

The probability current $J_0^{E} = G^{E}W_0 + G^{EE}(\partial W_0/\partial E)$ is therefore a constant, and if W_0 is to be integrable, it as well as its derivative must vanish at infinity, therefore, this constant must be zero, i.e., $J_0^{E} = 0$, hence

$$W_0(E) = N\exp\left[-\int^{E}\frac{G^{E}(E')}{G^{EE}(E')}dE'\right]. \tag{6}$$

In the case of a one-dimensional periodic system the coefficients G^{E}, G^{EE} are given by (Ref. 6c, Sec. III-C)

$$G^{E} = \tau\oint_E F\cdot\ddot{x}\,dt = \tau m\oint_E(\ddot{x})^2\,dt = 2\pi m\tau\omega^3\sum_{n=-\infty}^{+\infty} n^4|x_n|^2, \tag{7}$$

$$\begin{aligned} G^{EE} &= e^2\int_0^{\infty}\mathscr{B}_{\mathscr{E}}(\theta)\,d\theta\oint_E \dot{x}(t)\dot{x}(t-\theta)\,dt \\ &= 2\pi e^2\omega\int_0^{\infty}\mathscr{B}_{\mathscr{E}}(\theta)\,d\theta\sum_{n=-\infty}^{+\infty} n^2|x_n|^2e^{in\omega\theta} \\ &= \pi e^2\omega\sum_{n=-\infty}^{+\infty}\mathscr{S}_{\mathscr{E}}(n\omega)n^2|x_n|^2, \end{aligned} \tag{8}$$

where ω is the fundamental frequency of the system, $\mathscr{B}_{\mathscr{E}}(\theta)$ the correlation function of the electric field, and the x_n's are the Fourier components of the deterministic motion $\left[x(t) = \sum_{n=-\infty}^{+\infty} x_n e^{in\omega t}\right]$.

B. The stationary density and the averaged energy of the quartic anharmonic oscillator in SED

We consider now an anharmonic oscillator with a potential $V = \frac{1}{2}m\omega_0^2x^2 + \frac{1}{4}m\beta x^4$, where β is a small parameter.

In order to obtain the stationary density $W_0(E)$ from Eq. (6) we must calculate the Fourier components x_n and the fundamental frequency ω, appearing in the coefficients (7) and (8), as functions of E. To do this, we could use the Lindstedt–Poincaré method, which consists of eliminating the secular terms by expanding ω as well as x_n in powers of β (see, e.g., Ref. 24). In this way we should obtain x_n and ω as functions of x_0 (see Ref. 25 for a general algorithm to solve this problem). Using the resulting expression for x as a func-

tion of x_0, we have x_0 as a function of E and consequently x_n and ω as functions of E.

However, in our case it is possible to obtain x and ω in terms of elliptic functions.

Indeed, integrating the equation of motion along an orbit of energy E, we obtain

$$t=\int_x^{x_{\max}(E)}\frac{du}{[2/m(E-V(u))]^{1/2}}, \tag{9}$$

where we have taken the origin of time for

$$x(0)\equiv x_{\max}(E)=[(\omega_0^2/\beta)(-1+[1+(4E\beta/m\omega_0^4)]^{1/2})]^{1/2}$$

maximum value of x for a given energy E. Making in (9) the change of variable $s=\arccos[u/x_{\max}(E)]$, we get

$$t=\int_0^{\phi}\frac{ds}{\omega_0(1+4y)^{1/4}}\frac{1}{[1-k^2\sin^2 s]^{1/2}}, \tag{10}$$

where $y=(E\beta/m\omega_0^4)$, $k^2=\frac{1}{2}(1-(1+4y)^{-1/2})$, and $\phi=\arccos[x/x_{\max}(E)]$. The period T corresponds to making $\phi=2\pi$ in expression (10), hence

$$T=\frac{4}{\omega_0}\frac{\mathbb{K}(k)}{(1+4y)^{1/4}}, \tag{11}$$

where

$$\mathbb{K}(k)=\int_0^{\pi/2}\frac{ds}{[1-k^2\sin^2 s]^{1/2}} \tag{12}$$

is the complete elliptic integral of the first kind.

$\mathbb{K}$ is an analytic function for $|k|<1$ (Ref. 26, 21.72) given by the expansion (Ref. 27, Chap. 6, Sec. 3).

$$\mathbb{K}(k)=\frac{\pi}{2}\left(1+(\tfrac{1}{2})^2k^2+\cdots+\left[\frac{(2n-1)!!}{2^n n!}\right]^2k^{2n}+\cdots\right). \tag{13}$$

This function has no zeros at least for $|k|<a=0.94$, because $|\mathbb{K}(a)|<2|\mathbb{K}(0)|$. Therefore $\omega=(2\pi/T)$ is an analytic function of k for $|k|<a$, and then of y for $|y|<\frac{1}{4}(4a^4-2)/(4a^4-1)$ (the calculation is easy but somewhat lengthy). Using (13) and (11) we obtain

$$\omega=\omega_0(1+\tfrac{3}{4}y-\tfrac{69}{64}y^2+O(y^3)). \tag{14}$$

In this way we have obtained ω as a function of E. Now, from (15) we have $x=x_{\max}(E)\mathrm{cn}(\omega_0 t(1+4y)^{1/4},k)$, where $\mathrm{cn}(u,k)$ (with $u=\omega_0 t(1+4y)^{1/4}$) is the Jacobi elliptic function (see Ref. 26, Chap. XXII). This function may be written as the following Fourier series (Ref. 27, Chap. 6, Sec. 19):

$$x=x_{\max}(E)\mathrm{cn}(u,k)=4\left(\frac{2}{\beta}\right)^{1/2}\omega\sum_{n=0}^{\infty}\frac{q^{n+\frac{1}{2}}}{1+q^{2n+1}}\cos(2n+1)\omega t, \tag{15}$$

where $q=\exp[-\pi(\mathbb{K}(k')/\mathbb{K}(k))]$, $k'=(1-k^2)^{1/2}$.

q is an analytic function of k for $|k|<1$ (Ref. 26, 21.712) and therefore an analytic function of y for $|y|<\frac{1}{6}$ (this value is obtained by putting $a=1$ in the expression $(\frac{1}{4})(4a^4-2)/(4a^4-1)$ given above for the case $|k|<a$). Using an expansion due to Hermite we may express q as a function of k^2 (Ref. 27, Chap. 6, Sec. 20):

$$q=(k^2/2^4)|1+(k^2/2)+\tfrac{21}{64}k^4+O(k^6)|, \tag{16}$$

and therefore as a function of y,

$$q=(y/2^4)|1-\tfrac{5}{2}y+\tfrac{469}{64}y^2+O(y^3)|. \tag{17}$$

From Eqs. (7), (8), and (15) we have the reduced coefficients given by

$$G^E=32\pi m\tau\frac{\omega^5}{\beta}\sum_{n=0}^{\infty}(2n+1)^4\frac{q^{2n+1}}{[1+q^{2n+1}]^2}, \tag{18}$$

$$G^{EE}=16\pi e^2\frac{\omega^3}{\beta}\sum_{n=0}^{\infty}\mathscr{S}_z[(2n+1)\omega]\frac{(2n+1)^2q^{2n+1}}{[1+q^{2n+1}]^2}, \tag{19}$$

where we have used $\mathscr{S}_z(\omega)=\mathscr{S}_z(-\omega)$.

These coefficients, (18) and (19), are analytic functions of y for $|y|<\frac{1}{4}(4a^4-2)/(4a^4-1)$ because, on the one hand, ω is an analytic function of y for $y<(1/4)(4a^4-1)/(4a^4-2)$, as we saw above, and on the other hand, $|q|<1$ for $|k|<1$. The proof of the latter property is as follows.

In Ref. 26, Sec. 21.7, it is shown that $|q(k)|<1$, when a cut from 0 to $-\infty$ and from 1 to $+\infty$ is made in the complex plane of k^2, because in this region $\mathrm{Re}[\ln q(k)]<0$. Moreover, it is easy to see directly from the integral expressions of $\mathbb{K}(k)$ that this property [namely $|q(k)|<1$] is also true for $-\infty<k^2<0$. Therefore, $|q|<1$ when $|k|<1$, Q.E.D. Therefore, we have

$$\frac{G^E}{G^{EE}}=\frac{2m\tau\omega^2}{e^2\mathscr{S}_z(\omega)}\left[1+\tfrac{9}{256}y^2\left(9-\frac{\mathscr{S}_z(3\omega)}{\mathscr{S}_z(\omega)}\right)+O(y^3)\right]. \tag{20}$$

Using the spectral density (1) of the background field in SED and the expansion (14) for ω, in (20) we obtain

$$\frac{G^E}{G^{EE}}=\frac{2}{\hbar\omega_0}|1-\tfrac{3}{4}y+\tfrac{129}{128}y^2+O(y^3)|. \tag{21}$$

From expression (6) and (21) we obtain through order β^2

$$W_0(E)=N\exp[-2E/\hbar\omega_0]\left[1+\tfrac{3}{4}\beta\frac{E^2}{\hbar m\omega_0^5}+\frac{\beta^2E^3}{\hbar m^2\omega_0^9}\left(-\frac{43}{64}+\frac{E}{\hbar\omega_0}\frac{9}{32}\right)\right], \tag{22}$$

where N is the normalization constant in the phase space (see below). This expression (22) is valid at least for energies such that $(E^3\beta/\hbar^2m\omega_0^6)\ll 1$ [i.e., $y\ll(\hbar\beta/m\omega_0^3)^{2/3}$].

For extremely large energies ($y\gg 1$), for example $y\geqslant 1/\beta$, we may obtain the behavior of $W_0(E)$ from the coefficients (18) and (19). Using (1) as spectral density, according to (18) and (19) G^E/ω^5 and G^{EE}/ω^6 are analytic functions of q, when $|q|<1$, because the series involved in these coefficients are uniformly convergent in this region. Moreover, q and $\omega y^{-1/4}$ [see (11)] are analytic functions of $z=y^{1/2}$. Therefore, $G^E y^{-5/4}$ and $G^{EE}y^{-6/4}$ are analytic functions of z. The region of analyticity is given by $|k^2-\frac{1}{2}|<(a^2-\frac{1}{2})$, which corresponds to $|z|<\frac{1}{2}[1+(1/(2a^2-1)^2)]^{-1/2}$. We therefore have

$$\frac{G^E}{G^{EE}}=\frac{4(4y)^{-1/4}}{\pi\hbar\omega_0}\mathbb{K}(1/\sqrt{2})\left\{\sum_{n=0}^{\infty}(2n+1)^4\frac{e^{-(2n+1)\pi}}{[1+e^{-(2n+1)\pi}]^2}\right/\sum_{n=0}^{\infty}(2n+1)^5\frac{e^{-(2n+1)\pi}}{[1+e^{-(2n+1)\pi}]^2}\Bigg\}[1+O(y^{-1/2})], \tag{23}$$

where we have used $k(y=\infty)=1/\sqrt{2}$ and $q(y=\infty)=e^{-\pi}[\mathbb{K}(1/\sqrt{2})/\mathbb{K}(1/\sqrt{2})]=e^{-\pi}$. Hence from

expression (6) we obtain for large energies ($y \gg 1$)

$$W_0(E) \simeq \text{const} \cdot \exp[-(Cm^{1/4}/\hbar\beta^{1/4})E^{3/4}], \tag{24}$$

whereas from small energies we have from (22) the following behavior:

$$W_0(E) \simeq (1/\hbar\pi)\exp[-2E/\hbar\omega_0]. \tag{25}$$

For the intermediate range of energies, i.e., $(\hbar\beta/m\omega_0^3)^{-d} < E/\hbar\omega_0 < (\hbar\beta/m\omega_0^3)^{-1}$, with $0 < d < 1/3$, $W_0(E)$ is bounded by its value at the lower energy, namely $\hbar\omega_0(\hbar\beta/m\omega_0^3)^{-d}$, due to the fact that $W_0(E)$ is a monotonous decreasing function of E ($G^E/G^{EE} \geqslant 0$). This value may be obtained from Eq. (22), because this expression is valid at least for energies such that $E/\hbar\omega_0 \ll (\hbar\beta/m\omega_0^3)^{-1/3}$.

Thereafter, for any integration involving $W_0(E)$, we may decompose the integration interval $(0, \infty)$ into three parts. For the part corresponding to large energies ($y \geqslant 1/\beta$), we find a negligible contribution (smaller than any power of β), due to the behavior of $W_0(E)$ given by Eq. (24). The same result is found for the second part (intermediate energies) by using the bound given above. Therefore, the contribution of the range of energies such that $E/\hbar\omega_0 \gtrsim (\hbar\beta/m\omega_0^3)^{-b}$ ($b > 0$) will be negligible.

Then we may use Eq. (22) for obtaining through order β^2 the normalization constant, namely $N = (1/\hbar\pi) \times [1 - \frac{81}{256}(\hbar\beta/m\omega_0^3)^2]$ and the mean energy $\bar{E}$:

$$\bar{E} = \frac{\hbar\omega_0}{2}\left[1 + \tfrac{3}{8}\frac{\hbar\beta}{m\omega_0^3} + \tfrac{141}{256}\left(\frac{\hbar\beta}{m\omega_0^3}\right)^2\right] \tag{26}$$

which differs, even in the sign, from the quantum result [Ref. 28, formula (1.17)] at the second order in β: $\bar{E}_2(QT) = -\frac{21}{128}(\hbar\beta/m\omega_0^3)^2\hbar\omega_0$. However, the first-order perturbation energy $\bar{E}_1 = \frac{3}{16}\beta(\hbar^2/m\omega_0^2)$ is the same according to SED and quantum theory, but it must be pointed out that this result is not very specific[29]: whatever the damping force and the spectral density of the fluctuating force [provided they imply an integrable stationary density $P_0(x)$ for the unperturbed harmonic oscillator], the first-order correction to the energy is given by[29]

$$\bar{E}_1 = \int V_1(x)P_0(x)\,dx,$$

and this coincides with the quantum expression, provided $P_0(x)$ is the same Gaussian position density as in quantum theory (this is actually the case for SED).

III. THE ABSORPTION COEFFICIENT

Knowing $W_0(E)$, we may apply the Kubo linear response theory[22] in order to study the energy absorption of the anharmonic oscillator in the stationary state from an external electromagnetic field.

For a multiperiodic system in SED the absorption coefficient (ratio of the absorbed power to the incident power) $a(\omega)$ may be obtained in the form of a series over all harmonics.[30]

In the one-dimensional case the system is simply periodic, and this coefficient is given by

$$a(\omega) = -\frac{(2\pi)^3}{3}\omega e^2 \int_0^\infty dE \frac{dW_0}{dE} \sum_{n=-\infty}^{+\infty} n|x_n|^2 \delta(\omega - n\omega'(E)) = +\frac{(2\pi)^3}{3}\omega e^2 \sum_{n=0}^{\infty}\left[|x_n|^2 \frac{G_E}{G^{EE}}\frac{dE}{d\omega'}W_0\right]_{\omega'=\omega/n} \tag{27}$$

for $\omega \geqslant 0$, where we have used (6). In this expression $[\ ]_{\omega'=\omega/n}$ means that we take the energy which corresponds to $\omega' = \omega/n$.

In our case, namely an even potential for which the fundamental frequency ω' has a minimum value ω_0 (corresponding to $E = 0$), only a finite number of odd terms contribute to $a(\omega)$. Furthermore, we can restrict ourselves to just one term, because $a(\omega)$ is not negligible only for frequencies very close to $(2n+1)\omega_0$.

The essential point is that, for frequencies such that $\omega - \omega_0 \gg (\hbar\beta/m\omega_0^2)$, $W_0(E(\omega)) \ll (1/\hbar)(\hbar\beta/m\omega_0^3)^p\ \forall\ p \geqslant 0$, whereas G^E/G^{EE}, $dE/d\omega$, and $|x_n|^2$ are bounded by powers of β.

In order to prove this property, we consider first $dE/d\omega$, which is given by $dE/d\omega = -(2\pi/\omega^2)(dT/dE)^{-1}$. From (11) we have

$$\frac{dT}{dE} = \frac{4}{\omega_0}\left[\frac{d\mathbb{K}}{dk}\frac{dk}{dE} - \frac{\mathbb{K}}{1+4y}\frac{\beta}{m\omega_0^4}\right](1+4y)^{-1/4}. \tag{28}$$

Then, using in (28)

$$\frac{dk}{dE} = \frac{1}{2k}\frac{\beta}{m\omega_0^4}(1+4y)^{-3/2}$$

and

$$\frac{d\mathbb{K}}{dk} = \frac{\mathbb{E}}{kk'^2} - \frac{\mathbb{K}}{k}$$

(Ref. 27, Chap. 6 Sec. 3), where $\mathbb{E}$ is the complete elliptic integral of the second kind, we obtain

$$\frac{dT}{dE} = \frac{4\beta}{m\omega_0^5}\left[\frac{(1+4y)^{-1/2}}{2k}\left(\frac{\mathbb{E}}{kk'^2} - \frac{\mathbb{K}}{k}\right) - \mathbb{K}\right](1+4y)^{-5/4}; \tag{29}$$

but using in (29) $\mathbb{K} \geqslant \pi/2$ and $E \leqslant \pi/2$ (Ref. 27, Chap. 6, Sec. 3), we have

$$\frac{dT}{dE} \leqslant \frac{\pi\beta}{m\omega_0^5}(1+4y)^{-5/4}\left(-\frac{1}{k'^2}\right) \leqslant 0 \tag{30}$$

and then $d\omega/dE \geqslant 0$ and using $\omega \geqslant \omega_0$ and (30),

$$\frac{dE}{d\omega} \leqslant \frac{m\omega_0^3}{\beta}(1+4y)^{3/4}[1 + (1+4y)^{1/2}]. \tag{31}$$

On the other hand, we have from (18) and (19) $G^E/G^{EE} \leqslant (2/\hbar\omega_0)$ and

$$|x_n|^2 = \left|\frac{1}{T}\int_0^{2\pi/\omega} \exp(-i\omega\theta n)x(\theta)\,d\theta\right|^2 \leqslant |x_{\max}|^2 = \frac{\omega_0^2}{\beta}[-1 + (1+4y)^{1/2}].$$

Therefore, we have shown that G^E/G^{EE}, $|x_n|^2$, and $dE/d\omega$ are bounded by powers of β. Q.E.D. Now, from (14) we may obtain E as a function of ω:

$$E(\omega)=\frac{4}{3}\frac{m\omega_0^4}{\beta}\frac{\omega-\omega_0}{\omega_0}\left[1+\frac{23}{12}\left(\frac{\omega-\omega_0}{\omega_0}\right)+\frac{107}{72}\left(\frac{\omega-\omega_0}{\omega_0}\right)^2+O\left(\left(\frac{\omega-\omega_0}{\omega_0}\right)^3\right)\right]. \quad (32)$$

This inverse function $E(\omega)=(1/\beta)y(\omega)$ is analytic around ω_0 with a convergence radius independent of β. Then, for β small enough, it will be analytic for frequencies such that $|(\omega-\omega_0)/\omega_0| \lesssim (\hbar\beta/m\omega_0^3)^{1-c}$, for some c between 0 and 1. In this region, we have from expression (32) $E/\hbar\omega_0 \lesssim (\hbar\beta/m\omega_0^3)^{-c}$, and therefore, using the monotonous increasing character of E as a function of ω, for ω such that $(\omega-\omega_0)/\omega_0 \gtrsim (\hbar\beta/m\omega_0^3)^{1-c}$, we shall have $E/\hbar\omega_0 \gtrsim (\hbar\beta/m\omega_0^3)^{-c}$. Now, from the monotonous decreasing character of $W_0(E)$ we obtain for this range of energies

$$W_0(E) \lesssim \frac{1}{\hbar\pi}\exp\left[-2\left(\frac{\hbar\beta}{m\omega_0^3}\right)^{-c}\right] \ll \left(\frac{\hbar\beta}{m\omega_0^3}\right)^p\frac{1}{\hbar},$$

$\forall p \geqslant 0$. Thus we have achieved the proof that $W_0(E(\omega))$ is much smaller than any positive power of β for frequencies ω such that $\omega-\omega_0 \gg \hbar\beta/m\omega_0^2$.

Therefore, only for frequencies $\omega=(2n+1)\omega'$ such that $\omega' \geqslant \omega_0$ and $(\omega'-\omega_0)/\omega_0 \lesssim (\hbar\beta/m\omega_0^3)$, the terms in (27) are not negligible, and the contribution to this range of frequencies may come only from the term

$$\left[|x_{2n+1}|^2\frac{G^E}{G^{EE}}\frac{dE}{d\omega'}W_0\right]_{\omega'=\omega/(2n+1)}$$

of the sum (27). Then the first absorption line of the spectrum (frequencies very close to ω_0) is given by

$$a_1(\omega) = +\frac{8\pi^3}{3}e^2\omega\left[|x_1|^2\frac{G^E}{G^{EE}}\frac{dE}{d\omega'}W_0\right]_{\omega'=\omega} \quad (33)$$

Using (21), (22), (32), and $|x_1|^2$ obtained from (15) and (17) as a function of y, we obtain for frequencies $\omega \geqslant \omega_0$ such that $(\omega-\omega_0)/\omega_0 \ll (\hbar\beta/m\omega_0^3)^{2/3}$, the following expression for $a_1(\omega)$:

$$a_1(\omega) \doteq \frac{128}{27}\frac{\pi^2 m\omega_0^5 e^2}{\beta^2\hbar^2}N\Delta\left[1+\frac{17}{4}\Delta+\frac{719}{144}\Delta^2-\frac{34}{9}\frac{\delta\Delta^2}{\beta}-\frac{33}{2}\frac{\delta\Delta^3}{\beta}+\frac{578}{81}\frac{\delta^2\Delta^4}{\beta^2}\right]\times\exp\left[-\frac{8}{3}\frac{\delta\Delta}{\beta}\right], \quad (34)$$

where $\delta=m\omega_0^3/\hbar$ and $\Delta=(\omega-\omega_0)/\omega_0$.

The maximum of this first absorption line, namely $\omega_1=\omega_0+\frac{3}{8}\hbar\beta/m\omega_0^2$, does not coincide (already to first order in β) with the frequency of the first quantum transition [Ref. 28, formula (1.17)]:

$$\omega_1(QT)=\omega_0+\frac{3}{4}\frac{\beta\hbar}{m\omega_0^2}.$$

The power (denoted I_{a_1}) absorbed from the background field of SED in the frequency range corresponding to this first absorption line is given (up to terms of order β^2) by the expression

$$I_{a_1}=\int_{\omega_0}^{\omega_0+A}a_1(\omega)\rho(\omega)\,d\omega = \frac{\hbar\omega_0^3e^2}{3mc^3}\left[1+\frac{9}{4}\frac{\beta\hbar}{m\omega_0^3}-\frac{567}{512}\left(\frac{\beta\hbar}{m\omega_0^3}\right)^2\right], \quad (35)$$

where $\rho(\omega)=(3/4\pi^2)\mathcal{S}_z(\omega)=(\hbar/2\pi^2c^3)\omega^3$ is the spectral energy density of the random field (see, e.g., Ref. 3, Sec. 3.A).

The contribution of the range of frequencies $3\omega_0 > \omega \gtrsim \omega_0[1+(\hbar\beta/m\omega_0^3)^{1-c}]$ $(0<c<1)$ to I_{a_1} is negligible, because the integrand may be bounded, as we have shown above, by a power of β times a term coming from the probability density, which is given (due to the monotonous increasing character of $W_0[E(\omega)]$ as a function of ω) by

$$W_0(E) \lesssim \frac{1}{\hbar\pi}\exp\left[-2\left(\frac{\hbar\beta}{m\omega_0^3}\right)^{-c}\right] \ll \left(\frac{\hbar\beta}{m\omega_0^3}\right)^p\frac{1}{\hbar},$$

$\forall p \geqslant 0$. Therefore, the integral practically reaches its limit for $A \gg \omega_0(\hbar\beta/m\omega_0^3)$, and taking $A \ll \omega_0(\hbar\beta/m\omega_0^3)^{2/3}$, we may use (34) for getting (35).

Consequently, the linewidth is of order β, which is in contradiction with the experimental result of the existence of sharp transition frequencies (spectral lines), whose linewidth is not related with β. This result supports the view that it really appears nontrivial, in the framework of a classical stochastic theory such as SED, to account for sharp spectral lines.

In a similar way, we obtain the absorption coefficient for frequencies close to $3\omega_0$ (second "absorption line"):

$$a_3(\omega) = +\frac{8\pi^3}{3}e^2\omega\left[|x_3|^2\frac{G^E}{G^{EE}}\frac{dE}{d\omega'}W_0\right]_{\omega'=\omega/3} = \frac{8\pi^2}{81}\frac{m\omega_0^5e^2}{\beta^2\hbar^2}N\left(\frac{\omega-3\omega_0}{3\omega_0}\right)^3\times\exp\left[-\frac{8}{3}\frac{\delta}{\beta}\left(\frac{\omega-3\omega_0}{3\omega_0}\right)\right]\Bigg\}. \quad (36)$$

The maximum of this second line is again not the same as the one in quantum theory [Ref. 28, formula (1.17)]:

$$\omega_3^{\mathrm{SED}}=3\omega_0+\frac{27}{8}\frac{\hbar\beta}{m\omega_0^2}, \quad \omega_3^{\mathrm{QT}}=3\omega_0+\frac{9}{4}\frac{\hbar\beta}{m\omega_0^2}.$$

Proceeding as we did for the first line, we find that the absorbed power for the second line I_{a_3} is given by

$$I_{a_3}=\int_{3\omega_0}^{3\omega_0+A}a_3(\omega)\rho(\omega)\,d\omega=\frac{\hbar\omega_0^3e^2}{3mc^3}\left[\frac{729}{512}\left(\frac{\beta\hbar}{m\omega_0^3}\right)^2\right]. \quad (37)$$

We then develop the same kind of argument as for Eq. (35), and we find again a width of order β for this second line, because the integral in (37) converges for $A \gg 3\omega_0(\hbar\beta/m\omega_0^3)$. Finally, we may obtain the total absorbed power I_a through order β^2 from the contributions (35) and (37) of the first two absorption lines, because the contribution of the other lines to I_a is, at least of order β^4, due to the fact that $|x_{2n+1}|^2 = O(y^2)\,|x_{2n-1}|^2$. Therefore, by adding I_{a_1} and I_{a_3}, we get (up to terms of order β^2)

$$I_a=\frac{\hbar\omega_0^3e^2}{3c^3}\left[1+\frac{9}{4}\frac{\beta\hbar}{m\omega_0^3}+\frac{81}{256}\left(\frac{\beta\hbar}{m\omega_0^3}\right)^2\right]. \quad (38)$$

IV. ENERGY EMISSION AND THE PROBLEM OF "RADIATION BALANCE" IN SED

Thus far we have studied the absorption of energy. Now, in order to see whether the radiation balance at each frequency is satisfied or not for the anharmonic oscillator in SED, we calculate the emission of energy at each frequency.

From the total averaged emitted power,

$$I_e = -\langle \dot{x}F_d \rangle = \int_0^\infty dE\,\frac{2\pi}{\omega}W_0(E)\sum_{n=0}^{\infty}\frac{4e^2}{3c^3}(n\omega)^4|x_n|^2, \quad (39)$$

where $F_d = (2e^2/3c^3)\ddot{x}$ is the damping force, we may obtain the emitted power at a frequency $\omega \geqslant 0$ as given by

$$I_e(\omega) = \frac{8\pi e^2}{3c^3}\omega^3\sum_{n=0}^{\infty}\left[|x_n|^2\frac{dE}{d\omega'}W_0(E)\right]_{\omega'=\omega/n} \quad (40)$$

(see Ref. 31, Sec. 16 for the calculation of the total emitted power corresponding to one orbit). It should be noticed that, in order to get the correct result, given by (39), the halved sum (over *positive* frequencies), appearing in expressions (43), (44), and (46) of Ref. 31, Sec. 16, must be *multiplied* by 2 instead of being divided by 2. Moreover, in Van Vleck's work,[31] the symbol ω is used as a frequency, not as an angular frequency: $2\pi\omega$ (Van Vleck) $= \omega$ (present work).

In the same way as we did for $a(\omega)$, it may be shown that only for frequencies very close to $(2n+1)\omega_0$ the emission is not negligible. Therefore, for frequencies close to ω_0 (first "absorption line"), we have

$$\begin{aligned} I_{e_1}(\omega) &= \frac{8\pi}{3c^3}\omega^3\left[|x_1|^2\frac{dE}{d\omega'}W_0(E)\right]_{\omega'=\omega} \\ &= \frac{64}{27}\frac{m\omega_0^5e^2}{\beta^2\hbar c^3}N\omega^3\Delta\left[1+\frac{17}{4}\Delta+\frac{881}{144}\Delta^2\right. \\ &\quad \left.-\frac{34}{9}\frac{\delta\Delta^2}{\beta}-\frac{33}{2}\frac{\delta\Delta^3}{\beta^2}+\frac{578}{81}\frac{\delta^2\Delta^4}{\beta^2}\right] \\ &\quad \times\exp\left[-\frac{8}{3}\frac{\delta\Delta}{\beta}\right]. \end{aligned} \quad (41)$$

We note that there is *not* radiation balance at each frequency; the emitted power $I_{e_1}(\omega)$ and the absorbed power $I_{a_1}(\omega) = a_1(\omega)\rho(\omega)$ [see (34)] are not the same at each frequency. In actual fact, for $\Delta = (\omega-\omega_0)/\omega_0 = O(\hbar\beta/m\omega_0^3)$, $I_{e_1}(\omega) - I_{a_1}(\omega) = O[(e^2\hbar/m^2\omega_0)\beta]$.

For the second absorption line (frequencies close to $3\omega_0$), we have

$$I_{e_3}(\omega) = \tfrac{4}{243}\frac{m\omega_0^5e^2}{\hbar\beta^2c^3}N\omega^3\left(\frac{\omega-3\omega_0}{3\omega_0}\right)^3\exp\left\{-\tfrac{8}{3}\frac{\delta}{\beta}\left(\frac{\omega-3\omega_0}{3\omega_0}\right)\right\} \quad (42)$$

and therefore the radiation balance is again *not* satisfied for

$$\frac{\omega-3\omega_0}{3\omega_0} = O\left(\frac{\hbar\beta}{m\omega_0^3}\right),$$

$$I_{a_3}(\omega) - I_{e_3}(\omega) = 2I_{e_3}(\omega) = O\left(\frac{e^2\hbar}{m^2\omega_0}\beta\right).$$

These results show that the quartic anharmonic oscillator in the random field of SED absorbs energy at the frequencies close to $3\omega_0$, and radiates energy at the frequencies close to ω_0.

Following the same kind of argument developed for getting a_1 and I_{a_1} [Eqs. (35) and (37)] (the required integration interval is of the same order of magnitude) we get the total emitted power for the two first lines:

$$\begin{aligned} I_{e_1} &= \frac{\hbar\omega_0^3e^2}{3mc^3}\left[1+\frac{9}{4}\frac{\beta\hbar}{m\omega_0^3}-\frac{81}{512}\left(\frac{\beta\hbar}{m\omega_0^3}\right)^2\right], \\ I_{e_3} &= \frac{\hbar\omega_0^3e^2}{3mc^3}\left[\frac{243}{512}\left(\frac{\beta\hbar}{m\omega_0^3}\right)^2\right]. \end{aligned} \quad (43)$$

We recover the conservation of energy $I_{a_1}+I_{a_3} = I_{e_1}+I_{e_3}$, through order β^2. The amount of energy flowing towards the oscillator at frequencies close to $3\omega_0$ exactly balances that flowing away at frequencies close to ω_0, as it should be, since the *total* energy balance must obviously be satisfied as soon as the system is in its stationary state.

The quartic anharmonic oscillator is then *not* in *radiative* equilibrium with the zero-point radiation of SED.

Moreover, this system is in equilibrium only with the Rayleigh–Jeans spectrum. Using a general spectral density $\mathscr{S}_x(\omega)$ we obtain, for frequencies close to ω_0 to order β^2, from expressions (20), (33), and (40),

$$\begin{aligned} I_{a_1}(\omega) - I_{e_1}(\omega) &= \frac{8\pi}{3c^3}e^2\omega^3\left[|x_1|^2\frac{dE}{d\omega'}W_0\right]_{\omega'=\omega} \\ &\quad \times\left[\frac{9}{256}\gamma^2\left(9-\frac{\mathscr{S}_x(3\omega)}{\mathscr{S}_x(\omega)}\right)\right]. \end{aligned} \quad (44)$$

Therefore, in order to have radiation balance for this first line we must impose $\mathscr{S}_x(3\omega)/\mathscr{S}_x(\omega) - 9 = 0$ and hence $\mathscr{S}_x(\omega) = \text{const}\omega^2$, which is the Rayleigh–Jeans law. In order to get this result, it is enough to assume that $\mathscr{S}_x(\omega)$ is twice continuously differentiable. Indeed, by differentiating twice the relation $\mathscr{S}_x(3\omega) = 9\mathscr{S}_x(\omega)$, we get $\mathscr{S}_x''(3\omega) = \mathscr{S}_x''(\omega)$. Then, by using a sequence of frequencies $\{\omega/3^n\}$, with $n = 1,2,\ldots$, we get $\omega = 0$ as an accumulation point of this sequence of points for which $\mathscr{S}_x''(\omega)$ is constant. Therefore, $\mathscr{S}_x''(\omega)$ must be constant on a finite neighborhood of $\omega = 0$, and then we obtain by a suitable dilation $\mathscr{S}_x(\omega) = \text{const}\omega^2$ for any ω.

Note that the expression (44) is valid for any $\mathscr{S}_x(\omega)$. The only restriction is that $W_0(E)$ must be intergrable. In order to insure this property it is sufficient that $\mathscr{S}_x(\omega)$ behaves as a power of ω for large frequencies, because in this case we have always the result given by expression (24).

Using the Rayleigh–Jeans spectrum we may show that the radiation balance is always satisfied for any one-dimensional periodic system. Indeed, from (7) and (8), by using $\mathscr{S}_x(\omega) = C\omega^2$, we have $G_E/G_{EE} = 2\tau m/e^2C$ and then from (27) and (40)

$$\begin{aligned} I_a(\omega) - I_e(\omega) &= \frac{8\pi e^2}{3c^3}\omega^3\sum_{n=0}^{\infty}\left[|x_n|^2\frac{dE}{d\omega'}W_0(E)\right]_{\omega'=\omega/n} \\ &\quad \times\left(\pi^2c^3\frac{G_E}{G_{EE}}\frac{3C}{4\pi^2}-1\right)=0, \end{aligned}$$

where we have used $I_a(\omega) = a(\omega)\rho(\omega) = a(\omega)(3C/4\pi^2)\omega^2$.

Therefore, the radiation balance at each frequency is satisfied for any system for the Rayleigh–Jeans spectrum, and only for it. This result is in agreement with the one obtained by Boyer[32] for a harmonic oscillator perturbed by a cubic potential βx^3. In this case, however, it could be argued that no strictly stationary distribution exists, since the poten-

tial goes to $-\infty$ when $x \to -\infty$: thus the particle sooner or later escapes to $-\infty$, over the potential barrier of finite height h corresponding to the maximum of the potential curve which lies on the negative half of the x axis. As a consequence, there is only a *quasistationary* probability distribution (whose lifetime becomes longer and longer when the cubic perturbing term βx^3 goes to zero).

But no such objection is possible for the harmonic oscillator perturbed by a quartic potential, since in that case the problem admits in SED a genuine stationary density, as we have proved in Sec. II B above.

V. CONCLUSIONS

We have found that the results for the anharmonic oscillator, concerning the mean energy and the maximum absorption frequencies, are in disagreement with the quantum results as soon as $\beta \neq 0$.

Moreover, independently of this disagreement with quantum theory, the lack of radiation balance that we have found in SED is a serious defect of the theory, since the ratio of absorption and emission coefficients would not be the same function of the frequency for the anharmonic and harmonic oscillator and therefore Kirchhoff's law would be violated in SED (by contrast this law is actually satisfied in quantum theory[33]).

T. H. Boyer (Ref. 32a, Sec. VII; Ref. 32b, Sec. VI) conjectures that a relativistic mechanical system should leave invariant the Lorentz-invariant spectrum of classical electromagnetic zero-point radiation appearing in SED. Then, the Rayleigh–Jeans law found here as the only equilibrium spectrum would be tied specifically to the nonrelativistic character of the systems.

T. W. Marshall[34] suggests that the mechanical system is in equilibrium in the zero-point field of SED, while the field itself is not in equilibrium. The variables describing the mechanical system are "fast" ones with respect to those describing the field. Then the system realizes its stationary density according to the nearly instantaneous value of the field. Such a picture is perfectly reasonable, but it does not avoid the difficulty concerning Kirchhoff's law. Moreover, even if the problems concerning the lack of radiation balance in SED could be avoided, we would always be faced with the more difficult problem posed by the well established experimental evidence of the sharpness of spectral lines. It really appears nontrivial, in the framework of a classical stochastic theory such as SED, to account for these very sharp spectral lines.

These difficulties, together with the negative results for the hydrogen atom in SED,[10,11] make us adopt, at the present time, a very cautions opinion concerning the prospects of "simple" classical stochastic models in microphysics such as SED. By the qualification simple, we mean specifically that the stochastic field acts like a "thermostat", i.e., its spectrum is independent of the mechanical system; but, as pointed out, e.g., by Van Kampen (Ref. 16, Introduction), such an independence property is by no means trivial. Thus, less simple theories could be considered, for which the "effective" spectrum of the stochastic force would be dependent on the mechanical system upon which it is acting. In more physical terms, the reaction of the mechanical system upon the stochastic medium surrounding it would not be negligible. Such more general theories would be more involved, but at the same time they could offer possibilities of explanation for a wider class of physical properties and behaviors.

ACKNOWLEDGMENT

We acknowledge financial support from the Instituto de Estudios Nucleares, Madrid.

[1] T. H. Boyer, Phys. Rev. D **11**, 790 (1975).
[2] T. W. Marshall, Proc. R. Soc. London, Ser. A **276**, 475 (1963).
[3] P. Claverie and S. Diner, Int. J. Quantum Chem. **12**, Suppl. 1, 41 (1977).
[4] E. Santos, (a) Nuovo Cimento B **22**, 201 (1974); (b) B **19**, 57 (1974).
[5] G. N. Plass, Rev. Mod. Phys. **33**, 37 (1961).
[6] (a) P. Claverie, L. de la Peña-Auerbach, and S. Diner, "Stochastic electrodynamics of non-linear systems. II. Derivation of a reduced Fokker-Planck in terms of relevant constants of motion" (to be published); (b) P. Claverie, in *Dynamical Systems and Microphysics* edited by A. Blaquiere, F. Fer, and A. Marzollo, CISM courses and lectures No. 261 (Springer, Berlin, 1980), p. 111; (c) P. Claverie, "Stochastic Electrodynamics versus Quantum Theory: recent advances in the study of non-linear systems." [Contribution to the "Einstein Centennial Symposium on Fundamental Physics" Bogota, 1979 (in press)].
[7] L. de la Peña-Auerbach and A. M. Cetto, (a) Rev. Mex. Fis. **25**, 1 (1976); (b) J. Math. Phys. **20**, 469 (1979).
[8] P. Julg, Folia Chimica Theoretica Latina **6**, 99 (1978).
[9] L. Pesquera, in *Dynamical Systems and Microphysics*, edited by A. Blaquière, F. Fer, and A. Marzollo, CISM courses and lectures No. 261 (Springer, Berlin, 1980), p. 145.
[10] P. Claverie, L. Pesquera, and F. Soto, Phys. Lett. A **80**, 113 (1980).
[11] P. Claverie and F. Soto, J. Math. Phys. **23**, 753 (1982).
[12] M. Lax, Rev. Mod. Phys. **38**, 541 (1966).
[13] R. Z. Khas'minskii, Theory Prob. Appl. **11**, 390 (1966).
[14] G. C. Papanicolaou, Lect. Appl. Math. **16**, 109 (1977).
[15] R. L. Stratonovich, *Conditional Markov Processes and Their Application to the Theory of Optimal Control* (American Elsevier, New York, 1968), Chap. 1 [Russian original (Moscow University Press, Moscow, 1966)].
[16] N. G. Van Kampen, Phys. Rep. (Phys. Lett. C) **24**, 171 (1976).
[17] L. de la Peña-Auerbach and A. M. Cetto, J. Math. Phys. **18**, 1612 (1977).
[18] U. Frisch, in *Probabilistic Methods in Applied Mathematics*, edited by A. T. Bharucha-Reid (Academic, New York, 1968), p. 75.
[19] H. Haken, Rev. Mod. Phys. **47**, 67 (1975).
[20] L. Pesquera, "Etude des Equations Différentielles Stochastiques non Markoviennes et Application à l'Electrodynamique Stochastique." Thèse (Doctorat de 3ème Cycle) (Paris, 1980).
[21] (a) T. W. Marshall (unpublished work); (b) T. W. Marshall and P. Claverie, J. Math. Phys. **21**, 1819 (1980).
[22] R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957).
[23] L. Pesquera and P. Claverie, "Derivation of Fokker–Planck equations through response theory" (to be published).
[24] (a) H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Celeste* (Paris 1892, 1893; Dover reprint, 1957); (b) M. Lindstedt, Mem. Acad. R. Sci. St. Petersburg (Ser. 7) **31**, 438 (1883); (c) M. N. Minorsky, *Théorie des Oscillations* (Gauthier-Villars, Paris, 1967).
[25] (a) H. G. Helleman and E. W. Montroll, Physica **74**, 22 (1974); (b) P. J. Melvin, SIAM J. Appl. Math. **33**, 161 (1977).
[26] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U. P., Cambridge, 1969).
[27] H. T. Davis, *Introduction to non-Linear Differential and Integral Equations* (Dover, New York, 1962).
[28] F. T. Hioe, Don Mac Millen, and E. W. Montroll, Phys. Rep. (Phys. Lett. C) **43**, 305 (1978).
[29] P. Claverie, "Perturbation of the harmonic oscillator in stochastic mechanics" (to be published).

[30]A. Denis, L. Pesquera, and P. Claverie, "Linear response of stochastic multiperiodic systems in stationary states with application to stochastic electrodynamics" (submitted to Physica A).
[31]J. H. Van Vleck, Phys. Rev. **24**, 347 (1924).
[32]T. H. Boyer, (a) Phys. Rev. D **13**, 2832 (1976); (b) Phys. Rev. A **18**, 1228 (1978).
[33]G. Nienhuis, Physica **66**, 245 (1973).
[34]T. W. Marshall (a) in *Dynamical Systems and Microphysics*, edited by A. Blaquière, F. Fer, and A. Marzollo, CISM courses and lectures No. 261 (Springer, Berlin, 1980), p. 135; (b) Physica A **104**, 223 (1980), Sec. 4.

Continued fraction theory of the rotating harmonic oscillator

Virendra Singh
Tata Institute of Fundamental Research, Bombay 400005, India

S. N. Biswas and K. Datta
Department of Physics & Astrophysics, University of Delhi, Delhi 110007, India

(Received 25 June 1980; accepted for publication 24 April 1981)

We study the rotating–vibrating system, consisting of the rotating harmonic oscillator, using the analytic theory of continued fractions. We prove that there is a convergent continued fraction representation of the Green's function which is analytic in the complex coupling constant plane, except for a cut along the negative real axis. The perturbation series for the Green's function is unambiguously defined by the continued fraction but diverges on account of an essential singularity at the origin. An infinite but incomplete set of exact solutions for certain specific values of the coupling follows from the representation of the Green's function as a continued fraction. Finally, we use Worpitzky's theorem in continued fraction theory to show that in the strong coupling limit $\alpha \to 0^+$ (α being the inverse of the coupling parameter), there exists a lower bound to all energy eigenvalues for a given value of l, the orbital angular momentum.

PACS numbers: 03.65.Db, 02.30.Mv

1. INTRODUCTION

The rotating harmonic oscillator presents the simplest model of a rotating–vibrating molecule. Langer[1] noted that an exact solution for the bound state eigenvalue problem was not possible; he showed, however, that eigensolutions for which the reduced radial functions vanish both at the origin and at large r exist. He also obtained the result that the corresponding eigenvalues differ from integers by terms at most of the order $\alpha|\ln\alpha|$, α being the inverse of the coupling parameter assumed to satisfy $0 < \alpha \ll 1$. Subsequently, Fröman and Fröman[2] showed that the spectrum does not contain terms such as $\alpha|\ln\alpha|$, whose presence would invalidate strong coupling expansions, but only contains terms with integral powers of α. Using the phase integral method, they showed the eigenvalues to be of the form $\lambda = v + l(l+1)\alpha + \cdots$, where v is a positive integer and l the orbital angular momentum quantum number. Recently, Flessas[3] obtained one class of exact solutions of the problem: he suggested that there exists a solution of the radial equation which is analytic in the radial variable, provided that $\lambda = l + 1 (l \neq 0)$. Fröman *et al.*[4] have subsequently disproved this suggestion and obtained a sequence of exact eigenfunctions for the s-wave problem.

In this note, we present a study of the system using the analytic theory of continued fractions. The theory of continued fractions has been used earlier in the solution of several types of quantum mechanical problems.[5–8] In particular, the methods we use have been applied to the study of certain kinds of anharmonic oscillator systems.[9,10] We expand the wave function in a power series multiplied by a suitably chosen Gaussian and obtain for the coefficients of the power series a three term contiguous difference equation. The solution of this difference equation is written in the form of an infinite continued fraction. The continued fraction representation allows us to obtain the Green's function, whose poles are the energy eigenvalues. A suitable transformation allows us to write the Green's function as a Stieltjes S fraction. We prove that this S fraction converges and that in the complex $\hat{\alpha} \equiv 1/\alpha$ plane it is analytic except for a cut along the negative real axis. Further, the perturbation series in α diverges, the divergence arising from a branch point singularity at $\alpha = 0$. An infinite but incomplete sequence of exact solutions for specific values of the coupling is obtained by terminating the continued fraction for the Green's function. These exact solutions do not contain the eigenvalues $\lambda = l + 1$ suggested by Flessas[3]; in this respect our results therefore confirm those of Fröman *et al.*[4] Finally, we use Worpitzky's theorem[11] in continued fraction theory to show that in the strong coupling limit $\alpha \to 0^+$ there exists, for each given l, a lower bound to the energy eigenvalues.

2. THE DIFFERENCE EQUATION AND THE GREEN'S FUNCTION

In the notation of Ref. 1, we consider the radial Schrödinger equation for the reduced wave function,

$$\frac{d^2F}{dr^2} + \left(\frac{\lambda + \frac{1}{2}}{\alpha} - \frac{(r-1)^2}{4\alpha^2} - \frac{l(l+1)}{r^2}\right)F = 0. \tag{1}$$

r is the (dimensionless) radial variable, λ the eigenvalue, l the orbital quantum number, and α a coupling parameter: generally, one is interested in $0 < \alpha \ll 1$. We look for solutions which vanish at $r = 0$ and $r = \infty$; thus

$$F = r^{\sigma} \exp[-(\mu/2)r^2 - \nu r]\chi. \tag{2}$$

With

$$\begin{aligned} \sigma &= l + 1, \\ \mu &= 1/2\alpha, \\ \nu &= -1/2\alpha, \end{aligned} \tag{3}$$

and

$$\chi(r) = \sum_0^{\infty} a_n r^n, \tag{4}$$

we obtain for the a_n's the difference equation

 0022-2488/82/071323-04$02.50

$$\alpha(n+1)(n+2l+2)a_{n+1}+(n+l+1)a_n + (\lambda-n-l)a_{n-1}=0. \tag{5}$$

Defining

$$\begin{aligned} A_n &\equiv \alpha(n+1)(n+2l+2),\\ B_n &\equiv (n+l+1),\\ C_n &\equiv (\lambda-n-l), \end{aligned} \tag{6}$$

we write the solution to (5) in the form

$$\frac{a_n}{a_{n-1}} = \cfrac{-C_n}{B_n+\cfrac{-A_nC_{n+1}}{B_{n+1}+\cfrac{-A_{n+1}C_{n+2}}{\ddots}}}. \tag{7}$$

Noting from (5) that

$$\frac{a_1}{a_0} = -\frac{(l+1)}{\alpha(2l+2)} = -\frac{1}{2\alpha} = -\frac{B_0}{A_0}, \tag{8}$$

we have

$$-\frac{B_0}{A_0} = \cfrac{-C_1}{B_1+\cfrac{-A_1C_2}{B_2+\ddots}}.$$

Thus

$$0 = B_0 + \cfrac{-A_0C_1}{B_1+\cfrac{-A_1C_2}{B_2+\ddots}}. \tag{9}$$

We now apply an equivalence transformation and reduce this to an S fraction. We introduce a sequence of objects $\{\alpha_n\}$ such that

$$\alpha_n\alpha_{n+1}A_nC_{n+1} = -1 \tag{10}$$

so that (9) may be written as

$$0 = B_0\alpha_0 + \cfrac{1}{B_1\alpha_1+\cfrac{1}{B_2\alpha_2+\ddots}}.$$

We define the right hand side of this equation as the inverse of the Green's function suitably normalized.

$$\alpha_0^{-1}A_0^{-1}G(\alpha,\lambda) = \cfrac{1}{B_0\alpha_0+\cfrac{1}{B_1\alpha_1+\cfrac{1}{\ddots}}}. \tag{11}$$

This is a Stieltjes fraction if

$$B_n\alpha_n > 0. \tag{12}$$

Iteratively one finds

$$\begin{aligned} \alpha_{2n+1} = &-\frac{1}{\alpha_0}\frac{1}{\alpha}2^{2n-2}\frac{\Gamma(n+1)}{\Gamma(2n+2)}\frac{\Gamma(2l+2n+1)}{\Gamma(2l)}\\ &\times\left(\frac{\Gamma(l)}{\Gamma(l+n)}\right)^2\frac{\Gamma(\lambda-l-2n-1)}{\Gamma(\lambda-l)}\\ &\times\left(\Gamma\left(\frac{\lambda-l}{2}\right)\Big/\Gamma\left(\frac{\lambda-l}{2}-n\right)\right)^2, \end{aligned} \tag{13}$$

$$\alpha_1 = -\frac{1}{\alpha_0}\frac{1}{\alpha}\frac{1}{(2l+2)(\lambda-l-1)}, \tag{14}$$

$$\begin{aligned} \alpha_{2n} = &\alpha_0 2^{-2n-2}\frac{\Gamma(2n)}{(\Gamma(n+1))^2}\left(\frac{\Gamma(l+n-1)}{\Gamma(l+1)}\right)^2\\ &\times\frac{\Gamma(\lambda-l-2n+1)}{\Gamma(\lambda-l-1)}\\ &\times\left(\Gamma\left(\frac{\lambda-l}{2}-1\right)\Big/\Gamma\left(\frac{\lambda-l}{2}-n+1\right)\right). \end{aligned} \tag{15}$$

Introducing a new set of quantities k_n such that

$$\begin{aligned} B_{2n}\alpha_{2n} &= k_{2n},\\ B_{2n+1}\alpha_{2n+1} &= (1/\alpha)k_{2n+1}, \end{aligned} \tag{16}$$

we obtain for the Green's function $G(\alpha,\lambda)$ the continued fraction representation

$$G(\alpha,\lambda) = \cfrac{1}{k_0+\cfrac{1}{k_1\hat{\alpha}+\cfrac{1}{k_2+\cfrac{1}{k_3\hat{\alpha}+\ddots}}}}, \tag{17}$$

where $\hat{\alpha}\equiv 1/\alpha$.

3. ANALYTICITY OF THE GREEN'S FUNCTION: DIVERGENCE OF THE PERTURBATION SERIES

The representation of the Green's function as a continued fraction in (17) enables, as in the case of the doubly anharmonic oscillator, the use of analytic continued fraction theory to study the analytic structure of the Green's function in the coupling constant plane. From (13), (14), (15), and (6) it may be seen that it is possible to choose α_0 such that all the k_n (except, at best, a finite number) are positive. Further, using Stirling's approximation, we find that Σk_n diverges. We thus have[12]

Theorem I: $G(\alpha,\lambda)$ considered as a function of $\hat{\alpha}=1/\alpha$ for fixed λ is uniformly convergent over a finite closed domain of $\hat{\alpha}$, whose distance from the negative half of the real axis is positive. Its value is an analytic function of α for all $\hat{\alpha}$ not on the negative half of the real axis.

Next, we rewrite (17) in the form

$$\frac{1}{\hat{\alpha}}G(\alpha,\lambda) = \cfrac{1}{k_0\hat{\alpha}+\cfrac{1}{k_1+\cfrac{1}{k_2\hat{\alpha}+\ddots}}}. \tag{18}$$

Introducing new quantitites d_n through

$$d_n = 1/k_{n-1}k_n, \tag{19}$$

we write (17) in the form

$$k_0G(\alpha,\lambda) = \cfrac{1}{1-\cfrac{d_1(-\alpha)}{1-\cfrac{d_2(-\alpha)}{1-\ddots}}}. \tag{20}$$

This is an alternative form of the Stieltjes S fraction. From this representation it follows that this S fraction can be expanded in a power series in $(-\alpha)$ and the series is unambiguously defined by the S fraction.[13] We write the power series as

$$k_0G(\alpha,\lambda) = C_0 + C_1(-\alpha) + C_2(-\alpha)^2 + \cdots. \tag{21}$$

The coefficients C_n are all positive. This series may be regarded as the perturbation expansion of $G(\alpha,\lambda)$ in powers of α. To determine the convergence of this series, we note that $d_n = 1/k_{n-1}k_n \to \infty$ as $n \to \infty$. Then from the theorem of Stieltjes[14] we have

Theorem II: The perturbation series for $G(\alpha,\lambda)$ in powers of α is divergent, arising from an essential singularity at $\alpha = 0$.

4. EXACT SOLUTIONS: CONSTRAINTS ON THE COUPLING

An infinite sequence of exact energy eigenvalues for particular values of the coupling may be obtained from the continued fraction representation of the Green's function. From (7) we find that if $C_n = 0$, $C_{n-1} \neq 0$, $n = N$ (say), then $a_N = 0$. However, the difference equation (5) shows that $C_n = a_n = 0$ for $n = N$ requires that $a_{N+1} = 0$. Thus, all the $a_n, n \geqslant N$ vanish and the wavefunction reduces to a polynomial (weighted with the usual Gaussian). These are the exact polynomial solutions of the problem: the corresponding energy eigenvalues are given by

$$\lambda = n + l. \tag{22}$$

It must be emphasized that each of these exact energy eigenvalues corresponds to a specific value of the coupling α. This is most easily seen from (9). If $C_N = 0$, i.e., $\lambda = N + l$, we must ensure that

$$0 = B_0 + \cfrac{-A_0C_1}{B_1 + \cfrac{-A_1C_2}{\ddots \quad B_{N-2} + \cfrac{-A_{N-2}C_{N-1}}{B_{N-1}}}}. \tag{23}$$

Since λ has been constrained to a specific value (viz. $N + l$) this is an equation in the coupling α. Thus, e.g., $\lambda = l + 2$ is an exact eigenvalue only for $\alpha = 1/2(l + 2)$; $\lambda = l + 3$ requires $\alpha = (l + 2)(l + 3)/2(4l + 9)$. The solution $\lambda = l + 1$ must be excluded since from (6) we see that $B_0 = 0$ cannot be satisfied. We note that this set of exact solutions is incomplete: at each constrained value of α, there exists an infinity of energy eigenvalues of which only one (for each value of l) appears in (22). We note that the exact solutions obtained here specifically exclude the set $\lambda = l + 1$ suggested by Flessas[3]; further, the exact eigenvalues occur only for specific values of α and not for all $\alpha > 0$, as Flessas suggests. These results, therefore, confirm the remarks of Fröman *et al.*[4] regarding the validity of Flessas' results.

5. LOWER BOUND ON THE EIGENVALUES IN THE STRONG COUPLING LIMIT

We next show that in the strong coupling limit $\alpha \to 0^+$ the eigenvalues $\lambda(\alpha,l)$ are bounded from below. While a lower bound is to be expected on intuitive grounds and is reflected in the approximate results of Fröman and Fröman,[2] we use the theory of continued fractions to obtain a rigorous lower bound.

Consider the difference equation (5) in the form

$$\frac{\alpha(n+1)(2l+n+2)}{(n+l+1)}a_{n+1} + a_n + \frac{(\lambda-n-l)}{(n+l+1)}a_{n-1} = 0, \tag{24}$$

i.e.,

$$F_n a_{n+1} + a_n - H_n a_{n-1} = 0,$$

with

$$F_n \equiv \alpha(n+1)(n+2l+2)/(n+l+1),$$
$$H_n \equiv (n+l-\lambda)/(n+l+1). \tag{25}$$

Then

$$G(\alpha,\lambda) = \cfrac{1}{1 + \cfrac{F_0H_1}{1 + \cfrac{F_1H_2}{1 + \ddots}}}. \tag{26}$$

Define

$$b_{n+2} \equiv F_nH_{n+1}, \quad n = 0,1,2,\cdots. \tag{27}$$

Then

$$G(\alpha,\lambda) = \cfrac{1}{1 + \cfrac{b_2}{1 + \cfrac{b_3}{1 + \ddots}}}. \tag{28}$$

A continued fraction of the form (28) with partial quotients of the form $b_n/1$ satisfies Worpitzky's theorem, which is as follows.

Let $b_2, b_3, \cdots$ be functions of any variable over a domain D in which

$$|b_{p+1}| \leqslant \tfrac{1}{4}, \quad p = 1,2,\cdots. \tag{29}$$

Then the following statements hold.

(a) The continued fraction (28) converges uniformly over D.

(b) The values of the continued fraction and its approximants are in the circular domain

$$|z - \tfrac{4}{3}| \leqslant \tfrac{2}{3}. \tag{30}$$

(c) The constant $\frac{1}{4}$ is the "best" upper bound on $|b_{n+1}|$ which may be used in (29), and (30) is the best domain of values of the approximants.

Now, from (27) and (25),

$$\begin{aligned} b_{p+1} &= F_{p-1}H_p \\ &= \frac{\alpha p(p+2l+1)(p+l-\lambda)}{(p+l)(p+l+1)} \\ &< \frac{\alpha(p+l)(2p+2l+2)(p+l-\lambda)}{(p+l)(p+l+1)} \\ &= 2\alpha(p+l-\lambda). \end{aligned} \tag{31}$$

Defining $a \equiv (l - \lambda)$, we have

$$b_{p+1} < 2\alpha(p+a).$$

Assume that $a > 0$, i.e., all the eigenvalues λ are bounded above by l. Then, since $\alpha > 0$, (29) requires

$$2\alpha(p+a) < \tfrac{1}{4},$$

i.e.,

$$\alpha < 1/8(p+a).$$

Since $p = 1,2,\cdots$ and $a > 0$, a sufficient condition for Wor-

pitzky's theorem to hold is that

$$\alpha < \tfrac{1}{8}, \quad a > 0. \tag{32}$$

Worpitzky's theorem now ensures that $G(\alpha,\lambda)$ converges for all α,λ such that (32) holds. Since the eigenvalues are poles of $G(\alpha,\lambda)$ and since for $\alpha \to 0^+$ $G(\alpha,\lambda)$ is convergent for all $\lambda < l$, we therefore conclude that all the eigenvalues $\lambda(\alpha,l)$ are bounded *below* by l as $\alpha \to 0^+$. Indeed, (30) enables us to give an upper bound for $G(\alpha,\lambda)$ in the domain $\alpha < \frac{1}{8}, \lambda < l$:

$$G(\alpha,\lambda) < 2. \tag{33}$$

Thus, the orbital angular momentum l itself constitutes a lower bound for all eigenvalues for a given l in the strong coupling limit.

[1] R. E. Langer, Phys. Rev. **75**, 792 (1949).
[2] N. Fröman and P. O. Fröman, Ann. Phys. (N. Y.) **115**, 269 (1978).
[3] G. P. Flessas, Phys. Lett. A **71**, 315 (1979).
[4] N. Fröman, P. O. Fröman, and F. Karlsson, Phys. Lett. A **77**, 397 (1980).
[5] S. N. Biswas and T. Vidhani, J. Phys. A **6**, 468 (1973). See also J. P. Lavine and H. S. Picker, *ibid.* **6**, L139 (1973).
[6] S. C. Miller, Phys. Rev. D **12**, 3838 (1975).
[7] V. Singh, S. N. Biswas, and K. Datta, Lett. Math. Phys. **3**, 73 (1979).
[8] J. P. Vigneron and Ph. Lambin, J. Phys. A **13**, 1135 (1980).
[9] F. T. Hioe, D. MacMillen, and E. W. Montroll, J. Math. Phys. **17**, 1320 (1976). F. T. Hioe, J. Math. Phys. **19**, 1307 (1978).
[10] V. Singh, S. N. Biswas, and K. Datta, Phys. Rev. D **18**, 1901 (1978).
[11] H. S. Wall, *Analytic Theory of Continued Fractions* (Van Nostrand, New York, 1948), p. 42, Theorem 10.1.
[12] Reference 11, p. 120, Theorem 28.1.
[13] Reference 11, p. 202, Theorem 52.1.
[14] T. J. Stieltjes, "Recherches sur les fractions continues," Oeuvres (unpublished), Vol. 2, p. 402; L. R. Acad. Sci. C **XVIII**, 1401 (1894).

An $\mathscr{E}_6 \otimes \mathscr{U}(1)$ invariant quantum mechanics for a Jordan pair[a)]

P. Truini[b)] and L. C. Biedenharn[c)]
Stanford Linear Accelerator Center, Stanford University, Stanford, California 94305

(Received 15 September 1981; accepted for publication 29 January 1982)

Quantum mechanical spaces associated with geometries based on exceptional groups are of interest as models for internal (quark) symmetries. Using the concept of a Jordan pair, two copies of complex 3×3 octonionic Jordan algebras ($\mathscr{M}_3^8$) are shown to define a quantum mechanics over the complex octonionic plane having $\mathscr{E}_6 \otimes \mathscr{U}(1)$ as automorphism group. The unusual features of this new quantal structure (neither a projective geometry, nor a lattice) are discussed.

PACS numbers: 03.65.Fd, 11.30.Ly

1. INTRODUCTION AND SUMMARY

The unusual nature of the hypothetical quark degrees of freedom has been a strong incentive for the construction of new quantum mechanical models in which these strange properties are to appear naturally, and not as *ad hoc* artifacts. It has been stressed, both by Jordan[1] and more recently by Dirac,[2] that the most fruitful way to attempt any generalization of the standard Hilbert space structure of quantum mechanics lies in changing the basic algebraic structures; in particular, Dirac expressed the view that nonassociative algebraic systems still hold the most promise. Jordan formulated quantum mechanics[3] in terms of commutative, but not associative (finite-dimensional) algebras of observables (the so-called Jordan algebras); Jordan, von Neumann, and Wigner[4] showed that this approach is equivalent to (finite-dimensional) standard quantum mechanics with the single exception of $\mathscr{M}_3^8$, the algebra of 3×3 Hermitian matrices over octonions. It is a remarkable fact, emphasized by Faulkner and Ferrar,[5] that *all notions of exceptionality in algebra and in geometry are manifestations of one underlying structure*, that is, nonclassical Lie algebras; nonassociative alternative algebras; nonspecial Jordan algebras; and non-deSarguesian projective planes; are all related, in one way or another, to the *octonions* (Cayley numbers).

Current interest in the use of octonionic structures was initiated by Gürsey,[6] who noted that specializing one of the seven nonscalar Cayley units (to play the role of the imaginary unit) automatically achieves a rationale for $\mathrm{SU}(3)^{\mathrm{color}}$. In particular, the five exceptional Lie groups exhibit a color-flavor structure:

$$\mathscr{G}_2\colon \mathrm{SU}(3)^c,\quad \mathscr{F}_4\colon \mathrm{SU}(3)\times\mathrm{SU}(3)^c,\quad \mathscr{E}_7\colon \mathrm{SU}(6)\times\mathrm{SU}(3)^c,$$
$$\mathscr{E}_6\colon \mathrm{SU}(3)\times\mathrm{SU}(3)\times\mathrm{SU}(3)^c,\quad \mathscr{E}_8\colon \mathscr{E}_6\times\mathrm{SU}(3)^c,$$

Gürsey emphasizes that nonassociativity may be connected with the problem of confinement. The fact that the distance function for non-deSarguesian geometries contains a part directly due to nonassociativity is itself suggestive. One merit of this approach is that it leads to a small number of possible models, which can be tested against known results. In particular, $\mathscr{G}_2$ and $\mathscr{F}_4$ are eliminated (flavor group too small); $\mathscr{E}_6$ seems viable, but $\mathscr{E}_7$ seems to be ruled out.[7]

The exceptional quantum mechanics, $\mathscr{M}_3^8$, was reinvestigated by Günaydin, Piron, and Ruegg[8] and shown to accord with the standard propositional formulation, with a unique probability function for the Moufang (non-deSarguesian) plane. The automorphism group of this structure is $\mathscr{F}_4$, and, as mentioned, the color-flavor structure is too small to be acceptable.

The purpose of the present paper is to construct a quantum mechanics for the complexified octonionic plane, which, as we show, has the automorphism group $\mathscr{E}_6\times\mathrm{U}(1)$, a group large enough to accommodate—as finite-dimensional quantum-mechanical charge spaces—a color-flavor structure which is not ruled out by current experimental evidence. The construction makes essential use of recent technical advances in Jordan algebras, which we now sketch.

The Jordan algebraic approach attempted to capture the essence of Hermitian matrix algebra (which characterizes quantum mechanics) by eliminating all reference to the underlying wavefunction concept, by focusing attention only on the algebraic properties of observables, and by eliminating the explicit use of the imaginary unit i. (This latter via the "formally real axiom": $a^2+b^2=0\Rightarrow a=b=0$.) For Hermitian matrices the operations of multiplication by real scalars, $x\to\alpha x$; addition, $(x+y)$; and formation of powers, x^n, were all taken over, but the only allowed product is the symmetric one: $xy+yx$ (since the lack of i forbids commutators).

The axioms for a Jordan algebra were taken to be (1) $x\circ y\equiv y\circ x$ (commutativity) and (2) $(x^2\circ y)\circ x=x^2\circ(y\circ x)$ (Jordan axiom; nonassociativity). (The role of this second axiom is exactly the same as the Jacobi axiom in Lie algebras; it ensures that one has an integration process—the Jordan analog to the Baker–Campbell–Hausdorff identity.)

It is remarkable that this technique—which, by contrast to the Dirac q-number approach, exchanges commutativity for noncommutativity and nonassociativity for associativity—is essentially identical to standard quantum mechanics. The one exception, $\mathscr{M}_3^8$, is the first known example of a quantum mechanics for which there is no Hilbert space, and no wavefunction.

Although the Jordan program began in physics, most of the interest, and developments, in Jordan algebras have been in mathematics; progress here has led to considerable change in the basic viewpoints. We summarize two developments of

[a)] Supported in part by the Department of Energy under Contract No. DE-AC03-76SF00515, and partly by the NSF, Contract No. PHY8105333.
[b)] Fellow of the Fondazioni A. Della Riccia-Firenze (Italia).
[c)] On leave from Duke University, Durham, N. C. 27706.

 0022-2488/82/071327-19$02.50

this type:

(a) the concept of a *quadratic Jordan algebra*, and the related concept of *inner ideals*; and

(b) the concept of *structural group* and *Jordan pairs*.

Consider the concept of quadratic Jordan algebra. The idea here is to model everything on the product $U_x(y) = xyx$—which is quadratic in x—rather than on the linear Jordan product $x \circ y = \frac{1}{2}(xy + yx)$. The axioms for quadratic Jordan algebras were given by McCrimmon.[9] (These axioms are given in Sec. 2). These axioms appear complicated, and it is not clear that they really constitute a step forward! We can indicate that it is by noting these points:

(1) Nothing is lost—quadratic Jordan algebras are categorically equivalent to linear Jordan algebras whenever the latter is defined (i.e., characteristic not 2).

(2) The quadratic algebra allows composition with the "generalized determinant," the norm form $N(y)$. Thus, $N(U_x(y)) = (N(x))^2 N(y)$. There is nothing analogous to this in the linear case.

(3) There is a structure theory for the quadratic algebras which is closely analogous to that for associative algebras.[10]

Let us explain the significance of this last point. For a physicist the Jordan aproach is unhandy largely because it banishes the concept of wavefunction (more precisely, bra and ket vectors) with only the density matrix remaining. In mathematical language what has happened is this: the concept of a ket vector is the concept of a (left) *ideal*, a subset N of the associative algebra A such that $n \cdot A \subset N$ if $n \subset N$. In a nonassociative algebra there is *no* such concept. What replaces it comes from the *quadratic* algebra: the concept of an *inner ideal*. An inner ideal M is a subset of a quadratic algebra J such that $U_x(J) \subset M$ if $x \subset M$. The importance of this concept can be seen in this way: it is a fundamental result that the projective geometry of the space of n-tuples over a field Φ is isomorphic to the geometry of left ideals in the (associative) algebra of $n \times n$ matrices over Φ; for nonassociative algebras, inner ideals play an equivalent role in the construction of geometries.[5] It is our belief (noting the close relationship between geometries and quantum mechanics) that the concepts of quadratic Jordan algebras and inner ideals will be useful in physics.

Let us turn to the second conceptual development: the concept of a structural group (Koecher, Ref. 11) and Jordan pairs (Loos, Ref. 12). The automorphisms of a given physical structure are a well-known approach to the intrinsic properties of the structure. For an algebra, one studies the automorphisms which preserve the algebraic laws; accordingly, such transformations always map the unit element into itself.

How could one change the unit element? If u has an inverse, let us replace the product xy in an associative algebra by $xy \leftrightarrow xu^{-1}y$. The new unit element and its inverse are easily computed: $1^{(u)} = u$, $x^{-1(u)} = ux^{-1}u$.

For associative algebras this new algebra $A^{(u)}$ is, in fact, isomorphic to A but, remarkably, for nonassociative algebras this shift of the unit can produce a different algebra. Such a new algebra is called an isotope, $J^{(u)}$, of the original algebra J.

The desire to study not only the Jordan algebra J but all its isotopes as a single entity leads to the two concepts of structural group and of Jordan pair. The structural group, $\mathrm{Str}(J)$, is the group of isomorphic mappings of a Jordan algebra J and its isotopes onto itself: $J^{(u)} \underset{T}{\to} J^{(T_u)}$. The automorphism group $\mathrm{Aut}(J)$ is the subset of such mappings fixing the unit element.

The construction of quantum mechanics over a complex octonionic plane was begun by Gürsey,[13,14] but without using the concepts of inner ideals or Jordan pairs. (Appendix D discusses this work.) Let us indicate, briefly, how these concepts afford a more natural approach.

The work of Jordan, von Neumann, and Wigner really was categoric; within their axioms $\mathscr{M}_3^8$ is the *only* new quantum mechanics. Thus to go further one must drop one (or more) of their axioms: in the present case we drop the axiom of formal reality. The price one pays for this (in a direct approach, as in Ref. 13) is that the elements of the algebra become complex octonionic 3×3 matrices, which are Hermitian under octonionic conjugation, but *not* under complex conjugation. This destroys at once the *raison d'etre* for the Jordan approach, that is, the study of algebras of observables.

The use of Jordan pairs nicely remedies this difficulty: the pair consists of *two* complex $\mathscr{M}_3^8$ structures, and the concept of observable becomes the concept of Hermitian *pairs*.

Similarly the use of Jordan pairs allows one to take over the language of inner ideals and, equally importantly, the concept of a Peirce decomposition. It is through this latter concept (discussed in Sec. 4) that we are able to achieve, in a natural way, an orthocomplementation for the complex octonionic plane.

The use of Jordan pairs has several other intuitive advantages. Let us mention that the pair concept makes intuitively clear the structure of the projectors, which we construct for both points and lines. More importantly, the pair concept makes it clear why both *nilpotents* as well as idempotents of complex $\mathscr{M}_3^8$ are to be associated to points (both lead to idempotents of the pair). The association of nilpotents to quantal propositions is a new feature of the present structure.

As is to be expected, there are some unusual features of the quantum mechanics constructed for the complex octonionic plane. All these new features are, in one way or another, related to the existence of point spaces of dimension > 1, (in Sec. 5). There is a new relation in the geometry (see Sec. 4), called "connectedness" (by Springer, Ref. 15). The resulting geometry is no longer projective: two lines may intersect in more than one point. (If so, the two or more points of the intersection are then "connected" points.)

As a result of this new feature, the quantal structure must differ from the standard (propositional calculus) formulation: the only axiom which is dropped is the (lattice) axiom asserting the existence of a greatest lower bound for any two propositions. This axiom, as is well-known, is the axiom least justified experimentally, since it is nonconstructive. It is the merit of the present construction that it provides a model in which this axiom is denied in a natural way.

The lack of a lattice structure affects also the definition of "state," forcing us to define a "measure" with unusual

properties. Let us note, however, that this measure coincides with the unique probability function (defined by Günaydin, Piron, and Ruegg, Ref. 8, on the Moufang plane) when restricted to the real octonionic case. Moreover, when restricted to the purely complex case, the measure coincides with the usual modulus (squared) of complex three-dimensional Hilbert space quantum mechanics. [This is to be distinguished from the specialization of the Moufang plane itself (taking, say, e_7 as the complex unit) since, among other things, the pair structure is retained.]

The automorphism group of this quantum mechanics is $\mathscr{E}_{6,0} \otimes \mathscr{U}(1)$, which plays a role analogous to that of $\mathscr{F}_{4,0}$ for the Moufang plane (see Sec. 2 for notation on groups). The automorphism group $\mathscr{E}_{6,0} \otimes \mathrm{U}(1)$ preserves the trace $\mathrm{tr}(x,x^*)$, is transitive on points, and transitive on triples of orthogonal points. The group action can, however, only partially diagonalize the observables, leading to an unusual spectral theory. We show (in Appendix B) that the isotropy group of a point is $\mathrm{SO}(10) \otimes \tilde{\mathrm{U}}(1)$, so that this complex octonionic plane is the homogeneous space $\mathscr{E}_{6,0}/\mathrm{SO}(10) \otimes \tilde{\mathrm{U}}(1)$.

One of the advantages of the complexification is that we now recover the standard Wigner relation in which infinitesimal symmetry generators are directly related to observables. (This is lost in $\mathscr{M}_3^8$.) Similarly, time reversal can now be implemented.

All of the unusual features of the present structure have closely analogous features in the theory of reducible lattices. In the present model, for example, the superposition principle is not unrestrictedly valid. This is true also for reducible lattices, where there are superselection rules, but unlike reducible lattices (where there may be no superposition for two states) in our case, the failure (for two connected ponts) occurs because there are too many states (superposition is not uniquely defined). If one views reducible lattices as a structure between classical mechanics and standard quantum mechanics, in some sense the present structure lies on the opposite side of quantum mechanics.

It would be an interesting problem, we feel, to find some kind of physical understanding of the role of the connected points which are responsible for all the unusual features of this quantum mechanics.

2. THE MATHEMATICAL FRAMEWORK

We introduce here briefly the concepts which form the background needed for the understanding of our construction. (We assume the reader to be familiar with the theory of Lie groups.) Let us refer to the following review articles on the more recent developments in Jordan algebras: Refs. 16–19.

The first important step in the theory came with the definition of quadratic Jordan algebras. A quadratic[9,17] Jordan algebra J is based on a quadratic map U yielding $U_x\, y$ quadratic in x and linear in y satisfying:

$$\text{(Q1)} \quad U_I = Id,$$

$$\text{(Q2)} \quad U_x V_{y,x} = V_{x,y} U_x,$$

$$\text{(Q3)} \quad U_{U_x y} = U_x U_y U_x,$$

where $V_{x,y}(z) = (U_{x+z} - U_x - U_z)(y)$ and I is the unit in J.

In the associative case ($U_x\, y = xyx$) the theories of quadratic and linear Jordan algebras are equivalent (if the characteristic of the field is different from 2), but the theory of quadratic Jordan algebras leads naturally to the definition of new objects, of which the best example is the definition of inner ideal.[19] An inner ideal $\mathscr{B}$ is a subspace of J closed under the map U:

$$U_{\mathscr{B}} J \subset \mathscr{B}.$$

This new concept is very useful in the definition of geometries which can be entirely based on inner ideals.[20] Once the geometrical objects are identified with inner ideals the incidence relation is automatically given by set containment. Actually the geometrical objects are better identified with the *principal* inner ideals, that is, the inner ideals $\mathscr{B}$ generated by a single element b in J:

$$U_b J = \mathscr{B}.$$

The principal inner ideal plays the same role, in quadratic Jordan algebra, as that of the (one-sided) ideal in the associative case. (We have noted above that one-sided ideals are just the bras (or kets) in the standard theory of quantum mechanics.) Another important property which arises naturally in the quadratic formulation of Jordan algebra is the composition property with respect to the generic determinant[19]

$$N(U_x(y)) = (N(x))^2 N(y).$$

This has no analog in the linear case.

A further important concept is the definition of isotope. If v is an invertible element of J then we can define $U_x^{(v)} = U_x U_v$ for every x in J. The algebra J with quadratic map $U^{(v)}$ is a quadratic Jordan algebra with unit v^{-1}, called the v-isotope $J^{(v)}$ of J. Two algebras are called isotopic if there exist an isomorphism between one of them and an isotope of the other. Because two isotopic Jordan algebras need not be isomorphic, isotopism is more general than isomorphism, and one is led to the concept of structural group, $\mathrm{Str}(J)$, the group of isomorphisms of J with its isotopes. The automorphism group of J is therefore the subgroup of $\mathrm{Str}\, J$ preserving the unit. Many properties of a Jordan algebra can be found most easily by working in a particular isotope, but at the same time there are properties which may hold only up to isotopy. The concept of Jordan pair was introduced to define a larger structure which included a Jordan algebra along with all of its isotopes.

Before defining the Jordan pair let us define the prior concept of Jordan triples. A Jordan triple is a quadratic Jordan algebra without the unit element. The axioms are

$$\text{(JT1)} \quad V_{x,y} U_x = U_x V_{y,x},$$

$$\text{(JT2)} \quad V_{U_x y, y} = V_{x, U_y x},$$

$$\text{(JT3)} \quad U_{U_x y} = U_x U_y U_x.$$

[Just as Jordan algebras may be considered as (a generalization of) a way to multiply symmetric matrices, so may the Jordan triple structure be viewed as (a generalization of) a way to multiply *rectangular* matrices with $U_x\, y = xy^t x$, where t indicates transposition.]

A Jordan pair $V = (V^+, V^-)$ is a pair of spaces which

act on each other as a Jordan triple. The axioms are

(JP1) $V_{x^+,y^-}U_{x^+} = U_{x^+}V_{y^-,x^+}$,

(JP2) $V_{U_{x^+}y^-,y^-} = V_{x^+,U_{y^-}x^+}$,

(JP3) $U_{U_{x^+}y^-} = U_{x^+}U_{y^-}U_{x^+}$,

the same holding with the signs interchanged. For example, we can form a Jordan pair by taking V^+ as the set of matrices $m\times n$, V^- as the set of matrices $n\times m$ and $U_x\,y = xyx$. We can also get a Jordan pair by doubling a Jordan algebra. That is, we can take $V^+ = V^- = J$ and the same quadratic map U defined on J.

A homomorphism between two Jordan pairs V, W is a pair of linear maps $h = (h_+,h_-)$, $h_\sigma: V^\sigma \to W^\sigma$, $\sigma = \pm$ such that $h_\sigma(U_{x^\sigma}y^{-\sigma}) = U_{h_\sigma x}\sigma(h_{-\sigma}y^{-\sigma})$.

If J is a Jordan algebra and $V = (J,J)$ is the Jordan pair obtained by doubling J, then there exists a one-to-one correspondence between the structural group of J, Str J, and the automorphism group of V, Aut(V). If $g\in$Str J, then associated to it is the element $(g,U_{g(I)}^{-1}g)$ in Aut(V). (Thus, as noted, a Jordan pair includes J and its isotopes.)

In a Jordan pair it is possible to define idempotents and inner ideals. The idempotents of $V = (J,J)$ include the nilpotents of J (a nilpotent of J becomes an idempotent in a suitable isotope) and it is possible to define the Peirce decomposition[12] with respect to an idempotent of V in a way similar to that of the Jordan case. It follows that nilpotents and idempotents of J are treated in a *uniform* way in the Jordan pair (J,J). Inner ideals can also be defined and related to the Peirce decomposition. (We shall define and use these concepts in the next sections.)

Jordan pairs are strictly related to three-graded Lie algebras. Any three-graded Lie algebra $L = L_1 \oplus L_0 \otimes L_{-1}$ ($[L_i, L_j]\subset L_{i+j}$) can be obtained from a Jordan pair[19] and, conversely, a Jordan pair can be obtained from L by setting

$$L_1 = V^+,\ L_{-1} = V^-,\ V_{x^+,y^-}z^+ = [[x^+,y^-],z^+].$$

The map U is then obtained by

$$U_{x^+}\,y^- = \tfrac{1}{2}V_{x^+,y^-}x^+.$$

We shall examine in the sequel two different examples of Jordan pairs. The first—a didactical example to give the physicist reader some familiarity (and confidence) in the Jordan pair approach—is obtained from the three-grading of the complex Lie algebra A_2. We shall show that it is pairing of a triple system, and associated with 2×1 and 1×2 rectangular matrices.

The second example—the principal object of the paper—is obtained from the three-grading of the (complex) Lie algebra E_7, and is the pairing of a *complex* $\mathcal{M}_3^8$. The three-grading of E_7 coincides with the Tits–Koecher construction of the superstructural algebra[21,22]:

$$K = J + \text{Str}\,l(J) + J^*, \quad (133 = 27 + 79 + 27)$$

for the case in which J is a complex $\mathcal{M}_3^8$, a 27-dimensional representation of the exceptional Lie group $\mathcal{E}_6$; J^* is the conjugate representation of $\mathcal{E}_6$; and Str $l(J)$ is the complex Lie algebra of the structural group of J, $E_6\otimes\mathbb{C}$. From the discussion above, $\mathcal{E}_6\otimes\mathbb{C}$ is therefore isomorphic to the automorphism group of the Jordan pair (J,J). We note that $\mathcal{E}_6$ defined as the group preserving the generic determinant[19] of J; note also, for completeness, that the automorphism group of (complex) J is the exceptional Lie group $\mathcal{F}_4$, generated by the (complex) Lie algebra F_4.

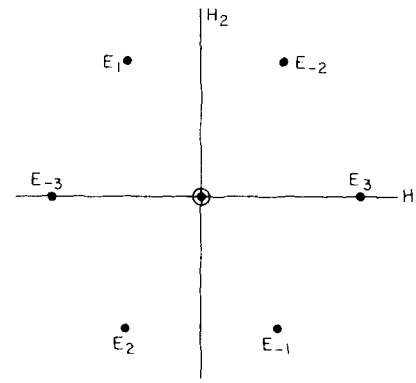
Remark on notation: We shall use in the sequel the Freudenthal notation[23] for the real forms of the Lie algebra H of type G_2, F_4, E_6, E_7. When no further index, other than the one specifying the rank of the group is written for H, it is meant that H is *complex*. The *compact real* form of H is denoted by $H_{r,0}$ (i.e., $E_{6,0}$, $F_{4,0}$,···). The group $H_{r,0}$ has a signature (the Cartan index) equal to minus the number of generators of H. For the *noncompact real* forms the signature must be specified (say, in Freudenthal's notation) to be meaningful.

In the sequel the *Lie group* associated with a certain Lie algebra will be denoted by the same, but script, letter.

3. AN INTRUCTIVE EXAMPLE FOR THE JORDAN PAIR STRUCTURE

As an example of a three-graded Lie algebra yielding a Jordan pair structure, we consider the (complex) Lie algebra A_2. We want to build a geometric structure on the generated pair and give to it a quantum mechanical meaning. We shall proceed along exactly the same lines we shall follow in the construction of a quantum mechanics for the Jordan pair (J,J).

Consider then the Lie algebra of A_2.



This is a three-graded Lie algebra with the identifications

L_1 = subalgebra generated by E_1,E_{-2},

L_{-1} = subalgebra generated by E_{-1},E_2,

L_0 = subalgebra generated by $E_{\pm 3}$,H_1,H_2.

We want to show that (L_1,L_{-1}) is indeed a Jordan pair. Let us calculate $U_x y = \frac{1}{2}[[x,y],x]$, where $x = \lambda_1 E_1 + \lambda_2 E_{-2}$; $y = \mu_1 E_{-1} + \mu_2 E_2$.

One finds

$$\begin{aligned} U_x y = {} & \tfrac{1}{2}(\lambda_1^2\mu_1[[E_1,E_{-1}],E_1] + \lambda_2^2\mu_2[[E_{-2},E_2],E_{-2}]) \\ & + \lambda_1\lambda_2\mu_1[[E_1,E_{-1}],E_{-2}] + \lambda_2\lambda_1\mu_2[[E_{-2},E_2],E_1] \\ & + \lambda_1\lambda_2\mu_2[[E_1,E_2]E_{-2}] + \lambda_1\lambda_2\mu_1[[E_{-2},E_{-1}],E_1]. \end{aligned}$$

Recall that (with conventional normalization) these double

commutators are, in the order in which they occur above, $\frac{1}{3}E_1;\frac{1}{3}E_{-2};\frac{1}{6}E_{-2};\frac{1}{6}E_1;\frac{1}{6}E_1$; and $\frac{1}{6}E_{-2}$. Thus (after some algebra),

$$U_x y = \tfrac{1}{6}(\lambda_1\mu_1 + \lambda_2\mu_2)(\lambda_1 E_1 + \lambda_2 E_{-2}) = x^\alpha y^\alpha x = (\mathrm{tr}\, xy)x,$$

where

$$x^\alpha = \lambda_\alpha/\sqrt{6},\ y^\alpha = \mu_\alpha/\sqrt{6}$$

(with $\alpha = 1,2$) are the "components" of x,y.
Linearizing this expression we get

$$\begin{aligned} V_{x,y}z &\equiv (U_{x+z} - U_x - U_z)y \\ &= (x^\alpha + z^\alpha)y^\alpha(x+z) - x^\alpha y^\alpha x - z^\alpha y^\alpha z \\ &= x^\alpha y^\alpha z + z^\alpha y^\alpha x. \end{aligned}$$

Remark: Note that these forms for U_x and $V_{x,y}$ are the same as one would get considering the pair obtained by doubling the *triple system* of rectangular matrices 1×2, 2×1. It is thus obvious that (V^+,V^-) is the pairing of a Jordan triple system.

It is easy to check that the Jordan pair axioms JP1–JP3 are satisfied.

To have a concrete example of the transformation U_x consider the three-dimensional representation of A_2. One finds

$$x = \begin{pmatrix} 0 & 0 & x^1 \\ 0 & 0 & x^2 \\ 0 & 0 & 0 \end{pmatrix}, \quad y = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ y^1 & y^2 & 0 \end{pmatrix}.$$

It is easily verified that $\frac{1}{2}[[x,y],x]$ $(\equiv U_x y)$
$= xyx = (x^1y^1 + x^2y^2)x = (\mathrm{tr}\, xy)x$.

We want now to give a quantum mechanical meaning to the Jordan pair $(V^+,V^-)\equiv V$. To do so, consider first the idempotents of V. They are given by the conditions $U_x y = x$, $U_y x = y$ with $x\in V^+$, $y\in V^-$. Since we have determined that $U_x y = \mathrm{tr}(xy)x$, and $U_y x = \mathrm{tr}(xy)y$, we find that the idempotents of (V^+,V^-) are just the elements (x,y) such that $\mathrm{tr}(xy) = 1$.

To interpret the idempotents as quantal objects we impose the normalization condition

$$\mathrm{tr}(xx^*) = \mathrm{tr}(yy^*) = 1,$$

where the asterisk denotes complex conjugation.

The interpretation of these idempotents is immediate if we think of x and y as two-dimensional complex vectors, one in the dual space of the other. Then $\mathrm{tr}(xy)$ is just the scalar product of x and y^*. The normalization condition implies $|\mathrm{tr}(xy)|\leqslant 1$ for any two vectors in $\mathbb{C}^2$, with the equality holding if and only if $x = e^{i\theta}y^*$. In particular if $\mathrm{tr}(xy) = 1$ it follows that $y = x^*$. The objects: (i) idempotent, normalized pair (x,y) and (ii) the projector $|\alpha\rangle\langle\alpha|$ of ordinary quantum mechanics, are thus all analogous.

In order to define the transition probability let us define the mapping Q: $V\to\mathrm{End}(V)$ as

$$Q_{(x_1,y_1)}(x_2,y_2) = (U_{x_1}U_{y_1}x_2, U_{y_1}U_{x_1}y_2),$$

and introduce the notation $P\equiv(x,y)$ if (x,y) is a normalized idempotent. Call P orthogonal to P' (denoted by $P\perp P'$) if

$$\mathrm{tr}(Q_P P') = 0.$$

We are now in a position to define the states of the system. The state α_p associated to the idempotent $P = (x,y)$ is defined as the probability function

$$\alpha_p(P') = \mathrm{tr}\, Q_p P',$$

where $P' = (x',y')$ is another normalized idempotent.

We now verify that α_p is a state, that is

(1) $0\leqslant\alpha_p(P')\leqslant 1$,

(2) $\alpha_p(P) = 1$,

(3) $\alpha_p(P_1\cup P_2) = \alpha_p(P_1) + \alpha_p(P_2)$ if $P_1\perp P_2$.

Proof: (1) $\mathrm{tr}(U_x U_y x', U_y U_x y')$
$= \mathrm{tr}(x'y)\mathrm{tr}(xy') = y^\alpha x'^\alpha x^\beta y'^\beta = x^{\alpha *}x'^\alpha x^\beta x'^{\beta *} = zz^*\geqslant 0$,
where $z\equiv x^{\alpha *}x'^\alpha\in\mathbb{C}$. The fact that $\alpha_p(P')\leqslant 1$ follows from the interpretation of $\mathrm{tr}(xy)$ as the scalar product of normalized vectors.

(2) $\mathrm{tr}(U_x U_y x, U_y U_x y) = \mathrm{tr}(xy) = 1$.

(3) We are obviously considering a "projective" complex line, hence the union of two orthogonal points must be defined as the line itself. So $\alpha_p(P_1\cup P_2) = 1$. On the rhs of condition (3) $\alpha_p(P_1) = |\mathrm{tr}\, y_1 x|^2$ and $\alpha_p(P_2) = |\mathrm{tr}\, y_2 x|^2$; hence we have the sum of the squared moduli of the components of x, which is (by assumption) equal to 1. □

Consider now the set of transformations in $\mathscr{L}_0$, the Lie group corresponding to the Lie algebra L_0, which maps points into points. Notice first that $\mathscr{L}_0$ is isomorphic to $GL(2,\mathbb{C})$ and that for g_+ in $\mathscr{L}_0$, g_0 is given by the condition

$$\mathrm{tr}(g_+(x),g_-(y)) = \mathrm{tr}(x,y).$$

Interpreting $\mathrm{tr}(x,y)$ as the scalar product in $\mathbb{C}^2$ of x with y^* it is easy to check that

$$g_- = (g'_+)^{-1},$$

where t indicates the transpose.

In order to have a mapping g from points into points we must have

$$g:(x,x^*)\to(g_+(x),(g_+(x))^*)\equiv(g_+(x),g_-(x^*)).$$

We thus get the condition

$$g_-(x^*) = (g_+(x))^*,$$

that is $g_+^\dagger = g_+^{-1}$ where $\dagger$ indicates the transposed conjugate. We have therefore proven that the subgroup of the automorphism group of the Jordan pair, mapping points into points is U(2). The action of U(2) on the points is obviously transitive.

The observables will then be the (Hermitian) generators of U(2), i.e., they can be represented by the linear combinations of the Pauli matrices and the identity matrix. The action of an observable H on the idempotent (and normalized) elements of V will be

$$H(x,y) = (Hx,Hy)$$

with the trace $\mathrm{tr}(x,Hy) = \mathrm{tr}(Hx,y)$ easily interpretable as the expectation of H in the "state" (x,y).

This example is, of course, much too simple to be of intrinsic interest, but it does provide a useful guide to the manipulations which follow for which there is no longer any possible underlying associative structure.

4. THE GEOMETRY OF THE COMPLEX OCTONIONIC PLANE

4.1. Some definitions and notations

We consider the Jordan Pair obtained by doubling the Jordan algebra J of the 3×3 Hermitian matrices over the complex octonions. The Hermiticity is considered only with respect to the octonionic conjugation and the Jordan product is the symmetrized product

$$x\cdot y = \tfrac{1}{2}(xy + yx) \quad \text{for} \quad x,y\in J \tag{1.1}$$

and xy the ordinary matrix product of x and y. The quadratic and trilinear operators defining the pair structure are

$$U_{x^\sigma}y^{-\sigma} = \mathrm{tr}(x^\sigma,y^{-\sigma})x^\sigma - x^{\sigma\#}\times y^{-\sigma}, \tag{1.2}$$

$$V_{x^\sigma,y^{-\sigma}}z^\sigma = (U_{x^\sigma+z^\sigma} - U_{x^\sigma} - U_{z^\sigma})y^{-\sigma}, \tag{1.3}$$

where $\sigma = \pm$ distinguishes the two copies of J. In some cases we use the notation $V = (V^+,V^-)$ for the pair (J,J).

We have, furthermore, the following definitions and notations:

$$\mathrm{tr}(x,y) = \mathrm{tr}(x\cdot y), \tag{1.4}$$

$$x^\# = x^2 - x\,\mathrm{tr}(x) - \tfrac{1}{2}I(\mathrm{tr}(x^2) - (\mathrm{tr}\,x)^2), \tag{1.5}$$

$$\begin{aligned} x\times y &= (x+y)^\# - x^\# - y^\# \\ &= 2x\cdot y - x\,\mathrm{tr}(y) - y\,\mathrm{tr}(x) \\ &\quad - I(\mathrm{tr}(x\cdot y) - \mathrm{tr}(x)\mathrm{tr}(y)), \end{aligned} \tag{1.6}$$

where I is the identity in J.

The following identities hold[24] in J:

$$x^3 - \mathrm{tr}(x)x^2 + Q(x)x - N(x)I = 0, \tag{1.7}$$

$\mathrm{tr}(x), Q(x), N(x)$ are complex numbers. If x is in the form

$$x = \begin{pmatrix} \alpha_1 & a & \bar{b} \\ \bar{a} & \alpha_2 & c \\ b & \bar{c} & \alpha_3 \end{pmatrix}, \quad \begin{aligned} &\alpha_i\in\mathbb{C}\ \ i = 1,2,3, \\ &a,b,c \text{ octonions}, \end{aligned} \tag{1.8}$$

where the bar is the octonionic conjugation,

$$\mathrm{tr}(x) = \alpha_1 + \alpha_2 + \alpha_3, \tag{1.9}$$

$$\begin{aligned} Q(x) &= \alpha_1\alpha_2 + \alpha_2\alpha_3 + \alpha_3\alpha_1 - n(a) - n(b) - n(c) \\ &= \tfrac{1}{2}[(\mathrm{tr}(x))^2 - \mathrm{tr}(x^2)], \end{aligned} \tag{1.10}$$

$$\begin{aligned} N(x) &= \alpha_1\alpha_2\alpha_3 - \alpha_1 n(c) - \alpha_2 n(b) - \alpha_3 n(a) + t(abc) \\ &= \tfrac{1}{6}[(\mathrm{tr}(x))^3 - 3\,\mathrm{tr}(x)\mathrm{tr}(x^2) + 2\,\mathrm{tr}(x^3)], \end{aligned} \tag{1.11}$$

where $n(a) = a\bar{a}$ is the norm of the octonion a and $t(a) = a + \bar{a}$ is its trace.

We shall make use of the following identities[25] of J:

$$x^{\#\#} = N(x)x, \tag{1.12}$$

$$I^\# = I, \tag{1.13}$$

$$I\times y = \mathrm{tr}(y)I - y, \tag{1.14}$$

$$(U_x y)^\# = U_{x^\#}y^\#, \tag{1.15}$$

$$N(x^\#) = N(x)^2, \tag{1.16}$$

$$\mathrm{tr}(x,y) = \mathrm{tr}(x)\mathrm{tr}(y) - \mathrm{tr}(x\times y), \tag{1.17}$$

$$\mathrm{tr}(x\times y,z) = \mathrm{tr}(x,y\times z), \tag{1.18}$$

$$x^\#\times(x\times y) = N(x)y + \mathrm{tr}(x^\#,y)x, \tag{1.19}$$

$$x^\#\times y^\# + (x\times y)^\# = \mathrm{tr}(y^\#,x)x + \mathrm{tr}(x^\#,y)y. \tag{1.20}$$

The quadratic and trilinear operators (1.2) and (1.3) satisfy the following properties, which define the Jordan Pair structure:

$$V_{x^\sigma,y^{-\sigma}}U_{x^\sigma} = U_{x^\sigma}V_{y^{-\sigma},x^\sigma}, \tag{1.21}$$

$$V_{U_{x^\sigma}y^{-\sigma},y^{-\sigma}} = V_{x,U_{y^{-\sigma}}x^\sigma}, \tag{1.22}$$

$$U_{U_{x^\sigma}y^{-\sigma}} = U_{x^\sigma}U_{y^{-\sigma}}U_{x^\sigma}. \tag{1.23}$$

We have, moreover, the following identities[18]:

$$U_x y = 2(x\cdot y)\cdot x - x^2\cdot y, \tag{1.24}$$

$$\tfrac{1}{2}V_{x,y}z = (x\cdot y)\cdot z + (z\cdot y)\cdot x - (x\cdot z)\cdot y. \tag{1.25}$$

Whenever it will not generate confusion we shall drop the symbol $\sigma = \pm$.

Throughout this section the asterisk will denote the complex conjugation of a matrix, as well as a number. We shall call *rank 1* element any element x of J such that $x^\# = 0$.

4.2. The idempotents of V

An element (x,y) of $V = (J,J)$ is an idempotent if

$$U_x y = x \quad \text{and} \quad U_y x = y. \tag{2.1}$$

We can classify the idempotents of V by considering the first element, x, of the pair; we may have one of the following cases:

$$\begin{aligned} &\text{(i)} \quad x^\# = 0, \\ &\text{(ii)} \quad x^\#\neq 0, \quad N(x) = 0, \\ &\text{(iii)} \quad N(x)\neq 0. \end{aligned} \tag{2.2}$$

In case (i) it follows from (1.16) that $N(x) = 0$, which shows that (2.2) covers all the possible cases. It is a well-known result[26] that in case (i) we can have

either (i') $x^2 = \lambda x$, $\lambda\neq 0$, or equivalently;

$x = \lambda u$ for u a primitive idempotent of J:

$$u^2 = u, \quad \mathrm{tr}\,u = 1. \tag{2.3}$$

or (ii') $x^2 = 0$.

For case (iii) it is equivalent to say that x is invertible.[25] We have the following proposition:

Proposition 2.1: Let (x,y) be an idempotent of V, then only the three following cases may occur:

$$\begin{aligned} &\text{(i)} \quad x^\# = 0, \quad y^\# = 0, \quad \mathrm{tr}(x,y) = 1; \\ &\text{(ii)} \quad x^\#\neq 0, \quad N(x) = 0, \\ &\qquad y^\#\neq 0, \quad N(y) = 0, \quad \mathrm{tr}(x,y) = 2; \\ &\text{(iii)} \quad N(x)\neq 0, \quad N(y)\neq 0, \quad \mathrm{tr}(x,y) = 3. \end{aligned} \tag{2.4}$$

To prove this proposition we need the following results:

Proposition 2.2: If (x,y) is an idempotent and $x^\# = 0$, then $y^\# = 0$ and $\mathrm{tr}(x,y) = 1$.

Proof: Let z be any element of J. Using the definition of idempotent we have

$$\begin{aligned} U_y z &= U_{U_y x}z = U_y U_x U_y z = U_y\,\mathrm{tr}(U_y z,x)x \\ &= \mathrm{tr}(U_y z,x)y \quad \text{for every } z \text{ in } J. \end{aligned}$$

That is, $U_y(J)\subseteq\mathbb{C}y$ (with $\mathbb{C}$ the complex field).
It is known that this implies $y^\# = 0$.[25] Furthermore $x = U_x y = \mathrm{tr}(x,y)x$ implies $\mathrm{tr}(x,y) = 1$ which concludes the proof. □

Proposition 2.3: If $x^\#\neq 0$ and $N(x) = 0$ then there exists

an element $z \in J$ such that $x^\# \times z = x$.

Proof: If $x^\# \neq 0$ then there exists y such that $\mathrm{tr}(x^\#, y) = 1$ by nondegeneracy of the trace.[27] Using (1.19) it follows that $x^\# \times (x \times y) = x$. The proposition is thus proven for $z = x \times y$. □

Proposition 2.4: Let (x,y) be an idempotent of V. If $x^\# \neq 0$ and $N(x) = 0$ then $y^\# \neq 0$, $N(y) = 0$ and $\mathrm{tr}(x,y) = 2$.

Proof: It follows from (1.15) that $x^\# = (U_x y)^\# = U_{x^\#} y^\# \neq 0$, hence, $y^\# \neq 0$. Furthermore, again from (1.15), $(x^\#, y^\#)$ is an idempotent and from (1.12) $x^\#$ is a rank 1 element. Thus, by Proposition (2.2), $y^\#$ is a rank 1 element; hence, $N(y) = 0$ by (1.16). Moreover, $\mathrm{tr}(x^\#, y^\#) = 1$. It remains to prove that $\mathrm{tr}(x,y) = 2$. To have this we calculate

$$\begin{aligned} x &= U_x y = \mathrm{tr}(x,y)x - x^\# \times y = \mathrm{tr}(x,y)x - x^\# \times U_y x \\ &= \mathrm{tr}(x,y)x - x^\# \times (\mathrm{tr}(x,y)y - y^\# \times x) \\ &= \mathrm{tr}(x,y)x - \mathrm{tr}(x,y)x^\# \times y + x^\# \times (y^\# \times x) \\ &= \mathrm{tr}(x,y)x - \mathrm{tr}(x,y)x^\# \times y + \mathrm{tr}(x^\#, y^\#)x, \end{aligned}$$

having used in the last step the identity (1.19) with $y^\#$ instead of y. Since $\mathrm{tr}(x^\#, y^\#) = 1$, it follows from the calculation above that

$$\mathrm{tr}(x,y)x = \mathrm{tr}(x,y)x^\# \times y.$$

Suppose $\mathrm{tr}(x,y) = 0$, then $y = U_y x = -y^\# \times x$ and $0 = \mathrm{tr}(x,y) = -\mathrm{tr}(x, y^\# \times x) = -\mathrm{tr}(x \times x, y^\#) = -2\mathrm{tr}(x^\#, y^\#) = -2$, a contradiction [we have used the identity (1.18)]. Thus,

$$\mathrm{tr}(x,y) \neq 0 \quad \text{and} \quad x = x^\# \times y.$$

So $x = U_x y = \mathrm{tr}(x,y)x - x$ which implies $\mathrm{tr}(x,y) = 2$ ending the proof of the present proposition. □

Proposition 2.5: Let (x,y) be an idempotent of V and let $N(x) \neq 0$. Then $y = x^{-1}$ and, in particular, $\mathrm{tr}(x,y) = 3$.

Proof: Since $N(x) \neq 0$ we can define an isotope of J in which x is the identity. Any two exceptional central simple Jordan algebras over an algebraically closed field are isomorphic. Therefore, there is just one orbit in $\mathrm{Str}(J)$. Thus there exists $g \in \mathrm{Str}\, J$ such that $g(x) = I$. We can thus define an automorphism of V by (g, g'), where[12] $g' = U_{\bar{g(I)}} g$, under which the trace $\mathrm{tr}(x,y)$, the "generic trace" of an element of a Jordan Pair, is preserved.[12] By definition of automorphism of V we have

$$I = g(x) = g(U_x y) = U_{g(x)} g'(y) = U_I g'(y) = g'(y)$$

(because $U_I z = z$ for every $x \in J$).
Hence, $\mathrm{tr}(x,y) = \mathrm{tr}(g(x), g'(y)) = \mathrm{tr}\, I = 3$. Therefore, $x = U_x y = 3x - x^\# \times y$, that is,

$$x^\# \times y = 2x \ \text{(and analogously } y^\# \times x = 2y). \tag{2.5}$$

Notice that, by (1.15) and (1.16), $(x^\#, y^\#)$ is idempotent and $N(x^\#) \neq 0$; hence, $\mathrm{tr}(x^\#, y^\#) = 3$ by the previous proof. From (2.5), (1.20), and (1.12) we get

$$\begin{aligned} 4x^\# &= (x^\# \times y)^\# = \mathrm{tr}(x^{\#\#}, y) \\ &\quad + \mathrm{tr}(y^\#, x^\#)x^\# - x^{\#\#} \times y^\# \\ &= N(x)\mathrm{tr}(x,y)y + 3x^\# - N(x)x \times y^\# \\ &= 3N(x)y + 3x^\# - 2N(x)y. \end{aligned}$$

That is, $x^\# = N(x)y$ or $y = N(x)^{-1}x^\# = x^{-1}$. The last step is indeed an identity.[25] This completes the proof of the present proposition. □

From the Propositions (2.2), (2.4), (2.5), together with the classification (2.2), the result given in Proposition (2.1) follows.

Remark 2.1: The classification of the idempotents of a Jordan Pair given in Proposition (2.1) is the exact analog of the classification of the idempotents in the Jordan Algebra J. In the case of J, indeed, the equation $x^2 = x$ has solutions only for $\mathrm{tr}\, x = 1$, $\mathrm{tr}\, x = 2$—in both cases $N(x) = 0$ (Ref. 24)—and $\mathrm{tr}\, x = 3$, in which case $x = I$.

Although the classification of the idempotents of V is analogous to that of J, the number of idempotents of V is far larger than the number of idempotents of J. Indeed *any* element of J can be completed to form an idempotent of V. [This can be easily shown using the fact that J is regular,[27] that is, for every element x in J there exists a y such that $x = U_x y$. In this case $(x, U_y x)$ is an idempotent of V.]

We can, though, restrict the class of idempotents whenever we can impose an "appropriate" normalization condition, as indicated by the following Proposition:

Proposition 2.6: Let x and y be any elements of J satisfying

$$\mathrm{tr}(x, x^*) = \mathrm{tr}(y, y^*) = \mathrm{tr}(x,y).$$

Then $x = y^*$.

Proof: Let us calculate explicitly the trace $\mathrm{tr}(x,y)$ for x as in (1.8), and y obtained from (1.8) by replacing α with β, a with d, b with f, and c with g. Then it is easily shown that

$$\begin{aligned} \mathrm{tr}(x,y) &= \alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3 + t(a\bar{d}) + t(b\bar{f}) + t(c\bar{g}) \\ &= \alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3 + 2a_\alpha d_\alpha + 2b_\alpha f_\alpha + 2c_\alpha g_\alpha, \end{aligned} \tag{2.6}$$

where a_α is the component of a for the octonionic unit e_α, with $e_0 = 1$, and where we have used the summation convention.

From (2.6) it follows that

$$\begin{aligned} \mathrm{tr}(x,x^*) &= \alpha_1\alpha_1^* + \alpha_2\alpha_2^* + \alpha_3\alpha_3^* \\ &\quad + 2a_\alpha a_\alpha^* + 2b_\alpha b_\alpha^* + 2c_\alpha c_\alpha^*, \end{aligned} \tag{2.7}$$

$$\begin{aligned} \mathrm{tr}(y,y^*) &= \beta_1\beta_1^* + \beta_2\beta_2^* + \beta_3\beta_3^* \\ &\quad + 2d_\alpha d_\alpha^* + 2f_\alpha f_\alpha^* + 2g_\alpha g_\alpha^*. \end{aligned} \tag{2.8}$$

We can now regard $A = (1/\sqrt{\lambda})(\alpha_1, \alpha_2, \alpha_3, \sqrt{2}a_\alpha, \sqrt{2}b_\alpha, \sqrt{2}c_\alpha)$ and $B = (1/\sqrt{\lambda})(\beta_1^*, +\beta_2^*, \beta_3^*, \sqrt{2}\, d_\alpha^*, \sqrt{2}f_\alpha^*, \sqrt{2}g_\alpha^*)$ as vectors in a 27-dimensional complex vector space $\mathbb{C}^{27}$ with the usual scalar product. If we put $\lambda = \mathrm{tr}(x,x^*) = \mathrm{tr}(y,y^*) = \mathrm{tr}(x,y)$, then (2.7) and (2.8) read: A and B are normalized to one; and (2.6) reads: the scalar product of A and B is equal to one. Using then the Schwartz inequality it follows that $A = B$ which immediately implies $x = y^*$. This completes the proof of Proposition 2.6.

Remark 2.2: As an immediate consequence of (2.7) we have that $\mathrm{tr}(x,x^*)$ is a real number, which is positive and, being a sum of positive numbers, is zero if and only if every summand is zero, that is if and only if x is zero. For this reason we can use $\mathrm{tr}(x,x^*)$ for a normalization condition on x.

Applying Propositions (2.1) and (2.6), the appropriate normalization conditions in the various cases of Proposition (2.1) are

for case (i) $\mathrm{tr}(x,x^*) = \mathrm{tr}(y,y^*) = 1,$ (2.9)

for case (ii) $\mathrm{tr}(x,x^*) = \mathrm{tr}(y,y^*) = 2,$ (2.10)

for case (iii) $\mathrm{tr}(x,x^*) = \mathrm{tr}(y,y^*) = 3.$ (2.11)

The normalized idempotent we get in the three cases is (x,y) with $y = x^*$. We shall call the idempotents of case (i) satisfying the normalization (2.9) "primitive normalized idempotents." The reason why we call them primitive is that x cannot be written as a sum of two orthogonal rank 1 elements in J. We shall define later on the concept of orthogonality between idempotents and it will then be easy to see that a primitive idempotent, as we have defined it now, cannot be written as a sum of two orthogonal idempotents.

4.3. The Peirce decomposition

In analogy with the theory of quadratic Jordan Algebras[17] it is possible to define the Peirce decomposition of a Jordan Pair with respect to an idempotent (x,y).[12] Define $F_2^+ = U_x U_y$; $F_1^+ = V_{x,y} - 2U_x U_y$; $F_0^+ = Id - V_{x,y} + U_x U_y$ with analogous definitions for F_i^-, $i = 0,1,2$. Then $F_2^\sigma, F_1^\sigma, F_0^\sigma$ are orthogonal projectors whose sum is the identity. Correspondingly we have the decomposition

$$V^\sigma = V_2^\sigma \oplus V_1^\sigma \oplus V_0^\sigma, \quad \text{where } F_i^\sigma V^\sigma = V_i^\sigma. \tag{3.1}$$

Denoting $V_i = (V_i^+, V_i^-)$ we can formally write

$$V = V_2 \oplus V_1 \oplus V_0. \tag{3.2}$$

Suppose now that (x,y) is a primitive normalized idempotent. That is $x^\# = y^\# = 0$, $\mathrm{tr}(x,x^*) = 1$, and $y = x^*$. Then the Peirce decomposition of V^+ with respect to it is given by

$$F_2^+ z = U_x U_y z = \mathrm{tr}(z,y)x, \tag{3.3}$$

$$F_1^+ z = (V_{x,y} - 2U_x U_y)z = z - \mathrm{tr}(z,y)x - (x \times z) \times y, \tag{3.4}$$

$$F_0^+ z = (x \times z) \times y, \tag{3.5}$$

for every z in J.

The derivation of these formulae is straightforward. For example, if we take $(x,y) = (E_1, E_1)$, where

$$E_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and we have

$$z = \begin{pmatrix} \alpha_1 & a & \bar{b} \\ \bar{a} & \alpha_2 & c \\ b & \bar{c} & \alpha_3 \end{pmatrix},$$

then

$$F_2^+ z = \begin{pmatrix} \alpha_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad F_1^+ z = \begin{pmatrix} 0 & a & \bar{b} \\ \bar{a} & 0 & 0 \\ b & 0 & 0 \end{pmatrix},$$

$$F_0^+ z = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \alpha_2 & c \\ 0 & \bar{c} & \alpha_3 \end{pmatrix}.$$

Given the Peirce decomposition of V with respect to the idempotent (x,y), the following results hold[12]:

$$V_2^+ = \mathrm{Im}(U_x), \tag{3.6}$$

$$U_{V_i^\sigma} V_j^{-\sigma} \subset V_{2i-j}^\sigma, \tag{3.7}$$

where $V_i^\sigma = 0$ for $i \neq 0,1,2$.

We now come to the main result of this section:

Proposition 3.1: Let (x,y) be a primitive normalized idempotent of V, and $V^\sigma = V_2^\sigma \oplus V_1^\sigma \oplus V_0^\sigma$ be the Peirce decomposition with respect to it. Then V_2^σ and V_0^σ are *principal inner ideals*. V_2^+ (respectively V_2^-) is the principal inner ideal generated by x (respectively y); V_0^+ (respectively V_0^-) is the principal inner ideal generated by an element $b \in V_0^+$ (resp. V_0^-) such that $b^\# \neq 0$, $N(b) = 0$; in the case $y = \lambda u$ (resp. $x = \lambda u$) (where $u^2 = u$, $\mathrm{tr}\, u = 1$), $b = I - u$; in the case $x^2 = 0$, $b = y + u$ (resp. $x + u$), where $u = -(x \times y)$ and $u^2 = u$, $\mathrm{tr}\, u = 1$; finally V_0^σ is a *maximal proper* inner ideal of J.

Proof: The fact that V_2^σ and V_0^σ are inner ideals is an immediate consequence of (3.7). Equation (3.6) implies that V_2^+ (resp. V_2^-) is the principal inner ideal generated by x (resp. y). We investigate now V_0^σ.

(A) Suppose first that $y = \lambda u$ where u is a primitive idempotent of J. Consider the element $b = I - u$, which is a trace 2 idempotent of J. We have that $b^\# = u \neq 0$ and $N(b) = 0$ by (1.16). Thus, by Proposition (2.3) given b_1 such that $\mathrm{tr}(b^\#, b_1) = 1$ we can write

$$b = b^\# \times (b \times b_1).$$

We can choose $b_1 = \lambda x$, since $\mathrm{tr}(\lambda x, b^\#) = \mathrm{tr}(x, \lambda u) = \mathrm{tr}(x,y) = 1$. Hence, $b = b^\# \times (b \times \lambda x) = y \times (b \times x)$ so that $b \in V_0^+$ by (3.5). We know already that V_0^+ is a proper inner ideal, therefore, $U_b J \subseteq V_0^+$ by definition of inner ideal. But it is a known result[27] that $U_b J$ is a maximal proper inner ideal whenever $b^\# \neq 0$, $N(b) = 0$. Hence $V_0^+ = U_b J$ and the proposition is proven in the case (A).

The result for V_0^- is indeed analogous.

(B) By (2.3) and Proposition (2.6) the only case remaining is $y^2 = 0 = x^2$. We have that $u = -x \times y = I - 2x \cdot y$ has trace 1 and is a rank 1 element by (1.20); hence it is a primitive idempotent of J. Furthermore, as we shall prove in Proposition 3.3, $u \cdot y = u \cdot x = 0$. Thus if we take $b = y + u$, we have

$$b^\# = u \times y = -y \neq 0 \quad \text{and } N(b) = N(b^\#)^{1/2} = 0.$$

We can now repeat the conclusion of part (A) of this proof, choosing $b_1 = -x$ to show that $b \in V_0^+$ and, therefore, $U_b J = V_0^+$ by maximality. The proof of Proposition 3.1 is, thus, complete. □

Proposition 3.2: Let (x,y) be a primitive normalized idempotent and let $x^2 = y^2 = 0$. Then $x \in V_0^-$, $y \in V_0^+$, for V_0 in the Peirce decomposition defined by (x,y).

Proof: We shall prove that $x \in V_0^-$, the other having an analogous proof. We define $u = -x \times y$. As we have seen in part (B) of the proof of the previous proposition, $u^2 = u$, $\mathrm{tr}\, u = 1$, and $u \cdot x = 0$. Thus $(x \times y) \times x = -u \times x = x$; that is, $x \in V_0^-$ by (3.5) and the proof is complete. □

Remark 3.1: For (x,y) as in Proposition 3.2, it follows from part B of the proof of Proposition 3.1 and the previous proposition that $u = -x \times y$ belongs to both V_0^+ and V_0^-.

Proposition 3.3: If $x^\# = 0$, then $(x \times z) \cdot x = 0$ for every z in J.

Proof: By (2.3) we can write $x^2 = \lambda x$, $\operatorname{tr} x = \lambda$, where the range of λ includes zero. Using the identity (1.24) we calculate

$$\begin{aligned}(x\times z)\cdot x &= 2(x\cdot z)\cdot x - x^2 \operatorname{tr} z - (z\cdot x)\operatorname{tr} x - x \operatorname{tr}(z\cdot x) \\ &\quad + x \operatorname{tr} x \operatorname{tr} z \\ &= U_x z + x^2\cdot z - x^2 \operatorname{tr} z - (z\cdot x)\operatorname{tr} x - x \operatorname{tr}(z\cdot x) \\ &\quad + x \operatorname{tr} x \operatorname{tr} z \\ &= \operatorname{tr}(x\cdot z)x + x^2\cdot z - x^2 \operatorname{tr} z \\ &\quad - (z\cdot x)\operatorname{tr} x - x \operatorname{tr}(z\cdot x) \\ &\quad + x \operatorname{tr} z \operatorname{tr} x \\ &= \lambda x\cdot z - \lambda x \operatorname{tr} z - \lambda x\cdot z + \lambda x \operatorname{tr} z = 0,\end{aligned}$$

which proves the Proposition. □

Proposition 3.4: $V_0^+ \subseteq \{z : z\cdot y = 0\}$, where V_0^+ is the Peirce decomposition with respect to a primitive normalized idempotent (x,y).

Proof: It follows from (3.5) and the previous proposition. □

We examine now the subpairs composed of inner ideals that we shall use in the next section to define the geometrical objects.

Definition 3.1: A pair of submodules (W^+, W^-) of (V^+, V^-) is a subpair if $U_{W^\sigma} W^{-\sigma} \subset W^\sigma$.

Proposition 3.5: The only pairs of $V_i^\sigma, V_j^{-\sigma}$ which form a subpair of V are $(V_i^\sigma, V_i^{-\sigma})$, $i = 0, 1, 2$.

Proof: It follows from the Definition 3.1 and from (3.7) that we have subpairs if $2i - j = i$, that is, $i = j$, which proves the proposition. □

We want to introduce the concept of orthogonality between idempotents. Before doing so we need the following result, proven in Ref. 12:

Proposition 3.6: Let $x = (x^+, x^-)$ and $y = (y^+, y^-)$ be idempotents of V such that $x^\sigma \in V_0^\sigma(y)$, where $V_0^\sigma(y)$ is in the Peirce decomposition defined by y. Then also, $y^\sigma \in V_0^\sigma(x)$ and $x + y = (x^+ + y^+, x^- + y^-)$ is an idempotent. Furthermore,

$$V_{x^\sigma, y} = V_{x, y^\sigma} = V_{y^\sigma, x} = V_{y, x^\sigma} = 0. \tag{3.8}$$

Whenever $x^\sigma \in V_0^\sigma(y)$ we shall write $x \in V_0(y)$.

Definition 3.2: Two nonzero idempotents x and y are called *orthogonal* if $y \in V_0(x)$.

It follows from the previous proposition that orthogonality is a symmetric relation. It is also evident that primitive idempotents cannot be written as a sum of two orthogonal idempotents.

4.4. Geometry of V

Given a primitive normalized idempotent $x = (x^+, x^-)$ in V, we associate to it

(i) a point $x_\star = V_2(x)$,

(ii) a line $x^\star = V_0(x)$.

We say that

(a) $x_\star$ is incident to $y^\star$ (written $x_\star | y^\star$) if $V_2(x) \subset V_0(y)$,

(b) $x_\star (x^\star)$ is connected to $y_\star (y^\star)$ (written $x_\star \cong y_\star$ $(x^\star \cong y^\star)$ if $V_2(x) \subset V_2(y) \oplus V_1(y)$,

(c) $x_\star$ is connected to $y^\star$ (written $x_\star \cong y^\star$) if $V_2(x) \subset V_0(y) + V_1(y)$.

Proposition 4.1: Let $z = (z^+, z^-)$ and $x = (x^+, x^-)$ be two primitive normalized idempotents. Then,

$z^+ \in V_i^+(x)$ if and only if $z^- \in V_i^-(x)$.

Proof: Suppose $z^+ \in V_2^+(x)$; then

$$z^+ = \operatorname{tr}(x^-, z^+)x^+,$$

$$z^- = (\operatorname{tr}(x^-, z^+))^* x^- = \operatorname{tr}(x^+, z^-)x^- \in V_2^-(x).$$

Suppose now $z^+ \in V_0^+(x)$; then,

$$\begin{aligned} z^+ &= (x^+ \times z^+) \times x^-, \\ z^- &= (x^+ \times z^+)^* \times (x^-)^* \\ &= ((x^+)^* \times (z^+)^*) \times (x^-)^* \\ &= (x^- \times z^-) \times x^+ \in V_0^-(x).\end{aligned}$$

From this it follows that if $z^+ \in V_1^+(x)$, then $z^- \in V_1^-(x)$. The same is obviously true inverting the signs; hence, the proposition is proven.

This proposition allows us to choose as representative of points and lines just the elements of J which generate V_2^+ and V_0^+.

As we did in the case of A_2 we want to consider the subgroup G of the structural group of J (which is isomorphic to the automorphism group of the Jordan Pair V) which maps primitive normalized idempotents into themselves.

The structural group of J has 79 generators, coming from the three-grading of the (complex) Lie algebra of $\mathscr{E}_7$. Out of these 79 generators, 78 form the complex Lie algebra E_6,[15] and the 79th is just a change of scale. When we consider the subgroup G of Str (J) we obviously get from this last generator the (compact real form) U(1), and from $\mathscr{E}_6$ a subgroup H such that, for any x in J,

$$(g_+(x))^* = g_-(x^*) \tag{4.1}$$

for every $g_+, g_- \in H$, where $(g_+, g_-) \in \operatorname{Aut}(V)$.

Proposition 4.2: Equation (4.1) holds for every primitive normalized idempotent (x, x^*) if and only if

$$\operatorname{tr}(x, x^*) = \operatorname{tr}(g_+(x), g_+(x)^*). \tag{4.2}$$

Proof: If (4.1) is true, then (4.2) follows because $\operatorname{Aut}(V)$ preserves the generic trace. Vice versa if (4.2) is true

$$\begin{aligned}\operatorname{tr}(g_+(x), (g_+(x))^*) = 1 &= \operatorname{tr}(g_-(x^*), (g_-(x^*))^*) \\ &= \operatorname{tr}(g_+(x), g_-(x^*)).\end{aligned}$$

Hence by Proposition 2.6, $g_+(x) = (g_-(x^*))^*$, and the present proposition is proved.

From Eq. (4.2) we derive that H must be compact. In fact it follows from Proposition 2.6 that H must be a subgroup of U(27), which is compact.

Furthermore if we denote by G_+ and G_- the generators of g_+ and g_-, we get

$$\operatorname{tr}(G_+(x), x^*) = -\operatorname{tr}(x, (G_+(x))^*) \tag{4.3}$$

which implies that the Lie algebra of H is real. In fact multiplying G_+ by λ we still have an equality, which implies λ real.

Therefore $H = \mathscr{E}_{6,0}$ (Ref. 23), the compact subgroup of $\mathscr{E}_6$, generated by the compact real form $E_{6,0}$ of E_6. $\mathscr{E}_{6,0}$ does indeed preserve $\operatorname{tr}(x, x^*)$ and is obviously the maximal compact subgroup of $\mathscr{E}_6$. $E_{6,0}$ can be realized via Tit's construction,[28] as the Lie algebra of the transformations

$$x \to (A, x, B) + iC\cdot x, \tag{4.4}$$

where A,B,C are traceless Hermitian matrices of the real $\mathcal{M}_3^8$ and $(\ ,\ ,\ )$ is the associator. If G_+ is the transformation (4.4), then

$$G_-(x^*) = (A,x^*,B) - iC\cdot x^*. \tag{4.5}$$

The 27-dimensional representation of $\mathscr{E}_{6,0}$ is complex and it is not equivalent to its complex conjugate.

Remark: The compact real form $E_{6,0}$ must not be confused with the Lie algebra of the structural group of the real $\mathcal{M}_3^8$, which is also the collineation group of the Moufang plane.[23] The latter is the *non*compact form $E_{6,0^*}$ of signature -26. $\mathscr{E}_{6,0^*}$ has only real representations.[29]

Proposition 4.3: If $g\in\mathscr{E}_{6,0}\otimes U(1)$, then $x\in V_i(e)$ implies $g(x)\in V_i(g(e))$ $(i=0,1,2)$ for x,e primitive normalized idempotents.

Proof: It follows directly from the definitions of automorphism and Peirce decomposition. □

Defining the following natural action of $\mathscr{E}_{6,0}\otimes U(1)$ on the points

$$g(x_\star) = (g(x))_\star, \tag{4.6}$$

it stems from Proposition 4.3 and the previous considerations that $\mathscr{E}_{6,0}\otimes U(1)$ preserves the relations in the geometry. [This is the reason for using $\mathscr{E}_{6,0}\otimes U(1)$ in Proposition 4.3, which is true $\forall g\in \mathrm{Aut}(V)$.]

Definition 4.1: Two points $x_\star$ and $y_\star$ are orthogonal $(x_\star \perp y_\star)$ if $x\in V_0(y)$.

We now state the following important result, that we shall prove in the Appendix B:

Proposition 4.4: $\mathscr{E}_{6,0}$ acts transitively on points and on triples of mutually orthogonal points. The maximal subgroup of $\mathscr{E}_{6,0}$ leaving a point invariant is $SO(10)\otimes\tilde{U}(1)$; therefore, the plane we are considering is the homogeneous space $\mathscr{E}_{6,0}/SO(10)\otimes U(1)$.

[Here we have used the tilde over U(1) to distinguish this as a subgroup of $\mathscr{E}_{6,0}$ from the "overall" phase group U(1) which is obviously outside $\mathscr{E}_{6,0}$. The action of $\tilde{U}(1)$ is written explicitly in the proof of Proposition 4.4 (see Appendix B).]

To proceed with the discussion of the geometry, we recall the following definitions:

Definition 4.2: A collineation (resp. correlation) is a bijective map of the points of a plane $\mathscr{P}$ onto the points (resp. lines) of a plane $\mathscr{P}'$, and of the lines of $\mathscr{P}$ onto the lines (resp. points) of $\mathscr{P}'$, preserving the incidence and connectedness relations.

Definition 4.3: A correlation of $\mathscr{P}$ onto itself is a *duality*.

Definition 4.4: A duality of order 2 is a *polarity*.

Definition 4.5: A point $x_\star$ in $\mathscr{P}$ is isotropic with respect to the polarity π if $x_\star | \pi(x_\star)$.

Definition 4.6: A polarity with respect to which no point is isotropic is called an orthocomplementation.

It is immediate to see that the correspondence we made $\pi: x_\star \leftrightarrow x^\star$ is an orthocomplementation.

This is a fundamental result we are going to use in building a quantum theory on the geometry we have defined. The orthocomplementation is indeed needed in defining both the proposition system and the states of the quantum logic.

An orthocomplementation is an "elliptic" polarity defining an "elliptic geometry".[23]

Planes, called Hjelmslev–Moufang planes, defined over an exceptional central simple Jordan algebra on a split Cayley algebra, and therefore including our complex J, have been investigated in Ref. 30. We want to stress that the Hjelmslev–Moufang plane, obtained by complexifying real $\mathcal{M}_3^8$,[13] although similar to ours, is defined over a hyperbolic polarity (i.e., a polarity admitting isotropic points). It is, therefore, difficult to give it a quantum mechanical interpretation. We shall compare the two planes in Appendix C, where we shall also see how certain results valid in the Hjelmslev–Moufang plane hold true in the plane we have defined. In particular we can reproduce the following results:

(1) If $x_\star \not\perp y_\star$ there is a unique line incident to both of them,

(2) If $x_\star \perp y_\star$ there is at least one line incident to both of them.

We shall see in the next section that two connected lines intersect in more than one point, which shows that this geometry is not projective. To do this we introduce the concept of point spaces.

5. THE POINT SPACES

We want to show in this section that the reason why we have a nonprojective geometry (in the sense that two lines may intersect in more than one point) is that in the structure we are dealing with, the following three related facts occur:

(1) the existence of nilpotent elements in J (due to the fact that J is defined over a split Cayley Algebra $\mathscr{C}$);

(2) the existence of connected points;

(3) the existence of *non*-principal inner ideals in J.

To do this we must go back to the general theory of the algebra $\mathscr{H}$ of the 3×3 Hermitian matrices over a Cayley algebra $\mathscr{C}$. There are just three kinds of inner ideals[27] in $\mathscr{H}$: (i) $\mathscr{H}$; (ii) $\mathscr{B} = b^\#\times\mathscr{H}$, where $N(b)=0, b^\#\neq 0$; (iii) $\mathscr{B}^\# = 0$ (i.e., $\mathscr{B}$ contains only rank 1 elements).

The first two ideals are the principal inner ideals generated by I (the identity in $\mathscr{H}$) and b (resp.).

If $\mathscr{C}$ is a division algebra then the third one is of the form ϕb, ϕ being the field on which $\mathscr{C}$ is defined, and in this case $\mathscr{B} = U_b(\mathscr{H})$ is principal.

The ideal $\mathscr{B}^\# = 0$ is called a point space. If $\mathscr{C}$ is a *division* algebra any point space is one dimensional. [The proof of this statement is very easy. Suppose that b and b_1 belong to a point space, then also, $b+b_1$ must belong to it by definition of ideal. In particular $(b+b_1)^\# = 0$, which implies $b\times b_1 = 0$. It is possible to show that the action of the structural group of $\mathscr{H}$ preserves the inner ideals, thus we can take $b = E_1$. The condition $b\times b_1 = 0$ implies then,

$$b_1 = \begin{pmatrix} \alpha & a & \bar{c} \\ \bar{a} & 0 & 0 \\ c & 0 & 0 \end{pmatrix},$$

and $b_1^{\#\#} = 0$ implies

$$n(a) = n(c) = ac = 0\,. \tag{5.1}$$

If $\mathscr{C}$ is a division algebra it follows from (5.1) that $a = c = 0$

and $b_1 = b$.] But if $\mathscr{C}$ is split, as in our case, the point space containing b is not one dimensional. It is possible to prove[27] that starting from b we can build a point space whose maximal dimension over ϕ is either 5 or 6. These are called the maximal point spaces of the first and second kind. For $b = E_1$ these can be written in the form

(i) $E_1 + \mathscr{C}\epsilon[1,2]$ first kind,

(ii) $E_1 + \epsilon\mathscr{C}[1,2] + \phi\epsilon[1,3]$ second kind,

where ϵ is a primitive idempotent of $\mathscr{C}$ [therefore $\dim_\phi \mathscr{C}\epsilon = 4$ (Ref. 31)]. We have used the notation $a[1,2]$ for the matrix with entry a in the (1,2) position and $\bar{a}$ in the (2,1) position and all others equal to zero.

We can extend the definition of point space to our Jordan pair just by doubling the point space in J with its conjugate. The group $\mathscr{E}_{6,0}$ being a subgroup of Str (J) will then preserve the inner ideals and in particular the point spaces.[27]

Thus we get that the necessary condition for two points $x = (x^+,x^-)$ and $y = (y^+,y^-)$ to belong to a point space is $x^+ \times y^+ = 0$. From Appendix C it follows that the points x and y are connected. Vice versa, if two points x and y are connected, they can always be imbedded in a point space.

We show now with an example that the maximal point spaces of the first kind (dim. 5) are just the inner ideals occurring in the intersection of two connected lines. (Note that the intersection of two inner ideals is an inner ideal). We take the lines e_3 and a_3 associated to the primitive normalized idempotents $e_3 = (E_3,E_3)$ and $a_3 = (\bar{a}^*[2,3],\bar{a}[2,3])$, where $n(a) = 0$ and $\bar{a}^*a + \bar{a}a^* = 1$.

We want to write explicitly $V_0^+(e_3)$ and $V_0^+(a_3)$. To do this we note that $x \in V_0^+(e)$ if and only if

$$(e^+ \times x) \times e^- = x\,. \tag{5.2}$$

It is easy to check that $x \in V_0^+(e_3)$ iff

$$x = \begin{pmatrix} \alpha_1 & x_3 & 0 \\ \bar{x}_3 & \alpha_2 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and that $x \in V_0^+(a_3)$ iff

$$x \equiv \begin{pmatrix} \alpha_1 & x_3 & \bar{x}_2 \\ \bar{x}_3 & \alpha_2 & x_1 \\ x_2 & \bar{x}_1 & \alpha_3 \end{pmatrix} = \begin{pmatrix} \alpha_1 & (x_3\bar{a}^*)a & (\bar{x}_2a^*)\bar{a} \\ \bar{a}(a^*\bar{x}_3) & 0 & t(\bar{a}^*\bar{x}_1)\bar{a} \\ a(\bar{a}^*x_2) & t(\bar{a}^*\bar{x}_1)a & 0 \end{pmatrix}, \tag{5.3}$$

using Eq. (5.2). We want to prove that (5.3) implies

(i) $\alpha_2 = \alpha_3 = 0$, (5.4)

(ii) $x_1 = \lambda\bar{a}$ for any $\lambda \in \mathbb{C}$, (5.5)

(iii) $x_3 = ya$ for any $y \in \mathscr{C}$, (5.6)

(iv) $x_2 = az$ for any $z \in \mathscr{C}$. (5.7)

Proof: (i) (5.4) is immediate.

(ii) From (5.3) it follows that $x_1 = t(\bar{a}^*\bar{x}_1)\bar{a}$; therefore, $x_1 = \lambda\bar{a}$ for $\lambda = t(\bar{a}^*\bar{x}_1)$; vice versa if $x_1 = \lambda\bar{a}$ for any $\lambda \in \mathbb{C}$ then $t(\bar{a}^*\bar{x}_1)\bar{a} = \lambda t(\bar{a}^*a)\bar{a} = \lambda\bar{a} = x_1$, which proves (5.5).

(iii) $x_3 = ya$ for $y = x_3\bar{a}^*$; vice versa if $x_3 = ya$ for every y in $\mathscr{C}$ we have, from the Moufang identity $[(yx)a]x = y(xax)$, that $[(ya)\bar{a}^*]a = y(a\bar{a}^*a) = y[a(\bar{a}^*a + \bar{a}a^* - \bar{a}a^*)]$ $= ya - (a\bar{a})a^* = ya$, which proves (5.6). □

(iv) is analogous to (iii).

We can therefore write the two lines as

$$V_0^+(e_3) = \begin{pmatrix} \mathbb{C} & \mathscr{C} & 0 \\ \mathscr{C} & \mathbb{C} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad V_0^+(a_3) = \begin{pmatrix} \mathbb{C} & \mathscr{C}a & \mathscr{C}\bar{a} \\ \bar{a}\mathscr{C} & 0 & \mathbb{C}\bar{a} \\ a\mathscr{C} & \mathbb{C}a & 0 \end{pmatrix}. \tag{5.8}$$

Since $n(a) = 0$ it follows that[31] the dimension of Ca is 4.

It is therefore possible to read directly from (5.8) that the dimension of the lines is 10 and that the dimension of their intersection is 5. The intersection is therefore a maximal point space of the first kind. It is proven in Ref. 30 that any maximal point space of the second kind can be obtained by considering the set of points $x_\star$ incident to a given line $y^\star$ and connected to a given $z_\star$ which is itself connected to $y^\star$. That is, $\{x_\star : x_\star \mid y^\star, x_\star \cong z_\star$, where $z_\star \cong y^\star\}$ contains a maximal point space of the second kind. This result, which is proved in Ref. 30 for the Hjelmslev–Moufang planes, can be translated to our geometry. We omit the proof. In ending this section, and with it the description of the mathematical framework of this paper, we want to emphasize that the point spaces are *the* new interesting object of the whole structure. Although they do not show up explicitly, neither in the geometry nor in the proposition system of the quantum mechanics (as we shall see in the next section), they are responsible for making the geometry nonprojective and, consequently, the proposition system not a lattice. Their appearance makes, in our opinion, this plane peculiar and interesting.

6. THE QUANTUM LOGIC

6.1. The proposition system

The language of quantum mechanics has always been identified with the language of projective geometry, the points of the geometry being identified to the density matrices of the (pure) states, and the lines and hyperplanes with the propositions which are not atoms. The automorphism group of the geometry—that is its collineation group—is, however, larger than the automorphism group of the quantal structure, because collineations need not preserve the traces (which are the canonical measures defining the quantal states) nor the orthogonality, which has no projective meaning. In mathematical language we can say that the quantum logic requires an automorphism group which preserves an elliptic polarity. For this reason, for instance, the automorphism group of the quantum system described in Ref. 8 for the Moufang plane is $\mathscr{F}_{4,0}$, whereas the collineation group of the plane itself is $\mathscr{E}_{6,0^*}$, which contains $\mathscr{F}_{4,0}$ as maximal compact subgroup. We need not investigate (although it is an interesting task) the collineation group of our plane, since we have determined the group which is needed in describing automorphisms of the quantum system. This is the compact group $\mathscr{E}_{6,0} \otimes U(1)$, which preserves the trace tr (x,x^*) and the orthocomplementation (i.e., our elliptic polarity).

The possibility of giving a quantum mechanical interpretation to our structure, despite the nonprojectivity, is

therefore beginning to appear.

Let us define the proposition system. We obviously identify the propositions with the geometrical objects: points and lines (the *principal* inner ideals of the pair). They form a partially ordered set, with ordering given by the set inclusion of the inner ideals. The plane itself, which is the principal inner ideal generated by an invertible element, is the trivial proposition.

We have an orthocomplementation $a \to a^{\perp}$ which is our polarity $a_{\star} \leftrightarrow a^{\star}$. Therefore, we can define orthogonality:

$$a \perp b \quad \text{if} \quad a < b^{\perp},$$

which is symmetric.

We have, indeed, the basic building blocks for constructing a quantum logic. The only concepts of the standard theory which are weakened are the concepts of greatest lower bound ("meet") and of least upper bound ("join"). *They are not defined here for every pair of propositions.* In fact, two connected lines do not intersect in a unique point and the "join" of two connected points is not a unique line (geometrically the second is indeed the dual of the first sentence). *Therefore we do not have a lattice structure.* But the subsets made of nonconnected points and lines *are* sublattices of the partially ordered set.

We want to stress, at this point, that the concept of "meet" (with its dual "join") is the only weak concept in the axiomatics of quantum theory. In the standard theory,[32] to define the proposition c which is the meet of a and b we are forced to consider an infinite sequence of propositions: $a,b,a,b, \cdots$. Indeed, if we just measure a and then b we can obtain a different result with respect to measuring b and then a, because the measurement of a proposition in general affects the initial state. This is just the peculiarity of quantum mechanics. The only way to overcome this obstacle is to say that c is true when the measure of the infinite sequence $a,b,a,b, \cdots$ gives the result 1. It is not possible physically to perform such a measurement; therefore, there cannot be any experimental evidence for assuming the existence of the meet of a and b as an axiom for a quantum theory. Therefore, a quantum logic which is not based on a lattice structure is perfectly conceivable.

Remark: Although it would violate the geometrical structure, and is not actually required for a quantum mechanics, one could still think of including the point spaces in the proposition system, in order to save the lattice structure. In this way one would indeed reach the goal of having a "meet" for any pair of lines (for the connected lines it would be a maximal point space of the first kind) but, at the same time, one would lose the orthocomplementation. Contrary to the concept of meet, the concept of orthocomplementation has an immediate intuitive meaning: it states the existence of the proposition "not a" for any proposition a.

The lack of a lattice structure affects the concepts of superposition between two propositions. By using the analogy between geometric points and density matrices of pure states, it is evident that the usual concept of superposition between states in a Hilbert space is translated and generalized in geometrical language by saying that the point P_3 is a superposition of the points P_1 and P_2 whenever P_3 is incident to *the* line incident to P_1 and P_2. In our geometry if P_1 and P_2 are connected points they do not define a unique line incident to both. Accordingly the concept of superposition of P_1 and P_2 fails. But we emphasize again that this happens only if P_1 and P_2 are connected (the *existence of point spaces of dimension greater than* 1 is responsible for this). If P_1 and P_2 are nonconnected, indeed, there exists just one line incident to both of them and any P_3 on this line is a well-defined superposition of P_1 and P_2.

Remark: We stress that this kind of failure of the concept of superposition has no analog in the usual quantum theory. In that case the concept of superposition fails only in the presence of superselection rules, which occur when the lattice is reducible. When we have a superselection rule we cannot define the superposition of P_1 and P_2, for P_1 and P_2 in two different irreducible components of the lattice, simply because there is no way of defining it as a point. In our case we have an *irreducible proposition system* (it is easy to see that the center is trivial) and the superposition of P_1 and P_2 is not defined if P_1 is connected to P_2 because it is not *unique*. Roughly speaking, in a reducible lattice we have no superposition of P_1 and P_2; in our irreducible proposition system we have too many.

A similar unusual feature shows up when we consider the sum of two connected points P_1 and P_2. We get that $P_1 + P_2$ is a point again. In the usual quantum theory the closest analog to this happens only in a reducible lattice and in this case the state related to the sum is a mixture of the states related to P_1 and P_2. But in our case we have an irreducible system and $P_1 + P_2$ is not related to a mixture but to another pure state. Once again the existence of point spaces of dimension greater than 1 is responsible for this peculiarity.

We now consider the concept of compatibility. This is essentially the concept which makes the distinction between a classical and a quantal system. In a classical system, indeed, all the propositions are compatible. In the language of lattice theory, a classical system is described as a Boolean lattice.[10] In a Boolean lattice any triple of propositions a,b,c satisfies the following distributive laws with respect to the "meet" ($\wedge$) and the "join" ($\vee$):

$$\begin{aligned} a \wedge (b \vee c) &= (a \wedge b) \vee (a \wedge c), \\ a \vee (b \wedge c) &= (a \vee b) \wedge (a \vee c). \end{aligned} \tag{6.1}$$

We shall define a subset $\mathcal{M}$ of the proposition system $\mathcal{P}$ as a compatible set of propositions if $\mathcal{M}$ generates a Boolean sublattice of $\mathcal{P}$.

It follows immediately from the definition that two connected points are not compatible (because they do not even generate a sublattice), and that two orthogonal points are compatible.

We want to show that any two propositions a,b, such that $a < b$, are compatible. If $a < b$, then $a \vee b = b$ and $a \wedge b = a$ by definition. Furthermore, a and $b^{\perp}$ are nonconnected, because they are orthogonal; therefore, $a \vee b^{\perp}$, $a^{\perp} \vee b$, $a \wedge b^{\perp}$, $a^{\perp} \wedge b$ exist. Hence, a,b generate a sublattice S and it is very easy to show, applying the usual arguments of projective geometry, that S is Boolean.

To the property of compatibility of two ordered propositions is given the name of orthomodularity or weak modularity. We can therefore state that our proposition system is orthomodular.

It is well known[33] that any projective geometry (of finite dimension) has a property stronger than orthomodularity. This is the so-called modular law:

$$a \vee (b \wedge c) = (a \vee b) \wedge c \quad \text{for all } a < c.$$

It is clear that we do not have this property because $a \vee b$ and $b \wedge c$ may not be defined. However, this law is satisfied by any sublattice of our proposition system.

We stress again that the modular law has no immediate physical meaning. It is even incompatible with the concept of localizability in quantum mechanics, due to the infinite dimensionality of the Hilbert space on which the position operator is defined. A Hilbert space is indeed orthomodular and it is modular only in finite dimension. The assumption of the modular law in a quantal system has therefore no physical justification, even in finite dimension. A nonmodular, but weakly modular, proposition system is perfectly conceivable.

6.2. The states

There is no quantum theory without the definition of a function (state) which "measures" the propositions in the system and is invariant under the automorphism group of the proposition system. Geometrically this function is related to the invariant distance function between two points (whenever it is possible to define it). The distance between two points x and y is interpreted physically as the measure $\alpha_x(y)$ of y in the "state" α_x associated to x. In this case x plays the role of the density matrix associated to a (pure) state in the usual quantum theory of the Hilbert spaces.

In the standard theory the state α_x is a probability function, and as such must satisfy

$$\begin{aligned} &(1)\quad 0 \leqslant \alpha_x(y) \leqslant 1 \quad \text{for every proposition } y, \\ &(2)\quad \alpha_x(x) = 1, \\ &(3)\quad \alpha_x(y_1 \vee y_2) = \alpha_x(y_1) + \alpha_x(y_2) \quad \text{if } y_1 \perp y_2. \end{aligned} \tag{6.2}$$

Having defined points and lines by use of the Peirce decomposition it is natural to associate them to the Peirce projectors F_2 and F_0 on the spaces V_2 and V_0. It is natural as well to make use of these projectors to determine the "distance" between two points and between a point and a line. What we have to do is to project one point on the Peirce space defining the other point (or the line) and take the square root of the trace. (The reason why we must take the square root will be clear in the sequel).

We thus define for any point x the following "measure":

$$\text{(i)}\quad \alpha_x(y_\star) = (\operatorname{tr}(U_{x^+}\, U_{x^-}\, y^+, U_{x^-}\, U_{x^+}\, y^-))^{1/2} = |\operatorname{tr}(x^+, y^-)|, \tag{6.3}$$

$$\text{(ii)}\quad \alpha_x(y^\star) = (\operatorname{tr}((y^+ \times x^+) \times y^-, (y^- \times x^-) \times y^+))^{1/2}.$$

Notice that $x = (x^+, x^-)$ is defined up to a phase, and so is y, but (6.3) is independent of the phases of x and y.

The "measure" (6.3) is manifestly $\mathscr{E}_{6,0} \otimes \mathrm{U}(1)$ invariant and, because it is defined by projectors and x and y are normalized, clearly satisfies (1) and (2) of (6.2). Furthermore, $\alpha_x(y^\star) = 1$ if and only if $x_\star \,|\, y^\star$.

To have an example of how this "measure" works consider $x_\star$ a generic point and $y_\star = (e_1)_\star$, where $e_1 = (E_1, E_1)$. This represents also the most general case because of the transitivity of $\mathscr{E}_{6,0} \otimes \mathrm{U}(1)$ on points and the invariance property of our "measure." In other words, we can always transform y into e_1 (thus changing also x) and still have the same measure. Writing x^+ as

$$x^+ = \begin{pmatrix} \alpha & a & \bar{b} \\ \bar{a} & \beta & c \\ b & \bar{c} & \gamma \end{pmatrix}, \tag{6.4}$$

and denoting by F_2, F_0 the projectors on $V_2(y)$, $V_0(y)$, we have (see Sec. 4.3)

$$F_2^+ x^+ = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad F_0^+ x^+ = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta & c \\ 0 & \bar{c} & \gamma \end{pmatrix}. \tag{6.5}$$

We therefore get

$$\alpha_x(y_\star) = (\alpha\alpha^*)^{1/2} = |\alpha|\,, \tag{6.6}$$

$$\begin{aligned} \alpha_x(y^\star) &= (\beta\beta^* + \gamma\gamma^* + c\bar{c}^* + c^*\bar{c})^{1/2} \\ &\leqslant (\operatorname{tr}(x^+, x^-))^{1/2} = 1\,. \end{aligned} \tag{6.7}$$

From (6.7) and (2.7) it follows that $\alpha_x(y^\star) = 1$ if and only if $\alpha = a = b = 0$, that is, if and only if $x_\star \,|\, y^\star$.

In particular, by taking x^+ and y^+ as real octonionic matrices [or such that can be mapped into real octonionic matrices by an $\mathscr{E}_{6,0} \otimes \mathrm{U}(1)$ transformation], then (6.6) and (6.7) imply that α_x restricted to the real octonionic case is just the *unique probability function* defined in Ref. 8 on the Moufang plane. In fact from (6.6) and (6.7), in the real octonionic case,

$$\begin{aligned} &\alpha_x(y_\star) = |\alpha| = \alpha = \operatorname{tr}(E_1, x^+), \\ &\alpha_x(y^\star) = (\beta^2 + \gamma^2 + 2c\bar{c})^{1/2} = (\beta^2 + \gamma^2 + 2\beta\gamma)^{1/2} \\ &\qquad\qquad = \beta + \gamma = \operatorname{tr}(x^+, E_2 + E_3), \end{aligned} \tag{6.8}$$

which is exactly the measure defined in Ref. 8.

In showing (6.8) we have used the fact that for a trace 1, real octonionic idempotent x^+ in the form (6.4), $c\bar{c} = \beta\gamma$. To see this suppose first that $\alpha = 1$; then

$$x^+ = \begin{pmatrix} 1 \\ \bar{a} \\ b \end{pmatrix} (1 \;\; a \;\; \bar{b}), \quad \operatorname{tr}(x^+) = 1. \tag{6.9}$$

Therefore $n(a) = n(b) = n(c) = 0$ and in particular $c = 0 = \beta = \gamma = \beta\gamma$. If $\alpha \neq 1$, then either β or γ or both are different from zero.

Suppose $\beta \neq 0$; then

$$x^+ = \begin{pmatrix} a/\sqrt{\beta} \\ \sqrt{\beta} \\ \bar{c}/\sqrt{\beta} \end{pmatrix} (\bar{a}/\sqrt{\beta} \quad \sqrt{\beta} \quad c/\sqrt{\beta}), \tag{6.10}$$

which implies $c\bar{c}/\beta = \gamma$, that is, $c\bar{c} = \beta\gamma$. This proves that (6.8) is, indeed, correct.

From what we have seen so far the "measure" defined in (6.3) is the most natural one and extends the probability function defined in Ref. 8 on the Moufang plane. Let us now consider the additivity property (3) of (6.2).

Consider the line $(e_1)^{\star}$ and the point $x_{\star}$ where

$$x^{+} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{6.11}$$

Take in $(e_1)^{\star}$ the following two pairs of orthogonal points: $(e_2)_{\star}$, $(e_3)_{\star}$ and $c_{\star}$, $(c^*)_{\star}$ where $c^{+} = a[2,3]$, $(c^*)^{+} = a^*[2,3]$, $a\bar{a} = 0$, and $a\bar{a}^* + a^*\bar{a} = 1$.

We then easily get

$$\alpha_x((e_1)^{\star}) = \tfrac{1}{2}, \tag{6.12}$$

$$\alpha_x((e_2)_{\star}) = \tfrac{1}{2}, \quad \alpha_x((e_3)_{\star}) = 0, \tag{6.13}$$

$$\alpha_x(c_{\star}) = \alpha_x((c^*)_{\star}) = 0. \tag{6.14}$$

This shows that, for our "measure," the condition of orthogonality of y_1 and y_2 in (3) of (6.2) is not sufficient to guarantee the additivity. In fact (6.12) is equal to the sum of the "measurements" in (6.13) but not of those in (6.14). The vice versa would hold if we took, for instance $x_{\star} = c_{\star}$. It is clear then that the condition for additivity must relate to the triple x, y_1, and y_2. Since the measure is additive whenever we can transform x, y_1, and y_2 into "real octonionic matrices" (by abuse of notation: x, y_1, y_2 are related to pairs not to matrices) then we know that $\alpha_x(y^{\star})$ is additive (because the probability function defined in Ref. 8 on the Moufang plane is additive). This condition is achieved[34] whenever

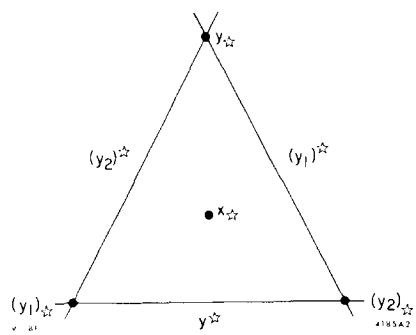
$$x_{\star} \cong (\hat{y})^{\star} \quad \text{implies} \quad x_{\star} \,|\, (\hat{y})^{\star}, \tag{6.15}$$

where $(\hat{y})^{\star}$ is any of the lines determined by y_1 and y_2, which are

$$y_1^{\star}, y_2^{\star}, y^{\star}. \tag{6.16}$$

A point is connected to a line whenever it is connected to *any* point of the line. By the condition (6.15) $x_{\star}$, $(y_1)_{\star}$, $(y_2)_{\star}$, $y_{\star}$ form a "four point" which can be imbedded into a projective geometry, because (6.15) avoids any possibility of having nontrivial point spaces.

This projective geometry is therefore isomorphic to the Moufang plane and the restriction of the "measure" (6.3) on it is additive. We can draw a picture for the condition (6.15):



(6.17)

$x_{\star}$ is either incident or not connected to any line forming the triangle.

In the language of propositions, denoting

$$x_{\star} = a, \ (y_i)_{\star} = b_i, \ i = 1,2, \tag{6.18}$$

we have

$$(y_i)^{\star} = b_i^{\perp}, \quad y_{\star} = b_1^{\perp} \wedge b_2^{\perp}, \quad y^{\star} = b_1 \vee b_2 \tag{6.19}$$

(since $b_1 \perp b_2$, $b_1 \vee b_2$ exists).

It follows that the condition (6.15) is translated in the language of quantum logic by saying that the "measure" (6.3) is additive whenever the associated propositions a, b_1 and b_2 generate a sublattice of the proposition system. Once again the existence of connected points, or equivalently of point spaces (of dimension greater than 1) which is a consequence of having a nondivision algebra of coefficients, affects the definition of quantum objects, causing, in this case, this particular behavior of the "measure" *naturally* defined in our system. We notice that also this unusual feature has an analog in the standard theory for a *reducible lattice*. In that case the "join" of two points in two different coherent components of the lattice is not a line and the additivity of the states cannot be phrased as in the irreducible case.

But what is more striking and interesting for our "measure" is that it becomes the usual measure in a three-dimensional Hilbert space when we restrict x and y to complex matrices. In fact as an immediate consequence of the results of Appendix A, any rank 1 element in J, composed just of complex numbers, can be written in the form

$$x = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix}(\lambda_1 \ \lambda_2 \ \lambda_3) \equiv |\lambda\rangle\langle\lambda|. \tag{6.20}$$

Notice that we do not have complex conjugation in passing from $|\lambda\rangle$ to $\langle\lambda|$; therefore, $|\lambda\rangle\langle\lambda|$ is *not* the usual projector in a Hilbert space which in the present notation is $|\lambda^*\rangle\langle\lambda|$. Using x and y in the form (6.20) and because of associativity we can write

$$\begin{aligned} |\mathrm{tr}\,(x^{+}, y^{-})| &= |\mathrm{tr}(|\lambda\rangle\langle\lambda|\mu^*\rangle\langle\mu^*|)| \\ &= (\langle\lambda|\mu^*\rangle^2\langle\lambda^*|\mu\rangle^2)^{1/2} = |\langle\lambda|\mu^*\rangle|^2. \end{aligned} \tag{6.21}$$

We also have that the restriction of $\mathscr{C}_{6,0}$ to purely complex matrices gives SU(3). To see this we consider the generic transformation of $\mathscr{C}_{6,0}$:

$$x \to (A, x, B) + i\,C{\cdot}x. \tag{6.22}$$

The only way we can obtain purely complex matrices is to take A, B, and C real. Then A, x, and B associate under the ordinary matrix product and the associator in (6.22) can be written as a commutator:

$$(A, x, B) = \tfrac{1}{4}[[A, B], x]. \tag{6.23}$$

Denoting $D = \frac{1}{4}[A, B]$ we can rewrite (6.22) as

$$x \to Dx - xD + (i/2)Cx + (i/2)xC \tag{6.24}$$

when D is a real antisymmetric traceless matrix and C is a real symmetric traceless matrix. Therefore (6.24) becomes

$$(D + (i/2)C)x + x(D + (i/2)C)^{\dagger}. \tag{6.25}$$

The matrix $D + (i/2)C$ is a 3×3 skew-adjoint traceless matrix and (6.25) is identified with the usual action of SU(3), once it is taken into account that we do not conjugate the complex numbers passing to the "dual space." By taking the tensor product of SU(3) with U(1) we get the group U(3). This shows that the system we have been describing, having $\mathscr{C}_{6,0} \otimes$ U(1) as automorphism group, reduces to the usual theory, in dimension 3, of a complex Hilbert space having as auto-

morphism group the unitary group.

It is well known that $\mathscr{E}_{6,0}$ has a maximal rank subgroup

$$\mathrm{SU}(3)\otimes\mathrm{SU}(3)\otimes\mathrm{SU}(3). \tag{6.26}$$

The only one of these SU(3) groups involving the imaginary unit i is the one we have here described. The other two are obtained by singling out an octonionic imaginary unit and are subgroups of $\mathscr{F}_{4,0}$, one of which is in $\mathscr{G}_{2,0}$.

We conclude this section with three remarks.

Remark 1: Notice the difference between the measure we have taken for our example in Sec. 3 and the present one. This difference (the appearance of a square root in the latter) is linked to the fact that in Sec. 3 we were pairing a Jordan triple system; here we are pairing a Jordan algebra.

Remark 2: Notice that we can have $\alpha_x(y)=0$ although $x_\star$ is not orthogonal to $y_\star$ (for instance taking $x^+=E_1$ and $y^+=a[1,2]$). This can only occur when $x_\star\cong y_\star$.

Remark 3: The "measure" we have defined, when restricted to the points [Eq. (1) of (6.3)] can be written as

$$\alpha_x(y)=\mathrm{Max}[\tfrac{1}{2}\,\mathrm{tr}\,(x^+y^-+y^+x^-)], \tag{6.27}$$

where the maximum is taken over the phase of y (or of x).

Suppose, indeed, that $\mathrm{tr}(x^+,y^-)=\lambda$; then varying the phase of y we get

$$\tfrac{1}{2}\,\mathrm{tr}(x^+y^-e^{-i\theta}+x^-y^+e^{i\theta})=\tfrac{1}{2}(e^{i\theta}\lambda+e^{i\theta}\lambda^*), \tag{6.28}$$

and the maximum will be reached when θ is such that $e^{i\theta}\lambda$ is real and therefore equal to its modulus.

It has been claimed that[7,13]

$$\tfrac{1}{2}\,\mathrm{tr}(x^+y^-+x^-y^+) \tag{6.29}$$

[That is, (6.27) *without the "maximum" condition*] is a positive definite measure. We point out that because of (6.28), the expression (6.29) is not uniquely determined on the points (because of the arbitrary phase) and, furthermore,

$$-1\leqslant\tfrac{1}{2}\,\mathrm{tr}(x^+y^-+x^-y^+)\leqslant 1. \tag{6.30}$$

We shall have further comments on this topic in Appendix D.

6.3. The observables

We can associate observables to the generators of the automorphism group $\mathscr{E}_{6,0}\otimes\mathrm{U}(1)$ exactly in the same way we do in the usual quantum theory, namely, by multiplying the skew-Hermitian generators by the imaginary unit i to obtain Hermitian operators. Consider, in fact, Eq. (4.3); denote the generator G_+ by K and multiply K by i. Then

$$\mathrm{tr}((iK)x),x^*)=-\mathrm{tr}(x,i(Kx)^*)=\mathrm{tr}(x,((iK)x)^*) \tag{6.31}$$

which shows that the operator $iK\equiv H$ is Hermitian with respect to $\mathrm{tr}(x,x^*)$ and that $\mathrm{tr}(Hx,x^*)$ is real.

In perfect analogy with the example of Sec. 3 (and with the usual quantum theory of the Hilbert spaces), $\mathrm{tr}(Hx,x^*)$ defines the expectation value of the observable H in the state associated to $x_\star$.

If we reduce the theory to the complex case, H becomes a self-adjoint operator and, for x in the form (6.20), we get

$$\mathrm{tr}(Hx,x^*)=\langle\lambda^*H\lambda\rangle. \tag{6.32}$$

The observables are also directly related to the pair via the three-grading $[L_1,L_{-1}]\subset L_0$. The elements of $\mathscr{E}_6\otimes\mathbb{C}$ are all generated by the set of transformations $V_{x^+,y}$, but out of these transformations only the skew-Hermitian ones are generators of $\mathscr{E}_{6,0}\otimes\mathrm{U}(1)$ and to these are related the observables, via a multiplication by i.

Notice, however, that the spectral theory of the observables thus defined is completely different from the usual one. The spectral family of an observable in this theory can have at most three points. They can always be transformed into e_1,e_2,e_3, where $e_i=(E_i,E_i)$, $i=1,2,3$, which generate a maximal Boolean sublattice. Therefore we can *simultaneously measure* only three eigenvalues of H, although H has many more independent parameters in it (the symmetry group being of rank $6+1$). This unusual behavior is already present in the real $\mathscr{M}_3^8$ case.[35] Other problems affecting the real $\mathscr{M}_3^8$ case,[35] however, are not present in this Jordan pair theory.

The Hamiltonian of the system will be one of the Hermitian generators of $\mathscr{E}_{6,0}\otimes\mathrm{U}(1)$. It will involve six independent frequencies and will be shift invariant because of the presence of the group U(1) of the "overall phase." Time reversal can be consistently defined because of the presence of the imaginary unit i. For a more detailed discussion of these topics we refer to Ref. 35.

ACKNOWLEDGMENTS

We wish to acknowledge our indebtedness to Professor John Faulkner (University of Virginia, Charlottesville) for his generous help. In particular, Faulkner is responsible for the counter example to complete additivity [discussed in Eqs. (6.11)–(6.14)], and for the proper formulation of the additivity restriction, Eq. (6.15), and discussion following. We wish to thank him for his many discussions and help.

APPENDIX A

We prove in this appendix the following two propositions:

Proposition A1: Any trace 1 idempotent of J can be written in the form

$$\begin{pmatrix}a\\b\\c\end{pmatrix}(\bar a\;\bar b\;\bar c)\,, \tag{A1}$$

where a, b, c are complex octonions, one of which is a complex number, and which satisfy

$$a\bar a+b\bar b+c\bar c=1\,. \tag{A2}$$

Proposition A2: Any nilpotent of J with nonzero elements on the diagonal can be written in the form (A1), where a, b, c are complex octonions, one of which is a complex number, and which satisfy

$$a\bar a+b\bar b+c\bar c=0\,. \tag{A3}$$

Using the associativity of the subalgebra generated by any two octonions (Artin theorem, Ref. 24) it is immediate to verify that an element of J which can be written in the form (A1) satisfying (A2) [resp. (A3)] is a trace 1 idempotent (resp. nilpotent).

We now give the proofs of the two propositions.

Proof of Proposition A1: Consider an element of J in the generic form

$$\begin{pmatrix} \alpha & a & \bar{b} \\ \bar{a} & \beta & c \\ b & \bar{c} & \gamma \end{pmatrix}. \tag{A4}$$

Then, in order to be a trace 1 idempotent, it must satisfy

(1) $\alpha^2 + a\bar{a} + b\bar{b} = \alpha$,
(2) $\beta^2 + a\bar{a} + c\bar{c} = \beta$,
(3) $\gamma^2 + b\bar{b} + c\bar{c} = \gamma$,
(4) $\bar{b}\bar{c} = \gamma a$,
(5) $ac = \beta\bar{b}$,
(6) $\bar{a}\bar{b} = \alpha c$,
(7) $\alpha + \beta + \gamma = 1$.

We distinguish various cases:

(A) Suppose $n(a) = n(b) = n(c) = 0$. Then we get $\alpha^2 = \alpha, \beta^2 = \beta, \gamma^2 = \gamma$. Thus either α or β or γ must be equal to 1. Because of (7), just one of them can be 1 and the others vanish. Take $\alpha = 1$ and $\beta = \gamma = 0$. Therefore, from (4), (5), and (6), $\bar{b}\bar{c} = 0 = ac$ and $\bar{a}\bar{b} = c$. Hence,

$$\begin{pmatrix} 1 \\ \bar{a} \\ b \end{pmatrix}(1\ a\ \bar{b}) = \begin{pmatrix} 1 & a & \bar{b} \\ \bar{a} & 0 & c \\ b & \bar{c} & 0 \end{pmatrix}$$

satisfies the required conditions.

(B) Suppose now that $n(a) \neq 0$ and $n(b) = n(c) = 0$. From (1), (2), and (3) we get

$$\alpha^2 - \beta^2 = \alpha - \beta \text{ and } \gamma = 0, 1. \tag{A5}$$

If $\gamma = 1$, then $\alpha + \beta = 0$ and, from (A5), $\alpha = \beta = 0$ which implies $a\bar{a} = 0$, contrary to the hypothesis.

If $\gamma = 0$ then $\alpha + \beta = 1$ and from (1), (2), and (7) we obtain $\alpha\beta = a\bar{a}$ (then $\alpha \neq 0$, $\beta \neq 0$) and from (4) $\bar{b}\bar{c} = 0$. Then

$$\begin{pmatrix} \sqrt{\alpha} \\ \dfrac{\bar{a}}{\sqrt{\alpha}} \\ \dfrac{b}{\sqrt{\alpha}} \end{pmatrix}\left(\sqrt{\alpha}\ \ \frac{a}{\sqrt{\alpha}}\ \ \frac{\bar{b}}{\sqrt{\alpha}}\right) = \begin{pmatrix} \alpha & a & \bar{b} \\ \bar{a} & \beta & c \\ b & \bar{c} & 0 \end{pmatrix}$$

satisfies the required conditions.

Notice that if $\alpha = 1$, condition (7) implies $n(a) = n(b) = 0$ and we are back in the first case.

(C) Suppose now that only $n(c) = 0$. If $ac \neq 0$, multiplying (4) by c on the right we get $\gamma = 0$ and, from (3), $b\bar{b} = 0$, contrary to the hypothesis.

If $ac = 0$ then $\alpha + \gamma = 1$ (being $b \neq 0$ by hypothesis). Hence, $\beta = 0$ and from (2) $a\bar{a} = 0$, contradicting again the hypothesis. Therefore, case (C) does not occur.

(D) Finally suppose $n(a) \neq 0$, $n(b) \neq 0$, $n(c) \neq 0$. From (7) it follows that either α or β or γ must differ from 0. Take $\alpha \neq 0$, then it is easy to see that

$$\begin{pmatrix} \sqrt{\alpha} \\ \dfrac{\bar{a}}{\sqrt{\alpha}} \\ \dfrac{b}{\sqrt{\alpha}} \end{pmatrix}\left(\sqrt{\alpha}\ \ \frac{a}{\sqrt{\alpha}}\ \ \frac{\bar{b}}{\sqrt{\alpha}}\right)$$

satisfies the required conditions (1)–(7). This ends the proof of Proposition A1. □

Proof of Proposition A2: For an element in the generic form (A4) to be a nilpotent of J, it must satisfy the following conditions:

(1) $\alpha^2 + a\bar{a} + b\bar{b} = 0$,
(2) $\beta^2 + a\bar{a} + c\bar{c} = 0$,
(3) $\gamma^2 + b\bar{b} + c\bar{c} = 0$,
(4) $\bar{b}\bar{c} = \gamma a$,
(5) $ac = \beta\bar{b}$,
(6) $\bar{a}\bar{b} = \alpha c$,
(7) $\alpha + \beta + \gamma = 0$.

We have furthermore assumed that there are nonzero elements on the diagonal. Suppose $\alpha \neq 0$. Dividing (1) by α we get

$$a\bar{a}/\alpha + b\bar{b}/\alpha = -\alpha = \beta + \gamma.$$

Multiplying (4) by b from the left we have

$$(b\bar{b}/\alpha)\bar{c} = \gamma\bar{c}.$$

If $c \neq 0$ we get $b\bar{b}/\alpha = \gamma$ and then $\beta = a\bar{a}/\alpha$; we can therefore write

$$\begin{pmatrix} \sqrt{\alpha} \\ \dfrac{\bar{a}}{\sqrt{\alpha}} \\ \dfrac{b}{\sqrt{\alpha}} \end{pmatrix}\left(\sqrt{\alpha}\ \ \frac{a}{\sqrt{\alpha}}\ \ \frac{\bar{b}}{\sqrt{\alpha}}\right) = \begin{pmatrix} \alpha & a & \bar{b} \\ \bar{a} & \beta & c \\ b & \bar{c} & \gamma \end{pmatrix}. \tag{A6}$$

If $c = 0$ then from (4) and (5) either β or γ must be equal to 0, otherwise $a = b = c = 0$, and would not have a nilpotent. If $\gamma = 0$ then $\beta = -\alpha$ and from (5), being $c = 0$, it follows that $b = 0$. It is easy to check that we can write (A6) in this case, just putting $b = 0$. Analogously, for $\beta = 0$, we get $a = 0$ in (A6). This ends the proof of Proposition A2. □

Remark A1: This way of writing the trace 1 idempotents and the nilpotents with nonzero elements on the diagonal shows immediately that the projective dimensionality of *any* rank 1 element is 16 (complex) (since any rank 1 element satisfies, indeed, the same conditions). It also shows that a rank 1 element behaves like a singlet and a spinor under SO(10). This fits with the known split

$$27 = 1 + 10 + 16$$

of J under SO(10) (Ref. 14).

If we take, for example, the element

$$\begin{pmatrix} \sqrt{\alpha} \\ \bar{a}/\sqrt{\alpha} \\ b/\sqrt{\alpha} \end{pmatrix}\ \ \left(\sqrt{\alpha}\ \ \frac{a}{\sqrt{\alpha}}\ \ \frac{\bar{b}}{\sqrt{\alpha}}\right),$$

we can identify the spinor with $(a,\bar{b})$, with spinor norm[36]

$$a\bar{a} + b\bar{b} = \sum_{\alpha=0}^{7} a_\alpha a_\alpha + b_\alpha b_\alpha,$$

and the singlet with E_1.

Extending these considerations to the pair, we get that SO(10) is a subgroup of $E_{6,0}$ leaving a point invariant. In the example just shown, this is the point $(e_1)_\star$, where $e_1 = (E_1, E_1)$.

APPENDIX B

We give the proof of Proposition 4.4. We shall exploit some results from the theory of the real $\mathcal{M}_3^8$, for which we refer to Ref. 8. In particular, we use the following facts:

(1) $\mathcal{F}_{4,0}$ is generated by the set of transformations $x \to (A, x, B)$, when A and B are traceless matrices in the (real) $\mathcal{M}_3^8$. Any element in the (real) $\mathcal{M}_3^8$ can be diagonalized by an $\mathcal{F}_{4,0}$ transformation.

(2) The maximal subgroup of $\mathcal{F}_{4,0}$ leaving a trace 1 idempotent of $\mathcal{M}_3^8$ invariant is SO(9).

(3) The maximal subgroup of $\mathcal{F}_{4,0}$ leaving the diagonal elements of any matrix of $\mathcal{M}_3^8$ invariant is SO(8). The action of SO(8) on the off-diagonal elements is

$$\mathrm{SO}(8): \begin{pmatrix} \alpha & a & \bar{b} \\ \bar{a} & \beta & c \\ b & \bar{c} & \gamma \end{pmatrix} \to \begin{pmatrix} \alpha & d_3 a & \overline{d_2 b} \\ \overline{d_3 a} & \beta & d_1 c \\ d_2 b & \overline{d_1 c} & \gamma \end{pmatrix},$$

where d_3, d_2, d_1 are related by the principle of triality

$$(d_1 a)(d_2 b) = \overline{d_3\,(\bar{a}\bar{b})}\,; \tag{B1}$$

given d_1, the SO(8) actions d_2 and d_3 are determined uniquely (up to a sign).

The following SO(8) actions obey the principle of triality

$$d_1 = R_{\bar{a}}, \quad d_2 = L_{\bar{a}}, \quad d_3 = L_a R_a\,, \tag{B2}$$

where R_a (L_a) indicates the right (left) multiplication by a unit real octonion a (i.e., $a\bar{a} = 1$).

We want to show first that any trace 1 idempotent in J can be brought into the form E_1 (and therefore into any other trace 1 idempotent).

Let us consider a trace 1 idempotent u in its most general form:

$$u = \begin{pmatrix} \alpha & a & \bar{b} \\ \bar{a} & \beta & c \\ b & \bar{c} & \gamma \end{pmatrix}. \tag{B3}$$

We can separate the real and imaginary parts of (B3) and then diagonalize the imaginary part via an $\mathcal{F}_{4,0}$ transformation. We thus get

$$u' = \begin{pmatrix} \alpha_1' + i\alpha_2' & a' & \bar{b}' \\ \bar{a}' & \beta_1' + i\beta_2' & c' \\ b' & \bar{c}' & \gamma_1' + i\gamma_2' \end{pmatrix}, \tag{B4}$$

where $\alpha_1', \alpha_2', \beta_1', \beta_2', \gamma_1', \gamma_2'$ are real and a', b', c' are real octonions.

Having used an $\mathcal{F}_{4,0}$ transformation (B4) must still be a trace 1, rank 1 element in J; therefore, we can apply the result of the Appendix A and write

$$u' = \begin{pmatrix} \sqrt{\alpha'} \\ \bar{a}'/\sqrt{\alpha'} \\ b'/\sqrt{\alpha'} \end{pmatrix} (\sqrt{\alpha'} \quad a'/\sqrt{\alpha'} \quad \bar{b}'/\sqrt{\alpha'}), \tag{B5}$$

where we have supposed $\alpha' \neq 0$ (we could do the same for $\beta' \neq 0$ or $\gamma' \neq 0$).

But we have that $c = \bar{a}'\bar{b}'/\alpha'$ is a real octonion. This can be the case only if $\alpha' \in \mathbb{R}$ or $a' = b' = 0$ or $a' = 0$, $b' \neq 0$ or $a' \neq 0$, $b' = 0$. In the first case, we have a real trace 1 idempotent which can be brought in the form E_1 by an $\mathcal{F}_{4,0}$ transformation.[8] In the second case, we have already $u' = E_1$. The remaining two cases are equivalent. Let us take, for instance, $b' = 0$. Then

$$u' = \begin{pmatrix} \alpha' & a' & 0 \\ \bar{a}' & \dfrac{a'\bar{a}'}{\alpha'} & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{B6}$$

Writing $\alpha' = re^{i\phi}$, it follows that $a'\bar{a}'/\alpha' = se^{-i\phi}$ where $s = a'\bar{a}'/|\alpha'| \in \mathbb{R}$.

We now apply the transformation generated by

$$iC = i\begin{pmatrix} -\phi & 0 & 0 \\ 0 & \phi & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{B7}$$

which, according to (4.4) is definitely an $\mathcal{E}_{6,0}$ transformation. In order to exponentiate (B7) we must consider it as a 27×27 matrix acting on the vector $(\alpha, a_\alpha, b_\alpha, c_\alpha, \beta, \gamma)$ where $a_\alpha, b_\alpha, c_\alpha$ are the components of a, b, c. It is then easy to see that $\alpha \to e^{-i\phi}\alpha$, $a \to a$, $b \to e^{-i\phi/2}b$, $c \to e^{i\phi/2}c$, $\beta \to e^{i\phi}\beta$, $\gamma \to \gamma$. The transformation (B7) is the generator of the $\tilde{\mathrm{U}}(1)$ group leaving E_3 fixed. Under such a transformation u' in the form (B6) becomes real (octonionic) and can therefore be brought into E_1 by an $\mathcal{F}_{4,0}$ transformation.

Let us consider now a nilpotent y in J. We know from Remark 3.1 of Sec. 4 that there is a trace 1 idempotent u of J such that $u \in V_0^+(y, y^*)$. As we have just proved, we can bring u into (a multiple) of E_3 so that y goes into

$$y' = \begin{pmatrix} \alpha & a & 0 \\ \bar{a} & \beta & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{B8}$$

If y' is a multiple of an idempotent then we can bring it into the form E_1. If it is a nilpotent, then $\beta = -\alpha$. We then split (B8) into its real and imaginary part:

$$y' = \begin{pmatrix} \alpha_1 & a_1 & 0 \\ \bar{a}_1 & -\alpha_1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + i\begin{pmatrix} \alpha_2 & a_2 & 0 \\ \bar{a}_2 & -\alpha_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{B9}$$

and use next the SO(8) transformation $d_3 = Ra_2/|a_2|$ bringing the second matrix in (B9) into a real matrix that can be diagonalized by a real rotation. Notice that in this way we have not affected E_3. We shall exploit this transformation in a moment. We then use SO(8) again to bring the remaining matrix into the form

$$y'' = \begin{pmatrix} \lambda & r & 0 \\ r & -\lambda & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{B10}$$

when $\lambda \in \mathbb{C}$ $r \in \mathbb{R}$.

Since y'' is a nilpotent we must have

$\lambda^2 + r^2 = 0$, that is, $\lambda = \pm ir$ and

$$y'' = \lambda\begin{pmatrix} 1 & \pm i & 0 \\ \pm i & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{B11}$$

We can now use the $\tilde{\mathrm{U}}(1)$ transformation (B7) with $\phi = \mp\pi/2$ and bring therefore y'' into

$$y''' = \pm i\lambda\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{B12}$$

which can be brought into a multiple of E_1 by an $\mathcal{F}_{4,0}$

transformation. □

This proves that $\mathscr{E}_{6,0}$ is transitive on the rank 1 elements (up to a scalar factor). It follows then from Proposition 4.2 and the considerations following it that $\mathscr{E}_{6,0}$ is transitive on points. To show that $\mathscr{E}_{6,0}$ is transitive on triples of mutually orthogonal points is now very easy.

Suppose u_1, u_2, u_3 are the primitive normalized idempotents generating the mutually orthogonal points $(u_1)_\star, (u_2)_\star$, $(u_3)_\star$. We can map u_3 into $e_3 \equiv (E_3, E_3)$ and then work with SO(8), $\tilde{\mathrm{U}}(1)$, and with real rotations in the 2×2 "block" $V_0^+(e_3)$, as we did in the final part of the previous proof, so as to leave e_3 invariant and bring u_1 into E_1. Once we do this it necessarily follows that $u_2 \to E_2$ and we have completed the proof. □

We finally prove that $\mathrm{SO}(10)\otimes\tilde{\mathrm{U}}(1)$ is the maximal subgroup of $\mathscr{E}_{6,0}$ leaving a point fixed. Combining this with the previous result it follows that the plane we are considering is the homogeneous space $\mathscr{E}_{6,0}/\mathrm{SO}(10)\otimes\tilde{\mathrm{U}}(1)$. We have noted already that SO(10) and $\tilde{\mathrm{U}}(1)$ leave a point invariant. We thus have only to prove that it is maximal. Consider the Lie algebra of $E_{6,0}$ in the Tits' form

$$E_{6,0} = iJ_0^{\mathbb{R}} \oplus \mathrm{Der}(J^{\mathbb{R}}),$$

where $J_0^{\mathbb{R}}$ is the set of traceless real octonionic matrices in J and $\mathrm{Der}(J^{\mathbb{R}})$ is the set of transformations $x \to (A,x,B)$, with A and B traceless real octonionic matrices. We know that the maximal subgroup generated by a Lie algebra in $\mathrm{Der}(J^{\mathbb{R}})$ (which is isomorphic to $F_{4,0}$), leaving a point invariant, is SO(9), with 36 generators.

The maximal number of parameters we can get from $iJ_0^{\mathbb{R}}$ in order to leave, say $(e_3)_\star$, invariant are the ten parameters of the transformation

$$iC = i\begin{pmatrix} \alpha & a & 0 \\ \bar{a} & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}, \quad \alpha+\beta+\gamma = 0,$$

which changes e_3 just by a phase, thus leaving the point $(e_3)_\star$ unchanged.

We have therefore proven that the maximal subgroup of $\mathscr{E}_{6,0}$ leaving a point invariant has 46 generators, which is the number of the generators of $\mathrm{SO}(10)\otimes\tilde{\mathrm{U}}(1)$. Therefore, these two groups coincide and the proof is complete. □

APPENDIX C

We want to compare the geometry we have defined in Sec. 4 to the Hjelmslev–Moufang plane defined in Refs. 25 and 30. This is defined on an exceptional Jordan algebra J over a split Cayley algebra. We shall, of course, make the comparison with the case in which the split Cayley algebra is the complex octonion algebra $\mathscr{C}$. The definition of the Hjelmslev–Moufang plane is the following.[25] For any rank 1 element x of J let $x_\star$ and $x^\star$ be two copies of the set $\{\alpha x | \alpha \in \mathbb{C}/\{0\}\}$. The plane is then defined as the set of points $x_\star$, and lines $x^\star$, and the following relations:

$$x_\star \,|\, y^\star \quad \text{if} V_{x,y} = 0,$$
$$x_\star \cong y_\star (x^\star \cong y^\star \quad \text{if } x\times y = 0,$$
$$x_\star \cong y^\star \quad \text{if } \mathrm{tr}(x,y) = 0.$$

In view of Proposition 4.1 we can compare these relations with ours by restricting our pairs, just to the V^+ part. We have the following results, whose proof we shall give at the end of this appendix ($\star$ will be associated to the H–M plane, $\stackrel{\star}{\star}$ to ours).

Proposition C 1: Let $x = (x^+, x^-)$ and $y = (y^+, y^-)$ be two primitive normalized idempotents. Then $V_2(x) \subset V_0(y)$ (i.e., $x_{\stackrel{\star}{\star}} | y^{\stackrel{\star}{\star}}$) if and only if $V_{x^+,y^-} = 0$.

Proposition C 2: $x_{\stackrel{\star}{\star}} \cong y_{\stackrel{\star}{\star}} (x^{\stackrel{\star}{\star}} \cong y^{\stackrel{\star}{\star}})$ iff $x^+ \times y^+ = 0$.

Proposition C 3: $x_{\stackrel{\star}{\star}} \cong y^{\stackrel{\star}{\star}}$ iff $\mathrm{tr}(x^+, y^-) = 0$.

It is clear from these results that there is a very close relationship between our plane $\mathscr{P}$ and the Hjelmslev–Moufang plane $\mathscr{P}'$ thus defined. This relationship is given by associating points of $\mathscr{P}'$ to points of $\mathscr{P}$ by

$$x_\star = V_2^+(x)$$

and lines of $\mathscr{P}'$ to lines of $\mathscr{P}$ by

$$x^\star = V_0^+(x^{\mathrm{op}}),$$

where $x^{\mathrm{op}} = (x^-, x^+)$. The objects, points, and lines (as well as the point spaces) are, therefore, essentially the same; hence, we can reproduce the following results:

(1) If $x_{\stackrel{\star}{\star}} \not\cong y_{\stackrel{\star}{\star}}$ there is a unique line incident to both of them.

(2) if $x_{\stackrel{\star}{\star}} \cong y_{\stackrel{\star}{\star}}$ there is at least one line intersecting both of them.

However, a big difference shows up when we consider transformations on points and lines. We have much more structure to preserve, namely, the pairing of a rank 1 element with its complex conjugate. This is reflected in the preservation of the standard polarity $\pi: x_{\stackrel{\star}{\star}} \to x^{\stackrel{\star}{\star}}$ we have chosen, and which is preserved by the group mapping points into points. In other words we can say that the geometry we have defined is a Hjelmslev–Moufang plane, carrying a further structure to be preserved: the standard (elliptic) polarity π.

We now give the proof of the preceding propositions.

Proof of Proposition C 1: It follows from Proposition 4.1 that $V_2(x) \subset V_0(y)$ if and only if $x^+ \in V_0^+(y)$. If $x^+ \in V_0^+(y)$ then, from Proposition 3.6, $V_{x^+,y^-} = 0$. Vice versa, suppose that $V_{x^+,y^-} = 0$. It is easy to see that this implies $V_{y^-,x^+} = 0$ and, therefore, $V_{y^-,x^+}\, y^- = 0$. That is $U_{y^-}\, x^+ = 0$ and $\mathrm{tr}(x^+, y^-) = 0$. Now let us calculate

$$\begin{aligned} &V_{x^+,y}\; y^+ \\ &= \mathrm{tr}(x^+ + y^+, y^-)(x^+ + y^+) - (x^+ \times y^+)\times y^- - y^+ \\ &= x^+ - (x^+ \times y^+)\times y^- = 0. \end{aligned}$$

Therefore

$$x^+ = (x^+ \times y^+)\times y^-,$$

that is $x^+ \in V_0^+(y)$ and the proposition is proven. □

Proof of Proposition C 2: We have

$$x_{\stackrel{\star}{\star}} \cong y_{\stackrel{\star}{\star}} (x^{\stackrel{\star}{\star}} \cong y^{\stackrel{\star}{\star}}) \quad \text{iff} \quad (x^+ \times y^+)\times y = 0.$$

Take $g_+ \in \mathscr{E}_{6,0}$ such that $g_+(y^+) = E_1$. It follows from Proposition 4.3 that if $x^+ \in V_2^+(y) \oplus V_1^+(y)$ then $g_+(x^+) \in V_2^+(g(y)) \oplus V_1^+(g(y))$, where $g = (g_+, g_-)$. Therefore, $(g_+(x^+)\times E_1)\times E_1 = 0$. Denote $d = g_+(x^+)\times E_1$.

Notice that $d^\# = 0$ [from (1.20)] and $d\cdot E_1 = 0$ by Proposition 3.3. Hence,

$$0 = d \times E_1 = -E_1 \operatorname{tr}(d) - d + I \operatorname{tr}(d);$$

that is,

$$d = \operatorname{tr}(d)(E_2 + E_3).$$

But $E_2 + E_3$ is not of rank 1; therefore, $\operatorname{tr}(d) = 0$ and $d = 0$. That is,

$$(g_+(x^+) \times E_1) \times E_1 = 0 \Rightarrow g_+(x^+) \times E_1 = 0.$$

Acting now with $(g_+)^{-1}$ and using the fact that the Freudenthal product is $\mathscr{E}_{6,0}$ convariant, we get

$$(x^+ \times y^+) \times y^- = 0 \Rightarrow x^+ \times y^+ = 0.$$

The vice versa is obvious and the proposition is proven. □

Proof of Proposition C 3: We have that $V_2(x) \subset V_0(y) \oplus V_1(y)$ implies that the component of x on $V_2(y) = 0$; that is, $\operatorname{tr}(x^+, y^-) = 0$ and vice versa. □

APPENDIX D

A possible quantum mechanics on a complex $\mathscr{M}_3^8$ has been already investigated by Gürsey,[13] from a different point of view. Although some results are similar, the survey of the structure in Ref. 13 is different from ours, being oriented towards the physical consequences and interpretations of the $\mathscr{E}_{6,0}$ symmetry rather than the rigorous examination of the quantum mechanical formalism. Because of the physical relevance of Gürsey's paper, it seems to us worth while to clarify some contradictory points in its mathematical formulation.

In Ref. 13 the states are related to the projectors $U_{x^\pm} U_{x^\pm}, (x^\pm)^\# = 0$, which are in our Jordan pair language the Peirce projectors on the $V_2^+(x)$ space. They seem therefore to be the same as ours. However, the distance function between x and y is defined as

$$\alpha_x(y) = \tfrac{1}{2} \operatorname{tr}(x^+ y^- + y^+ x^-)$$

which, as we have seen, is not positive definite and does not actually give a well-determined value on any given point. Any point, indeed, is defined only up to a phase.

As a consequence the definition of orthogonality in Ref. 13,

$$x \perp y \quad \text{if} \quad \alpha_x(y) = 0$$

is not well posed.

Another weak point in Ref. 13 is the definition of superposition. A point x_3 is the superposition of x_1 and x_2 if

$$\operatorname{tr}(x_3, x_1 \times x_2) = 0. \tag{D1}$$

We argue that if x_1 and x_2 are connected, $x_1 \times x_2 = 0$ and (D1) holds for any x_3. The concept of superposition is not well defined by the condition (D1).

Finally, we notice that there is no discussion in Ref. 13 about the group U(1) of the "overall phase" which is definitely a symmetry group of the system.

The paper by Gürsey is physically ingenious and path-breaking; we are indebted to this work for providing questions and techniques which have been incorporated in our work here.

[1] P. Jordan, Nachr. Ges. Wiss. Göttingen, 569–585 (1932).
[2] P. A. M. Dirac, Invited Lecture at the Southeastern Section of the American Physical Society, Columbia, SC, Nov. 1971.
[3] P. Jordan, Nachr. Ges. Wiss. Göttingen, 209 (1933).
[4] P. Jordan, J. von Neumann, and E. P. Wigner, Ann. Math. **35**, 29–64 (1934).
[5] J. R. Faulkner and J. C. Ferrar, Bull. London Math. Soc. **9**, 1–35 (1977).
[6] F. Gürsey, *International Symposium on Mathematical Problems in Theoretical Physics*, edited by A. Araki (Springer, New York, 1975), cf. pp. 189ff.
[7] L. C. Biedenharn and L. P. Horwitz, Invited paper at the Conference on Differential-Geometric Methods in Physics, 13–15 July 1978, Clausthal-Zellerfeld, West Germany.
[8] M. Günaydin, C. Piron, and M. Ruegg, Commun. Math. Phys. **61**, 69–85 (1978).
[9] K. McCrimmon, Proc. Natl. Acad. Sci. USA **56**, 1072–1079 (1966).
[10] N. Jacobson, *Structure and Representations of Jordan Algebras*, Amer. Math. Soc. Coll. Publ. (American Math. Soc., Providence, RI, 1968).
[11] M. Koecher, Amer. J. Math. **89**, 787–816 (1967).
[12] O. Loos, "Jordan pairs," *Lecture Notes in Mathematics* **460** (Springer, Berlin, Heidelberg, and New York, 1975).
[13] F. Gürsey, *Second Johns Hopkins Workshop on "Current Problems in High Energy Particle Theory,"* edited by G. Damokos and S. Kövesi-Domokos (Johns Hopkins Univ., Baltimore, MD, 1978), p. 3–15.
[14] F. Gürsey, *Group Theoretical Methods in Physics*, edited by W. Beiglaböck, A. Böhm, and E. Takasugi, *Lecture Notes in Physics* **94** (Springer, New York, 1978), pp. 508–521.
[15] T. A. Springer, Indag. Math. **24**, 259–265, 451–468 (1962).
[16] K. McCrimmon, Trans. Am. Math. Soc. **139**, 495–510 (1969).
[17] N. Jacobson, *Lectures on Quadratic Jordan Algebras*, Lecture Notes (Tata Institute, Bombay, 1969).
[18] K. McCrimmon, Trans. Am. Math. Soc. **153**, 265–278, 279–305 (1971).
[19] K. McCrimmon, Bull. Am. Math. Soc. **84**, 612,627,4 (1977).
[20] J. R. Faulkner, J. Algebra **26**, 1–9 (1973).
[21] M. Koecher, Am. J. Math. **90**, 476–510 (1968).
[22] M. Koecher, *An elementary approach to bounded symmetric domains* (Rice University, Houston, 1969).
[23] H. Freudenthal, Advan. Math. **1**, 145–190 (1964).
[24] R. D. Schafer, *An Introduction to Nonassociative Algebras* (Academic, New York and London, 1966).
[25] J. R. Faulkner, Mem. of Amer. Math. Soc. **104**, 1–71 (1970).
[26] T. A. Springer, F. D. Veldkamp, Indag. Math. **25**, 413–451 (1963).
[27] K. McCrimmon, Trans. Am. Math. Soc. **159**, 445–468 (1971).
[28] N. Jacobson, "Exceptional Lie Algebras," in *Lecture Notes in Pure and Applied Mathematics* **1** (Dekker, New York, 1971).
[29] J. Tits, "Tabellen zu den einfachen Lie Gruppen und ihren Darstellungen," *Lecture Notes in Mathematics* (Springer, Berlin, Heidelberg, and New York, 1967).
[30] T. A. Springer, Math. Z. **107**, 249–263 (1968).
[31] F. Van der Blij and T. A. Springer, Nieuw Arch. Wisk. (3) **8**, 158–169 (1960).
[32] J. M. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley, Reading, MA, 1968).
[33] V. S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand, Princeton, 1968), Vol. 1.
[34] J. R. Faulkner, private communication.
[35] P. Truini and L. C. Biedenharn, "A comment on the dynamics of M_3^8" in *Proceedings of the III Workshop on Lie Admissible Formulations* (Hadronic, Boston, 1981).
[36] E. Cartan, *The theory of spinors* (M.I.T., Cambridge, MA, 1966).

Linear invariants of a time-dependent quantal oscillator

R. Vaidyanathan
Reactor Research Centre, Kalpakkam-603 102, India

(Received 9 December 1980; accepted for publication 3 April 1981)

Explicitly time-dependent invariants, linear in x and p, are found to simplify the solution of the Schrödinger equation with oscillator-type Hamiltonian. As an example, exact dynamics is obtained of the time-dependent quantal oscillator, with damping and subject to external force. The invariant and hence the solution of the Schrödinger equation involve only the amplitude of the classical damped oscillator. The identity of our solution with a recently obtained result is illustrated. A digressive remark on quadratic invariants and the Dirac operator formalism is added.

PACS numbers: 03.65.Ge

I. INTRODUCTION

Exact dynamics of the time-dependent quantal harmonic oscillator (one with time-dependent frequency parameter) has recently been obtained, using the theory of explicitly time-dependent invariants.[1,2] Owing to the simple relation between the eigenstates of the invariant and the solution of the corresponding Schrödinger equation, there have also been attempts[3] to obtain general invariants, particularly for oscillator-type Hamiltonians. In the above works, the invariant is constructed as a quadratic function of x and p, originally by inspection of classical trajectories.[1] The invariant involves a single function $\rho(t)$, related to the amplitude of the classical free oscillator; $\rho(t)$ satisfies a nonlinear differential equation, which should be solved in practical applications. The spectrum and eigenstates of the invariant are obtained using an operator formalism analogous to Dirac's.[4]

We note that the solution can be accomplished, somewhat more directly, using invariants linear in x and p. In the first place, the linear invariant is readily diagonalized and involves only the amplitude $b(t)$ of the corresponding classical oscillator. The spectrum of the invariant is continuous, but with a suitable choice of the initial phase and initial conditions on $b(t)$, the evolution of an arbitrary state can be conveniently obtained through the usual Fourier expansion. It is hoped that the use of linear invariants will simplify problems with more general quadratic Hamiltonians.

In the following section, we consider, as an example, the Hamiltonian appropriate to a damped oscillator, subject to external force. The invariant is constructed as a linear function of x and p, its eigenfunctions readily found and the Schrödinger wave function compared with the recent result.[2] In Sec. 3, a remark on quadratic invariants and the Dirac operator formalism is made, which is of related interest.

II. THE INVARIANT AND SOLUTION OF THE SCHRÖDINGER EQUATION

We consider the Hamiltonian[2]

$$H(t) = e^{-rt}p^2/2 + e^{rt}\{\omega^2(t)x^2/2 + xf(t)\}, \tag{2.1}$$

where r is the damping constant, ω the time-dependent frequency parameter, $f(t)$ the external force, and the mass has been set equal to unity. We wish to construct a Hermitian invariant of the form

$$I(t) = a(t)x + b(t)p + c(t). \tag{2.2}$$

Requiring that I obeys the equation

$$\partial I/\partial t + (1/i\hbar)[I,H] = 0, \tag{2.3}$$

we are led to the conditions

$$\begin{aligned} \dot{a} &= be^{rt}\omega^2(t), \\ \dot{b} &= -ae^{-rt}, \\ \dot{c} &= be^{rt}f(t). \end{aligned} \tag{2.4}$$

The first two of these equations combine to read

$$\ddot{b} + r\dot{b} + \omega^2(t)b = 0, \tag{2.5}$$

which describes the amplitude of the classical damped oscillator. Any particular, real solution of this equation gives an invariant

$$I(t) = b(t)p - \dot{b}e^{rt}x + c(t), \tag{2.6}$$

where

$$c(t) = \int_t dt'\, b(t')e^{rt'}f(t').$$

Any real number λ is an eigenvalue of $I(t)$ and the corresponding eigenfunction

$$\psi_\lambda(x,t) = \left(\frac{1}{2\pi\hbar b}\right)^{1/2} \exp\left(\frac{i}{b\hbar}\{\dot{b}e^{rt}x^2/2 + (\lambda - c)x\}\right); \tag{2.7}$$

we have

$$\langle \psi_{\lambda'}(t)|\psi_\lambda(t)\rangle = \delta(\lambda - \lambda'). \tag{2.8}$$

The theory of explicitly time-dependent invariants[1] has shown that the eigenfunctions of the invariant, rephased by a time-dependent phase factor $\alpha_\lambda(t)$, satisfy the Schrödinger equation, where $\alpha_\lambda(t)$ are determined from

$$\hbar\frac{d\alpha_\lambda}{dt} = \left\langle \psi_\lambda(t)\left|i\hbar\frac{\partial}{\partial t} - H\right|\psi_\lambda(t)\right\rangle. \tag{2.9}$$

The matrix elements on the right hand side can be found from the explicit form of ψ_λ in Eq. (2.7). We get

$$\frac{d\alpha_\lambda}{dt} = -\frac{[\lambda - c(t)]^2}{2\hbar b^2}e^{-rt}. \tag{2.10}$$

 0022-2488/82/071346-03$02.50

The following rephased eigenstates of the invariant satisfy the Schrödinger equation:

$$\psi(\lambda;x,t)$$
$$\equiv\psi_\lambda(x,t)e^{i\alpha_\lambda(t)}$$
$$=(1/2\pi\hbar b)^{1/2}\times\exp\{i[\alpha_\lambda(t_0)+A(t)\lambda^2-B(t)\lambda-C(t)]\}$$
$$\times\exp(i/b\hbar)\{\dot{b}e^{rt}x^2/2+(\lambda-c)x\}), \tag{2.11}$$

where the first exponent represents the integral of Eq. (2.10) and $\alpha_\lambda(t_0)$ is an arbitrary initial phase. Note $A(t_0)=B(t_0)=C(t_0)=0$, t_0 being the initial time.

When the amplitude $b(t)$ vanishes, say at time t_1, the eigenstates of the invariant (2.6) reduce to

$$\psi_\lambda(x,t_1)=\frac{e^{-rt_1/2}}{\sqrt{\dot{b}(t_1)}}\delta\left\{x-\frac{(c(t_1)-\lambda)e^{-rt_1}}{\dot{b}(t_1)}\right\}, \tag{2.12}$$

which satisfy Eq. (2.8). This is also the limiting form of Eq. (2.7) as $b(t)\to 0$. The divergence occuring in Eq. (2.10) is resolved as follows: $\dot{\alpha}_\lambda$ diverges[5] as $(t-t_1)^{-2}$ and hence $f_\lambda\equiv\alpha_\lambda(t)b(t)$ is finite at all times. Invoking this in Eq. (2.10), we have

$$\dot{bf}_\lambda-\dot{b}f_\lambda=-[(\lambda-c(t))^2/2\hbar]e^{-rt}. \tag{2.13}$$

The first term is easily seen to vanish at $t=t_1$, leading to

$$f_\lambda(t_1)=\frac{(\lambda-c(t_1))^2}{2\hbar\dot{b}(t_1)}e^{-rt_1}. \tag{2.14}$$

The rephased eigenstates are

$$\psi(\lambda;x,t)$$
$$=\frac{1}{\sqrt{2\pi\hbar b}}\exp\left\{\frac{i}{b\hbar}[\dot{b}e^{rt}x^2/2+(\lambda-c)x+f_\lambda(t)\hbar]\right\}. \tag{2.15}$$

When $b(t)$ vanishes, the eigenstates can be cast in the form of the delta distribution as in Eq. (2.12). In practical applications, Eqs. (2.13)–(2.15) will be operative in place of Eqs. (2.10) and (2.11), for all times.

The evolution of a general Schrödinger state is described by a linear combination of the solutions in Eq. (2.15):

$$\psi(x,t)=\int_{-\infty}^{\infty}d\lambda\varphi_\lambda\psi(\lambda;x,t), \tag{2.16}$$

where

$$\phi_\lambda=\langle\psi_\lambda(t_0)|\psi_\lambda(t_0)\rangle,$$

$|\psi(t_0)\rangle$ being the initial state. As a convenient choice, let $f_\lambda(t_0)=\dot{b}(t_0)=c(t_0)=0$ and $b(t_0)=1$,

whence from Eq. (2.15), we see

$$\varphi_\lambda=\frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty}dx\,\exp(-i\lambda x/\hbar)\psi(x,t_0), \tag{2.17}$$

which are the Fourier components of the initial state.

Since choice of the invariant does not influence the solution obtained,[1] we can illustrate the identity of the above solution with the one recently reported[2] using quadratic invariants. Consider, for the initial state, a simple case

$$\psi(x,t_0)=Ke^{-irx^2/4}. \tag{2.18}$$

The mixing coefficients are then

$$\varphi_\lambda=\frac{K}{\sqrt{2\pi}}\left(\frac{4\pi}{r}\right)^{1/2}\exp\left(\frac{i\lambda^2}{r}\right),$$

where $\hbar$ is set equal to 1. The solution of the Schrödinger equation (2.11) reduces to

$$\psi(x,t)=K\left(\frac{i}{b(1+rA)}\right)^{1/2}\exp\left(i\left\{\frac{R}{2}-c(t)-\frac{rB^2}{4(1+rA)}\right\}\right),$$

where

$$R=x^2\left(\frac{\dot{b}e^{rt}}{b}-\frac{r}{2b^2(1+rA)}\right)+\frac{x}{b}\left(\frac{rB}{1+rA}-2C\right).$$

This is in complete agreement with the time-evolving Gaussian form given in Ref. 2.

III. REMARK ON QUADRATIC INVARIANTS

It is clear that one can construct a quadratic invariant as a product of two linear ones, corresponding to the linearly independent solutions of Eq. (2.5). Let I_1 and I_2 be the invariants in the form of Eq. (2.6), corresponding to solutions b_1 and b_2 of Eq. (2.5); their commutator

$$K_1\equiv[I_1,I_2]=i\hbar\{b_1(t)\dot{b}_2(t)-b_2(t)\dot{b}_1(t)\}e^{rt}, \tag{3.1}$$

is independent of time. This property of Eq. (2.5) is also required, since the commutator is (trivially) an invariant. Scaling down I_1 and I_2 by $(K_1)^{1/2}$ we have $[I_1,I_2]=1$. One can choose $I_2I_1+\frac{1}{2}$ as the required quadratic invariant, I_1 and I_2 being the annihilation and "creation" operators. Note, however, that one should not limit b_1 and b_2 of Eq. (2.5) to real solutions to obtain a Hermitian quadratic invariant. As the usual prescription, one can demand $I_2=I_1^+$, which is accomplished with the choice $b_1(t)=b_2^*(t)$. We now have the quadratic invariant in the standard form, for which the Dirac operator formalism applies.[1,2]

For the free, undamped oscillator [$r=0, f(t)=0$], $C(t)$ can be chosen to be zero [Eq. (2.4)], when we get a homogeneous quadratic invariant by the above procedure. In the more general case we considered, the quadratic invariant would be an inhomogeneous function of x and p. These are the forms used in Refs. 1 and 2, respectively, albeit starting directly with the quadratic form.

Below we present a digressive note on the Dirac method[4] of obtaining the spectrum and eigenstates of the Hamiltonian

$$H_0=(p^2+\omega_0^2x^2)/2. \tag{3.2}$$

Our aim is to show that the results can be deduced from considerations of the time-dependent invariants of the problem. The two linear invariants of this simple example, corresponding to the independent solutions of the classical equation [Eq. (2.5) with $r=0$, $\omega(t)=\omega_0$], are

$$I_1=(p-i\omega_0x)e^{i\omega_0t},$$
$$I_2=(p+i\omega_0x)e^{-i\omega_0t}, \tag{3.3}$$

which are mutually adjoint and proportional to the conventional annihilation and creation operators, respectively.

Let $|\psi\rangle$ be an arbitrary Schrödinger state, expanded in the complete orthornormal set of stationary states:

$$|\psi\rangle=\sum_n c_n|n\rangle e^{-iE_nt/\hbar}. \tag{3.4}$$

The energy eigenvalues E_n and stationary states $|n\rangle$ will be deduced from the invariant operators. The expectation value of the invariant in the above state should be independent of time, for arbitrary C_n. The expectation value of I_1 is

$$\sum_{nm} c_m^* c_n \langle m| -i\hbar\frac{\partial}{\partial x} - i\omega_0 x|n\rangle \times \exp\left\{i\left[\frac{E_m - E_n}{\hbar} + \omega_0\right]t\right\}. \tag{3.5}$$

This implies that either

$$\left\langle m\right| -i\hbar\frac{\partial}{\partial x} - i\omega_0 x\left|n\right\rangle = 0 \tag{3.6}$$

or

$$E_m - E_n + \hbar\omega_0 = 0. \tag{3.7}$$

Since $\omega_0 > 0$, Eq. (3.7) cannot be fulfilled if $E_m \geqslant E_n$. Therefore,

$$\left\langle m\right| i\hbar\frac{\partial}{\partial x} + i\omega_0 x\left|n\right\rangle = 0 \text{ if } E_m \geqslant E_n. \tag{3.8}$$

We note that the spectrum is bounded from below (H_0 being positive definite) and is discrete (due to bounded motion of the particle). If $|0\rangle$ is the ground state, by Eq. (3.8)

$$\left\langle m\right| i\hbar\frac{\partial}{\partial x} + i\omega_0 x\left|0\right\rangle = 0 \text{ for all } m. \tag{3.9}$$

Since the stationary states constitute a complete set, Eq. (3.9) implies

$$\left(\hbar\frac{\partial}{\partial x} + \omega_0 x\right)\psi_0(x) = 0, \tag{3.10}$$

which determines the ground state. If $|1\rangle$ be the first excited state, Eq. (3.8) requires

$$\left\langle m\right| i\hbar\frac{\partial}{\partial x} + i\omega_0 x\left|1\right\rangle = 0 \text{ if } m \neq 0. \tag{3.11}$$

One possibility is that $(i\hbar(\partial/\partial x) + i\omega_0 x)|1\rangle = 0$. However, the ground wavefunction is uniquely determined by the same equation (3.10). Hence we must have

$$\left\langle 0\right| i\hbar\frac{\partial}{\partial x} + i\omega_0 x\left|1\right\rangle \neq 0, \tag{3.12}$$

whence Eq. (3.7) gives

$$E_1 = E_0 + \hbar\omega. \tag{3.13}$$

Equation (3.11) and (3.12) imply

$$\left(i\hbar\frac{\partial}{\partial x} + i\omega_0 x\right)\psi_1(x) = c\psi_0(x), \tag{3.14}$$

where c is determined by a suitable phase convention and normalization. The adjoint form of Eqs. (3.11) and (3.14) leads to the useful equation

$$\left(i\hbar\frac{\partial}{\partial x} - i\omega_0 x\right)\psi_0(x) = c^*\psi_1(x), \tag{3.15}$$

which gets us the first excited state. The procedure is continued for higher states similarly.

The above results are readily deducible, starting from any invariant not proportional to the creation or annihilation operators. For example, one might consider the Hermitian invariant

$$I_3 = p\cos\omega t + x\omega\sin\omega t. \tag{3.16}$$

Constancy of the expectation value $\langle\psi|I_3|\psi\rangle$ in the arbitrary state in Eq. (3.4) again leads to Eqs. (3.6) and (3.7), on requiring that $(d/dt)\langle\psi|I_3|\psi\rangle$ vanish identically. This completes our remark on quadratic invariants and the Dirac procedure.

IV. CONCLUSIONS

Solution of the Schrödinger equation for oscillator type systems can be obtained rather directly using linear invariants. We consider the case of a damped oscillator, acted on by an external force and illustrated the identity of our solution with a recent result. A way of contructing quadratic invariants in the standard form, starting from linear ones, is noted. As a point of related interest, the spectrum and eigenstates of the ordinary oscillator Hamiltonian are deduced from consideration of time-dependent invariants.

[1] H. R. Lewis and W. B. Reisenfeld, J. Math. Phys. **10**, 1458 (1969).
[2] D. C. Khandekar and S. V. Lawande, J. Math. Phys. **20**, 1870 (1979).
[3] P. G. L. Leach, J. Math. Phys. **18**, 1608, 1902 (1977).
[4] P. A. M. Dirac, *Principles of Quantum Mechanics* (Clarendon, Oxford, 1958), 4th ed.

[5] This follows from the regularity of $b(t)$. If $\dot{b}(t)$ vanishes too, there is no future classical motion.

$\mathcal{HH}$ spaces with an algebraically degenerate right side

J. F. Plebański[a] and G. F. Torres del Castillo
Departamento de Física, Centro de Investigación y de Estudios Avanzados del I. P. N., Apdo. Postal 14-740, México 14, D. F., Mexico

(Received 19 August 1981; accepted for publication 29 January 1982)

We find the most general form of the key function for an $\mathcal{HH}$ space whose curvature on the right side is algebraically degenerate.

PACS numbers: 04.20.Cv, 02.40.Ky

I. INTRODUCTION

$\mathcal{HH}$ spaces are the complexified Riemannian structures which fulfill the Einstein equations and have algebraically degenerate conformal curvature from one side, e.g., the self-dual part (the "left side"). The structure of such spaces is determined by a potential function $\overline{W}$, the key function, which must satisfy a certain nonlinear differential equation.

All the algebraically degenerate real solutions of the Einstein equations correspond to a real slice of an $\mathcal{HH}$ space whose curvature is of the same algebraic type on both sides. Also, as Rózga[1] has shown, a necessary condition for the existence of a real slice with signature $(+\ +\ +\ -)$ of a complex solution is that its algebraic type be the same from both sides. Therefore, employing the theory of $\mathcal{HH}$ spaces in constructing algebraically degenerate real solutions, it is of particular interest to determine the $\mathcal{HH}$ spaces which also have the right side algebraically degenerate.

This paper constitutes an introductory step towards this end, by determining the most general form of the key function which makes the right conformal curvature algebraically degenerate.

In Sec. II we give a brief description of the results and formalism for $\mathcal{HH}$ spaces and deduce the basic equations. In Secs. III and IV we find the form of the key function for all possible special algebraic types on the right side.

II. THE h FUNCTION

As was shown in Ref. 2, and in a more detailed form in Ref. 3, the algebraic degeneracy of the conformal curvature of one side, together with Einstein's vacuum equations, imply the existence of a congruence of two-dimensional surfaces which are totally null ("null string"). In fact, if α_A is a multiple Debever–Penrose spinor for the left conformal curvature, i.e., $C_{ABCD} = \alpha_{(A}\alpha_B\beta_C\gamma_{D)}$, where C_{ABCD} are the spinorial components of the self-dual part of the conformal curvature, then the system of differential equations[4]

$$\alpha_A g^{AB} = 0 \tag{2.1}$$

defines a null string.

Associated with such a null string there exist[5] canonical coordinates q^A and p^A such that the metric has the form

$$ds^2 = -\tfrac{1}{2} g_{AB} \underset{s}{\otimes} g^{AB}, \tag{2.2}$$

with

$$\begin{aligned} g^{2A} &= -\sqrt{2}\phi^{-2} dq^A \\ g^{1A} &= -\sqrt{2}(dp^A - Q^{AB} dq_B)\,. \end{aligned} \tag{2.3}$$

The null string is then defined by $dq^A = 0$. Einstein's vacuum equations imply that $\phi = J_A p^A + k$, where J_A and k are functions of q^A only, which can be made constant by using the freedom in the choice of the canonical coordinates. Then, if K_A is another constant spinor such that $\tau \equiv K^A J_A$ does not vanish

$$Q^{AB} = -(\phi^4(\phi^{-3}\overline{W})^{,(B})^{,A)} + (\mu/\tau^2)\phi^3 K^A K^B, \tag{2.4}$$

where μ is a function of q^A only. The key function $\overline{W}$ must satisfy

$$\begin{aligned} &\tfrac{1}{2}\phi^4(\phi^{-2}\overline{W}^{,B})^{,A}(\phi^{-2}\overline{W}_{,B})_{,A} - \phi^{-1}\left(\frac{\partial \overline{W}}{\partial q}\right)^{,A} \\ &\quad - \mu\phi^4\partial_\phi\phi^{-1}\partial_\phi\phi^{-1}\overline{W} \\ &\quad + \frac{\eta}{2\tau^2}[\eta J^C - (\phi + k)K^C]\frac{\partial \mu}{\partial q^C} = N_A p^A + \gamma \end{aligned} \tag{2.5}$$

with the abbreviations $\eta \equiv K^A p_A$, $_{,A} = \partial/\partial p^A$, $\partial_\phi = (1/\tau)K^A \partial/\partial p^A$, and where N_A and γ are functions of q^A only.

The anti-self-dual part of the conformal curvature in the tetrad (2.3) is given by

$$C_{\dot A\dot B\dot C\dot D} = \phi^3(\overline{W} - (\mu/4\tau^2)\phi^2\eta^2)_{,\dot A\dot B\dot C\dot D}\,. \tag{2.6}$$

Therefore, if the conformal curvature is also right-degenerate, the key function must be such that

$$\left(\overline{W} - \frac{\mu}{4\tau^2}\phi^2\eta^2\right)_{,\dot A\dot B\dot C\dot D} = \alpha_{(\dot A}\alpha_{\dot B}\beta_{\dot C}\gamma_{\dot D)}\,. \tag{2.7}$$

This relation restricts the dependence of $\overline{W}$ on p^A only. Thus, at this stage, we can ignore the dependence on q^A, which has to be determined through the hyperheavenly equation (2.5).

Since we are interested in the p^A dependence only, and since in two dimensions each vector is proportional to a gradient, we can write

$$\tilde{W}_{,\dot A\dot B\dot C\dot D} = h_{,(\dot A}h_{,\dot B}\beta_{\dot C}\gamma_{\dot D)}\,, \tag{2.8}$$

where $\tilde{W} \equiv \overline{W} - (\mu/4\tau^2)\phi^2\eta^2$ and h is a function such that $h_{,\dot A} \neq 0$. The integrability conditions for (2.8) are $\tilde{W}_{,\dot A\dot B\dot C\dot D}{}^{,\dot D} = 0$; therefore, contracting with $h^{,\dot A}$, we have

$$0 = h^{,\dot A}\tilde{W}_{,\dot A\dot B\dot C\dot D}{}^{,\dot D} = (h^{,\dot A}\tilde{W}_{,\dot A\dot B\dot C\dot D})^{,\dot D} - h^{,\dot A\dot D}\tilde{W}_{,\dot A\dot B\dot C\dot D} \tag{2.9}$$

[a]On leave of absence from the University of Warsaw, Warsaw, Poland.

 0022-2488/82/071349-04$02.50

or

$$\tfrac{1}{4}[(h^{\cdot A}\beta_{\dot A})h_{(\dot B}h_{\cdot\dot C}\gamma_{\dot D)} + (h^{\cdot A}\gamma_{\dot A})h_{(\dot B}h_{\cdot\dot C}\beta_{\dot D)}]^{\cdot D} - h^{\cdot A\dot D}h_{(\dot A}h_{\cdot\dot B}\beta_{\dot C}\gamma_{\dot D)} = 0 . \tag{2.10}$$

If both spinors $\beta_{\dot A}$ and $\gamma_{\dot A}$ are proportional to $h_{\cdot\dot A}$, Eq. (2.10) implies

$$h_{\cdot\dot A\dot B}h^{\cdot\dot A}h^{\cdot\dot B} = 0 . \tag{2.11}$$

If either one of $\beta_{\dot A}$ or $\gamma_{\dot A}$ is proportional to $h_{\cdot\dot A}$, say $\beta_{\dot A}$, absorbing the proportionality factor in $\gamma_{\dot A}$ and contracting Eq. (2.10) with $h^{\cdot\dot B}$, the condition (2.11) is obtained. Finally, if none of the spinors $\beta_{\dot A}$ or $\gamma_{\dot A}$ is proportional to $h_{\cdot\dot A}$, contracting eq. (2.10) with $h^{\cdot\dot B}h^{\cdot\dot C}$, one obtains, once again, Eq. (2.11).

Since the anti-self-dual part of the conformal curvature is algebraically degenerate, there exists another null string defined by

$$h_{\cdot\dot B}\, g^{A\dot B} = 0 , \tag{2.12}$$

with $g^{A\dot B}$ given in (2.3). The two null strings intersect each other along a congruence of complex null geodesics which have a tangent vector defined by

$$V \equiv h^{\cdot A}\frac{\partial}{\partial p^A} . \tag{2.13}$$

If follows that $v[h] = 0$; therefore, the congruence of geodesics is given by $q^A = \text{const.}$, $h = \text{const.}$

The twist of this congruence is determined by the 3-form

$$T \equiv e \wedge de , \tag{2.14}$$

where $e = H_{\cdot A}, dq^A$. Thus

$$T = -h_{\cdot\dot A\dot B}h^{\cdot\dot A}dp^{\dot B}\wedge dq^1\wedge dq^2 . \tag{2.15}$$

We shall now integrate Eq. (2.11). First we write it in the equivalent form

$$h_{\cdot\dot A\dot B}h^{\cdot\dot A} = \lambda h_{\cdot\dot B} , \tag{2.16}$$

where λ is some function; multiplying this equation by $dp^{\dot B}$, one has

$$-h_{\cdot 2}\, dh_{\cdot 1} + h_{\cdot 1}dh_{\cdot 2} = \lambda\, dh . \tag{2.17}$$

By assuming $H_{\cdot 2} \neq 0$, Eq. (2.17) can be written as

$$d(h_{\cdot 1}/h_{\cdot 2}) = -[\lambda/(h_{\cdot 2})^2]\, dh \tag{2.18}$$

which implies that $h_{\cdot 1}/h_{\cdot 2}$ is some function of h. Therefore, we can write

$$h_{\cdot 1}/h_{\cdot 2} = -a^2(h)/a^1(h) , \tag{2.19}$$

where $a^A(h)$ is a nonvanishing spinor.

By setting $a_{\dot A} = \dot b_{\dot A}$ (a dot denotes derivation with respect to h), Eq. (2.19) takes the form $b^{\cdot\dot A}{}_{\cdot\dot A} = 0$. This means that the 1-form $b_{\dot A}dp^{\dot A}$ is closed; therefore, there exists, locally, a function b such that $b_{\dot A}\, dp^{\dot A} = db$. Thus

$$d(b - b_{\dot A}\, p^{\dot A}) + \dot b_{\dot A}p^{\dot A}\, dh = 0 . \tag{2.20}$$

This implies that $b_0 \equiv b - b_{\dot A}\, p^{\dot A}$ is a function of h. Substituting b_0 in (2.20), we get

$$(\dot b_0 + \dot b_{\dot A}\, p^{\dot A})dh = 0 . \tag{2.21}$$

Remembering that $\dot b_{\dot A} = a_{\dot A}$ and denoting $\dot b_0$ as a_0, we have

$$A(h) \equiv a_{\dot A}(h)p^{\dot A} + a_0(h) = 0 . \tag{2.22}$$

Differentiating (2.22) with respect to $p^{\dot A}$, one has

$$\dot A(h)h_{\cdot\dot A} + a_{\dot A}(h) = 0 , \tag{2.23}$$

where

$$\dot A(h) \equiv \dot a_{\dot A}(h)p^{\dot A} + \dot a_0(h) \tag{2.24}$$

is assumed to be different from zero. Under this last condition Eq. (2.22) defines h as a function of $p^{\dot A}$ giving the general solution of (2.11). In fact, from (2.23), it follows that

$$h_{\cdot\dot A\dot B}h^{\cdot\dot A} = [(\ln\dot A)^{\cdot\dot A}h_{\cdot\dot A}]h_{\cdot\dot B} . \tag{2.25}$$

This solution has two branches: the general, G, where $a^{\dot A}\dot a_{\dot A} \neq 0$, and the special, S, where $a^{\dot A}\dot a_{\dot A} = 0$. Due to (2.15) and (2.23), branch G corresponds to $T \neq 0$ while branch S corresponds to $T = 0$. In the case S, there is a function $o(h)$ such that $\dot a_{\dot A} = o a_{\dot A}$, and therefore $a_{\dot A} = \nu(h)H_{\dot A}$, where $H_{\dot A}$ is a nonvanishing constant (i.e., $p^{\dot A}$-independent) spinor and $\nu(h) \neq 0$. Hence in branch S, the solution is given by

$$A(h) \equiv H_{\dot A}\, p^{\dot A} + a_0(h) = 0 , \tag{2.26}$$

provided that

$$\dot A(h) \equiv \dot a_0(h) \neq 0 . \tag{2.27}$$

III. THE CASES N AND III

In this section we consider the case where the right conformal curvature has, at least, a triple Debever–Penrose spinor; therefore

$$\tilde W_{\cdot\dot A\dot B\dot C\dot D} = h_{(\dot A}\, h_{\cdot\dot B}\, h_{\cdot\dot C}\beta_{\dot D)} , \tag{3.1}$$

where h is given by (2.22). By using Eq. (2.25), the integrability condition $\tilde W_{\cdot\dot A\dot B\dot C\dot D}{}^{\cdot\dot D} = 0$ reduces the equation

$$\beta_{\dot B}{}^{\cdot\dot B}h_{\cdot\dot A} + 3\beta_{\dot B}\, h^{\cdot\dot B}_{\cdot\dot A} + 6h^{\cdot\dot B}(\ln\dot A)_{\cdot\dot B}\,\beta_{\dot A} + 3h_{\cdot\dot B}\,\beta_{\dot A}{}^{\cdot\dot B} = 0 . \tag{3.2}$$

By using the identity

$$\beta_{\dot A}{}^{\cdot\dot B} = \delta_{\dot A}{}^{\dot B}\beta_{\dot R}{}^{\cdot\dot R} + \beta^{\dot B}{}_{\cdot\dot A} , \tag{3.3}$$

Eq. (3.2) takes the form

$$6[\beta_{\dot B}\, h^{\cdot\dot B}_{\cdot\dot A} + h^{\cdot\dot B}(\ln\dot A)_{\cdot\dot B}\,\beta_{\dot A}] + 4\beta_{\dot B}{}^{\cdot\dot B}h_{\cdot\dot A} - 3(h^{\cdot\dot B}\beta_{\dot B})_{\cdot\dot A} = 0 , \tag{3.4}$$

which, as a consequence of Eq. (2.23), is equivalent to

$$[-12\beta_{\dot B}(\ln\dot A)^{\cdot\dot B} + 6(\ddot A/\dot A)\beta_{\dot B}\, h^{\cdot\dot B} + 4\beta_{\dot B}{}^{\cdot\dot B}] \times h_{\cdot\dot A} - 3(\beta_{\dot B}h^{\cdot\dot B})_{\cdot\dot A} = 0 , \tag{3.5}$$

where

$$\ddot A \equiv \ddot a_{\dot A}\, p^{\dot A} + \ddot a_0 . \tag{3.6}$$

Therefore, $\psi \equiv \tfrac{1}{4}\beta_{\dot B}h^{\cdot\dot B}$ is a function of h. If $\psi = 0$, $\beta_{\dot A}$ is proportional to $h_{\cdot\dot A}$ and the right conformal curvature is of type N. In this case, by writing $\beta_{\dot A} = -r\dot A^3 h_{\cdot\dot A}$, Eq. (3.5) implies

$$r^{\dot B}h_{\cdot\dot B} = 0 , \tag{3.7}$$

i.e., r is a function of h.

Using Eq. (2.23), we have

$$\tilde W_{\cdot\dot A\dot B\dot C\dot D} = r(h)a_{\dot A}a_{\dot B}a_{\dot C}h_{\cdot\dot D} = \left[\int^h r(t)a_{\dot A}(t)a_{\dot B}(t)a_{\dot C}(t)\, dt\right]_{\cdot\dot D} . \tag{3.8}$$

Hence, if $A(t)$ is defined by

$$A(t) \equiv a_{\dot A}(t)p^{\dot A} + a_0(t) , \tag{3.9}$$

it follows that, since $A(h)=0$

$$\tilde{W}_{,\dot{A}\dot{B}\dot{C}\dot{D}}=\left\{(1/3!)\int^{h} r(t)[A(t)]^3\,dt\right\}_{,\dot{A}\dot{B}\dot{C}\dot{D}}. \tag{3.10}$$

Thus, if the right curvature is of type N,[6]

$$\tilde{W}=\frac{1}{3!}\int^{h} r(t)[A(t)]^3 dt+\mathscr{P}\,, \tag{3.11}$$

where

$$\mathscr{P}\equiv(1/3!)A_{\dot{A}\dot{B}\dot{C}}\,p^{\dot{A}}p^{\dot{B}}p^{\dot{C}}+(1/2!)B_{\dot{A}\dot{B}}p^{\dot{A}}p^{\dot{B}}+C_{\dot{A}}\,p^{\dot{A}}+\bar{D} \tag{3.12}$$

and $A_{\dot{A}\dot{B}\dot{C}}, B_{\dot{A}\dot{B}}, C_{\dot{A}}$, and $\bar{D}$ are functions of the $q^{\dot{A}}$ only.

If $\psi\neq 0$, the right conformal curvature is of type III. Within the branch G one can write ψ as

$$\psi=a^{\dot{A}}\dot{a}_{\dot{A}}s\,, \tag{3.13}$$

where s is a function of h. Recalling the definition of ψ, (3.13) implies that $\beta_{\dot{A}}$ must be of the form

$$\beta_{\dot{A}}=-4s\dot{A}\dot{a}_{\dot{A}}+[s\ddot{A}-\dot{s}\dot{A}+r'\dot{A}^2]a_{\dot{A}}\,, \tag{3.14}$$

where r' is some function. Substituting (3.13) and (3.14) in (3.5), one concludes that r' is a function of h. By proceeding as in the previous case, it follows that (setting $r=r'$)

$$\tilde{W}=(1/2!)\int^{h} s(t)[A(t)]^2\,dt+(1/3!)\times\int^{h} r(t)[A(t)]^3\,dt+\mathscr{P}\,. \tag{3.15}$$

In branch S the function h is given by (2.26), and ψ can be represented as

$$\psi=-\dot{a}_0 s\,, \tag{3.16}$$

where $s=s(h)$. Introducing now a spinor $l_{\dot{A}}$ such that

$$H^{\dot{A}}l_{\dot{A}}=1\,, \tag{3.17}$$

one can express $\beta_{\dot{A}}$ as

$$\beta_{\dot{A}}=4s\dot{a}_0^2\,l_{\dot{A}}+[(s\ddot{a}_0-\dot{s}\dot{a}_0)l_{\dot{B}}\,p^{\dot{B}}+r''\dot{a}_0^2]H_{\dot{A}}\,. \tag{3.18}$$

Equation (3.5) implies that r'' is a function of h. Dropping primes, one finds that, in this case, $\tilde{W}$ must be of the form

$$\tilde{W}=(1/2!)(l_{\dot{B}}\,p^{\dot{B}})\int^{h} s(t)[H_{\dot{A}}\,p^{\dot{A}}+a_0(t)]^2dt+(1/3!)\int^{h} r(t)[H_{\dot{A}}\,p^{\dot{A}}+a_0(t)]^3dt+\mathscr{P}\,. \tag{3.19}$$

IV. CASES D AND II

We shall now assume the right conformal curvature to be of type D or II. Therefore,

$$\tilde{W}_{,\dot{A}\dot{B}\dot{C}\dot{D}}=h_{,(\dot{A}}h_{,\dot{B}}\beta_{\dot{C}}\gamma_{\dot{D})}\,, \tag{4.1}$$

where h is given by (2.22), and $h^{\cdot R}\beta_{\dot{R}}$ and $h^{\cdot R}\gamma_{\dot{R}}$ do not vanish. Using Eqs. (2.25) and (2.23) and the identity (3.3) for the derivatives of $\beta_{\dot{A}}$ and $\gamma_{\dot{A}}$ in the integrability condition of (4.1), and contracting then with $h^{\cdot\dot{B}}h^{\cdot\dot{C}}$, one has

$$h^{\cdot\dot{B}}h^{\cdot\dot{D}}\{(h^{\cdot\dot{C}}\gamma_{\dot{C}})[2\beta_{\dot{D},\dot{B}}+(\ln\dot{A})_{,\dot{B}}\beta_{\dot{D}}]+(h^{\cdot\dot{C}}\beta_{\dot{C}})[2\gamma_{\dot{D},\dot{B}}+(\ln\dot{A})_{,\dot{B}}\gamma_{\dot{D}}]\}=0 \tag{4.2}$$

This last equation implies that $X\equiv\frac{1}{12}\dot{A}^3h^{\cdot\dot{C}}h^{\cdot\dot{D}}\beta_{\dot{C}}\gamma_{\dot{D}}$ is a function of h. Within branch G the function X can be written as

$$X=-(a^{\dot{A}}\dot{a}_{\dot{A}})^2\,u\,, \tag{4.3}$$

where u is a function of h. This means that $\beta_{(\dot{C}}\gamma_{\dot{D})}$ is of the form

$$\beta_{(\dot{C}}\gamma_{\dot{D})}=-12\frac{u}{\dot{A}}\dot{a}_{\dot{C}}\dot{a}_{\dot{D}}-4\frac{u}{\dot{A}}a_{(\dot{C}}\ddot{a}_{\dot{D})}+4\left[3\frac{\ddot{A}u}{\dot{A}^2}-2\frac{\dot{u}}{\dot{A}}+v\right]a_{(\dot{C}}\dot{a}_{\dot{D})}+\left[3\frac{\dot{u}\ddot{A}}{\dot{A}^2}-3\frac{(\ddot{A})^2u}{\dot{A}^3}+\frac{\dddot{A}u}{\dot{A}^2}-\frac{\ddot{u}}{\dot{A}}+w\right]a_{\dot{C}}a_{\dot{D}}\,, \tag{4.4}$$

where v and w are some functions and $\dddot{A}=\dddot{a}_{\dot{R}}\,p^{\dot{R}}+\dddot{a}_0$.

Substituting (4.4) in (4.1), we have, after integration of the terms with u, $\dot{u}$ and $\ddot{u}$,

$$\tilde{W}_{,\dot{A}\dot{B}\dot{C}\dot{D}}=\left[\int^{h} u(t)A(t)\,dt\right]_{,\dot{A}\dot{B}\dot{C}\dot{D}}+h_{,(\dot{A}}h_{,\dot{B}}h_{,\dot{C}}\delta_{\dot{D})}\,, \tag{4.5}$$

where

$$\delta_{\dot{D}}=-\dot{A}(4v\dot{a}_{\dot{D}}+wa_{\dot{D}})\,. \tag{4.6}$$

Thus

$$\left[\tilde{W}-\int^{h} u(t)A(t)\,dt\right]_{,\dot{A}\dot{B}\dot{C}\dot{D}}=h_{,(\dot{A}}h_{,\dot{B}}h_{,\dot{C}}\delta_{\dot{D})}\,, \tag{4.7}$$

but this is precisely the problem solved in Sec. III; therefore, we conclude that

$$\tilde{W}=\int^{h} u(t)A(t)\,dt+(1/2!)\int^{h} s(t)[A(t)]^2\,dt+(1/3!)\int^{h} r(t)[A(t)]^3\,dt+\mathscr{P} \tag{4.8}$$

for some functions s and r.

In branch S, writing X as

$$X=-\dot{a}_0^2\,u\,, \tag{4.9}$$

where $u=u(h)$, one finds that $\beta_{(\dot{C}}\gamma_{\dot{D})}$ must be of the form

$$\beta_{(\dot{C}}\gamma_{\dot{D})}=-12u\dot{a}_0l_{\dot{C}}l_{\dot{D}}+[8\dot{a}_0(l_{\dot{B}}\,p^{\dot{B}})(u/\dot{a}_0)^{\cdot}+v]H_{(\dot{C}}l_{\dot{D})}+[-\dot{a}_0(l_{\dot{B}}\,p^{\dot{B}})^2(1/\dot{a}_0(u/\dot{a}_0)^{\cdot})^{\cdot}+w]H_{\dot{C}}H_{\dot{D}} \tag{4.10}$$

for some functions v and w. It follows that

$$\left\{\tilde{W}-(l_{\dot{B}}\,p^{\dot{B}})^2\int^{h} u(t)[H_{\dot{A}}\,p^{\dot{A}}+a_0(t)]\,dt\right\}_{,\dot{A}\dot{B}\dot{C}\dot{D}}=h_{,(\dot{A}}h_{,\dot{B}}h_{,\dot{C}}\delta'_{\dot{D})}\,, \tag{4.11}$$

where

$$\delta'_{\dot{D}}=-\dot{a}_0(vl_{\dot{D}}+wH_{\dot{D}})\,; \tag{4.12}$$

therefore

$$\tilde{W}=(l_{\dot{B}}\,p^{\dot{B}})^2\int^{h} u(t)[H_{\dot{A}}\,p^{\dot{A}}+a_0(t)]dt+(1/2!)(l_{\dot{B}}\,p^{\dot{B}})\int^{h} s(t)[H_{\dot{A}}\,p^{\dot{A}}+a_0(t)]^2dt+(1/3!)\int^{h} r(t)[H_{\dot{A}}\,p^{\dot{A}}+a_0(t)]^3dt+\mathscr{P}\,. \tag{4.13}$$

Expressing r, s, and u as

$$r(t) = \frac{d}{dt}\left(-\frac{1}{\dot{a}_0}\frac{d}{dt}\right)^3 R(t)$$
$$s(t) = \frac{d}{dt}\left(-\frac{1}{\dot{a}_0}\frac{d}{dt}\right)^2 S(t) \tag{4.14}$$
$$u(t) = \frac{d}{dt}\left(-\frac{1}{\dot{a}_0}\frac{d}{dt}\right) U(t)$$

and integrating by parts (4.13), one obtains

$$\tilde{W} = U(h)(l_{\dot{B}}\, p^{\dot{B}})^2 + S(h)(l_{\dot{B}}\, p^{\dot{B}}) + R(h) + \mathscr{P}\ . \tag{4.15}$$

V. CONCLUSIONS

The general expressions for $\tilde{W}$ given in (4.8) and (4.13) amount to the solution of a rather involved system of partial differential equations of fourth order, in two variables. This system is obtained from (2.6) by requiring that the Penrose polynomial

$$P(\zeta^{\dot{2}}/\zeta^{\dot{1}}) = (1/\zeta^{\dot{1}})^4\, C_{A\dot{B}\dot{C}\dot{D}}\, \zeta^{\dot{A}}\, \zeta^{\dot{B}}\, \zeta^{\dot{C}}\, \zeta^{\dot{D}} \tag{5.1}$$

has multiple roots. For example, if the Penrose polynomial (5.1) has a triple root (i.e., the right conformal curvature is of type III), $\tilde{W}$ must satisfy the system of equations

$$\begin{vmatrix} \tilde{W}_{,\dot{1}\dot{1}\dot{2}\dot{2}} & \tilde{W}_{,\dot{1}\dot{2}\dot{2}\dot{2}} & \tilde{W}_{,\dot{1}\dot{1}\dot{1}\dot{2}} \\ \tilde{W}_{,\dot{1}\dot{1}\dot{1}\dot{2}} & \tilde{W}_{,\dot{1}\dot{1}\dot{2}\dot{2}} & \tilde{W}_{,\dot{1}\dot{1}\dot{1}\dot{1}} \\ \tilde{W}_{,\dot{1}\dot{2}\dot{2}\dot{2}} & \tilde{W}_{,\dot{2}\dot{2}\dot{2}\dot{2}} & \tilde{W}_{,\dot{1}\dot{1}\dot{2}\dot{2}} \end{vmatrix} = 0 \tag{5.2}$$

and

$$\begin{vmatrix} \tilde{W}_{,\dot{1}\dot{1}\dot{2}\dot{2}} & \tilde{W}_{,\dot{1}\dot{1}\dot{1}\dot{1}} \\ \tilde{W}_{,\dot{2}\dot{2}\dot{2}\dot{2}} & \tilde{W}_{,\dot{1}\dot{1}\dot{2}\dot{2}} \end{vmatrix} = 4 \begin{vmatrix} \tilde{W}_{,\dot{1}\dot{1}\dot{2}\dot{2}} & \tilde{W}_{,\dot{1}\dot{1}\dot{1}\dot{2}} \\ \tilde{W}_{,\dot{1}\dot{2}\dot{2}\dot{2}} & \tilde{W}_{,\dot{1}\dot{1}\dot{2}\dot{2}} \end{vmatrix} \tag{5.3}$$

If the 2×2 determinants in (5.3) vanish, the right conformal curvature is of type N.

Our method has permitted us to solve these conditions determining the analytic dependence of $\tilde{W}$ on the variables $p^{\dot{A}}$—related to the affine parameters of the string—leaving open the dependence of $\tilde{W}$ on the variables q^R, which remains to be determined through the $\mathscr{H}\mathscr{H}$ equation (2.5). We have undertaken this task for various algebraic types on the right, and the results will be published elsewhere.[7]

ACKNOWLEDGMENT

One of us (G.F.T.dC.) acknowledges the Consejo Nacional de Ciencia y Tecnología for financial assistance through a doctoral fellowship.

[1] K. Rózga, Rep. Math. Phys. **11**, 197 (1977).
[2] J. F. Plebański and S. Hacyan, J. Math. Phys. **16**, 2403 (1975).
[3] C. P. Boyer, J. D. Finley, III, and J. F. Plebański, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980).
[4] The linear differential forms $g^{A\dot{B}}$ form a basis for the cotangent space; their scalar products are given by $g^{A\dot{B}}\cdot g^{C\dot{D}} = -2\epsilon^{AC}\epsilon^{\dot{B}\dot{D}}$.
[5] J. D. Finley, III, and J. F. Plebański, J. Math. Phys. **17**, 2207 (1976). See also J. D. Finley, III, and J. F. Plebański, J. Math. Phys. **18**, 1662 (1977).
[6] The expression of $\tilde{W}$, which makes the right curvature be of type N, was formerly obtained by J. D. Finley, III, and J. F. Plebański (private communication).
[7] In particular, the complete solution of (2.5) was obtained by Finley and Plebański when the right conformal curvature is zero. See J. D. Finley, III, and J. F. Plebański, J. Math. Phys. **22**, 667 (1981).

The investigation of some self-similar solutions of Einstein's equations

O. I. Bogoyavlensky[a] and G. Moschetti
Seminario Matematico, Università di Catania, Viale Andrea Doria 6 (95125) Catania, Italy

(Received 14 October 1981; accepted for publication 5 March 1982)

New application of self-similar generalizations of homogeneous cosmological models is shown in the problem of nonstationary nonspherically-symmetric accretion of self-gravitating gas on the center. For self-similar Bianchi types II and III, solutions of the system of Einstein equations are reduced to some dynamical system of small order. A number of exact solutions in empty space and for the stiff equation of state of matter is found.

PACS numbers: 04.20.Jb

1. THE MAIN EQUATIONS

In this paper we consider a class of solutions of the Einstein equations, for which the metric on the space-time manifold M^4 has a three-dimensional group of scale transformations:

$$G: M^4 \to M^4, \quad G_* \, ds^2 \to f(G) ds^2. \tag{1}$$

Here the map $G \to f(G)$ is a homomorphism of the Lie group G into the multiplicative group R^1_+. As was shown at first by Eardley,[1] for every three-dimensional group G of Bianchi type I–VII$_h$ there are self-similar generalizations of the kind (1) of homogeneous cosmological models, for which the group G acts on M^4 with three-dimensional spacelike orbits and the homomorphism $f(G)$ is a nontrivial one. In Ref. 1 solutions of type (1) were applied only in cosmology, as generalizations of homogeneous cosmological models. In this paper we investigate another application of solutions of the type (1) to the problem of nonstationary and nonspherical accretion of a self-gravitating gas onto an accreting center in general relativity.

Let us consider two metrics on M^4. The first metric is

$$ds^2 = g_{\mu\nu}\omega^\mu\omega^\nu, \quad \mu, \nu = 0,1,2,3, \tag{2}$$

where ω^μ are one-dimensional forms and $\omega^1 = d\rho$ (ρ is some function on M^4), and $g_{\mu\nu}$ are the metric components. The second metric is

$$ds^2 = e^{2\alpha\rho} g_{\mu\nu}\omega^\mu\omega^\nu, \tag{3}$$

where α is a constant. Let ω^i_j and $\tilde{\omega}^i_j$ denote the connection components of the metrics (2) and (3) which, as is well known,[2] satisfy the following equations:

$$d\omega^i + \omega^i_j \wedge \omega^j = 0, \tag{4}$$

$$dg_{ij} = \omega_{ij} + \omega_{ji}. \tag{5}$$

We obtain

Lemma 1: The connection components $\tilde{\omega}^i_j$ of the metric (3) are connected with connection components ω^i_j of the metric (2) by the following relations:

$$\tilde{\omega}^\mu{}_\nu = \omega^\mu{}_\nu + \alpha\delta^\mu{}_\nu d\rho - \alpha g^{\mu 1} g_{\nu\gamma}\omega^\gamma, \tag{6}$$

$$\tilde{\omega}^\mu{}_1 = \omega^\mu{}_1 + \alpha\omega^\mu - \alpha g^{\mu 1} g_{1\gamma}\omega^\gamma, \tag{7}$$

$$\tilde{\omega}^1{}_\mu = \omega^1{}_\mu - \alpha g^{11} g_{\mu\gamma}\omega^\gamma, \tag{8}$$

$$\tilde{\omega}^1{}_1 = \omega^1{}_1 + 2\alpha\, d\rho - \alpha g^{11} g_{1\gamma}\omega^\gamma. \tag{9}$$

where $\mu, \nu \neq 1$, $\gamma = 0,1,2,3$.

The proof of Lemma 1 consists simply in verifying the relations (4) and (5) for the connection $\tilde{\omega}^i_j$ and the metric (3). As it is well known,[2] the curvature forms $\mathbf{R}^\mu{}_\nu$ of the connection $\omega^\mu{}_\nu$ have the following form:

$$\mathbf{R}^\mu{}_\nu = d\omega^\mu{}_\nu + \omega^\mu{}_\beta \wedge \omega^\beta{}_\nu, \tag{10}$$

$$\mathbf{R}^\mu{}_\nu = R^\mu{}_{\nu\alpha\beta}\omega^\alpha \wedge \omega^\beta, \tag{11}$$

where $R^\mu{}_{\nu\alpha\beta}$ is the Riemann tensor.

Lemma 2: The curvature form $\tilde{\mathbf{R}}^\mu{}_\nu$ of the metric (3) is connected with the curvature form $\mathbf{R}^\mu{}_\nu$ of the metric (2) by the relations

$$\begin{aligned}\tilde{\mathbf{R}}^{\mu\nu} = (&\mathbf{R}^{\mu\nu} - \alpha^2 g^{11}\omega^\mu \wedge \omega^\nu + \alpha^2 g^{\mu 1} d\rho \wedge \omega^\nu \\ &- \alpha^2 g^{\nu 1} d\rho \wedge \omega^\mu + \alpha g^{\nu\beta}\omega^\mu \wedge \omega^1{}_\beta \\ &- \alpha g^{\mu\beta}\omega^\nu \wedge \omega^1{}_\beta) e^{-2\alpha\rho},\end{aligned} \tag{12}$$

where $\mu, \nu, \beta = 0,1,2,3$.

The proof of Lemma 2 consists in direct verification by using definitions (10) and (11) and relations (6), (7), (8), and (9). Note that the differential forms ω^μ in general are not closed and

$$d\omega^\alpha = -C^\alpha{}_{\mu\nu}\omega^\mu \wedge \omega^\nu. \tag{13}$$

The connection components ω^i_j of the metric (2) have the following form[2]:

$$\begin{aligned}\omega^\mu{}_\nu = &\tfrac{1}{2} g^{\mu\beta}(g_{\beta\nu,\alpha} + g_{\beta\alpha,\nu} - g_{\nu\alpha,\beta})\omega^\alpha \\ &+ \tfrac{1}{2}(-C^\mu{}_{\nu\alpha} + g^{\mu\beta}(g_{\nu\gamma}C^\gamma{}_{\beta\alpha} + g_{\alpha\gamma}C^\gamma{}_{\beta\nu}))\omega^\alpha,\end{aligned} \tag{14}$$

where $g_{\alpha\beta,\gamma}$ denotes the corresponding Lie derivative.

In the following we suppose that the metric (3) possesses a three-dimensional group of transformation (1), acting on M^4 with three-dimensional orbits, and the differential forms ω^μ are invariant under this action. The homomorphism $f(G)$ has the form

$$f(G) = e^{2\alpha\lambda(G)}, \tag{15}$$

where $\lambda(G)$ is some homomorphism $G \to R^1$. Under these assumptions the basis of G-invariant differential forms ω^μ may be chosen in such a way that

$$g_{00} = \pm 1, \quad g_{0i} = 0, \quad g_{ij} = g_{ij}(x^0), \quad i,j,\neq 0,$$

and the orbits of the Lie group G are given by the conditions $x^0 = \text{const}$. In this case the coefficient $C^\alpha{}_{\mu\nu}$ for $\alpha,\mu,\nu \neq 0$ are the structure constants of the Lie group G. By using the formulas (12) and (14) it is possible to derive the Ricci tensor

[a] Permanent address: V. A. Steklov Mathematical Institute of Academy of Sciences of USSR, Moscow, USSR.

 0022-2488/82/071353-06$02.50

components $R^i_j = R^{i\alpha}{}_{j\alpha}$ for the metric (3) and the Einstein tensor components, which have the following form:

$$\tilde{R}^0{}_0 - \tfrac{1}{2}\tilde{R} = e^{-2\alpha\rho}\{\tfrac{1}{8}g^{00}(\chi^\gamma{}_\gamma\chi^\beta{}_\beta - \chi^\beta{}_\gamma\chi^\gamma{}_\beta) - \tfrac{1}{2}P^\beta{}_\beta + \alpha^2 g^{11} - 2\alpha g^{1\gamma}C^\beta{}_{\gamma\beta}\}, \tag{16}$$

$$\tilde{R}_{0x} = -\tfrac{1}{2}(\chi^\gamma{}_\beta C^\beta{}_{\gamma x} - \chi^\gamma{}_x C^\mu{}_{\mu\gamma}) + \alpha\chi^1{}_x, \tag{17}$$

$$\begin{aligned}\tilde{R}^x{}_y - \tfrac{1}{2}\delta^x{}_y\tilde{R} = e^{-2\alpha\rho}\{&P^x{}_y - g^{00}(1/2|g|^{1/2})(|g|^{1/2}\chi^x{}_y) \\ &+ \alpha^2(-2g^{11}\delta^x{}_y + 2g^{x1}\delta^1{}_y) \\ &+ \alpha(2u^x{}_y + g^{1\gamma}C^\beta{}_{\gamma\beta}\delta^x{}_y) \\ &- \tfrac{1}{2}\delta^x{}_y(P^\beta{}_\beta - g^{00})(\dot{\chi}^\beta{}_\beta \\ &+ \tfrac{1}{4}(\chi^\beta{}_\beta\chi^\gamma{}_\gamma + \chi^\gamma{}_\beta\chi^\beta{}_\gamma)) \\ &- 6\alpha^2 g^{11} + 6\alpha g^{1\gamma}C^\mu{}_{\gamma\mu}\}.\end{aligned} \tag{18}$$

Here we use the following notations: $\beta,\gamma,\mu,\nu,x,y = 1,2,3$; $\chi^\gamma{}_\beta = g_{\beta\sigma}g^{\sigma\gamma}$; $\dot{g}_{\beta\sigma} = dg_{\beta\sigma}/dx^0$; $g = \det(g_{\mu\nu})$; the $P_{\beta\gamma}$ is the Ricci tensor of the metric $g_{\mu\nu}$, restricted to the orbits of the Lie group G:

$$P_{\beta\gamma} = -\hat{\Gamma}^x{}_{\beta\gamma}\hat{\Gamma}^y{}_{yx} + C^y{}_{yx}\hat{\Gamma}^x{}_{\beta\gamma}, \tag{19}$$

$$\hat{\Gamma}^x{}_{\beta\gamma} = \tfrac{1}{2}(-C^x{}_{\beta\gamma} + g^{xy}(g_{\beta z}C^z{}_{y\gamma} + g_{\gamma z}C^z{}_{y\beta})), \tag{20}$$

$$u_{xy} = \tfrac{1}{2}g^{1\gamma}(g_{x\beta}C^\beta{}_{\gamma y} + g_{y\beta}C^\beta{}_{\gamma x}), \tag{21}$$

$$u^x{}_y = \tfrac{1}{2}g^{1\gamma}(C^x{}_{\gamma y} + g^{xz}g_{y\beta}C^\beta{}_{\gamma z}). \tag{22}$$

Note that the Einstein tensor for the metric (3) has been calculated at first by Eardley,[1] by using another method. From formulas (16), (17), and (18) it follows that in the case under consideration the Einstein tensor components depend on only one variable x^0. If the distribution of matter is also G-invariant, in the sense that the stress-energy tensor T_{ij} depends only on one variable x, then the whole system of Einstein's equations is reduced to some system of ordinary differential equations.

The hydrodynamical equations

$$T^k{}_{i;k} = \partial_k T^k{}_i - \Gamma^m{}_{ik}T^k{}_m + \Gamma^k{}_{mk}T^m{}_i, \tag{23}$$

for the G-invariant solutions under consideration, have the following form:

$$T^k{}_{i;k} = \frac{1}{|g|^{1/2}}\left(\frac{\partial|g|^{1/2}T^0{}_i}{\partial x^0}\right) + 2\alpha T^1{}_i - C^\beta{}_{\gamma\beta}T^\gamma{}_i - \Gamma^m{}_{ik}T^k{}_m. \tag{24}$$

The last term in (23) has the form

$$\Gamma^m{}_{0k}T^k{}_m = \tfrac{1}{2}\chi^\gamma{}_\beta T^\beta{}_\gamma, \tag{25}$$

$$\Gamma^m{}_{1k}T^k{}_m = \alpha T + \hat{\Gamma}^m{}_{1k}T^k{}_m, \tag{26}$$

$$\Gamma^m{}_{\sigma k}T^k{}_m = \hat{\Gamma}^m{}_{\sigma k}T^k{}_m, \tag{27}$$

where $\sigma = 2,3$ and

$$\hat{\Gamma}^m{}_{\gamma k} = \tfrac{1}{2}(-C^m{}_{\gamma k} + g^{m\delta}(g_{\gamma x}C^x{}_{\delta k} + g_{kx}C^x{}_{\delta\gamma})). \tag{28}$$

Note that $C^i{}_{jk} = 0$ if one of i,j,k equals zero.

The procedure for deriving the system of ordinary differential equations from the Einstein equations

$$R^i{}_j - \tfrac{1}{2}\delta^i_j R = T^i{}_j, \tag{29}$$

for the hydrodynamical stress-energy tensor $T^i{}_j$, consists in the following: from the equations $R_{0i} = T_{0i}$ and $R_{00} - \tfrac{1}{2}g_{00}R = T_{00}$ we derive the expressions for the velocity components u^i and the energy density through the metric components $g_{\mu\nu}$ and their first derivatives $g_{\mu\nu}$. Then after substitution of the derived expressions into Eqs. (29) for $i,j,\neq 0$ we derive a closed system of second-order equations only for metric components. This system may be transformed by a standard way into a dynamical system (system of first-order equations) and determines completely the evolution of the self-similar metric (3).

2. THE MODELS OF NONSTATIONARY ACCRETION

In the paper by Eardley,[1] self-similar solutions of the type (3) were considered as generalizations of homogeneous cosmological models. In this case it is supposed that $g_{00} = +1$, that is the variable x^0 is interpreted as time and the metric g_{ij} depends essentially upon the time. In this work we consider also the other case, when $g_{00} = -1$, the function $e^{\alpha\rho}$ has the meaning of a radial coordinate $R = e^{\alpha\rho}$ and the metric g_{ij} depends essentially on the spacelike coordinate x^0; dependence on the other coordinates is determinated by the structure of the group G. For explanation of the physical meaning of these solutions, let us show two solutions in empty space for groups of Bianchi types I and III:

$$ds^2 = R^2\cos^2 x^0\,dt^2 - dR^2 - R^2(d(x^0)^2 + \sin^2 x^0\,d\phi^2), \tag{30}$$

$$ds^2 = R^2\sin^2 x^0\,dt^2 - dR^2 - R^2(d(x^0)^2 + e^t\sin^2 x^0\,d\phi^2). \tag{31}$$

From the exact form of these solutions it follows that the variable x^0 is an angle variable ($-\delta/2 \leqslant x^0 \leqslant \delta/2$) and that the metric of space at infinity ($R\to\infty$) is the Euclidean one. In the following we shall consider also self-similar solutions, for which the variable x^0 is an angle variable and the variable $R = e^{\alpha\rho}$ is a radial coordinate, and which are nonstationary, i.e., they depend on time t. So we are interested in solutions, which depend on three variables: x^0, $R = e^{\alpha\rho}$, and t. These solutions may be applied to models for nonspherically-symmetric accretion of gas onto the center. In these solutions the increasing of the gas mass in the neighborhood of the center as a result of gas accretion, leads to the nonstationarity of the metric. This nonstationarity may have, for example, the form (31). Note that such self-similar solutions have no self-similar analogs in Newtonian theory, because in Newtonian theory all self-similar solutions, which depend essentially on the angle coordinate, are stationary (there are the conical flows of gas). Let us find which nonstationary self-similar solutions of the form (3) at $g_{00} = -1$ can exist. For that we use the list of generators and invariant differential 1-forms for three-dimensional Lie algebras G.[3] Let ξ_1, ξ_2, ξ_3 be the basis of generators in G.[3] For construction of self-similar solutions the presence of a homomorphism

$$f_*: \mathbf{G} \to R^1 \tag{32}$$

is necessary. For this homomorphism f_* $[\mathbf{G},\mathbf{G}] = 0$. For all models of types IV–VII the generators ξ_1 and ξ_2 belong to the commutator subalgebra $[\mathbf{G},\mathbf{G}]$, so $f_*(\xi_1) = 0$ and $f_*(\xi_2) = 0$. Therefore, for models of types IV–VII there exists only the homomorphism $f_*(\xi_3) = \alpha$. The self-similar metric (3) in these cases has the form

$$ds^2 = e^{2\alpha x^1}g_{ij}(x^0)\omega^i\omega^j, \tag{33}$$

where all forms ω^i depend on the coordinate x^1 (see the list of

generators for three-dimensional algebras in Ref. 3). Therefore in these solutions (33), the metric depends on only two spacelike variables $R = e^{\alpha x^1}$ and x^0.

For groups of Bianchi type I all differential forms are $\omega^i = dx^i$; therefore, the self-similar metric (33) for model of type I is also stationary.

Let us consider now models of types II and III. The generators of the corresponding Lie algebras have the following form[3]:

$$\text{type II: } \xi_1 = \partial_2, \quad \xi_2 = \partial_3, \quad \xi_3 = \partial_1 + x^3\partial_2,$$
$$\omega_1 = dx^2 - x^1 dx^3, \quad \omega^2 = dx^3, \quad \omega^3 = dx^1,$$
$$d\omega^1 = \omega^2 \wedge \omega^3, \quad C^1{}_{23} = -C^1{}_{32} = 1. \tag{34}$$

$$\text{type III: } \xi_1 = \partial_2, \quad \xi_2 = \partial_3, \quad \xi_3 = \partial_1 + x^2\partial_2,$$
$$\omega^1 = e^{-x^1} dx^2, \quad \omega^2 = dx^3, \quad \omega^3 = dx^1,$$
$$d\omega^1 = \omega^1 \wedge \omega^2, \quad C^1{}_{13} = -C^1{}_{31} = 1. \tag{35}$$

For these two algebras the commutator subalgebra [G, G] is one-dimensional: all generators have the form $[G, G] = \mu\xi_1$. So there exist two essentially different homomorphisms:

$$f^{(1)}_*(\xi_1) = 0, \ f^{(1)}_*(\xi_2) = 0, \ f^{(1)}_*(\xi_3) = 0, \tag{36}$$

$$f^{(2)}_*(\xi_1) = 0, \ f^{(2)}_*(\xi_2) = 0, \ f^{(2)}_*(\xi_3) = 0, \tag{37}$$

The self-similar metrics, corresponding to the homomorphism $f^{(2)}_*$, have the form

$$ds^2 = e^{2\alpha x^1} g_{ij}(x^0)\omega^i\omega^j, \tag{38}$$

where the differential forms ω^i depend on the variable x^1 [see (34) and (35)]. Then these metrics depend on three variables x^0, $R = e^{\alpha x^1}$, and x^1. If the variable x^1 is timelike then these metrics are nonstationary. Therefore only the self-similar metrics of Bianchi types II and III may be used for modelling self-similar nonstationary accretion of gas onto the center (in a nonspherically-symmetric case). In the following sections we begin to study self-similar solutions of Bianchi type II and III in the most simple cases.

3. THE EINSTEIN EQUATIONS FOR SELF-SIMILAR BIANCHI TYPE II SOLUTIONS IN EMPTY SPACE

In this section we consider self-similar Bianchi type II solutions (3) with one nondiagonal metric component $g_{23}(x_0)$. The metric in this case has the form

$$ds = e^{2\alpha\rho}(g_{00}d(x^0)^2 + g_{11}(x^0)\,d\rho^2 + g_{22}(x^0)\,dt^2 + 2g_{23}(x^0)\,dt\,\omega^3 + g_{33}(x^0)(\omega^3)^2), \tag{39}$$

where $\omega^3 = d\phi + Bt\,d\rho$, $d\omega^3 = -B\,d\rho \wedge dt = -B\omega^1 \wedge \omega^2$, $C^3{}_{12} = B$. For the metric (39) we shall reduce Einstein's equations in empty space to the simplest possible form, prove that the variable x^0 for which $g_{00} = -1$ has the meaning of an angle coordinate [that is the determinant $g = \det(g_{ij}(x^0))$ has two zeros], and show some exact solutions for $g_{00} = 1$.

The Einstein equations in empty space for metrics (39), due to equations (16), (17), and (18), have the form

$$R^1{}_1 = -\tfrac{1}{2}B^2 g^{11}g^{22}g_{33} - g^{00}(1/2|g|^{1/2})(|g|^{1/2}\chi^1{}_1)^{\cdot} = 0, \tag{40}$$

$$R^2{}_2 = -\tfrac{1}{2}B^2 g^{11}g^{22}g_{33} - g^{00}(1/2|g|^{1/2})(|g|^{1/2}\chi^2{}_2)^{\cdot} - 2\alpha^2 g^{11} + \alpha B g^{11}g^{22}g_{23} = 0, \tag{41}$$

$$R^3{}_3 = \tfrac{1}{2}B^2 g^{11}g^{22}g_{33} - g^{00}(1/2|g|^{1/2})(|g|^{1/2}\chi^3{}_3)^{\cdot} - 2\alpha^2 g^{11} - \alpha B g^{11}g^{22}g_{33} = 0, \tag{42}$$

$$R^3{}_2 = B^2 g^{11}g^{22}g_{23} - g^{00}(1/2|g|^{1/2})(|g|^{1/2}\chi^3{}_2)^{\cdot} + \alpha B g^{11}(2 - g^{22}g_{22}) = 0, \tag{43}$$

$$R^2{}_3 = -g^{00}(1/2|g|^{1/2})(|g|^{1/2}\chi^2{}_3)^{\cdot} + \alpha B g^{11}g^{22}g_{33} = 0, \tag{44}$$

$$R^0{}_0 = \tfrac{1}{2}R = \tfrac{1}{8}g^{00}(\chi^\alpha{}_\alpha\chi^\beta{}_\beta - \chi_\alpha{}^\beta\chi^\alpha{}_\beta) + \tfrac{1}{2}(\tfrac{1}{2}B^2 g^{11}g^{22}g_{33} + 2\alpha^2 g^{11}) = 0, \tag{45}$$

$$R^0{}_0 = -g^{00}(\tfrac{1}{2}\dot{\chi}^\alpha{}_\alpha + \tfrac{1}{4}\chi_\alpha{}^\beta\chi_\beta{}^\alpha) - 2\alpha^2 g^{11} = 0, \tag{46}$$

$$R = -g^{00}(\dot{\chi}^\alpha{}_\alpha + \tfrac{1}{4}(\chi_\alpha{}^\alpha\chi_\beta{}^\beta + \chi_\alpha{}^\beta\chi_\beta{}^\alpha)) - \tfrac{1}{2}B^2 g^{11}g^{22}g_{33} - 6\alpha^2 g^{11} = 0, \tag{47}$$

$$R_{01} = \tfrac{1}{2}B\chi^2{}_3 + \alpha\chi^1{}_1 = 0. \tag{48}$$

From the equation $R^2{}_3 = 0$ it follows that there are no solutions with $g_{23} = 0$, so it is necessary to consider the nondiagonal case $g_{23} \neq 0$. The two equations $R^0{}_0 - \frac{1}{2}R = 0$ and $R_{01} = 0$ do not connect second derivatives of metric, so these two equations are constraints. All other equations are equivalent to a system of four independent differential equations of second order. Note, that the constraint $R_{01} = 0$ is satisfied by virtue of the other Einstein equations, because the following identity holds:

$$\alpha R^1{}_1 + \tfrac{1}{2}BR^2{}_3 = -g^{00}(1/2|g|^{1/2})(|g|^{1/2} \times (\alpha\chi^1{}_1 + \tfrac{1}{2}B\chi^2{}_3))^{\cdot} = -g^{00}(1/2|g|^{1/2})(|g|^{1/2}R_{01})^{\cdot} = 0. \tag{49}$$

Let us suppose that the coordinate ρ is spacelike, $g_{11} < 0$, and introduce the new independent variable τ:

$$\frac{d\tau}{dx^0} = |g^{11}|^{1/2} = 1/|g_{11}|^{1/2}. \tag{50}$$

In the new variable τ we deduce, from Eqs. (40)–(49), the following system (here $X = dX/d\tau$, $g_1 = |g_{22}g_{33} - g_{23}^2|$):

$$-\tfrac{1}{2}B^2 g^{22}g_{33} + g^{00}(1/2g_1^{1/2})(g_1^{1/2}\chi^1{}_1)^{\cdot} = 0, \tag{51}$$

$$-\tfrac{1}{2}B^2 g^{22}g_{33} + g^{00}(1/2g_1^{1/2})(g_1^{1/2}\chi^2{}_2)^{\cdot} - 2\alpha^2 + \alpha B g^{22}g_{23} = 0, \tag{52}$$

$$-\tfrac{1}{2}B^2 g^{22}g_{33} + g^{00}(1/2g_1^{1/2})(g_1^{1/2}\chi^3{}_3)^{\cdot} - 2\alpha^2 - \alpha B g^{22}g_{23} = 0, \tag{53}$$

$$g^{00}(1/2g_1^{1/2})(g_1^{1/2}\chi^2{}_3)^{\cdot} + \alpha B g^{22}g_{33} = 0, \tag{54}$$

$$B^2 g^{22}g_{23} + g^{00}(1/2g_1^{1/2})(g_1^{1/2}\chi^3{}_2)^{\cdot} + \alpha B(2 - g^{22}g_{22}) = 0, \tag{55}$$

$$\tfrac{1}{8}g^{00}(2\chi^1{}_1(\chi^2{}_2 + \chi^3{}_3) + 2\chi^2{}_2\chi^3{}_3 - 2\chi^3{}_2\chi^2{}_3) - \tfrac{1}{4}B^2 g^{22}g_{33} - \alpha^2 = 0. \tag{56}$$

By adding Eqs. (52) and (53), we obtain

$$g^{00}(1/2g_1^{1/2})(g_1^{1/2}(\chi^2{}_2 + \chi^3{}_3))^{\cdot} - 4\alpha^2 = 0, \tag{57}$$

which is equivalent to the equation

$$g^{00}(g_1^{1/2})^{\cdot\cdot}/g_1^{1/2} - 4\alpha^2 = 0. \tag{58}$$

For $g^{00} = -1$ all solutions of Eq. (58) have the form

$$g_1 = C\sin^2(2\alpha(\tau - \tau_0)). \tag{59}$$

So all solutions of system (51)–(56) have two zeros of g_1 at $\tau = \tau_0$ and $\tau = \tau_0 + \pi/2\alpha$. Hence the variable τ, and there-

fore the variable $\dot{x}^0$, also for $g_{00} = -1$, has the meaning of an angle coordinate and all solutions of system (51)–(56) are determined on the segment $\tau_0 < \tau < \tau_0 + \pi/2\alpha$.

Equation (51) separates from the others, which determine a closed system in the space of two-dimensional matrices (with coordinates g_{22}, g_{33}, g_{23}). Equation (51) follows from this system at the level of constraint $R_{01} = 0$:

$$\chi^1{}_1 = -(B/2\alpha)\chi^2{}_3. \tag{60}$$

Let us introduce in the space of two-dimensional matrices the following coordinates:

$$g_{23}, g_1 = g_{22}g_{33} - g_{23}^2, \quad h = g_{23}/g_{33}. \tag{61}$$

In these coordinates we have

$$\chi^2{}_3 = \dot{h}g_{33}^2/g_1, \tag{62}$$

$$\chi^3{}_2 = \dot{h} - h\dot{g}_1/g_1 + 2h\dot{g}_{33}/g_{33} - \dot{h}h^2 g_{33}^2/g_1, \tag{63}$$

$$\chi^2{}_2 = \dot{g}_1/g_1 - \dot{g}_{33}/g_{33} + \dot{h}hg_{33}^2/g_1, \tag{64}$$

$$\chi^3{}_3 = \dot{g}_{33}/g_{33} - \dot{h}hg_{33}^2/g_1. \tag{65}$$

The order of the system (51)–(56) falls after the transformation into the following new coordinates:

$$x = \dot{h} + (B/4\alpha)(\dot{g}_1/g_1), \quad z = \dot{g}_1/g_1, \quad u = g_{33}^2/g_1. \tag{66}$$

Equation (56) (equation $R^0{}_0 - \frac{1}{2}R = 0$) is equivalent to the equation

$$X = -\frac{1}{4}g^{00}\left(ux^2 + \frac{1}{4}\left(\frac{\dot{u}}{u}\right)^2\right) + \left(\frac{g^{00}}{16}z^2 - \alpha^2\right)\left(\frac{B^2}{4\alpha^2}u + 1\right) = 0. \tag{67}$$

All other equations (51)–(55) are equivalent to the following system:

$$g^{00}\frac{1}{2}\left(\frac{\dot{u}}{u}\right)^{\cdot} + \frac{g^{00}}{4}z\frac{\dot{u}}{u} - g^{00}ux^2 + g^{00}\frac{B}{2\alpha}uxz - \frac{B^2}{\alpha^2}u\left(\frac{g^{00}}{16}z^2 - \alpha^2\right) = 0, \tag{68}$$

$$\frac{g^{00}}{2}\dot{x} + \frac{g^{00}}{4}xz + g^{00}\frac{x}{2}\frac{\dot{u}}{u} - \frac{B}{8\alpha}g^{00}z\frac{\dot{u}}{u} = 0, \tag{69}$$

$$g^{00}\frac{\dot{z}}{2} + \frac{g^{00}}{4}z^2 - 4\alpha^2 = 0. \tag{70}$$

From the system (68), (69), and (70) it follows that

$$\dot{X} = -zX, \tag{71}$$

where X is determined by expression (67). Therefore the equation (67) is a constraint which is preserved by the system (68), (69), and (70). From Eq. (67) it is easy to find an expression of variable u through the variables x, z, and $\dot{u}/u$:

$$u = \frac{64\alpha^2}{-16\alpha^2 g^{00}x^2 + g^{00}B^2z^2 - 16\alpha^2B^2} \times\left(\alpha^2 + \frac{g^{00}}{16}\left(\left(\frac{\dot{u}}{u}\right)^2 - z^2\right)\right). \tag{72}$$

After substitution of this expression into system (68), (69), (70), we derive a closed system of three differential equations in the variables x, z, and $\dot{u}/u$. The last equation (for z) separates from the others and may be easily integrated:

$$g_{00} = -1, \quad z = 2\alpha\cot 2\alpha\tau. \tag{73}$$

After substitution of this solution into Eqs. (68) and (69), we obtain the following system of two differential equations:

$$\dot{y} = -2\alpha(\cot 2\alpha\tau)y + 2ux^2 - 4B(\cot 2\alpha\tau)ux + (2B^2/\sin^2 2\alpha\tau)u, \tag{74}$$

$$\dot{x} = -2\alpha(\cot 2\alpha\tau)x - xy + B(\cot 2\alpha\tau)y, \tag{75}$$

where

$$y = \frac{\dot{u}}{u} \quad u = \frac{1}{x^2\sin^2(2\alpha\tau) - B^2}\left(4\alpha^2 - \frac{y^2\sin^2 2\alpha\tau}{4}\right). \tag{76}$$

This system completely determines the evolution of all self-similar Bianchi type II solutions of Einstein equations in empty space.

Note that the system (68), (69), and (70) for $g_{00} = 1$ has the singular point

$$z = 4\alpha, \quad x = 0, \quad \dot{u}/u = 0. \tag{77}$$

The following exact solution of Einstein equations in empty space corresponds to this singular point:

$$ds^2 = e^{2\alpha\rho}\Big\{d(x^0)^2 - (x^0 + a_4)^2 d\rho^2 + \exp\left(\frac{8\alpha^2 a_2}{(Ba_1)^2}\ln(x^0 + a_4)\right) \times\Big[-\left(\frac{a_2}{a_1} + \frac{16\alpha^2 a_2^2}{Ba_1^3}(\ln(x^0 + a_4) + a_3)^2\right)dt^2 + \frac{8\alpha a_2}{Ba_1}(\ln(x^0 + a_4) + a_3)\,dt\,\omega^3 - a_1(\omega^3)^2\Big]\Big\}, \tag{78}$$

where $a_1 > 0$, $a_2 > 0$, a_3, a_4 are constants.

4. SOME EXACT SOLUTIONS FOR SELF-SIMILAR MODELS OF BIANCHI TYPE I AND III

The metric of the Bianchi type III self-similar solution with one nondiagonal term g_{12}, has the following form:

$$ds^2 = e^{2\alpha\rho}(g_{00}d(x^0)^2 + g_{11}(x^0)\,d\rho^2 + g_{12}(x^0)\,dt\,d\rho + g_{22}(x^0)\,dt^2 + g_{33}(x^0)(\omega^3)^2). \tag{79}$$

where

$$\omega^3 = e^{Bt}\,d\phi, \quad d\omega^3 = B\omega^2\wedge\omega^3, \quad C^3{}_{23} = -B. \tag{80}$$

Einstein's equations for metrics (79) in empty space have the form

$$R^1{}_1 = -g^{00}(1/2|g|^{1/2})(|g|^{1/2}\chi^1{}_1)^{\cdot} - \alpha Bg^{12} = 0, \tag{81}$$

$$R^2{}_1 = -g^{00}(1/2|g|^{1/2})(|g|^{1/2}\chi^2{}_1)^{\cdot} + 2\alpha^2 g^{12} = 0, \tag{82}$$

$$R^1{}_2 = -g^{00}(1/2|g|^{1/2})(|g|^{1/2}\chi^1{}_2)^{\cdot} - B^2 g^{12} = 0, \tag{83}$$

$$R^2{}_2 = -g^{00}(1/2|g|^{1/2})(|g|^{1/2}\chi^2{}_2)^{\cdot} - 2\alpha^2 g^{11} - B^2 g^{22} - \alpha Bg^{12} = 0, \tag{84}$$

$$R^3{}_3 = -g^{00}(1/2|g|^{1/2})(|g|^{1/2}\chi^3{}_3)^{\cdot} - 2\alpha^2 g^{11} - B^2 g^{22} - 3\alpha Bg^{12} = 0, \tag{85}$$

$$R_{01} = \alpha\chi^1{}_1 + \tfrac{1}{2}B\chi^2{}_1 = 0, \tag{86}$$

$$R_{02} = \tfrac{1}{2}B(\chi^2{}_2 - \chi^3{}_3) + \alpha\chi^1{}_2 = 0, \tag{87}$$

$$R^0{}_0 - \tfrac{1}{2}R = \tfrac{1}{8}g^{00}((\chi^\alpha{}_\alpha)^2 - \chi^\beta{}_\alpha\chi^\alpha{}_\beta) + \alpha^2 g^{11} + B^2 g^{22} + 2\alpha Bg^{12} = 0, \tag{88}$$

$$R = -T = -g^{00}(\dot{\chi}^\alpha{}_\alpha + \chi_\alpha{}^\beta\chi_\beta{}^\alpha) - 2B^2 g^{22} - 6\alpha^2 g^{11} - 6\alpha Bg^{12} = 0. \tag{89}$$

The system (81)–(89) is equivalent to a system of four differential equations of second order, because of three constraints $R_{01}=0, R_{02}=0, R^0{}_0-\frac{1}{2}R=0$. Conservation of the two constraints $R_{01}=0$ and $R_{02}=0$ by this system follows from the identities

$$\begin{aligned}&\tfrac{1}{2}B(R^2{}_2-R^3{}_3)+\alpha R^2{}_1\\&\quad=-g^{00}(1/2|g|^{1/2})(|g|^{1/2})(\tfrac{1}{2}B(\chi^2{}_2-\chi^3{}_3)+\alpha\chi^1{}_2))\\&\quad=-g^{00}(1/2|g|^{1/2})(|g|^{1/2}R_{02})^{\cdot}=0,\end{aligned}\tag{90}$$

$$\begin{aligned}&\alpha R^1{}_1+\tfrac{1}{2}BR^2{}_1\\&\quad=-g^{00}(1/2|g|^{1/2})(|g|^{1/2}(\alpha\chi^1{}_1+\tfrac{1}{2}B\chi^2{}_1))^{\cdot}\\&\quad=-g^{00}(1/2|g|^{1/2})(|g|^{1/2}R_{01})^{\cdot}=0,\end{aligned}\tag{91}$$

Conservation of the third constraint $R^0{}_0-\frac{1}{2}R=0$ follows on the levels $R_{01}=0$ and $R_{02}=0$ from the identity

$$(R^0{}_0-\tfrac{1}{2}R)^{\cdot}=-(\dot{g}/g)(R^0{}_0-\tfrac{1}{2}R).\tag{92}$$

In the whole, the system (81)–(89) may be reduced to some system of four equations of first order.

Diagonal solutions of the system (81)–(89) are reduced to exact solutions (30) [for type I ($B=0$)] and to (31) (for type III). Let us show some exact diagonal solutions of Einstein's equations with a hydrodynamical stress-energy tensor. Such solutions, at $g_{00}=-1$, exist only for $u^1=u^3=0$. Indeed, two Einstein's equations

$$R^1{}_2=(p+\epsilon)u_2u^1,\tag{93}$$

$$R^3{}_2=(p+\epsilon)u_2u^3,\tag{94}$$

for diagonal metrics are identically equal to zero. But for $g_{00}=-1$ we have $u_2\neq 0$; therefore, $u_1=u^1=u_3=u^3=0$. Further, in the diagonal case, from the equation

$$R_{01}=(p+\epsilon)u_0u_1=0,\tag{95}$$

we deduce $\chi^1{}_1=0$. From expressions (81)–(89) it follows in the diagonal case that $R^1{}_1=0$. So from Einstein's equation

$$R^1{}_1=T^1{}_1-\tfrac{1}{2}T=(p+\epsilon)u_1u^1+\tfrac{1}{2}(p-\epsilon)\tag{96}$$

and $u_1=0$, we deduce the necessary condition $p=\epsilon$. Thus, for $g_{00}=-1$ the diagonal solutions exist only for $u^1=u^3=0, p=\epsilon$. In this case the metric (79) has the form

$$\begin{aligned}ds^2=e^{2\rho}(&-d\theta^2+g_{11}(\theta)\,d\rho^2\\&+g_{22}(\theta)\,dt^2+g_{33}(\theta)e^{2Bt}\,d\phi^2).\end{aligned}\tag{97}$$

The 4-velocity vector of matter has the form $u^i=(u^0,0,u^2,0)$ $p=\epsilon$. Let us denote

$$g_{22}=z^2,\quad u=\dot{z}/z,\quad v=\tfrac{1}{2}\dot{g}_{33}/g_{33}.\tag{98}$$

From the equation $R_{01}=0$ it follows that $g_{11}=-A^2=\text{const}$. All other Einstein's equations are reduced to the system

$$\dot{z}=uz,\tag{99}$$

$$\dot{u}=-u^2-\frac{1}{A^2}\pm\left(\left(uv-\frac{B^2}{z^2}+\frac{1}{A^2}\right)^2-\frac{B^2}{z^2}(u-v)^2\right)^{1/2},\tag{100}$$

$$\dot{v}=-v^2-vu-\frac{2}{A^2}-\frac{B^2}{z^2}.\tag{101}$$

Energy density and pressure have the form

$$\epsilon=p=\frac{B^2}{2z^2}(u-v)^2\frac{z^2K^2-1}{L},\tag{102}$$

where

$$\begin{aligned}L=uv-\frac{B^2}{z^2}+\frac{1}{A^2}\\\pm\left(\left(uv-\frac{B^2}{z^2}+\frac{1}{A^2}\right)^2-\frac{B^2}{z^2}(u-v)^2\right)^{1/2},\end{aligned}\tag{103}$$

$$K=-\frac{L}{B(u-v)}.\tag{104}$$

The components of velocity of matter have the form

$$u^0=1/(z^2K^2-1)^{1/2},\quad u^2=1/(z^2K^2-1)^{1/2}.\tag{105}$$

Exact solution (31) may be obtained from the system (99), (100), (101) at $u-v=0$ and sign (-1) in Eq. (100). For self-similar Bianchi type I ($B=0$) diagonal solutions, from expressions (81)–(89), we have $R_{02}=0$, so from Einstein's equation $R_{02}=(p+\epsilon)u_0u_2$ it follows (at $g_{00}=-1$) that $u_0=0$. Together with $u^1=u^3=0$ it means that the motion of matter in these solutions is absent. Einstein's equations in this case have the following exact solutions:

$$ds^2=e^{2\rho}(-d\theta^2-A^2\,d\rho^2+g_{22}(\theta)\,dt^2+g_{33}(\theta)\,d\phi^2),\tag{106}$$

where

$$g_{22}=(z(\theta))^2,\ g_{33}=a_3\frac{dz}{d\theta}<0,\ g_{11}=-A^2=\text{const}\tag{107}$$

and function $z(\theta)$ is determined from the integral

$$\int\frac{dz}{(a_2+2a_1\ln(z)-(1/A^2)z^2)^{1/2}}=\theta-\theta_0.\tag{108}$$

Energy density and pressure have the form

$$\epsilon=p=a_1/z^2,\quad a_1>0.\tag{109}$$

Note that at $g_{00}=1$ there exist diagonal self-similar solutions of Bianchi type I and III with motion of matter $[u^i=(u^0,u^1,0,0)]$ and every equation of state $[p=k\epsilon, 0\leqslant k\leqslant 1]$. Such solutions of Bianchi type I at $g_{22}=g_{33}$ were investigated in Refs. 4 and 5. Let us show the new exact self-similar solution of Bianchi type III for $g_{00}=1, p=\epsilon$, $u^1=0$:

$$ds^2=e^{2\rho}(dt^2+g_{11}\,d\rho^2+g_{22}(t)\,dz^2+g_{33}(t)e^{2Bz}d\phi^2),\tag{110}$$

where

$$g_{11}=-A^2=\text{const},\tag{111}$$

$$g_{22}=\tfrac{1}{2}B^2A^2+C_1\exp(2/At)+C_2\exp(-(2/A)t)<0,\tag{112}$$

$$g_{33}=C_3g_{22},\quad C_3>0,\tag{113}$$

$$p=\epsilon=(4/A^2g_{22}^2)(\tfrac{1}{16}B^4A^4-C_1C_2).\tag{114}$$

The physical condition $\epsilon\geqslant 0$ denotes the permissible domain of constants values

$$C_1C_2\leqslant\tfrac{1}{16}B^4A^4.\tag{115}$$

For

$$C_1C_2=\tfrac{1}{16}B^4A^4.\tag{116}$$

we deduce from Eqs. (111)–(115) the exact solution in empty space, depending on three constants C_1,C_2, and C_3.

ACKNOWLEDGMENTS

This work has been performed under the auspices of GNFM-CNR of Italy. One of us (O.I.B.) acknowledges the National Committee for Mathematics of CNR for having supported his visit at the Seminario Matematico of Catania. Thanks are due to Professor Anile for his kind invitation and hospitality.

[1]D. M. Eardley, Comm. Math. Phys. **37**, 287 (1974)
[2]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1972).
[3]M. P. Ryan and L. C. Shepley, *Homogeneous Cosmological Models* (Princeton U. P., Princeton, N.J., 1975).
[4]M. E. Cahill and A. M. Taub, Comm. Math. Phys. **21**, 1 (1971).
[5]A. M. Taub, Comm. Math. Phys. **29**, 79 (1973).

Electrovac generalization of Neugebauer's $N = 2$ solution of the Einstein vacuum field equations [a]

Dong-sheng Guo [b] and Frederick J. Ernst
Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616

(Received 22 October 1981; accepted for publication 5 March 1982)

We show that the $N = 2$ Neugebauer solution of the Einstein vacuum field equations is easily reproduced by employing two successive Kinnersley–Chitre (K–C) transformations of a type considered earlier by I. Hauser. Furthermore, by employing two successive K–C transformations of a type considered recently by C. Cosgrove we are able to produce a new electrovac generalization of the $N = 2$ Neugebauer solution. In principle, an analogous approach could be employed for the explicit construction of electrovac generalizations of Neugebauer solutions corresponding to higher values of N.

PACS numbers: 04.20.Jb

I. INTRODUCTION

Not long ago Neugebauer[1] published a family of new solutions of the stationary axially symmetric vacuum field equations, which he had obtained by using a newly discovered Bäcklund transformation for the Ernst equation. His new solutions were, among other things, generalizations of the Tomimatsu–Sato (T–S) solutions. They also provided what may be called "multiple Kerr solutions."

Recently Cosgrove,[2] using the homogeneous Hilbert problem (HHP) formulation of Hauser and Ernst,[3] identified a Kinnersley–Chitre (K–C) transformation which provides an electrovac generalization of Neugebauer's Bäcklund transformation. Using this transformation Cosgrove was able to derive the charged Kerr solution for $a^2 + e^2 > m^2$.

In as yet unpublished work Hauser[4] developed a K–C transformation which yields the uncharged Kerr solution not only for $a > m$, but for $a = m$ and $a < m$ as well. This transformation, it turns out, was identified earlier by other methods. However, before we turn to Cosgrove's transformation, we shall illustrate our approach by showing that Hauser's transformation, when repeated twice, yields precisely the $N = 2$ Neugebauer solution. Then we shall evaluate for the first time the result of applying the Cosgrove transformation twice, thereby obtaining what may be considered to be an electrovac generalization of the $N = 2$ Neugebauer solution.

Our new solution should prove useful for finding electrovac generalizations of the Tomimatsu–Sato $\delta = 2$ solution analogous to the vacuum generalization which Kinnersley and Chitre[5] found. In this paper we shall show how the K–C generalization can be obtained from the $N = 2$ Neugebauer solution by an appropriate limiting process. How the (T–S) solution itself can be obtained was shown earlier by Kramer and Neugebauer[6] and was described fully by Sato.[7]

It remains a curious puzzle why no one has yet been able to solve an HHP for an electrovac transformation generalizing all three of Hauser's transformations.

II. KINNERSLEY–CHITRE TRANSFORMATIONS

The specification of a K–C transformation corresponds to the selection of a 3 by 3 matrix function $v(\tau)$ of a complex parameter τ, such that

$$v^\dagger(\tau)\mathfrak{E}v(\tau) = \mathfrak{E}, \quad \det v(\tau) = 1,$$

where

$$i\mathfrak{E} = \begin{pmatrix} 0 & i & 0 \\ -i & 0 & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

The parameter τ is related to the previously[3] used complex parameter t by $\tau = 1/(2t)$, while the matrix $v(\tau)$ is related to the previously used matrix $u(t)$ by

$$v(\tau) = \begin{pmatrix} 1/t & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} u(t) \begin{pmatrix} t & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Considered as a function of the complex parameter τ, $v(\tau)$ must be holomorphic in an open neighborhood of $\tau = 0$.

The $v(\tau)$ which we shall consider in this paper have an exponential form

$$v(\tau) = \exp[J\eta(\tau)],$$

where J is a constant 3 by 3 matrix and $\eta(\tau)$ is a real scalar function of the parameter τ. The group conditions require that J satisfy

$$\mathfrak{E}J + J^\dagger\mathfrak{E} = 0, \quad \mathrm{Tr}\, J = 0.$$

By explicit construction of the most general J we found that the relation

$$J^3 + aJ + ibI = 0$$

is satisfied, where a and b are real scalars. By rescaling $\eta(\tau)$ one may always arrange that $a = 1, 0$, or -1.

In the vacuum case considered by Hauser J was of the form

[a]Research supported in part by National Science Foundation grant PHY-79-08627.

[b]A more complete description of the research upon which this paper is based will appear in Mr. Guo's thesis, to be submitted to the I.I.T. graduate School in partial fulfillment of the requirements for the Ph.D. degree.

 0022-2488/82/071359-05$02.50

$$J=\begin{pmatrix} j & 0 \\ 0 & 0 \end{pmatrix},$$

where $j^2 = I$, 0, or $-I$, and where Hauser considered

$$\eta(\tau) = (\tfrac{1}{2})\ln[(\tau - K_2)/(\tau - K_1)]$$

for $a = 1$, and

$$\eta(\tau) = -(\tfrac{1}{2})i\ln[(\tau - K)/(\tau - K^*)]$$

for $a = -1$. This corresponds to $J^3 - aJ = 0$, i.e., to $b = 0$.

Thus far we have had no success solving the nonvacuum HHP with $J^3 - aJ = 0$ and

$$\eta(\tau) = (\tfrac{1}{2})\ln[(\tau - K_2)/(\tau - K_1)]$$

or

$$\eta(\tau) = -(\tfrac{1}{2})i\ln[(\tau - K)/(\tau - K^*)].$$

If one specializes J so that the minimal polynomial is of the second degree attempts to solve the resulting HHP, then, barring the discovery of a new method of solution, one is led to consider the transformation discovered earlier by Cosgrove, which with

$$\eta(\tau) = -i\ln[(\tau - K)/(\tau - K^*)]$$

corresponds to

$$J^2 + (\tfrac{1}{3})iJ + (\tfrac{2}{9})I = 0.$$

One may, in fact, write Cosgrove's J in the explicit form

$$J = (\tfrac{1}{3})iI + (2/E)hh^{\dagger}\mathfrak{F},$$

where h is an arbitrary 3 by 1 constant matrix, and where $E = 2ih^{\dagger}\mathfrak{F}h$.

III. THE $N = 2$ NEUGEBAUER SOLUTION

Applying Neugebauer's Bäcklund transformations, Kramer and Neugebauer obtained from Minkowski space a vacuum solution corresponding to a ξ potential which they expressed in the 4 by 4 determinant form

$$\xi = \frac{\begin{vmatrix} S_1 & S_2 & S_3 & S_4 \\ 1 & 1 & 1 & 1 \\ K_1 & K_2 & K_3 & K_4 \\ K_1^2 & K_2^2 & K_3^2 & K_4^2 \end{vmatrix}}{\begin{vmatrix} S_1 & S_2 & S_3 & S_4 \\ 1 & 1 & 1 & 1 \\ K_1 & K_2 & K_3 & K_4 \\ K_1S_1 & K_2S_2 & K_3S_3 & K_4S_4 \end{vmatrix}},$$

where the K_i $(i = 1,2,3,4)$ are arbitrary real constants, where

$$S_i = \exp(i\omega_i)r_i,$$

and where

$$r_i = [(z - K_i)^2 + \rho^2]^{1/2}.$$

The ω_i are arbitrary real phases.

In order to obtain from the $N = 2$ Neugebauer solution the generalization of the vacuum T–S solution discovered by Kinnersley and Chitre, one may proceed as follows. Select the phases ω_i so that

$$\exp(i\omega_1) = (p + iq)\exp(i\lambda_1 z_0),$$

$$\exp(i\omega_2) = -(p - iq)\exp(i\lambda_2 z_0),$$

$$\exp(i\omega_3) = (p + iq)\exp(i\lambda_3 z_0),$$

$$\exp(i\omega_4) = -(p - iq)\exp(i\lambda_4 z_0),$$

where p and q are constants with the property $p^2 + q^2 = 1$, and the λ_i are arbitrary constants. Then select the K_i so that

$$K_1 = mp + z_0,$$

$$K_2 = -mp + z_0,$$

$$K_3 = mp - z_0,$$

$$K_4 = -mp - z_0,$$

where m is an arbitrary constant (the mass parameter). If one now takes the limit as z_0 goes to zero, one obtains the ξ-potential of the K–C generalization of the T–S $\delta = 2$ solution. The K–C parameters α and β are related to our λ_i by

$$\alpha = (1/4)(\lambda_1 - \lambda_2 - \lambda_3 + \lambda_4)mp,$$

$$\beta = (1/4)(\lambda_1 + \lambda_2 - \lambda_3 - \lambda_4)mp.$$

IV. THE HAUSER TRANSFORMATION

In order to show that we get precisely the $N = 2$ Neugebauer solution by applying Hauser's transformation twice in succession to Minkowski space, we shall have to describe the as yet unpublished Hauser transformation in some detail. For $j^2 = I$ it is equivalent to the double Harrison[8] transformation, and for $j^2 = -I$ it is equivalent to the two-soliton transformation of Belinskii and Zakharov.[9]

As was remarked earlier, we now prefer to work in the τ-plane rather than the t-plane. Accordingly, in describing the HHP we shall introduce new symbols $Y_+(\tau) = X_-(t)$ and $Y_-(\tau) = X_+(t)$, as well as

$$P(\tau) = F(t)\begin{pmatrix} t & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

in place of the symbols which were employed in the earlier Hauser–Ernst papers. If the seed metric is Minkowski space, then (suppressing the third row and column)

$$P_0(\tau) = (1/(2r))\begin{pmatrix} -(r - \tau + z) & i(r + \tau - z) \\ -i & 1 \end{pmatrix},$$

where $r = [(z - \tau)^2 + \rho^2]^{1/2}$. The inverse of this matrix is given by

$$P_0(\tau)^{-1} = \begin{pmatrix} -1 & i(r + \tau - z) \\ -i & r - \tau + z \end{pmatrix}.$$

The HHP consists of finding $Y_+(\tau)$ and $Y_-(\tau)$ such that

$$Y_+ = Y_- P_0 \exp(j\eta)P_0^{-1}.$$

$Y_+(\tau)$ is holomorphic in τ except at the places where $\eta(\tau)$ has singularities, and $Y_-(\tau)$ is holomorphic in τ except at the places where $P_0(\tau)$ has singularities; namely, at the branch-points of the function $r(\tau)$. Furthermore, $Y_-(\tau)$ goes to I as τ goes to infinity.

By introducing the τ-dependent matrix field

$$\gamma = P_0 j P_0^{-1},$$

one can reexpress the HHP in the form

$$Y_+ = Y_- \exp(\gamma\eta),$$

or even better, one can write

$$Y_- = Y_+ \exp(-\gamma\eta),$$

where

$$\exp(-\gamma\eta) = [(I-\gamma)/2]\exp\eta + [(I+\gamma)/2]\exp(-\eta).$$

Hauser's solution of the HHP for the case $j^2 = I$ and

$$\eta(\tau) = (1/2)\ln[(\tau - K_2)/(\tau - K_1)]$$

is obtained by setting

$$Y_+ = N^0[A\exp\eta + B\exp(-\eta)],$$

where A and B are τ-independent matrix fields such that $A + B = I$,

$$A[I - \gamma(K_1)] = 0,$$

$$B[I + \gamma(K_2)] = 0,$$

and where

$$N^0 = \lim \exp[(\gamma\eta)]$$

as τ goes to infinity. The explicit evaluation of the matrices N^0, A, and B is very straightforward.

There are, of course, many ways to parametrize the constant j matrix, and which is best depends upon one's objectives. To facilitate comparison with the results of Kramer and Neugebauer we chose to express the j matrix of the first Hauser transformation in the form

$$j_1 = U_1 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} U_1^{-1},$$

where

$$U_1 = \begin{pmatrix} \sin\alpha_2 & \sin\alpha_1 \\ 1+\cos\alpha_2 & 1+\cos\alpha_1 \end{pmatrix},$$

and the j matrix of the second Hauser transformation in the form

$$j_2 = U_2 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} U_2^{-1},$$

where

$$U_2 = \begin{pmatrix} \sin\alpha_4 & \sin\alpha_3 \\ 1+\cos\alpha_4 & 1+\cos\alpha_3 \end{pmatrix}.$$

For the first Hauser transformation the real function $\eta(\tau)$ is given by

$$\eta_1 = (\tfrac{1}{2})\ln[(\tau - K_2)/(\tau - K_1)],$$

while for the second Hauser transformation

$$\eta_2 = (\tfrac{1}{2})\ln[(\tau - K_4)/(\tau - K_3)].$$

We shall also employ the notation

$$\gamma_n(\tau) = P_0 j_n P_0^{-1}.$$

(Note especially the subscript 0 rather than $n-1$ on P.)

Under the nth Hauser transformation the P potential is transformed from P_{n-1} to

$$P_n = N_n^0[A_n\exp\eta_n + B_n\exp(-\eta_n)]P_{n-1}\exp(-j_n\eta_n).$$

The $\mathscr{E}$-potential is easily obtained from

$$\mathscr{E} = \lim_{\tau\to\infty}[2\tau P_{44}(\tau)].$$

The actual calculations were carried out after having introduced certain quantities Q_i $(i = 1,2,3,4)$ which can be defined in terms of the eigenvectors

$$\begin{pmatrix} Q_1 \\ 1 \end{pmatrix}, \begin{pmatrix} Q_2 \\ 1 \end{pmatrix}, \begin{pmatrix} Q_3 \\ 1 \end{pmatrix}, \text{ and } \begin{pmatrix} Q_4 \\ 1 \end{pmatrix}$$

of $\gamma_1(K_1)$ corresponding to eigenvalue -1, of $\gamma_1(K_2)$ corresponding to eigenvalue $+1$, of

$$\exp[-\gamma_1(K_3)\eta_1(K_3)]\gamma_2(K_3)\exp[\gamma_1(K_3)\eta_1(K_3)]$$

corresponding to eigenvalue -1, and of

$$\exp[-\gamma_1(K_4)\eta_1(K_4)]\gamma_2(K_4)\exp[\gamma_1(K_4)\eta_1(K_4)]$$

corresponding to eigenvalue $+1$, respectively.

After a single application of the Hauser transformation to an arbitrary vacuum spacetime, we obtained the complex $\mathscr{E}$ potential

$$\mathscr{E} = \mathscr{E}(\text{seed}) - 2i(K_1 - K_2)/(Q_1 - Q_2),$$

while using two applications of the Hauser transformation to an arbitrary vacuum spacetime, we obtained the complex $\mathscr{E}$-potential

$$\mathscr{E} = \mathscr{E}(\text{seed}) - 2i\frac{\begin{vmatrix} Q_1 & Q_2 & Q_3 & Q_4 \\ 1 & 1 & 1 & 1 \\ K_1 & K_2 & K_3 & K_4 \\ K_1^2 & K_2^2 & K_3^2 & K_4^2 \end{vmatrix}}{\begin{vmatrix} Q_1 & Q_2 & Q_3 & Q_4 \\ 1 & 1 & 1 & 1 \\ K_1 & K_2 & K_3 & K_4 \\ K_1Q_1 & K_2Q_2 & K_3Q_3 & K_4Q_4 \end{vmatrix}}.$$

In the case when the seed metric is Minkowski space and $\mathscr{E}(\text{seed}) = 1$, one finds that the Q's have the form

$$Q_k = i[\exp(i\omega_k)r_k + (K_k - z)],$$

where $\omega_1 = \alpha_1$ and $\omega_2 = \alpha_2$, while ω_3 and ω_4 are more complicated functions of the α's and K's. The important thing to notice is that the ω's are unconstrained real constants, just as they were in Neugebauer's solution. In fact, it is quite easy to see that our result using Hauser's transformations is in complete agreement with the result of Kramer and Neugebauer using the Bäcklund transformation approach. In the case of a more general seed metric, our way of writing the $\mathscr{E}$-potential in terms of 4 by 4 determinants is an alternative to Neugebauer's manner of expression in terms of 5 by 5 determinants.

V. THE COSGROVE TRANSFORMATION

While we are not quite sure how Cosgrove himself solved the HHP corresponding to his transformations, we had no trouble reproducing his results, including the derivation of the charged Kerr metric with

$$a^2 + e^2 > m^2.$$

In particular, the HHP can be solved by noting that if

$$v^{-1} = L_2[(\tau - K^*)/(\tau - K)]^{1/3} + L_1[(\tau - K)/(\tau - K^*)]^{2/3},$$

where

$$L_1 = (2i/E)hh^{\dagger\mathfrak{G}}$$

and

$$L_2 = I - L_1$$

are projection operators onto one and two dimensional subspaces, respectively, then one may reasonably expect Y_+ to have the form

$$Y_+(\tau) = N^0 A\,[(\tau - K)/(\tau - K^*)]^{1/3} + N^0 B\,[(\tau - K^*)/(\tau - K)]^{2/3},$$

where A and B are τ-independent matrix fields such that $A + B = I$. In fact, if one writes out

$$Y_- = Y_+ P_0 v^{-1} P_0^{-1},$$

and

$$v^{-1} = \exp(-J\eta),$$

one sees that two of the terms have simple poles unless A and B are chosen so that

$$AP_0(K^*)L_1 = 0, \qquad BP_0(K)L_2 = 0.$$

These conditions are easily seen to require

$$B = \frac{EP_0(K^*)L_1P_0(K)^{-1}}{2h^\dagger i\mathfrak{G}P_0(K)^{-1}P_0(K^*)h}$$

and $A = I - B$. This is, in fact, the solution of the HHP.

Carrying out two successive Cosgrove transformations is a little more tedious than carrying out two successive Hauser transformations, because one deals in the electrovac case with 3 by 3 matrices, and because one must evaluate not only the gravitational complex potential but also the electromagnetic complex potential Φ. However, the method we used was basically the same.

The nth Cosgrove transformation corresponds to the J matrix

$$J_n = (1/3)iI + (2/E_n)h_n h_n^\dagger \mathfrak{G},$$

where

$$E_n = 2ih_n^\dagger \mathfrak{G} h_n,$$

and where h_n is a constant 3 by 1 matrix which is completely arbitrary. For the nth transformation we also write $v(\tau) = \exp(J_n\eta_n)$, where

$$\eta_n(\tau) = i\ln[(\tau - K_n^*)/(\tau - K_n)],$$

and we introduce the notation

$$\gamma_n(\tau) = P_0 J_n P_0^{-1}.$$

Under the nth Cosgrove transformation the P potential is transformed from P_{n-1} to

$$P_n = N_n^0[A_n\exp(i\eta_n/3) + B_n\exp(-2i\eta_n/3)] \times P_{n-1}\exp(-J_n\eta_n).$$

The $\mathscr{E}$ and Φ potentials are easily evaluated using

$$(-i,\mathscr{E},\Phi) = \lim_{\tau\to\infty}(0,2\tau,0)P(\tau).$$

As in the vacuum case, our actual calculations were performed after introducing certain quantities Q_i, R_i and S_i $(i = 1,\ldots,6)$ which may be defined (up to obvious linear combinations) in terms of the eigenvectors

$$\psi_i = \begin{pmatrix} Q_i \\ R_i \\ S_i \end{pmatrix}.$$

Here ψ_1 is an eigenvector of $\gamma_1(K_1^*)$ corresponding to eigenvalue $-2i/3$, ψ_2 and ψ_3 are eigenvectors of $\gamma_1(K_1)$ corresponding to eigenvalue $+i/3$, ψ_4 is an eigenvector of

$$\exp[-\gamma_1(K_2^*)\eta_1(K_2^*)]\gamma_2(K_2^*)\exp[\gamma_1(K_2^*)\eta_1(K_2^*)]$$

corresponding to eigenvalue $-2i/3$, and ψ_5 and ψ_6 are eigenvectors of

$$\exp[-\gamma_1(K_2)\eta_1(K_2)]\gamma_2(K_2)\exp[\gamma_1(K_2)\eta_1(K_2)]$$

corresponding to eigenvalue $+i/3$.

We applied two successive Cosgrove transformations to an arbitrary electrovac seed solution with $\mathscr{E}$- and Φ-potentials denoted by $\mathscr{E}$(seed) and Φ(seed), respectively. After a quite substantial calculation we obtained the $\mathscr{E}$- and Φ-potentials of the new electrovac spacetimes in the form

$$\mathscr{E} = \mathscr{E}(\text{seed}) - 2iN/D, \quad \Phi = \Phi(\text{seed}) - 2N'/D.$$

After the first Cosgrove transformation we obtained

$$D = \Delta_{123}, \quad N = (K_1^* - K_1)R_1\Delta_{23},$$
$$N' = (K_1^* - K_1)R_1\Delta'_{23}.$$

After the second Cosgrove transformation we obtained

$$D = \Delta_{423}\Delta_{156}(K_2 - K_1)(K_2^* - K_1^*) + \Delta_{523}\Delta_{164}(K_2^* - K_1)(K_2 - K_1^*) + \Delta_{623}\Delta_{145}(K_2^* - K_1)(K_2 - K_1^*),$$

while

$$N = \Delta_{456}\Delta_{32}R_1(K_1 - K_1^*)(K_1 - K_2^*)(K_1^* - K_2) + \Delta_{423}\Delta_{65}R_1(K_1 - K_1^*)(K_2 - K_2^*)(K_2 - K_1^*) + \Delta_{523}\Delta_{61}R_4(K_1 - K_1^*)(K_2^* - K_1)(K_2 - K_2^*) + \Delta_{623}\Delta_{51}R_4(K_1 - K_1^*)(K_2 - K_2^*)(K_1 - K_2^*) + \Delta_{123}\Delta_{65}R_4(K_1 - K_2)(K_1 - K_2^*)(K_2 - K_2^*)$$

and

$$N' = \Delta_{456}\Delta'_{32}R_1(K_1 - K_1^*)(K_1 - K_2^*)(K_1^* - K_2) + \Delta_{423}\Delta'_{65}R_1(K_1 - K_1^*)(K_2 - K_2^*)(K_2 - K_1^*) + \Delta_{523}\Delta'_{61}R_4(K_1 - K_1^*)(K_2^* - K_1)(K_2 - K_2^*) + \Delta_{623}\Delta'_{51}R_4(K_1 - K_1^*)(K_2 - K_2^*)(K_1 - K_2^*) + \Delta_{123}\Delta'_{65}R_4(K_1 - K_2)(K_1 - K_2^*)(K_2 - K_2^*).$$

In these expressions we have used the 3 by 3 determinants

$$\Delta_{ijk} = \det(\psi_i\psi_j\psi_k)$$

and the 2 by 2 determinants

$$\Delta_{ij} = \begin{vmatrix} R_i & R_j \\ S_i & S_j \end{vmatrix}$$

and

$$\Delta'_{ij} = \begin{vmatrix} Q_i & Q_j \\ R_i & R_j \end{vmatrix}.$$

It should be observed that the former satisfy the identity

$$\Delta_{423}\Delta_{156} + \Delta_{523}\Delta_{164} + \Delta_{623}\Delta_{145} = \Delta_{123}\Delta_{456}.$$

It is not difficult to evaluate the Q's, R's, and S's in terms of the components h^i $(i = 1,2,3)$ for the two successive Cosgrove transformations. In the special case when $h^3 = 0$ for both transformations, and when the seed metric is a vacuum space-time, we may set

$$Q_3 = Q_6 = R_3 = R_6 = S_1 = S_2 = S_4 = S_5 = 0,$$

$$R_1 = R_2 = R_4 = R_5 = S_3 = S_6 = 1.$$

The fields Q_1, Q_2, Q_4, and Q_5 then correspond to the fields Q_i ($i = 1, 2, 3, 4$) which we encountered in the consideration of the double Hauser transformation.

VI. POSSIBLE APPLICATIONS OF THESE RESULTS

While our new solution of the electrovac field equations, obtained by using Cosgrove's transformation twice in succession upon an arbitrary electrovac seed metric is perhaps interesting in its own right, since it is a natural generalization of Neugebauer's $N = 2$ vacuum solution, we anticipate that the solution will be of interest primarily as a source of additional electrovac solutions involving rational functions. In fact, it is our intention soon to publish T–S like solutions which are obtainable by limiting processes from our solution when the seed space-time is ordinary Minkowski space as well as generalized Plebański–Demiański solutions which are obtainable from our solution when the seed space-time is Minkowski space but an alternative pair of Killing vectors is focused upon. Naturally one may also contemplate the future construction of electrovac generalizations of Neugebauer's vacuum solutions with $N > 2$, although we shall probably leave that for someone else to do.

ACKNOWLEDGMENTS

We should like to express our appreciation to Professor I. Hauser for his interest in and suggestions concerning this work.

[1]G. Neugebauer, J. Phys. A: Math. Nucl. Gen. **13**, L19 (1980).
[2]C. Cosgrove, J. Math. Phys. **22**, 2624 (1981).
[3]I. Hauser and F. J. Ernst, J. Math. Phys. **21**, 1126 (1980); **21**, 1418 (1980); **22**, 1051 (1981).
[4]I. Hauser (private communication).
[5]W. Kinnersley and D. M. Chitre, J. Math. Phys. **19**, 2037 (1978).
[6]D. Kramer and G. Neugebauer, Phys. Lett. A **75**, 259 (1980).
[7]K. Oohara and H. Sato, Progr. Theor. Phys. **65**, 1891 (1981).
[8]B. K. Harrison, Phys. Rev. Lett. **41**, 1197 (1978).
[9]V. A. Belinskii and V. E. Zakharov, Sov. Phys. JETP **48**, 985 (1978).

New exact static solutions to Einstein's equations for spherically symmetric perfect fluid distributions

J. Ibañez
Departamento de Física, Facultad de Químicas, Universidad del Pais Vasco, Alza, San Sebastian, Spain
J. L. Sanz
Departamento de Física Teórica, Facultad de Ciencias, Universidad de Santander, Santander, Spain

(Received 19 February 1981; accepted for publication 22 May 1981)

New exact solutions to Einstein's equations are given which are spherically symmetric and static with perfect fluid distributions satisfying a linear equation of state $p = n\rho$ and $n\in(0,1]$. Heintzmann's generating method is then used to build up a family of new solutions for each value of n.

PACS numbers: 04.20.Jb

I. INTRODUCTION

Exact solutions to the equations of general relativity which are static, spherically symmetric, perfect-fluid distributions of matter have been considered in the literature (a review has been given by Kramer *et al.*[1]) to represent the interior of a relativistic star in equilibrium.

To get a semirealistic relativistic stellar interior, one should start with a reasonable equation of state for the distribution of matter. However, in practice most of people assume an ad hoc analytic expression for one of the components of the metric tensor or the density and then obtain the equation of state via the field equations. This computational method has the incovenience that one can easily obtain unphysical equations of state.

To avoid the last difficulty, we shall assume from the begining the equation of state of the distribution of matter; in fact, we shall consider a linear density-pressure relationship of the form $p = n\rho$, $n\in[0,1]$. After Einstein's equations are put into a form suitable for analysis, we obtain a particular solution (new as far as we know) for each $n\in(0,1]$. Next, we apply a method for generating a (possibly) new solution from another known solution due to Heintzmann[1,2] and obtain a one-parameter family of new solutions for each value of $n\in(0,1]$. Finally, some remarks concerning the regularity of these families are made, and we conclude with the hope that the solutions mentioned can be used as relativistic models in certain regions of the star.

II. BASIC EQUATIONS

The gravitational field being static and spherically symmetric, coordinates may be chosen so that the metric takes the form

$$ds^2 = -e^{2\nu}dt^2 + e^{2r}(e^{2\beta}dr^2 + d\Omega^2),$$
$$d\Omega^2 \equiv d\theta^2 + \sin^2\theta\, d\varphi^2, \tag{1}$$

where ν and β are each functions of the coordinate r.

As we are interested in perfect fluid distributions, the energy–momentum tensor is

$$T_{ab} = (\rho + p)u_a u_b + p g_{ab} \tag{2}$$

where ρ is density, p is pressure, and u^a is velocity.

In the obvious orthonormal tetrad ($\omega^0 = e^{\nu}dt$, $\omega^1 = e^{r+\beta}dr$, $\omega^2 = e^{r}d\theta$, $\omega^3 = e^{r}\sin\theta\, d\varphi$), Einstein's equations then are equivalent to the following set ($8\pi G = c = 1$):

$$u^a = \delta^a_0, \tag{3}$$

$$\rho = e^{-2(r+\beta)}(2\beta' + e^{2\beta} - 1)\left(' \equiv \frac{d}{dr}\right), \tag{4}$$

$$p = e^{-2(r+\beta)}(2\nu' - e^{2\beta} + 1), \tag{5}$$

$$\nu'' + \nu'^2 - (2+\beta')\nu' = 1 + \beta' - e^{2\beta}. \tag{6}$$

By assuming a linear density–pressure relationship

$$p = n\rho, \quad n\in[0,1], \tag{7}$$

Eqs. (4) and (5) imply

$$\rho = 2(1+n)^{-1}e^{-2(r+\beta)}(\nu' + \beta'), \tag{8}$$

$$\nu' = n\beta' + \tfrac{1}{2}(1+n)(e^{2\beta} - 1). \tag{9}$$

Thus, the problem of finding a static, spherically symmetric solution for a perfect-fluid distribution with a linear equation of state is reduced to obtain a solution for ν and β satisfying the ordinary differential equations (6) and (9). Once a solution has been found, the density and pressure are calculated via Eqs. (8) and (7), respectively.

III. EXACT ANALYTICAL SOLUTIONS

A. Case $n = 0$

For a dust distribution of matter ($p = 0$), Eq. (9) gives

$$\beta = \tfrac{1}{2}\ln(1 + 2\nu'). \tag{10}$$

Thus, by substituting Eq. (10) into Eq. (6), we obtain

$$\nu' = 0 \quad \text{or} \quad \nu'' = -\nu'(1 + 2\nu')$$

and these two equations obviously imply $\beta' = -\nu'$, i.e., $\rho \equiv 0$ taking into account the expression (8) for the density. Therefore, there exists no solution corresponding to a dust distribution such that $\rho > 0$. This is a very logical result from a physical point of view, because for such an equation of state there are no pressure gradients that can balance the attractive gravitational forces.

B. Case $n\in(0,1]$

A particular and simple solution to Eqs. (6) and (9), for any value of n, is given by

$$\nu = 2n(1+n)^{-1}r, \quad \beta = \tfrac{1}{2}\ln[1 + 4n(1+n)^{-2}]. \tag{11}$$

 0022-2488/82/071364-02$02.50

The corresponding density, obtained from Eq. (8), is

$$\rho = [1 + \tfrac{1}{4}(1+n)^2 n^{-1}]^{-1} e^{-2r}. \tag{12}$$

and the line element (1) then reads

$$ds^2 = -e^{2Ar}dt^2 + e^{2r}[(1 + A^2 n^{-1})dr^2 + d\Omega^2],$$
$$A \equiv 2n(1+n)^{-1}. \tag{13}$$

On choosing Schwarzschild (or canonical) coordinates, obtained making the change $e^r \to r$, our exact solution is

$$ds^2 = -r^{4n(1+n)^{-1}}dt^2 + [1 + 4n(1+n)^{-2}]dr^2 + r^2 d\Omega^2 \tag{14}$$
$$\rho = [1 + [\tfrac{1}{4}(1+n)^{-2}n^{-1}]^{-1}r^{-2}, \quad p = n\rho, \quad n \in (0,1].$$

As all the solutions violate the condition of regularity at the center insofar as the density and pressure become infinite at that point while g_{00} vanishes there, they can be considered unphysical solutions with regard to the possibility of representing stellar interiors globally. However, they can be used locally, then representing only certain regions of the star.

We shall comment that the 3-geometries $t = ct$ obtained from (14) can be embedded in four-dimensional Euclidean space. By setting $x^4 = 2(n)^{1/2}(1+n)^{-1}r$, one trivially obtain a surface that is a three-dimensional cone $C_3 \subset \mathbb{R}^4$.

IV. GENERATION OF NEW SOLUTIONS

By applying Heintzmann's generating method[1,2] to our exact analytical solution given by Eq. (14), we have obtained

$$ds^2 = -r^{4n(1+n)^{-1}}dt^2 + a[1 - Car^{2+b}]^{-1}dr^2 + r^2 d\Omega^2, \quad n \in (0,1],$$
$$\rho = 4n(1+n)^{-2}a^{-1}r^{-2} + C(3+b)r^b, \tag{15}$$
$$p = 4n^2(1+n)^{-2}a^{-1}r^{-2} - C[1 + 4n(1+n)^{-1}]r^b,$$

where

$$a \equiv 1 + 4n(1+n)^{-2}, \quad b \equiv 4n(1-n)(1+n)^{-1}(1+3n)^{-1},$$

and C is an arbitrary constant. Each solution obtained in this form is obviously new, being the old solution recovered for the value $C \equiv 0$.

We mention that this C-family of solutions, for each $n \in (0,1]$, violate the condition of regularity at the center because ρ and p become infinite at that point and g_{00} vanishes there (observe that all the solutions of the family behave as the $C = 0$ solution near the center).

Obviously for $C < 0$, the assumption $\rho > 0$ implies that the solution (15) can be used only in the region

$$r < [\tfrac{1}{4}|C|(1+n)^2 a(3+b)n^{-1}]^{-d}, \quad d \equiv (2+b)^{-1},$$

while, for $C > 0$, the assumptions $\rho > 0$ and $g_{00} > 0$ imply the following region of validity:

$$r < \min\{(Ca)^{-d}, [\tfrac{1}{4}Ca(1+n)(1+5n)n^{-2}]^{-d}\},$$
$$d \equiv (2+b)^{-1}.$$

Finally we remark that for the particular value $n = 1$ (a stiff equation of state for the old solution), Eq. (15) reads

$$ds^2 = r^2(-dt^2 + d\Omega^2) + 2(1 - 2Cr^2)^{-1}dr^2, \tag{16}$$
$$\rho = \tfrac{1}{2}r^{-2} + 3C, \quad p = \tfrac{1}{2}r^{-2} - 3C = \rho - 6C.$$

The solution generated in this form with $C > 0$ is exactly the irregular solution found by Buchdahl and Land[3] in their study of the most natural relativistic analog of the classical incompressible sphere.

ACKNOWLEDGMENTS

Thanks for stimulating and useful comments are due to R. Lapiedra.

[1] D. Kramer *et al.*, in *Exact solutions of Einstein's Field Equations* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1980), pp. 163–5.
[2] H. Heintzmann, Z. Phys. **228**, 489 (1969).
[3] H. A. Buchdahl and W. J. Land, Austr. Math. Soc. **8**, 6 (1968).

Static gravitational fields in a general class of scalar–tensor theories

S. B. Dutta Choudhury
Department of Physics, St. Anthony's College, Shillong 793003, India

D. Bhattacharya
Department of Physics, Rampurhat College, West Bengal, India

(Received 23 September 1980; accepted for publication 17 April 1981)

The static field equations are investigated within the framework of a general class of scalar–tensor theories of gravitation proposed by Nordtvedt. In the Brans–Dicke and Barker theories, it is shown that in vacuum g_{00} is functionally related to the scalar field. Two families of exact solutions are also obtained for a static spherically symmetric gravitational field in the Barker theory.

PACS numbers: 04.50. + h

1. INTRODUCTION

Nordtvedt[1] (1970) proposed a general class of scalar–tensor theories of gravitation in which the ω parameter of the Brans–Dicke theory is allowed to be an arbitrary function of the scalar field. This class includes the theories of Jordan[2] (1959) and Brans–Dicke[3] (1961) as special cases. In a recent communication Barker[4] (1978) considered another special case where $\omega = (4 - 3\psi)/(2\psi - 2)$, which ensures that the Newtonian gravitational constant G does not vary with time and also the strong principle of equivalence is not violated. In the same paper Barker discussed some interesting possibilities of the proposed theory and obtained an analytic solution for an empty universe from the standpoint of the theory.

Studies of static space-time within the framework of the general class of scalar–tensor theories have also received considerable attention in recent years. So far as the present paper is concerned, immediate reference may be made to the work of Banerjee and Bhattacharya[5] along with Dutta Choudhury and Banerjee[6] who, in their respective studies of static fields, in the Brans–Dicke and Barker theory, obtained expressions for g_{00} as a function of the scalar field, assuming beforehand that they are functionally related. Studies in this regard are presented in Sec. 2, where it is shown that the functional relationship between g_{00} and the scalar field in each theory follows directly from the corresponding field equations, without any prior assumption.

In Sec. 3 we have considered the case of the static spherically symmetric field about a point mass with ω in Barker's form. The field equations are investigated and the metric coefficients are at first expressed explicitly as a function of either the scalar field alone or both the scalar field and the radial coordinate. By virtue of these expressions, the whole system of field equations is effectively reduced to a single differential equation involving only a known function of the scalar field as the dependent variable, which on integration yields two families of exact solutions, which are probably the only known solutions in the present case.

2. THE STATIC FIELD EQUATIONS AND $g_{00} - \psi$ RELATIONS IN THE BARKER AND BRANS–DICKE THEORIES

The field equations in the metric formulation of Nordtvedt[1] can be expressed in the form

$$G_{\mu\nu} = -\frac{8\pi}{\psi}T_{\mu\nu} - \frac{\omega}{\psi^2}(\psi_{,\mu}\psi_{,\nu} - \tfrac{1}{2}g_{\mu\nu}\psi_{,\alpha}\psi^{,\alpha}) - \frac{1}{\psi}(\psi_{,\mu;\nu} - g_{\mu\nu}\Box\psi), \tag{2.1}$$

$$\Box\psi = \frac{8\pi T}{(2\omega+3)} - \frac{(d\omega/d\psi)}{(2\omega+3)}\psi_{,\alpha}\psi^{,\alpha}. \tag{2.2}$$

Equations (2.1) and (2.2) can be combined to yield[1]

$$\psi R_{\mu\nu} = -8\pi\left[T_{\mu\nu} - \frac{\omega+1}{(2\omega+3)}Tg_{\mu\nu}\right] + \frac{(d\omega/d\psi)}{2(2\omega+3)}\psi_{,\alpha}\psi^{\alpha}g_{\mu\nu} - \frac{\omega}{\psi}\psi_{,\mu}\psi_{,\nu} - \psi_{,\mu;\nu}. \tag{2.3}$$

The line element for a static space-time can be written as

$$ds^2 = g_{00}\,dt^2 + g_{ab}dx^a dx^b, \tag{2.4}$$

where g_{00} and g_{ab} are functions of space coordinates only, and a and b run from 1 to 3. For the static metric (2.4) it follows from the definition of $R_{\mu\nu}$ that

$$R^0_0 = F^{-1}F^{,a}_{;a}, \tag{2.5}$$

where

$$F^2 = g_{00} \tag{2.6}$$

and an index following a colon indicates covariant differentiation with respect to the metric tensor g_{ab}. Further, the field being static $\psi_{,0} = 0$, and one obtains

$$\psi^{,0}_{;0} = F^{-1}F^{,a}\psi_{;a}. \tag{2.7}$$

In view of (2.4)–(2.7) it follows from Eq. (2.3) that

$$8\pi\left[T^0_0 - \frac{(\omega+1)}{(2\omega+3)}T\right]F = -(\psi F^{;a})_{;a} + \frac{F(d\omega/d\psi)\psi_{,a}\psi^{,a}}{2(2\omega+3)}. \tag{2.8}$$

Equation (2.8) can also be written as

$$8\pi\left[T^0_0 - \frac{(\omega+0)}{(2\omega+3)}T\right]F = -\psi_{;a}F^{;a} - \frac{\psi}{\sqrt{-g^*}}\left(\sqrt{-g^*}F^{;a}\right)_{,a} + \frac{F(d\omega/d\psi)\psi_{,a}\psi^{,a}}{2(2\omega+3)}, \tag{2.9}$$

 0022-2488/82/071366-04$02.50

where g^* denotes the determinant of the g_{ab} matrix. Multiplying Eq. (2.9) throughout by $\sqrt{-g^*}$, one obtains

$$8\pi\left[T_0^0 - \frac{(\omega+1)}{(2\omega+3)}T\right]\sqrt{-g} = \frac{(d\omega/d\psi)\psi_{,a}\psi^{,a}\sqrt{-g}}{2(2\omega+3)} - \left(\sqrt{-g^*}\psi F^{,a}\right)_{,a}. \quad (2.10)$$

Again, in the static case under consideration, Eq. (2.2) yields

$$\frac{8\pi T\sqrt{-g}}{(2\omega+3)} = \frac{\sqrt{-g}(d\omega/d\psi)\psi_{,a}\psi^{,a}}{(2\omega+3)} + \left(\sqrt{-g^*}F\psi^{,a}\right)_{,a}. \quad (2.11)$$

Subtracting Eq. (2.10) from (2.11) one obtains

$$F^{-1}\nabla(F\psi) + \frac{(d\omega/d\psi)\psi_{,a}\psi^{,a}}{2(2\omega+3)} = 8\pi\left[\frac{(\omega+2)}{(2\omega+3)}T - T_0^0\right], \quad (2.12)$$

where ∇ is the covariant Laplace operator in the three dimensional, subspace whose metric tensor is g_{ab}. It may be noted here that when ω = constant, Eq. (2.12) reduces to the corresponding form in the Brans–Dicke theory (Buchdahl,[7] 1973).

When $T_{\mu\nu} = 0$ and $\omega = (4-3\psi)/(2\psi-2)$ (Barker's form), Eqs. (2.2) and (2.12) reduce, respectively, to

$$\Box(\sqrt{(\psi-1)} = 0, \quad (2.13)$$

$$4F^{-1}\nabla(F\psi) = g^{ab}(\psi-1)^{-1}\psi_{,a}\psi_{,b}, \quad (2.14)$$

where ψ now refers to the Barker scalar. Introducing now a new variable $f(\psi)$ by means of

$$\tan^{-1}\sqrt{\psi-1} = f, \quad (2.15)$$

one may express Eqs. (2.13) and (2.14), respectively, in the following forms:

$$(1/\sqrt{-g})(g^{ab}\sqrt{-g^*}F\sec^2 f f_{,b})_{,a} = 0, \quad (2.16)$$

$$(1/\sqrt{-g^*})\left[g^{ab}\sqrt{g^*}(F_{,b}\sec^2 f + 2F\sec^2 f\tan f f_{,b}\right]_{,a} = 0. \quad (2.17)$$

From Eq. (2.17) it is easy to see that

$$\frac{1}{\sqrt{-g}}(g^{ab}\sqrt{-g^*}F\sec^2 f f_{,b})_{,a}\left(\frac{F_{,b}}{Ff_{,b}} + 2\tan f\right) + g^{ab}\sec^2 f f_{,b}\left(\frac{F_{,b}}{Ff_{,b}} + \tan f\right)_{,a} = 0, \quad (2.18)$$

which in view of Eq. (2.16) yields

$$g^{ab}(\tan f)_{,b}(F_{,b}/Ff_{,b} + \tan f)_{,a} = 0. \quad (2.19)$$

Since, by hypothesis, ψ is not a constant, Eq. (2.19) implies that

$$F_{,b}/Ff_{,b} + \tan f = n, \quad (2.20)$$

where n is an arbitrary constant. Relation (2.20) shows that F and f are functionally related or equivalently, in view of (2.6) and (2.15), g_{00} and ψ are functionally related. This result, which appears as an assumption in the work of Dutta Choudhury and Banerjee[6] (1980), is thus found to be a natural consequence of the field equations. By virtue of this result one may write $F = F(f)$ and express Eq. (2.20) as

$$[\ln(F\sec f)]' = n, \quad (2.21)$$

where a prime indicates differentiation with respect to f. From Eq. (2.21) one immediately gets

$$F\sec f = A_1 e^{nf}, \quad (2.22)$$

where A_1 is an arbitrary constant of integration. Finally, in view of (2.6), (2.15), and (2.2), one obtains

$$g_{00}\psi = Ae^{2n\tan^{-1}(\sqrt{\psi-1})}, \quad A = \text{const} \quad (2.23)$$

which is the explicit relation between g_{00} and ψ previously obtained by Dutta Choudhury and Banerjee.[6]

Next, we consider the corresponding case in the Brans–Dicke theory, where ω is, however, a constant. In a vacuum, Eqs. (2.2) and (2.12) reduce, respectively, to

$$\Box\psi = 0, \quad (2.24)$$

$$F^{-1}\nabla(F\psi) = 0, \quad (2.25)$$

where ψ now refers to the Brans–Dicke scalar. Again, Eqs. (2.24) and (2.25) may also be written as

$$(1/\sqrt{-g})\left(g^{ab}\sqrt{-g^*}F\psi_{,b}\right)_{,a} = 0 \quad (2.26)$$

and

$$(1/\sqrt{-g})\left[g^{ab}\sqrt{-g^*}(F\psi)_{,b}\right]_{,a} = 0, \quad (2.27)$$

respectively.

From Eqs. (2.26) and (2.27), one obtains

$$g^{ab}\psi_{,b}(F_{,b}\psi/F\psi_{,b})_{,a} = 0, \quad (2.28)$$

which in turn implies that

$$F_{,b}\psi/F\psi_{,b} = \text{const}, \quad (2.29)$$

since ψ cannot be a constant by hypothesis. Relation (2.29) shows that $g_{00}(=F^2)$ and ψ are functionally related. This is the promised result. With Eq. (2.29) at hand, it is a simple matter to prove that the functional relationship is of the form

$$g_{00} = C_1\psi^{C_2}, \quad (2.30)$$

where C_1 and C_2 are arbitrary constants. The relation (2.30) in the Brans–Dicke theory was obtained earlier by Banerjee and Bhattacharya.[5] However, it may be noted that the proof of relation (2.30) given here follows directly from the field equations and does not presuppose functional dependence of g_{00} on ψ.

3. STATIC SPHERICALLY SYMMETRIC FIELD ABOUT A POINT MASS IN BARKER'S THEORY

We consider the line element in the isotropic form,

$$ds^2 = e^{\nu}dt^2 - e^{\mu}(dr^2 + r^2d\theta^2 + r^2\sin^2\theta d\phi^2), \quad (3.1)$$

where μ and ν are functions of r alone. Then with $\omega = (4-3\psi)/(2\psi-2)$ the field equations (2.1) and (2.2) take the form

$$\frac{\mu'^2}{4} + \frac{\mu'\nu'}{2} + \frac{\mu'\nu'}{r} = \frac{(4-3\psi)\psi'^2}{4(\psi-1)\psi^2} + \frac{\psi''}{\psi} - \frac{\mu'\psi'}{2\psi} - \frac{\psi'^2}{2\psi(\psi-1)}, \quad (3.2)$$

$$\frac{\mu''}{2} + \frac{\nu''}{2} + \frac{\nu'^2}{4} + \frac{\mu'+\nu'}{2r}$$

$$= -\frac{(4-3\psi)\psi'^2}{4(\psi-1)\psi^2} + \frac{\psi'}{r\psi} + \frac{\mu'\psi'}{2\psi} - \frac{\psi'^2}{2\psi(\psi-1)}, \tag{3.3}$$

$$\mu'' + \frac{\mu'^2}{4} + \frac{2\mu'}{r} = -\frac{(4-3\psi)\psi'^2}{4(\psi-1)\psi^2} + \frac{\nu'\psi'}{2\psi} - \frac{\psi'^2}{2\psi(\psi-1)}, \tag{3.4}$$

$$\frac{2\psi'}{r} + \frac{(\mu'+\nu')\psi'}{2} + \psi'' = \frac{\psi'^2}{2(\psi-1)}, \tag{3.5}$$

where ψ is a function of r and a prime indicates differentiation with respect to r.

Equation (3.5), when $\psi' \neq 0$, can be written in the form

$$\frac{2}{r} + \frac{(\mu'+\nu')}{2} + \frac{\psi''}{\psi'} = \frac{\psi'}{2(\psi-1)}, \tag{3.6}$$

which in turn yields on integration

$$r^2 e^{(\mu+\nu)/2}\psi' = a(\psi-1)^{1/2}, \tag{3.7}$$

where a is a arbitrary constant of integration.

Again, the operation 2 × Eq. (3.3) + Eq. (3.2) − Eq. (3.4), in view of Eq. (3.5), yields

$$\nu'' + \frac{\nu'\psi'}{\psi} + \frac{\psi''}{\psi} + \left(\nu' + \frac{\psi'}{\psi}\right)\left(\frac{\nu'}{2} + \frac{\nu'}{2} + \frac{2}{r}\right) = 0, \tag{3.8}$$

which on integration gives

$$r^2 e^{(\mu+\nu)/2}(\nu'\psi + \psi') = b, \tag{3.9}$$

where b is another arbitrary constant.

By means of Eqs. (3.7) and (3.9) one easily gets

$$\nu' + \frac{\psi'}{\psi} = \left(\frac{b}{a}\right)\frac{\psi'}{\psi(\sqrt{\psi-})}. \tag{3.10}$$

The integration of Eq. (3.10) gives

$$e^{\nu}\psi = Ae^{2(b/a)\tan^{-1}(\sqrt{\psi-1})}, \tag{3.11}$$

where A is an arbitrary constant. Although we are considering the spherically symmetric case, relation (3.11), which is already derived in Sec. 2 under general consideration, holds irrespective of any symmetry condition.

Again, in view of relation (3.11), Eq. (3.7) yields

$$e^{\mu}\psi = \frac{a^2}{4Ar^4}\left[(\tan^{-1}\sqrt{\psi-1})'\right]^{-2} e^{-2(b/a)\tan^{-1}(\sqrt{\psi-1})}, \tag{3.12}$$

which is another useful relation in this case.

Now, defining a new variable $f(\psi)$ by means of

$$\tan^{-1}\sqrt{\psi-1} = f \tag{3.13}$$

and substituting for the derivative of ν and μ in Eqs. (3.2)–(3.5) their corresponding expressions obtained from (3.11) and (3.12), it follows that Eqs. (3.2)–(3.4) reduce to

$$\frac{f''^2}{f'^2} + \frac{2}{r}\frac{f''}{f'} - \left(1+\frac{b^2}{a^2}\right)f'^2 = 0, \tag{3.14}$$

$$-\frac{f'''}{f'} + \frac{f''^2}{f'^2} - \frac{1}{r}\frac{f''}{f'} + \left(1+\frac{b^2}{a^2}\right)f'^2 = 0, \tag{3.15}$$

$$-\frac{2f'''}{f'} + \frac{3f''^2}{f'^2} + \left(1+\frac{b^2}{a^2}\right)f'^2 = 0, \tag{3.16}$$

respectively, whereas Eq. (3.5) is identically satisfied. Again, Eqs. (3.15) and (3.16) together can be reduced to Eq. (3.14). This may be easily verified by multiplying Eq. (3.15) throughout by 2 and then subtracting the resultant equation from Eq. (3.16). Thus, by virtue of relations (3.11) and (3.12), integration of the field equations (3.2)–(3.5) is reduced to the task of solving the differential equation (3.14).

By the substitution $\phi = rf'$, (3.17)

Eq. (3.14) may be expressed in the form

$$(r\phi'/\phi)^2 = 1 + m^2\phi^2, \tag{3.18}$$

where $m^2 = 1 + b^2/a^2$. (3.19)

Equation (3.18) in turn yields

$$\frac{d\phi}{\phi(1+m^2\phi^2)^{1/2}} = \pm\frac{dr}{r}. \tag{3.20}$$

When the positive sign in Eq. (3.20) is taken, the solution is

$$\phi = 2Cr/(m^2 - C^2r^2), \tag{3.21}$$

where C is an arbitrary constant. The negative sign gives

$$\phi = 2Cr/(m^2r^2 - C^2). \tag{3.22}$$

When the expression for ϕ in (3.21) is substituted in (3.17), one obtains after integration

$$f = \ln\left[B\left(\frac{m+Cr}{m-Cr}\right)^{1/m}\right], \tag{3.23}$$

where B is another arbitrary constant.

Again, when the expression given by (3.22) is used in (3.17) one obtains

$$f = \ln\left[B\left(\frac{mr-C}{mr+C}\right)^{1/m}\right]. \tag{3.24}$$

Both families of solutions satisfy Eq. (3.14). Accordingly, we also have two families of solutions of the field equations (3.2)–(3.5), which are now obtained from (3.11)–(3.3) by means of (3.23) and (3.24), respectively. These are as follows.

$$e^{\nu} = AB^{2(b/a)}\left(\frac{m+Cr}{m-Cr}\right)^{2b/ma}\cos^2\ln\left[B\left(\frac{m+Cr}{m-Cr}\right)^{1/m}\right], \tag{3.25}$$

$$e^{\mu} = A^{-1}B^{-2(b/a)}\frac{a^2(m^2-C^2r^2)^2}{16C^2r^4}\cdot\left(\frac{m+Cr}{m-Cr}\right)^{-2b/ma} \times\cos^2\ln\left[B\left(\frac{m+Cr}{m-Cr}\right)^{1/m}\right], \tag{3.26}$$

$$\psi = \sec^2\ln\left[B\left(\frac{m+Cr}{m-Cr}\right)^{1/m}\right]. \tag{3.27}$$

$$e^{\nu} = AB^{2(b/a)}\left(\frac{mr-C}{mr+C}\right)^{2b/ma}\cos^2\ln\left[B\left(\frac{mr-C}{mr+C}\right)^{1/m}\right], \tag{3.28}$$

$$e^{\mu} = A^{-1}B^{-2(b/a)}\cdot\frac{a^2(m^2r^2-C^2)^2}{16C^2r^4}\cdot\left(\frac{mr-C}{mr+C}\right)^{-2b/ma} \times\cos^2\ln\left[B\left(\frac{mr-C}{mr+C}\right)^{1/m}\right], \tag{3.29}$$

$$\psi = \sec^2\ln\left[B\left(\frac{mr-C}{mr+C}\right)^{1/m}\right]. \tag{3.30}$$

ACKNOWLEDGMENT

The authors wish to thank Dr. A. Banerjee for his suggestions.

[1]K. Nordtvedt, Astrophys. J. **161**, 1059 (1970).
[2]P. Jordan, Z. Phys. **157**, 112 (1959).
[3]C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).
[4]B. M. Barker, Astrophys. J. **219**, 5 (1978).
[5]A. Banerjee and D. Bhattacharya, J. Math. Phys. **20**, 1908 (1979).
[6]A. Banerjee and S. B. Dutta Choudhury, J. Math. Phys. **21**, 2799 (1980).
[7]H. A. Buchdahl, Gen. Relativ. Gravit. **4**, 319 (1973).

Path integrals and stationary phase approximation for quantum dynamical semigroups. Quadratic systems [a]

Robert Alicki [b]
Sektion Physik der Universität München Theoretische Physik, 8000 München 2, Theresienstr. 37, Federal Republic of Germany

(Received 7 November 1980; accepted for publication 11 December 1981)

The solution of the Markovian master equation for the quantum open system of n degrees of freedom is formally written in terms of a path integral and the stationary phase approximation is discussed. The exactly soluble models with generators quadratic in position and momentum operators are investigated and the explicit expressions for the space–time propagators for one-dimensional systems are derived.

PACS numbers: 05.30. – d

1. INTRODUCTION

In recent years many problems in nonequilibrium statistical mechanics were formulated and solved using the quantum Markovian master equations.[1–6] It is rather commonly believed that the dynamics of a quantum open system interacting with the environment can be approximated by the so-called completely positive dynamical semigroup under the assumption that the relaxation times of the correlation functions of the reservoir are much shorter than the natural time scale for the open system.

The dynamical semigroup $\{\exp tL,\ t\geqslant 0\}$ fulfills the Markovian master equation

$$\frac{d\rho(t)}{dt} = L\rho(t), \tag{1.1}$$

where $\rho(t)$ is a density matrix of the open system and L is the generator of the semigroup.

Lindblad proved[4] that the general form of the bounded generator L is the following:

$$L\rho = \frac{1}{i\hbar}[H,\rho] + \frac{\lambda}{2\hbar}\sum_{\alpha}\{[V_\alpha\rho, V_\alpha^*] + [V_\alpha, \rho V_\alpha^*]\}, \tag{1.2}$$

where $H = H^*$ and V_α, $\Sigma_\alpha V_\alpha^* V_\alpha$ are bounded operators on the Hilbert space $\mathscr{H}$ of the system.

This form is assumed to be valid (with H, V_α generally unbounded) also for the unbounded generators for the physically interesting examples.[2,5,6]

The $\hbar$ dependence of the dissipative part in (1.2) which is important for the semiclassical expansion[7] may be established by taking into account the method of derivation of (1.2). Namely, the origin of this part is a double commutator with respect to the interaction Hamiltonian[1,3] (it gives $1/\hbar^2$) and elimination of bath's variables gives the factors of the following form:

$$\int_0^\infty e^{-i\omega_\alpha t/\hbar}\langle A_\alpha(t)A_\alpha\rangle\, dt, \tag{1.3}$$

where

$$A_\alpha(t) = \exp[(i/\hbar)H_R t]\ A_\alpha \exp[-(i/\hbar)H_R t],$$

A_α—bath's operators, H_R—Hamiltonian of the bath, and $\langle\cdots\rangle$—mean value with respect to the bath's state. The integral (1.3) can be formally written as

$$\hbar\left\langle\left\{\frac{i}{\omega_\alpha - [H_R,\cdot]}A_\alpha\right\}A_\alpha\right\rangle, \tag{1.4}$$

and therefore we finally obtain the $1/\hbar$ factor.

Consider an open quantum system of n degrees of freedom. In this case we assume that the operators H, V_α, V_α^* in (1.2) are generally unbounded functions of position and momentum operators $\{\hat{q}_k, \hat{p}_k\}$ fulfilling the conditions of Ref. 6.

The solution of the equation of motion is described by the space–time propagator $\Lambda_t(q'', \bar{q}''|q', \bar{q}')$ in the Liouville space which is defined by the requirement that the density matrix of the system $\rho_t(q, \bar{q})$ evolves in time according to the integral transformation

$$\rho_t(q'', \bar{q}'') = \int \Lambda_t(q'', \bar{q}''|q', \bar{q}') \times \rho_0(q', \bar{q}')dq'd\bar{q}', \tag{1.5}$$

where $q = (q_1, q_2, \ldots, q_n)$, $\bar{q} = (\bar{q}_1, \bar{q}_2, \ldots, \bar{q}_n)$. By the semigroup property one can formally express the propagator Λ_t in terms of the path integral on the product phase space $\Gamma \times \bar{\Gamma} = \{(q_k, \bar{q}_k; p_k, \bar{p}_k), k = 1, 2, \ldots, n\}$. Such a formulation is presented in Sec. 2 in a formal way and we do not discuss in detail the problems concerning the discretization and limiting procedure in the definition of the path integral because they are similar to those in ordinary quantum mechanics.[8–11]

We are able to calculate the path integral for a few systems only, so various approximative methods are invoked. We apply here the method of the stationary phase approximation[9,11–15] which in our case of the dissipative quantum system leads to the notion of the complex extremal trajectories in the ficticious complex product phase-space (Sec. 3). These trajectories are the solutions of the corresponding "semiclassical" ordinary differential equations.

In some applications the Hamiltonian of the system can be approximated by the quadratic form and the dissipation effects are "linear," which means that the operators $\{V_\alpha\}$ in

[a] Work supported in parts by the Alexander von-Humboldt-Stiftung and the Polish Ministry of Higher Education, Science and Technology, project MRI. 7.

[b] Humboldt fellow, on leave of absence from University of Gdańsk, Institute of Theoretical Physics, Gdańsk, Poland.

 0022-2488/82/071370-06$02.50

(1.2) are linear functions of $\{\hat{p}, \hat{q}\}$. For such "quadratic" open systems the path integrals are Gaussian and the stationary phase approximation gives an exact result, so the problem of evaluation of the propagator Λ_t reduces to a solution of ordinary linear nonhomogeneous differential equations (Sec. 4). We solve such equations in the case of one degree of freedom and obtain the explicit expression for the propagator for the harmonic oscillator with damping and pumping (Sec. 5) and the Brownian particle in a constant field strength (Sec. 6). The similar one-dimensional models were already studied[2,5] in the Heisenberg picture and the dynamics were described by the expressions for the time evolution of Weyl operators $W(x, y) = \exp i(x\hat{p} + y\hat{q})$, but the manifest form of the propagators was not known.

Although the investigation of quadratic systems can be done without path integrals and stationary phase approximation we treat the presented results as a first step towards the investigation of more complicated nonlinear dissipative quantum systems. A nonlinear toy model is studied in Ref. 7.

2. PATH INTEGRALS IN LIOUVILLE SPACE

We study the quantum physical system described by the Hilbert space $\mathscr{H}$. It is convenient to introduce the so-called Liouville space[16] which is the Hilbert space $\mathscr{L}^2(\mathscr{H})$ of all Hilbert–Schmidt operators acting on $\mathscr{H}$ with the following scalar product:

$$\langle A \,|\, B \rangle = \mathrm{tr}(A^* B), \quad A, B \in \mathscr{L}^2(\mathscr{H}). \tag{2.1}$$

Liouville space contains both the mixed states of the system (density matrices) and those observables which are represented by the Hilbert–Schmidt operators.

Let us consider the quantum system S of n degrees of freedom with a configuration space $\mathbb{R}^n$. The Hilbert space is isomorphic to $L^2(\mathbb{R}^n)$ and the Liouville space as a space of square integrable integral kernels is isomorphic to $L^2(\mathbb{R}^{2n})$. We introduce the following notation and definitions.

$\rho(q|\bar{q})$—density matrix or alternatively the arbitrary vector of the Liouville space $L^2(\mathbb{R}^{2n})$, (2.2)

$Q_k, P_k, \bar{Q}_k, \bar{P}_k$, $k = 1,2,\ldots,n$—self-adjoint operators on Liouville space defined as

$$\begin{aligned}
&(Q_k \rho)(q, \bar{q}) = q_k \rho(q, \bar{q}), \\
&(P_k \rho)(q, \bar{q}) = -i\hbar \frac{\partial}{\partial q_k} \rho(q, \bar{q}), \\
&(\bar{Q}_k \rho)(q, \bar{q}) = \bar{q}_k \rho(q, \bar{q}), \\
&(\bar{P}_k \rho)(q, \bar{q}) = i\hbar \frac{\partial}{\partial \bar{q}_k} \rho(q, \bar{q}),
\end{aligned} \tag{2.3}$$

$|q, \bar{q}\rangle$, $|p, \bar{p}\rangle$—improper eigenvectors of $(Q, \bar{Q})$, $(P, \bar{P})$. Equation (1.2) for $H \equiv H(\hat{p}, \hat{q})$ and $V_\alpha \equiv V_\alpha(\hat{p}, \hat{q})$ can be written as a differential equation for the density matrix $\rho_t(q, \bar{q})$:

$$\frac{\partial}{\partial t} \rho_t(q, \bar{q}) = L\left(q, \frac{\partial}{\partial q}, \bar{q}, \frac{\partial}{\partial \bar{q}}\right) \rho_t(q, \bar{q}). \tag{2.4}$$

Let $A \equiv A(\hat{p}, \hat{q})$ and $B \equiv B(\hat{p}, \hat{q})$, then

$$(A \cdot \rho)(q, \bar{q}) = A\left(-i\hbar \frac{\partial}{\partial q}, q\right) \rho(q, \bar{q}) = A(P, Q) \rho(q, \bar{q}), \tag{2.5}$$

$$(\rho \cdot B^*)(q, \bar{q}) = \bar{B}\left(i\hbar \frac{\partial}{\partial \bar{q}}, \bar{q}\right) \rho(q, \bar{q}) = \bar{B}(\bar{P}, \bar{Q}) \rho(q, \bar{q}), \tag{2.6}$$

where if $B(\tilde{p}, \tilde{q}) = \Sigma a_{mn} \tilde{p}^m \tilde{q}^n$ then $\bar{B}(\tilde{p}, \tilde{q}) = \Sigma \bar{a}_{mn} \tilde{p}^m \tilde{q}^n$ with $\bar{a}_{mn}$ the complex conjugate of a_{mn} and $\{\tilde{p}, \tilde{q}\}$ denotes here $\{\hat{p}, \hat{q}\}$, $\{P, Q\}$, or $\{\bar{P}, \bar{Q}\}$. One can write Eq. (2.4) in a following form:

$$\frac{\partial \rho_t(q, \bar{q})}{\partial t} = -\frac{i}{\hbar} \hat{\mathscr{L}}(Q, P, \bar{Q}, \bar{P}) \rho_t(q, \bar{q}), \tag{2.7}$$

where

$$\hat{\mathscr{L}}(Q, P, \bar{Q}, \bar{P}) \doteq \{H(P, Q) - \bar{H}(\bar{P}, \bar{Q})\} + i\lambda \sum_\alpha \{2V_\alpha(P, Q) \bar{V}_\alpha(\bar{P}, \bar{Q}) - V_\alpha^*(P, Q) V_\alpha(P, Q) - \bar{V}_\alpha^*(\bar{P}, \bar{Q}) \bar{V}_\alpha(\bar{P}, \bar{Q})\}. \tag{2.8}$$

Following the analogy between standard Hilbert space (Schrödinger equation) and Liouville space [Eq. (2.4)] one can represent the propagator $\Lambda_t(q'', \bar{q}'' | q', \bar{q}') = \langle q'', \bar{q}'' | \exp\{-(i/\hbar) \hat{\mathscr{L}} t\} | q', \bar{q}' \rangle$ as the formal path integral,[9–15]

$$\Lambda_t(q'', \bar{q}'' | q', \bar{q}') = \int_{q', \bar{q}'}^{q'', \bar{q}''} Dq D\bar{q} Dp D\bar{p} \exp\left\{\frac{i}{\hbar} \int_0^t d\tau [p\dot{q} - \bar{p}\dot{\bar{q}} - \mathscr{L}^\gamma]\right\}, \tag{2.9}$$

where $\mathscr{L}^\gamma \equiv \mathscr{L}^\gamma(q, p, \bar{q}, \bar{p})$ depends on the limiting procedure[12–14] or discretization γ and differs in general from the "classical Liouville function"

$$\begin{aligned}
\mathscr{L}_{\mathrm{cl}}(q, p, \bar{q}, \bar{p}) &= \hat{\mathscr{L}}(Q \to q, P \to p, \bar{Q} \to \bar{q}, \bar{P} \to \bar{p})|_{\hbar = 0} \\
&= H_{\mathrm{cl}}(p, q) - H_{\mathrm{cl}}(\bar{p}, \bar{q}) + i\lambda \sum_\alpha \{2V_\alpha(p, q) \bar{V}_\alpha(\bar{p}, \bar{q}) - |V_\alpha(p, q)|^2 - |V_\alpha(\bar{p}, \bar{q})|^2\}
\end{aligned} \tag{2.10}$$

by terms of order $\hbar$.

For example, if $H(\hat{p}, \hat{q})$ and $V_\alpha(\hat{p}, \hat{q})$ are given in antistandard ordering ($\hat{p}$ at the left-hand side of $\hat{q}$) then the path integral (2.9) can be defined as a formal limit

$$\Lambda_t(q'',\bar q''|q',\bar q') = \lim_{N\to\infty}\int\prod_{j=1}^{N} dq^{(j)}d\bar q^{(j)}\prod_{j=1}^{N+1}\frac{dp^{(j)}}{(2\pi\hbar)^n}\frac{d\bar p^{(j)}}{(2\pi\hbar)^n}\exp\left\{\frac{i\epsilon}{\hbar}\sum_{j=1}^{N+1}\left[p^{(j)}\frac{q^{(j)}-q^{(j-1)}}{\epsilon} - \bar p^{(j)}\frac{\bar q^{(j)}-\bar q^{(j-1)}}{\epsilon} - H(p^{(j)},q^{(j)})\right.\right.$$
$$\left.\left. + H(\bar p^{(j)},\bar q^{(j)}) + i\lambda\sum_\alpha\left(2V_\alpha(p^{(j)},q^{(j-1)})\bar V_\alpha(\bar p^{(j)},\bar q^{(j-1)}) - \bar V_\alpha(p^{(j)},q^{(j)})V_\alpha(p^{(j)},q^{(j-1)}) - V_\alpha(\bar p^{(j)},\bar q^{(j)})\bar V_\alpha(\bar p^{(j)},\bar q^{(j-1)})\right)\right]\right\},$$
$$q^{(0)}\equiv q',\quad \bar q^{(0)}\equiv\bar q',\quad q^{(N+1)}\equiv q'',\quad \bar q^{(N+1)}\equiv\bar q'',\quad \epsilon=\frac{t}{N+1}. \tag{2.11}$$

The existence of the limit (2.11) is of course a major problem also for the standard path integrals in quantum mechanics so we only note that the path integral (2.9), because of an imaginary part in the "classical action," possesses the properties of both Feynman and Wiener integrals. This imaginary contribution may regularize the oscillatory behavior of the Feynman integral but a more rigorous treatment remains to be done.

3. STATIONARY PHASE APPROXIMATION

Very few path integrals can be evaluated exactly so we need various approximate methods. One such approach is the so-called stationary-phase or semiclassical approximation.[8,9,11–14] In this approach the propagator Λ_t is approximated by the expression Λ_t^{class}, which can be treated as a leading term in the formal expansion of Λ_t with respect to the "small parameter" $\hbar$. The formal but systematic approach to this problem is given in Refs. 11–14. Summarizing these results we can write for the case of Eq. (2.7)

$$\Lambda_t^{\text{class}}(q'',\bar q''|q',\bar q') = \mathcal{N}_t(q'',\bar q''|q',\bar q')\exp\left\{\frac{i}{\hbar}W_{\text{extr}}(t)\right\}, \tag{3.1}$$

where W_{extr} is a stationary value of the generalized action

$$W(t) = \int_0^t\left[(p\dot q - \bar p\dot{\bar q}) - \mathcal{L}_{\text{cl}}(q,p,\bar q,\bar p)\right]dt' \tag{3.2}$$

which fulfills the equation $\delta W(t)=0$ and is given by substituting in (3.2) the solution of the following equations[15,7]:

$$\dot q_k = \frac{\partial\mathcal{L}_{\text{cl}}}{\partial p_k},\quad \dot p_k = -\frac{\partial\mathcal{L}_{\text{cl}}}{\partial q_k},$$
$$\dot{\bar q}_k = -\frac{\partial\mathcal{L}_{\text{cl}}}{\partial\bar p_k},\quad \dot{\bar p}_k = \frac{\partial\mathcal{L}_{\text{cl}}}{\partial\bar q_k} \tag{3.3}$$

with the boundary conditions $q(0)=q'$, $\bar q(0)=\bar q'$, $q(t)=q''$, $\bar q(t)=\bar q''$.

The pre-exponential factor $\mathcal{N}_t$ contains the well-known Van Vleck's determinant[11] and some extra terms also (see detailed discussion in Refs. 12–14). The construction (3.1) relies on the existence of a unique classical solution of (3.3). Here the classical solution is complex and not unique. The similar problem was studied by Knoll and Schaeffer[17] in the case of the complex potential[18] and they have met the phenomenon that whereas for multiple real classical paths a simple summation has to be performed[11] for multiple complex classical paths the set of paths that contribute has to be determined separately. We expect here that for a large class of systems at least in the case of weak dissipation (small λ) the main contribution is given by the unique path with finite action in the limit $\lambda\to 0$.[7] The application of the full formulas (3.2) to the simple toy model and the discussion of the classical solutions are presented in Ref. 7. In the present paper we study the quadratic systems for which all difficulties disappear: the path integrals are Gaussian and therefore can be well defined; the stationary phase approximation is exact and the classical solution is unique, the pre-exponential factor $\mathcal{N}_t$ depends only on t and can be easily calculated using the trace preserving property

$$\int dq\,\Lambda_t(q,q|q',\bar q') = \delta(q'-\bar q'). \tag{3.4}$$

However, even in this case the formulas (3.1) provide the simplest method to derive the manifest form of the propagator.

4. OPEN SYSTEMS WITH QUADRATIC GENERATORS

Consider the quantum mechanical system of n degrees of freedom, the dynamics of which is described by the dynamical semigroup (1.2) with quadratic Hamiltonian (we use summation convention)

$$H(\hat p,\hat q) = \tfrac12 a_{kl}\hat p_k\hat p_l + \tfrac12 u_{kl}\hat q_k\hat q_l + \tfrac12\lambda_{kl}(\hat p_k\hat q_l + \hat q_l\hat p_k) + \sigma_k\hat p_k + \kappa_l\hat q_l \tag{4.1}$$

and operators $\{V_\alpha\}$ linear in $\hat p_k$, $\hat q_k$,

$$V_\alpha = a_k^\alpha\hat p_k + b_k^\alpha\hat q_k. \tag{4.2}$$

The Liouville function (up to some irrelevant constant) is given by the following expression ($\lambda\equiv 1$):

$$\mathcal{L}(p,q|\bar p,\bar q) = H(p,q) - H(\bar p,\bar q) - \frac{i}{2}\{A_{kl}(p_kp_l + \bar p_k\bar p_l - 2p_k\bar p_l) + B_{kl}(q_kq_l + \bar q_k\bar q_l - 2q_k\bar q_l) - 2iM_{kl}(\bar p_kq_l - p_k\bar q_l) + 2K_{kl}(p_kq_l + \bar p_k\bar q_l - p_k\bar q_l - \bar p_kq_l)\}, \tag{4.3}$$

Here

$$A_{kl} = \bar A_{lk} = \sum_\alpha a_k^\alpha\bar a_l^\alpha,\quad B_{kl} = \bar B_{lk} = \sum_\alpha b_k^\alpha\bar b_l^\alpha,$$
$$K_{kl} = \bar K_{kl} = \operatorname{Re}\sum_\alpha\bar a_k^\alpha b_l^\alpha,\quad M_{kl} = \bar M_{kl} = \operatorname{Im}\sum_\alpha\bar a_k^\alpha b_l^\alpha. \tag{4.4}$$

In order to evaluate the propagator for the quadratic system one must solve the "semiclassical" equations (3.3). Introducing the more convenient new variables

$$\xi_k = (q_k + \bar q_k),\quad \eta_k = (q_k - \bar q_k),$$
$$\pi_k = (p_k + \bar p_k),\quad \theta_k = (p_k - \bar p_k), \tag{4.5}$$

one can obtain the following equations:

$$\dot{\eta}_k = (\lambda_{kl} - M_{kl})\eta_l + (a_{kl} + iA^{a}_{kl})\theta_l, \tag{4.6}$$

$$\begin{aligned}
\dot{\theta}_k &= -(u_{kl} + iB^{a}_{kl})\eta_l - (\lambda_{kl} + M_{kl})\theta_l, \\
\dot{\xi}_k &= (\lambda_{kl} + M_{kl})\xi_l - 2iK_{kl}\eta_l \\
&\quad + (a_{kl} - iA^{a}_{kl})\pi_l - 2iA^{s}_{kl}\theta_l + 2\sigma_k,
\end{aligned} \tag{4.7}$$

$$\begin{aligned}
\dot{\pi}_k &= -(u_{kl} - iB^{a}_{kl})\xi_l + 2iB_{kl}\eta_l \\
&\quad - (\lambda_{kl} - M_{kl})\pi_l + 2iK_{kl}\theta_l - 2\kappa_k.
\end{aligned}$$

Here B^{a} (or A^{s}) denotes the antisymmetric (or symmetric) part of the matrix.

We prove in the Appendix that the extremal value of the generalized action $W_{\text{extr}}(t)$ is given by a very simple expression,

$$\begin{aligned}
W_{\text{extr}}(t) &= \tfrac{1}{2}\{\xi_k(t)\theta_k(t) - \xi_k(0)\theta_k(0) \\
&\quad + \eta_k(t)\pi_k(t) - \eta_k(0)\pi_k(0)\} \\
&\quad - \tfrac{1}{2}\sigma_k \int_0^t \theta_k(\tau)d\tau - \tfrac{1}{2}\kappa_k \int_0^t \eta_k(\tau)d\tau.
\end{aligned} \tag{4.8}$$

5. QUANTUM HARMONIC OSCILLATOR WITH PUMPING AND DAMPING

We apply the results of Sec. 4 to a system of one degree of freedom (harmonic oscillator). The special examples of such systems (with some simple choice of parameters) were studied in several papers using the operator formalism and the Heisenberg picture.[2,5]

The Liouville function for our model is given by

$$\begin{aligned}
\mathscr{L}(p, q|\bar{p}, \bar{q}) &= H(p, q) - H(\bar{p}, \bar{q}) \\
&\quad - (i/2)\{a^2(p - \bar{p})^2 + b^2(q - \bar{q})^2 \\
&\quad + 2k(q - \bar{q})(p - \bar{p}) - i2\mu(\bar{p}q - p\bar{q})\},
\end{aligned} \tag{5.1}$$

where

$$\begin{aligned}
H(p, q) &= (1/2m)p^2 + \tfrac{1}{2}m\omega_0^2 q^2 + \lambda pq \\
&\quad + \sigma p + \kappa q,
\end{aligned} \tag{5.2}$$

$0 \leqslant \omega_0$, $0 \leqslant m, \lambda, \sigma, \kappa, a, b, k, \mu \in \mathbb{R}$, and the matrix

$$\begin{bmatrix} a^2 & k + i\mu \\ k - i\mu, & b^2 \end{bmatrix}$$

is positive definite. Therefore, the semiclassical equations are the following.

$$\begin{aligned}
\dot{\eta} &= (\lambda - \mu)\eta + (1/m)\theta, \\
\dot{\theta} &= -m\omega_0^2\eta - (\lambda + \mu)\theta,
\end{aligned} \tag{5.3}$$

$$\begin{aligned}
\dot{\xi} &= (\lambda + \mu)\xi + (1/m)\pi + 2ik\eta - 2ia^2\theta + 2\sigma, \\
\dot{\pi} &= -m\omega_0^2\xi - (\lambda - \mu)\pi + 2ib^2\eta + 2ik\theta - 2\kappa.
\end{aligned}$$

One can solve these equations and put the solutions into the expression for the propagator (3.1). Using (3.1), (4.7), and (3.4) we obtain the following final result for the underdamped case ($\omega_0^2 > \lambda^2$):

$$\Lambda_t(q'', \bar{q}''|q', \bar{q}') = (m\omega e^{-\mu t}/2\pi|\sin\omega t|)\exp\left\{\frac{i}{\hbar}W(t)\right\}, \tag{5.4}$$

where

$$\begin{aligned}
W(t) &= (m\omega_0/2\sin\omega t)\{[(q'' - z)^2 - (\bar{q}'' - z)^2]\cos(\omega t + \varphi) + [(q' - z)^2 - (\bar{q}' - z)^2]\cos(\omega t - \varphi) - (q' + \bar{q}' - 2z)(q'' - \bar{q}'') \\
&\quad \times e^{\mu t}\cos\varphi - (q'' - \bar{q}'' - 2z)(q' - \bar{q}')e^{-\mu t}\cos\varphi\} - (im/4\sin^2\omega t) \\
&\quad \times \{A(t; \mu, \varphi, \chi, \psi)(q'' - \bar{q}'')^2 - A(t; -\mu, -\varphi, -\chi, -\psi)(q' - \bar{q}')^2 + B(t)(q'' - \bar{q}'')(q' - \bar{q}')\}
\end{aligned} \tag{5.5}$$

and

$$z = \frac{[\sigma(\lambda - \mu) - \kappa/m]}{\omega^2 + \mu^2}, \tag{5.6}$$

$$A(t, \mu, \varphi, \chi, \psi) = \{\Gamma\cos(\omega t + \varphi)\cos(\omega t + \chi) - \Gamma e^{2\mu t}\cos\varphi\cos\chi - D\cos(\omega t + \psi)\sin\omega t\}$$

$$\begin{aligned}
B(t) &= \{\Gamma e^{\mu t}\cos\varphi\cos(\omega t - \chi) - \Gamma e^{-\mu t}\cos\varphi\cos(\omega t + \chi) + \Gamma e^{\mu t}\cos\chi\cos(\omega t - \varphi) - \Gamma e^{-\mu t}\cos\chi\cos(\omega t + \varphi) \\
&\quad + D(e^{\mu t} + e^{-\mu t})\cos\psi\sin\omega t\}.
\end{aligned}$$

Here

$$\lambda = \omega_0\sin\varphi, \quad \omega = \omega_0\cos\varphi, \quad \varphi \in [0, \pi/2],$$

$$\begin{aligned}
\Gamma\cos\chi &= \frac{1}{(\omega^2 + \mu^2)\mu}\left\{k\omega(\mu - \lambda) + ma^2\left[\frac{\omega_0^2\omega}{2} + \mu\omega(\mu - \lambda)\right] + \frac{b^2}{2m}\omega\right\}, \\
\Gamma\sin\chi &= \frac{1}{\omega^2 + \mu^2}\left\{k(\lambda - \mu) + ma^2\left[(\omega^2 + \lambda\mu) - \frac{\omega_0^2}{2}\right] - \frac{b^2}{2m}\right\},
\end{aligned} \tag{5.7}$$

$$\begin{aligned}
D\cos\psi &= \frac{1}{(\omega^2 + \mu^2)\mu}\left\{k\frac{\lambda^2 - \mu^2}{\omega_0} - \frac{ma^2}{2}\omega_0\omega(\lambda - \mu) - \frac{b^2}{2m\omega_0}\omega(\lambda + \mu)\right\}, \\
D\sin\psi &= \frac{1}{(\omega^2 + \mu^2)\mu}\left\{-k\frac{(\lambda + \mu)(\omega^2 + \lambda\mu)}{\omega_0^2} + \frac{ma^2}{2}\omega_0(\omega^2 - \mu\lambda) + \frac{b^2}{2m\omega_0}[\omega^2 + \mu(\lambda + 2\mu)]\right\},
\end{aligned}$$

$$\Gamma, D \geqslant 0$$

As was pointed out by Lindblad,[5] if $\mu<0$ then we have the damping oscillator with the asymptotic stationary state ρ_∞. Therefore, for any initial state ρ we have

$$\lim_{t\to\infty} \Lambda_t \rho = \rho_\infty . \tag{5.8}$$

It is equivalent to the following relation for the propagator:

$$\lim_{t\to\infty} \Lambda_t(q'', \bar{q}''|q', \bar{q}') = \rho_\infty(q'', \bar{q}'')\delta(q' - \bar{q}'). \tag{5.9}$$

One can check after long but simple calculations that for our example, if $\mu<0$, $\omega_0^2 > \lambda^2$ then

$$\begin{aligned}\rho_\infty(q, \bar{q}) = \left(\frac{m\omega_0\cos\varphi}{-\hbar\pi\cos\chi}\right)^{1/2} \exp\frac{m\omega_0}{4\hbar}\Bigg\{&\frac{\cos\varphi}{\cos\chi}(q+\bar{q}-2z)^2 - \frac{1}{4\Gamma\cos\varphi\cos\chi}\\ &\times[\Gamma^2\sin^2(\varphi-\chi)+D^2\cos^2\psi+2\Gamma D(\sin(\varphi-\psi)\cos\chi+\sin(\chi-\psi)\cos\varphi)](q-\bar{q})^2\\ &+\frac{i[\sin(\varphi-\chi)+D\Gamma^{-1}\cos\psi]}{\cos\chi}(q+\bar{q}-2z)(q-\bar{q})\Bigg\}.\end{aligned} \tag{5.10}$$

The formula (5.10) describes a well-defined density matrix because the conditions $\mu<0$, $\omega_0^2>\lambda^2$ impose $\cos\varphi>0$, $\cos\chi<0$.

If $\mu>0$ we have a pumping oscillator and the energy grows to infinity. This describes the interaction with the reservoir at the negative temperature (e.g., laser).

6. QUANTUM BROWNIAN PARTICLE IN A CONSTANT FIELD STRENGTH

Consider a quantum particle in one dimension being under the influence of a translationally invariant reservoir described by the quadratic generator[5] and the constant force f. It follows that

$$\mu = -\lambda, \quad \omega_0 = 0, \quad \sigma = 0, \quad \kappa = -f \tag{6.1}$$

Moreover we assume for simplicity $k=0$. The free Brownian particle was studied by Lindblad in the Heisenberg picture.[5] We have the semiclassical equations

$$\dot{\eta} = \gamma\eta + (1/m)\theta, \quad \dot{\theta} = 0, \quad \dot{\xi} = (1/m)\pi - i2a^2\theta, \quad \dot{\pi} = -\gamma\pi + i2b^2\eta + 2f. \tag{6.2}$$

Here $\gamma = 2\lambda$.

Using the method presented in Sec. 4 we obtain the following propagator:

$$\Lambda_t(q'', \bar{q}''|q', \bar{q}') = \frac{m\gamma}{2\pi(1-e^{-\gamma t})}\exp\left\{\frac{i}{\hbar}W(t)\right\}, \tag{6.3}$$

where

$$\begin{aligned}W(t) = &\frac{m\gamma}{2(e^{\gamma t}-1)}\\ &\times\left\{[(q''-q')+(\bar{q}''-\bar{q}')][(q''-\bar{q}'')-(q'-\bar{q}')e^{\gamma t}]+\frac{2f}{m\gamma^2}[(e^{\gamma t}-1-\gamma t)(q''-\bar{q}'')+(1-e^{\gamma t}(1-\gamma t))(q'-\bar{q}')]\right\}\\ &-i[2\gamma(e^{\gamma t}-1)(1-e^{-\gamma t})]^{-1}\{[-b^2A(\gamma,t)-a^2m^2\gamma^3te^{-\gamma t}](q''-\bar{q}'')^2+[b^2A(-\gamma,t)-a^2m^2\gamma^3te^{\gamma t}](q'-\bar{q}')^2\\ &+[b^2(e^{-\gamma t}-e^{\gamma t}+2\gamma t)+2a^2m^2\gamma^3t](q''-\bar{q}'')(q'-\bar{q}')\}.\end{aligned} \tag{6.4}$$

Here

$$A(\gamma, t) = \tfrac{1}{2}(e^{-\gamma t}+e^{\gamma t}-2)+[e^{-\gamma t}(1+\gamma t)-1]. \tag{6.5}$$

These formulas can be checked by comparison with some particular simpler cases[19,20] (one can find that there are a few mistakes in a propagator for the particle in a linear stochastic potential given in the author's paper[20] which, however, do not change the physical suggestions).

In order to have a more physical picture one can evaluate the density matrix for very large t assuming that the initial state is a pure state given by a Gaussian packet with a mean position equal to zero.

For sufficiently large t and for $\gamma>0$ (damping case) we have

$$\rho_t(q, \bar{q}) \cong \mathcal{N}_t \exp\left\{-\left[\frac{m^2\gamma^2}{16\hbar\alpha t}\left(q+\bar{q}-\frac{2f}{m\gamma}t\right)^2-\frac{imb^2}{8\hbar\alpha t}\left(q+\bar{q}-\frac{2f}{m\gamma}t\right)(q-\bar{q})-\frac{if}{\gamma\hbar}(q-\bar{q})+\frac{b^2}{4\gamma\hbar}(q-\bar{q})^2\right]\right\}, \tag{6.6}$$

where $\alpha = \frac{1}{2}b^2 + \frac{1}{2}a^2m^2\gamma^2$.

The probability distribution in a position space has the classical form

$$P_t(q) = \rho_t(q, q)$$
$$= \mathcal{N}_t \exp\left\{ - \frac{m^2\gamma^2}{4\alpha\hbar t}(q - v_0 t)^2 \right\}, \tag{6.7}$$
$$v_0 = \frac{f}{m\gamma}.$$

One can transform the density matrix (6.6) to the momentum representation and obtain the probability distribution in momentum space,

$$\tilde{P}_t(p) = \mathcal{N}_0' \exp\left\{ - \frac{\gamma}{\hbar b^2}(p - p_0)^2 \right\}, \tag{6.8}$$

where $p_0 = umv_0$, $u = (1 - b^2/4\alpha)$, $\frac{1}{2} < u < 1$. Comparing with the classical results we have the diffusion constant

$$D = \hbar\alpha/m^2\gamma^2 \tag{6.9}$$

and temperature $kT = \hbar b^2/2m\gamma$.

Therefore, the classical relations

$$D = kT/m\gamma \quad \text{and} \quad u = 1$$

are never fulfilled exactly for this model because here $kT/m\gamma = D\,2(1 - u)$.

ACKNOWLEDGMENTS

The author would like to thank the referee for critical comments and also Dr. D. Roekaerts and Dr. F. Langouche for helpful discussions.

APPENDIX

The generalized action (3.2) for the quantum open system can be transformed to the following form:

$$W(t) = \tfrac{1}{2}(p_k q_k - \bar{p}_k \bar{q}_k)|_0^t + \int_0^t \{\tfrac{1}{2}[(p_k \dot{q}_k - \dot{p}_k q_k) - (\bar{p}_k \dot{\bar{q}}_k - \dot{\bar{p}}_k \bar{q}_k)] - \mathcal{L}_{\mathrm{cl}}(p, q|\bar{p}, \bar{q})\}\, dt'. \tag{A1}$$

For quadratic systems one can introduce the notation

$$x = \{x_1 \cdots x_{4n}\} \equiv \{p_k, q_k, \bar{p}_k, \bar{q}_k, \quad k = 1,2,\ldots,n\}$$

and the action $W(t)$ can be written as

$$W(t) = \tfrac{1}{2}(p_k q_k - \bar{p}_k \bar{q}_k)|_0^t + \int_0^t \{\tfrac{1}{2}J_{\alpha\beta}x_\alpha \dot{x}_\beta - \tfrac{1}{2}B_{\alpha\beta}x_\alpha x_\beta - c_\alpha x_\alpha\}\, dt', \tag{A2}$$

where

$$J_{\alpha\beta} = -J_{\beta\alpha}, \quad B_{\alpha\beta} = B_{\beta\alpha}.$$

Consider an extremal trajectory $\{\tilde{x}_\alpha(t')\}$ for which $\delta W(t) = 0$. Therefore, if $x_\alpha = \tilde{x}_\alpha + \delta x_\alpha$ with the boundary conditions $\delta x_\alpha(0) = \delta x_\alpha(t) = 0$ we also have

$$0 = \delta W(t) = \int_0^t \{(J_{\beta\alpha}\dot{\tilde{x}}_\alpha - B_{\beta\alpha}x_\alpha - c_\beta)\delta x_\beta\}\, dt'. \tag{A3}$$

It follows that

$$J_{\beta\alpha}\dot{\tilde{x}}_\alpha - B_{\beta\alpha}\tilde{x}_\alpha - c_\beta = 0. \tag{A4}$$

Substituting (A4) into (A2) and remembering that

$$c_\beta x_\beta = \sigma_k(p_k - \bar{p}_k) + \kappa_k(q_k - \bar{q}_k)$$

[c.f. (4.1) and (4.3)], we finally obtain (4.7).

[1]E. B. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976).
[2]A. Kossakowski, Rep. Math. Phys. **3**, 247 (1972).
[3]V. Gorini, A. Frigerio, M. Verri, A. Kossakowski, and E. C. G. Sudarshan, Rep. Math. Phys. **13**, 149 (1978).
[4]G. Lindblad, Commun. Math. Phys. **48**, 119 (1976).
[5]G. Lindblad, Rep. Math. Phys. **10**, 393 (1976).
[6]E. B. Davies, Rep. Math. Phys. **11**, 169 (1977).
[7]R. Alicki, F. Langouche, and D. Roekaerts, "Semiclassical approximation for a class of quantum dynamical semigroups," Physica (to appear).
[8]R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[9]J. Klauder, Phys. Rev. D **19**, 2349 (1979).
[10]F. Langouche, D. Roekaerts, and E. Tirapegui, Phys. Rev. D **20**, 419 (1979).
[11]M. S. Marinov, Phys. Rep. **60**, 1 (1980).
[12]F. Langouche, D. Roekaerts, and E. Tirapequi, in *Functional Integration, theory and applications*, edited by J.-P. Antoine and E. Tirapequi (Plenum, New York, 1980), pp. 191–204.
[13]M. M. Mizrahi, J. Math. Phys. **22**, 102 (1981).
[14]F. Langouche, D. Roekaerts, and E. Tirapequi, "WKB-expansion for arbitrary Hamiltonians," Nuovo Cimento B (to appear).
[15]R. Alicki, "Path Integrals for Open Quantum Systems," preprint (1979).
[16]G. G. Emch, Helv. Phys. Acta **37**, 532 (1964).
[17]J. Knoll and R. Schaeffer, Phys. Rep. **31**, 159 (1977).
[18]D. W. McLaughlin, J. Math. Phys. **13**, 784 (1972).
[19]T. Szczesny, M.D. thesis, Gdansk, 1980 (in Polish).
[20]R. Alicki, Phys. Lett. B **89**, 373 (1980).

On inverse problems for plane-parallel media with nonuniform surface illumination

C. E. Siewert
Mathematics and Nuclear Engineering Departments, North Carolina State University, Raleigh, North Carolina 27650

W. L. Dunn
Operations Analysis Division, Research Triangle Institute, Research Triangle Park, North Carolina 27709

(Received 1 December 1981; accepted for publication 5 February 1982)

Elementary considerations are used to solve the inverse problem in linear transport theory for the case of variable illumination over the surface of a plane-parallel layer. The developed formalism yields as a special case the inverse solution for the classical searchlight problem.

PACS numbers: 05.60. + w

I. INTRODUCTION

The inverse problem in radiation transport theory is concerned with the determination of scattering and absorbing properties of a medium from a set of measurable radiation quantities. In the past few years considerable work regarding exact solutions of such inverse problems has been reported.[1-8] However, all of these papers[1-8] have dealt with the case of an infinite plane-parallel layer illuminated uniformly over each of the two free surfaces. From a practical and/or experimental point of view, such problems cannot be easily realized, and so here we report a solution for a class of inverse problems that allows the incident radiation to vary over the surfaces.

We employ a notational scheme similar to that used by Rybicki[9] in a study of the searchlight problem, and thus we write the radiation transport equation as

$$\mu \frac{\partial}{\partial z} I(z, \boldsymbol{\rho}, \boldsymbol{\Omega}) + \boldsymbol{\omega} \cdot \frac{\partial}{\partial \boldsymbol{\rho}} I(z, \boldsymbol{\rho}, \boldsymbol{\Omega}) + I(z, \boldsymbol{\rho}, \boldsymbol{\Omega}) = \frac{c}{4\pi} \int\int I(z, \boldsymbol{\rho}, \boldsymbol{\Omega}') p(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}') d\boldsymbol{\Omega}', \tag{1}$$

where z and $\boldsymbol{\rho}$, which lies in the x–y plane, locate in optical units the position in the homogeneous medium and $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\mu, \phi)$, with $\mu = \cos(\theta)$, is a unit vector that defines the direction of propagation (see Fig. 1). In addition, $\boldsymbol{\omega}$ is the projection of $\boldsymbol{\Omega}$ in the x–y plane and $c < 1$ is the albedo for single scattering. We consider that $I(z, \boldsymbol{\rho}, \boldsymbol{\Omega})$ satisfies Eq. (1) subject to the boundary conditions

$$I(0, \boldsymbol{\rho}, \boldsymbol{\Omega}) = I_1(\boldsymbol{\rho}, \boldsymbol{\Omega}), \quad \mu > 0, \phi \in [0, 2\pi], \tag{2a}$$

and

$$I(a, \boldsymbol{\rho}, \boldsymbol{\Omega}) = I_2(\boldsymbol{\rho}, \boldsymbol{\Omega}), \quad \mu < 0, \phi \in [0, 2\pi], \tag{2b}$$

where $I_1(\boldsymbol{\rho}, \boldsymbol{\Omega})$ and $I_2(\boldsymbol{\rho}, \boldsymbol{\Omega})$ are assumed to be given and to have two-dimensional Fourier transforms. Expanding the scattering law in terms of Legendre polynomials, we write

$$p(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}') = \sum_{l=0}^{\infty} \beta_l P_l(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}'), \quad \beta_0 = 1, \tag{3}$$

or, if we use the addition theorem,

$$p(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}') = \sum_{l=0}^{\infty} \sum_{m=0}^{l} \beta_l^m P_l^m(\mu) P_l^m(\mu') \cos[m(\phi - \phi')]. \tag{4}$$

Here we use $P_l^m(\mu)$ to denote the associated Legendre functions,

$$P_l^m(\mu) = (1 - \mu^2)^{m/2} \frac{d^m}{d\mu^m} P_l(\mu) \tag{5}$$

and

$$\beta_l^m = (2 - \delta_{0,m}) \frac{(l-m)!}{(l+m)!} \beta_l. \tag{6}$$

We assume that, in general, the quantities $I(0, \boldsymbol{\rho}, \boldsymbol{\Omega})$, for $\mu < 0$ and $\phi \in [0, 2\pi]$, and $I(a, \boldsymbol{\rho}, \boldsymbol{\Omega})$, for $\mu > 0$ and $\phi \in [0, 2\pi]$, can be determined experimentally, and we seek to express the single-scattering albedo c and the coefficients β_l in the Legendre expansion of the scattering law in terms of these quantities.

II. ANALYSIS

We can multiply Eqs. (1) and (2) by $\exp(i\mathbf{k} \cdot \boldsymbol{\rho})$ and integrate, for fixed z, over the x–y plane to find

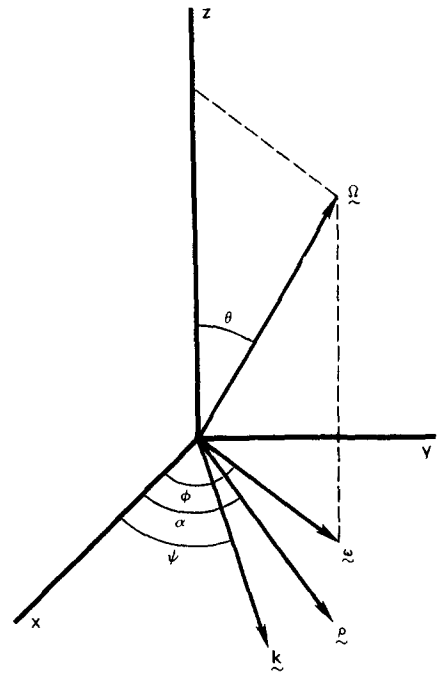


FIG. 1. The geometry for $\boldsymbol{\Omega}$, $\boldsymbol{\omega}$, $\boldsymbol{\rho}$, and $\mathbf{k}$.

 0022-2488/82/071376-03$02.50

$$\mu \frac{\partial}{\partial z} \Psi(z, \mu, \phi) + [1 - if(\mu, \phi)] \Psi(z, \mu, \phi)$$
$$= \frac{c}{4\pi} \int_0^{2\pi} \int_{-1}^{1} \Psi(z, \mu', \phi') p(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}') d\mu' d\phi' \tag{7}$$

and, for $\mu > 0$ and $\phi \in [0, 2\pi]$,

$$\Psi(0, \mu, \phi) = \Psi_1(\mu, \phi) \tag{8a}$$

and

$$\Psi(a, -\mu, \phi) = \Psi_2(\mu, \phi), \tag{8b}$$

where we suppress the dependence on the vector $\mathbf{k}$, which is in the x–y plane as shown in Fig. 1, and write

$$\Psi(z, \mu, \phi) = \int\int I(z, \boldsymbol{\rho}, \boldsymbol{\Omega}) e^{i\mathbf{k} \cdot \boldsymbol{\rho}} d\boldsymbol{\rho}, \tag{9}$$

$$\Psi_1(\mu, \phi) = \int\int I_1[\boldsymbol{\rho}, \boldsymbol{\Omega}(\mu, \phi)] e^{i\mathbf{k} \cdot \boldsymbol{\rho}} d\boldsymbol{\rho}, \tag{10a}$$

and

$$\Psi_2(\mu, \phi) = \int\int I_2[\boldsymbol{\rho}, \boldsymbol{\Omega}(-\mu, \phi)] e^{i\mathbf{k} \cdot \boldsymbol{\rho}} d\boldsymbol{\rho}. \tag{10b}$$

In addition,

$$f(\mu, \phi) = \mathbf{k} \cdot \boldsymbol{\omega} = k(1 - \mu^2)^{1/2} \cos(\phi - \psi), \tag{11}$$

with $k = |\mathbf{k}|$. We now follow an earlier work[4] and let

$$F(z, \mu, \phi) = \mu \frac{\partial}{\partial z} \Psi(z, \mu, \phi) \tag{12}$$

so that we can change μ to $-\mu$ in Eq. (7) and write

$$F(z, -\mu, \phi) + [1 - if(\mu, \phi)] \Psi(z, -\mu, \phi)$$
$$= \frac{c}{4\pi} \sum_{l=0}^{\infty} \sum_{m=0}^{l} (-1)^{l-m} \beta_l^m P_l^m(\mu)$$
$$\times \int_0^{2\pi} \Psi_l^m(z, \phi') \cos[m(\phi - \phi')] d\phi', \tag{13}$$

where

$$\Psi_l^m(z, \phi) = \int_{-1}^{1} P_l^m(\mu) \Psi(z, \mu, \phi) d\mu. \tag{14}$$

We can multiply Eq. (13) by $\Psi(z, \mu, \phi)$ and integrate over all μ and ϕ to find

$$T_0(z) + \int_0^{2\pi} \int_{-1}^{1} [1 - if(\mu, \phi)] \Psi(z, \mu, \phi) \Psi(z, -\mu, \phi) d\mu \, d\phi$$
$$= \frac{c}{4\pi} \sum_{l=0}^{\infty} \sum_{m=0}^{l} (-1)^{l-m} \beta_l^m [C_l^m(z) + S_l^m(z)], \tag{15}$$

where

$$C_l^m(z) = \left(\int_0^{2\pi} \Psi_l^m(z, \phi) \cos(m\phi) d\phi \right)^2, \tag{16a}$$

$$S_l^m(z) = \left(\int_0^{2\pi} \Psi_l^m(z, \phi) \sin(m\phi) d\phi \right)^2, \tag{16b}$$

and

$$T_0(z) = \int_0^{2\pi} \int_{-1}^{1} \Psi(z, \mu, \phi) F(z, -\mu, \phi) d\mu \, d\phi. \tag{17}$$

If we now differentiate Eqs. (15) and (17) and use Eq. (13) we can deduce that $T_0(z)$ is a constant, and on considering Eq. (15) at $z = 0$ and $z = a$ and subtracting the two resulting equations, we find

$$S_0 = \frac{c}{8\pi} \sum_{l=0}^{\infty} \sum_{m=0}^{l} (-1)^{l-m} \beta_l^m$$
$$\times [C_l^m(0) - C_l^m(a) + S_l^m(0) - S_l^m(a)], \tag{18}$$

where

$$S_0 = \int_0^{2\pi} \int_0^{1} [1 - if(\mu, \phi)] [\Psi(0, \mu, \phi) \Psi(0, -\mu, \phi)$$
$$- \Psi(a, \mu, \phi) \Psi(a, -\mu, \phi)] d\mu \, d\phi. \tag{19}$$

As we consider $I(z, \boldsymbol{\rho}, \boldsymbol{\Omega})$ to be known on the boundaries, $z = 0$ and $z = a$, the unknowns in Eq. (18) are c and the coefficients $\{\beta_l\}$. Thus we define

$$K_l^m = (2 - \delta_{0,m})(-1)^{l-m} \frac{(l-m)!}{(l+m)!}$$
$$\times [C_l^m(0) - C_l^m(a) + S_l^m(0) - S_l^m(a)] \tag{20}$$

and write Eq. (18) as

$$S_0 = \frac{c}{8\pi} \left\{ [\Psi_0(0)]^2 - [\Psi_0(a)]^2 + \sum_{l=1}^{\infty} \beta_l \sum_{m=0}^{l} K_l^m \right\}, \tag{21}$$

where

$$\Psi_0(z) = \int_0^{2\pi} \int_{-1}^{1} \Psi(z, \mu, \phi) d\mu \, d\phi. \tag{22}$$

Clearly for the case of isotropic scattering $\beta_l = 0$, $l \geqslant 1$, and Eq. (21) yields the concise result

$$c = 8\pi \{ [\Psi_0(0)]^2 - [\Psi_0(a)]^2 \}^{-1} S_0. \tag{23}$$

On the other hand, if we assume that $\beta_l = 0$ only for $l > L$, then Eq. (21) is a single equation for the $L + 1$ unknowns c and $\{\beta_l\}$. As the equation is linear in c and $\{c\beta_l\}$ we clearly can consider utilizing $L + 1$ different experiments, $L + 1$ different values of $\mathbf{k}$, or a combination of the two to generate $L + 1$ linear algebraic equations which, in principle, yield without approximation the desired solution of the inverse problem.

To complete this section we note for $k = 0$ that Eqs. (7) and (8) reduce to forms previously considered[1-8] and thus that in principle several solution techniques may apply, for example, in the event that a sufficient number of independent experiments is considered, that the boundary conditions lead to a radiation field that has sensitive dependence on the azimuthal angle, or that the scattering law is limited to three terms. By developing a solution here for the $k \neq 0$ case we clearly introduce the possibility of determining the required scattering coefficients from a single experiment, for a general class of boundary conditions and for a scattering law more general than the three-term model.

III. THE SEARCHLIGHT PROBLEM

As a special case of the foregoing we now consider the classical searchlight problem. We thus write

$$I_1(\boldsymbol{\rho}, \boldsymbol{\Omega}) = \frac{1}{2\pi\rho} \delta(\rho) \delta(\mu - \mu_0) \delta(\phi - \phi_0) \tag{24a}$$

and

$$I_2(\boldsymbol{\rho}, \boldsymbol{\Omega}) = 0, \tag{24b}$$

where we use the polar coordinates $\rho = |\boldsymbol{\rho}|$ and α to locate a field point in the x–y plane. Using Eqs. (24) in Eqs. (10), we obtain

$$\Psi_1(\mu, \phi) = \delta(\mu - \mu_0)\delta(\phi - \phi_0) \tag{25a}$$

and

$$\Psi_2(\mu, \phi) = 0, \tag{25b}$$

so that Eq. (19) becomes

$$S_0 = [1 - if(\mu_0, \phi_0)]\Psi(0, -\mu_0, \phi_0). \tag{26}$$

Equations (16) now yield

$$C_l^m(0) = \left(P_l^m(\mu_0)\cos(m\phi_0) + (-1)^{l-m} \times \int_0^{2\pi}\int_0^1 \Psi(0, -\mu, \phi)P_l^m(\mu)\cos(m\phi)d\mu\, d\phi\right)^2, \tag{27a}$$

$$S_l^m(0) = \left(P_l^m(\mu_0)\sin(m\phi_0) + (-1)^{l-m} \times \int_0^{2\pi}\int_0^1 \Psi(0, -\mu, \phi)P_l^m(\mu)\sin(m\phi)d\mu\, d\phi\right)^2, \tag{27b}$$

$$C_l^m(a) = \left(\int_0^{2\pi}\int_0^1 \Psi(a, \mu, \phi)P_l^m(\mu)\cos(m\phi)d\mu\, d\phi\right)^2, \tag{28a}$$

and

$$S_l^m(a) = \left(\int_0^{2\pi}\int_0^1 \Psi(a, \mu, \phi)P_l^m(\mu)\sin(m\phi)d\mu\, d\phi\right)^2, \tag{28b}$$

and Eq. (21) can be written as

$$8\pi[1 - if(\mu_0, \phi_0)]\,\Psi(0, -\mu_0, \phi_0) = c\left[C_0^0(0) - C_0^0(a) + \sum_{l=1}^{\infty}\beta_l \sum_{m=0}^{l} K_l^m\right]. \tag{29}$$

IV. CONCLUDING REMARKS

For the searchlight problem we note that Eq. (29) is our basic result for finding c and the coefficients $\{\beta_l\}$ in terms of the intensities on the two surfaces $z = 0$ and $z = a$. To use the equation we must in general be able to measure $I[0, \boldsymbol{\rho}, \boldsymbol{\Omega}(-\mu, \phi)]$ and $I[a, \boldsymbol{\rho}, \boldsymbol{\Omega}(\mu, \phi)]$, for $\mu \in [0, 1]$ and $\phi \in [0, 2\pi]$, experimentally and compute the quantities $C_l^m(0)$, $C_l^m(a)$, $S_l^m(0)$, and $S_l^m(a)$ with reasonable accuracy. As a first test of the solution we have considered the case of isotropic scattering and used the Monte Carlo method to solve, for given values of μ_0, ϕ_0, and c, the direct problem. For numerous cases studied we found that the value of c computed from

$$c = 8\pi[1 - if(\mu_0, \phi_0)]\,\Psi(0, -\mu_0, \phi_0)[C_0^0(0) - C_0^0(a)]^{-1} \tag{30}$$

agreed with the given value, for various choices of $\mathbf{k}$, with an accuracy consistent with the accuracy of the Monte Carlo results for the exiting intensities. More complete testing of the general formulation is clearly required in order to evaluate the extent to which basic results for practical experiments can be extracted from this exact solution.

It is clear that the inverse solution developed here for the infinite plane-parallel case requires that the incident radiation be specified over the entire boundary and that the exiting radiation be measured experimentally over the entire surface. However, in the event that there is absorption in the layer and the incident radiation is sufficiently localized (as, for example, in the searchlight problem) the case of a plane-parallel body finite in the transverse directions can be well approximated by the infinite plane-parallel case, and the developed inverse solution can be used with confidence.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation and by the National Institute of Health.

[1] C. E. Siewert, J. Math. Phys. **19**, 1587 (1978).
[2] C. E. Siewert, J. Math. Phys. **19**, 2619 (1978).
[3] N. J. McCormick, J. Math. Phys. **20**, 1504 (1979).
[4] C. E. Siewert, J. Quant. Spectrosc. Radiat. Transfer **22**, 441 (1979).
[5] W. L. Dunn and J. R. Maiorino, J. Quant. Spectrosc. Radiat. Transfer **24**, 203 (1980).
[6] E. W. Larsen, J. Math. Phys. **22**, 158 (1981).
[7] N. J. McCormick and R. Sanchez, J. Math. Phys. **22**, 199 (1981).
[8] R. Sanchez and N. J. McCormick, J. Math. Phys. **22**, 847 (1981).
[9] G. B. Rybicki, J. Quant. Spectrosc. Radiat. Transfer **11**, 827 (1971).

A Chern number for gauge fields on $\mathbb{R}^4$

R. Schlafly
Department of Mathematics, University of Chicago, 5734 University Avenue, Chicago, Illinois 60637

(Received 22 October 1980; accepted for publication 24 April 1981)

Let F be the curvature of some connection on some principal bundle over $\mathbb{R}^4$. I show that if F decays as fast as $(r^2 \ln r)^{-1}$ as r tends to infinity, then

$$\frac{1}{8\pi^2}\int \mathrm{tr}F^2$$

is an integer. If F decays like r^{-2}, any real value is possible. There is an analogous statement for $\mathbb{R}^{2n}(n>2)$, although it fails for $\mathbb{R}^2$.

PACS numbers: 11.10.Np

1. INTRODUCTION

Let G be a compact Lie group and let ρ: $G \to \mathrm{U}(N)$ be a representation. Let F be the curvature of some connection on some principal G bundle over S^4. The Chern number of the associated vector bundle is

$$c_2 = \frac{1}{8\pi^2}\int_{S^4} \mathrm{tr}\rho(F)^2.$$

According to the theory of characteristic classes, this quantity is always an integer.

Now consider the curvature F of a connection on a principal G bundle over $\mathbb{R}^4$. In this paper, we prove

Theorem 1.1: If

$$|F| \leqslant C/r^2 \ln r, \quad r \geqslant 2$$

for some constant C, then

$$\frac{1}{8\pi^2}\int_{\mathbb{R}^4} \mathrm{tr}\rho(F)^2$$

is an integer.

This formula is of interest in quantum field theory because the curvature F is the same as a Yang–Mills (gauge) field over Euclidean space. For discussions of related matters, see Refs. 1, 2, and 3.

If G is abelian or $N=1$, the integer obtained in Theorem 1.1 must be zero. If $G=\mathrm{SU}(N)(N>1)$, any integer value is possible.

It is not known whether the hypothesis of Theorem 1.1 can be relaxed to the energy $\int_{\mathbb{R}^4}|F|^2$ being finite, although this is suggested by Refs. 2, 4, and 5. However, it is shown in Sec. 6 that $(1/8\pi^2)\int_{\mathbb{R}^4}\rho(F)^2$ may not be an integer if we assume only that $|F| \leqslant C/r^2$. This example is equivalent to the one given in Ref. 5, and has infinite energy.

Theorem 1.1 is proved in Secs. 2, 3, 4, and 5. Section 3 gives a holonomy formula similar to the one in Ref. 6 and may be read independently of the rest of the paper. Section 7 shows how to handle dimensions other than four.

2. OUTLINE OF THE PROOF

For the proof of Theorem 1.1 it suffices to consider the associated $\mathrm{U}(N)$ bundle. Hence we may, without loss of generality, take $G=\mathrm{U}(N)$ and suppress mention of ρ.

Relative to some trivialization of the bundle, the connection A is a one-form on $\mathbb{R}^4$ with values in $\mathrm{u}(N)$. The curvature

$$F = dA + A^2$$

is a two-form on $\mathbb{R}^4$ with values in $\mathrm{u}(N)$. The Chern–Weil formalism (Ref. 7, p. 114) tells us that

$$\mathrm{tr}F^2 = d\,\mathrm{tr}(AF - \tfrac{1}{3}A^3).$$

Let S^3_r be the sphere of radius r in $\mathbb{R}^4$. If $|F| \leqslant C/r^2 \ln r$. Then $\mathrm{tr}F^2$ is integrable on $\mathbb{R}^4$. By Stokes' theorem,

$$\frac{1}{8\pi^2}\int_{\mathbb{R}^4}\mathrm{tr}F^2 = \lim_{r\to\infty}\frac{1}{8\pi^2}\int_{S^3_r}\mathrm{tr}(AF - \tfrac{1}{3}A^3).$$

Given $\epsilon>0$ and r sufficiently large, we will show that there exists a smooth map $\tilde{T}$: $S^3_r \to \mathrm{U}(N)$ such that

$$\left|\frac{1}{8\pi^2}\int_{S^3_r}\mathrm{tr}[\,AF - \tfrac{1}{3}A^3 - \tfrac{1}{3}(\tilde{T}^{-1}d\tilde{T})^3]\,\right| < \frac{\epsilon}{2}.$$

If we choose r large enough such that

$$\left|\frac{1}{8\pi^2}\int_{\mathbb{R}^4}\mathrm{tr}F^2 - \frac{1}{8\pi^2}\int_{S^3_r}\mathrm{tr}(AF - \tfrac{1}{3}A^3)\right| < \frac{\epsilon}{2},$$

then it will follow that

$$\left|\frac{1}{8\pi^2}\int_{\mathbb{R}^4}\mathrm{tr}F^2 - \frac{1}{24\pi^2}\int_{S^3_r}\mathrm{tr}(\tilde{T}^{-1}d\tilde{T})^3\right| < \epsilon.$$

We now recall a special case of Bott's work on periodicity in K-theory[8].

Theorem 2.1: For $N \geqslant 2$ we have an isomorphism $\pi_3\mathrm{U}(N) \simeq \mathbf{Z}$ given by assigning to a smooth map $\tilde{T}$: $S^3 \to \mathrm{U}(N)$ the integer

$$\frac{1}{24\pi^2}\int_{S^3}\mathrm{tr}(\tilde{T}^{-1}d\tilde{T})^3.$$

Thus for $\epsilon>0$ there exists an integer n such that

$$\left|\frac{1}{8\pi^2}\int_{\mathbb{R}^4}\mathrm{tr}F^2 - n\right| < \epsilon.$$

Hence $(1/8\pi^2)\int_{\mathbb{R}^4}\mathrm{tr}F^2$ must be an integer.

3. THE HOLONOMY PROPAGATOR

Let M be the sector in $\mathbb{R}^2$ described by polar coordinates (r,θ) with $0<r<r_0$ and $0<\theta<\theta_0$. Suppose we have a connection A on the trivial principal G bundle. For this paper we can assume that G is $\mathrm{U}(N)$, although Theorems 3.1 and 3.2 hold for any Lie group. A is a one-form on M with values in

 0022-2488/82/071379-06$02.50

$u(N)$. The curvature is $F = dA + A^2$.

For $x \in M$ let c_x: $[0, 1] \to M$ be the curve

$$c_x(t) = tx, \quad 0 \leqslant t \leqslant 1.$$

A section u: $M \to \mathbb{C}^N$ of the associated vector bundle is parallel along c_x iff it satisfies

$$c_x^*(du + Au) = 0.$$

The fundamental solution to this differential equation is the path ordered exponential

$$P\exp\left(-\int_{c_x} A\right).$$

If u is parallel along c_x then

$$u(x) = P\exp\left(-\int_{c_x} A\right)\cdot u(0).$$

$P\exp -(\int_{c_x} A)$ is called the holonomy from 0 to x along c_x. It may also be defined as a product integral (Ref. 9, p. 15).

Theorem 3.1: If

$$T(x) = P\exp\left(-\int_{c_x} A\right)$$

then

$$A(x) = -dT\,T^{-1} + T\left(\int_{c_x} T^{-1}FT\right)T^{-1}.$$

Proof: Plugging in the vector $\partial/\partial r$, we get

$$A\left(\frac{\partial}{\partial r}\right) = -\frac{\partial T}{\partial r}T^{-1},$$

which follows from the definition of T. The vector $\partial/\partial\theta$ gives

$$A_2 = -\frac{\partial T}{\partial \theta}T^{-1} + T\left(\int_0^r T^{-1}F_0 T\,dr\right)T^{-1},$$

where

$$A = A_1\,dr + A_2\,d\theta,$$

$$F = F_0 dr \wedge d\theta.$$

To prove this, we first observe that

$$-\frac{\partial A_1}{\partial \theta} = \frac{\partial}{\partial \theta}\left(\frac{\partial T}{\partial r}T^{-1}\right)$$

$$= \frac{\partial^2 T}{\partial \theta\,\partial r}T^{-1} - \frac{\partial T}{\partial r}T^{-1}\frac{\partial T}{\partial \theta}T^{-1}.$$

Hence

$$\frac{\partial}{\partial r}\left[T^{-1}\left(A_2 + \frac{\partial T}{\partial \theta}T^{-1}\right)T\right]$$

$$= -T^{-1}\frac{\partial T}{\partial r}T^{-1}A_2 T + T^{-1}\frac{\partial A_2}{\partial r}T + T^{-1}A_2\frac{\partial T}{\partial r}$$

$$- T^{-1}\frac{\partial T}{\partial r}T^{-1}\frac{\partial T}{\partial \theta} + T^{-1}\frac{\partial^2 T}{\partial r\,\partial \theta}$$

$$= T^{-1}A_1A_2T + T^{-1}\frac{\partial A_2}{\partial r}T$$

$$- T^{-1}A_2A_1T - T^{-1}\frac{\partial A_1}{\partial \theta}T$$

$$= T^{-1}F_0T.$$

Continuity of A and $dT\cdot T^{-1}$ at 0 requires that

$$\lim_{r\to 0} A_2 = 0,$$

$$\lim_{r\to 0}\frac{\partial T}{\partial \theta}T^{-1} = 0,$$

so

$$T^{-1}\left(A_2 + \frac{\partial T}{\partial \theta}T^{-1}\right)T = \int_0^r T^{-1}F_0 T\,dr$$

which gives the formula we wanted.

Theorem 3.2: The holonomy from 0 to 0 along ∂M is given by the θ-ordered exponential

$$P\exp\left(-\int_M T^{-1}FT\right).$$

Proof: Let

$$V(r, \theta) = P\exp\left(\int_0^\theta \int_0^r T^{-1}F_0 T\,dr\,d\theta\right)$$

with the ordering over θ, so

$$V(r_0, \theta_0) = P\exp\left(-\int_M T^{-1}FT\right).$$

By Theorem 3.1,

$$V = P\exp\left[-\int_0^{\theta_0}\left(T^{-1}A_2T + T^{-1}\frac{\partial T}{\partial \theta}\right)d\theta\right].$$

We compute

$$\frac{\partial}{\partial \theta}(TV)(TV)^{-1} = T\frac{\partial V}{\partial \theta}V^{-1}T^{-1} + \frac{\partial T}{\partial \theta}T^{-1}$$

$$= -T\left(T^{-1}A_2T + T^{-1}\frac{\partial T}{\partial \theta}\right)T^{-1} + \frac{\partial T}{\partial \theta}T^{-1}$$

$$= -A_2,$$

so by the uniqueness of solutions to this ordinary differential equation, there must be some W: $M \to G$ independent of θ such that

$$TV = P\exp\left(-\int_0^\theta A_2\,d\theta\right)\cdot W.$$

Writing $T = T(r, \theta)$ and letting $\theta = 0$ we find that $W = T(r, 0)$, so

$$T(r, \theta)V(r, \theta) = P\exp\left(-\int_0^\theta A_2\,d\theta\right)T(r, 0).$$

Setting $r = r_0$ and $\theta = \theta_0$ gives

$$V(r_0, \theta_0) = T(r_0, \theta_0)^{-1}P\exp\left(-\int_0^{\theta_0} A_2\,d\theta\right)T(r_0, 0)$$

which is precisely the holonomy from 0 to 0 along ∂M.

Proposition 3.3. For any one-form B and any curve c,

(i) $\left|P\exp\left(\int_c B\right)\right| \leqslant \exp\left(\int_c |B|\right)$,

(ii) $\left|1 - P\exp\left(\int_c B\right)\right| \leqslant \int_c |B|\exp\left(\int_c |B|\right)$.

Proof: Straightforward.

Corollary 3.4: If H is the holonomy along ∂M then

$$|1 - H| \leqslant \int_M |F|\cdot\exp\int_M |F|.$$

4. ESTIMATES IN THE RADIAL GAUGE

We now choose a gauge in which there is no holonomy in the radial direction. This means that $A(\partial/\partial r) = 0$ where r is the distance to the origin in $\mathbb{R}^4$. Define $G: \mathbb{R}^4 \to U(N)$ by

$$G(x) = P\exp\left(-\int_{c_x} A\right),$$

where

$$c_x(t) = tx, \quad 0 \leqslant t \leqslant 1.$$

Proposition 4.1: (i) Replacing A by $G^{-1}AG + G^{-1}\,dG$ (and F by $G^{-1}FG$) gives us a radial gauge.

(ii) In this gauge,

$$A(x) = \int_{c_x} F.$$

(iii) If C is a constant such that

$$|F| \leqslant C/r^2 \ln r$$

for $r \geqslant 2$ and

$$|A| \leqslant (C/e)\ln\ln 3$$

for $r = e = 2.718\cdots$, then for $r \geqslant 3$ we have

$$|A| \leqslant 2C(\ln\ln r)/r.$$

Proof: (i) and (ii) follow from Theorem 3.1. For (iii), write $F = F_0 dr \wedge d\theta$, so

$$A = \left(\int_0^r F_0\,dr\right)d\theta.$$

Since $|d\theta| = 1/r$ and $|F| = |F_0|/r$, we have

$$\begin{aligned}|A| &\leqslant \left|\int_0^e F_0\,dr\right||d\theta| + \left|\int_e^r F_0\,dr\right||d\theta| \\ &\leqslant \frac{C}{r}\ln\ln 3 + \frac{1}{r}\int_e^r \frac{C}{r^2\ln r}\,r\,dr \\ &= (C/r)\ln\ln 3 + (C/r)\ln\ln r \\ &\leqslant 2C(\ln\ln r)/r.\end{aligned}$$

Proposition 4.2: For $r \geqslant 3$, there exists a constant K_1 such that

$$\left|\int_{S_r^3} \mathrm{tr}AF\right| \leqslant K_1 \cdot \frac{\ln\ln r}{\ln r}.$$

Proof: The volume of S_r^3 is $2\pi r^3$, so

$$\begin{aligned}\left|\int_{S_r^3} \mathrm{tr}AF\right| &\leqslant \int_{S_r^3} |\mathrm{tr}\,AF| \\ &\leqslant N\int_{S_r^3} |A|\,|F| \\ &\leqslant N\,2\pi r^3\,2C\,[(\ln\ln r)/r](C/r^2\ln r) \\ &= 4\pi NC^2\,\frac{\ln\ln r}{\ln r}.\end{aligned}$$

■

Proposition 4.3: Suppose $r \geqslant 3$. If

$$|A - B| \leqslant \frac{2C}{r\ln r},$$

then there exists a constant K_2 such that

$$\left|\int_{S_r^3} \mathrm{tr}A^3 - \int_{S_r^3} \mathrm{tr}B^3\right| \leqslant K_2 \cdot \frac{(\ln\ln r)^2}{\ln r}.$$

Proof: We have

$$\begin{aligned}|B| &\leqslant |B - A| + |A| \\ &\leqslant 2C/r\ln r + 2C(\ln\ln r)/r \\ &\leqslant 4C(\ln\ln r)/r\end{aligned}$$

and

$$A^3 - B^3 = A^2(A - B) + A(A - B)B + (A - B)B^2,$$

so

$$\begin{aligned}|A^3 - B^3| &\leqslant \frac{2C}{r\ln r}\,3\left(4C\,\frac{\ln\ln r}{r}\right)^2 \\ &= \frac{96C^3(\ln\ln r)^2}{r^3\ln r}.\end{aligned}$$

Thus

$$\begin{aligned}\left|\int_{S_r^3} \mathrm{tr}A^3 - \int_{S_r^3} \mathrm{tr}B^3\right| &\leqslant N\int_{S_r^3} |A^3 - B^3| \\ &\leqslant N\,2\pi r^3\,96C^3(\ln\ln r)^2/r^3\ln r.\end{aligned}$$

5. THE HOLONOMY AT INFINITY

In this section, we restrict attention to S_r^3 for sufficiently large r. $C_1, C_2, \ldots, C_{13}$ will be constants independent of r.

Fix some "north pole" $p \in S_r^3$, and some great circle Γ from p to $-p$. If $x \in S_r^3$ is not on the "equator" S_r^2, let c_x be the shorter great arc from $\pm p$ to x. (There is a unique great circle in S_r^3 passing through $\pm p$ and x if $x \neq \pm p$.) For x in the "northern hemisphere," let $T(x)$ be the holonomy along c_x. For x in the "southern hemisphere", let $T(x)$ be the holonomy along Γ concatenated with c_x. T is discontinuous on the equator S_r^2.

Lemma 5.1: There exists a constant C_1 such that if T_1, T_2 are the two limiting values of T at a point x in S_r^2, then

$$|T_1 - T_2| \leqslant C_1/\ln r.$$

Proof: Choose a surface S in S_r^3 having area $\leqslant 2\pi r^2$ and boundary the union of Γ with the great semicircle from p to $-p$ through x. Then

$$\int_S |F| \leqslant 2\pi r^2\,C/r^2\ln r = 2\pi C/\ln r.$$

The holonomy around ∂S is $T_1T_2^{-1}$, so by Corollary 3.4,

$$|T_1 - T_2| = |1 - T_1T_2^{-1}| \leqslant \int_S |F| \exp\int_S |F|.$$

We now choose C_1 so that for large r,

$$\frac{2\pi C}{\ln r}\exp\left(\frac{2\pi C}{\ln r}\right) \leqslant \frac{C_1}{\ln r}.$$

Let $t: S_r^3 \to \mathbb{R}$ be the distance to p along S_r^3. Let $\theta = (\theta^1, \theta^2)$ be local coordinates on S_r^2, extended to S_r^3 by requiring θ to be constant along each c_x.

Lemma 5.2: For any $x \in S_r^3$,

$$\left|\int_{c_x} T^{-1}FT\right| \leqslant \frac{2C}{r\ln r}.$$

Proof: We suppose that x is on S_r^2, that c_x is taken to be the great arc from p to x, and that $d\theta^1, d\theta^2$ are orthonormal at x. The general case will fillow easily. Note that

$$|d\theta^i| = \csc(t/r), \quad |dt| = 1.$$

Let

$$F = F_{01}dt\wedge d\theta^1 + F_{02}dt\wedge d\theta^2 + F_{12}d\theta^1\wedge d\theta^2,$$

so

$$\int_{c_x} T^{-1}FT = \sum_{i=1}^{2}\left(\int_{c_x} T^{-1}F_{0i}T\,dt\right)d\theta^i.$$

Thus

$$\begin{aligned}\left|\int_{c_x} T^{-1}FT\right| &\leqslant \sum_{i=1}^{2}\int_{c_x}|F_{0i}|\,dt\\ &\leqslant 2\int_0^{\pi r/2}|F|\sin\left(\frac{t}{r}\right)dt\\ &\leqslant 2\frac{C}{r^2\ln r}\cdot r.\end{aligned}$$

If x is a point in the northern hemisphere (i.e., $t<\pi r/2$), then we have by Theorem 3.1,

$$A(x) = -dT\,T^{-1} + T\left(\int_{c_x} T^{-1}FT\right)T^{-1}. \tag{5.3}$$

If T_0 is constant, the substituting TT_0 for T leaves (5.3) unchanged. It follows that (5.3) is also valid if x is in the southern hemisphere. Hence we have

Corollary 5.4: (i) $|A + dT\,T^{-1}|\leqslant 2C/r\ln r$. (ii) $|dT| \leqslant C_2(\ln\ln r)/r$ for some constant C_2.

Given $\delta\in(0,1)$, let R be the set of points in S^3_r with $\frac{1}{2}\pi r - \delta < t < \frac{1}{2}\pi r + \delta$, and let

$$U(\theta) = \lim_{t\to\frac{1}{2}\pi r -} T(t,\theta).$$

From Corollary 5.4(ii), it follows that there exists a constant C_3 such that the inequality

$$|dU|\leqslant C_3(\ln\ln r)/r \tag{5.5}$$

holds on R. By Lemma 5.1, we can choose δ sufficiently small that on R,

$$|U - T|\leqslant 2C_1/\ln r.$$

We suppose that $2C_1/\ln r < \frac{1}{2}$ so that $f: R\to u(N)$ may be defined by the power series

$$f = \ln U^{-1}T = -\sum_{n=1}^{\infty}\frac{1}{n}(1-U^{-1}T)^n.$$

Then

$$\begin{aligned}|f| &\leqslant \sum_{n=1}^{\infty}\frac{1}{n}|1-U^{-1}T|^n\\ &\leqslant \sum_{n=1}^{\infty}\left(\frac{2C_1}{\ln r}\right)^n\\ &\leqslant \frac{4C_1}{\ln r}.\end{aligned} \tag{5.6}$$

Let ϕ: $\mathbf{R}\to[0,1]$ be a smooth function satisfying

$$\phi(t) = \begin{cases}1 & |t|\geqslant 1,\\ 0 & |t|\leqslant\frac{1}{2}.\end{cases}$$

Define $\tilde{T}: R\to U(N)$ by

$$\tilde{T}(t,\theta) = U(\theta)\exp[\phi((t-\tfrac{1}{2}\pi r)/\delta)f(t,\theta)]$$

and extend $\tilde{T}$ to be a smooth function on S^3_r by setting it equal to T on $S^3_r - R$.

Lemma 5.7: Let V and E be matrix valued functions.

(i) If $V = e^E$ then $|dV|\leqslant e^{|E|}|dE|$,

(ii) If $|1-V|\leqslant\mu<1$ and $E=\ln V$, then

$$|dE|\leqslant\frac{1}{1-\mu}|dV|.$$

Proof: This follows from differentiating the power series for exp and ln.

Lemma 5.8: There exist constants C_4, C_6, C_7 such that on R,

(i) $|df|\leqslant C_4\dfrac{\ln\ln r}{r}$,

(ii) $\left|\dfrac{\partial\tilde{T}}{\partial t}dt\right|\leqslant|d\tilde{T}|\leqslant\dfrac{C_6}{\delta\ln r}$,

(iii) $\left|\sum_{i=1}^{2}\dfrac{\partial\tilde{T}}{\partial\theta^i}d\theta^i\right|\leqslant C_7\dfrac{\ln\ln r}{r}$.

Proof: (i) Apply Lemma 5.7(ii) with $\mu=\frac{1}{2}$.

(ii) By Lemma 5.7(i),

$$|d\tilde{T}|\leqslant|dU|e^{|f|} + e^{|f|}|df| + e^{|f|}\left|d\left[\phi\left(\frac{t-\frac{1}{2}\pi r}{\delta}\right)\right]\right||f|.$$

Then, using (5.5), (5.6), and Lemma 5.8(i),

$$\begin{aligned}|d\tilde{T}| &\leqslant e^{4C_1/\ln r}\left(C_3\frac{\ln\ln r}{r} + C_4\frac{\ln\ln r}{r} + \frac{C_5}{\delta}\frac{4C_1}{\ln r}\right)\\ &\leqslant\frac{C_6}{\delta\ln r}.\end{aligned}$$

(iii) Similar, except that the term involving δ is absent.

Lemma 5.9: For some C_9,

$$\left|\int_{S^3_r}\mathrm{tr}(T^{-1}\,dT)^3 - \int_{S^3_r}\mathrm{tr}(\tilde{T}^{-1}\,d\tilde{T})^3\right|\leqslant C_9\frac{(\ln\ln r)^2}{\ln r}.$$

Proof:

$$\begin{aligned}&\left|\int_{S^3_r}\mathrm{tr}(T^{-1}dT)^3 - \int_{S^3_r}\mathrm{tr}(\tilde{T}^{-1}d\tilde{T})^3\right|\\ &\quad=\left|\int_R\mathrm{tr}(T^{-1}dT)^3 - \int_R\mathrm{tr}(\tilde{T}^{-1}d\tilde{T})^3\right|\\ &\quad\leqslant N\int_R|dT|^3 + \int_R|\mathrm{tr}(\tilde{T}^{-1}d\tilde{T})^3|.\end{aligned}$$

By Lemma 5.8(ii) and 5.8(iii), there exists C_8 such that

$$|\mathrm{tr}(\tilde{T}^{-1}d\tilde{T})^3|\leqslant C_8(1/\delta\ln r)[(\ln\ln r)/r]^2.$$

The volume of R is bounded by $4\pi r^2\delta$, so from Corollary 5.4(ii) and the above,

$$\begin{aligned}&\left|\int_{S^3_r}\mathrm{tr}(T^{-1}dT)^3 - \int_{S^3_r}\mathrm{tr}(\tilde{T}^{-1}d\tilde{T})^3\right|\\ &\leqslant N\,4\pi r^2\delta\,[(C_2\ln\ln r)/r]^3 + 4\pi r^2\delta\,C_8\,(1/\delta\ln r)[(\ln\ln r)/r]^2\\ &\leqslant C_9(\ln\ln r)^2/\ln r.\end{aligned}$$

We now complete the proof of Theorem 1.1. From $\mathrm{tr}F^2 = d\,\mathrm{tr}(AF - \frac{1}{3}A^3)$ and

$$|F|\leqslant C/r^2\ln r,$$

it follows that

$$\left|\frac{1}{8\pi^2}\int_{\mathbb{R}^4}\mathrm{tr}F^2-\frac{1}{8\pi^2}\int_{S^3_r}\mathrm{tr}(AF-\tfrac{1}{3}A^3)\right|$$
$$\leqslant\frac{N}{8\pi^2}\int_r^\infty\left(\frac{C}{\rho^2\ln\rho}\right)^2 2\pi\rho^3\,d\rho=\frac{C_{10}}{\ln r}.$$

From Proposition 4.2,

$$\left|\frac{1}{8\pi^2}\int_{\mathbb{R}^4}\mathrm{tr}F^2+\frac{1}{24\pi^2}\int_{S^3_r}\mathrm{tr}A^3\right|\leqslant C_{11}\frac{\ln\ln r}{\ln r}.$$

Corollary 5.4(i) shows that the hypothesis of Proposition 4.3 is satisfied if $B=-dT\cdot T^{-1}$, so

$$\left|\frac{1}{8\pi^2}\int_{\mathbb{R}^4}\mathrm{tr}F^2-\frac{1}{24\pi^2}\int_{S^3_r}\mathrm{tr}(T^{-1}dT)^3\right|\leqslant C_{12}\frac{(\ln\ln r)^2}{\ln r}.$$

By Lemma 5.9,

$$\left|\frac{1}{8\pi^2}\int_{\mathbb{R}^4}\mathrm{tr}F^2-n\right|\leqslant C_{13}\frac{(\ln\ln r)^2}{\ln r},$$

where n is the integer

$$\frac{1}{24\pi^2}\int_{S^3_r}\mathrm{tr}(\tilde{T}^{-1}d\tilde{T})^3.$$

The proof of Theorem 1.1 is now finished by letting r tend to infinity

6. AN EXAMPLE

Let $T: S^3\to\mathrm{SU}(2)$ be the standard identification. Let ω be the pull-back to $\mathbb{R}^4$ of $T^{-1}\,dT$ by the radial projection. The structural equation is

$$d\omega+\omega^2=0.$$

Let f: $[0,\infty)\to\mathbb{R}$ be a smooth function satisfying

$$f(r)=\begin{cases}0 & 0\leqslant r\leqslant\frac{1}{2},\\ a & r\geqslant 1,\end{cases}$$

for some real number a. Using r for the radial coordinate, $f(r)\omega$ is a well-defined smooth form on $\mathbb{R}^4$, and we define it to be a connection on the trivial SU(2) bundle.

We compute

$$F=f'dr\wedge\omega+(f^2-f)\omega^2,$$

so $F\in O(1/r^2)$. Also,

$$F^2=2(f^2-f)f'\,dr\wedge\omega^3$$

has compact support and

$$\frac{1}{8\pi^2}\int_{\mathbf{R}^4}\mathrm{tr}F^2=\frac{1}{8\pi^2}\int_0^\infty 2f'(r)(f^2(r)-f(r))\,dr\int_{S^3}\mathrm{tr}\omega^3$$
$$=\frac{1}{8\pi^2}\left[\tfrac{2}{3}f^3(r)-f^2(r)\right]_0^\infty\cdot 12\ \text{volume}\ S^3$$
$$=2a^3-3a^2.$$

By varying a, we may obtain any real number.

7. A GENERALIZATION

Theorem 1.1 generalizes to the following situation. Let G be a compact Lie group and let ρ: $G\to\mathrm{U}(N)$ be a representation. Let F be the curvature of some connection on some principal G bundle over $\mathbb{R}^{2n}$. Let

$$c_n=[(-1)^{n+1}/n]\mathrm{tr}(\rho(F)/2\pi i)^n.$$

If the connection extends to a connection on some bundle over S^{2n}, then the Chern number[10] of the associated vector bundle is $\int_{\mathbb{R}^{2n}}c_n$. This is an integer and is always divisible by $(n-1)!$ (see Ref. 7, p. 77 or Ref. 11, p. 156).

Theorem 7.1: Suppose $n>1$. If there exists a constant C such that

$$|F|\leqslant\frac{C}{r^2\ln r},\quad r\geqslant 2,$$

then

$$\frac{1}{(n-1)!}\int_{\mathbb{R}^{2n}}c_n$$

is an integer. Any integer value is possible if $G=\mathrm{SU}(N)$ and $N\geqslant n$.

Proof: The proof requires only minor modifications of the proof of Theorem 1.1, which we now discuss. According to the Chern–Weil theory (Ref. 7, p. 114),

$$\mathrm{tr}F^n=nd\int_0^1\mathrm{tr}\{A\,[tF+(t^2-t)A^2]^{n-1}\}dt.$$

It follows as before that in the radial gauge,

$$\int_{\mathbb{R}^4}\mathrm{tr}F^n=n\int_0^1(t^2-t)^{n-1}\,dt\lim_{r\to\infty}\int_{S^{2n-1}_r}\mathrm{tr}A^{2n-1}.$$

An elementary integration by parts gives

$$\int_0^1 t^{n-1}(1-t)^{n-1}\,dt=\frac{[(n-1)!]^2}{(2n-1)!}.$$

Approximating A by $-d\tilde{T}\,\tilde{T}^{-1}$ for some smooth $\tilde{T}$: $S^3_r\to\mathrm{U}(N)$, we find that

$$\frac{1}{(n-1)!}\int_{\mathbb{R}^{2n}}c_n$$

is approximated by

$$\frac{-1}{(2\pi i)^n}\frac{(n-1)!}{(2n-1)!}\int_{S^{2n-1}_r}\mathrm{tr}(\tilde{T}^{-1}\,d\tilde{T})^{2n-1}.$$

This is an integer because it is the integral that gives Bott periodicity,[8]

$$\pi_{2n-1}\mathrm{U}(N)\simeq\mathbb{Z},$$

for $n\leqslant N$. The integral is zero if $n>N$. ■

The proof of Theorem 7.1 breaks down if $n=1$. In order to estimate the holonomy around a closed curve near infinity, we expressed the curve as the boundary of a surface where the curvature was small and used Corollary 3.4. But the sphere of radius r in $\mathbb{R}^{2n}$ is simply connected only if $n>1$.

The following theorem shows that Theorem 7.1 actually fails if $n=1$.

Theorem 7.2: Let F be a two-form on $\mathbb{R}^2$ with values in some Lie algebra. Then F is the curvature of some connection on a principal bundle over $\mathbb{R}^2$.

Proof: Write

$$F=f(x,y)\,dx\wedge dy.$$

Then F is the curvature of the connection

$$A(x,y)=\left(\int_0^x f(x',y)\,dx'\right)dy.$$

ACKNOWLEDGMENT

This paper was motivated by a question raised by I. M. Singer. I would like to thank him for many interesting conversations.

[1]M. F. Atiyah, N. J. Hitchin, I. M. Singer, "Self-duality in four-dimensional Riemannian geometry," Proc. R. Soc. London A **362**, 425–461 (1978).
[2]A. A. Belavin, A. M. Polyakov, A. S. Schwartz, Yu. S. Tyupkin, "Pseudoparticle solutions of the Yang–Mills equations," Phys. Lett B **59**, 85–87 (1975).
[3]K. K. Uhlenbeck, "Removable singularities of Yang–Mills fields," Bull. Amer. Math. Soc. (new ser.) **1**, 579–581 (1979).
[4]E. C. Marino and J. A. Swieca, "Non-integer topological charge and infiniteness of action," P. U. C., Nota Cientifica 7/78.
[5]F. A. Schaposnik and J. E. Solomin "Gauge field singularities and noninteger topological charge," J. Math. Phys. **20**, 2110–2114 (1979).
[6]L. Schlesinger, "Parallelverschiebung und Krümmingstensor," Math. Ann. **99**, 413–434 (1928).
[7]S. S. Chern, *Complex Manifolds without Potential Theory*, 2nd Ed. (Springer, New York, 1979).
[8]R. Bott and R. Seeley, "Some remarks on the paper of Callias," Commun. Math. Phys. **62**, 235–245 (1978).
[9]E. Nelson, *Topics in Dynamics*, 1: *Flows* (Princeton U. P., Princeton, N. J., 1970).
[10]J. Milnor and J. Stasheff, *Characteristic Classes* (Princeton U. P., Princeton, N. J., 1974).
[11]P. Griffiths and J. Adams, *Topics in Algebraic and Analytic Geometry* (Princeton U. P., Princeton, N. J., 1974).

Nonlinear thermal evolution in an inhomogeneous medium[a)]

Shoshana Kamin
Department of Mathematics, Tel-Aviv University, Tel-Aviv, Israel

Philip Rosenau
Department of Mechanical Engineering, Technion, Israel Institute of Technology, Haifa, Israel

(Received 28 July 1981; accepted for publication 12 March 1982)

Various simple transport models of electron temperature in a confined plasma are reducible to the quasilinear equation $\rho(x)u_t = [c(x)u_x^n]_x + A(x)u^s$, $-1 < x < 1, u(\pm 1) = 0$. u is the temperature, $\rho(x)$ the density, and $c = g[\rho(x)]$ the density-dependent part of the thermal diffusion. $\rho(x)$ and $c(x)$ may vanish at the plasma edge, rendering the problem singular. The temporal behavior depends critically on the boundedness of $R = \int_{-1}^{+1} c^{-1}(x)\,dx$. If $R < \infty$ then in the absence of heat sources, $A \equiv 0$, every initially given state $u(x, 0)$ evolves toward an algebraically decaying, universal space-time separable solution. Its existence and uniqueness is proved. The method developed in this work may be used to show the equilibrization of the solution in the presence of a heat source of the form $A(x)u^s$, $s < n$, $\rho(x) > 0$. On the other hand, if $R = \infty$ and $A = 0$ then the system becomes isothermalized: $u \to \bar{u} = \int_{-1}^{+1} u(x, 0)\rho(x)\,dx / \int_{-1}^{1} \rho(x)\,dx > 0$. In such a case addition of heat sources will cause a thermal explosion.

PACS numbers: 52.25.Kn

I. INTRODUCTION

Mathematical modeling of radial transport in a confined plasma in its simplest formulation requires the solution of a quasilinear parabolic equation(s) in a fixed, bounded, domain.[1-10] In a tokamak, the presently most promising low β device, one distinguishes between energy and particle time scales, and it has become a common practice in theoretical calculations to treat each time scale separately. Thus calculations of thermal evolution,[2,3,10] or conditions for the set up of thermonuclear ignition[7,8] are done under the assumption of a stationary, homogeneous plasma. Here the rationale is that the particle confinement time is much larger than the energy confinement time. In the same spirit, in calculations of particle crossfield diffusion it is assumed that the plasma is in an isothermal state.[11,12]

In this work we will be mainly concerned with an asymptotic analysis of a certain mathematical model of a thermal evolution of a heated plasma. It will be assumed that the density is stationary but, unlike in previous studies, inhomogeneous. In fact, it is the study of the impact of inhomogeneity that distinguishes this work from previous studies. Since the density has a low value at the limiter, it is natural to assume it to vanish at the boundary. Moreover, since as a rule, present day theories predict the electron thermal diffusion to be density dependent such that it vanishes with density, the considered problem becomes singular. More will be said about this later.

Using an energy equation, the radial evolution of plasma temperature in a slab geometry, $x \in (-1, 1)$, may be described by

$$\rho \frac{\partial T}{\partial t} = \frac{\partial}{\partial x} K_\perp(\rho, x, T) \frac{\partial T}{\partial x} + \Phi(x, T, \rho), \tag{1.1}$$

where $\rho(x)$ is the particle density, T is the temperature, and $K_\perp$ is the perpendicular thermal conductivity; its form depends on the collision mechanism assumed. Φ represents the volumetric heating of the plasma.

We shall model $K_\perp$ assuming it to be of the form $K_\perp \sim \rho^\alpha T^{n-1}(x, t)$. Thus, for instance, $\alpha = 2$ in the classical, neoclassical, and banana diffusion or $\alpha = 1$ for the Bohm and plateau diffusion.

Similarly, the heat sources considered will be of the form $\Phi = A(x)T^s$, where $A = O(1)$, $s = 1.5$ for Ohmic heating in a constant electric field, and $A = O(\rho^2)$, $s = 2$ for alpha particle heating in the 6–20 keV regime.[13] Reduced to normalized units, the model equation to be studied is

$$\bar{\rho}(x) \frac{\partial u}{\partial t} = \frac{\partial}{\partial x} c(x) \frac{\partial u^n}{\partial x} + \bar{A}(x)u^s, \quad x \in (-1, 1). \tag{1.2}$$

Consider next

$$x \to y = R(x) = \int_0^x \frac{dx}{c(x)}. \tag{1.3}$$

If $c^{-1}(x)$ has integrable singularities, the case considered in this paper, $(-1, +1) \to (y_1, y_2)$, $|y_i| < \infty$. Rescaling the interval to $(-1, 1)$ and defining

$$x = y, \quad \rho(y) = c(x)\bar{\rho}(x),$$
$$A(y) = \bar{A}(x)c(x), \tag{1.4}$$

$y \to x$ and we obtain

$$\rho(x) \frac{\partial u}{\partial t} = \frac{\partial^2}{\partial x^2} u^n + A(x)u^s, \quad x \in (-1, 1), \tag{1.5}$$

$$u(-1, t) = u(+1, t) = 0 \tag{1.6}$$

together with

$$u(x, 0) = u_0(x), \quad x \in (-1, 1), \tag{1.7}$$

with $c(x)$, $A(x)$, and $\rho(x)$ that are known, bounded smooth functions of x and may vanish at $|x| = 1$. Equation (1.5) is in the form in which our results will be presented. Our main

[a)]This work was partially supported by U.S. Air Force Contracts AFOSR-78-3602 and AFOSR-76-2881.

 0022-2488/82/071385-06$02.50

concern will be with the purely diffusive case. Thus in what follows unless otherwise stated $A(x)\equiv 0$ will be assumed.

The asymptotic behavior of the solution of (1.5), (1.6) for $A(x)\equiv 0$ and $\rho(x)\equiv 1$ was studied in Refs. 14–17. (In Ref. 17 this problem is studied in $R^N, N>1$). It was shown there that the solution of such a problem converges to the separable solution of Eq. (1.5), namely,

$$t^{1/(n-1)}u(x,t)\to v_0(x), \quad t\to\infty, \tag{1.8}$$

where $v_0(x)$ is a nontrivial solution of

$$(v^n)'' + \frac{1}{n-1}v = 0, \quad v(\pm 1) = 0. \tag{1.9}$$

In the present work this result is generalized to the case $\rho(x)\neq$ const. In particular $\rho(x)$ may vanish at $|x|=1$. In the diffusive case, $A(x)\equiv 0$, we prove the uniqueness (Theorem 3.1) and existence (Theorem 3.2) of the positive separable solution of Eq. (1.5). This solution has the form

$$u(x,t) = \frac{1}{(a+t)^{1/(n-1)}}v_1(x), \quad a = \text{const},$$

where $v_1(x)$ satisfies Eq. (3.1). We prove also (Theorem 3.3) that if $u_0(x)\not\equiv 0$ then $t^{1/(n-1)}u(x,t)\to v_1(x)$, $t\to\infty$. Note that this result holds for $n>1$ and that the algebraic rate of decay depends on n. For $n<1$ the solution will be extincted within a finite time.[15,16]

The method for the diffusive case presented in Secs. 2 and 3 may be used to study the asymptotic behavior of Eq. (1.5). More will be said in Sec. 4.

Let us emphasize the importance of the integrability condition of $c^{-1}(x)$. [See Eq. (1.3).] Let c^{-1} be nonintegrable over $[-1,1]$. Now (1.3) maps $(-1,1)\to(-\infty,\infty)$ and (1.5) defined in $R^1\times(0,\infty)$ becomes a Cauchy problem. If also $A(x)\equiv 0$, then, as was proved by us elsewhere,[18]

$$u\to\bar{u} = \int_{-1}^{1} u(x,0)\rho(x)\,dx \Big/ \int_{-1}^{1}\rho(x)\,dx > 0 \quad \text{as } t\to\infty.$$

In physical interpretation, in the nonintegrable case, $c(x)$ insulates the system thermally and hence the isothermalization occurs. Clearly, if in this case a heat source is added, the solution of (1.5) will explode in time.

We note that the vanishing of $c(x)$, on the boundary causes the degeneration of the problem. For linear operators such types of problems are well known.

Finally, recall that our problem models transport in axial symmetry. Therefore, one is naturally interested in $c(x)$ being an even function of x. Consequently, c^{-1} satisfies the same integrability conditions in both ends; a tacitly used assumption throughout this work. As a mathematical problem, one may also be interested in the mixed case with c^{-1} being integrable only at one end. This problem will not be considered here.

2. SOME PROPERTIES OF THE SOLUTION

We consider the problem

$$\rho(x)\frac{\partial u}{\partial t} = \frac{\partial^2 u^n}{\partial x^2} \quad \text{in } Q = \{|x|<1, t>0\}, \tag{2.1}$$

$$u(-1,t) = u(1,t) = 0, \tag{2.2}$$

$$u(x,0) = u_0(x). \tag{2.3}$$

We assume that $\rho(x)$ is a smooth function and

$$\rho(\pm 1)\geqslant 0, \quad \rho(x)>0 \quad \text{if } |x|<1. \tag{2.4}$$

As for $u_0(x)$, we assume

$$u_0(x)\in C^1([-1,1]), \quad u_0(x)\geqslant 0, \quad u_0(\pm 1) = 0. \tag{2.5}$$

In this section we present some properties of the solution of (2.1)–(2.3) which we use in the next section for the asymptotic analysis.

We define first the weak solution for the problem (2.1)–(2.3).

Definition 1: A function $f(x,t)$ will be called a test function if it has continuous derivatives f_x, f_t, f_{xx} in $\bar{Q}$ and $f(\pm 1, t) = 0\ \forall\ t>0$.

Definition 2: The function $u(x,t)$ defined in Q is a weak solution of (2.1)–(2.3) if

(i) u is bounded, continuous, and non-negative in $\bar{Q}$;

(ii) the weak derivative $\partial u^n/\partial x$ exists, and for any test function $f(x,t)$ and any $t_1\geqslant 0$, $t_2\geqslant t_1$,

$$\int_{t_1}^{t_2}\int_{-1}^{1}(\rho u f_t + u^n f_{xx})\,dx\,dt = \int_{-1}^{1}\rho(x)[f(x,t_2)u(x,t_2) - f(x,t_1)u(x,t_1)]\,dx; \tag{2.6}$$

(iii) $u(x,t)$ satisfies the initial condition (2.3)

Theorem 2.1: Assume (2.4)–(2.5) to hold. Then the weak solution of the problem (2.1)–(2.3) exists and is unique.

Theorem 2.1 is proved in Ref. 19 for the case $\rho(x)\geqslant\rho_0>0$ for $|x|\leqslant 1$. For $\rho(x)$ satisfying (2.4), the uniquenesss is proved along the same lines as in Ref. 19. The existence may also be proved in a similar way to the one proposed in Ref. 19. We shall only outline it here.

Let

$$Q_{h,T} = \{(x,t): |x|\leqslant h,\ 0<t<T\}, \quad 0<h<1,$$

and consider the solution of (2.1) that satisfies the conditions

$$\begin{aligned} &u(x,0) = u_0(x) + (1-h) && \text{for } |x|\leqslant h,\\ &u(-h,t) = u_0(-h) + 1 - h && \text{for } t>0,\\ &u(h,t) = u_0(h) + 1 - h && \text{for } t>0. \end{aligned} \tag{2.7}$$

Such solution exists and is denoted by $u_h(x,t)$. The function $u_h(x,t)$ satisfies (2.1) in the classical sense. By the maximum principle

$$1 - h\leqslant u_h(x,t)\leqslant M = \max|u_0(x)|. \tag{2.8}$$

Moreover, for every $T>0$,

$$2\int_0^T\int_{-h}^{h}\rho\frac{\partial u_h}{\partial t}\frac{\partial (u_h)^n}{\partial t}\,dx\,dt + \int_{-h}^{h}\left[\frac{\partial (u_h)^n}{\partial x}\right]^2 dx \leqslant M_1, \tag{2.9}$$

where M_1 is a constant that does not depend on h.

The last inequality may be obtained by multiplying Eq. (2.1) for u_h by $\partial(u_h)^n/\partial t$ and integrating it over the domain $Q_{h,T}$.

Let $h\to 1$. It follows from (2.9) that the sequence of functions

$$w_h(x,t) = [u_h(x,t)]^n$$

is bounded in $W_1^2(Q_{b,T})$ for every $T>0$, $b\in(0,1)$. Thus, by the imbedding theorems there exists the subsequence $w_{h_i}(x,t)$ such that $w_{h_i}(\cdot,t)$ converges in $L_2(-b,b)$ for all $t\in[0,T]$ and all $b\in(0,1)$. Moreover, we can extract w_{h_i} in such a way that $w_{h_i}(x,t)$ converges a.e. in $Q_{1,T}$. The limit function, denote it by $w(x,t)$, is defined in the whole $Q_{1,T}$. It follows from (2.8) and (2.9) that $w(x,t)$ is bounded and continuous in $Q_{1,T}$.

Set $u=(w)^{1/n}$. Then $u_{h_i}(x,t)\to u(x,t)$ a.e. in $Q_{1,T}$ and for fixed t_1 and t_2 we may extract the subsequence h_i so that a.e. $u_{h_i}(x,t_k)\to u(x,t_k)$ $(k=1,2)$.

Let $f(x,t)$ be some test function. We have for every $t_1\geqslant 0$, $t_2>t_1$,

$$\int_{t_1}^{t_2}\int_{-h}^{h}\left[\rho u_h f_t+(u_h)^n f_{xx}\right]dx\,dt$$
$$=\int_{-h}^{h}\rho\left[f(x,t_2)u_h(x,t_2)-f(x,t_1)u_h(x,t_1)\right]dx$$
$$+\int_{t_1}^{t_2}\frac{\partial f}{\partial x}(u_h)^n\,dt\Big|_{-h}^{h}. \tag{2.10}$$

Passing to the limit as $h\to 1$ we obtain (2.6).

Corollary 2.1: Let $u_1(x,t)$ and $u_2(x,t)$ be two solutions of Eq. (2.1) that satisfy the conditions

$$u_1(\pm 1,t)=u_2(\pm 1,t)=0$$

and

$$u_1(x,0)\leqslant u_2(x,0).$$

Then $u_1(x,t)\leqslant u_2(x,t)$.

Let

$$Lz=(z^n)_{xx}-\rho z_t. \tag{2.11}$$

We show the following comparison principle.

Lemma 2.1: Let $z(x,t)\in C^2(\overline{Q}_{b,T})$, $b<1$, $z(x,t)>0$ in $Q_{b,T}$. Be $u(x,t)$ the weak solution of the problem (2.1)–(2.3) in $\overline{Q}_{1,T}$. Then

$$Lz\geqslant 0 \text{ in } Q_{b,T}, \tag{2.12}$$

$$z(\pm b,t)=0,\quad 0\leqslant t\leqslant T, \tag{2.13}$$

$$0<z(x,0)\leqslant u_0(x)\quad \text{for } |x|<b \tag{2.14}$$

implies

$$z(x,t)\leqslant u(x,t)\quad \text{in } Q_{b,T}. \tag{2.15}$$

For the proof we recall first that the solution $u(x,t)$ may be obtained as a limit of classical positive solutions. Thus it is enough to prove the assertion of Lemma 2.1 for the case that $u(x,t)>0$ in $\overline{Q}_{b,T}$ and satisfies (2.1) in the classical sense. We have

$$u(\pm b,t)>0=z(\pm b,t)\quad 0\leqslant t\leqslant T. \tag{2.16}$$

By continuity, the inequality (2.16) remains true for $|x|=b-\epsilon$ for ϵ small enough. Therefore

$$u[\pm(b-\epsilon),t]>z[\pm(b-\epsilon),t]>0. \tag{2.17}$$

Next we compare $u(x,t)$ and $z(x,t)$ in $Q_{b-\epsilon,T}$. Following Refs. 19 and 17 we set $U=u^n$, $Z=z^n$. Then

$$-\rho U_t+nU^{1-1/n}U_{xx}=0\quad \text{in } Q_{b-\epsilon,T},$$
$$-\rho Z_t+nZ^{1-1/n}Z_{xx}\geqslant 0\quad \text{in } Q_{b-\epsilon,T}.$$

Therefore, for $w=U-Z$ we obtain

$$nU^{1/n}w_{xx}-\rho w_t+n(U^{1/n}-Z^{1/n})Z_{xx}\leqslant 0$$

in $Q_{b-\epsilon,T}$. The last inequality may be written as

$$a(x,t)w_{xx}-w_t+b(x,t)w\leqslant 0, \tag{2.18}$$

where $a(x,t)$ and $b(x,t)$ are bounded in $Q_{b-\epsilon,T}$. By the classical maximum principle[20] it follows from (2.18), (2.17), and (2.14) that

$$w(x,t)=U-Z\geqslant 0$$

in $Q_{b-\epsilon,T}$. Therefore $U\geqslant Z$ and

$$u\geqslant z\quad \text{in } Q_{b-\epsilon,T}. \tag{2.19}$$

From (2.19) follows (2.15). Similarly, it may be proved that

Lemma 2.2: Let $z(x,t)\in C^2(Q_1,T)$, $z(x,t)>0$ in $\overline{Q}_{1,T}$ and $u(x,t)$ is a weak solution of (2.1)–(2.3). Then

$$Lz\leqslant 0\quad \text{in } Q_{1,T},$$
$$z(x,0)\geqslant u_0(x)\quad \text{for } |x|\leqslant 1$$

implies

$$z(x,t)\geqslant u(x,t)\quad \text{in } Q_{1,T}.$$

Remark 2.1: Lemma 2.1 is true if instead of $Q_{b,T}$ one considers the domain $|x-x_0|\leqslant b$, $0<t<T$, where $(x_0-b,x_0+b)\subset(-1,1)$ and the condition (2.13) is changed to

$$z(x_0\pm b,t)=0.$$

Set

$$z_-(x,t;x_0,\alpha,\beta)=\alpha\frac{[\cos(\pi/2\beta)(x-x_0)]^{1/n}}{(t+1)^{1/(n-1)}}, \tag{2.20}$$

where x_0, α, and β are constants such that $-1<x_0-\beta<x_0+\beta<1$ and α will be chosen later. Using (2.11) we have

$$Lz_-=-\alpha^n\left(\frac{\pi}{2\beta}\right)^2\frac{\cos(\pi/2\beta)(x-x_0)}{(t+1)^{n/(n-1)}}$$
$$+\frac{\rho\alpha}{n-1}\frac{[\cos(\pi/2\beta)(x-x_0)]^{1/n}}{(t+1)^{n/(n-1)}} \tag{2.21}$$

Set $\rho_0=\min\rho(x)$ for $|x-x_0|\leqslant\beta$ and

$$\alpha_0=\left[\frac{4\beta^2\rho_0}{(n-1)\pi^2}\right]^{1/(n-1)}. \tag{2.22}$$

It follows from (2.22) and (2.21) that for $\alpha\leqslant\alpha_0$, $Lz_-\geqslant 0$ for $x\in(x_0-\beta,x_0+\beta)$ and $t>0$. The function z_- defined in (2.20) with $\alpha\leqslant\alpha_0$ will be referred to as a subsolution of Eq. (2.1). Similarly, we set

$$z_+(x,t;\alpha)=\alpha\frac{[\cos\tfrac{1}{2}\pi x]^{1/n}}{(t+1)^{1/(n-1)}}, \tag{2.23}$$

$\rho_1=\max_{|x|<1}\rho(x)$ and

$$\overline{\alpha}=\left[\frac{16\rho_1}{(n-1)\pi^2}\right]^{1/(n-1)}\cdot(\sqrt{2})^{1/n}.$$

Then if $\alpha\geqslant\overline{\alpha}$

$$Lz_+\leqslant 0\quad \text{for } |x|<1,\ t>0.$$

The function z_+ defined in (2.23) with $\alpha\geqslant\overline{\alpha}$ is a supersolution of (2.1).

Theorem 2.2: Assume $u(x,t)$ is the weak solution of the problem (2.1)–(2.3) and $(x_0-\gamma,x_0+\gamma)\subset(-1,1)$. Suppose that

$u_0(x) \geqslant \delta$ for $|x - x_0| \leqslant \gamma$.

Then there exist two constants M_0 and m_0 such that

$$u(x, t) \leqslant \frac{M_0}{(t+1)^{1/(n-1)}} \quad \text{for } t > 0, \ |x| \leqslant 1, \tag{2.24}$$

$$u(x, t) \geqslant \frac{m_0}{(t+1)^{1/(n-1)}} \quad \text{for } t > 0, \ |x - x_0| \leqslant \frac{\gamma}{2}. \tag{2.25}$$

The proof of (2.24) follows applying Lemma 2.2 and using the function $z_+(x, t; \alpha)$ with a large enough α. The proof of (2.25) follows from Lemma 2.1 and Remark 2.1 using $z_-(x, t; x_0, \alpha, \gamma)$ with a small enough α.

Next we set

$$u(x, t) = \frac{v^*(x, t)}{(t+1)^{1/(n-1)}}. \tag{2.26}$$

If $u(x, t)$ is the classical solution of (2.1) then $v^*(x, t)$ satisfies the equation

$$\rho(x)(t+1)\frac{\partial v^*}{\partial t} = \frac{\partial^2 (v^*)^n}{\partial x^2} + \frac{1}{n-1}\rho v^*.$$

Introducing the new variable

$$\tau = \ln(t+1)$$

and denoting

$$v(x, \tau) = v^*(x, t), \tag{2.27}$$

we obtain the equation for $v(x, \tau)$,

$$\rho(x)\frac{\partial v}{\partial \tau} = \frac{\partial^2 v^n}{\partial x^2} + \frac{1}{n-1}\rho v. \tag{2.28}$$

In the general case $v(x, \tau)$ satisfies the integral identity

$$\int_{\tau_1}^{\tau_2}\int_{-1}^{1}\left[\rho v f_\tau + v^n f_{xx} + \frac{1}{n-1}\rho v f\right] dx\, d\tau = \int_{-1}^{1}\rho(x)[f(x, \tau_2)v(x, \tau_2) - f(x, \tau_1)v(x, \tau_1)]\, dx \tag{2.29}$$

for any test function $f(x, \tau)$. The integral identity (2.29) is easily obtained from (2.6) substituting (2.26) and (2.27). From the definition of $v(x, \tau)$ and Theorem 2.2 follows

Theorem 2.3: For $v(x, \tau)$ defined by (2.26), (2.27), and under the assumptions of Theorem 2.2,

$$v(x, \tau) \leqslant M_0 \quad \text{for } |x| \leqslant 1, \qquad \tau > 0, \tag{2.30}$$

$$v(x, \tau) \geqslant m_0 \quad \text{for } |x - x_0| \leqslant \tfrac{1}{2}\gamma, \quad \tau > 0. \tag{2.31}$$

We shall need also the following estimates.

Lemma 2.3: Let $u(x, t)$ be the solution of (2.1)–(2.3) and $v(x, \tau)$ defined in (2.26) and (2.27). Then there exist the weak derivative $\partial v^{(n+1)/2}/\partial \tau$ and, for every fixed τ, the weak derivative $\partial v^n/\partial x$. Moreover,

$$\int_0^T\int_{-1}^{1}\rho(x)\left(\frac{\partial v^{(n+1)/2}}{\partial \tau}\right)^2 dx\, d\tau \leqslant M_2, \tag{2.32}$$

$$\int_{-1}^{1}\left(\frac{\partial v^n}{\partial x}\right)^2 dx\Big|_{\tau = T} \leqslant M_2, \tag{2.33}$$

where M_2 is some constant depending only on $u_0(x)$.

We prove first (2.32) and (2.33) for $v = v_h$, where

$$v_h(x, \tau) = v_h^*(x, t) = u_h(x, t)\cdot(1+t)^{1/(n-1)},$$

where u_h is defined as in the proof of Theorem 2.1. To prove the integral estimates for v_h we multiply Eq. (2.28) by $\partial (v_h)^n/\partial \tau$ and integrate it over $Q_{h,T}$. Integrating by parts and using (2.30) we get

$$\iint_{Q_{h,T}}\rho(x)\left(\frac{\partial (v_h)^{(n+1)/2}}{\partial \tau}\right)^2 dx\, dt \leqslant M_2,$$

$$\int_{-h}^{h}\left(\frac{\partial (v_h)^n}{\partial x}\right)^2 dx\Big|_{\tau = T} \leqslant M_2.$$

Passing to the limit as $h \to 1$ we get the assertion of the lemma.

3. ASYMPTOTIC ANALYSIS

Suppose that the assumption on $\rho(x)$ and $u_0(x)$ stated in the beginning of Sec. 2 hold.

Theorem 3.1: The solution of the problem

$$(v^n)'' + [1/(n-1)]\,\rho(x)v = 0, \quad |x| < 1, \tag{3.1}$$

$$v(-1) = v(1) = 0, \tag{3.2}$$

$$v(x) \geqslant 0, \quad v(x) \not\equiv 0, \quad |x| \leqslant 1 \tag{3.3}$$

is unique.

Theorem 3.2: The nontrivial solution $v_1(x)$ of the problem (3.1)–(3.3) exists.

Theorem 3.3: Let $u(x, t)$ be the weak solution of the problem (2.1)–(2.3) and $v_0(x)$ the solution of (3.1)–(3.2). Then

$$t^{1/(n-1)}u(x, t) \to v_1(x) \quad \text{if } t \to \infty$$

uniformly for $|x| \leqslant 1$.

Proof of Theorem 3.1: Suppose there exist two solutions of (3.1)–(3.3), $v_1(x)$ and $v_2(x)$. By the maximum principle

$$v_1(x) > 0, \ v_2(x) > 0 \quad \text{for } |x| < 1. \tag{3.4}$$

Moreover,

$$\frac{dv_1^n}{dx}\Big|_{x = \pm 1} \neq 0, \quad \frac{dv_2^n}{dx}\Big|_{x = \pm 1} \neq 0. \tag{3.5}$$

It follows from (3.4) and (3.5) that there exist α_1, α_2 such that

$$\alpha_1^n v_1^n(x) \leqslant v_2^n(x) \leqslant \alpha_2^n v_2^n(x) \quad \text{for } |x| \leqslant 1$$

or

$$\alpha_1 v_1(x) \leqslant v_2(x) \leqslant \alpha_2 v_2(x). \tag{3.6}$$

Set

$$u_1(x, t) = \frac{v_1(x)}{(t+\xi_1)^{1/(n-1)}}, \quad \bar{u}(x, t) = \frac{v_2(x)}{(t+1)^{1/(n-1)}},$$

$$u_2(x, t) = \frac{v_1(x)}{(t+\xi_2)^{1/(n-1)}}.$$

The functions $u_1(x, t)$, $\bar{u}(x, t)$, and $u_2(x, t)$ are the solutions of Eq. (2.1) for every $\xi_1 > 0$ and $\xi_2 > 0$. Be $\xi_1 = \alpha_1^{1-n}$, $\xi_2 = \alpha_2^{1-n}$. Then by (3.6) we have

$$u_1(x, 0) \leqslant \bar{u}(x, 0) \leqslant u_2(x, 0). \tag{3.7}$$

Applying Corollary 2.1 we obtain

$$u_1(x, t) \leqslant \bar{u}(x, t) \leqslant u_2(x, t)$$

or

$$v_1(x)\left(\frac{t+1}{t+\xi_1}\right)^{1/(n-1)} \leqslant v_2(x) \leqslant v_1(x)\left(\frac{t+1}{t+\xi_2}\right)^{1/(n-1)}.$$

Passing to the limit as $t \to \infty$ we get $v_1(x) \equiv v_2(x)$.

Lemma 3.1: Let $v(x, \tau)$ be defined by (2.26) and (2.27).

Assume that $v(x, 0) = u(x, 0) \not\equiv 0$. Then there exists a sequence $\{\tau_i\}$, $\tau_i \to \infty$ such that the sequence of functions $v(x, \tau_i)$ $(-1 \leqslant x \leqslant 1)$ converges uniformly to the limit function $v_1(x)$. Moreover,

$$v_1(x) \not\equiv 0. \tag{3.8}$$

The proof follows from (2.33) and (2.31).

Lemma 3.2: Let $v(x, \tau_i)$ be the sequence defined in Lemma 3.1 and

$$v_1(x) = \lim_{\tau_i \to \infty} v(x, \tau_i). \tag{3.9}$$

Then $v_1(x)$ is the solution of the problem (3.1)–(3.3).

Proof: Let $\mathscr{D} = \{(x, \tau), |x| \leqslant 1, |\tau| \leqslant 1\}$. We prove first that

$$\|v^{(n+1)/2}(x, \tau + \tau_i) - v_1^{(n+1)/2}(x)\|_{L_2(\mathscr{D})} \to 0 \tag{3.10}$$

as $\tau_i \to \infty$. Let ϵ be some arbitrary, small number. By (2.33) there exists $b = b(\epsilon)$ such that

$$0 \leqslant v(x, \tau) \leqslant \epsilon, \quad \text{for } \tau > 0, \ b \leqslant |x| \leqslant 1 \tag{3.11}$$

and

$$0 \leqslant v_1(x) \leqslant \epsilon \quad \text{for } b \leqslant |x| \leqslant 1. \tag{3.12}$$

Next we choose $T = T(\epsilon)$ large enough so that

$$\int_{\tau_i}^{\tau_i + 1} \int_{-b}^{b} \left[\frac{\partial v^{(n+1)/2}}{\partial \tau}\right]^2 dx\, d\tau < \epsilon \quad \text{for } \tau_i \geqslant T. \tag{3.13}$$

Such T exists because of (2.32). It follows from (3.13) that

$$\int_{-b}^{b} |v^{(n+1)/2}(x, \tau_i + \tau) - v^{(n+1)/2}(x, \tau_i)|^2\, dx \leqslant \epsilon \tag{3.14}$$

for $\tau_i \geqslant T$ and all $\tau \in [0, 1]$. We conclude from (3.9) and (3.14) that for τ_i large enough and $\tau \in [0, 1]$

$$\int_{-b}^{b} |v^{(n+1)/2}(x, \tau_i + \tau) - v_1^{(n+1)/2}(x)|^2\, dx \leqslant 2\epsilon. \tag{3.15}$$

From (3.15), (3.11), and (3.12) we obtain that for τ_i large enough and all $\tau \in [0, 1]$

$$\int_{-1}^{1} |v^{(n+1)/2}(x, \tau_i + \tau) - v_1^{(n+1)/2}(x)|^2\, dx \leqslant \epsilon(2 + \sqrt{2}). \tag{3.16}$$

Hence (3.10) is proved.

Next we conclude from (3.16) and (3.10) that there exists a subsequence of τ_i such that

$$v^{(n+1)/2}(x, \tau_i + \tau) \to v_1^{(n+1)/2}(x) \quad \text{a.e. in } \mathscr{D}, \tag{3.17}$$

$$v^{(n+1)/2}(x, \tau_i + 1) \to v_1^{(n+1)/2}(x) \quad \text{a.e. for } |x| \leqslant 1. \tag{3.18}$$

if $\tau_i \to \infty$. We use the same notation for this subsequence. From (3.17) and (3.18) we have

$$v(x, \tau + \tau_i) \to v_1(x) \quad \text{a.e. in } \mathscr{D}, \tag{3.19}$$

$$v^n(x, \tau + \tau_i) \to v_1^n(x) \quad \text{a.e. in } \mathscr{D}, \tag{3.20}$$

$$v(x, \tau_i + 1) \to v_1(x) \quad \text{a.e. for } |x| \leqslant 1. \tag{3.21}$$

Let $f(x)$ be some test function depending only on x. Using (2.29) we have

$$\int_{-1}^{1} \rho(x) f(x)[v(x, \tau_i + 1) - v(x, \tau_i)]\, dx$$

$$= \int_{\tau_i}^{\tau_i + 1} \int_{-1}^{1} \left[v^n(x, \tau) f'' + \frac{1}{n-1} \rho v(x, \tau) f\right] dx\, d\tau. \tag{3.22}$$

It follows from (3.19)–(3.21), (2.30), and (3.9) that we can pass to the limit as $\tau_i \to \infty$ in (3.22). We obtain

$$\int_{\tau_i}^{\tau_i + 1} \int_{-1}^{1} \left[(v_1)^n f'' + \frac{1}{n-1} \rho v_1 f\right] dx\, dt = 0.$$

Therefore,

$$\int_{-1}^{1} \left(v_1^n f'' + \frac{1}{n-1} \rho v_1 f\right) dx = 0.$$

The last equality holds for every test function; thus $v_1(x)$ is a weak solution of Eq. (3.1). From the continuity of $v_1(x)$ it follows that $y(x) = v_1^n(x)$ is the unique classical solution of the problem

$$y'' = -[1/(n-1)]\, \rho(x) v_1(x),$$

$$y(\pm 1) = 0.$$

Therefore $v_1(x)$ is the classical solution of (3.1). Theorem 3.2 is thus proved.

By Theorem 3.1, $v_1(x)$ is the unique solution of (3.1)–(3.3) and thus the whole sequence $v(x, \tau)$ converges to $v_1(x)$ as $\tau \to \infty$.

Returning to the definition of $v(x, \tau)$ by (2.26) and (2.27) we obtain the assertion of Theorem 3.3

Remarks: 1. Theorem 3.3 may be also proved using a decreasing Liapunov functional (for a good survey and corresponding references see Ref. 21). Such an approach is equivalent to the one used in our work.

2. Note that via the study of the asymptotic behavior of the nonstationary problem, as a by-product we have obtained the existence of the solution of the eigenvalue problem (3.1), (3.2). In Ref. 17 the existence of the relevant eigenvalue problem is a consequence of a theorem by Amann.[22] This theorem is nonapplicable in our case because $\rho(x)$ may vanish on the boundary.

4. FINAL REMARKS

Consider the problem

$$\rho(x) \frac{\partial v}{\partial t} = \frac{\partial^2 v^n}{\partial x^2} + \frac{1}{n-1} \rho(x) v, \tag{4.1}$$

$$v(x, 0) = v_0(x), \tag{4.2}$$

$$v(\pm 1, t) = 0. \tag{4.3}$$

A weak solution of (4.1)–(4.3) can be defined by translating directly the definition of the weak solution of the problem (2.1)–(2.3). Similarly, as a consequence of Theorems (3.1)–(3.3) we obtain, if $v(x, t)$ is a solution of (4.1)–(4.3), then

$$v(x, t) \to v_1(x) \quad \text{as } t \to \infty$$

and $v_1(x)$ is the unique stationary solution of the Eq. (4.1). Note that the functions

$$v_+(x, t) \equiv \alpha[\cos \tfrac{1}{4}\pi x]^{1/n} \tag{4.4}$$

and

$$v_-(x, t) \equiv \alpha[\cos(\pi/2\beta)(x - x_0)]^{1/n} \tag{4.5}$$

are the appropriate super and subsolutions for (4.1).

The method we have presented may be used to study the stabilization of the solution of the equation

$$\rho(x)\frac{\partial v}{\partial t}=\frac{\partial^2 v^n}{\partial x^2}+\rho_1(x)v^s \tag{4.6}$$

with $\rho_1(x)$ satisfying the same conditions as $\rho(x)$ [see (2.3) and (2.4)] and $1\leqslant s<n$. Theorems (3.1) and (3.2) provide the existence and the uniqueness of the stationary solution of (4.6). The functions defined in (4.4) and (4.5) may be used here as super and subsolutions.

[1]P. H. Rutherford and D. F. Düchs, Nucl. Fusion **17**, 565 (1977).
[2]D. Dobrott, R. L. Miller, and J. M. Rawls, Phys. Fluids **20**, 1744 (1977).
[3]W. M. Manheimer and T. M. Antonsen, Phys. Fluids **22**, 957 (1979).
[4]H. Yamato, M. Ohta, and S. Mori, Nucl. Fusion **12**, 605 (1972).
[5]H. P. Furth, M. N. Rosenbluth, P. H. Rutherford, and W. Stodiek, Phys. Fluids **13**, 3020 (1970).
[6]N. Ohyabu, Nucl. Fusion **19**, 1491 (1979).
[7]L. Bromberg, D. R. Cohn, and J. Fischer, Nucl. Fusion **19**, 1359 (1979).
[8]Yu. I. Kolesnichenko and S. N. Reznik, in *Plasma Physics and Controlled Nuclear Fusion Research* (I.A.E.A., Vienna, 1977), p. 337.
[9]Ya. I. Kolesnichenko, S. N. Reznik, V. A. Yavorskij, Nucl. Fusion **16**, 105 (1976).
[10]P. Rosenau, Phys. Fluids **25**, 148 (1982).
[11]M. D. Rosen and J. M. Greene, Phys. Fluids **20**, 1466 (1977).
[12]C. L. Longmire, Elementary Plasma Physics (Wiley, New York, 1963), p. 228.
[13]R. A. Gerwin and R. C. Malone, Nucl. Fusion **19**, 155 (1979).
[14]J. G. Berryman, J. Math. Phys. **18**, 2108 1977).
[15]J. G. Berryman and C. J. Holland, Phys. Rev. Lett. **40**, 1720 (1978).
[16]J. G. Berryman and C. J. Holland, Arch. Rational Mech. Anal. **74**, 379 (1980).
[17]D. G. Aronson and L. A. Peletier, J. Differential Equations **39**, 378 (1981).
[18]P. Rosenau and S. Kamin, Commun. Pure Appl. Math. (to appear).
[19]O. A. Oleinik, A. S. Kalashnikov, and Yui-Lin Chzou, Izv. Akad. Nauk. SSSR, Ser. Mat. **22**, 667 (1958).
[20]M. Protter and H. Weinberger, *Maximum Principles in Differential Equations* (Prentice Hall, Englewood Cliffs, N.J., 1967).
[21]C. M. Dafermos, in *Asymptotic Behavior of Solutions of Evolution Equations, Nonlinear Evolution Equations*, edited by M. Crandall (Academic, New York, 1978), p. 103.
[22]H. Amann, Indiana Univ. Math. J. **21**, 125 (1971).