ANNOUNCEMENT

AIP Reduces Publication Charges

Lower page charges have been authorized for the six archival journals owned and published by the American Institute of Physics. The change, which amounts to a reduction of $20 per page, takes effect with articles published in February 1984 issues.

Substantial increases in subscription rates over the last three years were imposed as a necessary precursor to the loss of publication-charge income that will result from the lower page charges. The expense of publishing AIP journals has for many years been borne jointly by authors' institutions (through publication charges) and readers' institutions (through subscriptions). The effect of the recent changes is to alter the relative proportions of these two contributions to the total expense. Subscriptions to AIP and member-society journals, even after the increase imposed since 1981, are still two to five times less expensive per printed word than subscriptions to comparable journals of other publishers, and the number of copies distributed is much greater.

The publication charge for an article in an AIP journal is the sum of the article charge, which is unchanged at $20, and the page charge; the new rates are shown in the table below.

	Publication charge	
	Page Charge (per page)	Article Charge (per article)
Applied Physics Letters	$100	$20
Journal of Applied Physics	65	20
Journal of Mathematical Physics	65	20
The Physics of Fluids	65	20
Review of Scientific Instruments	65	20
The Journal of Chemical Physics	60	20

Authors whose institutions agree to honor the charges will continue to receive 100 free reprints of their articles.

H. WILLIAM KOCH,
Director,
American Institute of Physics

On the anomaly number of the classical groups

Hasan R. Karadayı[a)]
Institute for Nuclear Energy, Technical University of Istanbul, Yeni Levent, Istanbul, Turkey

(Received 8 June 1982; accepted for publication 21 January 1983)

The Adler–Bell–Jackiw (ABJ) anomalies for all representations of the classical groups are calculated with the aid of a technique proposed recently for the representation theory of these groups. The method presented here is useful also for the calculations of the eigenvalues of Casimir invariants with rank higher than 3.

PACS numbers: 02.20. + b

I. INTRODUCTION

Since 1971, we have known that the fundamental interactions can be described in a renormalizable manner[1] with the inclusion of the nonabelian local symmetry idea. However, there is another point which spoils renormalizability. This originates from a certain type of Feynman diagram and is consequently called "anomaly." There are two ways to prevent these triangle or ABJ anomalies[2] in a gauge model, one of which is the intrinsic L-R symmetry mechanism. In the absence of this mechanism, the representation content of

[a)] Supported by TBTAK, The National Science and Technology Council of Turkey.

the model should be chosen in such a way that its total contribution to anomalies would be equal to zero. Therefore, the calculation of the contribution for each irreducible representation has renewed interest[3] to the extent of the plethora of gauge models. This contribution is called "the anomaly number" of the corresponding irreducible representation. In this work, we give a simple formula for this anomaly number which operates all representations. Our basic framework for this is the novel method for the representation theory of groups that we recently proposed.[4] We must further emphasize that this simple method will prove useful also for the calculations of the eigenvalues of the Casimir invariants[5] with rank higher than 3.

II. ANOMALY NUMBER OF THE IRREDUCIBLE SU($N+1$) REPRESENTATIONS

The ABJ anomaly of a representation V of the Lie algebra

$$[T_a(V), T_b(V)] = if_{abc}T_c(V) \tag{2.1}$$

is defined to be

$$A(V)_{abc} = \mathrm{TR}\{T_a(V)(T_b(V)T_c(V) + T_c(V)T_b(V))\}, \tag{2.2}$$

where $T_a(V)$'s are the generators of this Lie algebra in the representation V. The main motivation here is the representation independence of $A(V)_{abc}$ up to a factor $A(V)$.[6] Hence, this factor $A(V)$ is determined with respect to a reference representation V_0 as being

$$A(V)_{abc} = A(V)\cdot A(V_0)_{abc} \tag{2.3}$$

and is called "the anomaly number" of the representation V. The reference representation V_0 is conventionally chosen to be the fundamental representation of the group.

In the first place, the fact that the real representations have always zero anomaly number is clear in view of definition (2.2). Consequently, only the representations of SU($N+1$) and spinorlike representations[7] of SO($4N+2$), $N=1,2,\ldots$ are subjects of the anomaly number calculations. It is a remarkable fact that E_6 has no anomaly in spite of the fact that it may have complex representations.

We now calculate the anomaly numbers of the elementary representations[8] of SU($N+1$). To calculate an $A(V)$, the restriction of the definition (2.2) to the Cartan subalgebra of the group is sufficient. Then the expression (2.3) will give

$$\sum_{A\in V} \mu_A{}^a\mu_A{}^b\mu_A{}^c = A(V)\sum_{A\in V_0}\mu_A{}^a\mu_A{}^b\mu_A{}^c, \tag{2.4}$$

where the weights $\mu_A{}^a$ are defined as

$$H(V)^a_{AB} \equiv \mu_A{}^a\delta_{AB}, \quad A, B = 1, 2, \ldots, \dim V \tag{2.5}$$

for the elements $H(V)^a$ of the Cartan subalgebra of the group. We formulate the weights for all irreducible SU($N+1$) representations in Ref. 4 in terms of the weights of the fundamental representation V_0. We now take the expression (2.4) in the form of

$$\sum_{A\in V}(M,\mu_A)^3 = A(V)\sum_{A\in V_0}(M,\mu_A)^3, \tag{2.6}$$

where the weight M is an appropriate one which we choose, say, as being the first weight μ_1 of the V_0. From now on, definition

$$(\mu_A,\mu_B) = \sum_{a=1}^{N}\mu_A{}^a\mu_B{}^a \tag{2.7}$$

are also adopted for any $A, B\in V$. On the other hand, we know from Ref. 4 that all weights of the elementary representations $V(\lambda_n)$ of SU($N+1$), where $n = 1, 2, \ldots, N$, can be expressed in the common form

$$\mu_{i_1} + \mu_{i_2} + \cdots + \mu_{i_n} \tag{2.8}$$

because the principal dominant weight of $V(\lambda_n)$ is expressed as

$$\lambda_n = \mu_1 + \mu_2 + \cdots + \mu_n. \tag{2.9}$$

Here the indices $i_1, i_2, \ldots, i_n$ take the values from 1 to $N+1$ while they all are different from each other within the same weight. With the aid of the scalar product (2.7), the expression

$$(\mu_A,\mu_B) = \delta_{AB} - 1/(N+1) \tag{2.10}$$

is clear. Then, the expression (2.6) immediately leads us to

$$A[V(\lambda_n)] \equiv A(\lambda_n) = \frac{(N-2)!}{(N-n)!(n-1)!}(N+1-2n), \tag{2.11}$$

which is the principal result of this article. Let us remark here that the following relation is also valid:

$$A(\lambda_{N+1-n}) = -A(\lambda_n). \tag{2.12}$$

We now see from these two expressions that $A(\lambda_2) = 1$ for $N=4$. This is the cancellation mechanism of the standard SU(5) model, which is shown first by Bouchiat *et al.*[9]

III. GENERALITIES

As is seen, our work does not terminate with a number of tables. The method presented in the last section would be applied equally well for the composite irreducible representations of SU($N+1$) and also for the Casimir invariants of all orders. What is needed in these computations is only the "orbital decomposition" of a composite representation. A composite representation originates from a dominant weight Λ having several subdominant weights Λ_i. Then, the decomposition of the representation $V(\Lambda)$ to the Weyl orbits $W(\Lambda_i)$ of each subdominant weight becomes

$$V(\Lambda) = W(\Lambda) + \sum_i m(\Lambda_i)W(\Lambda_i), \tag{3.1}$$

where $m(\Lambda_i)$'s are the multiplicities of the subdominants.

For instance, the computation of the quadratic and quartic indices,[10] which are defined to be

$$l_2(V) = \sum_{\mu\in V}(\mu,\mu)$$

and

$$l_4(V) = \sum_{\mu\in V}(\mu,\mu)^2,$$

are very clear in view of such an orbital decomposition. One remark for this is that all weights of a Weyl orbit have the same length with their dominant weight. Hence, for the representation having orbital decomposition (3.1), the quadratic and quartic indices will be obtained for $n = 1, 2$ as in the following:

$$l_{2n} = \dim W(\Lambda) \cdot (\Lambda, \Lambda)^n + \sum_i m(\Lambda_i) \cdot \dim W(\Lambda_i) \cdot (\Lambda_i, \Lambda_i)^n.$$

In Ref. 4, we show that the dimension formulas for the Weyl orbits are simple permutational computations in the light of a lemma for the weights of a Weyl orbit. On the other hand, we know that the scalar products between fundamental dominant weights will be calculated by the aid of the inverse Cartan matrix. Let us consider now the representation $V(\lambda_2)$ of E_6.[11] Its orbital decomposition is

$$V(\lambda_2) = W(\lambda_2) + 5W(\lambda_5), \tag{3.2}$$

and hence its indices are simply

$$l_{2n} = 216 \times (10/3)^{2n} + (351 - 216) \times (4/3)^{2n}$$

for $n = 1, 2$. This naive computation works similarly for all representations provided that the corresponding orbital decomposition is known. The cases for which the dimension of the corresponding representation is great is not a point here.

We think that there is no need to repeat the computations of the last section for composite or reducible SU$(N+1)$ representations. However, such an explicit example may be given for E_6. This is interesting because E_6 representations have no anomaly, even when they are complex. This fact can be investigated by the branchings SU(3)×SU(3)×SU(3) of irreducible E_6 representations.[12] We now explicitly calculate that a complex representation of E_6 has always zero anomaly number. There is an important remark here: every Weyl orbit always has zero anomaly number for E_6 and, consequently, every representation does the same thing. Let us consider again the representation $V(\lambda_2)$ of E_6. We will now see that $V(\lambda_2)$ has no anomaly number because the orbits $W(\lambda_2)$ and $W(\lambda_5)$ have zero anomaly number in spite of the fact they are complex. For this, we need only the suborbital decompositions of these orbits because the Weyl orbits other than SU$(N+1)$ expose a suborbital structure.[13] We have shown in Ref. 13 that the weights of the orbit $W(\lambda_5)$ can be grouped as

$$\begin{aligned} \mathrm{SW}(\lambda_5) &= \lambda_6 - \imath_1, \\ \mathrm{SW}(\lambda_2 - \lambda_6) &= \imath_1 + \imath_2 - \lambda_6, \\ \mathrm{SW}(\lambda_5 - \lambda_6) &= -\imath_1, \end{aligned} \tag{3.3}$$

while the ones for $W(\lambda_2)$ as

$$\begin{aligned} \mathrm{SW}(\lambda_2) &= \imath_1 + \imath_2, \\ \mathrm{SW}(\lambda_1 + \lambda_4 - \lambda_6) &= \imath_1 - \imath_2 - \imath_3 + \lambda_6, \\ \mathrm{SW}(2\lambda_1 - \lambda_6) &= 2\imath_1 - \lambda_6, \\ \mathrm{SW}(\lambda_3 + \lambda_5 - 2\lambda_6) &= \imath_1 + \imath_2 + \imath_3 - \imath_4 - \lambda_6, \\ \mathrm{SW}(\lambda_1 + \lambda_4 - 2\lambda_6) &= \imath_1 - \imath_2 - \imath_3, \\ \mathrm{SW}(\lambda_2 - 2\lambda_6) &= \imath_1 + \imath_2 - 2\lambda_6. \end{aligned} \tag{3.4}$$

Consequently, the dimensions of the orbits are, respectively,

$$6 + 15 + 6 = 27,$$

$$15 + 60 + 6 + 60 + 60 + 15 = 216.$$

In these expressions, all $\imath$ indices take the values from 1 to 6 and the corresponding definitions are

$$\begin{aligned} 1 &= \lambda_1, \\ 2 &= -\lambda_1 + \lambda_2, \\ 3 &= -\lambda_2 + \lambda_3, \\ 4 &= -\lambda_3 + \lambda_4 + \lambda_6, \\ 5 &= -\lambda_4 + \lambda_5 + \lambda_6, \\ 6 &= -\lambda_5 + \lambda_6, \end{aligned} \tag{3.5}$$

when they are expressed in terms of fundamental dominant weights λ_i of E_6. Their scalar products are subjects of the relation

$$(\imath, \jmath) = \delta_{ij} + \tfrac{1}{3}. \tag{3.6}$$

These can be investigated directly by the inverse Cartan matrix of E_6.

We now calculate, for example, the anomaly of $W(\lambda_5)$, just as in Sec. II. This anomaly will consist of three contributions coming from each suborbit in (3.3):

$$\sum_{\imath \in W(\lambda_5)} (1, \imath)^3 = A(\lambda_5) + A(\lambda_2 - \lambda_6) + A(\lambda_5 - \lambda_6),$$

where

$$A(\lambda_5) = \sum (1, \lambda_6 - \imath_1)^3,$$

$$A(\lambda_2 - \lambda_6) = \sum (1, \imath_1 \imath_2 - \lambda_6)^3,$$

$$A(\lambda_5 - \lambda_6) = \sum (1, -\imath_1)^3.$$

The result is

$$\sum_{\imath \in W(\lambda_5)} (1, \imath)^3 = \tfrac{13}{9} + \tfrac{10}{9} - \tfrac{23}{9} = 0,$$

where each contribution is calculated with the aid of (3.6). The same result would be obtained for any orbit of E_6.

[1] G. 't'Hooft, Nucl. Phys. B **35**, 167 (1971); B. W. Lee and J. Zinn-Justin, Phys. Rev. D **5**, 3121 (1972); **5**, 3137 (1972).

[2] S. L. Adler, Phys. Rev. **177**, 2426 (1969); J. S. Bell and R. Jackiw, Nucl. Phys. A **60**, 47 (1969); S. L. Adler and W. A. Bardeen, Phys. Rev. **182**, 1517 (1969).

[3] J. Patera and R. T. Sharp, J. Math. Phys. **22**, 2352 (1981); E. Eichten *et al.*, FERMILAB-Pub-81/83-THY BROWN-HET-464, 1981.

[4] H. R. Karadayı, ICTP, Trieste, Preprint IC/81/213.

[5] L. C. Biedenharn, J. Math. Phys. **4**, 436 (1963); S. Okubo, University of Rochester Report COO-3065-161, 1977; S. Okubo, University of Rochester Report, 1981.

[6] S. Okubo, Phys. Rev. D **16**, 3528 (1977); J. Banks and H. Georgi, Phys. Rev. D **14**, 1159 (1976).

[7] H. R. Karadayı, ICTP, Trieste, Preprint IC/81/224.

[8] As explained in Ref. 4, an elementary representation consists of the orbit of a dominant weight. These are the ones originating from the fundamental dominant weights $\lambda_1, \lambda_2, \ldots, \lambda_N$ for the chain SU$(N+1)$.

[9] C. Bouchiat *et al.*, Phys. Lett. B **38**, 519 (1972).

[10] E. B. Dynkin, Amer. Math. Soc. Transl. Ser. 2, **6**, 111 (1957); J. Patera *et al.*, J. Math. Phys. **17**, 1972 (1976); Erratum **18**, 1519 (1977).

[11] The definition of E_6 here is the same as in J. Patera and D. Sankoff, *Tables of Branching Rules for Representations of Simple Lie Algebras* (Les Presses de L' Université de Montreal, Montreal).

[12] J. Patera and R. T. Sharp, J. Math. Phys. **22**, 2352 (1981).

[13] H. R. Karadayı, Institute for Nuclear Energy, Report No. 30, 1982. The suborbital structure of the representations of the exceptional, orthogonal, and symplectic groups has been given in this paper.

Algebras with anticommuting basal elements, space-time symmetries, and quantum theory

G. P. Wene
Division of Mathematics, Computer Science and Systems Design, The University of Texas at San Antonio, San Antonio, Texas 78285

(Received 23 November 1982; accepted for publication 17 March 1983)

A collection $\mathscr{A}$ of algebras with anticommuting basal elements is investigated. It is shown that the collection $\mathscr{A}$ includes the quaternions, the octonions, the "algebra of color," as well as other algebras familiar to the physicist. Each algebra $\mathscr{A}$ is a quadratic, Jordan-admissible algebra and possesses a norm that is a generalization of the Minkowski metric. Using a Cayley–Dickson-like process, each such algebra A in $\mathscr{A}$ can be embedded into a larger algebra $\widehat{A}$ that is also in the collection $\mathscr{A}$. These algebras should provide candidates for models to describe observables, color, and other phenomena encountered in particle physics.

PACS numbers: 02.10.Tq, 02.10.Ws

I. INTRODUCTION

Let $\mathscr{A}$ denote the collection of algebras A such that each A in $\mathscr{A}$ is an algebra over a field Φ not of characteristic two. Furthermore, each such algebra A has a basis $\{e, u_1, u_2, ..., u_n\}$ over Φ such that e is the identity of A and

$$u_i u_j = -u_j u_i \quad \text{if } i \neq j, \quad i,j = 1, 2, ..., n, \tag{1.1}$$

and, for each $i = 1, 2, ..., n$

$$u_i^2 = \alpha_i e, \quad \alpha_i \in \Phi, \tag{1.2}$$

and each α_i is equal to $+1$ or -1.

Ilamed and Salingaros[1] construct all algebras in $\mathscr{A}$ which can arise in physics for $n = 3$. They are (i) the quaternions, (ii) the dihedral Clifford algebra N_1, which is related to the real 2-spinors, and (iii) the algebra of Pauli matrices S_1, which is related to the complex 2-spinors. Wene[2] shows that the algebra of color (see Domokos and Kövesi-Domokos[3]) is also in $\mathscr{A}$ for $n = 6$. There are many more algebras in $\mathscr{A}$ of interest to physicists besides these four.

We show in Sec. II that the so-called Cayley–Dickson process can be used to construct algebras in the collection $\mathscr{A}$. As a result, we note that most of the algebras used by physicists and constructed by the Cayley–Dickson process are in $\mathscr{A}$.

Section III is a discussion of the more significant properties of the algebras in $\mathscr{A}$. All the algebras in $\mathscr{A}$ are Jordan-admissible and a physical interpretation of this property is discussed.

Section IV gives necessary and sufficient conditions that these algebras be flexible.

II. A GENERALIZATION OF THE CAYLEY–DICKSON PROCESS

Let $A \in \mathscr{A}$ and $x \in A$. Let $-: A \rightarrow A$ be given by, for $x = x_0 e + \Sigma_{i=1}^{n-1} x_i u_i$,

$$\bar{x} = 2x_0 e - x \tag{2.1}$$

$$= x_0 e - \sum_{i=1}^{n-1} x_i u_i. \tag{2.2}$$

By the (generalized) Cayley–Dickson process construct an algebra $A(\mu)$ of dimension $2n$ over Φ having A as a subalgebra (with $e \in A$) as follows: $A(\mu)$ consists of all ordered pairs $x = (a_1, a_2)$, a_1, a_2 in A, addition and multiplication by scalars defined componentwise, and multiplication defined by

$$(a_1, a_2)(a_3, a_4) = (a_1 a_3 + \mu a_4 \bar{a}_2, \bar{a}_1 a_4 + a_3 a_2) \tag{2.3}$$

for all a_1, a_2, a_3, a_4 in A and some $\mu \in \Phi$. Then $e = (e, 0)$ is an identity element of $A(\mu)$, $A' = \{(a, 0) | a \in A\}$ is a subalgebra of $A(\mu)$ isomorphic to A, $v = (0, e)$ is an element of $A(\mu)$ such that $v^2 = \mu e$, and $A(\mu)$ is the vector space direct sum

$$A(\mu) = A + vA'. \tag{2.4}$$

If we identify A' with A, the elements of $A(\mu)$ are of the form

$$x = a_1 + va_2, \quad a_1, a_2 \text{ in } A, \tag{2.5}$$

and multiplication is given by

$$(a_1 + va_2)(a_3 + va_4) = a_1 a_3 + \mu a_4 \bar{a}_2 + v(\bar{a}_1 a_4 + a_3 a_2). \tag{2.6}$$

Theorem 2.1. If A is in $\mathscr{A}$, then so are $A(-1)$ and $A(1)$.

Proof. A basis for $A(-1)$ over Φ is

$$e, ve, u_i, vu_i, \quad i = 1, 2, ..., n-1.$$

By direct computations we see that

$$(ve)^2 = \mu = -1,$$
$$(vu_i)^2 = \mu u_i \bar{u}_i = -\mu \alpha_i = \alpha_i,$$
$$(vu_i)(vu_j) = -(vu_j)(vu_i),$$
$$(vu_i)u_j = -u_j(vu_i).$$

Similarly for $A(1)$.

If we begin with the real numbers $\mathbb{R}$, $\mathbb{R}(-1)$ is simply the complex numbers $\mathbb{C}$, $\mathbb{R}(-1)(-1) = \mathbb{R}(-1, -1)$ is the quaternions, $\mathbb{R}(-1, 1)$ the split quaternions, $\mathbb{R}(-1, -1, -1)$ the octonions, and $\mathbb{R}(-1, -1, -1, -1)$ the sedenions. Since $\mathbb{R}$ is (trivially) in $\mathscr{A}$, so is each of these algebras.

Conway[4] derived the Dirac equation and fine structure in terms of quaternions in 1948. Ilamed and Salingaros[1] demonstrate how the quaternions may arise in physics. The reader who is unfamiliar with the applications of the octonions in physics is referred to the article by Sorgsepp and Lôhmus.[5] For a discussion of the 16 dimensional sedenions see Sorgsepp and Lôhmus.[6] Brown[7] determines specific conditions that the 16-dimensional algebras be division algebras.

The algebra $\mathbb{R}(1)$ is the only abelian Clifford algebra other than $\mathbb{R}$ and $\mathbb{R}(-1)$. Salingaros[8] constructs the algebras $\mathbb{R}(1, -1)$ and $\mathbb{R}(1, -1, 1)$. The algebra $\mathbb{R}(1, -1)$ is isomorphic to the split quaternions, and $\mathbb{R}(1, -1, 1)$ is isomorphic to the split octonions. The split quaternions and space-time symmetries are discussed by Jantzen.[9] The article by Sorgsepp and Lôhmus[10] is a nice introduction to the real algebras constructed via the Cayley–Dickson process.

Each algebra A in $\mathscr{A}$ is a quadratic algebra in the sense of Theorem 2.2.

Theorem 2.2. Let A be an algebra in $\mathscr{A}$. Then A has an identity e and each x in A satisfies

$$x^2 = 2t(x)x - q(x)e, \tag{2.7}$$

where $t(x)$ and $q(x)$ are elements of Φ.

Proof. Let $x = x_0 e + \sum_{i=1}^{n} x_i u_i$. Then

$$x^2 = 2x_0 x + \left[\left(\sum_{i=1}^{n} x_i^2 \alpha_i\right) - x_0^2\right] e. \tag{2.8}$$

The trace of x, $t(x)$, is a linear functional on A. The trace is called associative if

$$t((xy)z) = t(x(yz)). \tag{2.9}$$

The trace induces a trace form $t(x,y)$ on A via

$$t(x,y) = t(xy) \tag{2.10}$$

for all x, y in A. Likewise the norm $q(x)$ defines a symmetric bilinear form $q(x,y)$ on A:

$$q(x,y) = q(x+y) - q(x) - q(y). \tag{2.11}$$

Say $q(x)$ is nondegenerate if $q(x,y)$ is.

Let $\mathscr{C}$ be a quadratic algebra and let $-: \mathscr{C} \to \mathscr{C}$ be given by

$$\bar{x} = 2t(x)e - x \tag{2.12}$$

for all x in $\mathscr{C}$. We can repeat the previous construction to get $\mathscr{C}(\mu)$ for $\mu \in \Phi$. The following two theorems describe the relation of $\mathscr{C}(\mu)$ to $\mathscr{C}$.

Theorem 2.3. Let $\mathscr{C}$ be a quadratic algebra and $\mu = 0$. Then $\mathscr{C}(0)$ contains an ideal N such that $N^2 = 0$ and

$$\mathscr{C}(0) \cong N \oplus \mathscr{C}(0)/N,$$

where $\mathscr{C}(0)/N$ is is isomorphic to $\mathscr{C}$.

Proof. The set $N = \{vc \mid c \in \mathscr{C}\}$ is an ideal and $N^2 = 0$.

Theorem 2.4. Let $\mathscr{C}$ **be a quadratic algebra over** Φ **and** $\mu \in \Phi$, $\mu \neq 0$.

(i) $\mathscr{C}(\mu)$ is a quadratic algebra over Φ and the linear form t can be extended to $\mathscr{C}(\mu)$ via

$$t(c_1 + vc_2) = t(c_1). \tag{2.13}$$

Similarly, $q(x,y)$ can be extended to $\mathscr{C}(\mu)$ by

$$q(c_1 + vc_2)(c_3 + vc_4) = q(c_1, c_3) - q(c_2, c_4). \tag{2.14}$$

(ii) If $\mathscr{C}$ has any one of the properties:

(1) $t(xy) = t(yx)$,

(2) $t((xy)z) = t(x(yz))$,

(3) $\mathscr{C}$ is flexible,

(4) q is nondegenerate,

(5) $t(x,y)$ is nondegenerate,

then $\mathscr{C}(\mu)$ has the same property.

(iii) If $c \to \bar{c}$ is an involution in $\mathscr{C}$, then $c_1 + vc_2 \to \bar{c}_1 - vc_2$ is an involution in $\mathscr{C}(\mu)$.

Proof. This is just Theorem 3.6 on page 218 of Braun and Koecher.[11]

III. THE ALGEBRAS IN $\mathscr{A}$

Theorem 2.2 says that each algebra A in $\mathscr{A}$ is quadratic. Each quadratic algebra of dimension 3 or more over its center is a central simple algebra; a quadratic algebra of dimension 5 or more over its center must be nonassociative. All quadratic algebras are strictly power associative. Gürsey[12] proposed that the nonobservability of isolated quarks must be associated with a nonassociative algebra.

If B is an algebra over Φ, let B^+ denote the vector space B over Φ with a new multiplication, $\circ$, given by

$$a \circ b = \tfrac{1}{2}(ab + ba) \tag{3.1}$$

for all a, $b \in B$. If B^+ is a Jordan algebra, we say that B is Jordan-admissible. Jordan[13] and Jordan, von Neumann, and Wigner[14] first showed that the set of observables form a Jordan algebra under the product (3.1). Faulkner,[15] starting with much weaker assumptions, shows that the set of observables has the structure of a quadratic Jordan algebra under the product (3.1) and is a normed linear space over the real numbers $\mathbb{R}$. If B^+ is a simple Jordan algebra, B is called J-simple. The classical discussions of Jordan algebras are Braun and Koecher[11] and Jacobson.[16]

All quadratic algebras are Jordan-admissible and, therefore, all algebras in $\mathscr{A}$ are quadratic, Jordan-admissible.

Many of the algebras in $\mathscr{A}$ arise in relativistic mechanics (see Kyrala[17] and Günaydin and Gürsey[18]). We observe that if A is an algebra in $\mathscr{A}$, and x is an element of A,

$$x = x_0 e + \sum_{i=1}^{n-1} x_i u_i. \tag{3.2}$$

The norm of x, $q(x)$, is given by

$$q(x) = x_0^2 - \sum_{i=1}^{n-1} x_i^2 \alpha_i, \tag{3.3}$$

where each α_i is $+1$ or -1. Recalling the Minkowski metric

$$d^2 s = c^2\, dt^2 - d\bar{R} \cdot d\bar{R}, \tag{3.4}$$

where $\bar{R}$ is a vector in 3-space, (3.3) is a Minkowski-like metric on a $n+1$ space.

The Cayley–Dickson process will allow one to construct algebras in $\mathscr{A}$ of very large dimensions. That algebras of large dimensions will probably be needed to describe observed phenomena is generally acknowledged. A specific example would be an algebra to model the hypercolor instantons in Weinberg[19]; with five hyperquarks for each of three colors and the corresponding anticolor components and a colorless natural element; such an algebra would require a basis of at least 31 elements. Horwitz and Biedenharn[20] show that a Hilbert space over the real Clifford algebra C_7 (a 128-dimensional algebra over $\mathbb{R}$) can provide models for the unification of weak, electromagnetic, and strong interactions utilizing the exceptional Lie groups.

We can attach yet another algebra to B. Denote by B^- the algebra with the same vector space as B but with a new product $[\ ,\ ]$ defined by

$$[a, b] = \tfrac{1}{2}(ab - ba), \tag{3.5}$$

where juxtaposition denotes the product of ab in B. If B^- is a Lie algebra, B is said to be Lie-admissible. The reader is referred to Myung, Okubo, and Santilli[21] for applications of Lie-admissible algebras in physics. Since the octonions are not Lie-admissible, the algebras in $\mathscr{A}$ are not, in general, Lie-admissible.

IV. THE FLEXIBLE ALGEBRAS IN $\mathscr{A}$

Albert[22] showed that each flexible, quadratic, J-simple algebra has a basis $u_0 = e, u_1, u_2, ..., u_n$, where

$$u_1^2 = \alpha_i e \quad \text{for } \alpha_i \neq 0, \quad \alpha_i \in \Phi$$

and $u_i u_j = -u_j u_i$ for $i \neq j, i, j = 1, 2, ..., n$. If the field Φ is algebraically closed, the u_i's can be chosen so that the α_i's are all $+1$. Particularly, over the field $\mathbb{C}$, the norm (3.3) becomes

$$q(x) = x_0^2 - \sum_{i=1}^{n-1} x_1^2. \tag{4.1}$$

We determine the flexible algebras in $\mathscr{A}$ and give necessary and sufficient conditions that an algebra in $\mathscr{A}$ be J-simple.

If B is an algebra of dimension n over Φ, let $u_0, u_1, u_2, ..., u_{n-1}$ be a basis of B over Φ. Then the multiplication in B is completely determined by the n^3 multiplication constants γ_{ijk} which appear in the products

$$u_i u_j = \sum_{k=0}^{n-1} \gamma_{ijk} u_k, \quad \gamma_{ijk} \in \Phi. \tag{4.2}$$

If A is an algebra in the collection $\mathscr{A}$, set $e = u_0$, then write (1.1) as

$$\gamma_{ijk} = -\gamma_{ijk}, \quad i \neq j, i, j = 1, 2, ..., n, \quad k = 0, 1, ..., n \tag{4.3}$$

and (1.2) gives

$$\gamma_{iik} = \delta_{0k} \alpha_i, \quad i, k = 0, 1, ..., n. \tag{4.4}$$

We recall that an algebra B is said to be flexible if, for each $x, y \in B$ we have

$$(x, y, x) = (xy)x - x(yx) = 0. \tag{4.5}$$

Expression (4.4) is equivalent to

$$(x, y, z) = -(z, y, x) \tag{4.6}$$

for all $x, y, z \in B$. We have the following:

Theorem 4.1. Let A be an algebra in $\mathscr{A}$. The following are equivalent:

(i) A is flexible.

(ii) The trace form $T(x, y) = T(xy)$ is symmetric and associative.

(iii) (a) $\gamma_{ij0} = \gamma_{iji} = \gamma_{ijj} = 0, \quad i \neq j, i, j = 1, ..., n$
(b) $\alpha_i \gamma_{jki} = \alpha_k \gamma_{ijk}, \quad i \neq j, i, j, k = 1, ..., n$
where a is satisfied if $k = i$ or $k = j$.

Proof. (i)$\Leftrightarrow$(ii). See Braun and Koecher.[11]
(i)$\Leftrightarrow$(iii). See Albert.[22]

Theorem 4.2. Let A be an algebra in $\mathscr{A}$. Then A is J-simple if and only if $t(x)$ is associative and $q(x)$ is nondegenerate.

Proof. Braun and Koecher,[11] p. 217.

CONCLUSION

We have seen that many of the models used in physics are members of the collection $\mathscr{A}$ of algebras. This collection should continue to provide models that reflect the Jordan algebra structure of observables.

Additionally, the Cayley–Dickson process can be used to construct new algebras with these same properties and that contain algebras with desirable properties.

[1] Y. Ilamed and N. Salingaros, "Algebras with three anticommuting elements. I. Spinors and quaternions," J. Math. Phys. **22**, 2091 (1981).
[2] G. P. Wene, "A generalization of the construction of Ilamed and Salingaros," J. Math. Phys. **24**, 221 (1983).
[3] G. Domokos and S. Kövesi," The algebra of color," J. Math. Phys. **19**, 1477–1481 (1978).
[4] Conway, Acta Pontificia Academia Scientiarum **XII**, 259–278 (1948).
[5] L. Sorgsepp and J. Lôhmus, "About nonassociativity in physics and Cayley–Graves octonions," Hadronic J. **2**, 1388–1459 (1979).
[6] L. Sorgsepp and J. Lôhmus, "Binary and ternary sedenions," Hadronic J. **4**, 327–353 (1981).
[7] R. B. Brown, "On generalized Cayley–Dickson algebras," Pac. J. Math. **20**, 415–422 (1967).
[8] N. Salingaros, "Algebras with three anticommuting elements. II. Two algebras over a singular field," J. Math. Phys. **22**, 2096–2100 (1981).
[9] R. T. Jantzen, "Generalized Quaternions and space-time symmetries," J. Math. Phys. 23, 1741–1746 (1982).
[10] L. Sorgsepp and J. Lôhmus, "Binary and ternary sedenions, "Hadronic J. **4**, 327–353 (1981).
[11] H. Braun and M. Koecher, *Jordan-Algebren* (Springer, Berlin, 1966).
[12] F. Gürsey, Orbis Scientiae Conference at Coral Gables, January 1976.
[13] P. Jordan, Z. Phys. **80**, 285 (1933).
[14] P. Jordan, J. von Neumann, and E. Wigner, "On an algebraic generalization of the quantum mechanical formalism," Ann. Math. **35**, 29 (1934).
[15] J. R. Faulkner, "Measurement systems and Jordan algebras," J. Math. Phys. **23**, 1617–1721 (1982).
[16] N. Jacobson, *Structures and Representations of Jordan Algebras*, A.M.S. Colloquiums Publ. Vol. 39 (A.M.S., Providence, RI, 1968).
[17] A. Kyrala, "Applications of Vectors, Matrices, Tensors and Quaternions," in *Theoretical Physics* (Saunders, Philadelphia, PA, 1967).
[18] M. Günaydin and F. Gürsey, "Quark statistics and octonions," Phys. Rev. D **9**, 3387 (1974).
[19] S. Weinberg, "Color and electroweak forces as a source of quark and lepton masses," University of Texas at Austin, 1981, preprint. See also M. Axenides, P. H. Cox, and A. Yildiz, Phys. Rev. Lett. **49**, 262 (1982).
[20] L. P. Horwitz and L. C. Biedenharn, "Exceptional gauge groups and quantum theory," J. Math. Phys. **20**, 269–298 (1979).
[21] H. C. Myung, S. Okubo, and R. M. Santilli, *Applications of Lie-admissible algebras in physics* (Hadronic, Nonamtum, MA, 1979), Vols. I and II.
[22] A. A. Albert, "Power-associative rings," Trans. Am. Math. Soc. 552–593 (1948).

Double-Gel'fand boson polynomials and the permutation group

R. S. Nikam and C. R. Sarma
Department of Physics, Indian Institute of Technology, Bombay 400 076, India

(Received 14 October 1982; accepted for publication 4 March 1983)

Double Gel'fand polynomials of boson operators spanning the irreducible representation $[m]$ of $U(n)$ in $U(n)*U(n)$ have been obtained using symmetrized linear combinations of Wigner operators of the permutation group. The normalized coefficients which occur in the polynomial representation have been expressed as linear combinations of the Young orthogonal representation matrix elements.

PACS numbers: 02.20. + b, 05.30.Jp

1. INTRODUCTION

The problem of boson polynomial representation of double Gel'fand states is one which has been extensively studied by Louck, Biedenharn, Moshinsky, and others,[1–3] over the last few years. These polynomials, which play an important role in generating a tensor algebra for unitary groups, are canonical basis states spanning an irreducible representation (irrep) $[m]$ of either $U(n)\otimes U(n)$, such that the product representation induces the symmetric representation $[N,0,\cdots,0]$ of $U(n^2)$. Basically the problem consists of determining the coefficients of the monomial terms $\pi^n_{i,j=1}(a_i^{\,j})^{\alpha_i^j}|0\rangle$ occurring in the boson polynomial (cf. Ref. 4 for notation) representation of the above states. Various schemes exist for determining these coefficients. A standard procedure[4] is to express these coefficients as matrix elements over a Gel'fand basis of a product of unit tensor operators. For a special class of terms such that $\alpha_i^{\,j}=0$ or 1 for all $[\alpha]$ and any $i,j=1,\ldots,n$, these matrix elements have been uniquely identified with the real orthogonal representation matrices of the permutation group S_N, on the subindices of either the upper or lower indices of the defining tensor monomials:

$$\{T^{(j)}_{(i)}|0\rangle \equiv T^{j_1j_2\cdots j_N}_{i_1i_2\cdots i_N}|0\rangle \equiv \prod_{k=1}^{n} a^{j_k}_{i_k}|0\rangle,$$
$$1\leqslant i_1\leqslant i_2\leqslant\cdots\leqslant i_N\leqslant n,$$
$$1\leqslant j_1\leqslant j_2\leqslant\cdots\leqslant j_N\leqslant n. \tag{1}$$

The natural question which arises is whether such a correspondence exists with a subset of permutations of S_N for cases where $\alpha_i^{\,j}>1$. If so, the next point to consider is whether, using the representations of S_N, a computationally simpler scheme results. In a recent paper, Antillon and Seligman[5] examined a number of alternative schemes for obtaining these coefficients and interrelationships among them. Though these alternatives recognize the dualism between unitary and permutation groups, they do not provide a direct answer to the above question.

In the present paper, we have attempted to utilize the dualism explicitly to generate an orthonormal double Gel'fand basis using doubly symmetrized linear combinations of the elements $\{e^m_{rs}|r,s=1,\ldots,f_m\}$ of the algebra of S_N for the representations $[m]$. An immediate fallout of this approach has been to identify the required coefficients as linear combinations of the representation matrix elements $[P]^m_{rs}$ of S_N over all P leading to the given monomial structure. These ideas have been developed in Sec. 2 and illustrated with examples. A brief discussion of the method is presented in Sec. 3.

2. DOUBLE GEL'FAND POLYNOMIALS

Consider the set of tensor monomials defined by Eq. (1), which are of rank N in both upper and lower sets of indices and span a representation space of $U(n^2)$. In view of the commutation relations satisfied by the boson operators

$$[\bar{a}^j_i, a^l_k] = \delta_{ik}\delta^{jl}, \quad [\bar{a}^j_i,\bar{a}^l_k] = [a^j_i, a^l_k] = 0, \tag{2}$$

we find that a simultaneous permutation of the upper and lower indices of $T^{j_1\cdots j_N}_{i_1\cdots i_N}$ leaves the monomial invariant:

$$PT^{(j)}_{(i)}|0\rangle = T^{P(j)}_{P(i)}|0\rangle = T^{j_{p(1)}j_{p(2)}\cdots j_{p(N)}}_{i_{p(1)}i_{p(2)}\cdots i_{p(N)}}|0\rangle = T^{(j)}_{(i)}|0\rangle. \tag{3}$$

This implies that only the symmetric representation $[N,0,\ldots,0]$ of $U(n^2)$ results from these tensor monomials. However, if the permutations $\bar{P}$ or $\underline{P}$ act only on the upper or lower indices, respectively, we obtain the results

$$\bar{P}T^{(j)}_{(i)}|0\rangle = \underline{P}^{-1}T^{(j)}_{(i)}|0\rangle, \tag{4a}$$

$$\underline{P}T^{(j)}_{(i)}|0\rangle = \bar{P}^{-1}T^{(j)}_{(i)}|0\rangle, \tag{4b}$$

following Eq. (3). The results of (4a) and (4b), in turn, lead to the fact that the Nth rank tensor space of upper and lower indices of $T^{(j)}_{(i)}$ are reducible and generate the irreps $[m]$ of $U(n)$. The reduction can be effected using suitable linear combinations of the Wigner operators of S_N,[6] defined in a normalized form as

$$e^m_{rs} = \sqrt{\frac{f_m}{N!}}\sum_P [P]^m_{rs} P, \tag{5}$$

where P is either $\bar{P}$ or $\underline{P}$ and $[P]^m_{rs}$ is the rsth element of the Young orthogonal representation matrix of $P\in S_N$ for the irrep $[m]$ of dimensionality f_m. Before examining the necessary linear combinations which generate the canonical basis for $[m]\times[m]\subset[N,0,\ldots,0]$ of $U(n^2)\downarrow U(n)\otimes U(n)$, consider the functions

$$\left|\begin{matrix}(r)\\ [\bar{m}]\\ (s)\end{matrix}\;;\; T^{(j)}_{(i)}\right\rangle = \sqrt{\frac{f_m}{N!}}\sum_P [P]^m_{rs}\bar{P}T^{(j)}_{(i)}|0\rangle, \tag{6}$$

where the bar on $[\bar{m}]$ is used to indicate that the $\bar{P}$ are being used. Using Eq. (4a), we obtain

 0022-2488/84/030417-05$02.50

$$\left|\begin{array}{l}(r)\\ [\bar{m}]\\ (s)\end{array}\ ;\quad T^{(j)}_{(i)}\right\rangle = \sqrt{\frac{f_m}{N!}}\sum_P [P]^m_{rs}\underline{P}^{-1}T^{(j)}_{(i)}|0\rangle$$

$$= \sqrt{\frac{f_m}{N!}}\sum_P [P]^m_{sr}\underline{P}T^{(j)}_{(i)}|0\rangle$$

$$= \left|\begin{array}{l}(s)\\ [\underline{m}]\\ (r)\end{array}\ ;\quad T^{(j)}_{(i)}\right\rangle, \tag{7}$$

thus showing that only one set of basis states resulting from the use of either $\bar{P}$ or $\underline{P}$ is at most linearly independent.

Operating a permutation $\bar{Q}$ on the right of the semicolon on the state defined by Eq. (6), we get the result:

$$\left|\begin{array}{l}(r)\\ [\bar{m}]\\ (s)\end{array}\ ;\quad \bar{Q}\,T^{(j)}_{(i)}\right\rangle = \sqrt{\frac{f_m}{N!}}\sum_P [P]^m_{rs}\bar{P}\bar{Q}\,T^{(j)}_{(i)}|0\rangle$$

$$= \sum_{t=1}^{f_m}[Q]^m_{st}\left|\begin{array}{l}(r)\\ [\bar{m}]\\ (t)\end{array}\ ;\quad T^{(j)}_{(i)}\right\rangle. \tag{8}$$

Thus, the basis defined by Eq. (7) transforms according to the Young orthogonal representation $[m]$ of S_N.

Similarly, the action of $\underline{Q}$ yields

$$\left|\begin{array}{l}(r)\\ [\bar{m}]\\ (s)\end{array}\ ;\quad \underline{Q}T^{(j)}_{(i)}\right\rangle = \left|\begin{array}{l}(s)\\ [\underline{m}]\\ (r)\end{array}\ ;\quad \underline{Q}T^{(j)}_{(i)}\right\rangle$$

$$= \sqrt{\frac{f_m}{N!}}\sum_P [P]^m_{sr}\underline{P}\underline{Q}T^{(j)}_{(i)}|0\rangle$$

$$= \sum_{t=1}^{f_m}[Q]^m_{rt}\left|\begin{array}{l}(s)\\ [\underline{m}]\\ (t)\end{array}\ ;\quad T^{(j)}_{(i)}\right\rangle$$

$$= \sum_{t=1}^{f_m}[Q]^m_{rt}\left|\begin{array}{l}(t)\\ [\bar{m}]\\ (s)\end{array}\ ;\quad T^{(j)}_{(i)}\right\rangle. \tag{9}$$

Thus a result similar to that of Eq. (8) again follows. Similar analyses and conclusions could be carried out for $\bar{Q}$ and $\underline{Q}$ acting from the left.

We now consider $T^{(j)}_{(i)}$ of Eq. (1), which is invariant under any $\bar{Q}\in S_{N'_1}\otimes S_{N'_2}\otimes\cdots\otimes S_{N'_n}$ and $\underline{Q}\in S_{N_1}\otimes S_{N_2}\otimes\cdots\otimes S_{N_n}$, where

$$\sum_{k=1}^{n}N'_k = \sum_{k=1}^{n}N_k = N. \tag{10}$$

This indicates that the indices i_k, j_k appear N_k, N'_k times, respectively, in the monomial. Using the results of Eqs. (8) and (9), it then follows that not all the basis states defined by Eq. (7) are linearly independent. In fact, all those standard Young tableaux (r) related by any $\bar{Q}\in S_{N'_1}\otimes S_{N'_2}\otimes\cdots\otimes S_{N'_n}$ yield only one independent state and, similarly, all those (s) related by $\underline{Q}\in S_{N_1}\otimes S_{N_2}\otimes\cdots\otimes S_{N_n}$ also yield just one state. Using arguments developed in earlier papers,[7,8] we can identify each of these sets leading to single Weyl tableaux. The question then arises as to what tableaux indices (r) and (s) to choose to indicate the Weyl states (m_1) and (m_2) of $[m]$. This has been considered in detail in the earlier paper[7] on Gel'fand basis. An extension of it leads us to define a doubly symmetrized Wigner operator as a linear combination

$$e^m_{(m_1)(m_2)} = \sum_{r\in(m_1)}\sum_{s\in(m_2)}a^m_{r(m_1)}a^m_{s(m_2)}e^m_{rs}, \tag{11}$$

where the summations are over all Young tableaux (r), (s) yielding the Weyl tableaux (m_1) and (m_2), respectively. A simple algorithm has been developed earlier[7,8] for determining $a^m_{r(m_1)}$, etc., using the invariance properties

$$e^m_{(m_1)(m_2)}\underline{Q} = e^m_{(m_1)(m_2)},\quad \underline{Q}\in S_{N_1}\otimes\cdots\otimes S_{N_n}, \tag{12}$$

$$\bar{Q}e^m_{(m_1)(m_2)} = e^m_{(m_1)(m_2)},\quad \bar{Q}\in S_{N'_1}\otimes\cdots\otimes S_{N'_n}. \tag{13}$$

Using these, we write the double Gel'fand state as

$$\left|\begin{array}{l}(m_1)\\ [\bar{m}]\\ (m_2)\end{array}\ ;\quad T^{(j)}_{(i)}\right\rangle$$

$$= \left[\prod_{k=1}^{n}N_k!\prod_{l=1}^{n}N'_l!\right]^{-1/2}(a^m_{r(m_1)}a^m_{s(m_2)})^{-1}\sqrt{\frac{f_m}{N!}}$$

$$\times\sum_P [P]^m_{rs}\bar{P}T^{(j)}_{(i)}|0\rangle. \tag{14}$$

We can now demonstrate that the above basis transforms as a set of basis states of the irrep $[m]$ under the action of the generators E^{cd} or E_{cd} of either component of $\mathrm{U}(n)\otimes\mathrm{U}(n)$. These generators are defined as (cf. Ref. 9, p. 122)

$$E^{cd} = \sum_{k=1}^{n}a^c_k\bar{a}^d_k, \tag{15a}$$

$$(c,d = 1,\ldots,n).$$

$$E_{cd} = \sum_{k=1}^{n}a^k_c\bar{a}^k_d, \tag{15b}$$

Considering the action of E^{cd} of Eq. (15a) on the state defined by Eq. (14), we find

$$E^{cd}\left|\begin{array}{l}(m_1)\\ [\bar{m}]\\ (m_2)\end{array}\ ;\quad T^{(j)}_{(i)}\right\rangle = N\sqrt{\frac{f_m}{N!}}\sum_P [P]^m_{sr}\underline{P}E^{cd}T^{(j)}_{(i)}|0\rangle, \tag{16}$$

where

$$N = \left[\prod_{k=1}^{n}N_k!\prod_{l=1}^{n}N_l!\right]^{-1/2}(a^m_{r(m_1)}a^m_{s(m_2)})^{-1}. \tag{17}$$

Observe that the right side of Eq. (16) is zero if $N_d = 0$. This follows from the definition of E^{cd} in Eq. (15a). If $N_d \neq 0$, then E^{cd} replaces, symmetrically, indexed by c, one at a time, so that we get N_d identical terms with $d\rightarrow c$, but not necessarily in the proper ordering defining a $T^{(j)}_{(i)}$ as in Eq. (1). A cyclic matching permutation, however, restores the order so that we have

$$E^{cd}\left|\begin{array}{l}(m_1)\\ [\bar{m}]\\ (m_2)\end{array}\ ;\quad T^{(j)}_{(i)}\right\rangle$$

$$= NN_d\sqrt{\frac{f_m}{N!}}\sum_P [P]^m_{sr}\underline{P}\underline{Q}T^{(j')}_{(i)}|0\rangle,$$

$$= NN_d\sum_{t=1}^{f_m}[Q]^m_{tr}\left|\begin{array}{l}(t)\\ [\underline{m}]\\ (s)\end{array}\ ;\quad T^{(j')}_{(i)}\right\rangle, \tag{18}$$

where the tensor monomial $T^{(j')}_{(i)}$ has all occupancies i_k, j_k the same as in $T^{(j)}_{(i)}$, except for c and d. For c and d, we have $N'_c = N'_{c+1}$, and $N'_d = N'_d - 1$. Using the procedure that is the inverse of the one leading to Eq. (14), we can express the right side of Eq. (18) as

$$E^{cd}\left|\begin{matrix}(m_1)\\ [\bar{m}] \\ (m_2)\end{matrix}\;;\quad T^{(j)}_{(i)}\right\rangle = [\,N_d(N_c+1)]^{1/2}(a^m_{r(m_1)})^{-1}$$
$$\times\sum_{m_3}\sum_{t\in m_3} a^m_{t(m_3)}[Q\,]^m_{tr}\left|\begin{matrix}(m_3)\\ [\bar{m}] \\ (m_2)\end{matrix}\;;\quad T^{(j')}_{(i)}\right\rangle. \tag{19}$$

Since the states $|{}^{[m]}_{(m_3)}\rangle$ are generated using the same Young diagrams as for $[m]$, which are stable under E^{cd}, these are the basis states for the irrep $[m]$ of $U(n)$ over the upper label of $T^{(j)}_{(i)}$. An exactly identical procedure can be carried out for E_{cd} of Eq. (15b), leading to the fact that the lower states $|{}^{[m]}_{(m_2)}\rangle$ provide basis for the irrep $[m]$ of $U(n)$ over the lower set of labels of $T^{(j)}_{(i)}$. Furthermore, considering the elementary generators $E^{c,c+1}$ and $E_{c,c+1}$, the equality can be demonstrated of the transformation coefficient in Eq. (19) to that of matrix elements of the elementary generators, calculated according to the prescription given in an earlier paper.[8] Thus Eq. (14) defines the double Gel'fand states terms of boson polynomials.

The occurrence of repeated indices in the upper and lower indices of $T^{(j)}_{(i)}$ can create difficulties in assessing the effect of $\bar{P}$ on the ordering of the upper indices. It is then more convenient to recast the right hand side of Eq. (14), using a set of matrices $[\alpha(P)]^{(j)}_{(i)}$, defined as

$$\begin{aligned}[\alpha(P)]^{(j)}_{(i)} &\equiv [\alpha(P)]^{(j_1j_2\cdots j_N)}_{(i_1i_2\cdots i_N)}\\ &= [\alpha]^N_{(i)}[P^{-1}]^{\mathrm{N}}_{\mathrm{N}}[\alpha]^{(j)}_N.\end{aligned} \tag{20}$$

Here $[\alpha]^N_i$ is an $(n\times N)$ matrix with zeros everywhere except unit entries at $(i_1, 1)$, $(i_2, 2)$,...,(i_N,N); $[P^{-1}]^N_N$ is an $(N\times N)$ matrix having zeros everywhere except unit entries at $(1, P^{-1}(1))$, $(2, P^{-1}(2))$,...,$(N, P^{-1}(N))$, and $[\alpha]^{(j)}_N$ is an $(N\times n)$ matrix with zeros everywhere except unit entries at $(1,j_1)$, $(2, j_2)$,...,(N,j_N). In terms of these $[\alpha(P)]^{(j)}_{(i)}$, we can reexpress Eq. (14) as

$$\left|\begin{matrix}(m_1)\\ [\bar{m}] \\ (m_2)\end{matrix}\;;\quad T^{(j)}_{(i)}\right\rangle = N\sqrt{\frac{f_m}{N!}}\sum_P [P\,]^m_{rs}\prod_{i,j=1}^{n}(a^j_i)^{\alpha^j_i(P)}|0\rangle, \tag{21}$$

where $\alpha^j_i(P)$ is the (i,j)th element of $[\alpha(P)]^{(j)}_{(i)}$.

Two cases of the matrices $[\alpha(P)]$ occurring in Eq. (21) need consideration. Firstly, if we consider a subset of monomials of Eq. (1) satisfying $1\leqslant i_1 < i_2z < \cdots < i_N \leqslant n$, $1\leqslant j_1 < j_2 \cdots < j_N \leqslant n$, we find that each $P\in S_N$ leads to distinct arrangements of the indices in $T^{(j)}_{(i)}$. This, in turn, means that there are $N!$ terms in each such boson polynomial and each coefficient is just $[P\,]^m_{rs}$ apart from an overall normalization constant $\sqrt{f_m/N!}$. This result is similar to the one established by Louck and Biedenharn.[2] The second case is that of the more general set of monomials $T^{(j)}_{(i)}$ of Eq. (1) with i_k, j_k $(k = 1,...,n)$ occurring, respectively, N_k, N'_k times. In this case, both lower and upper indices admit invariance groups of permutations as in Eqs. (12) and (13), respectively. This leads to the fact that a number of $[\alpha(P)]$ become identical and the corresponding coefficients of the monomials are linear combinations of Young representation matrices of S_N. Since these coefficients occur explicitly in the boson polynomials, it is necessary to identify the sets of permutations leading to distinct $T^{(j)}_{(i)}$. In order to do this, we proceed as follows:

(i) Determine the invariance group of permutations $\underset{\sim}{Q}\in S_{N_1}\otimes S_{N_2}\otimes\cdots\otimes S_{N_n}$ as defined in Eq. (12) and their right coset permutations in S_N.

(ii) Group these cosets into sets that lead to distinct arrangements of indices in $T^{(j)}_{(i)}$ taking into consideration the boson character of the $a_i{}^j$.

(iii) Operate with the permutations of Step (i) from the left on each of the elements of the sets defined as in Step (ii) to generate distinct arrangements of the upper indices of $T^{(j)}_{(i)}$ for a given ordering of the lower ones. Each set of these permutations $\{\bar{P}_{k\alpha}\,|\,k = 1,...,n_\alpha\}$ results in a single $[\alpha]$ matrix.

As an illustration of this procedure, consider $T^{1223}_{2233} = a_2{}^1a_2{}^2a_3{}^2a_3{}^3$ of $U(3)\otimes U(3)$. The invariance group as in Step (i) is $S_2\times S_2$: $\{e, (12), (34), (12)(34)\}$. The right coset permutations of this group may be chosen as $\{e, (13), (14), (23), (24), (13)(24)\}$. Using Step (ii), these are grouped as $\{e, (23)\}$, $\{(13)\}$, $\{(14, (13)(24)\}$, and $\{(24)\}$. Operating on these sets from the left with $\{e, (12), (34), (12)(34)\}$, as in Step (iii), we obtain the results:

$\{\bar{P}_1\} = \{e, (12), (34), (12)(34), (23), (123), (243), (1243)\}$,

$\{\bar{P}_2\} = \{(13), (132), (143), (1432)\}$,

$\{\bar{P}_3\} = \{(14), (142), (134), (1342), (13)(24), (1324), (1423), (14)(23)\}$,

$\{\bar{P}_4\} = \{(24), (124), (234), (1234)\}$.

It can be readily verified using Eq. (20), that these sets yield the following $[\alpha]$-matrices:

$$[\alpha_1] = \begin{pmatrix}0&0&0\\1&1&0\\0&1&1\end{pmatrix},\quad [\alpha_2] = \begin{pmatrix}0&0&0\\0&2&0\\1&0&1\end{pmatrix},$$
$$[\alpha_3] = \begin{pmatrix}0&0&0\\0&1&1\\1&1&0\end{pmatrix},\quad [\alpha_4] = \begin{pmatrix}0&0&0\\1&0&1\\0&2&0\end{pmatrix},$$

respectively.

Representing such sets of permutations as in Step (iii) by $\{P_{k\alpha}\,|\,k = 1,...,n_\alpha\}$, we can reexpress Eq. (21) as

$$\left|\begin{matrix}(m_1)\\ [\bar{m}] \\ (m_2)\end{matrix}\;;\quad T^{(j)}_{(i)}\right\rangle = N\sqrt{\frac{f_M}{N!}}\sum_{[\alpha]}A^{\,m}_{rs}(\alpha)\prod_{i,j=1}^{n}(a^j_i)^{\alpha_i{}^j}|0\rangle, \tag{22}$$

where N is as defined earlier in Eq. (17), $\alpha_i{}^j$ is the (i,j) element of $[\alpha]$ with

$$[\alpha] = [\alpha(\bar{P}_{k\alpha})]\quad (k = 1,...,n_\alpha), \tag{23}$$

and

$$A^{\,m}_{rs}(\alpha) = \sum_{k=1}^{n_\alpha}[P_{k\alpha}]^m_{rs}. \tag{24}$$

Before leaving this topic, it is worth illustrating the present approach using a simple example. Consider the basis state

$$\left| \begin{matrix} \left\{\begin{smallmatrix} 12 \\ 2 \end{smallmatrix}\right\} \\ [\,\overline{210}\,] \\ \left\{\begin{smallmatrix} 11 \\ 2 \end{smallmatrix}\right\} \end{matrix} \quad ; \quad T^{122}_{112} \right\rangle$$

of U(3) ⊗ U(3) expressed in Weyl tableaux notation. The T^{122}_{112} are symmetric under $(\underline{12})$ and $(\overline{23})$ so that symmetrization as in Eqs. (12) and (13) yields

$$e^{[21]}_{\left\{\begin{smallmatrix}12\\2\end{smallmatrix}\right\}\left\{\begin{smallmatrix}11\\2\end{smallmatrix}\right\}} = e^{[21]}_{\left\{\begin{smallmatrix}12\\2\end{smallmatrix}\right\}\begin{smallmatrix}12\\3\end{smallmatrix}} = \tfrac{1}{2} e^{[21]}_{\begin{smallmatrix}12 & 12\\2 & 3\end{smallmatrix}} + \frac{\sqrt{3}}{2} e^{[21]}_{\begin{smallmatrix}13 & 12\\2 & 3\end{smallmatrix}}.$$

Choosing $r = s = \begin{smallmatrix}12\\3\end{smallmatrix}$ for the representation [2, 1] of S_3, we have

$$a^{[2,1]}_{\begin{smallmatrix}12 & 12\\3 & 2\end{smallmatrix}} = \tfrac{1}{2} \text{ and } a^{[2,1]}_{\begin{smallmatrix}12 & 11\\3 & 2\end{smallmatrix}} = 1,$$

which yields the normalization factor of Eq. (17) to be

$$N = [2!2!]^{-1/2}(\tfrac{1}{2}\times 1)^{-1} = 1.$$

The invariance permutations as in Step (i) are $\{e, (12)\}$ and the right cosets of these in S_3 are $\{e, (13), (23)\}$. As per Step (ii), these can be grouped as $\{e, (23)\}$ and $\{(13)\}$. Operating on these sets from the left with $\{e, (12)\}$, as in Step (iii), we obtain

$$\{\bar{P}_1\} = \{e, (12), (23), (123)\},$$
$$\{\bar{P}_2\} = \{(13), (132)\}.$$

The distinct $[\alpha]$ matrices corresponding to these permutations are

$$[\alpha_1] = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad [\alpha_2] = \begin{pmatrix} 0 & 2 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Using the representation matrices for [2, 1] of S_3,[6] we readily obtain from Eq. (24) the result

$$A^{[21]}_{\begin{smallmatrix}12 & 12\\3 & 3\end{smallmatrix}}(\alpha_1) = 1, \quad A^{[21]}_{\begin{smallmatrix}12 & 12\\3 & 3\end{smallmatrix}}(\alpha_2) = -1$$

so that using Eq. (22) we have

$$\left| \begin{matrix} \left\{\begin{smallmatrix} 12 \\ 2 \end{smallmatrix}\right\} \\ [\,\overline{210}\,] \\ \left\{\begin{smallmatrix} 11 \\ 2 \end{smallmatrix}\right\} \end{matrix} \quad ; \quad T^{122}_{112} \right\rangle \equiv \left\langle \begin{matrix} & & 1 & & \\ & 2 & & 1 & \\ 2 & & 1 & & 0 \\ & 2 & & 1 & \\ & & 2 & & \end{matrix} \right\rangle = \frac{1}{\sqrt{3}}\{a_1{}^1 a_1{}^2 a_2{}^2 - a_1{}^2 a_1{}^2 a_2{}^1\}.$$

3. DISCUSSION

The approach outlined in Sec. 2 leading to the results of Eqs. (22)–(24) needs essentially the representation matrices of S_N. Procedures for handling general transpositions[10] and cyclic permutations[11,12] have now been programmed so that determining $A^{m}_{rs}(\alpha)$ of Eq. (24) is not too difficult a task. These coefficients are, apart from an overall multiplicative factor, identical with

$$C\begin{pmatrix} (m_1) \\ [m] \\ (m_2) \end{pmatrix}(\alpha)$$

obtained by Louck and Biedenharn.[2] Thus, the matrix elements of products tensor operators of U(n) over Gel'fand basis are basically linear combinations of Young orthogonal representations matrices over $\{\bar{P}_{k\alpha}\} | k = 1,...,n_\alpha\}$ satisfying Eq. (23). This result which is the generalization of the one obtained earlier[2] for $T^{(j)}_{(i)}$ $(1 \leqslant i_1 < i_2 < \cdots < i_N \leqslant n;$ $1 \leqslant j_1 < j_2 < \cdots < j_N \leqslant n)$, will, we hope, lead to a better understanding of the nature of the tensor operators of U(n). The interrelations obtained by Antillon and Seligman[5] between various procedures for obtaining the coupling coefficients

$$C\begin{pmatrix} (m_1) \\ [m] \\ (m_2) \end{pmatrix}(\alpha)$$

hold true in the present context also, since we have shown that it is related to one of them.

It should be pointed out that the present use of dualism between S_N and U(n) to generate double Gel'fand states is not entirely new. Bohr and Mottelson (cf. Ref. 9, pp. 130–131) dealt with a simple case of special Weyl tableaux states using such a dualism. Moshinsky[3] used this dualism in generating the special Gel'fand basis for the irreps of U(n). More recently, Patterson and Harter[13,14] have used unnormalized seminormal projection operators of S_N to generate canonical Weyl boson and fermion polynomials. These polynomials have to be individually normalized after collecting terms since the seminormal operators differ from permutation operators through a positive constant C_{rs} and no prescription has been given for determining them. Furthermore, an explicit identification of the permutations $\{P_{k\alpha}\} | k = 1,...,n_\alpha\}$ satisfying such an equation as our Eq. (23) has also not been carried out in their analysis; accordingly, a direct correspondence of their approach with other[5] is difficult.

ACKNOWLEDGMENTS

Our sincere thanks to Dr. J. D. Louck for his help in a clear understanding of the problem and to Dr. G. G. Sahasrabudhe for his help in the early stages of the study.

[1]J. D. Louck, Am. J. Phys. **38**, 3 (1970).
[2]J. D. Louck and L. G. Biedenharn, J. Math. Phys. **14**, 1336 (1973).

[3]M. Moshinsky, J. Math. Phys. **7**, 691 (1966).
[4]J. D. Louck, *Proceedings of Conference VII. International Colloquium on Group Theoretical Methods in Physics* (Springer-Verlag, Berlin, 1978).
[5]A. Antillon and T. H. Seligman, J. Math. Phys. **23**, 473 (1982).
[6]I. G. Kaplan, *Symmetry of Many Electron Systems* (Academic, New York, 1975).
[7]G. G. Sahasrabudhe, K. V. Dinesha, and C. R. Sarma, J. Phys. A **14**, 2189 (1981).
[8]C. R. Sarma and G. G. Sahasrabudhe, J. Math. Phys. **21**, 638 (1980).
[9]A. Bohr and B. R. Mottelson, *Nuclear Structure, Vol. I* (Benjamin, New York, 1969).
[10]S. Rettrup, Chem. Phys. Lett. **46**, 59 (1977).
[11]G. G. Sahasrabudhe, K. V. Dinesha, and C. R. Sarma, Theoret. Chim. Acta **54**, 333 (1980).
[12]S. Rettrup and C. R. Sarma, Theoret. Chim. Acta **46**, 73 (1977).
[13]C. W. Patterson and W. G. Harter, J. Math. Phys. **17**, 1125 (1976).
[14]C. W. Patterson and W. G. Harter, J. Math. Phys. **17**, 1137 (1976).

Infinitesimal operators and structure of the representations of the groups SO*(2n) and SO(2n) in a U(n) basis and of the groups SU*(2n) and SU(2n) in an Sp(n) basis

A. U. Klimyk and A. M. Gavrilik
Institute for Theoretical Physics, Kiev-130, USSR

(Received 13 October 1982; accepted for publication 18 February 1983)

The infinitesimal operators of the most degenerate representations of the groups SO*(2n) and SU*(2n) are found in a discrete basis. The structure (composition series) of these representations is studied. The classification of unitary irreducible representations of these groups which belong to most degenerate series is given. The infinitesimal operators of irredicuble unitary representations of SO(2n) in a U(n) basis and of SU(2n) in an Sp(n) basis are found for the cases of highest weights $(M, M, 0, ..., 0)$, $M \geqslant 0$.

PACS numbers: 02.20.Qs

I. INTRODUCTION

This paper deals with the representations of the groups SO*(2n) and SO(2n) in a U(n) basis, and of the groups SU*(2n) and SU(2n) in an Sp(n) basis.

The compact groups SO(n) and SU(n) are of great importance for different branches of physics (elementary particle theory, atomic physics, nuclear physics, and quantum chemistry). Higher unitary groups describe internal symmetries of elementary particles and their interactions, and the representations of these groups provide us with the corresponding quantum numbers. Different compact groups and their representations are involved in the nonabelian gauge theories which have been intensively developed during the last decade. Namely, the groups SU(n) and SO(n) underlie various models of grandunificating the fundamental forces (see, for example, Refs. 1–3). These groups also appear naturally in the extended supergravities.[4–6]

Considering physical applications of group representations we need various bases (corresponding to different subgroup reductions) of carrier spaces. For example, the reduction SO(10)⊃SU(5) is used to embed the Georgi–Glashow model into the Fritzsch–Minkowsky model. In our paper the unitary irreducible representations of SO(2n) are considered in a U(n) basis and that of SU(2n) in an Sp(n) basis. The highest weights are $(M,M,0,...,0)$, $M = 0,1,2,\cdots$.

On the other hand, the fact that the noncompact groups SO*(2n) and SU*(2n) have almost not been applied by physicists can be explained by complexity and poor knowledge of their representations. However, even the appearance of the groups SU*(6) and SO*(12), SO*(12)⊃U(6), as global symmetries[7] of the extended $N = 6$ supergravity in five- and six-dimensional space-time, respectively, supplies a good example of possible applications of these noncompact groups.

Here we shall investigate the representations of the most degenerate series (MDS) of the groups SO*(2n) and SU*(2n). The information on the degenerate series representations and their intertwining operators is contained in Refs. 8 and 9. We shall construct explicitly infinitesimal operators of the MDS representations of these groups in discrete bases. The formulas we obtain for the infinitesimal operators appear rather simple, and therefore the representations become visible and available for physical applications. Furthermore, the MDS representations of the groups SO*(2n) and SU*(2n) possess the feature that the spectrum of their restriction to the maximal compact subgroups U(n) and Sp(n), respectively, is simple (this fact is not valid for more general representations of these groups). It is this property which makes the use of the MDS representations of these groups in the dynamical group schemes possible.

Relative simplicity of the infinitesimal operators of the MDS representations allows us to study their structure and to obtain in the matrix form all the intertwining operators. By means of the latter we extract all the unitary representations which are contained in the MDS representations. In fact, we obtain the classification of the irreducible most degenerate unitary representations of SO*(2n) and SU*(2n).

In addition, we use the representations of SO*(2n) and SU*(2n) to derive the infinitesimal operators of the unitary irreducible representations of the group SO(2n) with highest weights $(M,M,0,...,0)$ in a U(n) basis and of the group SU(2n) with highest weights $(M,M,0,...,0)$ in an Sp(n) basis. For this purpose we use the method developed in Refs. 10–13. Let us note that the formulas for infinitesimal operators of the representations of SO*(2n), SO(2n), SU*(2n), and SU(2n) obtained here are valid for every K basis, where $K =$ U(n) for SO*(2n) and SO(2n), and $K =$ Sp(n) for SU*(2n) and SU(2n).

We remark that physicists sometimes denote the symplectic group Sp(n) as Sp(2n). Here we accept the notation and description of these and all other groups used by Helgason in Ref. 14. The concepts and statements of the representation theory of semisimple Lie groups can be found in Ref. 15.

Considering the groups SO*(2n) we put $n > 4$ whereas for SU*(2n) we put $n > 2$. The groups SO*(2n), $n \leqslant 4$, and SU*(2n), $n \leqslant 2$, are isomorphic to other groups[14] and their MDS representations have already been studied. Let us note that some of our results concerning SU*(2n) have been obtained independently (by other methods) by A. Guillemonat.[16]

 0022-2488/84/030422-09$02.50

II. THE REPRESENTATIONS OF THE MOST DEGENERATE SERIES OF THE GROUPS SO*(2n) and SU*(2n)

The group SO*(2n) consists of the matrices of SO(2n, C) which leave invariant the skew-Hermitian form[14]

$$-z_1\bar z_{n+1} + z_{n+1}\bar z_1 - z_2\bar z_{n+2} + z_{n+2}\bar z_2 - \cdots - z_n\bar z_{2n} + z_{2n}\bar z_n .$$

The group SU*(2n) consists of the matrices of SL(2n, C) which commute with the following transformation[14] of the space C^{2n}:

$$\psi:\ (z_1, \ldots, z_n, z_{n+1}, \ldots, z_{2n}) \to (\bar z_{n+1}, \ldots, \bar z_{2n}, -\bar z_1, \ldots, -\bar z_n).$$

The Lie algebra so*(2n) consists of the matrices[14]

$$\begin{pmatrix} Z_1 & Z_2 \\ -\bar Z_2 & \bar Z_1 \end{pmatrix}, \qquad (1)$$

where Z_1 are skew-symmetric complex $n\times n$ matrices and Z_2 Hermitian complex $n\times n$ matrices. The Lie algebra su*(2n) consists of the matrices[14] (1) for which Z_1 and Z_2 are complex $n\times n$ matrices with $\mathrm{Tr}(Z_1 + \bar Z_1) = 0$.

The maximal compact subgroup K of SO*(2n) is isomorphic to U(n). The group U(n) is embedded into SO*(2n) in the following manner: If $A + iB \in \mathrm{U}(n)$, where A and B are real matrices, then

$$A + iB \to \begin{pmatrix} A & B \\ -B & A \end{pmatrix} \in \mathrm{SO}^*(2n). \qquad (2)$$

The maximal compact subgroup K of SU*(2n) coincides with Sp(n).

Let G denote one of the groups SO*(2n), SU*(2n). Let $\mathfrak g$ be a Lie algebra of G and $\mathfrak k$ a Lie algebra of K. The algebra $\mathfrak g$ has the Cartan decomposition[14] $\mathfrak g = \mathfrak k + \mathfrak p$. For $\mathfrak g = \mathrm{so}^*(2n)$, $\mathfrak p$ consists of the matrices[14]

$$\begin{pmatrix} iX_1 & iX_2 \\ iX_2 & -iX_1 \end{pmatrix}, \quad X_1, X_2 \in \mathrm{so}(n),\ i = \sqrt{-1}. \qquad (3)$$

For $\mathfrak g = \mathrm{su}^*(2n)$, $\mathfrak p$ coincides with the set of matrices[14]

$$\begin{pmatrix} iZ_1 & Z_2 \\ -\bar Z_2 & -i\bar Z_1 \end{pmatrix}, \quad Z_1 \in \mathrm{su}(n),\ Z_2 \in \mathrm{so}(n, C). \qquad (4)$$

Let $G = ANK$ be an Iwasawa decomposition[15,17] of G, where A is a commutative subgroup and N a nilpotent subgroup. We shall consider the MDS representations of G. To construct them, we use the maximal parabolic subgroup

$$P = ANM_1(K) = A_1N_1M_1,$$

where A and N are the same as in the Iwasawa decomposition, A_1 is a one-dimensional subgroup of A, to be defined below, $M_1 \cap A = 1$, and $N_1 \subset N$. The M_1 is a maximal connected subgroup of G such that $m_1a_1 = a_1m_1$ for every $a_1 \in A_1$ and $m_1 \in M_1$. For $M_1(K)$ we have $M_1(K) = M_1 \cap K$.

The subgroup A can be defined as $A = \exp \mathfrak a$, where $\mathfrak a$ is a Lie algebra of A. For $G = \mathrm{SO}^*(2n)$ we can choose $\mathfrak a$ to consist of the matrices (3) for which $X_2 = 0$ and $X_1 = \mathrm{diag}(R_1, \ldots, R_{n/2})$, if n is even, and $X_1 = \mathrm{diag}(0, R_1, \ldots, R_{(n-1)/2})$, if n is odd. Here

$$R_j = \begin{pmatrix} 0 & t_j \\ -t_j & 0 \end{pmatrix}, \quad t_j \in R. \qquad (5)$$

For $A_1 = \exp \mathfrak a_1$ the Lie algebra $\mathfrak a_1$ consists of the matrices tH_1, $t \in R$, with

$$H_1 = \mathrm{diag}\left(0,\ldots,0,\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix},0,\ldots 0,\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\right). \qquad (6)$$

It is easy to verify that $M_1 \sim \mathrm{SO}^*(2n-4)\times \mathrm{SU}(2)\times A_1$, $M_1(K) \sim \mathrm{U}(n-2)\times \mathrm{SU}(2)$.

For $G = \mathrm{SU}^*(2n)$ we can choose $\mathfrak a$ to consist of the matrices

$$\mathrm{diag}(t_1,t_2,\ldots,t_n,t_1,t_2,\ldots,t_n), \quad t_j \in R, \qquad (7)$$

for which $t_1 + t_2 + \cdots + t_n = 0$. The Lie subalgebra $\mathfrak a_1$ consists of the matrices tH_1, $t \in R$, with

$$H_1 = \mathrm{diag}\left(\frac{-1}{n-1},\ldots,\frac{-1}{n-1},,\frac{-1}{n-1},\ldots,\frac{-1}{n-1},1\right). \qquad (8)$$

Now we have that $M_1 = \mathrm{SU}^*(2n-2)\times \mathrm{Sp}(1)\times A_1$, $M_1(K) = \mathrm{Sp}(n-1)\times \mathrm{Sp}(1)$.

The subgroup A of G can be represented as $A = A_1A_2$, where $A_2 = \exp \mathfrak a_2$, and $\mathfrak a_2$ consists of the matrices of $\mathfrak a$ which are orthogonal to $\mathfrak a_1$ (with respect to the Cartan–Killing bilinear form). In other words, $\mathfrak a_2$ consists of the matrices of $\mathfrak a$, for which the last coordinate t_j is equal to 0. It is clear that every element $h \in A$ can be decomposed uniquely into the product $h = h_1h_2$, $h_1 \in A_1$, $h_2 \in A_2$.

We consider the one-dimensional representation

$$h_1n_1m_1 \to \exp[\lambda(\log h_1)], \quad h_1 \in A_1,\ n_1 \in N_1,\ m_1 \in M_1,$$

of the parabolic subgroup P, where λ is a complex linear form on $\mathfrak a_1$. If $h_1 = \exp tH_1$, then

$$\exp[\lambda(tH_1)] = \exp \sigma t, \quad \sigma \in C. \qquad (9)$$

These representations of P are used to construct the MDS representations π_λ of G. They act in the space $L_0^2(K)$ which consists of the functions $f \in L^2(K)$ satisfying the condition

$$f(mk) = f(k), \quad m \in M_1(K). \qquad (10)$$

The operators $\pi_\lambda(g)$, $g \in G$, act upon $L_0^2(K)$ as

$$\pi_\lambda(g)f(k) = \exp[\lambda(\log h_1)]\,f(k_g), \qquad (11)$$

where $h_1 \in A_1$ and $k_g \in K$ are defined by the Iwasawa decomposition of the element kg: $kg = hnk_g = h_1h_2nk_g$, $h \in A$, $h_2 \in A_2$, $n \in N$. According to Eq. (9), π_λ are defined by one complex number σ. For pure imaginary $\sigma - \rho$ the representations π_λ form the principal unitary MDS (for definition of ρ, see Lemma 1 below).

III. PRELIMINARIES

In order to evaluate the infinitesimal operators for the representations π_λ, we shall make use of the Lemma 5.2 of Ref. 10. This lemma will be reproduced here. Let $B(.,.)$ be a Cartan–Killing form on $\mathfrak g$, and θ a Cartan involution related with the Cartan decomposition $\mathfrak g = \mathfrak k + \mathfrak p \equiv \mathbf k + \mathbf p$. Then

$$\langle x, y\rangle = -cB(x, \theta y), \quad c > 0, \qquad (12)$$

is a scalar product[14] on $\mathfrak g$. The adjoint representation of G in $\mathfrak g$ will be denoted by Ad. Now Lemma 5.2 of Ref. 10 for the representations π_λ can be formulated as follows.

Lemma 1: The infinitesimal operators $\pi_\lambda(Y)$, $Y \in \mathfrak p_c$ ($\mathfrak p_c$ is the complexification of $\mathfrak p$), of the representation π_λ act upon the infinitely differentiable functions of $L_0^2(K)$ as

$$\pi_\lambda(Y)f(k) = \langle(\mathrm{Ad}\,k)Y, H\rangle\lambda(H)f(k) - \langle(\mathrm{Ad}\,k)Y,\rho\rangle f(k) + \tfrac{1}{2}[Q,\langle(\mathrm{Ad}\,k)Y, h\rangle]f(k), \quad (13)$$

where H is a normalized element of $\mathbf{a}_1$, h is an element of $\mathbf{a}_1$ such that $\alpha(h) = 1$ [α is a simple restricted root[15] of the pair $(\mathbf{g}, \mathbf{a}_1)$], Q is identical to the operator Q_1 of formula (5) of Ref. 12, ρ is half the sum of the positive restricted roots of the pair $(\mathbf{g}, \mathbf{a}_1)$ (including multiple roots), and [.,.] denotes the commutator. Here $(\mathbf{g}, \mathbf{a}_1)\equiv(\mathfrak{g}, \mathfrak{a}_1)$.

Now we have to choose a basis of the space $L_0^2(K)$. This space has a basis which consists of all the matrix elements of irreducible unitary representations of K, which satisfy the condition (10). It is clear that this condition is satisfied by the matrix elements of those representations of K, which contain the identity representation of $M_1(K)$.

Lemma 2: The identity (one-dimensional) representation of $\mathrm{U}(n-2)\times\mathrm{SU}(2)$ is contained in the representations of $\mathrm{U}(n)$ with highest weights $(m_1,m_2,\dot{0},m_3,m_4)$, $m_1 - m_2 = m_3 - m_4$. (Here $\dot{0}$ denotes the part of the highest weight which consists of zeros.) The multiplicity of the identity representation is equal to 1.

Proof: The reduction $\mathrm{U}(n)\supset\mathrm{U}(n-1)\supset\mathrm{U}(n-2)\supset\cdots$ in the Gel'fand–Zetlin patterns shows that the identity representation of $\mathrm{U}(n-2)$ is contained in the representations of $\mathrm{U}(n)$ with highest weights $(m_1,m_2,\dot{0},m_3,m_4)$. Let us consider all the Gel'fand–Zetlin patterns of the form

$$\begin{bmatrix} m_1 & & m_2 & \dot{0} & m_3 & & m_4 \\ & m_1' & & \dot{0} & & m_2' & \\ & & & \dot{0} & & & \\ & & & \cdots & & & \end{bmatrix}.$$

From the action of the diagonal infinitesimal operators E_{jj} onto these patterns we can easily find the irreducible representations of the subgroup

$$\mathrm{U}(n-2)\times\mathrm{U}(1)\sim\mathrm{diag}(\mathrm{U}(n-2), u, u^{-1}), \quad u\in\mathrm{U}(1),$$

with the identity component for $\mathrm{U}(n-2)$, which are contained in the representation of $\mathrm{U}(n)$ with highest weight $(m_1,m_2,\dot{0},m_3,\mathrm{m}_4)$. Since the reduction $\mathrm{SU}(2)\supset\mathrm{U}(1)$ is well known, this result can be easily generalized to the reduction $\mathrm{U}(n)\supset\mathrm{U}(n-2)\times\mathrm{SU}(2)$. This gives the assertion of the lemma.

Lemma 3: The identity representation of $\mathrm{Sp}(n-1)\times\mathrm{Sp}(1)$ is contained in the representations of $\mathrm{Sp}(n)$ with highest weights $(m, m, \dot{0})$. The multiplicity is equal to 1.

The proof is given in Ref. 7. (This lemma follows also from Theorem 4.1 of Chap. X in Ref. 14 since $K/M_1(K)$ is a symmetric space.)

The irreducible representation of $\mathrm{U}(n)$ with highest weight $(m_1,m_2,\dot{0},m_3,m_4)$ and of $\mathrm{Sp}(n)$ with highest weight $(m, m, \dot{0})$ will be denoted by $[M]$. This implies that M is the corresponding highest weight. Let Ω be an orthonormal vector of the carrier space of $[M]$, which is invariant with respect to $M_1(K)$. Its uniqueness follows from the Lemmas 2 and 3. Let us choose an arbitrary, but fixed, orthonormal basis in the carrier space of $[M]$. Denote the elements of this basis by $|\Sigma\rangle$. The functions

$$(\dim[M])^{1/2}\langle\Omega\,|[M](k)|\Sigma\rangle = (\dim[M])^{1/2}D^M_{\Omega,\Sigma}(k) \quad (14)$$

satisfy the condition (10). The set of all functions (14), for all $[M]$ and Σ, forms an orthonormal basis of $L_0^2(K)$. The basis functions (14) will be denoted henceforth by $|M,\Sigma\rangle$. The restriction of π_λ onto K acts upon $L_0^2(K)$ according to the formula $\pi_\lambda(k_0)f(k) = f(kk_0)$. Therefore, $\pi_\lambda|_K$ does not change M in $|M,\Sigma\rangle$. The representation $\pi_\lambda|_K$ contains (with unit multiplicity) all the irreducible representations $[M]$ of Lemma 2 for $G = \mathrm{SO}^*(2n)$ and of Lemma 3 for $G = \mathrm{SU}^*(2n)$, and only these.

Now use Lemma 1 to derive an explicit expression for the infinitesimal operators $\pi_\lambda(Y)$ in the basis $|M,\Sigma\rangle$. The scalar product (12) is given by

$$\langle X, Y\rangle = b\,\mathrm{Tr}\,X\,\overline{Y}^T, \quad (15)$$

where we choose $b = \frac{1}{4}$ for $\mathrm{so}^*(2n)$ and $b = \frac{1}{2}$ for $\mathrm{su}^*(2n)$.

Let $G = \mathrm{SO}^*(2n)$. Determine the matrices H and h of Lemma 1. According to Eq. (15),

$$H = h = H_1, \quad (16)$$

where H_1 is given by Eq. (6), and a simple restricted root α of the pair $(\mathrm{so}^*(2n), \mathbf{a}_1)$ is defined by $\alpha(h) = 1$. The formula $\alpha(h') = \langle h_\alpha, h'\rangle, h'\in\mathbf{a}_1$, defines the correspondence between α and the element $h_\alpha\in\mathbf{a}_1$. It is clear that $h_\alpha = H_1$.

Now let $G = \mathrm{SU}^*(2n)$. We find that

$$H = \left(\frac{n-1}{n}\right)^{1/2} H_1, \quad h = \frac{n-1}{n} H_1, \quad h_\alpha = H_1, \quad (17)$$

where H_1 is given by Eq. (8), and a simple restricted root α is defined by $\alpha(h) = 1$.

Now we consider the summands of Eq. (13). It is easy to verify that (see Ref. 12)

$$\langle(\mathrm{Ad}\,k)Y, H\rangle\lambda(H) = \langle(\mathrm{Ad}\,k)Y, h\rangle\lambda(h_\alpha), \quad (18)$$

$$\langle(\mathrm{Ad}\,k)Y,\rho\rangle = \tfrac{1}{2}(r+2s)\langle h_\alpha, h_\alpha\rangle\langle(\mathrm{Ad}\,k)Y, h\rangle, \quad (19)$$

where r is the multiplicity of the root α and s the multiplicity of the root[15] 2α. The Lie algebras $\mathrm{so}^*(2n)$ and $\mathrm{su}^*(2n)$ are of the types DIII and AII, correspondingly. Using the root systems of these algebras (see Ref. 15, pp. 30–32, and Table 3 of Ref. 10) we find that

$$\mu\equiv\tfrac{1}{2}(r+2s)\langle h_\alpha, h_\alpha\rangle = \begin{cases} n - 3/2 & \text{for } \mathrm{so}^*(2n), \\ n & \text{for } \mathrm{su}^*(2n). \end{cases} \quad (20)$$

Since we consider the degenerate series of representations, the chain (2) of subgroups in Ref. 12 reduces to

$$\mathrm{U}(n) = K \equiv K_1 \supset K_2 = M_1(K) = \mathrm{U}(n-2)\times\mathrm{SU}(2) \quad \text{for } \mathrm{SO}^*(2n),$$

$$\mathrm{Sp}(n) = K \equiv K_1 \supset K_2 = M_1(K) = \mathrm{Sp}(n-1)\times\mathrm{Sp}(1) \quad \text{for } \mathrm{SU}^*(2n).$$

Moreover, in the case of the group $\mathrm{SO}^*(2n)$ there is the subgroup $K_{\frac{1}{2}} = \mathrm{U}(n-2)\times\mathrm{U}(2)$ between the subgroups K_1 and K_2 [see the chain (3) of subgroups in Ref. 12]. In the case of the group $\mathrm{SU}^*(2n)$ the subgroup $K_{\frac{1}{2}}$ is absent [as for the case of the group $\mathrm{SL}(n,R)$ in Ref. 13]. These chains of subgroups are used to define eigenvalues of the operator Q. The operator Q acts upon the state $|M, \Sigma\rangle$ as

$$Q|M, \Sigma\rangle = q(M)|M, \Sigma\rangle, \tag{21}$$

where $q(M)$ is a number. From Eqs. (13), (18)–(21) it follows that

$$\pi_\lambda(Y)|M, \Sigma\rangle = [\lambda(h_\alpha) - \mu + \tfrac{1}{2}Q - \tfrac{1}{2}q(M)] \times \langle(\mathrm{Ad}\, k)Y, h\rangle|M, \Sigma\rangle. \tag{22}$$

IV. INFINITESIMAL OPERATORS FOR THE REPRESENTATIONS π_λ of SO*(2n)

Since $[\mathbf{k}, \mathbf{p}] \subset \mathbf{p}$, a finite-dimensional representation of $K = \mathrm{U}(n)$ is realized in $\mathbf{p}_c$. In order to find this representation, we use the transformation $\varphi: g \to \tau g \tau^{-1}$, where

$$\tau = \frac{1}{2}\begin{pmatrix} E_n & iE_n \\ E_n & -iE_n \end{pmatrix}$$

and E_n is the unit $n \times n$ matrix. Under φ the matrices (2) of K transform into the matrices

$$\begin{pmatrix} A - iB & 0 \\ 0 & A + iB \end{pmatrix}, \quad A + iB \in \mathrm{U}(n), \tag{23}$$

and the matrices (3) of $\mathbf{p}$ transform into the matrices

$$\begin{pmatrix} 0 & iX_1 - X_2 \\ iX_1 + X_2 & 0 \end{pmatrix}. \tag{24}$$

Using the realizations (23) and (24) of K and $\mathbf{p}$, we find by direct evaluation that the representation Ad of $K = \mathrm{U}(n)$ in $\mathbf{p}_c$ is a direct sum of two irreducible representations of K with highest weights $(1,1,\dot{0})$ and $(\dot{0}, -1, -1)$. In order to construct the space $\mathbf{p}_c$, we have to complexify the matrices (24). Every element of $\mathbf{p}_c$ can be decomposed uniquely into a sum of matrices $\left(\begin{smallmatrix} 0 & X \\ 0 & 0 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 0 & 0 \\ Y & 0 \end{smallmatrix}\right)$, where X and Y are complex skew-symmetric $n \times n$ matrices. The representations of $\mathrm{U}(n)$ with highest weights $(1,1,\dot{0})$ and $(\dot{0}, -1, -1)$ are realized in the spaces $\mathbf{p}_1$ and $\mathbf{p}_2$ of matrices $\left(\begin{smallmatrix} 0 & 0 \\ X & 0 \end{smallmatrix}\right)$ and of matrices $\left(\begin{smallmatrix} 0 & Y \\ 0 & 0 \end{smallmatrix}\right)$, respectively. A similar situation appears in Ref. 18 for the group $\mathrm{Sp}(n, R)$.

By direct computation one can find the correspondence between the basis elements of $\mathbf{p}_1$ and $\mathbf{p}_2$, and Gel'fand–Zetlin patterns for the representations $(1,1,\dot{0})$ and $(\dot{0}, -1, -1)$, respectively. For example, the Gel'fand–Zetlin pattern

$$\begin{bmatrix} 1 & & 1 & & \dot{0} \\ & 1 & & \dot{0} & \\ & & \dot{0} & & \\ & & \cdots & & \end{bmatrix} \tag{25}$$

corresponds to the matrix $\left(\begin{smallmatrix} 0 & 0 \\ X & 0 \end{smallmatrix}\right)$, where $X = E_{21} - E_{12}$. Here E_{ij} is the matrix with elements $(E_{ij})_{st} = \delta_{is}\delta_{jt}$. The Gel'fand–Zetlin pattern

$$\begin{bmatrix} \dot{0} & & -1 & & -1 \\ & \dot{0} & & -1 & \\ & & \dot{0} & & \\ & & \cdots & & \end{bmatrix} \tag{26}$$

corresponds to the matrix $\left(\begin{smallmatrix} 0 & Y \\ 0 & 0 \end{smallmatrix}\right)$, where $Y = E_{21} - E_{12}$. Under φ the matrix $H_1 = h$, Eq. (6), transforms into a sum of the matrices corresponding to the Gel'fand–Zetlin patterns (25) and (26).

Elements of an orthonormal basis of the space $\mathbf{p}_1$ (which may be different from the Gel'fand–Zetlin basis) will be denoted by I_s^+, and of the space $\mathbf{p}_2$ by I_s^-, $s = 1,2,...,\dim \mathbf{p}$.

Now we apply the formula (22) to $Y = I_s^\pm$. The expression $\langle(\mathrm{Ad}\, k)I_s^+, h\rangle$ is a matrix element of the representation of $\mathrm{U}(n)$ with highest weight $(1,1,\dot{0})$. Since $|M,\Sigma\rangle$ in Eq. (14) is the matrix element of the representation $[M]$ of $\mathrm{U}(n)$, then

$$\begin{aligned}\langle(\mathrm{Ad}\, k)I_s^+, h\rangle|M, \Sigma\rangle &= \sum_{M',\Sigma'} (\dim[M]/\dim[M'])^{1/2} \\ &\quad \times \langle M, \Omega; \{1,1\}, h\,|M', \Omega\rangle \\ &\quad \times \langle M', \Sigma'|M, \Sigma; \{1,1\}, I_s^+\rangle|M', \Sigma'\rangle,\end{aligned} \tag{27}$$

where $\langle\cdots|\cdots\rangle$ are Clebsch–Gordan coefficients (CGC's) for the tensor product of the irreducible representations of $\mathrm{U}(n)$ with highest weights $M = (m_1, m_2, \dot{0}, m_3, m_4)$, $m_1 - m_2 = m_3 - m_4$, and $(1,1,\dot{0})$. [The second representation is denoted by $\{1,1\}$ in Eq. (27).] Let us note that the multiplicities in this tensor product do not exceed 1. The first CGC in Eq. (27) does not change Ω since h is invariant with respect to $M_1(K)$. For I_s^- we have to replace the representation $\{1,1\}$ in Eq. (27) by $(\dot{0}, -1, -1) \equiv \{-1, -1\}$.

The summation in Eq. (27) extends over all vectors $|M', \Sigma'\rangle$ for which the CGC's are nonzero. Because of the first CGC, M' is of the form $(m_1', m_2', \dot{0}, m_3', m_4')$, $m_1' - m_2' = m_3' - m_4'$. It is a consequence of an invariance of the vector $|M', \Omega\rangle$ with respect to $M_1(K)$. Using the Clebsch–Gordan series for the tensor product $[M] \otimes \{1,1\}$ [see Eq. (3.5) in Ref. 10], we find that the summation in Eq. (27) extends over the following, $M' = (m_1', m_2', \dot{0}, m_3', m_4')$:

$$\begin{aligned}&(m_1 + 1, m_2 + 1, \dot{0}, m_3, m_4), \quad (m_1 + 1, m_2, \dot{0}, m_3 + 1, m_4), \\ &(m_1, m_2, \dot{0}, m_3 + 1, m_4 + 1), \quad (m_1, m_2 + 1, \dot{0}, m_3, m_4 + 1).\end{aligned} \tag{28}$$

For I_s^- we have the tensor product $[M] \otimes \{-1, -1\}$ and the summation extends over the following M':

$$\begin{aligned}&(m_1 - 1, m_2 - 1, \dot{0}, m_3, m_4), \quad (m_1 - 1, m_2, \dot{0}, m_3 - 1, m_4), \\ &(m_1, m_2, \dot{0}, m_3 - 1, m_4 - 1), \quad (m_1, m_2 - 1, \dot{0}, m_3, m_4 - 1).\end{aligned} \tag{29}$$

We substitute Eq. (27) into Eq. (22). Now the operator Q acts upon the vectors $|M', \Sigma'\rangle$. We have to evaluate the numbers

$$\tfrac{1}{2}[q(M') - q(M)] \tag{30}$$

for all M' from Eqs. (28) and (29). The formula (50) of Ref. 12 can be used to find them. We use the notations from there and refer the reader to Ref. 12 for the meaning of these notations. For our case $\Lambda^1 = (m_1, m_2, \dot{0}, m_3, m_4)$ and $\Lambda^2 = m_1 + m_2 + m_3 + m_4$. [Since $K_2^1 = \mathrm{U}(n-2) \times \mathrm{U}(2) \sim M_1(K) \times \mathrm{U}(1)$, then Λ^2 is the highest weight for the group $M_1(K) \times \mathrm{U}(1)$. The identity representation corresponds to $M_1(K)$. Therefore, we leave in Λ^2 only the weight corresponding to the subgroup $\mathrm{U}(1)$.] The values of all summands of the formula (50) in Ref. 12 for our case are given in Table I.

Now we introduce for $|M, \Sigma\rangle$ the notation $|J, S, S'; \Sigma\rangle$, where

$$\begin{aligned}&J = m_1 - m_2 = m_3 - m_4, \quad S = m_1 + m_2, \\ &S' = -m_3 - m_4.\end{aligned} \tag{31}$$

TABLE I. Quantities for determining the operator Q for the case of the group SO*(2n).

$\Lambda^1+\tau^1$	$\langle\Lambda^1+\rho^1,\tau^1\rangle$	$\langle\tau^1,\tau^1\rangle/2$	$\langle\Lambda^2+\rho^2,\tau^2\rangle/2$	$\langle\tau^2,\tau^2\rangle/4$
m_1+1, m_2+1, m_3, m_4	m_1+m_2+n-2	1	$(m_1+m_2+m_3+m_4)/2$	$\frac{1}{2}$
m_1, m_2, m_3+1, m_4+1	m_3+m_4-n+2	1	$(m_1+m_2+m_3+m_4)/2$	$\frac{1}{2}$
m_1+1, m_2, m_3+1, m_4	m_1+m_3+1	1	$(m_1+m_2+m_3+m_4)/2$	$\frac{1}{2}$
m_1, m_2+1, m_3, m_4+1	m_2+m_4-1	1	$(m_1+m_2+m_3+m_4)/2$	$\frac{1}{2}$
m_1-1, m_2-1, m_3, m_4	$-m_1-m_2-n+2$	1	$-(m_1+m_2+m_3+m_4)/2$	$\frac{1}{2}$
m_1, m_2, m_3-1, m_4-1	$-m_3-m_4+n-2$	1	$-(m_1+m_2+m_3+m_4)/2$	$\frac{1}{2}$
m_1-1, m_2, m_3-1, m_4	$-m_1-m_3-1$	1	$-(m_1+m_2+m_3+m_4)/2$	$\frac{1}{2}$
m_1, m_2-1, m_3, m_4-1	$-m_2-m_4+1$	1	$-(m_1+m_2+m_3+m_4)/2$	$\frac{1}{2}$

Using Eqs. (27), (22), and Table I, we obtain the explicit expressions for the infinitesimal operators:

$$\pi_\lambda(I_s^+)|J,S,S';\Sigma\rangle$$
$$=\sum_{\Sigma'}\left(\sigma+\frac{S+S'}{2}\right)C^{JSS'}_{JS+2S'}(\Sigma',s)|J,S+2,S';\Sigma'\rangle$$
$$+\sum_{\Sigma'}\left(\sigma-\frac{S+S'}{2}-2n+4\right)C^{JSS'}_{JSS'-2}(\Sigma',s)$$
$$\times|J,S,S'-2;\Sigma'\rangle$$
$$+\sum_{\Sigma'}(\sigma+J-n+3)C^{JSS'}_{J+1S+1S'-1}(\Sigma',s)$$
$$\times|J+1,S+1,S'-1;\Sigma'\rangle$$
$$+\sum_{\Sigma'}(\sigma-J-n+1)C^{JSS'}_{J-1S+1S'-1}(\Sigma',s)$$
$$\times|J-1,S+1,S'-1;\Sigma'\rangle, \tag{32}$$

$$\pi_\lambda(I_s^-)|J,S,S';\Sigma\rangle$$
$$=\sum_{\Sigma'}\left(\sigma-\frac{S+S'}{2}-2n+4\right)C^{JSS'}_{JS-2S'}(\Sigma',s)$$
$$\times|J,S-2,S';\Sigma'\rangle$$
$$+\sum_{\Sigma'}\left(\sigma+\frac{S+S'}{2}\right)C^{JSS'}_{JSS'+2}(\Sigma',s)|J,S,S'+2;\Sigma'\rangle$$
$$+\sum_{\Sigma'}(\sigma-J-n+1)C^{JSS'}_{J-1S-1S'+1}(\Sigma',s)$$
$$\times|J-1,S-1,S'+1;\Sigma'\rangle$$
$$+\sum_{\Sigma'}(\sigma+J-n+3)C^{JSS'}_{J+1S-1S'+1}(\Sigma',s)$$
$$\times|J+1,S-1,S'+1;\Sigma'\rangle, \tag{33}$$

where $C^{JSS'}_{J_1S_1S_1'}(\Sigma',s)$ is a numerical coefficient at $|M',\Sigma'\rangle$ in (27) which includes CGC's and a dimensionality multiplier. The summation in Eqs. (32) and (33) extends over all Σ' for which CGC's are nonzero.

It is clear that the restriction $\pi_\lambda|_{U(n)}$ decomposes into a direct sum of all irreducible representations of U(n) with highest weights $(m_1,m_2,\dot{0},m_3,m_4)$, $m_1-m_2=m_3-m_4$. According to this fact the corresponding parameters J, S, and S' run over all nonnegative integers of the same evenness, such that $J\leqslant S$, $J\leqslant S'$. Evenness of $S+S'$ and $S-J$ follows from the relations $S-J=2m_2$ and $S+S'=(m_1-m_2)-(m_3-m_4)+2m_2-2m_4=2(m_2-m_4)$. The range of (J, S, S') will be used in what follows.

V. INFINITESIMAL OPERATORS FOR THE REPRESENTATIONS π_λ of SU*(2n)

The derivation is similar to that in the case of the group SO*(2n). Therefore, we give only the most important arguments. Now the irreducible representation of $K=\mathrm{Sp}(n)$ with highest weight $(1,1,\dot{0})\equiv\{1,1\}$ is realized in $\mathbf{p}_c$. This follows, for example, from comparing of dimensionalities. Elements of an orthonormal basis of $\mathbf{p}_c$ will be denoted by E_j. Since h is not an orthonormal element, we substitute instead of h in Eq. (22) the expression $h=[(n-1)/n]^{1/2}H$ [see Eq. (17)]. Now for the matrices E_j we can write down the relation similar to Eq. (27). For convenience we shall refer to this relation as to Eq. (27'). This relation contains the CGC's of Sp(n) for the tensor product $[M]\otimes\{1,1\}$, $M=(m,m,\dot{0})$. Multiplicities in this tensor product do not exceed 1. In Eq. (27') the summation extends over the following $M'=(m',m',\dot{0})$:

$$(m+1,m+1,\dot{0}),\quad (m-1,m-1,\dot{0}),\quad (m,m,\dot{0}). \tag{34}$$

The subgroup K_2^1 is absent in the case of the group SU*(2n). Therefore, instead of Table I we have a simpler one (see Table II).

From Eqs. (22) and (27') and Table II we obtain the explicit formula for the infinitesimal operators $\pi_\lambda(E_j)$ in the basis $|M,\Sigma\rangle\equiv|m,\Sigma\rangle$:

$$\pi_\lambda(E_j)|m,\Sigma\rangle=\sum_{\Sigma'}(\sigma+m)C^m_{m+1}(\Sigma',j)|m+1,\Sigma'\rangle$$
$$+\sum_{\Sigma'}(\sigma-m-2n+1)C^m_{m-1}(\Sigma',j)$$
$$\times|m-1,\Sigma'\rangle$$
$$+\sum_{\Sigma'}(\sigma-n)C^m_m(\Sigma',j)|m,\Sigma'\rangle, \tag{35}$$

where

$$C^m_{m'}(\Sigma,j)=\{[(n-1)/n]\dim[m,m,\dot{0}]/\dim[m',m',\dot{0}]\}^{1/2}$$
$$\times\langle m,\Omega;\{1,1\},H|m',\Omega\rangle$$
$$\times\langle m',\Sigma'|m,\Sigma;\{1,1\},E_j\rangle. \tag{36}$$

TABLE II. Quantities for determining the operator Q for the case of the group SU*(2n).

$\Lambda+\tau$	$\langle\Lambda+\rho,\tau\rangle$	$\langle\tau,\tau\rangle/2$
$m+1, m+1, \dot{0}$	$m+n-\frac{1}{2}$	$\frac{1}{2}$
$m-1, m-1, \dot{0}$	$-m-n+\frac{1}{2}$	$\frac{1}{2}$
$m, m, \dot{0}$	0	0

It is clear that the restriction $\pi_\lambda|_{\mathrm{Sp}(n)}$ decomposes into a direct sum of all irreducible representations of Sp(n) with highest weights $(m,m,\dot{0})$, $m = 0,1,2,\cdots$.

VI. THE STRUCTURE OF THE REPRESENTATIONS π_λ OF THE GROUPS SO*(2n) AND SU*(2n)

Now we shall determine the subset of irreducible representations from among the set of representations π_λ, and investigate the structure of the reducible representations π_λ. The procedure to be followed is completely analogous to that which was followed in Ref. 10 in the case of the groups U(n,1) and $\mathrm{SO}_0(n,1)$. It will be more convenient to use the notation π^σ for the representations π_λ, where σ is defined by Eq. (9). The parameter σ runs over all complex numbers.

Consider the representations π^σ of the group SO*(2n).

Theorem 1. The representation π^σ of SO*(2n) is irreducible if and only if σ is not an integer or σ equals to $n-2$ or $n-1$.

The irreducibility of these representations π^σ is proved in the same manner as in the cases of the groups U(n,1) and $\mathrm{SO}_0(n,1)$ (see Chap. 7 in Ref. 10), and we omit the proof. The reducibility of other representations π^σ will be shown below.

It is known (see Refs. 9 and 10) that for real σ the representations π^σ and $\pi^{-\sigma+2n-3} \equiv \pi^{-\sigma+2\rho}$ (ρ was defined in Lemma 1) are reducible or irreducible simultaneously. Moreover, if π^σ and $\pi^{-\sigma+2n-3}$ are irreducible, then they are equivalent; if they are reducible, then they consist of the same irreducible representations of SO*(2n) (see Chap. 5 in Ref. 10). Thus it is sufficient to consider the representations π^σ, $\sigma \leqslant n-3/2$.

Now we investigate the representations π^σ of SO*(2n) for which σ is an integer and $\sigma < n-3/2$. Let us equate the coefficients in parentheses on the right-hand side of Eqs. (32) and (33) to 0:

$$\sigma + (S+S')/2 = 0, \tag{37}$$

$$\sigma - (S+S')/2 - 2n + 4 = 0, \tag{38}$$

$$\sigma + J - n + 3 = 0, \tag{39}$$

$$\sigma - J - n + 1 = 0. \tag{40}$$

Above it was shown that $(S+S')/2 \geqslant 0$. Since $\sigma < n-3/2$, Eq. (38) cannot be fulfilled. Equation (40) can be valid only for $\sigma = n-1$, $J=0$; the coefficient $\sigma - J - n + 1$ stands at the vectors $|J-1, S+1, S'-1; \Sigma'\rangle$ and $|J-1, S-1, S'+1; \Sigma'\rangle$ in Eqs. (32) and (33). Since $J \geqslant 0$, these vectors at $J=0$ makes no sense. Hence we do not need Eq. (40) to study the structure of the reducible representations π_λ.

Now we give a graphic picture of the range of (J,S,S'). Since S and S' enter Eqs. (32) and (33) as $(S+S')/2$, we consider the plane $((S+S')/2, J)$. The integers S, S', and J are of the same evenness, and $(S+S')/2 \geqslant J$. Therefore, the range of (J, S, S') extends below the line $0A$ in Fig. 1, including this line. It is clear that every admissible point $((S+S')/2, J)$ corresponds to a number of points (J, S, S'), namely $(J, S+p, S'-p)$, $|p| = 0, 1, 2, \cdots$, give the same point $((S+S')/2, J)$.

Case 1: σ is an integer and $\sigma \leqslant 0$. The lines corresponding to Eqs. (37) and (38) are shown in Fig. 1. They divide the set of points $((S+S')/2, J)$ into three parts. We denote these

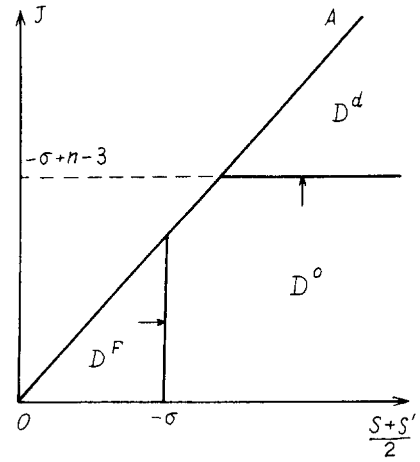


FIG. 1. Structure of the representation π^σ of SO*(2n) if σ is a nonpositive integer.

domains (parts) by D^F, D^0, D^d. The points which lie on the boundaries belong to that domain which contains the arrow pointing to the boundary. Note that at $\sigma = 0$ the line (37) reduces to the point (0,0).

Now consider the points from the domain D^F. From Eqs. (32) and (33) it is easy to see that the first summand of Eq. (32) and the second summand of Eq. (33) (and only these) increase the sum $(S+S')/2$ in the vector $|J,S,S';\Sigma\rangle$. These summands vanish at $-\sigma = (S+S')/2$ (see Fig. 1). Hence, the operators $\pi^\sigma(I_s^\pm)$ cannot transform the vectors $|J,S,S';\Sigma\rangle$, $((S+S')/2, J) \in D^F$, into the vectors $|J,S,S';\Sigma\rangle$, $((S+S')/2, J) \in D^0 \cup D^d$. This means that the vectors $|J,S,S';\Sigma\rangle$, $((S+S')/2, J) \in D^F$, form a basis of the subspace of $L_0^2(K)$, which is invariant with respect to $\pi^\sigma(I_s^\pm)$, $s = 1,2,\ldots,\dim \mathbf{p}$. Hence, π^σ realizes the finite-dimensional representation of SO*(2n) in this subspace; we denote it by D_σ^F. It is clear from Eqs. (32) and (33) that the operators $\pi^\sigma(I_s^\pm)$ can transform the vectors $|J,S,S';\Sigma\rangle$, $((S+S')/2, J) \in D^0$, into the vectors $|J,S,S';\Sigma\rangle$, $((S+S')/2, J) \in D^F$.

From Eqs. (32) and (33) it is easy to see that the third summand of Eq. (32) and the fourth summand of Eq. (33) (and only those) increase the value of $(S+S')/2$ in the vector $|J,S,S';\Sigma\rangle$. These summands vanish at $J = -\sigma + n - 3$ (see Fig. 1). Therefore, on the vectors $|J,S,S';\Sigma\rangle$, $((S+S')/2, J) \in D^F \cup D^0$, the representation π^σ realizes the subrepresentation; we denote it by $D_\sigma^{0,F}$. The quotient representation $D_\sigma^{0,F}/D_\sigma^F$ will be denoted by D_σ^0, and the quotient representation $\pi^\sigma/D_\sigma^{0,F}$ by D_σ^d.

The representations D_σ^F, D_σ^0, and D_σ^d are irreducible. The situation is completely similar to the case of the group U(p, q) in Ref. 19. Therefore, we omit the proof.

The representations π^σ in this case have the following structure:

$$\begin{pmatrix} D_\sigma^F & * & 0 \\ 0 & D_\sigma^0 & * \\ 0 & 0 & D_\sigma^d \end{pmatrix}, \tag{41}$$

where $*$ denotes a nonzero matrix.

Case 2: σ is an integer, and $0 < \sigma < n-3/2$. In this case Eq. (37) cannot be fulfilled. Therefore, now we have the situation shown in Fig. 2. The set of points $((S+S')/2, J)$ is divided into two domains D^0 and D^d. If $\sigma = n-3$, then the domain D^0 reduces to a line. The representation π^σ realizes on the vectors $|J,S,S';\Sigma\rangle$, $((S+S')/2, J) \in D^0$, irreducible representation. We denote it by D^0_σ. The quotient space π^σ/D^0_σ will be denoted by D^d_σ. It is irreducible. The representation π^σ in this case has the structure

$$\begin{pmatrix} D^0_\sigma & * \\ 0 & D^d_\sigma \end{pmatrix}. \tag{42}$$

The irreducible representations π^σ (see Theorem 1) and the representations $D^F_\sigma, D^0_\sigma, D^d_\sigma$ of the group SO*$(2n)$ (for the same or different σ) are pairwise infinitesimally nonequivalent. The proof follows from a comparison of U(n) spectra of these representations.

Now consider the representations of the group SU*$(2n)$.

Theorem 2: The representation π^σ of SU*$(2n)$ is irreducible if and only if σ is not an integer, or σ is equal to one of the integers $1, 2, ..., 2n-1$.

The proof is completely analogous to that of Ref. 10 for the groups U$(n, 1)$ and SO$_0(n, 1)$ (see also Sec. 2 in Ref. 19).

For real σ the representations π^σ and $\pi^{-\sigma+2n}$ are simultaneously irreducible or reducible.[9,10] In the first case they are equivalent; in the second case they consist of the same irreducible representations of SU*$(2n)$.

Now we consider the representations π^σ of SU*$(2n)$ for which σ is a nonpositive integer. They are studied by means of the infinitesimal operators (35). Now the infinitesimal operators are simpler than in the case of SO*$(2n)$. On the vectors $|m,\Sigma\rangle$, $m = 0, 1, 2, ..., -\sigma$, the representation π^σ realizes the irreducible representation of SU*$(2n)$. We denote it by D^F_σ. It is clear that this representation is finite-dimensional. The quotient representation π^σ/D^F_σ is also irreducible. We denote it by D_σ. Now the representation π^σ of SU*$(2n)$, where σ is a nonpositive integer, has the structure

$$\begin{pmatrix} D^F_\sigma & * \\ 0 & D_\sigma \end{pmatrix}.$$

The irreducible representations π^σ (see Theorem 2) and the representations D^F_σ, D_σ are pairwise infinitesimally nonequivalent.

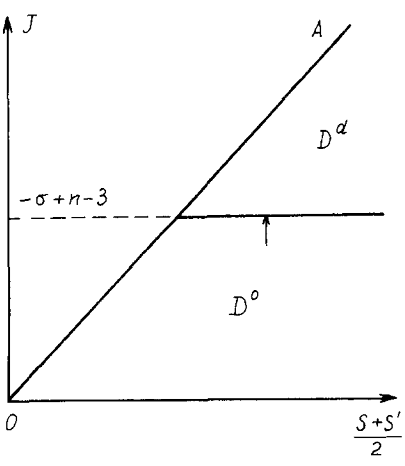


FIG. 2. Structure of the representation π^σ of SO*$(2n)$ if σ is an integer, $0<\sigma<n-\frac{3}{2}$.

VII. MOST DEGENERATE UNITARY SERIES (MDUS) OF REPRESENTATIONS OF SO*(2n) AND SU*(2n)

The group SO(2n)*: Now we extract the unitary representations and the representations which can be made unitary (unitarizable representations) from among the set $\mathfrak{R}$ of the irreducible representations π^σ and the representations D^0_σ, D^d_σ. These representations can be found by the unitarization procedure (see Refs. 8 and 9 and also Sec. 4 of Chap. 5 in Ref. 10). According to this procedure, we have to find the intertwining operators $\Pi \equiv \Pi(\sigma)$ for every pair π^σ, $\pi^{-\sigma+2n-3}$. Then we introduce the new basis

$$|J,S,S';\Sigma\rangle' = \Pi^{1/2}|J,S,S';\Sigma\rangle \tag{43}$$

into Eqs. (32) and (33). The representations which admit unitarization are unitary in this basis. It is easy to find that the unitarity condition is

$$\pi(Y)^* = -\pi(Y), \quad Y \in \mathbf{p}. \tag{44}$$

The intertwining operator $\Pi(\sigma)$ for the representations π^σ and $\pi^{-\sigma+2n-3}$ is defined as

$$\Pi(\sigma)\pi^\sigma = \pi^{-\sigma+2n-3}\Pi(\sigma), \tag{45}$$

where it is understood that both sides of this relation have to act upon $|J,S,S';\Sigma\rangle$. For matrix elements of $\Pi(\sigma)$ we have (see Refs. 8–10)

$$\langle J,S,S';\Sigma|\Pi(\sigma)|J_1,S_1,S'_1;\Sigma_1\rangle = a(J,S,S')\delta_{JJ_1}\delta_{SS_1}\delta_{S'S'_1}\delta_{\Sigma\Sigma_1},$$

where $a(J,S,S')$ is independent of Σ.

Substituting Eqs. (32) and (33) into Eq. (45) and comparing the coefficients at the same basis vectors, we obtain the system of equations for the matrix elements $a(J,S,S')$. Solving these equations, we derive that

$$a(J,S+2j,S') = a(J,S,S'+2j) = \prod_{r=0}^{j-1} \frac{-\sigma+2n-3+(S+S')/2+r}{\sigma+(S+S')/2+r}\, a(J,S,S'), \tag{46}$$

$$a(J,S-2j,S') = a(J,S,S'-2j) = \prod_{r=0}^{j-1} \frac{-\sigma-(S+S')/2+1+r}{\sigma-(S+S')/2-2n+4+r}\, a(J,S,S'), \tag{47}$$

$$a(J+j,S-j,S'+j) = a(J+j,S+j,S'-j) = \prod_{r=0}^{j-1} \frac{-\sigma+J+n+r}{\sigma+J-n+3+r}\, a(J,S,S'), \tag{48}$$

$$a(J-j,S-j,S'+j) = a(J-j,S+j,S'-j) = \prod_{r=0}^{j-1} \frac{-\sigma-J+n-2+r}{\sigma-J-n+1+r}\, a(J,S,S'). \tag{49}$$

Assigning a value to $a(J,S,S')$ for fixed $J=J_0$, $S=S_0$, $S'=S'_0$, from Eqs. (46)–(49), we can find uniquely the value of $a(J,S,S')$ for every (J,S,S'). Note that the intertwining operator is defined uniquely up to a numerical constant. If we fix $a(J_0,S_0,S'_0)$, then this constant is also fixed.

Theorem 3: The following representations of SO*(2n) from the set $\mathfrak{R}$ admit unitarization:

(a) The representations π^σ, for which $\sigma - n + 3/2$ are pure imaginary (principal MDUS);

(b) the representations π^σ, $0 < |\sigma - n + 3/2| < 3/2$ (supplementary MDUS);

(c) all the representations D^d_σ, σ an integer satisfying $\sigma < n - 3/2$ (discrete MDUS);

(d) the representation $D^0_{\sigma = n-3}$ (ladder representation).

Proof: If a representation admits unitarization, then the condition (44) is fulfilled in some basis. We have $I_s^\pm \in \mathbf{p}_c$ and $I_s^\pm \notin \mathbf{p}$. It is easy to verify that the condition (44) implies

$$\pi^\sigma(I_s^+)^* = -\pi^\sigma(I_s^-). \tag{50}$$

This condition is directly verified for the representations (a) with the use of Eqs. (32) and (33). The representations (b), (c), and (d) are unitary in the new basis (43). From Eqs. (32), (43), and (46)–(49) we find that

$$\begin{aligned}
&\pi^\sigma(I_s^+)|J,S,S';\Sigma\rangle' \\
&= \sum_{\Sigma'}\left[\left(\sigma + \frac{S+S'}{2}\right)\left(-\sigma + \frac{S+S'}{2} + 2n - 3\right)\right]^{1/2} C^{JSS'}_{JS+2S'}(\Sigma', s)|J, S+2, S';\Sigma'\rangle' \\
&+ \sum_{\Sigma'}\left[\left(\sigma + \frac{S+S'}{2} - 1\right)\left(-\sigma + \frac{S+S'}{2} + 2n - 4\right)\right]^{1/2} C^{JSS'}_{JSS'-2}(\Sigma', s)|J, S, S'-2;\Sigma'\rangle' \\
&+ \sum_{\Sigma'}[(\sigma + J - n + 3)(-\sigma + J + n)]^{1/2} C^{JSS'}_{J+1\,S+1\,S'-1}(\Sigma', s)|J+1, S+1, S'-1;\Sigma'\rangle' \\
&+ \sum_{\Sigma'}[(\sigma + J - n + 2)(-\sigma + J + n - 1)]^{1/2} C^{JSS'}_{J-1\,S+1\,S'-1}(\Sigma', s)|J-1, S+1, S'-1;\Sigma'\rangle',
\end{aligned} \tag{51}$$

$$\begin{aligned}
&\pi^\sigma(I_s^-)|J,S,S';\Sigma\rangle' \\
&= -\sum_{\Sigma'}\left[\left(-\sigma + \frac{S+S'}{2} + 2n - 4\right)\left(\sigma + \frac{S+S'}{2} - 1\right)\right]^{1/2} C^{JSS'}_{JS-2S'}(\Sigma', s)|J, S-2, S';\Sigma'\rangle' \\
&- \sum_{\Sigma'}\left[\left(-\sigma + \frac{S+S'}{2} + 2n - 3\right)\left(\sigma + \frac{S+S'}{2}\right)\right]^{1/2} C^{JSS'}_{JSS'+2}(\Sigma', s)|J, S, S'+2;\Sigma'\rangle' \\
&- \sum_{\Sigma'}[(-\sigma + J + n - 1)(\sigma + J - n + 2)]^{1/2} C^{JSS'}_{J-1\,S-1\,S'+1}(\Sigma', s)|J-1, S-1, S'+1;\Sigma'\rangle' \\
&- \sum_{\Sigma'}[(-\sigma + J + n)(\sigma + J - n + 3)]^{1/2} C^{JSS'}_{J+1\,S-1\,S'+1}(\Sigma', s)|J+1, S-1, S'+1;\Sigma'\rangle'.
\end{aligned} \tag{52}$$

Now we can easily verify that the representations of the theorem are the only representations which satisfy the condition (50). The theorem is proved.

The group SU(2n)*: This group (as every linear semisimple Lie group) has the principal MDUS (see, for example, Refs. 8–10). The representations π^σ, $\sigma = n + i\omega$, ω real, form the principal MDUS. These representations are unitary, and one does not need the unitarization procedure.

Let us construct the intertwining operator $\Pi \equiv \Pi(\sigma)$ for the pair π^σ and $\pi^{-\sigma+2n}$. We have that

$$\langle m, \Sigma|\Pi(\sigma)|m', \Sigma'\rangle = a(m)\delta_{mm'}\delta_{\Sigma\Sigma'}. \tag{53}$$

By the same procedure as in the case of the group SO*(2n) we find

$$a(m+j) = \prod_{r=0}^{j-1} \frac{-\sigma + 2n + m + r}{\sigma + m + r}\, a(m). \tag{54}$$

VIII. INFINITESIMAL OPERATORS OF UNITARY IRREDUCIBLE REPRESENTATIONS OF SO(2n) IN A U(n) BASIS AND OF SU(2n) IN AN Sp(n) BASIS

Consider the finite-dimensional representations D^F_σ, $\sigma = 0, -1, -2, \cdots$, of the groups SO*(2n) and SU*(2n) (see Sec. 6). It is clear that these representations lead to the irreducible representations of SO(2n) and SU(2n) (see Sec. 1 and Table 1 in Ref. 12). According to Theorem 5.13a in Ref. 10, the representation D^F_σ of SO(2n) and of SU(2n) has highest weight $(-\sigma, -\sigma, \dot{0})$. The weights are written down in the coordinate system $(m_1, m_2, \cdots)$, for which $m_1 \geqslant m_2 \geqslant \cdots$.

From the results of Sec. 6 we can determine the U(n) spectrum of the representation D^F_σ of SO(2n) and the Sp(n) spectrum of the representation D^F_σ of SU(2n). Namely, *the restriction $D^F_\sigma|_{U(n)}$ is characterized by all triples (J, S, S') of nonnegative integers of the same evenness, for which $J \leqslant S$, $J \leqslant S'$, and $(S+S')/2 \leqslant -\sigma$.* Every triple (J, S, S') corresponds to the irreducible representation of U(n) with highest weight $(m_1, m_2, \dot{0}, m_3, m_4)$, where integers m_i are defined by Eq. (31). *The representation D^F_σ of* SU(2n) *decomposes into a direct sum of irreducible representations of* Sp(n) *with highest weights $(m, m, \dot{0})$, $m = 0,1,2,\ldots,-\sigma$.*

Equations (32), (33), and (35) define infinitesimal operators of the representations D^F_σ of SO(2n) and SU(2n). But they do not satisfy the unitarity condition [which is similar to Eq. (44)]. In order to satisfy this condition, we have to introduce the new basis (43). In the new basis the infinitesimal operators of the irreducible unitary representation of SO(2n) with highest weight $(M, M, \dot{0})$ act as

$I_k^+|J,S,S';\Sigma\rangle$

$$=\sum_{\Sigma'} i\left[\left(M-\frac{S+S'}{2}\right)\left(M+\frac{S+S'}{2}+2n-3\right)\right]^{1/2} C^{JSS'}_{JS+2S'}(\Sigma',k)|J,S+2,S';\Sigma'\rangle$$

$$+\sum_{\Sigma'} i\left[\left(M-\frac{S+S'}{2}+1\right)\left(M+\frac{S+S'}{2}+2n-4\right)\right]^{1/2} C^{JSS'}_{JSS'-2}(\Sigma',k)|J,S,S'-2;\Sigma'\rangle$$

$$+\sum_{\Sigma'} i[(M-J+n-3)(M+J+n)]^{1/2} C^{JSS'}_{J+1S+1S'-1}(\Sigma',k)|J+1,S+1,S'-1;\Sigma'\rangle$$

$$+\sum_{\Sigma'} i[(M-J+n-2)(M+J+n-1)]^{1/2} C^{JSS'}_{J-1S+1S'-1}(\Sigma',k)|J-1,S+1,S'-1;\Sigma'\rangle, \tag{55}$$

$I_k^-|J,S,S';\Sigma\rangle$

$$=-\sum_{\Sigma'} i\left[\left(M-\frac{S+S'}{2}+1\right)\left(M+\frac{S+S'}{2}+2n-4\right)\right]^{1/2} C^{JSS'}_{JS-2S'}(\Sigma',k)|J,S-2,S';\Sigma'\rangle$$

$$-\sum_{\Sigma'} i\left[\left(M-\frac{S+S'}{2}\right)\left(M+\frac{S+S'}{2}+2n-3\right)\right]^{1/2} C^{JSS'}_{JSS'+2}(\Sigma',k)|J,S,S'+2;\Sigma'\rangle$$

$$-\sum_{\Sigma'} i[(M-J+n-2)(M+J+n-1)]^{1/2} C^{JSS'}_{J-1S-1S'+1}(\Sigma',k)|J-1,S-1,S'+1;\Sigma'\rangle$$

$$-\sum_{\Sigma'} i[(M-J+n-3)(M+J+n)]^{1/2} C^{JSS'}_{J+1S-1S'+1}(\Sigma',k)|J+1,S-1,S'+1;\Sigma'\rangle. \tag{56}$$

The infinitesimal operators of the irreducible unitary representation of SU($2n$) with highest weight $(M, M, \dot{0})$ act as

$$E_j|m,\Sigma\rangle=\sum_{\Sigma'} i[(M+m+2n)(M-m)]^{1/2} C^{m}_{m+1}(\Sigma',j)|m+1,\Sigma'\rangle$$

$$-\sum_{\Sigma'} i[(M+m+2n-1)(M-m+1)]^{1/2} C^{m}_{m-1}(\Sigma',j)|m-1,\Sigma'\rangle$$

$$-\sum_{\Sigma'} (M+n) C^{m}_{m}(\Sigma',j)|m,\Sigma'\rangle. \tag{57}$$

Note that CGC's for the group U(n) which are contained in Eq. (27) [and, therefore, in Eqs. (32), (33), (51), (55), and (56)] are known for the Gel'fand–Zetlin basis from Ref. 20. Unfortunately, CGC's for the group Sp(n) which are contained in Eqs. (35) and (57) are not known. They can be evaluated using the results of Ref. 21. They will be given in a separate paper.

[1]M. Gell-Mann, P. Ramond, and R. Slansky, Rev. Mod. Phys. **50**, 721 (1978).
[2]P. Langacker, Phys. Rep. **72**, 186 (1981).
[3]L. C. Biedenharn and P. Truini, Preprint SLAC-PUB-2802, 1982.
[4]S. Ferrara, J. Scherk, and B. Zumino, Phys. Lett. B **66**, 35 (1977).
[5]E. Cremmer, J. Scherk, and S. Ferrara, Phys. Lett. B **68**, 234 (1977).
[6]J. Ellis, M. K. Gaillard, and B. Zumino, Phys. Lett. B **94**, 343 (1980).
[7]E. Cremmer and B. Julia, Nucl. Phys. B **159**, 141 (1979); B. Julia, Preprint LPTENS 81/14, 1981; M. Günaydin and C. Saclioglu, Phys. Lett. B **108**, 180 (1982).
[8]A. W. Knapp and E. M. Stein, Ann. Math. **93**, 489 (1971).
[9]A. W. Knapp and E. M. Stein, Invent. Math. **60**, 9 (1980).
[10]A. U. Klimyk, *Matrix Elements and Clebsch–Gordan Coefficients of Group Representations* (Naukova Dumka, Kiev, 1979) (in Russian).
[11]A. U. Klimyk and A. M. Gavrilik, J. Math. Phys. **20**, 1624 (1979).
[12]A. U. Klimyk and B. Gruber, J. Math. Phys. **20**, 1995 (1979).
[13]B. Gruber and A. U. Klimyk, J. Math. Phys. **22**, 2762 (1981).
[14]S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[15]G. Warner, *Harmonic Analysis on Semisimple Lie Groups* (Springer, Berlin, 1972), Vol. I.
[16]A. Guillemonat, "Representations spheriques singulieres," preprint 1982.
[17]A. O. Barut and R. Rączka, *Theory of Group Representations and Applications* (PWN-Polish Scientific Publishers, Warsaw, 1977).
[18]A. U. Klimyk, J. Math. Phys. **24**, 224 (1983).
[19]A. U. Klimyk and B. Gruber, J. Math. Phys. **23**, 1399 (1982).
[20]L. C. Biedenharn and J. D. Louck, Commun. Math. Phys. **8**, 89 (1968).
[21]P. I. Holod and A. U. Klimyk, Rep. Math. Phys. (to be published).

Dyson representation of SU(3) in terms of five boson operators[a)]

Resit Dündarer and Feza Gürsey
Physics Department, Yale University, New Haven, Connecticut 06511

(Received 30 September 1983; accepted for publication 9 November 1983)

A representation of SU(3) in terms of five boson operators is proposed. It is a generalization of the Dyson–Maleev type representation used in nuclear physics with two boson operators related to the integers that label the irreducible representation of SU(3).

PACS numbers: 02.20.Qs, 11.30.Jw

I. INTRODUCTION

Several representations of SU(3) in terms of boson operators have been proposed in the past, but none of them use as many boson operators as the maximal number of commuting observables of the algebra. For any Lie group, the maximal number of commuting observables is $\delta = (r + d)/2$ where r is the rank of the group and d is the dimension of the adjoint representation (number of the parameters of the group). For SU(3), $\delta = (2 + 8)/2 = 5$. Since there is no five-dimensional irreducible representation of SU(3), the representation by means of five boson operators will not be linear. The Dyson–Maleev–Gelfand types of representations[1] in terms of differential operators are of this type. For their importance in nuclear physics we refer the reader to a review article by A. Klein.[2] In this note we propose a representation of SU(3) in terms of five commuting boson operators a_1, a_2, a_3, b_1, b_2 that satisfy $[a_i, a_j^+] = [b_i, b_j^+] = \delta_{ij}$, $[a_i, a_j] = [b_i, b_j] = [a_i, b_j] = [a_i, b_j^+] = 0$. The b_i will be chosen such that the eigenvalues of $b_i^+ b_i$ give the labels p and q of the irreducible representations of SU(3).

II. CONSTRUCTION OF THE REPRESENTATION

The most commonly chosen basis for the SU(3) algebra is the set of operators F_i that satisfy $[F_i, F_j] = if_{ijk}F_k$. In addition, in the fundamental representation of the algebra we also have the anticommutator algebra $\{F_i, F_j\} = \frac{1}{3}\delta_{ij} + d_{ijk}F_k$. In terms of 3×3 matrices, the F_i are represented by $\lambda_i/2$ where λ_i are the well-known Gell-Mann matrices. We will realize the algebra using a different basis related to the above operators as $I_\pm = F_1 \pm iF_2$, $U_\pm = F_6 \pm iF_7$, $V_\pm = F_4 \pm iF_5$, $I_3 = F_3$, $Y = 2/\sqrt{3}F_8$. We will proceed analogously to the 2-boson representation of SU(2) given in Ref. 3 which generalizes the Dyson–Maleev–Gelfand representation to all irreducible representations. Besides regarding the irreducible representation labels as eigenvalues of $b_i^+ b_i$, we also represent the commuting step-up operators of the algebra by different creation operators a_i^+ and derive the rest of the representation from the commutation relations of the SU(3) algebra. Our procedure parallels the work of I. Bars[4] on the representations of the noncompact group SL(3,C) based on the Z-operator formalism proposed earlier.[5] Here we summarize the procedure:

(1) Set $V_+ = a_1^+$,

(2) $[U_+, V_+] = 0$ suggests $U_+ = a_2^+$,

(3) $[I_+, V_+] = 0$ and $[I_+, U_+] = V_+$ suggests $I_+ = a_1^+ a_2 - a_3^+$,

(4) $[I_3, I_+] = I_+$, $[I_3, U_+] = -\frac{1}{2}U_+$ and $[I_3, V_+] = \frac{1}{2}V_+$ suggests $I_3 = \frac{1}{2}(a_1^+ a_1 - a_2^+ a_2) + a_3^+ a_3 + c_1 b_1^+ b_1 + c_2 b_2^+ b_2 + c_3$, where $c_1, c_2, c_3 \in R$,

(5) $[I_+, I_-] = 2I_3$, $[I_3, I_-] = -I_-$, $[U_+, I_-] = 0$ and $[I_-, V_+] = U_+$ suggests $I_- = a_2^+ a_1 + (a_3^+ a_3 + 2c_1 b_1^+ b_1 + 2c_2 b_2^+ b_2 + 2c_3)a_3$,

(6) $[V_+, U_-] = I_+$, $[I_3, U_-] = \frac{1}{2}U_-$, $[I_+, U_-] = 0$ suggests $U_- = a_3^+ a_1 + (a_3^+ a_3 - a_2^+ a_2 - a_1^+ a_1 + d_1 b_1^+ b_1 + d_2 b_2^+ b_2 + d_3)a_2$, where $d_1, d_2, d_3 \in R$,

(7) $[U_-, I_-] = V_-$ implies

$$V_- = -[a_1^+ a_1 + a_2^+ a_2 + a_3^+ a_3 + (2c_1 - d_1)b_1^+ b_1 + (2c_2 - d_2)b_2^+ b_2 + 2c_3 - d_3]a_1 - (a_3^+ a_3 + 2c_1 b_1^+ b_1 + 2c_2 b_2^+ b_2 + 2c_3)a_2 a_3.$$

This form satisfies the following relations:

$$[I_3, V_-] = -\tfrac{1}{2}V_-, [I_-, V_-] = [U_-, V_-] = 0,$$
$$[I_+, V_-] = -U_-, [U_+, V_-] = I_-,$$

(8) $[U_+, U_-] = \frac{3}{2}Y - I_3$ implies $Y = a_1^+ a_1 + a_2^+ a_2 + \frac{2}{3}(c_1 - d_1)b_1^+ b_1 + \frac{2}{3}(c_2 - d_2)b_2^+ b_2 + \frac{2}{3}(c_3 - d_3)$ and the following are satisfied:

$$[Y, I_3] = [Y, I_\pm] = 0, [Y, U_\pm] = \pm U_\pm,$$
$$[Y, V_\pm] = \pm V_\pm \text{ and } [V_+, V_-] = \tfrac{3}{2}Y + I_3.$$

To find the arbitrary coefficients c_i, d_i we demand that the labels of the representation $p, q (= 0, 1, 2, \ldots)$ be the eigenvalues of $b_1^+ b_1$, $b_2^+ b_2$, respectively. We find the quadratic Casimir operator $C_2 = \Sigma F_i^2$ and set it equal to $\frac{2}{3}(p^2 + q^2 + pq + 3(p + q))$ to get $c_1 = c_2 = \frac{1}{2}$, $c_3 = \frac{3}{2}$, $d_2 = 1$, $d_1 = d_3 = 0$. As a result we have (denoting the operators $b_1^+ b_1$ by P and $b_2^+ b_2$ by Q):

$$\begin{aligned}
V_+ &= a_1^+,\\
V_- &= -(a_1^+ a_1 + a_2^+ a_2 + a_3^+ a_3 + P + 3)a_1\\
&\quad - (a_3^+ a_3 + P + Q + 3)a_2 a_3,\\
U_+ &= a_2^+,\\
U_- &= a_3^+ a_1 + (a_3^+ a_3 - a_2^+ a_2 - a_1^+ a_1 + Q)a_2,\\
I_+ &= a_1^+ a_2 - a_3^+,\\
I_- &= a_2^+ a_1 + (a_3^+ a_3 + P + Q + 3)a_3,\\
I_3 &= \tfrac{1}{2}(a_1^+ a_1 - a_2^+ a_2) + a_3^+ a_3 + \tfrac{1}{2}(P + Q + 3),\\
Y &= a_1^+ a_1 + a_2^+ a_2 + \tfrac{1}{3}(P - Q + 3),
\end{aligned} \tag{1}$$

and the Casimir operators are

[a)] Research supported in part by the U.S. Department of Energy under Contract No. DE-AC02-76ERO3075.

 0022-2488/84/030431-02$02.50

$$C_2 = 2\sum F_i^2 = \tfrac{2}{3}(P^2 + Q^2 + PQ + 3(P+Q)),$$

$$\begin{aligned} C_3 &= d_{ijk}F_iF_jF_k \\ &= 2(P-Q)\left[\tfrac{2}{9}(P+Q)^2 + \tfrac{1}{9}PQ + P + Q + 1\right] \end{aligned} \tag{2}$$

and the dimension operator of the representation is

$$\begin{aligned} D &= \tfrac{1}{2}(P+1)(Q+1)(P+Q+2) \\ &= \tfrac{1}{2}(b_1 b_1^+)(b_2 b_2^+)(b_1 b_1^+ + b_2 b_2^+). \end{aligned} \tag{3}$$

A definite irreducible representation of SU(3) is characterized by the eigenvalues p and q of the number operators P and Q. On replacing P, Q by their integer eigenvalues in (1) one gets the Dyson–Maleev representation for SU(3) which is related to Bars' representation for SL(3, C). The eigenvalues of C_2, C_3, and D are then given by Eqs. (2) and (3) on making the same replacement for P and Q.

III. CONCLUDING REMARKS

In the representation given above, the five operators C_2, C_3, Y, I_3, and $\mathbf{I}\cdot\mathbf{I} = I_+I_- + I_3(I_3 - 1) = a_1^+ a_1(a_2^+ a_2 + 1) + (a_3^+ a_3 + P + Q + 3)a_1^+ a_2 a_3 - a_1 a_2^+ a_3^+ - (a_3^+ a_3 + P + Q + 2)a_3^+ a_3 + I_3(I_3 - 1)$ commute among themselves and thus can be used in labeling the states. However, unlike C_2, C_3, I_3 and Y, $\mathbf{I}^2$ does not consist only of number operators and makes it impossible to use the states $|n_1, n_2, n_3, p, q\rangle$ as our eigenstates. The simultaneous eigenstates of the five operators will instead have the form $G\,|n_1, n_2, n_3, p, q\rangle$, where G is the metric operator in this representation. The construction of the metric operator will be the object of a separate note.

ACKNOWLEDGMENTS

We would like to thank F. Iachello, Y. Alhassid, B. Balantekin, and C. Tze for stimulating discussions and references to the literature.

[1]F. Dyson, Phys. Rev. **102**, 1217 (1956).

[2]A. Klein, *Proceedings of the 11th International Colloquium on Group Theory Methods in Physics*, Istanbul, edited by M. Serdaroğlu and E. İnonü, (Springer-Verlag, Berlin, 1983).

[3]Y. Alhassid, F. Gürsey, and F. Iachello, "Group Theory Approach to Scattering," Yale preprint (1982).

[4]I. Bars, J. Math. Phys. **14**, 759 (1973).

[5]I. Bars and F. Gürsey, J. Math. Phys. **13**, 131 (1972).

Charge operators in simple Lie groups

A. Taormina[a]
Physique Théorique et Mathématique, Faculté des Sciences, Université de l'Etat à Mons, 7000 Mons, Belgium

(Received 16 September 1982; accepted for publication 11 February 1983)

Charge operators for representations of dimension less than or equal to 16 are computed in all simple Lie groups. The representations for which the charge operator reproduces the charge spectrum of leptons and quarks of one family are analyzed from a GUT point of view.

PACS numbers: 02.20.Qs, 12.10. – g

1. INTRODUCTION

Schemes for grand unified theories generally require one or more irreducible complex representations of a simple group, which are anomaly free and possess asymptotic freedom. The group has to contain the SU(2)×U(1) weak electromagnetic subgroup and the SU(3) color group thought to be responsible for the QCD strong interactions.

Although those conditions are by now considered as minimal,[1] it has seemed to us that it would be helpful to know exactly which groups one can select, asking only for a much weaker condition, namely the existence of a good electric charge operator. What we precisely mean by this is explained in Sec. 2.[2]

Since we now believe that there are 15 or 16 quarks and leptons in the first generation, and seven allowed charges (from -1 to $+1$, by steps of 1/3), we have restricted ourselves to compute the "good" charge operators for all irreducible representations of dimension less than or equal to 16 in simple Lie groups.[3] The values of the parameters are discussed quite generally according to the conditions of Sec. 2.

In Sec. 3, we compute the charge operators for all relevant groups. They appear in Table I.

It should be noted that this approach is valid for any additive quantum number, and not only for the electric charge. For this reason, our tables can be used more generally.[4]

In Sec. 4, we show that a great class of charge operators are obtained in a quite trivial way, and present the Table II where the less trivial cases are collected.

In Sec. 5, we discuss the opportunity of considering extra-abelian U(1) factors, which we call shifts.

Finally, we give in Sec. 6 the pathology of each group which has some good properties (but not all) to be a GUT candidate, and we give our conclusions in Sec. 7.

2. CHARGE RESTRICTION

Instead of using the embedding of the SU(3) color group in a larger group as the major condition to select a candidate for GUT, we restrict ourselves to the requirement of a good charge operator. We find all groups having such a good operator in a given irreducible representation. The latter obeys dimension limitations which we now describe.

The fermionic content of a GUT is described by quarks and leptons of several families. From the point of view of electric charge, these families are equivalent. We take into account one family in the representation(s) and conventionally use the labeling of the first family. The others are obtained by mere duplication of the constructed multiplet. The allowed charges in the spectrum are 0, $\pm 1/3$, $\pm 2/3$, ± 1, which means that we only consider charge operators with, at most, seven different eigenvalues.

The size of the representation will depend on the multiplicity of each eigenvalue. We ask the multiplicity of the $\pm 1/3$ and $\pm 2/3$ charges to be at most three, and the multiplicity of the ± 1 charges to be at most one. Moreover, we analyze the theories with one and two neutrinos.[5]

If the maximal allowed structure

$$(1e^-, 3\bar{u}, 3d, (1 \text{ or } 2)\nu, 3\bar{d}, 3u, 1e^+), \tag{1}$$

which limits the dimension of the representation to 16, is obtained, one can expect an underlying SU(3) color group. However, the restrictions above show that we are also interested in the cases where no SU(3) color appears. This may occur when the complete spectrum described in (1) is obtained by adding contributions of several representations of the same group. Then, the multiplicities may come out numerically (for instance, $2u$-quarks in an irrep, and $1u$-quark in another irrep of the same group make $3u$-quarks when considering the sum of the irreps). We think it can be interesting to investigate these groups even in the absence of any SU(3) color.

All these structures are presented in Table II, and we shall discuss them in Sec. 6.

Let us make a final comment on the possibility of U(1) shifts. Since the charge operator is a generator of a Lie group, it must be traceless. However, we also consider the opportunity of an extra U(1) factor, so that the charges could possibly be decomposed in two parts: the first one, traceless, belonging to the simple group and the second one to the extra U(1).

Table II gives the allowed structures when the charge operator is entirely within the simple group, while the shifts are discussed in Sec. 5.

3. CHARGE OPERATOR BUILDING

The first task is to construct the charge operators associated with the irreducible representations of dimension less than or equal to 16 for all allowed simple groups. The only groups to investigate are thus A_n ($2 \leqslant n \leqslant 15$), B_n ($2 \leqslant n \leqslant 7$), C_n ($3 \leqslant n \leqslant 8$), D_n ($4 \leqslant n \leqslant 8$), and the exceptional group G_2. Indeed, A_1 is clearly too small, $B_2 = C_2$, $D_2 = A_1 \times A_1$, and $D_3 = A_3$ locally.

Tables III recall the allowed representations of the al-

[a] Aspirant au Fonds National de la Recherche Scientifique, Faculté des Sciences, Université de l'Etat à Mons, Belgium.

 0022-2488/84/030433-10$02.50

TABLE I. Structure of charge operators in A_n, B_n, C_n, D_n, G_2, for d-dimensional irreducible representations ($d \leqslant 16$).[a]

$A_n \equiv SU(n+1)$, $2 \leqslant n \leqslant 15$		
$n=2$	3	$(0, a, b) - 1/3(a+b)$
	$\bar{3}$	$(0, -a+b, b) + 1/3(a-2b)$
	6	$(0, a, 2a, b, a+b, 2b) - 2/3(a+b)$
	$\bar{6}$	$(0, -a+b, b, -2a+2b, -a+2b, 2b) + 2/3(a-2b)$
	$8 \equiv \bar{8}$	$(0, a, -a+b, b^2, a+b, -a+2b, 2b) - b$
	10	$(0, a, 2a, 3a, b, a+b, 2a+b, 2b, a+2b, 3b) - (a+b)$
	$\overline{10}$	$(0, -a+b, b, -2a+2b, -a+2b, 2b, -3a+3b, -2a+3b, -a+3b, 3b) + (a-2b)$
	15	$(0, a, 2a, -a+b, b^2, (a+b)^2, 2a+b, -a+2b, (2b)^2, a+2b, -a+3b, 3b) - 1/3(a+4b)$
	$\overline{15}$	$(0, a, -a+b, b^2, a+b, -2a+2b, (-a+2b)^2, (2b)^2, a+2b, -2a+3b, -a+3b, 3b) + 1/3(a-5b)$
	$15'$	$(0, a, 2a, 3a, 4a, b, a+b, 2a+b, 3a+b, 2b, a+2b, 2a+2b, 3b, a+3b, 4b) - 4/3(a+b)$
	$\overline{15}'$	$(0, -a+b, b, -2a+2b, -a+2b, 2b, -3a+3b, -2a+3b, -a+3b, 3b, -4a+4b, -3a+4b, -2a+4b, -a+4b, 4b) + 4/3(a-2b)$
$n=3$	4	$(0, a, b, c) - 1/4(a+b+c)$
	$\bar{4}$	$(0, -b+c, -a+c, c) + 1/4(a+b-3c)$
	$6 \equiv \bar{6}$	$(0, -a+b, b, -a+c, c, -a+b+c) + 1/2(a-b-c)$
	10	$(0, a, 2a, b, a+b, 2b, c, a+c, b+c, 2c) - 1/2(a+b+c)$
	$\overline{10}$	$(0, -b+c, -a+c, c, -2b+2c, -a-b+2c, -b+2c, -2a+2c, -a+2c, 2c) + 1/2(a+b-3c)$
	$15 \equiv \overline{15}$	$(0, a, b, -b+c, a-b+c, -a+c, c^3, a+c, -a+b+c, b+c, -b+2c, -a+2c, 2c) - c$
$n=4$	5	$(0, a, b, c, d) - 1/5(a+b+c+d)$
	$\bar{5}$	$(0, -c+d, -b+d, -a+d, d) + 1/5(a+b+c-4d)$
	10	$(0, -a+b, b, -a+c, c, -a+b+c, -a+d, d, -a+b+d, -a+c+d) + 1/5(3a-2b-2c-2d)$
	$\overline{10}$	$(0, -b+c, -a+c, c, -b+d, -a+d, d, -a-b+c+d, -b+c+d, -a+c+d) + 1/5(2a+2b-3c-3d)$
	15	$(0, a, 2a, b, a+b, 2b, c, a+c, b+c, 2c, d, a+d, b+d, c+d, 2d) - 2/5(a+b+c+d)$
	$\overline{15}$	$(0, -c+d, -b+d, -a+d, d, -2c+2d, -b-c+2d, -a-c+2d, -c+2d, -2b+2d, -a-b+2d, -b+2d, -2a+2d, -a+2d, 2d) + 2/5(a+b+c-4d)$
$n=5$	6	$(0, a, b, c, d, e) - 1/6(a+b+c+d+e)$
	$\bar{6}$	$(0, -d+e, -c+e, -b+e, -a+e, e) + 1/6(a+b+c+d-5e)$
	15	$(0, -a+b, b, -a+c, c, -a+b+c, -a+d, d, -a+b+d, -a+c+d, -a+e, e, -a+b+e, -a+c+e, -a+d+e) + 1/3(2a-b-c-d-e)$
	$\overline{15}$	$(0, -c+d, -b+d, -a+d, d, -c+e, -b+e, -a+e, e, -b-c+d+e, -a-c+d+e, -c+d+e, -a-b+d+e, -b+d+e, -a+d+e) + 1/3(a+b+c-2d-2e)$
$6 \leqslant n \leqslant 15$	$n+1$	$(0, a_1, \ldots, a_n) - 1/(n+1) \sum_{i=1}^{n} a_i$
	$\overline{n+1}$	$(0, -a_{n-1}+a_n, \ldots, -a_1+a_n, a_n) + 1/(n+1)\left(\sum_{i=1}^{n-1} a_i - n a_n\right)$, where $0 < a_i \ll a_{i+1}$, for $1 \leqslant i \leqslant n$
$B_n \equiv SO(2n+1)$, $2 \leqslant n \leqslant 7$		
$n=2$	4_s	$(0, a, b, a+b) - 1/2(a+b)$
	5_v	$(0, -a+b, b, a+b, 2b) - b$
	10_v	$(0, a, 2a, b, (a+b)^2, 2a+b, 2b, a+2b, 2a+2b) - (a+b)$
	14_v	$(0, -a+b, b, a+b, -2a+2b, -a+2b, (2b)^2, a+2b, 2a+2b, -a+3b, 3b, a+3b, 4b) - 2b$
	16_s	$(0, a, -a+b, b^2, (a+b)^2, 2a+b, -a+2b, (2b)^2, (a+2b)^2, 2a+2b, 3b, a+3b) - 1/2(a+3b)$
$n=3$	7_v	$(0, -b+c, -a+c, c, a+c, b+c, 2c) - c$
	8_s	$(0, a, b, a+b, c, a+c, b+c, a+b+c) - 1/2(a+b+c)$
$n=4$	9_v	$(0, -c+d, -b+d, -a+d, d, a+d, b+d, c+d, 2d) - d$
	16_s	$(0, a, b, a+b, c, a+c, b+c, a+b+c, d, a+d, b+d, a+b+d, c+d, a+c+d, b+c+d, a+b+c+d) - 1/2(a+b+c+d)$
$5 \leqslant n \leqslant 7$	$(2n+1)_v$	$(0, -a_{n-1}+a_n, \ldots, -a_1+a_n, a_n, a_1+a_n, \ldots, a_{n-1}+a_n, 2a_n) - a_n$, where $0 < a_i \ll a_{i+1}$, for $1 \leqslant i \leqslant n$

TABLE I. (*Continued.*)

$C_n \equiv \mathrm{SP}(2n)$, $3 \leqslant n \leqslant 8$		
$n=3$	6	$(0,\ -b+c,\ -a+c, a+c, b+c, 2c)-c$
	14	$(0,\ -a+b, a+b, 2b,\ -a+c, a+c, (b+c)^2,\ -a+2b+c,$ $a+2b+c, 2c,\ -a+b+2c, a+b+2c, 2b+2c)-(b+c)$
	14′	$(0, 2a, a+b, 2b, 2a+2b, a+c, b+c, 2a+b+c, a+2b+c,$ $2c, 2a+2c, a+b+2c, 2b+2c, 2a+2b+2c)-(a+b+c)$
$4 \leqslant n \leqslant 8$	$2n$	$(0,\ -a_{n-1}+a_n,\ldots,-a_1+a_n, a_1+a_n,\ldots,$ $a_{n-1}+a_n, 2a_n)-a_n,$ where $0<a_i<a_{i+1}$, for $1 \leqslant i \leqslant n$
$D_n \equiv \mathrm{SO}(2n)$, $4 \leqslant n \leqslant 8$		
$n=4$	8_v	$(0,\ -c+d,\ -b+d,\ -a+d, a+d, b+d, c+d, 2d)-d$
	8_s	$(0, a+b, a+c, b+c, a+d, b+d, c+d, a+b+c+d)-1/2(a+b+c+d)$
	$8_s'$	$(0,\ -a+b,\ -a+c, b+c,\ -a+d, b+d, c+d,\ -a+b+c+d)$ $+1/2(a-b-c-d)$
$n=5$	10_v	$(0,\ -d+e,\ -c+e,\ -b+e,\ -a+e, a+e, b+e, c+e, d+e, 2e)-e$
	16_s	$(0,\ -a+b,\ -a+c, b+c,\ -a+d, b+d, c+d,\ -a+b+c+d,\ -a+e,$ $b+e, c+e,\ -a+b+c+e, d+e,\ -a+b+d+e,\ -a+c+d+e,$ $b+c+d+e)+1/2(a-b-c-d-e)$
	$\overline{16}_s$	$(0, a+b, a+c, b+c, a+d, b+d, c+d, a+b+c+d, a+e, b+e,$ $c+e, a+b+c+e, d+e, a+b+d+e, a+c+d+e, b+c+d+e)$ $-1/2(a+b+c+d+e)$
$6 \leqslant n \leqslant 8$	$(2n)_v$	$(0,\ -a_{n-1}+a_n,\ldots,-a_1+a_n, a_1+a_n,\ldots,$ $a_{n-1}+a_n, 2a_n)-a_n$ where $0<a_i<a_{i+1}$, for $1 \leqslant i \leqslant n$
G_2		
	7	$(0, 2a,\ -a+b, a+b, 3a+b, 2b, 2a+2b)-(a+b)$
	14	$(0,\ -3a+b,\ -a+b, a+b, 3a+b,\ -2a+2b, (2b)^2, 2a+2b,$ $-3a+3b,\ -a+3b, a+3b, 3a+3b, 4b)-2b$

[a] The charge operators are described by nonnegative eigenvalues, written in increasing order. The classification within an operator is based on the following assumption: If $a, b, c, \cdots$ are the free parameters of the operator, a is very much smaller than b, b very much smaller than c, and so on. We write, according to (2) when $K=0$,

$$(0^k 1, a^k 2, b^k 3, \cdots) + \lambda,$$

where k_i is the multiplicity of the ith eigenvalue. Then λ, $\lambda + a$, $\lambda + b$, $\cdots$ are the charges of the particles we consider.

lowed simple Lie groups, according to the restrictions of Sec. 2.

In order to minimize the notations, let us describe the charge values of a d-dimensional multiplet in a given group by

$$Q_d = (0^{k_0}, a_1^{k_1}, \ldots, a_p^{k_p}) + \lambda + K, \tag{2}$$

with

$$\lambda = -\frac{1}{d}\sum_{i=1}^{p} a_i k_i. \tag{3}$$

We have chosen to order the a_i's in the following way:

$$a_i < a_{i+1}, \quad 1 \leqslant i < p, \tag{4}$$

and k_i is the multiplicity of the charge eigenvalue

$$q_i = a_i + \lambda + K. \tag{5}$$

Clearly

$$d = \sum_{i=o}^{p} k_i. \tag{6}$$

In what follows, we will omit the symbol k_i in (2) when it is equal to 1, and introduce the square bracket notation

$$\mathbf{k} = [k_0, \ldots, k_p]. \tag{7}$$

In (3), λ has been defined in such a way as to insure zero trace for Q_d if $K=0$. Hence, Q_d can be, for $K=0$, the charge operator of a simple group.

When K, which we call the shift, is nonzero, the group must contain an extra U(1)factor. We will go back to the shifts in the discussion of Sec. 5.

To obtain the "good" Q_d's of specific representations of simple Lie groups, we use two different techniques: either the Kronecker products or the weight diagrams.

We illustrate the first technique in the case of A_n $(2 \leqslant n \leqslant 15)$, starting from the fundamental $(n+1)$-dimensional representation. We write Q_{n+1} when the $(n+1)$ eigenvalues are all different:

$$Q_{n+1} = (0, a_1, \ldots, a_n) - \frac{1}{n+1}\sum_{i=1}^{n} a_i, \tag{8}$$

TABLE II. "Good" charge operators (see Sec. 2).[a]

Multiplicities $\mathbf{m} = (m_0, m_1, ..., m_6)$ of the charges $(-3, -2, -1, 0, 1, 2, 3)$	Groups and representations	Values of the parameters
7 different charges		
[1, 1, 1, 1, 1, 1, 1]	G_2, 7	$a = 1/2,\ \ b = 5/2$
[1, 1, 1, 2, 1, 1, 1]	SU(3), 8	$a = 1,\ \ b = 3$
	SO(7), 8_s	$a = 1,\ \ b = 2,\ \ c = 3$
	SO(8), 8_s, $8_s'$	$a = 0,\ \ b = 1,\ \ c = 2,\ \ d = 3$
[1, 1, 2, 2, 2, 1, 1]	SU(3), 10	$a = 1,\ \ b = 2$
	SU(4), 10	$a = 1,\ \ b = 2,\ \ c = 3$
	SU(5), 10	$a = 1,\ \ b = 2,\ \ c = 3,\ \ d = 4$
	SO(5), 10_v	$a = 1,\ \ b = 2$
[1, 2, 3, 2, 3, 2, 1]	SP(6), 14	$a = 0,\ \ b = 1,\ \ c = 2$
[1, 3, 3, 2, 3, 3, 1]	SO(9), 16_s	$a = b = c = 1,\ \ d = 3$
	SO(10), 16_s	$a = 0,\ \ b = c = d = 1,\ \ e = 3$
6 different charges		
[1, 1, 1, 0, 1, 1, 1]	SU(4), 6	$a = 3,\ \ b = 4,\ \ c = 5$
[1, 2, 1, 0, 1, 2, 1]	SO(7), 8_s	$a = b = 1,\ \ c = 4$
	SO(8), 8_s, $8_s'$	$a = 0,\ \ b = c = 1,\ \ d = 4$
	SO(8), $8_s'$	$a = 1,\ \ b = c = 2,\ \ d = 3$
[1, 1, 2, 0, 2, 1, 1]	SO(8), 8_s	$a = b = 1/2,\ \ c = 3/2,\ \ d = 7/2$
	SO(8), $8_s'$	$a = 1/2,\ \ b = 3/2,\ \ c = d = 5/2$
5 different charges		
[0, 1, 1, 2, 1, 1, 0]	SU(3), 6	$a = 1,\ \ b = 2$
	SU(4), 6	$a = 1,\ \ b = 2,\ \ c = 3$
[1, 1, 0, 2, 0, 1, 1]	SU(4), 6	$a = 2,\ \ b = 3,\ \ c = 5$
[1, 0, 1, 2, 1, 0, 1]	SU(4), 6	$a = 1,\ \ b = 3,\ \ c = 4$
[0, 1, 2, 1, 2, 1, 0]	G_2, 7	$a = 1/2,\ \ b = 3/2$
[0, 1, 2, 2, 2, 1, 0]	SU(3), 8	$a = 1,\ \ b = 2$
	SO(7), 8_s	$a = b = 1,\ \ c = 2$
	SO(8), 8_s, $8_s'$	$a = 0,\ \ b = c = 1,\ \ d = 2$
[0, 2, 1, 2, 1, 2, 0]	SO(8), $8_s'$	$a = b = 1/2,\ \ c = 3/2,\ \ d = 5/2$
[1, 2, 0, 2, 0, 2, 1]	SO(8), $8_s'$	$a = 1/2,\ \ b = c = 3/2,\ \ d = 7/2$
[1, 0, 2, 2, 2, 0, 1]	SO(8), 8_s	$a = 1/2,\ \ b = c = 3/2,\ \ d = 5/2$
4 different charges		
[0, 1, 1, 0, 1, 1, 0]	SO(5), 4_s	$a = 1,\ \ b = 3$
[1, 1, 0, 0, 0, 1, 1]	SO(5), 4_s	$a = 1,\ \ b = 5$
[1, 0, 1, 0, 1, 0, 1]	SO(5), 4_s	$a = 2,\ \ b = 4$
[0, 1, 2, 0, 2, 1, 0]	SU(4), 6	$a = 2,\ \ b = c = 3$
[0, 2, 1, 0, 1, 2, 0]	SU(4), 6	$a = b = 3,\ \ c = 4$
[1, 2, 0, 0, 0, 2, 1]	SU(4), 6	$a = 4,\ \ b = c = 5$
[1, 0, 2, 0, 2, 0, 1]	SU(4), 6	$a = 2,\ \ b = c = 4$
[0, 1, 3, 0, 3, 1, 0]	SO(8), 8_s	$a = b = c = 1/2,\ \ d = 5/2$
	SO(8), $8_s'$	$a = 1/2,\ \ b = c = d = 3/2$
[0, 3, 1, 0, 1, 3, 0]	SO(8), $8_s'$	$a = b = c = 1/2,\ \ d = 7/2$
		$a = b = c = 3/2,\ \ d = 5/2$
[0, 2, 2, 0, 2, 2, 0]	SO(7), 8_s	$a = 0,\ \ b = 1,\ \ c = 3$
	SO(8), 8_s, $8_s'$	$a = b = 0,\ \ c = 1,\ \ d = 3$
[1, 3, 0, 0, 0, 3, 1]	SO(8), 8_s	$a = b = c = 1/2,\ \ d = 9/2$
	SO(8), $8_s'$	$a = 3/2,\ \ b = c = d = 5/2$
[1, 0, 3, 0, 3, 0, 1]	SO(7), 8_s	$a = b = c = 2$
	SO(8), 8_s, $8_s'$	$a = 0,\ \ b = c = d = 2$
	SO(8), 8_s	$a = b = c = 1,\ \ d = 3$
[1, 3, 0, 0, 3, 3, 0]	SU(5), 10	$a = 3,\ \ b = c = d = 4$
	SU(5), $\overline{10}$	$a = b = 0,\ \ c = 1,\ \ d = 4$
[0, 3, 3, 0, 0, 3, 1]	SU(5), 10	$a = b = 0,\ \ c = 1,\ \ d = 4$
	SU(5), $\overline{10}$	$a = 3,\ \ b = c = d = 4$

TABLE II. (*Continued*.)

Multiplicities $\mathbf{m} = (m_0, m_1,...,m_6)$ of the charges $(-3, -2, -1, 0, 1, 2, 3)$	Groups and representations	Values of the parameters
3 different charges		
[1, 0, 0, 2, 0, 0, 1]	SO(5), 4_s	$a = b = 3$
[0, 1, 0, 2, 0, 1, 0]	SO(5), 4_s	$a = b = 2$
[0, 0, 1, 2, 1, 0, 0]	SO(5), 4_s	$a = b = 1$
[0, 2, 0, 2, 0, 2, 0]	SU(4), 6	$a = b = 2, \quad c = 4$
[0, 0, 2, 2, 2, 0, 0]	SU(4), 6	$a = b = 1, \quad c = 2$
[0, 3, 0, 2, 0, 3, 0]	SO(8), $8'_s$	$a = b = c = 1, \quad d = 3$
[0, 0, 3, 2, 3, 0, 0]	SO(8), $8'_s$	$a = b = c = 1/2, \quad d = 3/2$
2 different charges		
[0, 2, 0, 0, 0, 2, 0]	SO(5), 4_s	$a = 0, \quad b = 4$
[0, 0, 2, 0, 2, 0, 0]	SO(5), 4_s	$a = 0, \quad b = 2$
[0, 3, 0, 0, 0, 3, 0]	SU(4), 6	$a = b = 0, \quad c = 4$
[0, 0, 3, 0, 3, 0, 0]	SU(4), 6	$a = b = 0, \quad c = 2$

[a] All the groups which don't need an extra U(1) factor to reproduce the charges of the quarks and the leptons of one family, with multiplicity 1 at most for the leptons (except for the neutrinos), and 3 at most for the quarks, are presented here. The values of the parameters must be divided by three in the table and must be put in the corresponding charge operator of Table I, to get the correct charge assignment.

To compute the charge operators described in the lemmas of the Sec. 4, recall that

$$Q_p = (0, a_1,...,a_{p-1}) + \lambda + K \quad \text{in SU}(p),$$

$$Q_{2n} = (-a_n,..., -a_1, a_1,...,a_n) \quad \text{in SO}(2n) \text{ and SP}(2n),$$

$$Q_{2n+1} = (-a_n,..., -a_1, 0, a_1,...,a_n) \quad \text{in SO}(2n+1).$$

If one has the structure

$$\mathbf{m} = [m_0, m_1,...,m_6], \quad \sum_{i=0}^{6} m_i = p$$

with m_f the first nonzero multiplicity, then the first m_f parameters of Q_p are equal to zero, the next $m_{(f+1)}$ are equal to 1/3, the next $m_{(f+2)}$ are equal to 2/3, and so on.

If $p = 2n + 1$, with $\mathbf{m}$ symmetric, then $m_3 = 1$, and the m_4 first parameters are equal to 1/3, the next m_5 are equal to 2/3, and the next m_6 are equal to 1. If $p = 2n$, with $\mathbf{m}$ symmetric, then $m_3 = 0$ or $m_3 = 2$. In the former case, the first m_4 parameters are equal to 1/3, the next m_5 are equal to 2/3, and the next m_6 are equal to 1. In the latter case, $a_1 = 0$, the next m_4 parameters are equal to 1/3, the next m_5 are equal to 2/3, and the next m_6 are equal to 1.

Example:

$$\mathbf{m} = [0, 1, 2, 1, 2, 1, 0], \quad \Sigma m_i = 7,$$

$$\text{SO}(7), \quad 7_v, \quad a_1 = a_2 = 1/3, \quad a_3 = 2/3,$$

$$\text{SU}(7), \quad 7, \quad a_1 = a_2 = 1/3, \quad a_3 = 2/3, \quad a_4 = a_5 = 1, \quad a_6 = 4/3.$$

TABLE IIIA. List of d-dimensional irreducible representations in simple Lie groups ($d \leqslant 16$).

SU(3)	1, 3, $\bar{3}$, 6, $\bar{6}$, $8 \equiv \bar{8}$, 10, $\overline{10}$, 15, $\overline{15}$, 15′, $\overline{15'}$
SU(4)	1, 4, $\bar{4}$, $6 \equiv \bar{6}$, 10, $\overline{10}$, $15 \equiv \overline{15}$
SU(5)	1, 5, $\bar{5}$, 10, $\overline{10}$, 15, $\overline{15}$
SU(6)	1, 6, $\bar{6}$, 15, $\overline{15}$
SU(N) ($7 \leqslant N \leqslant 16$)	1, N, $\bar{N}$
SO(5)	1, 4_s, 5_v, 10_v, 14_v, 16_s
SO(7)	1, 7_v, 8_s
SO(9)	1, 9_v, 16_s
SO($2N+1$) ($5 \leqslant N \leqslant 7$)	1, $(2N+1)_v$
SP(6)	1, 6, 14, 14′
SP($2N$) ($4 \leqslant N \leqslant 8$)	1, $2N$
SO(8)	1, 8_v, 8_s, $8'_s$
SO(10)	1, 10_v, 16_s, $\overline{16}_s$
SO($2N$) ($6 \leqslant N \leqslant 8$)	1, $2N$
G_2	1, $7 \equiv \bar{7}$, $14 \equiv \overline{14}$

TABLE IIIB. List of simple Lie Groups having d-dimensional irreducible representations ($d \leqslant 16$).

$d = 3$	SU(3) (2 types)
$d = 4$	SU(4) (2 types); SO(5)
$d = 5$	SU(5) (2 types); SO(5)
$d = 6$	SU(3) (2 types); SU(4); SU(6) (2 types); SP(6)
$d = 7$	SU(7) (2 types); SO(7); G_2
$d = 8$	SU(3); SU(8) (2 types); SO(7); SO(8) (3 types); SP(8)
$d = 9$	SU(9) (2 types); SO(9)
$d = 10$	SU(3) (2 types); SU(4) (2 types); SU(5) (2 types); SU(10) (2 types); SO(5); SO(10); SP(10)
$d = 11$	SU(11) (2 types); SO(11)
$d = 12$	SU(12) (2 types); SO(12); SP(12)
$d = 13$	SU(13) (2 types); SO(13)
$d = 14$	SU(14) (2 types); SO(5); SO(14); SP(6) (2 types); SP(14); G_2
$d = 15$	SU(3) (4 types); SU(4); SU(5) (2 types); SU(6) (2 types); SU(15) (2 types); SO(15)
$d = 16$	SU(16) (2 types); SO(5); SO(9); SO(10) (2 types); SO(16); SP(16)

according to formulas (2), (3), and (4), when $K=0$. From expression (8), it is easy to get $Q_{\overline{n+1}}$, if one notes that, under charge conjugation, the generic charge operator Q_d, given in (2), becomes

$$Q_{\bar{d}} = (0^{\bar{k}_0}, \bar{a}_1^{\bar{k}_1}, \ldots, \bar{a}_p^{\bar{k}_p}) + \bar{\lambda} + \bar{K}, \tag{9}$$

where

$$\bar{a}_i < \bar{a}_{i+1} \quad 1 \leqslant i \leqslant p \tag{10}$$

$$\bar{a}_i = a_p - a_{p-i}, \tag{11}$$

$$\bar{k}_i = k_{p-i}, \tag{12}$$

$$\bar{\lambda} = -a_p + \lambda, \tag{13}$$

$$\bar{K} = -K. \tag{14}$$

Any other representation of SU(n) is obtained by suitably Young-symmetrized Kronecker products of the fundamental representation.

To get the corresponding charge operator, just recall that for Q_d and $Q_{d'}$, whose eigenvalues are respectively X_i and Y_j, the eigenvalues of $Q_d \times Q_{d'}$ are $X_i + Y_j$. Now, any symmetrization of the $(d \times d')$ product gives a representation whose charge operator is obtained by picking up out of the set $(X_i + Y_j)$ the sums with the symmetries of the representation. Special cases when some of the a_i's in (8) are equal can be derived from the results above.

For the B_n, C_n, and D_n classes and G_2, we use the weight diagrams.

If a simple Lie group is of rank n, there exist n fundamental weights L^i, defined by the formula

$$2(L_i, \alpha_j)/(\alpha_j)^2 = \delta_{ij}, \tag{15}$$

where the α_i's are the simple roots and where the scalar product is defined in an n-dimensional Euclidean space, with orthogonal basis (e_i) and suitable normalization. We refer to it as to the weight space. Each representation has its highest weight L, of multiplicity one, expressed as a linear combination of the L^i's, with nonnegative integer coefficients (Dynkin labels):

$$L = \lambda_i L^i, \quad \lambda_i \in N, \quad 1 \leqslant i \leqslant n. \tag{16}$$

In order to find all the weights of 1, say, d-dimensional representation $\underline{d}$, we use the Weyl group. This one is generated by the reflections ρ acting in the weight space

$$\rho(\ell) = \ell - \mu\alpha, \tag{17}$$

where ℓ is a weight, α are the roots, and μ is an integer defined by

$$\mu = 2(\ell,\alpha)/(\alpha,\alpha). \tag{18}$$

All weights deduced from each other by a Weyl transform are equivalent, and have the same multiplicity. Starting from the highest weight L of multiplicity 1, we can find the set $W(L)$ of the weights equivalent to L. These weights are points of the Z-lattice obtained by adding arbitrary combinations of the n simple roots to the highest weight L. They enclose a finite number of points of the lattice, which are also weights of the representation. We write them as

$$\ell_{(f)} = b^j_{(f)} e_j, \quad 1 \leqslant j \leqslant n \text{ and } 1 \leqslant f \leqslant d, \tag{19}$$

$\ell_{(1)}$ being, by definition, the highest weight of the representation $\underline{d}$. The multiplicity $n(f)$ of the weight $\ell_{(f)}$ is given by Freudenthal's recursive formula

$$\{(\ell_{(1)} + \delta, \ell_{(1)} + \delta) - (\ell_{(f)} + \delta, \ell_{(f)} + \delta)\} n(f) = 2 \sum_{\alpha > 0} \sum_{i=1}^{\infty} n(f')(\ell_{(f)} + i\alpha, \alpha), \tag{20}$$

where α are the positive roots (linear combinations of the simple roots, with nonnegative coefficients);

$$\delta = \tfrac{1}{2} \sum_{\alpha > 0} \alpha \tag{21}$$

and $n(f')$ is the multiplicity of the weight

$$\ell_{(f')} = \ell_{(f)} + i\alpha. \tag{22}$$

Recall that

$$n(1) = 1. \tag{23}$$

The techniques and developments can be found in Ref. 6.

Now, the charge axis has *a priori* any direction in the n-dimensional weight space. Indeed by a good diagonalization any operator Q can be expressed as a linear combination of n operators H_i which form a suitably normalized Cartan basis:

$$Q_d = h_i H_i, \quad 1 \leqslant i \leqslant n. \tag{24}$$

(A generator Y has to be added to this expression if $K \neq 0$.) The h_i's are free parameters, and Q_d has eigenvalues given by [see (2) and (3)],

$$a_f + \lambda = (h, b_{(f)}), \tag{25}$$

i.e., q_f up to a factor K, by (5). The $b_{(f)}$ are fixed by the representation $\underline{d}$ [see (19)].

To get the charge operator we want, we have to discuss the values of the parameters h_i, components of an n-vector in the weight space, which we call charge axis. Its direction can be chosen within a closed Weyl chamber of the weight diagram. (An open Weyl chamber w being a connected component of $W - \cup_\alpha P_\alpha$, where W is the weight space and P_α the hyperplane orthogonal to the root α. Using Weyl transforms on a Weyl chamber, one finds $W - \cup_\alpha P_\alpha$). Let us now describe the limitations we can put on the charge axis in the weight space of B_n, C_n, D_n, and G_2.

A. B_n (roots: $\pm e_i$, $\pm e_i \pm e_j$) and C_n (roots: $\pm 2e_i$, $\pm e_i \pm e_j$)

Under a reflection through the hyperplane orthogonal to the root e_i (or $2e_i$), the charge axis transforms like

$$(h_1, h_2, \ldots, h_i, \ldots, h_n) \to (h_1, h_2, \ldots, -h_i, \ldots, h_n). \tag{26}$$

Thus we can consider that all the a_i's are nonnegative. Under a reflection through the hyperplane orthogonal to the root $e_i - e_j$, the charge axis transforms like

$$(h_1, \ldots, h_i, \ldots, h_j, \ldots, h_n) \to (h_1, h_2, \ldots, h_j, \ldots, h_i, \ldots, h_n). \tag{27}$$

Thus we can consider h_i smaller than h_j for $i<j$. The Weyl chamber we select is the connected part of the weight space delimited by the hyperplanes orthogonal to the roots $e_1, e_2 - e_1, e_3 - e_2, \ldots, e_n - e_{n-1}$, and whose vectors have a positive scalar product with all the roots above. The charge axis is such that

$$0 \leqslant h_1 \leqslant h_2 \leqslant \cdots \leqslant h_n. \tag{28}$$

B. D_n (roots: $\pm e_i \pm e_j$)

Under a reflection through the hyperplane orthogonal to the root $e_i - e_j$ the charge axis transforms like (27) while under a reflection through the hyperplane orthogonal to the root $e_i + e_j$, the charge axis transforms like

$$(h_1,h_2,\ldots,h_i,\ldots,h_j,\ldots,h_n)\mapsto(h_1,h_2,\ldots,-h_j,\ldots,-h_i,\ldots,h_n). \quad (29)$$

This transformation shows that we can take the h_i's as positive numbers except perhaps, h_n, whose sign is to be discussed. However, there exists a Z_2 symmetry which interchanges representations, and, in particular, the two spinorial representations whose dominant weights are

$$L = \tfrac{1}{2}(e_1 + e_2 + \cdots + e_n) \quad (30)$$

and

$$L = \tfrac{1}{2}(e_1 + e_2 + \cdots - e_n). \quad (31)$$

A negative value for a_n in the first spinorial representation corresponds to a positive value for a_n in the second one, of same dimension, and obtained by a Z_2 symmetry. The argument is easily generalized to any pair of Z_2-related representations. We then have

$$0 \leqslant h_1 \leqslant h_2 \leqslant \cdots \leqslant h_n. \quad (32)$$

The Weyl chamber we select here is the connected part of the weight space delimited by the hyperplanes orthogonal to the roots $e_1 - e_2, e_1 - e_2, e_2 - e_3, \ldots, e_{n-1} - e_n$, and whose vectors have a positive scalar product with all the roots above.

C. G_2 (roots: $\pm 2e_1'$, $\pm 3e_2'$, $\pm(e_1' + e_2')$, $\pm(e_1' + 3e_2')$, $\pm(-e_1' + 3e_2')$, $\pm(-e_1' + e_2')$)

Here

$$e_1' = \tfrac{1}{2}e_1 \quad \text{and} \quad e_2' = (\sqrt{3}/6)e_2. \quad (33)$$

The Weyl chamber we select is the connected part of the weight space delimited by the roots $2e_2'$ and $e_1' + 3e_2'$, and whose vectors have a positive scalar product with these two roots. Then we have

$$0 \leqslant h_1 \leqslant 3h_2. \quad (34)$$

The classification of Table I takes the restrictions (4), (28), (32), and (34) into account. Furthermore, we give there the eigenvalues in an increasing order, supposing a_{i+1} much greater than a_i. When some a_{i+1} become closer to a_i, the order of the eigenvalues can change, as is easily seen.

4. DISCUSSION OF PARAMETERS VALUES

We first deal with the trivial cases, which are essentially treated using the following lemmas. Let us slightly extend the notation **k** (7) introducing

$$\mathbf{m} = [m_0,\ldots,m_6], \quad (35)$$

where some m_i can be zero. They are the multiplicities of particles and antiparticles, listed in the following order $(e^-,\bar{u},d,(\nu \text{ or } \bar{\nu}),\bar{d},u,e^+)$.

Definition: The structure **m** is symmetric if $m_0 = m_6$, $m_1 = m_5$, $m_2 = m_4$.

Lemma 1: **m** symmetric implies that the sum of the charges of the particles put in the irreducible representation is zero.

Lemma 2: In all groups of Table I, nonsymmetric structures only appear in SU(n) ($3 \leqslant n \leqslant 16$) and in representation 16_s of SO(10).

Lemma 3: Each time one takes N particles (and/or antiparticles) so that the sum of electric charges is zero, with the multiplicity structure **m** there is a good N-dimensional charge operator in SU(N) associated with them.

Lemma 4: Each time one takes $2N$ particles (and/or antiparticles) plus one neutrino, with **m** symmetric, there is a good $(2N+1)$-dimensional charge operator in SO($2N+1$) associated with them.

Lemma 5: Each time one takes $2N$ particles (and/or antiparticles) with **m** symmetric, there is a good $2N$-dimensional charge operator in SO($2N$) and Sp($2N$) associated with them.

The proofs of the five lemmas are obvious when looking at the charge operator structures of Table I.

As a consequence of the lemmas, when one constructs a multiplicity structure **m**, four cases are possible.

(a) **m** is symmetric and $\Sigma_{i=0}^{6} m_i = 2N+1$. Then, there is at least the trivial charge operator Q_{2N+1} in Su($2N+1$) [resp. SO($2N+1$)] which describes the particles of the theory.

(b) **m** is symmetric and $\Sigma_{i=0}^{6} m_i = 2N$. Then, there is at least the trivial charge operator Q_{2N} in Su($2N$) [resp. SO($2N$), Sp($2N$)] which describes the particles of the theory.

(c) **m** is not symmetric, but the sum of the charge is zero and $\Sigma_{i=0}^{6} m_i = N$. Then, there is at least the trivial charge operator Q_N (and $Q_{\bar{N}}$) in SU(N) which describes the particles of the theory.

(d) **m** is not symmetric, and the sum of the charges is not zero. then, a shift is necessary, and this will be discussed in Sec. 5.

To construct explicitly the charge operators for the trivial cases, i.e., those described by *a* to *c*, one can simply use the prescriptions of the footnote to Table II.

The nontrivial cases can be found in Table II, and are obtained by the methods explained in Sec. 3.

5. THE SHIFTS

We look now at theories where the relevant group is $G \times U(1)$, with G a simple group. Indeed, if we want to consider groups where K, as defined in Eq. (2), is nonzero, this is the simplest possibility. In the preceding section, we have tried to present our results quite generally, so that they can apply to any representation of dimension less than or equal to 16, and for a set of at most seven different eigenvalues.

In this section, we will be more specific and limit ourselves to the actual problem of putting the known quarks and leptons in several representations of G.

A quick analysis shows that if we want the weak and strong interactions to act on the multiplets, the only partitions are:

9 + 6,	10 + 6:	$(e^-,\bar{u}^3,\nu^{1,2},\bar{d}^3,e^+) + (d^3,u^3)$,	(36a)
6 + 9,	7 + 9:	$(e^-,\nu^{1,2},\bar{d}^3,e^+) + (\bar{u}^3,d^3,u^3)$,	(36b)
6 + 10	:	$(e^-,\nu,\bar{d}^3,e^+) + (\bar{u}^3,d^3,\nu,u^3)$,	(36c)
5 + 10,	6 + 10:	$(e^-,\nu^{1,2},\bar{d}^3) + (\bar{u}^3,d^3,u^3,e^+)$,	(36d)
5 + 11,	:	$(e^-,\nu,\bar{d}^3) + (\bar{u}^3,d^3,\nu,u^3,e^+)$,	(36e)

$$8+7, \quad 9+7: \quad (e^-, \bar{u}^{\,3}, \nu^{1,2}, \bar{d}^{\,3}) + (d^{\,3}, u^3, e^+), \tag{36f}$$
$$8+8 \quad : \quad (e^-, \bar{u}^{\,3}, \nu, \bar{d}^{\,3}) + (d^{\,3}, \nu, u^3, e^+), \tag{36g}$$
$$1+14, \quad 1+15: \quad e^+ + (e^-, \bar{u}^{\,3}, d^{\,3}, \nu^{1,2}, \bar{d}^{\,3}, u^3), \tag{36h}$$
$$1+1+14 \quad : \quad e^+ + \nu + (e^-, \bar{u}^{\,3}, d^{\,3}, \nu, \bar{d}^{\,3}, u^3), \tag{36i}$$

with, in formulas (36b) and (36c) the possibility of interchanging $\bar{u}$ and $\bar{d}$.

Table III shows that the allowed sums are

$$6+10 \quad \text{in SU(3) and SU(4)}, \tag{37a}$$
$$7+8 \quad \text{in SO(7)}, \tag{37b}$$
$$8+\bar{8} \quad \text{in SU(8) and SO(8)}, \tag{37c}$$
$$1+14 \quad \text{in SU(14), SO(5), SO(14)}, \tag{37d}$$
$$1+1+14 \quad \text{SP(6), SP(14), } G_2, \tag{37e}$$
$$1+15 \quad \text{in SU(15),SO(15),SU(6),SU(5),SU(4),SU(3)}. \tag{37f}$$

Among these possibilities, only SU(8), SU(14), and SU(15) provide a good charge assignment. According to the notations of Sec. 3, and Table II, we write

$$8+\bar{8}, \quad \text{SU(8)}, \quad a=b=c=1/3, \quad d=1, \quad e=f=g=4/3, \quad K=-1/4, \quad \bar{K}=1/4, \tag{38a}$$

$$1+14, \quad \text{SU(14)}, \quad a=b=c=1/3, \quad d=e=f=2/3, \quad g=1, \quad h=i=j=4/3, \quad k=l=m=5/3, \quad K_{14}=1/14, \quad K_1=1, \tag{38b}$$

$$1+1+14, \quad \text{SU(14)}, \quad a=b=c=1/3, \quad d=e=f=2/3, \quad g=1, \quad h=i=j=4/3, \quad k=l=m=5/3, \quad K_{14}=-1/14, \quad K_1=1, \quad K_{1'}=0, \tag{38c}$$

$$1+15, \quad \text{SU(15)}, \quad a=b=c=1/3, \quad d=e=f=2/3, \quad g=h=1, \quad i=j=k=4/3, \quad l=m=n=5/3, \quad K_{15}=-1/15, \quad K_1=1. \tag{38d}$$

TABLE IV.

Group	Rep	Fermionic content	Values of parameters	Characteristics
SU(16)	16	$(e^-, \bar{u}^3, d^3, \nu^2, \bar{d}^3, u^3, e^+)$	$a=b=c=1, \quad d=e=f=2,$ $g=h=3, \quad i=j=k=4,$ $l=m=n=5, \quad o=6$	—Complex representation —Not anomaly free —Asymptotically free ($\lambda=176$) —SU(3) color present —SU(2) weak present
SO(16)	16	idem	$a=0, \quad b=c=d=1, \quad e=f=g=2,$ $h=3$	—Self-conjugated representation —Anomaly free —Asymptotically free ($\lambda=77$) —SU(3) color present —SU(2) acts on (u, d), (ν, e^-) and $(\bar{u}, \bar{d})$, (ν, e^+)
SP(16)	16	idem	idem	—Self-conjugated representation —Anomaly free —Asymptotically free ($\lambda=99$) —SU(3) color present —SU(2) acts on (u, d), (ν, e^-) and $(\bar{u}, \bar{d})$, (ν, e^+)
SO(10)	16_s	idem	$a=0, \quad b=c=d=1, \quad e=3$	—Complex representation —Anomaly free —Asymptotically free ($\lambda=22$) —SU(3) color present —SU(2) weak present
SO(9)	16_s	idem	$a=b=c=1, \quad d=3$	—Self-conjugated representation —Anomaly free —Asymptotically free ($\lambda=19$) —SU(3) color present —SU(2) acts on (u, d), (ν, e^-) and $(\bar{u}, \bar{d})$, (ν, e^+)
SU(15)	15	$(e^-, \bar{u}^3, d^3, \nu, \bar{d}^3, u^3, e^+)$	$a=b=c=1, \quad d=e=f=2, \quad g=3,$ $h=i=j=4, \quad k=l=m=5,$ $n=6$	—Complex representation —Not anomaly free —Asymptotically free ($\lambda=165$) —SU(3) color present —SU(2) weak present
SO(15)	15_v	idem	$a=b=c=1, \quad d=e=f=2,$ $g=1$	—Self-conjugated representation —Anomaly free —Asymptotically free ($\lambda=71$) —SU(3) color present —SU(2) acts on (u, d), $(\bar{u}, \bar{d})$ but not on (ν, e^-)

TABLE IV. (*Continued.*)

Group	Rep	Fermionic content	Values of parameters	Characteristics
SO(8)	$8'_s + 8_v$	$(\bar{u}^3, \nu^2, u^3)$ $+ (e^-, d^3, \bar{d}^3, e^+)$	$a=b=c=1,\ \ d=3$ $a=0,\ \ b=c=d=2$	—Self-conjugated representation —Anomaly free —Asymptotically free ($\lambda = 16$) —SU(3) color present —No SU(2) present
SU(8)	$8+\bar{8}$	$(\bar{u}^3, \nu, \bar{d}^3, e^+)$ $+ (e^-, d^3, \nu, u^3)$	$a=b=0,\ \ c=2,\ \ d=e=f=3$ $g=5$ or $a=b=c=2,\ \ d=3,\ \ e=f=g=5$	—Self-conjugated representation —Anomaly free —Asymptotically free ($\lambda = 44$) —SU(3) color present —SU(2) weak is taken to act only on 8
		$(e^-, \bar{u}, d, \nu, \bar{d}^2, u)$ $+ (\bar{u}^2, d^2, \nu, \bar{d}, u, e^+)$	$a=1,\ \ b=2,\ \ c=3,\ \ d=e=4,$ $f=g=5$	—Self-conjugated representation —Anomaly free —Asymptotically free ($\lambda = 44$) —No SU(3) color —No SU(2) weak present
SO(7)	$7_v + 8_s$	$(\bar{u}^3, \nu^2, u^3)$ $+ (e^-, d^3, \bar{d}^3, e^+)$	$a=b=c=2$	—Self-conjugated representation —Anomaly free —Asymptotically free ($\lambda = 13$) —SU(3) color present —No SU(2) weak
SU(7)	$7+\bar{7}+1$	$(e^-, \bar{u}, d, \bar{d}^2, u^2)$ $+ (\bar{u}^2, d^2, \bar{d}, u, e^+) + \nu$	$a=1,\ \ b=2,\ \ c=d=4,\ \ e=f=5$ or $a=0,\ \ b=c=1,\ \ d=3,\ \ e=4,\ \ f=5$	—Self-conjugated representation —Anomaly free —Asymptotically free ($\lambda = 38$) —No SU(3) color —No SU(2) weak
		(e^-, d^3, u^3) $+ (\bar{u}^3, \bar{d}^3, e^+) + \nu$	$a=b=c=2,\ \ d=e=f=5$ or $a=b=0,\ \ c=d=e=3,\ \ f=5$	—Self-conjugated representation —Anomaly free —Asymptotically free ($\lambda = 38$) —SU(3) color present —No SU(2) acting on (ν, e^-)
SU(5)	$\bar{5}+10$ $(5+\overline{10})$	$(e^-, \nu, \bar{d}^3)$ $+ (\bar{u}^3, d^3, u^3, e^+)$ (and c.c.)	$a=b=0,\ \ c=1,\ \ d=4$	—Complex representation —Anomaly free —Asymptotically free ($\lambda = 13$) —SU(3) color present —SU(2) weak present in $\bar{5}+10$
		(d^3, ν, e^+) $+ (e^-, \bar{u}^3, \bar{d}^3, u^3)$ (and c.c.)	$a=1,\ \ b=c=d=4$	—Complex representation —Anomaly free —Asymptotically free ($\lambda = 13$) —SU(3) color present —SU(2) weak present in $5+\overline{10}$
SU(4)	$4+\bar{4}+6+1$	$(d^3, e^+) + (e^-, \bar{d}^3)$ $+ (\bar{u}^3, u^3) + \nu$	$a=b=0,\ \ c=4$	—Self-conjugated representation —Anomaly free —Asymptotically free ($\lambda = 11$) —SU(3) color present —No SU(2) weak

[a] The values of the parameters must be divided by three to get the spectrum of the particles.

Let us note that a more general approach to the shift problem is to examine all the sums

$$n_1 + \cdots + n_i = 15 \quad (\text{or } 16), \tag{39}$$

where n_i is the dimension of the ith multiplet in a given simple group G. This analysis leads to extra cases like, for instance,

$$8_v + 8_s, \quad \mathrm{SO}(8), \quad a=b=c=1/6, \quad d=1/2, \quad K_{8_v} = -1/2, \quad K_{8_s} = 1/2 \tag{40}$$

the particle assignment being, for this, $(e^-, \bar{u}^3, d^3, \nu) + (\nu, \bar{d}^3, u^3, e^+)$.

Unfortunately, all these cases, deduced from too general assumptions, don't seem to have any hope of surviving.

6. SOME COMMENTS

Among all the charge operators described in Sec. 4 and Table II, we pick up those which reproduce the charge spectrum

$$e^-, \quad \bar{u}^3, \quad d^3, \quad \nu^{1,2}, \quad \bar{d}^3, \quad u^3, \quad e^+ \tag{41}$$

within one or several irreducible representations of a same Lie group.

We discuss in Table IV the SU(3) color content of the groups G when the charge operator comes out well, and look for an SU(2) group which could describe the weak interactions.

Beside the complex and anomaly-free character of the representations, we also give in each case the maximal multiplicity of these representations (λ), bounded by the asymptotic freedom requirement.

7. CONCLUSIONS

We have presented here all the simple Lie groups where the charge operator has 16 different eigenvalues at most. Our work has been guided by the idea of putting the leptons and the quarks of one family in a representation (reducible or irreducible) where the charge operator comes correctly out.

This can be seen as a classification of GUT candidate groups at the lowest level, i.e., with the charge restriction as unique condition on the representation. However, if we require SU(3) color and SU(2) weak to be subgroups of the relevant group G, SU(16) seems to be the best candidate [except, of course, the SU(5) and SO(10) groups].

A straightforward extension of this paper would be to allow several families in the representation. In particular, the extension of SU(16) to SU(48) or $SU(16)^3$, which are the maximal $SU(N)$ symmetries when one considers three generations of particles, becomes more and more popular.

On the other hand, the idea of inserting an extra $U(I)$ factor in the theory (shift problem) doesn't lead to any appealing alternative.

Anyhow, the above classification of representations is still useful if one succeeds in curing the weaknesses of the GUT candidates listed in Table IV, using, for instance, new concepts or new tools.

ACKNOWLEDGMENTS

I would like to thank Professor Jean Nuyts for helpful discussions and constant encouragement. I also acknowledge support from the Fonds National de la Recherche Scientifique.

[1]M. Gell-Mann, P. Ramond, and R. Slansky, Rev. Mod. Phys. **50**, 721 (1978); S. L. Adler, Phys. Rev. **177**, 2425 (1969); P. H. Frampton, Phys. Lett. B **88**, 299 (1979).

[2]We have recently studied this problem in composite models where representations of more than one simple group are considered. Part of our results are given in Appendix B of our paper [D. B. Fairlie, J. Nuyts, and A. Taormina, Phys. Rev. D **27**, 264 (1983).] One of the cases is the Rishon model [H. Harari, Phys. Lett. B **86**, 83 (1979)], which has met considerable interest even though it had no SU(3) color in the starting form.

[3]The reducible representations are analyzed in Sec. 6, when we add several irreducible representations.

[4]However, as will be seen in Sec. 2, a restriction on the multiplicity of each eigenvalue limits the generality of the tables.

[5]If the neutrinos are massless, one may consider one, two, or more left-handed neutrinos; if they have a nonzero mass, one must take a ν_L and a $\bar{\nu}_L$ which provides the ν_R.

[6]J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer-Verlag, New York, 1972).

Cohomology of Lie algebras with a nontrivial center

Bani Mitra, B. R. Sitaram, and K. C. Tripathy
Department of Physics and Astrophysics, University of Delhi, Delhi-110007, India

(Received 28 December 1982; accepted for publication 1 April 1983)

Let $\mathbf{g}$ be a Lie algebra and assume $X \in Z(\mathbf{g})$ (the center of $\mathbf{g}$). Let ϕ: $\mathbf{g} \to \operatorname{End} V$ be such that $\phi(X) = 1_V$. We show that $H^n_\phi(\mathbf{g}, V) = 0\ \forall n > 0$. The result applies to the nonrelativistic Poincaré, harmonic oscillator and Heisenberg algebras, and also to $\mathbf{g} \oplus \mathbf{u}(1)$ where $\mathbf{g}$ is a semisimple Lie algebra. We also give here the cohomology groups for the first three algebras with the adjoint action, giving explicit computations of H^1 and H^2, respectively, for the first two algebras.

PACS numbers: 02.20.Sv, 02.90. + p

I. INTRODUCTION

Cohomology theory has long been known to be of great importance in solving various problems in mathematics.[1] In recent years, some of these problems, e.g., classification of principal bundles (Chern class theory),[2] solutions of field equations in nonabelian gauge theories (de Rham theory),[3] classification of Lie algebra extensions and groups and Lie algebra contraction schemes (group/Lie algebra cohomology theory[4,5]) have been of interest to physicists also. The explicit computations of cohomology groups, remains, however, as a problem. In this paper, we wish to study some of the cohomology groups of Lie algebras with a nontrivial center. In this regard, it is worthwhile noting that complete results are known for finite dimensional representations of semisimple Lie algebras: $H^n_\phi(\mathbf{g}, V) = 0\ \forall n > 0$, $\mathbf{g}$ semisimple over $\mathbb{R}$ or $\mathbb{C}$, V being a $\mathbf{g}$ module.[4]

Our major result in this direction is the following.

Theorem: Let $\mathbf{g}$ be a Lie algebra over $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$) and let $Z(\mathbf{g})$ be its center. Assume $0 \neq X \in Z(\mathbf{g})$. Let ϕ: $\mathbf{g} \to \operatorname{End} V$ be a representation such that $\phi(X) = 1_V$. Then $H^n_\phi(\mathbf{g}, V) = 0\ \forall\ n > 0$. The proof of this theorem is given in Sec. II. The theorem can be applied immediately to several physically interesting Lie algebras, e.g., the nonrelativistic Poincaré algebras in $(m+n)$ dimensions, $\mathbf{nrp}(m,n)$, the Heisenberg algebra in n dimensions, $\mathbf{h}(n)$, the harmonic oscillator algebra in n dimensions, $\mathbf{ho}(n)$, and Lie algebras of the form $\mathbf{h} \oplus \mathbf{u}(1)$ where $\mathbf{h}$ is semisimple over $\mathbb{C}$. The common feature of the first three sets of these algebras is that representations satisfying $\phi(X) = 1_V$ are precisely the physically relevant ones as is shown in Sec. II.

In Sec. III, we give the groups $H^p_\phi(\mathbf{g}, \mathbf{g})$ where ϕ: $\mathbf{g} \to \operatorname{End} \mathbf{g}$ is the adjoint action and $\mathbf{g}$ is one of the following algebras: $\mathbf{nrp}(1,1)$, $\mathbf{ho}(1)$, $\mathbf{h}(1)$ and give the explicit calculations for $\mathbf{nrp}(1,1)$ ($p = 1$ case) and $\mathbf{h}(1)$ ($p = 2$ case).

II. RESUME OF COHOMOLOGY THEORY AND PROOF OF THE THEOREM

Let $\mathbf{g}$ be a Lie algebra over $\mathbb{F}$, and let V be a vector space. Let ϕ: $\mathbf{g} \to \operatorname{End} V$ be a representation. We define $C^n(\mathbf{g}, V) = \{ f: \mathbf{g}^n \to V: f$ is $\mathbb{F}$ linear and

$$f(X_1, ..., X_i, ..., X_j, ... X_n) = -f(X_1, ..., X_j, ..., X_i, ..., X_n)\ \forall X_i \in \mathbf{g}\}.$$

We define ∂^n: $C^n(\mathbf{g}, V) \to C^{n+1}(\mathbf{g}, V)$ by

$$\partial^n f(X_1, ..., X_{n+1}) = \sum_{i=1}^{n+1} (-1)^{i+1} \phi(X_i) f(X_1, ..., \hat{X}_i, ..., X_{n+1}) + \sum_{i<j}^{n+1} (-1)^{i+j} f([X_i, X_j]\ X_1, ..., \hat{X}_2, ..., \hat{X}_j, ..., X_{n+1}).$$

It follows that $\partial^{n+1}\partial^n = 0\ \forall n$. Further, we define

$C^0(\mathbf{g}, V) = \{ f$: $\mathbb{F} \to V$: f is $\mathbb{F}$ linear$\}$ and

$\partial^0 f(X)$: $\phi(X) f(1)\ \forall\ X \in \mathbf{g}$.

Let $Z^n_\phi(\mathbf{g}, V) = \operatorname{Ker} \partial^n$ and $B^n_\phi(\mathbf{g}, V) = \operatorname{Im} \partial^{n-1}$. Finally, $H^n_\phi(\mathbf{g}, V) = Z^n_\phi(\mathbf{g}, V)/B^n_\phi(\mathbf{g}, V)$.

Proof of the theorem

Let $f \in C^n(\mathbf{g}, V)$. Then

$$\partial^n f(Y_1, Y_2, ..., Y_{n+1}) = 0$$
$$\Rightarrow \sum_{i=1}^{n+1} (-1)^{i+1} \phi(Y_i) f(Y_1, ..., \hat{Y}_i, ..., Y_{n+1}) + \sum_{i<j}^{n+1} (-1)^{i+j} f([Y_i, Y_j], \hat{Y}_i, ..., \hat{Y}_j ..., Y_{n+1}) = 0 \tag{1}$$

and

$$\partial^n f(X, Y_1, ..., Y_n) = 0$$
$$\Rightarrow f(Y_1, ..., Y_n) + \sum_{i=1}^{n} (-1)^i \phi(Y_i) f(X, Y_1, ..., \hat{Y}_i, ..., Y_n) + \sum_{i<j}^{n} (-1)^{i+j} f([Y_i, Y_j], X, ..., \hat{Y}_i, ..., Y_j, ..., \hat{Y}_j, ..., Y_n) = 0$$
$$\forall\ Y_i \in \mathbf{g}/\{X\}, \tag{2}$$

where $\hat{Y}_i$ implies that Y_i is to be omitted. Using Eq. (2), we define

$$f(Y_1, ..., Y_n) = \sum_{i=1}^{n} (-1)^{i+1} \phi(Y_i) f(X, Y_1, ..., \hat{Y}_i, ..., Y_n) + \sum_{i<j}^{n} (-1)^{i+j+1} f([Y_i, Y_j], X, Y_1, ..., \hat{Y}_i, ..., \hat{Y}_j ..., Y_n). \tag{3}$$

It is possible to show that Eq. (1) is satisfied. Thus $f \in Z^n(\mathbf{g}, V)$ implies that (3) is satisfied. Now let $f' = f - \partial\omega$, i.e.,

 0022-2488/84/030443-02$02.50

$$f'(Y_1,...,Y_n) = f(Y_1,...,Y_n) - \sum_{i=1}^{n}(-1)^i\phi(Y_i)\omega(Y_1,...,\widehat{Y}_i,...,Y_n) - \sum_{i<j}^{n}(-1)^{i+j}\omega([Y_i,Y_j],Y_1,...,\widehat{Y}_i,...,\widehat{Y}_j,...,Y_n),$$

$$f'(X,Y_1,...,Y_{n-1}) = f(X,Y_1,...,Y_{n-1}) - \omega(Y_1,...,Y_{n-1}) - \sum_{i=1}^{n-1}(-1)^i\phi(Y_i)\omega(X,Y_1,...,\widehat{Y}_i,...,Y_{n-1}) - \sum_{i<j}^{n-1}(-1)^{i+j}\omega([Y_i,Y_j],X,...,\widehat{Y}_i,...,\widehat{Y}_j,...,Y_{n-1})$$

$$\forall\ Y_i \in \mathbf{g}/\{X\}. \tag{4}$$

Choose

$$\omega(Y_1,...,Y_{n-1}) = f(X,Y_1,...,Y_{n-1}), \tag{5a}$$

$$\omega(X,Y_1,...,Y_{n-2}) = 0 \quad \forall\ Y_i \in \mathbf{g}/\{X\}. \tag{5b}$$

Then $f'(X,Y_1,...,Y_{n-1}) = 0$. Also from (4), (5a), (5b), and (3), $f'(Y_1,...,Y_n) = 0$, and hence $f = \partial\omega$, i.e., $H^n_\phi(\mathbf{g},V) = 0\ \forall n>0$.

Remark 1: The Lie algebras $\mathbf{nrp}(m,n)$, $\mathbf{ho}(n)$, and $\mathbf{h}(n)$ share the property that $X\in[\mathbf{g},\mathbf{g}]$. As usual, this implies that if $\phi(X)\sim 1_V$ then V is necessarily infinite dimensional for $\operatorname{tr}\phi(X) = \dim V = \operatorname{tr}([\phi(\mathbf{g}),\phi(\mathbf{g})]) = 0$, which implies that either $\dim V$ is trivial or that $\operatorname{tr}\phi(X)$ is not well defined. On the other hand if $\mathbf{g} = \mathbf{h}\oplus\mathbf{u}(1)$, then $X\notin[\mathbf{g},\mathbf{g}]$ and the representation may be finite dimensional.

Remark 2: The above theorm may be generalized in the following sense.

Theorem: Let $\mathbf{g} = \mathbf{g}_0 + \mathbf{g}_1$ be a Z_2 graded Lie algebra over $\mathbb{F}$. Let $0\neq X\in Z(\mathbf{g})\cap\mathbf{g}_0$ and let ϕ: $\mathbf{g}\to\operatorname{End} V$ be a representation such that $\phi(X) = 1_V$. Then $H^n_\phi(\mathbf{g},V) = 0\ \ \forall n>0$. The proof of this theorem may be carried out on the same lines as above. In this connection, see also the explicit computations of $H^n_\phi(\mathbf{g},V)$ for the Dirac algebras reported earlier.[7]

III. $H^p_\phi(\mathbf{g},\mathbf{g})$ FOR THE ADJOINT ACTION

(i) $\mathbf{g} = \mathbf{nrp}(1,1)$. The Lie algebra is defined by $\{X_1,X_2,X_3,X_4\}$ with $[X_1,X_2] = X_4$, $[X_1,X_3] = X_2$, other commutators being zero, and with $Z(\mathbf{g}) = \{X_4\}$. Here X_1 is the velocity boost operator, X_2, X_3 the space and time translation operators, and X_4 is the identity operator.

Let $f\in Z^1(\mathbf{g},\mathbf{g})$ and let $f(X_i) = \alpha_{ij}X_j$. Then $\partial f(X_i,X_j) = 0 = [X_i,f(X_i)] + [f(X_i),X_j] - f([X_i,X_j])$.

Using the commutation relations, we get

$$\alpha_{31} = \alpha_{21} = \alpha_{23} = \alpha_{41} = \alpha_{42} = \alpha_{43} = 0,$$

$$\alpha_{44} = \alpha_{33} + 2\alpha_{11},\ \ \alpha_{22} = \alpha_{33} + \alpha_{11},\ \ \alpha_{24} = \alpha_{32}.$$

Thus, one is left with seven free parameters so $Z^1(\mathbf{g},\mathbf{g}) = \mathbb{F}^7$. Now let $f' = f - \partial\omega$ where $\omega\in C^0(\mathbf{g},\mathbf{g})$ and $f'(X_i) = \alpha'_{ij}X_i$, $\omega(1) = \beta_iX_i$. Then, from $f'(X_i) = f(X_i) - [X_i,\omega(1)]$, one finds $\alpha'_{13} = \alpha_{13} - \beta_4$, $\alpha'_{14} = \alpha_{14} - \beta_2$, $\alpha'_{32} = \alpha_{32} + \beta_1$, $\alpha'_{24} = \alpha_{24} + \beta_1$, $\alpha'_{ij} = \alpha_{ij}$ for all other i,j. Choosing $\beta_4 = \alpha_{13}$, $\beta_2 = \alpha_{14}$, $\beta_1 = -\alpha_{12}$, we may write f' as

$$f'(X_1) = \alpha'_{11}X_1 + \alpha'_{12}X_2,$$

$$f'(X_2) = \alpha'_{11}X_1 + \alpha'_{32}X_4,$$

$$f'(X_3) = \alpha'_{32}X_2 + \alpha'_{44}X_4,$$

$$f'(X_4) = 2\alpha'_{11}X_4.$$

Hence

$$H^1_\phi(\mathbf{g},\mathbf{g}) = \mathbb{F}^4.$$

Similarly one may show that $H^2 = \mathbb{F}^6$, $H^3 = \mathbb{F}^5$, $H^6 = \mathbb{F}^2$, and $H^n = 0\ \forall\ n>4$.

(ii) $\mathbf{g} = \mathbf{h}(1)$. The Lie algebra is defined by $\{X_1,X_2,X_3\}$ with $[X_1,X_2] = X_3$. Here, X_1 is the position, X_2 is the momentum, and X_3 is the identity operator. One can show that $H^1 = \mathbb{F}^4$.

Let $f\in Z^2(\mathbf{g},\mathbf{g})$ and let $f(X_i,X_j) = \alpha_{ijk}X_k$. Then

$$\partial f(X_1,X_2,X_3) = 0 = [X_1,f(X_2,X_3)] - [X_2,f(X_1,X_3)] + [X_3,f(X_1,X_2)].$$

Simplifying, we get $\alpha_{232} = -\alpha_{131}$. Thus $Z^2 = \mathbb{F}^8$. Let $f' = f - \partial\omega$, $\omega\in C^1(\mathbf{g},\mathbf{g})$ with $f'(X_i,X_j) = \alpha'_{ijk}X_k$, $\omega(X_i) = \beta_{ij}X_j$. Then we get

$$\alpha'_{121} = \alpha_{121} + \beta_{32},\quad \alpha'_{122} = \alpha_{122} + \beta_{31},$$

$$\alpha'_{123} = \alpha_{123} - \beta_{11} - \beta_{22} + \beta_{33},$$

$$\alpha'_{233} = \alpha_{233} - \beta_{31},\quad \alpha'_{133} = \alpha_{133} + \beta_{32},$$

$$\alpha'_{ijk} = \alpha_{ijk} \quad \forall \text{ other } i,j,k.$$

Choosing $\beta_{32} = -\alpha_{121}$, $\beta_{31} = -\alpha_{122}$, $\beta_{11} = \beta_{22} = 0$, $\beta_{33} = -\alpha_{123}$,

we may write f' as

$$f'(X_1,X_2) = 0,$$

$$f'(X_1,X_3) = \alpha'_{131}X_1 + \alpha'_{132}X_2 + \alpha'_{133}X_3,$$

$$f'(X_2,X_3) = \alpha'_{231}X_1 - \alpha'_{131}X_2 + \alpha'_{233}X_3,$$

and hence $H^2_\phi(\mathbf{g},\mathbf{g}) = \mathbb{F}^5$. Similarly, one gets $H^3 = \mathbb{F}^2$, and $H^n = 0\ \forall\ n>3$.

(iii) $\mathbf{g} = \mathbf{ho}(1)$.[6] The Lie algebra is defined by $\{X_1,X_2,X_3,X_4\}$ with $[X_1,X_2] = X_3$, $[X_1,X_3] = -X_2$, $[X_2,X_3] = X_4$, all other commutators being zero. The cohomology groups are given by $H^1 = \mathbb{F}^2$, $H^2 = \mathbb{F}^6$, $H^3 = \mathbb{F}^2$, $H^4 = \mathbb{F}^1$, and $H^n = 0$, $n>4$. (Here, X_1 is the harmonic oscillator Hamiltonian, X_2 is the position, X_3 the momentum, and X_4 the identity operator.)

[1] J. Erven and B. J. Falkowski, *Low Order Cohomology and Applications, Lecture Notes in Mathematics*, V. F. **877** (Springer-Verlag, Berlin, 1981).

[2] S. S. Chern, *Complex Manifolds without Potential Theory Universitext* (Springer-Verlag, Berlin, 1979).

[3] M. E. Mayer, in *Fibre Bundles and Gauge Theories, Lecture Notes in Physics*, Vol. 67 (Springer-Verlag, Berlin, 1977).

[4] C. Chevalley and S. Eilenberg, Trans. Am. Math. Soc. **63**, 85 (1948); N. Jacobson, *Lie Algebras* (Dover, New York, 1979).

[5] R. Herman, Commun. Math. Phys. **2**, 251 (1966); **3**, 53 (1966); **3**, 75 (1966).

[6] B. R. Sitaram and K. C. Tripathy "Geometric Quantization and the Harmonic Oscillator Group Revisted" (unpublished).

[7] B. R. Sitaram and K. C. Tripathy, Indian J. Pure Appl. Math. **13**, 672 (1982).

Specializations of integrable systems and affine Lie algebras

F. Guil Guerrero[a)]
Departamento de Métodos Matemáticos de la Física, Facultad de Ciencias Físicas, Universidad Complutense, Madrid-3, Spain

(Received 3 March 1983; accepted for publication 11 August 1983)

We analyze a number of restrictions on the evolution systems associated with the zero-curvature equations corresponding to the extended Dynkin diagram $A_{n-1}^{(1)}$. The resulting specialized evolution systems contain exponential terms, like a third-order differential equation previously derived in Ref. 5.

PACS numbers: 02.30.Jr, 02.20.Sv

1. INTRODUCTION

Recent work by Drinfel'd and Sokilov,[1] shows how equations of the Korteweg–de Vries (KdV) and modified KdV type can be generalized through the construction of zero-curvature equations[2] in the context of the affine Lie algebras (an adequate description of these algebras can be found in Helgason's book in Ref. 3). Even though the number of examples of two-dimensional integrable systems has been notably augmented and most of them conveniently classified,[4] there are still some equations, known to have an integrable behavior, which are not included in the general scheme of the theory. Among others, we can cite Eqs. (5.5)

$$p_t = 2p_{xxx} - p_x^3 - 3(c_1^2\ e^{2p} + c_0^2\ e^{-2p})\,p_x$$

first derived in Ref. 5, that will constitute, besides the generalizations of them, the main object of the present work.

The specialization of an integrable system consists in constructing another system compatible with the former and containing a number of functions less than those appearing in the original equations. This method has been widely discussed in the works of Refs. 5–10 from different viewpoints.

The main result of this paper establishes that the formula (4.4) determines a specialization for the zero-curvature equations associated with the diagram $A_{n-1}^{(1)}$. The method we have followed here makes use of what we call in Sec. 2 an invariant manifold: A condition over a duplicate evolution system which is to be specialized. These invariant manifolds constitute an infinite set, but only the first two elements of it are of practical use if one wants to maintain the discussion in an algebraic (local) setting. In Sec. 3 we give the necessary information about the symmetries of evolution equations based on automorphisms of order 2 of the extended Dynkin diagram of $A_{n-1}^{(1)}$, in order to derive in Sec. 4 the desired specializations of these equations by examining the solutions of a pair of coupled equations. Finally, in Sec. 5 we present some examples of the above in the simplest cases.

2. INVARIANT MANIFOLDS

In this section we shall determine a number of constraints (invariant manifolds) that will be used in the specialization of the zero-curvature equations. We start with some information about these equations that we shall be using in what follows, though the interested reader will learn correctly what we sketch here, in Refs. 1 and 2.

Let $\mathfrak{g}$ denote a simple Lie algebra and σ an automorphism of mth order in $\mathfrak{g}$ with the corresponding $\mathbb{Z}_m$-gradation induced by it: $\mathfrak{g} = \mathfrak{g}_0 \oplus \cdots \oplus \mathfrak{g}_{m-1}$ ($\mathfrak{g}_i$ is the subspace of $\mathfrak{g}$ where σ acts as ϵ^i, ϵ a primitive mth root of unity). Define the infinite-dimensional Lie algebra

$$L(\mathfrak{g}, \sigma) = \bigoplus_{-\infty}^{\infty} \lambda^i\, \mathfrak{g}_{i \bmod m},$$

its elements being of the form $\Sigma\, \lambda^i X_i$, $X_i \in \mathfrak{g}_{i \bmod m}$, understood as a formal power series in the parameter λ with coefficients in $\mathfrak{g}$. Let $F \in \mathfrak{g}_1$ be a semisimple element and $u \in \mathfrak{g}_0 \cap \operatorname{Im} \operatorname{ad} F$. We consider the algebra $B = \mathbb{C}[u_i^{(j)}]$ of differential polynomials in the variables u_i which determine u on some fixed set of vectors expanding $\mathfrak{g}_0 \cap \operatorname{Im} \operatorname{ad} F$ and define the usual derivation ∂,

$$\partial u_i^{(j)} = u_i^{(j+1)}.$$

We set $\tilde{\mathfrak{g}} = B \otimes \mathfrak{g}$, $\tilde{L} = L(\tilde{\mathfrak{g}}, \sigma)$ and extend ∂ to $\tilde{\mathfrak{g}}$, $\tilde{L}$ acting coefficientwise. Define the differential operator

$$\mathscr{L} = \partial - \mathscr{U}, \quad \mathscr{U} \equiv -u + \lambda F. \tag{2.1}$$

Thus, for each v_0 in the center of the centralizer of F we can form the element $\mathscr{V}$ of $\tilde{L}$, $\mathscr{V} = v_0 + v_1 \lambda^{-1} + \cdots$ satisfying

$$\partial \mathscr{V} = [\mathscr{U}, \mathscr{V}]. \tag{2.2}$$

Let k be some fixed positive integer, take $v_0 \in \mathfrak{g}_{k \bmod m}$ as above and set $V^+ = (\lambda^k \mathscr{V})^+$, $V^- = (\lambda^k \mathscr{V})^-$, i.e., the positive and negative parts in the powers of λ, respectively. The equivalence of

$$[\partial - \mathscr{U},\ \ \partial_t - V^+] = 0 \tag{2.3}$$

is then obtained with the evolution equations

$$-u_t = \partial v_k \tag{2.4}$$

defining the derivation ∂_t commuting with ∂. These are the zero-curvature equations described in Ref. 2.

We shall assume that some concrete faithful representation of $\mathfrak{g}$ by $n \times n$ matrices has been given for $\mathfrak{g} = \mathrm{sl}(n)$, and σ being the Coxeter transformation,[2] $L(\mathfrak{g}, \sigma)$ is the affine Lie algbera $A_{n-1}^{(1)}$.

Take $\mathscr{L}_u$, $\mathscr{L}_{\bar{u}}$ like the operator $\mathscr{L}$ in (2.1) but with different markings u and $\bar{u}$. We define $B_2 = \mathbb{C}[u_i^{(j)}, \bar{u}_i^{(j)}]$ the algebra of differential polynomials in u and $\bar{u}$ with the differential grading in which both u_i and $\bar{u}_i$ have degree 1 and ∂ increases the degree by 1 so that the degree of $u_i^{(j)}$ or $\bar{u}_i^{(j)}$ is equal to $j+1$. Let $\bar{B}_2$ be the extension of B_2 containing the

[a)] Partially supported by the Comisión Asesora de Investigación Científica y Técnica.

integrals of the elements in B_2 described in Ref. 11, being ∂ surjective and maintaining the grading.

Proposition 2.1: Let the operators $\mathscr{L}_u$ and $\mathscr{L}_{\bar{u}}$ be given. Then there exists an unique element K,

$$K = 1 + X_1 \lambda^{-1} + X_2 \lambda^{-2} + \cdots,$$

with X_i homogeneous of degree i in $\bar{B}_2 \times \mathfrak{g}$, such that

$$\mathscr{L}_{\bar{u}} = K \mathscr{L}_u K^{-1}.$$

Proof: We use the representation of $\mathfrak{g}$ for which $F = \operatorname{diag}(1, \epsilon, ..., \epsilon^{n-1})$, being then u and $\bar{u}$ circulants, as is done in Ref. 6. We want to prove that $\mathscr{L}_{\bar{u}} K = K \mathscr{L}_u$ determines the X_i of K, that satisfy

$$\bar{u} - u = [F, X_1],$$

$$\partial X_i + \bar{u} X_i - X_i u = [F, X_{i+1}], \quad i = 1,2,\ldots .$$

From the decomposition of $\mathfrak{g} = \ker \operatorname{ad} F \oplus \operatorname{Im} \operatorname{ad} F$ in diagonal and nondiagonal parts, it follows at once that X_i is determined by the equations above, except for integration constants that appear when calculating its diagonal part. However, such constants must be set equal to zero if we want to maintain the homogeneity requisite, so that the X_i are uniquely calculated in $\bar{B}_2 \otimes \mathfrak{g}$. ∎

We shall regard the element K (which could be interpreted in the gauge group for the manifold of operators $\mathscr{L}_u$) as a function of u and $\bar{u}$. Then the X_i are expressed in terms of u, $\bar{u}$, and their derivatives, but containing integrals as well, forcing us to introduce the extension $\bar{B}_2$ to deal with them.

We denote also by ∂_t the unique extension to $\bar{B}_2$ of the evolutionary derivation ∂_t of B_2, homogeneous of degree k, defined by (2.3).

Proposition 2.2: Let $\mathscr{L}_u$ and $\mathscr{L}_{\bar{u}}$ satisfy Eq. (2.3) for some fixed value of k. We have

$$\partial_t K = K V_u^- - V_{\bar{u}}^- K$$

and

$$\partial_t K = V_{\bar{u}}^+ K - K V_u^+. \tag{2.5}$$

Proof: Equation (2.3) can be written as

$$[\mathscr{L}_u, \partial_t + V_u^-] = 0$$

(this follows from (2.2), $[\mathscr{L}_u, \mathscr{V}_u] = 0$, and the decomposition of $\mathscr{V}\lambda^k$ in positive and negative parts). Conjugation by K gives us

$$[\mathscr{L}_{\bar{u}}, \partial_t + K V_u^- K^{-1} - (\partial_t K) K^{-1}] = 0,$$

which when compared with $[\mathscr{L}_{\bar{u}}, \partial_t + V_{\bar{u}}^-] = 0$ tells us that $K V_u^- K^{-1} - V_u^- - (\partial_t K) K^{-1}$ is a negative power series in λ commuting with $\mathscr{L}_{\bar{u}}$, say $\Sigma_{i>1} Y_i \lambda^{-i}$, with Y_i homogeneous of degree $i + k$. It is easily seen that $Y_i \equiv 0$ for all i, proving our first assertion. The second equation (2.5) results from the relation

$$\mathscr{V}_{\bar{u}} = K \mathscr{V}_u K^{-1}$$

that follows from the definition (2.2) of $\mathscr{V}_u$ and $\mathscr{V}_{\bar{u}}$. ∎

We pick out the coefficient of λ^{-i} on both sides of (2.5), obtaining

$$\partial_t X_i = \sum_{j=0}^{k} (v_j(\bar{u}) X_{i+j} - X_{i+j} v_j(u)).$$

Then, each $X_i(u,\bar{u})$ does satisfy an equation of the form

$$\partial_t X_i = \mathscr{F}(X_i, X_{i+1}, ..., X_{i+k}) \tag{2.6}$$

with a different $\mathscr{F}$ for each i, containing explicitly the $v_j(u)$, $v_j(\bar{u})$.

Lemma: If u, $\bar{u}$ are such that $X_i(u, \bar{u}) = 0$ for some fixed i, then $X_{i+l} = 0$ for $l = 0,1,\ldots$.

Proof: Our assertion, the vanishing of X_{i+l} with X_i, follows from the equations determining the elements X_j after imposing $X_i \equiv 0$. So, the nondiagonal part of X_{i+1} is zero, and the diagonal one constant, this constant being zero in order to maintain the homogeneity of K. ∎

Remark: We are not interested in analyzing at this point the solutions of $X_i(u, \bar{u}) = 0$ (which could be interpreted as an integro-differential equation), for arbitrary i. That will be done in Sec. 4 in a more concrete situation.

Now we are able to get the result we are interested in. From (2.6) we conclude that $\partial_t X_i = 0$ if $X_i = 0$, since $\mathscr{F}$ vanishes with $X_i, X_{i+1}, ..., X_{i+k}$ and that does happen if $X_i = 0$; thus we have

Proposition 2.3: The condition $X_i(u, \bar{u}) = 0$ is consistent with the evolution of u and $\bar{u}$ determined by (2.4) for $i = 1,2,\ldots$.

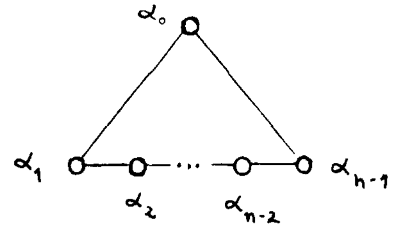
3. INVOLUTIONS

It is known that simple Lie algebras admit a set of outer automorphisms induced by an automorphism of their Dynkin diagrams.[3] It is clear that one can extend to the affine Lie algebras this procedure, in order to determine a collection of automorphisms through their corresponding system of roots. In turn we shall obtain a number of symmetries (often called invariance transformations) for the evolution equations associated with them. These symmetries are of importance for us not by themselves. They become relevant in our present context when combined with the invariant manifolds of Proposition 2.3. Then we shall obtain the specializations of the zero-curvature equations (2.4) that we shall perform in the next section.

Being more concretely interested in the algebra $A_{n-1}^{(1)}$, we give the relevant information about the generators and the root system in the representation we shall be using from now on. Let x_i, y_i, h_i, $i = 0,1,...,n-1$ be vectors of $\mathfrak{g} = \mathrm{sl}(n)$ defined by

$$x_i = E_{i-1,i}, \quad y_i = E_{i,i-1}, \quad h_i = E_{i-1,i-1} - E_{ii}, \quad i \in \mathbb{Z}_n,$$

where $E_{i,j}$ is the $n \times n$ matrix with 1 in place (i,j) and zero elsewhere. Thus $x_i \in \mathfrak{g}_1, y_i \in \mathfrak{g}_{n-1}, h_i \in \mathfrak{g}_0$, the subspaces $\mathfrak{g}_0$, $\mathfrak{g}_1$, $\mathfrak{g}_{n-1}$ being the eigenspaces corresponding to the eigenvalues $1, \epsilon, \epsilon^{n-1}$ for the Coxeter transformation acting by conjugation of the diagonal matrix $(1, \epsilon, \epsilon^2, ..., \epsilon^{n-1})$, $\epsilon^n = 1$. We set $e_i = \lambda x_i, f_i = \lambda^{-1} y_i$, $i = 0,1,...,n-1$. Therefore, e_i, f_i, h_i generates $A_{n-1}^{(1)}$ satisfying the defining relations of this algebra.[3] Further we introduce the linear forms, roots, over the linear span of vectors h_i, $\alpha_i(h_j) = a_{ij}$. The numbers a_{ij} determine the Cartan matrix of $A_{n-1}^{(1)}$ with values $a_{ij} = 2\delta_{ij} - \delta_{i,j-1} - \delta_{i-1,j}$, and the corresponding extended Dynkin diagram



Besides the rotational symmetry

$$\theta(\alpha_i) = \alpha_{i+1}$$

producing the "Bäcklund transformations" [12] for Eq. (2.4) associated with this diagram, we shall consider the following automorphisms of order 2, involutions.

Let $\bar{n}$ be one of the numbers $0,1,\dots,n-1$; we define the automorphism τ on $\mathbb{Z}_n$ by

$$\tau(i) = \bar{n} - i, \tag{3.1}$$

and extend it to the roots $\tau(\alpha_i) = \alpha_{\tau(i)}$ (we use the same symbol τ in both cases). We have a reflection of the root diagram for each $\bar{n}$, determined by one of the axes of symmetry of the corresponding regular polygon, with the roots placed in its vertices. Together with the rotation $\vartheta(i) = i+1$, they constitute the generators of the complete group of symmetries: the dihedral group.

We define automorphisms on $\mathfrak{g} = \mathrm{sl}(n)$ and $A^{(1)}_{n-1}$ in order to get the desired symmetries of Eqs. (2.4). Since the x_0, $x_1,\dots,x_{n-1}$ generates $\mathfrak{g}$, we determine τ on $\mathfrak{g}$ by setting $\tau(x_i) = -x_{\tau(i)}$. That fixes $\tau(y_i) = -y_{\tau(i)}$ and $\tau(h_i) = h_{\tau(i)}$. For $\lambda^i X$ in L_i, being $L = \oplus L_i$ and $L = A^{(1)}_{n-1}$, we make $\tau(\lambda^i X) = (-1)^i \lambda^i\, \tau(X)$, being the action of τ on the generators of L,

$$\tau(e_i) = e_{\tau(i)}, \quad \tau(f_i) = f_{\tau(i)}, \quad \tau(h_i) = h_{\tau(i)}. \tag{3.2}$$

Now, we make τ an automorphism on $B \otimes L$ in the natural way: $\tau(B \otimes L) = B \otimes \tau(L)$ and look at its action on Eq. (2.4). That equation $-u_t = \partial v_k$ results in

$$-\partial_t \tau(u) = \partial \tau(v_k). \tag{3.3}$$

Since τ preserves the $\mathbb{Z}_n$-grading of $\mathfrak{g}$, $\tau(u)$ plays the same role in (3.3) as u did in (2.4). Equation (3.3) is Eq. (2.4) for $\tau(u)$ in place of u, whenever $\tau(v_k)$ coincides with $v_k(\tau(u))$. We analyze the relation between $\tau(v_k)$ and $v_k(\tau(u))$ by studying Eq. (2.2) for $\mathscr{V} = v_0 + v_1 \lambda^{-1} + \cdots$.

Let $Z_{k \bmod n}(F)$ denote the centralizer of $F = \Sigma_i\, x_i$ in $\mathfrak{g}_{k \bmod m}$.

Proposition 3.1: For $v_0 \in Z_{k \bmod n}(F)$ the relation

$$\tau(v_k(u)) = (-1)^{k+1} v_k(\tau(u)) \tag{3.4}$$

holds, and τ defines a symmetry for Eq. (2.4), $-u_t = \partial v_k(u)$, with k odd.

Proof: Equation (2.2) transforms under τ in

$$\partial \tau(\mathscr{V}) = [\tau(\mathscr{V}),\ \tau(u) - \lambda F].$$

Now, $\tau(\mathscr{V}) = c\mathscr{V}(\tau(u))$ since $\tau(v_0) = cv_0$ with $c^2 = 1$. [$\tau(v_0)$ must be proportional to v_0, due to the fact that $\dim Z_{k \bmod n}(F) = 1$ and that $\tau(v_0) \in Z_{k \bmod n}(F)$]. But $Z_{k \bmod n}(F) = \mathbb{C}\Pi_\theta\ [x_1,\dots,x_{\bar{k}}]$ [$\bar{k} \equiv$ the integer on $(0,1,\dots,n-1)$ congruent with $k \bmod n$],

$$[x_1,\dots,x_{\bar{k}}] = \mathrm{ad}\, x_1\, \mathrm{ad}\, x_2 \cdots \mathrm{ad}\, x_{\bar{k}-1}\, x_{\bar{k}},$$

$$\Pi_\vartheta = e + \vartheta + \cdots + \vartheta^{n-1}.$$

Thus, $Z_{k \bmod n}(F)$ is generated by the power $\bar{k}$ of F in the representation introduced above. Because τ commutes with Π_θ, one easily gets (3.4) by taking into account the relation

$$\begin{aligned}\tau([x_1,\dots,x_{\bar{k}}]) &= (-1)^{\bar{k}}\, [x_{\tau(1)} \cdots x_{\tau(\bar{k})}] \\ &= (-1)^{\bar{k}}(-1)^{\bar{k}-1}\, [x_{\tau(\bar{k})}, x_{\tau(\bar{k})+1},\dots, x_{\tau(\bar{k})+\bar{k}-1}],\end{aligned}$$

telling us that $c = -1$. ■

4. SPECIALIZATION OF THE ZERO-CURVATURE EQUATIONS

Before going over the specialization of Eq. (2.4) we are really interested in, let us briefly comment on a more trivial situation closely related to our final result. For it, we consider the conditions $X_i(u, \bar{u}) = 0$ of Proposition 2.3, invariant under the flow (2.4), and use them to construct a set of relations containing a letter u only. This is easily done by taking advantage of our previous results of Sec. 3, and substituting $\tau(u)$ in place of $\bar{u}$, to get $X_i(u,\ \tau(u)) = 0$ that will be consistent for each fixed i with the evolution equations for which we have $\tau(u)$ still as a solution. However, condition $X_i = 0$ is nonlocal and in general is not of much use in constructing specializations of our original equations. But the nonlocality problems can be bypassed at least in cases $i = 1$ and $i = 2$. The first of these will produce nothing new, being of interest here because it is contained in the case $i = 2$.

For $i = 1$ we get simply

$$u = \tau(u), \tag{4.1}$$

which in accordance with Proposition 3.11 is a condition compatible with $-u_t = \partial v_k$ for k odd. Condition (4.1) specifies a collection of diagrams, namely, those corresponding to the subalgebras of $A^{(1)}_{n-1}$ determined by the fixed points of the automorphism τ. The specialized equations in this case are the corresponding ones to these diagrams in the general theory of Ref. 1.

Let us write what the generators' $\{\tilde{e}_i, \tilde{f}_i, \tilde{h}_i\}$ are in each case. For n odd, $n = 2m+1$ and $m > 1$, $\tau(i) = -i$, we have the set of generators $\tilde{e}_i, \tilde{f}_i, \tilde{h}_i$

$$\tilde{e}_0 = e_0, \quad \tilde{f}_0 = f_0,$$

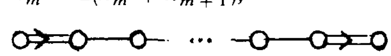
$$\tilde{e}_i = e_i + e_{2m+1-i}, \quad \tilde{f}_i = f_i + f_{2m+1-i},$$

$$\tilde{e}_m = e_m + e_{m+1}, \quad \tilde{f}_m = 2(f_m + f_{m+1}),$$

$$\tilde{h}_0 = h_0,$$

$$\tilde{h}_i = h_i + h_{2m+1-i}, \quad i = 1,\dots,m-1,$$

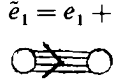
$$\tilde{h}_m = 2(h_m + h_{m+1}),$$



that realize $A^{(2)}_{2m}$ as it is directly checked. From $A^{(1)}_2$ we get $A^{(2)}_2$ by taking

$$\tilde{e}_0 = e_0, \quad \tilde{f}_0 = f_0, \quad \tilde{h}_0 = h_0,$$

$$\tilde{e}_1 = e_1 + e_2, \quad \tilde{f}_1 = 2(f_1 + f_2), \quad \tilde{h}_1 = 2(h_1 + h_2).$$



If we have n even, $n = 2m$, we can take two kinds of nonequivalent automorphisms. The first of them gives $C^{(1)}_m$, $m > 1$, from $A^{(1)}_{2m-1}$ with $\tau(i) = -i$ and setting

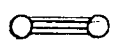
$$\tilde e_0 = e_0,\quad \tilde f_0 = f_0,\quad \tilde h_0 = h_0,$$
$$\tilde e_i = e_i + e_{2m-i},\quad \tilde f_i = f_i + f_{2m-i},$$
$$\tilde h_i = h_i + h_{2m-i},\quad i = 1,...,m-1,$$
$$\tilde e_m = e_m,\quad \tilde f_m = f_m,\quad \tilde h_m = h_m.$$

The vectors $\{\tilde e_i, \tilde f_i, \tilde h_i\}_0^m$ realize $C^{(1)}_m$. But instead of the τ above, we can take also in the diagram $A^{(1)}_{2m-1}$ the automorphism $\tau(i) = 2m-1-i$ producing $D^{(2)}_m$, $m > 2$ when we define

$$\tilde e_i = e_i + e_{2m-1-i},\quad i = 0,1,...,m-1,$$
$$\tilde f_0 = 2(f_0 + f_{2m-1}),\quad \tilde h_0 = 2(h_0 + h_{2m-1}),$$
$$\tilde f_i = f_i + f_{2m-1-i},\quad \tilde h_i = h_i + h_{2m-1-i},\quad i = 1,...,m-2,$$
$$\tilde f_{m-1} = 2(f_{m-1} + f_m),\quad \tilde h_{m-1} = 2(h_{m-1} + h_m).$$

And, in the case of $A^{(1)}_3$, $A^{(1)}_1$ will result, with the same τ, generated by

$$\tilde e_0 = e_0 + e_3,\quad \tilde f_0 = 2(f_0 + f_3),\quad \tilde h_0 = 2(h_0 + h_3),$$
$$\tilde e_1 = e_1 + e_2,\quad \tilde f_1 = 2(f_1 + f_2),\quad \tilde h_1 = 2(h_1 + h_2).$$



Now, we shall examine the next condition, derived from Proposition 2.3, that allows us to maintain our treatment in local terms. The description of $X_2(u, \tau(u)) = 0$ proves to be equivalent to the study of

$$\tau(u) - u = [F, X],$$
$$\partial X + \tau(u) X - Xu = 0, \tag{4.2}$$

where we write $X \equiv X_1$. Note that the obvious solution $X = 0$ to the second of the equations above leads to the case already examined $\tau(u) = u$. Thus the present specialized form of u determined by (4.2) should coincide with some of the previous u determined by (4.1) plus something going to zero with X.

We start by investigating an appropriate decomposition of the subalgebra $\mathfrak{g}_0$ of $n \times n$ diagonal matrices with null trace to which u belongs. In turn, such a decomposition will easily provide the solution of (4.2) we are looking for.

Let ν denote the automorphism of $A^{(1)}_{n-1}$ defined as

$$\nu = \tau\theta,$$

where τ is one of the automorphisms (3.1) and θ the rotation $\theta(i) = i + 1$. Clearly, ν is an automorphism of order 2 like τ.

Proposition 4.1: Let $u \in \mathfrak{g}_0$, then there exist elements $R, S \in \mathfrak{g}_0$ such that

$$\tau(R) = R,\quad \nu(S) = S,$$

providing the linear decomposition

$$u = R + S.$$

Proof: Consider the orthogonal decomposition of $u = u_0 + u_1$ in the eigenspaces of τ corresponding to the eigenvalues ± 1. We denote by Π_τ the orthogonal projector on the subspace of fixed points of τ, and by $\Pi^\perp_\tau = e - \Pi_\tau$. We set

$$a = (\theta + e)(\theta - e)^{-1}\, \Pi^\perp_\tau u,$$

$\theta - e$ being bijective (as linear application) over Ran $\Pi_\tau|_{\mathfrak{g}_0}$. It is easily seen that

$$R = u_0 + a,$$
$$S = u_1 - a$$

gives us the desired vectors R and S in terms of u. ■

The relevance of Proposition 4.1 in the context of Eqs. (4.2) is determined by the fact that after substitution of $u = r_x + S$ in (4.2) (where we have introduced the potential variables r: $R = r_x$) we find the equivalent system

$$(\vartheta - e) S = [F, X],$$
$$\partial X + [r_x, X] = 0, \tag{4.3}$$

The second of Eqs. (4.3) follows from the relation $\tau(S) X - XS = 0$ because $\nu(S) = S$. The first of Eqs. (4.3) allows us to calculate S as a function of X since $\theta - e$ is injective in $\mathfrak{g}_0$, X being determined by a differential equation of the same kind as those appearing in the 2-Toda lattice theory.[1,2] On the other hand, the appearance of the automorphism θ possessed by the diagram $A^{(1)}_{n-1}$ only, makes difficult the extension of our present situation to the remaining diagrams with automorphisms of order 2. Namely, the diagrams $D^{(1)}_{n+1}, B^{(1)}_n, \cdots$.

Let r denote one of the fixed points of τ in $\mathfrak{g}_0$ for one of the cases enumerated at the beginning of this section. We set $d_i \equiv E_{i,i}$ [the diagonal matrix with 1 in place of (i,i) and zero elsewhere]. With these notation conventions, the solution of Eqs. (4.3) is given by the following:

Proposition 4.2: The specialization of the zero-curvature equations (2.4), k odd in $A^{(1)}_{n-1}$, determined by Eqs. (4.3) corresponding to $X_2(u, \tau(u)) = 0$, is provided by

$$u = r_x + \sum_i c_i\, e^{\alpha_i(r)}\, d_{i-1}. \tag{4.4}$$

The c_i are constants satisfying

$$c_i + c_{\tau(i)} = 0,$$

and the α_i are the simple roots of $A^{(1)}_{n-1}$ acting on r viewed in $A^{(1)}_{n-1}$.

Proof: The solution of $\partial X + [r_x, X] = 0$ takes the form

$$X = \sum_i c_i\, e^{\alpha_i(r)}\, y_i,$$

$y_i \in \mathfrak{g}_{n-1}$ being defined in Sec. 3 and the c_i arbitrary constants. By introducing this X in $(\theta - e) S = [F, X]$ and developing the commutator we get

$$[F, X] = \sum_i c_i\, e^{\alpha_i(r)}\, h_i = (e - \theta) \sum_i c_i\, e^{\alpha_i(r)}\, d_{i-1}$$

due to the election made for $h_i = E_{i-1,i-1} - E_{ii} = (e - \theta)\, d_{i-1}$ and that θ extends to act on the E_{ii} by translation in i: $\theta(i) = i + 1$. The restriction of the c_i is obtained after imposing $\nu(S) = S$ in $(\theta - e) S = \Sigma\, c_i\, e^{\alpha_i(r)}\, h_i$. A change of sign in the c_i gives finally (4.4). ■

Corollary:

$$\bar{u} = \tau(u) = r_x + \tau(S) = r_x + \vartheta(S) = r_x + \sum c_i \, e^{\alpha_i(r)} \, d_i \quad (4.5)$$

produces the same specialization of (2.4) as that determined by (4.4).

That results from the fact that $\bar{u}$ satisfies (4.6) like u, as we next deduce.

Proposition 4.3: The evolution of r obeys the equations

$$\partial_t r = (\vartheta^{-1} - e)^{-1} \Pi_v^{\perp} \, w_k(r), \quad w_k(r) \equiv v_k(u). \quad (4.6)$$

Proof: From (2.4) and $u = r_x + S$ we have $-r_{xt} - S_t = \partial w_k$ or

$$-\Pi_v^{\perp} \, r_{xt} = \Pi_v^{\perp} \, w_{kx}$$

since $v(S) = S$. As $\Pi_v^{\perp} \, r = \frac{1}{2}(e - v)\,\tau(r) = -\frac{1}{2}(\theta^{-1} - e)\, r$, thanks to the relations $\tau(r) = r$ and $v\tau = \theta^{-1}$ (see the definition of the automorphism of order 2 v in Proposition 4.1), we obtain (4.6) after rescaling $t \to 2t$ and integrating in x. Note that integration constants do not appear since the evolution of S will determine precisely (4.6) without such constants. ■

For the proof of the Corollary it suffices to insert $\bar{u} = \tau(u)$ in (2.4) and perform analogous calculations to those done with u, to arrive at the same Eq. (4.6).

To conclude, let us observe that some kind of proof about the nontriviality of (4.6), in the sense that the right-hand side of it does in fact depend on the constants c_i, must be given. We postpone this, as well as the study of reductions[13] of these equations, to a future work. In the next section we calculate explicitly some concrete examples of the equations presented above.

5. EXAMPLES

Simplest cases in the specializations (equivalent) (4.4) and (4.5) arise when the resulting equations contain just one function. That situation does happen for $n = 3$ and $n = 4$ in $A_{n-1}^{(1)}$.

For $n = 3$ all the automorphisms of order 2 prove to be equivalent. By examining $\tau(i) = -i$, one finds diagonal matrices u and $\bar{u}$ of the form

$$u = (p_x + ce^p, \; -ce^p, \; -p_x), \quad (5.1)$$

$$\bar{u} = \tau(u) = (p_x, \; ce^p, \; -p_x - ce^p). \quad (5.2)$$

Here p is the dependent variable and c an arbitrary constant. According to Ref. 2, Eqs. (2.4) admit the Hamiltonian form

$$\partial_t \vec{u} = \mathscr{S} \, \partial \frac{\delta H}{\delta \vec{u}}, \quad (5.3)$$

with

$$\vec{u} = \begin{pmatrix} u_0 \\ u_1 \end{pmatrix}, \quad \mathscr{S} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

The first Hamiltonian for which the specializations (5.1), (5.2) apply is

$$H = \tfrac{1}{3} q_{1x}^2 - q_0 q_{1x} + q_0^2 - \tfrac{1}{3} q_1^3 \quad (5.4)$$

(see for example Ref. 14). The variables q_i are expressed in terms of the u_i by means of the generalized Miura transformation

$$\partial^3 + q_1 \partial + q_0 = (\partial - u_0 - u_1)(\partial + u_1)(\partial + u_0).$$

By using (5.2), i.e.,

$$u_0 = p_x, \quad u_1 = ce^p,$$

and (5.3) with H given by (5.4) we get for p

$$p_{xt} = -3(N_x + Np_x)_x,$$

$$p_t = -3(N_x + Np_x),$$

where

$$N = q_{0x} - \tfrac{2}{3} q_{1xx} - q_1^2 = -(\tfrac{1}{6} q_{1xx} + q_1^2),$$

$$q_1 = 2p_{xx} - p_x^2 - c^2 e^{2p}.$$

Note the reduction

$$q_0 = \tfrac{1}{2} q_{1x}$$

induced by the Miura transformation.

In the fifth-order equation

$$p_t = -3(N_x + Np_x),$$

we have for N

$$N = N(0) + \tfrac{1}{3} c^2 (13 p_{xx} - 4p_x^2)\, e^{2p} - c^4 e^{4p},$$

$$N(0) = N|_{c=0}.$$

A third-order equation is obtained by examining the case $n = 4$. The appropriate Hamiltonian

$$H = q_0 - \tfrac{1}{8} q_2^2,$$

when written in terms of the variables u_i through the Miura transformation

$$\partial^4 + q_2 \partial^2 + q_1 \partial + q_0$$
$$= (\partial - u_0 - u_1 - u_2)(\partial + u_2)(\partial + u_1)(\partial + u_0),$$

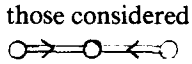
enables us to calculate the evolution equations for which $\tau(i) = 3 - i$, and induces specializations given by

$$u = (p_x + c_1 e^p, \; -c_1 e^p, \; -p_x - c_0 e^{-p}, \; c_0 e^{-p}),$$

$$\bar{u} = \tau(u) = (p_x + c_0 e^{-p}, \; c_1 e^p, \; -p_x - c_1 e^p, \; -c_0 e^{-p}),$$

c_0, c_1 arbitrary constants. The resulting equation[5] is

$$p_t = 2p_{xxx} - p_x^3 - 3(c_1^2 e^{2p} + c_0^2 e^{-2p})\, p_x. \quad (5.5)$$

In this case $n = 4$ too, we can consider the automorphism $\tau(i) = -i$ giving us a system of equations that reduce, for $c = 0$, to those considered in Ref. 4 in connection with the diagram  . First, the specialized forms are

$$u = (p_{0x} + c\, e^{p_0 - p_1}, \; p_{1x}, \; -p_{1x} - c\, e^{p_0 - p_1}, \; -p_{0x}),$$

$$\bar{u} = \tau(u) = (p_{0x}, \; p_{1x} + c\, e^{p_0 - p_1}, \; -p_{1x}, \; -p_{0x} - c\, e^{p_0 - p_1}),$$

where c is constant. With the same Hamiltonian

$H = q_0 - \frac{1}{8} q_2^2$, we get

$$\partial_t p_0 = \tfrac{1}{2} p_{0xxx} + \tfrac{3}{2} p_{1xxx} - 3(p_{0x} + p_{1x})\, p_{1xx} + 3p_{0x} p_{1x}^2$$
$$- p_{0x}^3 + c\left[\tfrac{3}{2} p_{0xx} - \tfrac{9}{2} p_{1xx} + 3p_{0x} p_{1x} - \tfrac{3}{2} p_{0x}^2 + \tfrac{9}{2} p_{1x}^2\right]$$
$$\times e^{p_0 - p_1} + 3c^2 p_{1x}\, e^{2(p_0 - p_1)},$$

$$
\begin{aligned}
\partial_t p_1 = {} & \tfrac{3}{2} p_{0xxx} + \tfrac{1}{2} p_{1xxx} + 3(p_{0x} + p_{1x}) p_{0xx} + 3p_{0x}^2 p_{1x} \\
& - p_{1x}^3 + c\left[\tfrac{9}{2} p_{0xx} - \tfrac{3}{2} p_{1xx} + 3p_{0x} p_{0x} p_{1x} + \tfrac{9}{2} p_{0x}^2 - \tfrac{3}{2} p_{1x}^2\right] \\
& \times e^{p_0 - p_1} + 3c^2 p_{0x} e^{2(p_0 - p_1)}.
\end{aligned}
$$

Analogous systems are found associated with the diagrams $\bigcirc \Rightarrow \bigcirc \Rightarrow \bigcirc$ and $\bigcirc \Leftarrow \bigcirc \Rightarrow \bigcirc$.

ACKNOWLEDGMENTS

It is a pleasure to thank George Wilson for valuable comments and useful suggestions on the subject of this paper. Conversations with Professors L. Abellanas and A. Galindo are also acknowledged.

[1]V. G. Drinfel'd and V. V. Sokolov, Sov. Math. Dokl. **23**, 457 (1981).
[2]G. Wilson, Ergodic Theory Dynamical Systems **1**, 361 (1981).
[3]S. Helgason, *Differential Geometry, Lie Groups and Symmetric Spaces* (Academic, New York, 1978).
[4]G. Wilson, Phys. Lett. A **89**, 332 (1982).
[5]F. Calogero and A. Degasperis, J. Math. Phys. **22**, 23 (1981).
[6]B. A. Kupershmidt and G. Wilson, Invent. Math. **62**, 403 (1981).
[7]B. A. Kupershmidt and G. Wilson, Commun. Math. Phys. **81**, 189 (1981).
[8]V. E. Zakharov and A. B. Shabat, Functional Anal. Appl. **13**, 166 (1979).
[9]A. V. Mikhailov, Physica D **3**, 73 (1981).
[10]A. V. Mikhailov, M. A. Olshanetsky, and A. M. Perelomov, Comm. Math. Phys. **79**, 473 (1981).
[11]G. Wilson, Math. Proc. Cambridge Math. Soc. **86**, 131 (1979).
[12]M. Adler, Commun. Math. Phys. **80**, 517 (1981).
[13]B. A. Kupershmidt, J. Math. Phys. **23**, 364 (1982).
[14]F. Guil Guerrero, J. Math. Phys. **23**, 211 (1982).

Supermanifolds and automorphisms of super Lie groups[a)]

Roberto Cianci
Istituto Matematico dell'Università, Genova, Italy

(Received 4 January 1983; accepted for publication 18 October 1983)

The geometrical theory of supermanifolds is developed and applied to the super Lie groups. The resulting structural equations and the adjoint representation are studied and used to find the automorphisms of super Lie groups. The $N = 1$ supergravity symmetry group (the so-called graded Poincaré group) is explicitly studied and possible applications are outlined.

PACS numbers: 02.40.Sf, 11.30.Pb, 02.20. + b

I. INTRODUCTION

In the last years various papers[1–5] have given contributions to the study of the geometrical theory of supermanifolds and, in particular, to the program of giving a clear, formal meaning to supersymmetric physical theories such as supergravity.[6,7] Particularly in Ref. 4 it has been shown how the idea of interpreting the anticommuting coordinates as elements of a Grassmann algebra could be extended by considering coordinates with values in a generic Banach–Grassmann algebra thus obtaining a non-necessarily denumerable set of odd generators. A satisfactory aspect of this approach is the fact that it avoids the restrictions on the nilpotence of the elements of the algebra.

In Sec. II, by using the resulting different treatment of the tangent space $T(M)$ to a supermanifold M, we develop the differential aspect of the geometry of M; subsequently in Sec. III, after having generalized some classical concepts and used the definition of a supergroup G, we show how the ordinary Lie algebra structure of $T_e(G)$ (e is the identity) induces, on a preferred set of nontangent vectors "basis" of $T_e(G)$, the commutation relations of a graded Lie algebra.

The problem of the determination of the automorphisms of a super Lie group is hence analyzed (Sec. IV) and solved by using a general definition of automorphism of a graded Lie algebra: the problem is reduced to find the even derivations of $T_e(G)$. The resulting equations (Sec. V) are applied to the symmetry group of the $N = 1$ supergravity, the so-called graded Poincaré group GP, and all the automorphisms are found.

Finally relations with the theory of fiber bundles built up with supergroups are suggested.

II. GENERAL GEOMETRICAL STRUCTURE

In this paper we use the notation introduced in Ref. 4. In brief, summing up, we introduce a Z_2-graded commutative Banach-Grassmann algebra $Q = Q_0 \oplus Q_1$ satisfying

$$\text{(i)}\quad a_i\, a_j = (-1)^{ij} a_j a_i \in Q_{i+j}\,,\quad a_i \in Q_i\,,\quad a_j \in Q_j\,,$$

$$\text{(ii)}\quad \|a_0 + a_1\| = \|a_0\| + \|a_1\|\,, \tag{2.1}$$

and we define the Q-module Q^{m+n} as the set of the $(m+n)$-tuples with values in Q,

$Q^{m+n} = \{a^A \,|\, A = 1,...,m+n;\ a^A \in Q\}$, for which the following conditions hold:

$$\text{(i)}\quad Q^{m+n} = (Q^{m+n})_0 \oplus (Q^{m+n})_1\,,$$

where

$$\begin{aligned}(Q^{m+n})_0 &\equiv Q^{m,n}\\ &= \{(a^i, a^\alpha), i = 1,...,m;\ \alpha = 1,...,n;\ a^i \in Q_0;\ a^\alpha \in Q_1,\},\\ (Q^{m+n})_1 &\equiv Q^{\bar m,\bar n}\\ &= \{(a^i, a^\alpha), i = 1,...,m;\ \alpha = 1,...,n;\ a^i \in Q_1:\ a^\alpha \in Q_0\};\end{aligned} \tag{2.2}$$

$$\text{(ii)}\quad a_r\, v_s = (-1)^{rs} v_s\, a_r\,;\quad a_i \in Q_i\,,\quad v_s \in (Q^{m+n})_s\,.$$

Now, we can define a supermanifold M as a Banach manifold with a supersmooth atlas of coordinate maps with values in $Q^{m,n}$.[4] In other words, M can be covered by open sets where convenient homeomorphisms (the coordinate maps) with open sets of $Q^{m,n}$ are defined; this definition is similar to the one used for ordinary manifolds. Finally, having defined a path γ on M as a continuous map $\gamma:[-\alpha,\alpha]\to M$, $\alpha\in\mathbb{R}$, introduce a local frame of coordinates ϕ and an equivalence relation R:$\gamma_1 \approx \gamma_2$ in $P \in M$ if $\exists\ \alpha', \alpha'' \in \mathbb{R}$ so that

$$\text{(i)}\ \gamma_1(\alpha') = \gamma_2(\alpha'') = P,$$

$$\text{(ii)}\ \frac{d}{d\alpha}\phi\gamma_1(\alpha)|_{\alpha=\alpha'} = \frac{d}{d\alpha}\phi\gamma_2(\alpha)|_{\alpha=\alpha''}\,.$$

Now, we can speak of tangent vectors in P as equivalence classes of paths in P.

Setting $T_P(M) = \gamma/\approx$ in P we have that, for every $V \in T_P(M)$ there exists $v \in Q^{m,n}$ so that we can write:

$$V = v^i \frac{\partial}{\partial x^i}\Big)_P + v^\alpha \frac{\partial}{\partial x^\alpha}\Big)_P$$

if $\phi(P) = (x^i, x^\alpha)$. This notation means that V belongs to the equivalence class of paths γ that, in P, satisfies: $(v^i, v^\alpha) = (d/d\alpha)\phi\gamma(\alpha)|_{\alpha=\bar\alpha}$ if $\gamma(\bar\alpha) = P$. Notice that no path can give $v^i = 0, v^\alpha = \delta^\alpha_\mu$ and, therefore, the set of tangent vectors is isomorphic only to the "even" derivations (i.e., not changing the grading of the functions).

In the remaining part of this section we shall give some results and some formulas that will be useful in the following. This matter is a natural development of the ideas of Ref. 4; however, since they are not available in the literature we shall spend some words on them. Let us give the

[a)] This work was carried out under the auspices of the GNFM of CNR and under the research program "Geometria e Fisica" of the "Ministero della Pubblica Istruzione."

 0022-2488/84/030451-05$02.50

Definition 2.1: If $\psi:M\to M$ is a continuous map with $v=\psi(u)$; the path $\tilde{\gamma}(t)=\psi(\gamma(t))$ is called the push-forward of γ by ψ.

By noticing that ψ preserves the relation R, one can define a map $\psi_*:T_u(M)\to T_v(M)$ as $\psi_*(X^A\,\partial/\partial x^A)_u = X^A(\partial/\partial x^A)\psi^B(u)\partial/\partial x^B)_v$, where $\psi^B=\phi^B(\psi)$ and $\partial\phi^B/\partial x^A)_u\in \mathrm{GL}(m,n)\forall\ u$.[8] Now let us introduce the dual space $T^*_P(M)$; $\omega\in T^*_P(M)$ can be written as $\omega=dx^i\omega_i + dx^\alpha\omega_\alpha$, $\omega_a\in Q$; for any $v\in T_P(M)$ we have the action

$$v(\omega)=\left(v^A\frac{\partial}{\partial x^A}\right)(dx^B\omega_B) = v^A\omega_A\,. \tag{2.3}$$

It is worth noticing that, following Ref. 9 we adopt the convention of making the vectors acting on the right (notice that $v^A\omega_A\neq\omega_A v^A$). In this way, we shall say that a form ω is "even" if $\omega_A\in Q^{m,n}$, thus inducing the grading $T^*_P(M)=(T^*_P(M))_0\oplus(T^*_P(M))_1$. We can now construct the Lie derivative with respect to tangent vector fields on the basis of the existence of a one-parameter group of diffeomorphisms ψ_t of M.[10] Consider the field $v\in T(M)$ tangent to ψ_t and define

$$\mathscr{L}_v u=\lim_{t\to 0}\frac{-\psi_{t_*}(u)+u}{t},\quad u\in\tilde{T}(M)\,.^{11} \tag{2.4}$$

This equation can be elaborated paying attention to the graded commutation relation rules in Q to obtain

$$\mathscr{L}_u v=\left(u^A\frac{\partial}{\partial x^A}v^B - v^A\frac{\partial}{\partial x^A}u^B\right)\frac{\partial}{\partial x^B}\bigg)_x$$
$$=uv-vu=[u,v]_-\,. \tag{2.5}$$

This result is obvious enough since, according to the grading of Ref. 4, v is an "even" field. The Lie derivative can be extended to functions $f\in F$ (Ref. 11) and, consequently to differential forms obtaining

$$\mathscr{L}_v(f)=v(f)\,,$$
$$\mathscr{L}_v\,\omega=dx^B\left[\left(\frac{\partial}{\partial x^B}v^A\right)\omega_A + v^A\frac{\partial}{\partial x^A}\omega_B\right], \tag{2.6}$$
$$\mathscr{L}_v\,u(\omega)=(\mathscr{L}_{vu})(\omega)+u(\mathscr{L}_v\,\omega)\,, \tag{2.7}$$
$$\mathscr{L}_v\omega=v\lrcorner d\omega+d(v\lrcorner\omega)\,, \tag{2.8}$$

III. THE GEOMETRY OF SUPER LIE GROUPS

The theory we have summed up and completed in Sec. II will be applied to the super Lie groups (SL-groups) thus extending some known results.[2] Briefly, an SL-group is an S-manifold G with a S-smooth composition law $\Lambda:G\times G\to G$ which makes G into an abstract group. It can be shown that the map Λ can be assumed S-analytic without any restriction,[12] since one can always give to G an SA-structure in which Λ is S-analytic. Like any tangent space to an S-manifold, the space

$$T_e(G)\equiv\left\{V=v^i\frac{\partial}{\partial x^i}\bigg)_e + v^\alpha\frac{\partial}{\partial\xi^\alpha}\bigg)_e,\ v^A\in Q^{m,n}\right\}\ \text{(Ref. 13)}$$

does not admit a module basis with respect to any algebra A; anyway, following Ref. 9, we can choose a set of nontangent fields $\frac{\partial}{\partial x^i})_e$, $\frac{\partial}{\partial\xi^a})_e\in T_e\dagger\dagger(G)$ (Ref. 14) whose graded linear span (3.1) generates exactly $T_e(G)$ and we can define the fields

$$D_A(a)=L_{a*}\frac{\partial}{\partial x^A}\bigg)_e\quad\forall\ a\in G\,,$$

L_{a*} denotes the push-forward given by the left transport map L_a.

Theorem 3.1: The linear span $\{V=v^A D_A\}$, with constant coefficients $v^a\in Q^{m,n}$, generates the class $\hat{g}$ of the left-invariant tangent vector fields on G; moreover $\hat{g}$ is the even sector of the super Lie module $W\ [\equiv T_e^{\dagger\dagger}(g)]$ of G.

The proof can be achieved according to the ideas of Theorem 3.4 of Ref. 2(b), where the definition of W (there called graded Lie module) is given; even if we have used a generic BL-algebra Q and so enlarged the category of SL groups, no problems arise since only the definition and the obvious properties of the map L_a are involved. It is worth noticing that, by using tangent fields $X, Y\in\hat{g}$, only the "ordinary" bracket $\mathscr{L}_X\,Y=[X,Y]=XY-YX$ and the "ordinary" Jacobi identities are involved. After introducing the fields D_A, one can ask if they are related to some graded Lie algebra g; in general the answer is negative.

Consider now $u, v\in\hat{g}$, set $u=u^A D_A$, $v=v^A D_A$ and calculate

$$\begin{aligned}\mathscr{L}_v\,u&=[v,u]\\&=v^A D_A u^B D_B - u^B D_B v^A D_A\\&=u^B v^A(D_A D_B - (-1)^{AB}D_B D_A)\\&=u^B v^A[D_A,D_B]_\pm\,.\end{aligned} \tag{3.1}$$

By setting $[D_A D_B]_\pm = C^C{}_{AB}D_C$, with $C^C{}_{AB}=-(-1)^{AB}C^C{}_{BA}$, it is easy to show that, the graded bracket being L_{a*}-invariant, the $C^C{}_{AB}$ are Q-valued constants satisfying grading of $C^C{}_{AB}=A+B+C$.

On the other hand, we have that

$$\mathscr{L}_v\,u=u^B v^A C^C{}_{AB}D_C\in\hat{g}\subset T_e(G)\,, \tag{3.2}$$

and therefore, whenever we can choose a basis $\{D_A\}$ of W on which the $C^C{}_{AB}$ are real, Eq. (3.2) gives

$$C^i_{\alpha_j}=C^\alpha_{\mu\nu}=C^\alpha_{ij}=0\,. \tag{3.3}$$

Theorem 3.2: If W admits a basis D_A for which the $C^C{}_{AB}$ are real, the $\mathbb{R}$-span of the D_A is a graded Lie algebra g.

Proof: Define $g_0=\{v^i D_i\}$, $g_1=\{v^\alpha D_\alpha\}$ with v^i, $v^\alpha\in\mathbb{R}$, and use Eqs. (3.2) to show that $[g_0,g_0]\in g_0$, $[g_0,g_1]\in g_1$, $[g_1,g_1]\in g_0$. The graded Jacobi identities for the $C^A{}_{BC}$ can be obtained by the component expression of the "ordinary" Jacobi identity for three fields $u, v, w\in\hat{g}$. We wish to stress that the correspondence between tangent space to the identity to SL-groups and g.L.a.'s is not one-to-one; so to say, there are "more" $T_e(g)$ than g.L.a.'s. Using the formalism of Ref. 2, we have that, under the same assumption of Theorem 3.2, W is decomposable: $W=Q\otimes g$.

The problem to investigate SL-groups with nondecomposable SL-modules is not easy, since decomposability is difficult to test in general. Actually, if, on one hand, it is easy to build simple nondecomposable modules, on the other hand, it is difficult to say if the cause of Q-valued $C^A{}_{BC}$ is only the use of a nonconvenient basis $\{D_A\}$ of W. To clarify this point, consider the GP group (see Sec. V for details on GP); the GL (4,4) transformation

$P_i \to P'_i = P_i + q_i^\alpha Q_\alpha$; $Q_\alpha \to Q'_\alpha = Q_\alpha$ gives a new basis of $W_{\rm GP}$ for which $C'^\alpha{}_{ij} = -C'^\alpha{}_{ji} \neq 0 \in Q_1$.

Now let us return to the general case and use the formalism of Ref. 9; namely set $z = A(x,y)$ and put

$$A^*dz^A = dx^B V_B{}^A(x,y) + dy^B W_B{}^A(x,y). \tag{3.4}$$

The "auxiliary" GL(m,n)-matrices V and W satisfy some useful properties often used in the following: see Eqs. (2.10) of Ref. 9. This formalism, although not strictly necessary, is very useful since it allows us to write

$$D_A(x) = W_A^B(x,e)\frac{\partial}{\partial x^B}\Big)_x, \quad \omega^B(x) = dx^A W_A^B(x^{-1},x), \tag{3.5}$$

where $\omega^B(x)$ denotes the left-invariant forms which are dual to the fields D_A:$D_A(\omega^B) = \delta_A^B$.

Theorem 3.3: The following equations hold:

$$[D_A, D_B](\omega^c) = D_A \lrcorner D_B \lrcorner d\omega^c,$$
$$d\omega^A = \tfrac{1}{2}\omega^B \wedge \omega^C C^A_{CB} \tag{3.6}$$

(Maurer–Cartan structural equations).

Outline of the proof: It is sufficient to carry on the calculation of $d\omega^A$ by using Eq. (2.7) Eqs. (2.10) of Ref. 9, and the natural extension of the □ operator given by

$$v \lrcorner \Omega = v^A \frac{\partial}{\partial x^A} \lrcorner (dx^Q \wedge dx^P \Omega_{QP})$$
$$= v^A dx^P \Omega_{AP} - (-1)^{AQ} v^A dx^Q \Omega_{QA}$$

for each $v \in T(G)$ and $\Omega \in T^{*2}(G)$. Since $d\,d\omega^A = 0$ Eq. (3.6) gives

$$\omega^A \wedge \omega^R \wedge \omega^Q C^B_{QR} C^C_{BA} = 0,$$

that is,

$$(\mathrm{gr}\,A)_{QRA} C^B_{QR} C^C_{BA} = 0,$$

where $(\mathrm{gr}\,A)_{QRA}$ denotes the graded antisymmetrization in QRA; this explicitly yields

$$(-1)^{C(A+B)} C^S_{BC} C^R_{SA} + (-1)^{B(C+A)} C^S_{AB} C^R_{SC}$$
$$+ (-1)^{A(B+C)} C^S_{CA} C^R_{SB} = 0. \tag{3.7}$$

Whenever C^S_{BC} are real we have the usual graded Jacobi identities.

Theorem 3.4: The fields $X \in \hat{g}$ generate a group of local, superregular diffeomorphisms of G: for any $X \in \hat{g}$ a group Γ_t:$G \to G$ exists so that

(i) $\Gamma_t \Gamma_s = \Gamma_{t+s}$,

(ii) $\Gamma_0 = e$,

(iii) $\dot{\Gamma}(0) = X$ for any $t \in (0,\epsilon)$ with $\epsilon \in \mathbb{R}$.

This result can be extended by using the formula $\Gamma_t = \Gamma_{n\epsilon}\Gamma_{t-n\epsilon}$ for $n\epsilon \le t \le (n+1)\epsilon$, $n > 0$. Leaving to the reader to verify that $\dot{\Gamma}_t = L_{\Gamma_t^*} X$ we can give

Definition 3.2: $\Gamma_t = \exp tX$ is called the exponential map of X.

On the other hand, calling ad_a:$G \to G$ $(a \in G)$ the map $\mathrm{ad}_a x = axa^{-1}$, we obtain ad_{a*}:$T(G) \to T(G)$ explicitly given by $\mathrm{ad}_{a*} D_A(x) = \mathrm{ad}(a)_A^B D_B(y)$ where $y = \mathrm{ad}_a x$ and $\mathrm{ad}(a)_A^B$ is a *constant* GL(m,n) matrix. This enables us to define ad_{a*}:$\hat{g} \to \hat{g}$ as

$$\mathrm{ad}_{a*} X^A D_A(x) = X^A \mathrm{ad}(a)_A^B D_B(y). \tag{3.8}$$

Theorem 3.5: The matrix $\mathrm{ad}(a)_A^B$ (the adjoint representation of G on $\hat{g}$) satisfies some useful properties which are listed here:

$$\mathrm{ad}(ab)_A^B = \mathrm{ad}(b)_A^C\,\mathrm{ad}(a)_C^B, \tag{3.9}$$

$$C^C_{FE}\,\mathrm{ad}(a^{-1})_C^A = \mathrm{ad}(a^{-1})_F^C\,\mathrm{ad}(a^{-1})_E^B C^A_{CB}(-1)^{C(E+B)}, \tag{3.10}$$

$$d\,\mathrm{ad}(a^{-1})_C^A = (-1)^{BC}\omega^B(a)\ \mathrm{ad}(a^{-1})_C^F C^A_{FB}. \tag{3.11}$$

Outline of the proof: Equation (3.9) is trivial; Eqs. (3.10) and (3.11) can be obtained calculating $A^*\omega^A$ by means of Eq. (3.6),

$$A^*d\omega^A = \tfrac{1}{2}A^*\omega^B \wedge A^*\omega^C C^A_{CB},$$

and by comparing this result with the equation

$$A^*\omega^A(z) = \omega^A(y) + \omega^C(x)\mathrm{ad}(y^{-1})_C^A, \quad z = A(x,y)$$

obtained directly by Eq. (3.4).

On the other hand, whenever $a = \exp tA, A \in \hat{g}$ the map $(d/dt)\mathrm{ad}_{a(t)*} X|_{t=0} = Y$ gives

$$Y = [A,X]. \tag{3.12}$$

It provides the adjoint representation of $\hat{g}$ on $\hat{g}$, also denoted

$$\mathrm{ad}_A X = [A,X]. \tag{3.13}$$

IV. AUTOMORPHISMS

The aim of this section is to determine the automorphism group Aut(G) of a given SL-group G whose module is decomposable. All results up to and including Eq. (4.9) are true also for nondecomposable modules. More precisely, $\psi \in$ Aut(G) is

(i) an automorphism of the topological group underlying G,

$$\psi: G \to G, \quad \psi(ab) = \psi(a)\psi(b) \quad \forall\, a,b \in G; \tag{4.1}$$

(ii) an SA-map with respect to the SA-structure which is used to make G an SL-group.

It is a remarkable result, analogous to the classical one, that the map ψ_* is a Q_0-linear automorphism of $\hat{g}$:

$$\psi_*[A,B] = [\psi_* A, \psi_* B] \quad \forall\ A,B \in \hat{g}. \tag{4.2}$$

The Q_0 linearity is a trivial consequence of the S-analyticity of ψ. Equation (4.2) can be obtained by a mere transposition of the classical proof,[15] or by performing the following calculation, which will be useful later on.

By differentiating Eq. (4.1) with respect to a and setting $a = e$ we obtain

$$V_C^B(e,x)\frac{\partial}{\partial x^B}\psi^A(x) = B_C^B V_B^A(e,y), \tag{4.3}$$

where $y = \psi(x)$ and $B_C^A = \dfrac{\partial}{\partial x^C}\psi^B(e) \in \mathrm{GL}(m,n)$. The analogous calculation for b gives

$$W_C^B(x,e)\frac{\partial}{\partial x^B}\psi^A(x) = B_C^B W_B^A(y,e). \tag{4.4}$$

By equating the term $\frac{\partial}{\partial x^B}\psi^A$ in both equations and using the expression $\mathrm{ad}(a^{-1})^B_A = V^C_A(e,a)W^B_C(a^{-1},a)$, we have

$$B^B_C = \mathrm{ad}(x^{-1})^E_C B^A_E \mathrm{ad}(y)^B_A , \quad (4.5)$$

from which we obtain

$$\psi_*(u^A D_A(x)) = u^B B^A_B D_A(y) . \quad (4.6)$$

Now, by differentiating Eq. (4.5) and using Eq. (3.11) we get

$$C^E_{QC} B^D_E = (-1)^{Q(A+C)} B^A_C B^B_Q C^D_{BA} \text{ (Ref. 16)}. \quad (4.7)$$

Equations (4.7) and (4.6) imply directly Eq. (4.2). Equation (4.7) implies that for any $S \in \mathrm{aut}(G)$ we have

$$C^E_{QC} S^D_E = S^B_Q C^D_{BC} - (-1)^{QC} S^A_C C^D_{AQ} . \quad (4.8)$$

(This can be obtained naively by putting $B^C_A = \delta^C_A + S^C_A$). Now, using Eq.(4.8) we can obtain

Theorem 4.1: The operator $S(D_A) = S^C_A D_C$ is a derivation of W, and $S^C_A \in \mathrm{gl}(m,n)$.

Proof: Calculating the expressions

$$S[D_Q, D_C] = S(C^E_{QC} D_E) = C^E_{QC} S^D_E D_D ,$$

$$[SD_Q, D_C] = S^B_Q [D_B, D_C] = S^B_Q C^D_{BC} D_D ,$$

$$[D_Q, SD_C] = [D_Q, S^A_C D_A] = (-1)^{Q(A+C)} S^A_C C^D_{QA} D_D ,$$

and using Eqs. (3.3) and (3.4) we prove that S is a derivation; on the other hand, $S^C_A \in \mathrm{gl}(m,n)$ in view of its definition.

Notice that, as in the case of the automorphisms of the classical Lie groups, S is an "even" derivation only if we consider as even derivation those preserving the grading in Q^{m+n}: in other words, we have to use a grading of the derivations S different from that of g.L.a. theory,[17,18] where S is called even if it satisfies the more restrictive condition

$$S: g_0 \to g_0, \quad S: g_1 \to g_1 .$$

Theorem 4.1 can be written as

$$S[A,B] = [SA,B] + [A,SB] \quad \text{or} \quad [S,\mathrm{ad}_A] = \mathrm{ad}_{SA} , \quad (4.9)$$

where the last commutator has to be regarded in $\mathrm{gl}(m,n)$. The problem of determining $\mathrm{Aut}(G)$ is reduced to finding the graded subalgebra of $\mathrm{gl}(m,n)$ of the matrices S satisfying Eq. (4.9). By denoting with $D(\hat{g})$ the derivations of $\hat{g}$ and with $\mathrm{Inder}(\hat{g})$ only the interior ones [i.e., $X \in \mathrm{Inder}(\hat{g})$ if $X = \mathrm{ad}_A$ for some $A \in \hat{g}$], we can see from Eq. (4.9) that $\mathrm{Inder}(\hat{g})$ is an ideal in $D(\hat{g})$ and use the result[18,17]

$$D(\hat{g}) = \mathrm{Inder}(\hat{g}) \oplus W, \quad (4.10)$$

where W is a subalgebra of $D(\hat{g})$ satisfying

$$\mathrm{ad}_{W(L)} = 0 . \quad (4.11)$$

In this equation L denotes the semisimple part of the ordinary Lie algebra g_0, the latter defined as the $\mathbb{R}$-linear span of the D_i. L is well characterized by means of Levi's decomposition theorem in g_0[19] and is unique modulo the action of ad_R, R being the soluble radical of g_0; the Malcev–Harish-Chandra theorem[19] has been used for g_0. Setting $g_0 = L \oplus R$ and denoting with $u, v, \ldots$ indices in L and with $a, b, \ldots$ indices in R, Eq. (4.11) becomes

$$W^A_u C^B_{AC} = 0 . \quad (4.12)$$

On the other hand, we have the following

Lemma 4.1: $W(L) = 0$ (i.e., $W^A_u = 0$). (4.13)

Proof: Using Eq. (4.12) when $C = v$ and $B = w$ we obtain $W^t_u C^w_{tv} = 0$; the properties of the graded Lie algebra D_A and of R have been used; since L is semisimple, it admits a nondegenerate Killing form which, in turn, yields the metric

$$g_{uv} = C^t_{uw} C^w_{vt} , \quad g_{uv} = g_{vu} , \quad (4.14)$$

thus allowing one to get $W^t_u = 0$. Now writing Eq. (4.8) for the derivation $W^D_E \in W$ we have

$$C^E_{QC} W^D_E = C^D_{BC} W^B_Q + (-1)^{Q(A+C)} C^D_{QA} W^A_C , \quad (4.15)$$

which furnishes $W^a_t = 0$ by setting $D = a$, $Q = u$, $C = v$. Finally setting $D = \alpha$, $Q = u$, $C = v$ we get

$$W^\beta_u = 0 .$$

The foregoing discussion can be summed up in

Lemma 4.2: The matrices S satisfying Eq. 4.9 are given by

$$S = \mathrm{ad}_A + W$$

for any $A \in \hat{g}$ and for any $W \in \mathrm{gl}(m,n)$ which is a solution of Eqs. (4.13) and (4.15).

The whole problem has been reduced to an explicit study of these equations for the specific SL-group which is under consideration.

V. THE CASE OF THE GRADED POINCARÉ GROUP GP

We wish to find the automorphism group of GP.[9] By denoting the "generators" with J_{ij}, P_i, Q_α, the graded Lie algebra[20] is

$$\left.\begin{aligned}
&[J_{pq}, J_{ij}] = C^{rs}_{pq\,ij} J_{rs} ,\\
&[J_{ij}, P_k] = \tfrac{1}{2}(\eta_{ik} P_j - \eta_{jk} P_i) ,\\
&[J_{ij}, Q_\alpha] = \tfrac{1}{2}(\sigma_{ji})_{\mu\alpha} Q_\mu ,\\
&[Q_\alpha, Q_\mu] = (C^{-1}\gamma^i)_{\alpha\mu} P_i ,\\
&[Q_\alpha, P_i] = 0 ,
\end{aligned}\right\} \quad (5.1)$$

where $C^{rs}_{pq\,ij} = \frac{1}{2}(\eta_{qj}\delta^{[r}_p\delta^{s]}_i + \eta_{pi}\delta^{[r}_q\delta^{s]}_j - \eta_{pj}\delta^{[r}_q\delta^{s]}_i - \eta_{qi}\delta^{[r}_p\delta^{s]}_j)$ are the Lorentz structural constants, C is the charge-conjugation matrix, γ^i ($i = 0, 1, 2, 3$) are the 4×4 Dirac matrices and $\sigma_{ij} = \frac{1}{4}[\gamma_i, \gamma_j]_-$.

In order to find the SL-group Aut(GP) we apply Lemma 4.2 and therefore, after having introduced a matrix $W \in \mathrm{gl}(10,4)$ satisfying Eq. (4.13), we write Eqs. (4.15) for our actual case; they are

$$C^\beta_{ij\,\alpha} W^{ks}_\beta = C^{ks}_{ij\,pt} W^{pt}_\alpha , \quad (5.2a)$$

$$C^a_{\beta\alpha} W^\beta_B + C^a_{b\,ij} W^{ij}_\alpha = 0 , \quad (5.2b)$$

$$C^b_{ij\,a} W^\alpha_b = C^\alpha_{ij\,\beta} W^\beta_a , \quad (5.2c)$$

$$C^a_{\mu\beta} W^\alpha_a = C^\alpha_{ij\,\beta} W^{ij}_\mu + C^\alpha_{ij\,\mu} W^{ij}_\beta , \quad (5.2d)$$

$$C^b_{ij\,d} W^a_b = C^a_{ij\,b} W^b_d , \quad (5.2e)$$

$$C^d_{\beta\alpha} W^a_d = C^a_{\gamma\alpha} W^\gamma_\beta + C^a_{\beta\gamma} W^\gamma_a , \quad (5.2f)$$

$$C^\gamma_{ij\,\beta} W^\alpha_\gamma = C^\alpha_{ij\,\gamma} W^\gamma_\beta , \quad (5.2g)$$

$$C^\beta_{ij\,\alpha} W^a_\beta = C^a_{ij\,b} W^b_a . \quad (5.2h)$$

To solve Eqs. (5.2a)–(5.2d) the use of a straightforward calculation gives $W^{ij}_{\alpha} = W^{\alpha}_{a} = 0$. On the other hand, Eq. (5.2h) yields $W^{a}_{\alpha} = 0$; finally Eqs. (5.2e)–(5.2g) gives

$$W^{b}_{a} = \lambda\delta^{b}_{a}, \quad W^{\beta}_{\alpha} = \tfrac{1}{2}\lambda\delta_{\alpha\beta} + \mu(\gamma_5)_{\alpha\beta}, \quad \lambda,\mu \in \mathbb{R}. \tag{5.3}$$

By recalling that $C^{i}_{\alpha\beta} = C^{i}_{\beta\alpha}$ one obtains the most general expression for $S \in \mathrm{gl}(10,4)$:

$$S^{B}_{C} = A^{A}C^{B}_{AC} + \delta^{B}_{i}\delta^{i}_{C}\lambda + \tfrac{1}{2}\delta^{B}_{\alpha}\delta^{\alpha}_{C}\lambda + \mu\delta^{B}_{\alpha}\delta^{\nu}_{C}(\gamma_5)_{\alpha\nu}, \quad A \in \hat{g}. \tag{5.4}$$

The matrix B can be obtained by exponentiation, $B = \exp S$, and finally we can obtain ψ by integration. This calculation can be performed easily using normal coordinates in G: if $z = \exp(z^A D_A(e))$ set $\phi^A(z) = z^A$. The exponential map assures us that an isomorphism between a neighborhood of $e \in G$ and a neighborhood of $a^A = 0$ in $Q^{m,n}$ exists. We have

$$\psi(z) = \psi(\exp(z^A D_A(e))) = \exp(z^A \psi_* D_A(e)) = \exp(z^a B^{C}_{A} D_C(e)) \tag{5.5}$$

from which

$$\phi^A(\psi(z)) = \phi^B(z) B^{A}_{B} = \phi^B(z)(\exp S)^{A}_{B} \tag{5.6}$$

follows. In the case of an infinitesimal automorphism ψ (where we can set $B^{C}_{A} = \delta^{C}_{A} + S^{C}_{A}$) we give explicitly the formula: if $\phi(z) = (L^{j}_{i}, v^i, \xi^{\alpha})$ then

$$\phi(\psi(z)) = (L^{j}_{i}, (1+\lambda)v^i, (1+\lambda)\xi^{\alpha} + \mu\xi^{\nu}(\gamma_5)_{\nu\alpha}).$$

Normal coordinates have been used; only the exterior derivations have been taken into account: in general, a term ad_A has to be added.

Eventually, we wish to notice that the study of the automorphisms of GP is particularly useful in the description of supergravity when one uses a principal fiber bundle whose structural group is GP.[21,22] In this framework, the automorphisms of GP can be linked with the fiber-bundle ones[23] and therefore, looking at the latter as gauge transformations, one can describe the symmetry of the theory.

[1] B. Kostant, "Graded manifolds, graded Lie theory and prequantization" in *Differential Geometrical Methods in Mathematical Physics Lecture Notes in Mathematics*, Vol. 570 (Springer, New York, 1977).
[2] A. Rogers, J. Math. Phys. **21**, 1352–1365 (1980) and **22**, 939–945 (1981).
[3] J. Hoyos, M. Quiros, J. R. Mittelbrunn, and F. J. de Urries "Generalized Supermanifold I," preprint, Instituto de Estructura de la Materia, Madrid, Spain, 1981.
[4] A. Jadczyk and K. Pilch, Commun. Math. Phys. **78**, 373–390 (1981).
[5] M. Batchelor, Trans. Am. Math. Soc. **253**, 329–338 (1979) and **258**, 257–270 (1981).
[6] J. Wess and B. Zumino, Nucl. Phys. B **70**, 39–50 (1974); A. Salam and J. Strathdee, Nucl. Phys. B **76**, 477–482 (1974).
[7] D. Z. Friedmann, P. van Niewenhuizen, and S. Ferrara, Phys. Rev. D **13**, 3214–3220 (1976).
[8] V. Rittenberg and M. Scheunert, J. Math. Phys. **19**, 709–713 (1978).
[9] Y. Neeman and T. Regge, Riv. Nuovo Cimento **1**, 1–43 (1978).
[10] S. Sternberg, *Lectures in Differential Geometry*, (Prentice Hall, Englewood Cliffs, NJ, 1964).
[11] For the definition of $\bar{T}(M)$ and F see Ref. 4.
[12] U. Bruzzo, and R. Cianci, "An existence result on super Lie groups" (in press).
[13] Henceforth we shall denote $\phi(P) = (x^i, \xi^{\alpha})$, $P \in G$. Moreover, with x^A $(A = 1,\ldots,m+n)$ we mean (x^i, ξ^{α}).
[14] See Ref. 4 for the definition of $T^{++}_{e}(G)$.
[15] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[16] Using Eq. (4.7) one can easily verify that the "Killing form" $B(X,Y) = \mathrm{S\text{-}tr}(\mathrm{ad}_X \mathrm{ad}_Y)$ for $X, Y \in \hat{g}$ gives rise to a graded-symmetric ψ^*-invariant two-form; setting $B(X,Y) = Y^B X^A g_{AB}$ we have: (i) $g_{AB} = (-1)^{AB} g_{BA}$, (ii) $\psi^* g = g$, (iii) $g_{AB} = C^{S}_{BL} C^{L}_{AS}(-1)^{L}$.
[17] V. G. Kaz, Adv. Math. **26**, 8–96 (1977).
[18] M. Scheunert, *The Theory of Lie Superalgebra*, Lecture Notes in Mathematics, Vol. 716 (Springer, New York, 1979).
[19] N. Jacobson, *Lie Algebras* (Wiley-Interscience, New York, 1962).
[20] In this section we use a couple of antisymmetrized Latin indices instead of u,v.
[21] R. G. Yates, Commun. Math. Phys. **76**, 255–268 (1980).
[22] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley-Interscience, New York, 1963), Vol. I.
[23] U. Bruzzo and R. Cianci, Lett. Math. Phys. **6**, 471–477 (1982).

Finite quantum processes

Stanley P. Gudder
Department of Mathematics and Computer Science, University of Denver, Denver, Colorado 80208

(Received 25 January 1983; accepted for publication 17 March 1983)

A model for quantum dynamics is presented in terms of a generalization of Markov chains. We first consider measurements and events on an amplitude space. Markov amplitude chains on an amplitude space are then studied. Quantum chains are defined and characterized by a Markov and weak stationarity property. We then consider random phase transformations and study the changes that result in an amplitude space and in the quantum chains due to such transformations. A perturbation expansion involving a potential perturbing the "free" motion is proved. Quantum processes are defined and their relation to quantum chains is discussed. The existence of an arbitrary quantum chain is proved using path chains.

PACS numbers: 02.50.Cw, 05.40. + j, 02.50.Ga, 03.65. – w

1. INTRODUCTION

Over thirty years ago, R. Feynman observed a basic difference between classical probability theory and quantum probability theory.[1–3] In classical probability theory one has a probability space (Ω, Σ, P), where Ω is the set of outcomes for a stochastic system, Σ is a σ-algebra of events, and P is a probability measure on Σ. For an event $E \in \Sigma$, $P(E)$ gives the probability that E occurs. In quantum probability theory, the situation is quite different. One still has a triple $(\Omega_q, \Sigma_q, P_q)$ consisting of an outcome space, an event set, and a probability function. However, in general, Σ_q is not a σ-algebra, and P_q is not a measure. As we shall discuss in detail later, Σ_q is a σ-algebra and P_q is a measure only when attention is restricted to a single measurement. When multiple measurements are considered, interference effects remove us from the realm of classical probability theory. The interference effects occur because, unlike P which can be quite arbitrary, the quantum probability function P_q is obtained in a specific way. It is basic to quantum mechanics that we begin with an amplitude function $A: \Omega_q \to C$. If $E \in \Sigma_q$, we define the amplitude of E as $\hat{A}(E) = \Sigma_{\omega \in E} A(\omega)$ and we define the probability that E occurs by

$$P_q(E) = |\hat{A}(E)|^2 = \left| \sum_{\omega \in E} A(\omega) \right|^2. \tag{1}$$

In this way, a quantum probability function is induced from an amplitude function. If $E, F \in \Sigma_q$ and $E \cap F = \varnothing$, we need not have that $P_q(E \cup F) = P_q(E) + P_q(F)$. Also, as we shall later show, we need not have that $E \cup F \in \Sigma_q$. It is because of such differences between classical and quantum probability theory, that we are confronted with certain so-called "paradoxes" of quantum mechanics.

One of these "paradoxes" involves the double-slit experiment. This experiment is discussed in detail in Ref. 2 so we shall only give what we feel is the essence of the argument. A source emits a beam of photons which impinge upon a screen with two slits which we designate as slits 1 and 2. After going through one of the slits, photons then strike a target screen. We seek to find the probability that a photon strikes the target screen within an area Δ on the screen. Denote the probability that a photon passes through slit 1 by $P(1)$ and the probability that a photon strikes within Δ given that it passes through screen 1 by $P(\Delta\,|1)$, and use similar notation for screen 2. Classical probability theory would then give

$$P(\Delta) = P(1)P(\Delta\,|1) + P(2)P(\Delta\,|2). \tag{2}$$

Now $P(\Delta\,|1)$ is the distribution obtained if slit 2 is closed and similarly for $P(\Delta\,|2)$. Thus Eq. (2) says that $P(\Delta)$ is a mixture of the two distributions obtained when slit 1 or slit 2 is closed. This does not agree with reality. In the corresponding quantum mechanical calculation, the probability P_q is induced by an amplitude function A. We use the notation $\hat{A}(\Delta)$, $\hat{A}(1)$, $\hat{A}(2)$, $\hat{A}(\Delta\,|1)$, $\hat{A}(\Delta\,|2)$ for the amplitudes of the various events, where $\hat{A}(\Delta\,|1) = \hat{A}(\Delta \cap 1)/\hat{A}(1)$ and $\hat{A}(\Delta\,|2) = \hat{A}(\Delta \cap 2)/\hat{A}(2)$. Assuming that the photons do not interact between the source and the first screen, we have $|\hat{A}(1)|^2 = P_q(1) = P(1)$ and $|\hat{A}(2)|^2 = P_q(2) = P(2)$. If slit 2 is closed, the photons must pass through slit 1 so there is no interference between the two slit measurements. We conclude that $|\hat{A}(\Delta\,|1)|^2 = P(\Delta\,|1)$, and similarly $|\hat{A}(\Delta\,|2)|^2 = P(\Delta\,|2)$. Notice that although P_q need not be additive, by its very definition in the previous paragraph A is additive. Hence, applying Eqs. (1) and (2), the quantum probability becomes

$$\begin{aligned} P_q(\Delta) &= |\hat{A}(\Delta)|^2 = |\hat{A}(1)\hat{A}(\Delta\,|1) + \hat{A}(2)\hat{A}(\Delta\,|2)|^2 \\ &= P(\Delta) + 2\,\mathrm{Re}\,\hat{A}(1)\hat{A}(2)\hat{A}(\Delta\,|1)\hat{A}(\Delta\,|2). \end{aligned} \tag{3}$$

Equation (3) shows that $P_q(\Delta)$ is $P(\Delta)$ plus an interference which can be positive or negative, so this latter term can cause reinforcement or cancellation. The quantum probability calculation does agree with experiment.

2. QUANTUM DYNAMICS

The main purpose of this paper is to present a model for quantum dynamics in terms of a generalization of Markov chains. In order to motivate this model, we first consider the usual formulation of quantum dynamics. Let K be a separable, complex Hilbert space, and let $e_j, j = 1,2, \ldots$ be an orthonormal basis for K. Suppose $\phi_0 \in K$, $\|\phi_0\| = 1$ represents the initial state of a quantum system and H is a self-adjoint operator on K representing the Hamiltonian of the system. If we establish a standard unit of time, then the probability amplitude that the system which is initially in the state ϕ_0 is

 0022-2488/84/030456-10$02.50

in the state e_j after n time steps is given by

$$\langle U^n\phi_0, e_j\rangle = \langle e^{-inH}\phi_0, e_j\rangle. \tag{4}$$

The probability that the above event occurs is given by the square of the modulus of the expression in (4). Expanding Eq. (4) in terms of U gives

$$\langle U^n \phi_0, e_j\rangle = \sum_{i_0,\dots,i_{n-1}} \langle \phi_0, e_{i_0}\rangle \langle Ue_{i_0}, e_{i_1}\rangle \times \langle Ue_{i_1}, e_{i_2}\rangle \cdots \langle Ue_{i_{n-1}}, e_j\rangle. \tag{5}$$

We now interpret Eq. (5) physically. Let $\omega(i_0, i_1,\dots,i_n)$ represent the outcome that a quantum system, which is in the initial state ϕ_0, begins in the state e_{i_0} at time zero, moves to state e_{i_1} in one time step, then to state e_{i_2} in one more time step, and continues in this way until it arrives at state e_{i_n} after a total of n time steps. We interpret $\langle\phi, e_{i_0}\rangle \langle Ue_{i_0}, e_{i_1}\rangle \cdots\langle Ue_{i_{n-1}},e_j\rangle$ as the probability amplitude $A\ [\omega(i_0,\dots,i_n)]$ of the outcome $\omega(i_0,\dots,i_n)$. Let $E(j)$ be the event that a quantum system in the initial state ϕ_0 moves through the various states e_i and arrives at the state e_j after n time steps. The probability amplitude of $E(j)$ is given by $\hat{A}\,[E(j)] = \langle U^n\phi_0, e_j\rangle$. We assume that the event $E(j)$ consists of the outcomes $\omega(i_0,\dots,i_{n-1},j)$, $i_0,\dots,i_{n-1} = 1,2,\dots$. Then Eq. (5) gives

$$\hat{A}\,[E(j)] = \sum_{i_0,\dots,i_{n-1}} A\ [\omega(i_0,\dots,i_{n-1},j)]; \tag{6}$$

that is, the probability amplitude of $E(j)$ is the sum of the probability amplitudes of the outcomes it contains. If $\Omega = \{\omega(i_0,\dots,i_n)\colon i_0,\dots,i_n = 1,2,\dots\}$ we call (Ω, A) an amplitude space. For $\omega\in\Omega$ we interpret $|A(\omega)|^2$ as the probability that the outcome ω occurs. We also interpret $|\hat{A}(E_j)|^2$ as the probability that the event $E(j)$ occurs. This gives a consistent probabilistic interpretation since

$$\sum_{i_0,\dots,i_n} |A\ [\omega(i_0,\dots, i_n)]|^2 = \|\phi_0\|^2 = 1 \tag{7}$$

and

$$\sum_j |\hat{A}\ [E(j)]|^2 = \|U^n\phi_0\|^2 = \|\phi_0\|^2 = 1. \tag{8}$$

In the sequel we shall call a set of outcomes or events which satisfy Eq. (7) or (8), respectively, a measurement. An important role will be played by real-valued functions defined on subsets of Ω whose inverse images form a measurement. These will be called observable functions.

In this paper we shall develop the theory of Markov chains on amplitude spaces. The simplest types of Markov chains will describe systems like the ones considered above. However, as we shall see, we can consider much more general systems. For simplicity, we shall only consider chains with a finite number of values. With careful attention to convergence matters, our work can be extended to chains with a countable number of values. However, we wish to avoid such matters in this study.

3. AMPLITUDE SPACES

Let Ω be a nonempty set and let $A\colon\ \Omega\to C$. We call the points $\omega\in\Omega$ *outcomes*, the function A an *amplitude function*, and (Ω, A) an *amplitude space*. We say that (Ω, A) is a *point amplitude space* if $\Sigma_{\omega\in\Omega}|A(\omega)|^2 = 1$. We call a subset $E\subseteq\Omega$ *summable* if $\Sigma_{\omega\in E}|A(\omega)| < \infty$ and denote the collection of summable sets by Σ_0. We define the set function $\hat{A}\colon\ \Sigma_0\to C$ by $\hat{A}(E) = \Sigma_{\omega\in E} A(\omega)$, $E\in\Sigma_0$ $[\hat{A}(\phi) = 0]$, and call $\hat{A}(E)$ the *amplitude* of E.

Lemma 1: Σ_0 is a ring of sets with the property that $E\in\Sigma_0$ and $F\subseteq E$ implies $F\in\Sigma_0$. Also, $\hat{A}$ is a σ-finite complex measure on Σ_0.

Proof: The proof that Σ_0 is a ring with the above property is straightforward. To show that $\hat{A}$ is a complex measure on Σ_0, suppose $E_i\in\Sigma_0$, $E_i\cap E_j = \phi$, $i\neq j = 1,2,\dots$, and $\cup E_i = E\in\Sigma_0$. Denumerate the elements of E_i on which A is nonvanishing by $\omega_{ij}, j = 1,2,\dots$. Then since $\Sigma_{i,j}|A(\omega_{ij})| = \Sigma_{\omega\in E}|A(\omega)| < \infty$, it follows from Fubini's theorem that

$$\hat{A}(E) = \sum_{i,j} A(\omega_{ij}) = \sum_i\Big[\sum_j A(\omega_{ij})\Big] = \sum_i \hat{A}(E_i).$$

Hence, $\hat{A}$ is countably additive on Σ_0. To show $\hat{A}$ is σ-finite, let $E\in\Sigma_0$ and let ω_i be the elements of E on which A is nonvanishing. Let $F = E - \{\omega_i : i = 1,2,\cdots\}$. Then $E = \cup\{\omega_i\}\cup F$ and $\hat{A}(F) = 0$, $\hat{A}(\{\omega_i\})\in C$ so $\hat{A}$ is σ-finite.■

Let Σ be the σ-ring generated by Σ_0. If $\hat{A}$ is a complex signed measure, then it follows from the Hahn extension theorem that there exists a unique extension of $\hat{A}$ to a σ-finite complex measure on Σ (which we also denote by $\hat{A}$). The measure space $(\Omega, \Sigma, \hat{A})$ will be important for more general studies. We shall mainly consider measurable functions with finite value space here, and $(\Omega, \Sigma_0, \hat{A})$ will suffice for our purposes.

A *measurement* on (Ω, A) is a collection of mutually disjoint sets $E_i\in\Sigma_0$ such that $\Sigma|\hat{A}(E_i)|^2 = 1$. Notice that (Ω, A) is a point amplitude space if and only if the set of outcomes is a measurement. Denote the set of measurements on (Ω, A) by $M(\Omega, A)$. An *event* is a set $E\in\Sigma_0$ such that $E\in M$ for some $M\in M(\Omega, A)$. An event can belong to more than one measurement. If a distinction is needed, we say that E is an event of the measurement M. Notice, if $M(\Omega, A)\neq\varnothing$, then ϕ is an event. Also, if $\hat{A}(E) = 0$, E is an event called a *trivial* event. A measurement is *proper* if it contains no trivial events. For $M\in M(\Omega, A)$ denote the Boolean algebra generated by the sets of M by $B(M)$. Define a probability distribution P_A^M on $B(M)$ as follows: if $E = \cup E_i$, $E_i\in M$, then $P_A^M(E) = \Sigma|\hat{A}(E_i)|^2$. Notice that if $E\in M$, then $P_A^M(E) = |A(E)|^2$. We call the elements of $B(M)$ *compound events* of M. For $E\in B(M)$ we interpret $P_A^M(E)$ as the probability that E occurs upon execution of the measurement M. An amplitude space can have various measurements and an outcome can refer (belong) to one measurement but not to another. The values that are obtained from a measurement are given by a certain function. This function is defined only on the outcomes that refer to the given measurement. Motivated by the above, we call a function $f\colon E\subseteq\Omega\to R$ *observable* if $\{f^{-1}(\lambda)\colon\lambda\in R\}\in M(\Omega, A)$; that is, $\Sigma_{\lambda\in R}|\hat{A}\ [f^{-1}(\lambda)]|^2 = 1$. A finite set of function $f_1,\dots,f_n$ is *jointly observable* if

$$\{f_1^{-1}(\lambda_1)\cap\cdots\cap f_n^{-1}(\lambda_n) : \lambda_1,\dots,\lambda_n\in R\}\in M(\Omega, A).$$

Example 1: Let (Ω, P) be a finite probability space, where $\Omega = \{\omega_1,\dots,\omega_n\}$. Define $A:\Omega\to C$ by $A(\omega) = P(\omega)^{1/2}$.

Then (Ω, A) becomes a point amplitude space. The collection $M = \{\{\omega_1\},\dots,\{\omega_i\}\}$ is a measurement and $P_A^M = P$ on the powerset $P(\Omega) = B(M)$. In general, M is the only proper measurement on (Ω, A) and the only nontrivial events are the singleton sets $\{\omega_i\}$, $i = 1,\dots,n$. Any $E \in P(\Omega)$ is a compound event. In general, the only observable functions are the injective functions defined on all of Ω.

Example 2: Let (X, Σ, P) be a probability space. Define $\Omega = \Sigma$ and $A : \Omega \to C$ by $A(\omega) = P(\omega)^{1/2}$. Then (Ω, A) is an amplitude space but is not a point amplitude space. If M is a measurable partition of X, then $M \in M(\Omega, A)$. For any measurable partition M, $P_A^M = P$. Any random variable with countably many values is observable.

Example 3: Let H be a separable, complex Hilbert space with unit sphere Ω and let $\phi \in \Omega$. Define $A : \Omega \to C$ by $A(\omega) = \langle \phi, \omega\rangle$. Then (Ω, A) is an amplitude space, but is not a point amplitude space. If e_i, $i = 1,2, \dots$ is an orthonormal basis for H, then $M = \{\{e_1\}, \{e_2\}, \dots\}$ is a measurement. Indeed, $\Sigma \mid A(e_i)|^2 = \Sigma \mid \langle \phi, e_i\rangle|^2 = \|\phi\|^2 = 1$. It follows that any outcome is an event. Let $E = \{e_i : i = 1,2, \dots\}$ and suppose $f(e_i) = \lambda_i \in R$ where $\lambda_i \neq \lambda_j$, $i \neq j$. Then $f: E \to R$ is observable. If P_{e_i} is the projection onto e_i we can identify f with the self-adjoint operator $\Sigma\, \lambda_i\, P_{e_i}$.

Example 4: Let H be a separable, complex Hilbert space and let Ω be the lattice of orthogonal projections on H. Let $\phi \in H$ with $\|\phi\| = 1$. Define $A : \Omega \to C$ by $A(\omega) = \langle \omega\phi, \phi\rangle^{1/2}$. Again (Ω, A) is an amplitude space, but is not a point amplitude space. Let $P_i \in \Omega$, with $P_i\, P_j = 0, i \neq j = 1,2, \dots, \Sigma\, P_i = I$. Then $M = \{\{P_1\}, \{P_2\}, \dots\}$ is a measurement. Any outcome is an event. Let $E = \{P_i : i = 1,2, \dots\}$ and suppose $f(P_i) = \lambda_i \in R$ where $\lambda_i \neq \lambda_j$, $i \neq j$. Then $f: E \to R$ is observable and can be identified with the self-adjoint operator $\Sigma\, \lambda_i\, P_i$.

Example 5: Let $\Omega = \{\omega_1,\omega_2,\omega_3\}$ and $A(\omega_1) = \frac{2}{3} - \frac{1}{3}i$, $A(\omega_2) = \frac{1}{3} + \frac{1}{3}i$, $A(\omega_3) = \sqrt{2}/3$. The (Ω, A) is a point amplitude space and the proper measurements are $M_1 = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}\}$ and $M_2 = \{\{\omega_1,\omega_2\}\}$. Notice that $\{\omega_1,\omega_2\}$ is an event in M_2 and a compound event in M_1. However, it has a different probability of occurring depending on which measurement is executed. Indeed $\frac{7}{9} = P_A^{M_1}[\{\omega_1,\omega_2\}] \neq P_A^{M_2}[\{\omega_1, \omega_2\}] = 1$. The only observable functions defined on all of Ω are the injective functions. If f and g are defined on all of Ω and one or the other is injective, then they are jointly observable. If f and g are constant on Ω, they are not jointly observable.

Let (Ω, A) be an amplitude space and let $E \in \Sigma_0$. We define $\hat{A}(\cdot \mid E)$: $P(\Omega) \to C$ by $\hat{A}(F \mid E) = \hat{A}(F \cap E)/\hat{A}(E)$ if $\hat{A}(E) \neq 0$ and $A(F \mid E) = 0$ otherwise. Notice that if $\hat{A}(E) \neq 0$, then $(\hat{A}\cdot \mid E)$ is a complex measure on $P(\Omega)$ with $\hat{A}(\Omega \mid E) = 1$. We call $\hat{A}(F \mid E)$ the *conditional amplitude of F given E* and E is the *conditioning* set. Care must be taken with conditional amplitudes since $\hat{A}(E) = 0$ need not imply that $\hat{A}(F \cap E) = 0$ even if the sets are events. Because of this possibility, formulas such as $\hat{A}(F \cap E) = \hat{A}(E)\,\hat{A}(F \mid E)$ need not hold when $\hat{A}(E) = 0$. However, when the conditioning sets have nonzero amplitude, the following formulas hold:

$$\hat{A}(E_1 \cap \cdots \cap E_n) = \hat{A}(E_1)\,\hat{A}(E_2 \mid E_1)\,\hat{A}(E_3 \mid E_1 \cap E_2) \cdots \hat{A}(E_n \mid E_1 \cap \cdots \cap E_{n-1}) \tag{9}$$

if $F_i \cap F_j = \phi, i \neq j = 1,\dots,n$, and $\cup F_i = \Omega$ then

$$\hat{A}(E) = \sum \hat{A}(F_i)\,\hat{A}(E \mid F_i). \tag{10}$$

If $f, g : E \subseteq \Omega \to R$, then f is g *observable* if $\Sigma_\lambda \mid \hat{A}[f^{-1}(\lambda) \mid g^{-1}(\alpha)]|^2 = 1$ for every $\alpha \in R$ such that $\hat{A}[g^{-1}(\alpha)] \neq 0$. If g is observable and f is g observable, then f and g are jointly observable. Indeed, when $g^{-1}(\alpha)$ is nontrivial we have $\hat{A}[f^{-1}(\lambda) \cap g^{-1}(\alpha)] = \hat{A}[g^{-1}(\alpha)] \times \hat{A}[f^{-1}(\lambda) \mid g^{-1}(\alpha)]$. We say that f is g *orthogonal* if $\Sigma_\lambda \hat{A}[f^{-1}(\lambda) \mid g^{-1}(\alpha)]\, \overline{\hat{A}[f^{-1}(\lambda) \mid g^{-1}(\alpha')]} = 0$ for every $\alpha \neq \alpha'$, where $\overline{}$ denotes the complex conjugate.

4. MARKOV AMPLITUDE CHAINS

Markov amplitude chains were considered a long time ago in Ref. 4. However, we shall treat them in more depth here. Throughout the sequel, (Ω, A) will denote an amplitude space and $S = \{a_1,\dots, a_r\}$ a fixed finite subset of R. If $h_j : E \subseteq \Omega \to S, j = 0,\dots,n$, are Σ_0 measurable we call $\{h_j\}_0^n$ an *amplitude n-chain* with *value space S*. We interpret S as the set of possible locations of a randomly moving physical system, and $h_0,\dots,h_n$ give the location of the system at the times $0,1,\dots,n$, respectively. The numbers in S need not correspond to position values, but could be interpreted as values for any physical quantity such as energy, momentum, spin, charge, color, flavor, etc. The possible n-step paths of the system are given by the set of sample paths $(h_0(\omega),\dots,h_n(\omega))$, $\omega \in E$. The *kth vector* of the n-chain is the vector $\phi_k \in C^r$ given by $\phi_k(j) = \hat{A}[h_k^{-1}(a_j)], j = 1,\dots,r;\ k = 0,\dots,n$. Notice that ϕ_k is a unit vector if and only if h_k is observable. We call ϕ_0 and ϕ_n the *initial* and *final vectors*, respectively.

We define the *amplitude matrix* of $\{h_j\}_0^n$ by $A(k,j) = \hat{A}[h_0^{-1}(a_j) \cap h_n^{-1}(a_k)], j,k = 1,\dots,r$. The matrix element $A(k,j)$ gives the amplitude that a system moves from a_j to a_k in n steps. This amplitude is the sum of the amplitudes for all possible n-step paths from a_j to a_k. Indeed, since $\hat{A}$ is a complex measure, it follows that $A(k,j) = \Sigma^r_{i_1,\dots,i_{n-1}=1} \times \hat{A}[h_0^{-1}(a_j) \cap h_1^{-1}(a_{i_1}) \cap \cdots \cap h_{n-1}^{-1}(a_{i_{n-1}}) \cap h_n^{-1}(a_k)]$. We also define the *conditional amplitude matrix* $A'(k,j) = \hat{A}[h_n^{-1}(a_k) \mid h_0^{-1}(a_j)], j, k = 1,\dots,r$. If the initial vector is nonvanishing $[\phi_0(j) \neq 0, j = 1,\dots,r]$, then $A(k,j) = \phi_0(j)\,A'(k,j)$ and hence

$$\begin{aligned}\phi_n(k) &= \hat{A}[h_n^{-1}(a_k)] = \sum_j \hat{A}[h_0^{-1}(a_j) \cap h_n^{-1}(a_k)] = \sum_j A(k,j)\\ &= \sum_j \phi_0(j)\,A'(k,j) = (A'\phi_0)(k).\end{aligned}$$

Hence, $\phi_n = A'\phi_0$.

An $r \times r$ matrix T is called a *stochastic matrix* if $\Sigma_k\ T(k,j) = 1, j = 1,\dots,r$ (notice this is different than the usual terminology where it is also assumed that $T(k,j) \geqslant 0$). An $r \times r$ matrix T is called a *stochastic amplitude matrix* if $|T(k,j)|^2$ is an amplitude matrix (that is, $\Sigma_k \mid T(k,j)|^2 = 1$, $j = 1,\dots,r$). Notice that h_n is h_0 observable if and only if ϕ_0 is nonvanishing and A' is a stochastic amplitude matrix. Suppose A is a stochastic amplitude matrix. This holds if and only if the collection of sets $\{h_0^{-1}(a_j) \cap h_n^{-1}(a_k) : k = 1,\dots,r\}$ is a measurement for each $j = 1,\dots,r$. We then call $\{h_j\}_0^n$ a *constrained n*-chain. In this case, we interpret $|A(k,j)|^2$ as the probability that a system which is initially constrained at a_j, then moves to a_k in n steps. Another important case is when h_0 and h_n are jointly observable. This holds if and only if $\Sigma_{j,k}|A(k,j)|^2 = 1$, and we then call $\{h_j\}_0^n$ a *definite n*-chain. We interpret $|A(k,j)|^2$, in this case, as the probability that a system begins at a_j and moves to a_k in n steps.

We call $\{h_j\}_0^n$ a *Markov amplitude n-chain if*

$$\hat{A}[h_j^{-1}(a_{i_j})|\,h_0^{-1}(a_{i_0})\cap h_{j-1}^{-1}(a_{i_{j-1}})] = \hat{A}[h_j^{-1}(a_{i_j})|\,h_{j-1}^{-1}(a_{i_{j-1}})]$$

for all $i_0,\ldots,i_j, j = 1,\ldots,n$. We say that $\{h_j\}_0^n$ is *stationary* if

$$\hat{A}[h_m^{-1}(a_k)|\,h_{m-1}^{-1}(a_j)] = \hat{A}[h_1^{-1}(a_k)|\,h_0^{-1}(a_j)]$$

for all $j,k = 1,\ldots,r$ $m = 1,\ldots,n$.

Theorem 2: Let $\{h_j\}_0^n$ be a stationary, Markov, amplitude n-chain for which ϕ_0 is nonvanishing and define the matrix $T(k,j) = \hat{A}[h_1^{-1}(a_k)|\,h_0^{-1}(a_j)]$.

(a) Then T is a stochastic matrix and

$$\hat{A}[h_0^{-1}(a_{i_0})\cap\cdots\cap h_n^{-1}(a_{i_n})] = \phi_0(i_0)\,T(i_1,i_0)\cdots T(i_n,i_{n-1}), \tag{11a}$$

$$A(k,j) = \phi_0(j)\,T^n(k,j), \tag{11b}$$

$$A'(k,j) = T^n(k,j), \tag{11c}$$

$$\phi_k = T^k\phi_0,\quad k = 0,1,\ldots,n. \tag{11d}$$

(b) If, in addition, h_0 is observable and h_1 is h_0 observable and orthogonal, then T is unitary and any subset of $\{h_j : j = 0,\ldots,n\}$ is jointly observable. In particular, each h_j is observable and $\{h_j\}_0^n$ is definite.

Proof: (a) Since $\hat{A}[h_0^{-1}(a_j)]\neq 0$ and $\hat{A}[\cdot\,|\,h_0^{-1}(a_j)]$ is a complex measure, we have

$$\sum_k T(k,j) = \hat{A}[\cup h_1^{-1}(a_k)|\,h_0^{-1}(a_j)] = \hat{A}[E\,|\,h_0^{-1}(a_j)] = 1.$$

Hence, T is a stochastic matrix. Equation (11a) follows from Markovicity, stationarity, and Eq. (9). Equation (11b) follows from letting $i_0 = j$, $i_n = k$ in Eq. (11a) and summing over $i_1,\ldots,i_{n-1}$. Equation (11c) follows directly from (11b). Replace n by k in (11a) and sum over $i_0,\ldots,i_{k-1}$ to obtain (11d).

(b) Since h_1 is h_0 observable and orthogonal, we have $\Sigma_k\,T(k,j)\,\bar{T}(k,j') = \Sigma_k\,T^*(j',k)\,T(k,j) = \delta_{jj'}$, and hence, T is unitary. We next show that T^m is a stochastic matrix for any $m = 0,1,\ldots$. Indeed, T^0 and T are stochastic and suppose T^m is stochastic. Then

$$\sum_k T^{m+1}(k,j) = \sum_k\sum_i T^m(k,i)\,T(i,j) = \sum_i T(i,j)\sum_k T^m(k,j) = \sum_i T(i,j) = 1.$$

Hence, the result follows by induction. Consider a subset $\{h_0, h_{h_1},\ldots,h_{j_k}\}$ of $\{h_j : j = 0,\ldots,n\}$ where $0 < j_1 < j_2 < \cdots < j_k < n$. Summing (11a) over the other indices gives

$$\hat{A}[h_0^{-1}(a_{i_0})\cap h_{j_1}^{-1}(a_{i_1})\cap\cdots\cap h_{j_k}^{-1}(a_{i_k})] = \phi_0(i_0)\,T^{j_1}(i_1,i_0)\,T^{j_2-j_1}(i_2,i_1)\cdots T^{j_k-j_{k-1}}(i_k,i_{k-1})\sum_i T^{n-j_k}k(i,i_k).$$

Since T^{n-j_k} is stochastic, the last sum is 1. Since T^k is unitary for $k = 0,1,\ldots$ and $\|\phi_0\| = 1$, taking the modulus squared and summing over $i_0,i_1,\ldots,i_k$ again gives 1. It follows that $\{h_0, h_{j_1},\ldots,h_{j_k}\}$ is jointly observable. The other cases follow in a similar manner. ■

The above theorem shows that stationary, Markov, amplitude n-chains can be very well behaved. In particular, for the situation described in part (b), we can not only observe any trajectory (sample path), but we can observe a sequence of "snap shots" in which time is stopped whenever we wish. We shall see in later sections that n-chains which describe quantum systems need not be stationary although they frequently are Markov. We then lose the ability to observe individual trajectories. However, frequently h_n is observable so we can observe where the system ends or h_0 and h_n are jointly observable so we can observe the beginning and the end. We shall also show that (11a)–(11d) hold in slightly altered form. The reader may wonder if it is possible for a nontrivial matrix to be both unitary and stochastic as in Theorem 2(b). Of course, the identity matrix has this property and it is easy to construct 0-1 matrices with this property. However, there are nontrivial examples of unitary stochastic matrices as the following examples show:

$$\begin{bmatrix}\frac{1}{2}+\frac{1}{2}i & \frac{1}{2}-\frac{1}{2}i\\ \frac{1}{2}-\frac{1}{2}i & \frac{1}{2}+\frac{1}{2}i\end{bmatrix},$$

$$\begin{bmatrix}\frac{1}{3}+(1/\sqrt{3})i & \frac{1}{3} & \frac{1}{3}-(1/\sqrt{3})i\\ \frac{1}{3}-(1/\sqrt{3})i & \frac{1}{3}+(1/\sqrt{3})i & \frac{1}{3}\\ \frac{1}{3} & \frac{1}{3}-(1/\sqrt{3})i & \frac{1}{3}+(1/\sqrt{3})i\end{bmatrix}.$$

5. QUANTUM *n*-CHAINS

We have seen that a stationary, Markov, amplitude n-chain has a joint amplitude satisfying Eq. (11a). We now use this equation to define an important class of n-chains. An n-chains $\{h_j\}_0^n$ is called a *quantum n*-chain if there exists a vector $0\neq\psi_0\in C^r$ and an $r\times r$ complex matrix T such that

$$\hat{A}[h_0^{-1}(a_{i_0})\cap\cdots\cap h_n^{-1}(a_{i_n})] = \psi_0(i_0)\,T(i_1,i_0)\cdots T(i_n,i_{n-1}). \tag{12}$$

We call ψ_0 and T the *vector* and *matrix* of $\{h_j\}_0^n$, respectively. We denote the vector of $\{h_j\}_0^n$ by ψ_0 instead of ϕ_0 since the vector may not equal the initial vector. The relationship between ψ_0 and ϕ_0 is given in Theorem 3. If $\|\psi_0\| = 1$, we call ψ_0 the *initial state*. If T is unitary we call $\{h_j\}_0^n$ a *closed* quantum n-chain.

Theorem 3: Let $\{h_j\}_0^n$ be a quantum n-chain with vector ψ_0 and matrix T.

(a) The amplitude matrix, conditional amplitude matrix, initial vector, and final vector satisfy

$$A(k,j) = \psi_0(j)\, T^n(k,j), \tag{13a}$$

$$A'(k,j) = T^n(k,j) \Big/ \sum_k T^n(k,j) \quad \text{if } \phi_0(j) \neq 0, \tag{13b}$$

$$\phi_0(j) = \psi_0(j) \sum_k T^n(k,j), \tag{13c}$$

$$\phi_n = T^n \psi_0. \tag{13d}$$

(b) If $\|\psi_0\| = 1$ and T^n is a stochastic amplitude matrix, then $\{h_j\}_0^n$ is definite (and hence, h_0 and h_n are jointly observable).

(c) If $\psi_0 \equiv 1$, then T^n is a stochastic amplitude matrix if and only if $\{h_j\}_0^n$ is constrained.

(d) If $\{h_j\}_0^n$ is closed, then $\|\psi_0\| = 1$ implies it is definite and h_n is observable, and $\psi_0 \equiv 1$ implies it is constrained.

Proof: (a) The proof of (13a) is the same as the proof of (11b). Sum (13a) over k to obtain (13c). (13b) follows from (13a) and (13c). Sum (13a) over j to obtain (13d).

(b) Take the square of the modulus of (13a) and sum over j and k.

(c) Same proof as (b).

(d) The first part follows from (b) and (13d). The second part follows from (c). ∎

Notice that formulas (12), (13a), and (13d) are analogous to formulas given in Sec. 2 on quantum dynamics. In that case the system was closed and T was given by a unitary operator U. In our present situation, suppose $\{h_j\}_0^n$ is closed and has unitary matrix U. By the spectral theorem there exists a unique self-adjoint $r \times r$ matrix H such that $U = e^{-iH}$. We call H the *Hamiltonian* for $\{h_j\}_0^n$. Observe from Theorem 3 that $\psi_0 = \phi_0$ if T (or T^n) is a stochastic matrix.

The next theorem summarizes some of the important properties of quantum n-chains. Among other things, it shows that if certain amplitudes are nonzero, then a quantum n-chain is Markov. Corresponding to an $r \times r$ matrix T we define vectors $T_m \in C^r, m = 0,\ldots,n$ by $T_m(j) = \sum_k T^{n-m}(k,j), j = 1,\ldots,r$. Notice that T is stochastic if and only if $T_m \equiv 1$, $m = 0,\ldots,n$.

Theorem 4: Let $\{h_j\}_0^n$ be a quantum n-chain with vector ψ_0 and matrix T.

(a) If ϕ_k is nonvanishing and $\hat{A}\left[h_0^{-1}(a_{i_0}) \cap \cdots \cap h_k^{-1}(a_{i_k})\right] \neq 0$ for all $k, i_0,\ldots,i_k$, then $\{h_j\}_0^n$ is Markov.

(b) If the terms are nonzero, then

$$\hat{A}\left[h_m^{-1}(a_k) \mid h_{m-1}^{-1}(a_j)\right] = T(k,j)\, T_m(k)/T_{m-1}(j).$$

(c) If the terms are nonzero, then

$$\begin{aligned}\hat{A}&\left[h_m^{-1}(a_k) \mid h_{m-1}^{-1}(a_j)\right] \\ &= \hat{A}\left[h_i^{-1}(a_k) \mid h_{i-1}^{-1}(a_j)\right] \\ &\quad \times T_m(k)\, T_{i-1}(j)/T_{m-1}(j)\, T_i(k).\end{aligned}$$

Proof: (a) Summing over various indices in (12) gives the following formulas:

$$\begin{aligned}\hat{A}&\left[h_0^{-1}(a_{i_0}) \cap \cdots \cap h_m^{-1}(a_{i_m})\right] \\ &= \psi_0(i_0)\, T(i_1,i_0) \cdots T(i_m,i_{m-1})\, T_m(i_m),\end{aligned} \tag{14a}$$

$$\hat{A}\left[h_m^{-1}(a_j)\right] = (T^m\psi_0)(j)\, T_m(j), \tag{14b}$$

$$\hat{A}\left[h_m^{-1}(a_j) \cap h_{m-1}^{-1}(a_k)\right] = (T^{m-1}\psi_0)(k)\, T(j,k)\, T_m(j). \tag{14c}$$

It follows from (14a) that

$$\begin{aligned}\hat{A}&\left[h_m^{-1}(a_{i_m}) \mid h_0^{-1}(a_{i_0}) \cap \cdots \cap h_{m-1}^{-1}(a_{i_{m-1}})\right] \\ &= T(i_m,i_{m-1})\, T_m(i_m)/T_{m-1}(i_{m-1}).\end{aligned}$$

Applying (14b) and (14c) we obtain

$$\begin{aligned}\hat{A}&\left[h_m^{-1}(a_{i_m}) \mid h_{m-1}^{-1}(a_{i_{m-1}})\right] \\ &= T(i_m,i_{m-1})\, T_m(i_m)/T_{m-1}(i_{m-1}).\end{aligned} \tag{14d}$$

Hence, $\{h_j\}_0^n$ is Markov.

(b) This follows from (14d).

(c) This follows from part (b). ∎

In Theorem 4(a) we gave a sufficient condition for a quantum n-chain to be Markov. In the following corollary we strengthen this to a characterization.

Corollary 5: A quantum n-chain with vector ψ_0 and matrix T is Markov if and only if the following two conditions hold for all $m = 1,2,\ldots,n$.

(1) $(T^m\psi_0)(i_m)\, T_m(i_m) = 0$ whenever $\psi_0(i_0)\, T(i_1,i_0) \cdots T(i_m,i_{m-1})\, T_m(i_m) = 0$ for some $i_0,\ldots,i_{m-1}$.

(2) $\psi_0(i_0)\, T(i_1,i_0) \cdots T(i_m,i_{m-1})\, T_m(i_m) = 0$ for every $i_0,\ldots,i_{m-1}$ whenever $(T^m\psi_0)(i_m)\, T_m(i_m) = 0$.

Proof: If follows from the proof of Theorem 4 that $\{h_j\}_0^n$ being Markov is equivalent to the two conditioning sets in the Markov definition having zero amplitude simultaneously. The result then follows from (14a) and (14b). ∎

We see from Theorem 4(b) and (c) that a quantum n-chain is stationary if its matrix T is stochastic. However, unlike the Markov condition, this is a very strong requirement, and in general we would not expect a quantum n-chain to be stationary. Nevertheless, the conditions in Theorem 4(b) and (c) are weak stationary conditions. Although $\hat{A}\left[h_m^{-1}(a_k) \mid h_{m-1}^{-1}(a_j)\right]$ is not independent of m, we can separate out some of the m dependence. We now make this idea precise. We call an n-chain $\{h_j\}_0^n$ *almost stationary* if there exist vectors $\delta_m \in C^r$, $m = 0,\ldots,n$, with $\delta_n \equiv 1$ such that

$$\begin{aligned}\delta_{m-1}(j)\,\delta_i(k)\,\hat{A}&\left[h_m^{-1}(a_k) \mid h_{m-1}^{-1}(a_j)\right] \\ &= \delta_m(k)\,\delta_{i-1}(j)\,\hat{A}\left[h_i^{-1}(a_k) \mid h_{i-1}^{-1}(a_j)\right]\end{aligned} \tag{15}$$

for all applicable i,j,k,m.

Lemma 6: (a) If $\{h_j\}_0^n$ is a quantum n-chain with matrix T and ϕ_m and T_m nonvanishing for $m = 0,\ldots,n$, then $\{h_j\}_0^n$ is almost stationary.

(b) If $\{h_j\}_0^n$ is an almost stationary, Markov n-chain with ϕ_m and δ_m nonvanishing for $m = 0,\ldots,n$, then $\{h_j\}_0^n$ is quantum with matrix $T(j,k) = \delta_0(k)\,\delta_1(j)^{-1}\hat{A} \times \left[h_1^{-1}(a_j) \mid h_0^{-1}(a_k)\right]$, vector $\psi_0(j) = \delta_0(j)^{-1}\phi_0(j)$, and $\delta_m = T_m$, $m = 0,\ldots,n$.

Proof: (a) This follows from Theorem 4(c).

(b) Using Markovicity and (15), the joint amplitude becomes

$$\hat{A}\left[h_0^{-1}(a_{i_0})\cap\cdots\cap h_n^{-1}(a_{i_n})\right]$$
$$=\hat{A}\left[h_0^{-1}(a_{i_0})\right]\hat{A}\left[h_1^{-1}(a_{i_1})|h_0^{-1}(a_{i_0})\right]$$
$$\cdots\hat{A}\left[h_n^{-1}(a_{i_n})|h_{n-1}^{-1}(a_{i_{n-1}})\right]$$
$$=(\delta_0(i_1)\cdots\delta_0(i_{n-1})/\delta_1(i_1)\cdots\delta_1(i_n))\,\hat{A}\left[h_0^{-1}(a_{i_0})\right]$$
$$\times\hat{A}\left[h_1^{-1}(a_{i1})|h_0^{-1}(a_{i_0})\right]\cdots\hat{A}\left[h_1(a_{i_n})|h_0^{-1}(a_{i_{n-1}})\right]$$
$$=\delta_0(i)^{-1}\hat{A}\left[h_0^{-1}(a_{i_0})\right]T(i_1,i_0)\cdots T(i_n,i_{n-1}),$$

where

$$T(i_j,i_{j-1})=\delta_0(i_{j-1})\delta_1(i_j)^{-1}\hat{A}\left[h_1^{-1}(a_{i_j})|h_0^{-1}(a_{i_{j-1}})\right].$$

This shows that $\{h_j\}_0^n$ is quantum with the prescribed matrix T and vector ψ_0. We prove that $\delta_m=T_m, m=0,\ldots,n$ by reverse induction. Clearly $\delta_n=T_n\equiv 1$. Suppose that $\delta_m=T_m$ for some integer m where $0<m\leqslant n$. By (15) we have

$$\hat{A}\left[h_m^{-1}(a_k)|h_{m-1}^{-1}(a_j)\right]$$
$$=\frac{\delta_m(k)\delta_0(j)}{\delta_{m-1}(j)\delta_1(k)}\hat{A}\left[h_1^{-1}(a_k)|h_0^{-1}(a_j)\right]$$
$$=\frac{\delta_m(k)}{\delta_{m-1}(j)}T(k,j).$$

Summing over k gives

$$\delta_{m-1}(j)$$
$$=\sum_k\delta_m(k)T(k,j)=\sum_k\sum_i T^{n-m}(i,k)T(k,j)$$
$$=\sum_i T^{n-m+1}(i,j)=T_{m-1}(j).\qquad\blacksquare$$

6. RANDOM PHASE TRANSFORMATIONS

In this section we shall study the changes that result in an amplitude space and in the n-chains due to a random phase transformation. In the sequel $\{h_j\}_0^n$ will be an n-chain on (Ω,A) with value space $S=\{a_1,\ldots,a_r\}$. For simplicity, we shall assume that the functions $h_j, j=0,\ldots,n$, are defined on all of Ω. In this case, we have $\Omega\in\Sigma_0$. We say that $\{h_j\}_0^n$ is *separating* if $h_j(\omega)=h_j(\omega')$ for $j=0,\ldots,n$, implies that $\omega=\omega'$. Define

$$L^1(\Omega,A)=\left\{\beta:\Omega\to C:\sum_{\omega\in\Omega}|\beta(\omega)A(\omega)|<\infty\right\}.$$

For $\beta\in L^1(\Omega,A)$ define

$$\sum\beta A\equiv\sum\beta(\omega)A(\omega)=\sum_{\omega\in\Omega}\beta(\omega)A(\omega)$$

and

$$\sum_{a_j}^{a_k}\beta A\equiv\sum_{a_j}^{a_k}\beta(\omega)A(\omega)$$
$$=\sum\{\beta(\omega)A(\omega):h_0(\omega)=a_j,h_n(\omega)=a_k\}.$$

Notice that $A(k,i)=\Sigma_{a_k}^{a_j}A$.

For $\alpha:\Omega\to R$, define $A^\alpha:\Omega\to C$ by $A^\alpha(\omega)=e^{-i\alpha(\omega)}A(\omega)$. We call $A\to A^\alpha$ a *random phase transformation*. Notice that $A^0=A$. If α is a constant function, then A^α and A are essentially the same since the probabilities are unchanged and $\hat{A}^\alpha(\cdot|\cdot)=\hat{A}(\cdot|\cdot)$. In general, (Ω,A^α) gives an entirely different amplitude space than (Ω,A). If (Ω,A) is a point amplitude space, then it is clear that (Ω,A^α) is also. In this case, the collection of singleton outcome sets is a measurement in both (Ω,A) and (Ω,A^α). In general, however, (Ω,A) and (Ω,A^α) have different measurements and events. If $\{h_j\}_0^n$ is separating, then it is jointly observable in (Ω,A) if and only if it is jointly observable in (Ω,A^α). Indeed, we then have

$$\hat{A}^\alpha\left[h_0^{-1}(a_{i_0})\cap\cdots\cap h_n^{-1}(a_{i_n})\right]$$
$$=e^{-i\alpha(\omega)}\hat{A}\left[h_0^{-1}(a_{i_0})\cap\cdots\cap h_n^{-1}(a_{i_n})\right],$$

where $h_j(\omega)=a_{i_j}, j=0,\ldots,n$ [if such an ω does not exist, then both A and A^α are zero on $h_0^{-1}(a_{i_0})\cap\cdots\cap h_n^{-1}(a_{i_n})$]. However, even when $\{h_j\}_0^n$ is separating, one can have h_j observable in (Ω,A) but not observable in (Ω,A^α).

If $\{h_j\}_0^n$ is Markov, stationary, almost stationary, quantum, etc., relative to A, it need not be relative to A^α. The most regular nontrivial case is when α has the following form. Let $v:S\to R$ and define $\alpha_v:\Omega\to R$ by $\alpha_v(\omega)=\Sigma_{j=0}^n v[h_j(\omega)]$. Define $A^v=A^{\alpha_v}$.

Theorem 7: Let $\{h_j\}_0^n$ be a quantum n-chain on (Ω,A) with vector ψ_0, and matrix T, and let $v:S\to R$. Let $\{h_j'\}_0^n$ be the same n-chain considred on (Ω,A^v).

(a) $\{h_j'\}_0^n$ is a quantum n-chain with vector $\psi_0'(j)=e^{-iv(a_j)}\psi_0(j)$ and matrix $T'=e^{-iv(Q)}T$, where Q is the $r\times r$ diagonal matrix $Q=\text{diag}(a_1,\ldots,a_r)$.

(b)

$$A^v(k,j)=\psi_0'T'^n(k,j)$$
$$=\sum_{a_j}^{a_k}\exp\left[-i\sum_{m=0}^n v(h_m(\omega))\right]A(\omega).$$

(c) If $\{h_j\}_0^n$ is closed, then so is $\{h_j'\}_0^n$. If $\{h_j\}_0^n$ is closed and definite, then so is $\{h_j'\}_0^n$.

Proof: (a) The joint amplitude of $\{h_j'\}_0^n$ becomes

$$\hat{A}^v\left[h_0^{-1}(a_{i_0})\cap\cdots\cap h_n^{-1}(a_{i_n})\right]$$
$$=\sum\{A^v(\omega):h_0(\omega)=a_{i_0},\ldots,h_n(\omega)=a_{i_n}\}$$
$$=\sum\{e^{-i\alpha_v(\omega)}A(\omega):h_0(\omega)=a_{i_0},\ldots,h_n(\omega)=a_{i_n}\}$$
$$=\exp i\left[-\sum_{j=0}^n v(a_{i_j})\right]\hat{A}\left[h_0^{-1}(a_{i_0})\cap\cdots\cap h_n^{-1}(a_{i_n})\right]$$
$$=\prod_{j=0}^n e^{-iv(a_{i_j})}\psi_0(i_0)T(i_1,i_0)\cdots T(i_n,i_{n-1})$$
$$=e^{-iv(a_{i_0})}\psi_0(i_0)\prod_{j=1}^n e^{-iv(a_{i_j})}T(i_j,i_{j-1})$$
$$=e^{-iv(a_{i_0})}\psi_0(i_0)\prod_{j=1}^n\left[e^{-iv(Q)}T\right](i_j,i_{j-1}).$$

The proofs of (b) and (c) are straightforward. $\blacksquare$

If $\{h_j\}_0^n$ is a closed quantum n-chain and has Hamiltonian H_0, then $T=e^{-iH_0}$ and $T'=e^{-iv(Q)}e^{-iH_0}$. The Hamiltonian H for the closed quantum n-chain $\{h_j'\}_0^n$, then satisfies $e^{-iH}=e^{-iv(Q)}e^{-iH_0}$. In practice, H_0 corresponds to the free Hamiltonian and $v(Q)$ to the potential.

Let (Ω,A) be a point amplitude space. As in Sec. 3, we define the probability distribution $P_A(E) = \Sigma_{\omega \in E} |A(\omega)|^2$, $E \in P(\Omega)$. Notice that $P_A(\omega) = \overline{A}(\omega) A(\omega)$ so P_A results from a random change in A although it is not a random phase change. If $\{h_j\}_0^n$ is an n-chain on (Ω, A) we call $\{h_j\}_0^n$ on (Ω, P_A) the *parallel classical* n-chain.

Lemma 8: Let $\{h_j\}_0^n$ be a separating, definite, closed quantum n-chain with vector ψ_0 and matrix T on a point amplitude space (Ω, A). Then the parallel classical n-chain is a stationary Markov n-chain with initial probability vector $\psi_0'(j) = |\psi_0(j)|^2$ and transition matrix $W(j,k) = |T(j,k)|^2$.

Proof: The result easily follows from

$$\begin{aligned}P_A&[h_0^{-1}(a_{i_0}) \cap \cdots \cap h_n^{-1}(a_{i_n})]\\&= |\widehat{A}[h_0^{-1}(a_{i_0}) \cap \cdots \cap h_n^{-1}(a_{i_n})]|^2\\&= |\psi_0(i_0)|^2 |T(i_1,i_0)|^2 \cdots |T(i_n,i_{n-1})|^2. \qquad \blacksquare\end{aligned}$$

The parallel classical n-chain gives entirely different probabilities than the original amplitude n-chain. For example, in the case considered in Lemma 8 we have

$$|\widehat{A}[h_n^{-1}(a_k)]|^2 = |(T^n \psi_0)(k)|^2$$

and

$$P_A[h_n^{-1}(a_k)] = (W^n \psi_0')(k).$$

In particular, if ψ_0 is the standard basis element e_j, then $|\widehat{A}[h_n^{-1}(a_k)]|^2 = |T^n(k,j)|^2$ and $P_A[h_n^{-1}(a_k)] = W^n(k,j)$. For 1-chains these agree. This indicates that quantum mechanical effects do not have time to occur during just one time step. However, for 2-chains we have

$$|\widehat{A}[h_2^{-1}(a_k)]|^2 = \left|\sum_i T(k,i)T(i,j)\right|^2,$$

while

$$P_A[h_2^{-1}(a_k)] = \sum_i |T(k,i)T(i,j)|^2,$$

which are quite different.

Also, as we have seen, a random phase transformation $A \to A^\alpha$ makes a profound change in an amplitude n-chain. However, since $P_A = P_{A^\alpha}$, random phase transformations cannot be distinguished in the parallel classical n-chain.

We now present a lemma which is interesting in its own right and which is needed to prove Theorem 10.

Lemma 9: Let $\{h_j\}_0^n$ be a quantum, n-chain with vector ψ_0 and matrix T. Let s be a positive integer and let F be a map from $S \times S \times \cdots \times S$ (s times) to R. If $0 \leqslant i_1 \leqslant \cdots \leqslant i_s \leqslant n$, then

$$\begin{aligned}\sum_{a_j}^{a_k} &F[h_{i_1}(\omega),\ldots,h_{i_s}(\omega)]A(\omega)\\&= \sum_{j_1,\ldots,j_s=1} F(a_{j_1},\ldots,a_{j_s})\psi_0(j)T^{i_1}(j_1,j)\\&\quad\times T^{i_2-i_1}(j_2,j_1)\cdots T^{n-i_s}(k,j_s).\end{aligned}$$

Proof:

$$\begin{aligned}\sum_{a_j}^{a_k} &F[h_{i_1}(\omega),\ldots,h_{i_s}(\omega)]A(\omega)\\&= \sum \{F[h_{i_1}(\omega),\ldots,h_{i_s}(\omega)]A(\omega): h_0(\omega) = a_j, h_n(\omega) = a_k\}\\&= \sum_{j_1,\ldots,j_s=1} \sum \{F[h_{i_1}(\omega),\ldots,h_{i_s}(\omega)]A(\omega): h_0(\omega)\\&= a_j, h_n(\omega) = a_k, h_{i_m}(\omega) = a_{j_m}, m = 1,\ldots,s\}\\&= \sum_{j_1,\ldots,j_s=1} F(a_{j_1},\ldots,a_{j_s}) \sum \{A(\omega): h_0(\omega)\\&= a_j, h_n(\omega) = a_k, h_{i_m}(\omega) = a_{j_m}, m = 1,\ldots,s\}\\&= \sum_{j_1,\ldots,j_s=1} F(a_{j_1},\ldots,a_{j_s})\\&\times \widehat{A}[h_0^{-1}(a_j) \cap h_{i_1}^{-1}(a_{j_1}) \cap \cdots \cap h_{i_s}^{-1}(a_{j_s}) \cap h_n^{-1}(a_k)]\\&= \sum_{j_1,\ldots,j_s=1} F(a_{j_1},\ldots,a_{j_s})\psi_0(j)T^{i}(j_1,j)\\&\quad\times T^{i_2-i_1}(j_2,j_1)\cdots T^{n-i_s}(k,j_s). \qquad \blacksquare\end{aligned}$$

Let $0 \leqslant i_1,\ldots,i_s \leqslant n$ be integers and place these integers in nondecreasing order. For $1 \leqslant m \leqslant s$, define $\hat{i}_m$ to be the mth element of the resulting ordered sequence. The next theorem gives a perturbation expansion for $A^v(k,j)$.

Theorem 10: If $\{h_j\}_0^n$ is a quantum n-chain with vector ψ_0 and matrix T, and $v: S \to R$, then

$$A^v(k,j) = A(k,j) + A^1(k,j) + A^2(k,j) + \cdots,$$

where

$$\begin{aligned}A^{(s)}(k,j) &= \frac{(-i)^s}{s!}\\&\times \sum_{i_1,\ldots,i_s=0} \sum_{j_1,\ldots,j_s=1} v(a_{j_1})\cdots v(a_{j_s})\psi_0(j)T^{\hat{i}_1}(j_1,j)\\&\times T^{\hat{i}_2-\hat{i}_1}(j_2,j_1)\cdots T^{n-\hat{i}_s}(k,j_s).\end{aligned}$$

Proof: Expanding the exponential in Theorem 7(b) gives

$$\begin{aligned}A^v(k,j) &= \sum_{a_j}^{a_k} \left\{1 - i \sum_m v[h_m(\omega)]\right.\\&\quad \left. + \cdots + \frac{(i)^s}{s!}\left[\sum_m v(h_m(\omega))\right]^s + \cdots \right\} A(\omega)\\&= \sum_{a_j}^{a_k} A(\omega) - i \sum_{a_j}^{a_k} \sum_m v[h_m(\omega)]A(\omega)\\&\quad + \cdots + \frac{(i)^s}{s!} \sum_{a_j}^{a_k} \left[\sum_m v(h_n(\omega))\right]^s A(\omega) + \cdots\\&= A(k,j) + \cdots + \frac{(i)^s}{s!} \sum_{i_1,\ldots,i_s=1} \sum_{a_j}^{a_k} v[h_{i_1}(\omega)]\\&\quad \cdots v[h_{i_s}(\omega)]A(\omega) + \cdots.\end{aligned}$$

An application of Lemma 9 gives

$$\begin{aligned}\sum_{a_j}^{a_k} &v[h_{i_1}(\omega)]\cdots v[h_{i_s}(\omega)]A(\omega)\\&= \sum_{j_1,\ldots,j_s=1} v(a_j)\cdots v(a_{j_s})\psi_0(j)T^{\hat{i}_1}(j_1,j)T^{\hat{i}_2-\hat{i}_1}(j_2,j_1)\\&\quad \cdots T^{n-\hat{i}_s}(k,j_s).\end{aligned}$$

The result now follows. $\blacksquare$

To get a better understanding of the perturbation expansion, let us write out $A^{(1)}(k,j)$ and $A^{(2)}(k,j)$ in detail:

$$A^{(1)}(k,j) = -i\sum_{m=0}^{n}\sum_{j_1=1}^{r}\psi_0(j)T^{m}(j_1,j)v(a_{j_1})T^{n-m}(k,j_1),$$

$$\begin{aligned}A^{(2)}(k,j) = &-\sum_{i_1<i_2=0}^{n}\sum_{j_1,j_2=1}^{r}\psi_0(j)T^{i_1}(j_1,j)v(a_{j_1})\\ &\times T^{i_2-i_1}(j_2,j_1)v(a_{j_2})T^{n-i_2}(k,j_2)\\ &-\frac{1}{2}\sum_{i_1=0}^{n}\sum_{j_1=1}^{r}\psi_0(j)T^{i_1}(j_1,j)v(a_{j_1})^2T^{n-i_1}(k,j_1).\end{aligned}$$

We follow Feynman and Hibbs[2] in interpreting the expansion. The amplitude matrix element $A^{v}(k,j)$ is a sum of alternative ways of going from a_j to a_k: not scattered at all $[A(k,j)]$, scattered once $[A^{(1)}(k,j)]$, scattered twice $[A^{(2)}(k,j)]$,... . Each of these alternatives is a sum of alternatives. For example, $A^{(1)}(k,j)$ is a sum of terms $\psi_0(j)T^{m}(j_1,j)v(a_{j_1})T^{n-m}(k,j_1)$. The system moves "freely" from a_j to a_{j_1} in m steps, is scattered by the potential $v(a_{j_1})$, and then moves "freely" from a_{j_1} to a_k in $n-m$ steps. The term $A^{(2)}(k,j)$ has a similar interpretation except something new happens in the second summation of this term. Here the system moves "freely" from a_{j_1} to a_{j_1} in i_1 steps, is doubly scattered at a_{j_1}, and then moves "freely" from a_{j_1} to a_k. This second summation does not appear in the continuum valued case.[2]

7. QUANTUM PROCESSES

A quantum n-chain describes a quantum system as it evolves in n-steps. Roughly speaking, a quantum process describes a system as it evolves in any number of steps. A *quantum process* is a sequence of chains $\{h_j^n\}_{j=0}^n$, $n=1,2,\ldots$ on amplitude spaces (Ω_n, A_n) with the same value space $S=\{a_1,\ldots,a_r\}\subseteq R$ such that the *product rule*

$$\begin{aligned}&\widehat{A}_n\left[(h_n^0)^{-1}(a_{i_0})\cap\cdots\cap(h_n^n)^{-1}(a_{i_n})\right]\\ &\quad=\widehat{A}_j\left[(h_0^j)^{-1}(a_{i_0})\cap\cdots\cap(h_j^j)^{-1}(a_{i_j})\right]\\ &\quad\quad\times\widehat{A}_{n-j}\left[(h_0^{n-j})^{-1}(a_{i_j})\cap\cdots\cap(h_{n-j}^{n-j})(a_{i_n})\right]\end{aligned}$$

holds for every $n, i_0,\ldots,i_n$, $1\leqslant j\leqslant n$. The next theorem shows that a quantum process consists of a sequence of quantum chains. [By convention $A_0(j,k)=\delta_{jk}$.]

Theorem 11: If $\{h_j^n\}_{j=0}^n$, $n=1,2,\ldots$ is a quantum process, then the n-chain $\{h_j^n\}_{j=0}^n$ is a quantum n-chain with vector $\psi_0\equiv1$ and matrix $A_1(j,k)$. Moreover,

$$A_n(k,j)=\sum_i A_m(i,j)A_{n-m}(k,i)$$

for every $j,k=1,\ldots,r$, and $m=0,\ldots,n$. Conversely, if $\{h_j^n\}_{j=0}^n$, $n=1,2,\ldots$ is a sequence of quantum chains with the same value space, vector $\psi_0\equiv1$, and matrix T, then this sequence is a quantum process.

Proof: We prove this using induction on n. For $n=1$, we have

$$\widehat{A}_1\left[(h_0^1)^{-1}(a_{i_0})\cap(h_1^1)^{-1}(a_{i_1})\right]=A_1(i_1,i_0).$$

Assume the result holds for $n=k-1$. Then

$$\begin{aligned}&\widehat{A}_k\left[(h_0^k)^{-1}(a_{i_0})\cap\cdots\cap(h_k^k)^{-1}(a_{i_k})\right]\\ &\quad=\widehat{A}_{k-1}\left[(h_0^{k-1})^{-1}(a_{i_0})\cap\cdots\cap(h_{k-1}^{k-1})^{-1}(a_{i_{k-1}})\right]\\ &\quad\quad\times\widehat{A}_1\left[(h_0^1)^{-1}(a_{i_{k-1}})\cap(h_1^1)^{-1}(a_{i_k})\right]\\ &\quad=A_1(i_1,i_0)\cdots A_1(i_{k-1},i_{k-2})A_1(i_k,i_{k-1}).\end{aligned}$$

This completes the proof by induction. For the next part we let A_1 be the matrix with components $A_1(j,k)$ and apply (13a) to obtain

$$\begin{aligned}A_n(k,j)=A_1^n(k,j)&=\sum_i A_1^{n-m}(k,i)A_1^m(i,j)\\ &=\sum_i A_m(i,j)A_{n-m}(k,i).\end{aligned}$$

The converse follows from

$$\begin{aligned}&\widehat{A}_n\left[(h_0^n)^{-1}(a_{i_0})\cap\cdots\cap(h_n^n)^{-1}(a_{i_n})\right]\\ &\quad=\left[T(i_1,i_0)\cdots T(i_j,i_{j-1})\right]\left[T(i_{j+1},i_j)\cdots T(i_n,i_{n-1})\right]\\ &\quad=\widehat{A}_j\left[(h_0^j)^{-1}(a_{i_0})\cap\cdots\cap(h_j^j)^{-1}(a_{i_j})\right]\\ &\quad\quad\times\widehat{A}_{n-j}\left[(h_0^{n-j})^{-1}(a_{i_j})\cap\cdots\cap(h_{n-j}^{n-j})(a_{i_n})\right].\end{aligned}$$ ■

Let $\{h_j^n\}_{j=0}^n$, $n=1,2,\ldots$ be a quantum process with value space S, and let $v: S\to R$. Let e_j, $j=1,\ldots,r$ be the standard basis for C^r. Applying Theorems 7 and 11, we have

$$\begin{aligned}A_n^v(k,j)&=\sum_{a_j}^{a_k}\exp\left[-i\sum_m v(h_m^n(\omega))\right]A_n(\omega)\\ &=e^{-iv(a_j)}\langle[e^{-iv(Q)}A_1]^n e_j,e_k\rangle.\end{aligned}$$

The phase factor $e^{-iv(a_j)}$ is inessential since it does not affect the probability $|A_n^v(k,j)|^2$. For a closed system, we have $A_1=e^{-iH_0}$, where H_0 is the "free" Hamiltonian, and then

$$\langle[e^{-v(Q)}e^{-iH_0}]^n e_j,e_k\rangle=e^{iv(a_j)}A_n^v(k,j). \tag{16}$$

Let us now make a scale change in time. We replace $A_1=e^{-iH_0}$ by $A_1(\tau)=e^{-i\tau H_0}$ and replace $e^{-iv(Q)}$ by $e^{-i\tau v(Q)}$, where $\tau\in R$ is fixed. We denote the corresponding quantum process by $\{h_j^n(\tau)\}_{j=0}^n$, $n=1,2,\ldots$. This process is interpreted as giving the location of the system at the time steps $0,\tau,2\tau,\ldots$. Equation (16) then becomes

$$\begin{aligned}&\langle[e^{-i\tau v(Q)}e^{-i\tau H_0}]^n e_j,e_k\rangle\\ &\quad=e^{i\tau v(a_j)}\sum_{a_j}^{a_k}\exp\left\{-i\tau\sum_m v[h_m^n(\tau)](\omega)\right\}A_n(\omega).\end{aligned} \tag{17}$$

For continuous time t, the perturbed Hamiltonian is taken to be $H_0+v(Q)$ and the evolution is given by $\exp(-it(H_0+v(Q)))$. The next result gives a Feynman formula[2] for the finite-dimensional case.

Theorem 12: For every $t\in R$ we have

$$\begin{aligned}&\langle e^{-it(H_0+v(Q))}e_j,e_k\rangle\\ &\quad=\lim_{n\to\infty}\sum_{a_j}^{a_k}\exp\left\{-i\frac{t}{n}\sum_m v\left[h_m^n\left(\frac{t}{n}\right)\right](\omega)\right\}A_n(\omega).\end{aligned}$$

Proof: Applying (17) with $\tau=t/n$ and Trotter's formula gives

$$\lim_{n\to\infty} \sum_{a_j}^{a_k} \exp\left\{ -i\frac{t}{n}\sum_m v\left(h_m^n\left(\frac{t}{n}\right)\right)(\omega)\right\} A_n(\omega)$$
$$= \lim_{n\to\infty} e^{-it(a_j)/n}\langle [e^{-itv(Q)/n}e^{-itH_0/n}]^n e_j, e_k\rangle$$
$$= \langle e^{-it(H_0+v(Q))}e_j,e_k\rangle. \qquad ■$$

8. PATH CHAINS

In this section we show, among other things, that quantum n-chains exist with any vector and matrix. Let $S=\{a_1,...,a_r\}\subseteq R$ and let $P_n = \{p:\{0,1,...,n\}\to S\}$. We call the functions in P_n *n-paths*. Notice that $|P_n| = r^{n+1}$. Let $\psi_0 \in C^r$ and let T be an $r\times r$ complex matrix. We then call (P_n, ψ_0, T) a *path space*. Define the map $I:S\to\{1,...,r\}$ by $Ia_j = j$. For $p\in P_n$ define

$$A_{\psi_0,T}(p) = \psi_0[Ip(0)]T[Ip(1),Ip(0)]\cdots T[Ip(n),Ip(n-1)]. \tag{18}$$

Then $(P_n, A_{\psi_0,T})$ becomes an amplitude space. If $\|\psi_0\| = 1$ and T is a stochastic amplitude matrix, the $(P_n, A_{\psi_0,T})$ is a point amplitude space. Define the *path n-chain* $\{h_j\}_0^n$ on $(P_n, A_{\psi_0,T})$ by $h_j(p) = p(j)$, $j=0,...,n$. Then the joint amplitude of $\{h_j\}_0^n$ becomes

$$\hat{A}_{\psi_0,T}[h_0^{-1}(a_{i_0})\cap\cdots\cap h_m^{-1}(a_{i_n})]$$
$$= A_{\psi_0,T}[(a_{i_0},...,a_{i_n})]$$
$$= \psi_0(i_0)T(i_1,i_0)\cdots T(i_n,i_{n-1}),$$

where $(a_{i_0},...,a_{i_n})$ denotes the path $p(j) = a_{i_j}$, $j=0,...,n$. Hence, $\{h_j\}_0^n$ is a quantum n-chain with vector ψ_0 and matrix T. We thus see that quantum n-chains with arbitrary vectors and matrices exist.

Let $\{h_j\}_0^n$ be an n-chain on an arbitrary amplitude space (Ω,A) with value space S. For each $\omega\in\Omega$, the sample path $p_\omega(j) = h_j(\omega)$ is an n-path $p_\omega:\{0,...,n\}\to S$. Notice that $\{h_j\}_0^n$ is separating if and only if the map $\omega\to p_\omega$ from Ω to P_n is injective. We call $\{h_j\}_0^n$ *conclusive* if the map $\omega\to p_\omega$ from Ω to P_n is bijective. We say that an *an*-chain $\{h_j\}_0^n$ on (Ω,A) is *isomorphic* to an n-chain $\{h'_j\}_0^n$ on (Ω',A') if there exists a bijection $K:\Omega\to\Omega'$ such that $A'(K\omega) = A(\omega)$ and $h'_j(K\omega) = h_j(\omega)$ for all $\omega\in\Omega$ and $j=0,...,n$. This is stronger than h_j and h'_j having the same joint amplitudes, $j=0,...,n$, which is called *stochastic equivalence.*

Lemma 13: An n-chain is a conclusive quantum n-chain if and only if it is isomorphic to a path n-chain.

Proof: Let $\{h_j\}_0^n$ be a conclusive quantum n-chain on (Ω,A) with vector ψ_0 and matrix T. Define $K:\Omega\to P_n$ by $K\omega = (h_0(\omega),...,h_n(\omega))$. Since $\{h_j\}_0^n$ is conclusive, K is bijective. Define $A_{\psi_0,T}:P_n\to C$ by (18). Then

$$A_{\psi_0,T}(K\omega) = A_{\psi_0,T}(h_0(\omega),....,h_n(\omega))$$
$$= \psi_0[Ih_0(\omega)]T[Ih_1(\omega),Ih_0(\omega)]\cdots T$$
$$\times[Ih_n(\omega),Ih_{n-1}(\omega)]$$
$$= \hat{A}[h_0^{-1}(h_0(\omega))\cap\cdots\cap h_n^{-1}(h_n(\omega))] = A(\omega).$$

Finally, if $\{h'_j\}_0^n$ is the pack n-chain on $(P_n, A_{\psi_0,T})$, then

$$h'_j(K\omega) = h'_j(h_0(\omega),...,h_n(\omega)) = h_j(\omega).$$

It follows that $\{h_j\}_0^n$ is isomorphic to the path n-chain $\{h'_j\}_0^n$. Since every path n-chain is a conclusive quantum n-chain, the converse follows easily. ■

In the rest of this section we shall study path n-chains (equivalently, conclusive quantum n-chains). Let $P'_n = \{\phi:P_n\to C\}$. Then P'_n is a complex vector space under pointwise addition and scalar multiplication of dimension r^{n+1}. Define an inner product on P'_n by

$$\langle\phi_1,\phi_2\rangle = \sum_{p\in P_n}\phi_1(p)\bar{\phi}_2(p).$$

For $p\in P_n$, define $p'\in P'_n$ by $p'(q) = \delta_{pq}$, $q\in P_n$. Then $\{p:p\in P_n\}$ is an orthonormal basis for the Hilbert space P'_n. For $\phi\in P'_n$ we have $\langle\phi,p'\rangle = \phi(p)$, $p\in P_n$. Hence, for every $\phi\in P'_n$

$$\phi = \sum_{p\in P_n}\langle\phi,p'\rangle p' = \sum_{p\in P_n}\phi(p)p'.$$

Let $V = C^r$ and for $p\in P_n$ define $Jp'\in V^{\otimes(n+1)}$ by

$$Jp' = e_{Ip(0)}\otimes\cdots\otimes e_{Ip(n)}.$$

Extending by linearity, we obtain the unitary transformations $J:P'_n\to V^{\otimes(n+1)}$ given by

$$J\phi = \sum_{p\in P_n}\phi(p)Jp'. \tag{19}$$

The above construction is used to prove the following theorem.

Theorem 14: Let $\{h_j\}_0^n$ be a conclusive quantum n-chain on (Ω,A) and let $V=C^r$. Then there exists a map $L:P(\Omega)\to V^{\otimes(n+1)}$ such that $\{L\{\omega\}:\omega\in\Omega\}$ is an orthonormal basis for $V^{\otimes(n+1)}$ and for every $E\in P(\Omega)$, $L(E) = \Sigma_{\omega\in E}L\{\omega\}$. There exists a vector $\bar{A}\in V^{\otimes(n+1)}$ and vectors $\bar{h}_j\in V^{\otimes(n+1)}$, $j=0,...,n$ such that $\hat{A}(E) = \langle\bar{A},L(E)\rangle$ for every $E\in P(\Omega)$ and $h_j(\omega) = \langle\bar{h}_j,L\{\omega\}\rangle$. If ψ_0,T are the vector and matrix of $\{h_j\}_0^n$, respectively, then

$$\bar{A} = \sum_{i_0,...,i_n=1}\psi_0(i_0)T(i_1,i_0)\cdots T(i_n,i_{n-1})e_{i_0}\otimes\cdots\otimes e_{i_n}$$

and

$$\check{h}_j = \sum_{i_0,...,i_n=1}a_{i_j}e_{i_0}\otimes\cdots\otimes e_{i_n}.$$

Proof: Applying Lemma 13, there exists a path chain $\{h'_j\}_0^n$ on $(P_n, A_{\psi_0,T})$ and a bijection $K:\Omega\to P_n$ such that $A(\omega) = A_{\psi_0,T}(K\omega)$ and $h_j(\omega) = h'_j(K\omega)$. For $\omega\in\Omega$ define $L\{\omega\} = J(K\omega)'$ and for $E\in P(\Omega)$ let $L(E) = \Sigma_{\omega\in E}L\{\omega\}$. It is clear that $\{L\{\omega\}:\omega\in\Omega\}$ is an orthonormal basis for $V^{\otimes(n+1)}$. Define the vectors $\bar{A},\bar{h}_j\in V^{\otimes(n+1)}$, $j=0,...,n$ by $\bar{A} = JA_{\psi_0,T}$ and $\bar{h}_j = Jh'_j$. Since J is unitary, we have

$$\langle\bar{A},L(E)\rangle = \sum_{\omega\in E}\langle\bar{A},L\{\omega\}\rangle = \sum_{\omega\in E}\langle JA_{\psi_0,T},J(K\omega)'\rangle$$
$$= \sum_{\omega\in E}\langle A_{\psi_0,T},(K\omega)'\rangle = \sum_{\omega\in E}A_{\psi_0,T}(K\omega)$$
$$= \sum_{\omega\in E}A(\omega) = \hat{A}(E)$$

and

$$\langle \bar{h}_j, L\{\omega\}\rangle = \langle Jh'_j, J(K\omega)'\rangle = \langle h'_j, (K\omega)'\rangle$$
$$= h'_j(K\omega) = h_j(\omega).$$

For the last part, using (19) gives

$$\bar{A} = JA_{\psi_0,T} = \sum_{p\in P_n} A_{\psi_0,T}(p)Jp'$$
$$= \sum_{p\in P_n} \psi_0[Ip(0)]T[Ip(1),Ip(0)]$$
$$\cdots T[Ip(n),Ip(n-1)]e_{Ip(0)}\otimes\cdots\otimes e_{Ip(n)}$$
$$= \sum_{i_0,\ldots,i_n=1} \psi_0(i_0)T(i_1,i_0)\cdots T(i_n,i_{n-1})e_{i_0}\otimes\cdots\otimes e_{i_n}$$

and

$$\bar{h}_j = Jh'_j = \sum_{p\in P_n} h'_j(p)Jp' = \sum_{p\in P_n} p(j)e_{Ip(0)}\otimes\cdots\otimes e_{Ip(n)}$$
$$= \sum_{i_0,\ldots,i_n=1} a_{i_j}e_{i_0}\otimes\cdots\otimes e_{i_n}. \qquad \blacksquare$$

For $p\in P_n$, define $C_i(p) = |p^{-1}(a_i)|$, $i = 1,\ldots r$. Notice that $\Sigma_{i=1}^r C_i(p) = n+1$. Let $C(p) = (C_1(p),\ldots,C_r(p))$ and write $p\sim q$ if $C(p) = C(q)$. Then $\sim$ is an equivalence relation. We say that $\phi\in P'_n$ is *symmetric* if $p\sim q$ implies that $\phi(p) = \phi(q)$. Notice if $p\in P_n$, then p' is symmetric if and only if $C_j(p) = n+1$ for some j. Denote the symmetric elements of P'_n by P^s_n and denote the equivalence class containing $p\in P_n$ by $[p]$. Notice that $|[p]| = (n+1)!/C_1(p)!\cdots C_r(p)!$. Denote the set of equivalence classes by $[P_n] = \{[p_1],\ldots,[p_s]\}$.

Theorem 15: (a) $\phi\in P^s_n$ if and only if $\phi = \Sigma_{j=1}^s \phi(p_j)\Sigma_{p\in[p_j]}p'$.

(b) $\phi\in P^s_n$ if and only if $J\phi\in V^{(n+1)}$, where Ⓢ denotes the symmetric tensor product.

Proof: (a) If $\phi = \Sigma_{j=1}^s \phi(p_j)\Sigma_{p\in[p_j]}p'$ and $g_1\sim g_2$ then $g_1, g_2\in[p_j]$ for some j. Hence, $\phi(g_1) = \phi(g_2) = \phi(p_j)$. Conversely, suppose that $\phi\in P^s_n$. Then

$$\phi = \sum_{p\in P_n}\phi(p)p' = \sum_{j=1}^s\sum_{p\in[p_j]}\phi(p)p'$$
$$= \sum_{j=1}^s \phi(p_j)\sum_{p\in[p_j]}p'.$$

(b) If $\phi\in P^s_n$, then by (a) we have

$$J\phi = \sum_{j=1}^s \phi(p_j)\sum_{p\in[p_j]}Jp'$$
$$= \sum_{j=1}^s \phi(p_j)|[p_j]|^{1/2}e_{Ip_j(0)}Ⓢ\cdotsⓈe_{Ip_j(n)}.$$

Hence, $J\phi\in V^{Ⓢ(n+1)}$. Conversely, suppose that $J\phi\in V^{Ⓢ(n+1)}$. Then we have

$$J\phi = \sum_{j=1}^s c_je_{Ip_j(0)}Ⓢ\cdotsⓈe_{Ip_j(n)}$$

for some $c_j\in C$, $j = 1,\ldots,s$. Hence,

$$\phi = \sum_{j=1}^s c_j J^{-1}[e_{Ip_j(0)}Ⓢ\cdotsⓈ e_{Ip_j(n)}]$$
$$= \sum_{j=1}^s c_j\frac{1}{|[p_j]|^{1/2}}\sum_{p\in[p_j]}p'$$
$$= \sum_{j=1}^s \phi(p_j)\sum_{p\in[p_j]}p'.$$

It follows from (a) that $\phi\in P^s_n$. $\blacksquare$

Let $(P_n,1,T)$, $n = 1,2,\ldots$ be path spaces with the vector $\psi_0\equiv 1$ and matrix T. The corresponding path chains $\{h^n_j\}_{j=0}^n$, $n = 1,2,\ldots$ then form a quantum process. Using the map J defined above, we can embed $\{(P_{n,1,T}):n = 1,2,\ldots\}$ into the tensor space

$$TV = C\oplus V\oplus V^{\otimes 2}\oplus V^{\otimes 3}\oplus\cdots$$

and the process then works on TV.

[1]R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).
[2]R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[3]L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).
[4]E. W. Montroll, Commun. Pure Appl. Math. **5**, 415 (1952).

Continued fraction expansions for the complete, incomplete, and relativistic plasma dispersion functions

Anthony L. Peratt

Advanced Concepts and Plasma Physics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 1 February 1983; accepted for publication 7 October 1983)

In our investigation of the linear theory of waves in plasma and the stability of relativistic beam-plasma systems, we have been led to consider methods for the evaluation of integrals of the form $\int d\chi(\chi-\zeta)^{-1}\exp(-\chi^2)$ and $\int d\chi(\chi-\zeta)^{-1}\exp[(1-\chi^2)^{-1/2}]$ for complex ζ. In this work, we report on the evaluation of this integral and its derivative by means of continued fraction expansions. The expressions derived allow the precise calculation of these integrals in previously inaccessible regions. Additionally, applications to non-Maxwellian particle distributions, such as those found in the analysis of plasma diodes, are included.

PACS numbers: 02.60.Gf, 02.70. + d, 52.35. − g, 52.60. + h

I. INTRODUCTION

In the linear theory of wave propagation in hot plasma, or in the case of a charged particle beam penetrating a magnetized plasma, one is confronted with integrals of the form

$$\int_{\chi_1}^{\chi_2} d\chi(\chi-\zeta)^{-m}\exp(-\chi^2), \tag{1.1}$$

for integer m. For the plasma wave case, ζ is usually a complex function dependent upon the constituent plasma parameters through the argument[1]

$$\zeta=\zeta_{nj}\equiv\frac{(\omega+i\nu+n\Omega_j)}{(k_r+ik_I)v_{\mathrm{th}_j}}, \tag{1.2}$$

where ω is the wave frequency, ν is a collisional frequency, k is the wave number, Ω is the cyclotron frequency at harmonic n, v_{th} is the thermal velocity, and the subscript j pertains to the particle type or species. Thus, in evaluating Eq. (1.1), one must contend with singularities and resonances associated with either complex ω, complex k, or both.

For the case where the distribution of electrons or ions is Maxwellian, the limits of integration of Eq. (1.1) may be taken as $\pm\infty$, and the integral, when $m=1$, is referred to as the plasma dispersion function $Z(\zeta)$.[2] Tabulated values for the plasma dispersion function and its first derivative, as well as for the related complex error function, $\omega(\zeta)=Z(\zeta)/i\pi^{1/2}$, are available, but only a finite number of the real and imaginary components of ζ can be selected from the tables.[2,3] A significant loss of accuracy can occur when interpolating between tabulated values, especially when $I_m\,\zeta<0$ [where zero's of $Z(\zeta)$ are known to be present]. In addition, some computation may be required to analytically continue tabulated values into other quadrants of the complex ζ plane.

Some steps have been taken to provide approximations to the plasma dispersion function,[4,5] but these generally result in large relative errors. In Sec. II we present continued fraction expansions and algorithms applicable for the quick and precise evaluation of $Z(\zeta)$ over all ζ.

When the distribution of particles is no longer completely Maxwellian, but rather truncated as is the case for plasma diodes, the limits of integration on Eq. (1.1) are χ and $+\infty$. This integral is known as the incomplete plasma dispersion relation.[6] Expansions useful for its calculation are given in Sec. III.

Lastly, in Sec. IV, the analysis is extended to include relativistic plasma supporting waves whose phase velocity is either above or below the speed of light.

II. THE COMPLETE PLASMA DISPERSION FUNCTION

The plasma dispersion function is defined by the integral

$$Z(\zeta)=\pi^{-1/2}\int_{-\infty}^{\infty}d\chi(\chi-\zeta)^{-1}\exp(-\chi^2),\ I_m\,\zeta>0, \tag{2.1}$$

and the analytic continuation of this for $I_m\,\zeta\leqslant 0$. The complex variable ζ, Eq. (1.2), has the physical significance of being the ratio of the phase velocity of a wave in plasma to the plasma thermal velocity.

A. Mathematical analysis of $Z(\zeta)$

We base our analysis of the function $Z(\zeta)$ upon the mathematical properties of the confluent hypergeometric functions, $M(a,b,\zeta)$ and $U(a,b,\zeta)$, by means of the expressions

$$Z(\zeta)=\zeta U(1,\tfrac{3}{2},-\zeta^2) \tag{2.2}$$

and

$$Z(\zeta)=i\pi^{1/2}\exp(-\zeta^2)-2\zeta M(1,\tfrac{3}{2},-\zeta^2). \tag{2.3}$$

Similarly, for the first derivative of $Z(\zeta)$,

$$Z'(\zeta)\equiv\frac{d}{d\zeta}Z(\zeta)=-2(1+\zeta Z), \tag{2.4}$$

we use the expressions[7]

$$Z'(\zeta)=-U(1,\tfrac{1}{2},-\zeta^2) \tag{2.5}$$

and

$$Z'(\zeta)=i2\pi^{1/2}\zeta\exp(-\zeta^2)-2M(1,\tfrac{1}{2},-\zeta^2). \tag{2.6}$$

 0022-2488/84/030466-03$02.50

For points at or near the origin we apply the power series and related continued fraction [Eq. (10.11) in Ref. 8]

$$M(1,c,z) = 1 + \frac{z}{c} + \frac{z^2}{c(c+1)} + \cdots + \frac{z^n}{c(c+1)\ldots(c+n-1)} + \cdots \tag{2.7}$$

and

$$M(1,c,z) = \frac{1}{1-}\,\frac{z}{c+}\,\frac{cz}{c+2+}\,\frac{2z}{c+3-} \cdots \frac{nz}{c+2n-1-}\,\frac{(c-1+n)z}{c+2n+}\cdots. \tag{2.8}$$

Expressions applicable to points away from the origin may be obtained by noting that Eqs. (2.2) and (2.5) are associated with Prym's function.[7] We employ the integral representation associated with $U(a,c,z)$ and Kummer's transformation [Eqs. (13.2.5) and (13.1.29) in Ref. 3], while making use of Eq. (11.17) of Khovanskii,[8] and find

$$U(1,c,z) = \frac{1}{\Gamma(2-c)} \int_0^\infty \frac{e^{-t}t^{1-c}}{z+t}\,dt = \frac{1}{z-c+2-}\,\frac{2-c}{z-c+4-}\,\frac{2(3-c)}{z-c+6-}\cdots \frac{n(n+1-c)}{z-c+2n+2-}\cdots. \tag{2.9}$$

Equation (2.9) immediately provides a contracted continued fraction upon replacement of the variable $z = -\zeta^2$. In addition, the following asymptotic expansion [Eq. (13.4.2), Ref. 3] is associated with Eq. (2.9):

$$U(1,c,z) = -\sum_{n=0}^{\infty} \frac{\Gamma(2-c+n)}{\Gamma(2-c)(-z)^{2n+2}}, \quad -\frac{3}{2}\pi < \arg z < \frac{3}{2}\pi. \tag{2.10}$$

Equations (2.7)–(2.10) form the basis for the algorithms that follow.

B. Description of the algorithms

When $I_m\,\zeta > 0$, one finds, in terms of the continued fraction expansion of Eq. (2.9), that for $c = 3/2$,

$$Z(\zeta) = \frac{a_1}{b_1+}\,\frac{a_2}{b_2+}\,\frac{a_3}{b_3+}\cdots, \tag{2.11}$$

where

$$a_1 = \zeta$$
$$a_{n+1} = (n/2)(1-2n), \quad n = 1,2,3,\ldots$$
$$b_{n+1} = -\zeta^2 + \tfrac{1}{2} + 2n, \quad n = 0,1,2,\ldots. \tag{2.12}$$

Similarly, for $c = \frac{1}{2}$,

$$Z'(\zeta) = \frac{a_1}{b_1+}\,\frac{a_2}{b_2+}\,\frac{a_3}{b_3+}\cdots, \tag{2.13}$$

where

$$a_1 = -1$$
$$a_{n+1} = -(n/2)(1+2n), \quad n = 1,2,3,\ldots$$
$$b_{n+1} = -\zeta^2 + \tfrac{3}{2} + 2n, \quad n = 0,1,2,\ldots. \tag{2.14}$$

Equations (2.11) and (2.13) converge rapidly outside a region linearly approximated by $|x| + 3|y| > 0$, $y \neq 0$ [convergence inside this region is possible with Eqs. (2.11) and (2.13), but only by retaining an increasing number of higher-order terms for arguments approaching the origin]. The continued fractions, (2.11) and (2.13), are evaluated using the recursion relations

$$A_{n+1} = b_{n+1}A_n + a_{n+1}A_{n-1}, \quad A_{-1} = 1,\ A_0 = 0,$$
$$B_{n+1} = b_{n+1}B_n + a_{n+1}B_{n-1}, \quad B_{-1} = 0,\ B_0 = 1, \tag{2.15}$$

and

$$Z(\zeta) = \lim_{n\to\infty} \frac{A_n}{B_n}. \tag{2.16}$$

For $I_m\,\zeta < 0$, we use

$$Z(x-iy) = 2i\pi^{1/2}\exp[-(x-iy)^2] - Z^*(x+iy). \tag{2.17}$$

The algorithms have been tested against programs for the evaluation of the complex error function.[9–11] A relative error of less than 4.2×10^{-8} was maintained.[12]

III. THE INCOMPLETE PLASMA DISPERSION FUNCTION

In situations where charged particles pass through potential barriers, the resulting distribution function, originally a complete Maxwellian, becomes truncated. This is the case for plasma diodes or triodes where ions may flow through grids or potential barriers unimpeded while the electrons may be reflected or trapped.[13] For this case, while the ions are Maxwellian, the limits of integration for the electron distribution may be $\chi = \nu$ and $+\infty$, respectively, and the integral Eq. (1.1) is referred to as the incomplete plasma dispersion function,[6]

$$Z(\nu,\zeta) = \frac{1}{\sqrt{\pi}} \int_\nu^\infty d\chi(\chi-\zeta)^{-1}\exp(-\chi^2). \tag{3.1}$$

It can be shown that Eq. (3.1) satisfies the differential equation

$$Z' + 2\zeta Z - \frac{\exp(-\nu^2)}{\pi^{1/2}(\nu-\zeta)} = \begin{cases} -\operatorname{erfc}(\nu), & \nu > 0, \\ -1-\operatorname{erf}|\nu|, & \nu < 0. \end{cases} \tag{3.2}$$

While Eq. (3.2) can be integrated numerically to obtain $Z(\nu,\zeta)$, difficulties exist because of the logarithmic singularity for $\nu \sim \zeta$ and also because of the resonant denominator when $\chi \sim R_e\zeta$ if $\operatorname{Im}\zeta$ is small and $R_e\,\zeta > \nu$. For these reasons, the following procedures have been used to calculate $Z(\nu,\zeta)$.

Since the complete plasma dispersion function is related to the incomplete function via the expression $Z(\zeta) = Z(-\infty,\zeta)$, then

$$Z(\nu,\zeta) = Z(\zeta) - \pi^{-1/2}\int_{-\infty}^{\nu} d\chi(\chi-\zeta)^{-1}\exp(-\chi^2), \tag{3.3}$$

where $Z(\zeta)$ is obtained from Sec. II and the second term in Eq. (3.3) has no resonant denominator if $R_e\,\zeta < \nu$. This important case is applicable to the study of slow waves in plasma diodes. Secondly, for the study of fast electromagnetic waves propagating within a nonrelativistic plasma whose particle distribution has become truncated, ζ is large and an asymptotic solution for $Z(\nu, \zeta)$ is readily found in terms of the chi-square probability function Q,

$$Z(\nu,\zeta) = \frac{-1}{2\zeta\pi^{1/2}} \sum_{l=1}^{\infty} \frac{\nu^l \exp(-\nu^2)}{\zeta^{(l-1)}} \times \left\{ \frac{1}{\nu^2+} \frac{1-l/2}{1+} \frac{1}{\nu^2+} \frac{2-l/2}{1+} \frac{2}{\nu^2+\cdots} \right\}. \tag{3.4}$$

This expansion is easily used via the algorithms given in Sec. II by simply replacing the coefficients Eq. (2.12) with those of Eq. (3.4) and carrying out the indicated summation.

IV. RELATIVISTIC PLASMA DISPERSION FUNCTION

For relativistic plasma, a Juttner–Synge distribution function may be used to describe the particle placement. The integral obtained using this distribution defines the relativistic plasma dispersion function,[14]

$$T(z,\zeta) \equiv \int_{-1}^{1} dv (v-z)^{-1} \exp[-\zeta(1-v^2)^{-1/2}], \quad \mathrm{Im}(\zeta) > 0. \tag{4.1}$$

The parameter $\zeta \equiv mc^2/k_B T$ represents the inverse of the plasma temperature normalized to the particle rest energy while z is defined as the Langmuir wave phase velocity,

$$z = \omega/kc, \tag{4.2}$$

where c is the speed of light.

Analytic extension into the lower-half complex z plane through the real axis, for $|z| > 1$, is possible with

$$T(z^*,\zeta) = T^*(z,\zeta). \tag{4.3}$$

For $|z| < 1$, a residue term due to the first-order pole in the integrand of Eq. (4.1) is acquired,

$$T(z^*,\zeta) = T^*(z,\zeta) - \{2\pi i \exp[-\zeta(1-z^2)^{-1/2}]\}^*. \tag{4.4}$$

A power series about $z = 0$, valid in the circle $|z| < 1$, has been derived,[14]

$$T(z,\zeta) = \sum_{n=0}^{\infty} a_n z^{2n+1} + i\pi e^{-\zeta(1-z^2)^{-1/2}}, \tag{4.5}$$

where

$$\begin{aligned} a_0 &= -2\zeta K_1(\zeta), \\ a_1 &= \tfrac{2}{3}[\zeta^2 K_0(\zeta) - \zeta K_1(\zeta)], \\ a_n &= [6(n-1)(2n-1)a_{n-1} + 3(\zeta^2-(2n-3)^2)a_{n-2} \\ &\quad + 2(n-2)(2n-5)a_{n-3}]/(4n^2-1), \quad n > 1, \end{aligned} \tag{4.6}$$

where K_0 and K_1 are modified Bessel functions.

For $|z| > 1$ there exists a series in z^{-1},

$$\begin{aligned} T(z,\zeta) &= \sum_{n=0}^{\infty} a_n z^{-2n-1}, \\ a_0 &= -2Ki_2(\zeta), \\ a_1 &= -\tfrac{1}{3}\zeta^2 K_0(\zeta) - \zeta K_1(\zeta) + \tfrac{1}{6}(3-\zeta^2)a_0, \\ a_2 &= -\tfrac{1}{10}\zeta K_1(\zeta) - \tfrac{3}{10}a_0 + \tfrac{1}{20}(27-\zeta^2)a_1, \\ a_n &= \{[3(2n-1)^2 - \zeta^2]a_{n-1} - 6(n-1)(2n-3)a_{n-2} \\ &\quad + (2n-3)(2n-5)a_{n-3}\}/2n(2n+1), \quad n > 2. \end{aligned} \tag{4.7}$$

Additionally, $T(z, \zeta)$ can be expanded about the branch point $z = 1$, yielding the continued fraction

$$T(z,\zeta) = \frac{-2K_0}{1+} \frac{-2K_2 v/K_0}{1+2K_2 v/K_0+} \times \frac{-K_4 v/K_2}{1+K_4 v/K_2+} \frac{-K_6 v/K_4}{1+K_8 v/K_6+} \cdots + i\pi\sigma \exp(-\zeta)(1-z^2)^{-1/2}, \tag{4.8}$$

for $\zeta \neq 0$ and $v = (1-z)/(3-z)$ about $z = 1$. The quantity σ is given by

$$\sigma = \begin{cases} 0 & z_I > 0;\ z_I = 0,\ |z_R| \geqslant 1 \\ 1 & z_I = 0,\ |z_R| < 1 \\ 2 & z_I < 0 \end{cases}. \tag{4.9}$$

The discontinuity across the real axis $|z| < 1$ is due to the Stokes phenomena as is also the case for the complete and incomplete plasma dispersion functions.

Equation (4.8) simplifies considerably the numerical evaluation of the relativistic plasma dispersion function near the branch points $z = \pm 1$.[14]

ACKNOWLEDGMENTS

The author acknowledges discussions with H. Derfler and T. C. Simonen and valuable comments from B. B. Godfrey. The author acknowledges the assistance of Nathana Haines in formatting and editing this report. This work was supported by the U. S. Department of Energy.

[1] A. L. Peratt, Phys. Fluids **18**, 57 (1975).
[2] B. D. Fried and S. D. Conte, *The Plasma Dispersion Function* (Academic, New York, 1961).
[3] W. Gautschi, in *Handbook of Mathematical Functions*, National Bureau of Standards Applied Mathematics Series No. 55 (U.S. GPO, Washington, D. C., 1964).
[4] B. D. Fried, C. L. Hedrick, and J. McCune, Phys. Fluids **11**, 249 (1968).
[5] A. L. Brinca, J. Plasma Phys. **10**, 123 (1973).
[6] R. N. Franklin, *Tenth International Conference on Phenomena in Ionized Gases*, Oxford, Sept. 13–18, 1971 (Donald Parsons, Oxford, 1971).
[7] H. Derfler and T. C. Simonen, Phys. Fluids **12**, 269 (1969).
[8] A. N. Khovanskii, *The Application of Continued Fractions and Their Generalizations to Problems in Approximation Theory* (Noordhoff, Groningen, 1963), pp. 139–148.
[9] W. Gautschi, SIAM J. Numer. Anal. 7, 187 (1970).
[10] W. Gautschi, Comm. ACM **12**, 635 (1969).
[11] K. S. Kolbig, Comm. ACM **15**, 465 (1972).
[12] A. L. Peratt, University of California, Lawrence Livermore Laboratory Report, UCRL-76705, 1975.
[13] A. L. Peratt, R. L. Watterson, and H. Derfler, Phys. Fluids **20**, 1900 (1977).
[14] B. B. Godfrey, B. S. Newberger, and K. A. Taggart, IEEE Trans. Plasma Sci. **PS-3**, 60 (1975).

Scattering by a penetrable body

A. G. Ramm
Mathematics Department, Kansas State University, Manhattan, Kansas 66506

(Received 23 November 1982; accepted for publication 11 February 1983)

An integral equation equivalent to the interface problem is derived. A numerical scheme for its solution is given. Convergence of the scheme is established.

PACS numbers: 02.60.Lj

1. INTRODUCTION

Consider the scattering problem:

$$(\nabla^2 + k_0^2)u = 0 \quad \text{in } \Omega, \quad k_0 > 0, \tag{1}$$

$$(\nabla^2 + k_1^2)u = 0 \quad \text{in } \mathscr{D}, \quad k_1 > 0, \tag{2}$$

$$u_+ = u_- \quad \text{on } \Gamma, \tag{3}$$

$$\rho \frac{\partial u_+}{\partial N} = \frac{\partial u_-}{\partial N} \quad \text{on } \Gamma, \tag{4}$$

$$u = u_0 + v, \tag{5}$$

$$\int_{|s| = R} \left| \frac{\partial v}{\partial r} - ik_0 v \right|^2 ds \to 0, \quad R \to \infty. \tag{6}$$

Here u_0 is the incident field which satisfies Eq. (1) in the whole space $\mathbb{R}^3$, $\mathscr{D}$ is a bounded obstacle with a smooth surface Γ, $\mathscr{D}$ is the interior domain, N is the exterior unit normal on Γ pointing into Ω, the exterior domain, k_0 (k_1) is the wave number in Ω ($\mathscr{D}$), $\rho = \text{const} > 0, \rho \neq 1$, the sign $+$ $(-)$ denotes the limit value on Γ from the interior (exterior) domain

$$\frac{\partial u_+}{\partial N} \equiv \left(\frac{\partial u}{\partial N}\right)_+.$$

There is an extensive literature on the exterior boundary value problem. The integral equation method is usually the tool of the studies.[1] The equivalence problem is important in these studies. For the Dirichlet and Neumann boundary conditions (corresponding to scattering by acoustically soft and hard obstacles), the integral equations obtained are not equivalent to the boundary value problems when k_0^2 belongs to some discrete set (the spectrum of the corresponding interior problems). Various ways to modify the integral equations at these exceptional values of k_0^2 were suggested by many authors. We will not discuss this question here since for the problem (1)–(6), the equivalence of the integral equation and the problem (1)–(6) is easy to establish. In Sec. 2 the uniqueness theorem is proved. This theorem is known,[1] but we included a very short proof of it for convenience of the reader. In Sec. 3 existence and uniqueness of the solution to the integral Eq. (17) and the equivalence of this equation to the problem (1)–(6) are proved. In Sec. 4 convergence of a numerical method of solving the integral equation is proved. This result is connected with the T-matrix approach.[2] The integral Eq. (17), which is used in this paper, is of Fredholm's type and is convenient for a numerical treatment. One can derive a boundary integral equation of the type used in the usual T-matrix scheme, but in this equation one has integral operators with strong singularities, and this fact makes the theoretical numerical analysis difficult.

Scattering by a permeable body was discussed recently in Ref. 3, where different integral equations were suggested. Numerical solution of these equations (which involve improper integrals) was not discussed. The basic integral equation in Ref. 3 is of the first kind and its kernel is weakly singular. Thus this equation presents difficulties from the numerical analysis viewpoint.

2. UNIQUENESS OF THE SOLUTION

Theorem 1: If $u_0 = 0$, then the only solution to problem (1)–(6) is $u \equiv 0$.

Proof: If $u_0 = 0$, then u satisfies the radiation condition (6) and $\bar{u}$ solves (1)–(4). Here and below the bar denotes complex conjugation. From Green's formula, it follows that

$$0 = \lim_{R \to \infty} \int_{|s| = R} \left(u \frac{\partial \bar{u}}{\partial r} - \bar{u} \frac{\partial u}{\partial r}\right) ds - \int_\Gamma \left(u \frac{\partial \bar{u}_-}{\partial N} - \bar{u} \frac{\partial u_-}{\partial N}\right) ds. \tag{7}$$

Applying (4) and Green's formula again, one obtains

$$\int_\Gamma \left(u \frac{\partial \bar{u}_-}{\partial N} - \bar{u} \frac{\partial u_-}{\partial N}\right) ds = \rho \int_\Gamma \left(u \frac{\partial \bar{u}_+}{\partial N} - \bar{u} \frac{\partial u_+}{\partial N}\right) ds = 0. \tag{8}$$

From Eq. (7) and (8) it follows that

$$\lim_{R \to \infty} \int_{|s| = R} \left(u \frac{\partial \bar{u}}{\partial r} - \bar{u} \frac{\partial u}{\partial r}\right) ds = 0. \tag{9}$$

Condition (6) for u and (9) yield

$$\lim_{R \to \infty} \int_{|s| = R} \left\{ \left| \frac{\partial u}{\partial r} \right|^2 + k_0^2 |u|^2 \right\} ds = 0. \tag{10}$$

From (1) and (10) it follows that $u \equiv 0$. The last conclusion is a well-known result. A short proof of this result can be found in Refs. 4 and 5.

3. BASIC INTEGRAL EQUATION, ITS EQUIVALENCE TO THE PROBLEM (1)–(6). EXISTENCE AND UNIQUENESS OF ITS SOLUTION

Let us rewrite (2) as

$$(\nabla^2 + k_0^2)u = \kappa u, \quad \kappa \equiv k_0^2 - k_1^2. \tag{2'}$$

From the Green's formula, it follows that

 0022-2488/84/030469-03$02.50

$$\int\{g(\nabla^2+k_0^2)u+u\delta(x-y)\}dy$$

$$=\kappa\int_{\mathscr{D}} gu\,dy+u(x)$$

$$=\lim_{R\to\infty}\int_{|s|=R}\Big(g\frac{\partial u}{\partial r}-u\frac{\partial g}{\partial r}\Big)ds-\int_\Gamma\Big(g\frac{\partial u_-}{\partial N}-u_-\frac{\partial g}{\partial N}\Big)ds+\int_\Gamma\Big(g\frac{\partial u_+}{\partial N}-u_+\frac{\partial g}{\partial N}\Big)ds$$

$$=u_0(x)+\int_\Gamma g\Big(\frac{\partial u_+}{\partial N}-\frac{\partial u_-}{\partial N}\Big)ds$$

$$=u_0(x)+(1-\rho)\int_\Gamma g\frac{\partial u_+}{\partial N}ds,\quad x\notin\Gamma,$$

$$g\equiv\frac{\exp(k_0|x-y|)}{4\pi|x-y|}. \tag{11}$$

This can be written as

$$u(x)=u_0(x)-\kappa Tu+(1-\rho)Q\sigma, \tag{12}$$

$$\sigma\equiv\frac{\partial u_+}{\partial N}, \tag{13}$$

$$Tu=\int_{\mathscr{D}}gu\,dy,\quad Q\sigma=\int_\Gamma g\sigma\,ds. \tag{14}$$

For any σ, any function u which solves (12) solves (1), (2), (3), (5), and (6). This function will solve (4) iff

$$0=(\rho-1)\frac{\partial u_0}{\partial N}-\kappa(\rho-1)\frac{\partial Tu}{\partial N}+(1-\rho)\Big(\rho\frac{A\sigma+\sigma}{2}-\frac{A\sigma-\sigma}{2}\Big)$$

or, which is the same,

$$\sigma=-\frac{2\kappa}{\rho+1}\frac{\partial Tu}{\partial N}+\frac{1-\rho}{1+\rho}A\sigma+\frac{2}{\rho+1}\frac{\partial u_0}{\partial N}, \tag{15}$$

where

$$A\sigma=2\int_\Gamma\frac{\partial g(s,s')}{\partial N_s}\sigma(s')ds',$$

and the known formulas were used:

$$\Big(\frac{\partial Q\sigma}{\partial N}\Big)_\pm=\frac{A\sigma\pm\sigma}{2}. \tag{16}$$

It is easy to check that (15) is equivalent to (13) if one takes as $u(x)$ in (13) the right-hand side of (12). Equations (12) and (15) can be written as

$$w=Bw+h, \tag{17}$$

where

$$w=\begin{pmatrix}u\\ \sigma\end{pmatrix},\quad B=\begin{pmatrix}-\kappa T & (1-\rho)Q\\ -\frac{2\kappa}{\rho+1}\frac{\partial}{\partial N}T & \frac{1-\rho}{1+\rho}A\end{pmatrix},$$

$$h=\begin{pmatrix}u_0\\ \frac{2}{\rho+1}\frac{\partial u_0}{\partial N}\end{pmatrix}. \tag{18}$$

Equation (17) is equivalent to the equation

$$u=u_0-\kappa Tu+(1-\rho)Q\frac{\partial u_+}{\partial N}. \tag{19}$$

Let us consider B as an operator from $H^q=H_q\oplus\widetilde{H}_{q-1/2}$ into H^q. Here $H_q=W_2^q(\mathscr{D})$, $q\geqslant 0$ is the Sobolev space of functions which are q times differentiable and their derivatives belong to $L^2(\mathscr{D})$, $\widetilde{H}_q=W_2^q(\Gamma)$. It is known[6] that the imbedding $H_{q'}\to\widetilde{H}_q$ is continuous if $q'\geqslant q+\frac{1}{2}$ and compact if $q'>q+\frac{1}{2}$. For $q<0$ the space H_q is dual to the space $H_{|q|}$ with respect to H_0. Symbol $\oplus$ means that any element $w\in H^q$ is uniquely representable as an ordered pair $\binom{u}{\sigma}$, where $u\in H_q$, $\sigma\in\widetilde{H}_{q-1/2}$, and the inner product in H^q is defined as $(w_1,w_2)=(u_1,u_2)_{H_q}+(\sigma_1,\sigma_2)_{\widetilde{H}_{q-1/2}}$.

Lemma 1: The operator $B:H^q\to H^q$ is compact.

Proof: This follows from the relations: $T:H_q\to H_{q+2}$ is continuous, $Q:\widetilde{H}_q\to H_{q+3/2}$ is continuous, $(\partial/\partial N)T:H_q\to H_{q+1}$ is continuous, $A:\widetilde{H}_q\to\widetilde{H}_{q+1}$ is continuous, and from the compactness of the embeddings: $H_{q'}\to H_q$ if $q'>q$, $H_{q'}\to\widetilde{H}_q$ if $q'>q+\frac{1}{2}$.

Lemma 2: Equation (17) and problem (1)–(6) are equivalent.

Proof: 1). (1)–(6)$\Rightarrow$(17). This was shown above in the process of deriving Eq. (17). 2). (17)$\Rightarrow$(1)–(6). If w satisfies (17), then (12) and (15) are satisfied. If u satisfies (12), then u solves (1), (2), (3), (5), and (6). If (15) holds, then $\sigma=\partial u_+/\partial N$ and (4) holds.

Lemma 3: Equation (17) has no more than one solution.

Proof: Equation (17) is equivalent to (1)–(6), and (1)–(6) has no more than one solution (by Theorem 1).

Theorem 2. If $h\in H^q$, then Eq. (1) has a solution in H^q and this solution is unique.

Proof: Theorem 2 follows from Lemmas 1 and 3, Fredholm's alternative, and the inclusion $h\in H^q$.

4. NUMERICAL SOLUTION

Since B is compact in H^0, the convergence of the projection method of solving Eq. (17) is easy to establish (Ref. 4, p. 192). Let us describe the projection method for Eq. (17). Let $\{\phi_j\}$ be a complete linearly independent system of functions in H_0, and $\{\psi_j\}$ be a similar system in $\widetilde{H}_{-1/2}$. The union of the systems

$$\begin{Bmatrix}\phi_j\\ 0\end{Bmatrix},\begin{Bmatrix}0\\ \psi_j\end{Bmatrix}$$

is a complete linearly independent system in H^0. Let us take

$$\psi_j=\Big(\frac{\partial\phi_j}{\partial N}\Big)_+.$$

As $\{\phi_j\}$, let us take the orthonormal system of eigenfunctions of the Dirichlet Laplacian in domain Δ, $\mathscr{D}\subset\Delta$. As Δ one can take, e.g., box or a ball, so that $\{\phi_j\}$ is given explicitly. The system $\{\phi_j\}$ is complete in H_0. The system $\{\partial\phi_j/\partial N\}$ is complete in $\widetilde{H}_{-1/2}$. Indeed, let

$$(*)\int_\Gamma f\overline{\frac{\partial\phi_j}{\partial N}}ds=0,\quad\forall j.$$

The bar denotes complex conjugation. Let us multiply (*) by $\phi_j(x)/k_j^2$, where $(\nabla^2+k_j^2)\phi_j=0$ in Δ, $\phi_j=0$ on $\partial\Delta$, and sum over j. Since $\Sigma_j\phi_j(x)\overline{\phi_j(s)}/k_j^2=G(x,s)$, $-\nabla^2G=\delta(x-y)$ in Δ, $G=0$ on $\partial\Delta$, this yields

$$v(x) = \int_\Gamma \frac{\partial G(x,s)}{\partial N_s} f(s)ds = 0, \quad x \in \Delta$$

and from jump relations for the potential of double layer, one sees that $f = 0$. Let

$$w_m = \begin{pmatrix} \sum_{j=1}^{m} c_j^{(m)} \phi_j \\ \sum_{j=1}^{m} d_j^{(m)} \frac{\partial \phi_j}{\partial N} \end{pmatrix}. \tag{20}$$

The projection method consists in finding $c_j^{(m)}$, $d_j^{(m)}$ from the linear system

$$(w_m - Bw_m - h, \eta_j)_{H^0} = 0, \quad 1 \leqslant j \leqslant 2m, \tag{21}$$

where

$$\eta_j = \begin{pmatrix} \phi_j \\ 0 \end{pmatrix}, \quad 1 \leqslant j \leqslant m,$$

$$\eta_j = \begin{pmatrix} 0 \\ \frac{\partial \phi_j}{\partial N} \end{pmatrix}, \quad m+1 \leqslant j \leqslant 2m. \tag{22}$$

The system (21) is a linear system of $2m$ equations for $2m$ unknowns $c_j^{(m)}, d_j^{(m)}$. From the known results [see Refs. 7 and 8 for general theory and Ref. 2(a) and Ref. 4, p. 192 for the problems similar to (21)], it follows that (21) is uniquely solvable for all sufficiently large m and $w_m \to w$ in H^0, where w solves (17).

We reduce by half the number of the unknowns if we use Eq. (15) [the second equation in the vector Eq. (17)] in the form (13), and set

$$d_j^{(m)} = c_j^{(m)}. \tag{23}$$

In this case, the system (21) takes the form

$$\sum_{j=1}^{m} a_{nj} c_j^{(m)} = u_{0n}, \quad 1 \leqslant n \leqslant m, \tag{24}$$

where

$$a_{nj} = (\phi_j, \phi_n) + \kappa (T\phi_j, \phi_n) - (1-\rho)\left(Q \frac{\partial \phi_j}{\partial N}, \phi_n\right), \tag{25}$$

$$u_{0n} = (u_0, \phi_n), \quad (f,g) \equiv (f,g)_{H_0}. \tag{26}$$

The system (24) one can also obtain by applying the projection method to Eq. (19). In this case, the proof of the convergence of the projection scheme requires further study. The reason is that the operator $Q(\partial/\partial N)$ is not compact in H^q.

Remark: If $\rho = 1$ [see (4)], then (19) becomes $u = u_0 - \kappa T u$ and the numerical scheme (24) [with the matrix a_{nj} defined by (25) with $\rho = 1$)] converges in H_q provided that $u_0 \in H_q$. If $\rho = 1$, then it follows from (12) that the values of u in $\mathscr{D}$ define u in the whole space. Therefore, if $u_m(x)$ is the approximate solution defined from (24) by the formula $u_m \equiv \Sigma_{j=4}^{m} c_j^{(m)} \phi_j$, then the function $u_0(x) - \kappa T u_m$ converges to the solution $u(x)$ of the problem (1)–(6) with $\rho = 1$ in H_q, in $\mathscr{C}(\Omega_R)$, where $\Omega_R \equiv \{x : |x| > R\}$ and $\mathscr{D} \subset \mathscr{B}_R$, $\mathscr{B}_R \equiv \{x : |x| \leqslant R\}$. If $q \geqslant 2$, then $u_0(x) - \kappa T u_m \to u(x)$ in $\mathscr{C}(\mathbb{R}^3)$ as $m \to \infty$.

[1] W. Kupradze, *Randwertaufgaben der Schwingungs theorie und integralgleichungen* (VEB, Berlin, 1956).
[2] A. G. Ramm, J. Math. Phys. (a) **22**, 1123 (1982); (b) **22**, 1112 (1982).
[3] E. Marx, J. Math. Phys. **23**, 1057 (1982).
[4] A. G. Ramm, *Theory and Applications of Some New Classes of Integral Equations* (Springer-Verlag, New York, 1980).
[5] A. G. Ramm, J. Math. Phys. **22**, 275 (1981).
[6] Ju. Berezanskii, *Expansions in Eigenfunctions of Selfadjoint Operators*, Transl. Math. Monographs, Vol. 17 (Am. Math. Soc., Providence, RI, 1968).
[7] M. A. Krasnoselskij *et al.*, *Approximate Solution of Operator Equations* (Walters Noordhoff, Groningen, 1972).
[8] I. Gohberg and I. Feldman, *Convolution Equations and Projection Methods for Their Solutions*, Transl. Math. Monographs, N41 (Am. Math. Soc., Providence, RI, 1974).

Three-term recursion relations for hydrogen wave functions: Exact calculations and semiclassical approximations

E. de Prunelé
Service de Physique des Atomes et des Surfaces, Centre d'Etudes Nucléaires de Saclay, 91191 Gif-sur-Yvette Cedex, France

(Received 6 May 1983; accepted for publication 7 October 1983)

Three-term recursion relations with respect to the angular momentum l are given for the normalized hydrogen wave functions associated with r, p_r, p (r is the radial polar coordinate in configuration representation, p_r is the momentum conjugated to r, p is the radial polar coordinate in momentum representation). These three-term recursion relations [Eqs. (4), (6), (16)] are found numerically stable in the order of decreasing l values, even for large quantum numbers. The three-term recursion relations in r and p are used to derive semiclassical approximations for the radial wave functions $P_{n,l}(p)$ and $R_{n,l}(r)$. These semiclassical approximations [Eqs. (67) and (84)] are valid even at the classical turning points and are still markedly good at small quantum numbers.

PACS numbers: 02.70. + d, 03.65.Ge, 03.65.Sq, 31.50. + w

I. INTRODUCTION

Atoms in highly excited, or Rydberg, states are now currently studied experimentally as well as theoretically. Computations based on the explicit expressions for the wave functions are time-consuming and may also introduce severe round-off errors (see, e.g., Ref. 1), due to the summation of numbers of nearly equal magnitude but of opposite signs. In the presence of an external field, and when the eigenstates are expressed in the basis of zero-field eigenstates,[1] numerically stable recursive algorithms between states of different angular momentum l appear desirable.

The purpose of this paper is twofold:

(a) In the first part (Sec. II), three-term recursion relations in l are given for normalized nonrelativistic hydrogenic wave functions: $R_{n,l}(r)$, $Q_{n,l}(p_r)$, $P_{n,l}(p)$. $R_{n,l}(r)$ is the usual radial wave function in configuration representation. The recursion relation for $R_{n,l}(r)$ corresponds therefore to the usual three-term recursion relation between Coulomb functions (see, e.g., Ref. 2). Up to now the numerical stability for the discrete spectrum has not been studied, at least to our knowledge. $Q_{n,l}(p_r)$ is the wave function associated with the conjugated momentum of r. An explicit expression for $Q_{n,l}(p_r)$ has been published recently by Lombardi.[3] $P_{n,l}(p)$ is the radial wave function in momentum representation. The explicit expression for $P_{n,l}(p)$ is known already for a long time (see Podolsky and Pauling[4]). To our knowledge the three-term recursion relations for $Q_{n,l}(p_r)$ and $P_{n,l}(p)$ were not given previously in the literature. As the most significant result, all these three-term recursion relations provide numerically stable algorithms when used in the order of decreasing l values.

(b) In the second part of this paper (Sec. III), the three-term recursion relations are used to derive semiclassical expressions for $P_{n,l}(p)$ and $R_{n,l}(r)$. To our knowledge, no semiclassical expressions for $P_{n,l}(p)$ have been given previously. The semiclassical expressions are valid even at the turning points and are accurate at small quantum numbers. The procedure involved is in close analogy to a previous work of Schulten and Gordon.[5] Their investigations show how one can derive semiclassical approximations for the Wigner's $3j$ and $6j$ coefficients starting from the three-term recursion relations satisfied by these coefficients. The derivation involves a discrete analog of the WKB method.

II. THREE-TERM RECURSION RELATIONS AND EXACT COMPUTATIONS

A. The three-term recursion relations

Atomic units will be used throughout and, for simplicity, we consider the case of unit charge ($Z = 1$). The starting point for the derivation is the factorization of the Schrödinger equation for the hydrogen atom (see, e.g., Ref. 6)

$$\left[1 + (l+1)\left(\frac{d}{dr} - \frac{l}{r}\right)\right]R_{n,l}(r) = E(n,l+1)R_{n,l+1}(r) \qquad (\text{if } 0 \leqslant l \leqslant n-2), \tag{1}$$

$$\left[1 - l\left(\frac{d}{dr} + \frac{(l+1)}{r}\right)\right]R_{n,l}(r) = E(n,l)R_{n,l-1}(r) \qquad (\text{if } 1 \leqslant l \leqslant n-1), \tag{2}$$

where

$$E(n,l) \equiv [1 - (l/n)^2]^{1/2} \tag{3}$$

and $R_{n,l}(r)$ are the normalized radial wave functions.

Equation (1) still remains valid for l equal to $n-1$ if the right-hand side is then taken equal to zero.[6]

Elimination of the derivative of $R_{n,l}(r)$ in Eqs. (1) and (2) leads to the well-known recursion relations between Coulomb functions[2]:

$$lE(n,l+1)R_{n,l+1}(r) + (l+1)E(n,l)R_{n,l-1}(r) - (2l+1)[1 - l(l+1)/r]R_{n,l}(r) = 0. \tag{4}$$

The operator p_r, which is not an observable[3] may be expressed in terms of the variable r^3:

$$p_r = \frac{1}{2}\left[\frac{\mathbf{r}}{r}\cdot\mathbf{p} + \mathbf{p}\cdot\frac{\mathbf{r}}{r}\right] = -i\left(\frac{d}{dr} + \frac{1}{r}\right). \tag{5}$$

Introducing p_r in Eqs. (1) and (2) and eliminating $1/r$ leads to a three-term recursion relation for $Q_{n,l}(p_r)$:

$$l^2E(n,l+1)Q_{n,l+1}(p_r)-(l+1)^2E(n,l)Q_{n,l-1}(p_r)$$
$$-(2l+1)[-1+il(l+1)p_r]Q_{n,l}(p_r)=0. \qquad (6)$$

The three-term recursion relation for $P_{n,l}(p)$ is more difficult to obtain, since the relation between $P_{n,l}(p)$ and $R_{n,l}(r)$ is l-dependent:

$$P_{n,l}(p)=(-i)^l\sqrt{\frac{2}{\pi}}\int_0^\infty dr\, r^2 j_l(pr)R_{n,l}(r), \qquad (7)$$

where $j_l(z)$ is the usual spherical Bessel function.[2] Multiplying Eq. (1) from the left by

$$(-i)^{l+1}\sqrt{\frac{2}{\pi}}\int_0^\infty dr\, r^2 j_{l+1}(rp),$$

Eq. (2) from the left by

$$(-i)^{l-1}\sqrt{\frac{2}{\pi}}\int_0^\infty dr\, r^2 j_{l-1}(rp)$$

and making use of the orthogonality condition,

$$\int_0^\infty dr\, r^2 j_\lambda(kr)j_\lambda(k'r)=\frac{\pi}{2k^2}\,\delta(k-k'), \qquad (8)$$

we obtain

$$E(n,l+1)P_{n,l+1}(p)$$
$$=-(i)\Big\{-(l+1)pP_{n,l}(p)+\frac{2}{\pi}\int_0^\infty dr\, r^2$$
$$\times j_{l+1}(rp)\int_0^\infty dp'\, p'^2 j_l(rp')P_{n,l}(p')\Big\} \qquad (9)$$

$$E(n,l)P_{n,l-1}(p)$$
$$=i\Big\{-lpP_{n,l}(p)+\frac{2}{\pi}\int_0^\infty dr\, r^2$$
$$\times j_{l-1}(rp)\int_0^\infty dp'\, p'^2 j_l(rp')P_{n,l}(p')\Big\}. \qquad (10)$$

In the derivation of Eqs. (9) and (10) use has been made of the facts that the operators $(-d/dz+l/z)$ and $(d/dz+(l+1)/z)$ appearing in Eqs. (1) and (2) are the raising and lowering operators associated with the spherical Bessel functions[2]:

$$\Big(-\frac{d}{dz}+\frac{l}{z}\Big)j_l(z)=j_{l+1}(z), \qquad (11)$$

$$\Big(\frac{d}{dz}+\frac{l+1}{z}\Big)j_l(z)=j_{l-1}(z). \qquad (12)$$

We now use the relations:

$$j_{l+1}(z)+j_{l-1}(z)=(2l+1)j_l(z)/z, \qquad (13)$$

$$\Big(\frac{p^2}{2}+\frac{1}{2n^2}\Big)P_{n,l}(p)=\frac{2}{\pi}\int_0^\infty dr\, rj_l(rp)$$
$$\times\int_0^\infty dp'\, p'^2 j_l(rp')P_{n,l}(p'). \qquad (14)$$

The last equation comprises the fact that the states $|n,l,m\rangle$ satisfy the Schrödinger equation:

$$\Big(\frac{p^2}{2}-\frac{1}{r}\Big)|n,l,m\rangle=-\frac{1}{2n^2}|n,l,m\rangle. \qquad (15)$$

Substracting Eq. (10) from Eq. (9) we finally obtain the desired recursion relation[7]:

$$E(n,l+1)P_{n,l+1}(p)-E(n,l)P_{n,l-1}(p)$$
$$-\frac{(2l+1)}{2n}\,i\Big(np-\frac{1}{np}\Big)P_{n,l}(p)=0. \qquad (16)$$

For the special value $p=1/n$ this recursion becomes a two-term recursion relation. The following section has the purpose of finding stable algorithms for the computation of $R_{n,l}(r)$, $Q_{n,l}(p_r)$, and $P_{n,l}(p)$ from the Eqs. (4), (6), and (16).

B. The algorithms

The numerical stability of a three-term recursion relation may depend critically on the direction of recursion. For the recursions given by Eqs. (4), (6), and (16), numerical stability can be *a priori* expected only if the recursive evaluation proceeds from the classically forbidden region towards the classically allowed region. The situation is quite analogous to the one corresponding to recursive evaluation of $3j$ and $6j$ Wigner's coefficients.[8]

The classical regions are determined as follows. For fixed values of n and l, there is associated with a classical motion on an ellipse of semimajor axis equal to n^2 and of eccentricity $e(n,l)$

$$e(n,l)=[1-l(l+1)/n^2]^{1/2}. \qquad (17)$$

Therefore, for a given value of r between 0 and $2n^2$, l belongs to the classical region if

$$e(n,l)\geqslant|r/n^2-1|. \qquad (18)$$

Using the classical relation between r, p_r, p,

$$-\frac{1}{2n^2}=\frac{p^2}{2}-\frac{1}{r}=\frac{p_r^2}{2}+\frac{l(l+1)}{2r^2}-\frac{1}{r}, \qquad (19)$$

the classical regions of l for fixed values of p_r and p are found to be

$$\frac{e(n,l)}{l}\geqslant|p_r|, \qquad (20)$$

$$e(n,l)\geqslant\Big|\frac{1-n^2p^2}{1+n^2p^2}\Big|. \qquad (21)$$

Equations (18), (20), and (21) show that progressing from the classically forbidden domain of l towards the allowed domain requires the use of recursion relations (4), (6), and (16) in the order of decreasing l values.

It can be seen that only one initial value is necessary to generate the recursion relations, namely $R_{n,n-1}(r)$, $Q_{n,n-1}(p_r)$, and $P_{n,n-1}(p)$. In fact, the recursion relations (4), (6), and (16) which are defined for $1\leqslant l\leqslant n-2$, are still valid for l equal to $n-1$ if the undefined products zero $\times R_{n,n}(r)$, zero $\times Q_{n,n}(p_r)$, and zero $\times P_{n,n}(p)$ appearing for l equal to $n-1$ are taken equal to zero. The reason for that is that $R_{n,n-1}(r)$ satisfies the following equation [see the remark just after Eq. (3)]:

$$\Big[1+n\Big(\frac{d}{dr}-\frac{(n-1)}{r}\Big)\Big]R_{n,n-1}(r)=0. \qquad (22)$$

Then it follows from Eq. (16) that $P_{n,l}(1/n)$ is zero if $n-l$ is even.

Still it remains to give the expressions for $R_{n,n-1}(r)$, $Q_{n,n-1}(p_r)$, and $P_{n,n-1}(p)$. Using the phase convention given in Ref. 6 the explicit expression for $R_{n,l}(r)$ is[6]:

$$R_{n,l}(r) = \frac{2(-1)^{n-l-1}}{n^2(2l+1)!}\left[\frac{(n+l)!}{(n-l-1)!}\right]^{1/2}\left(\frac{2r}{n}\right)^l \times {}_1F_1\left(-n+l+1, 2l+2; \frac{2r}{n}\right)\exp\left(-\frac{r}{n}\right) \tag{23}$$

which reduces for l equal to $n-1$ to

$$R_{n,n-1}(r) = \left(\frac{2}{n}\right)^{n+1/2}\frac{r^{n-1}}{\sqrt{(2n)!}}\exp\left(-\frac{r}{n}\right). \tag{24}$$

Lombardi[3] derives the expression for $Q_{n,l}(p_r)$ from its integro-differential equation. In order to obtain normalized expressions with a phase compatible with the choice of Eq. (23), it is more convenient to use the direct transformation between $R_{n,l}(r)$ and $Q_{n,l}(p_r)$. If p_r is given by Eq. (5), the transformation can be written as

$$Q_{n,l}(p_r) = \frac{1}{\sqrt{2\pi}}\int_0^\infty dr\, r\exp(-irp_r)R_{n,l}(r). \tag{25}$$

Because $R_{n,l}(r)$ is real [Eq. (23)], the relation

$$Q_{n,l}(-p_r) = Q_{n,l}(p_r)^* \tag{26}$$

holds, and from the orthogonality properties of $R_{n,l}(r)$ one obtains

$$\int_{-\infty}^{+\infty} dp_r\, Q^*_{n,l}(p_r)Q_{n',l}(p_r) = \delta_{nn'}. \tag{27}$$

Expressing the confluent hypergeometric function of Eq. (23) in terms of a Laguerre polynomial, and using the relation[9]:

$$\int_0^\infty dt\exp(-st)t^{\beta}L_n^{\alpha}(t) = \frac{\Gamma(\beta+1)}{n!}\frac{\Gamma(\alpha+n+1)}{\Gamma(\alpha+1)s^{\beta+1}} \times {}_2F_1\left(-n, \beta+1; \alpha+1; \frac{1}{s}\right) \quad (\text{if } \operatorname{Re}\beta > -1 \text{ and } \operatorname{Re}s > 0), \tag{28}$$

one obtains

$$Q_{n,l}(p_r) = \frac{(-1)^{n-l-1}}{2\sqrt{2\pi}}\frac{(l+1)!}{(2l+1)!} \times\left[\frac{(n+l)!}{(n-l-1)!}\right]^{1/2}\left(\frac{2}{1+inp_r}\right)^{l+2} \times {}_2F_1\left(-n+l+1, l+2; 2l+2; \frac{2}{1+inp_r}\right), \tag{29}$$

and therefore

$$Q_{n,n-1}(p_r) = \frac{1}{2\sqrt{2\pi}}\frac{n!}{\sqrt{(2n-1)!}}\left(\frac{2}{1+inp_r}\right)^{n+1}. \tag{30}$$

The expression for $P_{n,l}(p)$, as defined by Eqs. (7) and (23), is[4]

$$P_{n,l}(p) = (-i)^l\frac{2^{2l+3}}{\sqrt{2\pi}}n^2\frac{l!}{(2l+1)!} \times\left[\frac{(n+l)!}{(n-l-1)!}\right]^{1/2}\frac{(np)^l}{(1+n^2p^2)^{l+2}} \times {}_2F_1\left(-n+l+1, n+l+1; l+\frac{3}{2}; \frac{n^2p^2}{1+n^2p^2}\right), \tag{31}$$

and, therefore,

$$P_{n,n-1}(p) = (-i)^{n-1}\frac{2^{2n+1}}{\sqrt{2\pi}}\frac{nn!}{\sqrt{(2n-1)!}}\frac{(np)^{n-1}}{(1+n^2p^2)^{n+1}}. \tag{32}$$

It should be pointed out that for the special case of $l = 0$, the expression (29) for $Q_{n,l}(p_r)$ and (31) for $P_{n,l}(p)$ can be rewritten in forms directly suitable for numerical computation:

$$Q_{n,0}(p_r) = \frac{1}{2}\sqrt{\frac{n}{2\pi}}\left(\frac{2}{1+inp_r}\right)^2\left(\frac{1-inp_r}{1+inp_r}\right)^{n-1}, \tag{33}$$

$$P_{n,0}(p) = 4\sqrt{\frac{n}{2\pi}}\frac{1}{p(1+n^2p^2)} \times\sin\left[n\arccos\left(\frac{1-n^2p^2}{1+n^2p^2}\right)\right], \tag{34}$$

with the obvious symmetry properties

$$Q_{n,0}(1/(n^2p_r)) = (-1)^n(np_r)^2Q^*_{n,0}(p_r), \tag{35}$$

$$P_{n,0}(1/(n^2p)) = (-1)^{n-1}(np)^4P_{n,0}(p). \tag{36}$$

The symmetry given by Eq. (36) is also valid for $l \neq 0$ [with $(-1)^{n-l-1}$ in place of $(-1)^{n-1}$] and results from the dynamic symmetry of the hydrogen atom.[10]

The calculated values of $Q_{n,0}$ and $P_{n,0}$ obtained from the exact relations (33) and (34) have been used as references for checking the recursively computed values. A test has been made for $n = 100$ and the recursive algorithms have been found to be stable. Some numerical results are given in Table I, for values of p_r and p covering a large domain centered on $1/n$. All recursive calculations were made in single precision on a Cray computer. It was found that the recursive algorithms introduce only relative errors which are close to the machine accuracy, except, of course, in the closed vicinity of the nodes. For $R_{n,l}(r)$ the numerical stability of the recursively computed values has been tested by comparing $R_{n,0}(r)$ to $\tilde{R}_{n,0}(r)$ which is the value obtained after a relative error of 10^{-4} for $R_{n,n-2}(r)$ has been introduced. This error is not amplified during the recursive process. Some numerical results are reported in Table II for values of r covering a large domain centered on n^2.

III. SEMICLASSICAL EXPRESSIONS FOR $P_{n,l}(p)$ AND $R_{n,l}(r)$

A. Introduction

First we like to summarize some general results obtained by Schulten and Gordon,[5] which are essential for the present work. These authors have shown that approximate solutions of the following difference equation:

$$f(x+1) + f(x-1) - 2f(x)\cos(k(x)) = 0 \tag{37}$$

are also approximate solutions for

$$\left(\frac{d^2}{dx^2} + k^2(x)\right)\left(\frac{\sin k(x)}{k(x)}\right)^{1/2}f(x) = 0 \tag{38}$$

TABLE I. Numerical results for values of p_r and p covering a large domain centered on $1/n$. I computed from Eq. (33). II computed recursively. III computed from Eq. (34). $(a, b) = a + ib$.

p, p_r	$Q_{100,0}(p_r)$	$P_{100,0}(p)$
10^{-5}	I(0.781979219607396$E+01$, $-$0.158514981099567$E+01$) II(0.781979219607129$E+01$, $-$0.15851498199504$E+01$)	0.317029962198930$E+06$ III 0.317029962198891$E+06$ II
10^{-4}	I($-$0.331955574889624E01, $-$0.725463964515129$E+01$) II($-$0.331955574889447$E+01$, $-$0.725463964514978$E+01$)	0.145092792902893$E+06$ III 0.145092792902831$E+06$ II
10^{-3}	I(0.369430400728700$E+01$, $-$0.698281481097695$E+01$) II(0.369430400728943$E+01$, $-$0.698281481097101$E+01$)	0.139656296219514$E+05$ III 0.139656296219400$E+05$ II
10^{-2}	I(0.39894228041432$E+01$, 0.000000000000000$E+00$) II(0.398942280400938$E+01$, $-$0.241917597065822$E-12$)	0.000000000000000$E+00$ III 0.000000000000000$E+00$ II
10^{-1}	I(0.369430400728468$E-01$, 0.698281481097416$E-01$) II(0.369430400728419$E-01$, 0.698281481096714$E-01$)	$-$0.139656296219512$E+01$ III $-$0.139656296219354$E+01$ II
1	I($-$0.331955574889260$E-03$, 0.725463964514367$E-03$) II($-$0.331955574888673$E-03$, 0.725463964513437$E-03$)	$-$0.145092792902891$E-02$ III $-$0.145092792902661$E-02$ II
10	I(0.781979219607388$E-05$, 0.158514981099565$E-05$) II(0.781979219606661$E-05$, 0.158514981099378$E-05$)	$-$0.317029962198933$E-06$ III $-$0.317029962198682$E-06$ II

provided that $k(x)$ is a "slowly" varying function in x. As approximate solutions of the differential equation, they obtained[5]

$$f(x) = \begin{cases} C \dfrac{|\Omega(x)|^{1/4}}{[\sin k(x)]^{1/2}} F(\Omega(x)) & \text{if } k(x) \text{ is real} \quad (39a) \\ C \dfrac{|\Omega(x)|^{1/4}}{[\sinh|k(x)|]^{1/2}} F(\Omega(x)) & \text{if } k(x) \text{ is purely imaginary,} \quad (39b) \end{cases}$$

where C is a constant, F is a linear combination of regular and irregular Airy functions, denoted by Ai and Bi, respectively, and

$$\Omega(x) = \begin{cases} -\left(\dfrac{3}{2}\left|\displaystyle\int_{x_0}^{x} k(x')dx'\right|\right)^{2/3} & \text{if } k(x') \text{ is real} \quad (40a) \\ +\left(\dfrac{3}{2}\left|\displaystyle\int_{x_0}^{x} k(x')dx'\right|\right)^{2/3} & \text{if } k(x') \text{ is purely imaginary.} \quad (40b) \end{cases}$$

x_0 is either of the zeros of $\sin k(x)$. For more details we refer to Ref. 5.

TABLE II. Numerical results for values of r covering a large domain centered on n^2. I computed recursively. II computed recursively after a relative error of 10^{-4} has been introduced at the beginning of the recursion (see text).

r	$R_{100,0}(r)$
10^{-1}	$-$0.18066$E-02$ I $-$0.18067$E-02$ II
1	$-$0.56598$E-03$ I $-$0.56603$E-03$ II
10	$-$0.11252$E-03$ I $-$0.11253$E-03$ II
10^2	$-$0.15031$E-04$ I $-$0.15032$E-04$ II
10^3	0.20660$E-06$ I 0.20662$E-06$ II
10^4	0.77177$E-06$ I 0.77177$E-06$ II
10^5	0.37372 $-$ 270 I 0.37376 $-$ 270 II

In the following paragraphs it will be shown how the recursion relations (4) and (16) for $R_{n,l}(r)$ and $P_{n,l}(p)$ merge into the form of Eq. (37) for large quantum numbers. A similar procedure could be applied for $Q_{n,l}(p_r)$ [Eq. (6)] with the difference that $k(x)$ will be complex with both real and imaginary part being nonzero; this more complicated case will not be considered here.

We note that, for a fixed value of $E(n,l)$ [Eq. (3)], one obtains

$$E(n,l) = [E(n,l-\tfrac{1}{2})E(n,l+\tfrac{1}{2})]^{1/2} + O(1/n^2), \qquad (41a)$$

$$l^2 = (l-\tfrac{1}{2})(l+\tfrac{1}{2}) + O(1/n^2), \qquad (41b)$$

$$l(l+1) = (l+\tfrac{1}{2})^2 + O(1/n^2), \qquad (41c)$$

$O(1/n^2)$ denotes terms decreasing as $1/n^2$. On the other hand, for l near its maximum value, $n-1$, $E(n,l)$ cannot be considered invariable as n increases; in that case the correction term in Eq. (41a) decreases as $1/n$ only. Equations (41b) and (41c) are, of course, not valid for l near its minimum value ($l_{\min} = 0$).

Using the relations (41a, b, c) and setting

$$x = l + \tfrac{1}{2}, \qquad (42)$$

both recursions relations (4) and (16) are transformed into the form of Eq.(37), as will be shown in the following paragraphs. First x will be considered as varying continuously between zero and n; then it will be considered as verifying Eq. (42).

B. Semiclassical approximations for $P_{n,l}(p)$

Let us first consider the case $p < 1/n$. Setting

$$f(l+\tfrac{1}{2}, p) = i^l [E(n,l+\tfrac{1}{2})]^{1/2} P_{n,l}(p) \qquad (43)$$

and

$$\cos k(x) = \frac{x}{2nE(n,x)} (1/np - np) \qquad (44)$$

one verifies from Eq. (16) that $f(x, p)$ satisfies within the semiclassical approximations [Eqs. (41a, b, c,)] the difference equation (37). The condition $p < 1/n$ ensures that $\cos k(x)$ [Eq. (44)] increases from zero to unity as x varies from zero to its maximum classical value, denoted by x_0; hence $k(x)$ vanishes for x_0 and becomes purely imaginary for $x > x_0$. Replacing $l(l+1)$ by $(l+\frac{1}{2})^2$ in the Eq. (21) yields

$$x_0 = n[2np/(1+n^2p^2)]. \tag{45}$$

In the classical domain $(x \leqslant x_0)$

$$k(x) = \arccos[x\alpha/nE(n,x)], \tag{46}$$

with

$$\alpha = \tfrac{1}{2}(1/np - np) \tag{47}$$

and the analytic integration of $\int_{x_0}^{x} k(x')dx'$, abbreviated as $\omega(x,p)$, yields

$$\omega(x,p) = xk(x) - n\arctan\{[E(n,x)/\alpha]\sin k(x)\}. \tag{48}$$

For $x \leqslant x_0$ the semiclassical expression for $f(x,p)$ can therefore be written as [see Eq. (39)]:

$$f(x,p) = C(p)[\tfrac{3}{2}|\omega(x,p)|]^{1/6} \times F(-[\tfrac{3}{2}|\omega(x,p)|]^{2/3})/[\sin k(x)]^{1/2} \tag{49}$$

where $C(p)$ is a normalization constant. The analytic continuation into the classically forbidden domain $x > x_0$ yields [see Eqs. (39), (40)]:

$$f(x,p) = C(p)[\tfrac{3}{2}|\omega(x,p)|]^{1/6} \times F(+[\tfrac{3}{2}|\omega(x,p)|]^{2/3})/[\sinh|k(x)|]^{1/2}, \tag{50}$$

where now

$$|k(x)| = \operatorname{arccosh}[x\alpha/nE(n,x)] \tag{51}$$

$$|\omega(x,p)| = |x|k(x)| - n\operatorname{arctanh}\{[E(n,x)/\alpha]\sinh|k(x)|\}|. \tag{52}$$

To obtain a function $f(l+\frac{1}{2},p)$ which decreases towards zero as l penetrates into the classically forbidden domain, the irregular Airy function Bi must be suppressed for the discrete value $l+\frac{1}{2}$, where F is taken equal to the regular Airy function Ai. The same procedure is repeated for the case $p > 1/n$ by defining

$$g(l+\tfrac{1}{2},p) = (-1)^m(-i)^l[E(n,l+\tfrac{1}{2})]^{1/2}P_{n,l}(p), \tag{53}$$

$$\cos j(x) = -[x/2nE(n,x)](1/np - np), \tag{54}$$

in place of Eqs. (43), (44); again this ensures that the argument of the cosine, $j(x)$, is zero for x_0. The solution for $g(x,p)$ is given by the right-hand sides of Eqs. (48) and (49) where α [Eq. (47)] is now defined as

$$\alpha = \tfrac{1}{2}|1/np - np|. \tag{55}$$

The integer m in Eq. (53) will be determined in order to ensure approximate continuity of the semiclassical wave functions.

In the following the variable x will be treated as a parameter taking only the discrete values $l+\frac{1}{2}$ [Eq. (42)]; this requires

$$f(l+\tfrac{1}{2},1/n-\epsilon) \simeq (-1)^{l+m}g(l+\tfrac{1}{2},1/n+\epsilon) \tag{56}$$

in the limit where ϵ approaches zero from positive values. The value $1/n$ lies inside the classical domain in p, $p_1 \leqslant p \leqslant p_2$; with p_1, p_2 given by Eq. (21):

$$p_1 = \frac{1}{n}\left(\frac{1-E(n,l+\frac{1}{2})}{1+E(n,l+\frac{1}{2})}\right)^{1/2}, \tag{57a}$$

$$p_2 = \frac{1}{n}\left(\frac{1+E(n,l+\frac{1}{2})}{1-E(n,l+\frac{1}{2})}\right)^{1/2}. \tag{57b}$$

For p inside the classical domain $|\omega(x,p)|$ is large (except in the close vicinity of p_1 and p_2), and in that case[5]

$$\sqrt{\pi}\{[\tfrac{3}{2}|\omega(x,p)|]^{1/6}\mathrm{Ai}(-[\tfrac{3}{2}|\omega(x,p)|]^{2/3})\} \simeq \cos(-|\omega(x,p)| + \pi/4) \tag{58}$$

so that relation (56), together with the equality

$$\omega\left(l+\frac{1}{2},\frac{1}{n}-\epsilon\right) = \omega\left(l+\frac{1}{2},\frac{1}{n}+\epsilon\right) = \left(l+\frac{1}{2}-n\right)\frac{\pi}{2} \tag{59}$$

implies

$$\cos[(l+1-n)(\pi/2)] = (-1)^{l+m}\cos[(l+1-n)(\pi/2)]. \tag{60}$$

Therefore m can be chosen as

$$m = 1-n. \tag{61}$$

For the determination of $C(p)$ we use relation (58) and note that $\cos[-|\omega(x,p)| + \pi/4]$ oscillates rapidly inside the classical domain except for l near $n-1$. Normalizing the semiclassical wave function to unity inside this classical domain one may replace the squared cosine by its mean values and obtain:

$$\int_{p_1}^{p_2} dp\, p^2 \frac{|C(p)|^2}{2\pi\sin[k(x)]E(n,x)} = 1. \tag{62}$$

The integrand in Eq. (62) can be interpreted as the classical probability density in p. Classically, p varies between p_1 and p_2 monotonically as the time t progresses. Therefore the classical probability density, denoted by $D(p)$ which is proportional to the time elapsed between p and $p+dp$ is

$$D(p) = \frac{1}{T}\left|\frac{dt}{dp}\right|, \tag{63}$$

where T is the time elapsed between p_1 and p_2, that is, the half-period of motion. T is equal to πn^3.[11] Using Eq. (19) and the equality $p_r = dr/dt$, one obtains

$$\begin{aligned} D(p) &= \frac{1}{\pi n^3}\left|\frac{dt}{dr}\frac{dr}{dp}\right| \\ &= \frac{1}{\pi n^3}\frac{4}{(p^2+1/n^2)^2} \\ &\quad \times\left\{1-\left[\frac{l+\frac{1}{2}}{2n}\left(np+\frac{1}{np}\right)\right]^2\right\}^{-1/2} \end{aligned} \tag{64}$$

and therefore

$$|C(p)|^2 = 8/[n^3p^2(p^2+1/n^2)^2]. \tag{65}$$

At this stage, it is worthwhile to note that, for a straight line Bohr trajectory $[l(l+1)=0]$, one obtains from Eqs. (64) and(33):

$$D(p) = \frac{1}{\pi n^3}\frac{4}{(p^2+1/n^2)^2} = 2|Q_{n,0}(p)|^2. \tag{66}$$

Since classically p_r is equal to $\pm p$ for a straight-line trajectory [see Eq. (19)], $|Q_{n,0}(p_r)|^2$ corresponds exactly to the classical probability density associated with p_r. This is not

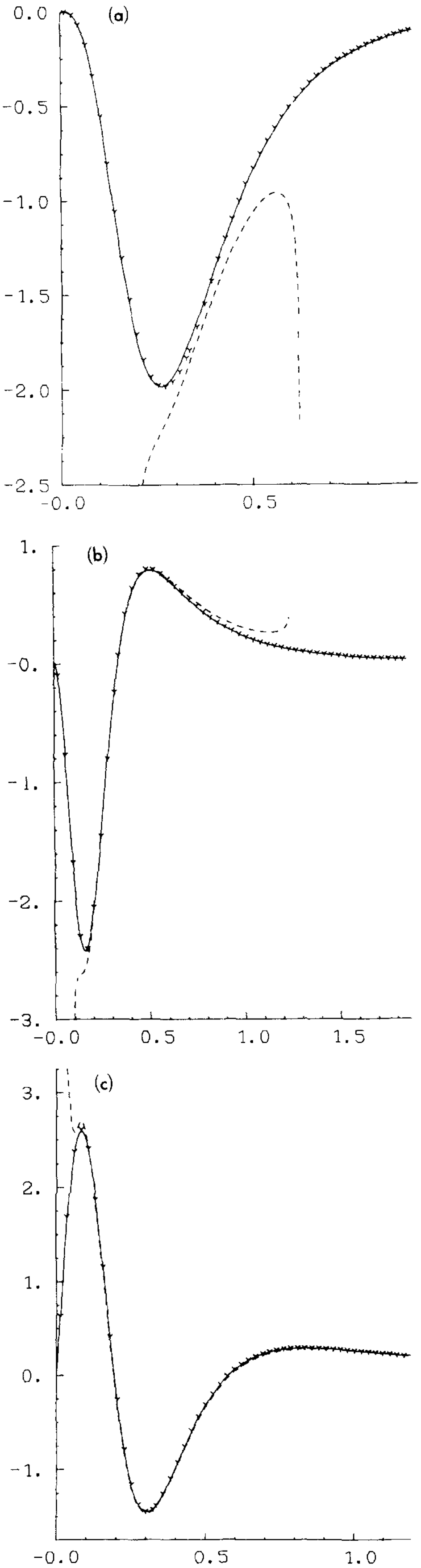


FIG. 1. Accuracy of the semiclassical approximation $p\times \mathscr{P}_{3,l}(p)$, $p\times W_{3,l}(p)$ with respect to the exact expression $p\times P_{3,l}(p)$. (a) $l=2$, (b) $l=1$ (y imaginary axis), (c) $l=0$. Y: $p\times \mathscr{P}_{3,l}(p)$; broken line – – –: $p\times \mathscr{W}_{3,l}(p)$; continuous line: $p\times P_{3,l}(p)$.

true for $p^2|P_{n,0}(p)|^2$ which differs from $D(p)$ [Eq. (66)] by oscillating terms.[12]

The semiclassical approximation to $P_{n,l}(p)$, denoted by $\mathscr{P}_{n,l}(p)$ is now summarized in a form convenient for direct application. Determining the overall phase factor by comparison with the exact result (31), and defining

$$\beta=\begin{cases}1 & \text{if } p<1/n\\ (-1)^{n-l-1} & \text{if } p>1/n,\end{cases} \tag{67a}$$

$$\alpha=\tfrac{1}{2}|1/np-np|, \tag{67b}$$

$$c=(l+\tfrac{1}{2})\alpha/[nE(n,l+\tfrac{1}{2})], \tag{67c}$$

one obtains

$$\mathscr{P}_{n,l}(p)=(-i)^l\beta\, 2\Big(\frac{2}{n^3}\Big)^{1/2}\frac{|z|^{1/4}\mathrm{Ai}(z)}{p(p^2+1/n^2)(y)^{1/2}}, \tag{67d}$$

with

$$y=E(n,l+\tfrac{1}{2})s, \tag{67e}$$

$$s=|1-c^2|^{1/2}, \tag{67f}$$

$$z=\begin{cases}-\{\tfrac{3}{2}[-(l+\tfrac{1}{2})k+n\arctan(y/\alpha)]\}^{2/3}\\ \qquad\text{if } c\leqslant 1\\ +(\tfrac{3}{2}\{+(l+\tfrac{1}{2})k-(n/2)\ln[(\alpha+y)/(\alpha-y)]\})^{2/3}\\ \qquad\text{if } c\geqslant 1,\end{cases} \tag{67g}$$

$$k=\begin{cases}\arccos(c) & \text{if } c\leqslant 1\\ \ln(c+(c^2-1)^{1/2}) & \text{if } c\geqslant 1.\end{cases} \tag{67h}$$

It can be seen that $\mathscr{P}_{n,l}(p)$ is continuous if $n-l$ is odd, and it exhibits a small discontinuity at the node $p=1/n$ for $n-l$ even. This discontinuity becomes smaller and smaller as $n-l$ increases as seen from Eq. (59) and relation (58). Using this latter relation, a continuous solution of the WKB type, denoted by $W_{n,l}(p)$ can be obtained in the classical domain ($c<1$):

$$W_{n,l}(p)=(-i)^l\beta\, 2\Big(\frac{2}{\pi n^3}\Big)^{1/2}\times\frac{\cos((l+\tfrac{1}{2})k-n\arctan(y/\alpha)+\pi/4)}{p(p^2+1/n^2)(y)^{1/2}}. \tag{68}$$

A systematic study of the accuracy of $\mathscr{P}_{n,l}(p)$, $W_{n,l}(p)$, with respect to all the parameters n, l, p is difficult. As an example we have considered these functions for $n=3$, and have compared them to the exact solutions [see Fig. 1(a), (b), (c)]. The accuracy of $\mathscr{P}_{3,l}(p)$ is remarkably good for such a low principal quantum number. The discontinuity of $\mathscr{P}_{3,1}$ at $p=\frac{1}{3}$ is negligible at the scale of the graph. The accuracy of the WKB type approximations is also good for $W_{3,0}(p)$ [Fig. 1(c)] and $W_{3,1}(p)$ [Fig. 1(b)] which are indistinguishable from the exact solutions at the scale of the graphs, except of course near the classical turning points. It should be noted that the divergence of $W_{3,0}(p)$ near the value $p_2\simeq 4$ [see Eq. (57b)] cannot be represented on Fig. 1(c) since p_2 is too large for the scale [the accuracy of $\mathscr{P}_{3,0}(p)$ for $p\simeq p_2$ has been verified]. The agreement of $W_{3,2}(p)$ compared to the exact solution is not very good. In that case $\mathscr{P}_{3,2}(p)$ provides a much better approximation.

C. Semiclassical approximations for $R_{n,l}(r)$

A similar procedure, as was outlined in III B for $\mathscr{P}_{n,l}(p)$, can be applied for $R_{n,l}(r)$; it will only be described briefly. Defining

$$f(l+\tfrac{1}{2},r)=(-1)^{l+m}[E(n,l+\tfrac{1}{2})/(l+\tfrac{1}{2})]^{1/2}\times R_{n,l}(r)\quad \text{if } r<n^2, \tag{69a}$$

$$g(l+\tfrac{1}{2},r)=[E(n,l+\tfrac{1}{2})/(l+\tfrac{1}{2})]^{1/2}\times R_{n,l}(r)\quad \text{if } r>n^2, \tag{69b}$$

where m is an integer, and

$$\cos k(x)=\alpha(x^2/r-1)/E(n,x), \tag{70}$$

with

$$\alpha=\begin{cases}+1 & \text{if } r<n^2\\ -1 & \text{if } r>n^2,\end{cases} \tag{71}$$

again it can be shown from Eq. (4) that, within the semiclassical approximations (41 a, b, c) $f(x,r)$ and $g(x,r)$ satisfy the difference equation (37). For $r<n^2$, $\cos k(x)$ increases from minus unity to unity as x varies from zero to its maximum value, x_0, which is equal to [see Eq. (18)]

$$x_0=[r(2-r/n^2)]^{1/2}. \tag{72}$$

For $r>n^2$, $\cos k(x)$ decreases from unity to $2(n^2/r)(r/n^2-1)^{1/2}$ as x increases from zero to $(r/n^2-1)^{1/2}$, and then rises to unity for x increasing to x_0. $f(x,r)$ and $g(x,r)$ are then both determined by the right-hand sides of Eqs. (48) and (49), where now r replaces p; $k(x)$ is determined by

$$k(x)=\arccos[\alpha(x^2/r-1)/E(n,x)]\quad \text{if } x<x_0 \tag{73}$$

$$|k(x)|=\operatorname{arccosh}[\alpha(x^2/r-1)/E(n,x)]\quad \text{if } x>x_0 \tag{74}$$

and, after analytic integration of $\int_{x_0}^{x}k(x')dx'$ denoted by $\omega(x,r)$:

$$\omega(x,r)=xk(x)-\alpha r[E(n,x)/x]\sin(k(x))-\alpha n\arctan\left(r\frac{E(n,x)}{x}\frac{\sin k(x)}{n-r/n}\right)\quad \text{if } x\leqslant x_0, \tag{75}$$

$$|\omega(x,r)|=x|k(x)|-\alpha r[E(n,x)/x]\sinh(|k(x)|)-\alpha n\operatorname{arctanh}\left(r\frac{E(n,x)}{x}\frac{\sinh|k(x)|}{n-r/n}\right)\quad \text{if } x\geqslant x_0. \tag{76}$$

The irregular Airy function Bi must be suppressed for the discrete values $x=l+\frac{1}{2}$ to ensure that $f(l+\frac{1}{2},p)$ and $g(l+\frac{1}{2},p)$ decrease towards zero as l crosses over into the classically forbidden domain. Approximate continuity at the point $r=n^2$ for $x=l+\frac{1}{2}$ requires

$$f(l+\tfrac{1}{2},n^2-\epsilon)\simeq(-1)^{l+m}g(l+\tfrac{1}{2},n^2+\epsilon) \tag{77}$$

in the limit where ϵ approaches zero from positive values. The value n^2 lies inside the classical domain in r, $r_1\leqslant r\leqslant r_2$ with r_1, r_2 given by Eq. (18):

$$r_1=n^2(1-E(n,l+\tfrac{1}{2})), \tag{78a}$$

$$r_2=n^2(1+E(n,l+\tfrac{1}{2})). \tag{78b}$$

Equation (75) yields

$$\begin{aligned}\omega(l+\tfrac{1}{2},n^2-\epsilon)&=(l+\tfrac{1}{2})\arccos[-E(n,l+\tfrac{1}{2})]-nE(n,l+\tfrac{1}{2})-n(\pi/2)\\&=-\omega(l+\tfrac{1}{2},n^2+\epsilon)+(l+\tfrac{1}{2}-n)\pi.\end{aligned} \tag{79}$$

Note that $\omega(x,r)$ is always negative inside the classical domain $x<x_0$ since $k(x)$ is positive. Using relation (58) requires

$$\begin{aligned}\cos[\omega(l+\tfrac{1}{2},n^2-\epsilon)+\pi/4]&=(-1)^{l+m}\cos[-\omega(l+\tfrac{1}{2},n^2-\epsilon)\\&\quad+(l+\tfrac{1}{2}-n)\pi+\pi/4].\end{aligned} \tag{80}$$

Therefore $(-1)^{1+m-n}$ must be equal to unity, and conveniently m may be choosen as

$$m=-n-1. \tag{81}$$

$C(r)$ is determined in the same way as for $C(p)$. This implies that we identify $r^2|C(r)|^2x/[2\pi E(n,x)\sin k(x)]$ with the classical probability density:

$$\begin{aligned}D(r)&=\frac{1}{T}\left|\frac{dt}{dr}\right|=\frac{1}{\pi n^3|p_r|}\\&=\left\{\pi n^3\left[-\frac{1}{n^2}-\left(\frac{l+\frac{1}{2}}{r}\right)^2+\frac{2}{r}\right]^{1/2}\right\}^{-1}\end{aligned} \tag{82}$$

and therefore

$$|C(r)|^2=2/(n^3r^2). \tag{83}$$

The overall phase factor can be determined by direct comparison between the exact result (23). Finally the semiclassical approximation to $R_{n,l}(r)$, denoted by $\mathscr{R}_{n,l}(r)$, is summarized in the following in a form convenient for direct application. Defining

$$\beta=\begin{cases}(-1)^{n-l-1} & \text{if } r<n^2\\ 1 & \text{if } r>n^2,\end{cases} \tag{84a}$$

$$\alpha=\begin{cases}1 & \text{if } r<n^2\\ -1 & \text{if } r>n^2,\end{cases} \tag{84b}$$

$$c=\alpha((l+\tfrac{1}{2})^2/r-1)/E(n,l+\tfrac{1}{2}), \tag{84c}$$

one obtains

$$\mathscr{R}_{n,l}(r)=\beta\left(\frac{2}{n^3}\right)^{1/2}\frac{|z|^{1/4}\mathrm{Ai}(z)}{r(y)^{1/2}}, \tag{84d}$$

with

$$y=E(n,l+\tfrac{1}{2})s/(l+\tfrac{1}{2}), \tag{84e}$$

$$s=|1-c^2|^{1/2}, \tag{84f}$$

$$z=\begin{cases}-\left\{\dfrac{3}{2}\left[-\left(l+\dfrac{1}{2}\right)k+\alpha ry+\alpha n\arctan\left(\dfrac{ry}{n-r/n}\right)\right]\right\}^{2/3} & \text{if } c\leqslant 1\\ +\left\{\dfrac{3}{2}\left[+\left(l+\dfrac{1}{2}\right)k-\alpha ry-\alpha\dfrac{n}{2}\ln\left(\dfrac{n-r/n+ry}{n-r/n-ry}\right)\right]\right\}^{2/3} & \text{if } c\geqslant 1,\end{cases} \tag{84g}$$

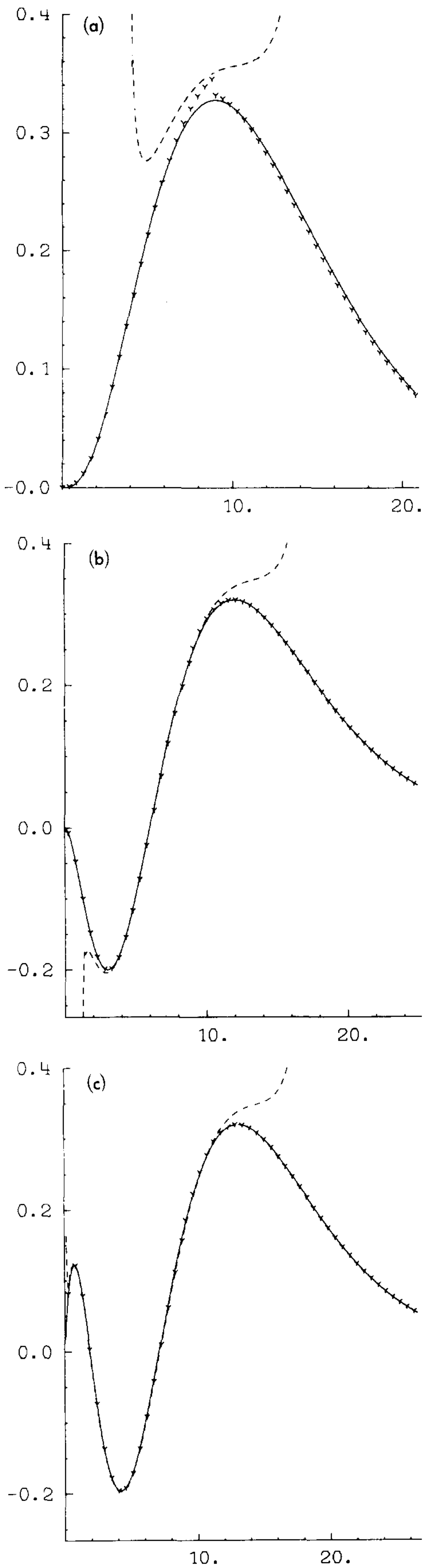


FIG. 2. Accuracy of the semiclassical approximations $r\times\mathscr{R}_{3,l}(r)$, $r\times\mathscr{W}_{3,l}(r)$ with respect to the exact expression $r\times R_{3,l}(r)$. (a) $l=2$, (b) $l=1$, (c) $l=0$. Y: $r\times\mathscr{R}_{3,l}(r)$; broken line ---: $r\times\mathscr{W}_{3,l}(r)$; continuous line: $r\times\mathscr{R}_{3,l}(r)$.

$$k=\begin{cases}\arccos(c) & \text{if } c\leqslant 1\\ \ln[c+(c^2-1)^{1/2}] & \text{if } c\geqslant 1.\end{cases} \tag{84h}$$

$\mathscr{R}_{n,l}(r)$ exhibits a small discontinuity for $r=n^2$ which becomes smaller and smaller as $n-l$ increases [see Eq. (79) and relation (58)]. Relation (58) may also be used to obtain a WKB type approximation denoted by $\mathscr{W}_{n,l}(r)$ in the classical domain $(c<1)$:

$$\begin{aligned}\mathscr{W}_{n,l}(r)&\\ &=(-1)^{n-l-1}\left(\frac{2}{\pi n^3 y}\right)^{1/2}\frac{1}{r}\\ &\quad\times\cos\left\{-ry-n\arcsin\left(\frac{r/n^2-1}{E(n,l+\frac{1}{2})}\right)\right.\\ &\quad\left.+\left(l+\frac{1}{2}\right)\arcsin\left(\frac{r-(l+\frac{1}{2})^2}{rE(n,l+\frac{1}{2})}\right)-(n-l-1)\frac{\pi}{2}\right\}.\end{aligned} \tag{85}$$

Equation (85) represents the usual WKB approximation.[13] This was expected right away recalling that [see Eq. (73)]

$$\int k(x')\,dx'=\int\left[\int\frac{d}{dr}(k(x'))\,dx'\right]dr=\int p_r\,dr. \tag{86}$$

Indeed the last integral corresponds to the classical action associated with the variable r. In the geometric picture the definition of $k(x)$ [Eq. (73)] corresponds to the polar angle in the plane of a classical trajectory.[11]

In Figs. 2(a), (b), (c) $\mathscr{R}_{n,l}(r)$ and $\mathscr{W}_{n,l}(r)$ are compared to the exact wave function for $n=3$. The agreement is quite good for $\mathscr{R}_{3,l}(r)$, except near the discontinuity at $r=n^2$ for maximum l, i.e., $l=2$. For this value of l the WKB approximation $\mathscr{W}_{3,2}(r)$ [Fig. 2(a)] is not good. The accuracy of $\mathscr{W}_{3,1}(r)$ [Fig. 2(b)] and $\mathscr{W}_{3,0}(r)$ [see Fig. 2(c)] is much better, except of course near the classical turning points.

IV. CONCLUSION

Highly excited atomic states require particular method of computation. Selected matrix elements may be obtained by simple analytic integration; also the classical evaluation applying the correspondence principle may lead to useful results. For more general or more complex cases, however, numerical integration may be necessary. The algorithms presented in this paper allow rapid and accurate calculations which are particularly suitable for Rydberg atoms in the presence of an external field when the basis is chosen to represent the zero-field eigenstates in polar coordinates. The recursion relations allow us to generate useful semiclassical approximation of the wave functions as outlined in Sec. III. The semiclassical approximations (67) and (84) extend the domain of validity of the WKB approximation in the region of the turning points. The semiclassical approximations are very useful since they provide direct information on the general shape of the wave functions (position of the nodes, etc...) which may not be obvious from the exact explicit expressions.

ACKNOWLEDGMENTS

The author would like to thank J. Pascale for helpful discussions and H. Telle for a careful reading of the manuscript.

[1]M. L. Zimmermann, M. G. Littman, M. M. Kash, and D. Kleppner, Phys. Rev. A **20**, 2251 (1979).
[2]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), 5th ed.
[3]J. R. Lombardi, Phys. Rev. A **22**, 797 (1980).
[4]B. Podolsky and L. Pauling, Phys. Rev. **34**, 109 (1929).
[5]K. Schulten and R. G. Gordon, J. Math. Phys. **16**, 1971 (1975).
[6]L. C. Biedenharn and J. D. Louck, "Angular momentum in quantum physics," in *Encyclopedia of Mathematics and Its Applications* (Addison-Wesley, Reading, MA, 1981), Vol. 8.
[7]Using a different phase convention, the recursion relation (16) can also be derived from Eqs. (27) and (27*) of Fock's classical paper, "Zur theorie des Wasserstoffatoms," Z. Phys. **98**, 745 (1935).
[8]K. Schulten and R. G. Gordon, J. Math. Phys. **16**, 1961 (1975).
[9]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals Series and Products* (Academic, New York, 1965), 4th ed.
[10]A. M. Perelomov and V. S. Popov, Zh. Eksp. Teor. Fiz. **50**, 179 (1966) [Sov. Phys. JETP **23**, 118 (1966)].
[11]L. D. Landau and E. M. Lifshitz, *Mécanique* (Edition Mir, Moscow, 1969), 3rd ed.
[12]E. de Prunelé, J. Phys. B: Atom. Mol. Phys. **13**, 3921 (1980).
[13]L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, New York, 1958), p. 170.

Explicit integrability for Hamiltonian systems and the Painlevé conjecture[a)]

B. Dorizzi and B. Grammaticos
Dépt. de Mathématiques Appliquées (MTI), Centre National d'Etudes des Télécommunications, 38-40 Ave. du Général Leclerc, 92131 Issy Les Moulineaux, France

A. Ramani
C.P.T. Ecole Polytechnique, 91128 Palaiseau, France

(Received 16 June 1983; accepted for publication 7 October 1983)

We analyze a class of Hamiltonian systems in two dimensions for which we proved that a second constant of motion exists. It is shown that, using the two first integrals, the equations of motion can be written in a form which allows their integration by quadratures. An analysis of the equations of motion in this reduced form establishes the behavior of the solutions in the complex-time plane. It is shown explicitly that the systems belonging to this class possess the "weak Painlevé" property, i.e., their solutions in complex time can present singularities of a specific algebraic type.

PACS numbers: 03.20. + i, 02.30. + g

In a recent work we have used the dual approach: Painlevé analysis of the equations of motion and direct search of constants of motion, to study two-dimensional Hamiltonian systems from the point of view of integrability.[1] Two interesting results were the fruit of this analysis. We were able to identify all the 2-D Hamiltonians which possess a second integral of motion quadratic in the velocities, generalizing the results of Darboux,[2] Winternitz *et al.*,[3] and Fokas *et al.*[4] Seven distinct classes of Hamiltonians were found in all, three of which correspond to complex potentials. The singularity analysis of the equations of motion related to these Hamiltonians revealed the following interesting feature. Although associated to an integrable system the solutions in the complex-time plane did not possess the Painlevé property, i.e., their movable singularities were not just poles. Rather the singularities were cuts of the type $(t-t_0)^{1/n}$.

This was in contradiction with a naive generalization of the ARS conjecture.[5] That conjecture associated the Painlevé property to integrability for PDE's. Our counterexample shows that it is not valid as it stands for ordinary differential systems, namely such systems can be integrable without having the Painlevé property. This has led to the formulation of the "weak Painlevé" conjecture.[6]

We have thus conjectured that a two-dimensional Hamiltonian system will be integrable whenever the solutions present movable singularities of the form $(t-t_0)^{1/n}$, where n is determined solely by the dominant behavior of the system at the singularity. The exponent $1/n$ is called then a "natural" power and determines fully the nature of the algebraic singularities of the solutions.

The aim of the present paper is to show that for the above classes of Hamiltonians the solution of the equations of motions can be reduced to quadratures, through a proper change of variables. The singularity structure in this new system of coordinates will be examined in particular in relation to the weak Painlevé property.

[a)] The authors dedicate this work to their Soviet colleague Alexander Yoffe.

I. EXPLICIT INTEGRATION OF THE EQUATIONS OF MOTION

In Ref. 1 we have analyzed the question of integrability for Hamiltonian systems in two dimensions, and found all the systems with a second integral quadratic in the velocities. Seven distinct classes of potentials have resulted, three of which corresponded to complex potentials. In this section we will show how the two integrals of motion can be combined in order to reduce the computations of the trajectories of the system to quadratures. We will limit ourselves to the more physical case of real potentials. (The treatment of the complex ones can be performed along the same lines).

The four classes of real potentials we have identified are the following:

$$V=F(x)+G(y), \tag{1}$$

$$V=F(\rho)+(1/\rho^2)G(\varphi), \tag{2}$$

$$V=\frac{F(\rho+y)+G(\rho-y)}{\rho}, \tag{3}$$

$$V=\frac{F(u)+G(v)}{u^2-v^2}, \tag{4}$$

where $x=\rho\sin\varphi$, $y=\rho\cos\varphi$, and
$2u^2=\rho^2+a^2+((\rho^2+a^2)^2-4a^2x^2)^{1/2}$ and
$2v^2=\rho^2+a^2-((\rho^2+a^2)^2-4a^2x^2)^{1/2}$.

The integrals of motion associated to these systems are the following:

$$C=\dot x^2+2F(x), \tag{5}$$

$$C=(\dot xy-\dot yx)^2+2G(\varphi), \tag{6}$$

$$C=\dot x(x\dot y-y\dot x)+\frac{(\rho+y)G(\rho-y)-(\rho-y)F(\rho+y)}{\rho}, \tag{7}$$

$$C=(x\dot y-y\dot x)^2+a^2\dot x^2+\frac{2[u^2F(u)-v^2G(v)]}{v^2-u^2}. \tag{8}$$

The integration of case (1) is immediate. The potential is separable in x and y. The Hamiltonian in each direction is a constant

$$E_1 = \tfrac{1}{2}\dot{x}^2 + F(x),$$
$$E_2 = \tfrac{1}{2}\dot{y}^2 + G(y),$$

and the integration for the trajectory straightforward. One thus obtains $t = t(x)$ and $t = t(y)$ which can in principle be inverted to give $x = x(t)$, $y = y(t)$.

Case (2) is also quite simple. Rewriting the Hamiltonian and the constant C in polar coordinates, we have

$$H = \tfrac{1}{2}(\dot{\rho}^2 + \rho^2\dot{\varphi}^2) + F(\rho) + (1/\rho^2)G(\varphi),$$
$$C = \rho^4\dot{\varphi}^2 + 2G(\varphi).$$

Combining the two constants of motion, we obtain

$$\dot{\rho}^2 = 2E - 2F(\rho) - C/\rho^2,$$
$$\dot{\varphi}^2 = \frac{C - 2G(\varphi)}{\rho^4}.$$

Although this is not completely separable in the usual sense, it is still true that the variable ρ separates out. Its integration can be performed by a simple quadrature. Once $\rho(t)$ is found, the equation for φ can also be reduced to a quadrature:

$$\int \frac{d\varphi}{\sqrt{C - 2G(\varphi)}} = \pm \int \frac{dt}{\rho^2(t)}.$$

Case (3) can be treated in a natural way in parabolic coordinates. We put

$$x = 2\xi\eta, \qquad y = \xi^2 - \eta^2. \tag{9}$$

The two constants of motion can be written in this coordinate system as

$$E = 2(\xi^2 + \eta^2)(\dot{\xi}^2 + \dot{\eta}^2) + \frac{F(\xi) + G(\eta)}{\xi^2 + \eta^2},$$
$$C = 2(\xi^2 + \eta^2)(\dot{\xi}^2\eta^2 - \dot{\eta}^2\xi^2) + \frac{\eta^2 F(\xi) - \xi^2 G(\eta)}{\xi^2 + \eta^2}. \tag{10}$$

Solving for $\dot{\xi}$, $\dot{\eta}$ we readily obtain

$$\frac{d\xi}{\sqrt{\xi^2 E + C - F(\xi)}} = \pm \frac{d\eta}{\sqrt{\eta^2 E - C - G(\eta)}} = \pm \frac{dt}{\sqrt{2}(\xi^2 + \eta^2)}. \tag{11}$$

We thus remark that although the *system* is not separable in parabolic coordinates, the equation for the *trajectories* does indeed separate. Once this equation is integrated to yield $\xi = \xi(\eta)$, the time dependence is then also reduced to a quadrature.

Finally case (4) can be treated in elliptical coordinates. We put

$$x = a\cosh\xi\cos\eta, \quad u = a\cos\eta,$$
$$y = a\sinh\xi\sin\eta, \quad v = a\cosh\xi.$$

We obtain thus for the two integrals of motion

$$E = \frac{a^2}{2}(\dot{\xi}^2 + \dot{\eta}^2)(\sinh^2\xi + \sin^2\eta) + \frac{F(\xi) - G(\eta)}{a^2(\sinh^2\xi + \sin^2\eta)},$$
$$C = a^4(\sinh^2\xi + \sin^2\eta)[\cos^2\eta\,\dot{\xi}^2 + \cosh^2\xi\,\dot{\eta}^2] + 2\frac{[\cos^2\eta\, F(\xi) - \cosh^2\xi\, G(\eta)]}{(\cosh^2\xi + \sin^2\eta)}.$$

Analogous to case (3) we obtain

$$\frac{d\xi}{\sqrt{-C + 2a^2E\cosh^2\xi - 2F(\xi)}} = \pm \frac{d\eta}{\sqrt{C - 2a^2E\cos^2\eta + 2G(\eta)}} = \pm \frac{dt}{a^2(\sinh^2\xi + \sin^2\eta)}.$$

Once again the equation for the trajectories separates and the solution is reduced to quadratures.

II. EXAMPLES

In what follows, we present some examples of integration of the equations of motion for potentials belonging to the case (3). We start with the reduced equations of motion (11). We restrict ourselves to potentials polynomial in x and y. This amounts to taking $F(\xi) = P(\xi^2)$ and $G(\eta) = -P(-\eta^2)$, where P is a polynomial. The reason for this choice is that this class is nontrivial while being simpler than case (4). Several systems of physical interest belong to this class (see Refs. 1 and 7) and moreover the discussion of the weak Painlevé property has been (and will be) based on examples from this class.

In this paragraph, we further restrict ourselves to the case where P is a monomial, i.e., $F(\xi) = \xi^{2(m+1)}$ and $G(\eta) = (-1)^m\eta^{2(m+1)}$, $C = 0$, in order to simplify the calculations:

$$\frac{d\xi}{\xi\sqrt{E - \xi^{2m}}} = \frac{d\eta}{\eta\sqrt{E - (-1)^m\eta^{2m}}} = \frac{dt}{\sqrt{2}(\xi^2 + \eta^2)} \tag{12}$$

[with the + choice of signs in (11)]. Putting $\xi^2 = \chi$, $\eta^2 = \psi$, we get

$$\frac{d\chi}{\chi\sqrt{E - \chi^m}} = \frac{d\psi}{\psi\sqrt{E - (-1)^m\psi^m}} = \frac{\sqrt{2}dt}{(\chi + \psi)}. \tag{13}$$

A remark can be made at this point, as far as the trajectory is concerned. If $m = 2n$, one remarks that Eq. (12) for ξ, η and n is the same as Eq. (13) for χ, ψ and m. So one obtains identical trajectories for potentials of degree n and $m = 2n$ for n even and, of course, $C = 0$.

We consider first a potential of degree $m = 2^p$. Equation (12) can be transformed, putting $\xi^m = \chi$, $\eta^m = \psi$ to the following:

$$\frac{d\chi}{\chi\sqrt{E - \chi^2}} = \frac{d\psi}{\psi\sqrt{E - \psi^2}}. \tag{14}$$

Integrating (14) we obtain for the trajectories

$$\psi^2\chi^2(\lambda^2 - 1)^2 = 4E[\lambda\chi\psi(\lambda - 1)^2 - \lambda^2(\chi - \psi)^2]. \tag{15}$$

Note that for real χ and ψ, the integration constant λ must be positive. If $m = 2$, $\chi = \xi^2$, $\psi = \eta^2$, one can easily find the trajectories in terms of x and y. Equation (15) becomes

$$x^4(\lambda^2 - 1)^2 = 16E(\lambda x^2(\lambda - 1)^2 - 4\lambda^2 y^2). \tag{16}$$

Indeed, for $m = 2$ the potential is harmonic,

$V = y^2 + x^2/4.$

The ratio of the frequencies is 2. Choosing $C = 0$ corresponds to taking the trajectory that goes through the origin and (16) can indeed be parametrized by

$$x = A \sin \Theta,$$
$$y = B \sin 2\Theta,$$

i.e., the correct Lissajoux curve for ratio 2.

In a similar way one can obtain the analytic form of the trajectories in terms of x and y for $m = 2p$ for any p. For example, for $m = 4$ we find

$$\left(\frac{x^2}{4}\right)^4(\lambda^2 - 1)^2 = 4E\left[\lambda\left(\frac{x^2}{4}\right)^2(\lambda - 1)^2 - \lambda^2 y^2(x^2 + y^2)\right].$$

However, further explicit integration (12) for the time dependence is not possible in general, unless $m = 2$ of course.

Another case where we can give an expression for the trajectories is the Henon–Heiles potential for the choice of parameters for which J. Greene[7] has given the second constant of motion. Treating just the homogeneous cubic part of the potential, we have $m = 3$ in expression (12). Introducing $E = a^3$, $\chi = (1/a)\xi^2$, $\psi = (1/a)\eta^2$, we obtain

$$\int_\chi^1 \frac{d\chi}{\chi\sqrt{1-\chi^3}} = \int_\psi^\infty \frac{d\psi}{\psi\sqrt{1+\psi^3}} + \text{const.} \tag{17}$$

The integration of (17) is straightforward but tedious. We obtain finally, the trajectory in terms of elliptic functions

$$(u + v)(1 + \alpha) + \alpha(f(u) - f(v))$$
$$- \Pi(\varphi,\gamma,k) - \Pi(\omega,\gamma,k) = c,$$

where

$$\cos\varphi = \text{cn}\, u = \frac{\sqrt{3} - 1 + \chi}{\sqrt{3} + 1 - \chi},$$
$$\cos\omega = \text{cn}\, v = \frac{\psi + 1 - \sqrt{3}}{\psi + 1 + \sqrt{3}},$$
$$\gamma = \frac{1}{2} + \frac{1}{\sqrt{3}}, \quad \alpha = \frac{1 + \sqrt{3}}{1 - \sqrt{3}},$$
$$k = \frac{\sqrt{2 + \sqrt{3}}}{2}\,(k'^2 + k^2 = 1),$$

and

$$f(w) = \left(\frac{\alpha^2 - 1}{k^2 + k'^2\alpha^2}\right)^{1/2}$$
$$\times \ln\left[\frac{\sqrt{k^2 + k'^2\alpha^2}\,\text{dn}\, w + \sqrt{\alpha^2 - 1}\,\text{sn}\, w}{\sqrt{k^2 + k'^2\alpha^2}\,\text{dn}\, w - \sqrt{\alpha^2 - 1}\,\text{sn}\, w}\right].$$

As in the $m = 2p$ case the integration for the time dependence does not seem possible, given the complexity of the expression for the trajectory.

III. SINGULARITY STRUCTURE OF THE EQUATIONS OF MOTION

The structure of singularities for case (c) where

$$V = \frac{F(\sqrt{x^2 + y^2} + y) + G(\sqrt{x^2 + y^2} - y)}{\sqrt{x^2 + y^2}} \tag{18}$$

has been investigated in Ref. 1, using the ARS algorithm,[5] in the special case where V is a polynomial in terms of x and y, i.e., $F(u)$ is a polynomial and $G(u) = -F(-u)$. In this section we will show directly how this structure follows from (11), which now writes

$$\frac{d\xi}{\sqrt{\xi^2 E + C - F(\xi^2)}} = \pm \frac{d\eta}{\sqrt{\eta^2 E - C + F(-\eta^2)}}$$
$$= \pm \frac{dt}{\sqrt{2}(\xi^2 + \eta^2)}, \tag{19}$$

where

$$y = \xi^2 - \eta^2,$$
$$x = 2\xi\eta.$$

The most general singularity arises when, for some value t_0 of t, one of the variables, say ξ, diverges while the other does not. If F is of degree $p + 3$ of its argument (i.e., degree $2p + 6$ in ξ) then the leading singularity is of the form

$$\xi \sim (t - t_0)^{-1/p}.$$

To obtain an expansion of ξ in terms of $(t - t_0)$, we first find η in terms of ξ, then $(t - t_0)$ in terms of ξ and $\eta(\xi)$ and invert. This expansion of ξ around t_0 will be of the form

$$\xi = (t - t_0)^{-1/p} \sum a_n (t - t_0)^{n/p}, \tag{20}$$

similarly

$$\eta = \sum b_n (t - t_0)^{n/p}. \tag{21}$$

If $p = 1$, this expansion is of Painlevé type. If $p = 2$, we will see later that only a_n's and b_n's for even n do not vanish and η and ξ^2 have Painlevé-type expansions, although ξ does not. For other values of p, there does not exist any power of ξ or η that exhibits the Painlevé property. Indeed ξ^p would have a pure pole only if all the a_n's vanish except those with n a multiple of p, and we will show that this is not possible.

In terms of x and y, one immediately sees that y diverges as $(t - t_0)^{-2/p}$ and x as $(t - t_0)^{-1/p}$. (Note that the degree of V in terms of x and y is $p + 2$). This corresponds to one of the possible singular behaviors studied in Ref. 1, namely the only one that has the maximal dimensionality, i.e., that depends on four arbitrary constants. In that paper we showed that two of these constants were t_0 and the coefficient of the dominant term of x, and that the two others were related to what we called the "resonances." Here we will show directly how these four arbitrary constants follow from (19) and thus illustrate the meaning of the "resonances."

The four free constants in integrating the equations of motion are E, C, and the two integration constants of (19). One will be taken as the value t_0 of t where ξ diverges, the other as the value η_0 of η at t_0. Where do these parameters enter in the expansion of ξ? The first one, t_0, is obvious. Then, η_0 enters because dt is proportional to $\xi^2 + \eta^2$. From this one sees that η_0 determines a_2 in (20).

Near t_0, at the lowest order, η behaves as

$$\eta \sim \eta_0 + (\eta_0^2 E - C + F(-\eta_0^2))(t - t_0)^{1 + 2/p}.$$

Thus, E and C will first appear in the expansion of η in b_{p+2}, through the combination $(\eta_0^2 E - C)$. In the expansion of ξ, because of the $(\xi^2+\eta^2)$ factor in dt, it will appear two orders later, namely in a_{p+4}. This is the only combination of E and C that will enter until the order $(t-t_0)^{2+4/p}$. At that order, because of the $\xi^2 E$ term in the square root under $d\xi$, E will determine a_{2p+4}. Indeed $\xi^2 E$ is smaller than $F(\xi^2)$ by ξ^{2p+4}, F being of degree $p+3$. Consider two solutions for which ξ diverges at the same value t_0 of t, that have the same value η_0 for η at t_0, and different values of E and C, chosen in such a way that $\eta_0^2 E - C$ has the same value. Then, their respective expansions for ξ will be identical up to a_{2p+3}.

In summary, the four arbitrary constants t_0, η_0, E, C, are equivalent to t_0, a_2, a_{p+4} (or b_{p+2}) and a_{2p+4}.

These are precisely the arbitrary constants which resulted from the singularity analysis in Ref. 1. Indeed, going back to the x,y coordinates, η_0 corresponds to the freedom of the coefficient of x, b_{p+2} (i.e., a_{p+4}) corresponds to the resonance $(1+2/p)$ in x, and a_{2p+4} corresponds to the resonance $2+4/\mathrm{p}$. The coefficients of the polynomial F will first appear in the expansion of ξ, one after the other, through a_{2k}, $k=2$ to $p+1$. As for the two lowest-order coefficients of F, their interpretation is the following. The only significant quantity is $F(u)-Eu-C$, as can be checked in (19). Changing the linear and constant terms in F just redefines E and C. One can immediately check that the corresponding potential V is shifted accordingly.

Note that if p is even, only a_n's and b_n's with even n will be nonvanishing, and thus η and ξ^2 have expansions in terms of $(t-t_0)^{1/p'}$, where $p'=p/2$, and in this case the "natural" power as defined in Ref. 8 is indeed $1/p'$ rather than $1/p$. In particular, if $p=2$, the expansion for η and ξ^2 is indeed of Painlevé type.

On the other hand, for $p>2$, ξ^p does not have a Painlevé-type expansion of full dimensionality, since the arbitrary parameter a_2 would destroy this property unless it is chosen equal to zero. For all p, however, the expansions are of what we called "weak Painlevé type."[6]

Similarly one could study the case where ξ and η diverge at the same value of t_0. These solutions have a lower dimensionality, namely three, because it corresponds to choosing η_0 equal to infinity. Again, one finds by direct study that the structure of the singularity in this case is exactly what has been found through the ARS algorithm in Ref. 1. There, the lower dimensionality was associated to the negative value of one of the resonances.

IV. DISCUSSION AND OUTLOOK

It is important to realize that the free parameters in the expansions of ξ and η appear at exactly the same powers of $(t-t_0)$ as in the expansions of x and y. This illustrates the intrinsic character of the resonances. The variables ξ and η are obviously the natural ones as they lead to an explicit integrability for the equations of motions, and even in these good variables, the weak Painlevé character persists (for $p>2$). From this, we surmise that no other change of variables will be able to restore the full Painlevé property. The answer to the question asked of us by M. D. Kruskal, namely whether the weak Painlevé character could not be converted to full Painlevé by an appropriate change of variable therefore appears to be negative.

However, it is possible to take the analysis one step further. If one takes F and G as two arbitrary functions of their respective arguments, rather than two closely related polynomials, we can introduce movable singularities of any character, without destroying the integrability of the system. Still, ξ and η will obviously remain the natural variables and it is quite improbable that these movable critical singularities will be removed by another choice of variables.

The fact that arbitrary movable singularities do not compromise integrability for two-dimensional Hamiltonian systems contradicts the weak-Painlevé conjecture. Therefore it would appear to make it useless as an integrability detector. This, however, is not true. The weak-Painlevé criterion, as a heuristic tool, keeps its predictive power.

Indeed suppose we start with two arbitrary functions $F(\xi)$ and $G(\eta)$. One can thus always construct a potential V for an integrable Hamiltonian problem in terms of x and y. However, in general, the variables ξ and η will be quite "visible" in the expression of V. Therefore, a cursory glance at V will suggest the right change of variables. Indeed, for general F and G we have

$$V=\left[F(y+\sqrt{x^2+y^2})+G(y-\sqrt{x^2+y^2})\right]/\sqrt{x^2+y^2}.$$

Except for very special F's and G's the radicals will never regroup. If this ever happens, as for example for

$$V=\frac{\ln x}{\sqrt{x^2+y^2}},$$

$(F(u)=G(u)=\frac{1}{2}\ln u)$ one would expect anyway the existence of a logarithmic singularity because the potential itself contains a logarithm. In that case, this singularity will not suggest nonintegrability.

On the other hand whenever V is such that x and y are the only "obvious" variables and moreover does not contain built-in singularity-generating terms, then integrability is conditioned by the "weak-Painlevé" property.

Another interesting question has been asked of us by A. Fokas, on whether the weak-Painlevé property was always associated to systems for which the second constant of motion is quadratic in velocities. In a future publication, we will show that this is not the case, by exhibiting several weak-Painlevé systems with nonquadratic second integrals of motion.

Finally, the above analysis will serve to shed a new light on the relationship between integrability and the analytic properties of the solutions of the equations of motion.[9]

ACKNOWLEDGMENTS

We wish to express our gratitude to M. D. Kruskal for many challenging discussions and for asking the question that this paper set up to answer.

We also acknowledge interesting discussions with P. Winternitz and A. Fokas.

[1]B. Dorizzi, B. Grammaticos, and A. Ramani, "A new class of integrable systems," J. Math. Phys. **24**, 2282 (1983).

[2]G. Darboux, Archives Néerlandaises (ii) VI, 371 (1901), E. T. Whittaker, *Analytical Dynamics of Particles*, 4th ed. (Cambridge University, Cam-

bridge, England, 1959), pp. 334–335.
[3]P. Winternitz, Ya A. Smorodinski, M. Uhlir, and I. Fris, Yad. Fiz. **4**, 625 (1966) [Sov. J. Nucl. Phys. **4**, 444 (1967)].
[4]A. S. Fokas, J. Math. Phys. **68**, 347–370 (1979).
[5]M. J. Ablowitz, A. Ramani, and H. Segur, Lett. Nuovo Cimento **23**, 333 (1978); M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715 (1980); M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 1006 (1980).
[6]A. Ramani, B. Dorizzi, and B. Grammaticos, Phys. Rev. Lett. **49**, 21, 1539 (1982).
[7]J. Greene, quoted in Y. F. Chang, M. Tabor, and J. Weiss, J. Math. Phys. **23**, 531 (1982).
[8]B. Grammaticos, B. Dorizzi, and A. Ramani, "Integrability of Hamiltonians with third- and fourth-degree polynomial potentials," J. Math. Phys. **24**, 2289 (1983).
[9]A. Ramani, B. Dorizzi, B. Grammaticos, and T. Bountis "Integrability and Painlevé property for low-dimensional systems," J. Math. Phys. (in press).

First integrals for some nonlinear time-dependent Hamiltonian systems

P. G. L. Leach
Department of Applied Mathematics, La Trobe University, Bundoora 3083, Australia
H. R. Lewis
CTR-6, MS-F642, Los Alamos National Laboratory, P.O. Box 1663, Los Alamos, New Mexico
W. Sarlet
Instituut voor Theoretische Mechanika, Rijksuniversiteit Gent, Krijgslaan 281-S9, B-9000 Gent, Belgium

(Received 20 July 1982; accepted for publication 11 March 1983)

An observation of a simple property of canonical transformations leads to a procedure for determining first integrals for classes of Hamiltonians. In illustrative examples the most general result presently known from other methods is recovered, a new result presented, and a generalization to more than one degree of freedom discussed.

PACS numbers: 03.20. + i, 02.30. + g

1. INTRODUCTION

If at a time, say $t = 0$, a Hamiltonian system is in the state $(q(0), p(0))$, it is a truism that the set of constants, $\{q_i(0), p_i(0): i = 1, n\}$, may be associated with the subsequent time evolution of the system through infinitesimal time translations generated by the Hamiltonian. In principle, we may write

$$\begin{aligned} q_i(0) &= f_i(q(t), p(t), t), \\ p_i(0) &= g_i(q(t), p(t), t) \end{aligned} \tag{1.1}$$

so that the f_i and g_i may then be called first integrals. Any other first integral of the motion may be expressed as a function of the f_i and g_i. Unfortunately, we must stress the phrase "in principle." When we come to treat an actual system, knowing that first integrals exist will not guarantee success in finding them.

Over the last several years there has grown a considerable literature devoted to the determination of first integrals for explicitly time-dependent dynamical systems. The methods used have been various: canonical transformations,[1-5] general dynamical symmetries and Noether's theorem,[6-11] direct *ad hoc* approaches,[12-14] and the study of systems with Ermakov-type coupling.[15] The usual goal when applying any of the methods is either to find explicit formulae for first integrals of Hamiltonian systems or to provide a basis for some computational procedure or both. The methods listed above may be stated in general terms. In order to obtain results with a particular method some restriction on the form of the first integral, the nature of the Hamiltonian or the way in which the method is applied must be introduced. By way of example, in the first paper of this series,[5] canonical transformations were used to find an exact invariant and the related potential for Hamiltonians of the form

$$H = \tfrac{1}{2}p^2 + V(q, t). \tag{1.2}$$

The restriction was to take the canonical transformation as

$$Q = Q(q, p, \rho(t)), \quad P = P(q, p, \rho(t), \dot\rho(t)), \tag{1.3}$$

where $\rho(t)$ is some function to be determined, the so-called auxiliary function. Furthermore, the transformed Hamiltonian was specified by

$$K = \beta(t)P. \tag{1.4}$$

The first integral was found to be quadratic in the momentum and a more direct approach to obtain first integrals polynomial in the momentum was adopted in the second paper of the series.[13] This simply required the solution of the equation

$$\frac{dI}{dt} = \frac{\partial I}{\partial t} + [I, H] = 0, \tag{1.5}$$

where

$$I(q, p, t) = \sum_{i=0}^{n} p^i f_i(q, t) \tag{1.6}$$

and H has the form of (1.2). The result provided a generalization of that found in Ref. 5 and has been confirmed subsequently in Refs. 14 and 9 by other methods.

Restrictions similar to those mentioned above can be seen in each of the methods which have been used to obtain a system for which a first integral is explicitly known. The application of canonical transformations in Refs. 2 and 5 was restricted in regard to the nature of the time dependence of the transformation. In Refs. 13–15 it is rather the assumed structure of the invariant or of the Hamiltonian which is restricted, while in Ref. 9 the infinitesimal transformation was confined to being a point transformation. There is now the question of whether the restrictions were *necessary* to obtain results. Furthermore, we may ask which method, under a relaxed restriction, will lead to a more general result with the least possible effort. Of one thing we may be nearly certain: If we perceive how to relax a restriction and achieve a more general result by one method, it will soon become obvious how to obtain the same result using other methods.

The present paper may be viewed as being motivated by the remarks at the end of the previous paragraph as applied to the first paper in this series.[5] The restrictions of that paper are summarized in Eqs. (1.2)–(1.4). That the theoretical content of the present paper is somewhat different from that of Ref. 5 is due to a simple observation to be made in Sec. 2 below. However, the starting point was to see how much could be determined when the restrictions (1.2) and (1.3)

 0022-2488/84/030486-05$02.50

were removed. The theoretical treatment is applied to a system of one degree of freedom. This is motivated solely by a desire for clarity. In the examples we show how the results of previous papers can be recovered, present a result not given in those papers, and indicate how the method may be extended to systems of more than one degree of freedom.

2. THEORETICAL DEVELOPMENT

As indicated above, for purposes of simplicity the discussion is limited to systems with one degree of freedom. There is no great difference for the case of a multidimensional system. The general result will be given in Sec. 4 in which we also treat a two-dimensional example.

Let

$$(q,p)\to(Q,P\colon Q=Q(q,p,t),\quad P=P(q,p,t),\quad [Q,P]_{qp}=1) \tag{2.1}$$

be a canonical transformation which transforms the Hamiltonian $H(q,p,t)$ to the new Hamiltonian $K(Q,P,t)$. In terms of the generating function $F(q,p,t)$ defined through

$$\frac{\partial F}{\partial q}=p-P\frac{\partial Q}{\partial q},\quad \frac{\partial F}{\partial p}=-P\frac{\partial Q}{\partial p}, \tag{2.2}$$

the two Hamiltonians are related according to Ref. 5:

$$K(Q(q,p,t),P(q,p,t),t)=H+P\frac{\partial Q}{\partial t}+\frac{\partial F}{\partial t}. \tag{2.3}$$

The integrability of (2.2) is guaranteed by the canonicity of the transformation so that (2.3) is a well-defined relation at least locally. If we define

$$\zeta(q,p,t)=-P\frac{\partial Q}{\partial t}-\frac{\partial F}{\partial t}, \tag{2.4}$$

(2.3) may be written as

$$H=K(Q(q,p,t),P(q,p,t),t)+\zeta(q,p,t). \tag{2.5}$$

We now make an observation so simple that it is not found in the standard texts. The new canonical variables and the function $\zeta(q,p,t)$ are related by

$$\frac{\partial Q}{\partial t}+[Q,\zeta]=0,\qquad \frac{\partial P}{\partial t}+[P,\zeta]=0. \tag{2.6}$$

One way of looking at (2.6) is to state that Q and P are first integrals of $\zeta(q,p,t)$ regarded as the Hamiltonian of a dynamical system. However, as we started with the canonical transformation (2.1), we may view (2.6) as defining the function $\zeta(q,p,t)$ in terms of its derivatives. Thus we may rearrange (2.6) to give

$$\frac{\partial \zeta}{\partial q}=[Q,P]_{qt},\quad \frac{\partial \zeta}{\partial p}=[Q,P]_{pt} \tag{2.7}$$

where

$$[Q,P]_{\alpha\beta}=\frac{\partial Q}{\partial \alpha}\frac{\partial P}{\partial \beta}-\frac{\partial Q}{\partial \beta}\frac{\partial P}{\partial \alpha}. \tag{2.8}$$

The consistency of the two equations in (2.7) follows from (2.1). Equation (2.7) defines $\zeta(q,p,t)$ up to an additive function of time which may be disregarded.

The basic theoretical result may now be stated. If $I(Q,P)$ is some function of Q and P and we may write

$$K(Q,P,t)=R(I(Q,P),t), \tag{2.9}$$

then $I(Q,P)$ is a first integral of $H(q,p,t)$. Alternatively we could state the result as follows. Given a canonical transformation of the type (2.1), we may construct a $\zeta(q,p,t)$ according to (2.7). Then there exists a family of Hamiltonians $H(q,p,t)$ with first integrals $I(Q(q,p,t),P(q,p,t))$, where I is an arbitrary function, defined by

$$H(q\,p,t)=R(I(Q,P),t)+\zeta(q,p,t). \tag{2.10}$$

This form of statement of our result emphasizes a constructive approach. In essence, we are determining the classes of Hamiltonians which possess a first integral associated with a particular selection of the canonical transformation (2.1).

The viewpoint of the previous paragraph is not the only one which can be adopted. We could, for instance, go back to (2.6) and imagine starting with an $\zeta(q,p,t)$ and solving those equations for Q and P. This is equivalent to solving Hamilton's equations with $\zeta(q,p,t)$ as the Hamiltonian. The difficulties associated with any attempting to solve Hamilton's equations for any but the simplest (time-dependent) systems suggest that this would not be a fruitful approach.

The general procedure to be followed in the applications given below is this. A pair of functions $Q(q,p,t)$ and $P(q,p,t)$ which are canonically conjugate are chosen. The function $\zeta(q,p,t)$ is determined from (2.7) and the family of integrable Hamiltonians follows from (2.10). (Strictly for one degree of freedom only; in the case of more than one degree of freedom it would be necessary to find more than one first integral.) In particular, because of the differentiation between q and p to be found in physical Hamiltonians, we could envisage establishing hierarchies of families of Hamiltonians according to the nature of the dependence on p chosen for Q and P.

We shall illustrate this point in Sec. 3 and conclude this theoretical discussion with the following remarks:

(i) In selecting the canonically conjugate variables Q and P, it is essential that they contain explicit time dependence when written in terms of q, p, and t. If Q and P do not contain t, then I is just a function of q and p and $\zeta(q,p,t)$, from (2.7), just an ignorable function of time. The relation (2.10) then does not produce a particularly interesting class of Hamiltonians.

(ii) In practical applications, the considerations above often would have to be supplemented by a third stage. Once a pair (Q,P) had been chosen and a family of appropriate Hamiltonians of the form (2.10) constructed, one would be left with the task of matching a given Hamiltonian $H(q,p,t)$ with a member of the family (2.10) for a suitable choice of the functions R and I. It would be aniticipated that this matching would lead to the introduction of auxiliary differential equations of the type which have played such an important role in previous work.

(iii) There is a context beyond the scope of the present work in which Eq. (2.6) could be preferred to (2.7). It may be imagined that similar ideas could be developed in the context of perturbation methods. The function $\zeta(q,p,t)$, treated as a

Hamiltonian, could be solved order by order with respect to some small parameter. Such a perturbation technique would be quite unconventional because one would not begin by solving the zeroth-order part of the equations of motion corresponding to H.

3. EXAMPLES WITH ONE DEGREE OF FREEDOM

In these illustrative examples we shall consider simple choices for Q and P and construct $\zeta(q,p,t)$. Rather than simply substitute this into the general relationship (2.10) we shall further impose a specific form on $H(q,p,t)$. This is the familiar

$$H=\tfrac{1}{2}p^2+V(q,t), \tag{3.1}$$

which arises so frequently in applications. Our intention is to go part-way along the procedure outlined in remark (ii) in Sec. 2 in that we shall determine the structure of the potential along with the associated first integral. In view of the opinion expressed in remark (i), we take p as a preferred coordinate.

Let the canonical variables be

$$Q=a(q,t),\quad P=b(q,t)p+c(q,t), \tag{3.2}$$

where

$$\frac{\partial a}{\partial q}b=1. \tag{3.3}$$

Equation (2.7) becomes

$$\frac{\partial\zeta}{\partial q}=p[a,b]_{qt}+[a,c]_{at}, \tag{3.4}$$

$$\frac{\partial\zeta}{\partial p}=-b\frac{\partial a}{\partial t}. \tag{3.5}$$

It follows that

$$\zeta(q,t)=-bp\frac{\partial a}{\partial t}+\int^q[a,c]_{q't}\,dq' \tag{3.6}$$

and so

$$H=R(I(Q,P),t)-bp\frac{\partial a}{\partial t}+\int^q[a,c]_{q't}\,dq'. \tag{3.7}$$

This gives, for various R, the class of Hamiltonians with first integral $I(Q,P)$.

We now look for a first integral quadratic in the momentum for a Hamiltonian of the type in (3.1). Without loss of generality, we may take

$$R(I,t)=I/\rho^2, \tag{3.8}$$

$$I(Q,P)=\tfrac{1}{2}A(Q)P^2+C(Q). \tag{3.9}$$

Substituting for H, R, I, P, and Q in (3.7) and separating by coefficients of powers of p, we find that

$$A^2b^2=\rho^2, \tag{3.10}$$

$$A^2c=\rho^2\frac{\partial a}{\partial t}\quad(b\neq 0), \tag{3.11}$$

$$V(q,t)=\frac{1}{\rho^2}(\tfrac{1}{2}A^2c^2+C)+\int^q[a,c]_{q't}\,dq'. \tag{3.12}$$

From (3.3) and (3.10) we see that

$$a(q,t)=E((q-\alpha)/\rho), \tag{3.13}$$

$$A(a)=E', \tag{3.14}$$

$$b(q,t)=\rho/E'((q-\alpha)/\rho), \tag{3.15}$$

where E is an arbitrary function of its argument and $\alpha(t)$ an arbitrary function of time. It follows immediately that

$$c(q,t)=-[\dot\alpha\rho+\dot\rho(q-\alpha)]/E' \tag{3.16}$$

and after some manipulation that

$$V(q,t)=(1/\rho^2)U((q-\alpha)/\rho)-\ddot\alpha\rho-\tfrac{1}{2}\rho\ddot\rho[(q-\alpha)/\rho]^2, \tag{3.17}$$

where U is an arbitrary function of its argument, to within an arbitrary function of time. The first integral is

$$I=\tfrac{1}{2}[\rho(p-\dot\alpha)-\dot\rho(q-\alpha)]^2+U((q-\alpha)/\rho). \tag{3.18}$$

This result replicates formulae found in Refs. 9, 11, and 12.

For our second example we intend to move firmly away from a first integral which is quadratic in the momentum. We take Q and P to both be linear in the momentum by defining them as

$$Q=a(q,t)p+b(q,t),\quad P=T(t)Q-c(q,t), \tag{3.19}$$

where the requirement of canonicity is satisfied by

$$\frac{\partial c}{\partial q}=\frac{1}{a} \tag{3.20}$$

and $T(t)$ a nonzero function of time. The first integral is defined by

$$I(Q,P)=P/Q. \tag{3.21}$$

Equation (2.7) is now

$$\begin{aligned}\frac{\partial\zeta}{\partial q}&=\frac{1}{2}\dot T\frac{\partial}{\partial q}(ap+b)^2-p[a,c]_{qt}-[b,c]_{qt},\\ \frac{\partial\zeta}{\partial p}&=\frac{1}{2}\dot T\frac{\partial}{\partial p}(ap+b)^2-a\frac{\partial c}{\partial t}\end{aligned} \tag{3.22}$$

so that

$$\zeta=\frac{1}{2}\dot T(ap+b)^2-ap\frac{\partial c}{\partial t}-\int^q[b,c]_{q't}\,dq'. \tag{3.23}$$

Again we confine our attention to a Hamiltonian of the type (3.1). The function $R(I,t)$ has to be such that we may write

$$\begin{aligned}R(I,t)&=R\left(T-\frac{c}{ap+b},t\right)\\&=\frac{1}{2}p^2+V(q,t)-\frac{1}{2}\dot T(ap+b)^2+ap\frac{\partial c}{\partial t}\\&\quad+\int^q[b,c]_{q't}\,dq'.\end{aligned} \tag{3.24}$$

As the right-hand side of (3.24) is polynomial in p, in terms of p, $R(I,t)$ must be polynomial in p. Observing that

$$T-I=c/(ap+b), \tag{3.25}$$

we may write

$$\begin{aligned}R(I,t)&=S((T-I)^{-1},t)\\&=S((ap+b)/c,t),\end{aligned} \tag{3.26}$$

where S is polynomial in its first argument. In fact it is sufficient to take S to be linear in $(ap+b)/c$, i.e.,

$$S((ap+b)/c,t)=\gamma(t)(ap+b)/c. \tag{3.27}$$

Substituting and separating coefficients of like powers of p, we have

$$1 - \dot{T}a^2 = 0, \tag{3.28}$$

$$-\dot{T}b + \frac{\partial c}{\partial t} = \frac{\gamma}{c} \quad (a \neq 0), \tag{3.29}$$

$$V = \frac{1}{2}\dot{T}b^2 - \int^q [b, c]_{q't} + \frac{\gamma b}{c}. \tag{3.30}$$

Defining $\rho(t)$ as

$$T(t) = \int^t \rho^{-2}(t')\,dt', \tag{3.31}$$

from (3.28) and (3.20), respectively,

$$a = \rho, \quad c = (q - \alpha)/\rho, \tag{3.32}$$

where $\alpha(t)$ is an arbitrary function of time. From (3.29)

$$b = -\alpha\dot{\rho} - \dot{\rho}(q - \alpha) - \gamma\rho^3/(q - \alpha) \tag{3.33}$$

and so the potential is

$$V(q, t) = \frac{1}{2}\frac{\ddot{\rho}}{\rho}(q - \alpha)^2 - \ddot{\alpha}(q - \alpha) - \dot{\sigma}\log(q - \alpha) - \frac{1}{2}\frac{\sigma^2}{(q - \alpha)^2}, \tag{3.34}$$

where

$$\sigma = \gamma\rho^2. \tag{3.35}$$

The first integral is

$$I = T - \frac{(q - \alpha)/\rho}{\rho(p - \dot{\alpha}) - \dot{\rho}(q - \alpha) - \sigma\rho/(q - \alpha)}. \tag{3.36}$$

It will be observed that the potential (3.34) is not of the class (3.17) because it contains a third arbitrary funciton of time, $\sigma(t)$. The same potential was derived earlier by Sarlet,[16] who used a generalization to time-dependent transformations of the techniques in Ref. 2. It has also been found by yet another method by Lewis and Leach.[17]

4. MORE THAN ONE DEGREE OF FREEDOM

The extension of the considerations of Sec. 2 to a system with more than one degree of freedom is straightforward. Under the canonical transformation

$$(\mathbf{q}, \mathbf{p}) \to (\mathbf{q}, \mathbf{P}\text{: } \mathbf{Q} = \mathbf{Q}(\mathbf{q}, \mathbf{p}, t), \quad \mathbf{P} = \mathbf{P}(\mathbf{q}, \mathbf{p}, t), \ [Q_i, P_j] = \delta_{ij}) \tag{4.1}$$

the old and new Hamiltonians are related according to

$$H(\mathbf{q}, \mathbf{p}, t) = K(\mathbf{Q}(\mathbf{q}, \mathbf{p}, t), \mathbf{P}(\mathbf{q}, \mathbf{p}, t), t) + \zeta(\mathbf{q}, \mathbf{p}, t), \tag{4.2}$$

where the function $\zeta(\mathbf{q}, \mathbf{p}, t)$ is defined according to the relations

$$\frac{\partial \zeta}{\partial q_i} = [Q_j, P_j]_{q_i t}, \tag{4.3}$$

$$\frac{\partial \zeta}{\partial p_i} = [Q_j, P_j]_{p_i t}. \tag{4.4}$$

The bracket relation has a slightly modified meaning compared with the given in (2.3). Now

$$[Q_j, P_j]_{\alpha t} = \sum_j \left(\frac{\partial Q_j}{\partial \alpha}\frac{\partial P_j}{\partial t} - \frac{\partial Q_j}{\partial t}\frac{\partial P_j}{\partial \alpha}\right). \tag{4.5}$$

Each $Q_j(\mathbf{q}, \mathbf{p}, t)$ and $P_j(\mathbf{q}, \mathbf{p}, t)$ is a first integral of $\zeta(\mathbf{q}, \mathbf{p}, t)$ regarded as a Hamiltonian system as is any function of them.

If we write K in the form

$$K(\mathbf{Q}, \mathbf{P}, t) = R(I_i(\mathbf{Q}, \mathbf{P}), t), \tag{4.6}$$

then the functions $I_i(\mathbf{Q}(\mathbf{q}, \mathbf{p}, t), \mathbf{P}(\mathbf{q}, \mathbf{p}, t))$ are first integrals of a class of Hamiltonians defined by

$$H(\mathbf{q}, \mathbf{p}, t) = R(I_i(\mathbf{Q}(\mathbf{q}, \mathbf{p}, t), \mathbf{P}(\mathbf{q}, \mathbf{p}, t)), t) + \zeta(\mathbf{q}, \mathbf{p}, t). \tag{4.7}$$

The remarks made at the end of Sec. 2 are equally applicable in the case of more than one degree of freedom. The Hamiltonian system will be completely integrable if there are n functions I_i in involution, where n is the number of degrees of freedom. More usually, we will be concerned with a Hamiltonian of specific structure, and the problem of integrability will be not so much the existence of n functions I_i in involution, but whether or not H can be expressed in terms of them by (4.7) with suitable selections of the function R.

We shall illustrate the procedure for determining the Hamiltonian and first integral with a particularly simple choice of a canonical transformation in an example with two degrees of freedom. Let the canonical transformation be

$$Q_1 = q_1/\alpha_1, \quad P_1 = \alpha_1 p_1 + \beta_1 q_1, \quad Q_2 = q_2/\alpha_2, \quad P_2 = \alpha_2 p_2 + \beta_2 q_2, \tag{4.8}$$

in which $\alpha_1, \alpha_2, \beta_1$, and β_2 are function of time only. Integration of Eqs. (4.3) and (4.4) yields

$$\zeta(q_1, q_2, p_1, p_2, t) = \frac{\dot{\alpha}_1}{\alpha_1}q_1p_1 + \frac{\dot{\alpha}_2}{\alpha_2}q_2p_2 + \frac{1}{2}\frac{(\alpha_1^{\cdot}\beta_1)}{\alpha_1^2}q_1^2 + \frac{1}{2}\frac{(\alpha_2^{\cdot}\beta_2)}{\alpha_2^2}q_2^2. \tag{4.9}$$

We shall consider the case of H of the form $T + V$ and desire to have one first integral quadratic in the momenta. The first integral will be of the form

$$I = \tfrac{1}{2}AP_1^2 + BP_1P_2 + \tfrac{1}{2}CP_2^2 + DP_1 + EP_2 + F, \tag{4.10}$$

where A to F are functions of Q_1 and Q_2 only. The function R will necessarily be of the form

$$R(I, t) = \beta(t)I \tag{4.11}$$

so that (4.7) becomes

$$\begin{aligned}\tfrac{1}{2}p_1^2 + \tfrac{1}{2}p_2^2 + V(q_1, q_2, t) &= \beta(t)(\tfrac{1}{2}AP_1^2 + BP_1P_2 + \tfrac{1}{2}CP_2^2 + DP_1 + EP_2 + F) \\ &\quad + \frac{\dot{\alpha}_1}{\alpha_1}q_1p_1 + \frac{\dot{\alpha}_2}{\alpha_2}q_2p_2 + \frac{1}{2}\frac{(\alpha_1\beta_1)^{\cdot}}{\alpha_1^2}q_1^2 \\ &\quad + \frac{1}{2}\frac{(\alpha_2\beta_2)^{\cdot}}{\alpha_2^2}q_2^2.\end{aligned} \tag{4.12}$$

Equating coefficients of like powers of p_1 and p_2 to zero, we have

$$\beta A\alpha_1^2 = 1, \quad \beta B\alpha_1\alpha_2 = 0, \quad \beta C\alpha_2^2 = 0,$$

$$\beta(A\alpha_1\beta_1q_1 + B\alpha_1\beta_2q_2 + D\alpha_1) + (\dot{\alpha}_1/\alpha_1)q_1 = 0,$$

$$\beta(B\beta_1\alpha_2q_1 + C\alpha_2\beta_2q_2 + E\alpha_2) + (\dot{\alpha}_2/\alpha_2)q_2 = 0,$$

$$\begin{aligned}V(q_1, q_2, t) &= \beta(\tfrac{1}{2}A\beta_1^2q_1^2 + B\beta_1\beta_2q_1q_2 \\ &\quad + \tfrac{1}{2}C\beta_2^2q_2^2 + D\beta_1q_1 + E\beta_2q_2 + F) \\ &\quad + \tfrac{1}{2}[(\alpha_1\beta_1)^{\cdot}/\alpha_1^2]q_1^2 + \tfrac{1}{2}[(\alpha_2\beta_2)^{\cdot}/\alpha_2^2]q_2^2.\end{aligned} \tag{4.13}$$

The first three of Eqs. (4.12) give

$$A = a^{-2}, \quad B = 0, \quad C = c^{-2},$$
$$\alpha_1 = a\rho, \quad \alpha_2 = c\rho, \tag{4.14}$$

where a and c are constants and $\beta(t)$ has been replaced by $\rho^{-2}(t)$. The next two equations give

$$D(Q_1, Q_2) = -Q_1(\rho\dot{\rho} + \rho\beta_1/a),$$
$$E(Q_1, Q_2) = -Q_2(\rho\dot{\rho} + \rho\beta_2/c), \tag{4.15}$$

and without loss of generality we may take D and E as zero by setting

$$\beta_1 = -a\dot{\rho}, \quad \beta_2 = -c\dot{\rho}. \tag{4.16}$$

The last of Eqs. (4.12) gives the potential as

$$V(q_1, q_2, t) = -\frac{1}{2}\frac{\ddot{\rho}}{\rho}q_1^2 - \frac{1}{2}\frac{\ddot{\rho}}{\rho}q_2^2 + \frac{1}{\rho^2}U\left(\frac{q_1}{\rho}, \frac{q_2}{\rho}\right), \tag{4.17}$$

where

$$U\left(\frac{q_1}{\rho}, \frac{q_2}{\rho}\right) = F\left(\frac{q_1}{a\rho}, \frac{q_2}{c\rho}\right). \tag{4.18}$$

The first integral is

$$I = \tfrac{1}{2}(\rho p_1 - \dot{\rho}q_1)^2 + \tfrac{1}{2}(\rho p_2 - \dot{\rho}q_2)^2 + U(q_1/\rho, q_2/\rho). \tag{4.19}$$

This result is easily generalized to n dimensions with translation included. We simply quote the result that a Hamiltonian of the form

$$H = \frac{1}{2}\mathbf{p}^2 + \frac{1}{\rho^2}V\left(\frac{\mathbf{q} - \boldsymbol{\alpha}}{\rho}\right) - \frac{1}{2}\frac{\ddot{\rho}}{\rho}(\mathbf{q} - \boldsymbol{\alpha})^2 - \ddot{\boldsymbol{\alpha}}\cdot\mathbf{q} \tag{4.20}$$

has the first integral

$$I = \tfrac{1}{2}[\rho(\mathbf{p} - \dot{\boldsymbol{\alpha}}) - \dot{\rho}(\mathbf{q} - \boldsymbol{\alpha})]^2 + V((\mathbf{q} - \boldsymbol{\alpha})/\rho). \tag{4.21}$$

5. CONCLUSION

In this paper we have outlined a procedure for the simultaneous construction of first integrals and Hamiltonians which possess those first integrals. In essence the method is based upon a reinterpretation of the relationship among an original Hamiltonian, the transformed Hamiltonian, and the canonical transformation relating the two. If the difference between the two Hamiltonians is treated as a Hamiltonian, it has as first integrals the new canonical coordinates. If the transformed Hamiltonian $K(Q, P, t)$ is written in the form $R(I(Q,P), t)$, then I is a first integral of the original Hamiltonian. In the first example considered we obtained the most general result [for Hamiltonians of type (1.2)] found in the recent literature. We have produced a new result with three arbitrary functions of time and have indicated how these results may be extended to problems of more than one degree of freedom.

In the particular case of Hamiltonians which have the form $T + V$, to extend the type of potential for which a first integral exits, it is necessary to go to a first ingetral which is not quadratic in the momentum. The extent of possible further results is determined by the choice of canonically conjugate variables from which ζ is deduced and by the choice of the form of the invariant, in particular, the nature of its p dependence. It may well be that further results will arise from choices which are not much more complicated than the choices made here.

We have not addressed ourselves to one task mentioned in Sec. 2 [remark (ii)], which is determining whether or not a *given* Hamiltonian belongs to one of the classes of Hamiltonians for which a formula for a first integral is available. This is a task of some complexity in itself as can be seen in a recent article by Sarlet and Bahar.[12] However, there is one area of possible application in which the simultaneous construction of Hamiltonian and first integral is actually desirable. This is the study of self-consistent problems such as the Vlasov–Poisson and Vlasov–Maxwell equations. In order to derive a set of nonlinear Maxwell equations, momentum integrals involving distribution functions (i.e., our first integrals) have to be performed as functionals of the potentials. If the momentum dependence of the first integrals is specified, it may be possible to manipulate the integrals before finding the potentials which are not specified *a priori* in self-consistent problems. It is due to the great interest in self-consistent particle dynamics problems (in plasma research, in particular) that we have stressed Hamiltonians of the form $\frac{1}{2}p^2 + V(q, t)$. We have already seen that the procedure outlined here provides something new for the class of admissible potentials.

ACKNOWLEDGMENTS

We wish to acknowledge the singular contribution made by the referee in bringing this paper to its present form. One of us (P. G. L. L.) wishes to express his appreciation of the generous hospitality extended him during the course of this work by the CTR Division of the Los Alamos National Laboratory and the Instituut voor Theoretische Mechanika of the Rijksuniversiteit Gent and the support provided by the U.S. Department of Energy through the Center for Nonlinear Studies of the Los Alamos National Laboratory, the National Fonds voor Wetenshappelijk Onderzoek (Belgium), and the Department of Mathematics of La Trobe University. We also thank Professor Mertens for his interest in this work.

[1] P. G. L. Leach, J. Math. Phys. **18**, 1902 (1977).
[2] W. Sarlet, J. Phys. A **11**, 843 (1978).
[3] P. G. L. Leach, J. Math. Phys. **20**, 86 (1979).
[4] P. G. L. Leach, J. Math. Phys. **21**, 21 (1980).
[5] H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 165 (1982).
[6] P. G. L. Leach, J. Math. Phys. **22**, 465 (1981).
[7] P. G. L. Leach, J. Math. Phys. **22**, 679 (1981).
[8] J. R. Ray and J. L. Reid, Nuovo Cimento A **59**, 134 (1980).
[9] P. G. L. Leach, Phy. Lett. A **84**, 161 (1981).
[10] Dj. Djukic, Arch. Ratl. Mech. Anal. **56**, 79 (1974).
[11] Dj. Djukic, Int. J. Non-Linear Mech. **16**, 489 (1981).
[12] W. Sarlet and L. Y. Bahar, Int. J. Non-Linear Mech. **15**, 133 (1980).
[13] H. R. Lewis and P. G. L. Leach, "A direct approach to finding exact invariants for one-dimensional time-dependent classical Hamiltonains," Preprint LA-UR-82-818, Los Alamos National Laboratory, 1982.
[14] W. Sarlet, "First integrals for one-dimensional particle motion in a nonlinear time-dependent potential field," Int. J. Non-Linear Mech. (to be published).
[15] W. Sarlet and J. R. Ray, J. Math. Phys. **22**, 2504 (1981).
[16] W. Sarlet, unpublished report.
[17] H. R. Lewis and P. G. L. Leach, "A resonance formulation for invariants of particle motion in a one-dimensional time-dependent potential," Preprint LA-UR-82-1417, Los Alamos National Laboratory, 1982.

Generalized variational principles and nondifferentiable potentials in analytical mechanics

Helge-Otmar May
School of Technology, Aristotelian University, Thessaloniki, Greece

(Received 24 March 1983; accepted for publication 13 May 1983)

The present paper deals with variational principles in terms of hemivariational inequalities and with multivalued differential equations which are called differential inclusions in analytical mechanics. Such inequalities and inclusions are received when no restrictions of differentiability are considered.

PACS numbers: 03.20. + i, 02.30. + g

I. INTRODUCTION

In analytical mechanics, the classical properties of differentiability are used in a multifarious manner.[1,2] Accordingly, the restrictions on the used functions are very severe. In this paper, however, variational principles should be summarized when no restrictions of differentiability are considered. For this, obviously, some new mathematical notions which are used in the theory of optimization and the calculus of variation are suitable.[3,4]

Let $f:\mathbb{R}^m \to [-\infty, +\infty]$ be a lower semicontinuous function and let epi f be the epigraph[4] of f. We define the generalized gradient (in the sense of Clarke[3]) of f at x by

$$\partial f(x) = \{z \in \mathbb{R}^m | (z, -1) \in N_{\text{epi} f}(x, f(x))\}, \tag{1}$$

where $N_{\text{epi} f}(x, f(x))$ is the normal cone to epi f at the point x. If f is continuously differentiable at x, then $\partial f(x) = \{\text{grad} f(x)\}$.

If f is convex, then $\partial f(x)$ denotes the classical subgradient in the sense of convex analysis. There is an equivalent definition of the generalized gradient through an inequality, as has been proposed by Rockafellar.[4] To this end, we define the upper derivative[4] by

$$f^{\uparrow}(x, y) = \limsup_{\substack{x' \to_f x \\ \lambda \downarrow 0}} \inf_{y' \to y} \frac{f(x' + \lambda y') - f(x')}{\lambda}. \tag{2}$$

Now for the generalized gradient, we have

$$\partial f(x) = \{z \in \mathbb{R}^m | f^{\uparrow}(x, y) \geqslant y_\lambda z^\lambda \quad \forall y \in \mathbb{R}^m\} \tag{3}$$

(summation convention with respect to the Greek indices). We shall call the inequality in (3) "hemivariational inequality".[5] If f is Lipschitzian around x, expression (2) reduces to

$$f^{\uparrow}(x, y) = \limsup_{\substack{x' \to x \\ \tau \downarrow 0}} \frac{f(x' + \tau y) - f(x')}{\tau}. \tag{4}$$

If, moreover, the function f is subdifferentially regular at x (which it is in the cases where f is convex or a "max function"[4]), then

$$f^{\uparrow}(x, y) = \lim_{\tau \downarrow 0} \frac{f(x + \tau y) - f(x)}{\tau}. \tag{5}$$

II. THE CENTRAL INEQUALITY IN PARAMETRIC FORM

Here we will use a parametric form of mechanics,[6-8] which has the following advantage. Usually, in analytical mechanics, the Hamiltonian function is introduced by a Legendre transformation applied to the Lagrangian function. For this, the differentiability of the Lagrangian must be presumed. But there is another possibility: to define a Hamiltonian function without any assumption concerning differentiability, where the property is used that the time-space velocities cannot be considered as independent variables.[9]

The time-space configuration of a mechanical system should be characterized by $n + 1$ variables $x = (x^0, x^1, \ldots, x^n)$, where x^0 is added to the other spatial variables in order to take the time into consideration. A curve $x:\tau \to x(\tau)$ is an absolutely continuous function from $[0,T]$ to $\mathbb{R}^{n+1}$ with derivative $\dot{x} = (\dot{x}^0, \dot{x}^1, \ldots, \dot{x}^n)$ almost everywhere (a.e.) in $[0,T]$.

Now we can introduce a time t along a curve by

$$\frac{dt}{d\tau} = \beta = f(x,\dot{x}) > 0, \tag{6}$$

where we suppose the function $\dot{x} \mapsto f(x,\dot{x})$ to be sublinear [the easiest case for f is $f(x,\dot{x}) = \dot{x}^0/c$, $\dot{x} > 0$, $c = \text{const} > 0$].

Let us define the velocities by

$$v^i = \frac{\dot{x}^i}{\beta} = \frac{dx^i}{dt}, \quad i = 0, 1, \ldots, n. \tag{7}$$

Because of the sublinearity for f, we have

$$1 = f(x, v), \tag{8}$$

so that the velocities cannot be considered as independent variables. We shall assume that the relation is solvable for v^0, i.e.,

$$v^0 = v^0(x, \tilde{v}), \tag{9}$$

where $\tilde{v} = (v^1, v^2, \ldots, v^n)$.

In analytical mechanics, we can start with a (generalized) principle of d'Alembert (compare Goldstein,[1] Lanczos,[7] and Heinz[9,10]),

$$\left(\frac{dp_\lambda}{dt} - K_\lambda\right)\delta x^\lambda = 0, \tag{10}$$

which must be valid for all $\delta x = (\delta x^0, \delta x^1, \ldots, \delta x^n)$ if constraints are avoided. We suppose the momenta $p = (p_0, p_1, \ldots, p_n)$ to be a function of the configuration and the velocities, i.e., $p_i = p_i(x, v)$. Because of (7) they could be considered as positively homogeneous functions of degree zero with respect to $\dot{x}$: $p_i = p_i(x, \dot{x})$. If the forces $K = (K_0, K_1, \ldots, K_n)$ could be derived from a generalized potential V, we have

$$-K \epsilon \partial V(x) \Leftrightarrow V^{\uparrow}(x, \delta x) \geqslant K_\lambda \delta x^\lambda \quad \forall \; \delta x, \tag{11}$$

 0022-2488/84/030491-03$02.50

and d'Alembert's principle yields a hemivariational inequality

$$V'(x,\delta x) \geqslant -\frac{dp_\lambda}{dt}\delta x^\lambda \quad \forall\ \delta x. \tag{12}$$

The differential equations of motion must be replaced now by a differential inclusion

$$-\frac{dp}{dt} \in \partial V(x). \tag{13}$$

From now on, we shall consider such systems for which

$$\delta p_\lambda v^\lambda \geqslant K_\lambda \delta x^\lambda \tag{14}$$

is fulfilled, i.e., for all $(\delta x,\delta \dot x)$, we have

$$p_\lambda^{\prime}(x,\dot x,\delta x,\delta \dot x)\dot x^\lambda \geqslant \beta(x,\dot x)K_\lambda(x,\dot x)\delta x^\lambda \quad \forall\ (\delta x,\delta \dot x) \tag{15}$$

and thus

$$p_\lambda^{\prime}(x,\dot x,0,\delta \dot x)\dot x^\lambda \geqslant 0 \quad \forall\ \delta \dot x. \tag{16}$$

If the p_i's are differentiable, we get from this,

$$\frac{\partial p_\lambda}{\partial \dot x^i}\dot x^\lambda = 0, \quad i = 0,1,\dots,n. \tag{17}$$

Instead of (14) or (15) we can also write, because of d'Alembert's principle,

$$\dot x^\lambda \delta p_\lambda \geqslant \dot p_\lambda \delta x^\lambda, \tag{18}$$

which is a generalization in space and time of Heun's central equation[2] for those systems which do not possess the usual properties of differentiability.

III. THE LAGRANGIAN FUNCTION AND THE EULER–LAGRANGE INCLUSION

Now we introduce a Lagrangian function $L(x,\dot x)$ which should satisfy

$$L'(x,\dot x,\delta x,\delta \dot x) \geqslant \dot x^\lambda p_\lambda^{\prime}(x,\dot x,\delta x,\delta \dot x) + p_\lambda(x,\dot x)\delta \dot x^\lambda \quad \forall\ (\delta x,\delta \dot x). \tag{19}$$

Using the central inequality (18), we have

$$L'(x,\dot x,\delta x,\delta \dot x) \geqslant \dot p_\lambda \delta x^\lambda + p_\lambda \delta \dot x^\lambda \quad \forall\ (\delta x,\delta \dot x) \tag{20}$$

or, equivalently,

$$(\dot p,p) \in \partial L(x,\dot x). \tag{21}$$

The Euler–Lagrange inclusion (21) is a counterpart of the usual Euler–Lagrange equation which is valid if L is differentiable. If the momenta and the Lagrangian, respectively, are Lipschitzian and subdifferentially regular, then (19) is fulfilled by $L = \dot x^\lambda p_\lambda$ because of

$$\begin{aligned} &L'(x,\dot x,\delta x,\delta \dot x) \\ &= \lim_{\tau \downarrow 0} \frac{(\dot x^\lambda + \tau\delta \dot x^\lambda)p_\lambda(x+\tau\delta x,\dot x+\tau\delta \dot x) - \dot x^\lambda p_\lambda(x,\dot x)}{\tau} \\ &= \lim_{\tau \downarrow 0} \{\dot x^\lambda\{p_\lambda(x+\tau\delta x,\dot x+\tau\delta \dot x) - p_\lambda(x,\dot x)\}/\tau \\ &\quad + \delta \dot x^\lambda p_\lambda(x+\tau\delta x,\dot x+\tau\delta \dot x)\} \\ &= \dot x^\lambda p_\lambda^{\prime}(x,\dot x,\delta x,\delta \dot x) + \delta \dot x^\lambda p_\lambda(x,\dot x). \end{aligned} \tag{22}$$

Here, we should remark that the Euler–Lagrange inclusion is connected with Hamilton's principle like in classical mechanics with the differentiable Lagrangian. Consider the following problem:

Minimize $\left\{\int_0^T L(x(\tau),\dot x(\tau))d\tau\right\}$ over all curves which satisfy $x(0) = x_0$, $x(T) = x_T$.

If $x(\cdot)$ is a (local) solution of the problem, then there is (under very mild assumptions, see Clarke[11,12]) an absolutely continuous function p: $[0,T] \to \mathbb{R}^{n+1}$ with

$$(\dot p(\tau),p(\tau)) \in \partial L(x(\tau),\dot x(\tau)) \quad \text{a.e. in } [0,T], \tag{23}$$

$$L(x(\tau),\dot x(\tau)) - p_\lambda(\tau)\dot x^\lambda(\tau) = \text{const a.e. in } [0,T], \tag{24}$$

$$L(x,(\tau),\dot x(\tau)+y) - L(x(\tau),\dot x(\tau)) \geqslant p_\lambda(\tau)y^\lambda \quad \forall\ y \text{ a.e. in } [0,T]. \tag{25}$$

IV. THE HAMILTONIAN FUNCTION AND THE CANONICAL INCLUSIONS

p_0 is a function of the configuration and the velocities. Considering (9), we get

$$p_0 = \pi(x,\tilde v). \tag{26}$$

We shall assume that the space momenta $\tilde p = (p_1,p_2,\dots,p_n)$ are functions of the configuration x and the space velocities only, i.e.,

$$\tilde p = \mathscr{P}(x,\tilde v). \tag{27}$$

If the mapping $\tilde v \mapsto \mathscr{P}(x,\tilde v)$ is Lipschitzian and if $(x,\tilde v)$ is a point so that each matrix $\mathscr{F}$, which is element of

$$\left\{(\mathscr{F}_{ik}) \,\middle|\, \exists\left\{\overset{l}{\tilde v}\right\} \to \tilde v \text{ such that } \left\{\frac{\partial \mathscr{P}_i}{\partial \tilde v^k}(x,\overset{l}{\tilde v})\right\} \to \mathscr{F}_{ik}\right\},$$

is nonsingular, then there exists a neighborhood of $\tilde v$ and a neighborhood of $\mathscr{P}(x,\tilde v)$ such that (27) is solvable for $\tilde p$,

$$\tilde v = Q(x,\tilde p), \tag{28}$$

where Q is Lipschitzian.[4] Therefore there must exist a function H, which we shall call Hamiltonian and for which we have

$$H(x,p) = p_0 - \pi(x,Q(x,\tilde p)) \equiv 0. \tag{29}$$

This should be valid for all admissible variations too, and we have to consider this relation as a constraint for the central inequality (18), where the δx^i's and δp_i's are not independent.

We shall assume that H, as a function of (x,p), is Lipschitzian. We may choose $(\delta x,\delta \tilde p)$ independently, and $\alpha(\epsilon)$ should be fixed by

$$p_0 + \alpha(\epsilon) - \pi(x+\epsilon\delta x,\tilde p+\epsilon\delta \tilde p) = 0, \tag{30}$$

where ϵ is given. Now we apply a generalized chain rule: Let $\alpha = G \circ F$ where G: $\mathbb{R}^m \to \mathbb{R}$ is a Lipschitzian function and F: $\mathbb{R} \to \mathbb{R}^m$ is continuously differentiable. Then,[4]

$$\partial\alpha(\epsilon) \subset \left\{z_i = \sum_{\lambda=1}^m w_\lambda \frac{\partial F_\lambda}{\partial x^i} \,\middle|\, w \in \partial G(F(\epsilon))\right\}. \tag{31}$$

From this, we have

$$\partial\alpha(0) \subset \left\{u_\lambda \delta x^\lambda + \sum_{\bar\lambda=1}^n q^{\bar\lambda}\delta p_{\bar\lambda}\right\}, \tag{32}$$

i.e., for all $(\delta x,\delta \tilde p)$, there is one $-\delta p_0 \in \partial\alpha(0)$ such that

$$u_\lambda \delta x^\lambda + q^\lambda \delta p_\lambda = 0$$

with

$$-(u,q) \in \partial H(x,p). \tag{33}$$

With a Lagrangian multiplier γ, which is arbitrary at first, and by the help of the central inequality, we get

$$(q^\lambda + \gamma \dot{x}^\lambda)\delta p_\lambda + (u_\lambda - \gamma \dot{p}_\lambda)\delta x^\lambda \geqslant 0 \quad \forall\ \delta x, \delta \bar{p}, \gamma. \tag{34}$$

Now we can choose γ to be such that $\gamma \dot{x}^0 = -q^0$ and, hence, it follows from (34) that

$$\gamma \dot{x}^{\bar{\imath}} = -q^{\bar{\imath}}, \quad \gamma \dot{p}_i = u_i, \quad i = 0,1,\ldots,n, \quad \bar{\imath} = 1,2,\ldots,n. \tag{35}$$

Finally, we get the canonical inclusion by the help of (33):

$$\gamma(-\dot{p},\dot{x}) \in \partial H(x,p). \tag{36}$$

Let us consider now an admissible transformation

$$\bar{x} = \bar{x}(x,p), \quad \bar{p} = \bar{p}(x,p), \tag{37}$$

i.e., that the inverse transformation

$$x = x(\bar{x},\bar{p}), \quad p = p(\bar{x},\bar{p}) \tag{38}$$

exists. If the first part of (38) is solvable for $\bar{x}$ too, i.e.,

$$\bar{x} = \bar{x}(x,\bar{p}), \tag{39}$$

and if there exists a function $W(x,\bar{p})$ for which we have

$$W^\uparrow(x,\bar{p},\delta x,\delta\bar{p}) \geqslant p_\lambda(x,\bar{p})\delta x^\lambda + \bar{x}^\lambda(x,\bar{p})\delta\bar{p}_\lambda \quad \forall\ (\delta x,\delta\bar{p}), \tag{40}$$

then we shall call this a "special generalized transformation." Here, it is

$$(p,\bar{x}) \in \partial W(x,\bar{p}). \tag{41}$$

W becomes a generating function in the classical sense if it is continuously differentiable. Now we shall find a special generalized transformation for which (41) is valid and for which W is independent of $\bar{p}$. We set

$$\overline{W}(x) = W(x,\bar{p} = a), \tag{42}$$

where a is a suitable constant. Then it is

$$q \in \partial \overline{W}(x) \Rightarrow q \in \{ p \mid \exists\ \bar{x} \text{ such that } (p,\bar{x}) \in \partial W(x,a)\},$$

and if (x,q) is given by a motion, the identity (29) for the Hamiltonian must be fulfilled, hence

$$H(x,\partial \overline{W}(x)) = 0, \tag{43}$$

which is a generalization of the Hamilton–Jacobi equation.

[1]H. Goldstein, *Classical Mechanics*, 2nd ed. (Addison-Wesley, Reading, MA, 1980).
[2]G. Hamel, *Theoretische Mechanik* (Springer-Verlag, New York, 1978).
[3]F. H. Clarke, Trans. Am. Math. Soc. **205**, 247–262 (1975).
[4]R. T. Rockafellar, *The Theory of Subgradients and its Applications to Problems of Optimization. Convex and Nonconvex Functions* (Heldermann-Verlag, Berlin, 1981).
[5]P. D. Panagiotopoulos, Mech. Res. Comm. **8**, 335–340 (1981).
[6]P. Mittelstaedt, *Klassische Mechanik* (Bibliographisches Institut, Mannheim, Wien, Zürich, 1969).
[7]C. Lanczos, *The Variational Principles of Mechanics* (University of Toronto, Toronto, 1966).
[8]C. Heinz, Acta Mech. **41**, 23–33 (1981).
[9]C. Heinz, "Vorlesungen über Analytische Mechanik," lecture given at RWTH Aachen (to be published).
[10]C. Heinz, Acta Mech. **10**, 110–129 (1970).
[11]F. H. Clarke, SIAM J. Control Optim. **14**, 682–699 (1976).
[12]F. H. Clarke, Can. J. Math. **32**, 494–509 (1980).

The direct correlation function for the hard-core potentials

Konstanty Mazuruk
Institute of Physics, Polish Academy of Sciences, 02-668 Warsaw, Poland[a)]

(Received 6 November 1981; accepted for publication 20 January 1983)

The Ornstein–Zernicke relation is examined for the class of intermolecular potentials with a hard-core term. The nonlinear integral equation relating the direct correlation function outside the hard-core region with the inside one is proposed. A class of analytical solutions is obtained.

PACS numbers: 03.20. + i, 05.20. − y, 02.30.Rz

I. INTRODUCTION

Modern methods of the statistical mechanical description of a classical system in equilibrium are based on the concept of two-particle correlation functions. Several approximate theories provide a functional relationship between the pair distribution function and the direct correlation function. The use of the exact Ornstein–Zernicke equation to these relationships yield nonlinear integral equations. The high machine-time requirements for numerical calculations limit the applications of these methods at present.

An important role in the development of the theory is played by analytical solutions for some special cases.[1–5] One well known and widely used example is the solution of the Percus–Yevick equation for the single-component system of hard spheres given by Wertheim[1] and Thiele.[2] In this case, the direct correlation function $C(r)$ vanishes for $r > R$, where R is the diameter of a sphere and the radial distribution function vanishes for $r < R$. The last condition simply reflects the impenetrability of a hard sphere and is called the core condition. In case of $C(r)$ for $r > R$ of a Yukawa potential form, the analytical solution for $C(r)$, $r < R$ has been found by Waisman.[3] Hoye and Blum[4] extended the solution to an arbitrary number N of Yukawas using the Baxters factorization technique. They transformed the initial Ornstein–Zernicke equation with the core-potential condition into a set of $N + 2$ nonlinear algebraic equations.

In this work, we give a detailed analysis of the properties of the direct correlation function of the system with the core condition imposed. In other words, we consider a system of molecules with a hard-core term in intermolecular pair potential. We develop a slightly different approach to that given by Wertheim.[1] As a consequence of the core condition, it is possible to eliminate the pair distribution function from the Ornstein–Zernicke equation and formulate an equivalent, quite new integral equation (56) concerning only $C(r)$ and relating $C(r)$ for $r < R$ to $C(r)$ for $r > R$. The relatively simple linear approximation (48) of this equation should be useful for practical purposes. In the case of N Yukawas, the solution of proposed equation (56) gives N nonlinear algebraic equations, instead of the $N + 2$ derived by Hoye and Blum.[4] We derive a new relationship between the inverse compressibility of the system and the integral characteristics of $C(r)$.

[a)] Present address: Department of Physics, University of Toronto, Toronto, Ontario, M5S 1A7, Canada.

II. GENERAL FORMALISM

The impenetrability property of a hard molecular core of diameter R can be expressed in terms of a total correlation function $h(r)$ as the so-called core condition

$$h(r) = -1, \quad r < R. \tag{1}$$

In the following, without loss of generality, the diameter R will be put equal to unity, $R = 1$. We introduce, for convenience, the normalized density of the system $\xi = \pi/6\rho$, where ρ is the number of molecules per unit volume. The Ornstein–Zernicke equation relating the direct correlation function $C(r)$ and the total correlation function $h(r)$ of a classical three-dimensional system of molecules can be expressed in bipolar coordinates in the form

$$C(r) = h(r) - \frac{12\xi}{r}\int_0^\infty th(t)dt\int_{|r-t|}^{r+t} xC(x)dx. \tag{2}$$

The core condition (1) leads to the discontinuity of $C(r)$ and $h(r)$ in $r = 1$. Let us define

$$rC(r) = \begin{cases} C_1(r), & r < 1, \\ C_2(r), & r > 1. \end{cases} \tag{3}$$

In this work, we consider the case of C_1 and C_2 being analytical functions in the region $r \in (0, \infty)$. The problem to be solved is to find $C_1(r)$ if $C_2(r)$ is known. In order to do this, we apply the Laplace-transform technique.

We introduce the following transforms:

$$\begin{aligned} C(s) &= \int_0^\infty rC(r)e^{-sr}\,dr \equiv \mathscr{L}\{rC(r)\}, \\ H(s) &= \mathscr{L}\{rh(r)\}, \\ H_1(s) &= \mathscr{L}\{rh(r)\theta(1-r)\}, \\ H_2(s) &= \mathscr{L}\{rh(r)\theta(r-1)\}, \\ A(s) &= \mathscr{L}\{C_1(r)\}, \\ B_1(s) &= -e^{-s}\mathscr{L}\{C_1(r)\theta(r-1)\}, \\ B_2(s) &= e^{s}\mathscr{L}\{C_2(r)\theta(r-1)\}, \\ B(s) &= B_1(s) + B_2(s), \end{aligned} \tag{4}$$

where

$$\theta(r) = \begin{cases} 1, & r \geqslant 0, \\ 0, & r < 0. \end{cases}$$

We note that

$$C(s) = A(s) + B(s)e^{-s}. \tag{5}$$

For s large enough, the expansion of $A(s)$ and $B(s)$ with re-

 0022-2488/84/030494-05$02.50

spect to inverse powers of s reads

$$A(s)=\frac{C_1(0)}{s}+\frac{C_1'(0)}{s^2}+\frac{C_1''(0)}{s^3}+\cdots, \quad (6)$$

$$B(s)=\frac{\Delta C}{s}+\frac{\Delta C'}{s^2}+\frac{\Delta C''}{s^3}+\cdots, \quad (7)$$

where

$$\Delta C^{(n)}=C_2^{(n)}(1)-C_1^{(n)}(1). \quad (8)$$

The Laplace transform of derivative of (2) with respect to r can be written in the form

$$\begin{aligned} sC(s)&=sH(s)+12\xi\Big[H(s)(C(s)-C(-s))\\ &\quad+B(s)\frac{e^{-s}}{s^2}-B(-s)\frac{e^{s}}{s^2}+\frac{B_2(s)+B_2(-s)}{s}\\ &\quad-\frac{B_2(s)-B_2(-s)}{s^2}+\frac{2}{s}\int_0^1 tC_1(t)\,dt\Big]\\ &\quad+\frac{12\xi}{s^2}(A(s)-A(-s))-12\xi\int_1^\infty th(t)\,dt\\ &\quad\times\int_0^\infty e^{-sr}[C_2(t+r)+C_2(t-r)]\,dr. \end{aligned} \quad (9)$$

We substract $sC(-s)$ from (9) and get

$$\begin{aligned} &H(s)-H(-s)\\ &\quad=(C(s)-C(-s))[1-(12\xi/s)(H(s)-H(-s))]. \end{aligned} \quad (10)$$

The choice of $C(r)$ in the form (3) leads to discontinuities of higher-order derivatives of $h(r)$ in points $r=2,3,\cdots$. Therefore, we can express $H(s)$ in the following way

$$H(s)=-1/s^2+Q(s)e^{-s}+Q_2(s)e^{-2s}+Q_3(s)e^{-3s}+\cdots, \quad (11)$$

where $Q_n(s)$ describes the discontinuity of $rh(r)$ and its derivatives of an arbitrary order in point $r=n$. This can be seen from the expansion

$$Q_n(s)=\frac{\Delta H(n)}{s}+\frac{\Delta H'(n)}{s^2}+\frac{\Delta H''(n)}{s^3}+\cdots, \quad (12)$$

where

$$\Delta H^{(l)}(n)=[rh(r)]^{(l)}_{r=n+0}-[rh(r)]^{(l)}_{r=n-0}. \quad (13)$$

Collecting terms with the same powers of e^{-s} in (10), we find

$$\begin{aligned} &A(s)-A(-s)\\ &\quad=-(12\xi/s)[B(s)Q(-s)+B(-s)Q(s)], \end{aligned} \quad (14)$$

$$B(s)=Q(s)+(12\xi/s)[D(s)Q(s)-B(-s)Q_2(s)], \quad (15)$$

and, for every $n=2,3,\cdots$,

$$\begin{aligned} Q_n(s)&=-(12\xi/s)\\ &\quad\times[D(s)Q_n(s)+B(s)Q_{n-1}(s)-B(s)Q_{n+1}(s)], \end{aligned} \quad (16)$$

where

$$D(s)=\tfrac{1}{2}(A(s)-A(-s)). \quad (17)$$

This set of equations has a convergent solution of the form

$$Q_n(s)=-\frac{s}{12\xi}\left(-\frac{12\xi Q(s)}{s}\right)^n \quad (18)$$

if

$$\left|\frac{12\xi}{s}Q(s)e^{-s}\right|<1. \quad (19)$$

Hence

$$B(s)=\frac{Q(s)}{1+(12\xi/s)^2Q(s)Q(-s)} \quad (20)$$

and

$$H(s)=-\frac{1}{s^2}+\frac{Q(s)e^{-s}}{1+(12\xi/s)Q(s)e^{-s}}. \quad (21)$$

The relation (20) can be practically applied to express the Taylor expansion of $h(r)$ in the region $r\in(1,2)$ in terms of $\Delta C^{(n)}$ [see Eq. (8)]. Expansion of (20) with respect to inverse powers of s gives directly

$$rh(r)=\sum_{n=0}^{\infty}H_n(r-1)^n/n!,\quad 1<r<2 \quad (22)$$

with

$$\begin{aligned} &H_0=\Delta C-1,\quad H_1=\Delta C'-1,\quad H_2=\Delta C'',\\ &H_3=\Delta C''',\quad H_4=\Delta C^{(4)}-\alpha_1\Delta C,\quad H_5=\Delta C^{(5)}-\alpha_1\Delta C',\\ &H_6=\Delta C^{(6)}-\alpha_1\Delta C''-\alpha_2\Delta C,\\ &H_7=\Delta C^{(7)}-\alpha_1\Delta C'''-\alpha_2\Delta C',\\ &H_8=\Delta C^{(8)}-\alpha_1\Delta C^{(4)}-\alpha_2\Delta C''-\alpha_3\Delta c,\\ &H_9=\Delta C^{(9)}-\alpha_1\Delta C^{(5)}-\alpha_2\Delta C'''-\alpha_3\Delta C',\\ &H_{10}=\Delta C^{(10)}-\alpha_1\Delta C^{(6)}-\alpha_2\Delta C^{(4)}-\alpha_3\Delta C''-\alpha_4\Delta C,\\ &H_{11}=\Delta C^{(11)}-\alpha_1\Delta C^{(7)}-\alpha_2\Delta C^{(5)}\\ &\qquad-\alpha_3\Delta C'''-\alpha_4\Delta C',\ \text{etc.}, \end{aligned} \quad (23)$$

where we introduced

$$\begin{aligned} &\alpha_1=(12\xi\Delta C)^2,\\ &\alpha_2=\alpha_1(2\Delta C\Delta C''-(\Delta C')^2),\\ &\alpha_3=\alpha_1(2\Delta C\Delta C^{(4)}-2\Delta C'\Delta C^{(3)}+(\Delta C'')^2-2\alpha_1(\Delta C)^4),\\ &\alpha_4=\alpha_1(2\Delta C\Delta C^{(6)}-2\Delta C'\Delta C^{(5)}+2\Delta C''\Delta C^{(4)}\\ &\qquad-(\Delta C''')^2-3\alpha_2(\Delta C)^2). \end{aligned} \quad (24)$$

Elimination of Q from (14) and (20) gives the following basic equation relating properties of $C(r)$ in $r=0^+$ with those in the vicinity of $r=1$ [see Eqs. (17), (6), and (7)]:

$$B(s)B(-s)=-(s/12\xi)D(s)-D^2(s). \quad (25)$$

This relation is equivalent to Eq. (51) given by Wertheim.[1] Adding $sC(-s)$ to (9) gives another basic relation:

$$\begin{aligned} &A(s)+A(-s)\\ &\quad=-\frac{2}{s^2}+\frac{24\xi}{s}\Big[\frac{B_2(s)+B_2(-s)}{s}\\ &\qquad-\frac{B_2(s)-B_2(-s)}{s^2}+\frac{2}{s}\int_0^1 tC_1(t)\,dt\Big]\\ &\qquad-\frac{24\xi}{s}\int_1^\infty th(t)\,dt\\ &\qquad\times\int_0^\infty e^{-sr}[C_2(t+r)+C_2(t-r)]\,dr. \end{aligned} \quad (26)$$

If we restrict our attention to the solutions for $h(r)$ with Fourier transforms, we must fulfill the condition $rh(r)=0$ for $r\to\infty$. The equivalent condition imposed on $H(s)$, namely,

$\lim_{s\to 0} sH(s) = 0$, leads directly to the following Taylor expansion of $Q(s)$ in the vicinity of $s = 0$:

$$Q(s) = q_1 s + q_2 s^2 + q_3 s^3 + q_4 s^4 + q_5 s^5 + \cdots, \tag{27}$$

where $q_1,\ldots,q_5$ are equal to

$$q_1 = -\xi/12, \quad q_2 = q_1, \quad q_3 = q_1/2,$$
$$q_4 = q_1/6 - q_1^2, \quad q_5 = q_1/24 - q_1^2. \tag{28}$$

Employing (20), (14), and (26), we prove that the Lorant series of $A(s)$ and $B(s)$ in the neighborhood of $s = 0$ have the form

$$A(s) = \frac{a_3}{s^5} + \frac{a_2}{s^3} + \frac{a_1}{s^2} + \frac{a_0}{s} U(s), \tag{29}$$

$$-B(s) = \frac{a_3}{s^5} + \frac{a_3}{s^4} + \frac{b_2}{s^3} + \frac{b_1}{s^2} + \frac{b_0}{s} + W(s), \tag{30}$$

where

$$b_0 = a_0 + a_1 + a_2/2 + a_3/24,$$
$$b_1 = a_1 + a_2 + a_3/6, \tag{31}$$
$$b_2 = a_2 + a_3/2,$$

and

$$a_3 = 12\xi a_1. \tag{32}$$

The functions $U(s)$ and $W(s)$ introduced above exist in $s = 0$ and can be expanded in the Taylor series

$$U(s) = U_0 + U_1 s + U_2 s^2 + U_3 s^3 + \cdots, \tag{33}$$

$$W(s) = W_0 + W_1 s + W_2 s^2 + W_3 s^3 + \cdots. \tag{34}$$

After substituting (29) and (30) with expansions (33) and (34) into (25) and collecting terms with square and 4th reciprocal powers of s, we find two conditions:

$$b_1^2 + a_1 + 2a_0 a_2 - 2b_0 b_2 + 2a_3(W_0 - W_1 + U_1) = 0,$$
$$a_0^2 - b_0^2 + 2b_1 W_0 + 2a_3(W_2 - W_3) - 2b_2 W_1 + a_2/12\xi + 2a_2(U_1 + U_3) = 0. \tag{35}$$

Hence, some simple but tedious algebra leads to the important relations

$$a_1 = -\frac{(1+2\xi)^2}{(1-\xi)^4}(1+\Psi) + \frac{12\xi a_0}{(1-\xi)^2} + \frac{72\xi^2}{(1-\xi)^4}\Phi, \tag{36}$$

$$a_2 = \frac{3\xi(\xi+2)^2}{(1-\xi)^4}(1+\Psi) + \frac{24\xi(1-2\xi^2-8\xi)}{(1-\xi)^4}\Phi - \frac{12\xi a_0(\xi+2)}{(1-\xi)^2}, \tag{37}$$

where

$$\Psi = 24\xi(W_0 + U_1 - W_1),$$
$$\Phi = W_0 + 12\xi(W_0/6 + W_2 - W_1/2 - W_3 + U_3). \tag{38}$$

For completeness, we can also note that

$$b_0 = a_0 + \frac{12\xi}{(1-\xi)^2}\Phi - \frac{2+\xi}{2(1-\xi)^2}(1+\Psi),$$

$$b_1 = \frac{24\xi(1-5\xi+4\xi^2)}{(1-\xi)^4}\Phi - \frac{12\xi a_0}{1-\xi} - \frac{1-6\xi+5\xi^3}{(1-\xi)^4}(1+\Psi), \tag{39}$$

$$b_2 = \frac{24\xi(1-8\xi+16\xi^2)}{(1-\xi)^4}\Phi + \frac{12\xi a_0(5\xi-2)}{(1-\xi)^2} + \frac{3\xi(2-4\xi-7\xi^2)}{(1-\xi)^4}(1+\Psi).$$

It can be seen from (26) that (terms $\sim s^{-2}$)

$$a_1 = -1 + 24\xi\int_0^\infty r^2 C(r)\,dr. \tag{40}$$

So, $-a_1$ is the inverse compressibility and thus (36) is the key expression of the theory. We remark that if W_i, $U_i \to 0$, Eq. (36) reproduces the Wertheim results for the hard-sphere system. A justification of this fact can be easily derived from Eqs. (26), (29), and (30). Hence $a_0, \Psi, \Phi \to 0$ if $C_2 \to 0$.

III. ANALYTICAL SOLUTIONS

In this section, we examine some particular cases of C_2. Let us consider $C_2(r)$ of the form

$$C_2(r) = \sum_{i=1}^{N} F_i e^{-x_i r}, \tag{41}$$

where F_i, x_i can, in general, be complex variables with C_2 real. We expect $C_1(r)$ to be of the form

$$C_1(r) = a_0 + a_1 r + a_2 r^2/2 + a_3 r^4/24 + \sum_{i=1}^{N}(A_i e^{-x_i r} + B_i e^{x_i r}). \tag{42}$$

The condition $C_1(0) = 0$ implies

$$a_0 = -\sum_{i=1}^{N}(A_i + B_i). \tag{43}$$

By performing a Laplace transform on (42), one gets

$$A(s) = \frac{a_3}{s^5} + \frac{a_2}{s^3} + \frac{a_1}{s^2} + \frac{a_0}{s} + \sum_{i=1}^{N}\left(\frac{A_i}{s+x_i} + \frac{B_i}{s-x_i}\right),$$

$$-B(s) = \frac{a_3}{s^5} + \frac{a_3}{s^4} + \frac{b_2}{s^3} + \frac{b_1}{s^2} + \frac{b_0}{s} + \sum_{i=1}^{N}\left(\frac{A_i - F_i}{s+x_i}e^{-x_i} + \frac{B_i}{s-x_i}e^{x_i}\right). \tag{44}$$

After substitution of these explicit forms (44) into Eq. (25) and a comparison of the terms with the same orders of singularity in $s = x_i$ one finds two conditions on A_i, B_i:

$$(A_i - B_i)^2 = -4B_i F_i, \tag{45}$$

$$\begin{aligned} m_{1i}(A_i - F_i)e^{-x_i} + M_{2i}B_i e^{x_i} &= \alpha_i(A_i + B_i) + (x_i^4/4)(A_i + B_i)^2 - (x_i^4/2) \\ &\times\left[(A_i - F_i)^2 e^{-2x_i} + B_i^2 e^{2x_i}\right] - x_i^5 \\ &\times\sum_{n\neq i}^{N}\left\{(A_i - F_i)e^{-x_i}\frac{A_n - F_n}{x_n + x_i}e^{-x_n} + B_i e^{x_i}\right. \\ &\times\frac{A_n - F_n}{x_i - x_n}e^{-x_n} + (A_i - F_i)e^{-x_i}\frac{B_n e^{x_n}}{x_i - x_n} \\ &\left.+ B_i e^{x_i}\frac{B_n}{x_i + x_n}e^{x_n} - x_i(A_i + B_i)\frac{A_n + B_n}{x_i^2 - x_n^2}\right\}, \end{aligned} \tag{46}$$

where

$$m_{1i} = b_0x_i^4 + b_1x_i^3 + b_2x_i^2 + a_3x_i + a_3,$$
$$m_{2i} = b_0x_i^4 - b_1x_i^3 + b_2x_i^2 - a_3x_i + a_3, \tag{47}$$
$$\alpha_i = x_i^6/24\xi + a_0x_i^4 + a_2x_i^2 + a_3.$$

For the simplest case, $N = 1$, it is seen that elimination of A_i from (45) and (46) yields a polynomial of 4th order with respect to B_i. This equation has for some range of changes of F, x four real solutions (if $F<0$); one of them fulfills condition (19). If $F>0$, we have only two real solutions. In this case, for a given x there exists $F_{\max}$ such that for $F>F_{\max}$, there are no real solutions. At this point $a_1 = 0$. These results were obtained from a numerical analysis.

The set of equations (45) and (46) can be solved by an iteration procedure. For this purpose we expand unknowns $A_i, B_i, m_{1i}, m_{2i}, \alpha_i$ in powers of F_j. The result consists of two terms, linear and nonlinear with respect to F_j. The nonlinear part gives us a perturbation term. Neglecting this term, we get a linear approximation. The solution in this case can be expressed as follows:

$$A = B + 2\sqrt{-BF},$$
$$B = -F\left(\frac{m_1e^{-x} - \alpha + x^6/24\xi}{m_1e^{-x} + m_2e^x - 2\alpha}\right)^2, \tag{48}$$

where

$$m_1 = \bar{b}_0x^4 + \bar{b}_1x^3 + \bar{b}_2x^2 + \bar{a}_3x + \bar{a}_3,$$
$$m_2 = \bar{b}_0x^4 - \bar{b}_1x^3 + \bar{b}_2x^2 - \bar{a}_2x + \bar{a}_3,$$
$$\alpha = x^6/24\xi + \bar{a}_2x^2 + \bar{a}_3,$$
$$\bar{b}_0 = -\frac{2+\xi}{2(1-\xi)^2}, \quad \bar{b}_1 = -\frac{1-6\xi+5\xi^3}{(1-\xi)^4}, \tag{49}$$
$$\bar{b}_2 = \frac{3\xi(2-4\xi-7\xi)^2}{(1-\xi)^4},$$
$$\bar{a}_1 = -\frac{(1+2\xi)^2}{(1-\xi)^4}, \quad \bar{a}_3 = 12\xi\bar{a}_1,$$
$$\bar{a}_2 = \frac{3\xi(\xi+2)^2}{(1-\xi)^4},$$

and where for convenience we have omitted the label i, because the same relations are valid for all components. We present here only this branch of solutions which fulfills condition (19). The exact numerical calculation given by Waisman[3] and the above approximate solution (48) with the same parameters differ by about 10%. So we expect that the iteration procedure applied to (45) and (46) with the first approximation of the form (48) can, in practice, be highly convergent.

Consider now a more general class of functions $C_2(r)$:

$$C_2(r) = \sum_{i=1}^{N} P_{ni}(r)e^{-x_ir}, \tag{50}$$

where $P_{ni}(r)$ are n-order polynomials,

$$P_{ni}(r) = P_{0i} + P_{1i}r + \cdots + P_{ni}r^n. \tag{51}$$

We expect $C_1(r)$ to be of the form

$$C_1(r) = a_0 + a_1r + a_2r^2/2 + a_3r^4/24 + \sum_{i=1}^{N} (A_{ni}(r)e^{-x_ir} + B_{ni}(r)e^{x_ir}), \tag{52}$$

where A_{ni}, B_{ni} are polynomials. Performing Laplace transforms on (50) and (52), we see that $A(s)$ and $B(s)$ are the only functions having only multiple poles in $s=0$, $s=\pm x_i$ as singularities. After a comparison of the coefficients in front of the same poles in (25), we can find a set of algebraic equations for unknown coefficients of polynomials A_{ni}, B_{ni}. The consistency between (25) and (26) can easily be proved, but we shall omit here the mathematical details of this problem. It can be shown, too, that for a given i, the polynomials A_i, B_i, and F_i are of the same order. It is impossible, unfortunately, to write down explicitly a set of algebraic equations relating A_{ni}, B_{ni}, and F_{ni} for arbitrary n. The existence of many solutions of this set and the choice of the one which fulfills condition (19) complicates the problem considered.

IV. INTEGRAL EQUATION RELATING C_1 WITH C_2

In the limit $x_{i-1} - x_i \to 0$, we recognize (41) as the Laplace transform representation of $C_2(r)$. Replacing the sums in (41), (42) by integrals, we can write

$$C_2(r) = \int_0^\infty F(x)e^{-xr}\,dx, \tag{53}$$
$$C_1(r) = a_0 + a_1r + a_2r^2/2 + a_3r^4/24 + \int_0^\infty (A(x)e^{-xr} + B(x)e^{xr})\,dx. \tag{54}$$

From (45), we have

$$(A(x) - B(x))^2 = -4B(x)F(x), \tag{55}$$

and from (46) we find the following basic integral equation relating unknown functions $A(x)$ and $B(x)$ to $F(x)$:

$$\begin{aligned} m_1(x)(A(x) - F(x))e^{-x} + m_2(x)B(x)e^x &= \alpha(x)(A(x)+B(x)) - x^5\int_0^\infty \Big\{(A(x)-F(x))e^{-x} \\ &\times \frac{A(x')-F(x')}{x+x'}e^{-x'} + B(x)e^x\frac{A(x')-F(x')}{x-x'}e^{-x'} \\ &+ (A(x)-F(x))e^{-x}\frac{B(x')}{x-x'}e^{x'} \\ &+ B(x)e^x\frac{B(x')}{x+x'}e^{x'} \\ &- x(A(x)+B(x))\frac{A(x')+B(x')}{x^2-x'^2}\Big\}\,dx', \end{aligned} \tag{56}$$

where

$$m_1(x) = b_0x^4 + b_1x^3 + b_2x^2 + a_3x + a_3,$$
$$m_2(x) = b_0x^4 - b_1x^3 + b_2x^2 - a_3x + a_3, \tag{57}$$
$$\alpha(x) = x^6/24\xi + a_0x^4 + a_2x^2 + a_3,$$

and

$$a_0 = -\int_0^\infty (A(x) + B(x))\,dx. \tag{58}$$

This equation can be solved by the iteration procedure described in the preceding section. A linear approximation leads to expressions (48) and seems to be accurate for a number of cases. In general, however, more detailed numerical calculations comparing this approximation with an exact solution are still needed.

V. SUMMARY

In this work, we have presented new exact relations concerning the direct correlation function $C(r)$ of a one-component system with a hard-core repulsive term in intermolecular potential. In particular, we have obtained a new expression relating inverse compressibility of the system with integral characteristics of the direct correlation function [see Eq. (36)]. We have considered $C(r)$ for $r > 1$ in the form of a sum of Yukawa potential-like functions and have found a solution for $C(r)$ in the region $0 < r < 1$. On the basis of this solution, it is possible to write a quite general integral equation which must fulfill $C(r)$. This equation appears as a consequence of the imposed core condition and, in contrast to the Ornstein–Zernicke equation, does not contain an unknown pair distribution function. The sum of a few "Yukawas" with, in general, complex x_i can be practically used as a good approximation to $C(r)$ for $r > 1$. The proposed iteration procedure of calculations, together with the explicitly given linear approximation, should substantially decrease the labor accompanying such numerical calculations.

ACKNOWLEDGMENT

The author wishes to thank Dr. W. Walukiewicz for useful remarks concerning this work.

[1]M. S. Wertheim, J. Math. Phys. **5**, 643 (1964).
[2]E. J. Thiele, J. Chem. Phys. **38**, 1959 (1963).
[3]E. Waisman, Molecular Phys. **25**, 45 (1973); D. Henderson, G. Stell, and E. Waisman, J. Chem. Phys. **62**, 4247 (1975).
[4]J. S. Hoye and L. Blum, J. Statist. Phys. **16**, 399 (1977).
[5]P. T. Cummings, C. C. Wright, J. W. Perram, and E. R. Smith, J. Statist. Phys. **21**, 659 (1979).

Lightlike contractions on Minkowski space-time

Ivor Robinson and Krzysztof Rózga[a)]
Program in Mathematical Sciences, The University of Texas at Dallas, Richardson, Texas 75080

(Received 9 November 1982; accepted for publication 4 March 1983)

The lightlike contractions of electromagnetic fields are discussed. In particular Dirac's delta type behavior on null hyperplanes is established and the lightlike limit of the electromagnetic field of a spinning, charged particle is obtained.

PACS numbers: 03.30. + p

1. INTRODUCTION

Consider a charged particle moving along a straight line in the Minkowski space-time, with the velocity $v<1$ (the velocity of light $c\equiv 1$). Then if the limit of the electromagnetic field, as $v\rightarrow 1$, is taken in the sense of generalized functions, one obtains the solution of the Maxwell equations which vanishes outside a null hyperplane and has Dirac's δ-type singularity on it.[1] This field can be interpreted as the electromagnetic field of a charged, lightlike particle.

Although there is no experimental evidence that the rest-mass-zero particles have electric charges or any type of an electromagnetic structure it is still of theoretical interest to study such fields, at least as approximations of electromagnetic fields of fast moving particles.

The proccess described above is called a contraction of the electromagnetic field. It can be stated equivalently as follows. Let F be an electromagnetic field on the Minkowski space-time M. Consider next a one-parameter group of motions $\Phi_v: M\rightarrow M, |v|<1$, corresponding to Lorentz boosts. Its action on F produces the one-parameter family of electromagnetic fields, namely $\Phi_v F$. Then the question is about the limits of that family as v approaches $+1$ or -1.

It is natural to generalize these concepts by considering an arbitrary one-parameter family of diffeomorphisms on the space-time which is not necessarily flat. The corresponding contractions may be expected to provide new solutions of the field equations or at least to throw new light on old ones.[2,3]

The purpose of this paper is to study the linear theory. In Sec. 2 the necessary concepts and means are introduced. In Sec. 3 an analysis of lightlike contractions for electromagnetic multipole structures is given. Section 4 deals with the problem of lightlike limits for the electromagnetic field of a spinning charged particle.

2. THE NOTION OF CONTRACTIONS AND RELATED TOPICS

Let (M,ds^2) denote a space-time, not necessarily that of Minkowski; that is, a pair consisting of a four-dimensional differentiable manifold and a smooth metric structure with the physical signature.

Let Φ_t, where $t\in)t_1,t_2($ be a one-parameter family of diffeomorphisms of M. (t_1 and t_2 are permitted to be equal to $\pm\infty$.) The action induced by Φ_t provides a one-parameter family of metrics on M, $\Phi_t\, ds^2$. If at the same time another tensor field (for example an electromagnetic one) F is given on M, then $\Phi_t F$ denotes the corresponding family. The metric ds^2 and the field F are subjected to some covariant, differential constraints, the field equations. It is clear, therefore, that the same relations hold between $\Phi_t\, ds^2$ and $\Phi_t F$ as t approaches t_1 or t_2. There are two questions related to the problem of limits:

(i) What type of a convergence does one have in mind?

(ii) Are the limits, if they exist, the solutions of the field equations?

The situation is especially simple if a linear theory on the Minkowski space-time is considered and $\{\Phi_t\}$ is assumed to be a one-parameter group of motions. The convergence of $\Phi_t F$ is understood then as the convergence of distributions on M.[4,5] The limit of $\Phi_t F$, if it exists, in that sense, fulfulls again the field equations. It follows from the fact that the linear algebraic operations as well as differentiations make sense for generalized functions and are continuous, e.g., they commute with the operations of taking limits.[4]

In the particular case when $)-1,1(\ni v\rightarrow\Phi_v$ is a one-parameter group of the Lorentz boosts, the corresponding limits of $\Phi_v F$ are referred to as *lightlike limits*. One can generalize slightly this construction, considering from the very beginning instead of a field F, a one-parameter family F_v and then a new family $\Phi_v F_v$. In that case the limits are called *lightlike contractions*.

Consider an electromagnetic field F, which is stationary in the Minkowski coordinate system $\{t,x,y,z\}$, and the one-parameter group Φ_v of the Lorentz boosts of the form

$$t\circ\Phi_v^{-1}=\frac{t-vz}{(1-v^2)^{1/2}},\quad x\circ\Phi_v^{-1}=x,$$
$$y\circ\Phi_v^{-1}=y,\quad z\circ\Phi_v^{-1}=\frac{z-vt}{(1-v^2)^{1/2}}. \tag{2.1}$$

Then the induced family of electromagnetic fields is

$$\begin{aligned}\Phi_v F={}&\frac{1}{(1-v^2)^{1/2}}F_{tx}\circ\Phi_v^{-1}(dt-vdz)\wedge dx\\&+\frac{1}{(1-v^2)^{1/2}}F_{ty}\circ\Phi_v^{-1}(dt-v\,dz)\wedge dy\\&+\frac{1}{(1-v^2)^{1/2}}F_{yz}\circ\Phi_v^{-1}dy\wedge(dz-v\,dt)\\&+\frac{1}{(1-v^2)^{1/2}}F_{zx}\circ\Phi_v^{-1}(dz-v\,dt)\wedge dx\\&+F_{tz}\circ\Phi_v^{-1}\,dt\wedge dz+F_{xy}\circ\Phi_v^{-1}\,dx\wedge dy.\end{aligned}\tag{2.2}$$

Because of further applications we assume for a moment that the components of F are locally integrable. Lemma 1 below

[a)] On leave of absence from Technical University of Kielce, Poland.

 0022-2488/84/030499-07$02.50

provides then the conditions on $F_{\mu\nu}$ under which the limit of (2.2) exists. Denote by f any component of $F_{\mu\nu}$.

Lemma 1: Let $f(x,y,z)$ be a locally integrable function on $\mathbb{R}^3$ such that for almost all (x,y) the function $f_{x,y}(z) := f(x,y,z)$ is an integrable function of z and the function

$$\int_{-\infty}^{+\infty} |f(x,y,z)| dz$$

is locally integrable on $\mathbb{R}^2$. Then for an arbitrary test function ψ (infinitely differentiable with a compact support)

$$\lim_{v\to\pm 1} \left\langle \frac{1}{(1-v^2)^{1/2}} f\circ\Phi_v^{-1},\psi \right\rangle = \int_{\mathbb{R}^3} dx\, dy\, d\sigma \left(\int_{-\infty}^{+\infty} f(x,y,z)dz \right) \psi(\sigma,x,y,\pm\sigma),$$

where σ is a new variable running from $-\infty$ to $+\infty$.

Proof: Indeed, taking into account (2.1), we have

$$\begin{aligned}
&\left\langle \frac{1}{(1-v^2)^{1/2}} f\circ\Phi_v^{-1},\psi \right\rangle \\
&= \frac{1}{(1-v^2)^{1/2}} \int_{\mathbb{R}^4} dt\, dx\, dy\, dz \times f\left(x,y,\frac{z-vt}{(1-v^2)^{1/2}}\right)\psi(t,x,y,z) \\
&= \frac{1}{(1-v^2)^{1/2}} \int_{\mathbb{R}^4} dt\, dx\, dy\, dz\, f(x,y,z) \times \psi\left(\frac{t+vz}{(1-v^2)^{1/2}},x,y,\frac{z+vt}{(1-v^2)^{1/2}}\right) \\
&= \int_{\mathbb{R}^4} d\sigma\, dx\, dy\, dz\, f(x,y,z)\psi(\sigma,x,y,\sigma v + (1-v^2)^{1/2}z).
\end{aligned} \tag{2.3}$$

Now we observe that the functions $f(x,y,z) \times \psi(\sigma,x,y,\sigma v + (1-v^2)^{1/2}z)$ are integrable on $\mathbb{R}^4$ and as $v\to\pm 1$ they converge pointwise to $f(x,y,z)\times\psi(\sigma,x,y,\pm\sigma)$ which are integrable. Besides that, $|f(x,y,z)\psi(\sigma,x,y,\sigma v + (1-v)^{1/2}z)| \leqslant C\chi(\sigma,x,y)|f(x,y,z)|$, where C is a constant and $\chi(\sigma,x,y)$ the characteristic function of the projection of $\mathrm{supp}(\psi)$ on three-dimensional space.

Then from the Lebesgue theorem[6] we infer that the right-hand side of (2.3) converges to

$$\int_{\mathbb{R}^3} d\sigma\, dx\, dy \left(\int_{-\infty}^{+\infty} f(x,y,z)dz \right) \psi(\sigma,x,y,\pm\sigma). \quad \blacksquare$$

3. LIGHTLIKE CONTRACTIONS OF THE ELECTROMAGNETIC MULTIPOLE FIELDS

Consider an electromagnetic field singularity, that of N-pole type, with time-independent structure. Then in its rest frame of reference the electromagnetic potential is of the form[7]

$$A^a = q^{ab_1\cdots b_N}\partial_{b_1}\cdots\partial_{b_N}(r^{-1}),^8 \tag{3.1}$$

where $r=(x^2+y^2+z^2)^{1/2}$, and constants $q^{ab_1\cdots b_N}$ are restricted by the following conditions:

(i) $q^{ab_1\cdots b_N} = \delta^a_0\epsilon^{b_1\cdots b_N} + \mu^{ab_1\cdots b_N}$, where $\mu^{ab_1\cdots b_N}$ is orthogonal to δ^a_0.

(ii) $\epsilon^{(b_1\cdots b_N)} = \epsilon^{b_1\cdots b_N}$ and $\mu^{a(b_1\cdots b_N)} = \mu^{ab_1\cdots b_N}$.

(iii) $\epsilon^{b_1\cdots b_N}$ and $\mu^{ab_1\cdots b_N}$ are orthogonal to δ^b_0 in all indices and are tracefree in $(b_1\cdots b_N)$ indices.

(iv) $\mu^{(ab_1\cdots b_N)} = 0$.

Next it is easy to construct $F=-dA$ and then $\Phi_v F$ with Φ_v given by (2.2). We also use the notation $(t,x,y,z)\equiv(x^0,x^1,x^2,x^3)$ and $(x^1,x^2)\equiv(x^B)$.

For convenience we discuss the electric and magnetic structures independently. To be as general as possible we admit the multipole moments to be dependent on v, which means that instead of one field F a family F_v is considered.

A. Electric type singularities

$$\Phi_v F = -\frac{1}{(1-v^2)^{1/2}}\epsilon^{b_1\cdots b_N}\partial_{b_1}\cdots\partial_{b_N}\left(\frac{1}{r}\right)_{,A}\circ\Phi_v^{-1}\,dx^A\wedge(dt-v\,dz) - \epsilon^{b_1\cdots b_N}\partial_{b_1}\cdots\partial_{b_N}\left(\frac{1}{r}\right)_{,z}\Phi_v^{-1}dz\wedge dt. \tag{3.2}$$

Consider the first part of (3.2). For an arbitrary test function ψ we have[4]

$$\begin{aligned}
&\left\langle -\frac{\epsilon^{b_1\cdots b_N}}{(1-v^2)^{1/2}}\partial_{b_1}\cdots\partial_{b_N}\left(\frac{1}{r}\right)_{,A}\circ\Phi_v^{-1},\psi\right\rangle \\
&= \left\langle -\frac{\epsilon^{b_1\cdots b_N}}{(1-v^2)^{1/2}}\partial_{b_1}\cdots\partial_{b_N}\left(\frac{1}{r}\right)_{,A},\psi\circ\Phi_v\right\rangle \\
&= \frac{\epsilon^{b_1\cdots b_N}}{(1-v^2)^{1/2}}(-1)^N\int_{\mathbb{R}^4}\frac{x^A}{r^3}\partial_{b_1}\cdots\partial_{b_N}\psi\left(\frac{t+vz}{(1-v^2)^{1/2}},x,y,\frac{z+vt}{(1-v^2)^{1/2}}\right)dt\,dx\,dy\,dz.
\end{aligned} \tag{3.3}$$

To discuss limits take at first the term in (3.3) of the form

$$I_v = \frac{\epsilon^{A_1\cdots A_N}}{(1-v^2)^{1/2}}(-1)^N\int_{\mathbb{R}^4}\frac{x^A}{r^3}\partial_{A_1}\cdots\partial_{A_N}\psi\left(\frac{t+vz}{(1-v^2)^{1/2}},x,y,\frac{z+vt}{(1-v^2)^{1/2}}\right)dt\,dx\,dy\,dz. \tag{3.4}$$

Then applying Lemma 1 with ψ replaced by $\psi_{,A_1\cdots A_N}$ for $f=x^A/r^3$, we obtain

$$\lim_{v\to\pm 1} I_v = 2(-1)^N\,\tilde{\epsilon}^{A_1\cdots A_N}\int_{\mathbb{R}^3}\frac{x^A}{\rho^2}\partial_{A_1}\cdots\partial_{A_N}\psi(\sigma,x,y,\pm\sigma)\,d\sigma\,dx\,dy = \langle 2\delta(t\mp z)\tilde{\epsilon}^{A_1\cdots A_N}\partial_{A_1}\cdots\partial_{A_N}(x^A/\rho^2),\psi\rangle, \tag{3.5}$$

where $\rho^2 = x^2+y^2$ and

$$\lim_{v\to\pm 1}\epsilon^{A_1\cdots A_N} := \tilde{\epsilon}^{A_1\cdots A_N}. \tag{3.6}$$

Next consider the term of (3.3) containing contractions with $\epsilon^{3A_1\cdots A_{N-1}}$.

$$I_v^{(1)} = \frac{(-1)^N N}{(1-v^2)^{1/2}}\epsilon^{3A_1\cdots A_{N-1}} \int_{\mathbf{R}^4} \frac{x^A}{r^3}\partial_{A_1}\cdots\partial_{A_{N-1}}\partial_z\psi\left(\frac{t+vz}{(1-v^2)^{1/2}},x,y,\frac{z+vt}{(1-v^2)^{1/2}}\right) dt\,dx\,dy\,dz$$
$$= \frac{(-1)^N N}{(1-v^2)^{1/2}}\epsilon^{3A_1\cdots A_{N-1}} \int_{\mathbf{R}^4} \frac{x^A}{r^3}\partial_{A_1}\cdots\partial_{A_{N-1}}(v\psi_{,1}+\psi_{,4})d\sigma\,dx\,dy\,dz, \tag{3.7}$$

where $\sigma = (t+vz)/(1-v^2)^{1/2}$ and $_{,1}$ and $_{,4}$ denote the partial derivatives with respect to the first and fourth arguments.

However, $\partial\psi/\partial\sigma = \psi_{,1} + v\psi_{,4}$; hence (3.7) takes the form of

$$I_v^{(1)} = (-1)^N N\epsilon^{3A_1\cdots A_{N-1}}(1-v^2)^{1/2}\int_{\mathbf{R}^4}\frac{x^A}{r^3}\partial_{A_1}\cdots\partial_{A_{N-1}}\psi_{,4}\,d\sigma\,dx\,dy\,dz. \tag{3.8}$$

Applying again Lemma 1 with $\partial_{A_1}\cdots\partial_{A_{N-1}}\psi_{,4}$ instead of ψ, for $f = x^A/r^3$ we infer that the limit of (3.8) as $v\to\pm 1$ is

$$\lim_{v\to\pm 1} I_v^{(1)} = 2(-1)^N N\tilde\epsilon^{3A_1\cdots A_{N-1}}\int_{\mathbf{R}^3}\frac{x^A}{\rho^2}\partial_{A_1}\cdots\partial_{A_{N-1}}\psi_{,4}(\sigma,x,y,\pm\sigma)d\sigma\,dx\,dy$$
$$= \langle \mp 2N\tilde\epsilon^{3A_1\cdots A_{N-1}}\delta'(t\mp z)(x^A/\rho^2)_{,A_1\cdots A_{N-1}},\psi\rangle, \tag{3.9}$$

where

$$\lim_{v\to\pm 1}(1-v^2)^{1/2}\epsilon^{3A_1\cdots A_{NN-1}} := \tilde\epsilon^{3A_1\cdots A_{N-1}}. \tag{3.10}$$

An application of the argument which led to (3.8), for the term of (3.3) containing contractions with $\epsilon^{3\cdots 3B_1\cdots B_{N-k}}$ for $1<k\leqslant N$, provides the following condition on the existence of the corresponding limit as $v\to\pm 1$

$$\lim_{v\to\pm 1}\epsilon^{3\cdots 3A_1\cdots A_{N-k}}(1-v^2)^{k/2} = \tilde\epsilon^{3\cdots 3A_1\cdots A_{N-k}}.$$

However, because of (ii) and (iii), $\epsilon^{3\cdots 3A_1\cdots A_{N-k}}$ is linearly dependent on $\epsilon^{A_1\cdots A_N}$ and $\epsilon^{3A_1\cdots A_{N-1}}$. Therefore [see (3.6) and (3.10)]

$$\lim_{v\to\pm 1}\epsilon^{3\cdots 3A_1\cdots A_{N-k}}(1-v^2)^{k/2} = 0 \quad \text{for } k>1. \tag{3.11}$$

Now consider the second part of (3.2).

$$\langle -\epsilon^{b_1\cdots b_N}\partial_{b_1}\cdots\partial_{b_N}(1/r)_{,z}\circ\Phi_v^{-1},\psi\rangle$$
$$= \langle -\epsilon^{b_1\cdots b_N}\partial_{b_1}\cdots\partial_{b_N}(1/r)_{,z},\psi\circ\Phi_v\rangle$$
$$= (-1)^N\epsilon^{b_1\cdots b_N}\int_{\mathbf{R}^4}\frac{z}{r^3}\partial_{b_1}\cdots\partial_{b_N}\psi\left(\frac{t+vz}{(1-v^2)^{1/2}},x,y,\frac{z+vt}{(1-v^2)^{1/2}}\right)dt\,dx\,dy\,dz. \tag{3.12}$$

The limit of (3.12) is equal to zero. Indeed (3.12) does not contain the factor $1/(1-v^2)^{1/2}$ as (3.3) does.

Thus the following theorem is true.

Theorem 1: Let F_v be a family of electromagnetic fields of electric N-poles, $N>0$. Then the lightlike contractions of the family $\Phi_v F_v$, as $v\to\pm 1$, exist, if there exist the limits

$$\lim_{v\to\pm 1}\epsilon^{A_1\cdots A_N} = \tilde\epsilon^{A_1\cdots A_N} \quad\text{and}\quad \lim_{v\to\pm 1}\epsilon^{3A_1\cdots A_{N-1}}(1-v^2)^{1/2} = \tilde\epsilon^{3A_1\cdots A_{N-1}}.$$

Then

$$\lim_{v\to\pm 1}\Phi_v F_v = 2\delta(t\mp z)\tilde\epsilon^{A_1\cdots A_N}[(x/\rho^2)_{,A_1\cdots A_N}dx + (y/\rho^2)_{,A_1\cdots A_N}dy]\wedge(dt\mp dz)$$
$$\mp 2N\delta'(t\mp z)\tilde\epsilon^{3A_1\cdots A_{N-1}}[(x/\rho^2)_{,A_1\cdots A_{N-1}}dx + (y/\rho^2)_{,A_1\cdots A_{N-1}}dy]\wedge(dt\mp dz).$$ ■

As a corollary we have

Corollary 1: Let F be an electromagnetic field of an electric N-pole, $N>0$. Then its lightlike limits as $v\to\pm 1$ are

$$\lim_{v\to\pm 1}\Phi_v F = 2\delta(t\mp z)\epsilon^{A_1\cdots A_N}[(x/\rho^2)_{,A_1\cdots A_N}dx + (y/\rho^2)_{,A_1\cdots A_N}dy]\wedge(dt\mp dz).$$ ■

B. Magnetic type singularities

$$\Phi_v F_v = \mu_A{}^{b_1\cdots b_N}\partial_{b_1}\cdots\partial_{b_N}(r^{-1})_{,B}\circ\Phi_v^{-1}\,dx^A\wedge dx^B$$
$$+\frac{1}{(1-v^2)^{1/2}}\mu_A{}^{b_1\cdots b_N}\partial_{b_1}\cdots\partial_{b_N}(r^{-1})_{,z}\circ\Phi_v^{-1}\,dx^A\wedge(dz - v\,dt)$$
$$-\frac{1}{(1-v^2)^{1/2}}\mu_z{}^{b_1\cdots b_N}\partial_{b_1}\cdots\partial_{b_N}(r^{-1})_{,A}\circ\Phi_v^{-1}\,dx^A\wedge(dz - v\,dt). \tag{3.13}$$

A discussion of the limits for (3.13) is similar to that for (3.2). Consider at first the second term in (3.13). The limit exists and is equal to zero if

$$\lim_{v\to\pm 1} \mu^{AB_1\cdots B_N} = \tilde{\mu}^{AB_1\cdots B_N} \tag{3.14}$$

and

$$\lim_{v\to\pm 1} \mu^{A\,3B_1\cdots B_{N-1}}(1-v^2)^{1/2} = \tilde{\mu}^{A\,3B_1\cdots B_{N-1}}. \tag{3.15}$$

[It vanishes because we have to integrate $(r^{-1})_{,z}$ with respect to z; compare with Lemma 1.] Then the limit of the first term is zero. The last term converges if

$$\lim_{v\to\pm 1} \mu^{3B_1\cdots B_N} = \tilde{\mu}^{3B_1\cdots B_N} \tag{3.16}$$

and

$$\lim_{v\to\pm 1} \mu^{33B_1\cdots B_{N-1}}(1-v^2)^{1/2} = \tilde{\mu}^{33B_1\cdots B_{N-1}}. \tag{3.17}$$

Of course we have to assume that the dependence on v is such that (3.15) and (3.16) are compatible [see (iv)]. Further, because of (ii), (iii), and (iv), (3.17) is a consequence of (3.14), and necessarily $\tilde{\mu}^{33B_1\cdots B_{N-1}} = 0$.

Hence the following theorem follows.

Theorem 2: Let F_v be a family of electromagnetic fields of magnetic N-poles, $N>0$. Then the lightlike contractions of the family $\Phi_v F_v$, as $v\to\pm 1$, exist if there exist the limits

$$\lim_{v\to\pm 1} \mu^{AB_1\cdots B_N} = \tilde{\mu}^{AB_1\cdots B_N}, \quad \lim_{v\to\pm 1} \mu^{3B_1\cdots B_N} = \tilde{\mu}^{3B_1\cdots B_N}$$

and

$$\lim_{v\to\pm 1} \mu^{A\,3B_1\cdots B_{N-1}}(1-v^2)^{1/2} = \tilde{\mu}^{A\,3B_1\cdots B_{N-1}},$$

where $\mu^{3B_1\cdots B_N}$ and $\mu^{A\,3B\cdots B_{N-1}}$ are related by (iv). Then

$$\lim_{v\to\pm 1} \Phi_v F_v = -2\delta(t\mp z)\tilde{\mu}^{3A_1\cdots A_N} \times [(x/\rho^2)_{,A_1\cdots A_N}dx + (y/\rho^2)_{,A_1\cdots A_N}dy] \wedge (dz\mp dt).$$ ∎

Corollary 2: Let F be an electromagnetic field of magnetic N-pole, $N>0$. Then its lightlike limits, as $v\to\pm 1$, are

$$\lim_{v\to\pm 1} \Phi_v F_v = -2\delta(t\mp z)\tilde{\mu}^{3A_1\cdots A_N} \times [(x/\rho^2)_{,A_1\cdots A_N}dx + (y/\rho^2)_{,A_1\cdots A_N}dy] \wedge (dz\mp dt).$$ ∎

4. LIGHTLIKE LIMITS FOR THE ELECTROMAGNETIC FIELD OF A SPINNING CHARGED PARTICLE

This field was discovered by P. Appel and it was discussed by many authors in various contexts.[9,10]

In the first part of this section we summarize its basic properties. We are interested in the global nature of the field; therefore the emphasis is put on the structure of singularities.

A. Derivation of the field

We recall at first that any electromagnetic field $F = \frac{1}{2} f_{\mu\nu}\, dx^\mu \wedge dx^\nu$ can be split into self-dual $F^{(+)}$ and anti-self-dual $F^{(-)}$ parts according to the formula

$$\begin{aligned} F &= \tfrac{1}{2} f_{\mu\nu}\, dx^\mu \wedge dx^\nu \\ &= \tfrac{1}{4}(f_{\mu\nu} + \check{f}_{\mu\nu})dx^\mu \wedge dx^\nu + \tfrac{1}{4}(f_{\mu\nu} - \check{f}_{\mu\nu})dx^\mu \wedge dx^\nu \\ &= F^{(+)} + F^{(-)}, \end{aligned} \tag{4.1}$$

where $\check{f}^{\mu\nu} = (i/2)\epsilon^{\mu\nu\rho\sigma} f_{\rho\sigma}$. Then

$$F = 2\,\mathrm{Re}\, F^{(+)}. \tag{4.2}$$

Consider next the Coulomb field for which

$$A = (e/r)\, dt \quad \text{and} \quad F = -dA. \tag{4.3}$$

One finds, then, that

$$2F^{(+)} = -d\omega, \tag{4.4}$$

where

$$\omega = \frac{e}{r}dt - \frac{eiz}{r}\frac{x\,dy - y\,dx}{x^2+y^2}. \tag{4.5}$$

The relation (4.4) unlike (4.3), does not hold globally, even if ω is considered as a generalized one-form. Indeed, (4.4) being satisfied would imply $j^\mu = 0$ (where j^μ denotes the current).

Now the self-dual part of the electromagnetic field of a spinning charged particle is obtained from (4.5) by a complex translation. Thus we have

$$\omega = \frac{e}{R}dt - ei\frac{z+ia}{R}\frac{x\,dy - y\,dx}{x^2+y^2}, \tag{4.6}$$

where $R = [x^2+y^2+(z+ia)^2]^{1/2}$.

B. Properties

Because of the remarks related to (4.4) we are interested, in fact, in the real part of (4.6) only. There is, however, an ambiguity in (4.6) which results from the fact that $R = [x^2+y^2+(z+ia)^2]^{1/2}$ has two branches as a complex function. (The branch point of the square root corresponds to the ring $x^2+y^2=a^2$, $z=0$.) Therefore, to construct a field one has to cut off a two-dimensional surface S spanned on that ring. Then those two branches can be joined appropriately, defining a field which is discontinuous on S. To see that explicitly let R^{-1} be represented as a sum of its real and imaginary parts. It occurs that

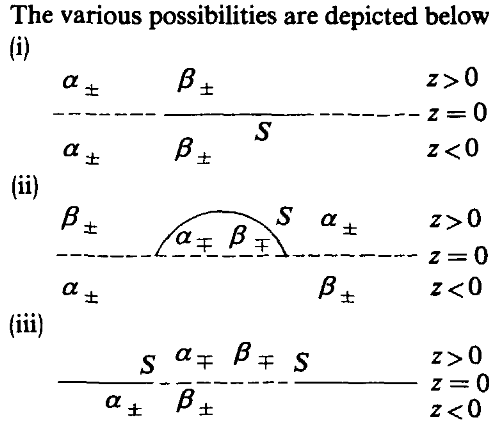
$$1/R = (\alpha_\pm + i\beta_\pm)[(r^2-a^2)^2 + 4a^2z^2]^{-1/2}, \tag{4.7}$$

where

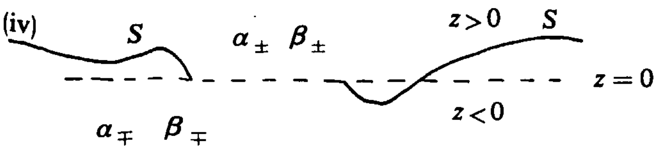
$$\alpha_\pm = \pm(1/\sqrt{2})\{r^2 - a^2 + [(r^2-a^2)^2 + 4a^2z^2]^{1/2}\}^{1/2}, \tag{4.8}$$

$$\beta_\pm = \mp(z/\sqrt{2}|z|)\{[(r^2-a^2)^2+4a^2z^2]^{1/2} - (r^2-a^2)\}^{1/2}. \tag{4.9}$$

The various possibilities are depicted below.



(iv)



Next we notice that there is a coordinate system well suited to the problem discussed here.[10] If (r,θ,Φ) denotes the usual spherical coordinates, then the new coordinates $(\tilde r,\tilde\theta,\tilde\Phi)$ are defined by the formulas

$$\tilde r=\{r^2-a^2+[(r^2-a^2)^2+4a^2r^2\cos^2\theta\,]^{1/2}\}^{1/2}\times 2^{-1/2}, \tag{4.10}$$

$$\cos\tilde\omega=r\cos\theta/\tilde r, \tag{4.11}$$

$$\cos\tilde\Phi=\frac{\tilde r\cos\Phi-a\sin\Phi}{(a^2+\tilde r^2)^{1/2}}, \tag{4.12}$$

$$\sin\tilde\Phi=\frac{\tilde r\sin\Phi+a\cos\Phi}{(a^2+\tilde r^2)^{1/2}}. \tag{4.13}$$

And the inverse transformation is

$$r=(\tilde r^2+a^2\sin^2\tilde\theta)^{1/2}, \tag{4.14}$$

$$\cos\theta=(\tilde r/r)\cos\tilde\theta, \tag{4.15}$$

$$\cos\Phi=\frac{\tilde r\cos\tilde\Phi+a\sin\tilde\Phi}{(\tilde r^2+a^2)^{1/2}}, \tag{4.16}$$

$$\sin\Phi=\frac{\tilde r\sin\Phi-a\cos\tilde\Phi}{(\tilde r^2+a^2)^{1/2}}. \tag{4.17}$$

The coordinates $(\tilde r,\tilde\theta,\tilde\Phi)$ as well as the spherical ones cover $\mathbb{R}^3$ almost everywhere. Indeed, one has to exclude from the domain of $(\tilde r,\tilde\theta,\tilde\Phi)$ the points on the z axis as well as those from the disk D: $z=0$, $x^2+y^2\leqslant a^2$. One can find out also that the surfaces of a constant $\tilde r$ are ellipsoids which degenerate to D as $\tilde r\to 0$.

The volume element in $(\tilde r,\tilde\theta,\tilde\Phi)$ coordinates is of the form

$$dV=(\tilde r^2+a^2\cos^2\tilde\theta)\sin\tilde\theta\,d\tilde r\,d\tilde\theta\,d\tilde\Phi. \tag{4.18}$$

The real part of ω (4.6) can be now represented in the form of

$$\begin{aligned} A=\operatorname{Re}\omega=&\pm\frac{e\tilde r}{(\tilde r^2+a^2\cos^2\tilde\theta}\,dt\\ &\pm\frac{ea\tilde r\sin\tilde\theta\,(\tilde r\cos\tilde\Phi+a\sin\tilde\Phi)}{(\tilde r^2+a^2\cos\tilde\theta)(\tilde r^2+a)}\,dy\\ &\mp\frac{ea\tilde r\sin\tilde\theta\,(\tilde r\sin\tilde\Phi-a\cos\tilde\Phi)}{(\tilde r^2+a^2\cos^2\tilde\theta)(\tilde r^2+a^2)}\,dx, \end{aligned} \tag{4.19}$$

where signs depend on the physical situation, one among (i)–(iv).

One observes then, that the components of A are locally integrable (compare with 4.18). One can show also that the same property holds for the electromagnetic field $F=-dA$. The field itself is singular on the ring $z=0$, $x^2+y^2=a^2$ only, and is discontinuous on S.

Henceforth we discuss the case (i) only, and we choose the branch of R^{-1} in (4.7) corresponding to α_+ and β_+, which are denoted further by α and β.

C. Limit transitions

The electromagnetic field of a spinning charged particle is axially symmetric. Therefore, to find its lightlike limits in an arbitrary direction, it suffices to consider a family of electromagnetic fields $(\Phi_v\circ R_\delta)F$, where R_δ is the rotation in the (x,z) plane by the angle δ, $0\leqslant\delta\leqslant\pi/2$, while Φ_v is as before in (2.1).

$$\begin{aligned} t\circ R_\delta^{-1}&=t,\\ x\circ R_\delta^{-1}&=x\cos\delta-z\sin\delta,\\ y\circ R_\delta^{-1}&=y,\\ z\circ R_\delta^{-1}&=x\sin\delta+z\cos\delta. \end{aligned} \tag{4.20}$$

Then one finds that

$$\begin{aligned} -(\Phi_v\circ R_\delta)F=&\frac{1}{(1-v^2)^{1/2}}\frac{\partial^\sim}{\partial x}A_t\,dx\wedge(dt-v\,dz)+\frac{1}{(1-v^2)^{1/2}}\frac{\partial^\sim}{\partial y}A_t\,dy\wedge(dt-v\,dz)\\ &-\frac{1}{(1-v^2)^{1/2}}\Big(\sin\delta\frac{\partial^\sim}{\partial x}A_x+\cos\delta\frac{\partial^\sim}{\partial z}A_x\Big)dx\wedge(dz-v\,dt)\\ &-\frac{1}{(1-v^2)^{1/2}}\Big(\frac{\partial^\sim}{\partial z}A_y+\sin\delta\frac{\partial^\sim}{\partial y}A_x\Big)dy\wedge(dz-v\,dt)\\ &+\frac{\partial^\sim}{\partial z}A_t\,dz\wedge dt+\Big(\frac{\partial^\sim}{\partial x}A_y-\cos\delta\frac{\partial^\sim}{\partial y}A_x\Big)dx\wedge dy, \end{aligned} \tag{4.21}$$

where

$$\frac{\partial^\sim}{\partial x}A_t:=\Big(\frac{\partial}{\partial x}(A_t\circ R_\delta^{-1})\Big)\circ\Phi_v^{-1},$$

$$\frac{\partial^\sim}{\partial y}A_t:=\Big(\frac{\partial}{\partial y}(A_t\circ R_\delta^{-1})\Big)\circ\Phi_v^{-1},$$

and so on.

Now it is clear that to find the limits of (4.21) we can apply Lemma 1.

It is immediately seen that the last two terms of (4.21) vanish as v approaches 1 or -1. Indeed, they do not contain the factor $1/(1-v^2)^{1/2}$ [see (2.3)]. Next we observe that the terms like $(1/(1-v^2)^{1/2})(\partial^\sim/\partial z)A_x$ also converge to zero. Indeed, let ψ be an arbitrary test function; then

$$\begin{aligned} &\lim_{v\to\pm1}\Big\langle\frac{1}{(1-v^2)^{1/2}}\frac{\partial^\sim}{\partial z}A_x,\psi\Big\rangle\\ &\quad=\int_{\mathbb{R}^3}d\sigma\,dx\,dy\Big(\int_{-\infty}^{+\infty}dz\,\frac{\partial}{\partial z}(A_x\circ R_\delta^{-1})\Big)\\ &\qquad\times\psi(\sigma,x,y,\pm\sigma)\equiv 0. \end{aligned}$$

The explicit form of the remaining limits depend on the integrals

$$\int_{-\infty}^{+\infty} dz \frac{\partial}{\partial z}(A_x \circ R_\delta^{-1}),\quad \int_{-\infty}^{+\infty} dz \frac{\partial}{\partial y}(A_t \circ R_\delta^{-1}),\quad \int_{-\infty}^{+\infty} dz \frac{\partial}{\partial x}(A_x \circ R_\delta^{-1}),$$

and

$$\int_{-\infty}^{+\infty} dz \frac{\partial}{\partial y}(A_x \circ R_\delta^{-1}).$$

Making use of (4.6) and (4.20) we easily find that

$$A_t \circ R_\delta^{-1} = \mathrm{Re}\, \frac{e}{[(x+ia\sin\delta)^2 + y^2 + (z+ia\cos\delta)^2]^{1/2}}, \tag{4.22}$$

and then

$$\frac{\partial}{\partial x}(A_t \circ R_\delta^{-1}) = -e\, \mathrm{Re}\, \frac{x+ia\sin\delta}{[y^2 + (x+ia\sin\delta)^2]} \left(\frac{z+ia\cos\delta}{[(x+ia\sin\delta)^2 + y^2 + (z+ia\cos\delta)^2]^{1/2}}\right)_{,z}, \tag{4.23}$$

$$\frac{\partial}{\partial y}(A_t \circ R_\delta^{-1}) = -e\, \mathrm{Re}\, \frac{y}{[y^2 + (x+ia\sin\delta)^2]} \left(\frac{z+ia\cos\delta}{[(x+ia\sin\delta)^2 + y^2 + (z+ia\cos\delta)^2]^{1/2}}\right)_{,z}. \tag{4.24}$$

Next one shows that

$$A_x \circ R_\delta^{-1} = e\, \mathrm{Re}\, \frac{iy[(x+ia\sin\delta)\sin\delta + (z+ia\cos\delta)\cos\delta]}{\{y^2 + [(x+ia\sin\delta)\cos\delta - (z+ia\cos\delta)\sin\delta]^2\}[(x+ia\sin\delta)^2 + y^2 + (z+ia\cos\delta)^2]^{1/2}}, \tag{4.25}$$

and after some additional work one obtains

$$\frac{\partial}{\partial x}(A_x \circ R_\delta^{-1}) = e\, \mathrm{Re} \left\{\frac{iy\{(\xi^2+y^2+\eta^2)(\eta\sin\delta - \xi\cos\delta) - \xi\eta(\eta\cos\delta + \xi\sin\delta)\}}{(\xi^2+y^2)(\xi^2+y^2+\eta^2)^{1/2}[y^2 + (\eta\sin\delta - \xi\cos\delta)^2]}\right\}_{,z}, \tag{4.26}$$

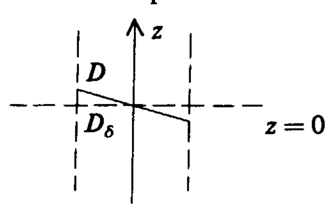
$$\frac{\partial}{\partial y}(A_x \circ R_\delta^{-1}) = -e\, \mathrm{Re} \left\{\frac{i\{\xi(\eta\sin\delta - \xi\cos\delta)(\xi^2+y^2+\eta^2) + \eta y^2(\eta\cos\delta + \xi\sin\delta)\}}{(\xi^2+y^2)(\xi^2+y^2+\eta^2)^{1/2}[y^2 + (\eta\sin\delta - \xi\cos\delta)^2]}\right\}_{,z}, \tag{4.27}$$

where $\xi = x + ia\sin\delta$ and $\eta = z + ia\cos\delta$.

A further comment is needed to clarify the formulas (4.23)–(4.27). All of them contain under the sign of differentiation with respect to z the term

$$\frac{1}{R} \circ R_\delta^{-1} = \frac{\alpha + i\beta}{[(r^2 - a^2)^2 + 4a^2z^2]^{1/2}} \circ R_\delta^{-1},$$

which is discontinuous on D. D is a disk region: $x\sin\delta + z\cos\delta = 0$, $(x\cos\delta - z\sin\delta)^2 + y^2 < a^2$. Thus, integrations with respect to z have to be performed cautiously. Indeed, the integration along a line within the cylinder defined by the projection of D on the $z=0$ plane has to be split into two parts defined by the point of intersection of that line with D



Then one obtains

$$G \equiv \int_{-\infty}^{+\infty} \left(\frac{\partial^\sim}{\partial x} A_t + \sin\delta \frac{\partial^\sim}{\partial x} A_x\right) dz = -\frac{2x}{x^2 + (y + a\sin\delta)^2} + \frac{2x(a + y\sin\delta)\Theta_\delta(x,y)}{(a^2\cos^2\delta - y^2\cos^2\delta - x^2)^{1/2}[x^2 + (y + a\sin\delta)^2]}, \tag{4.28}$$

$$H \equiv \int_{-\infty}^{+\infty} \left(\frac{\partial^\sim}{\partial y} A_t + \sin\delta \frac{\partial^\sim}{\partial y} A_x\right) dz = -\frac{2(y + a\sin\delta)}{x^2 + (y + a\sin\delta)^2} + \frac{2[a\cos^2\delta(y + a\sin\delta) - x^2\sin\delta]\Theta_\delta(x,y)}{(a^2\cos^2\delta - y^2\cos^2\delta - x^2)^{1/2}[x^2 + (y + a\sin\delta)^2]}, \tag{4.29}$$

where $\Theta_\delta(x,y)$ is a characteristic function of the subset $D_\delta \subset \mathbb{R}^2$, $D_\delta = \{(x,y): x^2 + y^2\cos^2\delta < a^2\cos^2\delta\}$.

The lightlike limits of the electromagnetic field are

$$\lim_{v \to \pm 1} F_v = -\delta(t \mp z)(G\,dx + H\,dy) \wedge (dt \mp dz). \tag{4.30}$$

The field corresponding to $\delta = \pi/2$ can be obtained from (4.30) by taking the limit $\delta \to \pi/2$.

D. Discussion

As one could expect, contracted electromagnetic fields are of Dirac's delta type on null hyperplanes. It was demonstrated also in Sec. 2 that during the process of contraction the multipole structures of higher order than monopole can survive. For the electromagnetic field of a spinning, charged particle the relation between them is, however, such that in the limit the external field differs from that of the Coulomb merely by the translation. Therefore, the external field does not contain information about the "spin" of a particle.

If the global nature of the field is taken into account then one observes that in fact we have a one-parameter family of fields, labeled by the parameter δ, with sources confined to the D_δ region.

[1]P. G. Bergmann, *Introduction to the Theory of Relativity* (Prentice-Hall, Englewood Cliffs, NJ, 1942).
[2]P. C. Aichelberg and R. V. Sexl, Lett. Nuovo Cimento **4**, 1316 (1970).
[3]J. F. Plebański, Ann. Phys. (N.Y.) **90**, 196 (1975).
[4]L. Schwartz, *Mathematics for the Physical Sciences* (Addison-Wesley, Reading, MA, 1966).
[5]The theory of distributions on the Minkowski space-time is in fact the theory of distributions on $\mathbf{R}^4$. Generalized tensor fields are defined by their components in global Minkowski coordinate systems; the components are assumed to be distributions.
[6]H. L. Royden, *Real Analysis* (Macmillan, New York, 1963).
[7]F. A. E. Pirani, *Lectures on General Relativity*, Vol. 1, 277; Brandeis Summer Institute in Theoretical Physics (Prentice-Hall, Englewood Cliffs, NJ, 1964).
[8]The derivatives of r^{-1}, which is locally integrable on $\mathbf{R}^3$, are understood in the distributional sense (Ref. 4).
[9]E. T. Whittacker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U.P., New York, 1969), p. 400.
[10]W. Israel, Phys. Rev. D **2**, 641 (1970).

Quantization, symmetry, and natural polarization

V. Aldaya, J. A. de Azcárraga, and K. B. Wolf[a)]
Departamento de Física Teórica, Facultad de C. Físicas, Burjasot (Valencia), Spain

(Received 21 December 1981; accepted for publication 11 February 1983)

We discuss the notion of polarization, as defined in a geometric quantization scheme recently introduced, in terms of the role played by the evolution operator of the quantum system. The analysis uses an integral transform representation of the group WSp(2,$\mathbb{R}$). This clarifies the group theoretic origin of the natural polarizations and the meaning of the polarization changing transformations.

PACS numbers: 03.65. – w, 02.20. + b, 02.30.Qy

1. INTRODUCTION AND STATEMENT OF THE PROBLEM

A method of geometric quantization has been introduced very recently[1], based on the definition of a group $\widetilde{G}$, the quantum group, from which all the essential ingredients of the theory are derived.[2] In the terminology of the usual quantization scheme[4] this means that given $\widetilde{G}$ there exists a procedure to construct the quantum manifold and the basic quantum operators (which prequantizes the system) and that a full quantization may be achieved by defining the appropriate polarization. This method uses the definition of the canonical left 1-form on $\widetilde{G}$, and in particular of its vertical component Θ {the "verticality" being defined by the fact that $\widetilde{G}$ [U(1), $\widetilde{G}$/U(1)$\equiv G$] is a principal bundle of structure group U(1)} as well as of the right and left invariant vector fields on $\widetilde{G}$ in order to define quantum operators and polarizations, respectively. In fact, the use of $\widetilde{G}$ as the starting point allows us to work directly on the evolution space, making it unnecessary to base the theory on the usual (contact) quantum manifold which, nevertheless may be obtained from $\widetilde{G}$ if one so wishes,[1] the contact 1-form being derived from Θ.

The so-called quantum group $\widetilde{G}$ is defined as a central extension by U(1) of a group $G_{(k)}$ (dynamical group[5] of the system under study) which contracts to the usual Galilei group for the case of the free system. If the problem under consideration is that of an interacting particle, the constant k in $G_{(k)}$ is a constant which switches off the potential in the limit $k \to 0$. Once $\widetilde{G}$ has been determined, the quantization may be performed by means of the following steps[1]:

(a) derivation of the left and the right invariant vector fields (LIVF and RIVF);

(b) construction of the canonical left 1-form;

(c) definition of the basic quantum operators (by means of RIVF);

(d) definition of the polarization (by means of LIVF).

In the usual approach to geometric quantization, the last step—that of defining a suitable polarization—is the least precise one. In the approach based on the quantum group $\widetilde{G}$, a polarization may be defined as a subspace of $\mathfrak{X}^{\mathrm{L}}(\widetilde{G})$—the space of L-vector fields on $\widetilde{G}$—which contains $\mathscr{C}_\Theta$—the characteristic module of Θ[6]—which is projected onto a subalgebra of the Lie algebra of $\mathfrak{X}^{\mathrm{L}}(G)$. It is the purpose of this paper to explore further the proposed definition to exhibit how it leads to the definition of a natural polarization associated with the system (and accordingly, with the corresponding quantum group) under consideration.

This will be done by considering the following four *one-dimensional* systems: the free particle, the free fall, the harmonic oscillator, and the repulsive "oscillator." The potentials which correspond to these situations are the representative of four orbits of the wsp(2,$\mathbb{R}$) algebra under the adjoint action of the corresponding group (Sec. 2). These are *all* orbits which include sp(2,$\mathbb{R}$) elements (containing the kinetic energy, as we shall see in Sec. 2). Thus, the group WSp(2,$\mathbb{R}$) includes the quantum dynamical groups of the above one-dimensional systems and will be used as the starting point for their study. To this aim, Sec. 2 will be devoted to describe the WSp(2,$\mathbb{R}$) group, its algebra, and the group of integral transform associated with it. Section 3 will present the simplest case of the free particle and the derivation of its Schrödinger equation. It will be found that the natural polarization leads to the momentum space Schrödinger equation, and that configuration space is obtained by means of an automorphism in Sp(2,$\mathbb{R}$) (the Fourier-integral transform). The essentially similar case of the free fall will be considered in Sec. 4. Sections 5 and 6 will be devoted to the harmonic oscillator and the repulsive "oscillator" (which is obtained through a change of sign in the harmonic oscillator potential). In all cases the vector fields of $\mathfrak{X}^{\mathrm{L}}(\widetilde{G})$ defining the polarizations will be found to define maximal invariant subgroups in the corresponding "classical groups" $G = \widetilde{G}/\mathrm{U}(1)$ in accordance with the given definition. Moreover, the condition that the polarization vector fields must include $\mathscr{C}_\Theta$ will turn out to be very helpful in finding the polarization. Indeed, $\mathscr{C}_\Theta$ characterizes the time evolution of the system, and so the determination of the polarization will be associated with the problem of diagonalizing the time evolution (Hamiltonian) part in $\widetilde{G}$. This will immediately yield the suitable polarization for each case. In particular, the polarization which corresponds to the harmonic oscillator case will be found to be the one which leads to the Bargmann–Fock–Segal picture, and the "rotation" which leads to the usual harmonic oscillator eigenfunctions will be seen to be the Bargmann transform.

2. THE WSp(2,$\mathbb{R}$) GROUP

This group is the semidirect product of the two-dimensional real symplectic group Sp(2,$\mathbb{R}$) [$\approx$SU(1,1)$\approx$SL(2,$\mathbb{R}$)]

a) Permanent address: IIMAS, Universidad Nacional Autónoma de México, Apdo. Postal 20-726, 01000, México D.F., Mexico.

by the Weyl group W as a normal factor. We shall present this group in the context of a particularly convenient basis. Consider the three abstract operators $\mathbb{Q}$, $\mathbb{P}$, $\mathbb{I}$, as a basis for a Lie algebra with brackets

$$[\mathbb{Q},\mathbb{P}] = i\mathbb{I}, \quad [\mathbb{Q},\mathbb{I}] = 0, \quad [\mathbb{P},\mathbb{I}] = 0, \tag{2.1}$$

which is the well-known Heisenberg–Weyl algebra w.[7] Consider now the universal enveloping algebra $\bar{w}$ as a Lie algebra induced by (2.1) through the Leibnitz rule and in particular the finite-dimensional subalgebra whose basis is the set of the three independent second-order generators

$$\mathbb{P}^2, \quad \{\mathbb{Q},\mathbb{P}\}_+ \equiv \mathbb{Q}\mathbb{P} + \mathbb{P}\mathbb{Q}, \quad \mathbb{Q}^2 \tag{2.2}$$

in the representation where $\mathbb{I}$ is the unit operator. The operators (2.2) constitute a basis for the metaplectic representation[8] of the symplectic algebra sp(2,$\mathbb{R}$). When we sum (2.1) and (2.2) as vector spaces, we naturally obtain the six-dimensional algebra wsp(2,$\mathbb{R}$), where w is a three-dimensional ideal. The exponential mapping of this Lie algebra yields the WSp(2,$\mathbb{R}$) group together with a local parametrization of its manifold[9]

$$\exp(i[\alpha\mathbb{P}^2 + \beta\{\mathbb{Q},\mathbb{P}\}_+ + \gamma\mathbb{Q}^2 + \delta\mathbb{Q} + \epsilon\mathbb{P} + \theta\mathbb{I}])$$
$$\equiv g\left\{\begin{pmatrix} a & b \\ c & d\end{pmatrix}, \ (p,q), \ z\right\}, \tag{2.3a}$$

where $ad - bc = 1$ and

$$\begin{aligned}
a &= \cos 2s - \beta s^{-1}\sin 2s, \quad s = \pm\sqrt{(\alpha\gamma - \beta^2)},\\
b &= -\alpha s^{-1}\sin 2s,\\
c &= \gamma s^{-1}\sin 2s,\\
d &= \cos 2s + \beta s^{-1}\sin 2s,\\
p &= \tfrac{1}{2}(\gamma\epsilon - \beta\delta)s^{-2}(1-\cos 2s) + \tfrac{1}{2}\delta s^{-1}\sin 2s,\\
q &= \tfrac{1}{2}(\beta\epsilon - \alpha\delta)s^{-2}(1-\cos 2s) + \tfrac{1}{2}\epsilon s^{-1}\sin 2s,\\
z &= \theta - \tfrac{1}{4}(\alpha\delta^2 + \gamma\epsilon^2 - 2\beta\delta\epsilon)s^{-2}(1 - \tfrac{1}{2}s^{-1}\sin 2s).
\end{aligned} \tag{2.3b}$$

The group composition law, $*$, may be established to be

$$g'\{\mathbf{M}',\mathbf{u}',z'\} * g\{\mathbf{M},\mathbf{u},z\}$$
$$= g''\{\mathbf{M}'\mathbf{M}, \mathbf{u}'\mathbf{M} + \mathbf{u}, \ z' + z + \tfrac{1}{2}\mathbf{u}'\mathbf{M}\Omega\mathbf{u}^T\}, \tag{2.4a}$$

where

$$\mathbf{M} \equiv \begin{pmatrix} a & b \\ c & d\end{pmatrix}, \quad \mathbf{u} \equiv (p,q), \quad \mathbf{u}^T \equiv \begin{pmatrix} p \\ q\end{pmatrix}, \quad \Omega = \begin{pmatrix} 0 & -1 \\ 1 & 0\end{pmatrix}. \tag{2.4b}$$

The unit element is given by $e = g\{\mathbb{I},\mathbf{0},0\}$ and the inverse is $[g\{\mathbf{M},\mathbf{u},z\}]^{-1} = g\{\mathbf{M}^{-1}, -\mathbf{u}\mathbf{M}^{-1}, -z\}$. The adjoint action of the group on the algebra may be found, and it can be ascertained to consist of six distinct orbits.[10] Representative of these orbits are $\mathbb{P}^2$, $\mathbb{P}^2 + \mathbb{Q}$, $\mathbb{P}^2 + \mathbb{Q}^2$, $\mathbb{P}^2 - \mathbb{Q}^2$, $\mathbb{P}$, and $\mathbb{I}$, the first four of which are germane to the present work since they include a $\mathbb{P}^2$ term. The algebra wsp(2,$\mathbb{R}$) is thus the common dynamical algebra for the free particle, the linear potential (or free fall, $\mathbb{P}^2 + \mathbb{Q}$), the harmonic ($\mathbb{P}^2 + \mathbb{Q}^2$) and the repulsive ($\mathbb{P}^2 - \mathbb{Q}^2$) oscillators. Moreover, this six-dimensional algebra is the largest finite algebra with a semisimple factor within $\bar{w}$.[11]

Let us now turn to the generalized representation basis where $\mathbb{Q}$ is diagonal, i.e., $\mathbb{Q}f(x) = xf(x)$. In this basis, the Weyl group acts as a Lie transformation group through

$$g{:}f(x) \mapsto [g\{\mathbb{I},(p,q),z\}f](x)$$
$$= \exp[i(z + \tfrac{1}{2}pq + px)]f(x+q). \tag{2.5a}$$

The Sp(2,$\mathbb{R}$) group is generated by *up-to-second*-order operators in these generators, and its action is that of a group of integral transforms[12]:

$$g{:}f(x) \mapsto \left[g\left\{\begin{pmatrix} a & b \\ c & d\end{pmatrix},0,0\right\}f\right](x) = \int_{\mathbb{R}} dx'\, C_{\mathbf{M}}(x,x')f(x') \tag{2.5b}$$

with kernel[13]

$$C_{\mathbf{M}}(x,x') = (2\pi b)^{-1/2}e^{-i\pi/4}\exp[i(ax'^2 - 2xx' + dx^2)/2b] \tag{2.5c}$$

which is unitary in $\mathscr{L}^2(\mathbb{R})$. In the two-parameter subgroup of Sp(2,$\mathbb{R}$) of lower triangular matrices [those with $b = 0$, generated by all algebra elements with no $\mathbb{P}^2$ summand in (2.3a), $\alpha = 0$] the integral transform (2.5b) collapses to a Lie transformation group action,

$$g{:}f(x) \mapsto \left[g\left\{\begin{pmatrix} a & 0 \\ b & a^{-1}\end{pmatrix},0,0\right\}f\right](x)$$
$$= a^{-1/2}\exp[(icx^2/2a)]f(x/a). \tag{2.6}$$

The full WSp(2,$\mathbb{R}$) group action may be obtained through the composition of (2.5a) and (2.5b). The composition of integral transforms (2.5) follows the group composition property (2.4) modulo a sign[12] [it is a faithful representation of the *two*-fold covering of Sp(2,$\mathbb{R}$), called the metaplectic group Mp(2,$\mathbb{R}$)]. Thus, there exists a local isomorphism between WSp(2,$\mathbb{R}$), the hiperdifferential operators (2.3a), and the integral transforms (2.5). In the limit $\mathbf{M}\rightarrow\mathbb{I}$ (or $b\rightarrow 0$, with $\arg b\in[-\pi,0]$), the integral kernel (2.5c) has as weak limit a Dirac δ, as can be inferred from (2.6). Finally, the inverse of the integral transform (2.5) associated with a matrix $\mathbf{M}$ is that corresponding to $\mathbf{M}^{-1}$; it has a kernel which is the complex conjugate of the original one.

One can subject the WSp(2,$\mathbb{R}$) integral transform action to analytic continuation in the complex parameter plane, with certain restrictions. In order that the domain remain $\mathscr{L}^2(\mathbb{R})$, one needs that $\operatorname{Im} b^*a \geqslant 0$ and, when $a = 0$, then $\operatorname{Im} b = 0$. [This defines a subsemigroup of Sp(2,$\mathbb{C}$), called HSp(2,$\mathbb{C}$).] The elements of the complex-extended integral transform semigroup can be made to have the *unitarity* property when we consider them as mappings between $\mathscr{L}^2(\mathbb{R})$ and Bargmann–Fock–Segal–type Hilbert spaces of analytic functions[14] $\mathscr{B}_{\mathbf{M}}$ whose defining inner product is on the complex plane:

$$(f,g)_{\mathbf{M}} = \int_C d\mu_{\mathbf{M}}(\xi,\xi^*)f^*(\xi)g(\xi), \tag{2.7a}$$

$$d\mu_{\mathbf{M}}(\xi,\xi^*) = 2(2\pi\nu)^{-1/2}\exp[(v\xi^2 - 2\xi\xi^* + v^*\xi^{*2})/2\nu]d\operatorname{Re}\xi\, d\operatorname{Im}\xi, \tag{2.7b}$$

$$v = a^*d - b^*c, \quad \nu = 2\operatorname{Im} b^*a. \tag{2.7c}$$

The transform kernel inverse to that corresponding to $\mathbf{M}$ [(2.7b)] is obtained by putting $\mathbf{M}^{-1}$ in (2.7a), and thus maps $\mathscr{B}_{\mathbf{M}}$ back to $\mathscr{L}^2(\mathbb{R})$.

In essence, we use HSp(2,$\mathbb{C}$) in order to act through similarity on the relevant ("Galilei"-type) quantum group $\widetilde{G}$, and correspondingly on its generators (adjoint action of the group on a subalgebra). This is required to perform the task of diagonalizing the Hamiltonian-generated time evolution group [see (2.3a)]. If this matrix cannot be diagonalized—the first two cases to be considered—then it should at least reduce to a triangular matrix having an invariant subspace.

In terms of generators, diagonalizing the Hamiltonian under consideration means rotating it onto the change-of-scale operator $\frac{1}{2}\{\mathbb{P},\mathbb{Q}\}_+ = -i(x\partial/\partial x + \frac{1}{2})$. This may be done using Sp(2,$\mathbb{R}$) transformations in the repulsive oscillator case, where $\frac{1}{2}(\mathbb{P}^2 - \mathbb{Q}^2)$ and $\frac{1}{2}\{\mathbb{P},\mathbb{Q}\}_+$ belong to the same orbit under Sp(2,$\mathbb{R}$). In the harmonic oscillator case, $\frac{1}{2}(\mathbb{P}^2 + \mathbb{Q}^2)$ and $\frac{1}{2}\{\mathbb{P},\mathbb{Q}\}_+$ are in different orbits under Sp(2,$\mathbb{R}$), but on the same orbit under HSp(2,$\mathbb{C}$). They are mapped onto each other by the Bargmann transform; this will change the structure of (i.e., the inner product defining) the Hilbert space.

Having defined this operational machinery, we now turn to the study of polarizations for the approach described in the previous section and for the four one-dimensional systems mentioned above.

3. THE CASE OF THE FREE PARTICLE

This is the simplest case; since there is no interaction, $k=0$ (Sec. 1), and $\widetilde{G}$ is directly given by the central extension of the ordianry Galilei group G by U(1), i.e., $\widetilde{G} \equiv \widetilde{G}_{(m)}$.[15] Since for the free particle only the kinetic term is relevant, the elements of $\widetilde{G}_{(m)}$ are given by the subgroup of WSp(2,$\mathbb{R}$), which is generated by $H_{\text{free}} = \mathbb{P}^2/2m$ $(\alpha = -t/2m)$ and the Weyl subgroup in (2.3a), i.e., by

$$\begin{aligned}\widetilde{G}_{(m)} &= g\left\{\begin{pmatrix}1 & t/m\\ 0 & 1\end{pmatrix},\ (p,q),\ \theta\right\}\\ &= \exp\left(-i\frac{\mathbb{P}^2}{2m}t\right)\exp[i(p\mathbb{Q} + q\mathbb{P} + \theta\mathbb{I})]. \qquad (3.1)\end{aligned}$$

The composition law (2.4) applied to (3.1) reproduces thus the usual one:

$$\begin{aligned}t'' &= t' + t,\\ p'' &= p' + p,\\ q'' &= q' + q + (p'/m)t,\\ \theta'' &= \theta' + \theta + \tfrac{1}{2}(pq' - p'q) + (1/2m)pp't\end{aligned} \qquad (3.2)$$

in terms of the evolution space variables (t, p, q) and the Bargmann cocyle[15] for the U(1) part of $\widetilde{G}_{(m)}$.

The left invariant vector fields are easily derived from (3.2) with the result

$$\begin{aligned}X_t^L &= \frac{\partial}{\partial t} + \frac{p}{m}\frac{\partial}{\partial q}, \quad [X_v^L, X_t^L] = X_q^L,\\ X_q^L &= \frac{\partial}{\partial q} - \frac{1}{2}p\Xi, \quad [X_q^L, X_v^L] = mX_\zeta^L,\\ X_v^L &= m\frac{\partial}{\partial p} + \frac{m}{2}q\Xi, \quad [\text{all others}] = 0,\\ X_\zeta^L &= i\zeta\frac{\partial}{\partial \zeta} \equiv \Xi,\end{aligned} \qquad (3.3)$$

where $\zeta = \exp i\theta$ and the subindices t, q, and v indicate that they parametrize translations in time, position, and velocity ("boosts"). The "vertical" component of the canonical form satisfies $\Theta(X_t; X_q; X_v) = 0$, $\Theta(X_\zeta) = 1$ and is given by $\Theta = \frac{1}{2}(p\,dq - q\,dp) - (p^2/2m)dt + d\zeta/i\zeta$; it is easily verified that $\mathscr{C}_\Theta$ generated by X_t.

It is now clear that the definition of polarization given in Sec. 1 requires us to take as wave functions $\mathbb{C}$-valued functions on $\widetilde{G}_{(m)}$ satisfying the conditions

$$X_t^L \cdot \psi(q,p,\zeta,t) = 0, \qquad (3.4a)$$

$$X_q^L \cdot \psi(q,p,\zeta,t) = 0, \qquad (3.4b)$$

and[16]

$$\Xi \cdot \psi = i\psi. \qquad (3.5)$$

Thus, X_q and X_t generate the polarizations (X_q and X_t determine a maximal—and abelian—invariant subgroup in G.[17]) In fact, X_t gives in (3.4a) the Schrödinger equation in momentum space $i(\partial/\partial t)\psi = (p^2/2m)\psi$ once (3.4b) and (3.5) have been taken into account. ($\hbar = h/2\pi$ has been put equal to unity throughout.)

The polarization defined by X_q we may now call the *natural* polarization by observing that the part of $\widetilde{G}_{(m)}$, which corresponds to the evolution operator

$$g\left\{\begin{pmatrix}1 & t/m\\ 0 & 1\end{pmatrix},(0,0),0\right\} \qquad (3.6)$$

leaves the subspace $(0, q)$ invariant; this is a way of rephrasing the definition of Sec. 1. The fact that the triangular matrix (3.6) is not diagonizable indicates that this is the only polarization which may be defined for the free particle in a natural way. Of course, we may now give the Schrödinger equation in configuration space. To do this, it is necessary to apply a transformation [an automorphism external to $\widetilde{G}_{(m)}$ but internal to Sp(2,$\mathbb{R}$)] capable of interchanging the roles of p and q. Such a transformation is (up to a phase $e^{-i\pi/4}$) the Fourier transform given by

$$g_{\mathscr{F}}\left\{\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix},0,0\right\}, \qquad (3.7)$$

which determines the well-known integral kernel $1/\sqrt{2\pi}\exp(-ipq)$ in (2.5c).

4. THE FREE FALL

We now turn to the quantization of the particle in free fall. The dynamical group for this case is the subgroup of WSp(2,$\mathbb{R}$) given by

$$\exp(-itH_{ff})\exp[i(p\mathbb{Q} + q\mathbb{P} + \theta\mathbb{I})], \qquad (4.1)$$

where

$$H_{ff} = \mathbb{P}^2/2m + F\mathbb{Q}. \qquad (4.2)$$

Using (2.3) for each factor and (2.4) for the product of the two exponentials, we obtain

$$\begin{aligned}\tilde{g}\Big\{&\begin{pmatrix}1 & t/m\\ 0 & 1\end{pmatrix},\ \left(p - Ft, q - \frac{F}{2m}t^2\right),\\ &\theta + \frac{1}{2}qFt - p\frac{Ft^2}{4m} + \frac{F^2t^3}{12m}\Big\}.\end{aligned} \qquad (4.3)$$

This group gives for the evolution variables the same composition law as for the free particle, except for the U(1) part for whose exponent θ we now obtain

$$\theta'' = \theta' + \theta + \frac{1}{2}\left[(pq' - p'q) + \frac{pp'}{m}t\right] - F\left(q't + \frac{1}{2m}p't^2\right). \quad (4.4)$$

The left invariant vector fields are given by

$$\begin{aligned} X_t^L &= \frac{\partial}{\partial t} + \frac{p}{m}\frac{\partial}{\partial q} - (Fq)\Xi, & [X_v^L, X_t^L] &= X_q^L, \\ X_q^L &= \frac{\partial}{\partial q} - \frac{1}{2}p\Xi, & [X_q^L, X_v^L] &= mX_\zeta^L, \\ X_v^L &= m\frac{\partial}{\partial p} + \frac{m}{2}q\Xi, & [X_t^L, X_q^L] &= FX_\zeta^L, \\ X_\zeta^L &= i\zeta\frac{\partial}{\partial \zeta} \equiv \Xi, & [\text{all others}] &= 0. \end{aligned} \quad (4.5)$$

It is clear that, when $F \to 0$, (4.5) reproduces (3.3). The fact that two parameters (m and F) label the extension ($\widetilde{G}_{(m,F)}$) of the Galilei group G is due to its one-dimensional character; this group has a *two*-dimensional space of extensions (the space of the extensions of the ordinary 10-dimensional G is labeled by the mass only). The fact that G is the common starting point for the free particle and the free fall case merely reflects that the invariance group of the *one-dimensional* equation $F = m\, d^2q/dt^2$ is the same for $F = 0$ and for $F = \text{const} \neq 0$.

To define polarization, we now observe that $[X_t^L, X_q^L] \neq 0$ contrarily to what happened in the free case. We now take the invariant subalgebra generated by X_q^L and

$$X_C^L = X_t^L + \frac{F}{m}X_v^L = \frac{\partial}{\partial t} + \frac{p}{m}\frac{\partial}{\partial q} + F\frac{\partial}{\partial p} - \frac{F}{2}q\Xi; \quad (4.6)$$

note that the basic commutator $[X_q, X_v] = m\Xi$ is not altered by the above redefinition, but that now

$$[X_C^L, X_q^L] = 0 \quad (4.7)$$

as in the free particle case.

To obtain the Schrödinger equation, we now require that the functions on $\widetilde{G}$ satisfy the polarization conditions

$$X_C^L \cdot \psi(q,p,t,\zeta) = 0, \quad X_q^L \cdot \psi = 0 \quad (4.8)$$

and the equivariance condition $\Xi \cdot \psi = i\psi$. This condition and $X_C^L \cdot \psi = 0$ give for ψ the form

$$\psi = \zeta\varphi(p,t)e^{iqp/2} \quad (4.9)$$

and then $X_C \cdot \psi = 0$ yields

$$\left(i\frac{\partial}{\partial t} - \frac{p^2}{2m} + Fi\frac{\partial}{\partial p}\right)\varphi(p,t) = 0, \quad (4.10)$$

i.e., the Schrödinger equation for the linear potential in the momentum representation. Again configuration space expressions are gained through the Fourier transform (3.7).

To fully characterize the process leading to the determination of the natural polarization, it is now important to ascertain the meaning of X_C^L, the polarization which determines the temporal evolution of the system and, accordingly, the quantum wave equation once all the other conditions have been taken into account. To do this, it is necessary to evaluate again the canonical 1-form on $\widetilde{G}$ and, more specifically, its "vertical" component Θ^L. The result is

$$\Theta^L = \tfrac{1}{2}(p\,dq - q\,dp) - (p^2/2m - Fq)dt + d\zeta/i\zeta. \quad (4.11)$$

Let us now derive the equations of the characteristic module generated by the vector fields such that

$$i_X\Theta = 0, \quad i_X\,d\Theta = 0, \quad (4.12a)$$

where X is the vector field of generic components

$$X = X^t\frac{\partial}{\partial t} + X^q\frac{\partial}{\partial q} + X^v\frac{\partial}{\partial p} + X^\zeta i\zeta\frac{\partial}{\partial \zeta}. \quad (4.12b)$$

Equation (4.12a) gives

$$X^t = 1, \quad X^q = p/m, \quad X^v = F, \quad X^\zeta = -\tfrac{1}{2}Fq, \quad (4.13)$$

i.e., the characteristic vector field is X_C^L, the part of the polarization (Sec. 1) which generates the quantum equations of motion. This was to be expected; the integral curves of X_C^L give also the *classical* equations of motion,

$$p = Ft + p_0, \quad q = \tfrac{1}{2}(F/m)t^2 + (p_0/m)t + q_0, \quad (4.14a)$$

plus

$$\zeta = z\exp\{-\tfrac{1}{2}iF[\tfrac{1}{6}(F/m)t^3 + (p_0/2m)t^2 + q_0t]\}. \quad (4.14b)$$

(4.14a) shows that one could have started from the quantum group $\widetilde{G}_{(m,F)}$ without physically identifying its parameters since the equations of motion provide directly the adequate correspondence with the evolution space variables.

It is interesting to remark at this stage that one can establish the connection with the usual quantization formalism by defining a contact 1-form on the manifold of solutions of the classical problem as parametrized by the constants of the motion (initial position q_0 and momentum p_0). Indeed one may check that on such manifold ($\widetilde{G}/\mathscr{C}_\Theta$, where $\mathscr{C}_\Theta$ is the characteristic manifold), Θ is written

$$\Theta = \tfrac{1}{2}(p_0\,dq_0 - q_0\,dp_0) + dz/iz. \quad (4.15)$$

We have not pursued this latter path so as to emphasize how the procedure outlined in Sec. 1[1] allows us to perform the quantization directly on $\widetilde{G}$.

To conclude this section, we mention that the basic quantum operators may be obtained from the right invariant vector fields: from

$$X_q^R = \frac{\partial}{\partial q} + \left(\frac{p}{2} - Ft\right)\Xi \quad (4.16a)$$

and

$$X_v^R = t\frac{\partial}{\partial q} + m\frac{\partial}{\partial p} + \left(-\frac{m}{2}q + \frac{p}{m}t + \frac{F}{2}t^2\right)\Xi \quad (4.16b)$$

it may be easily derived that, on $\varphi(p,t)$, $\hat{p}$ and $\hat{q}$ are represented by p and $i\partial/\partial p$ through imposing that the eigenvalues of $\hat{P} \equiv -iX_q^R$ and $\hat{K} \equiv (i/m)X_v^R$ be the corresponding constants of motion.

5. THE HARMONIC OSCILLATOR

For the case of the harmonic oscillator, the quantum dynamical group is the subgroup of WSp(2,$\mathbb{R}$) obtained from $\exp(-iH_{h_0}t)\cdot\exp[i(p\mathbb{Q} + q\mathbb{P} + \theta\mathbb{I})]$. Since the quantum

Hamiltonian is given by

$$H_{h_0} = \mathbb{P}^2/2m + \tfrac{1}{2}m\omega^2\mathbb{Q}^2, \tag{5.1}$$

the result of evaluating $\exp(-iH_{h_0}t)\exp[i(p\mathbb{Q}+q\mathbb{P}+\theta\mathbb{I})]$ is

$$\tilde{G}_{(m,\omega)} = \tilde{g}\left\{\begin{pmatrix} \cos\omega t & \frac{1}{m\omega}\sin\omega t \\ -m\omega\sin\omega t & \cos\omega t \end{pmatrix}, (p,q), \theta\right\} \tag{5.2}$$

[see Eqs. (2.3) with $\alpha = -t/2m, \gamma = -m\omega^2 t/2, s = \pm\omega t/2$], which characterizes the group we take as our starting point.

The composition law (2.4) induces the following one for the evolution space variables and θ:

$$\begin{aligned} t'' &= t' + t, \\ p'' &= p + p'\cos\omega t - m\omega q'\sin\omega t, \\ q'' &= q + (1/m\omega)p'\sin\omega t + q'\cos\omega t, \\ \theta'' &= \theta' + \theta' + \tfrac{1}{2}[(pq' - qp')\cos\omega t \\ &\quad + (pp'/m\omega - m\omega qq')\sin\omega t], \end{aligned} \tag{5.3}$$

which in the free limit $\omega \to 0$ gives (3.2) as it should. Clearly, we could continue and define left invariant vector fields on $\tilde{G}_{(m,\omega)}$. Nevertheless, it is already evident that the basis (5.3) of the evolution space is not the adequate one; the matrix which determines the evolution operator in (5.2) leaves neither the p nor the q subspaces invariant, indicating that a polarization is not immediately obtained. The matrix of the time evolution subgroup, however, is diagonalizable in this case, and the diagonalizing matrix is

$$\mathbf{B} = \frac{1}{\sqrt{2}}\begin{pmatrix} \sqrt{m\omega} & -i/\sqrt{m\omega} \\ -i\sqrt{m\omega} & 1/\sqrt{m\omega} \end{pmatrix}. \tag{5.4}$$

Again, this is an automorphism external to $\tilde{G}_{(m,\omega)}$ but internal to HSp(2,$\mathbb{C}$). Once $\mathbf{B}$ has been applied to (5.2), the elements of the transformed $\tilde{G}_{(m,\omega)}$ are written

$$g\left\{\begin{pmatrix} e^{i\omega t} & 0 \\ 0 & e^{-i\omega t} \end{pmatrix}, \sqrt{m/\omega}(iC^+, C), \theta\right\} = \mathbf{B}\tilde{g}\mathbf{B}^{-1}, \tag{5.5}$$

where $C = (1/m\sqrt{2})(m\omega q + ip)$ and the composition law is given by

$$\begin{aligned} C''^+ &= C'^+e^{i\omega t} + C^+, \\ C'' &= C'e^{-i\omega t} + C, \\ \theta'' &= \theta' + \theta + \tfrac{1}{2}[iC'C^+e^{-i\omega t} - iC'^+Ce^{i\omega t}], \\ t'' &= t' + t. \end{aligned} \tag{5.6}$$

It is now clear that X_t^L and X_C^L are appropriate to define the polarization, and since

$$\begin{aligned} X_t^L &= \frac{\partial}{\partial t} - i\omega C\frac{\partial}{\partial C} + i\omega C^+\frac{\partial}{\partial C^+}, \\ X_C^L &= \frac{\partial}{\partial C} - \frac{i}{2}C^+\Xi, \end{aligned} \tag{5.7}$$

the conditions $X_C^L\cdot\psi = 0$ and $X_t^L\cdot\psi = 0$ give

$$\psi = \zeta\varphi(C^+,t)e^{-C^+C/2} \tag{5.8a}$$

and

$$\frac{\partial\varphi}{\partial t} = -i\omega C^+\frac{\partial\varphi}{\partial t}, \tag{5.8b}$$

respectively. Again, X_t^L generates the characteristic module for this case; this may be seen by evaluating Θ from (5.7) [plus $X_{C^+}^L$ and X_ζ^L, cf. (4.5) and (4.11)], which turns out to be

$$\Theta = \tfrac{1}{2}i(C^+dC - C\,dC^+) - \omega CC^+dt + d\zeta/i\zeta. \tag{5.9}$$

The inner product for the ψ's now induces for the functions φ the following one:

$$\left(\frac{2}{\pi}\right)^{1/2}\int_{\mathbf{R}^2} d\,\mathrm{Re}\,C^+\,d\,\mathrm{Im}\,C^+\,\varphi_2^*(C^+,t)\varphi_1(C^+,t)e^{-|C|^2} \tag{5.10}$$

as obtained from (2.7) with $v = 0$, $\nu = 1$. This is the Bargmann scalar product,[14] as one should expect. The natural polarization leads thus to the BFS picture, and it may be checked that the evolution operator preserves the polarization X_C^L, i.e.,

$$[X_t^L, X_C^L] = i\omega X_C^L. \tag{5.11}$$

It should be noted that Eq. (5.8b) is not quite the Schrödinger equation for the harmonic oscillator. It corresponds to the Bohr–Wilson–Sommerfeld quantization, and the additive constant $\frac{1}{2}$ corresponding to the ground state energy is missing. In the usual approach to geometric quantization, this is remedied by looking to the transformation properties of the "half-forms" under the time evolution generator.[4] The result is that Eq. (5.8b) is corrected to

$$\frac{\partial\varphi}{\partial t} = -i\omega\left(C^+\frac{\partial}{\partial C} + \frac{1}{2}\right)\varphi, \tag{5.12}$$

where the functions $\varphi(C^+,t)$ belong to the Bargmann–Segal space $\mathscr{B}_{\mathbf{B}}$, with $\mathbf{B}$ given by (5.4). The solutions to (5.12) which separate into the product of C^+ times a function of t are, up to a multiplicative constant a_ν,

$$a_\nu(C^+)^\nu e^{-i(\nu+1/2)t} \tag{5.13}$$

for $\nu\in\mathbb{C}$. The condition that the solutions of (5.12) belong to $\mathscr{B}_{\mathbf{B}}$ (entire analytic functions with mild decrease conditions[14]) restricts the range of ν to the nonnegative integers, $\nu\in\{0,1,2,\cdots\}$. The normalization factor in (5.13) is found to be $a_n = (2\pi)^{-1/4}\cdot(n!)^{-1/2}$.[18] The set of eigenstates of the system is thus

$$u_n(C^+,t) = (2\pi)^{-1/4}(n!)^{-1/2}(C^+)^n e^{-i(n+1/2)t}, \quad n = 0,1,2,\cdots. \tag{5.14}$$

in BSF space. The eigenstates in the configuration space may be recovered through an inverse Bargmann transform [cf. (2.5)] on (5.14) yielding the familiar harmonic oscillator wave functions

$$\psi_n(q,t) = \int d^2\mu_{\mathbf{B}}(C^+,C)u_n(C^+,t)C_{\mathbf{B}}(C^+,q)^* \tag{5.15a}$$

$$= 2^{-n/2}\pi^{-1/4}(n!)^{-1/2}\exp(-\tfrac{1}{2}m\omega q^2) \times H_n(\sqrt{m\omega}q)e^{-i(n+1/2)t}. \tag{5.15b}$$

In essence, the Bargmann transform (for $m = \omega = 1$) is a group automorphism which "rotates" the usual harmonic oscillator Hamiltonian—the algebra generator $\frac{1}{4}(\mathbb{P}^2 + \mathbb{Q}^2)$—which appears in (5.1) onto the change of scale generator $\frac{1}{2}\{\mathbb{P},\mathbb{Q}\}_+$. As already mentioned, these two algebra elements belong to different orbits in sp(2,$\mathbb{R}$) under the adjoint action of Sp(2,$\mathbb{R}$), but may be brought into similarity through $g_{\mathbf{B}}$, a transformation in HSp(2,$\mathbb{C}$), obtained from (2.5) with $\mathbf{B}$ given by (5.4). This transformation corresponds to a rotation by $i\pi/4$ around the $\frac{1}{4}(\mathbb{P}^2 - \mathbb{Q}^2)$ axis (see Ref. 10, Sec. 9.2). The eigenfunctions of the change-of-scale generator are the (complex) power functions, i.e., essentially (5.13).

6. THE REPULSIVE "OSCILLATOR"

We complete our study of the one-dimensional potentials by considering the somewhat unphysical potential $-\frac{1}{2}m\omega^2\mathbb{Q}^2$, the repulsive "oscillator." For it we find

$$\exp\{-it[(1/2m)\mathbb{P}^2 - \tfrac{1}{2}m\omega^2\mathbb{Q}^2]\}$$
$$\times\exp[i(p\mathbb{Q} + q\mathbb{P} + \theta\mathbb{I})]$$
$$= g\left\{\begin{pmatrix} \cosh\omega t & \sinh\omega t/m\omega \\ m\omega\sinh\omega t & \cosh\omega t\end{pmatrix}, (p,q), \theta\right\} \tag{6.1}$$

and

$$\begin{aligned} t'' &= t' + t, \\ q'' &= q + q'\cosh\omega t + \frac{p'}{m\omega}\sinh\omega t, \\ p'' &= p + p'\cosh\omega t + m\omega q'\sinh\omega t, \\ \theta'' &= \theta' + \theta + \tfrac{1}{2}[(pq' - p'q)\cosh\omega t \\ &\quad + (pp'/m\omega - qq'm\omega)\sinh\omega t] \end{aligned} \tag{6.2}$$

for the group law.

To define the polarization adequate for this problem, we may follow the same pattern as for the harmonic oscillator. The time evolution operator $\exp(-iH_{r_0}t)$ is not diagonal, but diagonalizable by the Sp(2,$\mathbb{R}$) transformation

$$\mathbf{R} = \frac{1}{\sqrt{2}}\begin{pmatrix} \sqrt{m\omega} & 1/\sqrt{m\omega} \\ -\sqrt{m\omega} & 1/\sqrt{m\omega}\end{pmatrix}, \tag{6.3a}$$

which defines the new group variables

$$\alpha = \frac{1}{\sqrt{m\omega}}\frac{m\omega q + p}{\sqrt{2}}, \quad \nu = \frac{1}{\sqrt{m\omega}}\frac{m\omega q - p}{\sqrt{2}}, \tag{6.3b}$$

in terms of which the Hamiltonian is written $H = -\omega\alpha\nu$. In terms of these variables (6.1) reads

$$g\left\{\begin{pmatrix} e^{\omega t} & 0 \\ 0 & e^{-\omega t}\end{pmatrix}, (\alpha,\nu), \theta\right\}, \tag{6.4a}$$

where the time evolution part has been diagonalized, and (6.2) is given by

$$\begin{aligned} t'' &= t' + t, \\ \alpha'' &= \alpha + \alpha' e^{\omega t}, \\ \nu'' &= \nu + \nu' e^{-\omega t}, \\ \theta'' &= \theta + \theta' + \tfrac{1}{2}[\alpha\nu' e^{-\omega t} - \alpha'\nu e^{\omega t}]. \end{aligned} \tag{6.4b}$$

We now may proceed to evaluate the left-invariant vector fields, which are

$$\begin{aligned} X_t^L &= \frac{\partial}{\partial t} + \alpha\frac{\partial}{\partial\alpha} - \nu\frac{\partial}{\partial\nu}, \\ X_\alpha^L &= \frac{\partial}{\partial\alpha} + \frac{\nu}{2}\Xi, \\ X_\nu^L &= \frac{\partial}{\partial\nu} - \frac{\alpha}{2}\Xi, \\ X_\zeta^L &= \Xi \end{aligned} \tag{6.5}$$

and the vertical component of the canonical 1-form,

$$\Theta = \tfrac{1}{2}(\alpha\, d\nu - \nu\, d\alpha) + \omega\alpha\nu\, dt + d\zeta/i\zeta. \tag{6.6}$$

It is now easily seen that the vector fields defining a polarization are given by X_t^L and X_α^L and that, as it should, X_t^L is the characteristic vector field of Θ. The conditions $\Xi\cdot\psi = i\psi$ and $X_\alpha^L\cdot\psi = 0$ give

$$\psi = \zeta\varphi(\nu,t)e^{i\nu\alpha/2} \tag{6.7}$$

and $X_t^L\cdot\psi = 0$ finally yields

$$\frac{\partial\varphi}{\partial t} - \omega\nu\frac{\partial\varphi}{\partial\nu} = 0. \tag{6.8}$$

By a reasoning analogous to that leading to (5.12), the true Schrödinger equation is written

$$\frac{\partial\varphi}{\partial t} = \omega\left(\nu\frac{\partial}{\partial\nu} + \frac{1}{2}\right)\varphi. \tag{6.9}$$

In fact, the presence of the $\frac{1}{2}$ term can also be justified by unitarity considerations: The transformation $\mathbf{R}$ [(6.3a)] maps $\mathscr{L}^2(\mathbb{R})$ unitarily on $\mathscr{L}^2(\mathbb{R})$ in ν. The separated solutions of (6.9) are, again up to a constant a_μ,

$$a_\mu\nu^\mu e^{(\mu+1/2)t} \tag{6.10}$$

for μ complex. Although these functions do not belong to $\mathscr{L}^2(\mathbb{R})$, they constitute a generalized basis for $\mathscr{L}^2(\mathbb{R})$ eigenbasis for $\frac{1}{2}\{\mathbb{Q},\mathbb{P}\}_+$. This is the bilateral Mellin transform basis (Ref. 10, Sec. 8.2),

$$(2\pi)^{-1/2}\nu_\pm^{-1/2+i\lambda}, \quad \lambda\in\mathbb{R}, \tag{6.11a}$$

$$\nu_+ = \begin{cases}\nu & \nu\geqslant 0, \\ 0, & \nu<0,\end{cases} \quad \nu_- = \begin{cases} 0, & \nu\geqslant 0, \\ -\nu, & \nu<0,\end{cases} \tag{6.11b}$$

i.e., (6.10) with $a_\mu = (2\pi)^{-1/2}$ and $\mu = -\frac{1}{2} + i\lambda$. The Mellin basis is Dirac-orthonormal and complete, and its eigenvalues under $\frac{1}{2}\{\mathbb{Q},\mathbb{P}\}_+$ cover twice the real line. Hence,

$$\omega_{\sigma,\lambda}(\nu,t) = (2\pi)^{-1/2}\nu^{-1/2+i\lambda}e^{i\lambda t},$$
$$\sigma = \pm 1, \lambda\in\mathbb{R}, \tag{6.12}$$

is the $\mathscr{L}^2$-complete Dirac-orthonormal set of solutions of (6.9). [The importance of the $\frac{1}{2}$ additive term in the Schrödinger equation—which merely shifts the whole spectrum for the case of the harmonic oscillator—is seen in this example since, otherwise, the time part in (6.10) would adopt the unsuitable form $e^{(-1/2+i\lambda)t}$ instead of $e^{i\lambda t}$.]

In order to recover the wave functions in configuration space, we apply the inverse of the transform $\mathbf{R}$ of (6.3a) (for $m = \omega = 1$). This is a rotation by $\pi/4$ around the $\frac{1}{4}(\mathbb{P}^2 + \mathbb{Q}^2)$ axis—in fact, the square root of the Fourier transform—which through its action on the algebra, brings the repulsive oscillator $\frac{1}{2}(\mathbb{P}^2 - \mathbb{Q}^2)$ onto the change of scale operator

$\frac{1}{2}\{\mathbb{Q},\mathbb{P}\}_+$ appearing on the right-hand side of (6.9).

The configuration space eigenfunctions are thus the repulsive oscillator wavefunctions

$$\Upsilon_{\sigma,\lambda}(q,t) = \int_{-\infty}^{\infty} d\nu\, \omega_{\sigma,\lambda}(\nu,t) C_{\mathbf{M}}(\nu,q)^* = \exp[i\pi(\tfrac{1}{2} - i\lambda)] 2^{-3/4} \pi^{-1} \Gamma(\tfrac{1}{2} - i\lambda) \times D_{-1/2+i\lambda}(-\sigma 2^{1/2} e^{3i\pi/4} q) e^{i\lambda t}, \qquad (6.13)$$

where $D_\tau(x)$ is the parabolic cylinder function (see Ref. 19; Chap. 19 and Ref. 10, Secs. 7.5 and 8.2).

7. CONCLUSIONS

The above study of all inequivalent one-dimensional quantum systems with up-to-quandratic potentials, which admit a finite-dimensional group,[20] has shown how a general definition for a quantum manifold (in the sense of the geometric quantization) can be given based on the group. It has also been shown that for each case there exists in our formalism a natural polarization and that the role of the Blattner–Konstant polarization changing transformation (see, e.g., Ref. 4) is played by integral transforms of the group WSp(2,$\mathbb{R}$) or of the semigroup HS(2,$\mathbb{C}$), which include the four one-dimensional systems considered in this paper.

ACKNOWLEDGMENTS

This paper has been supported in part by the Commión Asesora de Investigación Científica y Técnica under Contract 4604–79.

[1]V. Aldaya and J. A. de Azcárraga, J. Math. Phys. **23**, 1297 (1982).

[2]This method assumes the fruitful belief that symmetry—and, in general, geometric—considerations are very important in constraining a theory, almost to the point of determining it completely. Nevertheless, and as pointed out by Yang,[3] the fact that the geometric approach to physical problems has been up to now a very sucessful one does not imply that every geometrical concept has its physical counterpart; the history of physics is littered with many useless geometries.

[3]C. N. Yang, talk given at the Chern Symposium, Berkeley, Calif., June 1979 and in *To Fulfill a Vision*, Proc. of the 1979 Jerusalem Einstein Centennial Symposium, edited by Y. Ne'eman (Addison-Wesley, Reading, MA, 1981).

[4]J. M. Souriau, *Structure des systèmes dynamiques* (Dunod, Paris, 1970); B. Kostant, *Quantization and Unitary Representations*, Lecture Notes in Mathematics **170** (Springer-Verlag, New York, 1970), pp. 87–208; see also: D. J.Simms and N. M. J. Woodhouse, *Lecture Notes in Geometric Quantization* (Springer-Verlag, Berlin, 1976); J. Śniatycki, *Geometric Quantization and Quantum Mechanics* (Springer-Verlag, New York, 1980).

[5]The dynamical algebra of a given classical (quantum) system is a Lie algebra under the Poisson (commutator) bracket such that the system Hamiltonian is an element of the algebra *not* belonging to the center. For quantum systems with a discrete energy spectrum this means that raising and lowering operators are to be found in the dynamical algebra. The dynamical algbra is in general not unique (in the harmonic oscillaor system, for instance, the Hamiltonian, creation, annihilation, and unit operators constitute a minimal four-dimensional algebra). The conditions that the dynamical algebra be of finite dimension and that it contain a subalgebra contractible to the Galilei algebra determine that the appropriate algebra for the four systems studied here be wsp(2,$\mathbb{R}$). The dynamical group is the exponentiation of the dynamical algebra.

[6]The characteristic module $\mathscr{C}_\Theta$ of a form Θ is the space of vector fields X such that $i_X \Theta = 0$ and $i_X\, d\Theta = 0$, where i_X indicates inner product.

[7]See, e.g., K. B. Wolf, "The Heisenberg-Weyl Ring in Quantum Mechanics," in *Group Theoretical Methods and Its Applications*, edited by E. M. Loebl (Academic, New York, 1975), Vol. III.

[8]A. Weil, Acta Math. **11**, 143 (1963).

[9]K. B. Wolf, SIAM J. Appl. Math. **40**, 419 (1981).

[10]K. B. Wolf, *Integral Transformations in Science and Engineering* (Plenum, New York, 1979), Sec. 10.2.

[11]A. Joseph, J. Math. Phys. **13**, 351 (1972).

[12]Ref. 10, Sec. 9.1.

[13]M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772, 1780 (1971); M. Moshinsky, SIAM J. Appl. Math. **25**, 193 (1973).

[14]V. Bargmann, Commun. Pure Appl. Math. 14, 187 (1961); **20**, 1 (1967).

[15]V. Bargmann, Ann. Math. **59**, 1 (1954).

[16]Condition (3.5), which constitutes the expression that ψ is a U(1) (equivariant) function may require some explanation. Let $H \xrightarrow{\tau} S$ be a vector bundle of base S and fiber F associated with a principal bundle $P(\mathfrak{G},X)$. Then, the modulus $\Gamma(H)$ of differentiable cross sections of $H \xrightarrow{\tau} S$ is isomorphic to the vector space of differentiable F-valued $\mathfrak{G}$ functions on P. In our case P is, as a manifold, $\widetilde{G}_{(m)}$; the structure group $\mathfrak{G}$ is U(1) and S is the Galilei group G. Thus, the space of the wave functions (cross sections of $H \xrightarrow{\tau} S$, and thus complex-valued functions of p, q, and t) may be defined as the space of $\mathbb{C}$-valued U(1) functions on $\widetilde{G}_{(m)}$.

[17]Note that the invariance condition for polarization is here oversimplified because of the monodimensional nature of the problem. When $\widetilde{G}_{(m)}$ is the eleven-dimensional extended Galilei group, X_t^L and X_q^L generate an invariant abelian subgroup of G including $\mathscr{C}_\Theta$.

[18]The normalization constants employed here are those of Ref. 10 and differ from those originally defined by Bargmann.[14] We follow that choice here since it assures us that the composition of transforms and the $\mathbf{M} \to 1$ limit are obtained without extra compensating factors.

[19]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).

[20]Only in dimension one is it possible to establish a one-to-one correspondence between the dynamical system and the dynamical group.

Orthogonality and orthocomplementations in the axiomatic approach to quantum mechanics: Remarks about some critiques

Gianpiero Cattaneo and Giuseppe Nistico'
Department of Mathematics, University of Calabria, Cosenza, Italy

(Received 24 March 1982; accepted for publication 29 June 1983)

The logic approach to axiomatic quantum mechanics via orthocomplemented partial ordered sets of yes–no measurements, which constitute the observing part of a concretely realizable experiment on microworld, has been criticized from the empirical point of view by Mielnik, which on the contrary privileges the convex scheme linked to the preparing part. In this work we do assume that a description of quantum phenomenology must take into account both these two parts in which every elementary experiment can be decomposed. According to this predecision, we develop an axiomatic approach based on indistinguishability principles of a quantum information system. The very general concept of yes–no measurement or "question" is accepted, and then the set of all questions is classified according to the behavior with respect to a phenomenological orthogonality relation. In particular, we single out the set F of fuzzy events or effects and the set $\mathbf{E} \subseteq \mathbf{F}$ of exact events. The Mielnik critique is then refused since it regards the order structure of E using counterexamples which pertain to F/E. The notions of physical property and noperty are then introduced and an axiomatic foundation of quantum mechanics based on a pre-Hilbert space is discussed.

PACS numbers: 03.65.Bz

1. INTRODUCTION TO AND CRITIQUE OF THE AXIOMATIC APPROACH

In this section we introduce the axiomatic approach to quantum mechanics taking into account the Mielnik critique.[1] It is well known that the scheme of states and observables with the Hilbert space at the bottom is the final result of a long historical development and "provides a good structural description of the existing theories. However, it seems to exclude the possibility of generalizations... . Moreover, (this scheme) is so compact that it is difficult to see in which point *it could be relaxed without denying something very fundamental.* This is sometimes taken as an argument against the possibility of further generalizations of the present day quantum scheme. However, the conclusion from (these results) might be just the opposite.

After all, the most of the essential progress in physics has been achieved by denying something apparently obvious. ...There is no reason to think that this process is ended." [1]

In the quoted work, Mielnik makes a critique to the traditional axiomatic approach to quantum mechanics observing, "According to a general accepted philosophy the 'quantum logic' is the set of all *questions* which may be put to micro-object. ...It is a specific status of quantum axiomatics that it should reflect phenomenology. In order to verify the phenomenological background of quantum logical axioms *a careful identification must be made in order to specify the elements of physical reality which correspond to the abstract 'questions.'* At this point the axiomatic theory is elusively elegant:

Definition 1: A *question* is an arbitrary macroscopic arrangement which, when interacting with a micro-object, may or may not produce a certain definite macroscopic (alternative) effect; the presence of the effect is conventionally taken as the answer 'yes' whereas its absence is 'no.' " [1]

An analysis of the methodologies the physical world is investigated with leads to the conclusion that the concretely realizable experiments may be considered as decomposed of two parts, each consisting of macroscopic apparatuses: the *preparing part* and the *observing part.*

According to Ludwig, "we have ... to return to experimental situations which everybody, physicist or layman, might examine in a laboratory as objectively given events.

Now, what kind of experimental situation should be selected as a starting point?...We find a *preparing part* which, via a *microscopic channel*, can act on a *signal part* (*observing part*), a part, that is, where the effects are produced. In the course of an experiment, the signal part will either respond or not. The preparing part, for example, might be an accelerator with the target placed in the beam; and the signal part might be a counter. ...

The preparing part shall be an apparatus, objectively given and technically describable and the same is required of the signal part together with its response or lack of response.

The term *microscopical channel* is only to indicate the *action* of the preparing part on the signal part, possibly producing a *response signal.*" [2]

The simpler interpretative scheme of these experiments is the one in which *ensembles* of physical objects are prepared by a preparation macroscopic apparatus and then are tested by an observation macroscopic arrangement which produces a certain macroscopic alternative direct confirmable effect.

The *measurement* "shall be understood as the action of a *single* object upon a measuring-device, so that a direct, objectively traceable effect occurs or does not occur (e.g., counter signal, a cloud-chamber-track, blackening of a photographic plate, etc.)." [2]

"Since different tests correspond in general to different experimental setups, it will in general be impossible to perform more than one test at once on the (object). And because, in general, the performance of a test changes the (object) and even sometimes destroys it, it will in general not be fruitful and sometimes impossible to perform more than one test on one (object)." [3]

Quoting Mielnik, "though the *question* may be put to any *single* micro-object, the answer becomes conclusive only if obtained for a great number of its *independent* replications. This leads to an abstract scheme where *questions* idealize the macroscopic devices used to test statistical ensembles of microsystems." [1]

For making more precise the Mielnik critique we now introduce some physical considerations which will lead to the fundamental *state–question structure*. From now on, for *elementary experiment* we mean any experiment consisting of a *preparing part* and a *signal* or *observing part*, and it will be denoted by the couple (q, x), where x represents the ensemble of micro-objects prepared according to a well-defined preparing procedure while q is the alternative effect or question x is investigated with, according to a well-defined observing procedure.

Experiments of this type present statistical regularities such that the quantitative analysis of the obtained results leads to a probability $P(q, x) \in [0,1]$ that the question q occurs once preparing the system according to x. To be precise, if x is a concrete method of preparation, q a concrete apparatus, let N objects be produced by x. During the action of those objects on q, a certain effect, denoted by q too, may occur one time or may not another time. If N_+ is the number of objects yielding the effect, the ratio N_+/N shall be called *frequency*.

"The frequency N_+/N proves, for large N, to be almost independent of N provided we consider experiments which are really reproducible. Hence it is reasonable to introduce a function $P(q, x)$ with $0 \leqslant P(q, x) \leqslant 1$, which, when comparing with experiment, is to be taken as $P(q, x) \sim N_+/N$ in 'physical approximation.' " [2]

Therefore, the analysis of a certain phenomenological world is made up of a set of elementary experiments and any scientific theory which describes the corresponding phenomena must involve a triple $(\mathbf{Q},\mathbf{S},P)$, called *question–state structure*, where:

(1) $\mathbf{S}$ is a nonempty set whose elements are said to be *preparation procedures of ensembles*, or *ensembles* for short. Sometimes, with a certain misuse, we shall also say that the ensemble is prepared in the *state* $x \in \mathbf{S}$.

(2) $\mathbf{Q}$ is a nonempty set whose elements are the *observation procedures of effects*, or *questions* for short.

(3) P: $\mathbf{Q} \times \mathbf{S} \to [0,1]$ is a function, called the *probability function*. The value $P(q, x)$ represents the *probability of occurrence* of the question q relative to the ensemble x, or to the ensemble prepared according to x.

Once a question $q \in \mathbf{Q}$ is fixed we can introduce the following subsets of $\mathbf{S}$:

(a) The *certainly yes* domain of q:

$\mathbf{S}_1(q) := \{x \in \mathbf{S}: P(q, x) = 1\}$.

(b) The *certainly no* domain of q:

$\mathbf{S}_0(q) := \{x \in \mathbf{S}: P(q, x) = 0\}$.

If $x \in \mathbf{S}_1(q)$ (resp. $x \in \mathbf{S}_0(q)$), then the question q is said to occur (resp. nonoccur) with certainty. In other words "when the physical system has been prepared in such a way (i.e., $x \in \mathbf{S}$) that the physicist may affirm that in the event of a measurement (of the question q) the result 'yes' is certain (i.e., $x \in \mathbf{S}_1(q)$), we shall say that the question is 'true.' If the outcome for the question q is not certain (i.e., $x \in \mathbf{S}/\mathbf{S}_1(q)$), the statement '$q$ is true' is false, but we do not say 'q is false.' " [4]

To be precise, we shall say that the question q is "false" iff the physical system has been prepared in such a way that once q is measured the result "no" is certain (i.e., $x \in \mathbf{S}_0(q)$).

Quoting Pool, "the phenomenological interpretation of the mathematical system, question–state structure, may be specified by selecting a collection of rules for the interpretation of the primitive entities: questions, states and probability functions.

The following collection is a possible (but obviously not the only) choice of these rules.

A question–state structure $(\mathbf{Q},\mathbf{S},P)$ is associated with the class of physical systems of a specific kind.

(i) A *state* may be identified with a 'state-preparation procedure,' that is, instruction for an apparatus which produces sample physical systems of the specific kind.

(ii) A *question* may be identified with the 'occurrence or nonoccurrence' of a particular phenomenon pertaining to physical systems of the specified kind.

More specifically, a question may be identified with an 'observation procedure,' that is, instructions for an apparatus which interacts with a sample physical system and indicates either yes or no corresponding to the occurrence or nonoccurrence of the phenomenon.

(iii) The interpretation of $P(q, x)$ for $x \in \mathbf{S}$ and $q \in \mathbf{Q}$ would then be the following:

(1) Prepare an ensemble of sample physical systems utilizing a state preparation procedure corresponding to x.

(2) Determine the occurrence or nonoccurrence of the question q utilizing an observation procedure for q with each sample of this ensemble.

(3) If the ensemble is sufficiently large, then the frequence of occurrence of q should be closed to $P(q, x)$." [5]

Of course, in studying a specific kind of physical system some additional principles and axioms must be introduced in order to describe the particular situation under examination. In the quantum logic case, Mielnik says that "now it is argued, the validity of the basic axioms of the quantum logic (apart from weak modularity) is almost a matter of tautology. For instance:

Axiom M.1: *The identity axiom*.

Two 'yes–no measurements' (or 'questions') with the identical 'certainly yes' domain are *obviously* testing for the same feature, and so the difference between them is not essential: this motivates the *identity law*:

(id) $\mathbf{S}_1(p) = \mathbf{S}_1(q)$ implies $p = q$.

Axiom M.2: *The uniqueness of the orthocomplementation.* The existence of an unique orthocomplement q' for an arbitrary 'yes–no' arrangement q is beyond discussion: q' is simply that same measuring arrangement with the role of 'yes' and 'no' interchanged."[1]

However, after an analysis of some concrete experimental situations, Mielnik reaches the following conclusions:

Conclusion 1: In spite of its elegant generality, the idea of a "question" as a quite arbitrary macroscopic arrangement which produces a certain macroscopic alternative effect is wrong.

Conclusion 2: Something would be broken in the assumed structure of **Q**:

(iia) either the identity axiom

(iib) or the uniqueness of the orthocomplement.

In this work we do accept the very general notion of "question" as a macroscopic arrangement which behaves according to Definition 1. On the contrary, once some empirically well-founded assumptions about the structure $(\mathbf{Q},\mathbf{S},P)$ are introduced, we partially agree with Conclusion 2.

In particular, as regards (iia) we shall substitute in Sec. 2 the identity axiom with an indistinguishability principle, according to Ludwig's approach to axiomatic quantum mechanics.[2]

Relative to (iib), it is possible to single out a particular subset **F** of questions for which the orthocomplement uniqueness holds; however, the corresponding orthocomplementation is quite different from the usually considered one. To be precise, the collection of all these questions has a structure $(\mathbf{F},\mathbb{0},\mathbb{1},\leqslant,')$ of a degenerate orthocomplemented partially ordered set (poset, for short), and using Ludwig terminology its elements are called *effects*. A particular subclass **E** of effects, whose elements are the events, is introduced in Sec. 7, and it is shown that this is the peculiar class of questions in which the identity axiom holds and the orthocomplementation is not degenerate.

The *quantum phenomenology* is then reflected by the general effects from **F**, which are the elements of physical reality that correspond to the abstract "questions," whereas the elements of the usual *quantum logic axiomatic* are the elements from **E**. These two "levels" of description of quantum phenomenology must not be confused.

In this context, Axioms M.1 and M.2, introduced by Mielnik in his critical reexamination of quantum logic, pertain to the events structure and not to the more general one involving effects. Therefore, from our point of view, it is not right to verify the phenomenological background of quantum logic axioms, related to events, using the phenomenology pertaining to general effects. In particular, the counterexamples presented by Mielnik against the logic approach are not acceptable for they are examples of general effects which are *not* events. For example, not all the identities which would "collapse" the quantum logic are right since they refer to effects for which the degeneration of orthocomplementation does not imply that $q \wedge q' = \mathbb{0}$ and $q \vee q' = \mathbb{1}$ (see Sec. 3).

Moreover, let us remark that the degeneration of the orthocomplementation of **F** implies that the corresponding orthogonality relation $q \perp p$ iff $q \leqslant p'$ (or, equivalently, $p \leqslant q'$) is degenerate too, and so there could exist elements $q \in \mathbf{F}$ such that $q \perp q$. The semitransparent mirror considered by Mielnik is an example of such a degenerate effect. Hence, this orthogonality is symmetric, but is *not* irreflexive (i.e., $p \perp q$ does not imply $p \neq q$) and so the Foulis–Randall approach to orthogonality,[6–11] where the irreflexivity is an essential condition for the development of the theory, cannot be applied.

In conclusion, the philosophical idea which underlies our work is to consider as a starting point a question–state structure with two indistinguishability principles. No other axiom is required (e.g., Ludwig's sensitivity increase of two effects,[2] sensitivity increase of one effect,[12] Piron's axiom A, axiom C, axiom P on the set of yes–no experiments,[13–16] Mielnik's convexity condition on the set of states,[1] and so on); rather we naturally single out some suitable substructures from the poset of all questions characterized by well-defined properties with respect to the original degenerate orthogonality.

To be precise, the posets of all:

(a) generalized effects $\mathbf{F}_g := \{q \in \mathbf{Q}\colon \exists\, q' = \max\{q\}^{\perp}, \exists\, q'' = \max\{q'\}^{\perp}\}$,

(b) effects or fuzzy events $\mathbf{F} := \{f \in \mathbf{F}_g, \exists\, f' \in \mathbf{Q} \ni P(f,x) + P(f',x) = 1, \forall\, x \in \mathbf{S}\}$,

(c) exact events $\mathbf{E} := \{a \in \mathbf{F}\colon (\mathbf{S}_1(f) = \mathbf{S}_1(a) \Rightarrow a \leqslant f)$ and $(\mathbf{S}_1(g) = \mathbf{S}_0(a) \Rightarrow a' \leqslant g)\}$.

Ludwig's approach refers to (a) [or, perhaps, to (b)], with the crucial assumption of sensibility increase of two effects; the usual quantum logic approach refers to (c) with some additional assumption in order to recover, also if in some weaker formulation, the mathematical Hilbert-space structure of which the quantum theory makes technical use for describing the microworld.

2. ORDERING, PHENOMENOLOGICAL IMPLICATION, AND ORTHOGONALITY IN THE STATE–QUESTION STRUCTURE

In spite of the fact that all the information we can obtain about the physical system is contained in the triple $(\mathbf{Q},\mathbf{S},P)$, the following two principles are physically well founded:

Axiom 2.1: Indistinguishability principle of preparation procedures:

If $P(q,x_1) = P(q,x_2)$ for all $q \in \mathbf{Q}$, then $x_1 = x_2$.

That is, if the previous condition holds the two preparation procedures x_1 and x_2 are called equivalent. They produce the *same* ensemble in the following sense: Both ensembles yield equal probabilities, i.e., information, for all available questions $q \in \mathbf{Q}$ and so we cannot obtain, in any way, any information about the system that allows us to distinguish between them. Consequently, equivalent x will not be distinguished from now on and two distinct preparation procedures x_1 and x_2 must assign different probability distributions at least to a question.

Axiom 2: Indistinguishability principle of observation procedures:

If $P(q_1,x) = P(q_2,x)$ for all $x \in \mathbf{S}$, then $q_1 = q_2$.

In the same way q_1 is called equivalent to q_2. To determine frequencies on the ensembles by use of one apparatus yielding q_1 or another one yielding q_2 makes no difference then. In the following we shall not discriminate q_1 and q_2 and so the set of states is *separating*. In this way, we do agree with the observation that "it seem reasonable to stipulate that a ... condition for logical equivalence of questions is that they have the same probability in every state (i.e., preparation procedure). Indeed, it seems monstrous to claim two questions are logically equivalent when they take different probabilities in every state." [17]

In conclusion, two preparation (respectively, observation) procedures are indistinguishable if they give the same experimental information once all possible elementary experiments are performed inside the question–state structure under examination.

Axiom 3: The existence of the certain and the impossible questions:
The set $\mathbf{Q}$ of all questions contains an element $\mathbb{1}$, called the *certain* question, such that

$$P(\mathbb{1}, x) = 1 \quad \text{for every} \quad x \in \mathbf{S},$$

i.e., the answer is always "yes" (this question consists of the observing procedure which verifies that the system exists) and contains an element $\mathbb{0}$, called the *absurd* or *impossible* question, such that

$$P(\mathbb{0}, x) = 0 \quad \text{for every} \quad x \in \mathbf{S},$$

i.e., the answer "no" is always given.

Let us notice that in the particular cases of these *trivial* questions we have

$$\mathbf{S}_1(\mathbb{0}) = \varnothing \quad \text{and} \quad \mathbf{S}_0(\mathbb{0}) = \mathbf{S}, \tag{2.1a}$$

$$\mathbf{S}_1(\mathbb{1}) = \mathbf{S} \quad \text{and} \quad \mathbf{S}_0(\mathbb{1}) = \varnothing. \tag{2.1b}$$

According to Ludwig, "now a remark is important for the following: we have avoided using the word 'measurement' quite intentionally hitherto. The concept of measurement could produce the idea that using an apparatus (for observing a question) q, something is measured or determined on the object, the question q revealing some property of the object such that if q occurs the object has *this* property. *We shall argue neither for nor against this kind of description.*

Primarily, we find effects (questions) q on certain devices constructed technically..., these effects (questions) saying nothing *immediate* about qualities of an 'object.' Hence the problem of 'object-properties' and their measurement shall be left aside and only be discussed later." [18]

From Axiom 2 it follows that the set of states is *order-determining*[5] in the sense that the binary relation defined on $\mathbf{Q}$ by

(or) $p \leqslant q$ iff $P(p, x) \leqslant P(q, x)$ for all $x \in \mathbf{S}$

is a partial order relation, that is, it satisfies the properties

$$p \leqslant p, \qquad \text{(reflexive)}$$

$$p \leqslant q \text{ and } q \leqslant p \text{ imply } p = q, \qquad \text{(antisymmetric)}$$

$$p \leqslant q \text{ and } q \leqslant r \text{ imply } p \leqslant r. \qquad \text{(transitive)}$$

Relative to this partial order, $(\mathbf{Q}, \mathbb{0}, \mathbb{1}, \leqslant)$ is a poset bounded by the *least* element $\mathbb{0}$ and the *greatest* element $\mathbb{1}$.

The phenomenological interpretation of the order relation may be briefly summarized: $p \leqslant q$ means that the probability of occurrence of the question p is equal or less than the probability of occurrence of the question q, whatever is the ensemble x in which the system is prepared. Of course, the following property holds:

$$p \leqslant q \quad \text{implies} \quad \mathbf{S}_1(p) \subseteq \mathbf{S}_1(q) \quad \text{and} \quad \mathbf{S}_0(q) \subseteq \mathbf{S}_0(p). \tag{2.2}$$

In particular $p \leqslant q$ *implies* that if p occurs with certainty (i.e., is "true"), then q also occurs with certainty (i.e., is "true").

If we have the situation that whenever a question p is "true," then the question q is also "true," we shall denote this as $p \subseteq q$, and we shall say "p is *stronger* than q" or that "q is *less* than p," i.e.,

(pi) $p \subseteq q$ iff $\mathbf{S}_1(p) \subseteq \mathbf{S}_1(q)$.

In general this relation has the following properties:

$$p \subseteq p, \qquad \text{(reflexive)}$$

$$p \subseteq q \text{ and } q \subseteq r, \text{ then } p \subseteq r, \qquad \text{(transitive)}$$

where p, q, r are questions of the system. So the binary relation $\subseteq$ is a *preorder* (in general, not antisymmetric) relation on the set of questions called the phenomenological relation of *implication*.[5]

People working on axiomatic quantum mechanics have been accustomed to think that at least two meaningful binary relations can be given in the set $\mathbf{Q}$ of all questions testable for a given physical system, viz., the one used by Mackey,[19] which is just (or), and the one used by Jauch[20] and Piron[21] expressed here in the equivalent form (pi).[22] In the framework of a question–state structure these two relations do not coincide in $\mathbf{Q}$: we can only state that

$$p \leqslant q \quad \text{implies} \quad p \subseteq q. \tag{2.3}$$

From our point of view, it is physically meaningless to ask whether or to assume that these binary relations coincide on $\mathbf{Q}$: more correctly, the problem is to single out a physically well-defined subset of $\mathbf{Q}$ in which the order relation (or) and the phenomenological implication (pi) coincide.

Notice that from the ordering properties of (or) we also deduce that:

$$\text{If } p \vee q \text{ exists, then } \mathbf{S}_1(p) \cup \mathbf{S}_1(q) \subseteq \mathbf{S}_1(p \vee q)$$
$$\text{and } \mathbf{S}_0(p \vee q) \subseteq \mathbf{S}_0(p) \cap \mathbf{S}_0(q), \tag{2.4a}$$

$$\text{If } p \wedge q \text{ exists, then } \mathbf{S}_1(p \wedge q) \subseteq \mathbf{S}_1(p) \cap \mathbf{S}_1(q)$$
$$\text{and } \mathbf{S}_0(p) \cup \mathbf{S}_0(q) \subseteq \mathbf{S}_0(p \wedge q). \tag{2.4b}$$

The idea of a pair of *mutually exclusive* or *mutually disjoint* questions p and q, written $p \perp q$, is formalized by the definition

(og) $p \perp q$ iff $P(p, x) + P(q, x) \leqslant 1$, for every $x \in \mathbf{S}$.

It is easy to prove that in the poset $\mathbf{Q}$ this binary relation is *a weak degenerate orthogonality*,[23,24] i.e., it satisfies the conditions:

(og1) $p \perp q$ implies $q \perp p$, (symmetry)

(og2) $\mathbb{0} \perp q$ for all $q \in \mathbf{Q}$, ($\mathbb{0}$-orthogonality)

(og3) $p_0 \leqslant p$ and $p \perp q$ imply $p_0 \perp q$. (absorption)

Let us remark that the degeneration property, i.e., the fact that there could be some elements $q \in \mathbf{Q}$ such that $q \perp q$, implies that this orthogonality relation is *not* irreflexive (i.e., such that $q \perp p$ implies $q \neq p$) and so it is *not* an orthogonality relation in the sense of the Foulis–Randall approach to operational statistics.[6–11] The set

$$\mathrm{Ker}(\perp) := \{ q \in \mathbf{Q}: q \perp q \}$$
$$:= \{ q \in \mathbf{Q}: P(q, x) \leqslant \tfrac{1}{2}, \quad \text{for every } x \in \mathbf{S} \}$$

is called the *kernel* of the orthogonality; the orthogonality is said to be *nondegenerate* iff $\mathrm{Ker}(\perp) = \{ \mathbb{0} \}$.[23,24]

From the properties of the orthogonality relation it follows that

$$p \perp q \quad \text{implies} \quad \mathbf{S}_1(p) \subseteq \mathbf{S}_0(q) \quad \text{and} \quad \mathbf{S}_1(q) \subseteq \mathbf{S}_0(p). \tag{2.5}$$

That is, $p \perp q$ implies that if q occurs with certainty (i.e., is "true"), then q nonoccurs with certainty (i.e., is "false"), and, if q occurs with certainty (i.e., is "true"), then p nonoccurs with certainty (i.e., is "false").

Notice that if $\{ q_1, q_2 \}$ is a pair of orthogonal questions then $0 \leqslant \Sigma_{i=1}^{2} P(q_i, x) \leqslant 1$. On the other hand, if $\{ q_1, q_2, \ldots, q_n \}$ is a set of pairwise orthogonal questions, then, in general, we cannot state that $0 \leqslant \Sigma_{i=1}^{n} P(q_i, x) \leqslant 1$.

Example: Let $\mathbf{S} = \{ r, y, v \}$ be a set of three ensembles and let $\{ q_1, q_2, q_3 \}$ be the set of questions schematically presented in

	q_1	q_2	q_3
r	1	0	0
y	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
v	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{8}$

Then the questions q_i are pairwise orthogonal but $\Sigma_{i=1}^{3} P(q_i, y) = \frac{3}{2}$ and $\Sigma_{i=1}^{3} P(q_i, v) = \frac{9}{8}$. Moreover, we have that $q_2 \perp q_2$.

Definition 2.1: A countable set of questions $\{ q_1, q_2, \ldots \}$ is said to be *globally orthogonal*, written $\{ q_1, q_2, \cdots \} \perp$, iff $0 \leqslant \Sigma P(q_i, x) \leqslant 1$.

Of course, $\{ q_1, q_2, \cdots \} \perp$ implies that every pair $\{ q_i, q_j \} \perp$, with $q_i \neq q_j$, but the converse, as the previous example shows, is not true.

3. GENERALIZED EFFECTS AND WEAK DEGENERATE ORTHOCOMPLEMENTATION

After these remarks about the poset with weak degenerate orthogonality $(\mathbf{Q}, \mathbb{0}, \mathbb{1}, \leqslant, \perp)$, we shall proceed to a classification of the elements from $\mathbf{Q}$ as to their orthogonality behavior.

Definition 3.1: In the poset with orthogonality $\mathbf{Q}$, for any fixed element $q \in \mathbf{Q}$, the set $\{ q \}^{\perp}$ defined as

$$\{ q \}^{\perp} := \{ p \in \mathbf{Q}: p \perp q \}$$
$$:= \{ p \in \mathbf{Q}: P(p, x) + P(q, x) \leqslant 1, \quad \text{for every } x \in \mathbf{S} \}$$

is called the *orthogonal* or *exclusive* domain of q.

Let $q \in \mathbf{Q}$; then we shall denote by q', if it exists, the element of $\mathbf{Q}$ such that $q' = \max \{ q \}^{\perp}$, i.e., such that the following two conditions are satisfied:

(i) $q' \perp q$;

(ii) let $p \perp q$; then $p \leqslant q'$.

Obviously, the element q', if it exists is also unique.

The element $q \in \mathbf{Q}$ is said to have *orthogonal complement* iff the following conditions are satisfied:

(ogco1) there exists $q' = \max \{ q \}^{\perp}$,

(ogco2) there exists $q'' = \max \{ q' \}^{\perp}$.

The collection of all questions q which possess orthogonal complement is denoted by $\mathbf{F}_g$. An element q from $\mathbf{F}_g$ is called a *generalized effect* or *generalized fuzzy event*. Obviously, the trivial questions $\mathbb{0}$ and $\mathbb{1}$ are generalized effects since $\mathbb{0}' = \mathbb{1}$ and $\mathbb{1}' = \mathbb{0}$.

Proposition 3.1: Let $q \in \mathbf{F}_g$ then there exists $q''' = \max \{ q'' \}^{\perp}$ and $q' = q'''$.

Proof: From (ogco2) it follows that $q' \perp q''$, that is $q' \in \{ q'' \}^{\perp}$. Let now $p \in \mathbf{Q}$ be such that $p \in \{ q'' \}^{\perp}$; we shall prove that $p \leqslant q'$. Since $q \perp q'$ follows by (ogco2), we have that $q \leqslant q''$, from which $\{ q'' \}^{\perp} \subseteq \{ q \}^{\perp}$ follows from the absorption property of orthogonality. We obtain that $p \in \{ q'' \}^{\perp}$ implies that $p \in \{ q \}^{\perp}$, and so it must be $p \leqslant q'$, since $q' = \max \{ q \}^{\perp}$.

Corollary: Let $q \in \mathbf{F}_g$; then also $q' \in \mathbf{F}_g$.

Proof: Obviously, $q'' = (q')'$ and $q' = (q'')' = (q')''$.

Remark: To be precise, $q \in \mathbf{F}_g$ iff there exist two questions $q', q'' \in \mathbf{Q}$ (and then, from previous results q' and $q'' \in \mathbf{F}_g$ too) such that

$$P(q, x) + P(q', x) \leqslant 1 \quad \text{for every} \quad x \in \mathbf{S}, \tag{3.1$'$}$$

$$P(q, x) + P(p, x) \leqslant 1 \quad \text{implies} \quad P(p, x) \leqslant P(q', x), \tag{3.1$''$}$$

$$P(q'', x) + P(q', x) \leqslant 1 \quad \text{for every} \quad x \in \mathbf{S}, \tag{3.2$'$}$$

$$P(q', x) + P(q, x) \leqslant 1 \quad \text{implies} \quad P(q, x) \leqslant P(q'', x). \tag{3.2$''$}$$

Proposition 3.2: The mapping $\mathbf{F}_g \to \mathbf{F}_g, q \to q'$ is a weak degenerate orthocomplementation, i.e., it satisfies the conditions:

(oc1) $q \leqslant q''$ for every $q \in \mathbf{F}_g$,

(oc2) $q \leqslant p$ implies $p' \leqslant q'$.

The kernel of the orthocomplementation $\mathrm{Ker}'(\mathbf{F}_g)$ is then

$$\mathrm{Ker}'(\mathbf{F}_g) = \{ q \in \mathbf{F}_g: q \leqslant q' \} = \{ q \in \mathbf{F}_g: q \perp q \}$$
$$= \{ q \in \mathbf{F}_g: P(q, x) \leqslant \tfrac{1}{2}, \quad \text{for every} \quad x \in \mathbf{S} \}.$$

Proof: Condition (3.2″), in the particular case of (3.1′), implies that $P(q, x) \leqslant P(q'', x)$ for every $x \in \mathbf{S}$, i.e., $q \leqslant q''$. Let now $q \leqslant p$ then $\{ p \}^{\perp} \subseteq \{ q \}^{\perp}$ is a consequence of the absorption property of the orthogonality and so $p' \leqslant q'$.

Remark: Equipped with the order relation induced from (or) the structure $(\mathbf{F}_g, \mathbb{0}, \mathbb{1}, \leqslant, ')$ of all generalized effects is therefore a *weak degenerate orthocomplemented poset*.

In Ref. 23, Proposition 1.6, it is shown that for such a poset the following properties are equivalent between them:

(og3a) $\mathrm{Ker}'(\mathbf{F}_g) = \{ \mathbb{0} \}$,

(og3b) $q \wedge q' = \mathbb{0}$ for every $q \in \mathbf{F}_g$.

If this is the case, the orthocomplementation is said to be *nondegenerate*.

From a phenomenological point of view, and comparing our approach with Mielnik's conclusions, we may say that:

(a) The generalized effects agree with the general definition of "questions" as proposed in Mielnik's Definition 1, quoted in Sec. 1.

(bi) The identity axiom (Axiom M.1) is substituted by the "stronger" indistinguishability principle (Axiom 2.2), which *implies* the identity axiom.

(bii) The orthocomplemented question q' of any element $q \in \mathbf{F}_g$ is unique, but the orthocomplementation is weak and degenerate.

The semitransparent mirror ST proposed by Mielnik as a counterexample is obviously an element of $\mathbf{F}_g$ belonging to the kernel since $ST = ST'$, and in this case as previously remarked, we cannot conclude that either $ST \wedge ST' = \mathbb{0}$ or $ST \vee ST' = \mathbb{1}$; on the contrary, for this particular question we rather have that $ST = ST \wedge ST' = ST \vee ST' = ST'$.

Therefore, the Mielnik conclusion that the whole structure of $\mathbf{F}_g$ would collapse is not correct since not all the following identities are right:

$$\mathbb{0} = ST \wedge ST' = ST \wedge ST = ST = ST \vee ST$$
$$= ST \vee ST' = \mathbb{1}.$$

On the other hand, starting from the previous analysis on the semitransparent mirror, Mielnik reports, as a possible answer to the inexact collapse result, the following as a general conclusion:

"One might reply, that the axioms of quantum logic are exact, but they must be properly understood. [More exactly]

(g) One feels that in order to be a quantum mechanical measuring device, the macroscopic arrangement should do something more specific than merely produce the 'yes' and 'no' effects in an arbitrary way. [Concluding that] not every arrangement producing a macroscopic alternative effect is a question." [1]

In the particular case of the semitransparent mirror:

"(g1) It is not a good example of a 'question' since it is not at all a measuring device: it does not verify any physical property of the transmitted photons.

(g2) In some axiomatic approaches the conclusion (g) is assured by requiring that the 'yes–no measurement' should have the nontrivial certainty domains: there should be some microsystems for which the answer 'yes' is certain (i.e., $\mathbf{S}_1(q) \neq \varnothing$) and some other for which the answer 'no' is certain (i.e., $\mathbf{S}_0(q) \neq \varnothing$)." [1]

He concludes that "this requirement eliminates the semitransparent window as an element of $\mathbf{Q}$." [1]

From our point of view, on the contrary, it is true that the semitransparent mirror *is* a question, i.e., an arrangement producing a macroscopic alternative effect. It possesses a *unique* orthocomplemented question, the semitransparent mirror itself; but the involved orthocomplementation is weak and degenerate.

More generally, we feel against (g) "that every arrangement producing a macroscopic alternative effect is a question" and this "means that the whole approach of 'quantum logic' *starts* from an information which is (in agreement with) the usually given." [1]

Moreover, as we shall show later, the semitransparent mirror, against (g1), also verifies a certain physical property of the transmitted photons, but in a manner which is deeply fuzzy. To be precise, it can be regarded as a fuzzy representation of the exact property associated to the impossible question $\mathbb{0}$ and this, against (g2), as a consequence of the fact that its certainly yes domain is empty, i.e., identical to the certainly yes domain of the impossible question $\mathbb{0}$.

4. THE DEGENERATE LOGIC OF EFFECTS

In this section we select from the set $\mathbf{Q}$ of the structure $(\mathbf{Q},\mathbf{S},P)$ a special collection of questions, called *effects* or *fuzzy events*, whose orthocomplemented ordered properties are studied.

Definition 4.1: An *effect* or *fuzzy event* of the structure $(\mathbf{Q},\mathbf{S},P)$ is any question $f \in \mathbf{Q}$ such that there exists a question $f' \in \mathbf{Q}$ which satisfies the condition

$$P(f, x) + P(f', x) = 1 \quad \text{for every} \quad x \in \mathbf{S}. \tag{4.1}$$

The set of all effects from $(\mathbf{Q},\mathbf{S},P)$ is denoted by $\mathbf{F}$. Obviously, every effect is also a generalized effect, i.e.,

$$\mathbf{F} \subseteq \mathbf{F}_g$$

and the trivial questions $\mathbb{0}$ and $\mathbb{1}$ are elements of $\mathbf{F}$.

In spite of the fact that if f "is an arbitrary macroscopic arrangement producing certain macroscopic alternative effects of which one is called 'yes' and the other is 'no', the f' is interpreted as essentially the same arrangement with an opposite convention determining what is 'yes' and what is 'no.' " [1] In fact the question f' is measured with exactly the same apparatus as f, the only difference being that the results are inverted: If f produces the answer "yes," then f' produces the answer "no" and vice versa.

Notice that

$$\mathbf{S}_1(f') = \mathbf{S}_0(f); \tag{4.2}$$

hence f' is true iff when we should decide to perform the test corresponding to f, we obtain with certainty the answer "no," i.e., f is false and vice versa.

Therefore, in conclusion, we have obtained a *state–effect–probability* structure $(\mathbf{F},\mathbf{S},P)$ which satisfies the following axioms introduced by Gunson[25] and which we quote in our notation and in a different order (we also quote in square brackets the corresponding Gunson numeration):

Axiom G.1[Axiom A.1]: $P(f_1, x) = P(f_2, x)$ for all $x \in \mathbf{S}$ iff $f_1 = f_2$; $f_1, f_2 \in \mathbf{F}$.

Axiom G.2[Axiom A.3]: $P(f, x_1) = P(f, x_2)$ for all $f \in \mathbf{F}$ iff $x_1 = x_2$; $x_1, x_2 \in \mathbf{S}$.

Axiom G.3 [Axiom A.]: There exists $\mathbb{0} \in \mathbf{F}$ such that $P(\mathbb{0}, x) = 0$ for all $x \in \mathbf{S}$.

Axiom G.4 [Axiom A.5]: For every $f \in \mathbf{F}$ there is an element $f' \in \mathbf{F}$ such that

$$P(f, x) + P(f', x) = 1 \quad \text{for all } x \in \mathbf{S}.$$

An order relation $\leqslant$ can be induced on $\mathbf{F}$ from the canonical order relation:

$$\text{(or)} \quad f_1 \leqslant f_2 \quad \text{iff} \quad P(f_1, x) \leqslant P(f_2, x) \quad \text{for all } x \in \mathbf{S}.$$

Let us remark that Axiom A.4 of the Gunson paper is just the definition of the order relation $\leqslant$ we have now introduced.

Of course, if $f \in \mathbf{F}$, then also $f' \in \mathbf{F}$ and the mapping $\mathbf{F} \to \mathbf{F}, f \to f'$ is a *degenerate orthocomplementation* on $\mathbf{F}$, that is, it satisfies the conditions:

(oc1) $f = f''$ for every $f \in \mathbf{F}$,

(oc2) $f_1 \leqslant f_2$ implies $f_2' \leqslant f_1'$.

The orthocomplementation kernel is the set

$$\mathrm{Ker}'(\mathbf{F}) = \{ f \in \mathbf{F}: f \leqslant f' \}$$
$$= \{ f \in \mathbf{F}: P(f, x) \leqslant \tfrac{1}{2} \quad \text{for every} \quad x \in \mathbf{S} \},$$

and in this case the following propositions are equivalent:

(oc3a) $\mathrm{Ker}'(\mathbf{F}) = \{ \mathbb{0} \}$,

(oc3b) $f \vee f' = \mathbb{0}$ for every $f \in \mathbf{F}$,

(oc3c) $f \wedge f' = \mathbb{1}$ for every $f \in \mathbf{F}$.

In this way $(\mathbf{F}, \mathbb{0}, \mathbb{1}, \leqslant, ')$ is a degenerate orthocomplemented poset in which the generalized de Morgan laws hold:

(DM1) If $f \vee g$ exists, then $f' \wedge g'$ exists and it is $f' \wedge g' = (f \vee g)'$.

(DM2) If $f \wedge g$ exists, then $f' \vee g'$ exists and it is $f' \vee g' = (f \wedge g)'$.

Example 4.1a: The semitransparent mirror discussed by Mielnik is an effect belonging to the orthocomplementation kernel and for this effect we cannot state neither $ST \vee ST' = \mathbb{1}$ nor $ST \wedge ST' = \mathbb{0}$. This result is inconsistent with Theorem 2.1 quoted by Gunson,[25] in which it is asserted that the axioms from G.1 to G.4 assure that the complementation $f \to f'$ is nondegenerate. This incongruence has already been recognized by Kupczynski,[26] which presents the following counterexample:

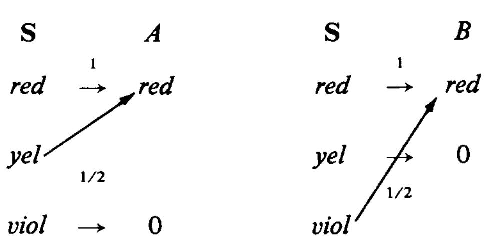
S consists of only two pure ensembles x_1, x_2, and F of two detectors d and l, their corresponding complements, and $\mathbb{0}$ *and* $\mathbb{1}$; the probability of occurrence is given by the following values:

	d	l	d'	l'
x_1	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{3}{4}$	$\frac{7}{8}$
x_2	$\frac{1}{3}$	$\frac{1}{7}$	$\frac{2}{3}$	$\frac{6}{7}$

The structure $(\mathbf{F},\mathbf{S},P)$ obviously satisfies the Gunson axioms, and we have in particular that $l \leqslant d$ according to the (or) order relation. The result that $d \leqslant d'$ (or $l \leqslant l'$) only tell us that d and $l \in \mathrm{Ker}'(\mathbf{F})$ and thus that the involved orthocomplementation is degenerate, but has nothing to do with the consideration that the "implication $\leqslant$ is physically completely unjustified."[26]

Example 4:1: The two hypothetical macroscopic devices A and B acting on mixtures of reds, yellow, and violet light proposed by Mielnik[1] can be considered as nondegenerate effects of a suitable structure $(\mathbf{F},\mathbf{S},P)$. In fact, "the device A transmits the red photons and absorbs the yellow and violet ones: however, it reemits an average 1/2 of the absorbed yellow photons in the form of red photons. The device B is also transparent for the red photons and absorbs the yellow and violet ones: now, however, 1/2 of the violet photons are reemitted in the form of red photons.

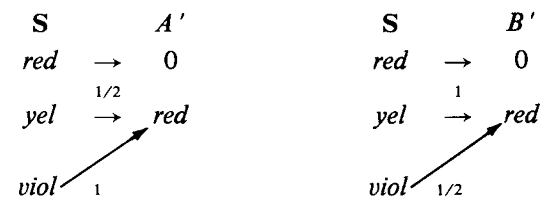
Schematically:



Both devices A and B have the common 'certainly yes' domain: they are completely transparent only to the red photons. ... However, A and B have different domains of 'certainly no,' and so, *they are not physically equivalent*."[1]

This last consideration does agree with our assumption of substituting the identity axiom with the indistinguishability principle: therefore A and B are two *different* questions, i.e., are arrangements producing two *different* macroscopic alternative effects, and cannot be considered as two different physical realizations of the same abstract question. Rather, as we shall see later, they are merely two different fuzzy realizations of the same *property*: "the light is red."

The corresponding orthocomplemented questions are the hypothetical macroscopic devices A' and B' schematically shown in the next figures:



These last are two essentially different prescriptions for producing the negative of the questions A and B, respectively. If S is the set of all possible mixtures of red, yellow, and violet light, and Q is the set of all detectors of red lights, then considering the structure $(\mathbf{F},\mathbf{S}, P)$ we have that:

(a) $A, B, \in \mathbf{F}$, i.e., are effects with $A \neq B$.

(b) Both the questions A and B have an unique orthocomplemented question, A' and B', respectively.

(c) A and B are not elements of $\mathrm{Ker}'(\mathbf{F})$.

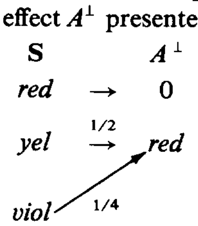
As usual, from the degenerate orthocomplementation on F we can induce a degenerate orthogonality relation $\perp$ of *mutually exclusivity* defined by one of the equivalent conditions:

$$f_1 \perp f_2 \quad \text{iff} \quad f_1 \leqslant f_2'$$
$$\text{iff} \quad f_2 \leqslant f_1'$$
$$\text{iff} \quad P(f_1, x) + P(f_2, x) \leqslant 1, \quad \textit{for all} \ x \in \mathbf{S}.$$

Once an effect $f \in \mathbf{F}$ is fixed, we shall denote by $\{ f \}^{\perp}$ the set of all effects $f^{\perp} \in \mathbf{F}$ which exclude f, i.e., such that $(f^{\perp}) \perp f$. Since $f' = \max \{ f \}^{\perp}$ we have that

$$\mathbf{S}_1(f^{\perp}) \subseteq \mathbf{S}_1(f') = \mathbf{S}_0(f),$$
$$\mathbf{S}_1(f) = \mathbf{S}_0(f') \subseteq \mathbf{S}_0(f^{\perp}).$$

Therefore, there exist as many possible effects which exclude the effect f, at least as many as are the elements of $\{ f \}^{\perp}$. The unique orthocomplemented effect f' of f has the property that it excludes f, it is the lub of all other effects which exclude f and so its certainly no (resp. yes) domain is the minimum (resp. maximum) as to all effects which are opposite to f. In the case of Example 1, an effect which is exclusive to A is the effect $A^{\perp}$ presented in the following scheme:



In conclusion, to sum up the discussion up to now we have seen that from every question–state structure $(\mathbf{Q},\mathbf{S},P)$ it is possible to single out the degenerate orthocomplemented poset $(\mathbf{F},\mathbb{0},\mathbb{1},\leqslant,')$ of all effects. The counterexamples presented by Mielnik are effects, and, using the conventional language of the axiomatic approach to quantum theory, we can say that $\mathbf{F}$ is the *"logic" of all effects* inside the question–state structure.

Orthogonality axioms

A state–effect structure satisfies the *strong orthogonality axiom* iff the following condition holds for the set of effects:

Axiom OG.S: For any finite set of pairwise orthogonal elements of $\mathbf{F}$, $\{f_i: i=1,2,...,n\}$, then

(1) $0 \leqslant \sum_{i=1}^{n} P(f_i,x) \leqslant 1$ for all $x \in \mathbf{S}$,

(2) there is an element $f \in \mathbf{F}$ (called the *sum* of the f_i) such that $P(f,x) = \sum_{i=1}^{n} P(f_i,x)$ for all $x \in \mathbf{S}$.

Let us remark that this strong orthogonality axiom is just Axiom A.6 of the Gunson paper and as Finch has shown in Ref. 27 if Axiom OG.S holds then the poset of all effects of a state–effect structure is an orthomodular orthoposet.

A state–effect structure satisfies the *weak orthogonality axiom* iff the set of effects is such that the following condition holds:

Axiom OG: For every finite sequence $\{f_1,f_2,...,f_n\}$ of globally orthogonal effects, there exists an effect $f \in \mathbf{F}$ (called the *sum* of the f_i) such that

$$P(f,x) = \sum_{i=1}^{n} P(f_i,x) \quad \text{for all } x \in \mathbf{S}.$$

Of course, the strong orthogonality axiom implies the weak orthogonality axiom.

5. YES–NO EXPERIMENTS AND JP AXIOMS

In their axiomatic approach to quantum mechanics, Jauch and Piron are mainly interested in the behavior of the certainly yes domains of questions rather than in their probability occurrence with respect to all preparation procedures.

"A certain class of measurements plays a particularly important role in the establishment of the physical properties of a system. It is the experiment (α) with only two possible results which may be denoted by 1 or 0 (yes or no). We ... shall refer to (such experiments) as yes–no experiments." [13]

We agree with the remark that "a formalization in the framework of mathematics ... of the intuitive sentence 'α is true' is only possible if one also has as a basic set the set $\mathbf{S}$ of states." [2] Indeed, in one of his works Jauch also explicitly asserts: "we shall say a yes–no measurement is 'true' in a *particular state* if its measurement will give the result yes with certainty (probability one)." [15]

We now try to translate Jauch–Piron theory in the context of state–effect structure, quoting at the beginning of any definition the original numeration presented in Refs. 28 and 29.

In our structure a possible interpretation of a yes–no experiment $\alpha(f)$ associated with the effect f could be a suitable pair of subsets of ensembles according to the following definition:

D_2 (*yes–no experiment*): $\alpha(f) := (\mathbf{S}_1(f), \mathbf{S}_0(f))$.

D_5 (*trivial yes–no experiments*): The yes–no experiments associated with the trivial effects are:

$$\alpha(\mathbb{0}) = (\varnothing,\mathbf{S}) \quad \text{and} \quad \alpha(\mathbb{1}) = (\mathbf{S},\varnothing).$$

In this way, the set of all yes–no experiments which can be performed on the system described by the state–effect structure $(\mathbf{F},\mathbf{S},P)$ is the collection of pairs of subsets of $\mathbf{S}$ defined as

$$\mathbf{Y} := \{(A_1,A_0): \exists f \in \mathbf{F} \ni A_1 = \mathbf{S}_1(f), A_0 = \mathbf{S}_0(f)\},$$

which satisfies the conditions

(i) $(A_1,A_0) \in \mathbf{Y}$ implies $(A_0,A_1) \in \mathbf{Y}$,

(ii) $A_1 \cap A_0 = \varnothing$.

An effect $f \in \mathbf{F}$ is an experimental equipment which measures or *realizes* the yes–no experiment $\alpha \in \mathbf{Y}$ iff $\alpha = \alpha(f)$. In general, there could exist several different effects which measure the same yes–no experiment. Hence, from D_2 we can introduce a mapping $\alpha: \mathbf{F} \to \mathbf{Y}$ which, in general, is not one-to-one, contrary to the assumption that "it is possible to identify $\mathbf{F}$ with $\mathbf{Y}$." [2] Following Ref. 29 "a bridge between (our model of) Piron's language and the ordinary predicate calculus (can be sketched) by a set of 'connecting definitions' *CD*."

CD_1: For any yes–no experiment α and any ensemble x and for any realization f of α [i.e., $\alpha = \alpha(f)$], we denote by $p_\alpha(x,f)$ the following proposition in the sense of the predicate calculus: The result of the realization f of the yes–no experiment α on the system prepared according to x is "yes" with certainty:

$$p_\alpha(x,f) \quad \text{iff} \quad \alpha = \alpha(f) \quad \text{and} \quad P(f,x) = 1.$$

CD_2: The proposition from the predicate calculus $\sim p_\alpha(x,f)$, called the *negation* in the sense of the ordinary logic, is the one obtained as CD_1, where the last word "yes" is changed to "no":

$$\sim p_\alpha(x,f) \quad \text{iff} \quad \alpha = \alpha(f) \quad \text{and} \quad P(f,x) = 0.$$

CD_3: [α is "true" for the ensemble x_0] iff $[(\forall f)\, p_\alpha(x_0,f)]$.

The yes–no experiment α is "true" for the particular ensemble x_0 iff its measurement by any realization f of α will give the result yes with certainty (i.e., probability 1).[15]

D_6: (*certain or true yes–no experiment*): "When the physical system has been prepared in such a way (i.e., $x_0 \in \mathbf{S}$) that the physicist can affirm that in the event of an experiment (i.e., $f \in \mathbf{F}$) corresponding to a yes–no experiment α [i.e., $\alpha = \alpha(f)$] the result will be "yes" [i.e., $P(f,x_0) = 1$], the yes–no experiment α is *certain* or the yes–no experiment α is *true*." [28]

The set of all yes–no experiments can be considered as a structure $(\mathbf{Y},\alpha(\mathbb{0}),\alpha(\mathbb{1}),\subseteq,^{\vee})$, once the following definitions are introduced:

D_7 (*preorder relation*): $\subseteq$ is the partial preorder relation defined in the following way: Let $\alpha = (A_1,A_0)$ and $\beta = (B_1,B_0)$; then $\alpha \subseteq \beta$ iff $A_1 \subseteq B_1$.

We shall say the yes–no measurement β is "stronger" than the yes–no measurement α iff whenever α is true then β is also true.

CD_4: $[\alpha \subseteq \beta]$ iff $(\forall x)[(\forall f) p_\alpha(x,f) \subseteq (\forall g) p_\beta(x,g)]$.

"A natural partial (pre)-ordering is defined on the set of all yes–no experiments. For example, if the system is *prepared* in such a way (i.e. $x \in A_1$) that whenever the yes–no experiment $\alpha = (A_1, A_0)$ is true, then there may be another $\beta = (B_1, B_0)$ which is also true with certainty (i.e., $x \in B_1$)." [14]

D_3 (*opposite or negation*): Let $\alpha = (A_1, A_0)$; then $\alpha^v = (A_0, A_1)$.

If α is a yes–no experiment, α^v is the yes–no experiment obtained by exchanging the terms yes and no of the alternative. Of course, if $\alpha = \alpha(f)$, then $\alpha^v = \alpha(f')$, and so, if the effect f realizes the yes–no experiment α, then the effect f' realizes the yes–no experiment α^v.

"If α is a yes–no measurement then there exists another one, denoted by α^v, obtained from α by inverting the results yes and no. Thus if the result of α is 'yes' that of α^v is 'no' and vice-versa. It is clear that α^v can be measured with the same physical equipment as that used for the measurement of α." [13]

So we have a mapping $\mathbf{Y} \to \mathbf{Y}, \alpha \to \alpha^v$ with the properties:

(i) $(\alpha^v)^v = \alpha$,

(ii) let $\beta = (B_1, B_0)$ and $\beta \subseteq \{\alpha, \alpha^v\}$; then $B_1 = \alpha(\mathbb{0})$.

In the Jauch–Prion approach a fundamental role is played by the axioms which we now get to introduce. Quoting Ref. 3, "in general the results of a test of one question are profoundly influenced by the testing of another question. In most cases it makes even no sense to perform two tests on the same (system).

There is indeed a way to construct a question that makes it possible to test several questions at once.

Let us analyze this first on an example:

We take ... a piece of wood as (system). We consider the following two questions:

(q) "set the wood on fire and give the answer 'yes' if it burns,"

(p) "make the wood float on water and give the answer 'yes' if it floats."

If we perform first the test p, and make the piece of wood float on water, we have brought the wood in the state of wet wood and as a result the wood will not burn anymore. On the other hand, if we perform the test q and burn the wood, it will not float anymore on water.

We shall make the following definition:

Given two questions q and p we define a new question, denoted by $p \cdot q$, that consists

(1) of choosing one of the two questions at random,

(2) performing the test corresponding to this chosen question,

(3) attributing the answer obtained in this way to the question $p \cdot q$.

We will call this question the *product* of p and q.

The product question $p \cdot q$ is well defined since the measuring apparata and the manuals that we need are just the measuring apparata and the manuals that we have for p and for q."

A possible realization of the product of two effects f_1 and f_2 is an effect $f_1 f_2$ such that

$$P(f_1 f_2, x) = \tfrac{1}{2}\,[P(f_1, x) + P(f_2, x)] \quad \text{for all} \quad x \in \mathbf{S}.$$

Indeed, this effect can be concretely executed preparing an ensemble of $2N$ sample objects according to the procedure x. In correspondence to each of these sample objects we choose at random one of the two equipments f_1 or f_2 and record the outcome of its measurement on the single object as the outcome of $f_1 f_2$. Of course, the test of f_1, on average, has been performed on N of the original sample objects and the test of f_2 on the other N sample objects. Let N'_+ and N''_+ be the numbers of sample objects yielding the effects f_1 and f_2, respectively. Then $N'_+ + N''_+$ is assumed to be the number of the original $2N$ sample objects which yield the effect $f_1 f_2$ and so, in physical approximation, we have that

$$P(f_1 f_2, x) \sim (N'_+ + N''_+)/2N.$$

Let us remark that according to this definition we have that

$$P(f_1 f_2, x) = 0 \quad \text{iff} \quad P(f_1, x) = P(f_2, x) = 0,$$

$$P(f_1 f_2, x) = 1 \quad \text{iff} \quad P(f_1, x) = P(f_2, x) = 1.$$

Hence, a state–effect structure satisfies the axiom JP iff the following statement on the set of effects holds:

Axiom JP: For any pair of effects $\{f_1, f_2\}$ there exists at least an effect denoted by $f_1 f_2$ such that

(1) $\mathbf{S}_1(f_1 f_2) = \mathbf{S}_1(f_1) \cap \mathbf{S}_1(f_2)$,

(2) $\mathbf{S}_0(f_1 f_2) = \mathbf{S}_0(f_1) \cap \mathbf{S}_0(f_2)$.

Remark: If $\{f_1, f_2, f_3\}$ are effects then there exist the effects $(f_1 f_2) f_3, f_3(f_1 f_2), f_1(f_2 f_3)$, and so on. *A priori*, these effects do not necessarily coincide, but any one of them is characterized by the same certainly yes domain $\cap\{\mathbf{S}_1(f_i)\colon i = 1,2,3\}$ and the same certainly no domain $\cap\{\mathbf{S}_0(f_i)\colon i = 1,2,3\}$.

In general, if we have any family of effects f_i, the effect $\Pi_i f_i$, called the product of the f_i, is the following:

We choose at random one of the f_i and measure it. The answer obtained by performing the test of this chosen effect is then attributed to $\Pi_i f_i$.

A state–effect structure satisfies the axiom c-JP iff the set of all effects is such that the following statement holds:

Axiom c-JP: For every set $\{f_j : j \in J\}$ of effects there exists at least one effect Πf_j such that

(1) $\mathbf{S}_1(\Pi f_j) = \cap \mathbf{S}_1(f_j)$,

(2) $\mathbf{S}_0(\Pi f_j) = \cap \mathbf{S}_0(f_j)$.

Let one of the JP axioms be true; then if $\{\alpha_j\}$ is a set of yes–no experiments, $\alpha_j = \alpha(f_j) = (\mathbf{S}_1(f_j), \mathbf{S}_0(f_j))$, the yes–no experiment associated with $\Pi(f_j)$ will be denoted by $\Pi\alpha_j$, i.e.,

D_4 (*product of yes–no experiments*):

$$\Pi\alpha_j := \alpha(\Pi f_j) = (\cap \mathbf{S}_1(f_j), \cap \mathbf{S}_0(f_j)).$$

"If $\{\alpha_j\}$ is a family of yes–no experiments, $\Pi\alpha_j$ is the yes–no experiment defined in the following manner: One measures an arbitrary one of the α_j (i.e., measures any $f_j \in \mathbf{F}$ such that $\alpha_j = \alpha(f_j)$) and then attributes to $\Pi\alpha_j$ the answer thus obtained (i.e., $\Pi\alpha_j = \alpha(\Pi f_j)$)." [28]

Of course, once $\{\alpha_j : j \in J\}$ is fixed there could exist several different effects which satisfy Axiom JP, but the corresponding yes–no experiment $\Pi\alpha_j$ is unique. At any rate, by starting from the definitions one can verify the following rule:

Rule R_1 (opposite of product yes–no experiments):

$$(\Pi\alpha_j)^{\vee} = \Pi\alpha_j^{\vee}.$$

Indeed, let $\alpha_j = \alpha(f_j)$; then $(\Pi\alpha_j)^{\vee} = (\cap \mathbf{S}_0(f_j), \cap \mathbf{S}_1(f_j))$. On the other hand, $\alpha_j^{\vee} = (\mathbf{S}_0(f_j), \mathbf{S}_1(f_j)) = \alpha(f_j')$ and so $\Pi\alpha_j^{\vee} = \alpha(\Pi f_j') = (\cap \mathbf{S}_0(f_j), \cap \mathbf{S}_1(f_j))$.

Remarks: (a) In our model of Jauch–Piron theory, if the state–effect structure satisfies Axiom c-JP, then the yes–no experiments $\alpha^{\vee}$ and $\Pi\alpha_j$ exist, contrary to Ludwig's model in which "the operations ... $\alpha^{\vee}$ and $\Pi\alpha_j$ in the theory of Jauch and Piron do not occur." [2]

(b) The model briefly reproduced above is built on two interconnected levels: the level of *effects* (or questions) and the level of *yes–no measurements*. Each yes–no measurement can be tested by a wide class of effects which represent it. These two levels give rise to two quite different structures.

The pre-Hilbert space model of quantum theory

Generalizing the Hilbert space approach to quantum mechanics we construct in this example a state-effect structure based on a pre-Hilbert space $\mathbf{k}$. To be precise, let $\mathbf{k}$ be a complex separable infinite-dimensional pre-Hilbert space, we do consider the triplet $(\mathbf{F}(\mathbf{k}),\mathbf{S}(\mathbf{k}),P)$, where

(a) $\mathbf{S}(\mathbf{k})$ is the set of all one-dimensional subspaces x of $\mathbf{k}$.

(b) $\mathbf{F}(\mathbf{k})$ is the set of all self-adjoint linear operators F defined on $\mathbf{k}$ such that the following condition holds:

$$0 \leqslant \langle F\psi|\psi\rangle \leqslant \|\psi\|^2 \quad \text{for every} \quad \psi \in \mathbf{k}.$$

(c) P is the mapping from $\mathbf{F}(\mathbf{k}) \times \mathbf{S}(\mathbf{k})$ into $[0,1]$ defined as

$$P(F,x) = \frac{\langle F\psi_x|\psi_x\rangle}{\|\psi_x\|^2} \quad \text{for any} \quad \psi_x \in x/\{\underset{\sim}{0}\}.$$

Any element x of $\mathbf{S}(\mathbf{k})$ represents an *ensemble* of physical systems and any element F of $\mathbf{F}(\mathbf{k})$ is a *question* which can be measured on the system. The couple (F, x) represents then an *elementary experiment* and the number $P(F, x)$ is the probability of occurrence of the question F when the system is prepared according to x.

Once fixed $F \in \mathbf{F}(\mathbf{k})$, we introduce

(1) The certainly yes domain of F:

$$\mathbf{S}_1(F) := \{x \in \mathbf{S} : P(F,x) = 1\},$$

which can be identified with $\equiv \{\psi \in \mathbf{k} : \langle F\psi|\psi\rangle = \|\psi\|^2\}$.

(2) The certainly no domain of F:

$$\mathbf{S}_0(F) := \{x \in \mathbf{S} : P(F,x) = 0\},$$

which can be identified with $\equiv \{\varphi \in \mathbf{k} : \langle F\varphi|\varphi\rangle = 0\}$.

The canonical order relation introduced on $\mathbf{F}(\mathbf{k})$ can be expressed in the following way:

$$F_1 \leqslant F_2 \quad \text{iff} \quad \langle F_1\psi|\psi\rangle \leqslant \langle F_2\psi|\psi\rangle \quad \text{for every} \quad \psi \in \mathbf{k}.$$

Taking into account this order relation, the condition (b) can be rewritten as:

$$\text{(bi)} \quad F \in \mathbf{F} \quad \text{iff} \quad F = F^* \quad \text{and} \quad \mathbb{0} \leqslant F \leqslant \mathbb{1}.$$

Of course, for every $F \in \mathbf{F}(\mathbf{k})$ the element $F' := \mathbb{1} - F$ belongs to $\mathbf{F}(\mathbf{k})$ too and is such that

$$P(F,x) + P(F',x) = 1 \quad \text{for every} \quad x \in \mathbf{S}.$$

In spite of this and following the terminology introduced in Sec. 4, we have that any element of $\mathbf{F}(\mathbf{k})$ is just an effect rather than a question. Therefore, according to our general results,

(*I*) $(\mathbf{F}(\mathbf{k}),\mathbb{0},\mathbb{1},\leqslant,')$ *is a degenerate orthocomplemented poset whose kernel is nonempty since the effect* $\frac{1}{2}F \in Ker'$ $(\mathbf{F})$ *for every* $F \in \mathbf{F}$.

We shall also remark that the set $\Pi(\mathbf{k})$ of all orthogonal projections is contained in $\mathbf{F}(\mathbf{k})$, i.e., any orthogonal projection is an effect too. For instance, in a pre-Hilbert description the semitransparent mirror could be described by an element of Ker'$(\mathbf{F})$, precisely by $ST := \frac{1}{2}\mathbb{1}$. In this case we have that $ST = ST' = \frac{1}{2}\mathbb{1}$ from which it follows that:

$$P(ST,x) = P(ST',x) = \tfrac{1}{2} \quad \text{for every} \quad x \in \mathbf{S}(\mathbf{k}).$$

Notice that $(ST) \perp (ST)$ and so even in the pre-Hilbert model of state–effect structure the orthogonality is not irreflexive.

(*II*) *The state–effect structure based on a pre-Hilbert space satisfies the strong orthogonality condition*:

(OG) If $\{F_1, F_2, ..., F_n\} \perp$ is a global orthogonal set of effects, then there exists the effect $F = \Sigma_{i=1}^n F_i$ such that

$$P(F,x) = \sum_{i=1}^{n} P(F_i, x).$$

Notice that in a state–effect based on a pre-Hilbert space the strong orthogonality condition (OG.S) does not hold. For instance, the set $\{F_1, F_2, F_3\}$, where each $F_i = \frac{1}{2}\mathbb{1}$ is a set of pairwise orthogonal effects for which $\Sigma_{i=1}^3 P(F_i, x) = \frac{3}{2}$. Since for every $F \in \mathbf{F}(\mathbf{k})$ we have that $\langle F\psi|\psi\rangle = 0$ iff $F\psi = \underset{\sim}{0}$, we can make the following further identifications:

$$\text{(1a)} \quad \mathbf{S}_0(F) \equiv \mathrm{Ker}(F),$$

$$\text{(2a)} \quad \mathbf{S}(F) \equiv \mathrm{Ker}(\mathbb{1} - F) = \{\psi \in \mathbf{k} : F\psi = \psi\}.$$

The yes–no experiment $\alpha(F)$ associated with the effect F is then identifiable with the pair of subspaces of $\mathbf{k}$:

$$\alpha(\mathbf{F}) \equiv (\mathrm{Ker}(\mathbb{1} - F), \ \mathrm{Ker}(F)).$$

Lemma: If A,B are positive operators defined on $\mathbf{k}$, then

(a) $\mathrm{Ker}(A + B) = \mathrm{Ker}(A) \cap \mathrm{Ker}(B)$,

(b) $\mathrm{Ker}\,(\alpha A) = \mathrm{Ker}(A)$.

Proof: If $x \in \mathrm{Ker}(A) \cap \mathrm{Ker}(B)$, then obviously $x \in \mathrm{Ker}(A + B)$. On the contrary, if $x \in \mathrm{Ker}(A + B)$, then $\langle Ax|x\rangle = -\langle Bx|x\rangle$, and, so from the positivity of the two operators, we get that $\langle Ax|x\rangle = \langle Bx|x\rangle = 0$, i.e., $x \in \mathrm{Ker}(A) \cap \mathrm{Ker}(B)$.

(*III*) *The state–effect structure* $(\mathbf{F}(\mathbf{k}),\mathbf{S}(\mathbf{k}),P)$ *based on a pre-Hilbert space* $\mathbf{k}$ *satisfies the Axiom JP*.

Indeed, for any pair effects $F_1, F_2 \in \mathbf{F}(\mathbf{K})$ the operator

$$F_1 \cdot F_2 := (F_1 + F_2)/2$$

is an effect too. This effect is the "product" of the two effects F_1, F_2 since

$$\begin{aligned}\mathrm{Ker}\,[\mathbb{1} - (F_1 + F_2)/2] &= \mathrm{Ker}\,\{\tfrac{1}{2}[(\mathbb{1} - F_1) + (\mathbb{1} - F_2)]\} \\ &= \mathrm{Ker}(\mathbb{1} - F_1) \cap \mathrm{Ker}(\mathbb{1} - F_2), \quad (5.1)\end{aligned}$$

$$\mathrm{Ker}[(F_1 + F_2)/2] = \mathrm{Ker}(F_1) \cap \mathrm{Ker}(F_2). \tag{5.2}$$

Let us remark that this "product" operation on effects is not associative. Indeed,

$$(F_1 F_2) F_3 = (F_1 + F_2 + 2F_3)/4,$$
$$F_1(F_2 F_3) = (2F_1 + F_2 + F_3)/4.$$

Hence, $(F_1 F_2) F_3 = F_1(F_2 F_3)$ iff $F_1 = F_3$.

Differently from Ref. 28, where the yes–no experiments "are represented by closed subspaces of a Hilbert space, [and] a product of yes–no experiments is represented by a sum [of sets of subspaces]," in our pre-Hilbertian model the yes–no experiments are represented by suitable pairs of closed subspaces of a pre-Hilbert space and a finite product of yes–no experiments by the pair of closed subspaces obtained intersecting the original ones; for instance,

$$\alpha(F_1 F_2) = (\mathrm{Ker}(\mathbb{1} - F_1) \cap \mathrm{Ker}(\mathbb{1} - F_2),\ \mathrm{Ker}(F_1) \cap \mathrm{Ker}(F_2)).$$

Each yes–no experiment is realized by at least one effect and the product of two yes–no experiments, realized by the effects F_1 and F_2, is realized by the effect $(F_1 + F_2)/2$. In this way no "drastic distortion" is introduced in the representation of the *finite* product of yes–no experiments concluding that there are no "difficulties in also mimicking the semantic structure associated with the quantum mechanical formalism." [28]

All the definitions $D_2, D_3, D_4, D_5, D_6, D_7$, the rule R_1, and the connecting definitions CD_1, CD_2, CD_3, CD_4 can be applied to this pre-Hilbertian model. In particular, contrary to Ref. 28, "the definition D_4 for a (finite) product of yes–no experiments, with the definition D_3 of the opposite yes–no experiment and with the consequent rule R_1 for the opposite of a product of yes–no experiments" do not raise any "grave formal problem," [28] and so there is no "source of a semantic barrier in the way of the interpretability of the (yes–no experiments structure) as a physical theory." [28]

(III–H) The state–effect structure $(\mathbf{F}(\mathbf{H}),\mathbf{S}(\mathbf{H}), P)$ *based on a Hilbert space* $\mathbf{H}$ *satisfies the axiom c-JP.*

Indeed, let $\{F_j : j \in J\}$ be any family of effects and let $M_1 = \cap\,\mathrm{Ker}(\mathbb{1} - F_j)$ and $M_2 = \cap\,\mathrm{Ker}(F_j)$. M_1 and M_2 are two subspaces of the Hilbert space $\mathbf{H}$ which are mutually orthogonal since $\mathrm{Ker}(\mathbb{1} - F_j) \perp \mathrm{Ker}(F_j)$ to each $j \in J$. Therefore, once the two orthogonal projections E_1 and E_2, which project onto M_1 and M_2, respectively, are introduced, we have that $E_1 \perp E_2$. The product $E_1 E_2 = (E_1 + E_2)/2$ is an effect which generates the yes–no experiment $\alpha(E_1 E_2) = (\cap\,\mathrm{Ker}(\mathbb{1} - F_j), \cap\,\mathrm{Ker}(F_j))$.

6. PROPERTIES AND PROPOSITIONS

It is now quite natural to face the following question: Can the "properties" testable on a physical system be derived inside the general question–state structure which describes the physical system under examination? In this section an affirmative answer to this question is given. To this end we introduce on the structure $(\mathbf{F},\mathbb{0},\mathbb{1},\leqslant,')$ the equivalence relation $\sim$ defined as:

$$f_1 \sim f_2 \quad \text{iff} \quad \mathbf{S}_1(f_1) = \mathbf{S}_1(f_2). \tag{6.1}$$

If $f \in \mathbf{F}$, we shall denote by $[f]_{(1)}$ the equivalence class generated by f, i.e.,

$$[f]_{(1)}: \{g \in \mathbf{F}: \mathbf{S}_1(g) = \mathbf{S}_1(f)\}.$$

In the particular cases of the trivial effects we have

$$[\mathbb{1}]_{(1)} = \{\mathbb{1}\},$$
$$[\mathbb{0}]_{(1)} = \{g \in F: P(g, x) \neq 1, \quad \text{for all } x \in \mathbf{S}\}.$$

This last is the collection of all effects that are never true.

Definition 5.1: Once the quotient set $\Pi = (\mathbf{F}/\sim)$ is considered, we define a *proposition* as an equivalence class of effects.

With each proposition $\hat{a} \in \Pi$ we associate the *certainly yes domain* $\mathbf{S}_1(\hat{a})$ defined as $\mathbf{S}_1(f)$ for any arbitrary $f \in \hat{a}$. Hence, we can say that the proposition is "true" if and only if any and therefore all of its effects are "true."

Definition 5.2: A *property* testable on the system is a pair $(\hat{a},\mathbf{S}_1(\hat{a}))$ consisting of

(1) a proposition $\hat{a}$, i.e., the set of all macroscopic apparatuses (effects) $f \in \hat{a}$ the property can be measured by,
(2) the certainly yes domain $\mathbf{S}_1(\hat{a})$, i.e., the subset of all ensembles $x \in \mathbf{S}_1(\hat{a})$ for which the answer "yes" to the proposition $\hat{a}$ is certain. In this case we shall say that the property is "true" for the ensemble x.

Therefore, if $f_1 \sim f_2$, then the effects f_1 and f_2 test the same property and generate the same proposition.

For sake of simplicity we shall identify the properties which can be tested on the system with the corresponding propositions owing to the one-to-one correspondence $\hat{a} \longleftrightarrow \mathbf{S}_1(\hat{a})$.

At this stage of the discussion the notion of property "cannot meaningfully be attributed to an individual system; it is a statistical concept, applicable only to a suitable chosen assembly of systems." [13] Therefore, linked to the question proposed at the beginning of the present section is the "controversial problem, whether the formalism of quantum theory can be used to describe the properties of a *single* micro-object as it is, in all its complexity.

The single act of measurement in quantum mechanics is not conclusive, and therefore, the direct interpretation of quantum mechanical formalism is that of a statistical scheme.

The notion of property of a *single* system can, however, be introduced as a next abstraction stage of the theory." [1]

If we now identify a property with its certainly yes domain, we also agree with Mielnik philosophy that: "if now the ensembles are represented by points of the ... set $\mathbf{S}$, the properties are just [suitable] subsets of $\mathbf{S}$. It is still an open question, whether a subset of $\mathbf{S}$ should fulfill some regularity requirements ... in order to represent a physically verifiable property." [1]

But from our point of view the subsets of $\mathbf{S}$ which represent properties must be just the element of $\mathbf{S}_1 := \{\mathbf{S}_1(\hat{a}): \hat{a} \in \Pi\}$, i.e., subsets of ensembles for which a collection of effects exists which allow one to measure the involving property giving the answer "yes" with certainty for each ensembles of the subset. Along this direction, properties are introduced departing "from the properties of statistical ensemble. Statistical ensembles are, in a way, macroscopic entities: *though it might be impossible to predict the behavior of a single micro-individual in a given physical situa-*

tion, one can predict the behavior of the ensemble as a whole.

Therefore, there is no difficulty in defining the physical properties of the ensembles.

By saying that a certain ensemble (i.e., x) has a certain property (i.e., $\hat{a}$) we simply have in mind that the ensemble behaves in a specified way in some definite physical circumstances (i.e., $P(f, x) = 1$ for every $f \in \hat{a}$).

... The main difficulty with the single individual in a statistical theory lies in the fact that there is no immediate correspondence between the properties of the ensembles and the properties of the individuals. In fact not every property is of such nature that it may be attributed to each single ensemble individual." [1]

From this discussion it is reasonable to introduce the following:

Definition 5.3: A property $\hat{a}$ of a statistical ensemble is a proper starting point for a definition of a certain property of the *single* micro-object of the ensemble x iff $P(f, x) = 1$ for all $f \in \hat{a}$.

For example, "one can have a beam of photons of which the average fraction 1/2 penetrates through a certain Nicol prism. However, it may be that the ability of penetrating through the prism with the probability 1/2 cannot be attributed to each *single* beam photon, for the beam is just a mixture of two types of photons one of which is certainly absorbed by the prism." [1] On the contrary, if the beam of photons penetrates the Nicol prism with average fraction 1, then the property of being transmitted can be attributed to *each single* photon of the beam.

In conclusion, "a proposition f of a physical system is said to be 'true' (for the ensemble x) and the corresponding property is said to be 'actual' iff when we should decide to perform the test proposed by f (once preparing the system according to the preparation procedure x) the expected answer 'yes' would come with certainty (i.e., $P(f, x) = 1$).

To exhibit an individual with a 'true' property we proceed for example as follows:

(1) We first prepare a collection of identical systems in a well-defined way (described by x),

(2) we make the test (described by f) on each element of the collection.

If we see by statistic that the probability of obtaining the answer 'yes' is 1 (i.e. $P(f, x) = 1$) then we claim that the one new system prepared in the same way has this property." [3]

On the set Π of all propositions of the system we can introduce the following binary relation:

$$\hat{a} \subseteq \hat{b} \quad \text{iff} \quad \mathbf{S}_1(\hat{a}) \subseteq \mathbf{S}_1(\hat{b}),$$

which is an order relation since it is easy to check that

$$\hat{a} \subseteq \hat{a} \qquad \text{(reflexive)}$$

$$\hat{a} \subseteq \hat{b} \quad \text{and} \quad \hat{b} \subseteq \hat{a} \quad \text{implies} \quad \hat{a} = \hat{b} \ \text{(antisymmetric)}$$

$$\hat{a} \subseteq \hat{b} \quad \text{and} \quad \hat{b} \subseteq \hat{c} \quad \text{implies} \quad \hat{a} \subseteq \hat{c}. \qquad \text{(transitive)}$$

This order relation is just the order relation defined as follows:

$$\hat{a} \subseteq \hat{b} \quad \text{iff} \quad \text{there exist } f \in \hat{a} \text{ and } g \in \hat{b} \text{ such that } \quad f \subseteq g.$$

Therefore $(\Pi, \varnothing, I, \subseteq)$ is a partial-ordered set bounded by the *absurd* proposition $\varnothing = [\mathbb{0}]_{(1)}$ and the *certain* proposition $I = [\mathbb{1}]_{(1)}$.

Notice that there is no "natural" procedure to introduce an orthocomplementation on Π starting from the orthocomplementation of **F** in spite of the fact that

$$f_1 \sim f_2 \quad \text{does not imply} \quad f_1' \sim f_2'. \tag{6.2}$$

In fact, the decomposition of **F** into equivalence classes from Π agree with the affirmation that "it is not a logical impossibility to imagine a hypothetical physical world where to every '*domain of micro-objects with a certain special property*' there would be many possible *complementing domains* corresponding to many possible ways of being *opposite* to that property." [1]

Quoting Mielnik, "a hypothetical sequence of such devices is represented below:

The devices $f_1, f_2, \ldots$ schematically presented in Fig. 1 choose the same domain of micro-objects on which the answer should be 'certainly yes' but [they have] various 'certainly no' domains $\mathbf{S}_0(f_1), \mathbf{S}_0(f_2), \ldots$. For each of these devices the verbal negation (yes $\longleftrightarrow$ no) could be easily performed leading to a sequence of devices $f_1', f_2', \ldots$ with different 'certainly yes' domains $\mathbf{S}_0(f_1), \mathbf{S}_0(f_2), \ldots$." [1]

Example 5.1: The devices A and B of Example 4.1 test the same property "the light is red," but the corresponding certainly no domains are different.

Example 5.2: In the pre-Hilbert state–effect structure we have that

$$\mathrm{Ker}(\mathbb{1} - \mathbb{0}) = \{\underset{\sim}{0}\}, \quad \mathrm{Ker}(\mathbb{0}) = \mathbf{S},$$

$$\mathrm{Ker}(\mathbb{1} - \tfrac{1}{2}\,\mathbb{1}) = \{\underset{\sim}{0}\}, \quad \mathrm{Ker}(\tfrac{1}{2}\,\mathbb{1}) = \{\underset{\sim}{0}\}$$

and so $\mathbf{S}_1(\mathbb{0}) = \mathbf{S}_1(\frac{1}{2}\,\mathbb{1}) \equiv \{\underset{\sim}{0}\}$ but $\mathbf{S}_0(\mathbb{0}) \neq \mathbf{S}_0(\frac{1}{2}\,\mathbb{1})$.

Of course, we agree with Mielnik's conclusion that "contrary to axiom M.1 section 1, the devices $f_1, f_2, \ldots$ would be physically different, and even if we tried to neglect the difference by insisting that (id) section 1 defines the right physical equivalence, the negatives $f_1', f_2', \ldots$ could no longer be identified on the same principle." [1]

However, from our point of view, these difficulties are resolved inside the more general logic of the effects of a question–state structure. No additional axiom is required on the

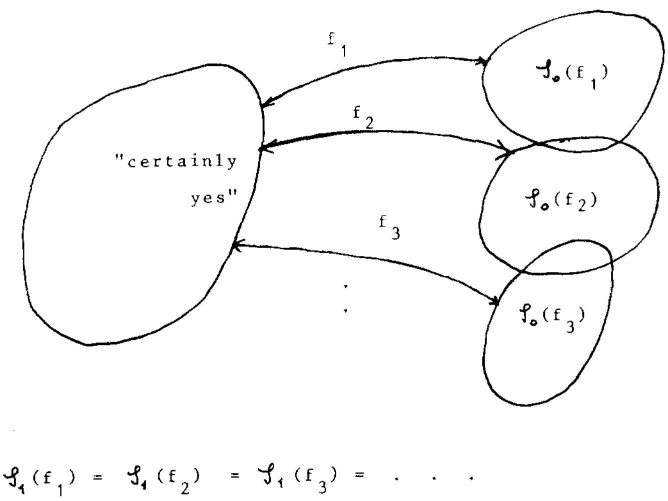


FIG. 1.

set of states, in particular that it is a convex structure, according to the consideration that *it is not a logical impossibility to imagine a hypothetical physical world where the mixtures of ensembles of physical systems are not always possible.*

We conclude this section underlining the following interesting result:

Theorem: If the set of all effects satisfies Axiom JP (Axiom c-JP), then the poset $(\Pi,\varnothing, I, \subseteq)$ of all propositions is a lattice (complete lattice).

More exactly, for any family $\{\hat{a}_i: i \in I\}$ of propositions, whose meet and join are written respectively as $\cap\, \hat{a}_i$, and $\cup\, \hat{a}_i$, we have that

$$\cap\, \hat{a}_i = [\Pi f_i]_{(1)},$$

where f_i is any representative of the equivalence class $\hat{a}_i$. Moreover,

$$\cup\, \hat{a}_i = \cap \{\hat{b} \in \Pi: \hat{a}_i \subseteq \hat{b}\}.$$

In particular, if Axiom JP holds, then $\mathbf{S}_1(\hat{a} \cap \hat{b}) = \mathbf{S}_1(\hat{a}) \cap \mathbf{S}_1(\hat{b})$, whereas in general we have $\mathbf{S}_1(\hat{a} \cup \hat{b}) \subseteq \mathbf{S}_1(\hat{a}) \cup \mathbf{S}_1(\hat{b})$, since the following theorem due to Piron holds:

Theorem: If $\mathbf{S}_1(\hat{a} \cup \hat{b}) = \mathbf{S}_1(\hat{a}) \cup \mathbf{S}_1(\hat{b})$ for every proposition $\hat{a}, \hat{b} \in \Pi$, then the lattice Π is distributive.

Yes–no experiments and JP propositions

Let $(\mathbf{Y}, \alpha(\mathbb{0}), \alpha(\mathbb{1}), \subseteq, ^{v})$ be the set of all yes–no experiments from a state–effect structure, where we have set the following:

(1) Let $\alpha = (A_1, A_0), \beta = (B_1, B_0)$, then $\alpha \subseteq \beta$ iff $A_1 \subseteq B_1$:

the yes–no experiment α stands in the relation $\alpha \subseteq \beta$ with respect to the yes–no experiment β iff whenever α is "true" then β is "true."

(2) Let $\alpha = (A_1, A_0)$, then $\alpha^{v}(A_0, A_1)$:

"If α is a yes–no experiment, then there exists another one α^{v}, measured with the same physical equipment, and such that if the outcome of the measurement α is 'yes' then it is 'no' for α^{v} and vice-versa. ...

D_6 (*equivalent yes–no experiments*): If two yes–no experiments α and β satisfy the relations $\alpha \subseteq \beta$ and $\beta \subseteq \alpha$, then we shall say that they are equivalent and we write for it $\alpha \sim \beta$. This is an equivalence relation ... .

D_9 (*JP proposition*): Let α be any yes–no experiment. The set of all yes–no experiments which are equivalent to it will be denoted by $\tilde{a} = \{\alpha\}$ and we call it a JP proposition. Thus more explicitly in a formula: $\tilde{a} = \{\alpha_i: \alpha_i \sim \alpha\} = \{\alpha\}$.

D_{10} (*true proposition*): It is easy to verify that if α is 'true' (in the sense of CD_3), then any $\alpha_i \sim \alpha$ is also 'true.' Hence we can say the JP proposition $\tilde{a}$ is 'true' if and only if any and hence all of its yes–no experiments are 'true.' If the JP proposition $\tilde{a}$ is 'true' we shall call it *a property* of the system." [15]

CD_5: $[\alpha \sim \beta]$

$$\text{iff}\quad (\forall x)[(\forall f)\, p_\alpha(x,f) \equiv (\forall g)\, p_\beta(x,g)]$$

$$\text{iff}\quad [(\forall f \ni \alpha = \alpha(f))(\forall g \ni \beta = \beta(g)), f \sim g].$$

D_{14} (*trivial propositions*): The equivalence class of the absurd yes–no experiment $\alpha(\mathbb{0}) = (\varnothing, \mathbf{S})$ defines the *absurd* proposition $\varnothing = \{(\varnothing, A): A \subseteq \mathbf{S}\}$ and the equivalence class of the certain yes–no experiment $\alpha(\mathbb{1}) = (\mathbf{S}, \varnothing)$ defines the *certain* proposition $I = \{(\mathbf{S},\varnothing)\}$.

D_{11} (*order relation between propositions*): If one has that for all $\alpha \in \tilde{a}$ and $\beta \in \tilde{b}$, $\alpha \subseteq \beta$, then the proposition $\tilde{a}$ is "stronger" than the proposition $\tilde{b}$, which is written $\tilde{a} \subseteq \tilde{b}$.[28]

Thus we can state the following:

Theorem T_1: The set of all JP propositions $(\mathbf{L},\varnothing,I,\subseteq)$ is a poset bounded by the least element $\varnothing$ and the greatest element I. Moreover, if $\mathbf{F}$ satisfies Axiom JP (Axiom c-JP), then $\mathbf{L}$ is a lattice (complete lattice). In particular, we have that

D_{12} ("*product*" *or* "*conjunction*" *of propositions*): Given any family of propositions $\{\tilde{b}_i\}$ from $\mathbf{L}$, $\cap\, \tilde{b}_i$ denotes the equivalence class containing the yes–no experiment $\Pi\beta_i$, where $\beta_i \in \tilde{b}_i$.

D_{13} ("*sum*" *of propositions*): Given any family $\{\tilde{b}_i\}$ of propositions from $\mathbf{L}$, $\cup\, \tilde{b}_i$ denotes the product $\cap\, \tilde{x}_\alpha$ of all the propositions $\tilde{x}_\alpha \in \mathbf{L}$ such that $\tilde{b}_i \subseteq \tilde{x}_{\alpha'}$, $\forall i$.[28]

Of course, using the set of "connection definitions" (CD), the remark that the operation of taking the orthocomplemented element f' of an effect f does not conserve the equivalence relation $\sim$ between effects [see Eq. (6.2)] is nothing else than the following:

Theorem T_2: The transposition CD_2—inside the predicate calculus—of Piron's operation of taking the inverse α^{v} of a yes–no experiment α does not conserve the transposition CD_5—inside the predicate calculus—of Piron's equivalence relation between yes–no experiments.[29]

From D_5 it obviously follows that

$$\begin{array}{ccc} \Pi & \rightarrow & \mathbf{L} \\ [f]_{(1)} & \rightarrow & \{\alpha(f)\} \end{array}$$

is a one-to-one and onto mapping for which the following statements hold:

$$\begin{array}{ccc} [\mathbb{0}]_{(1)} & \rightarrow & \{\alpha(\mathbb{0})\}, \\ [\mathbb{1}]_{(1)} & \rightarrow & \{\alpha(\mathbb{1})\}, \\ [f]_{(1)} \subseteq [g]_{(1)} & \text{iff} & \{\alpha(f)\} \subseteq \{\alpha(g)\}. \end{array}$$

Hence, the partial-order structures of all propositions and of all JP propositions are identifiable:

$$(\Pi,\varnothing, I, \subseteq) \equiv (\mathbf{L},\varnothing, I, \subseteq).$$

7. EVENTS

Once the notion of property is introduced, concretely measured by a set of different effects, we can single out the *events* of the usual approach to axiomatic quantum mechanics according to the following:

Definition 7.1: An element $\hat{a} \in \Pi$ is said to be a *real property* iff there exists an effect $a \in \mathbf{F}$ such that:

(i) $\mathbf{S}_1(a) = \mathbf{S}_1(\tilde{a})$,

(ii) $a \leqslant f$ for every $f \in \hat{a}$,

(iii) $a' \leqslant g$ for every $g \in [a']_{(1)}$.

If this is the case, the effect a is called the (exact) *event* associated with the property $\hat{a}$. The elements of Π which are not properties are then the very *fuzzy properties*. The set of all events in the following will be denoted by **E**. Condition (iii), Definition 7.1, assures us that the following statements are equivalent:

(1) a is an event corresponding to the property $[a]_{(1)}$ with certainly yes domain $\mathbf{S}_1(a)$;

(2) a' is an event corresponding to the property $[a']_{(1)}$ with certainly yes domain $\mathbf{S}_0(a)$.

From the previous definition we get that an event is a particular effect $a \in \mathbf{F}$ for which the following two conditions hold:

(e1) If there exists an effect $f \in \mathbf{F}$ such that $\mathbf{S}_1(a) = \mathbf{S}_1(f)$, then $a \leqslant f$,

(e2) if there exists an effect $g \in \mathbf{F}$ such that $\mathbf{S}_0(a) = \mathbf{S}_1(g)$, then $a' \leqslant g$.

Of course, (e1) and (e2) are the required conditions which allow one to *distinguish* the subclass of those macroscopic devices which more precisely correspond to events rather than to abstract "effects" and, moreover, give us some precise rules to select among all effects from **F** those which are accepted to represent propositions.

The distinction of the set of all events allows one also to single out the collection of all properties which can be tested starting from the structure $(\mathbf{F}, \mathbf{S}, P)$. Indeed to any event a is associated the property $\hat{a}$ with corresponding "certainly yes" $\mathbf{S}_1(a) = \mathbf{S}_1(\hat{a})$. The other elements of $\hat{a}$ can be considered as *fuzzy* representations of the unique exact event a. In a certain sense, by considering the event a as the effect which represents (or as the device which measures) the property $\hat{a}$ we eliminate from $\hat{a}$ the arrangements that present noise and imprecision.

Hence, "each event a determines a certain specific property of micro-objects (i.e., $\hat{a}$): the objects having the property are those for which the answer 'yes' is certain (i.e., $\mathbf{S}_1(a)$). Now, for each domain of micro-objects which possess a certain *property* there is a unique complementing domain of micro-objects with an *opposite property* (i.e., $\mathbf{S}_0(a)$): so that, once it is known for which objects the answer of the 'yes–no measurement' is 'certainly yes' it is also uniquely determined for which ones it should be 'certainly no.' "[1]

Moreover, if a is an event, we have then

$$\cup \{\mathbf{S}_0(f) : f \in \hat{a}\} = \mathbf{S}_0(a) \tag{7.1}$$

so that "the macroscopic yes–no measurement device (event) apart from possessing a nontrivial certainty domain must also have the property of minimizing the randomness of the 'yes' and 'no' answers.

A generalized version of this idea [is] that

(a) for a given 'certainly yes' domain, the 'yes–no measurement' ('event') should have a maximal possible 'certainly no' domain."[1]

Example 7.1: Taking into account the structure of light filters discussed in Example 4.1, we have that the macroscopic devices A and B have the same "certainly yes" domain, consisting of photon ensembles of red light, and are mutually opposed since they have *different* "certainly no" domains. They individuate the property: "the light is red" and the exact event associated with this property is schematically presented in the figure:

$$\begin{array}{ccc} \mathbf{S} & & R \\ red & \xrightarrow{1} & red \\ yel & \rightarrow & 0 \\ viol & \rightarrow & 0 \end{array} \qquad\qquad \begin{array}{ccc} \mathbf{S} & & R' \\ red & \rightarrow & 0 \\ yel & \xrightarrow{1} & red \\ viol & \nearrow^{1} & \end{array}$$

R is then the exact event: "the light is red" whereas A and B can be considered as "fuzzy" representations of this event. Indeed, R is characterized by the same "certainly yes" domain of A and B, but it has the maximal possible "certainly no" domain. The orthocomplement R' of the event R is the event: "the light is either yellow or violet" according to its "certainly yes" domain.

In this context, the semitransparent mirror has the behavior

$$\begin{array}{ccc} \mathbf{S} & & ST \\ red & \xrightarrow{1/2} & red \\ yel & \xrightarrow{1/2} & yel \\ viol & \xrightarrow{1/2} & viol \end{array}$$

and then $\mathbf{S}_1(ST) = \mathbf{S}_1(\mathbb{O}) = \varnothing$, e.g., the semitransparent mirror is a deeply "fuzzy" representation of the exact "impossible" event $\mathbb{O}$.

The previous considerations do not exclude that there could be some potential properties characterized by a well-defined certainly yes domain and in this case "there could exist many random minimalizing 'yes no measurements' with a common domain of 'certainly yes' and different 'certainly no' domains."[1]

An element $f \in \mathbf{F}$ is such a random minimalizing effect iff

$$f \in \hat{a} \quad \text{and} \quad (g \in \hat{a} \quad \text{with} \quad g \leqslant f \quad \text{implies} \quad g = f). \tag{7.2}$$

In this case, setting $(f)_{\leqslant} := \{h \in \hat{a} : f \leqslant h\}$ we have obviously

$$\mathbf{S}_1(f) = \mathbf{S}_1(h) \quad \text{for every} \quad h \in (f)_{\leqslant}, \tag{7.3}$$

$$\cup \{\mathbf{S}_0(h) : h \in (f)_{\leqslant}\} = \mathbf{S}_0(f). \tag{7.4}$$

Let us stress that the interactions between the preparing part and the observing part of an elementary experiment give us the only information about the system which can enable us to make physics. "Of course, the information about the (ensembles) depends essentially on the devices used, and vice versa the information about the devices depends essentially on the (ensembles) we have at our disposal. So everything we know is to a large extent relative"[26] to the triple state–effects.

For instance, if in the example of red, yel, viol photons, which constitute the set **S**, we consider the set of effects $\mathbf{F} = \{0, 1, A, A', B, B', RF, RF', G, G'\}$, where RF and G are schematically shown in the figures

$$
\begin{tikzcd}
\mathbf{S} & RF & \mathbf{S} & G \\
red \arrow[r, "1"] & red & red \arrow[r, "1/2"] & red \\
yel \arrow[ur, "1/4"] & & yel \arrow[ur, "1/4"] & \\
viol \arrow[r] & 0 & viol \arrow[uur, "1"'] &
\end{tikzcd}
$$

and no other effect is considered, then the property "the light is red" is a potential property since RF satisfies conditions (i) and (ii) but not (iii). Indeed, owing to the behavior of the effect

$$
\begin{tikzcd}
\mathbf{S} & RF' \\
red \arrow[r] & 0 \\
yel \arrow[r, "3/4"] & red \\
viol \arrow[ur, "1"'] &
\end{tikzcd}
$$

we have that $G \in [RF']_{(1)}$ but $P(RF',red) \leqslant P(G,red)$ and $P(G,yel) \leqslant (RF',yel)$ and so $RF' \nleqslant G$.

On the contrary, if we consider the set of all effects $\mathbf{F}_1 = \{0, 1, A, A', B, B', R, R', RF, RF', G, G'\}$, then in the triple $(\mathbf{F}_1, \mathbf{S}, P)$ the property "the light is red" can be tested by the effect R which is the event corresponding to this property.

8. THE LOGIC OF EVENTS

The set $\mathbf{E}$ of all events which can be detected in a state–effect structure $(\mathbf{F},\mathbf{S},P)$ is the natural framework of the usual logic approach to axiomatic quantum mechanics since we have that $\mathbf{E}$ satisfies:

(*a*) *The identity condition*: If $a, b \in \mathbf{E}$ and $\mathbf{S}_1(a) = \mathbf{S}_1(b)$, then $a = b$.

Indeed from (i) Definition 7.1 we have that $a \leqslant b$ and $b \leqslant a$ from which $a = b$ follows:

(*b*) *The uniqueness of nondegenerate orthocomplementation*.

Indeed, if $a \in \mathbf{E}$ with $a \leqslant a'$, then $P(a, x) \leqslant \frac{1}{2}$ for every $x \in \mathbf{S}$ that is $a \in [\mathbb{0}]_{(1)}$ so that $a = \mathbb{0}$, concluding that $\mathrm{Ker}'(\mathbf{E}) = \{\mathbb{0}\}$. From this last result it follows that $(\mathbf{E}, \mathbb{0}, \mathbb{1}, \leqslant, ')$ is an orthocomplemented poset, since the orthocomplementation $' \; \mathbf{E} \to \mathbf{E}$, $a \to a'$ satisfies the properties:

(oc1) $a'' = a$ for every $a \in \mathbf{E}$,

(oc2) $a \leqslant b$ implies $b' \leqslant a'$,

(oc3) $a \wedge a' = \mathbb{0}$ and $a \vee a' = \mathbb{1}$, for every $a \in \mathbf{E}$.

In conclusion, starting from the empirically well-founded state–question structure $(\mathbf{Q},\mathbf{S},P)$ we have singled out the logic $(\mathbf{F}, \mathbb{0}, \mathbb{1}, \leqslant, ')$ of all effects, for which the *indistinguishability principle* holds and the orthocomplementation is *degenerate*, and the logic $(\mathbf{E}, \mathbb{0}, \mathbb{1}, \leqslant, ')$ of all events, for which the *identity axiom* holds and the orthocomplementation is *nondegenerate*.

The usual logic approach to quantum mechanics refers to the poset $\mathbf{E}$ and the counterexamples presented by Mielnik do not scratch this approach for they pertain to the "degenerate" logic $\mathbf{F}$ rather than to the logic $\mathbf{E}$. This according to the interpretative rule for which the elements from $\mathbf{F}/\mathbf{E}$ are regarded as "fuzzy" representations either of an exact property from $\mathbf{E}$ or of a fuzzy property.

The orthogonality relation induced from this orthocomplementation, i.e., $a \perp b$ iff $a \leqslant b'$ iff $b \leqslant a'$, is then nondegenerate, i.e., it satisfies the following conditions:

(og1) $a \perp b$ implies $b \perp a$, (symmetric)

(og2) $\mathbb{0} \perp a$ for every $a \in \mathbf{E}$, (0-orthogonality)

(og3) $a \perp a$ implies $a = \mathbb{0}$, (nondegenerate)

(og4) $a_0 \leqslant a$ and $a \perp b$ imply $a_0 \perp b$, (absorption)

(og5) $a' = \max\{a\}^{\perp}$. (complete)

Hence, this orthogonality relation is not irreflexive and so is *not* a Foulis–Randall orthogonality relation. At any rate, if we consider the restriction of $\perp$ to the subset of events $\mathbf{E}_0 := \mathbf{E}/\{\mathbb{0}\}$, then it turns out to be a Foulis–Randall orthogonality.

A. Pool's approach and Axiom C

Let us remark that while the triple $(\mathbf{F},\mathbf{S},P)$ satisfies the Gunson conditions in order to have a structure which describes states and effects of quantum mechanics (see Sec. 4), the triple $(\mathbf{E},\mathbf{S},P)$ satisfies the most relevant axioms introduced by Pool to define an event–state structure.

To be precise, we have that:

(i) $\mathbf{E}$ is a set called the *logic* of the event–state structure and an element of $\mathbf{E}$ is called an *event*.

(ii) $\mathbf{E}$ is a set and an element of $\mathbf{S}$ is called a *state*.

(iii) P is a function P: $\mathbf{E} \times \mathbf{S} \to [0,1]$ called the *probability function* and if $a \in \mathbf{E}$ and $x \in \mathbf{S}$, then $P(a, x)$ is called the probability of occurrence of the event a in the state x.

(iv) If $a \in \mathbf{E}$, then subsets $\mathbf{S}_1(a)$ and $\mathbf{S}_0(a)$ are defined by

$$\mathbf{S}_1(a) = \{x \in \mathbf{S}: \; P(a, x) = 1\},$$

$$\mathbf{S}_0(a) = \{x \in \mathbf{S}: \; P(a, x) = 0\}.$$

(...)

(v) And the following conditions are satisfied (in the square brackets we quote the corresponding Pool numeration):

Axiom P.1 [Axiom I.1]: If $a, b \in \mathbf{E}$ and $\mathbf{S}_1(a) = \mathbf{S}_1(b)$, then $a = b$.

Axiom P.2 [Axiom I.6]: If $x_1, x_2 \in \mathbf{S}$ and $P(a, x_1) = P(a, x_2)$ for all $a \in \mathbf{E}$, then $x_1 = x_2$.

Axiom P.3 [Axiom I.2]: There exists an event $\mathbb{1} \in \mathbf{E}$ such that $\mathbf{S}_1(\mathbb{1}) = \mathbf{S}$.

Axiom P.4 [Axiom I.4]: If $a \in \mathbf{E}$, then there exists an event $a' \in \mathbf{E}$ such that

$$\mathbf{S}_1(a') = \mathbf{S}_0(a) \quad \text{and} \quad \mathbf{S}_0(a') = \mathbf{S}_1(a).$$

In spite of P.1, the relation of implication

(pi) $a, b \in \mathbf{E}$, $a \subseteq b$ iff $\mathbf{S}_1(a) \subseteq \mathbf{S}_1(b)$

is an order relation on $\mathbf{E}$ which corresponds to the phenomenological concept of implication more closely than the relation (or) on $\mathbf{E}$."[5]

Of course, it is always true that the (or) relation $a \leqslant b$ implies the (pi) relation $a \subseteq b$, but in general there is no reason to suppose that condition $a \subseteq b$ implies $a \leqslant b$.

Conditions P.1 and P.3 assert the existence of an unique event, the *certain* event, $\mathbb{1} \in \mathbf{E}$, which is the greatest element of $\mathbf{E}$ with respect to $\subseteq$ since

$$a \subseteq \mathbb{1} \quad \text{for all} \quad a \in \mathbf{E}.$$

From Conditions P.1, P.3, and P.4 we deduce the existence of the *impossible* event, $\mathbb{0} \in \mathbf{E}$, where $\mathbb{0} = \mathbb{1}'$, which is the least element of $\mathbf{E}$ with respect to $\subseteq$ since

$$\mathbb{0} \subseteq a \quad \text{for all} \quad a \in \mathbf{E}.$$

Notice that $\mathbf{S}_1(\mathbb{0}) = \varnothing$ and $\mathbf{S}_0(\mathbb{0}) = \mathbf{S}$.

If we denote by $a \cap b$ and $a \cup b$, if they exist, the glb and the lub of $\{a, b\}$ with respect to $\subseteq$, respectively, Axiom P.4 associates with any event $a \in \mathbf{E}$ the unique event $a' \in \mathbf{E}$, the *negation* of a, with the properties:

$$a = a'' \quad \text{for all} \quad a \in \mathbf{E}, \tag{8.1}$$

$$a \cap a' = \mathbb{0} \quad \text{for all} \quad a \in \mathbf{E}. \tag{8.2a}$$

Hence, the mapping $a \to a'$ is not an orthocomplementation since $a \subseteq b$ does not imply $b' \subseteq a'$ and in general $a \cup a' = \mathbb{1}$ does not hold.

No other result can be derived from a state–effect structure in a natural way. At any rate, if a state–event structure satisfies the following,

Axiom P.5: For any pair of events $a, b \in \mathbf{E}$, the condition $\mathbf{S}_1(a) \cup \mathbf{S}_0(a) \subseteq \mathbf{S}_1(b)$ implies $\mathbf{S}_0(b) = \varnothing$ (and so $b = \mathbb{1}$ or equivalently $\mathbf{S}_1(b) = \mathbf{S}$),

then we have that

$$a \cup a' = \mathbb{1} \quad \text{for all} \quad a \in \mathbf{E}. \tag{8.2b}$$

The event c is said to be a *complementary* event for a given event b iff $b \cup c = \mathbb{1}$ and $b \cap c = \mathbb{0}$. If this is the case, the events b and c are said to be *complements* of one another. So from (8.2a) and (8.2b) it follows that the event a' is a complement of a, which in general is not unique.

However, let us also remark that if Axiom P.5 holds, then $a \to a'$ is not an orthocomplementation. In the Pool approach this result is assured by the further introduction of an *ad hoc* axiom which is a stronger version of Axiom P.5: to be precise

Axiom P.5.S [Axiom I.3]: If $a, b \in \mathbf{E}$ and $\mathbf{S}_1(a) \subseteq \mathbf{S}_1(b)$, then $\mathbf{S}_0(b) \subseteq \mathbf{S}_0(a)$.

In conclusion, beside the orthocomplemented poset of all events $(\mathbf{E}, \mathbb{0}, \mathbb{1}, \leqslant, ')$, where $\leqslant$ is the Mackey (or) order relation, if a state–event structure satisfies Axiom P.5 (resp. P.5.S.) we can consider the complemented (resp. orthocomplemented) poset $(\mathbf{E}, \mathbb{0}, \mathbb{1}, \subseteq, ')$ where $\subseteq$ is the (pi) order relation also considered by Pool. At any rate, also in case Axiom P.5.S holds, the two order relations in general do not coincide.

Proposition: Once associated with any event $a \in \mathbf{E}$, the yes–no experiment $\alpha(a) = (\mathbf{S}_1(a), \mathbf{S}_0(a))$ and the property $\tilde{a} = \{\alpha(a)\} \equiv [a]_{(1)}$, if Axiom P.5 holds we have that:

(1) $\alpha(a) \in \tilde{a}$,

(2) $\beta = (B_1, B_0)$, $\beta \sim \alpha(a)$ implies $B_0 \subseteq \mathbf{S}_0(a)$,

(3) $\gamma = (C_1, C_0)$, $\gamma \sim \alpha(a)^{\vee}$ implies $C_0 \subseteq \mathbf{S}_1(a)$,

(4) let $\delta = (D_1, D_0)$ be such that if $\mathbf{S}_1(a) \cup \mathbf{S}_0(a) \subseteq D_1$ then $D_0 = \varnothing$.

Let us introduce the set of all:

exact yes–no experiments $\mathbf{k}(\mathbf{E})$:
$= \{\alpha \in \mathbf{Y}: \exists\, a \in \mathbf{E} \ni \alpha = \alpha(a)\}$,

exact JP propositions $\mathbf{L}(\mathbf{E})$:
$= \{\tilde{a} \in \mathbf{L}: \exists\, \alpha \in \mathbf{Y}(\mathbf{E}) \ni \alpha \in \tilde{a}\}$.

The structure $(\mathbf{L}(\mathbf{E}), \varnothing, I, \subseteq, ')$ is a bounded poset for which the following axiom introduced by Piron holds:

Axiom C: For any exact JP proposition $\tilde{a} \in \mathbf{L}(\mathbf{E})$ there exists at least one compatible complement $\tilde{a}' \in \mathbf{L}(\mathbf{E})$, i.e., an exact JP proposition $\tilde{a}'$ such that:

(1) $\tilde{a} \cup \tilde{a}' = I$ and $\tilde{a} \cap \tilde{a}' = \varnothing$,

(2) there exists an (exact) yes–no experiment $\alpha(a) \in \tilde{a}$ such that $\alpha(a)^{\vee} \in \tilde{a}'$,

The main difference between our approach and the Jauch and Piron one is that it is not true that any proposition is an exact proposition too, i.e., it is not true that $\mathbf{L}(\mathbf{E}) = \mathbf{L}$, which is the very formulation of Piron's Axiom C.

Definition 8.1: A state–effect structure is said to be *complete* iff

(1) Axiom c-JP holds,

(2) the underlying state–event structure satisfies the further conditions:

(2i) Axiom P.5.S is true,

(2ii) $\mathbf{L} = \mathbf{L}(\mathbf{E})$.

As a consequence of this definition, in the case of a complete state–effect structure the logicomathematical system of all JP propositions $\mathbf{L}$ is an orthocomplemented complete lattice, where

(pi) $\tilde{a} \subseteq \tilde{b}$ iff $\exists\, a, b \in \mathbf{E} \ni a \in \tilde{a}, b \in \tilde{b}$, and $a \subseteq b$,

(C) for any JP proposition $\tilde{a}$, realized by the event a, there corresponds the JP proposition $\tilde{a}'$, the negation of $\tilde{a}$, which is realized by the event a'.

Therefore, in the case of a complete state–effect structure the very "existence" of a compatible complement $\tilde{a}'$ for any $\tilde{a} \in \mathbf{L}$ is not at all questionable. Of course, the assumption that the compatible complement $\tilde{a}'$ of a JP proposition $\tilde{a}$ consists of the class of $\{\alpha^{\vee}\}$ of all negations $\alpha^{\vee}$ of the yes–no experiments $\alpha \in \tilde{a}$ is false.

To be more precise, it is false to assume that "if $\alpha \in \tilde{a}$ is a yes–no experiment in the class $\tilde{a}$, then $\tilde{a}'$ contains the yes–no experiment $\alpha^{\vee}$ (the strong negation of α) which is the same experiment of α but with its alternatives interchanged" [14] or that "every proposition has at least one compatible complement. This can be seen as follows: If $\tilde{a} \in \mathbf{L}$, let $\alpha \in \tilde{a}$ and let $\tilde{b}$ be the equivalence class containing $\alpha^{\vee}$, then $\tilde{b}$ is a compatible complement of $\tilde{a}$." [30]

On the contrary, if $\tilde{a}$ is represented by the exact yes–no experiment $\alpha = \alpha(a)$, where a is an exact event, then no other $\alpha^{\vee}$, where $\alpha \sim \alpha(a)$, except the only yes–no experiment $\alpha(a)^{\vee}$ belongs to $\tilde{a}'$ and so $\tilde{a}'$ is represented by the exact yes–no experiment $\alpha^{\vee} = \alpha(a')$.

"Thus contrary to a widely held opinion, the compatible complement a'—quite generally—cannot be formed as the class of the negations of the yes–no experiments. This, however, does not yet lead to doubts about the existence of a'. Indeed, so far there is still the possibility that for each given $a \in \mathbf{L}$ some method for constructing a nonvoid a' is specifiable, even if a' does not contain *all* the negations $\alpha^{\vee}$ of $\alpha \in a$." [28]

B. The pre-Hilbert space model

In the state–effect structure $(\mathbf{F}(\mathbf{k}),\mathbf{S}(\mathbf{k}), P)$ based on a pre-Hilbert space $\mathbf{k}$, for any effect $F \in \mathbf{F}(\mathbf{k})$ the "certainly yes" domain and the "certainly no" domain are identifiable with the subsets of $\mathbf{k}$:

$$\mathbf{S}_1(F) \equiv \{\psi \in k: F\psi = \psi\},$$

$$\mathbf{S}_0(F) \equiv \{\varphi \in \mathbf{k}: F\varphi = \underset{\sim}{0}\}.$$

The orthogonal projections from $\Pi(\mathbf{k})$ are particular effects, and for any $E \in \Pi(\mathbf{k})$ we have that

$$\mathbf{S}_1(E) \equiv \mathrm{Ran}(E),$$

$$\mathbf{S}_0(E) \equiv \mathrm{Ker}(E).$$

So that the "certainly yes" domain and the "certainly no" domain of E are mutually orthogonal orthosubspaces of $\mathbf{k}$, i.e.,

$$\mathrm{Ran}(E) = \mathrm{Ker}(E)^{\perp} \quad \text{and} \quad \mathrm{Ker}(E) = \mathrm{Ran}(E)^{\perp},$$

$$\mathbf{S} = \mathrm{Ran}(E) \oplus \mathrm{Ker}(E).$$

Let $E \in \Pi(\mathbf{k})$ be an orthogonal projection and $\hat{E}$ be the property associated with E with "certainly yes" domain $\mathbf{S}_1(E) = \mathrm{Ran}(E)$. If E_1 is another orthogonal projection belonging to $\hat{E}$, then $E_1 = E$, i.e.,

(1) *In any property there exists at most one orthogonal projection.* Moreover,

(2) *Any orthogonal projection E is an event corresponding to the property $\hat{E}$ with certainly yes domain $\mathbf{S}_1(E) \equiv Ran(E)$.*

Proof: Let $F \in \mathbf{F}(\mathbf{k})$ be such that $\mathbf{S}_1(F) = \mathbf{S}_1(\mathrm{E}) = \mathrm{Ran}(E)$; then for any $\psi \in \mathbf{K}$ we can set $\psi = \psi_1 + \psi_1^{\perp}$, where $\psi_1 \in \mathrm{Ker}(1-F) = \mathrm{Ran}(E)$ and $\psi_1^{\perp} \in \mathrm{Ker}(E)$ so that $\langle \psi_1 | \psi_1^{\perp} \rangle = 0$; thus

$$0 \leqslant \langle (F-E)\psi|\psi\rangle = \langle F\psi_1^{\perp}|\psi_1^{\perp}\rangle,$$

concluding that $E \leqslant F$ for every $F \in [E]_{(1)}$.

In the same manner the orthogonal projection $E' = 1 - E$, whose certainly yes domain is $\mathbf{S}_1(E') = \mathrm{Ker}(E)$, is such that $E' \leqslant G$ for every $G \in [E]_{(1)}$.

Therefore we have stated that every orthogonal projection is an event. On the other hand, we now prove that

(3) *In the state–effect structure based on a pre-Hilbert space each event is an orthogonal projection.*

Indeed, let $F \in \mathbf{E}(\mathbf{k})$, then we know that $0 \leqslant F^2 \leqslant F \leqslant 1$ and so $\mathbf{S}_1(F^2) \subseteq \mathbf{S}_1(F)$. Moreover, if $\psi \in \mathbf{S}_1(F)$, then $F\psi = \psi$, from which it follows $F^2\psi = F\psi = \psi$, i.e., $\psi \in \mathbf{S}_1(F^2)$, concluding that $F \sim F^2$. Therefore, $F = F^2$.

(IV) In a pre-Hilbert state–effect structure an effect is an exact event iff it is an orthogonal projection.

Since it is straightforward to prove that in a pre-Hilbert space the sum of a finite set of orthogonal projections is an orthogonal projection iff they are pairwise orthogonal, we can conclude that:

(4) *The set of all events from a pre-Hilbert state–effect structure satisfies the strong orthogonality axiom OG.S.*

Therefore, if we apply the Finch result[27] quoted in the subsection in Sec. 4 we have that

(V) The structure of all exact events from a pre-Hilbert state–effect structure, $(\mathbf{E}(\mathbf{k}),\mathbb{0},\mathbb{1},\leqslant,')$, *is an orthocomplemented orthomodular atomic orthoposet with covering property.*

Once an event $E \in \mathbf{E}(\mathbf{k})$ is singled out, any fuzzy representation F of this event is characterized by the property that

$$F \sim E \quad \text{iff} \quad \mathrm{Ker}(1-F) = \mathrm{Ran}(E).$$

(5) *Let $F \sim E$; then $F \neq E$ if $Ker(f) \subset Ker(E)$.*

(6) *The effects F for which $\|F\| \neq 1$ are all fuzzy representations of the absurd event $\mathbb{0}$.*

Remembering that

$$\|F\| = \sup\left\{\frac{\langle F\psi|\psi\rangle}{\|\psi\|^2} : \psi \in \mathbf{k} | \{\underset{\sim}{0}\}\right\},$$

the previous conclusion is a trivial consequence of the assumption $\|F\| \neq 1$.

(7) *In the pre-Hilbert state–effect structure, the physical order relation (or) and the phenomenological implication (pi) coincide on the set* $\mathbf{E}$ *of all events and so Axiom P.5.S holds.*

Indeed, it is easy to prove the following result:

Let $E_1, E_2 \in \mathbf{E}(k)$; then the following statements are equivalent:

(1) $\mathrm{Ran}(E_1) \subseteq \mathrm{Ran}(E_2)$;

(2) $\langle E_1\psi|\psi\rangle \leqslant \langle E_2\psi|\psi\rangle$ for every $\psi \in k$;

(3) $E_1 = E_1 E_2$;

(4) $E_1 = E_2 E_1$.

Lastly we have the following statement:

(8) *A Hilbert state–effect structure is complete.*

9. NOPERTIES AND NOVENTS

If we take into account the light filters of Example 6.1, we have that the certainly no domains of the exact events R and R' are, respectively,

$$\mathbf{S}_0(R) = \{\,yel,viol\,\} \quad \text{and} \quad \mathbf{S}_0(R') = \{red\},$$

and then the "no" property can be associated with the filter R: "the light is neither yellow nor violet" whereas to the filter R' corresponds the "no" property: "the light is not red."

According to this last remark, we introduce in a dual manner the following equivalence relation:

$$f_1 \equiv f_2 \quad \text{iff} \quad \mathbf{S}_0(f_1) = \mathbf{S}_0(f_2), \tag{1}$$

and we denote by $[f]_{(0)}$ the corresponding equivalence class generated by f, i.e.,

$$[f]_{(0)} := \{h \in \mathbf{F}: \mathbf{S}_0(h) = \mathbf{S}_0(f)\}.$$

For the trivial effects we have

$$[1]_{(0)} = \{h \in \mathbf{F}: P(f,x) \neq 0, \text{ for every } x \in \mathbf{S}\},$$

$$[0]_{(0)} = \{0\}.$$

Definition 9.1: An element $\tilde{a}$ of the quotient set $\mathbf{F}/\equiv$ is said to be a *noperty* iff there exists an effect $a \in \mathbf{F}$ such that

(i) $\mathbf{S}_0(a) = \mathbf{S}_0(\tilde{a})$,

(ii) $f \leqslant a$ for every $f \in \tilde{a}$,

(iii) $g \leqslant a'$ for every $g \in [a']_{(0)}$.

In this case a is called a *novent*.

Proposition 9.1: a is an event corresponding to the property $[a]_{(1)}$ with certainly yes domain $\mathbf{S}_1(a)$ iff a is a novent corresponding to the noperty $[a]_{(0)}$ with certainly no domain $\mathbf{S}_0(a)$.

Proof: Let a be an event. If $f \in [a]_{(0)}$, then $\mathbf{S}_1(f') = \mathbf{S}_0(f) = \text{bold } S_0(a) = \text{bold } S_1(a')$, i.e., $f' \in [a']_{(1)}$

and so $a' \leqslant f'$ follows from (ii) of Definition 9.1. Thus $f \leqslant a$ with $\mathbf{S}_0(f) = \mathbf{S}_0(a)$ concluding that $f \leqslant a$ for every $f \in [a]_{(0)}$. On the other hand, let $g \in [a']_{(1)}$; then $\mathbf{S}_1(g') = \mathbf{S}_0(g) = \mathbf{S}_0(a') = \mathbf{S}_1(a)$, i.e., $g' \in [a]_{(1)}$, and then $a \leqslant g'$ follows from (i) of Definition 9.1. Thus $g \leqslant a'$ with $\mathbf{S}_0(g) = \mathbf{S}_0(a')$, i.e., $g \in [a']_{(0)}$ for every $g \in [a']_{(0)}$.

The converse can be proved in the same manner.

From the previous result we have that $\mathbf{E}$ is also the collection of all novents.

Example: In the light filters structure, the filter R has the "certainly yes" domain {red} and then R is the event corresponding to the property "the light is red"; the "certainly no" domain of R is {yel,viol} and then R is the novent corresponding to the noperty "the light is neither yellow nor violet."

Analogously, the filter R' with "certainly yes" domain {yel,viol} is the vent corresponding to the property "the light is either yellow or violet"; since the R' "certainly no" domain is {red}, then R' is the novent corresponding to the noperty "the light is not red."

10. CONCLUSIONS

Generalizing the results obtained in the previous sections, particularly the behavior of the pre-Hilbert space model, we conclude with the proposal of the following definition to single out the peculiar state–effect structures which can be assumed to characterize quantum phenomenology.

Definition 10.1: A *quantum state–effect structure* is a triple $(\mathbf{F},\mathbf{S},P)$, where

(1) $\mathbf{S}$ is the nonempty set of all preparation procedures of ensembles,

(2) $\mathbf{F}$ is the nonempty set of all observation procedures of effects,

(3) P: $\mathbf{F}\times\mathbf{S} \to [0,1]$ is the probability function, which satisfies the following axioms:

Axiom G.1: $P(f_1, x) = P(f_2, x), \quad \forall\, x \in \mathbf{S}$ implies $f_1 = f_2$.

Axiom G.2: $P(f, x_1) = P(f, x_2), \quad \forall f \in \mathbf{F}$ implies $x_1 = x_2$.

Axiom G.3: There exists an effect, denoted by $\mathbb{0} \in \mathbf{F}$, such that $P(\mathbb{0}, x) = 0$ for all $x \in \mathbf{S}$.

Axiom G.4: For every effect $f \in \mathbf{F}$ there exists another effect $f' \in \mathbf{F}$ such that $P(f, x) + P(f', x) = 1$ for all $x \in \mathbf{S}$.

Axiom OG: For every finite sequence $\{f_1, f_2, \ldots, f_n\}$ of effects from $\mathbf{F}$ which satisfies the condition $0 \leqslant \Sigma_{i=1}^{n} P(f_i, x) \leqslant 1 \;\; \forall x \in \mathbf{S}$ there exists an effect $f \in \mathbf{F}$ such that $P(f, x) = \Sigma_{i=1}^{n} P(f_i, x) \;\; \forall x \in \mathbf{S}$.

Axiom JP: For every pair of effects f_1 and f_2 there exists an effect, denoted by $f_1 f_2$ and called the "product," such that

$$P(f_1 f_2, x) = \tfrac{1}{2}\,[P(f_1, x) + P(f_2, x)] \;\; \forall x \in \mathbf{S}.$$

Axiom E.OG.S: Once the state–event structure $(\mathbf{E},\mathbf{S},P)$ induced from the state–effect structure $(\mathbf{F},\mathbf{S},P)$ is singled out according to the procedure introduced in Secs. 7 and 8, to any finite sequence of pairwise orthogonal events $\{a_1, a_2, \ldots, a_n\}$ there corresponds an event $a \in \mathbf{E}$ such that $P(a, x) = \Sigma_{i=1}^{n} P(a_i, x), \forall x \in \mathbf{S}$.

The quantum phenomenology is then reflected by two "logical" structures:

(a) The logic of all effects: $(\mathbf{F},\mathbb{0},\mathbb{1},\leqslant,')$ which is a
   (i) bounded poset,
   (ii) with degenerate orthocomplementation,
   (iii) which satisfies the weak orthogonality axiom.

(b) The logic of all events $(\mathbf{E},\mathbf{S},P)$ which is a
   (i) bounded poset,
   (ii) with a real orthocomplementation,
   (iii) which satisfies the strong orthogonality axiom.

Therefore, the logic of all events turns out to be an orthocomplemented orthomodular orthoposet.

Summarizing the results obtained we can state the following:

The counterexamples discussed by Mielnik are effects which are not events and so they regard the quantum logic of effects whose orthocomplementation is degenerate.

The indistinguishability principle holds both for the logic of all effects and for the one of all events; this principle is the Mackey Axiom II or the Cooke–Hilgevoord Axiom 1. Moreover, in the logic of all events the identity axiom is true.

The logic of all effects satisfies the weak orthogonality axiom and not the strong one, which, on the contrary, holds in the logic of all events. This last is the Mackey Axiom V (or Cooke–Hilgevoord Axiom 2) and is peculiar to the logic of all events only. From our point of view, quantum logic must be practiced in such a way as to also take into account the logic of all effects with the corresponding order structure.

Of course, the sentence $(z_0 \in \{1,3\})$ considered in Ref. 17 is an effect, to be precise the effect $\frac{1}{2}\,\mathbb{1}$, which is not an event and so it is not at all surprising that the triple of effects $\{\frac{1}{2}\,\mathbb{1}, \frac{1}{2}\,\mathbb{1}, \frac{1}{2}\,\mathbb{1}\}$ does not satisfy Axiom E.OG.S (see subsection of Sec. 5).

Therefore, if we want to include the proposition $(z_0 \in \{1,3\})$ in a quantum logical scheme we must consider the logical structure of all effects. In this way, quantum logic completely reflects the full structure of quantum language since it contains those sentences and reproduces their evident logical relations.

[1]B. Mielnik, "Quantum logic: is it necessarily complemented?," in *Quantum Mechanics, Determinism, Causality and Particles*, edited by M. Flato *et al.* (Reidel, Dordrecht, The Netherlands, 1976).

[2]G. Ludwig, "The measuring process and the axiomatic foundations of quantum mechanics," in *Foundations of Quantum Mechanics*, edited by B. D'Espagnat (Academic, New York, 1971), p. 287.

[3]D. Aerts, "The one and the many," Ph.D. thesis, Vrije Universiteit, Brussels, 1980/81.

[4]C. Piron, "Survey of general quantum physics," Found. Phys. **2**, 287 (1972).

[5]J. C. T. Pool, "Baser ∗-semigroups and the logic of quantum mechanics," Comm. Math. Phys. **9**, 118 (1968).

[6]D. J. Foulis and C. H. Randall, "Operational statistics, I. Basic concepts," J. Math. Phys. **13**, 1667 (1972).

[7]C. H. Randall and D. J. Foulis, "Operational statistics, II. Manuals of operations and their logic," J. Math. Phys. **14**, 1472 (1973).

[8]D. J. Foulis and C. H. Randall, "The empirical logic approach to the physical sciences," in *Foundations of Quantum Mechanics and Ordered Linear Spaces*, edited by A. Hartkamper and H. Neumann (Springer-Verlag, Berlin, 1974).

[9]D. J. Foulis and C. H. Randall, "Empirical logic and quantum mechanics," Synthese **29**, 81 (1974).

[10]C. H. Randall and D. J. Foulis, "The operational approach to quantum mechanics," in *The Logic-Algebraic Approach to Quantum Mechanics, Vol. III*, edited by C. A. Hooker (Reidel, Dordrecht, The Netherlands, 1977).

[11]D. J. Foulis and C. H. Randall, "Empirical logic, operational statistics and tensor products," in *Interpretations and Foundations of Quantum Theory*, edited by H. Neumann (Wissenschaftsverlag, 1981).
[12]H. Neumann, "The structure of ordered Banach spaces in axiomatic quantum mechanics," in *Foundations of Quantum Mechanics and Ordered Linear Spaces*, edited by A. Hartkamper and H. Neumann (Springer-Verlag, Berlin, 1974).
[13]J. M. Jauch and C. Piron, "On the structure of quantal proposition systems," Helv. Phys. Acta **42**, 842 (1969).
[14]J. M. Jauch and C. Piron, "What is 'Quantum logic'?," in *Quanta*, edited by P. G. O. Freund, C. J. Goebel, and Y. Nambu (Chicago U.P.,Chicago, 1970).
[15]J. M. Jauch, "Foundations of quantum mechanics," in *Foundations of Quantum Mechanics*, edited by B. D'Espagnat (Academic, New York, 1971).
[16]C. Piron, "On the foundations of quantum physics," Ref. 1.
[17]R. M. Cooke and J. Hilgevoord, "A new approach to equivalence in quantum logic," in *Current Issues in Quantum Logic*, edited byE. G. Beltrametti and B. C. van Frassen (Plenum, New York, 1981).
[18]G. Ludwig, "Attempt of an axiomatic foundations of quantum mechanics and more general theories, I," guest lecture given at the University of Fribourg, 1964.
[19]G. W. Mackey, *The Mathematical Foundations of Quantum Mechanics* (Benjamin, New York, 1963).
[20]J. M. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley, Reading, MA, 1968).
[21]C. Piron, *Foundations of Quantum Physics* (Benjamin, Reading, MA, 1976).
[22]C. Garola, "Propositions and Orthocomplementation in Quantum Logic," Int. J. Theor. Phys. **19**, 369 (1980).
[23]G. Cattaneo and A. Mania, "Abstract orthogonality and orthocomplementation," Proc. Camb. Phil. Soc. **76**, 115 (1974).
[24]G. Cattaneo and G. Marino, "Weak orthocomplementation on posets, preclusivity structures, and preclusivity logics," preprint, Dept. Math. University of Calabria, Sept. 1983.
[25]J. Gunson, "On the algebraic structure of Quantum Mechanics," Comm. Math. Phys. **6**, 262 (1967).
[26]M. Kupczynski, "Is the Hilbert space language too rich?," Int. J. Theor. Phys. **10**, 297 (1974).
[27]P. Finch, "On the structure of quantum logic," J. Symb. Logic **34**, 275 (1969).
[28]N. Hadjivvas, F. Thieffine, and M. Mugur-Schacter, "Study of Piron's system of questions and propositions," Found. Phys. **10**, 751 (1980).
[29]N. Hadjivvas, F. Thieffine, and M. Mugur-Schacter, "Supplement to a critique of Piron's system of questions and propositions," Found. Phys. **11** 645 (1981).
[30]R. J. Greechie and S. P. Gudder, "Quantum logics," in *Contemporary Research in the Foundations and Philosophy of Quantum Theory*, edited by C. A. Hooker (Reidel, Dordrecht, The Netherlands, 1974).

Sharpening an inequality in quantum ergodic theory

Philip Pechukas
Department of Chemistry, Columbia University, New York, New York 10027

(Received 7 September 1983; accepted for publication 4 November 1983)

The fundamental inequality of quantum ergodic theory, which bounds the ensemble-averaged dispersion of a time-averaged occupation probability, is due to von Neumann and to Bocchieri and Loinger. Here this inequality is strengthened by combining the ensemble averaging of von Neumann and Bocchieri and Loinger. In this sharper form, the inequality says that the time average of an occupation probability is liable to be much closer to the statistical expectation than is its instantaneous value, the more so the less degenerate is the spectrum of the Hamiltonian.

PACS numbers: 03.65.Bz, 05.30.Ch, 02.50. + s

I. INTRODUCTION

When is a quantum system ergodic? The answer that comes first to mind is, when the spectrum of the Hamiltonian is nondegenerate, for then energy is the only constant of the motion, in the sense that any operator that commutes with the Hamiltonian can be expressed as a function of the Hamiltonian. Von Neumann[1] thought he had found a dynamical justification for this definition of quantum ergodicity in the following calculation: Let S be the N-dimensional "energy shell" spanned by N consecutive eigenvectors of H; let P be the projector onto an M-dimensional subspace of S; calculate the mean square deviation of the occupation probability $P(t) = (\psi(t), P\psi(t))$ from the statistical expectation $f = M/N$, where the mean is a time average,

$$\lim_{T\to\infty} T^{-1}\int_0^T dt\,(P(t)-f)^2 \equiv \overline{(P-f)^2}. \tag{1.1}$$

The result of course depends on the projector P. There being no reason to prefer one M-dimensional projector to another, von Neumann averaged over all P's, an operation we denote by $\langle\ \rangle_P$. He proved that if the energy eigenvalues in S are all different and if, in addition, the pairwise differences of these eigenvalues are all different, then the average relative dispersion $\langle (P-f)^2\rangle_P/f^2$ is small if M is large. Specifically, von Neumann showed that

$$\overline{\langle (P-f)^2\rangle}_P/f^2 < 2(1-f)/M. \tag{1.2}$$

The interpretation is that a typical "macro-observer," i.e., a typical projector P of large dimension M, will rarely find substantial deviation from the statistical expectation f in a system evolving under von Neumann's two spectral conditions.

The von Neumann result has not aged well. First Fierz[2] and ter Haar[3] showed that the assumption on pairwise differences of energy eigenvalues is unnecessary to establish the inequality (1.2); then Farquhar and Landsberg[4] showed that the assumption of eigenvalue nondegeneracy is likewise unnecessary; finally Bocchieri and Loinger[5] showed that time averaging—or, indeed, any dynamics whatsoever—is also inessential. The Hamiltonian may be identically equal to zero in the energy shell S and inequality (1.2) will still hold: von Neumann's result is entirely a consequence of the averaging over "macro-observers" and has nothing to do with quantum dynamics.

It is ironic that in the course of demolishing inequality (1.2) as a physical statement Bocchieri and Loinger strengthened it mathematically: They showed that for any normalized $\psi \in S$

$$\langle (\psi, P\psi)^2\rangle_P = M(M+1)/N(N+1), \tag{1.3}$$

a result that we shall use below and which implies inequality (1.2) in the sharper form

$$\overline{\langle (P-f)^2\rangle}_P/f^2 < (1-f)/M. \tag{1.4}$$

Subsequently Bocchieri and Loinger,[6] who objected to von Neumann's averaging over "macro-observers," suggested that one should instead average over all initial states belonging to S, an operation we denote by $\langle\ \rangle_\psi$. As they pointed out, there is little difference mathematically between averaging over P for fixed ψ and averaging over ψ for fixed P; it is a question of interpretation. For example, Eq. (1.3) still holds in the form

$$\langle (\psi, P\psi)^2\rangle_\psi = M(M+1)/N(N+1), \tag{1.5}$$

but the interpretation now is that, given P with $M \gg 1$, for most ψ the probability $(\psi, P\psi)$ does not differ much from the statistical expectation $f = M/N$. Inequality (1.4) still holds when the average over P for fixed initial ψ is replaced by the average over ψ for fixed P, and it is just as devoid of physical content.

It may seem doubly futile to combine the averaging procedures of von Neumann and of Bocchieri and Loinger, which is what is done in this paper, but we shall use these procedures for a slightly different calculation. By analogy with classical ergodic theory, the time average of real interest is not that of Eq. (1.1) but rather the time-averaged occupation probability

$$\lim_{T\to\infty} T^{-1}\int_0^T dt\, P(t) \equiv \bar P_\psi, \tag{1.6}$$

which of course depends on both the projector P and the initial state ψ. Since $\bar P_\psi^2 \leqslant \overline{P(t)^2}$, the dispersion $\langle (\bar P_\psi - f)^2\rangle_P$ or $\langle (\bar P_\psi - f)^2\rangle_\psi$ is subject to the same inequality (1.4), as Bocchieri and Loinger pointed out[5,6]:

$$\langle (\bar P_\psi - f)^2\rangle_P/f^2 < (1-f)/M, \tag{1.7a}$$

$$\langle (\bar P_\psi - f)^2\rangle_\psi/f^2 < (1-f)/M. \tag{1.7b}$$

We get a sharper result by calculating $\langle\langle (\bar P_\psi - f)^2\rangle_\psi\rangle_P$. We

 0022-2488/84/030532-03$02.50

find that this doubly averaged dispersion in the time-averaged occupation probability is smallest when the energy spectrum in S is nondegenerate and largest when the spectrum is completely degenerate. The interpretation is that a "typical" state subjected to a "typical" observation behaves "more ergodically" the less degenerate is the spectrum of its Hamiltonian.

Section II explains the calculation of various averages needed in Sec. III for the evaluation of $\langle\langle(\bar{P}_\psi - f)^2\rangle_\psi\rangle_P$. Section IV briefly discusses the significance—or lack thereof—of the result.

II. AVERAGES

Let $\{\phi_j\}, j = 1,...,N$, be a basis in S; then $\psi = \Sigma\, a_j\phi_j$ and the average over ψ, $\langle\ \rangle_\psi$, is an average over the uniform distribution of the $2N$ real variables $\mathrm{Re}\, a_1$, $\mathrm{Im}\, a_1$, ... on the sphere

$$(\mathrm{Re}\, a_1)^2 + (\mathrm{Im}\, a_1)^2 + \cdots = 1. \tag{2.1}$$

Thus if g is a function of $a_1,...,a_N$, we have

$$\langle g\rangle_\psi = \frac{\int d\,\mathrm{Re}\, a_1 d\,\mathrm{Im}\, a_1 \cdots g\delta(\Sigma|a_j|^2 - 1)}{\int d\,\mathrm{Re}\, a_1 d\,\mathrm{Im}\, a_1 \cdots \delta(\Sigma|a_j|^2 - 1)}. \tag{2.2}$$

The averages we need below are

$$\langle |a_j|^2\rangle_\psi = 1/N, \tag{2.3}$$

which is obvious, and

$$\langle |a_j|^2|a_k|^2\rangle_\psi = \begin{cases} 1/N(N+1) & \text{if } j\neq k\ , \\ 2/N(N+1) & \text{if } j = k\,, \end{cases} \tag{2.4}$$

which follow from an easy calculation.

The average over M-dimensional projectors P can be regarded as an average over all possible sets of M orthonormal vectors in S; we will have an integral over the expansion coefficients of M vectors with delta functions to enforce each normalization and orthogonality condition. Fortunately the averages that we need below can be evaluated without recourse to this formal definition. For example, since $\mathrm{tr}\, P = M$ we have

$$\langle P_{jj}\rangle_P = (\mathrm{tr}\, P)/N = M/N. \tag{2.5}$$

To evaluate $\langle P_{jj}^2\rangle_P$, we notice that in averaging $(\psi, P\psi)^2$ it clearly does not matter whether we average over all P for fixed ψ or over all ψ for fixed P, so we can replace $\langle P_{jj}^2\rangle_P$ by $\langle(\psi, P\psi)^2\rangle_\psi$, where P is any M-dimensional projector. Taking P to be the projector onto the subspace spanned by $\{\phi_1,...,\phi_M\}$, we therefore find

$$\begin{aligned}\langle(\psi, P\psi)^2\rangle_\psi &= \left\langle\left(\sum_{k=1}^{M} |a_k|^2\right)^2\right\rangle_\psi \\ &= M\left(\frac{2}{N(N+1)}\right) + M(M-1)\left(\frac{1}{N(N+1)}\right) \\ &= \frac{M(M+1)}{N(N+1)} = \langle P_{jj}^2\rangle_P\,,\end{aligned} \tag{2.6}$$

which is the result of Bocchieri and Loinger quoted above [Eq. (1.3)]. Finally, if $j\neq k$ we have

$$\begin{aligned}\langle P_{jj}P_{kk}\rangle_P &= \frac{1}{N(N-1)}\left\langle\sum_{\substack{j,k\\ j\neq k}} P_{jj}P_{kk}\right\rangle_P \\ &= \frac{1}{N(N-1)}\left\langle(\mathrm{tr}\, P)^2 - \sum_j P_{jj}^2\right\rangle_P \\ &= \frac{M^2}{N(N-1)} - \frac{N}{N(N-1)}\frac{M(M+1)}{N(N+1)} \\ &= \frac{M(NM-1)}{N(N^2-1)}.\end{aligned} \tag{2.7}$$

III. DISPERSION CALCULATIONS

Let $\{\phi_j\}, j = 1,...,N$, be a basis of eigenvectors of H in S, and assume first that the spectrum of H is nondegenerate. Then if $\psi = \Sigma\, a_k\,\phi_k$ is the initial state, we have

$$\psi(t) = \sum_k a_k\, e^{-iE_kt/\hbar}\phi_k, \tag{3.1a}$$

$$P(t) = \sum_{j,k} e^{i(E_j - E_k)t/\hbar} P_{jk}\, a_j^*\, a_k, \tag{3.1b}$$

$$\bar{P}_\psi = \lim_{T\to\infty} T^{-1}\int_0^T dt\, P(t) = \sum_j P_{jj}|a_j|^2, \tag{3.1c}$$

and

$$\langle\langle\bar{P}_\psi\rangle_\psi\rangle_P = \langle\bar{P}_\psi\rangle_\psi = \langle\bar{P}_\psi\rangle_P = M/N, \tag{3.2a}$$

$$\begin{aligned}&\langle\langle\bar{P}_\psi^2\rangle_\psi\rangle_P \\ &= N\langle P_{jj}^2\rangle_P\langle|a_j|^4\rangle_\psi \\ &\quad + N(N-1)\langle P_{jj}P_{kk}\rangle_P\langle|a_j|^2|a_k|^2\rangle_\psi \\ &= N\frac{M(M+1)}{N(N+1)}\frac{2}{N(N+1)} \\ &\quad + N(N-1)\frac{M(NM-1)}{N(N^2-1)}\frac{1}{N(N+1)} \\ &= \frac{M^2(N+2)+M}{N(N+1)^2} = \frac{M^2}{N^2}\left[1 + \frac{(N-M)}{M(N+1)^2}\right].\end{aligned} \tag{3.2b}$$

Therefore,

$$\langle\langle(\bar{P}_\psi - f)^2\rangle_\psi\rangle_P = f(1-f)/(N+1)^2\,, \tag{3.3}$$

which should be compared with the dispersion in the instantaneous value of $P(t)$,

$$\begin{aligned}\langle\langle(P(t) - f)^2\rangle_\psi\rangle_P &= \langle\langle(\psi, P\psi)^2 - f^2\rangle_\psi\rangle_P \\ &= \frac{M(M+1)}{N(N+1)} - \frac{M^2}{N^2} = \frac{f(1-f)}{(N+1)}.\end{aligned} \tag{3.4}$$

The dispersion in $P(t)$ is $O(1/N)$; the dispersion in the time average is $O(1/N^2)$.

Notice that Eq. (3.3) implies that inequality

$$\langle\langle(\bar{P}_\psi - f)^2\rangle_\psi\rangle_P/f^2 < (1-f)/NM, \tag{3.5}$$

which for large N is a considerable sharpening of inequality (1.7).

Let us now repeat the calculation for an arbitrarily degenerate spectrum. Let $E_\alpha, E_\beta,...$ be the distinct eigenvalues of H in S and $n_\alpha, n_\beta,...$ the dimensions of the corresponding eigenspaces $S_\alpha, S_\beta,...$. If the initial state is $\psi = \Sigma_\beta\,\psi_\beta$, where ψ_β is the component of ψ in S_β, we have

$$\psi(t) = \sum_\beta e^{-iE_\beta t/\hbar}\psi_\beta, \tag{3.6a}$$

$$P(t) = \sum_{\alpha,\beta} e^{i(E_\alpha - E_\beta)t/\hbar}(\psi_\alpha, P\psi_\beta), \tag{3.6b}$$

$$\bar{P}_\psi = \sum_\alpha(\psi_\alpha, P\psi_\alpha) = \sum_\alpha(\psi_\alpha, \psi_\alpha)(\hat{\psi}_\alpha, P\hat{\psi}_\alpha), \tag{3.6c}$$

where $\hat{\psi}_\alpha = \psi_\alpha/(\psi_\alpha, \psi_\alpha)^{1/2}$ is the unit vector along ψ_α. $\langle(\hat{\psi}_\alpha, P\hat{\psi}_\alpha)\rangle_P = M/N$, so

$$\langle\bar{P}_\psi\rangle_P = \frac{M}{N}\sum_\alpha(\psi_\alpha, \psi_\alpha) = \frac{M}{N} = \langle\langle\bar{P}_\psi\rangle_\psi\rangle_P \tag{3.7}$$

and

$$\langle\langle\bar{P}_\psi^2\rangle_\psi\rangle_P = \sum_{\alpha,\beta}\langle(\psi_\alpha, \psi_\alpha)(\psi_\beta, \psi_\beta)\rangle_\psi \times \langle(\hat{\psi}_\alpha, P\hat{\psi}_\alpha)(\hat{\psi}_\beta, P\hat{\psi}_\beta)\rangle_P. \tag{3.8}$$

As above, $\langle(\hat{\psi}_\alpha, P\hat{\psi}_\alpha)^2\rangle_P = M(M+1)/N(N+1)$, while if $\alpha \neq \beta$ $\hat{\psi}_\alpha$ and $\hat{\psi}_\beta$ are orthogonal and $\langle(\hat{\psi}_\alpha, P\hat{\psi}_\alpha)(\hat{\psi}_\beta, P\hat{\psi}_\beta)\rangle_P = M(NM-1)/N(N^2-1)$. To evaluate the averages over ψ, expand ψ_α in a basis for S_α, $\psi_\alpha = \Sigma_{k=1}^{n_\alpha} a_{\alpha k}\phi_{\alpha k}$; then

$$\langle(\psi_\alpha, \psi_\alpha)^2\rangle_\psi = \left\langle \sum_{k,l=1}^{n_\alpha} |a_{\alpha k}|^2|a_{\alpha l}|^2 \right\rangle_\psi = n_\alpha\frac{2}{N(N+1)} + \frac{n_\alpha(n_\alpha - 1)}{N(N+1)} = \frac{n_\alpha(n_\alpha+1)}{N(N+1)} \tag{3.9}$$

while if $\alpha \neq \beta$

$$\langle(\psi_\alpha, \psi_\alpha)(\psi_\beta, \psi_\beta)\rangle_\psi = \left\langle \sum_{k=1}^{n_\alpha}\sum_{l=1}^{n_\beta} |a_{\alpha k}|^2|a_{\beta l}|^2 \right\rangle_\psi = \frac{n_\alpha n_\beta}{N(N+1)}. \tag{3.10}$$

Therefore,

$$\begin{aligned}\langle\langle\bar{P}_\psi^2\rangle_\psi\rangle_P &= \frac{M(M+1)}{N(N+1)}\sum_\alpha\frac{n_\alpha(n_\alpha+1)}{N(N+1)} + \frac{M(NM-1)}{N(N^2-1)}\sum_{\substack{\alpha,\beta \\ \alpha\neq\beta}}\frac{n_\alpha n_\beta}{N(N+1)} \\ &= \frac{M(M+1)}{N(N+1)}\sum_\alpha\frac{n_\alpha(n_\alpha+1)}{N(N+1)} + \frac{M(NM-1)}{N(N^2-1)}\sum_\alpha\frac{n_\alpha(N-n_\alpha)}{N(N+1)}\end{aligned} \tag{3.11}$$

and, finally,

$$\langle\langle(\bar{P}_\psi - f)^2\rangle_\psi\rangle_P = \frac{f(1-f)}{(N-1)(N+1)^2}\left(\sum_\alpha n_\alpha^2 - 1\right). \tag{3.12}$$

Notice that Eq. (3.12) reduces to Eq. (3.3) when the spectrum is nondegenerate (all $n_\alpha = 1$) and to Eq. (3.4) in the opposite extreme of complete degeneracy ($n_\alpha = N$), when every state is stationary and $\bar{P}_\psi = (\psi, P\psi)$. Notice also that the dispersion is smaller the less degenerate the energy spectrum.

Finally, from Eq. (3.12) we get this sharpened form of the fundamental inequality of quantum ergodic theory,

$$\langle\langle(\bar{P}_\psi - f)^2\rangle_\psi\rangle_P/f^2 < (1-f)\left(\sum_\alpha n_\alpha^2\right)\Big/N^2M. \tag{3.13}$$

IV. DISCUSSION

The message of this calculation is that if one selects a state "at random" from an energy shell and determines, as a function of time, the probability that it lies in a "typical" subspace of the shell, the time average of this probability is liable to be much closer to the statistical expectation than is its instantaneous value, the more so the less degenerate the spectrum of the Hamiltonian.

One hesitates to call this hazy statement the solution to the problem of quantum ergodicity. Current opinion[7] has it that "ergodicity" or "chaos" in quantum mechanics must be associated with a characteristic pattern of energy level spacings—with a characteristic level spacing distribution—not just with the question of how many of these spacings happen to be zero. In classical mechanics time averaging over an infinite interval is essential to distinguish between ergodic and nonergodic motion; in quantum mechanics time averaging over an infinite interval destroys all dynamical distinctions except those associated with eigenvalue degeneracy. It is the behavior of quantum systems over finite intervals that is of most interest, not the infinite-time averages that von Neumann studied; still, it is comforting that the von Neumann approach to quantum ergodic theory, as embodied in Eq. (3.13), does provide some dynamical justification for our prejudice that nondegeneracy of the energy spectrum is the minimal requirement for ergodicity.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation.

[1] J. von Neumann, Z. Phys. **57**, 30 (1929).
[2] M. Fierz, Helv. Phys. Acta **28**, 705 (1955).
[3] D. ter Haar, Rev. Mod. Phys. **27**, 289 (1955).
[4] I. E. Farquhar and P. T. Landsberg, Proc. R. Soc. London Ser. A **239**, 134 (1957).
[5] P. Bocchieri and A. Loinger, Phys. Rev. **111**, 668 (1958).
[6] P. Bocchieri and A. Loinger, Phys. Rev. **114**, 948 (1959).
[7] See for instance M. V. Berry, Ann. Phys. (N.Y.) **131**, 163 (1981).

Quantization of three-dimensional space

Boris Leaf
Department of Physics, State University of New York, College at Cortland, Cortland, New York 13045

(Received 29 June 1983; accepted for publication 4 November 1983)

Geometrical properties of space are described by relations between quantum observables, functions of generalized coordinate and momentum operators of a physical system. Space comprises the set of eigenvalues of the coordinate observables, which may be a discrete set of points instead of the continuum of classical geometry. Differentiation on coordinates and momenta is defined by commutation operations. Well-known formulas from classical geometry for gradient and divergence operators, connections, covariant differentiation, tangent vectors, geodesics, and the Riemann curvature tensor are derived as relations between quantum operators. The quantum formulation retains the symmetry between coordinate and momentum representations required by the principle of complementarity.

PACS numbers: 03.65.Ca, 02.40. + m

1. INTRODUCTION

In accordance with quantum mechanical principles, the observable properties of a physical system are described by self-adjoint Hermitian operators. The eigenvalues of these observables are the measurable quantities which can be realized by the system. In particular, the position of a system is specified by a set of observables, three generalized coordinates in the case of a system with three degrees of freedom. These coordinates constitute a complete commuting set of operators, with nondegenerate spectrum. The system can be measured to be at only one of the points in the spectrum of these operators (with probability specified by the statistical operator which characterizes the state of the system). For generalized coordinates and momenta the spectra may be discrete[1,2]; a point in the continuum of real numbers which is not a spectral point of the coordinate observables has no physical reality as a possible location for a quantum mechanical system. In quantum mechanics, geometrical space comprises the set of eigenvalues of the coordinate observables, in contrast to classical geometry where a coordinate variable assumes all values on an interval of the continuum of real numbers. The geometrical properties of space should be described accordingly by functions which are realized only on the spectral values of the set of coordinate observables. These are the functions of the coordinate observables. In a geometry which is consistent with quantum mechanical principles, the real numbers of the classical description must be replaced by the self-adjoint Hermitian operators of which they are the eigenvalues; operations (functions) on the real numbers must be reformulated as operators on these self-adjoint operators.

In this paper we formulate such a quantized geometry for a system in three dimensions. Many of the equations of classical geometry reappear as relations between operators. Classical geometry is the classical limit of one representation (the coordinate representation) of an abstract quantum geometry. Our formulation retains the symmetry required by the principle of complementarity, and implicit in the fundamental commutation relations of coordinates and momenta, between the coordinate and momentum representations of the abstract geometry.

Considerable simplification of mathematical structure is achieved by expressing geometrical relations as equations between operators. Such equations hold regardless of the nature of the spectrum of the operators involved, whether continuous or discrete. This is particularly significant for differentiation. The derivative of a function $F(\hat{q})$ of a coordinate operator $\hat{q}$ will be defined as the commutator

$$F'(\hat{q}) = 2\pi i[\hat{\kappa},F(\hat{q})], \tag{1.1}$$

where $\hat{\kappa}$ is the momentum operator conjugate to $\hat{q}$. This definition extends the differential-calculus definition of the derivative of the function $F(q)$ at a point q in the spectrum of $\hat{q}$ to the general case where the spectrum of $\hat{q}$ need not be continuous. As shown in the Appendix the value of $F'(q)$ is simply the eigenvalue of the commutator $F'(\hat{q})$ in (1.1) at the spectral point q.

2. GEOMETRICAL OBSERVABLES

In three dimensions let $\{\hat{q}\} = \{\hat{q}^n; n = 1,2,3\}$ be a complete commuting set of generalized coordinate operators with the simultaneous eigenvector $|\{q\}\rangle$ belonging to the nondegenerate set of eigenvalues $\{q\}$, and similarly let $\{\hat{\kappa}\} = \{\hat{\kappa}_n; n = 1,2,3\}$ be the complete commuting set of conjugate generalized momenta (in units of $2\pi\hbar$), where (we employ the usual summation convention)

$$\langle\{q\}|\{\kappa\}\rangle = \exp(2\pi i q^n \kappa_n). \tag{2.1}$$

The fundamental commutation rules state that

$$2\pi i[\hat{\kappa}_n,\hat{q}^m] = \delta_n^m \hat{1}, \quad [\hat{q}^n,\hat{q}^m] = 0 = [\hat{\kappa}_n,\hat{\kappa}_m]. \tag{2.2}$$

The spectral representations of the coordinates and momenta are

$$\hat{q}^n = \int d^3\{q\}|\{q\}\rangle q^n\langle\{q\}|, \quad \hat{\kappa}_n = \int d^3\{\kappa\}|\{\kappa\}\rangle\kappa_n\langle\{\kappa\}|. \tag{2.3}$$

Also, let the vector position operator $\hat{\mathbf{x}}$ and momentum operator $\hat{\mathbf{k}}$ be designated as

$$\hat{\mathbf{x}} = \int d^3\mathbf{x}|\mathbf{x}\rangle\mathbf{x}\langle\mathbf{x}|, \quad \hat{\mathbf{k}} = \int d^3\mathbf{k}|\mathbf{k}\rangle\mathbf{k}\langle\mathbf{k}|, \tag{2.4}$$

where

 0022-2488/84/030535-06$02.50

$$\langle \mathbf{x}|\mathbf{k}\rangle = \exp(2\pi i \mathbf{x}\cdot\mathbf{k}). \tag{2.5}$$

The integration symbols in (2.3) and (2.4) indicate Stieltjes' integrations on the eigenvalues of the operators (spectral summations); the spectra need not be continuous.[1,2] Since we shall not be discussing spectral summations (unless specifically noted) we shall not use the caret to distinguish an operator from its eigenvalues in the following.

Define the partial derivative of an operator O with respect to coordinates and momenta as the commutation operations,

$$\partial_n O = \frac{\partial O}{\partial q^n} = 2\pi i[\kappa_n, O],$$
$$\partial^n O = \frac{\partial O}{\partial \kappa_n} = -2\pi i[q^n, O], \quad n = 1,2,3. \tag{2.6}$$

Accordingly, the rules of differentiation of operators are the rules of commutator algebra. In accordance with previous work in coordinate representation,[1,3] the vector momentum $\mathbf{k}$ is given by the Hermitian operator,

$$\mathbf{k} = \tfrac{1}{2}(\mathbf{e}^n\kappa_n + \kappa_n\mathbf{e}^n)$$
$$= \mathbf{e}^n\kappa_n + (4\pi i)^{-1}\partial_n\mathbf{e}^n = \kappa_n\mathbf{e}^n - (4\pi i)^{-1}\partial_n\mathbf{e}^n, \tag{2.7}$$

where $\mathbf{e}^n = \mathbf{e}^n(\{q\})$ is a vector-valued function of the coordinate operators. By the principle of complementarity, the position vector $\mathbf{x}$ is given by

$$\mathbf{x} = \tfrac{1}{2}(\boldsymbol{\epsilon}_n q^n + q^n\boldsymbol{\epsilon}_n)$$
$$= \boldsymbol{\epsilon}_n q^n - (4\pi i)^{-1}\partial^n\boldsymbol{\epsilon}_n = q^n\boldsymbol{\epsilon}_n + (4\pi i)^{-1}\partial^n\boldsymbol{\epsilon}_n, \tag{2.8}$$

where $\boldsymbol{\epsilon}_n = \boldsymbol{\epsilon}_n(\{\kappa\})$. The basis vectors $\{\mathbf{e}^n; n = 1,2,3\}$ of coordinate space commute with the coordinates $\{q\}$; the basis vectors $\{\boldsymbol{\epsilon}_n; n = 1,2,3\}$ of momentum space commute with $\{\kappa\}$. Accordingly, from (2.6), (2.7), (2.8), and (2.2),

$$\mathbf{e}^n(\{q\}) = -2\pi i[q^n, \mathbf{k}] = \partial^n\mathbf{k}, \tag{2.9}$$

$$\boldsymbol{\epsilon}_n(\{\kappa\}) = 2\pi i[\kappa_n, \mathbf{x}] = \partial_n\mathbf{x}, \tag{2.10}$$

and

$$\partial_n\partial_m\mathbf{x} = 0, \quad \partial^n\partial^m\mathbf{k} = 0, \quad n,m = 1,2,3. \tag{2.11}$$

Gradient operators ∇_x and ∇_k are defined as the commutation operators

$$\nabla_x = \mathbf{e}^n\partial_n, \quad \nabla_k = \boldsymbol{\epsilon}_n\partial^n. \tag{2.12}$$

If $F(\{q\})$ and $G(\{\kappa\})$ are scalar functions, then from (2.7) and (2.8)

$$\nabla_x F(\{q\}) = 2\pi i[\mathbf{k}, F(\{q\})],$$
$$\nabla_k G(\{\kappa\}) = -2\pi i[\mathbf{x}, G(\{\kappa\})], \tag{2.13}$$

so that the gradient of a scalar function in either coordinate or momentum space is a self-adjoint Hermitian operator (observable). Equations (2.9) and (2.10) can also be written as

$$\mathbf{e}^n = \nabla_x q^n, \quad \boldsymbol{\epsilon}_n = \nabla_k\kappa_n. \tag{2.14}$$

According to (2.12) and (2.6), $\nabla_x\kappa_n$ and $\nabla_k q^n$ both vanish; however, $[\mathbf{k},\kappa_n]$ and $[\mathbf{x},q^n]$ do not. For example,

$$2\pi i[\mathbf{k},\kappa_n] = -\tfrac{1}{2}[(\partial_n\mathbf{e}^m)\kappa_m + \kappa_m\partial_n\mathbf{e}^m] \neq 0. \tag{2.15}$$

The momentum vector operator does not commute with its generalized momentum components,[4] nor the position vector operator with its coordinate components.

In classical vector analysis the unit tensor is given by $\nabla_x\mathbf{x}$ and $\nabla_k\mathbf{k}$. However, from (2.9), (2.10), and (2.12),

$$\nabla_x\mathbf{x} = \mathbf{e}^n\boldsymbol{\epsilon}_n, \quad \nabla_k\mathbf{k} = \boldsymbol{\epsilon}_n\mathbf{e}^n, \tag{2.16}$$

which cannot represent the unit tensor, since $\mathbf{e}^n(\{q\})$ and $\boldsymbol{\epsilon}_n(\{\kappa\})$ do not commute. Neither can $2\pi i[\mathbf{k},\mathbf{x}]$ be the unit tensor, since

$$2\pi i[\mathbf{k},\mathbf{x}] = \tfrac{1}{2}(\mathbf{e}^n\boldsymbol{\epsilon}_n + \boldsymbol{\epsilon}_n\mathbf{e}^n) + \tfrac{1}{2}\pi i\{[\mathbf{e}^n,\boldsymbol{\epsilon}_m]q^m\kappa_n$$
$$+ \kappa_n q^m[\mathbf{e}^n,\boldsymbol{\epsilon}_m] + q^m[\mathbf{e}^n,\boldsymbol{\epsilon}_m]\kappa_n$$
$$+ \kappa_n[\mathbf{e}^n,\boldsymbol{\epsilon}_m]q^m\}. \tag{2.17}$$

To express the unit tensor, introduce the vectors $\boldsymbol{\xi} = \boldsymbol{\xi}(\{q\})$ and $\boldsymbol{\eta} = \boldsymbol{\eta}(\{\kappa\})$, and define

$$\mathbf{e}_n(\{q\}) = \partial_n\boldsymbol{\xi}, \quad \boldsymbol{\epsilon}^n(\{\kappa\}) = \partial^n\boldsymbol{\eta}, \tag{2.18}$$

where the unit tensor $\mathbf{I}$ is given by

$$\nabla_x\boldsymbol{\xi} = \mathbf{e}^n\mathbf{e}_n = \mathbf{I}, \quad \nabla_k\boldsymbol{\eta} = \boldsymbol{\epsilon}_n\boldsymbol{\epsilon}^n = \mathbf{I}. \tag{2.19}$$

Now the symmetric metric tensors can be defined as the operators

$$g_{nm}(\{q\}) = \mathbf{e}_n\cdot\mathbf{e}_m, \quad g^{nm}(\{q\}) = \mathbf{e}^n\cdot\mathbf{e}^m, \tag{2.20}$$

$$f^{nm}(\{\kappa\}) = \boldsymbol{\epsilon}^n\cdot\boldsymbol{\epsilon}^m, \quad f_{nm}(\{\kappa\}) = \boldsymbol{\epsilon}_n\cdot\boldsymbol{\epsilon}_m, \tag{2.21}$$

It follows accordingly that also

$$\mathbf{e}_n\mathbf{e}^n = \mathbf{I} = \boldsymbol{\epsilon}^n\boldsymbol{\epsilon}_n \tag{2.22}$$

and

$$\mathbf{e}^m\cdot\mathbf{e}_n = \delta^m_n = \mathbf{e}_n\mathbf{e}^m, \quad \boldsymbol{\epsilon}_m\cdot\boldsymbol{\epsilon}^n = \delta^n_m = \boldsymbol{\epsilon}^n\cdot\boldsymbol{\epsilon}_m. \tag{2.23}$$

The connection operators $\Gamma^k_{nm}(\{q\})$ and $\Gamma^{nm}_k(\{\kappa\})$ are defined by

$$\partial_m\mathbf{e}^k = -\mathbf{e}^n\Gamma^k_{nm}, \quad \partial_m\mathbf{e}_n = \mathbf{e}_k\Gamma^k_{nm}, \tag{2.24}$$

$$\partial^m\boldsymbol{\epsilon}_k = -\boldsymbol{\epsilon}_n\Gamma^{nm}_k, \quad \partial^m\boldsymbol{\epsilon}^n = \boldsymbol{\epsilon}^k\Gamma^{nm}_k. \tag{2.25}$$

Since, according to (2.18),

$$\partial_m\mathbf{e}_n = \partial_m\partial_n\boldsymbol{\xi} = \partial_n\mathbf{e}_m, \quad \partial^m\boldsymbol{\epsilon}^n = \partial^m\partial^n\boldsymbol{\eta} = \partial^n\boldsymbol{\epsilon}^m, \tag{2.26}$$

so that the sets $\{\mathbf{e}_n\}$ and $\{\boldsymbol{\epsilon}^n\}$ are coordinate bases,[5] therefore, the connections have the symmetries

$$\Gamma^k_{nm} = \Gamma^k_{mn}, \quad \Gamma^{mn}_k = \Gamma^{nm}_k. \tag{2.27}$$

From (2.24) and (2.12)

$$\nabla_x\mathbf{e}^n = -\mathbf{e}^m\mathbf{e}^k\Gamma^n_{km}, \quad \nabla_x\mathbf{e}_n = \mathbf{e}^m\mathbf{e}_k\Gamma^k_{nm}. \tag{2.28}$$

Because of the symmetry of Γ^k_{nm} in (2.25), $\nabla_x\mathbf{e}^n$ is a self-adjoint Hermitian operator; $\nabla_x\mathbf{e}_n$ is not. Similarly, from (2.25) and (2.12)

$$\nabla_k\boldsymbol{\epsilon}_n = -\boldsymbol{\epsilon}_m\boldsymbol{\epsilon}_k\Gamma^{km}_n, \quad \nabla_k\boldsymbol{\epsilon}^n = \boldsymbol{\epsilon}_m\boldsymbol{\epsilon}^k\Gamma^{nm}_k, \tag{2.29}$$

where $\nabla_k\boldsymbol{\epsilon}_n$ is self-adjoint, but $\nabla_k\boldsymbol{\epsilon}^n$ is not.

From (2.7) and (2.24), and from (2.8) and (2.25),

$$\mathbf{k} = \mathbf{e}^n(\kappa_n - (4\pi i)^{-1}\Gamma^m_{nm}) = (\kappa_n + (4\pi i)^{-1}\Gamma^m_{nm})\mathbf{e}^n, \tag{2.30}$$

$$\mathbf{x} = \boldsymbol{\epsilon}_n(q^n + (4\pi i)^{-1}\Gamma^{nm}_m) = (q^n - (4\pi i)^{-1}\Gamma^{nm}_m)\boldsymbol{\epsilon}_n. \tag{2.31}$$

Also, from (2.30) and (2.28)

$$\kappa_n = \tfrac{1}{2}(\mathbf{e}_n\cdot\mathbf{k} + \mathbf{k}\cdot\mathbf{e}_n)$$
$$= \mathbf{k}\cdot\mathbf{e}_n - (4\pi i)^{-1}\Gamma^m_{nm} = \mathbf{e}_n\cdot\mathbf{k} + (4\pi i)^{-1}\Gamma^m_{nm}, \tag{2.32}$$

$$\Gamma^{m}_{nm} = 2\pi i(\mathbf{k}\cdot\mathbf{e}_n - \mathbf{e}_n\cdot\mathbf{k}) = \nabla_x\cdot\mathbf{e}_n, \tag{2.33}$$

and from (2.31) and (2.29)

$$q^n = \tfrac{1}{2}(\epsilon^n\cdot\mathbf{x} + \mathbf{x}\cdot\epsilon^n)$$
$$= \mathbf{x}\cdot\epsilon^n + (4\pi i)^{-1}\Gamma^{nm}_{m} = \epsilon^n\cdot\mathbf{x} - (4\pi i)^{-1}\Gamma^{nm}_{m}, \tag{2.34}$$
$$\Gamma^{nm}_{m} = 2\pi i(\mathbf{x}\cdot\epsilon^n - \epsilon^n\cdot\mathbf{x}) = \nabla_k\cdot\epsilon^n. \tag{2.35}$$

It is a well-known consequence[6] of the properties of the determinant g of the metric tensor (g_{nm}) in (2.20), valid also when the elements of (g_{nm}) are commuting operators, that

$$\Gamma^{m}_{mn} = J^{-1}\partial_n J, \quad J(\{q\}) = g^{1/2}, \tag{2.36}$$

where J is a Jacobian operator. Therefore, from (2.27) and (2.33)

$$\nabla_x\cdot\mathbf{e}_n = \Gamma^{m}_{mn} = J^{-1}\partial_n J, \quad \partial_n(J\mathbf{e}^n) = 0. \tag{2.37}$$

Accordingly, from (2.30)

$$\mathbf{k} = \mathbf{e}^n(\kappa_n - (2\pi i)^{-1}J^{-1/2}\partial_n J^{1/2})$$
$$= \mathbf{e}^n(\kappa_n + J^{1/2}[\kappa_n, J^{-1/2}])$$
$$= \mathbf{e}^n J^{1/2}\kappa_n J^{-1/2} = J^{-1/2}\kappa_n J^{1/2}\mathbf{e}^n, \tag{2.38}$$

an equation expressing the vector momentum operator $\mathbf{k}$ in terms of its components $\{\kappa\}$. To see that this operator equation agrees with the results in the coordinate representation obtained previously, write the spectral summation (the caret again denotes an operator) using (2.4) and (2.5),

$$2\pi i\hat{\mathbf{k}} = \int d^3\mathbf{x}|\mathbf{x}\rangle(\partial/\partial\mathbf{x})\langle\mathbf{x}|, \tag{2.39}$$

and using (2.1) and (2.3),

$$2\pi i\hat{\kappa}_n = \int d^3\{q\}|\{q\}\rangle(\partial/q^n)\langle\{q\}|. \tag{2.40}$$

Since $d^3\mathbf{x} = J d^3\{q\}$, where J is the Jacobian in coordinate space and $|\mathbf{x}\rangle = J^{-1/2}|\{q\}\rangle$, (2.40) becomes

$$2\pi i\hat{\kappa}_n = \int d^3\mathbf{x}|\mathbf{x}\rangle J^{-1/2}\frac{\partial}{\partial q^n}J^{1/2}\langle\mathbf{x}|, \tag{2.41}$$

in agreement with Eq. (13) of Ref. 1. From (2.38), (2.39), and (2.41)

$$\int d^3\mathbf{x}|\mathbf{x}\rangle\frac{\partial}{\partial\mathbf{x}}\langle\mathbf{x}| = \int d^3\mathbf{x}|\mathbf{x}\rangle\mathbf{e}^n\frac{\partial}{\partial q^n}\langle\mathbf{x}|$$

so that the coordinate representative of the vector momentum $2\pi\hbar\mathbf{k}$ is

$$\frac{\hbar}{i}\frac{\partial}{\partial\mathbf{x}} = \mathbf{e}^n\frac{\hbar}{i}\frac{\partial}{\partial q^n}, \tag{2.42}$$

in agreement with Eq. (1) of Ref. 1. From (2.36) it follows that

$$\partial_m\Gamma^{s}_{sn} = \partial_n\Gamma^{s}_{sm}. \tag{2.43}$$

Similarly, as in (2.36),

$$\Gamma^{mn}_{m} = \mathscr{J}^{-1}\partial^n\mathscr{J}, \quad \mathscr{J}(\{\kappa\}) = f^{1/2}, \tag{2.44}$$

where f is the determinant of the tensor (f^{nm}) in (2.21); $\mathscr{J}$ is the Jacobian operator. Therefore, from (2.27) and (2.35)

$$\nabla_k\cdot\epsilon^n = \Gamma^{mn}_{m} = \mathscr{J}^{-1}\partial^n\mathscr{J}, \quad \partial^n(\mathscr{J}\epsilon_n) = 0. \tag{2.45}$$

From (2.31)

$$\mathbf{x} = \epsilon_n(q^n + (2\pi i)^{-1}\mathscr{J}^{-1/2}\partial^n\mathscr{J}^{1/2})$$
$$= \epsilon_n(q^n + \mathscr{J}^{1/2}[q^n, \mathscr{J}^{-1/2}])$$
$$= \epsilon_n\mathscr{J}^{1/2}q^n\mathscr{J}^{-1/2} = \mathscr{J}^{-1/2}q^n\mathscr{J}^{1/2}\epsilon_n, \tag{2.46}$$

which expresses vector position operator $\mathbf{x}$ in terms of its components $\{q\}$. This equation for $\mathbf{x}$ is the complement of (2.38) for $\mathbf{k}$. From (2.4) and (2.5)

$$-2\pi i\hat{\mathbf{x}} = \int d^3\mathbf{k}|\mathbf{k}\rangle\frac{\partial}{\partial\mathbf{k}}\langle\mathbf{k}|, \tag{2.47}$$

and from (2.1) and (2.3)

$$-2\pi i q^n = \int d^3\{\kappa\}|\{\kappa\}\rangle\frac{\partial}{\partial\kappa_n}\langle\{\kappa\}|$$
$$= \int d^3\mathbf{k}|\mathbf{k}\rangle\mathscr{J}^{-1/2}\frac{\partial}{\partial\kappa_n}\mathscr{J}^{1/2}\langle\mathbf{k}|, \tag{2.48}$$

since $d^3\mathbf{k} = \mathscr{J}\, d^3\{\kappa\}$, where $\mathscr{J}$ is the Jacobian in momentum space and $|\mathbf{k}\rangle = \mathscr{J}^{-1/2}|\{\kappa\}\rangle$. Therefore, from (2.46)

$$-2\pi i\hat{\mathbf{x}} = \int d_3\mathbf{k}|\mathbf{k}\rangle\frac{\partial}{\partial\mathbf{k}}\langle\mathbf{k}| = \int d^3\mathbf{k}|\mathbf{k}\rangle\epsilon_n\frac{\partial}{\partial\kappa_n}\langle k|,$$

so that the momentum (wave number) representative of $\hat{x}$ is

$$\frac{i}{2\pi}\frac{\partial}{\partial\mathbf{k}} = \epsilon_n\frac{i}{2\pi}\frac{\partial}{\partial\kappa_n}. \tag{2.49}$$

3. TANGENT VECTORS IN THE SPACE OF COORDINATE OPERATORS

Since the basis operators $\{\mathbf{e}^n, \mathbf{e}_n, n = 1,2,3\}$ commute with functions of the coordinates $\{q\}$ they can be used to construct a vector space, the space of coordinate operators, analogous to the coordinate space of classical geometry. The measurable values of functions of these operators are realized on the set of spectral points which are the eigenvalues of $\{q\}$.

Consider the vector operator

$$\mathbf{v} = \mathbf{e}_n v^n, \quad v^n = v^n(\{q\}). \tag{3.1}$$

From (2.12) and (2.24)

$$\partial_n\mathbf{v} = \mathbf{e}_m v^m_{;n}, \quad \nabla_x\mathbf{v} = \mathbf{e}^n\mathbf{e}_m v^m_{;n}, \tag{3.2}$$

where the convariant derivative of v^n is

$$v^m_{;n} = \partial_n v^m + v^s\Gamma^{m}_{sn}. \tag{3.3}$$

Alternatively,

$$\mathbf{v} = \mathbf{e}^n v_n, \quad v_n = g_{nm}v^m, \tag{3.4}$$

so that

$$\partial_n\mathbf{v} = \mathbf{e}^m v_{m;n}, \quad \nabla_x\mathbf{v} = \mathbf{e}^n\mathbf{e}^m v_{m;n}, \tag{3.5}$$

where

$$v_{m;n} = \partial_n v_m - v_s\Gamma^{s}_{mn}. \tag{3.6}$$

It is of interest that

$$[\mathbf{k},\mathbf{v}] = [\mathbf{e}^m\mathbf{e}_m\cdot\mathbf{k},\mathbf{v}] = \mathbf{e}^m[\kappa_m,\mathbf{v}] + [\mathbf{e}^m,\mathbf{e}^n]v_n\mathbf{e}_m\cdot\mathbf{k}, \tag{3.7}$$

so that from (2.38)

$$2\pi i[\mathbf{k},\mathbf{v}] = \nabla_x\mathbf{v} + 2\pi i[\mathbf{e}^m,\mathbf{e}^n]v_n J^{1/2}\kappa_m J^{-1/2}. \tag{3.8}$$

Accordingly, $\nabla_x\mathbf{v}$ differs from the self-adjoint Hermitian operator $2\pi i[\mathbf{k},\mathbf{v}]$ only because $[\mathbf{e}^m,\mathbf{e}^n]$ does not vanish; the noncommutivity of $\mathbf{e}^m$ and $\mathbf{e}^n$ for $m \neq n$ is not a quantum effect, but simply expresses the fact that the dyadic $\mathbf{e}^m\mathbf{e}^n$ is not symmetric. This result may be contrasted with (2.13) for the case of the gradient of a scalar function $F(\{q\})$. The average of $2\pi i[\mathbf{k},\mathbf{v}]$ in (3.8) with its Hermitian adjoint gives

$$2\pi i[\mathbf{k},\mathbf{v}] = \tfrac{1}{2}(\mathbf{e}^m\mathbf{e}^n + \mathbf{e}^n\mathbf{e}^m)v_{n;m} + \pi i([\mathbf{e}^m,\mathbf{e}^n]v_n\kappa_m + \kappa_m v_n[\mathbf{e}^m,\mathbf{e}^n]). \quad (3.9)$$

In contraction of $2\pi i[\mathbf{k},\mathbf{v}]$ in (3.8) the term in $[\mathbf{e}^m,\mathbf{e}^n]$ vanishes so that the divergence of $\mathbf{v}$ in (3.2) is

$$\nabla_x \cdot \mathbf{v} = 2\pi i(\mathbf{k}\cdot\mathbf{v} - \mathbf{v}\cdot\mathbf{k}) = v^n_{;n} = J^{-1}\partial_n(Jv^n). \quad (3.10)$$

Special cases, where $\mathbf{v} = \mathbf{e}^n$ or $\mathbf{e}_n$, appear in (2.24) and (2.28). In the case that $\mathbf{v}$ is the vector $\boldsymbol{\xi}$ in (2.18) and (2.19),

$$\nabla_x \boldsymbol{\xi} = \mathbf{e}^n\mathbf{e}_m\xi^m_{;n} = \mathbf{I}, \quad (3.11)$$

so that

$$\xi^m_{;n} = \delta^m_n, \quad \nabla_x \cdot \boldsymbol{\xi} = 3. \quad (3.12)$$

The Riemann curvature tensor in the space of coordinate operators is readily obtained from the defining condition,

$$(\partial_m\partial_n - \partial_n\partial_m)\mathbf{v} = \mathbf{e}_s R^s_{\ inm} v^i = \mathbf{e}^s R^{\ i}_{smn} v_i, \quad (3.13)$$

where $\mathbf{v}$ in (3.1) is any vector of the space. In particular, for $v^i = \delta^i_j$,

$$(\partial_m\partial_n - \partial_n\partial_m)\mathbf{e}_j = \mathbf{e}_s R^s_{\ jnm}, \quad (3.14)$$

so that from (2.24), the components of the Riemann tensor are

$$R^s_{\ inm} = \partial_m\Gamma^s_{\ in} - \partial_n\Gamma^s_{\ im} + \Gamma^k_{\ in}\Gamma^s_{\ km} - \Gamma^k_{\ im}\Gamma^s_{\ kn}, \quad (3.15)$$

the usual expression found in classical geometry. However, the connections, defined in (2.24) and (2.25), are quantum operators which are functions of the coordinate operators $\{q\}$. Accordingly, $R^s_{\ inm}$ is also a quantum operator, a function of $\{q\}$, whose measurable values are realized only on the spectral points which are the eigenvalues of $\{q\}$.

Using the components v^n of $\mathbf{v}$, define the corresponding linear derivation in the direction of $\mathbf{v}$ as the commutation operation,

$$\frac{d}{d\lambda(\mathbf{v})} = v^n\partial_n = \mathbf{v}\cdot\nabla_x. \quad (3.16)$$

In particular,

$$\frac{dq^n}{d\lambda(\mathbf{v})} = v^m\partial_m q^n = v^n, \quad (3.17)$$

so that $\mathbf{v}$ can be written as the tangent vector operator,

$$\mathbf{v}(\{q\}) = \mathbf{e}_n\frac{dq^n}{d\lambda(\mathbf{v})}, \quad (3.18)$$

with the corresponding derivation in the direction of $\mathbf{v}$,

$$\frac{d}{d\lambda(\mathbf{v})} = \frac{dq^n}{d\lambda(\mathbf{v})}\partial_n. \quad (3.19)$$

Also, from (3.11),

$$\mathbf{v} = \mathbf{v}\cdot\nabla_x\boldsymbol{\xi} = \frac{d\boldsymbol{\xi}}{d\lambda(\mathbf{v})} \quad (3.20)$$

and

$$\frac{d^2\boldsymbol{\xi}}{d\lambda^2(\mathbf{v})} = \frac{d\mathbf{v}}{d\lambda(\mathbf{v})} = \mathbf{v}\cdot\nabla_x\mathbf{v}. \quad (3.21)$$

On a geodesic in the space of coordinate operators, $\mathbf{v}\cdot\nabla_x\mathbf{v} = 0$, so that

$$\mathbf{v}\cdot\nabla_x\mathbf{v}\cdot\mathbf{e}^m = v^n v^m_{;n} = \frac{dv^m}{d\lambda(\mathbf{v})} + v^n v^s\Gamma^{\ m}_{sn} = 0. \quad (3.22)$$

Thus from (3.17), the differential equation for a geodesic is obtained as a relation on the coordinate operators,

$$\frac{d^2q^n}{d\lambda^2(\mathbf{v})} + \left[\frac{dq^m}{d\lambda(\mathbf{v})}\right]\left[\frac{dq^s}{d\lambda(\mathbf{v})}\right]\Gamma^{\ n}_{sm} = 0. \quad (3.23)$$

We may require that $\mathbf{v}$ be a unit tangent vector, so that from (3.18)

$$\mathbf{v}\cdot\mathbf{v} = v^n g_{nm} v^m = \left[\frac{dq^n}{d\lambda(\mathbf{v})}\right]\left[\frac{dq^m}{d\lambda(v)}\right]g_{nm} = 1, \quad (3.24)$$

in which case $d\lambda(\mathbf{v})$ appears as the differential arc-length tangent to $\mathbf{v}$. This result holds for all directions of the unit vector $\mathbf{v}$ [as does the geodesic equation (3.23)], so that, in place of $d\lambda(\mathbf{v})$, we write the Riemannian metric $d\lambda$ in the space of coordinate operators,

$$(d\lambda)^2 1 = g_{nm}(dq^n)(dq^m) = (d\boldsymbol{\xi})\cdot(d\boldsymbol{\xi}), \quad (3.25)$$

according to (3.20). In (3.25), 1 is the unit quantum operator; the metric $d\lambda$ is not a q-number.

Angular momentum[7] is defined $\mathbf{L} = 2\pi\hbar\boldsymbol{\xi}\times\mathbf{k}$.

4. TANGENT VECTORS IN THE SPACE OF MOMENTUM OPERATORS

The results obtained in the previous section for the space of coordinate operators can readily be duplicated, *mutatis mutandis*, for the complementary momentum space, where the operators $\{\boldsymbol{\epsilon}_n, \boldsymbol{\epsilon}^n\}$ commute with functions of $\{\kappa\}$. Let

$$\mathbf{w} = \boldsymbol{\epsilon}^n w_n, \quad w_n = w_n(\{\kappa\}). \quad (4.1)$$

Then

$$\partial^n\mathbf{w} = \boldsymbol{\epsilon}^m w^{;n}_m, \quad \nabla_k\mathbf{w} = \boldsymbol{\epsilon}_n\boldsymbol{\epsilon}^m w^{;n}_m, \quad (4.2)$$

$$w^{;n}_m = \partial^n w_m + w_s\Gamma^{sn}_{\ \ m}. \quad (4.3)$$

Or

$$\mathbf{w} = \boldsymbol{\epsilon}_n \mathbf{w}^n, \quad w^n = f^{nm}w_m, \quad (4.4)$$

$$\partial^n\mathbf{w} = \boldsymbol{\epsilon}_m w^{m;n}, \quad \nabla_k\mathbf{w} = \boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_m w^{m;n}, \quad (4.5)$$

$$w^{m;n} = \partial^n w^m - w^s\Gamma^{\ mn}_{s}. \quad (4.6)$$

Also,

$$\nabla_k\cdot\mathbf{w} = -2\pi i(\mathbf{x}\cdot\mathbf{w} - \mathbf{w}\cdot\mathbf{x}) = w^{;n}_n. \quad (4.7)$$

A curvature tensor in the space of momentum operators is defined by the condition

$$(\partial^m\partial^n - \partial^n\partial^m)\mathbf{w} = \boldsymbol{\epsilon}^s\mathcal{R}^{\ smn}_{i} w_i = \boldsymbol{\epsilon}_s\mathcal{R}^{smn}_{i} w^i, \quad (4.8)$$

with $\mathbf{w}$ given in (4.1). In particular,

$$(\partial^m\partial^n - \partial^n\partial^m)\boldsymbol{\epsilon}^i = \boldsymbol{\epsilon}^s\mathcal{R}^{\ inm}_{s}, \quad (4.9)$$

so that from (2.25)

$$\mathcal{R}^{\ inm}_{s} = \partial^m\Gamma^{\ in}_{s} - \partial^n\Gamma^{\ im}_{s} + \Gamma^{\ in}_{k}\Gamma^{\ km}_{s} - \Gamma^{\ im}_{k}\Gamma^{\ kn}_{s}. \quad (4.10)$$

$\mathcal{R}^{\ inm}_{s}$ is a quantum operator whose measurable values are realized only on the points which are eigenvalues of the momentum operators $\{\kappa\}$.

The linear derivation in the direction of $\mathbf{w}$ is defined as

$$\frac{d}{d\mu(\mathbf{w})} = w_n \partial^n = \mathbf{w} \cdot \nabla_k. \tag{4.11}$$

Since

$$\frac{d\kappa_n}{d\mu(\mathbf{w})} = w_n, \tag{4.12}$$

therefore, $\mathbf{w}$ is the tangent vector operator

$$\mathbf{w}(\{\kappa\}) = \epsilon^n \frac{d\kappa_n}{d\mu(\mathbf{w})}, \tag{4.13}$$

and

$$\frac{d}{d\mu(\mathbf{w})} = \frac{\partial\kappa_n}{\partial\mu(\mathbf{w})}\partial^n. \tag{4.14}$$

From (2.19)

$$\mathbf{w} = \mathbf{w} \cdot \nabla_k \boldsymbol{\eta} = \frac{d\eta}{d\mu(\mathbf{w})} \tag{4.15}$$

and

$$\frac{d^2\boldsymbol{\eta}}{d\mu^2(\mathbf{w})} = \frac{d\mathbf{w}}{d\mu(\mathbf{w})} = \mathbf{w} \cdot \nabla_k \mathbf{w}. \tag{4.16}$$

On a geodesic in the space of momentum operators, $\mathbf{w} \cdot \nabla_k \mathbf{w} = 0$, so that

$$\frac{dw_m}{d\mu(\mathbf{w})} + w_n w_s \Gamma^{sn}_{\ m} = 0 \tag{4.17}$$

or

$$\frac{d^2\kappa_m}{d\mu^2(\mathbf{w})} + \left[\frac{d\kappa_n}{d\mu(\mathbf{w})}\right]\left[\frac{d\kappa_s}{d\mu(\mathbf{w})}\right]\Gamma^{sn}_{\ m} = 0, \tag{4.18}$$

the geodesic equation in the space of momentum operators. The Riemannian metric in this space is $d\mu$, the differential arc length in an arbitrary direction where, if $\mathbf{w} \cdot \mathbf{w} = 1$,

$$(d\mu)^2 1 = f^{nm}(d\kappa_n)(d\kappa_m) = (d\eta) \cdot (d\eta). \tag{4.19}$$

5. DISCUSSION

Complete symmetry exists between the complementary spaces of coordinate operators and momentum operators. It is only necessary to replace functions of one set of operators $\{q\}$ or $\{\kappa\}$ by the corresponding functions of the other set, interchanging covariant and contravariant indices. This result is in agreement with the principle of complementarity between coordinate (particlelike) and momentum (wavelike) aspects of the physical system. The uncertainty principle of quantum mechanics will operate with respect to these complementary geometrical spaces. If the state of the system is specified by a statistical operator (state vector) which is closely approximated by an eigenstate of the coordinates $\{q\}$, then all spatial geometric properties are well determined (as in classical geometry), but the complementary momentum geometric properties have a high degree of uncertainty. If, however, the state vector is closely approximated by an eigenstate of the momentum components $\{\kappa\}$, then the momentum geometric properties are well determined and the high degree of uncertainty attaches to the spatial geometry. In the Schrödinger picture the state vector changes in time under the influence of a Hamiltonian $H(\{q\},\{\kappa\})$, or as a result of measurement. The degree of uncertainty in the spatial geometric properties, those associated with the particlelike aspect of the system, accordingly is variable and changes with the state of the system; similarly, for the momentum geometric properties.

In the context of the quantum geometrical formalism, the following identifications can be made with current terminology in differential geometry.[8,9] The set of eigenvalues of each tangent vector operator $\hat{\mathbf{v}}(\{\hat{q}\})$ (the caret denotes operator) specified a vector field, assigning a value of a tangent vector $\mathbf{v}(\{q\})$ to each point $\{q\}$ in the three-dimensional space of the system, i.e., the set of points in the spectrum of the coordinate operators $\{\hat{q}\}$. The set of all tangent vectors at a point $\{q\}$ is the collection of eigenvalues at $\{q\}$ of all operators $\hat{\mathbf{v}}(\{\hat{q}\})$; this is the tangent space at $\{q\}$. The union of all tangent spaces for all spectral points $\{q\}$ is the tangent bundle (whose projection onto the specific point $\{q\}$ is the tangent space at $\{q\}$); the tangent bundle is, therefore, the collection of eigenvalues of all operators $\hat{\mathbf{v}}(\{\hat{q}\})$ of the space of coordinate operators of the system. A section of the bundle is the set of eigenvalues of a particular $\hat{\mathbf{v}}(\{\hat{q}\})$, the aforesaid vector field. The cotangent bundle is the collection of eigenvalues of all operators $\hat{\mathbf{w}}(\{\hat{\kappa}\})$ in the space of momentum operators of the system. A section of the cotangent bundle is the set of eigenvalues of a particular operator $\hat{\mathbf{w}}(\{\hat{\kappa}\})$, a field $\mathbf{w}(\{\kappa\})$ in the space of momentum eigenvalues. However, because of the noncommutivity of $\{\hat{\kappa}\}$ and $\{\hat{q}\}$, one cannot specify the cotangent space at a point $\{q\}$ (or the tangent space at a point $\{\kappa\}$).

APPENDIX: DIFFERENTIATION ON THE EIGENVALUE q OF A COORDINATE OPERATOR $\hat{q}$

Let $F(\hat{q})$ be a self-adjoint Hermitian function of coordinate operator $\hat{q}$ (a complete commuting set), and let

$$F(\hat{q})|q\rangle = F(q)|q\rangle. \tag{A1}$$

Then[10]

$$[F(q) - F(q')]\langle q|q'\rangle = 0 \tag{A2}$$

and

$$F(q) = \int dq' \langle q|q'\rangle F(q'), \tag{A3}$$

so that $F(q)$ is in the class of test functionals of the Schwartz distribution $\langle q|q'\rangle$. The integration is a Stieltjes integral (spectral summation) on the set of (generalized) eigenvalues q of $\hat{q}$; the spectrum need not be continuous. If $\hat{\kappa}$ is the momentum operator conjugate to $\hat{q}$, then

$$[[\hat{\kappa},F(\hat{q})],\hat{q}] = -[[F(\hat{q}),\hat{q}],\hat{\kappa}] - [[\hat{q},\hat{\kappa}],F(\hat{q})] = 0, \tag{A4}$$

so that the derivative of $F(\hat{q})$ [see (2.6)], namely,

$$F'(\hat{q}) = 2\pi i[\hat{\kappa},F(\hat{q})], \tag{A5}$$

commutes with $\hat{q}$ and is also a self-adjoint function of $\hat{q}$. Accordingly,

$$\langle q_1|F'(\hat{q})|q_2\rangle = F'(q_2)\langle q_1|q_2\rangle$$

$$= [F(q_2) - F(q_1)]2\pi i\langle q_1|\hat{\kappa}|q_2\rangle. \tag{A6}$$

But, from (2.2),

$$2\pi i\langle q_i|[\hat{\kappa},\hat{q}]|q_2\rangle = \langle q_1|q_2\rangle = (q_2 - q_1)2\pi i\langle q_1|\hat{\kappa}|q_2\rangle, \tag{A7}$$

and therefore, substitution into (A6) gives

$$F'(q_2)\langle q_1|q_2\rangle = \{[F(q_2) - F(q_1)]/(q_2 - q_1)\}\langle q_1|q_2\rangle. \tag{A8}$$

Since $F'(q)$ is also in the class of test functionals of $\langle q|q'\rangle$ in (A3), spectral summation on q_2 gives

$$F'(q_1) = \int dq_2\{[F(q_2) - F(q_1)]/(q_2 - q_1)\}\langle q_1|q_2\rangle. \tag{A9}$$

$F'(q_1)$, the eigenvalue of the commutator $F'(\hat{q})$ in (A5), is the derivative on q, the eigenvalue of coordinate operator $\hat{q}$, of $F(q)$, the eigenvalue of $F(\hat{q})$, whether the spectrum of $\hat{q}$ is continuous or discrete. For example, if

$$\psi(q_1) = \sum_{n=0}^{\infty} \frac{\psi^{(n)}(q_0)(q_1 - q_0)^n}{n!}, \tag{A10}$$

a power series, for all points q_1, q_0 in the spectrum of $\hat{q}$, then $\psi(\hat{q})$ is given by the power series

$$\psi(\hat{q}) = \sum_{n=0}^{\infty} \frac{\psi^{(n)}(q_0)(\hat{q} - q_0\hat{1})^n}{n!} \tag{A11}$$

at each spectral point q_0. From (A5) in this case,

$$\psi'(\hat{q}) = \sum_{n=1}^{\infty} \frac{\psi^{(n)}(q_0)(\hat{q} - q_0\hat{1})^{n-1}}{(n-1)!}, \tag{A12}$$

and from (A9) and (A10)

$$\psi'(q_1) = \int dq_0\{[\psi(q_1) - \psi(q_0)]/(q_1 - q_0)\}\langle q_1|q_0\rangle$$
$$= \int dq_0\, \psi^{(1)}(q_0)\langle q_1|q_0\rangle = \psi^{(1)}(q_1), \tag{A13}$$

since for integral $n \geqslant 0$,

$$\int dq_0(q_1 - q_0)^n\langle q_1|q_0\rangle = \delta_{n,0}. \tag{A14}$$

Conversely, if $\psi(\hat{q})$ is given by a power series of the form (A11), the coefficient $\psi^{(n)}(q_0)$ is the nth derivative of $\psi(q)$ at the spectral point q_0, according to definition of the derivative, (A9).

Obviously analogous results follow for differentiation on the eigenvalue κ of a momentum operator $\hat{\kappa}$. The results apply to vector-valued functions as well as to the scalar functions $F(\hat{q})$ considered above.

[1]B. Leaf, Found. Phys. **10**, 581 (1980).
[2]B. Leaf, Found. Phys. **12**, 583 (1982).
[3]B. Leaf, Am. J. Phys. **47**, 811 (1979).
[4]Reference 1, Eq. (4), in coordinate representation.
[5]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), p. 210.
[6]H. Lass, *Elements of Pure and Applied Mathematics* (McGraw-Hill, New York, 1957), p. 98.
[7]$\boldsymbol{\xi}$ replaces $\mathbf{x}$ in the definition used in Ref. 3, Secs. III and IV.
[8]M. Spivak, *A Comprehensive Introduction to Differential Geometry* (Publish or Perish, Inc., Boston, 1970).
[9]W. Drechsler and M. E. Mayer, *Fiber Bundle Techniques in Gauge Theories*, Lecture Notes in Physics (Springer-Verlag, Berlin, 1977), Vol. 67.
[10]Reference 2, Eq. (24).

The effective potential in quantum mechanics[a)]

T. L. Curtright and C. B. Thorn
Physics Department, University of Florida, Gainesville, Florida 32611

(Received 7 December 1982; accepted for publication 11 March 1983)

General properties of the effective potential are discussed for quantum mechanical systems with a single degree of freedom. These properties are illustrated using specific one-dimensional potential models. In particular, it is stressed that the ground state for a system can exist even when the effective potential decreases monotonically towards a unique finite minimum at infinite $\langle x \rangle$.

PACS numbers: 03.65.Db, 11.10.Cd

I. INTRODUCTION

The effective potential V is a device widely used in quantum field theory to analyze global properties of the ground state (vacuum).[1] In general, $V(\phi)$ is defined as the minimum expectation value of the system's Hamiltonian, $V(\phi) = \min\langle \phi | H | \phi \rangle$, given the constraint that the field expectation value is held constant, $\phi = \langle \phi | \Phi(x) | \phi \rangle$. This constraint is invariably implemented in field theory through the Lagrange multiplier technique of introducing a linear coupling to a local external source.

The effective potential is not widely used to study nonrelativistic quantum mechanical models with a finite number of degrees of freedom. This is rightly so because the effective potential for a quantum mechanical system reveals very little of the system's complete dynamical content (as we shall see below). Several more refined methods usually exist which provide much more detailed information for systems with a finite number of degrees of freedom (e.g., direct numerical calculation of single-particle wave functions).

Nonetheless, by studying V for simple quantum mechanical systems, perhaps some intuition may be gained which can be used to clarify situations in field theory. Thus, in this paper we shall discuss some features of the effective potential for nonrelativistic systems with a single degree of freedom, x, whose dynamics are governed by the Schrödinger equation with an actual potential $U(x)$.

The paper is organized as follows. In Sec. II, we first define V and dispense with some trivial cases involving systems whose actual potentials are either localized or at least asymptotically sublinear. Then we consider less trivial cases where the actual potential is either asymptotically linear, or supralinear in at least one direction, or where U involves completely impenetrable regions. We prove some general theorems concerning the concavity and monotonicity of V, especially examining the effects of any impenetrable potential regions or bound states for U. In an appendix, we discuss other terms in the effective action which are needed to understand the physics of impenetrable potentials. Next, in Sec. III we discuss at length the asymptotic $\langle x \rangle$ behavior of V for various possible asymptotic x behaviors of the actual potential U. We relate such behavior to the scattering phase shifts at low momentum, paying particular attention to the effects of zero-energy solutions of the Schrödinger equation. In Sec. IV we discuss in detail some specific examples in order to illustrate some of the general properties of V. Those examples include actual potentials given by a simple harmonic oscillator, a linear potential, and a delta function located near an impenetrable wall. We also briefly consider "supersymmetric" potential models, which assign particular importance to states with zero energy. Finally, we close by commenting on the possible significance of our results for the two-dimensional Liouville quantum field theory.

II. DEFINITIONS AND GENERAL FEATURES

We shall consider nonrelativistic one-dimensional quantum mechanics for a single particle (with $m = \frac{1}{2}$) moving in an actual potential $U(x)$. Proceeding as in field theory, we define the effective potential V in the general case as

$$V(\langle x \rangle) = \min_{\{\psi\}} \langle \psi | H | \psi \rangle = \min_{\{\psi(x)\}} \int dx \left[\left| \frac{d\psi(x)}{dx} \right|^2 + U(x)|\psi(x)|^2 \right], \tag{1}$$

where we minimize over all wave functions $\psi(x)$ such that

$$\langle x \rangle = \int dx\, x |\psi(x)|^2, \quad 1 = \int dx |\psi(x)|^2. \tag{2}$$

The minimization procedure can always be carried out in principle by a thorough selection of trial wave functions. We shall sometimes employ this method of minimization in the following.

The minimization of $\langle H \rangle$ with $\langle x \rangle$ fixed can sometimes also be carried out as it is in field theory by using a Lagrange multiplier. We add to $U(x)$ the linear term Jx and solve the eigenvalue problem

$$\left(-\frac{d^2}{dx^2} + U(x) + Jx \right) \psi(x) = E(J) \psi(x). \tag{3}$$

We then choose $E(J)$ to be the lowest eigenvalue for this equation. As is well known, this procedure gives

$$\langle x \rangle = \frac{d}{dJ} E(J), \tag{4}$$

which implicitly specifies J as a function of $\langle x \rangle$, and

$$V(\langle x \rangle) = E(J) - J\langle x \rangle, \tag{5}$$

which yields the complement of (4),

$$J = -\frac{d}{d\langle x \rangle} V(\langle x \rangle). \tag{6}$$

Obviously, this Lagrange multipler method can be employed to calculate V only if the actual potential is either "linear" or

[a)] Work supported in part by the U.S. Department of Energy under Contract No. DE-AS-05-81-ER40008.

"supralinear" for asymptotic x, i.e., only if

$$\lim_{|x|\to\infty} U(x)/|x| > 0. \tag{7}$$

Otherwise, there is no ground state solution to (3). Actual potentials failing to satisfy (7) will be called "asymptotically sublinear." Such potentials require the use of the general prescription in (1) to compute V.

Let us first dispense with those trivial cases where the actual potential is asymptotically sublinear, i.e.,

$$\lim_{|x|\to\infty} U(x)/|x| = 0. \tag{8}$$

We do so by judiciously choosing a trial wave function

$$\psi(x) = \cos\theta\, \psi_1(x) + i \sin\theta\, \psi_2(x), \tag{9}$$

where ψ_1 and ψ_2 are both real wave packets and θ is a real parameter. For such a trial function, the hermiticity of H and x gives

$$\langle H\rangle = E_1\cos^2\theta + E_2\sin^2\theta = E_1 + (E_2 - E_1)\sin^2\theta, \tag{10}$$

$$\langle x\rangle = x_1\cos^2\theta + x_2\sin^2\theta = x_1 + (x_2 - x_1)\sin^2\theta, \tag{11}$$

where $(j = 1,2)$

$$x_j = \int dx\, x|\psi_j(x)|^2, \quad E_j = \int dx\, \psi_j^* H \psi_j. \tag{12}$$

Solving Eq. (11) for $\sin^2\theta$, we may write

$$\langle H\rangle = E_1 + (E_2 - E_1)(\langle x\rangle - x_1)/(x_2 - x_1). \tag{13}$$

Now for any $\epsilon > 0$ we may choose ψ_1 such that $E_1 < E_0 + \epsilon$, where E_0 is the greatest lower bound on the spectrum of H, and such that x_1 is finite. If we then take the limit $x_2/(\langle x\rangle - x_1)\to\infty$ with $\langle x\rangle \neq x_1$ fixed, but arbitrary, Eq. (13) gives

$$\langle H\rangle \xrightarrow[x_2/(\langle x\rangle - x_1)\to\infty]{} E_1 < E_0 + \epsilon.$$

The term involving E_2 becomes negligible in this limit due to the assumed localized character of the packet ψ_2 and the sublinearity of $U(x)$: $E_2/x_2 \approx [1/(\Delta x_2)^2 + U(x_2)]/x_2 \to 0$. Here Δx_2 is the width of the packet ψ_2. Thus we have shown that we can fix $\langle x\rangle$ arbitrarily and obtain $\langle H\rangle < E_0 + \epsilon$. From the general definition of $V(\langle x\rangle)$ in Eq. (1), we thus deduce the exact result

$$V(\langle x\rangle) = E_0, \quad \text{for all} \langle x\rangle, \tag{14}$$

when the actual potential is asymptotically sublinear. This simple result shows quite clearly that the exact effective potential may sometimes reveal practically nothing about the dynamical content of a model. Subsequently, we shall consider situations which are not so trivial as in (14).

We shall now establish the general result that V is concave upward. We shall use an argument similar to that which led to (14). (This result was previously established using a functional integration argument.[2]) Suppose that x_1 and x_2 are any two points where V is defined, with $x_1 < x_2$. From the definition of V, we know that for any $\epsilon > 0$ we can choose two real wave function packets ψ_1 and ψ_2 such that

$$E_i < V(x_i) + \epsilon, \tag{15}$$

where x_i and E_i are again given by (12). By choosing a linear combination of these packets, as in (9), it follows that Eqs. (10), (11), and (13) are again true, with $x_1 < \langle x\rangle < x_2$. From the general definition of V we then have

$$V(\langle x\rangle) \leqslant \langle H\rangle = \frac{(x_2 - \langle x\rangle)E_1 + (\langle x\rangle - x_1)E_2}{x_2 - x_1} < \frac{(x_2 - \langle x\rangle)[V(x_1) + \epsilon] + (\langle x\rangle - x_1)[V(x_2) + \epsilon]}{x_2 - x_1}. \tag{16}$$

Since this last inequality is true for all $\epsilon > 0$, it follows that V is concave upward.

A simple consequence of the concavity of V is that it limits the manner in which curve crossing can occur for different "V's" computed by varying J and following different stationary solutions of Eq. (3). Such "$V(\langle x\rangle)$'s" must cross tangentially because the true $V(\langle x\rangle)$ must always follow the lower of the two curves, and nontangential crossing would imply a forbidden convex cusp in V. Note further that Eq. (6) implies the intersection point of two such tangentially crossing "V" curves would correspond to equal J, as well as V and $\langle x\rangle$, and hence equal $E(J)$. For a one-dimensional system with an actual potential unbounded in one direction, every energy level is nondegenerate, assuming $U(x)$ is finite for finite x, so in this case such curve crossing is simply not allowed.

If $U(x)$ is not finite for finite x, degenerate energy levels are possible even in one dimension, and V appropriately exhibits some interesting behavior. This is the case if there are impenetrable, infinite potential regions at finite x. We now discuss an illustrative example involving one such impenetrable barrier. One can apply the Lagrange multiplier technique to each of the independent subsystems separated by the impenetrable barrier to obtain two disconnected $V(\langle x\rangle)$ curves computed by alternately taking vanishing wave functions either to the left or to the right of the barrier. The true $V(\langle x\rangle)$ for the combined system, however, must be computed using a superposition of two such wave functions. Let this superposition be given again by Eq. (9) with ψ_1 nonzero to the left of the barrier and ψ_2 nonzero to the right of the barrier. Then x_1 and x_2 as given by (12) will lie to the left and right of the barrier, respectively, and $\langle x\rangle$ for the superposition may be chosen to lie anywhere between x_1 and x_2, including within the impenetrable barrier.

The expectation of H for the superposition is again given as in Eq. (13) or (16). We may choose ψ_1 and ψ_2 to minimize E_1 and E_2 and obtain V_1 and V_2, the effective potentials for the separated subsystems. We may further choose x_1 and x_2 to minimize $\langle H\rangle$ for a fixed value of $\langle x\rangle$, with $x_1 < \langle x\rangle < x_2$. This requires

$$0 = \frac{\partial\langle H\rangle}{\partial x_1} = \frac{x_2 - \langle x\rangle}{(x_2 - x_1)^2}[V_1(x_1) - V_2(x_2)] + \frac{x_2 - \langle x\rangle}{x_2 - x_1}V_1'(x_1),$$

$$0 = \frac{\partial\langle H\rangle}{\partial x_2} = \frac{\langle x\rangle - x_1}{(x_2 - x_1)^2}[V_1(x_1) - V_2(x_2)] + \frac{\langle x\rangle - x_1}{x_2 - x_1}V_2'(x_2)$$

and thus specifies $V_1'(x_1) = V_2'(x_2) = [V_2(x_2) - V_1(x_1)]/(x_2 - x_1)$. This fixes x_1 and x_2 to be those points on the V_1 and

V_2 curves which have a common tangent given by the linear function of $\langle x \rangle$ in Eq. (13). The minimized $\langle H \rangle$ then gives for V

$$V(\langle x \rangle) = \begin{cases} V_1(\langle x \rangle) & \text{if } \langle x \rangle \leqslant x_1 \\ \dfrac{\langle x \rangle - x_1}{x_2 - x_1} V_2(x_2) + \dfrac{x_2 - \langle x \rangle}{x_2 - x_1} V_1(x_1) & \text{if } x_1 \leqslant \langle x \rangle \leqslant x_2. \\ V_2(\langle x \rangle) & \text{if } x_2 \leqslant \langle x \rangle \end{cases} \tag{17}$$

The effective potential for the combined system thus has a linear segment, extending across the impenetrable barrier.

At first encounter, this result appears to be paradoxical. The actual potential is impenetrable, but the effective potential exhibits absolutely no barrier, consistent with its being concave. To understand the physics of this situation, one must go beyond the effective potential and consider the nonstatic contributions to the effective action for the system. This is discussed in the Appendix.

III. ASYMPTOTIC BEHAVIOR

Henceforth, we always assume the actual potential is linear or supralinear for $x \to -\infty$, as defined in Eq. (7). The most interesting features of the effective potential will then be directly dependent on the behavior of $U(x)$ as $x \to +\infty$. We shall concentrate on how the asymptotic $\langle x \rangle \to +\infty$ behavior of V depends on that of U.

First consider the case where U is also linear, or supralinear, as $x \to +\infty$, and suppose U is finite for all finite x. Since U will dominate the kinetic energy of a localized packet for large $\langle x \rangle$, we shall apply the WKB approximation for $\langle x \rangle$ sufficiently large. The WKB quantization condition

$$\int_{x_{\text{low}}}^{x_{\text{high}}} dx \sqrt{E - U(x) - Jx} = \frac{\pi}{2} \tag{18}$$

then gives a corresponding approximation for $\langle x \rangle$,

$$\langle x \rangle = \int_{x_{\text{low}}}^{x_{\text{high}}} dy \frac{y}{\sqrt{E - U(y) - Jy}} \Bigg/ \int_{x_{\text{low}}}^{x_{\text{high}}} dy \frac{1}{\sqrt{E - U(y) - Jy}}. \tag{19}$$

If we now use the harmonic approximation of $U(x) + Jx$ about an assumed minimum x_m, we obtain

$$E(J) = U(x_m) + Jx_m + \sqrt{U''(x_m)}/2 + \cdots, \tag{20}$$

$$\langle x \rangle = x_m + \cdots. \tag{21}$$

Consequently, this will be a valid approximation giving negligible quantum corrections to the effective potential if

$$\frac{\sqrt{U''(x_m)}}{U(x_m) - x_m U'(x_m)} \ll 1.$$

More explicitly, if $U(x) \underset{x \to \infty}{\sim} x^p$, this condition becomes

$$x^{-1-p/2}/(1-p) \ll 1. \tag{22}$$

This is always satisfied as $x \to \infty$ if $p > 1$. Thus in this case we conclude that

$$V(\langle x \rangle) \underset{\langle x \rangle \to \infty}{\sim} U(\langle x \rangle). \tag{23}$$

Although the above argument fails for $p = 1$, we shall give an exact result later for the case of a linear potential which shows that the result in (23) still holds for that case [see Eqs. (47) and (48)]. Thus, if the actual potential is linear or supralinear, it provides the asymptotic form of the effective potential.

Next suppose that $U(x)$ is sublinear for large x, but still grows like a power, x^p, with $0 < p < 1$. In this case the Lagrange multiplier method fails to work for $J < 0$, and cannot be used to force $\langle x \rangle$ to be larger than x_0, the expectation in the $J = 0$ ground state of the sytem. Note that since the actual potential is unbounded in this case as $|x|$ becomes large, a normalizable ground state with energy E_0 and finite x_0 certainly exists. The general variation definition in Eq. (1) must be used to compute V for $\langle x \rangle > x_0$. An argument similar to that surrounding Eqs. (9)–(12) establishes that $V(\langle x \rangle > x_0) = E_0$ for this case.

For potentials that remain bounded at large positive x, we shall choose our zero of energy so that $U(x) \xrightarrow[x \to \infty]{} 0$. It follows that $V(\langle x \rangle)$ also asymptotes to zero. To be quantitative, let us consider potentials which behave like

$$U(x) \underset{\langle x \rangle \to \infty}{\sim} 1/x^p.$$

When $0 < p < 2$, the WKB analysis given above may then be applied to this class of potentials. If it is possible to obtain large values for $\langle x \rangle$ using the Lagrange multiplier J, then we conclude that Eq. (23) still holds. As above, it will not be possible to obtain arbitrarily large $\langle x \rangle$ using J if U possesses a bound state with energy $E_0 \leqslant 0$ and finite $\langle x \rangle = x_0$. To move to larger values of $\langle x \rangle$ would require $J < 0$, causing $U(x) + Jx$ to be unbounded below for large positive x. As in the case of the preceding paragraph, however, a variational argument establishes that $V(\langle x \rangle > x_0) = E_0$, where E_0 is again the lowest bound state energy.

In fact, it should be clear at this point that the effective potential always becomes a constant for $\langle x \rangle > x_0$ when the system admits a normalizable ground state with $\langle x \rangle = x_0$. Since this feature is completely insensitive to the local properties of the actual potential U, it provides another illustration of how V can be insensitive to the detailed dynamics of a model.

It should also be noted that if the actual potential is linear or supralinear for $x \to -\infty$, but sublinear for $x \to +\infty$, then the previous general result on the concave-upward character of the effective potential shows that $V(\langle x \rangle)$ will decrease monotonically towards its finite asymptotic value as $\langle x \rangle$ increases.

Continuing our discussion of bounded potentials, we next suppose that U goes to zero at least as fast as $1/x^2$, and we write

$$U(x) \underset{x \to \infty}{\sim} L(L+1)/x^2, \tag{24}$$

allowing the possibility of $L = 0$. If the nonleading asymptotic behavior of the actual potential satisfies the condition

$$[U(x) - L(L+1)/x^2]x^{2L+3} \xrightarrow[x \to \infty]{} 0, \tag{25}$$

then the $E=0, J=0$ solution of Eq. (3) has large-x behavior given by

$$\psi(x)_{E=0,J=0} \underset{x\to\infty}{\sim} a_1 x^{L+1} + a_2 x^{-L}, \tag{26}$$

where a_1 and a_2 will depend upon the specific details of the short-range part of U. On the other hand, the solution of (3) for large x will satisfy

$$\left(-\frac{d^2}{dx^2} + \frac{L(L+1)}{x^2} + Jx\right)\psi_{as}(x) = E(J)\psi_{as}(x), \tag{27}$$

which is simply the Airy differential equation generalized to include nonzero orbital angular momentum. In the limit of small J, and correspondingly small $E(J)$, the asymptotic solution in (27) should match onto the $E=0, J=0$ solution in (26) for a range of x: x should be large enough to neglect the short-range part of U, but much smaller than $J^{-1/3}$.

It facilitates further discussion to define

$$J = g^3, \quad E(g) = g^2 z(g), \quad \text{and} \quad y = gx. \tag{28}$$

In terms of these we have

$$\langle x\rangle = (1/3g)[gz'(g) + 2z(g)], \tag{29}$$

$$V(\langle x\rangle) = \tfrac{1}{3}g^2[z(g) - gz'(g)]. \tag{30}$$

The large-$\langle x\rangle$ behavior of V is then determined by the small-g behavior of $z(g)$. We also see that

$$\psi_{as}(x) = \phi(y), \tag{31}$$

were $\phi(y)$ satisfies

$$\left(-\frac{d^2}{dy^2} + \frac{L(L+1)}{y^2} + y\right)\phi(y) = z(g)\phi(y). \tag{32}$$

Thus we may write

$$\phi(y) = c_1(z)\phi_1(y) + c_2(z)\phi_2(y), \tag{33}$$

where the small y behavior for each of the two independent (assuming $L > -\frac{1}{2}$) solutions of (32) is given by

$$\phi_1(y) \underset{y\to 0}{\sim} y^{L+1}(1 + O(y^2)), \qquad \phi_2(y) \underset{y\to 0}{\sim} y^{-L}(1 + O(y^2)), \tag{34}$$

and where $c_1(z)/c_2(z)$ is fixed for each z so that $\phi(y) \xrightarrow[y\to\infty]{} 0$.

We now see that the required matching of the $E=0$, $J=0$ solution in (26) with $\psi_{as}(x)$, in the limit of vanishing g, determines the ratio of the coefficients in (33) in terms of the coefficients in (26). If $a_1 \neq 0$, then

$$c_2(z)/c_1(z) \underset{g\to 0}{\sim} g^{2L+1}a_2/a_1. \tag{35}$$

We conclude in this case that $z(g)$ approaches a zero of $c_2(z)/c_1(z)$ in the limit $g\to 0$.

The situation is more involved when $a_1 = 0$. To simultaneously treat all cases, we shall relate the small-g behavior of $z(g)$ to the low-energy behavior of the phase shifts for the potential $U(x)$, interpreting the asymptotic $L(L+1)/x^2$ term as the usual centrifugal barrier, but allowing L to take on continuous values. We also define the phase shifts in the usual way through the large-x behavior of the solution of (3) with $J=0$ [we assume $\delta_L(\infty) = 0$]:

$$\psi(x) \underset{J=0\ x\to\infty}{\sim} \sin\left[x\sqrt{E} - L\pi/2 + \delta_L(\sqrt{E})\right]. \tag{36}$$

Our discussion to follow makes use of some well-known results from single-channel scattering theory (e.g., see Ref. 3, especially Chaps. 11 and 12).

First note that the Jx term initially influences ϕ_1 and ϕ_2 at order y^3, so if $2L+1<3$ (i.e., $L<1$) a comparison of the scattering wave function to $\phi(y)$ in (33) yields

$$\frac{c_2(z)}{c_1(z)} \underset{g\to 0}{\sim} \frac{2^{2L+1}\Gamma(L+\frac{3}{2})\Gamma(L+\frac{1}{2})}{\pi z^{L+\frac{1}{2}}(\cot\delta_L(gz^{1/2}) - \cot\pi(L+\frac{1}{2}))}. \tag{37}$$

This result is contingent upon the nonleading asymptotic behavior of U satisfying Eq. (25).

Next, the zero-energy behavior of the phase shifts, which is needed to deal with the rhs of (37), is conveniently summarized by Levinson's theorem for continuous L. [Again this result can be proven if the condition in Eq. (25) holds.]

$$\frac{1}{\pi}\delta_L(0) = \begin{cases}(N+L+\frac{1}{2}), & a_1 = 0, \quad L<\frac{1}{2},\\ (N+1), & a_1 = 0, \quad L>\frac{1}{2},\\ N, & a_1 \neq 0.\end{cases} \tag{38}$$

Here N is the number of negative energy bound states. In view of the various cases in (38), we now consider the various possibilities for the effective potential.

If U admits any negative energy bound states, the effective potential is constant for $\langle x\rangle$ greater than the averaged position for the lowest bound state: $V(\langle x\rangle > x_0) = E_0$. We have previously argued this conclusion using trial wave functions. If $a_1 \neq 0$, and there are no negative energy bound states, then (38) gives

$$\cot\delta_L(g\sqrt{z}) \xrightarrow[g\to 0]{} \infty$$

and we have

$$c_2(z)/c_1(z) \to 0 \quad (a_1 \neq 0). \tag{39}$$

This confirms our previous result, Eq. (35). The possibility that c_2/c_1 remains finite while $z\to 0$ is ruled out, given the conditions in Eq. (25), because then $\delta_L(g\sqrt{z}) = O(g\sqrt{z})^{2L+1}$ as $g\to 0$.

If $a_1 = 0$ and $L<\frac{1}{2}$, Levinson's theorem implies that

$$c_1(z)/c_2(z)a^{L+\frac{1}{2}} \xrightarrow[g\to 0]{} 0.$$

In this case then

$$c_1(z)/c_2(z) \to 0 \quad (a_1 = 0, L<\tfrac{1}{2}). \tag{40}$$

We conclude in this case that $z(g)$ approaches a zero of $c_1(z)/c_2(z)$ in the limit $g\to 0$. If $a_1 = 0$ and $L>\frac{1}{2}$, Levinson's theorem implies once again that

$$\cot\delta_L(g\sqrt{z}) \xrightarrow[g\to 0]{} \infty.$$

However, since now $a_1 = 0$, it follows that $(g\sqrt{z})^{2l+1} \times \cot\delta_L(g\sqrt{z}) \to 0$, and hence the rhs of (37) can stay finite if $z \xrightarrow[g\to 0]{} 0$. Thus, we conclude that $z(g) \xrightarrow[g\to 0]{} 0$.

In all cases above where $z(g=0)$ is nonzero, the effective potential exhibits a universal $1/\langle x\rangle^2$ dependence for its asymptotic large positive $\langle x\rangle$ behavior. Explicitly,

$$V(\langle x\rangle) \underset{\langle x\rangle\to\infty}{\sim} \frac{4}{27}\frac{z^3(0)}{\langle x\rangle^2}, \tag{41}$$

where $z(0)$ is either the lowest zero of c_2/c_1 if $a_1\neq 0$, or the lowest zero of c_1/c_2 if $a_1=0$ and $L<\frac{1}{2}$.

If $a_1=0$ and $\frac{1}{2}<L<1$, given the conditions in Eq. (25), then

$$\cot\delta_L(k) \underset{k\to 0}{\sim} bk^{1-2L}$$

and therefore

$$z(g) \underset{g\to 0}{\sim} dg^{2L-1}, \tag{42}$$

where

$$d = \frac{2^{2L+1}\Gamma(L+\frac{3}{2})\Gamma(L+\frac{1}{2})}{\pi b}\frac{c_1(0)}{c_2(0)}.$$

The details of the short-range part of $U(x)$ determine the constant b (or d). In this case the effective potential behaves asymptotically as

$$V(\langle x\rangle) \underset{\langle x\rangle\to\infty}{\sim} \frac{2d(1-L)}{3}\left(\frac{d(2L+1)}{3\langle x\rangle}\right)^{(L+\frac{1}{2})/(1-L)}$$

$$(a_1=0,\quad \tfrac{1}{2}<L<1). \tag{43}$$

We emphasize that this formula only holds for $\frac{1}{2}<L<1$. For $L>1$ and $a_1=0$, the actual potential admits a zero-energy solution with $\langle x\rangle = x_0<\infty$, and so $V(\langle x\rangle > x_0)=0$ in this case.

Finally, we specialize to the case $L=0$. Then we have

$$c_1(z)/c_2(z) = \mathrm{Ai}'(-z)/\mathrm{Ai}(-z), \tag{44}$$

where Ai is the standard Airy function. Consequently, $V(\langle x\rangle)$ exhibits the behavior (41) with either $\mathrm{Ai}'(-z(0))=0$ if $a_1=0$, or $\mathrm{Ai}(-z(0))=0$ if $a_1\neq 0$. Referring to (25), the above analysis applies to this case only if $U(x)x^3\to 0$ as $x\to\infty$. However, if U has a power-law tail $1/x^p$, for $2<p<3$, one can treat the tail perturbatively as $g\to 0$ and again deduce that $-z(0)$ is either a zero of Ai or of Ai′. We shall say more about the $L=0$ case below in the context of an explicit example.

IV. SPECIFIC EXAMPLES

Some specific examples of simple potential models will serve to illustrate many of the general properties of the effective potential for quantum mechanical systems with a single degree of freedom. The most well-known example is that of the simple harmonic oscillator with actual potential

$$U(x) = x^2/4. \tag{45}$$

The corresponding effective potential is easily calculated using the linear Lagrange multiplier method, since $U(x)$ dominates Jx as $x\to+\infty$ or $-\infty$. The result is

$$V(\langle x\rangle) = \tfrac{1}{2} + \langle x\rangle^2/4 \tag{46}$$

and displays the familiar zero-point energy at the minimum, $\langle x\rangle = 0$.

A less familiar, but still exactly calculable example is that of a linearly rising potential attached to an impenetrable "brick wall" barrier. The actual potential is

$$U(x) = \begin{cases}\infty & \text{if } x<0\\ fx & \text{if } x>0,\end{cases} \tag{47}$$

where $f>0$. The corresponding effective potential is again calculable by the Lagrange multiplier method, provided that $J>-f$. All positive $\langle x\rangle$ expectation values are still obtainable by varying J. It is then straightforward to find

$$V(\langle x\rangle) = \begin{cases}\infty & \text{if } \langle x\rangle<0\\ f\langle x\rangle + 4z^3/27\langle x\rangle^2 & \text{if } \langle x\rangle>0,\end{cases} \tag{48}$$

where $-z = -2.338\ 107\ 41\cdots$ is the first zero of the Airy function, Ai. The form of $V(\langle x\rangle)$ follows straightforwardly from the lowest energy solution of the Schrödinger equation (3) in the presence of the linear potential plus external source $(f+J)x$, subject to the conditions that the wave function vanish at $x=0$ and $x=\infty$. That wave function solution is

$$\mathrm{Ai}(x(f+J)^{1/3}-z)$$

and has energy

$$E = z(f+J)^{2/3}.$$

It is remarkable that the result in Eq. (48) is simply the sum of the actual linear potential plus the effective potential for the impenetrable barrier alone (found by setting $f=0$).[4]

Another specific choice for U nicely illustrates many other general theorems, especially those concerning the effects of zero-energy states on the behavior of V. Consider the case where U is a delta-function potential situated near an impenetrable barrier.

$$U(x) = \begin{cases}\infty, & x<0,\\ -s\delta(x-1), & x>0.\end{cases} \tag{49}$$

The Lagrange multiplier method with $J>0$ may be used to compute $V(\langle x\rangle)$. Different qualitative behavior for the effective potential is determined by different ranges for the strength of the delta function, s, in accord with our general discussion. The results for various values of s are shown in the figure. For sufficiently small but positive $\langle x\rangle$, i.e., large positive J, the delta function is always insignificant and V approaches the general $1/\langle x\rangle^2$ form given in Eq. (41) [i.e., Eq. (48) with $f=0$]. For large values of $\langle x\rangle$, i.e., small positive J, the form of V is dependent on the occurrence of normalizable or bounded zero-energy solutions of the Schrödinger equation, (3), with $J=0$.

If $s>1$, the actual potential is sufficiently attractive for the existence of a single normalizable bound state with $E_B<0$ and finite $\langle x\rangle_B$. The effective potential decreases monotonically as $\langle x\rangle$ increases until $\langle x\rangle_B$ is reached, beyond which V is a constant, E_B. This is illustrated in the figure (curve H) for $s=2$, for which $E_B=-0.634\ 9095$, $\langle x\rangle_B = 1.342\ 284$.

If $s=1$, there is a bounded, zero-energy solution of the $J=0$ Schrödinger equation: $\psi_0(x)=x$ if $0\leqslant x\leqslant 1$, $\psi_0(x)=1$ if $x\geqslant 1$. In this case V decreases monotonically, as $\langle x\rangle$ increases, to the asymptotic form given in Eq. (41) with $-z(0) = -1.018\ 792\ 97\cdots$, the first zero of Ai′. We wish to emphasize that the effective potential in this case still has a

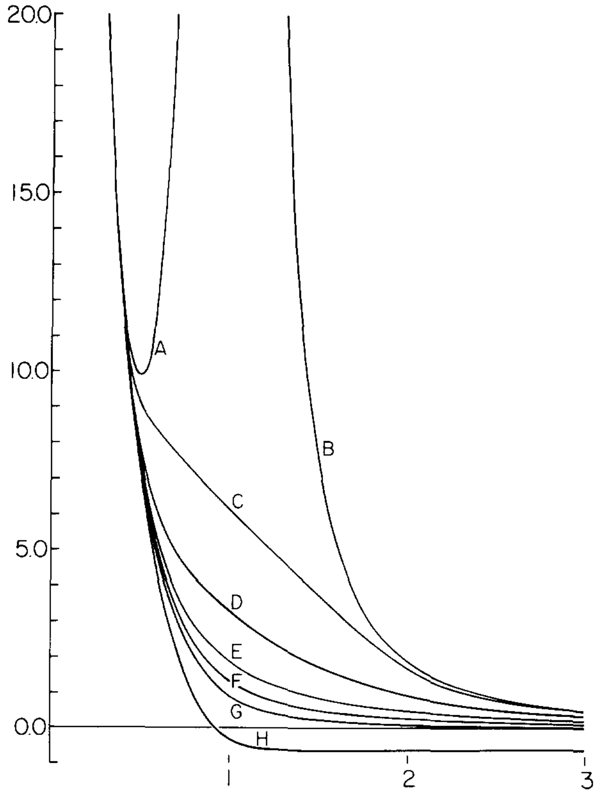


FIG. 1. $V(x)$ vs x for barrier plus δ function of strengths: (A) $s = -\infty$ (inside); (B) $s = -\infty$ (outside); (C) -20; (D) -2; (E) 0; (F) 0.6; (G) 1; (H) 2.

unique minimum, albeit at infinite $\langle x \rangle$. Again this case is shown in the figure (curve G).

If $s < 1$, neither normalizable nor bounded zero-energy solutions exist for the $J = 0$ Schrödinger equation. For increasing $\langle x \rangle$ the monotonically decreasing effective potential will once more approach the asymptotic form given in Eq. (41) with $-z(0)$ equal to the first zero of Ai. This is shown in the figure for the cases $s = 0.6, s = 0, s = -2$, and $s = -20$. Of course, for $s = 0$, Eq. (41) gives the exact form for V [see (48)]. For repulsive delta functions ($s < 0$), the effective potential is raised above the $s = 0$ case in the vicinity of the repulsive spike. Note that for an extremely repulsive (but finite s) case (e.g., curve C in the figure) the Lagrange multiplier calculation of V begins to approximate the linear segment which would be simultaneously tangent to the disconnected effective potentials for the two independent subsystems "inside" and "outside" of an impenetrable delta function ($s = \infty$). Those "inside" and "outside" subsystem effective potentials are shown in the figure as curves A and B. Recall the linear segment for the impenetrable delta function was established above, without detailed calculation, by using a simple variational argument.

The effective potentials shown in the figure were obtained using numerical methods supplemented by the following analytical results. The ground state wave function in the presence of the linear source term has the form

$$\psi(x) = \begin{cases} \mathrm{Ai}(gx - z(g))\mathrm{Bi}(-z(g)) & \text{if } 0 \leqslant x \leqslant 1 \\ -\mathrm{Ai}(-z(g))\mathrm{Bi}(gx - z(g)) & \\ K\,\mathrm{Ai}\,(gx - z(g)) & \text{if } x \geqslant 1, \end{cases} \tag{50}$$

where K is a constant, $J = g^3$, $E(J) = g^2 z(g)$, and Ai and Bi are the standard Airy functions. The continuity of ψ and the required discontinuity of $d\psi/dx$ at $x = 1$ fix the constant K and require that $z(g)$ be the lowest solution of the following equation:

$$\begin{aligned} &\mathrm{Ai}(-z(g))\mathrm{Bi}(g - z(g)) - \mathrm{Bi}(-z(g))\mathrm{Ai}(g - z(g)) \\ &\quad = (g/\pi s)\mathrm{Ai}(-z(g))/\mathrm{Ai}(g - z(g)). \end{aligned} \tag{51}$$

Given that $z(g)$ is such a solution, the expectation value $\langle x \rangle$ is determined using Eq. (29) with

$$z'(g) = \frac{s - 1 + 2g(d/dg)\ln[\mathrm{Ai}(g - z(g))]}{s - s\,\mathrm{Ai}^2(g - z(g))/\mathrm{Ai}^2(-z(g)) + 2g(d/dg)\ln[\mathrm{Ai}(g - z(g))]} \tag{52}$$

Combining these results, one obtains $V(\langle x \rangle)$ through the use of Eq. (30).

Zero-energy states play a natural role in "supersymmetric" potential models. As a final illustration of our general results on the effective potential, we briefly discuss a supersymmetric example.

In general, supersymmetric quantum mechanical potential models are defined by Hamiltonians of the form

$$H = Q^2, \tag{53}$$

where

$$Q = \sigma_1\left(-i\frac{d}{dx}\right) + \sigma_2 W(x) \tag{54}$$

and $W(x)$ is an arbitrary function.[5] H acts on a two-component wave function whose upper/lower components satisfy Schrödinger's equation with actual potentials given by

$$U_{\pm}(x) = W^2(x) \pm \frac{dW(x)}{dx}. \tag{55}$$

Since H is the square of an Hermitian operator, it follows that the energy spectrum is positive.

As with their field theoretic counterparts, a fundamental issue for supersymmetric quantum mechanics models is whether there is a zero-energy ground state. This can only happen if the ground state is supersymmetric, i.e.,

$$Q\psi_0(x) = 0. \tag{56}$$

If the ground state energy is nonzero, then Q does not annihilate the ground state, and supersymmetry is "spontaneously broken."

The formal solution to (56) is

$$\psi_0(x) = \left\{\exp\left[\int_0^x dy\; W(y)\sigma_3\right]\right\}\psi_0(0). \tag{57}$$

This is a valid solution if $\psi_0(x)$ remains finite for all x, and in that case we say ψ_0 exists. It is now straightforward to classify those functions $W(x)$ for which $\psi_0(x)$ remains finite[5] by examining the behavior of the exponential in (57).

Let us now consider the effective potential for the system defined by (53). Our general results hold. In particular, consider the simple case

$$W(x) = e^{-x}. \tag{58}$$

For this case a zero-energy ground state exists. [In view of the results in Ref. 5, $W(x)$ has a single zero at infinity.] That ground state is

$$\psi_0(x) = \begin{pmatrix} e^{-e^{-x}} \\ 0 \end{pmatrix}, \tag{59}$$

which dies rapidly for $x \to -\infty$, but asymptotes to $\binom{1}{0}$ as $x \to +\infty$. Thus $\psi_0(x)$ is not square integrable, but has the same boundedness properties as ordinary states in the continuum (cf. plane waves). Correspondingly, the effective potential behaves asymptotically as in Eq. (23), as $\langle x \rangle \to -\infty$, and as in Eq. (41), as $\langle x \rangle \to +\infty$, with $-z(0)$ again equal to the first zero of Ai′. Hence V decreases monotonically to a unique minimum $(=0)$ at $\langle x \rangle = +\infty$.

V. CONCLUSION

In conclusion, we wish to discuss a problem in field theory which provided the motivation for this analysis of the quantum mechanical effective potential. Recently the two-dimensional Liouville field theory defined by

$$\mathscr{L}(\phi) = -\tfrac{1}{2}(\partial\phi)^2 - (2m^2/g^2)e^{2g\phi} \tag{60}$$

has received considerable attention in connection with the relativistic string. The quantum correlation functions of the Liouville theory are needed to determine the scattering amplitudes of the quantum string. An operator analysis of the quantum Liouville theory has been given in pursuit of those correlation functions (see Ref. 6 and the references cited therein).

One of the results established by that operator analysis was that the energy eigenvalue spectrum of the quantum Liouville theory is continuous, including all $E > 0$. This spectrum is a direct consequence of the conformal invariance of the Lagrangian in (60). It then became a subtle problem to determine whether or not the state $|E = 0\rangle$ also exists, in the sense that the configuration space wave functional is bounded, as is true of the other members of the continuous spectrum, and in analogy with the existence of the ground state wave functions for the simple quantum mechanical models we have discussed above. While the detailed operator analysis in Ref. 6 is not quite complete enough to answer this question absolutely, there seems to be no reason in Ref. 6 to exclude $E = 0$. In fact, since the results in Ref. 6 provide an operator map identifying the Liouville and free pseudoscalar field Hamiltonians, the identification of the $E = 0$ state for the Liouville quantum field theory with the usual free pseudoscalar field $E = 0$ state is very strongly suggested.

The existence of the $E = 0$ state was challenged,[7] however, partly because of the structure of the exact effective potential for the Liouville model. The exact effective potential was computed by J. Goldstone.[8] It was shown to have the same exponential form as the actual potential in Eq. (60), namely

$$V(\phi) = \text{const}\ e^{2g\phi/(1+g^2/2\pi)}. \tag{61}$$

Thus a translationally invariant ground state for the quantum Liouville theory, corresponding to the minimum of this effective potential, would have infinite field expectation value. While this appears to be an unusual feature for the ground state of a quantum field theory, our results for simple quantum mechanical models show that it is not an impossible situation. In order to determine if a state actually exists with this property, whose wave functional is bounded as are the wave functionals for the other continuum states, it is necessary to investigate the dynamics of the theory more carefully. A cursory inspection of V is not adequate. We believe that further operator analysis, along the lines set out in Ref. 6, will eventually answer this question.

APPENDIX

Since the effective potential V is a concave function, it does not exhibit any barriers separating local minima, even though the actual potential U might have such a barrier. How then can one understand the phenomena of tunneling? What exhibits the inhibitory effect on wave packets that an actual potential barrier produces? An understanding of this effect requires consideration of the full effective action, not just the static effective potential but also nonstatic terms involving $dx(t)/dt$.

Recall[1] that the full quantum mechanical evolution of $x(t)$, in the presence of a time-dependent source $J(t)$, is governed by the effective action

$$\Gamma[x] = W[J] + \int_{-\infty}^{\infty} dt\, x(t)J(t), \tag{A1}$$

where W may be determined either by using a functional average or by taking an operator expectation value.

$$\begin{aligned} e^{iW[J]} &= \int \mathscr{D}x \exp\{i\textstyle\int dt\,[\tfrac{1}{4}\dot{x}^2(t) - U(x(t)) - x(t)J(t)]\} \\ &= \langle 0|T(\exp[-i\textstyle\int dt\, X(t)J(t)])|0\rangle. \end{aligned} \tag{A2}$$

We have denoted time ordering by T, represented the coordinate operator by X, and used $|0\rangle$ to represent the ground state of the $J = 0$ system. For simplicity here we assume $\langle 0|0\rangle = 1$ and suppose the energy spectrum is discrete.

In the limit of a static source, $W[J]$ and $\Gamma[x]$ reduce to $E(J)$ and $V(x)$, respectively,

$$\begin{aligned} W[J] &\xrightarrow[J(t)\to J]{} -E(J)\cdot\textstyle\int dt, \\ \Gamma[x] &\xrightarrow[x(t)\to\langle x\rangle]{} -V(\langle x\rangle)\cdot\textstyle\int dt, \end{aligned} \tag{A3}$$

and (A1) reduces to Eq. (5) of the text.

In standard fashion, one can obtain the expectation value and all the higher correlation functions of $X(t)$ in the presence of the source $J(t)$ by taking functional derivatives of (A2). Functional derivatives of (A1) then give the well-known inverses. For example,

$$x(t) = -\frac{\delta W[J]}{\delta J(t)} \tag{A4}$$

$$= \frac{\langle 0|T(e^{-i\int dt' X(t')J(t')}X(t))|0\rangle}{\langle 0|T(e^{-i\int dt'' X(t'')J(t'')})|0\rangle},$$

$$-iD(t_1,t_2) \equiv -i\frac{\delta^2 W[J]}{\delta J(t_1)\delta J(t_2)} = i\frac{\delta x(t_2)}{\delta J(t_1)} \tag{A5}$$

$$= \frac{\langle 0|T(e^{-i\int dt X(t)J(t)}X(t_1)X(t_2))|0\rangle}{\langle 0|T(e^{-i\int dt' X(t')J(t')})|0\rangle} - x(t_1)x(t_2),$$

$$J(t) = \frac{\delta \Gamma[x]}{\delta x(t)}, \tag{A6}$$

$$D(t_1,t_2)^{-1} = -\frac{\delta^2\Gamma[x]}{\delta x(t_1)\delta x(t_2)} = -\frac{\delta J(t_2)}{\delta x(t_1)}. \tag{A7}$$

Let us now consider a "quasistatic" situation where $J(t)$ is a constant J for large $|t|$ and varies very slowly for all t. The phase in (A2) then singles out the ground state of the system, $|0\rangle_J$, in the presence of a constant source, and (A5) gives

$$-iD(\omega) \equiv -i\int_{-\infty}^{\infty} dt\, e^{-i\omega t}D(t,0)$$

$$= \int_{-\infty}^{\infty} dt\, e^{-i\omega t}({}_J\langle 0|T(X(t)X(0))|0\rangle_J - {}_J\langle 0|X|0\rangle_J^2)$$

$$= \sum_{n\neq 0}\int_{-\infty}^{\infty} dt\, e^{-i\omega t}(\theta(t)\exp\{i[E_0(J) - E_n(J)]t\} + \theta(-t)\exp\{-i[E_0(J) - E_n(J)]t\})|{}_J\langle n|X|0\rangle_J|^2$$

$$= 2i\sum_{n\neq 0}\frac{[E_n(J) - E_0(J)]\,|{}_J\langle n|X|0\rangle_J|^2}{\omega^2 - [E_n(J) - E_0(J) - i\epsilon]^2}. \tag{A8}$$

Here $E_n(J)$ are the energy levels in the presence of the constant source. In the quasistatic case, as $\omega \to 0$, we therefore have

$$-iD(\omega) = -2i\sum_{n\neq 0}\frac{|{}_J\langle n|X|0\rangle_J|^2}{E_n(J) - E_0(J)} - 2i\omega^2\sum_{n\neq 0}\frac{|{}_J\langle n|X|0\rangle_J|^2}{[E_n(J) - E_0(J)]^3} + \mathscr{O}(\omega^4). \tag{A9}$$

Now in the same limit the effective action should have the form

$$\Gamma|x| \sim \int dt\,[-V(\langle x\rangle) + \tfrac{1}{4}M(\langle x\rangle)\dot{x}^2], \tag{A10}$$

where we have dropped all terms involving four or more time derivatives and $\langle x\rangle \equiv {}_J\langle 0|X|0\rangle_J$. Using this, (A7) gives

$$D^{-1}(\omega) = V''(\langle x\rangle) - \tfrac{1}{2}M(\langle x\rangle)\omega^2 + \mathscr{O}(\omega^4). \tag{A11}$$

Comparing (A9) and (A11), we obtain

$$[V''(\langle x\rangle)]^{-1} = 2\sum_{n\neq 0}\frac{|{}_J\langle n|X|0\rangle_J|^2}{E_n(J) - E_0(J)}, \tag{A12}$$

$$M(\langle x\rangle) = 4[V''(\langle x\rangle)]^2\sum_{n\neq 0}\frac{|{}_J\langle n|X|0\rangle_J|^2}{[E_n(J) - E_0(J)]^3}. \tag{A13}$$

These are exact results, given the limiting form of the effective action in (A10).

The result for V'' in (A12) clarifies the linear behavior exhibited by the effective potential when the actual potential has an impenetrable region, as displayed in Eq. (17), and the result for M clarifies the inhibiting effects on wave packets produced by a large barrier in the actual potential. Large barriers tend to reduce energy differences, for a particular value of J, thus causing V'' to become small and M to become large.

Denote the energy splitting between the ground state and the first excited state, in the presence of a constant source, by

$$\Delta E = E_1(J) - E_0(J). \tag{A14}$$

Then (A12) and (A13) yield the inequalities

$$\Delta E/\Delta x^2 \leqslant 2V'' \leqslant \Delta E/x_{10}^2, \tag{A15}$$

$$\Delta E\cdot\Delta x^4/x_{10}^2 \leqslant M^{-1} \leqslant \Delta E\cdot x_{10}^4/\Delta x^2, \tag{A16}$$

where

$$x_{10}^2 \equiv |{}_J\langle 1|X|0\rangle_J|^2, \tag{A17}$$

$$\Delta x^2 \equiv \sum_{n\neq 0}|{}_J\langle n|X|0\rangle_J|^2 = {}_J\langle 0|X^2|0\rangle_J - {}_J\langle 0|X|0\rangle_J^2. \tag{A18}$$

In the limit that the ground state becomes degenerate with the first excited state, with finite and nonzero Δx^2 and x_{10}^2, the inequalities in (A15) and (A16) imply that $V'' \to 0$ and $M \to \infty$. This degenerate situation is achieved, for a particular value of J, if the actual potential consists of two regions separated by a repulsive barrier in the limit where that barrier becomes impenetrable.

The motion of a wave packet in the presence of the actual potential U can be understood by considering the motion of a classical particle moving under the influence of the effective action. A classical particle governed by the action in (A10) has a conserved energy $\frac{1}{4}M(x)\dot{x}^2 + V(x)$. For a fixed energy E, such a particle will display varying $\dot{x}$ dependent on its position, in the usual way, according to

$$\tfrac{1}{4}\dot{x}^2 = [E - V(x)]/M(x). \tag{A19}$$

Thus the particle will slow down either as $E - V$ becomes small or as M becomes large. The particle has a turning point ($\dot{x} = 0$) either when $E = V$ or when M diverges, the latter being characteristic of the effective action in quantum mechanical models as discussed in this paper.

[1] Two elementary but thorough discussions of the effective potential in the context of quantum field theory are: S. Coleman, in *Laws of Hadronic Matter*, Proceedings of the 1973 International Summer School "Ettore Majorana," Erice, Italy, edited by A. Zichichi (Academic, New York, 1975), and J. Iliopoulos, C. Itzykson, and A. Martin, Rev. Mod. Phys. **47**, 165 (1975).

[2] K. Symanzik, Commun. Math. Phys. **16**, 48 (1970).

[3] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).

[4] The $f = 0$ case has also been discussed by E. D'Hoker and R. Jackiw, Phys. Rev. D **26**, 3517 (1982).

[5] E. Witten, Nucl. Phys. B **188**, 513 (1981).

[6] T. L. Curtright and C. B. Thorn, Phys. Rev. Lett. **48**, 1309 (1982). E. Braaten, T. Curtright, and C. Thorn, Phys. Lett. **118B**, 115 (1982); E. Braaten, T. Curtright, and C. Thorn, Ann. Phys. (N.Y.) **147**, 365 (1983).

[7] E. D'Hoker and R. Jackiw, Ref. 4.

[8] J. Goldstone, unpublished (as quoted in Ref. 4).

Instantons in quantum mechanics: Numerical evidence for a conjecture

J. Zinn-Justin
Commissariat a L'energie Atomique, Division de la Physique, CEN Saclay, 91191 Gif-sur-Yvette Cedex, France

(Received 20 January 1983; accepted for publication 4 March 1983)

In previous articles, we have given a proper definition of multi-instanton contributions in quantum mechanics and calculated these contributions to leading order. We have also presented a conjecture about the form of the expansion of these multi-instanton contributions to all orders in powers of the coupling constant. We give here some numerical results which support this conjecture.

PACS numbers: 03.65.Ge

I. INTRODUCTION

In a previous article,[1] we have presented the following conjecture[2]. Let $V(x)$ be an analytic potential with degenerate and symmetric minima

$$V(x) = \tfrac{1}{2}x^2 + O(x^3), \quad V(x) = V(x_0 - x). \tag{1}$$

We write the Schrödinger equation as

$$-\tfrac{1}{2}\psi''(x) + (1/g)V(x\sqrt{g})\psi(x) = E\psi(x). \tag{2}$$

With these conventions, the one-instanton contribution to the ground-state energy is, at leading order,

$$E^{(1)}(g) \sim -\lambda = -\sqrt{C/\pi g}\, e^{-a/g} \tag{3}$$

with

$$a = \int_0^{x_0} \sqrt{2V(x)}\, dx, \tag{4}$$

$$C = x_0^2 \exp\int_0^{x_0} dx\left[\frac{1}{\sqrt{2V(x)}} - \frac{1}{x} - \frac{1}{x_0 - x}\right]. \tag{5}$$

Consider now the two coupled equations for the two unknowns E and s:

$$s = E - \frac{1}{2} + \sum_{k=1}^{\infty} R_k(E) g^k, \tag{6}$$

$$\Gamma(-s)\mu^s \lambda e^{-A(E,g)} = \pm 1. \tag{7}$$

The sign $+$ corresponds to even states and the minus sign to odd states. The parameters are defined by

$$A(E,g) = \sum_{k=1}^{\infty} Q_k(E) g^k, \tag{8}$$

$$\mu = -2C/g. \tag{9}$$

The coefficients Q_k and R_k are polynomials of degree $k+1$ in E. From Eq. (7), we see that when g is small λ is small, and thus s must be close to a pole of the Γ function:

$$s = N + O(\lambda) \Rightarrow E = N + \tfrac{1}{2} + O(g, \lambda). \tag{10}$$

If we then expand systematically s and E in powers of λ and then the coefficients of λ^n in powers of g, we obtain the complete perturbative expansion to all orders, taking into account all instanton contributions to the Nth states. It has has the form[3,4]

$$E_N(g) = \sum_{k=0}^{\infty} E_{0,l}^{(N)} g^l + \sum_{n=1}^{\infty} \left(\frac{C}{\pi g}\right)^{n/2} e^{-na/g} \times \sum_{k=0}^{n-1} (\ln(-2C/g))^k \sum_{l=0}^{\infty} \epsilon_{n,k,l}^{(N)} g^l. \tag{11}$$

The various perturbation series have to be summed for g complex first and then continued to g real positive consistently with the function $\ln(-2C/g)$. The imaginary contributions coming both from the Borel sums and the logarithms cancel.

The implication of the conjecture as expressed by Eqs. (6) and (7) is that it is sufficient to know the perturbative and the one-instanton contributions to all orders and for all states, in order to be able to calculate the many-instanton contributions to all states and to all orders.

Equations (6) and (7) have been derived at leading order[1,3,4] and some numerical evidence for the conjecture can already be found in Ref. 3. We shall see that some consequences of this conjecture agree with the results of Damburg and Propin.[5,6] It is even possible that their methods can be extended to prove it.

We shall also present a large number of numerical data for the double-well potential and the periodic cosine potential, which confirm the conjecture up to order g^2.

II. THE PERTURBATIVE EXPANSION

We shall first explain how we can calculate numerically the perturbative coefficients in Eq. (6). We shall present the method for a general analytic potential and then show how the recursion formulas simplify for the double-well and the cosine potentials.

We shall use the standard trick of transforming the Schrödinger equation into a Riccati equation[7,8] by setting

$$\psi(x) = \exp\left(-\frac{1}{g}\int_0^{x\sqrt{g}} S(x'\sqrt{g})dx'\right), \tag{12}$$

which leads to

$$S' + (1/g)(2V(x) - S^2) = 2E. \tag{13}$$

We shall set

$$U = \sqrt{2V(x)} = x + O(x^2) \tag{14}$$

 0022-2488/84/030549-07$02.50

and expand S in powers of g:

$$S(x) = U(x) - \sum_{k=1}^{\infty} g^k S_k(x). \tag{15}$$

At order 1 we get

$$S_1 = U^{-1}(E - \tfrac{1}{2}U'), \tag{16}$$

and then

$$k>1, \quad 2US_k = S'_{k-1} + \sum_{l=1}^{k-1} S_l S_{k-l}. \tag{17}$$

We expand now $U(x)$ in powers of x;

$$U(x) = x + \sum_{r=2}^{\infty} U_r x^r, \tag{18}$$

and S_k in powers of x and E,

$$S_k = \sum_{\substack{r \geqslant 0 \\ 0 \leqslant s \leqslant k}} S^k_{r,s} x^{-2k+1+r} E^s. \tag{19}$$

Equation (17) generates then a set of recursion formulas

$$S^k_{r,s} = -\sum_{p=0}^{r-1} S^k_{p,s} U_{r+1-p} + \frac{1}{2} S^{k-1}_{r,s}(r+3-2k) + \frac{1}{2}\sum_{\substack{l=1 \\ p,q}}^{k-1} S^l_{p,q} S^{k-l}_{r-p,s-q}, \tag{20}$$

which allow us to calculate systematically $S(x)$.

The coefficients of the perturbative expansion (6) are obtained by observing that the wave function should be uniform at $x=0$. As a consequence, the coefficient of $\ln x$ in $\ln \psi(x)$ should be an integer. This means that the coefficient of $1/x$ in $S(x)$ should also be an integer. Setting

$$R_k(E) = \sum_{s=0}^{k+1} R^k_s E^s, \tag{21}$$

we obtain the relation

$$R^k_s = S^{k+1}_{2k,s}. \tag{22}$$

For example, the first terms are

$$R_0(E) = E - \tfrac{1}{2},$$

$$R_1(E) = 3(U_2^2 - \tfrac{1}{2}U_3)E^2 + \tfrac{1}{8}(2U_2^2 - 3U_3). \tag{24}$$

To obtain usual perturbation theory, one inverts the relation

$$N = E - \tfrac{1}{2} + gR_1(E) + \dots . \tag{25}$$

A. An algebraic property

The polynomials $R_k(E)$ are even or odd for k odd or even, respectively. To prove this property, let us decompose S in a sum of two terms:

$$S = S^+ + S^- \tag{26}$$

with

$$S^+(-g,-E) = S^+(g,E), \tag{27}$$

$$S^-(-g,-E) = -S^-(g,E). \tag{28}$$

Equation (13) leads then to two equations

$$S'^+ - (2/g)S_+S_- = 0, \tag{29}$$

$$S'^- + (1/g)(2V - S^2_+ - S^2_-) = 2E. \tag{30}$$

The first equation allows us to calculate S_+:

$$S_+ = \frac{g}{2}\frac{d}{dx}\ln S_-. \tag{31}$$

If we now expand S^- in powers of g, only the term of order 0 will have a logarithm:

$$\ln S_- = \ln x + O(x,g). \tag{32}$$

Therefore, S_+ gives a contribution to expansion (25) only at order 0 and is responsible for the term $-\tfrac{1}{2}$.

B. A special family of potentials

The double-well potential and the cosine potential belong to a special family of potentials for which the perturbative expansion in the form (15) is specially simple. We have noted this already before and used this property to make high-order calculations of the one-instanton contribution.[7]

Setting as before

$$U(x) = \sqrt{2V(x)},$$

we shall assume that $U(x)$ is the solution of the equation

$$U'^2 + 4U^m = 1 \tag{33}$$

(the 4 is a convenient normalization), in which m is a positive integer. For $m=1$, the solution is the double-well potential:

$$U(x) = x(1-x), \quad V(x) = \tfrac{1}{2}x^2(1-x^2). \tag{34}$$

For $m=2$, we obtain the cosine potential:

$$U(x) = \tfrac{1}{2}\sin 2x, \quad V(x) = \tfrac{1}{16}(1-\cos 4x). \tag{35}$$

Higher values of m correspond to elliptic functions. Only $m=1$ and 2 correspond to entire functions for the potential. For m even, the potentials are periodic and for m odd, they are of double-well type.

We shall now rewrite the recursion formula (17) by setting

$$S_k = T_k / 2^k U^{2k-1} \tag{36}$$

and taking as a new variable z,

$$z = -U'(x). \tag{37}$$

Equation (33) implies

$$UU'' = -(m/2)(1-z^2). \tag{38}$$

The resulting expression is

$$k>1, \quad T_k = (m/2)(1-z^2)T'_{k-1} + (2k-3)zT_{k-1} + 2\sum_{l=1}^{k-1} T_l T_{k-l} \tag{39}$$

and

$$T_1 = 2E + z. \tag{40}$$

It is easy to verify that T_k is a polynomial of degree k globally in the variables z and E such that

$$T_k(-z,-E) = (-1)^k T_k(z,E). \tag{41}$$

This leads to simple recursion formulas for the coefficients $T^k_{r,s}$:

$$T_k = \sum_{r+s \leqslant k} T^k_{r,s} z^r E^s. \tag{42}$$

To complete the calculation, one needs to find the residues

C_k^r of quantities of the form

$$C_k^r = \text{res}[U'^{2r}U^{-(2k-1)}]. \tag{43}$$

To calculate this residue, let us write the identity

$$\frac{d}{dx}\left[\frac{U'^r}{U^s}\right] = r\frac{U''U'^{r-1}}{U^s} - s\frac{U'^{r+1}}{U^{s+1}}. \tag{44}$$

Using then Eq. (38), we obtain

$$\frac{d}{dx}\left[\frac{U'^r}{U^s}\right] = -\frac{mr}{2}(1-U'^2)\frac{U'^{r-1}}{U^{s+1}} - s\frac{U'^{r+1}}{U^{s+1}}. \tag{45}$$

Since the derivative has no residue, we obtain a recursion formula

$$C_n^r = \frac{m(2r-1)}{m(2r-1)-4(k-1)}C_n^{r-1}. \tag{46}$$

It remains now to calculate C_m^0. To do this, we use Eq. (33) in the form

$$\frac{U'^2}{U^s} + \frac{4}{U^{s-m}} = \frac{1}{U^s}. \tag{47}$$

Taking the residues of both members and using the consequences of Eq. (45) for $r=1$, we obtain the following relation:

$$\text{res}[U^{-s}] = \frac{4(s-1-m/2)}{s-1}\text{res}[U^{m-s}]. \tag{48}$$

The recursion formulas (46) and (48) allow us to calculate the perturbative coefficients for the potentials solutions of Eq. (33). Of course, the method can be generalized for various polynomial relations between U and U'.

As a verification of our programs, we have compared the perturbation expansions for the double-well potential and the $O(2)$ anharmonic oscillator [using a suitable modification of formula (20)]. We have verified as expected[9,10] that the series are identical after the substitution

$$E[O(2)] = 2E[\text{d.w.}].$$

More precisely, the series are identical up to order 10 and from order 10–19 the relative difference is smaller than 10^{-21}, which is consistent with the numerical accuracy.

C. The one-instanton contribution

The algorithm that we had used previously[7] for $N=0$ can be extended to an arbitrary state N, but the calculation is done at N fixed and the complexity increases with N. On the other hand, in the method of Damburg and Propin,[5] N is just a parameter, and this method would have been more suited to our purpose. Since we have only expanded up to order g^2, we did not need it.

To find the expansion of the function $A(E,g)$ of Eq. (7) up to order g^2, we have assumed that $A(E,g)$ had the same parity property as the perturbative expansion

$$A(-E,-g) = -A(E,g). \tag{49}$$

We had previously calculated the one-instanton contribution to the ground-state energy. We have in addition expanded the W.K.B. expression for the one-instanton contribution in powers of g. This expansion has given us the term of highest degree in E at each order in g. The two pieces of information determine completely $Q_1(E)$ and $Q_2(E)$. For the next polynomial Q_3, we would need another constraint. We could have calculated the one-instanton contribution to the first excited state or used the general method of Ref. 5. We have not done it here.

Let us now give the W.K.B expression for $A(E,g)$:

$$\frac{a}{g} + A(E,g) \sim \frac{1}{g}\int\sqrt{2V(x)-2gE}\,dx. \tag{50}$$

The W.K.B. regime is a regime in which E becomes large at Eg fixed. To expand expression (50) in powers of g, one can use a Mellin transform.

The coefficient of g^K can then be obtained as the residue of the expression $I(\alpha)$ at $\alpha = K+1$:

$$I(\alpha) = \frac{1}{2}\sqrt{\pi}\frac{\Gamma(-\alpha)}{\Gamma(\frac{3}{2}-\alpha)}2^{-\alpha}\int_0^{x_0}[2V(x)]^{1/2-\alpha}\,dx. \tag{51}$$

Expression (51) has actually double poles at integer values of α. The coefficients of these double poles correspond to terms of highest degree of the perturbative expansion (6).

For the two cases we shall consider explicitly, the results are at order g^2:

For the double-well potential of Eq. (34),

$$A = g(\tfrac{17}{2}E^2 + \tfrac{19}{24}) + g^2(\tfrac{227}{2}E^2 + \tfrac{187}{8})E + O(g^3). \tag{52}$$

For the cosine potential of Eq. (35),

$$A = g(\tfrac{3}{2}E^2 + \tfrac{3}{8}) + g^2(\tfrac{11}{2}E^2 + \tfrac{23}{8})E + O(g^3). \tag{53}$$

These two expansions will allow us to test some consequences of our conjecture, as will be explained in the next section.

III. NUMERICAL CALCULATIONS

A. Large-order behavior and one-instanton contribution

A first simple consequence of our conjecture is a relation between the imaginary part of the two-instanton contribution $E^{(2)}(g)$ and the one-instanton contribution $E^{(1)}(g)$. The perturbative imaginary part of $E^{(2)}(g)$ cancels the imaginary part of the Borel sum of the perturbative expansion $E^{(0)}(g)$. This leads to a large-order estimate of the perturbative expansion[11]:

$$E^{(0)}(g) = E_k^{(0)}g^k, \tag{54}$$

$$E_k^{(0)} = -\frac{1}{\pi}\int_0 \frac{\text{Im}\,E^{(2)}(g)}{g^{k+1}}\,dg. \tag{55}$$

Since we can calculate many terms of the perturbative expansion, we can determine numerically the coefficients of a $1/k$ expansion of $E_k^{(0)}$ for k large,[7] and therefore the coefficients of the expansion of $\text{Im}\,E^{(2)}(g)$ for g small.

Let us rewrite Eq. (7) as

$$\sin\pi s/\pi = -\lambda\mu^s\Gamma(1+s)e^{-A}. \tag{56}$$

The one-instanton contribution to s is

$$s^{(1)} = -\lambda e^{-A(E,g)}, \tag{57}$$

in which E can just be replaced by the perturbative expansion obtained by inverting Eq.(25).

The imaginary part of the two-instanton contribution is

$$\operatorname{Im} s^{(2)} = \pi\lambda^{2} e^{-2A(E,g)}, \tag{58}$$

$$\operatorname{Im} s^{(2)} = \pi [s^{(1)}]^{2}. \tag{59}$$

Let us call $E^{(0)}(N,g)$ the perturbative expansion for the state N. Then Eqs. (58) and (59) yield

$$E^{(1)}(g) = -\frac{\partial E^{(0)}}{\partial N} \lambda e^{-A}. \tag{60}$$

This equation is very similar to Eq. (28) of Ref. 5.

$$\operatorname{Im} E^{(2)}(g) = \pi\lambda^{2} e^{-2A} \frac{\partial E^{(0)}}{\partial N}. \tag{61}$$

This leads to the general relation

$$\operatorname{Im} E^{(2)}(g) = \pi [E^{(1)}(g)]^{2} \left[\frac{\partial E^{(0)}}{\partial N}\right]^{-1}. \tag{62}$$

This relation was known at leading order[11] since

$$\frac{\partial E^{(0)}}{\partial N} = 1 + O(g). \tag{63}$$

We believe that in the case of the double-well potential, it follows from Ref. 6, where the corresponding expressions are given at order g for arbitrary N. We shall present some numerical verifications for the double-well and cosine potentials.

1. The double-well potential

More than 90 terms are known for the expansion of $E^{(1)}$ in the case $N = 0$. Using the correspondence between the double-well potential and the $O(2)$ anharmonic oscillator, it is possible to find 50 terms for the expansion of $\operatorname{Im} E^{(2)}$ for $N = 0$ in the literature.[12]

Finally, we have calculated here 46 terms of $E^{(0)}$ as an explicit function of N. It is easy to verify on a few orders that relation (62) is satisfied exactly. Furthermore, the series agree up to order 46 with a relative accuracy of 10^{-11}, which is consistent with the numerical accuracy of the calculation.

Using expansion (52), we can also verify relation (62) up to order g^2 for any value of N. To do this, we have first to analyze the large-order behavior of $E^{(0)}(N,g)$ as a function of N. Since we know the perturbative expansion as an explicit function of N, it is natural to calculate also with noninteger values of N. But an obvious problem arises. Let us write $\operatorname{Im} E^{(2)}$ at leading order

$$\operatorname{Im} E^{(2)} \sim \frac{1}{2} \frac{1}{(N!)^2} \left(\frac{2C}{g}\right)^{2n+1} e^{-2a/g}. \tag{64}$$

We have shown that the perturbative expansion has a parity property which on the $E^{(0)}(N,g)$ reads

$$E^{(0)}(N,g) = -E^{(0)}(-N-1,-g). \tag{65}$$

This property is not shared by expression (64). We shall therefore assume that it should be antisymmetrized:

$$\operatorname{Im} E^{(2)} = \frac{1}{2} \left[\frac{\partial E^{(0)}}{\partial N}(N) \frac{1}{(N!)^2} \left(\frac{2C}{g}\right)^{2N+1} e^{-(2a/g) - A(E,g)} - \frac{\partial E^{(0)}}{\partial N}(-N-1) \frac{1}{\Gamma^2(-N)} \times \left(\frac{g}{2C}\right)^{2N+1} e^{(2a/g) + A(E,g)}\right]. \tag{66}$$

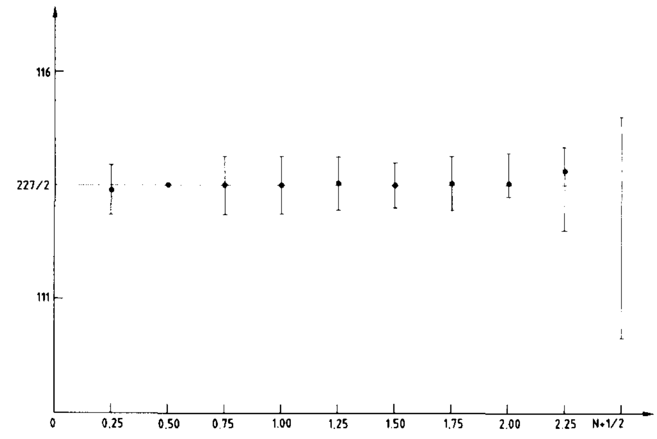


FIG. 1. The value of the coefficient of $E^3 g^2$ in $A(g,E)$ extracted from the large-order behavior of the perturbative expansion of $E^{(0)}(N,g)$ for various values of N, compared with the prediction. The numerical error for $N = 0$ is very small on this scale because the series is much longer. For $N = -\frac{1}{2}$ the coefficient of g^2 in $\operatorname{Im} E^{(2)}$ is predicted to be $\frac{1007}{288} = 3.4965\ldots$, while the numerical result is 3.50 ± 0.01.

We observe that the additional term has a nice property; it vanishes for integer values of N. It gives an oscillating contribution to the large-order behavior subleading by a power k^{-4N-2}. Since we use extrapolating methods based on the existence of an expansion in powers of $1/k$, we have restricted ourselves to values of N for which $4N$ is an integer, and extrapolated independently odd and even orders. The fact that the coefficients of the expansion in powers of $1/k$ have a unique limit justifies *a posteriori* our ansatz.

From relation (55), we see that if we know the expansion of $\operatorname{Im} E^{(2)}$ of up to order g^l, we know also the asymptotic expansion in powers of $1/k$ up to order $1/k^l$. At order $1/k$, the agreement between the predictions and the numerical results is excellent. At order $1/k^2$ or equivalently at order g^2, the numerical errors are larger, but the results are quite convincing, as Fig. 1 shows. We have given the value of the coefficient of $g^2 E^3$ in $A(g,E)$ extracted from the large-order behavior analysis, deriving all other coefficients from Eqs. (52) and (61). The check of our conjecture is that the numbers we obtain are within the numerical errors independent of N, and compatible with the prediction.

2. The cosine potential

Due to the fact that the cosine potential has an infinite number of minima, Expression (7) is modified and replaced[4]

TABLE I. The first line corresponds to the value of the coefficients of g^K for $\operatorname{Im} E^{(2)}$ calculated with expression (62). The second line is obtained from the extrapolation of the large-order behavior, the previous coefficients up to order 6 being fixed at their theoretical value.

K	7	8
$\operatorname{Im} E_2^{\mathrm{I}}$	$35\,323\,847 \times 2^{-11} = 17\,247.9721\ldots$	$162\,003.89\ldots$
$\operatorname{Im} E_2^{\mathrm{II}}$	$17\,247.974$	$162\,000.9$

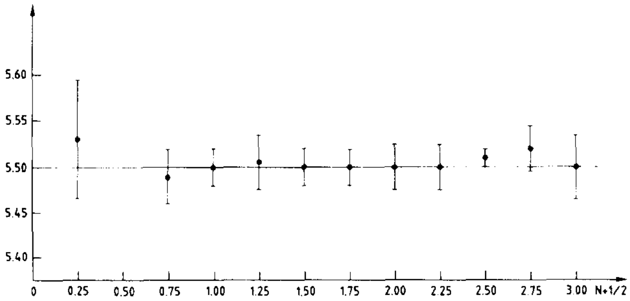


FIG. 2. The value of the coefficient of E^3g^2 in $A(E,g)$ extracted from the large-order behavior of the perturbative expansion for various values of N. For $N=-\frac{1}{2}$ the coefficient of g^2 in $\operatorname{Im} E^{(2)}$ is predicted to be 0.468 75 and found numerically as 0.468 ± 0.003.

by

$$\Gamma(-s)\mu^s\lambda e^{-A}\left[2\cos\varphi + \frac{2i\pi e^{-i\pi s}}{\Gamma(1+s)}\mu^s\lambda e^{-A}\right] = 1, \tag{67}$$

with

$$\lambda = (1/\sqrt{\pi g})e^{-(1/2)g}, \tag{68}$$

$$\mu = 2/g. \tag{69}$$

Since we know many terms of the perturbative expansion and of the one-instanton contribution (more than 90) for the ground state, we can verify the equivalent of relation (62) for $N=0$. Table I shows the comparison between the values of the coefficients at order 7 and 8, as obtained from the large-order behavior analysis (when the six first are fixed at their predicted values) and from relation (62). The terms up to order 6 are

$$\operatorname{Im} E^{(2)}(g) = \frac{2}{g}e^{-1/g}\Big[1 - \frac{5}{2}g - \frac{13}{8}g^2 - \frac{119}{16}g^3 - \frac{5225}{2^7}g^4 - \frac{68715}{2^8}g^5 - \frac{2079317}{2^{10}}g^6 + O(g^7)\Big]. \tag{70}$$

We have then performed the large-order behavior analysis for various values of N to order $1/k^2$. Figure 2 shows the coefficient of E^3g^2 in $A(E,g)$ as a function of N, when the coefficient Eg^2 has been eliminated by imposing the value at $N=0$. Again, the agreement between the prediction and the numerical result is very good.

As a consequence, we can safely assume that relation (62) is correct, independently of our general conjecture. Of course, one should verify it on other examples (or prove it) to make sure that the potentials considered here have not some exceptional property. We shall explore now some further consequences of our conjecture.

B. Some relations between instanton contribution

We shall give or derive some further consequences of Eqs. (6) and (7).

Let us set

$$A(N) = A(E^{(0)}(N)), \tag{71}$$

$$B(N) = -\ln\frac{\partial E^{(0)}}{\partial N}, \tag{72}$$

$$C_n(N) = e^{-B(N) - nA(N)}. \tag{73}$$

If we omit $A(E,g)$ and expand in powers of λ by solving Eq. (7), we obtain

$$s = N - \sum_{n=1}^{\infty}(-1)^n P_n(\ln\mu)\left[\frac{\lambda(-\mu)^N}{N!}\right]^n. \tag{74}$$

The four first polynomials $P_n(\mu)$ are for $N=0$

$$P_1(\ln\mu) = 1, \tag{75a}$$

$$P_2(\ln\mu) = \ln\mu + \gamma, \tag{75b}$$

$$P_3(\ln\mu) = \tfrac{3}{2}(\ln\mu+\gamma)^2 + \pi^2/12, \tag{75c}$$

$$P_4(\ln\mu) = \tfrac{8}{3}(\ln\mu+\gamma)^3 + (\pi^2/3)(\ln\mu+\gamma) + \tfrac{1}{3} + \zeta(3), \tag{75d}$$

where γ is Euler's constant.

With these notations and after some tedious calculations, we obtain the following expressions:

$$E^{(1)}(g) = -\lambda\frac{(-\mu)^N}{N!}C_1(N), \tag{76a}$$

$$E^{(2)}(g) = \lambda^2\frac{\mu^{2N}}{(N!)^2}\left[C_2(N)P_2 + \frac{1}{2}\frac{\partial C_2}{\partial N}\right], \tag{76b}$$

$$E^{(3)}(g) = -\lambda^3\frac{(-\mu)^{3N}}{(N!)^3}\left[C_3P_3 + \frac{\partial C_3}{\partial N}P_2 + \frac{1}{6}\frac{\partial^2 C_3}{(\partial N)^2}\right], \tag{76c}$$

$$E^{(4)}(g) = \lambda^4\frac{\mu^{4N}}{(N!)^4}\left[C_4P_4 + \left(P_3 + \frac{P_2^2}{2}\right)\frac{\partial C_4}{\partial N} + \frac{1}{2}P_2\frac{\partial^2 C_4}{(\partial N)^2} + \frac{1}{4!}\frac{\partial^3 C_4}{(\partial N)^3}\right]. \tag{76d}$$

Let us consider now again Eq. (62). This equation relates, in particular, the large-order behavior of perturbation theory of the one-instanton contribution and of $\operatorname{Im} E^{(2)}$. It is actually more convenient to take the logarithm of this equation. It will be useful from now on to distinguish between the perturbative imaginary part of a series, which we shall call Im, and the nonperturbative imaginary part of the Borel sum, which is connected with the large-order behavior and which we shall denote Ib. Taking then the imaginary part of the logarithm of the equation, we obtain

$$\operatorname{Ib}[\ln \operatorname{Im} E^{(2)}] = 2\operatorname{Ib}[\ln E^{(1)}] - \operatorname{Ib}\left[\ln\frac{\partial E^{(0)}}{\partial N}\right]. \tag{77}$$

The nonperturbative imaginary part of $E^{(1)}$ is cancelled by the perturbative imaginary part of $E^{(3)}$:

$$\operatorname{Ib}[\ln E^{(1)}] = -\lambda^2\frac{\mu^{2N}}{(N!)^2}e^{-2A(N)} \times\left[\operatorname{Im} P_3 - \left(\frac{\partial B}{\partial N} + 3\frac{\partial A}{\partial N}\right)\operatorname{Im} P_2\right], \tag{78}$$

$$\mathrm{Ib}[\ln E^{(1)}] = -\pi\lambda^2 \frac{\mu^{2N}}{(N!)^2} e^{-2A(N)} \times\left\{3[\ln(-\mu) - \psi(N+1)] - \frac{\partial B}{\partial N} - 3\frac{\partial A}{\partial N}\right\}. \tag{79}$$

The nonperturbative imaginary part of $E^{(0)}$ is compensated by $\mathrm{Im}\, E^{(2)}$:

$$\mathrm{Ib}\left[\ln \frac{\partial E^{(0)}}{\partial N}\right] = -\pi\left[2\ln(-\mu) - 2\psi(N+1) - \frac{\partial B}{\partial N} - 2\frac{\partial A}{\partial N}\right] \times \frac{\lambda^2\mu^{2N}}{(N!)^2} e^{-2A(N)}. \tag{80}$$

Adding the two contributions, we find

$$\mathrm{Ib}[\ln \mathrm{Im}\, E^{(2)}] = -\pi\left\{4[\ln(-\mu) - \psi(N+1)] - \frac{\partial B}{\partial N} - 4\frac{\partial A}{\partial N}\right\} \times \frac{\lambda^2\mu^{2N}}{(N!)^2} e^{-2A(N)}. \tag{81}$$

In particular, as was noted previously,[3] one linear combination has no logarithm, and therefore, the corresponding large-order behavior, an asymptotic expansion in powers of $1/k$:

$$\mathrm{Ib}\left[\ln \mathrm{Im}\, E^{(2)} - \frac{4}{3}\ln E^{(1)}\right] = -\frac{\pi}{3}\frac{\partial B}{\partial N} e^{-2A(N)} \frac{\lambda^2\mu^{2N}}{(N!)^2}. \tag{82}$$

C. The double-well potential

The right-hand side of Eq.(82) is known for $N = 0$ because it depends only on the perturbative expansion and the ratio $\mathrm{Im}\, E^{(2)}/E^{(1)}$. The result is

$$\mathrm{Ib}[\ln \mathrm{Im}\, E^{(2)} - \tfrac{4}{3}\ln E^{(1)}] = -(1/g)e^{-1/3g}[2g + \tfrac{34}{3}g^2 + \tfrac{7141}{36}g^3 + \tfrac{1331195}{324}g^4 + \ldots]. \tag{83}$$

The relative difference between the coefficients coming from the numerical extrapolation of the large-order behavior of the left-hand side and the predicted one is smaller than 3×10^{-3}, up to order 5. For the ground state, Eq. (80) can also be verified in the same manner up to order g^2, and the agreement between numerical values and prediction has the same accuracy.

Finally, using an old calculation of the Borel sum of perturbation theory and of the ground-state energy of the double-well potential,[9] it is possible to compare the coefficient of g of $\mathrm{Re}\, E^{(2)}$ to numerical values after subtraction of the terms proportional to $\ln g$. Since the coefficient of $\ln g$ is an asymptotic series which we have not tried to sum, the coefficient of $\ln g$ is only known with a finite accuracy. We have therefore given two results for each value of g to indicate the order of magnitude of the uncertainty. Figure 3 shows our results. Finally, let us note that our expression for $\mathrm{Re}\, E^{(2)}$ agrees at order g with the result of Ref. 6.

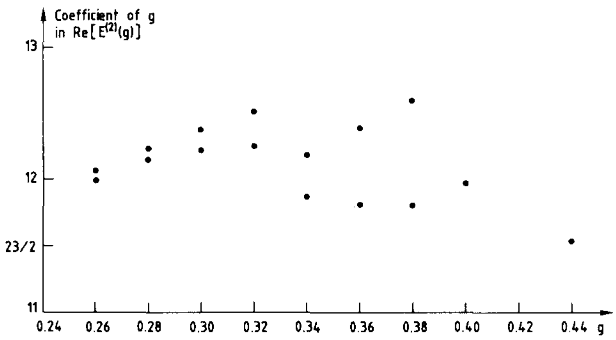


FIG. 3. The value of the coefficient of g in $E^{(2)}(g)$ obtained by subtracting from the half-sum of the ground state and first excited energy, the Borel sum of perturbation theory.

1. The cosine potential

The fact that in this case the spectrum is continuous, so that the energies depend on N and an angle φ, allows a direct evaluation of n-instanton contributions by solving the Schrödinger equation[13] and integrating over the angle φ. Let us write the energy as

$$E(\varphi) = \sum_{-\infty}^{+\infty} e^{in\varphi}E_n, \quad E_{-n} = E_n \tag{84}$$

in which φ characterizes the behavior of the wave function ψ under the translation of one period T of the potential

$$\psi_\varphi[x+T] = e^{i\varphi}\psi_\varphi(x). \tag{85}$$

It is easy to verify that E_n is dominated by an n-instanton contribution. It is possible to evaluate for g small E_n at least up to $n = 4$. Figure 4 shows the relative difference between the numerical value and the estimates obtained from Eqs. (67), (75), and (76), divided by g^3. The coefficient of the leading logarithm has been expanded only up to order 2. Thus the functions presented in Fig. 4 should decrease for

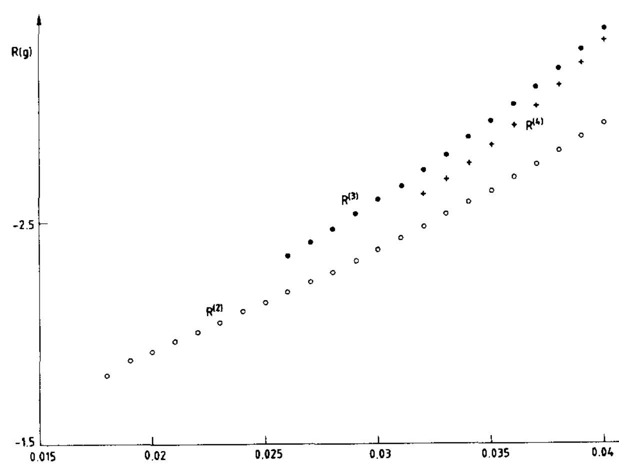


FIG. 4. $R^{(n)}$ is the relative difference between the n-instanton contribution and the theoretical estimates divided by g^3 in the case of the cosine potential.

g small as $1/\ln g$. The results are consistent with the expectation.

IV. CONCLUSION

We have presented very strong numerical evidence that our conjecture is correct for two potentials. It is therefore likely that our conjecture holds for all potentials with degenerate and symmetric minima. One would like now to prove it. Maybe the work of Refs. 5 and 6 will be useful in this respect. It would also be interesting to extend it to more general potentials, and to more than one degree of freedom. It remains to explore the consequences of this conjecture for the summation of the various perturbative and instanton contributions.

[1]J. Zinn-Justin, "Multi Instantons in Quantum Mechanics II," Preprint SPh-T:86, 1982.

[2]This conjecture has been partially inspired by the semiclassical calculation at order g of C. Bender and T. T. Wu, Phys. Rev. D **7**, 1620 (1973).

[3]J. Zinn-Justin, Nucl. Phys. B **192**, 125 (1981).

[4]J. Zinn-Justin, *Les Houches, Lecture Notes 1982* (North-Holland, Amsterdam, 1982) Preprint DPh-T/82-80.

[5]R. Damburg and R. Propin, J. Chem. Phys. **55**, 612 (1971).

[6]R. Damburg and R. Propin, Riga-Salaspils Preprint, 1982 (in Russian).

[7]J. Zinn-Justin, J. Math. Phys. **22**, 511 (1981).

[8]A. D. Dolgov and V. S. Popov, Phys. Lett. B **86**, 185 (1979).

[9]R. Seznec and J. Zinn-Justin, J. Math. Phys. **20**, 1398 (1979).

[10]J. Avron and R. Seiler, Phys. Rev. D **23**, 1316 (1981); S. Graffi and V. Grecchi, "The Borel sum of the double well...," Bologna preprint, 1982. A proof has been given by A. A. Andrianov, Ann. Phys. **140**, 82 (1982).

[11]E. Brézin, G. Parisi, and J. Zinn-Justin, Phys. Rev. D **16**, 408 (1977); E. B. Bogomolny and V. A. Fateev, Phys. Lett. B **71**, 93 (1977).

[12]H. J. Silverstone, E. Harrell, and C. Grot, Phys. Rev. A **24**, 1925 (1981).

[13]*Handbook of Mathematical functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1965), p. 721.

Perturbative solutions in two-channel Schrödinger and Klein–Gordon equations

V. P. Iyer and L. K. Sharma
Department of Applied Physics, Government Engineering College, Jabalpur 482011, India

(Received 3 August 1982; accepted for publication 20 January 1983)

Perturbative method has been applied for solving one- and two-channel Schrödinger and Klein–Gordon equations for general even power potentials, which can be expanded around a minimum. The eigenvalues and solutions have been derived up to and including the second order. Finally, this method has been applied for calculating the eigenvalues of the one-channel equations for the general even power potential and hence for its different particular cases.

PACS numbers: 03.65.Ge, 02.30.Hq

1. INTRODUCTION

Much interest has recently been shown in studying models which give rise to Schrödinger equations containing confining potentials but which also allow scattering to occur.[1] The simplest among such models contains two coupled channels, one of which is permanently closed and the other open for positive energies. Here although scattering occurs in the open channel, its amplitude is strongly influenced by the bound states appearing in the closed channel. Recently Von Gehlen and Rittenberg[2] have considered a two-channel potential scattering problem in three space dimensions for the case when one channel is permanently confined. The examples of confining potentials that they considered were harmonic oscillator and the infinite well. Karlsson and Kerbikov[3] for gaining insight into the widths of quasinuclear levels in the $B\bar{B}$ system investigated a simple multichannel model for the influence of decay channels on a bound state. They found the shift and width of a bound state level to decay strongly not only on the range of annihilation and the $B\bar{B}$ wave function at small distances but also on the position of the level relative to the thresholds. Dashen *et al.*[4] made a detailed investigation on a class of nonrelativistic multichannel potential scattering models. In these models a subset of the channels contained confinement potentials that allowed only a discrete spectrum with an accumulation at $+\infty$; the remaining channels contained the usual scattering states, which are allowed to communicate with the states of the permanently confined channel through an off diagonal local potential. Horn and Novoseller[5] discussed the possibility of existence of a narrow resonance in a multiresonance system above the threshold of an open decay channel.

Recently Müller-Kirsten and Müller[6] developed a general perturbative method for solving the coupled equations of the multichannel formalism. The method used by these authors was a direct generalization of the procedure applied previously[7–10] to a large number of single-channel equations and its extension to the multidimensional case.[11] Müller and Müller-Kirsten[12] later used this technique for iterating one- and two-channel Schrödinger equations for general power potentials. This general iteration procedure was then applied to the linear and logarithmic potentials and also their combinations with a Coulomb potential.

The fact that nonrelativistic models have been quite successful in reproducing the observed mass spectrum of heavy quark–antiquark states has led to a revival of interest[2,3,5,13] in the mulitchannel formalism.[14,15] The present work was thus motivated by the desire to explore the possibility of the existence of hadronic molecular states in the e^+e^- mass spectrum above the radial charmonium state. The search for such states has been activated by the recent observation of a rich closely spaced spectrum for "baryonium" states and the spectrum of $Q^2\bar{Q}^2$ mesons.[16] Once the success of above models have been firmly established the endeavor now is to go beyond the nonrelativistic approach. A complete treatment should actually incorporate both relativistic and quantum effects.

Further, in e^+e^- scattering, the region just above 3.7 GeV (center-of-mass energy) is of particular interest because this is where a $D\bar{D}^*$ molecular state[16] is most likely to show up (at around 3.85 GeV), and, if the charmonium model is reasonably correct, this state could not be mistaken for the next radial excitation, which is predicted to be around 4.0 GeV. Single-channel potential theory normally leads to broad widths, but narrow widths can be generated by the weak coupling to a second channel.[15] It is therefore worthwhile to investigate the two-channel problem defined by the transitions $C\bar{C} \rightarrow D\bar{D}^* \rightarrow D\bar{D}^*$.

In the present investigation we shall, of course, not be concerned with a specific application. In Secs. 2A and 2B we have applied the general perturbative method of Müller-Kirsten and Müller[6] for solving one- and two-channel Schrödinger and Klein–Gordon equations, respectively, for general even power potentials. The channel coupling is assumed to be weak. The general expressions for the eigenvalues and eigenfunctions for this class of potentials have been derived. These general expressions for the Schrödinger equation have then been applied to obtain corresponding expressions for Gauss and anharmonic oscillator potentials in both one and two channels as special cases. This has been done in Sec. 3.

2. DERIVATION OF ASYMPTOTIC EIGENSOLUTIONS

A. Schrödinger equation

We consider the two-channel problem defined by a system of coupled radial Schrödinger equations of the form

$$\begin{pmatrix} \frac{d^2}{dr^2}+E-\frac{l(l+1)}{r^2}-V_{11} & -V_{12} \\ -V_{21} & \frac{d}{dr^2}+E-\frac{l(l+1)}{r^2}-V_{22} \end{pmatrix}\begin{pmatrix}\psi_1(r)\\ \psi_2(r)\end{pmatrix}=0$$

$$(\hbar=c=1,\ m=\tfrac{1}{2}). \tag{2.1}$$

Here E is the total energy of the system, which we assume to be the same in both the channels in order to permit the system of channel 1 to convert into the system of channel 2 and vice versa.

The channel potentials are chosen as

$$V_{ii}(r)=g^2\sum_{p=0}^{\infty}(N_{2p})_{ii}r^{2p}\quad (i=1 \text{ or } 2), \tag{2.2}$$

where the coefficients $(N_{2p})_{ii}$ can be negative also. In particular, we require $(N_2)_{ii}$ to be negative so that the eigenvalues to be derived below are real. Further, for simplicity, we assume that

$$(N_0)_{11}=(N_0)_{22}=N_0$$

and

$$(N_2)_{11}=(N_2)_{22}=N_2. \tag{2.3}$$

The coupling potentials will be assumed to possess the following expansions:

$$V_{ij}(r)=\sum_{p=0}^{\infty}(N_{2p})_{ij}r^{2p}, \tag{2.4}$$

where $i\neq j$.

We now wish to determine the eigenenergies E under normal bound state boundary conditions for large values of the coupling constant g^2.

On substituting (2.2) and (2.4) in (2.1) and changing the independent variable to

$$z=(4g^2N_2)^{1/4}r, \tag{2.5}$$

one obtains

$$\begin{pmatrix} \frac{d^2}{dz^2}+\frac{E-g^2N_0}{(4g^2N_2)^{1/2}}-\frac{l(l+1)}{z^2}-\frac{z^2}{4} & 0 \\ 0 & \frac{d}{dz^2}+\frac{E-g^2N_0}{(4g^2N_2)^{1/2}}-\frac{l(l+1)}{z^2}-\frac{z^2}{4} \end{pmatrix}\begin{pmatrix}\psi_1(z)\\ \psi_2(z)\end{pmatrix}=\begin{pmatrix} -g^2\sum_{p=2}^{\infty}(N_{2p})_{11}\frac{z^{2p}}{(4g^2N_2)^{(p+1)/2}} & \sum_{p=0}^{\infty}\frac{(N_{2p})_{12}z^{2p}}{(4g^2N_2)^{(p+1)/2}} \\ \sum_{p=0}^{\infty}\frac{(N_{2p})_{21}z^{2p}}{(4g^2N_2)^{(p+1)/2}} & -g^2\sum_{p=2}^{\infty}(N_{2p})_{22}\frac{z^{2p}}{(4g^2N_2)^{(p+1)/2}} \end{pmatrix}\begin{pmatrix}\psi_1(z)\\ \psi_2(z)\end{pmatrix}. \tag{2.6}$$

We now assume that, as $g\to\infty$, $(E-g^2N_0)/(4g^2N_2)^{1/2}\sim$ finite and nonzero, i.e., of $O(g^0)$. Then the right-hand side of the above equation is of $O(1/g)$. Hence in the limit $g\to\infty$ (i.e., for small perturbing contributions of the two channels and for weak channel coupling) the right-hand side of (2.6) may be neglected to a first approximation, i.e., the solution

$$\psi(z)=\begin{pmatrix}\psi_1(z)\\ \psi_2(z)\end{pmatrix}$$

for the zeroth order is given by

$$\psi^{(0)}(z)=\begin{pmatrix}\psi^{(0)}(z)\\ \psi^{(0)}(z)\end{pmatrix}. \tag{2.7}$$

Setting $\psi^{(0)}(z)=z^{l+1}e^{-z^2/4}X^{(0)}(z)$ and

$$s=\tfrac{1}{2}z^2, \tag{2.8}$$

one gets (to a first approximation) from Eq. (2.6)

$$\begin{pmatrix} s\frac{d^2}{ds^2}+(b-s)\frac{d}{ds}-a & 0 \\ 0 & s\frac{d^2}{ds^2}+(b-s)\frac{d}{ds}-a \end{pmatrix}\begin{pmatrix}X^{(0)}(s)\\ X^{(0)}(s)\end{pmatrix}=0, \tag{2.9}$$

where

$$a=\frac{l}{2}+\frac{3}{4}-\frac{E-g^2N_0}{(4g^2N_2)^{1/2}},\quad b=l+\frac{3}{2}. \tag{2.10}$$

Then each solution of (2.9) is given by

$$X^{(0)}(s)=\phi^{(0)}(a,b;s), \tag{2.11}$$

where ϕ is a confluent hypergeometric function.

Then the first approximate solution of (2.6)

$$\psi^{(0)}(z)=z^{l+1}e^{-z^2/4}\phi(a,b;z^2/2) \tag{2.12}$$

will be a normalizable bound state wave function, if

$$a=-n\quad \text{for } n=0,1,2,\cdots. \tag{2.13}$$

Setting $q=4n+3$ gives

$$(E-g^2N_0)=g\sqrt{N_2}(2l+q). \tag{2.14}$$

Hence in our original problem we may write

$$(E-g^2N_0)=g\sqrt{N_2}(2l+q)+2N_2\Delta, \tag{2.15}$$

where Δ is an as yet undetermined expansion in descending powers of g.

Next we substitute (2.15) in (2.6) and multiply the equation by (-2). The resulting equation can then be written

$$\mathscr{D}_{qq}\psi=\mathscr{U}\psi, \tag{2.16}$$

where

$$\mathscr{D}_{qq}=\begin{pmatrix}\mathscr{D}_{qq} & 0\\ 0 & \mathscr{D}_q\end{pmatrix}, \tag{2.17}$$

$$\mathscr{D}_q=-2\left(\frac{d^2}{dz^2}+l+\frac{q}{2}-\frac{l(l+1)}{z^2}-\frac{z^2}{4}\right), \tag{2.18}$$

and

$$\mathscr{U}=\begin{pmatrix} 2\Delta h+\sum_{p=2}^{\infty}\frac{(N_{2p})_{11}h^{p-1}}{(N_2)^p}\left(\frac{z^2}{2}\right)^p & -\sum_{p=0}^{\infty}\frac{(N_{2p})_{12}h^{p+1}}{N_2^{p+1}}\left(\frac{z^2}{2}\right)^p \\ -\sum_{p=0}^{\infty}\frac{(N_{2p})_{21}h^{p+1}}{N_2^{p+1}}\left(\frac{z^2}{2}\right)^p & 2\Delta h+\sum_{p=2}^{\infty}\frac{(N_{2p})_{22}h^{p-1}}{N_2^p}\left(\frac{z^2}{2}\right)^p \end{pmatrix} \tag{2.19}$$

with

$$h=\sqrt{N_2}/g. \tag{2.20}$$

Thus the zeroth-order solution (2.16) is given by

$$\mathscr{D}_{qq}\psi_{qq}=0, \tag{2.21}$$

where

$$\psi_{qq}=\begin{pmatrix}\psi_q\\ \psi_q\end{pmatrix}=\begin{pmatrix}\psi^{(0)}\\ \psi^{(0)}\end{pmatrix}.$$

Since

$$\mathscr{D}_{q+i,q+i}\psi_{q+i,q+i}=0,$$
$$\mathscr{D}_{q+i,q+i}=\mathscr{D}_{qq}-\mathscr{I}_{ii}, \tag{2.22}$$

where $\mathscr{I}_{ii}$ is the unit matrix multiplied by i, we have

$$\mathscr{D}_{qq}\psi_{q+i,q+i}=\mathscr{I}_{ii}\psi_{q+i,q+i}. \tag{2.23}$$

This equation will now be used in the development of our iteration procedure.

Considering the right-hand side of (2.16), we first reexpress $\mathscr{U}\psi_{qq}$ as a sum over various $\psi_{q+i,q+i}$, i.e., we write

$$\mathscr{U}\psi_{qq}=\sum_i C(q,q+i)\psi_{q+i,q+i}, \tag{2.24}$$

where each coefficient C is a matrix. We next come to the perturbation procedure. Thus the zeroth-order solution $\psi^{(0)}=\psi_{qq}(z)$ leaves uncompensated on the right-hand side of (2.16) the contribution

$$R_{qq}^{(0)}=\mathscr{U}\psi_{qq}=\sum_i C(q,q+i)\psi_{q+i,q+i}. \tag{2.25}$$

Using (2.23), we see that a term $C(q,q+i)\psi_{q+i,q+i}$ of the sum can be taken care of by adding to $\psi^{(0)}$ the contribution $\mathscr{I}_{ii}^{-1}C(q,q+i)\psi_{q+i,q+i}$ except of course when $i=0$. This means that the first-order contribution of ψ is

$$\psi^{(1)}=\sum_{i\neq 0}C(q,q+i)\mathscr{I}_{ii}^{-1}\psi_{q+i,q+i}. \tag{2.26}$$

Following Müller-Kirsten *et al.*, we make an important observation that the contribution $\psi^{(1)}$ is obtained only by virtue of the fact that $\mathscr{D}_{qq}$ and $\mathscr{I}_{ii}$ are multiples of the unit matrix which commutes with $C(q,q+i)$.

We observe that the first-order contribution leaves uncompensated in (2.25) the term in ψ_{qq}. This will be used to determine Δ and hence E. Since $\psi^{(0)}=\psi_{qq}$ leaves uncompensated $R_{qq}^{(0)}$, the contribution $\psi^{(1)}$ leaves uncompensated

$$\begin{aligned}R_{qq}^{(1)}&=\sum_{i\neq 0}C(q,q+i)\mathscr{I}_{ii}^{-1}R_{q+i,q+i}\\ &=\sum_{i\neq 0}C(q,q+i)\mathscr{I}_{ii}^{-1}\\ &\quad\times\sum_j C(q+i,q+i+j)\psi_{q+i+j,q+i+j}.\end{aligned} \tag{2.27}$$

It follows that the next order contribution $\psi^{(2)}$ becomes

$$\begin{aligned}\psi^{(2)}=&\sum_{i\neq 0}C(q,q+1)\mathscr{I}_{ii}^{-1}\\ &\times\sum_{j+i\neq 0}C(q+i,q+i+j)\mathscr{I}_{i+j,i+j}^{-1}\psi_{q+i+j,q+i+j}.\end{aligned} \tag{2.28}$$

Thus finally we have the iterated sum

$$\psi=\psi^{(0)}+\psi^{(1)}+\psi^{(2)}+\cdots. \tag{2.29}$$

This would be a solution of our coupled equation provided the sum of the coefficients of the terms containing ψ_{qq} in $R_{qq}^{(0)}, R_{qq}^{(1)},\cdots$ left uncompensated so far is set equal to zero, i.e.,

$$0=\det\Big(C(q,q)+\sum_{i\neq 0}C(q,q+i)\mathscr{I}_{ii}^{-1}C(q+i,q)+\cdots\Big). \tag{2.30}$$

To calculate Δ and hence E, we now return to (2.24). For convenience we set

$$\psi_q(z)=\psi(a,b;z)=\psi(a) \tag{2.31}$$

and write the recurrence relation for $\psi(a)$ in the form

$$\begin{aligned}\tfrac{1}{2}z^2\psi(a)=&(a,a+1)\psi(a+1)+(a,a)\psi(a)\\ &+(a,a-1)\psi(a-1),\end{aligned} \tag{2.32}$$

where

$$\begin{aligned}(a,a+1)&=a=-\tfrac{1}{4}(q-3),\\ (a,a)&=b-2a=l+q/2,\\ (a,a-1)&=a-b=-\tfrac{1}{4}(q+3)-l.\end{aligned} \tag{2.33}$$

By repeated application of (2.32) we obtain the following general relation:

$$(\tfrac{1}{2}z^2)^m\psi(a)=\sum_{j=-m}^{m}S_m(a,a+j)\psi(a+j), \tag{2.34}$$

where the coefficients $S_m(a,a+r)$ satisfy the following recurrence relation:

$$\begin{aligned}S_m(a,a+r)=&S_{m-1}(a,a+r-1)(a+r-1,a+r)\\ &+S_{m-1}(a,a+r)(a+r,a+r)\\ &+S_{m-1}(a,a+r+1)(a+r+1,a+r)\end{aligned} \tag{2.35}$$

with $S_0(a,a)=1$; all $S_0(a,a+i)=0$ for $i\neq 0$ and $S_m(a,a+r)=0$, for $|r|>m$.

Using the relations (2.19), (2.24), and (2.34), we obtain (for $i,j=1,2$)

$$\begin{aligned}\sum_k C_{ij}(q,q+k)\psi(a+k)=&2\Delta h\psi(a)\\ &+\sum_{p=2}^{\infty}\frac{(N_{2p})_{ii}h^{p-1}}{N_2^p}\\ &\times\sum_{j=-p}^{p}S_p(a,a+j)\psi(a+j),\end{aligned} \tag{2.36}$$

and for $i\neq j$

$$\sum_k C_{ij}(q,q+k)\psi(a+k) = -\sum_{p=0}^{\infty}\frac{(N_{2p})_{ij}h^{p+1}}{N_2^{p+1}}\sum_{j=-p}^{p}S_p(a,a+j)\psi(a+j). \tag{2.37}$$

These general expressions (2.36) and (2.37) can now be used for writing different coefficients C's directly.

Inserting these coefficients into (2.30) and evaluating the determinant, we obtain

$$\begin{aligned}\Delta^2 = &-\frac{(N_0)_{12}(N_0)_{21}}{2^2N_2^2} + \left\{\frac{(N_4)_{11}(N_4)_{22}}{2^8N_2^4}\right\}\{9q^4+72q^3l+6q^2(32l^2-4l+3)+24ql(8l^2-4l+3)+(8l^2-4l+3)^2\}\\ &+\Delta\left[\left\{\frac{(N_4)_{11}+(N_4)_{22}}{2^4N_2^2}\right\}\{3q^2+12ql+8l^2-4l+3\}+h\left\{\left(\frac{(N_6)_{11}+(N_6)_{22}}{2^5N_2^3}\right)(5q^3+30q^2l+q(48l^2-12l+25)\right.\right.\\ &\left.\left.+16l^3-24l)-\left(\frac{(N_4)_{11}^2+(N_4)_{22}^2}{2^7N_2^4}\right)(17q^3+67q+2(51q^2-18q+67)l+24(7q-3)l^2+64l^3)\right\}+O(h^2)\right]\\ &+h\left[\left\{\frac{(N_4)_{11}(N_6)_{22}+(N_4)_{22}(N_6)_{11}}{2^9N_2^5}\right\}\{15q^5+150q^4l+q^3(544l^2-56l+90)\right.\\ &+q^2(864l^3-336l^2+540l)+q(576l^4-576l^3+992l^2-136l+75)+(128l^5-256l^4+544l^3-272l^2+150l)\}\\ &-\left\{\frac{(N_4)_{11}(N_4)_{22}^2+(N_4)_{22}(N_4)_{11}^2}{2^9N_2^6}\right\}\{51q^5+510q^4l+q^3(1864l^2-176l+252)+q^2(3024l^3-1056l^2+1512l)\\ &+q(2112l^4-1824l^3+2792l^2-376l+201)+512l^5-832l^4+1552l^3-752l^2+402l\}\\ &\left.-\left\{\frac{(N_0)_{12}(N_2)_{21}+(N_0)_{21}(N_2)_{12}}{2^3N_2^3}\right\}\{2l+q\}\right]+O(h^2).\end{aligned} \tag{2.38}$$

Solving for Δ by iteration, we finally obtain

$$E = g^2N_0+\sqrt{N_2}g(2l+g)+2N_2[(B+hD)^2+A+hF]^{1/2}, \tag{2.39}$$

where

$$\begin{aligned}A &= \left\{\frac{(N_4)_{11}(N_4)_{22}}{2^8N_2^4}\right\}\{9q^4+72q^3l+6q^2(32l^2-4l+3)+24ql(8l^2-4l+3)+(8l^2-4l+3)^2\}\\ &\quad-\frac{(N_0)_{12}(N_0)_{21}}{2^2N_2^2},\\ B &= \left\{\frac{(N_4)_{11}+(N_4)_{22}}{2^4N_2^2}\right\}\{3q^2+12ql+8l^2-4l+3\},\\ D &= \left\{\frac{(N_6)_{11}+(N_6)_{22}}{2^5N_2^3}\right\}\{5q^3+30q^2l+q(48l^2-12l+25)+16l^3-24l^2+50l\}\\ &\quad-\left\{\frac{(N_4)_{11}^2+(N_4)_{22}^2}{2^7N_2^4}\right\}\{17q^3+102q^2l+q(168l^2-36l+67)+64l^3-72l^2+134l\},\\ F &= \left\{\frac{(N_4)_{11}(N_6)_{22}+(N_4)_{22}(N_6)_{11}}{2^9N_2^5}\right\}\{15q^5+150q^4l+q^3(544l^2-56l+90)+q^2(864l^3-336l^2+540l)\\ &\quad+q(576l^4-576l^3+992l^2-136l+75)+(128l^5-256l^4+544l^3-272l^2+150l)\}\\ &\quad-\left\{\frac{(N_4)_{11}(N_4)_{22}^2+(N_4)_{22}(N_4)_{11}^2}{2^9N_2^6}\right\}\{51q^5+510q^4l+q^3(1864l^2-176l+252)+q^2(3024l^3-1056l^2+1512l)\\ &\quad+q(2112l^4-1824l^3+2792l^2-376l+201)+(512l^5-832l^4+1552l^3-752l^2+402l)\}\\ &\quad-\left\{\frac{(N_0)_{12}(N_2)_{21}+(N_0)_{21}(N_2)_{12}}{2^3N_2^3}\right\}\{2l+q\}.\end{aligned} \tag{2.40}$$

Equation (2.39) gives an explicit expression for the eigenenergies of our coupled equations.

B. Klein–Gordon equation

We next consider the two-channel problem defined by a system of coupled Klein–Gordon equations which we write as

$$\begin{pmatrix}\dfrac{d^2}{dr^2}+(E-V_{11})^2-m^2-\dfrac{l(l+1)}{r^2} & -V_{12}\\ -V_{21} & \dfrac{d^2}{dr^2}+(E-V_{22})^2-m^2-\dfrac{l(l+1)}{r^2}\end{pmatrix}\begin{pmatrix}\psi_1(r)\\ \psi_2(r)\end{pmatrix}=0 \quad (\hbar=c=1). \tag{2.41}$$

As assumed in Sec. 2A, we consider the total energy E of the system to be same for both the channels.

For the potentials given by Eqs. (2.2), (2.3), and (2.4), Eq. (2.41) takes the form

$$\begin{pmatrix} \dfrac{d^2}{dr^2}+k^2+\sum_{p=0}^{\infty}\{g^4(M_{2p})_{11}-2Eg^2(N_{2p})_{11}\}r^{2p}-\dfrac{l(l+1)}{r^2} & -\sum_{p=0}^{\infty}(N_{2p})_{12}r^{2p} \\ -\sum_{p=0}^{\infty}(N_{2p})_{21}r^{2p} & \dfrac{d^2}{dr^2}+k^2+\sum_{p=0}^{\infty}\{g^4(M_{2p})_{22}-2Eg^2(N_{2p})_{22}\}r^{2p}-\dfrac{l(l+1)}{r^2}\end{pmatrix}\begin{pmatrix}\psi_1(r)\\ \psi_2(r)\end{pmatrix}=0, \tag{2.42}$$

where

$$k^2=E^2-m^2 \quad \text{and} \quad (M_{2p})_{ii}=\sum_{s=0}^{p}\{N_{2(p-s)}\}_{ii}(N_{2s})_{ii}. \tag{2.43}$$

Now changing the independent variable to

$$z=(2ig\sqrt{g^2M_2-2EN_2})^{1/2}, \tag{2.44}$$

Eq. (2.42) reduces to

$$\begin{pmatrix} \dfrac{d^2}{dz^2}+\dfrac{k^2-g^4(2EN_0/g^2-M_0)}{2(ig)^2(2EN_2/g^2-M_2)^{1/2}}-\dfrac{z^2}{4}-\dfrac{l(l+1)}{z^2} & 0 \\ 0 & \dfrac{d^2}{dz^2}+\dfrac{k^2-g^4(2EN_0/g^2-M_0)}{2(ig)^2(2EN_2/g^2-M_2)^{1/2}}-\dfrac{z^2}{4}-\dfrac{l(l+1)}{z^2}\end{pmatrix}\begin{pmatrix}\psi_1(z)\\ \psi_2(z)\end{pmatrix}$$

$$=\begin{pmatrix} \dfrac{1}{2}\sum_{p=2}^{\infty}\dfrac{\{2E(N_{2p})_{11}/g^2-(M_{2p})_{11}\}(z^2/2)^p}{(ig)^{2p-2}\{2EN_2/g^2-M_2\}^{(p+1)/2}} & \dfrac{1}{2}\sum_{p=0}^{\infty}\dfrac{(N_{2p})_{12}(z^2/2)^p}{(ig)^{2p+2}\{2EN_2/g^2-M_2\}^{(p+1)/2}} \\ \dfrac{1}{2}\sum_{p=0}^{\infty}\dfrac{(N_{2p})_{21}(z^2/2)^p}{(ig)^{2p+2}\{2EN_2/g^2-M_2\}^{(p+1)/2)}} & \dfrac{1}{2}\sum_{p=2}^{\infty}\dfrac{\{2E(N_{2p})_{22}/g^2-(M_{2p})_{22}\}(z^2/2)^p}{(ig)^{2p-2}\{2EN_2/g^2-M_2\}^{(p+1)/2}}\end{pmatrix}\begin{pmatrix}\psi_1(z)\\ \psi_2(z)\end{pmatrix}. \tag{2.45}$$

We now assume that $g\to\infty$

$$\{k^2-g^4(2EN_0/g^2-M_0)\}/\{2g^2(2EN_2/g^2-M_2)^{1/2}\}$$

remains finite and nonzero, i. e., of $O(g^0)$.

Hence in the limit $g\to\infty$ (i. e., for small perturbing contributions of the two channels and for weak channel coupling) the right-hand side of (2.45) may be neglected to a first approximation, i. e., the solution

$$\psi(z)=\begin{pmatrix}\psi_1(z)\\ \psi_2(z)\end{pmatrix}$$

is given to zeroth order by

$$\psi^{(0)}(z)=\begin{pmatrix}\psi^{(0)}(z)\\ \psi^{(0)}(z)\end{pmatrix}. \tag{2.46}$$

Setting $\psi^{(0)}(z)=z^{l+1}e^{-z^2/4}X^{(0)}(z)$ and $s=\frac{1}{2}z^2$, one gets from (2.45) (to a first approximation)

$$\begin{pmatrix} s\dfrac{d^2}{ds^2}+(b-s)\dfrac{d}{ds}-a & 0 \\ 0 & s\dfrac{d^2}{ds^2}+(b-s)\dfrac{d}{ds}-a\end{pmatrix}\begin{pmatrix}X^{(0)}(s)\\ X^{(0)}(s)\end{pmatrix}=0, \tag{2.47}$$

where

$$a=\frac{l}{2}+\frac{3}{4}-\frac{k^2-g^4(2EN_0/g^2-M_0)}{4(ig)^2(2EN_2/g^2)-M_2)^{1/2}} \quad \text{and} \quad b=l+\frac{3}{2}. \tag{2.48}$$

The first approximate solution of (2.45)

$$\psi^{(0)}(z)=z^{l+1}e^{-z^2/4}\phi(a,b;z^2/2) \tag{2.49}$$

will be a normalizable bound state wave function, if

$$a=-n \quad \text{for} \quad n=0,1,2,\cdots. \tag{2.50}$$

Setting $q=4n+3$ gives

$$k^2-g^4(2EN_0/g^2-M_0)=(ig)^2(2EN_2/g^2-M_2)^{1/2}(2l+q). \tag{2.51}$$

Hence in our original problem we may write

$$k^2-g^4(2EN_0/g^2-M_0)=(ig)^2(2EN_2/g^2-M_2)^{1/2}(2l+q)+2ig(2EN_2/g^2-M_2)^{1/2}\Delta, \tag{2.52}$$

where Δ is an as yet undetermined expansion in descending powers of g.

Now we substitute (2.52) in (2.45) and multiply the equation by (-2). The resulting equation can be written as

$$\mathscr{D}_{qq}\psi = \mathscr{U}\psi, \tag{2.53}$$

where

$$\mathscr{D}_{qq} = \begin{pmatrix} \mathscr{D}_q & 0 \\ 0 & \mathscr{D}_q \end{pmatrix}, \tag{2.54}$$

$$\mathscr{D}_q = -2\left(\frac{d^2}{dz^2} + l + \frac{q}{2} - \frac{l(l+1)}{z^2} - \frac{z^2}{4}\right), \tag{2.55}$$

and

$$\mathscr{U} = \begin{pmatrix} 2\Delta h - \mathscr{V}_{11} & \mathscr{V}_{12} \\ \mathscr{V}_{21} & 2\Delta h - \mathscr{V}_{22} \end{pmatrix}, \tag{2.56}$$

where

$$\mathscr{V}_{ii} = \sum_{p=2}^{\infty} \frac{\{2E(N_{2p})_{ii}/g^2 - (M_{2p})_{ii}\}h^{2p-2}(z^2/2)^p}{\{2EN_2/g^2 - M_2\}^{(p+1)/2}}, \quad \mathscr{V}_{ij} = -\sum_{p=0}^{\infty} \frac{(N_{2p})_{ij}(z^2/2)^p h^{2p+2}}{\{2EN_2/g^2 - M_2\}^{(p+1)/2}}, \quad h = \frac{1}{ig}. \tag{2.57}$$

Thus the zeroth-order solution of (2.45) is given by

$$\mathscr{D}_{qq}\psi_{qq} = 0, \tag{2.58}$$

where

$$\psi_{qq} = \begin{pmatrix} \psi_q \\ \psi_q \end{pmatrix} = \begin{pmatrix} \psi^{(0)} \\ \psi^{(0)} \end{pmatrix}. \tag{2.59}$$

Proceeding exactly as in Sec. 2A, we finally obtain the following expression for Δ:

$$\begin{aligned} \Delta^2 = \frac{\Delta h}{2}\Bigg[&\left(\frac{(\alpha_2)_{22} + (\alpha_2)_{11}}{2^3\beta^{3/2}}\right)(3q^2 + 12lq + 8l^2 - 4l + 12) + h^2\Bigg\{\left(\frac{(\alpha_3)_{22} + (\alpha_3)_{11}}{2^4\beta^2}\right)(5q^3 + 30q^2l + q(48l^2 - 12l + 25) \\ &+ 16l^3 - 24l^2 + 50l) - \left(\frac{(\alpha_2)_{11}^2 + (\alpha_2)_{22}^2}{\beta^3}\right)\left(q^3 + 6q^2l + q\left(\frac{20l^2 - 4l + 7}{2}\right) - 4l^3 + 4l^2 - 7l\right)\Bigg\} + O(h^4)\Bigg] \\ &+ \frac{h^2}{4}\left[\frac{(N_0)_{12}(N_0)_{21}}{\beta} - \frac{(\alpha_2)_{11}(\alpha_2)_{22}}{2^6\beta^3}(3q^2 + 12ql + 8l^2 - 4l + 12)^2\right] + \frac{h^4}{4}\Bigg[\left\{\frac{(N_2)_{21}(N_0)_{12} + (N_2)_{12}(N_0)_{21}}{2\beta^{3/2}}\right\} \\ &\times\{2l + q\} - \left\{\frac{(\alpha_2)_{11}(\alpha_3)_{22} + (\alpha_3)_{11}(\alpha_2)_{22}}{2^7\beta^{7/2}}\right\}\Big\{15q^2 + 150q^4l + q^3(544l^2 - 56l + 90) + q^2(864l^3 - 336l^2 + 540l) \\ &+ q(576l^4 - 576l^3 + 992l^2 - 136l + 75) + (128l^5 - 256l^4 + 544l^3 - 272l^2 + 150l)\Big\}\Bigg] + O(h^6), \end{aligned} \tag{2.60}$$

where

$$(\alpha_p)_{ii} = [2E(N_{2p})_{ii}/g^2 - (M_{2p})_{ii}], \quad \text{where } p = 1, 2, \cdots, \quad \text{and} \quad \beta = [2EN_2/g^2 - M_2]. \tag{2.61}$$

3. APPLICATIONS OF THE GENERAL EIGENENERGY EXPANSION

We now apply the eigenenergy expansion (2.39) to some particular cases.

A. General even power potential in a single channel

For the case $V_{12} = V_{21} = V_{22} = 0$, i.e.,

$$(N_{2p})_{ij} = (N_{2p})_{22} = 0$$

and $\qquad$ (3.1)

$$(N_{2p})_{11} = N_{2p},$$

Eq. (2.1) reduces to a simple Schrödinger equation with a general even power potential. In this case the general eigenenergy expansion (2.39) reduces to

$$\begin{aligned} E = {} & g^2N_0 + g\sqrt{N_2}(2l + q) + (N_4/2^3N_2) \\ & \times[3(q^2 + 1) + 4(3q - 1)l + 8l^2] \\ & + (N_6/2^4N_2^{3/2}g)[5q(q^2 + 5) \\ & + 2(15q^2 - 6q + 25)l + 24(2q - 1)l^2 + 16l^3] \\ & - (N_4^2/2^6N_2^{5/2}g)[q(17q^2 + 67) \\ & + 2(51q^2 - 18q + 67)l + 24(7q - 3)l^2 + 64l^3] \\ & + O(1/g^2). \end{aligned} \tag{3.2}$$

This expansion is in exact agreement up to $O(1/g)$ with the expression derived previously.[17]

B. Harmonic oscillator

The harmonic oscillator potentials in the two channels will be given by

$$V_{11}(r) = V_{22}(r) = N_2 r^2. \tag{3.3}$$

With

$$(N_0)_{ii} = (N_{2p})_{ii} = 0 \quad \text{for } p \geqslant 2, \tag{3.4}$$

one obtains

$$E = g\sqrt{N_2}(2l+q) = -\lambda\sqrt{N_2}(2l+q), \tag{3.5}$$

which is the well-known result.

C. Gauss potentials

The Gauss potential is given by

$$V(r) = -g^2 e^{-\alpha^2 r^2},$$

so that

$$(N_{2p})_{ii} = (-1)^p \alpha^{2p}/p! \quad \text{and} \quad g^2 \to -g^2.$$

Hence the eigenenergy expansion in two-channel formalism is obtained from (2.39) by making the following substitutions:

$$\begin{aligned} A = {} & \frac{1}{2^{10}}\{9q^4 + 72q^3 l + 6q^2(32l^2 - 4l + 3) \\ & + 24ql(8l^2 - 4l + 3) + (8l^2 - 4l + 3)^2\} \\ & - \frac{(N_0)_{12}(N_0)_{21}}{2^2\alpha^4}, \end{aligned}$$

$$B = \frac{1}{2^4}\{3(q^2+1) + 4(3q-1)l + 8l^2\},$$

$$\begin{aligned} D = {} & \frac{-1}{3\times 2^8}[11q^3 + 6q^2 l + q(120l^2 - 12l + 1) \\ & + 64l^3 - 24l^2 + 2l], \end{aligned}$$

$$\begin{aligned} F = {} & \frac{-1}{3\times 2^{11}}[123q^5 + 1230q^4 l + q^3(4504l^2 \\ & - 416l + 576) + q^2(7344l^3 - 2496l^2 + 3456l) \\ & + q(5184l^4 - 4320l^3 + 6392l^2 - 856l + 453) \\ & + (1280l^5 - 1984l^4 + 3568l^3 - 1712l^2 + 906l)] \\ & + \left[\frac{(N_0)_{12}(N_2)_{21} + (N_0)_{21}(N_2)_{12}}{2^3\alpha^6}\right][2l+q]. \end{aligned}$$

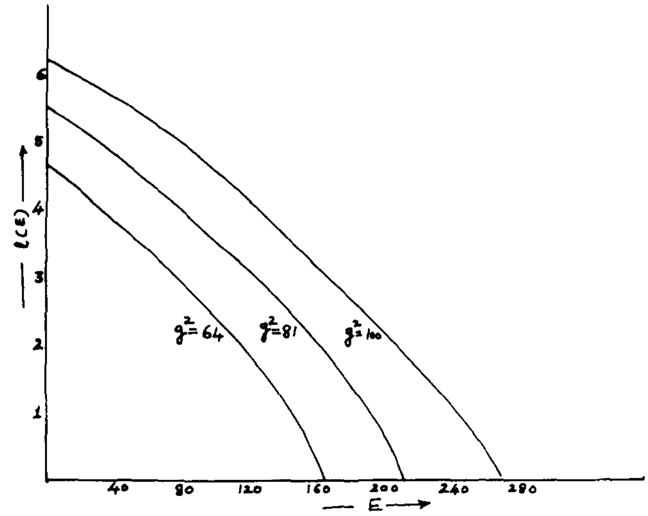


FIG. 1. Ground-state Regge trajectories for the two-channel Schrödinger equations with general even power potentials for different values of coupling constant g^2 with $N_0 = -3$, $N_2 = -1$, $(N_0)_{12} = (N_0)_{21} = 0.1$, $(N_{2p})_{11} = 1$ for $p \geqslant 2$ and $(N_{2p})_{22} = -1$ for $p \geqslant 2$.

D. The anharmonic oscillator

We consider the potentials

$$V_{11}(r) = N_0 + N_2 r^2 + (N_4)_{11} r^4,$$

$$V_{22}(r) = N_0 + N_2 r^2 + (N_4)_{22} r^4,$$

and

$$V_{ij}(r) = \sum_{p=0}^{2} (N_{2p})_{12} r^{2p}.$$

Hence the eigenenergy expansion is again given by (2.39) with the only difference that the term F is now given as

$$\begin{aligned} F = {} & -\left\{\frac{(N_4)_{11}(N_4)_{22}^2 + (N_4)_{11}^2(N_4)_{22}}{2^9 N_2^6}\right\}\{51q^5 + 510q^4 l \\ & + q^3(1864l^2 - 176l + 252) \\ & + q^2(3024l^3 - 1056l^2 + 1512l) \\ & + q(2112l^4 - 1824l^3 + 2792l^2 - 376l + 201) \\ & + (512l^5 - 832l^4 + 1552l^3 - 752l^2 + 402l)\} \\ & - \left\{\frac{(N_0)_{12}(N_2)_{21} + (N_0)_{21}(N_2)_{12}}{2^3 N_2^3}\right\}\{2l+q\}. \end{aligned}$$

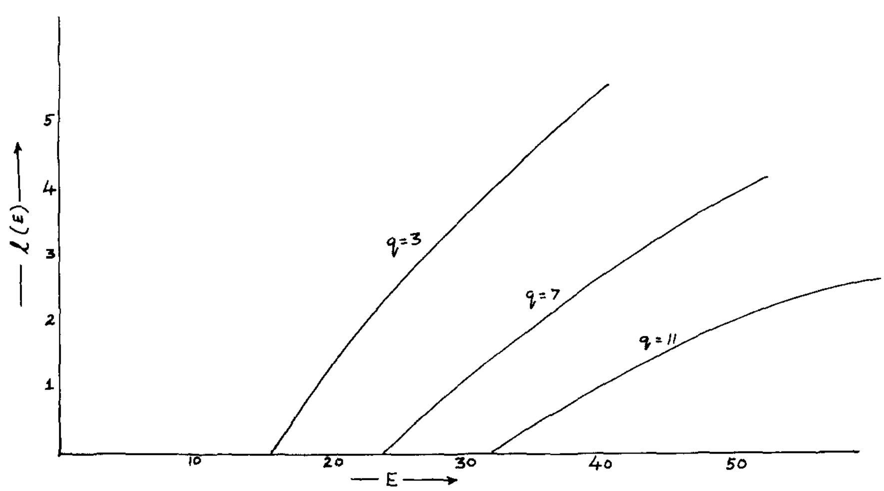


FIG. 2. Regge trajectories for the two-channel Klein–Gordon equations with general even power potentials for different values for the quantum number q. The other constants are $g^2 = -10$, $N_0 = -1$, $N_1 = 1$, $(M_p)_{ii} = 1$ for $p \geqslant 2$ and $i = 1, 2$; $(M_p)_{ij} = (M_p)_{ji} = 0.1$ for $i, j = 1, 2$, $p = 0, 1, 2, \cdots$.

4. CONCLUSIONS

In the preceding sections we have seen, without being concerned with any specific application, how the perturbation method can be used for solving explicitly the eigenvalue problem defined by the coupled equations of the multichannel formalism (both in Schrödinger and Klein–Gordon equations) using a general even power potential in the two channels. Our method is a direct generalization of the method applied previously to the single channel general even power potential problem.[17–19] However, in spite of its simplicity, it will be seen that this generalization is by no means trivial, since the procedure depends crucially on the construction of unperturbed "Hamiltonian," which is a multiple of the unit matrix and so commutes with each of the matrix coefficients of the perturbation. Although numerical methods are quite useful in such type of study, it has been thought worthwhile here to obtain analytical solutions using perturbation technique for answering questions pertaining to global analyticity. The eigenvalues are given by expansions (2.39) and (2.60) for the Schrödinger and Klein–Gordon equations, respectively. The Regge trajectories for the two cases, assuming the angular momentum to be the same in both the channels, are also shown in Figs. 1 and 2. While Fig. 1 is a study of Regge trajectories for different values of the coupling constants in the Schrödinger setup, Fig. 2 depicts rising trajectories in the Klein–Gordon equation for different values of the quantum number q.

In the preceding two-channel problems, for simplicity we have taken the reduced mass in each channel to be $\frac{1}{2}$. One could as well have taken different reduced masses μ_1 and μ_2 for the two channels. It is of interest to point out that for studying the meson spectra the problem of different masses in the two channels could be solved by proceeding along the lines suggested by Müller-Kirsten and Müller.[6] In the above calculations we have assumed the angular momentum also to be the same in both the channels. For different angular momenta we have to, however, treat the difference between channel angular momentum as a further perturbation.

ACKNOWLEDGMENT

One of us (L.K.S.) wishes to thank Professor H. J. W. Müller-Kirsten for many useful discussions while on a visit to his institute in the Federal Republic of Germany.

[1]C. Dullemond and E. Van Berveran, Ann. Phys. (N.Y.) **105**, 318 (1977).
[2]G. Von Gehlen and V. Rittenberg, Phys. Lett. B **71**, 373 (1977).
[3]B. R. Karlsson and B. Kerbikov, Nucl. Phys. B **141**, 241 (1978).
[4]R. F. Dashen, J. B. Healy, and I. J. Muzinich, Ann. Phys. (N.Y.) **102**, 1 (1976).
[5]D. Horn and D. E. Novoseller, Phys. Rev. D **17**, 1763 (1978).
[6]H. J. W. Müller-Kirsten and R. Müller, Phys. Rev. D **20**, 2541 (1979).
[7]R. B. Dingle and H. J. W. Müller, J. Reine Angew. Math. **216**, 123 (1964); H. J. W. Müller, *ibid.* **211**, 33 (1962).
[8]H. J. W. Müller, Ann. Phys. (Leipz.) **15**, 395 (1965).
[9]H. J. W. Müller-Kirsten and N. Vahedi-Faridi, J. Math. Phys. **14**, 1291 (1973).
[10]H. J. W. Müller, J. Math. Phys. **11**, 355 (1970).
[11]H. J. W. Müller-Kirsten, Phys. Lett. A **70**, 383 (1979).
[12]R. Müller and H. J. W. Müller-Kirsten, J. Math. Phys. **22**, 733 (1981).
[13]S. Nussinov and D. P. Sidhu, Fermilab Report No. 76/70-THY, 1976 (unpublished).
[14]The standard reference is R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[15]H. Feshbach, Ann. Phys. (N.Y.) **5**, 357 (1958); K. Gottfried, in *Elementary Particle Physics and Scattering Theory, 1967 Brandeis Summer Institute in Theoretical Physics*, edited by M. Chretien and S. S. Schweber (Gordon and Breach, New York, 1970), Vol. II, p. 148. A more recent treatment has been given by Cao-Xuan-Chuan and H. van Regemorter, ICTP Report No. 78/98, 1978 (unpublished).
[16]A. De Rújula and R. L. Jaffe, MIT Report No. CTP 658, 1977 (unpublished); A. De Rújula, H. Georgi, and S. L. Glashow, Phys. Rev. Lett. **38**, 317 (1977); R. L. Jaffe, Phys. Rev. D **15**, 281 (1977); W. Celmaster, Harvard University Report No. HUTP-77/A022, 1977 (unpublished); H. J. Lipkin, Phys. Lett. B **70**, 113 (1977); S. F. Tuan, Phys. Rev. D **15**, 3478 (1977); Ch. de la Vaissiere, CERN Report No. 77–33, 1977 (unpublished); D. Robson, Liverpool University Report No. LTH 35, 1977 (unpublished).
[17]L. K. Sharma, J. Choubey, and H. J. W. Müller-Kirsten, J. Math. Phys. **21**, 1533 (1980).
[18]L. K. Sharma and V. P. Iyer, J. Math. Phys. **23**, 1185 (1982).
[19]L. K. Sharma, S. Mehta, and V. P. Iyer, J. Math. Phys. (to be published).

Stochastic formulation of Feynman path integrals from the least action point of view

Lech Papiez[a)]
School of Theoretical Physics, Dublin Institute for Advanced Studies, 10 Burlington Road, Dublin 4, Ireland

(Received 17 September 1982; accepted for publication 20 January 1983)

It is shown that solutions of Schrödinger equation may be represented by the action of a stochastically perturbed classical system. In a sense this means that quantum mechanics could be regarded as being a consequence of a variational principle applied to randomly disturbed dynamical systems. A comparison with the Feynman–Kac formula is made and the classical limit is discussed.

PACS numbers: 03.65.Ge, 03.20. + i, 02.50.Ey

1. INTRODUCTION

This paper is devoted to a more careful analysis of the idea which was briefly sketched in Ref. 1. Let us remember that in Ref. 1 we showed that the Schrödinger equation could be regarded as being a consequence of the variational principle applied to a stochastically perturbed system. In this previous paper we argued rather heuristically, using an imaginary diffusion constant. Here we pay more attention to the mathematically proper formulation. Thus we use a real, positive diffusion constant to obtain the heat equation and then we exploit the idea of analytic continuation (with respect to a diffusion constant) to get the Schrödinger equation. The idea of analytic continuation for a heat equation is not new, and it was discussed many times, especially when the Feynman–Kac formula of path integrals approach to the formulation of quantum dynamics[2] was being considered. Therefore, we do not discuss this problem in full detail, assuming it is rather well known. More attention is paid to the problem of stochastic optimal control[3–6] which has been less studied by physicists.

This paper is constructed in the following way. First, to make motivation of this paper more clear, we give a short review of classical mechanics from the deterministic optimal control point of view and we introduce stochastic control methods. Then we apply these methods for stochastically perturbed (classical) systems, we derive a heat equation for such systems, and we discuss its analytic continuation to the Schrödinger equation. Next we try to find the place for our stochastic-variational formulation of quantum mechanics among the other methods, and consider its interpretation. We stress, first of all, that it may be seen as the realization of Feynman and Dirac ideas[7–11] of quantum mechanics' formulation. Comparing it with the Feynman–Kac formula we try to show its advantages. In this context we examine the classical limit which, as is well known, leads to difficult problems in the Feynman–Kac formulation of quantum mechanics. We show that using our method we can overcome these problems, i.e., that this limit is well defined and that it leads to reasonable and interesting implications. The main implication is that Maslov's WKB approximation of quantum mechanics[12] has its basis in the theory of stochastically perturbed dynamical systems.[3,6,13] Therefore, many results of stochastic perturbation theory[3,6,13] may be adjusted to quantum mechanics.

Finally we make some remarks about the interpretation of the proposed stochastic quantum mechanics, and about the possibility of extending our formulation to include relativistic and spin embedding systems.

2. OPTIMAL CONTROL THEORY AND HAMILTON–JACOBI FORMULATION OF CLASSICAL MECHANICS

It is well known that the Hamilton–Jacobi equation is obtained when the variational principle is applied to a classical dynamical system. In the optimal control language, the Hamilton–Jacobi theory may be very briefly reviewed in the following way.

Let us consider R^m both as a configuration space of a classical system and also as a space where values of control parameters belong. We shall denote points in the configuration space and in the control space by x and u, respectively, i.e., $(x,u) \in R^m \times R^m$. Later we shall use also the same notation x and u for appropriate functions (evolution and control, respectively) but it will not lead to misunderstandings. A dynamical equation is given which binds together a state of a system and a control function, namely,

$$\dot{x}(s) = u_s$$

subject to the initial condition

$$x(s_0) = x \in R^m, \tag{2.1}$$

where the dot denotes the derivative with respect to time s. Here the symbol u_s should be understood quite generally, i.e., that the value u_s at time s depends on complex information about a system (for example, u_s may be regarded as a functional of the trajectory of a system between the initial time s_0 and the time s). However, such a general point of view is not necessary. Namely, the deterministic nature of a classical system suggests that information which may be used to choose the control parameter u_s at any moment s is constrained to a state of a system at that moment. Thus, in practice, u_s must be a function of x and s only, i.e., $u_s = u(x,s)$. Hence, (2.1) may be written explicitly (and less generally) as

$$\frac{dx(s)}{ds} = u(x(s),s)$$

[a)]On leave of absence from Silesian University, Katowice, Poland.

subject to the initial condition

$$x(s_0) = x \in R^m. \tag{2.2}$$

Moreover, our system is characterized by a Lagrange function $L(x, u, s)$, where $x, u \in R^m$ and $s \in R$.

Now, we are ready to use the variational principle to decide which one of control functions $u(x, s)$ is to be chosen in order to determine the real evolution of our system. Let us define, firstly, S, the classical action of our system as a function which depends on initial conditions x and t (from now on, we make a substitution $s_0 = t$) as

$$S(x, t) = \inf_{u \in U} \left[\int_t^T L\,[x(s), u(x(s),s),s]ds + g(x(T),T) \right], \tag{2.3}$$

where T is some fixed, final moment of evolution, and U is a class of admitted control functions. The policy u^* for which the lower bound in (2.3) is attained is called the optimal control, and the solution of (2.2) for $u = u^*$ is called the optimal trajectory. The variational principle of classical mechanics is in fact a statement that the optimal trajectory is the real one for a classical system.

Here, some questions arise, however. The main one concerns the existence and regularity of functions $S(x, t)$ and $u^*(x, t)$. Fortunately, the above problem is quite simple from the optimal control point of view (the so called simplest problem of the calculus of variations) and some satisfactory answers may be formulated.[3] Briefly speaking, for sufficiently good functions L and g (and when the set U is not restricted too much, *a priori*) the action S and optimal policy u^* exist and are regular (continuous, differentiable, etc.). In this case, it is found also that the action S satisfies the nonlinear, differential equation[3]

$$\inf_{u \in R^m} \left[\frac{\partial S(x, t)}{\partial t} + L(x, u, t) + u\,\frac{\partial S(x, t)}{\partial x} \right] = 0, \qquad S(x, T) = g(x, T). \tag{2.4}$$

This equation will take a more familiar shape if we express the Lagrange function in the usual form as

$$L(x, u, t) = (m/2)u^2 - V(x), \tag{2.5}$$

where $V(x)$ is the potential and $m > 0$ is the mass. In this case, Eq. (2.4) yields

$$\frac{\partial S(x, t)}{\partial t} - \frac{1}{2m}\left(\frac{\partial S(x, t)}{\partial x}\right)^2 - V(x) = 0, \qquad S(x, T) = g(x, T) \tag{2.6}$$

as the value of the control function [at a point (x, t)] for which (2.4) attains the lower bound $u(x, t) = -(1/m)(\partial S(x, t)/\partial x)$.

This last equation is the well-known backward Hamilton–Jacobi equation. The reason why it is the backward equation is that the action S depends on initial conditions (not final, as is usual in classical mechanics[14,15]). Using defining equation (2.2) and the action S in (2.3) with final conditions we obtain the forward (usual) Hamilton–Jacobi equation; however, this construction is somewhat artificial from the optimal control point of view.

3. OPTIMAL CONTROL OF STOCHASTICALLY PERTURBED CLASSICAL SYSTEMS

Let us now consider a stochastically perturbed classical system. For such a system we will choose one which differs from the system described earlier only in that Eq. (2.2) embeds a (white) noiselike random disturbance.[3,4,6,13] The consequences of this perturbation will be crucial.

First of all, if we wish to apply the variational principle to such a system we have to use not deterministic, as before, but stochastic optimal control methods. These are well known to specialists but are not very accessible to many physicists, so we shall be more expository here.

The configuration space of our system is, as before, R^m. There are, also, no constraints on values of control functions, i.e., we assume that they take values in R^m. For convenience we assume that the potential V is a real-valued, $C^\infty(R^m)$ bounded function (such strong regularity is not necessary to obtain main results but simplifies our argument). We must also recognize here that the admitted stochastic nature of the perturbation will affect the control and the state, so both of them must be regarded, during the evolution, as stochastic processes. Taking into account the above remarks we see that, formally, our system is characterized by a stochastic equation

$$\dot{x}(s) = u_s + c\dot{W}(x)$$

subject to the initial condition

$$x(s_0) = x \in R^m, \tag{3.1}$$

and by a Lagrange function

$$L(x, u, s) = (m/2)u^2 - V(x), \tag{3.2}$$

where $V(x)$ is the potential discussed earlier, c is some real, positive constant which determines the variance of the $cW(s)$ process while $W(s)$ is the standard, m-dimensional Wiener process, and x is a nonrandom point in R^m from which our system starts at the moment s_0. It will be convenient for us to write the constant c in the form $c = \sqrt{\hbar'/m}$ where m is a mass [as in (2.5) or (3.2)] and $\hbar'$ some positive constant.

Equation (3.1) demands some further explanation. First of all, it is much easier to give a precise meaning to (3.1) when it is written in the integral form. Also, we have not so far explained how a control policy u should be understood when a stochastic perturbation is allowed. Intuitively one may argue that the Markovian nature of disturbance [$W(s)$ is a Wiener process] implies, as it does in the deterministic case, that only information about the state of the system at moment s matters for determination of the control parameter u at that moment. This means that as in the deterministic case, $u_s = \tilde{u}(x, s)$. Such policies $\tilde{u}$ are known as Markov strategies. Actually, it may be proved rigorously[3,4] that even when the most general strategies [i.e., when u_s are progressively measurable processes with respect to a system of σ-algebras $\{\mathscr{F}_s\}$ generated by a Wiener process $W(s)$] are admitted, the real evolution of a perturbed system derived with the help of (Lipshitz) Markov policies is not affected (this becomes clearer later). Thus Eq. (3.1) may be rewritten as the stochastic integral equation

$$x(s,\omega) = x + \int_{s_0}^{s} \tilde{u}(x(s',\omega),s')ds' + \sqrt{\hbar'/m}\,W(s,\omega), \quad s \geqslant s_0, \tag{3.3}$$

where, for the moment, to stress the random nature of our variables, we explicitly use the random parameter ω (the set Ω of all ω may be interpreted as the set of all Brownian motion trajectories). We are in a position now to apply the variational principle to our system. As in the deterministic case we start with a definition of the action S_p depending on initial conditions for x and t (similarily, as before, we make now a substitution $s_0 = t$). For our perturbed system, for any fixed x and t, we assign the value

$$S_p(x,t) = \inf_{\tilde{u}\in\tilde{U}} E_x\left[\int_t^T L\,(x(s),\tilde{u}(x(s),s),s)ds + g(x(T),T)\right],$$
$$t \leqslant T, \tag{3.4}$$

where g is some C^∞ function on R^m, $\tilde{U}$ is a set of Markov control policies suitable for a perturbed system, and E_x denotes the relevant probability average which, in a sense, may be interpreted as the average over all trajectories starting from $x \in R^m$ (for any fixed policy $\tilde{u}$). For the moment the behavior at infinity ($|x| \to \infty$) of the function $g(x,T)$ is not important—later, for obvious reasons, we shall put some constraints on it. Let us notice here that the action S_p of a perturbed system so defined is a nonrandom, real-valued function, depending only on the moment t and on the configuration space point x. The variational prinicple should be understood to mean that the real (stochastic) evolution of a perturbed system is determined, thanks to Eq. (3.3), by the control $\tilde{u}^*$ (optimal one) such that the lower bound in (3.4) is attained. What can we say, however, about the regularity of functions S_p and $\tilde{u}^*$? One may show, using the standard technique of step strategies,[4] that minimalization over these strategies in (3.4) gives us a function $S_p(x,t)$ which is no different from the one obtained by using the most general policies (or Markov policies, what was hinted earlier). This implies that, under our previous assumptions, differentiation of $S_p(x,t)$, to any order, is possible with respect to x and t. Details of differentiation for functions defined like $S_p(x,t)$ in (3.4) are exhaustively discussed in Ref. 4. These properties of the action S_p allow us to apply Ito's formula to this function.[16,17] We have for any (not necessarily the real) evolution $x(s)$ the identity

$$S_p(x(s),s) - S_p(x,t) = \int_t^s \left[\frac{\hbar'}{2m}\frac{\partial^2}{\partial x^2} + \tilde{u}(x(s'),s')\frac{\partial}{\partial x} + \frac{\partial}{\partial t}\right] S_p(x(s'),s')ds'. \tag{3.5}$$

Combining the last expression with (3.4), where $T = s$ and $g(x(s),s) = S_p(x(s),s)$, we get easily, for a standard form of a Lagrange function (3.2), the equality for S_p:

$$\inf_{\tilde{u}\in\tilde{u}} E_x\left(\int_t^s \left\{\left[\frac{\partial}{\partial t} + \frac{\hbar'}{2m}\frac{\partial^2}{\partial x^2} + \tilde{u}(x(s'),s')\frac{\partial}{\partial x}\right]S_p(x(s'),s') + \frac{m}{2}\tilde{u}^2(x(s'),s') - V(x(s'))\right\}ds'\right) = 0, \quad t < T, \tag{3.6}$$
$$S_p(x,T) = g(x,T).$$

When we divide throughout by $s - t$, and we let s tend to t, (3.6) yields

$$\inf_{u\in R^m}\left[\frac{\partial}{\partial t}S_p(x,t) + \frac{\hbar'}{2m}\frac{\partial}{\partial x^2}S_p(x,t)u\frac{\partial}{\partial x}S_p(x,t) + \frac{m}{2}u^2 - V(x)\right] = 0, \quad t < T, \tag{3.7}$$
$$S_p(x,T) = g(x,T).$$

This is the so-called Bellman equation. It may be regarded as a generalization of the backward Hamilton–Jacobi equation for a stochastically perturbed classical system. We notice that the lower bound in the bracket is attained if a control parameter, at any point (x,t), is equal to $-(1/m)(\partial S_p(x,t)/\partial x)$. Therefore, Eq. (3.7) may be also written as

$$\frac{\partial}{\partial t}S_p(x,t) + \frac{\hbar'}{2m}\frac{\partial^2}{\partial x^2}S_p(x,t) - \frac{1}{2m}\times\left(\frac{\partial S_p(x,t)}{\partial x}\right)^2 - V(x) = 0, \quad t < T, \tag{3.8}$$
$$S_p(x,T) = g(x,T).$$

Thus, we have found, in addition, that the optimal control $\tilde{u}^*(x,t)$ is a C^∞ function of x and t, and that

$$\tilde{u}^*(x,t) = -(1/m)(\partial S_p(x,t)/\partial x). \tag{3.9}$$

Incidentally, thanks to stochastic optimal technique, we have found that the nonlinear, differential equation (3.8) has a solution which is a C^∞ function of x and t jointly.

4. THE SCHRÖDINGER EQUATION AND STOCHASTICALLY PERTURBED CLASSICAL SYSTEMS

Let us consider a function $\varphi(x,t)$, depending on a configuration space point x and time t, defined as

$$\varphi(x,t) = \exp(-(1/\hbar')S_p(x,t)), \quad t \leqslant T. \tag{4.1}$$

This is a positive-definite nonrandom function of the class C^∞ with respect to x and t. As a consequence of (3.8), we have the following differential identity for this function:

$$\frac{\partial\varphi(x,t)}{\partial t} + \frac{\hbar'}{2m}\frac{\partial^2}{\partial x^2}\varphi(x,t) + \frac{1}{\hbar'}V(x)\varphi(x,t) = 0, \quad t < T, \tag{4.2}$$
$$\varphi(x,T) = \exp(-(1/\hbar')g(x,T)),$$

i.e., the function $\varphi(x,t)$, defined by (4.1), is the solution of the backward Cauchy problem for a heat equation (4.2) with a source $(1/\hbar')V(x)$.

We may interpret this result in the following way. It is well known that the backward Cauchy problem for a heat equation in $L^2(R^m)$ has a unique solution with properties of positivity and C^∞ regularity with respect to x and t. Thus, if we restrict ourselves to the case when $\varphi(x,T) \in L^2(R^m)$ [or more precisely, $\varphi(x,T) \in L^2(R^m) \cap C^\infty(R^m)$], then we have the solution of the backward Cauchy problem for a heat equation (4.2) in $L^2(R^m)$ represented, thanks to (4.1), by the action S_p of the perturbed classical system (3.4).

So far we have discussed only real-valued functions, and therefore real space $L^2(R^m)$. However, it is easy to see that the solution of (4.2) in $L^2(R^m)$-complex [let us denote this space $L_c^2(R^m)$] with the final condition $\varphi(x,T)\in L_c^2(R^m)\cap C_c^\infty(R^m)$ can also be represented by the action (3.4). It is enough to realize that a heat equation is linear, and that for any $\varphi\in L_c^2(R^m)$ we have $\varphi=\varphi_1+i\varphi_2$ where $\varphi_1,\varphi_2\in L^2(R^m)$-real.

Let us consider now, more carefully, Eq. (4.2) as the evolution equation on $L_c^2(R^m)$. We shall be interested especially in the likeness between this equation and Schrödinger equation. We shall use here the idea of the Schrödinger equation definition by "analytic continuation" of the heat equation.[2] In our case we may argue in the following way, using semigroup theory.[18,19] The first question is whether the operator

$$A=\frac{\hbar'}{2m}\frac{\partial^2}{\partial x^2}+\frac{1}{\hbar'}V(x) \tag{4.3}$$

on $L_c^2(R^m)$ with appropriately chosen domain $D(A)$ ($D(A)=W_2^2(R^m)$, a second-order Sobolev space with respect to $L_c^2(R^m)$), is the generator of the backward linear semigroup in $L_c^2(R^m)$. The answer will be obvious if one regards A as the sum of two linear operators, i.e., $A=A_1+A_2$, where $A_1=(\hbar'/2m)(\partial^2/\partial x^2)$, $A_2=(1/\hbar')V(x)$ and $D(A_1)=D(A_2)=D(A)$. The results of the perturbation theory of semigroups[18,19] immediately give us a positive answer [let us remember our assumptions about $V(x)$]. Let us denote this semigroup of operators (linear, continuous, and contractive) generated by A in $L_c^2(R^m)$ by $T_t^{\hbar'}$. In our case, $T_t^{\hbar'}$ is the backward semigroup "starting back" at moment T, so $T_t^{\hbar'}\varphi(x,T)=\varphi(x,t)$, and $T_T^{\hbar'}=1$ is the identity operator. But semigroup perturbation theory assures us also that $T_t^{\hbar'}$, as a function of $\hbar'$, has analytic extension to the right, complex semiplane P. We shall denote points in P by $\hbar''$, i.e., $\hbar''\in P=\{z:|\arg z|<\pi/2\}=\{z:\mathrm{Re}\,z>0\}$. It is enough to notice that $A_1'=(\hbar''/2m)(\partial^2/\partial x^2)$, $D(A_1')=D(A)$ generates a semigroup, that $A_2'=(1/\hbar'')V(x)$ is a bounded operator, and to recall the known results for the perturbation of semigroups,[19] to conclude that $A'=(\hbar''/2m)(\partial^2/\partial x^2)+(1/\hbar'')V(x)$ has $D(A)$ as its domain, and generates a semigroup $T_t^{\hbar''}$ on $L_c^2(R^m)$ $(T_t^{\hbar''}|_{\hbar''=\hbar'}=T_t^{\hbar'})$. Moreover, $T_t^{\hbar''}\varphi(x,T)$ is, in general, a holomorphic $L_c^2(R^m)$-valued function of $\hbar''$ (in P). From the theory of holomorphic functions we know that for every function $T_t^{\hbar''}\varphi(x,T)$ there exists a nontangential limit to almost every point on the imaginary axis $\mathrm{Re}\,z=0$ (excluding a set of Lebesgue measure 0). If we choose on this axis a point $z=i\hbar$, where $\hbar$ is the Planck constant, we shall obtain, thanks to this limit, a function $\psi(x,t)=T_t^{i\hbar}\varphi(x,T)$ which satisfies the equation

$$\frac{\hbar}{i}\frac{\partial}{\partial t}\psi(x,t)=-\frac{\hbar''}{2m}\frac{\partial^2}{\partial x^2}\psi(x,t)+V(x)\psi(x,t),\quad t<T,$$
$$\psi(x,T)=\varphi(x,T). \tag{4.4}$$

This equation is, of course, a consequence of (4.2), and the property that during the extension and continuation to the imaginary axis the differential operations remains valid.

Equation (4.4) is the backward Cauchy problem for the Schrödinger equation in $L_c^2(R^m)$. Thus, in the sense of definition by analytic continuation, quantum dynamics is obtained from a variational principle applied to a stochastically perturbed classical, Lagrangian system.

Let us stress here two problems associated with Eq. (4.4). The first is that we have got the backward Schrödinger equation: This is because the action $S_p(x,t)$ for a perturbed system was defined as a function of the initial conditions x and t. If the formulation is given through the action defined as a function of final conditions then the usual, forward Schrödinger equation will be obtained instead of (4.4). The next problem is whether we have to assume such strong regularity of functions $V(x)$ and $g(x,T)$ [or, as is equivalent, of $\varphi(x,T)$]. It turns out, for example, that all the steps we have taken can be repeated for $V(x)$ and $g(x,T)$ in $C^2(R^m)$—it is clear that the application of Itô's formula for $S_p(x,t)$ in (3.5), crucial to our arguments, is still admissable under these circumstances. It is possible to consider even more general cases, e.g., if $S_p(x,t)$ is nondifferentiable in the usual sense, and to apply the generalized Itô's formula[4] for $S_p(x,t)$ to get our results. Let us point out, however, that the principal, nontechnical, difficulty in treating such cases appears to arise when the action is no longer a common function but an element of some functional space. The point is that in this case the minimalization in the definition (3.4) of $S_p(x,t)$ loses its meaning! This last difficulty appears, in practice, only when we try to consider the final conditions $\varphi(x,T)$ [or $g(x,T)$] as functional space elements, for example, as elements of $L_c^2(R^m)$. Potential $V(x)$ is usually understood to be at least a continuous function in common sense. Therefore, a more attractive way to treat the general conditions for $\varphi(x,T)$ in Eq. (4.4) is to work initially with smooth final conditions in $L_c^2(R^m)\cap C_c^\infty(R^m)$, and then to extend this equation to general final conditions, which are elements of functional space $L_c^2(R^m)$, approaching them, in the sense of $L_c^2(R^m)$ limit, by functions from a dense set $L_c^2\cap C_c^\infty$.

It is worthwhile to notice at last that a standard technique exists which enables us to extend the stochastic control, as well as the semigroup theory considerations, for unbounded potentials $V(x)$.[2,20]

5. COMPARISONS WITH THE FEYNMAN–KAC FORMULA AND CLASSICAL LIMIT

The connections between our method and Feynman's path integral formulation of quantum mechanics are obvious. Especially, the stochastic representation of quantum dynamics, thanks to the Feynman–Kac formula, is in some ideas very similar to our above considerations, but with one important difference: the Feynman–Kac formula does not use the variational principle at all. This will be clearer if we put the Feynman–Kac quantization in our notation. In the Feynman–Kac method diffusion of the form

$$x(s,\omega)=x+\sqrt{(\hbar'/m)}\,W(s,\omega)$$

subject to the initial condition

$$x(t,\omega)=x \tag{5.1}$$

is assumed [here all symbols have the same meaning as in (3.3)]. This enables us to define a linear, holomorphic semigroup $F_t^{\hbar'}$ on $L_c^2(R^m)$, the so-called Feynman–Kac formula;

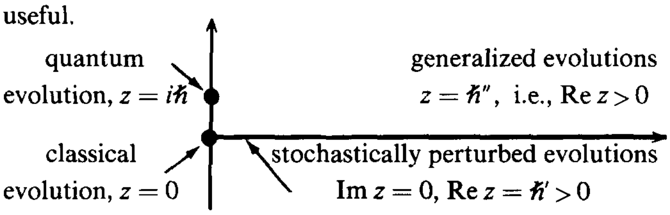
that is,

$$F_t^{\hbar'}\varphi(x,t)=E_x\left\{\exp\left[+\frac{1}{\hbar'}\int_t^T V(x(s))ds\right]\varphi(x(T),T)\right\},\tag{5.2}$$

$$F_T^{\hbar'}\varphi(x,T)=\varphi(x,T),$$

which, thanks to known results connecting diffusions with linear second-order differential equations,[16–18] fulfills an equation of the form (4.2). In particular, this equation may be considered as an evolution equation in $L_c^2(R^m)$. Thus, using the same analytic continuation in $\hbar'$ as before, the quantum dynamics is determined by the Feynman–Kac formula. As we see from (5.2) this formulation realizes, to some extent, the idea of "summation over all paths," but it completely ignores the variational principle. Also, the classical limit causes problems when (5.1) and (5.2) are to be interpreted for $\hbar'\to 0$. These shortcomings have no place in the stochastic optimal control formulation. The variational principle exists there by definition and the classical limit is an essential part of this formulation, too.

Before discussing some details of the classical limit, let us concentrate on the problem of how it should be understood in our formulation. For this purpose, a simple diagram giving interpretation of some points of a complex plane is useful.



Complex plane – z

Let us recall that in our formulation the quantum evolution ($z=i\hbar$) is a limit case of generalized evolutions ($z=\hbar''$, i.e., $\mathrm{Re}\,z>0$) and that the classical evolution ($z=0$) may be regarded as a limit case of generalized evolutions, too (this will be more clear later). These limits exists only when $\hbar''\to i\hbar$ and $\hbar''\to 0$ nontangentially to an imaginary axis. In our method it makes no sense therefore to approach classical evolution ($z=0$) converging to 0 over the imaginary axis ($\mathrm{Re}\,z=0$). The only reasonable definition of a classical limit emerges when the nontangential limit $\hbar''\to 0$ is applied to generalized evolutions. The most convenient way to approach classical evolution is, of course, over the real $\hbar''$ (when $\hbar''=\hbar'$), i.e., over stochastically perturbed evolutions. So let us discuss this limit more accurately.

First of all, we expect that the perturbed evolution will tend to a purely classical one if the perturbation parameter $\sqrt{\hbar'/m}$ converges to 0 (we will then be fully justified in calling the evolution obtained from the nontangential limit $\hbar''\to 0$ a classical evolution). This problem is one of stochastic control for "small noise intensities," and has been exhaustively investigated in the literature.[3,6,21] The rigorous results confirm the intuitive predictions. The stochastic optimal policy as a function of $\hbar'$ converges for $\hbar'\to 0$ to the classical optimal one, the optimal stochastic trajectory converges to the optimal (real) classical trajectory, and the action $S_p(x,t)$ tends to the classical action $S(x,t)$. Also, the generalized Hamilton–Jacobi equation (3.8) becomes the classical one (2.6). The more exact results[3,6,21] tell us that optimal stochastic policy $\tilde u^*$, optimal stochastic trajectory $\tilde x^*$, and the action $S_p(x,t)$ can all be expanded with respect to the value $\hbar'/m$ as follows:

$$\begin{aligned}\tilde u^*&=u^*+\frac{\hbar'}{m}u_1+\cdots\left(\frac{\hbar'}{m}\right)^k u_k+o\left(\left(\frac{\hbar'}{m}\right)^k\right),\\ \tilde x^*&=x^*+\frac{\hbar'}{m}x_1+\cdots+\left(\frac{\hbar'}{m}\right)^k x_k+o\left(\left(\frac{\hbar'}{m}\right)^k\right),\\ S_p&=S+\frac{\hbar'}{m}S_1+\cdots+\left(\frac{\hbar'}{m}\right)^k S_k+o\left(\left(\frac{\hbar'}{m}\right)^k\right),\end{aligned}\tag{5.3}$$

where u^*, x^*, and S are appropriate classical variables, and k is any integer. For us, the most interesting coefficients are S_i, $i=1,\dots,k$ of the action expansion. We know already that S is a classical action which fulfills the Hamilton–Jacobi equation (2.6). From the Bellman equation one may get also the equations determining functions S_i for $i=1,\dots,k$. For example, for $i=1$ (WKB approximation), we have

$$\frac{\partial S_1(x,t)}{\partial t}-\frac{1}{m}\frac{\partial S_1(x,t)}{\partial x}\frac{\partial S(x,t)}{\partial x}+\frac{\partial^2 S(x,t)}{\partial x^2}=0,\tag{5.4}$$

where S is a classical action and the final condition for $S_1(x,T)$ is determined by assumed final conditions for S_p and S.

But we must remember also that our definition of generalized evolutions admits the complex-valued final conditions $\varphi(x,T)$ and, *ipso facto*, complex-valued conditions $g(x,T)$ in the Bellman equation (3.7) or (3.8). These must be reflected in the classical limit case if we understand it in the sense given earlier. The most visible consequences are that functions S and S_i will be, respectively, solutions of Eqs. (2.6) and (5.4), but, in general, with complex-valued final conditions!

Let us notice here that our concept of classical limit for quantum evolution has much in common with the one used in Maslov's book,[12] where equations of type (2.6) and (5.4) are obtained in WKB approximation from the (usual, forward) Schrödinger equation, and are analyzed as equations with complex (initial) conditions. Our formulation enables us to see Maslov's approximation scheme from a new point of view, and justifies the application of stochastic perturbation theory in the classical limit of quantum mechanics.

6. DISCUSSION

Maybe it is worthwhile to forestall here some questions of interpretation which always arise when one attempts to embed quantum mechanics in the frame of stochastic processes. Here this problem is much more serious than for the Feynman–Kac formula; there the realization of the $x(t)$ process defined by (5.1) is not believed to be the real path of a particle. This realization has nothing in common with classical trajectory. It is purely random. The situation is, however, different when we consider the realization of a diffusion process $x(t)$ defined by (3.3). It is natural to see it as a possible, fluctuating trajectory of a particle (particles). The situation is, in fact, similar to Nelson's stochastic quantum theory.[22] Thus efforts made earlier to interpret Nelson's theory find application in our work, too. This interpretation is described clearly in Ref. 23 and I merely repeat here, briefly, the main points of that interpretation.

First of all, we are not claiming that particles and trajectories really exist in the physical sense. They should be treated rather as useful but additional construction, and without new information of a physical nature, it is impossible to confirm or exclude their existence. Thus there is no "stochastic interpretation" of quantum mechanics other than the standard one.

Let us say, finally, that the considerations of this paper apply only to a simple system of nonrelativistic, spinless particles in the field of the potential V. They may be generalized to the relativistic case. Some preliminary ideas were sketched in Ref. 24, and more careful analysis is in preparation. It seems, also, that the extension to general Lagrangian systems with configuration space given by a Riemannian manifold is merely a matter of technical complications. In this case, Eq. (3.3) has to be seen as an equation on such a manifold.[25,26] The papers on stochastic mechanics[26,27] suggest that in our formulation spin also may be embedded if configuration space R^3 (of one particle) is replaced by $R^3 \times SU(2)$.

[1]L. Papiez, J. Math. Phys. **23**, 1017 (1982).
[2]E. Nelson, J. Math. Phys. **5**, 332 (1964).
[3]W. H. Fleming and R. W. Rishel, *Deterministic and Stochastic Optimal Control* (Springer-Verlag, Berlin, 1975).
[4]N. V. Krylov, *Controlled Diffusion Processes* (Springer-Verlag, New York, 1980).
[5]W. H. Fleming, SIAM Rev. **11**, 470 (1969).
[6]W. H. Fleming, SIAM J. Control **9**, 473 (1971).
[7]R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).
[8]R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[9]P. A. M. Dirac, Phys. Z. Sowjetunion **3**(1), 64 (1933).
[10]P. A. M. Dirac, *Principles of Quantum Mechanics*, ed. IV (Clarendon, Oxford, 1958).
[11]P. A. M. Dirac, Rev. Mod. Phys. **17**, 195 (1945).
[12]W. P. Maslov, *Kompleksnyi Metod WKB v Nelineinykh Uravneniakh* (Nauka, Moskow, 1977).
[13]A. D. Wentzell and M. I. Freidlin, *Fluktuacii v Dinamitsheskikh Sistemakh pod Deistviem Malykh Slutshainykh Vozmushtshenii* (Nauka, Moscow, 1979).
[14]H. Goldstein, *Classical Mechanics* (Addison-Wesley, Cambridge, MA, 1951).
[15]L. Landau and E. Lifshitz, *Mechanics* (Pergamon, Oxford, 1960).
[16]A. Friedman, *Stochastic Differential Equations and Applications* (Academic, New York, 1975).
[17]A. D. Wentzell, *A Course in the Theory of Stochastic Processes* (McGraw-Hill, New York, 1981).
[18]K. Yosida, *Functional Analysis* (Springer-Verlag, Berlin, 1980).
[19]H. Tanabe, *Equations of Evolution* (Pitman, London, 1979).
[20]M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol. 2* (Academic, New York, 1975).
[21]W. H. Fleming, Rocky Mountain J. Math. **4**, 407 (1974).
[22]E. Nelson, Phys. Rev. **150**, 1079 (1966).
[23]F. Guerra, Phys. Rep. (A Review Section of Physics Letters) A **77**, 263 (1981).
[24]L. Papież, Preprint DIAS-STP-81-32.
[25]K. Itô, Nagoya Math. J. **1**, 35 (1950).
[26]J. P. Caubet, *Le Mouvement Brownian Relativiste* (Springer-Verlag, Berlin, 1976).
[27]D. Dohrn and F. Guerra, Lett. Nuove Cimento **22**, 121 (1978).

Non-self-adjoint Zakharov–Shabat operator with a potential of the finite asymptotic values. II. Inverse problem

Naruyoshi Asano and Yusuke Kato
Department of Engineering Mathematics, Utsunomiya University, Utsunomiya 321, Japan

(Received 29 September 1982; accepted for publication 29 January 1983)

The inverse spectral and scattering problem of the Zakharov–Shabat (ZS) operator is studied. The similarity transformation between ZS operators is examined when their potentials $Q(x)$ have the common nonvanishing asymptotic values $Q_\pm$ at the infinity. The Marchenko equation is derived from the Parseval equation. We give the necessary as well as the sufficient condition of the scattering data for the potential of the specified class.

PACS numbers: 03.65.Nk, 03.80. + r

1. INTRODUCTION

The direct and inverse spectral problem of the one-dimensional Dirac type operator with complex potential or the Zakharov–Shabat (ZS) operator L,

$$Lv \equiv i\sigma_3\{d/dx - Q(x)\}v(x), \quad x\in\mathbb{R} \tag{1.1}$$

$$v=(v_1,v_2)^T, \quad \sigma_3=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \quad Q=\begin{pmatrix}0 & q\\ r & 0\end{pmatrix},$$

was first investigated by Zakharov and Shabat[1] in their analysis of the nonlinear Schrödinger equation and examined systematically by Ablowitz, Kaup, Newell, and Segur[2] for Q vanishing rapidly as $|x|\to\infty$. The study is extended[3,4] to the case of nonvanishing asymptotic values $Q\to Q_\pm \neq 0$ as $x\to\pm\infty$ but under the restriction $q_+r_+ = q_-r_-$. Here and henceforth the double signs $\pm$ are ordered. In the preceding paper[5] (hereafter cited as I) we investigated the direct problem of the ZS operator for general asymptotic values $Q_\pm$.

The values $Q_\pm$ with $u_\pm^2 \equiv q_\pm r_\pm \neq 0$ are classified as C_s and C_d. In the case C_s, $u_+^2 - u_-^2 \in \mathbb{C}-\mathbb{R}$, the continuous spectrum of L is simple, whereas in the case C_d, $u_+^2 - u_-^2 \in\mathbb{R}$, L has the doubly degenerate continuous spectrum. Let us specify as in I the vector or matrix function $A(x)$ by the integrability of $\widetilde{A}_\pm(x) = A(x) - A_\pm$ in such a way that $\widetilde{A}_\pm \in CF_\pm(n)$ means

$$\pm\int_x^{\pm\infty} dy(1+|y|^n)\langle \widetilde{A}_\pm(y)\rangle < \infty, \quad x\in\mathbb{R}$$

for an integer n ($\geqslant 0$), where $\langle A\rangle$ denotes the maximum of the absolute values of the components of A. Further, when $A(x)$ is piecewise absolutely continuous, we write $\widetilde{A}_\pm \in CF'_\pm(n)$ ($n\geqslant 0$) if

$$\int_x^{\pm\infty} (1+|y|^n)\langle dA(y)\rangle < \infty, \quad x\in\mathbb{R}$$

where $dA(y) = A(y+dy) - A(y)$. It is shown in I that the direct spectral problem is solved for $\widetilde{Q}_\pm \in CF_\pm(1)\cap CF'_\pm(1)$ in both cases C_s and C_d under some additional conditions on the point spectrum and on the leaps of Q at its discontinuity. In the present paper we discuss the inverse problem of the same operator with nonvanishing $Q_\pm$ by means of the Marchenko equation[6] and exhibit especially the correspondence between the potential and the spectral function or the scattering data. The latter problem was studied for the case $Q_\pm = 0$ by the present authors[7,8] obtaining the correspondence of

$$Q\in CF_\pm(m)\cap CF'_\pm(n) \leftrightarrow F_\pm \in CF_\pm(m)\cap CF'_\pm(n), \quad m\geqslant 0,\ n\geqslant 1 \tag{1.2}$$

where $F_\pm$ is the kernel of the Marchenko equation and determined from the spectral function. A slightly different type of correspondence was pointed out by Eckhaus and van Harten.[9] In the present paper we establish the correspondence analogous to Eq. (1.2) in the analysis of the inverse problem for the potential of $Q_\pm \neq 0$.

As in I the Riemann plane R for two functions $\lambda_\pm = (\lambda^2 - u_\pm^2)^{1/2}$ is constructed from four complex λ planes connected along branch cuts Γ_j ($1\leqslant j\leqslant 4$) (see Fig. 1). Γ_j is the straight line from $u_\pm$ ($-u_\pm$) to the origin but some of them may be defined as a combination of line segments so as to include the other, as shown in Fig. 1 of I. Our analysis is performed on the first Riemann sheet $R_{\rm I}$, where $\lambda_\pm \to\lambda$ holds as $|\lambda|\to\infty$. The continuous spectrum σ_c of the ZS operator L consists of four curves C_j ($1\leqslant j\leqslant 4$) on $R_{\rm I}$ which originate at $u_\pm$ or $-u_\pm$ and approach asymptotically positive or negative real axes as $|\lambda|\to\infty$. The curves Γ_j and C_j divide $R_{\rm I}$ into four regions R_j ($1\leqslant j\leqslant 4$) in the case C_s and

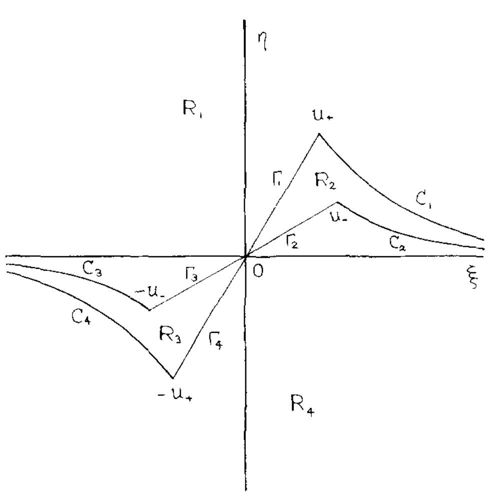


FIG. 1. The first Riemann sheet $R_{\rm I}$ ($\lambda = \xi + i\eta$).

 0022-2488/84/030570-19$02.50

two regions R_1 and R_4 in the case C_d, where C_1 and C_2 (C_3 and C_4) join into a curve to give the doubly degenerate continuous spectrum (see figures in I). The point spectrum σ_p of L consists of the zero points of the components of the matrix $S=\{\tilde{S}_{ij}(\lambda); i,j=1,2\}$, where $\tilde{S}_{11}, \tilde{S}_{21}, \tilde{S}_{12}, \tilde{S}_{22}$ are defined as the Wronskians of the Jost solutions of L and regular in R_1, R_2, R_3, R_4, respectively. Q^s stands for the class of potentials with only the finite number of simple point spectrums and without a point spectrum on C_j. In the case C_s, if the leaps $\delta Q_j \equiv Q(x_j+0)-Q(x_j-0)$ satisfy for some $\epsilon>0$,

$$\Big|\sum_j \delta Q_j e^{i\xi x_j}\Big| > \epsilon\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} \quad (R\ni\xi\to\pm\infty), \tag{1.3}$$

then we write $Q\in Q^\epsilon$. Here for any matrix (or vector) A we define a matrix (or vector) $|A|$ by $|A|_{ij}=|A_{ij}|$ (or $|A|_i=|A_i|$) and the inequality $|A|<|B|$ means $|A_{ij}|\leqslant|B_{ij}|$ ($|A|\neq|B|$) and so on. In this paper as in I we consider only the potential $Q\in Q^s$ in general and $Q\in Q^\epsilon$ in the case C_s.

In Sec. 2 we discuss the similarity transformation in the form of an integral operator $I+\tilde{K}_\pm$ with kernel

$$\tilde{U}_\pm(x,y)=\delta(x-y)+\tilde{K}_\pm(x,y),$$

where $\tilde{K}_\pm(x,y)$ is the Volterra kernel

$$\tilde{K}_\pm(x,y)=0, \quad x\gtrless y.$$

The Jost solution $\Phi^{(1)}_\pm(x,\lambda,\lambda_\pm)$ of the ZS operator L_1 with potential Q_1 is transformed by means of the above $\tilde{K}_\pm$ to the Jost solution $\Phi^{(2)}_\pm(x,\lambda,\lambda_\pm)$ of L_2 with potential Q_2

$$\Phi^{(2)}_\pm(x,\lambda,\lambda_\pm) = \Phi^{(1)}_\pm(x,\lambda,\lambda_\pm) \pm \int_x^{\pm\infty} dy\, \tilde{K}_\pm(x,y)\Phi^{(1)}_\pm(y,\lambda,\lambda_\pm),$$

when $Q_{1\pm}=Q_{2\pm}$. For the simple case where Q_1 is the step potential with a discontinuity at l,

$$Q_1(x)=Q_0(x;l)\equiv Q_\pm, \quad x\gtrless l$$

we estimate $\tilde{K}_\pm$ under the assumption that the perturbation Q_2-Q_1 is small in the sense specified there. In Sec. 3 we derive from the Parseval equation for L_j ($j=1,2$) the Marchenko equation for $\tilde{K}_\pm$ with the kernel $F_\pm$ and also give the estimate of $F_\pm$ in terms of Q. Sections 4 and 5 are devoted to the solution of the inverse problem. The existence, the uniqueness, and the estimate of the solution of the Marchenko equation are given in Sec. 4. In Sec. 5 the main results for the inverse problem are presented. We establish the necessary and sufficient condition of scattering data such that the potential which yields the scattering data belongs to the class $CF_\pm(m)\cap CF'_\pm(n)$ ($m,n\geqslant1$). In the Appendix the spectrum for the step potential $Q_0(x;l)$ is examined in detail, since Q_0 plays an important role corresponding to the constant potential of the Schrödinger operator in the quantum scattering problem.

2. SIMILARITY TRANSFORMATIONS

We study the ZS equation (1.1) with the potential $Q(x)$ having the asymptotic values

$$Q(x)\to Q=\begin{pmatrix}0 & q_\pm\\ r_\pm & 0\end{pmatrix}, \quad x\to\pm\infty. \tag{2.1}$$

Throughout the present paper we assume that $\tilde{Q}_\pm=Q-Q_\pm$ is piecewise differentiable and

$$\tilde{q}_\pm=q-q_\pm, \quad \tilde{r}_\pm=r-r_\pm\in CF'_\pm(0).$$

The matrix Jost solution $\Phi_\pm$ for L is introduced as

$$L\Phi_\pm(x,\lambda,\lambda_\pm)=\lambda\Phi_\pm(x,\lambda,\lambda_\pm), \tag{2.2}$$

$$\Phi_\pm(x,\lambda,\lambda_\pm)\to\Phi^{(0)}_\pm(x,\lambda,\lambda_\pm), \quad x\to\pm\infty, \tag{2.3}$$

where

$$\lambda\in C_1\cup C_4 \quad [\lambda_+=(\lambda^2-u_+^2)^{1/2}\in\mathbb{R}] \quad \text{for } \Phi_+,$$

$$\lambda\in C_2\cup C_3 \quad [\lambda_-=(\lambda^2-u_-^2)^{1/2}\in\mathbb{R}] \quad \text{for } \Phi_-.$$

The solution of ZS equations with a constant potential $Q_\pm$ is expressible as

$$\Phi^{(0)}_\pm(x,\lambda,\lambda_\pm)=T_\pm(\lambda,\lambda_\pm)J(x,\lambda_\pm), \quad \lambda_\pm\in\mathbb{R} \tag{2.4}$$

$$T_\pm(\lambda,\lambda_\pm) = \begin{pmatrix}1 & -iq_\pm/(\lambda+\lambda_\pm)\\ ir_\pm/(\lambda+\lambda_\pm) & 1\end{pmatrix}, \tag{2.5}$$

$$J(x,\lambda_\pm)=\begin{pmatrix}e^{-i\lambda_\pm x} & 0\\ 0 & e^{i\lambda_\pm x}\end{pmatrix}=\exp(-i\sigma_3\lambda_\pm x). \tag{2.6}$$

We also define a "free" Jost solution $\tilde{\Phi}^{(0)}_\pm(x,\lambda,\lambda_\pm)$ as an eigenfunction of the "free" ZS operator $L^{(0)}$ with a "step" potential $Q_0(x)\equiv Q_0(x;0)$,

$$L^{(0)}=i\sigma_3\partial/\partial x-i\sigma_3Q_0(x), \tag{2.7}$$

thus

$$\tilde{\Phi}^{(0)}_\pm(x,\lambda,\lambda_\pm)=\Phi^{(0)}_\pm(x,\lambda,\lambda_\pm), \quad x\gtrless0. \tag{2.8}$$

Hereafter it is assumed without loss of the generality that the step is located at the origin; $l=0$, and further we often write $\Phi^{(0)}_\pm(x)$ in place of $\Phi^{(0)}_\pm(x,\lambda,\lambda_\pm)$ and so on. As exhibited in the Appendix, there are three cases: (1) $L^{(0)}$ has two simple eigenvalue, (2) $L^{(0)}$ has a doubly degenerate eigenvalue, and (3) $L^{(0)}$ has no eigenvalue. In this paper we consider case (3). Case (1) can be discussed analogously. Theorems I-1, I-2, and their corollary in I show that the continuous spectrum of L is stable under such a perturbation of the potential that does not change the asymptotic values $Q_\pm$ if $\tilde{Q}_\pm\equiv Q-Q_\pm$ decreases rapidly as $x\to\pm\infty$. In this sense the free ZS operator $L^{(0)}$ characterizes a family of L with the same continuous spectrum and $Q_\pm$, and we show in this section that there exists a similarity transformation $\tilde{U}_\pm$ between L and $L^{(0)}$,

$$L\tilde{U}_\pm=\tilde{U}_\pm L^{(0)}, \tag{2.9}$$

which connects the "free" Jost solution to the Jost solution,

$$\Phi_\pm(x,\lambda,\lambda_\pm)=\tilde{U}_\pm\tilde{\Phi}^{(0)}_\pm(x,\lambda,\lambda_\pm) \equiv(I+\tilde{K}_\pm)\tilde{\Phi}^{(0)}_\pm(x,\lambda,\lambda_\pm). \tag{2.10}$$

$\tilde{U}_\pm$ should be an operator in the function space, which will be specified later, and (2.10) is expressible as

$$\Phi_\pm(x,\lambda,\lambda_\pm)=\tilde{\Phi}^{(0)}_\pm(x,\lambda,\lambda_\pm) + \int dy\,\tilde{K}_\pm(x,y)\tilde{\Phi}^{(0)}_\pm(y,\lambda,\lambda_\pm). \tag{2.11}$$

First, we introduce transformations $U_\pm=I+K_\pm$ and $U_{\pm0}=I+K_{\pm0}$ by

$$\Phi_\pm(x,\lambda,\lambda_\pm)=\Phi_\pm^{(0)}(x,\lambda,\lambda_\pm)+\int dy\,K_\pm(x,y)\Phi_\pm^{(0)}(y,\lambda,\lambda_\pm),\tag{2.12}$$

$$\tilde\Phi_\pm^{(0)}(x,\lambda,\lambda_\pm)=\Phi_\pm^{(0)}(x,\lambda,\lambda_\pm)+\int dy\,K_{\pm 0}(x,y)\Phi_\pm^{(0)}(y,\lambda,\lambda_\pm).\tag{2.13}$$

We denote a 2×2 matrix A as

$$A=A^D+A^N,\quad A^D=\begin{pmatrix}A_{11} & 0\\ 0 & A_{22}\end{pmatrix},$$
$$A^N=\begin{pmatrix}0 & A_{12}\\ A_{21} & 0\end{pmatrix}.$$

Lemma 2.1: Let $\tilde Q_\pm\in CF_\pm(1)\cap CF'_\pm(0)$, $Q_M=\sup\{|q|+|r|\}$. Then $K_\pm$ has the Volterra kernel

$$K_\pm(x,y)=0,\quad x\gtrless y,\tag{2.14}$$

and is piecewise differentiable with respect to x and y. We have the following estimates:

$$\begin{aligned}\langle K_\pm^N(x,y)-\tfrac12\tilde Q_\pm((x+y)/2)\rangle &\leqslant\tfrac12 Q_M I_\pm((x+y)/2)I_\pm(x)\exp[4Q_M\hat I_\pm(x)]\\ &\leqslant D_\pm(x)I_\pm(x)I_\pm((x+y)/2),\end{aligned}\tag{2.15}$$

$$\begin{aligned}\langle K_\pm^D(x,y)\rangle &\leqslant\tfrac12 Q_M I_\pm((x+y)/2)\exp[6Q_M\hat I_\pm(x)]\\ &\leqslant D_\pm(x)I_\pm((x+y)/2),\end{aligned}$$

$$\langle K_\pm(x,y)\rangle\leqslant D_\pm(x)\tau_\pm(x+y),$$

$$I_\pm(x)\equiv\pm\int_x^{\pm\infty}dy(|\tilde q_\pm(y)|+\tilde r_\pm(y)|),$$

$$\hat I_\pm(x)\equiv\pm\int_x^{\pm\infty}dy\,I_\pm(y)<\infty,$$

$$J_\pm(x)\equiv\pm\int_x^{\pm\infty}(|dq(y)|+|dr(y)|),$$

$$\tau_\pm(x)\equiv I_\pm(x/2)+J_\pm(x/2).$$

Further, let $\tilde Q_\pm\in CF_\pm(1)\cap CF'_\pm(1)$, then the estimates for the partial derivatives are given as

$$\begin{aligned}\langle K_{\pm A}^N(x,y)-\tfrac14 Q'((x+y)/2)\rangle &\leqslant Q_M I_\pm(x)J_\pm((x+y)/2)\\ &\quad+2\{Q_M J_\pm(x)+2Q_M^2 I_\pm^2(x)\}I_\pm((x+y)/2)\\ &\quad\times\exp[4Q_M\hat I_\pm(x)]\\ &\leqslant D_\pm(x)\tau_\pm(x)\tau_\pm(x+y),\end{aligned}\tag{2.16}$$

$$\begin{aligned}\langle K_{\pm A}^D(x,y)\rangle &\leqslant\tfrac14 Q'_M J_\pm((x+y)/2)+\tfrac12 Q_M^2 I_\pm((x+y)/2)\hat J_\pm(x)\\ &\quad+2Q_M Q'_M I_\pm((x+y)/2)I_\pm(x)\exp[Q_M\hat I_\pm(x)]\\ &\leqslant D_\pm(x)\tau_\pm(x+y),\end{aligned}$$

$$Q'_M=Q_M+J_\pm(\mp\infty),$$

$$\hat J_\pm(x)\equiv\pm\int_x^{\pm\infty}dy\,J_\pm(y)<\infty.$$

Here $K_{\pm A}$ denotes $K_{\pm x}\equiv(\partial/\partial x)K_\pm$ or $K_{\pm y}\equiv(\partial/\partial y)K_\pm$ and D_+ (D_-) denotes some decreasing (increasing) function.

The lemma is proved from the integral equation exhibited as I-[(2.24), (2.25)] for $K_\pm^D(M_\pm)$, $K_\pm^N(N_\pm)$. Equation (2.15) is already given in Lemma I-3 [I-(2.23)[10]]. Equation (2.16) is derived from the equation obtainable by differentiating I[(2.24), (2.25)] with respect to x and y and we omit the details of the proof here.

We define the inverse of $U_\pm(x,y)$,

$$U_\pm^{-1}(x,y)=\delta(x-y)+K_\pm^{(-1)}(x,y),$$

which transforms $\Phi_\pm$ to $\Phi_\pm^{(0)}$ as

$$\Phi_\pm^{(0)}(x,\lambda,\lambda_\pm)=\Phi_\pm(x,\lambda,\lambda_\pm)+\int dy\,K_\pm^{(-1)}(x,y)\Phi_\pm(y,\lambda,\lambda_\pm).\tag{2.17}$$

$K_\pm^{(-1)}$ is given formally by

$$\begin{aligned}K_\pm^{(-1)}(x,y)&=\sum_{n=1}^{\infty}(-1)^n K^{(n)}(x,y)\\ &=-K_\pm(x,y)-\int dz\,K_\pm(x,z)K_\pm^{(-1)}(z,y),\end{aligned}\tag{2.18}$$

$$K^{(n)}(x,y)=\int dz\,K_\pm(x,z)K_\pm^{(n-1)}(z,y),$$

$$K_\pm^{(1)}(x,y)=K_\pm(x,y).$$

Lemma 2.2: Under the condition $\tilde Q_\pm\in CF_\pm(1)\cap CF'_\pm(0)$ we have

$$K_\pm^{(-1)}(x,y)=0,\quad x\gtrless y\tag{2.19}$$

$$\langle K_\pm^{(-1)}(x,y)\rangle\leqslant D_\pm(x)\tau_\pm(x+y),\tag{2.20}$$

$$\langle K_\pm^{(-1)N}(x,y)+\tfrac12 Q((x+y)/2)\rangle\leqslant D_\pm(x)I_\pm(x)I_\pm((x+y)/2),\tag{2.21}$$

$$\langle K_\pm^{(-1)D}(x,y)\rangle\leqslant D_\pm(x)I_\pm((x+y)/2),\tag{2.22}$$

and under the condition $\tilde Q_\pm\in CF_\pm(1)\cap CF'_\pm(1)$ we have

$$\langle K_{\pm A}^{(-1)N}(x,y)+\tfrac14 Q'_\pm((x+y)/2)\rangle\leqslant D_\pm(x)\tau_\pm(x)\tau_\pm(x+y),\tag{2.23}$$

$$\langle K_{\pm A}^{(-1)D}(x,y)\rangle\leqslant D_\pm(x)\tau_\pm(x+y).\tag{2.24}$$

Here D_+ (D_-) denotes some decreasing (increasing) function.

Proof: Equation (2.19) is evident from Lemma 2.1 and (2.18). To obtain the estimate of $K_\pm^{(-1)}$, we derive from (2.15), (2.16), and (2.18) the inequalities

$$\begin{aligned}|K_\pm^{(-1)}+K_\pm-K_\pm^{(2)}| &\leqslant\int dz\sum_{n=0}^{\infty}|K_\pm K_\pm^{(n)}|\,|K_\pm^{(2)}|\\ &\leqslant D_\pm(x)\hat\tau_\pm(2x)\exp[D_\pm(x)\hat\tau_\pm(2x)]\\ &\quad\times\begin{pmatrix}1&1\\1&1\end{pmatrix}\sup_{z\gtrless x}\langle K_\pm^{(2)}(z,y)\rangle,\end{aligned}\tag{2.25}$$

$$\begin{aligned}\langle|K_\pm^{(2)}|^N\rangle &\leqslant\int dz(\langle K_\pm^D\rangle\langle K_\pm^N\rangle+\langle K_\pm^N\rangle\langle K_\pm^D\rangle)\\ &\leqslant D_\pm I_\pm(x)I_\pm((x+y)/2),\end{aligned}\tag{2.26}$$

$$\begin{aligned}\langle|K_\pm^{(2)}|^D\rangle &\leqslant\int dz(\langle K_\pm^D\rangle\langle K_\pm^D\rangle+\langle K_\pm^N\rangle\langle K_\pm^N\rangle)\\ &\leqslant D_\pm I_\pm((x+y)/2),\end{aligned}$$

where we used the relation $\langle \widetilde{Q}_\pm(x)\rangle \leqslant J_\pm(x)$. Equations (2.25) and (2.26) yield (2.20). Equation (2.22) is obtainable from the diagonal part of (2.25), (2.26), and (2.15). The off-diagonal part of (2.18) yields the equation for $K^{(-1)N}_\pm$:

$$K^{(-1)N}_\pm(x,y) + H_\pm(x,y) \pm \int_x^y dz\, K^D_\pm(x,z) K^{(-1)N}_\pm(z,y) = 0, \tag{2.27}$$

$$H_\pm(x,y) \equiv -K^N_\pm(x,y) \pm \int_x^y dz\, K^N_\pm(x,z) K^{(-1)D}_\pm(z,y),$$

where by the estimates of $K^N_\pm$ and $K^{(-1)D}_\pm$

$$\langle H_\pm(x,y) + \tfrac{1}{2}\widetilde{Q}_\pm((x+y)/2)\rangle \leqslant D_\pm I_\pm(x) I_\pm((x+y)/2).$$

Then by the iteration of (2.27) we have (2.21). Differentiating (2.18) by x, we have

$$K^{(-1)}_{\pm x}(x,y) + K_{\pm x}(x,y) \mp K_\pm(x,x) K^{(-1)}_\pm(x,y) \pm \int_x^y dz\, K_{\pm x}(x,z) K^{(-1)}_\pm(z,y) = 0, \tag{2.28}$$

and the off-diagonal and diagonal parts of the equation lead to (2.23) and (2.24) for $K^{(-1)}_{\pm x}$, respectively. By the differentiation by y of (2.18), we have

$$K^{(-1)}_{\pm y}(x,y) + L_\pm(x,y) \pm \int_x^y dz\, K_\pm(x,z) K^{(-1)}_{\pm y}(z,y) = 0, \tag{2.29}$$

$$L_\pm(x,y) \equiv K_{\pm y}(x,y) \pm K_\pm(x,y) K^{(-1)}_\pm(y,y),$$

and hence (2.24) for $K^{(-1)D}_{\pm y}$ by the iteration

$$\langle K^{(-1)D}_{\pm y}(x,y)\rangle \leqslant \langle K^{(-1)}_{\pm y}(x,y)\rangle \leqslant D_\pm(x)\tau_\pm(x+y).$$

Taking the off-diagonal part of (2.29), we derive the equation for $K^{(-1)N}_{\pm y}$:

$$K^{(-1)N}_{\pm y}(x,y) + M_\pm(x,y) \pm \int_x^y dz\, K^D_\pm(x,z) K^{(-1)N}_{\pm y}(z,y) = 0,$$

$$M_\pm(x,y) \equiv L^N_\pm(x,y) \pm \int_x^y dz\, K^N_\pm(x,z) K^{(-1)D}_{\pm y}(z,y),$$

which yields (2.23) for $K^{(-1)N}_{\pm y}$.

Lemma 2.3: $K_{\pm 0}(x,y)$ and $K^{(-1)}_{\pm 0}(x,y)$ are continuously differentiable in the region $x \lessgtr y$ except at $x+y=0$ and

$$K_{\pm 0}(x,y) = K^{(-1)}_{\pm 0}(x,y) = 0 \quad \text{for } x+y \gtrless 0 \text{ or } x \gtrless y,$$

$$\langle K_{\pm 0}(x,y)\rangle,\ \langle K^{(-1)}_{\pm 0}(x,y)\rangle \leqslant D_\pm(x)\tau_{\pm 0}(x+y) \quad \text{for } x+y \lessgtr 0,\ x \lessgtr y,$$

where τ_{+0} (τ_{-0}) and D_+ (D_-) are positive decreasing (increasing) and

$$\tau_{\pm 0}(\xi) \to 0, \quad \xi \to \pm\infty,$$

$$\hat{\tau}_{\pm 0}(\xi) \equiv \pm \int_\xi^{\pm\infty} d\eta\, \tau_{\pm 0}(\eta) < \infty, \quad \xi \in \mathbf{R}.$$

The kernel of $\widetilde{U}_\pm = I + \widetilde{K}_\pm$ defined by (2.9) and (2.10) is determined from the relation $\widetilde{U}_\pm = U_\pm U^{-1}_{\pm 0}$ or $\widetilde{K}_\pm = K_\pm + K^{(-1)}_{\pm 0} + K_\pm K^{(-1)}_{\pm 0}$ as

$$\widetilde{K}_\pm(x,y) = \begin{cases} K_\pm(x,y) + K^{(-1)}_{\pm 0}(x,y) \\ \quad \pm \int_x^y dz\, K_\pm(x,z) K^{(-1)}_{\pm 0}(z,y), & x \lessgtr y, \\ 0, & x \gtrless y. \end{cases} \tag{2.30}$$

Lemma 2.4: Let $K_\pm(x,y)$ be given as in Lemmas 2.1 and 2.2, then we have

$$\widetilde{K}_\pm(x,y) = \begin{cases} K_\pm(x,y), & x+y \gtrless 0 \\ 0, & x \gtrless y \end{cases}$$

and $\widetilde{K}_\pm$ is piecewise differentiable with respect to x (y) for each y (x)$\in\mathbf{R}$. The properties of $\widetilde{K}_\pm$ and $\widetilde{K}^{(-1)}_\pm$ are given by Lemmas 2.1 and 2.2 but replacing $K_\pm$ by $\widetilde{K}_\pm$ in (2.14)–(2.16) and $K^{(-1)}_\pm$ by $\widetilde{K}^{(-1)}_\pm$ in (2.19)–(2.24), respectively.

The proof is same as in Lemma 2.2.

3. DERIVATION OF MARCHENKO EQUATION

3.1. Parseval equation

In this subsection, we prove the identity

$$(f,g) = \lim_{\rho\to\infty} (f, \Delta_\rho g), \tag{3.1}$$

where $f \in L^1(\mathbf{R})$ and $g \in C^2_0(\mathbf{R})$ are vector functions, $\rho \in \mathbf{R}$, $(f,g) = \int_{-\infty}^{\infty} dx\, \bar{f}^T(x) g(x)$ with complex conjugate $\bar{f}$ of f, and Δ_ρ is the integral kernel to describe the expansion theorem. Hereafter, we call (3.1) the Parseval equation.

The Green function $G_\lambda(x,y)$ for the operator L is introduced in I as

$$\begin{array}{lll} x<y & x>y & \\ \dfrac{i}{\bar{S}_{11}}\phi_{-1}(x)\phi_{+2}(y)^A, & \dfrac{i}{\bar{S}_{11}}\phi_{+2}(x)\phi_{-1}(y)^A & (\lambda\in R_1), \\ -\dfrac{i}{\bar{S}_{21}}\phi_{-1}(x)\phi_{+1}(y)^A, & -\dfrac{i}{\bar{S}_{21}}\phi_{+1}(x)\phi_{-1}(y)^A & (\lambda\in R_2), \\ \dfrac{i}{\bar{S}_{12}}\phi_{-2}(x)\phi_{+2}(y)^A, & \dfrac{i}{\bar{S}_{12}}\phi_{+2}(x)\phi_{-2}(y)^A & (\lambda\in R_3), \\ -\dfrac{i}{\bar{S}_{22}}\phi_{-2}(x)\phi_{+1}(y)^A, & -\dfrac{i}{\bar{S}_{22}}\phi_{+1}(x)\phi_{-2}(y)^A & (\lambda\in R_4), \end{array} \tag{3.2}$$

where, as in I, for a vector $v = (v_1, v_2)^T$, $v^A = (v_2, v_1)$. An estimate in Lemma 3.1 is necessary for obtaining the expansion theorem.

Lemma 3.1: Let $\widetilde{Q}_\pm \in Q^c \cap Q^c \cap CF'_\pm(1)^{11}$ ($u_\pm \neq 0$), $g(x) \in C^2_0$, and $h(x) \equiv Lg$ with $h'(x) = dh/dx$. Then, we have an estimate for $\mathbf{R} \ni \rho \to \infty$,

$$\left\langle \oint_{C_\rho} \frac{d\lambda}{\lambda} \int_{-\infty}^{\infty} dy\, G_\lambda(x,y) h(y) \right\rangle \leqslant \frac{K}{\rho}(\|h\|_1 + \|h'\| + \|g\|_\infty),$$

where C_ρ is the circle with center at the origin and radius ρ, K a constant depending on Q and g, $\|g\|_1 = \|g_1\|_1 + \|g_2\|_1$, and $\|g\|_\infty = \sup\langle g(x)\rangle$ are $L^1(\mathbf{R})$ and $L^\infty(\mathbf{R})$ norms, respectively.

Proof: We give the details of the proof only for $\lambda \in R_1$.

For $\lambda \in R_1$, the Jost solutions $\phi_{-1}(x)$ and $\phi_{+2}(x)$ are estimated in Lemma I-2 as

$$\phi_{-1}(x) = \begin{pmatrix} 1 \\ \dfrac{ir_-}{\lambda+\lambda_-} \end{pmatrix} e^{-i\lambda_- x} + \psi_{-1}(x), \tag{3.3}$$

$$|\psi_{-1}(x)| < \begin{pmatrix} 1 \\ 1 \end{pmatrix} \frac{e^{\nu_- x}}{|\lambda_-|} C\,[I_-(x)+J_-(x)] \times e^{(C/|\lambda_-|)I_-(x)} \quad (\nu_- \geqslant 0), \tag{3.4}$$

$$\phi_{+2}(x) = \begin{pmatrix} -\dfrac{iq_+}{\lambda+\lambda_+} \\ 1 \end{pmatrix} e^{i\lambda_+ x} + \psi_{+2}(x), \tag{3.5}$$

$$|\psi_{+2}(x)| < \begin{pmatrix} 1 \\ 1 \end{pmatrix} \frac{e^{-\nu_+ x}}{|\lambda_+|} C\,[I_+(x)+J_+(x)] \times e^{(C/|\lambda_+|)I_+(x)} \quad (\nu_+ \geqslant 0), \tag{3.6}$$

where C is a real constant, $\nu_-(\nu_+)$ vanishing on the curve C_1 (C_2), and I_+, J_+ (I_-, J_-) are positive, monotone decreasing (increasing) functions. From these estimates and the asymptotic property of S_{11}, i.e., $S_{11}\to 1$ as $|\lambda|\to\infty$ (Lemma I-6), one has

$$\left\langle \frac{i}{\tilde S_{11}} \phi_{+2}(x) \int_{-\infty}^{x} dy\, \phi_{-1}(y)^A h(y) \right\rangle \leqslant \frac{K}{|\lambda|}(\|\delta h\|_1 + \|h\|_1) p(x), \quad |\lambda| \to \infty; \tag{3.7}$$

here $\|\delta h\|_1 \leqslant 2\int_a^b \langle d(Lg(x))\rangle$, supp $g \subset [a,b]$, K a constant depending on supp g and Q and $p(x)$ is a function $p(x) = e^{-\nu_+ x + \nu_- b}$ $(b \leqslant x)$, $= e^{(\nu_- - \nu_+)x}$ $(a \leqslant x < b)$, $= 0$ $(x < a)$. A similar calculation yields

$$\left\langle \frac{i}{\tilde S_{11}} \phi_{-1}(x) \int_{x}^{\infty} dy\, \phi_{+2}(y)^A h(y) \right\rangle \leqslant \frac{K}{|\lambda|}(\|\delta h\|_1 + \|h\|_1) q(x), \tag{3.8}$$

where $q(x)$ is defined as $q(x) = 0$ $(b < x)$, $= e^{(\nu_- - \nu_+)x}$ $(a < x \leqslant b)$, $= e^{\nu_- x - \nu_+ a}$ $(x \leqslant a)$. Since $|\nu_- - \nu_+| \to 0$ for $|\lambda| \to \infty$, the sum of (3.7) and (3.8) gives

$$\left\langle \int_{-\infty}^{\infty} dy\, G_\lambda(x,y) h(y) \right\rangle \leqslant \frac{K}{|\lambda|}(\|\delta h\|_1 + \|h\|_1) \tag{3.9}$$

for $|\lambda| \to \infty$ in R_1. The right-hand side of (3.9) can be replaced by

$$(K/|\lambda|)(\|h\|_1 + \|h'\|_1 + \|g\|_\infty)$$

via the inequality

$$\|\delta h\|_1 \leqslant \|h'\|_1 + (J_+(a) + J_-(b) + \sup\langle Q\rangle)\|g\|_\infty,$$

which completes the proof of Lemma 3.1 for $\lambda \in R_1$. The proof for λ in other regions R_i $(i \neq 1)$ is similar except at the point of the asymptotic property of $\tilde S_{12}$ $(\tilde S_{21})$ for $\lambda \in R_2$ (R_3): For $\tilde Q^{\pm} \in A^{\epsilon}$, $|\tilde S_{12}^{-1}|$ $(|\tilde S_{21}^{-1}|) \to O(\lambda)$ as $|\lambda| \to \infty$ by Lemma I-6 and I-(3.3) and the integral with respect to $\lambda \in R_2$ (R_3) vanishes as $O(1/\rho)$ for $\rho \to \infty$. Summing the integrals in R_i we establish the Lemma. ∎

The expansion theorem of the funciton $g(x) \in C_0^2(\mathbb{R})$ for the potential $\tilde Q_{\pm}(x) \in Q^{\epsilon} \cap Q^{s}$ proved in I takes the form

$$g(x) = \lim_{\rho\to\infty} \int dy\, \Delta_\rho(x,y) g(y), \tag{3.10}$$

where

$$\Delta_\rho(x,y) = \Delta_\rho^c(x,y) + \Delta_\rho^d(x,y), \tag{3.11}$$

$$\begin{aligned}\Delta_\rho^c(x,y) = \frac{1}{2\pi}\Bigg\{ & \int_{C_{1\rho}} d\lambda\, \frac{d_+}{\tilde S_{11}\tilde S_{21}} \phi_{-1}(x)\phi_{-1}^A(y) \\ & - \int_{C_{2\rho}} d\lambda\, \frac{d_-}{\tilde S_{21}\tilde S_{22}} \phi_{+1}(x)\phi_{+1}^A(y) \\ & - \int_{C_{3\rho}} d\lambda\, \frac{d_-}{\tilde S_{11}\tilde S_{12}} \phi_{+2}(x)\phi_{+2}^A(y) \\ & + \int_{C_{4\rho}} d\lambda\, \frac{d_+}{\tilde S_{12}\tilde S_{22}} \phi_{-2}(x)\phi_{-2}^A(y) \Bigg\},\end{aligned} \tag{3.12}$$

$$\begin{aligned}\Delta_\rho^d(x,y) = & -i\sum_{R_1} \frac{1}{\tilde S'_{11}} \phi_{-1}(x)\phi_{+2}^A(y) + i\sum_{R_2} \frac{1}{\tilde S'_{21}} \phi_{-1}(x)\phi_{+1}^A(y) \\ & - i\sum_{R_3} \frac{1}{\tilde S'_{12}} \phi_{-2}(x)\phi_{+2}^A(y) + i\sum_{R_4} \frac{1}{\tilde S'_{22}} \phi_{-2}(x)\phi_{+1}^A(y).\end{aligned}$$

The integration contour $C_{i\rho}$ $(1 \leqslant i \leqslant 4)$ is the part of C_i within the circle C_ρ in Lemma 3.1 and always directed from left to right. For the later use, we give another form of $\Delta_\rho^c(x,y)$, obtained by substituting the following relation (3.13) into (3.12):

$$\begin{aligned}
\phi_{-1}(x) &= (1/d_+)(\phi_{+1}(x)\tilde S_{11}(\lambda) + \phi_{+2}(x)\tilde S_{21}(\lambda)) && (\lambda \in C_1),\\
\phi_{+1}(x) &= (1/d_-)(\phi_{-1}(x)\tilde S_{22}(\lambda) - \phi_{-2}(x)\tilde S_{21}(\lambda)) && (\lambda \in C_2),\\
\phi_{+2}(x) &= (1/d_-)(-\phi_{-1}(x)\tilde S_{12}(\lambda) + \phi_{-2}(x)\tilde S_{11}(\lambda)) && (\lambda \in C_3),\\
\phi_{-2}(x) &= (1/d_+)(\phi_{+1}(x)\tilde S_{12}(\lambda) + \phi_{+2}(x)\tilde S_{22}(\lambda)) && (\lambda \in C_4).
\end{aligned} \tag{3.13}$$

In fact, we have

$$\begin{aligned}\Delta_\rho^c(x,y) = \frac{1}{2\pi}\int_{C_{1\rho}} \frac{d\lambda}{d_+}\left[\frac{\tilde S_{11}}{\tilde S_{21}}\phi_{+1}(x)\phi_{+1}^A(y) + \phi_{+1}(x)\phi_{+2}^A(y) + \phi_{+2}(x)\phi_{+1}^A(y) + \frac{\tilde S_{21}}{\tilde S_{11}}\phi_{+2}(x)\phi_{+2}^A(y)\right] \\
- \int_{C_{2\rho}} d\lambda\, \frac{d_-}{\tilde S_{21}\tilde S_{22}}\phi_{+1}(x)\phi_{+1}^A(y) - \int_{C_{3\rho}} d\lambda\, \frac{d_-}{\tilde S_{11}\tilde S_{12}}\phi_{+2}(x)\phi_{+2}^A(y) \\
+ \int_{C_{4\rho}} \frac{d\lambda}{d_+}\left\{\frac{\tilde S_{12}}{\tilde S_{22}}\phi_{+1}(x)\phi_{+1}^A(y) + \phi_{+1}(x)\phi_{+2}^A(y) + \phi_{+2}(x)\phi_{+1}^A(y) + \frac{\tilde S_{22}}{\tilde S_{12}}\phi_{+2}(x)\phi_{+2}^A(y)\right\}\end{aligned} \tag{3.14}$$

and also

$$\Delta_\rho^c(x,y)=\frac{1}{2\pi}\int_{C_{1\rho}} d\lambda\,\frac{d_+}{\tilde{S}_{11}\tilde{S}_{21}}\phi_{-1}(x)\phi^A_{-1}(y)+\frac{1}{2\pi}\int_{C_{4\rho}} d\lambda\,\frac{d_+}{\tilde{S}_{12}\tilde{S}_{22}}\phi_{-2}(x)\phi^A_{-2}(y)$$

$$-\frac{1}{2\pi}\int_{C_{2\rho}}\frac{d\lambda}{d_-}\Big[\frac{\tilde{S}_{22}}{\tilde{S}_{21}}\phi_{-1}(x)\phi^A_{-1}(y)-\phi_{-1}(x)\phi^A_{-2}(y)-\phi_{-2}(x)\phi^A_{-1}(y)+\frac{\tilde{S}_{21}}{\tilde{S}_{22}}\phi_{-2}(x)\phi^A_{-2}(y)\Big]$$

$$-\frac{1}{2\pi}\int_{C_{3\rho}}\frac{d\lambda}{d_-}\Big[\frac{\tilde{S}_{12}}{\tilde{S}_{11}}\phi_{-1}(x)\phi^A_{-1}(y)-\phi_{-1}(x)\phi^A_{-2}(y)-\phi_{-2}(x)\phi^A_{-1}(y)+\frac{\tilde{S}_{11}}{\tilde{S}_{12}}\phi_{-2}(x)\phi^A_{-2}(y)\Big].\qquad(3.15)$$

It is noted that (3.14) and (3.15) are expressed by ϕ_{+j} and ϕ_{-j} $(j=1,2)$, respectively.

Let us define the Fourier coefficients of the function $f=(f_1,f_2)^T\in C_0(\mathbb{R})$ by the Jost solutions;

$$\hat{f}_{\pm j}(\lambda)=\int_{-\infty}^{\infty}dy\{\phi_{\pm j_2}(y)f_1(y)+\phi_{\pm j_1}(y)f_2(y)\},\qquad(3.16)$$

$$\hat{f}^*_{\pm j}(\lambda)=\int_{-\infty}^{\infty}dy\{\phi_{\pm j_1}(y)\bar{f}_1(y)+\phi_{\pm j_2}(y)\bar{f}_2(y)\}$$

for $\lambda\in\sigma_c$ or $\lambda\in\sigma_p$. Note that these coefficients are also definable for $f\in L^1(\mathbb{R})$.

Lemma 3.2: Let $\tilde{Q}_\pm\in Q^c\cap Q^s\cap CF'_\pm(1)$, $\mathrm{Im}(u_+^2-u_-^2)\neq 0$, $u_\pm\neq 0$. Then we have the Parseval equation

$$(f,g)=\lim_{\rho\to\infty}(f,\Delta_\rho g)\qquad(3.17)$$

for $f\in L^1$, $g\in C_0^2$, where Δ_ρ is defined by (3.11) with Δ_ρ^c given by (3.12). In (3.17) we may interchange the function spaces of f and g, i.e., $f\in C_0^2$ and $g\in L^1$.

Proof: In terms of the Green function, the resolvent of L, we have obtained in I the identity

$$(1/\lambda)\delta(x-y)+G_\lambda(x,y)=(1/\lambda)G_\lambda(x,y)L\qquad(3.18)$$

in the Hilbert space $(\otimes L^2(\mathbb{R}))^2$ for $\tilde{Q}_\pm\in CF_\pm(0)$ and $\lambda\in\mathbb{C}-\sigma(L)$. Multiplying $f^*=\bar{f}^T\in L^1$ and $g\in C_0^2$ by (3.18) from the left and right sides, respectively, one has, after the integrations by $\lambda\in C_\rho$, $y\in\mathbb{R}$, and $x\in\mathbb{R}$,

$$(f,g)=-\frac{1}{2\pi i}\int dx\int dy\oint_{C_\rho}d\lambda\,f^*(x)G_\lambda(x,y)g(y)$$

$$+\frac{1}{2\pi i}\int dx\int dy\oint_{C_\rho}\frac{d\lambda}{\lambda}f^*(x)G_\lambda(x,y)L\,g(y).\qquad(3.19)$$

The second term of the right-hand side of (3.19) vanishes for $\rho\to\infty$ due to Lemma 3.1, while the substitutions of (3.1), (3.13), and (3.16) into the first term followed by the modification of the integral contour to $C_{j\rho}$ $(1\leqslant j\leqslant 4)$ yield

$$\frac{1}{2\pi}\Big[\int_{C_{1\rho}}d\lambda\,\frac{d_-}{\tilde{S}_{11}\tilde{S}_{11}}\hat{f}^*_{-1}(\lambda)\hat{g}_{-1}(\lambda)$$

$$-\int_{C_{2\rho}}d\lambda\,\frac{d_-}{\tilde{S}_{21}\tilde{S}_{22}}\hat{f}^*_{+1}(\lambda)\hat{g}_{+1}(\lambda)$$

$$-\int_{C_{3\rho}}d\lambda\,\frac{d_-}{\tilde{S}_{11}\tilde{S}_{12}}\hat{f}^*_{+2}(\lambda)\hat{g}_{+2}(\lambda)$$

$$+\int_{C_{4\rho}}d\lambda\,\frac{d_+}{\tilde{S}_{12}\tilde{S}_{22}}\hat{f}^*_{-2}(\lambda)\hat{g}_{-2}(\lambda)\Big]$$

$$-i\sum_{R_1}\frac{1}{\tilde{S}'_{11}}\hat{f}^*_{-1}\hat{g}_{+2}+i\sum_{R_2}\frac{1}{\tilde{S}'_{21}}\hat{f}^*_{-1}\hat{g}_{+1}$$

$$-i\sum_{R_3}\frac{1}{\tilde{S}'_{12}}\hat{f}^*_{-2}\hat{g}_{+2}+i\sum_{R_4}\frac{1}{\tilde{S}'_{22}}\hat{f}^*_{-2}\hat{g}_{+1}.$$

Equation (3.17) can easily be obtained from this expression by rewriting this in terms of Δ_ρ^c and Δ_ρ^d defined by (3.11) and (3.12). ■

For later use, we write the Parseval equation in terms of Δ_ρ^c [(3.14) or (3.15)] instead of (3.12).

Lemma 3.3: Let $\tilde{Q}_\pm\in Q^c\cap Q^s\cap CF'_\pm(1)$, $\mathrm{Im}(u_+^2-u_-^2)\neq 0$, $u_\pm\neq 0$, $g\in C_0^2$, and $f\in L^1$ with supp f bounded from below. Then we have the Parseval equation

$$(f,g)=\lim_{\rho\to\infty}(f,\Delta_\rho g),\qquad(3.20)$$

where Δ_ρ is defined by (3.11) and (3.14). If supp f is bounded from above, (3.20) still holds with Δ_ρ given (3.11) and (3.15).

Proof: For proof of the first part, it may suffice to note that, for $x\in\operatorname{supp}f\cap\operatorname{supp}g$, $\phi_{+1}(x)$ and $\phi_{+2}(x)$ are bounded on C_1, C_4, C_2 and C_1, C_4, C_3, respectively, and that $d\lambda/\lambda_+=d\lambda_+/\lambda$ is also finite on these curves. Then, the integrals defining the right-hand side of (3.20) converge to give the same value as that of (3.17). The second part of the lemma is similarly proved. ■

3.2. Marchenko equation

From the Parseval equations derived above, we obtain the Marchenko equation by means of the distributions. Let us define the convergence $f_m\Rightarrow 0$ $(m\to\infty)$ for a sequence

$\{f_m\} \in C_0^\infty$ as

(i) There exists a compact set D such that $\{\operatorname{supp} f_m\} \subset D$,

$$\text{(ii)} \sup_{x\in D} |f_m^{(p)}| \to 0 \quad (m\to\infty) \ (p=0,1,\ldots), \tag{3.21}$$

where p denotes the p-times differentiation. Then a functional $T(f) \in \mathbb{C}$ ($f \in C_0^\infty$) is said to be distribution iff it is continuous, i.e., $T(f_m) \to 0$ as $f_m \Rightarrow 0$.[12]

Lemma 3.4: Let $\tilde{Q}_\pm \in CF_\pm(1) \cap CF'_\pm(0)$, $u_+^2 - u_-^2 \in \mathbb{C} - \mathbb{R}$, and $u_\pm \neq 0$. Then the Fourier coefficients $\hat{f}_{m,\pm j}(\lambda)$ and $\hat{f}^*_{m,\pm j}(\lambda)$ of $f_m \in C_0^\infty$ $(m = 1, 2, \ldots)$ are analytic (resp. continuous) in $\lambda \in R_1$ for which $\phi_{\pm j}(x,\lambda)$ is analytic (resp. continuous) and have the properties

$$\hat{f}_{m,\pm j}, \hat{f}^*_{m,\pm j} = O(1/|\lambda|), \quad |\lambda| \to \infty, \tag{3.22}$$

and

$$|\hat{f}_{m,\pm j}|, \quad |\hat{f}^*_{m,\pm j}| \leqslant K \frac{\|f_m\|_\infty + \|f'_m\|_\infty}{1+|\lambda|} \to 0, \quad m\to\infty, \tag{3.23}$$

if $f_m \Rightarrow 0$ $(m\to\infty)$, where K is a constant depending on Q.

Proof: For definiteness, we consider the case $\operatorname{Im}(u_+^2 - u_-^2) > 0$ and give the proof for $\hat{f}_{m,+1}$ only. As defined by (3.16), $\hat{f}_{m,+1}$ is determined by $\phi_{+1}(y,\lambda,\lambda_+)$, which is analytic or continuous for λ such that $\operatorname{Im}\lambda_+ \leqslant 0$, i.e., $\lambda \in R_2 \cup R_4$ or $\lambda \in C_1 \cup C_4$. If we present $\phi_{+1}(x)$ in the form

$$\phi_{+1}(x) = \begin{pmatrix} 1 \\ \dfrac{ir_+}{\lambda+\lambda_+} \end{pmatrix} e^{-i\lambda_+ x} + \psi_{+1}, \tag{3.24}$$

ψ_{+1} is estimated, from Lemmas I-1 and I-2, as

$$|\psi_{+1}| < \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{\nu_+ x} c \int_x^\infty I_+(y)dy \exp\left(c\int_x^\infty I_+(y)dy\right) \quad (\nu_+ \leqslant 0) \tag{3.25}$$

or

$$|\psi_{+1}| < \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{\nu_+ x} \frac{c}{|\lambda_+|} [I_+(x) + J_+(x)] e^{(c/|\lambda_+|)I_+(x)} \quad (\nu_+ \leqslant 0). \tag{3.26}$$

Since the domain of the integration (3.16) is finite, (3.25) gives the absolutely convergent upper bound to the integration (3.16) and hence leads to the analyticity and continuity of $\hat{f}_{m,+1}$ according to those of ϕ_{+1}. The substitution of (3.24) into (3.16) yields

$$\begin{aligned} |\hat{f}_{m,+1}| &\leqslant \left| \int_{-\infty}^\infty dx \left(\frac{ir_+}{\lambda+\lambda_+} f_{m,1} + f_{m,2} \right) e^{-i\lambda_+ x} \right| \\ &\quad + \left| \int_{-\infty}^\infty dx (\psi_{+12} f_{m,1} + \psi_{+11} f_{m,2}) \right| \\ &\leqslant \frac{K}{|\lambda_+|} (\|f'_m\|_\infty + \|f_m\|_\infty) \quad (|\lambda| \to \infty), \end{aligned} \tag{3.27}$$

i.e., (3.22), where we have used the integration by parts and (3.26) to get the last estimate. It is clear that (3.16) is bounded by const $\times \|f_m\|_\infty$ for a finite $|\lambda_+|$ in virtue of (3.25); hence, taking (3.27) into account, we have (3.23). ■

Now let us introduce a function $F_{\pm\rho}(x,y)$, from which the integral kernel for the Marchenko equation is constructed. By $\Delta(f)$, we denote the difference of the function $f(\lambda)$ evaluated for the general potential $Q(x)$ and $f(\lambda)^{(0)}$ for the step potential $Q_0(x)$; $\Delta(f) = f(\lambda) - f(\lambda)^{(0)}$.

$$\begin{aligned} F_{+\rho}(x,y) = \frac{1}{2\pi} \Bigg\{ &\int_{C_{1\rho}} \frac{d\lambda}{d_+} \left[\Delta\left(\frac{\tilde{S}_{11}}{\tilde{S}_{21}}\right) \tilde{\phi}^{(0)}_{+1}(x) \tilde{\phi}^{(0)}_{+1}(y)^A + \Delta\left(\frac{\tilde{S}_{21}}{\tilde{S}_{11}}\right) \tilde{\phi}^{(0)}_{+2}(x) \tilde{\phi}^{(0)}_{+2}(y)^A \right] \\ &+ \int_{C_{4\rho}} \frac{d\lambda}{d_+} \left[\Delta\left(\frac{\tilde{S}_{12}}{\tilde{S}_{22}}\right) \tilde{\phi}^{(0)}_{+1}(x) \tilde{\phi}^{(0)}_{+1}(y)^A + \Delta\left(\frac{\tilde{S}_{22}}{\tilde{S}_{12}}\right) \tilde{\phi}^{(0)}_{+2}(x) \tilde{\phi}^{(0)}_{+2}(y)^A \right] \\ &- \int_{C_{2\rho}} d\lambda\, d_- \Delta\left(\frac{1}{\tilde{S}_{21}\tilde{S}_{22}}\right) \tilde{\phi}^{(0)}_{+1}(x) \tilde{\phi}^{(0)}_{+1}(y)^A - \int_{C_{3\rho}} d\lambda\, d_- \Delta\left(\frac{1}{\tilde{S}_{12}\tilde{S}_{11}}\right) \tilde{\phi}^{(0)}_{+2}(x) \tilde{\phi}^{(0)}_{+2}(y)^A \Bigg\} \\ &- i\sum_{R_1} \frac{1}{d_+} \left(\frac{\gamma_k}{\tilde{S}'_{11}}\right) \tilde{\phi}^{(0)}_{+2}(x) \tilde{\phi}^{(0)}_{+2}(y)^A + i\sum_{R_2} \frac{1}{d_+} \left(\frac{\gamma_k}{\tilde{S}'_{21}}\right) \tilde{\phi}^{(0)}_{+1}(x) \tilde{\phi}^{(0)}_{+1}(y)^A \\ &- i\sum_{R_3} \frac{1}{d_+} \left(\frac{\gamma_k}{\tilde{S}'_{12}}\right) \tilde{\phi}^{(0)}_{+2}(x) \tilde{\phi}^{(0)}_{+2}(y)^A + i\sum_{R_4} \frac{1}{d_+} \left(\frac{\gamma_k}{\tilde{S}'_{22}}\right) \tilde{\phi}^{(0)}_{+1}(x) \tilde{\phi}^{(0)}_{+1}(y)^A, \end{aligned} \tag{3.28}$$

$$\begin{aligned} F_{-\rho}(x,y) = \frac{1}{2\pi} \Bigg\{ &\int_{C_{1\rho}} d\lambda\, d_+ \Delta\left(\frac{1}{\tilde{S}_{11}\tilde{S}_{21}}\right) \tilde{\phi}^{(0)}_{-1}(x) \tilde{\phi}^{(0)}_{-1}(y)^A + \int_{C_{4\rho}} d\lambda\, d_+ \Delta\left(\frac{1}{\tilde{S}_{12}\tilde{S}_{22}}\right) \tilde{\phi}^{(0)}_{-2}(x) \tilde{\phi}^{(0)}_{-2}(y)^A \\ &+ \int_{C_{2\rho}} \frac{d\lambda}{d_-} \left[\Delta\left(\frac{\tilde{S}_{22}}{\tilde{S}_{21}}\right) \tilde{\phi}^{(0)}_{-1}(x) \tilde{\phi}^{(0)}_{-1}(y)^A + \Delta\left(\frac{\tilde{S}_{21}}{\tilde{S}_{22}}\right) \tilde{\phi}^{(0)}_{-2}(x) \tilde{\phi}^{(0)}_{-2}(y)^A \right] \\ &- \int_{C_{3\rho}} \frac{d\lambda}{d_-} \left[\Delta\left(\frac{\tilde{S}_{12}}{\tilde{S}_{11}}\right) \tilde{\phi}^{(0)}_{-1}(x) \tilde{\phi}^{(0)}_{-1}(y)^A + \Delta\left(\frac{\tilde{S}_{11}}{\tilde{S}_{12}}\right) \tilde{\phi}^{(0)}_{-2}(x) \tilde{\phi}^{(0)}_{-2}(y)^A \right] \Bigg\} \\ &- i\sum_{R_1} d_+ \left(\frac{1}{\gamma_k S'_{11}}\right) \tilde{\phi}^{(0)}_{-1}(x) \tilde{\phi}^{(0)}_{-1}(y)^A + i\sum_{R_2} \frac{1}{d_-} \left(\frac{1}{\gamma_k \tilde{S}'_{21}}\right) \tilde{\phi}^{(0)}_{-1}(x) \tilde{\phi}^{(0)}_{-1}(y)^A \\ &- i\sum_{R_3} \frac{1}{d_-} \left(\frac{1}{\gamma_k S'_{12}}\right) \tilde{\phi}^{(0)}_{-2}(x) \tilde{\phi}^{(0)}_{-2}(y)^A + i\sum_{R_4} d_+ \left(\frac{1}{\gamma_k \tilde{S}'_{22}}\right) \tilde{\phi}^{(0)}_{-2}(x) \tilde{\phi}^{(0)}_{-2}(y)^A, \end{aligned} \tag{3.29}$$

where γ_k is a constant defined at zeros of $\tilde{S}_{ij}$, i.e., σ_p by the relations

$$\phi_{-1}(x) = \gamma_k \phi_{+2}(x), \quad \lambda_k \in \sigma_p \subset R_1$$
$$\phi_{-1}(x) = \gamma_k \phi_{+1}(x), \quad \lambda_k \in \sigma_p \subset R_2,$$
$$\phi_{-2}(x) = \gamma_k \phi_{+2}(x), \quad \lambda_k \in \sigma_p \subset R_3,$$
$$\phi_{-2}(x) = \gamma_k \phi_{+1}(x), \quad \lambda_k \in \sigma_p \subset R_4.$$

We have assumed the absence of the point spectrum of $L^{(0)}$.

Lemma 3.5: Let $\tilde{Q}_\pm \in Q^\epsilon \cap Q^s \cap CF'_\pm(1)$, $\mathrm{Im}(u_+^2 - u_-^2) \neq 0$, $u_\pm \neq 0$. Then $F_{\pm\rho}(x,y)$ is continuous function of x and y and bounded for $x \gtrless x_0 \in \mathbb{R}$ at each fixed y and also bounded for $y \gtrless y_0 \in \mathbb{R}$ at each fixed x. Further, there exists

$$F_\pm(x,y) = \lim_{\rho\to\infty} F_{\pm\rho}(x,y) \tag{3.30}$$

in the sense of the distribution on y at each x, satisfying

$$\int F_\pm(x,y) f(y) dy < K \quad (x \gtrless x_0), \tag{3.31}$$

where $f \in C_0^\infty$ and K is a constant depending on x_0, y_0, and f. In (3.31) one may interchange the role of x and y.

Proof: We give the details of the proof only for $F_+(x,y)$. Since $\tilde{Q}_\pm \in CF_\pm(0) \cap CF'_\pm(0)$, Lemmas I-1 and I-2 in I give the estimates of the type (3.24)–(3.26) for $\tilde{\phi}^{(0)}_{+j}(x)$ $(j = 1, 2)$. Hence, $\tilde{\phi}^{(0)}_{+1}(x)$ is bounded and analytic for $x \geqslant x_0$ and $\lambda \in R_2 \cup R_4$ and continuous at $\lambda \in C_1 \cup C_2 \cup C_4$ while $\tilde{\phi}^{(0)}_{+2}(y)$ has the corresponding properties for $y \geqslant y_0$ and $\lambda \in R_1 \cup R_3$ and at $\lambda \in C_1 \cup C_3 \cup C_4$, respectively. The domains of the integrations in (3.28) are finite and $d_+ \propto (\lambda \mp u_+)^{1/2}$ near $\lambda = \pm u_+$ on $C_{1\rho}$ or $C_{4\rho}$; hence the integrations in (3.28) are convergent. The first half of the lemma can be obtained from these facts.

The later half of the lemma will be proved by seeing; for ${}^\forall f(x) \in C_0^\infty$ and $\mu > \rho \to \infty$,

$$\int F_{+\mu}(x,y) f(y) dy - \int F_{+\rho}(x,y) f(y) dy \to 0. \tag{3.32}$$

Let us first consider the term $\int_{C_{4\rho}} (d\lambda / d_+) \Delta (\tilde{S}_{12}/\tilde{S}_{22}) \tilde{\phi}^{(0)}_{+1}(x) \tilde{\phi}^{(0)}_{+1}(y)^A$ in (3.28). Noting the properties $|\tilde{S}_{12}| \sim O(1/|\lambda|)$, $|\tilde{S}_{22}| \sim 1$, $d\lambda / d_+ \sim d\lambda_+$ for $\lambda_+ \to \infty$ and $|\tilde{\phi}^{(0)}_{+1}(x)| < {}^\exists K$ for $x \geqslant x_0$, $|\lambda| \geqslant {}^\exists \rho_0$ as deduced from (3.24) and (3.26) and applying (3.22) of Lemma 3.4, one has

$$\left| \int_{C_{4\mu} - C_{4\rho}} \frac{d\lambda}{d_+} \Delta \left(\frac{\tilde{S}_{12}}{\tilde{S}_{22}} \right) \tilde{\phi}^{(0)}_{+1}(x) \hat{f}_{+1}(\lambda) \right|$$
$$\leqslant \int_\rho^\mu d\lambda_+ \frac{K}{|\lambda_+|^2} \to 0 \quad (\rho \to \infty)$$

as required. If we use the estimate of $\tilde{\phi}^{(0)}_{+2}(x)$ corresponding to (3.24) and (3.26), similar discussions also apply to the term $\int_{C_{1\rho}} (d\lambda / \lambda_+) \Delta (\tilde{S}_{21}/\tilde{S}_{11}) \tilde{\phi}^{(0)}_{+2}(x) \tilde{\phi}^{(0)}_{+2}(y)^A$ in (3.28). Next, we estimate the terms due to the sum $\int_{C_{1\rho}} (d\lambda / d_+) \Delta (\tilde{S}_{11}/\tilde{S}_{21}) \tilde{\phi}^{(0)}_{+1} \tilde{\phi}^{(0)A}_{+1} - \int_{C_{2\rho}} d\lambda d_- \Delta (1/\tilde{S}_{21}\tilde{S}_{22}) \tilde{\phi}^{(0)}_{+1} \tilde{\phi}^{(0)A}_{+1}$. As verified from the Lemma I-6, we may put $\tilde{S}_{11} = T + \delta\tilde{S}_{11}$ and $\tilde{S}_{22}^{-1} = T + \delta\tilde{S}_{22}^{-1}$, where $T = 1 - (\pi_-(0) - \pi_+(0))/2i\lambda$, while δS_{11} and δS_{22}^{-1} are of the order of $|\lambda|^{-2}$. Then we have an identity

$$\int_{C_{1\mu} - C_{1\rho}} \frac{d\lambda}{d_+} \Delta \left(\frac{\tilde{S}_{11}}{\tilde{S}_{21}} \right) \tilde{\phi}^{(0)}_{+1} \int \tilde{\phi}^{(0)A}_{+1} f dy$$
$$- \int_{C_{2\mu} - C_{2\rho}} d\lambda\, d_- \Delta \left(\frac{1}{\tilde{S}_{21}\tilde{S}_{22}} \right) \tilde{\phi}^{(0)}_{+1} \int \tilde{\phi}^{(0)}_{+1} f dy$$
$$= \int_{C_{1\mu} - C_{1\rho} - C_{2\mu} + C_{2\rho}} \frac{d\lambda}{d_+} \Delta \left(\frac{T}{\tilde{S}_{21}} \right) \tilde{\phi}^{(0)}_{+1} \hat{f}_{+1}$$
$$+ \int_{C_{1\mu} - C_{1\rho}} \frac{d\lambda}{d_+} \Delta \left[\frac{1}{\tilde{S}_{21}} (\tilde{S}_{11} - T) \right] \tilde{\phi}^{(0)}_{+1} \hat{f}_{+1}$$
$$- \int_{C_{2\mu} - C_{2\rho}} d\lambda\, d_- \Delta \left[\frac{1}{\tilde{S}_{21}} \left(\frac{1}{\tilde{S}_{22}} - \frac{T}{d_+ d_-} \right) \right] \tilde{\phi}^{(0)}_{+1} \hat{f}_{+1}, \tag{3.33}$$

where $\hat{f}_{+1}$ is given by (3.16) for $\phi_{+1} = \tilde{\phi}^{(0)}_{+1}$. The integrand in the first term of the right-hand side of (3.33) is analytic in R_2 and continuous on $C_{1\mu} - C_{1\rho}$ and $C_{2\mu} - C_{2\rho}$, and on two segments L_ρ and L_μ connecting these curves to form a closed contour. Thus the first term reduces to the integrations on L_ρ and L_μ by the Cauchy theorem, from which we have (see Fig. 2)

$$\left| \int_{C_{1\mu} - C_{1\rho} - C_{2\mu} + C_{2\rho}} \frac{d\lambda}{d_+} \Delta \left(\frac{T}{\tilde{S}_{21}} \right) \tilde{\phi}^{(0)}_{+1} \hat{f}_{+1} \right|$$
$$< \left| \int_{L_\rho} \frac{d\lambda}{d_+} \Delta \left(\frac{T}{\tilde{S}_{21}} \right) \tilde{\phi}^{(0)}_{+1} \hat{f}_{+1} \right| + \left| \int_{L_\rho} \frac{d\lambda}{d_+} \Delta \left(\frac{T}{\tilde{S}_{21}} \right) \tilde{\phi}^{(0)}_{+1} \hat{f}_{+1} \right|. \tag{3.34}$$

It is not difficult to see that the right-hand side of (3.34) tends to zero for $\rho \to \infty$ if one notes the length of $L_\rho \sim O(1/\rho)$, $|\Delta(1/\tilde{S}_{21})| \sim O(\rho)$, $|\hat{f}_{+1}| \sim O(1/\rho)$, $|\tilde{\phi}^{(0)}_{+1}| < K$ for $x \geqslant x_0$ and $|\rho| \geqslant \rho_0$, as shown by Lemmas I-6, 3.4, I-1, and I-2.

The second and third terms of the right-hand side of (3.33) are also shown to tend to zero as $\rho \to \infty$ if the integrands of these terms are $O(1/|\lambda|^2)$. In fact, the estimates of each factor given above imply the required property of the integrands. Thus the left-hand side of (3.33) tends to zero as $\rho \to \infty$.

We can repeat the similar method for the rest of the terms in (3.32), i.e., $\int_{C_{4\mu} - C_{4\rho}} (d\lambda / d_+) \Delta (\tilde{S}_{22}/\tilde{S}_{21}) \tilde{\phi}^{(0)}_{+2} \hat{f}_{+2} - \int_{C_{3\mu} - C_{3\rho}} d\lambda\, d_- \Delta (1/\tilde{S}_{12}\tilde{S}_{11}) \tilde{\phi}^{(0)}_{+2} \hat{f}_{+2}$, which completes the proofs of (3.32) and hence Lemma 3.5 for $F_+(x,y)$. ■

We are now in the position to describe the Marchenko equation. In Sec. 2, we have studied the properties of the kernels for the similarity transformation $K_\pm(x,y)$ and $K_\pm^{(-1)}(x,y)$. The Marchenko equation is the relation between $\tilde{K}_\pm$ and $F_\pm$ and will be used in Secs. 4 and 5 to determine $\tilde{K}_\pm$ from $F_\pm$ in the inverse problem. As shown in Sec. 2, the similarity transformations from $\tilde{\Phi}^{(0)}_+$ to Φ_+ and the converse are presented by the operators $\tilde{U}_+$ and $\tilde{U}_+^{(-1)}$, respectively. We rewrite (3.11) with $\Delta_\rho^c(x,y)$ given by (3.14) simply as $\Delta_\rho(x,y) = \Phi_+(x) \Omega_\rho \Phi_+^A(y)$, where Ω_ρ is a formal operator of the 2×2 matrix acting on both $\Phi_+(x)$ and $\Phi_+^A(y)$,

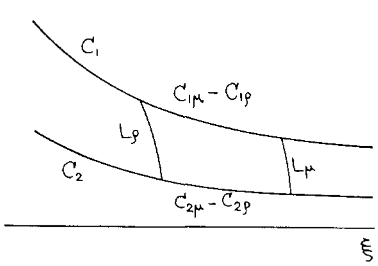


FIG. 2. Integration contour for Eq. (3.34).

$$\Omega_\rho = \frac{1}{2\pi}\begin{pmatrix} \overleftrightarrow{\int_{C_{4\rho}+C_{1\rho}}} d\lambda \frac{1}{d_+}, & \overleftrightarrow{\int_{C_{4\rho}}} d\lambda \frac{1}{d_+}\frac{\tilde S_{12}}{\tilde S_{22}} + \overleftrightarrow{\int_{C_{1\rho}}} d\lambda \frac{1}{d_+}\frac{\tilde S_{11}}{\tilde S_{21}} - \overleftrightarrow{\int_{C_{2\rho}}} d\lambda \frac{d_-}{\tilde S_{21}\tilde S_{22}} \\ \overleftrightarrow{\int_{C_{4\rho}}} d\lambda \frac{1}{d_+}\frac{\tilde S_{22}}{\tilde S_{12}} + \overleftrightarrow{\int_{C_{1\rho}}} d\lambda \frac{1}{d_+}\frac{\tilde S_{21}}{\tilde S_{11}} - \overleftrightarrow{\int_{C_{3\rho}}} d\lambda \frac{d_-}{\tilde S_{11}\tilde S_{12}}, & \overleftrightarrow{\int_{C_{4\rho}+C_{1\rho}}} d\lambda \frac{1}{d_+} \end{pmatrix}$$

$$+ i\begin{pmatrix} -\overleftrightarrow{\sum_{R_1}} \frac{\gamma_k}{\tilde S'_{11}}, & \overleftrightarrow{\sum_{R_2}} \frac{\gamma_k}{\tilde S'_{21}} \\ -\overleftrightarrow{\sum_{R_3}} \frac{\gamma_k}{\tilde S'_{12}}, & \overleftrightarrow{\sum_{R_4}} \frac{\gamma_k}{\tilde S'_{22}} \end{pmatrix}, \tag{3.35}$$

where $\leftrightarrow$ means that the functions on the left- and right-hand sides of $\leftrightarrow$ should be integrated or summed and Φ^A is the adjoint matrix of Φ:

$$\Phi^A = \begin{pmatrix} \phi_{22} & \phi_{12} \\ \phi_{21} & \phi_{11} \end{pmatrix}.$$

We denote Ω_ρ as $\Omega_\rho^{(0)}$ for the step potential $Q_0(x)$.

Lemma 3.6: For $\tilde Q_\pm \in Q^\epsilon \cap Q^s \cap CF'_\pm(1)$, $\mathrm{Im}(u_+^2 - u_-^2) \neq 0$, and $u_\pm \neq 0$, $F_\pm(x,y)$ is the function defined for almost all x, y and satisfies

$$\tilde K_\pm(x,y) + F_\pm(x,y) \pm \int_x^{+\infty} dz\, \tilde K_\pm(x,z) F_\pm(z,y) = 0 \quad (y \gtrless x). \tag{3.36}$$

$F_\pm(x,y)$ is bounded, integrable with respect to $y \in [x, \pm\infty)$ for fixed x.

Proof: Let us consider $F_+(x,y)$ and the case $\mathrm{Im}(u_+^2 - u_-^2) > 0$. Defining $\Delta_\rho^{(0)}$ by $\Delta_\rho^{(0)} = \Phi_+(x)\Omega_\rho \Phi_+^{(0)}(y)^A$, we can see that for $f, g \in C_0^2$, the form $(f, \Delta_\rho^{(0)} g) \equiv (f(x), \Phi_+(x)\Omega_\rho \tilde\Phi_+^{(0)}(y)^A g(y))$ has a finite value for $\rho < \infty$. Note that this $\Delta_\rho^{(0)}$ depends on y parametrically and yields two expressions $(f, \Delta_\rho^{(0)} g) = (f(x), \tilde U_+ \tilde\Phi_+^{(0)} \Omega_\rho \tilde\Phi_+^{(0)}(y)^A g(y)) = (f(x), \Phi_+(x)\Omega_\rho(\tilde U_+^{-1}\Phi_+)(y)^A G(y))$. Subtraction of these two expressions and a short calculation lead to

$$\begin{aligned} 0 &= (f, \tilde\Phi_+^{(0)} \Omega_\rho^{(0)} \tilde\Phi_+^{(0)A} g) + (f, \tilde K_+ \tilde\Phi_+^{(0)} \Omega_\rho^{(0)} \tilde\Phi_+^{(0)A} g) \\ &\quad + (f, \tilde\Phi_+^{(0)} \Delta\Omega_\rho \tilde\Phi_+^{(0)A} g) + (f, \tilde K_+ \tilde\Phi_+^{(0)} \Delta\Omega_\rho \tilde\Phi_+^{(0)A} g) \\ &\quad - (f, \Phi_+ \Omega_\rho \Phi_+^A g) - (f, \Phi_+ \Omega_\rho \Phi_+^A \tilde K_+^{(-1)A} g), \end{aligned} \tag{3.37}$$

where $\Delta\Omega_\rho = \Omega_\rho - \Omega_\rho^{(0)}$. In (3.37), let $\rho \to \infty$ and use Lemmas 3.2, 3.3, and 3.5, then

$$\begin{aligned} 0 &= (f,g) + (\tilde K_+^* f, g) + (f, F_+ g) \\ &\quad + (\tilde K_+^* f, F_+ g) - (f,g) - (f, \tilde K_+^{(-1)A} g), \end{aligned} \tag{3.38}$$

where each term corresponds to that in (3.37) in this order and $\tilde K_+^*$ is the adjoint operator of $\tilde K_+$. Since $g(x)$ and $(F_+ g)(x)$ are bounded for $x \geqslant x_0$ as shown in Lemma 3.5 and $\tilde K_+(\xi, \eta)$ is integrable with respect to η by Lemma 2.2, we have from (3.38) and Lemma 3.5

$$(\tilde K_+ + F_+ + \tilde K_+ F_+ - \tilde K_+^{(-1)A}) g = 0$$

for almost all x and

$$\tilde K_+(x,y) + F_+(x,y) + \int_x^\infty dz\, \tilde K_+(x,z) F_+(z,y) - \tilde K_+^{(-1)A}(y,x) = 0. \tag{3.39}$$

Hence, by Lemma 2.4, (3.36) holds in the sense of the distribution on y for almost all x. As shown in Sec. 3.3 the expression of F_+ in the Neumann series of $\tilde K_+$ by means of (3.36) and (3.39) converges with the upper bound $D_\pm(x)\tau_\pm(x+y) \times \exp[D_\pm(x)\hat\tau_\pm(2x)]$ by Lemma 2.4 and it is also easy to see by (3.31) that this gives the unique F_+ satisfying (3.36). ∎

3.3. Properties of $F_\pm(x,y)$

In this subsection we observe how the properties of $Q(x)$ yield those of $F_\pm(x,y)$ via the Marchenko equation (3.36).

Lemma 3.7: For $\tilde Q_\pm \in CF_\pm(1) \cap CF'_\pm(1)$, $F_\pm(x,y)$ has the partial derivatives $F_{\pm x}(x,y)$ and $F_{\pm y}(x,y)$ and the following estimates in terms of $\tilde Q_\pm$:

$$\langle F_\pm(x,y)\rangle \leqslant D_\pm(z)\tau_\pm(x+y), \tag{3.40}$$

$$\langle F_\pm^D(x,y)\rangle \leqslant D_\pm(z) I_\pm((x+y)/2), \tag{3.41}$$

$$\langle F_\pm^N(x,y) - \tfrac{1}{2}\tilde Q_\pm((x+y)/2)\rangle \leqslant D_\pm(z) I_\pm(z) I_\pm((x+y)/2), \tag{3.42}$$

$$\langle F_{\pm A}^D(x,y)\rangle \leqslant D_\pm(z)\tau_\pm(x+y), \tag{3.43}$$

$$\langle F_{\pm A}^N(x,y) - \tfrac{1}{4} Q'_\pm((x+y)/2)\rangle \leqslant D_\pm(z)\tau_\pm(z)\tau_\pm((x+y)/2), \tag{3.44}$$

where

$$z \equiv \begin{Bmatrix} \min \\ \max \end{Bmatrix}(x,y)$$

is defined corresponding to the double signs $\pm$ and D_+ (D_-) denotes some positive decreasing (increasing) function.

Proof: From (2.14), (2.19), and (3.39) we have the relation

$$F_\pm(x,y) \pm \int_x^{\pm\infty} dz\, \tilde K_\pm(x,z) F_\pm(z,y) = N_\pm(x,y) \equiv \begin{cases} -\tilde K_\pm(x,y), & y \gtrless x, \\ \tilde K_\pm^{(-1)A}(y,x), & y \lessgtr x, \end{cases} \tag{3.45}$$

in the sense of the distribution on y for each x. Hereafter we discuss F_+ as an example and omit the suffix $+$ for brevity. First consider the case $y \geqslant x$. As in (2.25) and (2.26) we have by the formal iteration of (3.45) in the operator form and by Lemma 2.4:

$$F = N - \tilde K N + \tilde K \tilde K F = N - \tilde K N - \sum_{n=1}^\infty (-1)^n \tilde K^n \tilde K N, \tag{3.46}$$

$$|F-N+\widetilde{K}N|(x,y)\leqslant\int_x^\infty dz\sum_{n=1}^\infty|\widetilde{K}^n|(x,z)|\widetilde{K}N|(z,y)$$
$$\leqslant D(x)\hat{\tau}(2x)\exp[D(x)\hat{\tau}(2x)]$$
$$\times\begin{pmatrix}1&1\\1&1\end{pmatrix}\sup_{z>x}\langle\widetilde{K}N(z,y)\rangle,$$
$$\langle|\widetilde{K}N|^N\rangle\leqslant D(x)I(x)I((x+y)/2),$$
$$\langle|\widetilde{K}N|^D\rangle\leqslant D(x)I((x+y)/2)$$

which, again by Lemma 2.4, yield (3.40) and (3.41) and thus we see that F is defined for $x, y\in\mathbb{R}$ as described in Lemma 3.6. The off-diagonal part of (3.45) gives the equation for F^N as

$$F^N+\int_x^\infty dz\,\widetilde{K}^DF^N=M\equiv N^N-\int_x^\infty dz\,\widetilde{K}^NF^D. \quad (3.47)$$

Since by Lemma 2.4 and (3.41) we have

$$\langle M(x,y)+\tfrac{1}{2}\widetilde{Q}((x+y)/2)\rangle<DI(x)I((x+y)/2),$$

the iteration of (3.47) leads to (3.42). Differentiating (3.45) with respect to x or y we have the relations

$$F_x(x,y)-\widetilde{K}(x,x)F(x,y)+\int_x^\infty dz\,K_x(x,z)F(z,y)$$
$$=N_x(x,y),$$
$$F_y(x,y)+\int_x^\infty dz\,K(x,z)F_y(z,y)=N_y(x,y).$$

The estimates of F_A^D, F_A^N $(A=x,y)$ are obtained from the relations using the estimates of $\widetilde{K}$ and $\widetilde{K}^{(-1)}$ in Lemma 2.4 and of F given by (3.40)–(3.42). The analysis is quite similar to that given for the derivation of (2.23) and (2.24) for $K_{\pm}^{(-1)}$ from (2.28) and (2.29) in Lemma 2.2 and we omit the details. We proved the estimates of $F_{\pm}$ (2.40)–(2.44) for $y\gtrless x$ $(z=x)$. The estimates for $y\lessgtr x$ $(z=y)$ are given from the definition of $F_{\pm}$ by (3.28)–(3.30). ■

4. SOLUTION OF THE MARCHENKO EQUATION

4.1. Marchenko equation

The Marchenko equation (3.36) is analyzed in the function space $L^\alpha(\mathbb{R}_{\pm,x})\equiv\{\otimes L^\alpha(\mathbb{R}_{\pm,x})\}^4$ $(\alpha=1,2)$, which is defined in the sense that $\psi\equiv\{\psi_{ij}\}\in L^\alpha(\mathbb{R}_{\pm,x})$ with its norm $\|\psi\|_\alpha\equiv\{\Sigma_{i,j=1}^2\|\psi_{ij}\|_\alpha^\alpha\}^{1/\alpha}$. $L_{1+2}(\mathbb{R}_{x,\pm})$ is the set of functions ψ such that $\psi=\psi_1+\psi_2$, $\psi_\alpha\in L^\alpha(\mathbb{R}_{\pm,x})$ $(\alpha=1,2)$. We also define $L_{1\cap 2}(\mathbb{R}_{\pm,x})\equiv L^1(\mathbb{R}_{\pm,x})\cap L^2(\mathbb{R}_{\pm,x})$. A linear operator $\mathbb{F}_{\pm}^x$ in $L^\alpha(\mathbb{R}_{\pm,x})$ with a parameter $x\in\mathbb{R}$ is defined in terms of the integral kernel $F_{\pm}(\xi,\eta)$ of the Marchenko equation (3.36) as

$$\psi=\phi\mathbb{F}_{\pm}^x$$

or

$$\psi(y)=\pm\int_x^{\pm\infty}dz\,\phi(x)F_{\pm}(z,y)$$

for $\phi\in L^\alpha(\mathbb{R}_{\pm,x})$ $(\alpha=1,2)$, where $\mathbb{F}_{\pm}^x$ is put on the right of ϕ, since $\phi F_{\pm}$ is the product of matrices ϕ and $F_{\pm}$.

The following conditions are introduced for the kernel of $\mathbb{F}_{\pm}^x$.

$[C_{\pm}]$ There exist positive decreasing (increasing) functions h_+ and c_+ (h_- and c_-) such that

$$\langle\hat{F}_{\pm}(\xi,\eta)\rangle<c_{\pm}(\zeta)h_{\pm}(\xi+\eta),\quad \zeta=\begin{Bmatrix}\min\\\max\end{Bmatrix}(\xi,\eta), \quad (4.1)$$

$$\hat{h}_{\pm}(x)\equiv\pm\int_x^{\pm\infty}dt\,h_{\pm}(t)<\infty,\quad z\in\mathbb{R}. \quad (4.2)$$

$[S_{\pm}]$ $F_{\pm}(\xi,\eta)$ is piecewise differentiable with respect to ξ and η, and there exist positive functions $f_{\pm}^{(0)}(t)$, $g_{\pm}^{(0)}(t)$, $f_{\pm}^r(s,t)$, $g_{\pm}^r(s,t)$, $h_{\pm}^{Df}(t)$, and $g_{\pm}^{Dg}(t)$ such that

$$\langle F_{\pm}^N(\xi,\eta)\rangle\leqslant f_{\pm}^{(0)}(\xi+\eta)+f_{\pm}^r(2\zeta,\xi+\eta),$$
$$\langle F_{\pm A}^N(\xi,\eta)\rangle\leqslant g_{\pm}^{(0)}(\xi+\eta)+g_{\pm}^r(2\zeta,\xi+\eta),$$
$$\zeta\equiv\begin{Bmatrix}\min\\\max\end{Bmatrix}(\xi,\eta), \quad (4.3)$$
$$\langle F_{\pm}^D(\xi,\eta)\rangle\leqslant h_{\pm}^{Df}(\xi+\eta),\quad \langle F_{\pm A}^D(\xi,\eta)\rangle\leqslant h_{\pm}^{Dg}(\xi+\eta),$$

where $f_{\pm}^r(s,t)$, $g_{\pm}^r(s,t)$ defined in $\{s,t; s\lessgtr t\}$ are monotone with respect to s and t and

$$f_{\pm}^{(0)},\ g_{\pm}^{(0)}\in CF_{\pm}(1),\quad h_{\pm}^{Df},\ h_{\pm}^{Dg}\in CF_{\pm}(0),$$
$$\int_a^{\pm\infty}ds\int_s^{\pm\infty}dt\,f_{\pm}^r(s,t),$$
$$\int_a^{\pm\infty}ds\int_s^{\pm\infty}dt\,g_{\pm}^r(s,t)<\infty,\quad a\in\mathbb{R}. \quad (4.4)$$

For $\rho_{\pm}^{(0)}, \rho_{\pm}^r$ $(\rho=f,g)$ introduced in the condition $S_{\pm}$, the functions

$$\hat{\rho}_{\pm}^{(0)}(s)\equiv\pm\int_s^{\pm\infty}du\,\rho_{\pm}^{(0)}(u),$$
$$\rho_{\pm}^{(1)}(s,t)\equiv\pm\int_s^t du\,\rho_{\pm}^r(u,t),\quad (\rho=f,g) \quad (4.5)$$
$$\rho_{\pm}^{(2)}(s)\equiv\pm\int_s^{\pm\infty}dv\,\rho_{\pm}^r(s,v),$$

exist with

$$\hat{\rho}_{\pm}^{(0)}(x),\ \rho_{\pm}^{(1)}(a,x),\ \rho_{\pm}^{(2)}(x)\in CF_{\pm}(0)\quad(\text{for each } a\in\mathbb{R}),$$
$$\rho_{\pm}^r(x,x)\in CF_{\pm}(1),$$

and the following lemma is obtainable.

Lemma 4.1.: Under condition $S_{\pm}$, condition $C_{\pm}$ is satisfied with

$$h_{\pm}(t)\equiv\sum_{\rho=f,g}\{h_{\pm}^{D\rho}(t)+h_{\pm}^{N\rho}(t)\}\in CF_{\pm}(0),\quad 2a\lessgtr t, \quad (4.6)$$

where $h_{\pm}^{D\rho}\in CF_{\pm}(0)$ is given by (4.3) and (4.4) and $h_{\pm}^{N\rho}\in CF_{\pm}(0)$ is defined as

$$h_{\pm}^{N\rho}(t)=\hat{\rho}_{\pm}^{(0)}(t)+\tfrac{1}{2}\rho_{\pm}^{(1)}(2a,t)+\rho_{\pm}^{(2)}(t)$$
$$(\rho=f,g),\quad 2a\lessgtr t. \quad (4.7)$$

Further, we have the following relations:

$$\langle F_{\pm}^N(\xi,\eta)\rangle\leqslant h_{\pm}^{Ng}(\xi+\eta),\quad \eta\gtrless\xi\gtrless a, \quad (4.8)$$

$$\pm\int_x^{\pm\infty}d\xi\,\langle F_{\pm}^N(\xi,y)\rangle\leqslant h_{\pm}^{Nf}(x+y),$$
$$\pm\int_x^{\pm\infty}d\xi\,\langle F_{\pm A}^N(\xi,y)\rangle\leqslant h_{\pm}^{Ng}(x+y), \quad (4.9)$$

$$\pm\int_x^{\pm\infty} dy\,\langle F^N_{\pm}(x,y)\rangle \leqslant h^{Nf}_{\pm}(2x), \qquad x \gtrless a. \tag{4.10}$$
$$\pm\int_x^{\pm\infty} dy\,\langle F^N_{\pm A}(x,y)\rangle \leqslant h^{Ng}_{\pm}(2x),$$

Note that $h^{D\rho}_{\pm}, h^{N\rho}_{\pm}$ $(\rho = f, g)$ exist for each $a\in\mathbb{R}$.

Proof: By (4.3) we have (4.8) for F^N_+ :

$$\begin{aligned}
&\langle F^N_+(\xi,\eta)\rangle \\
&\leqslant \int_\xi^\infty ds \langle F^N_{+s}(s,\eta)\rangle \quad (a\leqslant\xi\leqslant\eta) \\
&\leqslant \hat{g}^{(0)}_+(\xi+\eta) + \int_\xi^\eta ds\, g'_+(2s, s+\eta) \\
&\quad + \int_\eta^\infty ds\, g'_+(2\eta, s+\eta) \\
&\leqslant \hat{g}^{(0)}_+(\xi+\eta) + \int_\xi^\infty ds\, g'_+(2s, \xi+\eta) \\
&\quad + \int_\xi^\infty ds\, g'_+(\xi+\eta, s+\eta) \\
&\leqslant \hat{g}^{(0)}_+(\xi+\eta) + \tfrac{1}{2} g^{(1)}_+(2\xi, \xi+\eta) \\
&\quad + g^{(2)}_+(\xi+\eta) \leqslant h^{Ng}_+(\xi+\eta).
\end{aligned}$$

All the other relations are obtainable in an analogous way. ■

Lemma 4.2: Let $F_{\pm}$ satisfy $C_{\pm}$. Then $\mathbb{F}^x_{\pm}$ is completely continuous in $L^\alpha(\mathbb{R}_{\pm,x})$ for $\alpha = 1$ and 2 and

$$\left.\begin{aligned} &\|F^x_{\pm}\|_1 \leqslant 2\hat{h}_{\pm}(2x) \\ &\|F^x_{\pm}\|_2^2 \leqslant 4h_{\pm}(2x)\hat{h}_{\pm}(2x) \end{aligned}\right\} \to 0, \quad x\to\pm\infty. \tag{4.11}$$

There is $x^{(\alpha)}_{\pm}\in\mathbb{R}$ such that the inverse $(I+\mathbb{F}^x_{\pm})^{-1}$ in $L^\alpha(\mathbb{R}_{\pm,x})$ exists for $x \gtrless x^{(\alpha)}_{\pm}$ $(\alpha = 1, 2)$. Further

$$f^x_{\pm}(y) \equiv F_{\pm}(x,y) \in L_{1\cap 2}(\mathbb{R}_{\pm,x}). \tag{4.12}$$

Proof: We consider the case of $\mathbb{F}^x_-$. Put $\psi = \phi\mathbb{F}^x_-$ and let $\phi\in L^1(\mathbb{R}_{-,x})$, then

$$\begin{aligned}
\|\phi\mathbb{F}^x_-\|_1 &\leqslant 2\sum_{i,j}\int_{-\infty}^x dy \int_{-\infty}^x dz |\phi_{ij}(y)| h_-(y+z) \\
&= 2\sum \int_{-\infty}^x dy \int_{-\infty}^{x+y} dz |\phi_{ij}(y)| h_-(z) \\
&= 2\|\phi\|_1 \hat{h}_-(2x),
\end{aligned}$$

i.e., $\|\mathbb{F}^x_-\|_1 \leqslant 2\hat{h}_-(2x)$. Since by (4.1) $F_-(\xi,\eta)$ is dominated by means of the folding kernel $h_-(\xi+\eta)$ $(\xi,\eta<x)$ which is integrable with respect to $\eta\in\mathbb{R}_{-,x}$ by (4.2), the compactness of $\mathbb{F}^x_-$ in $L^1(\mathbb{R}_{-,x})$ follows in a standard way and we omit the details. Further,

$$\begin{aligned}
&\sum_{i,j}\iint dy\,dz |F_{-ij}(y,z)|^2 \\
&\quad \leqslant 4\int_{-\infty}^x \int_{-\infty}^x dy\,dz\, h^2_-(y+z) \leqslant 4\hat{h}^2_-(2x) < \infty.
\end{aligned}$$

Hence $\mathbb{F}^x_-$ is an operator of the Hilbert–Schmidt type in $L^2(\mathbb{R}_{-,x})$. The existence of $x^{(\alpha)}_{\pm}$ $(\alpha = 1, 2)$ is evident from the asymptotic properties (4.11). Equation (4.12) follows from (4.1) and (4.2). ■

We express the Marchenko equation (3.36) in the form

$$\phi + f^x_{\pm} + \phi\mathbb{F}^x_{\pm} = 0 \tag{4.13}$$

or

$$\phi(y) + f^x_{\pm}(y) \pm \int_x^{\pm\infty} dz\, \phi(z) F_{\pm}(z,y) = 0, \quad y \gtrless x, \tag{4.14}$$

and find the solution $\phi\in L^\alpha(\mathbb{R}_{\pm,x})$ for $\alpha = 1$ or 2.

4.2. Uniqueness and the estimate of the solution

Lemma 4.3: Let $F_{\pm}$ satisfy condition $C_{\pm}$, then the Marchenko equation (4.13) has at least a solution $\phi\in L^\alpha(\mathbb{R}_{\pm,x})$ for $\alpha = 1$ and 2. The solution ϕ is unique in $L_{1+2}(\mathbb{R}_{\pm,x})$, if and only if the homogeneous equation

$$\psi + \psi\mathbb{F}^x_{\pm} = 0 \tag{4.15}$$

in $L^2(\mathbb{R}_{\pm,x})$ has only the zero solution $\psi = 0$. We note $\phi\in L_{1\cap 2}(\mathbb{R}_{\pm,x})$.

Proof: By Lemma 4.2, (4.13) has at least a solution ϕ_α in $L^\alpha(\mathbb{R}_{\pm,x})$ for each α. Note that, by (4.1),

$$\begin{aligned}
\langle(\phi_1\mathbb{F}^x_{\pm})(y)\rangle &= \left\langle \int_x^{\pm\infty} dz\, \phi_1(z) F_{\pm}(z,y)\right\rangle \\
&\leqslant \sup_{z\gtrless x}\langle F_{\pm}(z,y)\rangle \|\phi_1\|_1 \leqslant h_{\pm}(x+y)\|\phi_1\|_1, \\
\|\phi_1\mathbb{F}^x_{\pm}\|_2^2 &\leqslant \pm 4\int_x^{\pm\infty} dy \langle(\phi_1 F^x_{\pm})(y)\rangle^2 \\
&\leqslant 4\hat{h}_{\pm}(2x) h_{\pm}(2x)\|\phi_1\|_1^2,
\end{aligned}$$

i.e., $\phi_1\mathbb{F}^x_{\pm}\in L^2(\mathbb{R}_{\pm,x})$. Hence, if (4.15) in $L^2(\mathbb{R}_{\pm,x})$ has only the zero solution $\psi = 0$, then (4.15) in $L^1(\mathbb{R}_{\pm,x})$ has only the zero solution and ϕ_α is the unique solution in $L^\alpha(\mathbb{R}_{\pm,x})$ $(\alpha = 1, 2)$. Since $\phi_1\in L^2(\mathbb{R}_{\pm,x})$ by $\phi_1\mathbb{F}^x_{\pm}, f^x_{\pm}\in L^2(\mathbb{R}_{\pm,x})$, and (4.13), we see $\phi_1\in L_{1\cap 2}(\mathbb{R}_{\pm,x})$ and thus is the unique solution of (4.13) in $L_{1+2}(\mathbb{R}_{\pm,x})$. ■

We define the open interval $I_{\pm}$ such that, for $x\in I_{\pm}$, (4.15) in $L^2(\mathbb{R}_{\pm,x})$ has only the zero solution. Let $I^{(0)}_{\pm}$ be the largest of $(x^{(2)}_{\pm}, \pm\infty)$, then

$$I^{(0)}_{\pm} \subset I_{\pm}. \tag{4.16}$$

The formal differentiation of (3.36) with respect to y or x leads to the following relations:

$$\tilde{K}_{\pm y}(x,y) + F_{\pm y}(x,y) \pm \int_x^{\pm\infty} dz\, \tilde{K}_{\pm}(x,z) F_{\pm y}(z,y) = 0, \quad y \gtrless x, \tag{4.17}$$

$$\tilde{K}_{\pm x}(x,y) + G_{\pm}(x,y) \pm \int_x^{\pm\infty} dz\, \tilde{K}_{\pm x}(x,z) F_{\pm}(z,y) = 0, \quad y \gtrless x, \tag{4.18}$$

$$G_{\pm}(x,y) \equiv F_{\pm x}(x,y) \mp \tilde{K}_{\pm}(x,x) F_{\pm}(x,y).$$

Lemma 4.4: Under condition $C_{\pm}$, (4.17) and (4.18) hold for $x\in I_{\pm}$ and

$$\tilde{K}_{\pm}(x,y),\ \tilde{K}_{\pm x}(x,y),\ \tilde{K}_{\pm y}(x,y) \in L_{1\cap 2}(\mathbb{R}_{\pm,x}) \quad \text{for } x\in I_{\pm} \tag{4.19}$$

as functions of y.

Proof: The existence of $\tilde{K}_{\pm A}$ for almost all (x, y) is shown by approximating $F_{\pm}$ by the infinite set $\{F^{(n)}_{\pm} \in C^\infty(\mathbb{R}^2); n = 1,2,\ldots\}$ and it can also be shown that $\tilde{K}_{\pm A}$ satisfies (4.17) and (4.18). The analysis is similar to the one given by Agranovich and Marchenko[13] and Kato[8] and we omit it. Equation (4.19) follows easily from $C_\pm$ and (4.17) and (4.18) for $x\in I_\pm$. ■

Lemma 4.5: Under condition $S_\pm$, the following estimates hold for the solution of (4.13), $\phi=\widetilde{K}_\pm(x,y)$:

$$\langle \widetilde{K}_\pm(x,x)+F_\pm(x,x)\rangle \leqslant D_\pm(x)h_\pm(2x),\quad x\in I_\pm^{(0)}, \tag{4.20}$$

$$\langle \widetilde{K}^N_\pm(x,x)+F^N_\pm(x,x)\rangle \leqslant D_\pm(x)h_\pm(2x)\{h^{Nf}_\pm(2x) + h_\pm(2x)\hat{h}^{Nf}_\pm(2x)\},\quad x\in I_\pm^{(0)}, \tag{4.21}$$

$$\langle \widetilde{K}_{\pm A}(x,x)+F_{\pm A}(x,x)\rangle \leqslant D_\pm(x)h_\pm(2x),\quad x\in I_\pm^{(0)}, \tag{4.22}$$

$$\langle \widetilde{K}^N_{\pm A}(x,x)+F^N_{\pm A}(x,x)\rangle \leqslant D_\pm(x)[h_\pm(2x)\{h^{Ng}_\pm(2x)+h^{Nf}_\pm(2x)\} + h^2_\pm(2x)\{\hat{h}^{Ng}_\pm(2x)+\hat{h}^{Nf}_\pm(2x)\}],\quad x\in I_\pm^{(0)}, \tag{4.23}$$

where, by (4.3), (4.6), and (4.8)

$$\langle F_\pm(x,x)\rangle \leqslant h_\pm(2x),\qquad \langle F^N_\pm(x,x)\rangle \leqslant \begin{cases} f^{(0)}_\pm(2x)+f^r_\pm(2x,2x),\\ h^{Ng}_\pm(2x),\end{cases} \tag{4.24}$$

$$\langle F^D_{\pm A}(x,x)\rangle \leqslant h^{Dg}_\pm(2x),\qquad \langle F^N_{\pm A}(x,x)\rangle \leqslant g^{(0)}_\pm(2x)+g^r_\pm(2x,2x), \tag{4.25}$$

and $D_+(x)$ $[D_-(x)]$ is some positive decreasing [increasing] function.

Proof: From (4.13) we have

$$\widetilde{K}_\pm = -f^x_\pm(I+\mathbb{F}^x_\pm)^{-1}\in L_{1\cap 2},\quad x\in I_\pm$$

and by (4.1)

$$\|\widetilde{K}_\pm\|_1 \leqslant \|f^x_\pm\|_1\|(I+\mathbb{F}^x_\pm)^{-1}\|_1 \leqslant D_\pm(x)\hat{h}_\pm(2x),\quad x\in I^{(0)}_\pm.$$

From (4.1) and (4.13) we obtain

$$\langle \widetilde{K}_\pm(x,y)+F_\pm(x,y)\rangle < \|\widetilde{K}_\pm\|_1 \sup_{z\gtrless x}\langle F_\pm(z,y)\rangle \leqslant D_\pm\hat{h}_\pm(2x)h_\pm(x+y)$$

and hence (4.20). The off-diagonal part of (4.13) gives the equation for $\widetilde{K}^N_\pm$,

$$\widetilde{K}^N_\pm + F^N_\pm \pm \int_x^{\pm\infty} dz\,\widetilde{K}^D_\pm F^N_\pm \pm \int_x^{\pm\infty} dz\,\widetilde{K}^N_\pm F^D_\pm = 0 \tag{4.26}$$

with the kernel of $\mathbb{F}^{x,D}_\pm$, $F^D_\pm(\xi,\eta)$, and the inhomogeneous term $F^N_\pm \pm \int dz\,\widetilde{K}^D_\pm F^N_\pm$. Solving (4.26), we have the unique solution

$$\widetilde{K}^N_\pm = -\left(F^N_\pm \pm \int dz\,\widetilde{K}^D_\pm F^N_\pm\right)(I+\mathbb{F}^{x,D}_\pm)^{-1},\quad x\in I_\pm \tag{4.27}$$

giving by (4.9) and (4.10)

$$\begin{aligned}\pm\int_x^{\pm\infty} dy\,\langle \widetilde{K}^N_\pm(x,y)\rangle &\leqslant \|(I+\mathbb{F}^{x,D}_\pm)^{-1}\|_1 \times\Big[\pm\int_x^{\pm\infty} dy\langle F^N_\pm(x,y)\rangle \\ &\quad + \sup_{z\gtrless x}\langle \widetilde{K}^D_\pm(x,z)\rangle \int_x^{\pm\infty}\int_x^{\pm\infty} dy\,dz\langle F^N_\pm(z,y)\rangle\Big] \\ &\leqslant D_\pm[h^{Nf}_\pm(2x)+h_\pm(2x)\hat{h}^{Nf}_\pm(2x)].\end{aligned} \tag{4.28}$$

Hence from (4.26), (4.9), (4.20), and (4.28), one has

$$\begin{aligned}\langle K^N_\pm(x,y)+F^N_\pm(x,y)\rangle &\leqslant \left\langle \int_x^{\pm\infty} dz(\widetilde{K}^D_\pm F^N_\pm + \widetilde{K}^N_\pm F^D_\pm)\right\rangle \\ &\leqslant \pm\sup_{z\gtrless x}\langle \widetilde{K}^D_\pm(x,z)\rangle\int_x^{\pm\infty} dz\langle F^N_\pm(z,y)\rangle \\ &\quad \pm\sup_{z\gtrless x}\langle F^D_\pm(z,y)\rangle\int_x^{\pm\infty} dz\langle \widetilde{K}^N_\pm(x,z)\rangle \\ &\leqslant D_\pm[h_\pm(2x)h^{Nf}_\pm(x+y)+h_\pm(x+y) \\ &\quad \times(h^{Nf}_\pm(2x)+h_\pm(2x)\hat{h}^{Nf}_\pm(2x))],\end{aligned} \tag{4.29}$$

giving (4.21) for $x=y$. After an analogous calculation using (4.20), (4.21), (4.24), (4.25), and (4.29) we obtain (4.22) and (4.23) from (4.17) and (4.18). We omit the details here. ■

5. INVERSE PROBLEM

5.1. Theorems of the inverse problem

We are considering the perturbation of the ZS operator under the restriction

$$\Delta Q \equiv Q(x)-Q_0(x)\to 0,\quad x\to\pm\infty. \tag{5.1}$$

In the inverse problem we determine L from a given $L^{(0)}$ [(2.7)] and given scattering data $S=\{\tilde{S}_{ij}\}$. The point spectrum of $L^{(0)}$ is examined in the Appendix for simple examples. Since it is determined from the zeroes of the algebraic equation of λ, its dimension is finite. The continuous spectrum of L is identical to that of $L^{(0)}$. A typical configuration of the continuous spectrum is given for $\mathrm{Re}\,u^2_- > \mathrm{Re}\,u^2_+ > 0$, $\mathrm{Im}\,u^2_+ > \mathrm{Im}\,u^2_- > 0$ from the group $C_s(u^2_+ - u^2_- \in \mathbb{C}-\mathbb{R})$ and $\mathrm{Re}\,u^2_- > \mathrm{Re}\,u^2_+ > 0$ ($\mathrm{Im}\,u^2_+ = \mathrm{Im}\,u^2_-$) from the group $C_d(u^2_+ - u^2_- \in\mathbb{R})$ and we discuss the details only for these cases. Most of the other cases can be treated similarly except the cases $u_+=0$ or $u_-=0$, where the separate study is needed. The continuous spectrum for the case C_s is illustrated in Fig. 1. The straight lines Γ_j $(1\leqslant j\leqslant 4)$ constitute the cut on the complex λ plane and λ_+ (λ_-) changes its sign when λ crosses Γ_1 and Γ_4 (Γ_2 and Γ_3) from the upper side, Γ_{1+} and Γ_{4+} (Γ_{2+} and Γ_{3+}), to the lower side, Γ_{1-} and Γ_{4-} (Γ_{2-} and Γ_{3-}).

The main results of the inverse problem are summarized in Theorems 5.1 and 5.2 under the following conditions I and II on the scattering data.

I. (i) $\tilde{S}_{ij}(\lambda)$ is regular in the following respective region R_i;

$$\tilde{S}_{11}\ \lambda\in R_1,\quad \tilde{S}_{21}\ \lambda\in R_2,\quad \tilde{S}_{12}\ \lambda\in R_3,\quad \tilde{S}_{22}\ \lambda\in R_4,$$

and has a set of finite number of simple zeros,

$\sigma_l \equiv \{\lambda_k;\ 1\leqslant k\leqslant N_l\}\in R_l \quad (1\leqslant l\leqslant 4)$,

$\tilde S_{11}=0$ for $\lambda_k\in\sigma_1$, $\tilde S_{12}=0$ for $\lambda_k\in\sigma_3$,

$\tilde S_{21}=0$ for $\lambda_k\in\sigma_2$, $\tilde S_{22}=0$ for $\lambda_k\in\sigma_4$.

(ii) $\tilde S_{ij}(\lambda)$ is continuous on the boundary of the region of analyticity given above and has no zero point on the boundary. Moreover it satisfies the following relation on Γ_j:

$$\begin{aligned}
\tilde S_{11} &= \frac{iq_+}{\lambda+\lambda_+}\tilde S_{21} && \text{on } \Gamma_1 \quad (\lambda_+\in\Gamma_{1+}),\\
\tilde S_{22} &= -\frac{iq_-}{\lambda+\lambda_-}\tilde S_{21} && \text{on } \Gamma_2 \quad (\lambda_-\in\Gamma_{2-}),\\
\tilde S_{11} &= \frac{ir_-}{\lambda+\lambda_-}\tilde S_{12} && \text{on } \Gamma_3 \quad (\lambda_-\in\Gamma_{3+}),\\
\tilde S_{22} &= -\frac{ir_+}{\lambda+\lambda_+}\tilde S_{12} && \text{on } \Gamma_4 \quad (\lambda_+\in\Gamma_{4-}).
\end{aligned} \tag{5.2}$$

(iii) $\tilde S_{ij}(\lambda)$ has the following asymptotic form:

$$\begin{aligned}
\tilde S_{11} &= 1 + A/\lambda + O(\lambda^{-2}), && \lambda\to\infty, && \lambda\in R_1 + C_1,\\
\tilde S_{21} &= O(\lambda^{-1}), && \lambda\to\infty, && \lambda\in R_2 + C_1 + C_2,\\
\tilde S_{12} &= O(\lambda^{-1}), && \lambda\to\infty, && \lambda\in R_3 + C_3 + C_4,\\
\tilde S_{22} &= 1 - A/\lambda + O(\lambda^{-2}), && \lambda\to\infty, && \lambda\in R_4 + C_4,
\end{aligned} \tag{5.3}$$

where A is some complex number. Note that the asymptotic forms of $\tilde S_{21}$ and $\tilde S_{12}$ are redundant in the case C_d.

II. (i)$_\pm$ $F_\pm$ satisfies conditions $S_\pm$.

(ii) $I_+\cup I_- = \mathbb{R}$, i.e., at least one of the inverses $(I+\mathbb{F}^x_\pm)^{-1}$ exists in $L^2(\mathbb{F}_{\pm,x})$ for each $x\in\mathbb{R}$.

We solve the inverse problem in two steps.[13,14]

First, by solving the Marchenko equation for $\widetilde K_+$ [(3.36 +)] we obtain the potential $Q^{(+)}(x) = \widetilde Q_+(x) + Q_+$, $\widetilde Q_+(x) = -2\widetilde K^N_+(x,x)$, and the Jost solution $\Phi_+ = (I+\widetilde K_+)\widetilde\Phi^{(0)}_+$ for $x\in I_+$. This process constitutes the right inverse problem. Similarly the left inverse problem is to determine $Q^{(-)}(x) = \widetilde Q_-(x) + Q_-$, $\widetilde Q_-(x) = 2\widetilde K^N_-(x,x)$, and $\Phi_- = (I+\widetilde K_-)\widetilde\Phi^{(0)}_-$ by solving the Marchenko equation (3.36 −) for $x\in I_-$. In the second step we establish the conditions for S such that $Q^{(+)}(x) = Q^{(-)}(x) = Q(x)$ holds and Q reproduces the given S. The results for the first and second steps of the inverse problem have their own role and are given as Theorems 5.1 and 5.2, respectively.

Theorem 5.1: Let $S(\lambda)$ satisfy condition I with $u_\pm \neq 0$ and $F_\pm$ given by (3.30) satisfy condition II-(i)$_\pm$. Then the potential

$$Q^{(\pm)}(x) \equiv \mp 2\widetilde K^N_\pm(x,x) + Q_0(x), \quad x\in I_\pm$$

is determined uniquely by the Marchenko equation (3.36 ±) with the property

$$\left.\begin{aligned}
&\pm\int_x^{\pm\infty} dx\,(1+|y|)|Q^{(\pm)}(y) - Q_0(y)| < \infty,\\
&\pm\int_x^{\pm\infty} dx\,(1+|y|)|dQ^{(\pm)}(y)| < \infty,
\end{aligned}\right\} \quad x\in I^{(0)}_\pm, \tag{5.4}$$

and the expansion theorem (3.10) of $f(x)\in C^1_0$ (supp $f\in I_\pm$) holds in terms of the Jost solution $\Phi_\pm = (I+\widetilde K_\pm)\widetilde\Phi^{(0)}_\pm$ for $Q^{(\pm)}(x)$, $x\in I_\pm$.

Proof: Under the assumptions for S, the unique $\widetilde K_\pm$ exists by Lemma 4.3 and (4.13) for $x\in I_\pm$ and has the properties given in Lemma 4.4 for $x\in I_\pm$ and Lemma 4.5 for $x\in I^{(0)}_\pm$. From (4.17)–(4.19) we obtain (5.4). Put

$$\begin{aligned}
\Phi_\pm(x,\lambda,\lambda_\pm) &= \widetilde\Phi^{(0)}_\pm(x,\lambda,\lambda_\pm)\\
&\quad \pm\int_x^{\pm\infty} dy\,\widetilde K_\pm(x,y)\widetilde\Phi^{(0)}_\pm(y,\lambda,\lambda_\pm),\\
&\lambda_\pm\in\mathbb{R} \quad x\in I_\pm
\end{aligned} \tag{5.5}$$

then we see easily that all the properties of $F_{\pm\rho}$, $F_\pm$, and $\widetilde K_\pm$ required for Lemma 3.6 hold and hence, by the reverse of the reduction in the proof of Lemma 3.6, the Marchenko equation (3.36 ±) yields the expansion theorem (3.17) for f under the restriction supp $f\in I_\pm$. Next, we show that Φ_+ is the Jost solution for the potential $Q^{(+)}$ for $x\in I_+$. By Lemma 4.4 and (3.28)–(3.30) we easily derive the relations

$$\begin{aligned}
&(\partial/\partial x + \partial/\partial y)F^D_+(x,y)\\
&\quad = Q_0(x)F^N_+(x,y) + F^D_+(x,y)Q_0(y),\\
&(\partial/\partial x - \partial/\partial y)F^N_+(x,y)\\
&\quad = Q_0(x)F^D_+(x,y) - F^D_+(x,y)Q_0(y).
\end{aligned} \tag{5.6}$$

We decompose the Marchenko equation

$$\widetilde K_+(x,y) + F_+(x,y) + \int_x^\infty dz\,\widetilde K_+(x,z)F_+(z,y) = 0, \quad y>x$$

into diagonal and off-diagonal parts:

$$\left.\begin{aligned}
&\widetilde K^D_+ + F^D_+ + \int_+^\infty ds\,(\widetilde K^D_+F^D_+ + \widetilde K^N_+F^N_+) = 0,\\
&\widetilde K^N_+ + F^N_+ + \int_x^\infty dz\,(\widetilde K^N_+F^D_+ + \widetilde K^D_+F^N_+) = 0,
\end{aligned}\right\} \quad y>x. \tag{5.7}$$

Differentiating (5.7) with respect to x or y, using (5.6) to eliminate F^D_{+y}, F^N_{+y} in the integrands, and integrating by parts with respect to z, we obtain

$$H(x,y) + \int_x^\infty dz\,H(x,z)F_+(z,y) = 0, \quad y>x \tag{5.8}$$

with

$$\begin{aligned}
H &= H^D + H^N,\\
H^D(x,y) &= \widetilde K^D_{+x} + \widetilde K^D_{+y} - Q^{(+)}(x)\widetilde K^N_+(x,y)\\
&\quad - \widetilde K^N_+(x,y)Q_0(y),\\
H^N(x,y) &= \widetilde K^N_{+x} - \widetilde K^N_{+y} - Q^{(+)}(x)\widetilde K^D_+(x,y)\\
&\quad + \widetilde K^D_+(x,y)Q_0(y).
\end{aligned}$$

Since H^D, $H^N\in L_{1\cap 2}$ by Lemma 4.4, and $\phi = \mathbb{F}^x_+\phi$ has only a zero solution in $L_{1\cap 2}$ for $x\in I_+$, we have $H=0$ from (5.8), i.e.,

$$\begin{aligned}&\widetilde{K}^D_{+x}+\widetilde{K}^D_{+y}\\&\quad=Q^{(+)}(x)\widetilde{K}^N_+(x,y)+\widetilde{K}^N_+(x,y)Q_0(y),\\&\widetilde{K}^N_{+x}-\widetilde{K}^N_{+y}\\&\quad=Q^{(+)}(x)\widetilde{K}^D_+(x,y)-\widetilde{K}^D_+(x,y)Q_0(y),\end{aligned}\qquad y>x\in I_+.$$

Then, a simple calculation shows that Φ_+ [(5.5)] is the Jost solution for the potential $Q^{(+)}$. The proof is similar for Φ_- and $Q^{(-)}$. ■

Theorem 5.2: Consider the case C_s $(u^2_+-u^2_-\notin\mathbb{R})$ with $u_\pm\neq 0$. If $Q\in Q^\epsilon\cap Q^s$ and both $\widetilde{Q}_\pm\in CF'_\pm(1)$, then the matrix S satisfies conditions I and II-(i)$_+$ and II-(i)$_-$. Conversely, if S satisfies conditions I and II, then Q is determined uniquely for $x\in\mathbb{R}$ and S is the scattering data with $Q\in Q^\epsilon\cap Q^s$ and $\widetilde{Q}_\pm\in CF'_\pm(1)$.

Consider the case C_d $(u^2_+-u^2_-\in\mathbb{R})$ with $u_\pm\neq 0$. If $Q\in Q^s$ and both $\widetilde{Q}_\pm\in CF'_\pm(1)$, then S satisfies conditions I and II-(i)$_+$ and II-(i)$_-$. Conversely, if S satisfies the conditions I and II, then Q is determined uniquely for $x\in\mathbb{R}$ and S is the scattering data with $Q\in Q^s$, $\widetilde{Q}_\pm\in CF'_\pm(1)$.

The proof of the Theorem 5.2 will be given in the last subsection.

5.2. Connection of the left and right inverse problems

We introduce a sufficient condition (i)–(iii)$_\pm$ for the function $\psi_{\pm j}(x,\lambda,\lambda_\pm)$ $(j=1,2)$ to be, similar to the Jost solution $\phi_{\pm j}(x,\lambda,\lambda_\pm)$ $(j=1,2)$, expressible in terms of some integral kernel $\widetilde{K}_\pm$ and the free Jost solution $\tilde{\phi}^{(0)}_{\pm j}(x,\lambda,\lambda_\pm)$ $(j=1,2)$.

(i)$_+$ ψ_{+1} (ψ_{+2}) is analytic in $R_2\cup R_4$ $(R_1\cup R_3)$ corresponding to $\mathrm{Im}\,\lambda_+<0$ (>0) and is continuous on the boundary $C_1\cup C_4\cup\Gamma_1\cup\Gamma_4$.

(ii)$_+$ $\psi_{+2}=-iq_+/(\lambda+\lambda_+)\psi_{+1}$ on $\Gamma_1\cup\Gamma_4(\mathrm{Re}\,\lambda_+=0,\ \mathrm{Im}\,\lambda_+\geqslant 0)$, (5.9)

(iii)$_+$ $$\left.\begin{matrix}\{\psi_{+1}(x,\lambda,\lambda_+)-\phi^{(0)}_{+1}(x,\lambda,\lambda_+)\}e^{i\lambda_+x}\\ \{\psi_{+2}(x,\lambda,\lambda_+)-\phi^{(0)}_{+2}(x,\lambda,\lambda_+)\}e^{-i\lambda_+x}\end{matrix}\right\}=O(1/\lambda_+)\ \begin{matrix}\text{in } R_2\cup R_4\\ \text{in } R_1\cup R_3\end{matrix}\quad|\lambda_+|\to\infty. \tag{5.10}$$

(i)$_-$ ψ_{-1} (ψ_{-2}) is analytic in $R_1\cup R_2$ $(R_3\cup R_4)$ corresponding to $\mathrm{Im}\,\lambda_->0$ (<0) and is continuous on the boundary $C_2\cup C_3\cup\Gamma_2\cup\Gamma_3$.

(ii)$_-$ $\psi_{-2}=iq_-/(\lambda+\lambda_-)\psi_{-1}$ on $\Gamma_2\cup\Gamma_3$ $(\mathrm{Re}\,\lambda_-=0,\ \mathrm{Im}\,\lambda_-\geqslant 0)$, (5.11)

(iii)$_-$ $$\left.\begin{matrix}\{\psi_{-1}(x,\lambda,\lambda_-)-\phi^{(0)}_{-1}(x,\lambda,\lambda_-)\}e^{-i\lambda_-x}\\ \{\psi_{-2}(x,\lambda,\lambda_-)-\phi^{(0)}_{-2}(x,\lambda,\lambda_-)\}e^{i\lambda_-x}\end{matrix}\right\}=O(1/\lambda_-)\ \begin{matrix}\text{in } R_1\cup R_2\\ \text{in } R_3\cup R_4\end{matrix}\quad|\lambda_-|\to\infty. \tag{5.12}$$

Lemma 5.1: Put $u_\pm\neq 0$ and consider both cases C_s and C_d. Let a matrix function $\Psi_\pm(x,\lambda,\lambda_\pm)=(\psi_{\pm 1},\psi_{\pm 2})$, satisfying conditions (i)–(iii)$_\pm$ for $x\in\mathbb{R}$. Then there exists the unique integral kernel $K_\pm(x,y)$ such that

$$\Psi_\pm(x,\lambda,\lambda_\pm)=\Phi^{(0)}_\pm(x,\lambda,\lambda_\pm)\pm\int_x^{\pm\infty}dy\,K_\pm(x,y)\Phi^{(0)}_\pm(y,\lambda,\lambda_\pm),\quad \lambda_\pm\in\mathbb{R}, \tag{5.13}$$

$$K_\pm(x,y)=0,\quad y\leqslant x, \tag{5.14}$$

$$\pm\int_x^{\pm\infty}dy\,\langle K_\pm(x,y)\rangle^2<\infty. \tag{5.15}$$

The expression (5.13) for $\psi_{\pm 1}$ and $\psi_{\pm 2}$ is valid also for $\mathrm{Im}\,\lambda_\pm\leqslant 0$ and $\mathrm{Im}\,\lambda_\pm\geqslant 0$, respectively.

Proof: We introduce the formal representation of Ψ_{+11} and Ψ_{+12} as

$$\begin{aligned}\Psi_{+11}(x)-\Phi^{(0)}_{+11}(x)&=\int dy\left[K_{+11}+\frac{ir_+}{(\lambda+\lambda_+)}K_{+12}\right]\times e^{-i\lambda_+y},\quad \mathrm{Im}\,\lambda_+\leqslant 0,\\ \Psi_{+12}(x)-\Phi^{(0)}_{+12}(x)&=\int dy\left[\frac{-iq_+}{(\lambda+\lambda_+)}K_{+11}+K_{+12}\right]\times e^{i\lambda_+y},\quad \mathrm{Im}\,\lambda_+\geqslant 0.\end{aligned} \tag{5.16}$$

Replacing λ by $-\lambda$ and λ_+ by $-\lambda_+$ in the first equation and using the identity $1+q_+r_+/(\lambda+\lambda_+)^2=2\lambda/(\lambda+\lambda_+)$, we have

$$\begin{aligned}\int dy\,K_{+11}e^{i\lambda_+y}&=(\lambda+\lambda_+)/(2\lambda)\{(\Psi_{+11}-\Phi^{(0)}_{+11})(x,-\lambda,-\lambda_+)\\&\quad+ir_+/(\lambda+\lambda_+)(\Psi_{+12}-\Phi^{(0)}_{+12})(x,\lambda,\lambda_+)\},\\&\qquad\lambda\in R_1\cup R_3\quad(\mathrm{Im}\,\lambda_+\geqslant 0).\end{aligned} \tag{5.17}$$

By (i)$_+$ the right-hand side is analytic in $R_1\cup R_3$. By (ii)$_+$ we have on $\Gamma_1\cup\Gamma_4$

$$\begin{aligned}(\Psi_{+11}-\Phi^{(0)}_{+11})(x,-\lambda,-\lambda_+)&=(-\lambda+\lambda_+)/(-iq_+)(\Psi_{+12}-\Phi^{(0)}_{+12})(x,-\lambda,\lambda_+)\\&=-ir_+/(\lambda+\lambda_+)(\Psi_{+12}-\Phi^{(0)}_{+12})(x,-\lambda,\lambda_+).\end{aligned}$$

Hence we have

$$\begin{aligned}\int dy\,K_{+11}e^{i\lambda_+y}&=ir_+/(2\lambda)\{-(\Psi_{+12}-\Phi^{(0)}_{+12})(x,-\lambda,\lambda_+)\\&\quad+(\Psi_{+12}-\Phi^{(0)}_{+12})(x,\lambda,\lambda_+)\}\\&\quad\text{on }\Gamma_1\cup\Gamma_4\quad\mathrm{Im}\,\lambda_+\geqslant 0.\end{aligned} \tag{5.18}$$

Since the right-hand side of (5.18) is invariant under the replacement of λ by $-\lambda$ on $\Gamma_1\cup\Gamma_4$, the right-hand side of (5.17) is an analytic function of λ_+ for $\mathrm{Im}\,\lambda_+>0$ except the isolated point $\lambda_+=(-u^2_+)^{1/2}$, $\mathrm{Im}\,\lambda_+>0$ (i.e., $\lambda=0$). But the quantity in { } of (5.17) vanishes at $\lambda = 0$, which cancels the singularity λ^{-1}, and the right-hand side of (5.17) is analytic for $\operatorname{Im}\lambda_+ > 0$. By $(\mathrm{iii})_+$, (5.17) is square integrable along any line with $\operatorname{Im}\lambda_+ \geqslant 0$ parallel to the real axis. Hence, by the Paley–Wiener theorem,[15] the formal representation (5.17) yields the lemma for K_{+11}. The proof is similar for the other components of $K_\pm$. ■

Lemma 5.2: Under the same conditions as in the preceding lemma we have the unique integral kernel $\widetilde{K}_\pm(x,y)$ such that

$$\Psi_\pm(x,\lambda,\lambda_\pm) = \widetilde{\Phi}^{(0)}_\pm(x,\lambda,\lambda_\pm) \pm \int_x^{\pm\infty} dy\, \widetilde{K}_\pm(x,y)\widetilde{\Phi}^{(0)}_\pm(y,\lambda,\lambda_\pm), \quad \lambda_\pm \in \mathbb{R} \tag{5.19}$$

$$\widetilde{K}_\pm(x,y) = 0, \quad y \lessgtr x \tag{5.20}$$

$$\pm\int_x^{\pm\infty} dy (\widetilde{K}_\pm(x,y))^2 < \infty. \tag{5.21}$$

Expression (5.19) for $\psi_{\pm 1}$ and $\psi_{\pm 2}$ is also valid for $\operatorname{Im}\lambda_\pm \leqslant 0$ and $\operatorname{Im}\lambda_\pm \geqslant 0$, respectively.

Proof: Into (5.13) we insert the expression of $\Phi^{(0)}_\pm$, which is derived by inverting (2.13), to get (5.19) and, by means of the estimate of $K^{(-1)}_{\pm 0}$ and (5.14) and (5.15), we obtain (5.20) and (5.21). ■

We now formulate the alternative method for the left and right inverse problems, where the Jost solution is determined directly from the matrix S using the theorems of analytic functions, whereas in Theorem 5.1 the potential and the Jost solution were determined in terms of the solution of the Marchenko equation (3.36). For this purpose we introduce a set of functions $\{\psi_{\pm j}(x,\lambda,\lambda_\pm), \hat{\psi}_{\pm j}(x,\lambda,\lambda_\pm); j = 1,2\}$. The following conditions (iv)–$(\mathrm{vi})_\pm$, together with conditions (i)–$(\mathrm{iii})_\pm$ imposed on $\psi_{\pm j}$, determine the unique set $\{\psi_{\pm j}, \hat{\psi}_\pm; j = 1,2\}$. Then the Jost solution $\phi_{\pm j}$ corresponding to S of the preceding subsection will be given by $\psi_{\pm j}$ $(j = 1,2)$.

$(\mathrm{iv})_+$ $\hat{\psi}_{+1}$ $(\hat{\psi}_{+2})$ is analytic in $R'_1 R'_3$ $(R'_2 R'_4)$, where $R'_j \equiv R_j - \sigma_j$, with boundary values

$$\hat{\psi}_{+1} = \begin{cases} (\psi_{+1} + \tilde{S}_{21}/\tilde{S}_{11}\psi_{+2})/d_+ & \text{on } C_1, \\ (\psi_{+1} + \tilde{S}_{22}/\tilde{S}_{12}\psi_{+2})/d_+ & \text{on } C_4, \end{cases}$$
$$\hat{\psi}_{+2} = \begin{cases} -(\tilde{S}_{11}/\tilde{S}_{21}\psi_{+1} + \psi_{+2})/d_+ & \text{on } C_1, \\ -(\tilde{S}_{12}/\tilde{S}_{22}\psi_{+1} + \psi_{+2})/d_+ & \text{on } C_4. \end{cases} \tag{5.22}$$

The singularity of $\hat{\psi}_{+j}$ consists of simple poles having residues

$$\operatorname*{Res}_{\lambda_k\in\sigma_1} \hat{\psi}_{+1} = \gamma_k \psi_{+2}/\tilde{S}'_{11}, \qquad \operatorname*{Res}_{\lambda_k\in\sigma_3} \hat{\psi}_{+1} = \gamma_k \psi_{+2}/\tilde{S}'_{12},$$
$$\operatorname*{Res}_{\lambda_k\in\sigma_2} \hat{\psi}_{+2} = -\gamma_k \psi_{+1}/\tilde{S}'_{21}, \qquad \operatorname*{Res}_{\lambda_k\in\sigma_4} \hat{\psi}_{+2} = -\gamma_k \psi_{+1}/\tilde{S}'_{22}, \tag{5.23}$$

with arbitrary γ_k $(\neq 0) \in \mathbb{C}$ and the discontinuity across C_2 and C_3 from the lower to the upper side,

$$\begin{aligned} \delta\hat{\psi}_{+1} &= -d_-/(\tilde{S}_{11}\tilde{S}_{12})\psi_{+2} \quad \text{on } C_3 \quad (\lambda_- < 0), \\ \delta\hat{\psi}_{+2} &= -d_-/(\tilde{S}_{21}\tilde{S}_{22})\psi_{+1} \quad \text{on } C_2 \quad (\lambda_- > 0), \end{aligned} \tag{5.24}$$

$(\mathrm{v})_+$ $\hat{\psi}_{+2} = -iq_+/(\lambda + \lambda_+)\hat{\psi}_{+1}$

on $\Gamma_1 \cup \Gamma_4$ $(\operatorname{Re}\lambda_+ = 0,\ \operatorname{Im}\lambda_+ \geqslant 0)$, (5.25)

$$(\mathrm{vi})_+ \quad (\hat{\psi}_{+2} - \phi^{(0)}_{+1})e^{i\lambda_+ x} = \begin{cases} O(1/\lambda_+), & \\ O(1), & \end{cases} |\lambda_+| \to \infty \quad \begin{matrix} \text{in } R_4 \\ \text{in } R_2, \end{matrix}$$
$$(\hat{\psi}_{+1} - \phi^{(0)}_{+2})e^{-i\lambda_+ x} = \begin{cases} O(1/\lambda_+), & \\ O(1), & \end{cases} |\lambda_+| \to \infty \quad \begin{matrix} \text{in } R_1 \\ \text{in } R_3, \end{matrix} \tag{5.26}$$

$(\mathrm{iv})_-$ $\hat{\psi}_{-1}$ $(\hat{\psi}_{-2})$ is analytic in $R'_3 \cup R'_4$ $(R'_1 \cup R'_2)$ with boundary values

$$\hat{\psi}_{-1} = \begin{cases} (-\psi_{-1} + \tilde{S}_{21}/\tilde{S}_{22}\psi_{-2})/d_- & \text{on } C_2, \\ (-\psi_{-1} + \tilde{S}_{11}/\tilde{S}_{12}\psi_{-2})/d_- & \text{on } C_3, \end{cases}$$
$$\hat{\psi}_{-2} = \begin{cases} (-\tilde{S}_{22}/\tilde{S}_{21}\psi_{-1} + \psi_{-2})/d_- & \text{on } C_2, \\ (-\tilde{S}_{12}/\tilde{S}_{11}\psi_{-1} + \psi_{-2})/d_- & \text{on } C_3. \end{cases} \tag{5.27}$$

The singularity of $\hat{\psi}_{-j}$ consists of simple poles with residues

$$\operatorname*{Res}_{\lambda_k\in\sigma_1} \hat{\psi}_{-1} = \gamma_k^{-1}\psi_{-2}/\tilde{S}'_{12}, \qquad \operatorname*{Res}_{\lambda_k\in\sigma_4} \hat{\psi}_{-1} = -\gamma_k^{-1}\psi_{-2}/\tilde{S}'_{22},$$
$$\operatorname*{Res}_{\lambda_k\in\sigma_1} \hat{\psi}_{-2} = \gamma_k^{-1}\psi_{-1}/\tilde{S}'_{11}, \qquad \operatorname*{Res}_{\lambda_k\in\sigma_2} \hat{\psi}_{-2} = -\gamma_k^{-1}\psi_{-1}/\tilde{S}'_{21}, \tag{5.28}$$

and the discontinuity across C_1 and C_4,

$$\begin{aligned} \delta\hat{\psi}_{-1} &= d_+/(\tilde{S}_{12}\tilde{S}_{22})\psi_{-2} \quad \text{on } C_4 \quad (\lambda_+ < 0), \\ \delta\hat{\psi}_{-2} &= d_+/(\tilde{S}_{11}\tilde{S}_{21})\psi_{-1} \quad \text{on } C_1 \quad (\lambda_+ > 0), \end{aligned} \tag{5.29}$$

$(\mathrm{v})_-$ $\hat{\psi}_{-2} = -iq_-/(\lambda + \lambda_-)\hat{\psi}_{-1}$ on $\Gamma_2 \cup \Gamma_3$

$(\operatorname{Re}\lambda_- = 0,\ \operatorname{Im}\lambda_- \geqslant 0)$, (5.30)

$$(\mathrm{vi})_- \quad (\hat{\psi}_{-2} - \phi^{(0)}_{-1})e^{i\lambda_- x} = \begin{cases} O(1/\lambda_-), & \\ O(1), & \end{cases} |\lambda_-| \to \infty \quad \begin{matrix} \text{in } R_1 \\ \text{in } R_2, \end{matrix}$$
$$(\hat{\psi}_{-2} - \phi^{(0)}_{-2})e^{-i\lambda_- x} = \begin{cases} O(1/\lambda_-), & \\ O(1), & \end{cases} |\lambda_-| \to \infty \quad \begin{matrix} \text{in } R_4 \\ \text{in } R_3, \end{matrix} \tag{5.31}$$

Lemma 5.3: Let S satisfy conditions I and II-$(\mathrm{i})_+$ with $u_+ \neq 0$. The set of functions $\{\psi_{+j}, \hat{\psi}_{+j}; j = 1,2\}$ satisfying conditions (i)–$(\mathrm{vi})_+$ is determined uniquely for $x \in I_+$. $\widetilde{K}_+$ given by Lemma 5.2 is the unique solution of the Marchenko equation (3.36 +) and ψ_{+j} $(j = 1,2)$ is the Jost solution of the potential $Q^{(+)} = -2K^N_+(x,x) + Q_+$. The analogous statement holds for the set $\{\psi_{-j}, \hat{\psi}_{-j}; j = 1,2\}$, $\widetilde{K}_-$, and $Q^{(-)} = 2K^N_-(x,x) + Q_-$ for $x \in I_-$.

Proof: Assume the existence of a set of functions $\{\psi_{+j}, \hat{\psi}_{+j}; j = 1,2\}$ satisfying conditions (i)–$(\mathrm{vi})_+$ except possibly the boundary condition of (5.22). We use the set $\{\tilde{\phi}^{(0)}_{+j}, \hat{\tilde{\phi}}^{(0)}_{+j}; j = 1,2\}$, where $\tilde{\phi}^{(0)}_{+j}$ is the Jost solution for the step potential Q_0 and

$$\hat{\tilde{\phi}}^{(0)}_{+1} = \begin{cases} \tilde{\phi}^{(0)}_{-1}/\tilde{S}^{(0)}_{11} & \text{in } R_1 \\ \tilde{\phi}^{(0)}_{-2}/\tilde{S}^{(0)}_{12} & \text{in } R_3, \end{cases}$$

$$\hat{\tilde{\phi}}^{(0)}_{+2} = \begin{cases} -\tilde{\phi}^{(0)}_{-1}/\tilde{S}^{(0)}_{21} & \text{in } R_2 \\ -\tilde{\phi}^{(0)}_{-2}/\tilde{S}^{(0)}_{22} & \text{in } R_4. \end{cases}$$

Evidently the set $\{\tilde{\phi}^{(0)}_{+j}, \hat{\tilde{\phi}}^{(0)}_{+j}\}$ satisfies conditions (i)–$(\mathrm{vi})_+$ with $S^{(0)}$ in place of S. Construct a function

$$\Xi(x,y,\lambda)\equiv\begin{cases}\Delta\hat{\psi}_{+1}(x,\lambda,\lambda_+)\tilde{\phi}^{(0)A}_{+2}(y,\lambda,\lambda_+), & \lambda\in R'_1\cup R'_3,\\ \Delta\hat{\psi}_{+2}(x,\lambda,\lambda_+)\tilde{\phi}^{(0)A}_{+1}(y,\lambda,\lambda_+), & \lambda\in R'_2\cup R'_4,\end{cases}$$

with

$$\Delta\hat{\psi}_{+j}\equiv\hat{\psi}_{+j}-\hat{\tilde{\phi}}^{(0)}_{+j}\quad(j=1,2).$$

Then under conditions (i)–(vi)$_+$, Ξ is analytic in $R'_1\cup R'_2\cup R'_3\cup R'_4$ with the singularity consisting of simple poles with residues

$$\begin{aligned}&\operatorname*{Res}_{\lambda_k\in\sigma_1}\Xi=\gamma_k\Delta\psi_{+2}\tilde{\phi}^{(0)A}_{+2}/\tilde{S}'_{11}, \qquad \operatorname*{Res}_{\lambda_k\in\sigma_3}\Xi=\gamma_k\Delta\psi_{+2}\tilde{\phi}^{(0)A}_{+2}/\tilde{S}'_{12},\\ &\operatorname*{Res}_{\lambda_k\in\sigma_2}\Xi=-\gamma_k\Delta\psi_{+1}\tilde{\phi}^{(0)A}_{+1}/\tilde{S}'_{21}, \quad \operatorname*{Res}_{\lambda_k\in\sigma_4}\Xi=-\gamma_k\Delta\psi_{+1}\tilde{\phi}^{(0)A}_{+1}/\tilde{S}'_{22},\\ &\Delta\psi_{+j}=\psi_{+j}-\tilde{\phi}^{(0)}_{+j}\quad(j=1,2),\quad\lambda=\lambda_k\end{aligned}\tag{5.32}$$

and the discontinuity across C_j from the lower to the upper side,

$$\delta\Xi=\begin{cases}\Delta(\psi_{+1}+\tilde{S}_{21}/\tilde{S}_{11}\psi_{+2})\tilde{\phi}^{(0)A}_{+2}/d_+ +\Delta(\tilde{S}_{11}/\tilde{S}_{21}\psi_{+1}+\psi_{+2})\tilde{\phi}^{(0)A}_{+1}/d_+ & \text{across } C_1\\ -d_-\Delta(\psi_{+1}\tilde{\phi}^{(0)A}_{+1}/(\tilde{S}_{21}\tilde{S}_{22})) & \text{across } C_2\\ -d_-\Delta(\psi_{+2}\tilde{\phi}^{(0)A}_{+2}/(\tilde{S}_{12}\tilde{S}_{11})) & \text{across } C_3\\ \Delta(\psi_{+1}+\tilde{S}_{22}/\tilde{S}_{12}\psi_{+2})\tilde{\phi}^{(0)A}_{+2}/d_+ +\Delta(\tilde{S}_{12}/\tilde{S}_{22}\psi_{+1}+\psi_{+2})\tilde{\phi}^{(0)A}_{+1}/d_+ & \text{across } C_4\end{cases}\tag{5.33}$$

where, for instance, $\Delta(\psi_{+1}+\tilde{S}_{21}/\tilde{S}_{11}\psi_{+2})=\psi_{+1}-\tilde{\phi}^{(0)}_{+1}+\tilde{S}_{21}/\tilde{S}_{11}\psi_{+2}-\tilde{S}^{(0)}_{21}/\tilde{S}^{(0)}_{11}\tilde{\phi}^{(0)}_{+2}$. Note that Ξ is regular on $\Gamma_1\cup\Gamma_4$ by (v)$_+$. In the Cauchy theorem with a contour C not containing the singularity of Ξ,

$$\oint_C d\lambda\,\Xi(x,y,\lambda)=0,\quad y>x,\tag{5.34}$$

we decompose $C=C_\rho+K_\rho$ into C_ρ and K_ρ [the path excluding the singularity C_j and σ_j $(1\leqslant j\leqslant4)$ contained in C_ρ]. Then the contribution from C_ρ vanishes as $\rho\to\infty$ by (vi)$_+$ and (5.34) leads to the relation among the singularities (5.32) and (5.33) of Ξ on $K\equiv K_\infty$. Expressing ψ_{+j} in terms of $\tilde{K}_+$ and $\tilde{\phi}^{(0)}_+$ by Lemma 5.2 and using (3.28) and (3.30) we obtain the Marchenko equation (3.36 +) which, for $x\in I_+$, has the unique solution $\tilde{K}_+$. Thus, (i)–(vi)$_+$ assure the existence of the unique set $\tilde{K}_+$ and ψ_{+j}. Using the Cauchy formula, we express Ξ $(\lambda\in R'_1\cup R'_2\cup R'_3\cup R'_4)$ in terms of the boundary value on K, or the singularity (5.32) and (5.33), which is composed of ψ_{+j}, $\tilde{\phi}^{(0)}_{+j}$, and S:

$$\begin{aligned}\Xi(x,y,\lambda)&=\frac{1}{2\pi i}\int_K d\lambda'\,\frac{\Xi(x,y,\lambda')}{\lambda-\lambda'}\\ &=\sum_j\frac{1}{2\pi i}\left(\int_{C_j}d\lambda'\,\frac{\delta\Xi}{\lambda-\lambda'}+\sum_{\sigma_j}\frac{\operatorname{Res}\Xi}{\lambda-\lambda'}\right),\\ &\qquad\lambda\in R'_1\cup R'_2\cup R'_3\cup R'_4.\end{aligned}\tag{5.35}$$

The existence of Ξ is verified if the integral on $\hat{C}_{j\mu}=C_j-C_{j\mu}$ $(1\leqslant j\leqslant4)$ for some large μ converges. We examine, for instance,

$$\begin{aligned}\Xi_{1,2}&\equiv\frac{1}{2\pi i}\int_{\hat{C}_{1\mu}+\hat{C}_{2\mu}}d\lambda'\,\frac{\delta\Xi}{\lambda-\lambda'}\\ &=\frac{1}{2\pi i}\left\{\int_{\hat{C}_{1\mu}}\frac{d\lambda'}{d_+}\frac{1}{\lambda-\lambda'}\left[\Delta\left(\psi_{+1}\frac{\tilde{S}_{21}}{\tilde{S}_{11}}\psi_{+2}\right)\tilde{\phi}^{(0)A}_{+2}\right.\right.\\ &\qquad\left.+\Delta\left(\frac{\tilde{S}_{11}}{\tilde{S}_{21}}\psi_{+1}+\psi_{+2}\right)\tilde{\phi}^{(0)A}_{+1}\right]\\ &\qquad\left.-\int_{\hat{C}_{2\mu}}d\lambda'\,\frac{d_-}{\lambda-\lambda'}\Delta\left(\frac{\psi_{+1}\tilde{\phi}^{(0)A}_{+1}}{\tilde{S}_{21}\tilde{S}_{22}}\right)\right\}.\end{aligned}\tag{5.36}$$

By the asymptotic forms

$$\begin{aligned}&\tilde{S}_{11},\tilde{S}_{22},\tilde{S}^{(0)}_{11},\tilde{S}^{(0)}_{22},d_\pm=1+O(1/\lambda),\\ &\tilde{S}_{21},\tilde{S}^{(0)}_{21}=O(1/\lambda),\end{aligned}\qquad|\lambda|\to\infty,$$

and by (iii)$_+$ the singularity of Ξ on both $\hat{C}_{1\mu}$ and $\hat{C}_{2\mu}$ is reducible to

$$\{\pm\}\Delta(\psi_{+1}\tilde{\phi}^{(0)A}_{+1}/\tilde{S}_{21})+O(1/\lambda')\quad\text{on}\begin{Bmatrix}\hat{C}_{1\mu}\\ \hat{C}_{2\mu}\end{Bmatrix},\quad|\lambda'|\to\infty\tag{5.37}$$

$$\Delta(\psi_{+1}\tilde{\phi}^{(0)A}_{+1}/\tilde{S}_{21})=O(1),\quad|\lambda'|\to\infty.\tag{5.38}$$

Since the left-hand side of (5.38) is analytic for $|\lambda'|>\mu$ in R_2,

$$\frac{1}{2\pi i}\int_{\hat{C}_{1\mu}+\hat{C}_{2\mu}}\frac{d\lambda'}{\lambda-\lambda'}\Delta(\psi_{+1}\tilde{\phi}^{(0)A}_{+1}/\tilde{S}_{21})=J_B+O(1/\lambda),$$

$$J_B = \frac{1}{2\pi i}\oint_B \frac{d\lambda'}{\lambda-\lambda'}\Delta\,(\psi_{+1}\tilde{\phi}^{(0)A}_{+1}/\tilde{S}_{21}) \tag{5.39}$$

$$= \begin{cases} \Delta\,(\psi_{+1}\tilde{\phi}^{(0)A}_{+1}/\tilde{S}_{21})_{\lambda'=\lambda} = O(1) & \text{as } |\lambda|\to\infty \\ & \text{if } \lambda \text{ is contained in } B, \\ 0 & \text{if } \lambda \text{ is outside of } B, \end{cases}$$

where $B = \hat{C}_{1\mu} + \hat{C}_{2\mu} + C'_\mu$ and C'_μ is the part of C_μ contained in R_2 and the integral on C'_μ gives $O(1/\mu\lambda)$ by (5.38). The contribution to $\Xi_{1,2}$ of the second term of (5.37) $[O(1/\lambda')]$ is easily estimated to give $O(1/\lambda)$. Hence we have

$$\Xi_{1,2} = \begin{cases} O(1) \\ O(1/\lambda) \end{cases} \text{as } |\lambda|\to\infty, \quad \begin{matrix} \lambda\in R_2, \\ \lambda\in R_1\cup R_4. \end{matrix}$$

This proves the existence of $\Xi_{1,2}$ and also its asymptotic estimate. The similar calculation is also performed on C_3, C_4 and we have the existence of Ξ. The asymptotic property of Ξ shown above implies (vi)$_+$ for $\hat{\psi}_j$. It remains to show the boundary value (5.22) of $\hat{\psi}_{+j}$. From (5.33) we have

$$\begin{aligned} &\Delta\hat{\psi}^{(+)}_{+1}(x)\tilde{\phi}^{(0)A}_{+2}(y) - \hat{\psi}^{(-)}_{+1}(x)\tilde{\phi}^{(0)A}_{+1}(y) \\ &\quad = \{\Delta\,(\psi_{+1} + \tilde{S}_{21}/\tilde{S}_{11}\psi_{+2})\tilde{\phi}^{(0)A}_{+2} \\ &\qquad + \Delta\,(\tilde{S}_{11}/\tilde{S}_{21}\psi_{+1} + \psi_{+2})\tilde{\phi}^{(0)A}_{+1}\}/d_+, \end{aligned} \tag{5.40}$$

where $\hat{\psi}^{(+)}_{+1}$ $(\hat{\psi}^{(-)}_{+1})$ denotes the boundary value of $\hat{\psi}_{+1}$ on the upper (lower) side of C_1. Since $\tilde{\phi}^{(0)A}_{+2}(y)$ and $\phi^{(0)A}_{+1}(y)$ depend on y only through $e^{i\lambda,y}$ and $e^{-i\lambda,y}$, respectively, for $y > l$ (l: the location of the step of Q_0), (5.40) leads to

$$\Delta\hat{\psi}^{(+)}_{+1} = \Delta\,(\psi_{+1} + \tilde{S}_{21}/\tilde{S}_{11}\psi_{+2})/d_+,$$

$$\Delta\hat{\psi}^{(-)}_{+1} = \Delta\,(\tilde{S}_{11}/\tilde{S}_{21}\psi_{+1} + \psi_{+2})/d_+,$$

which yield (5.22) on C_1 and similarly we have (5.22) on C_4. In this way the unique existence of $\{\psi_{+j}, \hat{\psi}_{+j}\}$ is shown and the right inverse problem is solved. ■

Thus by an elementary calculation we obtain the lemma connecting the left and right inverse problems.

Lemma 5.4: Put

$$\psi_{-1} = \begin{cases} \tilde{S}_{11}\hat{\psi}_{+2} & \text{in } R_1, \\ -\tilde{S}_{21}\hat{\psi}_{+2} & \text{in } R_2, \end{cases} \qquad \psi_{-2} = \begin{cases} \tilde{S}_{12}\tilde{\psi}_{+1} & \text{in } R_3, \\ -\tilde{S}_{22}\tilde{\psi}_{+2} & \text{in } R_4, \end{cases}$$

$$\hat{\psi}_{-1} = \begin{cases} \psi_{+2}/\tilde{S}_{12} & \text{in } R_3, \\ -\psi_{+1}/\tilde{S}_{22} & \text{in } R_4, \end{cases} \qquad \hat{\psi}_{-2} = \begin{cases} \psi_{+2}/\tilde{S}_{11} & \text{in } R_1, \\ -\psi_{+1}/\tilde{S}_{21} & \text{in } R_2, \end{cases} \tag{5.41}$$

then condition (i)–(vi)$_+$ for $\{\psi_{+j}, \hat{\psi}_{\pm j};\ j = 1, 2\}$ is equivalent to condition (i)–(vi)$_-$ for $\{\psi_{-j}, \hat{\psi}_{-j}; j = 1, 2\}$.

5.3 Proof of the Theorem 5.2

We prove Theorem 5.2 for the case C_s. In the direct problem let $\tilde{Q}_\pm \subset Q^\epsilon\cap Q^s\cap CF'_\pm(0)$, then S satisfies condition I by Lemmas I-2–7. If $\tilde{Q}\in Q^\epsilon\cap Q^s\cap CF'_\pm(1)$, then the kernel $F_\pm$ of the Marchenko equation satisfies, by Lemma 3.9, $S_\pm$, i.e., II-(i)$_\pm$. In the inverse problem let S satisfy conditions I and II and put $x\in I_0\equiv I_+\cap I_-$. Note that $I_\pm$ and accordingly I_0 $(\neq 0)$ are open intervals. By Lemma 4.3, the Marchenko equation with kernel $F_\pm$ has the unique solution $\tilde{K}_\pm$ giving $\Phi_\pm$ in the form (2.11) and the potential $\tilde{Q}_\pm(x) = \mp 2\tilde{K}^N(x, x)$ [I-(2.26)] for $x\in I_\pm$. On the other hand, since S satisfies conditions I and II-(i)$_+$, the unique set $\{\psi_{+j}, \hat{\psi}_{+j}; j = 1, 2\}$ satisfying the condition (i)–(vi)$_+$ exists and $\psi_{+j} = \phi_{+j}$ $(j = 1, 2)$ by Lemma 5.3 and Theorem 5.1. By Lemma 5.4 $\{\psi_{+j}, \hat{\psi}_{+j}; j = 1, 2\}$ determines the unique set $\{\psi_{-j}, \hat{\psi}_{-j}; j = 1, 2\}$ satisfying the condition (i)–(vi)$_-$. Since S satisfies conditions I and II-(i)$_-$, we get $\psi_{-j} = \phi_{-j}$ $(j = 1, 2)$ by Lemma 5.3 and Theorem 5.1. Hence, by (5.22) and (5.41) we have

$$\phi_{-1} = (1/d_+)(\tilde{S}_{11}\phi_{+1} + \tilde{S}_{21}\phi_{+2}) \quad \text{on } C_1,$$

$$\phi_{-2} = (1/d_+)(\tilde{S}_{12}\phi_{+1} + \tilde{S}_{22}\phi_{+2}) \quad \text{on } C_4.$$

ϕ_{-1} and ϕ_{-2} satisfy the same differential equation for $\lambda\in C_1$ and $\lambda\in C_4$, respectively, for $x\in I_0$. This means that the left and right inverse problems give the common potential $Q(x) = Q^{(\pm)}(x) \mp 2K^N_\pm(x, x) + Q_\pm$ for $x\in I_0$. Thus we can define a potential $Q(x)$ for $x\in(I_+\cup I_-) = \mathbb{R}$ by

$$Q(x) \equiv Q^{(\pm)}(x), \quad x\in I_\pm.$$

Since $F_\pm$ satisfies condition $S_\pm$ we see $\tilde{Q}_\pm \in CF_\pm(1) \cap CF'_\pm(1)$ by Lemma 4.5. Obviously, $\tilde{Q}_\pm \in Q^\epsilon\cap Q^s$ by condition I-(ii), I-(iii). Thus Theorem 5.2 for C_s is proved. The proof for C_d is analogous and is omitted here. ■

5.4 Correspondence between the potential and the scattering data

We introduce the function class $GCF_\pm(1)$ in such a sense that $\rho(s, t)\in GCF_\pm(1)$ means

$$\int_x^{+\infty} ds \int_x^{\pm\infty} dt\,|\rho(s, t)| < \infty, \quad x\in\mathbb{R}.$$

If $\rho(x)\in CF_\pm(1)$, then

$$\hat{\rho}_\pm(x) = \pm\int_x^{\pm\infty} dy\,|\rho(y)| < \infty,$$

$$\int_x^{\pm\infty} ds \int_x^{\pm\infty} dt\,|\rho(s+t)| = \pm\int_x^{\pm\infty} ds\,\hat{\rho}(2s) < \infty;$$

hence $\rho(s+t)\in GCF_\pm(1)$. This shows that $GCF_\pm(1)$ defined for the function of two variables is a natural generalization of $CF_\pm(1)$ defined for the function of a single variable. In an analogous way $GCF_\pm(n)$ $(n\geqslant 1)$ is defined as the class of functions of two variables; i.e., if $\rho(s, t)\in GCF_\pm(n)$, then

$$\int_x^{\pm\infty} ds \int_x^{\pm\infty} dt\,(1 + |t|^{n-1})|\rho(s, t)| < \infty, \quad x\in\mathbb{R}.$$

Evidently $GCF_\pm(n)\subset GCF_\pm(n')$ if $n < n'$. We note that if $\rho(s, t)\in GCF_+(n)$ $[GCF_-(n)]$ and $|\rho(s, t)|$ is nonincreasing [nondecreasing] with respect to s, then $\rho(t, t)\in CF_+(n)$ $[CF_-(n)]$ $(n\geqslant 1)$.

We deduce easily the following theorem from the estimates in Lemmas 3.7 and 4.5.

Theorem 5.3: In Theorems 5.1 and 5.2 the relations

$$\langle F^N_\pm(\xi, \eta)\rangle \leqslant f^{(0)}_\pm(\xi+\eta) + f'_\pm(2\zeta, \xi+\eta)\in GCF_\pm(m), \quad m\geqslant 1, \tag{5.42}$$

$$\langle F^N_{\pm A}(\xi, \eta)\rangle \leqslant g^{(0)}_\pm(\xi, \eta) + g'_\pm(2\zeta, \xi+\eta)\in GCF_\pm(n), \quad n\geqslant 1, \tag{5.43}$$

yield

$$\langle\tilde{Q}_\pm(x)\rangle \begin{cases} \in CF_\pm(m)\cap CF'_\pm(n) & \text{if } m\geqslant n\geqslant 1, \\ \in CF_\pm(m)\cap CF'_\pm(m) & \text{if } m = n-1\geqslant 1, \end{cases}$$

and conversely $\langle\tilde{Q}_\pm\rangle\in CF_\pm(m)\cap CF'_\pm(n)$ $(m\geqslant n\geqslant 1)$ implies

(5.42) and (5.43) for $m = n\ (\geqslant 1)$.

We note that due to the relation

$$|\tilde{Q}_\pm(x)| \leqslant \pm \int_x^{\pm\infty} |dQ_\pm(y)|,$$

$$|F_\pm(x,y)| \leqslant \pm \int_y^{\pm\infty} d\eta |F_{\pm\eta}(x,\eta)|, \quad y \gtrless x$$

$I_\pm \in CF_\pm(m)$ follows from $J_\pm \in CF_\pm(m+1)$ and (5.42) follows from (5.43) with $n = m+1$ and the cases with $m < n-1$ can be neglected in Theorem 5.3.

6. SUMMARY FOR THE DIRECT AND THE INVERSE PROBLEMS

Our study is performed under the assumption of the potential $Q \in Q^s$, which means that the point spectrum is separated from the continuous spectrum and composed of only a finite number of simple eigenvalues. In the case C_s we add the condition to the potential $Q \in Q^\epsilon$ [(1.3)]. This condition is satisfied if the perturbation from the step potential, $\tilde{Q}(x) = Q(x) - Q_0(x)$ $(q_+ \neq q_-, r_+ \neq r_-)$, is a continuous function, since then

$$\left|\sum \delta Q_j e^{i\xi x_j}\right| = |Q_+ - Q_-| \geqslant \delta \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

with

$$\delta = \min\{|q_+ - q_-|,\ |r_+ - r_-|\} > 0.$$

Thus the condition $Q \in Q^\epsilon$ is quite natural in the present study, which starts from the step potential Q_0.

The expansion theorems of cases C_s and C_d $(q_\pm r_\pm \neq 0)$ were considered in I for the potentials

$$\tilde{Q}_\pm \in CF_\pm(1) \cap CF'_\pm(0) \quad \text{(Theorems I-1, I-2)},$$

whereas the Marchenko equation is derived for $\tilde{Q}_\pm \in CF_\pm(1) \cap CF'_\pm(1)$ (Lemma 3.6). The inverse problem is discussed for the scattering data satisfying $S_\pm$ of Sec. 4.1, which corresponds to the potential

$$\tilde{Q}_\pm \in CF_\pm(1) \cap CF'_\pm(1) \quad \text{(Theorems 5.1 and 5.2)}.$$

Hence we obtain in our study I and II the mapping of the potential to the scattering data

$$\tilde{Q}_\pm \in CF_\pm(1) \cap CF'_\pm(1) \longleftrightarrow F^N_\pm,\ F^N_{\pm A} \in GCF_\pm(1). \tag{6.1}$$

Here two relations on the right-hand side mean the convenient expression of conditions $S_\pm$. Further, we obtain in Theorem 5.3 a more general correspondence

$$\tilde{Q}_\pm \in CF_\pm(m) \cap CF'_\pm(n) \longleftrightarrow \begin{array}{l} F^N_\pm \in GCF_\pm(m), \\ F^N_{\pm A} \in GCF_\pm(n), \end{array} \quad \text{for } m \geqslant n \geqslant 1 \tag{6.2}$$

with the convenient expression of (5.42) and (5.43) on the right-hand side.

The inverse problem for the potential Q with vanishing asymptotic values $Q_\pm = 0$ was studied by the present authors[7] for $Q \in Q^s$ with the resulting correspondence

$$Q \in CF'_\pm(1) \longleftrightarrow F^N_\pm = F_\pm \in CF'_\pm(1), \tag{6.3}$$

or more generally (1.2), that is,

$$Q \in CF_\pm(m) \cap CF'_\pm(n) \longleftrightarrow F^N_\pm = F_\pm \in CF_\pm(m) \cap CF'_\pm(n), \quad m \geqslant 0,\ n \geqslant 1.$$

From (6.1)–(6.4) we note that the inverse problem is manageable in terms of the Marchenko equation for $Q \in CF_\pm(0)$ when $Q_\pm = 0$ under some additional conditions, whereas it seems to require the condition $Q \in CF_\pm(1)$ when $q_\pm r_\pm \neq 0$.

There are λ-dependent potentials considered in the inverse scattering method.[16,17] The direct and inverse spectral problem of ZS operator with a potential polynomial in λ,

$$Q(x;\lambda) = \sum_{n=0}^{N-1} \lambda^{n/N} Q_n(x),$$

$$Q_n(s) \to Q_{n\pm}\ (\neq 0) \quad (x \to \pm\infty),$$

may be treated by the method similar to the present series I and II.

APPENDIX: SPECTRUM FOR THE STEP POTENTIAL

The step potential in this paper corresponds to the constant potential to define an unperturbed free state in the usual scattering problems in quantum mechanics. Let $Q(x)$ have a step at $x = l$ and be given by $Q(x) = Q_0(x;l) = Q_\pm$ $(\neq 0)$ $(x \gtrless l)$. The corresponding operator L is denoted as $L^{(0)} = i\sigma_3[d/dx - Q_0(x;l)]$ with the Jost solutions $\tilde{\Phi}^{(0)}_\pm = \Phi^{(0)}_\pm(x)$ $(x \gtrless l)$. The matrix $S^{(0)}$ is explicitly given by

$$S^{(0)} = \begin{pmatrix} \left\{1 - \dfrac{q_+ r_-}{(\lambda+\lambda_+)(\lambda+\lambda_-)}\right\} e^{i(\lambda_+ - \lambda_-)l}, & \left\{\dfrac{iq_+}{\lambda+\lambda_+} - \dfrac{iq_-}{\lambda+\lambda_-}\right\} e^{i(\lambda_+ + \lambda_-)l} \\ \left\{\dfrac{ir_+}{\lambda+\lambda_+} - \dfrac{ir_-}{\lambda+\lambda_-}\right\} e^{-i(\lambda_+ + \lambda_-)l}, & \left\{1 - \dfrac{q_- r_+}{(\lambda+\lambda_+)(\lambda+\lambda_-)}\right\} e^{-i(\lambda_+ - \lambda_-)l} \end{pmatrix} \tag{A1}$$

and analytic on our λ plane R_1 for $u_\pm \neq 0$. The spectrum $\sigma(L^{(0)})$ consists of the continuous part σ_c and the discrete part σ_p. As shown in I, σ_c is presented by the curves $C_1 - C_4$ in the λ plane while σ_p is determined by the zeros of $\tilde{S}^{(0)}_{ij}(\lambda)$. More specifically, σ_p is given by a set of zeros of $\tilde{S}^{(0)}_{11}$, $\tilde{S}^{(0)}_{21}$, $\tilde{S}^{(0)}_{12}$, and $\tilde{S}^{(0)}_{22}$ in the domains R_1, R_2, R_3, and R_4, respectively.

The conditions for $\tilde{S}^{(0)}_{11}$ $(\tilde{S}^{(0)}_{22})$ to have zeros in R_1 (R_4) and $\tilde{S}^{(0)}_{21}$ $(\tilde{S}^{(0)}_{12})$ in R_2 (R_3) are given by the equation

$$\pm \lambda_+ \lambda_- = \tfrac{1}{2}(q_+ r_- + q_- r_+) - \lambda^2 \tag{A2}$$

with the upper and the lower signs of the left-hand side, respectively, where λ gives the zero point for all cases and is

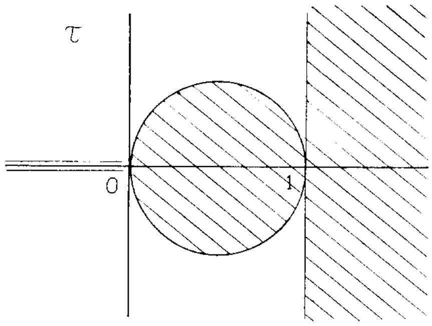


FIG. 3. Distribution of zeros for the case $q = \gamma \bar{r}$.

expressible by

$$\lambda^2 = \frac{1}{4} \frac{(q_+ r_- - q_- r_+)^2}{(q_+ - q_-)(r_- - r_+)} \tag{A3}$$

for $q_+ \neq q_-$ and $r_+ \neq r_-$. There is no zero point for $q_+ = q_-$ or $r_+ = r_-$ while two zeros $\notin \sigma_p$ at the branch points $\lambda_\pm = 0$ for $q_+ = q_-$ and $r_+ = r_-$. From (A2) and (A3), it is clear that there are possibly three cases; (1) nondegenerate two zeros symmetric with respect to the origin for $q_+ r_- - q_- r_+ \neq 0$, or (2) doubly degenerate zero at the origin $\lambda = 0$ for $q_+ r_- - q_- r_+ = 0$, or (3) no zero in the prescribed regions. We consider two simple examples.

1. The case $q(x) = \gamma \bar{r}(x)$ with $\gamma \in \mathbb{C}$ $(\gamma \neq 0)$

The classification of the three cases (1)–(3) above is conveniently made in terms of $\tau = r_+/r_- = \bar{q}_+/\bar{q}_-$. The domains of the τ plane where $S^{(0)}$ has two simple zeros $\pm \lambda$ are given by $\cos\theta < |\tau| < 1/\cos\theta$ ($|\theta| < \pi/2$) and $\pi/2 \leqslant |\theta| < \pi$ as shown in Fig. 3; here $\theta = \arg \tau$, while the domain of the τ plane with doubly degenerate zero is $\theta = \pi$, i.e., negative real axis. The domains in the τ plane without any zero in the λ plane are given by $\mathrm{Re}\,\tau > 1$ and $|\tau - \frac{1}{2}| < \frac{1}{2}$. It is noted that in the τ plane the classification does not depend on γ.

2. The case $q(x) = \gamma r(x)$ with $\gamma \in \mathbb{C}$ $(\gamma \neq 0)$

Since the right-hand side of (A3) identically vanishes, only a double zero $\lambda^2 = 0$ is possible in this case. The condition (A2) reduces to $\pm(\gamma^2 \tau^2)^{1/2} = \gamma \tau$ with the upper sign corresponding to zero of $\tilde{S}^{(0)}_{11}$ and $\tilde{S}^{(0)}_{22}$ and the lower sign to that of $\tilde{S}^{(0)}_{21}$ and $\tilde{S}^{(0)}_{12}$. Obviously one of these two cases is possible.

[1] V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys. JETP **34**, 62 (1972)].
[2] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[3] T. Kawata and H. Inoue, J. Phys. Soc. Jpn. **43**, 361 (1977); **44**, 1722 (1978).
[4] V. S. Gerdjikov and P. P. Kulish, Burg. J. Phys. **5**, 337 (1978).
[5] N. Asano and Y. Kato, J. Math. Phys. **22**, 2780 (1981).
[6] V. A. Marchenko, Dokl. Akad. Nauk SSSR **104**, 695 (1955).
[7] N. Asano and Y. Kato, in *Proceedings of International Conference on Plasma Physics* (Fusion Research Assoc. of Japan, Nagoya, 1980), Vol. I, p. 87.
[8] Y. Kato, *The Inverse Problem in Scattering Theory* (Iwanami, Tokyo, 1983).
[9] W. Eckhaus and A. van Harten, *The Inverse Scattering Transformation and the Theory of Solitons* (North-Holland, Amsterdam, 1981).
[10] Some numerical errors in factors on Q_M in I-(2.23) are corrected in (2.15).
[11] This is the convenient expression of $Q \in Q^c \cap Q^s$ and $\tilde{Q}_\pm \in CF'_\pm$ (1).
[12] L. Schwarz, *Théorie des distributions* (Hermann, Paris, 1966).
[13] Z. S. Agranovich and V. A. Marchenko, *The Inverse Problem of Scattering Theory* (Izd. Kharkov Univ., 1960), English translation by B. D. Seckler (Gordon and Breach, New York, 1963).
[14] L. D. Faddeev, Tr. Mat. Inst. Steklov **73**, 314 (1964).
[15] K. Yoshida, *Functional Analysis* (Springer, Berlin, 1978), 5th ed.
[16] D. J. Kaup and A. C. Newell, J. Math. Phys. **19**, 798 (1978).
[17] M. O. Guillaume and M. Jaulent, Lett. Math. Phys. **6**, 189 (1982).

On some family of congruences of null strings

Ivor Robinson and Krzysztof Rózga[a)]
Program in Mathematical Sciences, The University of Texas at Dallas, Richardson, Texas 75080

(Received 9 November 1982; accepted for publication 4 March 1983)

A definition of a special family of congruences of null strings is given, and a link between its existence and "projective extensions of heavens" is provided. Next the integrability conditions of the problem are studied, and a subclass of all "heavens" admitting these special congruences is described in terms of one function of three variables subject to a nonlinear, differential second-order constraint.

PACS numbers: 04.20.Cv, 02.40. + m, 04.20.Jb

1. INTRODUCTION

The concept of a congruence of null strings (two-dimensional totally null surfaces) turned out to be a key one in the studies of algebraically degenerate complex Einstein space-times.[1,2] The existence of a congruence of null strings is, however, something exceptional. Indeed, if such a congruence does exist then the spinor field k^A related to it (a dotted spinor field if the congruence is anti-self-dual[3]) turns out to be a Debever–Penrose spinor direction of the left conformal curvature C_{ABCD}. Therefore, the existence of five distinct, self-dual congruences already implies that $C_{ABCD} = 0$. An assumption that the Einstein equations are satisfied makes the situation even simpler.[1]

The spaces with $C_{ABCD} = 0$ became of independent interest due to Penrose[4]. Those among them which are Einstein space-times have been investigated by many authors from different points of view.[4–6] If, therefore, that specific subclass of left-flat spaces ("heavens") is expected to be related to a self-dual nonlinear graviton, more general complex space-times, restricted only by $C_{ABCD} = 0$ could correspond to a nonlinear graviton interacting with matter.

The condition $C_{ABCD} = 0$ turns out to be necessary and sufficient for the existence of a complete set of null self-dual strings; twistor surfaces in another terminology (see for example Ref. 4). Then one can construct congruences and families of congruences. Of special interest are one-parameter families.[4] In "heavens" there is a *canonical family*, the congruences of which are defined by covariantly constant undotted spinor fields. In a general, left conformally flat space (weak "heaven") there are it seems no distinguished congruences, with some exceptions. In this paper we discuss such an exceptional case.

In Sec. 1 a special family of congruences of null strings called a *generalized canonical family* is defined, and then it is shown, that the family can exist only in "projective extensions of heavens"—the spaces investigated some time ago by Finley and Plebański in Ref. 7.

In Sec. 3 the integrability conditions (and therefore necessary and sufficient conditions for the existence of generalized families) are established in a covariant form. In the last section a full characterization of "heavens" admitting those special families is given.

[a)] On leave of absence from Technical University of Kielce, Poland.

The formalism used throughout this paper has been developed by Plebański. For a summary we refer the reader to Ref. 8 and also to Ref. 3, where the definitions of basic concepts can be found.

Below some basic identities and facts are listed. Let (M, ds^2) be a complex space-time. Then the null-tetrad one-forms $\{e^a\}$, $a = 1, 2, 3, 4$, or $\{g^{A\dot{B}}\}$ are defined by

$$ds^2 = 2(e^1 \otimes e^2 + e^3 \otimes e^4) = -\tfrac{1}{2} g_{A\dot{B}} \underset{s}{\otimes} g^{A\dot{B}}, \tag{1.1}$$

and the convention is that

$$g^{A\dot{B}} = \sqrt{2}\begin{pmatrix} e^4, & e^2 \\ e^1, & -e^3 \end{pmatrix}. \tag{1.2}$$

The skew-symmetric Levi-Civita symbols are defined by

$$(\epsilon_{AB}) = (\epsilon^{AB}) = (\epsilon_{\dot{A}\dot{B}}) = (\epsilon^{\dot{A}\dot{B}}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{1.3}$$

and the spinorial indices are manipulated as follows:

$$\psi_A = \epsilon_{AB}\psi^B, \quad \psi^A = \psi_B \epsilon^{BA}. \tag{1.4}$$

The null-tetrad dual to $\{g^{A\dot{B}}\}$ consists of the vector fields $\{\partial_{A\dot{B}}\}$, such that $g^{A\dot{B}}(\partial_{B\dot{C}}) = -2\delta^A{}_B \delta^{\dot{B}}{}_{\dot{C}}$.

The two-forms S^{AB} and $S^{\dot{A}\dot{B}}$ are defined by the formula

$$g^{A\dot{B}} \wedge g^{C\dot{D}} = \epsilon^{AC} S^{\dot{B}\dot{D}} + \epsilon^{\dot{B}\dot{D}} S^{AC}. \tag{1.5}$$

The components of a one-form α are determined according to

$$\alpha = -\tfrac{1}{2}\alpha_{A\dot{B}} g^{A\dot{B}}. \tag{1.6}$$

Let Λ denote the Cartan–Grassman algebra of holomorphic forms on M, and let $\mathscr{S}$ be a module of holomorphic spinor fields on M.

Then $D: \Lambda \otimes \mathscr{S} \to \Lambda \otimes \mathscr{S}$ is defined according to the formula

$$DT^{\dots}_{\dots} = dT^{\dots}_{\dots} + \Gamma^A{}_S \wedge T^{\dots S \dots}_{\dots} - \Gamma_B{}^S \wedge T^{\dots}_{\dots S \dots} + \Gamma^{\dot{C}}{}_{\dot{S}} \wedge T^{\dots \dot{S} \dots}_{\dots} - \Gamma_{\dot{D}}{}^{\dot{S}} \wedge T^{\dots}_{\dots \dot{S} \dots} + \cdots \tag{1.7}$$

for $T \in \Lambda \otimes \mathscr{S}$, where $\Gamma_{AB} = \Gamma_{(AB)}$ and $\Gamma_{\dot{A}\dot{B}} = \Gamma_{(\dot{A}\dot{B})}$ are affine connection one-forms determined uniquely by the condition

$$Dg^{A\dot{B}} = 0. \tag{1.8}$$

As a consequence of (1.7) the Ricci identities follow:

$$D^2 T^{\dots}_{\dots} = R^A{}_S \wedge T^{\dots S \dots}_{\dots} - R_B{}^S \wedge T^{\dots}_{\dots S \dots} + R^{\dot{C}}{}_{\dot{S}} \wedge T^{\dots \dot{S} \dots}_{\dots} - R_{\dot{D}}{}^{\dot{S}} \wedge T^{\dots}_{\dots \dot{S} \dots}, \tag{1.9}$$

 0022-2488/84/030589-08$02.50

with $R^A{}_B$ and $R^{\dot A}{}_{\dot B}$ being curvature two-forms:

$$R_{AB} = d\Gamma_{AB} + \Gamma_{AS} \wedge \Gamma^S{}_B, \tag{1.10}$$

$$R_{AB} = -\tfrac{1}{2}C_{ABCD}S^{CD} + \tfrac{1}{24}RS_{AB} + \tfrac{1}{2}C_{AB\dot C\dot D}S^{\dot C\dot D}, \quad \text{etc.} \tag{1.11}$$

We notice also that

$$DS^{AB} = 0 = DS^{\dot A\dot B}. \tag{1.12}$$

Now let $T \in \mathscr{S}$; then $\nabla_{A\dot B}T^{\cdots}_{\cdots}$ is defined by

$$DT^{\cdots}_{\cdots} = -\tfrac{1}{2}(\nabla_{A\dot B}T^{\cdots}_{\cdots})g^{A\dot B}. \tag{1.13}$$

The Bianchi identities obtained by application of the operator D to both sides of (1.10) and (1.11) are

$$\begin{aligned}&\nabla^S{}_{\dot A}C_{BCDS} + \nabla^{\dot S}_{(B}C_{CD)\dot A\dot S} = 0,\\ &\nabla_A{}^{\dot S}C_{\dot B\dot C\dot D\dot S} + \nabla^S_{(\dot B}C_{|AS|\dot C\dot D)} = 0,\\ &\nabla^{R\dot S}C_{AR\dot B\dot S} + \tfrac{1}{8}\nabla_{A\dot B}R = 0.\end{aligned} \tag{1.14}$$

Next we notice that according to our conventions, for any one-form α its inner product "$\lrcorner$" with S^{AB} is determined by

$$\alpha \lrcorner S_{AB} = -\alpha_{(A}{}^{\dot N}g_{B)\dot N}. \tag{1.15}$$

We recall also that $g^{C\dot D} \wedge S^{AB}$ provide a basis for three-forms, denoted by $\check g^{A\dot B}$, such that

$$S^{AB} \wedge g^{C\dot D} = \epsilon^{BC}\check g^{A\dot D} + \epsilon^{AC}\check g^{B\dot D}. \tag{1.16}$$

2. A GENERALIZATION OF CANONICAL CONGRUENCES IN HEAVENS

We begin this section by recalling that the existence of self-dual congruence of null strings is equivalent to the existence of a spinor field k^A satisfying the equation

$$k^A k^B \nabla_{A\dot C} k_B = 0, \tag{2.1}$$

which is equivalent to the statement that the two-form

$$\Sigma := k_A k_B S^{AB} \tag{2.2}$$

is integrable in the sense that it vanishes on a two-parameter set (congruence) of null strings.[3] We notice also that (2.1) does not distinguish between spinor fields which are proportional.

Assume now, that there is a set of solutions of (2.1), which is at the same time a two-dimensional vector space F. Take two arbitrary, independent members $'k^A$ and $''k^A$. Then obviously

$$F = \{k^A : k^A = a'k_A + b''k_A, a,b \in \mathbb{C}\}. \tag{2.3}$$

Next we notice that from (2.1) it follows that

$$\nabla_{A\dot B}{}'k_C = {}'\rho_{A\dot B}{}'k_C + {}'\sigma_{C\dot B}{}'k_A \tag{2.4}$$

and

$$\nabla_{A\dot B}{}''k_C = {}''\rho_{A\dot B}{}''k_C + {}''\sigma_{C\dot B}{}''k_A \tag{2.5}$$

for some spinor fields $'\rho_{A\dot B}$, $''\rho_{A\dot B}$, $'\sigma_{A\dot B}$, and $''\sigma_{A\dot B}$.

Then substituting a general member of F, in the form of (2.3), into (2.1), after some elementary spinorial calculations one obtains the condition

$$''\rho_{A\dot B} + {}''\sigma_{A\dot B} = {}'\rho_{A\dot B} + {}'\sigma_{A\dot B}. \tag{2.6}$$

Define now the following two-forms:

$$'\Sigma = {}'k_A{}'k_B S^{AB}, \quad ''\Sigma = {}''k_A{}''k_B S^{AB}, \quad '''\Sigma = {}'k_A{}''k_B S^{AB}. \tag{2.7}$$

It is easy to show that

$$d'\Sigma = -2\alpha \wedge {}'\Sigma, \tag{2.8}$$

and

$$d''\Sigma = -2\beta \wedge {}''\Sigma. \tag{2.9}$$

In fact, those equations are equivalent to (2.1) with k^A replaced by $'k^A$ and $''k^A$, respectively.[3] To obtain an explicit form of α and β, we employ the formalism of Ref. 8:
$d'\Sigma = 2'k_A D'k_B \wedge S^{AB} = -'k_A(\nabla_{C\dot D}{}'k_B)g^{C\dot D} \wedge S^{AB}$
$= {}'k^C{}'k_B({}'\sigma_{C\dot D} - 2'\rho_{C\dot D})\check g^{B\dot D} = -2\alpha \wedge {}'\Sigma$
$= 2\alpha_{C\dot D}{}'k^C{}'k_B\check g^{B\dot D}$ [for the definition of $\check g^{A\dot B}$ see (1.16)].
Hence,

$$2\alpha_{C\dot D} = {}'\sigma_{C\dot D} - 2'\rho_{C\dot D} + {}'k_C\alpha_{\dot D} \tag{2.10}$$

and $\alpha = -\tfrac{1}{2}\alpha_{C\dot D}g^{C\dot D}$ ($\alpha_{\dot D}$ is an arbitrary spinor).

In a similar way one obtains

$$2\beta_{C\dot D} = {}''\sigma_{C\dot D} - 2''\rho_{C\dot D} + {}''k_C\beta_{\dot D} \quad \text{and} \quad \beta = -\tfrac{1}{2}\beta_{A\dot B}g^{A\dot B}. \tag{2.11}$$

It is important to notice now that, since $'k_C$ and $''k_C$ are independent, one can always find $\alpha_{\dot D}$ and $\beta_{\dot D}$ such that $\alpha = \beta$. Suppose, therefore, such a choice has been made. Then, employing (2.4)–(2.6) and (2.10) with that specific $\alpha_{\dot D}$, one checks that

$$d'''\Sigma = -2\alpha \wedge {}'''\Sigma. \tag{2.12}$$

Now we observe that a rescaling of all elements of F, by the same factor, results in a set $'F$ with the properties identical to those of F. Therefore, employing this freedom, as well as that of null-tetrad one-forms $g^{A\dot B}$, one can obtain $'k^A$ and $''k^A$ in the form of: $'k^A = \delta^A_1$ and $''k^A = \delta^A_2$. Then Eqs. (2.8), (2.9), and (2.12) are equivalent to

$$dS^{AB} = -2\alpha \wedge S^{AB}. \tag{2.13}$$

The condition (2.13) has been used by Finley and Plebański to define a class of spaces called "projective extensions of heavens."[7] The following lemma is true.

Lemma 2.1: There exists a set of solutions of (2.1) forming a two-dimensional vector space if and only if there is a null-tetrad gauge in which $dS^{AB} = -2\alpha \wedge S^{AB}$.

The ambiguity of that gauge consists of constant "heavenly" and arbitrary "hellish" SL(2,$\mathbb{C}$) transformations.

It is only "projective extensions of heavens" (among them heavens) which admit such gauges.

Proof: Indeed the assertion follows in one direction from considerations above. On the other hand, if there is a gauge in which $dS^{AB} = -2\alpha \wedge S^{AB}$, then $'k^A = \delta^A_1$ and $''k^A = \delta^A_2$ satisfy automatically (2.1) and the members of the corresponding set F have also that property. The remaining part of the lemma follows from the transformational properties of S^{AB}. ∎

A nontrivial solution of (2.1), and all others proportional to it, give rise to a congruence of null strings. Therefore, $F - \{0\}$ defines a *one-parameter* family of congruences of null strings parametrized by the points of $\mathbb{P}^1\mathbb{C}$ [ratios of the coefficients a and b in (2.3)]. We call it a *generalized canonical family* of congruences of null strings. By a *canonical fam-*

ily we mean one which corresponds to $\alpha = 0$. It exists only in heavens.

3. INTEGRABILITY CONDITIONS

The condition (2.13) is of great convenience if one is trying to find an explicit form of projective extensions of heavens.[7] On the other hand, to discuss the place of those spaces among all left conformally flat ones, a more covariant treatment is needed. The aim of this section is to establish the integrability conditions for the existence of that specific gauge (2.13).

It is convenient to change the notation slightly, endowing all quantities related to the gauge (2.13) with primes. Next, we notice an equivalence between (2.13) and the statement that

$$'\Gamma_{AB} = \tfrac{1}{2}(\alpha \lrcorner 'S_{AB}). \tag{3.1}$$

One proves this by means of $D\,'S_{AB} = 0$, (2.3), and the definition of the step product $\lrcorner$,[3,8] [see also (1.15)].

Further, from the transformation properties of "heavenly" connection one-forms Γ_{AB},[8] one infers that in an arbitrary tetradial gauge

$$g^{A\dot{B}} = l^{A}{}_{C}\,'g^{C\dot{B}}, \quad (l^{A}{}_{B}) \in \mathrm{SL}(2,\mathbb{C}), \tag{3.2}$$

the following relation holds:

$$\Gamma_{AB} = l^{-1C}{}_{A}\, l^{-1D}{}_{B}\,'\Gamma_{CD} - l^{-1S}{}_{B}\, dl_{AS}. \tag{3.3}$$

[One does not lose generality by restricting oneself to "heavenly" transformations (3.2) only, because the "hellish" ones do not effect Γ_{AB}'s.] Next, taking into account (3.1) and the formula

$$S_{AB} = l^{-1C}{}_{A}\, l^{-1D}{}_{B}\,'S_{CD}, \tag{3.4}$$

one obtains

$$dl_{AB} = \{\tfrac{1}{2}(\alpha \lrcorner S_{AC}) - \Gamma_{AC}\}\, l^{C}{}_{B}. \tag{3.5}$$

We now study the integrability conditions of (3.5). Applying the operator of external differentiation d to both sides of (3.5), and substituting into the resulting formula the right-hand side of (3.5) in place of dl_{MN}, one obtains

$$D(\alpha \lrcorner S_{AB}) + \tfrac{1}{2}(\alpha \lrcorner S^{K}{}_{B}) \wedge (\alpha \lrcorner S_{AK}) = 2R_{AB}. \tag{3.6}$$

Then making use of the formulas (1.7), (1.8), (1.11), and (1.15) one arrives at the conditions

$$C_{ABCD} = 0, \tag{3.7}$$

$$C^{AB}{}_{\dot{C}\dot{D}} = -\tfrac{1}{2}\{\nabla^{(A}{}_{(\dot{C}}\alpha^{B)}{}_{\dot{D})} - \alpha^{A}{}_{(\dot{C}}\alpha^{B}{}_{\dot{D})}\}, \tag{3.8}$$

$$R/3 = \nabla^{M\dot{N}}\alpha_{M\dot{N}} + \alpha^{M\dot{N}}\alpha_{M\dot{N}}, \tag{3.9}$$

$$\nabla_{(A}{}^{\dot{N}}\alpha_{B)\dot{N}} = 0. \tag{3.10}$$

Thus at the first step of the procedure all irreducible components of $\nabla^{A}{}_{\dot{C}}\alpha^{B}{}_{\dot{D}}$ but $\nabla_{M(\dot{C}}\alpha^{M}{}_{\dot{D})}$ have been expressed in terms of $\alpha_{A\dot{B}}$ and the curvature quantities.

Now, let $\rho_{\dot{A}\dot{B}}$ be defined by

$$\rho_{\dot{A}\dot{B}} = \nabla_{M(\dot{A}}\alpha^{M}{}_{\dot{B})}. \tag{3.11}$$

Then one easily finds that

$$\begin{aligned} D\alpha^{B}{}_{\dot{D}} &= -\tfrac{1}{2}(\nabla_{M\dot{N}}\alpha^{B}{}_{\dot{D}})g^{M\dot{N}} \\ &= C^{B}{}_{M\dot{N}\dot{D}}\,g^{M\dot{N}} - \tfrac{1}{2}\alpha_{M(\dot{N}}\alpha^{B}{}_{\dot{D})}g^{M\dot{N}} - \tfrac{1}{4}\rho_{\dot{N}\dot{D}}\,g^{B\dot{N}} \\ &\quad + \tfrac{1}{8}\epsilon_{\dot{N}\dot{D}}(R/3 - \alpha^{K\dot{L}}\alpha_{K\dot{L}})g^{B\dot{N}}. \end{aligned} \tag{3.12}$$

At the next step one acts on both sides of (3.12) with D. This time the calculation requires much more work, and it is not presented here. Some of the conditions resulting are fulfilled automatically because of the Bianchi identities; the rest are equivalent to the equation

$$\nabla_{A\dot{B}}\rho_{\dot{C}\dot{D}} = -4\nabla_{A}{}^{\dot{S}}C_{\dot{B}\dot{C}\dot{D}\dot{S}} + 4\alpha_{A\dot{S}}C^{\dot{S}}{}_{\dot{B}\dot{C}\dot{D}} + 3\alpha_{A(\dot{B}}\rho_{\dot{C}\dot{D})}. \tag{3.13}$$

Further integrability conditions are obtained by forming $D\rho_{\dot{C}\dot{D}}$ and acting on both sides of the resulting formula with D. The amount of work to be done at this stage is even greater than before; therefore only the final result is displayed,

$$\begin{aligned} &\nabla_{M(\dot{A}}\nabla^{M\dot{N}}C_{\dot{B}\dot{C}\dot{D})\dot{N}} + 2\rho_{(\dot{A}}{}^{\dot{N}}C_{\dot{B}\dot{C}\dot{D})\dot{N}} \\ &\quad - \tfrac{3}{4}\rho_{(\dot{A}\dot{B}}\rho_{\dot{C}\dot{D})} + \alpha^{M\dot{N}}\nabla_{M(\dot{A}}C_{\dot{B}\dot{C}\dot{D})\dot{N}} \\ &\quad - 3\alpha_{M(\dot{A}}\nabla^{M\dot{N}}C_{\dot{B}\dot{C}\dot{D})\dot{N}} + (R/6)C_{\dot{A}\dot{B}\dot{C}\dot{D}} \\ &\quad - 2\alpha^{M\dot{N}}\alpha_{M\dot{N}}C_{\dot{A}\dot{B}\dot{C}\dot{D}} = 0. \end{aligned} \tag{3.14}$$

Subsequent integrability conditions can be obtained from (3.14) by action with $\nabla_{P\dot{Q}}$ on both its sides, and then elimination of $\nabla_{P\dot{Q}}\rho_{\dot{C}\dot{D}}$ and $\nabla_{P\dot{Q}}\alpha_{A\dot{B}}$ by (3.12) and (3.13), respectively. Continuation of that process gives rise to a sequence of algebraic equations which involve $\rho_{\dot{C}\dot{D}}$, $\alpha_{A\dot{B}}$, and curvature quantities only. Let the conditions obtained on the Nth step be denoted by $(3.14)_N$, $N \geqslant 0$ [$(3.14)_0 \equiv (3.14)$]. We claim the following Lemma:

Lemma 3.1: (i) A left conformally flat space-time admits a *generalized canonical family* of congruences of null strings if and only if the set of the equations (3.12) and (3.13) is integrable for $\rho_{\dot{A}\dot{B}} = \rho_{(\dot{A}\dot{B})}$ and $\alpha_{A\dot{B}}$.

(ii) Let $C_{ABCD} = 0$. The set of equations (3.12) and (3.13) is integrable in a neighborhood of the point $p \in M$ if and only if there exists an integer number N, $0 \leqslant N \leqslant 7$ such that the equations of the conditions $(3.14)_0$, $(3.14)_1$, ..., $(3.14)_N$ are compatible in some neighborhood of p and the equations of the conditions $(3.14)_{N+1}$ are satisfied because of the former conditions.

If k is the number of independent equations in the first $(N+1)$ conditions, the solution involves $7-k$ arbitrary constants. $k \geqslant 3$ always and $k = 3$ for conformally flat spaces.

Proof: Indeed, the integrability of (3.5) is equivalent to the integrability of (3.12) and (3.13), and any solution of (3.5) which belongs to SL(2,$\mathbb{C}$) at some point p remains in SL(2,$\mathbb{C}$) at all points of its domain, because of the identity: $d(\epsilon_{AB}l^{A}{}_{C}l^{B}{}_{D})\,0$.

The second part of the lemma is a result of a more general theorem, given in Ref. 9. The lower bound on k can be obtained from $(3.14)_0$. Indeed, at least three of five equations on $\rho_{\dot{A}\dot{B}}$ and $\alpha_{M\dot{N}}$ represented by that condition can be solved for $\rho_{\dot{1}\dot{1}}$, $\rho_{\dot{1}\dot{2}}$ and $\rho_{\dot{2}\dot{2}}$ (locally). For conformally flat spaces, (3.14) reduces to $\rho_{(\dot{A}\dot{B}}\rho_{\dot{C}\dot{D})} = 0$, which constitute three independent equations: $\rho_{\dot{A}\dot{B}} = 0$. ■

Notice also the following simple lemma.

Lemma 3.2: (i) Let $d\tilde{s}^2$ and ds^2 be two conformally related metrics:

$$d\tilde{s}^2 = \phi^{-2}\, ds^2.$$

Then they have in common *generalized canonical families* of congruences of null strings, and for the corresponding objects the following relations hold:

$$\tilde{\alpha} = \alpha + d(\ln\phi), \quad \tilde{\rho}_{\dot{A}\dot{B}} = \rho_{\dot{A}\dot{B}}.$$

In particular, ds^2 is conformally related to a "heaven," if and only if there is a *generalized canonical family* with $\rho_{\dot{A}\dot{B}} = 0$ $(d\alpha = 0)$.

(ii) A heavenly metric ds^2 is conformally related to another heavenly one with a nontrivial conformal factor if and only if there exists a *generalized canonical family* such that

$$\rho_{\dot{A}\dot{B}} = 0 \quad \text{and} \quad \alpha_{A\dot{S}} C^{\dot{S}}{}_{\dot{B}\dot{C}\dot{D}} = 0,$$

where $\alpha \neq 0$. Then either $C_{\dot{A}\dot{B}\dot{C}\dot{D}} \neq 0$, $\alpha_{A\dot{B}} = \alpha_A k_{\dot{B}}$ and therefore $C_{\dot{A}\dot{B}\dot{C}\dot{D}}$ is of the algebraic type N, or $C_{\dot{A}\dot{B}\dot{C}\dot{D}} = 0$.

(iii) Any *generalized canonical family* in conformally flat space has $\rho_{\dot{A}\dot{B}} = 0$ $(d\alpha = 0)$.

Proof: Indeed, the first part of (i) is a consequence of (2.13) and the fact that $\tilde{S}^{AB} = \phi^{-2} S^{AB}$. The second part is obtained by combining the behavior of α under conformal transformations with the properties of heavens, which admit canonical families of congruences of null strings $(\alpha = 0)$.

To prove (ii), suppose at first that $d\tilde{s}^2 = \phi^{-2} ds^2$, with $d\tilde{s}^2$ and ds^2 being "heavens." Take then the canonical family corresponding to $d\tilde{s}^2$ for which $\tilde{\alpha} = 0, \tilde{\rho}_{\dot{A}\dot{B}} = 0$. Now, from (i), (3.13), and Bianchi's identities the assertion (ii) in one direction follows. Conversely, suppose that ds^2 is a "heavenly" metric with a generalized canonical family such that $\rho_{\dot{A}\dot{B}} = 0$ and $\alpha_{A\dot{S}} C^{\dot{S}}{}_{\dot{B}\dot{C}\dot{D}} = 0$. Then define ϕ according to the formula $d(\ln\phi) = -\alpha$. One easily infers that $d\tilde{s}^2 = \phi^{-2} ds^2$ is a "heavenly" metric structure. The second part of (ii) follows from the observation that $\alpha_{A\dot{S}} C^{\dot{S}}{}_{\dot{B}\dot{C}\dot{D}} = 0 \rightarrow (\alpha^{M\dot{N}} \alpha_{M\dot{N}}) C_{\dot{A}\dot{B}\dot{C}\dot{D}} = 0$. (iii) is a result of (3.14). ■

In fact the assertion (ii) of Lemma 3.2 can be stated more explicitly due to an old result of Brinkmann.[10]

Lemma 3.3: The most general "heaven" which admits a nontrivial conformal map into another "heaven" is a self-dual plane-fronted wave (in the simplest case a flat space).

Although it is rather obvious that "projective extensions of heavens" are only a subclass of one-side conformally flat spaces, it requires usually some work to demonstrate that for a given metric structure with the property that $C_{ABCD} = 0$, the appropriate integrability conditions cannot be satisfied. An example below illustrates such a situation.

Example: Consider the metric of the form[6]

$$ds^2 = 2(e^1 \otimes e^2 + e^3 \otimes e^4), \tag{3.15}$$

where

$$e^1 = du + (\tfrac{1}{2}u^2 a - uq)dr, \quad e^2 = dr,$$
$$e^3 = dq, \quad e^4 = dv + (-\tfrac{1}{2}av^2 + vr)dq,$$

and $a = \text{const}$.

The connection one-forms Γ_{AB} in this tetrad are given by

$$\Gamma_{11} = 0 = \Gamma_{22}, \quad \Gamma_{12} = \tfrac{1}{2}[(au - q)dr + (av - r)dq],$$
$$\Gamma_{\dot{1}\dot{1}} = 0, \quad \Gamma_{\dot{1}\dot{2}} = \tfrac{1}{2}[(av - r)dq + (q - au)dr],$$
$$\Gamma_{\dot{2}\dot{2}} = -u\,dr - v\,dq. \tag{3.16}$$

Next one shows that $C_{ABCD} = 0 = R$ as well as that the only nonvanishing components of $C_{\dot{A}\dot{B}\dot{C}\dot{D}}$ and $C_{AB\dot{C}\dot{D}}$ are

$$C_{\dot{1}\dot{2}\dot{2}\dot{2}} = -1, \quad C_{\dot{2}\dot{2}\dot{2}\dot{2}} = u(r - av) + v(q - au),$$
$$C_{12\dot{1}\dot{2}} = -\tfrac{1}{2}a. \tag{3.17}$$

Thus the conformal curvature is of the algebraic type $[-]\times\text{III}$, and in the limit $a \rightarrow 0$ the space becomes a "heaven."

Provided $a \neq 0$ there does not exist a *generalized canonical family* of congruences of null strings. For $a = 0$ there exists only that corresponding to $\alpha = 0$, the canonical one.

To justify this assertion one observes that (3.14) implies immediately that $\rho_{\dot{1}\dot{1}} = 0$. Then from Eqs. (3.13), and some of (3.12), one finds that $\rho_{\dot{1}\dot{2}} = 0$. Next employing again (3.13), it turns out, provided $a \neq 0$, that necessarily $\rho_{\dot{2}\dot{2}} = -4$. When $a = 0$ there is the second possibility $\rho_{\dot{2}\dot{2}} \neq -4$ and at the same time $\omega_{A\dot{1}} = 0$. Further, taking into account the remaining equations of (3.13) and their integrability conditions one arrives at a contradiction for $\rho_{\dot{2}\dot{2}} - 4$. The second possibility amounts to $\omega_{A\dot{B}} = 0 = \rho_{\dot{A}\dot{B}}$ which corresponds to the canonical family of congruences of null strings.

Examples of "projective extensions of heavens" can be found in Ref. 7. We close this section with remarks about *generalized canonical families* of congruences of null strings in spaces conformally related to complexified Minkowski space-time. It is clear (Lemma 3.2), that they are identical with those for the Minkowski space-time ds^2. Further, the same lemma implies that any *generalized canonical* family for ds^2 is a canonical one with respect to another metric structure $d\tilde{s}^2$, conformally related to ds^2: $d\tilde{s}^2 = \phi^{-2} ds^2$, where ϕ is defined by $d(\ln\phi) = -\alpha$. It is easy to find the general form of ϕ. Indeed one has to solve (3.12), where $\rho_{\dot{A}\dot{B}} = 0, \alpha = -d(\ln\phi)$, with respect to ϕ. Working in the tetradial gauge for ds^2 in which

$$g^{A\dot{B}} = dx^{A\dot{B}}, \tag{3.18}$$

one infers that ϕ is given by

$$\phi = x^{M\dot{N}} x_{M\dot{N}} \tag{3.19}$$

or

$$\phi = 1 + 2l_{M\dot{N}} x^{M\dot{N}}, \tag{3.20}$$

where $l_{A\dot{B}}$ is any constant, lightlike vector, or ϕ is of the form obtained from those above by an arbitrary translation of $x^{A\dot{B}}$: $x^{A\dot{B}} \rightarrow x^{A\dot{B}} + t^{A\dot{B}}$. Therefore, all congruences related to a fixed ϕ are defined only on a part of original Minkowski space-time, that without a null cone or a null hyperplane $(\alpha = -d\phi/\phi)$.

To obtain an explicit form of the families of congruences one has to perform the transformations, which in the two essentially different cases are

$$\tilde{x}^{A\dot{B}} = \frac{x^{A\dot{B}}}{(x^{M\dot{N}} x_{M\dot{N}})}, \tag{3.19$'$}$$

$$\tilde{x}^{A\dot{B}} = \frac{x^{A\dot{B}} + l^{A\dot{B}}(x^{M\dot{N}} x_{M\dot{N}})}{1 + 2l_{M\dot{N}} x^{M\dot{N}}}. \tag{3.20$'$}$$

Then $d\tilde{s}^2 = \phi^{-2} ds^2 = -\tfrac{1}{2} d\tilde{x}^{A\dot{B}} d\tilde{x}_{A\dot{B}}$.

The next observation is that the null tetrads $\tilde{g}^{A\dot{B}} := d\tilde{x}^{A\dot{B}}$ and $'\tilde{g}^{A\dot{B}} := \phi^{-1} dx^{A\dot{B}}$ related to $d\tilde{s}^2$ are of the same orientation if (3.20) holds, and of opposite ones for (3.19). It is therefore clear that provided (3.20) is satisfied, it is the canonical family of self-dual congruences of the null strings for $d\tilde{s}^2$ which is the *generalized canonical family* for ds^2.

Otherwise, in the case (3.19), one has to consider the canonical family of anti-self-dual congruences for $d\tilde{s}^2$.

Now let $\tilde{X}^{\dot{A}}$, $\tilde{Y}^{\dot{A}}$ and $\tilde{X}^{A}$, $\tilde{Y}^{A}$ be defined by

$$\begin{pmatrix} \tilde{Y}^{\dot{2}} & \tilde{Y}^{\dot{1}} \\ \tilde{X}^{\dot{2}} & \tilde{X}^{\dot{1}} \end{pmatrix} := (\tilde{X}^{A\dot{B}}) \tag{3.20$''$}$$

in the case (3.20)′ and

$$\begin{pmatrix} \tilde{X}^{1} & \tilde{Y}^{1} \\ \tilde{X}^{2} & \tilde{Y}^{2} \end{pmatrix} := (\tilde{X}^{A\dot{B}}) \quad \text{for (3.19)}'. \tag{3.19$''$}$$

Then the functions

$$\tilde{Z}^{\dot{A}} = \pi^{2}\tilde{X}^{\dot{A}} + \pi^{1}\tilde{Y}^{\dot{A}} \tag{3.20$'''$}$$

and

$$\tilde{Z}^{A} = \pi^{\dot{2}}\tilde{X}^{A} + \pi^{\dot{1}}\tilde{Y}^{A}, \tag{3.19$'''$}$$

where π^A and $\pi^{\dot{A}}$ are arbitrary nonzero constant spinors, define the corresponding families of congruences (compare Ref. 4).

4. GENERALIZED CANONICAL FAMILIES IN HEAVENS

The left-flat spaces (heavens) are those characterized by the existence of a canonical family of self-dual congruences of null strings, e.g., by the existence of a gauge such that (2.13) with $\alpha = 0$ holds. In the simplest case, that of the Minkowski space (Sec. 3), there is also a large class of generalized canonical families; for all of them $d\alpha = 0$.

Below we discuss the problem of generalized canonical families for an arbitrary heaven.

Any heaven can be described explicitly as follows.[11] There exists a coordinate system $\{q^{\dot{A}}, q^{\bar{B}}\}$ and a tetradial gauge $g^{A\dot{B}}$ such that

$$g^{1\dot{A}} = \sqrt{2}\Omega^{,\dot{A}}{}_{,\bar{B}}\, dq^{\bar{B}}, \tag{4.1}$$

$$g^{2\dot{A}} = \sqrt{2}\, dq^{\dot{A}} \tag{4.2}$$

for some function Ω, which fulfills additionally the equation

$$\Omega^{,\dot{A}\dot{B}}\Omega_{,\dot{A}\dot{B}} = -2. \tag{4.3}$$

The dual tetrad related to (4.1), (4.2) is of the form

$$\partial_{1\dot{A}} = -\sqrt{2}\Omega^{;\bar{B}}_{,\dot{A}}\frac{\partial}{\partial q^{\bar{B}}}, \tag{4.4}$$

$$\partial_{2\dot{A}} = -\sqrt{2}\frac{\partial}{\partial q^{\dot{A}}}. \tag{4.5}$$

Then

$$\Gamma_{AB} = 0, \tag{4.6}$$

and

$$\begin{aligned} S^{11} &= 2\, dq^{\bar{1}} \wedge dq^{\bar{2}}, \quad S^{22} = 2\, dq^{\dot{1}} \wedge dq^{\dot{2}}, \\ S^{12} &= \Omega_{,\dot{A}\bar{B}}\, dq^{\dot{A}} \wedge dq^{\bar{B}}. \end{aligned} \tag{4.7}$$

Now denote by $'g^{A\dot{B}}$ the tetradial gauge in which (3.1) is supposed to hold. Then there is a SL(2,C) transformation $(m^A{}_B)$ such that

$$'g^{A\dot{B}} = m^{A}{}_{C}g^{C\dot{B}}. \tag{4.8}$$

Then employing (3.5) and (4.6) properly, one obtains

$$dm_{AB} = -\tfrac{1}{2}(\alpha \lrcorner S_{BC})m^{-1C}{}_{A}. \tag{4.9}$$

An explicit form of this equation is

$$m^{C}{}_{A}\partial_{M\dot{N}}(m_{CB}) = -\tfrac{1}{2}(\epsilon_{MA}\alpha_{B\dot{N}} + \epsilon_{MB}\alpha_{A\dot{N}}). \tag{4.10}$$

It can be shown, by simple algebraic arguments, that (4.10) possesses a solution $\alpha_{A\dot{B}}$ for $(m^A{}_B)$ fixed if and only if

$$m^{C}_{(A}\partial_{M}{}^{\dot{N}}m_{|C|B)} = 0. \tag{4.11}$$

Then

$$\alpha_{A\dot{B}} = -\tfrac{2}{3}m^{C}{}_{D}\partial^{D}{}_{\dot{B}}(m_{CA}). \tag{4.12}$$

We shall now study the system of equations given by (4.11). It is convenient to introduce the notation

$$\begin{pmatrix} \mu & \nu \\ \rho & \sigma \end{pmatrix} := \begin{pmatrix} m^{1}{}_{1} & m^{1}{}_{2} \\ m^{2}{}_{1} & m^{2}{}_{2} \end{pmatrix}, \tag{4.13}$$

where μ, ν, ρ, and σ are subject to the constraint

$$\mu\sigma - \nu\rho = 1. \tag{4.14}$$

Then employing (4.4) and (4.5) one can rewrite (4.11) in the form

$$\mu\rho_{,\bar{N}} - \rho\mu_{,\bar{N}} = 0, \tag{4.15}$$

$$\nu\sigma_{,\dot{N}} - \sigma\nu_{,\dot{N}} = 0, \tag{4.16}$$

$$\mu\sigma_{,\bar{N}} - \sigma\mu_{,\bar{N}} + \nu\rho_{,\bar{N}} - \rho\nu_{,\bar{N}} + (\mu\rho_{,\dot{M}} - \rho\mu_{,\dot{M}})\Omega^{;\dot{M}}_{,\bar{N}} = 0, \tag{4.17}$$

$$\nu\rho_{,\dot{N}} - \rho\nu_{,\dot{N}} + \mu\sigma_{,\dot{N}} - \sigma\mu_{,\dot{N}} + (\nu\sigma_{,\bar{M}} - \sigma\nu_{,\bar{M}})\Omega^{;\bar{M}}_{,\dot{N}} = 0. \tag{4.18}$$

There are three cases to be discussed independently.

Case 1, $\mu \neq 0 \neq \nu$: Then the equations of the system (4.15) are integrated immediately, resulting in

$$\rho = -\mu F, \tag{4.19}$$

where F is a function of $q^{\dot{A}}$'s only.

Next from (4.14) σ can be calculated:

$$\sigma = -\nu F + 1/\mu. \tag{4.20}$$

Now employing (4.19) and (4.20) in (4.16)–(4.18) one obtains

$$(\mu\nu)_{,\dot{N}} = -(\mu\nu)^2 F_{,\dot{N}}, \tag{4.21}$$

$$\mu_{,\bar{N}}/\mu^3 = -\tfrac{1}{2}(F_{,\dot{M}}\Omega^{,\dot{M}})_{,\bar{N}}, \tag{4.22}$$

$$\mu_{,\dot{N}}/\mu = -\mu\nu F_{,\dot{N}} - (1/2\mu^2)(\mu\nu)_{,\bar{M}}\Omega^{;\bar{M}}_{,\dot{N}}. \tag{4.23}$$

As a consequence of (4.21) it follows that

$$\mu\nu = 1/(F + \bar{H}), \tag{4.24}$$

while (4.22) results in

$$\mu = (G + F_{,\dot{M}}\Omega^{,\dot{M}})^{-1/2}, \tag{4.25}$$

where G is a function of $q^{\dot{A}}$'s and $\bar{H}$ a function of $q^{\bar{A}}$'s only. From (4.24), (4.19), and (4.20) we infer also that

$$\nu = \frac{(G + F_{,\dot{C}}\Omega^{,\dot{C}})^{1/2}}{F + \bar{H}}, \tag{4.26}$$

$$\rho = -F(G + F_{,\dot{C}}\Omega^{,\dot{C}})^{-1/2}, \tag{4.27}$$

$$\sigma = \frac{\bar{H}}{F + \bar{H}}(G + F_{,\dot{C}}\Omega^{,\dot{C}})^{1/2}. \tag{4.28}$$

Now let the function f be defined by

$$f:=\frac{G+F_{,\dot M}\Omega^{,\dot M}}{(F+\overline{H})^2}. \tag{4.29}$$

It is not difficult to show that (4.23) is equivalent to

$$f_{,\dot N}/f^2=-(\overline{H}_{,\overline{\dot M}}\Omega^{,\overline{\dot M}})_{,\dot N}. \tag{4.30}$$

Thus $f=(\overline{E}+\overline{H}_{,\overline{\dot M}}\Omega^{,\overline{\dot M}})^{-1}$, where $\overline{E}$ is a function of $q^{\overline{A}}$'s only, and because of (4.29) one arrives at the following constraint on Ω:

$$(G+F_{,\dot M}\Omega^{,\dot M})(\overline{E}+\overline{H}_{,\overline{\dot M}}\Omega^{,\overline{\dot M}})=(F+\overline{H})^2. \tag{4.31}$$

Case 2, $\mu\neq 0=\nu$: Then (4.21) becomes an identity. The formulas for μ and ρ preserve their form (4.25) and (4.27), respectively, whereas σ is determined from (4.20):

$$\sigma=(G+F_{,\dot M}\Omega^{,\dot M})^{1/2}. \tag{4.32}$$

Equation (4.23) amounts to the following condition on Ω:

$$\overline{E}(G+F_{,\dot M}\Omega^{,\dot M})=1. \tag{4.33}$$

Case 3, $\mu=0$ and consequently (4.14) $\nu\neq 0$: This case is similar to the former one.

$$\nu=G^{1/2}, \tag{4.34}$$

$$\rho=-G^{-1/2}, \tag{4.35}$$

$$\sigma=\overline{H}G^{1/2}, \tag{4.36}$$

and the condition on Ω is of the form

$$G(\overline{E}+\overline{H}_{,\overline{\dot M}}\Omega^{,\overline{\dot M}})=1. \tag{4.37}$$

These results are summarized in a lemma below.

Lemma 4.1: Let ds^2 be a heavenly metric structure. There exists a generalized canonical family of self-dual congruences of null strings with $\alpha\neq 0$, if and only if for a pair of canonical congruences of null strings Σ_1: $q^A=$ const and Σ_2: $q^{\overline{A}}=$ const there exist four functions $F, G, \overline{E}$, and $\overline{H}$ (not all of them being constant), the first two of them constant along the leaves of Σ_1 and the remaining constant along the leaves of Σ_2 such that (4.31) or (4.33) or (4.37) holds.

$F,G,\overline{E},\overline{H}=$ const iff $(m^A{}_B)=$ const iff $d\,'S^{AB}=0$.

Further we recall that, given a pair of canonical congruences, there is an ambiguity in q^A, $q^{\overline{A}}$ as well as in Ω,[11] which does not effect (4.3) and the form of ds^2,

$$'q^A='q^A(q^B),\quad 'q^{\overline{A}}='q^{\overline{A}}(q^{\overline{B}}), \tag{4.38}$$

such that

$$\frac{\partial('q^1,'q^2)}{\partial(q^1,q^2)}\,\frac{\partial('q^{\overline{1}},'q^{\overline{2}})}{\partial(q^{\overline{1}},q^{\overline{2}})}=1 \tag{4.39}$$

and

$$\widetilde{\Omega}=\Omega+K(q^A)+\overline{L}(q^{\overline{A}}). \tag{4.40}$$

That permits us to simplify (4.31), (4.33), and (4.37) remarkably. At first, the form of $\overline{E}$ and G can be simplified by means of (4.40).

There are four situations to be considered:

(i) $dF\neq 0$, $d\overline{H}\neq 0$,
(ii) $dF\neq 0$, $d\overline{H}=0$,
(iii) $dF=0$, $d\overline{H}\neq 0$,
(iv) $dF=0$, $d\overline{H}=0$.

The specifications of K and $\overline{L}$ for the corresponding cases, as well as a new form of (4.31), (4.33), or (4.37) are given below.

Case 1: (i) K and $\overline{L}$ are determined by the equations $K^{,\dot M}F_{,\dot M}=G$, $\overline{L}^{,\overline{\dot M}}\overline{H}_{,\overline{\dot M}}=\overline{E}$, and (4.31) takes the form

$$(F_{,\dot M}\widetilde{\Omega}^{,\dot M})(\overline{H}_{,\overline{\dot M}}\widetilde{\Omega}^{,\overline{\dot M}})=(F+\overline{H})^2, \tag{4.41}$$

(ii) $K^{,\dot M}F_{,\dot M}=G$.

$$\overline{E}(F_{,\dot M}\widetilde{\Omega}^{,\dot M})=(F+\overline{H})^2, \tag{4.42}$$

(iii) $(\overline{L}^{,\overline{\dot M}}\overline{H}_{,\overline{\dot M}})=\overline{E}$.

$$G(\overline{H}_{,\overline{\dot M}}\widetilde{\Omega}^{,\overline{\dot M}})=(F+\overline{H})^2. \tag{4.43}$$

(iv) (4.31) implies that G and $\overline{E}$ are constant; thus, this situation can be excluded from further discussion (Lemma 4.1).

Case 2: The only interesting subcase corresponds to $dF\neq 0$ (Lemma 4.1). Then K is chosen according to $K^{,\dot M}F_{,\dot M}=G$ and Eq. (4.33) is reduced to

$$\overline{E}(F_{,\dot M}\widetilde{\Omega}^{,\dot M})=1. \tag{4.44}$$

Case 3: The only nontrivial subcase corresponds to $d\overline{H}\neq 0$. Then $\overline{L}$ is subject to $\overline{L}^{,\overline{\dot M}}\overline{H}_{,\overline{\dot M}}=\overline{E}$ and (4.37) takes the form

$$G(\overline{H}_{,\overline{\dot M}}\widetilde{\Omega}^{,\overline{\dot M}})=1. \tag{4.45}$$

Now, the coordinate freedom (4.38) can be used to simplify the form of F and $\overline{H}$. We shall denote a new coordinate system obtained from $\{q^A; q^{\overline{A}}\}$ by $\{x, y, \bar x, \bar y\}$. Again the discussion is split into three cases. In each of them the definitions of new coordinates are given. It is to be noticed that at most two of four coordinates are specified precisely. The remaining are subject to (4.39) only.

Case 1: (i) $x:=F$, $\bar x:=\overline{H}$, and (4.41) are reduced to

$$\widetilde{\Omega}_{,y}\widetilde{\Omega}_{\bar y}=(x+\bar x)^2. \tag{4.46}$$

It is remarkable that (4.46) can be integrated. Indeed the substitution of $\Theta:=(x+\bar x)^{-1}\widetilde{\Omega}$ reduces it to

$$\Theta_{,y}\Theta_{\bar y}=\tfrac{1}{2}, \tag{4.47}$$

which can be solved by the method of characteristics.[12] The solution of (4.46) is described below.

Let τ be a function of $(\varphi, x, \bar x)$ such that $\tau':=\partial\tau/\partial\varphi\neq 0$. Then

$$\widetilde{\Omega}=\sqrt{2}(x+\bar x)(\bar y/\tau'+\tau), \tag{4.48}$$

where φ is understood here as a function of $(x, y, \bar x, \bar y)$ determined by the functional equation

$$y=\bar y/2\tau'^2+\varphi. \tag{4.49}$$

Still $\widetilde{\Omega}$ is subject to (4.3), which now is reduced to

$$\begin{aligned}(x+\bar x)&(\tau''\tau_{x\bar x}-\tau_x'\tau_{\bar x}')\\&+\tau''(\tau_x+\tau_{\bar x})+\tau'(\tau_{\bar x}'-\tau_x')=0.\end{aligned} \tag{4.50}$$

We notice also that the formula for $(m^A{}_B)$ [(4.25)–(4.28)] takes the form

$$(m^A{}_B)=\begin{pmatrix}(-\widetilde{\Omega}_y)^{-1/2}, & (x+\bar x)^{-1}(-\widetilde{\Omega}_y)^{1/2}\\ -x(-\widetilde{\Omega}_y)^{-1/2}, & \dfrac{\bar x}{x+\bar x}(-\widetilde{\Omega}_y)^{1/2}\end{pmatrix}, \tag{4.51}$$

and α [Eq. (4.12)] is expressed by

$$\alpha = (\ln \tilde{\Omega}_y)_x\, dx + (\ln \tilde{\Omega}_y)_y\, dy - (\ln \tilde{\Omega}_y)_{\bar{x}}\, d\bar{x} - (\ln \tilde{\Omega}_y)_{\bar{y}}\, d\bar{y} - 2/(x+\bar{x})\, dx. \quad (4.52)$$

(ii) If $d\bar{E} = 0$, then one arrives at a contradiction with (4.3). Assuming, therefore, $d\bar{E} \neq 0$ and $x := F + \bar{H}, \bar{x} := \bar{E}$, it follows from (4.42) that

$$\tilde{\Omega}_y = -x^2/\bar{x}. \quad (4.53)$$

Next taking into account (4.3) and (4.53), one infers that

$$\tilde{\Omega} = -yx^2/\bar{x} - \bar{x}^2\bar{y}/x + w(x,\bar{x}) + \tilde{w}(\bar{x},\bar{y}). \quad (4.54)$$

The metric structure

$$ds^2 = 2w_{x\bar{x}}\, dx\, d\bar{x} + 2d(x^2y)d(-1/\bar{x}) + 2d(\bar{x}^2\bar{y})d(-1/x) \quad (4.55)$$

is that of the self-dual plane-fronted wave. Indeed, making the transformation $u = -1/\bar{x}, \zeta = -1/x, v = x^2y + w_u$, $\bar{\zeta} = \bar{x}^2\bar{y}$ and denoting $h := -2w_{uu}$ one arrives at $ds^2 = 2(du\, dv + d\zeta\, d\bar{\zeta}) + h(u,\zeta)du^2$.

We notice, however, that the self-dual plane-fronted wave can be obtained from (4.48) and (4.49) as well. For this purpose it suffices to take τ in the form of $\tau = \varphi + g(x,\bar{x})$.

Case 2: Let $d\bar{E} = 0$, and introduce $x := F$. Then (4.44) and (4.3) are contradictory. Suppose therefore that $d\bar{E} \neq 0$, and define $x, \bar{x}$ according to $x := F, \bar{x} := -\bar{E}^{-1}$. Then from (4.44) and (4.3) it follows that

$$\tilde{\Omega} = x\bar{y} + y\bar{x} + w(x,\bar{x}) + \tilde{w}(\bar{x},\bar{y}), \quad (4.56)$$

and so the metric structure is again that of a self-dual plane-fronted wave.

Case 3: From the form of (4.45) it follows that the result is exactly the same as in case 2.

The discussion above together with Lemma 4.1 and Lemma 3.3, amounts to a theorem.

Theorem 4.1: Let ds^2 be a heavenly metric structure.

(i) It admits a nontrivial ($\alpha \neq 0$) generalized canonical family of left congruences of null strings if and only if for some pair of canonical congruences Σ_1 and Σ_2 there exists an admissible coordinate transformation (4.38), (4.39) and an admissible Ω-gauge (4.40) such that Ω is determined by (4.48)–(4.50).

(ii) There exist generalized canonical families of self-dual congruences of null strings with $\alpha \neq 0$ and $d\alpha = 0$ if and only if ds^2 is a metric structure of a self-dual plane-fronted wave.

Remark: Although (ii) is a result of Lemma 3.3 one can provide also its independent proof. By forcing the one-form α [Eq. (4.52)] to be closed and at the same time $\tilde{\Omega}$ to be a solution of (4.3) and (4.46), one arrives at the required result. It can be checked also for $\tilde{\Omega}$ in the form (4.54) or (4.56), directly from (4.12). Below we present some examples.

Example 1 (self-dual plane-fronted wave): Take ds^2 in a symmetric form (4.56): $ds^2 = 2\, d\xi\, dp + 2\, d\eta\, dq + 2h(\xi,\eta)d\xi\, d\eta$. There is a natural choice for a null tetrad $e^1 = dp + h\, d\eta$, $e^2 = d\xi$, $e^3 = d\eta$, $e^4 = dq$. The only nonvanishing component of the curvature is that of the anti-self-dual Weyl spinor $C_{\dot{2}\dot{2}\dot{2}\dot{2}} = h_{\xi\eta} \neq 0$ [Eq. (1.11)]. (If this vanishes we have a flat space.) Assuming $d\alpha = 0$, one can solve (3.12) easily. Indeed, from Lemma 3.2 (ii) it follows that $\alpha_{A\dot{B}} = \alpha_A k_{\dot{B}}$, where $k_{\dot{B}}$ can be taken in the form of $k^{\dot{A}} = \delta^{\dot{A}}_{\dot{1}}$, and Eq. (3.12) is reduced to $d\alpha^B = (1/\sqrt{2})\alpha^B(\alpha^1\, d\eta + \alpha^2\, d\xi)$. Let $q_1 := \eta$ and $q_2 := \xi$. Then $\alpha = -d(\ln\phi)$, where ϕ is of the form

$$\phi = 1 + l^A q_A, \quad (4.57)$$

with l^A being a constant "spinor" or any other obtained from (4.57) by a translation $q_A \to q_A + t_A$, $t_A = \text{const}$ (this agrees with that in Ref. 10).

One can check also, that a system of equations (3.12), (3.13) does not have solutions such that $d\alpha \neq 0$.

Example 2: We present now an example of a heavenly space admitting a nontrivial generalized canonical family of self-dual congruences of null strings such that $d\alpha \neq 0$.

Take a solution of (4.5) in the form $\tau = (\sqrt{2}/3)\varphi^{3/2}$. Then from (4.48) and (4.49) $\tilde{\Omega}$ can be found:

$$\tilde{\Omega} = \frac{\sqrt{2}}{3}(x+\bar{x})\frac{y^2 + 4\bar{y} + y(y^2 - 4\bar{y})^{1/2}}{[y + (y^2 - 4\bar{y})^{1/2}]^{1/2}}. \quad (4.58)$$

Next the expression for ds^2 follows.

$$ds^2 = \frac{\sqrt{2}}{[y + (y^2 - 4\bar{y})^{1/2}]^{1/2}}\left\{2\, dx\, d\bar{y} + dy\left[(y + (y^2 - 4\bar{y})^{1/2})^{1/2}\, d\bar{x} - \frac{x+\bar{x}}{(y^2 - 4\bar{y})^{1/2}}\, d\bar{y}\right]\right\}.$$

Then working in a null tetrad gauge of the form

$$e^1 = \psi^{-1}\, dx, \quad e^2 = \psi^{-1}\, d\bar{y}, \quad e^3 = \tfrac{1}{2}\psi^{-1}\, dy,$$

$$e^4 = \psi^{-1}\left\{[y + (y^2 - 4\bar{y})^{1/2}]d\bar{x} - \frac{x+\bar{x}}{(y^2 - 4\bar{y})^{1/2}}\, d\bar{y}\right\},$$

where ψ is defined by $\psi^2 = (1/\sqrt{2})[y + (y^2 - 4\bar{y})^{1/2}]^{1/2}$, one infers that the only nonvanishing components of $C_{\dot{A}\dot{B}\dot{C}\dot{D}}$ are

$$C_{\dot{1}\dot{2}\dot{2}\dot{2}} = \frac{2\psi^2}{(y^2 - 4\bar{y})^{3/2}}$$

and

$$C_{\dot{2}\dot{2}\dot{2}\dot{2}} = \frac{2(x+\bar{x})}{(y^2 - 4\bar{y})^{5/2}}[6y + 5(y^2 - 4\bar{y})^{1/2}]\psi^2.$$

Hence the algebraic type of the conformal curvature is $[-] \times \mathrm{III}$. We notice also that α [Eq. (4.52)] is of the form

$$\alpha = -d\ln(x+\bar{x}) + \tfrac{1}{2}\ln[y + (y^2 - 4\bar{y})^{1/2}]_{,y}\, dy - \tfrac{1}{2}\ln[y + (y^2 - 4\bar{y})^{1/2}]_{,\bar{y}}\, d\bar{y},$$

and so $d\alpha \neq 0$.

5. CONCLUSIONS

The "projective extensions of heavens" do not exhaust all left conformally flat spaces. In this context it is important to have covariant means to distinguish them among all those spaces. The integrability conditions as stated in Sec. 3 provide such means.

What makes the "projective extensions of heavens" especially interesting is the fact that they admit generalized canonical families of congruences of null strings which are "very close" to nonexpanding congruences in "heavens." One can think therefore about twistor's constructions as those for heavens,[4] in which the structure of a projective twistor space is determined by the null strings organized into congruences of a generalized canonical family. In this respect it is interesting to point out that for some heavens, specified in Sec. 4, more than one twistor's construction could be available. It is to be noticed also that to describe the congruences of a generalized canonical family in a heaven one could follow the method used in Ref. 13, where nonexpanding congruences have been investigated. In fact, in Sec. 4 a null tetrad has been found in which (2.13) holds [formulas (4.8) and (4.51)]. Performing then arbitrary, constant heavenly transformations one could generate all congruences in terms of their two-forms ($\Sigma = k_A k_B S^{AB}$).

[1]J. F. Plebański and S. Hacyan, J. Math. Phys. **16**, 2403 (1975).
[2]J. F. Plebański and I. Robinson, Phys. Rev. Lett. **37**, 493 (1976).
[3]J. F. Plebański and K. Rózga (to appear).
[4]R. Penrose, Gen. Rel. Grav. **7**, 31 (1976).
[5]E. T. Newman, Gen. Rel. Grav. **7**, 107 (1976).
[6]J. F. Plebański, J. Math. Phys. **16**, 2395 (1975).
[7]J. D. Finley III and J. F. Plebański, J. Math. Phys. **17**, 585 (1976).
[8]J. F. Plebański, "Spinors, Tetrads and Forms," Instituto de Investigaciones en Matematicas Aplicadas y en Sistemas, UNAM Centro de Investigacion y Estudios Avanzados, IPN, Mexico City, 1981.
[9]L. P. Eisenhart, *Non-Riemannian Geometry* (American Math. Soc., New York, 1968), p. 15.
[10]H. W. Brinkmann, Math. Annalen **94**, 119 (1925).
[11]Ch. P. Boyer and J. F. Plebański, J. Math. Phys. **18**, 1022 (1977).
[12]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. 2.
[13]K. Rózga, Gen. Rel. Grav. **14**, 741 (1982).

On some theorems of Geroch and Stiefel

Phillip E. Parker
Department of Mathematics, The University of Iowa, Iowa City, Iowa 52242

(Received 21 July 1982; accepted for publication 4 February 1983)

Modern proofs of two theorems of Geroch (spinor spacetimes and globally hyperbolic spacetimes are parallelizable) and a theorem of Stiefel (orientable 3-manifolds are parallelizable) are given, using the computationally efficient obstruction theory of algebraic topology. These techniques also easily show that, in fact, Geroch's second theorem is a corollary of Stiefel's theorem.

PACS numbers: 04.20.Cv, 02.40.Re

I. INTRODUCTION

One of the fundamental theorems in the development of differential topology is that of Stiefel[1]: every compact orientable 3-manifold is parallelizable. In general relativity, Geroch[2] proved two similar theorems: every spinor spacetime and every orientable globally hyperbolic spacetime is parallelizable. As Geroch pointed out, due to a peculiarity of 4-manifolds, his first theorem actually shows that every noncompact spin 4-manifold is parallelizable. In this paper, I give a unified exposition of these theorems using modern obstruction theory, and show that Geroch's second theorem is a corollary of Stiefel's theorem, as extended to possibly noncompact 3-manifolds.

That a modern treatment might be desirable was pointed out to be by several people, including a few physicists, who had encountered some difficulties in trying to understand the original proofs. Stiefel, in fact, introduced his famous characteristic class to prove his theorem, and did not have any obstruction theory available; he was creating it. Geroch's proofs use obstruction theory, but the modern form of the theory was only beginning to take shape at that time.

The background necessary for this paper consists of basic algebraic topology and some knowledge of classifying spaces for vector bundles. At present, there is no one source for all this, but the present author and C. T. J. Dodson are currently preparing an introductory treatment for physicists and other nonspecialists. Our attitude is that one might as well have the best available apparatus for computations, whether or not one proves all the big theorems. In this spirit, Sec. 1 contains the necessary theoretical equipment for making the computations, which are then carried out in Secs. 2 and 3.

Notation and terminology are standard; see Refs. 3–5. For example, $\pi_k(X)$ denotes the k th homotopy group of X with respect to the fixed base point (see Sec. 1), and $H^k(X;\pi)$ denotes the k th cohomology group of X with coefficients in the (abelian if $k>1$) group π. As the notation suggests, our coefficients will frequently be homotopy groups. The classifying space for the group G is denoted by BG.

II. SOME OBSTRUCTION THEORY

All spaces throughout this paper are assumed to be CW complexes. We shall also assume that some 0-cell has been distinguished as the base point. All maps are assumed to be base-point preserving.

Lemma 1.1: Given a space X and an integer $n\geqslant 0$, there exists a space $X^{[n]}$, the nth *Postnikov section*, $\iota_n: X\to X^{[n]}$, and $\rho_n: X^{[n]}\to X^{[n-1]}$ for $n\geqslant 1$, such that

(1) $(X^{[n]}, X)$ is a relative CW complex with cells in dimensions $\geqslant n+2$;

(2) $\pi_k(X^{[n]})=0$ for $k\geqslant n$;

(3) $\iota_{n\#}: \pi_k(X)\mapsto \pi_k(X^{[n]})$ is an isomorphism for $k\geqslant n$;

(4) ρ_n is a fibration with fiber $K(\pi_n(X), n)$.

For a proof, see, for exmaple, Ref. 3.

Recall that $K(\pi, n)$ is the unique (up to homotopy type) space with $\pi_n(K(\pi,n))=\pi$ and $\pi_k(K(\pi,n))=0$ for $k\neq n$. The spaces and maps $\{X^{[n]},\rho_n\}$ fit together to form the *Postnikov tower* of X. It is not difficult to show that $X\cong \varprojlim X^{[n]}$, where $\cong$ denotes homotopy equivalence, so the $X^{[n]}$ can be considered as approximations to X. It is convenient to think of them as "dual" to the n-skeleta of X.

There is a generalization of Postnikov towers to fibrations, the *Moore–Postnikov decomposition*, which is used in studying the existence of liftings.

Lemma 1.2: If $p: E\to B$ is a fibration with fiber F and B connected, then there exist fibrations $p_n: E^{[n]}\to E^{[n-1]}$ for all $n\geqslant 1$ and maps $h_n: E\to E^{[n]}$ with $p_n h_n = h_{n-1}$ such that

(1) $E^{[0]}=B$ and $h_0=p$;

(2) the fiber of p_n is $K(\pi_n(F), n)$;

(3) $h_{n\#}: \pi_k(E)\to\pi_k(E^{[n]})$ is an isomorphism for $k<n$;

(4) $q_n := p_1p_2\cdots p_n: E^{[n]}\to B$ induces isomorphisms,

$$\pi_k(E^{[n]})\to\pi_k(B)\quad\text{for}\quad k>n;$$

(5) if $F^{[n]}$ is the fiber of q_n, then $\{h_n\mid F: F\to F^{[n]}\}$ defines a Postnikov tower of F.

For a proof, see, for example, Refs. 3 and 6.

There is also the relation of $K(\pi, n)$ to reduced ordinary cohomology.

Lemma 1.3: $\tilde H^n(X;\pi)\cong [X, K(\pi,n)]$ and this isomorphism is natural. Here [,] denotes homotopy classes of maps. Recall that $\tilde H^n = H^n$ for $n\geqslant 1$ and that $\operatorname{rank}(\tilde H^0) = \operatorname{rank}(H^0)-1$.

Now consider the problem: Is $f: X\to Y$ inessential (i.e., homotopic to the constant map)? We shall say that f is n-*trivial* iff $\iota_n f: X\to Y^{[n]}$ is inessential.

Lemma 1.4: If $\dim X = n<\infty$, then f is n-trivial iff f is inessential. If f is $(n-1)$-trivial, there is defined $\mathscr{O}_n(f)\subseteq \tilde H^n(X;\pi_n(Y))$, the *n-dimensional obstruction*; f is n-trivial iff $0\in\mathscr{O}_n(f)$. See Ref. 3, p. 165f for a proof.

We shall also need the *first Postnikov invariant* of a fibration. In the Moore–Postnikov decomposition of Lemma 1.2, let X be a space and consider the map $[X, E] \rightarrow [X, E^{[n]}]$ induced by h_n. It follows from (3) of Lemma 1.2 that this map is bijective if dim $X \leqslant n$. Thus in this dimension range a map $X \rightarrow B$ lifts to $X \rightarrow E$ iff it lifts to $X \rightarrow E^{[n]}$. Suppose now that the first nonzero homotopy group of F is $\pi_{n_1}(F)$. By the universal coefficient theorem,[5] the Kronecker index $\langle\ ,\ \rangle$ provides an isomorphism $H^{n_1}(F, \pi) \rightarrow \mathrm{Hom}(H_{n_1}(F), \pi)$ for abelian π. Let $h: \pi_{n_1}(F) \rightarrow H_{n_1}(F)$ be the Hurewicz isomorphism.[3] An element $v_1 \in H^{n_1}(F; \pi_{n_1}(F))$ such that $\langle v_1, \ \rangle = h^{-1} \in \mathrm{Hom}(\pi_{n_1}(F), H_{n_1}(F))$ is called a *fundamental class* of F.

To make use of fundamental classes, we shall need the *transgression* $\tau: D \rightarrow H^{n+1}(B; \pi)$, where $D \subseteq H^n(F; \pi)$. For a complete discussion, see [Ref. 7, p. 132ff]; other than the fact that it is a homomorphism as indicated, the only thing we shall need is the *Serre exact sequence*.

Lemma 1.5: If $F \rightarrow E \rightarrow B$ is a fibration with F and B connected, if $\tilde{H}_i(F; \pi) = 0$ for $i \leqslant t$, and if $\tilde{H}_i(B; \pi) = 0$ for $i \leqslant s$, then for $n \leqslant s + t - 1$ the following sequence is exact:

$$\cdots \rightarrow H^{n-1}(E; \pi)$$
$$\rightarrow H^{n-1}(F; \pi) \xrightarrow{\tau} H^n(B; \pi) \rightarrow H^n(E; \pi) \cdots .$$

Here and in what follows we assume that $\pi_1(B)$ acts trivially on $H^*(F; \pi)$ i.e., that the fibration is *orientable* with respect to the coefficients π. This will always be the case in our applications and implies that a fundamental class v is *transgressive*; i.e., lies in D.

Returning now to a fundamental class $v_1 \in H^{n_1}(F; \pi_{n_1}(F))$ as above, define the first *Postnikov invariant*

$$k^1 := \tau(v_1) \in H^{n_1+1}(B; \pi_{n_1}(F)).$$

Consider $k_1(f) := f^*(k^1) \in H^{n_1+1}(X; \pi_{n_1}(F))$.

Lemma 1.6: f lifts to $E^{[n_1]}$ iff $k_1(f) = 0$. There are higher order Postnikov invariants k^i, and corresponding $k_i(f)$ which are *sets* of cohomology classes. One can show that, up to sign conventions, k^1 is the classical obstruction to extending a section to the $(n_1 + 1)$-skeleton. See Ref. 6 for details. All we need is

Lemma 1.7: If $k_1(f) = 0$ and if $0 \in k_i(f)$ for $i \geqslant 2$, then f lifts to E.

III. SPIN 4-MANIFOLDS AND PARALLELIZABILITY

Let X be a (smooth, paracompact) 4-manifold. Recall that (the isomorphism class of) the tangent bundle TX is determined by its *classifying map* $T: X \rightarrow \mathrm{BO}(4)$, where BO(4) is the *classifying space* for vector bundles of fiber dimension 4; similarly, there are classifying spaces BSO(4) for oriented bundles and BSpin(4) for spin bundles.[8] Thus X is orientable iff T lifts to BSO(4) and further admits a spin structure iff T lifts to BSpin(4). Recall that $H^*(\mathrm{BO}(4); Z_2) \cong Z_2[w_1, w_2, w_3, w_4]$, $H^*(\mathrm{BSO}(4); Z_2) \cong Z_2[w_2, w_3, w_4]$, and $H^*(\mathrm{BSpin}(4); Z_2) \cong Z_2[w_4]$, where w_i is the ith universal Stiefel–Whitney class and the notation indicates polynomial algebras in certain of these classes.[4]

Theorem 2.1: A 4-manifold X is orientable iff $w_1(X) = 0$ and further admits a spin structure iff $w_2(X) = 0$.

Proof: To be orientable, there must be a lifting of T from BO(4) to BSO(4). We have a fibration $\mathrm{O}(1) \rightarrow \mathrm{BSO}(4) \rightarrow \mathrm{BO}(4)$. Now O(1) is a $K(Z_2, 0)$, and one may check that $\tau: \tilde{H}^0(\mathrm{O}(1); Z_2) \rightarrow H^1(\mathrm{BO}(4); Z_2)$ is an isomorphism so $k^1 = w_1$. Then $k_1(T) = w_1(X)$.

If now X is orientable, then we may assume $T: X \rightarrow \mathrm{BSO}(4)$. Recall that we have an exact sequence $1 \rightarrow \mathrm{O}(1) \rightarrow \mathrm{Spin}(4) \rightarrow \mathrm{SO}(4) \rightarrow 1$, which gives rise to a fibration $\mathrm{BO}(1) \rightarrow \mathrm{BSpin}(4) \rightarrow \mathrm{BSO}(4)$, and that BO(1) is a $K(Z_2, 1)$. Now $H^1(\mathrm{BSpin}(4); Z_2) = H^2(\mathrm{BSpin}(4); Z_2) = 0$ so by Lemma 1.5, $\tau: H^1(\mathrm{BO}(1); Z_2) \rightarrow H^2(\mathrm{BSO}(4); Z_2)$ is an isomorphism. Thus $k^1 = w_2$ and $k_1(T) = w_2(X)$.

In either case, the conclusion now follows via Lemma 1.7 and the fact that the existence of a lifting implies the vanishing of the relevant classes.

Theorem 2.2: A noncompact spin 4-manifold is parallelizable.

Proof: Let $T: X \rightarrow \mathrm{BSpin}(4)$ be the classifying map as before. Recall that the first nonzero homotopy group is $\pi_4(\mathrm{BSpin}(4)) \cong Z \oplus Z$. But X is noncompact so $H^4(X; Z \oplus Z) = 0$, whence $0 \in \mathscr{O}_n(T)$ for $1 \leqslant n \leqslant 4$ and by Lemma 1.4, T is inessential. Therefore, X is parallelizable.

Corollary 2.3 (Geroch[2]): A noncompact 4-manifold which is orientable admits a spin structure iff it is parallelizable.

Geroch was actually looking at noncompact Lorentzian 4-manifolds X which were time and space orientable, so that $T: X \rightarrow L(4)$, the identity component of the Lorentz group. Instead of Spin(4), one considers SL(2,$\mathbb{C}$) and refers to *spinor* structures rather than *spin* structures. Here it is also true that the first nonzero homotopy group is $\pi_4(\mathrm{BSL}(2, \mathbb{C})) \cong Z$, so the argument of Theorem 2.2 still applies. Letting $\cong$ denote homeomorphism, $\mathrm{SL}(2, \mathbb{C}) \cong \mathrm{Spin}(3) \times \mathbb{R}^3$ and $\mathrm{Spin}(4) \cong \mathrm{Spin}(3) \times S^3$. It follows that for noncompact 4-manifolds, there is a bijective correspondence between spin and spinor structures, which fact was pointed out by Geroch (Ref. 2, p. 1740).

IV. ORIENTABLE 3-MANIFOLDS AND GLOBALLY HYPERBOLIC SPACETIMES

We shall need the fact that if X is a 3-manifold, then $w_2(X) = w_1(X)^2$. For compact X, one may show this using Wu's formula (Ref. 4, p. 132). For noncompact X, the same argument goes through provided that singular homology with finite chains (the usual kind) is replaced by singular homology with infinite chains. Since Poincaré duality is used to establish Wu's formula, it must also be redone. But it is easy to check that one need only replace compactly supported cohomology with ordinary cohomology and ordinary homology with infinite chain homology in A9 (Ref. 4, p. 278), and then "dualize" their proof.

Theorem 3.1 (Stiefel[1]): An orientable 3-manifold X is parallelizable.

Proof: Since X is orientable, it will suffice to produce a global 2-frame; i.e., a section of the bundle with fiber $V_2(\mathbb{R}^3)$, the set of 2-frames in $\mathbb{R}^3$. Indeed, this splits TX into an orientable plane bundle plus a line bundle which is thus also orientable, hence trivial.

Let $T: X \rightarrow \mathrm{BSO}(3)$ be the classifying map for the tangent bundle. Now $\pi_1(V_2(\mathbb{R}^3)) \cong Z_2$ and this diagram commutes [recalling $V_2(\mathbb{R}^3) \cong \mathrm{SO}(3)$],

$$\begin{tikzcd} H^1(\mathrm{SO}(3); Z_2) \arrow[d, "1"'] & H^2(\mathrm{BSO}(3); Z_2) \arrow[l, "\tau"'] \arrow[d, "T^*"] \\ H^1(V_2(\mathbb{R}^3); Z_2) \arrow[r] & H^2(X; Z_2) \end{tikzcd}$$

so $k_1(T) = w_2(X) = w_1(X)^2 = 0$. Since $\pi_2(V_2(\mathbb{R}^3)) = 0$, it follows by Lemma 1.7 that a lifting of 1_X, or a global 2-frame, exists.

Corollary 3.2 (Geroch[2]): An orientable globally hyperbolic 4-manifold is parallelizable.

Proof: We know that $X \cong \mathbb{R} \times S$, where S is an orientable (since X is) 3-manifold. By Theorem 3.1, S is parallelizable; hence so is X.

ACKNOWLEDGMENT

I would like to thank the members of the Geometry Seminar at the University of Missouri, J. Beem, B. DeFacio, P. Ehrlich, M. England, M. Funderburg, T. Powell, and D. Retzloff for their enthusiastic encouragement.

[1] E. Stiefel, "Richtungsfelder und Fernparallelismus in Mannigfaltigkeiten," Comment. Math. Helv. **8**, 3–51 (1936).

[2] R. Geroch, "Spinor structure of space-times in general relativity," J. Math. Phys. **9**, 1739–1744 (1968).

[3] B. Gray, *Homotopy Theory* (Academic, New York, 1975).

[4] J. Milnor and J. Stasheff, *Characteristic Classes* (Princeton U.P., Princeton, N.J., 1974).

[5] J. Vick, *Homology Theory* (Academic, New York, 1973).

[6] E. Thomas, *Seminar on Fiber Spaces* (Springer-Verlag, Berlin, 1966). p. 13.

[7] A. Borel, "Sur la cohomologie des espaces fibres principaux...," Ann. Math. **57**, 115–207 (1953).

[8] D. Husemoller, *Fiber Bundles* (McGraw-Hill, New York, 1966).

Isometries compatible with gravitational radiation

J. Bičák[a] and B. G. Schmidt
Max-Planck-Institut für Physik und Astrophysik, 8046 Garching bei München, Federal Republic of Germany

(Received 4 October 1982; accepted for publication 25 February 1983)

Isometries compatible with asymptotic flatness and admitting radiation are studied by using Bondi's formalism. In axially symmetric space-times, the only second allowable symmetry that does not exclude radiation is boost symmetry. The boost-rotation symmetric solutions describe "uniformly accelerated particles" of various kinds. The news function is restricted by a differential equation; however, it need not vanish, as has been claimed in the literature. If two Killing fields corresponding to null rotations at null infinity are present, then it is shown that the vacuum field equations imply a further isometry. The resulting space-time is a plane wave.

PACS numbers: 04.20.Cv, 04.20.Jb, 04.30. + x

1. INTRODUCTION AND SUMMARY

It is of interest to know which isometries are compatible with asymptotic flatness when radiation is present.[1,2] Berezdivin and Herrera[3,4] claim that "in an axially symmetric asymptotically flat vacuum space-time, the existence of another isometry causes all possible solutions to be non-radiative." Work by Bičák,[5] the example of the C-metric,[6,7] and more general solutions[8] are in striking contradiction to this claim. The main purpose of this paper is to resolve this contradiction and to demonstrate that if, in addition to the axial symmetry, another isometry is assumed in an asymptotically flat space-time, then this has to be a boost symmetry in order that radiation may exist. The boost-rotation symmetry is present in all the examples mentioned above. What we mean precisely by an "asymptotically flat space-time" is explained at the beginning of Sec. 2 [below Eqs. (2)–(7)]; our assumptions imply that at least "the piece" of the null bondary $\mathscr{I}^+$, as defined in Ref. 2, exists.

Rather than employing quantities defined directly on $\mathscr{I}^+$, as the authors of Refs. 1 and 2 do, Berezdivin and Herrera[3] analyze the asymptotic form of the Killing equations in space-time, extending the original work of Sachs.[9] By using this method we show in Sec. 2 that in an axially symmetric space-time with at least a local $\mathscr{I}^+$, another Killing vector has to be either a translation—which then (Sec. 3) implies that the Bondi news function[10] vanishes (i.e., the absence of radiation)—or it has to be a boost Killing field. In the case of boost symmetry, the news function is only restricted by a differential equation,but it need not vanish. The work of Berezdivin and Herrera[3,4] contains errors in the asymptotic expansions of the Killing equations which lead to the incorrect conclusion that the news function vanishes in an axially symmetric space-time when another isometry is present. Moreover, Berezdivin and Herrera do not find the solution of the Killing equations—even in the leading terms in the expansions in powers of a luminosity distance—so that the character of the additional isometry remained undetermined in their work.

Space-times admitting boost-rotation symmetries are described by metrics[11,5,8]

$$g = -e^{\lambda}\, d\rho^2 - \rho^2 e^{-\mu}\, d\phi^2 + (z^2 - t^2)^{-1}[(z^2e^{\mu} - t^2e^{\lambda})\, dt^2 - (z^2e^{\lambda} - t^2e^{\mu})\, dz^2 + 2zt\,(e^{\lambda} - e^{\mu})\, dz\, dt\,], \tag{1}$$

where $\mu(\rho^2, z^2 - t^2)$ is a solution of the flat space wave equation and λ is determined in terms of μ by quadrature; in Minkowski space, $\{t,\rho,z,\phi\}$ are cylindrical coordinates. Solutions exist admitting $\mathscr{I}^+$.[8] The detailed analysis of boost-rotation symmetric space-times will be given elsewhere; some new solutions of this type are contained in Ref. 12. Here (in Sec. 3) we shall just write down the news function which can be inferred from Ref. 5. Although that work was motivated by specific boost-rotation symmetric solutions of Bonnor and Swaminarayan,[11] the concrete form of functions μ and λ in (1) was not used during the derivation of the news function. This news function is shown to satisfy the differential equation implied by the existence of the boost Killing vector field.

If one assumes that Bondi's form of the metric takes values for all azimuthal angles ϕ but only for some interval of lattitudes θ (i.e., not necessarily on a whole sphere), $\mathscr{I}^+$ has only a "local" character and it need not even have the topology $S^2\times R$. In Refs. 1 and 2, a general, detailed study of symmetries compatible with $\mathscr{I}^+$ and radiation is given (without the assumption of axial symmetry). Most of theorems proved there assume that $\mathscr{I}^+$ is toplogically $S^2\times R$. Then the Lie algebra $\mathscr{L}$ of symmetries is a subalgebra of the Bondi–Metzner–Sachs Lie algebra, and $\mathscr{L}/\tau$ (where τ is the space of translational Killing fields) must be a Lie subalgebra of the Lorentz Lie algebra.[1,2] All three-parameter subgroups of the Lorentz group not treated in Refs. 1 and 2 contain an abelian two-parameter group of null rotations.[13] In Sec. 4 we show that this case leads only to a plane wave.

The case of a "local $\mathscr{I}^+$" with a preferred rotational, hypersurface orthogonal Killing vector is exceptional. In Bondi *et al.*'s paper[10] it is established, that in this case the asymptotic symmetry group is a subgroup of the BMS group. This is done purely locally, i.e., the whole S^2 is never used. The geometrical reason for this possibility is that conformal rescalings of $d\theta^2 + \sin^2\theta\, d\phi^2$, which preserve $\partial/\partial\phi$ as a Killing vector, are of the form $k(\theta)$, where $k(\theta)$ is determined by a boost. Hence relative to a chosen $\partial/\partial\phi$ Killing vector, Bondi news and translations have an invariant meaning even locally.

Let us also note that, assuming Bondi's metric and expansions, we exclude the cylindrical case (cf. Ref. 14 for the

[a] Permanent address: Department of Mathematical Physics, Faculty of Mathematics and Physics, Charles University, V. Holešovičkách 2, 18000 Prague 8, Czechoslovakia.

 0022-2488/84/030600-07$02.50

modifications of Bondi's method for cylindrically symmetric metrics).

The results of Refs. 1 and 2 and of this paper show that (1) axisymmetry generated by a hypersurface orthogonal rotational Killing vector, (2) existence of at least a piece of $\mathscr{I}^+$, (3) presence of radiation, and (4) the assumption of a further Killing field lead uniquely to the boost-rotation symmetry. We conjecture that the same is true also in the case when the rotational Killing vector field is not hypersurface orthogonal. An example of an exact solution with this property is available: the *C*-metric with rotation.[15]

If $\mathscr{I}^+$ has topology $S^2 \times R$, the only remaining case of a two-parameter group is a nonabelian group acting on $\mathscr{I}^+$ as a null rotation and a boost. Nothing seems to be known about this case. If it turns out to lead to flat space-time (as in the case of two null rotations), then the boost-rotation symmetric solutions will probably long remain as the only examples of exact radiative solutions available for testing both the general theory of the asymptotic structure of space-time and various approximation methods.

2. AXIALLY SYMMETRIC SPACE-TIMES WITH ANOTHER ISOMETRY

Consider an axially symmetric space-time with circular group orbits; denote the corresponding Killing vector field by $\partial/\partial\phi$. We assume this vector field to be hypersurface orthogonal. Assuming then that at least the "piece of $\mathscr{I}^+$" (as defined in Ref. 2) exists, one can introduce Bondi's coordinate system $\{u,r,\theta,\phi\} \equiv \{x^0,x^1,x^2,x^3\}$, and the metric satisfying vacuum field equations can be written in the form[10]

$$ds^2 = (Vr^{-1}e^{2\beta} - U^2r^2e^{2\gamma})\,du^2 + 2e^{2\beta}\,du\,dr + 2Ur^2e^{2\gamma}\,du\,d\theta - r^2(e^{2\gamma}\,d\theta^2 + e^{-2\gamma}\sin^2\theta\,d\phi^2). \quad (2)$$

Functions entering the metric have the following asymptotic forms at $r \to \infty$ (commas denote partial derivatives):

$$\gamma = c(u,\theta)r^{-1} + O(r^{-3}), \quad (3)$$

$$U = -(c_{,\theta} + 2c\cot\theta)r^{-2} + O(r^{-3}), \quad (4)$$

$$V = r - 2M(u,\theta) + O(r^{-1}), \quad (5)$$

$$\beta = -\tfrac{1}{4}c^2r^{-2} + O(r^{-4}). \quad (6)$$

The mass aspect $M(u,\theta)$ is connected with the news function $c_{,u}$ by the relation

$$M_{,u} = -c_{,u}^2 + \tfrac{1}{2}(c_{,\theta\theta} + 3c_{,\theta}\cot\theta - 2c)_{,u}. \quad (7)$$

However, we shall *not* assume that the space-time admits $\mathscr{I}^+$ with topology $S^2 \times R$. We assume that Eqs. (2)–(7) are valid for all $\phi \in [0,2\pi)$ but only in some open interval of θ, i.e., not necessarily on the whole sphere. In particular, the "axis of symmetry" ($\theta = 0,\pi$) may be singular; thus, the regularity conditions that

$$V,\quad \beta,\quad U/\sin\theta,\quad \gamma/\sin^2\theta \quad (8)$$

are regular functions of $\cos\theta$ at $\cos\theta = \pm 1$ need not be satisfied for any retarded time u. If, for example, Eqs. (2)–(7) were satisfied for all u and all θ except for $\theta = 0,\pi$, then only "local" $\mathscr{I}^+$ as defined in Ref. 2 would exist, because two generators of $\mathscr{I}^+$ would be missing, and $\mathscr{I}^+$ therefore would not even be topologically $S^2 \times R$.

A simple example of such a space-time is the special case of the metric of Bonnor and Swaminarayan[11] describing two "particles," symmetrically located with respect to $z = 0$, and "uniformly accelerated" along the z axis. In general, stresses (usually referred to as "conical singularities") extending *both* to infinity *and* between the particles exist (Fig. 1). The metric is of the form (1) with specific functions μ and λ; however, these are determined up to additive constants, and the above situation arises with a general, arbitrary choice of these constants.

Let us now assume that another Killing vector field η exists in space-time and forms together with $\xi = \partial/\partial\phi$ a two-parameter group. Consequently, their commutator is of the form

$$[\eta,\xi] = a\eta + b\xi, \quad (9)$$

where a,b are constants. It is then easy to prove the following:

Lemma: Let ξ be a vector field with circles as integral curves; η another vector field such that $[\eta,\xi] = a\eta + b\xi$, a and b constants. Then ξ and η determine an abelian Lie algebra.

Proof: If we choose coordinates adapted to ξ such that $\xi^\alpha = (0,0,0,1)$, then (9) implies

$$-\eta^i{}_{,\phi} = a\eta^i \quad \text{for } i = 0,1,2$$

and

$$-\eta^3{}_{,\phi} = a\eta^3 + b.$$

Integrating, we find

$$\eta^i = \tilde{\eta}^i(x^j)e^{-a\phi}, \quad \eta^3 = \tilde{\eta}^3(x^j)e^{-a\phi} - b/a.$$

Since we can always consider the Killing vector field $\eta^\alpha + (b/a)\xi^\alpha$ instead of the original η^α, we may put $b = 0$ without loss of generality so that

$$\eta^\alpha = \tilde{\eta}^\alpha(x^j)e^{-a\phi}. \quad (10)$$

However, since axial symmetry requires $\eta^\alpha(\phi) = \eta^\alpha(\phi + 2\pi)$, the constant a must be zero.

Therefore, we assume $[\eta,\xi] = 0$ and in Bondi's coordinates (which are adapted to ξ) the components η^α are independent of ϕ.

Introduce the standard null tetrad field[16] $\{k^\alpha, m^\alpha, t^\alpha, \bar{t}^\alpha\}$ (with bar denoting complex conjugation), where

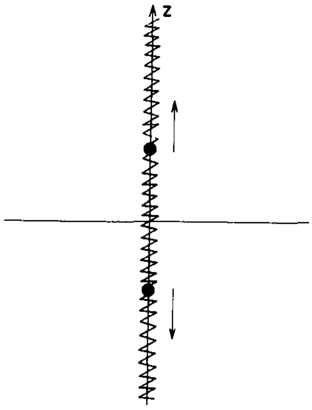


FIG. 1. Two particles, symmetrically located with respect to $z = 0$ and uniformly accelerated along the z axis. Stresses (conical singularities) extending both to infinity and between the particles exist in general; two generators of $\mathscr{I}^+$ are then missing.

$$k_\alpha = \partial_\alpha u, \quad m^\alpha k_\alpha = 1, \quad m^\alpha m_\alpha = 0.$$

The complex vector (subscripts R and I denoting the real and imaginary parts)

$$t^\alpha = t^\alpha_R + i t^\alpha_I$$

satisfies $t^\alpha \bar{t}_\alpha = -1$, $t^\alpha t_\alpha = t^\alpha k_\alpha = t^\alpha m_\alpha = 0$. A convenient choice of the tetrad vectors is given by

$$\begin{aligned} &k_\alpha = [1,0,0,0], \quad k^\alpha = [0, e^{-2\beta}, 0, 0], \\ &m_\alpha [(V/2r)e^{2\beta}, e^{2\beta}, 0, 0], \quad m^\alpha = [1, -V/2r, U, 0], \\ &t_\alpha = [\tfrac{1}{2}(1+i)Ure^\gamma, 0, -\tfrac{1}{2}(1+i)re^\gamma, -\tfrac{1}{2}(1-i)r\sin\theta\, e^{-\gamma}], \\ &t^\alpha = [0,0,\tfrac{1}{2}(1+i)e^{-\gamma}/r, \tfrac{1}{2}(1-i)e^\gamma/r\sin\theta]. \end{aligned} \tag{11}$$

It can be easily checked that $g_{\alpha\beta} = 2k_{(\alpha}m_{\beta)} - 2\bar{t}_{(\alpha}t_{\beta)}$, where $g_{\alpha\beta}$ is given by (2).

Now decompose the Killing vector field η^α into this tetrad[3]

$$\eta^\alpha = Ak^\alpha + Bm^\alpha + Ct^\alpha + \bar{C}\bar{t}^\alpha; \tag{12}$$

here, $A,B,C = C_R + iC_I$ are general functions of u,r,θ. One can write the Killing equations

$$L_\eta g_{\alpha\beta} = 0 \tag{13}$$

explicitly by inserting (12) for η^α and using the simple properties of the Lie derivative ($L_{x+y} = L_x + L_y$ and $L_{\alpha x} g_{\mu\nu} = \alpha L_x g_{\mu\nu} + 2x_{(\mu}\alpha_{,\nu)}$, x and y vector fields, α = a scalar function). Defining functions

$$f = C_R - C_I, \quad g = C_R + C_I, \tag{14}$$

one can rewrite the Killing equations (13) as follows:

$$\begin{aligned} 0 = L_\eta g_{\alpha\beta} &= AL_k g_{\alpha\beta} + BL_m g_{\alpha\beta} + 2A_{,(\alpha}k_{\beta)} + 2B_{,(\alpha}m_{\beta)} \\ &+ f[(L_t g_{\alpha\beta})_R + (L_t g_{\alpha\beta})_I] + 2f_{,(\alpha}[t_{R\beta)} + t_{I\beta)}] \\ &+ g[(L_t g_{\alpha\beta})_R - (L_t g_{\alpha\beta})_I] + 2g_{,(\alpha}[t_{R\beta)} - t_{I\beta)}]. \end{aligned} \tag{15}$$

All expressions $L_k\, g_{\alpha\beta}$, $L_m g_{\alpha\beta}$, and $L_t g_{\alpha\beta}$, with $g_{\alpha\beta}$ given by (2), are contained in the Appendix of the paper by Berezdivin and Herrera.[3] In our appendix we give corrections to their expressions; and in the same appendix we write down explicitly all the Killing equations (13) because (half of) these equations as given in Ref. 3 contain errors. In the Appendix we also allow ϕ-dependent η^α. However, as consequence of our Lemma, we assume here $\eta^\alpha_{,\phi} = 0$.

Inspecting now the expressions for $L_\eta\, g_{\alpha\beta}$ given in the Appendix, we observe that equations $L_\eta g_{03} = L_\eta g_{13} = L_\eta g_{23} = 0$ can be rewritten as

$$(ge^\gamma/r\sin\theta)_{,u} = (ge^\gamma/r\sin\theta)_{,r} = (ge^\gamma/r\sin\theta)_{,\theta} = 0 \tag{16}$$

so that $g = \text{const}\, r\sin\theta\, e^{-\gamma}$. Then (11) and (12) imply that the contribution of g to the vector field η^α is just $\eta^\phi = \text{const}$, which is a constant multiple of the axial Killing vector $\partial/\partial\phi$. Without loss of generality, we may thus put

$$g = 0. \tag{17}$$

Further, we see that $L_\eta g_{11} = 0$ immediately implies that the function B is indepenent of r:

$$B = B(u,\theta). \tag{18}$$

The other Killing equations are much more complicated, and we shall solve them asymptotically only by assuming that the functions A and f can be expanded in powers of r. Denote by $A^{(k)}$ and $f^{(k)}$ the coefficients of r^{-k} in the expansions. By examining the asymptotic forms of functions (3)–(6) entering the metric (2) and the Killing equation $L_\eta g_{12} = 0$ in the terms proportional to r^k with $k \geqslant 2$, we find out that $f^{(-k)} = 0$; similarly, the equation $L_\eta g_{22} = 0$ then implies $A^{(-k)} = 0$ if $k \geqslant 2$. Hence, the asymptotic forms of A and f at $r \to \infty$ are

$$\begin{aligned} A &= A^{(-1)}r + A^{(0)} + A^{(1)}r^{-1} + O(r^{-2}), \\ f &= f^{(-1)}r + f^{(0)} + f^{(1)}r^{-1} + O(r^{-2}), \end{aligned} \tag{19}$$

where $A^{(i)}$ and $f^{(i)}$ are functions of u and θ.

Now after substituting these expansions into the left-hand sides of the Killing equations given in the Appendix, taking into account (17) and (18) and the expansions of the metric functions (3)–(6), we find out that, in the leading orders in powers of r, the Killing equations lead to constraints on the leading terms in A and f but do not restrict any of metric functions (3)–(6); in particular, the news function $c_{,u}$ remains arbitrary. The powers of r appearing in the first nonvanishing terms on the "left-hand sides" of the Killing equations are different for different equations and are given in the brackets. The resulting system of equations has the following simple form:

$$L_\eta g_{00} = 0 \quad (r^1): \quad A^{(-1)}_{,u} = 0, \tag{20}$$

$$L_\eta g_{01} = 0 \quad (r^0): \quad B_{,u} + A^{(-1)} = 0, \tag{21}$$

$$L_\eta g_{02} = 0 \quad (r^2): \quad f^{(-1)}_{,u} = 0, \tag{22}$$

$$L_\eta g_{22} = 0 \quad (r^2): \quad f^{(-1)}_{,\theta} + A^{(-1)} = 0, \tag{23}$$

$$L_\eta g_{33} = 0 \quad (r^2): \quad f^{(-1)}\cos\theta + A^{(-1)}\sin\theta = 0. \tag{24}$$

In the equation $L_\eta g_{12} = 0$, the first nonzero terms appearing at r^1 cancel out; the other equations are satisfied in all orders as a consequence of (17) and (18). Now Eqs. (20) and (21) imply that $A^{(-1)}$ and $f^{(-1)}$ are functions of θ only. Applying $\partial/\partial\theta$ to (24) and using (23), we find a simple differential equation for $A^{(-1)}$,

$$A^{(-1)}_{,\theta}\cos\theta + A^{(-1)}\sin\theta = 0,$$

the solution of which is

$$A^{(-1)} = k\cos\theta, \quad k = \text{const}. \tag{25}$$

Then (23) gives

$$f^{(-1)} = -k\sin\theta \tag{26}$$

and (21) leads to

$$B = -ku\cos\theta + \alpha(\theta), \tag{27}$$

where α is an arbitrary function of θ. Therefore, (25)–(27) together with (17) and (11), (12), (14) imply that the Killing vector field η^α corresponding to another isometry in an axially symmetric space-time has to have an asymptotic form

$$\eta^\alpha = [-ku\cos\theta + \alpha(\theta),\ kr\cos\theta,\ -k\sin\theta,\ 0]. \tag{28}$$

Now, when $k = 0$, we obtain the vector field that generates supertranslations.[9] As we shall show in the following section, in the case $k = 0$, this vector field is in fact the generator of translations. The news functions then vanishes, and the

space-time is nonradiative. So assume $k \neq 0$. Then we can find a (Bondi) coordinate system in which $\alpha = 0$. Indeed, suppose we make a supertranslation $\tilde{u} = u + \hat{\alpha}(\theta)$, $\tilde{r} = r$, $\tilde{\theta} = \theta$, $\tilde{\phi} = \phi$. This implies $\eta^{\tilde{r}} = \eta^r$, $\eta^{\tilde{\theta}} = \eta^\theta$, $\eta^{\tilde{\phi}} = \eta^\phi$, and $\eta^{\tilde{u}} = \eta^u + \hat{\alpha}_{,\theta}\eta^\theta = -k\cos\theta\,(\tilde{u} - \hat{\alpha}) + \alpha - \hat{\alpha}_{,\theta}\, k \sin\theta$.
If the function $\hat{\alpha}(\theta)$ satisfies the differential equation

$$-\hat{\alpha}_{,\theta} k \sin\theta + \hat{\alpha} k \cos\theta = -\alpha, \tag{29}$$

then we obtain $\eta^{\tilde{u}} = -k\tilde{u}\cos\theta$. The solution of Eq. (29) is

$$\hat{\alpha} = k\sin\theta + \sin\theta \int \frac{\alpha}{\sin^2\theta}\, d\theta,$$

so that requiring η^α to be bounded for all θ, i.e., $\alpha(\theta)$ to be bounded, we can find the function $\hat{\alpha}$ even globally on S^2.

Hence, we put $\alpha = 0$ in (28) and, clearly, without any loss of generality one may choose $k = 1$. We conclude that in an axially symmetric space-time admitting another isometry (which, as we shall see in the next section, does not exclude radiation), the asymptotic form of the Killing vector field corresponding to this isometry is determined uniquely to be

$$\eta^\alpha = [\,-u\cos\theta,\ r\cos\theta,\ -\sin\theta, 0\,]. \tag{30}$$

This is the "*boost Killing vector*" if, as usual, we adopt the terminology from flat space-time: It generates the Lorentz transformations along the axis of axial symmetry. Indeed, the boost Killing vector generating Lorentz transformations along the z axis in Minkowski space-time is

$$\eta^\alpha_M = (z,0,t,0) \tag{31}$$

in cylindrical coordinates (t,ρ,z,ϕ). Introducing spherical coordinates (r,θ,ϕ) by $z = r\cos\theta$, $\rho = r\sin\theta$ and retarded time $u = t - r$, we get the vector field

$$\eta^\alpha_M = [\,-u\cos\theta,\ r(1 + u/r)\cos\theta,\ -\sin\theta\,(1 + u/r), 0\,], \tag{32}$$

which goes over to (30) as $r \to \infty$.

We have thus demonstrated the following.

Theorem: Suppose that an axially symmetric vacuum space-time (with circular group orbits and with a hypersurface orthogonal Killing vector) admits a "piece" of $\mathscr{I}^+$; i.e., suppose that the Bondi coordinates can be introduced and that the metric is of the form (2)–(6). Suppose that this space-time admits an additional Killing vector forming with the axial Killing vector a two-dimensional Lie algebra. Then the additional Killing vector has asymptotically the form (28). If $k = 0$, it generates a supertranslation; if $k \neq 0$, it is the boost Killing field.

It is of some interest to see what the form of this boost Killing vector on $\mathscr{I}^+$ is. Introducing, instead of $\hat{r}$, an inverted radial coordinate $l = r^{-1}$ (see Ref. 17, where l is used as the simplest conformal factor for obtaining the conformal Bondi frame on $\mathscr{I}^+$), we find $\xi^u = -u\cos\theta$, $\xi^l = -l\cos\theta \times (1 + lu)$, $\xi^\theta = -\sin\theta\ (1 + lu)$, $\xi^\phi = 0$, so that on $\mathscr{I}^+$ (where $l = 0$) in $\{u,l,\theta,\phi\}$ coordinates one obtains

$$\eta^\alpha_{/\mathscr{I}^+} = [\,-u\cos\theta,\ 0,\ -\sin\theta,\ 0\,]. \tag{33}$$

From the form of the boost symmetry group orbits (Fig. 2), it is seen that, within the null cone of the origin (i.e., $u = 0$), the boost Killing vector generates only rotation and, in particular, the points $z = \pm t$ (i.e., $\theta = 0$) at the null boundary, $\mathscr{I}^+$, represent the fixed points of the group orbits. Thus, components $\eta^\alpha_{/\mathscr{I}^+}$, $z = \pm t$, should be zero there. This is precisely what happens with the expression (33) for $u = 0$ and $\theta = 0$. Of course, in the case of the boost-rotation symmetric solutions (1) the boost Killing vector field also has the asymptotic form (32); however, this is not so easy to see since the introduction of Bondi's coordinate system is not straightforward in this case.[5] We shall turn to this question in the next section.

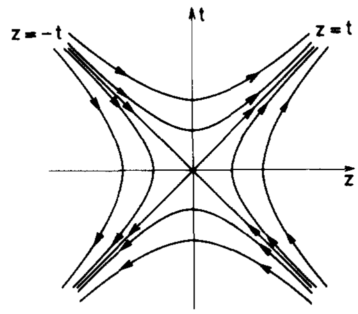


FIG. 2. The boost symmetry group orbits. The points $z = \pm t$ at $\mathscr{I}^+$ represent the fixed points.

3. CONSTRAINTS ON THE NEWS FUNCTION

Let us now investigate the asymptotic form of the Killing equations in the next orders in r^k when in these equations the components of the Killing vector field already get mixed with the metric functions appearing in (3)–(6). Recall that all functions are independent of ϕ and that the equations $L_\eta g_{03} = L_\eta g_{13} = L_\eta g_{23} = 0$ are satisfied exactly by $g = 0$. Expanding the other Killing equations (as given in the Appendix) to one order in r^k beyond that used to derive Eqs. (20)–(24), and recalling our previous result $f^{(-1)}_{,u} = A^{(-1)}_{,u} = 0$ [see (20), (22)], we find the following system of equations:

$$L_\eta g_{00} = 0 \quad (r^0){:}\quad B_{,u} + A^{(0)}_{,u} = 0, \tag{34}$$

$$L_\eta g_{02} = 0 \quad (r^1){:}\quad f^{(-1)}c_{,u} - f^{(0)}_{,u} + A^{(-1)}_{,\theta} = 0, \tag{35}$$

$$L_\eta g_{12} = 0 \quad (r^0){:}\quad -f^{(-1)}c + f^{(0)} + B_{,\theta} = 0, \tag{36}$$

$$L_\eta g_{22} = 0 \quad (r^1){:}\quad A^{(0)} - \tfrac{1}{2}B = -f^{(0)}_{,\theta} - Bc_{,u}, \tag{37}$$

$$L_\eta g_{33} = 0 \quad (r^1){:}\quad A^{(0)} - \tfrac{1}{2}B = 2cf^{(-1)}\cot\theta + c_{,\theta}f^{(-1)} - f^{(0)}\cot\theta + Bc_{,u}. \tag{38}$$

The equation $L_\eta g_{01} = 0$ is satisfied automatically for the terms proportional to r^{-1}, which is the next order in comparison with that from which (21) was obtained. If (20)–(24) and (34)–(38) are satisfied, then the Killing equations are satisfied asymptotically for the first two nontrivial orders of the expansions in r^k. Now the general solution (25)–(27) of (20)–(24) has not yet been used in (34)–(38). In Sec. 2 we saw that it is necessary to distinguish two cases: $k = 0$, when the Killing vector field is asymptotically a supertranslation, and $k \neq 0$, when it is the boost Killing vector.

A. Supertranslational Killing field

If $k=0$, then from (25)–(27) we have

$$A^{(-1)}=f^{(-1)}=0, \quad B=\alpha(\theta), \tag{39}$$

where α is an arbitrary function of θ; as we shall see, however, α will be restricted by Killing equations in further orders. First notice that (34), (35), and (36) imply

$$A^{(0)}_{,u}=f^{(0)}_{,u}=0, \quad f^{(0)}=-B_{,\theta}. \tag{40}$$

Taking $\partial/\partial u$ of either (37) or (38), we see that

$$c_{,uu}=0. \tag{41}$$

Hence, the radiation field—the leading term ($\sim r^{-1}$) in the expansion of the Riemann tensor—vanishes.[10] We shall now show that $c_{,u}$, the news function itself, has to vanish. In order to prove this, we have to solve the Killing equations in further orders. Assuming (37), (40), (41) and expanding $L_\eta g_{\mu\nu}$ into the higher orders in r^k, we find [examining again (3)–(6)] the following equations:

$$L_\eta g_{00}=0 \quad (r^{-1}): \quad A^{(1)}_{,u}=0, \tag{42}$$

$$L_\eta g_{01}=0 \quad (r^{-2}): \quad -B_{,\theta}(c_{,\theta}+2c\cot\theta)-Bcc_{,u} - A^{(1)}-BM=0, \tag{43}$$

$$L_\eta g_{02}=0 \quad (r^{0}): \quad -B_{,\theta}c_{,u}-f^{(1)}_{,u}+\tfrac{1}{2}B_{,\theta}+A^{(0)}_{,\theta}=0, \tag{44}$$

$$L_\eta g_{12}=0 \quad (r^{-1}): \quad f^{(1)}-B(c_{,\theta}+2c\cot\theta)=0, \tag{45}$$

$$L_\eta g_{22}=0 \quad (r^{0}): \quad f^{(1)}-B_{,\theta\theta}+A^{(1)}+cA^{(0)}-\tfrac{1}{2}Bc + BM-B(c_{,\theta}+2c\cot\theta)_{,\theta}=0, \tag{46}$$

$$L_\eta g_{33}=0 \quad (r^{0}): \quad f^{(1)}\cot\theta+B_{,\theta}(c_{,\theta}+\cot\theta) - c(A^{(0)}-\tfrac{1}{2}B)+A^{(1)}+BM - B\cot\theta(c_{,\theta}+2c\cot\theta)=0. \tag{47}$$

Now, the last equation simplifies considerably if we substitute for $f^{(1)}$ from (45) and for $A^{(1)}+BM$ from (43). In this manner we obtain

$$[B_{,\theta}\cot\theta+Bc_{,u}+A^{(0)}-\tfrac{1}{2}B]c=0. \tag{48}$$

Therefore, either $c=0$ or the expression in the square brackets has to vanish. Consider first $c=0$. Then, of course, the news function vanishes. Comparing (37) with (38) (with $f^{(-1)}=c_{,u}=0$), we find

$$f^{(0)}=a\sin\theta, \quad a=\text{const}, \tag{49}$$

and (40) gives

$$B=a\cos\theta+b, \quad b=\text{const}. \tag{50}$$

Since the general form of the Killing vector field η^α is [see (11), (12), (14), and (17)]

$$\eta^\alpha=[B,\ Ae^{-2\beta}-BV/2r,\ BU+fe^{-\gamma}/r,\ 0], \tag{51}$$

we find that, asymptotically, it now reads

$$\eta^\alpha=(a\cos\theta+b,\ -a\cos\theta,\ a\sin\theta/r,\ 0). \tag{52}$$

Hence, the Killing vector field generating supertranslations actually only generates translations (with $a=0$ the time translation, with $b=0$ the translation along the z axis).

Returning to Eq. (48), we next suppose that the expression in the square bracket vanishes:

$$B_{,\theta}\cot\theta+Bc_{,u}+A^{(0)}-\tfrac{1}{2}B=0. \tag{53}$$

From (39), (40), and (38) it then follows that

$$A^{(0)}-\tfrac{1}{2}B=0. \tag{54}$$

Equation (37) with $f^{(0)}=-B_{,\theta}$ [see (40)] gives

$$B_{,\theta\theta}=Bc_{,u} \tag{55}$$

and the comparison with (53) and (54) leads to the equation for $B(\theta)$:

$$B_{,\theta\theta}+B_{,\theta}\cot\theta=0. \tag{56}$$

There are two independent solutions. The first one, $B=\text{const}$, gives immediately [for example, from (55)] $c_{,u}=0$; from (51) and (54) we see that the Killing field η^α is of the form (52) with $a=0$; i.e., it is a time translation. The other independent solution of (56) is

$$B=\text{const}\ \log(\tan\theta/2). \tag{57}$$

This solution is irregular on the axis $\theta=0,\pi$. However, it is not even locally the solution of the system (42)–(47). In fact, a nonlinear differential equation of the fourth order can be derived for function $B(\theta)$ if we apply $\partial/\partial u$ to (43), express $M_{,u}$ from (7), $A^{(1)}_{,u}=0$ by (42), and express $c_{,u}$ in terms of B using (55); one finds $c_{,u}=-\cos\theta[\sin^2\theta\log(\tan\theta/2)]^{-1}$. It can be checked by straightforward calculations that the fourth-order equation for B is satisfied by the solution (50) but not by (57). Therefore, we have demonstrated the following.

Theorem: If in an axially symmetric vacuum space-time with at least local $\mathscr{I}^+$ another Killing field exists of the form (28) with $k=0$ (i.e., it generates a supertranslation), then this field must, in fact, be the generator of a translation, and the news function must vanish.

Analogous results were proved in Refs. 1 and 2 using concepts defined directly on $\mathscr{I}^+$ and without assuming axial symmetry (see Lemma 1.4 in Ref. 1 and Lemma 3.5 in Ref. 2). Although our derivation is restricted to axial symmetry, it is local. Thus, we did not even have to assume that $\mathscr{I}^+$ is topologically $S^2\times R$ as in Refs. 1 and 2.

B. The boost Killing vector

Now consider the case when $k\neq 0$ in (28). We know already that we may put $k=1$ and $\alpha=0$ without loss of generality. The Killing vector field has asymptotically the form (30)—it generates the boost along the z axis. Before writing down the differential equation which the news function has to satisfy and demonstrating that the news function of the boost-rotation symmetric solutions (1) really does satisfy this equation, let us show how the wrong conclusion of Berezdivin and Herrera[3,4] (that the news function has to vanish) was reached. Berezdivin and Herrera do not give the solutions (25)–(27) of the Killing equations (in the first orders in r^j) but they apply $\partial^2/\partial u^2$ to both (37) and (38), and use the remaining Eqs. (20)–(24) and (34)–(36). However, their Eq. (40d) in Ref. 3 [and Eq. (26) in Ref. 4], corresponding to our Eq. (38), is incorrect, since the term $2cf^{(-1)}\cot\theta$ is missing there. Performing the same procedure as Berezdivin and Herrera[3] with the correct Eq. (38), one finds that the equations obtained by applying $\partial^2/\partial u^2$ to (37) and (38) are identical, so that their comparison does not give $c_{,uu}=0$ (or $f^{(-1)}=0$) as claimed in Ref. 3.

Nevertheless, the system (34)–(38) leads to a constraint on the news function. Regarding (20)–(24) and (34)–(38), we see that only Eqs. (35)–(38) contain the news function. Equation (35) can be obtained by applying $\partial/\partial u$ to (36) and using (21). Now apply $\partial/\partial u$ to (37) and express $f^{(0)}_{,u}$ from (35). Do the same with (38). The left-hand sides of the equations so obtained are identical, and the comparison of the right-hand sides gives

$$c_{,u}\left[f^{(-1)}\cot\theta+f^{(-1)}_{,\theta}+2B_{,u}\right]+2f^{(-1)}c_{,u\theta}$$
$$-A^{(-1)}_{,\theta}\cot\theta+A^{(-1)}_{,\theta\theta}+2Bc_{,uu}=0.$$

Substituting for $f^{(-1)}$, $A^{(-1)}$, and B, the solutions (25)–(27) with $k=1$ and $\alpha=0$, we arrive at the following simple constraint on the news function:

$$(u\cot\theta)c_{,uu}+c_{,u\theta}+(2\cot\theta)c_{,u}=0. \tag{58}$$

In Ref. 5 the news function was derived for the solutions of Bonnor and Swaminarayan,[11] which have the form (1) with specific functions μ and λ. However, the concrete form of μ and λ was not used during the derivation. Thus expression (26) in Ref. 5 represents the news function for a general boost-rotation symmetric space-time (1) that is asymptotically flat in the sense explained at the beginning of Sec. 2. A detailed analysis of the properties of this news function will be given in a paper on boost-rotation symmetric space-times[18]; for our present purposes, it is sufficient to observe that the news function has to be of the form

$$c_{,u}=(\sin\theta)^{-2}F(\widetilde{U}/\sin\theta), \tag{59}$$

where F is a general function of the argument $\widetilde{U}/\sin\theta$ and $\widetilde{U}$ is "flat-space retarded time" defined by $\widetilde{U}=t-r$; the time coordinate t enters the metric (1), and ρ,z,ϕ in (1) are connected with Bondi's r,θ,ϕ just by $\rho=r\sin\theta$, $z=r\cos\theta$, $\phi=\phi$ at large r. (The notation in Ref. 5 is different, however: Bondi's retarded time is denoted by $\bar{u}$ there and $\widetilde{U}$ is denoted by u.) The connection between $\widetilde{U}$ and Bondi's u and θ, i.e., the function $\widetilde{U}(u,\theta)$ (denoted by $\mathring{\pi}$ in Ref. 5), is given implicitly by the equation [see (25) in Ref. 5]

$$u=\int\exp\left(\lambda_0\frac{\widetilde{U}}{\sin\theta}\right)d\widetilde{U}+\Theta(\theta), \tag{60}$$

where λ_0 is the leading term in the expansion of the function λ [entering (1)] for large r with $\widetilde{U},\theta,\phi$ fixed, $\lambda=\lambda_0+O(r^{-1})$, $\Theta(\theta)$ is an arbitrary function. As with the function F in (59), λ_0 depends only on the combination $\widetilde{U}/\sin\theta$, owing to the boost symmetry of the metric (1). [This point was not realized in Ref. 5 and will be explained in detail elsewhere,[18] but the reader can easily convince himself that the function λ_0 (denoted by β in Ref. 5) does have this property and that the news function, given in (26) in Ref. 5, is of the form (59).] From (60) it follows that

$$\widetilde{U}_{,u}e^{[\lambda_0(\widetilde{U}/\sin\theta)]}=1, \tag{61}$$

where $\widetilde{U}_{,u}=\partial\widetilde{U}/\partial u$. Using this and calculating $c_{,uu}$ and $c_{,u\theta}$ from (59), we arrive at the equation

$$(\widetilde{U}_{,u})^{-1}(\widetilde{U}\cot\theta-\widetilde{U}_{,\theta})c_{,uu}+c_{,u\theta}+2c_{,u}\cot\theta=0. \tag{62}$$

The equation obtained is identical to Eq. (58), which the news function has to satisfy as a consequence of the Killing equations, because

$$u\cot\theta=(\widetilde{U}_{,u})^{-1}(\widetilde{U}\cot\theta-\widetilde{U}_{,\theta}). \tag{63}$$

The last relation can easily be verified by applying $\partial/\partial u$ to both sides and taking into account (61), where it is again important that λ_0 depends only on the variable $\widetilde{U}/\sin\theta$. A possible additive function of θ is equal to zero, which corresponds to the fact that we choose that system for which the boost Killing vector is given by (30) with the additive function $\alpha(\theta)$ [cf. (28)] equal to zero.

Indeed, in the case of the boost-rotation symmetric solutions (1), the boost Killing vector in coordinates $\{t,\rho,z,\phi\}$ has the same form as in flat space-time:

$$\eta^\alpha=(z,0,t,0). \tag{64}$$

Introducing "flat-space spherical" coordinates (r,θ,ϕ) and $\widetilde{U}=t-r$ as before, we get

$$\eta^\alpha=[-\widetilde{U}\cos\theta,\quad r(1+\widetilde{U}/r)\cos\theta,$$
$$-\sin\theta(1+\widetilde{U}/r),\ 0],$$

as in (32). Now since in the leading order Bondi's r and θ are identical with "flat-space" r and θ, whereas Bondi's u is related with $\widetilde{U}$ by $\widetilde{U}=\widetilde{U}(u,\theta)$, it follows that $\eta^{\widetilde{U}}=-\widetilde{U}\times\cos\theta=\widetilde{U}_{,u}\eta^u-\widetilde{U}_{,\theta}\eta^\theta=-\widetilde{U}_{,u}u\cos\theta+\widetilde{U}_{,\theta}\sin\theta$, where we have substituted for components, η^u, η^θ in Bondi's coordinates from (30). Dividing the last relation by $\widetilde{U}_{,u}\sin\theta$, we obtain (63).

Therefore, we can conclude that the boost Killing vector of the space-times described by the metrics (1) is, asymptotically, really of the form (30) and the news function satisfies the differential equation (58).

4. ABELIAN GROUP OF NULL ROTATIONS

Since isometries in asymptotically flat space-times with $\mathscr{I}^+$ with topology $S^2\times R$ act on $\mathscr{I}^+$ in the same way as in the Minkowski space-time, one may ask for other two-dimensional abelian subgroups besides those with the boost and the rotation. The only possibility is the group of null rotations. In Minkowski space-time a basis of Killing vectors is

$$u\frac{\partial}{\partial y}+2y\frac{\partial}{\partial v},\quad u\frac{\partial}{\partial x}+2x\frac{\partial}{\partial v}, \tag{65}$$

where the coordinates are such that

$$ds^2=du\,dv-dx^2-dy^2 \tag{66}$$

is the metric. Coordinates adapted to the Killing fields are $\bar{u}$, $\bar{v}$, ϕ, χ, defined by

$$u=\bar{u},\quad x=\bar{u}\phi,\quad y=\bar{u}\chi,\quad v=\bar{u}(\phi^2+\chi^2)+\bar{v}. \tag{67}$$

Then one gets

$$ds^2=d\bar{u}\,d\bar{v}-\bar{u}^2(d\phi^2+d\chi^2). \tag{68}$$

Comparing this with the general form of the metric admitting two computing spacelike, hypersurface orthogonal Killing vectors,[19] one observes that (68) is a special case of the metric.

$$ds^2=e^\lambda\,d\bar{u}\,d\bar{v}-\bar{u}^2(e^{-\mu}\,d\phi^2+e^\mu\,d\chi^2), \tag{69}$$

where $\lambda = \lambda(\bar u,\bar v)$ and $\mu = \mu(\bar u,\bar v)$. The metric (69) has the property that the gradient of the volume of the group orbits is a null vector. The vacuum field equations turn out to be (the conventions follow Ref. 20)

$$\begin{aligned} R_{\bar v\bar v} &= -\tfrac{1}{2}(\mu_{,\bar v})^2 = 0,\\ R_{\bar v\bar u} &= -\tfrac{1}{2}\mu_{,\bar u}\mu_{,\bar v} - \lambda_{,\bar u\bar v} = 0,\\ R_{\bar u\bar u} &= 2\lambda_{,\bar u}/\bar u - \tfrac{1}{2}(\mu_{,\bar u})^2,\\ -R_{\phi\phi}e^{\lambda+\mu} &= R_{\chi\chi}e^{\lambda-\mu}\\ &= 2[\bar u^2\mu_{,\bar u\bar v} + \bar u\mu_{,\bar v}] = 0. \end{aligned} \tag{70}$$

They imply that μ is independent of $\bar v$ and that $\lambda = A(\bar u) + B(\bar v)$. Rescaling $\bar v$, one observes that $\partial/\partial\bar v$ is a further Killing vector field. Space-times with an abelian three-dimensional group acting on a null hypersurface are plane waves.[19] Hence, we obtain space-times which do not have $\mathscr{I}^+$ topologically $S^2\times R$.

This result also sheds some light on the question posed in Ref. 2 in the Summary. It is not known whether space-times with $\mathscr{I}^+$ topologically $S^2\times R$ exist admitting a three-parameter group acting on $\mathscr{I}^+$ different from both the three-dimensional rotation group and the three-dimensional Lorentz group. All the three remaining groups contain a two-parameter group of null rotations.[13] Hence, no examples exist with hypersurface orthogonal Killing vectors.

It can be shown that a further Killing vector field is still implied by the field equations if one generalizes the metric (69) to the case when two Killing fields are not hypersurface orthogonal but still orthogonally transitive. Nothing is known in the fully general case in which Killing fields are not even orthogonally transitive. We conjecture that a further Killing field will exist in this case, too.

ACKNOWLEDGMENTS

One of us (J. B.) is grateful to Professor Jürgen Ehlers and all members of the Munich relativity group for kind hospitality. We thank Michael Streubel and Leoš Dvořák for checking the field equations in the case of two null rotations on the computer, and Ron Kates for improving our English.

APPENDIX

Assuming that the metric is in the Bondi form (2) and the vector field η^α is given by (11)–(14), the left-hand sides of the Killing equations $L_\eta g_{\alpha\beta} = 0$ read as follows:

$$\begin{aligned} L_\eta g_{00} = {}& B(Vr^{-1}e^{2\beta} - U^2r^2e^{2\gamma})_{,u}\\ &+ (fr^{-1}e^{-\gamma} + BU)(Vr^{-1}e^{2\beta} - U^2r^2e^{2\gamma})_{,\theta}\\ &+ (Ae^{-2\beta} - \tfrac{1}{2}BVr^{-1})(Vr^{-1}e^{2\beta} - U^2r^2e^{2\gamma})_{,r}\\ &+ 2B_{,u}(Vr^{-1}e^{2\beta} - U^2r^2e^{2\gamma})\\ &- 2Ufre^{\gamma}\gamma_{,u} + 2Ure^{\gamma}f_{,u} + 2Ur^2e^{2\gamma}(BU)_{,u}\\ &- e^{2\beta}r^{-1}(BV)_{,u} + 2A_{,u} - 4A\beta_{,u},\\ L_\eta g_{01} = {}& e^{2\beta}\beta_{,\theta}(fr^{-1}e^{-\gamma} + 2BU) + (Be^{2\beta})_{,u}\\ &+ Ue^{\gamma}(rf_{,r} - f - fr\gamma_{,r}) + Ur^2e^{2\gamma}BU_{,r}\\ &+ (A - \tfrac{1}{2}BVr^{-1}e^{2\beta})_{,r},\\ L_\eta g_{02} = {}& fr[(Ue^{\gamma})_{,\theta} + e^{\gamma}\gamma_{,u}] - re^{\gamma}f_{,u}\\ &+ re^{\gamma}Uf_{,\theta} + 2Br^2Ue^{2\gamma}\gamma_{,u}\\ &+ (Ae^{-2\beta} - \tfrac{1}{2}BVr^{-1})(Ur^2e^{2\gamma})_{,r}\\ &+ Vr^{-1}e^{2\beta}B_{,\theta} - \tfrac{1}{2}r^{-1}e^{2\beta}(BV)_{,\theta}\\ &+ 2Ur^2e^{2\gamma}BU_{,\theta} + 2Br^2U^2e^{2\gamma}\gamma_{,\theta}\\ &+ A_{,\theta} - 2A\beta_{,\theta},\\ L_\eta g_{03} = {}& A_{,\phi} - re^{-\gamma}\sin\theta\,(g_{,u} + g\gamma_{,u})\\ &+ Ure^{\gamma}f_{,\phi} + \tfrac{1}{2}Vr^{-1}e^{2\beta}B_{,\phi},\\ L_\eta g_{11} = {}& e^{2\beta}B_{,r},\\ L_\eta g_{12} = {}& e^{\gamma}(f + fr\gamma_{,r} - rf_{,r})\\ &+ e^{2\beta}B_{,\theta} - r^2e^{2\gamma}BU_{,r},\\ L_\eta g_{13} = {}& e^{-\gamma}\sin\theta\,(g + gr\gamma_{,r} - rg_{,r}) + e^{2\beta}B_{,\phi},\\ L_\eta g_{22} = {}& -2re^{\gamma}[f_{,\theta} + rBe^{\gamma}\gamma_{,u} + Ure^{\gamma}\gamma_{,\theta} + Bre^{\gamma}U_{,\theta}\\ &+ e^{\gamma}(Ae^{-2\beta} - \tfrac{1}{2}BVr^{-1})(1 + r\gamma_{,r})],\\ L_\eta g_{23} = {}& -re^{-\gamma}[g(\gamma_{,\theta}\sin\theta - \cos\theta)\\ &+ g_{,\theta}\sin\theta + e^{2\gamma}f_{,\phi}],\\ L_\eta g_{33} = {}& -2re^{-2\gamma}\sin^2\theta\,[(\cot\theta - \gamma_{,\theta})(rBU + fe^{-\gamma})\\ &- rB\gamma_{,u} + (Ae^{-2\beta} - \tfrac{1}{2}BVr^{-1})(1 - r\gamma_{,r})]\\ &- 2re^{-\gamma}\sin\theta\, g_{,\phi}. \end{aligned}$$

In Ref. 3, there are errors in $L_\eta g_{00}$, $L_\eta g_{12}$, $L_\eta g_{23}$, $L_\eta g_{02}$, and $L_\eta g_{03}$. In the Appendix of Ref. 3 [where the Lie derivatives of $g_{\mu\nu}$ with respect to the Sachs tetrad (11) are given] the following errors appear: the second term in $L_t g_{01}$ contains the factor $(1-i)$, but it should be $(1+i)$; $L_k g_{22}$ has an opposite sign; the second term in the brackets in $L_m g_{00}$ contains $-U^2r^2e^{2\beta}$—that should be $-U^2r^2e^{2\gamma}$; the factor B in $L_m g_{33}$ should be omitted and in $L_m g_{02}$ an additional term $(Ur^2e^{2\gamma})_{,\theta}U$ is omitted. Most of these are misprints; the crucial error appeared in the expansion of the Killing equation $L_\eta g_{33} = 0$, as explained in the main text.

[1] A. Ashtekar and B. C. Xanthopoulos, J. Math. Phys. **19**, 2216 (1978).
[2] A. Ashtekar and B. G. Schmidt, J. Math. Phys. **21**, 862 (1980).
[3] R. Berezdivin and L. Herrera, Phys. Rev. D **11**, 2063 (1975).
[4] R. Berezdivin and L. Herrera, Lett. Nuovo Cimento **16**, 92 (1976).
[5] J. Bičák, Proc. Roy. Soc. London A **302**, 201 (1968).
[6] W. Kinnersley and M. Walker, Phys. Rev. D **2**, 1359 (1970).
[7] A. Ashtekar and T. Dray, Commun. Math. Phys. **79**, 581 (1981).
[8] B. G. Schmidt, Commun. Math. Phys. **78**, 447 (1981).
[9] R. Sachs, Phys. Rev. **128**, 2851 (1962).
[10] H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner, Proc. Roy. Soc. London A **269**, 21 (1962).
[11] W. B. Bonnor and N. S. Swaminarayan, Z. Phys. **177**, 240 (1964).
[12] J. Bičák, C. Hoenselaers, and B. G. Schmidt, Proc. R. Soc. London A (to appear).
[13] See, e.g., J. Segercrantz, Ann. Acad. Sci. Fennicae A **233**, 1 (1967) and B. G. Schmidt, thesis (University of Hamburg, 1968).
[14] J. Stachel, J. Math. Phys. **7**, 1321 (1966).
[15] J. F. Plebański and M. Demianski, Ann. Phys. (N.Y.) **98**, 98 (1978).
[16] We keep the original notation of Sachs [Proc. Roy. Soc. London A **270**, 103 (1962)], which is also used by Berezdivin and Herrera in Refs. 3,4, or in Ref. 16. Today one rather uses the Newman–Penrose notation $\{1^\alpha, n^\alpha, m^\alpha, \bar m^\alpha\}$ for the null tetrad. Notice, however, that usually one chooses $l^r = 1$, where r is an affine parameter along outgoing null geodesic $\{u,\theta,\phi\}$ = const while in Bondi's formalism r is the luminosity distance and $k^r \neq 1$ in general.
[17] L. A. Tamburino and J. H. Winicour, Phys. Rev. **150**, 1039 (1966).
[18] J. Bičák and B. G. Schmidt, in preparation.
[19] D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980), p. 175.
[20] Ch. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

Accelerating black hole in a magnetic field

K. D. Krori
Mathematical Physics Forum, Cotton College, Gauhati-781 001, India and Institute of Advanced Study in Science and Technology, Assam Science Society, Gauhati-781 001, India

Sumita Chaudhury and S. Dowerah
Mathematical Physics Forum, Cotton College, Gauhati-781 001, India

(Received 19 October 1982; accepted for publication 11 March 1983)

This paper presents an investigation of the geometrical properties of an accelerating black hole embedded in a magnetic field.

PACS numbers: 04.20.Jb, 97.60.Lf

I. INTRODUCTION

Two most important vacuum solutions of the Einstein field equations are the Schwarzschild and Kerr solutions. Perhaps another very important vacuum solution is that of a uniformly accelerating object. The vacuum C-metric representing this solution was first derived by Levi-Cività[1] and was rediscovered over forty years later by Newman and Tamburino.[2] The mechanism causing the acceleration of an object is an interesting subject of study. Kinnersley and Walker[3] have pointed out that the two-surface surrounding it has a conical singularity at the north or the south pole. They have suggested that the nodal singularity appears due to the neglect of the force necessary to accelerate an object. Transforming the charged C-metric into another exact solution of the Einstein–Maxwell field equations corresponding to a massive charged particle accelerated by an electric field, Ernst[4] has shown that when the appropriate equation of motion is satisfied, the nodal singularity associated with the C-metric disappears.

Farhoosh and Zimmerman[5] are, however, of the view that the nodal singularity is a manifestation of the uniform acceleration, and is not a direct consequence of the neglect of the force necessary to cause acceleration. They suggest[5,6] that the acceleration of an object is caused by the reaction of the emission of gravitational radiation that it anisotropically emits. They back up this suggestion by constructing an interior C-metric. At the boundary of the interior solution, there exists a discontinuity in the pressure which is, they claim, responsible for the uniform acceleration of the object.

Much interest is being evinced now in the study of black holes under realistic conditions, viz., in the presence of matter or external fields. Recently Wild and Kerns[7] have studied the surface geometry of a Schwarzschild black hole in a magnetic field. This has been followed up by Wild, Kerns, and Drish[8] in an investigation of the geometry of the event horizon of a Kerr black hole embedded in an external magnetic field oriented along the axis of symmetry. In view of what has been stated in the preceding paragraph, the next problem of immediate interest would be the study of an accelerating black hole embedded in a magnetic field oriented along the axis of symmetry. An investigation of this subject is presented in this paper.

II. MAGNETIZED ACCELERATING BLACK HOLE

The static vacuum C-metric describing the gravitational field of a uniformly accelerating Schwarzschild-like object is[9]

$$ds^2 = r^2(F\,dt^2 - F^{-1}\,dq^2 - G^{-1}\,dp^2 - G\,d\phi^2), \tag{1}$$

where

$$G(p) = 1 - p^2 - 2Amp^3 = \sin^2\theta, \tag{2}$$

$$F(q) = -1 + q^2 - 2Amq^3, \tag{3}$$

$$r = 1/(A(p+q). \tag{4}$$

Here A is the acceleration and m is the mass of the object. We shall be interested in the range of the coordinate p from p_0 to p_π and that of q from q_r to q_s, where[6]

$$p_0 = -(1/6Am)[2\cos(\lambda/3 + 4\pi/3) + 1], \tag{5}$$

$$p_\pi = -(1/6Am)[2\cos(\lambda/3 + 2\pi/3) + 1], \tag{6}$$

$$\cos\lambda = 1 - 54A^2m^2, \tag{7}$$

$$q_r = -(1/6Am)[2\cos(\delta/3 + 4\pi/3) - 1], \tag{8}$$

$$q_s = -(1/6Am)[2\cos(\delta/3 + 2\pi/3) - 1], \tag{9}$$

$$\cos\delta = -(1 - 54A^2m^2). \tag{10}$$

For $A^2m^2 < \frac{1}{27}$, there are two physically meaningful event horizons. One is analogous to Schwarzschild surface and is given by $q = q_s$. The other is called the Rindler surface and corresponds to $q = q_r$.

When embedded in a magnetic field, the metric (1) becomes (following a method due to Ernst[10])

$$ds^2 = \Lambda^2 r^2(F\,dt^2 - F^{-1}\,dq^2 - G^{-1}\,dp^2) - Gr^2\Lambda^{-2}\,d\phi^2, \tag{11}$$

where

$$\Lambda = 1 + \tfrac{1}{4} r^2 B_0^2\, G(p)$$

and B_0 is the magnetic field parameter.

III. SCHWARZSCHILD SURFACE

A. Gaussian curvature

From (11) we get the two-dimensional Riemannian metric intrinsically defining the Schwarzschild event hori-

 0022-2488/84/030607-05$02.50

TABLE I. Expression for $m^2K_{\theta=\pi/2}$ as a function of β for various choices of $54A^2m^2$.

$54A^2m^2$	$m^2K_{\theta=\pi/2}$	$\beta_c^2(\beta_c=$ critical value of $\beta)$
0	$-\frac{4(\beta^2 0.25)(\beta^2+0.25)}{(1+4\beta^2)^4}$	0.25
0.25	$-\frac{4.3988(\beta^2-0.2497)(\beta^2+0.2147)}{(1+4.1586\beta^2)^4}$	0.2497
0.5	$-\frac{4.8658(\beta^2-0.2487)(\beta^2+0.1826)}{(1+4.3421\beta^2)^4}$	0.2487
0.75	$-\frac{5.4251(\beta^2-0.2468)(\beta^2+0.1534)}{(1+4.5591\beta^2)^4}$	0.2468
1	$-\frac{6.1154(\beta^2-0.2436)(\beta^2+0.1268)}{(1+4.8231\beta^2)^4}$	0.2436
1.25	$-\frac{7.0061(\beta^2-0.2385)(\beta^2+0.1022)}{(1+5.1583\beta^2)^4}$	0.2385
1.5	$-\frac{8.2422(\beta^2-0.2305)(\beta^2+0.0791)}{(1+5.6149\beta^2)^4}$	0.2305
1.75	$-\frac{10.2305(\beta^2-0.2170)(\beta^2+0.0565)}{(1+6.3322\beta^2)^4}$	0.2170

zon as

ds^2(event horizon)

$$=A_{01}^2 r_{01}^2 G^{-1} dp^2 + G r_{01}^2 A_{01}^{-2} d\phi^2, \tag{12}$$

where

$$r_{01} = 1/A(p+q_s), \tag{13}$$

$$A_{01} = 1 + \tfrac{1}{4} r_{01}^2 G(p) B_0^2. \tag{14}$$

The expression for the Gaussian curvature for the surface defined by the line element (12) is

$$K = -\frac{1}{2(E^*G^*)^{1/2}} \frac{d}{dp}\left(\frac{1}{(E^*G^*)^{1/2}} \frac{dG^*}{dp}\right), \tag{15}$$

where

$$E^* = A^2 r_{01}^2 G^{-1}, \tag{16}$$

$$G^* = A^{-2} r_{01}^2 G. \tag{17}$$

Using (15), (16), and (17) a very complicated expression for the Gaussian curvature K is obtained. For the equator this

TABLE II. Expression for $m^2K_{\theta=0}$ as a function of β for various choices of $54A^2m^2$.

$54A^2m^2$	$m^2K_{\theta=0}$
0	$0.25 + 8\beta^2$
0.25	$0.1447 + 5.6930\beta^2$
0.5	$0.1022 + 4.6399\beta^2$
0.75	$0.0717 + 3.7768\beta^2$
1	$0.0481 + 3.0000\beta^2$
1.25	$0.0296 + 2.2660\beta^2$
1.5	$0.0154 + 1.5466\beta^2$
1.75	$0.0052 + 0.8138\beta^2$
1.99	$0.000\,04 + 0.037\,99\beta^2$

TABLE III. Expression for $m^2K_{\theta=\pi}$ as a function of β for various choices of $54A^2m^2$.

$54A^2m^2$	$m^2K_{\theta=\pi}$
0	$0.25 + 8\beta^2$
0.25	$0.3409 + 10.0791\beta^2$
0.5	$0.3676 + 10.8913\beta^2$
0.75	$0.3808 + 11.4983\beta^2$
0	$0.3849 + 12.0000\beta^2$
1.25	$0.3808 + 12.4350\beta^2$
1.5	$0.3676 + 12.8230\beta^2$
1.75	$0.3409 + 13.1754\beta^2$

expression reduces to

$$m^2K_{\theta=\pi/2} = -\frac{\left[\left(\frac{q_s+3}{q_s^2 A^2m^2}\right)\beta^4 - \frac{8}{q_s^2}\beta^2 + (Am)^2(1-q_s^2)\right]}{\left\{1+\frac{\beta^2}{(Amq_s)^2}\right\}^4}, \tag{18}$$

Here $\beta = mB_0/2$ is defined to be a dimensionless distortion parameter analogous to the quantity used by Smarr[11] in his discussion of the Kerr solution.

From Table I we observe that, for a fixed value of Am, there is a critical value β_c of the distortion parameter such that

$$K_{\theta=\pi/2} > 0, \quad \text{if } \beta<\beta_c,$$
$$=0, \quad \text{if } \beta=\beta_c,$$
$$<0, \quad \text{if } \beta>\beta_c.$$

Thus for a fixed value of Am, the equatorial zone will exhibit negative Gaussian curvature if the distortion parameter β exceeds the corresponding critical value β_c. As the acceleration increases, the critical value β_c gradually decreases. As $A^2m^2 \to \frac{1}{27}$, β_c approaches 0.172 068. In the absence of the

TABLE IV. Expression for C_E/m as a function of β for various choices of $54A^2m^2$.

$54A^2m^2$	C_E/m
0	$\frac{12.5664}{(1+4\beta^2)}$
0.25	$\frac{13.0345}{(1+4.1586\beta^2)}$
0.5	$\frac{13.5791}{(1+4.3421\beta^2)}$
0.75	$\frac{14.2271}{(1+4.5591\beta^2)}$
1	$\frac{15.0222}{(1+4.8231\beta^2)}$
1.25	$\frac{16.0434}{(1+5.1583\beta^2)}$
1.5	$\frac{17.4570}{(1+5.6149\beta^2)}$
1.75	$\frac{19.7369}{(1+6.3322\beta^2)}$

TABLE V. Bivariate table showing various numerical values of C_E/m.

$\beta\rightarrow$ / $54A^2m^2\downarrow$	0	0.1	0.2	0.3	0.4	0.5
0	12.5664	12.0830	10.8331	9.239 98	7.6624	6.2832
0.25	13.0345	12.5141	11.1755	9.4847	7.8268	6.3906
0.5	13.5791	13.0141	11.5697	9.7636	8.0125	6.5111
0.75	14.2271	13.6067	12.0328	10.0879	8.2263	6.6489
1	15.0222	14.3311	12.5928	10.4752	8.4790	6.8104
1.25	16.0434	15.2564	13.2993	10.9567	8.7893	7.0071
1.5	17.457	16.5289	14.2553	11.5967	9.1957	7.2625
1.75	19.7369	18.5616	15.7481	12.5721	9.8040	7.6409

netic field ($\beta = 0$), the Gaussian curvature is positive in the equatorial zone for any value of the acceleration ($A^2m^2 < \frac{1}{27}$). At the pole $\theta = 0$, the Gaussian curvature is given by

$$m^2K_{\theta=0} = \frac{(Am)^2[(1 + 6Amp_0) - 2Ar_{02}(p_0 + 3Amp_0^2)]}{(Ar_{02})^2} + 8(p_0 + 3Am\,p_0^2)^2\beta^2, \tag{19}$$

where

$$r_{02} = 1/A\,(p_0 + q_s). \tag{20}$$

Table II shows that the Gaussian curvature is always positive in the vicinity of the pole $\theta = 0$. As acceleration increases the curvature decreases. As $A^2m^2 \rightarrow \frac{1}{27}$, $m^2K_{\theta=0} \rightarrow 28.8\times 10^{-13}\beta^2$, that is, $K_{\theta=0} \rightarrow 7.2\times 10^{-13}B_0^2$ which is negligibly small for $B_0^2 \ll 10^{13}$. So an increase of acceleration results in the flattening of the polar region $\theta = 0$. The Gaussian curvature at the pole $\theta = \pi$ is given by

$$m^2K_{\theta=\pi} = \frac{(Am^2)[(1 + 6Amp_\pi) - 2Ar_{03}(p_\pi + 3Amp_\pi^2)]}{(Ar_{03})^2} + 8(p_\pi + 3Am\,p_\pi^2)^2\beta^2, \tag{21}$$

where

$$r_{03} = 1/A\,(p_\pi + q_s). \tag{22}$$

From Table III we see that the curvature goes on increasing as $54A^2m^2$ varies from 0 to 1 for any value of the distortion parameter β. In the absence of the magnetic field ($\beta = 0$), at first the curvature gradually increases with $54A^2m^2$, attains its maximum value 0.3849 (approximately), and then decreases symmetrically about this maximum value if the acceleration is further increased. Also for a fixed value of Am, $K \rightarrow \infty$ as $\beta \rightarrow \infty$ and a cusplike singularity will develop at the pole $\theta = \pi$.

B. Range of coordinate ϕ

For $A \rightarrow 0$, the metric (12) reduces to the form

$$ds^2 = A_0^2 r_0^2\, d\theta^2 + A_0^{-2}\sin^2\theta\, r_0^2\, d\phi^2,$$

where

$$A_0 = 1 + \tfrac{1}{4}B_0^2\, r_0^2 \sin^2\theta \quad \text{and} \quad r_0 = 2m,$$

which is the line element describing magnetized Schwarzschild black hole. For this surface, the range of ϕ is known to be $0 < \phi \leqslant 2\pi$. For $A = 0$, $B_0 = 0$, we get the geometry of a simple sphere. However, for $A \neq 0$, $B_0 \neq 0$, we cannot presume $0 < \phi \leqslant 2\pi$. Let us identify 0 with $2\pi F$, where F is yet unknown. The effect of A and B_0 is to distort the Schwarzschild surface from the spherical shape. For small of A and B_0, distortion from spherical geometry will be small, and the event horizon will be a compact differential manifold homeomorphic to a sphere. Hence by the Gauss–Bonnet theorem,

$$\iint K\cdot ds = 4\pi$$

or

$$F[1 + 3Am(p_0 + p_\pi)](p_\pi - p_0) = 1. \tag{23}$$

Using (5) and (6) we get

$$F = 4\sqrt{3}\,Am/\sin\tfrac{2}{3}\lambda. \tag{24}$$

Thus $0 < \phi \leqslant 2\pi F$, where F is given by (24). Using (7) it is easily shown that $F \rightarrow 1$ as $A \rightarrow 0$.

It should be pointed out that our choice for the range of ϕ leads to a nodal singularity at both $\theta = 0$ and $\theta = \pi$. One could, however, adopt the method of Kinnersley and Walker[3] for picking the range of ϕ. But in that method also, a node persists at the north or the south pole. As mentioned in the introduction of the paper, Ernst[4] removed, in the case of the charged C-metric, the nodal singularity by the addition of an external electric field.

C. Polar and equatorial circumferences

To have an idea of the surface geometry of the event horizon, we calculate the equatorial and the polar circumfer-

TABLE VI. Expression for C_E/m as a function of β for various choices of $54A^2m^2$.

$54A^2m^2$	C_P/m as function of β^2 for different values of $54A^2m^2$
0	$12.566 + 25.133\,\beta^2$
0.25	$13.379 + 28.331\,\beta^2$
0.5	$14.387 + 32.470\,\beta^2$
0.75	$15.676 + 38.044\,\beta^2$
1	$17.432 + 45.983\,\beta^2$
1.25	$19.911 + 58.272\,\beta^2$
1.5	$24.036 + 80.109\,\beta^2$
1.75	$32.979 + 131.747\,\beta^2$

TABLE VII. Bivariate table showing various numerical values of D.

$\beta\rightarrow$ / $54A^2m^2$ ↓	0	0.1	0.2	0.3	0.4	0.5
0	0	0.0608	0.2528	0.6048	1.1648	2.0000
0.25	0.0265	0.0918	0.2986	0.6795	1.2886	2.2020
0.5	0.0595	0.1305	0.3558	0.7729	1.444	2.4564
0.75	0.1018	0.1800	0.4293	0.8934	1.6455	2.7882
1	0.1604	0.2485	0.5304	1.0592	1.9236	3.2476
1.25	0.2411	0.3433	0.6724	1.2959	2.3261	3.9205
1.5	0.3769	0.5026	0.9109	1.6944	3.008	5.0672
1.75	0.6710	0.8477	1.4288	2.5664	4.5140	7.6267

ence. The relative magnitude of these two quantities will give us an intuitive idea of the departure of the surface from spherical geometry. The equatorial circumference is given by

$$C_E = \frac{8\pi\sqrt{3}m}{q_s \sin\frac{2}{3}\lambda\,\{1+\beta^2/(Amq_s)^2\}}. \tag{25}$$

Equation (25) shows that C_E contracts as the magnetic distortion parameter β increases and $C_E \rightarrow 0$ as $\beta \rightarrow \infty$ for any value of the acceleration.

To analyze the effect of acceleration more closely, we construct a bivariate table (Table IV) showing various numerical values of C_E/m. Along each row, $54A^2m^2$ is constant while β varies from 0 to 0.5. Along each column, β is constant and $54A^2m^2$ varies from 0 to 1.75.

In Table V, C_E/m increases for each fixed value of β as $54A^2m^2$ increases from 0 to 1.75. However, it can be shown from Table IV that for sufficiently large values of β, the equatorial circumference decreases with the increase of $54A^2m^2$. The polar circumference C_P is given by

$$\frac{C_P}{m} = \frac{2\alpha}{Am} + \frac{2\delta}{(Am)^3}\beta^2, \tag{26}$$

where

$$\alpha = \int_{p_0}^{p_\pi} \frac{dp}{(p+q_s)\sqrt{G}}, \tag{27}$$

$$\delta = \int_{p_0}^{p_\pi} \frac{\sqrt{G}}{(p+q_s)^3}\,dp. \tag{28}$$

The integrand in (27) has singularities at both the end points as $G(p_0) = G(p_\pi) = 0$. But the integral is convergent for $A^2m^2 < \frac{1}{27}$. The integrand in (28) has singularities in its derivatives at both the end points. These two integrals can be evaluated numerically by Gauss–Legendre quadrature[12] with proper weight functions. It is obvious from (26) that C_P increases as β^2 for any fixed value of Am. The effect of an increase in the acceleration of the particle on the polar circumference can be seen from Table VI.

Table VI shows that for any fixed value of the distortion parameter β, the polar circumference increases as $54A^2m^2$ increases from 0 to 1.75. As $A^2m^2 \rightarrow \frac{1}{27}$, $C_P \rightarrow \infty$ as can easily be seen from (5), (9), and (26). We define a dimensionless quantity D by $D = (C_P - C_E)/C_E$. Then D will be a measure of deviation of the event horizon from spherical shape. D will depend on the acceleration as well as the magnetic distortion parameter β.

From Table VII it is obvious that $D > 0$ for any value of the magnetic distortion parameter and the acceleration $(A^2m^2 < \frac{1}{27})$ of the object. So the polar circumference always exceeds the equatorial circumference. Table VII shows that for any fixed value of β, D gradually increases as $54A^2m^2$ increases from 0 to 1.75. As $A^2m^2 \rightarrow \frac{1}{27}$, $D \rightarrow \infty$. Again for a fixed value of $54A^2m^2$, D increase with β. So the effect of increasing the acceleration of the object or the distortion parameter is to make the event horizon more and more prolate. For small values of A and β, the event horizon tends to assume prolate sphere-like shape.

D. Surface area

The surface area of the event horizon is given by

$$S = \iint (E^*G^*)^{1/2}\,dp\,d\phi$$
$$= \frac{8\sqrt{3}\pi}{Am\sin\frac{2}{3}\lambda}\left[\frac{p_\pi - p_0}{(p_0 - q_s)(p_\pi + q_s)}\right]m^2. \tag{29}$$

TABLE VIII. Expression of S for different choices $54A^2m^2$.

$54A^2m^2$	Surface area S
0	$16\pi m^2 \simeq 50.265\,81\,m^2$
0.25	$16\pi m^2 \simeq 56.011\,m^2$
0.5	$16\pi m^2 \simeq 63.376\,m^2$
0.75	$16\pi m^2 \simeq 73.201\,m^2$
1	$16\pi m^2 \simeq 87.062\,m^2$
1.25	$16\pi m^2 \simeq 108.348\,m^2$
1.5	$16\pi m^2 \simeq 146.108\,m^2$
1.75	$16\pi m^2 \simeq 237.712\,m^2$

TABLE IX. Expressions for $m^2K(\pi)$ as a function of β^2 for different choices of $54A^2m^2$.

$54A^2m^2$	$m^2K_{\sigma=\pi}$
0.25	$0.005\,227\,45 + 10.0791\,\beta^2$
0.75	$0.029\,6346 + 11.4983\,\beta^2$
1.25	$0.071\,6883 + 12.4351\,\beta^2$
1.75	$0.144\,6961 + 13.1754\,\beta^2$

TABLE X. Expression for $m^2K_{\sigma=\pi/2}$ for different choices of $54A^2m^2$.

$54A^2m^2$	$m^2K_{\sigma=\pi/2}$	β^2(β_c = critical value of β)
0.25	$\dfrac{655.240(\beta^2-0.0105)(\beta^2+0.000\,12)}{(1+184.167\beta^2)}$	0.0105
0.75	$\dfrac{164.868(\beta^2-0.0360)(\beta^2+0.0009)}{(1+52.038\beta^2)}$	0.0360
1.25	$\dfrac{74.693(\beta^2-0.0683)(\beta^2+0.0029)}{(1+26.377\beta^2)}$	0.0683
1.75	$\dfrac{36.225(\beta^2-0.1146)(\beta^2+0.0084)}{(1+14.833\beta^2)}$	0.1146

Equation (29) shows that the surface area S is independent of the magnetic field parameter and depends only on the mass and the acceleration of the object.

We observe that the area S increases with the increase of the acceleration (see Table VIII). As $A^2m^2 \to \frac{1}{27}$, $S \to \infty$. Thus the surface area of an accelerated black-hole in a uniform magnetic field always exceeds that of magnetized Schwarzschild black hole ($A = 0$, $B_0 \neq 0$).

IV. RINDLER SURFACE

For the Rindler surface the Gaussian curvature at the pole $\theta = \pi$ is

$$\begin{aligned} m^2\kappa_{\theta=\pi} = (Am)^2\ [(1+6Amp_\pi) \\ - 2Ar_{04}(p_\pi + 3Am\,p_\pi^2)] \\ + 8(p_\pi + 3Am\,p_\pi^2)^2\beta^2. \end{aligned} \tag{30}$$

Table IX shows that curvature increases as β^2 for a fixed value of Am. For fixed value of β, K increases with $54A^2m^2$.

The Gaussian curvature at the equator is

$$K_{\theta=\pi/2} = \frac{\left\{\dfrac{q_r^2+3}{q_r^4A^2m^2}\beta^4 - \dfrac{8}{q_r^2}\beta^2 + (Am)^2(1-q_r^2)\right\}}{\left\{1+\dfrac{\beta^2}{(Am\,q_r)^2}\right\}^4}. \tag{31}$$

We observe that the equatorial zone will exhibit negative Gaussian curvature if the magnetic parameter exceeds the corresponding critical value β_c (see Table X). The critical value β_c increases as $54A^2m^2$ increases from 0.25 to 1.75.

ACKNOWLEDGMENTS

The authors are grateful to the Government of Assam for all facilities provided at Cotton College, Gauhati-781 001 (India) for carrying out this piece of work. They thank Professor B. K. Barua for encouragement. One of them (S.C.) expresses her profound gratitude to CSIR, New Delhi, for financial assistance in the form of a senior research fellowship.

[1]T. Levi-Cività, Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Nat. Rend. **27**, 343 (1981).
[2]E. Newman and L. Tamburino, J. Math. Phys. **2**, 667 (1961).
[3]W. Kinnersley and M. Walker, Phys. Rev. D **2**, 1359 (1970).
[4]F. J. Ernst, J. Math. Phys. **17**, 515 (1976).
[5]H. Farhoosh and R. L. Zimmerman, Phys. Rev. D **23**, 299 (1981).
[6]H. Farhoosh and R. L. Zimmerman, Phys. Rev. D **22**, 797 (1980).
[7]W. J. Wild and R. M. Kerns, Phys. Rev. D **21**, 332 (1980).
[8]W. J. Wild, R. M. Kerns, and W. F. Drish, Jr., Phys. Rev. D **23**, 829 (1981).
[9]H. Farhoosh and R. L. Zimmerman, Phys. Rev. D **21**, 317 (1980).
[10]F. J. Ernst, J. Math. Phys. **17**, 54 (1976).
[11]L. Smarr, Phys. Rev. D **7**, 289 (1973).
[12]A. Ralston, *A First Course in Numerical Analysis* (McGraw-Hill Kogakusha, Tokyo, 1965), p. 100.

The metric dependence of four-dimensional formulations of electromagnetism

E. J. Post
8634-3 Pershing Drive, Playa del Rey, California 90291

(Received 25 January 1983; accepted for publication 4 March 1983)

The following is a comment on a recent paper by Birss [R. R. Birss, J. Math. Phys. **23**, 1153 (1982)] under the above title. The objective is to supplement the conclusion arrived therein and to confront an ensuing conflict between flux quantization and the recently revived hypothesis of magnetic charge quantization.

PACS numbers: 04.40. + c, 03.50.De

More than half a century ago Kottler,[1] Cartan,[2] and van Dantzig[3] (KCD) established, independently of one another, the premetric (i.e., metric-independent) properties of the *preconstitutive* (i.e., medium-independent) form of the Maxwellian laws. The discovery came as an *afterthought*; say, as an addendum to an early history of electromagnetic theory, which, until that time, had been firmly imbedded in the use of metric concepts pertaining to space as well as to space-time.

The premetric condition, seen in this light, is a necessary ingredient to secure a general applicability of the preconstitutive Maxwell laws to all conceivable media, including the medium known as free-space with its physical properties determined by the metric. Hence what emerged as an afterthought should, in retrospect, have been injected as a forethought when the theory was being constructed. In fact, Maxwell came very close to doing just that!

Birss makes this exchange of afterthought by forethought the basis of his philosophy of approach. In so doing, he finds that the Maxwellian laws can be made to be independent of the spatial metric, but not independent of the space-time metric. Since the KCD argument is not known to be in error, the physical origin of this defect in forethought-afterthought symmetry deserves further scrutiny. The following examples are believed to help in identifying the underlying causes.

Since counting quanta cannot be expected to depend on metric specifics, Birss' observations may be naturally illustrated by the London–Aharonov–Bohm (LAB) law counting magnetic flux quanta h/e:

$$\oint_{c_1} A = \frac{h}{e}\sum_k n_k \quad (c_1 = 1-\text{cycle residing in } dA = 0) \tag{1}$$

(A is the three-dimensional 1-form defined by the vector potential) and Gauss' law of electrostatics counting (stationary) charge quanta e:

$$\oint_{c_2} D = e\sum_k s_k \quad (c_2 = 2\text{-cycle residing in } dD = 0) \tag{2}$$

(D is the three-dimensional 2-form defined by the displacement vector). The sums Σn_k and Σs_k count the number of quanta linked or enclosed by c_1 and c_2, respectively.

At this point in time, only Eq. (2) stands firm as a law of physics. Equation (1) has a nearly firm status; there are questions though as to when the sum Σn_k involves integers or rational fractions. As good counting laws, though, Eqs. (1) and (2) both confirm Birss' observation on the preconstitutive Maxwellian laws' independence from three-dimensional metric specifics.

Yet neither Eq.(1) nor (2) are known to have conspicuous space-time counterparts that presently have found acceptance as counting quanta in the dynamic context of space-time. Birss' reservations with respect to a space-time metric independence of the preconstitutive Maxwellian laws reflect exactly this state of affairs. Let us examine the viability of such space-time extensions.

A space-time extension of Eq. (1) involves an integration of the electric field in the time direction and in a spatial direction. The value of such integrals can manifest a quantization if magnetic charge is locally absent. An example is discussed in Birss' Ref. 3, p. 3384. The ac Josephson effect can be shown to be a consequence of this type of "electroflux" quantization.[4] Note, however, that this electroflux quantization graduates from being a local *ad hoc* oddity to being a global law if magnetic charge is taken to be globally absent. In general, one can then speak of flux quantization as having magnetic-, electric-, or both components.

A space-time extension of Eq. (2) requires the counting of electrons that are in a collective dynamic state. A Gaussian form of Ampère's law is needed to do so; the latter is discussed in a recent paper by the present author.[5] In this manner, one can count the number of charge carriers participating in a current that is circulating in a ring that is being kept in a superconducting state. Since this number is an adiabatic invariant, temperature changes within the superconducting interval do not affect its magnitude. It then follows that in superconductors, the density of charge carriers in the super-current should be inversely proportional to the penetration depth of the super-current. A result of this kind is known, but obtained by a different rationale.

A comparison between the spatial and the space-time cases of counting quanta illustrates the essence of the metric distinction made by Birss. In the spatial cases quanta can frequently be directly perceived; say, in terms of discrete levels of magnetic flux, which can be recorded regardless of the choice of units or the calibration of the instrument, provided the sensitivity is there. In the space-time cases, quanta are being perceived indirectly through the intermediary of metric measurements; e.g., the ac Josephson effect requires a

 0022-2488/84/030612-02$02.50

frequency-plus-voltage measurement to identify a quantum. The metric specifics, though, drop out in the end and honor the KCD symmetry between space and space-time.

Let us now confront a major mathematical physical question which emerges from the preceding considerations. With the presently available abundance of experimental and theoretical evidence of an extended type of flux quantization, how can contemporary physics still tolerate the notion of a quantized magnetic charge? It has been known for a long time that the two are globally incompatible! Flux is the period of a 1-form A, magnetic charge is the period of 2-form, say F. Since F and A are definitely not independent but related by $F = dA$, it follows that F is exact and cannot have nonzero periods.

The simultaneous acceptance of flux and magnetic charge quantization is an amazing dichotomy of contemporary physics. While there is no rule against physics having its own little pet dichotomy, there is a common sense rule against not identifying a dichotomy when it has become apparent.

Proponents of magnetic charge can perhaps think of many reasons why the here-presented arguments do not apply. I could think of a few myself. Yet by the same token, I can think of reasons why they could apply; for one, the experimental situation of flux versus monopole observations is rather eloquent testimony to this effect. It is disturbing that monopole proponents rarely acknowledge these matters. In fact, I do not know of any instances where they are mentioned at all, but I will stand corrected if wrong.

[1]F. Kottler, Sitzungsber. Akad. Wien IIa **131**, 119 (1922).
[2]E. Cartan, Ann. Ec. Norm. Super. **41**, 1, 12 (1924).
[3]D. van Dantzig, Proc. Cambridge Philos. Soc. **30**, 421 (1934); Proc. Acad. Amsterdam **37**, 521, 526, 644, 825 (1934).
[4]E. J. Post, Phys. Rev. D **9**, 3379 (1974), see p. 3384. In fact, as shown in Phys. Lett. **92**, 224 (1982), the Bohr–Sommerfeld conditions are a case of electroflux quantization. Here again, the quanta are being perceived through the intermediary of metric measurements.
[5]E. J. Post, J. Math. Phys. **19**, 347 (1978).

The detailed balance condition in quantum statistical mechanics

W. A. Majewski[a)]
Department of Pure Mathematics, University of New South Wales, P.O. Box 1, Kensington 2033, Sydney, Australia

(Received 23 December 1980; accepted for publication 4 February 1983)

The generalization of the detailed balance condition is discussed in the framework of algebraic quantum statistical mechanics.

PACS numbers: 05.30. – d, 03.65.Fd

INTRODUCTION

A system in nonequilibrium statistical mechanics is not truly isolated being subject to usually uncontrollable random influences from outside. For this reason such models are called open systems. One of the ways of treating their interaction with the outside is to regard the open system as a subsystem of a larger closed system. On this basis one can derive various types of master equation describing the dynamics of the system, e.g., the Pauli equation. One of the consequences of such an approach is the fact that evolution is described by a dynamical semigroup. In classical statistical mechanics, in order to get a certain link between equilibrium and nonequilibrium statistical mechanics (the dynamics is still described by a classical version of a master equation) it is necessary to add an auxiliary condition—usually the detailed balance condition. This condition describes the fact that a steady state is maintained by "detailed balance."

In recent works[1,2] this pure classical concept was reformulated for the quantum mechanical case in terms of correlation functions of operators at two different times. The purpose of this paper is to give the generalization of Agarwal's definition for algebraic quantum statistical mechanics and to study its consequences. The main result is that the dynamical semigroup on W^*-algebra satisfies the detailed balance condition with respect to the faithful, normal, time-invariant state ω if there is a one-parameter $\mathscr{J}_\sigma$-self-adjoint semigroup acting over the Hilbert space of the representation associated with ω and preserving the natural cone and vacuum. Some conclusions also are given. Furthermore, we close our paper with a model.

1. DEFINITIONS AND NOTATIONS

Let $\mathfrak{A}$ be a W^*-algebra. A dynamical semigroup of $\mathfrak{A}$ is a σ-weak continuous one-parameter semigroup τ_t, $t \geqslant 0$, of positive identity preserving contractions on $\mathfrak{A}$ with $\tau_0 = I$ the identity map. Let ω be a faithful normal state on $\mathfrak{A}$ which is stationary under τ_t and denote by $(\mathscr{H}, \Pi, \Omega)$ the GNS representation associated with ω. For simplicity we will denote $\Pi(\mathfrak{A})$ by $\mathfrak{A}_\Pi$. As Ω is cyclic and separating for $\mathfrak{A}_\Pi$ there exists a modular operator Δ and also a modular conjugation $\mathscr{J}$ associated with the pair $(\mathfrak{A}_\Pi, \Omega)$ by Tomita–Takesaki theory. Furthermore, one can define the natural positive (self-dual) cone P in $\mathscr{H}$ by the following formula:

$$P = \overline{\Delta^{1/4}\mathfrak{A}_\Pi^+ \Omega} \quad (\mathfrak{A}_\Pi^+ = \{A \in \mathfrak{A}_\Pi,\ A \geqslant 0\}).$$

[a)] Postal address: Institute of Theoretical Physics and Astrophysics, Gdańsk University, Gdańsk, Wita Stwosza 57, Poland.

An operator $\mathscr{J}$ on a complex Hilbert space $\mathscr{H}$ is called a conjugation if it is an involution and $(\mathscr{J}x, \mathscr{J}y) = (y,x)$ for all $x,y \in \mathscr{H}$. Clearly, the modular conjugation $\mathscr{J}$ fulfills the above-given definition. On the other hand, we will deal with conjugation $\mathscr{J}_\sigma$ such that $\mathscr{J}_\sigma \mathfrak{A}_\Pi^+ \Omega \subseteq \mathfrak{A}_\Pi^+ \Omega$ (where $\mathfrak{A}_\Pi^+ = \{A \in \mathfrak{A}_\Pi, A \geqslant 0\}$). Therefore, to avoid future confusion, we emphasize that the conjugation $\mathscr{J}_\sigma$ is not the modular conjugation $\mathscr{J}$.

Following Glazman,[3] a linear operator A with a domain of definition $D(A)$ dense in $\mathscr{H}$ is said to be $\mathscr{J}$-self-adjoint if $\mathscr{J}A\mathscr{J} = A^*$ for some conjugation $\mathscr{J}$ on $\mathscr{H}$.

Let σ be an antilinear Jordan automorphism on $\mathfrak{A}$, i.e., σ is antilinear, one-to-one, onto, $*$-preserving map of $\mathfrak{A}$ such that $\sigma(AB + BA) = \sigma(A)\sigma(B) + \sigma(B)\sigma(A)$ for $A, B \in \mathfrak{A}$. If, additionally, $\omega(\sigma(A)\sigma(B)) = \omega(AB)$ and $\sigma \cdot \sigma = I$, we say that σ is a reversing operation for the triple $(\mathfrak{A}, \tau_t, \omega)$.

Remark: Arguments similar to those given in the next section (compare Lemma 1) imply that σ induces the conjugation $\mathscr{J}_\sigma$ on $\mathscr{H}$ such that $\mathscr{J}_\sigma \Delta^{it} \mathscr{J}_\sigma = \Delta^{-it}$. Therefore, the reversing operation σ acts on the modular (Hamiltonian) dynamics of the system $(\mathfrak{A}, \tau_t, \omega)$ but *does not change* the time direction of the semigroup τ_t.

Finally, let us note that the dynamical semigroup τ_t on $\mathfrak{A}$ induces in the representation $(\mathscr{H}, \Pi, \Omega)$ a semigroup $\hat{\tau}_t$ on $\mathscr{H}$ which enjoys the properties of dynamical semigroup, i.e., $\hat{\tau}_t$ is the weak continuous one-parameter uniformly bounded semigroup over $\mathscr{H}$ such that $\hat{\tau}_t\Omega = \Omega$ for $t \geqslant 0$. Its definition is given by

$$\hat{\tau}_t \Pi(A)\Omega = \Pi(\tau_t(A))\Omega$$

for all $A \in \mathfrak{A}$, $t \geqslant 0$.

Definition: We say the dynamical semigroup τ_t on $\mathfrak{A}$ satisfies the detailed balance condition with respect to the stationary, faithful, normal state ω on $\mathfrak{A}$ if the following equality is satisfied:

$$\omega(A^*\tau_t(B)) = \omega(\sigma(B^*)\tau_t\sigma(A))$$

for all $A, B \in \mathfrak{A}$, where σ is a reversing operation such that $\omega(\sigma(AB)) = \omega(\sigma(A)\sigma(B))$.

Remark: The above definition is a generalization of the Agarwal's definition of detailed balance for open Markovian systems.[1,2] To check this fact, it is enough to restrict the algebraic framework to the scheme of the ordinary quantum mechanics.

2. MAIN RESULT

The aim of this section is to prove:

Theorem: Let τ_t be a dynamical semigroup on the W^*-

 0022-2488/84/030614-03$02.50

algebra $\mathfrak{A}$ with normal, stationary, and faithful state ω. Further, let $(\mathscr{H}, \Pi, \Omega)$ denote the representation associated with ω, and let $\mathscr{J}_\sigma$, be a conjugation on $\mathscr{H}$ such that $\mathscr{J}_\sigma \Omega = \Omega$ and $\mathscr{J}_\sigma \mathfrak{A}_\Pi^+ \Omega \subseteq \mathfrak{A}_\Pi^+ \Omega$.

If τ_t satisfies the detailed balance condition with respect to ω, then there is $\mathscr{J}_\sigma$-self-adjoint semigroup $\hat{\tau}_t$, $\hat{\tau}_t \Pi(A)\Omega = \Pi(\tau_t(A))\Omega$, such that $\hat{\tau}_t P \subseteq P$, $\hat{\tau}_t \Omega = \Omega$, and $\hat{\tau}$ commutes strongly with the modular operator Δ.

Conversely, for an arbitrary $\mathscr{J}_\sigma$-self-adjoint dynamical semigroup $\hat{\tau}_t$ on $\mathscr{H}$, leaving invariant the cone P, $\tau_t \Omega = \Omega$, and, commuting strongly with the modular operator Δ for $t \geqslant 0$, there exists a dynamical semigroup τ_t on $\mathfrak{A}$ satisfying the detailed balance condition with respect to $\omega(\cdot) = (\Omega, \cdot \Omega)$ for the reversing operation σ induced by the conjugation $\mathscr{J}_\sigma$.

Remark: The proof will be divided into three lemmas. The conclusions and model are given in the next section.

Lemma 1: Adopt the assumptions of theorem. Further let τ_t satisfy the detailed balance condition. Then the following statements are valid:

(i) $\hat{\tau}_t$ is a $\mathscr{J}_\sigma$-self-semigroup, where

$$\hat{\tau}\Pi(A)\Omega = \Pi(\tau_t(A))\Omega.$$

(ii) $\hat{\tau}$ leaves invariant the cone P and strongly commutes with the modular operator Δ for all $t \geqslant 0$.

Proof: It is clear that the equality

$$\mathscr{J}_\sigma \Pi(A)\Omega = \Pi(\sigma(A))\Omega$$

gives the correct definition of antilinear operator on $\mathscr{H}$. Let us note

$$(\mathscr{J}_\sigma \Pi(A)\Omega, \mathscr{J}_\sigma \Pi(B)\Omega) = \omega(\sigma(A^*)\sigma(B)) = \omega(B^*A) = (\Pi(B)\Omega, \Pi(A)\Omega).$$

So one can extend $\mathscr{J}_\sigma$ to a conjugation on the Hilbert space $\mathscr{H}$ (for simplicity we will use the same symbol for the extension). Moreover,

$$(\hat{\tau}_t^* \Pi(A)\Omega, \Pi(B)\Omega) = \omega(A^* \tau_t(B))$$
$$= \omega(\sigma(B^*)\tau_t \sigma(A)) = (\mathscr{J}_\sigma \Pi(B)\Omega, \hat{\tau}_t \mathscr{J}_\sigma \Pi(A)\Omega)$$
$$= (\mathscr{J}_\sigma \hat{\tau}_t \mathscr{J}_\sigma \Pi(A)\Omega, \Pi(B)\Omega),$$

which proves the statement (i). The above arguments imply also that

$$\hat{\tau}_t^* \mathfrak{A}_\Pi^+ \Omega \subseteq \mathfrak{A}_\Pi^+ \Omega, \quad t \geqslant 0.$$

Since it is clear that, for all $t \geqslant 0$, $\hat{\tau}_t \mathfrak{A}_\Pi^+ \Omega \subseteq \mathfrak{A}_\Pi^+ \Omega$, statement (ii) follows directly from Lemmas 2 and 3 of the Bratteli and Robinson paper.[4]

Now we will study the implications of the assumed properties of semigroup $\hat{\tau}_t$.

Lemma 2: Let $\hat{\tau}_t$ be a dynamical semigroup on the Hilbert space $\mathscr{H}$ such that

(i) $\hat{\tau}_t^* P \subseteq P$ for all $t \geqslant 0$;

(ii) $\hat{\tau}_t$ commutes strongly with the modular operator Δ for all $t \geqslant 0$;

(iii) $\hat{\tau}_t \Omega = \Omega$, $t \geqslant 0$.

Then $\hat{\tau}_t$ induces a dynamical semigroup on $\mathfrak{A}$.

Proof: Let us note that assumption (i), Langer's theorem,[5] and the structure of the cone P imply

$$0 \leqslant (\Delta^{1/4} b\Omega, \hat{\tau}_t^* \Delta^{-1/4} a\Omega) = (b\Omega, \hat{\tau}_t^* a\Omega)$$

for all $a \in (\mathfrak{A}'_\Pi)^+$ and all $b \in \mathfrak{A}_\Pi^+$. One then has

$$0 \leqslant (b\Omega, \hat{\tau}_t^* a\Omega) = \lim(\Delta^{1/4} b\Omega, \Delta^{-1/4} a_\alpha \Omega)$$
$$= \lim(\Omega, a_\alpha^{1/2} b a_\alpha^{1/2} \Omega) \leqslant \|b\| \lim(\Omega, a_\alpha \Omega)$$
$$= \|b\| (\Omega, a\Omega),$$

where in the second step we have used $\{a_\alpha\} \subset (\mathfrak{A}'_\Pi)^+$ such that $\Delta^{-1/4} a_\alpha \Omega \to \hat{\tau}_t^* \Delta^{-1/4} a\Omega$, which is possible by assumption (i). Hence[6] there exists a $c_t \in \mathfrak{A}_\Pi^+$ such that $\|c_t\| \leqslant \|b\|$ and

$$(\hat{\tau}_t b\Omega, a\Omega) = (c_t \Omega, a\Omega)$$

for all $a \in \mathfrak{A}'_\Pi$. Now let us define

$$\Pi(\tau_t(A))\Omega = \hat{\tau}_t \Pi(A)\Omega \quad (= c_t \Omega).$$

The above equality gives the correct definition of linear positive maps on $\mathfrak{A}$ since Π is the faithful representation. But the semigroup properties are evident, weak and strong continuity are equivalent for semigroup so the proof of Lemma 2 is finished.

Now we will describe the reversing operation.

Lemma 3: Let $\mathfrak{R}$ be a von Neumann algebra on a Hilbert space $\mathscr{H}$ with a cyclic and separating vector Ω. If $\mathscr{J}_\sigma$ is a conjugation such that $\mathscr{J}_\sigma \Omega = \Omega$ and $\mathscr{J}_\sigma \mathfrak{R}^+ \Omega \subseteq \mathfrak{R}^+ \Omega$ ($\mathfrak{R}^+ = \{A \in \mathfrak{R}; A \geqslant 0\}$), then there exists a unique antilinear Jordan automorphism σ on $\mathfrak{R}$ such that $\mathscr{J}_\sigma A\Omega = \sigma(A)\Omega$ for all $A \in \mathfrak{R}$.

Proof: It is enough to use, with obvious modifications, arguments given in the Bratteli–Robinson book.[7]

Proof of Theorem: The first statement of theorem follows from Lemma 1. Further, $\hat{\tau}_t P \subseteq P$ for $t \geqslant 0$ implies $0 \leqslant (x, \hat{\tau}_t y) = (\hat{\tau}_t^* x, y)$ for all $x, y \in P$ and $t \geqslant 0$. Hence $\hat{\tau}_t^* P \subseteq P$ and, to end the proof of theorem, it is enough to show that the semigroup τ_t, described by Lemma 2, satisfies the detailed balance condition. Let us take the antilinear Jordan automorphisms σ given by Lemma 3. It is easy to see that $(\Omega, \sigma(A)\sigma(B)\Omega) = \overline{(\Omega, AB\Omega)} = \overline{\omega(AB)}$ and so we can consider σ as a reversing operation. Moreover,

$$\omega(A^* \tau_t(B)) = (\hat{\tau}_t^* \Pi(A)\Omega, \Pi(B)\Omega)$$
$$= (\mathscr{J}_\sigma \Pi(B)\Omega, \hat{\tau}_t \mathscr{J}_\sigma \Pi(A)\Omega) = \omega(\sigma(B^*)\tau_t \sigma(A)).$$

Hence the detailed balance condition is satisfied, and the proof is completed.

3. CONCLUSIONS AND MODEL

First, we want to point out that our theorem, giving the representation of a dynamical semigroup in terms of a $\mathscr{J}_\sigma$-self-adjoint semigroup on the Hilbert space, is in some sense the extension of the similar problem for reversible dynamical systems. Namely, each automorphism of a von Neumann algebra with cyclic and separating vector can be represented by a unitary operator with special properties.[8] To get a better understanding of the present situation, we want to present the following model.

Model: Let us consider an n-level system S, i.e., the system whose C^*-algbra of observables can be identified to the set $\mathscr{L}(\mathscr{H})$ of all linear operators on the n-dimensional Hilbert space, (with n a finite number). Further, let us assume that the dynamics is given by a completely positive dynamical semigroup τ_t. In particular, the infinitesimal generator of τ_t has the following general form[9]:

$$L: A \mapsto L(A) = -i[h,A] + \sum_{i,j=1}^{n^2-1} c_{ij}\{[f_i, Af_j^*] + [f_iA, f_j^*]\}$$

for $A \in \mathscr{L}(\mathscr{H})$, where $h = h^*$, $\operatorname{Tr} h = 0$, $\operatorname{Tr} f_i = 0$, $\operatorname{Tr} f_i^* f_j = \delta_{ij}$ for $i,j = 1,2,...,n^2-1$ and $\{c_{ij}\}_{i,j=n}^{n^2-1}$ is a complex positive matrix.

Let ρ be a strictly positive density matrix on $\mathscr{H}$. Then $\omega(A) = \operatorname{Tr} \rho A$, $A \in \mathscr{L}(\mathscr{H})$ defines the faithful (obviously normal) state on $\mathscr{L}(\mathscr{H})$. In order to define the reversing operator for the dynamical system $(\mathscr{L}(\mathscr{H}), \tau_t, \omega)$ let us choose a conjugation K on $\mathscr{H}$ such that $K\rho = \rho K$ (it is always possible). Then we define σ as follows $\sigma(A) = KAK$ for $A \in \mathscr{L}(\mathscr{H})$.

The equation defining the detailed balance can be rewritten in the terms of infinitesimal generator L of τ_t; then

$$\operatorname{Tr} \rho [L(B)]^* A = \operatorname{Tr} \rho B^* \sigma \circ L \circ \sigma(A).$$

Consequently, if one assumes

(1) $KhK = h$ and $[\tau, h] = 0$,

(2) $\sigma L_S \sigma = L_S$,

(3) $c_{ij} = c_{ji}$, $i,j = 1,2,...,n^2-1$, and (c_{ij}) real,

(4) $\omega(L_S(B)A) = \omega(BL_S(A))$,

where L_S denotes the "dissipative" part of the infinitesimal generator L, i.e., $L_S(A) = \frac{1}{2}\sum_{i,j=1}^{n^2-1} c_{ij}\{[f_i, Af_j^*] + [f_iA, f_j^*]\}$, then the detailed balance condition is satisfied for $(\mathscr{L}(\mathscr{H}), \tau_t)$ with respect to ω. Condition (1) is the restriction on the "Hamiltonian" part of dynamics while conditions (2), (3), and (4) are restrictions on the "dissipative" part of dynamics. Let us note that condition (2) is satisfied if, for example, $Kf_lK = f_l$ for indices l such that $c_{lk} \neq 0$, where k is an arbitrary index and we have used condition (3). The general characterization of operators L_S satisfying condition (4) was given by Alicki.[10]

ACKNOWLEDGMENTS

It is a pleasure to thank Professor D. W. Robinson for his kind hospitality at the University of New South Wales and for arranging my visit.

[1] G. Z. Agarwal, Z. Physik **258**, 409–22 (1973).

[2] H. J. Carmichael and D. F. Walls, Z. Physik B **23**, 299–306 (1976).

[3] I. M. Glazman, *Direct Methods of Qualitative Spectral Analysis of Singular Differential Operators* (Jerusalem, 1965); see definition on p. 76.

[4] O. Bratteli and D. W. Robinson, Ann. Inst. H. Poincaré **25**, 139–64 (1976).

[5] T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, New York, 1966); see Theorem 3.35, p. 281.

[6] O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics* (Springer-Verlag, New York, 1979), Vol. I; compare Theorem 2.3.19.

[7] See Theorem 3.2.18 in Ref. 6.

[8] A. Connes, Ann. Inst. Fourier, Grenoble **24**, 121–55 (1974), or Sec. 2.5.4 in Ref. 6.

[9] V. Gorini, A. Kossakowski, and G. G. Sudarshan, J. Math. Phys. **17**, 821–25 (1976).

[10] R. Alicki, Rep. Math. Phys. **10**, 249–58 (1976).

Borchers algebra formulation of an indefinite inner product quantum field theory

Lech Jakóbczyk
Institute of Theoretical Physics, University of Wrocław, 50-205 Wrocław, Cybulskiego 36, Poland

(Received 10 December 1982; accepted for publication 11 March 1983)

Formulation of an indefinite inner product quantum field theory in terms of so-called Strocchi–Wightman states is investigated. In particular, the reconstruction theorem for such states is discussed. We consider also general properties of field theory reconstructed from a Strocchi–Wightman state.

PACS numbers: 11.10.Cd

1. INTRODUCTION

It is known that quantum field theory can be defined by the set of Wightman functions $\{W_n(x_1,...,x_n)\}$.[1] From general physical assumptions concerning field theory, it follows that the set $W_n(x_1,...,x_n)$ ought to satisfy:

(I) $W_n(Lx_1,...,Lx_n) = W_n(x_1,...,x_n)$, $L\in P^{\uparrow}_{+}$;

(II) $W_n(x_1,...,x_i, x_{i+1},...,x_n) = W_n(x_1,...,x_{i+1}, x_i,...,x_n)$ if $(x_{i+1} - x_i)^2 < 0$;

(III) The Fourier transform $\widetilde{\mathscr{W}}_{n-1}(q_1,...,q_{n-1})$ of $\mathscr{W}_{n-1}(y_1,...,y_{n-1}) = W_n(x_1,...,x_n)$ with $y_i = x_i - x_{i+1}$, vanishes if q_i does not satisfy $q_i^2 \geqslant 0$, $(q_0)_i \geqslant 0$.

It is also assumed that the set $\{W_n(x_1,...,x_n)\}$ satisfies the following positive definiteness property:

$$\text{(IV)} \sum_{n,m} W_{n+m}(f_n^* \otimes f_m) \geqslant 0, \quad f_k \in S(R^{4K}). \tag{1.1}$$

(1.1) allows us to introduce the Hilbert space topology in the space of states reconstructed from $\{W_n(x_1,...,x_n)\}$.[2]

As was shown by Strocchi,[3] in quantum theory of gauge fields, the set $\{W_n(x_1,...,x_n)\}$ cannot satisfy (I)–(III) *and* (IV). A promising possibility is to give up positivity (IV) and preserve (I)–(III).[4] This choice has some advantages. For example, we can apply the technique of axiomatic field theory based on locality, spectrality, and Poincaré covariance.[5] In this context, Morchio and Strocchi[6] introduced the following *Hilbert space structure condition* to replace (IV):

(IV′) There exists a set of Hilbert seminorms $\{p_n\}$ defined on $S(R^{4n})$ such that

$$|W_{n+m}(f_n^* \otimes g_m)| \leqslant p_n(f_n) p_m(g_m). \tag{1.2}$$

We propose here another (stronger) Hilbert space structure condition. It is called *α-positivity* of $W_n(x_1,...,x_n)$ (see Sec. 2). $\{W_n(x_1,...,x_n)\}$ satisfying (I)÷(III) and α-positivity, define a functional ω on a Borchers algebra $A(S(R^4))$ (we will call it a *Strocchi–Wightman state*). This state defines, via GNS construction, a representation of $A(S(R^4))$ on an indefinite inner product space. A field theory defined by this representation will be called the *Strocchi–Wightman quantum field theory*.

In this paper we consider structural properties of the Strocchi–Wightman field theory. α-positive functionals on $A(S(R^4))$ are analyzed in detail in Sec. 2. It is shown that there exist α-positive functionals on $A(S(R^4))$ and that every such functional is continuous. An example of a Strocchi–Wightman state corresponding to free field theory is also constructed. Section 3 contains the general theory of the so-called $\mathscr{J}^*$ representations of a Borchers algebra on an indefinite inner product space. As a conclusion, we prove the reconstruction theorem for Strocchi–Wightman states. Simple structural properties of a field theory reconstructed from a Strocchi-Wightman state are collected in Sec. 4. For example, the *PCT* theorem, the spin and statistic theorem for scalar fields, and the general version of Haag's theorem are discussed. An outline of a theory of indefinite inner product spaces is presented in the Appendix.

2. STROCCHI–WIGHTMAN STATES

Let $A(S(R^4))$ be a Borchers algebra.[7] By definition, it is the set of sequences

$$a = (f_0, f_1(x_1),...,f_n(x_1,...,x_n),\cdots) \tag{2.1}$$

with properties:

(1) $f_0 \in C$, $f_n \in S(R^{4n})$;

(2) only a finite number of f_n's are different from zero.

Let

$$a + b = (f_0 + g_0, f_1 + g_1,...,f_n + g_n,\cdots) \tag{2.2}$$

$$za = (zf_0, zf_1,..., zf_n,\cdots) \tag{2.3}$$

$$ab = \Big(f_0 g_0, f_0 g_1 + g_1 f_1 g_0,..., \sum_{i+k=n} f_i(x_1,...,x_i) \times g_k(x_{i+1},...,x_n),\dots\Big) \tag{2.4}$$

$$a^* = (\bar f_0, \bar f_1(x_1),...\bar f_n(x_n,...,x_1),\cdots). \tag{2.5}$$

The space $A(S(R^4))$ is a topological $*$-algebra under the operations (2.2)–(2.4) with involution (2.5) and with topology of direct sum of $S(R^{4n})$. Let I: $S(R^4) \to S(R^4)$ be a continuous linear operator, such that $I^2 = 1$ and $I \neq 1$. Define a $*$-automorphism α_I of $A(S(R^4))$ by

$$\alpha_I(a) = (f_0, If_1,...,I_n f_n,\cdots), \tag{2.6}$$

where I_n is the extension by continuity of $\otimes^n I$. Let $\mathscr{J}_A$ be the set of all such $*$-automorphisms of $A(S(R^4))$.

Definition 2.1: A continuous linear functional ω on $A(S(R^4))$ is *α-positive* ($\alpha \in \mathscr{J}_A$) if, for every $a \in A(S(R^4))$, $\omega(\alpha(a^*)a) \geqslant 0$ and $\omega \circ \alpha = \omega$.

Definition 2.2: A continuous linear functional ω on $A(S(R^4))$ is called *Strocchi–Wightman state* ($\omega \in$ SW) if:

(1) $\omega(1) = 1$,

(2) $\omega \circ \tau_L = \omega$ for every $L \in P^{\uparrow}_{+}$,

(3) $\omega(I_{\rm loc}) = 0$,

(4) $\omega(I_{\rm sp}) = 0$,

(5) $P(\omega) = \{\alpha \in \mathscr{J}_A : \omega \text{ is } \alpha\text{-positive}\}$ is not empty,

 0022-2488/84/030617-06$02.50

where τ_L is a $*$-automorphism of $A(A(R^4))$ representing $L\in P^{\uparrow}_+$, $\tau_L(f_n(x_1,...,x_n)) = f_n(L^{-1}x_1,...,L^{-1}x_n)$, I_{loc} is the smallest closed two-sided ideal containing all elements of the form $ab-ba$ (a and b have spacelike separated supports), I_{sp} is the smallest closed left ideal containing the set of elements $\int d^4_x p(x)\tau_x(a)$, $a\in A(S(R^4))$ with $f_0=0$, and $p(x)\in$ Fourier transform of $\{f_1\in S(R^4): f_1(q)=0$, for $q\in V^+\}$. Simple properties of $\omega\in$ SW are collected in the following:

Lemma 2.1: If $\omega\in$ SW, then:

(1) ω is $*$-Hermitian,

(2) for every $\alpha\in P(\omega)$, $|\omega(a^*b)|^2 \leqslant \omega(\alpha(a^*)a)\omega(\alpha(b^*),b)$,

(3) for every $\alpha\in P(\omega)$, the set $L_\alpha(\omega)$ $=\{a\in A(S(R^4)):\omega(\alpha(a^*)a)=0\}$ is a closed left ideal in $A(S(R^4))$.

Proof: (1) Let $[a,b]_\alpha := \omega(\alpha(a^*)b)$. $[\cdot,\cdot]_\alpha$ is a sesquilinear positive-definite form on $A(S(R^4))$. It follows that $[a,b]_\alpha = \overline{[b,a]_\alpha}$, so $\overline{\omega(\alpha(b^*)a)} = \omega(\alpha(a^*)b)$ and $\omega(a^*) = \omega(a)$.

(2) By Cauchy–Schwarz inequality

$$|[a,b]_\alpha|^2 \leqslant [a,a]_\alpha[b,b]_\alpha$$

$$\begin{aligned}|\omega(a^*b)|^2 &= |\omega(\alpha\{(a^*)\}b)|^2 = |[\alpha(a),b]_\alpha| \\ &\leqslant [\alpha(a),\alpha(a)]_\alpha[b,b]_\alpha \\ &= \omega(\alpha(a^*)a)\omega(\alpha(b^*)b).\end{aligned}$$

(3) $L_\alpha(\omega)$ is a vector space. By (2) it is also a left ideal in $A(S(R^4))$. Let $\{a_n\}\subset L_\alpha(\omega)$ be a Cauchy sequence and let $a=\lim a_n$. $\omega(\alpha(b^*)a) = \lim\omega(\alpha(b^*)a_n)=0$ and $a\in L_\alpha(\omega)$. Let us define

$$A_{h,\alpha} = \{a\in A(S(R^4)):\alpha(a^*)=a\}, \tag{2.7}$$

$$A_{+,\alpha} = \overline{S}_\alpha \text{ with } S_\alpha = \{a\in A(S(R^4)):$$

$$a=\sum_j \alpha(b_j^*)b_j;\ b_j\in A(S(R^4))\}. \tag{2.8}$$

One can easily prove:

Lemma 2.2: (1) $A_{h,\alpha}$ is a closed subspace of $A(S(R^4))$.

(2) $A(S(R^4)) = A_{h,\alpha} + iA_{h,\alpha}$.

(3) $A_{h,\alpha} = A_{+,\alpha} - A_{+,\alpha}$.

It is known that $A(S(R^4))$ is a metrizable topological vector space.[8] $A_{h,\alpha}\subset A(S(R^4))$ with induced topology is also metrizable, so $A_{h,\alpha}$ is bornological.[9] For bornological vector spaces we have:

Property[10]: Let E be a bornological space and F a locally convex topological space. A linear mapping $u:E\to F$ is continuous iff for every bounded set $B\subset E$, $u(B)$ is bounded in F. Every bounded set $B\subset A_{h,\alpha}$ is of the form

$$B\subset B_1\cap A_{+,\alpha} - B_1\cap A_{+,\alpha},$$

where B_1 is bounded. Consider the bounded set $C\subset A_{+,\alpha}$. The α-positive functional ω is positive on elements from $A_{+,\alpha}$. One can show that ω is bounded on C.[11] Thus we have:

Corollary 2.1: If a functional ω on $A(S(R^4))$ is α-positive, then it is continuous on $A(S(R^4))$.

Let us consider now a structure of the set $P(\omega)$ ($\omega\in$SW). We see that if $\alpha\in P(\omega)$, then $\alpha_L := \tau_L\circ\alpha\circ\tau_L^{-1}\in P(\omega)$. If $\beta\in P(\omega)$ and $\beta\neq\alpha$, then either $\beta\in M(\alpha)=\{\alpha_L: L\in P^{\uparrow}_+\}$ or $M(\beta)\cap M(\alpha)=\varnothing$.

Lemma 2.3: $P(\omega)=\cup_\rho M_\rho$, where $\{M_\rho\}$ denotes the set of orbits of a Poincaré group $P^{\uparrow}_+$ in $\mathscr{J}_A$.

Now we construct an example of a state $\omega\in$ SW. Let $(K,\langle\cdot,\cdot\rangle)$ be a Krein space (Appendix). For any fundamental symmetry $I\in\mathscr{J}(K)$ the I-inner product turns K into a Hilbert space K_I. Let $\widetilde{F}_I$ be a Fock space over K_I. Define the sesquilinear form

$$\langle\chi,\psi\rangle_{\widetilde{F}_I} = (\chi,\Gamma(I)\psi)_{\widetilde{F}_I}, \tag{2.9}$$

where $\Gamma(I)$ is the second quantization of I. $(\widetilde{F}_I,\langle\cdot,\cdot\rangle_{\widetilde{F}_I})$ is a Krein space.[12] Now we construct the field operator $F(f)$, $f\in K$.[13] It is easy to check that:

(1) $F(f)$ is closable in $\widetilde{F}_I$.

(2) If $f_n\to f$ in K_I then $F(f_n)\chi\to F(f)\chi$ in $\widetilde{F}_I$ for any $\chi\in\widetilde{F}_0$ ($\widetilde{F}_0$ denotes a finite particle subspace).

(3) $\Omega_F=(1,0,...,0,...)\in\widetilde{F}_0$ is $\Gamma(I)$-cyclic, i.e., $\{F(f_1)\cdots F(f_n)\Omega_F;\ n=1,2,\cdots\}$ is total in $\widetilde{F}_I$.

(4) $F(f)F(g)\chi - F(g)F(f)\chi = i\,\mathrm{Im}\langle f,g\rangle\chi$, $\chi\in\widetilde{F}_0$.

Suppose that to any $L\in P^{\uparrow}_+$ there corresponds a linear operator $\mathscr{U}_L$ on K such that:

(a) $\langle\mathscr{U}_Lf,\mathscr{U}_Lg\rangle = \langle f,g\rangle$, $f,g\in K$,

(b) for every $L\in P^{\uparrow}_+$, $\mathscr{U}_L$ is bounded on K_I.

Let $\Gamma(\mathscr{U}_L)$ be the second quantization of $\mathscr{U}_L$. It may be that $\|\mathscr{U}_L\|_I>1$ so $\Gamma(\mathscr{U}_L)$ is generally defined only on $\widetilde{F}_0$.[14] One can easily check that

$$\Gamma(\mathscr{U}_L)F(f)\Gamma(\mathscr{U}_L)^{-1}\chi = F(\mathscr{U}_Lf)\chi,\quad \chi\in\widetilde{F}_0. \tag{2.10}$$

Consider now the concrete realization of the "one-particle" Krein space $(K,\langle\cdot,\cdot\rangle)$. Let

$$\begin{aligned}&K=L_2(H_0,d\Omega_0)\otimes C^4,\\ &H_0=\{p\in R^4: p^2=0, p_0>0\},\ d\Omega_0(p)=d^3p/|\vec{p}|\end{aligned} \tag{2.11}$$

and

$$\langle f,g\rangle = -\int_{H_0} d\Omega_0(p)\bar{f}_\mu(p)g^\mu(p),\quad \mu=0,1,2,3, \tag{2.12}$$

$$\mathscr{U}_L: f_\mu\to e^{ipx}\Lambda_\mu{}^\nu f_\nu(\Lambda^{-1}p),\ L=(\Lambda,x). \tag{2.13}$$

Let $\widetilde{F}_I, F, \Gamma(\mathscr{U}_L)$ be constructed as above. Define also a continuous mapping

$$\begin{aligned}&T: S^{(4)}(R^4)\to L_2(H_0,d\Omega_0)\otimes C^4,\\ &(Tf)_\mu(p) = \int e^{ipx}f_\mu(x)\,d^4x \restriction H_0,\end{aligned} \tag{2.14}$$

where

$$S^{(4)}(R^4) = S(R^4)\otimes C^4. \tag{2.15}$$

For $f_n(x_1,...,x_n) = f_1(x_1)\otimes\cdots\otimes f_n(x_n)$ with $f_k\in S^{(4)}(R^4)$ define

$$W_n(f_n) := \langle\Omega_F, F(Tf_1)\cdots F(Tf_n)\Omega_F\rangle_{\widetilde{F}_I}. \tag{2.16}$$

(2.16) is a tempered distribution in every f_k, so there exists tempered distribution W_n on $S^{(4)}(R^{4n})$ which is an extension of (2.16). Let

$$\omega_F(a) := \sum_n W_n(f_n). \tag{2.17}$$

Put

$$\tau_L: f_\mu(y)\to\Lambda_\mu{}^\nu f_\nu(\Lambda^{-1}(y-x)),\quad L=(\Lambda,x). \tag{2.18}$$

τ_L can be extended to a continuous $*$-automorphism of $A(S^{(4)}(R^4))$; moreover, $\mathscr{U}_L\circ T = T\circ\tau_L$.

Let $\mathscr{I}_0(K) \subset \mathscr{I}(K)$ be such that:

(1) for any $I \in \mathscr{I}_0(K)$, $IS^{(4)}(R^4) \subset S^{(4)}(R^4)$;

(2) for any $I \in \mathscr{I}_0(K)$, $I \circ T = T \circ I$.

Now we can prove the following:

Theorem 2.1: ω_F defined by (2.17) has the following properties:

(1) $\omega_F \circ \alpha_I = \omega_F$, $I \in \mathscr{I}_0(K)$;

(2) $\omega_F(\alpha_I(a^*)a) \geqslant 0$, $a \in A\,(S^{(4)}(R^4))$, $I \in \mathscr{I}_0(K)$;

(3) $\omega_F \circ \tau_L = \omega_F$, $L \in P^{\uparrow}_{+}$;

(4) $\omega_F(I_{\rm loc}) = 0$,

(5) $\omega_F(I_{\rm sp}) = 0$.

Hence $\omega_F \in$ SW with $P(\omega_F) = \{\alpha_I : I \in \mathscr{I}_0(K)\}$.

Proof: (1) For $\mathbf{f}_n(x_1,...,x_n)$ we have

$$\begin{aligned} W_n(I_n\mathbf{f}_n) &= W_n(If_1 \otimes \cdots \otimes If_n) \\ &= \langle \Omega_F, F(T \circ If_1) \cdots F(T \circ If_n)\Omega_F \rangle \\ &= \langle \Omega_F, F(I \circ Tf_1) \cdots F(I \circ Tf_n)\Omega_F \rangle \\ &= \langle \Omega_F, \Gamma(I) F(Tf_1) \cdots F(Tf_n)\Omega_F \rangle = W_n(\mathbf{f}_n). \end{aligned}$$

So $\omega_F \circ \alpha_I = \omega_F$.

(2) $\omega_F(\alpha_I(a^*)a) = \sum_{n,m} W_{n+m}(I_n f_n^* \otimes f_m) \geqslant 0$.

(3)–(5) can be proved as in the standard case.[15]

It is, of course, not proved that there exist *nontrivial* Strocchi–Wightman states. The existence problem is as complex as in the standard case[16] (probably even more complex). However, for a fixed $\alpha \in \mathscr{I}_A$, we can always construct many α-positive functionals on $A(S(R^4))$. The construction is the following: For every $b \in A\,(S(R^4))$ there exists a continuous linear functional $\omega_b^{(+)}$ on $A\,(S(R^4))$ such that $\omega_b^{(+)}(b) \neq 0$ and $\omega_b^{(+)}(a^*a) \geqslant 0$.[17] Fix $b \in A\,(S(R^4))$ and put

$$\omega^{(+)(\alpha)} := \tfrac{1}{2}\{\omega_b^{(+)}(\alpha(a)) + \omega_b^{(+)}(a)\}. \tag{2.19}$$

For $a \in A_{+,\alpha}$ of the form $a = \Sigma_j^{\alpha}(b_j^*)b_j$ define

$$\omega^{(\alpha)}(a) := \Sigma_j\, \omega^{(+)(\alpha)}(b_j^* b_j). \tag{2.20}$$

For the remaining $a \in A_{+,\alpha}$, $\omega^{(\alpha)}$ is the extension of (2.20). If $a = \sum_k z_k a_k$, $z_k \in C$, $a_k \in A_{+,\alpha}$,

$$\omega^{(\alpha)}(a) := \sum_k z_k \omega^{(\alpha)}(a_k). \tag{2.21}$$

(2.21) defines $\omega^{(\alpha)}$ for any $a \in A\ (S(R^4))$ since, according to Lemma 2.2. $a = a_1 - a_2 + ia_3 - ia_4$ with $a_k \in A_{+,\alpha}$. Such defined $\omega^{(\alpha)}$ is linear, continuous and α-positive. Hence:

Theorem 2.2: For every positive functional ω on $A\,(S(R^4))$ and every $\alpha \in \mathscr{I}_A$ we can construct α-positive functional $\omega^{(\alpha)}$.

3. RECONSTRUCTION THEOREM

Now we want to prove that every Strocchi–Wightman state ω on $A\,(S(R^4))$ defines a field theory with an indefinite inner product state space. To do this, we have to develop first general theory of representations of Borchers algebra on an indefinite inner product space.

Definition 3.1: A pair (R, D_R) is called a *$\mathscr{I}^*$-representation* of $A\,(S(R^4))$ if:

(1) $R: A\,(S(R^4)) \rightarrow {\rm op}(D_R)$ is a mapping of $A\,(S(R^4))$ into linear operators all defined on a linear space D_R with an indefinite, nondegenerate inner product $\langle \cdot,\cdot \rangle_R$; R is such that:

(1a) for any $a \in A\,(S(R^4))$, $R_a D_R \subset D_R$;

(1b) for every $a,b \in A\,(S(R^4))$, $w,z \in C$, $k \in D_R$,

$R_{ab}k = R_a R_b k$, $R_{za+wb}k = zR_a k + wR_b k$;

(1c) for every $a \in A\,(S(R^4))$, $k,l \in D_R$,

$\langle k, R_a l \rangle_R = \langle R_{a^*} k, l \rangle_R$.

(2) There exists a set $\mathscr{I}_R \subset \mathscr{I}_A$ such that:

(2a) for any $\alpha \in \mathscr{I}_R$ there is a linear operator $I_R(\alpha): D_R \rightarrow D_R$ with properties: $I_R(\alpha)^2 = 1$,

$\langle I_R(\alpha)k, l \rangle_R = \langle k, I_R(\alpha) l \rangle_R$, $(\cdot,\cdot)_\alpha = \langle \cdot, I_R(\alpha) \cdot \rangle_R$

is positive definite.

(2b) for every $\alpha \in \mathscr{I}_R$, $a \in A\,(S(R^4))$, $k \in D_R$,

$R_{\alpha(a)}k = I_R(\alpha) R_a I_R(\alpha) k$.

Remarks: (1) Let $K_\alpha = \overline{D}_R^{\|\cdot\|_\alpha}$ $[\,\|\cdot\|_\alpha^2 = (\cdot,\cdot)_\alpha\,]$ and $\langle \cdot,\cdot \rangle_\alpha$ is the extension of $\langle \cdot,\cdot \rangle_R$ to K_α. $(K_\alpha, \langle \cdot,\cdot \rangle_\alpha)$ is a Krein space.

(2) For a fixed $\alpha \in \mathscr{I}_R$, (R, D_R) can be considered as a standard representation of $A\,(S(R^4))$, considered as the algebra with involution $a \rightarrow (a)^\alpha := \alpha(a^*)$ in a Hilbert space $(K_\alpha, (\cdot,\cdot)_\alpha)$.

Definition 3.2: A $\mathscr{I}^*$-representation (R, D_R) is called *α-cyclic* $(\alpha \in \mathscr{I}_R)$ if there is such a vector $k_0 \in D_R$ that:

(1) $\{R_a k_0 : a \in A\,(S(R^4))\}$ is dense in K_α.

(2) $I_R(\alpha) k_0 = k_0$.

Let (R, D_R) be α-cyclic; then the functional

$$\omega_0(a) := \langle k_0, R_A k_0 \rangle_R \tag{3.1}$$

is α-positive. The following proposition shows that every α-positive functional ω on $A\,((S(R^4))$ defines a $\mathscr{I}^*$-representation of $A\,(S(R^4))$.

Proposition 3.1: To every α-positive functional ω on $A\,(S(R^4))$ there corresponds the $\mathscr{I}^*$-representation $(R^{(\omega)}, D_{R^{(\omega)}})$ of $A\,(S(R^4))$. It is α-cyclic for any $\alpha \in \mathscr{I}_{R^{(\omega)}}$.

Proof: It is easy to show that $L(\omega) = \{a \in A\,(S(R^4)) : \omega(ba) = 0;\ b \in A\,(S(R^4))\} = L_\alpha(\omega)$. Put

$$D_{R^{(\omega)}} := A\,(S(R^4))/L(\omega), \tag{3.2}$$

$$\langle a_\omega, b_\omega \rangle_\omega := \omega(a^*b),\ \ a_\omega, b_\omega \in D_{R^{(\omega)}}, \tag{3.3}$$

$$R_a^{(\omega)} b_\omega := (ab)_\omega, \tag{3.4}$$

$$I_{R^{(\omega)}}(\alpha) b_\omega := (\alpha(b))_\omega. \tag{3.5}$$

Let $\mathscr{I}_{R^{(\omega)}} = P(\omega)$. Since $\{R_a^{(\omega)} 1_\omega : a \in A\,(S(R^4))\} = D_{R^{(\omega)}}$, this representation is α-cyclic for any $\alpha \in \mathscr{I}_{R^{(\omega)}}$.

Definition 3.3: Let (R, D_R) and $(\overline{R}, D_{\overline{R}})$ be such $\mathscr{I}^*$-representations of $A\,(S(R^4))$ that $\mathscr{I}_R = \mathscr{I}_{\overline{R}}$. Assume that for every $\alpha \in \mathscr{I}_R$ these representations are α-cyclic. Such two representations are called *equivalent* if there is a linear operator

$$V: D_R \rightarrow D_{\overline{R}}$$

with properties:

(1) $VD_R = D_{\overline{R}}$;

(2) $\langle Vk, Vl \rangle_{\overline{R}} = \langle k, l \rangle_R$, $k,l \in D_R$;

(3) for every $k \in D_R$, $a \in A\,(S(R^4))$, $VR_a k = \overline{R}_a Vk$;

(4) for every $\alpha \in \mathscr{I}_R$, $VI_R(\alpha)V^* = I_{\overline{R}}(\alpha)$.

It is easy to prove the following:

Proposition 3.2: Assume that the $\mathscr{J}^*$-representation (R,D_R) is α-cyclic for every $a \in \mathscr{J}_R$. If there is such an α-positive functional ω on $A(S(R^4))$ that $P(\omega) = \mathscr{J}_R$ and

$$\langle k_0,R_a\, k_0\rangle_R = \langle 1_\omega,R_a^{(\omega)}1_\omega\rangle_\omega, \tag{3.6}$$

then the $\mathscr{J}^*$-representations (R,D_R) and $(R^{(\omega)},D_{R^{(\omega)}})$ are equivalent.

Remark: From the above proposition follows that α-positive functional defines $\mathscr{J}^*$-representation up to equivalence.

Now we are prepared to formulate and prove the reconstruction theorem for $\omega \in \mathrm{SW}$.

Theorem 3.1: To every $\omega \in \mathrm{SW}$ and $\alpha \in P(\omega)$ there correspond

$$\{(K_\alpha,\langle\ ,\ \rangle_\alpha), D_0, F, \Omega_0, \{\mathscr{U}(L): L \in P_+^\uparrow\}\}$$

such that

(1) $(K_\alpha,\langle\ ,\ \rangle_\alpha)$ is a Krein space.

(2) $D_0 \subset K_\alpha$ is a dense subspace.

(3) F: $S(R^4) \to \mathrm{op}(D_0)$ is a linear mapping into linear operators all defined on D_0, with properties:

(3a) for every $f \in S(R^4)$, $F(f)D_0 \subset D_0$;

(3b) a mapping $f \to \langle F(f)k,l\rangle_\alpha$ is continuous for every $k, l \in D_0$;

(3c) $\langle F(f)k,l\rangle_\alpha = \langle k,F(\bar{f})l\rangle_\alpha$ for every $k, l \in D_0$;

(3d) there is an $I(\alpha) \in \mathscr{J}(K_\alpha)$ such that, for every $f \in S(R^4)$, $k \in D_0$,

$F(\alpha(f))k = I(\alpha)F(f)I(\alpha)k$;

(3e) $F(f_1)F(f_2)k - F(f_2)F(f_1)k = 0$ for $f_1, f_2 \in S(R^4)$ with spacelike separated supports.

(4) $\Omega_0 \in D_0$ is such that $\{F(f_1)\cdots F(f_n)\Omega_0;\ n = 1,2,\cdots\}$ is total in K_α.

(5) $\{\mathscr{U}(L): L \in P_+^\uparrow\}$ is a representation of a Poincaré group $P_+^\uparrow$ with properties:

(5a) $\mathscr{U}(L)D_0 \subset D_0$ for every $L \in P_+^\uparrow$;

(5b) $\langle \mathscr{U}(L)k,\mathscr{U}(L)l\rangle_\alpha = \langle k,l\rangle_\alpha$ for every $k,l \in D_0$;

(5c) $\mathscr{U}_L: K_\alpha \to K_{\alpha_L}$ and $\|\mathscr{U}_L k\|_{\alpha_L} = \|k\|_\alpha$ for every $k \in K_\alpha$;

(5d) a mapping $L \to \langle \mathscr{U}(L)k,l\rangle_\alpha$ is continuous for every $k,l \in D_0$;

(5e) for every $f \in S(R^4)$, $k \in D_0$, $L \in P_+^\uparrow$, $\mathscr{U}(L)F(f)\mathscr{U}(L)^*k = F(f_L)k$ with $f_L(y) = f(\Lambda^{-1}(y-x))$, $L = (\Lambda,x)$.

(6) $\int p(x)\langle k,\mathscr{U}(x)l\rangle_\alpha\, d^4x = 0$ for $p \in$ Fourier transform of $\{f \in S(R^4): f(q) = 0 \text{ for } q \in V^+\}$.

Proof: By Proposition 3.1 there is the $\mathscr{J}^*$-representation $(R^{(\omega)},D_{R^{(\omega)}})$ defined by $\omega \in \mathrm{SW}$.

(1) and (2) Put $D_0 = D_{R^{(\omega)}}$ and $K_\alpha = \overline{D}_{R^{(\omega)}}^{\|\cdot\|_\alpha}$.

(3) Put $F(f) = R_a^{(\omega)}$ with $a \in A_1(S(R^4))$. By $A_1(S(R^4))$ we denote the set of elements $a \in A(S(R^4))$ of the form

$$a = (0,f(x),0,\ldots,0,\cdots).$$

(3a)–(3d) follow from Proposition 3.1.

(3e) Let $a, b \in A_1(S(R^4))$ have spacelike separated supports. For every $c, d \in A(S(R^4))$

$$\langle d_\omega,[R_a^{(\omega)},R_b^{(\omega)}]c_\omega\rangle_\omega = \omega(d^*(ab-ba)c).$$

Since ab-$ba \in I_{\mathrm{loc}}$, $\omega(d^*(ab-ba)c) = 0$ and $[R_a^{(\omega)},R_b^{(\omega)}]c_\omega \in D_{R^{(\omega)}}^\perp$. $D_{R^{(\omega)}}$ is dense in K_α; hence $D_{R^{(\omega)}}^\perp = \{0\}$.[18] Thus $[F(f_1),F(f_2)]k = 0$ for $f_1,f_2 \in S(R^4)$ with spacelike separated supports and $k \in D_0$.

(4) Put $\Omega_0 := 1_\omega$. Since the representation $(R^{(\omega)},D_{R^{(\omega)}})$ is α-cyclic $\{F(f_1)\cdots F(f_n)\Omega_0;\ n = 1,2,\cdots\}$ is total in K_α.

(5) Since ω is $P_+^\uparrow$-invariant, there exists a family

$$\{\mathscr{U}(L): L \in P_+^\uparrow\};\quad \mathscr{U}(L)b_\omega := (\tau_L(b))_\omega.$$

(5a)–(5e) are trivially fulfilled.

(6) $\int p(x)\langle k,\mathscr{U}(x)l\rangle_\alpha d^4x = \omega(a^*\int d^4x\, p(x)\tau_x(b)) = 0$ with $k = a_\omega$, $l = b_\omega$.

Remarks: (1) A standard (Wightman) quantum field theory is defined by a field operator, a space of states, and a representation of *the Poincaré* group. In the case of Strocchi–Wightman field theory, we must also specify an indefinite inner product in a space of states. Thus, in general, to every $\alpha \in P(\omega)$ there corresponds the field theory with the space of states $(K_\alpha,\langle\cdot,\cdot\rangle_\alpha)$.

(2) A representation $\{\mathscr{U}(L): L \in P_+^\uparrow\}$ is generally defined only on a dense domain D_0. It is connected with the fact that α does not commute with every $*$-automorphism from $\{\tau_L: L \in P_+^\uparrow\}$.[19]

4. ELEMENTARY STRUCTURAL PROPERTIES OF STROCCHI–WIGHTMAN FIELD THEORY

In this section we collect some structural properties of the field theory defined by Theorem 3.1. Since the theory is local and the spectral property is fulfilled, we are able to adopt some standard results of axiomatic quantum field theory. The main point is that the Wightman functions of this theory are also boundary values of analytic functions.[20]

A. Theorem of Reeh and Schlieder[21]

Theorem 4.1: Let $\mathscr{O}$ be an open set in R^4. Vectors of the form

$$\sum_j F(f_1^{(j)}\cdots F(f_j^{(j)})\Omega_0 \tag{4.1}$$

with supp $f_j^{(k)} \subset \mathscr{O}$ are dense in K_α.

Proof: Let L denote the set of vectors (4.1). If $k \in L^\perp$, i.e., $\langle k,l\rangle_\alpha = 0$ for every $l \in K$, then $k = 0$.[22] Since K_α is a Krein space, L is dense in K_α (Appendix).

B. *PCT* invariance

Consider the case of scalar field. By standard arguments[23] it follows that

$$\langle \Omega_0,F(x_1)\cdots F(x_n)\Omega_0\rangle_\alpha = \langle \Omega_0,F(-x_n)\cdots F(-x_1)\Omega_0\rangle_\alpha. \tag{4.2}$$

Define

$$\delta: f_n(x_1,\ldots,x_n) \mapsto \bar{f}_n(-x_1,\ldots,-x_n). \tag{4.3}$$

(4.2) can be expressed as

$$\omega(\delta(a)) = \omega(a^*). \tag{4.4}$$

Hence we have:

Theorem 4.2: If $\omega \in \mathrm{SW}$, then there is an operator Θ such that:

(1) Θ: $D_0 \to D_0$;

(2) $\langle \Theta k, \Theta l\rangle = \langle \overline{k,l}\rangle, k,l \in D_0$;
(3) if δ does not commute with $\alpha \in P(\omega)$, then

$\Theta: K_\alpha \to K_\beta,\ \|\Theta k\|_\beta = \|k\|_\alpha$ with $\beta = \delta \circ \alpha \circ \delta^{-1}$.

C. Spin and statistics theorem for the scalar Strocchi–Wightman field

Theorem 4.3: Let F be the scalar field. Suppose that

$$F(x)F(y)^* + F(y)^*F(x) = 0 \text{ if } (x-y)^2 < 0. \tag{4.5}$$

Then $F(f)\Omega_0 \in K_\alpha^0 = \{k \in K_\alpha: \langle k,k\rangle_\alpha = 0\}$.

Proof: (4.5) leads to equality[24]

$$\langle \Omega_0, F(x)F(y)^*\Omega_0\rangle_\alpha + \langle \Omega_0, F(-y)^*F(-x)\Omega_0\rangle_\alpha = 0.$$

It is equivalent to

$$\omega(aa^*) + \omega(\delta(aa^*)) = 0, \quad a = (0,f(x),0,\dots,0,\cdots).$$

By PCT invariance $\omega(\delta(aa^*)) = \omega(aa^*)$, so $\omega(aa^*) = 0$ and $F(f)\Omega_0 \in K_\alpha^0$.

Remark: In the Strocchi–Wightman field theory there may exist scalar fields which anticommute on spacelike distances, since from $F(f)\Omega_0 \in K_\alpha^0$ it does not follow that $F(f)k = 0$ for every $k \in D_0$. In a formal canonical quantization procedure such fields are called "ghost fields."

Now we pass to the theorem which may be called the "abstract version" of Haag's theorem.

D. Theorem 4.4

Let $\omega_1, \omega_2 \in \mathrm{SW}$. Suppose that $P(\omega_1) = P(\omega_2)$ and ω_1 is such that, in the representation of $A\,(S(R^4))$ defined by ω_1, there is only one invariant state (for some symmetry group G). Let $(R^{(\omega_i)}, D_{R^{(\omega_i)}})$ be $\mathscr{J}^*$-representations of $A\,(S(R^4))$ defined by ω_i $(i = 1,2)$. Similarly, let $\{\mathscr{U}_0^{(i)}(g): g \in G\}$ $(i = 1,2)$ be a representation of G. Assume that there is a family $\{\mathscr{U}^{(i)}(t)\}_{t \in R}$ $(i = 1,2)$ of operators such that

$$\mathscr{U}^{(i)}(t): D_{R^{(\omega_i)}} \to D_{R^{(\omega_i)}}, \tag{4.6}$$

$$\langle \mathscr{U}^{(i)}(t)k_i, \mathscr{U}^{(i)}(t)l_i\rangle_{\omega_i} = \langle k_i, l_i\rangle_{\omega_i},$$

$$k_i, l_i \in D_{R^{(\omega_i)}}. \tag{4.7}$$

Let us define

$$R_a^{(\omega_i)}(t) := \mathscr{U}^{(i)}(t) R_a^{(\omega_i)} \mathscr{U}^{(i)}(t)^*. \tag{4.8}$$

Suppose that for $t = t_0$ there is a mapping

$$V(t_0): D_{R^{(\omega_1)}} \to D_{R^{(\omega_2)}} \tag{4.9}$$

satisfying

$$\langle V(t_0)k_1, V(t_0)l_1\rangle_{\omega_2} = \langle k_1, l_1\rangle_{\omega_1}, \quad k_1, l_1 \in D_{R^{(\omega_1)}}, \tag{4.10}$$

$$V(t_0) I_{R^{(\omega_1)}}(\alpha) V(t_0)^* = I_{R^{(\omega_2)}}(\alpha), \quad \alpha \in P(\omega_1), \tag{4.11}$$

$$V(t_0) R_a^{(\omega_1)}(t_0) V(t_0)^* k_2 = R_a^{(\omega_2)}(t_0) k_2, \quad k_2 \in D_{R^{(\omega_2)}}. \tag{4.12}$$

Then $\omega_1 = \omega_2$.

Proof: Our argumentation is based on the proof of analogous theorem in the book of Emch.[25] Let $V(t) = \mathscr{U}^{(2)}(t-t_0) V(t_0) \mathscr{U}^{(1)}(t_0 - t)$. It is obvious that

$$\langle V(t)k_1, V(t)l_1\rangle_{\omega_2} = \langle k_1, l_1\rangle_{\omega_1} \tag{4.13}$$

and

$$V(t) I_{R^{(\omega_1)}}(\alpha) V(t)^* = I_{R^{(\omega_2)}}(\alpha). \tag{4.14}$$

Define $V := V(0)$. Of course, $R_a^{(\omega_2)} = V R_a^{(\omega_1)} V^*$. Let $\widetilde{\Omega}_0^1 = V^*\Omega_0^2$, where Ω_0^2 denotes a cyclic vector for $(R^{(\omega_2)}, D_{R^{(\omega_2)}})$. Let

$$\widetilde{\omega}_1(a)' = \langle \widetilde{\Omega}_0^1, R_a^{(\omega_1)} \widetilde{\Omega}_0^1\rangle_{\omega_1}. \tag{4.15}$$

Since $\langle \widetilde{\Omega}_0^1, R_a^{(\omega_1)}\widetilde{\Omega}_0^1\rangle_{\omega_1} = \langle \Omega_0^2, R_a^{(\omega_2)}\Omega_0^2\rangle_{\omega_2}$, $\widetilde{\omega}_1 = \omega_2$. $\widetilde{\omega}_1$ is G-invariant since ω_2 is G-invariant. If ω_1 is the only G-invariant state in the representation $(R^{(\omega_1)}, D_{R^{(\omega_1)}})$ then $\widetilde{\omega}_1 = \omega_1$. Hence $\omega_1 = \omega_2$.

Remark: Let ω_1 correspond to the free field theory. It is known that there are models of indefinite inner product free field theory with nonunique vacuum.[26] Thus, the main assumption of Theorem 4.4 may not be fulfilled.

ACKNOWLEDGMENTS

Discussions with Professor F. Strocchi, Professor G. Morchio and Professor W. Karwowski are kindly acknowledged.

APPENDIX

We collect here some facts concerning a theory of indefinite inner product spaces.[27]

1. Indefinite inner product spaces

Let E be a vector space with an inner product $\langle\cdot,\cdot\rangle$. A pair $(E,\langle\cdot,\cdot\rangle)$ will be called an inner product space. If the quantity $\langle x,x\rangle$, $x \in E$, may be positive, negative, or zero, $(E,\langle\cdot,\cdot\rangle)$ is called an *indefinite inner product space*. $(E,\langle\cdot,\cdot\rangle)$ is *decomposable* and *nondegenerate* if

$$E = E^+ \oplus E^-, \quad E^{\pm} = \{x \in E: \langle x,x\rangle \gtrless 0\}. \tag{A1}$$

2. Topology on an indefinite inner product space

Let T be a topology on $(E,\langle\cdot,\cdot\rangle)$. We say that T is a *majorant* of the inner product $\langle\cdot,\cdot\rangle$ if

—T is locally convex,

—$\langle\cdot,\cdot\rangle$ is jointly T-continuous.

A topology T is said to be *admissible* if:

—for any fixed $y \in E, x \to \langle y,x\rangle$ is T-continuous,

—for any T-continous linear form l on E, there is an element $y_0 \in E$ such that $l(x) = \langle y_0, x\rangle$.

Let $F \subset E$ be the subspace of E. Define $F^\perp = \{x \in E: \langle x,y\rangle = 0; y \in F\}$.

If T is admissible, the T-closure of any subspace F of E coincides with $F^{\perp\perp}$.

A topology T is called a *Hilbert majorant* if it is defined by a Hilbert space norm. A Hilbert majorant topology is admissible.

Let $F \subset E$ be such that $\langle\cdot,\cdot\rangle$ is definite on F. The equality

$$|x|_F^2 := |\langle x,x\rangle|, \quad x \in F, \tag{A2}$$

defines the norm on F. Topology induced by $|\cdot|_F$ is called an intrinsic topology on F. If an indefinite inner product space $(E,\langle\cdot,\cdot\rangle)$ admits a decomposition of the form $E = E^+ \oplus E^-$, where E^+ and E^- are intrinsically complete, then it is called a *Krein space*.

Let $P^{\pm}E = E^{\pm}$. We set

$$I = P^+ - P^- \tag{A3}$$

and say that I is a *fundamental symmetry* of E. For any $I \in \mathscr{J}(E)$ [$\mathscr{J}(E)$ denotes the set of all fundamental symmetries of E] define the *I-inner product*

$$(x,y)_I := \langle x,Iy \rangle. \tag{A4}$$

A decomposable, nondegenerate, indefinite inner product space $(E,\langle \cdot,\cdot \rangle)$ is a Krein space iff, for every $I \in \mathscr{J}(E)$, the I-inner product turns E into a Hilbert space.

[1]R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).
[2]See Ref. 1, Chap. 3.
[3]F. Strocchi, "Gauss' law in local quantum field theory," in *Field Theory, Quantization and Statistical Physics*, edited by E. Tirapegi (D. Reidel, Dordrecht, 1981).
[4]For an interesting discussion of this point see G. Morchio and F. Strocchi, Ann. Inst. H. Poincaré **33**, 251 (1980).
[5]See Ref. 1.
[6]See Ref. 4.
[7]H. J. Borchers, "Algebraic aspects of Wightman field theory," in *Statistical Mechanics and Field Theory*, edited by R. N. Sen and C. Weil (Israel U.P., Jerusalem, 1972).
[8]W. Wyss, Commun. Math. Phys. **27**, 223 (1972).
[9]H. H. Schaefer, *Topological Vector Spaces* (McMillian, New York, 1969).
[10]See Ref. 9, Chap. 2.
[11]See Ref. 9, Chap. 5.
[12]L. Jakobczyk, Bull. Acad. Pol.: Phys. Astron. **28**, 167 (1980).
[13]M. Mintchev, J. Phys. A **13**, 1841 (1980); see also Ref. 12.
[14]If A is such that $\|A\| > 1$, then $\Gamma(A)$ is not bounded operator on a Fock space [E. Nelson, J. Funct. Anal. **12**, 211 (1973)].
[15]See, e.g., M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. II, Chap. X.7.
[16]G. Hofman and G. Lassner, "Existence proofs for Wightman type functionals," preprint, JINR, Dubna, 1973.
[17]G. Lassner and A. Uhlmann, Commun. Math. Phys. **7**, 152 (1968).
[18]Since the topology defined by $\|\cdot\|_\alpha$ is admissible $D_{R^{(\omega)}}$ is dense in K_α iff $D^{\perp}_{R^{(\omega)}} = \{0\}$ (Appendix).
[19]In the case of the algebra $A(S^{(4)}(R^4))$ the $*$-automorphism α_g, defined by $\alpha_g(a) := (f_0, \mathbf{g}f_1, \dots, \mathbf{g}^{\otimes n}f_n, \dots)$, where $\mathbf{g}$ denotes the matrix $g_{\mu\nu}$, does not commute with Lorentz boosts. It is well known that in the case of free electromagnetic field, Lorentz boosts are represented by unbounded operators on a Fock space. See, e.g., J.-P. Antoine, "Indefinite metric and Poincaré covariance," in *Mathematical Aspects of Quantum Field Theory*, Acta Universitatis Wratislaviensis No. 519 (University Press, Wrocław, 1979).
[20]F. Strocchi, Phys. Rev. D **17**, 2010 (1978).
[21]See also Ref. 20 and T. Kugo and I. Ojima, Suppl. Progr. Theor. Phys. No. 66, 1 (1979).
[22]See Ref. 1, Theorem 4-2.
[23]See Ref. 1, Theorem 4-6.
[24]See Ref. 1, Theorem 4-9.
[25]G. G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley, New York, 1972), Chap. 3, Theorem 9.
[26]See Ref. 4.
[27]A theory of indefinite inner product spaces is developed in the monograph, J. Bognár, *Indefinite Inner Product Spaces* (Springer-Verlag, Berlin, 1974). See also A. Z. Jadczyk, Rep. Math. Phys. **2**, 263 (1971).

Spin-$\frac{1}{2}$ quantum field theory in nonstatic space-times. I. Construction of states

Amir-Homayoon Najmi
Department of Physics, University of Utah, Salt Lake City, Utah 84112

(Received 27 June 1983; accepted for publication 19 October 1983)

The problem of constructing states for a Hermitian spin-$\frac{1}{2}$ quantum field in a nonstatic background geometry is considered and the solution corresponding to the extremization of the total energy at an initial time is obtained.

PACS numbers: 11.10.Ef

INTRODUCTION

The problem of constructing states of quantum fields in a curved nonstatic background geometry has been the subject of much research in the past few years. Many different but equivalent formalisms have been developed to deal with this problem. These formalisms are, of course, incomplete unless a criterion for choosing a preferred vacuum state is stated. The most aesthetically pleasing of these criteria (though not necessarily with physically acceptable consequences) is the requirement that the expectation value of the total energy of the field when restricted to an initial hypersurface should be a minimum with respect to infinitesimal variations of the state, or alternatively, that the total energy at an initial time should be diagonal when written in terms of creation and annihilation operators.[1] This problem has been solved exactly for a scalar field in a general nonstatic background space-time.[2] However, it is observed that the states thus constructed suffer from the so-called Fulling anomaly, i.e., in a generic case, the states not only evolve in a nonunitary fashion, they also acquire an infinite amount of renormalized energy density to the future of the hypersurface on which the minimization took place. This pathological behavior was first pointed out by Parker (Ref. 3, see also Ref. 4).

In an attempt to investigate the generality of this result we are led to consider the problem of construction of the states of a Hermitian spin-$\frac{1}{2}$ field in a nonstatic background geometry using the criterion of energy minimization. In this paper we present the most general solution to the latter problem, but postpone an investigation of the evolution of the states thus constructed to a future publication. The main result is rather simple and shows a great deal in common with proofs of the positive mass theorems which have been of interest recently.[5]

I. PRELIMINARIES

We assume throughout that the manifold M in addition to being globally hyperbolic is also orientable, i.e., the first Steifel–Whitney class of M vanishes.[6] Then a spin structure can be defined on the whole of M. Equivalently, we may require that the second Steifel–Whitney class of M vanish.[6]

The Dirac matrices γ^μ satisfy

$$\{\gamma^\mu,\gamma^\nu\} = 2g^{\mu\nu}1, \tag{1}$$

where $g_{\mu\nu}$ is the metric on M, and has signature $(-,+,+,+)$. With this signature we may take the γ^μ to be real.[7] Then there exists a real antisymmetric matrix γ such that

$$\gamma^{\mu\sim} = -\gamma\gamma^\mu\gamma^{-1}. \tag{2}$$

The γ^μ are related to flat-space Dirac matrices γ^p by the Vierbein $L_{p\mu}$ (the summation convention is used throughout):

$$\gamma^\mu = L_p{}^\mu\gamma^p, \tag{3}$$

where the indices p and μ are raised and lowered by means of η_{pq} (the Minkowski metric) and $g_{\mu\nu}$, respectively. Under a Lorentz transformation $\mathscr{L}^s_r$ the matrices γ^p become

$$\gamma'^r = \mathscr{L}^r_p\gamma^p, \tag{4}$$

leaving the anticommutation relations (1) unchanged. It follows that there exists a real matrix S with the property that

$$\gamma'^r = S^{-1}\gamma^r S, \tag{5}$$

or equivalently

$$\gamma^r = \mathscr{L}^r_q S\gamma^q S^{-1}. \tag{6}$$

The spinor field ψ provides a spin representation of the Vierbein group according to the transformation law[8]

$$\psi' = S\psi. \tag{7}$$

The contragradient representation is obtained by taking

$$\bar\psi \equiv \psi^\sim\gamma, \tag{8}$$

which transforms as

$$\bar\psi' = \bar\psi S^{-1}. \tag{9}$$

The covariant derivative of a spinor field in the coordinate basis is

$$\psi_{\cdot\mu} \equiv \psi_{,\mu} + G^{\ \nu}_{\sigma}\Gamma^{\sigma}_{\mu\nu}\psi, \tag{10}$$

where the $\Gamma^\sigma_{\mu\nu}$ are the usual Christoffel symbols. In the local frame basis we can write the covariant derivative as

$$\psi_{\cdot\mu} = \psi_{,\mu} + \tfrac{1}{2}G_{[pq]}\Gamma^{pq}_{\mu}\psi, \tag{11}$$

where

$$G_{[pq]} \equiv \tfrac{1}{4}[\gamma_p,\gamma_q] \tag{12}$$

and Γ^{pq}_μ are determined by the requirement that covariant differentiation commute with the conversion of coordinate into frame indices[8]

$$L^{p\sigma}{}_{,\mu} + L^{p\nu}\Gamma^\sigma_{\mu\nu} + L^{\ \sigma}_{q}\Gamma^{pq}_{\mu} = 0. \tag{13}$$

 0022-2488/84/030623-06$02.50

Equation (13) has the following solution:

$$\Gamma^{pq}_\mu = \tfrac{1}{2}L^{p\nu}(L^{\ q}_{\nu,\mu} - L^{\ q}_{\mu,\nu}) - \tfrac{1}{2}L^{q\nu}(L^{\ p}_{\nu,\mu} - L^{\ p}_{\mu,\nu}) + \tfrac{1}{2}L^{q\sigma}L^{p\nu}(L^{m}_{\ \ \nu,\sigma} - L^{m}_{\ \ \sigma,\nu})L_{m\mu}. \tag{14}$$

Hence

$$\Gamma_\nu \equiv -\tfrac{1}{2}G_{[pq]}\Gamma^{pq}_\nu = \tfrac{1}{4}L_{p\mu\cdot\nu}L_q{}^\mu\gamma^p\gamma^q. \tag{15}$$

It is, of course, straightforward to verify that γ^μ and γ are covariantly constant:

$$\gamma^\mu_{\cdot\nu} \equiv \gamma^\mu_{,\nu} + \Gamma^\mu_{\ \nu\sigma}\gamma^\sigma - \tfrac{1}{8}[\gamma^\mu[\gamma_p,\gamma_q]\Gamma^{pq}_\nu] = 0, \tag{16i}$$

$$\gamma_{\cdot\mu} \equiv \gamma_{,\mu} - \tfrac{1}{8}\gamma\,[\gamma_p,\gamma_q]\Gamma^{pq}_\mu - \tfrac{1}{8}[\gamma_p,\gamma_q]^{\sim}\gamma\Gamma^{pq}_\mu = 0. \tag{16ii}$$

The action functional for the free, massive, Hermitian fermion field $\Psi(x)$ is[7]

$$S[\Psi] = \tfrac{1}{2}i\int_M \overline{\Psi}(\gamma^\mu\Psi_{\cdot\mu} + m\Psi)g^{1/2}\,d^4x. \tag{17}$$

The stationarity of the action with respect to arbitrary variations $\delta\psi(x)$ in the form of $\Psi(x)$ lead to the field equations

$$S[\Psi]\frac{\overleftarrow{\delta}}{\delta\Psi^{\sim}} = ig^{1/2}\gamma(\gamma^\mu\Psi_{\cdot\mu} + m\Psi) = 0, \tag{18i}$$

$$\frac{\vec{\delta}}{\delta\overline{\Psi}}S[\Psi] = ig^{1/2}(-\overline{\Psi}_{\cdot\mu}\gamma^\mu + m\overline{\Psi}) = 0, \tag{18ii}$$

where the covariant derivative of the adjoint spinor $\overline{\Psi}$ is

$$\overline{\Psi}_{\cdot\mu} \equiv \Psi^{\sim}_{,\mu}\gamma + \Psi^{\sim}\gamma\Gamma_\mu. \tag{19}$$

It is useful to define the intrinsic covariant derivative of a spinor restricted to a spacelike hypersurface $\mathscr{S}_t$ whose metric $h_{\mu\sigma}$ is defined by

$$h_{\mu\sigma} = g_{\mu\sigma} + n_\mu n_\sigma, \tag{20}$$

where n^μ is the unit future-pointing normal to the surface $\mathscr{S}_t$. We define this derivative by requiring the covariant derivative, in the hypersurface, of the projection of the γ^μ matrices into $\mathscr{S}_t$, to vanish. The calculation is performed in Appendix A and here we state the result. Denoting the intrinsic derivative by D_σ we have[9]

$$D_\sigma\Psi|_{\mathscr{S}_t} = h_\sigma{}^\mu\Psi_{\cdot\mu} + \tfrac{1}{2}\chi_{\sigma\mu}n_\nu\gamma^\mu\gamma^\nu\Psi, \tag{21}$$

where $\chi_{\sigma\mu}$ is the second fundamental form of $\mathscr{S}_t$.

The intrinsic spatial covariant Dirac operator on $\mathscr{S}_t$ is then given by

$$\not{D}\Psi|_{\mathscr{S}_t} \equiv \gamma^\mu h_\mu{}^\sigma D_\sigma\Psi = \gamma^\mu h_\mu{}^\sigma\Psi_{\cdot\sigma} + \tfrac{1}{2}\chi n_\nu\gamma^\nu\Psi, \tag{22}$$

where we have used the symmetry of $\chi_{\mu\sigma}$ and Eq. (1).

II. QUANTIZATION, THE MINIMIZATION OF THE TOTAL ENERGY AND HAMILTONIAN DIAGONALIZATION

The quantum field theory of $\Psi(x)$ is constructed by defining the operator-valued distribution $\hat{\Psi}(x)$ which satisfies the field equation and imposing the equal-time anticommutation relations

$$\{\hat{\Psi}(x'),\hat{\overline{\Psi}}(x)\}|_{x,x'\in\mathscr{S}_t} = -n_\nu\gamma^\nu\delta(\underset{\sim}{x}',\underset{\sim}{x})\hat{\mathbb{1}}. \tag{23}$$

The δ-function on the right-hand side is an invariant bispinor density on $\mathscr{S}_t$, for example it transforms like

$$h^{1/4}(\underset{\sim}{x}')\Psi(\underset{\sim}{x}')\overline{\Psi}(\underset{\sim}{x})h^{1/4}(\underset{\sim}{x}). \tag{24}$$

The field is expanded in terms of positive and negative frequency spinors $\{\psi_k(x)\}$ and $\{\psi^*_k(x)\}$,

$$\hat{\Psi}(x) = \sum_k \{\hat{a}_k\psi_k(x) + \hat{a}^*_k\psi^*_k(x)\}, \tag{25}$$

where k includes at least one discrete spinor index taking four distinct values.[10] The creation and annihilation operators $\hat{a}^*_k$ and $\hat{a}_k$ are now required to satisfy the anticommutation relations

$$\{\hat{a}_k, \hat{a}^*_{k_1}\} = \delta_{kk_1}\hat{\mathbb{1}}, \tag{26i}$$

$$\{\hat{a}_k,\hat{a}_{k_1}\} = \{\hat{a}^*_k,\hat{a}^*_{k_1}\} = \hat{0}, \tag{26ii}$$

which together with the anticommutation relations (23) lead to the completeness relations for the positive and negative frequency spinors

$$\sum_k \{\psi_k(\underset{\sim}{x}')\bar{\psi}^*_k(\underset{\sim}{x}) + \psi^*_k(\underset{\sim}{x}')\bar{\psi}_k(\underset{\sim}{x})\} = -n_\nu\gamma^\nu\delta(\underset{\sim}{x},\underset{\sim}{x}'). \tag{27}$$

The equivalent orthonormality relations are[7]

$$\int_{\mathscr{S}_t}\bar{\psi}^*_{k_1}\gamma^\mu\psi_{k_2}\,d\Sigma_\mu = \delta_{k_1k_2}, \tag{28i}$$

$$\int_{\mathscr{S}_t}\bar{\psi}_{k_1}\gamma^\mu\psi_{k_2}\,d\Sigma_\mu = 0, \tag{28ii}$$

which are independent of the hypersurface $\mathscr{S}_t$, thanks to the field equations.

The Dirac inner product, unlike the Klein–Gordon inner product, is positive both for positive and negative frequency solutions.[11]

The one-particle Hilbert space H^1 is taken to be $H^+ \oplus \overline{H}^-$, where H^+ and H^- are the Hilbert spaces of positive frequency and negative frequency solutions of the Dirac equation and $\overline{H}^-$ is the dual space to H^-. However, the Hilbert space of all states is now taken to be the antisymmetric Fock space constructed from H^1,[12]

$$\mathscr{F}_A = C \oplus H^1 \oplus \cdots \oplus \bigotimes_{j=1}^{n} H^j \cdots. \tag{29}$$

Two sets of positive frequency solutions are related to each other by the Bogoliubov transformations[12,13]

$$\eta_k = \sum_{k_1}(\alpha_{k_1k}\psi_{k_1} + \beta_{k_1k}\psi^*_{k_1}), \tag{30}$$

where the Bogoliubov relations differ from those of the scalar field by an important sign difference[13]:

$$\sum_k (\alpha^*_{kk_1}\alpha_{kk_2} + \beta^*_{kk_1}\beta_{kk_2}) = \delta_{k_1k_2}, \tag{31i}$$

$$\sum_k (\alpha_{kk_1}\beta_{kk_2} + \beta_{kk_1}\alpha_{kk_2}) = 0. \tag{31ii}$$

The complex structure J[11,14] is defined to be equivalent to multiplying by $+i$ when acting on the positive frequency solutions and by $-i$ when acting on the negative frequency solutions. However, in the context of the procedure of energy minimization, the complex structure J need not be employed as a characterization of the states of the quantum field.

The stress-energy tensor of the Hermitian spin-$\frac{1}{2}$ field is[7]

$$\widehat{T}_{\mu\nu} \equiv g^{-1/2} L_{\rho\nu} \frac{\delta}{\delta L_p{}^{\mu}} \widehat{S} = \tfrac{1}{8} i (\widehat{\overline{\Psi}}_{,\mu} \gamma_\nu \widehat{\Psi} + \widehat{\overline{\Psi}}_{,\nu} \gamma_\mu \widehat{\Psi} - \widehat{\overline{\Psi}} \gamma_\mu \widehat{\Psi}_{,\nu} - \widehat{\overline{\Psi}} \gamma_\nu \widehat{\Psi}_{,\mu}). \tag{32}$$

We consider the energy-density operator defined by

$$\hat{\rho}(x) \equiv n^\mu n^\nu \widehat{T}_{\mu\nu} = \tfrac{1}{4} i (\widehat{\overline{\Psi}}_{,\mu} n^\mu n^\nu \gamma_\nu \widehat{\Psi} - \widehat{\overline{\Psi}} n_\nu \gamma^\nu n^\mu \widehat{\Psi}_{,\mu}). \tag{33}$$

The total energy operator at the instant t is

$$\widehat{E}(t) = \int_{\mathscr{S}_t} \hat{\rho}(x) d\underset{\sim}{x}, \tag{34}$$

where $d\underset{\sim}{x}$ is the invariant volume element on $\mathscr{S}_t$, i.e., $h^{1/2}(\underset{\sim}{x}) d^3\underset{\sim}{x}$.

We can eliminate all the normal derivative terms with the aid of the field equations (18) and arrive at

$$\widehat{E}(t) = \tfrac{1}{4} i \int_{\mathscr{S}_t} (\widehat{\overline{\Psi}}_{,\sigma} h_\mu{}^\sigma \gamma^\mu \widehat{\Psi} - \widehat{\overline{\Psi}} \gamma^\mu h_\mu{}^\sigma \widehat{\Psi}_{,\sigma} - 2m \widehat{\overline{\Psi}} \widehat{\Psi}) d\underset{\sim}{x}. \tag{35}$$

In order to simplify (35) we prove an integration-by-parts formula. The latter will hold provided we assume the usual conditions (such as square integrability) on the functions $\psi_k(x)$, or equivalently we may consider smeared field operators. Since we shall want to use this formula in connection with a variational problem in which the variations of the functions $\psi_k(x)$, i.e., $\delta\psi_k(x)$, vanish on the boundary of $\mathscr{S}_t$, we shall not need the surface terms arising from total divergences and shall set them equal to zero in what follows.

The expression $\bar{\psi}_1 h^{\tau\sigma} \gamma_\sigma \psi_2$ when restricted to the hypersurface $\mathscr{S}_t$ is a vector whose intrinsic covariant divergence is

$$(\bar{\psi}_1 h^{\tau\sigma} \gamma_\sigma \psi_2)_{|\tau} \equiv (\bar{\psi}_1 h^{\nu\sigma} \gamma_\sigma \psi_2)_{,\mu} h_\tau{}^\mu h_\nu{}^\tau = \bar{\psi}_{1,\tau} h^{\tau\sigma} \gamma_\sigma \psi_2 + \bar{\psi}_1 h^{\tau\sigma} \gamma_\sigma \psi_{2,\tau} + \chi \bar{\psi}_1 n_\nu \gamma^\nu \psi_2, \tag{36}$$

where χ is the trace of the second fundamental form of $\mathscr{S}_t$.[15]

Integrating both sides of Eq. (36) over the hypersurface $\mathscr{S}_t$ and ignoring the boundary term arising from the left-hand side we obtain

$$\int_{\mathscr{S}_t} (\bar{\psi}_{1,\tau} h^{\tau\sigma} \gamma_\sigma \psi_2 + \bar{\psi}_1 h^{\tau\sigma} \gamma_\sigma \psi_{1,\tau} + \chi \bar{\psi}_1 n_\sigma \gamma^\sigma \psi_2) d\underset{\sim}{x} = 0. \tag{37}$$

Using the expression for the intrinsic covariant Dirac operator given in Eq. (22) and the above formula, the expression for the total energy operator can be cast in the following form:

$$\widehat{E}(t) = -\tfrac{1}{2} i \int_{\mathscr{S}_t} \widehat{\overline{\Psi}} (\mathbb{D} + m) \widehat{\Psi} \, d\underset{\sim}{x}. \tag{38}$$

We substitute the field expansion (28) into the expression above to obtain

$$\begin{aligned} \widehat{E}(t) = -\tfrac{1}{2} i \int_{\mathscr{S}_t} d\underset{\sim}{x} \sum_{k_1,k_2} \{ & (\bar{\psi}_{k_1} (\mathbb{D} + m) \psi_{k_2}) \hat{a}_{k_1} \hat{a}_{k_2} \\ & + (\bar{\psi}^*_{k_1} (\mathbb{D} + m) \psi_{k_2}) \hat{a}^*_{k_1} \hat{a}_{k_2} + (\bar{\psi}_{k_1} (\mathbb{D} + m) \psi^*_{k_2}) \hat{a}_{k_1} \hat{a}^*_{k_2} \\ & + (\bar{\psi}^*_{k_1} (\mathbb{D} + m) \psi^*_{k_2}) \hat{a}^*_{k_1} \hat{a}^*_{k_2} \}. \end{aligned} \tag{39}$$

Let us also give the expectation value of (39) in the $\hat{a}$- vacuum state (the total energy), viz.,

$$\langle \hat{a};\text{vac} | \widehat{E}(t) | \hat{a};\text{vac} \rangle = -\tfrac{1}{2} i \int_{\mathscr{S}_t} d\underset{\sim}{x} \sum_k \bar{\psi}_k (\mathbb{D} + m) \psi^*_k. \tag{40}$$

We are now ready to solve the problem of energy minimization. The latter requires the choice of initial values for the positive frequency spinors so as to minimize Expression (40), subject to the constraints implied by (28). In order to solve this variational problem we introduce the Lagrange multipliers λ_{kk_1} and μ_{kk_1} and minimize the expression

$$\begin{aligned} \mathrm{Tr} [& -\tfrac{1}{2} i \bar{\psi}^*_k (\mathbb{D} + m) \psi_k + \lambda_{kk_1} (\bar{\psi}^*_k n_\nu \gamma^\nu \psi_{k_1} - \delta_{kk_1}) \\ & + \mu_{kk_1} (\bar{\psi}_k n_\nu \gamma^\nu \psi_{k_1})] \end{aligned} \tag{41}$$

with respect to infinitesimal variations of the ψ_k. Tr denotes an integral over $\mathscr{S}_t$ and a sum over the k indices, as well as a trace over the spin indices.

This problem is solved by taking the initial values of the positive frequency spinors, i.e., $\psi_k(\underset{\sim}{x})$, to be the eigenvectors of the operator $\vec{0}$ defined as follows:

$$\vec{0} \equiv i n_\nu \gamma^\nu (\mathbb{D} + m). \tag{42}$$

We shall now verify that the eigenvectors of the spinorial operator $\vec{0}$ (defined entirely in terms of quantities in the hypersurface $\mathscr{S}_t$) satisfy the constraint (28). Before we do so we point out that these same eigenvectors diagonalize the total energy operator of Eq. (39) in the sense that when $\psi_k(\underset{\sim}{x})$ in that equation are chosen as eigenvectors of $\vec{0}$, the coefficients of the operators $\hat{a}_k \hat{a}_{k_1}$ and $\hat{a}^*_k \hat{a}^*_{k_1}$ vanish.

First we show that the spectrum of $\vec{0}$ is contained in $\mathbb{R} \setminus \{0\}$. Consider the square of $\vec{0}$

$$\vec{0}^2 = -n_\nu \gamma^\nu (\mathbb{D} + m) n_\sigma \gamma^\sigma (\mathbb{D} + m). \tag{43}$$

For an arbitrary spinor ψ the quantity $n_\nu \gamma^\nu \psi$ transforms like a spinor. Therefore,

$$\begin{aligned} \mathbb{D}(n_\nu \gamma^\nu \psi) & \equiv \gamma^\mu h_\mu{}^\sigma (n_\nu \gamma^\nu \psi)_{,\sigma} + \tfrac{1}{2} \chi n_\nu \gamma^\nu n_\sigma \gamma^\sigma \psi \\ & = \gamma^\mu h_\mu{}^\sigma (n_{\nu,\sigma} \gamma^\nu \psi + n_\nu \gamma^\nu \psi_{,\sigma}) - \tfrac{1}{2} \chi \psi \\ & = \gamma^\mu \chi_{\mu\sigma} \gamma^\sigma \psi + \gamma^\mu h_\mu{}^\sigma n_\nu \gamma^\nu \psi_{,\sigma} - \tfrac{1}{2} \chi \psi \\ & = h_\mu{}^\sigma n_\nu \gamma^\mu \gamma^\nu \psi_{,\sigma} + \tfrac{1}{2} \chi \psi. \end{aligned} \tag{44}$$

It follows that

$$\begin{aligned} n_\nu \gamma^\nu \mathbb{D}(n_\sigma \gamma^\sigma \psi) & = \tfrac{1}{2} \chi n_\sigma \gamma^\sigma \psi + h_\mu{}^\nu n_\delta n_\sigma \gamma^\delta \gamma^\mu \gamma^\sigma \psi_{,\nu} \\ & = \tfrac{1}{2} \chi n_\sigma \gamma^\sigma \psi + h_\mu{}^\nu n_\delta n_\sigma \gamma^\delta (2g^{\mu\sigma} - \gamma^\sigma \gamma^\mu) \psi_{,\nu} \\ & = \tfrac{1}{2} \chi n_\sigma \gamma^\sigma \psi + 0 - \tfrac{1}{2} h_\mu{}^\nu n_\delta n_\sigma \{\gamma^\delta, \gamma^\sigma\} \gamma^\mu \psi_{,\nu} \\ & = h_\mu{}^\nu \gamma^\mu \psi_{,\nu} + \tfrac{1}{2} \chi n_\nu \gamma^\nu \psi \equiv \mathbb{D} \psi. \end{aligned} \tag{45}$$

Thus Eq. (43) becomes

$$\vec{0}^2 = -(\mathbb{D} - m)(\mathbb{D} + m) = -\mathbb{D}^2 + m^2. \tag{46}$$

It is known that the square of the intrinsic hypersurface Dirac operator $\mathbb{D}$ is related to the Laplace–Beltrami operator of the hypersurface $\mathscr{S}_t$ by a "Weitzenbock formula"[9]

$$\mathbb{D}^2 = {}^{(3)}\Delta + \tfrac{1}{4} {}^{(3)}R, \tag{47}$$

where ${}^{(3)}\Delta$ is the Laplace–Beltrami operator of the hypersurface $\mathscr{S}_t$ and ${}^{(3)}R$ is its Ricci scalar.

Here we derive an integration-by-parts formula for the operator $\not{D}$. It follows from (22) that

$$n_\nu \gamma^\nu \not{D}\psi = n_\nu \gamma^\nu \gamma^\sigma h_\sigma{}^\tau \psi_{,\tau} - \tfrac{1}{2}\chi\psi. \tag{48}$$

Taking the spinor adjoint of both sides we obtain

$$-\bar{\psi}\overleftarrow{\not{D}} n_\nu \gamma^\nu = \bar{\psi}_{,\tau} h_\sigma{}^\tau \gamma^\sigma n_\nu \gamma^\nu - \tfrac{1}{2}\chi\bar{\psi}, \tag{49}$$

where we have used (2) and defined the adjoint of $\not{D}$,

$$\overleftarrow{\not{D}} \equiv \gamma^{-1} \not{D}^{\sim} \gamma. \tag{50}$$

Multiplying the two sides of (49) by some spinor ϕ and then integrating the trace of both sides over the hypersurface $\mathscr{S}_t$ we obtain

$$-\int_{\mathscr{S}_t} \bar{\psi}\overleftarrow{\not{D}}\gamma^\mu \phi \, d\Sigma_\mu = \int_{\mathscr{S}_t} \bar{\psi}_{,\sigma} h_\nu{}^\sigma \gamma^\nu \gamma^\mu \phi \, d\Sigma_\mu - \tfrac{1}{2}\int_{\mathscr{S}_t} \chi \bar{\psi}\phi \, d\underline{x}. \tag{51}$$

Let us write down the covariant divergence in the hypersurface of the quantity $\bar{\psi} h_\tau{}^\sigma \gamma^\tau n_\nu \gamma^\nu \phi_{|\mathscr{S}_t}$ which is a vector in $\mathscr{S}_t$

$$\begin{aligned}
(\bar{\psi} h_\nu{}^\sigma \gamma^\nu n_\mu \gamma^\mu \phi)_{|\sigma}
&= (\bar{\psi} h_\tau{}^m \gamma^\tau n_\mu \gamma^\mu \phi)_{;\nu} h_\sigma{}^\nu h_m{}^\sigma \\
&= \bar{\psi}_{;\sigma} h_\tau{}^\sigma \gamma^\tau n_\nu \gamma^\nu \phi + \bar{\psi} h_\tau{}^\sigma \gamma^\tau n_\nu \gamma^\nu \phi_{;\sigma} \\
&\quad + \bar{\psi}(n_{\mu;\nu} h_\tau{}^\sigma + n_\mu n^\sigma{}_{;\nu} n_\tau + n_\mu n^\sigma n_{\tau;\nu}) h_\sigma{}^\nu \gamma^\tau \gamma^\mu \phi \\
&= \bar{\psi}_{;\sigma} h_\tau{}^\sigma \gamma^\tau n_\nu \gamma^\nu \phi + \bar{\psi} h_\tau{}^\sigma \gamma^\tau n_\nu \gamma^\nu \phi_{;\sigma} + 0,
\end{aligned} \tag{52}$$

where the symmetry of $\chi_{\mu\sigma}$ and Eq. (1) have been used. Now we integrate both sides over $\mathscr{S}_t$ and neglect the boundary term arising from the total divergence on the left-hand side to obtain

$$0 = \int_{\mathscr{S}_t} (\bar{\psi}_{;\sigma} h_\tau{}^\sigma \gamma^\tau n_\nu \gamma^\nu \phi + \bar{\psi} h_\tau{}^\sigma \gamma^\tau n_\nu \gamma^\nu \phi_{;\sigma}) d\underline{x}, \tag{53}$$

or equivalently

$$\begin{aligned}
\int_{\mathscr{S}_t} \bar{\psi}_{;\sigma} h_\tau{}^\sigma \gamma^\tau \gamma^\mu \phi \, d\Sigma_\mu
&= -\int_{\mathscr{S}_t} \bar{\psi} h_\tau{}^\sigma \gamma^\tau \gamma^\mu \phi_{;\sigma} \, d\Sigma_\mu \\
&= +\int_{\mathscr{S}_t} \bar{\psi} \gamma^\mu h_\tau{}^\sigma \gamma^\tau \phi_{;\sigma} \, d\Sigma_\mu.
\end{aligned} \tag{54}$$

In the last line we have used

$$h_\sigma{}^\tau n_\mu \gamma^\sigma \gamma^\mu = h_\sigma{}^\tau n_\mu (2g^{\mu\sigma} - \gamma^\mu \gamma^\sigma) = -h_\sigma{}^\tau n_\mu \gamma^\mu \gamma^\sigma. \tag{55}$$

We use Eq. (54) to rewrite the integral on the right-hand side of Eq. (51) in the following form:

$$-\int_{\mathscr{S}_t} \bar{\psi}\overleftarrow{\not{D}}\gamma^\mu \phi \, d\Sigma_\mu = \int_{\mathscr{S}_t} \bar{\psi}\gamma^\mu (h_\sigma{}^\tau \gamma^\sigma \phi_{;\tau} + \tfrac{1}{2}\chi n_\nu \gamma^\nu \phi) d\Sigma_\mu. \tag{56}$$

The expression in brackets on the right-hand side of the above is precisely $\not{D}\phi$. Finally,

$$-\int_{\mathscr{S}_t} \bar{\psi}\overleftarrow{\not{D}}\gamma^\mu \phi \, d\Sigma_\mu = \int_{\mathscr{S}_t} \bar{\psi}\gamma^\mu \overleftarrow{\not{D}}\phi \, d\Sigma_\mu, \tag{57}$$

which is the formula we were seeking. This equation together with the positivity of the Dirac inner product is sufficient to show that the square of $\overleftarrow{\not{D}}$ is positive definite. It follows that the square of the operator $\vec{0}$ is positive and hence the spectrum of $\vec{0}$ is contained in $\mathbb{R}\setminus\{0\}$.

We note that if $\psi_k(\underline{x})$ is an eigenvector of $\acute{0}$ with the (real) eigenvalue λ_k, then $\psi_k^*(x)$ is an eigenvector with the eigenvalue $-\lambda_k$.

Finally we are in a position to verify that the eigenvectors of $\vec{0}$ satisfy (28ii). To this end let

$$i n_\nu \gamma^\nu (\not{D} + m)\psi_k(\underline{x}) = \lambda_k \psi_k(\underline{x}). \tag{58}$$

Taking the spinor adjoint of both sides we obtain

$$-i\bar{\psi}_k(\underline{x})(\overleftarrow{\not{D}} + m) n_\nu \gamma^\nu = \lambda_k \bar{\psi}_k(\underline{x}), \tag{59}$$

or equivalently

$$i\bar{\psi}_k(\underline{x})(\overleftarrow{\not{D}} + m) = \lambda_k \bar{\psi}_k(\underline{x}) n_\nu \gamma^\nu. \tag{60}$$

If we now take the trace of the product of Eqs. (58) and (60) (for different k's) and integrate over the surface $\mathscr{S}_t$, we obtain

$$\begin{aligned}
\lambda_{k_1}\lambda_{k_2} \int_{\mathscr{S}_t} \bar{\psi}_{k_1} \gamma^\mu \psi_{k_2} \, d\Sigma_\mu
&= -\int_{\mathscr{S}_t} \bar{\psi}_{k_1} (\overleftarrow{\not{D}} + m)\gamma^\mu (\overrightarrow{\not{D}} + m)\psi_{k_2} \, d\Sigma_\mu \\
&= -\int_{\mathscr{S}_t} \bar{\psi}_{k_1} \gamma^\mu (-\overrightarrow{\not{D}}^2 + m^2)\psi_{k_2} \, d\Sigma_\mu \\
&= -\lambda_{k_2}^2 \int_{\mathscr{S}_t} \bar{\psi}_{k_1} \gamma^\mu \psi_{k_2} \, d\Sigma_\mu.
\end{aligned} \tag{61}$$

Equation (28ii) follows from the above and the fact that λ_{k_1} and λ_{k_2} are real and have the same sign.

To determine the initial vacuum state completely we need to show which sign of the spectrum of $\vec{0}$ we should assign to positive frequency. We note that in Minkowski space-time positive frequency spinors are taken to be[10]

$$\psi_k(t,\underline{x}) = (2\pi)^{-3/2}(\underline{k}^2 + m^2)^{-1/2} \exp(-ikt + i\underline{k}\cdot\underline{x})\psi_{\underline{k}}, \tag{62}$$

where $\psi_{\underline{k}}$ satisfy

$$(-ik\gamma^o + i\underline{k}\cdot\underline{\gamma} + m)\psi_{\underline{k}} = 0. \tag{63}$$

The latter equation can be put in the form of Eq. (58)

$$-i\gamma^o(i\underline{k}\cdot\underline{\gamma} + m)\psi_{\underline{k}} = -k\psi_{\underline{k}}, \tag{64}$$

from which it follows that the positive frequency spinor solutions of the Dirac equation, minimizing the total energy at an instant $\mathscr{S}_t$, have the negative-eigenvalue eigenvectors of the spatial operation $\vec{0}$ as their initial data.

III. DISCUSSION

To compare the result of Sec. II to the case of a scalar field we may obtain an expression for the expectation value of the energy-density operator in the state given by Eq. (58), when the metric is a Bianchi Type I whose line element squared is

$$ds^2 = -dt^2 + \sum_{j=1}^{3} a_j^2(t)(dx^j)^2. \tag{65}$$

The vierbein may be chosen as

$$L_{p\mu} = \mathrm{diag}(-1, a_1(t), a_2(t), a_3(t)), \tag{66}$$

and using the expressions in Appendix B we find

$$\vec{0} = -i\gamma^o\left[\sum_{j=1}^{3} a_j^{-1}\gamma^j\left(\frac{\partial}{\partial x^j} - \Gamma_{x^j}\right) - \frac{1}{2}\sum_{j=1}^{3}\left(\frac{\dot{a}_j}{a_j}\right)\gamma^o + m\right]. \tag{67}$$

The positive frequency solutions $\psi_{rk}(x)$ can be taken as

$$\psi_{rk}(t,\underline{x}) = \psi_{rk}(t)e^{ik\cdot x}. \tag{68}$$

Eigenvalue Eq. (58) then becomes (at $t = t_o$)

$$-i\gamma^o\left[i\sum_{j=1}^{3}\frac{k_j}{a_j}\gamma^j + m\right]\psi_{rk}(t_0) = \lambda_k\psi_{rk}(t_0), \tag{69}$$

where the eigenvalues λ_k can be obtained by squaring $\vec{0}$

$$\lambda_k = \pm\left[\sum_{j=1}^{3}\left(\frac{k_j}{a_j}\right)^2 + m^2\right]^{1/2}. \tag{70}$$

Hence our positive frequency solutions satisfy the following equation at $t = t_0$:

$$-i\gamma^o\left[i\sum_{j=1}^{3}\frac{k_j}{a_j}\gamma^j + m\right]\psi_{rk}(t_0) = -\left[\sum_{j=1}^{3}\left(\frac{k_j}{a_j}\right)^2 + m^2\right]^{1/2}\psi_{rk}(t_0). \tag{71}$$

Using the above and Eq. (40) we obtain the following expression for the expectation value of the energy density at t_0:

$$\rho(t_0) = -\frac{1}{2}\sum_{r,k}\bar{\psi}_{rk}(t_0)\gamma^o\psi^*_{rk}(t_0)\left[\sum_{j=1}^{3}\left(\frac{k_j}{a_j}\right)^2 + m^2\right]. \tag{72}$$

Now the orthonormality relations reduce to

$$(a_1a_2a_3)^3\int_{t=t_0}\psi^*_{rk_1}(-\gamma^o)\psi_{sk_2}\,e^{-ix\cdot(k_1-k_2)}\,d^3\underline{x} = \delta_{rs}\delta(\underline{k}_1-\underline{k}_2), \tag{73}$$

or equivalently

$$\bar{\psi}^*_{rk}(-\gamma^o)\psi_{sk} = 4(2\pi)^{-3}(a_1a_2a_3)^{-1}\delta_{rs}. \tag{74}$$

Using the above equation we can perform the sum over the spinor index r in Eq. (72) and obtain the following formally divergent expression:

$$\rho(t_0) = 2(2\pi)^{-3}(a_1a_2a_3)^{-3}\int d^3\underline{k}\left[\sum_{j=1}^{3}\left(\frac{k_j}{a_j}\right)^2 + m^2\right]^{1/2}, \tag{75}$$

which is precisely twice the result for the conformally coupled scalar field.[4] We may employ usual methods of regularization and renormalization in Eq. (75) but we shall not pursue the matter further here.

The next and perhaps the most important task is to calculate the effects of the evolution of these states, such as the renormalized expectation value of the energy-density operator at some later time t_1. It is of great interest to see if the pathologies exhibited by the scalar field persist in this case also. Work on this subject is under progress and will be reported at a later date.[16]

ACKNOWLEDGMENTS

I would like to thank Professor Michael Atiyah, Dr. Paul Todd, and Dr. Nicholas Woodhouse of the Mathematical Institute of Oxford University, Dr. Martin Brown and Dr. Adrian Ottewill of the Astrophysics Department of Oxford University, and Professor Karel Kuchař of the University of Utah for helpful discussions. This work has been supported in part by the National Science Foundation under Contract No. Phy 81-06909.

APPENDIX A

In this Appendix we show that the intrinsic covariant derivative of a spinor restricted to a hypersurface $\mathscr{S}_t$ [Eq. (21)] can be obtained by requiring that the projection of the γ^μ matrices into the hypersurface vanish. Let us assume that

$$D_\alpha\psi_{|\mathscr{S}_t} = h_\alpha{}^\mu\psi_{;\mu} + h_\alpha{}^\mu K_\mu\psi, \tag{A1}$$

where K_μ is to be determined from the requirement below:

$$\vec{D}_\alpha(h_\mu{}^\sigma\gamma^\mu\psi)_{|\mathscr{S}_t} = (\vec{D}_\alpha h_\mu{}^\sigma\gamma^\mu)\psi + h_\mu{}^\sigma\gamma^\mu\vec{D}_\alpha\psi = h_\mu{}^\sigma\gamma^\mu\vec{D}_\alpha\psi. \tag{A2}$$

Since $h_\mu{}^\sigma\gamma^\mu\psi$ is a vector-spinor we have

$$\vec{D}_\alpha(h_\mu{}^\sigma\gamma^\mu\psi) = h_\nu{}^\sigma h_\alpha{}^\tau(h_\mu{}^\nu\gamma^\mu\psi)_{;\tau} + h_\alpha{}^\tau K_\tau\gamma^\mu h_\mu{}^\sigma\psi. \tag{A3}$$

The left-hand side of (A2) is equal to

$$\begin{aligned} &n_\nu{}^\sigma h_\alpha{}^\tau(\gamma^\mu h_\mu{}^\nu\psi_{;\tau} + \gamma^\mu(n_{\mu;\tau}n^\nu + n_\mu n^\nu{}_{;\tau})\psi) + h_\alpha{}^\tau K_\tau\gamma^\mu h_\mu{}^\sigma\psi \\ &= \gamma^\mu h_\mu{}^\sigma h_\alpha{}^\tau\psi_{;\tau} + \gamma^\mu n_\mu h_\nu{}^\sigma h_\alpha{}^\tau n^\nu{}_{;\tau}\psi + h_\alpha{}^\tau K_\tau\gamma^\mu h_\mu{}^\sigma \\ &= \gamma^\mu h_\mu{}^\sigma h_\alpha{}^\tau\psi_{;\tau} + n_\mu\gamma^\mu\chi_\alpha{}^\sigma\psi + h_\mu{}^\sigma h_\alpha{}^\tau K_\tau\gamma^\mu\psi. \end{aligned} \tag{A4}$$

Equation (A2) then becomes

$$\begin{aligned} &\gamma^\mu h_\mu{}^\sigma h_\alpha{}^\tau\psi_{;\tau} + n_\mu\gamma^\mu\chi_\alpha{}^\sigma\psi + h_\mu{}^\sigma h_\alpha{}^\tau K_\tau\gamma^\mu\psi \\ &= h_\mu{}^\sigma\gamma^\mu\vec{D}_\alpha\psi = h_\mu{}^\sigma\gamma^\mu h_\alpha{}^\tau\psi_{;\tau} + h_\mu{}^\sigma\gamma^\mu h_\alpha{}^\tau K_\tau\psi. \end{aligned} \tag{A5}$$

Hence,

$$h_\mu{}^\sigma h_\alpha{}^\tau\gamma^\mu K_\tau = h_\mu{}^\sigma h_\alpha{}^\tau K_\tau\gamma^\mu + n_\mu\gamma^\mu\chi_\alpha{}^\sigma, \tag{A6}$$

or equivalently

$$h_\mu{}^\sigma h_\alpha{}^\tau[\gamma^\mu,K_\tau] = n_\mu\gamma^\mu\chi_\alpha{}^\sigma. \tag{A7}$$

The solution to this equation is

$$h_\alpha{}^\tau K_\tau = \tfrac{1}{2}\chi_{\alpha\tau}\gamma^\tau n_\nu\gamma^\nu. \tag{A8}$$

To verify that the above is indeed the solution of Eq. (A7) we substitute (A8) into the left-hand side of (A7) and obtain

$$\begin{aligned} &h_\mu{}^\sigma(\tfrac{1}{2}\gamma^\mu\chi_{\alpha\tau}\gamma^\tau n_\nu\gamma^\nu - \tfrac{1}{2}\chi_{\alpha\tau}\gamma^\tau n_\nu\gamma^\nu\gamma^\mu) \\ &= \tfrac{1}{2}h_\mu{}^\sigma\chi_{\alpha\tau}n_\nu(\gamma^\mu\gamma^\tau\gamma^\nu - \gamma^\tau\gamma^\nu\gamma^\mu) \\ &= \tfrac{1}{2}h_\mu{}^\sigma\chi_{\alpha\tau}n_\nu(\gamma^\mu\gamma^\tau\gamma^\nu + \gamma^\tau\gamma^\mu\gamma^\nu - 2g^{\mu\nu}\gamma^\tau) \\ &= \tfrac{1}{2}h_\mu{}^\sigma\chi_{\alpha\tau}n_\nu\{\gamma^\mu,\gamma^\tau\}\gamma^\nu - 0 \\ &= \tfrac{1}{2}h_\mu{}^\alpha\chi_{\alpha\tau}n_\nu 2g^{\mu\tau}\gamma^\nu = \chi_\alpha{}^\sigma n_\nu\gamma^\nu, \end{aligned} \tag{A9}$$

which is precisely the right-hand side of (A7).

APPENDIX B

We use the conventions of Hawking and Ellis.[15] For the metric of Eq. (65) we find the following Christoffel symbols:

$$\Gamma^o_{ii} = \dot{a}_i a_i, \qquad \Gamma^i_{io} = \dot{a}_i/a_i \quad (i = 1,2,3). \tag{B1}$$

The trace of the second fundamental form of t = constant hypersurfaces is

$$\chi = \sum_{i=1}^{3}\frac{\dot{a}_i}{a_i}. \tag{B2}$$

Using Eq. (15) we find the following spinor connection coefficients:

$$\Gamma_t = 0,$$
$$\Gamma_{x^i} = \tfrac{1}{2}\,\dot{a}_i \gamma^0 \gamma_i \quad (i = 1,2,3). \tag{B3}$$

[1] S. A. Fulling, Ph.D. thesis, Princeton University, 1972.
[2] D. Deutsch and A. H. Najmi, Phys. Rev. D **28**, 1907 (1983).
[3] L. Parker, Phys. Rev. **183**, 1065 (1969).
[4] S. A. Fulling, Gen. Relativ. Gravit. **10**, 807 (1979).
[5] E. Witten, Commun. Math. Phys. **80**, 301 (1981).
[6] R. Penrose, in *Battelle Recontres 1967*, edited by C. M. DeWitt and J. A. Wheeler (Benjamin, New York, 1968).
[7] B. S. DeWitt and C. J. Isham, in *Quantum Mechanics in Strong Gravitational Fields* (University of Texas, Austin, TX, 1978), Chap. I.
[8] B. S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1967).
[9] T. Parker and C. H. Taubes, Commun. Math. Phys. **84**, 223 (1982).
[10] D. Lurie, *Particles and Fields* (Interscience, New York, 1968).
[11] N. Woodhouse, Proc. R. Soc. London. Ser. A **378**, 119 (1981).
[12] R. M. Wald, Ann. Phys. **118**, 490 (1979).
[13] L. Parker, Phys. Rev. D **3**, 346 (1971).
[14] A. Ashtekar and A. Magnon, Proc. R. Soc. London Ser. A **346**, 375 (1975).
[15] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge University, Cambridge, 1973).
[16] A. H. Najmi and A. C. Ottewill, paper in preparation (1984).

On zeta function regularization for operators with continuous spectra

G. M. Gallatin
Department of Physics, Fairfield University, Fairfield, Connecticut 06430

(Received 13 September 1983; accepted for publication 19 October 1983)

We present the generalization of the method of zeta function regularization for operators with continuous spectra where one cannot interchange the order of integration and limit processes. A sample calculation is presented.

PACS numbers: 11.10.Gh

1. INTRODUCTION

In recent years zeta function regularization has become a popular method for finding the finite part of the one-loop correction to the classical action. Unfortunately the standard references,[1] with the exception of DeWitt[2] describe the method for a discrete but not for a continuous spectrum. DeWitt discusses the case of a continuous spectrum where it is valid to interchange orders of integration and limit processes in defining the determinant, and for such operators the definition of the generalized zeta function is the same for the discrete as well as the continuous spectrum. In general this is not so. We will present the generalization of DeWitt's work which is valid when one cannot interchange orders of integration and limit processes.

One way of avoiding this problem is to place the system in a "box" by putting boundaries on the manifold to obtain a discrete spectrum, and the boundaries can be sent to infinity at the end of the calculation. In flat space this is a trivial procedure but in curved space the introduction of boundaries can lead to problems.[3] Thus it should be useful to develop the generalization of the zeta function method for continuous spectra directly.

The paper is organized as follows. In Sec. 2 the method for discrete spectra is reviewed and the method for continuous spectra is presented and discussed. In Sec. 3 a sample calculation is done which illustrates the care that must be taken to avoid getting nonsense.

2. ZETA FUNCTION REGULARIZATION FOR CONTINUOUS SPECTRA

We begin by reviewing, briefly, the method for discrete spectra (see Ref. 1 for a more detailed treatment).

In general what one wants to calculate is the determinant of a given differential operator. In a typical loop expansion this operator is given by the second variation of the action evaluated at the classical solution.[4] Here we will consider any operator which has a complete orthonormal set of eigenfunctions with eigenvalues which may be discrete or continuous. We work in N-dimensional flat space. The generalization to curved space should be straightforward.

If the operator $A(\partial,x)$ has a discrete spectrum then its determinant is given by the product of all the eigenvalues, E_n,

$$\det A = \det[A(\partial,x)\delta^N(x,x')] = \prod_n E_n = \exp\sum_n \ln E_n, \tag{2.1}$$

where n in general ranges over some countably infinite set. As might be expected, this yields infinity for the value of the determinant and the result must be regularized to obtain a finite answer. The zeta function method accomplishes this by analytic continuation. Consider the generalized zeta function defined by

$$\zeta(z) = \sum_n E_n^{-z}. \tag{2.2}$$

When the eigenvalues E_n increase without bound it can be shown that the sum will converge for $\mathrm{Re}(z) > 2$, and it can be analytically continued to a meromorphic function of z which is regular at $z = 0$ and has poles at $z = 1$ and 2.[1] Noticing that

$$\zeta'(0) \equiv \partial_z \zeta(z)|_{z=0} = -\sum_n \ln E_n, \tag{2.3}$$

one *defines* the determinant to be given by

$$\det A = \exp(-\zeta'(0)). \tag{2.4}$$

To generalize this procedure to the case where the eigenvalues E form a continuous spectrum, one must replace the sum in (2.2) by an integral. But this requires the multiplicity or degeneracy of the eigenvalues, i.e., the measure. In the discrete case discussed above, it is tacitly assumed that n labels the eigenvalues in a one-to-one manner; that is, for each n there is only one eigenvalue. Thus, if there is an m-fold multiplicity or degeneracy of the eigenvalues, there are m values of n each of which has the same value of E_n. If instead we wished to label only distinct values of E, we would of course have to insert the appropriate multiplicity $m(E)$, so that (2.1) would be replaced by

$$\det A = \prod_E E^{m(E)} = \exp\left[\sum_E m(E)\ln(E)\right], \tag{2.5}$$

and (2.2) would be replaced by

$$\zeta(z) = \sum_E m(E)E^{-z}, \tag{2.6}$$

where the product and the sum run over distinct values of E. Equations (2.5) and (2.6) are the appropriate point of departure for generalizing the method to continuous spectra. We will see below that this will involve more than just replacing the sum in (2.5) and (2.6) by an integral. This is because in general one must be careful about orders of integration and limit processes. This will lead us to define a zeta function which depends on x and x'.

Consider the case where $A(\partial,x)$ has a complete set of orthonormal eigenfunctions $\eta(\alpha,x)$ with continuous eigenvalues $E(\alpha)$, i.e.,

 0022-2488/84/030629-04$02.50

$$A(\partial,x)\eta(\alpha,x) = E(\alpha)\eta(\alpha,x), \tag{2.7}$$

$$\sum\!\!\!\!\!\!\int d^N\alpha\, W(\alpha,x')\eta(\alpha,x)\eta^*(\alpha,x') = \delta^N(x,x'), \tag{2.8}$$

$$\int d^Nx\, W(\alpha',x)\eta(\alpha,x)\eta^*(\alpha',x) = \delta^N(\alpha,\alpha'), \tag{2.9}$$

where $W(\alpha,x)$ is a combination of the "weight function"[5] and normalization factors. Note that we have assumed that there will be N parameters α_i, $i = 1,2,...,N$ needed to label the eigenfunctions and eigenvalues. Certainly this will be true in general. The parameter α_i may be continuous or discrete. This is the reason for the notation $\sum\!\!\!\!\!\!\int$ in (2.8); it indicates summation for the discrete α_i and integration for the continuous α_i. Also $\delta_N(\alpha,\alpha')$ is a product of Kronecker δ-functions for the discrete α_i and Dirac δ-functions for the continuous α_i. We are of course considering the case where at least one of the α_i is continuous so that the spectrum is continuous.

Using (2.8) and the power series for ln we get

$$\ln[A(\partial,x)\delta^N(x,x')] = -\sum_{n=0}^{\infty}\frac{1}{n+1}\sum\!\!\!\!\!\!\int d^N\alpha_1\cdots d^N\alpha_{n+1}\left[\prod_{i=1}^{n+1}(1-E(\alpha_i))\right] \times \int d^Nx_1\cdots d^Nx_n \prod_{i=1}^{n+1}[W(\alpha_i,x_i)\eta(\alpha_i,x_{i-1})\eta^*(\alpha_i,x_i)],$$

where $x_0 = x$ and $x_{n+1} = x'$. Using (2.9) we see that the d^Nx_i integrations yield a product of δ-functions on the α_i and thus

$$\ln[A(\partial,x)\delta^N(x,x')] = \sum\!\!\!\!\!\!\int d^N\alpha\, W(\alpha,x')\eta(\alpha,x)\eta^*(\alpha,x')\ln E(\alpha). \tag{2.10}$$

Now since $\det A = \exp \operatorname{tr} \ln A$ we have

$$\det A = \exp \operatorname{tr} \sum\!\!\!\!\!\!\int d^N\alpha\, W(\alpha,x')\eta(\alpha,x)\eta^*(\alpha,x')\ln E(\alpha). \tag{2.11}$$

This calculation has appeared before in the literature.[6]

The symbol tr in the above formula denotes the trace, i.e.,

$$\operatorname{tr} f(x,x') = \int d^Nx\, d^Nx'\, \delta^N(x,x')f(x,x') = \int d^Nx\, f(x,x). \tag{2.12}$$

It is tempting to do either of two things; interchange the order of the $d^N\alpha$ integration with the trace and use (2.9) to obtain

$$\det A = \exp\left[\delta^N(0)\sum\!\!\!\!\!\!\int d^N\alpha \ln E(\alpha)\right], \tag{2.13}$$

or maintain the order of integration but use $\delta^N(x,x')$ to write

$$\det A = \exp\left[\int d^Nx \sum\!\!\!\!\!\!\int d^N\alpha\, W(\alpha,x)|\eta(\alpha,x)|^2 \ln E(\alpha)\right]. \tag{2.14}$$

We will show in the next section by examining a particular example that it is not valid in general to use either (2.13) or (2.14) to calculate the determinant.

Following DeWitt, but maintaining the order of integration and limit processes, let us define

$$\bar\zeta(z,x,x') = \sum\!\!\!\!\!\!\int d^N\alpha\, W(\alpha,x')\eta(\alpha,x)\eta^*(\alpha,x')E^{-z}(\alpha) \tag{2.15}$$

and

$$\zeta(z,x) = \bar\zeta(z,x,x) = \lim_{x\to x'}\zeta(z,x,x'). \tag{2.16}$$

Note that

$$\bar\zeta(n,x,x') = G^n(x,x') = \int d^Nx_1\cdots d^Nx_n\, G(x,x_1)\cdots G(x_{n-1},x') \tag{2.17}$$

and

$$\zeta(n,x) = G^n(x,x) = \lim_{x\to x'} G^n(x,x'), \tag{2.18}$$

where

$$G(x,x') = \sum\!\!\!\!\!\!\int d^N\alpha\, W(\alpha,x')\eta(\alpha,x)\eta^*(\alpha,x')E^{-1}(\alpha) \tag{2.19}$$

is the Green's function for $A(\partial,x)$.

If $A(\partial,x)$ is an elliptic operator, then $G(x,x')$ will be analytic but if $A(\partial,x)$ is a hyperbolic operator then $G(x,x')$ will be given by a generalized function or distribution.[7] Thus as long as one is working in the Euclidean sector[1] where the wave equation is elliptic, then $\bar\zeta(n,x,x')$ will be analytic since it is the integrated product of analytic functions as shown in (2.17). If we now analytically continue n to z, then $\bar\zeta(z,x,x')$ will be analytic except for branch points since it is essentially an analytic function raised to the power z. Note that usually these branch points will be at infinity because this is where $G(x,x')$ vanishes.

Thus we define the determinant by

$$\det A = \exp\int d^Nx(-\partial_z\zeta(z,x))_{z=0}. \tag{2.20}$$

This is the most general and in that sense the most correct way of defining the determinant because it does not involve interchanging orders of integration and limit processes. The reason for keeping the d^Nx integration for last is that sometimes $(\partial_z\zeta(z,x))_{z=0}$ is independent of x and thus $\det A$ is infinite, but (2.20) shows that in this case infinity comes from the volume of space-time and is not due to a short distance divergence.

Before ending this section let us examine (2.11) and compare it with (2.5). We see from this that one must consider $W(\alpha,x')\eta(\alpha,x)\eta^*(\alpha,x')$ as the multiplicity density of the eigenvalues. This is as close as one can get to $m(E)$ when the orders of integration and limit processes cannot be interchanged. To see what $m(E)$ would be in terms of $W(\alpha,x')\eta(\alpha,x)\eta^*(\alpha,x')$ when we can interchange one order of integration with the rest, make the following change of variables: α_i goes to $E,\beta_a,a = 1,...,N-1$. Then $d^N\alpha = J(E,\beta)d^{N-1}\beta$ with $J(E,\beta)$ being the Jacobian of the transformation. Now assume that we can interchange the dE integration with the others to get

$$\det A = \exp\int dE\left[\operatorname{tr}\sum\!\!\!\!\!\!\int d^{N-1}\beta\, J(\beta,E) \times \widetilde W(\beta,E,x')\tilde\eta(\beta,E,x)\tilde\eta^*(\beta,E,x')\right]\ln E,$$

which shows the quantity in square brackets to be $m(E)$.

It should be noted that (2.17) indicates that it is not necessary to know the eigenfunctions and eigenvalues in order to find $\bar{\zeta}(z,x,x')$; one need only know the Green's function.

3. AN ILLUSTRATIVE EXAMPLE

Here we will calculate the one-loop correction to the standard ϕ^4 field theory in the presence of a massless plane wave background field. The reason for doing this particular calculation is that it illustrates very nicely just how careful one must be. We will use (2.20) to get the correct answer which has been calculated by different methods,[8,9] and then show how and why (2.13) and (2.14) would yield the wrong result. We will follow the steps in Ref. 6 for calculating the unrenormalized result.

Consider the action for the standard ϕ^4 field theory where ϕ is a real scalar field,

$$S[\phi] = \int d^4x \, \mathscr{L}(x) = \int d^4x \left[\tfrac{1}{2}\partial_\mu\phi\partial^\mu\phi - \tfrac{1}{2}M^2\phi^2 - \tfrac{1}{4}g\phi^4\right]. \quad (3.1)$$

The one-loop correction to this action in the presence of a massless plane wave background $\phi_c(k_\mu x^\mu)$ is

$$S_1 = \tfrac{1}{2}i \ln \det\left[(\partial^2 + M^2 + 3g\phi_c^2(k_\mu x^\mu))\delta^4(x,x')\right], \quad (3.2)$$

where k_μ is a given constant four-vector with $k_\mu k^\mu = 0$. Without loss of generality we can take $k_1 = k_2 = 0$ and thus $k_\mu x^\mu = k_0x_0 - k_3x_3 = k_0(x_0 + x_3)$. Thus it will be useful to work in null plane coordinates[9]:

$$x_\pm \equiv (1/\sqrt{2})(x_0 \pm x_3) = (1/\sqrt{2})(x^0 \mp x^3) \equiv x^\mp.$$

In terms of these coordinates inner products take the form

$$A_\mu B^\mu = A_+B_- + A_-B_+ - A_i B_i,$$

where i is summed over the values 1,2. In the rest of this section $i,j,\ldots$ take the values 1,2 unless otherwise specified. The eigenvalue equation we need to solve is

$$\left[2\partial_+\partial_- - \partial_i\partial_i + M^2 + 3g\phi_c^2(k_-x_+)\right]\eta(p,x) = E(p)\eta(p,x), \quad (3.3)$$

where α has been replaced with p.

The eigenfunctions are easily seen to be

$$\eta(p,x) = (2\pi)^{-2} \exp\left[-ip_\mu x^\mu - \frac{i}{2p_+}\int^{x_+} 3g\phi_c(k_-y_+)dy_+\right] \quad (3.4)$$

with eigenvalues

$$E(p) = -2p_+p_- + p_i p_i + M^2. \quad (3.5)$$

They can be shown to satisfy

$$\int_{-\infty}^{+\infty} d^4p \, \eta(p,x)\eta^*(p,x') = \delta^4(x,x'), \quad (3.6)$$

$$\int_{-\infty}^{+\infty} d^4x \, \eta(p,x)\eta^*(p',x) = \delta^4(p,p'), \quad (3.7)$$

and thus $W(p,x) = 1$.

Substituting these results into (2.15) and using

$$\lambda^{-z} = \frac{(-i)^z}{\Gamma(z)}\int_0^\infty s^{z-1}\exp(is\lambda)ds, \quad (3.8)$$

we get

$$\begin{aligned}\bar{\zeta}(z,x,x') &= \frac{(-i)^{z+1}}{\Gamma(z)}\int_0^\infty ds\, s^{z-1} \times \int_{-\infty}^{+\infty} d^4p\, \eta(p,x)\eta^*(p,x')e^{isE(p)} \\ &= \frac{(-i)^{z+1}}{\Gamma(z)16\pi^2}\int_0^\infty ds\, s^{z-3}\exp\left[\frac{i}{4s}(x-x')^2 + isM^2 + \frac{1}{x_+ - x'_+}\int_{x'_+}^{x_+} 3g\phi_c^2(k_-y_+)dy_+\right], \quad (3.9)\end{aligned}$$

and therefore

$$\zeta(z,x) = \frac{i}{16\pi^2(z-1)(z-2)}(M^2 + 3g\phi_c^2(k_-x_+))^{2-z}, \quad (3.10)$$

after using (3.8) again. We now have

$$(\partial_z\zeta(z,x))_{z=0} = \frac{i}{32\pi^2}(M^2 + 3g\phi_c^2(k_-x_+))^2 \times \left[\tfrac{3}{2} - \ln(M^2 + 3g\phi_c^2(k_-x_+))\right],$$

and finally

$$S_1 = \frac{-1}{64\pi^2}(M^2 + 3g\phi_c^2(k_-x_+))^2 \times \left[\tfrac{3}{2} - \ln(M^2 + 3g\phi_c^2(k_-x_+))\right]. \quad (3.11)$$

This result matches those obtained by other methods[10] up to a finite renormalization.

Note that if we had tried using (2.14) to define $\zeta(z,x)$ rather than (2.16), then in (3.9) we would have $|\eta(p,x)|^2$ rather than $\eta(p,x)\eta^*(p,x')$. But $|\eta(p,x)|^2$ is completely independent of ϕ_c and we would not have obtained the correct answer for S_1. The reason why we cannot set $x = x'$ before the d^4p integration in (3.9) is obvious in this case: $\eta(p,x)\eta^*(p,x')$ has an essential singularity at $p_+ = 0$ which contributes to the integral, setting $x = x'$ in the integrand eliminates this singularity which changes the result for the integral. In general it would seem that if $W(\alpha,x')\eta(\alpha,x)\eta^*(\alpha,x')$ has poles and/or singularities in the complex α plane and if the poles and/or singularities are not present in $W(\alpha,x)|\eta(\alpha,x)|^2$, then one must do the $d^N\alpha$ integration before setting $x = x'$. If there are no poles or singularities which contribute to the integral, then one can set $x = x'$ before doing the $d^N\alpha$ integration.

Note next that the eigenvalues (3.5) are the same as the eigenvalues for $\phi_c = 0$, and thus had we tried using (2.13) to define the determinant (through the zeta function) we would have obtained the result one gets in the absence of any background. Furthermore, using (2.13) would imply that all operators with the same spectrum would have the same determinant regardless of the form of the operator. This is obviously nonsense and is reminiscent of the "joke proof" that the S matrix must be unity.[11] It is well known[12] that the open channel elements of the S matrix can be calculated directly from $\det[(H_0 + gV)/H_0]$, where H_0 and $(H_0 + gV)$ are assumed to have the same spectrum. Using (2.13) would then give $\det[(H_0 + gV)/H_0] = 1$ independent of the value of g, the coupling constant. This then implies that the S matrix is independent of g and thus, for any g, has the same value. Since $S(g = 0) = 1$, then $S = 1$ for any g, which is obviously incorrect.

ACKNOWLEDGMENTS

The author would like to thank the Lewes Center for Physics (where this work was begun) for their hospitality, and Alan Chodos and Mike Vaughn for many helpful and stimulating discussions.

[1]N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge U.P., Cambridge, 1982), Chap. 6; S. W. Hawking, Commun. Math. Phys. **55**, 133 (1977); S. W. Hawking, in *General Relativity, An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U.P., Cambridge, 1979), p. 746.

[2]B. DeWitt, in *General Relativity, An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U.P., Cambridge, 1979), p. 680.

[3]S. W. Hawking, in *General Relativity, An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U.P., Cambridge, 1979), p. 746.

[4]See, for example, R. Jackiw, Phys. Rev. D **9**, 1686 (1947).

[5]See, for example, G. Arfken, *Mathematical Methods for Physicists* (Academic, New York, 1970) 2nd ed., Chap. 9.

[6]G. M. Gallatin, Phys. Rev. D **25**, 434 (1982).

[7]L. Schwartz, *Theorie des distributions*, Vol. 2, *Actualities scientifiques et industrieles* (Hermann and Cie, Paris, 1950, 1951), pp. 1091, 1122.

[8]S. G. Matinyan and G. K. Savvidi, Yad. Fiz. **25**, 218 (1977) [Sov. J. Nucl. Phys. **25**, 118 (1977)].

[9]See, for example, H. Bacry and N. P. Chang, Ann. Phys. (N.Y.) **47**, 407 (1968); or J. B. Kogut and D. E. Soper, Phys. Rev. D **1**, 2901 (1970), and references therein.

[10]C. Itzykson and J. -B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980), p. 454.

[11]M. Vaughn (private communication).

[12]See, for example, R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966), p. 515.

A possible approach to the construction of the :exp ξ:$_d$ quantum field theory in a finite volume

Edward P. Osipov
Department of Theoretical Physics, Institute for Mathematics, 630090, Novosibirsk, 90, USSR

(Received 4 February 1983; accepted for publication 25 February 1983)

Using a solution of some Euclidean invariant problem of moments, we construct a Euclidean realization (= a Euclidean field) for the Wightman–Jaffe quantum field $:\exp\varphi(x):_d$, where $\varphi(x)$ is the free quantum field. With the help of this Euclidean realization we propose a new nonperturbative mathematically rigorous approach to the construction of the quantum field theory with the exponential interaction in a finite volume of d-dimensional space-time ($d\geqslant 2$) and without ultraviolet cutoffs. In particular, for pure imaginary values of the coupling constant the generating functional for Schwinger functions is constructed. The expansion of the theory which is constructed by using this method in powers of the coupling constant gives the renormalized perturbation series.

PACS numbers: 11.10.Mn, 11.10.Gh

1. INTRODUCTION

Constructive quantum field theory has by now provided several nontrivial examples of models of quantum field theory in two- and three-dimensional space-time satisfying all Wightman and/or Haag–Kastler axioms (see, for instance, Refs. 1–6 and references therein).

At the same time the investigation of models for quantum fields in four-dimensional space-time is hampered by the presence of ultraviolet divergences. The analysis of perturbation theory makes it more or less clear that every bare interaction has to be modified at small distances; in other words, it has to involve counterterms (but cf. the paper by Petrina and Rebenko.[7])

The structure of counterterms is very complicated and allows an evident nonperturbative description for superrenormalizable theories only. The description of counterterms is renormalizable and, moreover, in nonrenormalizable theories is very, very complicated. A nonperturbative attempt to take into account counterterms in renormalizable theories was undertaken by Schrader.[8–10]

In the present paper we undertake a new mathematically rigorous nonperturbative attempt to construct the $:\exp\xi:_d$ quantum field theory in a finite volume of d-dimensional space-time ($d\geqslant 2$) and without ultraviolet cutoffs. Our method manifestly takes into account the presence of ultraviolet divergences and counterterms.

The $:\exp\xi:_4$ quantum field theory was thoroughly investigated in perturbation theory; see, for instance, Refs. 11 and 12. Using the technique of superpropagators, it was shown that the Green functions exist in the sense of perturbation theory. These Green functions have a nonpolynomial increase in momentum space and are Gel'fand–Shilov–Jaffe-type ultradistributions. Moreover, calculations have been made in the chiral quantum field theory of phases of $\pi\pi$ scattering, of weak decays of mesons, and of the mass difference of K_L and K_S mesons. These calculations used the technique of superpropagators and are in fair agreement with observation. For the application of models with nonpolynomial Lagrangians for the description of real physical phenomena, see the monograph by Volkov and Pervushin.[13]

In Refs. 14 and 15 was proved the necessity of an introduction of counterterms in order to obtain a nontrivial theory with the exponential interaction in four-dimensional space-time.

The main idea of our approach consists in making a formal change of variables $\xi(x)\rightarrow :\exp\xi(x):_d$ + *counterterms* and in the mathematically rigorous definition of generated formal expressions. For this purpose we construct a Euclidean realization, that is, a Euclidean field (see Sec. 5), for the quantum Wightman–Jaffe field $:\exp\varphi(x):_d$, and $\varphi(x)$ is the free massive quantum field. To construct the Euclidean realization, we formulate some Euclidean invariant problem of moments, that is, the problem of the representation of a sequence of ultradistributions as moments of a Euclidean invariant complex measure. This problem of moments consists of the following. Let $\xi(x)$ be the free Euclidean field and μ_0 be the corresponding Gaussian measure. There will be found a Euclidean invariant measure μ, defined on some space of generalized functions and such that the following formal equalities are fulfilled:

$$\int d\mu\, \Phi(x_1)\otimes\cdots\otimes\Phi(x_n)=\int d\mu_0(\xi)\prod_{i=1}^{n}:\exp\xi(x_i):$$

for all $x_i\neq x_j$ for $i\neq j$ and all n. A rigorous definition is given in Theorem 2.1. We reduce the solution of this problem of moments to the problem of extension of a linear functional defined on a subspace and use the Hahn–Banach theorem. Then we use the obtained Euclidean realization to construct the $:\exp\xi:_d$ quantum field theory. In other words, to construct the $:\exp\xi:_d$ theory, we start not from the Euclidean realization $\xi(x)$ (= the Gaussian generalized random process) of the free quantum field $\varphi(x)$, but we construct the interacting theory with the help of a Euclidean realization of some representative of the Borchers class of the free field, namely, with the help of a Euclidean realization $\Phi(x)$ of the quantum Wightman–Jaffe field $:\exp\varphi(x):_d$. Formally, $\Phi(x)$ is equal to $:\exp\xi(x):_d$ + *counterterms.*

Unlike the noncoincident (i.e., defined at noncoinciding points) Schwinger functions of the free field, the noncoincident Schwinger functions of the quantum field $:\exp\varphi(x):_d$

 0022-2488/84/030633-16$02.50

are nonintegrable and for their extension at coinciding points one has to consider them as ultradistributions and to apply the Hahn–Banach theorem. In addition, Schwinger functions only define moments of a measure and, while in the case of the free field the measure is uniquely reconstructed by their moments, in the general case this is not so. Therefore, in general, a Euclidean field is not given uniquely by its Schwinger functions alone. For uniqueness one has to require the fulfilment of some more conditions. The problems connected with uniqueness of the construction of a Euclidean field require an additional study.

Using the method proposed in this paper, we construct the generating functional for Schwinger functions for pure imaginary values of the coupling constant of the $:\exp \xi:_d$ Euclidean theory without an ultraviolet cutoff and with a fixed space-time one. The expansion of the theory, which is constructed by using this method in powers of the coupling constant, gives the renormalized perturbation series.

This slightly generalized method is also applicable to polynomial interactions in d-dimensional space-time,[16] although in this case the problem of the connection with the usual consideration of the ξ_3^4 model[2–5] is not clear.

Analogous arguments are also applicable to the construction of Green functions,[17,18] that is, to the construction of a Feynman integral.

We use the opportunity to correct mistakes contained in Refs. 17 and 18. In the proof of Theorem 2.1[18] the equality after formula (2.4) is only valid for $k_1 = \cdots = k_n$. One should exclude the remark after Theorem 2.1[18] and assume that the spaces $\mathscr{C}^{\alpha}_{\mathrm{Re}}(\mathbb{R}^4)$ (see Sec. 2) with $1<\alpha<\frac{3}{2}$ are used in Theorem 3.1.[17] In the proof of Theorem 3.1[18] the assertion about the τ continuity follows now from Theorem 2.1,[18] from inequality 3.1,[18] and from the inequality

$$\int d^4x \left| \alpha\left(2^{j/2}Bi\frac{\partial}{\partial x^0}\right) f(x^0,x)\right| \leqslant c_j \int d^4x \left| \beta\left(i\frac{\partial}{\partial x^0}\right) f(x^0,x)\right|,$$

where $\alpha(\cdot)$ and $\beta(\cdot)$ are entire functions of the form [Ref. 17, formula (2.1)] with indices $\alpha\leqslant\frac{3}{2}$, $1<\beta<\alpha$, $c_j = \int dx^0\, |\gamma_j(x^0)|$, and $\gamma_j(x^0)\in L_1(\mathbb{R})$ is the function with the Fourier transform $\gamma_j^{\sim}(p^0) = \beta(p^0)^{-1}\alpha(2^{j/2}Bp^0)$ (the inclusion $\gamma_j\in L_1$ follows from Lemma 1.3[18]). In addition in the inequality before (1.1)[18] c^k should be replaced by $c_k(f)$.

However, in the case of a Feynman integral a location of this approach as a whole is less clear. In particular, there exists no translation invariant measure whose moments are equal to the Green functions of the free field.[19] Then, while in contrast to the case of the theory in the Euclidean region there does not appear a problem of integrability of an unbounded exponent for the Feynman integral (more precisely, it can be avoided), but (for the theory in the infinite volume) a problem arises about the Poincaré invariance of the obtained Green functions. In the two-dimensional case the Poincaré group is amenable,[20] and, using an invariant mean on the two-dimensional Poincaré group, one can try to obtain a Poincaré-invariant (finitely additive) measure, but the measure obtained in such a way has, most likely, no moments and so it cannot be used to obtain an integral representation for Green functions.

Our paper is organized as follows.

In Sec. 2 we formulate the main results of our paper, outline their proofs and give a short discussion.

In Sec. 3 we investigate linear functionals on the symmetric tensor algebra $S(\mathscr{C})$ over the Jaffe-type space $\mathscr{C}$, which are given by measures on $\mathscr{C}'_{\mathrm{Re}}$.

In Sec. 4 we characterize the topology which is weaker than the topology in which linear continuous functionals on $S(\mathscr{C})$ are given by measures on $\mathscr{C}'_{\mathrm{Re}}$.

In Sec. 5 we formulate sufficient conditions for the solvability of a Euclidean invariant problem of moments and prove the existence of a Euclidean realization for the quantum field $:\exp \varphi(x):_d$.

By c with or without indices we denote various strictly positive constants, possibly depending on unessential variables, $\langle\cdot,\cdot\rangle$ denotes the action of an ultradistribution on a test function.

2. MAIN RESULTS

In this section we present the main results of our paper. We use (in configuration space) the spaces of test functions which coincide with Jaffe-type spaces $\mathscr{C}^{\alpha}(\mathbb{R}^{dn})$ of infinitely differentiable complex functions on $\mathbb{R}^{dn}$ for which the following equivalent sequences of seminorms are finite (cf. Refs. 21–23):

$$\|f\|^{(1)}_{A,k} = \sup_{m,x}\{A^{-|m|}|m|^{-|m|}|x^k\,\partial^m f(x)| \times |m\in\mathbb{Z}^{dn},\ m\geqslant 0;\ x\in\mathbb{R}^{dn}\},$$

$$\|f\|^{(2)}_{k,1} = \sup_{p,m}\{|\alpha(p)^k\partial^m f^{\sim}(p)| \times |m\in\mathbb{Z}^{dn},\ m\geqslant 0, |m|\leqslant 1;\ p\in\mathbb{R}^{dn}\}. \qquad (2.1)$$

Here $A>0$, $k,l\in\mathbb{Z}$, $k,l\geqslant 0$, we have used multiindex denotations, $f^{\sim}(p)$ is the Fourier transform of the function $f(x)$, and $\alpha(p)$ is an entire function,

$$\alpha(p) = \sum_{r=0}^{\infty}(r!)^{-2\alpha}|p|^{2r}, \qquad (2.2)$$

of the order $1/\alpha$, which gives the character of possible singularities of ultradistributions. The natural topology on $\mathscr{C}^{\alpha}(\mathbb{R}^{dn})$ can be defined by one of equivalent sequences of seminorms (2.1).

We note that other types of test functions spaces can also be used, for instance, the Gel'fand–Shilov spaces $S^{\alpha,\dots,\alpha}$.[24]

By $\mathscr{C}^{\alpha}_{\mathrm{Re}}(\mathbb{R}^d)$ we denote the (topological) subspace of real functions from $\mathscr{C}^{\alpha}(\mathbb{R}^d)$ with induced topology.

The denotations $\mathscr{C}^{\alpha}(\mathbb{R}^d)$ and $\mathscr{C}^{\alpha}_{\mathrm{Re}}(\mathbb{R}^d)$ we shall often abbreviate to $\mathscr{C}^{\alpha}$ and $\mathscr{C}^{\alpha}_{\mathrm{Re}}$ or to $\mathscr{C}$ and $\mathscr{C}_{\mathrm{Re}}$.

We introduce on $\mathscr{C}'_{\mathrm{Re}}$ the weak topology $\sigma(\mathscr{C}'_{\mathrm{Re}},\mathscr{C}_{\mathrm{Re}})$ and use for integration the theory presented in the Bourbaki's tractate,[25] that is, the measures used by us are Radon measures, in other words, they are countably additive inner regular (weakly) Borel measures, cf. Ref. 25, Chap. IX, §3, no. 2, Theorem 2). We also remark that for the spaces $\mathscr{C}'_{\mathrm{Re}}$

all the natural definitions of integrals and σ-algebras are identical.

Let us denote by $\mathscr{P}_0$ the subspace of polynomial functions on $\mathscr{C}'_{\rm Re}$ of the form

$$f(\Phi) = f_0 + \langle \Phi, f_1 \rangle + \sum_{n=2}^{\infty} \langle \Phi \otimes \cdots \otimes \Phi, f_n \rangle, \tag{2.3}$$

where $f_0 \in \mathbb{C}$, $f_1 \in \mathscr{C}$, $f_n \in \mathscr{C}_0(\mathbb{R}^{dn})$ for $n \geqslant 2$ and only a finite number of f_n is not equal to zero. Here, $\mathscr{C}_0^\alpha(\mathbb{R}^{dn}) = \{ f \in \mathscr{C}^\alpha(\mathbb{R}^{dn}) | \partial^k f(x_1,...,x_n) = 0$ for all multiindices $k \geqslant 0$ if some $x_i = x_j$ for $i \neq j\}$. $\mathscr{C}_0^\alpha(\mathbb{R}^{dn})$ is a closed subspace, and we equip it with the induced topology.

We remark that for $\alpha > 1$ the subspaces $\mathscr{C}_0^\alpha(\mathbb{R}^{dn})$ are nontrivial (see, for instance, Refs. 22 and 23).

The main idea of our approach consists of constructing a Euclidean field $\Phi(x)$ corresponding to the normal ordered exponent of the free quantum field and of using this Euclidean field for the construction of the quantum field theory with the exponential interaction. The following assertion about the existence of a Euclidean field for the quantum field $:\exp \varphi(x):_d$ is valid.

Theorem 2.1: Let $1 < \alpha < d - 1/d - 2$. There exists a real Euclidean invariant measure μ on $\mathscr{C}^{\alpha\prime}_{\rm Re}$ such that the unity and all functions of the form $\langle \Phi \otimes \cdots \otimes \Phi, h_n \rangle$, $h_n \in \mathscr{C}^\alpha(\mathbb{R}^{dn})$, are integrable, and, if $f \in \mathscr{P}_0$ and has the form 2.3) then

$$\int d\mu(\phi) \Big(f_0 + \langle \phi, f_1 \rangle + \sum_{n=2}^{\infty} \langle \Phi \otimes \cdots \otimes \Phi, f_n \rangle \Big) = S_0^0 f_0 + \int d^d x\, S_1^0(x) f_1(x) + \sum_{n=2}^{\infty} \int d^{dn} x\, S_n^0(x_1,...,x_n) f_n(x_1,..., x_n). \tag{2.4}$$

Here S_n^0, $S_0^0 = 1$, $S_1^0(x) = 1$,

$$S_n^0(x_1,..., x_n) = \exp\Big[\sum_{1 < i < j < n} G(x_i - x_j) \Big] \quad \text{for } n \geqslant 2,$$

$$G(x) = (2\pi)^{-d} \int d^d p\, (p^2 + m^2)^{-1} \exp(ipx), \quad m > 0,$$

are the noncoincident n-point Schwinger functions of the strictly localizable Wightman–Jaffe quantum field $:\exp \varphi(x):_d$, where $\varphi(x)$ is the d-dimensional free scalar massive quantum field and double colons : : denote the Wick ordering.

Theorem 2.1 will be proved in Sec. 5.

With the help of the measure μ, whose existence is proved in Theorem 2.1, one may represent the partition function of the theory with the exponential interaction and with a space-time cutoff Λ in the following form:

$$Z_\Lambda = \int d\mu(\Phi) \exp(-g\langle \Phi, \Lambda \rangle), \tag{2.5}$$

where g is the coupling constant. For $\Lambda \in \mathscr{C}^\alpha_{\rm Re}$, $\alpha < d - 1/d - 2$, the integrand is a μ-measurable unbounded function. Unfortunately, we are not able to prove μ-integrability of this function and so, *a priori*, (2.5) is only defined for pure imaginary values of the coupling constant.

With the help of the measure μ one may also write the expression for the generating functional of Schwinger functions of the quantum field theory with the exponential interaction. Indeed, with the help of the Wick theorem[26] the formal expression

$$\int d\mu_0(\xi) \exp\Big[-g \int d^d x :\exp \xi(x): \Lambda(x) + \int d^d x\, \xi(x) f(x) \Big]$$

for the unnormalized generating functional of Schwinger functions of the $:\exp \xi:_d$ theory can be represented in the form of a formal continual integral

$$\exp[-\tfrac{1}{2}(fGf)] \int d\mu_0(\xi) \times \exp\Big[-g \int d^d x :\exp \xi(x): \Lambda_f(x) \Big], \tag{2.6}$$

whose integrand contains $\xi(x)$ in the form $:\exp \xi(x):$ only. Here μ_0 is the Gaussian measure, corresponding to the free Euclidean $\xi(x)$, $(fGf) = \int d^d x\, d^d y\, f(x) G(x-y) f(y)$, $\Lambda_f(x) = \Lambda(x) \exp[\int d^d y\, f(x-y) G(y)]$.

If one tries to define (2.6) as a limit of the corresponding expressions with ultraviolet cutoffs, then these expressions tend to the corresponding expressions for the free theory when the ultraviolet cutoffs are removed.[14,15] On the other hand, in the integrand (2.6) $\xi(x)$ is involved in the form $:\exp \xi(x):$ only, and so it is natural to use Theorem 2.1 and to try to define the unnormalized generating functional for Schwinger functions as

$$J_\Lambda(f) = \exp[-\tfrac{1}{2}(fGf)] \int d\mu(\Phi) \exp(-g\langle \Phi, \Lambda_f \rangle). \tag{2.7}$$

The measurability of the integrand in (2.7) follows from:

Lemma 2.2: Let $f(x), g(x) \in \mathscr{C}^\alpha$, $\alpha \geqslant 1$; then $g(x) \exp[f(x)] \in \mathscr{C}^\alpha$.

Proof: Let ∂_i be the partial derivative with respect to the ith coordinate. We prove by induction on n that

$$\sup_{x \in \mathbb{R}^d} \Big| \Big(\prod_{i=1}^d \partial_i^{m_i} \Big) f(x)^n \Big| \leqslant c^n \prod_{i=1}^d (A^{m_i} m_i!^\alpha 2^{n+m_i}) \tag{2.8}$$

for every $A > 0$ and some c depending on A. Since $f(x) \in \mathscr{C}^\alpha$,

$$\sup_{x \in \mathbb{R}^d} \Big| \Big(\prod_{i=1}^d \partial_i^{m_i} \Big) f(x) \Big| \leqslant c \prod_{i=1}^d (A^{m_i} m_i!^\alpha) \tag{2.9}$$

for every $A > 0$ and some c depending on A, and, thus, for $n = 1$ assumption (2.8) is fulfilled. Assuming that it is fulfilled for n, we have for $n+1$

$$\left|\left(\prod_{i=1}^{d} \partial_i^{m_i}\right) f(x)^{n+1}\right|$$
$$= \left| \sum_{i_1=0}^{m_1} \cdots \sum_{i_d=0}^{m_d} \frac{m_1! \cdots m_d!}{i_1!(m_1-i_1)! \cdots i_d!(m_d-i_d)!} \times \partial_1^{i_1} \cdots \partial_d^{i_d} f(x)^n \, \partial_1^{m_1-i_1} \cdots \partial_d^{m_d-i_d} f(x) \right|$$
$\leqslant$[by the induction hypothesis and (2.9)]
$$c^{n+1} \prod_{i=1}^{d} (A^{m_i} m_i!^{\alpha} 2^{n+m_i}) \times \sum_{i_1=0}^{m_1} \cdots \sum_{i_d=0}^{m_d} 2^{i_1-m_1+\cdots+i_d-m_d} \times \left[\frac{m_1! \cdots m_d!}{i_1!(m_1-i_1)! \cdots i_d!(m_d-i_d)!}\right]^{1-\alpha}.$$

Since $\alpha \geqslant 1$ and the expression in square brackets is an integer, summing over $i_1,\ldots,i_d$, we obtain the estimate

$$\left|\left(\prod_{i=1}^{d} \partial_i^{m_i}\right) f(x)^{n+1}\right| \leqslant c^{n+1} \prod_{i=1}^{d} (A^{m_i} m_i!^{\alpha} 2^{n+1+m_i}),$$

that is, the induction hypothesis for $n+1$. Thus, (2.8) is proved. (2.8) implies that

$$\sup_{x\in\mathbb{R}^d} \left|\left(\prod_{i=1}^{d} \partial_i^{m_i}\right) \exp[f(x)]\right| \leqslant \sum_{n=0}^{\infty} n!^{-1} \sup_{x\in\mathbb{R}^d} \left|\left(\prod_{i=1}^{d} \partial_i^{m_i}\right) f(x)^n\right| \leqslant \sum_{n=0}^{\infty} \frac{c^n}{n!} \prod_{i=1}^{d} (A^{m_i} m_i!^{\alpha} 2^{n+m_i}) = \exp(c2^d) \prod_{i=1}^{d} (A^{m_i} m_i!^{\alpha} 2^{m_i})$$

for every $A>0$ and some c depending on A. This estimate and the statement of the Gel'fand and Shilov's book [Ref. 20, Chap. IV, §4.2b, §9) imply that $g(x)\exp[f(x)] \in \mathscr{C}^{\alpha}$. Lemma 2.2 is proved.

Lemma 2.2 implies that for $\Lambda, f \in \mathscr{C}^{\alpha}_{\rm Re}$, $1 \leqslant \alpha < d-1/d-2$, the function $\exp(-g\langle \Phi, \Lambda_f\rangle)$ is μ-measurable and for pure imaginary values of the coupling constant g is bounded. Therefore, (2.5) and (2.7) are defined correctly for pure imaginary values of the coupling constant.

By definition we set that the (unnormalized) generating functional for Schwinger functions of the :exp ξ:$_d$ quantum field theory with the space-time cutoff Λ is given by (2.7). Thus, we obtain an integral representation in the region of pure imaginary values of the coupling constant g.

Unfortunately, we are not able to prove that the function $\exp(-g\langle \Phi, \Lambda_f\rangle)$ is μ-integrable for positive values of the coupling constant g. We also are not able to give a rigorous meaning as a change of variables to the transition from the free field $\xi(x)$ to the field $\Phi(x)$. Formally, these fields are connected with each other by a relation $\Phi(x) =$:exp $\xi(x)$: + *counterterms*. In order to interpret this relation literally as a change of variables, one has to overcome great difficulties of both combinatorial and analytical nature.

Analogous constructions can be performed for a polynomial interaction as well. The difference is that, for the polynomial interaction in the integral representation of the generating functional of Schwinger functions corresponding to (2.7), an integral arises over several fields $\Phi_k(x)$, corresponding to :$\xi(x)^k$: + *counterterms*.

We note that the arising Euclidean field $\Phi(x)$ [in the polynomial case several fields $\Phi_k(x)$] is a Euclidean field for the quantum field :exp $\varphi(x)$:$_d$ belonging to the Borchers class of equivalence of the free field.

Our construction is analogous to that of the renormalized Hilbert space for the ξ^4_3 theory, where the Hamiltonian is a densely defined symmetric operator. The integrability of $\exp(-\langle \Phi, \Lambda_f\rangle)$ corresponds to the boundedness from below of the renormalized Hamiltonian and is an open problem.

The proof of Theorem 2.1 is based on the Hahn–Banach theorem. We present here a sketch of the proof of Theorem 2.1.

Let $\mathscr{F}$ be the set of functions on $\mathscr{C}'_{\rm Re}$ of the form $\Phi \mapsto g(a_1(\Phi),\ldots,a_n(\Phi))$, where the g are polynomially bounded continuous functions,

$$a_i(\Phi) = (a_i)_0 + \sum_{k=1}^{\infty} \langle \underbrace{\Phi \otimes \cdots \otimes \Phi}_{k}, (a_i)_k\rangle,$$
$$(a_i)_0 \in \mathbb{C},\ (a_i)_k \in \mathscr{C}(\mathbb{R}^{dk})$$

and only a finite number of $(a_i)_k$ (depending on a_i) are nonzero.

Using the noncoincident Schwinger functions of the quantum field :exp $\varphi(x)$:$_d$ one can define a linear functional $\mu_{\mathscr{P}_0}$ on the subspace $\mathscr{P}_0 \subset \mathscr{F}$. For this purpose for $f \in \mathscr{P}_0$ we set $\mu_{\mathscr{P}_0}(f) =$ rhs (2.4). Then $\mu_{\mathscr{P}_0}$ is correctly defined and gives a linear functional on $\mathscr{P}_0$.

Now let us choose a topology on the space of functions $\mathscr{F}$ such that the dual space would be given by measures (= would be given by linear functionals which are defined by measures), then, for the linear functional $\mu_{\mathscr{P}_0}$ to be extended to a measure on $\mathscr{C}'_{\rm Re}$, it is sufficient by the Hahn–Banach theorem that it satisfy some continuity conditions. Namely, it should be continuous in the topology which is induced by the above topology (that is, in the topology in which the dual to $\mathscr{F}$ is given by measures). Sufficient conditions are formulated in Theorem 5.1 Thereby we prove the existence of a measure satisfying equalities (2.4).

Further, we prove the existence of a measure with Euclidean invariant estimates and use an invariant mean on the amenable group of Euclidean transformations[20] to construct a Euclidean invariant measure, satisfying equality (2.4).

In general, the measure whose existence is asserted in Theorem 2.1 is both nonpositive and nonunique. This nonuniqueness reflects both the nonuniqueness of renormalization and the nonuniqueness of the reconstruction of the theory by its perturbation series.

From the general point of view, the measure corresponding to the physical theory has to be complex due to the breaking of time-reversal symmetry.[27,28] Further, on the one hand, the nonpositivity of the measure corresponds to the old idea of an introduction into the theory of states with negative probabilities in order to remove ultraviolet divergences (these negative probabilities do not mean an introduction of negative probabilities into the physical space of states). On the other hand, the condition of (enlarged) physical positivity of the measure (= the Osterwalder–Schrader positivity condition for bounded functions) and the invariance of the measure under temporal translations imply positivity of the measure itself (see Refs. 29 and 30). Are there among measures of Theorem 2.1 measures satisfying such enlarged physical positivity conditions? Moreover, expression (2.7) corresponds to the generating functional, i.e., to the Laplace transformation of the measure corresponding to the interacting theory, and the connection of its properties with those of the measure from Theorem 2.1 requires further investigation.

To construct the theory for real (positive) values of the coupling constant, it would be sufficient to prove that the function $\exp(-g\langle \Phi,\Lambda_f\rangle)\in L_1(\mathscr{C}'_{\mathrm{Re}},|\mu|)$. Properly speaking, it would be sufficient to prove that the generating functional $J_\Lambda(f)$ is analytic in g in the region $\operatorname{Re} g>0$. We note that, due to the fast increase of the noncoincident Schwinger functions of the field $:\exp\varphi(x):_d$ with the increase of n, one is to expect that the generating functional $J_\Lambda(f)$ is nonanalytic in g at zero.

The effective construction of approximations of measures from Theorem 2.1 would advance a further investigation of the theory with the exponential interaction in our approach.

It would be interesting to investigate Markov properties[1] of measures from Theorem 2.1.

If one compares our approach with the usual one, then, first of all, in the free case the measure constructed by using our approach (and integrable Schwinger functions) gives the usual Gaussian measure.

For the exponential interaction in two-dimensional space-time[1]

$$V_\Lambda(\xi)=\int d^2x\ :\exp\lambda\xi(x):_2\Lambda(x),\quad \lambda\in(-\sqrt{4\pi},\sqrt{4\pi}),$$

one can easily show that there exists the positive Euclidean invariant measure μ defined on $\mathscr{S}_{\mathrm{Re}}(\mathbb{R}^2)'$ ($\mathscr{S}_{\mathrm{Re}}(\mathbb{R}^n)$ ($\mathscr{S}(\mathbb{R}^n)$) is the Schwartz space of rapidly decreasing infinitely differentiable real (complex) valued functions on $\mathbb{R}^n$) satisfying the Osterwalder–Schrader positivity conditions for bounded functions and such that for $f\in\mathscr{S}_{\mathrm{Re}}(\mathbb{R}^2)$, nonnegative g, and nonnegative $\Lambda\in L_\infty(\mathbb{R})^2\cap L_1(\mathbb{R}^2)$,

$$\int d\mu_0(\xi)\exp[\langle\xi,f\rangle-gV_\Lambda(\xi)]$$

$$=\exp[-\tfrac{1}{2}(fGf)]\int d\mu(\Phi)\exp(-g\langle\Phi,\Lambda_f\rangle),$$

$$\exp(-g\langle\Phi,\Lambda_f\rangle)\in L_1(\mathscr{S}_{\mathrm{Re}}(\mathbb{R}^2)',\mu).$$

It should be noted that since $V_\Lambda(\xi)\notin L_p(\mathscr{S}_{\mathrm{Re}}(\mathbb{R}^2)',\mu_0)$ for sufficiently large p, the measure μ has moments only of the order not larger than some p_0 [and coinciding with the corresponding (integrable) Schwinger functions of the quantum field $:\exp\lambda\varphi(x):_{1+1}$]. Thus, this measure μ is not a Euclidean realization of the field $:\exp\lambda\varphi(x):_{1+1}$ in the sense of the definition of Sec. 5, where the existence of all moments is needed.

In the case of the zero-dimensional field theory, the method proposed in the paper is reduced to the solvability of the problem of moments in a class of complex countably additive measures on $\mathbb{R}$, for which all polynomials are integrable[31] (see also Sec. 4). Such a problem of moments is always solvable, and its solution is based on the fact that any (real) sequence may be represented as a difference of two positively definite sequences growing sufficiently fast (see Lemma 4.3).

3. MEASURES ON $\mathscr{C}'_{\mathrm{Re}}$ AS CONTINUOUS FUNCTIONALS ON A FUNCTION ALGEBRA

Let us introduce some definitions, see also Ref. 17.

The symmetric tensor algebra over $\mathscr{C}$ is the direct sum

$$S(\mathscr{C})=\mathop{\oplus}_{n=0}^{\infty}S_n(\mathscr{C}),$$

where $S_0(\mathscr{C})=\mathbb{C}$ and $S_n(\mathscr{C})$ is the completed n-fold symmetric tensor power of $\mathscr{C}$ equipped with the topology induced by $\mathscr{C}(\mathbb{R}^{dn})$. $S(\mathscr{C})$ is a commutative $*$-algebra where the $*$-operation is defined as complex conjugation. We introduce on $S(\mathscr{C})$ the direct sum topology and denote it by τ.

The dual space $S(\mathscr{C})'$ consists of all sequences $T=(T_0,T_1,\cdots)$ with $T_0\in\mathbb{C}$, $T_n\in\mathscr{C}(\mathbb{R}^{dn})'$ and such that T_n is invariant under permutations of arguments.

The algebra $S(\mathscr{C})$ may also be considered as the algebra of polynomial functions on $\mathscr{C}'_{\mathrm{Re}}$. The isomorphism of these algebras is defined by the relation $a\leftrightarrow\chi_\Phi(a)\equiv a(\Phi)$, where $a=(a_0,a_1,\cdots)\in S(\mathscr{C})$, χ_Φ is a continuous real character on $S(\mathscr{C})$ defined by $\chi_\Phi=(1,\Phi,\Phi\otimes\Phi,\ldots)\in S(\mathscr{C})'$, $\Phi\in\mathscr{C}'_{\mathrm{Re}}$, $\chi_\Phi(a)=a(\Phi)=a_0+\Sigma_{n=1}^{\infty}\langle\underbrace{\phi\otimes\cdots\otimes\phi}_{n},a_n\rangle$. The fact that χ_Φ is a character means that $\chi_\Phi(ab)=\chi_\Phi(a)\chi_\Phi(b)$, a, $b\in S(\mathscr{C})$. The correspondence $a\leftrightarrow\chi_\Phi(a)=a(\Phi)$ is bijective, which follows from a general polarization identity (Ref. 32, Lemma 1.5.4).

In this section we consider the following question: When a functional $T=(T_0,\cdots,T_n,\cdots)\in S(\mathscr{C})'$ has a representation of the form

$$T(a)\equiv T_0a_0+\sum_{n=1}^{\infty}\langle T_n,a_n\rangle=\int d\mu(\Phi)\,a(\Phi),$$

where μ is a complex measure on $\mathscr{C}'_{\mathrm{Re}}$ (cf. Ref. 33).

For this purpose, following Ref. 33, we introduce a space $\mathscr{F}$ of functions on $\mathscr{C}'_{\mathrm{Re}}$ of the form $f(\Phi)=g(a_1(\Phi),\ldots,a_n(\Phi))$, where $a_i\in S(\mathscr{C})$ and g are polynomially bounded continuous functions on $\mathbb{C}^n$ (naturally, a_i and n depend on f).

Following Ref. 33, we define the topology $\hat{\tau}$ on $\mathscr{F}$ by the collection of seminorms of the form

$$\|f\|_{F_p}=\sup\{|f(\Phi)|F_p(\Phi)^{-1}\,|\Phi\in\mathscr{C}'_{\mathrm{Re}}\},\qquad(3.1)$$

where each F_p is a function $\mathscr{C}'_{\mathrm{Re}}\to[0,\infty)$ of the form

$$F_p(\Phi)=\sup\{|a(\Phi)|\,|\,p(a)\leqslant1\},\qquad(3.2)$$

$a\in S(\mathscr{C})$ and p is a seminorm on $S(\mathscr{C})$ continuous in the direct sum topology τ. We note that $F_p(\Phi)\geqslant c_p>0$.

The bipolar theorem (Ref. 34, Chap. II, no. 4, Theorem 4, Corollary 1) implies that for each continuous on $S(\mathscr{C})$ seminorm p we have

$$p(a) = \sup\{|T(a)|\, p^\circ(T)^{-1}\, | T\in S(\mathscr{C})',\ T\neq 0\}, \tag{3.3}$$

where $p^\circ(T) = \sup\{|T(a)|\ \ | p(a)\leqslant 1\}$.

Indeed, the set $\{a|\, p(a)\leqslant 1\}$ is closed and convex and its polar is the set $\{T\,|\, p^\circ(T)\leqslant 1\}$. The bipolar theorem implies that the set $\{a|\, p(a)\leqslant 1\}$ is the polar of the set $\{T\,|\, p^\circ(T)\leqslant 1\}$. Equality (3.3) follows from Ref. 34, Chap. I, no. 4, Lemma 2.

Since $F_p(\Phi) = p^\circ(\chi_\Phi)$, this and equality (3.3) imply that $\|a\|_{F_p}\leqslant p(a)$. Thus, the topology induced on $S(\mathscr{C})$ by the topology $\hat{\tau}$ is weaker than the topology τ. This equality also implies that (3.1) are really seminorms on $\mathscr{F}$ (that is, $\|\mathscr{F}\|_{F_p} < \infty$).

The topology $\hat{\tau}$ is Hausdorff; this easily follows from Lemma 4.4(a).

For the space $\mathscr{F}$ the following assertions are valid.

Theorem 3.1: Let T be a linear functional on $\mathscr{F}$ continuous with respect to the seminorm $\|\cdot\|_{F_p}$ of the form (3.1). Then

$$T(f) = \int d\mu(\Phi) f(\Phi),\quad f\in\mathscr{F},$$

where μ is a unique complex measure on $\mathscr{C}'_{\rm Re}$ such that every $f\in\mathscr{F}$ is integrable and

$$\int d\,|\mu|(\Phi)\,|f(\Phi)|\leqslant c\|f\|_{F_p},\quad f\in\mathscr{F}. \tag{3.4}$$

By the Hahn–Banach theorem, Theorem 3.1 has the following corollary:

Theorem 3.2: For a linear functional T defined on some subspace $M\subset S(\mathscr{C})$ having a representation

$$T(a) = \int d\mu(\Phi)\, a(\Phi),\quad a\in M\subset S(\mathscr{C})\subset\mathscr{F},$$

with a complex measure μ for which all $a(\Phi)\in\mathscr{F}$ are integrable, it would be sufficient that T be continuous in the topology on M induced by the topology $\hat{\tau}$. In addition, if $|T(a)|\leqslant \|a\|_{F_p}$ for $a\in M$ and a seminorm $\|\cdot\|_{F_p}$ of the form (3.1), then there exists a measure μ, representing T on M, such that for $f\in\mathscr{F}$

$$\left|\int d\mu(\Phi) f(\Phi)\right|\leqslant\int d\,|\mu|(\Phi)\,|f(\Phi)|\leqslant c\|f\|_{F_p}.$$

Remark. We note without a proof that assumptions of the theorem are also necessary.

Proof of Theorem 3.1: Let $\mathscr{F}_{\rm Re}$ be the set of real functions from $\mathscr{F}$. Since $\mathscr{F}_{\rm Re}$ is a Riesz space [with respect to the natural order $f\leqslant g \Leftrightarrow f(\Phi)\leqslant g(\Phi)$] (Ref. 25, Chap. II, §1, no. 1, Definition 1) and $|f|\leqslant|g|$ implies $\|f\|_{F_p}\leqslant\|g\|_{F_p}$, so a linear functional T defined on $\mathscr{F}$ and continuous with respect to the seminorm $\|\cdot\|_{F_p}$ can be written as $T = T_1 - T_2 + i(T_3 - T_4)$ with linear and positive functionals T_i, continuous with respect to $\|\cdot\|_{F_p}$ [Ref. 25, Chap. II, §2, no. 2, Theorem 1 and Formula (1)].

Let us consider, therefore, a positive functional T. In this case $T(\exp(i\langle\Phi,h\rangle))$ defines a positive definite function on $\mathscr{C}_{\rm Re}$. We have

$$\begin{aligned}&|T(\exp(i\langle\Phi,h_1\rangle)) - T(\exp(i\langle\Phi,h_2\rangle))|\\ &\quad= |T(\exp(i\langle\Phi,h_1\rangle) - \exp(i\langle\Phi,h_2\rangle))|\\ &\quad\leqslant c\,\|\exp(i\langle\Phi,h_1\rangle) - \exp(i\langle\Phi,h_2\rangle)\|_{F_p}\\ &\quad\leqslant c\|\langle\Phi,h_1-h_2\rangle\|_{F_p}\leqslant cp(h_1-h_2)\end{aligned}$$

for the τ-continuous seminorm p on $S(\mathscr{C})$. Since the restriction of the seminorm p on $\mathscr{C}$ is a continuous seminorm on $\mathscr{C}$, this estimate implies that $T(\exp(i\langle\Phi,h\rangle))$ defines a continuous positively definite function on $\mathscr{C}_{\rm Re}$. Since $\mathscr{C}_{\rm Re}$ is barreled (Ref. 34, Chap. IV, no. 1, Theorem 2) and nuclear,[21] by the Minlos theorem (Ref. 5, Chap. IX, §6, no. 12, Corollary) there exists a unique positive measure on $\mathscr{C}'_{\rm Re}$ such that

$$T(\exp(i\langle\Phi,h\rangle)) = \int d\mu(\Phi)\exp(i\langle\Phi,h\rangle).$$

Let us show that every $f\in\mathscr{F}$ is μ-integrable and

$$T(f) = \int d\mu(\Phi) f(\Phi). \tag{3.5}$$

To prove equality (3.5), we show that the linear combinations of exponents, $\exp(i\langle\Phi,h\rangle)$, $h\in\mathscr{C}_{\rm Re}$, are dense in $\mathscr{F}$ both in the topology defined by $\|\cdot\|_{F_p}$ and in the topology of $L_1(\mathscr{C}'_{\rm Re},\mu)$. It is sufficient to prove that they are dense in the set of bounded functions from $\mathscr{F}$. Let a bounded function $f\in\mathscr{F}$, then $f(\Phi) = f(a_1(\Phi),\dots,a_m(\Phi))$. We set $f_L(\Phi) = f(\Phi)\chi_L(\Sigma_{i=1}^m |a_i(\Phi)|)$, where $\chi_L(x) = \chi(x/L)$, $\chi(x)\in\mathscr{S}(\mathbb{R})$, $0\leqslant\chi(x)\leqslant 1$, $\chi(x) = 1$ for $x\in[0,1]$ and $=0$ for $x\notin[-1,2]$. We also introduce $f_{LM}(\Phi) = f_{LM}(a_1(\phi),\dots,a_m(\Phi))$, where $f_{LM}(x) = \int d^m y f_L(x-y)\sigma_M(y)$ and $\sigma_M(y) = M^m\sigma(y/M)$, $\sigma(y)\in\mathscr{S}(\mathbb{R}^m)$, $\sigma(y)\geqslant 0$, $\int d^m y\,\sigma(y) = 1$, $\operatorname{supp}\sigma(y)\subset[-1,1]$.

Since $S_n(\mathscr{C})$ is a Fréchet space and the completed n-fold symmetric tensor product of spaces $\mathscr{C}$ and the set $\{T\in S(\mathscr{C})'|\, p^\circ(T)\leqslant K<\infty\}$, where a seminorm p is τ-continuous, is equicontinuous, so every $a(\Phi)\equiv\chi_\Phi(a)$ is the point limit of a sequence of cylinder (and measurable) functions $a_N(\Phi)$ as $N\to\infty$ and the convergence $a_N(\Phi)\to a(\Phi)$ is uniform on any set $\mathrm{K}_{F_p,K} = \{\Phi\in\mathscr{C}'_{\rm Re}|F_p(\Phi)\equiv p^\circ(\chi_\Phi)\leqslant K\}$. Thus, for every $f_{LM}(\Phi)$ the sequence of cylinder functions $f_{LMN}(\Phi) = f_{LM}(a_{1N}(\Phi),\dots,a_{mN}(\Phi))$ converges as $N\to\infty$ to $f_{LM}(\Phi)$ uniformly on any $\mathrm{K}_{F_p,K}$. Denoting $f_{LMN}(\Phi) = g(\Phi) = g(\langle\Phi,b_1\rangle,\dots,\langle\Phi,b_k\rangle)$, $b_i\in\mathscr{C}_{\rm Re}$, and $g_{L'}(\Phi) = g_{L'}(\langle\Phi,b_1\rangle,\dots,\langle\Phi,b_k\rangle) = g(\Phi)\chi_{L'}(\Sigma_{i=1}^k|\langle\Phi,b_i\rangle|)$, we have that $g_{L'}(x)\in\mathscr{S}(\mathbb{R}^k)$ and its Fourier transform also belongs to $\mathscr{S}(\mathbb{R}^k)$. Hence $g_{L'}(x)$, and therefore $g_{L'}(\Phi)$, is the uniform limit of linear combinations of exponents.

Since

$$\|f - f_L\|_{F_p} = \|(1-\chi_L)f\|_{F_p} \leqslant L^{-1}\left\|\sum_{i=1}^m |a_i(\phi)|f\right\|_{F_p}$$

and

$$\|g - g_{L'}\|_{F_p}\leqslant L'^{-1}\left\|\sum_{i=1}^k|\langle\phi,b_i\rangle|\,g\right\|_{F_p},$$

one can easily see that the linear hull of exponents is dense in $\mathscr{F}$ in the topology defined by $\|\cdot\|_{F_p}$. Its density in the topology of $L_1(\mathscr{C}'_{\rm Re}, \mu)$ follows from the Lebesgue dominated and montone convergence theorems.

This proves the existence of a measure. Uniqueness of the measure follows, for example, from Ref. 25, Chap. IX, §6, no. 12, Corollary, and from simple arguments. Estimate (3.4) follows from the existence and properties of the absolute value of a complex (Radon) measure [cf. Ref. 25, Chap. III, §1, no. 6; see also Chap. II, §2, no. 2, Theorem 1 and formula (1)]. Theorem 3.1 is proved.

4. THE TOPOLOGY $\hat{\tau}$ ON THE SUBSPACE OF POLYNOMIAL FUNCTIONS

In this section we describe the topology which is weaker than that induced by $\hat{\tau}$ on the subspace $S(\mathscr{C})$, and this topology is used in the following (in Sec. 5) to solve a Euclidean-invariant problem of moments.

Let us denote by $p_1 \otimes_\epsilon p_2$ the (bi-) equicontinuous tensor product of seminorms p_1, p_2.[34,35]

Theorem 4.1: Let

$$p'(a) = c_0|a_0| + c_1 p(a_1) + \sum_{n=2}^{\infty} c_n (p \otimes_\epsilon \cdots \otimes_\epsilon p)(a_n), \quad a \in S(\mathscr{C}), \quad (4.1)$$

where p is a continuous Euclidean invariant seminorm on $\mathscr{C}$ (that is, invariant under transformations of $\mathscr{C}$ generated by translations and Euclidean rotations of $\mathbb{R}^d$) and $p(f) = p(f^*)$. Then

$$p'(a) \leqslant \sup_{\Phi \in \mathscr{C}'_{\rm Re}} |a(\Phi)| F_{p''}(\Phi)^{-1}, \quad a \in S(\mathscr{C}),$$

for a Euclidean invariant function $F_{p''}(\Phi)$ of the form (3.2) given by a Euclidean-invariant τ-continuous seminorm p'' on $S(\mathscr{C})$,

$$p''(a) = \sup_n \max(c_0'|a_0|, c_1' p(a_1), c_n' p \otimes_\epsilon \cdots \otimes_\epsilon p(a_n)). \quad (4.2)$$

Remark: Is the assertion of Theorem 4.1 valid for other Euclidean-invariant seminorms? We note that in Ref. 18 we have considered an analogous topology [with the projective tensor product of seminorms and with a Euclidean-noninvariant function $F_{p''}(\Phi)$].

To prove Theorem 4.1, we first formulate a number of lemmas. First we consider an analogous problem for the symmetric tensor algebra over a one-dimensional space. Such an algebra is the same as the algebra $\mathbb{P}(\mathbb{R})$ of complex polynomials of one real variable. The analog of the topology τ is defined by the seminorms

$$\|P\|_{\{c_n\}} = \sum_{n=0}^{N} c_n |\alpha_n|,$$

where $P(x) = \Sigma_{n=0}^N \alpha_n x^n$ and $\{c_n\}$ is an arbitrary sequence with $0 \leqslant c_n < \infty$. The analog of the topology $\hat{\tau}$ is given by the seminorms

$$\|P\|_F = \sup_{x \in \mathbb{R}} |P(x)| F(x)^{-1},$$

where $F(x) = \Sigma_{n=0}^\infty d_n |x|^n$ with some constants $0 < d_n \leqslant \infty$. Since $F(x)$ is a function which grows faster than any polynomial, $\|\mathbb{P}(\mathbb{R})\|_F < \infty$.

We shall show that both sets of seminorms define the same topology on $\mathbb{P}(\mathbb{R})$.

Lemma 4.2[33]: Both sets of seminorms, $\|\cdot\|_{\{c_n\}}$ and $\|\cdot\|_F$, define the same topology on $\mathbb{P}(\mathbb{R})$.

The proof of Lemma 4.2 is based on Lemma 4.3.

A sequence $\{\mu_n\}$ is called positive definite if

$$\sum_{i,j=0}^{n} \mu_{i+j} \alpha_i \alpha_j^* \geqslant 0$$

for each finite number of complex $\alpha_0,\ldots,\alpha_n$ (here α_j^* denotes the complex conjugate of α_j).

Lemma 4.3: Let a sequence $\{c_n\}$, $0 \leqslant c_n < \infty$, be given. Then there exists a sequence $\{c_n'\}$ such that, for every sequence $\{b_n\}$, $|b_n| \leqslant c_n$, there exists a decomposition $b_n = \lambda_n - \mu_n + i(\nu_n - \xi_n)$, where sequences $\{\lambda_n\}$, $\{\mu_n\}$, $\{\nu_n\}$, $\{\xi_n\}$ are positive definite and $|\lambda_n| + |\mu_n| + |\nu_n| + |\zeta_n| \leqslant c_n'$.

Proof: The proof of the lemma is in fact a slightly stronger version of the Boas' proof[31] (cf. also Ref. 33).

We use the following notation:

$$\det(\lambda)_{2n} = \begin{vmatrix} \lambda_0, \lambda_1, \ldots, \lambda_n \\ \lambda_1, \lambda_2, \ldots, \lambda_{n+1} \\ \cdots\cdots\cdots\cdots \\ \lambda_n, \lambda_{n+1}, \cdots, \lambda_{2n} \end{vmatrix}.$$

It is sufficient to consider real sequences $\{b_n\}$.

For positive definiteness of a sequence $\{\lambda_n\}$ it is sufficient that all consequent principal minors $\det(\lambda)_{2n}$ of the matrix λ_{i+j} are strictly positive (Ref. 36, Chap. X, §4, Theorem 3).

Let a sequence $\{b_n\}$, $|b_n| \leqslant c_n$, be given. We define the sequences $\{\lambda_n\}$, $\{\mu_n\}$, $\{c_n'\}$ by induction. Let $\lambda_0 = b_0 + c_0 + 1$, $\mu_0 = c_0 + 1$, $c_0' = 3c_0 + 2$. Then $\lambda_0 - \mu_0 = b_0$ and $|\lambda_0| + |\mu_0| \leqslant c_0'$, and c_0' only depends on c_0.

We suppose that $\lambda_k - \mu_k = b_k$ for $k \leqslant 2n-2$, that determinants $\det(\lambda)_{2k} \geqslant 1$, $\det(\mu)_{2k} \geqslant 1$ for $k \leqslant n-1$, and that $|\lambda_k| + |\mu_k| \leqslant c_k'$, $k \leqslant 2n-2$, where the sequence $(c_0', \cdots, c_{2n-2}')$ depends only on $(c_0,\ldots,c_{2n-2})$.

Let $\lambda_{2n-1} = b_{2n-1}$, $\mu_{2n-1} = 0$, and $c_{2n-1}' = c_{2n-1}$. We write (with so far undetermined λ_{2n})

$$\det(\lambda)_{2n} = \lambda_{2n} \det(\lambda)_{2n-2} + P(\lambda),$$

where $P(\lambda)$ is a polynomial in $\lambda_0,\ldots,\lambda_{2n-1}$, and the corresponding relation for $\det(\mu)_{2n}$,

$$\det(\mu)_{2n} = \mu_{2n} \det(\mu)_{2n-2} + P(\mu).$$

We denote $P = \sup\{|P(\lambda)| \mid |\lambda_k| \leqslant c_k', k \leqslant 2n-1\}$ and choose $\lambda_{2n} = b_{2n} + c_{2n} + 1 + P$, $\mu_{2n} = c_{2n} + 1 + P$, $c_{2n}' = 3c_{2n} + 2 + 2P$; then we obtain that $\lambda_{2n} - \mu_{2n} = b_{2n}$, $\det(\lambda)_{2n} \geqslant 1$, $\det(\mu)_{2n} \geqslant 1$, and $|\lambda_{2n}| + |\mu_{2n}| \leqslant c_{2n}'$, where c_{2n}' depends only on $(c_0,\ldots,c_{2n})$. This completes the induction and the proof of Lemma 4.3

Proof of Lemma 4.2 (cf. Ref. 33): Let a sequence $\{c_n\}$, $0 \leqslant c_n < \infty$, be given. Lemma 4.3 implies that there exists a sequence $\{c_n'\}$ such that each $T \in \mathbb{P}(\mathbb{R})'$ with $|T(P)| \leqslant \|P\|_{\{c_n\}}$ can be represented in the form $T = T_1 - T_2 + i(T_3 - T_4)$ with positive functionals T_i (i.e., they are positive on positive polynomials) such that $\Sigma_{i=1}^4 |T_i(P)| \leqslant \|P\|_{\{c_n'\}}$. Now if T is such that $|T(P)| \leqslant \|P\|_{\{c_n\}}$, then the above arguments and the theorem (Ref. 37, XII.8.1) about the Hamburger moment problem imply that, for some measure μ on $\mathbb{R}$, such that $\int d|\mu|(x)\,|x^n|\leqslant c_n'$, $T(P)=\int d\mu(x)\,P(x)$. With $d_n = 2^{-n-1}(c_n'+1)^{-1}$ and $F(x)=\Sigma_{n=0}^{\infty} d_n|x|^n$, we have

$$|T(P)|\leqslant\int d\,|\mu|\;|P(x)|\leqslant\|P\|_F\int d\,|\mu|\,F(x)\leqslant\|P\|_F$$

[the integrability of $F(x)$ follows from the Lebesgue monotone convergence theorem].

This inequality means that $\|P\|_{\{c_n\}} = \sup\{|T(P)|\;|T(P)|\leqslant\|P\|_{\{c_n\}}\leqslant\|P\|_F$. The converse inequality, $\|\cdot\|_F\leqslant\|\cdot\|_{\{c_n\}}$, for some sequence $\{c_n\}$ is obvious. Lemma 4.2 is proved.

Now we consider the algebra $S(\mathscr{C})$.

Lemma 4.4: Seminorms of the form

$$\|f\|_F=\sup_{\Phi\in\mathscr{C}'_{\mathrm{Re}}}|f(\Phi)|F(\Phi)^{-1},\quad f\in\mathscr{F},$$

with the functions F of one of the following two types:

(a) $$F_p(\Phi)=\sum_{n=0}^{\infty}p^\circ_n(\Phi)^n=\sup_{p(a)\leqslant 1}|a(\Phi)|,$$

$$p(a)=\sup_n\max(|a_0|,p_n\otimes_\epsilon\cdots\otimes_\epsilon p_n(a_n))$$

and $p^\circ_0(\Phi)^0=1$,

(b) $$F(\Phi)=\prod_{n=1}^{\infty}[1+p^\circ_n(\Phi)],$$

form a basis of the topology $\hat{\tau}$. Here $\{p_n\}$ is any sequence of continuous seminorms on $\mathscr{C}$ and $p^\circ_n(\Phi) = \sup\{|\langle\Phi,f\rangle|\;|p_n(f)\leqslant 1,f\in\mathscr{C}\}$ is the dual Minkowski functional of the seminorm p.

Proof: One can easily see that the seminorms of the form $p_n'(a_n)=p_n\otimes_\epsilon\cdots\otimes_\epsilon p_n(a_n)$, where p_n are continuous seminorms on $\mathscr{C}$, define the topology of $\mathscr{C}(\mathbb{R}^{dn})$ and the seminorms of the $p(a)=\sup_n p_n'(a_n)$, where $p_0'(a_0)=|a_0|$, $p_1'(a_1)=p_1(a_1)$, define the topology τ. Thus, the seminorms of the form (3.1) with functions F, corresponding to the seminorms p, define the topology $\hat{\tau}$. Calculating a value of the dual Minkowski functional $p^\circ(\cdot)$ of the seminorm p for $\chi_\Phi = (1,\Phi,\ldots,\Phi\otimes\cdots\otimes\Phi,\cdots)\in S(\mathscr{C})'$, $\Phi\in\mathscr{C}'_{\mathrm{Re}}$, we obtain

$$F(\Phi)=p^\circ(\chi_\Phi)=\sum_{n=0}^{\infty}p^\circ_0(\Phi)^n.$$

Indeed,

$$\sup_{p(a)\leqslant 1}\left|a_0+\sum_{n=0}^{\infty}\langle\Phi\otimes\cdots\otimes\Phi,a_n\rangle\right|$$

$$\leqslant\sup_{p(a)\leqslant 1}|a_0|+\sum_{n=1}^{\infty}\sup_{p(a)\leqslant 1}|\langle\Phi\otimes\cdots\otimes\Phi,a_n\rangle|,$$

$$\sup_{p(a)\leqslant 1}|\langle\Phi\otimes\cdots\otimes\Phi,a_n\rangle|$$

$$=\sup_{p_n\otimes_\epsilon\cdots\otimes_\epsilon p_n(a_n)\leqslant 1}|\langle\Phi\otimes\cdots\otimes\Phi,a_n\rangle|$$

$$\leqslant p^\circ_0(\Phi)^n.$$

The last inequality follows from the definition of the ϵ-tensor product,

$$p_n\otimes_\epsilon\cdots\otimes_\epsilon p_n(a_n)=\sup\{|\langle\Phi_1\otimes\cdots\otimes\Phi_n,a_n\rangle|\;|\;\Phi_1,\ldots,\Phi_n\in\mathscr{C}'_{\mathrm{Re}},p^\circ_n(\Phi_i)\leqslant 1\}.$$

So,

$$F(\Phi)\leqslant\sum_{n=0}^{\infty}p^\circ_n(\Phi)^n.$$

On the other hand, if all $p^\circ_n(\Phi)<\infty$, then we choose $a=(a_0,a_1,\ldots,a_n,\cdots)$, $a_0=1$, $a_n=b_n\otimes\cdots\otimes b_n$ for $1\leqslant n\leqslant N$ and $a_n=0$ for $n>N$ and choose b_n such that $p_n\,(b_n)\leqslant 1$,

$$p^\circ_n(\Phi)^n-2^{-n}\epsilon\leqslant\langle\Phi\otimes\cdots\otimes\Phi,a_n\rangle$$
$$=\langle\Phi,b_n\rangle^n\leqslant p^\circ_n(\Phi)^n.$$

Note that $p(a)=\sup_n\max(1,p_n\,(b_n)^n)\leqslant 1$. Then, we obtain

$$F(\Phi)\geqslant\left|1+\sum_{n=1}^{\infty}\langle\Phi\otimes\cdots\otimes\Phi,a_n\rangle\right|$$
$$\geqslant-\epsilon+\sum_{n=0}^{\infty}p^\circ_n\,(\Phi)^n$$

for each $\epsilon>0$ and natural N. This implies that $F(\Phi)\geqslant\Sigma_{n=0}^{\infty}p^\circ_n(\Phi)^n$, that is, $F(\Phi)=\Sigma_{n=0}^{\infty}p^\circ_n(\Phi)^n$.

If $p^\circ_n(\Phi)=\infty$ for some n, then analogously we obtain that $F(\Phi)=\infty$, too.

Thus, case (a) of Lemma 4.4 is proved.

Functions of the type (b) are not in general functions of the form (3.2).

We assume without loss of generality that $p_n\leqslant p_{n+1}$, which implies $p^\circ_n(\Phi)\geqslant p^\circ_{n+1}(\Phi)$.

It is obvious that seminorms with functions from (a) dominate seminorms with functions from (b).

On the other hand,

$$\sum_{n=0}^{\infty}p^\circ_n(\Phi)^n\geqslant\sum_{m=1}^{\infty}2^{-m}\sum_{n=0}^{2^m}p^\circ_n(\Phi)^n$$
$$\geqslant\sum_{m=1}^{\infty}2^{-m}\sum_{n=0}^{2^m}p^\circ_{2^m}(\Phi)^n$$
$$\geqslant\sum_{m=1}^{\infty}2^{-m}[1+2^{-m}p^\circ_{2^m}(\Phi)]^{2^m}.\quad(4.3)$$

We have used the inequality for binomial coefficients $c_N^n = N!/n!(N-n)!\leqslant N^n$ for $N=2^m$.

Using the inequality

$$\ln\left(\sum_{m=1}^{\infty}2^{-m}x_m\right)\geqslant\sum_{m=1}^{\infty}2^{-m}\ln x_m,\quad x_m\geqslant 1,$$

which simply follows from the convexity and continuity of the logarithm, we obtain that (4.3) is larger than

$$\prod_{m=1}^{\infty}[1+2^{-m}p^\circ_{2^m}(\Phi)]=\prod_{n=1}^{\infty}[1+q^\circ_m(\Phi)]$$

with $q_m=2^m p_{2^m}$.

This proves that seminorms with functions from (b) dominate seminorms with functions from (a). Lemma 4.4 is proved.

Lemma 4.5: Let $F(\Phi)=\sum_{n=0}^{\infty}p^\circ_n(\Phi)^n$, where p_n is a sequence of continuous seminorms on $\mathscr{C}$ and $p^\circ_0(\Phi)^0=1$. Let $a=(a_0,\ldots,a_n,\cdots)\in S(\mathscr{C})$, then for any sequence $\{c_n\}$,

$0 \le c_n < \infty,$

$$\sum_{n=0}^{\infty} c_n \|a_n\|_F \le \|a\|_{F'},$$

where $F'(\Phi) = \Sigma_{n=0}^{\infty} q^\circ_n(\Phi)^n$, $q_n = c'_n p_n$ for some sequence $\{c'_n\}$, $0 < c'_n < \infty$.

Proof: Lemma 4.2 implies that for any sequence $\{c_n\}$ there exists a function $H(x) = \Sigma_{n=0}^{\infty} d_n |x|^n$ with $0 < d_n < \infty$, such that

$$\sum_{n=0}^{\infty} 2^{n+1} c_n |\alpha_n| \le \sup_{x \in \mathbf{R}} \left| \sum_{n=0}^{\infty} \alpha_n x^n \right| H(x)^{-1}$$

for all terminating sequences $\{\alpha_n\}$. Hence,

$$\begin{aligned}
\sum_{n=0}^{\infty} c_n \|a_n\|_F &\le \sup_n 2^{n+1} c_n \|a_n\|_F \\
&= \sup_{\Phi} \sup_n 2^{n+1} c_n |\langle \Phi \otimes \cdots \otimes \Phi, a_n \rangle| F(\Phi)^{-1} \\
&\le \sup_{\Phi} \sum_{n=0}^{\infty} 2^{n+1} c_n |\langle \Phi \otimes \cdots \otimes \Phi, a_n \rangle| F(\Phi)^{-1} \\
&\le \sup_{\Phi} \sup_{x} \left| \sum_{n=0}^{\infty} \frac{\langle \Phi \otimes \cdots \otimes \Phi, a_n \rangle x^n}{F(\Phi) H(x)} \right| \\
&\le \sup_{\Phi} \sup_{x} \left| \sum_{n=0}^{\infty} \frac{\langle \Phi \otimes \cdots \otimes \Phi, a_n \rangle x^n}{F'(x\Phi)} \right| \\
&= \sup_{\Phi} \left| \sum_{n=0}^{\infty} \frac{\langle \Phi \otimes \cdots \otimes \Phi, a_n \rangle}{F'(\Phi)} \right|,
\end{aligned}$$

where $F'(\Phi) = \Sigma_{n=0}^{\infty} d_n p^\circ_n(\Phi)^n$ and we have used the relations $p^\circ_n(x\Phi) = |x| p^\circ_n(\Phi)$ and $F(\Phi) H(x) \ge \Sigma_{n=0}^{\infty} d_n p^\circ_n(x\Phi)^n = F'(x\Phi)$.

Hence, it follows that the assertion of the lemma is fulfilled for $F'(\Phi) = \Sigma_{n=0}^{\infty} q^\circ_n(\Phi)^n$, $q_n = d_n^{-1/n} p_n$. Lemma 4.5 is proved.

Lemma 4.6: Let $\{p_k\}$ be a sequence of continuous seminorms on $\mathscr{C}$. Then

$$\sup_{\Phi \in \mathscr{C}'_{\mathrm{Re}}} \frac{|\langle \Phi \otimes \cdots \otimes \Phi, a_n \rangle|}{p^\circ_1(\Phi) \cdots p^\circ_n(\Phi)} \le \|a_n\|_H, \quad a_n \in \mathbf{S}_n(\mathscr{C}),$$

for a seminorm $\|f\|_H = \sup_{\Phi} |f(\Phi)| H(\Phi)^{-1}$, $f \in \mathscr{F}$, with $H(\Phi) = \Pi_{k=1}^{\infty} [1 + q^\circ_k(\Phi)]$ with $q_k = 2^{k+2} p_k$.

Proof: Again without restriction, we assume that $p^\circ_n \ge p^\circ_{n+1}$ and define $H(\Phi) = \Pi_{k=1}^{\infty} [1 + q^\circ_k(\Phi)]$, where $q_k = 2^{K+2} p_k$, that is, $q^\circ_k(\Phi) = 2^{-k-2} p^\circ_k(\Phi)$. Then, denoting $a_n(\Phi) = \langle \Phi \otimes \cdots \otimes \Phi, a_n \rangle$, we have

$$\begin{aligned}
&\sup_{\Phi} |a_n(\Phi)| \prod_{k=1}^{\infty} [1 + q^\circ_k(\Phi)]^{-1} \\
&\ge \sup_{\Phi} \left| a_n\left(\frac{\Phi}{q^\circ_n(\Phi)}\right) \right| \prod_{k=1}^{\infty} \left[1 + q^\circ_k \left(\frac{\Phi}{q^\circ_n(\Phi)} \right) \right] \\
&= \sup_{\Phi} |a_n(\Phi)| q^\circ_k(\Phi)^{-n} \\
&\quad \times \prod_{k=1}^{\infty} [1 + q^\circ_k(\Phi) q^\circ_n(\Phi)^{-1}]^{-1} \\
&= \sup_{\Phi} |a_n(\Phi)| \prod_{k=1}^{\infty} [q^\circ_k(\Phi) + q^\circ_n(\Phi)]^{-1} \\
&\quad \times \prod_{k=n+1}^{\infty} [1 + q^\circ_k(\Phi) q^\circ_n(\Phi)^{-1}]^{-1} \\
&\ge \sup_{\Phi} |a_n(\Phi)| \prod_{k=1}^{n} [2q^\circ_k(\Phi)]^{-1} \\
&\quad \times \prod_{k=n+1}^{\infty} (1 + 2^{-(k-n)})^{-1} \\
&= \sup_{\Phi} |a_n(\Phi)| (p^\circ_1(\Phi) \cdots p^\circ_n(\Phi))^{-1} \\
&\quad \times 2^{n(n+3)/2} \prod_{k=1}^{\infty} (1 + 2^{-k})^{-1} \\
&\ge \sup_{\Phi} |a_n(\Phi)| (p^\circ_1(\Phi) \cdots p^\circ_n(\Phi))^{-1},
\end{aligned}$$

since $2^{n(n+3)/2} \Pi_{k=1}^{\infty} (1 + 2^{-k})^{-1} \ge 2^{n(n+3)/2} \exp(-\Sigma_{k=1}^{\infty} 2^{-k}) \ge 1$ for $n \ge 1$. Lemma 4.6 is proved.

Proof of Theorem 4.1: To prove Theorem 4.1, it is sufficient to prove that

$$(p \otimes_\epsilon \cdots \otimes_\epsilon p)(a_n) \le c'_n \sup_{\Phi \in \mathscr{C}'_{\mathrm{Re}}} \frac{|\langle \Phi \otimes \cdots \otimes \Phi, a_n \rangle|}{p^\circ(\Phi)^n}, \quad a \in \mathbf{S}(\mathscr{C}), \tag{4.4}$$

where $p^\circ(\Phi)$ is the dual Minkowski functional of the seminorm p. Indeed, if (4.4) is valid, then

$$\begin{aligned}
c_0|a_0| + c_1 p(a_1) + \sum_{n=2}^{\infty} c_n (p \otimes_\epsilon \cdots \otimes_\epsilon p)(a_n) &\le c''_0 |a_0| + \sum_{n=0}^{\infty} c''_n \sup_{\Phi} \frac{|\langle \Phi \otimes \cdots \otimes \Phi, a_n \rangle|}{p^\circ(\Phi)^n} \quad \text{[inequality (4.4)]} \\
&\le \sum_{n=0}^{\infty} c''_n \|a_n\|_H \left(H(\Phi) = \prod_{n=1}^{\infty} [1 + 2^{-n-2} p^\circ(\Phi)], \quad \text{Lemma 4.6} \right) \\
&\le \sum_{n=0}^{\infty} c''_n \|a_n\|_F \left(F(\Phi) = \sum_{n=0}^{\infty} 2^{-n^2-2n} p^\circ(\Phi)^n \right) \le \|a\|_{F'} \left(F'(\Phi) = \sum_{n=0}^{\infty} c'''_0 p^\circ(\Phi)^n, \quad \text{Lemma 4.5} \right).
\end{aligned}$$

Lemma 4.4(a) now implies that this is an estimate of Theorem 4.1 with Euclidean-invariant functions $F'(\Phi)$ defined by the seminorm of the form (4.2).

The proof of inequality (4.4) is the following. For every $\delta>0$, $a_n \in S_n(\mathscr{C})$

$$(p \otimes_\epsilon \cdots \otimes_\epsilon p)(a_n)$$

$$= \sup_{\Phi_i,\, p^\circ(\Phi_i)\leqslant 1} |\langle \Phi_1 \otimes \cdots \otimes \Phi_n, a_n\rangle|$$

$$\leqslant |\langle \bar{\Phi}_1^\delta \otimes \cdots \otimes \Phi_n^\delta, a_n\rangle| + \delta$$

$$\leqslant 2^n |\langle \Phi_1^\delta \otimes \cdots \otimes \Phi_n^\delta, a_n\rangle| + \delta$$

for some $\bar{\Phi}_1^\delta,\ldots,\bar{\Phi}_n^\delta \in \mathscr{C}'$, $p^\circ(\bar{\Phi}_1^\delta)\leqslant 1$, and $\bar{\Phi}_1^\delta,\ldots,\bar{\Phi}_n^\delta \in \mathscr{C}'_{\rm Re}$, $p^\circ(\Phi_i^\delta)\leqslant 1$.

The general polarization identity (Ref. 32, Lemma 1.5.4) implies that

$$\sum_{\pi\in P_n} \Phi^\delta_{\pi(1)} \otimes \cdots \otimes \Phi^\delta_{\pi(n)}$$

is a linear combination of terms of the form $\Phi_\vartheta^\delta \otimes \cdots \otimes \Phi_\vartheta^\delta$, where P_n is the set of all permutations of n elements and $\Phi_\vartheta^\delta = \Sigma_{i=1}^n \vartheta_i \Phi_i^\delta$ with ϑ_i equal to 0 or 1. Using the general polarization identity and taking into account that a_n is symmetric, we obtain

$$|\langle \Phi_1^\delta \otimes \cdots \otimes \Phi_n^\delta, a_n\rangle|$$

$$= n!^{-1}\Big|\sum_{\pi\in P_n}\langle \Phi^\delta_{\pi(1)} \otimes \cdots \otimes \Phi^\delta_{\pi(n)}, a_n\rangle\Big|$$

$$\leqslant 2^n n!^{-1} \max_{\vartheta_1\cdots\vartheta_n=0,1} |\langle \Phi_\vartheta^\delta \otimes \cdots \otimes \Phi_\vartheta^\delta, a_n\rangle|$$

$$= 2^n n!^{-1} |\langle \Phi_{\vartheta'}^\delta \otimes \cdots \otimes \Phi_{\vartheta'}^\delta, a_n\rangle|$$

for some $\vartheta' = (\vartheta'_1,\ldots,\vartheta'_n)$. Then we have $p^\circ(\Phi^\delta_{\vartheta'})\leqslant n \max_j p^\circ(\Phi_j^\delta)\leqslant n$ and

$$|\langle \Phi_1^\delta \otimes \cdots \otimes \Phi_n^\delta, a_n\rangle|$$
$$\leqslant 2^n n^n n!^{-1} |\langle \Phi_{\vartheta'}^\delta \otimes \cdots \otimes \Phi_{\vartheta'}^\delta, a_n\rangle| / p^\circ(\Phi^\delta_{\vartheta'})^n.$$

Thus,

$$(p \otimes_\epsilon \cdots \otimes_\epsilon p)(a_n)$$

$$\leqslant 4^n n^n n!^{-1} \sup_{\Phi\in\mathscr{C}'_{\rm Re}} [|\langle \Phi \otimes \cdots \otimes \Phi, a_n\rangle| / p^\circ(\Phi)^n + \delta].$$

Since δ is arbitrarily small, this implies inequality (4.4). Theorem 4.1 is proved.

5. EXISTENCE OF A EUCLIDEAN REALIZATION FOR THE QUANTUM FIELD :exp $\varphi(x)$:$_d$

Definition: By a Euclidean realization (= a Euclidean field) for a d-dimensional quantum Wightman–Jaffe field we mean a 2-tuple $(\mathscr{C}^{\alpha\prime}_{\rm Re}, \mu)$, where $\mathscr{C}^{\alpha\prime}_{\rm Re}$ is some space of real ultradistributions of a Jaffe type with $\alpha > 1$ and μ is a Euclidean-invariant complex measure on $\mathscr{C}^{\alpha\prime}_{\rm Re}$ having all moments and such that its moments at noncoinciding points are equal to the Schwinger functions S_n of the considered quantum field, that is, for $f\in\mathscr{P}_0$ and having the form (2.3).

$$\int d\mu(\Phi)\Big(f_0 + \langle \Phi, f_1\rangle + \sum_{n=2}^\infty \langle \Phi\otimes\cdots\otimes\Phi, f_n\rangle\Big)$$
$$= S_0 f_0 + \int d^d x\, S_1(x) f_1(x)$$
$$+ \sum_{n=2}^\infty \int d^{dn}{}_x\, S_n(x_1,\ldots,x_n) f_n(x_1,\ldots,x_n).$$

In this section we prove Theorem 2.1, that is, we prove the existence of a Euclidean realization on a space $\mathscr{C}^{\alpha\prime}_{\rm Re}$ with $1<\alpha<(d-1)(d-2)^{-1}$ for the quantum Wightman–Jaffe field :exp $\varphi(x)$:$_d$, where $\varphi(x)$ is the free massive Hermitian scalar quantum field and double colons denote normal ordering.

The proof of the existence of a Euclidean realization is based on the following theorem.

Theorem 5.1: Let $T=(T_0,T_1,\cdots)$, $T_0\in\mathbb{C}$,| $T_1\in\mathscr{C}'$, $T_n \in \mathscr{C}_0(\mathbb{R}^{dn})'$ for $n\geqslant 2$, be a sequence of symmetric Euclidean-invariant ultradistributions, defined on $\mathscr{C}_0(\mathbb{R}^{dn})$ and

$$|\langle T_n, f_n\rangle| \leqslant c_n (p\otimes_\epsilon\cdots\otimes_\epsilon p)(f_n),$$
$$f_n \in \mathscr{C}_0(\mathbb{R}^{dn})\cap S_n(\mathscr{C}), \tag{5.1}$$

for some Euclidean-invariant continuous seminorm p on $\mathscr{C}$ such that $p(f) = p(f^*)$. Then there exists a Euclidean-invariant complex measure on $\mathscr{C}'_{\rm Re}$ such that every $f\in\mathscr{F}$ is integrable and for $f\in\mathscr{P}_0$ and having the form (2.3):

$$\int d\mu(\Phi)\Big(f_0 + \langle \Phi, f_1\rangle + \sum_{n=2}^\infty \langle \Phi\otimes\cdots\otimes\Phi, f_n\rangle\Big)$$
$$= T_0 f_0 + \sum_{n=1}^\infty \langle T_n, f_n\rangle. \tag{5.2}$$

Proof of Theorem 5.1: Theorem 4.1 and 3.2 imply the existence of a measure μ, such that every $f\in\mathscr{F}$ is integrable and satisfying inequalities (5.2) and an estimate

$$\Big|\int d\mu(\Phi) f(\Phi)\Big| \leqslant \int d\,|\mu|(\Phi)\,|f(\Phi)| \leqslant c\|f\|_{F_{p'}} \tag{5.3}$$

with a Euclidean-invariant function $F_{p'}(\Phi)$, defined by a Euclidean-invariant τ-continuous seminorm p',

$$p'(a) = \sup_{n\geqslant 2}\max(c'_0|a_0|,\ c'_1\, p(a_1),\ c'_n(p\otimes_\epsilon\cdots\otimes_\epsilon p)(a_n)),$$
$$a_n \in S_n(\mathscr{C}).$$

Let $\mu = \mu_1 - \mu_2 + i(\mu_3 - \mu_4)$, be the Hahn–Jordan decomposition of a measure μ, $\mu_1 = ({\rm Re}\,\mu)^+$, $\mu_2 = ({\rm Re}\,\mu)^-$, $\mu_3 = ({\rm Im}\,\mu)^+$, $\mu_4 = ({\rm Im}\,\mu)^-$ [Ref. 25, Chap. II, §2, no. 2, Theorem 1 and formula (1); see also Chap. III, §1, no. 6, Theorem 3]. Positive measures μ_j satisfy estimate (5.3).

Let $IO(d)$ be the inhomogeneous Euclidean group, i.e., the group of translations and rotations of the Euclidean space $\mathbb{R}^d$. One defines naturally the representation of the group $IO(d)$ into the group of automorphisms of $\mathscr{C}'_{\rm Re}$. For $(a,R)\in IO(d)$, $a\in\mathbb{R}^d$, $R\in O(d)$, we define the map $\Phi\mapsto\Phi_{(a,R)}$ as follows:

$$\langle \Phi_{(a,R)}, h\rangle = \langle \Phi, h_{(a,R)}\rangle,$$
$$\Phi\in\mathscr{C}'_{\rm Re},\ h\in\mathscr{C}(\mathbb{R}^d),$$
$$h_{(a,R)} = h_{(a,R)}(x) = h(R^{-1}(x-a)).$$

Let CB($IO(d)$) be the space of bounded continuous functions on $IO(d)$. We assert that for $g \in \mathscr{F}$

$$\int d\mu_j(\Phi)\, g(\Phi_{(a,R)}) \in \mathrm{CB}(IO(d)).$$

Estimate (5.3) with a Euclidean-invariant function $F_{p'}(\Phi)$ implies that the function $\int d\mu_j(\Phi)\, g(\Phi_{(a,R)})$ is bounded.

To prove the continuity of this function, we use the fact, as in the proof of Theorem 3.1, that the linear hull of the exponents is dense in $\mathscr{F}$ with respect to the seminorm $\|\cdot\|_{F_{p'}}$. For every $\epsilon > 0$

$$\begin{aligned}
\int d\mu_j(\Phi)\, | g(\Phi_{(a,R)}) - g(\Phi_{(a',R')})|
&\leqslant \epsilon + \sum_{j\in J} |\alpha_j|\, \|\exp(i\langle \Phi_{(a,R)}, h_j\rangle) \\
&\quad - \exp(i\langle \Phi_{(a',R')}, h_j\rangle)\|_{F_{p'}} \\
&\leqslant \epsilon + \sum_{j\in J} |\alpha_j|\, \|\langle \Phi_{(a,R)} - \Phi_{(a',R')}, h_j\rangle\|_{F_{p'}} \\
&\leqslant \epsilon + \sum_{j\in J} |\alpha_j|\, p'(h_{j,(a,R)} - h_{j,(a',R')}) \qquad (5.4)
\end{aligned}$$

for some J, α_j, h_j (depending on ϵ). Since the restriction of the seminorm p' on $\mathscr{C}$ is the continuous seminorm p, so, using Euclidean invariance of the seminorm p, we have

$$\begin{aligned}
p'(h_{j,(a,R)} - h_{j,(a',R')}) &= p(h_{j,(a,R)} - h_{j(a',R')}) \\
&= p(h_j - h_{j,(b,Q)}), \qquad (5.5)
\end{aligned}$$

where $b = R^{-1}a' - R^{-1}a$, $Q = R^{-1}R'$.

The seminorm p is dominated by a seminorm q from (2.1). We denote the Fourier transform of h by $h^{\sim}$, the components of the matrix Q^{-1} by Q_{ij}^{-1}, and the Kronecker symbol by δ_{ij}. So we write

$$\begin{aligned}
h_j^{\sim} - h^{\sim}_{j(b,Q)} &= [h_j^{\sim} - h^{\sim}_{j(0,Q)}] \\
&\quad + [1 - \exp(ibp)]\, h^{\sim}_{j(0,Q)}, \\
h_j^{\sim} - h^{\sim}_{j(0,Q)} &= \int_0^1 ds \sum_{i,j} (\delta_{ij} - Q_{ij}^{-1})\, p_j\, \partial_i h^{\sim}(p_s), \\
p_s &= Q^{-1}p + s(p - Q^{-1}p), \\
1 - \exp(ibp) &= -ibp \int_0^1 ds\, \exp[i(1-s)bp],
\end{aligned}$$

and we have that (5.5) is bounded by

$$\sum_{i,j} |\delta_{ij} - Q_{ij}^{-1}| \sup_s q(p_j\, \partial_i h^{\sim}(p_s)) + \sum_j |b_j| \sup_s q(p_j \exp[i(1-s)bp]\, h^{\sim}(Q^{-1}p)),$$

where b_j are components of the vector b. Since $|p_s| \geqslant |p| - \|1 - Q^{-1}\|\, |p|$, where $\|1 - Q^{-1}\|$ is the norm of the operator given by the matrix $\delta_{ij} - Q_{ij}^{-1}$ in the Euclidean space $\mathbb{R}^d$, so $|p| \leqslant 2|p_s|$ for $\|1 - Q^{-1}\| \leqslant \frac{1}{2}$. The explicit form of the seminorm q and the inclusion $h \in \mathscr{C}$ imply easily that $\sup_s q(p_j\, \partial_i h^{\sim}(p_s))$ is bounded uniformly in (b,Q) for (b,Q) sufficiently close to the unity of the group $IO(d)$. Now, the obtained estimates and estimate (5.4) imply that for $g \in \mathscr{F}$

$$\int d\mu_j(\Phi)\, g(\Phi_{(a,R)}) \in \mathrm{CB}(IO(d)).$$

The inhomogeneous Euclidean group $IO(d)$ is amenable [since it is a semidirect product of two amenable groups, of the commutative group $\mathbb{R}^d$ and the compact group of rotations $O(d)$ (see Ref. 20, Theorems 2.3.3 and 3.6.2 and Ref. 37, Theorem XI, 1.1)]. So there exists an invariant mean for this group.

Let m be an invariant mean on $\mathrm{CB}(IO(d))$. Then

$$m\Big(\int d\mu_j(\Phi)\, g(\Phi_{(a,R)})\Big)$$

is correctly defined and gives a linear positive functional on $\mathscr{F}$ continuous with respect to the seminorm $\|\cdot\|_{F_{p'}}$.

By Theorem 3.1 there exists a (positive) measure μ_{j*} such that every $g \in \mathscr{F}$ is integrable and

$$m\Big(\int d\mu_j(\Phi)\, g(\Phi_{(a,R)})\Big) = \int d\mu_{j*}(\Phi)\, g(\Phi).$$

One can easily see that the measure μ_{j*} is Euclidean-invariant. Let now $\mu_* = \mu_{1*} - \mu_{2*} + i(\mu_{3*} - \mu_{4*})$; then μ_* is a Euclidean-invariant measure such that every $g \in \mathscr{F}$ is integrable and

$$m\Big(\int d\mu(\Phi)\, g(\Phi_{(a,R)})\Big) = \int d\mu_*(\Phi)\, g(\Phi), \quad g \in \mathscr{F}.$$

For $f \in \mathscr{C}_0(\mathbb{R}^{dn})$

$$\begin{aligned}
\langle T_n, f\rangle &= \langle T_n, f_{(a,R)}\rangle = m(\langle T_n, f_{(a,R)}\rangle) \\
&= m\Big(\int d\mu(\Phi)\, \langle \Phi_{(a,R)} \otimes \cdots \otimes \Phi_{(a,R)}, f\rangle\Big) \\
&= \int d\mu_*(\Phi)\, \langle \Phi \otimes \cdots \otimes \Phi, f\rangle,
\end{aligned}$$

that is, the measure μ_* satisfies the conditions of Theorem 5.1. Theorem 5.1 is proved.

Let us now proceed to obtaining estimates of the form (5.1) for the Schwinger functions of the quantum field $:\exp \varphi(x):_d$. These estimates allow us to prove Theorem 2.1.

The expression $:\exp \varphi(x):_d$ is correctly defined as a Wightman–Jaffe quantum field. Its Schwinger functions S_n^0 can easily be calculated, and their explicit forms are given in the formulation of Theorem 2.1. Moreover, at noncoinciding points, naturally,

$$S_n^0(x_1,\ldots,x_n) = \lim_\sigma \int d\mu_0(\xi) \prod_{i=1}^n :\exp \xi_\sigma(x_i):,$$

where σ is an ultraviolet cutoff and μ_0 is the Gaussian measure with covariance $G(x-y)$ and defined on the Schwartz space of tempered distributions.

The following estimates are valid:

Theorem 5.2:

$$\Big|\int d^{dn}x\, S_n^0(x_1,\ldots,x_n)\, f_n(x_1,\ldots,x_n)\Big| \leqslant c_n \sup_{p_1\cdots p_n \in \mathbb{R}^d} |\alpha(p_1)\cdots\alpha(p_n)\, f_n^{\sim}(p_1,\ldots,p_n)|,$$

$f_n \in \mathscr{C}_0^\alpha(\mathbb{R}^{dn}) \cap S_n(\mathscr{C})$, $\alpha(\cdot)$ is an entire function of the form (2.2) where $1 < \alpha < (d-2)(d-1)^{-1}$.

First of all we show that the estimate of Theorem 5.2 is that of the form (5.1).

Lemma 5.3: Let $q(f) = \sup_{p\in\mathbb{R}^d} |\alpha(p) f^{\sim}(p)|$, $f \in \mathscr{C}^\alpha$;

then

$$(q \otimes_\epsilon \cdots \otimes_\epsilon q)(f_n)$$

$$= \sup_{p_1 \cdots p_n \in \mathbb{R}^d} |\alpha(p_1) \cdots \alpha(p_n) f_n^{\sim}(p_1,\ldots,p_n)|,$$

$$f_n \in \mathscr{C}^\alpha(\mathbb{R}^{dn}). \tag{5.6}$$

Proof: This equality may be proved as follows. Since the image of the space $\mathscr{C}(\mathbb{R}^{dn})$ under the Fourier transformation is dense in the space of continuous functions with compact supports and with the topology of the inductive limit, which is described in Ref. 25, (Chap. III, §1, no. 1), the Fourier transform of the set $\{ \mu \in \mathscr{C}' | q^\circ(\mu) < \infty \}$ consists of Radon measures on $\mathbb{R}^d$. Here q° is the dual Minkowski functional of the seminorm q. The proposition (Ref. 25, Chap. IV, §4, no. 7, Proposition 12) implies that

$$q^\circ(\mu) = \sup\left\{ |\langle \mu, f\rangle| \ \ f \in \mathscr{C} \sup_{p \in \mathbb{R}^d} |\alpha(p) f^{\sim}(p)| \leqslant 1 \right\}$$

$$= \int \frac{d\,|\sigma(\mu)|\,(p)}{\alpha(p)},$$

where $\sigma(\mu)$ is the measure corresponding to the Fourier transform of the ultradistribution μ.

Hence, for $f_n = \Sigma_j f_1^j \otimes \cdots \otimes f_n^j$

$$(q \otimes_\epsilon \cdots \otimes_\epsilon q)(f_n)$$

$$= \sup_{q^\circ(\mu_i) \leqslant 1} \left| \sum_j \prod_{i=1}^n \int d\sigma(\mu_i)(p_i) f_i^j(p_i) \right|$$

$$\leqslant \sup_{p_1} \cdots \sup_{p_n} \left| \alpha(p_1) \cdots \alpha(p_n) \sum_j f_1^j(p_1) \cdots f_n^j(p_n) \right|$$

$$= \sup_{p_1 \cdots p_n} |\alpha(p_1) \cdots \alpha(p_n) f_n^{\sim}(p_1,\ldots,p_n)|\,.$$

On the other hand, choosing $d\sigma(\mu_i)(p') = \alpha(p_i)\delta(p' - p_i) \times d^d p'$, we have $q^\circ(\mu_i) \leqslant 1$ and

$$(q \otimes_\epsilon \cdots \otimes_\epsilon q)(f_n) \geqslant |\alpha(p_1) \cdots \alpha(p_n) f_n^{\sim}(p_1,\ldots,p_n)|.$$

Taking into account that $\mathscr{C}(\mathbb{R}^{dn})$ is the completed tensor product of $\mathscr{C}$ and $q \otimes_\epsilon \cdots \otimes_\epsilon q$ is a continuous seminorm on $\mathscr{C}(\mathbb{R}^{dn})$ we obtain equality (5.6). Lemma 5.3 is proved.

Now we proceed to the proof of Theorem 5.2.

Let us recall the definition of truncated functions; see, for example, Ref. 38, Ch. III, 5.C. Let a finite ordered set of indices be given. We identify this set with an interval of natural numbers $\{1,\ldots,n\}$. Let $\Pi = \bigcup_{k=1}^n \Pi_k$, where Π_k is the set of all partitions of the ordered set $\{1,\ldots,n\}$ on k nonempty and nonintersecting subsets $I_1,\ldots,I_k$, and in each subset I_j the indices $i_{j_1},\ldots,i_{j_{|I_j|}} \in I_j$ are naturally ordered $(i_{j_1} < i_{j_2} < \cdots)$. Here $|I_j|$ is the number of elements of I_j. The subsets I_j we call parts of the corresponding partition.

Truncated Schwinger functions T_n are defined in terms of Schwinger functions by the following recurrence formula:

$$S_n(x_1,\ldots,x_n) = \sum_{k=1}^n \sum_{\{I_j\} \in \Pi_k} \prod_{j=1}^k T_{|I_j|}(x_{i_{j_1}},\ldots x_{i_{j_{|I_j|}}}).$$

One can easily see that symmetry and Euclidean invariance of Schwinger functions imply symmetry and Euclidean invariance of truncated Schwinger functions. For Euclidean invariance this easily follows by induction, and for symmetry this follows from simple arguments. Indeed, using induction on the number of arguments, it is sufficient to consider a permutation of neighboring arguments x_i and x_{i+1}. If indices of these arguments belong to the same part I, belonging to a partition $\rho \in \Pi$, then the symmetry of this term follows from the induction hypothesis. If these indices belong to different parts I' and I'', $i_1 \in I'$, $i_2 \in I''$, of a given partition ρ, then let ρ' be the partition obtained by permuting $i \leftrightarrow i+1$ from the partition ρ. If $\rho' = \rho$ then the term corresponding to ρ is symmetric under the permutation $x_i \leftrightarrow x_{i+1}$. If $\rho' \neq \rho$, then the sum of terms corresponding to ρ and ρ' is symmetric.

Let T_n^0 be the truncated Schwinger functions of the quantum field $:\exp \varphi(x):_d$ (defined at noncoinciding points).

Lemma 5.4:

$$|T_n^0(x_1,\ldots,x_n)| \leqslant c_n \exp\Big(-mc_n' \max_{i,j} |x_i - x_j| \Big)$$

$$\times \max_{i \neq j} \exp[c_n''\, G(c_n'''(x_i - x_j))] \tag{5.7}$$

$$\leqslant c_n \exp\Big(-\frac{mc_n'}{n} \sum_{i=2}^n |x_i - x_{i-1}| \Big)$$

$$\times \sum_{i<j} \exp[c_n''\, G(c_n'''(x_i - x_j))], \tag{5.8}$$

where c_n, c_n', c_n''' are strictly positive constants, m is the mass appearing in the two-point function $G(x)$.

Proof of Lemma 5.4: Let us define Jost vectors. Let $x_1,\ldots,x_k$ be noncoinciding points and $(x_1,\ldots,x_k) = (R(t_1, u_1) - a, R(t_2, u_2) - a, \cdots)$, where $R \in O(d)$, $a \in \mathbb{R}^d$, $u_i \in \mathbb{R}^{d-1}$, and $0 < t_{\pi(1)} \leqslant t_{\pi(2)} \cdots \leqslant t_{\pi(k)}$ for some permutation π. Then for any Wightman–Jaffe field A one can naturally define a Jost vector $J(x_1,\ldots,x_k)$ [more precisely, the Jost vector $J_{a,R}(x_1,\ldots,x_k)$ with respect to the coordinate system with the origin at the point a and with a temporal direction $R(1,0)$],

$$J(x_1,\ldots,x_k) = \exp(-t_{\pi(1)} H) A(0,u_{\pi(1)})$$
$$\times \exp[(t_{\pi(1)} - t_{\pi(2)})H] A(0,u_{\pi(2)}) \cdots \Omega,$$

belonging to the Hilbert space of the field A. Here H is the Hamiltonian and Ω is the vacuum of the Wightman–Jaffe field A. The correctness of this definition follows from the smoothness of Schwinger functions at noncoinciding points (cf. also formulas II.27–II.28 in Simon's book[1]).

Before proceeding to the proof of Lemma 5.4, we formulate a statement we need.

Lemma 5.5 (Ruelle): Let X' be a subset of indices $(1,\ldots,n)$ and all the indices are naturally ordered. Let $X'' = (1,\ldots,n) \setminus X'$ be the complement of X' and again indices are naturally ordered. For each configuration $(x_1,\ldots,x_n)$, $x_i \in \mathbb{R}^d$, there is such a decomposition X', X'' and such a hyperplane that

$$\min_{i' \in X', i'' \in X''} |x_{i'} - x_{i''}| \geqslant (n-1)^{-1} \max_{i,j} |x_i - x_j| \qquad (5.9)$$

and the sets $\{x_{i'} | i' \in X'\}$ and $\{x_{i''} | i'' \in X''\}$ lie at a distance $\geqslant (2n-2)^{-1} \max_{i,j} |x_i - x_j|$ from the hyperplane in different half-spaces of this hyperplane.

Proof [see Jost's book (Ref. 38, Chap. VI, §5, Lemma 1)]: Let $\Delta = |x_k - x_l| = \max_{i,j} |x_i - x_j|$. $(x_1,...,x_n)$ are points in the d-dimensional Euclidean space. Hyperplanes containing the point x_k or x_l and orthogonal to the vector $x_k - x_l$ are supporting hyperplanes of the convex set generated by points $(x_1,...,x_n)$. Hence, hyperplanes orthogonal to $x_k - x_l$ and containing points x_i intersect the segment $x_k + s(x_l - x_k)$, $0 \leqslant s \leqslant 1$, and decompose it in at most $n-1$ intervals, with the total length Δ. The length of at least one of the intervals must be no less than $\Delta / n - 1$. Then the hyperplane intersecting the middle point of this interval and orthogonal to $x_k - x_l$ gives a decomposition X', X'' and a hyperplane satisfying the condition of the lemma. Lemma 5.5 is proved.

Now we proceed to estimate the truncated functions T_n^0 and to obtain bounds (5.7) and (5.8). Since $G(x) = G(-x)$ and

$$\max_{i,j} |x_i - x_j| \geqslant (n-1)^{-1} \sum_{i=2}^{n} |x_i - x_{i-1}|,$$

bound (5.8) follows from bound (5.7).

We prove bound (5.7) by induction.

For given $(x_1,...,x_n)$, let $\Delta = \max_{i,j} |x_i - x_j| = |x_k - x_l|$, X' and $X'' = (1,...,n) \backslash X'$ be a decomposition satisfying (5.9), and O be the point of the intersection of the segment $x_k + s(x_l - x_k)$, $0 \leqslant s \leqslant 1$, and the hyperplane such that $\{x_i | i' \in X'\}$ and $\{x_{i''} | i'' \in X''\}$ lie at a distance $\geqslant \Delta / 2(n-1)$, respectively, in the positive and negative half-space of the hyperplane (the existence of such a hyperplane is proved in Lemma 5.5).

The definition of truncated Schwinger functions and their symmetry imply that

$$\begin{aligned} T_n^0(x_1,...,x_n) &= S_n^0(x_1,...x_n) - S_{k'}^0(x_{1'},...,x_{k'}) S_{k''}^0(x_{1''},...,x_{k''}) \\ &\quad - \sum{}' \prod T_{|I_j|}^0, \quad 1',...,k' \in X', 1'',...,k'' \in X'', \end{aligned} \qquad (5.10)$$

where the summation Σ' runs over all partitions Π', which cannot be represented as a union of partitions of sets X' and X'', that is, in other words, the summation Σ' runs over all partitions Π' such that for each partition $\rho \in \Pi'$ there exists a part I', which is contained in the partition ρ such that $I \cap X' \neq \varnothing$, $I \cap X'' \neq \varnothing$.

Using the Osterwalder–Schrader reconstruction theorem,[39,40,21] symmetry and Euclidean invariance of Schwinger functions, the expression for $S_n^0(x_1,...,x_n)$ can be represented as the inner product of Jost vectors $J(x_{1'},...,x_{k'})$, $J(x_{1''},...,x_{k''})$ with respect to the coordinate system with the origin at the point O and with the vector $x_k - x_l$ as a temporal direction,

$$S_n^0(x_1,...,x_n) = (J(\vartheta x_{1'},..., \vartheta x_{k'}),\ J(x_{1''},...,x_{k''})),$$

where ϑ is the reflection of the direction given by the vector $x_k - x_l$ with respect to the point O, and where the inner product is taken in the Hilbert space corresponding, by the reconstruction theorem, to the quantum field $:\exp \varphi(x):_d$, that is, in the subspace of the Fock space of the free quantum field $\varphi(x)$, generated by smoothed polynomials in the field $:\exp \varphi(x):_d$.

Let $E_0^{\perp}$ be the projection on the subspace orthogonal to the vacuum. Then

$$\begin{aligned} &S_n^0(x_1,...,x_n) - S_{k'}^0(x_1,...,x_{k'}) S_{k''}^0(x_{1''},...,x_{k''}) \\ &\quad = (J(\vartheta x_{1'},...,\vartheta x_{k'}),\ E_0^{\perp} J(x_{1''},...,x_{k''})). \end{aligned} \qquad (5.11)$$

Since the Hamiltonian of the field $:\exp \varphi(x):_d$ coincides with the free one (more precisely, with the restriction of the free Hamiltonian on the corresponding subspace), so it has a mass gap $\geqslant m$ at the bottom part of the spectrum and (5.11) is estimated by the spectral theorem by

$$\begin{aligned} &\exp[-m\Delta / 2(n-1)] \| J(\vartheta(x_{1'} + y),...,\vartheta(x_{k'} + y)) \| \\ &\quad \times \| J(x_{1''} - y,...,x_{k''} - y) \|, \end{aligned} \qquad (5.12)$$

where $y = (x_k - x_l)(n-4)^{-1}$, that is, the vector of a length $\Delta / (4n-4)$ in the direction $x_k - x_l$, and $\|\cdot\|$ is the norm in the Fock space. Since for $i_1, i_2 \in X'$

$$|\vartheta(x_{i_1} + y) - (x_{i_2} + y)| \geqslant \Delta / 2n - 2 \geqslant (2n-2)^{-1} \min_{i \neq j} |x_i - x_j|,$$

and, analogously, for $i_1, i_2 \in X''$

$$|\vartheta(x_{i_1} - y) - (x_{i_2} - y)| \geqslant (2n-2)^{-1} \min_{i \neq j} |x_i - x_j|,$$

so expressing norms of Jost vectors in (5.12) in terms of functions S_n^0 and taking into account the explicit expression for S_n^0, the positivity and monotone dependence of $G(x)$ on x which follow from the representation

$$\begin{aligned} &G(x) = (2\pi)^{1-d} \int d\mathbf{p}\ [2\omega(\mathbf{p})]^{-1} \exp[-|x| \omega(\mathbf{p})], \\ &\omega(\mathbf{p}) = (\mathbf{p}^2 + m^2)^{1/2}, \end{aligned} \qquad (5.13)$$

we obtain the following bound:

$$\begin{aligned} &\| J(\vartheta(x_{1'} + y),...,\vartheta(x_{k'} + y)) \|\ \| J(x_{1''} - y,...,x_{k''} - y) \| \\ &\quad \leqslant c_n \max_{i \neq j} \exp[c_n'\ G(c_n''(x_i - x_j))]. \end{aligned} \qquad (5.14)$$

In items contained in the sum Σ' in (5.10) each factor $T_{|I|}^0$ in the product is dominated by the induction hypothesis by

$$\begin{aligned} &c_{|I|} \exp(-mc_{|I|}' \max_{i,j} |x_i - x_j|) \\ &\quad \times \max_{\substack{i,j \in I \\ i \neq j}} \exp[c_{|I|}''\ G(c_{|I|}'''(x_i - x_j))]. \end{aligned}$$

For each partition ρ over which the summation Σ' runs in (5.10), there exists a part I' entering the partition ρ such that $I \cap X' \neq \varnothing$ and $I \cap X'' = \varnothing$. Due to the choice of X', X'',

$$\max_{i,j\in I}|x_i-x_j|\geqslant \min_{\substack{i\in I\cap X'\\ j\in I\cap X''}}|x_i-x_j|$$

$$\geqslant (n-1)^{-1}\max_{i,j}|x_i-x_j|.$$

The obtained bounds imply the existence of such strictly positive constants c_n, c_n', c_n'', c_n''' that inequalities (5.7) are fulfilled. Lemma 5.4 is proved.

Now we proceed to obtaining bounds of Theorem 5.2.

We first remark that Lemma 5.4 and Ref. 21, Lemma 7 (or arguments given below) imply that for $f_n\in\mathscr{C}_0^\alpha(\mathbb{R}^{dn})$, $\alpha\leqslant(d-1)(d-2)^{-1}$, the expression

$$\langle T_n^0,f_n\rangle=\int d^{dn}x\, T_n^0(x_1,\dots,x_n)f_n(x_1,\dots,x_n)$$

is well defined.

Let us introduce relative coordinates $\xi_1=x_1$, $\xi_i=x_i-x_{i-1}$ for $2\leqslant i\leqslant n$.

Since $T_n^0(x_1,\dots,x_n)$ is translation invariant, there exists such a function $t_n^0(\xi_2,\dots,\xi_n)$ that at noncoinciding points, $x_i\neq x_j$ for $i\neq j$,

$$T_n^0(x_1,\dots,x_n)=t_n^0(\xi_2,\dots,\xi_n).$$

Hence,

$$|\langle T_n^0,f_n\rangle|\leqslant\sum_{i<j}c_n^{(1)}\int d^{dn-d}\xi\exp\Big(-c_n^{(2)}\sum_{i=2}^n|\xi_i|\Big)\times\exp\Big(c_n^{(3)}G\Big(c_n^{(4)}\sum_{k=i+1}^{j}\xi_k\Big)\Big)\times\Big|\int d^d\xi_1 f_n(\xi_1,\dots,\xi_1+\dots+\xi_n)\Big|. \quad (5.15)$$

Expanding the exponent and using an estimate

$$0<G(x)\leqslant c|x|^{2-d-\epsilon}\exp(-m|x|/2),$$

where $\epsilon=0$ for $d>2$ and $\epsilon>0$ for $d=2$, and which follows from representation (5.13), we obtain

$$\exp(c_n^{(3)}G(c_n^{(4)}x))\leqslant\sum_{l=0}^{\infty}\frac{c_n^{(5)l}}{l!}|x|^{-\{(d-2+\epsilon)l\}},$$

where by $\{(d-2+\epsilon)l\}$ we denote the smallest integer which is larger or equal to $(d-2+\epsilon)l$.

Using the Taylor expansion we write for $f_n\in\mathscr{C}_0(\mathbb{R}^{dn})$

$$\Big|\int d^d\xi_1 f_n(\xi_1,\dots,\xi_1+\dots+\xi_i,\dots,\xi_1+\dots+\xi_i+y,\dots,\xi_1+\dots+\xi_n)\Big|$$

$$\leqslant\sum_{l_1+\dots+l_d=\{(d-2+\epsilon)l\}}\frac{|y_1|^{l_1}\cdots|y_d|^{l_d}}{l_1!\cdots l_d!}\sup_\xi\Big|\int d^d\xi_1\,\partial_{j_1}^{l_1}\cdots\partial_{j_d}^{l_d}f_n(\xi_1,\dots,\xi_1+\dots+\xi_n)\Big|$$

$$\leqslant d^{\{(d-2+\epsilon)l\}}\frac{|y|^{\{(d-2+\epsilon)l\}}}{\{(d-2+\epsilon)l\}!}\sup_{\substack{\xi\\ l_1+\dots+l_d=\{(d-2+\epsilon)l\}}}\Big|\int d^d\xi_1\partial_{j_1}^{l_1}\cdots\partial_{j_d}^{l_d}f_n(\xi_1,\dots,\xi_1+\dots+\xi_n)\Big|,$$

where ∂_{j_i} is the partial derivative with respect to the ith component of the variable x_j.

Using the obtained bounds and continuing bound (5.15), we obtain

$$|\langle T_n^0,f_n\rangle|\leqslant\sum_{i<j}\sum_{l=0}^{\infty}c_n^{(6)}\frac{c_n^{(6)l}}{l!\{(d-2+\epsilon)l\}!}\times\sup_{\substack{\xi\\ l_1+\dots+l_d=\{(d-2+\epsilon)l\}}}\Big|\int d^d\xi_1\,\partial_{j_1}^{l_1}\cdots\partial_{j_d}^{l_d}f_n(\xi_1,\dots,\xi_1+\dots+\xi_n)\Big|.$$

Further,

$$\int d^d\xi_1\,\partial_{j_1}^{l_1}\cdots\partial_{j_d}^{l_d}f_n(\xi_1,\dots,\xi_1+\dots+\xi_n)=(2\pi)^{-d(n-2)/2}\,i^{l_1+\dots+l_d}\times\int d\sigma_n(p)\,p_{j_1}^{l_1}\cdots p_{j_d}^{l_d}f_n^{\sim}(p_1,\dots,p_n),$$

where $f_n^{\sim}(p_1,\dots,p_n)$ denotes the Fourier transform of the function $f_n(x_1,\dots,x_n)$ and the positive $d\sigma_n(p)=d^{dn}p\times\delta(\Sigma_{i=1}^n p_i)$.

Hence,

$$\Big|\int d^d\xi_1\,\partial_{j_1}^{l_1}\cdots\partial_{j_d}^{l_d}f_n(\xi_1,\dots,\xi_1+\dots+\xi_n)\Big|\leqslant\int d\sigma_n(p)\,|p_j|^{l_1+\dots+l_d}|f_n^{\sim}(p_1,\dots,p_n)|. \quad (5.16)$$

Using estimate (5.16), we obtain

$$|\langle T_n^0,f_n\rangle|\leqslant\sum_{i<j}c_n^{(6)}\int d\sigma_n(p)\sum_{l=0}^{\infty}\frac{c_n^{(6)l}|p_j|^{\{(d-2+\epsilon)l\}}}{l!\{(d-2+\epsilon)l\}!},$$

$$|f_n^{\sim}(p_1,\dots,p_n)|\leqslant\sum_{i<j}c_n^{(6)}\int d\sigma_n(p)\,\beta(c_n^{(7)}p_j)|f_n^{\sim}(p_1,\dots,p_n)|$$

for an entire function $\beta(\cdot)$ of form (2.2) and of order equal to $(d-2+\epsilon)(d-1+\epsilon)^{-1}$.

Thus, taking into account that the measure $d\sigma_n(p)$ is symmetric, we have for $f_n\in\mathscr{C}_0(\mathbb{R}^{dn})\cap\mathrm{S}_n(\mathscr{C})$

$$|\langle T_n^0, f_n\rangle| \leqslant \sum_{i<j} c_n^{(6)} \int d\sigma_n(p)\beta(c_n^{(7)}p_j)|f_n^{\sim}(p_1,...,p_n)|$$

$$= \tfrac{1}{2} n(n-1)c_n^{(6)} \int d\sigma_n(p)\beta(c_n^{(7)}p_n)|f_n^{\sim}(p_1,...,p_n)|$$

$$\leqslant c_n^{(8)} \sup_p \left| \prod_{i=1}^n \beta(c_n^{(7)}p_i)\, f_n^{\sim}(p_1,...,p_n)\right|,$$

where $c_n^{(8)} = \frac{1}{2} n(n-1)\, c_n^{(6)} \int d\sigma_n(p)\, \Pi_{i=1}^{n-1}\beta(c_n^{(7)}p_i)^{-1} < \infty$.

Since $\beta(c_n^{(7)}p) \leqslant c_n^{(9)}\alpha(p)$, where $\alpha(p)$ is an entire function of the form (2.2) and of the order larger than $(d-2+\epsilon)(d-1+\epsilon)^{-1}$, then, choosing an appropriate ϵ, we finally have the bound

$$|\langle T_n^0, f_n\rangle| \leqslant c_n \sup_p \left| \prod_{i=1}^n \alpha(p_i)\, f_n^{\sim}(p_1,...,p_n)\right|,$$

$$f_n \in \mathscr{C}_0(\mathbb{R}^{dn}) \cap S_n(\mathscr{C}), \qquad (5.17)$$

for an entire function of the form (2.2) and of the order larger than $(d-2)(d-1)^{-1}$.

We use the bound (5.17) to obtain bounds for $S_n^0(x_1,...,x_n)$. For this purpose we do the following. Let T_n be an extension on $\mathscr{C}(\mathbb{R}^{dn})$ of the linear functional T_n^0, given on the subspace $\mathscr{C}_0(\mathbb{R}^{dn}) \cap S_n(\mathscr{C})$, which satisfies an estimate

$$|\langle T_n, f_n\rangle| \leqslant c_n \sup_p \left| \prod_{i=1}^n \alpha(p_i)\, f_n^{\sim}(p_1,...,p_n)\right|,$$

$$f_n \in \mathscr{C}(\mathbb{R}^{dn}). \qquad (5.18)$$

By the Hahn–Banach theorem such extensions exist. Let

$$S_n = \sum_{k=1}^n \sum_{\{I_j\}\in\rho_k} \prod_{j=1}^k T_{|I_j|d},$$

then S_n as a sum of direct products of ultradistributions $T_{|I_j|}$ is a well-defined ultradistribution and its restriction on $\mathscr{C}_0(\mathbb{R}^{dn}) \cap S_n(\mathscr{C})$ coincides with S_n^0.

For the Fourier transform $S_n^{\sim}$ we have

$$S_n^{\sim} = \sum_{k=1}^n \sum_{\{I_j\}\in\rho_k}^{\infty} \prod_{j=1}^k T_{|I_j|}^{\sim}.$$

The Fourier image of the space $\mathscr{C}(\mathbb{R}^{dn})$ is dense in the space of continuous functions with compact supports and the topology of the inductive limit described in Ref. 25, Chap. III, §1, no. 1. Thus, this density, bound (5.18) and the proposition (Ref. 25, Chap. IV, §4, no. 7, Proposition 12) imply that $T_{|I_j|}$ are given by Radon measures $\sigma(T_{|I_j|}^{\sim})$ on $\mathbb{R}^{d|I_j|}$ such that

$$\int d|\sigma(T_{|I_j|}^{\sim})|(p) \prod_{i=1}^{|I_j|} \alpha(p_i)^{-1} < \infty.$$

Since $S_n^{\sim}$ is a sum of direct products of $T_{|I_j|}^{\sim}$, that is, a sum of tensor products of the corresponding measures, the Fubini theorem (Ref. 25, Chap. III, §4, no. 1, Theorem 2) and bound (5.18) imply that

$$\left|\left\langle \prod_{j=1}^k T_{|I_j|}^{\sim}, f_n^{\sim}\right\rangle\right|$$

$$= \left|\int \bigotimes_{j=1}^k d\sigma(T_{|I_j|}^{\sim}) f_n^{\sim}(p_1,...,p_n)\right|$$

$$\leqslant \sup_p \left| \prod_{j=1}^k c_{|I_j|} \prod_{i\in I_j} \alpha(p_i)\, f_n^{\sim}(p_1,...,p_n)\right|$$

$$\leqslant c_n \sup_p \left| \prod_{i=1}^n \alpha(p_i)\, f_n^{\sim}(p_1,...,p_n)\right|.$$

This implies also that

$$|\langle S_n^{\sim}, f_n^{\sim}\rangle| \leqslant c_n' \sup_p \left| \prod_{i=1}^n \alpha(p_1\, f_n^{\sim}(p_1,...,_n)\right|.$$

Thus, for $f_n \in \mathscr{C}_0(\mathbb{R}^{dn}) \cap S_n(\mathscr{C})$,

$$|\langle S_n^0, f_n\rangle| = |\langle S_n, f_n\rangle|$$

$$\leqslant c_n' \sup_p \left| \prod_{i=1}^n \alpha(p_i)\, f_n^{\sim}(p_1,...,p_n)\right|.$$

The obtained bound proves Theorem 5.2.

Proof of Theorem 2.1: Theorem 2.1 follows now from Theorems 5.1, 5.2, and Lemma 5.3. The assertion that a measure can be chosen to be real follows from the reality of the Schwinger functions S_n^0. Theorem 2.1 is proved.

ACKNOWLEDGMENTS

The author is indebted to R. L. Dobrushin for the possibility of participating in the Fifth International Symposium on Information Theory, where a preliminary version of this paper was reported. The author is also indebted to J. Yngvason for having sent his preprint[19] prior to publication.

[1]B. Simon, *The P(φ)₂ Euclidean (Quantum) Field Theory* (Princeton U. P., Princeton, NJ, 1974).

[2]V. A. Malyshev, "Probability aspects of quantum field theory (in Russian)," in *Probability Theory. Mathematical Statistics. Theoretical Cybernetics. Vol. 14 (Review of Science and Technics)* (Viniti, Moscow, 1977).

[3]J. S. Feldman and K. Osterwalder, "The Wightman axioms and the mass gap for weakly coupled $(\varphi^4)_3$ quantum field theories," Ann. Phys. (N.Y.) **97**, 80–135 (1976).

[4]J. Magnen, and R. Sénéor, "The infinite volume limit of the φ_3^4 model," Ann. Inst. H. Poincaré A **24**, 95–159 (1976).

[5]Y. M. Park, "Convergence of lattice approximations and infinite volume limit in the $(\lambda\Phi^4 - \sigma\Phi^2 - \mu\Phi)_3$ field theory," J. Math. Phys. **18**, 354–66 (1977).

[6]J. Glimm and A. Jaffe, *Quantum Physics. A Functional Integral Point of View* (Springer, New York, 1981).

[7]D. Ya. Petrina and A. L. Rebenko, "Projection-iteration method of solving quantum field theory equations and its relation to renormalization theory. Quantum field theory equations and improperly stated problems of mathematical physics (in Russian)," Teor. Math. Phys. **42**, 167–83 (1980).

[8]R. Schrader, "A possible constructive approach to $\Phi_4{}^4$," Commun. Math. Phys. **49**, 131–53 (1976).

[9]R. Schrader, "A possible constructive approach to $\Phi_4{}^4$ II," Ann. Inst. H. Poincaré **26**, 295–301 (1977).

[10]R. Schrader, "A possible constructive approach to $\varphi_4{}^4$ III," Commun. Math. Phys. **50**, 97–102 (1976).

[11]H. Lehmann and K. Pohlmeyer, "On the superpropagator of fields with exponential coupling," Commun. Math. Phys. **20**, 101–10 (1971).

[12]H. Epstein, and V. Glaser: "Renormalization of nonpolynomial Lagrangians in Jaffe's class," Commun. Math. Phys. **27**, 181–94 (1972).

[13]M. K. Volkov and V. N. Pervushin, *Essentially Non-Linear Quantum Theories, Dynamical Symmetries and Physics of Mesons* (in Russian) (Atomizdat, Moscow, 1978).

[14]S. Albeverio, G. Gallavotti, and R. Høegh-Krohn, "Some results for the exponential interaction in two or more dimensions," Commun. Math. Phys. **70**, 187–92 (1979).
[15]E. P. Osipov, "On triviality of the :exp $\lambda\varphi$:$_4$ quantum field theory in a finite volume," Preprint TPh-103, Institute of Mathematics, Novosibirsk, USSR, 1979, and Rept. Math. Phys. **19**, No. 1 145–50 (1981).
[16]E. P. Osipov, "A constructive approach to the polynomial interaction in four-dimensional space-time," Preprint TPh-108, Institute for Mathematics, Novosibirsk, USSR, 1979.
[17]E. P. Osipov, "Feynman integral for the exponential interaction in four-dimensional space-time, I" (in Russian), Teor. Math. Phys. **47**, 307–14 (1981).
[18]E. P. Osipov, "Feynman integral for the exponential interaction in four-dimensional space-time, II" (in Russian), Teor. Math. Phys. **48**, 156–67 (1981).
[19]J. Yngvason, "Translationally invariant states and the spectrum ideal in the algebra of test functions for quantum fields," preprint, Reykjavik, 1980.
[20]F. P. Greenleaf, *Invariant Means on Topological Groups and Their Applications* (Van Nostrand Reinhold, New York, 1969).
[21]F. Constantinescu and W. Thalheimer, "Euclidean Green's functions for Jaffe fields," Commun. Math. Phys. **38**, 299–316 (1974).
[22]G. Björk, "Linear partial differential operators and generalized distributions," Ark. Math. **6**, 351–407 (1966).
[23]A. M. Jaffe, "High energy behavior in quantum field theory. I. Strictly localizable fields," Phys. Rev. **158**, 1454 (1967).
[24]I. M. Gel'fand, and G. E. Shilov, *Generalized functions* (Academic, New York, 1964), Vol. II.
[25]N. Bourbaki, *Integration* (Hermann, Paris, 1952–1969), Livre VI, Chaps. 1–9.
[26]G. V. Efimov, *Nonlocal Interactions of Quantized Fields* (Nauka, Moscow, 1977) (in Russian).
[27]K. W. Symanzik, "A modified model of Euclidean quantum field theory," Report IMM–NYU 327, Courant Institute of Mathematical Sciences, New York University, 1964.
[28]J. Challifour and S. Slinker, "Euclidean field theory I. The moment problem," Commun. Math. Phys. **43**, 41–58 (1975).
[29]J. Yngvason, "On the decomposition of Wightman functionals in the Euclidean framework," Rept. Math. Phys. **13**, 101–15 (1978).
[30]E. P. Osipov, paper in preparation.
[31]R. P. Boas, Jr., "The Stieltjes moment problem for functions of bounded variation," Bull. Am. Soc. **45**, 399–404 (1939).
[32]K. A. Čevlakov, A. M. Slinko, I. P. Shestakov, and A. I. Shirshov, *Rings Close to Associative* (Nauka, Moscow, 1978) (in Russian).
[33]H. J. Borchers, and J. Yngvason, "Necessary and sufficient conditions for integral representations of Wightman functionals at Schwinger points," Commun. Math. Phys. **47**, 197–213 (1967).
[34]A. P. Robertson and W. J. Robertson, *Topological Vector Spaces*, (Cambridge U. P., Cambridge, 1964).
[35]A. Pietsch, *Nuclear Locally Convex Spaces* (Springer, Berlin, 1972).
[36]F. P. Gantmakher, *Theory of Matrices* (Nauka, Moscow, 1966) (in Russian).
[37]N. Dunford and J. T. Schwartz, *Linear Operators. Part II. Spectral Theory. Self Adjoint Operators in Hilbert Space* (Interscience, New York, 1963).
[38]R. Jost, *The General Theory of Quantized Fields* (Am. Math. Soc., Providence, RI, 1965).
[39]K. Osterwalder and R. Schrader, "Axioms for Euclidean Green's functions," Commun. Math. Phys. **31**, 83–112 (1973).
[40]K. Osterwalder and R. Schrader, "Axioms for Euclidean Green's functions, II," Commun. Math. Phys. **42**, 281–305 (1975).

Separable systems for the Dirac equation in curved space-times

Roland Rüdiger[a)]
Institut für Astronomie und Astrophysik der Universität Würzburg, Am Hubland, 8700 Würzburg, Federal Republic of Germany

(Received 9 August 1982; accepted for publication 11 February 1983)

In this paper, the consequences of the existence of a Killing–Yano tensor for the separability of the general relativistic Dirac equation in an exterior electromagnetic field are investigated. Those properties of Killing–Yano tensors, which are relevant for the subject of this work, are reviewed; it is then shown that for any class of metrics admitting this type of tensor the Dirac equation can be separated at least once. Furthermore, all separable systems which are obtained in this way are stated explicitly. Finally, for the special case of the Kerr solution, the formalism of the present paper is compared with Chandrasekhar's work on the separability of the Dirac equation.

PACS numbers: 11.10.Qr

I. INTRODUCTION

The solution of wave equations by separation-of-variables methods in a curved background has played an important role in the analysis of the Kerr geometry. Carter[1,2] has shown that the existence of his fourth constant of the motion of test particles also leads to a separation of the Klein–Gordon equation. This result has been extended to the massless spin-1 and spin-2 equations and later to the massless spin-$\frac{1}{2}$ equation by Teukolski[3,4] and Unruh.[5] Using a different method, Chandrasekhar[6] resolved the remaining problem of separating the Dirac equation in a Kerr background. His work was generalized to a class of type-D vacuum space-times by Güven,[7] and to the charged case by Toop[8] and Page.[9]

The Killing tensor[10] of the Kerr metric has the peculiar property that it can be written as the square of a skew tensor f^*_{ab}:$K_{ab} = f^*_{ak} f^{*k}{}_b$. Tensors of this type have been called Killing–Yano tensors. Carter and McLenaghan[11] discovered the remarkable fact that the separation of the Dirac equation in a curved background is connected to this quantity rather than to the Killing tensor itself. They succeeded in constructing an operator, commuting with the Dirac operator, with the property that the Chandrasekhar separation constants are interpretable as eigenvalues of this operator. In connection with this work, McLenaghan and Spindel[12] determined the general self-adjoint first-order differential operator which commutes with the Dirac operator. They found that this operator can be constructed from Killing–Yano tensors of different valences (in the nomenclature of Dietz and Rüdiger,[13,14] referred to henceforth as DR I and DR II).

Therefore, it appears to be an interesting problem to investigate the separability of the Dirac equation under the sole assumption that the underlying space-time geometry admits one Killing–Yano tensor. To do this, one has to construct a coordinate system and a set of operators which together guarantee the separability. For a Killing vector field the corresponding problem is trivial: with respect to a coordinate system with one of the coordinates adapted to the Killing vector field, one can always separate one factor of the wave function; the corresponding operator is the Lie derivative. A corresponding statement for a single irreducible Killing tensor seems not to exist for any type of wave equation. General relations between the existence of Killing tensors and the separability of the Klein–Gordon equation have been investigated by many authors; see the review article by Benenti and Francaviglia.[15] The corresponding problem concerning Killing–Yano tensors and the Dirac equation, however, can be solved because these tensors have been classified, and canonical line elements for each type are available.[13,14]

So, in this paper, the sole assumption concerning the background geometry will be that there exists a Killing–Yano tensor. In particular, there will be no explicit restrictions on the Ricci tensor and no explicit assumptions on the existence of isometries. We shall investigate the consequences of this assumption for the separability of the Dirac equation along the lines of the Chandrasekhar procedure.

The case of a Killing–Yano tensor of valence 3 will not be treated in this paper because the corresponding canonical metric is of the form of the Robertson–Walker metric, but with an arbitrary 3-metric. The way in which the Dirac equation can be separated in this type of metric has already been known for a long time.[16] Killing–Yano tensors of valence 1 and valence 4 are trivial. Therefore, in this paper, we restrict ourselves to the valence 2 case.

In the case of an algebraically general Killing–Yano tensor of valence 2, the relevant operators will be constructed in such a way that they possess a definite GHP type.[17] In particular, the procedure of this paper is invariant under the GHP operation prime in contrast to the formalisms of Chandrasekhar[6] and Güven.[7] It turns out that for all types of Killing–Yano tensors of valence 2, the Dirac equation can be decoupled into at least two pairs of coupled equations for two spinor components. The details of this separation depend on the type of the Killing–Yano tensor. According to the classification of these tensors, as given in DR I and DR II, which will be summarized as far as it is relevant for this paper in Sec. 2, there are 3, 2, 2, and 1 separation constants for the types (4, I), (4, II), (4, III), and (4, IV), respectively (see

a) Permanent address: Hochschulrechenzentrum, Universität Essen-GHS, Schützenbahn 70, 4300 Essen 1, Federal Republic of Germany.

Table I), one of which is nontrivial, i.e., does not belong to a Killing vector field. In cases (2, I), (2, II), and (2, III) there exists at least one separation constant. In the last case (of an algebraically special Killing–Yano tensor), this is trivial.

The separation procedure of this paper is strictly local. There exist important and more difficult open questions concerning the global nature of the functions obtained by this separation procedure, in particular, the question whether or under what conditions these functions form a complete set. These questions are beyond the scope of this paper.

II. KILLING–YANO TENSORS OF VALENCE 2

In this section, we summarize some of those results of DR I and DR II on Killing–Yano tensors of valence 2 which will be needed in the following sections. The defining equation of this type of quantity is

$$\nabla_{(a} f^{*}{}_{b)c} = 0, \tag{2.1}$$

where the skew tensor $f^{*}{}_{ab}$ can be translated into spinor language by

$$f^{*}{}_{ab} = i(\epsilon_{AB}\bar{\chi}_{A'B'} - \epsilon_{A'B'}\chi_{AB}).$$

Written in terms of the symmetric spinor χ_{AB}, Eq. (2.1) is equivalent to the two spinor equations

$$\nabla^{A'}{}_{(A}\chi_{BC)} = 0, \quad \nabla_{A'}{}^{R}\chi_{RA} = -\nabla_{A}{}^{R'}\bar{\chi}_{A'R'}. \tag{2.2}$$

If χ_{AB} is algebraically general, it can be written in the form

$$\chi_{AB} = \psi o_{(A}\iota_{B)} \tag{2.3}$$

or, equivalently,

$$f^{*}{}_{ab} = \operatorname{Im}\psi D_{ab} + \operatorname{Re}\psi D^{*}_{ab},$$

where $D_{ab} = 2k_{[a}n_{b]}$ and $\{k_a, n_a, m_a, \bar{m}_a\}$ is the canonical null tetrad connected with the dyad o_A, ι_A. The normalization $o_K\iota^K = 1$ implies that the right-hand side of Eq. (2.3) is unique up to transformations of the form $o_A \to \lambda o_A$, $\iota_A \to \lambda^{-1}\iota_A$, $\psi \to \psi$ or $o_A \to \lambda\iota_A$, $\iota_A \to -\lambda^{-1}o_A$, $\psi \to -\psi$, where $\lambda \in \mathbb{C}$. Equations (2.2) written in the GHP formalism take the form

$$\begin{aligned}
&\kappa = \sigma = 0,\\
&\text{þ}\psi = -\rho\psi,\\
&\eth\psi = -\tau\psi,\\
&\rho\psi = \bar{\rho}\bar{\psi},\\
&\tau\psi = -\bar{\tau}'\bar{\psi},
\end{aligned} \tag{2.4}$$

together with the primed versions of these equations. Note that the amplitude ψ of the Killing–Yano tensor is of type (0,0) and that $\psi' = -\psi$. The corresponding equations in the algebraically special case will not be needed here.

Table I summarizes the relevant results on Killing–Yano tensors. By definition, the parameter ν of column 5 takes the values $-1, 0,$ or $+1$ if $d(\operatorname{Re}\psi)$ is timelike, null, or spacelike, respectively. So, in cases (4, III) and (4, IV), ν is not defined. In DR I, an arbitrary parameter (called e) has been introduced in these cases to obtain line elements which are formally symmetric to those of the other cases. To simplify some of the equations, this parameter has been set to zero in the present paper. Table I and the following list also incorporate some further slight modifications and simplifications of the results of DR I and DR II, which simplify some of the final equations of this paper.

The directional derivatives $D, D' = \Delta$, $\delta, \delta' = \bar{\delta}$ will be needed in Sec. IV. We list these because they have not been

TABLE I. Classification of Killing–Yano tensors and properties of canonical metrics.

		Type	Defining property of type	Form of ψ	Arbitrary functions in canonical metric	Value of ν	Killing vector fields ∂_v	Killing vector fields ∂_y
Algebraically general	Rank 4	(4,I)	$d(\operatorname{Re}\psi)\wedge d(\operatorname{Im}\psi)\neq 0$	$\psi = u + ix$	$A(u), B(x)$	-1	spacelike	spacelike
						0	spacelike	spacelike
						$+1$	a	b
		(4,II)	$d(\operatorname{Re}\psi)\neq 0$, $d(\operatorname{Im}\psi) = 0$	$\psi = u + il$	$C(x,y), D(x,y)$, $A(u)$	-1	spacelike	
						0	null	
						$+1$	timelike	
		(4,III)	$d(\operatorname{Re}\psi) = 0$, $d(\operatorname{Im}\psi)\neq 0$	$\psi = k + ix$	$C(u,v), D(u,v)$, $B(x)$	undef.		spacelike
		(4,IV)	$d(\operatorname{Re}\psi) = 0$, $d(\operatorname{Im}\psi) = 0$	$\lvert\psi\rvert = 1$, $\operatorname{Re}\psi\operatorname{Im}\psi\neq 0$	$E(u,v), F(\zeta,\bar{\zeta})$	undef.		
	Rank 2	(2,I)	$f^{*}{}_{ab}$ spacelike	$\psi(u,v)$ real, >0 arbitrary	$A(u,v), F(\zeta,\bar{\zeta})$			
		(2,II)	$f^{*}{}_{ab}$ timelike	$\psi(\zeta,\bar{\zeta})/i$ real, >0 arbitrary	$C(u,v), G(\zeta,\bar{\zeta})$			
Algebr. special		(2,III)	$f^{*}{}_{ab}$ null	$\psi = 1$	$P(u,x)$, $H(u,x,y)$, $\Omega(u,x,y)$		null	

[a] Spacelike if $A^2 - B^2 < 0$, timelike if $A^2 - B^2 > 0$, null if $A^2 = B^2$.
[b] Spacelike if $B^2u^4 - A^2x^4 > 0$, timelike if $B^2u^4 - A^2x^4 < 0$, null if $B^2u^4 - A^2x^4 = 0$.

given explicitly in DR I and DR II.

Type (4,I):

$$D = 2^{-1/2}|\psi|^{-1}\{-A\partial_u + \nu u^2 A^{-1}\partial_v + \nu A^{-1}\partial_y\},$$

$$D' = 2^{-1/2}|\psi|^{-1}\{\nu A\partial_u + 2u^2(1+\nu^2)^{-1}A^{-1}\partial_v + 2(1+\nu^2)^{-1}A^{-1}\partial_y\},$$

$$\delta = 2^{-1/2}|\psi|^{-1}\{ix^2B^{-1}\partial_v + B\partial_x - iB^{-1}\partial_y\}.$$

Type (4,II):

$$D = 2^{-1/2}|\psi|^{-1}\{-A\partial_u + \nu|\psi|^2A^{-1}\partial_v\},$$

$$D' = 2^{-1/2}|\psi|^{-1}\{\nu A\partial_u + 2(1+\nu^2)^{-1}|\psi|^2A^{-1}\partial_v\},$$

$$\delta = 2^{-1/2}|\psi|^{-1}(\partial_x C)^{-1/2}\{2ilCD\partial_v + D^{-1}\partial_x - iD\partial_y\}.$$

Type (4,III):

$$D = -2^{-1/2}|\psi|^{-1}(\partial_u C)^{-1/2}D^{-1}\partial_u,$$

$$D' = 2^{-1/2}|\psi|^{-1}(\partial_u C)^{-1/2}\{D\partial_v + 4kCD\partial_y\},$$

$$\delta = 2^{-1/2}|\psi|^{-1}\{B\partial_x + i|\psi|^2B^{-1}\partial_y\}.$$

Type (4,IV):

$$D = -2^{-1/2}E^{-1}\partial_u,$$

$$D' = 2^{-1/2}E^{-1}\partial_v,$$

$$\delta = 2^{-1/2}F^{-1}\partial_{\bar\xi}.$$

Type (2,I):

$$D = -2^{-1/2}A^{-1}\partial_u,$$

$$D' = 2^{-1/2}A^{-1}\partial_v,$$

$$\delta = 2^{-1/2}(\psi F)^{-1}\partial_{\bar\xi}.$$

Type (2,II):

$$D = -2^{-1/2}(C\psi/i)^{-1}\partial_u,$$

$$D' = 2^{-1/2}(C\psi/i)^{-1}\partial_v,$$

$$\delta = 2^{-1/2}G^{-1}\partial_{\bar\xi}.$$

Type (2,III):

$$D = \partial_v,$$

$$D' = P^{-2}\{\partial_u - H\partial_v + (\partial_{\bar\xi}\Omega)\partial_\xi + (\partial_\xi\Omega)\partial_{\bar\xi}\},$$

$$\delta = -P^{-1}\partial_{\bar\xi}.$$

III. SEPARATION OF THE DIRAC EQUATION

According to Table I the canonical metrics of an algebraically special Killing–Yano tensor [Type (2,III)] possess a null Killing vector field. Apart from the trivial separability no further separation is possible, in general, because the function Ω depends on all of the remaining three coordinates. This case will not be considered further.

The Weyl tensors of all space–times admitting an algebraically general Killing–Yano tensor of valence 2 are of type D. Therefore, it is most convenient to work in the GHP formalism. The following equations, (3.1), together with their primed versions, constitute the Dirac equation in an exterior electromagnetic field (for $\phi = 0$, see Güven[7])

$$\begin{aligned}(\tilde{\text{Þ}} - ie\phi_2)f + i(\tilde\eth' + ie\phi_4')f' &= i\mu_e g,\\ (\tilde{\text{Þ}} - ie\phi_2)g' + i(\tilde\eth + ie\phi_4)g &= i\mu_e f'.\end{aligned} \tag{3.1}$$

Here the ϕ_a are the components of the electromagnetic four-potential: $\phi_1 = \phi_2' = \boldsymbol{\phi}\cdot\mathbf{n}$, $\phi_3 = \phi_4' = -\boldsymbol{\phi}\cdot\bar{\mathbf{m}}$, and the operators $\tilde{\text{Þ}}$ and $\tilde\eth$[18,19] are defined by

$$\tilde{\text{Þ}}\eta = (\text{Þ} + p\rho + q\bar\rho)\eta,$$

$$\tilde\eth\eta = (\eth + p\tau - q\bar\tau')\eta,$$

where η is a scalar of type (p,q). The components f, f', g, g' of the Dirac spinor are of types $(-1,0)$, $(1,0)$, $(0,1)$, $(0,-1)$, respectively.

To separate the Dirac equation under the assumption that a Killing–Yano tensor exists, we introduce the following operator

$$\mathscr{T}\eta = |\psi|\{\tilde{\text{Þ}}\eta - (1/m)(p\rho + q\bar\rho)\eta\}, \tag{3.2}$$

where $2m$ is the matrix rank of the Killing–Yano tensor. It is easy to see that $\mathscr{T}$ is a derivation. Applying the Sachs symmetry operation $(\ )^*$ to this operator and the operation $(\ )'$ to the resulting two operators, we obtain another three operators, the explicit forms of which are

$$\begin{aligned}\mathscr{T}'\eta &= |\psi|\{\tilde{\text{Þ}}'\eta + (1/m)(p\rho' + q\bar\rho')\eta\},\\ \mathscr{L}\eta &= |\psi|\{\tilde\eth\eta - (1/m)(p\tau - q\bar\tau')\eta\},\\ \mathscr{L}'\eta &= |\psi|\{\tilde\eth'\eta + (1/m)(p\tau' - q\bar\tau)\eta\}.\end{aligned} \tag{3.3}$$

By use of the Killing–Yano tensor equations (2.4), it turns out that the operators $\tilde{\text{Þ}}$, $\tilde\eth$ and their primed versions take a very condensed form if they are expressed in terms of the new operators $\mathscr{T}$, $\mathscr{T}'$, $\mathscr{L}$, and $\mathscr{L}'$. This is the key point for the following formalism. The result is

$$\tilde{\text{Þ}}\eta = |\psi|^{-1}\psi^{p/m}\bar\psi^{q/m}\mathscr{T}(\psi^{-p/m}\bar\psi^{-q/m}\eta), \tag{3.4}$$

$$\tilde{\text{Þ}}'\eta = |\psi|^{-1}\psi^{-p/m}\bar\psi^{-q/m}\mathscr{T}'(\psi^{p/m}\bar\psi^{q/m}\eta), \tag{3.5}$$

$$\tilde\eth\eta = |\psi|^{-1}\psi^{p/m}\bar\psi^{-q/m}\mathscr{L}(\psi^{-p/m}\bar\psi^{q/m}\eta), \tag{3.6}$$

$$\tilde\eth'\eta = |\psi|^{-1}\psi^{-p/m}\bar\psi^{q/m}\mathscr{L}'(\psi^{p/m}\bar\psi^{-q/m}\eta). \tag{3.7}$$

Obviously, Eqs. (3.5) and (3.7) are obtained from (3.4) and (3.6) under the operation prime. Henceforth, all equations have to be supplemented by their primed versions, which will not be written out explicitly. Inserting these expressions into the Dirac equation (3.1), we obtain

$$\begin{aligned}&(\mathscr{T} - ie|\psi|\phi_2)(\psi^{1/m}f) + i(\mathscr{L}' + ie|\psi|\phi_4')(\psi^{1/m}f')\\ &\quad = i\mu_e|\psi|\psi^{1/m}g,\\ &(\mathscr{T} - ie|\psi|\phi_2)(\bar\psi^{1/m}g') + i(\mathscr{L} + ie|\psi|\phi_4)(\bar\psi^{1/m}g)\\ &\quad = i\mu_e|\psi|\bar\psi^{1/m}f.\end{aligned} \tag{3.8}$$

The form of these equations suggests a separation ansatz for the functions $\psi^{1/m}f$, $\psi^{1/m}f'$, $\bar\psi^{1/m}g$, and $\bar\psi^{1/m}g'$. Although this ansatz can be made internally consistent, it leads to some inconvenient formal complications in the case $m = 2$ because of the transformation property of ψ under the prime operation. In particular, there is no natural way to assign one of the two branches of $\psi^{1/2}$ and $\psi'^{1/2}$ to each other because the operation prime is a discrete transformation. Although these difficulties could be circumvented by some additional conventions, the final equations become considerably simpler if one proceeds as follows. Select one special decomposition $\chi_{AB} = \psi^{(0)}o_{(A}{}^{(0)}\iota_{B)}{}^{(0)}$. Then, according to Sec. II, any decomposition satisfies either $\psi = \psi^{(0)}$, $D_{ab} = D_{ab}^{(0)}$ or $\psi = -\psi^{(0)}$, $D_{ab} = -D_{ab}{}^{(0)}$. Now suppose that, by definition, the operation prime acts on all quantities without the superscript $^{(0)}$ as usual, whereas the quantities with the superscript are invariant. Geometrically, this means that on the two-surface elements spanned by D_{ab} an orientation has been defined which is invariant under the prime operation, whereas the orientation induced by the four-dimensional orientation changes its sign under this operation. (The four-dimensional orientation itself is, of course, invariant.) The way in which this orientation is defined is arbitrary because a change of orientation results in a change of a common factor of the four functions which, of course, is irrelevant. Obviously, in any of the Eqs. (2.4) and (3.8), one has the choice to write ψ or $\psi^{(0)}$.

Equations (3.8) can now be separated as follows. We write

$$(\psi^{(0)})^{1/m} f = e^{i\kappa} Z W,$$
$$(\bar{\psi}^{(0)})^{1/m} g = a e^{i\kappa} Z W', \tag{3.9}$$

where, for $m = 2$ and $m = 1$, with ψ real we put $a = i$, and for $m = 1$, with ψ imaginary we put $a = 1$. Here, the function κ is an arbitrary linear combination with constant coefficients of the ignorable coordinates which are adapted to the Killing vector fields according to Table I. The ansatz (3.9) necessarily implies that the types of the functions Z and W are $(-\frac{1}{2},\frac{1}{2})$ and $(-\frac{1}{2},-\frac{1}{2})$ respectively. Note that the transformation property of Z' and W' is given by $Z'' = Z$ and $W'' = -W$, respectively.

Furthermore, the functions Z, Z', W, and W' are supposed to satisfy

$$\mathcal{T} Z = \mathcal{T}' Z = 0, \quad \mathcal{L} W = \mathcal{L}' W = 0. \tag{3.10}$$

In Sec. IV we shall show that, with respect to the gauge and the canonical coordinates of DR I and DR II, these differential equations are trivial in the sense that they can be fulfilled for arbitrary functions provided these depend on at most two of the coordinates. The details depend on the type of the Killing–Yano tensor and will be given in Sec. IV.

The conditions (3.10) allow the Dirac equation (3.8) to be separated. The electromagnetic field has to satisfy the condition that the components $|\psi|\phi_1$, $|\psi|\phi_1'$ and $|\psi|\phi_3$, $|\psi|\phi_3'$ depend on the same coordinates as the functions W, W' and Z, Z', respectively. Henceforth, the superscript$^{(0)}$ in the amplitude ψ will be omitted. Similarly to Chandrasekhar's work,[6] the four separation constants reduce to one single constant, called λ. The resulting equations, which are valid for any type of Killing–Yano tensor of valence 2, are

$$(\hat{\mathcal{T}} - ie|\psi|\phi_2) W - i(\lambda + i\mu_e \operatorname{Re}\psi) W' = 0,$$
$$(\hat{\mathcal{L}} + ie|\psi|\phi_4) Z - (\lambda - \mu_e \operatorname{Im}\psi) Z' = 0, \tag{3.11}$$
$$\lambda' = -\lambda.$$

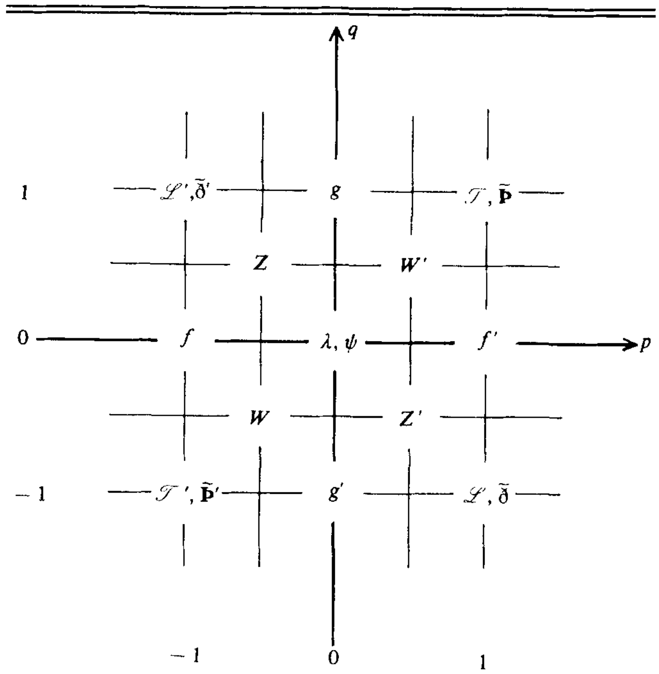
Here, the operators $\hat{\mathcal{T}}$ and $\hat{\mathcal{L}}$ are defined by

$$\hat{\mathcal{T}} = e^{-i\kappa} \mathcal{T} e^{i\kappa} \text{ and } \hat{\mathcal{L}} = e^{-i\kappa} \mathcal{L} e^{i\kappa}.$$

The GHP types of all quantities introduced in this paper are summarized in Table II.

No complete decoupling of these equations is possible, in general, i.e., under the sole assumption that a Killing–Yano tensor exists, except in case (4,I), which contains the types of metrics treated by Chandrasekhar[6] and Güven.[7]

TABLE II. Summary of GHP types.



IV. EXPLICIT FORM OF THE OPERATORS $\mathcal{T}, \mathcal{T}', \mathcal{L}, \mathcal{L}'$. THE SEPARABLE SYSTEMS

In this section we shall demonstrate the way in which the differential equations (3.10) can be integrated explicitly. Furthermore, the operators will be constructed in terms of the canonical coordinates. The subsequent calculations are greatly simplified by the following observation. If one introduces the boost weight $n = (p+q)/2$ and the spin weight $s = (p-q)/2$ instead of p and q, then Eqs. (3.2) and (3.3) take the following forms:

$$\mathcal{T} = |\psi| \{ D - 2n \operatorname{Re}(\epsilon - (1 - m^{-1})\rho) - 2is \times \operatorname{Im}(\epsilon - (1 - m^{-1})\rho) \},$$
$$\mathcal{L} = |\psi| \{ \delta - n(\beta - \bar{\beta}' - (1 - m^{-1})(\tau - \bar{\tau}')) - s(\beta + \bar{\beta}' - (1 - m^{-1})(\tau + \bar{\tau}')) \}. \tag{4.1}$$

It turns out that the integrability conditions of the Killing–Yano tensor equations (2.4), together with the gauge conditions used in DR I and DR II, imply that the coefficients of s in $\mathcal{T}$ (and $\mathcal{T}'$) and the coefficients of n in $\mathcal{L}$ (and $\mathcal{L}'$) vanish. For $m = 2$, the relevant equations are Eqs. (5.4) and (5.9) in DR I; for $m = 1$, the corresponding equations have been stated in Sec. 2 of DR II. Therefore, by use of the fact that the spin weight of W and W' and the boost weight of Z and Z' vanish, we deduce that Eqs. (3.10) reduce to

$$DZ = DZ' = 0, \quad \delta W = \delta W' = 0.$$

From the explicit form of the directional derivatives as given in Sec. II, and the following list of the functions W, W', Z, Z', it can be easily checked that these conditions are fulfilled.

Moreover, from Eqs. (4.1) the explicit form of the operators $\mathcal{T}, \mathcal{T}', \mathcal{L}, \mathcal{L}'$ will be obtained straightforwardly using the N.P. coefficients as given in DR I and DR II, which will not be repeated here. On inserting the operators in their subsequent explicit coordinate forms into Eqs. (3.11), one obtains the Dirac equation in its separated forms. In the following list, the quantities κ_y and κ_v are the separation constants associated with the Killing vector fields. Note that not all of the gauge conditions in DR I which have proven to be most convenient for constructing the coordinate systems are prime invariant. (However, the combination of conditions used to simplify the Expressions (4.1) are invariant under the prime.) Therefore, in the following, all of the operators $\mathscr{T}$, $\mathscr{T}'$, $\mathscr{L}$, $\mathscr{L}'$ will be stated explicitly.

Type (4,I) [$W = W(u)$, $W' = W'(u)$, $Z = Z(x)$, $Z' = Z'(x)$]:

$$\hat{\mathscr{T}} = 2^{-1/2}\{-A\partial_u + ivA^{-1}\kappa_y + ivu^2A^{-1}\kappa_v + n\,dA/du\},$$
$$\hat{\mathscr{T}}' = 2^{-1/2}\{vA\partial_u + 2i(1+v^2)^{-1}A^{-1}\kappa_y + 2i(1+v^2)^{-1}u^2A^{-1}\kappa_v + nv\,dA/du\},$$
$$\hat{\mathscr{L}} = 2^{-1/2}\{B\partial_x + B^{-1}\kappa_y - x^2B^{-1}\kappa_v - s\,dB/dx\},$$
$$\hat{\mathscr{L}}' = 2^{-1/2}\{B\partial_x - B^{-1}\kappa_y + x^2B^{-1}\kappa_v + s\,dB/dx\}.$$

Type (4,II) [$W = W(u)$, $W' = W'(u)$, $Z = Z(x,y)$, $Z' = Z'(x,y)$]:

$$\hat{\mathscr{T}} = 2^{-1/2}\{-A\partial_u + iv|\psi|^2A^{-1}\kappa_v + n\,dA/du\},$$
$$\hat{\mathscr{T}}' = 2^{-1/2}\{vA\partial_u + 2i(1+v^2)^{-1}|\psi|^2A^{-1}\kappa_v + nv\,dA/du\},$$
$$\hat{\mathscr{L}} = 2^{-1/2}(\partial_x C)^{-1/2}\{D^{-1}\partial_x - iD\partial_y - 2lCD\kappa_v - s(\partial_x C)^{-1/2}[\partial_x((\partial_x C)^{1/2}D^{-1}) - i\partial_y(D(\partial_x C)^{1/2})]\},$$
$$\hat{\mathscr{L}}' = 2^{-1/2}(\partial_x C)^{-1/2}\{D^{-1}\partial_x + iD\partial_y + 2lCD\kappa_v + s(\partial_x C)^{-1/2}[\partial_x((\partial_x C)^{1/2}D^{-1}) + i\partial_y(D(\partial_x C)^{1/2})]\}.$$

Type (4,III) [$W = W(u,v)$, $W' = W'(u,v)$, $Z = Z(x)$, $Z' = Z'(x)$]:

$$\hat{\mathscr{T}} = 2^{-1/2}(\partial_u C)^{-1/2}\{-D^{-1}\partial_u + n(\partial_u C)^{-1/2}\partial_u((\partial_u C)^{1/2}D^{-1})\},$$
$$\hat{\mathscr{T}}' = 2^{-1/2}(\partial_u C)^{-1/2}\{D\partial_v + 4ikCD\kappa_y + n(\partial_u C)^{-1/2}\partial_v(D(\partial_u C)^{1/2})\},$$
$$\hat{\mathscr{L}} = 2^{-1/2}\{B\partial_x - |\psi|^2B^{-1}\kappa_y - s\,dB/dx\},$$
$$\hat{\mathscr{L}}' = 2^{-1/2}\{B\partial_x + |\psi|^2B^{-1}\kappa_y + s\,dB/dx\}.$$

Type (4,IV) [$W = W(u,v)$, $W' = W'(u,v)$, $Z = Z(\zeta,\bar{\zeta})$, $Z' = Z'(\zeta,\bar{\zeta})$]:

$$\mathscr{T} = 2^{-1/2}E^{-1}\{-\partial_u + nE^{-1}\partial_u E\},$$
$$\mathscr{T}' = 2^{-1/2}E^{-1}\{\partial_v + nE^{-1}\partial_v E\},$$
$$\mathscr{L} = 2^{-1/2}F^{-1}\{\partial_{\bar{\zeta}} - sF^{-1}\partial_{\bar{\zeta}}F\},$$
$$\mathscr{L}' = 2^{-1/2}F^{-1}\{\partial_\zeta + sF^{-1}\partial_\zeta F\}.$$

Type (2,I) [$W = W(u,v)$, $W' = W'(u,v)$, $Z = Z(\zeta,\bar{\zeta})$, $Z' = Z'(\zeta,\bar{\zeta})$]:

$$\mathscr{T} = 2^{-1/2}\psi A^{-1}\{-\partial_u + nA^{-1}\partial_u A\},$$
$$\mathscr{T}' = 2^{-1/2}\psi A^{-1}\{\partial_v + nA^{-1}\partial_v A\},$$
$$\mathscr{L} = 2^{-1/2}F^{-1}\{\partial_{\bar{\zeta}} - sF^{-1}\partial_{\bar{\zeta}}F\},$$
$$\mathscr{L}' = 2^{-1/2}F^{-1}\{\partial_\zeta + sF^{-1}\partial_\zeta F\}.$$

Type (2,II) [$W = W(u,v)$, $W' = W'(u,v)$, $Z = Z(\zeta,\bar{\zeta})$, $Z' = Z'(\zeta,\bar{\zeta})$]:

$$\mathscr{T} = 2^{-1/2}C^{-1}\{-\partial_u - nC^{-1}\partial_v C\},$$
$$\mathscr{T}' = 2^{-1/2}C^{-1}\{\partial_v - nC^{-1}\partial_u C\},$$
$$\mathscr{L} = 2^{-1/2}(\psi/i)G^{-1}\{\partial_{\bar{\zeta}} - sG^{-1}\partial_{\bar{\zeta}}G\},$$
$$\mathscr{L}' = 2^{-1/2}(\psi/i)G^{-1}\{\partial_\zeta + sG^{-1}\partial_\zeta G\}.$$

V. CONCLUSION

It is instructive to compare the present formalism with the original work of Chandrasekhar.[6] The type D vacuum space-times considered by Güven[7] and, in particular, the Kerr metric, belong to the type (4,I) metrics of Sec. II. In Boyer–Lindquist coordinates, the Kerr metric will be obtained by the following definitions: $u = r$, $v = -t + a\phi$, $x = -a\cos\theta$, $y = -a^{-1}\phi$, $A = -\sqrt{\Delta}$, where $\Delta = r^2 + a^2 - 2Mr$, $B = a\sin\theta$, and $\nu = +1$. In this way, however, one does not obtain the Kinnersley tetrad,[20] which has been used by Chandrasekhar and which, in fact, destroys the prime invariance. Therefore, we shall apply the following spin transformation:

$$o^A \to \hat{o}^A = (2/\Delta)^{1/4}\psi^{1/2}o^A,$$
$$\iota^A \to \hat{\iota}^A = (\Delta/2)^{1/4}\psi^{-1/2}\iota^A,$$

which yields the new spinor components

$$\hat{f} = (\Delta/2)^{1/4}\psi^{-1/2}f,$$
$$\hat{f}' = (2/\Delta)^{1/4}\psi^{1/2}f',$$
$$\hat{g} = (2/\Delta)^{1/4}\bar{\psi}^{1/2}g,$$
$$\hat{g}' = (\Delta/2)^{1/4}\bar{\psi}^{-1/2}g'.$$

Applying the Chandrasekhar separation ansatz in his notation to these new quantities according to

$$\hat{f} = \psi^{-1}R_{-1/2}S_{-1/2},$$
$$\hat{f}' = -iR_{+1/2}S_{+1/2},$$
$$\hat{g} = R_{+1/2}S_{-1/2},$$
$$\hat{g}' = -i\bar{\psi}^{-1}R_{-1/2}S_{+1/2},$$

we find, apart from a common constant factor, the following relations between his functions $R_{\pm 1/2}$, $S_{\pm 1/2}$, and our functions W, W', Z, Z':

$$R_{+1/2} = 2^{-1/4}i(2/\Delta)^{1/4}W',$$
$$R_{-1/2} = 2^{-1/4}(\Delta/2)^{1/4}W,$$
$$S_{+1/2} = -2^{1/4}Z',$$
$$S_{-1/2} = -2^{1/4}Z.$$

Now it is straightforward to show that on inserting these relations into the differential equations (3.11), with $\phi = 0$ when specialized to the Kerr metric, Chandrasekhar's[6] equations (40) and (41) are reproduced.

To conclude, we mention two ways of possible generalizations of the foregoing formalism which might be of interest.

(i) The formalism appears to be sufficiently compact so that a generalization to massive fields with a spin different from $\frac{1}{2}$ might be worth considering.

(ii) In the U_4 theory of gravitation,[21] the Dirac equation plays a role analogous to the equation of geodesics in general relativity. So, one could speculate that an exterior field which allows the Dirac equation to be separated in this theory might be as interesting as the corresponding fields which allow the equation of geodesics in general relativity to be separated. The present paper could be considered as a preliminary step towards this end.

ACKNOWLEDGMENTS

The author wishes to thank Professor J. Ehlers and Dr. M. Walker for their interest in this work and for suggesting several improvements of the manuscript. The hospitality of the Max-Planck-Institute for Physics and Astrophysics, Garching, Federal Republic of Germany, is gratefully acknowledged. This work has been supported by the Deutsche Forschungsgemeinschaft.

[1]B. Carter, Phys. Rev. **174**, 1559 (1968).
[2]B. Carter, Commun. Math. Phys. **10**, 280 (1968).
[3]S. A. Teukolski, Phys. Rev. Lett. **29**, 114 (1972).
[4]S. A. Teukolski, Astrophys. J. **185**, 283 (1973).
[5]W. Unruh, Phys. Rev. Lett. **31**, 1265 (1973).
[6]S. Chandrasekhar, Proc. R. Soc. London Ser. A **349**, 571 (1976).
[7]R. Güven, Proc. R. Soc. London Ser. A **356**, 465 (1977).
[8]N. Toop, work cited in Refs. 9 and 11.
[9]D. N. Page, Phys. Rev. D **14**, 1509 (1976).
[10]M. Walker and R. Penrose, Commun. Math. Phys. **18**, 265 (1970).
[11]B. Carter and R. G. McLenaghan, Phys. Rev. D **19**, 1093 (1979).
[12]R. G. McLenaghan and Ph. Spindel, Phys. Rev. D **20**, 409 (1979).
[13]W. Dietz and R. Rüdiger, Proc. R. Soc. London Ser. A **375**, 361 (1981).
[14]W. Dietz and R. Rüdiger, Proc. R. Soc. London Ser. A **381**, 315 (1982).
[15]S. Benenti and M. Francaviglia, "The theory of separability of the Hamilton–Jacobi equation and its applications to General Relativity," in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. I, Chap. 13.
[16]E. Schrödinger, Proc. Roy. Irish. Acad. Sect. A **46**, 25 (1940).
[17]R. Geroch, A. Held, and R. Penrose, J. Math. Phys. **14**, 874 (1973).
[18]A. Held, Commun. Math. Phys. **37**, 311 (1974).
[19]J. Ehlers, Commun. Math. Phys. **37**, 327 (1974).
[20]W. Kinnersley, J. Math. Phys. **10**, 1195 (1969).
[21]F. W. Hehl in *Cosmology and Gravitation. Spin, Torsion, Rotation, and Supergravity*, edited by P. G. Bergmann and V. De Sabbata (Plenum, New York, 1980), pp. 5–61.

On the Weyl group, its little groups, and associated massless particle descriptions

V. Hussin[a] and J. Beckers
Physique théorique et mathématique, Université de Liège, Institut de Physique au Sart Tilman, Bâtiment B.5, B-4000 Liège 1, Belgium

(Received 8 December 1982; accepted for publication 17 March 1983)

The Weyl or similitude group of space-time is studied in order to get invariant (free) wave equations for massless particles. We deduce the extended Wigner little groups of the Weyl group and define new operators related to the generators of these little groups in order to get wave equations for arbitrary (but nonzero) helicity particles. The Maxwell theory is also discussed through the action of our operators on electromagnetic tensors.

PACS numbers: 11.10.Qr, 02.20. + b

1. INTRODUCTION

Scale transformations (or dilatations) combined with inhomogeneous restricted Lorentz (or Poincaré) transformations form a well-known structure—the so-called Weyl group or similitude group SIM(3,1) of space-time—which is very often regarded as a (maximal) subgroup of the conformal group of space-time. Mathematically the subgroup or subalgebraic contents of SIM(3,1) has been studied by Patera, Winternitz, and Zassenhaus[1] and, physically, such a group spans an important part of the description of free massless particles due to the impact of scale transformations. In fact, we recently discussed[2] "extended Wigner little groups" of the Weyl group. We also proposed[2] exploiting such considerations with the aim of deriving wave equations for massless particles and, in this way, relating recent works which start from different points of view.[3–6] This last point is the purpose of our paper. First we want to recall (Sec. 2) some basic features about the Weyl group and its algebra and to add a few points about transformation laws (under the Weyl group) of second-rank tensors. Then the discussion of "extended little groups" will lead to interesting new operators (Sec. 3) related to physical quantities like four-momenta and helicity for example. In Sec. 4, we will get simple wave equations for massless particles and will establish the connection between our group theoretical point of view and the results obtained by Bacry,[3] Stepanovskii,[4] and Bracken and Jessup.[6] Finally, we will exploit the action of the new operators on electromagnetic tensors (Sec. 5) in order to get a necessary and sufficient condition in connection with the Maxwell theory.

2. THE WEYL GROUP, ITS LIE ALGEBRA, AND SOME TRANSFORMATION LAWS

In Minkowski space-time characterized by the metric tensor $G_M \equiv \{g^{\mu\nu}\,(\mu,\nu=0,1,2,3)\} = \mathrm{diag}(1,-1,-1,-1)$, the events $x \equiv (x^\mu) = \{x^0, x^i\,(i=1,2,3)\} \equiv (t,\mathbf{x})$ can be subjected to Poincaré (or inhomogeneous restricted Lorentz) transformations—i.e., space-time translations (α^μ) and restricted homogeneous Lorentz transformations $(\Lambda^\mu{}_\nu, \Lambda^0{}_0 \geqslant 1, \det\Lambda = 1)$—*and* to dilatations (σ) (or scale transformations), so that

$$x'^\mu = \sigma\Lambda^\mu{}_\nu x^\nu + \alpha^\mu, \quad \mu = 0, 1, 2, 3. \tag{2.1}$$

The set of $(\alpha, \Lambda, \sigma)$-transformations form an *eleven* parameter Lie group called the Weyl group or SIM(3,1), the similitude group of space-time. Its multiplication law is given by

$$(\alpha', \Lambda', \sigma')(\alpha, \Lambda, \sigma) = (\alpha' + \sigma'\Lambda'\alpha, \Lambda'\Lambda, \sigma'\sigma). \tag{2.2}$$

The Lie algebra of the Weyl group is generated by the eleven generators $\{P^\mu, M^{\mu\nu}, D\}$ associated with *infinitesimal* space-time translations, restricted homogeneous Lorentz transformations and dilatations characterized by infinitesimal parameters a^μ, $\omega^{\mu\nu} = -\omega^{\nu\mu}$, and ρ, respectively, corresponding to Eq. (2.1), in the following form:

$$x'^\mu = x^\mu + \rho x^\mu + \omega^\mu{}_\nu x^\nu + a^\mu = x^\mu - \xi^\mu. \tag{2.3}$$

The commutation relations of this Lie algebra read

$$\begin{aligned} &[M^{\mu\nu}, M^{\rho\sigma}] = i(g^{\mu\sigma}M^{\nu\rho} + g^{\nu\rho}M^{\mu\sigma} - g^{\mu\rho}M^{\nu\sigma} - g^{\nu\sigma}M^{\mu\rho}),\\ &[M^{\mu\nu}, P^\rho] = i(g^{\nu\rho}P^\mu - g^{\mu\rho}P^\nu),\\ &[P^\mu, D] = iP^\mu, \quad [P^\mu, P^\nu] = 0. \end{aligned} \tag{2.4}$$

This is a (maximal) subalgebra of the conformal algebra already studied among authors by Mack and Salam.[7] Following their conventions, a realization of the Weyl generators is given by

$$\begin{aligned} &P^\mu = i\partial^\mu = i\frac{\partial}{\partial x_\mu}, \quad M^{\mu\nu} = i(x^\mu\partial^\nu - x^\nu\partial^\mu) + \Sigma^{\mu\nu},\\ &D = ix^\mu\partial_\mu + \Delta, \end{aligned} \tag{2.5}$$

where the operators $\Sigma^{\mu\nu}$ and Δ refer to the internal structure of the Weyl context or, following Mack and Salam,[7] to the corresponding stability subgroup of $x = 0$. The Lie algebra of this subgroup is isomorphic to $\mathrm{so}(3,1) \otimes \{D\}$ and its commutation relations are

$$\begin{aligned} &[\Sigma^{\mu\nu}, \Sigma^{\rho\sigma}] = i(g^{\mu\sigma}\Sigma^{\nu\rho} + g^{\nu\rho}\Sigma^{\mu\sigma} - g^{\mu\rho}\Sigma^{\nu\sigma} - g^{\nu\sigma}\Sigma^{\mu\rho}),\\ &[\Sigma^{\mu\nu}, \Delta] = 0, \end{aligned} \tag{2.6}$$

leading to the three usual Casimir operators

$$C_1 = \tfrac{1}{2}\Sigma^{\mu\nu}\Sigma_{\mu\nu}, \quad C_2 = -\tfrac{1}{4}i\tilde{\Sigma}^{\mu\nu}\Sigma_{\mu\nu}, \quad \Delta, \tag{2.7}$$

where $\tilde{\Sigma}$ is the dual of Σ:

$$\tilde{\Sigma}^{\mu\nu} = -\tfrac{1}{2}\epsilon^{\mu\nu}{}_{\rho\sigma}\Sigma^{\rho\sigma}, \quad \epsilon^{0123} = 1. \tag{2.8}$$

[a] Boursier I.R.S.I.A. 1982–1983.

Let us also recall[8] that, in the context of so(3,1), the invariants C_1 and C_2 take very simple values multiplying the identity operator when finite irreducible (m, n)-representations, $m, n \geqslant 0$ are considered. In this case, we have indeed

$$C_1 = [2m(m+1) + 2n(n+1)]\mathbb{1},\qquad C_2 = [m(m+1) - n(n+1)]\mathbb{1}. \tag{2.9}$$

Furthermore, it can easily be shown from (2.6)–(2.8) that

$$\Sigma_{\alpha\mu}\widetilde{\Sigma}^{\mu\nu} = \widetilde{\Sigma}_{\alpha\mu}\Sigma^{\mu\nu} = i\widetilde{\Sigma}_\alpha{}^\nu - iC_2\delta_\alpha{}^\nu, \tag{2.10}$$

a relation which will be useful in the following.

Under arbitrary coordinate transformations, a *covariant* second-rank tensor $T \equiv \{T_{\mu\nu}(x)\}$ becomes

$$T'_{\mu\nu}(x') = \frac{\partial x^\alpha}{\partial x'^\mu}\frac{\partial x^\beta}{\partial x'^\nu}T_{\alpha\beta}(x). \tag{2.11}$$

When the infinitesimal transformations (2.3) are considered, we explicitly get (at the same point x)

$$T'_{\mu\nu}(x) = (1-2\rho)T_{\mu\nu}(x) - (\rho x^\alpha + \omega^\alpha{}_\beta x^\beta + a^\alpha)\partial_\alpha T_{\mu\nu}(x) + \omega_\mu{}^\alpha T_{\alpha\nu}(x) + \omega_\nu{}^\alpha T_{\mu\alpha}(x) \tag{2.12}$$

or

$$T'_{\mu\nu}(x) = T_{\mu\nu}(x) + \xi^\alpha\partial_\alpha T_{\mu\nu}(x) + (\partial_\mu\xi^\alpha)T_{\alpha\nu}(x) + (\partial_\nu\xi^\alpha)T_{\mu\alpha}(x). \tag{2.13}$$

In terms of Lie derivatives,[9] these relations are nothing but

$$T'_{\mu\nu}(x) = T_{\mu\nu}(x) + L_X T_{\mu\nu}(x), \tag{2.14}$$

where the vector fields are given by $X \equiv \xi^\alpha\partial_\alpha$ as usual. If invariance conditions on T are required, we recover[10] the annulation of its Lie derivative with respect to the vector fields X.

Now, let us rewrite Eq. (2.13) with the Weyl generators, i.e.,

$$T'_{\mu\nu}(x) = T_{\mu\nu}(x) + ia_\alpha(P^\alpha T)_{\mu\nu}(x) + \tfrac{1}{2}i\omega_{\beta\alpha}(M^{\alpha\beta}T)_{\mu\nu}(x) + i\rho(DT)_{\mu\nu}(x), \tag{2.15}$$

which emphasizes the action of those generators on the tensor T by

$$(P^\alpha T)_{\mu\nu}(x) = i\partial^\alpha T_{\mu\nu}(x), \tag{2.16}$$

$$(M^{\alpha\beta}T)_{\mu\nu}(x) = i(x^\alpha\partial^\beta - x^\beta\partial^\alpha)T_{\mu\nu}(x) + (\Sigma^{\alpha\beta}T)_{\mu\nu}(x), \tag{2.17}$$

and

$$(DT)_{\mu\nu}(x) = ix^\alpha\partial_\alpha T_{\mu\nu}(x) + (\Delta T)_{\mu\nu}(x), \tag{2.18}$$

where

$$(\Sigma^{\alpha\beta}T)_{\mu\nu}(x) = i(g^\alpha{}_\mu T^\beta{}_\nu(x) - g_\nu{}^\beta T_\mu{}^\alpha(x) + g_\nu{}^\alpha T_\mu{}^\beta(x) - g^\beta{}_\mu T^\alpha{}_\nu(x)) \tag{2.19}$$

and

$$(\Delta T)_{\mu\nu}(x) = 2iT_{\mu\nu}(x). \tag{2.20}$$

The point of view expressed by the relations (2.15)–(2.20) has recently been developed in the Poincaré context.[5]

Parallel transformation laws can evidently be obtained for a *contravariant* tensor $T \equiv \{T^{\mu\nu}(x)\}$, so that in correspondence with Eq. (2.12) we get

$$T'^{\mu\nu}(x) = (1+2\rho)T^{\mu\nu}(x) - (\rho x^\alpha + \omega^\alpha{}_\beta x^\beta + a^\alpha)\partial_\alpha T^{\mu\nu}(x) + \omega^\mu{}_\alpha T^{\alpha\nu}(x) + \omega^\nu{}_\alpha T^{\mu\alpha}(x) \tag{2.21}$$

or

$$T'^{\mu\nu}(x) = T^{\mu\nu}(x) + \xi^\alpha\partial_\alpha T^{\mu\nu}(x) - (\partial_\alpha\xi^\mu)T^{\alpha\nu}(x) - (\partial_\alpha\xi^\nu)T^{\mu\alpha}(x) \tag{2.22}$$

$$= T^{\mu\nu}(x) + L_X T^{\mu\nu}(x). \tag{2.23}$$

The corresponding action of the Weyl generators on T is similar to Eqs. (2.16)–(2.20). In fact, apart from raising the indices μ, ν, the only modification lies in Eq. (2.20) which becomes

$$(\Delta T)^{\mu\nu}(x) = -2iT^{\mu\nu}(x). \tag{2.24}$$

In particular, if T is the *metric* tensor G_M, we immediately get

$$g'_{\mu\nu} = (1-2\rho)g_{\mu\nu},\quad g'^{\mu\nu} = (1+2\rho)g^{\mu\nu}. \tag{2.25}$$

Such relations express the well-known fact that G_M is not scale invariant while being invariant under the Poincaré transformations. This implies that contravariant and covariant components of space-time events transform according to (2.3) *and*

$$x'_\mu = x_\mu - \rho x_\mu + \omega_\mu{}^\nu x_\nu + a_\mu, \tag{2.26}$$

respectively.

3. EXTENDED LITTLE GROUPS AND THEIR GENERATORS

In the Weyl context, four-momenta and their contra- or covariant components have to transform according to (2.3) or (2.26), so that we have

$$p'^\mu = p^\mu + \rho p^\mu + \omega^\mu{}_\nu p^\nu \tag{3.1}$$

or

$$p'_\mu = p_\mu - \rho p_\mu + \omega_\mu{}^\nu p_\nu. \tag{3.2}$$

In a quantum mechanical context, a consistent set of transformation laws is evidently given by (2.3) and (3.2), remembering that coordinates and derivatives transform oppositely. So, from (3.2), the invariance conditions on p under the Weyl group read

$$\omega_\mu{}^\nu p_\nu = \rho p_\mu. \tag{3.3}$$

These four conditions are meaningful and can lead to the little group of p inside the Weyl structure, our so-called "extended little group."

In order to put the analysis of extended little groups in correspondence with Wigner's well-known results,[11] let us first recall some Poincaré considerations (Sec. 3A) and then develop our Weyl comments (Sec. 3B).

A. Poincaré case

If $\rho = 0$, Eq. (3.3) reduces to the conditions leading to the definition[11] of the so-called "Wigner little groups of p" and to the associated Pauli–Lubanski or Bargmann–Wigner operators[12]:

$$W^\mu = -\tfrac{1}{2}\epsilon^\mu{}_{\nu\rho\sigma}P^\nu M^{\rho\sigma} = \widetilde{\Sigma}^{\mu\nu}P_\nu. \tag{3.4}$$

If as usual we refer to spatial rotations (**J**) and boosts (**K**) through the definitions

$$M^{oi} = K^i \quad \text{and} \quad J^i = \tfrac{1}{2}\epsilon^i_{\ jk} M^{jk}, \quad \epsilon^{123} = 1, \tag{3.5}$$

the four components of W are given by

$$W^0 = \mathbf{P}\cdot\mathbf{J}, \quad \mathbf{W} = P^0\mathbf{J} - \mathbf{P}\wedge\mathbf{K}. \tag{3.6}$$

These four operators do form a four-vector

$$[M^{\mu\nu}, W^\rho] = i(g^{\nu\rho}W^\mu - g^{\mu\rho}W^\nu), \tag{3.7}$$

and only three of them are independent,

$$W^\mu P_\mu = 0. \tag{3.8}$$

In fact, the other interesting commutation relations are

$$[W^\mu, W^\nu] = -i\epsilon^{\mu\nu}_{\ \ \rho\sigma} W^\rho P^\sigma, \quad [W^\mu, P^\rho] = 0. \tag{3.9}$$

Finally, let us remember that there are three nonisomorphic little groups of p according to the time-, light-, or spacelike character of four-momenta, i.e., SO(3), E(2), or SO(2,1), respectively. Moreover, in the lightlike case, we know that

$$W^\mu = \lambda P^\mu, \tag{3.10}$$

where λ is the eigenvalue of the invariant helicity operator.

B. Weyl case

Let us exploit the invariance conditions (3.3).

(i) It is straightforward to notice that the conditions (3.3) imply that the four-momentum has to be a lightlike four-vector

$$p^\mu p_\mu = 0, \tag{3.11}$$

which shows that we are dealing here only with massless particles. A representative example of such a class of lightlike four-vectors is

$$\hat{p} \equiv \{\hat{p}^\mu\} = (|\mathbf{p}|, 0, 0, |\mathbf{p}|). \tag{3.12}$$

(ii) The extended little group dimension is four, i.e., among the seven parameters $\omega^{\mu\nu}$ and ρ, only four are linearly independent. In order to establish this property, let us parametrize Lorentz transformations according to the definitions (3.5) in terms of rotational parameters $\boldsymbol{\theta}$ and boost parameters $\boldsymbol{\phi}$:

$$\omega^{oi} = \phi^i \quad \text{and} \quad \theta^i = \tfrac{1}{2}\epsilon^i_{\ jk}\omega^{jk}. \tag{3.13}$$

Then, Eq. (3.3) becomes

$$\boldsymbol{\phi}\cdot\mathbf{p} + \rho p^0 = 0, \quad \boldsymbol{\phi} p^0 - \boldsymbol{\theta}\wedge\mathbf{p} + \rho\mathbf{p} = 0, \tag{3.14}$$

where the first relation is easily recovered by taking the scalar product of the second one with $\mathbf{p}(p^0)^{-1}$.

(iii) The generators of the extended little group can be explicitly determined. In correspondence with the four linearly independent parameters, we get the four operators

$$G' = \mathbf{P}\cdot\mathbf{J} = W^0, \quad \mathbf{G} = \mathbf{J}\wedge\mathbf{P} - \mathbf{K}P^0 + D\mathbf{P}. \tag{3.15}$$

They generate the Lie algebra

$$\begin{aligned}[G^i, G^j] &= i\epsilon^{ij}_{\ \ k}(\mathbf{P}\wedge\mathbf{G})^k, \\ [G', \mathbf{G}] &= -i\mathbf{P}\wedge\mathbf{G}.\end{aligned} \tag{3.16}$$

If the eigenvalues of the translation operators are given by $\hat{p}\equiv$(3.12), we finally get the generators of the extended little group of $\hat{p}$:

$$\{J^3, A^1 = K^1 - J^2, A^2 = K^2 + J^1, K^3 - D\}, \tag{3.17}$$

with the commutation relations

$$[J^3, A^1] = iA^2, \quad [J^3, A^2] = -iA^1, \quad [A^1, A^2] = 0, \tag{3.18}$$

$$[K^3 - D, A^1] = iA^1, \quad [K^3 - D, A^2] = iA^2. \tag{3.19}$$

Obviously, it represents a subgroup of the *homogeneous* part of the Weyl group and it contains the E(2)-Wigner little group generated by (J^3, A^1, A^2) as a subgroup [cf. (3.18)] as expected. All the extended Wigner little groups of arbitrary lightlike four-momenta will be isomorphic to (3.17) and their algebras to Eqs. (3.18) and (3.19).

(iv) The operators (3.15) do *not* form a four-vector. This property is straightforward because $G' \equiv W^0 = \mathbf{P}\cdot\mathbf{J}$ but $\mathbf{G}\neq\mathbf{W}$. Here we want to exploit the three-vector character of the G^i's in order to find a *four-vector* V with $V^i = G^i$ ($i = 1,2,3$). This four-vector must obey the commutation relations

$$[M^{\mu\nu}, V^\rho] = i(g^{\nu\rho}V^\mu - g^{\mu\rho}V^\nu), \tag{3.20}$$

from which we get

$$V^0 = DP^0 - \mathbf{K}\cdot\mathbf{P}, \quad \mathbf{V} = \mathbf{G} \tag{3.21}$$

or, in a contravariant notation,

$$V^\mu = M^{\mu\nu}P_\nu + DP^\mu. \tag{3.22}$$

Using the commutation relations (2.4), we also obtain

$$V^\mu = P_\nu M^{\mu\nu} + P^\mu D + 2iP^\mu. \tag{3.23}$$

Then, with the realization (2.5), the operators (3.22) or (3.23) are given by

$$V^\mu = \Sigma^{\mu\nu}P_\nu + \Delta P^\mu. \tag{3.24}$$

In fact, such operators V^μ correspond to a *particular* parametrization which is a solution of Eq. (3.3). It is given by

$$\omega_{\mu\nu} = n_\mu p_\nu - n_\nu p_\mu, \quad \rho = -n^\mu p_\mu, \tag{3.25}$$

where the n_μ's are arbitrary parameters, three of them being independent since

$$V^\mu P_\mu = 0. \tag{3.26}$$

Together with (3.20), we also have

$$\begin{aligned}&[P^\mu, V^\nu] = 0, \\ &[D, V^\mu] = -iV^\mu, \\ &[V^\mu, V^\nu] = i(V^\mu P^\nu - V^\nu P^\mu).\end{aligned} \tag{3.27}$$

So now we easily see that the four operators G' and $\mathbf{G}$ are in fact constructed from some components of the *two* four-vectors V and W.

Let us also mention that with the choice (3.12), the operators V^μ reduce to the three generators A^1, A^2, and $K^3 - D$ which form a closed structure having the nonvanishing commutators (3.19), a substructure of our extended little algebra.

As a last result, let us exploit the property (3.10) of W in order to find a similar property for V. In fact, from the lightlike character of V and the orthogonality condition (3.26), we learn[13] that V and P are collinear. Consequently we can set

$$V^\mu = \eta P^\mu, \tag{3.28}$$

where the parameter η is determined as follows. From Eqs. (3.4), (3.10), and (2.10), we get

$$\begin{aligned}\lambda\Sigma_{\alpha\mu}P^\mu = \Sigma_{\alpha\mu}W^\mu &= \Sigma_{\alpha\mu}\tilde{\Sigma}^{\mu\nu}P_\nu \\ &= i(\tilde{\Sigma}_\alpha^{\ \nu} - C_2\delta_\alpha^{\ \nu})P_\nu.\end{aligned} \tag{3.29}$$

With (3.24), we finally obtain

$$V^\mu = i(\mathbb{1} - C_2\lambda^{-1} - i\Delta)P^\mu, \quad \lambda \neq 0, \tag{3.30}$$

which fixes the value of η.

4. MASSLESS PARTICLES AND WAVE EQUATIONS

Since here the lightlike character of four-momenta has been explicit from the start [see property 3B(i)], our considerations apply only to the description of massless particles. Now, by using (nonunitary) finite-dimensional (m,n)-representations of the Lorentz group, we can get *ad hoc* wave equations describing particles of nonzero helicity λ. Let us recall, following Weinberg,[14] that the helicity is given in terms of m and n by

$$\lambda = m - n, \quad \lambda \neq 0, \tag{4.1}$$

so that, in particular, the Casimir invariant C_2 takes the form

$$C_2 = \lambda(m + n + 1)\mathbb{1}, \tag{4.2}$$

a useful result for our further developments.

From our group theoretical considerations, let us apply the operators V_μ to a finite-component wave function ψ and use the results (3.24) and (3.30). We have

$$\Sigma^{\mu\nu}P_\nu\psi = i(\mathbb{1} - C_2\lambda^{-1})P^\mu\psi \tag{4.3}$$

or, with (4.2),

$$\Sigma^{\mu\nu}P_\nu\psi = -i(m+n)P^\mu\psi. \tag{4.4}$$

This equation has been derived by Stepanovskii.[4] Here its origin may be traced back via our Eqs. (3.28) and (3.30). In the Poincaré context and via Eq. (3.10), we get in a similar way

$$\widetilde{\Sigma}^{\mu\nu}P_\nu\psi = (m-n)P^\mu\psi, \tag{4.5}$$

a set of equations which is the one given by Bacry.[3]

Now, with the specific choice

$$C_2\lambda^{-1} = -i\Delta \tag{4.6}$$

in Eq. (3.30), we get

$$V^\mu = iP^\mu \tag{4.7}$$

and the wave equation

$$\Sigma^{\mu\nu}P_\nu\psi = (i - \Delta)P^\mu\psi, \tag{4.8}$$

a result obtained by Bracken and Jessup[6] in the discussion of local conformal invariance. Let us finally notice that the choice (4.6) corresponds with a (m, n)-representation to

$$\Delta = i(m+n+1)\mathbb{1} = -il\mathbb{1}, \tag{4.9}$$

where l is the scale dimension of the field under consideration, as discussed by Mack and Salam.[7] Such representations are of special interest as also discussed by Bracken and Jessup.[6]

Explicit examples for arbitrary nonzero helicity particles can easily be given. Among these, the cases of neutrinos or photons are the physically interesting ones. Let us only mention that in the neutrino case we are dealing with the $(0, \frac{1}{2})$-representation which in terms of the Pauli matrices can be realized by

$$(\Sigma^{\mu\nu}) \equiv (\Sigma^{oi}, \Sigma^{ij}) = (-i\boldsymbol{\sigma}/2, \boldsymbol{\sigma}/2). \tag{4.10}$$

Then, Eqs. (4.4) and (4.5) are, respectively,

$$\Sigma^{\mu\nu}P_\nu\psi = -\tfrac{1}{2}iP^\mu\psi, \quad \widetilde{\Sigma}^{\mu\nu}P_\nu\psi = -\tfrac{1}{2}P^\mu\psi. \tag{4.11}$$

These sets are obviously equivalent due to the fact that $\widetilde{\Sigma}^{\mu\nu} = -i\Sigma^{\mu\nu}$. They give the well-known Weyl equation and, with the choice (4.6), we recover

$$\Delta = \tfrac{3}{2}i \Rightarrow l = -\tfrac{3}{2}, \tag{4.12}$$

the scale dimension of the neutrino field. The corresponding remarks in the photon context are straightforward after noticing that

$$(\Sigma^{\mu\nu}) \equiv (i\mathbf{S}, \mathbf{S}), \quad \widetilde{\Sigma}^{\mu\nu} = i\Sigma^{\mu\nu}, \quad \Delta = 2i, \quad l = -2,$$

where the S-matrices are associated with the (1, 0)-representation of the homogeneous Lorentz group.

5. ACTION OF WEYL OPERATORS ON ELECTROMAGNETIC TENSORS

Let us now discuss the Maxwell theory in the framework of the Weyl context by considering the action of our operators V^μ defined by Eq. (3.23) using, for example, the realization (2.5). Such a discussion adopts a point of view similar to the one developed recently[5] for helicity operators for integral spin fields.

From Sec. 2, we know the action of the Weyl generators on an arbitrary second-rank tensor T. If we consider the covariant skew-symmetric Maxwell tensor (two-form) $F \equiv \{F_{\alpha\beta}(x)\}$, the action of $V^\mu \equiv (3.23)$ on F yields

$$\begin{aligned}(V^\mu F)_{\alpha\beta}(x) = P_\nu(M^{\mu\nu}F)_{\alpha\beta}(x) + P^\mu(DF)_{\alpha\beta}(x) \\ + 2iP^\mu F_{\alpha\beta}(x),\end{aligned} \tag{5.1}$$

or, with (2.16)–(2.18),

$$(V^\mu F)_{\alpha\beta}(x) = P_\nu(\Sigma^{\mu\nu}F)_{\alpha\beta}(x) + P^\mu(\Delta F)_{\alpha\beta}(x). \tag{5.2}$$

With Eqs. (2.19) and (2.20), we finally get

$$\begin{aligned}(V^\mu F)_{\alpha\beta}(x) = g^\mu{}_\beta\partial_\nu F^\nu{}_\alpha(x) - g^\mu{}_\alpha\partial_\nu F^\nu{}_\beta(x) \\ + \partial_\alpha F^\mu{}_\beta(x) - \partial_\beta F^\mu{}_\alpha(x) - 2\partial^\mu F_{\alpha\beta}(x).\end{aligned} \tag{5.3}$$

In such a context, the operator equation (4.7) plays a particularly prominent role. Indeed we shall prove the following statement:

"the action on F of Eq. (4.7) gives a necessary and sufficient condition for obtaining the free Maxwell theory."

Let us show that

$$(V^\mu F)_{\alpha\beta}(x) = i(P^\mu F)_{\alpha\beta}(x) \tag{5.4}$$

is such a necessary and sufficient condition. If the Maxwell equations are satisfied, we have with $F_{\alpha\beta} = -F_{\beta\alpha}$,

$$\partial_\mu F^\mu{}_\alpha = 0, \quad \partial_\alpha F^\mu{}_\beta + \partial_\beta F_\alpha{}^\mu = \partial^\mu F_{\alpha\beta}, \tag{5.5}$$

so that Eq. (5.3) gives

$$(V^\mu F)_{\alpha\beta}(x) = -\partial^\mu F_{\alpha\beta}(x) = i(P^\mu F)_{\alpha\beta}(x).$$

Conversely, if Eq. (5.4) is satisfied, we get through (5.3),

$$g^\mu{}_\beta\partial_\nu F^\nu{}_\alpha - g^\mu{}_\alpha\partial_\nu F^\nu{}_\beta = \partial^\mu F_{\alpha\beta} + \partial_\alpha F_\beta{}^\mu + \partial_\beta F^\mu{}_\alpha. \tag{5.6}$$

This set is valid for arbitrary μ, α, β so that we easily recover the Maxwell equations (5.5) in agreement with the above statement.

[1]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1615 (1975).
[2]V. Hussin, Lett. Nuovo Cimento **35**, 305 (1982).
[3]H. Bacry, Nuovo Cimento A **32**, 448 (1976).
[4]Yu. P. Stepanovskii, Theor. Math. Phys. **47**, 501 (1981).
[5]I. Bialynicki-Birula, E. T. Newman, J. Porter, J. Winicour, B. Lukas, Z. Perjes, and A. Sebestyen, J. Math. Phys. **22**, 2530 (1981).
[6]A. Bracken and B. Jessup, J. Math. Phys. **23**, 1925 (1982).
[7]G. Mack and A. Salam, Ann. Phys. (N.Y.) **53**, 174 (1969).
[8]I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon, London, 1963).
[9]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963).
[10]J. Beckers, J. Harnad, M. Perroud, and P. Winternitz, J. Math. Phys. **19**, 2126 (1978).
[11]E. P. Wigner, Ann. Math. **40**, 149 (1939); S. Gasiorowicz, *Elementary Particle Physics* (Wiley, New York, 1966), Chap. 4.
[12]J. K. Lubanski, Physica **8**, 44 (1941); V. Bargmann and E. P. Wigner, Proc. Natl. Acad. USA **34**, 211 (1948).
[13]H. Bacry, *Leçons sur la théorie des groupes et les symétries des particules élémentaires* (Gordon and Breach, Paris, 1967).
[14]S. Weinberg, Phys. Rev. **134**, B882 (1964).

Bifurcation and the magnetic multipole solutions of the Yang–Mills equations with sources

C. H. Oh[a)]
School of Physics, University of Science of Malaysia, Penang, Malaysia

(Received 18 January 1983; accepted for publication 18 February 1983)

It is shown that any Sikivie–Weiss-type magnetic multipole solution possesses a bifurcation point and it has the corresponding abelian Coulomb solution as its partner branch, which is unstable. The magnetic multipole solution is the stable branch.

PACS numbers: 11.15. – q, 11.10.Lm

I. INTRODUCTION

Some time ago Mandula[1] showed that the abelian Coulomb solution of the Yang–Mills (YM) equations is stable only if the external source strength is weak. Subsequently Sikivie and Weiss[2] constructed a total screening solution and magnetic dipole solution, both possessing less energy than the corresponding abelian Coulomb solution. Since then, many other solutions of the YM equations have been discovered.[3,8] Particularly, Jackiw *et al.*[4] found that there are bifurcating solutions when the external source strength is large enough.

Since the abelian Coulomb solution becomes unstable when the external source strength is large[1] and as bifurcation is related to instability, one is led to study whether there are solutions which bifurcate from the Coulomb solution. Indeed it is shown in Ref. 6 that a special magnetic multipole solution bifurcates from the abelian Coulomb solution. We now wish to demonstrate generally that all magnetic multipole solutions[2,5,6] possess bifurcation points and their respective accompanying bifurcation branch with higher energy is the abelian Coulomb solution. This is performed in Sec. II. In Sec. III, explicit expressions for the critical charge and various terms of the total energy are written down in the simple case of the magnetic dipole solution. We end with some remarks in Sec. IV.

II. BIFURCATION

The SU(2) YM equations in the presence of an external source are

$$(D_\mu F^{\mu\nu})_a = j_a^\nu = \delta_0^\nu \rho_a, \tag{1a}$$

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g\epsilon^{abc} A_\mu^b A_\nu^c, \tag{1b}$$

and our metric is $g_{ii} = -g_{00} = 1$. Here ρ_a is the external charge density and g is the coupling constant. Following Sikivie and Weiss,[2] we substitute the following ansatz into the YM equations:

$$A_0^a = \delta_3^a \phi(\rho,x_3)/g, \tag{2a}$$

$$A_i^a = \delta_1^a \epsilon_{i3j}(x_j/g\rho) A(\rho,x_3), \tag{2b}$$

$$\rho^a = \delta_3^a q/g, \tag{2c}$$

where $\rho^2 = x_1^2 + x_2^2$, and obtain two coupled nonlinear equations for the function $\phi(\rho,x_3)$ and $A(\rho,x_3)$,

$$-\nabla^2\phi + \phi A^2 = q, \tag{3a}$$

$$\nabla^2 A - A/\rho^2 + \phi^2 A^2 = 0. \tag{3b}$$

[a)] Present address: Physics Dept., National University of Singapore, Singapore.

Magnetic multipole solutions will result if we put[2,5,6]

$$A(\rho,x_3) = \frac{c}{a}\frac{P_n^1(\cos\theta)}{y^{n+1}} f(y,\theta) = \frac{c}{a}\bar{A}(y,\theta),$$
$$y = r/a, \quad r^2 = x_3^2 + \rho^2. \tag{4}$$

Here c is the norm of $A(\rho,x_3)$, a indicates the size of the external charge distribution, and $P_n^1(\cos\theta)$ is the Legendre polynomial. To ensure finite energy, the function $f(y,\theta)$ must tend to one sufficiently fast as $y \to \infty$ and tend to zero as $y \to 0$. The magnetic multipole solution in fact interpolates between the following two trivial solutions:

$$f(y,\theta) = 0, \quad \phi(\rho,x_3) = 0 \quad \text{for all } x_i, \tag{5a}$$

and

$$f(y,\theta) = 1, \quad \phi(\rho,x_3) = 0 \quad \text{for all } x_i, \tag{5b}$$

and is a solution with critical source strength.

The energy of the magnetic mulitpole solution can be written as

$$H = \frac{1}{2}\int d^3x (E_i^{a^2} + B_i^{a^2})$$
$$= \int d^3x \left[\frac{1}{2}(\nabla\phi)^2 + \phi^2 A^2\right]. \tag{6}$$

The function $\phi(\rho,x_3)$ is determined by $f(y,\theta)$ through Eq. (3b) and can be written as

$$\phi(\rho,x_3) = (1/a)\mathscr{F}(y,\theta), \tag{7}$$

where $\mathscr{F}(y,\theta)$ is a dimensionless function. The energy H then depends quadratically on the parameter c,

$$H = (H_1 + c^2 H_2)/a, \tag{8a}$$

$$H_1 = \frac{1}{2}\int d^3y (\nabla_y \mathscr{F})^2, \tag{8b}$$

$$H_2 = \int d^3y\, \bar{A}^2 \mathscr{F}^2, \tag{8c}$$

where ∇_y means differentiation with respect to y. The charge density q for the solution (4) and (7) can be evaluated from Eq. (3a), and with this charge density so obtained, one can construct the abelian Coulomb solution. It energy is

$$H_c = \frac{1}{2}\int d^3x (\nabla^{-1} q)^2 \tag{9a}$$

$$= \frac{1}{8\pi}\int d^3x\, d^3x' \frac{q(\mathbf{x})q(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}, \tag{9b}$$

where $\nabla^{-1} = \nabla/\nabla^2$. Substituting the expression (3a) into Eq. (9a), we find after partial integration,

$$H_c = \frac{1}{2}\int d^3x[(\nabla\phi)^2 + 2\phi^2A^2 + (\nabla^{-1}(\phi A^2))^2]. \quad (10)$$

Using expressions (4) and (7), this becomes

$$H_c = (H_1 + c^2H_2 + c^4H_3)/a, \quad (11)$$

where

$$H_3 = \frac{1}{2}\int d^3y[\nabla_y^{-1}(\mathscr{F}\overline{A}^2)]^2. \quad (12)$$

Evidently the Coulomb energy H_c is greater than the magnetic multipole energy H as H_3 is positive.

A gauge-invariant characterization of the total external source is

$$Q = \int d^3x(\rho_a\rho_a)^{1/2},$$

$$= \frac{1}{g}\int d^3x\,|q|, \quad (13)$$

where in the present case q is given by

$$q = (-\nabla_y^2\mathscr{F} + c^2F\overline{A}^2)/a^3,$$

$$= [q_1(y,\theta) + c^2q_2(y,\theta)]/a^3. \quad (14)$$

In the neighborhood of a fixed c_0, Eq. (13) takes the form

$$Q = Q_1 + c^2Q^2, \quad (15)$$

and at $c = 0$,

$$Q = \frac{1}{g}\int d^3y\,|q_1(y,\theta)| \equiv Q_c. \quad (16)$$

For small Q, the abelian Coulomb solution is stable.[1] As Q increases, bifurcation will result if respective extremum of the total energy and the total charge occurs at the same value of the c.[4,7] Clearly at $c = 0$, $H_c = H = H_1/a$ and $\partial H_c/\partial c = \partial H/\partial c = 0$, but $\partial^2H_c/\partial c^2 = \partial^2H/\partial c^2 = 2H^2/a \neq 0$; that is, the energy is minimized at $c = 0$. Furthermore,

$$\frac{\partial Q}{\partial c} = 2c\int d^3y\,qq_2/|q|, \quad (17)$$

which vanishes at $c = 0$ but not $\partial^2Q/\partial c^2$,

$$\frac{\partial^2Q}{\partial c^2} = 2\int d^3y\,\frac{q_1q_2}{|q|}. \quad (18)$$

Hence $c = 0$ cannot be an inflection point and bifurcation takes place at $Q = Q_c$, $H = H_c = H_1/a$. We have thus shown that for $c > 0$, the abelian Coulomb solution bifurcates into a set of two branches: the stable magnetic multipole solution and the unstable abelian Coulomb solution. Note that the relation between bifurcation and stability is discussed in detail in Refs. 4 and 7.

III. MAGNETIC DIPOLE SOLUTION

We now restrict the above discussions to the magnetic dipole solution.[2,5] In this case, the expression $\overline{A}(y,\theta)$ in Eq. (14) is given by

$$\overline{A}(y,\theta) = (\sin\theta)f(y)/y^2, \quad (19)$$

and the function $\mathscr{F}$ in Eq. (7) depends on y only[5]:

$$\mathscr{F}(y) = 3u^{2/3}\left(\frac{-1}{f}\frac{d^2f}{du^2}\right)^{1/2}, \quad u = y^3. \quad (20)$$

From (14), we find

$$q_1 = -9\frac{d}{du}\left(u^{4/3}\frac{d\mathscr{F}}{du}\right), \quad q_2 = u^{-4/3}f^2F. \quad (21)$$

With these, the expressions for the total energy and total charge can be simplified:

$$H_1 = 54\pi\int_0^\infty du\,u^{4/3}\frac{d}{du}\left[u^{2/3}\left(\frac{-1}{f}\frac{d^2f}{du^2}\right)^{1/2}\right], \quad (22)$$

$$H_2 = 8\pi\int_0^\infty du\left(-f\frac{d^2f}{du^2}\right), \quad (23)$$

$$H_3 = \frac{16\pi}{81}\left[\int_0^\infty du\,u^{-5/3}f^2\mathscr{F}\int_0^u du'\,u'^{-4/3}f^2\mathscr{F} + \frac{1}{25}\int_0^\infty du\,u^{-7/3}f^2\int_0^u du'\,u'^{-2/3}f^2\mathscr{F}\right], \quad (24)$$

$$Q_c = \frac{12\pi}{g}\int_0^\infty du\left|\frac{d}{du}\left(u^{4/3}\frac{d\mathscr{F}}{du}\right)\right|. \quad (25)$$

Thus given any explicit solution for $f(y)$, all the above quantities can be evaluated.

IV. REMARKS

(1) If instead of the gauge-invariant measure for the total charge as given by expression (13), we define the total charge as [2,7]

$$\overline{Q} = \frac{1}{g}\int d^3x\,q,$$

then from Eq. (14),

$$\overline{Q} = \frac{c^2}{g}\int d^3y\,\overline{A}^2. \quad (26)$$

The critical value $\overline{Q}_c$ vanishes. For the magnetic multipole solution, we still have $\overline{Q} > 0$ since $c = 0$ corresponds to the abelian Coulomb solution.

(2) We stress that the analysis in Sec. II does indicate that the abelian Coulomb solution is unstable for large external source strength and the magnetic multipole solution is linearly stable.

(3) The type II solution obtained in Ref. 4 necessarily requires the support of nonzero critical source strength, but it need not always bifurcate.[8] In contrast, the magnetic multipole solution, which also requires critical source strength, always possesses a bifurcation point.

(4) Near $c = 0$, H varies linearly as Q, while H_c varies quadratically.

ACKNOWLEDGMENT

We wish to thank W. K. Koo and R. Teh for helpful discussions.

[1] J. E. Mandula, Phys. Rev. D **14**, 3497 (1976); Phys. Lett. B **67**, 175 (1977); M. Magg, Phys. Lett. B **78**, 481 (1978).

[2] P. Sikivie and N. Weiss, Phys. Rev. Lett. **40**, 1411 (1978); Phys. Rev. D **18**, 3809 (1978).

[3]K. Cahill, Phys. Rev. Lett. **41**, 599 (1978); M. Kalb, Phys. Rev. D **18**, 2909 (1978); P. Pirila and P. Presnajder, Nucl. Phys. B **142**, 229 (1978); Y. Leroyer and A. Raychaudhuri, Nucl. Phys. B **156**, 465 (1979); Phys. Rev. D **20**, 2086 (1979); P. Sikivie and N. Weiss, Phys. Rev. D **20**, 487 (1979); W. B. Campbell, D. W. Joseph, and T. A. Morgan, Phys. Lett. B **96**, 352 (1980); C. H. Oh, R. Teh, and W. K. Koo, Phys. Rev. D **24**, 2305 (1981).
[4]R. Jackiw, L. Jacobs, and C. Rebbi, Phys. Rev. D **20**, 474 (1979); R. Jackiw and P. Rossi, Phys. Rev. D **21**, 426 (1980).
[5]R. Teh, W. K. Koo, and C. H. Oh, Phys. Rev. D **23**, 3046 (1981); C. H. Oh, R. Teh, and W. K. Koo, Phys. Lett. B **101**, 3376 (1981); C. H. Oh, Phys. Rev. D **25**, 2194 (1982).
[6]A. Raychaudhuri and U. Sarkar, Phys. Lett. B **102**, 421 (1981).
[7]L. Jacobs and J. Wudka, Phys. Rev. D **25**, 1114 (1982).
[8]C. H. Oh, R. Teh, and W. K. Koo, Phys. Rev. D **25**, 3263 (1982); R. Teh, C. H. Oh, and W. K. Koo, Phys. Rev. D **26**, 3649 (1982).

Reduction of the Poincaré gauge field equations by means of duality rotations

Eckehard W. Mielke
International Centre for Theoretical Physics, Trieste, Italy

(Received 22 June 1982; accepted for publication 21 January 1983)

A rather general procedure is developed in order to reduce the two field equations of the Poincaré gauge theory of gravity by a modified ansatz for the curvature tensor involving double duality. In the case of quasilinear Lagrangians of the Yang–Mills type it is shown that nontrivial torsion solutions with duality properties necessarily "live" on an Einstein space as metrical background.

PACS numbers: 11.15. – q, 11.30.Cp, 04.50. + h

I. INTRODUCTION

In recent years Poincaré gauge theories (PG) of gravitation have become a viable alternative to Einstein's theory of general relativity.[1,2] It involves a Riemann–Cartan space-time (U_4) with the tetrad coefficients $e_i^{\cdot\alpha}$ and the connection coefficients $\Gamma_i^{\cdot\alpha\beta} = -\Gamma_i^{\cdot\beta\alpha}$ regarded as independent geometrical variables. (The conventions of Ref. 1 are used throughout.) In PG theories these geometrical objects are considered as translational and (Lorentz-) rotational gauge potentials with the corresponding gauge field strength given by the *torsion*

$$F_{ij}^{\cdot\cdot\alpha} := 2D_{[i}e_{j]}^{\cdot\alpha} := 2(\partial_{[i}e_{j]}^{\cdot\alpha} + \Gamma_{[i|\beta}^{\cdot\cdot\alpha}e_{|j]}^{\cdot\beta}) \tag{1.1}$$

and the U_4 *curvature*

$$F_{ij\alpha}^{\cdot\cdot\cdot\beta} := 2(\partial_{[i}\Gamma_{j]\alpha}^{\cdot\cdot\beta} + \Gamma_{[i|\gamma}^{\cdot\cdot\beta}\Gamma_{|j]\alpha}^{\cdot\cdot\gamma}). \tag{1.2}$$

In accordance with the general ideas on Yang–Mills gauge fields[3] we adopt

$$\mathscr{L}_g = eL_g(e_i^{\cdot\alpha}, F_{ij}^{\cdot\cdot\alpha}, F_{ij\alpha}^{\cdot\cdot\cdot\beta}) \tag{1.3}$$

as our most general, gauge-invariant Lagrangian density for the gravitational field.

The corresponding canonical field momenta are defined according to

$$\mathscr{H}_\alpha^{\cdot ij} := 2\frac{\partial\mathscr{L}_g}{\partial F_{ji}^{\cdot\cdot\alpha}}, \quad \mathscr{H}_{\alpha\beta}^{\cdot\cdot ij} := 2\frac{\partial\mathscr{L}_g}{\partial F_{ji}^{\cdot\cdot\alpha\beta}}. \tag{1.4}$$

Improving earlier attempts of, e.g., Lanczos[4] and Lopez,[5] von der Heyde[6] could correctly deduce the Euler–Lagrange equations for such *general* Lagrangians. The two first-order field equations of the Poincaré gauge theory are of the following, elegant Maxwell-type form:

$$D_j\mathscr{H}_\alpha^{\cdot ij} - \epsilon_\alpha^{\cdot i} = e\Sigma_\alpha^{\cdot i}, \tag{1.5}$$

$$D_j\mathscr{H}_{\alpha\beta}^{\cdot\cdot ij} - \mathscr{H}_{[\beta\alpha]}^{\cdot\cdot i} = e\tau_{\alpha\beta}^{\cdot\cdot i}. \tag{1.6}$$

The sources on the right-hand side of (1.5) and (1.6) are the canonical momentum currents and the canonical spin currents of the matter fields, if present. Compared to Maxwell's theory of electromagnetism, additional terms arise in the field equations of PG theory due to its inherent nonabelian group structure. The *nonlinear* contribution

$$\epsilon_\alpha^{\cdot i} = e_{\cdot\alpha}^{i}\mathscr{L}_g - F_{\alpha j}^{\cdot\cdot\gamma}\mathscr{H}_\gamma^{\cdot ji} - F_{\alpha j}^{\cdot\cdot\gamma\delta}\mathscr{H}_{\gamma\delta}^{\cdot\cdot ji} \tag{1.7}$$

comprise the energy-momentum currents of the translational and rotational gauge fields, whereas $\mathscr{H}_{[\alpha\beta]}^{\cdot\cdot i}$ corresponds to the spin current of the translational gauge field.

It has been suggested earlier[7,8] that the tetrad field $e_i^{\cdot\alpha}$ should be regarded as a gravitational Higgs field. This view gets further support after having a closer look at the basic set (1.5) and (1.6) of field equations. By reformulating PG theory in terms of globally defined differential forms[9] it can be made particularly transparent that they correspond to a coupled Yang–Mills–*Higgs* system rather than to pure gauge field equations. Note that the field equations are supplemented by the two *Bianchi identities*

$$D_j(e^*F^{ji\alpha}) \equiv e^*F_{\cdot\cdot j}^{ji\cdot\alpha}, \tag{1.8}$$

$$D_j(e^*F_{\cdot\cdot\alpha\beta}^{*ij}) \equiv 0 \tag{1.9}$$

in a Riemann–Cartan space-time as well as by the Noether identities[1]

$$D_j(e\Sigma_\alpha^{\cdot j}) \equiv F_{\alpha i}^{\cdot\cdot\beta}e\Sigma_\beta^{\cdot i} + F_{\alpha i}^{\cdot\cdot\beta\gamma}e\tau_{\beta\gamma}^{\cdot\cdot i}, \tag{1.10}$$

$$D_j(e\tau_{\alpha\beta}^{\cdot\cdot j}) - e\Sigma_{[\alpha\beta]} \equiv 0. \tag{1.11}$$

In flat space-time the latter degenerate to the conservation laws of energy-momentum and angular momentum currents, respectively.

In order to keep as close as possible to the original Yang–Mills theory[3] as well as to Einstein's theory of general relativity, the canonical field momenta (1.4) are usually[1,3] taken as linear in torsion

$$\mathscr{H}_\alpha^{\cdot ij} = (e/l^{*2})(d_1F_{\cdot\cdot\alpha}^{ji} + d_2F_{[\alpha}^{\cdot ij]} + d_3e_{\cdot\alpha}^{[i}F_{\cdot\cdot\gamma}^{j]\gamma}) \tag{1.12}$$

and at most linear with respect to the curvature

$$\begin{aligned}\mathscr{H}_{\alpha\beta}^{\cdot\cdot ij} = (e/\kappa)[&F_{\cdot\cdot\alpha\beta}^{ij} + f_1F_{\cdot\alpha\cdot\beta}^{[i\cdot j]} + f_2F_{\alpha\beta}^{\cdot\cdot ij} \\ &+ f_3e_{\cdot[\beta}^{[i}F_{\cdot\cdot\alpha]\mu}^{|\mu|j]} + f_4e_{\cdot[\beta}^{[i}F_{\cdot\alpha]\cdot\mu}^{|\mu|\cdot j]} + f_5e_{\cdot[\beta}^{i}e_{\cdot\alpha]}^{j}F] \\ &+ (e/\chi l^{*2})e_{\cdot[\alpha}^{i}e_{\cdot\beta]}^{j}.\end{aligned} \tag{1.13}$$

This leads to

$$\begin{aligned}\mathscr{L}_g = \tfrac{1}{4}[&F_{ji}^{\cdot\cdot\alpha}\mathscr{H}_\alpha^{\cdot ij} + F_{ji}^{\cdot\cdot\alpha\beta}\mathscr{H}_{\alpha\beta}^{\cdot\cdot ij}] \\ &+ (e/4\chi l^{*2})F + \Lambda e/l^{*4}\end{aligned} \tag{1.14}$$

as our most general *quasilinear*[1] Lagrangian density. Here $l^* = (8\pi\hbar G/c^3)^{1/2}$ denotes the modified Planck length, $e = \det e_i^{\cdot\alpha}$, $F := F_{\beta\alpha}^{\cdot\cdot\alpha\beta}$ the scalar curvature, and Λ a cosmological constant. (The particle content of such quasilinear models of gravity has been analyzed recently[10,11] for $\Lambda = 0$.) However, also in PG theory, one should be open to more general gauge invariant Lagrangians capable of generating an essentially nonlinear gauge dynamics similar to those proposed, e.g., by Born and Infeld[12] for electromagnetism and more recently discussed by Mills[13] in the case of nonabelian gauge fields in order to achieve confinement.

II. DOUBLE DUALITY ANSATZ FOR THE CURVATURE

In order to solve the general *vacuum* field equations (1.5) and (1.6) the strategy of Belavin *et al.*[14] used in the derivation of the instanton solution will be adopted. Such a procedure has been applied by the author[15] to Yang's theory of gravity[16] in the search for pseudoparticle solutions, as well as by Baekler *et al.*[17] in order to rederive Baekler's nontrivial torsion solution[18] for a certain class of quasilinear PG Lagrangians. Recently[19,20] multimonopole solutions of the Yang–Mills–Higgs equations have been constructed using likewise a duality ansatz as a first step. In our case this amounts to employing the (modified) *double duality* ansatz

$$\mathscr{H}_{\alpha\beta}^{\cdot\cdot ij} = \zeta\,(e/\kappa)^*F^{*ij}_{\cdot\cdot\alpha\beta} + 2\gamma(e/l^{*2})e^{i}_{\cdot[\alpha}e^{j}_{\cdot\beta\,]} \tag{2.1}$$

for the *rotational* field momenta $\mathscr{H}_{\alpha\beta}^{\cdot\cdot ij}$. The right, left and double dual tensors are defined according to the following conventions:

$$F^*{}_{\alpha\beta\mu\nu} := (i/2)e\epsilon_{\mu\nu\sigma\tau}F_{\alpha\beta}{}^{\sigma\tau}, \tag{2.2}$$

$$^*F^{\alpha\beta\mu\nu} := (i/2)(1/e)\epsilon^{\alpha\beta\gamma\delta}F_{\gamma\delta}{}^{\mu\nu}, \tag{2.3}$$

$$^*F^{*\alpha\beta}_{\cdot\cdot\mu\nu} := -\tfrac{1}{4}\epsilon^{\alpha\beta\gamma\delta}F_{\gamma\delta}{}^{\sigma\tau}\epsilon_{\sigma\tau\mu\nu}, \tag{2.4}$$

$$^*F^{\alpha\beta\gamma} := (i/2)(1/e)\epsilon^{\alpha\beta\mu\nu}F_{\mu\nu}{}^{\gamma}, \tag{2.5}$$

where $\epsilon^{\alpha\beta\gamma\delta}$ with $\epsilon^{\hat{0}\hat{1}\hat{2}\hat{3}} = -1$ is the anholonomic Levi-Civitá tensor.

It should be remarked that the ansatz (2.1) may be regarded as a generalization of the *duality rotation*

$$F_{\mu\nu} \rightarrow e^{*\delta}F_{\mu\nu} := F_{\mu\nu}\cos\delta + {}^*F_{\mu\nu}\sin\delta \tag{2.6}$$

of a gauge field strength $F_{\mu\nu}$, which in the case of electromagnetism plays a decisive role in the Rainich geometrization[21,22] of the coupled Einstein–Maxwell system. It is the general virtue of such a duality ansatz that it will "rotate" the dynamical part of a gauge field equation into the corresponding Bianchi identity, thereby facilitating considerably the search for *exact* solutions.

III. SOLVING THE SECOND FIELD EQUATION

Let us first turn to the second field equation (1.6) and apply the ansatz (2.1). Since its constant curvature piece after covariant differentiation produces the modified torsion

$$T_{\alpha\beta}^{\cdot\cdot i} := F_{\alpha\beta}^{\cdot\cdot i} + 2e^{i}_{\cdot[\alpha}F_{\beta]\gamma}^{\cdot\cdot\gamma} = (2/e)D_j(ee^{i}_{\cdot[\alpha}e^{j}_{\cdot\beta\,]}), \tag{3.1}$$

the left-hand side of (1.6) as a consequence of the Bianchi identity (1.9) turns out to be

$$D_j\mathscr{H}_{\alpha\beta}^{\cdot\cdot ij} - \mathscr{H}_{[\beta\alpha\,]}^{\cdot\cdot i} = \gamma(e/l^{*2})T_{\alpha\beta}{}^{i} - \mathscr{H}_{[\beta\alpha\,]}^{\cdot\cdot i}. \tag{3.2}$$

For the second field equation (1.6) to be satisfied we have to require

$$\mathscr{H}_{[\alpha\beta\,]}^{\cdot\cdot i} = -\gamma(e/l^{*2})T_{\alpha\beta}^{\cdot\cdot i} + e\tau_{\alpha\beta}^{\cdot\cdot i}. \tag{3.3}$$

Therefore, in vacuum the ansatz (2.1) for the rotational field momenta—without assuming this beforehand—imposes a certain quasilinearity on the antisymmetric part of the translational momenta! At a second thought this appears to be not that surprising, since the duality ansatz (2.1) involves likewise a kind of "linearization" for the rotational momenta.

In order to see whether the condition (3.3) can be satisfied for the quasilinear choice (1.12) the identity

$$F^{i}_{\cdot[\alpha\beta\,]} = \tfrac{3}{2}F^{[i}_{\cdot\alpha\beta\,]} - \tfrac{1}{2}F_{\alpha\beta}^{\cdot\cdot i} \tag{3.4}$$

has to be employed. Then in vacuum (3.3) assumes the following algebraic condition on the torsion tensor:

$$(\gamma + d_1/2)F_{\alpha\beta}^{\cdot\cdot i} + (d_2 - \tfrac{3}{2}d_1)F^{\cdot\cdot i]}_{[\alpha\beta} + (2\gamma - d_3/2)e^{i}_{\cdot[\alpha}F_{\beta]\nu}^{\cdot\cdot\nu} = 0. \tag{3.5}$$

As already remarked by the double dual curvature part $^*F^{*ij}_{\cdot\cdot\alpha\beta}$ of the ansatz (2.1) will be mapped into the second Bianchi identity (1.9) and consequently drops out, whereas the "unit" curvature term in (2.1) produces a certain amount of torsion, which may be adjusted by a proper choice of the free constant γ so as to make the right-hand side of (3.2) vanish.

Three cases have to be distinguished. For

(A) *"spherical" torsion*

$$F_{[\alpha\beta\gamma\,]} = 0 \quad \text{and} \quad 2\gamma = d_3/2 = -d_1 \tag{3.6}$$

has to be required, whereas for

(B) *purely axial torsion*

$$F_{\alpha\beta\gamma} = F_{[\alpha\beta\gamma\,]} \quad \text{and} \quad \gamma = d_1 - d_2 \tag{3.7}$$

is mandatory in order to satisfy the second field equation.

(C) The trivial case $\mathscr{H}_{\alpha}^{\cdot ij} = 0$ requires $\gamma = 0$. Consequently, it yields no restriction on the symmetry of those torsion terms being concealed in the U_4 curvature. Note that in this case, the constant curvature piece in the related "modified" double duality ansatz of Ref. 23 is solely a consequence of the choice (1.13) of the rotational field momenta. This unit tensor will not be differentiated and consequently does not contribute to the spin current of the gauge field, in contrast to the generic situation.

An irreducible decomposition of the torsion shows that in vacuum these are the only possible solutions of (3.3) in the case of quasilinear translational momenta.

IV. REDUCTION OF THE FIRST FIELD EQUATION

The next step is to reduce the *first* field equation (1.5). However, its *antisymmetric* part

$$D_j\mathscr{H}_{[\alpha\beta\,]}^{\cdot j} - \tilde{\epsilon}_{[\alpha\beta\,]} = e\Sigma_{[\alpha\beta\,]}, \tag{4.1}$$

written anholonomically by introducing

$$\tilde{\epsilon}_{\alpha\beta} := \epsilon_{\alpha\beta} - \tfrac{1}{2}F_{ij\beta}\mathscr{H}_{\alpha}^{\cdot ij} = -(e/l^{*2})X_{\alpha\beta} + \overset{\text{ROT}}{\epsilon_{\alpha\beta}}, \tag{4.2}$$

turns out to be redundant for vacuum solutions with duality properties, except for a certain topological information.

To this end, let us split the anholonomic energy-momentum curent as indicated into a translational part

$$X_{\alpha\beta} := -(l^{*2}/e)\overset{\text{TR}}{\tilde{\epsilon}_{\alpha\beta}} = (l^{*2}/e)[-\tfrac{1}{4}\eta_{\alpha\beta}F_{ij}^{\cdot\cdot\gamma}\mathscr{H}_{\gamma}^{\cdot ji} + F_{\alpha j}^{\cdot\cdot\gamma}\mathscr{H}_{\gamma\beta}^{\cdot\cdot j} + \tfrac{1}{2}F_{ij\beta}\mathscr{H}_{\alpha}^{\cdot ij}] \tag{4.3}$$

depending only on the torsion and a remaining contribution from the rotational field strength. Upon inserting the ansatz (3.3)—or its resolved form (4.12) given later—into (4.3) the expression

$$X_{[\alpha\beta\,]} = -(l^{*2}/e)[\tfrac{1}{2}F_{ij[\alpha}\mathscr{H}_{\beta\,]}^{\cdot ij} - F_{i[\alpha|j}\mathscr{H}_{\cdot|\beta\,]}^{j\cdot i}] = 2\gamma F_{\mu[\alpha\beta\,]}F^{\mu\nu}_{\cdot\cdot\nu} \tag{4.4}$$

of the antisymmetrized translational energy-momentum current shows clearly that it vanishes in the cases A and B due to the torsional symmetry requirements (3.6) and (3.7) as well as in case C for trivial reasons. Consequently, only the curvature is instrumental in the antisymmetrical part of the energy-momentum current. Linearizing it by means of the duality ansatz (2.1) a calculation reveals that

$$\overset{\mathrm{ROT}}{\epsilon}_{[\alpha\beta]} = -2\gamma(e/l^{*2})F_{[\alpha\beta]} + {}^*\epsilon^*{}_{[\alpha\beta]} \tag{4.5}$$

is proportional to the antisymmetric Ricci tensor in a U_4, notwithstanding the "dual" current

$$\begin{aligned} {}^*\epsilon^*{}_{\alpha\beta} &= \zeta\,(e/2\kappa)(F_{\alpha}^{\cdot\mu\nu\sigma}{}^*F^*{}_{\beta\mu\nu\sigma} - {}^*F^*{}_{\alpha}^{\cdot\mu\nu\sigma}F_{\beta\mu\nu\sigma}) \\ &= -\zeta\,(e/\kappa)(FF_{[\alpha\beta]} + 2F^{\mu\cdot\cdot\nu}_{\cdot[\alpha\beta]}F_{\nu\mu}). \end{aligned} \tag{4.6}$$

Note that this current which involves the double dual curvature tensor is antisymmetric, i.e., ${}^*\epsilon^*{}_{(\alpha\beta)} = 0$, and according to the given equivalent expression [compare with (4.12) of Ref. 17] vanishes in a Riemannian space-time (V_4).

Using this information, the derived condition (3.3) on the translational field momenta as well as (1.11), we find that the field equations (4.1) by means of a duality rotation are transformed into the *contracted* form

$$D_j(eT^{\cdot\cdot j}_{\alpha\beta}) \equiv 2eF_{[\alpha\beta]} \tag{4.7}$$

of the first Bianchi identity (1.8) *except* for the requirement

$${}^*\epsilon^*{}_{[\alpha\beta]} = 0. \tag{4.8}$$

In order to understand the meaning of this remainder of the first field equation (occurring only in a U_4!) note that (4.6) is formally the energy-momentum current corresponding to the Lagrangian density

$$-\frac{\zeta}{\kappa}\mathscr{L}_{\mathrm{Euler}} := -\frac{\zeta}{4}\frac{e}{\kappa}F^{\alpha\beta}_{\cdot\cdot\gamma\delta}{}^*F^{*\cdot\cdot\gamma\delta}_{\alpha\beta} \tag{4.9}$$

of the Euler type. It is known that this term contributes only a boundary terms to the gravitational action. More precisely, upon integration over a *closed* four-dimensional manifold it yields the Euler number

$$\chi(M^4) = \frac{(-1)^{3/2}}{2(2\pi)^2}\int_{M^4}\mathscr{L}_{\mathrm{Euler}}\,d^4x \tag{4.10}$$

due to the generalized Gauss–Bonet theorem (see, e.g., Levine and Zund,[24] also Mielke,[15] for the sign conventions for spaces with Lorentzian signature).

A *sufficient* condition for (4.8) to hold is therefore the vanishing of (4.9), or equivalently, the restriction to nontrivial torsion configurations for which the Euler number is zero. It is gratifying to note that this restriction on the global topology is, e.g., met[25] by Baekler's solution.[18]

Then we are left with the *symmetric* part

$$D_j\mathscr{H}^{\cdot\cdot j}_{(\alpha\beta)} + (e/l^{*2})X_{(\alpha\beta)} - \overset{\mathrm{ROT}}{\epsilon}_{(\alpha\beta)} = e\Sigma_{(\alpha\beta)}. \tag{4.11}$$

of the first field equation (1.5). From the ansatz (3.3) the translational field momenta it may be deduced with the help of (3.4) that

$$\begin{aligned} \mathscr{H}^{\cdot\cdot j}_{\alpha\beta} &= -2\gamma(e/l^{*2})(T^{j}_{\cdot\beta\alpha} + \tfrac{3}{2}T^{[j}_{\cdot\alpha\beta]}) \\ &\quad + 2e\tau^{j}_{\cdot\beta\alpha} + 3e\tau^{[j}_{\cdot\alpha\beta]}, \end{aligned} \tag{4.12}$$

i.e., that the symmetric part is reduced to

$$\mathscr{H}^{\cdot\cdot j}_{(\alpha\beta)} = -2\gamma(e/l^{*2})T^{j}_{\cdot(\alpha\beta)} + 2e\tau^{j}_{\cdot(\alpha\beta)}. \tag{4.13}$$

Furthermore, the curvature dependent part $\overset{\mathrm{ROT}}{\epsilon}{}^{\cdot i}_{\alpha}$ of the gauge momentum currents (1.7) after substitution of the duality ansatz (2.1) assumes the form

$$\overset{\mathrm{ROT}}{\epsilon}_{(\alpha\beta)} = -2\gamma\frac{e}{l^{*2}}\not{F}_{(\alpha\beta)} + \frac{e}{l^{*2}}\left(\frac{A}{l^{*2}} + \frac{F}{4\chi}\right)\eta_{\alpha\beta}, \tag{4.14}$$

i.e., it involves the symmetric *traceless* Ricci tensor $\not{F}_{(\alpha\beta)}$ and a trace part. Due to the duality ansatz (2.1) the occurring scalar curvature is a constant, i.e.,

$$F = -(\kappa/Al^{*2})(12\gamma + 6/\chi), \tag{4.15}$$

where

$$A := 1 + \tfrac{1}{3}f_1 + f_2 + f_3 + f_4 + 6f_5 - \zeta \tag{4.16}$$

provided our analysis has been restricted to a theory which is also quasilinear in the curvature, i.e., to a model for which (1.13) holds.

The crucial step in the reduction of (4.11) is to note the identity [Ref. 17, Eq. (4.5)]

$$(1/e)D_j(eT^{j}_{\cdot(\alpha\beta)}) = -G_{\alpha\beta}(\{\ \}) + G_{(\alpha\beta)} - Y_{\alpha\beta} \tag{4.17}$$

for the difference between the Einstein tensor $G_{\alpha\beta}(\{\ \})$ with respect to a symmetric (Christoffel) connection $\{{}^{i}_{\alpha\beta}\}$ of general relativity and the corresponding tensor

$$G_{\alpha\beta} := -{}^*F^{*\mu}{}_{\beta\alpha\mu} = F_{\alpha\beta} - \tfrac{1}{2}\eta_{\alpha\beta}F \tag{4.18}$$

in a Riemann–Cartan space-time. The additional symmetric tensor $Y_{\alpha\beta}$ in (4.17) depends solely on torsion

$$\begin{aligned} Y_{\alpha\beta} := {} & \tfrac{1}{2}T_{\alpha\mu\nu}(T^{\cdot\mu\nu}_{\beta} + T^{\cdot\nu\mu}_{\beta}) - \tfrac{1}{4}T_{\mu\nu\alpha}T^{\mu\nu}_{\cdot\cdot\beta} - \tfrac{1}{2}T_{\alpha}T_{\beta} \\ & + \tfrac{1}{4}\eta_{\alpha\beta}\{T_{\mu}T^{\mu} - T_{\mu\nu\kappa}(T^{\mu\nu\kappa} + \tfrac{1}{2}T^{\mu\kappa\nu})\}. \end{aligned} \tag{4.19}$$

Then, on account of (4.13) the relation (4.17) can be employed to eliminate the covariant derivative in the first field equation (4.11). Using also (4.18) the Einstein-type field equations

$$\begin{aligned} & G_{\alpha\beta}(\{\ \}) + (1/l^{*2})\Lambda_{\mathrm{eff}}\eta_{\alpha\beta} \\ & \quad = \frac{l^{*2}}{2\gamma}\left[\Sigma_{(\alpha\beta)} - \frac{2}{e}D_j(e\tau^{j}_{\cdot(\alpha\beta)})\right] + \tilde{T}_{\alpha\beta} \end{aligned} \tag{4.20}$$

remain with an effective "cosmological constant"

$$\begin{aligned} \Lambda_{\mathrm{eff}} &= -\frac{A}{2\gamma} - \frac{3\kappa}{A}\left(1 - \frac{1}{2\gamma\chi}\right)\left(\gamma + \frac{1}{2\chi}\right) \\ &= -\frac{A}{2\gamma} + \frac{2\gamma\chi - 1}{8\gamma\chi}l^{*2}F \end{aligned} \tag{4.21}$$

of partially microscopic[26] origin.

The insertion of the resolved ansatz (4.12) for the translational momenta into (4.3) shows that the additional "source" tensor

$$\begin{aligned} \tilde{T}_{\alpha\beta} &:= -((1/2\gamma)X_{(\alpha\beta)} + Y_{\alpha\beta}) \\ &= -\tfrac{1}{2}{}^*T_{\alpha}{}^*T_{\beta} - {}^*T^{\gamma}_{\cdot(\alpha\beta)}{}^*T_{\gamma} + \tfrac{1}{4}\eta_{\alpha\beta}{}^*T^{\gamma*}T_{\gamma} \end{aligned} \tag{4.22}$$

for the Einstein equations (4.20) vanishes only in the case of spherical or zero torsion (case A) but not for purely axial torsion. (Remarkably enough, the combination

$$(1/2\gamma)X_{(\alpha\beta)} - Y_{\alpha\beta} = \tfrac{1}{4}T^{k}T_{k(\alpha\beta)} \tag{4.23}$$

has the opposite property with respect to the symmetry of the torsion).

V. SOLUTIONS WITH SPHERICAL TORSION

Let us now consider the case A in which the antisymmetric part of the torsion vanishes. The preceeding calculation reveals that in vacuum the right-hand side of (4.20) is zero. This is not surprising, since the conditions (3.6) relate our case to the special one studied by Baekler *et al.*[17] in which the same result holds. Then, it follows for (4.20) that the symmetric part of the first field equation reduces completely to the Einstein field equations

$$G_{ij}(\{\ \}) + (\Lambda_{\text{eff}}/l^{*2})g_{ij} = 0 \tag{5.1}$$

of general relativity written holonomically.

In order to obtain nontrivial torsion solutions, we may integrate the duality ansatz (2.1) on an Einstein space as metrical background. This still difficult task has been undertaken in Ref. 17 for a particular model specified by the choice

$$\Lambda = 1/\chi = f_i = 0, \quad d_1 = -1,\ d_2 = 0,\ d_3 = 2 \tag{5.2}$$

of the characteristic parameters of the quasilinear Lagrangian (1.14).

In a spherically symmetric background, among others, a new derivation of the Baekler solution[18] is thereby achieved. This solution is reminiscent of the self-dual instanton solutions[27] of the coupled Einstein–Yang-Mills system inasmuch as in both theories the metric is not deformed off from an Einstein space and both configurations also have no flat space analogs.

Benn *et al.*[28] have found a similar spherical solution as Baekler's in a model with $\mathscr{H}_\alpha^{ij} = 0$ (case C with $\gamma = 0$). Then the symmetric part (4.11) of the first field equation is identically fulfilled merely by requiring $\Lambda = -(\kappa/A)(3/2\chi^2)$. In view of this it may be questioned whether an Einstein space is really the most general starting point of such a deduction.

VI. CONFIGURATIONS WITH AXIAL TORSION

In the case B of purely axial torsion, the symmetrized translational field momenta (4.13) have to vanish on account of (3.7). Consequently, the first field equation reduces in vacuum to

$$F_{(\alpha\beta)} = -\frac{1}{2\gamma}\left\{X_{(\alpha\beta)} + \left(\frac{\Lambda}{l^{*2}} + \frac{F}{4\chi}\right)\eta_{\alpha\beta}\right\}. \tag{6.1}$$

For a purely axial torsion tensor this implies the constraint

$$X^{\alpha}{}_{\alpha} = \tfrac{1}{2}(d_2 - d_1)F_{[\mu\nu\kappa]}F^{[\mu\nu\kappa]} = -\left(\frac{4\Lambda}{l^{*2}} + \frac{F}{\chi}\right) \tag{6.2}$$

on the trace of the translational part (4.3) of the energy-momentum current

Consequently, the axial torsion turns out to be dual to a lightlike axial vector in those particular models, studied, e.g., by Baekler *et al.*,[17] for which $\Lambda = 0$ and $1/\chi = 0$ hold. Otherwise, there is no further information contained in (6.1) compared with the relation (4.20) which is true in general.

Solutions of the Einstein equations (4.20) with a nonvanishing "source" term given by (4.22) provide the metrical background geometry in the search for solutions with axial torsion. [Although given by the same formal expression (4.21) the occuring "cosmological" constant in (4.20) takes on a different value due to $\gamma = d_1 - d_2$].

So far no explicit solutions of the duality ansatz (2.1) have emerged in this case. For a background with axial symmetry the techniques described by Tomimatsu[29] or, alternatively, Bogomol'nyi's[30] method of generating monopole solutions of the coupled Yang–Mills–Higgs system could possibly provide a way out if transferred to PG theory.

VII. THE REDUCED LAGRANGIAN

In order to conclude this paper we would like to examine how these reductions of the field equations are reflected in the quasilinear Lagrangian (1.14) of the quadratic Poincaré gauge theory (PG^2).

The insertion of the double-duality ansatz (2.1) yields

$$\begin{aligned}\mathscr{L}_g = \frac{e}{l^{*2}}\left[\frac{1}{2}\left(\gamma + \frac{1}{2\chi}\right)F + \frac{\Lambda}{l^{*2}}\right] \\ -\frac{\zeta}{4}\frac{e}{\kappa}F^{\alpha\beta}{}_{\cdot\cdot\gamma\delta}{}^{*}F^{*\cdot\cdot\gamma\delta}{}_{\alpha\beta} \\ +\tfrac{1}{4}F_j{}^{\cdot\alpha\beta}\mathscr{H}^{\cdot\cdot j}_{\beta\alpha}.\end{aligned} \tag{7.1}$$

For a comparison with the field equation (4.20) it is crucial to pass on to the *effective* "cosmological" constant Λ_{eff} in the Lagrangian. This may be achived by noting the second part of (4.21), i.e., the formal relation

$$\Lambda = -2\gamma\Lambda_{\text{eff}} + (\gamma/2 - (1/4\chi))l^{*2}F. \tag{7.2}$$

Furthermore, we now make use of the resolved ansatz (4.12) for the translational momenta with respect to the torsion. Then (7.1) reads

$$\begin{aligned}\mathscr{L}_g = \gamma\frac{e}{l^{*2}}\Big[F - \frac{2}{l^{*2}}\Lambda_{\text{eff}} - \tfrac{1}{2}F_j{}^{\cdot\alpha\beta}(T^j{}_{\cdot\alpha\beta} \\ -\tfrac{3}{2}T^{[j}{}_{\cdot\alpha\beta\,]})\Big] - \frac{\zeta}{\kappa}\mathscr{L}_{\text{Euler}}.\end{aligned} \tag{7.3}$$

From the trace of (4.17) one may deduce the splitting of the occurring U_4 scalar curvature F into the corresponding Riemannian scalar R plus related torsion-dependent terms:

$$\begin{aligned}F = R + \tfrac{1}{4}F_j{}^{\cdot\alpha\beta}(2T^j{}_{\cdot\alpha\beta} - 3T^{\cdot j]}{}_{[\alpha\beta}) \\ + (2/e)\partial_j(eF^{j\cdot\nu}{}_{\cdot\nu}).\end{aligned} \tag{7.4}$$

It should be noted that this identity is at the heart of the equivalence proof relating the teleparallelism theory [characterized by a vanishing U_4 curvature and the choice $f_i = 1/\chi = \Lambda = 0, d_1 = -1, d_2 = \frac{3}{2}, d_3 = 2$ of the dynamical parameters in (1.12) and (1.13)] to Einstein's standard theory of general relativity. For linear translational momenta (1.12) the Lagrangian (7.3) may also be written as follows:

$$\begin{aligned}\mathscr{L}_g = \frac{e}{l^{*2}}\Big\{\gamma R - \frac{2\gamma}{l^{*2}}\Lambda_{\text{eff}} + 2\gamma\frac{1}{e}\partial_j(eF^{j\cdot\nu}{}_{\cdot\nu}) \\ +\tfrac{1}{4}F_j{}^{\cdot\alpha\beta}[(2\gamma + d_1)F^j{}_{\cdot\alpha\beta} - (3\gamma + d_2)F^{\cdot j]}{}_{[\alpha\beta} \\ -(4\gamma - d_3)\eta_{[\alpha|\beta}F^{j]\cdot\mu}{}_{\cdot\mu}]\Big\} - \frac{\zeta}{\kappa}\mathscr{L}_{\text{Euler}}.\end{aligned} \tag{7.5}$$

The constraints (3.6) or (3.7), respectively, on the symmetry of the torsion tensor together with the condition for the free constant γ forces the quadratic torsion terms to vanish in both cases A and B (C is trivial). Consequently, we are left with an Einstein–Hilbert type Lagrangian density

$$\mathscr{L}_g = 2\gamma \frac{e}{l^{*2}} \left\{ \frac{1}{2} R - \frac{\Lambda_{\text{eff}}}{l^{*2}} + \frac{1}{e} \partial_j (e F^{j\gamma}_{\cdot\cdot\gamma}) \right\} - \frac{\zeta}{\kappa} \mathscr{L}_{\text{Euler}} \tag{7.6}$$

in a *Riemannian* space-time except for two boundary terms.

These additional terms are the relicts of the Riemann–Cartan space-time (U_4) we started with and, according to our discussion in Sec. IV, would have been expected to occur. One of these boundary terms corresponding to a local expression of the Euler–Poincaré characteristic and is of significance for the global topology. On the other hand, the term linear in the torsion does not seem to have a similar interpretation. However, for solutions with axial torsion the total derivative with respect to the torsion is zero anyhow, whereas the vanishing of the Euler–Poincaré characteristic is a necessary condition in order to satisfy the remainder (4.8) of the antisymmetrized first field equation. Therefore, the reduction of quadratic Poincaré gauge theories of gravity to Einstein's theory by means of a double duality rotation is rather complete even in the generic case. Configurations with nontrivial torsion do exist,[17,23] the information on the latter being coded into the duality ansatz. Inserting the decomposition

$$F_{ij\alpha}{}^{\beta} = R_{ij\alpha}{}^{\beta} - 2D_{[i}K_{j]\alpha}{}^{\beta} + 2K_{[i|\alpha}{}^{\epsilon}K_{j]\epsilon}{}^{\beta} \tag{7.7}$$

of the U_4 curvature, (2.1) entails partial differential equations for the torsion on the Riemannian background given by (5.1).

VIII. OUTLOOK

As a gauge field theory, PG theory is founded on the quantum realm of microphysics. Consequently, the nontrivial torsion solutions emerging from the described reduction procedure would be expected to gain a meaning only after fully quantizing the theory.

Following covariant quantization methods[31] the calculation of the scattering matrix for asymptotic states proceeds via Feynman path integrals. As in Yang–Mills theories[32] also for a quantized, quadratic Poincaré gauge theory (QPG^2) one would suppose that the contributions from the gravitational configurations of the *instanton* type (see, e.g., Ref. 15) dominate in the transition amplitudes. This could be particularly important if the contributions of classical solutions with nontrivial topology[33] as well as with nontrivial torsion[34] are considered.

Already in the "one-loop" or WKB approximations to these path integrals the divergences stemming from the conventional Einstein–Hilbert Lagrangian density of general relativity have to be compensated by counter terms which are quadratic in the curvature tensor. For a Riemannian space-time, Stelle[35] could prove that a model containing quadratic curvature terms of the Yang–Mills type in the gravitational action is renormalizable in each order of the perturbative expansion. However, such a modified Stephenson–Kilmister–Yang (SKY) gravity would give rise to "ghosts." These states can be suppressed in the special cases of QPG^2 listed in Refs. 10 and 11. Moreover, under certain conditions[36] such tensorial states having a negative norm in the quantal Hilbert space are innocuous, since they do not necessarily invalidate the Froissart unitarity boundedness of cross sections.

Generalizing the definition given in Ref. 15 to PG theory, those configurations should be regarded as *gravitational instantons* which are nonsingular in an "Euclidean" space-time and satisfy duality ansätze such as (2.1) and (3.3). Since these solutions occur on Einstein spaces as metrical background, it follows from (7.6) that the classical gravitational action given by (1.14) takes an extremum. Apart from a torsional boundary term the precise value of which is determined by the Euler number (4.10) which is instrumental for classifying instantons topologically.

Interesting enough, the generation of a "cosmological" or "bag"[37] term of microscopical origin, and in the wake of this, a *"spontaneous compactification"*[38] *of space-time* appears to be unavoidable in such constructions. Macroscopically the astrophysical data indicate that the cosmological constant is rather small if not zero. Similarly, as in the Weinberg-Salam model in which the Higgs fields after symmetry breaking generate a huge vacuum expectation value,[39] in the QPG^2 theory, being dominated by an "instanton gas," there remains then the problem of compensating for the "induced" cosmological term. Superficially, the *effective* cosmological constant (4.21) can be given any value (even zero!) by properly adjusting the "bare" constant Λ. On a deeper level this could be viewed as an indication that a "confining phase" may occur in such geometrodynamical models.[9,26]

ACKNOWLEDGMENTS

The author expresses his sincere gratitude to Professor Friedrich W. Hehl and Diplom-Physiker Peter Baekler for many stimulating discussions on this subject and the warm hospitality during his stay at Cologne. He would also like to thank Professor Abdus Salam, the International Atomic Energy Agency and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste. This work was supported by a grant of the Deutsche Forschungsgemeinschaft, Bonn.

[1]F. W. Hehl, *Four Lectures on Poincaré Gauge Field Theory*, given at Erice, May 1979, edited by P. G. Bergmann and V. De Sabbata (Plenum, New York, 1980), p. 5.

[2]F. W. Hehl,in *Grundlagen probleme der modernen Physik*, edited by J. Nitsch, J. Pfarr, and E.-W. Stachow (Bibliographisches Institut, Mannheim, 1981), p.103.

[3]C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).

[4]C. Lanczos, Rev. Mod. Phys. **57**, 336 (1957).

[5]C. A. Lopez, Lett. Nuovo Cimento **6**, 608 (1973); C. A. Lopez, Int. J. Theor. Phys. **16**, 167 (1977).

[6]P. von der Heyde, Phys. Lett. A **58**, 141 (1976); Z. Naturforschung A **31**, 1725 (1977).

[7]C. J. Isham, Abdus Salam, and J. Strathdee, Ann. Phys. (N.Y.) **62**, 98 (1971).

[8]Y. Nambu, Ann. New York Acad. Sci. **294**, 74 (1977).

[9]E. W. Mielke, "Über die Hypothesen, welcher der Geometrodynamik zugrunde liegen," Habilitation-thesis, Kiel, March 1982.

[10]E. Sezgin and P. van Nieuwenhuizen, Phys. Rev. **D21**, 3269 (1980).

[11]E. Sezgin, Phys. Rev. D **24**, 1677 (1981).

[12]M. Born and L. Infeld, Proc. Roy. Soc. London Ser. A **144**, 425 (1934).

[13]R. Mills, Phys. Rev. Lett. **43**, 549 (1979).

[14]A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin. Phys. Lett. B **53**, 85 (1975).
[15]E. W. Mielke, Gen. Relativ. Gravit. **13**, 175 (1981).
[16]C. N. Yang, Phys. Rev. Lett. **33**, 445 (1974).
[17]P. Baekler, F. W. Hehl, and E. W. Mielke, *Proceedings of the Second Marcel Grossmann Meeting on General Relativity*, edited by R. Ruffini (North-Holland, Amsterdam, 1982), p. 413.
[18]P. Baekler, Phys. Lett. B **99**, 329 (1981).
[19]M. K. Prasad and P. Rossi, Phys. Rev. Lett. **46**, 806 (1981).
[20]M. K. Prasad, Commun. Math. Phys. **80**, 137 (1981).
[21]G. Y. Rainich, Trans. Am. Math. Soc. **27**, 106 (1925).
[22]C. W. Misner and J. A. Wheeler, Ann. Phys. (N.Y.) **2**, 525 (1957).
[23]I. M. Benn, T. Dereli, and R. W. Tucker, Gen. Relativ. Gravit. **13**, 581 (1981).
[24]J. Levine and J. D. Zund, Tensor N.S. **21**, 250 (1970).
[25]F. W. Hehl, private communication (1980).
[26]It has been speculated [F. W. Hehl, Y. Ne'eman, J. Nitsch, and P. von der Heyde, Phys. Lett. B **78**, 102 (1978)] that the "rotational" part of the Lagrangian (1.3) describes a *strong* spin-2 interaction [see, e.g., E. W. Mielke, Int. J. Theor. Phys. **19**, 189 (1980)] in the subnuclear domain.
[27]J. M. Charap and M. J. Duff, Phys. Lett. B **71**, 219 (1977).
[28]I. M. Benn, T. Dereli, and R. W. Tucker, J. Phys. A **15**, 849 (1982), compare with P. Baekler, "Spherically symmetric double-selfdual solutions of gravitational gauge field theories with quadratic field lagrangian," University of Cologne preprint, 1982.
[29]A. Tomimatsu, Progr. Theor. Phys. **66**, 526 (1981).
[30]E. B. Bogomol'nyi, Yad. Fiz. **24**, 861 (1976) [Sov. J. Nucl. Phys. **24**, 449 (1976)].
[31]B. S. DeWitt, in *Magic Without Magic: John Archibald Wheeler*, edited by J. R. Klauder (Freeman, San Francisco, 1972), p. 409.
[32]R. Jackiw, Rev. Mod. Phys. **49**, 681 (1977); **52**, 661 (1980).
[33]S. W. Hawking, Nucl. Phys. B **144**, 349 (1978).
[34]A. A. Tseytlin, Phys. Rev. D **26**, 3327 (1982).
[35]K. S. Stelle, Phys. Rev. D **16**, 953 (1977).
[36]Abdus Salam and J. Strathdee, Phys. Rev. D **18**, 4480 (1978).
[37]See, e.g., the review of P. Hasenfratz and J. Kuti, Phys. Rep. **40**, 77 (1978).
[38]E. Cremmer and J. Scherk, Nucl. Phys. B **108**, 409 (1976); L. V. Volkov and V. I. Tkach, Pis'ma Zh. Eksp. Teor. Fiz. **32**, 681 (1980) [JETP Lett. **32**, 668 (1980)]; compare also with Abdus Salam and J. Strathdee, Trieste Preprint IC/81/211.
[39]P. C. W. Davis and S. D. Unwin, Proc. Roy. Soc. London Ser. A **377**, 147 (1981).

Bäcklund transformations for spherically symmetric solutions to the self-dual Yang–Mills equations

R. K. Dodd and H. C. Morris
School of Mathematics, Trinity College, Dublin 2, Ireland

(Received 18 August 1982; accepted for publication 28 January 1983)

Prolongation techniques are used to derive a linear deformation problem for self-dual spherically symmetric solutions of the SU(2) Yang–Mills equations. Bäcklund transformations are then obtained which generate solutions to these equations. We use the transformations to generate some new solutions to the equations.

PACS numbers: 11.15. – q, 11.30.Jw

1. INTRODUCTION

The SU(2) Yang–Mills field equations are derived from the Lagrangian density[1]

$$\mathscr{L} = -\tfrac{1}{4}G^{\alpha\beta}G_{\alpha\beta},$$

$$\text{where } G^{\alpha\beta} = \partial_\alpha A_\beta - \partial_\beta A_\alpha - A_\alpha \wedge A_\beta. \tag{1.1}$$

The Euclidean Yang–Mills field equations are solved by self-dual gauge fields

$$G_{\alpha\beta} = {}^*G_{\alpha\beta}, \quad {}^*G_{\alpha\beta} = \epsilon_{\alpha\beta\gamma\delta}G^{\gamma\delta}, \tag{1.2}$$

which alternatively can be written as

$$G_{ij} = -\epsilon_{ijk}G_{k0}. \tag{1.3}$$

Witten[2] has obtained a large class of spherically symmetric instanton solutions by using an ansatz in the self-dual equations:

$$\begin{aligned} A^a_i &= [(\phi_2+1)/r^2]\epsilon_{iak}x_k + (\phi_1/r^3)(\delta_{ia}r^2 - x_ix_a) \\ &\quad + A_1x_ix_a/r^2, \\ A^a_0 &= A_0x_a/r \quad \text{with } \phi_i = \phi_i(x,t),\ A_i = A_i(x,t), \end{aligned} \tag{1.4}$$

which reduces the self-duality equations to the set

$$\begin{aligned} &\phi_{1,t} + A_0\phi_2 = \phi_{2,r} - A_1\phi_1, \\ &\phi_{1,r} + A_1\phi_2 = -\phi_{2,t} + A_0\phi_1, \\ &r^2(A_{1,t} - A_{0,r}) = 1 - \phi_1^2 - \phi_2^2. \end{aligned} \tag{1.5}$$

Several authors have investigated this set of equations besides Witten. Manton[3] has discussed the relation between Witten's multi-instanton solution and the Prasad–Sommerfield monopole solution of the classical monopole equations. Leznov and Saveliev,[4] using group-theoretic methods, have obtained an (L, A) pair of operators which can be used to investigate solutions to (1.5) with the techniques developed by Zakharov.[5]

In the second section of this paper we present a direct method using prolongation techniques[6] for determining a linear deformation problem which can be associated with (1.5). Then in Sec. 3 we derive a Bäcklund transformation for (1.5) and use it to derive some new solutions.

2. LINEAR DEFORMATION PROBLEMS

On the manifold M with local coordinates $(r,t,\phi_1,\phi_2, A_0,A_1)$ introduce the set of 2-forms

$$\begin{aligned} \alpha_1 &= d\phi_2\wedge dt + d\phi_1\wedge dr - (A_1\phi_1 + A_0\phi_2)\,dr\wedge dt, \\ \alpha_2 &= d\phi_1\wedge dt - d\phi_2\wedge dr - (A_0\phi_1 - A_1\phi_2)\,dr\wedge dt, \\ \alpha_3 &= dA_0\wedge dt + dA_1\wedge dr + r^{-2}(1-\phi_1^2-\phi_2^2)\,dr\wedge dt. \end{aligned} \tag{2.1}$$

The set of forms $\{\alpha_i, i = 1,2,3\}$ generate a differential ideal $E(M)$ of the exterior algebra of M called an exterior system. A solution manifold of $E(M)$ is a two-dimensional manifold S conveniently parametrized by x and t together with a map $f: S\to M$, which annihilates $E(M)$, $f^*E(M) = 0$. Then, provided x and t are in involution ($dx\wedge dt \neq 0$ on a solution manifold), the solutions of (2.1) are equivalent to the solutions of the original set of equations (1.5). To determine a scattering problem for (1.5), we require the Wahlquist–Estabrook prolongation of $E(M)$ to form an exterior system. Thus, we introduce the vector-valued 1-form

$$\omega = dy - Fy\,dr - Gy\,dt, \tag{2.2}$$

where $y = (y_1,...,y_n)$ and F and G are matrix-valued functions on M. Then, on the manifold N with local coordinates $(r,t,\phi_1,\phi_2,A_0,A_1,y)$, $\{\alpha_i,\omega\}$ generates an exterior system $E(N)$ on N provided

$$F = (X_0 + \phi_1X_1 + \phi_2X_2 + A_1X_3), \tag{2.3a}$$

$$G = (Y_0 - \phi_1X_2 + \phi_2X_1 + A_0X_3) \tag{2.3b}$$

and

$$X_{0,t} - Y_{0,r} + [X_0,Y_0] + r^{-2}X_3 = 0, \tag{2.4a}$$

$$X_{1,t} + X_{2,r} + [X_0,X_2] + [X_1,Y_0] = 0, \tag{2.4b}$$

$$X_{2,t} - X_{1,r} + [X_0,X_1] + [X_2,Y_0] = 0, \tag{2.4c}$$

$$X_{3,t} + [Y_0,X_3] = 0, \tag{2.4d}$$

$$X_{3,r} - [X_0,X_3] = 0, \tag{2.4e}$$

$$X_1 - [X_2,X_3] = 0, \tag{2.5a}$$

$$X_2 + [X_1,X_3] = 0. \tag{2.5b}$$

The r,t dependency of the matrix functions can be removed from these relations by defining new constant matrices,

$$\begin{aligned} &\overline{X}_0 = rX_0, \quad \overline{X}_1 = rX_1, \quad \overline{X}_3 = rX_3, \\ &\overline{X}_3 = X_3, \quad \overline{Y}_0 = rY_0. \end{aligned} \tag{2.6}$$

In terms of these new matrices (2.4) and (2.5) become

$$[\overline{X}_0,\overline{Y}_0] + \overline{X}_3 + \overline{Y}_0 = 0, \tag{2.7a}$$

$$[\overline{X}_0,\overline{X}_2] + [\overline{X}_1,\overline{Y}_0] - \overline{X}_2 = 0, \tag{2.7b}$$

$$[\overline{X}_0,\overline{X}_1] + [\overline{X}_2,\overline{Y}_0] + \overline{X}_1 = 0, \tag{2.7c}$$

$$[\overline{Y}_0,\overline{X}_3] = 0, \tag{2.7d}$$

$$[\overline{X}_0,\overline{X}_3] = 0, \tag{2.7e}$$

$$[\overline{X}_2,\overline{X}_3] - \overline{X}_1 = 0, \tag{2.8a}$$

$$[\overline{X}_1,\overline{X}_3] + \overline{X}_2 = 0. \tag{2.8b}$$

The set $\{\overline{X}_i,\overline{Y}_0\}$ thus generates a prolongation algebra, that is, a free Lie algebra with constraints, which is associated with Eq. (1.5). We now determine a homomorphism of the prolongation algebra into a finite-dimensional Lie algebra. A representation of this together with (2.2) then leads on solution manifolds of the exterior system to the required linear deformation problem. The simplest such homomorphism which yields nontrivial results is given by choosing

$$\overline{Y}_0 = -\overline{X}_3, \quad \overline{X}_0 = 0. \tag{2.9}$$

We are therefore left with the two relations of (2.8) which can be expressed in terms of a basis for sl (2,$\mathbb{C}$). Thus with

$$e_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad e_{-1} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad h = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.10}$$

we find that

$$\overline{X}_1 = \tfrac{1}{2}(e_1 + e_{-1}), \quad \overline{X}_2 = \tfrac{1}{2}i(e_{-1} - e_1), \quad \overline{X}_3 = -\tfrac{1}{2}ih. \tag{2.11}$$

We can introduce a parameter into the representation by the automorphism

$$e_1 \mapsto \zeta e_1, \quad e_{-1} \mapsto \zeta^{-1} e_{-1}, \quad h \mapsto h. \tag{2.12}$$

Thus on a solution manifold the 1-form ω is completely integrable, and the deformation problem we have obtained is

$$y_{,r} = \tfrac{1}{2}\begin{pmatrix} -iA_1 & \Phi^*\zeta r^{-1} \\ \Phi\zeta^{-1}r^{-1} & iA_1 \end{pmatrix} y,$$
$$y_{,t} = \tfrac{1}{2}\begin{pmatrix} -i(A_0 - r^{-1}) & i\zeta\Phi^* r^{-1} \\ -i\zeta^{-1}r^{-1}\Phi & i(A_0 - r^{-1}) \end{pmatrix} y, \tag{2.13}$$

where $\Phi = \phi_1 + i\phi_2$. The complete integrability conditions on (2.13) yield the equations (1.5). Other deformation problems can also be associated with (1.5). For example, the linear deformation problem

$$(r\partial_r - \eta\partial_\eta)w = \tfrac{1}{2}\begin{pmatrix} -irA_1 & \Phi^* r\eta \\ \Phi r^{-1}\eta^{-1} & iA_1 r \end{pmatrix} w,$$
$$w_{,t} = \tfrac{1}{2}\begin{pmatrix} -i(A_0 - r^{-1}) & i\Phi^*\eta \\ -r^{-2}\Phi\eta^{-1} & i(A_0 - r^{-1}) \end{pmatrix} w \tag{2.14}$$

corresponds to a solution of (2.5) in which $X_0 \neq 0$.

3. SOLUTIONS TO THE SELF-DUAL YANG–MILLS EQUATIONS

Before investigating the equations in detail it is worth establishing their relationship with some of the known solutions. If we introduce the complex coordinate $z = r + it$, then (2.13) can be written as

$$Y_{,z} = \tfrac{1}{4}\begin{pmatrix} -iA_1 - (A_0 - r^{-1}) & 2\Phi^*\zeta r^{-1} \\ 0 & iA_1 + (A_0 - r^{-1}) \end{pmatrix} Y,$$
$$Y_{,z^*} = \tfrac{1}{4}\begin{pmatrix} -iA_1 + (A_0 - r^{-1}) & 0 \\ 2\Phi\zeta^{-1}r^{-1} & iA_1 - (A_0 - r^{-1}) \end{pmatrix} Y. \tag{3.1}$$

In terms of the inhomogeneous coordinate $w = y_1 y_2^{-1}$, $Y = (y_1, y_2)^T$, the system can be written as

$$w_{,z} = 2Aw + B\zeta, \tag{3.2a}$$

$$w_{,z^*} = -2A^*w - B^*\zeta^{-1}w^2, \tag{3.2b}$$

where $A = -\tfrac{1}{4}[iA_1 + (A_0 - r^{-1})]$ and $B = \tfrac{1}{2}\Phi^* r^{-1}$. If we impose the condition that w is analytic, then $w_{,z^*} = 0$ and we get that

$$B = -2A\zeta^* w^{*-1}. \tag{3.3}$$

Then from (3.2a)

$$A = \frac{1}{2}\frac{d}{dz}\ln(|w|^2 - |\zeta^2|),$$
$$B = \frac{-\zeta^* w_{,z}}{(|w|^2 - |\zeta|^2)}. \tag{3.4}$$

If we now put $w = \zeta g$, then $B = g_{,z}(1 - |g|^2)^{-1}$. Define functions χ_1, χ_2, and ψ by

$$g_{,z} = \chi_1 - i\chi_2, \quad \psi = -\ln[(1 - |g|^2)/2r], \tag{3.5}$$

and we find that Eqs. (3.4) can be written as

$$\phi_1 = e^{\psi}\chi_1 \quad \text{and} \quad \phi_2 = e^{\psi}\chi_2. \tag{3.6}$$

This contains the multi-instanton solution of Witten.[2]

The one-monopole solution is defined by the functions

$$\phi_1 = 0, \quad \phi_2 = -\beta r/\sinh\beta r,$$
$$A_0 = r^{-1} - \beta + \coth\beta r, \quad A_1 = 0. \tag{3.7}$$

It is interesting to attempt to find a Bäcklund transformation which generates this solution from some initial state. To do this, we put the value of the functions given by (3.7) into (2.13) and integrate the equations to determine the wave functions. The result is the following:

$$Y = G(I + R_1\zeta + R_2\zeta^{-1})Y_0, \tag{3.8}$$

where

$$G = \tfrac{1}{2}[\tanh(\tfrac{1}{2}\beta r)]^{-1/2}[1 + \tanh(\tfrac{1}{2}\beta r)]I,$$
$$R_1 = -ie^{-\beta r}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$
$$R_2 = ie^{\beta r}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$
$$Y_0 = \begin{pmatrix} e^{i\beta t/2} & 0 \\ 0 & e^{-i\beta t/2} \end{pmatrix}. \tag{3.9}$$

Thus G is just a gauge transformation whereas $I + R_1\zeta + R_2\zeta^{-1}$ is a singular matrix function of ζ. The monopole solution is therefore generated from the initial solution

$$\phi_1 = 0, \quad \phi_2 = 0, \quad A_0 = r^{-1} - \beta, \quad A_1 = 0. \tag{3.10}$$

We present here a modification of the methods in Ref. 5 for obtaining Bäcklund transformations from the linear deformation problem given by Eq. (2.13). Let Y be a fundamental matrix solution of (2.13). Then we can write the equations as

$$Y_{,r}(\zeta)=A(\zeta)Y(\zeta),\quad Y_{,t}(\zeta)=B(\zeta)Y(\zeta), \tag{3.11a}$$

with

$$A=\begin{pmatrix}\alpha_1 & \beta_1\zeta\\ \beta_1^*\zeta^{-1} & \alpha_1^*\end{pmatrix},\quad B=\begin{pmatrix}\alpha_2 & \beta_2\zeta\\ \beta_2^*\zeta^{-1} & \alpha_2^*\end{pmatrix} \tag{3.11b}$$

and

$$\alpha_1=-\tfrac{1}{2}iA_1,\quad \beta_1=\tfrac{1}{2}\Phi^*r^{-1},$$
$$\alpha_2=-\tfrac{1}{2}i(A_0-r^{-1}),\quad \beta_2=i\beta_1. \tag{3.11c}$$

An investigation of (2.13) shows that it possesses the following symmetry:

$$Y(\zeta)\sigma Y^T(-\zeta)\sigma=I,\quad \sigma=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}. \tag{3.12}$$

Furthermore, a new matrix solution $\widetilde{Y}$ can be defined in terms of the given solution Y by the formula

$$\widetilde{Y}(\zeta^{*-1})=\sigma Y^*(\zeta^*). \tag{3.13}$$

Assume that (A_0,B_0,Y_0) and (A,B,Y) are Bäcklund-related. Then there exists a singular function $H(\zeta)$ such that

$$Y(\zeta)=H(\zeta)Y_0(\zeta). \tag{3.14}$$

Moreover, the Bäcklund transformation is given by

$$A=H_{,r}H^{-1}+HA_0H^{-1},$$
$$B=H_{,t}H^{-1}+HB_0H^{-1}. \tag{3.15}$$

The conditions (3.12) and (3.13) require that H must satisfy the relations

$$H(\zeta)\sigma H^T(-\zeta)\sigma=I, \tag{3.16a}$$
$$H(\zeta)=\sigma H^*(\zeta^{-1})\sigma. \tag{3.16b}$$

The relation (3.16b) forces H to have the representation

$$H=G\Big(I+\frac{R}{(\zeta-\mu)}+\frac{\zeta\sigma R^*\sigma}{(1-\zeta\mu^*)}\Big), \tag{3.17}$$

where $G=\sigma G^*\sigma$, $G=G(r,t)$, and $R=R(r,t)$.

The function R is a singular matrix function. Consider the case first of all when $\mu=0$. Then (3.17) becomes

$$H=G(I+R/\zeta+\zeta\sigma R^*\sigma), \tag{3.18}$$

which is easily seen to agree with (3.8). If we implement the condition (3.16a), then we find that

$$I=G\sigma(I-RR^{*T}-R^*R^T)G^T\sigma, \tag{3.19a}$$
$$R\sigma-\sigma R^T=0, \tag{3.19b}$$
$$R\sigma R^T=0. \tag{3.19c}$$

The conditions (3.19b) and (3.19c) require that

$$R=\begin{pmatrix}0 & b\\ c & 0\end{pmatrix}\quad \text{with } bc=0, \tag{3.20}$$

and (3.19a) imposes the constraints,

$$g_1g_2=0\quad\text{and}\quad (|g_1|^2+|g_2|^2)(1-|b|^2-|c|^2)=1,$$

where

$$G=\begin{pmatrix}g_1 & g_2\\ g_2^* & g_1^*\end{pmatrix}. \tag{3.21}$$

Introduce the notation $x^i=(r,t)$ and α_{0i},β_{0i} and α_i,β_i for, respectively, the components of A_0,B_0 and A_1,B_1. There are two possible Bäcklund transformations: case (a) $b=0,\beta_{0i}\neq 0$; case (b) $\beta_{0i}=0$. Case (a) further resolves into two subcases:

(ai) $c_{,x_i}+2\alpha_{0i}c+\beta_{0i}^*-\beta_{0i}c^2=0.$

In this case the Bäcklund transformation can be written as

$$\alpha_i=(1-|c|^2)(g_{2,x_i}g_2^*+g_{1,x_i}g_1^*)+|g_1|^2\lambda_i+|g_2|^2\lambda_i^*, \tag{3.22}$$
$$\beta_i=0,$$

where

$$\lambda_i=\alpha_{0i}+\alpha_{0i}|c|^2-cc_{,x_i}^*+\beta_{0i}^*c^*-\beta_{0i}c,\ g_1g_2=0,$$
$$(1-|c|^2)(|g_1|^2+|g_2|^2)=1. \tag{3.23}$$

First observe that the equation defining c is equivalent to (2.17) for the initial variables (those with a 0 suffix) upon introducing homogeneous coordinates

$$c=\Big(\frac{\delta y_{01}+\omega\bar{y}_{01}}{\delta y_{02}+\omega\bar{y}_{02}}\Big)^*_{\zeta=-1}. \tag{3.24}$$

The functions $(y_{0i},\bar{y}_{0i})$ are the components of the column vectors $(y_0,\bar{y}_0)$ comprising Y_0, the fundamental matrix solution for the initial functions. Since the gauge is at our disposal, we choose a suitable gauge function so that the transformation (3.22) assumes its simplest form. There is no loss in generality here because the transformations we are considering are only defined up to an arbitrary gauge. We choose $g_1\neq 0$ and real so that (3.23) defines g_1. The final form of the Bäcklund transformation becomes

$$\alpha\beta_i=\alpha_{0i}+\tfrac{1}{2}(\beta_{0i}^*c^*-\beta_{0i}c),$$
$$\beta_i=0, \tag{3.25}$$

where c is defined by (3.24).

(aii) $g_2=0.$

In this case the Bäcklund transformation becomes

$$\alpha_i=(1-|c|^2)(g_{1,x_i}g_1^*)+|g_1|^2\lambda_{1i},$$
$$\beta_i=g_1^2\lambda_{2i},$$

where

$$\lambda_{1i}=\alpha_{0i}+\alpha_{0i}|c|^2-\beta_{0i}c+\beta_{0i}^*c^*-cc_{,x_i}^*,$$
$$\lambda_{2i}=-2\alpha_{0i}c^*+\beta_{0i}-\beta_{0i}^*c^*2+c_{,x_i}^*, \tag{3.26}$$
$$(1-|c|^2)|g_1|^2=1.$$

For this case the constraint $\beta_2=i\beta_1$ imposes the additional condition

$$c_{,z}+c(\alpha_{01}-i\alpha_{02})-\beta_{01}c^2=0. \tag{3.27}$$

This equation can be solved up to an arbitrary function of z^*,

$$c=\exp\Big[-\int^z a(z')\,dz'\Big]\Big\{h(z^*)-\int^z\beta_{01}(z')\times\exp\Big[-\int^{z'}a(z'')\,dz''\Big]dz'\Big\}^{-1}, \tag{3.28}$$

where h is arbitrary and $a=\alpha_{01}-i\alpha_{02}$. If we choose the gauge so that (3.26) assumes its simplest form, we take g_1 real, and the transformation becomes

$$\alpha_i = (1-|c|^2)^{-1}(\tfrac{1}{2}(c^* c_{,x_i} - cc^*_{,x_2}) + \beta^*_{01} c^* - \beta_{0i} c + \alpha_{0i}(1+|c|^2),$$

$$\beta_i = (1-|c|^2)^{-1}(c^*_{,x_i} + \beta_{0i} - \beta^*_{0i} c^{*2} - 2\alpha_{0i} c^*). \quad (3.29)$$

Case (*b*) $\beta_{0i} = 0$: The Bäcklund transformation is

$$\alpha_i = (1-|b|^2-|c|^2)(g_{2,x_i} g_2^* + g_{1,x_i} g_1^*) + p_{3i}|g_1|^2 + p^*_{3i}|g_2|^2, \quad (3.30)$$

$$\beta_i = p_{2i} g_1^2 + p^*_{1i} g_2^2,$$

where

$$p_{1i} = b_{,x_i} - 2\alpha_{0i} b,$$
$$p_{2i} = c^*_{,x_i} - 2\alpha_{0i} c^*, \quad (3.31)$$
$$p_{3i} = \alpha_{0i} + \alpha_{0i}(|b|^2 + |c|^2) - (cc^*_{,x_i} + b^* b_{,x_i})$$

and

$$p_{1i} g_1^2 + p^*_{2i} g_2^2 = 0,$$
$$(|g_1|^2 + |g_2|^2)(1-|b|^2-|c|^2) = 1. \quad (3.32)$$

Since $g_1 g_2 = 0$ and $bc = 0$, we select the gauge so that $g_2 = 0$ and $g_1 \neq 0$ and real. Again there are two subcases to consider:

(bi) $c = 0$ and $(\ln b)_{,x_i} = 2\alpha_{0i}$:

$$\alpha_i = (1-|b^2|)^{-1}\left[\tfrac{1}{2}(bb^*_{,x_i} - b^* b_{,x_i}) + \alpha_{0i}(1+|b|^2)\right], \quad (3.33)$$

$$\beta_i = 0.$$

(bii) $b = 0$:

$$\alpha_i = (1-|c|^2)^{-1}\left[\tfrac{1}{2}(c^* c_{,x_i} - cc^*_{,x_i}) + \alpha_{0i}(1+|c|^2)\right], \quad (3.34)$$

$$\beta_i = (1-|c|^2)^{-1}(c^*_{,x_i} - 2\alpha_{0i} c^*).$$

It is easy to see that (bi) is just the identity transformation

$$\beta_i = \beta_{0i} \equiv 0, \quad \alpha_i = \alpha_{0i}. \quad (3.35)$$

For (bii) the relation $\beta_2 = i\beta$ leads to the following form for c:

$$c = h(z^*)\exp\left[-\int^x a(z')dz'\right], \quad (3.36)$$

where h is an arbitrary function and $a = \alpha_{01} - i\alpha_{02}$.

We shall only consider here some simple applications of these transformations. Transformations which belong to case (bii) we shall call B_+ transformations and those for case (a) B_- transformations. We shall denote transformations of type (ai) by an index 1 (B^1_-).

Consider the starting solution

$$(\mathrm{I}_+) \quad \alpha_{01} = 0, \quad \alpha_{02} = \tfrac{1}{2}i\beta, \quad \beta_{0i} = 0 \quad (3.37)$$

and apply a B_+ transformation such that c [Eq. (3.36)] only depends upon $x_1 \equiv r$. Then we find that

$$c = e^{\delta - \beta r}, \quad (3.38)$$

where δ is a complex number. The choice $\delta = -i\pi/2$ results in (3.34) generating the monopole:

$$(\mathrm{II}) \quad \alpha_1 = 0, \quad \alpha_2 = \tfrac{1}{2}i\beta\coth\beta r, \quad \beta_1 = \tfrac{1}{2}i\beta\,\mathrm{csch}\,\beta r. \quad (3.39)$$

An application of the B^1_- transformation to the monopole solution requires as we quickly see from (3.8), (3.9), and (3.24) that one of the following possibilities arise:

(i) $\delta = 0, \quad \omega \neq 0, \quad c = -ie^{-\beta r}$,

(ii) $\delta \neq 0, \quad \omega = 0, \quad c = -ie^{\beta r}$,

(iii) $\delta \neq 0, \quad \omega \neq 0$,

$$c = -i\cosh(\alpha - \tfrac{1}{2}i\beta t + \tfrac{1}{2}\beta r)\,\mathrm{sech}(\alpha - \tfrac{1}{2}i\beta t - \tfrac{1}{2}\beta r),$$

where

$$e^\alpha = (i\delta/\omega)^{1/2}.$$

Notice that from case (i) and (ii) we regain the starting solution

$$(\mathrm{I}_\pm) \quad \alpha_1 = 0, \quad \alpha_2 = \pm\tfrac{1}{2}i\beta, \quad \beta_i = 0. \quad (3.40)$$

Thus $B^1_- B_+$ for case (i) is the identity transformation. For a nontrivial transformation we consider case (iii). The result is the following solution:

$$\alpha_1 = \frac{i\beta}{2}\frac{\sin(\gamma + \beta t)}{\cosh(\beta r - \tau) + \cos(\gamma + \beta t)},$$
$$\alpha_2 = \frac{i\beta}{2}\frac{\sinh(\beta r - \tau)}{\cosh(\beta r - \tau) + \cos(\gamma + \beta t)}, \quad (3.41)$$
$$\beta_i = 0,$$

where $i\gamma = \alpha^* - \alpha, \quad \tau = \alpha^* + \alpha$.

The combination of functions a which occurs in (3.28) can be written in terms of the starting solution (3.41) as

$$\alpha_{10} - i\alpha_{20} = (\beta/2)\tanh(\beta z/2). \quad (3.42)$$

Equation (3.28) can now be solved to give

$$c = h(z)^*\,\mathrm{sech}(\beta z/2). \quad (3.43)$$

The particular choice $h(z^*) \equiv 1$ results in the solution

$$\alpha_1 = \tfrac{1}{4}i\beta\Delta^{-1}\sin\beta t,$$
$$\alpha_2 = \tfrac{1}{4}i\beta\Delta^{-1}\sinh\beta r,$$
$$\beta_1 = -\tfrac{1}{2}\beta\Delta^{-1}[\sinh(\beta r/2)\cos(\beta t/2) + i\cosh(\beta r/2)\sin(\beta t/2)], \quad (3.44)$$

where $\Delta = \sinh^2(\beta r/2) - \sin^2(\beta t/2)$.

We are currently investigating the significance of this and some other solutions.

[1]Greek indices range from 0 to 3, Latin indices from 1 to 3. If (x_α) is a local chart then $x_0 \equiv t$.

[2]E. Witten, Phys. Rev. Lett. **38**, 121–4 (1977).
[3]N. S. Manton, Nucl. Phys. B **135**, 319–32 (1978).
[4]A. N. Leznov and M. V. Saveliev, Commun. Math. Phys. **74**, 111–8 (1980).
[5]V. E. Zakharov, "The Inverse Scattering Method," in *Solitons*, edited by R. K. Bullough and P. J. Caudrey, Vol. 17 in Topics in Current Physics (Springer-Verlag, Berlin, 1980).
[6]H. D. Wahlquist and F. B. Estabrook, J. Math. Phys. **16**, 1–7 (1976).

Compact gauge group representations without triangular anomalies

L. Saliu and L. Tătaru
Department of Physics, University of Craiova, Craiova, Romania

(Received 8 February 1980; accepted for publication 17 March 1983)

By formulating the anomaly-free condition in terms of the fully symmetric third-order Casimir operators, we find all safe algebras and the algebraic equations satisfied by the highest weights of the anomaly-free representations of the only nonsafe algebras A_n, $n\geqslant 2$. By solving these equations for the irreducible representations of A_{n-1} [SU(n)], $n = 3, 4, 5$, and 6, we obtain the generating formulas of the highest weights for all anomaly-free representations of these groups. It turns out that for SU(n), $n\geqslant 5$ there is an infinite set of anomaly-free complex irreducible representations grouped as infinite series of such representations. Using the same technique, the infinite series of complex anomaly-free representations containing the lowest-dimension ones for SU(n), $n = 7, 8, 9$, and 10 are determined.

PACS numbers: 11.30.Ly, 11.15. – q, 11.10.Gh

I. INTRODUCTION

The condition of the renormalizability of gauge theories requires use of anomaly-free representations.[1-3]

The general condition for cancellation of the anomalies[3,4] is

$$\mathrm{Tr}(\{T_i^+, T_j^+\}T_k^+)\equiv C_{ijk}^+ = C_{ijk}^- \equiv \mathrm{Tr}(\{T_i^-, T_j^-\}T_k^-), \tag{1.1}$$

with $T_i^\pm$, the right- and left-handed parts of the Hermitian matrices T_i, specifying the couplings of gauge bosons to spinor-fermion fields through the interaction Lagrangian $gW_\mu^i\bar\psi\gamma^\mu T_i\psi$. $\{T_i^+\}$ and $\{T_i^-\}$ are representations not necessarily irreducible and equivalent to the same compact semisimple Lie algebras

$$[T_i^\pm, T_j^\pm] = c_{ij}{}^k T_k^\pm, \tag{1.2}$$

with the negative definite symmetric metric tensor $g_{ij} = c_{in}{}^m c_{jm}{}^n$. The condition (1.1) may be satisfied in the following different ways:

$$C_{ijk}^\pm = 0 \quad \text{on all representations;} \tag{1.3}$$

$$C_{ijk}^\pm = 0 \quad \text{on some particular representations } T_i^\pm; \tag{1.4a}$$

$$C_{ijk}^+ = C_{ijk}^- \neq 0 \quad \text{on some particular representations } T_i^\pm. \tag{1.4b}$$

An irreducible representation (IR) satisfying $C_{ijk} = 0$ is called a safe representation and an algebra with all safe representations is called a safe algebra.[4] On the basis of their condition (I.1) Georgi and Glashow[4] have established that all compact algebras except E_6 and A_n ($n\geqslant 2$) are safe.

Using different methods in Refs. 5 and 6 it has been proved that all safe algebras are characterized by the nonexistence of a genuine third-order Casimir operator, the algebra E_6 being also of this type. Also, in Refs. 5, 6, and 7 it has been proved that for the only nonsafe algebras A_n ($n\geqslant 2$), the coefficients C_{ijk} have the remarkable property that they can be factorized in two parts, a tensorial one which depends only on the algebra and a scalar one depending on IR; this last part, called anomaly, and consequently Eqs. (1.4a) and (1.4b) have been expressed in terms of the eigenvalues of the fully symmetric third-order Casimir operators.

The purpose of this paper is, in the light of some general arguments from group theory, to present a short and quite complete proof of the vanishing of C_{ijk}'s for all simple compact groups except SU(n), $n\geqslant 3$ [and SO(6) which is locally isomorphic to SU(4)], to obtain in the same manner the standard formula[5-7] for the anomaly of SU(n), and to discuss in detail, on this basis, the complex safe representations of these groups.

This paper is divided as follows. In Sec. II we show that the algebras without third-order invariant polynomials are safe ones. In Sec. III a compelling form of condition (1.4a) is given for IR's of SU(n). It consists in the vanishing of the eigenvalue of the fully symmetric third-order Casimir operator and it represents a constrained algebraic (Diophantic) equation in integers. Using this equation it turns out that there is an essential difference between $n = 3, 4$ and $n\geqslant 5$ cases. For SU(3) and SU(4) only real (self-contragradient) representations are safe. For SU(n), $n\geqslant 5$, in addition to the real representation, there is an infinite set of safe complex representations. For SU(5) and SU(6) we give the generating formulas for the highest weights of all safe representations, and for SU(n), $n = 7, 8, 9$, and 10 we offer some examples of infinite series of complex safe representations which contain the ones with minimal dimensions. The Appendix contains certain details associated with the proofs.

II. LIE GROUPS WITH SAFE ALGEBRAS

In order to obtain all safe algebras it is useful following Gruber and O'Raifeartaigh[8] to build up the Casimir operators

$$S_3 = C_{ijk}(\hat T)T^iT^jT^k = C^{ijk}(\hat T)T_iT_jT_k, \tag{2.1}$$

$$C_{ijk}(\hat T) = \mathrm{Tr}(\{\hat T_i, \hat T_j\}\hat T_k), \tag{2.2}$$

with $\hat T_i$ ($\hat T^i$) the matrix of the group generator T_i ($T^i = g^{ij}T_j$) in any particular representations $\hat T$. S_3's represent the main objects for finding the safe algebras because they vanish identically iff all representations $\hat T$ are safe. Therefore all we

 0022-2488/84/030673-05$02.50

have to do is to find all algebras for which any S_3 vanishes. In order to obtain all safe algebras one may use the Cartan–Weyl canonical basis of the algebra $\mathscr{G}$ of the group G, $\{H_i, E_\alpha, E_{-\alpha}\}$, $1\leqslant i\leqslant l\equiv\text{rank }\mathscr{G}$, $\alpha=(\alpha_1,\alpha_2,...,\alpha_l)\neq 0$, $\pm\alpha\in\Delta$, where Δ is the set of the roots of the algebra $\mathscr{G}$:

$$[H_i, H_j]=0,\quad [H_i,E_\alpha]=\alpha_i E_\alpha$$
$$[E_\alpha,E_\beta]=N_{\alpha\beta}E_{\alpha+\beta},\quad \text{where } N_{\alpha\beta}=0 \text{ if } \alpha+\beta\notin\Delta, \tag{2.3}$$
$$[E_\alpha,E_{-\alpha}]=\alpha^i H_i,\quad \text{where } \alpha^i=g^{ik}\alpha_k.$$

Let $\lambda=(\lambda_1,...,\lambda_l)$ be the weights of an IR of $\mathscr{G}$ with λ_i the eigenvalues of the representation operators H_i, and let $\tilde\lambda=(\tilde\lambda_1,...,\tilde\lambda_l)$ be its highest weight. In what follows we shall specify an IR of $\mathscr{G}$ by $\tilde\lambda$ or by nonnegative integers $(\Lambda_1,...,\Lambda_l)=\Lambda$ defined by $\tilde\lambda=\Sigma_1^l\Lambda_i\tilde\lambda^{(i)}$, where $\tilde\lambda^{(i)}$, $i=1,2,...,l$ are the highest weights of the fundamental IR's of $\mathscr{G}$. As S_3 is a Casimir operator, it has a unique eigenvalue $S_3(\tilde\lambda)$ on the space of IR. A Lie algebra is safe iff $S_3(\tilde\lambda)=0$ for all $\tilde\lambda$. If one defines s_3 as the reduction of S_3 on the Cartan subalgebra spanned by H_i's, i.e.,

$$s_3\equiv 2\,\mathrm{Tr}(\hat H_i\hat H_j\hat H_k)H^iH^jH^k, \tag{2.4}$$

then its eigenvalue $s_3(\tilde\lambda)$ on an IR is a homogeneous polynomial of third degree in $\tilde\lambda$.

The vanishing of $S_3(\tilde\lambda)$ for all algebras except A_n, $n\geqslant 2$, may be shown by using the well-known S Theorem[8–10] which for $S_3(\tilde\lambda)$ reads

Theorem: Let $S_3'(\sigma)$ be the polynomial $S_3(\tilde\lambda)$ written as a polynomial of $\sigma\equiv\tilde\lambda+\frac{1}{2}\Sigma$, $S_3'(\sigma)\equiv S_3(\sigma-\frac{1}{2}\Sigma)$, where Σ is sum of the positive roots for $\mathscr{G}$. The polynomial $S_3'(\sigma)$ is invariant under the Weyl group of $\mathscr{G}$. Moreover, $S_3(\tilde\lambda)$ and $s_3(\tilde\lambda)$ have the same homogeneous term of the highest degree.

On the other hand, according to a theorem due to Chevalley,[10] $S_3'(\sigma)$ as a polynomial of third degree invariant under the Weyl group generated by the reflections in the planes orthogonal to the roots, is a polynomial in so-called invariant fundamental polynomials of degree $\leqslant 3$. The degrees of the fundamental polynomials for all compact simple Lie algebras are listed below.[10,11]

$$A_n: 2,3,...,n+1\quad B_n \text{ and } C_n: 2,4,...,2n\quad D_n: 2,4,...,2(n-1) \text{ and } n\quad G_2: 2,6\quad F_4: 2,6,8,12\quad E_6: 2,5,6,8,9,12\quad E_7: 2,6,8,10,12,14,18\quad E_8: 2,8,12,14,18,20,24,30. \tag{2.5}$$

One may see from this list that for all simple Lie algebras with the possible exception of A_n, $n\geqslant 2$ and $D_3\approx A_3$, $S_3(\tilde\lambda)$ must be a polynomial of degree $\leqslant 2$ and consequently $s_3(\tilde\lambda)\equiv 0$, i.e., $\mathrm{Tr}(\hat H_i\hat H_j\hat H_k)=0$ for any representation of the former algebras.

The final step of the proof is contained in the following *Lemma*: If the symmetric invariant tensor $C_{ijk}(\hat T)\equiv\mathrm{Tr}(\{\hat T_i,\hat T_j\}\hat T_k)$ vanishes on the Cartan subalgebra then it vanishes on the whole algebra.

A possible proof of this lemma consists in using straightforwardly the Cartan–Weyl commutation relations (2.3) to show that

$$\mathrm{Tr}(E_\alpha H_jH_k)=\mathrm{Tr}([H_i,E_\alpha]H_jH_k)=0,$$
$$(\alpha_i+\beta_i)\mathrm{Tr}(E_\alpha E_\beta H_k)=\mathrm{Tr}([H_i,E_\alpha]E_\beta H_k)+\mathrm{Tr}(E_\alpha[H_i,E_\beta]H_k)=0, \tag{2.6}$$
$$\alpha_j\,\mathrm{Tr}(H_j\{E_\alpha,E_{-\alpha}\})=\sum_i\alpha^i\,\mathrm{Tr}(H_iH_j^2),$$
$$\gamma_i\,\mathrm{Tr}(\{E_\alpha,E_\beta\}E_\gamma)=N_{\alpha\beta}\,\mathrm{Tr}(\{E_{\gamma+\beta},E_\alpha\}H_i)+N_{\gamma\alpha}\,\mathrm{Tr}(\{E_{\gamma+\alpha},E_\beta\}H_i).$$

Now it is obvious that if $\mathrm{Tr}(H_iH_jH_k)=0$, then $\mathrm{Tr}(\{E_\alpha,H_j\}H_k)=\mathrm{Tr}(\{E_\alpha,E_\beta\}H_k)=\mathrm{Tr}(\{E_\alpha,E_\beta\}E_\gamma)=0$.

In conclusion, all the algebras A_1, D_n $(n\neq 3)$, B_n, C_n, G_2, F_4, $E_{6,7,8}$, i.e., of the groups SU(2), SO($2n$) $(n\neq 3)$, SO($2n+1$), Sp($2n$), and of all compact special groups are safe. Obviously, the most general safe algebra is a direct sum of simple safe algebras with no abelian components.

III. SAFE REPRESENTATIONS FOR SU(n)

According to the previous arguments the only Lie algebras nonsafe in all IR's are A_{n-1} algebras of the SU(n) groups. In order to obtain all safe representations and the anomaly-free condition for nonsafe representations, we must write down a relevant form for $C_{ijk}(\Lambda)$ in an appropriate basis. Obviously, the safe property of $C_{ijk}(\Lambda)$ does not depend on the chosen basis. On the other hand, we are interested in a safe condition having a general and simple algebraic form in the highest weight of IR. This is possible due to factorization of $C_{ijk}(\Lambda)$ in two parts: a tensorial one depending only on the algebra and the other depending on the representation, more exactly on the eigenvalue of an appropriate Casimir operator of third degree. This task may be more easily accomplished in the tensorial basis. Moreover, in this basis the eigenvalues of all Casimir operators take a form studied by many authors.[12–14]

The tensorial basis $\{A^i_j\}_{1<i,j<n}$ for A_{n-1} algebras is defined by

$$[A^i_j,A^k_l]=\delta^k_jA^i_l-\delta^i_lA^k_j,\quad \sum_i A^i_i=0. \tag{3.1}$$

(For the sake of simplicity we shall use the same symbols A^i_j for both the Lie group generators as well as their representation.) We shall prove in the Appendix that for an IR of SU(n) defined by the highest weight $\Lambda=(\Lambda_1,\Lambda_2,...,\Lambda_{n-1})$, one has the factorization[5]

$$\mathrm{Tr}(\{A^i_j,A^k_l\}A^m_n)\equiv\mathscr{T}^{ikm}_{jln}=N(\Lambda)I_3(\Lambda)(8/n(n^2-1)(n^2-4))\times[\delta^i_j\delta^k_l\delta^m_n-\tfrac{1}{2}n(\delta^i_j\delta^m_l\delta^k_n+\delta^k_l\delta^m_j\delta^i_n+\delta^m_n\delta^k_j\delta^i_l)+\tfrac{1}{2}n^2(\delta^k_j\delta^m_l\delta^i_n+\delta^m_j\delta^i_l\delta^k_n)], \tag{3.2}$$

where

$$N(\Lambda)=\prod_{m<k}\Big[1+(k-m)^{-1}\sum_{j=m}^{k-1}\Lambda_j\Big],\quad k=1,2,...,n \tag{3.3}$$

is the dimension of the IR and $I_3(\Lambda)$ is the eigenvalue of the third-degree symmetric Casimir operator

$$I_3=(1/3!)P(A^{i_1}_{i_2}A^{i_2}_{i_3}A^{i_3}_{i_1}), \tag{3.4}$$

P denoting the sum over all permutations of Λ's. Thus, the vanishing of $I_3(\Lambda)$ is a necessary and sufficient condition for an IR of A_{n-1} [SU(n)] to be safe. It is worthwhile noticing that Eq. (3.2) represents an explicit proof of the fact that the groups SU(n), $n \geqslant 3$ can have only one symmetric invariant tensor. The right hand of Eq. (3.2) is manifestly split in two parts, one which depends only on the representation, namely, $N(\Lambda) I_3(\Lambda)$, and one contained in the bracket which depends only on the algebra and which can be evaluated from $\mathcal{T}^{ikm}_{jln}$ on the fundamental n-dimensional representaion of SU(n). $I_3(\Lambda)$ has the form (see Appendix)

$$I_3(\Lambda) = \sum_1^n l_i^3, \tag{3.5}$$

with l_i completely determined in terms of nonnegative integers $(\Lambda_1, \Lambda_2, ..., \Lambda_{n-1})$ by the equation

$$\Lambda_i = l_i - l_{i+1} - 1, \quad l_i > l_{i+1} \quad \sum_1^n l_i = 0, \tag{3.6}$$

i.e., by

$$l_i = \frac{1}{n} \sum_{k=1}^{n} (n-k)(\Lambda_k + 1) - \sum_{k=1}^{i-1} k(\Lambda_k + 1). \tag{3.7}$$

It follows now that in order to obtain all safe representations for SU(n) we have to solve the algebraic equations in integers

$$\sum_1^n l_i^3 = 0, \quad \sum_1^n l_i = 0, \quad l_i > l_{i+1}. \tag{3.8}$$

A real (self-contragradient) representation corresponds to a highest weight which satisfies $\Lambda_i = \Lambda_{n-i}$; hence, $l_i = -l_{n-i+1}$, i.e., any real IR is a safe one. The general scheme for obtaining safe representations of SU(n) consists of two steps:

(i) Solve the homogeneous equations (3.8) with l_i integers.

(ii) Take l_i such that $l_i > l_{i+1}$, and for any distinct multiplet of relatively prime integers l_i, $i = 1,2,...,n$ associate an infinite series of safe representations having the highest weights $\Lambda_j = (l_j - l_{j+1})t - 1, j = 1,2,...,n-1$, where the homogeneity parameter t takes all values leading to nonnegative integers Λ_j. We note that Eqs. (3.8) become

$$l_1 l_2 l_3 = 0, \quad l_1 + l_2 + l_3 = 0 \quad \text{for SU(3)} \tag{3.9}$$

and

$$\begin{gathered}(l_1 + l_2)(l_2 + l_3)(l_3 + l_4) = 0,\\ l_1 + l_2 + l_3 + l_4 = 0 \quad \text{for SU(4)}.\end{gathered} \tag{3.10}$$

The solution of these equations have the form $(l_1, l_2, l_3) = (l, 0, -l)$, $(\Lambda_1, \Lambda_2) = (l-1, l-1)$, $l = 1, 2, ...$ for SU(3) and $(l_1, l_2, l_3, l_4) = (l, k, -k, -l)$, $(\Lambda_1, \Lambda_2, \Lambda_3) = (l-k-1, 2k-1, l-k-1)$, $l > k$, both integers or half-integers for SU(4). It turns out that for SU(3) and SU(4) only real representations are safe.

Equations (3.8) become

$$\begin{gathered}(l_1 + l_2)(l_2 + l_3) + l_4 l_5 (l_4 + l_5) = 0,\\ \sum_1^5 l_i = 0 \quad \text{for SU(5)}\end{gathered} \tag{3.11}$$

and

$$\begin{gathered}(l_1 + l_2)(l_1 + l_3)(l_2 + l_3) + (l_4 + l_5)(l_5 + l_6)(l_4 + l_6) = 0,\\ \sum_1^6 l_i = 0\end{gathered} \tag{3.12}$$

for SU(6).

In order to solve these equations it is convenient to rewrite them as

$$\begin{gathered}\begin{vmatrix} l_1 + l_2 & l_4 & 0 \\ 0 & l_2 + l_3 & l_5 \\ l_4 + l_5 & 0 & l_3 + l_1 \end{vmatrix} = 0,\\ (l_1 + l_2) p + l_4 q = 0,\\ (l_2 + l_3) q + l_5 r = 0,\\ (l_4 + l_5) p + (l_3 + l_1) r = 0\end{gathered} \tag{3.13}$$

for SU(5) and

$$\begin{gathered}\begin{vmatrix} l_1 + l_2 & l_4 + l_5 & 0 \\ 0 & l_2 + l_3 & l_5 + l_6 \\ l_6 + l_4 & 0 & l_3 + l_1 \end{vmatrix} = 0,\\ (l_1 + l_2) p + (l_4 + l_5) q = 0,\\ (l_2 + l_3) q + (l_5 + l_6) r = 0,\\ (l_6 + l_4) p + (l_3 + l_1) r = 0\end{gathered} \tag{3.14}$$

for SU(6). The general solutions of these equations in integers are

$$\begin{aligned} l_1 &= k(q^2 r - q r^2 + r^2 p - r p^2),\\ l_2 &= k(q^2 r - r^2 p + r p^2 - p q^2),\\ l_3 &= k(-q^2 r + q r^2 - r^2 p + p q^2),\\ l_4 &= k(p^2 q + r^2 p - 2pqr),\\ l_5 &= k(-p^2 q - q^2 r + 2pqr),\end{aligned} \tag{3.15}$$

with p, q, r relatively prime integers and k a rational number for SU(5), and

$$\begin{aligned} l_1 &= s((r-q)/q) P(mR + nQ) - kRQ,\\ l_2 &= s((p-r)/r) Q(kR + nP) - mPR,\\ l_3 &= s((q-p)/p) R(kQ + mP) - nPQ,\\ l_4 &= s\,k\,Q\,R,\\ l_5 &= s\,m\,P\,R,\\ l_6 &= s\,n\,P\,Q,\end{aligned} \tag{3.16}$$

where

$$\begin{aligned} P &= q(p^2 + qr - 2pr),\\ Q &= r(q^2 + pr - 2pq),\\ R &= p(r^2 + pq - 2qr),\end{aligned} \tag{3.17}$$

with p, q, r relatively prime integers, s a rational number, and k, m, n three rational numbers satisfying the constraint $k + m + n = 0$ for SU(6).

These equations will give us all anomaly-free IR's for SU(5) and SU(6) if we follow the second step described above. According to Eqs. (3.15)–(3.17), in addition to the real IR's for SU(5) and SU(6) there are infinite series of complex anomaly-free IR's. From these equations one can obtain easily the following examples of infinite series of complex safe representations, together with their contragradients

		Λ_1	Λ_2	Λ_3	Λ_4	Λ_5	$N(\Lambda)$
SU(5)	$t-1$	$8t-1$	$4t-1$	$4t-1$	–	$1\,357\,824\,t^{10}$
	$2t-1$	$9t-1$	$2t-1$	$6t-1$	–	$3\,048\,474\,t^{10}$
	$2t-1$	$25t-1$	$16t-1$	$10t-1$	–	$92\,925\,085\,500\,t^{10}$
SU(6)	$t-1$	$6t-1$	$t-1$	$t-1$	$5t-1$	$374\,556\,t^{15}$
	$t-1$	$11t-1$	$t-1$	$t-1$	$9t-1$	$108\,645\,537\,t^{15}$
	$t-1$	$7t-1$	$t-1$	$4t-1$	$4t-1$	$28\,514\,304\,t^{15}$

Similar methods for solving Eqs. (3.8) can be developed for SU(n), $n \geqslant 7$. Here we shall give only some examples of complex safe representations of SU(n), $n = 7, 8, 9, 10$, $t = 1, 2, \ldots$.

n^Λ	Λ_1	Λ_2	Λ_3	Λ_4	Λ_5	Λ_6	Λ_7	Λ_8	Λ_9	$N(\Lambda)$
7	$t-1$	$5t-1$	$3t-1$	$t-1$	$3t-1$	$4t-1$				$823\,350\,528\,t^{21}$
8	$t-1$	$2t-1$	$4t-1$	$t-1$	$t-1$	$t-1$	$4t-1$			$12\,360\,348\,t^{28}$
9	$t-1$	$4t-1$	$t-1$	$3t-1$	$t-1$	$t-1$	$2t-1$	$4t-1$		$12\,322\,252\,800\,t^{36}$
10	$t-1$	$t-1$	$t-1$	$4t-1$	$t-1$	$t-1$	$t-1$	$t-1$	$3t-1$	$19\,423\,404\,t^{45}$

It is interesting to note that for $t = 1$ all these representations and the first ones from those given for SU(5) and SU(6) are the complex safe representations with the minimal dimensions for the corresponding groups. The fact that the dimensions of the complex safe representations of SU(n), $n \geqslant 5$ are extremely large is a very strong argument against the unification theories in which the fermionic multiplet belongs to a single complex irreducible representation of SU(n). On the other hand, for reducible representations the anomaly-free condition can be satisfied for very low dimensions, as for example in the model of Georgi and Glashow[15] based on SU(5) where the fermions belongs to the anomaly-free complex reducible representation $5^* + 10$.

So far we have studied only safe representations but an anomaly-free theory can be obtained also for nonsafe irreducible representations of SU(n) if they satisfy [see Eqs. (1.5) and (3.2)]

$$N(\Lambda^+)I_3(\Lambda^+) = N(\Lambda^-)I_3(\Lambda^-), \tag{3.18}$$

where $\Lambda^{\pm}$ are the highest dominant weights of the right and left IR's. Of course this condition is trivially satisfied when the right and left IR's are equivalent, but it may be satisfied also for some inequivalent IR's.

For the reducible representation the anomaly-free condition for the algebra A_n takes the form

$$\sum_{\Lambda^+} N(\Lambda^+)I_3(\Lambda^+) = \sum_{\Lambda^-} N(\Lambda^-)I_3(\Lambda^-) \begin{cases} = 0, \\ \neq 0, \end{cases} \tag{3.19}$$

where $\Lambda^{\pm}$ are the highest weights of all IR's contained in the right and left reducible representations.

When the abelian components occur, none of the previous formulas, except the general condition (1.1) can be applied. From Eq. (1.1) it may be shown that in this case an anomaly-free representation must be reducible or its left and right representations must be equivalent. The right and left components $T_0^{\pm}$ of the generator of the representation of the abelian group have to satisfy the anomaly-free condition

$$C^+_{ij0} = C^-_{ij0}, \quad C^+_{000} = C^-_{000}, \tag{3.20}$$

where $\{T_i\}$, $i = 1,2,\ldots$ are the anomaly-free representation generators of the simple group G. Choosing the normalization $\mathrm{Tr}\, T_i T_j \sim \delta_{ij}$, Eqs. (3.20) become

$$\begin{aligned} &\sum N(\Lambda^{(a)}_+)I_2(\Lambda^{(a)}_+)t^{(a)} = \sum N(\Lambda^{(b)}_-)I_2(\Lambda^{(b)}_-)t^{(b)}_-, \\ &\sum N(\Lambda^{(a)}_+)t^{(a)^3}_+ = \sum N(\Lambda^{(b)}_-)t^{(b)^3}_-, \end{aligned} \tag{3.21}$$

where $\Lambda^{(a,b)}_{\pm}$ denote the IR's of G and $I_2(\Lambda)$ and t are the eigenvalues of the second-degree Casimir operator and T_0, respectively, on the IR's Λ of G.

ACKNOWLEDGMENTS

We are grateful to Professor O. Gherman and Dr. G. Steinbrecher for stimulating discussions.

APPENDIX

In this Appendix we shall prove Eqs. (3.2) and (3.5). The tensor

$$\mathscr{T}^{ikm}_{jln} = \mathrm{Tr}(\{A^i_j, A^k_l\}A^m_n) \tag{A1}$$

in an IR, as an invariant tensor, can be expressed by products of Kronecker δ's. This tensor has the properties

$$\mathscr{T}^{ikm}_{iln} = 0, \tag{A2a}$$

$$\mathscr{T}^{ikl}_{jmn} = \mathscr{T}^{kim}_{ljn} = \mathscr{T}^{imk}_{jnl}, \tag{A2b}$$

so that it must have the structure

$$\begin{aligned} \mathscr{T}^{ikm}_{jln}(\Lambda) = {} & a_1\delta^i_j\delta^k_l\delta^m_n + a_2(\delta^i_j\delta^m_l\delta^k_n \\ & + \delta^k_l\delta^m_j\delta^i_n \\ & + \delta^m_n\delta^k_j\delta^i_l) + a_3(\delta^k_j\delta^m_l\delta^i_n \\ & + \delta^m_j\delta^i_l\delta^k_n), \end{aligned} \tag{A3}$$

where owing to (A2a) and (A2b)

$$a_2 = -\tfrac{1}{2}n\, a_1, \quad a_3 = \tfrac{1}{4}n^2\, a_1. \tag{A4}$$

The only unknown coefficient in Eq. (A3) may be obtained by relating $\mathscr{T}$ to the eigenvalues $I_3(\Lambda)$ of the third-degree symmetric Casimir operator (3.4). In this way one obtains Eqs. (3.2).

The eigenvalue $I_3(\Lambda)$ may be calculated by using the tensorial basis (3.1) and the definition of the highest weight.

By inspection of Eqs. (3.1) and (2.3) one can easily see that $A^{\,i}_{\,i}$ (no sum), $A^{\,i}_{\,j}$ and $A^{\,j}_{\,i}$, $i<j$ correspond to the generators H_i of the Cartan subalgebra, to the rising operators E_α and the lowering operators $E_{-\alpha}$, respectively.

For the highest weight we shall define the vector $|\psi_0\rangle$ satisfying

$$A^{\,i}_{\,i}|\psi_0\rangle = m_i|\psi_0\rangle \ \text{(no sum)}, \quad A^{\,i}_{\,j}|\psi_0\rangle = 0 \ \text{for } i<j,$$
$$\sum_i m_i = 0, \quad m_i \geqslant m_{i+1}. \tag{A5}$$

We note that for the Young tableaux $[f_1, f_2, ..., f_n]$ associated with the previous IR we have $m_i = f_i$ for U(n) and $m_i = f_i - (1/n)\sum f_i$ for SU(n).

In terms of the Casimir operators C_2 and C_3 defined as

$$C_2 = A^{\,i_1}_{\,i_2} A^{\,i_2}_{\,i_1}, \quad C_3 = A^{\,i_1}_{\,i_2} A^{\,i_2}_{\,i_3} A^{\,i_3}_{\,i_1}, \tag{A6}$$

I_3 can be written as

$$I_3 = C_3 - \tfrac{1}{2} n\, C_2. \tag{A7}$$

The eigenvalue $C_p(\Lambda)$ of C_p can be calculated by noticing that[14]

$$C_p = (B^{(p-1)})^i_j A^{\,j}_{\,i}, \quad [A^{\,i}_{\,j}, B^{\,k}_{\,l}] = \delta^k_j B^{\,i}_{\,l} - \delta^i_l B^{\,k}_{\,j}, \tag{A8}$$
$$(B^{(p-1)})^i_j|\psi_0\rangle = 0 \ \text{for } i<j,$$
$$(B^{(q)})^i_i|\psi_0\rangle = (B^{(q-1)})^i_j A^{\,j}_{\,i}|\psi_0\rangle = \sum_{j=1}^{n} a_{ij}(B^{(q-1)})^j_j|\psi_0\rangle,$$

where

$$(B^{(p-1)})^i_j = A^{\,i}_{\,i_2} A^{\,i_2}_{\,i_3} \cdots A^{\,i_{p-1}}_{\,j} \tag{A9}$$

and

$$a_{ij} = (m_i + n - i)\delta_{ij} = (l_i + (n-1)/2)\delta_{ij} - \theta_{ij}, \tag{A10}$$
$$l_i = m_i + \frac{n+1}{2} - i, \quad \sum_1^n l_i = 0, \quad \theta_{ij} = \begin{cases} 1 & \text{for } i>j, \\ 0 & \text{for } i\leqslant j. \end{cases}$$

Therefore

$$C_p|\psi_0\rangle = \sum_{i=1}^{n}(m_i + n + 1 - 2i)(B^{(p-1)})^i_i|\psi_0\rangle = \sum_{i,j=1}^{n}(a^p)_{ij} \tag{A11}$$

and

$$I_3(\Lambda) = \sum_{i,j=1}^{n}\left[(a^3)_{ij} - \frac{n}{2}(a^2)_{ij}\right] = \sum_1^n l_i^3, \tag{A12}$$

where the nonnegative integers $\{\Lambda_j\}$,

$$\Lambda_j = m_j - m_{j+1} = l_j - l_{j+1} - 1, \quad j = 1,2,...,n-1, \tag{A13}$$

as well as $\{m_i\}$, $m_i \geqslant m_{i+1}$ or $\{l_i\}$, $l_i > l_{i+1}$, $i = 1,2,...,n$ characterize an IR of SU(n).

Finally we note that Eqs. (3.2) may be rewritten in the standard form given by Banks and Georgi[7] and Okubo[6] as follows. Define

$$T_a = \tfrac{1}{2}(\lambda_a)_{ij} A^{\,j}_{\,i}, \tag{A14}$$

where the standard Hermitian traceless $n\times n$ matrices $\{\lambda_a\}$, $a = 1,2,...,n^2-1$ satisfy

$$\begin{aligned} &\mathrm{Tr}(\lambda_a \lambda_b) = 2\delta_{ab}, \\ &(\lambda_a)_{ij}(\lambda_b)_{kl} = 2(\delta_{il}\delta_{jk} - (1/n)\delta_{ij}\delta_{kl}). \end{aligned} \tag{A15}$$

Thus Eq. (3.2) may be put in the form

$$\mathrm{Tr}(\{T_a, T_b\}T_c) = K(\Lambda)d_{abc}, \tag{A16}$$

where

$$d_{abc} \equiv \tfrac{1}{4}\mathrm{Tr}(\{\lambda_a, \lambda_b\}\lambda_c) \tag{A17}$$

and

$$K(\Lambda) \equiv (n/(n^2-1)(n^2-4))N(\Lambda)I_3(\Lambda) \tag{A18}$$

has been called the anomaly coefficient.[6,7]

[1] C. Bouchiat, J. Iliopoulos, and Ph. Mayer, Phys. Lett. B **38**, 519 (1972); D. Gross and R. Jackiw, Phys. Rev. D **6**, 477 (1972).
[2] S. L. Adler, Phys. Rev. **177**, 2426 (1969); J. S. Bell and R. Jackiw, Nuovo Cimento **51**, 47 (1969); J. Wess and B. Zumino, Nucl. Phys. B. **37**, 95 (1971).
[3] W. A. Bardeen, Phys. Rev. **184**, 1848 (1969).
[4] H. Georgi and S. L. Glashow, Phys. Rev. D **6**, 429 (1972).
[5] L. Saliu and L. Tătaru, Lett. Nuovo Cimento **7**, 253 (1973).
[6] S. Okubo, Phys. Rev. D **16**, 3528 (1977).
[7] J. Banks and H. Georgi, Phys. Rev. D **14**, 1159 (1976).
[8] B. Gruber and L. O'Raifeartaigh, J. Math. Phys. **5**, 1796 (1964); see also L. C. Biedenharn, J. Math. Phys. **4**, 436 (1963).
[9] G. Racah, *Group Theory and Spectroscopy* (Princeton Lecture Notes, 1951).
[10] C. Chevalley, *Theory of Lie Groups*, Vol. I (Princeton U.P., Princeton, NJ, 1946), Vols. II and III (Hermann, Paris, 1955); D. P. Zelobenko, *Compact Lie Groups and Their Representations*, Trans. Math. Monographs, No. 40 (American Math. Society, Providence, RI, 1973).
[11] A. Borel and C. Chevalley, Mem. Amer. Math. Soc. **14**, 1 (1955).
[12] M. Micu, Nucl. Phys. **60**, 353 (1964).
[13] M. Umezawa, Nucl. Phys. **48**, 111 (1963); **53**, 54 (1964); **57**, 65 (1964).
[14] A. M. Perelomov and V. S. Popov, Yad. Fiz. **3**, 924, 1127 (1966) [Sov. J. Nucl. Phys. 3, 676, 819 (1966)].
[15] H. Georgi and S. L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).

The theory of superselection rules. II. Sectors of the free electromagnetic field

G. O. S. Ekhaguere
Department of Mathematics, University of Ibadan, Ibadan, Nigeria

(Received 24 August 1982; accepted for publication 25 February 1983)

In the first part of this study [J. Math. Phys. **19**, 1751 (1978)], a wide class of inequivalent irreducible $*$-representations of the C^*-algebra $\mathfrak{A}$ of quasilocal observables of the free electromagnetic field was constructed and analyzed. A $*$-representation π is called *positive* if it satisfies the spectrum condition, i.e., if space-time automorphisms are implemented in π by a strongly continuous unitary representation of Minkowski space M^4 whose infinitesimal generator has spectrum contained in the forward light cone. Here, we characterize the subclass of the class of $*$-representations constructed in Ref. 1, which consists of positive $*$-representations. This then leads to the exhibition of new superselection sectors, i.e., equivalence classes of positive $*$-representations.

PACS numbers: 11.90. + t

I. INTRODUCTION

In this paper, unless there is a statement to the contrary, our notation and assumptions will be as in Ref. 1. However, for the convenience of the reader, we begin the discussion with a cursory review of the definitions of some of the function spaces and operators employed in Ref. 1, which will also be used in the following.

Let $\mathscr{D}(\mathbb{R}^3)$ be the real Schwartz space of C^∞ functions with compact support in $\mathbb{R}^3$, and let $\mathscr{D}^4(\mathbb{R}^3)$ denote the fourfold Cartesian product of $\mathscr{D}(\mathbb{R}^3)$ with itself. We shall denote by $\mathscr{D}_0^4(\mathbb{R}^3)$ the subspace of $\mathscr{D}^4(\mathbb{R}^3)$ consisting of all $f=(f_0,f_1,f_2,f_3)\in\mathscr{D}^4(\mathbb{R}^3)$ such that each component f_μ of f has a Fourier transform $\tilde{f}_\mu$ given by

$$\tilde{f}_\mu(\mathbf{p}) = |\mathbf{p}|\tilde{f}_{\mu 0}(\mathbf{p}) + \sum_{j=1}^{3} p_j\tilde{f}_{\mu j}(\mathbf{p}),$$
$$\mathbf{p}=(p_1,p_2,p_3)\in\mathbb{R}^3,$$

where $f_{\mu\nu}\in\mathscr{D}(\mathbb{R}^3)$ and $f_{\mu\nu}+f_{\nu\mu}=0$. Clearly, each $f\in\mathscr{D}_0^4(\mathbb{R}^3)$ satisfies $\tilde{f}(0)=0$.

Let Δ denote the Laplace operator in three variables and set $(-\Delta)^{1/4}=C$. We may now introduce two Hilbert spaces $\mathscr{H}^*$ and $\mathscr{H}$ defined as follows.

$\mathscr{H}^*$ is the completion of $\mathscr{D}_0^4(\mathbb{R}^3)$ in the topology derived from the norm

$$\|\cdot\|_{\mathscr{H}^*}:\mathscr{D}_0^4(\mathbb{R}^3)\to\mathbb{R}_+=[0,\infty),$$
$$f=(f_0,f_1,f_2,f_3)$$
$$\mapsto\|f\|_{\mathscr{H}^*}=\left[\sum_{\mu=0}^{3}\int d\mathbf{x}\,|(C^{-1}f_\mu)(\mathbf{x})|^2\right]^{1/2}.$$

Similarly, $\mathscr{H}$ is the completion of $\mathscr{D}_0^4(\mathbb{R}^3)$ in the topology derived from the norm

$$\|\cdot\|_{\mathscr{H}}:\mathscr{D}_0^4(\mathbb{R}^3)\to\mathbb{R}_+=[0,\infty),$$
$$f=(f_0,f_1,f_2,f_3)$$
$$\mapsto\|f\|_{\mathscr{H}}=\left[\sum_{\mu=0}^{3}\int d\mathbf{x}\,|(Cf_\mu)(\mathbf{x})|^2\right]^{1/2}.$$

The Hilbert space $\mathscr{H}^*$ is dual to the Hilbert space $\mathscr{H}$ in the pairing

$$\langle\cdot,\cdot\rangle:\mathscr{H}^*\times\mathscr{H}\to\mathbb{R},$$
$$(f,g)\mapsto\langle f,g\rangle=\sum_{\mu=0}^{3}\langle C^{-1}f_\mu,Cg_\mu\rangle_0,$$

where $\langle\cdot,\cdot\rangle_0$ is the inner product of $L^2(\mathbb{R}^3,d\mathbf{x})$.

In Ref. 1, we constructed a class of inequivalent irreducible $*$-representations ρ_T of the C^*-algebra $\mathfrak{A}$ of quasilocal observables of the free electromagnetic field by means of $*$-automorphisms γ_T of $\mathfrak{A}$; the $*$-automorphisms γ_T are induced by simplectic transformations T which belong to a certain class $\mathfrak{G}_0(\mathscr{H}^*,\mathscr{H})$ defined in Sec. 4 of Ref. 1. The $*$-representations ρ_T and the $*$-automorphisms γ_T of $\mathfrak{A}$ are related as follows:

$$\rho_T=\rho_0\circ\gamma_T,$$

where ρ_0 is the identity $*$-representation of $\mathfrak{A}$. Each simplectic transformation T in $\mathfrak{G}_0(\mathscr{H}^*,\mathscr{H})$ is an operator-valued matrix whose full representation is given in (A1).

In this paper, we are interested in a characterization of the subset of $\{\rho_T:T\in\mathfrak{G}_0(\mathscr{H}^*,\mathscr{H})\}$, consisting of $*$-representations of $\mathfrak{A}$, which are positive in the sense of Definition (1.5) below.

We may view the abstract C^*-algebra $\rho_T(\mathfrak{A})$ as a concrete C^*-algebra whose representation Fock space $\mathscr{F}^{\rho_T}$, $T\in\mathfrak{G}_0(\mathscr{H}^*,\mathscr{H})$, may be described as follows.

Let T^+ and T^- denote the linear transformations

$$T^+:\quad \mathscr{H}^*\oplus\mathscr{H}\to\mathscr{H}^*$$
$$f\oplus g\mapsto T^+(f\oplus g)=C^{-1}(I+\mathscr{A})Cf+C^{-1}\mathscr{L}C^{-1}g$$

and

$$T^-:\quad \mathscr{H}^*\oplus\mathscr{H}\to\mathscr{H}$$
$$f\oplus g\mapsto T^-(f\oplus g)=C\mathscr{B}Cf+C(I+\mathscr{T})C^{-1}g\cdot$$

Recall that to each fixed pair (f,g) in $\mathscr{H}^*\times\mathscr{H}$, there corresponds a unique solution (F,G) of the equations

$$\frac{\partial F}{\partial t}=G,\quad \frac{\partial G}{\partial t}=\Delta F \tag{1.1}$$

with Cauchy data

$$F(0,\mathbf{x})=f(\mathbf{x}),\quad G(0,\mathbf{x})=g(\mathbf{x}),\quad \mathbf{x}\in\mathbb{R}^3.$$

 0022-2488/84/030678-06$02.50

Hence, there is a unique pair (F_T, G_T) which solves (1.1) such that

$$F_T(0,\mathbf{x}) = T^+(f\oplus g)(\mathbf{x}), \quad \mathbf{x}\in\mathbb{R}^3, \tag{1.2}$$

and

$$G_T(0,\mathbf{x}) = T^-(f\oplus g)(\mathbf{x}), \quad \mathbf{x}\in\mathbb{R}^3, \tag{1.3}$$

for each fixed pair (f,g) in $\mathscr{H}^*\times\mathscr{H}$. We mention next the following relationship between the vector functions F_T and $T^+(f\oplus g)$. Let $T^+(f\oplus g)^\sim$ be the uniquely determined four-component vector function, with complex-valued components, given by

$$T^+(f\oplus g)(\mathbf{x}) = \int d\mathbf{p}\; T^+(f\oplus g)^\sim(\mathbf{p})e^{i\mathbf{p}\cdot\mathbf{x}}, \tag{1.4}$$

$(f,g)\in\mathscr{H}^*\times\mathscr{H}$. Then, we have

$$F_T(x_0,\mathbf{x}) = \int d\mathbf{p}\; T^+(f\oplus g)^\sim(\mathbf{p})e^{-i[p,x]},$$

$$(f,g)\in\mathscr{H}^*\times\mathscr{H},$$

$(x_0,\mathbf{x}) = x\in\mathbb{M}^4$, and p is the 4-vector $p = (|\mathbf{p}|,\mathbf{p})$, $\mathbf{p}\in\mathbb{R}^3$. Here and hereafter $[.,.]$ denotes the indefinite inner product of $\mathbb{M}^4$. It is clear that since the components of F_T are real generalized functions, we must have that

$$T^+(f\oplus g)^\sim(-p) = \overline{T^+(f\oplus g)^\sim(p)}, \quad \mathbf{p}\in\mathbb{R}^3,$$

$$(f,g)\in\mathscr{H}^*\times\mathscr{H}.$$

Next, let $\mathscr{F}_T$ denote the collection of all four-component vector functions $T^+(f\oplus g)^\sim$, $(f,g)\in\mathscr{H}^*\times\mathscr{H}$, related to $T^+(f\oplus g)\in\mathscr{H}^*$, as in (1.4), which satisfy the following two conditions:

$$\text{(i)}\ \ |\mathbf{p}|T^+(f\oplus g)_0^\sim(\mathbf{p}) - \sum_{i=1}^{3} p_i T^+(f\oplus g)_i^\sim(\mathbf{p}) = 0,$$

$$\text{(ii)}\ \ \|T^+(f\oplus g)\|^2 \equiv \int d\mathbf{p}[T^+(f\oplus g)^\sim(\mathbf{p}),\quad T^+(f\oplus g)^\sim(\mathbf{p})] < \infty,$$

for all $(f,g)\in\mathscr{H}^*\times\mathscr{H}$. Then the map

$$\|\cdot\|:\mathscr{F}_T\to[0,\infty)$$

[where $\|\cdot\|$ is defined as in (ii)] is a norm on $\mathscr{F}_T/\mathrm{Ker}(\|\cdot\|)$. Denote by $\mathscr{F}_T^0$ the real Hilbert space obtained by completing $\mathscr{F}_T/\mathrm{Ker}(\|\cdot\|)$.

Let $\langle .,.\rangle_T$ be the inner product of $\mathscr{F}_T^0$. Introduce the linear operator j on $\mathscr{F}_T^0$ with the following properties:

(a) $j^2 = -I$, where I is the identity operator on $\mathscr{F}_T^0$,

(b) $\langle jF_1,jF_2\rangle_T = \langle F_1,F_2\rangle_T$, for all $(F_1,F_2)\in\mathscr{F}_T^0\times\mathscr{F}_T^0$,

(c) $\langle jF,F\rangle_T > 0$ if and only if $F\neq 0$.

Then $\mathscr{F}_T^0$ becomes a complex Hilbert space, denoted by $\mathscr{F}_1^{\rho_T}$, relative to the inner product $\langle\langle .,.\rangle\rangle_T$ given by

$$\langle\langle F_1,F_2\rangle\rangle_T = \langle F_1,F_2\rangle_T - i\langle jF_1,F_2\rangle_T,$$

where i is the imaginary unit.

Set $\mathscr{F}_0^{\rho_T} = C$, the complex numbers, and let $\mathscr{F}_n^{\rho_T}$, $n = 1,2,\ldots$, be the n-fold symmetric tensor product of $\mathscr{F}_1^{\rho_T}$ with itself. Then the Hilbert space

$$\mathscr{F}^{\rho_T} = \bigoplus_{n=0}^{\infty} \mathscr{F}_n^{\rho_T}$$

is the representation space for the $*$-representation

$$\rho_T:\mathfrak{A}\to\rho_T(\mathfrak{A}) = \rho_0(\gamma_T(\mathfrak{A})).$$

For T belonging to $\mathfrak{G}_0(\mathscr{H}^*,\mathscr{H})$ the $*$-representation $(\rho_T,\mathscr{F}^{\rho_T})$ is unitarily *inequivalent* to the $*$-representation $(\rho_0,\mathscr{F})$, which is the Fock representation.

(1.5) *Definition*: Identify Minkowski space $\mathbb{M}^4$ with the subgroup of $\mathscr{P}_+^\uparrow$, the Poincaré group, consisting of all space-time translations and let $a\mapsto\alpha_a$, $a\in\mathbb{M}^4$, be a representation of $\mathbb{M}^4$ in $*$-Aut($\mathfrak{A}$). Then a $*$-representation ϕ of $\mathfrak{A}$ is said to be *positive* (*a la* Borchers[2]) if and only if the map $\phi(\mathfrak{A})\mapsto\phi(\alpha_a(\mathfrak{A}))$ is implementable by a strongly continuous unitary representation of $\mathbb{M}^4$ in $\mathscr{U}(\mathscr{F}^{\rho_T})$ (the group of unitary operators on $\mathscr{F}^{\rho_T}$) which satisfies the spectrum condition in the usual sense.

(1.6) *Remark*: (i) The representation $a\mapsto\alpha_a$ of $\mathbb{M}^4$ in $*$-Aut($\mathfrak{A}$) induces, and is induced by, a representation $a\mapsto\tau_a$ of $\mathbb{M}^4$ in $B(\mathscr{H}^*\oplus\mathscr{H})$ (the Banach algebra of all bounded linear operators on the Hilbert space $\mathscr{H}^*\oplus\mathscr{H}$), defined as follows:

$$(\tau_a f\oplus g)^\sim(\mathbf{p}) = e^{i(a_0|\mathbf{p}| - \mathbf{a}\cdot\mathbf{p})}(f\oplus g)^\sim(\mathbf{p}),$$

$$a = (a_0,\mathbf{a})\in\mathbb{M}^4,$$

$$|\mathbf{p}|^{-1}\tilde{f}(\mathbf{p}) = \int d\mathbf{x}\, f(\mathbf{x})e^{-i\mathbf{p}\cdot\mathbf{x}}, \quad f\in\mathscr{H}^*$$

and

$$|\mathbf{p}|\tilde{g}(\mathbf{p}) = \int d\mathbf{x}\, g(\mathbf{x})e^{-i\mathbf{p}\cdot\mathbf{x}}, \quad g\in\mathscr{H}, \ \mathbf{p}\in\mathbb{R}^3.$$

It is clear that $a\mapsto\tau_a$ extends by linearity and continuity to a strongly continuous unitary representation of $\mathbb{M}^4$ in $\mathscr{U}(\mathscr{H}^*\oplus\mathscr{H})$:

(ii) In Sec. 3, we shall also employ a representation $\Lambda\mapsto\tau_\Lambda$ of the restricted Poincaré group in $\mathscr{U}(\mathscr{H}^*\oplus\mathscr{H})$ defined as follows:

$$(\tau_\Lambda f\oplus g)^\sim(\mathbf{p}) = ((\Lambda\oplus\Lambda)(f\oplus g))^\sim(\Lambda^{-1}\mathbf{p}),$$

where $\Lambda^{-1}\mathbf{p}$ denotes the space component of $\Lambda^{-1}(p_0,\mathbf{p})$. The representation $\Lambda\mapsto\tau_\Lambda$ induces, and is induced by, a representation $\Lambda\mapsto\alpha_\Lambda$ of the restricted Poincaré group in $*$-Aut($\mathfrak{A}$).

2. The set $\mathbb{K}$ of admissible kernels

In our subsequent characterization of the subset of the $*$-representations $\{\rho_T = \rho_0\circ\gamma_T: T\in\mathfrak{G}_0(\mathscr{H}^*,\mathscr{H})\}$ of $\mathfrak{A}$ which are positive in the sense of Definition (1.5), we shall utilize a certain distinguished set $\mathbb{K}$ of kernels. In the following, we characterize the members of $\mathbb{K}$.

For each $a\in\mathbb{M}^4$, $a = (a_0,\mathbf{a})$, let $\mathbb{K}_a$ denote the set of all Lebesgue-measurable kernels K which satisfy the following two conditions:

$$\int d\mathbf{p}\, d\mathbf{q}\, |K(\mathbf{p},\mathbf{q})|^2 = \infty, \tag{2.1}$$

$$\int d\mathbf{p}\, d\mathbf{q}\, |K(\mathbf{p},\mathbf{q})|^2 \sin^2\{\tfrac{1}{2}[a_0(|\mathbf{p}| + |\mathbf{q}|) - \mathbf{a}\cdot(\mathbf{p}+\mathbf{q})]\} < \infty. \tag{2.2}$$

(2.3) Theorem: The set $\mathbb{K}_a$, $a\in\mathbb{M}^4$, is nonempty.

Proof: To prove the assertion, it suffices to exhibit a class of members of $\mathbb{K}_a$, $a\in\mathbb{M}^4$. Thus, let $\mathbb{K}_a^0$, $a\in\mathbb{M}^4$, be the set of all Lebesgue-measurable kernels k which are:

(2.4) asymptotically absolutely square summable with respect to $d\mathbf{p}\, d\mathbf{q}$;

(2.5) such that $|k(\mathbf{p},\mathbf{q})|^2$ behaves like

$$|\mathbf{p}|^{-2}|\mathbf{q}|^{-2}[(|\mathbf{p}|+|\mathbf{q}|)^2+|\mathbf{p}+\mathbf{q}|^2]^{-1}$$

in a neighborhood of the origin.

The condition (2.1) above is clearly satisfied because in a neighborhood of the origin, we have

$$\begin{aligned}\int d\mathbf{p}\, d\mathbf{q}\, |k(\mathbf{p},\mathbf{q})|^2 &\sim \int d|\mathbf{p}|\, d|\mathbf{q}|\, d\Omega_{\mathbf{p}}\, d\Omega_{\mathbf{q}}\, |\mathbf{p}|^2|\mathbf{q}|^2|\mathbf{p}|^{-2}|\mathbf{q}|^{-2} \\ &\quad\times[(|\mathbf{p}|+|\mathbf{q}|)^2+|\mathbf{p}+\mathbf{q}|^2]^{-1} \\ &= \int d|\mathbf{p}|\, d|\mathbf{q}|\, d\Omega_{\mathbf{p}}\, d\Omega_{\mathbf{q}}\, [(|\mathbf{p}|+|\mathbf{q}|)^2+|\mathbf{p}+\mathbf{q}|^2]^{-1} \\ &= \infty.\end{aligned}$$

(Here $d\Omega_{\mathbf{p}}$ denotes the element of surface measure on the unit sphere of $\mathbb{R}^3$ expressed in the variables $p'=\mathbf{p}/|\mathbf{p}|$, $|\mathbf{p}|\neq 0$, $\mathbf{p}\in\mathbb{R}^3$.)

Condition (2.2) is also satisfied for, in a neighborhood $\mathscr{N}$ of the origin of $\mathbb{R}^3\times\mathbb{R}^3$, we have that

$$\begin{aligned}&\int_{\mathscr{N}} d\mathbf{p}\, d\mathbf{q}\, |k(\mathbf{p},\mathbf{q})|^2 \sin^2\{\tfrac{1}{2}[a_0(|\mathbf{p}|+|\mathbf{q}|)-\mathbf{a}\cdot(\mathbf{p}+\mathbf{q})]\} \\ &\quad\sim \int_{\mathscr{N}} d|\mathbf{p}|\, d|\mathbf{q}|\, d\Omega_{\mathbf{p}}\, d\Omega_{\mathbf{q}}\, ((|\mathbf{p}|+|\mathbf{q}|)^2 \\ &\qquad + |\mathbf{p}+\mathbf{q}|^2)^{-1}\sin^2\{\tfrac{1}{2}[a_0(|\mathbf{p}|+|\mathbf{q}|)-\mathbf{a}\cdot(\mathbf{p}+\mathbf{q})]\} \\ &\quad= \tfrac{1}{4}|a|^2\int_{\mathscr{N}} d|\mathbf{p}|\, d|\mathbf{q}|\, d\Omega_{\mathbf{p}}\, d\Omega_{\mathbf{q}}\, \cos^2\lambda \\ &\qquad\times\frac{\sin^2\frac{1}{2}[a_0(|\mathbf{p}|+|\mathbf{q}|)-\mathbf{a}\cdot(\mathbf{p}+\mathbf{q})]}{\frac{1}{4}[a_0(|\mathbf{p}|+|\mathbf{q}|)-\mathbf{a}\cdot(\mathbf{p}+\mathbf{q})]^2} \\ &\quad\leqslant 4\pi^2|a|^2\int d|\mathbf{p}|\, d|\mathbf{q}| < \infty, \quad \text{for all } a\in\mathbb{M}^4,\end{aligned}$$

where $|a|^2=a_0^2+|\mathbf{a}|^2=a_0^2+a_1^2+a_2^2+a_3^2$, and λ is the angle between the vectors $a=(a_0,\mathbf{a})$ and $b=(|\mathbf{p}|+|\mathbf{q}|,-(\mathbf{p}+\mathbf{q}))$, considered as elements of the four-dimensional Euclidean space $(\mathbb{R}^4,|\cdot|)$. Thus, members of $\mathbb{K}_a^0$, $a\in\mathbb{M}^4$, are square-integrable in a neighborhood $\mathscr{N}$ of the origin with respect to the measure $(\sin^2\{\frac{1}{2}[a_0(|\mathbf{p}| + |\mathbf{q}|)-\mathbf{a}\cdot(\mathbf{p}+\mathbf{q})]\})\, d\mathbf{p}\, d\mathbf{q}$. Since, by the definition of $\mathbb{K}_a^0$, $a\in\mathbb{M}^4$, each member of $\mathbb{K}_a^0$, $a\in\mathbb{M}^4$, is asymptotically absolutely square-summable with respect to $d\mathbf{p}\, d\mathbf{q}$, it follows that each member of $\mathbb{K}_a^0$, $a\in\mathbb{M}^4$, satisfies condition (2.2). This concludes proof of the nonemptiness of $\mathbb{K}_a^0$, $a\in\mathbb{M}^4$, and hence of $\mathbb{K}_a$, $a\in\mathbb{M}^4$.

(2.6) Remark: A specific member of $\mathbb{K}_a^0$, $a\in\mathbb{M}^4$, is given by the kernel:

$$(\mathbf{p},\mathbf{q})\mapsto k(\mathbf{p},\mathbf{q}) = |\mathbf{p}|^{-2}|\mathbf{q}|^{-2}K_2([(|\mathbf{p}|+|\mathbf{q}|)^2+|\mathbf{p}+\mathbf{q}|^2]^{1/2}),$$

where $t\mapsto K_\nu(t)$ is the *modified* Bessel function of the second kind of order ν.

(2.7) Definition: In the following, we set

$$\bigcap_{a\in\mathbb{M}^4}\mathbb{K}_a=\mathbb{K}.$$

Kernels from the set $\mathbb{K}$ will play a role in the sequel.

3. SUPERSELECTION SECTORS OF THE FREE ELECTROMAGNETIC FIELD

In this section, we characterize the subset of $\{\rho_T=\rho_0\circ\gamma_T : T\in\mathfrak{G}(\mathscr{H}^*,\mathscr{H})\}$, which consists of positive $*$-representations.

In the sequel, the operator-valued matrix $Y=T'T-I$ intervenes in our considerations. The matrix representation of Y is given by (A3). Notice that the matrix of Y involves the operators A_μ, B_μ, C_μ, and D_μ, $\mu=0,1,2,3$, which are described in (A4). The operators A_μ, B_μ, C_μ, and D_μ are bounded integral operators.

By studying the above kernels, we readily obtain the following result:

(3.1) Theorem: For $(\mathscr{A},\mathscr{B},\mathscr{I},\mathscr{L})\in\mathfrak{G}(\mathscr{H}^*,\mathscr{H})$ the bounded integral operators A_μ, B_μ, C_μ, and D_μ, $\mu=0,1,2,3$, do not belong to the Hilbert–Schmidt class.

Proof: The assertion is established in the same way as Theorem 1 of Ref. 1. Indeed, the operator A_1 of the present theorem is precisely the operator Y_{11} of Theorem 1 of Ref. 1. Hence we shall omit the details.

(3.2) Remark: (i) The class $\mathfrak{G}(\mathscr{H}^*,\mathscr{H})$ is defined in the course of the discussion immediately preceding Theorem 1 of Ref. 1. In the following, we denote the Fourier transforms of the kernels of the integral operators A_μ, B_μ, C_μ, and D_μ, by $\widetilde{A}_\mu(.,.)$, $\widetilde{B}_\mu(.,.)$, $\widetilde{C}_\mu(.,.)$, and $\widetilde{D}(.,.)$, respectively. (ii) The next result marks the beginning of our characterization of the subset of the set $\{\rho_T=\rho_0\circ\gamma_T : T\in\mathfrak{G}_0(\mathscr{H}^*,\mathscr{H})\}$, which consists of positive $*$-representations of $\mathfrak{A}$.

(3.3) Theorem: There exists a continuous projective unitary representation $a\mapsto\cup_T(a)$ of $\mathbb{M}^4$ in $\mathscr{U}(\mathscr{F}^{\rho_T})$, which implements the $*$-automorphism

$$\rho_T(\mathfrak{A})\mapsto(\rho_T\circ\alpha_a)(\mathfrak{A}), \quad a\in\mathbb{M}^4, \quad T\in\mathfrak{G}_0(\mathscr{H}^*,\mathscr{H}),$$

i.e., $(\rho_T\circ\alpha_a)(\mathfrak{A})=\cup_T(a)\rho_T(\mathfrak{A})\cup_T(a)^{-1}$, if and only if each of the kernels $\widetilde{A}_\mu(.,.)$, $\widetilde{B}_\mu(.,.)$, $\widetilde{C}_\mu(.,.)$, and $\widetilde{D}_\mu(.,.)$ belongs to $\mathbb{K}$. Furthermore, $\cup_T$ has an associated true unitary representation.

Proof: Let $a\mapsto\nu_a$ denote the representation of $\mathbb{M}^4$ in $*$-Aut$(\rho_T(\mathfrak{A}))$ given by

$$\nu_a(\rho_T(\mathfrak{A}))=(\rho_T\circ\alpha_a)(\mathfrak{A}).$$

Then ν_a, $a\in\mathbb{M}^4$, is unitarily implementable if and only if

$$(\rho_T\circ\alpha_a)(\mathfrak{A})=(\rho_0\circ\gamma_T\circ\alpha_a)(\mathfrak{A})\cong\rho_T(\mathfrak{A})=(\rho_0\circ\gamma_T)(\mathfrak{A}),$$

where $\cong$ denotes unitary equivalence. Thus ν_a, $a\in\mathbb{M}^4$, is unitarily implementable if and only if

$$(\rho_0\circ\gamma_T\circ\alpha_a\circ\gamma_T^{-1})(\mathfrak{A})\cong\rho_0(\mathfrak{A}).$$

But, by Shale's theorem, the latter will be the case if and only if the operator-valued matrix

$$\begin{aligned}&(T\tau_a T^{-1})'(T\tau_a T^{-1})-I \\ &\quad=(T')^{-1}\tau_a^{-1}T'T\tau_a T^{-1}-I \\ &\quad\equiv\theta(a)\end{aligned}$$

has entries each of which is Hilbert–Schmidt. Since τ_a, T', and T are bounded operators, with bounded inverses, it follows that $\theta(a)$ has entries which are Hilbert–Schmidt if and only if the entries of the operator-valued matrix $\theta_T(a)=Y\tau_a-\tau_a Y$ are Hilbert–Schmidt. Here $Y=T'T-I$. From the matrix representation for Y [see

(A3)], we conclude that the entries of the operator-valued matrix $\theta_T(a)$ are Hilbert–Schmidt if and only if each of the operators

(i) $A_\mu \tau_a - \tau_a A_\mu$,

(ii) $B_\mu \tau_a - \tau_a B_\mu$,

(iii) $C_\mu \tau_a - \tau_a C_\mu$,

(iv) $D_\mu \tau_a - \tau_a D_\mu$, $\mu = 0,1,2,3$,

is Hilbert–Schmidt, for all $a \in \mathbb{M}^4$.

The method of establishing the assertion of the theorem for each of the listed operators is the same; hence it suffices only to consider the operators $A_\mu \tau_a - \tau_a A_\mu$, $\mu = 0,1,2,3$, for example.

Let $(\mathbf{p},\mathbf{q}) \mapsto \widetilde{L}_{a\mu}(\mathbf{p},\mathbf{q})$, $(\mathbf{p},\mathbf{q}) \mapsto \widetilde{J}_{a\mu}(\mathbf{p},\mathbf{q})$, and $(\mathbf{p},\mathbf{q}) \mapsto \widetilde{K}_{a\mu}(\mathbf{p},\mathbf{q})$ denote the Fourier transforms of the integral operators $A_\mu \tau_a - \tau_a A_\mu$, $A_\mu \tau_a$, and $\tau_a A_\mu$, respectively, $\mu = 0,1,2,3$. Then

$$\widetilde{L}_{a\mu}(\mathbf{p},\mathbf{q}) = \widetilde{J}_{a\mu}(\mathbf{p},\mathbf{q}) - \widetilde{K}_{a\mu}(\mathbf{p},\mathbf{q}).$$

Furthermore, one readily checks that, in fact,

$$\widetilde{J}_{a\mu}(\mathbf{p},\mathbf{q}) = \widetilde{A}_\mu(\mathbf{p},\mathbf{q}) e^{-i(a_0|\mathbf{q}| - \mathbf{a}\cdot\mathbf{q})}$$

and

$$\widetilde{K}_{a\mu}(\mathbf{p},\mathbf{q}) = \widetilde{A}_\mu(\mathbf{p},\mathbf{q}) e^{i(a_0|\mathbf{p}| - \mathbf{a}\cdot\mathbf{p})}.$$

Hence

$$\widetilde{L}_{a\mu}(\mathbf{p},\mathbf{q}) = \widetilde{A}_\mu(\mathbf{p},\mathbf{q})(e^{-i(a_0|\mathbf{q}| - \mathbf{a}\cdot\mathbf{q})} - e^{i(a_0|\mathbf{p}| - \mathbf{a}\cdot\mathbf{p})}),$$

$a = (a_0,\mathbf{a}) \in \mathbb{M}^4$.

Finally, the operators $A_\mu \tau_a - \tau_a A_\mu$, $\mu = 0,1,2,3$, are Hilbert–Schmidt if and only if

$$\int d\mathbf{p}\, d\mathbf{q}\, |\widetilde{L}_{a\mu}(\mathbf{p},\mathbf{q})|^2 < \infty, \quad \mu = 1,2,3,$$

i.e., if and only if

$$\int d\mathbf{p}\, d\mathbf{q}\, |\widetilde{A}_\mu(\mathbf{p},\mathbf{q})|^2 |e^{-i(a_0|\mathbf{q}| - \mathbf{a}\cdot\mathbf{q})} - e^{i(a_0|\mathbf{p}| - \mathbf{a}\cdot\mathbf{p})}|^2$$

$$= 4 \int d\mathbf{p}\, d\mathbf{q}\, |\widetilde{A}_\mu(\mathbf{p},\mathbf{q})|^2 \times \sin^2\{\tfrac{1}{2}[a_0(|\mathbf{p}| + |\mathbf{q}|) - \mathbf{a}\cdot(\mathbf{p}+\mathbf{q})]\} < \infty,$$

$\mu = 0,1,2,3$ and for all $a = (a_0,\mathbf{a}) \in \mathbb{M}^4$. But this is condition (2.2) in the definition of the set $\mathbb{K}_a \subset \mathbb{K}$, $a = (a_0,\mathbf{a}) \in \mathbb{M}^4$. Also, since T belongs to $\mathfrak{S}_0(\mathscr{H}^*,\mathscr{H})$, we know that $\int d\mathbf{p}\, d\mathbf{q} |\widetilde{A}_\mu(\mathbf{p},\mathbf{q})|^2 = \infty$, $\mu = 0,1,2,3$, by theorem (3.1): This is condition (2.1) in the definition of $\mathbb{K}_a \subset \mathbb{K}$, $a = (a_0,\mathbf{a}) \in \mathbb{M}^4$. [Analogous proofs are valid for the operators $B_\mu \tau_a - \tau_a B_\mu$, $C_\mu \tau_a - \tau_a C_\mu$, and $D_\mu \tau_a - \tau_a D_\mu$, $\mu = 0,1,2,3$, and $a = (a_0,\mathbf{a}) \in \mathbb{M}^4$.] Hence the assertion of the theorem concerning the existence of the unitary operator $\cup_T(a)$, $a \in \mathbb{M}^4$, has been established.

Next, let $a \mapsto \cup(a) \in \mathscr{U}(\mathscr{F})$ be the strongly continuous unitary representation of $\mathbb{M}^4$ which implements space-time automorphisms in the $*$-representation ρ_0, i.e.,

$$\rho_0(\sigma_a \mathfrak{A}) = \cup(a)\rho_0(\mathfrak{A})\cup(a)^{-1}, \quad a \in \mathbb{M}^4,$$

where $a \mapsto \alpha_a$ is a representation of $\mathbb{M}^4$ in $*$-Aut($\mathfrak{A}$).

Let $\mathfrak{S}_0(\mathscr{H}^*,\mathscr{H},\mathbb{K})$ denote the set of all $T \in \mathfrak{S}_0(\mathscr{H}^*,\mathscr{H})$ such that the Fourier transform of the kernel of each of the entries of the operator-valued matrix $Y = T^t T - I$ belongs to $\mathbb{K}$. Since ρ_0 is an irreducible $*$-representation of $\mathfrak{A}$, it follows that $\cup(a)$ belongs to $\rho_0(\mathfrak{A})''$ for each $a \in \mathbb{M}^4$. Furthermore, notice that the $*$-representation $\rho_T = \rho_0 \circ \gamma_T$, $T \in \mathfrak{S}_0(\mathscr{H}^*,\mathscr{H},\mathbb{K})$ being faithful, is nondegenerate. Hence by a well-known theorem (Ref. 3, Theorem 3.7.7.), there is a uniquely determined normal $*$-morphism $\boldsymbol{\rho}_T$ of the enveloping W^*-algebra env($\mathfrak{A}$) (of the C^*-algebra $\mathfrak{A}$) onto the enveloping W^*-algebra env($\rho_T(\mathfrak{A})$) [of the C^*-algebra $\rho_T(\mathfrak{A})$], which extends ρ_T, $T \in \mathfrak{S}_0(\mathscr{H}^*,\mathscr{H},\mathbb{K})$. [Recall that the enveloping W^*-algebra env($\mathscr{C}$) of a C^*-algebra $\mathscr{C}$ is the strong closure in $B(\mathbb{H})$ of the C^*-algebra $\oplus_{s \in S} \rho_s(\mathscr{C})$, where $\mathbb{H} = \oplus_{s \in S} \mathbb{H}_s$, the pair $(\rho_s, \mathbb{H}_s)$ is the Gel'fand–Naimark–Segal $*$-representation of $\mathscr{C}$ associated with the state s, and S is the state space of $\mathscr{C}$.]

It was previously shown that the $*$-automorphism

$$\rho_T(\mathfrak{A}) \mapsto \nu_a(\rho_T(\mathfrak{A})) = \rho_T(\alpha_a, \mathfrak{A}), \tag{3.4}$$

$T \in \mathfrak{S}_0(\mathscr{H}^*,\mathscr{H},\mathbb{K})$ is implemented by a unitary operator $\cup_T(a)$, $a \in \mathbb{M}^4$. Now, (3.4) is also implemented by the unitary operator $\boldsymbol{\rho}_T(\cup(a))$, $a \in \mathbb{M}^4$. To see this, notice that

$$\begin{aligned}\rho_T(\alpha_a \mathfrak{A}) &= \boldsymbol{\rho}_T(\alpha_a \mathfrak{A}), \quad \text{since } \boldsymbol{\rho}_T \text{ extends } \rho_T \\ &= \boldsymbol{\rho}_T(\cup(a)\mathfrak{A}\cup(a)^{-1}) \\ &= \boldsymbol{\rho}_T(\cup(a))\rho_T(\mathfrak{A})\boldsymbol{\rho}_T(\cup(a))^{-1}.\end{aligned}$$

Hence, since $\rho_T(\mathfrak{A})$ is *irreducible* for each $T \in \mathfrak{S}_0(\mathscr{H}^*,\mathscr{H},\mathbb{K})$, it follows that

$$\cup_T(a) = e^{i\zeta_T(a)}\boldsymbol{\rho}_T(\cup(a)), \quad (a,T) \in \mathbb{M}^4 \times \mathfrak{S}_0(\mathscr{H}^*,\mathscr{H},\mathbb{K}),$$

where $a \mapsto \zeta_T(a)$ is some real-valued function on $\mathbb{M}^4$. The unitary operator $\cup_T(a)$, $a \in \mathbb{M}^4$, satisfies the group property:

$$\cup_T(a)\cup_T(b) = e^{i\sigma_T(a,b)}\cup_T(a+b),$$

where

$$\sigma_T(\mathrm{a,b}) = \zeta_T(\mathrm{a}+\mathrm{b}) - \zeta_T(\mathrm{a}) - \zeta_T(\mathrm{b}). \tag{3.5}$$

Hence $\cup_T(a)$, $a \in \mathbb{M}^4$, is a *projective unitary representation*. Furthermore, defining $V_T(a)$ by

$$V_T(a) = e^{i\zeta_T(a)}\cup_T(a), \quad a \in \mathbb{M}^4, \ T \in \mathfrak{S}_0(\mathscr{H}^*,\mathscr{H},\mathbb{K}),$$

one readily checks that V_T is a true unitary representation of $\mathbb{M}^4$.

Finally, the strong continuity of $\cup_T(a)$, $a \in \mathbb{M}^4$, implies and is implied by the continuity of $a \mapsto \zeta_T(a)$, $a \in \mathbb{M}^4$.

(3.6) Remarks: (i) Notice that since $\cup_T(a)\cup_T(-a) =$ the identity operator, it follows that ζ_T is an odd function on $\mathbb{M}^4$. From this fact, one infers that the mapping $(a,b) \mapsto \sigma_T(a,b)$ is also an odd function on $\mathbb{M}^4 \times \mathbb{M}^4$.

(ii) In the functional equation (3.5), suppose that $\sigma_T(.,.)$ is known. Then a continuous solution ζ_T of (3.5) is of the form

$$\zeta_T(a) = [p^{(T)},a] - \int \mu_T(db)\sigma_T(b \times a,a), \quad a \in \mathbb{M}^4,$$

where $p^{(T)}$ is some fixed member of $\mathbb{M}^4$, μ_T is some finite measure on $\mathbb{M}^4$, $b \times a = (b_0a_0, b_1a_1, b_2a_2, b_3a_3)$, and $[.,.]$ denotes the indefinite inner product on $\mathbb{M}^4$.

Let

$$c_\mu^{(T)} = \left[\frac{\partial}{\partial a_\mu}\int \mu_T(db)\sigma(b \times a,a)\right]\Big|_{a=0},$$

$$T \in \mathfrak{S}_0(\mathscr{H}^*,\mathscr{H},\mathbb{K}), \tag{3.7}$$

and let $\mathbf{G}$ be the Minkowskian metric tensor with components: $G_{00}=1$, $G_{ij}=-\delta_{ij}$, $G_{0i}=0=G_{i0}$, $i,j=1,2,3$.

(3.8) Definition: Let $\mathfrak{S}_0(\mathscr{H}^*,\mathscr{H},\mathbb{K})$ denote as before the set of all $T\in\mathfrak{S}_0(\mathscr{H}^*,\mathscr{H})$ such that the Fourier transform of the kernel of each of the entries of the operator-valued matrix $Y=T^tT-I$ belongs to $\mathbb{K}$. Set

$$\mathfrak{S}_+(\mathscr{H}^*,\mathscr{H},\mathbb{K})=\{T\in\mathfrak{S}_0(\mathscr{H}^*,\mathscr{H},\mathbb{K}):$$

$p^{(T)}-\mathbf{G}c^{(T)}$ belongs to the closed forward light cone of $\mathbb{M}^4\}$.

(3.9) Theorem: The energy–momentum spectrum for the $*$-representation $\rho_T, T\in\mathfrak{S}_+(\mathscr{H}^*,\mathscr{H},\mathbb{K})$, is contained in the closed forward light cone of $\mathbb{M}^4$.

Proof: Let $H_0^{(T)}$ denote the infinitesimal generator of the unitary group $\boldsymbol{\rho}_T(\cup(a))$, $a\in\mathbb{M}^4$. Then since the spectrum of the infinitesimal generator of the unitary group $\cup(a)$, $a\in\mathbb{M}^4$, is contained in the closed forward light cone of $\mathbb{M}^4$, so is the spectrum of $H_0^{(T)}$. Next, the infinitesimal generator $H^{(T)}$ of the unitary group $\cup_T(a)=e^{i\zeta_T(a)}\boldsymbol{\rho}_T(\cup(a))$, $a\in\mathbb{M}^4$, is related to $H_0^{(T)}$ as follows:

$$H^{(T)}=H_0^{(T)}+(p^{(T)}-\mathbf{G}c^{(T)})\mathbb{1},\quad T\in\mathfrak{S}_+(\mathscr{H}^*,\mathscr{H},\mathbb{K}),$$

where $\mathbb{1}$ is the identity operator. Hence the spectrum of $H^{(T)}$ is contained in the closed forward light cone of $\mathbb{M}^4$ if and only if $p^{(T)}-\mathbf{G}c^{(T)}$ belongs to the closed forward light cone of $\mathbb{M}^4$. But $p^{(T)}-\mathbf{G}c^{(T)}$ belongs to the closed foward light cone of $\mathbb{M}^4$ since $T\in\mathfrak{S}_+(\mathscr{H}^*,\mathscr{H},\mathbb{K})$. This concludes the proof.

(3.10) Remark: We turn next to the question of unitary implementability, in the $*$-representation ρ_T of $\mathfrak{A}$, $T\in\mathfrak{S}_0(\mathscr{H}^*,\mathscr{H})$, of the representation $\Lambda\mapsto\alpha_\Lambda$, defined in Remark (1.6) (ii), of the restricted Poincaré group in $*$-Aut($\mathfrak{A}$). By arguing as in Theorem (3.3), we readily also establish the following assertion, whose proof we therefore omit:

(3.11) Theorem: The $*$-automorphism $\rho_T(\mathfrak{A})\mapsto(\rho_T\circ\alpha_\Lambda)(\mathfrak{A})$, $T\in\mathfrak{S}_0(\mathscr{H}^*,\mathscr{H})$, $T\neq$ identity operator, with Λ belonging to the restricted Poincaré group, is unitarily implementable if and only if Λ is the identity matrix.

(3.12) Remark: We infer from Theorem (3.9) that restricted Poincaré automorphisms are *not* unitarily implementable in any of the $*$-representations ρ_T, $T\in\mathfrak{S}_0(\mathscr{H}^*,\mathscr{H})$, $T\neq$ identity operator. Hence angular momentum operators are not defined in any of these $*$-representations. This difficulty, arising from infrared problems, is amply noted in the literature.[4–6]

(3.13) Remarks: (i) The $*$-representations forming the set

$$\mathfrak{S}_+(\mathscr{H}^*,\mathscr{H},\mathbb{K})\equiv\{\rho_T: T\in\mathfrak{S}_+(\mathscr{H}^*,\mathscr{H},\mathbb{K})\}$$

are the *positive* $*$-representations of $\mathfrak{A}$, and $\mathfrak{S}_+(\mathscr{H}^*,\mathscr{H},\mathbb{K})$ is the physically interesting subset of the set $\{\rho_T:T\in\mathfrak{S}_0(\mathscr{H}^*,\mathscr{H})\}$. In the preceding, we have achieved a characterization of the physically relevant $*$-representations of $\mathfrak{A}$.

(ii) Two representations ρ_{T_1} and ρ_{T_2} belonging to $\mathfrak{S}_+(\mathscr{H}^*,\mathscr{H},\mathbb{K})$ are *unitarily equivalent* if and only if each of the entries of the operator-valued matrix $T_1^tT_1-T_2^tT_2$ is of Hilbert–Schmidt class.

(3.14) Definition: By a *superselection sector* we mean a unitary equivalence class of members of $\mathfrak{S}_+(\mathscr{H}^*,\mathscr{H},\mathbb{K})$.

(3.15) Comments: Let $\mathbb{Y}$ be the image in $B(\mathscr{H}^*\oplus\mathscr{H})$ of the operator Y which maps $\mathfrak{S}_0(\mathscr{H}^*,\mathscr{H},\mathbb{K})$ into $B(\mathscr{H}^*\oplus\mathscr{H})$ as follows:

$$Y:\ \mathfrak{S}_0(\mathscr{H}^*,\mathscr{H},\mathbb{K})\to B(\mathscr{H}^*\oplus\mathscr{H})$$
$$T\mapsto Y(T)=T^tT-I.$$

In $\mathbb{Y}$, we introduce a relation $\doteq$ thus: for Y_1, Y_2 in $\mathbb{Y}$, write $Y_1\doteq Y_2$ if and only if each of the entries of the operator-valued matrix Y_1-Y_2 is Hilbert–Schmidt. By Remark (3.13), if ρ_{T_1} and ρ_{T_2} belong to $\mathfrak{S}_+(\mathscr{H}^*,\mathscr{H},\mathbb{K})$, then $\rho_{T_1}\cong\rho_{T_2}$ if and only if the entries of the operator-valued matrix $T_1^tT_1-T_2^tT_2$ are Hilbert–Schmidt. Setting $Y_1\doteq Y(T_1)$ and $Y_2\doteq Y(T_2)$, one sees that $\rho_{T_1}\cong\rho_{T_2}$ if and only if $Y_1\doteq Y_2$. Hence the disjoint equivalence classes $\mathfrak{S}_+^*(\mathscr{H}^*,\mathscr{H},\mathbb{K})$ of $*$-representations in $\mathfrak{S}_+(\mathscr{H}^*,\mathscr{H},\mathbb{K})$, i.e., the *superselection sectors,* are labeled by the set $\mathbb{Y}^*$ of disjoint equivalence classes of members of $\mathbb{Y}$. Consequently, $\mathfrak{S}_+^*(\mathscr{H}^*,\mathscr{H},\mathbb{K})$ is a set of the form

$$\mathfrak{S}_+^*(\mathscr{H},\mathscr{H},\mathbb{K})=\{\tilde{\rho}_\nu:\nu\in\mathbb{Y}^*\}$$

in which, for each $\nu\in\mathbb{Y}^*$, $\tilde{\rho}_\nu$ is an equivalence class of positive $*$-representations. Next, set

$$\tilde{\rho}=\bigoplus_{\nu\in\mathbb{Y}^*}\tilde{\rho}_\nu$$

and

$$\tilde{\mathscr{F}}=\bigoplus_{\nu\in\mathbb{Y}^*}\tilde{\mathscr{F}}_\nu.$$

Here, $\tilde{\mathscr{F}}_\nu$, $\nu\in\mathbb{Y}^*$, is a Hilbert space which is isomorphic to $\mathscr{F}$, the Fock space of the free, electromagnetic field; in our analysis, $\tilde{\mathscr{F}}_\nu$ is the representation Hilbert space for $\tilde{\rho}_\nu$, $\nu\in\mathbb{Y}^*$. The pair $(\tilde{\rho}_\nu,\tilde{\mathscr{F}}_\nu)$ is, for each $\nu\in\mathbb{Y}^*$, what we have referred to above as a *superselection sector* of the free electromagnetic field. The couple $(\tilde{\rho},\tilde{\mathscr{F}})$, which clearly has the structure of a *vector bundle*, is a superselection theory representation of the canonical commutation relations of the free electromagnetic field.

The superselection sectors constructed above still constitute a fairly wide class. We obtain an interesting subclass of $\mathfrak{S}_+^*(\mathscr{H}^*,\mathscr{H},\mathbb{K})$ as follows:

Let $\mathbb{Y}_p$ be the subset of $\mathbb{Y}$ defined thus:

$\mathbb{Y}_p=\{X\in\mathbb{Y}: X \text{ has pure point spectrum}\}$.

In view of the separability of $\mathscr{H}^*\oplus\mathscr{H}$, the set of eigenvalues of each operator in $\mathbb{Y}_p$ is countable. For $X\in\mathbb{Y}_p$, let $(q_1(X),q_2(X),\ldots,q_n(X),\ldots)$ be an enumeration of the eigenvalues of X, with

$$q_1(X)\geqslant q_2(X)\geqslant\cdots\geqslant q_n(X)\cdots.$$

Let

$$\mathbb{Q}=\{(q_1(X),q_2(X),\ldots q_n(X),\ldots):X\in\mathbb{Y}_p\}.$$

For $\mathbf{q}^{(1)}=(q_1^{(1)},q_2^{(1)},\ldots,q_n^{(1)},\ldots)$ and $\mathbf{q}^{(2)}=(q_1^{(2)},q_2^{(2)},\ldots,q_n^{(2)},\ldots)$ belonging to $\mathbb{Q}$, we shall write $\mathbf{q}^{(1)}\approx\mathbf{q}^{(2)}$ if and only if $\Sigma_{n=1}^{\infty}|q_n^{(1)}-q_n^{(2)}|^2<\infty$. We denote $\mathbb{Q}/\approx$ by $\mathbb{Q}^*$. Let

$$\mathfrak{S}_{+p}(\mathscr{H}^*,\mathscr{H},\mathbb{K})$$
$$=\{\rho_T:T\in\mathfrak{S}_0(\mathscr{H}^*,\mathscr{H},\mathbb{K})\ \text{and}\ T^tT-I\in\mathbb{Y}_p\}$$

and let $\mathfrak{S}_{+p}^*(\mathscr{H}^*,\mathscr{H},\mathbb{K})$ be the set of equivalence classes of $*$-representations in $\mathfrak{S}_{+p}(\mathscr{H}^*,\mathscr{H},\mathbb{K})$. Clearly, $\mathfrak{S}_{+p}^*(\mathscr{H}^*,\mathscr{H},\mathbb{K})\subset\mathfrak{S}_+^*(\mathscr{H}^*,\mathscr{H},\mathbb{K})$. The superselection sectors arising from this construction are now pairs of the form $(\tilde{\rho}_{q_1,q_2,\ldots,q_n,\ldots},\tilde{\mathscr{F}}_{q_1,q_2,\ldots,q_n,\ldots})$, which are labeled by sequences of real numbers. We remark that, as mentioned in Ref. 1, Streater and Wilde[7] and Bonnard and Streater,[8] in their studies of superselection rules in two-dimensional space-time, have also obtained superselection sectors labeled by continuous real numbers.

APPENDIX

For convenience, we collect together in this Appendix some of the undefined notation which we have employed in the foregoing sections.

Let $\mathfrak{S}_0(\mathscr{H}^*,\mathscr{H})$ be as defined in Sec. 4 of Ref. 1. Then each $T\in \mathfrak{S}_0(\mathscr{H}^*,\mathscr{H})$ is of the form

$$T=\begin{bmatrix} M_0 & 0 & 0 & 0 & Q_0 & 0 & 0 & 0\\ 0 & M_1 & 0 & 0 & 0 & Q_1 & 0 & 0\\ 0 & 0 & M_2 & 0 & 0 & 0 & Q_2 & 0\\ 0 & 0 & 0 & M_3 & 0 & 0 & 0 & Q_3\\ P_0 & 0 & 0 & 0 & N_0 & 0 & 0 & 0\\ 0 & P_1 & 0 & 0 & 0 & N_1 & 0 & 0\\ 0 & 0 & P_2 & 0 & 0 & 0 & N_2 & 0\\ 0 & 0 & 0 & P_3 & 0 & 0 & 0 & N_3 \end{bmatrix}, \quad \text{(A1)}$$

where

$$\{M_\mu = C^{-1}(I+\mathscr{A}_\mu)C, \quad N_\mu = C(I+\mathscr{T}_\mu)C^{-1}, \quad \text{(A2)}$$
$$\{P_\mu = C\mathscr{B}_\mu C, \quad Q_\mu = C^{-1}\mathscr{L}_\mu C^{-1}, \quad \mu=0,1,2,3.$$

With T as above, let $Y=T^tT-I$. Then Y has a matrix representation given by

$$Y=\begin{bmatrix} A_0 & 0 & 0 & 0 & B_0 & 0 & 0 & 0\\ 0 & A_1 & 0 & 0 & 0 & B_1 & 0 & 0\\ 0 & 0 & A_2 & 0 & 0 & 0 & B_2 & 0\\ 0 & 0 & 0 & A_3 & 0 & 0 & 0 & B_3\\ C_0 & 0 & 0 & 0 & D_0 & 0 & 0 & 0\\ 0 & C_1 & 0 & 0 & 0 & D_1 & 0 & 0\\ 0 & 0 & C_2 & 0 & 0 & 0 & D_2 & 0\\ 0 & 0 & 0 & C_3 & 0 & 0 & 0 & D_3 \end{bmatrix}, \quad \text{(A3)}$$

$$\begin{aligned} A_\mu &= C^{-1}\mathscr{A}_\mu C + C\mathscr{A}^t_\mu C^{-1}\\ &\quad + C\mathscr{A}^t_\mu C^{-2}\mathscr{A}_\mu C + C^{-1}\mathscr{L}^t_\mu \mathscr{B}_\mu C,\\ B_\mu &= C^{-1}\mathscr{L}_\mu C^{-1} + C^{-1}\mathscr{L}^t_\mu C^{-1}\\ &\quad + C\mathscr{A}^t_\mu C^{-2}\mathscr{L}_\mu C^{-1} + C^{-1}\mathscr{L}^t_\mu \mathscr{T}_\mu C^{-1},\\ C_\mu &= C\mathscr{B}_\mu C + C\mathscr{B}^t_\mu C + C^{-1}\\ &\quad \times \mathscr{T}^t_\mu C^2\mathscr{B}_\mu C + C\mathscr{B}^t_\mu \mathscr{A}_\mu C,\\ D_\mu &= C\mathscr{T}_\mu C^{-1} + C^{-1}\mathscr{T}^t_\mu C\\ &\quad + C^{-1}\mathscr{T}^t_\mu C^2\mathscr{T}_\mu C^{-1} + C\mathscr{B}^t_\mu \mathscr{L}_\mu C^{-1},\\ \mu &= 0,1,2,3, \end{aligned} \quad \text{(A4)}$$

$C=(-\Delta)^{1/4}$ and X^t denotes the transpose of X.

One sees that the entries of the matrix for T involve the operators $\mathscr{A}_\mu$, $\mathscr{B}_\mu$, $\mathscr{T}_\mu$, and $\mathscr{L}_\mu$, $\mu=0,1,2,3$; these operators are defined on p. 1754 of Ref. 1. We remark that the operators A_μ, B_μ, C_μ, and D_μ, which appear in the matrix representation of Y, are bounded integral operators on $L^2(\mathbf{R}^3,d\mathbf{x})$. The Fourier transforms of the kernels of the operators are denoted by $\widetilde{A}_\mu(.,.)$, $\widetilde{B}_\mu(.,.)$, $\widetilde{C}_\mu(.,.)$, and $\widetilde{D}_\mu(.,.)$, respectively.

[1] G. O. S. Ekhaguere, "The theory of superselection rules. I. A class of inequivalent, irreducible *-representations of the canonical commutation relations of the free electromagnetic field," J. Math. Phys. **19**, 1751–7 (1978).

[2] H. J. Borchers, "On Groups of Automorphisms with Semi-bounded Spectrum," in *Systems á un nombre infini de degree de liberté* (editions du Centre National de la Recherche Scientifique, Paris, 1970), pp. 125–42.

[3] G. K. Pedersen, *C*-algebras And Their Automorphism Groups* (Academic, New York, 1979).

[4] G. Roepstorff, "Coherent Photon States and Spectral Condition," Commun. Math. Phys. **19**, 301–14 (1970).

[5] S. Doplicher, R. Hagg, and J. E. Roberts, "Fields, Observables and Gauge Transformations II," Commun. Math. Physics, **15**, 173–200 (1969).

[6] L. Polley, G. Reents, and R. F. Streater, "Some Covariant Representations of Massless Boson Fields," J. Phys. A: Math. Gen. **14**, 2479–88 (1981).

[7] R. F. Streater and I. F. Wilde, "Fermion States of a Boson Field," Nucl. Phys. **24**, 561–74 (1970).

[8] J. L. Bonnard and R. F. Streater, "Local Gauge Models Predicting Their Own Superselection Rules," paper presented at the International Conference on Mathematical Problems of Quantum Field Theory and Statistical Mechanics, Moscow, December 1972.

Coulomb-modified scattering parameters for Coulomb-plus-separable potentials for all l

J. W. de Maag and L. P. Kok
Institute for Theoretical Physics, P. O. Box 800, Groningen, The Netherlands
H. van Haeringen
Department of Mathematics, Delft University of Technology, Delft, The Netherlands

(Received 17 March 1983; accepted for publication 17 June 1983)

For three different separable-potential models, closed analytical expressions are presented for the Coulomb-modified scattering length and effective range for all values of the angular-momentum quantum number, $l = 0,1,2,...$. In the derivation of these results, use is made of the regular and irregular Coulomb wave functions that are entire analytic in k^2. It is shown that the Coulomb-modified effective-range function can be written as a simple expression involving these entire-analytic functions.

PACS numbers: 12.40.Qq, 03.80. + r

I. INTRODUCTION

Effective-range theory has been highly successful in the analysis and interpretation of low-energy two-particle scattering data. Usually, in this theory the effect of scattering by a short-range potential V_s is described essentially by two parameters, the scattering length and the effective range, see, for example, Refs. 1 and 2.

The modification of effective-range theory in the presence of a long-range (Coulomb) tail of the interaction has been studied extensively.[1–15] Thereby the Coulomb-modified effective-range function K_{csl} has been introduced, which is a real-meromorphic function of the energy variable k^2 in a large part of the k^2 plane, for each value of the angular-momentum variable l. Roughly speaking, it describes, in a practical way, the deviation from pure-Coulomb scattering due to the addition of V_s to the pure-Coulomb potential V_c.

Separable potentials have been, since the appearance of Yamaguchi's original paper,[16] an immensely popular tool in dynamical calculations. In this paper we show how, when V_s is a separable potential, the Coulomb-modified effective-range function and scattering parameters are obtained in a convenient and relatively simple way. An essential ingredient in our method is the use of radial Coulomb wave functions ϕ_{cl} and χ_{cl} that are entire-analytic functions of k^2. The entire analyticity of the regular function ϕ_{cl} is well known, whereas Lambert[9] has given the explicit prescription to construct χ_{cl}. The analytic structure of K_{csl} is clearly displayed in Eq. (3.17), which is basic for this paper.

The organization of this paper is as follows. In Sec. II we recall some notations and conventions. In Sec. III the Coulomb-modified effective-range function is obtained. Choosing three particular shapes for the form factor $\langle r|g_l\rangle$, we derive, in Sec. IV, closed expressions for the Coulomb-modified scattering length a_{csl} and effective range r_{csl} in each case, for all l, and for Coulomb repulsion. After a minor modification, which we explain in Sec. V, the same formulas hold for the case of Coulomb attraction as well. By considering the limit $V_c \to 0$, we have obtained also the expressions for the effective-range parameters a_{sl} and r_{sl} (Sec. IV). This enables us to study the relationship between purely-short-range and Coulomb-modified low-energy scattering parameters for these form factors. Section VI concludes this paper with a discussion. Thereby the relation between a_{csl} and a_{sl} for the three form factors considered is compared with other known relations.

Most of our results are new. In this respect, our work extends the results of previous work: For the Yamaguchi-like form factor, we mention Ref. 17 (a_{cs0} and r_{cs0}), Ref. 18 (a_{cs1} and r_{cs1}), and Ref. 19 (a_{csl} for all l, in perhaps a slightly less elegant form, but not r_{csl}), whereas for the δ-shell form factor, we recall Ref. 20 (a_{csl} and r_{csl} for all l). The results for the third form factor have not been obtained before.

II. NOTATIONS

We shall use notations and conventions that have been developed previously.[13,21] Units are such that $\hbar = 1 = 2m$, where m is the reduced mass, and the energy variable is denoted by $E = k^2$. The Coulomb potential is given in the coordinate representation by

$$V_c(r) = -2s/r = 2k\gamma/r, \tag{2.1}$$

where s is the Coulomb strength and γ is Sommerfeld's parameter. It is convenient to consider solutions of the complete partial-wave projected Schrödinger equation

$$H_l|\psi_l\rangle = k^2|\psi_l\rangle, \tag{2.2}$$

where

$$H_l = H_{0l} + V_l, \tag{2.3}$$

and V_l is restricted to be a rotationally invariant potential. In the representation $[rl]H_{0l}$ is given by

$$H_{0l} = -\frac{1}{r}\frac{d^2}{dr^2}r + \frac{l(l+1)}{r^2}. \tag{2.4}$$

Three different solutions of Eq. (2.2) in coordinate space are $\langle r|kl+\rangle$, $\langle r|kl\uparrow\rangle$, and $\langle r|kl\downarrow\rangle$. Their connection with the regular solution $\phi_l(k,r)$ and the Jost solutions $f_{l\pm}(k,r)$ of the radial Schrödinger equation is given by

$$\langle r|kl+\rangle = \frac{(2/\pi)^{1/2}(ik)^l}{(2l+1)!!f_l(k)}\frac{\phi_l(k,r)}{r}, \tag{2.5}$$

$$\langle r|kl\uparrow\rangle = (2/\pi)^{1/2}f_{l+}(k,r)/(kr), \tag{2.6}$$

$$\langle r|kl\downarrow\rangle = (-)^l(2/\pi)^{1/2}f_{l-}(k,r)/(kr), \tag{2.7}$$

 0022-2488/84/030684-09$02.50

respectively. We recall that the regular solution satisfies the boundary condition

$$\lim_{r\downarrow 0} \phi_l(k,r) r^{-l-1} = 1 \,. \tag{2.8}$$

Furthermore, $f_l(k)$ is the Jost function,

$$f_l(k) = \lim_{r\to 0} \langle r|kl\uparrow\rangle / \langle r|kl\uparrow\rangle_0 \,, \tag{2.9}$$

and $\langle r|kl\uparrow\rangle_0$ is the solution corresponding to $V\equiv 0$. The three solutions (2.5)–(2.7) are interrelated by

$$2i\langle r|kl+\rangle = \exp(2i\delta_l(k))\langle r|kl\uparrow\rangle - \langle r|kl\downarrow\rangle \,, \tag{2.10}$$

where the phase shift $\delta_l(k)$ is given by

$$\delta_l(k) = -\arg f_l(k) \,, \quad k>0 \,. \tag{2.11}$$

For the pure Coulomb case ($V = V_c$), all these functions are well known in closed form,

$$\begin{aligned}\langle r|kl+\rangle_c &= (-)^l{}_c\langle kl-|r\rangle \\ &= (2/\pi)^{1/2} \frac{\Gamma(l+1+i\gamma)}{\Gamma(2l+2)} e^{-1/2\pi\gamma + ikr} \\ &\quad \times (2ikr)^l {}_1F_1(l+1+i\gamma; 2l+2; -2ikr) \,, \end{aligned}\tag{2.12}$$

$$\begin{aligned}\langle r|kl\uparrow\rangle_c &= (-)^l \langle r|k^*l\downarrow\rangle_c^* \\ &= (2/\pi)^{1/2} e^{1/2\pi\gamma + ikr}(-2ikr)^{l+1} \\ &\quad \times U(l+1+i\gamma, 2l+2, -2ikr)/(kr) \,, \end{aligned}\tag{2.13}$$

where the subscript c stands for Coulomb, and ${}_1F_1(.;.;.)$ and $U(.,.,.)$ are the confluent hypergeometric functions.[22–25] The Coulomb Jost function obeys (2.9) and reads

$$f_{cl}(k) = e^{1/2\pi\gamma}\Gamma(l+1)/\Gamma(l+1+i\gamma) \,; \tag{2.14}$$

the pure Coulomb phase shift $\sigma_l(k)$ follows from (2.11),

$$\sigma_l(k) = -\arg f_{cl}(k) = \arg \Gamma(l+1+i\gamma) \,, \quad k>0 \,. \tag{2.15}$$

The regular solution $\phi_{cl}(k,r)$ satisfies the boundary condition (2.8). Since this boundary condition is independent of k, ϕ_{cl} is an entire-analytic function of k^2 for every nonnegative value of r. This follows from a theorem by Poincaré.[26] From the relationship

$$\phi_{cl}(k,r) = r^{l+1} e^{ikr} {}_1F_1(l+1+i\gamma; 2l+2; -2ikr) \,, \tag{2.16}$$

it is manifestly clear that Eq. (2.8) is satisfied.

An irregular solution $\chi_{cl}(k,r)$ of the radial Schrödinger equation, which is an entire analytic function of k^2 for every positive r, has already been constructed by Lambert.[9] We shall use

$$\begin{aligned}\chi_{cl}(k,r) &= (\pi/2)^{1/2}(2l+1)!!(-ik)^l \left[\frac{kr\langle r|kl\uparrow\rangle_c}{f_{cl}(k)} \right. \\ &\quad \left. - \binom{l+i\gamma}{l}\binom{l-i\gamma}{l} 2\gamma H(\gamma) f_{cl}(k) kr\langle r|kl+\rangle_c \right] \\ &= [(2l+1)!!(2l-1)!!] e^{ikr} r^{-l} \frac{\Gamma(l+1+i\gamma)}{\Gamma(2l+1)} \\ &\quad \times (-2ikr)^{2l+1} U(l+1+i\gamma, 2l+2, -2ikr) \\ &\quad - \binom{l+i\gamma}{l}\binom{l-i\gamma}{l} 2\gamma H(\gamma) k^{2l+1} \phi_{cl}(k,r) \,. \end{aligned}\tag{2.17}$$

Here $H(\gamma)$ is given by[18]

$$H(\gamma) = \psi(i\gamma) + (2i\gamma)^{-1} - \ln(-i\gamma\,\mathrm{sgn}(s)) \,, \tag{2.18}$$

$\psi(z) = \Gamma'(z)/\Gamma(z)$. Note that χ_{cl} is not the only entire-analytic irregular solution. It has been constructed in such a way that the Coulomb-modified effective-range function can be given by a simple and convenient formula, see Eq. (3.17). From the relation

$$\begin{aligned}&\lim_{z\to 0} \frac{\Gamma(l+1+i\gamma)}{\Gamma(2l+1)} z^{2l+1} U(l+1+i\gamma, 2l+2, z) = 1 \,, \\ &l = 0,1,2,\ldots, |\arg z| < \pi \,, \end{aligned}\tag{2.19}$$

the behavior of χ_{cl} near $r = 0$ is easily derived:

$$\chi_{cl}(k,r) = (2l+1)!!(2l-1)!! r^{-l}[1 + O(r)] \,, \quad l = 1,2,\ldots, \tag{2.20}$$

whereas for $l = 0$, we have $\chi_{c0}(k,r) = 1 + O(r\ln r)$. We stress that for real k and $r>0$, both ϕ and χ are real. This is not apparent from the forms given in Eqs. (2.16) and (2.17).

III. THE COULOMB-MODIFIED EFFECTIVE-RANGE FUNCTION

In this section we consider the potential

$$V = V_c + V_s \,, \tag{3.1}$$

where V_c is the Coulomb potential and V_s is a short-range potential. In the Gell-Mann–Goldberger[27] two-potential formalism, it follows that the total T operator for this potential can be written as

$$\begin{aligned}T &= T_c + T_{cs} \\ &= T_c + (1 + T_c G_0) t_{cs} (1 + G_0 T_c) \,. \end{aligned}\tag{3.2}$$

Here, T_c is the Coulomb T operator, t_{cs} is a short-range operator satisfying the equation

$$t_{cs} = V_s + V_s G_c t_{cs} \,, \tag{3.3}$$

and G_0 and G_c are the free and Coulomb resolvents, respectively. The partial-wave analogs of these equations have the same form: One merely attaches an extra subscript l to each operator.

The Coulomb-modified phase shifts $\delta_{csl}(k)$ are connected to the Coulomb-modified physical on-shell T-matrix elements[13,28]

$$\begin{aligned}\langle kl\infty -|T_{csl}|kl\infty +\rangle &= {}_c\langle kl-|t_{csl}|kl+\rangle_c \\ &= -(2/\pi k)\exp(2i\sigma_l) \\ &\quad \times [\cot\delta_{csl}(k) - i]^{-1} \,. \end{aligned}\tag{3.4}$$

For separable potentials of the type

$$V_{sl} = -\lambda_l |g_l\rangle\langle g_l| \,, \tag{3.5}$$

the Coulomb-modified t operator and the phase shift can be obtained explicitly:

$$t_{csl} = -|g_l\rangle\langle g_l|/(\lambda_l^{-1} + \langle g_l|G_{cl}|g_l\rangle) \,, \tag{3.6}$$

$$\cot\delta_{csl} - i = \frac{2}{\pi k}\exp(2i\sigma_l) \frac{\lambda_l^{-1} + \langle g_l|G_{cl}|g_l\rangle}{{}_c\langle kl-|g_l\rangle\langle g_l|kl+\rangle_c} \,. \tag{3.7}$$

Furthermore, it is known that for a large class of potentials, the so-called Coulomb-modified effective-range function, defined as

$$K_{csl}(k^2) := k^{2l+1}\binom{l+i\gamma}{l}\binom{l-i\gamma}{l} \times\left[2\gamma H(\gamma) + \frac{2\pi\gamma}{\exp(2\pi\gamma)-1}(\cot\delta_{csl} - i)\right], \tag{3.8}$$

is a real-meromorphic function of k^2 in a neighborhood of $k = 0$, see Cornille and Martin,[10] Hamilton *et al.*,[11] and van Haeringen.[18] Its expansion coefficients are related to the Coulomb-modified low-energy scattering parameters.

$$K_{csl}(k^2) = -1/a_{csl} + \tfrac{1}{2}r_{csl}k^2 - P_{csl}r_{csl}^3k^4 + Q_{csl}r_{csl}^5k^6 - \cdots. \tag{3.9}$$

The definition of $H(\gamma)$ has already been recalled in Eq. (2.18). For potentials of the type (3.5), it follows from Eqs. (3.7) and (3.8) that

$$K_{csl}(k^2) = k^{2l+1}\binom{l+i\gamma}{l}\binom{l-i\gamma}{l}\left[2\gamma H(\gamma) + \frac{2\pi\gamma}{\exp(2\pi\gamma)-1} \times \frac{2}{\pi k}\exp(2i\sigma_l)\frac{\lambda_l^{-1} + \langle g_l|G_{cl}|g_l\rangle}{{}_c\langle kl-|g_l\rangle\langle g_l|kl+\rangle_c}\right]. \tag{3.10}$$

The RHS of this equation has a seemingly complicated structure in the variable k. For example, $H(\gamma)$ is not analytic at $k = 0$. The analytic structure in k can be made more transparent, as we shall show now. The following procedure is also a convenient method to calculate the Coulomb-modified low-energy parameters $a_{csl}, r_{csl}, \ldots$.

First, we observe that G_{cl} in Eq. (3.10) is the usual outgoing Green function. In the coordinate representation, its matrix elements can be expressed in terms of the solution regular at $r = 0$, and the outgoing irregular solution, as follows [cf. Ref. 29, Eq. (14.50)],

$$\langle r'|G_{cl}|r\rangle = -(\pi k/2)(-)^l\langle r_>|kl\uparrow\rangle_c\langle r_<|kl+\rangle_c, \tag{3.11}$$

where $r_< = \min(r',r)$, and $r_> = \max(r',r)$.

Second, both the regular and the irregular solution can be expressed in terms of the entire-analytic radial solutions $\phi_{cl}(k,r)$ and $\chi_{cl}(k,r)$, as is clear from Eqs. (2.12), (2.13), (2.16), and (2.17):

$$\langle r|kl+\rangle_c = \frac{(2/\pi)^{1/2}(ik)^l}{(2l+1)!!f_{cl}(k)}\frac{\phi_{cl}(k,r)}{r}, \tag{3.12}$$

$$\langle r|kl\uparrow\rangle_c = \frac{(2/\pi)^{1/2}f_{cl}(k)}{(2l+1)!!(-ik)^l}\left[\frac{\chi_{cl}(k,r)}{kr} + \binom{l+i\gamma}{l}\binom{l-i\gamma}{l}2\gamma H(\gamma)k^{2l}\frac{\phi_{cl}(k,r)}{r}\right]. \tag{3.13}$$

Third, we introduce the notation

$$\mathscr{I}_{\phi\phi}(k^2) = \int_0^\infty dr'\, r'\phi_{cl}(k,r')\langle r'|g_l\rangle \int_0^\infty dr\, r\phi_{cl}(k,r)\langle g_l|r\rangle, \tag{3.14}$$

and

$$\mathscr{I}_{\chi\phi}(k^2) = \int_0^\infty dr'\, r' \int_0^\infty dr\, r\langle r'|g_l\rangle\chi_{cl}(k,r_>)\phi_{cl}(k,r_<)\langle g_l|r\rangle. \tag{3.15}$$

Using Eqs. (3.11)–(3.15), and

$$\{f_{cl}(k)\}^{-2} = \binom{l+i\gamma}{l}\binom{l-i\gamma}{l}\frac{2\pi\gamma}{\exp(2\pi\gamma)-1}\exp(2i\sigma_l), \tag{3.16}$$

it is straightforward to rewrite the RHS of Eq. (3.10), so as to obtain

$$K_{csl}(k^2) = \frac{1}{\mathscr{I}_{\phi\phi}(k^2)}\left[\frac{\{(2l+1)!!\}^2}{\lambda_l} - \mathscr{I}_{\chi\phi}(k^2)\right]. \tag{3.17}$$

Equation (3.17) is central to this paper. Barring convergence difficulties for large r,r' the integrals $\mathscr{I}_{\phi\phi}$ and $\mathscr{I}_{\chi\phi}$ are analytic functions of k^2. Note that $\mathscr{I}_{\phi\phi}$ is the product of two one-dimensional integrals. Equation (3.17) clearly displays the analytic structure of the Coulomb-modified effective-range function for separable potentials: The RHS is a real-meromorphic function of k^2. The region of meromorphicity of K_{csl} can be obtained by means of estimates and inequalities as have been used by Cornille and Martin.[10]

Here, and in the next section, we have tacitly assumed the form factors $\langle r|g_l\rangle$ and strength parameters λ_l to be independent of k^2. This assumption is unnecessary and the extension to form factors and strength parameters that are functions of k^2 is immediate.

The main purpose of this paper is the calculation in closed form of the Coulomb-modified effective-range parameters for certain separable potentials with simple form factors. For the expansion in k^2 of the Coulomb-modified effective-range function, the use of Eq. (3.10) seems to be less convenient.

The functions ϕ and χ are, by construction, entire-analytic functions of k^2. Their expansions in powers of k^2 are given in Appendix A. Upon interchanging the order of summation and integration (with due care!) the calculation of the Coulomb-modified effective-range parameters is immediate if the remaining integrals are known. Therefore Eq. (3.17) will be our starting point in the next section.

Finally, we note that in the limit of vanishing Coulomb potential ($s \to 0$), we retrieve expressions for the effective-range function $K_{sl}(k^2)$:

$$K_{sl}(k^2) = k^{2l+1}\cot\delta_{sl}(k). \tag{3.18}$$

Clearly this function is real meromorphic in k^2, too. Its expansion coefficients are related to the scattering length a_{sl}, the effective range $r_{sl}, \ldots$, according to

$$K_{sl}(k^2) = -a_{sl}^{-1} + \tfrac{1}{2}r_{sl}k^2 - P_{sl}r_{sl}^3k^4 + Q_{sl}r_{sl}^5k^6 - \cdots. \tag{3.19}$$

IV. THE COULOMB-MODIFIED SCATTERING PARAMETERS FOR COULOMB REPULSION

In this section we shall utilize a method by which the Coulomb-modified effective-range parameters can be obtained in closed form. The method will be illustrated by choosing three particular simple separable-potential form factors $\langle r|g_l\rangle$. We stress that the method is quite general: It works for all l, and both for Coulomb repulsion and Coulomb attraction.

Starting point is the observation that the integrals $\mathscr{I}_{\phi\phi}(k^2)$ and $\mathscr{I}_{\chi\phi}(k^2)$, under some mild restrictions on the form factor, are analytic functions of k^2 in a region containing the origin. This follows from the fact that $\phi_{cl}(k,r)$ and $\chi_{cl}(k,r)$ have this property for every $r>0$. In particular, $\phi_{cl}(k,r)$ and $\chi_{cl}(k,r)$ are uniformly represented by their Taylor series in k^2 in the whole k plane. These series are given in Appendix A, which heavily relies on Ref. 30. Replacing $\phi_{cl}(k,r)$ and $\chi_{cl}(k,r)$ by their Taylor series in k^2, we have to evaluate integrals of the type

$$\int_0^b dt\, f(z,t) = \int_0^b dt \left\{ \sum_{n=0}^{\infty} f_n(t)(z-z_0)^n \right\}, \qquad (4.1)$$

where b is either finite or infinite. Conditions for interchanging the order of summation and integration are given in Appendix B. Interchanging the summation over the first N terms and the integration is already allowed if only $f(z,t)$ and $f_n(t)$, $0 \leqslant n \leqslant N$, are integrable as functions of t.

We shall consider, for each value of l, three different form factors, viz.,

$$\text{(i)}\ \langle r|g_l\rangle = (\tfrac{1}{2}i)^l (l!)^{-1} r^{l-1} \delta(1-r/R), \quad R>0, \qquad (4.2)$$

$$\text{(ii)}\ \langle r|g_l\rangle = (\tfrac{1}{2}i)^l (l!)^{-1} r^{l-1} \Theta(1-r/R), \quad R>0, \qquad (4.3)$$

$$\text{(iii)}\ \langle r|g_l\rangle = (\tfrac{1}{2}i)^l (l!)^{-1} r^{l-1} \exp(-r/R), \quad R>0. \qquad (4.4)$$

In all three cases, R plays the role of a typical (short-) range parameter. In the momentum representation, these form factors take the following form:

$$\text{(i)}\ \langle p|g_l\rangle = (2/\pi)^{1/2} 2^{-l} (l!)^{-1} R^{l+2} j_l(pR), \qquad (4.5)$$

$$\text{(ii)}\ \langle p|g_l\rangle = (2/\pi)^{1/2} \{(2l+2)!\}^{-1} R^{2l+2} p^l \times {}_1F_2(l+1; l+\tfrac{3}{2}, l+2; -\tfrac{1}{4}p^2R^2), \qquad (4.6)$$

$$\text{(iii)}\ \langle p|g_l\rangle = (2/\pi)^{1/2} p^l (p^2 + R^{-2})^{-l-1}, \qquad (4.7)$$

respectively. In each case, the threshold behavior p^l is apparent. The δ-shell potential (in a slightly different normalization) has already been considered in Ref. 20, where formulas for a_{csl} and r_{csl} are presented. The calculation, in closed form, of these parameters for cases (ii) and (iii) can be carried out along the lines sketched in Appendix C and Appendix D, respectively.

We shall first consider a *repulsive* Coulomb potential $(-s>0)$. The following formulas hold.

(i) Defining $\nu = -2sR$, one has, for the δ-shell potential,

$$-a_{csl}^{-1} = \frac{\nu^{2l+1}}{(I_{2l+1})^2} \times \left\{ \lambda_l^{-1} R^{-4l-4} - \frac{(2R)^{-2l-1}}{(l!)^2} 4 I_{2l+1} K_{2l+1} \right\}, \qquad (4.8)$$

$$\frac{1}{2} r_{csl} = \frac{\nu^{2l+1}}{6(I_{2l+1})^2} \left[4\lambda_l^{-1} R^{-4l-2} \left(\frac{l I_{2l+3} + \sqrt{\nu} I_{2l+2}}{\nu I_{2l+1}} \right) - \frac{(2R)^{-2l+1}}{(l!)^2} \nu^{-2} \{ -2l(l+1) \times (1 - (4l+2) I_{2l+1} K_{2l+1}) - (I_{2l+1})^2 + \nu \} \right]. \qquad (4.9)$$

(ii) Using again $\nu = -2sR$, one has for the unit-step form factor

$$-a_{csl}^{-1} = \frac{\nu^{2l+2}}{(I_{2l+2})^2} \left\{ \lambda_l^{-1} R^{-4l-4} - \frac{(2R)^{-2l-1}}{(l!)^2} \nu^{-1} \left(\frac{1}{l+1} - 4 I_{2l+2} K_{2l+2} \right) \right\}, \qquad (4.10)$$

$$\frac{1}{2} r_{csl} = \frac{\nu^{2l+2}}{6(I_{2l+2})^2} \left[4\lambda_l^{-1} R^{-4l-2} \left(\frac{(l-1) I_{2l+4} + \sqrt{\nu} I_{2l+3}}{\nu I_{2l+2}} \right) - \frac{(2R)^{-2l+1}}{(l!)^2} \times \nu^{-3} \left\{ l(2l+1)(1 - 4(l+1) I_{2l+2} K_{2l+2}) - (I_{2l+2})^2 + 2\nu \left(-\frac{l}{2l+3} + \frac{(l-1) I_{2l+4} + \sqrt{\nu} I_{2l+3}}{(2l+2) I_{2l+2}} \right) \right\} \right]. \qquad (4.11)$$

In Eqs. (4.8)–(4.11), the shorthand notations I_n and K_n have been used for $I_n(2\sqrt{\nu})$ and $K_n(2\sqrt{\nu})$, respectively.

(iii) For the Yamaguchi-like form factor, we use $\nu = -sR$ (and *not* $\nu = -2sR$ as in the two previous cases!), and obtain

$$-a_{csl}^{-1} = e^{-4\nu} \lambda_l^{-1} R^{-4l-4} + (2(2l+1)!/(l!)^2) s^{2l+1} \Gamma(-2l-1, 4\nu) \qquad (4.12a)$$

$$= e^{-4\nu} \{ \lambda_l^{-1} R^{-4l-4} - (2(2l+1)!/(l!)^2) \times R^{-2l-1} \nu^{2l+1} U(2l+2, 2l+2, 4\nu) \}, \qquad (4.12b)$$

$$\tfrac{1}{2} r_{csl} = e^{-4\nu} \lambda_l^{-1} R^{-4l-2} (2l+2+\tfrac{4}{3}\nu) - \tfrac{1}{6} s^{2l-1}/(l!)^2 + ((2l+1)! s^{2l-1}/6(l!)^2) \{ \tfrac{1}{4} e^{-4\nu} (4\nu)^{-2l+1} + (l+1) \Gamma(-2l+1, 4\nu) \}. \qquad (4.13)$$

In the limit of vanishing Coulomb strength ($s \to 0$, i.e., $\nu \to 0$ for fixed R) we retrieve the effective-range parameters a_{sl} and r_{sl} for the following.

(i) For the δ-shell potential,

$$-a_{sl}^{-1} = [(2l+1)!]^2\lambda_l^{-1}R^{-4l-4} - \frac{[(2l+1)!!]^2}{2l+1}R^{-2l-1}, \qquad (4.14)$$

$$\tfrac{1}{2}r_{sl} = [(2l+1)!]^2\lambda_l^{-1}R^{-4l-2}/(2l+3) - \frac{[(2l+1)!!]^2}{(2l-1)(2l+3)}R^{-2l+1}. \qquad (4.15)$$

(ii) For the unit-step form factor,

$$-a_{sl}^{-1} = [(2l+2)!]^2\lambda_l^{-1}R^{-4l-4} - \frac{[(2l+1)!!]^2(4l+4)}{(2l+1)(2l+3)}R^{-2l-1}, \qquad (4.16)$$

$$\frac{1}{2}r_{sl} = \frac{[(2l+2)!]^2(2l+2)}{(2l+3)(2l+4)}\lambda_l^{-1}R^{-4l-2} - \frac{[(2l+1)!!]^2(4l+4)(4l^2+16l+11)}{(2l-1)(2l+3)^2(2l+4)(2l+5)}R^{-2l+1}. \qquad (4.17)$$

(iii) For the Yamaguchi-like form factor,

$$-a_{sl}^{-1} = \lambda_l^{-1}R^{-4l-4} - 2\binom{2l}{l}(4R)^{-2l-1}, \qquad (4.18)$$

$$\frac{1}{2}r_{sl} = (2l+2)\lambda_l^{-1}R^{-4l-2} - \frac{1}{8}\binom{2l}{l}\frac{(2l+1)^2}{2l-1}(4R)^{-2l+1}. \qquad (4.19)$$

Equations (4.10)–(4.19) enable us to study the relationship between purely short-range and Coulomb-modified low-energy scattering parameters. To lowest orders in ν (the ratio of a typical short range and the typical long-range parameter $|s|^{-1} = a_B$, the Bohr radius), we get the following.

(i) For the δ-shell potential ($\nu = -2sR$),

$$-a_{cs0}^{-1} = -a_{s0}^{-1}[1-\nu+O(\nu^2)] - R^{-1}\nu[2C - \tfrac{1}{2} + \ln\nu + O(\nu)], \qquad (4.20)$$

$$-a_{csl}^{-1} = -a_{sl}^{-1}\left[1 - \frac{\nu}{l+1} + O(\nu^2)\right] + \frac{\{(2l+1)!!\}^2R^{-2l-1}}{2l(2l+1)(2l+2)} \times 2\nu\left[1 - \frac{2l(4l+7)}{(2l-1)(2l+2)(2l+3)}\nu + O(\nu^2\ln\nu)\right], \quad (l\geqslant 1). \qquad (4.21)$$

(ii) For the unit-step form factor ($\nu = -2sR$),

$$-a_{cs0}^{-1} = -a_{s0}^{-1}[1-\tfrac{2}{3}\nu+O(\nu^2)] - R^{-1}\nu[2C - \tfrac{11}{12} + \ln\nu + O(\nu)], \qquad (4.22)$$

$$-a_{csl}^{-1} = -a_{sl}^{-1}\left[1 - \frac{2\nu}{2l+3} + O(\nu^2)\right] + \frac{[(2l+1)!!]^2(2l+2)R^{-2l-1}}{2l(2l+1)(2l+3)(2l+4)} \times 6\nu\left[1 - \frac{2}{3}\frac{2l(6l+17)}{(2l-1)(2l+3)(2l+5)}\nu + O(\nu^2\ln\nu)\right], \quad (l\geqslant 1). \qquad (4.23)$$

(iii) For the Yamaguchi-like form factor ($\nu = -sR$),

$$-a_{cs0}^{-1} = -a_{so}^{-1}\exp(-4\nu) - 2R^{-1}\nu\left[C + \ln 4\nu + \sum_1^\infty(-4\nu)^n/(n\,n!)\right], \qquad (4.24)$$

$$-a_{csl}^{-1} = -a_{sl}^{-1}[1-4\nu+O(\nu^2)] + \frac{8(2l-1)!}{(l!)^2}(4R)^{-2l-1} \times\left[1 - \frac{2l}{2l-1}4\nu + O(\nu^2\ln\nu)\right], \quad (l\geqslant 1). \qquad (4.25)$$

In Eqs. (4.21), (4.23), and (4.25), the order terms $O(\nu^2\ln\nu)$ may be replaced by order terms $O(\nu^2)$ in case $l\geqslant 2$. The constant C is Euler's constant, $C\approx 0.5772$. Note that Eq. (4.24) is without any approximation.

V. THE COULOMB-MODIFIED SCATTERING PARAMETERS FOR COULOMB ATTRACTION

In the previous section we have discussed the case of Coulomb repulsion. We now briefly discuss the case of Coulomb attraction, i.e., $s>0$. All formulas in Sec. III remain valid. In particular, Eq. (3.17) is again a convenient starting point.

The function $\phi_{cl}(k,r)$ is an entire, real-analytic function of s. Likewise, the expansion coefficients in its Taylor expansion in the variable k^2 are entire, real-analytic functions of s. For example, $\phi_{cl}(k,r)$ for $k=0$, i.e., $\phi_{cl}(0,r)$, is entire real-analytic in s, and $(2l+1)!(2s)^{-l-1}(2sr)^{1/2}J_{2l+1}(2(2sr)^{1/2})$ is the entire-analytic continuation to the positive real s axis of $(2l+1)!(-2s)^{-l-1}(-2sr)^{1/2}I_{2l+1}(2(-2sr)^{1/2})$, which is $\phi_{cl}(k=0,r)$ on the negative real s axis, cf. Eqs. (A1) and (A4).

The function $\chi_{cl}(k,r)$ has been given by expansions in Eqs. (A2) and (A5) for Coulomb repulsion and Coulomb attraction, respectively. It is not analytic in the variable s. For $s>0$, it is given in (A5). One observes that for $s>0$, we have

$$-(2s)^l(2sr)^{1/2}\pi N_{2l+1}(2(2sr)^{1/2}) = \mathrm{Re}[(2s)^l(2sr)^{1/2}\pi i H_{2l+1}^{(1)}(2(2sr)^{1/2})], \qquad (5.1)$$

and similar relations hold for the expressions involving $N_{2l+1+q}(\cdot)$, $q=1,2,\ldots$. From this it may be concluded that $\chi_{cl}(k,r)$ for s on the positive real axis ($s>0$) is the real part of the analytic continuation in the complex s plane of $\chi_{cl}(k,r)$ for s on the negative real s axis ($s<0$). Indeed, upon defining $-s = e^{-\pi i}s$, by using (Ref. 22, p. 67),

$$K_\nu(ze^{-(1/2)\pi i}) = \tfrac{1}{2}\pi i e^{1/2\pi\nu} H_\nu^{(1)}(z)\,, \tag{5.2}$$

we have that

$$\mathrm{Re}\big[(2s)^l(2sr)^{1/2}\pi i H^{(1)}_{2l+1}(2(2sr)^{1/2})\big]$$
$$= \mathrm{Re}\big[(2se^{-\pi i})^l(2sre^{-\pi i})2K_{2l+1}(2(2sre^{-\pi i})^{1/2})\big]\,. \tag{5.3}$$

Hence, in Eqs. (3.17) and (3.15), for the case of Coulomb attraction ($s>0$), we can still use Eq. (A2), provided we take $\mathrm{Re}\,\chi_{cl}$ instead of χ_{cl}.

As a consequence, all formulas which have been given in the previous section for the case of Coulomb repulsion ($s<0$) hold also for the case of Coulomb attraction ($s>0$) after only a minor modification: The RHS of each of the Eqs. (4.8)–(4.13) and (4.20)–(4.25) has to be replaced by its real part. This only affects the functions $K_n(\cdot)$, $\Gamma(\cdot,\cdot)$, and $\ln(\cdot)$.

For the δ-shell potential, some further practical details are given in Ref. 20.

VI. DISCUSSION

In Secs. III and IV, we have given a simple prescription for evaluating the Coulomb-modified effective-range function K_{csl} and parameters a_{csl} and r_{csl} for the Coulomb-plus-separable potential. It can handle a separable potential of arbitrary form, and it works for all values of the angular-momentum quantum number l. The derivation of the central formula of this paper, given by Eq. (3.17) and Eqs. (3.14) and (3.15), utilizes the entire analyticity of the functions $\phi_{cl}(k,r)$ and $\chi_{cl}(k,r)$, which are regular and irregular at the origin $r=0$, respectively.[9]

In Sec. IV, closed expressions are given for a_{csl} and r_{csl} for three choices (i)–(iii) for the form factor $\langle r|g_l\rangle$. These results generalize and extend previous work in several respects. First, our formulas incorporate the effect of the Coulomb potential to all orders in the Coulomb coupling parameter s, whereas recent work of Nogami and van Dijk[31] includes only effects to first order in the fine structure constant. When evaluated for our form factors, their first-order formulas give results in agreement with Eqs. (4.20), (4.22), and (4.24). Second, the results of Refs. 17–20 are extended. In Refs. 17 and 18, only in the cases $l=0$ and $l=1$, are closed formulas for the scattering parameters given. In Ref. 19, in a momentum-representation formulation, closed expressions for K_{csl} and a_{csl} for the form factor (iii) have been given. Our formula (4.12) agrees with the result of Ref. 19; at the same time our expression is appreciably simpler and very elegant. In addition, the result (4.13) for r_{csl} is new and relatively simple. The results for form factor (ii) are new. They are rather similar to the results for case (i). This is, perhaps, not surprising in view of the intimate connection between the function $\Theta(1-r/R)$ and the distribution $\delta(1-r/R)$.

In passing, we note that Eq. (3.17) is useful also in the chargeless case. In Eqs. (3.14) and (3.15) the entire real-analytic functions (in the variable k^2)

$$\phi_{0l}(k,r) = \lim_{s\to 0}\phi_{cl}(k,r) = (2l+1)!!rk^{-l}j_l(kr)\,, \tag{6.1}$$

$$\chi_{0l}(k,r) = \lim_{s\to 0}\chi_{cl}(k,r) = -(2l+1)!!k^l krn_l(kr)\,, \tag{6.2}$$

have to be used. We note here that Lambert[9] uses Messiah's convention[32] for the spherical Neumann function, which differs by a minus sign from $n_l(\cdot)$ in Eq. (6.2), where we have used the convention of Ref. 23.

The formulas of Sec. IV are useful in checking charge symmetry or charge independence of the nuclear forces. Illustrative checks have been made for NN scattering in Ref. 17 and for $N\alpha$ scattering in Ref. 20 for $l=0$ and $l=1$, respectively.

From the explicit formulas for a_{csl} and a_{sl}, one can find a relation between these two quantities. For low orders in ν (the ratio of a typical short-range parameter and the typical long-range parameter $|s|^{-1}=a_B$, the Bohr radius), we have obtained Eqs. (4.20)–(4.25). In the literature, one finds a number of relations between a_{csl} and a_{sl}. For the S-wave case, we mention an old and famous formula given three decades ago. Chew and Goldberger[33] then derived a relation valid for Coulomb repulsion and attraction:

$$a_{cs0}^{-1} = a_{s0}^{-1} - 2s[\ln 2R/a_B + \text{series in powers of } 2R/a_B]\,. \tag{6.3}$$

Here R stands for an effective range. Blatt and Jackson[1] quote a formula, with Coulomb repulsion in mind, from Schwinger.[6] Writing it for repulsion and attraction, we have

$$a_{cs0}^{-1} = a_{s0}^{-1} - 2s[\ln 2r_{cs0}/a_B + 2C - 0.824]\,. \tag{6.4}$$

Following these authors, we have taken the quantity r_{cs0} in the logarithmic term. The constant 0.824 is appropriate for a local square-well potential. It would be somewhat larger (but $\lesssim 1$) for a Yukawa-like potential, i.e., it is shape dependent. In fact, Sauer[12] has observed a considerable model dependence of the relation between a_{cs0} and a_{s0}. He considers certain unitary transformations which greatly affect the interior part of the radial wave functions.

Recently, Popov and collaborators[34] have presented a relation relating a_{csl} and a_{sl} for $l\geqslant 1$, see Ref. 14, p. 124. It reads

$$a_{csl}^{-1} = a_{sl}^{-1} - 2s[(2l-1)!!]^2\int_0^\infty dr(r\psi_l(r))^2/r\,. \tag{6.5}$$

Here $r\psi_l(r)$ is the radial wave function of the short-range interaction with zero binding energy, normalized according to $\lim_{r\to\infty} r^{l+1}\psi_l(r)=1$. Equation (6.5) can be easily checked in our explicit models (i)–(iii). For example, in case (i) we have $\psi_l(r) = r^lR^{-2l-1}\Theta(R-r) + r^{-l-1}\Theta(r-R)$. Hence it is trivial to verify that Eq. (6.5) leads to a relation which results after the two forms between square brackets in Eq. (4.21) are deleted.

Equations (4.20)–(4.25) exemplify what has already been stressed in Ref. 14 [Sec. 4.2.2, and, in particular, Eq. (4.44)]: The appearance of a correction factor of the type $1 + c_l\nu + O(\nu^2)$ multiplying the quantity a_{sl}^{-1}. Here c_l is some real constant. Such a factor is absent in relations (6.3)–(6.5). It becomes particularly important when the scattering lengths are small. Among others, this is the case in the limit of weak short-range coupling ($V_{sl}\to 0$).

The connection between the spectrum of $H = H_0 + V_c + V_s$ for hadronic atoms and (Coulomb-modified) scattering parameters has been studied by many authors.[35,7,8,36–38,20] Bound-state positions $-\kappa^2$ follow from the solution of $\cot \delta_{csl} - i = 0$, or [cf. Eq. (3.8) with $k \to + i\kappa$]

$$K_{csl}(-\kappa^2) = -2sH(is/\kappa) \prod_{m=1}^{l} (-\kappa^2 + s^2/m^2) . \quad (6.6)$$

This equation is rigorous, and it has been the starting point of many approximation schemes. These schemes relate the energy shift (due to V_s) to *Coulomb-modified* scattering parameters. By measuring the energy shifts experimentally, one can gain information about the purely internal scattering parameter a_{sl}, by invoking the relations connecting a_{csl} and a_{sl}. Inclusion of the correction factor discussed in the previous paragraph in that case may be very important.

APPENDIX A: TAYLOR EXPANSIONS OF ϕ AND χ IN k^2

The functions $\phi_{cl}(k,r)$ and $\chi_{cl}(k,r)$ are entire analytic in the variable k^2, for every $r > 0$. Their Taylor expansions are easily derived from the results given in Ref. 30.

For the repulsive Coulomb potential ($s < 0$), one has

$$\phi_{cl}(k,r) = (2l+1)! \left(\frac{x}{2}\right)^{-2l-1} r^{l+1} \sum_{n=0}^{\infty} k^{2n} s^{-2n} \times \sum_{m=2n}^{3n} a_{n,m}^{(l)} \left(-\frac{x}{2}\right)^m I_{2l+1+m}(x) , \quad (A1)$$

$$\chi_{cl}(k,r) = \frac{(2l+1)!}{(l!)^2} s^{2l+1} \left(\frac{x}{2}\right)^{-2l-1} r^{l+1} \times \sum_{n=0}^{\infty} k^{2n} s^{-2n} \sum_{p=0}^{n} \left[\sum_{m=1}^{p} \frac{|B_{2m}|}{m} c_{p-m}(l) \times \sum_{q=2n-2p}^{3n-3p} a_{n-p,q}^{(l)} \left(-\frac{x}{2}\right)^q I_{2l+1+q}(x) \quad (A2) - c_p(l) \sum_{q=2n-2p}^{3n-3p} a_{n-p,q}^{(l)} \left(\frac{x}{2}\right)^q 4K_{2l+1+q}(x) \right] .$$

Here $a_{n,m}^{(l)}$ and $c_n(l)$ are defined (and partly tabulated) in Sec. 7 of Ref. 30, see also the Appendix of Ref. 20. Furthermore,

$$x = 2\sqrt{-2sr} , \quad (A3)$$

and $\Sigma_{m=1}^{p} \cdots = 0$ if $p = 0$ has been assumed in writing down Eq. (A2). The B_{2m} are the Bernoulli numbers.[23] Note that in Ref. 30 a different convention for these numbers has been used.

For the attractive Coulomb potential ($s > 0$), one has

$$\phi_{cl}(k,r) = (2l+1)! \left(\frac{\xi}{2}\right)^{-2l-1} r^{l+1} \sum_{n=0}^{\infty} k^{2n} s^{-2n} \times \sum_{m=2n}^{3n} a_{n,m}^{(l)} \left(\frac{\xi}{2}\right)^m J_{2l+1+m}(\xi) , \quad (A4)$$

$$\chi_{cl}(k,r) = \frac{(2l+1)!}{(l!)^2} s^{2l+1} \left(\frac{\xi}{2}\right)^{-2l-1} r^{l+1} \times \sum_{n=0}^{\infty} k^{2n} s^{-2n} \sum_{p=0}^{n} \left[\sum_{m=1}^{p} \frac{|B_{2m}|}{m} c_{p-m}(l) \times \sum_{q=2n-2p}^{3n-3p} a_{n-p,q}^{(l)} \left(\frac{\xi}{2}\right)^q J_{2l+1+q}(\xi) - c_p(l) \times \sum_{q=2n-2p}^{3n-3p} a_{n-p,q}^{(l)} \left(\frac{\xi}{2}\right)^q 2\pi N_{2l+1+q}(\xi) \right] , \quad (A5)$$

where

$$\xi = 2\sqrt{2sr} . \quad (A6)$$

APPENDIX B: ON THE INTERCHANGE OF THE ORDER OF SUMMATION AND INTEGRATION ON THE RHS OF EQ. (4.1)

Without loss of generality, we can take $b = \infty$. When either

$$\sum_{n=0}^{\infty} |z - z_0|^n \int_0^{\infty} dt\, |f_n(t)| \quad (B1)$$

converges, or

$$\sum_{n=0}^{\infty} |z - z_0|^n |f_n(t)| \quad (B2)$$

converges pointwise to an integrable function h on $[0, \infty]$, then the series

$$\sum_{n=0}^{\infty} f_n(t)(z - z_0)^n \quad (B3)$$

converges pointwise almost everywhere on $[0, \infty]$ to an integrable function $f(z,t)$ and

$$F(z) := \int_0^{\infty} dt\, f(z,t) = \int_0^{\infty} dt \sum_{n=0}^{\infty} f_n(t)(z - z_0)^n = \sum_{n=0}^{\infty} (z - z_0)^n \int_0^{\infty} dt\, f_n(t) . \quad (B4)$$

For the proof of this, we refer to Ref. 39. Suppose further that F is analytic on $|z - z_0| < \rho$. Then

$$F(z) = \sum_{n=0}^{\infty} F_n (z - z_0)^n , \quad |z - z_0| < \rho . \quad (B5)$$

Since a power series for an analytic function is unique, it follows that

$$F_n = \int_0^{\infty} dt\, f_n(t) . \quad (B6)$$

APPENDIX C: DERIVATION OF THE EFFECTIVE-RANGE PARAMETERS FOR THE UNIT-STEP FORM FACTOR

In this Appendix we give some essential steps in the derivation of the Coulomb-modified effective-range parameters for the separable potential given by Eq. (3.5) with form factor (4.3). The effective-range function K_{csl} is given by Eq. (3.17). The functions $\mathscr{I}_{\phi\phi}(k^2)$ and $\mathscr{I}_{\chi\phi}(k^2)$ have Taylor-series expansions in k^2. Given these expansions it is straightforward to compute the coefficients $-a_{csl}^{-1}$, $\frac{1}{2}r_{csl}$,... on the RHS of Eq. (3.9).

The expansion for $\mathscr{J}_{\phi\phi}$ is obtained by using the indefinite integral of Ref. 22, p. 87. It follows that

$$\int_0^\infty dr\, r\phi_{cl}(k,r) r^{l-1}\Theta(1-r/R) = (2l+1)!(4s)^{-2l-2}\sum_{n=0}^{\infty} k^{2n}s^{-2n} \times \sum_{m=2n}^{3n} a_{n,m}^{(l)}(-2)^{-m}x_0^{2l+2+m}I_{2l+2+m}(x_0), \tag{C1}$$

with

$$x_0 = 2\sqrt{-2sR}\,. \tag{C2}$$

The expansion for $\mathscr{J}_{\chi\phi}$ can be obtained in closed form, too, from

$$\int_0^\infty dr'\, r' \int_0^\infty dr\, r\, r'^{l-1}\Theta(1-r'/R)\chi_{cl}(k,r_>)\phi_{cl}(k,r_<)r^{l-1}\Theta(1-r/R)$$

$$= 2\left[\frac{(2l+1)!}{l!}\right]^2 s^{2l+1}(4s)^{-4l-4}\sum_{u=0}^{\infty}k^{2u}s^{-2u}\sum_{v=2u}^{3u}a_{u,v}^{(l)}(-2)^{-v}\sum_{n=0}^{\infty}k^{2n}s^{-2n}\sum_{p=0}^{n}\left[\sum_{m=1}^{p}\frac{|B_{2m}|}{m}c_{p-m}(l)\right.$$
$$\times\sum_{q=2n-2p}^{3n-3p}a_{n-p,q}^{(l)}(-2)^{-q}\frac{x_0^{4l+5+q+v}}{2(4l+4+q+v)}\Big(I_{2l+1+q}(x_0)I_{2l+2+v}(x_0)+I_{2l+2+q}(x_0)I_{2l+3+v}(x_0)\Big)-c_p(l)$$
$$\left.\times\sum_{q=2n-2p}^{3n-3p}a_{n-p,q}^{(l)}2^{2-q}\frac{x_0^{4l+5+q+v}}{2(4l+4+q+v)}\Big(K_{2l+1+q}(x_0)I_{2l+2+v}(x_0)+K_{2l+2+q}(x_0)I_{2l+3+v}(x_0)\Big)\right], \tag{C3}$$

cf. Eq. 11.3.31 of Ref. 23, p. 484.

APPENDIX D: DERIVATION OF a_{csl}, r_{csl},... FOR YAMAGUCHI-TYPE POTENTIALS

In this appendix, we give some essential steps in the derivation of the Coulomb-modified effective-range parameters for the separable potential given by Eq. (3.5) with form factor (4.4). The effective-range function K_{csl} is given by Eq. (3.17). The full integral $\mathscr{J}_{\phi\phi}$ factorizes into two factors which are easily calculated (cf. Ref. 22, p. 278) using

$$\int_0^\infty dr\, r^{2l+1}\exp(ikr - r/R)\,{}_1F_1(l+1+i\gamma;2l+2;-2ikr)$$
$$= \frac{(2l+1)!R^{2l+2}}{(1+k^2R^2)^{l+1}}\left(\frac{1-ikR}{1+ikR}\right)^{i\gamma},\quad R^{-1} > \mathrm{Re}(-ik)\,,$$
$$|R^{-1}-ik| > |-2ik|\,. \tag{D1}$$

This leads immediately to an expression for the first term on the RHS of (3.17), namely, $\lambda_l^{-1}(1+k^2R^2)^{2l+2}R^{-4l-4}B^{2i\gamma}$, with $B=(1+ikR)/(1-ikR)$. This expression is easily expanded in a (Taylor) series of powers of k^2.

The first term in this series can also be directly derived from the series representation (A1), by using (cf. Ref. 22, p. 93)

$$\int_0^\infty dr\, r(2l+1)!(-2s)^{-l-1}(-2sr)^{1/2}$$
$$\times I_{2l+1}((-8sr)^{1/2})r^{l-1}e^{-r/R}$$
$$= (2l+1)!R^{2l+2}\exp(-2sR)\,. \tag{D2}$$

Similarly, the lowest-order term in the Taylor-series expansion of $\mathscr{J}_{\chi\phi}$ can be computed directly by using the lowest-order terms of the expansions (A1) and (A2) in Eq. (3.15). To this end, we first note that (recall $\nu \equiv -sR \equiv k\gamma R$)

$$\int_0^r dr'\, r'(-2sr')^{1/2}I_{2l+1}(\sqrt{-8sr'})r'^{l-1}e^{-r'/R}$$
$$= R^{l+1}(2\nu)^{-l}\sum_{m=0}^{\infty}\frac{(2\nu)^{2l+1+m}}{m!(2l+1+m)!}$$
$$\times\gamma(2l+2+m,r/R)\,, \tag{D3}$$

see Ref. 22, p. 337, where an integral representation for $\gamma(a,x)$ is given. We then use the integral representation for the Bessel function of the third kind (see Ref. 22, p. 85),

$$2(z/\zeta)^{\mu/2}K_\mu[2(z\zeta)^{1/2}] = \int_0^\infty dt\, e^{-zt}e^{-\zeta/t}t^{-\mu-1}\,,$$
$$\mathrm{Re}\,z>0,\quad \mathrm{Re}\,\zeta>0\,, \tag{D4}$$

and, furthermore (Ref. 24, Formula 6.451.1),

$$\int_0^\infty dt\, e^{-t(1+u)}\gamma(m+2l+2,t)$$
$$= (1+u)^{-1}\Gamma(m+2l+2)(2+u)^{-m-2l-2}\,, \tag{D5}$$

to obtain

$$\int_0^\infty dr\, r(-2sr)^{1/2}K_{2l+1}(\sqrt{-8sr})r^{l-1}e^{-r/R}$$
$$\times\int_0^r dr'\, r'(-2sr')^{1/2}I_{2l+1}(\sqrt{-8sr'})$$
$$\times r'^{l-1}e^{-r'/R} = \tfrac{1}{4}(2s)^{2l+2}R^{4l+4}(2l+1)!$$
$$\times U(2l+2,2l+2,4\nu)$$
$$= \tfrac{1}{4}(2s)^{2l+2}R^{4l+4}(2l+1)!e^{4\nu}\Gamma(-2l-1,4\nu)\,. \tag{D6}$$

Collection of the lowest-order terms on the RHS of Eq. (3.17), with use of the results (D2) and (D6) gives immediately the result (4.12) for a_{csl}.

Using similar relations, it is not difficult (but somewhat laborious) to obtain the general formula (4.13) for r_{csl}.

[1]J. M. Blatt and J. D. Jackson, Phys. Rev. **76**, 18 (1949); J. D. Jackson and J. M. Blatt, Rev. Mod. Phys. **22**, 77 (1950).
[2]H. A. Bethe, Phys. Rev. **76**, 38 (1949).
[3]G. Breit, E. Condon, and R. Present, Phys. Rev. **50**, 825 (1936); G. Breit, B. Thaxton, and L. Eisenbud, Phys. Rev. **55**, 1018 (1939).
[4]L. D. Landau and Ya. A. Smorodinsky, Zh. Eksp. Teor. Fiz. **14**, 269 (1944).
[5]G. F. Chew and M. L. Goldberger, Phys. Rev. **75**, 1637 (1949).
[6]J. Schwinger, Phys. Rev. **78**, 135 (1950).
[7]T. L. Trueman, Nucl. Phys. **26**, 57 (1961).
[8]A. Partensky and M. Ericson, Nucl. Phys. B **1**, 382 (1967).
[9]E. Lambert, Helv. Phys. Acta **42**, 667 (1969). Lambert's function $\Theta_l(k,r)$ equals $[(2l+1)!!]^{-2}$ times our function $\chi_{cl}(k,r)$.
[10]H. Cornille and A. Martin, Nuovo Cimento **26**, 298 (1962).
[11]J. Hamilton, I. Øverbö, and B. Tromborg, Nucl. Phys. B **60**, 443 (1973).
[12]P. U. Sauer, Phys. Rev. Lett. **32**, 626 (1974); Phys. Rev. C **11**, 1786 (1975).
[13]H. van Haeringen, Ph.D. thesis (Free University, Amsterdam, 1978) (unpublished).
[14]A. M. Badalyan, L. P. Kok, M. I. Polikarpov, and Yu. A. Simonov, Phys. Rep. **82**, 31 (1982).
[15]H. van Haeringen and L. P. Kok, Phys. Rev. A **26**, 1218 (1982).
[16]Y. Yamaguchi, Phys. Rev. **95**, 1628 (1954).
[17]H. van Haeringen, Nucl. Phys. A **253**, 355 (1975).
[18]H. van Haeringen, J. Math. Phys. **18**, 927 (1977).
[19]H. van Haeringen and L. P. Kok, Phys. Lett. A **82**, 317 (1981).
[20]L. P. Kok, J. W. de Maag, H. H. Brouwer, and H. van Haeringen, Phys. Rev. C **26**, 2381 (1982).
[21]H. van Haeringen and L. P. Kok, Reports 128–151 (University of Groningen, 1979–1980).
[22]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, Berlin, 1966).
[23]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1965). In this reference the functions n_l and $N_{l+1/2}$ are denoted by Y_l and $Y_{l+1/2}$, respectively; cf. pp. 358 and 437.
[24]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products, Corrected and Enlarged Edition* (Academic, New York, 1980).
[25]Bateman Manuscript Project: *Higher Transcendental Functions*, edited by A. Erdélyi *et al.* (McGraw-Hill, New York, 1953), Vols. I and II.
[26]H. Poincaré, Acta Math. **4**, 213 (1884).
[27]M. Gell-Mann and M. L. Goldberger, Phys. Rev. **91**, 398 (1953).
[28]H. van Haeringen, J. Math. Phys. **17**, 995 (1976).
[29]R. G. Newton, *Scattering Theory of Waves and Particles*, 2nd ed. (Springer, New York, 1982).
[30]M. H. Hull, Jr. and G. Breit, in *Encyclopedia of Physics*, edited by S. Flügge (Springer, Berlin, 1959), Vol. XLI/1, p. 408. Note that the convention for the modified Bessel function K_n used by Hull and Breit differs from that used in, for example, Magnus *et al.*, cf. Ref. 22, by a factor $(-)^n$, for integral n.
[31]Y. Nogami and W. van Dijk, Can. J. Phys. **60**, 205 (1982).
[32]A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1969), Vols. 1 and 2.
[33]G. F. Chew and M. L. Goldberger, Phys. Rev. **75**, 1637 (1949).
[34]V. S. Popov *et al.*, Zh. Eksp. Teor. Fiz. **80**, 1271 (1981) [Sov. Phys. JETP **53**, 650 (1981)]; Preprint ITEP-180, Moscow, 1980 (Russian).
[35]S. Deser, M. L. Goldberger, K. Baumann, and W. Thirring, Phys. Rev. **96**, 774 (1954).
[36]V. B. Mandelzweig, Nucl. Phys. A **292**, 333 (1977).
[37]A. E. Kudryavtsev, V. I. Lisin, V. D. Mur, and V. S. Popov, Preprint ITEP-58, Moscow, 1982.
[38]A. Deloff, Phys. Rev. C **13**, 730 (1976); A. Deloff and J. Law, Phys. Rev. C **20**, 1597 (1979).
[39]J. Korevaar, *Mathematical Methods* (Academic, New York, 1968), Vol. 1, cf. pp. 366–395.

On the survival of primordial tachyons up to the present epoch

Sushil K. Srivastava
Government College, Port Blair 744104, India

(Received 29 October 1982; accepted for publication 25 February 1983)

In the present paper, a spinless primordial tachyon is considered in the background of Robertson–Walker (RW) cosmology. Various physical parameters, such as probability density, energy, and dissipation of energy are investigated in open, flat, and closed RW models. For this purpose, the Klein–Gordon equation, generalized to curved space-time and having a term proportional to the scalar curvature of the space-time, is solved for the spacelike scalar field $\psi(t,r,\theta,\phi)$. It is speculated that if a primordial spinless tachyon survives up to the present epoch, its metamass would be less than 2.447×10^{-93} g.

PACS numbers: 14.80.Pb, 98.80.Dr

1. INTRODUCTION

Now it is widely accepted that existence of tachyons does not violate the theory of relativity. But so far these particles could not be detected and produced in the laboratory. Some particle physicists, such as Arons and Sudarshan,[1] Feinberg,[2] and Dhar and Sudarshan,[3] have discussed quantum-mechanical properties of these particles which have some relevance to their production and detection. In 1976, Narlikar and Sudarshan[4] have suggested that as far as production of tachyons is concerned, we should not confine our attempts to the terrestrial laboratory alone, but we should pay some attention to astronomical discoveries also. There have been many astronomical discoveries of such particles that could not be produced in the laboratory. One of the extreme events is the big bang, if the universe originated in a big bang. They (Narlikar and Sudarshan) have assumed that tachyons also were produced at or just after the epoch of big bang along with many other particles.

If one assumes the production of tachyons like Narlikar and Sudarshan,[4] and as we do here, he will naturally ask "What happened to primordial tachyons? Do they survive up to the present epoch?" Narlikar and Sudarshan[4] have discussed many features of primordial tachyons in cosmological background taking the Robertson–Walker model of zero space curvature. To answer the above questions, they infer from their investigations that if tachyons survive up to the present epoch, their metamass would be less than even the rest mass of an electron.

In the present paper, we also deal with the question of survival of spinless primordial tachyons in the background of Robertson–Walker cosmology in more detail taking open, flat as well as closed models. The author[5] has published one paper in which he has considered a spinless tachyon in the closed RW model. In another paper[6] he has considered spin-$\frac{1}{2}$ primordial tachyons in RW cosmology and discussed similar questions. This paper is different from the above papers in certain aspects. In this paper we have solved the Klein–Gordon equation generalized to curved space-time of the Friedmann universe and having a term proportional to scalar curvature R which enables us to deal with the case of conformal coupling of the scalar and spacelike field ψ to the gravitational field. Thus here we deal with the more generalized case.

We consider the Robertson–Walker line element

$$ds^2 = c^2\,dt^2 - S^2(t)[dr^2/(1-kr^2) + r^2(d\theta^2 + \sin^2\theta\,d\phi^2)], \tag{1.1}$$

where t is the cosmic time along the hypersurface $x^2 = \text{const}$ $(i = 1,2,3)$ and k is the space curvature with possible values $-1, 0, +1$ for open, flat, and closed models, respectively. Under the coordinate transformations

$$\tau = \int_0^t \frac{dt}{S(t)} \quad \text{and} \quad \sigma = \int_0^r \frac{dr}{\sqrt{1-kr^2}} = \frac{1}{\sqrt{k}}\sin^{-1}(r\sqrt{k}), \tag{1.2}$$

the line element (1.1) is written as

$$dS^2 = \Omega^2(\tau)[d\tau^2 - d\sigma^2 - (1/k)\sin^2(\sigma\sqrt{k})(d\theta^2 + \sin^2\theta d\phi^2)], \tag{1.3}$$

where $\Omega(\tau) = S(t)$.

In Sec. 2, we solve the Klein–Gordon equation

$$(-g)^{-1/2}\frac{\partial}{\partial x^\mu}\left[(-g)^{1/2}g^{\mu\nu}\frac{\partial}{\partial x^\nu}\right]\psi + \frac{1}{\hbar^2}(\gamma R + m^2)\psi = 0 \tag{1.4}$$

for different values of k. Here R is the scalar curvature and γ is a real constant. For our investigations, like Parker,[7] we take $\gamma = \frac{1}{6}$. As it appears from the generalized Klein–Gordon equation given above, the gravitational effect is incorporated through the metric tensor $g^{\mu\nu}$ and scalar curvature R. $\hbar$, in the above equation, is Planck's constant divided by 2π.

In Sec. 3, we derive expressions for probability density of primordial tachyons in open, flat, and closed models of the universe.

In Sec. 4, we derive expressions for energy as well as dissipation of energy of primordial tachyons in different models. It is observed that dissipation of energy is fast in the beginning but it slows down when t is large.

In Sec. 5, with the help of the uncertainty principle we speculate that the metamass of a primordial tachyon surviving up to the present epoch would be less than 2.447×10^{-93} g.

2. SOLUTION OF THE KLEIN–GORDON EQUATION FOR DIFFERENT VALUES OF k

Under coordinate transformations (1.2), the Klein–Gordon equation (1.4) is written as

$$(-\bar{g})^{-1/2}\frac{\partial}{\partial x^{\mu}}\Big[(-\bar{g})^{1/2}\bar{g}^{\mu\nu}\frac{\partial}{\partial x^{\nu}}\Big]\bar{\psi}+\frac{1}{\hbar^2}\times(\gamma\bar{R}+m^2)\Omega^2\bar{\psi}=0, \tag{2.1}$$

where $\bar{\psi}=\Omega(\tau)\psi$ and $\bar{g}^{\mu\nu}$ is the metric tensor provided by the line element (1.3) and $\bar{R}$ is the scalar curvature of the space-time derived from $\bar{g}^{\mu\nu}$.

On substituting $\bar{g}^{\mu\nu}$ and

$$m=iM \tag{2.2}$$

(where M is the metamass[8] of the tachyon) we have Eq. (2.1) as

$$\frac{1}{c^2}\frac{\partial^2\bar{\psi}}{\partial\tau^2}+\frac{2}{c^2\Omega}\frac{\partial\Omega}{\partial\tau}\frac{\partial\bar{\psi}}{\partial\tau}-\frac{\partial^2\bar{\psi}}{\partial\sigma^2}-2\sqrt{k}\cot(\sigma\sqrt{k})\frac{\partial\bar{\psi}}{\partial\sigma}-\frac{k}{\sin^2(\sigma\sqrt{k})}\Big[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\Big(\sin\theta\frac{\partial\bar{\psi}}{\partial\theta}\Big)+\frac{1}{\sin^2\theta}\frac{\partial^2\bar{\psi}}{\partial\phi^2}\Big]+\frac{1}{\hbar^2}\Big[\frac{k+2}{3}+\frac{\dot{\Omega}^2}{3\Omega^2}+\frac{\ddot{\Omega}}{\Omega}-M^2\Omega^2\Big]\bar{\psi}=0, \tag{2.3}$$

where the dot over the variable denotes differentiation with respect to τ.

$\bar{\psi}$ can be expanded in terms of a complete set of eigenfunctions of the angular momentum operator. Hence, setting

$$\bar{\psi}=\sum_{l=0}^{\infty}\bar{\psi}_l P_l(\cos\theta), \tag{2.4a}$$

we have the partial differential equation of $\bar{\psi}_l(\tau,\sigma)$ as

$$\frac{1}{c^2}\frac{\partial^2\bar{\psi}_l}{\partial\tau^2}+\frac{2}{c^2\Omega}\frac{\partial\Omega}{\partial\tau}\frac{\partial\bar{\psi}_l}{\partial\tau}-\frac{\partial^2\bar{\psi}_l}{\partial\sigma^2}-2\sqrt{k}\cot(\sigma\sqrt{k})\frac{\partial\bar{\psi}_l}{\partial\sigma}+\frac{kl(l+1)}{\sin^2(\sigma\sqrt{k})}\bar{\psi}_l+\frac{1}{\hbar^2}\times\Big[\frac{k+2}{3}+\frac{\dot{\Omega}^2}{3\Omega^2}+\frac{\ddot{\Omega}}{\Omega}-M^2\Omega^2\Big]\bar{\psi}_l=0. \tag{2.4b}$$

The velocity of the tachyon is given by

$$v=\frac{S(t)}{\sqrt{1-kr^2}}\frac{dr}{dt}=S(t)\frac{d\sigma}{dt}>1; \tag{2.5a}$$

hence,

$$\sigma>\int\frac{dt}{S(t)}. \tag{2.5b}$$

This shows that σ will increase rapidly. Hence there is no harm in taking the limit $\sigma\to\infty$ as

$$\cot(\sigma\sqrt{k})=\begin{cases}0 & \text{when } k=+1,\\ i^3 & \text{when } k=-1,\end{cases} \tag{2.6}$$

where $i=\sqrt{-1}$.

Also we find that

$$\frac{1}{\Omega}\frac{\partial\Omega}{\partial\tau}=H\Omega=HS, \tag{2.7}$$

where H is Hubble's constant.

After the epoch of big bang, we assume that at time t_e the entire energy of the universe came into thermal equilibrium. Now approximating HS near $t=t_e$ we have[5]

$$HS=H_eS_e+(t-t_e)\Big[H_e\Big(\frac{\partial S}{\partial t}\Big)_{t_e}+\Big(\frac{\partial H}{\partial t}\Big)_{t_e}S_e\Big].$$

But up to the time t_e expansion of the universe would have been very small; hence S_e may be approximated to zero. As a result, we have

$$HS\approx 0. \tag{2.8}$$

Hence the partial differential equation (2.4) becomes

$$\frac{1}{c^2}\frac{\partial^2\bar{\psi}_l}{\partial\tau^2}-\frac{\partial^2\bar{\psi}_l}{\partial\sigma^2}-2\sqrt{k}\cot(\sigma\sqrt{k})\frac{\partial\bar{\psi}_l}{\partial\sigma}+\frac{kl(l+1)}{\sin^2(\sigma\sqrt{k})}\bar{\psi}_l+\frac{1}{\hbar^2}\Big[\frac{k+2}{3}+\frac{\dot{\Omega}^2}{3\Omega^2}+\frac{\ddot{\Omega}}{\Omega}-M^2\Omega^2\Big]\bar{\psi}_l=0. \tag{2.9}$$

Now we suppose the solution of the partial differential equation (2.9) to be

$$\bar{\psi}_l=\Phi(\sigma)\exp(-i\nu\tau). \tag{2.10}$$

Combining Eqs. (2.9) and (2.10) we have the ordinary differential equation

$$\frac{d^2\Phi}{d\sigma^2}+2\sqrt{k}\cot(\sigma\sqrt{k})\frac{d\Phi}{d\sigma}+\Big[\frac{\nu^2}{c^2}+\frac{kl(l+1)}{\sin^2(\sigma\sqrt{k})}-\frac{1}{\hbar^2}\Big(\frac{k+2}{3}+\frac{\dot{\Omega}^2}{3\Omega^2}+\frac{\ddot{\Omega}}{\Omega}-M^2\Omega^2\Big)\Big]\Phi=0. \tag{2.11}$$

Substituting different values of k we solve the ordinary differential equation (2.11).

Case I. When $k=-1$, with the help of approximation (2.6) we have the differential equation (2.11) as

$$\frac{d^2\Phi}{d\sigma^2}+2\frac{d\Phi}{d\sigma}+\Big[\frac{\nu^2}{c^2}-\frac{1}{3\hbar^2}-\frac{\dot{\Omega}^2}{3\Omega^2\hbar^2}-\frac{\dot{\Omega}}{\Omega\hbar^2}+\frac{M^2\Omega^2}{\hbar^2}\Big]\Phi=0. \tag{2.12}$$

This equation is integrated into

$$\Phi=A\exp\Big[-\sigma\pm i\sigma\Big(\frac{\nu^2}{c^2}-\frac{1}{3\hbar^2}-\frac{\dot{\Omega}^2}{3\Omega^2\hbar^2}-\frac{\ddot{\Omega}}{\Omega\hbar^2}+\frac{M^2\Omega^2}{\hbar^2}\Big)^{1/2}\Big]. \tag{2.13}$$

Hence

$$[\bar{\psi}_l]_{k=-1}A\exp\Big[-\sigma-i\nu\tau+i\sigma\Big(\frac{\nu^2}{c^2}-\frac{1}{3\hbar^2}-\frac{\dot{\Omega}^2}{3\Omega^2\hbar^2}-\frac{\ddot{\Omega}}{\Omega\hbar^2}+\frac{M^2\Omega^2}{\hbar^2}\Big)^{1/2}\Big]. \tag{2.14}$$

Also

$$[\psi_l]_{k=-1}=A\Omega^{-1}\exp\Big[-\sigma-i\nu\tau+i\sigma\Big(\frac{\nu^2}{c^2}-\frac{1}{3\hbar^2}-\frac{\dot{\Omega}^2}{3\Omega^2\hbar^2}-\frac{\ddot{\Omega}}{\Omega\hbar^2}+\frac{M^2\Omega^2}{\hbar^2}\Big)^{1/2}\Big]. \tag{2.15}$$

Case II: When $k = 0$, on taking the limit $r \rightarrow \infty$ the differential equation (2.11) reduces to

$$\frac{d^2\Phi}{dr^2} + \left[\frac{v^2}{c^2} - \frac{2}{3\hbar^2} - \frac{\dot{\Omega}^2}{3\Omega^2\hbar^2} - \frac{\ddot{\Omega}}{\Omega\hbar^2} + \frac{M^2\Omega^2}{\hbar^2}\right]\Phi = 0. \tag{2.16}$$

This equation is integrated into

$$\Phi = A' \exp\left[\pm ir\left(\frac{v^2}{c^2} - \frac{2}{3\hbar^2} - \frac{\Omega^2}{3\Omega^2\hbar^2} - \frac{\Omega}{\Omega\hbar^2} + \frac{M^2\Omega^2}{\hbar^2}\right)^{1/2}\right], \tag{2.17}$$

which yields

$$[\bar{\psi}_l]_{k=0} = A' \exp\left[-iv\tau \pm ir\left(\frac{v^2}{c^2} - \frac{2}{3\hbar^2} - \frac{\dot{\Omega}^2}{3\Omega^2\hbar^2} - \frac{\ddot{\Omega}}{\Omega\hbar^2} + \frac{M^2\Omega^2}{\hbar^2}\right)^{1/2}\right] \tag{2.18}$$

and

$$[\psi_l]_{k=0} = A'\Omega^{-1} \exp\left[-iv\tau + ir\left(\frac{v^2}{c^2} - \frac{2}{3\hbar^2} - \frac{\dot{\Omega}^2}{3\Omega^2\hbar^2} - \frac{\ddot{\Omega}}{\Omega\hbar^2} + \frac{M^1\Omega^2}{\hbar^2}\right)^{1/2}\right]. \tag{2.19}$$

Case III: When $k = +1$, on using the approximation (2.6), the differential equation (2.11) reduces to

$$\frac{d^2\Phi}{d\sigma^2} + \left[\frac{v^2}{c^2} - \frac{1}{\hbar^2} - \frac{\dot{\Omega}^2}{3\Omega^2\hbar^2} - \frac{\ddot{\Omega}}{\Omega\hbar^2} + \frac{M^2\Omega^2}{\hbar^2}\right]\Phi = 0. \tag{2.20}$$

This equation is integrated into

$$\Phi = A'' \exp\left[\pm i\sigma\left(\frac{v^2}{c^2} - \frac{1}{\hbar^2} - \frac{\dot{\Omega}^2}{3\Omega^2\hbar^2} - \frac{\ddot{\Omega}}{\Omega\hbar^2} + \frac{M^2\Omega^2}{\hbar^2}\right)^{1/2}\right], \tag{2.21}$$

which yields

$$[\bar{\psi}_l]_{k=+1} A'' \exp\left[-iv\tau \pm i\sigma\left(\frac{v^2}{c^2} - \frac{1}{\hbar^2} - \frac{\dot{\Omega}^2}{3\Omega^2\hbar^2} - \frac{\ddot{\Omega}}{\Omega\hbar^2} + \frac{M^2\Omega^2}{\hbar^2}\right)^{1/2}\right]. \tag{2.22}$$

Also

$$[\psi_l]_{k=+1} A''\Omega^{-1} \exp\left[-iv\tau \pm i\sigma\left(\frac{v^2}{c^2} - \frac{1}{\hbar^2} - \frac{\dot{\Omega}^2}{3\Omega^2\hbar^2} - \frac{\ddot{\Omega}}{\Omega\hbar^2} + \frac{M^2\Omega^2}{\hbar^2}\right)^{1/2}\right]. \tag{2.23}$$

The scalar wavefunction ψ_l, in this case, can be normalized to unity as

$$\int \psi_l{}^*\psi_l \, d^3V = 1$$

($*$ on ψ_l denotes the usual meaning), i.e.,

$$\int \frac{A''^2}{\Omega^2}\Omega^2 \, d\Omega \, d\omega = 1$$

(here $d\omega = \sin\theta \, d\theta \, d\phi$), which yields

$$A'' = \frac{1}{2\sqrt{\pi}}\Omega^{-1/2}. \tag{2.24}$$

Hence

$$[\psi_l]_{k=+1} = \frac{1}{2\sqrt{\pi}}\Omega^{-3/2} \exp\left[-iv\tau \pm i\sigma\left(\frac{v^2}{c^2} - \frac{1}{\hbar^2} - \frac{\dot{\Omega}^2}{3\Omega^2\hbar^2} - \frac{\ddot{\Omega}}{\Omega\hbar^2} + \frac{M^2\Omega^2}{\hbar^2}\right)^{1/2}\right]. \tag{2.25}$$

The solution (2.22) is different from the author's solution for this case in his previous[5] paper due to the presence of an extra term proportional to the scalar curvature.

Our solution in the case $k = 0$ is different from that of Narlikar and Sudarshan[4] due to two reasons: (1) they have omitted the term proportional to the scalar curvature of the space-time in the Klein–Gordon equation and (2) they have shown that no tachyon can live in this universe beyond an epoch $t = t_{\text{maximum}}$ and have used this result throughout their entire paper, whereas we do not start our investigations with such notions.

3. PROBABILITY DENSITY OF PRIMORDIAL TACHYONS IN RW MODELS

The probability density is defined as

$$\rho = \psi^*\psi. \tag{3.1}$$

Case I: When $k = -1$, we have the probability density from Eq. (2.15) as

$$[\rho]_{k=-1} = A^2/e^{2\sigma}\Omega^2(\tau). \tag{3.2}$$

Case II: When $k = 0$ we have the probability density from Eq. (2.19) as

$$[\rho]_{k=0} = A'^2/\Omega^2(\tau). \tag{3.3}$$

Case III: When $k = +1$, we have the probability density, as given by Eq. (2.25),

$$[\rho]_{k=+1} = 1/4\pi\Omega^3. \tag{3.4}$$

Einstein field equations with suitable equations of state yield[9]

$$\Omega(\tau) = S(t) = t^q, \tag{3.5}$$

where q has different values $\frac{2}{3}$, $\frac{1}{2}$, $\frac{1}{3}$, and $\frac{2}{3}$ for RW models containing different perfect fluids as bradyon dust, radiation, superdense bradyonic matter, and nonrelativistic matter, respectively.

Connecting Eqs. (3.2), (3.3), and (3.4) with Eq. (3.5), we can easily study how the probability density for a primordial tachyon will decrease in different RW models containing different perfect fluids. Moreover, we find that if we take the limit $\sigma \to \infty$, as earlier, the probability density in the open model of the Friedmann universe will be almost zero. Also, it is interesting to note that the decay of the probability is faster in the case of the closed ($k = +1$) model than the flat ($k = -1$) model. Any way, the probability density of a primodial tachyon decreases very fast in every model.

4. ENERGY AND DISSIPATION OF ENERGY OF PRIMORDIAL TACHYON IN RW MODELS CONTAINING DIFFERENT PERFECT FLUIDS

Connecting Eqs. (1.2) and (2.5) with (3.5) we have

$$\tau = t^{1-q}/(1-q)$$

and

$$\sigma = v \int \frac{dt}{t^q} = \frac{v}{1-q} t^{1-q}, \tag{4.1}$$

where v is the velocity of a free tachyon.

Substituting σ in Eq. (2.14) we have

$$[\bar{\psi}_l]_{k=-1} = A \exp\left[-\frac{vt^q}{1-q} - \frac{ivt^q}{1-q} \pm \frac{ivt^q}{1-q} \right]\left(\frac{v^2}{c^2} - \frac{1}{3\hbar^2} - \frac{\dot{\Omega}^2}{3\Omega^2\hbar^2} - \frac{\ddot{\Omega}}{\Omega\hbar^2} + \frac{M^2\Omega^2}{\hbar^2} \right)^{1/2}. \tag{4.2}$$

The magnitude of energy associated with this wave is given by

$$E = \left| \frac{d}{dt}\left[\frac{-ivt^{1-q}}{1-q} \pm iv\frac{t^{1-q}}{1-q}\left(\frac{v^2}{c^2} - \frac{1}{3\hbar^2} - \frac{\dot{\Omega}^2}{3\Omega^2\hbar^2} - \frac{\ddot{\Omega}}{\Omega\hbar^2} + \frac{M^2\Omega^2}{\hbar^2} \right)^{1/2} \right] \right|$$

$$= \frac{1}{\hbar}\left| \left[-vt^{-q}\hbar \pm \frac{vt^{-2-3q}}{6(1-q)M}(q - 2q2 - q3 - 2M^2qt^{2q+2}) \right.\right.$$

$$\left.\left. \pm \frac{v}{M}t^{-1-3q}\left(\frac{1}{3}t^2 + \frac{4q^2}{3} - q - M^2t^{2q+2} \right) \right]\left(1 - \frac{1}{6M^2}t^{-2q} - \frac{1}{6q^2}t^{-2-2q} - q(q-1)t^{-2q-2} \right) \right|$$

$$\approx \frac{1}{\hbar}\left| \left[\pm \frac{vt^{-2-3q}}{6(1-q)M}(q - 2q^2 - q^3 - 2M^2qt^{2q+2}) \pm \frac{v}{M}t^{-1-3q}\left(\frac{1}{3}t^2 + \frac{4q^2}{3} - q - M^2t^{2q+2} \right) \right] \right.$$

$$\left. \times \left(1 - \frac{1}{6M^2}t^{-2q} - \frac{1}{6q^2}t^{2-2q} - q(q-1)t^{-2q-2} \right) \right|$$

(neglecting the term containing $\hbar$ within the modulus)

$$\approx \frac{vMq}{3(1-q)\hbar}t^{-q} + \frac{v}{M\hbar}\left(\frac{t^{1-3q}}{3} - M^2t^{1-q} \right) \tag{4.3}$$

(neglecting terms containing higher powers of t^{-1}).

Employing the same method as above we find that the expression for the energy of a primordial tachyon in case of flat and closed models is also given by (4.3) as in the case of the open model.

When models contain bradyon dust, $q = \frac{2}{3}$. Hence

$$E = \frac{2vM}{3\hbar}t^{-2/3} + \frac{v}{3M\hbar}t^{-1} - \frac{vM}{\hbar}t^{1/3}. \tag{4.4}$$

Now we have

$$E = \begin{cases} \dfrac{2v}{3\hbar}t^{-2/3} + \dfrac{v}{3M\hbar}t^{-1} & \text{when } t \text{ is small,} \\ -\dfrac{vM}{\hbar}t^{1/3} & \text{when } t \text{ is large.} \end{cases} \tag{4.5}$$

The rate of emission of energy of the primordial tachyon is given by

$$-\frac{dE}{dt} = \begin{cases} \dfrac{4vM}{9\hbar}t^{-5/3} + \dfrac{v}{3M\hbar}t^{-2} & \text{when } t \text{ is small,} \\ \dfrac{vM}{3\hbar}t^{-2/3} & \text{when } t \text{ is large.} \end{cases} \tag{4.6}$$

When the model contains radiation $q = \frac{1}{2}$. Hence

$$E = \frac{vM}{3\hbar}t^{-1/2} + \frac{v}{3M\hbar}t^{-1/2} - \frac{vM}{\hbar}t^{1/2}. \tag{4.7}$$

It yields

$$E = \begin{cases} \dfrac{v}{3M\hbar}(1 + M^2)t^{-1/2} & \text{when } t \text{ is small,} \\ -\dfrac{vM}{\hbar}t^{1/2} & \text{when } t \text{ is large} \end{cases} \tag{4.8}$$

and

$$-\frac{dE}{dt} = \begin{cases} \dfrac{v}{6M\hbar}(1 + M^2)t^{-3/2} & \text{when } t \text{ is small,} \\ \dfrac{vM}{2\hbar}t^{-1/2} & \text{when } t \text{ is large.} \end{cases} \tag{4.9}$$

When models contain superdense matter $q = \frac{1}{3}$. Hence

$$E = \frac{vM}{6\hbar}t^{-1/3} + \frac{v}{\hbar M}(1 - M^2t^{2/3}), \tag{4.10}$$

which yields

$$E=\begin{cases}\dfrac{vM}{6\hbar}t^{-1/3}+\dfrac{v}{\hbar M} & \text{when } t \text{ is small,}\\ -\dfrac{vM}{\hbar}t^{2/3} & \text{when } t \text{ is large}\end{cases} \tag{4.11}$$

and

$$-\frac{dE}{dt}=\begin{cases}\dfrac{vM}{18E}t^{-4/3} & \text{when } t \text{ is small,}\\ \dfrac{2vM}{3\hbar}t^{-1/3} & \text{when } t \text{ is large.}\end{cases} \tag{4.12}$$

When models contain nonrelativistic matter $q=\frac{2}{3}$. Hence

$$E=\frac{2vM}{9\hbar}t^{-2/5}+\frac{v}{\hbar M}\left(\frac{t^{-1/5}}{3}-M^2t^{3/5}\right), \tag{4.13}$$

which yields

$$E=\begin{cases}\dfrac{2vM}{9\hbar}t^{-2/5}+\dfrac{v}{3\hbar M}t^{-1/5} & \text{when } t \text{ is small,}\\ -\dfrac{vM}{\hbar}t^{3/5} & \text{when } t \text{ is large.}\end{cases} \tag{4.14}$$

and

$$-\frac{dE}{dt}=\begin{cases}\dfrac{4vM}{45\hbar}t^{-7/5}+\dfrac{v}{15\hbar M}t^{-6/5} & \text{when } t \text{ is small,}\\ \dfrac{3vM}{5\hbar}t^{-2/5} & \text{when } t \text{ is large.}\end{cases} \tag{4.15}$$

Thus we find that dissipation of energy of a primordial tachyon is large in the beginning but this phenomena slows down later on, in every phase of the universe.

5. PRIMORDIAL TACHYON AT THE PRESENT EPOCH

It is noted that the expansion of the universe has been governed by its nonrelativistic matter content at least since the time when $S(t)$ was one-hundredth its present value.[10] It is also the view of cosmologists that the universe is 1.3×10^{10} years old. Hence, it is justified to take the expression for the energy of a primordial spinless tachyon as given by Eq. (4.14),

$$E=-(vM/\hbar)t^{3/5}. \tag{5.1}$$

Suppose T is the life of a primordial tachyon such that $T>t_0$ where t_0 is the present age of the universe. Now by the uncertainty principle we have

$$(vMt_0^{3/5}/\hbar)T\approx\hbar.$$

It yields

$$M\approx\hbar^2/vt_0^{3/5}T.$$

Hence

$$M\leqslant\hbar^2/vt_0^{8/5}. \tag{5.2}$$

In the case of tachyons we can write

$$v=3\times10^{10}a, \tag{5.3}$$

where $a>1$. Combining Eqs. (5.2) and (5.3),

$$M\leqslant\hbar^2/3\times10^{10}at_0^{8/5}. \tag{5.4}$$

Substituting the value of t_0 and $\hbar$ as

$$t_0=401\times10^{17}\ \text{s},$$

$$\hbar=1.055\times10^{-27}\ \text{erg s}$$

in (5.4) we have

$$M\leqslant2.447\times10^{-93}/a\ \text{g}. \tag{5.5}$$

Thus we find that if a spinless primordial tachyon survives up to the present epoch as well as into the future, its metamass would be less than 2.447×10^{-93} g. We also find that primordial tachyons moving with higher speeds, if they survive up to the present epoch and into the future, should be much lighter than those moving with low speeds. One thing is also interesting to note here—that the metamass of a spinless primordial tachyon surviving up to the present epoch would be very much less than that of spin-$\frac{1}{2}$ primordial tachyons, because it has been estimated earlier[6] that the metamass of a spin-$\frac{1}{2}$ primordial tachyon surviving up to the present epoch should be less than or almost equal to 8.77×10^{-54} g. Thus we find that the possibility of survival of spin-$\frac{1}{2}$ primordial tachyons is more than the spinless ones. However, our investigations agree with Narlikar and Sudarshan's result that if primordial tachyons survive up to the present epoch, their metamass should be very much less than the rest mass of an electron.[4]

[1] M. E. Arons and E. C. G. Sudarshan, Phys. Rev. **173**, 1622 (1968).
[2] G. Feinberg, Phys. Rev. **159**, 1089 (1967); Science **166**, 879 (1969).
[3] J. Dhar and E. C. G. Sudershan, Phys. Rev. **174**, 1808 (1968).
[4] J. V. Narlikar and E. C. G. Sudarshan, Mon. Not. R. Astron. Soc. **175**, 105 (1976).
[5] S. K. Srivastava, J. Math. Phys. **23**, 1981 (1982).
[6] S. K. Srivastava, "Friedmann universe and superluminal Dirac particles," J. Math. Phys. **24**, 996 (1983).
[7] L. Parker, "The production of Elementary particles by Strong Gravitational Fields," in *Asymptotic Structure of Space-time*, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977) and references therein.
[8] O. M. P. Bilaniuk, V. K. Deshpande, and E. C. G. Sudarshan, Am. J. Phys. **30**, 718 (1962).
[9] M. P. Pathak, Ph.D. thesis, Gorakhpur University, Gorakhpur, 1975.
[10] S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).

Generating differential operators for the basis functions in the variational LCAO-type methods

A. W. Niukkanen
V. I. Vernadsky Institute of Geochemistry and Analytical Chemistry of the Academy of Sciences of the USSR, Moscow, USSR

(Received 21 September 1982; accepted for publication 18 February 1983)

The analytical expressions for operators, which allow us to generate the basis functions from the simplest (ground) states by means of differentiating by parameters, are obtained. Some important properties of these operators are indicated. This approach allows us to simplify the structure of molecular matrix elements, whose calculation presents considerable difficulties in the variational LCAO-type methods.

PACS numbers: 31.20. – d, 03.65. – w

1. INTRODUCTION

The application of variational methods of quantum mechanics to many-particle systems with several force centers is related, as a rule, to the use of a multicenter basis set, i.e., instead of the set of functions $\Psi_{\gamma,j}(\mathbf{r})$ (γ, j are the continuous and discrete parameters, respectively), which are sufficient for the case of one force center; one should consider the states

$$\Psi^{\mathbf{R}}_{\gamma,j}(\mathbf{r}) = \Psi_{\gamma,j}(\mathbf{r}-\mathbf{R}) = \mathscr{T}(\mathbf{R})\Psi_{\gamma,j}(\mathbf{r}), \tag{1}$$

where $\mathscr{T}(\mathbf{R})$ is the translation operator

$$\mathscr{T}(\mathbf{R}) = \exp[-\mathbf{R}\cdot\nabla(\mathbf{r})] = \exp\left(-X\frac{\partial}{\partial x} - Y\frac{\partial}{\partial y} - Z\frac{\partial}{\partial z}\right), \tag{2}$$

which correspond to force centers localized at some fixed points $\mathbf{R}$. Atoms or cluster groups in molecules and crystals, nucleons or nucleon associations in nuclei, point defects in solids, etc., may serve the examples of such systems. We shall restrict ourselves to the case of atomic orbitals used in the MO LCAO SCF methods,[1] which contains all the typical problems inherent in the multicenter basis set problem.

As an example of typical quantities, which have to be dealt with in the LCAO scheme, one may consider matrix elements

$$\mathscr{O}^{\omega\gamma}_{ij}(\mathbf{R}) = \langle \Psi_{\omega,i}|\mathscr{O}|\Psi^{\mathbf{R}}_{\gamma,j}\rangle = \int d\mathbf{r}\, \Psi_{\omega,i}(\mathbf{r})\mathscr{O}(\mathbf{r})\Psi_{\gamma,j}(\mathbf{r}-\mathbf{R}) \tag{3}$$

related to the mean values of some operator $\mathscr{O}$ as well as the coefficients of an expansion

$$\Psi_{\gamma,j}(\mathbf{r}-\mathbf{R}) = \sum_i C^{\omega\gamma}_{ij}(\mathbf{R})\Psi_{\omega,i}(\mathbf{r}) \tag{4}$$

that allows us to express multicenter electron repulsion integrals[1] as expansions over simpler integrals with fewer centers. Obviously, the coefficients $C(\mathbf{R})$, in the case of orthogonal basis functions Ψ are directly related to integrals (3) for the case of the unit operator $\mathscr{O} = I$.

As a rule, the state $\Psi_{\gamma,0}$ corresponding to the lowest possible quantum numbers has a simple structure, which considerably facilitates the calculation of corresponding matrix elements. If there exists such a differential operator $\mathscr{D}_j(\gamma,\mathbf{R})$ that

$$\Psi_{\gamma,j}(\mathbf{r}-\mathbf{R}) = \mathscr{D}_j(\gamma,\mathbf{R})\Psi_{\gamma,0}(\mathbf{r}-\mathbf{R}), \tag{5}$$

there arises as well an obvious possibility of reducing the corresponding matrix elements to simpler form. For example, in the case of the matrix element I_{ij}, using Eq. (5) consecutively—first, directly for the function $\Psi_{\gamma,j}(\mathbf{r}-\mathbf{R})$ in Eq. (3) and, then, after changing $\mathbf{r}\to\mathbf{r}+\mathbf{R}$, for the function $\Psi_{\omega,i}(\mathbf{r}+\mathbf{R})$—we have the following expression:

$$I^{\omega\gamma}_{ij}(\mathbf{R}) = \mathscr{D}_j(\gamma,\mathbf{R})\mathscr{D}_i(\omega,-\mathbf{R})I^{\omega\gamma}_{00}(\mathbf{R}). \tag{6}$$

Since for the quantity $I^{\omega\gamma}_{00}(\mathbf{R})$ (not only in this illustrative example, but also in other, more complicated cases) one succeeds frequently in obtaining relatively simple expressions, Eq. (6) allows us, in principle, to get the "explicit" expressions, i.e., those not containing quadratures, and leads to the advised breaking up of the initial complicated problem into a series of more simple steps. Since the operators $\mathscr{D}$ and matrix elements I_{00} are found to be, as a rule, associated with some system of special functions, the practical value of such an approach is stipulated not only (and not predominantly) by the possibility of obtaining the "explicit expressions," but also by the possibility of using the analytical properties of functions $\mathscr{D}$ and I_{00} for establishing the analytical properties of matrix elements, for example, recurrence relations, differential and asymptotic properties, and so on.

Thus, establishing the analytical form of the generating operators $\mathscr{D}$ in Eq. (5), which relate the basis function $\Psi_{\gamma,j}$ to the "ground" state $\Psi_{\gamma,0}$ by means of differentiating by parameters, may have significance for the development of more effective methods of calculating matrix elements in the LCAO scheme. In the following we shall consider all the main types of basis functions used in quantum chemical applications (the Slater, hydrogenlike, Gaussian, and so-called reduced Bessel functions[2,3]). Two methods of constructing the generating differential operators (GDO) will be used. The first approach is based on the Fourier-transformation method and results in GDO's containing the operators $\nabla(\mathbf{R})$ only. The second method utilizes commutation properties of operators and some integral transformations and leads to GDO's which are simpler in a form, their arguments, however, containing not only the operators $\nabla(\mathbf{R})$, but differentiations by the scaling parameter ω as well.

 0022-2488/84/030698-08$02.50

2. THE FOURIER-TRANSFORMATION METHOD

Consider the general Fourier-transformation formulas

$$\overline{F}(\mathbf{k}) = (2\pi)^{-3/2} \int d\mathbf{r} \exp(i\mathbf{k}\cdot\mathbf{r}) F(\mathbf{r}) \equiv \langle \mathscr{F} F | \mathbf{k} \rangle , \quad (7)$$

$$F(\mathbf{r}) = (2\pi)^{-3/2} \int d\,\mathbf{k} \exp(-i\mathbf{k}\cdot\mathbf{r}) \overline{F}(\mathbf{k}) \equiv \langle \mathscr{F} \overline{F} | -\mathbf{r} \rangle . \quad (8)$$

Let

$$\overline{F}(\mathbf{k}) = \mathscr{D}(\mathbf{k}) \overline{\Phi}(\mathbf{k}) , \quad (9)$$

where $\overline{\Phi}(\mathbf{k})$ is the Fourier transform of some known function Φ. Using the identity

$$\mathbf{k} \exp(-i\mathbf{k}\cdot\mathbf{r}) = i\nabla(\mathbf{r}) \exp(-i\mathbf{k}\cdot\mathbf{r}) , \quad (10)$$

we have

$$\mathscr{D}(\mathbf{k}) \exp(-i\mathbf{k}\cdot\mathbf{r}) = \mathscr{D}(i\nabla(\mathbf{r})) \exp(-i\mathbf{k}\cdot\mathbf{r}) . \quad (11)$$

Thus, the substitution of Eq. (9) into Eq. (8) yields the following differential representation:

$$F(\mathbf{r}) = \mathscr{D}(i\nabla(\mathbf{r})) \Phi(\mathbf{r}) , \quad (12)$$

wherefrom

$$F(\mathbf{r} - \mathbf{R}) = \mathscr{D}(-i\nabla(\mathbf{R})) \Phi(\mathbf{r} - \mathbf{R}) . \quad (13)$$

Obviously, Eq. (13) gives a certain form of the GDO for the function F. The practical value of the representation (13) depends on how simple the functions $\mathscr{D}$ and Φ are. For example, if $\mathscr{D}(\mathbf{k})$ is a polynomial in Cartesian coordinates of the vector $\mathbf{k}$, then this eliminates the problem of analyzing the convergence of a series that would take place otherwise. Moreover, the function Φ should have a simpler form than the function F; otherwise we would have met with complication rather than simplification of the initial problem.

As the first step, we shall consider the Shavitt, Filter, and Steinborn functions (SFSF)[2,3]:

$$\Xi_{\omega nlm}(\mathbf{r}) = \xi_{\omega nl}(r) Y_{lm}(\mathbf{r}) = (\omega r)^l K^{+}_{n+1/2}(\omega r) Y_{lm}(\mathbf{r}) , \quad (14)$$

where $K^{+}_{\nu}(x) = x^{\nu} K_{\nu}(x)$, $K_{\nu}(x)$ is the Macdonald function, and $Y_{lm}(\mathbf{r})$ is the spherical function. One may easily show[4] that the Fourier transform $\overline{\Xi}$ is

$$\overline{\Xi}_{\omega nlm}(\mathbf{k}) = \overline{\xi}_{\omega nl}(k) Y_{lm}(\mathbf{k}) = a(\omega,n,l)[k^l/(k^2+\omega^2)^{n+l+2}] Y_{lm}(\mathbf{k}) , \quad (15)$$

where

$$a(\omega,n,l) = 2^{l+n+1}(l+n+1)!\,\omega^{2n+l+1} i^l . \quad (16)$$

Taking into account the corollary that follows from Eq. (15) in the case $l = m = 0$ ($Y_{00} \equiv 1/\sqrt{4\pi}$), we obtain

$$\overline{\Xi}_{\omega nlm}(\mathbf{k}) = \sqrt{4\pi}(i/\omega)^l\, \overline{\Xi}_{\omega,n+l,00}(\mathbf{k}) \mathscr{Y}_{lm}(\mathbf{k}) , \quad (17)$$

where $\mathscr{Y}_{lm}(\mathbf{k}) = k^l Y_{lm}(\mathbf{k})$ is the homogeneous harmonic polynomial in k_x, k_y, k_z. Writing down Eq. (17) in the form of Eq. (9) with $\mathscr{D}(\mathbf{k}) = \mathscr{Y}(\mathbf{k})$, taking into account Eq. (13) and noting that $\mathscr{Y}_{lm}(c\mathbf{k}) = c^l \mathscr{Y}_{lm}(\mathbf{k})$, we obtain

$$\Xi_{\omega nlm}(\mathbf{r} - \mathbf{R}) = \sqrt{4\pi}\,\omega^{-l} \mathscr{Y}_{lm}(\nabla(\mathbf{R})) \Xi_{\omega,n+l,00}(\mathbf{r} - \mathbf{R}) . \quad (18)$$

Thus, in the case of SFSF the simplest form of GDO, $\mathscr{Y}(\nabla)$, allows us to relate the irreducible spherical tensor of rank l on the left-hand side of Eq. (18) with the spherical scalar, i.e., with the function with $l = 0$. In its turn, using the standard differential properties of the function[5] $K_\nu(z)$ associated with the shift operator

$$\mathscr{D}(\omega) = \frac{1}{\omega}\frac{\partial}{\partial\omega} = \frac{\partial}{\partial(\omega^2/2)} , \quad (19)$$

the function $\Xi_{\omega N00}$ in Eq. (18) may be related to the simplest functions of this type with $N = 0$ or $N = -1$. The case of $N = -1$ is more preferable, since, by virtue of the relation

$$\Xi_{\omega,-1,00}(\mathbf{r} - \mathbf{R}) = (1/2^{3/2})\omega^{-1} G_\omega(\mathbf{r} - \mathbf{R}) , \quad (20)$$

which involves the Green function

$$G_\omega(\mathbf{r} - \mathbf{R}) = \exp(-\omega|\mathbf{r} - \mathbf{R}|)/|\mathbf{r} - \mathbf{R}| \quad (21)$$

for the Helmholtz equation, it becomes possible to use the unique property

$$\Delta(\mathbf{R}) G_\omega(\mathbf{r} - \mathbf{R}) = \omega^2 G_\omega(\mathbf{r} - \mathbf{R}) - 4\pi\delta(\mathbf{r} - \mathbf{R}) , \quad (22)$$

which is not inherent in the other functions of this type. The necessary reduction is given by the formula[6]

$$\Xi_{\omega N00}(\mathbf{r}) = (-1)^{N+1}\omega^{2N+1}[\mathscr{D}(\omega)]^{N+1}[\omega\Xi_{\omega,-1,00}(\mathbf{r})] \quad (23)$$

and, hence,

$$\Xi_{\omega nlm}(\mathbf{r} - \mathbf{R}) = \sqrt{\pi/2}(-1)^{n+l+1}\omega^{2n+l+1}\mathscr{Y}_{lm} \times (\nabla(\mathbf{R}))[\mathscr{D}(\omega)]^{n+l+1} G_\omega(\mathbf{r} - \mathbf{R}) . \quad (24)$$

Thus, integrals with the SFSF Ξ may be obtained by means of GDO (24) from the simplest matrix elements with the functions G_ω. In its turn, Eq. (22) may be used for evaluating and analyzing the properties of these simplest elements. For example, in the case of the overlap integral,

$$I^{\gamma\omega}(\mathbf{R}) = \int d\mathbf{r}\, G_\gamma(\mathbf{r}) G_\omega(\mathbf{r} - \mathbf{R}) \quad (25)$$

by virtue of Eq. (21), we have

$$\Delta(\mathbf{R}) I^{\gamma\omega}(\mathbf{R}) = \omega^2 I^{\gamma\omega}(\mathbf{R}) - 4\pi G_\gamma(\mathbf{R}) . \quad (26)$$

On the other hand, after replacing $\mathbf{r} \to \mathbf{r} + \mathbf{R}$, we get

$$\Delta(\mathbf{R}) I^{\gamma\omega}(\mathbf{R}) = \gamma^2 I^{\gamma\omega}(\mathbf{R}) - 4\pi G_\omega(\mathbf{R}) \quad (27)$$

since $G_\omega(-\mathbf{R}) = G_\omega(\mathbf{R})$. It follows from Eqs. (26) and (27) that

$$I^{\gamma\omega}(\mathbf{R}) = 4\pi \frac{G_\omega(\mathbf{R}) - G_\gamma(\mathbf{R})}{\gamma^2 - \omega^2} = \frac{4\pi}{(\gamma^2 - \omega^2)} \frac{1}{R} [\exp(-\omega R) - \exp(-\gamma R)] . \quad (28)$$

If $\omega = 0$, Eq. (28) yields the expression for the simplest integral of attraction to the nucleus, into which the overlap integral is transformed in this case. Thus, in special cases the Helmholtz equation (22) allows the immediate evaluation of the matrix elements. In the case of more complicated integrand functions, Eq. (22) can be used for deriving the recurrence relations for molecular integrals, which should apparently result in the particular case of the Coulomb integrals with the Slater functions, in a more clear and compact alternative to the Harris method.[7] Note also that the relation between the matrix elements of exponential class and the

Helmholtz equation generalizes and supplements the relation of the Coulomb integrals with the Laplace equation[8] [into which Eq. (22) is transformed in the case of $\omega = 0$] and, apparently, may offer some new possibilities for the evaluation of matrix elements following the general lines of the earlier approaches.[7–10]

Consider the other basis functions of exponential class. In the case of hydrogenlike functions (HLF),

$$H_{\omega nlm}(\mathbf{r}) = h_{\omega nl}(r) Y_{lm}(\mathbf{r}) = (\omega r)^l L_n^{2l+1}(2\omega r)\exp(-\omega r) Y_{lm}(\mathbf{r}), \tag{29}$$

the Fourier transform $\overline{H}$ can be written as follows[4]:

$$\overline{H}_{\omega nlm}(\mathbf{k}) = \overline{h}_{\omega nl}(k) Y_{lm}(\mathbf{k}) = b(\omega,n,l) \times F(-n, -n-l-\tfrac{1}{2}; l+\tfrac{3}{2}; -k^2/\omega^2) \times \mathscr{Y}_{lm}(\mathbf{k})/(k^2+\omega^2)^{n+l+2}, \tag{30}$$

$$b(\omega,n,l) = (-1)^n (n+l+1)(1/\sqrt{\pi}) 2^{l+3/2} l! \omega^{l+1} \times [(2l+2)_n/n!] i^l, \tag{31}$$

where $F = {}_2F_1$ is the hypergeometric Gauss function. It follows from comparison of Eq. (30) with Eq. (15) that the function $\overline{H}$ is proportional to the product of the Fourier transform $\overline{\Xi}$ by the polynomial $F(-k^2/\omega^2)$.

Therefore, using Eqs. (7)–(13), we have

$$H_{\omega nlm}(\mathbf{r}-\mathbf{R}) = \frac{b(\omega,n,l)}{a(\omega,n,l)} F\left(-n, -n-l-\frac{1}{2}; l+\frac{3}{2}; -\frac{\Delta(\mathbf{R})}{\omega^2}\right) \times \Xi_{\omega nlm}(\mathbf{r}-\mathbf{R}) \tag{32}$$

$$= \sqrt{4\pi}\, \frac{b(\omega,n,l)}{a(\omega,n,l)} F\left(-n, -n-l-\frac{1}{2}; l+\frac{3}{2}; -\frac{\Delta(\mathbf{R})}{\omega^2}\right) \times \mathscr{Y}_{lm}(-i\nabla(\mathbf{R})) \Xi_{\omega,n+l,00}(\mathbf{r}-\mathbf{R}). \tag{33}$$

Equations (32) and (33) yield the expressions for the GDO's which relate HLF H to SFSF Ξ. With due respect to relation (23), we obtain, just as in the case of functions Ξ, the reduction of the functions H to the functions $G_\omega(\mathbf{r}-\mathbf{R})$. Note that polynomials $F(-\Delta(\mathbf{R})/\omega^2)\mathscr{Y}_{lm}(-i\nabla(\mathbf{R}))$, which enter into Eq. (33), by virtue of connection of the denominator parameter c in $F(a,b;c;z)$ with the rank l ($c = l + \frac{3}{2}$), possess some special properties,[11] which facilitate the handling of these functions.

In the case of the Slater functions,

$$\Psi_{\omega nlm}(\mathbf{r}) = \psi_{\omega nl}(r) Y_{lm}(\mathbf{r}) = (\omega r)^l (\omega r)^n \exp(-\omega r) Y_{lm}(\mathbf{r}) \tag{34}$$

we have[4]

$$\overline{\Psi}_{\omega nlm}(\mathbf{k}) = c(\omega,n,l) F\left(-\frac{n}{2}, -\frac{n+1}{2}; l+\frac{3}{2}; -\frac{k^2}{\omega^2}\right) \times \frac{\mathscr{Y}_{lm}(\mathbf{k})}{(k^2+\omega^2)^{n+l+2}}, \tag{35}$$

$$c(\omega,n,l) = \frac{1}{\sqrt{\pi}} \frac{2^{l+1/2} l!}{(2l+1)!} (n+2l+2)! \omega^{2n+l+1} i^l. \tag{36}$$

Since the general structure of the expression (35) is quite similar to Eq. (30), we obtain

$$\Psi_{\omega nlm}(\mathbf{r}-\mathbf{R}) = \frac{c(\omega,n,l)}{a(\omega,n,l)} F\left(-\frac{n}{2}, -\frac{n+1}{2}; l+\frac{3}{2}; -\frac{\Delta(\mathbf{R})}{\omega^2}\right) \times \Xi_{\omega nlm}(\mathbf{r}-\mathbf{R}) \tag{37}$$

$$= \sqrt{4\pi}\, \frac{c(\omega,n,l)}{a(\omega,n,l)} \times F\left(-\frac{n}{2}, -\frac{n+1}{2}; l+\frac{3}{2}; -\frac{\Delta(\mathbf{R})}{\omega^2}\right) \times \mathscr{Y}_{lm}(-i\nabla(\mathbf{R})) \Xi_{\omega,n+l,00}(\mathbf{r}-\mathbf{R}). \tag{38}$$

All the above remarks, including the special property of the polynomial $F\mathscr{Y}$ on the right-hand side of Eq. (38), are also valid for the GDO's (37) and (38).

Thus, the Fourier-transformation method allows us to obtain, in a relatively regular and simple way, the expressions for GDO's relating any functions of exponential class to the simplest functions Ξ or G_ω belonging to the same class. All these GDO's have a form of irreducible tensor polynomials in Cartesian components of the gradient operator $\nabla(\mathbf{R})$, the radial parts of these tensors being expressed in a class of standard special functions ${}_2F_1$, which satisfy the quadratic transformations.[12]

3. THE COMMUTATION RELATIONS METHOD: INTEGRAL REPRESENTATIONS OF GDO's

Obviously, the Fourier transformation method, combined with the differential properties of the basis functions with respect to the parameter ω, leads to the special GDO's which have a factorized form relative to the operators $\nabla(\mathbf{R})$ and $\mathscr{D}(\omega)$. In this sense, the Fourier-transformation method is not flexible enough and does not provide a possibility of constructing the mixed, nonfactorized operators. We shall consider here an alternative method which is based on the commutation properties of operators and leads to GDO's of simpler form.

Consider the function

$$e = e(\omega;x,y,z) = \exp(-\omega r), \tag{39}$$

$$r = r(x,y,z) = (x^2+y^2+z^2)^{1/2}. \tag{40}$$

Since

$$\frac{\partial e}{\partial x} = -\omega e \frac{\partial r}{\partial x}, \quad \frac{\partial r}{\partial x} = \frac{x}{r}, \tag{41}$$

then

$$xe = -\frac{1}{\omega} r \frac{\partial}{\partial x} e. \tag{42}$$

Using on the right-hand side of Eq. (42) the commutation relation

$$\left[r, \frac{\partial}{\partial x}\right] \equiv r\frac{\partial}{\partial x} - \frac{\partial}{\partial x} r = -\frac{x}{r}, \tag{43}$$

we have

$$xe = -\frac{1}{\omega}\frac{\partial}{\partial x} re + \frac{1}{\omega}\frac{x}{r} e. \tag{44}$$

Writing down the expression re on the right-hand side of Eq. (44) as

$$re = -\frac{\partial}{\partial\omega} e \tag{45}$$

and transforming the function xe/r by means of Eq. (42), we obtain

$$xe = T(\omega)\, d(x)\, e\,, \tag{46}$$

where $d(x) \equiv \partial/\partial x$ and

$$T(\omega) = \frac{1}{\omega}\frac{\partial}{\partial\omega} - \frac{1}{\omega^2} \equiv \mathscr{D}(\omega) - \frac{1}{\omega^2}\,. \tag{47}$$

Obviously,

$$T(\omega) = \frac{\partial}{\partial\omega}\frac{1}{\omega}\,. \tag{48}$$

Using Eq. (46), one may prove by induction that

$$x^n e = [T(\omega)]^{n/2} H_n(d(x)\sqrt{T(\omega)})e\,, \tag{49}$$

where $H_n(z)$ is the Hermite polynomial.[13] Indeed, since $H_0(z) = 1$ and $H_1(z) = z$ then Eq. (49) is valid for $n = 0$, $n = 1$ [see Eq. (46)]. Multiplying Eq. (49) by the variable x from the left, taking into account the formal relation

$$[x, f(d)] = -f'(d)\,, \tag{50}$$

where $d \equiv d(x)$, using the identity[13]

$$H_n'(z) = nH_{n-1}(z) \tag{51}$$

and transforming the term xe arising after commuting x with H_n we obtain, by means of Eq. (46),

$$x^{n+1}e = [T(\omega)]^{(n+1)/2}\big[d(x)\sqrt{T(\omega)}H_n(d(x)\sqrt{T(\omega)}) - nH_{n-1}(d(x)\sqrt{T(\omega)})\big]e\,. \tag{52}$$

By virtue of the recurrence relation for the Hermite polynomials,[13]

$$H_{n+1}(z) - zH_n(z) + nH_{n-1}(z) = 0\,, \tag{53}$$

we obtain

$$x^{n+1}e = [T(\omega)]^{(n+1)/2}H_{n+1}(d(x)\sqrt{T(\omega)})e\,, \tag{54}$$

which proves, by induction, the validity of Eq. (49). Since[13]

$$H_n(z) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} dz'\,(z + iz')^n \exp(-\tfrac{1}{2}z'^2)\,, \tag{55}$$

then it follows from Eq. (49) that

$$x^n e = (2\pi)^{-1/2}\int_{-\infty}^{\infty} dx'\,\big[d(x)T(\omega) + i\sqrt{T(\omega)}x'\big]^n \times \exp(-\tfrac{1}{2}x'^2)\,e\,. \tag{56}$$

Therefore, for the function $P(\mathbf{r}) = P(x,y,z)$, which can be represented by the series in powers of x,y,z, we have

$$P(\mathbf{r})e = (2\pi)^{-3/2}\int d\mathbf{r}'\,P(T(\omega)\nabla(\mathbf{r}) + i\sqrt{T(\omega)}\mathbf{r}') \times \exp(-\tfrac{1}{2}r'^2)\,e\,. \tag{57}$$

By formal change of the integration variable, Eq. (57) can be reduced to the form

$$P(\mathbf{r})e = (2\pi)^{-3/2}\int d\boldsymbol{\rho}\,\exp\{-\tfrac{1}{2}[\boldsymbol{\rho} + i\sqrt{T(\omega)}\nabla(\mathbf{r})]^2\} \times P(i\sqrt{T(\omega)}\,\boldsymbol{\rho})\,e\,. \tag{58}$$

4. THE ALGEBRAIC EXPRESSION FOR GDO

Any of two integral representations, Eq. (57) or Eq. (58), can be used for the derivation of the algebraic expression for the GDO which relates the function $P(\mathbf{r})e$ to e. For this purpose in the case of Eq. (57) one should use the addition theorem for the integrand function P and, in the case of Eq. (58), the addition theorem for the Gaussian function. The latter case leads to a more general relation, since there is no need for the explicit form of an operator at the initial step.

We present here the derivation of the addition theorem for the Gaussian function, which differs from that given in Ref. 14. Our proof is more strictly related to the one-dimensional case and allows an immediate generalization to the multidimensional case as well. Besides, we indicate here a more general addition theorem, which is absent in Ref. 14 and which allows us to obtain the "exponential" addition theorem as a natural particular case.

Writing down the Gaussian function in the form

$$\exp[-\tfrac{1}{2}(\boldsymbol{\rho} - \mathbf{R})^2] = \exp(R\rho c - \tfrac{1}{2}R^2)\exp(-\tfrac{1}{2}\rho^2)\,, \tag{59}$$

where $c \equiv \cos(\boldsymbol{\rho},\mathbf{R})$, and, taking into account that the first multiplier on the right-hand side of Eq. (59) as a form of the generating function for the Hermite polynomials H_k,[13] we have

$$\exp[-\tfrac{1}{2}(\boldsymbol{\rho} - \mathbf{R})^2] = \sum_{k=0}^{\infty}\frac{1}{k!}R^k H_k(\rho c)\exp(-\tfrac{1}{2}\rho^2)\,. \tag{60}$$

Expanding the quantity $H_k(\rho c)$ in a series over the Legendre polynomials $P_l(c)$,

$$H_k(\rho c) = \sum_l F_{kl}(\rho)P_l(c)\,, \tag{61}$$

and using the standard orthogonality relation for functions $P_l(c)$, we obtain

$$F_{kl}(\rho) = \Big(l + \frac{1}{2}\Big)\int_{-1}^{1} dc\,H_k(\rho c)P_l(c)\,. \tag{62}$$

Using the Rodrigues formula for the function $P_l(c)$, integrating by parts l times, and taking into consideration that

$$\frac{d^l}{dc^l}H_k(\rho c) = \frac{k!}{(k-l)!}\rho^l H_{k-l}(\rho c)\,, \tag{63}$$

we have

$$F_{kl}(\rho) = 2^{-l}\Big(l + \frac{1}{2}\Big)\rho^l\frac{k!}{l!(k-l)!} \times \int_{-1}^{1} dc\,H_{k-l}(\rho c)(1-c^2)^l\,. \tag{64}$$

Any classical polynomial is orthogonal, with the proper definition of a scalar product, to any polynomial of a lesser order. Thus, the polynomial P_l in Eq. (62) is orthogonal to $H_k(\rho c)$, if $k < l$. Thus, it follows that $F_{kl} \neq 0$, if $l \leqslant k$. Besides, since $H_q(-x) = (-1)^q H_q(x)$, it results from Eq. (64) that the quantity $k - l$ assumes only even values, i.e., $k - l = 2n$, where $n \geqslant 0$. Under these conditions, the Uspensky formula[15] is valid for the integral in Eq. (64). As a result, we have

$$F_{kl}(\rho) = \sqrt{\pi}\Big(l+\frac{1}{2}\Big)(-2)^{-(k-l)/2}\frac{k!}{\Gamma((k+l+3)/2)} \times\Big(\frac{\rho}{2}\Big)^l L^{l+1/2}_{(k-l)/2}\Big(\frac{\rho^2}{2}\Big). \quad (65)$$

Substituting Eq. (65) into Eq. (61) and introducing the new variable, $n=(k-l)/2$, we obtain

$$\exp[-\tfrac{1}{2}(\boldsymbol{\rho}-\mathbf{R})^2] = \sum_{n,l}(-1)^n\frac{\sqrt{\pi}(2l+1)}{2^{n+l+1}\Gamma(n+l+\frac{3}{2})}\rho^l L^{l+1/2}_n \times\Big(\frac{\rho^2}{2}\Big)\exp(-\tfrac{1}{2}\rho^2)R^{l+2n}P_l(c). \quad (66)$$

Evidently, Eq. (66) is valid for vectors $\boldsymbol{\rho}$ and $\mathbf{R}$ of any dimension. The addition theorem

$$P_l[\cos(\boldsymbol{\rho},\mathbf{R})] = \sum_m\frac{4\pi}{2l+1}Y_{lm}(\boldsymbol{\rho})Y^*_{lm}(\mathbf{R}) \equiv\frac{4\pi}{\sqrt{2l+1}}(-1)^l\{Y_l(\boldsymbol{\rho})\otimes Y_l(\mathbf{R})\}_{00} \quad (67)$$

takes place in the three-dimensional case. Here the symbol $\{\ \otimes\ \}$ denotes the irreducible tensor product of two spherical tensors.[16] Introducing Eq. (67) into Eq. (66) and denoting

$$\mathscr{Y}^n_{lm}(\boldsymbol{\rho}) = \rho^{2n}\mathscr{Y}_{lm}(\boldsymbol{\rho}), \quad (68)$$

$$\mathscr{L}^n_{lm}(\boldsymbol{\rho}) = L^{l+1/2}_n(\rho^2)\mathscr{Y}_{lm}(\boldsymbol{\rho}), \quad (69)$$

$$\Lambda^n_{lm}(\boldsymbol{\rho}) = \mathscr{L}^n_{lm}(\boldsymbol{\rho})\exp(-\rho^2), \quad (70)$$

we obtain a completely factorized, in $\boldsymbol{\rho}$ and $\mathbf{R}$, expansion of the Gaussian function in the form

$$\exp[-\tfrac{1}{2}(\boldsymbol{\rho}-\mathbf{R})^2] = \sum_{n,l}\alpha(n,l)\{\mathscr{Y}^n_l(\mathbf{R}/\sqrt{2})\otimes\Lambda^n_l(\boldsymbol{\rho}/\sqrt{2})\}_{00}, \quad (71)$$

where

$$\alpha(n,l) = (-1)^{l+n}2\pi^{3/2}\sqrt{2l+1}/\Gamma(n+l+\tfrac{3}{2}). \quad (72)$$

The polynomials $\mathscr{L}$, appearing in the definition of the function Λ, are closely related to the eigenfunctions of the isotropic harmonic oscillator,

$$\Gamma_{nlm}(\boldsymbol{\rho}) = \mathscr{L}^n_{lm}(\boldsymbol{\rho})\exp(-\tfrac{1}{2}\rho^2) \equiv\Lambda^n_{lm}(\boldsymbol{\rho})\exp(\tfrac{1}{2}\rho^2). \quad (73)$$

Note that the expansion (71) is a particular case of a more general addition theorem for functions Λ:

$$\Lambda^N_{LM}\Big(\frac{\boldsymbol{\rho}-\mathbf{R}}{\sqrt{2}}\Big) = \sum_{n,l,\lambda}\alpha(n,l)\frac{1}{\sqrt{4\pi}}\langle l0L0|\lambda 0\rangle(-1)^{\omega+l}\frac{(N+\nu)!}{\nu!} \times\Big\{\mathscr{Y}^n_l\Big(\frac{\mathbf{R}}{\sqrt{2}}\Big)\otimes\Lambda^{N+\nu}_\lambda\Big(\frac{\boldsymbol{\rho}}{\sqrt{2}}\Big)\Big\}_{LM}, \quad (74)$$

where $\langle a\alpha b\beta|c\gamma\rangle$ is the Clebsch–Gordan coefficient[16] and

$$\nu = n+(l+L-\lambda)/2, \quad (75)$$

$$\omega = n+(l+\lambda-L)/2. \quad (76)$$

The numbers l,λ,L satisfy the triangle condition

$$|\lambda-l|\leqslant L\leqslant\lambda+l, \quad (77)$$

with the quantity $\lambda+l-L$ assuming only even values. Equation (74) can be derived with the help of a general method outlined in Ref. 11. If $N=L=M=0$, Eq. (74) transfers into Eq. (71), since the sum over λ is reduced to one term with $\lambda=l$, the equalities $\nu=n,\omega=n+l$, and $\langle l000|l0\rangle=1$ being satisfied in this case.

Since the basis functions in the LCAO method have a form of irreducible spherical tensors, the function $P(\mathbf{r})$ in Eqs. (57) and (58) is expressed as $P(\mathbf{r})=p(r)Y_{lm}(\mathbf{r})$, where $p(r)$ is some polynomial in r. Substituting this expression into Eq. (58) and using also the expansion (71), we obtain, after integration in spherical coordinates,

$$P(\mathbf{r})e = \mathscr{Y}_{lm}(-i\sqrt{T(\omega)}\boldsymbol{\nabla}(\mathbf{r}))\sum_n\frac{[T(\omega)\Delta(\mathbf{r})]^n}{2^{l+n+1/2}\Gamma(n+l+3/2)} \times\int_0^\infty d\rho\,\rho^{l+2}L^{l+1/2}_n(\tfrac{1}{2}\rho^2)\exp(-\tfrac{1}{2}\rho^2)p(\rho)\,e. \quad (78)$$

5. GENERATING OPERATORS FOR THE SLATER AND THE GAUSSIAN FUNCTIONS

Consider the Slater function in Eq. (34). If $n=2k$, where k is a nonnegative integer, then

$$\Psi_{\omega,2k,lm}(\mathbf{r}) = \mathscr{Y}^k_{lm}(\mathbf{r})\exp(-\omega r), \quad (79)$$

so that we have in Eqs. (57), (58), and (78)

$$P(\mathbf{r}) = \mathscr{Y}^k_{lm}(\mathbf{r}),\quad p(r)=r^{l+2k}. \quad (80)$$

Calculating the coefficients in the expansion (78) or using the addition theorem for polynomials[17,18] $\mathscr{Y}^k_{lm}(\mathbf{a}+\mathbf{b})$ directly in Eq. (57), we obtain

$$\mathscr{Y}^k_{lm}(\mathbf{r})\exp(-\omega r) = (-1)^k k!2^k[T(\omega)]^{l+k}L^{l+1/2}_k(\tfrac{1}{2}T(\omega)\Delta(\mathbf{r})) \times\mathscr{Y}_{lm}(\boldsymbol{\nabla}(\mathbf{r}))\exp(-\omega r), \quad (81)$$

which yields, with due respect to Eqs. (12) and (13), the necessary expression for the GDO which relates an arbitrary Slater function with the $1s$ state e. In the particular cases of $k=0$ and $l=m=0$ we have, respectively,

$$\mathscr{Y}_{lm}(\mathbf{r})\exp(-\omega r) = [T(\omega)]^l\mathscr{Y}_{lm}(\boldsymbol{\nabla}(\mathbf{r}))\exp(-\omega r), \quad (82)$$

$$r^{2k}\exp(-\omega r) = (-1)^k k!2^k[T(\omega)]^k \times L^{1/2}_k(\tfrac{1}{2}T(\omega)\Delta(\mathbf{r}))\exp(-\omega r). \quad (83)$$

In the case $n=2k-1$ in Eq. (34) we shall write down the function Ψ in the form

$$\Psi_{\omega,2k-1,lm}(\mathbf{r}) = \mathscr{Y}^k_{lm}(\mathbf{r})\exp(-\omega r)/r \equiv\mathscr{Y}^k_{lm}(\mathbf{r})G_\omega(\mathbf{r}). \quad (84)$$

By analogy with Eq. (46), we have

$$xG_\omega(\mathbf{r}) = \mathscr{D}(\omega)\,d(x)\,G_\omega(\mathbf{r}). \quad (85)$$

Applying exactly the same reasoning, as in Secs. 3 and 4, we obtain

$$\mathscr{Y}^k_{lm}(\mathbf{r})G_\omega(\mathbf{r}) = (-1)^k k!2^k[\mathscr{D}(\omega)]^{l+k}L^{l+1/2}_k(\tfrac{1}{2}\mathscr{D}(\omega)\Delta(\mathbf{r})) \times\mathscr{Y}_{lm}(\boldsymbol{\nabla}(\mathbf{r}))G_\omega(\mathbf{r}). \quad (86)$$

Thus, just as in Sec. 2, we get a possibility of generating the basis function from the Green function for the Helmholtz equation. The function G_ω can also be introduced onto the right-hand side of Eq. (81). Indeed, taking into account the identity

$$\exp(-\omega r) = -\frac{d}{d\omega} G_\omega(\mathbf{r}), \tag{87}$$

noting that

$$T^q(\omega)\, d(\omega) = d(\omega)\, \mathscr{D}^q(\omega) \tag{88}$$

and then using the equality

$$d(\omega) = \omega\, \mathscr{D}(\omega), \tag{89}$$

we have

$$\begin{aligned}\mathscr{Y}^k_{lm}(\mathbf{r})\exp(-\omega r) &= \omega(-1)^{k+1} k!2^k [\mathscr{D}(\omega)]^{l+k+1} L_k^{l+1/2} \\ &\quad \times (\tfrac{1}{2}\mathscr{D}(\omega)\Delta(\mathbf{r}))\mathscr{Y}_{lm}(\nabla(\mathbf{r}))G_\omega(\mathbf{r}).\end{aligned} \tag{90}$$

Exactly the same reasoning is applicable to the case of the Gaussian basis functions. Indeed, introducing the function g instead of the function e in Eq. (39)

$$g \equiv g(\omega;x,y,z) = \exp(-\tfrac{1}{2}\omega r^2), \tag{91}$$

we obtain, instead of Eq. (42),

$$xg = -\frac{1}{\omega}\frac{\partial}{\partial x} g, \tag{92}$$

and instead of Eq. (49) we have

$$x^n g = \omega^{-n/2} M_n(-d(x)\omega^{-1/2}) g, \tag{93}$$

where $M_n(x) = i^{-n} H_n(ix)$ is the modified Hermite polynomial. By means of the arguments which were used in the proof of the Eq. (57), we get

$$\begin{aligned}P(\mathbf{r})g &= (2\pi)^{-3/2} \int d\boldsymbol{\rho}\, P\left(-\frac{1}{\omega}\nabla(\mathbf{r}) + \frac{1}{\sqrt{\omega}}\boldsymbol{\rho}\right) \\ &\quad \times \exp(-\tfrac{1}{2}\rho^2)\, g.\end{aligned} \tag{94}$$

In its turn, this gives, by analogy with Eq. (81),

$$\begin{aligned}\mathscr{Y}^k_{lm}(\mathbf{r})\exp(-\omega r^2/2) &= (-1)^l k!2^k \omega^{-l-k} L_k^{l+1/2}(-\tfrac{1}{2}(1/\omega)\Delta(\mathbf{r})) \\ &\quad \times \mathscr{Y}_{lm}(\nabla(\mathbf{r}))\exp(-\omega r^2/2).\end{aligned} \tag{95}$$

Since in variational calculations only the Gaussian functions covered by Eq. (95) are exclusively used, there is no necessity in considering the case of the odd values of k.

The comparison of Eqs. (81) and (95) reveals the remarkable analogy between the Slater and the Gaussian GDO's. Indeed, in both these cases the general structure of GDO is described by the function $\mathscr{L}^k_{lm}(\boldsymbol{\rho})$, i.e., by the polynomial part of the harmonic oscillator eigenfunction (73).

6. SOME PROPERTIES OF POLYNOMIALS $\mathscr{L}$

We consider here some properties of the polynomials $\mathscr{L}$ which may occur to be useful in applying the GDO method for calculating matrix elements with the Slater and Gaussian functions.

First, the action of the operator $\mathscr{L}(\nabla)$ on some function $\varphi(r)$ of the scalar argument is described by the equation

$$\begin{aligned}L_k^{l+1/2}(t\Delta(\mathbf{r}))\mathscr{Y}_{lm}(\nabla(\mathbf{r}))\phi(r) &= \mathscr{Y}_{lm}(\mathbf{r}) L_k^{l+1/2}\left(\underset{2}{r^2}\frac{\mathscr{D}^2(r)}{1-2t\underset{1}{\mathscr{D}}(r)}\right) \\ &\quad \times (1-2t\mathscr{D}(r))^k \mathscr{D}^l(r)\varphi(r),\end{aligned} \tag{96}$$

where the operator r^2 in the argument of the polynomial L and the operators $\mathscr{D}(r)$ should be considered to be ordered in such a way, that the operators $\mathscr{D}(r)$ should be applied to the function $\varphi(r)$ prior to multiplication by powers of r^2. In accordance with the Feinmann and Maslov[19] notations, the operator ordering is indicated in Eq. (96) by indexes under operators. For example, we have

$$\underset{1}{\mathscr{D}}\underset{2}{r^2} = r^2\mathscr{D}, \quad (\underset{1}{\mathscr{D}}\underset{2}{r^2})^n = r^{2n}\mathscr{D}^n \tag{97}$$

and so on. We shall not present here the proof of the formula (96) confining ourselves to the observation that Eqs. (81) and (95) are the particular cases of the more general equation (96). To deduce these corollaries, one should bear in mind, both in the Slater and Gaussian cases, that the application of the operator $1-2t\mathscr{D}(r)$, where $t = \frac{1}{2}T(\omega)$ or $t = -\frac{1}{2}(1/\omega)$, respectively, to the function e or g gives a zero value. Therefore, in both cases it is sufficient to take into account only the term with the highest power in the Laguerre polynomial on the right-hand side of Eq. (96).

Second, consider the action of the operator $\mathscr{L}(\nabla)$ on the product of functions $f\varphi$. Generalizing the Leibnitz rule, we have the following formal relation:

$$\mathscr{L}(\nabla)f\varphi = \mathscr{L}(\nabla_f + \nabla_\varphi)f\varphi, \tag{98}$$

where the subscripts f and φ indicate the function to which the corresponding gradient operator should be applied. It follows from this relation that, to formulate the generalized Leibnitz rule in an explicit way, one should make use of some addition theorem which would allow us to represent the function $\mathscr{L}(\mathbf{a}+\mathbf{b})$ as a superposition of contributions factorized in $\mathbf{a}$ and $\mathbf{b}$. To this end, rewrite the expansion (71) in the form

$$\exp[\gamma(\mathbf{a},\mathbf{r})] = \sum_{n,l} \alpha(n,l)\{\mathscr{Y}^n_l(\mathbf{a}) \otimes \mathscr{L}^n_l(\mathbf{b})\}_{00}, \tag{99}$$

where $\gamma(\mathbf{a},\mathbf{r}) = -a^2 + 2\mathbf{ar}$. Note, that

$$\exp[\gamma(\mathbf{a},\mathbf{r}_1+\mathbf{r}_2)] = \exp[\gamma(\alpha_1\mathbf{a},\mathbf{r}_1/\alpha_1)]\exp[\gamma(\alpha_2\mathbf{a},\mathbf{r}_2/\alpha_2)], \tag{100}$$

where α_1 and α_2 are arbitrary numbers satisfying the condition $\alpha_1^2 + \alpha_2^2 = 1$. Using the expansion (99) for each of the three exponents in Eq. (100), by means of algebraic transformations, equivalent to those used in Ref. 18 for the derivation of the addition theorem for polynomials $\mathscr{Y}(\mathbf{r}_1+\mathbf{r}_2)$, we obtain

$$\begin{aligned}\mathscr{L}^n_{lm}(\mathbf{r}_1+\mathbf{r}_2) &= \sum_{n_1l_1n_2l_2} \frac{\alpha(n_1,l_1)\alpha(n_2,l_2)}{\alpha(n,l)}\frac{1}{\sqrt{4\pi}}\langle l_1 0 l_2 0 | l 0\rangle \\ &\quad \times \alpha_1^{2n_1+l_1}\alpha_2^{2n_2+l_2} \\ &\quad \times \left\{\mathscr{L}^{n_1}_{l_1}\left(\frac{\mathbf{r}_1}{\alpha_1}\right) \otimes \mathscr{L}^{n_2}_{l_2}\left(\frac{\mathbf{r}_2}{\alpha_2}\right)\right\}_{lm}.\end{aligned} \tag{101}$$

Obviously, the case of $\alpha_1 = \alpha_2 = \frac{1}{2}$ leads to a symmetric form of an expansion. In the case of $\alpha_1 = 0$ and $\alpha_2 = 1$, using the limiting transition in Eq. (101),

$$\lim_{\alpha_1 \to 0} \alpha_1^{2n_1 + l_1} \mathscr{L}^{n_1}_{l_1 m_1}\left(\frac{\mathbf{r}_1}{\alpha_1}\right) = \frac{(-1)^{n_1}}{n_1!} \mathscr{Y}^{n_1}_{l_1 m_1}(\mathbf{r}_1), \tag{102}$$

we have

$$\mathscr{L}^n_{lm}(\mathbf{r}_1 + \mathbf{r}_2) = \sum_{n_1 l_1 n_2 l_2} \frac{\alpha(n_1,l_1)\alpha(n_2,l_2)}{\alpha(n,l)} \frac{1}{\sqrt{4\pi}} \langle l_1 0 l_2 0 | l 0 \rangle \times \frac{(-1)^{n_1}}{n_1!} \left\{ \mathscr{Y}^{n_1}_{l_1}(\mathbf{r}_1) \otimes \mathscr{L}^{n_2}_{l_2}(\mathbf{r}_2) \right\}_{lm}. \tag{103}$$

Summations in n_1, l_1, n_2, l_2 in Eqs. (102) and (103) are restricted by the triangle condition, $\mathbf{l}_1 + \mathbf{l}_2 = \mathbf{l}$ ($l_1 + l_2 - l$ is even), as well as by the relation

$$2n_1 + l_1 + 2n_2 + l_2 = 2n + l. \tag{104}$$

Finally, we establish here another important property of the functions $\mathscr{L}(\mathbf{r})$, viz., an addition theorem of the Clebsch–Gordan type. For this purpose, it is expedient to use the integral representation

$$\mathscr{L}^n_{lm}((i/\sqrt{2})t\mathbf{r}) = (i^l / n! 2^{n + l/2})(2\pi)^{-3/2} \times \int d\boldsymbol{\rho} \mathscr{Y}^n_{lm}(\boldsymbol{\rho}) \exp[-\tfrac{1}{2}(\boldsymbol{\rho} - t\mathbf{r})^2], \tag{105}$$

which follows, for example, by comparison of Eq. (58) with Eq. (81) for the case $P = \mathscr{Y}$. The multiplier $i/\sqrt{2}$ on the left-hand side of Eq. (105) is introduced to simplify the structure of the right-hand side of the equation. Introducing the irreducible tensor product of functions $\mathscr{L}$ and using, on one hand, the addition theorem for spherical functions[16] appearing in the definition (69),

$$\{ Y_{l_1}(\mathbf{r}) \otimes Y_{l_2}(\mathbf{r}) \}_{lm} = H(l_1,l_2,l) Y_{lm}(\mathbf{r}) = (1/\sqrt{4\pi}) \sqrt{(2l_1+1)(2l_2+1)/(2l+1)} \times \langle l_1 0 l_2 0 | l 0 \rangle Y_{lm}(\mathbf{r}), \tag{106}$$

we have

$$\left\{ \mathscr{L}^{n_1}_{l_1}((i/\sqrt{2})t_1\mathbf{r}) \otimes \mathscr{L}^{n_2}_{l_2}((i/\sqrt{2})t_2\mathbf{r}) \right\}_{lm} = (i/\sqrt{2})^{l_1 + l_2} t_1^{l_1} t_2^{l_2} H(l_1,l_2,l)(r^2)^{(l_1 + l_2 - l)/2} \times L^{l_1 + 1/2}_{n_1}(-\tfrac{1}{2}t_1^2 r^2) L^{l_2 + 1/2}_{n_2}(-\tfrac{1}{2}t_2^2 r^2) \mathscr{Y}_{lm}(\mathbf{r}). \tag{107}$$

On the other hand, using for each multiplier on the left-hand side of Eq. (107) the integral representation (105), we obtain

$$\left\{ \mathscr{L}^{n_1}_{l_1}\left(\frac{i}{\sqrt{2}} t_1 \mathbf{r}\right) \otimes \mathscr{L}^{n_2}_{l_2}\left(\frac{i}{\sqrt{2}} t_2 \mathbf{r}\right) \right\}_{lm} = \frac{i^{l_1 + l_2}}{n_1! n_2! 2^{n_1 + n_2 + (l_1 + l_2)/2}} (2\pi)^{-3} \times \iint d\boldsymbol{\rho}_1 \, d\boldsymbol{\rho}_2 \left\{ \mathscr{Y}^{n_1}_{l_1}(\boldsymbol{\rho}_1) \otimes \mathscr{Y}^{n_2}_{l_2}(\boldsymbol{\rho}_2) \right\}_{lm} \times \exp[-\tfrac{1}{2}(\boldsymbol{\rho} - t_1\mathbf{r})^2 - \tfrac{1}{2}(\boldsymbol{\rho} - t_2\mathbf{r})^2]. \tag{108}$$

Denoting

$$t = \sqrt{t_1^2 + t_2^2}, \quad \tau_1 = t_1/t, \quad \tau_2 = t_2/t \tag{109}$$

and introducing new integration variables

$$\mathbf{r}_1 = \tau_1 \boldsymbol{\rho}_1 + \tau_2 \boldsymbol{\rho}_2, \quad \mathbf{r}_2 = \tau_2 \boldsymbol{\rho}_1 - \tau_1 \boldsymbol{\rho}_2, \tag{110}$$

we obtain, with due respect to the orthogonality of the transformation (110), the following expression for the integral I on the right-hand side of Eq. (108):

$$I = \iint d\mathbf{r}_1 \, d\mathbf{r}_2 \left\{ \mathscr{Y}^{n_1}_{l_1}(\tau_1\mathbf{r}_1 + \tau_2\mathbf{r}_2) \otimes \mathscr{Y}^{n_2}_{l_2}(\tau_2\mathbf{r}_1 - \tau_1\mathbf{r}_2) \right\}_{lm} \times \exp[-\tfrac{1}{2}(\mathbf{r}_1 - t\mathbf{r})^2] \exp(-\tfrac{1}{2}r_2^2). \tag{111}$$

The tensor polynomial in the integrand expression (111) can be expressed as a linear combination of polynomials $\{ \mathscr{Y}(\mathbf{r}_1) \otimes \mathscr{Y}(\mathbf{r}_2) \}_{lm}$ with some coefficients depending on continuous parameters τ_1 and τ_2. Such a procedure corresponds to the standard Talmi transformation:

$$\{ \mathscr{Y}^{n_1}_{l_1}(\tau_1\mathbf{r}_1 + \tau_2\mathbf{r}_2) \otimes \mathscr{Y}^{n_2}_{l_2}(\tau_2\mathbf{r}_1 - \tau_1\mathbf{r}_2) \}_{lm} = \sum_{N_1 L_1 N_2 L_2} T^{n_1 l_1 n_2 l_2}_{N_1 L_1 N_2 L_2}(l; \tau_1, \tau_2) \{ \mathscr{Y}^{N_1}_{L_1}(\mathbf{r}_1) \otimes \mathscr{Y}^{N_2}_{L_2}(\mathbf{r}_2) \}_{lm}. \tag{112}$$

The coefficients of this transformation have been calculated in the general case by Smirnov.[20] Various definitions for such coefficients are used, which differ, mainly, in phase and normalizing multipliers. Besides, various algebraic and recurrence formulas have been derived for these coefficients (see, for example, Refs. 21 and 22). The coefficients T in Eq. (112) may be easily related to a more standard definition of the Talmi–Smirnov coefficients.[21,22] One may also use the special representation for these coefficients,[23] which takes into account the specific features inherent in the molecular case in a more thorough way.

Substituting Eq. (112) into Eq. (111), performing the explicit integration in $\mathbf{r}_2$, and representing the integral by $\mathbf{r}_1$, with the help of Eq. (105), we obtain the following addition theorem:

$$\left\{ \mathscr{L}^{n_1}_{l_1}\left(\frac{i}{\sqrt{2}} t_1 \mathbf{r}\right) \otimes \mathscr{L}^{n_2}_{l_2}\left(\frac{i}{\sqrt{2}} t_2 \mathbf{r}\right) \right\}_{lm} = \frac{(-1)^{l_1 + l_2}}{\pi n_1! n_2! 2^{n_1 + n_2 + (l_1 + l_2 - l)/2}} \sum_{N_1, N_2} T^{n_1 l_1 n_2 l_2}_{N_1 l N_2 0}(l; \tau_1, \tau_2) \times 2^{N_1 + N_2} N_1! \Gamma\left(N_2 + \frac{3}{2}\right) \mathscr{L}^{N_1}_{lm}\left(\frac{i}{\sqrt{2}} t\mathbf{r}\right). \tag{113}$$

Note that the summation variables N_1 and N_2 are interrelated by the condition

$$N_1 + N_2 = n_1 + n_2 + \tfrac{1}{2}(l_1 + l_2 - l). \tag{114}$$

The particular type of coefficients T with $L_1 = l$, $L_2 = 0$ in Eq. (113) has a more simple form in comparison with the general case (see, for example, Refs. 22 and 24). Note that these particular formulas allow us to express the coefficients T in Eq. (113) in the form of the standard hypergeometric functions.[24]

7. CONCLUSIONS

Thus, to derive algebraic representations for GDO's, one may use either the Fourier-transformation method, or the commutation relations method, which leads to a convenient integral representation for GDO. The Fourier-transformation method is more simple in handling; however, it leads to specific GDO's which are factorized in operators $\nabla(\mathbf{R})$ and $d(\omega)$. In this case the radial part of the GDO for two important types of basis functions—the Slater and hydrogenlike ones—is expressed in the form of hypergeometric functions $_2F_1$, which satisfy the quadratic transformations. The relations, obtained in Ref. 4, allow us to extend such an approach to more general basis functions of exponential class. The commutation relations method allows us to introduce the mixed, nonfactorized GDO's, whose radial part has a more simple analytic structure. In the case of the Slater and the Gaussian functions, for example, the radial part of GDO is associated with the Laguerre polynomials, and the GDO's have a form of polynomial parts, $\mathscr{L}$, of the harmonic oscillator eigenfunctions. First, it leads to a remarkable analogy between two important classes of basis functions. Second, it allows us to use in calculations a number of important relations for functions $\mathscr{L}$, which have been established in Sec. 6. Apparently, it is worth noting that there exists the remarkable relation between the coefficients of the Clebsch–Gordan series for functions $\mathscr{L}$ and the particular type of the Talmi–Smirnov coefficients T. Note also that Eq. (107) allows one, with due respect of the addition theorems for the Laguerre polynomials,[25] to obtain an alternative expression for the corresponding coefficients in the form of the generalized hypergeometric functions, ^NF [26] of two variables, which, therefore, turn out to be related to the particular type of coefficients T, mentioned above.

The reduction of the basis functions in the integrand expressions for matrix elements to the functions G_ω allows us to use for calculating and analyzing the properties of molecular integrals, along with the Laplace equation for the Coulomb potential,[8] also the Helmholtz equation for functions G_ω, which makes it possible to improve some earlier approaches.[7–10]

Note that the possibility of generating basis functions from the simplest ones through differentiating by parameters has been indicated by Boys.[27] Shavitt[2] had used this possibility for some simple functions with small values of quantum numbers and had pointed out also the possibility of utilizing computer techniques for the generation of more complicated functions. The proposed method, which is the direct extension of the earlier work,[28] yields a far reaching generalization of such an approach to the case of any basis function and any quantum number and may be used in various methods associated with the procedure of differentiating functions by parameters (see, for example, Refs. 2 and 29).

ACKNOWLEDGMENTS

The author is indebted to Professor L. A. Gribov for valuable advice and discussions. The author highly acknowledges Professor Yu. F. Smirnov's comments on computational problems in applied quantum mechanics.

[1] MO LCAO SCF = molecular orbital—linear combination of atomic orbitals—self-consistent field. See J. C. Slater, *Electronic Structure of Molecules* (McGraw-Hill, New York, 1963).

[2] I. Shavitt, in *Methods in Computational Physics*, edited by B. Alder, S. Fernbach, and M. Rotenberg (Academic, New York, 1963), p. 16.

[3] E. Filter and E. O. Steinborn, J. Math. Phys. **19**, 79 (1978).

[4] A. W. Niukkanen, Int. J. Quant. Chem. (to be published). A general method of calculating Fourier transforms of basis functions of exponential class, including particular cases and possible transformations of radial parts, is considered in this paper.

[5] *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill, New York, 1953), Vol. 2, Eqs. 7.11 (21) and 7.11 (22).

[6] This relation may be easily obtained also as the direct consequence of Eq. (15) since, with respect to the parameter ω, any linear differential relation for $\bar{\Xi}$ is valid for the function Ξ as well as due to linearity of the Fourier transformation.

[7] F. E. Harris, J. Chem. Phys. **51**, 4770 (1969).

[8] K. O-Ohata and K. Ruedenberg, J. Math. Phys. **7**, 547 (1966).

[9] T. Szondy, Acta. Phys. Hung. **18**, 381 (1965).

[10] D. M. Silver and K. Ruedenberg, J. Chem. Phys. **49**, 4306 (1968).

[11] A. W. Niukkanen, J. Math. Phys. (to be published).

[12] See Ref. 5, Vol. 1, Secs. 2.1.5 and 2.11.

[13] J. Kampé de Fériet, R. Campbell, G. Petiau, and T. Vogel, *Fonctions de la physique matematique* (Dunod, Paris, 1957).

[14] K. Kumar, J. Math. Phys. **7**, 671 (1966).

[15] See Ref. 5, Vol. 2, Eq. 10.13(34).

[16] D. A. Varshalovich, A. N. Moskalyov, and V. K. Khersonsky, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975) (in Russian).

[17] K. G. Kay, H. D. Todd, and H. J. Silverstone, J. Chem. Phys. **51**, 2359 (1969).

[18] A. W. Niukkanen and L. A. Gribov, Theor. Chim. Acta (Berlin) **62**, 443 (1983).

[19] V. P. Maslov, *Operator Methods* (Nauka, Moscow, 1973) (in Russian).

[20] Yu. F. Smirnov, Nucl. Phys. **27**, 177 (1961); **39**, 346 (1962).

[21] M. Moshinsky, *The Harmonics Oscillator in Modern Physics: From Atoms to Quarks* (Gordon and Breach, New York, 1969).

[22] G. Fieck, J. Phys. B: Atom. Molec. Phys. **12**, 1063 (1979).

[23] A. W. Niukkanen, Chem. Phys. Lett. **69**, 174 (1980).

[24] L. Trlifaj, Phys. Rev. C **5**, 1535 (1972).

[25] A. W. Niukkanen, submitted to J. Phys. A: Math. Gen. Phys.

[26] A. W. Niukkanen, J. Phys. A: Math. Gen. Phys. **16**, 1813 (1983).

[27] S. F. Boys, Proc. R. Soc. London, Ser. A **200**, 542 (1950).

[28] A. W. Niukkanen, reports at the Conference on Atomic and Molecular Theory, Vilnyus, 1979 (in Russian).

[29] M. N. Adamov, R. A. Evarestov, A. I. Ivanov, and V. Z. Slonim, Preprint ITP 75-20, Kiev, 1975 (in Russian).

Relativistic motion of a charged particle, the Lorentz group, and the Thomas precession

Nikos Salingaros
Department of Mathematics, The University of Iowa, Iowa City, Iowa 52242

(Received 30 December 1982; accepted for publication 18 February 1983)

The equation of motion of a classical charged particle in a homogeneous, static electromagnetic field is solved exactly in terms of a Lorentz transformation. All higher-order corrections to the motion follow from the Baker–Campbell–Hausdorff formula. Some of these terms, as well as corrections to the Thomas precession, are calculated here for the first time. The derivation is based on an intrinsic formulation of the Lorentz group using differential forms and the Clifford algebra in Minkowski space-time.

PACS numbers: 41.70. + t, 02.10. + w, 02.20. + h, 03.30. + p,

1. INTRODUCTION

In this paper, we show how the classical motion of a charged particle in an external electromagnetic field can be derived directly from the space-time structure. Our treatment does not involve curvature or gravitation, but is strictly concerned with flat four-dimensional space-time with the Lorentz–Minkowski metric. The only assumption is an antisymmetric tensor structure in terms of the differential form basis,[1] and an algebraic product between the forms which realizes a Clifford algebra.[2–6] This geometrical matrix-free field description was previously discussed in Refs. 7 and 8, and is reviewed in Sec. 2 of this paper.

In Sec. 3, we derive the Lorentz group[9–12] from the intrinsic algebraic structure. We review some standard results cast in this particular formalism, and then describe a convenient method of performing finite Lorentz transformations and spatial rotations (Sec. 4). This is more general than the usual infinitesimal treatments.[13–16]

In Sec. 5 we present a general solution of the equation $d\alpha/dt = [\beta,\alpha]$, where α and β are elements in the Clifford algebra. This equation includes the Heisenberg and rotation equations, and hence describes the behavior of a large class of physical systems. As an illustration, we discuss the spin precession of a particle in a magnetic field in the nonrelativistic case.

The Lorentz force law is derived directly in the Clifford algebra in Sec. 6. Our key result is in showing that the Lorentz force law is a special case of the equation of Sec. 5, and that the general solution is a Lorentz transformation of the initial particle velocity. Hence, the motion of a charged particle in an external electromagnetic field can be written down directly (Sec. 7). The separation of the rotational motion from the linear motion follows from an application of the Baker–Campbell–Hausdorff formula.[17,18] Here, we display terms up to third order explicitly; terms of order three and higher are not usually calculated in the standard treatments.[14,19,20] When they have been calculated by other methods,[21,22] the expressions do not appear as general as those obtained here.

Finally, in an entirely distinct application of the Baker–Campbell–Hausdorff formula, we give a simple derivation of the Thomas precession.[15,16] Because of the generality of the formalism, we can calculate the next-order term explicitly and show that it is a small correction to the net Lorentz boost which does not affect the rotation (Sec. 8). A related but distinct discussion of the topics in this paper is given in Refs. 23 and 24.

It is appropriate at this point to recall the differential form basis of space-time as used in the text.[7,8] Space-time is described by the four coordinates x^1, x^2, x^3, and $x^4 = t$. The differential 1-forms[1] dx^μ, $\mu = 1, \ldots ,4$, define an orthogonal basis frame for vectors. The Grassmann (or exterior) product $\wedge$ is used to construct area elements and volume elements in space-time from the basis 1-forms.[1] The collection of all possible geometrical objects is the following set of 16 basis forms:

$$\begin{aligned}&\{1, dx^\mu, dx^\mu \wedge dx^\nu, dx^\mu \wedge dx^\nu \wedge dx^\lambda,\\ &\quad dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4 = \omega\},\\ &\quad \mu,\nu,\lambda = 1,2,3,4,\ \mu \neq \nu \neq \lambda.\end{aligned} \tag{1.1}$$

We have included the scalar unit 1 as the zero-rank basis form. The rank of each type of form in (1.1) is, respectively, 0, 1, 2, 3, and 4, and there are 1,4,6,4,1 basis forms of each corresponding rank. It is convenient to label the basis 1-forms by the symbol σ, and also to label the three-dimensional and four-dimensional volume elements as η and ω, respectively:

$$\sigma^\mu = dx^\mu, \tag{1.2a}$$

$$\eta = dx^1 \wedge dx^2 \wedge dx^3, \tag{1.2b}$$

$$\omega = dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4. \tag{1.2c}$$

An inner (scalar) product can be defined in this space in terms of the Lorentz–Minkowski metric:

$$g^{\mu\nu} = (\sigma^\mu, \sigma^\nu) = 0, \quad \mu \neq \nu, \tag{1.3a}$$

$$g^{\mu\mu} = (\sigma^\mu, \sigma^\mu) = \begin{cases} -1, & \mu = 1,2,3, \\ +1, & \mu = 4. \end{cases} \tag{1.3b}$$

Using the differential form basis (1.1) and the metric (1.3), we construct a geometrical realization of the Clifford algebra in Minkowski space-time. This is detailed in Ref. 7 and 8. Here, we will need to manipulate tensor fields defined in the Clifford algebra, which is reviewed in the following section.

A note on units: We employ the physicist's convention of setting the speed of light c equal to 1. One may rewrite equations given in these "natural" units by inserting factors of c as follows: the Thomas precession [Eq. (8.10)] acquires a factor of $1/c^2$ on the right-hand side, and the angular frequencies [Eqs. (5.9b) and (7.4)] a factor of $1/c$.

2. PROPERTIES OF THE $\vee$ ALGEBRA IN SPACE-TIME

In this section, we review briefly the description of fields in four-dimensional space-time introduced in Ref. 7 and discussed in detail in Ref. 8. The elements of this algebraic framework are called "tensor types." They are real, antisymmetric, tensor fields, which are expanded on the differential form basis (1.1). Since the rank of antisymmetric tensors in four dimensions can be either zero, one, two, three, or four, these are precisely the "types" that are possible in space-time. We can display representative tensor types as follows:

$$a_0, \qquad \text{scalar} \tag{2.1a}$$

$$a=\sum a^{\mu}\sigma^{\mu}, \qquad \text{vector type,} \tag{2.1b}$$

$$F=\tfrac{1}{2}\sum F^{\mu\nu}\sigma^{\mu}\wedge\sigma^{\nu}, \qquad \text{tensor type 2,} \tag{2.1c}$$

$$M=\frac{1}{3!}\sum M^{\mu\nu\lambda}\sigma^{\mu}\wedge\sigma^{\nu}\wedge\sigma^{\lambda}, \qquad \text{tensor type 3,} \tag{2.1d}$$

$$b=b_0\,\omega=b_0\,\sigma^1\wedge\sigma^2\wedge\sigma^3\wedge\sigma^4, \qquad \text{tensor type 4,} \tag{2.1e}$$

$$\mu,\nu,\lambda=1,\ldots,4,\ \mu\neq\nu\neq\lambda.$$

The components of tensor types a_0, a^{μ}, $F^{\mu\nu}$, $M^{\mu\nu\lambda}$, and b_0 are all real scalars.

The tensor types correspond directly to the physical fields in space-time which we wish to describe. For example, the particle 4-momentum p, the electromagnetic potential a, and the electromagnetic current j are all vector types, as in (2.1b). The electromagnetic field F is a tensor of type 2 as in (2.1c). The 4-dual of the current $\overset{*}{_4}j$ is a tensor of type 3. The 4-duals are defined in the usual way using the Levi-Civita entirely antisymmetric index symbol. Indices are lowered using the metric $g^{\mu\nu}=g_{\mu\nu}$ [(1.3)]:

$$\overset{*}{_4}(j^{\mu}\sigma^{\mu})=(1/3!)j^{\mu}\epsilon^{\mu}{}_{\nu\lambda\rho}\sigma^{\nu}\wedge\sigma^{\lambda}\wedge\sigma^{\rho}, \tag{2.2a}$$

$$\overset{*}{_4}(F^{\mu\nu}\sigma^{\mu}\wedge\sigma^{\nu})=\tfrac{1}{2}F^{\mu\nu}\epsilon^{\mu\nu}{}_{\lambda\rho}\sigma^{\lambda}\wedge\sigma^{\rho}. \tag{2.2b}$$

(See Refs. 1 and 8 for a full discussion of duality in both three and four dimensions.)

It is possible to describe the tensor types (2.1) using ordinary vector algebra, by means of the following decomposition.[7,8] The usual notation for vectors in the three-dimensional Euclidean subspace of space-time can be utilized to write a tensor type 1, or vector type, (2.1b) as

$$a=\mathbf{a}+a^4\sigma^4,\quad \mathbf{a}=\sum a^i\sigma^i. \tag{2.3}$$

In the case of a tensor type 2, we can decompose the components of F [(2.1c)] using two spatial vectors $\mathbf{e}$ and $\mathbf{b}$, and duality in the three-dimensional spatial subspace. This is entirely analogous to separating the electromagnetic field tensor into the electric and magnetic field vectors[5,7]

$$F=\mathbf{e}\wedge\sigma^4-\overset{*}{_3}\mathbf{b},\quad e^i=F^{i4},\quad b^i=-\tfrac{1}{2}\sum\epsilon^{ijk}F^{jk}. \tag{2.4}$$

An important and useful property of the $\vee$ product is the ability to express the duality operation algebraically. We recall the "duality theorem" from Ref. 5. For any tensor type α, the 3- and 4-duals are obtained by $\vee$ multiplication with the three- and four-dimensional volume elements, respectively:

$$\overset{*}{_3}\alpha=(\pm)\,\eta\vee\alpha,\quad \overset{*}{_4}\alpha=(\pm)\,\omega\vee\alpha. \tag{2.5}$$

The signs in (2.5) are determined by the type of the tensor in each case. The cases of immediate interest are the tensor types in (2.3) and (2.4), whose duals are given as follows:

$$\overset{*}{_3}\mathbf{a}=-\eta\vee\mathbf{a}, \tag{2.6a}$$

$$\overset{*}{_4}a=\omega\vee a, \tag{2.6b}$$

$$\overset{*}{_4}F=\omega\vee F. \tag{2.6c}$$

The product $\vee$ establishes an "algebra of tensor types" that is both associative and has an inverse.[5,8] We recall from Refs. 7 and 8 the $\vee$ product rules between tensor types which will be needed in the sequel. First, the $\vee$ product of a vector type a with another vector type b as in (2.1b) and (2.3) is given in the usual vector notation as

$$\begin{aligned}a\vee b&=(a,b)+a\wedge b\\&=a^4b^4-(\mathbf{a}\cdot\mathbf{b})-\eta\vee\mathbf{a}\times\mathbf{b}+(b^4\mathbf{a}-a^4\mathbf{b})\vee\sigma^4,\end{aligned} \tag{2.7}$$

$$(a,b)=\sum a_{\mu}b^{\mu},\quad (\mathbf{a}\cdot\mathbf{b})=\sum a^ib^i.$$

The square of a vector type in the $\vee$ algebra is just the Minkowski quadratic form

$$a\vee a=(a,a)=(a^4)^2-(\mathbf{a}\cdot\mathbf{a}). \tag{2.8}$$

These expressions demonstrate how the $\vee$ product generalizes the ordinary vector algebra from three to four dimensions. The use of the traditional vector notation in describing the $\vee$ products of arbitrary tensor types is made possible by the consistent use of the space-time decomposition (2.3), (2.4).

The $\vee$ product of a vector type a with a tensor of type-2 F is given in terms of the decomposition (2.4) with (2.6a) as

$$\begin{aligned}a\vee F&=a\vee(\mathbf{e}\vee\sigma^4+\eta\vee\mathbf{b})\\&=-a^4\mathbf{e}-\mathbf{a}\times\mathbf{b}-(\mathbf{a}\cdot\mathbf{e})\sigma^4\\&\quad+\omega\vee[-(\mathbf{a}\cdot\mathbf{b})\sigma^4+\mathbf{a}\times\mathbf{e}-a^4\mathbf{b}].\end{aligned} \tag{2.9}$$

The $\vee$ product of two tensors of type 2, F and G, is similarly obtained. We have decomposed both F and G as in (2.4) and applied (2.6a). (Full details may be found in Ref. 8.)

$$\begin{aligned}F\vee G&=(\mathbf{e}\vee\sigma^4+\eta\vee\mathbf{b})\vee(\mathbf{g}\vee\sigma^4+\eta\vee\mathbf{h})\\&=(\mathbf{e}\cdot\mathbf{g})-(\mathbf{b}\cdot\mathbf{h})+(-\mathbf{e}\times\mathbf{h}-\mathbf{b}\times\mathbf{g})\vee\sigma^4\\&\quad+\eta\vee(\mathbf{e}\times\mathbf{g}-\mathbf{b}\times\mathbf{h})-\omega[(\mathbf{e}\cdot\mathbf{h})+(\mathbf{b}\cdot\mathbf{g})].\end{aligned} \tag{2.10}$$

A special product which is useful is that of F with itself.

TABLE I. Commutators and anticommutators of tensor types.

$\{a,b\} = 2(a,b) = 2[a^4b^4 - (\mathbf{a}\cdot\mathbf{b})]$
$[a,b] = 2a\wedge b = 2(-\eta\vee\mathbf{a}\times\mathbf{b} + (b^4\mathbf{a} - a^4\mathbf{b})\vee\sigma^4)$
$\{a,F\} = 2[-(\mathbf{a}\cdot\mathbf{b})\eta + (a^4\mathbf{b} - \mathbf{a}\times\mathbf{e})\vee\omega]$
$[a,F] = 2[-(\mathbf{a}\cdot\mathbf{e})\sigma^4 - a^4\mathbf{e} - \mathbf{a}\times\mathbf{b}]$
$\{F,G\} = 2\{(\mathbf{e}\cdot\mathbf{g}) - (\mathbf{b}\cdot\mathbf{h}) - \omega[(\mathbf{e}\cdot\mathbf{h}) + (\mathbf{b}\cdot\mathbf{g})]\}$
$[F,G] = 2[-(\mathbf{e}\times\mathbf{h} + \mathbf{b}\times\mathbf{g})\vee\sigma^4 + \eta\vee(\mathbf{e}\times\mathbf{g} - \mathbf{b}\times\mathbf{h})]$
$a = \mathbf{a} + a^4\sigma^4, \quad b = \mathbf{b} + b^4\sigma^4$
$F = \mathbf{e}\vee\sigma^4 + \eta\vee\mathbf{b}, \quad G = \mathbf{g}\vee\sigma^4 + \eta\vee\mathbf{h}$

From (2.10), we have

$$F\vee F = |\mathbf{e}|^2 - |\mathbf{b}|^2 - 2\omega(\mathbf{e}\cdot\mathbf{b}), \tag{2.11}$$

$$|\mathbf{e}| = \sqrt{\mathbf{e}\cdot\mathbf{e}}.$$

In the case of the electromagnetic field, the two parts (scalar and pseudoscalar) of this product (2.11) are precisely the two combinations of the electric and magnetic fields which are invariant under Lorentz transformations.[14,16]

It can be seen that the manipulations of tensor types in space-time reduce to vector rules (2.7), and extensive applications of the duality theorem (2.6). For this reason, all that one needs in practice are the above multiplication rules for the fields, along with rules for manipulating the volume elements. These are summarized in the following theorem:

Theorem 1: (a) η commutes with all spatial tensor types and anticommutes with σ^4 and ω.

(b) ω commutes with tensors or even type, and anticommutes with tensors of odd type.

(c) $\eta\vee\sigma^4 = \omega, \quad \eta\vee\eta = 1, \quad \omega\vee\omega = -1.$

(d) $\{\sigma^1,\sigma^2,\sigma^3,\sigma^4,\omega\}$ mutually anticommute. (2.12)

In the later sections, it will be necessary to calculate commutator $\vee$ products of particular tensor types. Commutators and anticommutators can be defined in terms of the $\vee$ product as follows:

Definition 1:

$$\{\alpha,\beta\} = \alpha\vee\beta + \beta\vee\alpha, \tag{2.13a}$$

$$[\alpha,\beta] = \alpha\vee\beta - \beta\vee\alpha. \tag{2.13b}$$

Here α and β are arbitrary tensor types. For convenience, we have included some identities for commutator and anticommutator $\vee$ products in Table I, which will be useful later. This concludes the review of tensor types and their algebra.

3. AUTOMORPHISMS OF TENSOR TYPES AND THE LORENTZ GROUP

In this section, we will use exponentials of tensor types to define the automorphisms of tensor types. We can define the exponential function on the tensor types by using the standard expression for the exponential interpreted in the $\vee$ algebra:

$$\exp(\beta) = \sum_{n=0}^{\infty} \frac{(\beta)^n}{n!}. \tag{3.1}$$

Here, β is any tensor type. The nth power of β is defined to be the $\vee$ product of n copies of β (which is associative):

Definition 2:

$$(\beta)^n = \beta\vee\beta\vee\cdots\vee\beta \quad (n \text{ times}). \tag{3.2}$$

From the closure of the $\vee$ algebra, $(\beta)^n$ will always be a sum of tensor types. One can therefore use (3.2) explicitly to calculate the exponential of any tensor type β. (The convergence of this expression will not be discussed here.)

In order to evaluate the exponentials of the tensor types a [(2.3)] and F [(2.4)], we use the products (2.8) and (2.11). We calculate the exponentials of the vector type a (as well as for the space vector $\mathbf{a}$), and the exponentials of the two parts of F [(2.4)] $\mathbf{e}\wedge\sigma^4$ and $\overset{*}{_3}\mathbf{b}$, separately. A straightforward calculation gives closed-form expressions in terms of trigonometric or hyperbolic functions. The notation used is $|a| = (a_\mu a^\mu)^{1/2}$ and $|\mathbf{a}| = (a^i a^i)^{1/2}$.

Theorem 2:

$$\exp(a) = \cosh|a| + (a/|a|)\sinh|a|, \tag{3.3a}$$

$$\exp(\mathbf{a}) = \cos|\mathbf{a}| + (\mathbf{a}/|\mathbf{a}|)\sin|\mathbf{a}|, \tag{3.3b}$$

$$\exp(\mathbf{e}\wedge\sigma^4) = \cosh|\mathbf{e}| + [(\mathbf{e}\wedge\sigma^4)/|\mathbf{e}|]\sinh|\mathbf{e}|, \tag{3.3c}$$

$$\exp(\overset{*}{_3}\mathbf{b}) = \cos|\mathbf{b}| + (\overset{*}{_3}\mathbf{b}/|\mathbf{b}|)\sin|\mathbf{b}|. \tag{3.3d}$$

These calculations show that the exponentials of these types are expressible as the sum of a scalar with a tensor of the same type. The above expressions are analogous to the Euler formula for the exponential of a complex number—the analogy follows since the $\vee$ algebra is a generalization of the complex algebra.[5,6] [Note also that the distinction between three- and four-dimensional vectors (3.3a) and (3.3b) generates an extra minus sign due to the metric, changing the hyperbolic functions to trigonometric functions.]

We list some useful identities for the exponentials in the algebra. They are easily verified using expressions (3.3); α and β are any tensor types; the commutator and anticommutator are defined as in (2.13).

Lemma:

$$\exp(\beta)\vee\exp(-\beta) = \exp(-\beta)\vee\exp(\beta) = 1, \tag{3.4a}$$

$$[\alpha,\beta] = 0 \Leftrightarrow \exp(\beta)\vee\alpha = \alpha\vee\exp(\beta), \tag{3.4b}$$

$$\{\alpha,\beta\} = 0 \Leftrightarrow \exp(\beta)\vee\alpha = \alpha\vee\exp(-\beta). \tag{3.4c}$$

The object of this section is to define the automorphisms of the tensor types, using exponentials such as (3.3). By analogy to the definition of a similarity transformation in the matrix language, we write an expression for the transformation of α as:

Definition 3:

$$\alpha' = \exp(\beta)\vee\alpha\vee\exp(-\beta). \tag{3.5}$$

This general expression will be an automorphism if, and only if, it is *type-preserving*. The determination of which fields α and β satisfy (3.5) such that the type of α and α' is the same can be accomplished by direct substitution of all possible tensor types into (3.5). The type-preserving condition distinguishes the following cases: β can only be a tensor of type two; α can be either a vector type, a tensor type 2, or a combination of vector and tensor type 3. This is an exhaustive result in four-dimensional space-time. Identity (3.4a) guarantees that they are indeed inner automorphisms.

What we have actually done is to obtain the Lorentz group directly from four-dimensional space-time. Moreover, this is the unique inner automorphism of fields in space-time, as may be verified from the Clifford group automorphism theorem.[2,3,6,12]

Theorem 3: (i) The basis 2-forms of a Clifford algebra in the space $M^{p,q}$ define a Lie algebra via a commutator product which is locally isomorphic to so(p,q).

(ii) The group of (inner) automorphisms of this Clifford algebra is the Lie group associated with the Lie algebra so(p,q).

In the case of space-time $M^{1,3}$, the automorphism group is the Lie group corresponding to the Lie algebra so(1,3). This is precisely the Lorentz group with elements $\exp(\beta)$, where β is a tensor type 2.

We now proceed to relate our derivation to the usual treatments of the Lorentz group.[9–12] The six basis 2-forms $\sigma^\mu \wedge \sigma^\nu$ in (1.1) can be shown to satisfy the following commutation relations in the $\vee$ algebra, by labeling them $J^{\mu\nu} = \frac{1}{2}\sigma^\mu \wedge \sigma^\nu$ [5]:

$$[J^{\mu\nu}, J^{\lambda\rho}] = g^{\nu\lambda} J^{\mu\rho} - g^{\mu\lambda} J^{\nu\rho} - g^{\nu\rho} J^{\mu\lambda} + g^{\mu\rho} J^{\nu\lambda},$$
$$\mu,\nu,\lambda,\rho = 1,2,3,4. \tag{3.6}$$

Here, $g^{\mu\nu}$ is the Lorentz–Minkowski metric (1.3), and relations (3.6) define the Lie algebra so(1,3). The space 2-forms $J^{ij} = \frac{1}{2}\sigma^i \wedge \sigma^j$ define the closed subalgebra so(0,3)$\simeq$so(3) when the commutation relations (3.6) are restricted to the three space indices. To see this, label the spatial basis 2-forms as $L^1 = J^{23}$, $L^2 = J^{31}$, $L^3 = J^{12}$; i.e., $L^k = \frac{1}{2}{}^{*}_{3}\sigma^k$. We can then rewrite the spatial part of (3.6) as

$$[L^i, L^j] = \epsilon^{ijk} L^k, \quad i,j,k = 1,2,3. \tag{3.7}$$

These commutation relations describe the rotation algebra in three dimensions. The physical consequence of this result is that the exponentials of $\beta = {}^{*}_{3}\mathbf{b}$ in the automorphism (3.5) describe spatial rotations in three-dimensional space. The other transformations, involving exponentials of $\beta = \mathbf{e} \wedge \sigma^4$, mix the space with the time components of tensor fields, and therefore describe Lorentz boosts, as we show in the next section.

We note that the basis forms $J^{i4} = \frac{1}{2}\sigma^i \wedge \sigma^4$ corresponding to pure Lorentz boosts do not define a closed subalgebra. Hence, an algebraic combination of two Lorentz boost operators will in general create an additional spatial rotation term. This is physically manifested in the Thomas precession, which is discussed in detail in Sec. 8.

For completeness, we can relabel the space-time 2-forms as $K^i = J^{i4} = \frac{1}{2}\sigma^i \wedge \sigma^4$, in order to write the commutation relations of the Lorentz Lie algebra (3.6) in the standard manner.[9–12] Note that, in contrast to other treatments, the commutation relations as given here are strictly real:

$$\begin{aligned} [L^i, L^j] &= \epsilon^{ijk} L^k, \\ [L^i, K^j] &= \epsilon^{ijk} K^k, \\ [K^i, K^j] &= -\epsilon^{ijk} L^k. \end{aligned} \tag{3.8}$$

These results demonstrate that the Lorentz group is a consequence of a Clifford algebraic structure in four-dimensional space-time and does not have to be an additional assumption.

4. ROTATIONS AND LORENTZ TRANSFORMATIONS

In this section, spatial rotations and Lorentz transformations are described in some detail. The algebraic manipulations are simple and straightforward and lead to quite general results. Our description is finite, and is valid for transformations along any axes. This stands in contrast to other treatments which examine only infinitesimal transformations, or finite transformations along one particular axis only.

Following the discussion in the preceding section, we define a rotation operator $\mathbb{R}$ as the exponential of the space-dual of a three-dimensional vector $\boldsymbol{\theta}$. The three components $\theta^1, \theta^2, \theta^3$ are the three spatial rotation parameters.

Definition 4:

$$\mathbb{R}(\boldsymbol{\theta}) = \exp(\tfrac{1}{2}\,{}^{*}_{3}\boldsymbol{\theta}). \tag{4.1}$$

The definition of automorphisms in the $\vee$ algebra implies that a tensor type α is transformed under a spatial rotation according to (3.5):

Theorem 4:

$$\alpha' = \mathbb{R}(\boldsymbol{\theta}) \vee \alpha \vee \mathbb{R}^{-1}(\boldsymbol{\theta}). \tag{4.2}$$

The inverse rotation operator is $\mathbb{R}^{-1}(\boldsymbol{\theta}) = \mathbb{R}(-\boldsymbol{\theta})$ [(3.5) and (3.4a)]. An algebraic expression for the rotation operator (4.1) can be obtained from (3.3d). We proceed to write down the explicit forms that transformation (4.2) assumes in the different cases of interest. In the case of a vector type (necessarily in three dimensions) $\mathbf{a} = a^i\sigma^i$, Eq. (4.2) becomes the usual formula for the conical rotation of a vector $\mathbf{a}$ about $\boldsymbol{\theta}$, by an angle $|\boldsymbol{\theta}|$.[13,25] This is illustrated in Fig. 1. Using (3.3d), (4.1), (4.2), and some $\vee$ algebra, one obtains

$$\begin{aligned} \mathbf{a}' = {} & \mathbf{a}\cos|\boldsymbol{\theta}| + [(\boldsymbol{\theta}\times\mathbf{a})/|\boldsymbol{\theta}|]\sin|\boldsymbol{\theta}| \\ & + (\boldsymbol{\theta}/|\boldsymbol{\theta}|^2)(\boldsymbol{\theta}\cdot\mathbf{a})(1-\cos|\boldsymbol{\theta}|). \end{aligned} \tag{4.3}$$

What is important to note is that spatial rotations of a type 2 field are also described by the transformation (4.2).

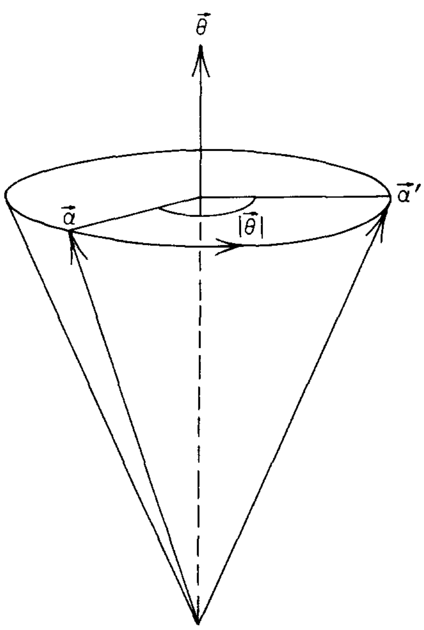


FIG. 1. Conical rotation of the vector $\mathbf{a}$ about $\boldsymbol{\theta}$.

For the type 2 field $F = \mathbf{e} \wedge \sigma^4 - \overset{*}{_3}\mathbf{b}$ (2.4), the rotation of the field about $\boldsymbol{\theta}$ by an angle $|\boldsymbol{\theta}|$ is obtained from (3.3d), (4.1), (4.2), and $\vee$ algebra as

$$\begin{aligned} F' = (&\mathbf{e}\cos|\boldsymbol{\theta}| + [(\boldsymbol{\theta}\times\mathbf{e})/|\boldsymbol{\theta}|]\sin|\boldsymbol{\theta}| \\ &+ \boldsymbol{\theta}[(\boldsymbol{\theta}\cdot\mathbf{e})/|\boldsymbol{\theta}|^2][1-\cos|\boldsymbol{\theta}|]) \wedge \sigma^4 \\ &- \overset{*}{_3}(\mathbf{b}\cos|\boldsymbol{\theta}| + [(\boldsymbol{\theta}\times\mathbf{b})/|\boldsymbol{\theta}|]\sin|\boldsymbol{\theta}| \\ &+ \boldsymbol{\theta}[(\boldsymbol{\theta}\cdot\mathbf{b})/|\boldsymbol{\theta}|^2][1-\cos|\boldsymbol{\theta}|]). \end{aligned} \tag{4.4}$$

It is instructive to consider the space-time decomposition of the transformed field F' in terms of two new vectors $\mathbf{e}'$ and $\mathbf{b}'$ such that $F' = \mathbf{e}' \wedge \sigma^4 - \overset{*}{_3}\mathbf{b}'$ (2.4). We can set this expression equal to (4.4) and then separate the components of $\mathbf{e}'$ and $\mathbf{b}'$. As a result, one sees that $\mathbf{e}$ and $\mathbf{b}$ transform independently, each as in (4.3). The physical consequence of this point is clear if one identifies $\mathbf{e}$ with the electric field $\mathbf{E}$ and $\mathbf{b}$ with the magnetic field $\mathbf{B}$: It is an experimental fact that $\mathbf{E}$ and $\mathbf{B}$ transform independently under spatial rotations. They do mix, however, under Lorentz transformations (see below).

A rotation in four dimensions will in general be a Lorentz transformation. By analogy to (4.1), in which a spatial tensor type two defines the rotation operator in 3-space, a space-time tensor type 2 is used to describe a pure Lorentz boost. Define, therefore, a Lorentz boost operator $\mathbb{L}$ as:

Definition 5:

$$\mathbb{L}(\mathbf{b}) = \exp(-\tfrac{1}{2}\mathbf{b}\wedge\sigma^4). \tag{4.5}$$

The boost vector $\mathbf{b}$ is in the direction of the relative frame velocity $\mathbf{V}$, and its components are the three boost parameters b^1, b^2, b^3. In the Appendix, we provide a derivation of these identities:

$$\begin{aligned} &\gamma = (1 - |\mathbf{V}|^2)^{-1/2} = \cosh|\mathbf{b}|, \quad |\mathbf{V}| = \tanh|\mathbf{b}|, \\ &\gamma|\mathbf{V}| = \sinh|\mathbf{b}|. \end{aligned} \tag{4.6}$$

The Lorentz boost of any tensor type α to a frame moving with relative velocity $\mathbf{V}$ is accomplished by the transformation $[\mathbb{L}^{-1}(\mathbf{b}) = \mathbb{L}(-\mathbf{b})]$:

Theorem 5:

$$\alpha' = \mathbb{L}(\mathbf{b}) \vee \alpha \vee \mathbb{L}^{-1}(\mathbf{b}). \tag{4.7}$$

Our notation is the standard one: Primed fields are moving with respect to the unprimed observer's frame with velocity $\mathbf{V}$. As a first case, consider the boost of the vector field $a = a^\mu \sigma^\mu$. An expression for (4.7) is obtained from (2.3), (3.3c), (4.5), and $\vee$ algebra:

$$\begin{aligned} a' = \mathbf{a} &- \frac{\mathbf{b}}{|\mathbf{b}|} a^4 \sinh|\mathbf{b}| + \frac{(\mathbf{b}\cdot\mathbf{a})}{|\mathbf{b}|^2}\mathbf{b}(\cosh|\mathbf{b}| - 1) \\ &+ a^4\sigma^4\cosh|\mathbf{b}| - \frac{(\mathbf{b}\cdot\mathbf{a})}{|\mathbf{b}|}\sigma^4\sinh|\mathbf{b}|. \end{aligned} \tag{4.8}$$

One can conveniently separate the space and time components of the vector field a in order to write Eq. (4.8) in the familiar component form.[13-15] Using identities (4.6), we have

$$\begin{aligned} &a^{i'} = a^i - \gamma V^i a^4 - V^i[(\mathbf{V}\cdot\mathbf{a})/|\mathbf{V}|^2](1-\gamma) \\ &a^{4'} = \gamma a^4 - \gamma(\mathbf{V}\cdot\mathbf{a}). \end{aligned} \tag{4.9}$$

As in the case of spatial rotations, the Lorentz boost equation (4.7) can also be used to describe transformations of type 2 fields. This is particularly useful in the case of the electromagnetic field $f = \mathbf{E}\wedge\sigma^4 - \overset{*}{_3}\mathbf{B}$. The transformation of f described by (4.7) becomes, after some $\vee$ algebra,

$$\begin{aligned} f' = (&\mathbf{E}\cosh|\mathbf{b}| + [(\mathbf{b}\times\mathbf{B})/|\mathbf{b}|]\sinh|\mathbf{b}| \\ &+ \mathbf{b}[(\mathbf{b}\cdot\mathbf{E})/|\mathbf{b}|^2][1-\cosh|\mathbf{b}|]) \wedge \sigma^4 \\ &- \overset{*}{_3}(\mathbf{B}\cosh|\mathbf{b}| - [(\mathbf{b}\times\mathbf{E})/|\mathbf{b}|]\sinh|\mathbf{b}| \\ &+ \mathbf{b}[(\mathbf{b}\cdot\mathbf{B})/|\mathbf{b}|^2][1-\cosh|\mathbf{b}|]). \end{aligned} \tag{4.10}$$

If one proceeds as in the case of rotations, and decomposes the transformed electromagnetic field as $f' = \mathbf{E}'\wedge\sigma^4 - \overset{*}{_3}\mathbf{B}'$, one can then separate the transformation of the electric and magnetic field components. Using identities (4.6), we obtain the well-known transformation rules for the electric and magnetic fields[13-15]:

$$\begin{aligned} &E^{i'} = \gamma E^i + \gamma(V\times B)^i + V^i[(\mathbf{V}\cdot\mathbf{E})/|\mathbf{V}|^2](1-\gamma), \\ &B^{i'} = \gamma B^i - \gamma(V\times E)^i + V^i[(\mathbf{V}\cdot\mathbf{B})/|\mathbf{V}|^2](1-\gamma). \end{aligned} \tag{4.11}$$

The electric and magnetic fields do not transform independently, because electromagnetism is an intrinsically relativistic phenomenon.

The novelty of the above description of rotations and Lorentz transformations lies in the use of a single formula to describe transformations of tensors of any rank. This is in contrast to the more traditional index description, where a distinct transformation operator is required for each tensor index.

5. CANONICAL SOLUTIONS OF THE EQUATIONS OF MOTION

In this section, we use the result established so far to give a general formulation of a class of differential equations, along with their solutions. These include equations of motion of both classical and quantum systems. The mathematical framework is the connection between a Lie group and its Lie algebra, or the relationship between the Lie automorphism group and its derivation.[18] While the general theory is well known, methods of obtaining explicit solutions have to be developed separately.

Consider the automorphisms of a field α, described by Eq. (3.5). Assume, moreover, that the tensor type α is a function of the scalar parameter t (not necessarily the time). We propose the following result:

Theorem 6: The canonical solution to the system of equations

$$\frac{d\alpha}{dt} = [\beta,\alpha] \quad \text{is}$$

$$\alpha(t) = \exp\Big(\int_0^t \beta\, dt\Big) \vee \alpha(0) \vee \exp\Big(-\int_0^t \beta\, dt\Big). \tag{5.1}$$

The commutator is understood to be defined in the $\vee$ algebra. The proof is easily obtained by direct differentiation with respect to t, and using the commutation rules for the exponentials in the vee product (3.4). In the special case when β is not explicitly dependent on the parameter t, the canonical solution assumes the simpler form:

$$\frac{d\alpha}{dt} = [\beta,\alpha] \Leftrightarrow \alpha(t) = \exp(\beta t) \vee \alpha(0) \vee \exp(-\beta t). \tag{5.2}$$

As an illustrative example, we can apply this theorem to solve the rotation equation in three dimensions, which describes the rotation of any vector $\mathbf{a} = a^i\sigma^i$ about a direction given by the angular velocity $\mathbf{\Omega}$. In this case, t is indeed the ordinary time[25]:

$$\frac{d\mathbf{a}}{dt} = \mathbf{\Omega}\times\mathbf{a}. \tag{5.3}$$

From the $\vee$ product rules (2.9) and Table I, we can express the vector cross product as a commutator,

$$\mathbf{\Omega}\times\mathbf{a} = [\tfrac{1}{2}\,{}^*_3\mathbf{\Omega},\mathbf{a}]. \tag{5.4}$$

(The asterisk denotes the three-dimensional dual.[8]) Applying Theorem 6, Eq. (5.2) to (5.3), (5.4) immediately gives the canonical solution:

$$\mathbf{a}(t) = \exp(\tfrac{1}{2}{}^*_3\mathbf{\Omega}t)\vee\mathbf{a}(0)\vee\exp(-\tfrac{1}{2}{}^*_3\mathbf{\Omega}t). \tag{5.5}$$

From the discussion in Sec. 4, we see that Eq. (5.5) explicitly describes the conical rotation of the vector $\mathbf{a}$ about the angle $\mathbf{\theta} = \mathbf{\Omega}t$ (Fig. 1). This is precisely what was expected, but we have obtained here the rotation as a solution of the equation of motion (5.3), rather than the usual other way around. Moreover, this is an exact and finite result.

As an example of the motion of a quantum system, consider the precession of a particle with spin S in a uniform magnetic field $\mathbf{B}$.[26] Neglecting any orbital motion of the particle, the Hamiltonian $\mathscr{H}$ is given as the scalar product of the intrinsic magnetic moment $\mathbf{M}$ of the particle with the magnetic field $\mathbf{B}$:

$$\mathscr{H} = -\mathbf{M}\cdot\mathbf{B}, \tag{5.6a}$$

$$\mathbf{M} = (gq/2m)\mathbf{S}. \tag{5.6b}$$

Here, q is the charge of the particle, m its mass, and g a numerical constant, the "Landé g factor," characteristic of each particle. The motion of the particle is given by the Heisenberg equation of motion for the components of the spin, as follows:

$$i\hbar\frac{dS^i(t)}{dt} = [S^i(t),\mathscr{H}(t)]. \tag{5.7}$$

The three components of the spin S^i, $i = 1,2,3$, satisfy the usual spin commutation relations,[10,26] enabling us to translate (5.7) into the $\vee$ algebra by treating the S^i as scalar components of a vector type $\mathbf{S}$ in three dimensions:

$$[S^i,S^j] = i\hbar\epsilon^{ijk}S^k \tag{5.8a}$$

$$\Rightarrow\frac{d\mathbf{S}}{dt} = -\frac{gq}{2m}\mathbf{B}\times\mathbf{S},\quad \mathbf{S} = S^i\sigma^i. \tag{5.8b}$$

The canonical solution of Eq. (5.8b) is obtained in exactly the same way as in the first example. It describes a rotation about the direction of $\mathbf{B}$, with angular frequency $\mathbf{\Omega}$ given as follows:

$$\mathbf{S}(t) = \mathbb{R}(\mathbf{\Omega}t)\vee\mathbf{S}(0)\vee\mathbb{R}^{-1}(\mathbf{\Omega}t), \tag{5.9a}$$

$$\mathbf{\Omega} = -(qg/2m)\mathbf{B}. \tag{5.9b}$$

This rotation is in the positive sense (counterclockwise, as in Fig. 1) for negatively charged particles; for positively charged particles it is clockwise about the magnetic field direction.[25,26] These two examples illustrate the application of the differential formulation of automorphisms to obtaining direct solutions to physical problems in two simple cases.

We now proceed to apply this method to an intrinsically relativistic problem in four dimensions.

6. CHARGED PARTICLE IN AN ELECTROMAGNETIC FIELD: DERIVATION OF THE EQUATIONS OF MOTION

We consider the motion of a particle in an external electromagnetic field. In this section, we derive the classical equations of motion using the formalism of the $\vee$ algebra, and then obtain a general solution by applying the method of the previous section.

As a first step, we calculate the Lie derivative of a field in the direction of motion as follows: Consider the tensor type α, which is a function of all space-time variables, and take its derivative with respect to the time $t = x^4$. The total derivative is given by the chain rule as[14]

$$\frac{d\alpha}{dt} = \frac{dx^\mu}{dt}\frac{\partial\alpha}{\partial x^\mu} = [(\mathbf{V}\cdot\mathbf{\nabla}) + \partial_4]\alpha, \tag{6.1a}$$

$$\mathbf{r} = x^i\sigma^i,\quad \mathbf{V} = \frac{d\mathbf{r}}{dt},\quad \mathbf{\nabla} = \partial_i\sigma^i. \tag{6.1b}$$

It is convenient to introduce the relativistic factor γ as the derivative of the ordinary time t with respect to the invariant line element $|r| = \sqrt{x_\mu x^\mu} = \sqrt{t^2 - (\mathbf{r}\cdot\mathbf{r})}$[14] (see Appendix).

Definition 6:

$$\gamma = \frac{dt}{d\,|r|} = (1 - |\mathbf{V}|^2)^{-1/2}. \tag{6.2}$$

Using the relativistic factor γ given as above, we can define the relativistic four-dimensional velocity u [14,15] as a vector type in four dimensions which has unit length (2.8) (also see Appendix):

$$u = \frac{dr}{d\,|r|} = \gamma(\mathbf{V} + \sigma^4),\quad r = x^\mu\sigma^\mu, \tag{6.3a}$$

$$u\vee u = (u,u) = 1. \tag{6.3b}$$

The intrinsic differential operator acting on tensor types in four-dimensional space-time is the Dirac operator $D = \partial^\mu\sigma^\mu$. This was constructed and studied as a vector type operator in the $\vee$ algebra in Ref. 7.

We apply the operator D here to compute the Lie derivative of the field α along the relativistic velocity u. Using (6.1), (6.2), and (6.3), we have:

Theorem 7:

$$\frac{d\alpha}{d\,|r|} = \frac{d\alpha}{dt}\frac{dt}{d\,|r|} = \gamma(\mathbf{V} + \sigma^4,D)\alpha = (u,D)\alpha, \tag{6.4a}$$

$$D = \partial^\mu\sigma^\mu = -\mathbf{\nabla} + \partial^4\sigma^4. \tag{6.4b}$$

This result (6.4) shows that the Lie derivative in the direction of the relativistic motion is equivalent to the total derivative with respect to the line element $|r|$.

Now consider the motion of an electron with a four-dimensional relativistic momentum $p = mu$ in an external electromagnetic field. The Minkowski relativistic force exerted on the particle is defined as the derivative of the momentum with respect to the line element[14,15]:

Definition 7:

$$\mathscr{F} = \frac{dp}{d\,|r|}. \tag{6.5}$$

The electromagnetic interaction can be obtained by using minimal coupling, i.e., the gauge identification between the particle momentum p and the electromagnetic vector potential a as in quantum electrodynamics.[14,20] (Here, q is the charge of the particle, which for an electron equals $-e$.)

$$\begin{aligned} p(\text{total}) &= p(\text{original}) + p(\text{electromagnetic}) \\ &= p(\text{original}) - qa. \end{aligned} \tag{6.6}$$

The force due to just the external electromagnetic field corresponds to the derivative of the vector potential in (6.6). From (6.4), (6.5), and (6.6) we obtain

$$\mathscr{F} = \frac{dp}{d\,|r|} = -q(u,D)a. \tag{6.7}$$

The relativistic Minkowski force can be written in terms of the commutator of u with the electromagnetic field f as follows: In the $\vee$ algebra, the field f is the D derivative of the electromagnetic potential a, $f = D \vee a$, along with the Lorentz condition $(D,a) = 0$.[7] The Lie derivative of the electromagnetic potential is written in terms of the field f as

$$(u,D)a = \tfrac{1}{2}[u,f]. \tag{6.8}$$

The commutator can be calculated using the formulas given in Sec. 2. The Minkowski relativistic force (6.7) due to the external electromagnetic field is therefore equal to[14,15,20]

$$\mathscr{F} = -\tfrac{1}{2}q[u\,f] = -qu_\mu f^{\mu\lambda}\sigma^\lambda \tag{6.9a}$$

$$\Rightarrow \mathscr{F}^\lambda = q f^{\lambda\mu} u_\mu. \tag{6.9b}$$

We may use the space-time decomposition of the electromagnetic field (2.4) and the relativistic velocity u [(6.3a)] to write the Minkowski force (6.9a) in vector form, from the commutators of Table I.

$$\mathscr{F} = q\gamma[\mathbf{E} + \mathbf{V}\times\mathbf{B} + (\mathbf{E}\cdot\mathbf{V})\sigma^4]. \tag{6.10}$$

We note that the measured force on a moving particle is given by Newton's law as the ordinary time derivative of the momentum; it is given by expression (6.10) without the factor of γ. Decomposing the particle's four-dimensional momentum into the three-dimensional momentum $\mathbf{p}$ and the energy ϵ as $p = \mathbf{p} + \epsilon\sigma^4$, the usual Lorentz force law is obtained from (6.2), (6.4a), (6.5), and (6.10) as[14]

$$\frac{d\mathbf{p}}{dt} = q(\mathbf{E} + \mathbf{V}\times\mathbf{B}), \quad \frac{d\epsilon}{dt} = q(\mathbf{E}\cdot\mathbf{V}). \tag{6.11}$$

In this manner, the classical electrodynamic equations of motion arise in the formalism of the $\vee$ algebra.

7. CHARGED PARTICLE IN AN ELECTROMAGNETIC FIELD: SOLUTION OF THE EQUATIONS OF MOTION

In this section, we apply the results of Sec. 5 to solve the electromagnetic equations of motion derived in the previous section. The solution follows immediately, once we express the Lorentz force law in the commutator form (5.1).

The classical equation of motion can be written from Eqs. (6.5) and (6.9a) as a commutator of tensor types (using $p = mu$) as follows:

$$\frac{du}{d\,|r|} = \frac{q}{2m}[f,u]. \tag{7.1}$$

We assume for this discussion that the electromagnetic field f is static, homogeneous, and ignore any self-interaction effects. In that case, the equation of motion (7.1) has a canonical solution given by Eq. (5.2), where the parameter is the line element $|r| = \sqrt{x_\mu x^\mu}$.

Theorem 8:

$$u(|r|) = \exp[(q|r|/2m)f] \vee u(0) \vee \exp[-(q|r|/2m)f]. \tag{7.2}$$

This is a general solution of the motion of a charged particle in a static, homogeneous electromagnetic field, given in terms of the line element $|r|$. We proceed to discuss the physical meaning of this result.

Since f is a tensor of type 2, the canonical solution (7.2) is in the form of a general Lorentz transformation (3.5). The motion of a charged particle is given by a simultaneous boost and rotation of the initial velocity. The simplest case is that of a pure magnetic field. The canonical solution (7.2) then describes a rotation of the space part of the velocity about the magnetic field direction. Therefore, the particle describes a helical path which twists about the magnetic field line (Figure 2). From (4.1), (4.2), (7.2), and the decomposition $f = \mathbf{E}\wedge\sigma^4 - {}^{*}_{3}\mathbf{B}$ we have, when $\mathbf{E} = 0$:

$$\mathbf{u}(|r|) = \mathbb{R}(\boldsymbol{\theta}) \vee \mathbf{u}(0) \vee \mathbb{R}^{-1}(\boldsymbol{\theta}), \tag{7.3a}$$

$$\boldsymbol{\theta} = -(q|r|/m)\mathbf{B}. \tag{7.3b}$$

The angular velocity of the rotation is obtained as the time derivative of the angle by using (6.2) and the chain rule:

$$\boldsymbol{\Omega} = \frac{d\boldsymbol{\theta}}{dt} = -\frac{q\mathbf{B}}{\gamma m}. \tag{7.4}$$

This is a standard result[14,19,20] [note that Fig. 1–1(a) in Ref. 20 describes the motion of a *negatively* charged particle].

In the presence of both electric and magnetic fields, Eq. (7.2) gives a general description of the motion. It is instruc-

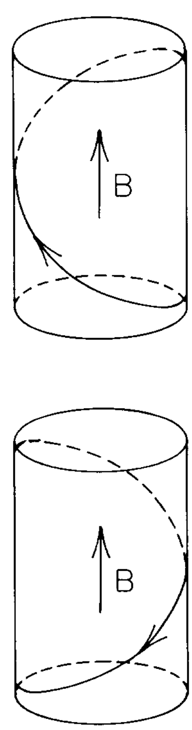


FIG. 2. Trajectory of positively charged particle in a magnetic field: (i) Initial velocity with component along magnetic field; (ii) initial velocity with component antiparallel to magnetic field.

tive to discuss the rotation and boost components of the canonical solution separately, even though they are acting simultaneously. To this effect, we separate the exponential operator in (7.2) into space and space-time components. We use the following identity, which is easily verified via the Baker–Campbell–Hausdorff formula[17,18]:

$$\exp(x+y)\approx\exp(x-\tfrac{1}{2}[x,y]+\tfrac{1}{6}[[x,y],y]) \times\exp(y+\tfrac{1}{12}[[x,y],x]). \tag{7.5}$$

With the identification $x=(q|r|/2m)\mathbf{E}\wedge\sigma^4$ and $y=-(q|r|/2m)\overset{*}{_3}\mathbf{B}$, we have the separation of the exponential operator in (7.2) into Lorentz boost and spatial rotation operators (4.1) and (4.5) as follows:

$$\exp\left(\frac{q|r|}{2m}f\right)=\mathbb{L}(-\mathbf{b})\vee\mathbb{R}(\boldsymbol{\theta}), \tag{7.6a}$$

$$\begin{cases}\mathbf{b}=\dfrac{q|r|}{m}\mathbf{E}+\dfrac{q^2|r|^2}{2m^2}\mathbf{E}\times\mathbf{B}+\dfrac{1}{6}\dfrac{q^3|r|^3}{m^3}(\mathbf{E}\times\mathbf{B})\times\mathbf{B}+\cdots, & (7.6b)\\ \boldsymbol{\theta}=-\dfrac{q|r|}{m}\mathbf{B}+\dfrac{1}{12}\dfrac{q^3|r|^3}{m^3}(\mathbf{E}\times\mathbf{B})\times\mathbf{E}+\cdots. & (7.6c)\end{cases}$$

We can evaluate the gyration frequency in the presence of an electric field; this is the ordinary time derivative of the rotational angle (7.6c) and is equal to (up to third order):

$$\boldsymbol{\Omega}\approx\frac{1}{\gamma}\left[\left(-\frac{q}{m}+\frac{q^3|r|^2}{4m^3}|\mathbf{E}|^2\right)\mathbf{B}-\frac{q^3|r|^2}{4m^3}(\mathbf{E}\cdot\mathbf{B})\mathbf{E}\right]. \tag{7.7}$$

We see that the magnitude of the frequency as well as the direction of the axis of rotation depend upon the line element $|r|$. This implies a continuously changing helical motion. In the absence of an electric field, expression (7.7) reduces to (7.4), as it should. The magnitude of the gyration frequency is given by (again, to third order):

$$|\boldsymbol{\Omega}|\approx\frac{q}{\gamma m}\left[|\mathbf{B}|^2+\left(\frac{q^4|r|^4|\mathbf{E}|^2}{16m^4}-\frac{q^2|r|^2}{2m^2}\right)|\mathbf{E}\times\mathbf{B}|^2\right]^{1/2}. \tag{7.8}$$

A distinct expression for $|\boldsymbol{\Omega}|$ is given in Refs. 21, 22, and 27 from a solution of the scalar coefficent equation (6.9b). This is a system of linear ordinary differential equations with constant coefficients, for which one assumes solutions of the form $\exp(\pm\nu|r|)$ and $\exp(\pm i\lambda|r|)$. We have instead solved Eq. (7.1), which is an algebraic-differential equation in the Lorentz group basis, and have, in contrast, obtained an infinite series in $|r|$ in the exponent. Using the congruent identity to (7.5),

$$\exp(x+y)\approx\exp(y+\tfrac{1}{12}[[x,y],x]) \times\exp(x+\tfrac{1}{2}[x,y]+\tfrac{1}{6}[[x,y],y]), \tag{7.9}$$

we can verify that the inverse operator to (7.6a) separates as

$$\exp[-(q|r|/2m)f]=\mathbb{R}^{-1}(\boldsymbol{\theta})\vee\mathbb{L}^{-1}(-\mathbf{b}). \tag{7.10}$$

Hence, the complete solution (7.2) becomes, with (7.6b) and (7.6c):

$$u(|r|)=\mathbb{L}(-\mathbf{b})\vee\mathbb{R}(\boldsymbol{\theta})\vee u(0)\vee\mathbb{R}^{-1}(\boldsymbol{\theta})\vee\mathbb{L}^{-1}(-\mathbf{b}). \tag{7.11}$$

Now, we can apply the transformation equations from Sec. 4 to write down (7.11) in vector form. Denote the initial velocity as

$$u(0)=(\mathbf{V}_0+\sigma^4)/\sqrt{1-|\mathbf{V}_0|^2}=\gamma_0(\mathbf{V}_0+\sigma^4). \tag{7.12}$$

The canonical solution (7.11) is written down in two steps. First, the rotation of the space part of the initial velocity $\mathbf{V}_0$ [(7.12)] is described by a vector $\mathbf{K}$, which is obtained from (7.2), (7.11), and (4.3) as a function of $\boldsymbol{\theta}$ and $\mathbf{V}_0$:

$$\mathbf{K}(\mathbf{V}_0,\boldsymbol{\theta})=\mathbf{V}_0\cos|\boldsymbol{\theta}|+[(\boldsymbol{\theta}\times\mathbf{V}_0)/|\boldsymbol{\theta}|]\sin|\boldsymbol{\theta}| +[\boldsymbol{\theta}(\boldsymbol{\theta}\cdot\mathbf{V}_0)/|\boldsymbol{\theta}|^2](1-\cos|\boldsymbol{\theta}|). \tag{7.13}$$

After the rotation (7.13), the initial velocity (7.12) becomes $\gamma_0(\mathbf{K}+\sigma^4)$. The other part of the motion is a Lorentz boost of this 4-vector. Using (7.12), (7.13), and (4.8) gives an expression which may be separated into space and time components using (6.3a) to obtain

$$\begin{aligned}\mathbf{u}(|r|)&=\gamma_0\{\mathbf{K}+(\mathbf{b}/|\mathbf{b}|)(\sinh|\mathbf{b}| +[(\mathbf{b}\cdot\mathbf{K})/|\mathbf{b}|](\cosh|\mathbf{b}|-1))\},\\ \gamma(|r|)&=\gamma_0\{\cosh|\mathbf{b}|+[(\mathbf{b}\cdot\mathbf{K})/|\mathbf{b}|]\sinh|\mathbf{b}|\}.\end{aligned} \tag{7.14}$$

This expression describes the velocity of a charged particle as a function of the initial velocity, the line element $|r|$, and the external electric and magnetic fields, via (7.6b), (7.6c), (7.12), and (7.13). The boost vector $\mathbf{b}$ and the gyration angle $\boldsymbol{\theta}$ are in general given as an infinite series in the electric and magnetic fields (7.6). The only case when the series is finite is when $\mathbf{E}$ is parallel to $\mathbf{B}$: then, $\mathbf{E}\times\mathbf{B}=0$ and all the higher-order terms in (7.6) which necessarily contain $\mathbf{E}\times\mathbf{B}$ are automatically zero. In that case, Eq. (7.14) is an exact solution of the equation of motion.

An expression may be obtained directly from the canonical solution (7.11) in the case of a particle which is initially at rest. The initial relativistic velocity (7.12) is then equal to $u(0)=\sigma^4$, and the velocity is given by

$$u(|r|)=\mathbb{L}(-\mathbf{b})\vee\mathbb{R}(\boldsymbol{\theta})\vee\sigma^4\vee\mathbb{R}^{-1}(\boldsymbol{\theta})\vee\mathbb{L}^{-1}(-\mathbf{b}). \tag{7.15}$$

The commutation rules (3.4) applied to the rotation and Lorentz boost operators (4.1) and (4.5) give the following expression:

$$u(|r|)=\mathbb{L}(-2\mathbf{b})\vee\sigma^4=\exp(\mathbf{b}\wedge\sigma^4)\vee\sigma^4. \tag{7.16}$$

Hence, the velocity of a particle initially at rest in a constant, homogeneous electromagnetic field is given in terms of the boost vector $\mathbf{b}$ (7.6b) [cf. Appendix, Eq. (A1)]. Using identity (3.3c) gives

$$\begin{aligned}&\mathbf{u}=(\mathbf{b}/|\mathbf{b}|)\sinh|\mathbf{b}|,\quad \mathbf{V}=\mathbf{u}/\gamma=(\mathbf{b}/|\mathbf{b}|)\tanh|\mathbf{b}|,\\ &u^4=\gamma=\cosh|\mathbf{b}|.\end{aligned} \tag{7.17}$$

We should point out that even though (7.17) is written down in closed form, it is not an exact expression, since $\mathbf{b}$ is in general given as a series (7.6b). Expression (7.17) can alternately be obtained by setting $\mathbf{K}=0$ and $\gamma_0=1$ in the general velocity (7.14). It is instructive to write down the velocity (7.17) in polynomial form. The velocity components are functions of the line element $|r|=\sqrt{t^2-(\mathbf{r}\cdot\mathbf{r})}$, and the electric and magnetic field components, and are given as follows:

$$\mathbf{u}(|r|)\approx(q|r|/m)\mathbf{E}+(q^2|r|^2/2m^2)\mathbf{E}\times\mathbf{B} +\tfrac{1}{6}(q^3|r|^3/m^3)[(|\mathbf{E}|^2-|\mathbf{B}|^2)\mathbf{E}+(\mathbf{E}\cdot\mathbf{B})\mathbf{B}], \tag{7.18a}$$

$$\gamma=u^4(|r|)\approx1+(q^2|r|^2/2m^2)|\mathbf{E}|^2 +(q^4|r|^4/24m^4)(|\mathbf{E}|^4-|\mathbf{E}\times\mathbf{B}|^2). \tag{7.18b}$$

The first term in the velocity (7.18a) is due to the Lorentz force; the second term is the "drift velocity," which, since it involves q^2, is independent of the sign of the charge.[14,19,20] Expression (7.18) is given to second order in Ref. 20. The third and fourth order terms are, to our knowledge, new.

8. THE THOMAS PRECESSION

In this section, we present a simple derivation of the Thomas precession, which is a well-known consequence of relativity. In contrast to the usual derivations given in textbooks, we do not resort to an infinitesimal description, but derive a finite result. The generality of this result allows the calculation of a higher-order term for the first time. This new term is a very small correction to the boost, and is not a correction to the rotational (precessional) motion. We emphasize that these are not actual physical effects at all, yet are genuinely perceived by an observer in another frame of reference.

We essentially wish to evaluate the net result of two consecutive Lorentz boosts in different directions. This can be done as follows: Consider a physical system described by a tensor type α' that is moving with a velocity $\mathbf{V}$ with respect to an observer. The observed tensor type α in the laboratory system is described by an inverse Lorentz transformation, obtained from (4.7) as

$$\alpha = \mathbb{L}^{-1}(\mathbf{b}) \vee \alpha' \vee \mathbb{L}(\mathbf{b}), \tag{8.1a}$$

$$\tanh|\mathbf{b}| = |\mathbf{V}|, \quad \cosh|\mathbf{b}| = (1-|\mathbf{V}|^2)^{-1/2} = \gamma(\mathbf{V}). \tag{8.1b}$$

Now suppose that the first observer is moving with a velocity $\mathbf{W}$ with respect to a second observer. How does the physical tensor type α appear to the second observer? This is a result of two consecutive Lorentz transformations in the different directions $\mathbf{V}$ and $\mathbf{W}$, described as follows:

$$\alpha_{(\text{observed})} = \mathbb{L}^{-1}(\mathbf{d}) \vee [\mathbb{L}^{-1}(\mathbf{b}) \vee \alpha' \vee \mathbb{L}(\mathbf{b})] \vee \mathbb{L}(\mathbf{d}), \tag{8.2a}$$

$$\tanh|\mathbf{d}| = |\mathbf{W}|, \quad \cosh|\mathbf{d}| = (1-|\mathbf{W}|^2)^{-1/2} = \gamma(\mathbf{W}). \tag{8.2b}$$

The physics of the system $\alpha_{(\text{observed})}$ should be the result of a Lorentz transformation by the combined boost $\mathbf{b}+\mathbf{d}$. A correction term which corresponds to a rotation in three-dimensional space arises naturally from the algebraic structure, as can be seen below. First observe that the combination of the two boosts (8.2a) can be written using associativity as

$$\alpha_{(\text{observed})} = [\mathbb{L}^{-1}(\mathbf{d}) \vee \mathbb{L}^{-1}(\mathbf{b})] \vee \alpha' \vee [\mathbb{L}(\mathbf{b}) \vee \mathbb{L}(\mathbf{d})]. \tag{8.3}$$

The net Lorentz boost operator corresponds to the explicit expression:

$$\mathbb{L}^{-1}(\mathbf{d}) \vee \mathbb{L}^{-1}(\mathbf{b}) = \exp(\tfrac{1}{2}\mathbf{d} \wedge \sigma^4) \vee \exp(\tfrac{1}{2}\mathbf{b} \wedge \sigma^4). \tag{8.4}$$

We now apply another identity derived from the Baker–Campbell–Hausdorff formula to evaluate (8.4). In a procedure akin to that in the previous section, we combine the commutator terms in the following manner:

$$\exp(x)\exp(y) \approx \exp(x+y+\tfrac{1}{6}[[x,y],x] + \tfrac{1}{3}[[x,y],y]) \cdot \exp(\tfrac{1}{2}[x,y]). \tag{8.5}$$

Substituting $x = \frac{1}{2}\mathbf{d} \wedge \sigma^4$ and $y = \frac{1}{2}\mathbf{b} \wedge \sigma^4$ in (8.5) gives, to second order,

$$\mathbb{L}^{-1}(\mathbf{d}) \vee \mathbb{L}^{-1}(\mathbf{b}) \approx \exp[\tfrac{1}{2}(\mathbf{d}+\mathbf{b}) \wedge \sigma^4] \vee \exp(\tfrac{1}{8}[\mathbf{d} \wedge \sigma^4, \mathbf{b} \wedge \sigma^4]). \tag{8.6}$$

The correction term is easily evaluated from Table I to be

$$[\mathbf{d} \wedge \sigma^4, \mathbf{b} \wedge \sigma^4] = -2\, {}^{*}_{3}(\mathbf{d} \times \mathbf{b}). \tag{8.7}$$

From (4.1) and (4.2) we see that this correction term describes a spatial rotation in the three-dimensional subspace of space-time. The derivation is now complete, and the net result of two consecutive Lorentz transformations is shown to be a boost by the sum of the two separate boosts, along with a spatial rotation in the plane defined by the two boosts. From (8.6), (8.7), (4.1), and (4.5) we have, to second order

$$\mathbb{L}^{-1}(\mathbf{d}) \vee \mathbb{L}^{-1}(\mathbf{b}) \approx \mathbb{L}^{-1}(\mathbf{d}+\mathbf{b}) \vee \mathbb{R}^{-1}(\tfrac{1}{2}\mathbf{d} \times \mathbf{b}). \tag{8.8}$$

The rotation part of (8.8) is important because in the limit of nonrelativistic velocities, $\mathbf{b} \approx \mathbf{V}$ and $\mathbf{d} \approx \mathbf{W}$, the angle of rotation is given by

$$\boldsymbol{\theta} \approx \tfrac{1}{2}\mathbf{W} \times \mathbf{V} \quad \text{(nonrelativistic)}. \tag{8.9}$$

The Thomas precession is the result of observing a system moving with a velocity $\mathbf{V}$, which is continually changing in direction. At each point in time, the system moving with velocity $\mathbf{V}$ is further boosted by the velocity $\mathbf{W} = \dot{\mathbf{V}}t$ where $\dot{\mathbf{V}} = d\mathbf{V}/dt$. The net result is a boost by the sum of the two velocities, combined with a time-dependent rotation. The (nonrelativistic) angle of the rotation is given by (8.9) as $\boldsymbol{\theta} \approx \frac{1}{2}\dot{\mathbf{V}} \times \mathbf{V}t$. If one assumes a constant rate of rotation, this angle is equal to $\boldsymbol{\Omega}t$, where $\boldsymbol{\Omega}$ is called the Thomas precession[15,16]:

$$\boldsymbol{\Omega} = \tfrac{1}{2}\dot{\mathbf{V}} \times \mathbf{V}. \tag{8.10}$$

A related, but distinct derivation is given in Refs. 28 and 29.

Note that the rotation effect in (8.8) is correct to third order, since the operator expansion of (8.5) gives not an additional correction to the rotation, but a small correction to the boost. We can calculate the next-order corrections to (8.4) as exponentials of the two terms $[[x,y],x]$ and $[[x,y],y]$, as follows:

$$\begin{aligned} [[x,y],x] &= \tfrac{1}{2}[(\mathbf{b} \times \mathbf{d}) \times \mathbf{d}] \wedge \sigma^4, \\ [[x,y],y] &= \tfrac{1}{2}[(\mathbf{b} \times \mathbf{d}) \times \mathbf{b}] \wedge \sigma^4. \end{aligned} \tag{8.11}$$

Following (8.5), the net Lorentz boost of a system which is a result of two successive boosts in different directions is obtained from (8.11) and (4.5) as

$$\mathbb{L}^{-1}(\mathbf{b}+\mathbf{d}+\tfrac{1}{6}[2|\mathbf{b}|^2\mathbf{d} - |\mathbf{d}|^2\mathbf{b} + (\mathbf{b}\cdot\mathbf{d})(\mathbf{d}-2\mathbf{b})]). \tag{8.12}$$

The third-order finite boost correction in (8.12) appears here explicitly for the first time. It is interesting that this expression is not symmetrical between the two consecutive boosts $\mathbf{b}$ and $\mathbf{d}$. This concludes our derivation of the Thomas precession.

9. DISCUSSION

In this paper we have constructed an intrinsic algebraic setting for the Lorentz group using differential forms and the

Clifford algebra. We then used this formalism to perform spatial rotations and Lorentz transformations. As a practical method, it is easier to use and is more general than the traditional methods. Contrast, for example, the description of rotations in three dimensions in terms of Euler angles or the Cayley–Klein parameters given in Ref. 25. Also, contrast the usual treatments of Lorentz transformations for motion along only one particular coordinate axis with our general description.

The Thomas precession effect is usually either omitted from discussions of Lorentz transformations, or is given a somewhat disconnected treatment, always in terms of infinitesimal parameters.[15,16] We were able to derive the Thomas precession as a direct consequence of the Lorentz group structure. Furthermore, we obtained a finite description, enabling us to calculate all the previously overlooked correction terms. The term which is next higher in order after the usual Thomas precession term was shown to be a boost correction, and not a correction to the rotational motion.

Our key contribution in this paper is that we have treated the Lorentz group as a finite Lie transformation group. By this, we mean manipulating the exponentials which are the elements of the Lie group, as opposed to just the infinitesimal generators, which are the elements of the Lie algebra.[18] The formalism obtained by combining the Clifford algebra with the differential form basis enabled us to realize the exponentials in closed form, and to use them to calculate results of physical interest. All our results followed essentially from evaluating combinations of exponentials using the Baker–Campbell–Hausdorff formula. The utilization of a global instead of a local description of the Lorentz group in physical applications gives results of greater generality.

As a consequence of the above formulation, we were able to write down the solution to the canonical system of equations $d\alpha/dt = [\beta,\alpha]$, where α and β are tensor fields in the Clifford algebra. This equation includes the rotation, Heisenberg, and Lorentz force equations as special cases. Three cases were solved explicitly by this method: (i) a rotating system in space, (ii) the precession of a spin in a homogeneous magnetic field, and (iii) the motion of a charged particle in a constant homogeneous electromagnetic field.

This paper supplied a mathematical framework in which to solve case (iii) in general. Our results are in agreement with the relativistic velocity of a charged particle in a constant, homogeneous electromagnetic field given to second order in Refs. 19 and 20. We were able to include all the higher-order terms in a straightforward manner; the closed-form expressions were given in this paper, and some special cases were exhibited to fourth order.

We also obtained an expression for the gyration frequency of a particle's orbit, which, however, differs from the usual one.[21,22,27] The reason for this discrepancy is that the usual derivation assumes a solution which is linear in the exponent. We, in contrast, made no such assumption, and were forced by the Lorentz group structure into a power series in the exponent. Therefore, while the usual solution is a perfectly valid one, it does not appear to be the most general solution of the problem which takes into account the Lorentz group and the geometry of the Minkowski space.

Our method is strictly distinct from the usual "guiding center" or "adiabatic" approximation to the motion of a particle in an electromagnetic field.[19,21,22] Nevertheless, the motivation of separating the gyrational motion of the particle from the total motion is shared by both methods. We believe that the results obtained in this paper indicate the applicability of this mathematical formalism to the description of physics in four-dimensional space-time.

APPENDIX: DERIVATION OF THE RELATIVISTIC PARAMETERS

For completeness, we review some elementary results in order to show how the relativistic parameters are obtained in the algebraic framework of this paper. The velocity of a particle in its own rest frame is equal to σ^4. To an observer, its relativistic velocity u is just the inverse Lorentz transformation of σ^4 to his frame, described by (4.5), (4.7), and (3.4c):

$$u = \mathbb{L}^{-1}(\mathbf{b}) \vee \sigma^4 \vee \mathbb{L}(\mathbf{b}) = \mathbb{L}(-2\mathbf{b}) \vee \sigma^4. \qquad \text{(A1)}$$

Using (3.4a), (3.4c) and the identity $\sigma^4 \vee \sigma^4 = 1$, we verify the unit norm of the 4-velocity u:

$$\begin{aligned} u \vee u &= \mathbb{L}(-2\mathbf{b}) \vee \sigma^4 \vee \mathbb{L}(-2\mathbf{b}) \vee \sigma^4 \\ &= \mathbb{L}(-2\mathbf{b}) \vee \mathbb{L}(2\mathbf{b}) = 1. \end{aligned} \qquad \text{(A2)}$$

The relativistic velocity is alternately defined as the derivative of the position 4-vector r with respect to the line element $|r|$, with $\gamma = dt/d|r|$:

$$u = \frac{dr}{d|r|} = \frac{d\mathbf{r}}{dt}\frac{dt}{d|r|} + \frac{dt}{d|r|}\sigma^4 = \gamma(\mathbf{V} + \sigma^4). \qquad \text{(A3)}$$

The usual formula for γ is obtained from (A2), (A3), and $\vee$ algebra, as follows:

$$u \vee u = \gamma(\mathbf{V} + \sigma^4) \vee \gamma(\mathbf{V} + \sigma^4) = \gamma^2(1 - |\mathbf{V}|^2) = 1. \qquad \text{(A4)}$$

Finally, the parametrization of the Lorentz boost in terms of hyperbolic functions is obtained from (A1). Using (3.3c), (4.5), (A1), and (A3), we have

$$\begin{aligned} u &= \{\cosh|\mathbf{b}| + [(\mathbf{b} \wedge \sigma^4)/|\mathbf{b}|]\sinh|\mathbf{b}|\} \vee \sigma^4 \\ &= \sigma^4 \cosh|\mathbf{b}| + (\mathbf{b}/|\mathbf{b}|)\sinh|\mathbf{b}| \\ &= \gamma(\mathbf{V} + \sigma^4) \\ &\Rightarrow \gamma = \cosh|\mathbf{b}|, \quad \gamma\mathbf{V} = (\mathbf{b}/|\mathbf{b}|)\sinh|\mathbf{b}|. \end{aligned} \qquad \text{(A5)}$$

[1] H. Flanders, *Differential Forms, with Applications to the Physical Sciences* (Academic, New York, 1963).
[2] C. Chevalley, *The Algebraic Theory of Spinors* (Columbia U. P., New York, 1954).
[3] M. F. Atiyah, R. Bott, and A. Shapiro, "Clifford modules," Topology **3** (Suppl. 1), 3–38 (1964).
[4] N. Jacobson, *Basic Algebra* (Freeman, San Francisco, 1980), Vol. 2.
[5] N. Salingaros, "Realization, extension, and classification of certain physically important groups and algebras," J. Math. Phys. **22**, 226–232 (1981).
[6] N. Salingaros, "On the classification of Clifford algebras and their relation to spinors in *n* dimensions," J. Math. Phys. **23**, 1–7, 1231 (1982).
[7] N. Salingaros, "Electromagnetism and the holomorphic properties of spacetime," J. Math. Phys. **22**, 1919–1925 (1981).
[8] N. Salingaros and M. Dresden, "Physical algebras in four dimensions I. The Clifford algebra in Minkowski spacetime," Adv. Appl. Math. **4**, 1–30 (1983).
[9] E. P. Wigner, "On unitary representations of the inhomogeneous Lorentz group," Ann. Math. **40**, 149–204 (1939).
[10] V. Bargmann and E. P. Wigner, "Group theoretical discussion of relativistic wave equations," Proc. Natl. Acad. Sci. USA **34**, 211–23 (1948).

[11]S. Weinberg, "Feynman rules for any spin," Phys. Rev. **133**, B 1318–32 (1964).
[12]H. Boerner, *Representations of Groups* (North-Holland, Amsterdam, 1967).
[13]A. Kyrala, *Theoretical Physics* (Saunders, Philadelphia, 1967).
[14]L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Pergamon, Oxford, 1971), 3rd ed.
[15]C. Møller, *The Theory of Relativity* (Oxford U. P., London, 1972), 2nd ed.
[16]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[17]W. Magnus, "On the exponential solution of differential equations for a linear operator," Comm. Pure Appl. Math. **7**, 649–73 (1954).
[18]V. S. Varadarajan, *Lie Groups, Lie Algebras, and Their Representations* (Prentice-Hall, Englewood Cliffs, NJ, 1974).
[19]H. Alfvén and C. G. Fälthammar, *Cosmical Electrodynamics* (Oxford U. P., London, 1963), 2nd ed.
[20]C. Itzykson and J. B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980).
[21]P. O. Vandervoort, "The relativistic motion of a charged particle in an inhomogeneous electromagnetic field," Ann. Phys. (N.Y.) **10**, 401–53 (1960).
[22]T. G. Northrop, *The Adiabatic Motion of Charged Particles* (Interscience, New York, 1963).
[23]D. Hestenes, "Proper particle mechanics," J. Math. Phys. **15**, 1768–77 (1974).
[24]D. Hestenes, "Proper dynamics of a rigid point particle," J. Math. Phys. **15**, 1778–86 (1974).
[25]H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1980), 2nd ed.
[26]G. Baym, *Lectures in Quantum Mechanics* (Benjamin, New York, 1969).
[27]G. Hellwig, "Über die Bewegung geladener Teilchen in schwach veränderlichen Magnetfeldern," Z. Naturforsch. **10a**, 508–16 (1955).
[28]D. Shelupsky, "Derivation of the Thomas precession formula," Am. J. Phys. **35**, 650–1 (1967).
[29]R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974).

Theory of nonlocal piezoelectricity

A. Cemal Eringen
Princeton University, Princeton, New Jersey 08544

(Received 20 July 1982; accepted for publication 4 March 1983)

Constitutive equations are derived for nonlinear and linear, nonlocal piezoelectric elastic solids. Restrictions arising from the second law of thermodynamics are determined. The theory of nonlocal em elastic solids is introduced and applied to the discussion of Debye screening of an electron in an elastic solid with defect and dispersion of optical and piezoelectric waves.

PACS numbers: 77.60. + v, 05.70. − a, 78.90. + t

1. INTRODUCTION

According to Maxwell's electromagnetic theory, plane waves are nondispersive. Therefore, the refractive index in an isotropic nondissipative medium is constant. Yet experiments show that the refractive index depends on frequency and wavelength. Consequently, the *dispersion* is the rule rather than the exception. Electromagnetic dispersive phenomena are felt strongly in the high frequency region. For most media, as the frequency approaches 10^{14} Hz, the electron clouds lag in the adaptation to the electric field. In the region between microwaves and infrared frequencies, atomic distortions cause dispersion.

At transition frequencies to exciton states in a crystal, the dependence of exciton energy on the wave vector cannot be neglected. In this case, corresponding resonant frequencies depend on the wave number. Even for the static fields of stationary electrons, we have the Debye screening resulting from nonlocal interactions.

In plasma physics, at low temperature, it has been observed that the effect of the spatial dispersion on electromagnetic properties of metals is considerable.

In classical electromagnetism, to take care of the frequency dependence of the refractive index, usually, excursions are made to spring and dashpot models.[1] In this way, an admixture of continuum and atomic models are brought together. While this curve-fitting process gives satisfactory results for the refractive index in various frequency ranges, it is not based on a fundamental theory which can explain other physical phenomena without further modifications. More satisfactory quantum mechanical approaches, on the other hand, present major mathematical difficulties.

In the prediction of magnetic phenomena, similar situations are encountered. For example, the effects of magnetic domains and spin waves cannot be explained by means of classical electromagnetism. Much of the published work in this area makes use of the ideas of inner structure and domains that exist in materials, either in the form of multipoles, microstructures, or atomic structure (cf. Brown,[2] Maugin and Eringen,[3] Kittel,[4] and Bloch[5]).

There exists a large literature on the subject of wave-number-dependent dielectric functions which is based on quantum and statistical mechanical consideration (Pines,[6] Lindhard,[7] Ehrenreich and Cohen,[8] Adler,[9] Penn,[10] and Wiser[11]). These works are concerned only with the dielectric function of nondeformable solids. A survey on the electrodynamics of media with spatial dispersion was also published by Rukhadze and Silin.[12] This study is also concerned with the linear theory of nondeformable media, and it has some contacts with the present approach for the case of *rigid* solids. However, thermodynamical considerations were not studied. Moreover, the Fourier domain formalism used there fail to apply for nonlinear theory.

Recently, we gave a general theory of nonlocal, nonlinear electromagnetic theory of elastic solids.[13] Specific constitutive equations were not developed, however, to discuss piezoelectricity and piezomagnetism. In the case of non-heat- and electric-conducting materials, the theory can be simplified a great deal.

Although nonlinear theory is difficult to deal with, it should have significant applications on surface phenomena and in phase transition. Linear theory, however, has many practical applications and can be used to study a large class of phenomena in the molecular scale. In other works,[14,15] we have shown that, by means of nonlocal elasticity theory, the dispersion of elastic waves can be predicted in the entire Brillouin zone. Moreover, nonlocality eliminates unphysical stress singularity at crack tips[16,17] so that a natural fracture criterion based on the cohesive stress can be used to predict the crack instability. The present paper is intended for the construction of a theory of nonlocal piezoelectricity which has similar possibilities for waves and for electromagnetic singularities.

While any theory involving electromagnetic interactions should be relativistic, it is possible to construct a rational theory on nonrelativistic grounds for small material velocities v as compared to the speed of light c in a vacuum ($v^2/c^2 \ll 1$).

In Sec. 2, we present local balance laws which were obtained before.[13] The second law of thermodynamics essential to our development is presented in Sec. 3. In Sec. 4, we begin the development of the nonlocal constitutive theory. Section 5 employs a special representation for the response functionals. The linear theory is presented in Sec. 6. Constitutive equations of isotropic, uniaxial, and anisotropic elastic dielectrics are the subject of Sec. 7.

With Sec. 8, we begin the treatment of some problems. The Debye screening of the field of a stationary electron in an elastic solid with defect is obtained in Sec. 8. Section 9 discusses dispersion of optical modes, and Sec. 10, piezoelectric waves. It is shown that nonlocal theory leads to dispersive waves, and there is no necessity for recourse to any discrete spring models. In fact, once the nonlocal material moduli are determined, the theory can be used to solve prob-

 0022-2488/84/030717-11$02.50

lems without any further adjustment.

The theory in the long-wave limit naturally converts to the classical theory and, in the short wavelength limit, it can account for the local distortions and dispersions arising from the discrete nature of the atomic structure. In the last section, we make some remarks on the subject of memory-dependent materials.

2. BALANCE LAWS

The body at the natural state occupies a region $V-\Sigma$, the volume V excluding a discontinuity surface Σ. The motion carries a material point $\mathbf{X}\in V-\Sigma$ to a spatial place $\mathbf{x}\in\mathscr{V}-\sigma$, where $\mathscr{V}-\sigma$ is the image of $V-\Sigma$ at time t. The motion is a bijective mapping expressed by

$$\mathbf{x}=\mathbf{x}(\mathbf{X},t)\leftrightarrow\mathbf{X}=\mathbf{X}(\mathbf{x},t). \tag{2.1}$$

We employ a rectangular frame of reference so that rectangular coordinates of $\mathbf{x}$ and $\mathbf{X}$ are denoted by x_k and X_K, respectively ($k, K=1,2,3$). Since (2.1) is bijective, the Jacobian must be positive,

$$J=\det(x_{k,K})>0. \tag{2.2}$$

Henceforth, we employ a comma to denote partial derivative and a dot to express the material derivative. The usual summation convention on repeated indices is also assumed, e.g.,

$$x_{k,K}=\frac{\partial x_k}{\partial X_K},\quad \dot{x}_k=\left.\frac{\partial x_k}{\partial t}\right|_{\mathbf{X}}=v_k(\mathbf{x},t),$$
$$a_k=\dot{v}_k=\frac{\partial v_k}{\partial t}+v_{k,l}v_l. \tag{2.3}$$

Elastic dielectrics are nonconductors and carry no free charge. Therefore, for piezoelectric solids, electromagnetic (em) balance laws are identical to those of the classical (local) theory, since the free-charge density, current, and the pole strength (local and nonlocal) vanish. We have, therefore, Maxwell's equations, expressed in Lorentz–Heaviside units, in $\mathscr{V}-\sigma$:

$$\nabla\cdot\mathbf{D}=0, \tag{2.4}$$

$$\nabla\times\mathbf{E}+\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t}=\mathbf{0}, \tag{2.5}$$

$$\nabla\cdot\mathbf{B}=0, \tag{2.6}$$

$$\nabla\times\mathbf{H}-\frac{1}{c}\frac{\partial\mathbf{D}}{\partial t}=\mathbf{0}, \tag{2.7}$$

where $\mathbf{D}$, $\mathbf{E}$, $\mathbf{B}$, and $\mathbf{H}$ are, respectively, the electric displacement vector, electric vector, magnetic induction vector, and magnetic field vector. c is the speed of light in vacuum.

Maxwell's equations are supplemented with the mechanical balance laws. We assume that the body is inert, and gravitation force and couple residuals are negligible. Consequently, mechanical balance laws (conservation of mass, balance of momenta and energy) are expressed as[13,18]

$$\rho_0/\rho=\det(x_{k,K}), \tag{2.8}$$

$$t_{kl,k}+\rho(f_l-\dot{v}_l)+{}_Mf_l=0, \tag{2.9}$$

$$t_{kl}+P_k\mathscr{E}_l+\mathscr{M}_kB_l=t_{lk}+P_l\mathscr{E}_k+\mathscr{M}_lB_k\equiv{}_Et_{kl}, \tag{2.10}$$

$$\rho\dot{\epsilon}-t_{kl}v_{l,k}-q_{k,k}-\rho h-\rho\mathscr{E}\cdot(\mathbf{P}/\rho)^{\boldsymbol{\cdot}}+\mathscr{M}\cdot\dot{\mathbf{B}}=0, \tag{2.11}$$

where ρ_0 is the mass density in $V-\Sigma$ and ρ, t_{kl}, f_l, v_l, ϵ, q_k, and h are, respectively, the mass density, stress tensor, body force density, velocity vector, internal energy density, heat vector, and the heat source in $\mathscr{V}-\sigma$. $\mathscr{E}$ and $\mathscr{M}$ are the electric and magnetization vectors in the *proper* (comoving) frame as defined by

$$\mathscr{E}=\mathbf{E}+(1/c)\mathbf{v}\times\mathbf{B},\quad \mathscr{M}=\mathbf{M}+(1/c)\mathbf{v}\times\mathbf{P}. \tag{2.12}$$

Here, $\mathbf{M}$ and $\mathbf{P}$ are, respectively, the magnetization and polarization vectors in the fixed frame so that

$$\mathbf{D}=\mathbf{E}+\mathbf{P},\quad \mathbf{B}=\mathbf{H}+\mathbf{M}. \tag{2.13}$$

The em body force ${}_M\mathbf{f}$, in the absence of the charge and current, is given by[18]

$${}_M\mathbf{f}=(\nabla\mathbf{E})\cdot\mathbf{P}+(\nabla\mathbf{B})\cdot\mathbf{M}+\frac{1}{c}\left[(\mathbf{P}\times\mathbf{B})v_k\right]_{,k}$$
$$+\frac{1}{c}\frac{\partial}{\partial t}(\mathbf{P}\times\mathbf{B}). \tag{2.14}$$

Accompanying Maxwell's equations and mechanical balance laws, we have the *jump* conditions across σ. These conditions give boundary conditions when σ is made to coincide with the surface of the body. For brevity, we do not list these conditions here. They can be found in Ref. 18, Sec. 10.17.

3. SECOND LAW OF THERMODYNAMICS

In classical field theories (local continuum theories), the local form of the entropy inequality plays a central role. The second law of thermodynamics is a statement about the dissipative process that takes place in the entire body. The localization used in continuum physics is a reinterpretation of this law, which produces severe restrictions to the thermodynamic process. This fact has come to the surface clearly in the case of contemporary mixture theories, where often the mixture law is used instead of the entropy production law for each species. While this question still remains open, it is clear that the global entropy production must be nonnegative in any case. It is also less restrictive, allowing possible entropy exchanges among various points of the body.

For chemically inert bodies, the second law of thermodynamics can be expressed as

$$\frac{d}{dt}\int_{\mathscr{V}-\sigma}\rho\eta\,dv-\int_{\partial\mathscr{V}-\sigma}\frac{1}{\theta}\mathbf{q}\cdot\mathbf{n}\,da-\int_{\mathscr{V}-\sigma}\frac{\rho h}{\theta}\,dv\geqslant0, \tag{3.1}$$

where η is the entropy density and $\theta>0$ is the absolute temperature. From this, by means of Green–Gauss and the transport theorems, we derive[18]

$$\int_{\mathscr{V}-\sigma}\left[\rho\dot{\eta}-\nabla\cdot(\mathbf{q}/\theta)-(\rho h/\theta)\right]dv$$
$$+\int_\sigma \boldsymbol{[}\rho\eta(\mathbf{v}-\boldsymbol{\nu})-\mathbf{q}/\theta\boldsymbol{]}\cdot\mathbf{n}\,da\geqslant0, \tag{3.2}$$

where $\boldsymbol{\nu}$ is the velocity of σ and a boldface bracket is used to indicate the jump across σ. These expressions are equivalent to

$$\rho\dot{\eta} - \nabla\cdot(\mathbf{q}/\theta) - (\rho h/\theta) - \rho\hat{s} \geqslant 0, \quad \text{in } \mathscr{V} - \sigma, \tag{3.3}$$

$$[\rho\eta(\mathbf{v} - \nu) - \mathbf{q}/\theta]\cdot\mathbf{n} = \hat{N}, \quad \text{on } \sigma, \tag{3.3'}$$

where $\hat{s}$ and $\hat{N}$ are called *body* and *surface entropy residuals* which are subject to the restrictions

$$\int_{\mathscr{V} - \sigma} \rho\hat{s}\, dv = 0, \quad \int_\sigma \hat{N}\, da = 0. \tag{3.4}$$

If we eliminate h between (3.3) and (2.11), we obtain

$$-\frac{\rho}{\theta}(\dot{\psi} + \eta\dot{\theta}) + \frac{1}{\theta} t_{kl} v_{l,k} \frac{1}{\theta^2} q_k \theta_{,k}$$

$$+ \frac{\rho}{\theta}\mathscr{E}\cdot(\mathbf{P}/\rho)^\cdot - \frac{1}{\theta}\mathscr{M}\cdot\dot{\mathbf{B}} - \rho\hat{s} \geqslant 0, \tag{3.5}$$

where we introduced the Helmholtz free energy by

$$\psi = \epsilon - \theta\eta. \tag{3.6}$$

We now introduce various fields in the reference frame by

$$T_{KL} = JX_{K,k}X_{L,l}t_{kl}, \quad Q_K = JX_{K,k}q_k,$$

$$\Pi_K = JX_{K,k}P_k, \quad M_K = JX_{K,k}\mathscr{M}_K, \tag{3.7}$$

$$C_{KL} = x_{k,K}x_{k,L}, \quad \mathscr{E}_K = \mathscr{E}_k x_{k,K},$$

$$\theta_{,K} = \theta_{,k}x_{k,K}, \quad B_K = B_k x_{k,K}.$$

With these, (3.5) can be transformed into

$$-\rho_0(\dot{\Psi} + \eta\dot{\theta}) + \tfrac{1}{2}\,{}_E T_{KL}\dot{C}_{KL} + (1/\theta)Q_K\theta_{,K}$$

$$- \Pi_K\dot{\mathscr{E}}_K - M_K\dot{B}_K - \rho_0\theta\hat{s} \geqslant 0, \tag{3.8}$$

where we use $J = \rho_0/\rho$ and

$$\dot{C}_{KL} = (v_{k,l} + v_{l,k})x_{k,K}x_{l,L},$$

and set

$$\Psi = \psi - \rho_0^{-1}\Pi_K\mathscr{E}_K = \epsilon - \theta\eta - \rho_0^{-1}\Pi_K\mathscr{E}_K. \tag{3.9}$$

Since we assume that the heat and electric conduction are negligible, $Q_K = 0$, and θ is uniform throughout $V - \Sigma$, the volume integral of (3.8), upon using (3.4), gives

$$\int_{V-\Sigma} [-\rho_0(\dot{\Psi} + \eta\dot{\theta}) + \tfrac{1}{2}\,{}_E T_{KL}\dot{C}_{KL}$$

$$- \Pi_K\dot{\mathscr{E}}_K - M_K\dot{B}_K]\, dV \geqslant 0. \tag{3.10}$$

It is posited that the inequality (3.10) must not be violated for any thermodynamic process that is physically admissable. In Sec. 4, we employ this axiom to derive the constitutive equations of nonlocal piezoelectric solids.

4. CONSTITUTIVE EQUATIONS

In accordance with the axiom of causality,[18,19] all physical processes that take place in a body are the result of motions (deformations). When the intrinsic motions of subbodies in a volume element are taken into account, this implies centroidal motions of the volume element and dependence on temperature, polarization, and magnetization. Ignoring the memory dependence and conduction, this is equivalent to the selection of the independent constitutive variables.

$$\mathscr{Y}' \equiv \{\mathbf{x}(\mathbf{X}'), \mathscr{E}_K(\mathbf{X}'), B_K(\mathbf{X}'); \theta\}, \tag{4.1}$$

where $\mathbf{x}(\mathbf{X}')$, $\mathscr{E}_K(\mathbf{X}')$, $B_K(\mathbf{X}')$ represents the motions, electric fields, and magnetic inductions of *all* points $\mathbf{X}'$ of the body at time t. The dependence on t is suppressed for brevity. Similarly, when $\mathbf{X}'$ is taken to be the fixed reference point $\mathbf{X}$, we suppress it, e.g., we write $\theta \equiv \theta(\mathbf{X},t)$.

Constitutive equations express the functional dependence of the set

$$Z = \{\Psi, \eta, {}_E T_{KL}, \Pi_K, M_K\} \tag{4.2}$$

at a reference point $\mathbf{X}$ at time t to the set (4.1), e.g.,

$$\Psi(\mathbf{X},t) = \mathscr{F}[\mathbf{x}(\mathbf{X}'), \mathscr{E}(\mathbf{X}'), \mathbf{B}(\mathbf{X}'); \theta], \tag{4.3}$$

where $\mathscr{F}$ is a functional of the first three functions and a function of θ. Expressions of this type are written for all members of Z. Response functionals (such as $\mathscr{F}$) must be form-invariant under arbitrary spatial translations and rotation. This implies that $\mathscr{F}$ will depend on $\mathbf{x}(\mathbf{X}')$ only through $|\mathbf{x}(\mathbf{X}') - \mathbf{x}(\mathbf{X})|$. Since the distance can be expressed as a functional of C_{KL}, for the elastic bodies, it proves to be helpful to replace Eq. (4.3) by

$$\Psi(\mathbf{X},t) = \mathscr{F}[G(\mathbf{X}'); G, \theta], \tag{4.4}$$

where

$$G' \equiv G(\mathbf{X}') \equiv \{C_{KL}(\mathbf{X}'), \mathscr{E}_K(\mathbf{X}'), B_K(\mathbf{X}')\},$$

$$G \equiv \{C_{KL}, \mathscr{E}_K, B_K\}. \tag{4.5}$$

We assume that $G(\mathbf{X}')$ and G possess continuous partial derivatives with respect to their arguments.

In order to introduce a topology to the space of functions $G(\mathbf{X}')$, we define the inner product of two such sets by

$$(G'_1, G'_2)_H = \int_{V-\Sigma} H(|\mathbf{X}' - \mathbf{X}|)G_1(\mathbf{X}')\cdot G_2(\mathbf{X}')\, dV(\mathbf{X}'), \tag{4.6}$$

where

$$G_1(\mathbf{X}')\cdot G_2(\mathbf{X}') = \mathrm{tr}(\mathbf{C}'_1\mathbf{C}'_2) + \mathscr{E}'_1\cdot\mathscr{E}'_2 + \mathbf{B}'_1\cdot\mathbf{B}'_2 \tag{4.7}$$

and the *influence function* $H(|\mathbf{X}' - \mathbf{X}|)$ is a positive, decreasing function of $|\mathbf{X}' - \mathbf{X}|$ such that $H(0) = 1$. It emphasizes the influence of motions and em fields near the reference point $\mathbf{X}$ over the distant points from $\mathbf{X}$. This is in accordance with the *attenuating neighborhood* hypothesis[18,19] based on the nature of long-range intermolecular forces. The norm of the set G' is defined by

$$\|G'\cdot G'\| = (G', G')_H^{1/2}. \tag{4.8}$$

The space of functions G' is now a Hilbert space.

There exist many choices for the influence function. We mention two such functions as examples:

$$H(|\mathbf{X}' - \mathbf{X}|) = [1 + \alpha(\mathbf{X})|\mathbf{X}' - \mathbf{X}|]^{-1/2},$$

$$H(|\mathbf{X}' - \mathbf{X}|) = \exp[-\alpha(\mathbf{X})|\mathbf{X}' - \mathbf{X}|], \quad \alpha(\mathbf{X}) > 0. \tag{4.9}$$

It is now possible to calculate

$$\dot{\Psi} = \frac{\partial\Psi}{\partial\theta}\dot{\theta} + \frac{\partial\Psi}{\partial G}\dot{G} + \int_{V-\Sigma}\frac{\delta\Psi}{\delta G'}\dot{G}'\, dV', \tag{4.10}$$

where the term $\delta\Psi/\delta G'$ denotes the Fréchet partial derivative. We express (4.10) in the equivalent form

$$\dot{\Psi} = \frac{\partial\Psi}{\partial\theta}\dot{\theta} + \left[\frac{\partial\Psi}{\partial G} + \int_{V-\Sigma}\left(\frac{\delta\Psi}{\delta G'}\right)^* dV'\right]\dot{G} + \mathscr{D}, \tag{4.11}$$

where

$$\mathscr{D} \equiv \int_{V-\Sigma} \left[\frac{\delta \Psi}{\delta G'} \dot{G}' - \left(\frac{\delta \Psi}{\delta G'} \right)^{*} \dot{G} \right] dV' . \tag{4.12}$$

Here, an asterisk is used to indicate the interchange of $\mathbf{X}'$ and $\mathbf{X}$, i.e.,

$$\overset{*}{A}(\mathbf{X}',\mathbf{X}) = A(\mathbf{X},\mathbf{X}') . \tag{4.13}$$

Note that, because of antisymmetry of $\mathscr{D}$ in $\mathbf{X}$ and $\mathbf{X}'$, we have

$$\int_{V-\Sigma} \mathscr{D}\, dV = 0 . \tag{4.14}$$

Substituting (4.11) into (3.10) and using (4.5) and (4.14), we obtain

$$\begin{aligned} \int_{V-\Sigma} \Big\{ & \rho_0 \Big(\eta + \frac{\partial \Psi}{\partial \theta} \Big) \dot{\theta} \\ & + \frac{1}{2} \Big[{}_E T_{KL} - 2\rho_0 \frac{\partial \Psi}{\partial C_{KL}} - 2\rho_0 \int_{V-\Sigma} \Big(\frac{\delta \Psi}{\delta C'_{KL}} \Big)^{*} dV' \Big] \dot{C}_{KL} \\ & - \Big[\Pi_K + \rho_0 \frac{\partial \Psi}{\partial \mathscr{E}_K} + \rho_0 \int_{V-\Sigma} \Big(\frac{\partial \Psi}{\partial \mathscr{E}'_K} \Big)^{*} dV' \Big] \dot{\mathscr{E}}_K \\ & - \Big[M_K + \rho_0 \frac{\partial \Psi}{\partial B_K} + \rho_0 \int_{V-\Sigma} \Big(\frac{\delta \Psi}{\delta B'_K} \Big)^{*} dV' \Big] \dot{B}_K \Big\} dV \geqslant 0 . \end{aligned} \tag{4.15}$$

Since $\dot{\theta}$, $\dot{C}_{KL}$, $\dot{\mathscr{E}}_K$, and $\dot{B}_K$ can be varied independently and arbitrarily throughout $V-\Sigma$ from a theorem of calculus, it follows that inequality (4.15) will not be violated if and only if

$$\eta = -\frac{\partial \Psi}{\partial \theta} , \tag{4.16}$$

$${}_E T_{KL} = 2\rho_0 \frac{\partial \Psi}{\partial C_{KL}} + 2\rho_0 \int_{V-\Sigma} \left(\frac{\delta \Psi}{\delta C'_{KL}} \right)^{*} dV' , \tag{4.17}$$

$$\Pi_K = -\rho_0 \frac{\partial \Psi}{\partial \mathscr{E}_K} - \rho_0 \int_{V-\Sigma} \left(\frac{\delta \Psi}{\delta \mathscr{E}'_K} \right)^{*} dV' , \tag{4.18}$$

$$M_K = -\rho_0 \frac{\partial \Psi}{\partial B_K} - \rho_0 \int_{V-\Sigma} \left(\frac{\delta \Psi}{\delta B'_K} \right)^{*} dV' . \tag{4.19}$$

These are the nonlinear constitutive equations for nonlocal piezoelectric solids. The spatial forms of these equations follow from Eq. (3.7):

$$\eta = -\frac{\partial \Psi}{\partial \theta} , \tag{4.20}$$

$${}_E t_{kl} = \frac{\rho}{\rho_0} {}_E T_{KL} x_{k,K} x_{l,L} , \tag{4.21}$$

$$P_k = \frac{\rho}{\rho_0} \Pi_K x_{k,K} , \tag{4.22}$$

$$\mathscr{M}_k = \frac{\rho}{\rho_0} M_K x_{k,K} . \tag{4.23}$$

Since the entropy production vanishes, we conclude that nonconducting piezoelectric solids are in thermodynamic equilibrium. Equations (4.17) to (4.23) provide the source for approximate theories.

5. ADDITIVE FUNCTIONALS

For most materials, it is not necessary to consider a general functional, for the description of free energy. *Additive functionals*, in the sense of Friedman and Katz,[20] are adequate to characterize most piezoelectric substances. For such functionals, we have the representation

$$\rho_0 \Psi = \frac{1}{2} \int_{V-\Sigma} S(E'_{KL}, \mathscr{E}'_K, B'_K; E_{KL}, \mathscr{E}_K, B_K, \theta)\, dV' , \tag{5.1}$$

where we also introduced Lagrangian strain tensor E_{KL} by

$$2E_{KL} = C_{KL} - \delta_{KL} . \tag{5.2}$$

The total free energy of the body is given by

$$\int_{V-\Sigma} \Sigma\, dV = \frac{1}{2} \int_{V-\Sigma} \int_{V-\Sigma} S\, dV'\, dV . \tag{5.3}$$

From Eq. (5.3), it is clear that only the symmetric part of S in $\mathbf{X}$ and $\mathbf{X}'$ contributes to the total energy. Thus, we may select $S = \overset{*}{S}$ and we have

$$\eta = -\frac{1}{2\rho_0} \int_{V-\Sigma} \frac{\partial S}{\partial \theta} dV' , \tag{5.4}$$

$${}_E T_{KL} = \int_{V-\Sigma} \frac{\partial S}{\partial E_{KL}} dV' , \tag{5.5}$$

$$\Pi_K = -\int_{V-\Sigma} \frac{\partial S}{\partial \mathscr{E}_K} dV' , \tag{5.6}$$

$$M_K = -\int_{V-\Sigma} \frac{\partial S}{\partial B_K} dV' . \tag{5.7}$$

Polynomial constitutive equations of various degree may be derived from Eqs. (5.4)–(5.7) by expressing S as a polynomial in the vector and tensor variables.

6. LINEAR CONSTITUTIVE EQUATIONS

To obtain linear constitutive equations, we express S as a second-degree polynomial in the variables $(E_{KL}, \mathscr{E}_K, B_K)$ and $(E'_{KL}, \mathscr{E}'_K, B'_K)$:

$$S = S' + \overset{*}{S}{}' \tag{6.1}$$

$$\begin{aligned} S' = {} & \tfrac{1}{2}\Sigma^0 + \Sigma^0_{KL} E_{KL} + \tfrac{1}{2}\Sigma^0_{KLMN} E_{KL} E_{MN} \\ & + \tfrac{1}{2}\Sigma'_{KLMN} E_{KL} E'_{MN} - E^0_{KLM} \mathscr{E}_K E_{LM} \\ & - E'_{KLM} \mathscr{E}_K E'_{LM} - H^0_{KLM} B_K E_{LM} \\ & - H'_{KLM} B_K E'_{LM} - \chi^{0E}_K \mathscr{E}_K - \tfrac{1}{2}\chi^{0E}_{KL} \mathscr{E}_K \mathscr{E}_L \\ & - \tfrac{1}{2}\chi'^{E}_{KL} \mathscr{E}_K \mathscr{E}'_L - \chi^{0B}_K B_K - \tfrac{1}{2}\chi^{0B}_{KL} B_K B_L - \tfrac{1}{2}\chi'^{B}_{KL} B_K B'_L \\ & - \Lambda^0_{KL} \mathscr{E}_K B_L - \Lambda'_{KL} \mathscr{E}_K B'_L , \end{aligned} \tag{6.2}$$

where the material moduli $\Sigma^0, \Sigma^0_{KL}, \ldots, \Lambda'_{KL}$ are functions of θ, $\mathbf{X}$, and $\mathbf{X}'$ and they possess the following symmetry relations:

$$\Sigma^0 = \overset{*}{\Sigma}{}^0, \Sigma^0_{KL} = \Sigma^0_{LK},$$

$$\Sigma^0_{KLMN} = \Sigma^0_{LKMN} = \Sigma^0_{KLNM} = \Sigma^0_{MNKL},$$

$$E^0_{KLM} = E^0_{KML}, \quad H^0_{KLM} = H^0_{KML},$$

$$\chi^{0E}_{KL} = \chi^{0E}_{LK}, \quad \chi^{0B}_{KL} = \chi^{0B}_{LK}, \tag{6.3}$$

$$\Sigma'_{KLMN} = \Sigma'_{LKMN} = \Sigma'_{KLNM} = \overset{*}{\Sigma}{}'_{MNKL},$$

$$E'_{KLM} = E'_{KML} = \overset{*}{E}{}'_{KLM},$$

$$H_{KLM} = H'_{KML} = \overset{*}{H}{}'_{KLM}, \chi'^E_{KL} = \overset{*}{\chi}{}'^E_{LK},$$

$$\chi'^B_{KL} = \overset{*}{\chi}{}'^B_{LK}.$$

Substituting (6.1) into (5.4) to (5.7), we obtain

$$\eta = -\frac{1}{2\rho_0}\left(\frac{\partial \Sigma}{\partial \theta} + \frac{\partial \Sigma_{KL}}{\partial \theta} E_{KL} - \frac{\partial \chi^E}{\partial \theta} \mathscr{E}_K - \frac{\partial \chi^B_K}{\partial \theta} B_K\right)$$

$$-\frac{1}{2\rho_0}\int_{V-\Sigma}\left(\frac{\partial \overset{*}{\Sigma}{}^0_{KL}}{\partial \theta} E'_{KL} + \frac{\partial \overset{*}{\chi}{}^{0E}_K}{\partial \theta} \mathscr{E}'_K - \frac{\partial \overset{*}{\chi}{}^{0B}_K}{\partial \theta} B'_K\right) dV', \tag{6.4}$$

$$_E T_{KL} = \Sigma_{KL} + \Sigma_{KLMN} E_{MN} - E_{MKL} \mathscr{E}_M - H_{MKL} B_M$$
$$+ \int_{V-\Sigma} (\Sigma'_{KLMN} E'_{MN} - E'_{MKL} \mathscr{E}'_M - H'_{MKL} B'_M)\, dV', \tag{6.5}$$

$$\Pi_K = \chi^E_K + \chi^E_{KL} \mathscr{E}_L + E_{KLM} E_{LM} + \Lambda_{KL} B_L$$
$$+ \int_{V-\Sigma} (\chi'^E_{KL} \mathscr{E}'_L + E'_{KLM} E'_{LM} + \Lambda'_{KL} B'_L)\, dV', \tag{6.6}$$

$$M_K = \chi^B_K + \chi^B_{KL} B_L + H_{KLM} E_{LM} + \Lambda_{LK} \mathscr{E}_L$$
$$+ \int_{V-\Sigma} (\chi'^B_{KL} B'_L + H'_{KLM} E'_{LM} + \overset{*}{\Lambda}{}'_{LK} \mathscr{E}'_L)\, dV', \tag{6.7}$$

where we set

$$\{\Sigma, \Sigma_{KL}, \Sigma_{KLMN}, E_{MKL}, H_{MKL}, \chi^E_K, \chi^B_K, \chi^E_{KL}, \chi^B_{KL}, \Lambda_{KL}\}$$
$$= \int_{V-\Sigma} \{\Sigma^0, \Sigma^0_{KL}, \Sigma^0_{KLMN}, E^0_{MKL},$$
$$H^0_{MKL}, \chi^{0E}_K, \chi^{0B}_K, \chi^{0E}_{KL}, \chi^{0B}_{KL}, \Lambda^0_{KL}\}\, dV' \tag{6.8}$$

and, in Eq. (6.4), we dropped quadratic terms.

To obtain the spatial forms of these equations, we employ Eqs. (4.20) to (4.23) and use

$$E_{KL} = e_{kl} x_{k,K} x_{l,L}, \quad R_{KL} = r_{kl} x_{k,K} x_{l,L}, \tag{6.9}$$

$$\rho_0/\rho \simeq 1 - e_{rr}, \quad x_{k,K} = (\delta_{MK} + E_{MK} + R_{MK})\delta_{Mk},$$

where δ_{Mk} is the Kronecker delta, when the spatial and material frames are coincident, and e_{kl} and r_{kl} are, respectively, the linear strain and rotation measures which are defined, in terms of the spatial displacement vector u_k, by

$$e_{kl} = \tfrac{1}{2}(u_{k,l} + u_{l,k}), \quad r_{kl} = \tfrac{1}{2}(u_{k,l} - u_{l,k}). \tag{6.10}$$

We also introduce spatial material moduli by

$$\Sigma'_{KL} = \sigma'_{kl}\delta_{kK}\delta_{lL}, \quad \Sigma'_{KLMN} = \sigma'_{klmn}\delta_{kK}\delta_{lL}\delta_{mM}\delta_{nN}, \tag{6.11}$$
$$E'_{KLM} = e'_{klm}\delta_{kK}\delta_{lL}\delta_{mM}, \quad \chi'^E_K = \chi'^E_k \delta_{kK}, \ldots$$

and drop nonlinear terms in the expressions of η, $_E t_{kl}$, P_k, and $\mathscr{M}_k$, resulting in

$$S' = \Sigma^0 + \sigma^0_{kl}[e_{kl} + 2e_{km}(e_{ml} + r_{ml})] + \tfrac{1}{2}\sigma^0_{klmn} e_{kl} e_{mn}$$
$$- e^0_{klm} \mathscr{E}_k e_{lm} - h^0_{klm} B_k e_{lm}$$
$$- \chi^{0E}_k[\mathscr{E}_k + \mathscr{E}_l(e_{lk} + r_{lk})]$$
$$- \tfrac{1}{2}\chi^{0E}_{kl}\mathscr{E}_k\mathscr{E}_l - \chi^{0B}_k[B_k + B_l(e_{lk} + r_{lk})]$$
$$- \tfrac{1}{2}\chi^{0B}_{kl} B_k B_l - \lambda^0_{kl}\mathscr{E}_k B_l$$
$$+ \tfrac{1}{2}\sigma'_{klmn} e_{kl} e'_{mn} - e'_{klm}\mathscr{E}_k e'_{lm}$$
$$- h'_{klm} B_k e'_{lm} - \tfrac{1}{2}\chi'^E_{kl}\mathscr{E}_k\mathscr{E}'_l$$
$$- \tfrac{1}{2}\chi'^B_{kl} B_k B'_l - \lambda'_{kl}\mathscr{E}_k B'_l, \tag{6.12}$$

$$\eta = -\frac{1}{2\rho_0}\left(\frac{\partial \Sigma}{\partial \theta} + \frac{\partial \sigma_{kl}}{\partial \theta} e_{kl} - \frac{\partial \chi^E_k}{\partial \theta} \mathscr{E}_k - \frac{\partial \chi^B_k}{\partial \theta} B_k\right)$$

$$-\frac{1}{2\rho_0}\int_{\mathscr{V}-\sigma}\left(\frac{\partial \overset{*}{\sigma}{}^0_{kl}}{\partial \theta} e'_{kl} - \frac{\partial \overset{*}{\chi}{}^{0E}_k}{\partial \theta} \mathscr{E}'_k - \frac{\partial \overset{*}{\chi}{}^{0B}_k}{\partial \theta} B'_k\right) dv', \tag{6.13}$$

$$_E t_{kl} = (1 - e_{rr})\sigma_{kl} + \sigma_{ml}(e_{km} + r_{km}) + \sigma_{km}(e_{lm} + r_{lm})$$
$$+ \sigma_{klmn} e_{mn} - e_{mkl}\mathscr{E}_m - h_{mkl} B_m$$
$$+ \int_{\mathscr{V}-\sigma} (\sigma'_{klmn} e'_{mn} - e'_{mkl}\mathscr{E}'_m - h'_{mkl} B'_m)\, dv', \tag{6.14}$$

$$P_k = (1 - e_{rr})\chi^E_k + \chi^E_l(e_{kl} + r_{kl}) + \chi^E_{kl}\mathscr{E}_l + e_{klm} e_{lm}$$
$$+ \lambda_{kl} B_l + \int_{\mathscr{V}-\sigma} (\chi'^E_{kl}\mathscr{E}'_l + e'_{klm} e'_{lm} + \lambda'_{kl} B'_l)\, dv', \tag{6.15}$$

$$\mathscr{M}_k = (1 - e_{rr})\chi^B_k + \chi^B_l(e_{kl} + r_{kl}) + \chi^B_{kl} B_l + h_{klm} e_{lm}$$
$$+ \lambda_{lk}\mathscr{E}_l + \int_{\mathscr{V}-\sigma} (\chi'^B_{kl} B'_l + h'_{klm} e'_{lm} + \lambda'_{lk}\mathscr{E}'_l)\, dv'. \tag{6.16}$$

Local (unprimed) material moduli ($\Sigma, \sigma_{kl}, e_{klm}, \ldots$) are functions of $\mathbf{x}$ and θ and the nonlocal (primed) moduli ($\sigma'_{kl}, e'_{klm}, \sigma'^E_k, \ldots$) are functions of $\mathbf{x}$, $\mathbf{x}'$, and θ. For homogeneous materials the local moduli are independent of $\mathbf{x}$ and the nonlocal moduli are functions of $\mathbf{x}' - \mathbf{x}$. Physical meaning of various nonlocal moduli are the same as in the local theory,[18] except that they are volume densities, i.e.,

σ'_{klmn} = elastic moduli volume density,

e'_{klm} = piezoelectric moduli volume density,

h'_{klm} = piezomagnetic moduli volume density,

χ'^E_{kl} = dielectric susceptibility density,

χ'^B_{kl} = magnetic susceptibility density,

λ'_{kl} = magnetic polarizability density.

If the natural state is stress-free, unpolarized, and unmagnetized, then $\sigma_{kl} = \chi_k^E = \chi_k^B = 0$. In this case, we can absorb the local terms into nonlocal ones by redefining them, and then Eqs. (6.12)–(6.16) may be expressed as

$$S' = \Sigma^0 + \tfrac{1}{2}\sigma'_{klmn}e_{kl}e'_{mn} - e'_{klm}\mathscr{E}_k e'_{lm} - h'_{klm}B_k e'_{lm} - \tfrac{1}{2}\chi'^E_{kl}\mathscr{E}_k\mathscr{E}'_l - \tfrac{1}{2}\chi'^B_{kl}B_kB'_l - \lambda'_{kl}\mathscr{E}_kB'_l, \quad (6.17)$$

$$\eta = -\frac{1}{2\rho_0}\frac{\partial\Sigma}{\partial\theta}, \quad (6.18)$$

$$_Et_{kl} = \int_{\mathscr{V}-\sigma}(\sigma'_{klmn}e'_{mn} - e'_{mkl}\mathscr{E}'_m - h'_{mkl}B'_m)\,dv', \quad (6.19)$$

$$P_k = \int_{\mathscr{V}-\sigma}(\chi'^E_{kl}\mathscr{E}'_l + e'_{klm}e'_{lm} + \lambda'_{kl}B'_l)\,dv', \quad (6.20)$$

$$\mathscr{M}_k = \int_{\mathscr{V}-\sigma}(\chi'^B_{kl}B'_l + h'_{klm}e'_{lm} + \lambda'_{lk}\mathscr{E}'_l)\,dv', \quad (6.21)$$

where now the nonlocal moduli must be a Dirac-delta function sequence, so that, in the limit when the nonlocality vanishes, these equations must revert to classical (local) forms.

If the material possesses certain symmetry represented by a group of orthogonal transformations $\{\mathbf{S}\}$, then the material moduli must obey the following types of functional relations:

$$\begin{aligned} &S_{kp}S_{lq}\chi'^E_{pq}(\boldsymbol{\kappa},\theta) = \chi'^E_{kl}(\mathbf{S}\boldsymbol{\kappa},\theta),\\ &S_{kp}S_{lq}S_{mr}e'_{pqr}(\boldsymbol{\kappa},\theta) = e'_{klm}(\mathbf{S}\boldsymbol{\kappa},\theta),\\ &S_{kp}S_{lq}S_{mr}S_{ns}\sigma'_{pqrs}(\boldsymbol{\kappa},\theta) = \sigma'_{klmn}(\mathbf{S}\boldsymbol{\kappa},\theta), \end{aligned} \quad (6.22)$$

where $\boldsymbol{\kappa} = \mathbf{x}' - \mathbf{x}$, for all members of the group $\{\mathbf{S}\}$. As a consequence of these, the material moduli will be restricted in their dependence on $\mathbf{x}' - \mathbf{x}$. For example, for the isotropic dielectrics, these imply that

$$\begin{aligned} \chi'^E_{kl} &= \chi'_1\delta_{kl} + \chi'_2(x'_k - x_k)(x'_l - x_l),\\ e'_{klm} &= e'_1(x'_k - x_k)\delta_{lm} + e'_2[(x'_l - x_l)\delta_{km} + (x'_m - x_m)\delta_{kl}]\\ &\quad + e'_3(x'_k - x_k)(x'_l - x_l)(x'_m - x_m), \end{aligned} \quad (6.23)$$

$$\begin{aligned} \sigma'_{klmn} &= \lambda'\delta_{kl}\delta_{mn} + \mu'(\delta_{km}\delta_{ln} + \delta_{kn}\delta_{lm})\\ &\quad + \sigma'_1[(x'_m - x_m)(x'_n - x_n)\delta_{kl}\\ &\quad + (x'_k - x_k)(x'_l - x_l)\delta_{mn}]\\ &\quad + \sigma'_2[(x'_k - x_k)(x'_m - x_m)\delta_{nl}\\ &\quad + (x'_k - x_k)(x'_n - x_n)\delta_{ml}\\ &\quad + (x'_l - x_l)(x'_m - x_m)\delta_{kn} + (x'_l - x_l)(x'_n - x_n)\delta_{km}]\\ &\quad + \sigma'_3(x'_k - x_k)(x'_l - x_l)(x'_m - x_m)(x'_n - x_n). \end{aligned}$$

Similar expressions are valid for other moduli. Here the coefficients $\chi'_1, \chi'_2, e'_1, e'_2, e'_3, \lambda', \mu', \sigma'_1, \sigma'_2$, and σ'_3 are functions of $|\mathbf{x}' - \mathbf{x}|$ and θ, e.g.,

$$\chi'_1 = \chi'_1(|\mathbf{x}' - \mathbf{x}|,\theta), \quad \sigma'_\alpha = \sigma'_\alpha(|\mathbf{x}' - \mathbf{x}|,\theta), \quad \alpha = 1,2,3. \quad (6.24)$$

The appearance of the material moduli χ'_2, e'_α, and σ'_α indicates that, even though the material may be considered isotropic in the macroscopic scale (as in the classical theory of isotropic dielectrics), the atomic and molecular orientation may induce directional dependence. These additional terms, however, are expected to be small when macroscopic characteristic lengths guide the physical phenomena. It must be noted that, for other microscopic symmetry, Eqs. (6.22) lead to more complicated anisotropies and the consequence of Eqs. (6.22) requires more detailed study. For vector and second-order tensor functions, these restrictions are known for 32 crystal classes.[21]

Finally, we note that when the material moduli become Dirac-delta measure, all constitutive equations revert to their classical forms. Thus, we expect that the material moduli must be delta sequence in an internal length parameter so that when this parameter approaches zero, classical local field theories are obtained, e.g.,

$$\lambda' = \lambda'(\mathbf{x}' - \mathbf{x}, a), \quad \lim_{a\to 0}\lambda' = \lambda\delta(\mathbf{x}' - \mathbf{x}). \quad (6.25)$$

If we also recall the attenuating neighborhood hypothesis as formalized by an influence function, we may use such forms as

$$\lambda' = A\exp[-(k^2/a^2)(\mathbf{x}' - \mathbf{x})\cdot(\mathbf{x}' - \mathbf{x})], \quad (6.26)$$

where k and A are constants and a is an internal characteristic length (e.g., lattice parameter, granular distance, distance of fibers, etc.). The constant A may be obtained by the normalization

$$\int_{\mathscr{V}-\sigma}\lambda'\,dv' = \lambda. \quad (6.27)$$

For instance, for a body of infinite extent in N dimensions, Eq. (6.26) and (6.27) give

$$A = \lambda\pi^{-N/2}(k/a)^N. \quad (6.28)$$

Of course, other possibilities exist. We may, for example, determine λ' by comparing the dispersion curves obtained in lattice dynamics with those calculated by means of nonlocal theory (cf. Eringen[14,15]; see also Sec. 11).

7. DIELECTRICS

Most dielectric materials are nonmagnetizable. For these materials, the effect of the B-field can be discarded. Since the free energy Ψ is now independent of $\mathbf{B}$, the entropy inequality (4.15) gives

$$M_K = 0, \quad (7.1)$$

and constitutive equations (5.1) and (5.4), (5.5), and (5.6) do not contain $\mathbf{B}$. We list below linear constitutive equations for the case when the natural state is free of stress and em fields (i.e., $\sigma_{kl} = 0$, $\chi_k^E = 0$, $\chi_k^B = 0$).

$$\rho_0\Psi = \Sigma + \frac{1}{2}\int_{\mathscr{V}-\sigma}[\sigma'_{klmn}e_{kl}e'_{mn} - e'_{klm}(\mathscr{E}'_k e_{lm} + \mathscr{E}_k e'_{lm}) - \chi'^E_{kl}\mathscr{E}_k\mathscr{E}'_l]\,dv', \quad (7.2)$$

$$\eta = -\frac{1}{\rho_0}\frac{\partial\Sigma}{\partial\theta}, \quad (7.3)$$

$$_Et_{kl} = \int_{\mathscr{V}-\sigma}(\sigma'_{klmn}e'_{mn} - e'_{mkl}\mathscr{E}'_m)\,dv', \quad (7.4)$$

$$P_k = \int_{\mathscr{V}-\sigma} (\chi'^{E}_{kl} \mathscr{E}'_l + e'_{klm} e'_{lm})\, dv' . \tag{7.5}$$

For *rigid dielectric*, the dependence on the strain tensor is eliminated and we have for the polarization, the linear constitutive equation

$$P_k = \int_{\mathscr{V}-\sigma} \chi'^{E}_{kl} \mathscr{E}'_l \, dv' . \tag{7.6}$$

For unbounded solids, the Fourier transforms of Eqs. (7.4) and (7.5) are useful:

$$\begin{aligned} {}_E\bar{t}_{kl} &= \bar{\sigma}'_{klmn} \bar{e}'_{mn} - \bar{e}'_{mkl} \bar{\mathscr{E}}'_m , \\ \bar{P}_k &= \bar{\chi}'^{E}_{kl} \bar{\mathscr{E}}'_l + \bar{e}'_{klm} \bar{e}'_{lm} . \end{aligned} \tag{7.7}$$

The nonlocal moduli $\bar{\sigma}'_{klmn}$, $\bar{e}'_{mkl}$, and $\bar{\chi}'^{E}_{kl}$, are functions of $\boldsymbol{\xi}$ and θ only, where $\boldsymbol{\xi}$ is the wave vector and a superposed bar indicates the Fourier transform, e.g.,

$$\bar{F}(\boldsymbol{\xi}) = (2\pi)^{-3/2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\mathbf{x}) e^{i\boldsymbol{\xi}\cdot\mathbf{x}}\, dx_1\, dx_2\, dx_3 . \tag{7.8}$$

The general forms of the nonlocal moduli can be obtained in terms of $\boldsymbol{\xi}$ by using the invariant theory for each class of crystal symmetry. Below, we give these expressions for isotropic and uniaxial crystals.

A. Isotropic dielectrics

For isotropic solids, the nonlocal moduli are in the forms as given by Eqs. (6.23) with $(\mathbf{x}' - \mathbf{x})$ replaced by $\boldsymbol{\xi}$. Consequently,

$$\begin{aligned} \bar{\epsilon}_{kl}(\boldsymbol{\xi},\theta) &= \delta_{kl} + \bar{\chi}'^{E}_{kl} = (\delta_{kl} - \xi_k \xi_l/\xi^2)\epsilon_T(\xi^2,\theta) \\ &\quad + (\xi_k\xi_l/\xi^2)\epsilon_L(\xi^2,\theta) , \end{aligned} \tag{7.9}$$

$$\begin{aligned} \bar{e}'_{klm}(\boldsymbol{\xi},\theta) &= \xi^{-1}(\gamma_1 \xi_k \delta_{lm} + \gamma_2 \xi_l \delta_{km} + \gamma_2 \xi_m \delta_{kl}) \\ &\quad + \gamma_3 \xi^{-3} \xi_k \xi_l \xi_m , \end{aligned} \tag{7.10}$$

$$\begin{aligned} \bar{\sigma}'_{klmn}(\boldsymbol{\xi},\theta) &= \bar{\lambda}'\delta_{kl}\delta_{mn} + \bar{\mu}'(\delta_{km}\delta_{ln} + \delta_{kn}\delta_{lm}) \\ &\quad + \lambda'_1 \xi^{-2}(\xi_m\xi_n\delta_{kl} + \xi_k\xi_l\delta_{mn}) \\ &\quad + \lambda'_2 \xi^{-2}(\xi_k\xi_m\delta_{ln} + \xi_k\xi_n\delta_{lm} \\ &\quad + \xi_l\xi_m\delta_{kn} + \xi_l\xi_n\delta_{km}) \\ &\quad + \lambda'_3 \xi^{-4}\xi_k\xi_l\xi_m\xi_n , \end{aligned} \tag{7.11}$$

where ϵ_T and ϵ_L represent the transverse and longitudinal dielectric moduli. These and γ_α, $\bar{\lambda}'$, $\bar{\mu}'$, and λ'_α are functions of ξ^2 and θ only. The dependence of $\bar{\epsilon}_{kl}$ on the wave vector indicates the spatial dispersion of optical waves. The presence of ϵ_L can be shown to lead to the Debye screening of the field of a stationary point charge in the medium (see Sec. 8).

From Eq. (7.10), it is clear that elastic strains can cause polarization in an *isotropic* solid. Moreover, since Eq. (7.10) is an odd function of $\boldsymbol{\xi}$, the polarization is reversed by reversing the direction of the wave vector. Thus, an isotropic nonlocal electromagnetic elastic solid can display piezoelectric effect.

Similarly, Eq. (7.11) indicates that, in an isotropic solid, the stress at a point depends on orientation. This fact is in accordance with the physics of matter at the atomic scale.

According to classical theory of piezoelectricity, for isotropic solids $e'_{mkl} = 0$ so that the electric field cannot contribute to the stress ${}_Et_{kl}$. From Eq. (7.10), it is clear that there will be a contribution to the elastic stress field from the electric field for small wavelengths, i.e., near the boundary of the Brillouin zone. At this region, Brillouin scattering from an exciton has been observed. Such an effect can be explained in terms of $\bar{\epsilon}_{kl}$ alone. However, optical activity, anisotropic stress–optic effects, in an isotropic solid requires the interaction of the strain field with the polarization, i.e., the presence of the material moduli e'_{klm}.

B. Uniaxial crystals

For uniaxial crystals, the material moduli can be obtained by determining the general form of a second-order symmetric isotropic tensor that depends on $\boldsymbol{\xi}$ and a unit vector (say $\mathbf{i}_3$) directed along the optic axis, i.e.,

$$\bar{\epsilon}_{kl} = \bar{\epsilon}_{kl}(\boldsymbol{\xi},\mathbf{i}_3) . \tag{7.12}$$

To determine $\bar{e}'_{klm}$ and $\bar{\sigma}'_{klmn}$, we form

$$\begin{aligned} \bar{f}_{kl}(\boldsymbol{\xi},\mathbf{i}_3;\mathbf{v}) &= \bar{e}'_{klm}(\boldsymbol{\xi},\mathbf{i}_3) v_m , \\ \bar{\sigma}_{kl}(\boldsymbol{\xi},\mathbf{i}_3;\tau_{mn}) &= \bar{\sigma}'_{klmn}(\boldsymbol{\xi},\mathbf{i}_3)\tau_{mn} . \end{aligned} \tag{7.13}$$

Generators of $\bar{\epsilon}_{kl}$, $\bar{f}_{kl}$, and $\bar{\sigma}_{kl}$ can be read from tables available (cf. Eringen,[18] p. 534), retaining only the linear terms in v_m and τ_{mn}. Once this is done, it follows that

$$\bar{e}'_{klm} = \left.\frac{\partial \bar{f}_{kl}}{\partial v_m}\right|_{v_m = 0} , \quad \bar{\sigma}'_{klmn} = \left.\frac{\partial \bar{\sigma}_{kl}}{\partial \tau_{mn}}\right|_{\tau_{mn}} = 0 . \tag{7.14}$$

Resulting expressions are

$$\begin{aligned} \bar{\epsilon}_{kl} &= (\delta_{kl} - \xi_k\xi_l/\xi^2)\epsilon_T + (\xi_k\xi_l/\xi^2)\epsilon_L \\ &\quad + \epsilon_0 \delta_{3k}\delta_{3l} + \epsilon_R \xi^{-1}(\xi_k\delta_{3l} + \xi_l\delta_{3k}) , \end{aligned} \tag{7.15}$$

$$\begin{aligned} \bar{e}'_{klm} &= \xi^{-1}[\gamma_1\xi_k\delta_{lm} + \gamma_2(\xi_l\delta_{km} + \xi_m\delta_{kl})] \\ &\quad + \gamma_3\xi^{-3}\xi_k\xi_l\xi_m + \gamma_4\delta_{k3}\delta_{lm} + \gamma_5(\delta_{kl}\delta_{m3} \\ &\quad + \delta_{km}\delta_{l3}) + \gamma_6\delta_{k3}\delta_{l3}\delta_{m3} + \xi^{-1}(\gamma_7\xi_k\delta_{l3}\delta_{m3} \\ &\quad + \gamma_8\xi\delta_{k3}\delta_{m3} + \gamma_8\xi_m\delta_{k3}\delta_{l3}) + \xi^{-2}(\gamma_9\xi_l\xi_m\delta_{k3} \\ &\quad + \gamma_{10}\xi_k\xi_l\delta_{m3} + \gamma_{10}\xi_k\xi_m\delta_{l3}) , \end{aligned} \tag{7.16}$$

$$\begin{aligned} \sigma'_{klmn} &= \bar{\lambda}'\delta_{kl}\delta_{mn} + \bar{\mu}'(\delta_{km}\delta_{ln} + \delta_{kn}\delta_{lm}) \\ &\quad + \lambda'_1\xi^{-2}(\xi_m\xi_n\delta_{kl} + \xi_k\xi_l\delta_{mn}) \\ &\quad + \lambda'_2\xi^{-2}(\xi_k\xi_m\delta_{ln} + \xi_k\xi_n\delta_{lm} + \xi_l\xi_m\delta_{kn} + \xi_l\xi_n\delta_{km}) \\ &\quad + \lambda'_3\xi^{-4}\xi_k\xi_l\xi_m\xi_n + \lambda'_4(\delta_{kl}\delta_{m3}\delta_{n3} + \delta_{mn}\delta_{k3}\delta_{l3}) \\ &\quad + \lambda'_5(\delta_{lm}\delta_{k3}\delta_{n3} + \delta_{km}\delta_{l3}\delta_{n3} + \delta_{ln}\delta_{k3}\delta_{m3} \\ &\quad + \delta_{kn}\delta_{l3}\delta_{m3}) + \lambda'_6\delta_{k3}\delta_{l3}\delta_{m3}\delta_{n3} + \xi^{-1}\lambda'_7(\xi_l\delta_{mn}\delta_{k3} \\ &\quad + \xi_k\delta_{mn}\delta_{l3} + \xi_n\delta_{kl}\delta_{m3} + \xi_m\delta_{kl}\delta_{n3}) \\ &\quad + \xi^{-1}\lambda'_8(\xi_l\delta_{k3}\delta_{m3}\delta_{n3} + \xi_k\delta_{l3}\delta_{m3}\delta_{n3} \\ &\quad + \xi_m\delta_{k3}\delta_{l3}\delta_{n3} + \xi_n\delta_{k3}\delta_{l3}\delta_{m3}) , \end{aligned} \tag{7.17}$$

where ϵ_T, ϵ_L, ϵ_0, ϵ_R, γ_α, $\bar{\lambda}'$, $\bar{\mu}'$, and λ'_α are functions of ξ^2, ξ_3, and θ.

8. POINT CHARGE IN AN ELASTIC DIELECTRIC

Consider a point charge e located at $\mathbf{x} = \mathbf{0}$ in an isotropic elastic dielectric of infinite extent. We would like to determine the electric field $\mathbf{E}$ and the elastic displacement caused by this charge. Two of the surviving Maxwell's equations are

$$\nabla \cdot (\mathbf{E} + \mathbf{P}) = q, \quad \nabla \times \mathbf{E} = \mathbf{0}, \tag{8.1}$$

where

$$q = e\delta(\mathbf{x}). \tag{8.2}$$

In addition, we need equations of equilibrium (2.9),

$${}_E t_{kl,k} = 0, \tag{8.3}$$

in which we have dropped the nonlinear em terms. From $(8.1)_2$, we have

$$\mathbf{E} = -\nabla \phi, \tag{8.4}$$

where ϕ is the electric potential. First, we take the Fourier transform of (8.2) and (8.4) and then substitute these and the Fourier transform of e_{kl} given by (6.10) into (7.7). Carrying (7.7) into the Fourier transforms of (8.1) and (8.3) leads to

$$\begin{aligned} &\bar{e}'_{klm}\xi_m\xi_k\bar{u}_l - \bar{\epsilon}_{kl}\xi_k\xi_l\bar{\phi} = -\bar{q}, \\ &\bar{\sigma}'_{klmn}\xi_k\xi_n\bar{u}_m + \bar{e}'_{mkl}\xi_k\xi_m\bar{\phi} = 0, \end{aligned} \tag{8.5}$$

where

$$\bar{\epsilon}_{kl} = \delta_{kl} + \bar{\chi}'^{E}_{kl}. \tag{8.6}$$

For isotropic materials using (7.9)–(7.11), Eqs. (8.5) can be reduced to the forms

$$\begin{aligned} &\gamma\xi^{-1}\boldsymbol{\xi}\cdot\mathbf{u} - \epsilon_L\bar{\phi} = -\bar{q}\xi^{-2}, \\ &\alpha\boldsymbol{\xi}\boldsymbol{\xi}\cdot\bar{\mathbf{u}} + \beta\xi^2\bar{\mathbf{u}} + \gamma\boldsymbol{\xi}\xi\bar{\phi} = 0, \end{aligned} \tag{8.7}$$

where

$$\begin{aligned} &\gamma(\xi^2,\theta) = \gamma_1 + 2\gamma_2 + \gamma_3, \\ &\alpha(\xi^2,\theta) = \bar{\lambda}' + \bar{\mu}' + 2\lambda'_1 + 3\lambda'_2 + \lambda'_3, \\ &\beta(\xi^2,\theta) = \bar{\mu}' + \bar{\lambda}'_2. \end{aligned} \tag{8.8}$$

The scalar product of Eq. $(8.7)_2$ by $\boldsymbol{\xi}$ gives

$$\boldsymbol{\xi}\cdot\mathbf{u} = -[\gamma\xi/(\alpha+\beta)]\bar{\phi} \tag{8.9}$$

provided $\xi^2(\alpha+\beta)\neq 0$. Substituting this into Eq. $(8.7)_1$, we obtain

$$\begin{aligned} &\bar{\phi} = \bar{q}\xi^{-2}[\epsilon_L + \gamma^2/(\alpha+\beta)]^{-1}, \\ &\bar{\mathbf{u}} = -[\gamma/(\alpha+\beta)](\boldsymbol{\xi}/\xi)\bar{\phi}. \end{aligned} \tag{8.10}$$

The inverse Fourier transforms of these give ϕ and $\mathbf{u}$. We consider the following two special cases.

A. Rigid dielectric

In this case, $\gamma = 0$. To simplify the matter, we also select

$$\epsilon_L/\epsilon_0 = 1 + (r_s^2\xi^2)^{-1}. \tag{8.11}$$

This implies that the field of the point charge in the medium differs from the Coulomb field. The inverse Fourier transform of $(8.10)_1$ gives

$$\phi(\mathbf{x}) = (e/4\pi\epsilon_0 r)e^{-r/r_s}, \tag{8.12}$$

where $r = |\mathbf{x}|$ is the radial distance. The fact that such a potential corresponds to the Debye screening of the field is well known.

B. Elastic dielectric

Suppose that, again, ϵ_L is given by (8.11) and the piezoelectric moduli γ, α, and β are constants. The elastic displacement field will be in the direction of the electric field if $\gamma/(\alpha+\beta)$ is a positive, imaginary scalar, i.e.,

$$\gamma/(\alpha+\beta) = ib/\xi, \tag{8.13}$$

where, in general, $b = b(\xi^2,\theta)$. Further, if we take b real and positive constant and use (8.11), then $(8.10)_1$ reduces to

$$\bar{\phi} = \bar{q}/\epsilon_0(\xi^2 + r_m^{-2}), \tag{8.14}$$

where

$$r_m^{-2} = r_s^{-2} - b^2(\alpha+\beta)/\epsilon_0. \tag{8.15}$$

Assuming that $b^2(\alpha+\beta)/\xi_0 < r_s^{-2}$, the inverse Fourier transform of (8.14) is found to be

$$\phi = (e/4\pi\epsilon_0 r)e^{-r/r_m}. \tag{8.16}$$

Compared to (8.12), this result indicates that the elastic deformations due to the electric field of the point charge increases the Debye screening radius r_s. By measuring this change, it should be possible to determine the material moduli $b^2(\alpha+\beta)/\epsilon_0$. It appears that the Debye screening radius may be made very large if $b^2(\alpha+\beta)/\epsilon_0 = r_s^{-2}$. Whether this is possible or not, depends on the magnitudes of material moduli.

The elastic displacement field is obtained by inverting $(8.10)_2$. To this end, we employ (8.9) and (8.13), i.e.,

$$\overline{u_{l,l}} = -b\bar{\phi}, \tag{8.17}$$

which gives

$$\nabla\cdot\mathbf{u} = -b\phi. \tag{8.18}$$

This can be integrated to give

$$u_r = \frac{b}{r^2}\int^r \rho^2\phi(\rho)\,d\rho + \frac{C}{r^2}, \tag{8.19}$$

where C is a constant of integration. Substituting for ϕ from (8.16), we obtain

$$u_r = \frac{be}{4\pi\epsilon_0}\frac{r_m}{r}\left(1 + \frac{r_m}{r}\right)e^{-r/r_m} + \frac{C}{r^2}. \tag{8.20}$$

If we assume that the point charge is in a small spherical inclusion with radius $r = r_0$, then the volume change due to the inclusion is given by $4\pi r_0^2 u_r(r_0) = \delta v$. This determines the "strength" of singularity C introduced by the point defect. As $r_0 \to 0$, we have

$$(ber_m^2/\epsilon_0) + 4\pi C = \delta v, \tag{8.21}$$

which indicates a decrease (ber_m^2/ϵ_0) in the strength of singularity over the case of point defect with no charge.

Employing (8.16) and (8.20), it is not difficult to determine the stress, strain, and electric fields.

It is clear from this analysis that nonlocal effects are important even for static problems. In the discussion of point defects, impurities, dislocations, cracks, space charge singularities, and magnetic dipoles, clearly nonlocal theory can be an effective tool circumventing major difficulties due to the discrete nature of materials. Especially for imperfect materials, this model should bear much fruit.

9. DISPERSION OF OPTICAL MODES

Here, we consider the propagation of plane harmonic waves in an anisotropic nonmagnetizable crystal of infinite extent. In this case, $M_k = 0$ and the three-dimensional Fourier transform of **P** is given by

$$\overline{P}_k(\xi,t) = \overline{\chi}_{kl}^{\prime E}\overline{E}_l^{\prime} . \tag{9.1}$$

The Fourier transform of the two Maxwell equations (2.5) and (2.7) with $\mathbf{M} = \mathbf{0}$ read

$$\begin{aligned} &\boldsymbol{\xi}\times\overline{\mathbf{E}} + (\omega/c)\overline{\mathbf{H}} = \mathbf{0} , \\ &\boldsymbol{\xi}\times\mathbf{H} - (\omega/c)(\overline{\mathbf{E}} + \overline{\mathbf{P}}) = \mathbf{0} . \end{aligned} \tag{9.2}$$

Eliminating $\overline{\mathbf{H}}$ between these two equations and using (9.1), we will have

$$[\xi^2\delta_{kl} - \xi_k\xi_l - (\omega^2/c^2)\overline{\epsilon}_{kl}]\overline{E}_l = 0 . \tag{9.3}$$

A nontrivial solution of Eq. (9.3) exists if

$$\det[\xi^2\delta_{kl} - \xi_k\xi_l - (\omega^2/c^2)\overline{\epsilon}_{kl}] = 0 . \tag{9.4}$$

This is the dispersion relation for optical modes. We examine two special cases.

A. Isotropic solids

In this case, $\overline{\epsilon}_{kl}$ is given by (7.9) and (9.4) leads to the roots

$$\xi^2c^2/\omega^2 = \epsilon_T(\xi^2,\theta) , \tag{9.5}$$

$$\xi^2c^2/\omega^2 = \epsilon_T(\xi^2,\theta) - \epsilon_L(\xi^2,\theta) . \tag{9.6}$$

If we substitute (7.9) into (9.3), we obtain

$$\begin{aligned} &[\xi^2 - (\omega^2/c^2)\epsilon_T]\overline{\mathbf{E}} - \boldsymbol{\xi}(\boldsymbol{\xi}\cdot\overline{\mathbf{E}}) \\ &\quad\times[1 + (\omega^2/c^2\xi^2)(\epsilon_L - \epsilon_T)] = 0 . \end{aligned} \tag{9.7}$$

From this, it is clear that (9.5) corresponds to $\boldsymbol{\xi}\cdot\overline{\mathbf{E}} = 0$, indicating that the waves are transverse to the direction of propagation. The root (9.6), on the other hand, is not possible for $\mathbf{E}\neq\mathbf{0}$, unless (9.5) is also satisfied. This leads to

$$\epsilon_L(\xi^2,\theta) = 0 . \tag{9.8}$$

From (9.5), it is clear that optical waves are dispersive. By matching the index of refraction $n = \xi c/\omega = c/v$, where $v = \omega/\xi$ is the phase velocity, with experimental results or theoretical formulas based on atomic models, we can determine the dielectric function ϵ_T as a function of wavelength $2\pi/\xi$.

B. Uniaxial crystals

Substituting (7.15) into (9.3), we have

$$\begin{aligned} &[\xi^2 - (\omega^2/c^2)\epsilon_T]\overline{\mathbf{E}} - \boldsymbol{\xi}(\boldsymbol{\xi}\cdot\overline{\mathbf{E}})[1 + (\epsilon_L - \epsilon_T)\omega^2/c^2\xi^2] \\ &\quad + \epsilon_0\overline{E}_3\mathbf{i}_3 + (\epsilon_R/\xi)[\overline{E}_3\boldsymbol{\xi} + (\boldsymbol{\xi}\cdot\overline{\mathbf{E}})\mathbf{i}_3] = \mathbf{0} . \end{aligned} \tag{9.9}$$

From this, for the transverse waves $(\boldsymbol{\xi}\cdot\overline{\mathbf{E}} = 0)$, we have either

$$(c^2\xi^2/\omega^2) = \epsilon_T(\xi^2,\xi_3,\theta) \quad (\overline{E}_3 = 0) \tag{9.10}$$

or

$$\epsilon_0\xi_3 + \epsilon_R\xi = 0 \quad (E_3\neq 0) . \tag{9.11}$$

For longitudinal waves $(\boldsymbol{\xi}\times\overline{\mathbf{E}} = \mathbf{0})$ and we have

$$\mathbf{i}_3\times\boldsymbol{\xi}[\epsilon_0\overline{E}_3 + (\epsilon_R/\xi)\boldsymbol{\xi}\cdot\overline{\mathbf{E}}] = \mathbf{0} . \tag{9.12}$$

Hence, either $\mathbf{i}_3\times\boldsymbol{\xi} = \mathbf{0}$ and the propagation is in the $\mathbf{i}_3$ direction, or else

$$\epsilon_0\xi_3 + \epsilon_R\xi = 0 . \tag{9.13}$$

Equations (9.9)–(9.13) acquire particularly important significance when the memory effects are included in ϵ_{kl}, i.e., where $\overline{\epsilon}_{kl}$ depends on $\boldsymbol{\xi}$ and ω. This will be the case when the nonlocal moduli are considered to depend on the past times as well, e.g.,

$$\epsilon_{kl}^{\prime} = \epsilon_{kl}^{\prime}(\mathbf{x}' - \mathbf{x}, t - t') , \quad -\infty\leqslant t'\leqslant t . \tag{9.14}$$

For a few remarks on the topic, see Sec. 11.

C. A special anisotropic solid

Suppose that the dependence of the nonlocal moduli on the direction of the wave vector is small. In this case, we may write

$$\chi_{kl}^{\prime E} = \chi_{kl}^0\,\eta(|\mathbf{x}' - \mathbf{x}|/\epsilon l) , \tag{9.15}$$

where χ_{kl}^0 is the classical susceptibility tensor, l is an external characteristic length, and

$$\epsilon = \beta a/l , \tag{9.16}$$

in which β is a constant appropriate to each material, a is an internal characteristic length, e.g., lattice parameter, granular distance. We know that $\eta(|\mathbf{x}' - \mathbf{x}|/\epsilon l)$ must be a delta sequence, which is invariant under translation and rotations. Such a sequence may be obtained by solving

$$(1 - \epsilon^2l^2\nabla^2)\eta = \delta(|\mathbf{x}' - \mathbf{x}|) . \tag{9.17}$$

For example, in two-dimensional space extending to infinity, an appropriate solution of Eq. (9.17) is

$$\eta = (2\pi l^2\epsilon^2)^{-1}K_0(|\mathbf{x}' - \mathbf{x}|/l\epsilon) , \tag{9.18}$$

where K_0 is the modified Bessel function.

Note that Eq. (9.18) satisfies the translational invariance and leads to solutions which are delta sequence. Moreover, as $\epsilon\to 0$, Eq. (9.18) reverts to the classical (local) constitutive equation.

The Fourier transform of Eq. (9.15), with the use of Eq. (9.17), gives

$$\overline{\chi}_{kl}^{\prime E} = \chi_{kl}^0(1 + \epsilon^2l^2\xi^2)^{-1} . \tag{9.19}$$

Equation (9.4) now takes a special form:

$$(\delta_{kl} - s_ks_l)\overline{E}_l^0 = \frac{1}{n^2}\frac{\epsilon_{kl}^0\overline{E}_l^0}{1 + \epsilon^2l^2\xi^2} , \tag{9.20}$$

where

$$\epsilon_{kl}^0 = \delta_{kl} + \chi_{kl}^0 , \quad s_k = \xi_k/\xi , \quad n = \xi c/\omega = c/v . \tag{9.21}$$

Equation (9.20) has the classical (local) form[22] with equivalent dielectric constants defined by

$$\boldsymbol{\epsilon}^{\text{eq}} = \boldsymbol{\epsilon}^0/(1 + \epsilon^2l^2\xi^2) . \tag{9.22}$$

Note that, unlike classical theory, $\boldsymbol{\epsilon}^{\text{eq}}$ depends on the wave number ξ. We can now borrow all the classical results with the provision equation (9.22), e.g., the orthogonality relations

$$\bar{E}^0_\alpha \cdot \epsilon^{eq} \bar{E}^0_\beta = \delta_{\alpha\beta} , \tag{9.23}$$

the dispersion relations

$$\frac{1}{n^2} = \sum_{k=1}^{3} \frac{s_k}{n^2 - \epsilon^{eq}_{kk}} . \tag{9.24}$$

An alternative form of Eq. (9.24) is

$$\sum_{k=1}^{3} \frac{s_k^2}{v^2 - v_k^2} = 0 , \tag{9.25}$$

where v_k are called *principal speeds*, which do not form a vector. They are given by

$$v_k = c/\sqrt{\epsilon^{eq}_{kk}} , \quad k = 1,2,3 . \tag{9.26}$$

The phase velocity v is obtained by solving Eq. (9.25) for v:

$$v = \{\tfrac{1}{2}[c_1 \pm (c_1^2 - 4c_2)]^{1/2}\}^{1/2}, \tag{9.27}$$

where

$$\begin{aligned} c_1 &= s_1^2(v_2^2 + v_3^2) + s_2^2(v_3^2 + v_1^2) + s_3^2(v_1^2 + v_2^2) , \\ c_2 &= s_1^2 v_2^2 v_3^2 + s_2^2 v_3^2 v_1^2 + s_3^2 v_1^2 v_2^2 . \end{aligned} \tag{9.28}$$

Note that v_k and c_k depend on the wave number ξ; consequently, the phase velocity is a function of the wavelength. The space dispersion is then clearly indicated.

10. PIEZOELECTRIC WAVES

Field equations of linear piezoelectric waves are obtained by substituting Eqs. (7.4) and (7.5) into Eqs. (2.4) and (2.9), with $\mathbf{B} = \mathbf{0}$, and the electric field is determined by a potential ϕ, i.e.,

$$\mathbf{E} = -\nabla\phi . \tag{10.1}$$

Upon writing

$$\begin{aligned} \frac{\partial \chi'^E_{kl}(\mathbf{x}' - \mathbf{x})}{\partial x_m} \frac{\partial \phi'}{\partial x'_r} &= -\frac{\partial \chi'^E_{kl}(\mathbf{x}' - \mathbf{x})}{\partial x'_m} \frac{\partial \phi'}{\partial x'_r} \\ &= -\frac{\partial}{\partial x'_m}\left(\chi'^E_{kl} \frac{\partial \phi'}{\partial x'_r}\right) + \chi'^E_{kl} \frac{\partial^2 \phi}{\partial x'_r \, \partial x'_m} , \end{aligned}$$

in the volume integrals, we can convert the first term to a surface integral by means of the Green–Gauss theorem. Employing this procedure for the other terms and substituting for the linear strain measures

$$e_{kl} = \tfrac{1}{2}(u_{k,l} + u_{l,k}) , \tag{10.2}$$

we obtain the field equations of piezoelectricity

$$\begin{aligned} -\nabla^2\phi &- \int_{\mathscr{V}} (\chi'^E_{kl} \phi'_{,lk} - e'_{klm} u'_{m,lk}) \, dv' \\ &+ \oint_{\partial\mathscr{V}} (\chi'^E_{kl} \phi'_{l} - e'_{klm} u'_{m,l}) n'_k \, da' = 0 , \end{aligned} \tag{10.3}$$

$$\begin{aligned} &\int_{\mathscr{V}} (\sigma'_{klmn} u'_{m,nk} + e'_{mkl} \phi'_{,mk}) \, dv' \\ &\quad - \oint_{\partial\mathscr{V}} (\sigma'_{klmn} u'_{m,n} + e'_{mkl} \phi'_{,m}) n'_k \, da' + \rho(f_l - \ddot{u}_l) = 0 . \end{aligned} \tag{10.4}$$

These two integro–partial-differential equations must be solved to determine the electric potential $\phi(\mathbf{x},t)$ and the displacement field $u_k(\mathbf{x},t)$. It is interesting to note that these equations contain surface integrals over $\partial\mathscr{V}$. These are the effects of "surface electric field" and "surface tensions" which are not included in the classical (local) theory of piezoelectricity. Therefore, nonlocal theory accounts for the *surface phenomena*, and it presents interesting possibilities for the discussion of surface physics.

We now consider a solid of infinite extent and investigate plane wave propagations. In this case, the surface terms will vanish at infinity, and the Fourier transforms of (10.3) and (10.4) give

$$(\xi^2 + \bar{\chi}'^E_{kl} \xi_k \xi_l)\bar{\phi} - \bar{e}'_{klm} \xi_k \xi_l \bar{u}_m = 0 , \tag{10.5}$$

$$\bar{\sigma}'_{klm} \xi_k \xi_m \bar{u}_m + \bar{e}'_{mkl} \xi_m \xi_k \bar{\phi} - \rho\omega^2 u_l = 0 , \tag{10.6}$$

where a superposed bar represents the Fourier transform, as defined by Eq. (7.8). From Eq. (10.5), we solve

$$\bar{\phi} = (\bar{e}'_{klm} \xi_k \xi_l / \bar{\epsilon}_{pq} \xi_p \xi_q)\bar{u}_m , \tag{10.7}$$

where we have introduced dielectric moduli $\bar{\epsilon}_{pq} = \delta_{pq} + \bar{\chi}'^E_{pq}$.

Carrying (10.7) into (10.6), we obtain

$$(Q_{lm} - \rho c^2 \delta_{lm})\bar{u}_m = 0 , \tag{10.8}$$

where $c^2 = \omega^2/\xi^2$ is the phase velocity and

$$Q_{lm} = \tau_{klmn} s_k s_n , \quad s_k \equiv \xi_k/\xi , \tag{10.9}$$

is the *acoustical tensor* expressed in terms of the *piezoelectrically stiffened stiffness* tensor $\bar{\tau}$ defined

$$\tau_{klmn} = \bar{\sigma}'_{klmn} + (\mathbf{s} \cdot \boldsymbol{\epsilon} \cdot \mathbf{s})^{-1} \bar{e}'_{pkl} s_p \bar{e}'_{qmn} s_q . \tag{10.10}$$

Nontrivial solutions of Eq. (10.8) exists if

$$\det(\mathbf{Q} - \rho c^2 \mathbf{1}) = 0 . \tag{10.11}$$

The three roots c of this bicubic equation determine c^2.

Equations (10.7) and (10.8) are familiar to us from classical piezoelectricity (cf. Ref. 11, p. 488). The main difference is in that c is a function of the wavelength vector, i.e.,

$$c = c(\boldsymbol{\xi}) , \tag{10.12}$$

while, in classical theory, the dependence on the wave vector is missing. The dispersion of waves having short wavelengths is, of course, well known in the atomic theory of lattices.

Uniaxial crystals

As a special case, consider the antiplane elastic displacement field in the direction of the x_1 axis in which u_1 depends only on x_3 and t, with x_3 being the optic axis of a uniaxial crystal. The electric field has the only nonvanishing component E_1. In this case, Eqs. (10.5) and (10.6) reduce to

$$\begin{aligned} &e_1 \bar{u}_1 - \epsilon\bar{\phi} = 0 , \\ &\sigma\bar{u} + e_2\bar{\phi} - (\rho\omega^2/\xi_3^2) = 0 , \end{aligned} \tag{10.13}$$

where from (7.9)–(7.11)

$$\epsilon \equiv (1 - \xi_3^2/\xi^2)\epsilon_T + (\xi_3^2/\xi^2)\epsilon_L + \epsilon_0 + 2\epsilon_R \xi_3/\xi ,$$

$$e_1 = (\gamma_1 + \gamma_7)\xi^{-1}\xi_1 + 2\gamma_{10}\xi^{-2}\xi_1\xi_3 + \gamma_3\xi^{-3}\xi_1\xi_3^2 , \quad (10.14)$$

$$e_2 = (\gamma_2 + \gamma_8)\xi^{-1}\xi_1 + (\gamma_9 + \gamma_{10})\xi^{-2}\xi_1\xi_3 + \gamma_3\xi^{-3}\xi_1\xi_3^2 ,$$

$$\sigma = \overline{\mu}' + \lambda_5' + \lambda_2'\xi^{-2}(\xi_1^2 + \xi_3^2) + \lambda_3'\xi^{-4}\xi_1^2\xi_3^2 .$$

From (10.13), the dispersion relations follow:

$$\rho\omega^2/\xi_3^2 = \sigma + e_1 e_2/\epsilon . \quad (10.15)$$

By measuring the index of refraction, the combined piezoelectric moduli appearing in these equations can be determined as functions of the wave number.

Dispersion of elastic waves is well known and observed for many crystals. The fact that significant deviations which occur from the classical results as the wavelength becomes shorter, approaching the boundaries of the Brillouin zone, is a clear indication of the fact that piezoelectric waves are also dispersive. Importance of the dispersion in the microwave region needs no comment.

11. MEMORY-DEPENDENT MATERIALS

A thorough discussion of electrodynamics of nonlocal continua that possess memory, requires a separate study, which is under preparation. For the linear theory, we can state a few results by inspection. As in the case of viscoelastic solids, clearly the nonlocal moduli will depend also on past history, i.e., the time interval, $t - t'$, e.g.,

$$\chi_{kl}^E = \chi_{kl}^E(\mathbf{x}' - \mathbf{x}, t - t') , \quad -\infty \leqslant t' \leqslant t . \quad (11.1)$$

Consequently, constitutive equations will also involve integrals over the time domain, e.g.,

$$P_k = \int_{-\infty}^{t} dt' \int_{\mathscr{V}} \chi_{kl}'^E(\mathbf{x}' - \mathbf{x}, t - t') E_l'(\mathbf{x}', t') \, dv(x') . \quad (11.2)$$

The four-dimensional Fourier transform of (11.2) replaces (9.1) by

$$\overline{P}_k(\boldsymbol{\xi}, \omega) = \overline{\chi}_{kl}'^E(\boldsymbol{\xi}, \omega)\overline{E}_l(\boldsymbol{\xi}, \omega) . \quad (11.3)$$

This situation is valid for all other constitutive equations. We note, however, that the complex susceptibility tensor $\overline{\chi}_{kl}'^E$ must satisfy certain symmetry requirements, since the time domain is cut off at t, i.e., $-\infty \leqslant t' \leqslant t$. This emanates from the fact that $\chi_{kl}^E(\mathbf{x}' - \mathbf{x}, t - t')$ must be a real function.

By considering all material functions as functions of ω also, we obtain, from our previous results, equations that are valid for memory-dependent materials. Consider, for instance, Eqs. (9.5) and (9.8):

$$\xi^2 c^2/\omega^2 = \epsilon_T(\xi^2, \omega) \quad \text{(transverse optical modes)}, \quad (11.4)$$

$$\epsilon_L(\xi^2, \omega) = 0 \quad \text{(longitudinal optical modes)}. \quad (11.5)$$

Both wavelength and frequency dispersion are now accounted for.

According to an atomic theory, the dielectric constant is given by the classical formula (cf. Hodgson[23])

$$\epsilon(\xi, \omega) = \epsilon(\infty) + \alpha(\omega_0^2 - \omega^2 + B\xi^2)^{-1} , \quad (11.6)$$

where $\epsilon(\infty)$ is the dielectric constant for frequencies well above ϵ_0 and α and B are known constants. Equation (11.6) is valid when ξ^4 is negligible as compared to ξ^2 in the parenthesis.

If we now equate (11.6) to (11.3), we determine $\epsilon_T(\xi^2, \omega)$. Once this is done, other problems can be solved by using ϵ_T or its inverse Fourier transform.

From these and other considerations, as discussed in our various previous works, it is clear that the nonlocal continuum theory of electromagnetism can be used to discuss problems all the way from atomic to macroscopic scale phenomena. The nonlinear theory can be used for nonlinear optics, magnetism, and electromagneto-elasticity. In the nonlinear case, however, the Fourier-domain formalism is no longer valid, and we must employ the full constitutive equations such as (5.4)–(5.7), in the domain of space-time.

A systematic study of conduction and nonlinear electromagnetic theory of memory-dependent nonlocal continua is left to future work.

ACKNOWLEDGMENT

I wish to thank Professor P. W. Anderson for a helpful discussion on space dispersion of optical modes.

[1] Max Born and Emil Wolf, *Principle of Optics* (Pergamon, New York, 1970), 4th ed., p. 91.
[2] W. F. Brown, *Micromagnetics* (Wiley, New York, 1963).
[3] Gérard A. Maugin and A. Cemal Eringen, J. Math. Phys. **6**, 779 (1965).
[4] C. Kittel, Rev. Mod. Phys. **21**, 541 (1949).
[5] F. Bloch, Z. Phys. **49**, 295 (1932).
[6] David Pines, *Elementary Excitations in Solids* (Benjamin, New York, 1954).
[7] J. Lindhard, Kgl. Danske Videnskab, Mat.-Fys. Medd. **28**(8) (1954).
[8] H. Ehrenreich and M. H. Cohen, Phys. Rev. **115**, 786 (1959).
[9] Stephen L. Adler, Phys. Rev. **126**, 413 (1962).
[10] David R. Penn, Phys. Rev. **128**, 2098 (1962).
[11] Nathan Wiser, Phys. Rev. **129**, 62 (1963).
[12] A. A. Rukhadze and V. P. Silin, Usp. Fiz. Nauk. **74**, 223 (1961) [Sov. Phys. Usp. **4**, 459 (1961)].
[13] A. Cemal Eringen, J. Math. Phys. **14**, 733 (1973).
[14] A. Cemal Eringen, Int. J. Eng. Sci. **10**, 425 (1972).
[15] A. Cemal Eringen, *Nonlinear Equations in Physics and Math*, edited by A. O. Barut (D. Reidel, Dordrecht, 1978), p. 271.
[16] A. Cemal Eringen, J. Phys. D Appl. Phys. **10**, 671 (1977).
[17] A. Cemal Eringen, C. Speziale, and B. S. Kim, J. Mech. Phys. Solids **25**, 339 (1977).
[18] A. Cemal Eringen, *Mechanics of Continua* (Krieger, Melbourne, FL, 1980), 2nd ed., p. 446.
[19] A. Cemal Eringen, Int. J. Eng. Sci. **4**, 179 (1966).
[20] N. Friedman and M. Katz, Arch. Ratl. Mech. Anal. **21**, 49 (1960).
[21] A. J. M. Spencer, *Continuum Physics*, edited by A. C. Eringen (Academic, New York, 1971), Vol. I, Part iii.
[22] Donald F. Nelson, *Electric, Optic, and Acoustic Interactions in Dielectrics* (Wiley, New York, 1979), p. 166.
[23] J. N. Hodgson, *Optical Absorption and Dispersion in Solids* (Chapman and Hall, London, 1970), p. 107.