Generalized quaternions and spacetime symmetries[a)]

Robert T. Jantzen
Harvard–Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138

(Received 2 March 1982; accepted for publication 7 May 1982)

The construction of a class of associative composition algebras q_n on R^4 generalizing the well-known quaternions Q provides an explicit representation of the universal enveloping algebra of the real three-dimensional Lie algebras having tracefree adjoint representations (class A Bianchi type Lie algebras). The identity components of the four-dimensional Lie groups $GL(q_n,1) \subset q_n$ (general linear group in one generalized quaternion dimension) which are generated by the Lie algebra of this class of quaternion algebras are diffeomorphic to the manifolds of spacetime homogeneous and spatially homogeneous spacetimes having simply transitive homogeneity isometry groups with tracefree Lie algebra adjoint representations. In almost all cases the complete group of isometries of such a spacetime is isomorphic to a subgroup of the group of left and right translations and automorphisms of the appropriate generalized quaternion algebra. Similar results hold for the single class B Lie algebra of Bianchi type V, characterized by its "pure trace" adjoint representation.

PACS numbers: 02.10. + w, 02.20. + b, 98.80.Dr

1. INTRODUCTION

Generalized quaternions were first employed in the description of a spacetime isometry group by Kurt Gödel in his 1949 paper presenting his famous cosmological solution of the Einstein field equations.[1] These "Gödel quaternions" (also called "split quaternions" or "antiquaternions"[2]) belong to a real subalgebra of the complexified quaternion algebra which is not equivalent to the ordinary real quaternion algebra. The equally historically important static Einstein cosmological solution of 1917 provides a corresponding application for the ordinary quaternions, as described in detail by Ozsváth and Schücking in their treatment of a generalization of this solution.[3] Ozsváth generalized both the Einstein and Gödel solutions by considering spacetime homogeneous solutions of the Einstein equations, leading to four classes I–IV of solutions.[4,5] The homogeneity groups of the first three classes have tracefree Lie algebra adjoint representations but those of the class IV solutions do not. The quaternions are relevant to the class I solutions which generalize the Einstein static solution while the Gödel quaternions are relevant to the class II and III generalizations of the Gödel solution.[5,6]

The quaternions and Gödel quaternions are representations of the universal enveloping algebras[7,8] of the two inequivalent real three-dimensional semisimple Lie algebras, namely the Lie algebras of the groups SU(2) and SL(2,R), respectively, both of which have tracefree adjoint representations. The construction of the present paper[9] extends this representation of the universal enveloping algebra to all of the real three-dimensional Lie algebras with tracefree adjoint representations, namely the class A Bianchi type Lie algebras.[10,11] In addition to the semisimple Bianchi types VIII ($\mathfrak{sl}(2,R)$) and IX($\mathfrak{su}(2)$), these include the Bianchi types I (abelian case), II (Heisenberg Lie algebra of supertriangular 3×3 matrices), VI_0 (Poincaré Lie algebra in two dimensions), and VII_0 (Lie algebra of the Euclidean group in two dimensions). The corresponding Lie groups are the homogeneity isometry groups of the class A spatially homogeneous spacetimes.[5,12] Taub and Misner made use of the ordinary quaternions in their study of the Bianchi type IX Taub–NUT spacetime.[13] The same construction also applies to the Lie algebras with "pure trace" adjoint representations in all dimensions $n > 1$, which for $n = 3$ includes the single class B Lie algebra of Bianchi type V.

The ordinary quaternion algebra Q (Bianchi type IX) is a division algebra,[8] namely an associative algebra in which every nonzero element has an inverse. The generalized quaternions of the remaining Bianchi types are examples of "singular division algebras" which are associative algebras in which almost every nonzero element (all but a set of measure zero) has an inverse. This term was introduced by Illamed and Salingaros,[14] who studied the real and complex three-dimensional division and singular division algebras which possess a nondegenerate norm (Bianchi type IX and VIII quaternions in the real case). Familiar examples of singular division algebras are the general linear group Lie algebras $\mathfrak{gl}(n,R)$, $\mathfrak{gl}(n,C)$, and $\mathfrak{gl}(n,Q)$ of real dimension n^2, $2n^2$, and $4n^2$, respectively. The cases $n = 1$, i.e., R, C, and Q, are well known to be the only finite-dimensional real division algebras.[8] The Lie algebra $\mathfrak{gl}(2,R)$ is in fact isomorphic to the Gödel quaternions. The $n = 1$ construction given below involves a single real parameter γ_{11} whose three inequivalent values -1, 1, and 0 lead to the complex numbers C and two singular division algebras,[2] the nondegenerate case $\gamma_{11} = 1$ having been discussed by Salingaros.[14] For $n = 2$ the pure trace construction given below is the general case for a Clifford-like multiplication. This case should have been given by Campbell in his classification of all three-dimensional real associative algebras[15] but an error resulted in the omission of a large class of examples. His discussion is apparently a slightly different presentation of earlier work by Lie.[16] In the older literature, associative algebras are referred to as higher complex numbers or hypercomplex numbers. A classification of all hypercomplex number systems of dimension less than seven was given by Pierce,[17] but his results are not very

[a)] Supported in part by National Science Foundation Grant No. PHY-80-07351.

transparent since he does not use bases containing the unit element.

2. THE CONSTRUCTION

Consider a real n-dimensional Lie algebra $\mathfrak{g}$ with basis $\{e_a\}$, dual basis $\{\omega^a\}$, structure constant tensor components $C^a{}_{bc} = \omega^a([e_b,e_c]) = C^a{}_{[bc]}$, and adjoint matrices $\mathbf{k}_a \equiv ad_e(e_a) = C^b{}_{ac}\mathbf{e}^c{}_b$, which span the matrix representation of the adjoint representation[18] of $\mathfrak{g}$ with respect to the basis $\{e_a\}$. Here $\{\mathbf{e}^a{}_b\}$ is the natural basis of $\mathfrak{gl}(n,R)$ in terms of which an $n\times n$ matrix is given by $\mathbf{A} = A^a{}_b\mathbf{e}^b{}_a$. Note that the Jacobi identity $C^d{}_{[ab}C^e{}_{c]d} = 0$ is equivalent to the matrix identity $[\mathbf{k}_a,\mathbf{k}_b] = C^c{}_{ab}\mathbf{k}_c$. For $n>1$ define

$$(n-1)C_a = C^b{}_{ab} = \mathrm{Tr}\, ad(e_a) = \mathrm{Tr}\,\mathbf{k}_a,$$
$$(n-1)\gamma_{ab} = \mathrm{Tr}\, ad(e_a)ad(e_b) = \mathrm{Tr}\,\mathbf{k}_a\mathbf{k}_b = (n-1)\gamma_{(ab)},$$
$$(n-1)C_{abc} = (n-1)\gamma_{ad}C^d{}_{bc} = \mathrm{Tr}\,\mathbf{k}_a[\mathbf{k}_b,\mathbf{k}_c] = (n-1)C_{[abc]}. \tag{2.1}$$

$(n-1)\gamma_{ab}$ are the components of the Killing form of $\mathfrak{g}$,[18] while the antisymmetry of C_{abc} follows from the properties of the trace operation Tr. Note that by taking the trace of the Jacobi identity in matrix form, one obtains the relation $C_dC^d{}_{ab} = 0$.

Define an algebra on R^{n+1} with natural basis $\{\mathbf{e}_\alpha\} = \{\mathbf{e}_0,\mathbf{e}_a\}$ by introducing the following Clifford-like multiplication of the basis elements:

$$\mathbf{e}_0\mathbf{e}_\alpha = \mathbf{e}_\alpha\mathbf{e}_0 = \mathbf{e}_\alpha,\quad \mathbf{e}_a\mathbf{e}_b = \gamma_{ab}\mathbf{e}_0 + C^c{}_{ab}\mathbf{e}_c, \tag{2.2}$$

or, in a more uniform notation,

$$\mathbf{e}_\alpha\mathbf{e}_\beta = \gamma_{\alpha\beta}\mathbf{e}_0 + C^\gamma{}_{\alpha\beta}\mathbf{e}_\gamma \equiv M^\gamma{}_{\alpha\beta}\mathbf{e}_\gamma,$$
$$\gamma_{\alpha\beta} \equiv \delta^0{}_\alpha\delta^0{}_\beta + \gamma_{ab}\delta^a{}_\alpha\delta^b{}_\beta,$$
$$C^\alpha{}_{\beta\gamma} \equiv \delta^\alpha{}_a\delta^b{}_\beta\delta^c{}_\gamma C^a{}_{bc} = M^\alpha{}_{[\beta\gamma]}. \tag{2.3}$$

The unit element of this algebra is $\mathbf{e}_0$ and the multiplication of two arbitrary elements $\mathbf{a} = a^\alpha\mathbf{e}_\alpha$ and $\mathbf{b} = b^\alpha\mathbf{e}_\alpha$ of R^{n+1} is given by

$$\mathbf{ab} = M^\gamma{}_{\alpha\beta}a^\alpha b^\beta\mathbf{e}_\gamma. \tag{2.4}$$

Then in the following three cases this is an associative algebra:

$$\text{(i)}\ n=1,\quad \gamma_{11}\in R,$$
$$\text{(ii)}\ n>1,\quad C^a{}_{bc} = C_d\delta^{da}_{bc} \equiv 2C_d\delta^d{}_{[b}\delta^a{}_{c]},$$
$$\text{(iii)}\ n=3,\quad C_a = 0. \tag{2.5}$$

For an associative algebra, the associator[2,8]

$$(\mathbf{e}_\alpha,\mathbf{e}_\beta,\mathbf{e}_\gamma) \equiv (\mathbf{e}_\alpha\mathbf{e}_\beta)\mathbf{e}_\gamma - \mathbf{e}_\alpha(\mathbf{e}_\beta\mathbf{e}_\gamma) = (M^\delta{}_{\alpha\beta}M^\epsilon{}_{\delta\gamma} - M^\epsilon{}_{\alpha\delta}M^\delta{}_{\beta\gamma})\mathbf{e}_\epsilon \tag{2.6}$$

vanishes identically. Since this vanishes if any index is zero due to the properties of the unit element, it is sufficient to consider only $(\mathbf{e}_a,\mathbf{e}_b,\mathbf{e}_c)$.

In the case (i), $C^1{}_{11} = 0$ and (e_1,e_1,e_1) trivially vanishes. In the remaining cases using the Jacobi identity and antisymmetry of C_{abc} one easily finds

$$(\mathbf{e}_a,\mathbf{e}_b,\mathbf{e}_c) = (C^d{}_{be}C^e{}_{ca} - 2\gamma_{b[c}\delta^d{}_{a]})\mathbf{e}_d. \tag{2.7}$$

In the case (ii), use of the relation $\gamma_{ab} = C_aC_b$ leads immediately to the vanishing of this expression.

For $n=3$ one has the well-known decomposition[10,11]

$$C^a{}_{bc} = \epsilon_{bcd}n^{ad} + a_d\delta^{da}_{bc},\quad a_d \equiv C_d,$$
$$\gamma_{ab} = a_aa_b - \tfrac{1}{2}\epsilon_{acd}\epsilon_{bfg}n^{cf}n^{dg}. \tag{2.8}$$

The vanishing or nonvanishing of the Lie algebra representation $\mathrm{Tr}\, ad{:}\mathfrak{g}\to R$ (i.e., $a_d = 0$ or $a_d \neq 0$) divides all three-dimensional real Lie algebras into two classes called class A and class B, respectively. To show that $(\mathbf{e}_a,\mathbf{e}_b,\mathbf{e}_c) = 0$ when $a_d = 0$, assume that the structure constant tensor components are in standard diagonal form,[19] i.e., $\mathbf{n} \equiv n^{ab}\mathbf{e}^b{}_a = \mathrm{diag}(n^{(1)},n^{(2)},n^{(3)})$, in which case $\gamma_{ab}\mathbf{e}^b{}_a = -\,\mathrm{diag}(n^{(2)}n^{(3)},n^{(3)}n^{(1)},n^{(1)}n^{(2)})$ and $C^A{}_{BC} = n^{(A)} = -C^A{}_{CB}$ for each cyclic permutation (A,B,C) of $(1,2,3)$ are the only nonvanishing components. The class A identity $\mathbf{k}_a^2 = \gamma_{aa}(1-\mathbf{e}^a{}_a)$ (no sum on a) shows that $(e_a,e_a,e_b) = 0 = (\mathbf{e}_b,\mathbf{e}_a,\mathbf{e}_a)$ (no sum on a). Since $\mathbf{e}_{(a}\mathbf{e}_{b)} = 0$ for $a\neq b$, this implies left and right alternativity,[2] and since in an alternative algebra the alternator is totally antisymmetric,[2,8] it suffices to examine $(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3)$, which is easily seen to vanish.

In what follows only $n=3$ will be considered. Denote the algebra on R^4 resulting from case (iii) by $q_{\mathbf{n}}$ and call its elements generalized quaternions. It is sufficient to consider only a canonical set of values of the structure constant tensor components for each Bianchi type Lie algebra. The following canonical values of $(n^{(1)},n^{(2)},n^{(3)})$ for the class A Bianchi types will be assumed here: IX(1,1,1), VIII(1,1, − 1), VII_0(1,1,0), VI_0(1, − 1,0), II(0,0,1), I(0,0,0). When these values are understood the quaternion algebra will be denoted by q_Z, where Z is the Roman numeral Bianchi type, and will be referred to as the canonical Bianchi type Z generalized quaternion algebra on R^4. Thus $q_{\mathrm{IX}} = Q$ and q_{VIII} is the Gödel quaternion algebra while q_{VII_0} has been called the "semiquaternions."[2] In the nonabelian case, again letting (A,B,C) be a cyclic permutation of (1,2,3), at least one component, say $n^{(A)}$, is nonzero, in which case the algebra $q_Z(Z\neq I)$ is isomorphic to the Clifford algebra generated by $\mathbf{e}_B$ and $\mathbf{e}_C$.[20]

Case (ii) for $n=3$ yields generalized quaternions for the single class B Lie algebra of Bianchi type V whose canonical structure constant tensor components may be taken to be $n=0$ and $a_b = \delta^3{}_b$ so that $\gamma_{ab} = \delta^3{}_a\delta^3{}_b$. Denote the corresponding algebra by q_{V} and let q denote all of the canonical generalized quaternion algebras for $n=3$.

Define quaternion conjugation* (an involutive antiautomorphism[2]) by

$$\mathbf{a} = a^\alpha\mathbf{e}_\alpha \to \mathbf{a}^* = a^0\mathbf{e}_0 - a^a\mathbf{e}_a,$$
$$(\mathbf{ab})^* = \mathbf{b}^*\mathbf{a}^*. \tag{2.9}$$

The real valued quaternion norm and trace are defined by

$$N(\mathbf{a})\mathbf{e}_0 \equiv |\mathbf{a}|^2\mathbf{e}_0 = \mathbf{a}^*\mathbf{a} = \mathbf{aa}^*,$$
$$N(\mathbf{a}) = (a^0)^2 - \gamma_{ab}a^aa^b \equiv \gamma^*{}_{\alpha\beta}a^\alpha a^\beta,$$
$$\mathrm{Tr}(\mathbf{a})\mathbf{e}_0 \equiv a + a^* = 2a^0\mathbf{e}_0. \tag{2.10}$$

These appear in the characteristic equation satisfied by every quaternion,[21]

$$\mathbf{a}^2 - \mathrm{Tr}(\mathbf{a})\mathbf{a} + N(\mathbf{a})\mathbf{e}_0 = 0. \tag{2.11}$$

For canonical structure constant components one has in the class A case

$$|\mathbf{a}|^2 = (a^0)^2 + n^{(2)}n^{(3)}(a^1)^2 + n^{(3)}n^{(1)}(a^2)^2 + n^{(1)}n^{(2)}(a^3)^2, \tag{2.12}$$

and in the type V case

$$|\mathbf{a}|^2 = (a^0)^2 - (a^3)^2. \tag{2.13}$$

The norm satisfies the relation

$$N(\mathbf{ab}) = N(\mathbf{a})N(\mathbf{b}), \tag{2.14}$$

i.e., is an algebra homomorphism into the real numbers. An algebra with such a norm is called a composition algebra.[21]

The generalized quaternions with vanishing norm are called singular quaternions. If $|\mathbf{a}|^2 \neq 0$, then $\mathbf{a}$ is called nonsingular since it has the inverse

$$\mathbf{a}^{-1} = |\mathbf{a}|^{-2}\mathbf{a}^*, \quad \mathbf{a}^{-1}\mathbf{a} = \mathbf{a}\mathbf{a}^{-1} = \mathbf{e}_0. \tag{2.15}$$

Because of (2.14), the open submanifold of R^4 consisting of nonsingular quaternions forms a Lie group which is generated by "the Lie algebra"[7] of the quaternion algebra q, namely q with the Lie bracket given by the ordinary commutator, denoted by $\mathfrak{gl}(1,q)$. This four-dimensional Lie group $GL(1,q) = \{\mathbf{a}\in q \mid |\mathbf{a}|^2 \neq 0\}$ has a natural three-dimensional subgroup $SL(1,q) = \{\mathbf{a}\in q \mid |\mathbf{a}|^2 = 1\}$ of elements with unit norm whose Lie algebra consists of tracefree quaternions $\mathfrak{sl}(1,q) = \{\mathbf{a}\in q \mid \mathrm{Tr}(\mathbf{a}) = 0\} = \mathrm{span}\{\mathbf{e}_a\}$. This latter Lie algebra is isomorphic to the original Lie algebra $\mathfrak{g}$ with which the construction began since the basis $\{\mathbf{e}_a\}$ satisfies

$$[\tfrac{1}{2}\mathbf{e}_a,\tfrac{1}{2}\mathbf{e}_b] = C^c{}_{ab}\tfrac{1}{2}\mathbf{e}_c. \tag{2.16}$$

Thus $e_a \to \frac{1}{2}\mathbf{e}_a$ is a Lie algebra isomorphism and q is a representation of the universal enveloping algebra of $\mathfrak{g}$.[7,8]

$GL(1,q)$ and $SL(1,q)$ will be called the general and special linear groups in one generalized quaternion dimension. This terminology arises from the fact that the natural action of $GL(1,q)$ on q by left or right multiplication is linear, while the subgroup $SL(1,q)$ leaves the quaternion norm invariant under this action due to (2.14).[22] Let $GL(1,q)^+$ and $SL(1,q)^+$ be the identity components of these Lie groups. Locally these are the images of their Lie algebras by the generalized quaternion exponential map

$$\exp \mathbf{a} \equiv \sum_{n=0}^{\infty} (n!)^{-1}(\mathbf{a})^n, \quad (\mathbf{a})^0 \equiv \mathbf{e}_0. \tag{2.17}$$

In fact using the identity $a^a a^b \mathbf{e}_a \mathbf{e}_b = \gamma_{ab} a^a a^b \mathbf{e}_0 \equiv u\mathbf{e}_0$, one may easily obtain the formulas

$$\exp(\tfrac{1}{2}a^\alpha \mathbf{e}_\alpha) = e^{(1/2)a^0}(\mathbf{e}_0 \cosh\tfrac{1}{2}u^{1/2} + a^a\mathbf{e}_a u^{-1/2}\sinh\tfrac{1}{2}u^{1/2}),$$
$$N(\exp \mathbf{a}) = e^{\mathrm{Tr}(\mathbf{a})}. \tag{2.18}$$

The second formula shows that exp maps $\mathfrak{sl}(1,q)$ into $SL(1,q)$. The first formula reflects the natural isomorphism $GL(1,q)^+ \cong R \times SL(1,q)^+$,

$$\mathbf{a}\in GL(1,q)^+ \to (\ln|\mathbf{a}|^2, |\mathbf{a}|^{-1}\mathbf{a}) \in R \times SL(1,q)^+. \tag{2.19}$$

The matrix representations with respect to the basis $\{\mathbf{e}_\alpha\}$ of the left and right regular representations of q which are obtained by letting q act on itself by left and right multiplication are two mutually commuting four-dimensional subalgebras of $\mathfrak{gl}(4,R)$ with bases $\{M_\alpha\}$ and $\{\widetilde{M}_\alpha\}$, respectively,

$$\mathbf{M}_\alpha = M^\beta{}_{\alpha\gamma}\mathbf{e}^\gamma{}_\beta, \quad \widetilde{\mathbf{M}}_\alpha = M^\beta{}_{\gamma\alpha}\mathbf{e}^\gamma{}_\beta,$$
$$\mathbf{M}_\alpha\mathbf{M}_\beta = M^\gamma{}_{\alpha\beta}\mathbf{M}_\gamma, \quad \widetilde{\mathbf{M}}_\alpha\widetilde{\mathbf{M}}_\beta = M^\gamma{}_{\beta\alpha}\widetilde{\mathbf{M}}_\gamma,$$
$$[\tfrac{1}{2}\mathbf{M}_\alpha,\tfrac{1}{2}\mathbf{M}_\beta] = C^\gamma{}_{\alpha\beta}\tfrac{1}{2}\mathbf{M}_\gamma,$$
$$[\tfrac{1}{2}\widetilde{\mathbf{M}}_\alpha,\tfrac{1}{2}\widetilde{\mathbf{M}}_\beta] = -C^\gamma{}_{\alpha\beta}\tfrac{1}{2}\widetilde{\mathbf{M}}_\gamma,$$
$$[\mathbf{M}_\alpha,\widetilde{\mathbf{M}}_\beta] = 0. \tag{2.20}$$

These are all equivalent to the associativity condition that (2.6) vanish together with the relation $C^\alpha{}_{\beta\gamma} = M^\alpha{}_{[\beta\gamma]}$ of (2.3). If $\mathbf{a} = a^\alpha \mathbf{e}_\alpha \in q$, then the left multiplication $L_\mathbf{a}$ and right multiplication $R_\mathbf{a}$ are represented by the matrices

$$L_\mathbf{a} \to a^\alpha \mathbf{M}_\alpha \equiv \mathbf{M}(\mathbf{a}),$$
$$R_\mathbf{a} \to a^\alpha \widetilde{\mathbf{M}}_\alpha \equiv \widetilde{\mathbf{M}}(\mathbf{a}). \tag{2.21}$$

Some of the properties of these matrices are

$$\mathrm{Tr}\,\mathbf{M}(\mathbf{a}) = \mathrm{Tr}\,\widetilde{\mathbf{M}}(\mathbf{a}) = 2\,\mathrm{Tr}\,\mathbf{a} = 4a^0,$$
$$\det \mathbf{M}(\mathbf{a}) = \det \widetilde{\mathbf{M}}(\mathbf{a}) = |a|^4,$$
$$\{\mathbf{M}(\mathbf{a}),\mathbf{M}(\mathbf{b})\} = 2\gamma_{\alpha\beta}a^\alpha b^\beta = \mathbf{1} = \{\widetilde{\mathbf{M}}(\mathbf{a}),\widetilde{\mathbf{M}}(\mathbf{b})\}. \tag{2.22}$$

The matrix groups generated by these two matrix algebras are both isomorphic to $GL(1,q)$.

Suppose $\{a^\alpha\}$ are now interpreted as the Cartesian coordinates on R^4 associated with the natural basis $\{\mathbf{e}_\alpha\}$, and set $\partial_\alpha = \partial/\partial a^\alpha$. Then the right and left action of $GL(1,q)$ on $q = R^4$ is generated by the Lie algebras $\{e_\alpha\}$ and $\{\tilde{e}_\alpha\}$, respectively, of generating vector fields

$$e_\alpha = \tfrac{1}{2}M^\gamma{}_{\beta\alpha}a^\beta\partial_\gamma, \quad \tilde{e}_\alpha = \tfrac{1}{2}M^\gamma{}_{\alpha\beta}a^\beta\partial_\gamma,$$
$$[e_\alpha,e_\beta] = C^\gamma{}_{\alpha\beta}e_\gamma, \quad [\tilde{e}_\alpha,\tilde{e}_\beta] = -C^\gamma{}_{\alpha\beta}\tilde{e}_\gamma,$$
$$[e_\alpha,\tilde{e}_\beta] = 0. \tag{2.23}$$

In fact when restricted to the Lie group $LG(1,q)$, $\{e_\alpha\}$ and $\{\tilde{e}_\alpha\}$ are bases for the Lie algebras of, respectively, left and right invariant vector fields on the group. The corresponding dual bases of left and right invariant 1-forms on the group are given, respectively, by

$$\omega^\alpha = 2|\mathbf{a}|^{-2}M^\alpha{}_{\beta\gamma}a^{*\beta}da^\gamma, \quad \tilde{\omega}^\alpha = 2|\mathbf{a}|^{-2}M^\alpha{}_{\gamma\beta}a^{*\beta}da^\gamma, \tag{2.24}$$

which may be written in terms of quaternion valued 1-forms as

$$\boldsymbol{\omega} \equiv \omega^\alpha \mathbf{e}_\alpha = 2\mathbf{a}^{-1}d\mathbf{a}, \quad \tilde{\boldsymbol{\omega}} \equiv \tilde{\omega}^\alpha \mathbf{e}_\alpha = 2d\mathbf{a}\,\mathbf{a}^{-1}. \tag{2.25}$$

The present notation identifies the original Lie algebra $\mathfrak{g}$ having basis $\{e_a\}$ with the Lie algebra of left invariant vector fields on the Lie group $SL(1,q)$.

The exponential formula (2.18) may be used to parametrize $GL(1,q)^+$ in R^4 using various types of canonical coordinates on this group. Define the real valued functions

$$c_a(x) = \cosh((\gamma_{aa})^{1/2}\tfrac{1}{2}x), \quad c_a = c_a(x^a),$$
$$s_a(x) = (\gamma_{aa})^{-1/2}\sinh((\gamma_{aa})^{1/2}\tfrac{1}{2}x), \quad s_a = s_a(x^a), \tag{2.26}$$

which appear in the formula

$$\exp(\tfrac{1}{2}x\mathbf{e}_a) = \mathbf{e}_0 c_a(x) + \mathbf{e}_a s_a(x). \tag{2.27}$$

When $\gamma_{aa} = 0$, these formulas are understood to hold in the limit $\gamma_{aa} \to 0$, i.e., $c_a = 1, s_a = \frac{1}{2}x^a$. A parametrization involving canonical coordinates of the second kind on $GL(1,q)$ is obtained by expanding the product

$$a^\alpha \mathbf{e}_\alpha = \exp(\tfrac{1}{2}x^0\mathbf{e}_0)\exp(\tfrac{1}{2}x^1\mathbf{e}_1)\exp(\tfrac{1}{2}x^2\mathbf{e}_2)\exp(\tfrac{1}{2}x^3\mathbf{e}_3), \quad (2.28)$$

leading in the class A case to the result

$$\begin{aligned} a^0 &= e^{(1/2)x^0}(c_1c_2c_3 - n^{(1)}n^{(2)}n^{(3)}s_1s_2s_3), \\ a^1 &= e^{(1/2)x^0}(s_1c_2c_3 + n^{(1)}c_1s_2s_3), \\ a^2 &= e^{(1/2)x^0}(c_1s_2c_3 - n^{(2)}s_1c_2s_3), \\ a^3 &= e^{(1/2)x^0}(c_1c_2s_3 + n^{(3)}s_1s_2c_3), \end{aligned} \quad (2.29)$$

and in the type V case

$$a^0 = e^{(1/2)x^0}c_3, \quad a^3 = e^{(1/2)x^0}s_3,$$
$$a^1 = \tfrac{1}{2}x^1e^{(1/2)x^0}(c_3 - s_3), \quad a^2 = \tfrac{1}{2}x^2e^{(1/2)x^0}(c_3 - s_3). \quad (2.30)$$

A parametrization generalizing the Euler angles of the type IX case is valid in the class A case as long as $n^{(2)} \neq 0$.

$$a^\alpha \mathbf{e}_\alpha = e^{(1/2)x^0}\exp(\tfrac{1}{2}x^2\mathbf{e}_3)\exp(\tfrac{1}{2}x^1\mathbf{e}_1)\exp(\tfrac{1}{2}x^3\mathbf{e}_3), \quad (2.31)$$

or

$$\begin{aligned} a^0 &= e^{(1/2)x_0}c_1c_3(x^2 + x^3), \quad a^1 = e^{(1/2)x^0}s_1c_3(x^2 - x^3), \\ a^2 &= n^{(2)}e^{(1/2)x^0}s_1s_3(x^2 - x^3), \quad a^3 = e^{(1/2)x^0}c_1s_3(x^2 + x^3). \end{aligned} \quad (2.32)$$

These apply only in the nonabelian class A case, although if one assumes the canonical components $n^{(a)} = \delta^a{}_3$ in the Bianchi type II case, one must cyclically permute the above formulas so that they apply to the case $n^{(3)} \neq 0$. Similar formulas hold for the class B type V case.

In each of these parametrizations $\{x^\alpha\}$ may be interpreted as local coordinates on $\mathrm{GL}(1,q)^+$ for certain ranges of their values. These coordinates are adapted to the direct product structure (2.19), with $x^0 = \ln|\mathbf{a}|^2$ being a homomorphism onto the additive group of real numbers and $\{x^a\}$ being local coordinates on the factor manifold $\mathrm{SL}(1,q)^+$. Note also that

$$\tilde{\omega}^0 = \omega^0 = \mathrm{Tr}\, \mathbf{a}^{-1}d\mathbf{a} = d \ln |\mathbf{a}|^2 = dx^0. \quad (2.33)$$

The manifold of $\mathrm{SL}(1,q)$ is a certain quadratic surface in R^4 given by the equation $\gamma^*_{\alpha\beta}a^\alpha a^\beta = 1$. For Bianchi type IX this is just the unit sphere S_3 and so all of the coordinates $\{x^a\}$ must be restricted to finite intervals (integral multiples of π depending on the exponential parametrization). For Bianchi types VIII and VII_0 one sees that the canonical coordinate of the second kind x^3 must be restricted to an interval of length 4π since $\exp(2\pi\mathbf{e}_3) = \mathbf{e}_0$, i.e., $\mathrm{SL}(1,q)^+$ has one compact direction, being the hyperboloid $(a^0)^2 + (a^3)^2 - (a^1)^2 - (a^2)^2 = 1$ for type VIII and the cylinder $(a^0)^2 + (a^3)^2 = 1$ for type VII_0. In these two cases $\mathrm{SL}(1,q)^+ = \mathrm{SL}(1,q)$ is not simply connected but has a simply connected covering group[22–24] $\overline{\mathrm{SL}(1,q)}$ obtained by extending the range of the canonical coordinate of the second kind x^3 to the real line. The same extension yields the simply connected covering group $\overline{\mathrm{GL}(1,q)^+}$. For Bianchi types V and VI_0, $\mathrm{SL}(1,q)^+$ is one sheet of the hyperbolic cylinder $(a^0)^2 - (a^3)^2 = 1$ and for Bianchi types I and II it is the hyperplane $a^0 = 1$, all of which are simply connected.

The group $\mathrm{Aut}(q)$ of automorphisms of the algebra q is that subgroup of $\mathrm{GL}(4,R)$ acting naturally on R^4 which leaves the group multiplication invariant. In particular, the identity must remain fixed, while the Cliffordlike multiplication requires that an algebra automorphism be an automorphism of the Lie algebra $\mathfrak{sl}(1,q)$, i.e., the algebra homomorphisms coincide with the automorphisms of the Lie algebra from which the quaternion algebra is constructed. This is to be expected since q is a representation of the universal enveloping algebra of this Lie algebra.

The geometry of the generalized quaternion algebra is related to the quadratic form γ^*

$$\gamma^*(\mathbf{a},\mathbf{b}) = \gamma^*{}_{\alpha\beta}a^\alpha b^\beta = \tfrac{1}{2}\,\mathrm{Tr}(\mathbf{ab}^*). \quad (2.34)$$

Using the trace symmetry $\mathrm{Tr}(\mathbf{ab}) = \mathrm{Tr}(\mathbf{ba})$, it is simple to show that this quadratic form is invariant under the independent left and right translation action of $\mathrm{SL}(1,q)$. Similarly one may introduce the following bi-invariant symmetric tensor field on $\mathrm{GL}(1,q)$,

$$\gamma^* = \tfrac{1}{4}\gamma^*{}_{\alpha\beta}\omega^\alpha \otimes \omega^\beta = \tfrac{1}{8}\,\mathrm{Tr}\,\boldsymbol{\omega} \otimes \boldsymbol{\omega} = \tfrac{1}{8}\mathrm{Tr}\,\tilde{\boldsymbol{\omega}} \otimes \tilde{\boldsymbol{\omega}}. \quad (2.35)$$

where the final equality follows from the trace symmetry and the definitions (2.25). For the semisimple Bianchi types this is nondegenerate and therefore a metric tensor field. Its restriction to $\mathrm{SL}(1,q)$ is the metric induced on $\mathrm{SL}(1,q)$ by the inner product space (R^4,γ^*); in fact the submanifolds of constant nonzero quaternion norm are all isometric due to the bi-invariance of the metric γ^*. In the Bianchi type IX case of ordinary quaternions, the inner product space (R^4,γ^*) is Euclidean space and the Riemannian manifold $[\mathrm{SL}(1,q),\gamma^*]$ is the 3-sphere S^3 with its natural metric.[25]

It is worth pointing out the fact that the generalized quaternion algebras q_{V}, q_{VIII}, and $q_{\mathrm{IX}} = Q$ (as is well known in the latter two cases) have matrix representations in two dimensions as real subalgebras of $\mathfrak{gl}(2,C)$. If $\{\boldsymbol{\sigma}_a\}$ are the standard Pauli matrices, then the quaternion basis $\{\mathbf{e}_\alpha\}$ corresponds respectively to $\{\mathbf{1},\mathbf{e}^2{}_1,i\mathbf{e}^2{}_1,\boldsymbol{\sigma}_3\}$, $\{\mathbf{1},\boldsymbol{\sigma}_3,\boldsymbol{\sigma}_1,-i\boldsymbol{\sigma}_2\}$, and $\{\mathbf{1},-i\boldsymbol{\sigma}_a\}$. The latter two bases generate the matrix subalgebras $\mathfrak{gl}(2,R) \simeq R \oplus \mathfrak{sl}(2,R)$ and $\mathfrak{u}(2) \simeq R \oplus \mathfrak{su}(2)$, respectively.

3. HOMOGENEITY GROUPS AND SPACETIME SYMMETRIES

The spacetime homogeneous cosmological models with simply transitive isometry groups may be defined as spacetimes (M,g) whose manifold M is that of a connected four-dimensional Lie group M and whose metric g is a left invariant Lorentz metric on this Lie group.[26,5] Solutions of the Einstein equations with a dust source were studied by Ozsváth,[4,6] Ozsváth and Schücking,[3] and Farnsworth and Kerr,[26] while the results for a general perfect fluid source are quoted by Ryan and Shepley.[5] Of the four classes of solutions, the four-dimensional homogeneity group M is $\mathrm{GL}(1,q)$ of Bianchi type IX for class I solutions and the simply connected covering group $\overline{\mathrm{GL}(1,q)^+}$ of Bianchi type VIII for class II and III solutions. The class IV solutions have groups M which have three-dimensional subgroups of the class B Bianchi types and so do not involve quaternions except in certain degenerate cases.

The Einstein static solution (class I) and the Gödel solution (classes II–IV) are the only solutions with additional continuous symmetry. The Einstein static solution may be written $\mathrm{g} = \mathscr{R}^2\eta_{\alpha\beta}\omega^\alpha \otimes \omega^\beta$, with $\eta_{\alpha\beta}\mathbf{e}^\beta{}_\alpha = \mathrm{diag}\,(-1,1,1,1)$ and $\mathscr{R}$ a constant. This is bi-invariant under the action of

$SL(1,q)$ as well as under the scaling of the quaternion norm so the full isometry group is $GL(1,q)_L \times SL(1,q)_R \cong R \times SL(1,q)_L \times SL(1,q)_R \cong R \times SO(4,R) \times Z_2$. Here $Z_2 = \{\pm \mathbf{e}_0\}$ is the discrete parity subgroup associated with reflection of R^4 about the origin, while the subscripts L and R refer to the left and right actions of $GL(1,q)$ or its subgroups on q. The action of $SL(1,q)_L \times SL(1,q)_R \cong O(4,R)$ is equivalent to the natural action of the orthogonal group on R^4.

The Gödel solution may be written[6]

$$g = \mathscr{R}^2[-\omega^3 \otimes \omega^3 + \tfrac{1}{2}(\omega^1 \otimes \omega^1 + \omega^2 \otimes \omega^2) + \omega^0 \otimes \omega^0], \tag{3.1}$$

where $\{\omega^a\}$ is understood to be extended to the simply connected covering group $\overline{GL(1,q)^+}$. If the factor of $\frac{1}{2}$ were not present the full isometry group would again be $GL(1,q)_L^+ \times SL(1,q)_R$, but its presence limits the additional symmetry to local rotational symmetry, the full group being $\overline{GL(1,q)_L^+} \times \overline{\exp(\text{span}\{\mathbf{e}_3\})_R}$. However, $SL(1,q)_L \times \exp(\text{span}\{\mathbf{e}_3\})_R$ contains an $SO(2,R)$-parametrized family of three-dimensional subgroups G_{III} of Bianchi type III $= VI_{-1}$, which act simply transitively on $SL(q,R)$ and hence identifying the group manifolds of $SL(q,R)$ and G_{III}, one may express the metric in terms of left invariant 1-forms on $R \times \overline{G_{III}}$ [identified with $\overline{GL(1,q)^+}$],[27]

$$g = \mathscr{R}^2[-(\sigma^3 + \sigma^2) \otimes (\sigma^3 + \sigma^2) + \tfrac{1}{2}(\sigma^1 \otimes \sigma^1 + \sigma^2 \otimes \sigma^2) + \omega^0 \otimes \omega^0],$$
$$d\sigma^1 = 0 = d\sigma^3, \quad d\sigma^2 = \sigma^1 \wedge \sigma^3. \tag{3.2}$$

These 1-forms are given explicitly by (A4) in canonical coordinates $\{y^a\}$ of the second kind on $\overline{G_{III}}$, while $\omega^0 + dx^0 = d \ln|\mathbf{a}|^2$. The Appendix clarifies this point. Equation (3.2) is the form of the metric originally given by Gödel,[1] apart from an interchange of x^0 and y^3 and a scaling of x^0, y^2, and y^3 by $\sqrt{2}$.

The spatially homogeneous cosmological models whose homogeneity isometry groups act simply transitively on spacelike hypersurfaces may be defined as spacetimes (M,g) whose manifold M is that of the four-dimensional Lie Group $R \times G$ and whose Lorentz metric g is invariant under the natural left action of the connected three-dimensional subgroup G, the copies of which in the product manifold $R \times G$ are assumed to be spacelike. For the Lie groups of class A Bianchi types as well as Bianchi type V, the spacetime manifold may therefore be identified with $GL(1,q)^+ \cong R \times SL(1,q)^+$ with $G = SL(1,q)^+$. For Bianchi types VI_0 and VIII one must use the simply connected covering groups if simply connected spatial slices are desired.

The quaternion norm parametrizes the family of spatially homogeneous hypersurfaces. The coordinate $x^0 = \ln|\mathbf{a}|^2$ is a homomorphism onto the additive group of real numbers, which is relevant to the Lie group isomorphism $GL(1,q)^+ \cong R \times SL(1,q)^+$. For all but a special class of spatially homogeneous spacetimes with whimper singularities,[28] this coordinate is timelike on the entire spacetime. A spatially homogeneous metric on $GL(1,q)^+$ is of the general form

$$g = g_{\alpha\beta}(x^0)\omega^\alpha \otimes \omega^\beta, \tag{3.3}$$

with $\{\bar{e}_a\}$ a basis of Killing vector fields. Provided x^0 is always timelike one can always choose the simpler form

$$g = -N(x^0)^2 dx^0 \otimes dx^0 + g_{ab}(x^0)\omega^a \otimes \omega^b. \tag{3.4}$$

The additional continuous spacetime symmetries possible for these spacetimes which act within the homogeneous hypersurfaces are local rotational symmetry for all Bianchi types except VI_0 and isotropy for Bianchi types I, VII_0, V, and IX. The local rotational symmetry for Bianchi types VII_0, VIII, and IX corresponds to a one-dimensional isometry subgroup of the group of inner automorphisms of q

$$\mathbf{a} \in q \to AD(\mathbf{b})\mathbf{a} = \mathbf{bab}^{-1}, \quad \mathbf{b} \in GL(1,q). \tag{3.5}$$

The complete isometry group (ignoring discrete symmetries) of these three types is $SL(1,q)^+{}_L \times H_r \cong SL(1,q)^+ x_s AD(H)$, where H is any one-dimensional subgroup of $SL(1,q)$ for Bianchi type IX and the one-dimensional subgroup $\exp \text{span}\{\mathbf{e}_3\}$ for the other two types. Here "x_s" denotes the semidirect product group and the inner automorphism subgroup $AD(H)$ is the isotropy group at the "identity line" $\{t\mathbf{e}_0 | t \in R\}$. For Bianchi types I, II, and V, this subgroup is replaced by a one-dimensional subgroup H of the group of automorphisms of q (not an inner automorphism subgroup). In the case of isotropy the identity component of the isometry group for Bianchi type IX is $SL(1,q)_L \times SL(1,q)_R$ corresponding to bi-invariance of the metric with respect to the subgroup $SL(1,q)$, while the quaternion conjugation map* is a discrete reflection symmetry. For Bianchi type I, the one-dimensional automorphism subgroup H of local rotational symmetry enlarges to a three-dimensional automorphism subgroup, but for Bianchi types VII_0 and V, the additional two dimensions of the three-dimensional isotropy subgroup are not related to automorphisms of the quaternion algebra.[9] Most of the discrete symmetries possible for these spatially homogeneous spacetimes are also directly related to automorphisms of the Lie algebra $\mathfrak{sl}(1,q)$ and hence of the quaternion algebra itself.[29] Schmidt has considered a special class of such examples.[30]

APPENDIX

Consider the following $SO(2,R)$-parametrized family of parametrizations of $SL(1,q_{VIII})$ due to Ozsváth[6]

$$\mathbf{a} = \exp(-\tfrac{1}{2}\theta\mathbf{e}_3)[\tfrac{1}{2}y^2 e^{(1/2)y^1}(\sin\tfrac{1}{2}y^3(\mathbf{e}_1 - \mathbf{e}_0) + \cos\tfrac{1}{2}y^3(\mathbf{e}_2 + \mathbf{e}_3)) + \exp(\tfrac{1}{2}y^1\mathbf{e}_1)\exp(\tfrac{1}{2}y^3\mathbf{e}_3)]\exp(\tfrac{1}{2}\theta\mathbf{e}_3). \tag{A1}$$

This is the sum of a unit quaternion and a null quaternion which is orthogonal to it (with respect to the quadratic form γ^*) and represents a 2-parameter family of straight lines in $SL(1,q_{VIII}) \subset R^4$ for each value of the additional parameter θ. Computing the basis $\{\omega^a\}$ of left invariant 1-forms on $SL(1,q_{VIII})$ using the restriction of (2.24) to the group (i.e., set $\omega^0 = 0$) and the constructing the basis $\{e_a\}$ using duality one finds

$$\omega^1 = c_\theta dy^1 + s_\theta e^{y^1} dy^2,$$
$$\omega^2 = -s_\theta dy^1 + c_\theta e^{y^1} dy^2,$$
$$\omega^3 = dy^3 + e^{y^1} dy^2,$$
$$e_1 = c_\theta \partial_1 + s_\theta (e^{-y^1}\partial_2 - \partial_3),$$
$$e_2 = -s_\theta \partial_1 + c_\theta (e^{-y^1}\partial_2 - \partial_3),$$
$$e_3 = \partial_3, \tag{A2}$$

where $c_\theta = \cos(x^3 + \theta)$ and $s_\theta = \sin(x^3 + \theta)$. [It is helpful to use the isomorphism with SL(2,R).] Similarly one finds the right invariant basis

$$\overline{E}_1 = \partial_1 - y^2\partial_2 = \tilde{e}_1^\theta,$$
$$\overline{E}_2 = y^2\partial_1 + \tfrac{1}{2}(e^{-2y^1} - (y^2)^2)\partial_2 - e^{-y^1}\partial_3 = \tfrac{1}{2}(\tilde{e}_2^\theta - \tilde{e}_3),$$
$$\overline{E}_3 = \partial_2 = \tilde{e}_2^\theta + \tilde{e}_3, \tag{A3}$$

where $\tilde{e}_1^\theta = \cos\theta\, \tilde{e}_1 - \sin\theta\, \tilde{e}_2$ and $\tilde{e}_2^\theta = \sin\theta\, \tilde{e}_1 + \cos\theta\, \tilde{e}_2$.

Introducing the Bianchi type III = VI_{-1} invariant fields

$$\sigma^1 = dy^1, \quad \sigma^2 = e^{y^1}dy^2, \quad \sigma^3 = dy^3,$$
$$\epsilon_1 = \partial_1, \quad \epsilon_2 = e^{-y^1}\partial_2, \quad \epsilon_3 = \partial_3 = \tilde{\epsilon}_3 = e_3,$$
$$\tilde{\epsilon}_1 = \partial_1 - y^2\partial_2, \quad \tilde{\epsilon}_2 = \partial_2, \tag{A4}$$

which depend on θ through the parametrization (A1), one sees that any locally rotationally symmetric left invariant metric on SL(1,q_{VIII}) may be written as a left invariant metric on a Bianchi type III Lie group G_{III} with the same base manifold,

$$g = g_{11}(\omega^1 \otimes \omega^1 + \omega^2 \otimes \omega^2) + g_{33}\omega^3 \otimes \omega^3$$
$$= g_{11}(\sigma^1 \otimes \sigma^1 + \sigma^2 \otimes \sigma^2)$$
$$+ g_{33}(\sigma^3 + \sigma^2) \otimes (\sigma^3 + \sigma^2). \tag{A5}$$

The SO(2,R)-parametrized family of coordinates $\{y^a\}$ on SL(1,q_{VIII}) provides a corresponding family of identifications of the manifold of G_{III} with that of SL(1,q_{VIII}); these coordinates are global coordinates on the manifold R^3 of the simply connected covering groups of both types on which y^3 assumes all real values rather than being restricted to an interval of 4π.

[1]K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).
[2]M. L. Tomber, Had. J. **2**, 1252 (1979); L. Sorgsepp and J. Lôhmus, *ibid.* **2**, 1388 (1979).
[3]I. Ozsváth and E. L. Schücking, Ann. Phys. **55**, 166 (1969).
[4]I. Ozsváth, J. Math. Phys. **6**, 590 (1965).
[5]M. P. Ryan, Jr., and L. C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U. P., Princeton, 1975).
[6]I. Ozsváth, J. Math. Phys. **11**, 2871 (1970).
[7]N. Jacobson, *Lie Algebras* (Wiley, New York, 1962).
[8]A. G. Kurosh, *Lectures in General Algebra* (Pergamon, Oxford, 1965).
[9]R. T. Jantzen, unpublished undergraduate thesis, Princeton University, 1974; "Spatially Homogeneous Cosmology: Background and Dynamics," unpublished report, U. C. Berkeley, 1977.
[10]F. B. Estabrook, H. D. Wahlquist, and C. G. Behr, J. Math. Phys. **9**, 497 (1968).
[11]G. F. R. Ellis and M. A. H. MacCallum, Commun. Math. Phys. **12**, 108 (1969).
[12]A. H. Taub, Ann. Math. **53**, 472 (1951).
[13]C. W. Misner and A. H. Taub, Sov. Phys. JETP **28**, 122 (1969).
[14]Y. Llamed and N. Salingaros, J. Math. Phys. **22**, 2091 (1981); N. Salingaros, *ibid.* **22**, 2096 (1981); **22**, 226 (1981).
[15]J. E. Campbell, *Introductory Treatise on Lie's Theory of Finite Continuous Transformation Groups* (Oxford U. P., Oxford, 1903), Chap. 25.
[16]S. Lie and G. Scheffers, *Vorlesungen über continuierliche gruppen* (Teubner, Berlin, 1893), Chap. 21.
[17]B. Pierce, Am. J. Math. **4**, 97 (1881).
[18]A. A. Sagle and R. E. Walde, *Introduction to Lie Groups and Lie Algebras* (Academic, New York, 1973).
[19]R. T. Jantzen, Commun. Math. Phys. **64**, 211 (1979).
[20]C. Chevalley, *The Construction and Study of Certain Important Algebras* (Mathematical Society of Japan, Tokyo, 1955), p. 40.
[21]N. Jacobson, Rend. Circ. Mat. Palermo **7**, 55 (1958).
[22]R. Gilmore, *Lie Groups, Lie Algebras and Some of their Applications* (Wiley, New York, 1974).
[23]J. G. Miller, Commun. Math. Phys. **22**, 98 (1976).
[24]S. T. C. Siklos, Phys. Lett. **A59**, 173 (1976).
[25]R. T. Jantzen, J. Math. Phys. **19**, 1163 (1978).
[26]D. L. Farnsworth and R. P. Kerr, J. Math. Phys. **7**, 1625 (1966).
[27]L. Bianchi, "Sugli spazi a tre dimensioni che ammettono un gruppo continuo di movimenti," *Opere* (Edizione Cremonese, Rome, 1958), Vol. 9.
[28]S. T. C. Siklos, Commun. Math. Phys. **58**, 255 (1978).
[29]R. T. Jantzen, "Perfect Fluid Sources for Spatially Homogeneous Spacetimes," in preparation.
[30]B. G. Schmidt, Commun. Math. Phys. **15**, 329 (1969).

Some comments on finite subgroups of SU(3)

W. M. Fairbairn
University of Lancaster, Lancaster, England
T. Fulton
The Johns Hopkins University, Baltimore, Maryland

(Received 2 February 1982; accepted for publication 28 May 1982)

A recently published set of finite subgroups of SU(3) is shown to contain some groups which are not subgroups of SU(3). The others are subgroups of one of the dihedral-like family of SU(3) subgroups Δ, of order $3n^2$. Some comments are made also on the structure of other finite subgroups previously listed.

PACS numbers: 02.20 + b

I. INTRODUCTION

In a recent publication, Bovier *et al.*[1] (BLW), it has been claimed that a new class of finite subgroups of SU(3) has been obtained. This new set of subgroups is said to exist in addition to those listed in Fairbairn *et al.*[2] (FFK). Their structure is that of the semidirect product $z_m \circledS Z_3$ of two cyclic groups Z_m and Z_3, of order m and 3, respectively, where m must contain at least one prime factor of the form $(3k + 1)$, with k a positive integer.

If m contains a factor q which is the product of powers of primes which are equal neither to 3 nor $(3k + 1)$ then the subgroup of order $3m$ can be written as the direct product $(Z_{3^r p} \circledS Z_3) \otimes Z_q$, where $m = 3^r pq$ with r equal to zero or a positive integer and p is the product of powers of primes of the form $(3k + 1)$. Because the direct product of subgroups is automatically a subgroup and because Z_q is a trivial finite subgroup of SU(3) we need concentrate only on the semidirect products $(Z_{3^r p} \circledS Z_3)$ of order $3^{r+1}p$. BLW claim that this group is a finite subgroup of SU(3) for all integral r. We assert that it is a finite subgroup of SU(3) only for $r = 0$ and $r = 1$ and, for these values of r, it is a subgroup of $\Delta\,(3n^2)$ for an appropriate value of n. The groups Δ are defined in FFK, who call them "dihedral-like," while BLW refer to them as "trihedral".

II. DISCUSSION

A. For $r = 0$ the subgroup is isomorphic to $(Z_p \circledS Z_3)$. A three-dimensional defining (and irreducible) representation of this group is generated by the 3×3 unitary matrices

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

and

$$\begin{pmatrix} e^{2\pi i/p} & 0 & 0 \\ 0 & e^{2\pi i a/p} & 0 \\ 0 & 0 & e^{2\pi i a^2/p} \end{pmatrix},$$

where $(1 + a + a^2) = 0$, mod p. This is the representation [0,1] as defined by (2.3) of BLW.

The three-dimensional matrices which generate the subgroups given by FFK are enumerated in Table I of that paper. The generators for $\Delta\,(3n^2)$ are

$$E\,(0,0),\ A\left(\frac{2\pi}{n},\ 0\right) \quad \text{and} \quad A\left(0,\ \frac{2\pi}{n}\right),$$

where $E\,(0,0)$ is identical to the first (3×3) matrix listed above and

$$A\,(\alpha,\beta) = \begin{pmatrix} e^{i\alpha} & 0 & 0 \\ 0 & e^{i\beta} & 0 \\ 0 & 0 & e^{-i(\alpha+\beta)} \end{pmatrix}.$$

Thus, the second generator of $(Z_p \circledS Z_3)$ is $A\,(2\pi/p, 2\pi a/p)$ and is an element of $\Delta\,(3p^2)$. The subgroup isomorphic to $(Z_p \circledS Z_3)$ is therefore a subgroup of $\Delta\,(3p^2)$. The values of the number a for some smaller integral values of p of appropriate form are given in Table I.

B. For $r = 1$ the appropriate subgroup is isomorphic to $(Z_p \circledS Z_3) \otimes Z_3$. The generators of this group can be represented by (3×3) unitary matrices using again a positive integer a, which in this case satisfies $(1 + a + a^2) = 0$, mod $3p$. This three-dimensional representation of the group $(Z_{3p} \circledS Z_3)$ then contains the center of SU(3), which has three elements, and the group can be factored into the direct product $(Z_p \circledS Z_3) \otimes Z_3$. The smallest example of such a structure is the group $(Z_{21} \circledS Z_3)$ of order 63. Here $p = 7$ and $a = 4$. The generators are $E\,(0,0)$ and $A\,(2\pi/21, 8\pi/21)$, with both A^7 and A^{14} diagonal and equal to $e^{2\pi i/3}$ and $e^{4\pi i/3}$, respectively, times the three-dimensional unit matrix.

These SU(3) subgroups are in turn subgroups of $\Delta\,(27p^2)$ and are therefore subgroups of the set of groups $\Delta\,(3n^2)$ where n is a multiple of 3. It was noted by FFK that such groups did always contain the center of SU(3).

Other similar finite subgroups of not too large order have $p = 13$, $a = 16$ (order 117) and $p = 19$, $a = 7$ (order

TABLE I. Appropriate value of the integer a for some small values of the integer p.

p	7	13	19	31	$91 = (7\times 13)$	$133 = (7\times 19)$
a	2	3	7	5	9	11

171). Both are of the type $(Z_p \circledS Z_3) \otimes Z_3$, and both contain the center of SU(3).

C. For $r > 1$ it is claimed by BLW that a faithful three-dimensional representation can be found which defines a finite subgroup of SU(3) of order $3^{r+1}p$. They denote this subgroup by $G(3^r p, 3, a)$ and its defining representation is $[0, a-1]$, where the number a satisfies the condition of their Lemma 3.

One of these conditions is that $a^3 = 1$, mod $3^r p$, and the other is that $a(a-1) \neq (a-1)$, mod $3^r p$. Because $(a^3 - 1) = (a-1)(a^2 + a + 1) = 0$, mod $3^r p$ and because $(a^2 + a + 1)$ can have at most one factor 3 we can deduce that if $r > 1$ then $(a-1)$ contains at least one factor 3. It follows that the representation $[0, a-1]$ is not faithful because the factor $\omega^{ja^{(s-\alpha)}\beta}$, where $\omega = \exp(2\pi i/3^r p)$, will be identical for values of β differing by $3^{r-1}p$. There will, in fact, be greater duplication of matrices if $r > 2$. Since the defining representation of the proposed subgroup must be faithful, we see that the class of finite subgroups of SU(3) constructed in this way for $r > 1$ does not exist. The smallest such subgroup would be of the order $189 = 3^3 \times 7$. Possible values for a would be $a = 4$ or $a = 58$, for which $(a-1) = 3$ or $(a-1) = 57$; in both cases $(a-1)$ is divisible by 3. The orbits, of length three because the representation is three-dimensional, are 3, 12, 48 and 57, 30, 39, respectively. Again we notice that all of these numbers are divisible by 3 and only 63 distinct matrices are obtained to define a group of order 189. Similarly for $r = 2$, $p = 13$ (a group of order 351) an appropriate value is $a = 16$ and the orbit is 15, 6, 96; only 117 distinct (3×3) matrices are generated.

D. More recently Bovier and Wyler[3] have shown that the Hessian group of order 216 and its subgroups of order 72, 36, and 18 can be written in the form of semidirect products. For all of these groups the appropriate normal subgroup is $(Z_3 \otimes Z_3)$ of order 9 and for the group of order 216 the other factor is the double tetrahedral group T' (see FFK, p. 1043); for the group of order 72 the other factor is Q, the quaternion group of order 8. This group can be generated by the permutations (1234), (5678), (9) and (1537), (2846), (9) on nine symbols,[4] and the character tables (VII and VI in FFK) show explicitly the elements 14^2 and 12^4 of this type contained in these two finite subgroups of SU(3). Note that $T' = Q \circledS Z_3$.

Because the other three subgroup (of orders 36, 168, and 360) listed by FFK are simple, it is not possible to express them as either direct or semidirect products.

The $\Delta(6n^2)$ groups can be expressed as semidirect products. For even n, as stated by BLW, the order of the group is $6n^2$ and it is isomorphic to $(Z_n \otimes Z_n) \circledS S_3$, where S_3 is the symmetric group on three symbols of order 6. However, for odd n, as mentioned by FFK the group is of order $24n^2$ and it is isomorphic to $(Z_n \otimes Z_n) \circledS W$, where W is a group of order 24. This group is isomorphic to the semidirect product $V \circledS S_3$, where V is the well-known four-group. For both even and odd n, S_3 consists of the six elements $\{A(0,0), C(0,0), E(0,0), B(\pi,\pi), D(\pi,\pi), F(\pi,\pi)\}$ and for odd n, $V = \{A(0,0), A(0,\pi), A(\pi,0), A(\pi,\pi)\}$ where $E(0,0)$ and the 3×3 matrices $A(\alpha, \beta)$ have been defined previously, and the others are given in Tables I and VIII of FFK.

CONCLUSIONS

The finite groups $G(m, n, a)$ with $n = 3$ and $m = 3^r pq$, with p and q powers of appropriate primes (see Introduction), can be divided into two categories. Whereas all groups of this type have been proposed by BLW as a new class of finite subgroups of SU(3), we have shown that the first category, those with $r = 0$ or 1, are subgroups of an appropriate $\Delta(3n^2)$, in the notation of FFK. (It is of interest that the smallest such group of order 21 is a subgroup also of the group $\Sigma(168)$ listed by FFK.) The second category $(r > 1)$ does not define finite subgroups of SU(3).

It must be remarked that the analysis by BLW of the structure of the irreducible representations of both $\Delta(3n^2)$ and $\Delta(6n^2)$ is a new and considerable achievement and enables one, as they show, to write down the Clebsch–Gordan coefficients for these groups. Using these methods, it should also be possible to find the coefficients for the various subgroups of the "dihedral-like" Δ groups.

Note added in proof: It has been called to our attention that an erratum to Ref. 1 had been submitted by its authors, referring to the preprint version of this paper, and accepting the conclusions of Sec. C, above.

ACKNOWLEDGMENT

We would like to thank Professor V. Rittenberg for calling the work of BLW to our attention.

[1]A. Bovier, M. Luling, and D. Wyler, J. Math. Phys. **22**, 1543 (1981).
[2]W. M. Fairbairn, T. Fulton, and W. H. Klink, J. Math. Phys. **5**, 1038 (1964).
[3]A. Bovier and D. Wyler, J. Math. Phys. **22**, 2108 (1981).
[4]M. Hamermesh, *Group Theory and its Application to Physical Problems* (Addison-Wesley, Reading, Mass. 1963), p. 28.

Some spectral properties of the Kostant complex

Mircea Puta
University of Timişoara, Seminarul de Geometrie, Timişoara 1900, Romania

(Received 29 January 1981; accepted for publication 24 July 1981)

In this paper we establish some spectral properties of an elliptic complex introduced by Kostant in the context of geometric quantization.

PACS numbers: 02.40.Sf

I. INTRODUCTION

Kostant[1] and Souriau[2] independently developed a theory of geometric quantization. One seeks to associate differential operators with functions on a symplectic manifold so as to preserve as much as possible of the Poisson bracket structure of the functions. The Kostant–Souriau quantization is usually made in three steps

(*a*) *Prequantization*: Let (M,ω) be a symplectic manifold, L^{ω} a line bundle over M having a connection ∇^{ω} whose curvature is $2\pi i\omega$ and such that for any section s of L^{ω} and any pair ξ,η of vector fields on M the following relation is satisfied:

$$[\nabla^{\omega}_{\xi},\nabla^{\omega}_{\eta}]s - \nabla^{\omega}_{[\xi,\eta]}s = 2\pi i\omega(\xi,\eta)s.$$

This is only possible when ω is integral which means has integral periods over integral homology cycles in M. It is in fact the famous Weil lemma[3] modified by Kostant.[1] In the case when this condition is satisfied the set of isomorphism classes of such bundles can be identified with $H^1(M,S^1)$. (Here S^1 is the group of complex numbers of modules one.)

The prequantization of $C(M)$ is constructed on the space $\Gamma(L^{\omega})$ of smooth sections of L^{ω} as follows. With each function $\varphi \in C(M)$ we associate a first-order operator $\delta(\varphi)$: $\Gamma(L^{\omega}) \rightarrow \Gamma(L^{\omega})$ by setting

$$\delta(\varphi) = \nabla^{\omega}_{\xi_\varphi} + 2\pi i\varphi,$$

where ξ_φ is the Hamiltonian vector field associated to φ (i.e., $i_{\xi_\varphi}\omega = d\varphi$). δ is a homomorphism of Lie algebra where the operators on $\Gamma(L^{\omega})$ are given their usual commutator bracket Lie algebra structure.

(*b*) *Polarization*: Let (M,ω) be a symplectic manifold. A polarization of (M,ω) is a maximally isotropic involutive complex subtangent bundle F. If L^{ω} is a Hermitian line bundle on M as above, we denote by $\Gamma_F(L)$ the space of polarized sections of L (i.e., the space of smooth sections of the bundle $L = L^{\omega} \otimes N_F^{1/2}$ covariant constant along F, and where $N_F^{1/2}$ means the bundle of 1/2 forms normal to F.

(*c*) *Quantization*: Let C_F^1 be the Lie algebara under Poisson bracket of all functions on M whose Hamiltonian vector fields are infinitesimal automorphism of F. There is a natural Lie derivative action of ξ_φ in $\Gamma(N_F^{1/2})$ (see Ref. 4). Combining $\delta(\varphi)$ with this Lie derivative gives a differential operator $\delta_F(\varphi)$: $\Gamma(L) \rightarrow \Gamma(L)$ which preserves $\Gamma_F(L)$. This action of $\delta_F(\varphi)$ on $\Gamma_F(L)$ is known as quantization and is defined for any φ in C_F^1.

Other extensions of these notions can be founded in: Kostant,[5] Onofri and Pauri,[6,7] Rawnsley,[8,9] Renouard,[10] Simms,[11–13] Sniatycki.[14–16]

In connection with the geometric quantization, Fischer and Williams[17] introduced the notion of "complex foliated structure" and in a particular case they refined the Kostant complex.

In this note using this complex foliated structure we shall make some remarks on the spectral properties of the Kostant complex.

II. THE KOSTANT COMPLEX

Let M be an orientable, smooth ($= C^{\infty}$) paracompact manifold in $n+m$ dimensions. $C(M)$ shall denote the space of smooth functions on M, $T(M)$ shall denote the tangent, space of M, and $T(M)_{\mathrm{C}}$ its complexification.

Definition 2.1[7]: A "complex foliated structure" of M is a complex subbundle $F \subset T(M)_{\mathrm{C}}$ satisfying the following conditions:

(1) $F \cap \overline{F}$ is of constant rank;

(2) F and $F + \overline{F}$ are integrable.

We shall suppose in all that follows that rank $(F) = n$.

Choosing a direct summand $F^{\perp}$ of F in $T(M)_{\mathrm{C}}$ with respect to some Hermitian structure on $T(M)_{\mathrm{C}}$, we obtain

$$T(M)_{\mathrm{C}} = F \oplus F^{\perp}.$$

Examples: (1) If $T(M)_{\mathrm{C}} = F \oplus \overline{F}$, then, defining $J \in \mathrm{End}(T(M))$ as $-i$ on F and i on $\overline{F}$, M becomes a complex manifold.

(2) if M is a symplectic manifold and F is a polarization of M, then M is a Kähler manifold.

Let $\Omega_F^q \overset{\mathrm{def}}{=} \Omega^{(0,q)}$ be the space of differential forms on M of the type $(0,q)$. The "exterior derivative along F," d_F, is given by

$$d_F{:}\alpha \in \Omega_F^q \rightarrow d_F{:}\alpha \in \Omega_F^{q+1},$$

where for any vector fields along F, $X_1,\dots,X_{q+1}$, we take

$$d_F\alpha(X_1,\dots,X_{q+1}) = \sum_{i=1}^{q} (-1)^{i+1} X_i(\alpha(X_1,\dots,\widehat{X}_i,\dots,X_{q+1})) + \sum_{i<j} \alpha([X_i,X_j],X_1,\dots,\widehat{X}_i,\dots,\widehat{X}_j,\dots,X_{q+1})$$

Definition 2.2: The sheaves of forms $\{\underline{\Omega}_F^q\}$ and d_F now yield a sheaf complex which we shall call the Kostant complex.

It yields a fine resolution of the sheaf $\underline{\mathscr{A}}_F^0 =$

$$= \ker d_F^0 \subset \underline{\Omega}_F^0, 0 \rightarrow \underline{\mathscr{A}}_F^0 \xrightarrow{i} \underline{\Omega}_F^0 \xrightarrow{d_F} \underline{\Omega}_F^1 \xrightarrow{d_F^1} \cdots \underline{\Omega}_F^n \rightarrow 0$$

 0022-2488/82/101749-03$02.50

This result was pointed out by Kostant.[1] In a more general context it was obtained by Fischer and Williams[7] and a somewhat different proof may also be found in Rawnsley.[8]

Hereafter we shall assume that M is an orientable compact manifold and the complex foliated structure of M is elliptic (i.e., see Ref. 17: $T(M)_{\mathbb{C}} = F + \bar{F}$).

We choose a Riemannian metric on M and extend it to a Hermitian structure on $T(M)_{\mathbb{C}}$. Then the operator d_F has an adjoint d_F^* defined using the given Hermitian structure and the Kostant complex is an elliptic complex.

It is known that under these conditons the Hodge–De Rham theorem[18] guarantees that the cohomology groups are finite-dimensional. In particular, it makes sense to define the Euler–Poincaré characteristic associated with F:

$$\chi_F(M) = \sum_{q=0}^{n} (-1)^q \dim(\mathscr{H}_F^q),$$

where $\Delta_F^q = d_F^{q+1} d_F^* + d_F^* d_F^q$ and $\mathscr{H}_F^q = \ker(\Delta_F^q)$.

For any $0 \leqslant q \leqslant n$, Δ_F^q is an elliptic, autoadjoint positive operator. Then the F-spectrum of M denoted by $\mathrm{Spec}_F^q(M)$ is the set of eigenvalues of Δ_F^q, i.e., λ's such that there exist an $\omega \in \Omega_F^q$, $\omega \neq 0$, with $\Delta_F^q \omega = \lambda\omega$. We write

$$\mathrm{Spec}_F^q(M) = \{0 \leqslant \lambda_0 < \lambda_1 < \cdots \rightarrow +\infty\},$$

each λ being written a number of times equal to its multiplicity.

It is easy to see that the F-spectrum of M depends only on the foliated and Riemannian structure of M:

Let $V_F^q(\lambda) = \{\omega \in \Omega_F^q \mid \Delta_F^q \omega = \lambda\omega\}$

and $m_F^q(\lambda) = \dim V_F^q(\lambda)$ be the eigenspace of λ and the multiplicity of λ, respectively. It is easy to see that $\dim(\mathscr{H}_F^q) = m_F^q(0)$. We will adopt the convention $m_F^q(\lambda) = 0$ if $\lambda \neq \mathrm{Spec}_F^q(M)$.

Proposition 2.1 (*F-télescopage of McKean–Singer*): Let M be a smooth manifold as above and F an elliptic complex foliated structure on M. Then

$$\sum_{q=0}^{n} (-1)^q m_F^q(\lambda) = \begin{cases} \chi_F(M) & \text{if } \lambda = 0 \\ 0 & \text{if } \lambda \neq 0. \end{cases}$$

Proof: The first equality follows immediately from the Hodge–De Rham decomposition theorem. If $\lambda > 0$, then we can verify that the following identities hold:

$$V_F^q(\lambda) \cap d_F^{-1}(0) = V_F^q(\lambda) \cap d_F \Omega_F^{q-1},$$

$$V_F^q(\lambda) \cap d_F^{*-1}(0) = V_F^q(\lambda) \cap d_F^* \Omega_F^{q+1}.$$

Setting

$$A_F^q \overset{\text{def}}{=} V_F^q(\lambda) \cap d_F^{-1}(0),$$

$$B_F^q \overset{\text{def}}{=} V_F^q(\lambda) \cap d_F^{*-1}(0),$$

we obtain in a natural way the orthogonal decomposition of the $V_F^q(\lambda)$:

$$V_F^q(\lambda) = A_F^q(\lambda) \oplus B_F^q(\lambda).$$

On the other hand, $d_F: B_F^{q-1}(\lambda) \rightarrow A_F^q(\lambda)$ is an isomorphism and therefore

$$\sum_{q=0}^{n} (-1)^q m_F^q(\lambda) = 0.$$

Proposition 2.2: For any $\lambda > 0$, $1 \leqslant q \leqslant n-1$, we have

(1) $m_F^{q+1}(\lambda) + m_F^{q-1}(\lambda) \geqslant m_F^q(\lambda)$

(2) $m_F^1(\lambda) \geqslant m_F^0(\lambda)$.

Proof: Let Z_F^q be the space of d_F-closed q-forms, $\{\omega_1,\ldots,\omega_r\}$ be a basis for $Z_F^q \cap V_F^q(\lambda)$, and $\{\omega_{r+1},\ldots,\omega_k\}$ be a basis for the orthogonal complement of $Z_F^q \cap V_F^q(\lambda)$ in $V_F^q(\lambda)$ so that $m_F^q(\lambda) = k$; since Δ_F^q commutes with d_F^*, $\{d_F^*\omega_1,\ldots,d_F^*\omega_r\}$ is contained in $V_F^{q-1}(\lambda)$. In fact, it is claimed that it is a linearly independent set. Therefore,

$$m_F^{q-1}(\lambda) \geqslant r.$$

A similar arguemnt shows that

$$m_F^{q+1}(\lambda) \geqslant k - r.$$

Adding these inequalities yields immediately the desired result. Q.E.D.

Proposition 2.3: Let M be as above and $n = 2k$. If, for any q, $0 \leqslant q \leqslant n$, we have

$$\mathrm{Spec}_F^q(M) = \mathrm{Spec}_F^{n-q}(M),$$

then $m_F^k(\lambda)$ is an even number.

Proof: Using Proposition 2.1, we have

$$0 = \sum_{q,\ \text{even}} m_F^q(\lambda) - \sum_{q,\ \text{odd}} m_F^q(\lambda)$$

and then

$$m_F^k(\lambda) = (-1)^{k+1} 2 \left\{ \sum_{\substack{q \geqslant 0 \\ q,\ \text{even}}}^{q<k} m_F^q(\lambda) - \sum_{\substack{q > 0 \\ q,\ \text{odd}}}^{q<k} m_F^q(\lambda) \right\}.$$

Q.E.D.

In view of the general theory of elliptic complexes the Minakshisundaram–Pleijel theorem[19] can be extended in a natural way to $\mathrm{Spec}_F^q(M)$. More precisely, setting $Z_F^q(t) \overset{\text{def}}{=} \Sigma_{i=0}^{\infty} e^{-\lambda i t}$, we have

Proposition 2.4: For any $q = 0,1,2,\ldots,Z_F^q(t)$ has the following asymptotic development:

$$Z_F^q(t) \underset{t\to 0}{\sim} (4\pi t)^{-n/2}(a_0^q(F) + t a_1^q(F) \cdots)$$

As a consequence of this propositon and of the F-télescopage formula, we can prove:

Proposition 2.5: Let M be as above. Then $\chi_F(M) = 0$ if and only if for any i we have

$$\sum_{q=0}^{n} (-1)^q a_i^q(F) = 0.$$

Proof: Indeed we can write

$$\sum_{q=0}^{n} (-1)^q Z_F^q(t) = \sum_{\lambda} \left[\sum_{q=0}^{n} (-1)^q m_F^q(\lambda) \right] e^{-\lambda t} = \chi_F(M).$$

Therefore,

$$\chi_F(M) \underset{t\to 0}{\sim} (4\pi t)^{-n/2} \left[\sum_{q=0}^{n} a_0^q(F) + \cdots + t^i \sum_{q=0}^{m} (-1)^q a_i^q(F) + \cdots \right]$$

from which the announced result. Q.E.D.

III. KOSTANT COMPLEX OF A LINE BUNDLE L

In this section we shall make some remarks on complex line bundles in view of their applications in geometric quantization.

Let M be an orientable, compact smooth manifold with a Riemannian metric, F an elliptic complex foliated structure on M, and L a line bundle over M such that it is F-holomorphic and F is compatible with the linear connection of L [i.e., $\nabla_F s = 0$, for any $s \in \Gamma(L)$]. We write $\Gamma_F(L)$ for the F-holomorphic sections of L, and $\Omega^q_L(L)$ for $\Omega^{0,q}(L) = \Omega^q_F \otimes \Gamma_F(L)$ (see Refs. 8, 13).

Under these assumptions the sheves of bundle values forms $\{\underline{\Omega}^q_F(L)\}$ yields an elliptic sheaf complex, which we shall call the Kostant complex of a line bundle L. It yields a fine resolution of the sheaf $\mathscr{A}^0_{\nabla_F}(L) = \{s \in \Gamma(L) | \nabla_F s = 0\}$ $0 \rightarrow \underline{\mathscr{A}}^0_{D_F}(L) \rightarrow \underline{\Omega}^n_F(L) \rightarrow \cdots \rightarrow \underline{\Omega}^n_F(L) \rightarrow 0$.

Some spectral properties of this complex can be founded in the following propositions:

Proposition 3.1: Under the above restrictions we have

$$\sum_{q=0}^{n} (-1)^q m^q_F(\lambda, L) = \begin{cases} \chi_F(M,L) & \text{if } \lambda = 0, \\ 0 & \text{if } \lambda \neq 0. \end{cases}$$

Proposition 3.2: For any q, $1 \leqslant q \leqslant n-1$ we have

(1) $m^{q+1}_F(\lambda, L) + m^{q-1}_F(\lambda, L) \geqslant m^q_F(\lambda, L)$,

(2) $m^1_F(\lambda, L) \geqslant m^0_F(\lambda, L)$,

where $m^q_F(\lambda, L) = \dim\{\omega \in \Omega^q_F(L) | \Delta^q_F \omega = \lambda\omega\}$.

Remark: There is an open and very tempting problem, to see if the geometric quantization is completey determined by the spectrum of the Kostant complex?

[1]B. Kostant, *Lecture Notes in Mathematics*, Vol. 170 (Springer–Verlag, Berlin, Heidelberg, New York, 1974).
[2]J. M. Souriau, *Structure des systems dynamiques* (Dunod, Paris, 1970).
[3]A. Weil, *Variété Kähleriennes* (Herman, Paris, 1958).
[4]R. J. Blattner, Colloque Symplectic, Aix-en-Provence, 1974.
[5]B. Kostant, Colloque Symplectique, Aix-en-Provence, 1974.
[6]E. Onofri and M. Pauri, J. Math. Phys. **13**, 533–43 (1972).
[7]E. Onofri, "Quantization theory for homogeneous Kähler manifolds," preprint, Parma (1974).
[8]J. H. Rawnsley, Trans. Math. Soc. **230**, 235–55 (1977).
[9]J. H. Rawnsley, "Some applications of quantization," thesis, Oxford University (1972).
[10]P. Renouard, "Variétés symplectiques et quantification," thesis, Orsay (1969).
[11]D. J. Simms, International Symposium Mathematical Physics, Warsaw, 1974.
[12]D. J. Simms, Colloque Symplectique, Aix-en-Provence, 1974.
[13]D. J. Simms, Rep. Math. Phys. **12**, 213–17 (1977).
[14]J. Sniatycki, Rep. Math. Phys. **7**, 127–35 (1974).
[15]J. Sniatycki, Third International Colloqium on Group Theoretical Methods in Physics, Marseille, 1974.
[16]J. Sniatycki, Int. J. Theoret. Phys. **14**, 277–88 (1975).
[17]H. R. Fischer and F. L. Williams, Trans. Am. Math. Soc. **252**, 163–95 (1979).
[18]G. De Rham, *Variétés differentiables* (Herman, Paris, 1955).
[19]M. Berger, P. Gauduchon, and E. Mazet, *Le spectre d'une variété Riemannienne, Lecture Notes in Mathematics*, Vol. 194 (Springer–Verlag, Berlin, Heidelberg, New York, 1974).

Analytic continuation from data points with unequal errors

S. Ciulli[a] and T. D. Spearman
School of Mathematics, Trinity College, Dublin, Ireland

(Received 16 April 1981; accepted for publication 20 July 1981)

The problem solved in this paper is that of constructing zero-free holomorphic functions which will (a) assume specified values at a set of discrete data points in a data region Γ_1 inside the holomorphy domain, and which at the same time will (b) provide the optimum solution to various stabilizing (boundedness or smoothness) conditions on the boundaries Γ_R. The immediate motivation to this problem arose from the need to renormalize data with unequal errors by using a holomorphic weight function to bring all the errors to the same value: this was a preliminary step before making an analytic continuation off the data region Γ_1. Since a stabilizing condition has to be imposed on the boundaries Γ_R, the weight function must be chosen so as to introduce the minimum additional instability on Γ_R. Although this was the specific motivation, other interesting applications suggest themselves and some of these are discussed. The stability conditions on Γ_R which are treated may all be expressed in terms of the real parts, or of the normal derivative of the real parts. Γ_1 is taken to be on the real axis and the functions considered satisfy a reflection principle which means that the data values are real. It follows that the results obtained may be expressed in terms of the real parts alone—in other words the problem solved here, is in fact that of obtaining harmonic functions which take specific values inside their harmonicity domain and which satisfy the appropriate extremum condition on the boundary.

PACS numbers: 02.60.Ed

1. INTRODUCTION

We shall consider extremum problems of the following type: A finite set $\{z_i\}$ of data points is given on the interval $(-1,1)$. Real values a_i are assigned to each of these points. It is then required to construct a complex function $X(z)$ which will have the following properties:

(i) it will be holomorphic in the unit disk $|z| < 1$ and will satisfy $X(z) = \overline{X}(\bar{z})$;

(ii) it will assume the specified values a_i: $X(z_i) = a_i$;

(iii) it will satisfy some extremum requirements on the unit circle $|z| = 1$, namely *either*

(*Problem A*):

$$\frac{1}{2\pi}\int_0^{2\pi} |\mathrm{Re}\, X(e^{i\phi})|^2 \sigma(\phi)\, d\phi \to \text{least},$$

where $\sigma(\phi)$ is some given positive and even function of ϕ, or

(*Problem B*):

$$\frac{1}{2\pi}\int_0^{2\pi} \Big| \frac{d}{d\phi}\,\mathrm{Im}\, X(e^{i\phi})\Big|^2 \sigma(\phi)\, d\phi \to \text{least}.$$

Such extremal functions could be valuable tools in strong interactions phenomenology or theory, in quantum chromodynamics, or in other branches of physics where one wishes to construct scattering amplitudes, form factors, vacuum polarization tensors,[1] or any other function of interest with known holomorphy domain, either from experimental or theoretical[2] data available at some given points z_i inside the holomorphy domain.

However, from a mathematical point of view such a continuation is highly unstable, both because of the finite number of the data points and of the uncertainties—theoretical[3] or experimental—of the data. Indeed, it is well known that in continuations off open contours, the errors propagate in an explosive and extremely anisotropic way in the function space under consideration,[4] as they are enhanced by factors which grow progressively (usually exponentially) in the far dimensions of this function space. Therefore, to make the extrapolation stable, we have to add some supplementary information such as boundedness or smoothness, which mainly has the effect of confining the output function to a region of the function space which is progressively flattened along the higher dimensions. In this way, the number of dimensions of the function space which really matter is very much reduced; this effectively counteracts the instabilities of the extrapolation, provided of course that one uses at the same time an adequate continuation technique.

The method of accelerated convergence expansions (ACE) introduced by Cutkosky and his collaborators[5,6] provides such a continuation technique. The ACE are polynomial expansions $P_n^{(F)}(w)$, in some special variable $w(s) = \frac{1}{2}(W(s) + 1/W(s))$ (see Refs. 5 and 6) which is chosen to optimize the convergence of these polynomials along some given (open) curve Γ_1 to functions $F(s)$ analytic in the cut s plane. The cuts will be denoted by Γ_R; in the $z = z(s)$ plane Γ_R is mapped onto the unit circle. A data function $D(s)$ representing $F(s)$, together with its associated errors $\epsilon(s)$, is given on Γ_1; notice that the *exact* expansion polynomials $P_n^{(F)}(w)$ of $F(s)$ are as unknown as the exact function $F(s)$ itself is. The unique analytic functions at hand are the computer-constructed polynomials $P_n^{(D)}(w)$ which best fit the (error-affected) data $D(s)$ on Γ_1. It can then be shown[6] that one can find an upper bound E_n for the deviations of these polynomials $P_n^{(D)}(w)$ with respect to the unknown function $F(s)$ at any point $s = s_p$. If the errors of the data $D(s)$ on Γ_1 are constant [if $\epsilon(s) = \bar{\epsilon}$], and if η_n is the rate of convergence *on the cuts*

[a]Present address: School of Theoretical Physics, Dublin Institute for Advanced Studies, Dublin, Ireland.

Γ_R of the unknown expansion $P_n^{(F)}(w)$ to $F(s)$, then the error bound E_n consists of two terms

$$E_n = 3\eta_n(\rho(s)/R)^n + 2\bar{\epsilon}\rho(s)^n \qquad (1)$$

where $\rho(s) = |W(s_P)|$ depends on the position $s = s_P$ of the point where the extrapolation is carried out. Since the optimal variable $W(s)$ maps the analyticity $F(s)$ onto an annular region such that the data region Γ_1 maps onto the circle $|W| = 1$ and the cuts Γ_R onto the circle $|W| = R$, $\rho(s)$ is in general greater than one, and hence the second term of Eq. (1) explodes when n increases too much

The optimal polynomial extrapolation $P_{n_{\rm opt}}^{(D)}(w)$ is obtained[7] for precisely that $n = n_{\rm opt}$ for which E_n reaches its minimum; it is obvious that the value of $n_{\rm opt}$ depends strongly both on the magnitude of the imprecisions ϵ of the data on Γ_1 *and* on the error-bound η_n which measures the way in which the exact polynomials $P_n^{(F)}(w)$ converge to the function $F(s)$ on the cuts Γ_R. It is clear that the smoother the function $F(s)$ on Γ_R, the more rapid will be the convergence of the $P_n^{(F)}(w)$ there, and hence the smaller η_n will be. Since according to the theory of maximally converging polynomials on Γ_1 we have $|P_n^{(F)}(w) - F(w(s))|_{\Gamma_1} < \eta_n/R^n$, it follows that for a smooth $F(s)$ one should be able to approximate the data $D(s)$ on Γ_1 with low-order (with low n) polynomials $P_n^{(D)}(w)$. Hence the exploding term $\bar{\epsilon}\rho(s)^n$ of Eq. (1) may also be kept small, and thus one takes full advantage of the optimality of the variable $w(s)$: Indeed, the latter is optimal only for n greater than some $N_{\rm asmp}$, since the optimality of the $P_n^{(F)}(w)$ expansion of $F(s)$ is proven only in the asymptotic case.

The value $n = N_{\rm asmp}$ where this asymptotic behavior begins depends in an essential way on the smoothness of $F(s)$. In the following we shall refer to the situation when $n_{\rm opt} > N_{\rm asmp}$ as the "near asymptotic situation" (NAS), to be contrasted with the "far asymptotic situation" (FAS) to be discussed later.

Let us begin with the NAS case, when, as shown, one is able to take full advantage of the qualities of the accelerated convergence expansions. One should, however, remember that the ACE, as well as the theory of maximally converging polynomials[8] on which they are based, were primarily meant for the constant error case. The theory, as it stands, may be extended also to the nonuniform error case, but this involves some loss of information since the error propagation is governed by the maximum modulus theorem[9] and by other[10] maximum principles. It is important, therefore, to be able to reduce the nonconstant error case to the constant one without loss of information, by means of a suitable holomorphic and zero-free weight function.

The main motivation behind the present paper was to solve this problem.

Case α (*NAS case*): An important use of the functions $X(z)$ which will be constructed in this paper will be to provide the zero-free weight function

$$C(z) = \exp\{X(z)\} \qquad (2)$$

which will transform a set of data with unequal errors into a new set $\widetilde{D}(z) = C(z)D(z)$ with a constant error $\bar{\epsilon}$, while the boundedness/smoothness condition on the cuts, which provides the stability of the whole approach, is altered as little as possible. Note that Problem A means in fact a condition for the modulus $|C(z)| = \exp\{\mathrm{Re}\, X(z)\}$ on the cuts,[11] while Problem B limits the variation of the phase of $C(z)$. Notice also that $C(z) = \exp\{X(z)\}$ has no cuts at all in the data region Γ_1 and does not alter the analyticity domain of the initial problem.

If the function $F(s)$ has a lot of "structure" so that the asymptotic behavior of η_n is reached only for $n > N_{\rm asmp}$ where $N_{\rm asmp}$ is large, then, in order to get $n_{\rm opt} > N_{\rm asmp}$ (i.e., to have the NAS case again), one has to have extremely precise data (very small ϵ's). If this condition is satisfied then one can proceed as above and use the function $C(z)$ in the standard way, (case α) to bring the errors to a constant value.

But this might not be the case. The FAS situation will arise if, on the one hand, the data have quite a lot of structure (for instance if they appear to have an exponential-like forward peak) so that it is obvious that they could not be approximated satisfactorily by a low degree polynomial $P_n^{(D)}(w)$, and if, on the other hand, ϵ is not small enough to prevent $\epsilon\rho^n$ from becoming exceptionally large. In this case one way to proceed would be as follows:

Case β (*FAS case*): A weight function $C_{\rm str}(z)$ could be constructed to contain as much of the structure of the initial data as we choose to remove, leaving a relatively smooth weighted data function $D_{\rm sm}(z) = C_{\rm str}(z)D(z)$, while preserving as far as possible the boundedness or smoothness on the cuts in order to make η_n fall as quickly as possible. Again, the role of the conditions A and B is obvious. Notice also that, in contrast with some factors like plain exponentials which we might have used to remove structure, the function $C_{\rm str} = \exp\{X(z)\}$ has the advantage that it does not introduce any spurious singularities (in particular essential singularities) which might be theoretically unacceptable. On the contrary, by its very method of construction, it is the most bounded or smoothest function on Γ_R which one can find having the prescribed structure in the physical region Γ_1.

A few words upon other possible applications of the solutions of the problems A and B will be given. Because $X(z) = \overline{X}(\bar{z})$, the conditions $X(z_i) = a_i$ are in fact conditions satisfied by the *harmonic* function $\mathrm{Re}\, X(z)$. Moreover, we may use the Cauchy equations in Problem B to make the replacement

$$\frac{\partial \mathrm{Im}\, X(z)}{\partial \phi} = \frac{\partial \mathrm{Re}\, X(z)}{\partial r}$$

on the boundary Γ_R, and hence both Problems A and B are in fact extremal problems for the *harmonic* function $\mathrm{Re} X(z)$ alone. The solutions of these extremal harmonic-functions problems are likely to be of value in potential theory or in dealing with the types of inverse problems which arise, for example, in geophysics and in heat-transfer theory.

2. USE OF DUALITY

In this section we show how the duality theorem, which is a direct consequence of the Hahn–Banach lemma, enables us to reformulate and solve the extremum problems A and B. A derivation of the duality theorem is given in Appendix A. We shall apply this procedure initially to Problem A, as this is the easier of the two cases to handle.

A. Notation

We shall use capitals, e.g., $X(z)$, $Y(z)$, $M(z)\cdots$ to represent analytic functions, and correspondingly subscripted symbols $X_{\text{Re}}(z)$, $X_{\text{Im}}(z),\cdots$, to denote their real and imaginary parts. z' will be used to denote points on the unit circle $z' = e^{i\phi}$, and we shall frequently write

$$X_{\text{Re}}(z') \equiv X_{\text{Re}}(e^{i\phi}) \equiv x(\phi), \tag{3}$$

using the lower case letter to denote real functions of ϕ obtained as shown. The functions $x(\phi)$ will always be periodic so that $x(2\pi - \phi) = x(-\phi)$; since $X(z) = \overline{X}(\bar{z})$, $x(\phi)$ is even, $x(\phi) = x(-\phi)$. Further, $\{z_i\} = \{\text{Re}\, z_i\}$ is the set of given points on the real axis and $\{a_i\}$ the real values specified.

It will become apparent as the calculation proceeds that the boundary value functions $x(\phi)$ are of central importance both for defining linear functionals and also norms for $X(z)$, each chosen in such a way as to suit the extremum problem under consideration. For instance, one may define a norm for $F(z)$ related to the L^2 norm of $f(\phi)$

$$\|F(z)\| = \frac{1}{2\pi}\int_0^{2\pi} (f(\phi))^2 \sigma(\phi)\, d\phi, \tag{4}$$

where, following the above notation,

$$f(\phi) = F_{\text{Re}}(e^{i\phi}) = \text{Re}\, F(e^{i\phi}), \tag{5}$$

and where $\sigma(\phi)$ is a real, positive weight function satisfying the condition

$$\sigma(\phi) = \sigma(-\phi). \tag{6}$$

In Sec. 4. we shall also consider spaces of analytic functions whose norms are related to the (tangential) derivative of the boundary values of their imaginary part, as these are relevant for problems in which solutions of maximum smoothness (least phase variation) are sought.

We shall also have to deal with linear functionals Y^* acting on the analytic functions $X(z)$. Since the functions $X(z)$ may be expressed linearly in terms of their boundary values $x(\phi)$, the functionals Y^* may be seen as functionals y^* acting on these boundary value functions. The Riesz theorem (see Appendix A) may then be used to associate each linear functional y^* with a real function $y(\phi)$ (even in ϕ) as follows:

$$\langle X, Y^*\rangle \equiv \langle x, y^*\rangle = \frac{1}{2\pi}\int_0^{2\pi} y(\phi) x(\phi) \sigma(\phi)\, d\phi, \tag{7}$$

where we have introduced the following notation for functionals,

$$Y^* \equiv \langle \cdot, Y^*\rangle.$$

For more details, the reader is referred to Appendix A.

B. Formulation of Problem A; Duality

The objective is to construct a function $X(z)$ with the properties

(i) $X(z_i) = a_i$ where the points z_i are real and the values a_i are also real;

(ii) $X(z)$ is holomorphic in the unit disk, and $X(\bar{z}) = \overline{X}(z)$;

(iii) subject to (i) and (ii) above, $X(z)$ should satisfy the condition that $\|X(z)\|$ should have the least possible value. Here the norm $\|X\|$ is defined according to which of the conditions A or B we wish to implement.

We shall first solve Problem A and return later to Problem B, which will be solved in Sec. 4. The procedure we adopt is the following. A particular function $X^{(1)}(z)$ is constructed to possess properties (i) and (ii); this is really an easy task. For instance we may make the choice

$$X^{(1)}(z) = \sum_i a_i \frac{(z - z_1)(z - z_2)\cdots\widehat{(z - z_i)}\cdots(z - z_n)}{(z_i - z_1)(z_i - z_2)\cdots\widehat{(z_i - z_i)}\cdots(z_i - z_n)} \tag{8}$$

with the convention that the factors marked $\widehat{\ }$ are to be deleted. Now if $M(z)$ is any function which has value 0 at each of the points z_i, is holomorphic in the unit disk and satisfies $\overline{M}(z) = \overline{M}(z)$, then the function $X^{(1)}(z) - M(z)$ also possesses properties (i) and (ii); conversely, any function with those two properties may be represented in this form. So the function $X^{(0)}(z)$ possessing properties (i), (ii), and (iii), (which gives the solution to our problem) is

$$X^{(0)}(z) \equiv X^{(1)}(z) - M^{(0)}(z), \tag{9}$$

where $M^{(0)}(z)$ is the solution to the minimization problem

$$\delta_0 = \inf_M \|X^{(1)}(z) - M(z)\|, \tag{10}$$

the minimization being with respect to the class of functions $M(z)$ which satisfy

$$\begin{aligned} &M(z) \text{ holomorphic for } |z| < 1,\\ &M(\bar{z}) = \overline{M}(z),\\ &M(z_i) = 0, \quad i = 1,\ldots,n. \end{aligned} \tag{11}$$

The duality theorem (see Appendix A), when applied to this minimization problem reads

$$\delta_0 = \inf_M \|X^{(1)} - M\| = \sup_{Y^*} \langle X^{(1)}, Y^*\rangle, \tag{12}$$

where the functional $\langle ., Y^*\rangle$ is as defined in Eq. (7) and the extremum problem is now with respect to the class of functionals $\langle ., Y^*\rangle$ which satisfy the conditions

$$\text{(a)}\quad \frac{1}{2\pi}\int_0^{2\pi} (y(\phi))^2 \sigma(\phi)\, d\phi = 1, \tag{13a}$$

$$\text{(b)}\ \langle M, Y^*\rangle = 0, \tag{13b}$$

for all $M(z)$ satisfying the conditions (11),

$$\text{(c)}\ y(\phi) = y(-\phi). \tag{13c}$$

C. An explicit representation for the functionals $\langle ., Y^*\rangle$

We need to identify the class of functionals satisfying conditions (13), but before doing this it is necessary to look more closely at the class of functions $M(z)$. Associated with each holomorphic function $M(z)$ there is a real function $m(\phi)$, which, following the notation we have established, is

$$m(\phi) = M_{\text{R}}(e^{i\phi}) \equiv \text{Re}\, M(e^{i\phi}). \tag{14}$$

Conversely, once $m(\phi)$ is specified {any real, even [i.e., $m(\phi) = m(-\phi)$] square-integrable function on $[0,2\pi]$}, the analytic function $M(z)$ is completely[12] determined and may be expressed as

$$M(z) = \frac{1}{2\pi}\int_0^{2\pi} \frac{e^{i\phi} + z}{e^{i\phi} - z}\, m(\phi)\, d\phi. \tag{15}$$

Equation (15) is the Schwartz–Villat formula, which is simply the complex extension of the well-known Poisson integral by means of which harmonic functions $M_R(z)$ are constructed from their boundary values:

$$M_R(z) = \frac{1}{2\pi}\int_o^{2\pi} P(z,z')m(\phi)\,d\phi. \tag{16a}$$

Here $P(z,z')$ is the Poisson kernel

$$P(z,z') = \mathrm{Re}\left(\frac{e^{i\phi}+z}{e^{i\phi}-z}\right) \equiv \frac{1-r^2}{1+r^2-2r\cos(\theta-\phi)}. \tag{16b}$$

Consider now the set of functionals $\langle .,Y^*\rangle$ defined by the following special functions:

$$y(\phi) = \sum_i y_i P(z_i,e^{i\phi})(\sigma(\phi))^{-1}, \tag{17}$$

where the y_i are arbitrary real constants. We first observe that

$$\begin{aligned}\langle M,Y^*\rangle &= \frac{1}{2\pi}\sum_i y_i \int P(z_i,e^{i\phi})(\sigma(\phi))^{-1}m(\phi)\sigma(\phi)\,d\phi \\ &= \sum_i y_i M_R(z_i) \\ &= 0\end{aligned} \tag{18}$$

since $M_R(z_i) = \mathrm{Re}\, M(z_i) = 0$.

We have thus shown that any functional $\langle .,Y^*\rangle$, constructed by means of the functions $y(\phi)$ defined in Eq. (17), automatically satisfies the requirements (13b). We shall prove at the end of this section that conversely, any linear functional $\langle .,Y^*\rangle$ satisfying (13b) can be expressed in terms of a function $y(\phi)$ having the form (17). But before doing this consider first the normalization (13a), which becomes

$$\sum_{ij}\alpha_{ij}y_iy_j = 1, \tag{19}$$

where the constants α_{ij} are

$$\alpha_{ij} = \frac{1}{2\pi}\int_0^{2\pi} P(z_i,e^{i\phi})P(z_j,e^{i\phi})\sigma(\phi)^{-1}\,d\phi. \tag{20}$$

To enable us to evaluate this integral explicitly we first introduce a holomorphic function $S(z)$ whose real part has the value $(\sigma(\phi))^{-1}$ when $z' = e^{i\phi}$. This may be done immediately using the Schwartz–Villat formula

$$S(z) = \frac{1}{2\pi}\int_o^{2\pi}\frac{e^{i\phi}+z}{e^{i\phi}-z}\frac{1}{\sigma(\phi)}\,d\phi. \tag{21}$$

The weight function $\sigma(\phi)$ was restricted to be a positive function so that $(\sigma(\phi))^{-1}$ is bounded. It is also required to satisfy the condition $\sigma(\phi) = \sigma(-\phi)$ [Eq. (8)] from which it follows that

$$S(\bar z) = \bar S(z),$$

and in particular we note that when z is real so is $S(z)$, so that for each of the points z_i,$S(z_i)$ is real. Again because $\sigma(\phi) = \sigma(-\phi)$, we may write

$$\begin{aligned}S(z) &= \frac{1}{2\pi}\int_o^{\pi}\left\{\frac{e^{i\phi}+z}{e^{i\phi}-z} + \frac{e^{-i\phi}+z}{e^{-i\phi}-z}\right\}\frac{1}{\sigma(\phi)}\,d\phi \\ &= \frac{1}{\pi}\int_0^{\pi}\frac{(1-z^2)}{1-2z\cos\phi+z^2}\frac{1}{\sigma(\phi)}\,d\phi.\end{aligned} \tag{22}$$

For real z with $|z|<1$ the integrand is positive since $\sigma(\phi)$ is positive and it follows that $S(z)$ is positive too.

As an alternative to Eqs. (21) and (22), $S(z)$ may be constructed by expanding $\sigma^{-1}(\phi)$ in a Fourier series (only the cosine terms are present, since $\sigma^{-1}(\phi)$ is even),

$$\sigma^{-1}(\phi) = s_0 + \sum_1^{\infty} s_n\cos(n\phi),$$

$$s_n \equiv \frac{2-\delta_{no}}{\pi}\int_0^{\pi} d\phi\,\cos(n\phi)\sigma^{-1}(\phi), \tag{23a}$$

and then writing $S(z)$ as

$$S(z) = \sum_0^{\infty} s_n e^{in\phi}. \tag{23b}$$

From Eqs. (20) and (21) we obtain

$$\begin{aligned}\alpha_{ij} &= \frac{1}{2\pi}\int_0^{2\pi} P(z_i,e^{i\phi})P(z_j,e^{i\phi})(\sigma(\phi))^{-1}\,d\phi \\ &= \mathrm{Re}\left\{\frac{1}{2\pi}\int_0^{2\pi} P(z_i,e^{i\phi})P(z_j,e^{i\phi})S(z'\equiv e^{i\phi})\,d\phi\right\}.\end{aligned} \tag{24}$$

Now,

$$P(z_i,z') = \mathrm{Re}\left(\frac{z'+z_i}{z'-z_i}\right) = \tfrac{1}{2}\left\{\frac{z'+z_i}{z'-z_i} + \frac{\bar z'+z_i}{\bar z'-z_i}\right\}, \tag{25}$$

since the z_i are real. Further, since $z' = e^{i\phi}$, Eq. (25) becomes

$$P(z_i,z') = \tfrac{1}{2}\left(\frac{z'+z_i}{z'-z_i} + \frac{1+z'z_i}{1-z'z_i}\right). \tag{26}$$

Substitution into Eq. (24) yields

$$\begin{aligned}\alpha_{ij} = \tfrac{1}{4}\mathrm{Re}\Bigg[\frac{1}{2\pi i}\int_C &\left(\frac{z'+z_i}{z'-z_i} + \frac{1+z'z_i}{1-z'z_i}\right) \\ &\times\left(\frac{z'+z_j}{z'-z_j} + \frac{1+z'z_j}{1-z'z_j}\right)S(z')\frac{dz'}{z'}\Bigg].\end{aligned} \tag{27}$$

Deforming the contour C around the poles and noting that the residue at $z'=0$ vanishes we get

$$\begin{aligned}\alpha_{ij} &= \tfrac{1}{4}\mathrm{Re}\Bigg(\frac{S(z_i)}{z_i}2z_i\left(\frac{z_i+z_j}{z_i-z_j} + \frac{1+z_iz_j}{1-z_iz_j}\right) \\ &\quad - \frac{S(z_j)}{z_j}2z_j\left(\frac{z_i+z_j}{z_i-z_j} - \frac{1+z_iz_j}{1-z_iz_j}\right)\Bigg) \\ &= \tfrac{1}{2}\left\{\frac{S(z_i)-S(z_j)}{z_i-z_j}(z_i+z_j) + (S(z_i)+S(z_j))\left(\frac{1+z_iz_j}{1-z_iz_j}\right)\right\}.\end{aligned} \tag{28a}$$

For $i=j$ the result of the integration is $(S'\equiv dS/dz)$

$$\alpha_{ii} = z_iS'(z_i) + S(z_i)((1+z_i^2)/(1-z_i^2)). \tag{28b}$$

Note that if the weight function $\sigma(\phi)$ were a constant, which we would set to be 1 so that $S(z_i) = S(z_j) = 1$, we would have for all i and j,

$$\alpha_{ij} = (1+z_iz_j)/(1-z_iz_j). \tag{29}$$

In this case, all the coefficients α_{ij} are clearly positive; in fact, since $\Sigma\alpha_{ij}y_iy_j$ is *a norm* and hence cannot vanish unless all y_i are identically zero, the matrix α_{ij} is always positive definite and $\Sigma\alpha_{ij}y_iy_j = 1$ represents an ellipsoid.

We shall now show that Eq. (17) includes all possible functionals satisfying Eqs. (13b) and (13c). For if this is not so, suppose that $\langle .,\Gamma^*\rangle$ is a functional which is not included in the set defined by Eq. (17) but which none the less satisfies

(13b), that is,

$$\langle M,\Gamma^*\rangle \equiv \frac{1}{2\pi}\int_0^{2\pi} \gamma(\phi)m(\phi)\sigma(\phi)\,d\phi = 0 \qquad (30)$$

for all $M(z)$ satisfying the conditions (11).

Since all z_i occurring in Eq. (11) are real and hence the imaginary part of $M(z)$ vanishes there identically, and remembering that $M_R(z)$ may be expressed in terms of $m(\phi)$ by means of the Poisson integral (16a), it is equivalent to say that the integral from Eq. (30) has to vanish for all $m(\phi)$ satisfying

$$\frac{1}{2\pi}\int_0^{2\pi} m(\phi)P(z_i,e^{i\phi})\,d\phi = 0, \quad i = 1, 2,...,n, \qquad (31)$$
$$m(\phi) = m(-\phi).$$

Now define

$$\tilde{\gamma}_i = \frac{1}{2\pi}\int_0^{2\pi} \gamma(\phi)P(z_i,e^{i\phi})\,d\phi \qquad (32)$$

and

$$\gamma^\perp(\phi) = \gamma(\phi) - \sum_{i,j=1}^{n} \tilde{\gamma}_i(\alpha^{-1})_{ij}P(z_j,e^{i\phi})(\sigma(\phi))^{-1} \qquad (33)$$
$$= \gamma(\phi) - \sum_j \gamma_j P(z_j,e^{i\phi})(\sigma(\phi))^{-1},$$

where

$$\gamma_j \equiv \sum_k \tilde{\gamma}_k(\alpha^{-1})_{kj}. \qquad (34)$$

In order that the function $\gamma(\phi)$ should have a form different from that of the $y(\phi)$'s, (Eq. 17) it is necessary that $\gamma^\perp(\phi)$ should not be identically zero. But from Eq. (33) one sees that for any $z = z_k$ [cf. also (32)],

$$\frac{1}{2\pi}\int_0^{2\pi} \gamma^\perp(\phi)P(z_k,e^{i\phi})d\phi$$
$$= \tilde{\gamma}_k - \sum_{ij}\tilde{\gamma}_i(\alpha^{-1})_{ij}\frac{1}{2\pi}\int_0^{2\pi} P(z_j,e^{i\phi})P(z_k,e^{i\phi})\frac{d\phi}{\sigma(\phi)}$$
$$= \tilde{\gamma}_k - \sum_{ij}\tilde{\gamma}_i(\alpha^{-1})_{ij}\alpha_{jk} = 0, \qquad (35)$$

so that the function $\gamma^\perp(\phi)$ satisfies all the necessary requirements [see Eq. (31); note also that $\gamma^\perp(\phi)$ is even] "to be an $m(\phi)$" defining a holomorphic function $M(z)$ satisfying all the conditions (11). Hence, putting $m_1(\phi) \equiv \gamma^\perp(\phi)$, we get

$$\langle M_1,\Gamma^*\rangle = \frac{1}{2\pi}\int_0^{2\pi} \gamma(\phi)m_1(\phi)\sigma(\phi)\,d\phi$$
$$= \frac{1}{2\pi}\int_0^{2\pi} \gamma^\perp(\phi)m_1(\phi)\sigma(\phi)\,d\phi$$
$$= \frac{1}{2\pi}\int_0^{2\pi} (m_1(\phi))^2\sigma(\phi)\,d\phi > 0; \qquad (36)$$

[unless $\gamma^\perp(\phi)$ is identically zero] a result which is in direct contradiction with the conditions (13b). Hence *$\gamma^\perp(\phi)$ has to vanish identically.* We have thus shown that the set of functionals $\langle .,Y^*\rangle$ defined by Eq. (7) and (17), is precisely the set of functionals satisfying Eqs. (13b) and (13c).

3. EXPLICIT SOLUTION OF THE EXTREMUM PROBLEM A

We wish to determine δ_0, where

$$\delta_0 \equiv \inf_M \|X^{(1)} - M\|, \qquad (37)$$

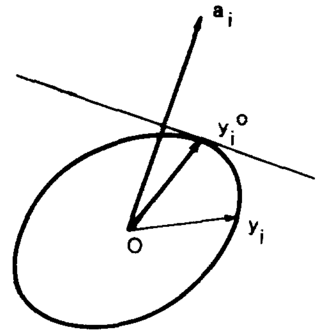


FIG. 1. The ellipsoid $\Sigma\alpha_{ij}y_iy_j = 1$ and the vector $y_i^{(0)}$ whose projection on a_i is largest.

and we also want to know the function $M^{(0)}$ giving the minimum, so that we will have obtained the function $X^{(0)} \equiv X^{(1)} - M^{(0)}$. We saw from Eq. (12) that the extremum problem (37) could be replaced by

$$\delta_0 = \sup_{Y^*}\langle X^{(1)},Y^*\rangle, \qquad (38)$$

where the supremum is taken with respect to the set of functionals satisfying Eqs. (13a)–(13b). We have been able to specify this set of functionals and to show that it can be represented by the set of functions $y(\phi)$, defined in Eq. (17). When we substitute for $\langle .,Y^*\rangle$ in Eq. (38) we see that

$$\delta_0 = \frac{1}{2\pi}\sup_{y_i}\sum_i y_i\int_0^{2\pi} P(z_i,e^{i\phi})(\sigma(\phi))^{-1}(\phi)\sigma(\phi)d\phi$$
$$= \sup_{y_i}\Sigma y_i a_i, \qquad (39)$$

since the integrals appearing here yield by definition the values of $X^{(1)}(z)$ at $z = z_i$, i.e., the constants a_i. On the other hand, the coefficients y_i must satisfy also the condition

$$\sum_{ij}\alpha_{ij}y_iy_j = 1, \qquad (19)$$

where the constants α_{ij} are given by Eqs. (28a) and (28b).

In geometric terms the problem is illustrated in Fig. 1. Since α_{ij} is positive-definite it represents an n-dimensional ellipsoid, and we look for that vector y_i on the ellipsoid whose component in the direction **a** is a maximum.

We can solve the problem analytically using Lagrange multipliers. We put

$$\Phi = \sum_i y_ia_i - \lambda\left(\sum_{ij}\alpha_{ij}y_iy_j - 1\right) \qquad (40)$$

and differentiate to get

$$\frac{\partial\Phi}{\partial y_i} = a_i - 2\lambda\sum_j\alpha_{ij}y_j = 0. \qquad (41)$$

It follows that

$$y_i = \frac{1}{2\lambda}\sum_j(\alpha^{-1})_{ij}a_j. \qquad (42)$$

λ is determined from Eq. (19) by substituting from Eq. (42) for y_i; the result is

$$\lambda = \tfrac{1}{2}\left\{\sum_{ij}(\alpha^{-1})_{ij}a_ia_j\right\}^{1/2}. \qquad (43)$$

The required vector y_i is

$$y_i^{(0)} = \Sigma(\alpha^{-1})_{ij}a_j/\{\Sigma(\alpha^{-1})_{ij}a_ia_j\}^{1/2}, \tag{44}$$

and so the corresponding optimal functional $\langle .,y^{(0)*}\rangle$ is defined by the function

$$y^{(0)}(\phi) = \Sigma y_i^{(0)}P(z_i,e^{i\phi})(\sigma(\phi))^{-1}. \tag{45}$$

Substitution from Eq. (44) into Eq. (39) gives the value of δ_0:

$$\delta_0 = \{\Sigma(\alpha^{-1})_{ij}a_ia_j\}^{1/2}. \tag{46}$$

The fact that we have already an explicit form [Eqs. (44) and (45)] for $y^{(0)}(\phi)$ allows us to determine also the optimal function $M^{(0)}(z)$: indeed, if the extremum is realized with $M^{(0)}$ and $y^{(0)}$, we simply have

$$\delta_0 = \|X^{(1)} - M^{(0)}\| = \langle X^{(1)},Y^{(0)*}\rangle = \langle X^{(1)} - M^{(0)},Y^{(0)*}\rangle, \tag{47}$$

where the last step follows from Eq. (13b). But [put $x^0(\phi) \equiv x^1(\phi) - m^0(\phi)$] from Schwarz's inequality we have usually [see also the definition (A3)]

$$\langle x^0,y^0\rangle < \|x^0\|\,\|y^0\| \equiv \|x^0\| \tag{48}$$

(the norm of $\|y^0\| \equiv 1$) unless $x^0(\phi) = \text{const}\cdot y^{(0)}(\phi)$, when equality occurs. Hence, Eq. (47) tells that $x^0(\phi)$ and $y^0(\phi)$ should be "aligned"; further, since [again (47)] $\|x^{(0)}\| = \delta_0$ and $\|y^{(0)}\| = 1$, const equals δ_0 and hence,

$$x^{(1)}(\phi) - m^{(0)}(\phi) \equiv x^{(0)}(\phi) = \delta_0 y^{(0)}(\phi). \tag{49}$$

So the optimal function $x^{(0)}(\phi)$ can be written in terms of known entities alone :

$$x^{(0)}(\phi) = \sum_{i,j}(\alpha^{-1})_{ij}a_jP(z_i,e^{i\phi})(\sigma(\phi))^{-1}. \tag{50}$$

Finally, the corresponding complex function $X^{(0)}(z)$ is

$$X^{(0)}(z) = \sum_{i,j}(\alpha^{-1})_{ij}a_j\frac{1}{2\pi} \times \int_0^{2\pi}\left(\frac{e^{i\phi}+z}{e^{i\phi}-z}\right)P(z_i,e^{i\phi})(\sigma(\phi))^{-1}\,d\phi. \tag{51}$$

We can immediately verify from Eq. (51) using Eq. (24), that, as must be the case, $X^{(0)}(z_i)$ does indeed have the value a_i.

A modified extremum problem (Problem A')

In Sec. 1 we referred to the function $C(z) \equiv \exp\{X(z)\}$. In specifying the values of $C(z)$ at the points x_i we are often only concerned with the ratios of the values $C(x_i), C(x_j),\cdots$, that is, we may replace $C(x_i) = c_i$ by $C(x_i) = \gamma c_i$. For $X(z)$ this means that the value a_i of $X(z)$ at x_i may be replaced by $a_i + \alpha$. This modified extremum problem takes the form

$$\delta_0 = \inf_{\alpha,M}\|X^{(1)} - \alpha - M\| = \sup_{Y^*}\langle X^{(1)},Y^*\rangle, \tag{52}$$

where the functions M are defined as before and α is a constant. The infimum is with respect to all the functions M *and* all possible constants α. Since the set of functions $M(z) + \alpha$ is larger than that of problem A, in evaluating the supremum the functionals $\langle .,Y^*\rangle$ must satisfy a further restriction in addition to Eqs. (13a)–(13b), namely ,

$$\langle \alpha,Y^*\rangle = 0\,. \tag{53}$$

If we take the set of functionals defined by Eq. (17) then the additional condition (53) becomes

$$\sum_i y_i = 0. \tag{54}$$

This may be treated as an additional auxiliary condition and as such is incorporated through a second Lagrange multiplier μ. We write, in this case,

$$\Phi = \sum_i y_ia_i - \lambda\,(\sum_{i,j}\alpha_{ij}y_iy_j - 1) - \mu\sum_i y_i. \tag{55}$$

Then

$$\frac{\partial\Phi}{\partial y_i} = a_i - \mu - 2\lambda\sum_j\alpha_{ij}y_j = 0, \tag{56}$$

which gives

$$y_i = \frac{1}{2\lambda}\sum_j(\alpha^{-1})_{ij}(a_j - \mu). \tag{57}$$

Equation (54) may be used to eliminate μ; substituting from Eq. (57) yields

$$\mu = \sum_{ij}(\alpha^{-1})_{ij}a_j\Big/\sum_{ij}(\alpha^{-1})_{ij}. \tag{58}$$

If we introduce the vector $a_i' = a_i - \mu$, with the constant μ given by Eq. (58), then Eqs. (43) and (44) give us λ and $y_i^{(0)}$ provided that we replace a_i by a_i'. To evaluate δ_0 we write, using Eq. (54),

$$\delta_0 = \sum_i y_i^{(0)}a_i = \sum_i y_i^{(0)}a_i', \tag{59}$$

and substitution for $y_i^{(0)}$ gives

$$\delta_0 = \left\{\sum_{ij}(\alpha^{-1})_{ij}a_i'a_j'\right\}^{1/2}. \tag{60}$$

Proceeding as before we obtain the results

$$x^{(0)}(\phi) = \sum_{ij}(\alpha^{-1})_{ij}a_j'P(z_i,e^{i\phi})(\sigma(\phi))^{-1}, \tag{61}$$

$$X^{(0)}(z) = \sum_{ij}(\alpha^{-1})_{ij}a_j'\frac{1}{2\pi}\int_0^{2\pi}\frac{e^{i\phi}+z}{e^{i\phi}-z}P(z_i,e^{i\phi})(\sigma(\phi))^{-1}\,d\phi. \tag{62}$$

Again we observe that Eq. (62) gives

$$X^{(0)}(z_i) = a_i' \equiv a_i - \sum_{mn}(\alpha^{-1})_{mn}a_n/\sum_{mn}(\alpha^{-1})_{mn}. \tag{63}$$

It is, of course, clear that the relaxation of the original problem introduced here, and the corresponding restriction of the supremum problem, must lead to a value of δ_0 as given by Eq. (60) which is less than that in Eq. (46). This result may also be seen from a direct comparison of the right-hand sides of the two equations. Equation (54) allows us to write

$$\sum_{ij}(\alpha^{-1})_{ij}a_ia_j = \sum_{ij}(\alpha^{-1})_{ij}a_i'a_j' + \mu^2\sum_{ij}(\alpha^{-1})_{ij}. \tag{64}$$

Since α_{ij} is positive definite, so is $(\alpha^{-1})_{ij}$, and hence, the second term on the right of Eq. (64) is positive and will only be zero if $\mu = 0$.

IV. THE NEUMANN BOUNDARY CONDITION

In this section we consider the alternative problem where the selection criterion A, is replaced by B (see the

Introduction). The integral, taken over the unit circle, which we wish to minimize is

$$\frac{1}{2\pi}\int_0^{2\pi}\left|\frac{\partial}{\partial\phi}\operatorname{Im}X(e^{i\phi})\right|^2\sigma(\phi)\,d\phi. \tag{65}$$

Now the Cauchy–Riemann relations imply that

$$\frac{1}{r}\frac{\partial X_{\rm Im}}{\partial\phi}=\frac{\partial X_{\rm R}}{\partial r}=\operatorname{Re}\left(\frac{dX}{dz}\frac{\partial z}{\partial r}\right)\equiv\frac{1}{r}\operatorname{Re}(X'(z)z), \tag{66}$$

where $z=re^{i\phi}$. Equation (66) allows us to write the optimization condition B as

$$\frac{1}{2\pi}\int_0^{2\pi}\left|\frac{\partial X_{\rm R}(re^{i\phi})}{\partial r}\right|^2\sigma(\phi)\,d\phi\,|_{r=1}\to\text{least.} \tag{67}$$

In order to proceed we need to be able to construct a complex function $X(z)$ which is holomorphic in the unit disk, when the radial derivative of the real part, $\partial X_{\rm R}/\partial r$, is specified on the unit circle. This Neumann type problem may be solved in terms of a Green's function analogous to the Poisson kernel of Eq. (16b). The Neumann kernel is derived in Appendix B where we see that a subtraction is required. If we introduce the following notation

$$\left.\frac{\partial F_{\rm R}(re^{i\phi})}{\partial r}\right|_{r=1}=f_{,r}(\phi), \tag{68}$$

using $f_{,r}(\phi)$ to denote the radial derivative of the real part of the function $F(z)$ at the point $z=e^{i\phi}$, the required result takes the form [Cf. Eqs. (B8)–(B10)]

$$F_{\rm R}(z)=F_{\rm R}(z_0)+\frac{1}{2\pi}\int_0^{2\pi}f_{,r}(\phi)N(z_0;z,e^{i\phi})\,d\phi, \tag{69}$$

where

$$\begin{aligned}N(z_0;z,e^{i\phi})&=-2\ln\left|\frac{e^{i\phi}-z}{e^{i\phi}-z_0}\right|\\&=-2\operatorname{Re}\left\{\ln\left(\frac{e^{i\phi}-z}{e^{i\phi}-z_0}\right)\right\}.\end{aligned} \tag{70}$$

Further, the complex function $F(z)$ is given by

$$F(z)=F(z_0)-\frac{1}{\pi}\int_0^{2\pi}f_{,r}(\phi)\ln\left(\frac{e^{i\phi}-z}{e^{i\phi}-z_0}\right)d\phi. \tag{71}$$

If we are to proceed in analogy with the Dirichlet case, the next step should be to try to define a norm for $F(z)$ by means of the boundary function $f_{,r}(\phi)$:

$$\|F(z)\|\overset{(?)}{=}\left\{\frac{1}{2\pi}\int_0^{2\pi}(f_{,r}(\phi))^2\sigma(\phi)\,d\phi\right\}^{1/2}. \tag{72}$$

The difficulty which immediately faces us is that because of the subtraction in Eq. (71), the right-hand side of (72) does not define a valid norm for $F(z)$ since the latter might be $\neq 0$ even if the right-hand side of (72) is zero.

Fortunately we can circumvent this difficulty by restricting ourselves to the space[13] $\{F(z)\}$ of $F(z)$ vanishing at $z=z_0$. Moreover, we choose one of the points z_i at which the values a_i of $X(z)$ are prescribed as subtraction point z_0; to be specific we take $z_0\equiv z_1$. Indeed, we may immediately see that the optimization condition, Eq. (67), is not altered if we replace the initial function $X(z)$ by $X(z)-a$; that is, if we replace the set of values $a_1,\dots,a_n$ by $0, a_2-a_1,\dots,a_n-a_1$. Further, the functions $M(z)$ will be required, as usual, to be zero at all the points $z_1,\dots,z_n$. In this way we have a unique function $F(z)$ associated with each real radial derivative function $f_{,r}(\phi)$:

$$F(z)=-\frac{1}{\pi}\int_0^{2\pi}f_{,r}(\phi)\ln\left(\frac{e^{i\phi}-z}{e^{i\phi}-z_1}\right)d\phi \tag{73}$$

and the integral from the right-hand side of Eq. (72) is now indeed *a norm* for the $F(z)$'s.

We proceed as before and we start with a function $X^{(2)}(z)$ defined to be holomorphic and to take the values $0,a_2-a_1,\dots,a_n-a_1$ at the points z_i. Specifically we choose

$$X^{(2)}(z)=X^{(1)}(z)-a_1, \tag{74}$$

where $X^{(1)}(z)$ is defined in Eq. (8). We then want to solve the infimum problem

$$\delta_0\equiv\inf_M\|X^{(2)}-M\|, \tag{75}$$

where the norm is defined as in Eq. (72). The infimum is with respect to the set of functions $M(z)$ defined in Eq. (11). The function $M(z)$ which gives the least value of δ will be denoted by $M^{(0)}(z)$, and the corresponding $X(z)$ by $X^{(0)}(z)$;

$$X^{(0)}(z)=X^{(2)}(z)-M^{(0)}(z). \tag{76}$$

As before we use the Duality theorem to replace the infimum problem by a supremum one

$$\delta_0\equiv\inf_M\|X^{(2)}-M\|=\sup_{Y^*}\langle X^{(2)},Y^*\rangle. \tag{77}$$

The supremum is with respect to the set of functionals $\langle .,Y^*\rangle$ defined by Eqs. (13a)–(13c). We shall show that in this case the set of functionals satisfying Eqs. (13a)–(13c) is precisely the set defined by

$$y(\phi)=\sum_{i=2}^{n}y_iN(z_1;z_i,e^{i\phi})(\sigma(\phi))^{-1}, \tag{78}$$

where the constants y_i are real and take all possible values subject to the normalization condition (13a). The summation is from 2 to n since

$$N(z_1;z_1,e^{i\phi})=0. \tag{79}$$

To show that the set of functionals with respect to which the supremum is to be taken, is that defined by Eq. (78) we proceed in much the same way as in Sec. 2. Some care is needed however, particularly in view of the effective elimination of the point z_1. We first observe that for any $y(\phi)$ given by Eq. (78),

$$\begin{aligned}\langle M,Y^*\rangle&=\sum_{i=2}^{n}y_i\frac{1}{2\pi}\int_0^{2\pi}N(z_1;z_i,e^{i\phi})m_{,r}(\phi)\,d\phi &(80)\\&=\sum_{i=2}^{n}y_iM_{\rm R}(z_i) &(81)\\&=0.\end{aligned}$$

Now consider the normalization condition (13a). When we substitute for $y(\phi)$ from Eq. (78), Eq. (13a) becomes

$$\sum_{i,j=2}^{n}\alpha_{ij}y_iy_j=1, \tag{82}$$

where

$$\begin{aligned}\alpha_{ij}=\frac{1}{2\pi}\int_0^{2\pi}&N(z_1;z_ie^{i\phi})\\&\times N(z_1;z_j,e^{i\phi})(\sigma(\phi))^{-1}\,d\phi.\end{aligned} \tag{83}$$

Note that $\alpha_{1j} = \alpha_{j1} = 0$. The constants α_{ij} are evaluated in Appendix C. Since

$$\sum_{i,j=2}^{n} \alpha_{ij} y_i y_j = \frac{1}{2\pi}\int_0^{2\pi} |y(\phi)|^2 \sigma(\phi)\, d\phi \qquad (84)$$

it follows that the $n-1$ by $n-1$ matrix α_{ij} (where i and j range from 2 to n) is positive definite and the surface given by Eq. (82) is ellipsoidal.

Following an argument similar to that from Sec. 2 [Eqs. (30)–(36)], one can show that the set of functionals defined by Eq. (78) is precisely the set of functionals satisfying Eqs. (13b) and (13c).

Explicit solution of the extremum problem for the Neumann case

The extremum problem expressed by Eq. (77) may now be written as

$$\delta_0 = \sup \sum_{i=2}^{n} y_i \frac{1}{2\pi}\int_0^{2\pi} N(z_1;z_i,e^{i\phi}) x^{(2)}{}_{,r}(\phi)\, d\phi \qquad (85)$$

$$= \sup \sum_{i=2}^{n} y_i (a_i - a_1).$$

The coefficients y_i must satisfy Eq. (82):

$$\sum_{i,j=2}^{n} \alpha_{ij} y_i y_j = 1, \qquad (82)$$

where the coefficients α_{ij} are constants whose values are given by Eq. (C11). In geometric and analytic terms this problem is completely analogous to that described by Eqs. (39) and (19) (see Sec. 3) except that it is now $n-1$ dimensional. In Fig. 1 the ellipsoid is now in an $(n-1)$-dimensional space and the n vector a_i is replaced by the $(n-1)$-dimensional vector $(a_i - a_1)$, $i = 2$ to n. The extremum calculation, using a Lagrange multiplier to take account of Eq. (82), yields the value of δ_0 and the functional $\langle ., Y^{(0)*}\rangle$ which gives the supremum

$$\delta_0 = \sup \langle X^{(2)}, Y^*\rangle = \langle X^{(2)}, Y^{(0)*}\rangle. \qquad (86)$$

The results obtained are

$$\delta_0 = \Big(\sum_{i,j=2}^{n} (\alpha^{-1})_{ij} (a_i - a_1)(a_j - a_1)\Big)^{1/2}, \qquad (87)$$

and $\langle ., Y^{(0)*}\rangle$ is defined by means of the function

$$y^{(0)}(\phi) = \sum_{i,j=2}^{n} y_i{}^{(0)} N(z_1,z_i,e^{i\phi})(\sigma(\phi))^{-1}, \qquad (88)$$

where the coefficients $y_i{}^{(0)}$ have the values

$$y_i{}^{(0)} = \sum_{i,j=2}^{n} (\alpha^{-1})_{ij}(a_j - a_1) \Big/ \Big(\sum_{i,j=2}^{n} (\alpha^{-1})_{ij}(a_i - a_1)(a_j - a_1)\Big)^{1/2}. \qquad (89)$$

The required function $X^{(0)}(z) \equiv X^{(2)}(z) - M^{(0)}(z)$ may now be determined. We observe that

$$\delta_0 = \|X^{(2)} - M^{(0)}\| = \langle X^{(2)}, Y^{(0)*}\rangle = \langle X^{(0)}, Y^{(0)*}\rangle, \qquad (90)$$

where the last step follows from Eq. (13b). Now the equality between norm and functional

$$\|X^{(0)}\| = \langle X^{(0)}, Y^{(0)*}\rangle = \delta_0, \qquad (91)$$

together with the fact that the functional $\langle ., Y^{(0)*}\rangle$ has unit norm implies, by the Schwarz inequality, that

$$x^{(0)}{}_{,r}(\phi) = \delta_0 y^{(0)}(\phi).$$

So we have the following result:

$$x^{(0)}{}_{,r}(\phi) = \sum_{i,j=2}^{n} (\alpha^{-1})_{ij}(a_j - a_1) N(z_1;z_i e^{i\phi})(\sigma(\phi))^{-1} \qquad (92)$$

and the corresponding complex function $X^{(0)}(z)$ is

$$X^{(0)}(z) = \sum_{i,j=2}^{n} (\alpha^{-1})_{ij}(a_j - a_1) \times \frac{1}{2\pi}\int_0^{2\pi}\Big\{-2\ln\Big(\frac{e^{i\phi} - z}{e^{i\phi} - z_1}\Big) N(z_1;z_i e^{i\phi})(\sigma(\phi))^{-1}\, d\phi\Big\}. \qquad (93)$$

Equations (87), (92), and (93) represent the solution to the problem with the Neumann type boundary condition. Because of the choice made in the derivation, to single out the point z_1, this solution is not symmetic in form between the points z_i. It is clear, however, from the derivation that this solution has no specific dependence on z_1 or a_1; in other words, if any other one of the points z_i were selected instead of z_1 the result would be the same, although this is not manifestly evident from the form of the solution given above. Note that the matrix α_{ij} is also explicitly dependent on z_1. It is interesting and also of some practical value to recast this solution in a form which involves each of the z_i in an equivalent way and is thus manifestly symmetric in form under interchanges of the $\{z_i\}$.

We have seen that the set of functionals with respect to which the supremum of Eq. (77) is determined, is defined by Eq. (78) where the y_i range over all real values subject to the normalization condition (13a). We rewrite Eq. (78) as

$$y(\phi) = \sum_{i \neq 1} y_{(1)i} N(z_1;z_i,e^{i\phi})(\sigma(\phi))^{-1}, \qquad (94)$$

where the subscript (1) is inserted to indicate the special role of z_1 in $N(z_1;z_i,e^{i\phi})$. The normalization condition (82) will similarly be written as

$$\sum_{i,j \neq 1} \alpha_{ij}^{(1)} y_{(1)i} y_{(1)j}, \qquad (95)$$

emphasizing the fact that the coefficients $\alpha_{ij}^{(1)}$ depend in a special way on z_1 [Eq. (83)]. We now introduce a point z_0 on the real axis which does not coincide with any of the z_i. It now turns out that each $y(\phi)$ given by Eq. (94) can be expressed in the form

$$y(\phi) = \sum_{i=1}^{n} y_{(0)i} N(z_0;z_i,e^{i\phi})(\sigma(\phi))^{-1}. \qquad (96)$$

This follows from the relationship

$$N(z_1;z_i,e^{i\phi}) = N(z_0;z_i,e^{i\phi}) - N(z_0;z_1,e^{i\phi}). \qquad (97)$$

When Eq. (97) is substituted into Eq. (94) the result takes the form of Eq. (96), the coefficients $y_{(0)i}$ being related to the $y_{(1)i}$ as follows:

$$y_{(0)i} = y_{(1)i} \quad \text{for } i \neq 1,$$

$$y_{(0)1} = -\sum_{i \neq 1} y_{(1)i}. \qquad (98)$$

Equation (98) implies that

$$\sum_{i=1}^{n} y_{(0)i} = 0. \qquad (99)$$

This is an essential constraint on the $y_{(0)i}$: it comes from the fact that only $n-1$ of the $y_{(1)i}$ contribute, since $y_{(1)1}$ plays no role, and hence only $n-1$ of the $y_{(0)i}$ should be linearly independent. The normalization constraint Equation (95) comes from

$$\frac{1}{2\pi}\int_0^{2\pi}(y(\phi))^2\sigma(\phi)d\phi = 1, \tag{13a}$$

and this may be expressed in terms of the coefficients $y_{(0)i}$ by substituting from Eq. (96) for $y(\phi)$. The result is

$$\sum_{i,j=1}^{n}\alpha_{ij}^{(0)}y_{(0)i}y_{(0)j} = 1, \tag{100}$$

where

$$\alpha_{ij}^{(0)} = \frac{1}{2\pi}\int_0^{2\pi}N(z_0;z_i,e^{i\phi})N(z_0;z_j,e^{i\phi})(\sigma(\phi))^{-1}\,d\phi. \tag{101}$$

Equation (98) allows us to rewrite Eq. (85) in terms of the coefficients $y_{(0)i}$. Note the following equality

$$\sum_{i=1}^{n}y_{(0)i}a_i = \sum_{i\neq 1}y_{(1)i}a_i + \Big(-\sum_{1\neq 1}y_{(1)i}\Big)a_1$$
$$= \sum y_{(1)i}(a_i - a_1). \tag{102}$$

The supremum problem of Eq. (85), where the $y_{(1)i}(i\neq 1)$ take all real values subject to the normalization condition (95) may now be expressed in terms of the $y_{(0)i}$ as follows:

$$\delta_0 = \sup\sum_{i=1}^{n}y_{(0)i}a_i, \tag{103}$$

where the constants $y_{(0)i}$ take all real values subject to the two constraints

$$\sum_{i=1}^{n}y_{(0)i} = 0, \tag{104}$$

$$\sum_{i,j=1}^{n}\alpha_{ij}^{(0)}y_{(0)i}y_{(0)j} = 1. \tag{105}$$

5. CONCLUSION

In this paper we have been concerned with the problem of constructing zero-free holomorphic functions which assume specified values at some finite set of data points. Defining the function on a finite data set is, of course, a nonunique prescription, and further specification is required if a particular function is to be selected. The additional constraints which have been considered here and which are of particular interest, correspond to stabilizing conditions on the analytic continuation process which typically are boundedness or smoothness requirements. Hence, the following optimization problems have been solved: Values are specified at a finite set of points within a data region Γ_1. A boundedness or smoothness condition is specified on the cuts Γ_R and we look for the holomorphic function which takes the specified values in Γ, *and* which best satisfies the stabilizing condition on Γ_R. Since we are looking for zero-free functions we worked directly with their logarithm. Minimal conditions involving the real part of the latter or the tangential derivative of its imaginary part, were treated separately. A solution expressed in closed form is obtained in each case.

The initial motivation for solving this problem was in connection with the method of accelerated convergence expansions (ACE) devices by Cutkosky *et al.*[5–6] for constructing optimum polynomial expansions in order to perform an analytic continuation from a region Γ_1 to other points of the analytically domain, subject to stability constraints on the cuts Γ_R. In order to apply this method to a set of data with unequal errors, without incurring an unncessary loss of information, we need to construct a weight function which allows us to renormalize the data so that the errors will all assume the same value. This must be achieved however, without introducing additional instability in Γ_R. The problem of constructing an appropriate weight function is precisely that which has been solved here.

An important by-product of the above procedure is an explicit numerical expression for δ_0, the minimal distance (δ) between some given function $X^{(1)}$ assuming the specified values a_i at $z = z_i$ and the function $m(z)$, vanishing at each z_i. In other words, depending on the norm in use (either problem A or B), the value of δ_0 represents the least L^2-norm for the real part of $X(z)$ or for the (tangential) derivative of its imaginary part still compatible with the data and with the analyticity of $X(z)$. Hence this quantity (δ_0) could be a sensitive device for detecting bumps due to resonances and for use in other similar problems. Cutkosky was the first[14] to recognize, ten years ago, the importance of supplementing the usual χ^2 test used in fitting data by including a term which is related to the predictive power of the continuation procedure under consideration. This led to his well-known[14] "modified χ^2 test." The quantity δ_0 gives a measure of the quality of the functions in terms of the stability criterion and we shall show in a later paper how the results obtained here may be used to solve the continuation problem, without involving an expansion procedure.

Although the problem has been posed in terms of analytic continuation and the construction of holomorphic functions, it is important to observe that the results could equally well have been expressed in terms of harmonic functions. This is possible because first of all the stability conditions which have been considered may all be expressed in terms of the real parts of the function on Γ_R. Then secondly, the data region Γ_1 is taken to be a segment of the real axis and the holomorphic functions are required to possess the reflection symmetry $X(z) = X(\bar{z})$, which means that the data values are real. Consequently the results obtained for holomorphic functions may be expressed in terms of the real parts alone. These are real harmonic functions and so the problem which we have solved is in fact that of obtaining harmonic functions which take specific boundary values on Γ_1 and which satisfy various minimum conditions on Γ_R.

APPENDIX A: THE HAHN-BANACH LEMMA AND THE DUALITY THEOREM

The functional-analytic techniques used in this paper depend on a theorem, a consequence of the Hahn–Banach lemma, which allows nonlinear optimization problems expressed in terms of norms to be recast into linear integral form. This "Duality theorem" states that

$$\delta_0 \equiv \inf_{m\in M}\|x-m\| = \sup_{\substack{y^*\{\|y^*\|=1 \text{ and}\\ \langle m,y^*\rangle = 0 \text{ for all } m\in M\}}}\langle x,y^*\rangle, \tag{A1}$$

where $\|x-m\|$ is the distance δ between the function $x(\phi)$, *not* contained in the function subspace M, and the function $m(\phi)$ from M. Here $\langle .,y^*\rangle$ (we use the mathematicians' way of denoting linear functionals, which is the reverse of Dirac's one) is a linear functional which "annihilates" every function $m(\phi)$ from M. But, at least for the spaces in which we shall be interested [the L^P spaces, see below Eq. (A4)], such a functional can always be expressed as an integral (*F. Riesz*)

$$\langle .,y^*\rangle \equiv \frac{1}{b-a}\int_a^b d\phi\, y(\phi)\cdot(\phi) \tag{A2}$$

(replace the dot "·" by the name of any L^P function). Hence, in order to determine the functional y^*, it is sufficient to give the corresponding function $y(\phi)$.

The norm *of a linear functional* y^* acting on some functions $n(\phi)\in N$ is defined by means of the numbers[15] $\langle n,y^*\rangle$ as follows:

$$\|y^*\|_N \equiv \sup_{n\in N}\{\langle n,y^*\rangle/\|n\|\}, \tag{A3}$$

where $\|n\|$ is the norm of the function $n(\phi)$. However, in a practical situation one does not use the definition (A3) directly, but takes advantage of some theorems which in some cases (e.g., L^P spaces, see below) provide simple formulae for the computation of the functional norms (A3). Indeed, if the function space N coincides with an L^P function space, i.e., if the norm of the *functions* $n(\phi)$ are defined to be

$$\|n\| = \left\{\frac{1}{b-a}\int_a^b d\phi\, |n(\phi)|^p\right\}^{1/p}, \quad 1<p<\infty, \tag{A4}$$

the norm (A3) of the corresponding linear *functionals* y^* is then given by the simple formula

$$\|y^*\| = \left\{\frac{1}{b-a}\int_a^b d\phi\, |y(\phi)|^q\right\}^{1/q} \quad \text{with } \frac{1}{q} = 1-\frac{1}{p}. \tag{A5}$$

Hence, if the functions $m(\phi)$ in (A1) are of class L^P, the set of functionals with respect to which the supremum in Eq. (A1) is taken, is defined [see (A2)] in terms of *functions* $y(\phi)$, which satisfy the two conditions

$$\frac{1}{b-a}\int_a^b d\phi\, y(\phi)m(\phi) = 0 \quad [\text{for all } m(\phi)\in M], \tag{A6}$$

$$\frac{1}{b-a}\int_a^b d\phi\, |y(\phi)|^q = 1. \tag{A7}$$

$\mathscr{A}$. The Hahn–Banach theorem

Before proving the duality relation (A1), we shall sketch here a proof of the Hahn–Banach theorem for a separable space X, i.e., for a space in which there exists a dense countable set of points.

The Hahn–Banach theorem states that, given a *linear* functional y_N^* defined on some subspace N of X [in other words it is supposed that the number $\langle n,y_N^*\rangle$ is known for each function $n(\phi)$ in N], one can extend y_N^* [i.e., define $\langle x,y^*\rangle$ for $x(\phi)$'s outside N] to the whole space X in such a way that its norm over X should not exceed[16] the initial one, over N. The construction of this "most economical extension" of y_N^* is done by induction, adding one new dimension after another to the subspace N.

We can suppose without loss of generality that the initial norm of y_N^* over N is equal to 1; i.e., that [cf. (A3)] for any $n(\phi)\in N$ we have

$$\langle n,y_N^*\rangle \leqslant \|n\|. \tag{A8}$$

To extend y_N^* to a functional y^* acting on the space $[x+N]$ of vectors $\alpha x(\phi)+n(\phi)$ it is sufficient to give the value of the constant $c\equiv\langle x,y^*\rangle$. Since by definition the extension y^* of y_N^* acting on elements $n(\phi)$ of N yields the same values as y_N^* itself,

$$\langle n,y^*\rangle \equiv \langle n,y_N^*\rangle \quad [n(\phi)\in N], \tag{A9a}$$

we have

$$\langle(\alpha x+n),y^*\rangle = \alpha\langle x,y^*\rangle + \langle n,y^*\rangle = \alpha c + \langle n,y_N^*\rangle, \tag{A9b}$$

and hence the action of y^* on vectors from $[x+N]$ is completely defined once the number c is given. For different c's we get, of course, different extensions of y_N^*.

Our aim is to define the number c so that the supremum (A3) taken over all vectors $\alpha x+n$ should not exceed 1, in order that the norm of the corresponding y^* on $[x+N]$ should not exceed that of y_N^* on N. We shall see that this is, indeed, possible. To proceed we need the following inequality which comes from the linearity of y^*, from Eq. (A8) and from the "triangle inequality" of the norm:

$$\begin{aligned}\langle n_1,y_N^*\rangle + \langle n_2,y_N^*\rangle &= \langle(n_1+n_2),y_N^*\rangle \leqslant \|n_1+n_2\|\\ &\equiv \|n_1+x+n_2-x\| \leqslant \|x+n_1\| + \|x-n_2\|\end{aligned} \tag{A10}$$

(in the last step, the usual properties of a norm were used). Separating now the terms depending on n_1 from those depending on n_2, one gets

$$-\{\|x-n_2\| - \langle n_2,y_N^*\rangle\} \leqslant \{\|x+n_1\| - \langle n_1,y_N^*\rangle\}. \tag{A11}$$

This inequality remains valid if one takes the supremum over n_2 and the infimum over n_1. Then we choose a constant $c\equiv\langle x,y^*\rangle$ to satisfy

$$\sup_{n_2}(-\{\|x-n_2\| - \langle n_2,y_N^*\rangle\}) \leqslant c \leqslant \inf_{n_1}\{\|x+n_1\| - \langle n_1,y_N^*\rangle\}. \tag{A12}$$

One can verify now that this choice for c is a good one. For $\alpha>0$ and any $n\in N$, we see that [taking $n_1 = n/\alpha$ in Eq. (A12)]

$$\begin{aligned}&\langle(\alpha x+n),y^*\rangle = \alpha(\{c+\langle n/\alpha,y_N^*\rangle\})\\ &\leqslant \alpha\{\|x+n/\alpha\| - \langle n/\alpha,y_N^*\rangle + \langle n/\alpha,y_N^*\rangle\} = \|\alpha x+n\|.\end{aligned} \tag{A13}$$

A similar inequality is obtained for $\alpha<0$, using the second side of Eq. (A12). But Eq. (A13) tells us that

$$\sup_{\alpha,n}\{\langle(\alpha x+n),y^*\rangle/\|\alpha x+n\|\} \leqslant 1. \tag{A14}$$

Actually in (A14) we have equality since we have supposed that $\sup\{\langle n,y_N^*\rangle/\|n\|\} \equiv \|y_N^*\|_N = 1$. In other words, the norm on the space $[x+N]$ of this special [see Eqs. (A9) and (A12)] extension y^* of the initial functional y_N^*, does not exceed (is equal to) the norm of y_N^* on the space N:

$$\|y^*\|_{[x+N]} = \|y_N^*\|_N = 1. \tag{A15}$$

Now, since we have supposed that the large space X contains a countable dense set of vectors, we first select from those an independent basis, to which we extend our construction in a recurrent way. Since our functional is now defined for any vector in this dense set, we may extend it by continuity to the whole space X, its norm always remaining equal to 1, the norm of the initial functional y_N^* defined on the subspace N.

$\mathscr{B}$. The Duality theorem

The proof of the "duality" relation (A1) is now straightforward.

(a) First of all, let us remark that the word "inf" in (A1) means that for any positive ϵ there exists at least one element $m_\epsilon(\phi)\in M$ such that

$$\delta_0 < \|x - m_\epsilon\| < \delta_0 + \epsilon. \tag{A16}$$

But since by definition

$$\langle m_\epsilon, y^*\rangle \equiv 0 \tag{A17a}$$

and

$$\|y^*\| = 1,\text{i.e.}, \langle (x - m_\epsilon), y^*\rangle \leqslant \|x - m_\epsilon\| \tag{A17b}$$

[see the definition (A3) of the norm], we have that

$$\langle x,y^*\rangle = \langle (x - m_\epsilon), y^*\rangle \leqslant \|x - m_\epsilon\| \leqslant \delta_0 + \epsilon. \tag{A18}$$

Since this inequality is valid for any positive ϵ, however small, and since $\langle x,y^*\rangle$ itself does not depend on ϵ, Eq. (A18) means that

$$\langle x,y^*\rangle \leqslant \delta_0, \tag{A19}$$

which is equivalent to the statement that $\sup\langle x,y^*\rangle$ from the right-hand side of (A1) is smaller or at most equal to $\delta_0 (\equiv \inf_{m\in M}\|x - m\|)$. Actually these two entities are equal, and this will be proved by means of an effective construction (using the Hahn–Banach lemma) of a functional y_0^* saturating the inequality (A19).

(b) This construction is done in two steps. First we take into consideration the subspace $N \equiv [x + M]$ of functions $n(\phi) = \alpha x(\phi) + m(\phi)$, and here we define the linear functional $y_{0,N}^*$ by

$$\langle n, y_{0,N}^*\rangle = \langle (\alpha x + m), y_{0,N}^*\rangle \equiv \alpha\delta_0. \tag{A20}$$

It is obvious that (putting $\alpha = 0$)

$$\langle m, y_{0,N}^*\rangle = 0 \quad \text{for all } m(\theta)\in M, \tag{A21}$$

and, moreover, that the norm (computed on the subspace N) is equal to 1:

$$\|y_{0,N}^*\|_N$$
$$\equiv \sup_n\{|\langle n, y_{0,N}^*\rangle|/\|n\|\} = \sup_{\alpha,m}\{|\alpha|\delta_0/(|\alpha|\cdot\|x + m/\alpha\|)\}$$
$$= \delta_0/\inf_{m_1}\|x + m_1\| = \delta_0/\delta_0 = 1 \tag{A22}$$

Hence the functional $y_{0,N}^*$ would have all the properties required by (A1) if it were possible to extend it to the whole space X, without altering its norm. But this is exactly the effect of the Hahn–Banach lemma described above!

Combining the conclusions of section (a) above with the existence of such a functional y_0^* saturating the inequality (A19), we have proved that

$$\delta_0 \equiv \inf_{m\in M}\|x - m\| = \max_{\substack{\|y^*\| = 1\\ \langle m,y^*\rangle = 0}} \langle x,y^*\rangle. \tag{A23}$$

The word "sup" has been replaced here by "max" in order to point out that this maximal functional y_0^* really exists.

This completes the proof of the Duality theorem.

APPENDIX B: THE GREEN'S FUNCTION FOR THE NEUMANN PROBLEM WITH THE UNIT CIRCLE AS BOUNDARY

The problem is to construct a real function $X_R(z)$, *harmonic* in the unit disk $|z| < 1$, when we are given the boundary values of its radial derivative $\partial X_R/\partial r$ on the unit circle $|z| = 1$. Clearly the function $X_R(z)$ is determined this way only up to a constant, and hence a "subtraction" is required.

We start from Cauchy's theorem for the (complex) *holomorphic* function $X(z) = X_R(z) + iX_{Im}(z)$, using C to denote the unit circle:

$$\begin{aligned} X(z) &= \frac{1}{2\pi i}\int_C \frac{X(z')}{z' - z}\,dz' \\ &= \frac{1}{2\pi i}\int_C X(z')d(\ln(z' - z)) \\ &= \left[\frac{1}{2\pi i}X(z')\ln(z' - z)\right]_C - \frac{1}{2\pi i}\int_C X'(z')\ln(z' - z_0)\,dz', \end{aligned} \tag{B1}$$

where $X'(z') \equiv dX(z')/dz'$. If we now select a point z_0 inside the disk as subtraction point, we write

$$\begin{aligned} X(z_0) &= \frac{1}{2\pi i}[X(z')\ln(z' - z_0)]_C \\ &\quad - \frac{1}{2\pi i}\int_C X'(z')\ln(z' - z_0)\,dz', \end{aligned}$$

and when this is subtracted from Eq. (B1) we get the result

$$X(z) - X(z_0) = -\frac{1}{2\pi i}\int_C X'(z')\ln\left(\frac{z' - z}{z' - z_0}\right)dz' \tag{B2}$$

since

$$\left[\frac{1}{2\pi i}X(z')\ln\left(\frac{z' - z}{z' - z_0}\right)\right]_C = 0. \tag{B3}$$

Equation (B3) follows from the fact that the logarithmic term is single-valued: if we make a cut from z to z_0, this cut does not intersect C since both z and z_0 are internal points of the disk (Fig. 2).

Now introduce the points Z and Z_0, the images of z and z_0 in C:

$$Z = 1/\bar{z}, \quad Z_0 = 1/\bar{z}_0,$$

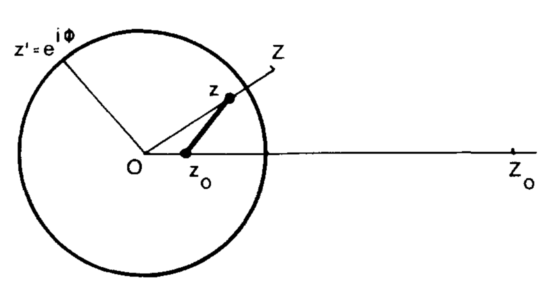


FIG. 2. The cuts of $\ln((z' - z)/(z' - z_0))$; $z' = e^{i\phi}$, $Z_0 = 1/\bar{z}_0$, $Z = 1/\bar{z}$.

and notice that

$$\frac{1}{2\pi i}\int_C X'(z')\ln\left(\frac{z'-Z}{z'-Z_0}\right)dz'=0, \tag{B4}$$

since both $dX(z')/dz'$ and $\ln((z'-Z)(/z'-Z_0))$ are holomorphic inside the unit z' disk.

Now, since on the unit circle C we have $z'=1/\bar{z}'$, we see that

$$\frac{z'-Z}{z'-Z_0}=\frac{1/\bar{z}'-1/\bar{z}}{1/\bar{z}'-1/\bar{z}_0}=\frac{\bar{z}'-\bar{z}}{\bar{z}'-\bar{z}_0} \tag{B5}$$

so that (B4) may be written as

$$0=\frac{1}{2\pi i}\int_C X'(z')\ln\left(\frac{\bar{z}'-\bar{z}}{\bar{z}'-\bar{z}_0}\right)dz'. \tag{B6}$$

Equation (B6) may now be combined with Eq. (B2) to yield the result

$$\begin{aligned}X(z)-X(z_0)&=\frac{1}{\pi i}\int_C X'(z')\ln\left|\frac{z'-z}{z'-z_0}\right|dz'\\&=-\frac{1}{\pi}\int_C\frac{\partial X}{\partial r}\ln\left|\frac{z'-z}{z'-z_0}\right|d\phi\end{aligned} \tag{B7}$$

where we have used $\partial X/\partial r=(dX/dz)(\partial z/\partial r)\equiv X'e^{i\phi}$.

If we take the real part of each side of this equation, and, since on the right-hand side of (B7) only $\partial X/\partial r$ is complex, we find

$$X_{\rm R}(z)=X_{\rm R}(z_0)-\frac{1}{\pi}\int_0^{2\pi}\frac{\partial X_{\rm R}(e^{i\phi})}{\partial r}\ln\left|\frac{e^{i\phi}-z}{e^{i\phi}-z_0}\right|d\phi, \tag{B8}$$

which is the analog for the Neumann problem of the Poisson formula (16) which solves the usual Dirichlet problem. So we have shown that the Neumann-kernel, which we shall denote by $N(z_0;z,e^{i\phi})$, is

$$N(z_0;z,e^{i\phi})=-2\ln|(e^{i\phi}-z)/(e^{i\phi}-z_0)|. \tag{B9}$$

Equation (B8) clearly implies the following representation for the holomorphic (*complex*) function $X(z)$ in terms of the *normal derivative* of its *real* part $X_{\rm R}$:

$$X(z)=X(z_0)-\frac{1}{\pi}\int_0^{2\pi}\frac{\partial X_{\rm R}(e^{i\phi})}{\partial r}\ln\left(\frac{e^{i\phi}-z}{e^{i\phi}-z_0}\right)d\phi \tag{B10}$$

Equation (B10) is the analog for the Neumann problem of the Schwarz–Villat formula (15).

APPENDIX C: CALCULATION OF THE COEFFICIENTS α_{ij} IN THE NEUMANN CASE

We have to evaluate the coefficients α_{ij} defined in Eq. (83),

$$\alpha_{ij}=\frac{1}{2\pi}\int_0^{2\pi}N(z_1;z_i,e^{i\phi})N(z_1;z_j,e^{i\phi})(\sigma(\phi))^{-1}\,d\phi, \tag{83}$$

where, from Appendix B and Eq. (70), and putting $e^{i\phi}=z'$, the Neumann kernel is

$$\begin{aligned}N(z_1;z,z')&=-2\ln\left|\frac{z'-z}{z'-z_1}\right|\\&=-2\mathrm{Re}\left\{\ln\left(\frac{z'-z}{z'-z_1}\right)\right\}\\&=-\left\{\ln\left(\frac{z'-z}{z'-z_1}\right)+\ln\left(\frac{1-zz'}{1-z_1z'}\right)\right\}.\end{aligned} \tag{C1}$$

As in Sec. 2, we shall introduce the holomorphic function $S(z)$ [Eqs. (21–23)], in order "to extend" the function $1/\sigma(\phi)$ defined initially on the unit circle $z'=e^{i\phi}$ also to points z with $|z|<1$.

$$S(z)=\frac{1}{2\pi}\int_0^{2\pi}\frac{e^{i\phi}+z}{e^{i\phi}-z}\frac{1}{\sigma(\phi)}d\phi, \tag{C2}$$

$$1/\sigma(\phi)=\mathrm{Re}\{S(e^{i\phi})\}. \tag{C3}$$

Noticing that both N factors appearing in (83) are real, substituting this in Eq. (38) and changing the variable integration from ϕ to $z'=e^{i\phi}$ we obtain

$$\begin{aligned}\alpha_{ij}=\mathrm{Re}\frac{1}{2\pi i}\int_C&\left\{\ln\left(\frac{z'-z_i}{z'-z_1}\right)+\ln\left(\frac{1-z'z_i}{1-z'z_1}\right)\right\}\\&\times\left\{\ln\left(\frac{z'-z_j}{z'-z_1}\right)+\ln\left(\frac{1-z'z_j}{1-z'z_1}\right)\right\}\frac{S(z')}{z'}dz',\end{aligned} \tag{C4}$$

where the contour C is the unit circle taken in a counterclockwise direction. It is convenient to choose the point z_1 such that $z_1<z_r$ for $r=2,\ldots,n$, and also for the moment we assume that $z_i<z_j$. Now the only singular factors in the integrand which lie within the disk are:

(a) $\ln((z'-z_i)/(z'-z_1))$ with branch points z_1,z_i;

(b) $\ln((z'-z_j)/(z'-z_1))$ with branch points z_1,z_j;

$(P)\,1/z'$, a simple pole at $z'=0$.

This means that α_{ij} may be written in the following form:

$$\alpha_{ij}=\{\alpha_{ij}^{(a)}+\alpha_{ij}^{(b)}+\alpha_{ij}^{(ab)}+\alpha_{ij}^{(P)}\}, \tag{C5}$$

where $\alpha_{ij}^{(a)}$ and $\alpha_{ij}^{(b)}$ contain either the factor (a) or (b) [see below Eqs. (C6) and (C8)], $\alpha_{ij}^{(ab)}$ contains both of them [see Eq. (C9)], while $\alpha_{ij}^{(P)}$ is the pole contribution. Since the residue at $z'=0$ of $\ln((1-z'z_k)/(1-z'z_1))$ is nil, only the term in (C4) containing $\ln((z'-z_i)/(z'-z_1))\ln((z'-z_j)/(z'-z_1))$ contributes to $\alpha_{ij}^{(P)}$ [(C10)]. The term containing $\ln((1-z'z_i)/(1-z'z_1))\ln((1-z'z_j)/(1-z'z_1))$ having no singularities inside the unit circle, vanishes identically.

By moving the integration contour we may write

$$\alpha_{ij}^{(a)}=\frac{1}{2\pi i}\int_{C_{z_1z_i}}\ln\left(\frac{z'-z_i}{z'-z_1}\right)\ln\left(\frac{1-z'z_j}{1-z'z_1}\right)S(z')\frac{dz'}{z'}, \tag{C6}$$

$C_{z_1z_i}$ being a contour which encircles the cut from z_1 to z_i in a counterclockwise direction. The discontinuity of $\ln((z'-z_i)/(z'-z_1))$ across the cut is $2\pi i$, so that

$$\alpha_{ij}^{(a)}=-P\int_{z_1}^{z_i}\ln\left|\frac{1-z'z_j}{1-z'z_1}\right|S(z')\frac{dz'}{z'}. \tag{C7}$$

The principal value of the integral is to be taken if $0\in(z_1,z_i)$.

The second term in Eq. (C5) is

$$\alpha_{ij}^{(b)} = \frac{1}{2\pi i}\int_{C_{z_j z_1}} \ln\left(\frac{z'-z_j}{z'-z_1}\right)\ln\left(\frac{1-z'z_i}{1-z'z_1}\right)S(z')\frac{dz}{z'}$$
$$= -P\int_{z_1}^{z_j}\ln\left|\frac{1-z'z_i}{1-z'z_1}\right|S(z')\frac{dz'}{z'}. \quad \text{(C8)}$$

The third term is

$$\alpha_{ij}^{(ab)} = \frac{1}{2\pi i}\int_{C_{z_i z_j}} \ln\left(\frac{z'-z_i}{z'-z_1}\right)\ln\left(\frac{z'-z_j}{z'-z_1}\right)S(z')\frac{dz}{z'}$$
$$= -P\int_{z_i}^{z_j}\ln\left|\frac{z'-z_i}{z'-z_1}\right|S(z')\frac{dz'}{z'}$$
$$+\frac{1}{2\pi i}\int_{C_{z_1 z_i}} \ln\left(\frac{z'-z_i}{z'-z_1}\right)\ln\left(\frac{z'-z_j}{z'-z_1}\right)S(z')\frac{dz'}{z'}$$
$$= -P\int_{z_i}^{z_j}\ln\left|\frac{z'-z_i}{z'-z_1}\right|S(z')\frac{dz'}{z'}$$
$$-P\int_{z_1}^{z_i}\left\{\ln\left|\frac{z'-z_i}{z'-z_1}\right| + \ln\left|\frac{z'-z_j}{z'-z_1}\right|\right\}S(z')\frac{dz'}{z'}$$
$$= -P\int_{z_1}^{z_i}\ln\left|\frac{z'-z_j}{z'-z_1}\right|S(z')\frac{dz'}{z'}$$
$$-P\int_{z_1}^{z_j}\ln\left|\frac{z'-z_i}{z'-z_1}\right|S(z')\frac{dz'}{z'}. \quad \text{(C9)}$$

In evaluating the above integral we noted that the discontinuity in $\ln((z'-z_i)/(z'-z_1))\ln((z'-z_j)/(z'-z_1))$ across the cut form z_1 to z_i is $2\pi i(\ln|(z'-z_i)/(z'-z_1)| + \ln|(z'-z_j)/(z'-z_1)|)$ and $2\pi i\ln|(z'-z_i)/(z'-z_1)|$ across the cut from z_i to z_j.

The fourth term is the residue at $z' = 0$. A little care is needed in defining this due to the presence of the cut. It is not difficult to see that the correct result is

$$\alpha_{ij}^{(P)} = \{\ln((-z_i)/(-z_1))\ln((-z_j)/(-z_1))\}S(0)$$
$$= \{\ln|z_i/z_1|\ln|z_j/z_1| - \pi^2\theta(-z_1)\theta(z_i)\theta(z_j)\}S(0), \quad \text{(C10)}$$

where the factor $\theta(-z_1)\theta(z_i)\theta(z_j)$ is equal to *one* if $z = 0$ lies between z_1 and min (z_i, z_j), and is *zero* otherwise.

We may combine these terms to get

$$\alpha_{ij} = \{\ln|z_i/z_1|\ln|z_j/z_1| - \pi^2\theta(-z_1)\theta(z_i)\theta(z_j)\}S(0) + A_{ij} + A_{ji}, \quad \text{(C11)}$$

where

$$A_{ij} = -P\int_{z_1}^{z_i}\left\{\ln\left|\frac{1-z'z_j}{1-z'z_1}\right| + \ln\left|\frac{z'-z_j}{z'-z_1}\right|\right\}\frac{S(z')}{z'}dz' \quad \text{(C12)}$$

and where $S(z)$ is given by Eq. (C2). The symbol "Re" present in Eq. (C5) is here unnecessary, since all quantities appearing in (C11) are real. Nor is the restriction $z_1 < z_j$ any longer necessary, since Eq. (C11) is symmetric in the variables z_i and z_j. It may also be shown that this formula gives the correct result also in the case $i = j$.

[1] See for instance R. Oehme and W. Zimmerman, preprint (University of Chicago, 1980).

[2] Q.C.D. perturbative calculations are carried out in spacelike regions, but in order to produce predictions also in the timelike region one might use the spacelike information in conjunction with some adequate continuation technique. See also I. Caprini and C. Verzegnassi, ICTP preprint (1980).

[3] For instance in perturbative calculations, there are uncertainties since any practical computation takes into account only a finite number of graphs.

[4] See for instance the review paper S. Ciulli, C. Pomponiu, and I. S. Stefanescu, Phys. Rep. **17**, 133–224 (1975). A simple example of the explosive propagation of error is provided by the analytic continuation procedures (see Ref. 5,6) which make use of the mapping $W(s)$ which maps the data region Γ_1 onto the unit circle $|W(s)| = 1$ and the cuts, Γ_R, onto the circle $|W(s)| = R$ (R is here a conformal invariant). A given point s_p of the cut-plane will then map at some point $W(s_p)$ on the circle Γ_p of radius $\rho_p \equiv |W(s_p)|$.

Now, on each of these circles the amplitude $F(s)$ might be seen as a function of $\theta = \arg W$, i.e., the function $F(s)$ is an element (a point) in the function space spanned by the infinite basis $\{e^{in\phi}\}$. Hence analytic continuation from Γ_1 towards Γ_p and Γ_R can be viewed as a flow inside this space of functions. This flow is extremely anisotropic and divergent, as a spacing ϵ on two points along the n^{th} dimension, at $\rho = 1$, will become $\epsilon\rho_p''$ at $\rho = \rho_p$.

[5] R. E. Cutkosky and B. B. Deo; Phys. Rev. Lett. **22**, 1272 (1968) and Phys. Rev. **174**, 1859 (1968). R. E. Cutkosky, J. Math. Phys. **14**, 1231 (1973); see also next reference.

[6] S. Ciulli, Nuovo Cimento A **61**, 787 (1969); **62**, 301 (1969).

[7] A rough estimate of this number is given by the minimal number of terms in $P_n^{(D)}(w)$ necessary to approximate the data better than ϵ.

[8] J. L. Walsh, *Interpolation and Approximation by Rational Functions*, Vol. 20, 2nd ed. (Am. Math. Soc.,Providence 6, RI, 1956).

[9] The derivation of the expression (1) for the error bound is primarily based on (Ref. 6) the computation of the bound $(\epsilon + \eta_n/R^n)$ for the deviation between the exact but unknown polynomials $P_n^{(F)}(w)$ and the data-constructed ones $P_n^{(D)}(w)$ in the data region Γ_1. One recognizes next the analyticity of $(P_n^{(F)}(w) - P_n^{(D)}(w))/W^n$ outside the unit disk (i.e., also in the cuts region, $|W| = R$ and even for $W \to \infty$), such that one might use the maximum modulus principle.

[10] See for instance Chap. V of Ref. 8. Here the weight $n(z)$ is essentially $1/\epsilon(z)$. There is a factor $1/N_1 \equiv 1/\min(n(z)) = \max\epsilon(z)$ which appears in the right-hand side of the inequality (5) (page 91 of Ref. 8), which worsens hence the error bound η_n/R^n of the $P_n^{(F)}(w)$ in the data region $\Gamma_1(\equiv c)$.

[11] From now on it is assumed that the cut complex s plane has been mapped onto the unit disk $|z| < 1$, by means of standard methods [see for instance S. Ciulli and J. Fischer, Nucl. Phys. B **24**, 537 (1970)].

[12] An analytic function is determined by its real part up to a pure imaginary constant, but the latter vanishes in our case because of the condition $M(z) = \overline{M}(\bar{z})$. Strictly speaking, to ensure one-to-one correspondence between the functions $M(z)$ and their boundary values $m(\phi)$, the interior function $M_R(z)$ must be restricted to the class h^P (i.e. $\sup_{0\leqslant r\leqslant 1}\|M_R(z = re^{i\phi})\|_{L^P}$ has to be *finite*) for $P > 1$, when $m(\phi) \in L^P$. In this paper we have been concerned with $P = 2$.

[13] The set $\{F(z)\}$ of analytic functions having a representation (71) but vanishing at $z = z_0$ indeed forms a space. Further, there is a linear one-to-one correspondence between the elements of this space and the L^2 space of the functions $f_{,r}(\phi)$ (the boundary values of the normal derivatives), each linear functional over $\{F\}$ being a linear functional over $\{f_{,r}\}$ and vice versa. As we have already stressed in Sec. 2, this is essential for the use of duality in conjunction with the Riesz representation (which is valid for L^P spaces).

[14] R. E. Cutkosky, Ann. of Phys. (NY) **54**, 350 (1969); see also R. E. Cutkosky, J. Math. Phys. **14**, 1231 (1973).

[15] The effect of the functional $\langle .,y^*\rangle$ on the function $n(\phi)$, is, by definition, the number $\langle n,y^*\rangle$.

[16] Since the supremum in (A3) is taken over the larger set X, it cannot be smaller than that taken over the set N.

Initial-boundary-value problem for diffusion of magnetic fields into conductors with external electromagnetic transients

H. E. Wilhelm
Michelson Laboratory, Naval Weapons Center, China Lake, California 93555

(Received 22 September 1981; accepted for publication 14 May 1982)

The initial-boundary-value problem for the diffusion of an initially homogeneous magnetic field into a slab of conductivity $\sigma < \infty$ and width $\Delta x = 2a$ is solved, under consideration of the electromagnetic wave pulses generated at the surfaces of the conductor by its interaction with the external magnetic field, which propagate into the surrounding vacuum. The analytical solutions show that (i) the external electromagnetic transients are necessary in order to correctly satisfy the boundary conditions for the tangential electric and magnetic field components, and (ii) the spatial and temporal development of the electromagnetic field and electric current in the conductor is quantitatively determined by a new dimensionless parameter group $\mathscr{R} = \mu_0\sigma ac[c = (\mu_0\epsilon_0)^{-1/2}]$. This "magnetic Reynolds number of the vacuum" determines the coupling between the transient fields in the conductor $\sigma > 0$ and the ambient space $(\sigma = 0)$.

PACS numbers: 02.60.Lj, 41.90. + e

1. INTRODUCTION

The diffusion of electromagnetic fields $\mathbf{B}(\mathbf{r}, t)$, $\mathbf{E}(\mathbf{r}, t)$ in a conductor of finite conductivity σ and normal surface vector $\mathbf{n}(\mathbf{s})$, when the electromagnetic field $\mathbf{B}_0(\mathbf{r}, t)$ and $\mathbf{E}_0(\mathbf{r}, t)$ outside of the conductor are known, is in general described by Maxwell's equations without displacement current, where the tangential field components are assumed to satisfy the boundary conditions[1,2] $\mathbf{n}\times[\mathbf{B}(\mathbf{r}, t) - \mathbf{B}_0(\mathbf{r}, t)] = \mathbf{0}$ and $\mathbf{n}\times[\mathbf{E}(\mathbf{r}, t) - \mathbf{E}_0(\mathbf{r}, t)] = \mathbf{0}$. If the external electromagnetic field is time-independent and electric potential fields are absent, then $\mathbf{B}_0 = \mathbf{B}_0(\mathbf{r})$ and $\mathbf{E}_0 = \mathbf{0}$ (since $\nabla\times\mathbf{E}_0 = -\partial\mathbf{B}_0/\partial t = \mathbf{0}$ and $\mathbf{E}_0 = -\nabla\Phi_0 \equiv 0$), so that the tangential boundary conditions reduce to[1,2] $\mathbf{n}\times[\mathbf{B}(\mathbf{r}, t) - \mathbf{B}_0(\mathbf{r})] = \mathbf{0}$ and $\mathbf{n}\times\mathbf{E}(\mathbf{r}, t) = \mathbf{0}$. These boundary conditions have found widespread use in mathematical physics,[1] electromagnetic theory,[2] and the theory of magnetic flux compression (at the outside surface of the liners).[3,4] However, these boundary conditions are questionable approximations, since they do not take into consideration the wave fields $\mathbf{B}(\mathbf{r}, t)$, $\mathbf{E}(\mathbf{r}, t)$ propagating away from the conductor into the surrounding medium, which have their sources in the transient current fields $\mathbf{j} = \mu_0^{-1}\nabla\times\mathbf{B}$ of the conductor.

For a concrete illustration of the problematics, consider the diffusion of an external (homogeneous) magnetic field, $\mathbf{B}_0 = \{0, B_0, 0\}$ for $|x| > a$, into a conducting slab in the region $|x| < a$ which is field-free at time $t = 0$ (Fig. 1). Using the conventional boundary conditions, the transient magnetic field $\mathbf{B}(x, t) = \{0, B(x, t), 0\}$ in the conductor is determined by the parabolic initial-boundary-value problem[5]:

$$\partial B/\partial t = \kappa\partial^2 B/\partial x^2, \quad |x| < a, \quad t > 0, \tag{1}$$

$$B(x, t = 0) = 0, \quad |x| < a, \tag{2}$$

$$B(x = \pm a, t) = B_0, \quad t > 0, \tag{3}$$

where $\kappa = 1/\mu_0\sigma$. By means of Fourier's method, the general solution of Eqs. (1)–(3) is obtained as a superposition of eigenfunctions[5]:

$$B(x, t) = B_0\left[1 - \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{(2n-1)} \times e^{-\kappa(2n-1)^2\pi^2 t/4a^2}\cos\frac{2n-1}{2a}\pi x\right],$$
$$|x| \leqslant a, \quad t \geqslant 0, \tag{4}$$

with $B(x, t) \to B_0$ in $|x| < a$ for $t \to \infty$. Since $\nabla\times\mathbf{B} = \mu_0\sigma\mathbf{E}$, the electric field $\mathbf{E}(x, t) = \{0,0,E(x, t)\}$ in the conductor is

$$E(x, t) = \frac{2B_0}{\mu_0\sigma a}\sum_{n=1}^{\infty}(-1)^{n-1} \times e^{-\kappa(2n-1)^2\pi^2 t/4a^2}\sin\frac{2n-1}{2a}\pi x,$$
$$|x| \leqslant a, \quad t \geqslant 0. \tag{5}$$

In accordance with the boundary conditions (3), the space outside of the conductor remains unperturbed while the electromagnetic field diffuses into the conductor,

$$B_0(x, t) = B_0, \quad E_0(x, t) = 0, \quad |x| \geqslant a, \quad t \geqslant 0. \tag{6}$$

The transient currents $\mathbf{j} = \nabla\times\mathbf{B}/\mu_0$ in the conductor are "eddy currents," and, therefore, cannot produce transient magnetic fields $\widetilde{B}_0(x, t) = B_0(x, t) - B_0 \neq 0$ in the outside region $|x| > a$. The net current $I(t)$ through any cross section $z =$ const vanishes, due to the boundary conditions (3):

$$I(t)/\Delta y = \mu_0^{-1}\int_{-a}^{+a}[\partial B(x, t)/\partial x]\, dx$$
$$= \mu_0^{-1}\int_{B_0}^{B_0} dB = 0. \tag{7}$$

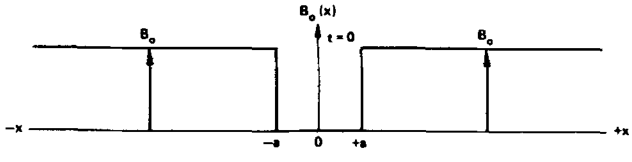


FIG. 1. Magnetic field $B_0(x)$ for $t = 0$.

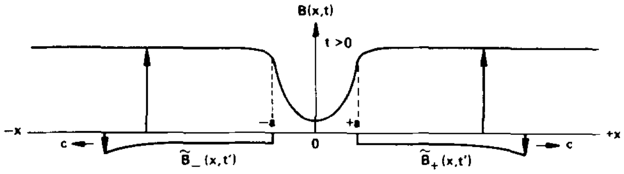


FIG. 2. Diffusion field $B(x,t)$ and transients $\tilde{B}_{\pm}(x,t)$ for $t>0$ and $t'>0$ (qualitative).

By comparing the conductor solutions (4) and (5) with the vacuum solutions (6), it is seen that $B(x=\pm a, t)-B_0=0$, but $E(x=\pm a,t)-E_0(x=\pm a, t)=E(x=\pm a,t)\neq 0$! Thus, the conventional boundary conditions[1,2] lead to a violation of the fundamental law of the continuity of the tangential electric fields at interfaces.

The correct formulation of the boundary conditions requires consideration of the simultaneous wave fields $\tilde{B}_{\pm}(x,t)$, $\tilde{E}_{\pm}(x,t)$ propagating with the speed of light c into the positive and negative half-spaces $x>+a$ and $x<-a$ surrounding the conductor (Fig. 2), which are excited by the transient current fields $j(x,t)=\mu_0^{-1}\partial B(x,t)/\partial x$ in the (space-charge free) conductor. No matter how small the external transients $\tilde{B}_{\pm}(x,t)$ and $\tilde{E}_{\pm}(x,t)$ are (in comparison with $B_0\neq 0$ and $E_0=0$), they have to be taken into account in order to rigorously satisfy the boundary conditions $\mathbf{n}\times[\mathbf{B}]=\mathbf{0}$ and $\mathbf{n}\times[\mathbf{E}]=\mathbf{0}$ for the continuity of the tangential electromagnetic fields at conductor interfaces.

The quantitative assessment of the significance of the external transients of the diffusion process leads to the discovery of a new dimensionless parameter combination, which has the physical meaning of a "magnetic Reynolds number of the vacuum":

$$\mathscr{R}=\mu_0\sigma ac, \quad c=(\mu_0\epsilon_0)^{-1/2}=3\times 10^8 \text{ m/s}. \tag{8}$$

In the following, the formulation of the initial-boundary-value problem for diffusion processes with external transients and its analytical solutions for the transient electromagnetic fields inside and outside the conductor are presented. The qualitative and quantitative importance of the new boundary conditions and the external wave fields are discussed in terms of $\mathscr{R}$.

The presented theory has important implications for the evaluation fo the flux losses through the liners of magnetic field compressors,[3,4] the electromagnetic acceleration of conducting macroparticles,[6,7] the electromagnetic induction in conductors moving relative to external magnetic fields,[8,9] and for the interaction of transient plasma shock waves with external magnetic fields.[10,11] The general significance for theoretical physics is obvious.

2. INITIAL-BOUNDARY-VALUE PROBLEM

The subject of the following considerations is the diffusion of the magnetic field into a conducting slab $|x|<a$, which is initially embedded in a homogeneous magnetic field $\mathbf{B}_0=\{0,B_0,0\}$, under simultaneous emission of electromagnetic waves from the conductor surfaces $x=\pm a$ (Fig. 1).

The transient electromagnetic fields $\mathbf{B}_{\pm}=\{0, B_{\pm}(x,t),0\}$ and $\mathbf{E}_{\pm}=\{0,0,E_{\pm}(x,t)\}$ in the infinite vacuum half-spaces ($\sigma=0, \mu=\mu_0$) $x>+a$ and $x<-a$ are determined by the initial-boundary-value problems ($\pm$) for the wave equation:

$$\partial^2 B_{\pm}/\partial t^2=c^2\partial^2 B_{\pm}/\partial x^2, \quad \pm x>a,\ t>0, \tag{9}$$

$$B_{\pm}(x,t=0)=B_0, \quad \pm x>a, \tag{10}$$

$$B_{\pm}(x=\pm a,t)=B_0+\psi_{\pm}(t), \quad t>0, \tag{11}$$

since

$$\partial E_{\pm}/\partial t=c^2\partial B_{\pm}/\partial x, \quad \partial E_{\pm}/\partial x=\partial B_{\pm}/\partial t, \tag{12}$$

by Maxwell's equations with displacement current. The solutions of Eqs. (9)–(11) for the still undetermined boundary values $\psi_{\pm}(t)$ are

$$\begin{aligned}B_{\pm}(x,t)&\\ &=B_0+\psi_{\pm}(t\mp(x\mp a)/c), \quad a<\pm x<a+ct,\\ &=B_0, \quad a+ct<\pm x<\infty,\end{aligned} \tag{13}$$

and

$$\begin{aligned}E_{\pm}(x,t)&\\ &=\mp c\psi_{\pm}(t\mp(x\mp a)/c), \quad a<\pm x<a+ct,\\ &=0, \quad a+ct<\pm x<\infty.\end{aligned} \tag{14}$$

These solutions are typical for hyperbolic equations, i.e., the boundary values $\psi_{\pm}(t)$ are "transported" into the half-spaces $\pm x>a$ with the speed of light c, so that discontinuous wave fronts result at $x=\pm(a+ct)$.

Let the external magnetic field $\mathbf{B}_0$ be switched on at $t=0$ so that no electromagnetic fields exist in the conductor for $t<0$.[1,2] The conductor has a finite conductivity σ and can, therefore, not carry surface current densities (Ref. 12), $\mathbf{j}^*=\lim_{\Delta x\to 0}\sigma\mathbf{E}\Delta x=\mathbf{0}$ for $\sigma<\infty$ and $\mathbf{E}$ bounded. Accordingly, the boundary conditions for the tangential electric and magnetic field components at the conductor vacuum interfaces are

$$B(x=\pm a,t)=B_0+\psi_{\pm}(t), \quad t>0, \tag{15}$$

$$E(x=\pm a,t)=\mp c\psi_{\pm}(t), \quad t>0, \tag{16}$$

where

$$E(x,t)=\kappa\partial B(x,t)/\partial x, \quad |x|<a,\ t>0, \tag{17}$$

by Ohm's law is the electric field in the conductor, and $B(x,t)$ is the magnetic field in $|x|<a$. Furthermore,

$$\kappa=1/\mu_0\sigma>0. \tag{18}$$

(The boundary conditions $\mathbf{n}\cdot[\epsilon\mathbf{E}]=0$ and $n\cdot[\mathbf{B}]=\mathbf{0}$ are satisfied since $\mathbf{E}$ and $\mathbf{B}$ have no normal components.) By elimination of the unknown boundary values $E(x=\pm a,t)$ and $\psi_{\pm}(t)$ from Eqs. (15)–(17), boundary conditions involving only the magnetic field $B(x,t)$ in the conductor are obtained:

$$\frac{\partial B(x=\pm a,t)}{\partial x}\pm\frac{c}{\kappa}B(x=\pm a,t)=\pm\frac{c}{\kappa}B_0, \quad t>0. \tag{19}$$

These are the new boundary conditions for the diffusion of magnetic fields $B(x,t)$ into conductors. They differ from the conventional boundary conditions[1,2] through the curl terms $\partial B(x=\pm a,t)/\partial x\neq 0$, which consider the emission of magnetic dilution waves from the conductor surfaces $x=\pm a$ into the vacuum $|x|>a$.

Within the conducting slab of finite width $2a$, the propagation of the magnetic field can be treated in the nonrelativistic or diffusion approximation.[1,2,12] Accordingly, $B(x,t)$ in the initially field-free conductor is determined by the parabolic initial-boundary-value problem:

$$\partial B/\partial t = \kappa \partial^2 B/\partial x^2, \quad |x|<a, \ t>0, \tag{20}$$

$$B(x, t=0)=0, \quad |x|<a, \tag{21}$$

$$\partial B(x=\pm a, t)/\partial x \pm hB(x=\pm a, t) = \pm hB_0, \quad t>0, \tag{22}$$

where

$$h = c/\kappa > 0. \tag{23}$$

The transformation

$$B(x,t) = B_0 + \widetilde{B}(x,t), \quad |x|\leqslant a, \ t\geqslant 0, \tag{24}$$

reduces Eqs. (21)–(22) to an initial-boundary-value problem with standard "radiation" boundary conditions:

$$\partial \widetilde{B}/\partial t = \kappa \partial^2 \widetilde{B}/\partial x^2, \quad |x|<a, \ t>0, \tag{25}$$

$$\widetilde{B}(x, t=0) = -B_0, \quad |x|<a, \tag{26}$$

$$\partial \widetilde{B}(x=\pm a, t)/\partial x \pm h\widetilde{B}(x=\pm a, t) = 0, \quad t>0. \tag{27}$$

In accordance with Fourier's theorem, the solution of Eqs. (25)–(27) is obtained as a superposition of eigenfunctions $\widetilde{B}_n(x,t)$ for the region $|x|<a$ which satisfy the boundary conditions (27):

$$\widetilde{B}(x,t) = -2B_0 \sum_{n=1}^{\infty} \frac{(h^2+k_n^2)k_n^{-1}\sin k_n a}{[(h^2+k_n^2)a+h]} \times e^{-\kappa k_n^2 t}\cos k_n x, \quad |x|\leqslant a, \ t\geqslant 0, \tag{28}$$

where

$$k_n a tg(k_n a) = ha, \quad n=1,2,3,\cdots \tag{29}$$

gives the eigenvalues k_n associated with the boundary conditions (27).

A. Conductor Solutions

For a compact representation of the field solutions, dimensionless independent (ξ, τ) and dependent variables are introduced,

$$\xi = x/a, \quad \tau = \kappa t/a^2, \quad \alpha_n = k_n a, \tag{30}$$

$$\mathscr{B}(\xi,\tau) = B(x,t)/B_0, \quad \mathscr{E}(\xi,\tau) = E(x,t)/(\kappa B_0/a), \tag{31}$$

$$\not{j}(\xi,\tau) = j(x,t)/(B_0/\mu_0 a).$$

According to Eqs. (24) and (28), the solutions for the dimensionless fields $\mathscr{B}(\xi,\tau)$, $\mathscr{E}(\xi,\tau) = \partial\mathscr{B}(\xi,\tau)/\partial\xi$, and $\not{j}(\xi,\tau)$ in the conductor are given by

$$\mathscr{B}(\xi,\tau) = 1 - 2\sum_{n=1}^{\infty} \frac{(\mathscr{R}^2+\alpha_n^2)\alpha_n^{-1}\sin\alpha_n}{[(\mathscr{R}^2+\alpha_n^2)+\mathscr{R}]} \times e^{-\alpha_n^2\tau}\cos\alpha_n\xi, \quad |\xi|\leqslant 1, \ \tau\geqslant 0, \tag{32}$$

$$\mathscr{E}(\xi,\tau) = 2\sum_{n=1}^{\infty} \frac{(\mathscr{R}^2+\alpha_n^2)\sin\alpha_n}{[(\mathscr{R}^2+\alpha_n^2)+\mathscr{R}]} \times e^{-\alpha_n^2\tau}\sin\alpha_n\xi, \quad |\xi|\leqslant 1, \ \tau\geqslant 0, \tag{33}$$

$$\not{j}(\xi,\tau) = \mathscr{E}(\xi,\tau), \quad |\xi|\leqslant 1, \ \tau\geqslant 0, \tag{34}$$

where

$$\alpha_n tg\alpha_n = \mathscr{R}, \quad n=1,2,3,\ldots, \quad \mathscr{R} = ac/\kappa = \mu_0\sigma ac, \tag{35}$$

by Eqs. (8) and (29). For sufficiently large times $\tau \gg \alpha_1^{-2}$, the homogeneous magnetic field has diffused completely into the conductor,

$$\mathscr{B}(\xi,\tau)\to 1, \quad \mathscr{E}(\xi,\tau)\to 0, \quad \not{j}(\xi,\tau)\to 0, \quad \tau\to\infty. \tag{36}$$

In the hypothetical limit of infinite magnetic vacuum Reynolds number $\mathscr{R}$, Eq. (32) reduces to the known solution (4) for the conventional boundary conditions,[5]

$$\lim_{\mathscr{R}\to\infty} \mathscr{B}(\xi,\tau) = 1 - \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{(2n-1)} \times e^{-(2n-1)^2\pi^2\tau/4}\cos\frac{2n-1}{2}\pi\xi, \quad |\xi|\leqslant 1, \ \tau\geqslant 0, \tag{37}$$

since

$$\lim_{\mathscr{R}\to\infty}\alpha_n = \frac{2n-1}{2}\pi, \quad n=1,2,3,\cdots. \tag{38}$$

Comparison of Eq. (32) with Eq. (37) indicates that the $\mathscr{B}(\xi,\tau)$ solutions with the new and conventional boundary conditions differ not much if $\mathscr{R} \gg \alpha_1 = \pi/2$.

B. Vacuum Solutions

In view of the boundary conditions (15), Eq. (32) yields for the boundary values $\Psi_\pm(\tau) = \mathscr{B}(\xi=\pm 1,\tau) - 1$. Accordingly, Eqs. (13) and (14) give for the electromagnetic fields in the positive $(\xi > +1)$ and negative $(\xi < -1)$ half-spaces the solutions

$$\mathscr{B}_\pm(\xi,\tau) = 1 + \Psi_\pm(\tau \mp (\xi \mp 1)/\mathscr{R}), \quad 1 < \pm\xi < 1+\mathscr{R}\tau, \\ = 1, \quad 1+\mathscr{R}\tau < \pm\xi < \infty, \tag{39}$$

and

$$\mathscr{E}_\pm(\xi,\tau) = \mp\mathscr{R}\Psi(\tau\mp(\xi\mp 1)/\mathscr{R}), \quad 1<\pm\xi<1+\mathscr{R}\tau, \\ = 0, \quad 1+\mathscr{R}\tau < \pm\xi < \infty, \tag{40}$$

where

$$\Psi_\pm\left(\tau \mp \frac{\xi\mp 1}{\mathscr{R}}\right) = -2\sum_{n=1}^{\infty}\frac{(\mathscr{R}^2+\alpha_n^2)\sin\alpha_n\cos\alpha_n}{\alpha_n[(\mathscr{R}^2+\alpha_n^2)+\mathscr{R}]} e^{-\alpha_n^2[\tau\mp(\xi\mp 1)/\mathscr{R}]}. \tag{41}$$

Equations (39) and (40) represent electromagnetic wave pulses which propagate with the dimensionless speed $\mathscr{R}(c)$ from the conductor surfaces $\xi = \pm 1$ into the vacuum spaces $\pm\xi > 1$ with discontinuous wave fronts at $\xi = \pm(1+\mathscr{R}\tau)$. They are kicked on at $\tau = 0$ and their emission lasts to the end $(\tau\to\infty)$ of the diffusion process in the conductor. The vacuum fields $\mathscr{B}_\pm(\xi,\tau)$ are in the opposite direction of $\mathbf{B}_0$, i.e., they represent dilution waves (Fig. 2).

In the case of large coupling numbers, $\mathscr{R} \gg 1$, Eqs. (39) and (40) yield with Eq. (41)

$$\mathscr{B}_{\pm}(\xi,\tau) \cong 1 + O[\mathscr{R}^{-1}], \quad 1 < \pm\xi < 1 + \mathscr{R}\tau, \tag{42}$$

$$\mathscr{E}_{\pm}(\xi,\tau) \cong \pm 2 \sum_{n=1}^{\infty} e^{-\left(\frac{2n-1}{2}\pi\right)^2[\tau \mp (\xi \mp 1)/\mathscr{R}]}$$

$$1 < \pm\xi < 1 + \mathscr{R}\tau, \tag{43}$$

since $\cos\alpha_n \cong (-1)^{n-1}(2n-1)\pi/2\mathscr{R}$ for $\mathscr{R} \gg 1$ by Eq. (35).

The magnetic field $\mathbf{B}_0$ outside of the conductor remains nearly unperturbed during the diffusion process, $\widetilde{\mathscr{B}}_{\pm} \sim \mathscr{R}^{-1}$, whereas the external electric transients $\mathscr{E}_{\pm} \neq 0$ are of order $\mathscr{R}^0$ behind the wave fronts, $\xi = \pm(1 + \mathscr{R}\tau)$, for $\mathscr{R} \gg 1$. However, since $\nabla \times \mathbf{B}_{\pm} = c^{-2}\partial\mathbf{E}_{\pm}/\partial t$, not only $\mathscr{E}_{\pm} \sim \mathscr{R}^0$ but also $\widetilde{\mathscr{B}}_{\pm} \sim \mathscr{R}^{-1}$ cannot be neglected for $\mathscr{R} \gg 1$.

Thus, we have shown how self-consistent solutions can be obtained for the electric and magnetic fields in the conductor and the surrounding vacuum, which satisfy the boundary conditions for the continuity of the tangential electric and magnetic fields at the conductor-vacuum interfaces. The conventional boundary conditions for electromagnetic diffusion processes,[1,2] ignore the external electromagnetic transients, violate the boundary condition for the tangential electric field, and permit no Poynting vector $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ outside of the conductor. As a result, the conventional boundary conditions[1,2] make it impossible for electromagnetic fields to diffuse through conductors and to escape into the ambient space.

For both the conductor and vacuum solutions, the limit $\mathscr{R} \to 0$, which implies $\sigma \to 0$ since $a \neq 0$, is not realizable since the conductivity of conductors is by definition large. For insulators or extremely poor conductors ($\mathscr{R} \propto \sigma \to 0$), the nonrelativistic or parabolic diffusion equation is inapplicable.[12] Therefore, the investigation of the limit $\mathscr{R} \to 0$ would require solution of Maxwell's equations with displacement current in the slab $|x| < a$ given elsewhere.[13]

The generalizations of the theory required for conductors and external media (vacuum, gases, fluids) with different permittivities ϵ and μ are trivial but complicate the notation.

3. DISCUSSION

It is known that Maxwell's equations with displacement current and the nonrelativistic Ohm's law $\mathbf{j} = \sigma\mathbf{E}$ combine to a hyperbolic diffusion equation for the magnetic field $\mathbf{B}(\mathbf{r}, t)$ in conductors,[12]

$$\frac{\partial^2\mathbf{B}}{\partial t^2} + \frac{1}{\tau_R}\frac{\partial\mathbf{B}}{\partial t} = c^2\nabla^2\mathbf{B}, \tag{44}$$

where

$$\tau_R = \epsilon_0/\sigma \tag{45}$$

is the field relaxation time, e.g., $\tau_R \cong (10^{-9}/36\pi)/6\times10^7 \sim 10^{-19}$ s for copper with $\sigma = 6\times10^7\,\Omega^{-1}$/m. Equation (44) reduces to the parabolic diffusion equation in the limit $\tau_R \ll a/c$:

$$\frac{\partial\mathbf{B}}{\partial t} = \tau_R c^2\nabla^2\mathbf{B}, \quad \tau_R \ll a/c. \tag{46}$$

The parabolic diffusion equation is an excellent approximation, since the relaxation time of conductors is extremely small, $\tau_R \ll a/c$. By Eqs. (45) and (46), the field relaxation time τ_R and the diffusion time τ_D are interrelated by

$$\tau_D^{-1} = \tau_R c^2/a^2, \quad \tau_D = \mu_0\sigma a^2, \tag{47}$$

where a is the extension of the conductor. For conductors, the diffusion time is relatively large if a is not microscopically small, e.g., $\tau_D = 4\pi\times10^{-7}\times6\times10^7\times10^{-4} \sim 10^{-2}$ s for a copper slab of width $a = 10^{-2}$ m.

Comparison of the neglected term $\partial^2\mathbf{B}/\partial t^2$ with the second and third terms of Eq. (44) reveals the relation of the parabolic diffusion approximation to the new coupling number $\mathscr{R} = \mu_0\sigma ac$:

$$\left|\frac{\partial^2\mathbf{B}}{\partial t^2}\right| \Big/ \left|\tau_R^{-1}\frac{\partial\mathbf{B}}{\partial t}\right| \sim \left|\frac{\partial^2\mathbf{B}}{\partial t^2}\right| \Big/ |c^2\nabla^2\mathbf{B}|$$

$$\sim \frac{\tau_R}{\tau_D} = \frac{\epsilon_0\mu_0}{(\mu_0\sigma a)^2} = \mathscr{R}^{-2}. \tag{48}$$

This result again confirms the validity of the parabolic diffusion equation for conductors, for which $\mathscr{R} = \mu_0\sigma ac \gg 1$. E.g., $\mathscr{R} \cong 4\pi\times10^{-7}\times6\times10^7\times10^{-2}3\times10^8 \sim 10^8$ for a copper slab $a = 10^{-2}$ m. More important, Eq. (48) demonstrates that the neglected relativistic term $\partial^2\mathbf{B}/\partial t^2$ in the conductor is small of order $\mathscr{R}^{-2} <<<< 1$, whereas the calculated electromagnetic fields in the conductor are of order $\mathscr{B} \sim \mathscr{E} \sim \mathscr{R}^0$ [Eqs. (32)–(33)], and the external electromagnetic transients are of order $\widetilde{\mathscr{B}}_{\pm} \sim \mathscr{R}^{-1}$ and $\mathscr{E}_{\pm} \sim \mathscr{R}^0$ [Eqs. (39)–(40)], since in Eq. (41) for large $\mathscr{R}$

$$|\cos\alpha_n| = [1 + (\mathscr{R}/\alpha_n)^2]^{-1/2} \cong \alpha_n/\mathscr{R} \sim \tfrac{1}{2}(2n-1)\pi\mathscr{R}^{-1},$$

$$\mathscr{R} \gg 1, \quad n = 1, 2, 3, \cdots. \tag{49}$$

In conclusion, it is noted that, in conductors, magnetic field diffusion is a nonrelativistic process (as in electric conduction, $\mathbf{j} = \sigma\mathbf{E}$). The electric transients $\mathscr{E}_{\pm}$ in vacuum must be of the same order as the electric field $\mathscr{E}$ in the conductors, $\mathscr{E}_{\pm} \sim \mathscr{E} \sim \mathscr{R}^0$, since otherwise the tangential electric field is not continuous across the conductor-vacuum interface. On the other hand, the external magnetic transients $\widetilde{\mathscr{B}}_{\pm}$ are small of order $\mathscr{R}^{-1} = (\mu_0\sigma a)^{-1}c^{-1}$ since the magnetic field energy flows with the speed of light towards the conducting cavity. The deeper physical reason for these electromagnetic transients is to be seen in the conservation laws for electromagnetic energy and momentum, which follow from Maxwell's equations.[12]

[1] B. M. Budak, A. A. Samarskii, and A. N. Tikhonov, *Mathematical Physics* (Pergamon, New York, 1964).
[2] W. W. Batygin and I. N. Toptygin, *Problems on Electrodynamics* (Academic, New York, 1964).
[3] G. Lehner, J. G. Linhart, and J. P. Somon, Nucl. Fusion **4**, 362 (1964).
[4] G. Lehner, Ergeb. Exakten Naturawiss. **40**, 67 (1968).
[5] Reference 1, p. 347.
[6] F. Winterberg, Atomkernenergie **35**, 34 (1980).
[7] S. C. Rashleigh and R. A. Marshall, J. Appl. Phys. **49**, 2540 (1978).
[8] C. M. Fowler, W. B. Garn, and R. S. Caird, J. Appl. Phys. **31**, 588 (1960).
[9] V. V. Poljudov, V. M. Titov, and G. A. Shvetsov, J. Appl. Mech. Tech. Phys. **6**, 41 (1973).
[10] H. Matsuo, Phys. Fluids **22**, 1618 (1979).
[11] A. Sakurai and T. Takao, J. Phys. Soc. Jpn. **28**, 1329 (1970).
[12] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1949).
[13] H. E. Wilhelm, Technical Report, Naval Weapons Center, 1981.

Some comments about the tensor virial theorem and orthogonal linear transformations

Francisco M. Fernández and Eduardo A. Castro
INIFTA, Sección Química Teórica, Sucursal 4, Casilla de Correo 16, La Plata 1900, Argentina

(Received 11 March 1980; accepted for publication 24 July 1981)

The tensor virial theorem is analyzed in relation to orthogonal linear transformations. The physical implications are discussed.

PACS numbers: 03.20. + i

I. INTRODUCTION

The tensorial generalization of the virial theorem (TVT) was introduced in classical mechanics by Parker[1] and it was recently reformulated by Miglietta.[2] The corresponding quantum tensor virial theorem was deduced by Pandres[3] and Cohen.[4]

The former author used a linear coordinate transformation plus the variational principles to make the demonstration. Such procedure allows one to obtain the equations that have to satisfy the eigenfunctions of the Hamiltonian operator, and furthermore the conditions under which approximate functions satisfy those equations. Later, Cohen deduced the TVT from the Heisenberg equations of motion. Recently, we have discussed the importance of the groups of transformation in relation to the TVT.[5,6]

Pandres' procedure[3] consists of the insertion of n^2 independent parameters in the trial wave function, through an $n\times n$ square matrix, and *a posterior* imposition of the extremum conditions of the energy functional with respect to such parameters. This method gives a set of n^2 independent equations which compose the TVT.

Naturally, when the number of independent parameters is less than n^2, the TVT will be satisfied in a partial and incomplete way.

The purpose of this communication is to show which is the class of equations that will be fulfilled when an orthogonal matrix is used. The physical implications will be self-evident.

II. ORTHOGONAL MATRIX

An orthogonal matrix $C_{n\times n}$ satisfies the well-known relationships

$$C^tC = CC^t = I_{n\times n}, \tag{1}$$

$$C_i^tC_j = \delta_{ij}, i,j = 1,2,...,n, \tag{2}$$

where C^t is the transpose matrix of C, I is the identity matrix, and C_i represents the ith column of the matrix C. Equation (2) follows from Eq. (1) and it assures us that C contains just $n(n-1)/2 \equiv s$ independent elements. These independent elements will be denoted by $z_1,z_2,...,z_s$.

From Eq. (1) it is deduced that the s matrices A_i defined by the formulas

$$A_i = C^t\frac{\partial C}{\partial z_i}, \quad i = 1,2,...,s, \tag{3}$$

are antisymmetric.

III. TENSOR VIRIAL THEOREM AND ORTHOGONAL TRANSFORMATION

Let us consider a system composed of N identical particles, whose Hamiltonian is

$$H = \frac{1}{2m}\sum_{a=1}^{N}\sum_{j=1}^{3} p_{aj}^2 + V(X), \tag{4}$$

where m is the mass of each particle, and p_{aj} is the conjugated momentum of the jth Cartesian coordinate x_{aj} corresponding to the ath particle. X represents the set of coordinates of all the particles.

$V(X)$ takes into account internal as well as external potentials, i.e.,

$$V(X) = V_i(X) + V_e(X). \tag{5}$$

When $V(X)$ possesses a well defined symmetry, some relationships which constitute the TVT are trivial.[7,8] In order to be sure of the nontriviality of the relations that we will discuss, we assume that $V_e(X)$ is asymmetrical enough.

From a normalized well-behaved function $\phi(X)$, we can define the variational function $\phi(Y)$, where Y symbolizes the set of variables $\{y_{ai}\}$ given by the following orthogonal transformation

$$y_{ai} = \sum_{j=1}^{3} c_{ij}x_{aj}, \quad i = 1,2,3. \tag{6}$$

The value of the Jacobian of the transformation (6) is one, so that $\phi(Y)$ is normalized. The derivative of $\phi(Y)$ with respect to the s-variational parameters z_i ($i = 1,...,s$) is given by

$$\frac{\partial\phi(Y)}{\partial z_i} = \sum_a\sum_j\frac{\partial\phi(Y)}{\partial y_{aj}}\frac{\partial y_{aj}}{\partial z_i} = \sum_k\sum_t (A_i)_{tk}v_{kt}\phi(Y)$$

$$= \sum_k\sum_{t>k}(A_i)_{tk}(v_{kt} - v_{tk})\phi(Y), \tag{7}$$

where

$$v_{ij} = \sum_a x_{ai}\frac{\partial}{\partial x_{aj}}. \tag{8}$$

Imposing the extremum conditions to the functional E,

$$E(z_1,z_2,...,z_s) = \langle\phi(Y)|H\phi(Y)\rangle, \tag{9}$$

with respect to the s parameters z_i, we obtain at once

$$\frac{\partial E}{\partial z_i} = \sum_k\sum_{t>k}\langle[H,v_{kt} - v_{tk}]\rangle(A_i)_{tk} = 0. \tag{10}$$

If L_i denotes the ith component of the angular momentum operator of the system, then Eq. (10) is transformed into

$$\langle[H,L_i]\rangle = 0, \quad i = 1,2,3. \tag{11}$$

Evidently, the set of parameters introduced via an orthogonal matrix implies the conservation of the expectation value of the angular momentum.

ACKNOWLEDGMENT

We wish to thank the referee for illuminating and interesting comments and suggestions with regard to this paper.

[1] E. Parker, Phys. Rev. **96**, 1686 (1954).
[2] F. Miglietta, Lett. Nuovo Cimento **19**, 497 (1977).
[3] D. Pandres, Jr., Phys. Rev. **131**, 886 (1963).
[4] L. Cohen, J. Math. Phys. **19**, 1838 (1978).
[5] E. A. Castro and F. M. Fernández, Lett. Nuovo Cimento **28**, 484 (1980).
[6] F. M. Fernández and E. A. Castro, Rev. Bras. Fis. **11**, 145 (1981).
[7] S. T. Epstein, Theor. Chim. Acta **52**, 89 (1979).
[8] F. M. Fernández and E. A. Castro, Theor. Chim. Acta **59**, 87 (1981).

Gradient theorem for completely integrable Hamiltonian systems[a)]

M. D. Arthur and K. M. Case
Center for Studies of Nonlinear Dynamics, La Jolla Institute, Post Office Box 1434, La Jolla, California 92038
The Rockefeller University,[b)] New York, New York 10021

(Received 2 February 1982; accepted for publication 21 May 1982)

For evolution equations which can be written in Hamiltonian form two ways, there exists a relation between two functions $Q^{(1)}$ and $Q^{(2)}$, both of which are gradients of conserved functionals. The relation can be extended to define (recursively) functions $Q^{(n)}$. It is shown that the $Q^{(n)}$ corresponding to the general evolution equation associated with the Zakharov–Shabat eigenvalue problem are all gradients of conserved functionals. This in turn implies all these functionals are in involution.

PACS numbers: 03.40. – t, 03.65.Fd

I. INTRODUCTION

Previously[1] we have seen that the simple properties of most of the known completely integrable Hamiltonian systems follow directly from the existence of two Hamiltonian formulations.

Specifically, we have the following situation in mind: There is a nonlinear evolution equation of the form

$$u_t = K(u), \quad -\infty < x < \infty. \tag{1}$$

Between any two functionals F_i, F_j two different Poisson brackets of the form

$$[F_i,F_j] = \int_{-\infty}^{\infty} \frac{\delta F_i}{\delta u} L \frac{\delta F_j}{\delta u}\, dx$$

and

$$[F_i,F_j]' = \int_{-\infty}^{\infty} \frac{\delta F_i}{\delta u} M \frac{\delta F_j}{\delta u}\, dx \tag{2}$$

are defined. (Naturally these must be antisymmetric and satisfy the Jacobi identity).

Further, there exist two constants of motion of Eq. (1) (H,H') such that Eq. (1) can be written as

$$u_t = [u,H]$$

or

$$u_t = [u,H']'. \tag{3}$$

If we define

$$Q^{(1)} = \delta H'/\delta u,$$
$$Q^{(2)} = \delta H/\delta u,$$

these equations imply

$$LQ^{(2)} = MQ^{(1)}.$$

More generally, we can define functions $Q^{(n)}$ by the recursion relation

$$LQ^{(n+1)} = MQ^{(n)}. \tag{4}$$

Now if the $Q^{(n)}$ are gradients, i.e.,

$$Q^{(n)} = \frac{\delta I_n}{\delta u}, \tag{5}$$

it follows[1,2] that the I_n are constants of motion for Eq. (1) and they are in involution, i.e.,

$$[I_n,H] = 0 = [I_n,I_m]. \tag{6}$$

Here we give a proof that the $Q^{(n)}$ arising from all evolution equations associated with the Zakharov–Shabat[3] eigenvalue equation are indeed gradients.

The method of proof is the following. First, it is shown that a generating function for the $Q^{(n)}$ satisfies a set of three linear coupled first order equations. The idea is then to show that any solution of these equations is a gradient. This is done by relating solutions of the third order system to bilinear combinations of the Z–S eigenvalue problem. Using variational principles,[4] it is shown that these bilinear combinations are indeed gradients.

A special case here is the Kortweg–de Vries equation. Since, however, the mathematics for this case is possibly more familiar to the reader, we treat this first in Sec. II. In Sec. III, the general evolution equation associated with the Z–S problem is introduced. The related hierarchy of evolution equations is discussed in Sec. IV. The proof of the general gradient theorem is then given in Sec. V.

II. THE K–deV EQUATION

Two Hamiltonian forms for the K–deV equation have already been presented in Ref. 1. They are obtained from the Poisson brackets defined by Eq. (2) when

$$V = u, \tag{7}$$

$$L = \partial_x, \tag{8}$$

$$M = -2\partial_x^3 - \tfrac{4}{3}u\partial_x - \tfrac{2}{3}u_x, \tag{9}$$

$$H = \int_{-\infty}^{\infty} \left[\frac{u^3}{3} - (u_x)^2\right] dx, \tag{10}$$

and

$$H' = \int_{-\infty}^{\infty} \tfrac{1}{2}u^2\, dx. \tag{11}$$

Then it is easy to verify that

$$\partial_t u = [u,H] = [u,H']' = -2uu_x - 2u_{xxx}, \tag{12}$$

which established the specific form of the K–deV equation to be considered. The Q_n are generated by the recursion formula given by Eq. (4). It is to be shown that they are gradients.

---

a) Supported in part by the National Science Foundation under grant MCS 80-17781 and in part by the Office of Naval Research under Contract No. NOO 14-79-0537.
b) Permanent address.

 0022-2488/82/101771-07$02.50

We now consider the linear eigenvalue problem for the function defined by

$$\Psi = \sum_{n=-\infty}^{\infty} \left(\frac{1}{2\lambda}\right)^n Q_n. \tag{13}$$

Using Eqs. (4) and (13), we obtain

$$M\Psi = 2\lambda L\Psi, \tag{14}$$

or specifically

$$\lambda\Psi_x = \Psi_{xxx} + \tfrac{2}{3}u\Psi_x + \tfrac{1}{3}u_x\Psi. \tag{15}$$

This last equation is a third order linear differential equation for Ψ and has three linearly independent solutions.

The eigenvalue problem for Ψ, Eq. (15), is of a somewhat unfamiliar form. This is readily remedied, as in Ref. 1, by noticing that if $\hat{\Psi}$ satisfies the Schrödinger equation

$$\left(\partial_x^2 + \frac{u}{6}\right)\hat{\psi} = \frac{\lambda}{4}\hat{\psi}, \tag{16}$$

then

$$\Psi = \hat{\psi}^2 \tag{17}$$

satisfies Eq. (15).

We find it more convenient to consider two integral forms of the Schrödinger equation,

$$\hat{\psi}^{\pm} = \hat{\psi}_0^{\pm} + \mathscr{G}^{\pm}v\hat{\psi}^{\pm}, \tag{18}$$

where the operators $\mathscr{G}^{\pm}$ are defined by

$$(\mathscr{G}^{+}f)(x) = -\frac{1}{k}\int_x^{\infty} \sin k(x-x')f(x')\,dx', \tag{19a}$$

$$(\mathscr{G}^{-}f)(x) = \frac{1}{k}\int_{-\infty}^{x} \sin k(x-x')f(x')\,dx', \tag{19b}$$

where

$$k^2 = -\lambda/4 \tag{20}$$

and the $\hat{\psi}_0^{\pm}$ satisfy the equation

$$(\partial_x^2 + k^2)\hat{\psi}_0^{\pm} = 0. \tag{21}$$

Note that the substitution

$$v = -\frac{u}{6} \tag{22}$$

has been used. The $\pm$ superscripts denote two related scattering problems, each of which has two forms. Corresponding to the $+(-)$ superscript, we associate two functions ψ, $\bar{\psi}(\phi,\bar{\phi})$ defined by their asymptotic values at $+\infty$ $(-\infty)$,

$$\psi \to e^{ikx}, \quad x \to +\infty, \tag{23a}$$

$$\bar{\psi} \to e^{-ikx}, \quad x \to +\infty, \tag{23b}$$

$$\phi \to e^{ikx}, \quad x \to -\infty, \tag{23c}$$

$$\bar{\phi} \to e^{-ikx}, \quad x \to -\infty. \tag{23d}$$

We define reflection and transmission coefficients by the asymptotic relations

$$\psi \to \frac{1}{T}e^{ikx} + \frac{R}{T}e^{-ikx}, \quad x \to -\infty, \tag{24a}$$

$$\bar{\psi} \to \frac{1}{T^*}e^{-ikx} + \frac{R^*}{T^*}e^{ikx}, \quad x \to -\infty. \tag{24b}$$

Using the invariance of the Wronskian,

$$W(\psi^{(1)},\psi^{(2)}) = \psi^{(1)}\psi_x^{(2)} - \psi^{(2)}\psi_x^{(1)} \tag{25}$$

for all mixed pairs of the functions ψ, $\bar{\psi}$, ϕ, and $\bar{\phi}$, we find that

$$\phi \to \frac{1}{T^*}e^{ikx} - \frac{R}{T}e^{-ikx}, \quad x \to +\infty, \tag{26}$$

$$\bar{\phi} \to +\frac{1}{T}e^{-ikx} - \frac{R^*}{T^*}e^{ikx}, \quad x \to +\infty, \tag{27}$$

and

$$RR^* + TT^* = 1. \tag{28}$$

Functional forms for the reflection and transmission coefficients may be obtained from Eq. (18), thus

$$\frac{1}{T} = 1 - \frac{1}{2ik}\int_{-\infty}^{\infty} e^{-ikx'}v(x')\psi(x')\,dx' \tag{29a}$$

$$= 1 - \frac{1}{2ik}\int_{-\infty}^{\infty} e^{ikx'}v(x')\bar{\phi}(x')\,dx' \tag{29b}$$

$$= 1 - \frac{1}{2ik}\int_{-\infty}^{\infty} \psi(x')(v - v\mathscr{G}^{-}v)\bar{\phi}(x')\,dx', \tag{29c}$$

$$\frac{R}{T} = \frac{1}{2ik}\int_{-\infty}^{\infty} e^{ikx'}v(x')\psi(x')\,dx' \tag{30a}$$

$$= \frac{1}{2ik}\int_{-\infty}^{\infty} e^{ikx'}v(x')\phi(x')\,dx' \tag{30b}$$

$$= \frac{1}{2ik}\int_{-\infty}^{\infty} \phi(x')(v - v\mathscr{G}^{+}v)\psi(x')\,dx', \tag{30c}$$

and

$$\frac{R^*}{T^*} = -\frac{1}{2ik}\int_{-\infty}^{\infty} e^{-ikx'}v(x')\bar{\psi}(x')\,dx' \tag{31a}$$

$$= -\frac{1}{2ik}\int_{-\infty}^{\infty} e^{-ikx'}v(x')\bar{\phi}(x')\,dx' \tag{31b}$$

$$= -\frac{1}{2ik}\int_{-\infty}^{\infty} \bar{\psi}(x')(v - v\mathscr{G}^{-}v)\bar{\phi}(x')\,dx'. \tag{31c}$$

In order to obtain expressions which are stationary with respect to variations of ψ, $\bar{\psi}$, ϕ, and $\bar{\phi}$, we define

$$A \equiv \frac{\int_{-\infty}^{\infty} e^{-ikx'}v(x')\psi(x')\,dx' \int_{-\infty}^{\infty} e^{ikx'}v(x')\bar{\phi}(x')\,dx'}{\int_{-\infty}^{\infty} \psi(x')(v - v\mathscr{G}^{-}v)\bar{\phi}(x')\,dx'} = 2ik\left(1 - \frac{1}{T}\right), \tag{32}$$

$$B \equiv \frac{\int_{-\infty}^{\infty} e^{ikx'}v(x')\psi(x')\,dx' \int_{-\infty}^{\infty} e^{ikx'}v(x')\phi(x')\,dx'}{\int_{-\infty}^{\infty} \phi(x')(v - v\mathscr{G}^{+}v)\psi(x')\,dx'} = 2ik\frac{R}{T}, \tag{33}$$

and

$$B^* \equiv \frac{\int_{-\infty}^{\infty} e^{-ikx'}v(x')\bar{\phi}(x')\,dx' \int_{-\infty}^{\infty} e^{-ikx'}v(x')\bar{\psi}(x')\,dx'}{\int_{-\infty}^{\infty} \bar{\psi}(x')(v - v\mathscr{G}^{-}v)\bar{\phi}(x')\,dx'} = 2ik\frac{R^*}{T^*}. \tag{34}$$

That A, B, and B^* are stationary as claimed may be verified by direct computation. Therefore, the variational derivatives of A, B, and B^* may be computed by considering only the explicit dependence on v. From Eq. (32) we have

$$\delta A = \int_{-\infty}^{\infty} e^{-ikx'}\delta v(x')\psi(x')\,dx' + \int_{-\infty}^{\infty} e^{ikx'}\delta v(x')\bar{\phi}(x')\,dx'$$
$$- \int_{-\infty}^{\infty} \psi(x')\delta v(x')\bar{\phi}(x')\,dx'$$
$$+ \int_{-\infty}^{\infty} \psi(x')\delta v(x')\mathscr{G}^{-} v\bar{\phi}(x')\,dx'$$
$$+ \int_{-\infty}^{\infty} \psi(x')v(x')\mathscr{G}^{-}\delta v\bar{\phi}(x')\,dx'. \tag{35}$$

From Eqs. (18) and (23d) we have

$$\mathscr{G}^{-}v\bar{\phi} = \bar{\phi} - e^{-ikx}, \tag{36}$$

so that Eq. (35) reduces to

$$\delta A = \int_{-\infty}^{\infty} e^{ikx'}\delta v(x')\bar{\phi}(x')\,dx'$$
$$+ \int_{-\infty}^{\infty} \psi(x')v(x')\mathscr{G}^{-}\delta v\bar{\phi}(x')\,dx'. \tag{37}$$

The second term on the right-hand side of Eq. (37) may be integrated by parts,

$$\int_{-\infty}^{\infty} \Big[-\frac{d}{dx'} \int_{x'}^{\infty} \psi(x)v(x)e^{ikx}\,dx\,\frac{1}{2ik}$$
$$\times \int_{-\infty}^{x'} e^{-ikx''}\delta v(x'')\bar{\phi}(x'')\,dx'' \Big]\,dx'$$
$$- \int_{-\infty}^{\infty} \Big[-\frac{d}{dx'} \int_{x'}^{\infty} \psi(x)v(x)e^{-ikx}\,dx\,\frac{1}{2ik}$$
$$\times \int_{-\infty}^{x'} e^{ikx''}\delta v(x'')\bar{\phi}(x'')\,dx'' \Big]\,dx'$$
$$= \int_{-\infty}^{\infty} (\mathscr{G}^{+}v\Psi)(x')\delta v(x')\bar{\Phi}(x')\,dx'. \tag{38}$$

Now, from Eqs. (18) and (23a) we have

$$\mathscr{G}^{+}v\psi = \psi - e^{ikx}, \tag{39}$$

so that Eq. (37) becomes

$$\delta A = \int_{-\infty}^{\infty} \psi(x')\delta v(x')\bar{\phi}(x')\,dx', \tag{40}$$

and we conclude that

$$\frac{\delta A}{\delta v} = \psi\bar{\phi}. \tag{41}$$

In a similar manner we find

$$\frac{\delta B}{\delta v} = \phi\psi \tag{42}$$

and

$$\frac{\delta B^*}{\delta v} = \bar{\phi}\bar{\psi}. \tag{43}$$

There are only two linearly independent solutions of Eq. (16). From the asymptotic forms (23a), (23b), (26), and (27), and Eq. (28), we may write

$$\phi = \frac{1}{T^*}\psi - \frac{R}{T}\bar{\psi} \tag{44a}$$

and

$$\bar{\phi} = \frac{R^*}{T^*}\psi + \frac{1}{T}\bar{\psi}. \tag{44b}$$

Therefore, ϕ and $\bar{\phi}$ may be eliminated from Eqs. (41), (42), and (43). If we let S denote some linear combination of the scattering data, then $\delta S/\delta v$ will be a linear combination of ψ^2, $\psi\bar{\psi}$, and $\bar{\psi}^2$. Now, $\bar{\psi}^2$ and ψ^2 are solutions of Eq. (15) since each of ψ and $\bar{\psi}$ is a solution of the Schrödinger equation. But, $(\psi + \bar{\psi})$ is also a solution of the Schrödinger equation and therefore $(\psi + \bar{\psi})^2$ is a solution of Eq. (15). Thus, by the linearity of Eq. (15), $\psi\bar{\psi}$ is a solution. The three solutions ψ^2, $\psi\bar{\psi}$, and $\bar{\psi}^2$ are linearly independent since the Wronskian

$$\mathscr{W}(\psi^2,\psi\bar{\psi},\bar{\psi}^2) = \left\| \begin{matrix} \psi^2 & \psi\bar{\psi} & \bar{\psi}^2 \\ (\psi^2)_x & (\psi\bar{\psi})_x & (\bar{\psi}^2)_x \\ (\psi^2)_{xx} & (\psi\bar{\psi})_{xx} & (\bar{\psi}^2)_{xx} \end{matrix} \right\| = 2ik \tag{45}$$

does not vanish. Thus, the general solution of Eq. (15) satisfies

$$\Psi = \frac{\delta S}{\delta v}. \tag{46}$$

This last equation may be used to compare the Laurent expansion for S with Eq. (13). This yields the result that each Q_n is immediately seen to be the variational derivative of some functional. By virtue of the discussion in Sec. I, we have proved that the dual Hamiltonian structure of the K–deV equation implies that the K–deV equation constitutes a completely integrable Hamiltonian system.

III. EQUATIONS ASSOCIATED WITH THE ZAKHAROV–SHABAT PROBLEM

Consider the system of two coupled nonlinear evolution equations given by

$$\partial_x q = -\alpha(q_{xx} - 2q^2 r), \tag{47}$$
$$\partial_x r = \alpha(r_{xx} - 2r^2 q). \tag{48}$$

Since there are two unknowns, we must extend our notation for variational derivatives. Let

$$\mathbf{V} = \begin{pmatrix} 0 & q \\ r & 0 \end{pmatrix}. \tag{49}$$

Then, for any functional F, define

$$\frac{\delta F}{\delta \mathbf{V}} = \begin{pmatrix} \delta F/\delta q \\ \delta F/\delta r \end{pmatrix}. \tag{50}$$

We shall demonstrate that the eigenvalue problem associated with the system of equations (47) and (48) is the Zakharov–Shabat problem. First, the dual Hamiltonian structure of the nonlinear system may be verified by choosing

$$\mathbf{L} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{51}$$

$$\mathbf{M} = \alpha\begin{pmatrix} 2q\partial_x^{-1}q & \partial_x - 2q\partial_x^{-1}r \\ \partial_x - 2r\partial_x^{-1}q & 2r\partial_x^{-1}r \end{pmatrix}, \tag{52}$$

$$\partial_x^{-1} = \frac{1}{2}\Big(\int_{-\infty}^{x} \cdots dx' - \int_{x}^{\infty} \cdots dx'\Big), \tag{53}$$

$$H = \alpha\int_{-\infty}^{\infty} (q_x r_x + q^2 r^2)\,dx, \tag{54}$$

$$H' = -\frac{1}{2}\int_{-\infty}^{\infty} (rq_x - qr_x)\,dx, \tag{55}$$

and

$$\mathbf{u} = \begin{pmatrix} q \\ r \end{pmatrix}, \tag{56}$$

so that

$$\mathbf{u}_t = [\mathbf{u}, H] = [\mathbf{u}, H']. \tag{57}$$

Here, the two Poisson brackets are the natural extention of Eq. (2), i.e., for any two functionals F_i, F_j,

$$[F_i, F_j] = \int_{-\infty}^{\infty} \left(\frac{\delta F_i}{\delta q}, \frac{\delta F_i}{\delta r}\right) \mathbf{L} \begin{pmatrix} \delta F_j/\delta q \\ \delta F_j/\delta r \end{pmatrix} dx, \tag{58}$$

and similarly for $[,]'$ with $\mathbf{L}$ replaced by $\mathbf{M}$.

We now define an infinite sequence of two component functions, $\mathbf{Q}_n$, by the recursion formula

$$\mathbf{L}\mathbf{Q}_{n+1} = \mathbf{M}\mathbf{Q}_n. \tag{59}$$

Analogous to the K–deV case, we consider a two-component function

$$\mathbf{\Psi} = \sum_{n=-\infty}^{\infty} \left(\frac{1}{\lambda}\right)^n \mathbf{Q}_n. \tag{60}$$

The linear eigenvalue problem for $\mathbf{\Psi}$ is obtained from Eqs. (59) and (60),

$$\mathbf{M}\bar{\mathbf{\Psi}} = \lambda \mathbf{L}\bar{\mathbf{\Psi}} \tag{61}$$

or, in component form,

$$-2i\zeta\Psi_2 = 2q\partial_x^{-1}(q\Psi_1) + \partial_x\Psi_2 - 2q\partial_x^{-1}(r\Psi_{2x}), \tag{62}$$

$$2i\zeta\Psi_1 = \partial_x\Psi_1 - 2r\partial_x^{-1}(q\Psi_1) + 2r\partial_x^{-1}(r\Psi_2), \tag{63}$$

where

$$\zeta = -\tfrac{1}{2}i\alpha\lambda. \tag{64}$$

The third-order character of Eqs. (62) and (63) becomes obvious when we define Ψ_3 by

$$\partial_x\Psi_3 = q\Psi_1 - r\Psi_2, \tag{65}$$

so that we obtain

$$-\partial_x\Psi_2 - 2i\zeta\Psi_2 = 2q\Psi_3 \tag{66}$$

and

$$\partial_x\Psi_1 + 2i\zeta\Psi_1 = 2r\Psi_3. \tag{67}$$

A second-order system of equations may be related to Eqs. (66) and (67) by the substitution

$$\Psi_1 = \hat{\psi}_2^2, \quad \Psi_2 = -\hat{\psi}_1^2, \quad \text{and } \Psi_3 = \hat{\psi}_1\hat{\psi}_2, \tag{68}$$

so that we find

$$\partial_x\hat{\psi}_1 + i\zeta\hat{\psi}_1 = q\hat{\psi}_2 \tag{69}$$

and

$$\partial_x\hat{\psi}_2 - i\zeta\hat{\psi}_2 = r\hat{\psi}_1. \tag{70}$$

These last two equations constitute the Zakharov–Shabat eigenvalue problem for $\bar{\psi}_1$ and $\bar{\psi}_2$.

IV. THE HIERARCHY OF EQUATIONS

If we choose

$$\mathbf{Q}_0 = \begin{pmatrix} r \\ q \end{pmatrix} \tag{71}$$

and define

$$\mathscr{L} = (\alpha)^{-1}\mathbf{L}^{-1}\mathbf{M}, \tag{72}$$

then it is easy to verify that

$$\mathscr{L}Q_0 = -\frac{\delta H'}{\delta \mathbf{V}}. \tag{73}$$

Using Eqs. (57), (58), and (72), the system of equations (47) and (48) may be written as

$$\mathbf{u}_t = -\alpha\mathbf{L}\mathscr{L}^2\mathbf{Q}_0. \tag{74}$$

The recursion formula for the $\mathbf{Q}_n$ and the definition of $\mathscr{L}$ lead us to consider a hierarchy of equations given by

$$\mathbf{u}_t = (-\alpha)L\mathscr{L}^j\mathbf{Q}_0, \quad j = \lambda \pm 1, \lambda \pm 2, \ldots . \tag{75}$$

Each equation in this hierarchy is of dual-Hamiltonian form and each has associated with it the same set of $\mathbf{Q}$'s defined by Eq. (59). The system of equations we started with corresponds to $j = 2$. In that system, the substitution $\alpha = -i$, $r = -\sigma q^*$ ($\sigma = \pm 1$) gives the NLS equation

$$\partial_t q = i(q_{xx} + 2\sigma q^* q^2) \tag{76}$$

and its complex conjugate

$$\partial_t q^* = -i(q^*_{xx} + 2\sigma q q^{*2}). \tag{77}$$

Other equations in the hierarchy are

$$j = 3, \quad \alpha = -2, \quad r = -\tfrac{1}{6}, \quad q = u \leftrightarrow \text{K–deV}$$

[cf. Eq. (12)] and

$$j = -1, \quad \alpha = \tfrac{1}{2}, \quad q = -r = u_x/\sqrt{2} \leftrightarrow \text{sine-Gordon}.$$

These classic examples of nonlinear evolution equations are well known to be completely integrable. However, the hierarchy of equations considered here contains infinitely many nonlinear evolution equations, each of which is completely integrable provided we can prove that the $\mathbf{Q}_n$ are gradients. Furthermore, these properties of dual-Hamiltonian form and complete integrability would be shared by any system of two coupled nonlinear evolution equations that can be put in the form

$$\mathbf{u}_t = \mathbf{L}f(\mathscr{L})\mathbf{Q}_0, \tag{78}$$

where f is any entire function. In the next section we shall prove that the $\mathbf{Q}_n$ are gradients of the scattering data for the scattering problem associated with the Zakharov–Shabat eigenvalue problem.

V. ZAKHAROV-SHABAT SCATTERING PROBLEM

The coupled system given by Eqs. (69) and (70) may be written in vector form as

$$\hat{\boldsymbol{\psi}}_x + i\zeta\boldsymbol{\sigma}_3\hat{\boldsymbol{\psi}} = \mathbf{V}\hat{\boldsymbol{\psi}}, \tag{79}$$

where

$$\hat{\boldsymbol{\psi}} = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \tag{80}$$

and

$$\boldsymbol{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{81}$$

There are two useful integral forms of Eq. (79) given by

$$\hat{\boldsymbol{\psi}}^{\pm} = \hat{\boldsymbol{\psi}}_0^{\pm} + \mathscr{G}^{\pm}\mathbf{V}\hat{\boldsymbol{\psi}}^{\pm}, \tag{82}$$

where

$$(\mathscr{G}^{+}\mathbf{f})(x) = -\int_{x}^{\infty} \exp[i\zeta(x-x')\boldsymbol{\sigma}_3]\,\mathbf{f}(x')\,dx', \tag{83}$$

$$(\mathscr{G}^{-}\mathbf{f})(x) = \int_{-\infty}^{x} \exp[i\zeta(x-x')\boldsymbol{\sigma}_3]\,\mathbf{f}(x')\,dx', \tag{84}$$

and $\hat{\boldsymbol{\psi}}_0^{\pm}$ satisfies

$$(\partial_x + i\zeta\boldsymbol{\sigma}_3)\hat{\boldsymbol{\psi}}_0^{\pm} = 0. \tag{85}$$

The $\pm$ superscripts denote two related scattering problems, each of which has two forms. Corresponding to the $-\,(+)$ superscript, we associate two vector functions $\boldsymbol{\phi},\bar{\boldsymbol{\phi}}(\boldsymbol{\psi},\bar{\boldsymbol{\psi}})$ defined by their asymptotic values at $-\infty\,(+\infty)$,

$$\boldsymbol{\phi} \to \begin{pmatrix}1\\0\end{pmatrix} e^{-i\zeta x} \equiv \hat{\boldsymbol{\psi}}_0^{-(1)}, \quad x \to -\infty, \tag{86a}$$

$$\bar{\boldsymbol{\phi}} \to \begin{pmatrix}0\\-1\end{pmatrix} e^{i\zeta x} \equiv \hat{\boldsymbol{\psi}}_0^{-(2)}, \quad x \to -\infty, \tag{86b}$$

$$\boldsymbol{\psi} \to \begin{pmatrix}0\\1\end{pmatrix} e^{i\zeta x} \equiv \hat{\boldsymbol{\psi}}_0^{+(1)}, \quad x \to +\infty, \tag{86c}$$

and

$$\hat{\boldsymbol{\psi}} \to \begin{pmatrix}1\\0\end{pmatrix} e^{-i\zeta x} \equiv \hat{\boldsymbol{\psi}}_0^{+(2)} x \to +\infty. \tag{86d}$$

[Note that these asymptotic forms are solutions of Eq. (85) and have been so indicated by additional superscripts.] We define the scattering data by the asymptotic relations

$$\boldsymbol{\phi} \to \begin{pmatrix} a e^{-i\zeta x} \\ b e^{i\zeta x}\end{pmatrix}, \quad x \to +\infty, \tag{87a}$$

$$\bar{\boldsymbol{\phi}} \to \begin{pmatrix} \bar{b} e^{-i\zeta x} \\ -\bar{a} e^{i\zeta x}\end{pmatrix}, \quad x \to +\infty. \tag{87b}$$

Using the invariance of the Wronskian,

$$\hat{W}(\boldsymbol{\psi}^{(1)},\boldsymbol{\psi}^{(2)}) = \psi_1^{(1)}\psi_2^{(2)} - \psi_2^{(1)}\psi_1^{(2)}, \tag{88}$$

for all mixed pairs of vectors $\phi,\bar{\phi},\psi$, and $\bar{\psi}$, we find that

$$\boldsymbol{\psi} \to \begin{pmatrix} \bar{b} e^{-i\zeta x} \\ a e^{i\zeta x}\end{pmatrix}, \quad x \to -\infty, \tag{89a}$$

$$\bar{\boldsymbol{\psi}} \to \begin{pmatrix} \bar{a} e^{-i\zeta x} \\ -b e^{i\zeta x}\end{pmatrix}, \quad x \to -\infty, \tag{89b}$$

and

$$a\bar{a} + b\bar{b} = 1. \tag{90}$$

Functional forms for the scattering data may be obtained from Eqs. (82), (86), (87), and (89),

$$a = 1 + \int_{-\infty}^{\infty} e^{+i\zeta x} q(x)\phi_2(x)\,dx \tag{91a}$$

$$= 1 - \int_{-\infty}^{\infty} e^{-i\zeta x} r(x)\psi_1(x)\,dx, \tag{91b}$$

$$b = \int_{-\infty}^{\infty} e^{-i\zeta x} r(x)\phi_1(x)\,dx \tag{92a}$$

$$= \int_{-\infty}^{\infty} e^{-i\zeta x} r(x)\bar{\psi}_1(x)\,dx, \tag{92b}$$

$$\bar{a} = 1 - \int_{-\infty}^{\infty} e^{-i\zeta x} r(x)\bar{\phi}_1(x)\,dx \tag{93a}$$

$$= 1 - \int_{-\infty}^{\infty} e^{i\zeta x} q(x)\bar{\psi}_2(x)\,dx, \tag{93b}$$

$$\bar{b} = \int_{-\infty}^{\infty} e^{i\zeta x} q(x)\bar{\phi}_2(x)\,dx \tag{94a}$$

$$= -\int_{-\infty}^{\infty} e^{i\zeta x} q(x)\psi_2(x)\,dx. \tag{94b}$$

Additional functional forms for the scattering data may be obtained if we consider an equation for adjoint functions defined by

$$\hat{\boldsymbol{\psi}}_x^{\dagger} - i\zeta\boldsymbol{\sigma}_3\hat{\boldsymbol{\psi}}^{\dagger} = -V^T\hat{\boldsymbol{\psi}}^{\dagger}. \tag{95}$$

If Eq. (95) is written in component form, it is easy to verify that the choice

$$\hat{\psi}_1^{\dagger} = \hat{\psi}_2, \quad \hat{\psi}_2^{\dagger} = -\hat{\psi}_1 \tag{96}$$

returns us to Eq. (79). We will have need of integral forms of the adjoint equation (95). The formula

$$\int_{-\infty}^{\infty} \mathbf{f}(x)\cdot\mathscr{G}^{\pm}\boldsymbol{g}(x)\,dx = -\int_{-\infty}^{\infty} \bar{\mathscr{G}}^{\mp}\mathbf{f}(x)\cdot\boldsymbol{g}(x)\,dx \tag{97}$$

may be established by integrating by parts (here, $\bar{\mathscr{G}}^{\mp}$ is the complex conjugate of $\mathscr{G}^{\mp}$). This suggests defining the integral form of the adjoint equations to be

$$\boldsymbol{\phi}^{\dagger} = (\hat{\boldsymbol{\psi}}_0^{-(1)})^{\dagger} - \bar{\mathscr{G}}^{-}\mathbf{V}^T\boldsymbol{\phi}^{\dagger}, \tag{98a}$$

$$\bar{\boldsymbol{\phi}}^{\dagger} = (\hat{\boldsymbol{\psi}}_0^{-(2)})^{\dagger} - \bar{\mathscr{G}}^{-}\mathbf{V}^T\bar{\boldsymbol{\phi}}^{\dagger}, \tag{98b}$$

$$\hat{\boldsymbol{\psi}}^{\dagger} = (\hat{\boldsymbol{\psi}}_0^{+(1)})^{\dagger} - \bar{\mathscr{G}}^{+}\mathbf{V}^T\boldsymbol{\psi}^{\dagger}, \tag{98c}$$

and

$$\bar{\boldsymbol{\psi}}^{\dagger} = (\hat{\boldsymbol{\psi}}_0^{+(2)})^{\dagger} - \bar{\mathscr{G}}^{+}\mathbf{V}^T\bar{\boldsymbol{\psi}}^{\dagger}. \tag{98d}$$

We now observe that

$$\int_{-\infty}^{\infty} \boldsymbol{\psi}^{\dagger}(x)(\mathbf{V} - \mathbf{V}\mathscr{G}^{-}\mathbf{V})\phi(x)\,dx$$

$$= -\int_{-\infty}^{\infty} \boldsymbol{\psi}^{\dagger}(x)\hat{\mathbf{V}}\psi_0^{-(1)}(x)\,dx \quad \text{[by Eq. (82)]}$$

$$= -\int_{-\infty}^{\infty} \psi_1(x)r(x)e^{-i\zeta x}\,dx \quad \text{[by Eqs. (86a) and (96)]}$$

$$= a - 1 \quad \text{[by Eq. (91b)]}. \tag{99}$$

In an analgous manner, we find additional forms for the rest of the scattering data,

$$\bar{a} - 1 = \int_{-\infty}^{\infty} \bar{\psi}^{\dagger}(x)\cdot(\mathbf{V} - \mathbf{V}\mathscr{G}^{-}\mathbf{V})\bar{\phi}(x)\,dx, \tag{100}$$

$$b = -\int_{-\infty}^{\infty} \boldsymbol{\phi}^{\dagger}(x)\cdot(\mathbf{V} - \mathbf{V}\mathscr{G}^{+}\mathbf{V})\bar{\boldsymbol{\psi}}(x)\,dx, \tag{101}$$

$$b = -\int_{-\infty}^{\infty} \bar{\boldsymbol{\phi}}^{\dagger}(x)\cdot(\mathbf{V} - \mathbf{V}\mathscr{G}^{+}\mathbf{V})\boldsymbol{\psi}(x)\,dx. \tag{102}$$

We now combine the three forms for each of $a-1$, $\bar{a}-1$, b, and $\bar{b}$ to obtain

$$a - 1 = \frac{\int_{-\infty}^{\infty}(\hat{\boldsymbol{\psi}}_0^{+(1)}(x))^{\dagger}\cdot\mathbf{V}\boldsymbol{\phi}(x)\,dx\int_{-\infty}^{\infty}(\hat{\boldsymbol{\psi}}_0^{-(1)}(x))^{\dagger}\cdot\mathbf{V}\boldsymbol{\psi}(x)\,dx}{\int_{-\infty}^{\infty}\boldsymbol{\psi}^{\dagger}(x)\cdot(\mathbf{V} - \mathbf{V}\mathscr{G}^{-}\mathbf{V})\phi(x)\,dx}, \tag{103}$$

$$\bar{a}-1=\frac{\int_{-\infty}^{\infty}(\hat{\boldsymbol\psi}_0^{+(2)}(x))^\dagger\cdot\mathbf{V}\bar{\boldsymbol\phi}(x)\,dx\int_{-\infty}^{\infty}(\hat{\boldsymbol\psi}_0^{-(2)}(x))^\dagger\cdot V\bar{\boldsymbol\psi}(x)\,dx}{\int_{-\infty}^{\infty}\bar{\phi}^\dagger(x)\cdot(\mathbf{V}-\mathbf{V}\mathscr{G}^+\mathbf{V})\bar{\psi}(x)\,dx},\tag{104}$$

$$b=-\frac{\int_{-\infty}^{\infty}(\hat{\boldsymbol\psi}_0^{-(1)}(x))^\dagger\mathbf{V}\bar{\psi}(x)\,dx\int_{-\infty}^{\infty}(\hat{\boldsymbol\psi}_0^{-(2)}(x))^\dagger\mathbf{V}\boldsymbol\phi(x)\,dx}{\int_{-\infty}^{\infty}\boldsymbol\Psi^\dagger(x)\cdot(\mathbf{V}-\mathbf{V}\mathscr{G}^-\mathbf{V})\phi(x)\,dx},\tag{105}$$

$$\bar{b}=\frac{\int_{-\infty}^{\infty}(\hat{\boldsymbol\psi}_0^{+(1)}(x))^\dagger\cdot\mathbf{V}\bar{\phi}(x)\,dx\int_{-\infty}^{\infty}(\hat{\boldsymbol\psi}_0^{-(2)}(x))^\dagger\cdot\mathbf{V}\boldsymbol\psi(x)\,dx}{\int_{-\infty}^{\infty}\bar{\boldsymbol\phi}^\dagger(x)\cdot(\mathbf{V}-\mathbf{V}\mathscr{G}^+\mathbf{V})\psi(x)\,dx},\tag{106}$$

The advantage of these seemingly more complicated expressions for the scattering data [Eqs. (103)–(106)] is that each one is stationary with respect to variations of $\boldsymbol\phi$, $\bar{\boldsymbol\phi}$, $\boldsymbol\psi$, and $\bar{\boldsymbol\psi}$ [this can easily be proved using Eqs. (97) and (98)]. Therefore, in order to compute the variational derivatives, we need only consider the explicit dependence on V. We shall perform the calculation for $(a-1)$. First,

$$\begin{aligned}\delta(a-1)=&\int_{-\infty}^{\infty}(\hat{\boldsymbol\psi}_0^{+(1)}(x))^\dagger\cdot\delta\mathbf{V}\boldsymbol\phi(x)\,dx\\&+\int_{-\infty}^{\infty}(\hat{\boldsymbol\psi}_0^{-(1)}(x))^\dagger\cdot\delta\mathbf{V}\boldsymbol\psi(x)\,dx\\&-\int_{-\infty}^{\infty}\boldsymbol\psi^\dagger(x)\cdot\delta\mathbf{V}\boldsymbol\phi(x)\\&+\int_{-\infty}^{\infty}\boldsymbol\psi^\dagger(x')\cdot\delta\mathbf{V}\mathscr{G}^-\mathbf{V}\boldsymbol\phi(x)\,dx\\&+\int_{-\infty}^{\infty}\psi^\dagger(x)\cdot\mathbf{V}\mathscr{G}^-\delta\mathbf{V}\boldsymbol\phi(x)\,dx.\end{aligned}\tag{107}$$

But, from Eq. (82) we have

$$\mathscr{G}^-\mathbf{V}\boldsymbol\phi=\boldsymbol\phi-\hat{\boldsymbol\psi}_0^{-(1)},\tag{108}$$

so that with Eq. (97) we find

$$\begin{aligned}\delta(a-1)=&\int_{-\infty}^{\infty}(\hat{\boldsymbol\psi}_0^{+(1)}(x)^\dagger\cdot\delta\mathbf{V}\boldsymbol\phi(x)\,dx\\&-\int_{-\infty}^{\infty}(\overline{\mathscr{G}}^+\mathbf{V}^T\boldsymbol\psi^\dagger(x))\cdot\delta\mathbf{V}\boldsymbol\phi(x)\,dx.\end{aligned}\tag{109}$$

Now we use Eq. (98c) to obtain

$$\delta(a-1)=\int_{-\infty}^{\infty}\boldsymbol\psi^\dagger(x)\cdot\delta V\boldsymbol\phi(x)\,dx;\tag{110}$$

thus, using Eqs. (96),

$$\frac{\delta a}{\delta q}=\psi_2\phi_2\quad\text{and}\quad\frac{\delta a}{\delta r}=-\psi_1\phi_1,\tag{111}$$

and from Eq. (50) we may write

$$\frac{\delta a}{\delta v}=\begin{pmatrix}\psi_2\psi_2\\-\psi_1\phi_1\end{pmatrix}.\tag{112}$$

Similarly, one finds

$$\frac{\delta\bar{a}}{\delta v}=\begin{pmatrix}\bar{\psi}_2\bar{\psi}_2\\-\bar{\psi}_1\bar{\phi}_1\end{pmatrix},\tag{113}$$

$$\frac{\delta b}{\delta v}=\begin{pmatrix}\bar{\psi}_2\hat{\phi}_2\\-\bar{\psi}_1\hat{\phi}_1\end{pmatrix},\tag{114}$$

and

$$\frac{\delta\bar{b}}{\delta v}=\begin{pmatrix}\psi_2\bar{\phi}_2\\-\psi_1\bar{\phi}_1\end{pmatrix}.\tag{115}$$

Now, since $\boldsymbol\psi$ and $\boldsymbol\phi$ are solutions of Eq. (79), the components of the vectors,

$$\hat{\boldsymbol\Psi}^{(1)}=\begin{pmatrix}\psi_2^2\\-\psi_1^2\\\psi_1\psi_2\end{pmatrix}\quad\text{and}\quad\hat{\boldsymbol\psi}^{(2)}=\begin{pmatrix}\phi_2^2\\-\phi_1^2\\\phi_1\phi_2\end{pmatrix},\tag{116}$$

are solutions of the system of equations (65)–(67). But, $\boldsymbol\psi+\boldsymbol\phi$ is also a solution of Eq. (79). Therefore, by the linearity of the system [Eqs. (65)–(67)],

$$\hat{\boldsymbol\psi}^{(3)}=\begin{pmatrix}\psi_2\phi_2\\-\psi_1\phi_1\\\frac{1}{2}[\psi_1\phi_2+\psi_2\phi_1]\end{pmatrix}\tag{117}$$

is also a solution of the third-order system. Note that the first and second components of $\hat{\Psi}^{(3)}$ [cf. Eq. (117)] are precisely the first and second components of $\delta(a-1)/\delta\mathbf{V}$ given by Eq. (112). Therefore, for the purpose of discussing linear independence, we will define

$$\Psi^{(a)}=\begin{pmatrix}\psi_2\phi_2\\-\psi_1\phi_1\\\frac{1}{2}(\psi_1\phi_2+\psi_2\phi_1)\end{pmatrix},\quad\Psi^{(b)}=\begin{pmatrix}\psi_2\phi_2\\-\psi_1\phi_1\\\frac{1}{2}(\psi_1\phi_2+\psi_2\phi_1)\end{pmatrix},$$

$$\Psi^{(\bar{a})}=\begin{pmatrix}\bar{\psi}_2\bar{\phi}_2\\-\bar{\psi}_1\bar{\phi}_1\\\frac{1}{2}(\bar{\psi}_1\bar{\phi}_2\bar{\psi}_2\bar{\phi}_1)\end{pmatrix},\quad\Psi^{(\bar{b})}=\begin{pmatrix}\psi_2\bar{\phi}_2\\-\psi_1\bar{\phi}_1\\\frac{1}{2}(\psi_1\bar{\phi}_2+\psi_2\bar{\phi}_1)\end{pmatrix}.$$

Then, the Wronskian of any three of these vectors may be computed. In particular we find

$$\begin{aligned}&\hat{\mathscr{W}}(\Psi^{(a)},\Psi^{(b)},\Psi^{(b)})\\&=\left|\left|\begin{matrix}\psi_2\phi_2&\bar{\psi}_2\phi_2&\psi_2\bar{\phi}_2\\-\psi_1\phi_1&-\bar{\psi}_1\phi_1&\psi_1\bar{\phi}_1\\\frac{1}{2}(\psi_1\phi_2+\psi_2\phi_1)&\frac{1}{2}(\bar{\psi}_1\phi_2+\psi_2\phi_1)&\frac{1}{2}(\psi_1\bar{\phi}_2+\psi_2\bar{\phi}_1)\end{matrix}\right|\right|\\&=\frac{a}{2},\end{aligned}\tag{118}$$

$$\mathscr{W}(\Psi^{(a)},\Psi^{(\bar{a})},\Psi^{(b)})=\frac{b}{2},\tag{119}$$

and

$$\mathscr{W}(\Psi^{(a)},\Psi^{(\bar{a})},\Psi^{(\bar{b})})=-\frac{\bar{b}}{2}.\tag{120}$$

From Eq. (90) it is not possible for both a and b, or a and $\bar{b}$, to be zero; thus we may always find three linearly independent variational derivatives of the scattering data.

The general solution to the eigenvalue problem for $\boldsymbol\Psi$ [cf. Eq. (61)] may be written as a linear combination of three of the variational derivatives of the scattering data. Then we may write

$$\boldsymbol\Psi=\frac{\delta S}{\delta V},\tag{121}$$

where S is some linear combination of those three terms of the scattering data. We expect the scattering data (and hence S) to have a Laurent expansion about $\lambda = 0$. S is a functional since the scattering data may be expressed as functionals [cf. Eqs. (103)–(106)]; therefore, each term in the Laurent expansion of S may be regarded as a functional, and comparison with Eqs. (60) and (121) shows that each $\mathbf{Q}_n$ is the variational derivative of some functional. As a consequence of the discussion in Sec. I, we have proved that the dual-Hamiltonian structure of the coupled pair of nonlinear evolution equations, (47) and (48), implies that those equations and all the equations in the related hierarchy are completely integrable Hamiltonian systems.

VI. CONCLUSION

It has been shown that the general evolution equation related to the Zakharov–Shabat eigenvalue problem can be written in Hamiltonian form two ways. This then implies a recursion relation for functions $Q^{(n)}$. It is proved that these are functional gradients. This in turn implies that the corresponding functionals are all constants and are in involution.

The key point in the proof is the existence of variational principles for the scattering problem for the Z–S equations.

A preliminary look at other completely integrable Hamiltonian systems suggests that similar proofs of the gradient property are also possible—and simple.

[1]K. M. Case, "Dual Hamiltonian Formalisms for Nonlinear Evolution Equations" M.I.T. Press (to be published); K. M. Case and A. Roos, J. Math. Phys. **23**, 392 (1982).

[2]F. Magri, J. Math. Phys. **19**, 1156–1162 (1978).

[3]V. E. Zakharov and A. B. Shabat, Sov. Phys. JETP **34**, 62 (1972); M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **LIII**, 249–336 (1974). We follow the notation in this paper.

[4]J. Schwinger, lecture notes (unpublished). This is discussed in M. L. Goldberger and K. M. Watson, *Collision Theory* (Wiley, New York, 1964), p. 320.

Wigner approach to quantization. Noncanonical quantization of two particles interacting via a harmonic potential

Tchavdar D. Palev[a)]

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

(Received 7 December 1981; accepted for publication 9 April 1982)

Following the ideas of Wigner, we quantize noncanonically a system of two nonrelativistic point particles, interacting via a harmonic potential. The center of mass phase-space variables are quantized in a canonical way, whereas the internal momentum and coordinates are assumed to satisfy relations, which are essentially different from the canonical commutation relations. As a result, the operators of the internal Hamiltonian, the relative distance, the internal momentum, and the orbital momentum commute with each other. The spectrum of these operators is finite. In particular, the distance between the constituents is preserved in time and can take at most four different values. The orbital momentum is either zero or one (in units $\hbar/2$). The operators of the coordinates do not commute with each other and, therefore, the position of any one of the constituents cannot be localized; the particles are smeared with a certain probability in a finite space volume, which moves together with the center of mass. In the limit $\hbar \to 0$ the constituents "fall" into their center of mass and the composite system behaves as a free point particle.

PACS numbers: 03.65. — w

INTRODUCTION

In ordinary, canonical quantum mechanics the operators of the Cartesian coordinates $\hat{q}_1,...,\hat{q}_n$ and momenta $\hat{p}_1,...,\hat{p}_n$, corresponding to a classical system with a Hamiltonian

$$H = \sum_{i=1}^{n} \frac{p_i^2}{2m_i} + U(q_1,...,q_n), \tag{1}$$

satisfy the canonical commutation relations (CCR's)

$$[\hat{q}_j,\hat{p}_k] = i\hbar\delta_{jk}, \quad [\hat{q}_j,\hat{q}_k] = [\hat{p}_j,\hat{p}_k] = 0. \tag{2}$$

The quantization with CCR's can be applied to any classical system, independently of the dynamics, i.e., for every Hamiltonian, and in this sense it is universal. In 1950 Wigner (See. Ref. 1, hereafter referred as to I) pointed out, however, that for a given Hamiltonian the canonical scheme can be in principle generalized. In particular, he has shown that the one-dimensional harmonic oscillator can be quantized in several noncanonical ways, i.e., with position and momentum operators that do not satisfy the CCR's (2). In the present paper we shall consider another example, quantizing noncanonically a system of two nonrelativistic point particles, interacting via a harmonic potential.

In order to motivate the definition of the noncanonical quantization, which we shall follow, consider the canonical quantum mechanics in the Heisenberg picture. In this case the time evolution of a given system is described by the Heisenberg equations of motion[2]

$$\dot{\hat{p}}_k = -\frac{i}{\hbar}[\hat{p}_k,\hat{H}], \quad \dot{\hat{q}}_k = -\frac{i}{\hbar}[\hat{q}_k,\hat{H}]. \tag{3}$$

The use of the CCR's then yield[3]

$$\frac{i}{\hbar}[\hat{p}_k,\hat{H}] = \frac{\partial \hat{U}}{\partial \hat{q}_k}, \quad \frac{i}{\hbar}[\hat{q}_k,\hat{H}] = -\frac{\hat{p}_k}{m_k}. \tag{4}$$

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

The relations (3) and (4) lead to operator equations, which formally coincide with the classical Hamiltonian equations[4]

$$\dot{\hat{p}}_k = -\frac{\partial \hat{U}}{\partial \hat{q}_k}, \quad \dot{\hat{q}}_k = \frac{\hat{p}_k}{m_k}. \tag{5}$$

Hence, the classical equations[5] are a simple consequence of the quantum equations (3) and the canonical commutation relations (2). The key point for a generalization of the concept of a quantization comes now from the observation of Wigner[1] that the Heisenberg Eqs. (3) and the Hamiltonian Eqs. (5) have a more immediate physical significance than the CCR's. From this point of view the CCR's appear only as a tool to derive the Hamiltonian equations. Therefore, it is logically justified to postulate from the very beginning Eqs. (5), instead of the conditions that lead to them, namely the CCR's.

On grounds of the above considerations, we define a (noncanonical) quantization of a given mechanical system with a Hamiltonian (1) as a replacement,

$$q_k \to \hat{q}_k, \quad p_k \to \hat{p}_k, \tag{6}$$

of the classical canonical variables by operators, so that the Heisenberg Eqs. (3) and the Hamiltonian Eqs. (5) will be simulaneously fulfilled.

The first question that arises in connection with the above definition is whether the new definition is more general than the canonical one. This can be the case if the compatibility relations (4), considered as equations with respect to the unknown operators $\hat{q}_1,...,\hat{q}_n,\hat{p}_1,...,\hat{p}_n$, also have solutions which are different from the canonical solution (2). In I Wigner has studied this problem in the case of a one-dimensional harmonic oscillator with a Hamiltonian ($\hbar = 1$)

$$\hat{H} = \tfrac{1}{2}(\hat{p}^2 + \hat{q}^2). \tag{7}$$

The solutions he found are labelled by one arbitrary nonnegative integer E_0, the energy of the ground state. The operators $\hat{p}$ and $\hat{q}$, corresponding to different E_0, are nonequivalent; their representation spaces $W(E_0)$ are infinite-dimensional. If $|E_0;n\rangle$, $n = 1,2,...$, is a basis in $W(E_0)$, then

$$\hat{q}|E_0;n\rangle = x_{n-1,n}|E_0;n-1\rangle + x_{n,n+1}|E_0;n+1\rangle,$$
$$\hat{p}|E_0;n\rangle = -ix_{n-1,n}|E_0;n-1\rangle + ix_{n,n+1}|E_0;n+1\rangle, \quad (8)$$

where

$$x_{n,n+1} = (E_0 + n/2)^{1/2} \quad \text{for even } n,$$
$$x_{n,n+1} = (n/2 + \tfrac{1}{2})^{1/2} \quad \text{for odd } n.$$

Only in the case $E_0 = \frac{1}{2}$ do the operators $\hat{p}$ and $\hat{q}$ satisfy the CCR's (2).

To establish some further algebraical properties of the operators $\hat{p}$ and $\hat{q}$, introduce the creation ($\xi = +$) and annihilation ($\xi = -$) operators (CAO's)

$$a^{\xi} = \frac{1}{\sqrt{2}}(\hat{q} - i\xi\hat{p}). \quad (9)$$

Then

$$\hat{H} = \tfrac{1}{2}\{a^+,a^-\} \quad (10)$$

and the CAO's transform the basis vectors as follows.

$$a^-|E_0;2n\rangle = (2n)^{1/2}|E_0;2n-1\rangle,$$
$$a^-|E_0;2n+1\rangle = (2n+2E_0)^{1/2}|E_0;2n\rangle,$$
$$a^+|E_0;2n\rangle = (2n+2E_0)^{1/2}|E_0;2n+1\rangle,$$
$$a^+|E_0;2n+1\rangle = (2n+2)^{1/2}|E_0;2n+2\rangle. \quad (11)$$

By a straightforward computation one shows that for every E_0 the operators a^+ and a^- satisfy one and the same relations, namely

$$[\{a^{\xi},a^{\eta}\},a^{\epsilon}] = (\epsilon - \xi)a^{\eta} + (\epsilon - \eta)a^{\xi}. \quad (12)$$

Here and throughout the paper $\xi,\eta,\epsilon,\delta = \pm$ or ± 1; $[x,y] = xy - yx$ and $\{x,y\} = xy + yx$.

Thus, the solutions for different values E_0 of a^+ and a^- appear as different irreducible representations of operators that satisfy the above equality (12). To every such representation there corresponds a self-consistent generalization of the ordinary quantization with position and momentum operators that are not unitarily equivalent to the canonical $\hat{q}$ and $\hat{p}$.

The results of Wigner can be easily extended to quantize noncanonically also a system of n noninteracting oscillators with a Hamiltonian

$$\hat{H} = \sum_{j=1}^{n}\left(\frac{1}{2m_j}\hat{p}_j^2 + \frac{m_j\omega_j^2}{2}\hat{q}_j^2\right). \quad (13)$$

In terms of the CAO's

$$a_k^{\xi} = \left(\frac{m_k\omega_k}{2\hbar}\right)^{1/2}\hat{q}_k - i\xi(2m_k\omega_k\hbar)^{-1/2}\hat{p}_k, \quad (14)$$

Eq. (4), which is a compatibility condition for Eqs. (3) and (5), reads ($\xi = \pm$)

$$\sum_{i=1}^{n}\frac{\omega_i\hbar}{2}[\{a_i^+,a_i^-\},a_k^{\xi}] = \xi\frac{\omega_k\hbar}{2}a_k^{\xi}. \quad (15)$$

One solution (among others) of the above equation is given with operators, which are a straightforward generalization of (12),

$$[\{a_i^{\xi},a_j^{\eta}\},a_k^{\epsilon}] = \delta_{ik}(\epsilon - \xi)a_j^{\eta} + \delta_{jk}(\epsilon - \eta)a_i^{\xi}. \quad (16)$$

The operators (16) are known in quantum field theory. They were introduced by Green[6] as a possible generalization of the statistics of integer-spin fields and are called para-Bose operators. The irreducible representations of the para-Bose operators, corresponding to Hermitian position and momentum operators and a nondegenerate ground state $|0\rangle$, are labelled by one non-negative integer p, the order of the statistics,[7] which is defined by the relations

$$a_i^-|0\rangle = 0, \quad a_i^-a_j^+|0\rangle = p\delta_{ij}|0\rangle, \quad p = 1,2,\ldots. \quad (17)$$

Only in the case $p = 1$ do the position and momentum operators, corresponding to a_i^{ξ}, obey the CCR's (2).

We see that the Wigner quantization of the one-dimensional harmonic oscillator (with $n = \hbar = \omega = 1$) is in fact a quantization with para-Bose operators. Therefore, it generalizes quantum mechanics along the same line as the para-Bose statistics extends quantum field theory. Different aspects of the Wigner quantization of the one-dimensional oscillator were studied in Ref. 8 and more recently in Refs. 9–11.

Since Eq. (4), which have to be satisfied by $\hat{p}_i$ and $\hat{q}_i$, depend on the Hamiltonian, the properties of the position and momentum operators may depend on the interaction. This is a particular feature of the noncanonical quantization, which is not of geometrical origin, but rather of a dynamical one. Because of this propery we often refer to the noncanonical quantization as a dynamical one.

In the canonical case the mapping (6) defines uniquely the quantum Hamiltonian $\hat{H}$, corresponding to (1), and the derivatives $\partial\hat{U}/\partial\hat{q}_i$ [this is the reason to write down the Hamiltonian in the form (1)]. This is also true for any noncanonical operators, if the classical potential can be represented in the form

$$U(q_1,\ldots,q_n) = \sum_{j=1}^{n}U(q_j). \quad (18)$$

For an arbitrary interaction, however, since $\hat{q}_1,\ldots,\hat{q}_n$ may not commute, one has to give a rule for an ordering of the operators when passing from $U(q_1,\ldots,q_n)$ to the quantum potential $\hat{U}$. This procedure, which is also not unique for arbitrary functions $F(\hat{p},\hat{q})$ of canonical variables,[12–14] has to be defined for every interaction. Here we will not go into a discussion of this important point. Instead, we shall consider another example of noncanonical quantization with a potential of the form (18), which exibits some new features and indicates that the ideas of Wigner in this respect deserve to be investigated further.

We consider a system of two nonrelativistic point particles, interacting via a harmonic potential. Assuming that the center of mass variables are quantized canonically and commute with the internal variables, we reduce the problem to a quantization of a three-dimensional harmonic oscillator for the internal degrees of freedom (Sec. IIA). Then (Sec. IIB) we quantize noncanonically a more general n-dimensional oscillator and study in more detail the two particle system (Sec. IIC). Section III is independent of the other part of the paper. It contains a motivation for the quantization of the oscillalator we consider, which is of the Lie superalgebraical origin. Finally, we investigate the behavior of the system in the classical limit $\hbar \to 0$ and give one possible interpretation of the results.

II. DYNAMICAL QUANTIZATION OF TWO POINT PARTICLES INTERACTING VIA A HARMONIC POTENTIAL

A. Reduction of the problem

Consider in the frame of nonrelativistic mechanics two point particles with masses m_1 and m_2 and a Hamiltonian

$$H_{\text{tot}} = \frac{\mathbf{p}_1^2}{2m_1} + \frac{\mathbf{p}_2^2}{2m_2} + \frac{m\omega^2}{2}(\mathbf{r}_1 - \mathbf{r}_2)^2. \tag{19}$$

Introduce the center of mass (CM) coordinates

$$\mathbf{R} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{m_1 + m_2}, \quad \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2 \tag{20}$$

and let μ and m be, respectively, the total and the reduced masses, $\mathbf{P}$ and $\mathbf{p}$ be the total mometum and the internal (the conjugate to $\mathbf{r}$) momentum, respectively; $r = |\mathbf{r}_1 - \mathbf{r}_2|$. Then the energy is a sum of the CM energy H_{CM} and the internal energy H,

$$H_{\text{tot}} = H_{CM} + H, \tag{21}$$

where

$$H_{CM} = \frac{\mathbf{P}^2}{2\mu}, \quad H = \frac{\mathbf{p}^2}{2m} + \frac{m\omega^2 r^2}{2}. \tag{22}$$

Similarly, the angular momentum

$$\mathbf{M}_{\text{tot}} = \mathbf{M}_{CM} + \mathbf{M}, \tag{23}$$

with

$$\mathbf{M}_{CM} = \mathbf{R}\times\mathbf{P}, \quad \mathbf{M} = \mathbf{r}\times\mathbf{p}. \tag{24}$$

According to the definition we have accepted, to quantize the system we have first to find simultaneous solutions of the Hamiltonian equations, replacing in them the classical variables $\mathbf{R}$, $\mathbf{P}$, $\mathbf{r}$, $\mathbf{p}$ by operators, i.e.,

$$\dot{\hat{\mathbf{P}}} = 0, \quad \dot{\hat{\mathbf{R}}} = \hat{\mathbf{P}}/\mu, \tag{25}$$

$$\dot{\hat{\mathbf{p}}} = -m\omega^2\hat{\mathbf{r}}, \quad \dot{\hat{\mathbf{r}}} = \hat{\mathbf{p}}/m, \tag{26}$$

and of the Heisenberg equations

$$\dot{\hat{\mathbf{P}}} = -\frac{i}{\hbar}[\hat{\mathbf{P}},\hat{H}_{\text{tot}}], \quad \dot{\hat{\mathbf{R}}} = -\frac{i}{\hbar}[\hat{\mathbf{R}},\hat{H}_{\text{tot}}], \tag{27}$$

$$\dot{\hat{\mathbf{p}}} = -\frac{i}{\hbar}[\hat{\mathbf{p}},\hat{H}_{\text{tot}}], \quad \dot{\hat{\mathbf{r}}} = -\frac{i}{\hbar}[\hat{\mathbf{r}},\hat{H}_{\text{tot}}]. \tag{28}$$

The operators $\hat{\mathbf{R}}$, $\hat{\mathbf{P}}$, $\hat{\mathbf{r}}$, $\hat{\mathbf{p}}$ should be determined to give a solution of the above Eqs. (25)–(28). By $\hat{H}_{\text{tot}}$ we denote the operator, obtained from the classical Hamiltonian after the replacement

$$(\mathbf{R}, \mathbf{P}, \mathbf{r}, \mathbf{p}) \to (\hat{\mathbf{R}}, \hat{\mathbf{P}}, \hat{\mathbf{r}}, \hat{\mathbf{p}}). \tag{29}$$

Independently of the dynamics, Eqs. (25)–(28) are satisfied with canonical operators. We wish to study some other, dynamically dependent solutions. Our purpose is not the investigation of all possible operators (29) that satisfy the Eqs. (25)–(28). Rather than that, we restrict ourselves only to one particular noncanonical solution for the internal variables $\hat{\mathbf{r}}$ and $\hat{\mathbf{p}}$ and study its properties. To this end we first assume that the CM observables can be measured simultaneously with the internal observables. Thus, we accept

Assumption 1: The CM variables commute with the internal variables, i.e.,

$$[\hat{\mathbf{R}},\hat{\mathbf{r}}] = [\hat{\mathbf{R}},\hat{\mathbf{p}}] = [\hat{\mathbf{P}},\hat{\mathbf{r}}] = [\hat{\mathbf{P}},\hat{\mathbf{p}}] = 0. \tag{30}$$

Under this assumption the quantization equations resolve into two independent groups. The first one, consisting of Eqs. (25) and (27), depends only upon the CM coordinate operator $\hat{\mathbf{R}}$ and the momentum operator $\hat{\mathbf{P}}$. Here we make

Assumption 2: The center of mass coordinates and momenta are quantized in a canonical way,

$$[\hat{\mathbf{R}}_j,\hat{\mathbf{P}}_k] = i\hbar\delta_{jk}, \quad [\hat{\mathbf{R}}_j,\hat{\mathbf{R}}_k] = [\hat{\mathbf{P}}_j,\hat{\mathbf{P}}_k] = 0. \tag{31}$$

Thus, we are left with the equations

$$\dot{\hat{\mathbf{p}}} = -m\omega^2\hat{\mathbf{r}}, \quad \dot{\hat{\mathbf{r}}} = \hat{\mathbf{p}}/m, \tag{32}$$

$$\dot{\hat{\mathbf{p}}} = -\frac{i}{\hbar}[\hat{\mathbf{p}},\hat{H}], \quad \dot{\hat{\mathbf{r}}} = -\frac{i}{\hbar}[\hat{\mathbf{r}},\hat{H}] \tag{33}$$

for the operators $\hat{\mathbf{r}}$ and $\hat{\mathbf{p}}$, which follow from Eqs. (26), (28), and (30).

Equations (32)–(33) coincide with the Hamiltonian and the Heisenberg equations of a three-dimensional harmonic oscillator. We now proceed to quantize it noncanonically. Since the generalization to the case of any dimension is straightforward, in the next section (IIB) we quantize an n-dimensional harmonic oscillator ($n > 1$) instead of a three-dimensional one.

B. Noncanonical quantization of an n-dimensional harmonic oscillator

Consider an n-dimensional harmonic oscillator with a Hamiltonian

$$H = \sum_{i=1}^{n}\left(\frac{1}{2m}p_i^2 + \frac{m\omega^2}{2}r_i^2\right). \tag{34}$$

To quantize it we have to replace as a first step the classical phase-space variables $(r_1,\ldots,r_n,p_1,\ldots,p_n)$ with operators that have to satisfy the operator Hamiltonian equations ($i = 1,2,\ldots,n$)

$$\dot{\hat{p}}_i = -m\omega^2\hat{r}_i, \quad \dot{\hat{r}}_i = \frac{\hat{p}_i}{m}, \tag{35}$$

and simultaneously the Heisenberg equations

$$\dot{\hat{p}}_i = -\frac{i}{\hbar}[\hat{p}_i,\hat{H}], \quad \dot{\hat{r}}_i = -\frac{i}{\hbar}[\hat{r}_i,\hat{H}]. \tag{36}$$

These equations are compatible only if

$$[\hat{H},\hat{p}_k] = i\hbar m\omega^2\hat{r}_k, \quad [\hat{H},\hat{r}_k] = -\frac{i\hbar}{m}\hat{p}_k. \tag{37}$$

Introduce in place of $\hat{r}_i,\hat{p}_i, i = 1,\ldots,n$ new operators ($\xi = \pm$)

$$a_k^\xi = \left(\frac{(n-1)m\omega}{4\hbar}\right)^{1/2}\hat{r}_k + i\xi\left(\frac{n-1}{4m\omega\hbar}\right)^{1/2}\hat{p}_k, \tag{38}$$

which will be referred to as creation ($\xi = +$) and annihilation ($\xi = -$) operators (CAO's). In terms of these operators the Hamiltonian (34) and the compatibility conditions (37) read

$$\hat{H} = \frac{\omega\hbar}{n-1}\sum_{i=1}^{n}\{a_i^+,a_i^-\}, \tag{39}$$

$$\sum_{i=1}^{n}[\{a_i^+,a_i^-\},a_k^\xi] = -\xi(n-1)a_k^\xi. \tag{40}$$

As a solution of Eq. (40) we choose operators $a_1^\pm,...,a_n^\pm$ satisfying the relations

$$[\{a_i^+,a_j^-\},a_k^+] = \delta_{jk}a_i^+ - \delta_{ij}a_k^+,$$
$$[\{a_i^+,a_j^-\},a_k^-] = -\delta_{ik}a_j^- + \delta_{ij}a_k^-,$$
$$\{a_i^+,a_j^+\} = \{a_i^-,a_j^-\} = 0. \tag{41}$$

We recall that all considerations are in the Heinsenberg picture. The position and the momentum operators depend on time and they also have to satisfy the Hamiltonian Eqs. (35), which read in terms of the CAO's

$$\dot{a}_k^\xi(t) = -i\xi\omega a_k^\xi(t). \tag{42}$$

Hence,

$$a_k^\xi(t) = \exp(-i\xi\omega t)a_k^\xi(0), \tag{43}$$

and, therefore, if the defining relations (41) for the CAO's hold at a certain time $t = 0$, i.e., for $a_k^\xi = a_k^\xi(0)$, then they hold as equal time relations for any other time t. One can easily check that the Heinsenberg Eqs. (36) (written in terms of the CAO's),

$$\dot{a}_k^\xi(t) = \frac{i\omega}{n-1}[\{a_j^+(t),a_j^-(t)\},a_k^\xi(t)], \tag{44}$$

agree at any time with the Hamiltonian Eqs. (42).

It remains to define the position and the momentum operators $\hat{r}_k$ and $\hat{p}_k$, correpsonding to the CAO's (41), as a linear Hermitian operators in a Hilbert space, which will be the space of the states of the oscillator. In terms of the CAO's this means that the Hermitian conjugate to a_k^+ should be equal to a_k^-, i.e.,

$$(a_k^+)^\dagger = a_k^-. \tag{45}$$

One can find several spaces where the operators (41) are linear and satisfy (45). Since *a priori* there exists no reason to exclude any of the possible spaces, one has to determine and study all of them and subsequently rule out those that are not appropriate for physical applications, in other words, one has to determine those representations of the CAO's (41) for which the condition (45) also holds. Here we shall consider only representations which are obtained by the usual Fock space technique. These Fock representations are labelled by one non-negative integer $p = 0,1,...$. To construct them assume (as in case of the para-Bose statistics) that the corresponding space $W(n;p)$ contains a single vector (up to a multiple) $|0\rangle$, called a vacuum, such that

$$a_i^-|0\rangle = 0 \text{ and } a_i^-a_j^+|0\rangle = p\delta_{ij}|0\rangle, \quad i,j = 1,2,...,n. \tag{46}$$

Since $(a_i^+)^2 = 0$, from (46) one derives that the vectors

$$|p;\theta_1,...,\theta_n\rangle = (p!)^{-1/2}\left(\left(p - \sum_{i=1}^n \theta_i\right)!\right)^{1/2}(a_1^+)^{\theta_1}\cdots(a_n^+)^{\theta_n}|0\rangle, \tag{47}$$

with $\theta_i = 0,1$ and $\sum_{i=1}^n \theta_i \leqslant p$, constitute an orthonormal basis in $W(n;p)$ with respect to a scalar product, defined usually with "bra" and "ket" vectors and $\langle 0|0\rangle = 1$.

The CAO's transform the basis vectors according to

$$a_k^-|\cdots,\theta_k,\cdots\rangle$$
$$= \theta_k(-1)^{\theta_1+\cdots+\theta_{k-1}}\left(p - \sum_i\theta_i + 1\right)^{1/2}|\cdots,\theta_k - 1,\cdots\rangle,$$
$$a_k^+|\cdots,\theta_k,\cdots\rangle \tag{48}$$
$$= (1-\theta_k)(-1)^{\theta_1+\cdots+\theta_{k-1}}\left(p - \sum_i\theta_i\right)^{1/2}|\cdots,\theta_k + 1,\cdots\rangle.$$

One can check that within any Fock space the relation (45) holds, so that $\hat{r}_i$ and $\hat{p}_i$ are Hermitian operators. The Hamiltonian (39) is diagonal in the basis (47). To determine its spectrum, call the vector $|p;\theta_1,...,\theta_n\rangle \in W(n,p)$ an m state and denote it by $|p;m)$ if $\Sigma_{i=1}^n\theta_i = m$. Then from (39) and (48) one obtains

$$\hat{H}|p;m) = E_m|p;m), \tag{49}$$

where

$$E_m = \frac{\omega\hbar}{n-1}(np - nm + m). \tag{50}$$

Since m can run only through the values $0,1,...,\min(n,p)$, the energy of the n-dimensional oscillator for a statistics of order p has $\min(n+1,p+1)$ different values. The dimension of the subspace $W_m(n;p)$ of all m states is

$$\dim W_m(n;p) = \binom{n}{m}, \tag{51}$$

so that the different (linearly independent) states with energy E_m are $\binom{n}{m}$. In particular, the state $|p,0)$ with the highest energy is nondegenerate. A given ground state $|p,\min(n,p))$ is nondegenerate only if $p \geqslant n$.

C. Quantization of the two-particle system

Here we apply the results of the previous section to quantize the internal motion of the two-particle system. In this case $n = 3$ and in terms of the CAO's

$$a_k^\xi = (2\hbar)^{-1/2}(m\omega)^{1/2}\hat{r}_k + i\xi(2m\omega\hbar)^{-1/2}\hat{p}_k, \tag{52}$$

the internal Hamiltonian reads

$$\hat{H} = \frac{1}{2m}\hat{\mathbf{p}}^2 + \frac{m\omega^2}{2}\hat{\mathbf{r}}^2 = \frac{\omega\hbar}{2}\sum_{i=1}^3\{a_i^+,a_i^-\}. \tag{53}$$

For the operators of the squared distance between the particles, $\hat{\mathbf{r}}^2 = \hat{r}_1^2 + \hat{r}_2^2 + \hat{r}_3^2$, and the squared internal momentum, $\hat{\mathbf{p}}^2 = \hat{p}_1^2 + \hat{p}_2^2 + \hat{p}_3^2$, one obtains

$$\hat{\mathbf{r}}^2 = \frac{\hbar}{2m\omega}\sum_{i=1}^3\{a_i^+,a_i^-\}, \tag{54}$$

$$\hat{\mathbf{p}}^2 = \frac{m\omega\hbar}{2}\sum_{i=1}^3\{a_i^+,a_i^-\}. \tag{55}$$

Inserting in the classical expression (24) for the internal angular momentum $\hat{\mathbf{M}}$ the operators $\hat{r}_k$ and $\hat{p}_k$ (in terms of the CAO's), one obtains

$$\hat{M}_k = \frac{i\hbar}{2}\sum_{l,m}\epsilon_{klm}\{a_l^+,a_m^-\}. \tag{56}$$

These operators satisfy the commutation relations for the generators of the rotation group

$$[\hat{M}_j,\hat{M}_k] = -\tfrac{1}{2}i\hbar\epsilon_{jkl}\hat{M}_l. \tag{57}$$

We remark, however, that the angular momentum is measured in units of $\hbar/2$.

As it should be, the position and the momentum operators transform as vectors under rotations,

$$[\hat{M}_j,\hat{r}_k] = -\tfrac{1}{2}i\hbar\epsilon_{jkl}\hat{r}_l,$$
$$[\hat{M}_j,\hat{p}_k] = -\tfrac{1}{2}i\hbar\epsilon_{jkl}\hat{p}_l. \tag{58}$$

It is a straightforward calculation to show that all operators

$$\hat{H},\hat{\mathbf{r}}^2,\hat{\mathbf{p}}^2,\hat{\mathbf{M}}^2,\hat{M}_3,$$

commute with each other and, therefore, can be measured simultaneously. If $|p;k)$ is a k-state in a representation for a statistic of order p, then

$$\hat{H}\,|\,p;k) = \tfrac{1}{2}\omega\hbar(3p-2k)|\,p;k), \tag{59}$$
$$\hat{\mathbf{r}}^2|p;k) = (\hbar/2m\omega)(3\,p-2k)|\,p;k), \tag{60}$$
$$\hat{\mathbf{p}}^2|p;k) = (m\omega\hbar/2)(3\,p-2k)|\,p;k), \tag{61}$$
$$\hat{\mathbf{M}}^2|\,p;k) = 0 \quad \text{for } k=0,3,$$
$$= \tfrac{1}{2}\hbar^2|p;k) \quad \text{for } k=1,2. \tag{62}$$

There is only one state, the state $|p;0,0,0\rangle$, corresponding to the maximum distance between the constituents and to the maximum of the internal energy,

$$r_{\max} = \left(\frac{3\hbar p}{2m\omega}\right)^{1/2}, \quad E_{\max} = \tfrac{3}{2}\,\omega\hbar p. \tag{63}$$

This state carries momentum zero. If $p\geqslant 3$, then $|\,p;1,1,1\rangle$ is the ground state; it is nondegenerate, with zero momentum, and corresponds to the minimal distance and energy

$$r_{\min} = \left(\frac{3\hbar(p-2)}{2m\omega}\right)^{1/2}, \quad E_{\min} = \tfrac{3}{2}\omega\hbar(p-2). \tag{64}$$

If, however, $p=1$ or 2 then the ground state is degenerate; there are three different states with the same energy and orbital momentum 1 (in units $\hbar/2$). In this case

$$r_{\min} = \left(\frac{\hbar p}{2m\omega}\right)^{1/2}, \quad E_{\min} = \frac{\omega\hbar}{2}p, \quad p=1,2. \tag{65}$$

We see that after the quantization the two particles are bound to each other; they are movir ; together with their center of mass in such a way that the distance between them is fixed. The position, however, of any one of the constituents cannot be localized in the space. The latter follows from the observation that the operators $\hat{r}_k$ of the internal coordinates do not commute with each other,

$$[\hat{r}_i,\hat{r}_j]\neq 0, \quad i\neq j = 1,2,3,$$

and, therefore, they cannot be diagonalized simultaneously. Thus, trying to visualize the picture, one can say that the two particles are moving as the ends of a massless ridged stick, whose length depends on the internal energy and can take no more than four different values. The stick itself is rotating around the center of mass of the system; however, its orientation in the space cannot be localized.

III. QUANTIZATION, STATISTICS, AND LIE SUPERALGEBRAS

The results, obtained in the previous section, were essentially based on the properties of the creation and annihilation operators (41). The latter appear as one possible way to satisfy the compatibility equations (40). The reason for selection of the CAO's (41) as a solution of Eqs. (40) is of a Lie superalgebraical origin; these operators generalize in a natural way some known Lie superalgebracial properties of the Bose and the para-Bose operators, which will be reviewed now shortly.

To this end we recall that the set of operators A is a Lie superalgebra (LS) with a product $[\![,]\!]$ if[15]

(a) the set A is a linear space (with respect to the usual sum between operators and multiplication by numbers), which is a direct sum of its subspaces A_0 and A_1, $A=A_0+A_1$. The elements $a_\alpha\in A_\alpha$ are called homogeneous even ($\alpha=0$) and odd ($\alpha=1$) elements, respectively.

(b) for any two homogeneous elements $a_\alpha\in A_\alpha$ and $b_\beta\in A_\beta$ the product is defined as

$$[\![a_\alpha,b_\beta]\!] = a_\alpha b_\beta - (-1)^{\alpha\beta}b_\beta a_\alpha, \tag{66}$$

and is extended by linearity to arbitrary elements from A.

(c) if $\alpha+\beta=\gamma(\text{mod } 2)$, then

$$[\![a_\alpha,b_\beta]\!]\in A_\gamma. \tag{67}$$

The algebra is simple if it has no nontrivial ideals. A representation of the LS A is a linear map θ of A onto another LS $\bar{A}$, which preserves the product $[\![,]\!]$.

Consider now n pairs $a_1^\pm,\ldots,a_n^\pm$ of para-Bose operators (16) and let (sum over repeated indices; $i,j=1,\ldots,n$; $\xi,\eta=\pm$; $\mathbb{C}$: complex numbers)

$$A_0 = \{\alpha_{ij}^{\xi\eta}\{a_i^\xi,a_j^\eta\}\,|\,\alpha_{ij}^{\xi\eta}\in\mathbb{C}\}, \tag{68}$$
$$A_1 = \{\alpha_i^\xi a_i^\xi\,|\,\alpha_i^\xi\in\mathbb{C}\}, \tag{69}$$
$$A = \{\alpha_i^\xi a_i^\xi + \alpha_{ij}^{\xi\eta}\{a_i^\xi,a_j^\eta\}\,|\,\alpha_i^\xi,\,\alpha_{ij}^{\xi\eta}\in\mathbb{C}\}. \tag{70}$$

We now show that the set (70) of linear operators is a Lie superalgebra. Clearly A, A_0, and A_1 are linear spaces and $A=A_0+A_1$. Consider two arbitrary odd elements

$$a_1 = \alpha_i^\xi a_i^\xi\in A_1 \quad\text{and}\quad b_1 = \beta_j^\eta a_j^\eta\in A_1, \tag{71}$$

and two arbitrary even elements

$$a_0 = \alpha_{jk}^{\eta\epsilon}\{a_j^\eta,a_k^\xi\}\in A_0, \quad b_0 = \beta_{il}^{\xi\delta}\{a_i^\xi,a_l^\delta\}\in A_0. \tag{72}$$

From the definition of the product (66) one has

$$[\![a_1,b_1]\!] = \{a_1,b_1\} = \alpha_i^\xi\beta_j^\eta\{a_i^\xi,a_j^\eta\}\in A_0.$$

The relation (16) yields

$$[\![a_0,a_1]\!] = [a_0,a_1] = \alpha_{jk}^{\eta\epsilon}\alpha_i^\xi[\{a_j^\eta,a_k^\xi\},a_i^\xi]$$
$$= \alpha_{jk}^{\eta\epsilon}\alpha_i^\xi((\xi-\eta)\delta_{ij}a_k^\epsilon + (\xi-\epsilon)\delta_{ik}a_j^\eta)\in A_1.$$

Finally, using the equality

$$[\{a_i^\xi,a_j^\eta\},\{a_k^\epsilon,a_l^\delta\}]$$
$$= (\epsilon-\xi)\delta_{ik}\{a_j^\eta,a_l^\delta\} + (\epsilon-\eta)\delta_{jk}\{a_i^\xi,a_l^\delta\}$$
$$+ (\delta-\xi)\delta_{il}\{a_j^\eta,a_k^\epsilon\} + (\delta-\eta)\delta_{jl}\{a_i^\xi,a_k^\epsilon\}, \tag{73}$$

which is a consequence of (16), one easily shows that

$$[\![a_0,b_0]\!] = [a_0,b_0]\in A_0.$$

Hence, A is a lie superalgebra.[16] A more detailed investigation shows that this LS is isomorphic to the simple orthosymplectic LS osp $(1,2n)$.[17] Since, moreover, the elements of A—see (70)—are polynomials of the para-Bose operators, to every (irreducible) respresentation of $a_1^\pm,\ldots,a_n^\pm$ there corresponds an (irreducible) representation of osp $(1,2n)$ and vice versa.

The representation of the CAO's (12), corresponding to an order of the statistic $p = 1$ in (17), is of particular importance for the quantum physics, since in this case the para-Bose operators reduce to Bose creation and annihilation operators,

$$[a_i^\xi, a_j^\eta] = \tfrac{1}{2}(\eta - \xi)\delta_{ij}.$$

Inserting these operators into (70) one obtains an infinite-dimensional irreducible respresentation of the LS osp(1,2n). The canonical quantum mechanics is essentially based on this particular representation. On the other hand, the noncanonical quantization ($p \neq 1$) of Hamiltonian (13) with para-Bose operators (and in particular the Wigner quantization of the one-dimensional oscillator) is a quantization according to some other representations of the same orthosymplectic Lie superalgebra. Having observed that, one may wonder why among the several available Lie superalgebras the orthosymplectic one plays such a distinguished role in quantum physics. One may also ask whether it is not possible to quantize with position and momentum operators that lead to representations of other LS's, and in particular [since osp (1,2n) is simple] other simple LS's. The example we have considered gives a positive answer to this question. The operators (41) were chosen in such a way that when inserted in (70) instead of the para-Bose operators, they give a simple Lie superalgebra, which is isomorphic to the special linear LS sl (1,n).[18]

Since $\{a_i^\xi, a_j^\xi\} = 0, \xi = \pm$, the even part (68) of A is ($i,j = 1,...,n$)

$$A_0 = \{\alpha_{ij}\{a_i^+, a_j^-\} | \alpha_{ij} \in \mathbb{C}\}. \tag{74}$$

If

$$e_{ij} = \{a_i^+, a_j^-\}, \tag{75}$$

then one obtains from (41)

$$[e_{ij}, e_{kl}] = \delta_{jk} e_{il} - \delta_{il} e_{kj}, \tag{76}$$

which are the commutation relations for the generators of the Lie algebra gl(n); $A_0 = $ gl(n). The CAO's $a_1^\pm,...,a_n^\pm$ are the odd generators; they define the even generators (75) and, hence, the whole algebra uniquely. Therefore, also in this case to every irreducible representation of the creation and the annihilation operators (41) there corresponds an irreducible representation of sl (1,n) and vice versa. Thus, the quantization of the n-dimensional oscillator, considered in Sec. IIB, is according to a set of finite-dimensional irreducible representations of the LS sl(1,n). The nonequivalent finite-dimensional irreducible representations of this LS are labelled by $n + 1$ numbers $(\alpha_0, \alpha_1, ..., \alpha_n)$, where α_0 is an arbitrary complex number and $\alpha_1,...,\alpha_n$ are arbitrary non-negative integers.[19] The representations (48) of the CAO's or, equivalently, of the position and the momentum operators $\hat{q}_1,...,\hat{q}_n,\hat{p}_1,...,\hat{p}_n$, are labelled only with one integer p and, therefore, described a small part of all possible representations.

From (43) one concludes that the generators (75) of gl (n) and, hence, the even part of sl(1,n), is preserved in time. The Hamiltonian (39) is an element from the center of gl(n) and, therefore, commutes with the even subalgebra (this is another way to conclude that A_0 is preserved in time).

Up to now we have not specified whether we consider sl(1,n) as a real or a complex LS. One of the real forms of A_0 is given with the linear envelope of all operators $\{\hat{q}_i, \hat{q}_j\}$, $\{\hat{p}_i, \hat{p}_j\}$, and $\{\hat{q}_i, \hat{p}_j\}$. The latter is isomorphic to the algebra of the unitary group U (n) and commutes with $\hat{H}$. Therefore, as in the canonical case we obtain that U (n) is a symmetry group of the oscillator. At the same time the odd generators $a_i^\pm$ are shifting the energy so that, starting with a given energy state, one can obtain a state with any other energy from the spectrum of $\hat{H}$. Hence, the LS sl(1,n) appears as a spectrum generating algebra of the n-dimensional oscillator. In the case of the two-particle system the internal symmetry group is U (3) and since every irreducible U (n) representation is also SU (n) irreducible, the internal symmetry group of the composite system is SU (3).

IV. A POSSIBLE INTERPRETATION AND FURTHER GENERALIZATIONS

One of the interesting features of the example, considered in Sec. IIC, is that after quantization the initial two particles are bound to each other, the distance between them is bounded from above. The position of any one of the initial constituent particles cannot be localized in the space and, therefore, the particles are smeared with a certain probability within a finite volume. The composite system exhibits an internal structure, it has eight different states, characterized completely by the internal energy, the orbital momentum, and its third projection. If $k = \theta_1 + \theta_2 + \theta_3$ then, measuring the energy E in units $\omega\hbar/2$ and the orbital momentum in units $\hbar/2$, one has [$k \leqslant \min(p,3)$]

k	number of the states	E	M	M_3
0	1	$3p$	0	0
1	3	$3p-2$	1	$0, \pm 1$
2	3	$3p-4$	1	$0, \pm 1$
3	1	$3p-6$	0	0.

Using a particle terminology, i.e., intepreting the internal energy as a mass (the picture is, however, nonrelativistic!) and the orbital momentum as a spin of the composite system, we may say that the quantized system behaves as a multiplet of two spin zero particles and two spin one particles, all of them with different masses.

The noncanonical and the canonical quantizations of the two particles lead to essentially different pictures. In the noncanonical case the internal state space is finite-dimensional; as a result (contrary to the canonical one) the spectrum of the internal energy is finite and the composite system occupies a finite space volume. The difference becomes even more evident in the classical limit $\hbar \to 0$. Since ($k = 1,2,3$)

$$\hat{r}_k = (\hbar/2m\omega)^{1/2}(a_k^- + a_k^+),$$
$$\hat{p}_k = i(m\omega\hbar/2)^{1/2}(a_k^- - a_k^+), \tag{77}$$

in the limit $\hbar \to 0$ the internal position and momentum operators tend to zero. As a consequence—see also (62)–(63)—the internal energy, the distance between the particles, and the relative momentum converge to zero. Hence,

$$\lim E_{\max} = \lim r_{\max} = 0 \quad \text{when } \hbar \to 0. \tag{78}$$

The radius vectors of both particles coincide in the limit with the radius vector of the center of mass, i.e., if $\hbar \to 0$, then

$$\lim \hat{\mathbf{r}}_1 = \lim \left(\widehat{\mathbf{R}} + \frac{m_2}{m_1 + m_2} \hat{\mathbf{r}} \right) = \widehat{\mathbf{R}},$$

$$\lim \hat{\mathbf{r}}_2 = \lim \left(\widehat{\mathbf{R}} - \frac{m_1}{m_1 + m_2} \hat{\mathbf{r}} \right) = \widehat{\mathbf{R}}. \tag{79}$$

Thus, in the classical limit the composite system collapses into a point. Since the equation of motion of this particle is (25), it moves as a free classical point particle with a mass $m_1 + m_2$. We remark, however, that the internal variables are zero operators in an (no more than) eight-dimensional space, i.e., every point of the space preserves in the limit $\hbar \to 0$ its internal structures.

The above results hold only for the Fock representations (48). In he general case (we omit the proof, which will be given elsewhere) the representation space L of the position and the momentum operators $\hat{q}_1, \hat{q}_2, \hat{q}_3, \hat{p}_1, \hat{p}_2, \hat{p}_3$ is a direct sum of (at most) four SU (3) irreducible subspaces,

$$L = L_1 + L_2 + L_3 + L_4.$$

Within every subspace L_i the operators $\widehat{H}$, $\hat{\mathbf{r}}$, and $\hat{\mathbf{p}}^2$ are proportional to the unity and, hence, the energy, the distance between the initial constituents, the internal momentum have as before (no more than) four different values, and the composite system is smeared in a finite volume of space. The orbital momentum of the system can take, however, arbitrarily large integer values and the SU (3)-irreducible subspace L_i may contain states with different orbital momenta.

The present investigation makes no pretentions to being a generalization of the canonical quantization for the case of an arbitrary interaction. For certain potentials it could be a difficult problem to find simultaneous solutions of the quantum Eqs. (3) and the classical Eqs. (5). To develop an approach, which can be applied for any interaction, one has to define in addition a general rule for ordering of the position operators, when replacing them in the classical potential $\widehat{U}(q_1,...,q_n)$ and to give a precise meaning to the derivates $\partial \widehat{U} / \partial \hat{q}_k$. In order to interpret the states of the composite system as a multiplet of particles one has, as a next step, to develop a relativistic analog of the present approach. In this paper we have considered only an example of a noncanonical quantization, which exhibits some new features and shows to our mind that the Wigner ideas for generalization of the ordinary quantization deserve further investigation.

ACKNOWLEDGMENTS

The author is grateful to Professor H. D. Doebner for the warm hospitality at the Institute of Theoretical Physics A in Clausthal-Zellerfeld and for many valuable discussions during the course of this work. I wish to thank Professor S. Ali and Professor K. H. Mariwalla for the critical reading of the manuscript.

[1]E. P. Wigner, Phys. Rev. 77, 711 (1950).
[2]Throughout the paper the caret denotes an operator. Functions of operators, e.g., $H(\hat{q},\hat{p})$, will be denoted by $\widehat{H}$.
[3]The relations (4) hold for a large class of potentials, namely such that $U(q_1,...,q_n)$ can be expanded in a Laurent series around zero for every q_i.
[4]P. Ehrenfest, Z. Phys. 45, 455 (1927).
[5]Note that by classical or Hamiltonian equations we always understand the operator Eqs. (5).
[6]H. S. Green, Phys. Rev. 90, 270 (1953).
[7]O. W. Greenberg and A. M. L. Messiach, Phys. Rev. 138B, 1155 (1965).
[8]L. M. Yang, Phys. Rev. 84, 788 (1951).
[9]Y. Ohnuki and S. Kamefuchi, J. Math. Phys. 19, 67 (1978).
[10]J. K. Sharma, C. L. Mehta, and E. C. G. Sudarshan, J. Math. Phys. 19, 2089 (1978).
[11]N. Mukunda, E. C. G. Sudarshan, J. K. Sharma, and C. L. Mehta, J. Math. Phys. 21, 2386 (1980).
[12]J. R. Shewell, Am. J. Phys. 27, 16 (1959).
[13]C. L. Mehta, J. Math. Phys. 5, 677 (1964).
[14]G. S. Agarwal and E. Wolf, Phys. Rev. D 2, 2164 (1970).
[15]This definition agrees with the abstract one. For the latter see, for instance, L. Corwin, J. Ne'eman, and S. Sternberg, Rev. Mod. Phys. 47, 573 (1975).
[16]M. Omote, Y. Ohnuki, and S. Kamefuchi, Prog. Theor. Phys. 56, 1948 (1976).
[17]A. Ganchev and T. D. Palev, J. Math. Phys. 21, 797 (1980).
[18]T. D. Palev, J. Math. Phys. 21, 1293 (1980).
[19]V. G. Kac, *Representations of classical Lie superalgebras*, Lecture Notes in Mathematics, Vol. 676 (Springer, New York, 1978), p. 597.

Quantum and classical mechanics on homogeneous Riemannian manifolds

Gérard G. Emch
Departments of Mathematics and of Physics, The University of Rochester, Rochester, New York 14627

(Received 15 January 1982; accepted for publication 12 March 1982)

Starting from the axioms of quantum mechanics as formalized by the systems of imprimitivity for homogeneous Riemannian manifolds, the classical theory is derived as a consequence, complete with: its phase space realized as the space of pure classical states; a generalized version of the Wigner–Moyal correspondence rule; the Jordan and Lie algebra structures of functions on the cotangent bundle, given by point-wise multiplication and Poisson bracket; and the momentum map. A comparison is also given of the quantum and classical dynamics and equilibrium statistical mechanics of free particles on compact manifolds of constant negative curvature.

PACS numbers: 03.65. – w, 02.40.Ky

I. INTRODUCTION

1. The purpose of this paper is to propose a solution to the Dirac problem when the configuration space manifold M is not flat. Briefly stated, the problem is to build up the classical and quantum theories from an identification of the fundamental observables, in such a manner that a correspondence principle be established giving a phenomenological meaning to the formal analogies between the Jordan and Lie structures of quantum and classical mechanics. In quantum theory both of these structures originate in the noncommutative operator product AB: the Jordan structure is obtained by forming the symmetric product $A\circ B = (AB + BA)/2$ (or $[(A+B)^2 - A^2 - B^2]/2$), whereas the Lie structure is defined by the quantum commutator $\{A,B\}_\hbar = [A,B]/i\hbar$ [i.e., $(AB - BA)/i\hbar$]. In classical theory the Jordan product is the point-wise multiplication of functions on the cotangent bundle T^*M, whereas the Lie structure, defined by the canonical symplectic form ω ($= \Sigma_{k=1}^{d} dp_k \wedge dq_k$ in local coordinates) on T^*M, is given by the Poisson bracket

$$\{f,g\} = \sum_{k=1}^{d} \partial_{q_k} f\cdot\partial_{p_k} g - \partial_{q_k} g\cdot\partial_{p_k} f.$$

Some serious objections to a straightforward mathematization of the Dirac problem, and the associated putative correspondence principle, have been raised[1] and formalized as no-go theorems. To illustrate the essence of this type of argument in the simplest possible case, we sketch the proof of the following result.

2. *Scholium*: For $M = R^d$, and $d = 1$, no linear map can exist between the classical observables ($f,g,\ldots$) and the quantum observables ($F,G,\ldots$) and sastisfy the following conditions: (i) the classical Poisson bracket $\{f,g\}$ corresponds to the quantum commutator $[F,G]/i\hbar$; (ii) the identity function 1 corresponds to the identity operator I; (iii) the operators P (corresponding to the classical momentum p) and Q (corresponding to the classical position coordinate q) together act irreducibly on some Hilbert space $\mathfrak{H}$.

Proof: Suppose that such a putative correspondence principle exists. From $\{q,p\} = 1$, we have $[Q,P]/i\hbar = I$, and by recursion $[Q^n,P^m]/i\hbar = m\Sigma_{k=1}^{n} Q^{n-k}P^{m-1}Q^{k-1} = n\Sigma_{j=1}^{m} P^{m-j}Q^{n-1}P^{j-1}$, a basic formula already noticed in Ref. 2. This relation, together with the putative correspondence principle and the irreducibility of the representation, would make Q^n (respectively, P^m and $[Q^n,P^m]/i\hbar nm$) the quantum observables corresponding to the classical observables q^n (respectively, p^m and $q^{n-1}p^{m-1}$). Since $\{q^3,p^3\} - 3\{q^2p,qp^2\} = 0$, this correspondence principle would give $[Q^3,P^3] - 3[PQ^2 + Q^2P, P^2Q + QP^2]/4 = 0$ instead of the correct value $3i\hbar I$. ■

Remarks: This proof generalizes immediately to $M = R^d$ with $1 < d < \infty$; and the no-go theorem still remains true when one weakens its condition (iii) from irreducibility to finite multiplicity.[3] It, however, breaks down when infinite multiplicity is allowed, as shown by the existence of the "prequantization map"[4]; in that case, however, the recovery of the usual Schroedinger representation involves the machinery of the "geometric quantization programme."[5] The present paper shows than an alternate route can be travelled in the opposite direction, thus providing under rather general circumstances another, and perhaps more natural, solution to the Dirac problem.

3. An indication of what might have gone astray in the assumptions of the no-go theorems alluded to above, is provided (in case $M = R^d$) by the Wigner–Moyal correspondence principle, which attributes to the classical function

$$f(p,q) = \int\int da\, db\, \tilde f(a,b)\exp\{-i(a\cdot p + b\cdot q)\}$$

the quantum operator

$$F(P,Q) = \int\int da\, db\, \tilde f(a,b)\exp\{-i(a\cdot P + b\cdot Q)\}.$$

The quantum commutator $[F,G]/i\hbar$ then induces the pairing

$$\{\tilde f,\tilde g\}_\hbar : \zeta\in R^d\times R^d \rightarrow \int dz\, \tilde f(z)\tilde g(\zeta - z)\pi_\hbar(z,\zeta),$$

with

$$\pi_\hbar(z,\zeta) = [\chi_\hbar(z,\zeta) - \chi_\hbar(\zeta,z)]/i\hbar,$$
$$\chi_\hbar(z,\zeta) = \exp\{i\sigma(z,\zeta)\hbar/2\},$$
$$\sigma(z,\zeta) = a\cdot\beta - b\cdot\alpha,$$

where

$$z = (a,b),\ \zeta = (\alpha,\beta)\in R^d\times R^d.$$

Clearly, $\{\tilde f,\tilde g\}_\hbar$ differs from the Fourier transform of the usual Poisson bracket $\{f,g\}$, but converges to it as $\hbar\to 0$. The reasons originally given by Wigner[6] and Moyal[7] for their

 0022-2488/82/101785-07$02.50

correspondence principle were merely its formal simplicity and its convenience. Its phenomenological meaning in terms of the convergence of expectation values has recently been analyzed in Ref. 8. It is argued there that classical mechanics can be derived completely, with both its Jordan and Lie structures, from the more fundamental setting of quantum theory. This argument is generalized here, away from the flat situation $M = R^d$, while the latter naturally still appears as a particular case.

4. The classical geodesic flow on a compact surface of constant negative curvature is a typical model[9] of an Anosov flow which is a Kolmogorov flow and thus enjoys very strong ergodic properties. These spaces being obtained as quotient of the Poincaré half-plane by a discrete, nonabelian, cocompact subgroup Γ of SL(2,R), we first discuss the connection between classical and quantum mechanics on their universal covering space. In a wider context, we first consider models of space where M is a connected, simply connected, d-dim Riemann manifold M on which a symmetry group G acts transitively. M is therefore geodesically complete[10] and of constant curvature K. We restrict our analysis to the cases where $K<0$ and, for notational simplicity, $d=2$.

II. QUANTUM AND CLASSICAL MECHANICS ON LOBATCHEVSKI SPACE

1. The formulation of the quantum theory of a particle whose configuration space is a homogeneous Riemannian manifold M, with symmetry group G, is based on Mackey's notion of an irreducible system of imprimitivity[11] resulting from the action of G on M. We therefore have a Hilbert space $\mathfrak{H}$, a unitary representation $U:G\to\mathfrak{U}(\mathfrak{H})$, and a projection-valued measure $Q:\mathfrak{B}(M)\to\mathfrak{P}(\mathfrak{H})$ based on the σ-algebra $\mathfrak{B}(M)$ of the Borel subsets of M, satisfying

(i) $U(g)Q(\Delta)U(g^{-1}) = Q(g[\Delta]) \quad \forall(g,\Delta)\in G\times\mathfrak{B}(M)$,

(ii) $\mathfrak{L}(\mathfrak{H}) = \{U(g),Q(\Delta)\,|\,g\in G,\Delta\in\mathfrak{B}(M)\}''$.

As usual Q describes the "position" observable on M; and the generators of $U(G)$, defined for every one-parameter subgroup T of G by

$$U(a) = \exp\{-iPa/\hbar\},$$

are identified as the corresponding "momentum" observables. To assume that $\mathfrak{H} = \mathfrak{L}^2(M)$ amounts then to restricting one's attention to particles with no internal degree of freedom (e.g., spin). The aim of this section is to derive, from this framework alone, the corresponding classical theory, i.e., to control the limit $\hbar\to 0$, and to exploit systematically its consequences to establish the correspondence principle which will solve the Dirac problem.

We now specify the manifold M and the symmetry group H we want to consider, and we describe them in a parametrization which is convenient for our purposes (although the final results are ultimately independent of the coordinate system thus initially singled out).

2. *Scholium*: (i) For $0<c<\infty$, $M_c \equiv \{\xi = (x,y)\,|\,x,y\in R\}$ equipped with the metric

$$ds^2 = \exp(-y/c)\{dx^2 + (x/c)dx\,dy + [(x^2/4c^2) + \exp(y/c)]dy^2\}$$

is a Riemann surface of constant negative curvature $K_c = -c^{-2}$.

(ii) Every connected, simply connected, geodesically complete, two-dimensional, Riemannian manifold of constant negative curvature K_c is isometric to (M_c, ds^2).

(iii) For $0<\alpha<\infty$, $H_\alpha \equiv \{\xi = (s,t)\,|\,s,t\in R\}$ equipped with the composition law

$$\gamma_\alpha:(\xi_1,\xi_2)\in H_\alpha\times H_\alpha \to (s_1\exp(-t_2/2\alpha) + s_2\exp(+t_1/2\alpha), t_1+t_2)$$

is a noncommutative Lie group extension $1\to R\to H_\alpha \to R\to 1$, with Haar measure $d\mu_\alpha(s,t) = \exp(-t/2\alpha)ds\,dt$, and modular function $\Delta_\alpha(\xi^{-1}) = \exp(t/\alpha)$.

(iv) The map $\gamma_c:(\xi,\zeta)\in H_c\times M_c\to\gamma_c(\xi,\zeta)\in M_c$ defines a transitive, free, and isometric action of H_c on (M_c, ds^2).

Proof: To prove (i) and (ii), it is sufficient[12] to verify that (M_c, ds^2) is isometric to the one-sheet hyperboloid

$$M'_c = \{(x',y',t)\in R^3\,|\,x'^2 + y'^2 - c^2t^2 = -c^2; t>0\}$$

equipped with the Riemannian metric g obtained by restricting to M'_c the Minkowski metric on R^3 given by

$$(ds)^2 = (dx')^2 + (dy')^2 - c^2(dt)^2.$$

The natural symmetry group G of M'_c is the homogeneous Lorentz group of the linear transformation of R^3 leaving invariant the quadratic form $x'^2 + y'^2 - c^2t^2$. Upon eliminating t, M'_c is identified with $\{(x',y')\,|\,x',y'\in R\}$ with metric

$$(ds')^2 = (c^2 + x'^2 + y'^2)^{-1}\{(c^2 + y'^2)(dx')^2 - 2x'y'dx'dy' + (c^2 + x'^2)(dy')^2\}.$$

Upon identifying the elements $(x',y')\in M'_c$ with the 2×2 matrices of the form

$$X' \equiv \begin{pmatrix} -x' & y' + (c^2 + x'^2 + y'^2)^{1/2} \\ y' - (c^2 + x'^2 + y'^2)^{1/2} & x' \end{pmatrix},$$

G is identified with SL(2,R)/Z_2 by identifying the elements $A\in$ SL(2,R)/Z_2 with the Lorentz transformations $X'\to AX'A^{-1}$. The Lie subgroup H'_α of G, identified in this manner with

$$H'_\alpha = \left\{\begin{pmatrix} \exp(t/2\alpha) & s/\alpha \\ 0 & \exp(-t/2\alpha)\end{pmatrix}\,\middle|\, s,t\in R\right\},$$

is clearly isomorphic to the group H_α defined in (iii). Moreover, H'_c acts transitively, freely, and isometrically on M'_c; we can therefore use the parametrization of H'_c to define a new coordinate system on M'_c. A straightforward computation shows that the metric of M'_c, when expressed in these new coordinates, coincide with the metric of the Riemann manifold M_c defined in (i). The reminder of the scholium follows by immediate inspection. ■

3. *Remarks*: (a) The proof of the scholium was presented here in a manner which emphasizes the explicit connection with the Lobatchevski plane in its usual presentation: The parametrization chosen here happens to be more convenient for our purposes. (b) In particular, in the limit $\alpha\to\infty$ (respectively, $c\to\infty$), H_α (respectively, M_c) reduces to the translation group R^2 (respectively, to the flat Riemannian manifold R^2) with its usual Euclidean group structure and its invariant Haar measure $d\xi = ds\,dt$ (respectively, with its usual Euclid-

ean metric $ds^2 = dx^2 + dy^2$). (c) For c finite, however, H_c differs from R^2 in at least two mathematically important aspects: It is not abelian, and its one parameter subgroups cannot in general be identified with geodesics in M_c; in this latter respect, $A \equiv \{(0,t) | t \in R\} \subset H_c$ is an exception. One should notice also that A normalizes $N \equiv \{(s,0) | s \in R\}$, and that H_c can be written as $H_c = AN$. (d) The structures singled out in the scholium carry over to higher dimensions [e.g., for $d = 3$, where $G = \mathrm{SL}(2,C)$, see the seminal work of Mackey[13]], where one can take advantage of the Iwasawa decomposition theory, which is in fact what is going on already in the case $d = 2$ chosen here for its explicitness. (e) Since H_c (and not merely G) acts transitively on M_c, we have already that $\{U(\xi), Q(\Delta) | \xi \in H_c, \Delta \in \mathfrak{B}(M_c)\}$ acts irreducibly on $\mathfrak{H}_c = \mathfrak{L}^2(M_c, d\mu_c)$. Consequently, every observable of the quantum theory can be obtained as a function of the position observables *and* of the momenta relative to H_c alone. This is in particular the case for the angular momentum, thus generalizing to our curved configuration spaces an essential feature of the theory of a spinless particle. (f) For the passage to the classical limit, it is important to concentrate on the functional dependence on the momenta P rather than merely on the variables $P/\hbar$. The parametrization of H_c, introduced in the scholium, allows one to achieve this by introducing, on the one hand, the unitary operators $\{U_\hbar(\xi) = U(\xi_\hbar) | \xi \in H_c\}$, where $\xi_\hbar = (\hbar s, \hbar t)$ for $\xi = (s,t)$. We have thus on $\mathfrak{H}_c$

$$(U_\hbar(\xi)\psi)(\zeta) = \psi(\gamma_c(\xi_\hbar^{-1}, \zeta)).$$

Since

$$\gamma_c(\xi_\hbar, \eta_\hbar) = \gamma_{c/\hbar}(\xi, \eta)_\hbar,$$

$$U_\hbar : \xi \in H_{c/\hbar} \to U_\hbar(\xi) \in \mathfrak{U}(\mathfrak{H}_c)$$

is a unitary representation of $H_{c/\hbar}$. On the other hand, the functions $f(Q)$ of position are defined as usual by

$$(f(Q)\psi)(\zeta) = f(\zeta)\psi(\zeta).$$

This suggests the following realization of the algebra of quantum observables.

4. *Lemma*: Let $\mathfrak{A}_{c,\hbar}$ be the vector space of functions

$$f:(\xi,\zeta) \in H_{c/\hbar} \times M_c \to f(\xi,\zeta) \in C$$

which are continuous of compact support in ξ, and bounded $\mathscr{C}^\infty$ in ζ. Let further $f \in \mathfrak{A}_{c,\hbar} \to f^* \in \mathfrak{A}_{c,\hbar}$, and $(f,g) \in \mathfrak{A}_{c,\hbar} \times \mathfrak{A}_{c,\hbar} \to \gamma_{c,\hbar}(f,g) \in \mathfrak{A}_{c,\hbar}$ be defined by

$$f^*:(\xi,\zeta) \in H_{c/\hbar} \times M_c \to f(\xi^{-1}, \gamma_c(\xi_\hbar^{-1}, \zeta))^* \Delta_{c/\hbar}(\xi^{-1}),$$

$$\gamma_{c,\hbar}(f,g):(\xi,\zeta) \in H_{c/\hbar} \times M_c \to \int d\mu_{c/\hbar}(\eta) f(\eta,\zeta) g(\gamma_{c/\hbar}(\eta^{-1},\xi), \gamma_c(\eta_\hbar^{-1},\zeta)).$$

Then the map $f \in \mathfrak{A}_{c,\hbar} \to F \in \mathfrak{L}(\mathfrak{H}_c)$ defined by

$$(F\psi)(\zeta) = \int d\mu_{c/\hbar}(\xi) f(\xi,\zeta) \psi(\gamma_c(\xi_\hbar^{-1},\zeta))$$

is a *-algebra isomorphism from $\mathfrak{A}_{c,\hbar}$ onto a dense *-subalgebra of $\mathfrak{L}(\mathfrak{H}_c)$ such that

$$(f \circ g)_{c,\hbar} \equiv \{\gamma_{c,\hbar}(f,g) + \gamma_{c,\hbar}(g,f)\}/2,$$

respectively,

$$\{f,g\}_{c,\hbar} \equiv \{\gamma_{c,\hbar}(f,g) - \gamma_{c,\hbar}(g,f)\}/i\hbar,$$

corresponds to $(FG + GF)/2$ (respectively, $[F,G]/i\hbar$).

5. *Remarks*: (a) The proof of the above lemma is made by straightforward inspection. (b) The assertion of the lemma would carry over if $\mathfrak{A}_{c,\hbar}$ were to be replaced by $\overline{\mathfrak{A}}_{c,\hbar} = \mathfrak{L}^1(H_{c/\hbar}) \times \mathfrak{L}^\infty(M_c)$, with the composition laws extended accordingly; the formulation chosen in the lemma is, however, better adapted to the sequel (see, e.g., Lemma 7 and Theorem 11 below). (c) In case f and g are functions of ξ only, the involution $f \to f^*$ [respectively, the twisted convolution product $\gamma_{c,\hbar}(f,g)$] reduces to the ordinary involution (respectively, the ordinary convolution product) usually defined on $\mathfrak{L}^1(H_{c/\hbar})$. (d) At fixed c, but with $\hbar$ free to run over, say $(0,1]$, the objects in $\mathfrak{A}_{c,\hbar}$ do not change: only their composition laws do. One then verifies easily the following assertion.

6. *Lemma*: As $\hbar \to 0$, the nonabelian *-algebra $\mathfrak{A}_{c,\hbar}$ contracts to the abelian *-algebra $\mathfrak{A}_c$, whose elements are the functions

$$f:(\xi,\zeta) \in R^2 \times M_c \to f(\xi,\zeta) \in C,$$

which are continuous of compact support in ξ, and bounded $\mathscr{C}^\infty$ in ζ; and whose composition laws are

$$f^*:(\xi,\zeta) \in R^2 \times M_c \to f(\xi^{-1},\zeta)^* \in C,$$

$$\gamma_c(f,g):(\xi,\zeta) \in R^2 \times M_c \to \int d\eta f(\eta,\zeta) g(\eta^{-1}\xi,\zeta) \in \mathfrak{A}_c,$$

where $d\eta$ (respectively, $\eta^{-1}\xi$) is the Euclidean measure $ds\,dt$ [respectively, the Euclidean addition $(-s_1 + s_2, -t_1 + t_2)$] in R^2.

The contraction of the Lie structure of $\mathfrak{A}_{c,\hbar}$ as $\hbar \to 0$ requires a more detailed argument, which we now give.

7. *Lemma*: for every $f \in \mathfrak{A}_c$ define

$$\tilde{f}:(p_x,p_y;x,y) \in R^2 \times M_c \to \int\int dx'dy' \exp\{-i(p_x x' + p_y y')\} f(x',y';x,y).$$

Then (i) the Jordan product $(f \circ g)_{c,\hbar}$ on $\mathfrak{A}_{c,\hbar}$ induces an $\mathfrak{A}_c$ the composition law given by the point-wise multiplication in all variables $(p_x,p_y;x,y)$;

(ii) the Lie bracket $\{f,g\}_{c,\hbar}$ on $\mathfrak{A}_{c,\hbar}$ induces on $\mathfrak{A}_c$ the Poisson bracket

$$\{\tilde{f},\tilde{g}\}_c = -\omega_c(X_{\tilde{f}}, X_{\tilde{g}}) \quad \text{with} \quad X_{\tilde{f}} \lrcorner \omega_c = -d\tilde{f},$$

where ω_c is the symplectic form with Darboux coordinates

$$\{\bar{p}_x = \exp(y/2c)p_x, \bar{p}_y = p_y - (x/2c)\exp(y/2c)p_x; x,y\}.$$

Proof: Since the convolution product on $\mathfrak{A}_c$ (see Lemma 6) is abelian, it coincides with the composition law induced on $\mathfrak{A}_c$ by the Jordan product on $\mathfrak{A}_{c,\hbar}$; assertion (i) then follows by usual Fourier transform (in R^2!). A straightforward computation shows that, as $\hbar \to 0$, the quantum Lie bracket reduces, in the Fourier transform realization of $\mathfrak{A}_c$, to

$$\begin{aligned}\{\tilde{f},\tilde{g}\}_c(p_x,p_y;x,y) &= \exp(-y/2c)\{\partial_x\tilde{f}\cdot\partial_{p_x}\tilde{g} - \partial_x\tilde{g}\cdot\partial_{p_x}\tilde{f}\}(p_x,p_y;x,y) \\ &+ (x/2c)\{\partial_x\tilde{f}\cdot\partial_{p_y}\tilde{g} - \partial_x\tilde{g}\cdot\partial_{p_y}\tilde{f}\}(p_x,p_y;x,y) \\ &+ 1\{\partial_y\tilde{f}\cdot\partial_{p_y}\tilde{g} - \partial_y\tilde{g}\cdot\partial_{p_y}\tilde{f}\}(p_x,p_y;x,y) \\ &- (p_x/c)\{\partial_{p_x}\tilde{f}\cdot\partial_{p_y}\tilde{g} - \partial_{p_x}\tilde{g}\cdot\partial_{p_y}\tilde{f}\}(p_x,p_y;x,y).\end{aligned}$$

One then verifies that

$$\{x,y\}_c = 0,$$
$$\{x,\bar{p}_y\}_c = 0 = \{y,\bar{p}_x\}_c,$$
$$\{\bar{p}_x,\bar{p}_y\}_c = 0,$$
$$\{x,\bar{p}_x\}_c = 1 = \{y,\bar{p}_y\}_c.$$

With the change of variables $(p_x,p_y;x,y)\mapsto(\bar{p}_x,\bar{p}_y;x,y)$, the expression for $\{\tilde{f},\tilde{g}\}_c$ becomes

$$\{\tilde{f},\tilde{g}\}_c = -\omega_c(X_{\tilde{f}},X_{\tilde{g}}) \text{ with } X_{\tilde{f}}\lrcorner\omega_c = -d\tilde{f},$$

where

$$\omega_c = d\bar{p}_x\wedge dx + d\bar{p}_y\wedge dy. \qquad ■$$

8. *Corollary*: $\{\lambda p_x + \mu p_y | \lambda,\mu\in R\}$ equipped with the Poisson bracket $\{.,.\}_c$ is a representation of the Lie algebra $\mathfrak{h}_c$ of H_c.

Proof: On the one hand $\{p_x, p_y\}_c = -p_x/c$. On the other hand, with our parametrization of H_c, we construct a basis for $\mathfrak{h}_c$ as follows:

$$\Xi_x \equiv c^{-1}\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix} = \frac{d}{dx}\begin{pmatrix}1 & x/c\\ 0 & 1\end{pmatrix}\Bigg|_{x=0},$$

$$\Xi_y \equiv (2c)^{-1}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$$
$$= \frac{d}{dy}\begin{pmatrix}\exp(y/2c) & 0\\ 0 & \exp(-y/2c)\end{pmatrix}\Bigg|_{y=0}.$$

We have then $[\Xi_x,\Xi_y] = -\Xi_x/c$, so that the representation is given explicitly by $\Phi(\lambda\Xi_x + \mu\Xi_y) = \lambda p_x + \mu p_y$. ■

9. The Fourier transform, which appeared in Lemma 7, was introduced there as a mere mathematical convenience, although that lemma, and its corollary 8, do indicate that a deeper phenomenological meaning should be sought. The remainder of this section proposes such an interpretation.

Let $\mathfrak{A}_{c,\hbar}$ (respectively, $\mathfrak{A}_c$) be equipped with the topology it inherits from $\mathfrak{L}^1(H_{c/\hbar})\times\mathfrak{L}^\infty(M_c)$ [respectively, $\mathfrak{L}^1(R^2)\times\mathfrak{L}^\infty(M_c)$]. A *quantum state* is defined as a positive, continuous linear functional of norm 1 on $\mathfrak{A}_{c,\hbar}$. A family $\{\varphi_\hbar|\hbar\in(0,1]\}$ of quantum states is said to be *classical* whenever

$$\varphi_0: f\in\mathfrak{A}_c\mapsto\lim_{\hbar\to 0}\langle\varphi_\hbar;f\rangle\in C$$

defines a positive, continuous linear functional of norm 1 on $\mathfrak{A}_c$. Equivalent definitions of a classical family of quantum states, and several classes of examples, have been described in Ref. 8 for the flat case ($c=\infty$), but the above condition will suffice for our present purpose.

10. *Lemma*: To every classical family $\{\varphi_\hbar|\hbar\in(0,1]\}$ of quantum states corresponds a positive measure $d\varphi$, of norm 1, concentrated on the space $\mathfrak{X}$ ($\simeq R^2\times M_c$) of the pure states of $\mathfrak{A}_c$ such that

$$\langle\varphi_0;f\rangle = \int d\varphi(p_x,p_y;x,y)\tilde{f}(p_x,p_y;x,y),$$

where

$$\tilde{f}:(p_x,p_y;x,y)\in\mathfrak{X}$$
$$\to\int dx'dy'\exp\{-i(x'p_x + y'p_y)\}f(x',y';x,y).$$

Proof: Upon writing

$$\varphi_\hbar: f\in\mathfrak{A}_{c,\hbar}\to\int d\mu_{c/\hbar}(\xi)d\mu_c(\zeta)f(\xi,\zeta)\varphi_\hbar(\xi,\zeta)$$

we notice that the condition that $\varphi_\hbar$ be a positive functional on $\mathfrak{A}_{c,\hbar}$ is equivalent to

$$\int d\mu_c(\zeta)\sum_{j,k=1}^{n} g_k(\zeta)^* g_j(\zeta)\varphi_\hbar(\gamma_{c/\hbar}(\xi_k^{-1},\xi_j),\gamma_c(\xi_{k\hbar}^{-1},\zeta))\geqslant 0.$$

This implies that

$$\varphi_0: f\in\mathfrak{A}_c\to\int d\xi d\mu_c(\zeta)f(\xi,\zeta)\varphi_0(\xi,\zeta),$$

where φ_0 satisfies the positivity condition

$$\int d\mu_c(\zeta)\sum_{j,k=1}^{n} g_k(\zeta)^* g_j(\zeta)\varphi_0(\xi_k^{-1}\xi_j,\zeta)\geqslant 0.$$

From the generalized Bochner theorem, there exists therefore a positive measure $d\varphi$ of norm 1, concentrated on the pure states of $\mathfrak{A}_c$, such that

$$\varphi_0(\xi,\zeta) = \int d\varphi(p,\zeta')e_{p,\zeta'}(\xi,\zeta).$$

Since $\mathfrak{A}_c$ is abelian, its pure states are product states of the form

$$e_{p,\zeta'}(\xi,\zeta) = e_p(\xi)\delta^c_{\zeta'}(\zeta),$$

where

$$e_p(\xi) = \exp\{-i(x'p_x + y'p_y)\};$$
$$\int d\mu_c(\zeta)\delta^c_{\zeta'}(\zeta)g(\zeta) = g(\zeta').$$

Hence

$$\langle\varphi_0;f\rangle = \int d\varphi(p,\zeta)\int d\xi f(\xi,\zeta)e_p(\zeta).$$

We collect our results in the following summary.

11. **Theorem**: To every quantum observable

$$F = \int d\mu_{c/\hbar}(\xi)f(\xi,.)U_\hbar(\xi)$$

[where $(f(\xi,.)\psi)(\zeta) = f(\xi,\zeta)\psi(\zeta)$; $(U_\hbar(\xi)\psi)(\zeta) = \psi(\gamma_c(\xi_\hbar^{-1},\zeta))\ \forall\psi\in\mathfrak{H}_c = \mathfrak{L}^2(M_c,d\mu_c)$] corresponds a classical observable, namely a function $\tilde{f}$ on $\mathfrak{X} = \{(p,\zeta)|\,p=(p_x,p_y)\in R^2;\ \zeta=(x,y)\in M_c\}$ given by

$$\tilde{f}(p,\zeta) = \int d\xi f(\xi,\zeta)e_p(\xi)$$

[where $e_p(\xi) = e_{p_x,p_y}(x',y') = \exp\{-i(x'p_x + y'p_y)\}$] such that, for every classical family $\{\varphi_\hbar|\hbar\in(0,1]\}$ of quantum states, one has

$$\text{(i) } \lim_{\hbar\to 0}\langle\varphi_\hbar;F\rangle = \int d\varphi(p,\zeta)\tilde{f}(p,\zeta),$$

where $d\varphi$ is a positive measure of norm 1 on $\mathfrak{X}$;

$$\text{(ii) } \lim_{\hbar\to 0}\langle\varphi_\hbar;F\circ G\rangle = \int d\varphi(p,\zeta)(\tilde{f}\cdot\tilde{g})(p,\zeta),$$

where $(\tilde{f}\cdot\tilde{g})$ is the point-wise multiplication of functions on $\mathfrak{X}$;

$$\text{(iii) } \lim_{\hbar\to 0}\langle\varphi_\hbar;[F,G]/i\hbar\rangle = \int d\varphi(p,\zeta)\{\tilde{f},\tilde{g}\}_c(p,\zeta),$$

where $\{\tilde{f},\tilde{g}\}_c$ is the Poisson bracket associated with the symplectic form $\omega_c = d\bar{p}_x \wedge dx + d\bar{p}_y \wedge dy$ on $\mathfrak{X}$, with Darboux coordinates given by

$$\bar{p}_x = \exp(y/2c)p_x \text{ and } \bar{p}_y = p_y - (x/2c)\exp(y/2c)p_x.$$

12. *Remarks*: (a) This theorem solves the Dirac problem for M_c by constructing, via the classical limit $\hbar \to 0$ of the initial quantum theory, a correspondence principle $F \to \tilde{f}$ between quantum and classical observables. In this limiting procedure the Jordan and Lie structures of the classical theory are obtained from purely quantum premises; so is the momentum map of Corollary 8. In that sense, the classical mechanics of a particle with curved configuration space M_c is completely derived from the primary theory, namely from the quantum mechanics of such a particle, as axiomatized by the system of imprimitivity formulation of Mackey.

(b) The classical limit selects unambiguously the variables p_μ as the physical observables respectively associated to the coordinates x^μ of M_c. These momenta have to be distinguished from the variables $\bar{p}_\mu$ which only appear as a mathematical convenience, namely as Darboux coordinates in the diagonalization of the symplectic form ω_c canonically associated with the classical Poisson bracket derived from the quantum theory. This distinction is intimately linked to the nonvanishing curvature of the homogeneous Riemannian manifold M_c; indeed the explicit form of the $\bar{p}_\mu$ in terms of the p_μ shows that this distinction disappears in the limit $c \to \infty$, i.e., $K_c \to 0$. In this limit, the results of Ref. 8 are completely recovered, thus showing that the geometric dequantization program outlined there has a nontrivial extension to nonflat homogeneous Riemannian manifolds.

(c) The starting point for the generalization to higher dimensions of the considerations presented in this section is indicated in Remark 3(d).

III. QUANTUM GEODESIC FLOWS

1. In this section we bring in perspective the role of the Laplace–Beltrami operator

$$\Delta_c = g_c^{-1/2}\partial_\mu g_c^{\mu\nu} g_c^{1/2}\partial_\nu$$

in the quantization of classical geodesic flows, with special attention to the case where the configuration space is a Riemannian manifold of constant negative curvature. We first discuss the case where the manifold is simply connected, as exemplified by the Lobatchevski plane; we then look at the modifications to be brought to the theory for multiply connected manifolds which are compact and without boundaries.

Our first remark is to notice that in the coordinates of the Lobatchevski plane with metric

$$(g_{c,\mu\nu}) = (c^2 + x^2 + y^2)^{-1}\begin{pmatrix} c^2 + y^2 & -xy \\ -xy & c^2 + x^2 \end{pmatrix},$$

the operator $H_0 = -\hbar^2 \Delta_c/2$ takes the form

$$H_0 = (\pi_x^2 + \pi_y^2 - c^{-2}J^2)/2,$$

where

$$\pi_\mu = -i\hbar g_c^{-1/2}\partial_\mu,$$
$$J = -i\hbar(x\partial_y - y\partial_x)$$

are, respectively, the momenta canonically associated with the Lorentz boosts in the directions x^μ, and to the rotations. H_0 is thus invariant under the action of the unitary representation

$$(U(g)\psi)(\mathbf{x}) = \psi(g^{-1}[\mathbf{x}])$$

of the group $G = \mathrm{SL}(2,R)/Z_2$; and Δ_c is in fact the corresponding representative of the Casimir operator which generates the center of the universal enveloping algebra of the Lie algebra sl(2,R). The questions then are (i) to derive the form of H_0 from first (i.e., quantum) principles; and (ii) to relate this operator to the classical Hamiltonian of the corresponding geodesic flow on T^*M_c.

2. *Scholium*: For H self-adjoint in $\mathfrak{H}_c = \mathfrak{L}^2(M_c, d\mu_c)$ the following conditions are equivalent:

(i) $H = -\hbar^2\Delta_c/2 + V(x)$ with $V: x \in M_c \to V(x) \in R$,

(ii) $[x^\mu, H]/i\hbar = g_c^{\mu\nu} g_c^{1/2} \circ \pi_\nu$,

where $\pi_\mu = -i\hbar g_c^{-1/2}\partial_\mu$ and $A \circ B = (AB + BA)/2$.

Proof: from $[x^\mu, \pi_\nu]/i\hbar = g_c^{-1/2}\delta^\mu{}_\nu I$ and $H_0 = -\hbar^2\Delta_c/2 = \pi_\mu g_c^{\mu\nu} g_c \pi_\nu/2$, we obtain $[x^\mu, H_0]/i\hbar = g_c^{\mu\nu} g_c^{1/2} \circ \pi_\nu$. From (ii) we have thus $[x^\mu, H - H_0] = 0$, i.e., $(H - H_0)$ is affiliated with $\{Q(\Delta) | \Delta \in \mathfrak{B}(M_c)\}'$; since $\{Q(\Delta) | \Delta \in \mathfrak{B}(M_c)\}''$ acting on $\mathfrak{L}^2(M_c, d\mu_c)$ is maximal abelian, we thus obtain (i). The converse implication (i)$\Rightarrow$(ii) is trivial. ∎

In conformity with the correspondence principle established in Sec. II, the commutation relation (ii) in the above scholium, with the Jordan product appearing in the rhs, reflects the classical relation $\{x, \bar{H}_0\}_c = g_c^{\mu\nu}\bar{p}_\nu$ for the Darboux coordinates $(x^\mu, \bar{p}_\nu)$, with $\bar{H}_0 = g_c^{\mu\nu}\bar{p}_\mu\bar{p}_\nu/2$, which is the Hamiltonian defining the classical geodesic flow on T^*M_c. In particular, in the flat case $(c \to \infty)$, π_μ (respectively, H_0; $\bar{p}_\mu$ and $\bar{H}_0$) reduces to $-i\hbar\,\partial_\mu$ [respectively, $(\pi_x^2 + \pi_y^2)/2$; p_μ and $(p_x^2 + p_y^2)/2$]. In this limiting case, the commutation relation (ii) can be interpreted[14] as an expression of the restriction further imposed on the theory by the condition that it be Galilean invariant.

3. *Scholium*: For a continuous one-parameter group $\{V(t) | t \in R\}$ of unitary operators on $\mathfrak{H}_c = \mathfrak{L}^2(M_c, d\mu_c)$ the following conditions are equivalent:

(i) $V(t) = \exp\{-iHt\}$ with $H = f(\Delta_c)$,

(ii) $U(g)V(t)U(g^{-1}) = V(t)\,\forall t \in R, g \in G = \mathrm{SL}(2,R)/Z_2$.

Proof: Let K be the stabilizer of the origin in the Lobatchevski plane; we have then $M_c \simeq G/K$, and

$$K = \left\{\begin{pmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{pmatrix} \middle| \vartheta \in [0, 2\pi]\right\}.$$

$U(G)$ is the representation of G induced by the identity representation of K, and thus $\mathfrak{H}_c$ can be seen as the space of functions $\psi: G \to C$ satisfying $\psi(gk) = \psi(g)\,\forall (g,k) \in G \times K$, and which are square integrable over G/K with respect to the canonical measure μ_c. With X defined by

$$(X\psi)(g) = \int d\mu_c(g')X(g,g')\psi(g'),$$

the condition that X belong to $U(G)'$ is equivalent to $X(g,g') = \chi(g^{-1}g')$, where χ is a spherical function on G with respect to K, i.e., that χ is constant on the double cosets $K{:}K$. As a consequence of the following three facts, the algebra $\mathscr{A}_K(G)$ of spherical functions is abelian[15]: (a) K is compact; (b) K is maximal, i.e., $K = \{g\in G\,|gKg^{-1} = K\}$; and (c) the transposition $g\to\bar{g}$ in SL(2,R) is an antiautomorphism and satisfies $\bar{k} = k^{-1}$ for all k in K. We thus have $U(G)'\subseteq U(G)''$, i.e., $U(G)$ is multiplicity free, and $V(R) \subseteq U(G)'\cap U(G)''$. From this and the remark at the end of Sec. III. 1, condition (i) of the scholium follows immediately. The converse implication, (i)$\Rightarrow$(ii), is trivial.

4. *Corollary*: For H self-adjoint in $\mathfrak{H}_c = \mathfrak{L}^2(M_c,d\mu_c)$ the following conditions are equivalent:

(i) $H = -\hbar^2\Delta_c/2 + \lambda I$ with $\lambda\in R$,

(ii) $V(R)\equiv\{\exp(-iHt/\hbar|t\in R\}$

commutes with

$U(G)\restriction H_c$,

and

$$[x^\mu,H]/i\hbar = g_c^{\mu\nu}g_c^{1/2}\circ\pi_\nu.$$

Proof: This follows immediately from Scholium 2 and the fact that H_c acts transitively on M_c. ■

Each of the equivalent conditions of this corollary thus implies that $V(R)$ commutes with the full $U(G)$ (and in particular that the Hamiltonian H is rotation-invariant). As shown by Scholium 3 the latter condition, namely that $V(R)$ commutes with $U(G)$, is weaker than the conditions of Corollary 4. Via the correspondence principle established in Sec. II each of the equivalent conditions of Corollary 4 characterizes completely the classical geodesic flow on T^*M_c. The ambiguity left by Scholium 3 also has its equivalent in the classical limit: It only places in a general setting such well-known facts as, for instance, that the trajectories of a free particle in flat configuration space are straight lines for special relativity, i.e., $H = (\Sigma_k p_k^2 + m^2)^{1/2}$, as well as for Galilean relativity, i.e., $H = \Sigma_k p_k^2/2$.

5. We now turn to the formulation of the quantum and classical theories on compact Riemannian manifolds of constant negative curvature. These manifolds are obtained as quotients $\Gamma\backslash M_c = \Gamma\backslash G/K$ of the Lobatchevski plane $M_c\simeq G/K$ by a discrete subgroup $\Gamma\subset G$. This is quite analogous to the identification of the flat torus as the quotient of the Euclidean plane $R^2\simeq E^2/K$ by the discrete subgroup $Z^2\subset E^2$. There are, however, two essential differences: (i) the restriction imposed on the volume of $\Gamma\backslash G/K$ by the Gauss–Bonnet formula; and (ii) the fact that, whereas the torus still has a continuous group of smooth isometries (namely $Z^2\backslash R^2$, where R^2 is the translation subgroup of the Euclidean group E^2), this is no more the case for the curved manifolds $\Gamma\backslash G/K$. This latter fact can be obtained as follows. Since Γ has no fixed point (it identifies opposite sides by pairs) and cannot be simply conjugate to a subgroup of either one of the maximal abelian subgroups

$$A = \left\{\begin{pmatrix}\lambda & 0\\ 0 & \lambda^{-1}\end{pmatrix}\Big|\lambda>0\right\} \quad\text{or}\quad N = \left\{\begin{pmatrix}1 & \mu\\ 0 & 1\end{pmatrix}\Big|\mu\in R\right\},$$

we can assume without loss of generality that Γ is non-abelian and contains an element $\gamma\in A$ with $\gamma\neq e$. As a consequence, one obtains by contradiction (see, e.g., Ref. 16) that $N_G(\Gamma)\equiv\{g\in G\,|g\Gamma g^{-1} = \Gamma\}$ is discrete. So is then the group $\mathscr{A}_G(\Gamma) = \Gamma\backslash N_G(\Gamma)$ of smooth isometries of $\Gamma\backslash G/K$. Since this manifold is compact, $\mathscr{A}_G(\Gamma)$ is in fact finite.

This precludes any direct use of the system of imprimitivity approach to formulate the quantum theory of a particle moving on $\Gamma\backslash G/K$. Since the covering space M_c is nevertheless available, we start with the theory on M_c, as established in Sec. II, and restrict it to $\Gamma\backslash M_c$ by making use of the periodic boundary conditions provided by Γ. We obtain in this manner the projection valued measure Q_Γ: $E\in\mathfrak{B}(\Gamma\backslash M_c)\to Q_\Gamma(E)\in\mathfrak{P}(\mathfrak{H}_\Gamma)$, where $\mathfrak{H}_\Gamma$ is the Hilbert space of square-integrable functions on $\Gamma\backslash M_c$ with respect to the measure $d\mu_\Gamma$ canonically inherited from $d\mu_c$. The Laplace–Beltrami operator Δ_Γ^c is similarly obtained, and its properties have been extensively studied.[17]

6. *Scholium*: With $H_\Gamma = -\hbar^2\Delta_\Gamma^c/2$, and $V_\Gamma(t) = \exp\{-iH_\Gamma t/\hbar\}$, the von Neumann algebra $\{Q_\Gamma(E), V_\Gamma(t)|E\in\mathfrak{B}(\Gamma\backslash M_c), t\in R\}''$ coincides with $\mathfrak{L}(\mathfrak{H}_\Gamma)$.

Proof: Impose first $BQ_\Gamma(E) = Q_\Gamma(E)B$ for all E in $\mathfrak{B}(\Gamma\backslash M_c)$. Since the $Q_\Gamma(E)$'s generate a maximal abelian von Neumann subalgebra of $\mathfrak{L}(\mathfrak{H}_\Gamma)$, we have $(B\psi)(x) = B(x)\psi(x)$ with $B(.)\in\mathfrak{L}^\infty(\Gamma\backslash M_c)$. Since $\Gamma\backslash M_c$ has finite volume, we have $B(.)\in\mathfrak{L}^2(\Gamma\backslash M_c)$. Impose moreover $BV_\Gamma(t) = V_\Gamma(t)B$ for all t in R. From the fact that 0 is a nondegenerate eigenvalue of Δ_Γ^c with eigenvector 1, we obtain that

$$\begin{aligned}(V_\Gamma(t)B)(x) &= (V_\Gamma(t)B\,1)(x)\\ &= (BV_\Gamma(t)1)(x) = (B\,1)(x) = B(x)\end{aligned}$$

implies $B(x) = \lambda\cdot 1(x)$ with $\lambda\in C$, and thus $B = \lambda I$. Hence $\{Q_\Gamma(E), V_\Gamma(t)|E\in\mathfrak{B}(\Gamma\backslash M_c), t\in R\}$ acts irreducibly on $\mathfrak{H}_\Gamma$. ■

In physical terms, this scholium asserts that all the observables of the theory can be expressed as functions of two kinds of fundamental observables: the position observables Q_Γ *and* the energy H_Γ. This generalizes to the curved manifolds $\Gamma\backslash M_c$ a result well known for the flat torus. Here, however, the *scholium* is important in that it allows one to bypass the absence of the momentum observables, a difficulty inherent to the curved space situation where the geometrical symmetry group $\mathscr{A}_G(\Gamma)$ is discrete.

7. Whereas the spectrum of $-\Delta_\Gamma^c$ is discrete, and positive, some degeneracies might occur amongst its strictly positive eigenvalues, so that the quantum systems, in contradistinction with the corresponding classical systems (with $c<\infty$), are not strictly ergodic. Still, as $\hbar\to 0$, the Weyl formula (in the form given for instance in Ref. 18) can be used to show that the number $N_\hbar(E)$ of eigenvalues, for $H_\Gamma = -\hbar^2\Delta_\Gamma^c/2$, contained in any interval $(E,E+h^2)$ with $E>0$, satisfies

$$\nu(E)\equiv\lim_{\hbar\to 0}N_\hbar(E) = 2\pi\mu_\Gamma(\Gamma\backslash M_c).$$

This again generalizes to the curved manifolds $\Gamma\backslash M_c$ a result well known for the flat torus. Here, however, there is a difference in interpretation: Since, for $c<\infty$, the classical system is ergodic, the rhs of the above expression is equal to the volume of $T_1^*(\Gamma\backslash M_c)$, i.e., to the volume which is "occupied" by the classical particle in the phase space $T^*(\Gamma\backslash M_c)$. Moreover, by the Gauss–Bonnet formula, this volume is determined, at fixed curvature $K_c = -c^{-2}$, by the genus g of the surface; we have thus

$$\nu(E)/8\pi^2c^2 = g - 1\in Z_+.$$

For finite, but small, values of $\hbar$, this result holds to $O(\hbar^4)$. One can actually do somewhat better, as the next result shows, when we compare the quantum partition function

$$Z_\Gamma(\beta,\hbar,c) = \mathrm{Tr}\exp(-\beta H^c_\Gamma)$$

with

$$H^c_\Gamma = -\hbar^2(\Delta^c_\Gamma - \alpha I)/2$$

and the classical partition function

$$\begin{aligned} Z_\Gamma(\beta,c) &= \int\int dp_x dp_y \int\int_{\Gamma\setminus M_c} dx dy \exp\{-\beta g_c^{\mu\nu}(x,y) p_\mu p_\nu\} \\ &= 2\pi\mu^c_\Gamma(\Gamma\setminus M_c)/\beta. \end{aligned}$$

8. *Proposition*: With $H^c_\Gamma = -\hbar^2(\Delta^c_\Gamma - \alpha I)/2$ and $\alpha = \alpha_c \equiv K_c/6$

$$\frac{h^2 Z_\Gamma(\beta,\hbar,c)}{Z_\Gamma(\beta,c)} - 1 = O((\beta\hbar^2)^2) \quad \text{as} \quad \beta\hbar^2 \to 0,$$

whereas one has only $O(\beta\hbar^2)$ when $\alpha \neq \alpha_c$.

Proof: The Selberg asymptotic formula reads[17]

$$\sum_n \exp(-\epsilon\lambda_n) \simeq (4\pi\epsilon)^{-d/2}(a_0 + a_1\epsilon + \cdots + a_n\epsilon^n + \cdots),$$

where λ_n are the eigenvalues $0 = \lambda_0 < \lambda_1 \leqslant \lambda_2 \leqslant \cdots$ of the Laplace–Beltrami operator $-\Delta^c_\Gamma$ for the two-dimensional ($d = 2$) manifold $\Gamma\setminus M_c$;

$$a_0 = \mu^c_\Gamma(\Gamma\setminus M_c) \quad \text{and} \quad a_1 = \frac{1}{6}\int_{\Gamma\setminus M_c} d\mu^c_\Gamma K_c.$$

From this one computes

$$\frac{h^2 Z_\Gamma(\beta,\hbar,c)}{Z_\Gamma(\beta,c)} - 1 = \tfrac{1}{2}\beta\hbar^2(\tfrac{1}{6}K_c - \alpha) + O((\beta\hbar^2)^2). \quad \blacksquare$$

The interest of this result is that the curvature correction $\alpha_c = K_c/6$ to the Laplace–Beltrami operator $-\Delta^c_\Gamma$ provides the best possible fit of the quantum partition function to its classical limit. This is to be compared with the original proposal[19] concerning the possible dynamics on curved, infinite, and simply connected spaces; see also Ref. 20 for a connection with the BKS-kernel of the geometric quantization program, and the results[21] based on the Feynman path integral.

ACKNOWLEDGMENTS

This work was done while the author was on academic leave of absence from the University of Rochester. He wishes to thank Professor G. W. Mackey and Professor S. Sternberg for making possible his visit with the Mathematics Department at Harvard during the Spring of 1981, Professor K. Schmidt for the time spent at the Mathematics Research Centre of Warwick University in the Summer of 1981, and Professor R. V. Kadison for his invitation to join his group at the Mathematics Department of the University of Pennsylvania for the Fall of 1981. The research presented here benefited from stimulating discussions with R. J. Blattner (UCLA), P. Chernoff (Berkeley), V. Guillemin (MIT), S. Helgason (MIT), R. Kadison (Penn.), G. W. Mackey (Harvard), and S. Sternberg (Harvard).

[1]H. J. Groenewold, Physica **12**, 405–460 (1946); L. van Hove, Acad. R. Belg. Bull. Classe Sci. Mem. (5) **37**, 610–620 (1951).
[2]M. Born and P. Jordan, Z. Phys. **33**, 858–888 (1925).
[3]P. R. Chernoff, "*Difficulties of Canonical Quantization*" (unpublished lecture notes, Berkeley, California); see also R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin, Reading, Mass., 1978), Theorem 5.4.9.
[4]L. van Hove, Ref. 1; see also R. Abraham and J. E. Marsden, Ref. 3, Theorem 5.4.13.
[5]J. -M. Souriau, *Structure des Systèmes Dynamiques* (Dunod, Paris, 1970); B. Kostant, in *Lecture Notes in Mathematics* Vol. 170 (Springer, New York, 1970), pp. 87–208; and in *Géometrie Symplectique et Physique Mathématique* (CNRS, Paris, 1975), pp. 187–210; see also R. Abraham and J. E. Marsden, Ref. 3, Sec. 5.4.
[6]E. P. Wigner, Phys. Rev. **40**, 749–759 (1932).
[7]J. E. Moyal, Proc. Cambridge Philos. Soc. **45**, 99–124 (1949).
[8]G. G. Emch, "*A Geometric Dequantization Programme for the Solution of the Dirac Problem*," preprint, Rochester, New York, 1981.
[9]V. I. Arnold and A. Avez, *Problémes Ergodiques de la Mécanique Classique* (Gauthier-Villars, Paris, 1967), Cor. 17.12.
[10]R. Abraham and J. E. Marsden, Ref. 3, Prop. 3.1.11.
[11]G. W. Mackey, *Mathematical Foundations of Quantum Mechanics* (Benjamin, New York, 1963); *Induced Reoresentations of Groups and Quantum Mechanics* (Benjamin, New York, 1968).
[12]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley–Interscience, New York, 1963), Theorems V.3.1 and VI.7.10.
[13]G. W. Mackey, in *Lie Groups and their Representations*, edited by I. M. Gel'fand (Hilgar, London, 1975), pp. 339–363.
[14]G. W. Mackey, *Unitary Group Representations in Physics, Probability and Number Theory* (Benjamin, Reading, Mass., 1978). See also Ref. 11, and R. Abraham and J. E. Marsden, Ref. 3, Theorem 5.4.21.
[15]R. Godement, Trans. Am. Math. Soc. **73**, 496–556 (1952).
[16]I. M. Gel'fand, M. I. Graev, and I. I. Pyateskii-Shapiro, *Representation Theory and Automorphic Functions* (Saunders, Philadelphia, 1969).
[17]M. Berger, P. Granduchon, and E. Mazed, *Le Spectre d'une Variété Riemannienne, Lecture Notes in Mathematics* Vol. 194 (Springer, New York, 1971); D. A. Hejhal, *The Selberg Trace Formula for PSL (2,R), Lecture Notes in Mathematics* Vol. 548 (Springer, New York, 1976); R. Osserman and A. Weinstein, *Geometry of the Laplace Operator*, Proceedings of the Symposia in Pure Mathematics, Vol. 36 (American Mathematics Society, Providence, R. I., 1980).
[18]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Wiley, New York, 1962), Vol. I, Theorem VI.16–18.
[19]B. S. DeWitt, Rev. Mod. Phys. **29**, 337–397 (1957).
[20]J. Sniatycki, *Geometric Quantization and Quantum Mechanics* (Springer, New York, 1980), Sec. 7.2
[21]K. S. Cheng, J. Math. Phys. **13**, 1723–1726 (1972); D. Elworthy and A. Truman, in *Functional Integration*, edited by J. -P. Antoine and E. Tirapegui (Plenum, New York, 1980).

Quantum evolution in the presence of additive conservation laws and the quantum theory of measurement[a)]

G. C. Ghirardi
Istituto di Fisica Teorica dell'Università, Trieste, Italy, International Centre for Theoretical Physics, Trieste, Italy, and International School for Advanced Studies, Trieste, Italy

A. Rimini
Istituto di Fisica dell'Università, Salerno, Italy

T. Weber
Istituto di Fisica Teorica dell'Università, Trieste, Italy

(Received 22 February 1982; accepted for publication 4 June 1982)

In this paper we derive in a completely rigorous way a family of inequalities holding for proper combinations of the squared norms of the states generated by the quantum evolution of a compound quantum system in the presence of additive conservation laws. The application of these inequalities to the quantum theory of measurement yields lower bounds for the malfunctioning of a measuring apparatus, which are valid under more general mathematical conditions and for a larger variety of physical situations than those considered up to now in the literature.

PACS numbers: 03.65.Bz

1. INTRODUCTION

In this paper we consider the evolution induced by the interaction of two quantum systems S and s when, in the interaction process, some additive quantity is conserved.

The existence of additive conserved quantities gives rise to constraints on the norms of the states which are generated by the application of the unitary evolution operatior U to the states of the system $S + s$. The importance of this fact was first recognized by Wigner[1] in connection with the quantum theory of measurement. In fact, he has pointed out that the existence of additive conserved quantities by itself entails limitations upon the measurability of certain observables. To be more precise, let us denote by H and h the Hilbert spaces of the systems S and s, respectively, and identify S with a measuring apparatus devised to measure the observable $\mathscr{M}$ associated to the self-adjoint operator M of h, and s with the measured system. In the Wigner formulation one assumes that M has a purely discrete spectrum with eigenvalues m_l and a complete set of eigenstates ψ_l, and that the unitary operator U describing the system–apparatus interaction commutes with an additive quantity $N = N_H \otimes \mathbb{I}_h + \mathbb{I}_H \otimes N_h$. Then one can prove that, unless the operators M and N_h commute, it is impossible for U to act as required by the ideal measurement scheme hypothesized by von Neumann[2]

$$U(\phi_0 \otimes \psi_\mu) = \phi_\mu \otimes \psi_\mu. \tag{1.1}$$

where ϕ_0 is the initial state of the apparatus and the ϕ_μ's are the final states satisfying $(\phi_\mu, \phi_\nu) = \delta_{\mu\nu}$. Araki and Yanase,[3] and subsequently Stein and Shimony,[4] have shown that the theorem originally suggested by Wigner can be rigorously proved when N_h is a bounded operator and the initial apparatus state ϕ_0 belongs to the domain of definition of N_H. Moreover, Stein and Shimony have made conceptually clear what has to be the mathematical requirement corresponding to the physical occurrence of an additive conservation law. In fact when N is unbounded (as happens in almost all physical cases) the sense of the statement "U commutes with N" has to be made mathematically precise.

In the case in which the measured observable of the system s does not commute with the part N_h of the additive conserved quantity N, since (1.1) cannot hold, we must resort to a nonideal measurement scheme. To this purpose, let us start by writing an equation expressing the most general type of evolution which can be induced by U on the state $\phi_0 \otimes \psi_m$:

$$U\phi_0 \otimes \psi_m = \phi_m \otimes \psi_m + \sum_{n \neq m} \phi_{mn} \otimes \psi_n. \tag{1.2}$$

In the case of the quantum theory of measurement the problem then arises of making as small as possible the norms of the states ϕ_{mn}, and almost orthogonal the states ϕ_m. In fact, both the nonorthogonality of the states ϕ_m and the presence of the states ϕ_{mn} give rise to errors, ambiguities, and distortions in the measurement process, as discussed in detail in Ref. 5. This problem has been the subject of various investigations. First of all, it can be easily proved that the norms of the states ϕ_{mn} can be made very small only by making very large the expectation value of the operator N_H^2 on the state ϕ_0.[1,3,6] Besides this general result, it then becomes relevant to obtain a quantitative estimate of the minimal deviation from the ideal measurement scheme, and this is usually expressed through the derivation of bounds for proper linear combinations of the squared norms of the states ϕ_{mn}. Up to now, such bounds have been obtained only in the case in which the conserved quantity is the total angular momentum, the Hilbert space is the finite-dimensional spin space of a particle of given spin, and the measured quantity is a spin component. More precisely, in Ref. 7 the case of spin $\frac{1}{2}$ has been dealt with in a rather heuristic way. In Ref. 5 the Yanase bound has been rederived in a simpler and more rigorous way and equations defining an optimal measuring apparatus have been obtained. In Ref. (8) we have built an explicit example of a

[a)] Work supported in part by Istituto Nazionale di Fisica Nucleare, Sezioni di Trieste e Pavia.

physically reasonable measuring apparatus working almost ideally. Finally, in Ref. 9 the derivation of Ref. 5 has been generalized to the case of a particle of arbitrary spin.

Summarizing, the bounds for the terms related to the malfunctioning of the apparatus have been obtained up to now under the following rather restrictive assumptions:

(i) the conserved quantity is the total angular momentum;

(ii) the measured quantity is a component of the spin;

(iii) the spin space, i.e., the space h, is finite dimensional.

It is then interesting to derive explicit bounds for the general case of an arbitrary additive conserved quantity, of an arbitrary measured observable and for an infinite Hilbert space h. We do this in Sec. 3 by making use of a theorem which will be proved in Sec. 2 under rather general assumptions. This theorem consists in the derivation of a family of inequalities which must be satisfied by proper linear combinations of the squared norms of the states ϕ_{mn} appearing in (1.2). These inequalities are valid for any unitary operator U which commutes with an additive conserved quantity N satisfying some general requirements. Therefore, the theorem could find applications to other quantum problems besides the quantum theory of measurement.

Before coming to the statement and to the proof of the theorem, we want to comment briefly about the correct mathematical formulation of the assumption that an additive conservation law holds for U. In fact, while for a bounded self-adjoint operator A the fact that A is conserved can be unambiguously expressed as $[U, A] = 0$, in the case in which A is unbounded this relation can be meaningless. In such a case, by the statement "A is conserved," we will mean that for every real number r

$$[U,e^{irA}] = 0. \tag{1.3}$$

As discussed in Ref. 4, this condition, from the point of view of every essential mathematical and physical consideration, should be regarded as the meaning of the proposition that the operators U and A commute.

2. A GENERAL THEOREM

Before stating our theorem let us specify the notations we will use.

We shall deal with the tensor product $\mathscr{H} = H \otimes h$ of two separable Hilbert spaces. In h we consider a complete orthonormal set of states $\{\psi_l\}$ and denote by P_l the projection operator on ψ_l.

Given a pair of integers i and j, $i \neq j$, we define two projection operators $\mathscr{P}_i$ and $\mathscr{P}_j$ according to

$$\mathscr{P}_l = P_l + Q_l, \quad l = i,j, \tag{2.1}$$

with

$$Q_l = \sum_{k \in \mathbb{N}_l} P_k, \quad l = i,j. \tag{2.2}$$

Here and in the following $\mathbb{N}_i$ and $\mathbb{N}_j$ are two subsets of the set $\mathbb{N}$ of the positive integers which are arbitrary except for the condition that $\mathbb{N}_i, \mathbb{N}_j$ and the pair i, j form a partition of $\mathbb{N}$. It follows from the definition that

$$\mathscr{P}_i \mathscr{P}_j = \mathscr{P}_j \mathscr{P}_i = 0, \tag{2.3}$$

$$\mathscr{P}_i + \mathscr{P}_j = \mathbb{I}_h. \tag{2.4}$$

Given a unitary operator U onto $\mathscr{H}$ and a normalized vector $\phi_0 \in H$, one can write with complete generality

$$U\phi_0 \otimes \psi_m = \phi_m \otimes \psi_m + \sum_{n \neq m} \phi_{mn} \otimes \psi_n, \tag{2.5}$$

where the states ϕ_m and ϕ_{mn} belong to H. In connection with the expansion (2.5) we define the quantity η_l according to

$$\eta_l^2 = \sum_{k \neq l} \|\phi_{lk}\|^2. \tag{2.6}$$

For the sake of brevity we shall use shortened notations such as, e.g., $\mathscr{P}_l$ for $\mathbb{I}_H \otimes \mathscr{P}_l$ and N_H for $N_H \otimes \mathbb{I}_h$. Furthermore, the same symbol $\|\cdot\|$ will be used to denote operator norms and vector norms in $\mathscr{H}$, H, and h.

Theorem: Let U be a unitary operator on $\mathscr{H} = H \otimes h, \{\psi_l\}$ a complete orthonormal set in h and ϕ_0 a normalized vector in H. If there exist two self-adjoint operators N_H and N_h acting on H and h, respectively, such that

(i) N_h is bounded,

(ii) ϕ_0 belongs to the domain D_{N_H} of N_H,

(iii) for all real numbers r, $[U,e^{i(N_H + N_h)r}] = 0$,

then for any pair of positive integers i and j, $i \neq j$, the following inequalities hold:

$$\begin{aligned}
&|1 - (\phi_i,\phi_j)| \cdot |(\psi_i, N_h \psi_j)| \\
&\leqslant \|N_H \phi_0\| \cdot \left(\left(\sum_{l \in \mathbb{N}_i} \|\phi_{jl}\|^2 + \|\phi_{ji}\|^2 \right)^{1/2} \right. \\
&\quad \left. + \left(\sum_{l \in \mathbb{N}_j} \|\phi_{il}\|^2 + \|\phi_{ij}\|^2 \right)^{1/2} \right) \\
&\quad + \|\mathbb{N}_h\| (2\eta_i + 2\eta_j + \eta_i \eta_j),
\end{aligned} \tag{2.7}$$

where the vectors ϕ_k, ϕ_{kl} of H are defined by the expansion (2.5), the positive real numbers η_l are given by Eq. (2.6), and the sets of positive integers $\mathbb{N}_i$ and $\mathbb{N}_j$ are arbitrary except for the condition that their union covers the whole set of positive integers except the pair i and j.

Proof: Having chosen the pair i, j and defined the operators $\mathscr{P}_l$ through Eqs. (2.1), Eq. (2.4) allows us to write the operator identity

$$\begin{aligned}
e^{i(N_H + H_h)r} &= e^{i(N_H + H_h)r} \mathscr{P}_i + \mathscr{P}_j e^{i(N_H + N_h)r} \\
&\quad - \mathscr{P}_j e^{i(N_H + N_h)r} \mathscr{P}_i + \mathscr{P}_i + \mathscr{P}_i e^{i(N_H + N_h)r} \mathscr{P}_j.
\end{aligned} \tag{2.8}$$

Multiplying by U^+ on the left and by U on the right, taking into account the commutativity condition (iii) one gets the relation

$$\begin{aligned}
e^{i(N_H + N_h)r} &= e^{i(N_H + N_h)r} U^+ \mathscr{P}_i U + U^+ \mathscr{P}_j U e^{i(N_H + N_h)r} \\
&\quad - U^+ \mathscr{P}_j e^{iN_h r} \mathscr{P}_i e^{-iN_h r} U e^{i(N_H + N_h)r} \\
&\quad + U^+ \mathscr{P}_i e^{iN_h r} \mathscr{P}_j e^{-iN_h r} U e^{i(N_H + N_h)r}.
\end{aligned} \tag{2.9}$$

In the right-hand side of the above equality we have brought the operator $e^{iN_H r}$ either to the left or to the right in each term appearing in it; this has been done since we will sandwich this equation between the states $\phi_0 \otimes \psi_i$ and $\phi_0 \otimes \psi_j$ and we will subsequently take the limit for $r \rightarrow 0$. We obtain in this way a correct expression, since the operator N_H will then act on the vector $\phi_0 \in D_{N_H}$.

For $r\neq 0$ let us define two bounded operators $g(r)$ and $G(r)$ through the equations

$$e^{iN_h r} = \mathbb{I}_h + irg(r),$$
$$e^{iN_H r} = \mathbb{I}_H + irG(r), \qquad (2.10)$$

and note that under the limit $r\to 0$ the operator $g(r)$ converges strongly to N_h and the state $G(r)\phi_0$ converges strongly to $N_H\phi_0$. Sandwiching (2.9) between the states $\phi_0\otimes\psi_i$ and $\phi_0\otimes\psi_j$ with $i\neq j$ and using (2.10) we have

$$\begin{aligned}
(\phi_0,[\mathbb{I}_H + irG(r)]\phi_0)\cdot(\psi_i,[\mathbb{I}_h + irg(r)]\psi_j)
&= ([\mathbb{I}_H - irG^+(r)]\phi_0\otimes\psi_i, \\
&\quad [\mathbb{I}_h + irg(r)]U^+\mathscr{P}_i U\phi_0\otimes\psi_j) \\
&\quad + (\mathscr{P}_j U\phi_0\otimes\psi_i, U[\mathbb{I}_h + irg(r)] \\
&\quad \times[\mathbb{I}_H + irG(r)]\phi_0\otimes\psi_j) \\
&\quad - (\mathscr{P}_j U\phi_0\otimes\psi_i,[\mathbb{I}_h + irg(r)] \\
&\quad \times\mathscr{P}_i[\mathbb{I}_h - irg^+(r)]U[\mathbb{I}_h + irg(r)] \\
&\quad \times[\mathbb{I}_H + irG(r)]\phi_0\otimes\psi_j) \\
&\quad + (\mathscr{P}_i U\phi_0\otimes\psi_i,[\mathbb{I}_h + irg(r)] \\
&\quad \times\mathscr{P}_j[\mathbb{I}_h - irg^+(r)]U[\mathbb{I}_h + irg(r)] \\
&\quad \times[\mathbb{I}_H + irG(r)]\phi_0\otimes\psi_j).
\end{aligned} \qquad (2.11)$$

Due to Eqs. (2.3) and (2.4) all terms not containing r cancel. Dividing by r and taking the limit for $r\to 0$, we get

$$\begin{aligned}
(\psi_i,N_h\psi_j) &= (UN_H\phi_0\otimes\psi_i,\mathscr{P}_i U\phi_0\otimes\psi_j) \\
&\quad + (\mathscr{P}_j U\phi_0\otimes\psi_i, UN_H\phi_0\otimes\psi_j) \\
&\quad + (UN_h\phi_0\otimes\psi_i\,\mathscr{P}_i U\phi_0\otimes\psi_j) \\
&\quad + (\mathscr{P}_j U\phi_0\otimes\psi_i, UN_h\phi_0\otimes\psi_j) \\
&\quad - (\mathscr{P}_j U\phi_0\otimes\psi_i, N_h\mathscr{P}_i U\phi_0\otimes\psi_j) \\
&\quad + (U\phi_0\otimes\psi_i,\mathscr{P}_i N_h\mathscr{P}_j U\phi_0\otimes\psi_j).
\end{aligned} \qquad (2.12)$$

For the last term in the r.h.s. of (2.12), recalling (2.1), we write the identity

$$\mathscr{P}_i N_h\mathscr{P}_j = P_i N_h P_j + P_i N_h Q_j + Q_i N_h\mathscr{P}_j. \qquad (2.13)$$

Observing that

$$P_l U\phi_0\otimes\psi_l = \phi_l\otimes\psi_l, \qquad (2.14)$$

we get from Eq. (2.12)

$$\begin{aligned}
[1-(\phi_i,\phi_j)]\cdot(\psi_i,N_h\psi_j) &= (UN_H\phi_0\otimes\psi_i,\mathscr{P}_i U\phi_0\otimes\psi_j) \\
&\quad + (\mathscr{P}_j U\phi_0\otimes\psi_i, UN_H\phi_0\otimes\psi_j) \\
&\quad + (UN_h\phi_0\otimes\psi_i,\mathscr{P}_i U\phi_0\otimes\psi_j) \\
&\quad + (\mathscr{P}_j U\phi_0\otimes\psi_i, UN_h\phi_0\otimes\psi_j) \\
&\quad - (\mathscr{P}_j U\phi_0\otimes\psi_i, N_h\mathscr{P}_i U\phi_0\otimes\psi_j) \\
&\quad + (P_i U\phi_0\otimes\psi_i, N_h Q_j U\phi_0\otimes\psi_j) \\
&\quad + (Q_i U\phi_0\otimes\psi_i, N_h\mathscr{P}_j U\phi_0\otimes\psi_j).
\end{aligned} \qquad (2.15)$$

Using the Schwarz inequality for all terms in the r.h.s., we get

$$\begin{aligned}
|1-(\phi_i,\phi_j)|\cdot|(\psi_i,N_h\psi_j)| &\leqslant \|N_H\phi_0\|\cdot\|\mathscr{P}_i U\phi_0\otimes\psi_j\| \\
&\quad + \|N_H\phi_0\|\cdot\|\mathscr{P}_j U\phi_0\otimes\psi_i\| + \|N_h\|\cdot\{\|\mathscr{P}_i U\phi_0\otimes\psi_j\| \\
&\quad + \|\mathscr{P}_j U\phi_0\otimes\psi_i\| + \|\mathscr{P}_j U\phi_0\otimes\psi_i\|\cdot\|\mathscr{P}_i U\phi_0\otimes\psi_j\| \\
&\quad + \|P_i U\phi_0\otimes\psi_i\|\cdot\|Q_j U\phi_0\otimes\psi_j\| \\
&\quad + \|Q_i U\phi_0\otimes\psi_i\|\cdot\|\mathscr{P}_j U\phi_0\otimes\psi_j\|\}.
\end{aligned} \qquad (2.16)$$

For the various vector norms appearing in the curly brackets we use the following relations, which can be easily obtained from (2.1), (2.2), (2.5), and (2.6):

$$\|\mathscr{P}_i U\phi_0\otimes\psi_j\| = \Big\{\sum_{l\in \mathbf{N}_i}\|\phi_{jl}\|^2 + \|\phi_{ji}\|^2\Big\}^{1/2} \leqslant\eta_j, \qquad (2.17a)$$

$$\|\mathscr{P}_j U\phi_0\otimes\psi_i\| = \Big\{\sum_{l\in \mathbf{N}_j}\|\phi_{il}\|^2 + \|\phi_{ij}\|^2\Big\}^{1/2} \leqslant\eta_i, \qquad (2.17b)$$

$$\|P_i U\phi_0\otimes\psi_i\| = \|\phi_i\|\leqslant 1, \qquad (2.17c)$$

$$\|Q_j U\phi_0\otimes\psi_j\| = \Big\{\sum_{l\in \mathbf{N}_j}\|\phi_{jl}\|^2\Big\}^{1/2} \leqslant\eta_j, \qquad (2.17d)$$

$$\|Q_i U\phi_0\otimes\psi_i\| = \Big[\sum_{l\in \mathbf{N}_i}\|\phi_{il}\|^2\Big]^{1/2} \leqslant\eta_i, \qquad (2.17e)$$

$$\|\mathscr{P}_j U\phi_0\otimes\psi_j\| = \Big\{\sum_{l\in \mathbf{N}_j}\|\phi_{jl}\|^2 + \|\phi_j\|^2\Big\}^{1/2} \leqslant 1. \qquad (2.17f)$$

Use of (2.17a)–(2.17f) in (2.16) yields immediately Eq. (2.7).

Note that this theorem is meaningful only when N_h is not diagonal on the basis $\{\psi_l\}$. We note also that, if in Eq. (2.16) we had used, instead of the simple quantities η_i and η_j which give a majorization of the norms in the l.h.s. of (2.17a)–(2.17f), the actual expressions of such norms [also shown in Eqs. (2.17a)–(2.17f)], we would have obtained a bound more stringent than (2.7). However this bound takes a rather complicated form and, for the use we intend to make of our theorem in the next section, the result (2.7) is sufficient.

3. LIMITATIONS ON QUANTUM MEASUREMENTS

As already stated in the Introduction, the theorem of Sec. 2 has a straightforward application to the quantum theory of measurement. In fact, let us suppose we want to measure an observable $\mathscr{M}$ associated to the self-adjoint operator M (with purely discrete nondegenerate spectrum)[10] of the Hilbert space of the measured system. The measurement is obtained through an interaction of the system with another quantum system acting as a measuring apparatus. Suppose also that there exists an additive quantity which is conserved during the evolution induced by the system—apparatus interaction. If the part of the additive quantity referring to the measured system is bounded and the initial state of the apparatus belongs to the domain of definition of the part referring to the apparatus, we can then identify the states ψ_l of the theorem of Sec. 2 with the complete set of eigenstates of M, the quantity $N = N_h\otimes\mathbb{I}_H + N_H\otimes\mathbb{I}_h$ with the additive conserved quantity, the operator U with the unitary evolution operator describing the system—apparatus interaction, and the state ϕ_0 with the initial state of the apparatus.

In the case in which M does not commute with N_h, Eq. (2.7) shows that an ideal measurement scheme of the type (1.1) is impossible. One is then forced to describe the measurement process through Eq. (1.2). Since, as already discussed, the appearance of the states ϕ_{mn} and the nonorthogonality of the states ϕ_i correspond to a malfunctioning of the apparatus,[5] to have a physically acceptable measuring apparatus we have to make as small as possible all the norms $\|\phi_{lk}\|$ and the scalar products (ϕ_i,ϕ_j). As shown by Eq. (2.7) this can only be obtained by monitoring the physically meaningful quantity $\|N_H\phi_0\|$, making it larger and larger. To see that this is actually the case, first of all one has to note that the left-hand side of Eq. (2.7) cannot be zero; in fact the quantity (ϕ_i,ϕ_j) must remain different from 1, since, being the norms of ϕ_i and ϕ_j bounded by 1, it could attain the value 1 only for $\phi_i = \phi_j$. This must be excluded, otherwise the measuring apparatus would respond exactly in the same way when detecting the states ψ_i and ψ_j associated to different eigenvalues of the measured quantity. Actually, as already stated, we need to make (ϕ_i,ϕ_j) become infinitesimal in order to have a correctly functioning apparatus. Secondly, the quantities $(\psi_i,N_h\psi_j)$ cannot all be zero, since N_h cannot be diagonal in the representation given by the vectors ψ_l. Then the only way to obtain an acceptable measuring apparatus is to make $\|N_H\phi_0\|$ larger and larger. In such a case we can treat as infinitesimals the quantities η_l (of order $1/\|N_H\phi_0\|$) and (ϕ_i,ϕ_j). Then, keeping in Eq. (2.7) only the dominant terms, we get

$$\frac{|(\psi_i,N_h\psi_j)|}{\|N_H\phi_0\|} \leqslant \Big(\sum_{l\in \mathrm{N}_i}\|\phi_{jl}\|^2 + \|\phi_{ji}\|^2\Big)^{1/2} + \Big(\sum_{l\in \mathrm{N}_j}\|\phi_{il}\|^2 + \|\phi_{ij}\|^2\Big)^{1/2}. \tag{3.1}$$

Using the fact that $(A)^{1/2} + (B)^{1/2} \leqslant \sqrt{2}(A+B)^{1/2}$ and taking the square of both sides of (3.1), we then get for the norms of the unwanted states the family of inequalities

$$\sum_{l\in \mathrm{N}_i}\|\phi_{jl}\|^2 + \sum_{l\in \mathrm{N}_j}\|\phi_{il}\|^2 + \|\phi_{ji}\|^2 + \|\phi_{ij}\|^2 \geqslant |(\psi_i,N_h\psi_j)|^2/2\|N_H\phi_0\|^2. \tag{3.2}$$

We recall that in Eq. (3.2) i and j ($i\neq j$) are arbitrary, and that N_i and N_j are two arbitrary subsets of the set of positive integers, such that their union reproduces all the positive integers except for the pair i,j. Formula (3.2) summarizes a set of inequalities which cannot be violated when one wants, for physical reasons, to make as small as possible the norms of the states ϕ_{mn} appearing in an evolution equation like (1.2).

As particular cases of (3.2), it is interesting to consider the following ones:

(a) Fix the pair i,j, with $i<j$, and choose then N_i to be the set of integers going from 1 to t ($i\leqslant t<j$) excluding i.

Equation (3.2) then becomes

$$\sum_{l\leqslant t}\|\phi_{jl}\|^2 + \sum_{l>t}\|\phi_{il}\|^2 \geqslant |(\psi_i,N_h\psi_j)|^2/2\|N_H\phi_0\|^2. \tag{3.3}$$

If h is the two-dimensional spin space of a spin-$\frac{1}{2}$ particle and the measured quantity is the z component of the spin, we must choose in Eq. (3.3) $i=1$, $t=1$, $j=2$. Moreover, if the additive conserved quantity not commuting with S_z is either the x or the y component of the total angular momentum $\bar{J}$, N_h is correspondingly either the x or the y component of the spin and N_H is the same component of $\bar{L} = \bar{J} - \bar{S}$. Equation (3.3) then becomes

$$\|\phi_{21}\|^2 + \|\phi_{12}\|^2 \geqslant \hbar^2/8\|L_{x,y}\phi_0\|^2. \tag{3.4}$$

This is the bound expressing the total malfunctioning of the apparatus obtained in Refs. 7 and 5.

In the analogous case of a particle of spin s we allow for convenience the indices to run from $-s$ to s. Then choosing $t=i$ arbitrarily, and $j=i+1$, we get

$$\sum_{l=-s}^{i}\|\phi_{i+1,l}\|^2 + \sum_{l=i+s}^{s}\|\phi_{il}\|^2 \geqslant \hbar^2/8\|L_{x,y}\phi_0\|^2(s+i+1)(s-i), \tag{3.5}$$

which is the bound (3.13) of Ref. 9. We note that the bound (3.15) of Ref. 9 for the total amount of distortion ε^2 (defined in general as $\varepsilon^2 = \Sigma_{l,m,l\neq m}\|\phi_{lm}\|^2$) can be obtained summing Eqs. (3.5) over all values of i from $-s$ to $s-1$.

(b) Fix the pair i,j and then choose N_j to be the empty set. Then from Eq. (3.2) we get

$$\sum_{l\neq j}\|\phi_{jl}\|^2 + \|\phi_{ij}\|^2 \geqslant |(\psi_i,N_h\psi_j)|^2/2\|N_H\phi_0\|^2. \tag{3.6}$$

The interest of inequality (3.6) lies in the fact that its left-hand side, apart from the term $\|\phi_{ij}\|^2$, is the sum of the square norms of all the unwanted terms in Eq. (1.2) with $m=j$. This quantity represents the probability of changing the state of the system when a measurement of $\mathscr{M}$ is performed, the system being initially in the jth eigenstate of M. Equation (3.6), when used for a measurement of a spin component of a spin $\frac{1}{2}$ particle, gives the result (3.4). For spin greater than $\frac{1}{2}$ Eq. (3.6) is a relation essentially different from (3.5). In particular, if one uses Eq. (3.6) to derive a bound for the total amount of distortion ϵ^2, one gets a less stringent bound than the one obtained by (3.5).

With reference to the general Eq. (3.2) we observe that, from a physical point of view, the quantities which must be made as small as possible are the $\eta_m^2 = \Sigma_{n\neq m}\|\phi_{mn}\|^2$. Unfortunately, from Eq. (3.2) one cannot get an inequality involving only a given η_m^2. On the contrary, it is quite easy to get inequalities for the total amount of distortion $\epsilon^2 = \Sigma_m \eta_m^2$. This quantity has been considered in the literature as expressing the overall malfunctioning of the apparatus. We stress, however, that this parameter is not particularly significant in the theory of measurement. In fact, in the infinite-dimensional case it can very well happen that, even though all the η_m^2 are made very small (and therefore the apparatus works in a physically acceptable way), ϵ^2 turns out to be infinite. Taking $t=i$ and $j=i+1$ in Eq. (3.3) and performing the summation over i, we get

$$\epsilon^2 \geqslant \frac{1}{2\|N_H\phi_0\|^2}\sum_i|(\psi_i,N_h\psi_{i+1})|^2. \tag{3.7}$$

For a bounded operator the series appearing in the r.h.s. can be divergent. (Incidentally, we remark that if N_h, besides being bounded is of the Hilbert-Schmidt type, then the series is convergent. Therefore in such a case apparatuses could exist leading to an arbitrarily small total amount of distortion.)

If, instead of using Eq. (3.3), we sum Eq. (3.6) for a fixed i over all j's different from i, we get in place of (3.7)

$$\epsilon^2 > (\Delta N_h)_i^2 / 2 \| N_H \phi_0 \|^2, \tag{3.8}$$

where $(\Delta N_h)_i^2$ is the mean square deviation of N_h is the state ψ_i. Since Eq. (3.8) holds for any i, ϵ^2 turns out to be larger that the maximum possible value of the r.h.s. of (3.8) when ψ_i runs over the set of the eigenstates of M. Since for a bounded operator $(\Delta N_h)_i^2 < \| N_h \|^2$, the bound (3.8) is useless when the sum of the series at the r.h.s. of (3.7) is larger than $\| N_h \|^2$.

Concluding we have proved a theorem having a straightforward application in the quantum theory of measurement. In the well-known situation in which there exists an additive quantity which is conserved during the system-apparatus interaction and does not commute with the measured quantity, the theorem allows to prove the set of inequalities (3.2) which puts lower bounds to the norms of the unwanted terms in the nonideal measurement scheme (1.2).

The assumptions under which (3.2) has been obtained are much more general than those under which bounds for the malfunctioning have been obtained in the literature up to now. In fact, in our treatment both Hilbert spaces for the system and for the apparatus are allowed to be infinite-dimensional and the apparatus part N_H of the conserved quantity is allowed to be unbounded. On the other hand, the system part N_h of the conserved quantity is assumed to be bounded.

The relevance of (3.2) lies both in the fact that its derivation is rigorous under rather general assumptions and in the fact that through the choice of the pair i, j and subsequently of the sets N_i, N_j a fairly large set of conditions is obtained. All known bounds obtained in the literature are particular cases of (3.2).

[1]E. P. Wigner, Z. Phys. **133**, 101 (1952).

[2]J. von Neumann, *Mathematical Grundlagen der Quantenmechanik* (Springer, Berlin, 1932; English editon: Princeton U. P., Princeton, 1955).

[3]H. Araki and M. M. Yanase, Phys. Rev. **120**, 622 (1960).

[4]H. Stein and A. Shimony, in *Foundations of Quantum Mechanics*, edited by B. d'Espagnat (Academic, New York, 1971), p. 56.

[5]G. C. Ghirardi, F. Miglietta, A. Rimini, and T. Weber, Phys. Rev. D **24**, 347 (1981).

[6]M. M. Yanase in *Foundations of Quantum Mechanics* edited by B. d'Espagnat (Academic, New York, 1971), p. 77.

[7]M. M. Yanase, Phys. Rev. **123**, 666 (1961).

[8]G. C. Ghirardi, F.Miglietta, A. Rimini, and T. Weber, Phys. Rev. D **24**, 353 (1981).

[9]F. Crisciani, G. C. Ghirardi, A. Rimini, T. Weber, Nuovo Cimento B **64**, 338 (1981).

[10]One can generalize in a straightforward way the results of this section to the case in which degeneracy is present, provided the measurement scheme be approximately of the state preserving type. This means, introducing the index α of degeneracy, that we assume $U\psi_{m\alpha} \otimes \phi_0 = \psi_{m\alpha} \otimes \phi_m^{(\alpha)} + X$ and $\| X \| \ll 1$. However, when degeneracy is present, it is more interesting and natural to consider as ideal a scheme which is eigenvalue-preserving rather than state-preserving. One then starts from a modification of such an ideal scheme, allowing in the evolved state small terms corresponding to eigenvalues different from the initial one and tries to derive lower bounds for their norms. How his can be done will be shown in a forthcoming paper.

Constructing measures for spin-variable path integrals

John R. Klauder[a)]
Max-Planck-Institut für Physik und Astrophysik, Munich, Federal Republic of Germany

(Received 10 February 1982; accepted for publication 23 April 1982)

By exploiting the overcompleteness of the spin-coherent states we derive expressions for spin-kinematics path integrals (specifically for spin 1/2 and spin 1) in terms of genuine (Wiener) measures on continuous paths lying on the unit sphere and for certain dynamical systems which when projected onto the subspace spanned by the proper spin-coherent-state matrix elements yield the appropriate quantum-mechanical propagator.

PACS numbers: 03.65.Bz, 02.20.Sv

I. INTRODUCTION

Quantum-mechanical path integrals have almost invariably been expressed as formal relations in view of the nonexistence of the "measure" in such formulas. Proper definitions typically involve approximating the time integrals as Riemann sums, integrating over the finitely many variables that result, and subsequently performing a limit as the mesh approximating the time integral is made infinitely fine.[1] No genuine measure on a path space emerges since the limit of the approximating measures is not countably additive. For imaginary-time quantum mechanics, on the other hand, the situation for canonical variables is quite different as one has the well-known Wiener measure on continuous configuration paths as embodied in the Feynman–Kac formula.[2]

In some recent work[3,4] it was shown that by exploiting the overcompleteness properties of the usual, canonical coherent states it was possible (in a certain projection sense) to formulate the quantum-mechanical propagator for a restricted set of dynamical systems as a well-defined integral involving genuine Wiener measures on continuous phase-space paths. In this paper we wish to show that an analogous formulation exists for other kinematical variables, in particular those associated with (half-) integer spin. Thus we shall construct (again in a certain projection sense to be defined below) a representation of the quantum-mechanical propagator for a restricted class of dynamical systems involving spin-kinematical variables as a well-defined integral involving Wiener measure on a spherical manifold. For spin 1/2 the operator structure is identical to that for a fermion degree of freedom, and thus in that case our Wiener integral provides a representation of the dynamical evolution of fermion degrees of freedom as well. We feel compelled to emphasize that this representation does *not* involve Grassmann variables or anticommuting c-numbers, but only ordinary, classical functions, the general possibility for which has been shown some time ago by the author.[5]

In Sec. II we review basic properties of the spin-coherent states with a special emphasis on consequences of their overcompleteness. In Sec. III we show how positive-definite functions involving these states lead to the basic Wiener measure on the sphere; more specifically we are led to a realization of the Green's function of the Laplacian on the sphere as a Wiener integral over the space of continuous paths on the sphere. In so doing we need the results of a certain extension of the Itô differential calculus, the details of which are presented elsewhere.[6] In Sec. IV we show how the basic Wiener measure can be combined with other factors to generate a representation of the quantum-mechanical propagator for certain dynamical systems as genuine path integrals over continuous paths on a spherical manifold.

II. SPIN COHERENT STATES

A. Aspects of the rotation group

Let $S_j, j = 1,2,3$, denote an irreducible representation of the Lie algebra of the group SU(2), which satisfies $[S_1,S_2] = iS_3$ plus cyclic permutations.[7] As an irreducible representation it follows that

$$\sum S_j^2 = s(s+1)I_s, \tag{2.1}$$

where $s = 0,1/2,1,3/2,\cdots$, and I_s is the unit operator in the $(2s+1)$-dimensional representation space $\mathfrak{H}_s$. Let $|s_m\rangle$ denote a normalized vector in $\mathfrak{H}_s$ with the property that

$$S_3|s_m\rangle = m|s_m\rangle, \tag{2.2}$$

where m is one of the variables, $-s,\ldots,s-1,s$. Let θ, φ denote the usual coordinates on the unit sphere, where $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \varphi < 2\pi$, and define[8] spin coherent states (for state s_m) by

$$\begin{aligned}|\theta,\varphi,s_m\rangle &\equiv U[\theta,\varphi]|s_m\rangle\\ &\equiv e^{-i\varphi S_3}e^{-i\theta S_2}|s_m\rangle\end{aligned} \tag{2.3}$$

for all points on the sphere. The representations of $\vec{S}$ and U will always be fixed by the vector $|s_m\rangle$. We frequently will use the notation Ω for the pair θ,φ, and denote the unitary operator by $U[\Omega]$ and the spin coherent states by $|\Omega,s_m\rangle$. In many applications only one value of s and m is needed and it suffices to simplify the notation to $|\theta,\varphi\rangle$ or $|\Omega\rangle$; we are not so fortunate.

It follows from properties of the rotation group that in each space $\mathfrak{H}_s$ the spin-coherent states admit a resolution of unity in the form

$$\left(\frac{2s+1}{4\pi}\right)\int_0^\pi\int_0^{2\pi}|\theta,\varphi,s_m\rangle\langle\theta,\varphi,s_m|\sin\theta\,d\theta\,d\varphi = I_s. \tag{2.4}$$

Henceforth we shall set

$$d\Omega \equiv \sin\theta\,d\theta\,d\varphi, \tag{2.5}$$

and we assume the integration limits are implicit. If $|\lambda,s\rangle$ and

[a)] Permanent address: Bell Laboratories, Murray Hill, New Jersey 07974, U.S.A.

$|\chi,s\rangle$ denote arbitrary vectors in $\mathfrak{H}_s$, then we can restate (2.4) as

$$\left(\frac{2s+1}{4\pi}\right)\int\langle\lambda,s|\theta,\varphi,s_m\rangle\langle\theta,\varphi,s_m|\chi,s\rangle\,d\Omega = \langle\lambda,s|\chi,s\rangle. \tag{2.6}$$

Moreover there are group orthogonality conditions which state that

$$\int\langle\lambda,s|\theta,\varphi,s_m\rangle\langle\theta,\varphi,s'_m|\chi,s'\rangle\,d\Omega=0\,, \tag{2.7}$$

provided $|s-s'|$ is a positive (nonzero) integer; note that the m values are equal for both vectors.[9]

It follows from the foregoing that for $s>0$ the spin-coherent states satisfy infinitely many conditions of linear dependency of the form

$$\int|\theta,\varphi,s_m\rangle g(\theta,\varphi)\,d\Omega=0. \tag{2.8}$$

Another manifestation of their linear dependency is the fact that two spin-coherent states are generally not orthogonal. In particular, for the special case $m=s$, it follows that

$$\langle\theta'',\varphi'',s_s|\theta',\varphi',s_s\rangle = \left[\cos\left(\frac{\theta''-\theta'}{2}\right)\cos\left(\frac{\varphi''-\varphi'}{2}\right) + i\cos\left(\frac{\theta''+\theta'}{2}\right)\sin\left(\frac{\varphi''-\varphi'}{2}\right)\right]^{2s}, \tag{2.9}$$

and these vectors are orthogonal only in the case that the two variables Ω'' and Ω' label diametrically opposite points on the sphere. We note in passing that if $m=-s$ then it follows that

$$\langle\theta'',\varphi'',s_{-s}|\theta',\varphi',s_{-s}\rangle = \langle\theta'',\varphi'',s_s|\theta',\varphi',s_s\rangle^*. \tag{2.10}$$

B. Ambiguity of operator representation

The overcompleteness of the spin-coherent states introduces an ambiguity in the representation of operators. Let B be a linear operator on $\mathfrak{H}_s$ which we wish to represent in the form

$$B=\left(\frac{2s+1}{4\pi}\right)\int\int|\theta'',\varphi'',s_m\rangle \times K_B(\theta'',\varphi'';\theta',\varphi')\langle\theta',\varphi',s_m|\,d\Omega''\,d\Omega'. \tag{2.11}$$

In view of the resolution of unity (2.4) one acceptable integral kernel K_B is always given by

$$K_B(\theta'',\varphi'';\theta',\varphi')=\left(\frac{2s+1}{4\pi}\right) \times\langle\theta'',\varphi'',s_m|B|\theta',\varphi',s_m\rangle. \tag{2.12}$$

However in view of the linear dependencies among the spin-coherent states there are infinitely many other linearly independent integral kernels K_B that yield the same operator B when inserted into (2.11). All such integral kernels form an equivalence class labelled by B, and it is convenient to denote a generic element of that class by

$$((2s+1)/4\pi)\langle\theta'',\varphi'',s_m|B|\theta',\varphi',s_m\rangle_{E.C.}. \tag{2.13}$$

As a simple example consider the unit operator $B=I_s$. One acceptable integral kernel is clearly given by

$$((2s+1)/4\pi)\langle\theta'',\varphi'',s_m|\theta',\varphi',s_m\rangle, \tag{2.14}$$

while according to (2.4) another member of the equivalence class of the unit operator is the δ-function distribution on the sphere,

$$\delta(\Omega''-\Omega')\equiv\delta(\cos\theta''-\cos\theta')\delta(\varphi''-\varphi'), \tag{2.15}$$

which has the property that

$$\int f(\Omega'')\delta(\Omega''-\Omega')\,d\Omega''=f(\Omega'), \tag{2.16}$$

for continuous functions f. In fact, as a preparation for the introduction of dynamics one of our principal tasks will be to find other elements of the equivalence class of the unit operator that admit representations as Wiener integrals over continuous paths on the sphere. To this end we first construct the basic Wiener measure on the sphere starting from positive-definite group functions.

III. BASIC WIENER MEASURE

A. Positive-definite group functions

A positive-definite group function is a continuous function $F(\theta'',\varphi'';\theta',\varphi')$ that is invariant under left group multiplication and satisfies the inequality

$$\sum_{j,k=1}^{M}\alpha_j^*\alpha_k F(\theta_j,\varphi_j;\theta_k,\varphi_k)\geqslant 0, \tag{3.1}$$

for any choice of complex numbers $\alpha_1,\ldots,\alpha_M$, and for any $M<\infty$.[10] Evidently $\langle\Omega'',s_m|\Omega',s_m\rangle$ is such a function since

$$\begin{aligned}\langle s_m|U^\dagger[\Omega\cdot\Omega'']U[\Omega\cdot\Omega']|s_m\rangle &= \langle s_m|U^\dagger[\Omega'']U^\dagger[\Omega]U[\Omega]U[\Omega']|s_m\rangle\\ &= \langle s_m|U^\dagger[\Omega'']U[\Omega']|s_m\rangle,\end{aligned} \tag{3.2}$$

and

$$\sum\alpha_j^*\alpha_k\langle\Omega_j,s_m|\Omega_k,s_m\rangle=\|\sum\alpha_k|\Omega_k,s_m\rangle\|^2\geqslant 0. \tag{3.3}$$

Consequently any sum of the form

$$F(\Omega'';\Omega')=\sum_{s,m}\rho_{s,m}\langle\Omega'',s_m|\Omega',s_m\rangle, \tag{3.4}$$

with $\rho_{s,m}\geqslant 0$ is a positive-definite group function, and a fundamental theorem[11] asserts that all positive-definite group functions can be represented in this manner.

Now if F is a positive-definite group function then so too is F^* [cf. (2.10)]. Since the product of positive-definite group functions is again one, then

$$J(\Omega'';\Omega')\equiv|F(\Omega'';\Omega')|^2, \tag{3.5}$$

and the positive sum given by

$$\begin{aligned}R(\Omega'';\Omega') &= Ae^{-\alpha}\sum_{n=0}^{\infty}\alpha^nJ^n(\Omega'';\Omega')/n!\\ &= Ae^{\alpha[J(\Omega'';\Omega')-1]},\end{aligned} \tag{3.6}$$

are positive-definite group functions for arbitrary $A>0$ and $\alpha>0$. For our purposes we shall choose F to be the expression in (2.9) for some $s>0$, and consequently

$$J(\Omega'';\Omega')=(1/2^{2s})(1+\cos\beta)^{2s}, \tag{3.7}$$

where

$$\cos\beta \equiv \cos\theta''\cos\theta' + \sin\theta''\sin\theta'\cos(\varphi'' - \varphi'). \tag{3.8}$$

It follows that

$$R(\Omega'';\Omega') = A\exp\{\alpha[2^{-2s}(1+\cos\beta)^{2s} - 1]\}. \tag{3.9}$$

We now specialize further to the choice $s = 1/2$; we shall return later to indicate what changes occur for $s \gg 1$.

The function

$$R(\Omega'';\Omega') = A\exp[\tfrac{1}{2}\alpha(\cos\beta - 1)] \tag{3.10}$$

is a positive-definite group functional, and therefore admits the general representation

$$R(\Omega'';\Omega') = \sum \rho_{s,m}\langle\Omega'', s_m|\Omega', s_m\rangle, \tag{3.11}$$

where the coefficients $\rho_{s,m} \geqslant 0$. If we examine the special case where $\theta'' = \theta' = 0$ it follows that $\beta = 0$ and thus

$$\begin{aligned} A &= \sum \rho_{s,m}\langle s_m|e^{i(\varphi''-\varphi')S_3}|s_m\rangle \\ &= \sum \rho_{s,m} e^{i(\varphi''-\varphi')m}. \end{aligned} \tag{3.12}$$

However, this can hold only if $\Sigma_s \rho_{s,m} = 0$ for all $m \neq 0$, and since the terms are all nonnegative it means that $\rho_{s,m} = 0$ for $m \neq 0$. As a consequence it follows that $R(\Omega'';\Omega')$ has the special representation

$$R(\Omega'';\Omega') = (4\pi)^{-1}\sum_{l=0}^{\infty}(2l+1)r_l\langle\Omega'', l_0|\Omega', l_0\rangle, \tag{3.13}$$

in which only integer spins contribute, and all with $m = 0$; the factor $(2l+1)/4\pi$ has been introduced for later convenience. When $\beta = 0$ it now follows that

$$R(\Omega;\Omega) = (4\pi)^{-1}\sum(2l+1)r_l = A. \tag{3.14}$$

We next wish to link A and α by the requirement that

$$\int R(\Omega;\Omega')\,d\Omega = 1. \tag{3.15}$$

Since this expression is invariant under orientation of the z-axis it is convenient to set $\Omega' = 0$, i.e., $\theta' = \varphi' = 0$, in which case

$$\begin{aligned} \int R(\Omega;0)\,d\Omega &= 2\pi A\int_0^{\pi}\exp[\tfrac{1}{2}\alpha(\cos\theta - 1)]\sin\theta\,d\theta \\ &= 4\pi A\int_0^1\exp(-\alpha y)\,dy \\ &= 4\pi A\alpha^{-1}(1 - e^{-\alpha}). \end{aligned} \tag{3.16}$$

Consequently we choose

$$A = (\alpha/4\pi)(1 - e^{-\alpha})^{-1}. \tag{3.17}$$

Later when we are interested in the case where $\alpha \gg 1$ it will be adequate to choose

$$A = \alpha/4\pi. \tag{3.18}$$

The function $R(\Omega'';\Omega')$ admits several equivalent and useful representations as given by

$$\begin{aligned} &R(\Omega'';\Omega') \\ &= (4\pi)^{-1}\sum_{l=0}^{\infty}(2l+1)r_l\langle l_0|U^\dagger[\Omega'']U[\Omega']|l_0\rangle \\ &= (4\pi)^{-1}\sum_{l=0}^{\infty}(2l+1)r_l P_l(\cos\beta) \\ &= (4\pi)^{-1}\sum_{l,m}(2l+1)r_l\langle l_0|U^\dagger[\Omega'']|l_m\rangle\langle l_m|U[\Omega']|l_0\rangle \\ &= \sum_{l,m} r_l Y^*_{l,m}(\Omega'')Y_{l,m}(\Omega'), \end{aligned} \tag{3.19}$$

where the P_l are the usual Legendre functions, and the $Y_{l,m}$ are the usual spherical harmonics which fulfill the orthogonality relation

$$\int Y^*_{l,m}(\Omega)Y_{l',m'}(\Omega)\,d\Omega = \delta_{ll'}\delta_{mm'}. \tag{3.20}$$

From (3.19) it follows that

$$\begin{aligned} r_l &= \int R(\Omega;\Omega')P_l(\cos\beta)\,d\Omega \\ &= 2\pi A\int_{-1}^{1}\exp[\tfrac{1}{2}\alpha(x-1)]P_l(x)\,dx. \end{aligned} \tag{3.21}$$

Since $|P_l(x)| \leqslant P_0(x) \equiv 1$ we learn that

$$r_l \leqslant r_0 \equiv 1. \tag{3.22}$$

More specifically, we shall need the evaluation of r_l for large α, in which case it follows that

$$\begin{aligned} r_l &= \frac{1}{2}\alpha\int_{-1}^{1}\exp[\tfrac{1}{2}\alpha(x-1)] \\ &\quad\times[1 - \tfrac{1}{2}l(l+1)(x-1)]\,dx \\ &= 1 - l(l+1)/\alpha, \end{aligned} \tag{3.23}$$

accurate to the indicated order.

Suppose we consider another positive-definite group function of similar form,

$$S(\Omega'';\Omega') = \sum s_l Y^*_{l,m}(\Omega'')Y_{l,m}(\Omega'), \tag{3.24}$$

where

$$\begin{aligned} &s_0 \equiv 1, \\ &s_l \geqslant 0, \\ &\sum s_l < \infty. \end{aligned} \tag{3.25}$$

Then it follows that the group convolution of R and S becomes

$$\begin{aligned} T(\Omega'';\Omega') &\equiv \int R(\Omega'';\Omega)S(\Omega;\Omega')\,d\Omega \\ &= \sum r_l s_l Y^*_{l,m}(\Omega'')Y_{l,m}(\Omega') \\ &\equiv \sum t_l Y^*_{l,m}(\Omega'')Y_{l,m}(\Omega'), \end{aligned} \tag{3.26}$$

which leads to a new positive-definite group function where

$$\begin{aligned} &t_0 = r_0 s_0 = 1, \\ &t_l = r_l s_l \geqslant 0, \\ &\sum t_l = \sum r_l s_l < \infty. \end{aligned} \tag{3.27}$$

Consequently functions of the sort as S form a closed class with the properties (3.25) under group convolution.

Now let us consider the N-fold convolution of R with itself given by

$$R_{(N)}(\Omega'';\Omega')\equiv\int R(\Omega';\Omega_N)\cdots \times R(\Omega_2;\Omega_1)R(\Omega_1;\Omega')\,d\Omega_N\cdots d\Omega_2\,d\Omega_1, \tag{3.28}$$

which evidently has the expansion

$$R_{(N)}(\Omega'';\Omega')=\sum r_l^{(N+1)}Y^*_{l,m}(\Omega'')Y_{l,m}(\Omega'). \tag{3.29}$$

We are interested in the limit of this expression as $N\to\infty$, and we see that we can obtain a meaningful and nontrivial limit if we link the value of α and N according to the rule

$$\alpha=2(N+1)/\nu T, \tag{3.30}$$

where ν and T are two positive parameters the significance of which will become clear later. With this identification, and the elementary fact that

$$\lim_{N\to\infty}\left[1-\frac{l(l+1)\nu T}{2(N+1)}\right]^{(N+1)}=e^{-(1/2)l(l+1)\nu T}, \tag{3.31}$$

it follows that $R_{(N)}$ converges at $N\to\infty$ to a positive-definite group function given by

$$R_T(\Omega'';\Omega')\equiv\lim_{N\to\infty}R_{(N)}(\Omega'';\Omega') =\sum e^{-(1/2)l(l+1)\nu T}Y^*_{l,m}(\Omega'')Y_{l,m}(\Omega'). \tag{3.32}$$

Furthermore it is clear that

$$\int R_T(\Omega;\Omega')\,d\Omega=1, \tag{3.33}$$

and

$$R_{T_2+T_1}(\Omega'';\Omega')=\int R_{T_2}(\Omega'';\Omega)R_{T_1}(\Omega;\Omega')\,d\Omega. \tag{3.34}$$

In addition,

$$\lim_{T\to 0}R_T(\Omega'';\Omega')=\sum Y^*_{l,m}(\Omega'')Y_{l,m}(\Omega') =\delta(\Omega''-\Omega'). \tag{3.35}$$

Consequently we may regard the function $R_T(\Omega'';\Omega')$ as the integral kernel of a semigroup (in T), which is normalized for all T in the manner of (3.33). Now let us interpret T as a *time interval*; then since by construction R, $R_{(N)}$ and thus R_T are nonnegative real functions, it follows that we can regard $R_T(\Omega'';\Omega')$ as a Markov transition probability density governing an underlying stochastic process. Thus we see the possibility of representing the positive-definite group function $R_T(\Omega'';\Omega')$ by means of a path integral over an appropriate path-space measure determined by the underlying stochastic process.

Heat kernel

It is not difficult to see that

$$R_T(\Omega'';\Omega')=(e^{(1/2)\Delta\nu T})(\Omega'';\Omega'); \tag{3.36}$$

namely, that R_T is the Green's function G for the "heat" equation, where

$$\frac{\partial}{\nu\partial T}G(\Omega,\Omega',T)=\tfrac{1}{2}\Delta G(\Omega,\Omega',T), \tag{3.37}$$

subject to the initial condition

$$G(\Omega,\Omega',0)=\delta(\Omega-\Omega'), \tag{3.38}$$

and where Δ is the Laplacian on the unit sphere,

$$\Delta=\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}. \tag{3.39}$$

Thus the representation of R_T by a path integral is at the same time a representation of the heat kernel by a path integral.

B. Spherical Wiener measure

If we set $\epsilon\equiv T/(N+1)$, then by definition we have

$$R_T(\Omega'';\Omega')=\lim_{\epsilon\to 0}\int\cdots\int\frac{1}{(2\pi\nu\epsilon)^{(N+1)}} \times\exp\left\{\frac{1}{\nu\epsilon}\sum_{k=0}^{N}[\cos\theta_{k+1}\cos\theta_k +\sin\theta_{k+1}\sin\theta_k\cos(\varphi_{k+1}-\varphi_k)-1]\right\} \times\prod_{k=1}^{N}\sin\theta_k\,d\theta_k\,d\varphi_k, \tag{3.40}$$

where $\Omega'\equiv\theta_0,\varphi_0$ and $\Omega''\equiv\theta_{N+1},\varphi_{N+1}$. We assert that this expression defines the integral of a pinned Wiener measure on the unit sphere,[12]

$$R_T(\Omega'';\Omega')\equiv\int d\mu^T_W(\theta,\varphi), \tag{3.41}$$

where it is implicit that the paths are pinned at $t=0$ and $t=T$ so that

$$\Omega'=\theta',\varphi'\equiv\theta(0),\varphi(0),$$
$$\Omega''=\theta'',\varphi''\equiv\theta(T),\varphi(T). \tag{3.42}$$

Formally the expressions (3.40) and (3.41) may be stated as

$$R_T(\Omega'';\Omega')=\mathcal{N}\int\exp\left\{-\frac{1}{2\nu}\int_0^T [\dot\theta^2+\sin^2\theta\,\dot\varphi^2]\,dt\right\}\prod_t d\Omega, \tag{3.43}$$

where $\mathcal{N}$ is a formal normalizing factor. The form of (3.43) correctly suggests that, like ordinary Wiener measure, the measure μ^T_W is concentrated on continuous (but nowhere differentiable) paths on the unit sphere. This feature has the consequence that

$$d\theta^2(t)=\nu\,dt,$$
$$\sin^2\theta(t)\,d\varphi^2(t)=\nu\,dt, \tag{3.44}$$

insofar as infinitesimal integration measures go; these are just two of the symbolic rules of the Itô differential calculus[13] for the problem at hand. We also see that ν represents the temporal scale factor in the Wiener process.

The relations (3.44) may be regarded as consequences of the appropriate Prokhorov formula,[14] which asserts, for example, for arbitrary smooth functions $B(\theta,\varphi)$ and $C(\theta,\varphi)$ that

$$E\left(\exp\left\{\frac{1}{\nu}\int B(\theta,\varphi)\,d\theta+\frac{1}{\nu}\int C(\theta,\varphi)\sin\theta\,d\varphi-\frac{1}{2\nu}\int[B^2(\theta,\varphi)+C^2(\theta,\varphi)]\,dt\right\}\right)=1, \qquad (3.45)$$

where E denotes expectation with respect to μ_W^T including integration over Ω'', and the stochastic integrals are interpreted in the sense of Itô. We illustrate this formula for a special choice of B and C in Sec. IV.

Other starting points

We next wish to outline the argument that the result for R_T did not depend on our specific assumption that $s=1/2$ in the function F in (3.5). Let us examine the result for the general case that $2s=p$, and therefore that

$$R(\Omega'';\Omega')=A\exp\{(\alpha/p)[2^{-p}(1+\cos\beta)^p-1]\}, \qquad (3.46)$$

where we have scaled α to α/p to make the comparison for different p values easier. In the present situation A is chosen so that

$$1=\int R(\Omega;0)\,d\Omega=4\pi A\int_0^1\exp[(\alpha/p)(y^p-1)]\,dy, \qquad (3.47)$$

which for large α we can approximate as

$$1=4\pi A\int_0^1\exp[-\alpha(1-y)-\tfrac{1}{2}\alpha(p-1)(1-y)^2]\,dy, \qquad (3.48)$$

which then leads to

$$A=(\alpha/4\pi)[1-(p-1)/\alpha] \qquad (3.49)$$

correct to the indicated order. This modification of A as compared with (3.18) leads to an extra overall factor

$$[1-\tfrac{1}{2}(p-1)\nu\epsilon]^{(N+1)}\rightarrow e^{-(1/2)(p-1)\nu T}. \qquad (3.50)$$

compared to the previous construction of R_T. However there is another factor that just cancels this one.

For brevity let

$$\cos\beta_k\equiv\cos\theta_{k+1}\cos\theta_k+\sin\theta_{k+1}\sin\theta_k\cos(\phi_{k+1}-\phi_k). \qquad (3.51)$$

Then the extra factor arises from the property that, to the necessary accuracy, the integrand has the form (apart from $A^{(N+1)}$)

$$\prod_{k=0}^N\exp[(\alpha/p)\{[1+\tfrac{1}{2}(\cos\beta_k-1)]^p-1\}]=\prod_{k=0}^N\exp[\alpha(\cos\beta_k-1)/2+\alpha(p-1)(\cos\beta_k-1)^2/8]. \qquad (3.52)$$

The first term leads to the Wiener measure μ_W^T on the sphere just as before, while the extra term becomes

$$\exp\left[(p-1)/(4\nu\epsilon)\sum(\cos\beta_k-1)^2\right]\rightarrow\exp\left[(p-1)/(16\nu)\int(d\theta^2+\sin^2\theta\,d\varphi^2)^2/dt\right]. \qquad (3.53)$$

For Wiener paths such a factor does not vanish, and we have shown elsewhere[6] in an extension of the Itô differential calculus that, almost surely,

$$\int\frac{(d\theta^2+\sin^2\theta\,d\varphi^2)^2}{dt}=\int\left[\frac{d\theta^4}{dt}+2\sin^2\theta\,d\theta^2\frac{d\varphi^2}{dt}+\sin^4\theta\frac{d\varphi^4}{dt}\right]=(3+2+3)\nu^2\int dt=8\nu^2T. \qquad (3.54)$$

Consequently the extra factor that comes from the integrand, namely

$$e^{(1/2)(p-1)\nu T}, \qquad (3.55)$$

exactly cancels the additional term that comes from A, leading in the end to the identical expression for R_T independent of the starting point. For convenience we shall hereafter adopt our original definition of R based on spin 1/2.

IV. PATH-INTEGRAL REPRESENTATION

A. The unit operator

Spin 1/2

We start our discussion of dynamics by constructing integral kernels that are in the equivalence class of the unit operator for spin s, specifically for spin $s=1/2$ and 1. To this end consider the positive-definite group function

$$\tilde{R}(\Omega'';\Omega')=\langle\Omega'',\tfrac{1}{2}{}_{1/2}|\Omega',\tfrac{1}{2}{}_{1/2}\rangle R(\Omega'';\Omega'), \qquad (4.1)$$

where according to (3.19)

$$R(\Omega'';\Omega')=(4\pi)^{-1}\sum_{l=0}^\infty(2l+1)r_l\langle l_0|U^\dagger[\Omega'']U[\Omega']|l_0\rangle. \qquad (4.2)$$

Since

$$\langle\Omega'',\tfrac{1}{2}{}_{1/2}|\Omega',\tfrac{1}{2}{}_{1/2}\rangle=\langle\tfrac{1}{2}{}_{1/2}|U^\dagger[\Omega'']U[\Omega']|\tfrac{1}{2}{}_{1/2}\rangle, \qquad (4.3)$$

it follows from the properties of the rotation group that $\tilde{R}$ admits an expansion given by

$$\tilde{R}(\Omega'';\Omega')=(4\pi)^{-1}\sum_{l=0}^\infty(2l+2)\tilde{r}_{l+1/2}\times\langle l+\tfrac{1}{2}{}_{1/2}|U^\dagger[\Omega'']U[\Omega']|l+\tfrac{1}{2}{}_{1/2}\rangle, \qquad (4.4)$$

where the coefficients $\tilde{r}_{l+1/2}$ are given in terms of the Clebsch-Gordan coefficients $C(s_1,s_2,j;m_1,m_2)$ by the relation

$$(2l+2)\tilde{r}_{l+1/2}\equiv(2l+1)r_l|C(l,\tfrac{1}{2},l+\tfrac{1}{2};0,\tfrac{1}{2})|^2+(2l+3)r_{l+1}|C(l+1,\tfrac{1}{2},l+\tfrac{1}{2};0,\tfrac{1}{2})|^2. \qquad (4.5)$$

This relation reduces simply to

$$\tilde{r}_{l+1/2}=\tfrac{1}{2}(r_l+r_{l+1}). \qquad (4.6)$$

We renormalize the series in (4.4) by dividing both sides by $\tilde{r}_{1/2}$, thereby introducing

$$\bar{r}_{l+1/2}\equiv\tilde{r}_{l+1/2}/\tilde{r}_{1/2}, \qquad (4.7)$$

and

$$\overline{R}(\Omega'';\Omega')\equiv(4\pi)^{-1}\sum_{l=0}^{\infty}(2l+2)\overline{r}_{l+1/2}\times\langle l+\tfrac{1}{2}{}_{1/2}|U^{\dagger}[\Omega'']U[\Omega']|l+\tfrac{1}{2}{}_{1/2}\rangle. \tag{4.8}$$

The leading term in this sum is then

$$(4\pi)^{-1}2\langle\tfrac{1}{2}{}_{1/2}|U^{\dagger}[\Omega'']U[\Omega']|\tfrac{1}{2}{}_{1/2}\rangle = (2\pi)^{-1}\langle\Omega'',\tfrac{1}{2}{}_{1/2}|\Omega',\tfrac{1}{2}{}_{1/2}\rangle, \tag{4.9}$$

and thus it follows that $\overline{R}(\Omega'',\Omega')$ is in the equivalence class of the spin-1/2 kernel (4.9) which is a member of the equivalence class of the unit operator. In the notation introduced in Sec. II,

$$\overline{R}(\Omega'';\Omega')=(2\pi)^{-1}\langle\Omega'',\tfrac{1}{2}{}_{1/2}|\Omega',\tfrac{1}{2}{}_{1/2}\rangle_{\text{E.C.}}. \tag{4.10}$$

It follows that the N-fold convolution of $\overline{R}$ with itself,

$$\overline{R}_{(N)}(\Omega'';\Omega')\equiv\int\overline{R}(\Omega'';\Omega_N)\cdots\times\overline{R}(\Omega_2;\Omega_1)\overline{R}(\Omega_1;\Omega')\,d\Omega_N\cdots d\Omega_2\,d\Omega, \tag{4.11}$$

is given by the expression

$$\overline{R}_{(N)}(\Omega'';\Omega')=(4\pi)\sum_{l=0}^{\infty}(2l+2)\overline{r}^{(N+1)}_{l+1/2}\times\langle l+\tfrac{1}{2}{}_{1/2}|U^{\dagger}[\Omega'']U[\Omega']|l+\tfrac{1}{2}{}_{1/2}\rangle, \tag{4.12}$$

which, since $\overline{r}_{1/2}\equiv 1$, is still in the equivalence class of the spin-1/2 unit operator. We are again interested in the limit $N\rightarrow\infty$ subject to the choice of $\alpha=2(N+1)/\nu T=2/\nu\epsilon$. For large α it follows from (3.23) and (4.6) that

$$\tilde{r}_{l+1/2}=1-[l(l+1)+(l+1)(l+2)]/2\alpha, \tag{4.13}$$

and in particular that

$$\tilde{r}_{1/2}=1-1/\alpha. \tag{4.14}$$

Consequently, we find that

$$\overline{R}_T(\Omega'';\Omega')\equiv\lim_{N\rightarrow\infty}\overline{R}_{(N)}(\Omega'';\Omega') =(4\pi)^{-1}\sum_{l=0}^{\infty}(2l+2)e^{-1/2l(l+2)\nu T}\times\langle l+\tfrac{1}{2}{}_{1/2}|U^{\dagger}[\Omega'']U[\Omega']|l+\tfrac{1}{2}{}_{1/2}\rangle, \tag{4.15}$$

which is still an element of the equivalence class of the spin-1/2 unit operator. Let us now seek a path-integral representation of this expression.

By definition

$$\overline{R}_T(\Omega'';\Omega')=\lim_{N\rightarrow\infty}\tilde{r}_{1/2}^{-(N+1)}\int\cdots\int\times\prod_{k=0}^{N}\langle\Omega_{k+1},\tfrac{1}{2}{}_{1/2}|\Omega_k,\tfrac{1}{2}{}_{1/2}\rangle\times R(\Omega_{k+1};\Omega_k)\prod_{k=1}^{N}d\Omega_k, \tag{4.16}$$

where $\Omega''\equiv\Omega_{N+1}$ and $\Omega'\equiv\Omega_0$. In order to evaluate this limit we first observe that by themselves the R factors along with the $d\Omega$ factors lead to the Wiener measure on the sphere. The rest of the expression leads to some weighting of this measure. Clearly the initial factor

$$\tilde{r}_{1/2}^{-(N+1)}\rightarrow e^{\nu T/2}. \tag{4.17}$$

For the remaining factor we first note for general s_m (which we momentarily suppress) and for continuous Wiener paths that[15]

$$\prod_{k=0}^{N}\langle\Omega_{k+1}|\Omega_k\rangle \rightarrow\exp\Big[-\int\langle\Omega(t)|d\Omega(t)\rangle -\tfrac{1}{2}\int\langle d\Omega(t)|(1-|\Omega(t)\rangle\langle\Omega(t)|)|d\Omega(t)\rangle\Big], \tag{4.18}$$

where $|d\Omega(t)\rangle\equiv d\,|\Omega(t)\rangle$. When $m=s$ it follows, for the problem at hand,[16] that this expression becomes

$$\exp\Big\{is\int\cos\theta(t)\,d\varphi(t) -\frac{1}{4}s\int[d\theta^2(t)+\sin^2\theta(t)\,d\phi(t)]\Big\}. \tag{4.19}$$

As a consequence of Itô's law (3.44) we learn that, for general s, this factor in the integrand becomes

$$\exp\Big(-s\nu T/2+is\int\cos\theta\,d\varphi\Big), \tag{4.20}$$

where the remaining integral represents a well-defined stochastic integral.

Combining the relevant factors in the case $s=1/2$ we learn that

$$\overline{R}_T(\Omega'';\Omega') =e^{\nu T/4}\int\exp\Big[i1/2\int\cos\theta(t)\,d\varphi(t)\Big]d\mu^T_W(\theta,\varphi) \tag{4.21}$$

provides a path integral representation for an element in the equivalence class of the spin-1/2 unit operator.

Although this formula holds for all T, the part of interest to us—the element of the equivalence class of the spin-1/2 unit operator—is *independent* of T. Were it not for the prefactor it is apparent that as T grows large the entire integral would decrease. Thus the need for an increasing prefactor is apparent, and as a semigroup representation even its structure is fixed, only the specific factor $(\nu/4)$ remaining undetermined. In this way we can understand the need for a prefactor here and in subsequent relations as well.

Spin 1

Let us now outline the analogous calculation for a spin-1 case. Thus we now set

$$\tilde{R}(\Omega'';\Omega')=\langle\Omega'',1_1|\Omega',1_1\rangle R(\Omega'';\Omega') =(4\pi)^{-1}\sum_{l=1}^{\infty}(2l+1)\tilde{r}_l\times\langle l_1|U^{\dagger}[\Omega'']U[\Omega']|l_1\rangle, \tag{4.22}$$

where the sum starts at $l=1$ since there is no spin-zero state with $m=1$. The coefficients $\tilde{r}_l$ are again given by Clebsch-Gordan coefficients, and it follows for $l\geqslant 1$ that

$$\tilde{r}_l=\tfrac{1}{2}\{r_l+(2l+1)^{-1}[(l+1)r_{l-1}+lr_{l+1}]\}. \tag{4.23}$$

We renormalize the coefficients according to

$$\overline{r}_l\equiv\tilde{r}_l/\tilde{r}_1, \tag{4.24}$$

and let

$$\bar R(\Omega'';\Omega')\equiv(4\pi)^{-1}\sum_{l=1}^{\infty}(2l+1)\bar r_l \times\langle l_1|U^\dagger[\Omega'']U[\Omega']|l_1\rangle, \tag{4.25}$$

which is then in the same equivalance class as

$$(4\pi)^{-1}3\langle l_1|U^\dagger[\Omega'']U[\Omega']|l_1\rangle = (3/4\pi)\langle\Omega'',l_1|\Omega',l_1\rangle \tag{4.26}$$

appropriate to the spin-1 unit operator. The N-fold convolution of this $\bar R$ leads to

$$\bar R_{(N)}(\Omega'';\Omega')=(4\pi)^{-1}\sum_{l=1}^{\infty}(2l+1)\bar r_l^{(N+1)} \times\langle l_1|U^\dagger[\Omega'']U[\Omega']|l_1\rangle, \tag{4.27}$$

and again we take the limit as $N\to\infty$. For large α it follows that

$$\bar r_1 = 1-2/\alpha,$$
$$\bar r_l = 1+[2-l(l+1)]/\alpha, \tag{4.28}$$

and thus in the present case

$$\bar R_T(\Omega'';\Omega')\equiv\lim_{N\to\infty}\bar R_{(N)}(\Omega'';\Omega') = (4\pi)^{-1}\sum_{l=1}^{\infty}(2l+1)e^{(1/2)(l-1)(l+2)\nu T} \times\langle l_1|U^\dagger[\Omega'']U[\Omega']|l_1\rangle, \tag{4.29}$$

which is still a function in the desired equivalence class.

For the path-integral representation of (4.29) we need to consider the result of

$$\prod_{k=0}^{N}\bar r_1^{-1}\langle\Omega_{k+1},1_1|\Omega_k,1_1\rangle, \tag{4.30}$$

which according to the remarks made earlier contributes in the limit $N\to\infty$ the factor

$$e^{\nu T}\exp\Big(-\nu T/2+i\int\cos\theta\,d\varphi\Big) \tag{4.31}$$

to the integrand. Consequently we find for the spin-1 case that

$$\bar R_T(\Omega'';\Omega') = e^{\nu T/2}\int\exp\Big[i\int\cos\theta(t)\,d\varphi(t)\Big]d\mu_W^T(\theta,\varphi) \tag{4.32}$$

provides a path-integral representation for an element in the equivalence class of the spin-1 unit operator.

Limiting form

We should observe that in each of the indicated cases the deviation of the path-integral expression from the true expression is at least of order $e^{-\nu T}$. Consequently the limit $\nu\to\infty$ will yield just the leading term, which is the desired integral kernel. In particular, for spin 1/2 for example, we have the relation

$$(2\pi)^{-1}\langle\theta'',\varphi'',\tfrac{1}{2}{}_{1/2}|\theta',\varphi',\tfrac{1}{2}{}_{1/2}\rangle = \lim_{\nu\to\infty}e^{\nu T/4}\int\exp\Big[i\frac{1}{2}\int\cos\theta(t)\,d\varphi(t)\Big]d\mu_W^T(\theta,\varphi). \tag{4.33}$$

Of course it should be recalled that the Wiener measure depends on ν as well. The intuitive reason for the validity of this expression is perhaps best seen from the formal expression (3.43); for if we formally take the limit $\nu\to\infty$, then according to (3.43) it follows that

$$\langle\theta'',\varphi'',\tfrac{1}{2}{}_{1/2}|\theta',\varphi',\tfrac{1}{2}{}_{1/2}\rangle = \mathscr{N}\int\exp\Big(i\frac{1}{2}\int\cos\theta\,\dot\varphi\,dt\Big)\prod_t d\Omega/(t), \tag{4.34}$$

which is just the usual formal path-integral expression for the spin-1/2 propagator for vanishing dynamics, where $\mathscr{N}$ is a formal normalization constant.[1,5] We remark in addition that the leading correction to this expression which is suggested by (4.33) has been useful in interpreting the stationary-phase approximation to the formally defined path-integral expression for spin-coherent states.[16] Here we obtain confirmation of that procedure on the basis of the general approach adopted in this paper.

B. Nonvanishing Hamiltonians

For the introduction of dynamics let us concentrate on the spin-1/2 case and initially take as a Hamiltonian

$$\mathscr{H}=\mathscr{H}(t)=\gamma(t)S_3, \tag{4.35}$$

where γ is a suitably smooth function. In a certain sense there is no loss of generality (when γ is constant) in assuming $\mathscr{H}=\gamma S_3$ since in a two-dimensional space any Hermitian operator when diagonalized can be written as a multiple of S_3 and the identity operator (I_2). At any rate, this is the simplest example to treat; other examples are treated later.

If we set

$$\Gamma=\int_0^T\gamma(t)\,dt, \tag{4.36}$$

then it follows for spin 1/2 that [cf. (4.15)]

$$\bar R_T^\Gamma(\Omega'';\Omega') = (4\pi)^{-1}\sum_{l=0}^{\infty}(2l+2)e^{-(1/2)l(l+2)\nu T} \times\langle l+\tfrac{1}{2}{}_{1/2}|U^\dagger[\Omega'']e^{-i\Gamma S_3}U[\Omega']|l+\tfrac{1}{2}{}_{1/2}\rangle \tag{4.37}$$
$$= (4\pi)^{-1}\sum_{l=0}^{\infty}(2l+2)e^{-(1/2)l(l+2)\nu T} \times\langle l+\tfrac{1}{2}{}_{1/2}|U^\dagger[\theta'',\varphi'']U[\theta',\varphi'+\Gamma]|l+\tfrac{1}{2}{}_{1/2}\rangle$$

provides a function which is in the equivalence class of the evolution operator. To determine a path-integral representation of this expression we return to (4.16) and simply add $\gamma_k\epsilon$, where $\gamma_k\equiv\gamma(k\epsilon)$, to the appropriate φ_k factors (those in kets) in the integrand. Effectively only two changes take place, one in the spin-1/2 factors and one in the R factors. In the former we find a change in the phase factor to

$$\exp\Big(i\frac{1}{2}\int\cos\theta\,d\varphi-i\frac{1}{2}\int\gamma\cos\theta\,dt\Big), \tag{4.38}$$

which for $s>1/2$ has an obvious generalization according to (4.20).

The change introduced by the R factors is somewhat more complicated. To find this change observe, for large N and in the presence of γ, that

$$R^{\Gamma}_{(N)}(\Omega'';\Omega') = \int\cdots\int\left(\frac{1}{2\pi\nu\epsilon}\right)^{(N+1)} \times\exp\left[\frac{1}{\nu\epsilon}\sum_{k=0}^{N}(\cos\beta'_k - 1)\right]\prod_{k=1}^{N} d\Omega_k, \quad (4.39)$$

where

$$\cos\beta^1_k = \cos\theta_{k+1}\cos\theta_k + \sin\theta_{k+1}\sin\theta_k\cos(\varphi_{k+1} - \varphi_k - \gamma_k\epsilon). \quad (4.40)$$

According to (3.40) we know how to describe the limit as $N\to\infty$ when $\gamma\equiv 0$, so let us concentrate on the additional γ-dependent terms. Since

$$\cos(\varphi_{k+1} - \varphi_k - \gamma_k\epsilon) = \cos(\varphi_{k+1} - \varphi_k)\cos(\gamma_k\epsilon) + \sin(\varphi_{k+1} - \varphi_k)\sin(\gamma_k\epsilon), \quad (4.41)$$

the difference in the exponent from the vanishing-γ case is given by

$$\frac{1}{\nu\epsilon}\sum_{k=0}^{N}(\sin\theta_{k+1}\sin\theta_k\{\cos(\varphi_{k+1} - \varphi_k)[\cos(\gamma_k\epsilon) - 1] + \sin(\varphi_{k+1} - \varphi_k)\sin(\gamma_k\epsilon)\}). \quad (4.42)$$

In the limit $N\to\infty$ the only nonvanishing contributions that survive for continuous Wiener paths on the sphere are those that arise from expanding $\cos(\gamma_k\epsilon)$ to second order in ϵ, and from expanding $\sin(\varphi_{k-1} - \varphi_k)$ and $\sin(\gamma_k\epsilon)$ each to first order. As a consequence, in the appropriate limit, (4.42) leads to

$$\frac{1}{\nu}\int\gamma(t)\sin^2\theta(t)\,d\varphi(t) - \frac{1}{2\nu}\int\gamma^2(t)\sin^2\theta(t)\,dt, \quad (4.43)$$

which evidently are well-defined stochastic integrals. Thus the final form for the path-integral expression of a function in the equivalence class of the evolution operator is given by

$$(2\pi)^{-1}\langle\Omega'',\tfrac{1}{2}1/2|\exp\left[-i\int_0^T\gamma(t)S_3\,dt\right]|\Omega',\tfrac{1}{2}1/2\rangle_{\text{E.C.}} = e^{\nu T/4}\int\exp\left[i\frac{1}{2}\int_0^T(\cos\theta\,d\varphi - \gamma\cos\theta\,dt)\right] \times\exp\left[\frac{1}{\nu}\int_0^T(\gamma\sin^2\theta\,d\varphi - \tfrac{1}{2}\gamma^2\sin^2\theta\,dt)\right]d\mu^T_W(\theta,\varphi). \quad (4.44)$$

Expressed in this form it is clear that this representation holds for all γ that are locally square integrable. To describe a similar expression for spin 1 it is only necessary to double the phase and double the exponent in the prefactor ($e^{\nu T/4}$).

Several remarks regarding (4.44) are in order. We have already commented on the need and form of the prefactor. Observe next that the phase factor is just the classical action,[15]

$$I = \frac{1}{2}\int[\cos\theta\,\dot\varphi - \gamma\cos\theta]\,dt \quad (4.45)$$

for a driven spin-1/2 variable. In terms of

$$p \equiv \tfrac{1}{2}\cos\theta, \qquad H \equiv \tfrac{1}{2}\gamma\cos\theta = \gamma p, \quad (4.46)$$

this relation reads

$$I = \int[p\dot\varphi - H]\,dt, \quad (4.47)$$

and implies the classical Hamiltonian equations of motion

$$\dot p = -\tfrac{1}{2}\sin\theta\,\dot\theta = -\partial H/\partial\varphi = 0, \qquad \dot\varphi = \frac{\partial H}{\partial p} = \gamma. \quad (4.48)$$

Finally let us examine the measure $\mu^{T,\gamma}$, where

$$d\mu^{T,\gamma}(\theta,\varphi) \equiv \exp\left[\frac{1}{\nu}\int\gamma\sin^2\theta\,d\varphi - \frac{1}{2\nu}\int\gamma^2\sin^2\theta\,dt\right]d\mu^T_W(\theta,\varphi). \quad (4.49)$$

First, we observe that (4.39) implies that

$$1 = \lim_{N\to\infty}\int R^{\Gamma}_{(N)}(\Omega'';\Omega')\,d\Omega'' = \int d\mu^{T,\gamma}(\theta,\varphi)\,d\Omega'', \quad (4.50)$$

and thus like μ^T_W we can regard $\mu^{T,\gamma}$ as a probability measure when the final integral is also included. This formula may also be stated as

$$E\left(\exp\left[\frac{1}{\nu}\int\gamma\sin^2\theta\,d\varphi - \frac{1}{2\nu}\int\gamma^2\sin^2\theta\,dt\right]\right) = 1, \quad (4.51)$$

which is then seen as a special case of the general Prokhorov formula (3.45) when $B = 0$ and $C = \gamma\sin\theta$. It is important to appreciate that, in the above sense, the measure $\mu^{T,\gamma}$ only *redistributes* the weight of the Wiener paths *without* affecting the total weight, which remains unchanged. In probability language we may say that the stochastic variable

$$\exp\left[\frac{1}{\nu}\int\gamma\sin^2\theta\,d\varphi - \frac{1}{2\nu}\int\gamma^2\sin^2\theta\,dt\right] \quad (4.52)$$

is a martingale.

It is also instructive to understand this result regarding $\mu^{T,\gamma}$ in a formal manner, and we observe, in the same sense as (3.43), that

$$\int d\mu^{T,\gamma}(\theta,\varphi) = \mathcal{N}\int\exp\left[\frac{1}{\nu}\int\gamma\sin^2\theta\,d\varphi - \frac{1}{2\nu}\int\gamma^2\sin^2\theta\,dt\right] \times\exp\left[-\frac{1}{2\nu}\int(\dot\theta^2 + \sin^2\theta\,\dot\varphi^2)\,dt\right]\prod_t d\Omega(t) = \mathcal{N}\int\exp\left\{-\frac{1}{2\nu}\int \times[\dot\theta^2 + \sin^2\theta(\dot\varphi - \gamma)^2]\,dt\right\}\prod_t d\Omega(t). \quad (4.53)$$

We see then, in this case, that the proper result for this part of the integrand is obtained by replacing $\dot\theta$ and $\dot\varphi$ by the relevant classical equations of motion, $\dot\theta$ and $\varphi - \gamma$, in the presence of the classical Hamiltonian $H = \tfrac{1}{2}\gamma\cos\theta$. Similar conclusions are found in the case of canonical variables.[3]

Other Hamiltonians

The particular Hamiltonian S_3 is especially simple because it is an element of the Lie algebra. Equally simple expressions should exist for the other elements of the Lie algebra, e.g., S_1 and S_2. However, since we are only working with two of the three variables of the rotation group, we can expect a somewhat different relation for S_1 and S_2 than for S_3. To discuss the general case where

$$\mathscr{H} = \mathscr{H}(t) = \alpha(t)S_1 + \beta(t)S_2 + \gamma(t)S_3 \tag{4.54}$$

we note, for $\alpha_k \equiv \alpha(k\epsilon)$, etc., that

$$\begin{aligned} &e^{-i\epsilon(\alpha_k S_1 + \beta_k S_2 + \gamma_k S_3)} e^{-i\varphi_k S_3} e^{-i\theta_k S_2} |s_m\rangle \\ &\quad = e^{-i[\varphi_k + \epsilon\gamma_k - \epsilon(\alpha_k \cos\varphi_k + \beta_k \sin\varphi_k)\cot\theta_k] S_3} \\ &\quad \times e^{-i[\theta_k + \epsilon(\beta_k \cos\varphi_k - \alpha_k \sin\varphi_k)] S_2} |s_m\rangle \\ &\quad \times e^{-im\epsilon(\alpha_k \cos\varphi_k + \beta_k \sin\varphi_k)\csc\theta_k} \end{aligned} \tag{4.55}$$

valid to first order in ϵ, which up to a phase factor is again a spin-coherent state. As a consequence, and based on our previous results, it is not difficult to establish the path-integral expression for spin 1/2 (or 1) for an element in the equivalence class of this evolution operator. For convenience we indicate this expression formally as

$$\begin{aligned} &(2\pi)^{-1}\langle \theta'',\varphi'',\tfrac{1}{2}{}_{1/2}| T\exp\Big\{ -i\int_0^T [\alpha(t)S_1 + \beta(t)S_2 + \gamma(t)S_3]\, dt \Big\} |\theta',\varphi',\tfrac{1}{2}{}_{1/2}\rangle_{\text{E.C.}} \\ &\quad = \mathscr{N}\int \exp\Big\{ i\frac{1}{2}\int [\cos\theta\,\dot\phi - (\alpha\cos\varphi + \beta\sin\varphi)\sin\theta - \gamma\cos\theta]\, dt \Big\} \\ &\quad\quad \times \exp\Big(-\frac{1}{2\nu}\int \{ [\dot\theta + \alpha\sin\varphi - \beta\cos\varphi]^2 + \sin^2\theta\,[\dot\varphi + (\alpha\cos\varphi + \beta\sin\varphi)\cot\theta - \gamma]^2 \}\, dt \Big) \prod_t d\Omega(t), \end{aligned} \tag{4.56}$$

where T denotes the time-ordering operator. Moreover, exactly as before, it follows that the drift terms in the Wiener measure are restatements of the classical equations of motion as determined here from the classical Hamiltonian

$$\begin{aligned} H &\equiv \langle \theta,\varphi,\tfrac{1}{2}{}_{1/2}|\mathscr{H}|\theta,\varphi,\tfrac{1}{2}{}_{1/2}\rangle \\ &= \alpha\sin\theta\cos\varphi + \beta\sin\theta\sin\varphi + \gamma\cos\theta. \end{aligned} \tag{4.57}$$

We may note as before that the limit $\nu \to \infty$ yields the appropriate matrix element, i.e., as in (4.56) but without "E.C."

It is also natural to consider Hamiltonians that are not simply elements of the Lie algebra. Unfortunately, at present, our methods are unable to generate path-integral expressions for general Hamiltonias. One could imagine integrating a formula such as (4.56) over the variables α, β, and γ with some suitable weight factor. The principal roadblock in instrumenting this proposal is the need for α, β, and γ to be locally square integrable. This precludes any rigorous integration to generate a Hamiltonian formulation, i.e., a Markov dynamics. However, we may expect that we can couple the parameter controlling the necessary nonlocality in time with the parameter ν in such a way that in the limit $\nu \to \infty$ not only does the desired matrix element emerge directly but also the nonlocalities of the dynamics disappear at the same time. The question of how best to determine genuine path-integral expressions for general Hamiltonians is clearly a problem of some interest and one that deserves further attention.

ACKNOWLEDGMENTS

It is a pleasure to thank M. Sirugue for a helpful remark, and H. Grosse for a stimulating discussion.

[1]See, e.g., J. R. Klauder, Acta Phys. Austriaca, Suppl. XXII, 3 (1980).

[2]See, e.g., B. Simon, *Functional Integration and Quantum Physics* (Academic, New York, 1979).

[3]J. R. Klauder and I. Daubechies, Phys. Rev. Letters **48**, 117 (1982).

[4]I. Daubechies and J. R. Klauder, "Constructing Measures for Path Integrals," J. Math. Phys. **23**, 1806 (1982).

[5]J. R. Klauder, Ann. Phys. **11**, 123 (1960).

[6]J. R. Klauder, "Extension of the Itô Calculus" (in preparation).

[7]There are many excellent sources for properties of SU(2) among which are: M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957); I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Representations* (Pergamon, New York, 1977); M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Addison-Wesley, Reading, MA, 1962); E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).

[8]See, e.g., J. R. Klauder, J. Math. Phys. **4**, 1058 (1963). Compare J. M. Radcliffe, J. Phys. A **4**, 313 (1971) and F. T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas, Phys. Rev. A **6**, 2211 (1972) for related states (making allowance for a different phase convention).

[9]The orthogonality relations of (2.7) are a subset of those of the group SU(2). To discuss the full set requires the use of all three Euler angles so that

$$U[\theta,\varphi,\psi]|s_m\rangle = U[\theta,\varphi]|s_m\rangle e^{-im\psi},$$

where ψ is an auxiliary angle such that $0 < \psi < 4\pi$ (double covering). Thus integration over ψ ensures orthogonality (in a rather trivial way) whenever $m \neq m'$. When $m = m'$, as in (2.7), the variable ψ drops out and can be ignored as we do.

[10]Strictly speaking we should include the third variable ψ (see Ref. 9) of the group to allow for a functional statement of left group multiplication. We only use the group invariance implicitly, and content ourselves with the variables θ and φ.

[11]See, e.g., M. A. Neumark, *Normierte Algebren* (VEB Deutscher Verlag der Wissenschaften, 1959).

[12]It is not difficult to see that (3.40) represents just the angular part of the usual defining measure for a three-dimensional Brownian motion, a property which strongly suggests that it leads to Wiener measure on the sphere.

[13]See, e.g., K. Itô in *Mathematical Problems in Theoretical Physics* (Springer New York, 1975), p. 218.

[14]Yu. V. Prokhorov, Teor. Verojatnost. i Primenen. **1**, 77 (1956) [Theory Probab. Appl. **1**, 157 (1956)].

[15]J. R. Klauder, in *Path Integrals*, edited by G. J. Papadopoulos and J. T. Devreese (Plenum, New York, 1978), p. 5.

[16]J. R. Klauder, Phys. Rev. D **19**, 2349 (1979).

Constructing measures for path integrals

Ingrid Daubechies[a),b)]
Department of Physics, Princeton University, Princeton, New Jersey 08544
John R. Klauder
Bell Laboratories, Murray Hill, New Jersey 07974

(Received 15 December 1981; accepted for publication 26 March 1982)

The overcompleteness of the coherent states for the Heisenberg–Weyl group implies that many different integral kernels can be used to represent the same operator. Within such an equivalence class we construct an integral kernel to represent the quantum-mechanical evolution operator for certain dynamical systems in the form of a path integral that involves genuine (Wiener) measures on continuous phase-space paths. To achieve this goal it is necessary to employ an expression for the classical action different from the usual one.

PACS numbers: 03.65.Db

I. INTRODUCTION

As usually formulated, quantum mechanical path integrals are physically elegant but unfortunately are mathematically inelegant as well. The apparently closed form of solution path integrals provide to many problems is tempered by the ambiguities inherent in giving the path integral a meaningful definition, and this aspect has been carefully documented.[1] There have been several attempts to introduce genuine measures and thereby restore order in path-integral formulations. In the works of Albeverio and Høegh-Kröhn[2] and of Combe *et al.*,[3] for example, effort is concentrated on multiplicative potentials which have the property that their Fourier transform is a bounded measure. While this limitation leads to well-defined path integrals the measures involved are Poisson measures for which the paths are not continuous but rather entail discontinuities. In addition this limited class of potentials does not include the harmonic oscillator which, to be incorporated, must be dealt with in an alternative fashion.

In this paper we present a detailed analysis of a quantum mechanical path integral formulation that involves genuine (Wiener) measures concentrated on continuous paths, which deals in a natural way with harmonic-oscillator potentials; a summary of our principal results has already appeared in Ref. 4. We are able to handle directly an essentially arbitrary quadratic Hamiltonian of the harmonic-oscillator type involving quite general time-dependent coefficients, all with one and the same Wiener measure. Superpositions over the time-dependent coefficients significantly widen the class of systems we are able to consider.

Our approach and analysis is based on coherent states and their special properties, and differs considerably from the viewpoint adopted in Ref. 2 or Ref. 3. Before undertaking our detailed analysis we sketch the general mathematical setting of our approach.

A. Consequences of coherent-state overcompleteness

Coherent states are conventionally defined in an abstract Hilbert space $\mathscr{H}$ by

$$|p{\cdot}q\rangle \equiv e^{i(pQ-qP)}|0\rangle \tag{1}$$

for all real p and q, where Q and P are an irreducible Heisenberg pair, and $|0\rangle$ denotes the normalized solution of the equation $(Q+iP)|0\rangle = 0$.[5] These states admit the fundamental resolution of unity

$$\mathbb{1} = \int |p,q\rangle\langle p,q|(dp\,dq/2\pi) \tag{2}$$

when integrated over all phase space. As a consequence we may conveniently represent the vectors of the abstract Hilbert space by bounded, continuous functions

$$\psi(p,q) \equiv \langle p,q|\psi\rangle, \tag{3}$$

with an inner product given by

$$\langle\phi|\psi\rangle = \int \phi^*(p,q)\psi(p,q)(dp\,dq/2\pi). \tag{4}$$

If $|\phi\rangle = |p',q'\rangle$ it follows that each function $\psi(p,q)$ satisfies the identity

$$\psi(p',q') = \int \mathscr{K}(p',q';p,q)\psi(p,q)(dp\,dq/2\pi), \tag{5}$$

where

$$\mathscr{K}(p',q';p,q) \equiv \langle p',q'|p,q\rangle = \exp\{\tfrac{1}{2}i(pq'-qp') - \tfrac{1}{4}[(p'-p)^2 + (q'-q)^2]\} \tag{6}$$

plays the role of a reproducing kernel. Thus the set of functions of the form (3) with the inner product (4) comprise a reproducing-kernel Hilbert space $\mathscr{C}_0$.[6] The reproducing kernel projects out a closed subspace of the space $L^2(\mathbb{R}^2)$ of all square-integrable functions, and there remain infinitely many linearly independent square-integrable functions orthogonal to all elements of $\mathscr{C}_0$. This feature has important consequences for the representation of operators on $\mathscr{C}_0$ by integral kernels.

[a)] On leave from Dienst voor Theoretische Natuurkunde, Vrije Universiteit Brussels, Belgium.
[b)] Scientific collaborator at the Interuniversitair Instituut voor Kernwetenschappen (Interuniversitary Institute for Nuclear Sciences), Belgium.

 0022-2488/82/101806-17$02.50

Consider the expression

$$\langle \phi | B | \psi \rangle = \int \phi^*(p'',q'') K_B(p'',q'';r',q') \times \psi(p',q')(dp''dq''/2\pi)(dp'dq'/2\pi) \quad (7)$$

for arbitrary vectors $|\phi\rangle$ and $|\psi\rangle$, and an arbitrary but fixed bounded operator B. One integral kernel that satisfies (7) is always given by

$$K_B(p'',q'';p',q') = \langle p'',q''|B|p',q'\rangle, \quad (8)$$

but in view of the foregoing remarks there are infinitely many other kernels that serve equally well to represent the operator B. As an example we note that all kernels of the form

$$F_\lambda(p'',q'';p',q') = \tfrac{1}{2}(1+4\lambda) \times \exp\{\tfrac{1}{2}i(p'q''-q'p'') - \lambda[(p''-p')^2 + (q''-q')^2]\}, \quad (9)$$

where $\lambda > -\frac{1}{8}$ serve to represent the unit operator, even including the limiting distribution as $\lambda \to \infty$,

$$F_\lambda(p'',q'';p',q') \to 2\pi\delta(p''-p')\delta(q''-q'), \quad (10)$$

which also serves the same purpose.

All kernels that satisfy (7) for a given B form an equivalence class labeled by the operator B, and which we shall denote by $\mathscr{C}(B)$. Thus the examples F_λ in (9) and (10) all belong to the equivalence class $\mathscr{C}(\mathbb{1})$. A generic element of $\mathscr{C}(B)$ is conveniently denoted by $\langle p'',q''|B|p',q'\rangle_{\text{E.C.}}$ (where E. C. represents equivalence class). Any such kernel can serve to represent the operator B in the context of (7), or stated otherwise, in the form

$$B = \int |p'',q''\rangle\langle p'',q''|B|p',q'\rangle_{\text{E.C.}}\langle p',q'| \times (dp''dq''/2\pi)(dp'dq'/2\pi). \quad (11)$$

It is by exploiting this freedom of representation that we shall achieve our goal of representing the quantum mechanical propagator by means of a path integral involving genuine (Wiener) measures.

In the next section, Sec. 2, we detail the construction of the path integral for a special class of dynamical systems, following closely but with significant differences, the usual method of construction. In Sec. 3 we evaluate the path integrals constructed in Sec. 2, while in Sec. 4 we prove that each of the evaluated path integrals is indeed an element of the equivalence class (in the sense described above) of the evolution operator for the particular Hamiltonian in question. A brief conclusion follows in Sec. 5, and the Appendix contains some details needed for Sec. 3.

2. CONSTRUCTION OF THE PATH INTEGRAL

We start by recalling some more properties of the coherent states and the Weyl operators.

A. Basic properties and notations

We take $\mathscr{H}$ to be a separable Hilbert space, on which we define the Weyl operators $W(p,q)$ as

$$W(p,q) = \exp[i(pQ - qP)], \quad (12)$$

where P,Q are an irreducible Heisenberg pair on $\mathscr{H}$, chosen in such a way that

$$W(p',q')W(p'',q'') = \exp[\tfrac{1}{2}i(p'q''-p''q')]W(p'+p'',q'+q''). \quad (13)$$

The operators $W(p,q)$ then act on $\mathscr{H}$ in an irreducible way. Additional properties of the $W(p,q)$ are

$$W(p,q)^\dagger = W(-p,-q) \quad (14)$$

and, for any operator formally written as $F(P,Q)$,

$$W(p,q)^\dagger F(P,Q) W(p,q) = F(P+p,Q+q). \quad (15)$$

We shall use the fact that any (bounded) operator is completely characterized by its diagonal matrix elements between coherent states

$$B \in \mathscr{B}(\mathscr{H}), \forall p,q: \langle p,q|B|p,q\rangle = 0 \Leftrightarrow B \equiv 0, \quad (16)$$

where the coherent states (c.s) are defined as (see Sec. 1)

$$|p,q\rangle = W(p,q)|0\rangle.$$

One can also use diagonal matrix elements between coherent states to evaluate traces.

$$A \text{ trace-class} \Rightarrow \operatorname{Tr} A = \int \frac{dp\,dq}{2\pi}\langle p,q|A|p,q\rangle. \quad (17)$$

Using the product rule (13) for the Weyl operators, one may show that (4) can be rewritten in the following form:

$$\forall \phi,\psi \in \mathscr{H}: \int \frac{dp\,dq}{2\pi} W(p,q)|\phi\rangle\langle\psi|W(p,q)^\dagger = \langle\psi|\phi\rangle \mathbb{1}_{\mathscr{H}} \quad (18)$$

[the easiest way to verify (18) is to check that the diagonal matrix elements between c.s. (coherent states) of the two sides are the same].

Defining $|n\rangle$, $n = 0,1,2,\ldots$, to be the $(n+1)$th normalized eigenstate of $\frac{1}{2}(P^2+Q^2-1)$ (which is consistent with the definition of $|0\rangle$ in the Introduction), we have in particular

$$\int \frac{dp\,dq}{2\pi} W(p,q)|n\rangle\langle m|W(p,q)^\dagger = \delta_{mn}\mathbb{1}_{\mathscr{H}}. \quad (19)$$

The usual technique in the construction of a c.s. path integral for an evolution operator U_t is to reexpress the evolution operator as a product $U_t = (U_{t/n})^n$, to insert the resolution of the identity (2) between each two factors, and to take the limit as $n \to \infty$ (see Ref. 1)

$$\langle p'',q''|U_t|p',q'\rangle = \lim_{n\to\infty}\int \frac{dp_{n-1}dq_{n-1}}{2\pi} \cdots \int \frac{dp_1 dq_1}{2\pi} \times \prod_{j=0}^{n-1}\langle p_{j+1},q_{j+1}|U_{t/n}|p_j,q_j\rangle, \quad (20)$$

$$p_n = p'',\ q_n = q'',$$
$$p_0 = p',\ q_0 = q'.$$

For a time-ordered product $T\exp[-i\int_{t'}^{t''}H(t)\,dt]$, the same technique is used [put $\epsilon = (t''-t')/n$]

$$T\exp\left[-i\int_{t'}^{t''}H(t)\,dt\right] = \lim_{n\to\infty}\{\exp[-iH(t''-\epsilon)\epsilon] \times \exp[-iH(t''-2\epsilon)\epsilon]\cdots\exp[-iH(t')\epsilon]\},$$

which then implies that

$$\langle p'',q''|T\exp\left[-i\int_{t'}^{t''}H(t)dt\right]|p',q'\rangle$$
$$=\lim_{n\to\infty}\int\frac{dp_{n-1}dq_{n-1}}{2\pi}\cdots\int\frac{dp_1dq_1}{2\pi}\times\prod_{j=0}^{n-1}\langle p_{j+1},q_{j+1}|\exp[-iH(t'+j\epsilon)\epsilon]|p_j,q_j\rangle, \tag{21}$$
$$p_n=p'',q_n=q'';p_0=p',q_0=q'.$$

Basically we shall do the same here; however, instead of (2) we shall insert some more complicated object, and the results of our manipulations will no longer be the matrix elements $\langle p'',q''|U_{t'',t'}|p',q'\rangle$, but some other element of $\mathscr{C}(U_{t'',t'})$ {From now on, we shall use the symbol $U_{t'',t}$ to denote the evolution operator $T\exp[-i\int_{t'}^{t''}H(t)\,dt]$}.

B. The "big" space $\hat{\mathscr{H}}$ and the vectors $|p,q;\beta\rangle\rangle$

We define a "big" Hilbert space $\hat{\mathscr{H}}$ by

$$\hat{\mathscr{H}}=\bigoplus_{n=0}^{\infty}\mathscr{H}_n,$$

where each $\mathscr{H}_n$ is isomorphic with $\mathscr{H}$ (we shall not write out these isomorphisms explicitly, but shall always assume them tacitly understood). We define canonical projections P_m from $\hat{\mathscr{H}}$ to $\mathscr{H}$ as follows:

$$\forall\phi=\bigoplus_{n=0}^{\infty}\phi_n\in\hat{\mathscr{H}}:P_m\phi=\phi_m.$$

The conjugate operators to these P_m are the canonical injections I_m; these are the maps from $\mathscr{H}$ to $\hat{\mathscr{H}}$ defined as follows:

$$\forall\phi\in\mathscr{H}:I_m\phi=\bigoplus_{n=0}^{\infty}\psi_n,$$

where all but the mth ψ_n are zero:

$$\psi_n=\delta_{nm}\phi.$$

The following properties of the P_m, I_m are easy to check:

(a) $P_mI_n=\delta_{mn}\mathbb{1}_{\mathscr{H}}$,

(b) $I_mP_m=\bigoplus_{n=0}^{\infty}(\delta_{mn}\mathbb{1}_n)$

(this operator is zero on all the $\mathscr{H}_n$ with $n\neq m$, and $\mathbb{1}$ on $\mathscr{H}_m$; it is the orthogonal projection operator in $\hat{\mathscr{H}}$ with image $\{0\}\oplus\cdots\oplus\{0\}\oplus\mathscr{H}_m\oplus\{0\}\oplus\cdots$),

(c) $$\mathbb{1}_{\hat{\mathscr{H}}}=\bigoplus_{n=0}^{\infty}\mathbb{1}_n=\sum_m I_mP_m \tag{22}$$

(as a sum of mutually orthogonal projection operators, this sum is well defined in the strong topology),

(d) $I_m=P_m^\dagger,\quad P_m^\dagger=I_m$.

For any $\beta\in[0,1)$ we define a set of normalized vectors $|p,q;\beta\rangle\rangle$ in $\hat{\mathscr{H}}$ by the rule

$$|p,q;\beta\rangle\rangle\equiv(1-\beta)^{1/2}\sum_{n=0}^{\infty}\beta^{n/2}I_n(W(p,q)|n\rangle)=(1-\beta)^{1/2}\bigoplus_{n=0}^{\infty}[\beta^{n/2}W(p,q)|n\rangle]. \tag{23}$$

These vectors $|p,q;\beta\rangle\rangle$ have the following overlap function:

$$\langle\langle p'',q'';\beta|p',q';\beta\rangle\rangle=(1-\beta)\sum_n\beta^n\langle n|W(p'',q'')^\dagger W(p',q')|n\rangle=(1-\beta)\mathrm{Tr}[\beta^NW(p'',q'')^\dagger W(p',q')],$$

where $N=\frac{1}{2}(P^2+Q^2-1)$. For the evaluation of this expression we refer to the Appendix [see (A9)]; the result is

$$\langle\langle p'',q'';\beta|p',q';\beta\rangle\rangle=\exp\left\{\frac{i}{2}(p'q''-p''q')-\frac{1+\beta}{4(1-\beta)}[(p''-p')^2+(q''-q')^2]\right\}. \tag{24}$$

One can easily calculate the overlap function $\langle\langle p'',q'';0|p',q';\beta\rangle\rangle$ as

$$\langle\langle p'',q'';0|p',q'';\beta\rangle\rangle=(1-\beta)^{1/2}\langle0|W(p'',q'')^\dagger W(p',q')|0\rangle=(1-\beta)^{1/2}\langle p'',q''|p',q'\rangle. \tag{25}$$

Another property of the $|p,q;\beta\rangle\rangle$ is the following [use (19)]:

$$\int\frac{dp\,dq}{2\pi}|p,q;\beta\rangle\rangle\langle\langle p,q;\beta|=(1-\beta)\sum_{n,m}\beta^{(n+m)/2}I_n\times\int\frac{dp\,dq}{2\pi}W(p,q)|n\rangle\langle m|W(p,q)^\dagger P_m=(1-\beta)\sum_n\beta^nI_nP_n=(1-\beta)\bigoplus_n(\beta^n\mathbb{1}_n). \tag{26}$$

This operator is a multiple of the identity on each of the $\mathscr{H}_n$-spaces with, however, difficult constants on different spaces. We shall use this "generalized effective resolution of the identity" to replace (2) in the construction of (20) or (21). Note that (26) holds for any $\beta\in[0,1)$ [for $\beta=0$, it essentially gives (2) again], which allows us to adjust β when needed; this feature will turn out to be important in our construction of the path integrals below.

C. Construction of elements of $\mathscr{C}(B)$ for $B\in\mathscr{B}(\mathscr{H})$

Let us now see how (26) can be useful for our purposes.

Every bounded operator $\mathbf{B}$ on $\hat{\mathscr{H}}$ is completely characterized by the sequence $(B_{kl}=P_k\mathbf{B}I_l)_{k,l=0}$.

Using (26) twice, we obtain

$$\int\frac{dp''dq''}{2\pi}\int\frac{dp'dq'}{2\pi}|p'',q'';\beta\rangle\rangle\times\langle\langle p'',q'';\beta|\mathbf{B}|p',q';\beta\rangle\rangle\langle p',q'|p_1,q_1\rangle=(1-\beta)^2\sum_{n,m}\beta^{n+m}I_nB_{nm}P_m,$$

Sandwiching this between $\langle\langle p_2,q_2;0|$ and $|p_1,q_1;0\rangle\rangle$ and using (25), we find

$$\int\frac{dp''dq''}{2\pi}\int\frac{dp'dq'}{2\pi}\langle p_2,q_2|p'',q''\rangle\times\langle\langle p'',q'';\beta|\mathbf{B}|p',q';\beta\rangle\rangle\langle p',q'|p_1,q_1\rangle=(1-\beta)\langle p_2,q_2|B_{00}|p_1,q_1\rangle.$$

Hence

$$B_{00}=\frac{1}{1-\beta}\int\frac{dp''dq''}{2\pi}\int\frac{dp'dq'}{2\pi}|p'',q''\rangle \langle\langle p'',q'';\beta|\mathbf{B}|p',q';\beta\rangle\rangle\langle p',q'| \tag{27}$$

or, stated otherwise,

$$1/(1-\beta)\langle\langle p'',q'';\beta|\mathbf{B}|p',q';\beta\rangle\rangle\in\mathscr{C}(B_{00}).$$

What has happened in this construction is really a projection such as described in the Introduction: The matrix element $\langle\langle p'',q'';\beta|\mathbf{B}|p',q';\beta\rangle\rangle$ is a sum of different matrix elements:

$$\langle\langle p'',q'';\beta|\mathbf{B}|p',q';\beta\rangle\rangle$$
$$=(1-\beta)\sum_{n,m}\beta^{(n+m)/2}\langle n|W(p'',q'')^\dagger B_{nm}W(p',q')|m\rangle.$$

By virtue of (19) all matrix elements with $n\neq 0$ and/or $m\neq 0$ give no contribution whatever when the projection is carried out:

$$n\neq 0 \quad\text{or}\quad m\neq 0$$
$$\Rightarrow\forall\phi,\psi\in\mathscr{H}:\int\frac{dp''dq''}{2\pi}\int\frac{dp'dq'}{2\pi}$$
$$\times\langle\phi|W(p'',q'')|0\rangle\langle n|W(p'',q'')^\dagger B_{nm}W(p',q')|m\rangle$$
$$\times\langle 0|W(p',q')^\dagger|\psi\rangle=0.$$

Therefore, all these terms drop out when the projection is performed, and the only relevant term is the B_{00} term.

To apply the same argument to a product of operators $\mathbf{B}_n\cdots\mathbf{B}_1$, we must restrict ourselves to diagonal operators. An operator $\mathbf{B}\in\mathscr{B}(\hat{\mathscr{H}})$ is called "diagonal" if $\forall k\neq l, B_{kl}=0$ (i.e., the operator $\mathbf{B}$ does not mix the different $\mathscr{H}_n$, $\mathbf{B}=\oplus_{n=0}^{\infty}B_{nn}$). Let $\mathbf{B}_1,\ldots,\mathbf{B}_n$ be diagonal operators on $\mathscr{H}$; then

$$P_0\bigg[\frac{1}{(1-\beta)^{n-1}}\int\frac{dp_{n-1}dq_{n-1}}{2\pi}\cdots\int\frac{dp_1dq_1}{2\pi}\mathbf{B}_n|p_{n-1},q_{n-1};\beta\rangle\rangle$$
$$\langle\langle p_{n-1},q_{n-1};\beta|\mathbf{B}_{n-1}|p_{n-2},q_{n-2};\beta\rangle\rangle\cdots|p_1,q_1;\beta\rangle\rangle\langle\langle p_1,q_1;\beta|\mathbf{B}_1\bigg]I_0$$
$$=\sum_{l_1,\ldots,l_{n-1}}\beta^{(l_1+\cdots+l_{n-1})}P_0\mathbf{B}_nI_{l_{n-1}}P_{l_{n-1}}\mathbf{B}_{n-1}I_{l_{n-2}}\cdots P_{l_1}\mathbf{B}_1I_0=B_{n,00}B_{n-1,00}\cdots B_{1,00};$$

hence [apply (27)]

$$\frac{1}{(1-\beta)^n}\int\frac{dp_{n-1}dq_{n-1}}{2\pi}\cdots\int\frac{dp_1dq_1}{2\pi}\prod_{j=0}^{n-1}\langle\langle p_{j+1},q_{j+1};\beta|\mathbf{B}_{j+1}|p_j,q_j;\beta\rangle\rangle\in\mathscr{C}(B_{n,00}B_{n-1,00}\cdots B_{1,00}) \tag{28}$$
$$(p_n=p'',q_n=q'';p_0=p',q_0=q').$$

Again, one can easily understand what has happened; since all the $\mathbf{B}_j$ are diagonal, the insertion of the generalized effective resolution of the identity (26) does not mix the $B_{j,kk}$ with different k and, as before in the linear combination of functions in the left-hand side of (27), only one term, the term corresponding to $B_{n,00}\cdots B_{1,00}$, is not orthogonal to $\{\langle\phi|p'',q''\rangle\langle p',q'|\psi\rangle;\ \psi,\phi\in\mathscr{H}\}$.

D. Application to the evolution operator

It is now easy to apply (28) to the propagator $U_{t'',t'}=T\exp[-i\int_{t'}^{t''}H(t)\,dt]$. We have

$$U_{t'',t'}=\lim_{n\to\infty}U_n(t'',t') \tag{29}$$

with $U_n(t'',t')=\exp[-iH(t''-\epsilon)\epsilon]\exp[-iH(t''-2\epsilon)\epsilon]\times\exp[-iH(t')\epsilon]$ [where $\epsilon=(t''-t')/n$]. Let $\mathbf{H}(t)$ be a self-adjoint, diagonal operator on $\mathscr{H}$ satisfying

$$H_{00}(t)=H(t).$$

Then (28) implies

$$\frac{1}{(1-\beta)^n}\int\frac{dp_{n-1}dq_{n-1}}{2\pi}\cdots\int\frac{dp_1dq_1}{2\pi}\prod_{j=0}^{n-1}$$
$$\times\langle\langle p_{j+1},q_{j+1};\beta|\exp[-i\mathbf{H}(t'+j\epsilon)\epsilon]$$
$$\times|p_j,q_j;\beta\rangle\rangle\in\mathscr{C}(U_n(t'',t')) \tag{30}$$
$$(p_n=p'',q_n=q'';p_0=p',q_0=q').$$

As yet we are still free to choose β and all the H_{kk} for $k\neq 0$; (30) holds for all possible choices. We shall see that, at least for certain quadratic Hamiltonians $H(t)$, it is possible to choose the β, H_{kk} in such a way that the functions (30) converge for $n\to\infty$, and give rise to an element of $\mathscr{C}(U_{t'',t'})$.

To show how the Wiener integral emerges, we first study the case $H(t)=0$. In this simple case, we take $\mathbf{H}(t)=0$ [i.e., $\forall k$, $H_{kk}(t)=H_{00}(t)=H(t)=0$]; the function (30) becomes [use (24)]

$$\frac{1}{(1-\beta)^n}\int\frac{dp_{n-1}dq_{n-1}}{2\pi}\cdots\int\frac{dp_1dq_1}{2\pi}$$
$$\times\prod_{j=0}^{n-1}\langle\langle p_{j+1},q_{j+1};\beta|p_j,q_j;\beta\rangle\rangle$$
$$=\frac{1}{(1-\beta)^n}\int\frac{dp_{n-1}dq_{n-1}}{2\pi}\cdots\int\frac{dp_1dq_1}{2\pi}$$
$$\times\prod_{j=0}^{n-1}\exp\bigg\{\frac{i}{2}(p_jq_{j+1}-p_{j+1}q_j)$$
$$-\frac{(1+\beta)}{4(1-\beta)}[(p_{j+1}-p_j)^2+(q_{j+1}-q_j)^2]\bigg\} \tag{31}$$
$$(p_n=p'',q_n=q'';p_0=p',q_0=q').$$

Recall[7] that the joint probability density for a Wiener process $x(t)$ to be at the points x_j at times t_j ($j=1,\ldots,m$; $t_m\geqslant t_{m-1}\geqslant\cdots\geqslant t_1\geqslant t_0$), having started at x_0 at time t_0, is given by

$$p(x_n,t_n;x_{n-1},t_{n-1};\cdots;x_0,t_0) = \prod_{j=1}^{n}\left\{\frac{1}{[2\pi(t_j-t_{j-1})]^{1/2}}\exp\left[-\frac{1}{2}\frac{(x_j-x_{j-1})^2}{t_j-t_{j-1}}\right]\right\}.$$

If all the time intervals are equal, $\forall j$: $t_j - t_{j-1} = \epsilon = (t_n - t_0)/n$ this becomes

$$p(x_n,t_0+n\epsilon;x_{n-1},t_0+(n-1)\epsilon;\cdots;x_0,t_0) = \frac{1}{(2\pi\epsilon)^{n/2}}\prod_{j=1}^{n}\exp\left[-\frac{1}{2}\frac{(x_j-x_{j-1})^2}{\epsilon}\right]. \tag{32}$$

In order to fit (31) with (32), we choose $t_0 = t'$, $t_n = t''$ and $(1+\beta)/[2(1-\beta)] = 1/\epsilon$ or $\beta = (1-\epsilon/2)/(1+\epsilon/2)$ [where $\epsilon = (t''-t')/n$]. With this choice for β, it is now clear that the Gaussian factors in (31) are correctly chosen to generate independent (non-normalized) pinned Wiener measures in p and q, pinned at the starting points so that $p(t') = p'$, $q(t') = q'$, and at the final points so that $p(t'') = p''$, $q(t'') = q''$. We shall denote these pinned Wiener measures by $d\mu_W{}^{t''\leftarrow t'}_{p''\leftarrow p'}(p)$ and $d\mu_W{}^{t''\leftarrow t'}_{q''\leftarrow q'}(q)$. It is also easy to see what the other factors in (31) become in the limit for $n\to\infty$, namely,

$$(2\pi\epsilon)^n\frac{1}{(2\pi)^{n-1}(1-\beta)^n} = 2\pi(1+\tfrac{1}{2}\epsilon)^n = 2\pi\left(1+\frac{t''-t'}{2n}\right)^n \underset{n\to\infty}{\longrightarrow} 2\pi e^{(t''-t')/2},$$

$$\sum[p_jq_{j+1} - p_{j+1}q_j] = \sum[p_j(q_{j+1}-q_j) - (p_{j+1}-p_j)q_j] \to \int [p(t)dq(t) - q(t)dp(t)].$$

Thus (31) becomes

$$2\pi e^{(t''-t')/2}\int d\mu_{W,}{}^{t''\leftarrow t'}_{p''\leftarrow p'}(p)d\mu_{W,}{}^{t''\leftarrow t'}_{q''\leftarrow q'}(q) \times\exp\left\{\frac{i}{2}\int [p(t)dq(t) - q(t)dp(t)]\right\}. \tag{33}$$

It is clear that the integrand in (33) may be given a well-defined meaning in terms of stochastic integrals. Moreover, since p and q are independent stochastic variables, all prescriptions for defining the stochastic integral in (33) are equivalent, which means that this integral is a perfectly well-defined path integral over a genuine measure. We shall evaluate this integral in the next section, and show that it is indeed an element of $\mathscr{C}(\mathbb{1})$.

In the simple case $H(t) = 0$ above, we chose all the H_{kk} identical, i.e., $H_{kk} = H_{00} = 0$. Although this is of course the simplest choice, there is no *a priori* reason to choose all the H_{kk} identical. Indeed, considering $H(t)\neq 0$ below, we shall see that in general the choice of identical H_{kk} does not lead to a well-defined limit of the expressions (30) as $n\to\infty$. On the other hand, it may well be possible that two different sequences of H_{kk}, H'_{kk}, with the same zeroth component $H_{00} = H'_{00} = H$, both lead to well-defined but different limits of (30), both of which are elements of $\mathscr{C}(U_{t'',t'})$.

For H linear in P,Q,

$$H(t) = s(t)Q + r(t)P,$$

the choice $H_{kk}(t) = H_{00}(t) = H(t)$ is satisfactory, and the result is

$$2\pi e^{(t''-t')/2}\int d\mu_{W,p''p'}(p)d\mu_{W,q''q'}(q)\exp\left(\frac{i}{2}\int [p(t)dq(t) - q(t)dp(t)] + \int [r(t)dq(t) - s(t)dp(t)] + \int dt\{-i[s(t)q(t) + r(t)p(t)] - \tfrac{1}{2}[s^2(t) + r^2(t)]\}\right), \tag{34}$$

where we assume s and r to be square integrable. Again, we shall check below that (34) is indeed an element of $\mathscr{C}(T\exp\{-i\int_{t'}^{t''} \times dt\,[s(t)Q + r(t)P]\})$.

For quadratic Hamiltonians, the choice of identical H_{kk} leads to convergence problems. We illustrate this by means of the simple time-independent quadratic Hamiltonian ($\alpha = \text{constant}\neq 0$)

$$H = (\alpha/2)(P^2 + Q^2 - 1).$$

Let us first try the choice $H_{kk} = H_{00} = H$. We get [$\epsilon = (t''-t')/n$]

$$\langle\langle p_{j+1},q_{j+1};\beta|e^{-i\mathbf{H}\epsilon}|p_j,q_j;\beta\rangle\rangle$$
$$= (1-\beta)\mathrm{Tr}[\beta^N W(p_{j+1},q_{j+1})^\dagger e^{-i\alpha N\epsilon}W(p_j,q_j)]$$
$$= (1-\beta)(1-\beta e^{-i\alpha\epsilon})^{-1}\exp\left\{\frac{i}{2}[q_{j+1}(-q_j\sin\alpha\epsilon + p_j\cos\alpha\epsilon) - p_{j+1}(q_j\cos\alpha\epsilon + p_j\sin\alpha\epsilon)] - \frac{1+\beta e^{-i\alpha\epsilon}}{4(1-\beta e^{-i\alpha\epsilon})}[(q_{j+1} - q_j\cos\alpha\epsilon - p_j\sin\alpha\epsilon)^2 + (p_{j+1} + q_j\sin\alpha\epsilon - p_j\cos\alpha\epsilon)^2]\right\}; \tag{35}$$

see (A8) in the Appendix.

To generate a measure in the limit $n\to\infty$, we have to choose

$$\beta = 1 - b\epsilon + o(\epsilon),$$

which leads to

$$\frac{1+\beta e^{-i\alpha\epsilon}}{1-\beta e^{-i\alpha\epsilon}}=\frac{1}{\epsilon}\left(\frac{2}{b+i\alpha}\,o(1)\right).$$

For $\beta\in[0,1)$, b is real, and the factor $2(b+i\alpha)^{-1}$ in front of ϵ^{-1} has a nonzero imaginary part as long as $\alpha\neq 0$, which means that (35) cannot generate a genuine path integral measure. At first sight, it seems that this problem can be circumvented by allowing β to be complex: $\beta=e^{-i\alpha\epsilon}(1-\epsilon/2)/(1+\epsilon/2)$; however, going back to (23), one sees that for β complex, the factor $(1+\beta^{-i\alpha\epsilon})/(1-\beta e^{-i\alpha\epsilon})$ would have to be replaced by $(1+|\beta|e^{-i\alpha\epsilon})/(1-|\beta|e^{-i\alpha\epsilon})$, which shows that a complex choice for β does not solve the convergence problems.

All convergence problems are avoided if the H_{kk} are chosen in the following way:

$$H_{kk}=(\alpha/2)(P^2+Q^2-1)-\alpha k\mathbf{1}=H-\langle H\rangle_k.$$

For this choice of (nonidentical!) H_{kk}, we obtain $[\epsilon=(t''-t')/n]$

$$\begin{aligned}&\langle\langle p_{j+1},q_{j+1};\beta|e^{-i\mathbf{H}\epsilon}|p_j,q_j;\beta\rangle\rangle\\&=(1-\beta)\mathrm{Tr}[\beta^N W(p_{j+1},q_{j+1})^\dagger e^{-i\alpha N\epsilon}W(p_j,q_j)e^{i\alpha N\epsilon}]\\&=\exp\left\{\frac{i}{2}[q_{j+1}(-q_j\sin\alpha\epsilon+p_j\cos\alpha\epsilon)-p_{j+1}(q_j\cos\alpha\epsilon+p_j\sin\alpha\epsilon)]\right.\\&\left.-\frac{1+\beta}{4(1-\beta)}[(q_{j+1}-q_j\cos\alpha\epsilon-p_j\sin\alpha\epsilon)^2+(p_{j+1}+q_j\sin\alpha\epsilon-p_j\cos\alpha\epsilon)^2]\right\};\end{aligned}\tag{36}$$

see (A8) in the Appendix.

We can now again choose $\beta=(1-\epsilon/2)/(1+\epsilon/2)$ $[\epsilon=(t''-t')/n]$; the substitution of (36) into (30) again leads to an integral w.r.t. the pinned Wiener measure, and we obtain

$$\begin{aligned}&2\pi e^{(t''-t')/2}\int d\mu_{W,\,p''\leftarrow p'}^{t''\leftarrow t'}(p)\,d\mu_{W,\,q''\leftarrow q'}^{t''\leftarrow t'}(q)\\&\times\exp\left\{\left(\frac{i}{2}+\alpha\right)\int[p(t)\,dq(t)-q(t)\,dp(t)]-\frac{\alpha}{2}(i+\alpha)\int dt\,[p^2(t)+q^2(t)]\right\}.\end{aligned}\tag{37}$$

The same technique of choosing

$$H_{kk}(t)=H(t)-\langle H(t)\rangle_k$$

works also for the time-dependent quadratic Hamiltonian

$$H(t)=[\alpha(t)/2](P^2+Q^2-\mathbf{1}+s(t)Q+r(t)P.$$

For this Hamiltonian we obtain

$$\begin{aligned}&2\pi e^{(t''-t')/2}\int d\mu_{W,\,p''\leftarrow p'}^{t''\leftarrow t'}(p)\,d\mu_{W,\,q''\leftarrow q'}^{t''\leftarrow t'}(q)\\&\times\exp\left[\int\left\{\left[\frac{i}{2}+\alpha(t)\right][p(t)\,dq(t)-q(t)\,dp(t)]-\frac{\alpha(t)}{2}[i+\alpha(t)][p(t)^2+q(t)^2]\,dt\right.\right.\\&\left.\left.-[s(t)\,dp(t)-r(t)\,dq(t)]-[i+\alpha(t)][s(t)q(t)+r(t)p(t)]\,dt-\tfrac{1}{2}[s(t)^2+r(t)^2]dt\right\}\right];\end{aligned}\tag{38}$$

the trace to be calculated is slightly more complicated than for (38); see (A10) and (A11) in the Appendix. Note that to give a sense to (38) or (34) we have to take s and r square integrable in $[t',t'']$. For (38) and (37) additional conditions on α will be introduced in Sec. 3 where needed.

In the next section (Sec. 3) we shall evaluate the path integrals (33), (34), (37), and (38). In Sec. 4 we show that they are indeed elements of the corresponding $\mathscr{C}(U_{t'',t'})$ for the Hamiltonians in question.

Remark: It is not really essential in the construction of (28) that the states $|n\rangle$ are the eigenstates of the harmonic oscillator; the only properties used are

$1^\circ)\langle m|n\rangle=\delta_{mn}$,

$2^\circ)|p,q\rangle=W(p,q)|0\rangle$.

We could therefore replace the vectors $|n\rangle$ by any orthonormal set ϕ_n in $\mathscr{H}$, as long as $\phi_0=|0\rangle$; the functions β^n can also be replaced by a positive function $\rho(\lambda_n)$, where λ_n are the eigenvalues of an operator A with eigenvectors ϕ_n: $A\phi_n=\lambda_n\phi_n$ with, however, the restriction that $\Sigma_n\rho(\lambda_n)<\infty$. This would allow us to try the same technique

$$H_{kk}(t)=H(t)-\langle H(t)\rangle_k$$

or, even more generally,

$$H_{kk}(t)=H(t)-g(\lambda_k,t),\quad\text{with }g(\lambda_0,t)=0.$$

for Hamiltonians different from the harmonic oscillator; the problem is then to choose $A,f,$ and g in such a way that the traces

$$\mathrm{Tr}[\rho(A)W(p_{j+1},q_{j+1})^\dagger e^{-iH(t)\epsilon}W(p_j,q_j)e^{ig(A,t)\epsilon}]$$

still have the right form to generate a genuine measure in the limit $n\rightarrow\infty$.

In the case of ν degrees of freedom ($\nu > 1$), a class of Hamiltonians for which the procedure above clearly works is given by

$$H(t) = \tfrac{1}{2}\alpha(t)\Big\{\sum_{i,j=1}^{\nu}[A_{ij}(P_iP_j + Q_iQ_j) + 2B_{ij}P_iQ_j]\Big\} + \sum_{j=1}^{\nu}[S_j(t)Q_j + R_j(t)P_j], \tag{39}$$

where A, B are $\nu\times\nu$ matrices with $A^t = A$, $B^t = -B$. The path integral corresponding to the Hamiltonian (39) is given below [expression (44)] in a more intrinsic and shorter notation system than we have used up to now (see Sec. 3A). The proof, given in Sec. 4, that the path integral (38) really is an element of the equivalence class $\mathscr{C}(U_{t'',t'})$ for the Hamiltonian $H(t) = \frac{1}{2}\alpha(t)(P^2 + Q^2 - \mathbb{1}) + s(t)Q + r(t)P$, easily extends to this multidimensional case.

One can show that the path integral (44) also gives an element of $\mathscr{C}(U_{t'',t'})$ for the more general quadratic Hamiltonian

$$H(t) = \frac{1}{2}\sum_{i,j=1}^{\nu}[A_{ij}(t)(P_iP_j + Q_iQ_j) + 2B_{ij}(t)P_iQ_j] + \sum_{j=1}^{\nu}[S_j(t)Q_j + R_j(t)P_j],$$

with $A_{ji}(t) = A_{ij}(t)$, $B_{ij}(t) = -B_{ji}(t)$, and A_{ij}, B_{ij} almost everywhere differentiable and piecewise continuous. The only change needed in (44) is the replacing of the constant matrices A, B by time-dependent ones. Note, however, that this is a generalization on the level of the path integral only, while for the Hamiltonian (39) (i.e., constant matrices A, B) the complete construction in Sec. 2 can be generalized; this is not true for the case where A, B are time-dependent; it would then be necessary to choose also the basis vectors $|n\rangle$ time-dependent, and the evaluation of the resulting formulas as traces (see above) would no longer hold.

3. EVALUATION OF THE PATH INTEGRALS

Since the path integrals (37), (34), and (33) can all be obtained from (38) [by putting, respectively, $r = s = 0$ for (37), $\alpha = 0$ for (34), and $r = s = \alpha = 0$ for (33)], we shall only evaluate (38) here.

A. Notations

For reasons of convenience, and to shorten the calculations, we shall use the more condensed symplectic notation system, introduced in Ref. 8 and frequently used thereafter in, e.g., studies of Weyl quantization[9]:

$$\begin{aligned}
&(p,q) = v,\\
&\sigma(v',v'') = \tfrac{1}{2}(p',q'' - p'',q'),\\
&Jv = J(p,q) = (-q,p),\\
&s(v',v'') = \sigma(v',Jv'') = \tfrac{1}{2}(p'p'' + q'q'').
\end{aligned}\tag{40}$$

Some simple and useful properties of σ, s, an J are

$$\begin{aligned}
&\sigma(v,v) = 0,\\
&J^2 = -\mathbb{1},\\
&\sigma(Jv',Jv'') = \sigma(v',v'') = s(Jv',v'') = -s(v',Jv''),\\
&e^{\gamma J} = \cos\gamma\mathbb{1} + \sin\gamma J \quad (\gamma\in\mathbb{C}),\\
&\sigma(e^{J\gamma}v',e^{\gamma J}v'') = \sigma(v',v'') \quad (\gamma\in\mathbb{C}),\\
&s(e^{\gamma J}v',e^{\gamma J}v'') = s(v',v'') \quad (\gamma\in\mathbb{C}).
\end{aligned}\tag{41}$$

We shall also use the following consequence of the properties of J:

$$(\mathbb{1} - e^{iJ\alpha})^{-1}(a - e^{iJ\alpha}b) = \frac{-i}{2}\coth\frac{\alpha}{2}J(a-b) + \frac{a+b}{2}. \tag{42}$$

Furthermore, we introduce the notations

$$|v|^2 = s(v,v) = \tfrac{1}{2}(p^2 + q^2),$$

$$\omega(v) = \exp(-\tfrac{1}{2}|v|^2).$$

In these notations, (6) and (13), e.g., become

$$\langle v''|v'\rangle = e^{i\sigma(v',v'')}\omega(v' - v''),$$

$$W(v')W(v'') = e^{i\sigma(v',v'')}W(v' + v'').$$

In the symplectic notation system, we can rewrite (38) as

$$2\pi e^{(t''-t')/2}\int d\mu_{W,\,v''\leftarrow v'}^{\,t''\leftarrow t'}(v) \times\exp\Big\{\int[(i+2\alpha)\sigma(v,dv) + 2\sigma(b,dv)] + \int dt\,[-\alpha(i+\alpha)s(v,v) - 2(i+\alpha)s(b,v) - s(b,b)]\Big\}, \tag{43}$$

where we have written $b(t)$ for $(r(t), s(t))$; in general, both α and b are time-dependent.

Almost the same integral can be written for the more complicated Hamiltonian (39); while this integral would be rather lengthy to write in the conventional notations, in the symplectic notation system it becomes

$$2\pi e^{(t''-t')/2}\int d\mu_{W,\,v''\leftarrow v'}^{\,t''\leftarrow t'}(v) \times\exp\Big\{\int[\sigma(iv + 2\alpha Cv + 2b,dv)] + \int dt\,[-s(v,i\alpha Cv + \alpha^2C^2v) - 2s(ib + \alpha Cb,v) - s(b,b)]\Big\}, \tag{44}$$

where now v is a 2ν-dimensional pinned Wiener process, where b is the 2ν-dimensional vector $(R_1,\ldots,R_\nu; S_1,\ldots,S_\nu)$, and where C is the $2\nu\times 2\nu$ matrix $\begin{pmatrix}A & B\\ -B & A\end{pmatrix}$.

In evaluating any of these Gaussian integrals, the result will be given by the contribution of the extremal path, multiplied by a suitable constant.

B. Contribution of the extremal path

To determine the extremal path, we proceed formally and extract the $\exp[-\frac{1}{2}(\dot{p}^2+\dot{q}^2)]$ from the Wiener measure, and (43) then becomes

$$2\pi e^{(t''-t')/2}\int \mathcal{D}v_{t''\leftarrow t'}^{v''\leftarrow v'}\exp\left\{\int_{t'}^{t''} dt\,[i\sigma(-\dot{v}+\alpha Jv+2Jb,v)-|-\dot{v}+\alpha Jv+b|^2]\right\}. \tag{45}$$

We can rewrite the integrand in the exponent as

$$\begin{aligned} F(v,\dot{v},t) &= -\alpha(i+\alpha)s(v,v)+s(v,-2(i+\alpha)b-(i+2\alpha)J\dot{v})+f(\dot{v},t)\\ &= -s(\dot{v},\dot{v})+s(\dot{v},2\alpha Jv+2Jb+iJv)+g(v,t). \end{aligned}$$

The variational equations are therefore (we assume α to be differentiable a.e.)

$$-2\alpha(i+\alpha)v-2(i+\alpha)b-(i+2\alpha)J\dot{v}-\frac{d}{dt}(-2\dot{v}+2\alpha Jv+2Jv+iJv)=0$$

or

$$\ddot{v}-(i+2\alpha)J\dot{v}-\alpha(i+\alpha)Jv-\dot{\alpha}Jv=(i+\alpha)b+J\dot{b},$$

which can be rewritten as

$$\left[\frac{d}{dt}-(i+\alpha)J\right]\left(\frac{d}{dt}-\alpha J\right)v=J\left[\frac{d}{dt}-(i+\alpha)J\right]b.$$

The extremal path is therefore given by

$$v(t)=e^{a(t)J}[c+e^{iJ(t-t')}d+JB(t)],$$

where

$$a(t)\equiv\int_{t'}^{t} ds\,\alpha(s),\quad B(t)\equiv\int_{t'}^{t} ds\,e^{-a(s)J}b(s) \tag{46}$$

(to define a, we assume α to be locally L^1), while the boundary conditions $v(t')=v'$, $v(t'')=v''$ impose

$$c+d=v',\quad c+e^{iJ(t''-t')}d=e^{-a(t'')J}v''-JB(t''). \tag{47}$$

We now evaluate the exponential in (45) for this extremal path. Since [from (46)]

$$-\dot{v}+\alpha Jv+Jb=-iJe^{[a(t)+i(t-t')]J}d,$$

we have [use (41)]

$$\begin{aligned} &i\sigma(\dot{v}+\alpha J+2Jb,v)-s(-\dot{v}+\alpha Jv+Jb,-\dot{v}+\alpha Jv+Jb)\\ &\quad=i\sigma(-iJe^{iJ(t-t')}d+J\dot{B}(t),c+e^{iJ(t-t')}d+JB(t))+s(d,d)\\ &\quad=i\sigma(\dot{B}(t),B(t))-i\frac{d}{dt}s(B(t),c+e^{iJ(t-t')}d)-s(e^{iJ(t-t')}d,c). \end{aligned}$$

Integrating this we obtain

$$\begin{aligned} &\int_{t'}^{t''} dt\,[i\sigma(-\dot{v}+\alpha Jv+2Jb,v)-|-\dot{v}+\alpha Jv+b|^2]\\ &\quad=i\int_{t'}^{t''}\sigma(\dot{B}(t),B(t))dt-is(B(t''),c+e^{iJ(t''-t')}d)-i\sigma(e^{iJ(t''-t')}d-d,c)\\ &\quad=i\int_{t'}^{t''} dt\,\sigma(\dot{B}(t),B(t))-is(B(t''),e^{-a(t'')J}v'')\\ &\quad\quad-i\sigma(e^{-a(t'')J}v''-JB(t'')-v',[1-e^{iJ(t''-t')}]^{-1}[e^{-a(t'')J}v''-JB(t'')-e^{iJ(t''-t')}v']). \end{aligned}$$

Using (42), this becomes

$$\begin{aligned} \int_{t'}^{t''} dt\cdots &= i\int_{t'}^{t''} dt\sigma(\dot{B}(t),B(t))-is(B(t''),v'+e^{-a(t'')J}v'')\\ &\quad-i\sigma(e^{-a(t'')J}v'',v')-\tfrac{1}{2}\coth\frac{t''-t'}{2}|e^{-a(t'')J}v''-JB(t'')-v'|^2. \end{aligned}$$

So finally (43) is equal to

$$\begin{aligned} (43)&=2\pi e^{(t''-t')/2}A_{t'',t'}\cdot\exp\Big[i\int_{t'}^{t''} dt\,\sigma(\dot{B}(t),B(t))\\ &\quad-is(B(t''),v'+e^{-a(t'')J}v'')-i\sigma(e^{-a(t'')J}v'',v')\Big]\omega[e^{-a(t'')J}v''-v'-JB(t'')]^{\coth[(t''-t')/2]}, \end{aligned} \tag{48}$$

where the multiplicative constant $A_{t'',t'}$ still has to be determined, and where $a(t)$, $B(t)$ are given by (46).

C. Determination of the multiplicative constant

In this and in some of the following subsections we shall use the shorthand dv for the measure $dp\, dq/2\pi$.

With respect to this measure, we have

$$\int dv\, \omega^2(v) = 1.$$

$$\int dv\, e^{i\beta\sigma(v',v)}\omega(v)^\alpha = \frac{2}{\alpha}\,\omega(v')^{\beta^2/\alpha}. \tag{49}$$

We shall also use the following property:

Take any complex $2n\times 2n$ matrix A (matrix elements A_{ij}) satisfying

$$A^t = A,$$

$$\operatorname{Re} A = \frac{A + A^\dagger}{2} > 0;$$

let a_{ij} be the 2×2 matrices

$$a_{ij} = \begin{pmatrix} A_{2i-1,2j-1} & A_{2i-1,2j} \\ A_{2i,2j-1} & A_{2i,2j} \end{pmatrix},$$

then

$$\int dv_1 \cdots \int dv_n \exp\left[-\sum_{i,j=1}^{n} s(v_i, a_{ij} v_j)\right] = \frac{1}{(\det A)^{1/2}}. \tag{50}$$

If A is real, the square root to be chosen is the positive one; if A is not a real matrix, the sign of $(\det A)^{1/2}$ is determined as follows:

$$(\det A)^{1/2} = \lim_{\lambda\to 1} f(\lambda), \text{ with } \begin{cases} f:[0,1]\to\mathbb{C} \text{ continuous,} \\ f(0)\in\mathbb{R}_+ \\ f(\lambda)^2 = \det(\operatorname{Re} A + i\lambda \operatorname{Im} A). \end{cases}$$

Let us now proceed to the determination of $A_{t'',t'}$ under the assumption that $\alpha(t)$ is a continuous function (this condition will be relaxed at the end of this section). As usual, the constant $A_{t'',t'}$ can be shown (by a variational argument) to be independent of the boundary conditions v',v'' and of the linear parts of the integrand in the exponential in (45). Hence

$$A_{t'',t'} = \int Dv_{0\leftarrow 0}^{t''\leftarrow t'} \times \exp\left\{\int_{t'}^{t''} dt\, [i\sigma(-\dot v + \alpha J v, v) - |\dot v - \alpha J v|^2]\right\}.$$

Writing this out as a limit, we obtain [as before $\epsilon = (t'' - t')/n$]

$$\begin{aligned} A_{t'',t'} &= \lim_{n\to\infty} \frac{1}{2\pi\epsilon^n}\int dv_1 \cdots \int dv_{n-1} \prod_{j=0}^{n-1} \exp\left[i\sigma(-v_{j+1} + \alpha(t'+\epsilon)Jv_j\epsilon, v_j) - \frac{1}{\epsilon}|v_{j+1} - v_j - \alpha(t'+j\epsilon)Jv_j\epsilon|^2\right](v_n = v_0 = 0) \\ &= \lim_{n\to\infty} \frac{1}{2\pi\epsilon}\int du_1 \cdots du_{n-1} \prod_{j=0}^{n-1} \exp[-is(Ju_{j+1} + \alpha(t'+j\epsilon)u_j\epsilon, u_j)\epsilon \\ &\quad - s(u_{j+1} - u_j - \alpha(t'+j\epsilon)Ju_j, u_{j+1} - u_j - \alpha(t'+j\epsilon)u_j)](u_n = u_0 = 0) \\ &= \lim_{n\to\infty} \frac{1}{2\pi\epsilon}\int du_1 \cdots \int du_{n-1} \exp\left[-\sum_{i,j=1}^{n-1} s(u_j, m_{i,j} u_j)\right], \end{aligned} \tag{51}$$

where $m_{i,j}$ are 2×2 matrices defined by

$$m_{k,k} = [2 + \epsilon^2\alpha(t'+k\epsilon)(i + \alpha(t'+k\epsilon))]\,\mathbb{1},$$

$$m_{k,k+1} = -\mathbb{1} + \left[\frac{i}{2} + \alpha(t'+k\epsilon)\right]\epsilon J,$$

$$m_{k+1,k} = -\mathbb{1} - \left[\frac{i}{2} + \alpha(t'+k\epsilon)\right]\epsilon J,$$

$$m_{k,l} = 0 \quad \text{if} \quad |k-l| > 1,$$

with J [as in (40)] given by

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Applying (50) to (51) we see that

$$A_{t'',t'} = \lim_{n\to\infty} \frac{1}{2\pi}\frac{1}{\epsilon(\det M_n)^{1/2}}, \tag{52}$$

where M_n is the $2(n-1)\times 2(n-1)$ matrix constituted by the 2×2 blocks $m_{i,j}$. We shall calculate the limit (52) in the standard way, i.e., by constructing a recursion formula for $\det M_n$. In the limit for $n\to\infty$, this recursion formula will become a differential equation, and the solution of this differential equation then gives an explicit expression for (52). Due to the particular structure of the matrix M_n, and the continuity of α, all $\alpha(t)$ dependence will cancel from the differential equation, leading to a constant $A_{t'',t'}$ independent of $\alpha(t)$!

For n fixed we define $M_{n,k}$ to be the $2k\times 2k$ matrix constructed from the 2×2 blocks $m_{i,j}$ with $i,j\leqslant k$ in the fashion

$$M_{n,k} = \begin{pmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,k} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,k} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ m_{k,1} & m_{k,2} & \cdots & m_{k,k} \end{pmatrix}.$$

We furthermore define

$D_{n,k} = \det M_{n,k}$,

$D_{n,k,1}$ = minor in $M_{n,k}$ of the matrix element $(M_{n,k})_{2k,2k}$,

and

$D_{n,k,2}$ = minor in $M_{n,k}$ of the matrix element $(M_{n,k})_{2k-1,2k-1}$.

Developing the determinants $D_{n,k}$, $D_{n,k,1}$, and $D_{n,k,2}$ into products of matrix elements with the appropriate minors (we develop these determinants along the last row of columns), we obtain the following two recursion relations:

$$\begin{aligned} D_{n,j} = {} & [2+\alpha_j(i+\alpha_j)\epsilon^2](D_{n,j,1}+D_{n,j,2}) \\ & + [2+\alpha_j(i+\alpha_j)\epsilon^2]^2 D_{n,j-1} + \left[1+\left(\frac{i}{2}+\alpha_j\right)^2\epsilon^2\right]^2 D_{n,j-2}, \end{aligned} \tag{53}$$

$$D_{n,j,1}+D_{n,j,2} = 2[2+\alpha_j i+\alpha_j)\epsilon^2]D_{n,j-1} - \left[1+\left(\frac{i}{2}+\alpha_j\right)^2\epsilon^2\right](D_{n,j-1,1}+D_{n,j-1,2}),$$

where we have written α_k for $\alpha(t'+k\epsilon)$. The quantity of interest to us is

$$f_n[t'+(j-1)\epsilon] = \epsilon^2 D_{n,j}. \tag{54}$$

In the limit for $n\to\infty$, this relation defines a (continuous) function f_∞ on $[t',t'']$, and we see from (52) that

$$A_{t'',t'} = \frac{1}{2\pi}\frac{1}{[f_\infty(t'')]^{1/2}}, \tag{55}$$

where the procedure discussed below (50) has to be applied to determine the sign of the square root. By analogy with (54), we define

$$g_n[t'+(j-1)\epsilon] = \epsilon^2(D_{n,j,1}+D_{n,j,2}).$$

In terms of the g_n, f_n the recursion equations (53) can be written as $[t_j = t'+(j-1)\epsilon]$

$$f_n(t_j) = \frac{g_n(t_{j+1}) + \{1+\epsilon^2[i/2+\alpha(t_{j+1})]^2\}g_n(t_j)}{2\{2+\alpha(t_{j+1})[i+\alpha(t_{j+1})]\epsilon^2\}},$$

$$g_n(t_j) = \frac{f_n(t_j) + \{2+\alpha(t_j)[i+\alpha(t_j)]\epsilon^2\}^2 f(t_{j-1}) - \{1+\epsilon^2[i/2+\alpha(t_j)]^2\}^2 f(t_{j-2})}{2+\alpha(t_j)[i+\alpha(t_j)]\epsilon^2}.$$

Substituting the expression for g_n into the equation for f_n, and grouping the terms of order 1, ϵ^2, ϵ^4,$\cdots$ together, we obtain from these equations the relation

$$\begin{aligned} -2f_n(t_{j+1}) + 6f_n(t_j) - 6f_n(t_{j-1}) + 2f_n(t_{j-2}) = {} & \epsilon^2[\xi_j f_n(t_{j+1}) + (2\zeta_{j+1} - 7\xi_{j+1} - 4\xi_j)f_n(t_j) \\ & + (4\zeta_{j+1} + 4\xi_{j+1} + 7\xi_j)f_n(t_{j-1}) + (-2\zeta_{j+1} - \xi_{j+1} - 4\xi_j)f_n(t_{j-2})] + O(\epsilon^4), \end{aligned} \tag{56}$$

where

$$\xi_j = \alpha(t_j)[i+\alpha(t_j)], \quad \zeta_j = \left[\frac{i}{2}+\alpha(t_j)\right]^2.$$

Using the fact that $\zeta_j = \xi_j - \frac{1}{4}$, (56) can be rewritten as

$$\begin{aligned} -2[f_n(t_{j+1}) - 3f_n(t_j) + 3f_n(t_{j-1}) - f_n(t_{j-1})] = {} & \epsilon^2\{-\tfrac{1}{2}[f_n(t_j) + 2f_n(t_{j-1}) - 3f_n(t_{j-2})] \\ & + \xi_{j+1}[-5f_n(t_j) + 8f_n(t_{j-1}) - 3f_n(t_{j-2})] + \xi_j[f_n(t_{j+1}) - 4f_n(t_j) + 7f_n(t_{j-1}) - 4f_n(t_{j-2})]\} + O(\epsilon^4). \end{aligned} \tag{57}$$

Equation (57) holds for fixed n, and for j:$2\to n-1$, again with $\epsilon = (t''-t')/n$. In the limit for $n\to\infty$, (57) will lead to a differential equation for f_∞, and as an intermediate step we obtain

$$\begin{aligned} -2f_n'''(t_j)\epsilon^3 + O(\epsilon^4) = {} & \epsilon^2[-\tfrac{1}{2}\cdot 4f_n'(t_j)\epsilon + O(\epsilon^2) + \xi_{j+1}(-2)f_n'(t_j)\epsilon + O(\epsilon^2) \\ & + \xi_j\cdot 2f_n'(t_j)\epsilon + O(\epsilon^2)] + O(\epsilon^4), \end{aligned}$$

hence

$$f_n'''(t_j) = f_n'(t_j) + (\xi_{j+1} - \xi_j)f_n'(t_j) + O(\epsilon). \tag{58}$$

It is here that the "miracle" happens: If $\alpha(t)$ is a continuous function, then

$$\xi_{j+1} - \xi_j = \alpha(t_{j+1})[i+\alpha(t_{j+1})] - \alpha(t_j)[i+\alpha(t_j)] = o(1),$$

which means that the α-dependent terms ξ_j drop out of the equation. In the limit $n\to\infty$, Eq. (58) becomes

$$f'''_{\infty}(t) = f'_{\infty}(t),$$

where all the α dependence has vanished!

To determine the initial conditions for this differential equation, we go back to the definition (54) of the f_n, and we easily obtain

$$f_{\infty}(t') = \lim_{n\to\infty} f_n(t_0) = \lim_{n\to\infty} \epsilon^2 \cdot 1 = 0,$$

$$f'_{\infty}(t') = \lim_{n\to\infty} \frac{f_n(t_1) - f_n(t_0)}{\epsilon} = \lim_{n\to\infty} \epsilon[(2 + A_1\epsilon^2)^2 - 1] = 0,$$

$$f''_{\infty}(t') = \lim_{n\to\infty} \frac{f_n(t_2) - 2f_n(t_1) + f_n(t_0)}{\epsilon^2} = \lim_{n\to\infty} \{ [(2 + \xi_2\epsilon^2)(2 + \xi_1\epsilon^2) - \zeta_2\epsilon^2 - 1]^2 - 2(2 + \xi_1\epsilon^2)^2 + 1 \} = 2.$$

With these initial conditions, the solution of the differential equation is

$$f_{\infty}(t) = 2[\cosh(t - t') - 1] = 4\left(\sinh\frac{t - t'}{2}\right)^2.$$

One can easily check that the procedure sketched under (50) to determine the sign of the square root of $\det M_n$ gives

$$\lim_{n\to\infty} \sqrt{\det M_n} = \lim_{\lambda\to 1} \frac{2}{\lambda} \sinh\frac{(t'' - t')\lambda}{2} = 2 \sinh\frac{t'' - t'}{2};$$

so finally [from (55)] we find

$$A_{t'',t'} = \left(4\pi \sinh\frac{t'' - t'}{2}\right)^{-1}. \tag{59}$$

Substituting expression (59) for $A_{t'',t'}$ into (48), we have as a final result

$$2\pi e^{(t''-t')/2}\int d\mu_{W,\,{}^{t''\leftarrow t'}_{v''\leftarrow v'}} \exp\left[\int\int [(i + 2\alpha)\sigma(v,dv) + 2\sigma(b,dv)] + \int dt\, [-\alpha(i + \alpha)s(v,v) - 2(i + \alpha)s(b,v) - s(b,b)]\right]$$

$$= \frac{1}{1 - e^{-(t''-t')}} \exp\Big\{i\int_{t'}^{t''} dt_1 \int_{t'}^{t_1} dt_2 [\sigma(b(t_1),b(t_2))\cos(\beta(t_1,t_2)) + s(b(t_1),b(t_2))\sin(\beta(t_1,t_2))]$$

$$- i\int_{t'}^{t''} dt\, [s(b(t),v')\cos(\beta(t,t')) - \sigma(b(t),v')\sin(\beta(t,t'))$$

$$+ s(b(t),v'')\cos(\beta(t'',t')) + \sigma(b(t),v'')\sin(\beta(t'',t))] - i\sigma(v'',v')\cos(\beta(t'',t')) - is(v'',v')\sin\beta(t'',t')\Big\}$$

$$\times\omega\left\{v''\cos[\beta(t'',t')] - Jv''\sin[\beta(t'',t')] - v' - \int_{t'}^{t} dt\, [Jb(t)\cos(\beta(t,t')) + b(t)\sin(\beta(t,t'))]\right\}^{\coth[(t''-t')/2]}, \tag{60}$$

where

$$\beta(t_1,t_2) = \int_{t_2}^{t_1} dt\, \alpha(t).$$

Putting $b(t) = [0,s(t)]$, $\alpha(t) = \alpha$ (time independent), (60) can easily be seen to lead to expression (15) in Ref. 4.

Remark: In what follows, we shall denote the right-hand side of (60) by $F_{t'',t'}(v'',v')$:

$$F_{t'',t'}(v'',v') = [1 - e^{-(t'',t')}]^{-1}\exp\left[i\int_{t'}^{t''} dt\, \sigma(\dot{B}(t),B(t))\right.$$

$$\left. - i\sigma(v' + e^{-a(t'')J}v'',JB(t'')) - i\sigma(e^{-a(t'')J}v'',v')\right]\omega(v' + JB(t'') - e^{-a(t'')J}v'')^{\coth[(t''-t')/2]}.$$

One can easily show that these $F_{t'',t'}$ have an interesting property,

$$\forall\tilde{t}\in[t,t'']: \int d\tilde{v}\, F_{t',\tilde{t}}(v,\tilde{v})F_{\tilde{t},t''}(\tilde{v},v'') = F_{t',t''}(v',v'').$$

This can be proven by direct computation; it can also be considered as a consequence of the fact that

$$\forall\tilde{t}\in[t',t'']: \int d\tilde{v} \int d\mu'_{W,\,{}^{t''\leftarrow\tilde{t}}_{v''\leftarrow\tilde{v}}}\, d\mu_{W,\,{}^{\tilde{t}\leftarrow t'}_{\tilde{v}\leftarrow v'}} = d\mu_{W,\,{}^{t''\leftarrow t'}_{v''\leftarrow v'}}.$$

This last property can be used to relax the continuity requirement on α (see above); for piecewise continuous α, we can cut the path integral into different pieces corresponding to time intervals on which α is continuous. For each of these pieces, our evaluations as carried out above hold without any problem. We can then use the "chain property" of the $F_{t'',t'}$ as stated above (the direct proof of which does not require α to be continuous) to show that even for piecewise continuous α the result (60) still holds.

Bringing all our conditions on α, r, and s together, we see now that
—α has to be piecewise continuous, a.e. differentiable and locally L^1
—r and s have to be in $L^2([t',t''])$.

Now that we have calculated the integral, we shall verify in the next section that the result is indeed an element of $\mathscr{C}(U_{t'',t'})$ for the corresponding Hamiltonian; we shall also discuss in what respect it differs from the matrix element $\langle v''|U_{t'',t'}|v'\rangle$.

4. THE PATH INTEGRALS YIELD ELEMENTS OF THE PROPAGATOR EQUIVALENCE CLASS

Let us denote the function defined by (60) by $F_{t'',t'}(v'',v')$. We claim that $F_{t'',t'}\in\mathscr{C}(U_{t'',t'})$, i.e., that

$$\int dv''\int dv'|v\rangle F_{t'',t'}(v'',v')\langle v'| = U_{t'',t'} = T\exp\left[-i\int_{t'}^{t''}H(t)\,dt\right], \tag{61}$$

with $H(t)=\frac{1}{2}\alpha(t)(P^2+Q^2-\mathbb{1})+r(t)P+s(t)Q$. To prove (61) it is sufficient to show for all v that

$$\int dv''\int dv'\langle v|v''\rangle F_{t'',t'}(v'',v')\langle v'|v\rangle = \langle v|U_{t'',t'}|v\rangle. \tag{62}$$

The equivalence of (61) and (62) follows from a standard analyticity argument: Since $F_{t'',t'}$ is uniformly bounded,

$|F_{t'',t'}(v'',v')|\leqslant[1-e^{-(t''-t')}]^{-1}$ (we always assume $t''>t'$),

one can use [see (16)]

$$\langle a|b\rangle = e^{(p_a+ix_a)(p_b-ix_b)/2}e^{(x_a^2+p_a^2+x_b^2+p_b^2)/4}$$

to show that the function

$$\phi(v_1,v_2)=\int dv''\int dv'\langle v_2|v''\rangle F_{t'',t'}(v'',v')\langle v'|v_1\rangle$$

can be written as a product,

$$\phi(v_1,v_2)=f(v_1,v_2)e^{(x_1^2+p_1^2+x_2^2+p_2^2)/4},$$

where f is a complex analytic function in the variables p_1+ix_1, p_2-ix_2. The matrix element $\langle v_1|U_{t'',t'}|v_2\rangle$ is a function of the same type:

$$\langle v_1|U_{t'',t'}|v_2\rangle = g(v_1,v_2)e^{-(x_1^2+p_1^2+x_2^2+p_2^2)/4}$$

with g complex analytic in p_1+ix_1, p_2-ix_2. Equation (62) can be rewritten as

$$\forall v{:}f(v,v)=g(v,v).$$

Because of their analyticity, this condition forces f and g to be identical:

$$\forall v_1,v_2{:}f(v_1,v_2)=g(v_1,v_2)$$

or

$$\forall v_1,v_2{:}\int dv''\int dv'\langle v_2|v''\rangle F_{t'',t'}(v'',v')\langle v'|v_1\rangle = \langle v_2|U_{t'',t'}|v_1\rangle. \tag{63}$$

Now using the fact that $|F_{t'',t'}|$ is an L^1 function in $v''-v'$, and the density of the linear span of the c.s. in $\mathscr{H}$, one sees that (63) implies (61).

Note that the argument above only uses properties of the "small" space $\mathscr{H}$. The "big" space $\hat{\mathscr{H}}$ was only introduced as a device to define $F_{t'',t'}$ as a path integral with respect to a genuine measure. Once $F_{t'',t'}$ is found, we no longer concern ourselves with $\hat{\mathscr{H}}$ or the $|p,q;\beta\rangle\rangle$, but simply prove directly that $F_{t'',t'}\in\mathscr{C}(U_{t'',t'})$.

We now proceed to prove (62). Using $\langle v|v''\rangle\langle v'|v\rangle = e^{i\sigma(v,v'-v'')}\,\omega(v-v')\omega(v''-v)$ (see Sec. 3A and (60)), we have

$$\begin{aligned}\int dv''\int dv'\langle v|v''\rangle F_{t'',t'}(v'',v')\langle v'|v\rangle &= \int dv''\int dv' e^{i\sigma(v,v'-v'')}\omega(v-v')\omega(v-v'')[1-e^{-(t''-t')}]^{-1}\\ &\times\exp\left[i\int_{t'}^{t''}dt\,\sigma(\dot B(t),B(t))-i\sigma(v'+e^{-a(t'')J}v'',JB(t''))\right]\\ &\times\exp[-i\sigma(e^{-a(t'')J}v'',v')]\,\omega[e^{-a(t'')J}v''-v'-JB(t'')]^{\coth[(t''-t')/2]},\end{aligned} \tag{64}$$

where

$$a(t)=\int_{t'}^{t}ds\,\alpha(s),\quad B(t)=\int_{t'}^{t}ds\,e^{-a(s)J}b(s).$$

Introducing the change of variable $u'' = e^{-a(t'')J}v''$, and using (41), Eq. (64) becomes

$$(64) = [1 - e^{-(t''-t')}]^{-1}\exp\left[i\int_{t'}^{t''} dt\,\sigma(\dot B(t),B(t))\right]\int du''\int dv'\exp[-i\sigma(u'',JB(t'') - e^{-a(t'')J}v)]$$
$$\times e^{-i\sigma(v',JB(t'')+v-u'')}\omega(v'-v)\omega[u''-v'-JB(t'')]^{\coth[(t''-t')/2]}\omega[u''-e^{-a(t'')J}v].$$

Taking the Gaussian in v' together and completing the squares, we have

$$\omega(v'-v)\omega[v'-u''+JB(t'')]^{\coth[(t''-t')/2]}$$
$$= \omega\left\{v' - \left[v + \coth\frac{t''-t'}{2}\cdot(u''-JB(t''))\right]\Big/\left[1+\coth\frac{t''-t'}{2}\right]\right\}^{1+\coth[(t''-t')/2]}$$
$$\cdot\,\omega[v-(u''-JB(t''))]^{\coth[(t''-t')/2]/\{1+\coth[(t''-t')/2]\}}.$$

Substituting this in the integral above, and making the change of variable

$$u' = v' - \frac{1}{1+\coth[(t''-t')/2]}\left[v + \coth\frac{t''-t'}{2}(u''-JB(t''))\right],$$

(64) becomes

$$(64) = [1 - e^{-(t''-t')}]^{-1}\exp\left[i\int_{t'}^{t''} dt\,\sigma(\dot B(t),B(t))\right]\int du''\int du'\exp[-i\sigma(u'',JB(t'') - e^{-a(t'')J}v)]e^{-i\sigma(v,JB(t'')-u'')}$$
$$\times e^{-i\sigma(u',JB(t'')-u''+v)}\omega[u''-e^{-a(t'')J}v]\omega(u')^{1+\coth[(t''-t')/2]}\cdot\omega[v-u''+JB(t'')]^{\coth[(t''-t')/2]/\{1+\coth[(t''-t')/2]\}}.$$

Applying (49) to the u' integral yields

$$(64) = \frac{2}{[1-e^{-(t''-t')}](1+\coth[(t''-t')/2])}\exp\left[i\int_{t'}^{t''} dt\,\sigma(\dot B(t),B(t))\right]$$
$$\times\int du''\exp[-i\sigma(u'',JB(t'') - e^{-a(t'')J}v + v)]e^{-i\sigma(v,JB(t''))}\omega[u''-e^{-a(t'')J}v]\omega[v-u''+JB(t'')].$$

Again we group the Gaussians in u'', and complete the squares

$$\omega[u''-e^{-a(t'')J}v]\omega[u''-v-JB(t'')] = \omega\{u''-\tfrac{1}{2}[e^{-a(t'')J}v+v+JB(t'')]\}^2\omega[v+JB(t'')-e^{-a(t'')J}v]^{1/2}.$$

Substituting this into the integral, and making the change of variable

$$u = u'' - \tfrac{1}{2}[e^{-a(t'')J}v + v + JB(t'')],$$

we obtain

$$(64) = \frac{2[e^{(t''-t')/2} - e^{-(t''-t')/2}]}{[1-e^{-(t''-t')}]\cdot 2e^{(t''-t')/2}}\exp\left[i\int_{t'}^{t''} dt\,\sigma(\dot B(t),B(t))\right]\int du\,\exp[-i\sigma(e^{-a(t'')J}v,v+JB(t''))]e^{-i\sigma(v,JB(t''))}$$
$$\times\omega[v+JB(t'')-e^{-a(t'')J}v]^{1/2}\exp[-i\sigma(u,v+JB(t'')-e^{-a(t'')J}v)]\omega^2(u).$$

Applying (49) again, this becomes

$$(64) = \exp\left[i\int_{t'}^{t''} dt\,\sigma(\dot B(t),B(t))\right]e^{-i\sigma(v,JB(t''))}\exp[-i\sigma(e^{-a(t'')J}v,v+JB(t''))]\omega[v+JB(t'')-e^{-a(t'')J}v]$$
$$= \exp\left[i\int_{t'}^{t''} dt\,\sigma(\dot B(t),B(t))\right]e^{-i\sigma(v,JB(t''))}\langle e^{-a(t'')J}v|v+JB(t'')\rangle \tag{65}$$

(see Sec. 3A).

Since (see Appendix) $e^{-i\beta N}|v\rangle = |e^{\beta J}v\rangle$, we thus have

$$\int dv''\int dv'\langle v|v''\rangle F_{t'',t'}(v'',v')\langle v'|v\rangle = \exp\left[i\int_{t'}^{t''} dt\,\sigma(\dot B(t),B(t))\right]e^{-\sigma(v,JB(t''))}\langle v|e^{-ia(t'')N}|v+JB(t'')\rangle$$
$$= \exp\left[i\int_{t'}^{t''} dt\,\sigma(\dot B(t),B(t))\right]\langle v|e^{-ia(t'')N}W(JB(t''))|v\rangle.$$

Our claim (61) thus reduces to

$$\exp\left[i\int_{t'}^{t''} dt\,\sigma(\dot B(t),B(t))\right]e^{-ia(t'')N}W(JB(t'')) = U_{t'',t'} = T\exp\left[-i\int_{t'}^{t''} dt\,H(t)\right], \tag{66}$$

where [see (46)]

$$a(t) = \int_{t'}^{t} ds\,\alpha(s),\quad B(t) = \int_{t'}^{t} ds\,e^{-a(s)J}b(s).$$

We shall prove (66) by differentiation with respect to t. We have, first of all,

$$\frac{d}{dt}W(JB(t)) = \lim_{\epsilon\to 0}\frac{1}{\epsilon}[W(JB(t+\epsilon))W(-JB(t)) - \mathbf{1}]W(JB(t))$$
$$= \lim_{\epsilon\to 0}\frac{1}{\epsilon}[e^{-i\sigma(B(t+\epsilon),B(t))}W(JB(t+\epsilon)-JB(t)) - \mathbf{1}]W(JB(t))$$
$$= -i\sigma(\dot B(t),B(t))W(JB(t)) - 2is(\dot B(t),V)W(JB(t))$$

[the second term can be obtained by putting $B=(R,S)$]; then

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon}W(JB(t+\epsilon)-JB(t))$$
$$= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\exp(-i\{[R(t+\epsilon)-R(t)]P + [S(t+\epsilon)-S(t)]Q\}) = -i[\dot R(t)P + \dot S(t)Q] \equiv -2is(\dot B(t),V)\}.$$

Moreover, $e^{-i\beta N}s(B,V) = s(e^{\beta J}B,V)e^{-i\beta N}$ [this can be obtained from (A5) by differentiation].

After these preliminaries we are now ready to evaluate the time derivative of the left-hand side of (66)

$$i\frac{d}{dt}\left\{\exp\left[i\int_{t'}^{t} ds\,\sigma(\dot B(s),B(s))\right]e^{-i\alpha(t)N}W(JB(t))\right\}$$
$$= [-\sigma(\dot B(t),B(t)) + \dot a(t)N + \sigma(\dot B(t),B(t)) + 2s(e^{a(t)J}\dot B(t),V)]\exp\left[i\int_{t'}^{t} ds\,\sigma(\dot B(s),B(s))\right]e^{-a(t)N}W(JB(t))$$
$$= [\alpha(t)N + 2s(b(t),V)]\exp\left[i\int_{t'}^{t} ds\,\sigma(\dot B(s),B(s))\right]e^{-i\alpha(t)N}W(JB(t)). \qquad (67)$$

Since $\alpha(t)N + 2s(b(t),V) = \frac{1}{2}\alpha(t)(P^2+Q^2-1) + r(t)P + s(t)Q = H(t)$, we see that $\exp[i\int_{t'}^{t} ds\,\sigma(\dot B(s),B(s))]e^{-i\alpha(t)N}W(JB(t))$ and $T\exp[-i\int_{t'}^{t} H(t)dt]$ satisfy the same first-order differential equation in t. Since both operators have the same initial value (at $t=t'$, they are both equal to $\mathbf{1}$), they are therefore equal for all times

$$\exp\left[i\int_{t'}^{t''} ds\,\sigma(\dot B(s),B(s))\right]e^{-i\alpha(t'')N}W(JB(t'')) = T\exp\left[-i\int_{t'}^{t''} H(t)dt\right].$$

This completes our proof that $F_{t'',t'}\in\mathscr{C}(U_{t'',t'})$.

Remarks: Comparing $F_{t'',t'}(v'',v')$,

$$F_{t'',t'}(v'',v')$$
$$= \frac{1}{1-e^{-(t''-t')}}\exp\left[i\int_{t'}^{t''} dt\,\sigma(\dot B(t),B(t))\right]e^{-i\sigma(v',JB(t''))}\exp[-i\sigma(e^{-a(t'')J}v'',v' + JB(t''))]$$
$$\times\omega(v' + JB(t'') - e^{-a(t'')J}v'')^{\coth[(t''-t')/2]},$$

with the true matrix element $\langle v''|T\exp[-i\int_{t'}^{t''} H(t)dt]|v'\rangle$ [see (65)], we immediately see that the two expressions are very similar; there are only two differences: an overall extra factor $[1-e^{-(t''-t')}]^{-1}$ in $F_{t'',t'}$, and an exponent $\coth[(t''-t')/2]$ for the ω-factor in $F_{t'',t'}$, where this exponent is 1 in the true matrix element. [These similarities were already noticed in Ref. 4 for the slightly simpler Hamiltonian $H(t) = \alpha(P^2+Q^2-1)/2 + s(t)Q$]. In the limit where the time integral diverges, $t''-t'\to\infty$, both these differences disappear

$$1-e^{-(t''-t')}\to 1,\quad \coth\frac{t''-t'}{2}\to 1,$$

which means that as $t''-t'\to\infty$, the function $F_{t'',t'}$ approaches the true matrix element; its component orthogonal to $\{\langle p'',q''|\phi\rangle\langle\psi|p',q'\rangle;\ \phi,\psi\in\mathscr{H}\}$ vanishes!

This is easily understood if one tries to analyze what happens for $t''-t'\to\infty$ to the construction we made in Sec. 2. As an example we take the time-independent Hamiltonian $H=(\alpha/2)(P^2+Q^2-\mathbf{1})$. Then

$$F_{t'',t'}(p'',q'';p',q') = \lim_{k\to\infty}\frac{1}{(1-\beta)^k}\sum_{l=0}^{\infty}\left[\beta^{kl}(1-\beta)^k\int dv_{k-1}\cdots dv_1\prod_{j=0}^{k-1}\langle l|W(v_{j+1})^\dagger e^{-i\alpha(N-l)\epsilon}W(v_j)|l\rangle\right.$$

and

$$= \lim_{k\to\infty}\sum_{l=0}^{\infty}\beta^{kl}\langle l|W(v'')^\dagger e^{-i\alpha(N-l)(t''-t')}W(v')|l\rangle.$$

The term corresponding to $l=0$ is simply $\langle v''|U_{t'',t'}|v'\rangle$. For $l\neq 0$, however, we also have to take into account a factor

$$\lim_{k\to\infty}\beta^{kl} = \lim_{k\to\infty}\left(1-\frac{t''-t'}{2k}\right)^{kl}\left(1+\frac{t''-t'}{2n}\right)^{-kl}$$
$$= e^{-(t''-t')l}.$$

In the limit $t''-t'\to\infty$, these factors $e^{-(t''-t')l}\to 0$ for $l\neq 0$, which means that all the contributions to $F_{t'',t'}$ from terms with $l\neq 0$ disappear. Only the $l=0$ term is left over; since this $l=0$ term is exactly $\langle v''|U_{t'',t'}|v'\rangle$, we see that

$$F_{t'',t'}(v'',v')\underset{t''-t'\to\infty}{\longrightarrow}\langle v''|U_{t'',t'}|v'\rangle.$$

In the limit where $t'' - t' \to 0$, $F_{t''-t'}(v'',v')$ approaches $\delta(v''-v')$, i.e., a specific member of the equivalence class of the unit operator (see Sec. 1). Again, this can easily be understood from our construction. For the example $H = (\alpha/2) \times (P^2 + Q^2 - \mathbb{1})$, we now have

$$F_{t'',t'}(v'',v') = \sum_l e^{-(t''-t')l'} \times \langle l \,|\, W(v'')^\dagger e^{-i\alpha(N-l)(t''-t')} W(v') | l \rangle.$$

As $t'' - t' \to 0$, we see that $e^{-(t''-t')l} \to 1$, $e^{-i\alpha(N-l)(t''-t')} \to \mathbb{1}$, and thus as a distributional limit

$$F_{t'',t'}(v'',v') \to \sum_l \langle l \,|\, W(v'')^\dagger W(v') | l \rangle = \text{"Tr"}(W(v'')^\dagger W(v')) = \delta(v''-v').$$

5. CONCLUSION

In this paper we have stressed the basic feature of overcompleteness of coherent states, and have used this fact to construct integral kernels to represent the evolution operator for a limited class of dynamical systems in the form of path integrals expressed in terms of Wiener measure.

Equation (38) presents the path integral representation for the most general one-dimension dynamical system that we are able to treat. This equation provides a novel formulation of the (equivalence-class) propagator and suggests a variety of further directions for study in addition to providing an alternative computational scheme for such propagators. However, our results are less than optimal in one sense. The necessary restriction that s and r be square integrable prohibits our results from describing local-in-time potentials when integrated over the external fields but only leads to nonlocal potentials. It is important to learn if and how this limitation can be overcome, and this problem may be clarified by using basic states other than the harmonic oscillator eigenstates. As observed in Ref. 4, the complex expression that plays the role of the classical action in the Wiener measure formulation of quantum mechanical path integrals may be formally interpreted in a natural way: The phase of the integrand is such as to form a martingale in which the phase-space motion is driven by the classical equations of motion. It is interesting to add that an entirely analogous type of construction can be given for kinematical groups other than the Heisenberg–Weyl group, and in particular for the kinematics of the SU(2) spin group.[10]

APPENDIX

We calculate the traces needed in Sec. 2:

$$\mathrm{Tr}[\beta^N W(p'',q'')^\dagger W(p',q')], \tag{A1}$$

$$\mathrm{Tr}[\beta^N W(p'',q'')^\dagger e^{-i\alpha N\epsilon} W(p',q')], \tag{A2}$$

$$\mathrm{Tr}[\beta^N W(p'',q'')^\dagger e^{-i\alpha N\epsilon} W(p',q') e^{i\alpha N\epsilon}], \tag{A3}$$

$$\mathrm{Tr}[\beta^N W(p'',q'')^\dagger e^{-i(\alpha N + sQ + rP)\epsilon} W(p',q') e^{i\alpha N\epsilon}], \tag{A4}$$

where $N = \frac{1}{2}(P^2 + Q^2 - \mathbb{1})$. The last trace (A4) corresponds to the choice $H_{kk}(t) = H(t) - \langle H(t)\rangle_k$ with $H(t) = \frac{1}{2}\alpha(t)(P^2 + Q^2 - \mathbb{1}) + s(t)Q + r(t)P$ (see the end of Sec. 2); since $\langle Q\rangle_k = \langle P\rangle_k = 0$, the term $-\langle H(t)\rangle_k$ leads to the factor $e^{i\alpha N\epsilon}$ in (A4). We start by proving two lemmas.

Lemma 1:

$$\text{(i) } e^{-itN} W(p,q) = W(p_t,q_t) e^{-itN} \tag{A5}$$

with

$$p_t = -q \sin t + p \cos t,$$

$$q_t = q \cos t + p \sin t.$$

$$\text{(ii) } e^{-it(\alpha N + sQ + rS)} = e^{(i/2\alpha)(s^2+r^2)t} W\left(-\frac{r}{\alpha}, \frac{-s}{\alpha}\right) e^{-i\alpha t N} W\left(\frac{r}{\alpha}, \frac{s}{\alpha}\right). \tag{A6}$$

Proof:

(i) This property can be proved by direct differentiation, but it can also be considered to be a consequence of the properties of homogeneous quadratic Hamiltonians in general. Indeed, it is well known that for a homogeneous quadratic Hamiltonian H, the quantum evolution of a coherent state $|p,q\rangle$ is given by the classical evolution, under the same Hamiltonian, of the labels p,q

$$e^{-itH(P,Q)}|p,q\rangle = |p_t,q_t\rangle,$$

where p_t,q_t are the solutions for the Hamiltonian equations for $H(p,q)$, with initial conditions $p_0 = p$, $q_0 = q$. Hence $e^{-itN}|p,q\rangle = |p_t,q_t\rangle$, with p_t,q_t as in (A5). Consequently

$$\begin{aligned} e^{-itN} W(p,q)|p',q'\rangle &= e^{-itN} e^{(1/2)i(pq'-p'q)} |p+p',q+q'\rangle \\ &= e^{(1/2)i(pq'-p'q)} |p_t + p'_t, q_t + q'_t\rangle \\ &= e^{(1/2)i(pq'-p'q)} e^{-(1/2)i(p_t q'_t - p'_t q_t)} W(p_t,q_t)|p'_t,q'_t\rangle \\ &= W(p_t,q_t) e^{-itN}|p',q'\rangle. \end{aligned}$$

Since the linear span of the c.s. is dense, (A5) follows.

(ii) We prove (A6) by differentiation. Take any ψ in the linear span of the c.s. Then

$$\begin{aligned} &i\frac{d}{dt}\left[e^{(i/2\alpha)(s^2+r^2)t} W\left(-\frac{r}{\alpha}, -\frac{s}{\alpha}\right) e^{-i\alpha t N} W\left(\frac{r}{\alpha},\frac{s}{\alpha}\right)\right]\psi \\ &= -\frac{1}{2\alpha}(s^2+r^2)[\cdots]\psi + e^{(i/2\alpha)(s^2+r^2)t}\, W\left(-\frac{r}{\alpha}, -\frac{s}{\alpha}\right)\frac{\alpha}{2}(P^2+Q^2-\mathbb{1}) e^{-i\alpha t N} W\left(\frac{r}{\alpha},\frac{s}{\alpha}\right)\psi \\ &= -\left\{\frac{1}{2\alpha}(s^2+r^2) + \frac{\alpha}{2}\left[\left(P+\frac{r}{\alpha}\right)^2 + \left(Q+\frac{s}{\alpha}\right)^2 - \mathbb{1}\right]\right\}[\cdots]\psi \\ &= \left[\frac{\alpha}{2}(P^2+Q^2-\mathbb{1}) + rP + sQ\right]\left[e^{(i/2\alpha)(s^2+r^2)t} W\left(-\frac{r}{\alpha} - \frac{s}{\alpha}\right) e^{-i\alpha t N} W\left(\frac{r}{\alpha},\frac{s}{\alpha}\right)\right]\psi. \end{aligned}$$

Since the linear span of the c.s. is a core for N, (A6) follows.

Note: (A6) is still true for $\alpha \to 0$; for $\alpha = 0$ the left-hand side is equal to $W(-st, -rt)$, while the right-hand side gives

$$w - \lim_{\alpha\to 0} e^{(i/2\alpha)(s^2+r^2)t} W\left(-\frac{r}{\alpha}, \frac{-s}{\alpha}\right) W\left(\frac{r_t}{\alpha}, \frac{s_t}{\alpha}\right) e^{-i\alpha t N}$$

$$= w - \lim_{\alpha\to 0} e^{(i/2\alpha)(s^2+r^2)(t-(1/\alpha)\sin\alpha t)} W\left(\frac{r_t - r}{\alpha}, \frac{s_t - s}{\alpha}\right)$$

$$= W(-st, -rt).$$

Lemma 2: For all $\gamma \in \mathbb{C}$ such that $|\gamma| < 1$: γ^N is trace-class, and

$$\mathrm{Tr}(\gamma^N W(p,q)) = \frac{1}{1-\gamma} \exp\left[-\frac{1+\gamma}{4(1-\gamma)}(p^2+q^2)\right]. \tag{A7}$$

Proof: Since γ^N has only a discrete spectrum, with eigenvalues γ^n, it is obvious that γ^N is trace-class for $|\gamma| < 1$. On the other hand, direct calculation from (A5) yields

$$\langle p'',q''|e^{-itN}|p',q'\rangle = \exp[\tfrac{1}{2}e^{-it}(p''+iq)(p'-iq') - \tfrac{1}{4}(p''^2+q''^2+p'^2+q'^2)],$$

hence, by analytic continuation (the c.s. are analytic vectors for N),

$$\langle p'',q''|\gamma^N|p',q'\rangle = \exp[(\gamma/2)(p''+iq'')(p'-iq') - \tfrac{1}{4}(p''^2+q''^2+p'^2+q'^2)].$$

Using (17) we now evaluate $\mathrm{Tr}(\gamma^N W(p,q))$:

$$\mathrm{Tr}(\gamma^N W(p,q)) = \int \frac{dp'dq'}{2\pi}\langle p',q'|\gamma^N W(p,q)|p',q'\rangle = \int \frac{dp'dq'}{2\pi} e^{(i/2)(pq'-p'q)}\langle p',q'|\gamma^N|p+p',q+q'\rangle$$

$$= \int \frac{dp'dq'}{2\pi}\exp\left\{\frac{i(\gamma+1)}{2}(pq'-qp') - \frac{1-\gamma}{2}[(p'+p/2)^2+(q'+q/2)^2] - \frac{1+\gamma}{8}(p^2+q^2)\right\}$$

$$= \frac{1}{1-\gamma}\exp\left[-\frac{1+\gamma}{8}(p^2+q^2) - \frac{(1+\gamma)^2}{8(1-\gamma)}(p^2+q^2)\right] = \frac{1}{1-\gamma}\exp\left[-\frac{1+\gamma}{4(1-\gamma)}(p^2+q^2)\right].$$

Using this result it is now very easy to calculate the traces (A1)$\to$(A4). Since (A1) can be obtained from (A2) by taking the limit $\alpha \to 0$, and (A3) from (A4) by putting $r = s = 0$, we shall only evaluate (A2) and (A4) explicitly. For (A2) we get

$$\mathrm{Tr}[\beta^N W(p'',q'')^\dagger e^{-i\alpha N\epsilon} W(p',q')] = \mathrm{Tr}[\beta^N W(p'',q'')^\dagger W(p'_\alpha,q'_\alpha) e^{-i\alpha N\epsilon}]$$

[with $p'_\alpha = p'\cos\alpha\epsilon - q'\sin\alpha\epsilon, \quad q'_\alpha = p'\sin\alpha\epsilon + q'\cos\alpha\epsilon$]

$$= e^{(i/2)(p'_\alpha q'' - q'_\alpha p'')}\mathrm{Tr}[(\beta e^{-i\alpha\epsilon})^N W(p'_\alpha - p'', q'_\alpha - q'')]$$

$$= \frac{1}{1-\beta e^{-i\alpha\epsilon}} e^{(i/2)(p'_\alpha q'' - q'_\alpha p'')}\exp\left\{-\frac{1+\beta e^{-i\alpha\epsilon}}{4(1-\beta e^{-i\alpha\epsilon})}[(p''-p'_\alpha)^2 + (q''-q'_\alpha)^2]\right\}. \tag{A8}$$

This is exactly what was used in (35); in the limit for $\alpha \to 0$, we have

$$\mathrm{Tr}[\beta^N W(p'',q'')^\dagger W(p',q')] = \frac{1}{1-\beta} e^{(i/2)(p'q''-p''q')}\exp\left\{-\frac{1+\beta}{4(1-\beta)}[(p''-p')^2+(q''-q')^2]\right\}, \tag{A9}$$

which yields (24). The evaluation of (A4) gives

$$\mathrm{Tr}[\beta^N W(p'',q'')^\dagger e^{-i(\alpha N + sQ + rP)\epsilon} W(p',q') e^{i\alpha N\epsilon}]$$

$$= \mathrm{Tr}\left[\beta^N W(-p'',-q'') e^{(i/2\alpha)(s^2+r^2)\epsilon} W\left(-\frac{r}{\alpha},-\frac{s}{\alpha}\right) e^{-i\alpha N\epsilon} W\left(\frac{r}{\alpha},\frac{s}{\alpha}\right) W(p',q') e^{i\alpha N\epsilon}\right]$$

$$= \exp\left[\frac{i}{2\alpha}(s^2+r^2)\epsilon + \frac{i}{2\alpha}(p''s - q''r) + \frac{i}{2\alpha}(rq'-sp')\right]\mathrm{Tr}\left[\beta^N W\left(-p''-\frac{r}{\alpha}, -q''-\frac{s}{\alpha}\right)\right.$$

$$\left.\times e^{-i\alpha N\epsilon} W\left(p'+\frac{r}{\alpha}, q'+\frac{s}{\alpha}\right) e^{i\alpha N\epsilon}\right]$$

$$= \exp\left[\frac{i}{2\alpha}(s^2+r^2)\epsilon + \frac{i}{2\alpha}(p''s-q''r) + \frac{i}{2\alpha}(rq'-sp')\right]$$

$$\times \mathrm{Tr}\left\{\beta^N W\left(-p''-\frac{r}{\alpha}, -q''-\frac{s}{\alpha}\right) W([p'+r/\alpha]\cos\alpha\epsilon - [q'+s/\alpha]\sin\alpha\epsilon,\right.$$

$$[p'+r/\alpha]\sin\alpha\epsilon+[q'+s/\alpha]\cos\alpha\epsilon)\Big\}$$

$$=\exp\left[\frac{i}{2\alpha}(s^2+r^2)\epsilon+\frac{i}{2\alpha}(p''s-q''r)+\frac{i}{2\alpha}(rq'-sp')\right]\cdot\frac{1}{1-\beta}$$
$$\times\exp((i/2)\{q''+s/\alpha)[(p'+r/\alpha)\cos\alpha\epsilon-(q'+s/\alpha)\sin\alpha\epsilon]\}-(p''+r/\alpha)[(p'+r/\alpha)\sin\alpha\epsilon+(q'+s/\alpha)\cos\alpha\epsilon]\}$$
$$-\frac{1+\beta}{4(1-\beta)}\{[(p''+r/\alpha)-(p'+r/\alpha)\cos\alpha\epsilon+(q'+s/\alpha)\sin\alpha\epsilon]^2$$
$$+[(q''+s/\alpha)-(p'+r/\alpha)\sin\alpha\epsilon-(q'+s/\alpha)\cos\alpha\epsilon]^2\}) \qquad \text{(A10)}$$

[use (A5) and (A9)]. Putting $r=s=0$, this leads to (35).

Finally, we show how, under the assumption $\beta=(1-\epsilon/2)/(1+\epsilon/2)$, and in the approximation that ϵ is small, (A10) leads to (38). (A10) becomes now

$$\exp\left[\frac{i}{2\alpha}(s^2+r^2)+\frac{i}{2\alpha}(p''s-q''r)+\frac{i}{2\alpha}(rq'-sp')\right]\frac{1}{1-\beta}$$
$$\times\exp\{(i/2)[(q''+s/\alpha)(p'+r/\alpha)-(q''+s/\alpha)(q'+s/\alpha)\alpha\epsilon-(p''+r/\alpha)(p'+r/\alpha)\alpha\epsilon-(p''+r/\alpha)(q'+s/\alpha)]\}$$
$$\times\exp(-(1/2\epsilon)\{[p''-p'+(q'+s/\alpha)\alpha\epsilon]^2+[(q''-q')-(p'+r/\alpha)\alpha\epsilon]^2\})$$
$$=\frac{1}{1-\beta}\exp\{-(1/2\epsilon)[(p''-p')^2+(q''-q')^2]+(i/2+\alpha)[p'(q''-q')-q'(p''-p')]$$
$$-(\alpha^2/2)(q'^2+p'^2)\epsilon-(i\alpha/2)(p''p'+q''q')\epsilon-(i/2)[s(q''+q')+r(p''+p')]\epsilon$$
$$-\alpha(sq'+rp')\epsilon-s(p''-p')+r(q''-q')-\tfrac{1}{2}(s^2+r^2)\epsilon\}.$$

Putting $p_{j+1}=p''$, $q_{j+1}=q''$ and $p_j=p'$, $q_j=q'$, and assuming $p_{j+1}-p_j=O(\sqrt{\epsilon})$, $q_{j+1}-q_j=O(\sqrt{\epsilon})$, we finally obtain

$$\mathrm{Tr}\{\beta^N W(p_{j+1},q_{j+1})^\dagger e^{-i[\alpha(t_j)N+r(t_j)P+s(t_j)Q]\epsilon}W(p_j,q_j)e^{i\alpha(t_j)N\epsilon}\}$$
$$=(1+\epsilon/2)\frac{1}{\epsilon}\exp\{-\frac{1}{2\epsilon}[(p_{j+1}-p_j)^2+(q_{j+1}-q_j)^2]\}$$
$$+[i/2+\alpha(t_j)][p_j(q_{j+1}-q_j)-q_j(p_{j+1}-p_j)]-\tfrac{1}{2}\alpha(t_j)[i+\alpha(r_j)][p_j^2+q_j^2]\epsilon$$
$$-[s(t_j)(p_{j+1}-p_j)-r(t_j)(q_{j+1}-q_j)]-[i+\alpha(t_j)][s(t_j)p_j+r(t_j)q_j]\epsilon-\tfrac{1}{2}[s(t_j)^2+r(t_j)^2]\epsilon+O(\epsilon^{3/2}), \qquad \text{(A11)}$$

which can easily be seen to lead to (38).

[1]See, e.g., J. R. Klauder, Acta Phys. Austriaca, Suppl. XXII, 3 (1980).

[2]S. Albeverio and R. Høegh-Kröhn, *Mathematical Theory of Feynman Path Integrals*, Lecture Notes in Mathematics, No. 523 (Springer, Berlin, 1976).

[3]Ph. Combe, R. Høegh-Kröhn, R. Rodrigues, M. Sirugue, and M. Sirugue-Collin, Commun. Math. Phys. **77**, 269 (1980); J. Math. Phys. **23**, 405 (1982); see also *Proceedings of the International Workshop "Functional Integration: Theory and Applications,"* edited by J. P. Antoine and E. Tirapegui (Plenum, New York, 1980).

[4]J. R. Klauder and I. Daubechies, Phys. Rev. Lett. **48**, 117 (1982); see also "Wiener Measures for Quantum Mechanical Path Integrals," in *Proceedings of the International Workshop Stochastic Processes in Quantum Theory and Statistical Physics: Recent Progress and Applications*, Lecture Notes in Physics (Springer, New York) (to be published).

[5]See, e.g., J. R. Klauder and E. C. G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968), Chap. 7.

[6]J. McKenna and J. R. Klauder, J. Math. Phys. **5**, 878 (1964).

[7]See, e.g., B. Simon, *Functional Integration and Quantum Mechanics*, (Academic, New York, 1979).

[8]D. Kastler, Commun. Math. Phys. **1**, 14 (1965).

[9]A. Grossmann, Commun. Math. Phys. **48**, 191 (1976).

[10]J. R. Klauder, "Constructing measures for spin-variable path integrals," J. Math. Phys. **23**, 1797 (1982).

Properties of solutions for N-body Yakubovskii–Faddeev equations

P. Benoist-Gueutal and M. L'Huillier
Division de Physique Théorique,[a] *Institut de Physique Nucléaire, F-91406 Orsay Cedex, France*

(Received 14 April 1981; accepted for publication 30 July 1981)

We give a revised presentation of the Yakubovskii–Faddeev formalism based on a systematic study of the N-body system chain structure. Completeness properties of the corresponding equations in differential form are considered. The expressions of physical and spurious solutions are given in terms of the N-body asymptotic partition Hamiltonians eigenvectors.

PACS numbers: 03.65.Ge

I. INTRODUCTION

Consider a nonrelativistic system of N particles interacting via short range two-body potentials.

The method proposed by Faddeev[1] and Yakubovskii[2] to calculate the bound states and scattering states, consists in decomposing the wave function into components corresponding to all various chains of partitions of the N-body system.

Partition a_p, $1 \leqslant p \leqslant N$, is a way of grouping the N particles into p disjoint subsets called clusters. If a_p can be obtained by fusing the clusters of another partition b_r we write $b_r \subset a_p$. A chain of embedded partitions will be denoted

$$A_p^r = \{a_r \subset a_{r-1} \subset \cdots \subset a_{p+1} \subset a_p\}, \quad 1 \leqslant p < r \leqslant N. \tag{1.1}$$

Notations $A_{a_p}^r$, $A_p^{a_r}$, or $A_{a_p}^{a_r}$ will be used to display one or both of the external partitions of the chain. We eventually simplify notation by omitting upper indices N and a_N, and lower indices 1 and a_1 since the total breakup partition a_N and the one-cluster partition a_1 are unique

$$A^r \equiv A_1^r, \quad A_p \equiv A_p^N, \quad A \equiv A_1^N. \tag{1.2}$$

Let V_{a_p}, $1 \leqslant p \leqslant N-1$ be the internal interaction of partition a_p, that is, the sum of all two-body potentials internal to the p clusters of a_p.

The Yakubovskii–Faddeev (YF) equations can be derived by defining interactions $(v_p)_{AB}$ which realize a chain decomposition of V_{a_p}

$$\sum_{A_{a_p}} (v_p)_{AB} = V_{a_p} \delta_{A^{p} B^{p}}, \tag{1.3}$$

where the summation runs over chains A_{a_p} contained in a_p. Equation (1.3) does not uniquely define $(v_p)_{AB}$. The YF-formalism corresponds to the choice

$$(v_p)_{AB} = V_{a_{N-1}} (Y_p^N)_{AB}, \tag{1.4}$$

where $V_{a_{N-1}}$ is the pair interaction internal to the partition a_{N-1} contained in chain A. Numbers $(Y_p^N)_{AB}$ are either 0 or 1 and correspond to a specific correlation between chains A and B. They are defined in Eq. (2.13) and their properties are studied in Appendix A, Sec. 3.

The partition Hamiltonian $H_{a_p} = H_0 + V_{a_p}$, where H_0 is the kinetic energy operator with center-of-mass energy removed, can be decomposed using (1.3) and (1.4) as

$$H_{a_p} \delta_{A^{p} B^{p}} = \sum_{A_{a_p}} (h_p)_{AB}, \tag{1.5}$$

$$(h_p)_{AB} = H_0 \delta_{AB} + (v_p)_{AB}. \tag{1.6}$$

Let $\mathscr{H}$ be the Hilbert space spanned by column vectors $|\Psi\rangle$ with components $|\psi_A\rangle$ in the Hilbert space $\mathscr{H}_N$ of the N-body system, where A runs over the total set of chains. The scalar product and norm on $\mathscr{H}$ are defined by

$$\langle \varphi | \Psi \rangle = \sum_A \langle \varphi_A | \psi_A \rangle = \langle \Psi | \varphi \rangle^*, \tag{1.7}$$

where $\langle \varphi_A | \psi_A \rangle$ is a scalar product on $\mathscr{H}_N$. The set $(h_p)_{AB}$ defines a non-self-adjoint operator h_p on $\mathscr{H}$.

The YF equations are usually considered in an integral form which results from the existence of the resolvent operator

$$g_p(z) = (z - h_p)^{-1}, \tag{1.8}$$

z being a complex number. The properties of $g_p(z)$ follow from the spectral properties of h_p. They are studied from the equation on $\mathscr{H}$

$$(z - h_p)|\Psi\rangle = 0, \tag{1.9a}$$

called a YF equation in differential form. This equation has the structure of a system of coupled equations on $\mathscr{H}_N$

$$z|\psi_A\rangle - \sum_B (h_p)_{AB} |\psi_B\rangle = 0. \tag{1.9b}$$

Summing (1.9b) over all chains A which contain a_p, one gets through (1.5)

$$(z - H_{a_p}) \sum_{A^{a_p}} \sum_{A_{a_p}} |\psi_A\rangle = 0. \tag{1.10}$$

It follows that either $\Sigma_{A^{a_p}} \Sigma_{A_{a_p}} |\psi_A\rangle$ vanishes or it satisfies the eigenvalue equation of H_{a_p}. The solutions of Eq. (1.9) (and the corresponding spectra) will be called "spurious" in the former case and "physical" in the latter case. Obviously, the physical spectrum is real. Evans and Hoffman[3,4] consider the properties of the solution of (1.9) in the three-body case. They show that the spurious spectrum is real and prove the completeness property of the physical and spurious solutions on the chain-space $\mathscr{H}$. In the present paper we extend the results of Evans and Hoffman to the N-body case. The spectral properties of h_p depend on the nature of the two-body interactions and a rigorous proof of these properties is difficult. Our study relies on the assumption of completeness

[a]Laboratoire associé au C.N.R.S.

of all partition Hamiltonians H_{a_p} and of existence, for $\mathrm{Im}\, z \neq 0$, of the resolvent operator

$$G_{a_p}(z) = (z - H_{a_p})^{-1}. \tag{1.11}$$

The present paper is organized as follows: We first review the YF formalism. A systematic study of the N-body system chain structure, of the operators Y_p^N introduced in Eq. (1.4) and related correlation operators clarifies the presentation. This part, which contains no dynamical aspects and is relevant to any N-body system, is discussed in Sec. II and Appendices A and B. In Sec. III we define Hilbert chain spaces $\mathscr{H}_{a_p}^{a_r}$ and chain Hamiltonians h_p. We make explicit the structure of the matrix elements $[g_p(z)]_{AB}$ of the resolvent operator defined by Eq. (1.8) in terms of the operators

$$\Gamma_{A_q}(z) = G_0(z) V_{a_{N-1}} G_{a_{N-1}}(z) V_{a_{N-2}}^{a_{N-1}} \cdots G_{a_{q+1}}(z) V_{a_q}^{a_{q+1}} G_{a_q}(z), \tag{1.12a}$$

$$V_{a_q}^{a_{q+1}} = V_{a_q} - V_{a_{q+1}}. \tag{1.12b}$$

This result is of interest to understand the structure of the YF equations in integral form and to study the asymptotic properties of their solutions.

In Sec. IV spectral properties of chain Hamiltonians $h_p, 1 \leqslant p \leqslant N$, are studied. We present in Sec. IV B some properties of the solutions of the Schrödinger equation

$$(E - H_{a_p}) |\psi\rangle = 0, \tag{1.13}$$

which follow from the initial assumption of asymptotic completeness on the Hilbert space $\mathscr{H}_N$. Consequently we prove there exists a one-to-one correspondence between the solutions of (1.13) and a choice of physical eigenvectors of the restriction of h_p to the Hilbert space $\mathscr{H}_{a_p}$. The explicit relation between the chain components of a physical eigenvector and the corresponding eigenvector of H_{a_p} is given. In Sec. IV C we construct spurious eigenvectors of h_p from physical eigenvectors of the subsystem chain Hamiltonian $h_q, p+2 \leqslant q \leqslant N$. Thus we are in position to prove that the set of physical and spurious eigenvectors of h_p forms a complete basis in the chain space $\mathscr{H}$ and that the corresponding eigenvalues are real. Chain Hamiltonian h_p being not self-adjoint, this basis is not orthogonal. We show how to construct a biorthogonal basis and its relation to the eigenvectors of adjoint chain Hamiltonian $\tilde{h}_p$.

II. CHAIN SPACE $\mathbb{C}_{a_p}^{a_r}$

A. Definitions

Let a_p and b_r be two partitions of the N-particle system. We use the symbol

$$\delta_{a_p}^{b_r} = 1 \text{ if } b_r \text{ is contained in or is identical to } a_p, \quad = 0 \text{ otherwise}. \tag{2.1}$$

When $r = p$, $\delta_{a_p}^{b_r}$ is identical to the Kronecker symbol $\delta_{a_p b_p}$. The symbol

$$\bar{\delta}_{a_p b_p} = 1 - \delta_{a_p b_p} \tag{2.2}$$

will also be used.

The chain of partitions A_p^r is defined in Eq. (1.1). Let

$$\delta_{A_p^r B_p^r} = \delta_{a_r b_r} \prod_{i=p}^{r-1} \delta_{a_i b_i} \delta_{a_i}^{a_{i+1}} \delta_{b_i}^{b_{i+1}}; \tag{2.3}$$

then $\delta_{A_p^r B_p^r} = 1$ means chains A_p^r and B_p^r are identical and $\delta_{A_p^r B_p^r} = 0$ means they differ by at least one partition. Let $\mathscr{C}_{a_p}^{a_r} = \{A_{a_p}^{a_r}, B_{a_p}^{a_r} \cdots\}$ be the set of $n_{a_p}^{a_r}$ chains A_p^r which contain a_r and are contained in a_p, classified in a specific order. (Formulas related to the counting of chains are given in Appendix A1.) Let $\mathbb{C}_{a_p}^{a_r}$ be the set of the applications of $\mathscr{C}_{a_p}^{a_r}$ on the complex number field. A vector $|F\rangle$ of $\mathbb{C}_{a_p}^{a_r}$ is a sequence of $n_{a_p}^{a_r}$ complex numbers $F_{A_{a_p}^{a_r}}, F_{B_{a_p}^{a_r}}, \cdots$ which can be written as a column matrix. The space $\mathbb{C}_{a_p}^{a_r}$ is a complex vectorial space of dimension $n_{a_p}^{a_r}$. In particular the dimension of $\mathbb{C} \equiv \mathbb{C}_{a_1}^{a_N}$ is the total number n of chains.

The norm and scalar product on $\mathbb{C}_{a_p}^{a_r}$ are defined by

$$\langle G | F \rangle = \sum_{A_{a_p}^{a_r}} G^*_{A_{a_p}^{a_r}} F_{A_{a_p}^{a_r}} = \langle F | G \rangle^*. \tag{2.4}$$

B. Operators on $\mathbb{C}_{a_p}^{a_r}$

Operators O on $\mathbb{C}_{a_p}^{a_r}$ are represented by $n_{a_p}^{a_r} \times n_{a_p}^{a_r}$ matrices of complex numbers $O_{A_{a_p}^{a_r} B_{a_p}^{a_r}}$. The adjoint operator $\tilde{O}$ is defined by

$$\tilde{O}_{A_{a_p}^{a_r} B_{a_p}^{a_r}} = (O_{B_{a_p}^{a_r} A_{a_p}^{a_r}})^*. \tag{2.5}$$

1. $\binom{r}{p}$ operators

Given $1 \leqslant p < r \leqslant N$, $\binom{r}{p}$ operators are defined on the chain space $\mathbb{C}$ by matrix elements in the form

$$(O_p^r)_{AB} = \delta_{A_r B_r} (O_p^r)_{A_p^r B_p^r} \delta_{A^p B^p}. \tag{2.6}$$

It results from this definition that one can consider the restriction of O_p^r, denoted by the same symbol, to any subspace $\mathbb{C}_{a_q}^{a_s}$ with $q \leqslant p < r \leqslant s$,

$$(O_p^r)_{A_{a_q}^{a_s} B_{a_q}^{a_s}} = \delta_{A_r^{a_s} B_r^{a_s}} (O_p^r)_{A_p^r B_p^r} \delta_{A_{a_q}^p B_{a_q}^p}. \tag{2.7}$$

In particular,

$$(O_p^r)_{A_{a_p}^{a_r} B_{a_p}^{a_r}} = \delta_{a_r b_r} (O_p^r)_{A_p^r B_p^r} \delta_{a_p b_p}. \tag{2.8}$$

Eventually the upper index will be suppressed in the notation of $\binom{r}{p}$ operators if $r = N$.

The $\binom{r}{p}$ operators satisfy the commutation relations

$$O_p^q O_r^s = O_r^s O_p^q \quad \text{if} \quad p < q \leqslant r < s \quad \text{or} \quad r < s \leqslant p < q. \tag{2.9}$$

2. Chain correlation operators

In order to introduce in Sec. III the YF chain decomposition of the interactions of the N-particle system, we study in Appendix A correlation operators O on $\mathbb{C}$, the elements O_{AB} of which are equal to 0 or 1. The value $O_{AB} = 1$ implies the realization of a definite correlation between chains A and B. In the present section we summarize the main results of interest for the analysis of the solutions of the YF equation in differential form.

Summation operators S_p^r are defined for $1 \leqslant p < r \leqslant N$ by

$$(S_p^r)_{AB} = \delta_{A^p B^p} \delta_{A_r B_r}. \tag{2.10}$$

They are self-adjoint

$$\tilde{S}_p^r = S_p^r. \tag{2.11}$$

In particular,

$$(S_p^{p+1})_{AB} = \delta_{AB},\ (S_p^N)_{AB} = \delta_{A^pB^p},\ (S_1^r)_{AB} = \delta_{A_rB_r}. \tag{2.12}$$

We define for $1 \leqslant p < q \leqslant N$ Yakubovskii operators Y_p^r through recurrence relations

$$Y_p^{p+1} = 1,\ Y_p^{p+2} = S_p^{p+2},\ Y_p^r = 1 + (Y_{r-2}^r - 1)Y_p^{r-1}. \tag{2.13}$$

One deduces the factorization property

$$Y_p^r - Y_q^r = (Y_{q-1}^r - Y_q^r)Y_p^q,\quad p < q < r \leqslant N. \tag{2.14}$$

Given a chain B and any partition a_{r-1} with $\delta_{b_p}^{a_{r-1}} = \delta_{a_{r-1}}^{b_r} = 1$, there exists one and only one chain A containing a_{r-1} in such a way that $(Y_p^r)_{AB} = 1$. This fundamental property of Y_p^r is expressed by the relation

$$S_p^{r-1}Y_p^r = S_p^r. \tag{2.15}$$

III. HILBERT CHAIN SPACE $\mathscr{H}_{a_p}^{a_r}$

A. Definition

Let $\mathscr{H}_{a_p}^{a_r}$ be the vectorial space spanned by the set of the applications $|\Psi\rangle$ of $\mathscr{C}_{a_p}^{a_r} = \{A_{a_p}^{a_r}, B_{a_p}^{a_r}, \cdots\}$ on the Hilbert space $\mathscr{H}_N$ of the N-body system. Vector $|\Psi\rangle$ is a column of $n_{a_p}^{a_r}$ components $|\psi_{A_{a_p}^{a_r}}\rangle, |\psi_{B_{a_p}^{a_r}}\rangle, \cdots$ which are vectors of $\mathscr{H}_N$.

Space $\mathscr{H}_{a_p}^{a_r}$ can be considered as the tensor product

$$\mathscr{H}_{a_p}^{a_r} = \mathscr{H}_N \otimes \mathbb{C}_{a_p}^{a_r}. \tag{3.1}$$

The norm and scalar product on $\mathscr{H}_{a_p}^{a_r}$ are defined by

$$\langle\varphi|\Psi\rangle = \sum_{A_{a_p}^{a_r}} \langle\varphi_{A_{a_p}^{a_r}}|\psi_{A_{a_p}^{a_r}}\rangle = \langle\Psi|\varphi\rangle^*, \tag{3.2}$$

where $\langle\varphi_{A_{a_p}^{a_r}}|\psi_{A_{a_p}^{a_r}}\rangle$ is a scalar product on $\mathscr{H}_N$.

If $\mathscr{H}_N$ is separable, $\mathscr{H}_{a_p}^{a_r}$ is a separable Hilbert space. Operators on $\mathscr{H}_{a_p}^{a_r}$ are represented by $n_{a_p}^{a_r} \times n_{a_p}^{a_r}$ matrices of operators on $\mathscr{H}_N$. Definition (2.6) of $\binom{r}{p}$ operators can be extended to the full Hilbert chain space $\mathscr{H} \equiv \mathscr{H}_{a_1}^{a_N}$ and to $\mathscr{H}_{a_q}^{a_s}$, with $q \leqslant p < r \leqslant s$.

B. Chain Hamiltonians

Let us define v_p operators on $\mathscr{H}$ for $1 \leqslant p \leqslant N$ by

$$v_N = 0,\ (v_{N-1})_{AB} = V_{a_{N-1}}\delta_{AB},\ v_p = v_{N-1}Y_p^N. \tag{3.3}$$

Operator v_p is a $\binom{N}{p}$ operator on $\mathscr{H}$, non-Hermitian if $p \leqslant N-2$. Elements $(v_p)_{AB}$ are 0 or $V_{a_{N-1}}$ according to the correlation property of Y_p^N.

We define chain Hamiltonians h_p for $1 \leqslant p \leqslant N$ by

$$(h_N)_{AB} = H_0\delta_{AB},\ h_p = h_N + v_p. \tag{3.4}$$

Using Eqs. (2.10) and (2.15) one gets

$$\begin{aligned}(S_p^N v_p)_{AB} &= \sum_C \delta_{A^pC^p} V_{c_{N-1}}(Y_p^N)_{CB} = \delta_{A^pB^p}\sum_C V_{c_{N-1}}(Y_p^N)_{CB} \\ &= \delta_{A^pB^p}\sum_{c_{N-1}} V_{c_{N-1}}\delta_{c_{N-1}}^{b_p} \\ &= \delta_{A^pB^p}V_{a_p}\end{aligned} \tag{3.5}$$

Similarly,

$$(S_p^N h_p)_{AB} = \delta_{A^pB^p}H_{a_p}. \tag{3.6}$$

Operators v_p and h_p realize the YF chain decomposition of the internal interactions and of the partition Hamiltonians of the p-cluster partitions.

Using Eq. (3.3) and (2.14) with $r = N$, one obtains

$$v_p - v_q = (v_{q-1} - v_q)Y_p^q,\quad p < q \leqslant N, \tag{3.7}$$

and from (3.4)

$$h_p = h_q + (v_{q-1} - v_q)Y_p^q,\quad p < q \leqslant N. \tag{3.8}$$

C. Green's operators

Let z be a complex number belonging to the resolvent of every Hamiltonian $h_p, 1 \leqslant p \leqslant N$. Then there exists a resolvent operator

$$g_p(z) = (z - h_p)^{-1}. \tag{3.9}$$

Left multiplying (3.9) by $S_p^N(z - h_p)$ one gets from Eq. (3.6)

$$\begin{aligned}[S_p^N(z - h_p)g_p(z)]_{AB} &= (z - H_{a_p})[S_p^N g_p(z)]_{AB} \\ &= (S_p^N)_{AB} = \delta_{A^pB^p}.\end{aligned}$$

Left multiplying by $G_{a_p}(z) = (z - H_{a_p})^{-1}$ one obtains

$$[S_p^N g_p(z)]_{AB} = G_{a_p}(z)\delta_{A^pB^p}, \tag{3.10a}$$

which is equivalent to

$$\sum_A [g_p(z)]_{AB} = G_{b_p}(z). \tag{3.10b}$$

Thus, $g_p(z)$ can be interpreted as a chain decomposition of the p-cluster partition Green functions. It satisfies the resolvent identities

$$g_p = g_q + g_q(v_p - v_q)g_p, \tag{3.11a}$$

$$\quad = g_q + g_p(v_p - v_q)g_q, \tag{3.11b}$$

where z has been omitted for notational convenience. First we shall use (3.10) and (3.11b) to obtain relations between the partition Green's functions which may be of use.

Upon iterative use of (3.11b) for $q = r, r-1, \cdots, p+1$, one obtains

$$g_p = \sum_{q=p}^{r} g_p^{(r,q)}, \tag{3.12}$$

with $g_p^{(r,q)}$ recursively defined by

$$g_p^{(r,r)} = g_r,\quad g_p^{(r,q)} = g_q(v_p - v_{q+1})g_p^{(r,q+1)}. \tag{3.13}$$

Using (3.7) and the commutation relations (2.9) one gets

$$g_p^{(r,q)} = \gamma_q^{(r)} Z_p^{r(q)}, \tag{3.14}$$

where $Z_p^{r(q)}$ is a correlation operator studied in Appendix A3 e

$$Z_p^{r(q)} = Y_p^{q+1}Y_p^{q+2}\cdots Y_p^{r-1}Y_p^r, \tag{3.15}$$

and $\gamma_q^{(r)}$ is recursively defined by

$$\gamma_r^{(r)} = g_r,\quad \gamma_q^{(r)} = g_q(v_q - v_{q+1})\gamma_{q+1}^{(r)}. \tag{3.16}$$

Using (3.5) and (3.10) one obtains

$$\sum_A (\gamma_q^{(r)})_{AC} = \tilde{\Gamma}_{c_q^r}, \tag{3.17}$$

$$\tilde{\Gamma}_{c_q^r}(z) = G_{c_q}(z)V_{c_q}^{c_{q+1}}G_{c_{q+1}}(z)V_{c_{q+1}}^{c_{q+2}}\cdots G_{c_{r-1}}(z)V_{c_{r-1}}^{c_r}G_{c_r}(z). \tag{3.18}$$

From (3.14), using formulas (A39) and (3.17) one obtains

$$\sum_A [g_p^{(r,q)}]_{AB} = \sum_A \sum_C (\gamma_q^{(r)})_{AC} (Z_p^{r(q)})_{CB} = \sum_{b_q} \delta_{b_p}^{b_q} \sum_{B_{b_q}^{b_r}} \widetilde{\Gamma}_{B_q^r}, \quad (3.19)$$

and from (3.12) and (3.10b)

$$G_{b_p}(z) = \sum_{q=p}^{r} \sum_{b_q} \delta_{b_p}^{b_q} \sum_{B_{b_q}^{b_r}} \widetilde{\Gamma}_{B_q^r}(z). \quad (3.20)$$

The Green's operator $G_{b_p}(z)$ satisfies

$$G_{b_p}(z) = \widetilde{G}_{b_p}(z^*). \quad (3.21)$$

From (3.18) and (3.20) one deduces

$$G_{b_p}(z) = \sum_{q=p}^{r} \sum_{b_q} \delta_{b_p}^{b_q} \sum_{B_{b_q}^{b_r}} \Gamma_{B_q^r}(z), \quad (3.22)$$

$$\Gamma_{B_q^r}(z) = G_{b_r}(z) V_{b_{r-1}}^{b_r} G_{b_{r-1}}(z) \cdots G_{b_{q+1}}(z) V_{b_q}^{b_{q+1}} G_{b_q}(z). \quad (3.23)$$

Relations (3.20) and (3.22) are algebraical indentities. For a given $G_{b_p}(z)$ they can be satisfied for any partition b_r contained in b_p and involve the resolvents corresponding to the partitions of the various chains. Of particular use are the relations (3.20) and (3.23) with the choice $b_r = b_N$, which read

$$G_{b_p}(z) = \sum_{q=p}^{N} \sum_{b_q} \delta_{b_p}^{b_q} \overline{G}_{b_q}(z), \quad (3.24)$$

$$\overline{G}_{b_q}(z) = \sum_{B_{b_q}} \widetilde{\Gamma}_{B_q}(z) = \sum_{B_{b_q}} \Gamma_{B_q}(z). \quad (3.25)$$

and correspond to the decomposition of $G_{b_p}(z)$ into a sum of operators of definite connectivity.

Relation (3.12) will now be used with $r = N$ to obtain the expression of the matrix elements $(g_p)_{AB}$ in terms of partition Green's operators. To simplify the notation index N will be omitted in the following.

$$g_p = \sum_{q=p}^{N} g_p^{(q)}. \quad (3.26)$$

Using Eqs. (3.13) and (3.11) we find

$$g_p^{(p)} = g_p(v_p - v_{p+1}) g_{p+1}(v_p - v_{p+2}) g_{p+2} \cdots g_{N-1} v_p g_N,$$
$$= g_{p+1}(v_p - v_{p+1}) g_{p+2}(v_p - v_{p+2}) g_{p+3} \cdots g_N v_p g_p.$$

Using Eqs. (3.7), (2.9), and (A36) one obtains

$$g_p^{(p)} = g_{p+1}(v_p - v_{p+1}) g_{p+2}(v_{p+1} - v_{p+2}) g_{p+3} \cdots g_N v_{N-1} S_p^N g_p. \quad (3.27)$$

From (3.7) and (A18) we deduce

$$(v_{p+r} - v_{p+r+1}) = (v_{p+r+1} - v_{p+r+2}) I_{p+r}^{p+r+2} \quad (3.28)$$

Substituting in (3.27) and using Eqs. (2.9), (3.16), (A22), (A30), and (3.3) one obtains

$$g_p^{(p)} = \gamma_{p+1}(\tilde{v}_p - \tilde{v}_{p+1}) S_p^N g_p, \quad (3.29)$$

which implies, from (3.10a) and (3.5),

$$(g_p^{(p)})_{AB} = \sum_C (\gamma_{p+1})_{AC} V_{a_p}^{a_{p+1}} G_{a_p} \delta_{A^p B^p}. \quad (3.30)$$

On the other hand, from (3.14) and (A38)

$$(g_p^{(p)})_{AB} = \sum_C (\gamma_p)_{AC} \delta_{A^p B^p}, \quad (3.31)$$

which gives with (3.30)

$$\sum_C (\gamma_p)_{AC} = \sum_C (\gamma_{p+1})_{AC} V_{a_p}^{a_{p+1}} G_{a_p}.$$

It is then straightforward by recursion to obtain

$$\sum_C (\gamma_p)_{AC} = \Gamma_{A_p}, \quad (3.32)$$

where Γ_{A_p} is given by (3.23).

Since matrix element $(Z_p^{(q)})_{CB}$ does not depend on C_{q+1} one obtains from (3.14) and (3.32)

$$(g_p^{(q)})_{AB} = \sum_C (\gamma_q)_{AC} (Z_p^{(q)})_{CB} = (Z_p^{(q)})_{AB} \sum_C (\gamma_q)_{AC}, = (Z_p^{(q)})_{AB} \Gamma_{A_q}, \quad (3.33)$$

which gives with Eqs. (3.26) and (3.23) the explicit expression of $(g_p(z))_{AB}$ in terms of the Green's operators $G_{a_q}(z)$ corresponding to the partitions a_q contained in a_p. We remark from (3.32) that $(g_p(z))_{AB}$ depends on B through purely geometrical factors $(Z_p^{(q)})_{AB}$.

Definition (3.23) displays the fact that $\Gamma_{A_q}(z)$ is an a_q-connected operator. Then the decomposition (3.26) of $g_p(z)$ corresponds to the decomposition of the matrix elements $(g_p(z))_{AB}$ into a sum of operators of definite connectivity. From (3.17), (A39), and (3.25) one gets

$$\sum_A (g_p^{(q)})_{AB} = \sum_{a_q} \delta_{b_p}^{a_q} \overline{G}_{a_q}(z). \quad (3.34)$$

The self-adjoint partition Hamiltonian H_{a_p} has a real spectrum and $G_{a_p}(z) = (z - H_{a_p})^{-1}$ is defined for any z with $\text{Im} z \neq 0$. Equations (3.26), (3.33), and (3.23) imply that $g_p(z)$ is defined for any nonreal z number. The spectrum of operators h_p is then real. This result will be derived in Sec. IV from an explicit construction of the eigenvectors of h_p.

IV. YAKUBOVSKII–FADDEEV EQUATIONS IN DIFFERENTIAL FORM

A. Introduction

Consider the restriction to the Hilbert chain space $\mathscr{H}_{a_p} = \mathscr{H}_N \otimes \mathbb{C}_{a_p}$ of the Hamiltonian h_p defined in (3.4) and the equation

$$(z - h_p)|\Psi\rangle = 0, \quad (4.1)$$

which is a Yakubovskii–Faddeev (YF) equation in differential form.

Spectral properties of h_p derive from the study of this equation and depend on the nature of the two-body interactions. A rigorous proof of these properties is difficult. In the present paper we start from the assumption of asymptotic completeness for any partition Hamiltonian H_{a_p} which is believed to hold for reasonably short-range interactions. The main consequences of this assumption relevant to the properties of the solutions of (4.1) are presented in Sec. IV B1 and in the present section we show they are enough to derive the spectral properties of any chain Hamiltonian. Left multiplying (4.1) by S_p^N and with the help of (3.6) one gets

$$(z - H_{a_p}) S_p^N |\Psi_{a_p}\rangle = 0. \quad (4.2)$$

From definition (2.10), $S_p^N|\Psi_{a_p}\rangle$ is a vector which has its n_{a_p} components equal to the sum of $|\psi_{A_{a_p}}\rangle$ components. As a result the YF equation may have two kinds of solutions.

(1) A solution $|\Psi_{a_p}\rangle$ such that $S_p^N|\Psi_{a_p}\rangle \neq 0$ will be called a "physical eigenvector." From (4.2) the sum of its components has to be an eigenvector of H_{a_p}. Then the corresponding "physical eigenvalue" z is real.

(2) A solution $|\Phi_{a_p}\rangle$ such that $S_p^N|\Phi_{a_p}\rangle = 0$ will be called a "spurious eigenvector." It may be arbitrarily closely approximated in the norm by a vector of the null space $\mathscr{O}_{a_p}$ of S_p^N on $\mathscr{H}_{a_p}$,

$$\mathscr{O}_{a_p} = \mathscr{H}_N \otimes \mathrm{O}_{a_p}, \tag{4.3}$$

where O_{a_p}, the null space of S_p^N on $\mathbb{C}_{a_p}$ is studied in Appendix B, Sec. 1.

In the particular cases $p = N$ and $p = N-1$, vectors $|\Psi_{a_N}\rangle$ and $|\Psi_{a_{N-1}}\rangle$ have only one component and Eq. (4.1) reads, respectively,

$$(z - H_0)|\psi_{a_N}\rangle = 0, \tag{4.4}$$

$$(z - H_{a_{N-1}})|\psi_{a_{N-1}}\rangle = 0. \tag{4.5}$$

Then for h_N and h_{N-1} the eigenvalues are real and there is no spurious solution. Thanks to the assumption of asymptotic completeness for H_0 and $H_{a_{N-1}}$, h_N and h_{N-1} have complete sets of eigenvectors on $\mathscr{H}$.

Under spectral properties of the partition Hamiltonian H_{a_p} we prove in Sec. C the existence of a physical solution $|\Psi_{a_p}\rangle$ of (4.1) in correspondence with any eigenvector of H_{a_p}. Obviously, eventual spurious solutions of (4.1) for the same eigenvalue may be added to $|\Psi_{a_p}\rangle$ to give another physical solution corresponding to the same eigenstate of H_{a_p}. Let $\mathscr{P}_{a_p}$ be a subspace of $\mathscr{H}_{a_p}$ spanned by some choice of physical solutions of (4.1) in one-to-one correspondence with the complete set of eigenvectors of H_{a_p}. Let $|\chi\rangle \in \mathscr{H}_{a_p}$ not contained in $\mathscr{O}_{a_p}$. The sum $\Sigma_{A_{a_p}}|\chi_{A_{a_p}}\rangle$ of its components may be approximated in the norm by a linear combination of eigenvectors of H_{a_p}. Let $|\tilde{\chi}\rangle \in \mathscr{P}_{a_p}$ be the corresponding linear combination of eigenvectors of h_p. Since $\Sigma_{A_{a_p}}|\chi_{A_{a_p}}\rangle = \Sigma_{A_{a_p}}|\tilde{\chi}_{A_{a_p}}\rangle$ one has

$$|\chi\rangle - |\tilde{\chi}\rangle = |\Phi\rangle \in \mathscr{O}_{a_p}. \tag{4.6}$$

Then, if the set of spurious solutions of (4.1) is a complete basis on $\mathscr{O}_{a_p}$, any vector $|\chi\rangle \in \mathscr{H}_{a_p}$ can be uniquely approximated in the norm by a linear combination of eigenvectors of h_p (cf. Refs. 3 and 4). To sum up, if the YF equation (4.1) admits a complete set of spurious solutions on $\mathscr{O}_{a_p}$, and if there exists a set of physical solutions in one-to-one correspondence with the eigenvectors of H_{a_p}, which span $\mathscr{P}_{a_p}$, h_p admits a complete set of solutions on $\mathscr{H}_{a_p}$ and

$$\mathscr{H}_{a_p} = \mathscr{P}_{a_p} + \mathscr{O}_{a_p}. \tag{4.7}$$

In Sec. IV D we prove that the set of spurious solutions of Eq. (4.1) is a complete basis on $\mathscr{O}_{a_p}$ if chain Hamiltonians h_q, $p + 2 \leqslant q \leqslant N$ have a complete set of solutions on $\mathscr{H}_{a_q}$ spaces. Since it is true for $q = N$ and $q = N-1$ this ends recursively the proof that the YF equation (4.1) admits a complete set of solutions on $\mathscr{H}_{a_p}$, therefore on $\mathscr{H}$, and the spectrum of any chain Hamiltonian is real.

B. Physical solutions of YF equation

We now prove by an explicit construction the existence of solution of YF equation (4.1) in correspondence to every eigenstate of H_{a_p}.

1. Eigenstates of H_{a_p} and asymptotic completeness assumption

With a suitable system of Jacobi coordinates, partition Hamiltonian H_{a_p} can be written as a sum of $p+1$ commuting operators

$$H_{a_p} = \sum_{i=1}^{p} \overline{H}_i + T_{a_p}, \tag{4.8}$$

where $\overline{H}_i$ is the internal Hamiltonian of the ith fragment of partition a_p and T_{a_p} is associated with the center-of-mass kinetic energies of the p clusters.

Consider solutions of the Schrödinger equation

$$(E - H_{a_p})|\psi^\alpha_{a_p}(E)\rangle = 0, \tag{4.9}$$

written as a product of the internal eigenstates of the p clusters and of their free relative-motion state. As usual in physics, "eigenstate" of H_{a_p} means any solution of Eq. (4.9) which is of finite norm or normalizable in the delta function. In $|\psi^\alpha_{a_p}(E)\rangle$ the index α stands for the remaining set of quantum numbers (possibly continuous) which, besides energy, specify the solution.

Any solution of Eq. (4.9) satisfies for any $\epsilon \neq 0$

$$i\epsilon G_{a_p}(z)|\psi^\alpha_{a_p}(E)\rangle = |\psi^\alpha_{a_p}(E)\rangle, \tag{4.10}$$

where $z = E + i\epsilon$.

The eigenstates of H_{a_p} are of two kinds:

(1) States $|\varphi^\alpha_{a_p}(E)\rangle$ corresponding to bound states of the p clusters are called "bound eigenstates" of H_{a_p}, though they correspond, if $p > 1$, to a continuous part of the spectrum of H_{a_p} and are not square integrable. We assume they satisfy the normalization relation

$$\langle\varphi^\beta_{a_p}(E')|\varphi^\alpha_{a_p}(E)\rangle = \delta_{\alpha\beta}\delta(E - E'), \tag{4.11}$$

where the Dirac distribution $\delta(E - E')$ is replaced by the Kronecker symbol $\delta_{EE'}$ if $p = 1$.

(2) States corresponding to at least one cluster in a scattering state are referred as "scattering eigenstates" of H_{a_p}. We assume they evolve asymptotically from a bound state $|\varphi^\alpha_{a_R}(E)\rangle$ of another partition Hamiltonian H_{a_r}, where partition a_r is contained in a_p. There are two such linearly independent solutions $|\psi^{a_r,\alpha}_{a_p}(E^+)\rangle$ and $|\psi^{a_r,\alpha}_{a_p}(E^-)\rangle$, defined by

$$|\psi^{a_r,\alpha}_{a_p}(E^\pm)\rangle = \lim_{\epsilon \to 0\pm} |\psi^{a_r,\alpha}_{a_p}(z)\rangle, \tag{4.12a}$$

$$|\psi^{a_r,\alpha}_{a_p}(z)\rangle = i\epsilon G_{a_p}(z)|\varphi^\alpha_{a_r}(E)\rangle. \tag{4.12b}$$

This definition is meaningful only when the forces are sufficiently short-range. We assume the eigenstates of H_{a_p} form a complete set in $\mathscr{H}_N$, which corresponds to the closure relation

$$1 = \sum_\alpha \int_{E(a_p,\alpha)}^{\infty} |\varphi^\alpha_{a_p}(E')\rangle\langle\varphi^\alpha_{a_p}(E')|\, dE' + \sum_{a_r}\delta^{a_r}_{a_p}\sum_\alpha \int_{E(a_r,\alpha)}^{\infty} |\psi^{a_r,\alpha}_{a_p}(E'^\pm)\rangle\langle\psi^{a_r,\alpha}_{a_p}(E'^\pm)|\, dE', \tag{4.13}$$

written in a condensed notation,

$$1=\sum_{\alpha}\int_{E(\alpha)}^{\infty}\left|\psi_{a_p}^{\alpha}(E')\right\rangle\left\langle\psi_{a_p}^{\alpha}(E')\right|dE'. \qquad (4.14)$$

2. Bound physical solutions of the YF equation

Let $|\varphi_{a_p}^{\alpha}(E)\rangle$ be a bound eigenstate of H_{a_p} at energy E and $|\chi_{a_p}^{\alpha}(E)\rangle$ be an n_{a_p} component vector with a single nonzero component $|\chi_{B_{a_p}}^{\alpha}\rangle=|\varphi_{a_p}^{\alpha}(E)\rangle$.

$$|\chi_{A_{a_p}}^{\alpha}(E)\rangle=\delta_{A_{a_p}B_{a_p}}|\varphi_{a_p}^{\alpha}(E)\rangle, \qquad (4.15)$$

where B_{a_p} is an arbitrary chain contained in a_p.

Using the restriction to $\mathscr{H}_{a_p}$ of $g_p(z)=(z-h_p)^{-1}$, we define an n_{a_p}-component vector

$$|\varphi_{a_p}^{\alpha}(z)\rangle=i\epsilon g_p(z)|\chi_{a_p}^{\alpha}\rangle. \qquad (4.16)$$

From Eqs. (3.26) and (3.33) the components of $|\varphi_{a_p}^{\alpha}(z)\rangle$ are

$$\begin{aligned}|\varphi_{A_{a_p}}^{\alpha}(z)\rangle &= i\epsilon(g_p(z))_{A_{a_p}B_{a_p}}|\varphi_{a_p}^{\alpha}(E)\rangle\\ &= i\epsilon\Gamma_{A_p}(z)|\varphi_{a_p}^{\alpha}(E)\rangle\\ &\quad+\sum_{q=p+1}^{N}(Z_p^{(q)})_{A_{a_p}B_{a_p}}i\epsilon\Gamma_{A_q}(z)|\varphi_{a_p}^{\alpha}(E)\rangle. \qquad (4.17)\end{aligned}$$

Let us consider

$$\begin{aligned}\Gamma_{A_q}(z)|\varphi_{a_p}^{\alpha}(E)\rangle &= G_0(z)V_{a_{N-1}}G_{a_{N-1}}(z)\cdots G_{a_s}(z)\\ &\quad\cdots G_{a_{q+1}}(z)V_{a_q}^{a_{q+1}}G_{a_q}(z)|\varphi_{a_p}^{\alpha}(E)\rangle,\end{aligned}$$

where any partition a_s of chain A_q is contained in a_p. The partition Hamiltonian H_{a_s} can be decomposed into two commuting parts, $H_{a_s}=\overline{H}_{as}+T_{a_p}$, where T_{a_p} is defined in (4.8). Any factor in $\Gamma_{A_q}(z)$ commutes with T_{a_p}. Since $|\varphi_{a_p}^{\alpha}(E)\rangle$ is an eigenstate of T_{a_p} corresponding to the eigenvalue E_{a_p}, operator T_{a_p} can be replaced by E_{a_p} in any resolvent $G_{a_s}(z)$. Since $|\varphi_{a_p}^{\alpha}(E)\rangle$ is a bound eigenstate of H_{a_p}, the energy $E-E_{a_p}$ is less than the least eigenvalue of $\overline{H}_{a_s}$. Then $G_{a_s}(z)$ and consequently $\Gamma_{A_q}(z)$ is not singular when ϵ goes to zero. This proves the existence of

$$\lim_{\epsilon\to 0}\Gamma_{A_q}(z)|\varphi_{a_p}^{\alpha}(E)\rangle\equiv\Gamma_{A_q}(E)|\varphi_{a_p}^{\alpha}(E)\rangle, \qquad (4.18)$$

if $\delta_{a_p}^{a_q}=1$, $\delta_{a_pa_q}=0$.

Then

$$\lim_{\epsilon\to 0}i\epsilon\Gamma_{A_q}(z)|\varphi_{a_p}^{\alpha}(E)\rangle=0 \qquad (4.19)$$

if $\delta_{a_p}^{a_q}=1$, $\delta_{a_pa_q}=0$.

Using now (4.10), one obtains from (4.17)

$$\lim_{\epsilon\to 0}|\varphi_{A_{a_p}}^{\alpha}(z)\rangle=\Gamma_{A_{p+1}}(E)V_{a_p}^{a_{p+1}}|\varphi_{a_p}^{\alpha}(E)\rangle. \qquad (4.20)$$

This proves the existence of

$$|\varphi_{a_p}^{\alpha}(E)\rangle=\lim_{\epsilon\to 0}|\varphi_{a_p}^{\alpha}(z)\rangle.$$

From (4.9) it is a solution of the YF equation in differential form

$$(E-h_p)|\varphi_{a_p}^{\alpha}(E)\rangle=0. \qquad (4.21)$$

From (4.17), (3.10), and (4.10)

$$\sum_{A_{a_p}}|\varphi_{A_{a_p}}^{\alpha}(z)\rangle=|\varphi_{a_p}^{\alpha}(E)\rangle=\sum_{A_{a_p}}|\varphi_{A_{a_p}}^{\alpha}(E)\rangle. \qquad (4.22)$$

Therefore $|\varphi_{a_p}^{\alpha}(E)\rangle$ is a physical solution of (4.1) corresponding to the bound eigenstate $|\varphi_{a_p}^{\alpha}(E)\rangle$ of H_{a_p}. Let us remark from (4.20) that $|\varphi_{a_p}^{\alpha}(E)\rangle$ does not depend on the arbitrary chain B_{a_p} in (4.15).

3. Scattering physical solutions

Let $|\psi_{a_p}^{b_r\beta}(E)\rangle$ be a scattering eigenstate of H_{a_p}

$$|\psi_{a_p}^{b_r\beta}(E)\rangle=\lim_{\epsilon\to 0^+}i\epsilon G_{a_p}(z)|\varphi_{b_r}^{\beta}(E)\rangle. \qquad (4.23)$$

Let $|\chi_{a_p}^{b_r\beta}(E)\rangle$ be an n_{a_p}-component vector with a single nonzero component

$$\begin{aligned}|\chi_{B_{a_p}}^{b_r\beta}(E)\rangle &= |\varphi_{b_r}^{\beta}(E)\rangle,\\ |\chi_{A_{a_p}}^{b_r\beta}(E)\rangle &= \delta_{A_{a_p}B_{a_p}}|\varphi_{b_r}^{\beta}(E)\rangle, \qquad (4.24)\end{aligned}$$

where B_{a_p} is an arbitrary chain contained in a_p, which contains b_r.

We define an n_{a_p}-component vector

$$|\Psi_{a_p}^{b_r\beta}(z)\rangle=i\epsilon g_p(z)|\chi_{a_p}^{b_r\beta}\rangle, \qquad (4.25)$$

with components

$$|\psi_{A_{a_p}}^{b_r\beta}(z)\rangle=\sum_{q=p}^{N}(Z_p^{(q)})_{A_{a_p}B_{a_p}}i\epsilon\Gamma_{A_q}(z)|\varphi_{b_r}^{\beta}(E)\rangle. \qquad (4.26)$$

From (4.25)

$$\sum_{A_{a_p}}|\psi_{A_{a_p}}^{b_r\beta}(z)\rangle=i\epsilon G_{a_p}(z)|\varphi_{b_r}^{\beta}(E)\rangle=|\psi_{a_p}^{b_r\beta}(z)\rangle. \qquad (4.27)$$

Then

$$\lim_{\epsilon\to 0^+}\sum_{A_{a_p}}|\psi_{A_{a_p}}^{b_r\beta}(z)\rangle=|\psi_{a_p}^{b_r\beta}(E)\rangle. \qquad (4.28)$$

If one admits that the vector $|\Psi_{a_p}^{b_r\beta}(z)\rangle$ has a limit when ϵ goes to 0^+, then this limit $|\Psi_{a_p}^{b_r\beta}(E^+)\rangle$ is a physical solution of the YF equation (4.1) corresponding to the scattering eigenstate $|\psi_{a_p}^{b_r\beta}(E)\rangle$ of H_{a_p}. It remains to prove that this limit exists.

We remark that instead of the vector $|\chi_{a_p}^{b_r\beta}\rangle$ defined by (4.24) one might have chosen any chain decomposition of $|\varphi_{b_r}^{\beta}(E)\rangle$. Indeed spurious solutions at energy E may be added to $|\Psi_{a_p}^{b_r\beta}(E^+)\rangle$ to give another physical solution.

C. Spurious solutions of YF equation

With the help of (3.8), Eq. (4.1) can be written

$$(z-h_q)|\Psi_{a_p}\rangle=(v_{q-1}-v_q)Y_p^q|\Psi_{a_p}\rangle,\quad p+2\leqslant q\leqslant N-1. \qquad (4.29)$$

Consider the subspace of $\mathscr{H}_{a_p}$ defined by the tensor product $\mathscr{H}_{a_q}\otimes\mathbb{Q}_{a_p}^{a_q}$, where $\mathbb{Q}_{a_p}^{a_q}$ defined by (B7) is a subspace of the null space of Y_p^q on $\mathbb{C}_{a_p}^{a_q}$.

We have shown in Sec. IV A that h_N and h_{N-1} admit a complete set of eigenvectors on $\mathscr{H}_N$ and $\mathscr{H}_{N-1}$, respectively corresponding to real eigenvalues. Let us assume any h_q with $p+2\leqslant q\leqslant N$ has the same properties on $\mathscr{H}_{a_q}$. Then one can build a complete basis in the subspace $\mathscr{H}_{a_q}\otimes\mathbb{Q}_{a_p}^{a_q}$ of vectors $|\Phi_{a_p}(a_q)\rangle$ with components $|\Phi_{B_{a_p}}(a_q)\rangle$ corresponding to chains B_{a_p} containing a_q which satisfy

$$(E-h_q)|\Phi_{a_p}(a_q)\rangle=0, \qquad (4.30)$$

$$Y_p^q|\Phi_{a_p}(a_q)\rangle = 0. \tag{4.31}$$

Using (4.29) one verifies that vectors $|\Phi_{a_p}(a_q)\rangle$ are solutions of (4.1). They are spurious solutions since from (A35) and (4.31)

$$S_p^N|\Phi_{a_p}(a_q)\rangle = \widetilde{Y}_p^N\widetilde{Y}_p^{N-1}...\widetilde{Y}_p^{q+1}Y_p^{p+1}Y_p^{p+2}...Y_p^q|\Phi_{a_p}(a_q)\rangle = 0.$$

Then h_p admits on $\mathscr{H}_{a_q}\otimes\mathbb{Q}_{a_p}^{a_q}, p+2\leqslant q\leqslant N$, a complete set of spurious solutions corresponding to real engenvalues.

From Eq. (B21) with $a_r = a_n$ one obtains the decomposition of the null space $\mathscr{O}_{a_p} = \mathscr{H}_N\otimes\mathbb{O}_{a_p}$ of S_p^N on $\mathscr{H}_{a_p}$.

$$\mathscr{O}_{a_p} = \sum_{q=p+2}^{N}\sum_{a_q}\mathscr{H}_{a_q}\otimes\mathbb{Q}_{a_p}^{a_q}. \tag{4.32}$$

Then h_p admits a complete set of spurious solutions corresponding to real eigenvalues on $\mathscr{O}_{a_p}$. Coming back to the discussion at the end of Sec. IV A we have then proved that YF equation (4.1) admits a complete set of eigenvectors on $\mathscr{H}_{a_p}$ corresponding to real eigenvalues.

We shall now exhibit the structure of the spurious solutions of (4.1) in a somewhat different matter using Eqs. (B20) and (4.3) which gives

$$\mathscr{O}_{a_p} = \mathscr{H}_N\otimes\mathbb{K}_{a_p} + \sum_{q=p+2}^{N-1}\sum_{a_q}\mathscr{H}_N\otimes\mathbb{R}_{a_q}\otimes\mathbb{K}_{a_p}^{a_q}, \tag{4.33}$$

where $\mathbb{K}_{a_p}^{a_q}$ is the null space of Y_p^q on $\mathbb{C}_{a_p}^{a_q}$ and $\mathbb{R}_{a_q}$ is the eigenspace of Y_q^N defined in Eq. (B6). In (4.33) $\mathscr{H}_N\otimes\mathbb{R}_{a_q}$ can be replaced by any subspace $\mathscr{P}_{a_q}$ of $\mathscr{H}_{a_q}$ spanned by a set of physical solutions of h_q in one to one correspondence with the eigenvectors of H_{a_q}. Since $\mathscr{H}_N\equiv\mathscr{P}_{a_N}$, Eq. (4.33) reads

$$\mathscr{O}_{a_p} = \sum_{q=p+2}^{N}\sum_{a_q}\mathscr{P}_{a_q}\otimes\mathbb{K}_{a_p}^{a_q}. \tag{4.34}$$

One can build a complete basis of vectors in $\mathscr{P}_{a_q}\otimes\mathbb{K}_{a_p}^{a_q}$ from the physical solutions of h_q and $\mathscr{H}_{a_q}$ and the null eigenvectors of Y_p^q on $\mathbb{C}_{a_p}^{a_q}$. These vectors satisfy Eqs. (4.30)–(4.31). Therefore they are spurious solutions of (4.1) corresponding to the eigenvalues which belong to the spectrum of H_{a_q}. Formula (4.34) displays the fact that the complete set of spurious solutions of h_p on $\mathscr{H}_{a_p}$ can be constructed from the physical solutions of h_q and the null space of Y_p^q with $p+2\leqslant q\leqslant N$.

In particular case of the three-body problem and $p=1$, Eq. (4.34) reduces to

$$\mathscr{O}_{a_1} = \mathscr{P}_{a_3}\otimes\mathbb{K}_{a_1}^{a_3}. \tag{4.35}$$

$\mathbb{K}_{a_1}^{a_3}$ is a two dimensional space. To each solution of $(E-H_0)|\psi_N^\alpha(E)\rangle = 0$ correspond two independent spurious solutions of h_1

$$|\Psi_1^{\alpha,2}(E)\rangle = \begin{pmatrix} |\psi_N^\alpha(E)\rangle \\ -|\psi_N^\alpha(E)\rangle \\ 0 \end{pmatrix}, |\Psi_1^{\alpha,3}(E)\rangle = \begin{pmatrix} |\psi_N^\alpha(E)\rangle \\ 0 \\ -|\psi_N^\alpha(E)\rangle \end{pmatrix}. \tag{4.36}$$

Observe that spurious solutions exist only for eigenvalues E above the three particles breakup threshold. There is no spurious solutions when one is looking for bound states of the three-body system or scattering states at energies below the three particle breakup threshold. These properties of YF equation in differential form are given in Ref. 3 for $N=3$. In the general N-body case the equation $(E-h_1)|\Psi_1\rangle = 0$ admits no spurious solution if E is less than the three cluster breakup threshold.

D. Dual solutions

Let us construct a complete basis on $\mathscr{H}_{a_p}$ of physical and spurious eigenvectors of the restriction of h_p to $\mathscr{H}_{a_p}$ along the decomposition of $\mathscr{H}_{a_p}$

$$\mathscr{H}_{a_p} = \mathscr{P}_{a_p} + \sum_{q=p+2}^{N}\sum_{a_q}\mathscr{P}_{a_q}\otimes\mathbb{K}_{a_p}^{a_q}, \tag{4.37}$$

given by Eqs. (4.7) and (4.34).

Let $|\Psi_{a_p}^{\alpha,\tau}(E)\rangle$ be any member of this complete set. Index τ runs from 1 to n_{a_p}, the value $\tau=1$ corresponding to physical solutions. Each $\tau>1$ value specifies both a partition a_q and a basis vector in $\mathbb{K}_{a_p}^{a_q}$. To each eigenvector $|\Psi_{a_p}^{\alpha,\tau}(E)\rangle$ corresponds an eigenvector $|\psi_{a_q}^\alpha(E)\rangle$ of H_{a_q},

$$(E-H_{a_q})|\psi_{a_q}^\alpha(E)\rangle = 0,$$
$$\langle\psi_{a_q}^{\alpha'}(E')|\psi_{a_q}^\alpha(E)\rangle = \delta_{\alpha\alpha'}\delta(E-E'). \tag{4.38}$$

Index α stands for the remaining set of quantum numbers possibly continuous which besides energy E specify the solution. If $p>1$, for each τ value, energy E belongs to the continuum $E\geqslant E_\tau$, where E_τ is the least eigenvalue of the corresponding partition Hamiltonian H_{a_q}.

The complete set of eigenvectors $|\Psi_{a_p}^{\alpha,\tau}(E)\rangle$ is not orthogonal since h_p is not self-adjoint and we proceed in the following to the construction of the associated biorthogonal basis. Assume that the N-body Hilbert space $\mathscr{H}_N$ is separable. Let $|u_i\rangle\in\mathscr{H}_N$ be a member of a complete orthonormalized basis

$$\langle u_i|u_j\rangle = \delta_{ij}. \tag{4.39}$$

Let ρ be an index running from 1 to n_{a_p}, which specifies a particular chain A_{a_p} denoted $A_{a_p}^{(\rho)}$. Let $|\chi_{i,\rho}\rangle\in\mathscr{H}_{a_p}$ with components

$$|\chi_{i,\rho A_{a_p}}\rangle = \delta_{A_{a_p}A_{a_p}^{(\rho)}}|u_i\rangle. \tag{4.40}$$

It results

$$\langle\chi_{i,\rho}|\chi_{j,\rho'}\rangle = \delta_{ij}\delta_{\rho\rho'}, \tag{4.41}$$

and the closure relation on $\mathscr{H}_{a_p}$

$$\mathbb{1} = \sum_i\sum_\rho|\chi_{i,\rho}\rangle\langle\chi_{i,\rho}|. \tag{4.42}$$

Since eigenvectors $|\Psi_{a_p}^{\alpha,\tau}(E)\rangle$ form a complete basis on $\mathscr{H}_{a_p}$ any vector $|\chi_{i,\rho}\rangle$ can be expressed in a unique way as

$$|\chi_{i,\rho}\rangle = \sum_\tau\sum_\alpha\int_{E_\tau}^{\infty}|\Psi_{a_p}^{\alpha,\tau}(E)\rangle a_{i,\rho}^{\alpha,\tau}(E)dE, \tag{4.43}$$

which defines functions $a_{i,\rho}^{\alpha,\tau}$ of E. Let $\langle\widetilde{\Psi}_{a_p}^{\alpha,\tau}(E)|$ be the distribution on $\mathscr{H}_{a_p}$ defined by

$$\langle\widetilde{\Psi}_{a_p}^{\alpha,\tau}(E)|\chi_{i,\rho}\rangle = a_{i,\rho}^{\alpha,\tau}(E), \tag{4.44}$$

which is equivalent to

$$\langle\widetilde{\Psi}_{a_p}^{\alpha,\tau}(E)\Big| = \sum_i\sum_\rho\langle\chi_{i,\rho}\Big|a_{i,\rho}^{\alpha,\tau}(E). \tag{4.45}$$

Substituting (4.45) into (4.43) gives

$$|\chi_{i,\rho}\rangle = \sum_\tau\sum_\alpha\int_{E_\tau}^{\infty} |\Psi_{a_p}^{\alpha,\tau}(E)\rangle\langle\widetilde{\Psi}_{a_p}^{\alpha,\tau}(E)|\chi_{i,\rho}\rangle dE. \quad (4.46)$$

The relation being true for any $|\chi_{i,\rho}\rangle$ implies the closure relation

$$\mathbb{1} = \sum_\tau\sum_\alpha\int_{E_\tau}^{\infty} |\Psi_{a_p}^{\alpha,\tau}(E)\rangle\langle\widetilde{\Psi}_{a_p}^{\alpha,\tau}(E)|\, dE. \quad (4.47)$$

The sequence of pairs $\{|\widetilde{\Psi}_{a_p}^{\alpha,\tau}(E)\rangle, \langle\widetilde{\Psi}_{a_p}^{\alpha,\tau}(E)|\}$ defines a complete biorthogonal basis on $\mathscr{H}_{a_p}$. From Eq. (4.47) one obtains

$$h_p = \sum_\tau\sum_\alpha\int_{E_\tau}^{\infty} E\, |\Psi_{a_p}^{\alpha,\tau}(E)\rangle\langle\widetilde{\Psi}_{a_p}^{\alpha,\tau}(E)|\, dE$$

$$= \sum_\tau\sum_\alpha\int_{E_\tau}^{\infty} |\Psi_{a_p}^{\alpha,\tau}(E)\rangle\langle\widetilde{\Psi}_{a_p}^{\alpha,\tau}(E)|\, h_p\, dE.$$

After substraction of these two relations, one obtains from the linear independence of the vectors $|\Psi_{a_p}^{\alpha,\tau}(E)\rangle$

$$\langle\widetilde{\Psi}_{a_p}^{\alpha,\tau}(E)|h_p = E\,\langle\widetilde{\Psi}_{a_p}^{\alpha,\tau}(E)|, \quad (4.48)$$

which implies that $|\widetilde{\Psi}_{a_p}^{\alpha,\tau}(E)\rangle$ is an eigenvector of the adjoint chain Hamiltonian $\tilde{h}_p$.

The explicit construction of the $|\widetilde{\Psi}_{a_p}^{\alpha,1}(E)\rangle$ eigenvector of $\tilde{h}_p$ corresponding to the physical eigenvector $|\Psi_{a_p}^{\alpha,1}(E)\rangle$ of h_p is very easy. Let us add up the n_{a_p} components of the vectorial relation (4.43). On the left-hand side from (4.40) one obtains $|u_i\rangle$. On the right-hand side, spurious solutions $\tau > 1$ are eliminated. One obtains

$$|u_i\rangle = \sum_\alpha\int_{E_{a_p}}^{\infty} |\psi_{a_p}^{\alpha}(E)\rangle a_{i,\rho}^{\alpha,1}(E)\, dE, \quad (4.49)$$

and from (4.38)

$$a_{i,\rho}^{\alpha,1}(E) = \langle\psi_{a_p}^{\alpha}(E)|u_i\rangle \quad (4.50)$$

does not depend on ρ. From (4.45) and (4.40) one verifies that the n_{a_p} components of $|\widetilde{\Psi}_{a_p}^{\alpha,1}(E)\rangle$

$$|\widetilde{\Psi}_{A_{a_p}}^{\alpha,1}(E)\rangle = \sum_i\langle u_i|\psi_{a_p}^{\alpha}(E)\rangle\sum_\rho\delta_{A_{a_p}A_{a_p}^{(\rho)}}|u_i\rangle = |\psi_{a_p}^{\alpha}(E)\rangle \quad (4.51)$$

are equal.

V. CONCLUDING REMARKS

The Yakubovskii–Faddeev method solves the many body problem in a natural way by decomposing the Hamiltonian according to all chains of partitions corresponding to all possible paths which can be followed to fuse the particles together. Our presentation outlines a structural analogy between the YF formalism and the channel-coupling array theory[4,5] which follows from the introduction in both theories of decomposed interactions, chain decomposition in the first case [Eq. (1.3)] and partition decomposition in the second case. This analogy involves similar properties such as, for instance, the existence of spurious solutions for the equations in differential form (Sec. IV C) and the fact that the physical duals of both theories have similar equal component nature [Eq. (4.51)]. However there is a main difference. In channel-coupling array theories spurious solutions can occur for complex energies. We have shown, on the contrary, that the spectrum of any YF chain Hamiltonian h_p is real and that the set of its eigenvectors (physical and spurious) is complete.

The resolvent operator $g_p(z) = (z - g_p)^{-1}$ exists for any complex number $z = E + i\epsilon$. The resolvent identity

$$g_p(z) = g_q(z) + g_q(z)(v_p - v_q)g_p(z). \quad (5.1)$$

can be considered as an integral equation to calculate $g_p(z)$ and correspondingly any many-body wave function $|\Psi_{a_p}^{b,\beta}(z)\rangle$ in the chain space $\mathscr{H}_{a_p}$. The matrix notation permits us to stress the formal analogy between Eq. (5.1) and a Lippmann–Schwinger equation in the space of partitions

$$G_{a_p}(z) = G_{a_q}(z) + G_{a_q}(z)(V_{a_p} - V_{a_q})G_{a_p}(z). \quad (5.2)$$

Both equations have a unique solution for z complex, but there is an essential difference.

Equation (5.2) is not well behaved because its kernel cannot be connected after a finite number of iterations. On the contrary, Eq. (5.1) considered in the particular case $q = p + 1$,

$$g_p(z) = g_{p+1}(z) + g_{p+1}(z)(v_p - v_{p+1})g_p(z), \quad (5.3)$$

has its $N - p - 1$ iterated kernel $[g_{p+1}(z)(v_p - v_{p+1})]^{N-P}$ with the connectivity of a p-cluster partition. Indeed, the matrix element of the coupling potential

$$(v_p - v_{p+1})_{A_pB_p} = V_{a_N}\,,(Y_p^N - Y_{p+1}^N)_{A_pB_p} \quad (5.4)$$

connects two clusters of partition b_{p+2} to form a new partition a_{p+1}, where $a_{p+1} \supset b_{p+2}$ but $a_{p+1} \neq b_{p+1}$. This string of restrictions on the coupling of chains ensures an increasing connectivity with each iteration.

Thus the N-body problem can be solved in a systematic manner by induction through the hierarchy of connected-kernel YF equations (5.3) starting from $p = N - 1$ up to $p = 1$.

Consider, for instance, the YF integral equation

$$|\Psi^{b_2,\beta}(z)\rangle = |\varphi_{b_2}^{\beta}(E)\rangle + g_2(z)(v_1 - v_2)|\Psi^{b_2,\beta}(z)\rangle, \quad (5.5)$$

where $|\varphi_{b_2}^{\beta}(E)\rangle$ is the physical eigenstate of h_2 corresponding to a two-cluster initial bound state $|\varphi_{b_2}^{\beta}(E)\rangle$. It can be formally solved by

$$|\Psi^{b_2,\beta}(z)\rangle = i\epsilon g_1(z)|\varphi_{b_2}^{\beta}(E)\rangle,$$

which proves the uniqueness of the solution.

Adding up all chain components one gets

$$\sum_A|\psi^{b_2,\beta}{}_A(z)\rangle = i\epsilon G(z)\sum_A|\varphi_{b_2,A}^{\beta}(E)\rangle,$$

$$= i\epsilon G(z)|\varphi_{b_2}^{\beta}(E)\rangle = |\psi^{b_2,\beta}(z)\rangle, \quad (5.6)$$

that is the N-body wave function.

In Sec. IV C, we found that the YF equation in differential form

$$(E - h_1)|\Psi(E)\rangle = 0 \quad (5.7)$$

may have "spurious" solutions (i.e, $\Sigma_A|\psi_A(E)\rangle = 0$). These solutions can be obtained from the YF integral equation (5.5) when one substitutes for the inhomogeneous term a spurious state.

In Sec. III C we have analyzed the connectivity structure of the Green's operator

$$g_p(z) = \sum_{q=p}^{N} g_p^{(q)}(z), \quad (5.7a)$$

$$(g_p^{(q)}(z))_{AB} = (Z_p^{(q)})_{AB}\Gamma_{A_q}(z), \quad (5.7b)$$

$$\Gamma_{A_q} = G_0 V_{a_{N-1}} G_{a_{N-1}} V^{a_{N-1}}_{a_{N-2}} \cdots G_{a_{q+1}} V^{a_{q+1}}_{a_q} G_{a_q}. \quad (5.7c)$$

Dynamics enters into the matrix elements of $g_p^{(q)}$ only through the factor $\Gamma_{A_q}(z)$, the other factor $(Z_p^{(q)})_{AB}$ being purely algebraic. Term $\Gamma_{A_q}(z)$ is the most connected part of $(g_q(z))_{AB}$. Then if one assumes g_q is known for $q>2$, the unknown part of g_1 is only its most connected piece $(g_1^{(1)})_{AB} = \Gamma_A$. From (3.11) and (3.16), it satisfies an integral equation with the same kernel as the YF equation, and a connected homogeneous term

$$g_1^{(1)} = g_2(v_1 - v_2)g_2^{(2)} + g_2(v_1 - v_2)g_1^{(1)}. \quad (5.8)$$

Relations (5.7) are of interest to analyze the asymptotic structure of the YF chain components of the wave function.

Actually one cannot hope for an exact solution of the hierarchic system (5.3) for $N>4$. An approximation scheme may consist in a truncation of the hierarchy at some level p with a given model for $g_{p+1}(z)$. Such a model may be derived from the analysis of (5.7) or from the spectral expansion of $g_{p+1}(z)$ which follows from the closure relation (4.47) in terms of the biorthogonal basis.

ACKNOWLEDGMENT

The authors are grateful to Professor F. S. Levin for bringing to our attention the work of J. W. Evans and D. K. Hoffman.

APPENDIX A: CHAIN PROPERTIES

1. Counting numbers

Given any pair or partitions a_p and a_r restricted to

$$\delta^{a_r}_{a_p} = 1, \quad r>p+1, \quad (A1)$$

let $\nu^{a_r}_{a_p}$ be defined by

$$\nu^{a_r}_{a_p} = \sum_{a_{r-1}} \delta^{a_r}_{a_p} \delta^{a_{r-1}}_{a_p}, \quad (A2)$$

where the symbol $\delta^{a_r}_{a_p}$ is defined in Eq. (2.1). The number $\nu^{a_r}_{a_p}$, $r>p+2$, counts the partitions a_{r-1} which are contained in a_p and which contain a_r. Note that $\nu^{a_{p+1}}_{a_p}$ reduces to $\delta^{a_{p+1}}_{a_p}$.

Let $n^{a_r}_{a_p}$ denote the number of chains $A^{a_r}_{a_p}$ which are contained in a_p and which contain a_r. Then

$$n^{a_{p+1}}_{a_p} = \delta^{a_{p+1}}_{a_p}, \quad (A3)$$

and $n^{a_r}_{a_p}$ can be computed from either of the recurrence relations

$$n^{a_r}_{a_p} = \sum_{a_{r-1}} \delta^{a_r}_{a_{r-1}} n^{a_{r-1}}_{a_p}, \quad (A4)$$

$$n^{a_r}_{a_p} = \sum_{a_{p+1}} n^{a_r}_{a_{p+1}} \delta^{a_{p+1}}_{a_p}, \quad (A5)$$

which gives in particular

$$n^{a_{p+2}}_{a_p} = \nu^{a_{p+2}}_{a_p}. \quad (A6)$$

From (A2) and (A5) one deduces

$$\sum_{a_q} n^{a_r}_{a_q} \nu^{a_q}_{a_p} = \sum_{a_{q-1}} \sum_{a_q} n^{a_r}_{a_q} \delta^{a_q}_{a_{q-1}} \delta^{a_{q-1}}_{a_p} = \sum_{a_{q-1}} n^{a_r}_{a_{q-1}} \delta^{a_{q-1}}_{a_p}. \quad (A7)$$

It will prove useful to introduce counting numbers $m^{a_r}_{a_p}$ defined by

$$m^{a_{p+1}}_{a_p} = 0, \quad (A8)$$

$$n^{a_r}_{a_p} = \delta^{a_r}_{a_p} + \sum_{q=p+2}^{r} \sum_{a_q} \delta^{a_r}_{a_q} m^{a_q}_{a_p}. \quad (A9)$$

From (A6) one derives, in particular,

$$m^{a_{p+2}}_{a_p} = \nu^{a_{p+2}}_{a_p} - 1. \quad (A10)$$

Using (A4) and (A9) one obtains

$$n^{a_r}_{a_p} = \sum_{a_{r-1}} \delta^{a_r}_{a_{r-1}} \delta^{a_{r-1}}_{a_p} + \sum_{a_{r-1}} \sum_{q=p+2}^{r-1} \sum_{a_q} \delta^{a_r}_{a_{r-1}} \delta^{a_{r-1}}_{a_q} m^{a_q}_{a_p},$$

and through (A2),

$$n^{a_r}_{a_p} = \nu^{a_r}_{a_p} + \sum_{q=p+2}^{r-1} \sum_{a_q} \nu^{a_r}_{a_q} m^{a_q}_{a_p}. \quad (A11)$$

From (A9) and (A11) one gets

$$m^{a_r}_{a_p} = \nu^{a_r}_{a_p} - \delta^{a_r}_{a_p} + \sum_{q=p+2}^{r-2} \sum_{a_q} (\nu^{a_r}_{a_q} - \delta^{a_r}_{a_q}) m^{a_q}_{a_p}. \quad (A12)$$

Note also the relation

$$m^{a_r}_{a_p} = \nu^{a_r}_{a_p} - \delta^{a_r}_{a_p} + \sum_{q=p+2}^{r-2} \sum_{a_q} m^{a_r}_{a_q} (\nu^{a_q}_{a_p} - \delta^{a_q}_{a_p}), \quad (A13)$$

which can be proved by identifying the iterative solutions of (A12) and (A13).

In Appendix B we show that $m^{a_r}_{a_p}$ is the dimension of the null space in $\mathbb{C}^{a_r}_{a_p}$ of the correlation operator Y^r_p defined in Eq. (2.13)

2. Fundamental property of chains

The main properties of the chain correlation operators defined in Sec. II B derive from the following property: given three partitions b_r, a_{r-1}, and b_q with $q<r-1$ such that $\delta^{b_r}_{a_{r-1}} = \delta^{b_r}_{b_q} = 1 - \delta^{a_{r-1}}_{b_q} = 1$, there exists one and only one partition a_{q-1} such that $\delta^{a_{r-1}}_{a_{q-1}} = \delta^{b_q}_{a_{q-1}} = 1$.

Indeed, let us denote by "1", "2",...,"r," the r clusters in b_r labeled in such a way that "1" and "2" are fused to obtain a_{r-1}. Since $\delta^{a_{r-1}}_{b_q} = 0$, "1" and "2" belong to different clusters in b_q. Then by fusing these two clusters of b_q one obtains the unique a_{q-1} partition which contains both b_q and a_{r-1}.

As a consequence a partition b_{q-1}, which contains b_q and is different from a_{q-1}, does not contain a_{r-1}.

3. Chain correlation operators

We define various operators on the chain space $\mathbb{C}$ having all their matrix elements equal to 0 or 1.

(*a*) *Summation operators* S^q_p. They are defined for $1<p<q<N$ through their matrix elements

$$(S^q_p)_{AB} = \delta_{A_q B_q} \delta_{A^p B^p}. \quad (A14)$$

Operator S^q_p is self-adjoint

$$S^q_p = \tilde{S}^q_p, \quad (A15)$$

and behaves like a projector

$$(S^q_p)^2_{AB} = n^{a_q}_{a_p} (S^q_p)_{AB}, \quad (A16)$$

where $n^{a_q}_{a_p}$ is defined in Appendix A, Sec. 1. Its main interest results from the summation property

$$(S_p^q O)_{AB} = \sum_{A_{p+1}^{q-1}} O_{AB},$$
$$(OS_p^q)_{AB} = \sum_{B_{p+1}^{q-1}} O_{AB}. \qquad \text{(A17)}$$

(*b*) *Operators* I_p^{p+2}. They are defined for $1 \leqslant p \leqslant N-2$ by

$$I_p^{p+2} = S_p^{p+2} - 1, \qquad \text{(A18)}$$

which corresponds to matrix elements

$$(I_p^{p+2})_{AB} = \delta_{A_{p+2}B_{p+2}} \bar{\delta}_{a_{p+1}b_{p+1}} \delta_{A^pB^p}. \qquad \text{(A19)}$$

Operator I_p^{p+2} is self-adjoint

$$I_p^{p+2} = \tilde{I}_p^{p+2}. \qquad \text{(A20)}$$

(*c*) *Operators* D_p^r. They are defined for $1 \leqslant p < r \leqslant N$ by resursion relations

$$D_p^{p+1} = 1,\ D_p^r = I_{r-2}^r D_p^{r-1} \quad \text{if} \quad p \leqslant r-2. \qquad \text{(A21)}$$

In particular $D_p^{p+2} = I_p^{p+2}$.
By iterating (A21) one obtains

$$D_p^r = I_{r-2}^r I_{r-3}^{r-1} \cdots I_{p+1}^{p+3} I_p^{p+2} = D_{p+1}^r I_p^{p+2}, \qquad \text{(A22)}$$

and for the adjoint operator

$$\tilde{D}_p^r = I_p^{p+2} I_{p+1}^{p+3} \cdots I_{r-3}^{r-1} I_{r-2}^r = I_p^{p+2} \tilde{D}_{p+1}^r. \qquad \text{(A23)}$$

From (A19) and (A22) matrix elements have the form

$$(D_b^r)_{AB} = \delta_{A_rB_r}\Big(\prod_{q=p+1}^{r-1} \bar{\delta}_{a_qb_q}\Big)\Big(\prod_{q=p+1}^{r-2} \delta_{a_q}^{b_{q+1}}\Big)\delta_{A^pB^p}, \qquad \text{(A24)}$$

which proves that $(D_p^r)_{AB}$ equals 0 or 1.

Property of D_p^r *operators*: Given a chain B and a partition a_{r-1} which satisfies

$$\delta_{a_{r-1}}^{b_r} = \delta_{b_p}^{a_{r-1}} = 1 - \delta_{b_{p+1}}^{a_{r-1}} = 1. \qquad \text{(A25)}$$

there is a unique chain A containing a_{r-1} such that $(D_p^r)_{AB} = 1$. This property is derived from the property of chains described in Appendix A, Sec. 2 in the following way: We first remark from Eq. (A24) that $(D_p^r)_{AB} = 1$ implies $a_s = b_s$ for $s \leqslant p$ and $s \geqslant r$ which is consistent with conditions (A25) $\delta_{a_{r-1}}^{b_r} = \delta_{b_p}^{a_{r-1}} = 1$. Condition (A25) $\delta_{b_{p+1}}^{a_{r-1}} = 0$ implies $\delta_{b_q}^{a_{r-1}} = 0$ for $p+2 \leqslant q \leqslant r-1$. Relation $(D_p^r)_{AB} = 1$ implies $\delta_{a_{q-1}}^{a_{r-1}} = \delta_{a_{q-1}}^{b_q} = 1$. Then from the property of chains, partition a_{q-1} exists and is unique. This property is expressed by any of the relations

$$(S_p^{r-1}D_p^r)_{AB} = \delta_{A_rB_r}(1 - \delta_{b_{p+1}}^{a_{r-1}})\delta_{A^pB^p}, \qquad \text{(A26)}$$

$$(S_p^{r-1}D_q^r)_{AB} = \delta_{A_rB_r}(\delta_{b_q}^{a_{r-1}} - \delta_{b_{q+1}}^{a_{r-1}})\delta_{A^pB^p} \quad \text{if} \quad p < q < r-1, \qquad \text{(A27)}$$

$$(S_p^r D_p^r)_{AB} = \delta_{A_rB_r}(\nu_{b_p}^{a_r} - \nu_{b_{p+1}}^{a_r})\delta_{A^pB^p}. \qquad \text{(A28)}$$

where $\nu_{a_p}^{a_r}$ is defined by Eq. (A2).

(*d*) *Yakubovskii operators* Y_p^r. They are defined for $1 \leqslant p < r \leqslant N$ by

$$Y_p^{p+1} = 1, \quad Y_p^r = 1 + I_{r-2}^r Y_p^{r-1}. \qquad \text{(A29)}$$

They satisfy the relations

$$Y_p^r = \sum_{q=p}^{r-1} D_q^r = D_p^r + Y_{p+1}^r, \qquad \text{(A30)}$$

and the factorization property

$$Y_p^r - Y_q^r = (Y_{q-1}^r - Y_q^r)Y_p^q, \qquad \text{(A31)}$$
$$\text{if } p < q < r \leqslant N.$$

From (A30), (A26), and (A27) one obtains

$$\sum_{A_{p+1}^{r-2}} (Y_p^r)_{AB} = (S_p^{r-1}Y_p^r)_{AB} = \delta_{A_rB_r}\delta_{A^pB^p}. \qquad \text{(A32)}$$

Since from its definition $(Y_p^r)_{AB}$ is a non-negative integer, Eq. (A32) proves that $(Y_p^r)_{AB}$ equals 0 or 1. It proves also that given a chain B and a partition a_{r-1} such that

$$\delta_{b_p}^{a_{r-1}} = \delta_{a_{r-1}}^{b_r} = 1, \qquad \text{(A33)}$$

there exists an unique chain A containing a_{r-1} for which $(Y_p^r)_{AB} = 1$. From (A14) we can write (A32) as

$$S_p^{r-1}Y_p^r = S_p^r. \qquad \text{(A34)}$$

From (A15) one obtains

$$\tilde{Y}_p^r S_p^{r-1} = S_p^r. \qquad \text{(A35)}$$

By iteration one gets

$$S_p^r = Y_p^{p+2}Y_p^{p+3} \cdots Y_p^r = \tilde{Y}_p^r \tilde{Y}_p^{r-1} \cdots \tilde{Y}_p^{p+3}\tilde{Y}_p^{p+2}. \qquad \text{(A36)}$$

(*e*) *Operators* $Z_p^{r(q)}$. They are defined for $1 \leqslant p \leqslant q < r \leqslant N$ by

$$Z_p^{r(q)} = Y_p^{q+1}Y_p^{q+2} \cdots Y_p^{r-1}Y_p^r. \qquad \text{(A37)}$$

Note that

$$Z_p^{r(p)} = Z_p^{r(p+1)} = S_p^r. \qquad \text{(A38)}$$

From (A36) and (A37) one gets

$$S_p^q Z_p^{r(q)} = S_p^r = \tilde{Z}_p^{r(q)} S_p^q. \qquad \text{(A39)}$$

We now prove that $(Z_p^{r(q)})_{AB}$ does not depend on chain A_{q+1}^{r-1}, which can be expressed by

$$(S_q^r Z_p^{r(q)})_{AB} = n_{a_q}^{a_r}(Z_p^{r(q)})_{AB}. \qquad \text{(A40)}$$

Let us first prove (A40) is true for $Z_p^{r(r-2)} = Y_p^{r-1}Y_p^r$. From (A29) and (A23) one derives

$$Z_p^{r(r-2)} = (1 + I_{r-2}^r)Y_p^{r-1} + I_{r-3}^{r-1}I_{r-2}^r Y_p^{r-2}Y_p^{r-1}$$
$$= S_{r-2}^r Y_p^{r-1} + \tilde{D}_{r-3}^r Z_p^{r-1(r-3)}. \qquad \text{(A41)}$$

From (A26) one gets

$$(\tilde{D}_{r-3}^r S_{r-3}^{r-1})_{AB} = \delta_{A_rB_r}(1 - \delta_{a_{r-2}}^{b_{r-1}})\delta_{A^{r-3}B^{r-3}},$$

which proves that $(\tilde{D}_{r-3}^r S_{r-3}^{r-1})_{AB}$ does not depend on a_{r-1}. Let us assume that $(Z_p^{r-1(r-3)})_{AB}$ does not depend on a_{r-2}. Then $(\tilde{D}_{r-3}^r Z_p^{r-1(r-3)})_{AB}$ does not depend on a_{r-1}. By (A41) it follows that $(Z_p^{r(r-2)})_{AB}$ does not depend on a_{r-1}. Since it is true for $(Z_p^{p+2(p)})_{AB} = \delta_{A^{p+2}B^{p+2}}\delta_{A_pB_p}$, it is true for any r. From (A37) one gets

$$Z_p^{r(q)} = Z_p^{q+2(q)}Z_p^{r(q+2)}, \qquad \text{(A42)}$$

which proves that $(Z_p^{r(q)})_{AB}$ does not depend on a_{q+1}. Using the commutation relation (2.9) one obtains

$$(S_q^r Z_p^{r(q)})_{AB} = \sum_{a_{q+1}} (S_{q+1}^r Z_p^{r(q)})_{AB}$$
$$= \sum_{a_{q+1}} (Y_p^{q+1} S_{q+1}^r Z_p^{r(q+1)})_{AB}.$$

Let us assume that (A40) holds for $Z_p^{r(q+1)}$. Then

$$(S_q^r Z_p^{r(q)})_{AB} = \sum_{a_{q+1}} n_{a_{q+1}}^{a_r} (Z_p^{r(q)})_{AB}.$$

Since $(Z_p^{r(q)})_{AB}$ does not depend on a_{q+1}, one obtains (A40) from (A5).

APPENDIX B: CONSTRUCTION OF A COMPLETE BASIS IN $\mathbb{C}_{a_p}^{a_r}$

Our purpose is to construct, in the chain space $\mathbb{C}_{a_p}^{a_r}$ defined in Sec. II, a basis which is adapted, in the particular case $r = N$, to the solution of the YF equation (4.1). The useful results are expressed by Eqs. (B19), (B20), and (B21).

1. Definition of various subspaces of $\mathbb{C}_{a_p}^{a_r}$

We define in this section subspaces $\widetilde{\mathbb{R}}_{a_p}^{a_r}, \mathbb{O}_{a_p}^{a_r}, \widetilde{\mathbb{Q}}_{a_p}^{a_r}, \overline{\mathbb{O}}_{a_p}^{a_r}, \mathbb{R}_{a_p}^{a_r}, \mathbb{Q}_{a_p}^{a_r}$, and $\mathbb{K}_{a_p}^{a_r}$.

Let us first consider the summation operator S_p^r defined by Eq. (A14) which is represented on $\mathbb{C}_{a_p}^{a_r}$ by a square matrix with its $n_{a_p}^{a_r}$ elements equal to 1. Its eigenvalues are $n_{a_p}^{a_r}$ and 0 with degeneracy equal to 1 and $n_{a_p}^{a_r} - 1$, respectively. Let $\widetilde{\mathbb{R}}_{a_p}^{a_r}$ be the one-dimensional subspace corresponding to the eigenvalue $n_{a_p}^{a_r}$. It is the set of vector of $\mathbb{C}_{a_p}^{a_r}$ which have all their components equal.

Let $\mathbb{O}_{a_p}^{a_r}$ be the null space of S_p^r. Since S_p^r is self-adjoint $\mathbb{O}_{a_p}^{a_r}$ is orthogonal to $\widetilde{\mathbb{R}}_{a_p}^{a_r}$ and its dimension is $n_{a_p}^{a_r} - 1$.

$$\mathbb{C}_{a_p}^{a_r} = \widetilde{\mathbb{R}}_{a_p}^{a_r} \oplus \mathbb{O}_{a_p}^{a_r}. \tag{B1}$$

Consider now the set of vectors $|\widetilde{E}\rangle$ of $\mathbb{C}_{a_p}^{a_r}$ with components $\widetilde{E}_{A_{a_p}^{a_r}} = \widetilde{E}_{a_{r-1}}$ depending only on partition a_{r-1} contained in chain $A_{a_p}^{a_r}$. They span a subspace of $\mathbb{C}_{a_p}^{a_r}$ of dimension $\nu_{a_p}^{a_r}$, where $\nu_{a_p}^{a_r}$ is defined by Eq. (A2). This subspace can be decomposed into the direct sum, $\widetilde{\mathbb{R}}_{a_p}^{a_r} \oplus \widetilde{\mathbb{Q}}_{a_p}^{a_r}$ of two orthogonal subspaces

$$\sum_{a_{r-1}} \widetilde{E}_{a_{r-1}} = 0 \quad \text{if} \quad |\widetilde{E}\rangle \in \widetilde{\mathbb{Q}}_{a_p}^{a_r}. \tag{B2}$$

Any vector $|\widetilde{E}\rangle \in \widetilde{\mathbb{R}}_{a_p}^{a_r} \oplus \widetilde{\mathbb{Q}}_{a_p}^{a_r}$ is orthogonal to the null space $\overline{\mathbb{O}}_{a_p}^{a_r}$ of S_p^{r-1} on $\mathbb{C}_{a_p}^{a_r}$.

Considering $\widetilde{Y}_p^r$ defined by Eq. (A29), one obtains

$$(\widetilde{Y}_p^r|\widetilde{E}\rangle)_{A_{a_p}^{a_r}} = \sum_{B_{a_p}^{a_r}} (\widetilde{Y}_p^r)_{A_{a_p}^{a_r} B_{a_p}^{a_r}} \widetilde{E}_{b_{r-1}}$$

$$= \sum_{b_{r-1}} (\widetilde{Y}_p^r S_p^{r-1})_{A_{a_p}^{a_r} B_{a_p}^{a_r}} \widetilde{E}_{b_{r-1}}.$$

Using (A35) one gets

$$(\widetilde{Y}_p^r|\widetilde{E}\rangle)_{A_{a_p}^{a_r}} = \sum_{b_{r-1}} \widetilde{E}_{b_{r-1}}, \tag{B3}$$

$$\widetilde{Y}_p^r|\widetilde{E}\rangle = \nu_{a_p}^{a_r}|\widetilde{E}\rangle \quad \text{if} \quad |\widetilde{E}\rangle \in \widetilde{\mathbb{R}}_{a_p}^{a_r}, \tag{B4}$$

$$\widetilde{Y}_p^r|\widetilde{E}\rangle = 0 \quad \text{if} \quad |\widetilde{E}\rangle \in \widetilde{\mathbb{Q}}_{a_p}^{a_r}. \tag{B5}$$

If $|\widetilde{E}\rangle$ is an eigenvector of $\widetilde{Y}_p^r$ corresponding to eigenvalue λ, then from properties of finite dimensional square matrices it follows that Y_p^r admits an eigenvector $|E\rangle$ corresponding to the same eigenvalue λ, which is not orthogonal to $|\widetilde{E}\rangle$. Thus $|E\rangle$ does not belong to $\overline{\mathbb{O}}_{a_p}^{a_r}$, i.e., $S_p^{r-1}|E\rangle \neq 0$. To sum up, from the existence of these two eigenspaces of $\widetilde{Y}_p^r$, orthogonal to $\overline{\mathbb{O}}_{a_p}^{a_r}$ follows the existence of two corresponding eigenspaces of $Y_p^r, \mathbb{R}_{a_p}^{a_r}$, and $\mathbb{Q}_{a_p}^{a_r}$ of dimension 1 and $\nu_{a_p}^{a_r} - 1$, respectively,

$$Y_p^r|E\rangle = \nu_{a_p}^{a_r}|E\rangle, \; S_p^{r-1}|E\rangle \neq 0 \quad \text{if} \quad |E\rangle \in \mathbb{R}_{a_p}^{a_r}, \tag{B6}$$

$$Y_p^r|E\rangle = 0, S_p^{r-1}|E\rangle \neq 0 \quad \text{if} \quad |E\rangle \in \mathbb{Q}_{a_p}^{a_r}. \tag{B7}$$

Finally, let $\mathbb{K}_{a_p}^{a_r}$ be the null space of Y_p^r. Then

$$\mathbb{Q}_{a_p}^{a_r} \subseteq \mathbb{K}_{a_p}^{a_r}. \tag{B8}$$

2. Eigenspaces of Y_p^r

For $p = r-2$ one has $Y_{r-2}^r = S_{r-2}^r$. Then $\mathbb{K}_{a_{r-2}}^{a_r} \equiv \mathbb{O}_{a_{r-2}}^{a_r}$ and the dimension of

$$\mathbb{K}_{a_{r-2}}^{a_r} = \nu_{a_{r-2}}^{a_r} - 1 = m_{a_{r-2}}^{a_r}, \tag{B9}$$

where the counting numbers $m_{a_p}^{a_r}$ are defined by Eqs. (A8) and (A9).

Let us assume that the dimension of $\mathbb{K}_{a_q}^{a_r}$ is $m_{a_q}^{a_r}$ if $p+2 \leqslant q \leqslant r-2$. We shall then prove recursively that dimension of

$$\mathbb{K}_{a_p}^{a_r} = m_{a_p}^{a_r}. \tag{B10}$$

Given any partition a_q with $\delta_{a_p}^{a_q} = \delta_{a_q}^{a_r} = 1, p+2 \leqslant q \leqslant r-2$, consider the tensor product $\mathbb{K}_{a_q}^{a_r} \otimes \mathbb{Q}_{a_p}^{a_q}$. From Eq. (B7) one has

$$Y_q^r|F\rangle = Y_p^q|F\rangle = 0, \; S_p^{q-1}|F\rangle \neq 0 \quad \text{if} \quad |F\rangle \in \mathbb{K}_{a_q}^{a_r} \otimes \mathbb{Q}_{a_p}^{a_q}. \tag{B11}$$

From the factorization property (A31) and from Eqs. (A34) and (A35) it follows that

$$Y_p^r|F\rangle = 0, S_p^s|F\rangle = 0, q \leqslant s \leqslant r \quad \text{if} \quad |F\rangle \in \mathbb{K}_{a_q}^{a_r} \otimes \mathbb{Q}_{a_p}^{a_q}. \tag{B12}$$

Let $\overline{\mathbb{K}}_{a_p}^{a_r}$ be defined by

$$\overline{\mathbb{K}}_{a_p}^{a_r} = \mathbb{Q}_{a_p}^{a_r} + \sum_{q=p+2}^{r-2} \sum_{a_q} \mathbb{K}_{a_q}^{a_r} \otimes \mathbb{Q}_{a_p}^{a_q}. \tag{B13}$$

From (B11) and (B12) the subspaces which are added in (B13) to give $\overline{\mathbb{K}}_{a_p}^{a_r}$ are linearly independent. The dimension of $\mathbb{Q}_{a_p}^{a_q}$ is $\nu_{a_p}^{a_q} - 1$, we assumed that the dimension of $\mathbb{K}_{a_q}^{a_r}$ is $m_{a_q}^{a_r}$ if $p+2 \leqslant q \leqslant r-2$. Then from (A13)

$$\text{dimension of } \overline{\mathbb{K}}_{a_p}^{a_r} = m_{a_p}^{a_r}. \tag{B14}$$

Since, from (B12), $\overline{\mathbb{K}}_{a_p}^{a_r}$ is a subspace of $\mathbb{K}_{a_p}^{a_r}$, one has

$$\text{dimension of } \mathbb{K}_{a_p}^{a_r} \geqslant m_{a_p}^{a_r}. \tag{B15}$$

We obtain from (B15) that dimension of $\mathbb{R}_{a_q}^{a_r} \otimes \mathbb{K}_{a_p}^{a_q} \geqslant m_{a_p}^{a_q}$ and from (B6) and (A31)

$$Y_p^r|F\rangle = \nu_{a_q}^{a_r}|F\rangle \quad \text{if} \quad |F\rangle \in \mathbb{R}_{a_q}^{a_r} \otimes \mathbb{K}_{a_p}^{a_q}. \tag{B16}$$

Using (A9) one obtains that the number of linearly independent eigenvectors of Y_p^r corresponding to a nonzero eigenvalue is at least equal to $1 + \Sigma_{q=p+2}^{r-1} \Sigma_{a_q} m_{a_p}^{a_q} = n_{a_p}^{a_r} - m_{a_p}^{a_r}$.

Then

$$\text{dimension of } \mathbb{K}_{a_p}^{a_r} \leqslant m_{a_p}^{a_r}. \tag{B17}$$

From (B15) and (B17) one obtains (B10) and, by the way, we have proved that the set of eigenvectors of Y_p^r is complete on $\mathbb{C}_{a_p}^{a_r}$.

We obtain from (B13) and (B14)

$$\mathbb{K}_{a_p}^{a_r} = \mathbb{Q}_{a_p}^{a_r} + \sum_{q=p+2}^{r-2} \sum_{a_q} \mathbb{K}_{a_q}^{a_r} \otimes \mathbb{Q}_{a_p}^{a_q}. \tag{B18}$$

3. Decomposition of $\mathbb{C}_{a_p}^{a_r}$

Since the set of eigenvectors of Y_p^r is complete, $\mathbb{C}_{a_p}^{a_r}$ can be decomposed into the sum of the eigenspaces of Y_p^r, which reads, using (B6) and (B16),

$$\mathbb{C}_{a_p}^{a_r} = \mathbb{R}_{a_p}^{a_r} + \mathbb{O}_{a_p}^{a_r}, \tag{B19}$$

$$\mathbb{O}_{a_p}^{a_r} = \mathbb{K}_{a_p}^{a_r} + \sum_{q=p+2}^{r-1} \sum_{a_q} \mathbb{R}_{a_q}^{a_r} \otimes \mathbb{K}_{a_p}^{a_q}. \tag{B20}$$

Substituting (B18) into (B20) and using (B19) one obtains another decomposition of the null space of S_p^r

$$\mathbb{O}_{a_p}^{a_r} = \mathbb{Q}_{a_p}^{a_r} + \sum_{q=p+2}^{r-1} \sum_{a_q} \mathbb{C}_{a_q}^{a_r} \otimes \mathbb{Q}_{a_p}^{a_q}. \tag{B21}$$

[1]L. D. Faddeev, Zh. Eksp. Teor. Fiz. **39**, 1459 (1960) [Sov. Phys. JETP **12**, 1014 (1961)].

[2]O. A. Yakubovskii, Sov. J. Nucl. Phys. **5**, 937 (1967).

[3]J. W. Evans and D. K. Hoffman, J. Math. Phys. **22**, 2858 (1981).

[4]J. W. Evans, J. Math. Phys. **22**, 1672 (1981).

[5]M. Baer and D. J. Kouri, J. Math. Phys. **14**, 1639 (1973); D. J. Kouri and F. S. Levin, Phys. Lett **50 B**, 421 (1974); W. Tobocman, Phys. Rev. C **9**, 2466 (1974); F. S. Levin, Phys. Rev. C **21**, 2199 (1980).

Perturbative technique as an alternative to the WKB method applied to the double-well potential

H. M. M. Mansour[a] and H. J. W. Müller-Kirsten
Department of Physics, University of Kaiserslautern, 6750 Kaiserslautern, West Germany

(Received 22 September 1981; accepted for publication 7 January 1982)

We give an explicit and complete perturbation theoretical analysis of the solutions and eigenvalues of the Schrödinger equation for the double-well potential. In particular we demonstrate the matching of various branches of the solutions over the entire range of the independent variable, and we calculate the splitting of eigenvalues due to the finite height of the central hump of the potential.

PACS numbers: 03.65.Ge

1. INTRODUCTION

It has been recognized recently that the physical vacuum of a nonabelian gauge theory is degenerate, and that the true vacuum will therefore have to be taken as a superposition of these degenerate vacua.[1] An understanding of the physical consequences of this phenomenon requires an understanding of the tunneling of a particle through a barrier from one vacuum to another. Coleman[2] has discussed the stability of the false vacuum which is related to the eigenvalue gap. He also compared instanton and WKB analyses of the splitting for one dimensional anharmonic oscillators. A rigorous WKB analysis of the gap for the one dimensional case was made by Harrell.[3] The double-well potential has also been treated by Fröman *et al.*[4] by means of certain phase-integral approximations. To first order their phase-integral approximation is identical with the first order WKB approximation. However, as explained in Ref. 5, higher order phase-integral approximations differ from the corresponding higher order WKB approximations. In other approaches, Isaacson[6] discussed singular perturbations resulting in asymptotic eigenvalue degeneracy for the ordinary differential operator

$$H_\nu = \frac{1}{2}\left[-\frac{d^2}{dq^2} + \nu q^4 - q^2 + \frac{1}{4\nu}\right]$$

when $\nu \to 0$, and Brézin *et al.*[7] showed how the large order behavior of perturbation theory is affected by the presence of pseudoparticle–antipseudoparticle contributions to the relevant path integral. Brézin *et al.*[7] confirmed the result obtained previously[8] that in the case of degenerate minima the perturbation series is not Borel-summable. The relation of tunneling solutions to Borel summability has also been studied using the simple quantum-mechanical model of the double-well potential.[9] For such a system the solution corresponding to an instanton is known analytically. Other attempts have been made to study the analytic structure and Borel summability of the perturbation series for the double-well potential.[10,11] In particular, Caswell[12] showed that there exists a summable perturbation series in terms of an effective coupling.

An investigation of the tunneling phenomena may be subdivided into two stages, the first stage consisting in the study of analogous potential models which exhibit most of the essential physical aspects of the problem, and the second in extending or applying the methods used for investigating these simple models to the case of real nonabelian field theories. In the present investigation we are concerned with the first stage. In fact, since the WKB method has been used as the most important tool for investigating tunneling phenomena[13–15] in potential as well as field theoretic models, our main objective here is to demonstrate the usefulness of an alternative procedure which we believe has definite advantages over the WKB method and in particular makes the problem of the matching of various branches of the overall solution particularly transparent. This alternative procedure has been applied previously to a large number of examples such as the Mathieu[16] and other equations,[17,18] Schrödinger equations with Yukawa,[19] Gauss,[20] logarithmic[21] and quark-confining power potentials,[22] and multidimensional and multichannel[23] equations.

In the following we consider first the nonsymmetric double-well potential. Our main objective is to calculate—in the form of asymptotic expansions—the eigenvalues of the wave equation and in particular to investigate the amplitude which describes the tunneling from one well to the other. This investigation is neither simple nor trivial. A secondary objective is to demonstrate the usefulness of our technique.

In Sec. 2 we present the solutions together with their respective eigenvalues derived in the neighborhood of the minima of the potential (regions I and II of Fig. 1). In Sec. 3 we calculate the solutions around the instability point at $x = 0$ (region III). Sections 4 and 5 deal with the solutions in the remaining or intermediate domains (regions IV, V, VI, and VII). Section 6 is devoted to a discussion of the matching of these solutions to one another in their regions of common validity. Finally, in Sec. 7, we discuss the symmetric double well, and we compute the resulting splitting of the energy eigenvalues. This last aspect has also been considered by Damburg and Propin.[24] However, our method is different from theirs and, we believe, more straightforward.

2. EIGENSOLUTIONS AROUND THE MINIMA OF THE POTENTIAL

We derive first the solutions in domains I and II of Fig. 1. In Fig. 1 the potential has the following form:

[a]On leave from the Department of Physics, University of Cairo, Cairo, Egypt.

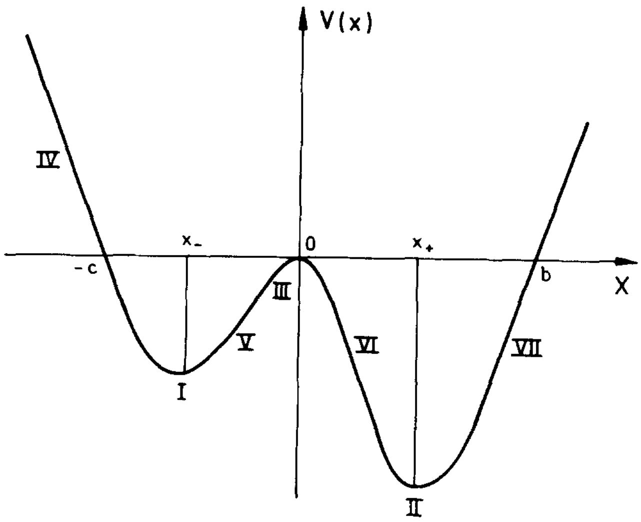


FIG. 1. The double-well potential.

$$V(x) = -ax^2(b-x)(c+x), \quad (1)$$

where $a,b,c > 0$. The first derivative is

$$V^{(1)}(x) = ax(x - x_+)(x - x_-),$$

and

$$x_\pm(b,c) = \tfrac{1}{8}\{3(b-c) \pm [9(b-c)^2 + 32bc]^{1/2}\} \quad (2)$$

(observe that $x_+ > 0$ and $x_- < 0$).

It is readily seen that $V^{(2)}(0) < 0$ and $V^{(2)}(x_\pm) > 0$. Also we observe that for $b = c$, $x_\pm = \pm c/\sqrt{2}$.

Our problem now is to calculate the eigensolutions in the neighborhood of the minima of the potential together with their respective eigenvalues. We use the S-wave equation in the form

$$\frac{d^2\psi}{dx^2} + [\lambda - V(x)]\psi = 0, \quad (3)$$

where $\lambda = 2\mu E/\hbar^2$ in the customary notation. We observe that Eq. (3) is invariant under the combined interchanges $x \leftrightarrow -x, b \leftrightarrow c$, i.e., $x_\pm \to \bar{x}_\pm = -x_\mp$, where $\bar{x}_\pm(b,c) = x_\pm(c,b)$. Expanding $V(x)$ around the minima at $x_\pm$ we have

$$V(x) = V(x_\pm) + \sum_{i=2}^{\infty} \frac{(x - x_\pm)^i}{i!} V^{(i)}(x_\pm). \quad (4)$$

Inserting (4) into (3) we obtain

$$\frac{d^2\psi}{dx^2} + [\lambda - V(x_\pm) - \tfrac{1}{2}(x - x_\pm)^2 V^{(2)}(x_\pm)]\psi = \sum_{i=3}^{\infty} \frac{(x - x_\pm)^i}{i!} V^{(i)}(x_\pm)\psi. \quad (5)$$

We now set

$$h_\pm(x_\pm) = \{2V^{(2)}(x_\pm)\}^{1/4}, \quad (6)$$

where $V^{(2)}(x_\pm)$ is positive, and change the independent variable to

$$\omega_\pm(x) = h_\pm(x_\pm)(x - x_\pm). \quad (7)$$

The equation then becomes

$$\frac{d^2\psi}{d\omega_\pm^2} + \left[\frac{\lambda - V(x_\pm)}{h_\pm^2} - \frac{\omega_\pm^2}{4}\right]\psi = \sum_{i=3}^{\infty} \frac{V^{(i)}(x_\pm)}{2V^{(2)}(x_\pm)} \frac{\omega_\pm^i}{i!\,h_\pm^{i-2}}\psi. \quad (8)$$

For large values of $h_\pm$ the right-hand side of Eq. (8) can—to a first approximation—be neglected. The corresponding behavior of the "eigenvalues" $[\lambda - V(x_\pm)]/h_\pm^2$ can then be determined by comparing the equation with the equation of parabolic cylinder functions. The solutions are normalizable only if

$$\frac{\lambda - V(x_\pm)}{h_\pm^2} = \tfrac{1}{2}q_\pm,$$

where $q_\pm$ is approximately an odd integer, i.e., $2n + 1, n = 0,1,2,\ldots$. The wave functions are correspondingly the parabolic cylinder functions $D_{(1/2)(q_\pm - 1)}(\omega_\pm)$. For the complete solutions we set

$$\frac{1}{h_\pm^2}(\lambda - V(x_\pm)) = \tfrac{1}{2}q_\pm + \frac{\Delta_\pm}{h_\pm}, \quad (9)$$

where $\Delta_\pm$ remains to be determined. We proceed in the standard way.[17]

By substituting Eq. (9) into (8) we obtain

$$D_q\psi = \frac{2\Delta_\pm}{h_\pm}\psi - \sum_{i=3}^{\infty} \frac{V^{(i)}(x_\pm)}{V^{(2)}(x_\pm)} \frac{\omega_\pm^i}{i!\,h_\pm^{i-2}}\psi, \quad (10)$$

where

$$D_q = -2\frac{d^2}{d\omega_\pm^2} - q + \tfrac{1}{2}\omega_\pm^2.$$

Equation (10) can be solved by the perturbation method explained in Refs. 17–22. The first approximation $\psi = \psi^{(0)}$ is the parabolic cylinder function $D_{(q-1)/2}(\omega)$, i.e.,

$$\psi^{(0)} = \psi_q = D_{(q-1)/2}(\omega) \text{ with } D_q\psi_q = 0.$$

For simplicity the subscripts $\pm$ have been dropped. The function ψ_q obeys the recurrence formula[25]

$$\omega\psi_q = (q,q+2)\psi_{q+2} + (q,q-2)\psi_{q-2}, \quad (11)$$

where

$$(q,q+2) = 1, \quad (q,q-2) = \tfrac{1}{4}(q-1).$$

For higher powers we have

$$\omega^i\psi_q = \sum_{j=2i,2i-4,\ldots}^{-2i} S_i(q,j)\psi_{q+j} \quad (12)$$

and a recurrence relation can be written down for the coefficients S_i.[22] The first approximation $\psi = \psi^{(0)}$ leaves uncompensated terms amounting to

$$R_q^{(0)} = \left[\frac{2\Delta}{h} - \sum_{i=3}^{\infty} \frac{V^{(i)}(x_\pm)}{V^{(2)}(x_\pm)} \frac{\omega^i}{i!\,h^{i-2}}\right]\psi_q(\omega)$$

$$= \frac{2\Delta}{h}\psi_q - \sum_{i=3}^{\infty} \frac{1}{h^{i-2}} \sum_{j=2i,2i-4,\ldots}^{-2i} \tilde{S}_i(q,j)\psi_{q+j}(\omega), \quad (13)$$

where

$$\tilde{S}_i(q,j) = \frac{V^{(i)}(x_\pm)}{V^{(2)}(x_\pm)} \frac{1}{i!} S_i(q,j).$$

Hence, the expression for $R_q^{(0)}$ may be written in the form

$$R_q^{(0)} = \sum_{i=3}^{\infty} \frac{1}{h^{i-2}} \sum_{j=2i}^{-2i} [q,q+j]_i \psi_{q+j}(\omega), \tag{14}$$

where

$$[q,q]_3 = 2\Delta - \tilde{S}_3(q,0),$$

and for $j \neq 0$

$$[q,q+j]_3 = -\tilde{S}_3(q,j),$$

and for $i > 3$, $-2i \leqslant j \leqslant 2i$,

$$[q,q+j]_i = -\tilde{S}_i(q,j).$$

Since $D_{q+j} = D_q - j$ and $D_q \psi_{q+j} = j\psi_{q+j}$, a term $\mu\psi_{q+j}$ in $R_q^{(0)}$ can be removed if we add to $\psi^{(0)}$ a term $(\mu/j)\psi_{q+j}$ except when $j = 0$. Hence, the next order contribution of ψ becomes

$$\psi^{(1)} = \sum_{i=3}^{\infty} \frac{1}{h^{i-2}} \sum_{\substack{j=2i,2i-4,\ldots \\ j\neq 0}}^{-2i} \frac{[q,q+j]_i}{j} \psi_{q+j}(\omega), \tag{15}$$

which in turn leaves uncompensated a term

$$R_q^{(1)} = \sum_{i=3}^{\infty} \frac{1}{h^{i-2}} \sum_{\substack{j=2i,\ldots \\ j\neq 0}}^{-2i} \frac{[q,q+j]_i}{j} R_{q+j}^{(0)}. \tag{16}$$

This yields the next contribution of ψ:

$$\psi^{(2)} = \sum_{i=3}^{\infty} \frac{1}{h^{i-2}} \sum_{\substack{j=2i,\ldots \\ j\neq 0}}^{-2i} \frac{[q,q+j]_i}{j} \sum_{i'=3}^{\infty} \frac{1}{h^{i'-2}} \times \sum_{\substack{j'=2i' \\ j+j'\neq 0}}^{-2i'} \frac{[q+j,q+j+j']_{i'}}{j+j'} \psi_{q+j+j'}. \tag{17}$$

Proceeding in this way we obtain the expansion

$$\psi = \psi^{(0)} + \psi^{(1)} + \psi^{(2)} + \cdots, \tag{18}$$

which is an asymptotic expansion in descending powers of h, valid for[23]

$$x - x_\pm = O\left(\frac{1}{h_\pm^{2/3}}\right), \quad |\omega_\pm(x)| \ll h_\pm^{1/3},$$

i.e., around $x = x_\pm$. Another solution in the same domain, i.e., in the region around $x = x_\pm$, is obtained by changing throughout the signs of q and h^2. However, the sum $\psi^{(0)} + \psi^{(1)} + \psi^{(2)} + \cdots$ is a solution only if the sum of the terms in ψ_q in $R_q^{(0)}, R_q^{(1)}, \cdots$ (left unaccounted for so far) is set equal to zero. Hence,

$$0 = \sum_{i=3}^{\infty} \frac{1}{h^{i-2}} [q,q]_i + \sum_{i=3}^{\infty} \frac{1}{h^{i-2}} \sum_{\substack{j=2i \\ j\neq 0}}^{-2i} \frac{[q,q+j]_i}{j} \times \sum_{i'=3}^{\infty} \frac{1}{h^{i'-2}} [q+j,q]_{i'} + \cdots \tag{19}$$

or

$$0 = \frac{1}{h}[q,q]_3 + \frac{1}{h^2}\left\{[q,q]_4 + \sum_{\substack{j=6,4,\ldots \\ j\neq 0}}^{-6} \frac{[q,q+j]_3}{j} [q+j,q]_3\right\} + O\left(\frac{1}{h^3}\right). \tag{20}$$

From this equation we obtain Δ and hence the eigenvalues.

Thus,

$$2h_\pm \Delta_\pm = \left[\frac{q_\pm^2 + 1}{2^4} \frac{V^{(4)}(x_\pm)}{V^{(2)}(x_\pm)} - \frac{15q_\pm^2 + 7}{2^4 . 3^2} \left(\frac{V^{(3)}(x_\pm)}{V^{(2)}(x_\pm)}\right)^2\right] + O\left(\frac{1}{h_\pm^2}\right), \tag{21}$$

where

$$\frac{V^{(4)}(x_\pm)}{V^{(2)}(x_\pm)} = \frac{12}{-3bx_\pm + 3cx_\pm - bc + 6x_\pm^2},$$

and

$$\frac{V^{(3)}(x_\pm)}{V^{(2)}(x_\pm)} = \frac{-3b + 3c + 12x_\pm}{-3bx_\pm + 3cx_\pm - bc + 6x_\pm^2}.$$

Knowing Δ we can obtain λ, i.e.,

$$\lambda_\pm = V(x_\pm) + \tfrac{1}{2}q_\pm h_\pm^2 + h_\pm \Delta_\pm, \tag{22}$$

corresponding to the two minima, where

$$V(x_\pm) = -ax_\pm^2 (b - x_\pm)(c + x_\pm),$$

and

$$h_\pm^4 = -24ax_\pm^2 + 12abx_\pm - 12acx_\pm + 4abc.$$

We observe that these expressions possess the symmetry discussed earlier, i.e., the "$\pm$" versions become the reflected "$\pm$" versions and vice versa under the combined interchanges $x_\pm \leftrightarrow \bar{x}_\pm = -x_\mp$, $b \leftrightarrow c$. The expansion for λ provided by (22) is valid for large values of $h_\pm$ and small q.

For convenience we write the solutions just derived for the domains

$$x - x_\pm = O(1/h_\pm^{2/3}),$$

$$\psi_B(q_\pm, h_\pm; \omega_\pm(x, x_\pm)).$$

It follows from the symmetry of our original Eq. (3) that knowing these solutions an associated solution $\bar{\psi}_B$ is obtained by the combined replacements $x \leftrightarrow -x$ and $x_\pm \leftrightarrow \bar{x}_\pm$, i.e.,

$$\bar{\psi}_B = \psi_B(q_\pm, h_\pm; \omega_\pm(-x, \bar{x}_\pm)).$$

These solutions are valid in the domain

$$-x - \bar{x}_\pm = O\left(\frac{1}{h^{2/3}}\right), \quad h = h(x_\pm),$$

which is in general the mirror reflection of the domain of the solutions ψ_B.

Now, from the symmetry of Eq. (10) it can be seen that a further pair of solutions is obtained by the combined interchanges

$$\omega \to i\omega, \quad q \to -q, \quad h \to ih,$$

or

$$\omega \to -i\omega, \quad q \to -q, \quad h \to -ih.$$

The ensuing solutions require a careful specification. We use formulas given in Ref. 25. Using formula 19.4.6 of Ref. 25, one can show that

$$\mathrm{Re}(-i)^{-(1/2)(q+1)}D_{-(1/2)(q+1)}(-i\omega)$$
$$= \frac{\Gamma[-\frac{1}{2}(q-1)]}{(2\pi)^{1/2}}\{D_{(1/2)(q-1)}(-\omega) - \sin(\tfrac{1}{2}q\pi)D_{(1/2)(q-1)}(\omega)\}\,, \tag{23}$$

where Re means "real part of." We now set

$$\bar{\psi}_q(\omega) = \mathrm{Re}\,\Gamma[\tfrac{1}{2}(q+1)](-i)^{-(1/2)(q+1)}D_{-(1/2)(q+1)}(-i\omega)\,. \tag{24}$$

Then (see Ref. 25, line following formula 19.6.4) $\bar{\psi}_q(\omega)$ again satisfies the recurrence relation (11). Moreover (see Ref. 25, formula 19.4.2), it is that solution which together with $D_{(1/2)(q-1)}(\omega)$ forms a linearly independent pair [here normalized so that their Wronskian is 1, not $(2/\pi)^{1/2}$ as in Ref. 25]. We write these solutions, valid for $x - x_\pm = O(1/h^{2/3})$,

$$\bar{\psi}_c(q_\pm, h_\pm; \omega_\pm(x, x_\pm))\,.$$

Another set of solutions ψ_c, valid in the domains $-x-\bar{x}_\pm = O(1/h^{2/3})$, is obtained by the replacements $x \to -x$, $x_\pm \to \bar{x}_\pm$. The general solution in the first domain around a minimum is then given by the linear combination

$$\psi = \alpha\psi_B + \beta\bar{\psi}_c\,, \tag{25}$$

where α and β are constants.

For later convenience we note here the following particular expressions [see Eq. (24) and Ref. 25, formulas 19.3.5 and 19.3.6]:

$$\psi_q(0) = \frac{\pi^{1/2}2^{(1/4)(q-1)}}{\Gamma[-\frac{1}{4}(q-3)]}\,, \tag{26a}$$

$$\left(\frac{d}{d\omega}\psi_q(\omega)\right)_0 = -\frac{\pi^{1/2}2^{(1/4)(q+1)}}{\Gamma[-\frac{1}{4}(q-1)]}\,, \tag{26b}$$

$$\bar{\psi}_q(0) = \frac{\pi^{1/2}\sin\{(\pi/4)(q+3)\}\Gamma[\frac{1}{2}(q+1)]}{2^{(1/4)(q+1)}\Gamma[\frac{1}{4}(q+3)]}\,, \tag{26c}$$

$$\left(\frac{d}{d\omega}\bar{\psi}_q(\omega)\right)_0 = \frac{\pi^{1/2}\sin\{(\pi/4)(q+1)\}\Gamma[\frac{1}{2}(q+1)]}{2^{(1/4)(q-1)}\Gamma[\frac{1}{4}(q+1)]}\,. \tag{26d}$$

The asymptotic behavior of $\psi_q, \bar{\psi}_q$ is given by (Ref. 25, formulas 19.8.1, 19.8.2)

$$\psi_q(\omega) \simeq e^{-(1/4)\omega^2}\omega^{(1/2)(q-1)}\left[1 + O\left(\frac{1}{\omega^2}\right)\right]\,, \tag{27a}$$

$$\bar{\psi}_q(\omega) \simeq \Gamma[\tfrac{1}{2}(q+1)]e^{+(1/4)\omega^2}\omega^{-(1/2)(q+1)}\left[1 + O\left(\frac{1}{\omega^2}\right)\right]\,. \tag{27b}$$

These expressions will be needed below.

3. WKB-LIKE SOLUTIONS NEAR THE MINIMA OF THE POTENTIAL

We now derive the solutions in regions V and VI of Fig. 1 as reached from the minima. The actual matching will be considered later. Substituting (9) into (3) we have

$$\frac{d^2\psi}{dx^2} + [\tfrac{1}{2}q_\pm h_\pm^2 + \Delta_\pm h_\pm - \tfrac{1}{4}h_\pm^4 v(x)]\psi = 0\,, \tag{28}$$

where

$$v(x) = \frac{4}{h_\pm^4}[V(x) - V(x_\pm)] \simeq (x - x_\pm)^2 + \cdots\,. \tag{29}$$

Since $V^{(1)}(x_\pm) = 0$ it is readily seen that $v(0) = 0$ and $v'(0) = 0$ and $v^{(i)}(0)$ for $i \geqslant 2$ follows from Eq. (32) below. Next we set

$$\psi(x) = \chi(x)\exp\left[\pm\tfrac{1}{2}h_\pm^2\int^x v^{1/2}(x)dx\right]\,. \tag{30}$$

The function $\chi(x)$ satisfies the following equation:

$$\frac{d^2\chi}{dx^2} \pm h_\pm^2 v^{1/2}\frac{d\chi}{dx} \pm \tfrac{1}{4}h_\pm^2\frac{v'}{v^{1/2}}\chi + (\tfrac{1}{2}qh_\pm^2 + \Delta_\pm h_\pm)\chi = 0\,, \tag{31}$$

where

$$v(x) = (x - x_\pm)^2(1 + \alpha_1(x_\pm)(x - x_\pm) + \alpha_2(x_\pm)(x - x_\pm)^2)$$
$$= \sum_{i=2}^{4}\frac{(x - x_\pm)^i}{i!}v^{(i)}(0)\,, \tag{32}$$

$$\alpha_1 \equiv \alpha_1(x_\pm) = \frac{2a}{V^{(2)}(x_\pm)}(4x_\pm - b + c)\,,$$

and

$$\alpha_2 \equiv \alpha_2(x_\pm) = \frac{2a}{V^{(2)}(x_\pm)}\,.$$

Hence we make the important observation that under the combined interchanges $b \leftrightarrow c$, $x_\pm \to -x_\pm$,

$\alpha_1(x_\pm) \mapsto -\alpha_1(x_\mp)$ and $\alpha_2(x_\pm) \mapsto +\alpha_2(x_\mp)$.

Thus [see Eq. (31)], replacing $v^{1/2}(x)$ by $-v^{1/2}(x)$ is equivalent to the replacements $x \to -x$, $b \leftrightarrow c$, and $x_\pm \to -x_\pm$.

In the following we consider the solutions of Eq. (31) with the upper signs. For lower signs the solution in the same domain is obtained by changing the signs of q and h^2 throughout or by replacing $v^{1/2}$ by $-v^{1/2}$.

Equation (31) can then be written in the form

$$D_q\chi = \frac{2}{h^2}\left(\frac{d^2\chi}{dx^2} + \Delta h\chi\right)\,, \tag{33}$$

where

$$D_q = -2v^{1/2}\frac{d}{dx} - \frac{1}{2}\frac{v'}{v^{1/2}} - q\,. \tag{34}$$

We observe that Δh is at most of $O(0)$ in h^2 when $h^2 \to \infty$. Thus, to a first approximation we can neglect the terms on the right-hand side and the solution to that order is

$$\chi^{(0)} = \chi_q\,,$$

where χ_q is a solution of $D_q\chi_q = 0$, i.e.,

$$\chi_q(x) = \frac{C}{v^{1/4}}\exp[K(q,x)]\,, \tag{35a}$$

and

$$K(q,x) = -\frac{q}{2}\int^x\frac{dx}{v^{1/2}(x)}$$
$$= -q\left(\tfrac{1}{2}\ln(x - x_\pm) + \sum_{i=1}^{\infty}\gamma_i(x - x_\pm)^i\right)\,, \tag{35b}$$

and the γ_i's are coefficients which can easily be calculated, e.g., $\gamma_1 = -\alpha_1/4$, etc. Thus

$$\frac{d^2\chi_q}{dx^2}+\Delta h\chi_q=\Big(\Delta h+\frac{5}{16}\frac{v'^2}{v^2}+\frac{q}{2}\frac{v'}{v^{3/2}}+\frac{q^2}{4v}-\frac{v''}{4v}\Big)\chi_q\,. \tag{36}$$

Proceeding as in Ref. 17 we now express the right-hand side of Eq. (36) as a sum over various χ_{q+j} (since the perturbation procedure then becomes particularly simple); we then have the following expansion:

$$\frac{d^2\chi_q}{dx^2}+\Delta h\chi_q=\sum_{j=\bar{2},\bar{1},0}^{\bar{\infty}}(q,q+2j)\chi_{q+2j}\,, \tag{37}$$

where for $i\neq 0$

$$(q,q+2i)=\tfrac{5}{16}\tau_{2i}+\frac{q}{2}\kappa_{2i}+\frac{q^2}{4}\delta_{2i}-\tfrac{1}{4}\epsilon_{2i}\,,$$

and the coefficients τ, κ, δ, and ϵ are obtained in a similar way as in Ref. 23, and for $i=0$

$$\begin{aligned}(q,q)&=\Delta h+\tfrac{5}{16}\tau_0+\frac{q}{2}\kappa_0+\frac{q^2}{4}\delta_0-\tfrac{1}{4}\epsilon_0\\&=\Delta h+\tfrac{5}{16}(16\gamma_2+8\alpha_1\gamma_1-3\alpha_1^2-24\alpha_2)\\&\quad+\frac{q}{2}(8\gamma_2+\alpha_2-\tfrac{3}{4}\alpha_1^2)+\frac{q^2}{4}(4\gamma_2-2\gamma_1\alpha_1+\alpha_1^2-\alpha_2)\\&\quad-\tfrac{1}{4}(8\gamma_2+8\alpha_1\gamma_1+10\alpha_2-4\alpha_1^2)\,.\end{aligned} \tag{38}$$

The first approximation $\chi^{(0)}=\chi_q$ leaves uncompensated on the right-hand side of Eq. (28) a sum of terms amounting to

$$R_q^{(0)}=\frac{2}{h^2}\sum_{j=\bar{2},\bar{1},0}^{\bar{\infty}}(q,q+2j)\chi_{q+2j}\,. \tag{39}$$

Following Ref. 17 the next order contribution to χ is

$$\chi^{(1)}=\frac{2}{h^2}\sum_{\substack{j=\bar{2},\bar{1},\ldots\\ j\neq 0}}^{\bar{\infty}}\frac{(q,q+2j)}{2j}\chi_{q+2j}\,. \tag{40}$$

The coefficient of χ_q in $R_q^{(0)}$ set equal to zero, i.e.,

$$(q,q)=0$$

yields an expression for Δ which is identical to that obtained in the previous section to the same order of iteration (this is an important aspect of our procedure).

The complete solution is obtained in the standard fashion, leading to the sum

$$\chi=\chi^{(0)}+\chi^{(1)}+\chi^{(2)}+\cdots$$

in descending powers of h^2. The corresponding equation for Δ is

$$0=(q,q)+\frac{2}{h^2}\sum_{\substack{j=\bar{2},\bar{1},\ldots\\ j\neq 0}}^{\bar{\infty}}\frac{(q,q+2j)}{2j}(q+2j,q)+\cdots\,. \tag{41}$$

Successive contributions $\chi^{(0)},\chi^{(1)},\ldots$ of χ form a rapidly decreasing sequence, provided that

$$\frac{\chi_{q+2j}}{\chi_q}=\exp\Big[-j\int^x\frac{dx}{v^{1/2}(x)}\Big]<\tfrac{1}{2}h^2\,, \tag{42}$$

which clearly excludes the region around $x=x_\pm$. As mentioned previously, a second solution valid in the same domain is obtained by changing the signs of q and h^2 throughout or by replacing $v^{1/2}$ by $-v^{1/2}$ throughout. We write the solutions, respectively, $\psi_A(q,h;v^{1/2})$ and $\bar{\psi}_A(q,h,v^{1/2})$, where $\psi_A=\chi(x)\exp[+\tfrac{1}{2}h^2\int v^{1/2}dx]$ and

$$\begin{aligned}\bar{\psi}_A(q,h,v^{1/2})&=\psi_A(q,h;-v^{1/2})\\&=\psi_A(q\to -q,h^2\to -h^2;v^{1/2})\,.\end{aligned}$$

4. SOLUTIONS AROUND THE INSTABILITY POINT

Our next step is to derive the solutions in region III of Fig. 1. Clearly, if we reverse the sign of $V(x)$ the potential will have a minimum at this point, and we can calculate the eigenvalues by the method of Sec. 2. This time, however, we expand the potential around $x=0$, i.e., instead of Eq. (5) we have

$$\frac{d^2\psi}{dx^2}+[\lambda+abcx^2]\psi=(ax^4-a(b-c)x^3)\psi\,. \tag{43}$$

Again, we observe the invariance of the equation under the combined interchanges $x\leftrightarrow -x$, $b\leftrightarrow c$. Now we set

$$h=\{-4abc\}^{1/4}\,, \tag{44}$$

and change the independent variable to

$$\omega=hx\,. \tag{45}$$

The equation then becomes

$$\frac{d^2\psi}{d\omega^2}+\Big[\frac{\lambda}{h^2}-\frac{\omega^2}{4}\Big]\psi=\frac{1}{4abc}\Big\{(b-c)\frac{\omega^3}{h}-\frac{\omega^4}{h^2}\Big\}\psi\,. \tag{46}$$

For large values of $|h|$, i.e., a, the right-hand side of the above equation can—to a first approximation—be neglected. The remaining equation has the solution $\psi^{(0)}=D_{(1/2)(q-1)}(\omega)$, where q is given by

$$\lambda/h^2\simeq\tfrac{1}{2}q\,. \tag{47}$$

It should be observed that since the instability point does not support bound states, q is not an integer here (as in Sec. 2); instead it is an auxilliary parameter determined by this equation. For the complete solution we set

$$\frac{\lambda}{h^2}=\tfrac{1}{2}q+\frac{\Delta}{h}\,, \tag{48}$$

where Δ remains to be determined by iteration.

The procedure of solution is now similar to that of Sec. 2. Thus, proceeding as above, we find

$$2h\Delta=\Big[-\frac{1}{2^3}\beta(q^2+1)-\frac{\alpha^2}{2^4.3^2}(15q^2+7)\Big]+O\Big(\frac{1}{h^2}\Big)\,, \tag{49}$$

with

$$\alpha=3\Big(\frac{1}{c}-\frac{1}{b}\Big)\,,$$

and

$$\beta=6/bc\,,$$

and the complete solution of Eq. (37) is

$$\psi_B=\psi^{(0)}+\psi^{(1)}+\psi^{(2)}+\cdots\,, \tag{50}$$

with

$$\psi^{(0)} = \psi_q = D_{(1/2)(q-1)}(\omega),$$

and

$$\psi^{(1)} = \sum_{i=3}^{\infty} \frac{1}{h^{i-2}} \sum_{\substack{j=2i \\ j\neq 0}}^{-2i} \frac{[q,q+j]_i}{j} \psi_{q+j}(\omega),$$

and so on, where

$$[q,q]_3 = 2\Delta - \tilde{S}_3(q,0), \tag{51}$$

$$\tilde{S}_i(q,j) = \frac{v^{(i)}(0)}{v^{(2)}(0)} \frac{1}{i!} S_i(q,j), \tag{52}$$

and the coefficients S_i are defined as in Ref. 17. Also for $j\neq 0$

$$[q,q+j]_3 = -\tilde{S}_3(q,j) \tag{53}$$

and for $i>3$, $-2i<j<2i$,

$$[q,q+j]_i = -\tilde{S}_i(q,j). \tag{54}$$

The solution ψ_B is valid for

$$x = O\left(\frac{1}{h^{2/3}}\right), \quad |\omega(x)| < h^{1/3}.$$

The associated solution obtained by replacing in ψ_B x by $-x$ and b,c by c,b, respectively, is seen to be identical with ψ_B in view of Eq. (50) and the fact that q, i.e., Eq. (48), is invariant under the interchange $b\leftrightarrow c$.

Proceeding as in Sec. 2, a further solution $\bar{\psi}_C$ is obtained by effectively replacing h and ω by $\pm ih$ and $\pm i\omega$ and q by $-q$, this solution being valid for

$$|\omega(x)| < h^{1/3}.$$

The solutions $\psi_B, \bar{\psi}_C$ form a linearly independent pair in the domain of the point of instability.

5. WKB-LIKE SOLUTIONS NEAR THE INSTABILITY POINT

To derive the solutions in regions V and VI of Fig. 1 as reached from the instability point we proceed as in Sec. 3.

Thus, we insert (48) into (46) and obtain

$$\frac{d^2\psi}{dx^2} + \left[\tfrac{1}{2}qh^2 + \Delta h + \frac{h^4}{4bc} v(x)\right]\psi = 0, \tag{55}$$

where

$$v(x) = x^2(x-b)(x-c), \tag{56}$$

and

$$h = \{-4abc\}^{1/4}.$$

The method of solution of Eq. (55) now parallels the method of solution of (28).

Hence, following the same procedure as in Sec. 3, we put

$$\psi(x) = \chi(x) \exp\left[\pm \frac{h^2}{2\lambda^{1/2}} \int^x v^{1/2}(x)\, dx\right], \tag{57}$$

where $\lambda = -bc$ and the function $\chi(x)$ satisfies the following equation:

$$\frac{d^2\chi}{dx^2} \pm \frac{h^2}{\lambda^{1/2}} v^{1/2}(x) \frac{d\chi}{dx} \pm \frac{h^2}{4\lambda^{1/2}} \frac{v'(x)}{v^{1/2}(x)} \chi + (\tfrac{1}{2}qh^2 + \Delta h)\chi = 0. \tag{58}$$

We consider the solutions of the above equation with the upper signs only since for the lower signs the solution of the equation can be obtained by changing the signs of q and h^2 or by the combined interchanges $x\to -x$, $b\leftrightarrow c$. Equation (45) may be written

$$D_q\chi = \frac{2}{h^2}\left(\frac{d^2\chi}{dx^2} + \Delta h\chi\right), \tag{59}$$

where

$$D_q = -\frac{2}{\lambda^{1/2}} v^{1/2} \frac{d}{dx} - \frac{1}{2\lambda^{1/2}} \frac{v'}{v^{1/2}} - q. \tag{60}$$

Thus, to a first approximation we can neglect the terms on the right-hand side and the solution is

$$\chi^{(0)} = \chi_q,$$

where χ_q is a solution of $D_q\chi_q = 0$, i.e.,

$$\chi_q(x) = \frac{c'}{v^{1/4}} \exp[-K(q,x)], \tag{61}$$

where c' is a normalization constant and

$$K(q,x) = \frac{q\lambda^{1/2}}{2} \int^x \frac{dx}{v^{1/2}(x)} = q\left(\tfrac{1}{2}\ln x + \sum_{i=0}^{\infty} \gamma_i' x^i\right) \tag{62}$$

and the γ_i's are given by

$$\gamma_1' = \frac{1}{4}\left(\frac{1}{b} - \frac{1}{c}\right), \text{ etc.}$$

Proceeding as in Sec. 3, we obtain

$$\frac{d^2\chi_q}{dx^2} + \Delta h\chi_q = \left(\Delta h + \frac{5}{16}\frac{v'^2}{v^2} + q\frac{\lambda^{1/2}}{2}\frac{v'}{v^{3/2}} + \frac{q^2\lambda}{2}\frac{1}{v} - \frac{v''}{4v}\right)\chi_q, \tag{63}$$

which can be written in the form

$$\frac{d^2\chi_q}{dx^2} + \Delta h\chi_q = \sum_{j=2,1,\ldots}^{-\infty} (q,q+2j)\chi_{q+2j}, \tag{64}$$

where for $i=0$

$$\begin{aligned}(q,q) = \Delta h &+ \frac{5}{16}\left(16\gamma_2 - 8\gamma_1\frac{(b-c)}{\lambda} + \frac{8}{\lambda} - \frac{3(b-c)^2}{\lambda^2}\right) \\ &+ \frac{q}{2}\left[\frac{4}{\lambda} - \frac{9}{2\lambda^3}(b-c)^2 - \frac{3}{\lambda^2} + \frac{3(b-c)^2}{\lambda^2}\right. \\ &+ \frac{3}{4}\frac{(b-c)^2}{\lambda^4} + (2\gamma_2 - \gamma_1^2)\left(-\frac{3(b-c)}{\lambda} + \frac{3(b-c)}{\lambda^2}\right) \\ &+ \frac{q^2}{4}\left(\frac{4\gamma_2}{\lambda} + \frac{2\gamma_1(b-c)}{\lambda^2} + \frac{(b-c)^2}{\lambda^3} - \frac{1}{\lambda^2}\right) \\ &- \frac{1}{4}\left(-\frac{8\gamma_1}{\lambda}(b-c) + 8\gamma_2 + \frac{10}{\lambda} - \frac{4(b-c)^2}{\lambda^2}\right).\end{aligned} \tag{65}$$

The complete solution $\psi_A(q,h,v^{1/2})$ is now obtained in the standard fashion leading to the sum[17]

$$\psi_A = \chi = \chi^{(0)} + \chi^{(1)} + \chi^{(2)} + \cdots, \tag{66}$$

where

$$\chi^{(1)} = \frac{2}{h^2} \sum_{\substack{j=2,1,\ldots \\ j\neq 0}}^{-\infty} \frac{(q,q+2j)}{2j} \chi_{q+2j} , \tag{67}$$

and so on. Successive contributions $\chi^{(0)}, \chi^{(1)}, \cdots$ of χ form a rapidly decreasing sequence provided that

$$\exp\left[-\lambda^{1/2} \int^x \frac{dx}{v^{1/2}(x)} \right] < \tfrac{1}{2} h^2 , \tag{68}$$

which excludes the region around $x = 0$. A second solution $\bar{\psi}_A(q,h,v^{1/2})$ in the same domain is obtained by changing the signs of q and h^2 or by the replacement $x \to -x$, $b \leftrightarrow c$ in the above solution, i.e.,

$$\bar{\psi}_A(q,h^2;v^{1/2}) = \psi_A(-q,-h^2;v^{1/2}) .$$

6. MATCHING OF SOLUTIONS IN REGIONS OF COMMON VALIDITY

Knowing the solutions in every region of the independent variable we can now consider their continuation in adjoining domains. Since the functions to be joined are asymptotic expansions, we do not call this continuation analytic. In the preceding sections we derived the following solutions:

in Sec. 2,

$$\psi_B(q_\pm, h_\pm; \omega_\pm),$$

and

$$\bar{\psi}_C(q_\pm, h_\pm; \omega_\pm) \propto \psi_B(-q_\pm, ih_\pm; i\omega_\pm) ,$$

which are valid in the domains around the minima, i.e.,

$$x - x_\pm < O\left(\frac{1}{h_\pm^{2/3}}\right),$$

and $\bar{\psi}_B, \psi_C$, which are valid in the corresponding mirror imaged domains

$$-x - \bar{x}_\pm < O\left(\frac{1}{h_\pm^{2/3}}\right);$$

in Sec. 3,

$$\psi_A(q_\pm, h_\pm; v^{1/2}(x)) ,$$

and

$$\bar{\psi}_A(q_\pm, h_\pm; v^{1/2}(x)) = \psi_A(q_\pm, h_\pm; -v^{1/2}(x)) ,$$

which are valid in the domains away from the minima, i.e.,

$$x - x_\pm > O\left(\frac{1}{h^{2/3}}\right);$$

this condition excludes the points $x = x_\pm$;

in Sec. 4,

$$\psi_B(q,h;\omega),$$

and

$$\bar{\psi}_C(q,h;\omega) \propto \psi_B(-q, ih; i\omega) ,$$

which are valid around the instability point, i.e.,

$$|x| < O\left(\frac{1}{h^{2/3}}\right);$$

in Sec. 5,

$$\psi_A(q,h;v^{1/2})$$

and

$$\bar{\psi}_A(q,h;v^{1/2}) ,$$

which are valid for

$$x > O\left(\frac{1}{h^{2/3}}\right),$$

thus excluding the point $x = 0$.

We dub all solutions ψ_B and ψ_C which involve Hermite or parabolic cylinder functions "oscillatorlike" and the solutions ψ_A and $\bar{\psi}_A$ "WKB-like." These solutions can now be matched to each other in the following way (see Fig. 2).

First of all we note that the general solution in the domains of ψ_A and $\bar{\psi}_A$ can be written

$$\psi = \alpha\psi_A + \beta\bar{\psi}_A , \tag{69}$$

where α and β are constants. The matching of the solutions $\psi_A, \bar{\psi}_A$ to the oscillatorlike solutions around a minimum can be easily obtained by going to a common region of validity of $\psi_A, \bar{\psi}_C$ and $\bar{\psi}_A, \psi_B$.[23]

Considering the dominant contribution of ψ_A for x approaching $x_\pm$, we have (with $q = q_\pm, h = h_\pm$)

$$\psi_A \simeq \frac{C}{v^{1/4}(x-x_\pm)^{q/2}} \exp\left[\tfrac{1}{2} h^2 \int^x v^{1/2} dx\right]$$

$$\simeq \frac{Ce^{-(1/4)h^2 x_\pm^2}}{(x-x_\pm)^{(1/2)(q+1)}} \exp\left[\tfrac{1}{4} h^2 (x-x_\pm)^2\right] .$$

On the other hand,

$$\bar{\psi}_C \simeq \bar{\psi}_q(\omega) \simeq \frac{\Gamma[\frac{1}{2}(q+1)] e^{(1/4)h^2(x-x_\pm)^2}}{[h(x-x_\pm)]^{(1/2)(q+1)}} .$$

Hence, in their common region of validity

$$\psi_A = \gamma\bar{\psi}_C , \tag{70}$$

where

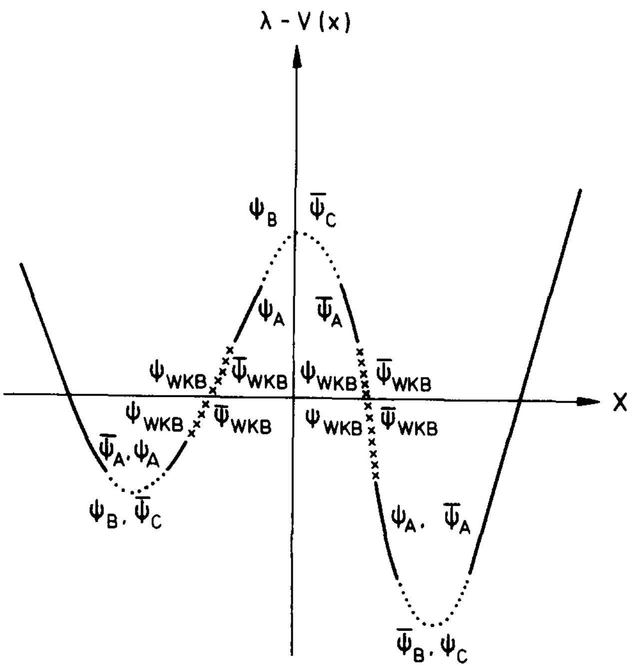


FIG. 2. The domains of various solutions.

$$\gamma = \frac{Ch^{(1/2)(q+1)}e^{-(1/4)h^2x_\pm^2}}{\Gamma[\frac{1}{2}(q+1)]}\left[1+O\left(\frac{1}{h}\right)\right]. \quad (71)$$

Similarly, we obtain

$$\bar{\psi}_A = \bar{\gamma}\psi_B, \quad (72)$$

where

$$\bar{\gamma} = \frac{Ce^{(1/4)h^2x_+^2}}{h^{(1/2)(q-1)}}\left[1+O\left(\frac{1}{h}\right)\right] \quad (73)$$

(C being an overall constant which, like the eigenvalue, remains unchanged under the replacements $q \to -q$, $h \to ih$).

It is not too difficult to calculate the first few terms[23] of the expansions on the right-hand side of these equations. This establishes the continuation of the oscillatorlike solutions at the minima to the WKB-like solutions of Sec. 3. Hence $\psi = \alpha\gamma\bar{\psi}_C + \beta\bar{\gamma}\psi_B$ is the continuation of ψ of Eq. (69) into the region of a minimum. Similarly, we can obtain the relationship between the solutions ψ_B and ψ_C at the instability point and the WKB-like solutions ψ_A and $\bar{\psi}_A$ of Sec. 5. Hence, the problem which remains is the matching of the WKB-like exponential type of solutions across the Stokes singularities corresponding to classical turning points.

We proceed as follows. Equation (3) can be written

$$\frac{d^2y}{dx^2} = X(x)y, \quad (74)$$

with

$$\begin{aligned} X(x) &= V(x) - \lambda \\ &= V(x) - [V(x_\pm) + \tfrac{1}{2}q_\pm h_\pm^2 + \Delta_\pm h_\pm] \\ &= v(x)\frac{h_\pm^4}{4} - \tfrac{1}{2}q_\pm h_\pm^2 - \Delta_\pm h_\pm \end{aligned}$$

on using (9) and (29) with (6). The solutions to this equation above the minimum but just below the turning point at which $\lambda = V(x)$ are

$$\psi_{\rm WKB}(x,q,h) = \frac{1}{X^{1/4}}\left[\exp\left(\int^x X^{1/2}dx\right)\right]Y(x,q,h), \quad (75)$$

and

$$\bar{\psi}_{\rm WKB}(x,q,h) = \frac{1}{X^{1/4}}\left[\exp\left(-\int^x X^{1/2}dx\right)\right]\bar{Y}(x,q,h), \quad (76)$$

where Y satisfies an equation which has been solved by Dingle.[26] We have

$$X^{1/4} = \frac{h_+}{2^{1/2}}[v(x)]^{1/4}\left\{1+O\left(\frac{1}{h_+^2}\right)\right\}, \quad (77)$$

and

$$X^{1/2} = \frac{h_+^2}{2}[v(x)]^{1/2}\left\{1 - \frac{q}{h_+^2 v(x)} + O\left(\frac{1}{h_+^3}\right)\right\}, \quad (78)$$

where $v(x)$ is given by Eq. (29). Comparison with (30) and (35) shows that the WKB solutions join smoothly to the solutions ψ_A and $\bar{\psi}_A$ of Sec. 3. Thus, in their common domain [choosing the overall constant C in (35a) equal to 1]

$$\psi_A = \frac{h_+}{2^{1/2}}\psi_{\rm WKB}\left[1+O\left(\frac{1}{h_+^2}\right)\right], \quad (79)$$

and

$$\bar{\psi}_A = \frac{h_+}{2^{1/2}}\bar{\psi}_{\rm WKB}\left[1+O\left(\frac{1}{h_+^2}\right)\right]. \quad (80)$$

Similarly, we can obtain the relationship between the corresponding solutions $\psi_A, \bar{\psi}_A$ and the WKB solutions above the turning point, i.e., far below the maximum of $V(x)$. In fact, each of the solutions ψ_A and $\bar{\psi}_A$ shifts the turning point at $\lambda = V(x)$ to an extremum, thereby hiding it in the WKB-like solutions of Secs. 3 and 5.

Now the matching of the WKB-like solutions below and above the turning point can be achieved by matching the WKB solutions in the usual way and then matching these to our WKB-like solutions. The connection formulas read[26] to leading order

$$\psi_{\rm WKB}(x,q,h) \leftrightarrow \tfrac{1}{2}(\psi_{\rm WKB}(x,q,h) + i\bar{\psi}_{\rm WKB}(x,q,h)),$$

and

$$\psi_{\rm WKB}(x,q,h) \leftrightarrow (i\psi_{\rm WKB}(x,q,h) + \bar{\psi}_{\rm WKB}(x,q,h)),$$

i.e.,

$$X^{-1/4}\exp\left[-\left|\int^x X^{1/2}dx\right|\right] \leftrightarrow 2(-X)^{-1/4}\sin\left\{\left|\int^x(-X)^{1/2}dx\right| + \frac{\pi}{4}\right\},$$

and

$$X^{-1/4}\exp\left[+\left|\int^x X^{1/2}dx\right|\right] \leftrightarrow (-X)^{-1/4}\cos\left\{\left|\int^x(-X)^{1/2}dx\right| + \frac{\pi}{4}\right\}.$$

Above the turning point (i.e., on the oscillatory side) each of the solutions $\psi_{\rm WKB}$ on the right-hand side of the above relations is to be continued to ψ_A and $\bar{\psi}_A$. In this case one has

$$(-X)^{1/2} = \frac{h^2}{2}(v'(x))^{1/2}\left\{1 - \frac{q}{h^2 v'(x)} + O\left(\frac{1}{h^3}\right)\right\} \quad (81)$$

and

$$(-X)^{1/4} = \frac{h}{2^{1/2}}(v'(x))^{1/2}\left\{1+O\left(\frac{1}{h^2}\right)\right\}, \quad (82)$$

where $v'(x)$ is related to $v(x)$ of Sec. 5 by

$$v'(x) = -v(x)/bc. \quad (83)$$

Hence, the relations between the WKB-like solutions $\psi_A, \bar{\psi}_A$ and the WKB solutions above the turning point are

$$\psi_A = \frac{h}{2^{1/2}}\psi_{\rm WKB}\left[1+O\left(\frac{1}{h^2}\right)\right], \quad (84)$$

and

$$\bar{\psi}_A = \frac{h}{2^{1/2}}\bar{\psi}_{\rm WKB}\left[1+O\left(\frac{1}{h^2}\right)\right]. \quad (85)$$

From the above results we observe that the solutions around the instability point are trigonometric functions in order to match the WKB-like solutions above the turning point. We indicate here schematically how the matching is actually done,[27] starting from the region around x_+:

$$\psi_B(x_+) \to \bar{\psi}_A(x_+) \to \bar{\psi}_{\rm WKB} \leftrightarrow (i\psi_{\rm WKB} + \bar{\psi}_{\rm WKB}) \to (i\psi_A(0) + \bar{\psi}_A(0)) \to (i\gamma\bar{\psi}_C(0) + \bar{\gamma}\psi_B(0)) .$$

Also

$$\bar{\psi}_C(x_+) \to \psi_A(x_+) \to \psi_{\rm WKB} \leftrightarrow \tfrac{1}{2}(\psi_{\rm WKB} + i\bar{\psi}_{\rm WKB}) \to \tfrac{1}{2}(\psi_A(0) + i\bar{\psi}_A(0)) \to \tfrac{1}{2}(\gamma\bar{\psi}_C(0) + i\bar{\gamma}\psi_B(0)) .$$

The matching to the left-hand side of the instability point can be done in the same manner as explained previously.

7. THE SYMMETRIC DOUBLE-WELL POTENTIAL

The special case of a symmetric double well has some particularly interesting features, as we will discuss below.

First of all we note that the general potential given by Eq. (1) may be cast into a symmetric form either by putting $b = c$ or by the transformation

$$x = \frac{b+c}{2d} y - \frac{c-b}{2} . \quad (86)$$

In this case the potential becomes symmetric with two minima at $\pm d$ and a maximum at the point $d(c-b)/(c+b)$, i.e.,

$$ax^2(b-x)(c+x) = a\left(\frac{b+c}{2d}\right)^4 \times \left\{y - \frac{d(c-b)}{c+b}\right\}^2 (y+d)(y-d) . \quad (87)$$

The symmetric potential has two nearly harmonic wells as we can see from the solutions obtained previously, i.e., the parameter

$$q \approx \text{odd integer} = 2n + 1 .$$

A rough explanation for the asymptotic degeneracy is that the two wells asymptotically decouple into independent oscillators and the effect of a second well on the energy (eigenvalue) is inversey proportional to the time a quantum mechanical particle would take to tunnel through the barrier, by the uncertainty principle.[28] Hence, the tunneling probability $\approx \exp[-2\int_{t_1}^{t_2}(V(x) - E)^{1/2}\,dx]$, where t_1 and t_2 are the classical turning points. This in turn is proportional to the energy gap. The energy splitting has been calculated previously[29] and has been treated rigorously by Harrell.[30] The wave functions for a symmetric double oscillator using a phase-integral approximation have been discussed in Refs. 4 and 31 and their normalization is considered in Ref. 32. In Ref. 31 the quantization condition was applied to calculate some energy eigenvalues for a special double-well potential for which the energy eigenvalues have been calculated numerically by Chan and Stelman.[33]

For the symmetric double-well potential we also observe that all the solutions are the same as those obtained previously if we put $b = c$. The only difference is that the domains of validity of the solutions $\psi_B, \bar{\psi}_C$ collapse to one domain, i.e., $x - x_0 = O(1/h_0^{2/3})$, where in this case $x_+ = -x_- = \bar{x}_+ = -\bar{x}_- = x_0$ (say) and $h_+ = h_- = h_0$.

Now, the actual eigenstates must be even or odd about the axis passing through the central maximum of the potential. The degeneracy is then split by the perturbation (i.e., finite height and thus finite cross sectional area of the central hump of the potential) which couples the otherwise independent oscillators, so that the symmetric state lies slightly below the antisymmetric one.

We now proceed to calculate the deviation of q from an exact odd integer q_0 and thence the splitting of the asymptotically degenerate energy levels. For this purpose it is necessary to construct wave functions $\psi_\pm$ which are, respectively, even or odd under the interchange $x \to -x$. Here we make the important observation that this symmetry of the wave functions which is related to the symmetry of the potential is retained only as long as we do not expand around a particular point such as one of the minima of the potential. As soon as we select a minimum and expand the potential or wave function in its neighborhood, this symmetry is violated. It is therefore essential for the construction of the even and odd wave functions to consider solutions which are pure functions of x. Solutions of this type are the WKB-like solutions (30), which we now write $\psi^{(\pm)}$. Thus

$$\psi_\pm = \psi^{(+)}(x) \pm \psi^{(-)}(x) \quad (88)$$

apart from an overall constant, with

$$\psi^{(+)}(x) = \psi_A(q,h;v^{1/2}) ,$$
$$\psi^{(-)}(x) = \psi^{(+)}(-x) . \quad (89)$$

We have, therefore, on using (70) and (72),

$$\psi_\pm = \gamma\bar{\psi}_C(q,h,\omega(x)) \pm \bar{\gamma}\psi_B(q,h,\omega(x)) \quad (90)$$

in the neighborhood of the minimum of the potential at x_0 (specified by appropriate indices attached to q, h, etc.). In the neighborhood of the other minimum these solutions involve ψ_C and $\bar{\psi}_B$ (see Fig. 2).

Figure 3 shows the typical shape of the wave functions $\psi_\pm$ along with the potential $V(x)$. The wave functions $\psi_\pm$

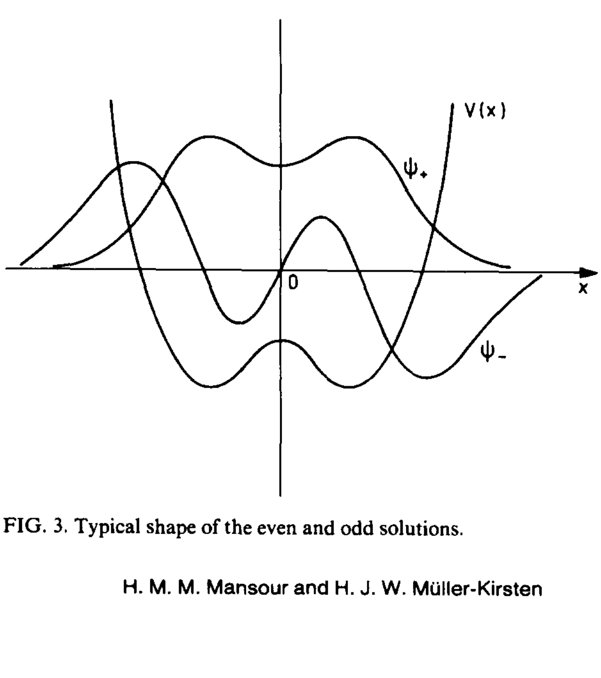


FIG. 3. Typical shape of the even and odd solutions.

characterized by integers q_0 are defined by[30]

$$\left(\frac{\partial\psi_+}{\partial x}\right)_{\pm x_0}=0\,,\quad (\psi_-)_{\pm x_0}=0\,,$$

$$(\psi_+)_{\pm x_0}=+1\,,\quad \left(\frac{\partial\psi_-}{\partial x}\right)_{\pm x_0}=-1. \tag{91}$$

These conditions are such that the Wronskian of ψ_-,ψ_+ is 1 [the constant C in (71) and (73) is chosen to be 1].

Consider ψ_- at $x=x_0$. Then $\omega(x_0)=0$ and $(\psi_-)_{x_0}=0$ yields, to leading order in h,

$$\gamma\bar{\psi}_q(0)\simeq\bar{\gamma}\psi_q(0)\,, \tag{92}$$

which [on using (26), (70), and (72)] leads to

$$\tan\left\{\frac{\pi}{4}(q+1)\right\}=\pi\left(\frac{h^2}{2}\right)^{q/2}\frac{e^{-(1/2)h^2x_0^2}}{\Gamma[\frac{1}{4}(q+1)]\Gamma[\frac{1}{4}(q+3)]}$$
$$=\left(\frac{\pi}{2}\right)^{1/2}h^q\frac{e^{-(1/2)h^2x_0^2}}{\Gamma[\frac{1}{2}(q+1)]}\,. \tag{93}$$

Performing the corresponding calculation for $(\partial\psi_+/\partial x)_{x_0}=0$, we obtain to leading order

$$\gamma\left(\frac{\partial\bar{\psi}_q(\omega)}{\partial\omega}\right)_{x_0}+\bar{\gamma}\left(\frac{\partial\psi_q(\omega)}{\partial\omega}\right)_{x_0}\simeq 0\,, \tag{94}$$

which leads to

$$\cot\left\{\frac{\pi}{4}(q+1)\right\}\simeq\left(\frac{\pi}{2}\right)^{1/2}h^q\frac{e^{-(1/2)h^2x_0^2}}{\Gamma[\frac{1}{2}(q+1)]}\,. \tag{95}$$

Applying the same procedure to $(\psi_+)_{x_0}=+1$, we obtain

$$\tan\left\{\frac{\pi}{4}(q+1)\right\}\simeq-\left(\frac{\pi}{2}\right)^{1/2}h^q\frac{e^{-(1/2)h^2x_0^2}}{\Gamma[\frac{1}{2}(q+1)]}+\pi^{1/2}\left(\frac{h^2}{2}\right)^{(1/4)(q-1)}\frac{e^{-(1/4)h^2x_0^2}}{\cos\{(\pi/4)(q+1)\}\Gamma[\frac{1}{4}(q+1)]}\,. \tag{96}$$

Similarly, $(\partial\psi_-/\partial x)_{x_0}=-1$ yields

$$\cot\left\{\frac{\pi}{4}(q+1)\right\}\simeq-\left(\frac{\pi}{2}\right)^{1/2}h^q\frac{e^{-(1/2)h^2x_0^2}}{\Gamma[\frac{1}{2}(q+1)]}-\frac{\pi^{1/2}h^{(1/2)(q-3)}e^{-(1/4)h^2x_0^2}}{2^{(1/4)(q+1)}\sin\{(\pi/4)(q+1)\}\Gamma[\frac{1}{4}(q+3)]}\,. \tag{97}$$

We now expand $\tan\{(\pi/4)(q+1)\}$ around $q=q_0=3,7,11,\ldots$ so that

$$\tan\left\{\frac{\pi}{4}(q+1)\right\}\simeq(q-q_0)\frac{\pi}{4}+O[(q-q_0)^2]\,. \tag{98}$$

Then, from (93) and (96)

$$q-q_0\simeq\pm 2\left(\frac{2}{\pi}\right)^{1/2}h^{q_0}\frac{e^{-(1/2)h^2x_0^2}}{\Gamma[\frac{1}{2}(q_0+1)]}\,, \tag{99}$$

where $q_0=3,7,11,\ldots$ and the signs $+,-$ apply to ψ_-, ψ_+, respectively. In writing down (99) for ψ_+ we have neglected the second term in (96), which is of lower order in h than the first term.

Expanding $\cot\{(\pi/4)(q+1)\}$ around $q_0=1,5,9,\ldots$, we obtain

$$\cot\left\{\frac{\pi}{4}(q+1)\right\}\simeq-(q-q_0)\frac{\pi}{4}+O[(q-q_0)^2]\,, \tag{100}$$

so that

$$q-q_0\simeq\mp 2\left(\frac{2}{\pi}\right)^{1/2}h^{q_0}\frac{e^{-(1/2)h^2x_0^2}}{\Gamma[\frac{1}{2}(q_0+1)]}\,, \tag{101}$$

where $q_0=1,5,9,\ldots$ and the signs $-,+$ apply to ψ_+,ψ_-, respectively.

We have seen previously that the eigenvalue λ is obtained as a function of q [see, e.g., (21) and (22)]. Thus, expanding λ around q_0 we have

$$\lambda(q)\simeq\lambda(q_0)+\left(\frac{\partial\lambda}{\partial q}\right)_{q_0}(q-q_0)\,. \tag{102}$$

From (22) we see that $(\partial\lambda/\partial q)_{q_0}$ is positive. It follows that the even states for which $q-q_0$ is negative lie below the odd states for which $q-q_0$ is positive. Our formulas thus permit the explicit calculation of this splitting of the asymptotically degenerate energy levels. They also demonstrate the enormous usefulness of the parameter q.

ACKNOWLEDGMENT

One of the authors (HMMM) is indebted to the Deutsche Akademische Austauschdienst (DAAD) for the award of a visiting fellowship.

[1]C. G. Callan, R. Dashen, and D. J. Gross, Phys. Rev. D **17**, 2717 (1978).
[2]S. Coleman, "The uses of instantons," in *Proceedings of the International School of Physics*, Erice, Italy, 1977 (to appear).
[3]E. Harrell, Commun. Math. Phys. **60**, 73 (1978).
[4]N. Fröman, P.-O. Fröman, U. Myhrman, and R. Paulsson, Ann. Phys. (NY) **74**, 314 (1972).
[5]N. Fröman, Ann. Phys. (NY) **61**, 451 (1970).
[6]D. Isaacson, Commun. Pure Appl. Math. **29**, 531 (1976).
[7]E. Brézin, G. Parisi, and J. Zinn-Justin, Phys. Rev. D **16**, 408 (1978).
[8]E. Brézin, J. C. Le Guillon, and J. Zinn-Justin, Phys. Rev. D **15**, 1558 (1977).
[9]E. B. Bogomolny and V. A. Fateyev, Phys. Lett. B **71**, 93 (1977).
[10]W. Y. Crutchfield, Phys. Lett. B **77**, 109 (1978).
[11]J. J. Loeffel, A. Martin, B. Simon, and A. S. Wightman, Phys. Lett. B **30**, 656 (1969).
[12]W. E. Caswell, Ann. Phys. (NY) **123**, 153 (1979).
[13]J.-L. Gervais and B. Sakita, Phys. Rev. D **16**, 3507 (1977).
[14]K. M. Bitar and S.-J. Chang, Phys. Rev. D **18**, 435 (1978).
[15]S.-J. Chang, in *Proceedings of 1979 NATO International Summer Institute for Theoretical Physics, University of Kaiserslautern*, Kaiserslautern, West Germany, edited by W. Rühl (Plenum, New York, 1980).
[16]R. B. Dingle and H. J. W. Müller, J. Reine Angew. Math. **211**, 11 (1962); **216**, 123 (1964); H. J. W. Müller, *ibid.* **211**, 179 (1962).
[17]H. J. W. Müller-Kirsten, Phys. Rev. D **22**, 1962 (1980); **22**, 1952 (1980).

[18]H. J. W. Müller, J. Reine Angew. Math. **211**, 33 (1962); **212**, 26 (1963); Math. Nachr. **31**, 89 (1966); **32**, 49, 157 (1966); Z. Angew. Math. Mech. **44**, 371 (1964); **45**, 29 (1965).
[19]H. J. W. Müller, Ann. Phys. (Leipzig) **15**, 395 (1965); H. J. W. Müller and K. Schilcher, J. Math. Phys. **9**, 255 (1968); H. J. W. Müller-Kirsten and N. Vahedi-Faridi, *ibid.* **14**, 1291 (1973).
[20]H. J. W. Müller, J. Math. Phys. **11**, 355 (1970).
[21]H. J. W. Müller-Kirsten and S. K. Bose, J. Math. Phys. **20**, 2471 (1979).
[22]H. J. W. Müller-Kirsten, C. E. Hite, and S. K. Bose, J. Math Phys. **20**, 1878 (1979); R. S. Kaushal and H. J. W. Müller-Kirsten, *ibid.* **20**, 2233 (1979).
[23]H. J. W. Müller-Kirsten, Phys. Lett. A **70**, 383 (1979); Nuovo Cimento A **63**, 1 (1981); H. J. W. Müller-Kirsten and R. Müller, Phys. Rev. D **20**, 2541 (1979).
[24]R. J. Damburg and R. K. Propin, J. Chem. Phys. **55**, 612 (1971).
[25]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964), Chap. 19.
[26]R. B. Dingle, *Asymptotic Expansions: Their Derivation and Interpretation* (Academic, London, 1973).
[27]H. M. M. Mansour and H. J. W. Müller-Kirsten, Phys. Rev. D **24**, 1033 (1981).
[28]E. M. Harrell, Commun. Math. Phys. **60**, 73 (1978).
[29]E. M. Harrell, "Double Wells," Johns Hopkins University preprint, 1980 (to be published).
[30]E. M. Harrell, Ann. Phys. (NY) **119**, 351 (1979).
[31]N. Fröman and U. Myhrman, Ark. Fys. **40**, 497 (1970).
[32]S. Yngve, J. Math. Phys. **13**, 324 (1972).
[33]S. I. Chan and D. J. Stelman, J. Chem. Phys. **39**, 545 (1963).

Demiański-type metric in Brans–Dicke theory

K. D. Krori
Mathematical Physics Forum, Cotton College, Gauhati-781001 Assam, India
D. R. Bhattacharjee
St. Edmund's College, Shillong-793003, Meghalaya, India

(Received 18 June 1981; accepted for publication 20 July 1981)

A Demiański-like metric is obtained by means of a complex coordinate transformation in the Brans–Dicke theory.

PACS numbers: 04.20.Cv

Newman and Janis[1] (NJ) have given a derivation of the Kerr metric by performing a complex coordinate transformation on the Schwarzschild metric. Newman *et al.*[2] also applied a similar technique to obtain the charged Kerr metric (Kerr–Newman solution). Later Demiański, following the same technique, developed a general metric which contains both Kerr and NUT metrics as special cases. In this paper, we have followed a similar technique and obtained a Demiański-like metric in Brans–Dicke (BD) theory,[4] showing that NJ[1] technique may also be applied to BD theory in a wider context.

There is a BD version of the NUT metric obtained by Sneddon and McIntosh[5] by a method developed by them. They could not obtain a BD version of the Kerr metric by the same method. McIntosh[6] employed a different method to achieve this end. But his Kerr-like family of solutions does not have the spherical symmetry when the rotation is zero and the scalar field is not constant. The metric derived in this paper does not only pass over to NUT-like and Kerr-like ones in BD theory, as special cases, but also formally goes over to BD, NUT, and Kerr solutions and is free from the shortcoming appearing in the solution of McIntosh.[6] The Brans–Dicke line element in isotropic form[4] may be written as

$$ds^2 = e^{2\alpha_0}\left[\frac{1-B/r}{1+B/r}\right]^{2\eta} dt^2 - e^{2\beta_0}(1+B/r)^4\left[\frac{1-B/r}{1+B/r}\right]^{2\xi} \times [dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)], \tag{1}$$

where $\eta = 1/\lambda$ and $\xi = (\lambda - c - 1)/\lambda$.
The line element (1) may be written in the form

$$ds^2 = (1 - 2r_0/\bar r)^\eta du^2 - 2(1 - 2r_0/\bar r)^\sigma du\, d\bar r - \bar r^2(1 - 2r_0/\bar r)^\xi [d\theta^2 + \sin^2\theta\, d\phi^2], \tag{2}$$

where

$$r \to r e^{\beta_0}, \quad t \to t e^{\alpha_0}, \quad r_0 = 2Be^{\beta_0}, \quad \bar r = (1 + r_0/2r)^2, \quad du = dt + (1 - 2r_0/\bar r)^{\xi - \sigma - 1} d\bar r, \quad \sigma = (\eta + \xi - 1)/2. \tag{3}$$

The contravariant components of the metric coefficients of (2) may be written in the form

$$g^{\mu\nu} = l^\mu n^\nu + l^\nu n^\mu - m^\mu \bar m^\nu - m^\nu \bar m^\mu \tag{4}$$

with

$$l^\mu = \delta_1^\mu, \quad n^\mu = (1 - 2r_0/\bar r)^{-\sigma}\delta_0^\mu - \tfrac12(1 - 2r_0/\bar r)^{1-\xi}\delta_1^\mu,$$
$$m^\mu = (1/2^{1/2}\bar r)(1 - 2r_0/\bar r)^{-\xi/2}[\delta_2^\mu + (i/\sin\theta)\delta_3^\mu], \tag{5}$$

where $\bar m^\mu$ is the complex conjugate of m^μ (the bars on r are dropped).

The coordinate r is now allowed to take complex values and the tetrad is rewritten in the form (r' indicates complex value of r)

$$l^\mu = \delta_1^\mu,$$
$$n^\mu = \left[1 - \left\{r_0\left(\frac{1}{r'} + \frac{1}{r}\right) + \frac{b}{i}\left(\frac{1}{r'} - \frac{1}{r}\right)\right\}\right]^{-\sigma}\delta_0^\mu - \tfrac12\left[1 - \left\{r_0\left(\frac{1}{r'} + \frac{1}{r}\right) + \frac{b}{i}\left(\frac{1}{r'} - \frac{1}{r}\right)\right\}\right]^{1-\xi}\delta_1^\mu,$$
$$m^\mu = \frac{1}{2^{1/2}r'}\left[1 - \left\{r_0\left(\frac{1}{r'} + \frac{1}{r}\right) + \frac{b}{i}\left(\frac{1}{r'} - \frac{1}{r}\right)\right\}\right]^{-\xi/2} \times [\delta_2^\mu + (i/\sin\theta)\delta_3^\mu], \tag{6}$$
$$\bar m^\mu = \frac{1}{2^{1/2}r}\left[1 - \left\{r_0\left(\frac{1}{r'} + \frac{1}{r}\right) + \frac{b}{i}\left(\frac{1}{r'} - \frac{1}{r}\right)\right\}\right]^{-\xi/2} \times [\delta_2^\mu - (i/\sin\theta)\delta_3^\mu],$$

where b is an arbitrary constant defined in (8).

We now formally perform the complex coordinate transformation

$$r' = r + iF(\theta,\phi), \quad \theta' = \theta, \quad u' = u + iG(\theta,\phi), \quad \phi' = \phi, \tag{7}$$

on the tetrad vectors, where $F(\theta,\phi)$ and $G(\theta,\phi)$ are real functions of θ and ϕ and are given as

$$F = a\cos\theta + c\cos\theta \ln\tan\theta/2 + c + b,$$
$$G = -a\cos\theta - 2b\ln\sin\theta - c\cos\theta\ln\tan\theta/2, \tag{8}$$

where a, b, and c are constants. If we now restrict r' and u' to be real, we obtain the following tetrad

$$l^{\mu'} = \delta_1^{\mu'},$$
$$n^{\mu'} = \left[1 - \frac{2r_0r' + 2bF}{r'^2 + F^2}\right]^{-\sigma}\delta_0^{\mu'} - \frac12\left[1 - \frac{2r_0r' + 2bF}{r'^2 + F^2}\right]^{1-\xi}\delta_1^{\mu'},$$
$$m^{\mu'} = \frac{1}{2^{1/2}(r' + iF)}\left[1 - \frac{2r_0r' + 2bF}{r'^2 + F^2}\right]^{-\xi/2}[iH\csc\theta\,\delta_0^{\mu'} - i(H + 2b\cos\theta)\csc\theta\,\delta_1^{\mu'} + \delta_2^{\mu'} + (i/\sin\theta)\delta_3^{\mu'}], \tag{9}$$

where

$$H = a\sin^2\theta - 2b\cos\theta + c\sin^2\theta\ln\tan\theta/2 - c\cos\theta. \tag{10}$$

The metric coefficient $g^{\mu\nu}$ now takes the form

$$g^{\mu'\nu'} = l^{\mu'}n^{\nu'} + l^{\nu'}n^{\mu'} - m^{\mu'}\bar m^{\nu'} - m^{\nu'}\bar m^{\mu'}, \tag{11}$$

where $\bar{m}^{\mu'}$ is the complex conjugate on $m^{\mu'}$. Using (9), (10), and (11), the metric coefficient of the Demiański-like metric in BD theory may easily be obtained.

A further simplification is now made by another coordinate transformation[7] so as to bring the required line element as close as possible to the standard form.

The desired line element may finally be written as (dropping the primes on r)

$$ds^2\left[1-\frac{2r_0r+2bF}{r^2+F^2}\right]^{\eta}(dt-H\,d\phi)^2 - \left[1-\frac{2r_0r+2bF}{r^2+F^2}\right]^{\xi}(r^2+F^2)\left(\frac{dr^2}{\Delta}+d\theta^2+\sin^2\theta\,d\phi^2\right) + 2\left[1-\frac{2r_0r+2bF}{r^2+F^2}\right]^{\sigma}(H+2b\cos\theta)(dt-H\,d\phi)d\phi, \quad (12)$$

where

$$\Delta=(r^2+F^2-2r_0r+2bF)+(H\csc\theta+2b\cot\theta)^2.$$

The expression for Φ (scalar field) is

$$\Phi=\Phi_0[1-(2r_0r+2bF)/(r^2+F^2)]^{\sigma}, \quad (13)$$

A check has been made on the Brans–Dicke field equations and it has been found that (12) and (13) satisfy them.

If $b=c=0$, the metric (12) and the scalar field (13) pass over to a Kerr-like metric in BD theory, and in addition if $\eta=\lambda=1$ and $\xi=0$, the Kerr metric is recovered.

If $a=c=0$, the metric (12) and the scalar field (13) pass over to a NUT-like metric in BD theory, and further, when $\eta=\lambda=1$, $\xi=0$, the NUT metric is readily obtained.

Finally, with $a=b=c=0$, the metric (12) and the scalar field (13) go over to the BD solution.

[1]E. T. Newman and A. I. Janis, J. Math. Phys. **6**, 915 (1965).
[2]E. T. Newman, E. Couch, R. Chinnapared, A. Exton, A. Prakash, and R. Torrence, J. Math. Phys. **6**, 918 (1965).
[3]M. Demiański, Phys. Lett. **42**, 157 (1972).
[4]C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).
[5]G. E. Sneddon and C. B. G. McIntosh, Aust. J. Phys. **27**, 411 (1974).
[6]C. B. G. McIntosh, Commun. Math. Phys. **37**, 335 (1974).
[7]R. Adler, M. Bazin, M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1965).

On the invertibility of Møller morphisms

Hans Maassen
Institute of Theoretical Physics, Rijksuniversiteit Groningen, Groningen, The Netherlands

(Received 20 January 1982; accepted for publication 26 March 1982)

Local perturbations of the dynamics of infinite quantum systems are considered. It is known that, if the Møller morphisms associated to the dynamics and its perturbation are invertible, the perturbed evolution is isomorphic to the unperturbed one, and thereby shares its ergodic properties. It was claimed by V. Ya. Golodets [Theor. Math. Phys. **23**, 525 (1975)] that the above condition holds whenever the observable algebra is asymptotically abelian for the unperturbed evolution, and the perturbed evolution has a KMS state. The present paper contains a counterexample to this statement, and a construction of a spatial representation of the Møller morphisms.

PACS numbers: 05.30.Ch, 05.70.Ln

I. INTRODUCTION

Let us consider a quantum mechanical system that can be described by a C^*-algebra $\mathscr{A}$ and a group $\{\alpha_t\}_{t\in\mathbb{R}}$ of *-automorphisms of $\mathscr{A}$. We interpret $\mathscr{A}$ as the set of (bounded) observables of the system, and $\{\alpha_t\}$ as its dynamics. For $A\in\mathscr{A}$, $t\to\alpha_t(A)$ is the time evolution of the observable A. In Refs. 1 and 2 it is assumed that $t\to\alpha_t(A)$ is continuous. This seemingly innocent assumption excludes many important cases from the discussion as, for instance, the free Bose gas. It is, however, not vital for the conclusions to be drawn here, so let us also make the assumption, for the sake of simplicity. Being strongly continuous, the group $\{\alpha_t\}$ has an infinitesimal generator, δ say,

$$\alpha_t = \exp(t\delta). \tag{1}$$

Now, let V be any self-adjoint element of $\mathscr{A}$, and define

$$\widetilde{\alpha}_t = \exp[t(\delta + [iV,\cdot])]. \tag{2}$$

$\{\widetilde{\alpha}_t\}_{t\in\mathbb{R}}$ is another strongly continuous group of *-automorphisms of $\mathscr{A}$, which we shall call "the perturbed dynamics". Now suppose that $\{\alpha_t\}$ has some nice ergodic property. It may be that $\{\alpha_t\}$ is ergodic:

$$\{A\in\mathscr{A}\,|\,\forall_t:\alpha_t(A)=A\} = \mathbb{C}\mathbf{1}, \tag{3}$$

i.e., $\{\alpha_t\}$ has no nontrivial fixed points ("constants of the motion"). Or it may be that $\{\mathscr{A},\alpha\}$ is asymptotically abelian, i.e.,

$$\forall_{A,B\in\mathscr{A}}:\|[A,\alpha_t(B)]\| \underset{|t|\to\infty}{\longrightarrow} 0. \tag{4}$$

In these cases, it is interesting to know whether or not $\{\widetilde{\alpha}_t\}$ shares the ergodic property. In order to answer these, and related questions, it was proposed by Robinson[1] to study the limits

$$\gamma_\pm(A) = \lim_{t\to\pm\infty} \widetilde{\alpha}_{-t}\circ\alpha_t(A) \tag{5}$$

in the norm topology of $\mathscr{A}$. Suppose these limits exist for all $A\in\mathscr{A}$. (A sufficient condition for this was given in Ref. 1). Then $\gamma_\pm$ are isometric *-morphisms of $\mathscr{A}$, intertwining α and $\widetilde{\alpha}$:

$$\gamma_\pm\circ\alpha_t = \widetilde{\alpha}_t\circ\gamma_\pm\,. \tag{6}$$

Clearly, if γ_+ or γ_- is invertible, $\{\widetilde{\alpha}_t\}$ is similar to $\{\alpha_t\}$, and inherits its ergodic properties.

The maps $\gamma_\pm$ are called the "Møller morphisms", by analogy with the Møller operators in scattering theory. Now, in scattering theory, the nonunitary of the Møller operators is generally thought of as due to to the existence of bound states for the perturbed Hamiltonian. It turns out that, analogously, we may consider the noninvertibility of $\gamma_\pm$ as roughly equivalent to the existence of nontrivial fixed points of $\{\widetilde{\alpha}_t\}$, i.e., constants of the motion for the perturbed evolution. In fact, if $\{\widetilde{\alpha}_t\}$ has a fixed point that is not a fixed point of $\{\alpha_t\}$, then γ_+ are not invertible.

It follows from a result of Araki[3] that, whenever there exists an $\{\alpha,\beta\}$-KMS state ω on $\mathscr{A}$ for some $\beta>0$ (i.e., a state, satisfying the Kubo–Martin–Schwinger condition[4] at inverse temperature β w.r.t. $\{\alpha_t\}$), there also is an $\{\widetilde{\alpha},\beta\}$-KMS state, and it is quasiequivalent to ω. This holds regardless of the existence of fixed points for $\{\widetilde{\alpha}_t\}$.

In view of the above remarks, Theorem 3 of Ref. 2 is surprising. Indeed, we shall see that it is not valid.

II. A COUNTEREXAMPLE

Let us assume that

(I) $\{\mathscr{A},\alpha\}$ is asymptotically abelian [i.e., (4) holds],
(II) the limits $\gamma_\pm(A)$ in (5) exists for all $A\in\mathscr{A}$,
(III) for some $\beta>0$ there is an $\{\widetilde{\alpha},\beta\}$-KMS state $\widetilde{\omega}$ on $\mathscr{A}$.

Let $\widetilde{\pi}$ be the representation of $\mathscr{A}$ determined by $\widetilde{\omega}$ according to the Gel'fand–Naimark–Segal (GNS) construction. It is the content of Theorem 3 of Ref. 2 that, under these assumptions, there exist *-automorphisms $\bar{\gamma}_\pm$ of $\widetilde{\pi}(\mathscr{A})''$, such that for all $A\in\mathscr{A}$

$$\bar{\gamma}_\pm\circ\widetilde{\pi}(A) = \widetilde{\pi}\circ\gamma_\pm(A). \tag{7}$$

The following example shows that this cannot be true.

Let H be the self-adjoint operator $-\partial^2/\partial x^2$ on $L^2(\mathbb{R})$. For $g\in L^2(\mathbb{R})$, let P_g denote the orthogonal projection on g. If $g\in L^1\cap L^2(\mathbb{R})$ is such that $\int g\,dx\neq 0$, the operator $\widetilde{H} := H - P_g$ has an eigenvector $h\neq 0$.

Let $\mathscr{B}$ be the C^*-algebra, embodying the canonical anticommutation relations (CAR) over $L^2(\mathbb{R})$, and let $\mathscr{A}$ be its even subalgebra. Then the groups $\{\alpha_t\}$ and $\{\widetilde{\alpha}_t\}$ of automorphisms of $\mathscr{B}$, defined by

$$\alpha_t(a(f)) = a(e^{itH}f) \quad\text{and}\quad \widetilde{\alpha}_t(a(f)) = a(e^{it\widetilde{H}}f), \tag{8}$$

 0022-2488/82/101848-04$02.50

are related by (1) and (2), with $V = -a(g)^*a(g)/\|g\|^2$. $\{\alpha_t\}$ and $\{\tilde{\alpha}_t\}$ both leave $\mathscr{A}$ invariant, and V is in $\mathscr{A}$.

The system $\{\mathscr{A},\alpha\}$ is asymptotically abelian because $\exp(itH)$ tends to zero weakly as $|t|\to\infty$. Furthermore, by Kato's theorem[5] on perturbations of rank one, the strong limits

$$W_\pm = \lim_{t\to\pm\infty} e^{-it\tilde{H}}e^{itH} \tag{9}$$

exist, and $W_\pm$ are isometries onto the absolutely continuous spectral subspace of $\tilde{H}$. Because h is an eigenvector of $\tilde{H}$, it follows that

$$h \perp \text{Range}(W_\pm). \tag{10}$$

Now, define the *-morphisms $\gamma_\pm : \mathscr{B}\to\mathscr{B}$ by

$$\gamma_\pm(a(f)) = a(W_\pm f).$$

Then $\gamma_\pm(A)$ are indeed the norm limits of $\tilde{\alpha}_{-t}\circ\alpha_t(A)$ as $t\to\pm\infty$ because of (8) and (9) and the continuity of $a(f)$ in f. Moreover, for any $\beta>0$ there is an $\{\tilde{\alpha},\beta\}$-KMS state on $\mathscr{A}$; namely the gauge invariant quasifree state $\tilde{\omega}$ with two-point function:

$$\tilde{\omega}(a(f_1)^*a(f_2)) = \langle f_2, F(\tilde{H})f_1\rangle,$$

where

$$F(x) = (1+e^{\beta x})^{-1}.$$

Now, because $h\perp \text{Range}(W_\pm)$, $a(h)^*a(h)$ commutes with any element of the range of $\gamma_\pm$, a fact that contradicts (7). Indeed, let $\tilde{\pi}$ be the representation determined by $\tilde{\omega}$, and assume that automorphisms $\bar{\gamma}_\pm$ satisfying (7) exist. Then $\tilde{\pi}(\gamma_\pm(\mathscr{A}))'' = \tilde{\pi}(\mathscr{A})''$, and this would lead to the conclusion that $\tilde{\pi}(a(h)^*a(h))$ commutes with $\tilde{\pi}(\mathscr{A})''$, a contradiction.

Remark: The above example describes a noninteracting one-dimensional Fermi gas in a rank one "potential" P_g. The perturbed one-particle evolution has a bound state h, and consequently there is a constant of the motion for the perturbed evolution of the gas, namely, the observable $a(h)^*a(h)$, counting the particles in the bound state. As the unperturbed evolution is ergodic (i.e., has no constants of the motion), the two evolutions are not isomorphic, and $\gamma_\pm$ cannot be invertible. If the claim to be disproved had been that $\gamma_\pm$ are automorphisms of $\mathscr{A}$, our argument could stop here. However, only the existence and invertibility of $\bar{\gamma}_\pm$ is actually asserted, and it could be that $\bar{\gamma}_\pm^{-1}$, mapping $\tilde{\pi}(\mathscr{A})''$ into itself, did not leave $\tilde{\pi}(\mathscr{A})$ invariant. Therefore we need a slightly different argument, the one presented above, based on the fact that any fixed point of $\tilde{\alpha}$ commutes with $\gamma_\pm(\mathscr{A})$ if $\mathscr{A}$ is asymptotically abelian for α.

III. A PRELIMINARY RESULT

In what follows, we will have a closer look at the action of $\tilde{\pi}(\gamma_\pm(\mathscr{A}))$ on the Hilbert space $\tilde{H}$. The following result, taken from Ref. 2, will enable us to do this:

Lemma 1: Suppose conditions (I), (II), and (III) hold. Let $\{\tilde{H},\tilde{\pi},\tilde{\xi}\}$ be the GNS-triple associated to $\{\mathscr{A},\tilde{\omega}\}$. Then there is $\xi\in\tilde{H}$, cyclic and separating for $\tilde{\pi}(\mathscr{A})''$, such that the state ω, defined by

$$\omega(A) = \langle \xi, \tilde{\pi}(A)\xi\rangle, \tag{11}$$

is an $\{\alpha,\beta\}$-KMS state, and

$$\omega = \tilde{\omega}\circ\gamma_\pm. \tag{12}$$

The vector ξ can be chosen to lie in the positive cone[3] $\mathscr{P}_{\tilde{\xi}}$ of $\tilde{\xi}$.

Remark: The reverse is also true: If there is an $\{\alpha,\beta\}$-KMS state ω on $\mathscr{A}$, then there is a $\tilde{\xi}$ in the GNS-space of $\{\mathscr{A},\omega\}$ that implements an $\{\tilde{\alpha},\beta\}$-KMS state $\tilde{\omega}$, satisfying (12). In fact, this reversed statement is the more useful one. In examples where α is "simple", the existence of ω is easier to establish than that of $\tilde{\omega}$. I choose to state the less useful version in order to agree with Ref. 2. Let me emphasize on the other hand that it would certainly not be advisable to entirely interchange α and $\tilde{\alpha}$, and to replace condition (II) of the existence of $\gamma_\pm$ by a condition ($\tilde{\text{II}}$), the existence of

$$\tilde{\gamma}_\pm(A) = \lim_{t\to\pm\infty} \alpha_{-t}\circ\tilde{\alpha}_t(A).$$

In examples where α is "simple", ($\tilde{\text{II}}$) is much harder to test than (II).

Proof: Let $\mathscr{Z}$ be the center of $\tilde{\pi}(\mathscr{A})''$, i.e., $\mathscr{Z} = \tilde{\pi}(\mathscr{A})''\cap\tilde{\pi}(\mathscr{A})'$. By the perturbation theory of KMS states,[3] there exists $\eta\in\mathscr{P}_{\tilde{\xi}}$, cyclic and separating for $\tilde{\pi}(\mathscr{A})''$, such that $A\to\langle\eta,\tilde{\pi}(A)\eta\rangle$ is $\{\alpha,\beta\}$-KMS. Now consider the states $Z\to\langle\eta,Z\eta\rangle$ and $Z\to\langle\tilde{\xi},Z\tilde{\xi}\rangle$. They are both faithful normal states on $\mathscr{Z}$. It follows that there is a vector $\xi\in\overline{\mathscr{Z}^+\eta}$ such that

$$\forall_{Z\in\mathscr{Z}}: \langle\xi,Z\xi\rangle = \langle\tilde{\xi},Z\tilde{\xi}\rangle. \tag{13}$$

This vector ξ is also in the cone $\mathscr{P}_{\tilde{\xi}}$, and is cyclic and separating for $\mathscr{Z}$. Let ω be given by ξ as in (11). It is not hard to show that, because $\xi\in\overline{\mathscr{Z}\eta}$, ω is $\{\alpha,\beta\}$-KMS. And then ξ must be cyclic and separating for the whole of $\tilde{\pi}(\mathscr{A})''$.

Let us now prove that

$$\forall_{A\in\mathscr{A}}: \lim_{t\to\pm\infty} \tilde{\omega}\circ\alpha_t(A) = \omega(A). \tag{14}$$

Suppose the contrary. Then there are $A\in\mathscr{A}$, $\epsilon>0$, and a sequence $\{t_n\}$ of times, such that $|t_n|\to\infty$ and

$$|\tilde{\omega}\circ\alpha_{t_n}(A) - \omega(A)| \geqslant \epsilon, \tag{15}$$

Now, the sequence $\{\tilde{\pi}(\alpha_{t_n}(A))\}_{n\in\mathbb{N}}$ must have a w^*-converging subnet, because it remains inside the w^*-compact set $\{X\in\tilde{\pi}(\mathscr{A})'' \mid \|X\|\leqslant\|A\|\}$. So let $\{n(\sigma)\}$ be a net in $\mathbb{N}$, such that $\lim_\sigma n(\sigma) = \infty$ and $w^*-\lim_\sigma \tilde{\pi}(\alpha_{t_{n(\sigma)}}(A)) = Z\in\tilde{\pi}(\mathscr{A})''$. Then for all $B\in\mathscr{A}$

$$\begin{aligned}[\tilde{\pi}(B),Z] &= w^* - \lim_\sigma [\tilde{\pi}(B),\tilde{\pi}(\alpha_{t_{n(\sigma)}}(A))] \\ &= w^* - \lim_\sigma \tilde{\pi}([B,\alpha_{t_{n(\sigma)}}(A)]) = 0,\end{aligned}$$

because $\{\mathscr{A},\alpha\}$ is asymptotically abelian. So $Z\in\mathscr{Z}$, and we can apply (13). But then

$$\begin{aligned}\lim_\sigma \tilde{\omega}(\alpha_{t_{n(\sigma)}}(A)) &= \lim_\sigma\langle\tilde{\xi},\tilde{\pi}(\alpha_{t_{n(\sigma)}}(A))\tilde{\xi}\rangle = \langle\tilde{\xi},Z\tilde{\xi}\rangle = \langle\xi,Z\xi\rangle \\ &= \lim_\sigma \langle\xi,\tilde{\pi}(\alpha_{t_{n(\sigma)}}(A))\xi\rangle = \lim_\sigma \omega(\alpha_{t_{n(\sigma)}}(A)) = \omega(A),\end{aligned}$$

because ω is α-invariant. This contradicts (15) and we conclude that (14) holds. Finally, note that for all $A\in\mathscr{A}$,

$$\tilde{\omega}\circ\gamma_{\pm}(A) = \lim_{t\to\pm\infty} \tilde{\omega}\circ\tilde{\alpha}_{-t}\circ\alpha_t(A)$$
$$= \lim_{t\to\pm\infty} \tilde{\omega}\circ\alpha_t(A) = \omega(A). \blacksquare$$

IV. INVERTIBILITY OF γ AND EXISTENCE OF $\bar{\gamma}$

For ease of notation, let us from now on identify $A\in\mathscr{A}$ with the operator $\tilde{\pi}(A)$ on $H=\tilde{H}$, so that $\mathscr{A}$ becomes a C^*-algebra of bounded operators on a Hilbert space. Moreover, let us focus our attention on only one of the Møller morphisms: γ_+, and call it γ.

In Ref. 2 a counterpart θ to the map γ: $\mathscr{A}\to\mathscr{A}$ is introduced. θ acts on the commutant $\mathscr{A}'$, which is a von Neumann algebra, unlike $\mathscr{A}$ itself. I shall give a direct construction of θ below.

Lemma 2: Suppose conditions (I), (II), and (III) hold. Then there is an isometry Ω: $H\to H$, such that for all $A\in\mathscr{A}$,

$$\gamma(A)\Omega = \Omega A.$$

Proof: Define Ω_0: $\mathscr{A}\xi\to\gamma(\mathscr{A})\tilde{\xi}$ by

$$\Omega_0 A\xi = \gamma(A)\tilde{\xi}.$$

Then for all $A\in\mathscr{A}$, $\|\Omega_0 A\xi\|^2 = \|\gamma(A)\tilde{\xi}\|^2 = \langle\tilde{\xi},\gamma(A)^*\gamma(A)\tilde{\xi}\rangle = \langle\tilde{\xi},\gamma(A^*A)\tilde{\xi}\rangle = \tilde{\omega}\circ\gamma(A^*A) = \omega(A^*A) = \langle\xi,A^*A\xi\rangle = \|A\xi\|^2$. As $\overline{\mathscr{A}\xi} = H$, Ω_0 extends continuously to an isometry $\Omega{:}H\to H$ with range, $\gamma(\mathscr{A})\tilde{\xi}$. Now for all $A,B\in\mathscr{A}$,

$$\gamma(A)\Omega B\xi = \gamma(A)\gamma(B)\tilde{\xi} = \gamma(AB)\tilde{\xi} = \Omega AB\xi,$$

and the statement follows from the cyclicity of ξ for $\mathscr{A}$. $\blacksquare$

Lemma 3: Suppose (I), (II), and (III) hold. Let ξ be given by Lemma 1 and Ω by Lemma 2. Let J be the modular conjugation $H\to H$, associated with $\{\mathscr{A}'',\xi\}$. Then

$$J\Omega = \Omega J.$$

Proof: Let Δ and $\tilde{\Delta}$ be the modular operators associated with $\{\mathscr{A}'',\xi\}$ and $\{\mathscr{A}'',\tilde{\xi}\}$ according to the Tomita–Takesaki theory[6] and J and $\tilde{J}$ the corresponding modular conjugations.

Let $A\in\mathscr{A}$ be analytic for $\{\alpha_t\}$. Then by the intertwining property (6) of γ, $\gamma(A)$ is analytic for $\{\tilde{\alpha}_t\}$ and

$$\begin{aligned}\tilde{J}\Omega\alpha_{i\beta/2}(A)\xi &= \tilde{J}\gamma(\alpha_{i\beta/2}(A))\tilde{\xi}\\ &= \tilde{J}\tilde{\alpha}_{i\beta/2}(\gamma(A))\tilde{\xi} = \tilde{J}\tilde{\Delta}^{1/2}\gamma(A)\tilde{\xi}\\ &= \gamma(A)^*\tilde{\xi} = \gamma(A^*)\tilde{\xi} = \Omega A^*\xi\\ &= \Omega J\Delta^{1/2}A\xi = \Omega J\alpha_{i\beta/2}(A)\xi.\end{aligned}$$

Now, the linear space $\{\alpha_{i\beta/2}(A)\xi \mid A\in\mathscr{A}$ analytic for $\alpha\}$ is dense in H. Therefore $\tilde{J}\Omega = \Omega J$. And because ξ and $\tilde{\xi}$ are in the same positive cone,[3] J and $\tilde{J}$ coincide, and the statement follows. $\blacksquare$

Lemma 4: Suppose (I), (II), and (III) hold. Let J, Ω be as defined before, and let

$$\theta{:}\mathscr{A}'\to\mathscr{L}(H){:}A\to\Omega^*A\Omega,$$

$$j{:}\mathscr{A}''\to\mathscr{A}'{:}A\to JAJ.$$

Then

$$\theta\circ j\circ\gamma = j\restriction\mathscr{A}.$$

Moreover,

$$\theta(\mathscr{A}')\subset\mathscr{A}'.$$

Proof: First we show that $\theta(\mathscr{A}')\subset\mathscr{A}'$. Let $B\in\mathscr{A}'$, $A\in\mathscr{A}$. Then, by Lemma 2,

$$\begin{aligned}[\theta(B),A] &= [\Omega^*B\Omega,A] = \Omega^*B\Omega A - A\Omega^*B\Omega\\ &= \Omega^*B\gamma(A)\Omega - \Omega^*\gamma(A)B\Omega\\ &= \Omega^*[B,\gamma(A)]\Omega = 0.\end{aligned}$$

So $\theta(B)\in\mathscr{A}'$ for all $B\in\mathscr{A}'$. Furthermore, it follows from Lemmas 2 and 3 that, if $A\in\mathscr{A}$,

$$\begin{aligned}\theta\circ j\circ\gamma(A) &= \Omega^*J\gamma(A)J\Omega = \Omega^*J\gamma(A)\Omega J\\ &= \Omega^*J\Omega AJ = \Omega^*\Omega JAJ = JAJ = j(A). \quad\blacksquare\end{aligned}$$

Lemma 5: Again suppose that (I), (II), (III) hold. Then θ, defined in Lemma 4, is the unique map $\mathscr{A}'\to\mathscr{A}'$ satisfying

$$\forall_{A\in\mathscr{A}}\forall_{B\in\mathscr{A}'}{:}\langle\tilde{\xi},B\gamma(A)\tilde{\xi}\rangle = \langle\xi,\theta(B)A\xi\rangle. \tag{16}$$

Moreover, θ is linear, $*$-preserving, w^*-continuous, and surjective.

Proof: Let $A\in\mathscr{A}$, $B\in\mathscr{A}'$. Then, by Lemma 2, and because $\tilde{\xi} = \Omega\xi$,

$$\begin{aligned}\langle\tilde{\xi},B\gamma(A)\tilde{\xi}\rangle &= \langle\Omega\xi,B\gamma(A)\Omega\xi\rangle\\ &= \langle\xi,\Omega^*B\Omega A\xi\rangle = \langle\xi,\theta(B)A\xi\rangle.\end{aligned}$$

Uniqueness of θ follows from the cyclicity of ξ for $\mathscr{A}$. Clearly, θ is linear, $*$-preserving, and w^*-continuous. It remains to prove surjectivity. So let $B\in\mathscr{A}'$; $\|B\| = 1$, say. We look for an $X\in\mathscr{A}'$ such that $B=\theta(X)$. Now, because j is a bijection $\mathscr{A}''\to\mathscr{A}'$, $j^{-1}(B)$ is a well-defined element of $\mathscr{A}''$; $\|j^{-1}(B)\| = 1$. By Kaplanski's density theorem the unit sphere in $\mathscr{A}$ is dense in the unit sphere in $\mathscr{A}''$. So there is a net $\{B_\sigma\}$ in $\mathscr{A}$ with $\|B_\sigma\|\leqslant 1$ and $w^*\text{-}\lim B_\sigma = j^{-1}(B)$. Now consider the net $\{j\circ\gamma(B_\sigma)\}$. Being included in the w^*-compact unit ball of $\mathscr{A}'$, it must have a w^*-converging subnet $\{B_{\sigma(\tau)}\}$,

$$w^*\text{-}\lim_\tau j\circ\gamma(B_{\sigma(\tau)}) = X\in\mathscr{A}'.$$

But then it follows from Lemma 4 that

$$\theta(X) = w^*\text{-}\lim_\tau \theta\circ j\circ\gamma(B_{\sigma(\tau)})$$

$$= w^*\text{-}\lim_\tau j(B_{\sigma(\tau)}) = j(j^{-1}(B)) = B,$$

because both θ and j are w^*-continuous. $\blacksquare$

Remark: In Ref. 2, (16) is the defining property of θ. The w^*-continuity and surjectivity of θ are also proved there. But, in addition, it is claimed that θ has the morphism property

$$\forall_{A,B\in\mathscr{A}'}{:}\theta(AB) = \theta(A)\theta(B), \tag{17}$$

which is now easily seen not to hold if Ω is not unitary, i.e., if $\gamma(\mathscr{A})\tilde{\xi}$ is not dense in H. And, indeed, a close look at the proof of (17) in Ref. 2 reveals that the w^*-density of $\gamma(\mathscr{A})$ in $\mathscr{A}''$ is implicitly assumed there. Once accepting (17), Golodets can prove that $\bar{\gamma}$ exists as an automorphism of $\mathscr{A}''$ by turning the argument around that has proved the existence of θ as an automorphism of $\mathscr{A}'$. Actually, the existence of the automorphism $\bar{\gamma}$ and (17) are equivalent:

Theorem 6: Suppose that the conditions (I), (II), (III) hold, and let Ω be given by Lemma 2, and θ be as defined in Lemma 4. Then the following statements are equivalent:

(i) There is a *-automorphism $\bar{\gamma}$ of $\mathscr{A}''$, such that $\bar{\gamma} \restriction \mathscr{A} = \gamma$,

(ii) $\gamma(\mathscr{A})' \subset \mathscr{A}'$,

(iii) $\Omega H = H$,

(iv) For all $A,B \in \mathscr{A}' : \theta(AB) = \theta(A)\theta(B)$.

Proof: (i)⟹(ii): Suppose (i) holds. Then $\bar{\gamma}$ is w^*-continuous, and therefore $\bar{\gamma}(\mathscr{A})'' \supset \bar{\gamma}(\mathscr{A}'') = \mathscr{A}''$. It follows that $\gamma(\mathscr{A})' = \bar{\gamma}(\mathscr{A})' = (\bar{\gamma}(\mathscr{A})'')' \subset (\mathscr{A}'')' = \mathscr{A}'$.

(ii)⟹(iii): Suppose (ii) holds. Let $P = \Omega\Omega^*$. P is the orthogonal projection on $\overline{\gamma(\mathscr{A})\tilde{\xi}}$, so $P \in \gamma(\mathscr{A})'$, and therefore $P \in \mathscr{A}'$ by (ii). Now $P\tilde{\xi} = \tilde{\xi}$, so $(P - \mathbf{1})\tilde{\xi} = 0$, and because $\tilde{\xi}$ is separating for $\mathscr{A}'$, $P = \mathbf{1}$. It follows that $\Omega H = H$.

(iii)⟹(i): Suppose Ω is unitary; define $\bar{\gamma}(A) = \Omega A \Omega^*$ for all $A \in \mathscr{A}''$. Then for $A \in \mathscr{A}$ we have $\bar{\gamma}(A) = \Omega A \Omega^* = \gamma(A)\Omega\Omega^* = \gamma(A)$, so $\bar{\gamma} \restriction \mathscr{A} = \gamma$. $\bar{\gamma}$ is clearly a *-morphism, and we have to show that it is onto. Let $A \in \mathscr{A}''$ and let $B = \Omega^* A \Omega$. As J commutes with Ω, $B = J\Omega^* JAJ\Omega J = j^{-1} \circ \theta \circ j(A)$; so $B \in \mathscr{A}''$ by Lemma 4. Moreover, $\bar{\gamma}(B) = \Omega B \Omega^* = \Omega\Omega^* A \Omega\Omega^* = A$. We conclude that any $A \in \mathscr{A}''$ is of the form $\bar{\gamma}(B)$, $B \in \mathscr{A}''$.

(iii)⟹(iv): If Ω is unitary then, for all $A,B \in \mathscr{A}'$, $\theta(A)\theta(B) = \Omega^* A \Omega\Omega^* B \Omega = \Omega^* AB \Omega = \theta(AB)$.

(iv)⟹(iii): Suppose (iv) holds. Then for all $A \in \mathscr{A}'$ $\|\Omega^* A\tilde{\xi}\|^2 = \langle \tilde{\xi}, A^* \Omega\Omega^* A\tilde{\xi} \rangle = \langle \xi, \Omega^* A^* \Omega\Omega^* A \Omega \xi \rangle = \langle \xi, \theta(A^*)\theta(A)\xi \rangle = \langle \xi, \theta(A^* A)\xi \rangle = \langle \xi, \Omega^* A^* A \Omega \xi \rangle = \langle \tilde{\xi}, A^* A\tilde{\xi} \rangle = \|A\tilde{\xi}\|^2$, and because $\tilde{\xi}$ is cyclic for $\mathscr{A}'$, Ω^* is an isometry. Hence Ω is unitary, and $\Omega H = H$. ■

ACKNOWLEDGMENT

I wish to thank Professor N. M. Hugenholtz for many helpful discussions.

[1] D. W. Robinson, "Return to Equilibrium," Commun. Math. Phys. **31**, 171–189 (1973).

[2] V. Ya. Golodets, "Local perturbations of the dynamics of infinite systems," Theor. Math. Phys. **23**, 525–532 (1975) [Theor. Mat. Fiz. **23**, 300 (1975)].

[3] H. Araki, "Positive Cone, Radon–Nikodym theorems, relative Hamiltonian and the Gibbs condition in statistical mechanics. An Application of the Tomita–Takesaki Theory," in *C*-Algebras and Their Applications to Statistical Mechanics and Quantum Field Theory*, edited by D. Kastler (Editrice Composition, Bologna, 1975), pp. 64–99.

[4] R. Haag, N. M. Hugenholtr, and M. Winnink, "On the equilibrium states in quantum statistical mechanics," Commun. Math. Phys. **5**, 215–236 (1967).

[5] T. Kato, *Perturbation theory for linear operators* (Springer, Berlin, 1966).

[6] M. Takesaki, "Tomita's theory of modular Hilbert algebras and its applications," in *Lecture Notes in Mathematics* (Springer, Berlin, 1970), Vol. 128.

On the relativistic Bose–Einstein integrals

Howard E. Haber and H. Arthur Weldon
Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104

(Received 4 August 1981; accepted for publication 23 December 1981)

Two integrals which appear in the study of the relativistic Bose gas are analyzed. The complete low-temperature and high-temperature expansions are computed.

PACS numbers: 05.30.Jp, 03.30. + p, 05.70. − q, 02.30. + g

1. INTRODUCTION

There have been many attempts in the past to study the properties of a relativistic ideal Bose gas.[1–8] It is fairly easy to derive integral expressions for the various thermodynamic quantities. Unfortunately, the integrals obtained cannot be evaluated exactly and various approximation schemes must be employed. In the past, a number of authors attempted[1,2,4,7] to obtain high-temperature expansions for the thermodynamic variables. The leading term of the high-temperature expansions were easily obtained, but their methods failed, in general, to determine further terms in the expansion. In Ref. 8, we pointed out that the expressions used by past authors failed to include the possibility of particle–antiparticle pair production. When this feature was included in the analysis, we were able to derive the relevant high-temperature expansions. Here we will provide a detailed analysis on how to obtain the full high-temperature expansions.

The plan of the paper is as follows. In Sec. 2, we introduce the integrals to be studied. Sections 3 and 4 discuss the high temperature expansions of those integrals. Our technique is to reduce the integral expressions to contour integrals in the complex plane, which may then be computed by summing over residues of single and double poles. This we do in Sec. 3. In Sec. 4, we briefly describe the computations which lead to the desired expansions. For completeness, we also discuss the full low-temperature expansion in Sec. 5 and indicate its use in obtaining the first relativistic corrections to the standard nonrelativistic thermodynamic results. We have included some relevant mathematical information in Appendices A–C and have collected the results of the high temperature expansion in Appendix D.

2. THE INTEGRALS

We consider the following general problem: Calculate all thermodynamic quantities for a relativistic ideal Bose gas in n space dimensions. To do this, we introduce two functions:

$$g_n(y,r)=\frac{1}{\Gamma(n)}\int_0^\infty x^{n-1}dx\left[\frac{1}{\exp[(x^2+y^2)^{1/2}-ry]-1}\right], \tag{1}$$

$$h_n(y,r)=\frac{1}{\Gamma(n)}\times\int_0^\infty \frac{x^{n-1}dx}{(x^2+y^2)^{1/2}}\left[\frac{1}{\exp[(x^2+y^2)^{1/2}-ry]-1}\right]. \tag{2}$$

These functions are related to the thermodynamic potential $\Omega(T,V,\mu)$ in n space dimensions

$$\begin{aligned}\frac{\Omega}{V}&=\frac{2T^{n+1}}{(4\pi)^{n/2}\Gamma(n/2)}\int_0^\infty x^{n-1}dx\\&\quad\times\{\ln[1-\exp[-(x^2+y^2)^{1/2}+ry]]+(r\to -r)\}\\&=-2\pi^{-[(n+1)/2]}\Gamma[(n+3)/2]T^{n+1}\\&\quad\times\{h_{n+2}(y,r)+h_{n+2}(y,-r)\},\end{aligned} \tag{3}$$

where $y\equiv m/T$ and $r\equiv\mu/m$ (μ is the chemical potential). The two terms in (3) correspond to the contribution of particles and antiparticles, respectively. [Those authors who ignored the antiparticles did not include the second terms in brackets in (3)]. Given Ω, one may calculate all other thermodynamic quantities by computing certain derivatives.[9] It is therefore useful to obtain the following relations satisfied by g_n and h_n:

$$\frac{\partial}{\partial y}g_{n+1}=\frac{-y}{n}g_{n-1}+rnh_{n+1}+\frac{ry^2}{n}h_{n-1}, \tag{4}$$

$$\frac{\partial}{\partial r}g_{n+1}=ynh_{n+1}+\frac{y^3}{n}h_{n-1}, \tag{5}$$

$$\frac{\partial}{\partial y}h_{n+1}=\frac{-y}{n}h_{n-1}+\frac{r}{n}g_{n-1}, \tag{6}$$

$$\frac{\partial}{\partial r}h_{n+1}=\frac{y}{n}g_{n-1}. \tag{7}$$

Note that the recursion relations above connect g_n and h_n with even n among themselves and also connect g_n and h_n with odd n among themselves. This suggests that we will have to analyze separately the cases of even and odd n. Furthermore, it is sufficient to compute the expansions for $h_n(y,r)$; then the expansions for $g_n(y,r)$ may be obtained by using (7). We therefore turn to the computation of the high-temperature expansion of $h_n(y,r)$.

3. THE HIGH TEMPERATURE EXPANSION: PART I

We now analyze (2) in the limit of $y\to 0$ at fixed r. Recall that $y\equiv m/T$ and $r\equiv\mu/m$ so that this limit corresponds to the high-temperature limit. For convenience, we will always take $y\geqslant 0$. Consider then h_n as a complex function of r. It is easy to see that it has branch points at $r=1$ and is analytic in the complex r plane cut from $r=1$ to $r=\infty$. [It follows that Ω given by (3) is analytic in the complex r plane cut from $r=1$ to ∞ and $r=-1$ to $-\infty$. In particular, Ω is real valued only for real r satisfying $-1\leqslant r\leqslant 1$. This condition corresponds to the physical requirement that the occupation numbers n_k of particles and antiparticles be positive for all momenta k.[8]]

The first step in the computation involves expanding

 0022-2488/82/101852-07$02.50

$$\frac{1}{\exp[(x^2+y^2)^{1/2}-ry]-1}=\sum_{p=1}^{\infty}e^{ryp}e^{-p(x^2+y^2)^{1/2}}. \tag{8}$$

Inserting (8) into (2), we may integrate term by term (see Appendix A) and obtain

$$h_n(y,r)=\frac{1}{\Gamma[(n+1)/2]}\left(\frac{y}{2}\right)^{(n-1)/2}\times\sum_{p=1}^{\infty}e^{ryp}p^{(1-n)/2}K_{(n-1)/2}(yp), \tag{9}$$

where K is a modified Bessel function. To insure convergence, $|r|<1$ and $n>2$. In fact, we will be able to relax these requirements later by the method of analytic continuation.

We are interested in obtaining an analytic expression for the high-temperature (small y) expansion of h_n. As it stands, (9) is inconvenient for this purpose. For example, if $r>0$, one easily sees that as y is made smaller, more terms in the sum must be kept. Our goal is to obtain an expansion for h_n, where fewer terms need be kept as $y\to 0$. To accomplish this goal, we begin by making use of the Mellin summation formula[10] to evaluate the sum over p in (9). Using (A2) for the Mellin transform of the summand, we obtain

$$h_n(y,r)=\frac{y^{n-1}}{i2\pi^{1/2}\Gamma[(n+1)/2]}\times\int_{c-i\infty}^{c+i\infty}ds\,\frac{\Gamma(s)\Gamma(s-n+1)\zeta(s)}{[y(1-r)]^s\Gamma(s-\frac{1}{2}n+1)}\times{}_2F_1\left(s,\frac{n}{2};s-\frac{n}{2}+1;\frac{r+1}{r-1}\right), \tag{10}$$

where $\operatorname{Re}c>n-1$. To evaluate the contour integral, we may close the path to the left (see Fig. 1). The value of the contour integral around the large arc is zero in the limit of infinite radius. Hence by the residue theorem, the value of the integral (10) is given by the sum of the residues of the poles of the integrand. (Note that all the poles are enclosed by the contour of Fig. 1.) We now apply some hypergeometric function identities to (10) to separate terms even or odd in r:

$${}_2F_1\left(s,\frac{n}{2};s-\frac{n}{2}+1;\frac{r+1}{r-1}\right)=\pi^{1/2}\left(\frac{1-r}{2}\right)^s\Gamma\left(s-\frac{n}{2}+1\right)\times\left\{\frac{{}_2F_1\left(\frac{s}{2},\frac{s-n+1}{2};\frac{1}{2};r^2\right)}{\Gamma\left(\frac{s+1}{2}\right)\Gamma\left(\frac{s-n+2}{2}\right)}+2r\frac{{}_2F_1\left(\frac{s+1}{2},\frac{s-n+2}{2};\frac{3}{2};r^2\right)}{\Gamma\left(\frac{s}{2}\right)\Gamma\left(\frac{s-n+1}{2}\right)}\right\}. \tag{11}$$

This is valid for $|\arg(1\pm r)|<\pi$ (which is equivalent to $|r|<1$ if r is real). Note, however, that the left-hand side of (11) is analytic in the r plane cut from $r=1$ to $r=\infty$. Thus, later on we should be able to analytically continue our results to the region on the real axis where $-\infty<r<-1$. For now, we will assume that r is real such that $|r|<1$.

Using (11) in (10) allows us to break up h_n into pieces even and odd in r. Therefore, we define

$$h_n^e(y,r)=\tfrac{1}{2}[h_n(y,r)+h_n(y,-r)], \tag{12}$$

$$h_n^o(y,r)=\tfrac{1}{2}[h_n(y,r)-h_n(y,-r)]. \tag{13}$$

We then find

$$h_n^e(y,r)=\frac{1}{4\Gamma[(n+1)/2]}\left(\frac{y}{2}\right)^{n-1}\sum_{\text{Res}}\left(\frac{y}{2}\right)^{-s}\Gamma\left(\frac{s}{2}\right)\Gamma\left(\frac{s-n+1}{2}\right)\zeta(s){}_2F_1\left(\frac{s}{2},\frac{s-n+1}{2};\frac{1}{2};r^2\right), \tag{14}$$

$$h_n^o(y,r)=\frac{r}{2\Gamma[(n+1)/2]}\left(\frac{y}{2}\right)^{n-1}\sum_{\text{Res}}\left(\frac{y}{2}\right)^{-s}\Gamma\left(\frac{s+1}{2}\right)\Gamma\left(\frac{s-n+2}{2}\right)\zeta(s){}_2F_1\left(\frac{s+1}{2},\frac{s-n+2}{2};\frac{3}{2};r^2\right). \tag{15}$$

The evaluation of (14) and (15) requires us to sum over the residues (Res) of the poles of the functions specified. In general, there are both single and double poles although this, in part, depends on whether n is even or odd. We will therefore evaluate the two cases of n even and odd spearately. Equations (14) and (15) are the basic equations which will provide the high-temperature ($y\ll 1$) expansion; we now turn to the calculation.

4. THE HIGH TEMPERATURE EXPANSION: PART II

In this section we will summarize the details of the calculation of $h_{2l+1}^e(y,r)$ (i.e., $n=2l+1$ is odd). We emphasize that this function when $l=2$ is relevant for the calculation of the ideal Bose gas in three space dimensions [see (3)]. We will then make some brief comments on the calculation of the other functions of interest. For the reader's convenience, we have displayed the final results for the high-temperature (small y) expansions of the h_n in Appendix D.

To compute the residues of the poles of $h_{2l+1}^e(y,r)$ we need to study the singularities in s of

$$f(s)=\left(\frac{y}{2}\right)^{-s}\Gamma\left(\frac{s}{2}\right)\Gamma\left(\frac{s-2l}{2}\right)\zeta(s){}_2F_1\left(\frac{s}{2},\frac{s-2l}{2};\frac{1}{2};r^2\right). \tag{16}$$

We are assuming (for now) that $|r|<1$; hence the only singularities of $f(s)$ are due to the gamma and the Riemann zeta function. Specifically, $f(s)$ has single poles at $s=1,2,4,\ldots,2l$ and double poles at $s=0,-2,-4,\ldots$. The residues at the single poles are easy to calculate; we find that

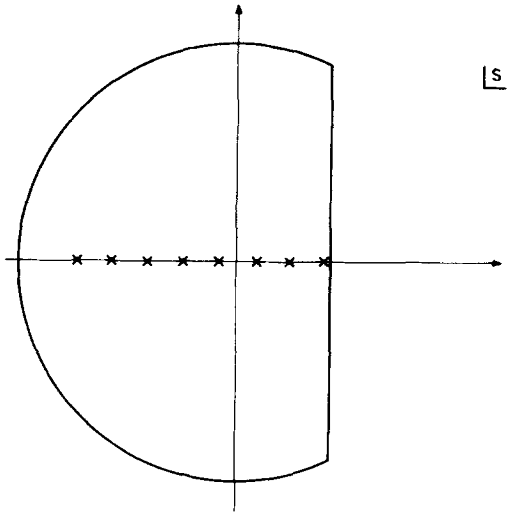


FIG. 1. The contour specified in integral (10) may be closed in the left half-plane since the integrand vanishes asymptotically on the arc at infinity. The residues of the poles (either single or double) at integer s must be evaluated.

$$\operatorname*{Res}_{s=1}[f(s)] = \frac{2\pi(-1)^l 2^{2l}\Gamma(l+1)}{y\Gamma(2l+1)}(1-r^2)^{l-1/2}, \quad (17)$$

$$\operatorname*{Res}_{s=2k}[f(s)] = \left(\frac{2}{y}\right)^{2k}\frac{2\Gamma(k)(-1)^{l-k}}{\Gamma(l-k+1)}\zeta(2k)\,{}_2F_1(k,k-l;\tfrac{1}{2};r^2), \quad (18)$$

where $k = 1,2,\ldots,l$. In obtaining (17) we have used a number of properties of the gamma function to simplify the expression.[11] Note in particular that the hypergeometric function in (18) is simply a polynomial in r^2 because $k-l$ is either zero or a negative integer.

The computation of the residues at the double poles is somewhat more complicated. First, at $s=0$ we find

$$\operatorname*{Res}_{s=0}[f(s)] = \frac{2(-1)^l}{l!}\Big\{\tfrac{1}{2}\gamma - \tfrac{1}{2}\psi(l+1) + \ln\left(\frac{y}{4\pi}\right) - \left[\frac{d}{ds}\,{}_2F_1\left(\frac{s}{2},\frac{s}{2}-l;\tfrac{1}{2};r^2\right)\right]_{s=0}\Big\}. \quad (19)$$

where γ is Euler's constant. The derivative of the hypergeometric function can be computed by using the series definition of ${}_2F_1$. At the end of the computation, the resulting series can be resummed and we find

$$\left[\frac{d}{ds}\,{}_2F_1\left(\frac{s}{2},\frac{s}{2}-l;\tfrac{1}{2};r^2\right)\right]_{s=0} = -lr^2\,{}_3F_2(1,1,1-l;\tfrac{3}{2},2;r^2). \quad (20)$$

Finally, consider the residues as $s=-2k$, $k=1,2,\ldots$. The calculation is easy because $\zeta(-2k)=0$. The result then is

$$\operatorname*{Res}_{s=-2k}[f(s)] = \frac{4(-1)^l}{k!(k+l)!}\left(\frac{y}{2}\right)^{2k}{}_2F_1(-k,-k-l;\tfrac{1}{2};r^2) \times\zeta'(-2k), \quad (21)$$

where the prime refers to differentiation and k is a positive integer. Using the reflection formula for the zeta function,[11] we may show that

$$\zeta'(-2k) = \tfrac{1}{2}(-1)^k(2\pi)^{-2k}\Gamma(2k+1)\zeta(2k+1). \quad (22)$$

Using (14), we add the results obtained in (17)–(22). The end result is the expression for $h^e_{2l+1}(y,r)$ given by (D1) of Appendix D.

At first glance, (D1) is quite a formidable looking expression [especially when compared with (9)]. However, we emphasize the advantages of our result. First, it is indeed the high-temperature (small y) expansion we were seeking; in the limit of small y, (D1) reduces to a very manageable result. Second, the analytic structure in y and r is easily analyzed. Note that all the hypergeometric functions which appear in (D1) are polynomials in r^2 (we assume l is a positive integer). Actually, one must recall that (D1) was derived under the assumption that $|r|<1$ (and $y \geqslant 0$[12]). However, one may extend the results to the remainder of the complex plane by analytic continuation [by noting that $h^e_{2l+1}(y,r)$ is analytic in the r plane with cuts running from $r=-\infty$ to $r=-1$ and $r=1$ to $r=\infty$.]

We now turn to the other functions; first consider $h^o_{2l+1}(y,r)$. The main complication arises in the calculation of the residues of the double poles. In particular, one needs to compute various derivatives of ${}_2F_1$'s with respect to their arguments. Such computations are far from trivial and require extensive manipulations of special functions. We give one such example:

$$\begin{aligned}\left[\frac{d}{ds}\,{}_2F_1\left(\frac{s+1}{2},\frac{s-2l+1}{2};\tfrac{3}{2};r^2\right)\right]_{s=1} &= -\gamma\,{}_2F_1(1,1-l;\tfrac{3}{2};r^2) \\ &\quad - \frac{\Gamma(\frac{3}{2})\Gamma(l)}{2\Gamma(l+\frac{3}{2})}(-1)^l r^{2l}\,{}_2F_1(1,1;l+\tfrac{3}{2};r^2) \\ &\quad + \tfrac{1}{2}\Gamma(\tfrac{3}{2})\Gamma(l)\sum_{k=0}^{l-1}\frac{(1)^k[S_k+S_{l-k-1}]r^{2k}}{\Gamma(k+\frac{3}{2})\Gamma(l-k)},\end{aligned} \quad (23)$$

where we have written

$$S_k \equiv \gamma + \psi(k+1) = \sum_{p=1}^{k}\frac{1}{p}. \quad (24)$$

The end result is that the expression for $h^o_{2l+1}(y,r)$ is the most complicated one of all and is given by (D2) in Appendix D.

We may now add the two expressions (D1) and (D2) to obtain $h_n(y,r)$ [see (12) and (13)]. It is of interest to check the analytic structure of h_n which should be analytic in the r plane cut from $r=1$ to $r=\infty$ (whereas h^e_n and h^o_n separately require an additional cut from $r=-\infty$ to $r=-1$). To demonstrate that the sum of (D1) and (D2) has no cut along the negative r axis, let us isolate the terms in $h_{2l+1}(y,r)$ which contribute to the cut structure. From (D1) and (D2) they are

$$h^e_{2l+1}(y,r) = \frac{\pi y^{2l-1}}{2\Gamma(2l+1)}(-1)^l(1-r^2)^{l-1/2} + \text{reg}, \quad (25)$$

$$h^o_{2l+1}(y,r) = \frac{y^{2l-1}}{\Gamma(2l+2)}r^{2l+1}\,{}_2F_1(1,1;l+\tfrac{3}{2};r^2) + \text{reg}, \quad (26)$$

where reg indicates pieces which are nonsingular for all real r. We may simplify (26) by using (B4) to extract the term with the singularity. Because

$$\arcsin(r) = \pm \frac{\pi}{2} \mp \arccos(\pm r) = \pm \frac{\pi}{2} \pm i \ln[\pm r + (r^2 - 1)^{1/2}], \tag{27}$$

we note that when $r > 0$ (excluding the point $r = \pm 1$) $\arccos r/(1 - r^2)^{1/2}$ is real valued, and when $r \leqslant 0$, $\arccos(-r)/(1 - r^2)^{1/2}$ is real valued. It follows that

$$_2F_1(1,1;l+\tfrac{3}{2};r^2) = \pm (1 - r^2)^{l - 1/2} \frac{\pi^{3/2}(l + \frac{1}{2})}{\Gamma(l + \frac{1}{2})} \left(\frac{d}{dr^2}\right)^l \frac{1}{r} + \text{reg}, \tag{28}$$

which after some manipulation becomes

$$_2F_1(1,1;l+\tfrac{3}{2};r^2) = \pm (1 - r^2)^{l - 1/2} \frac{\pi(-1)^l(l + \frac{1}{2})}{r^{2l+1}} + \text{reg}. \tag{29}$$

The sign ($\pm$) which appears in (28) and (29) is to be taken positive when $r > 0$ and negative when $r < 0$. Inserting (29) into (26), we see that $h_{2l+1}(y,r)$ is regular for $-\infty < r < 1$, as we originally claimed.

Lastly, we comment briefly on the case of $h_{2l}(y,r)$. It is readily apparent from (14) and (15) that for h^o_{2l} we need only calculate the residues of single poles. For h^e_{2l}, in addition to the single poles there is one double pole at $s = 1$. Thus, it is fairly simple to compute the expansions for h^e_{2l} and h^o_{2l} which we have written down in (D3) and (D4), respectively. We may repeat the arguments of the previous paragraph to show that $h_{2l}(y,r)$ has the correct analytic structure.

It is interesting to note that one can derive explicit expressions for $h_{2l}(y,r)$ and $g_{2l}(y,r)$ in terms of elementary functions and polylogarithms. In Section 5 we will derive the general expansions, which turn out to be useful in the low temperature expansion as well. For now, we will provide a simple example when $l = 1$. By making the substitution $\omega = \exp[-(x^2 + y^2)^{1/2}]$ in (1) and (2), we readily find

$$g_2(y,r) = \text{Li}_2[e^{(r-1)y}] - y \ln[1 - e^{(r-1)y}], \tag{30}$$

$$h_2(y,r) = -\ln[1 - e^{(r-1)y}]. \tag{31}$$

The high temperature ($y \to 0$) expansions may be easily worked out by using (C6).

5. THE LOW TEMPERATURE EXPANSION

We now consider (1) and (2) in the limit of $y \to \infty$ at fixed r, which corresponds to the low temperature limit. We may derive expansions for g_n and h_n directly by making the substitution $\omega = \exp[y - (x^2 + y^2)^{1/2}]$ in (1) and (2). The results are

$$g_n(y,r) = \frac{1}{\Gamma(n)} \times \int_0^1 d\omega \frac{(-\ln\omega)^{n/2-1}(2y - \ln\omega)^{n/2-1}(y - \ln\omega)}{\exp[(1-r)y] - \omega}, \tag{32}$$

$$h_n(y,r) = \frac{1}{\Gamma(n)} \int_0^1 d\omega \frac{(-\ln\omega)^{n/2-1}(2y - \ln\omega)^{n/2-1}}{\exp[(1-r)y] - \omega}. \tag{33}$$

Expanding the numerators under the assumption that $|\ln\omega/2y| < 1$, we may use (C1) to integrate term by term and obtain

$$g_n(y,r) = \frac{\Gamma(n/2)}{\Gamma(n)} \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(n/2-k)} \left(\frac{1}{2y}\right)^{k+1-n/2} \times \left\{ y\Gamma\left(\frac{n}{2}+k\right) \text{Li}_{k+n/2}(e^{(r-1)y}) + \Gamma\left(\frac{n}{2}+k+1\right) \text{Li}_{k+n/2+1}(e^{(r-1)y}) \right\}. \tag{34}$$

$$h_n(y,r) = \frac{\Gamma(n/2)}{\Gamma(n)} \sum_{k=0}^{\infty} \frac{\Gamma(n/2+k)}{\Gamma(n/2-k)k!} \left(\frac{1}{2y}\right)^{k+1-n/2} \times \text{Li}_{k+n/2}(e^{(r-1)y}). \tag{35}$$

If n is odd then (34) and (35) are asymptotic series in y as $y \to \infty$. In the case of n even, then the sums in (34) and (35) contain only a finite number of terms up to $k = n/2 - 1$. This verifies the remarks made at the end of the last section, and confirms (30) and (31). Therefore, by using (C6), one can actually use (34) and (35) as high temperature expansions, provided n is even. On the other hand, one can always use (34) and (35) as low temperature expansions by making use of the series expansion given in (C1).

We illustrate the use of (34) by computing the nonrelativistic limit of the charge density ρ of an ideal Bose gas. It was shown in Ref. 8 that ρ is equal to the difference between the number density of particles (n) and antiparticles ($\bar{n}$), where

$$n = \frac{T^3}{\pi^2} g_3(y,r), \tag{36}$$

$$\bar{n} = \frac{T^3}{\pi^2} g_3(y,-r), \tag{37}$$

with $y = m/T$ and $r = \mu/m$. Using (34), we see that the leading term in the low temperature expansion of g_3 is

$$g_3(y,r) \approx \tfrac{1}{4}(2\pi y^3)^{1/2} \text{Li}_{3/2}(e^{(r-1)y}). \tag{38}$$

As discussed in Ref. 8, the nonrelativistic chemical potential is $\mu_{NR} = \mu - m$. Let us define

$$z_{NR} = e^{(r-1)y} = e^{\mu_{NR}/T}. \tag{39}$$

Then

$$g_3(y,-r) \approx \tfrac{1}{4}(2\pi y^3)^{1/2} \text{Li}_{3/2}(z_{NR}^{-1} e^{-2y}) \approx \frac{1}{4z_{NR}}(2\pi y^3)^{1/2} e^{-2y}, \tag{40}$$

which is exponentially small as $y \to \infty$. That is, the contribution of the antiparticles is exponentially small in the nonrelativistic limit. Therefore, in this limit, $\rho = n$ and therefore by (36) and (38)

$$n = \left(\frac{mT}{2\pi}\right)^{3/2} \text{Li}_{3/2}(z_{NR}), \tag{41}$$

which is the standard result.[9] The first relativistic corrections to (41) may be easily calculated using (34).

6. SUMMARY

The solution to the complete high temperature ($y \to 0$) expansion of integrals (1) and (2) has been obtained. The results for (2) have been explicitly written down in Appendix D. For thermodynamic applications, this allows us to obtain the thermodynamic potential (3) (in an arbitrary number of dimensions) from which all thermodynamic quantities may be computed. In three spatial dimensions this requires evaluating $h_5^\epsilon(y,r)$ from (D1) and yields

$$\begin{aligned}\frac{\Omega}{VT^4} = \frac{-\pi^2}{45} + \frac{y^2}{12}(1-2r^2) - \frac{y^3}{6\pi}(1-r^2)^{3/2} \\ + \frac{y^4}{16\pi^2}\left[\ln\left(\frac{4\pi}{y}\right) - \gamma + \tfrac{3}{4} - 2r^2 + \tfrac{2}{3}r^4\right] \\ - \frac{2}{\pi^2}\left(\frac{y}{2}\right)^4 \sum_{k=1}^{\infty}(-1)^k\left(\frac{y}{4\pi}\right)^{2k}\frac{\Gamma(2k+1)\zeta(2k+1)}{\Gamma(k+1)\Gamma(k+3)} \\ \times {}_2F_1(-k,-k-2;\tfrac{1}{2};r^2),\end{aligned} \quad (42)$$

where $y = m/T$, $r = \mu/m$. The hypergeometric function is a polynomial in r^2 of order k. By using (42) all thermodynamic functions immediately follow from the relations

$$\begin{aligned} S &= -\left(\frac{\partial\Omega}{\partial T}\right)_{V,\mu}, \\ P &= -\left(\frac{\partial\Omega}{\partial V}\right)_{T,\mu}, \\ Q &= -\left(\frac{\partial\Omega}{\partial \mu}\right)_{T,V}, \\ U &= TS - PV + \mu Q. \end{aligned} \quad (43)$$

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation.

APPENDIX A: SOME IMPORTANT INTEGRALS

We list here two integrals used in the text.[11] In deriving (9), we need

$$\int_0^\infty dx \frac{x^{2n}e^{-p(x^2+y^2)^{1/2}}}{(x^2+y^2)^{1/2}} = \pi^{-1/2}\Gamma(n+\tfrac{1}{2})\left(\frac{2y}{p}\right)^n K_n(py), \quad (A1)$$

where Re $n > -\frac{1}{2}$, Re $y > 0$, and $p > 0$.

Second, in deriving (10) it is sufficient to know

$$\begin{aligned}\int_0^\infty dp\, p^{\mu-1} e^{ryp} K_N(yp) \\ = \frac{\pi^{1/2}(2y)^N}{[y(1-r)]^{N+\mu}} \frac{\Gamma(\mu+N)\Gamma(\mu-N)}{\Gamma(\mu+\frac{1}{2})} \\ \times {}_2F_1\left(\mu+N, N+\tfrac{1}{2};\ \mu+\tfrac{1}{2};\ \frac{r+1}{r-1}\right),\end{aligned} \quad (A2)$$

where Re $\mu > |\mathrm{Re}\, N|$ and Re $y(1-r) > 0$.

APPENDIX B: SOME PROPERTIES OF HYPERGEOMETRIC FUNCTIONS

We quote some very useful relations for Gauss' hypergeometric function ${}_2F_1$.[13,14] ${}_2F_1(a, b; c; z)/\Gamma(c)$ is an entire analytic function of its parameters a, b, c for fixed $|z| < 1$. An important relation used in deriving (11) is[15]

$$\begin{aligned}{}_2F_1\left(2a,\ 2b;\ a+b+\tfrac{1}{2};\ \frac{1-z}{2}\right) \\ = \frac{\Gamma(\frac{1}{2})\Gamma(a+b+\frac{1}{2})}{\Gamma(a+\frac{1}{2})\Gamma(b+\frac{1}{2})}\,{}_2F_1(a,b;\tfrac{1}{2};z^2) \\ + z\frac{\Gamma(-\frac{1}{2})\Gamma(a+b+\frac{1}{2})}{\Gamma(a)\Gamma(b)}\,{}_2F_1(a+\tfrac{1}{2},b+\tfrac{1}{2};\tfrac{3}{2};z^2).\end{aligned} \quad (B1)$$

where $|\arg(1\pm z)| < \pi$ and $a+b+\frac{1}{2} \neq 0, -1, -2, \ldots$.

In using the recursion relations (4)–(7) the following two results are particularly useful:

$$\frac{d}{dz}\,{}_2F_1(a,b;c;z) = \frac{ab}{c}\,{}_2F_1(a+1,b+1;c+1;z), \quad (B2)$$

$$\frac{d}{dz}\left[z^c\,{}_2F_1(a,b;c+1;z)\right] = cz^{c-1}\,{}_2F_1(a,b;c;z). \quad (B3)$$

We note that all hypergeometric functions which appear in Appendix D are simple polynomials with the exception of

$$\begin{aligned}{}_2F_1(1,1;l+\tfrac{3}{2};r^2) = (1-r^2)^{l-1/2}\frac{\pi^{1/2}(2l+1)}{\Gamma(l+\frac{1}{2})} \\ \times\left(\frac{d}{dr^2}\right)^l \frac{\arcsin r}{r},\end{aligned} \quad (B4)$$

$$\begin{aligned}{}_2F_1(1,\tfrac{3}{2}-l;\tfrac{3}{2};r^2) = (1-r^2)^{l-1}\frac{1}{2\Gamma(l)}\left(\frac{d}{dr^2}\right)^{l-1}\Big[r^{2l-3} \\ \times\ln\left(\frac{1+r}{1-r}\right)\Big].\end{aligned} \quad (B5)$$

Finally, we briefly mention some properties of ${}_3F_2(a_1, a_2, a_3; b_1, b_2; z)$. First, suppose $a_3 = b_2$; then ${}_3F_2(a_1,a_2,a_3;b_1,a_3;z) = {}_2F_1(a_1,a_2;b_1;z)$. Second, it will be useful to express ${}_3F_2$ as a sum over ${}_2F_1$'s when possible. This makes it easier to check recursion relations (4)–(7). We give one such example:

$${}_3F_2(1,1,1-l;\tfrac{3}{2},2;r^2) = \frac{1}{l}\sum_{j=0}^{l-1}{}_2F_1(-j,1;\tfrac{3}{2};r^2), \quad (B6)$$

which is useful in working with (D1). Finally, we note the following useful relations found in Ref. 14:

$$\begin{aligned}{}_3F_2(a,b,c;c+1,d+1;z) = \frac{c}{c-d}\,{}_2F_1(a,\ b;\ d+1;\ z) \\ - \frac{d}{c-d}\,{}_3F_2(a,\ b,\ c;\ c+1,\ d;\ z)\end{aligned} \quad (B7)$$

and

$$\begin{aligned}{}_3F_2(a,b,c;c+1,d+1;z) = \frac{c-1}{z(a-1)(b-1)} \\ \times\left[{}_2F_1(a-1,b-1;c-1;z) - 1\right].\end{aligned} \quad (B8)$$

APPENDIX C: THE POLYLOGARITHM

The polylogarithm $\mathrm{Li}_n(x)$ is defined (for $n > 0$) as[16,17]

$$\mathrm{Li}_n(x) = \frac{-1}{\Gamma(n)}\int_0^1 dt\,\frac{(-\ln t)^{n-1}}{t-x^{-1}} = \sum_{p=1}^{\infty}\frac{x^p}{p^n}. \quad (C1)$$

The following properties are useful:

$$\frac{d}{dx}\,\mathrm{Li}_n(x) = \frac{\mathrm{Li}_{n-1}(x)}{x}, \tag{C2}$$

$$\mathrm{Li}_1(x) = -\ln(1-x), \tag{C3}$$

$$\mathrm{Li}_n(1) = \zeta(n) \quad (n > 1). \tag{C4}$$

It is useful to deive a series expansion for $\mathrm{Li}_n(e^{-y})$ about $y = 0$. To do this, start with the well known identity

$$\frac{1}{\exp(y) - 1} = \frac{1}{y} - \tfrac{1}{2} + \frac{1}{\pi}\sum_{k=1}^{\infty}(-1)^{k+1} \times \zeta(2k)\left(\frac{y}{2\pi}\right)^{2k-1}. \tag{C5}$$

Successive integration of (C5) n times will yield $\mathrm{Li}_n(e^{-y})$:

$$(-1)^n \mathrm{Li}_n(e^{-y}) = \sum_{k=0}^{n-2} \frac{(-1)^{n-k}\zeta(n-k)}{k!} y^k + \frac{y^{n-1}}{(n-1)!}[\ln y - S_{n-1}] - \frac{y^n}{2n!} + 2y^{n-1}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}(2k-1)!\zeta(2k)}{(2k+n-1)!}\left(\frac{y}{2\pi}\right)^{2k}, \tag{C6}$$

where S_n is the sum of the first n reciprocals [see (24)]. Note that as it stands, (C6) is only valid when n is a non-negative integer. The proper analytic continuation of (C6) is derived in Ref. 18 and we quote it here:

$$\mathrm{Li}_\sigma(e^{-y}) = \Gamma(1-\sigma)\,y^{\sigma-2} + \sum_{k=0}^{\infty}\frac{(-1)^k\zeta(\sigma-k)}{k!}y^k. \tag{C7}$$

APPENDIX D: TABULATION OF RESULTS

We list here the complete high temperature expansion of $h_n(y, r)$ that results from the calculations described in Secs. 3 and 4, where y is assumed to be positive. [The functions $g_n(y, r)$ are immediately obtained by computing one r derivative as in (7).] There are four possiblities since the part of $h_n(y, r)$ that is even in r or odd in r is calculated separately [see (14) and (15)] and the index n may be either even or odd. Note that S_k is the sum of the first k reciprocals [see (24)].

$$\begin{aligned}
h^e_{2l+1}(y, r) = {} & \frac{\pi y^{2l-1}}{2\Gamma(2l+1)}(-1)^l(1-r^2)^{l-\frac{1}{2}} + \frac{(-1)^l}{2[\Gamma(l+1)]^2}\left(\frac{y}{2}\right)^{2l} \\
& \times\left\{\ln\left(\frac{y}{4\pi}\right) + \tfrac{1}{2}[\gamma - \psi(l+1)] + l r^2 {}_3F_2(1,1,1-l;\tfrac{3}{2},2;r^2)\right\} \\
& + \frac{1}{2\Gamma(l+1)}\sum_{k=0}^{l-1}(-1)^k\left(\frac{y}{2}\right)^{2k}\frac{\Gamma(l-k)\zeta(2l-2k)}{\Gamma(k+1)}{}_2F_1(-k,l-k;\tfrac{1}{2};r^2) \\
& + \frac{(-1)^l}{2\Gamma(l+1)}\left(\frac{y}{2}\right)^{2l}\sum_{k=1}^{\infty}(-1)^k\left(\frac{y}{4\pi}\right)^{2k}\frac{\Gamma(2k+1)\zeta(2k+1)}{\Gamma(k+1)\Gamma(k+l+1)}{}_2F_1(-k,-l-k;\tfrac{1}{2};r^2),
\end{aligned} \tag{D1}$$

$$\begin{aligned}
h^o_{2l+1}(y,r) = {} & \\
& \frac{r(-1)^{l+1}}{\Gamma(l+1)}\left(\frac{y}{2}\right)^{2l-1}\left\{\frac{-1}{\Gamma(l)}\ln\left(\frac{y}{2}\right){}_2F_1(1,1-l;\tfrac{3}{2};r^2)\right. \\
& + \sum_{k=1}^{l-1}(-1)^k\frac{\Gamma(k+1)\zeta(2k+1)}{\Gamma(l-k)}\left(\frac{y}{2}\right)^{-2k}{}_2F_1(k+1,k+1-l;\tfrac{3}{2};r^2) + \frac{(-1)^{l+1}\Gamma(\frac{3}{2})}{2\Gamma(l+\frac{3}{2})}r^{2l}\,{}_2F_1(1,1;l+\tfrac{3}{2};r^2) \\
& + \tfrac{1}{2}\Gamma(\tfrac{3}{2})\sum_{k=0}^{l-1}(-1)^k\frac{r^{2k}(S_k+S_{l-k-1})}{\Gamma(k+\frac{3}{2})\Gamma(l-k)} + 4\sum_{k=1}^{\infty}(-1)^k\frac{\Gamma(2k)\zeta(2k)}{\Gamma(k)\Gamma(l+k)}\left(\frac{y}{4\pi}\right)^{2k}\left[\psi(2k) + \tfrac{1}{2}\gamma + \ln\left(\frac{y}{4\pi}\right) + \frac{\zeta'(2k)}{\zeta(2k)}\right] \\
& \times {}_2F_1(1-k,1-k-l;\tfrac{3}{2};r^2) + \frac{2}{\Gamma(l)}\sum_{k=1}^{\infty}(-1)^k\frac{\Gamma(2k)\zeta(2k)}{\Gamma(2k+2)}\left(\frac{ry}{2\pi}\right)^{2k}{}_3F_2(1,1,1-l;k+1,k+\tfrac{3}{2};r^2) \\
& \left. - 2\Gamma(\tfrac{3}{2})\sum_{k=1}^{\infty}(-1)^k\Gamma(2k)\zeta(2k)\left(\frac{y}{4\pi}\right)^{2k}\sum_{j=0}^{k-1}\frac{r^{2j}(S_{k-j-1}+S_{l+k-j-1})}{\Gamma(j+1)\Gamma(j+\frac{3}{2})\Gamma(k-j)\Gamma(l+k-j)}\right\},
\end{aligned} \tag{D2}$$

$$\begin{aligned}
h^e_{2l}(y, r) = {} & \frac{y^{2l-2}(-1)^l(1-r^2)^{l-1}}{2\Gamma(2l)}\{\ln[y^2(1-r^2)] - \gamma - \psi(l) + 2r^2(1-l)\,{}_3F_2(1,1,2-l;2,\tfrac{3}{2};r^2)\} \\
& - \frac{1}{4}\left(\frac{y}{2}\right)^{2l-1}\frac{\Gamma(\frac{1}{2}-l)}{\Gamma(\frac{1}{2}+l)} + \frac{1}{2\Gamma(l+\frac{1}{2})}\sum_{k=0}^{l-2}(-1)^k\frac{\Gamma(l-k-\frac{1}{2})\zeta(2l-2k-1)}{\Gamma(k+1)}\left(\frac{y}{2}\right)^{2k}{}_2F_1(-k,l-k-\tfrac{1}{2};\tfrac{1}{2};r^2) \\
& + (-1)^{l+1}\frac{\Gamma(l)}{\Gamma(2l)}y^{2l-2}\sum_{k=1}^{\infty}(-1)^k\frac{\Gamma(k)\Gamma(2k)}{\Gamma(k+l)}\left(\frac{y}{2\pi}\right)^{2k}{}_2F_1(-k-l+1,\tfrac{1}{2}-k;\tfrac{1}{2};r^2),
\end{aligned} \tag{D3}$$

$$\begin{aligned}
h^o_{2l}(y, r) = {} & \frac{r}{\Gamma(l+\frac{1}{2})}\sum_{k=0}^{l-1}(-1)^k\frac{\Gamma(l-k-\frac{1}{2})\Gamma(2l-2k-2)}{\Gamma(k+1)}\left(\frac{y}{2}\right)^{2k+1}{}_2F_1(-k,l-k-\tfrac{1}{2};\tfrac{3}{2};r^2) \\
& - \frac{r}{\Gamma(l+\frac{1}{2})}\left(\frac{y}{2}\right)^{2l-2}\sum_{k=0}^{\infty}\frac{\Gamma(2k+1)\Gamma(\frac{3}{2}-l-k)\zeta(2k)}{\Gamma(k+1)}\left(\frac{y}{4\pi}\right)^{2k}{}_2F_1(1-k,\tfrac{3}{2}-l-k;\tfrac{3}{2};r^2)
\end{aligned} \tag{D4}$$

[1]F. Juttner, Z. Phys. **47**, 542 (1928).
[2]W. Glaser, Z. Phys. **94**, 677 (1935).
[3]Robert M. May, Phys. Rev. **135A**, 1515 (1964).
[4]P. T. Landsberg and J. Dunning-Davies, Phys. Rev. **A138**, 1049 (1965).
[5]M. M. Nieto, Lett. Nuovo Cimento **1**, 677 (1969); J. Math. Phys. **11**, 1346 (1970).
[6]R. Bechmann, F. Karsch, and D. E. Miller, Phys. Rev. Lett. **43**, 1277 (1979).
[7]Carlos Aragao de Carvalho and S. Goulart Rosa, Jr., J. Phys. A **13**, 3233 (1980).
[8]Howard E. Haber and H. Arthur Weldon, Phys. Rev. Lett.**46**, 1497 (1981); Phys. Rev. D **25**, 502 (1982).
[9]See, for example, K. Huang, *Statistical Mechanics* (Wiley, New York, 1963).
[10]See, for example, B. Davies, *Integral Transforms and Their Applications* (Springer, New York, 1978).
[11]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1980).
[12]The restriction $y \geqslant 0$ is made as a matter of convenience in the calculation; the behavior for $y < 0$ is easily obtained. For example, from (2) one deduces that $h_n(-y, r) = h_n(y, -r)$ from which it follows that $h_n^c(y, r)$ is an even function of y. Note that because of the term in (D1) proportional to y^{2l-1}, we conclude that $\partial h_n^c/\partial y$ is discontinuous at $y = 0$. As far as the physics is concerned, $y = m/T$ so that only positive y is physical. In addition, the zero mass limit clearly must be taken from the positive side.
[13]Arthur Erdelyi *et al.*, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I.
[14]Yudell L. Luke, *The Special Functions and Their Approximations* (Academic, New York, 1969), Vol. I.
[15]N. N. Lebedev, *Special Functions and Their Applications* (Dover, New York, 1972).
[16]L. Lewin, *Dilogarithms and Associated Functions* (MacDonald, London, 1958).
[17]K. S. Kolbig, J. A. Mignaco, and E. Remiddi, B. I. T. **10**, 38 (1970).
[18]J. E. Robinson, Phys. Rev. **83**, 678 (1951).

Rigorous iterated solutions to a nonlinear integral evolution problem in particle transport theory

V. C. Boffi and G. Spiga
Laboratorio di Ingegneria Nucleare dell'Università di Bologna, Bologna, Italy

(Received 11 December 1981; accepted for publication 28 May 1982)

After a preliminary functional study of the operator associated with the relevant Boltzmann equation, which is shown to be a contraction operator, a nonlinear integral evolution problem occurring in the diffusion of the particles of a mixture is solved by resorting to a rigorous iterative scheme, in the case without removal. According to this scheme, an explicit recursive representation for the general iterated solution of order n is developed. Structure and behavior of the solution so obtained are investigated and commented on.

PACS numbers: 05.60. + w, 51.10. + y, 02.30. − f

INTRODUCTION

We refer to the following physical situation. At the time $t = 0$, a spatially uniform, pulsed source, say $Q(\mathbf{v},t) = Q_0 S(\mathbf{v})\delta(t)$, injects Q_0 test particles (t.p.) (per unit volume) with the velocity distribution $S(\mathbf{v})$ and such that $\int_{\mathbb{R}_3} d\mathbf{v} S(\mathbf{v}) = 1$ in an unbounded host medium, consisting, in turn, of some other particles to be distinguished as field particles (f.p.). The host medium is taken to be free of t.p. up to $t = 0$ so that for the initial t.p. distribution function $f(\mathbf{v},t)$, right after the pulsed injection, we have $f(\mathbf{v},0) = Q_0 S(\mathbf{v})$. With respect to this initial datum we want to study the behavior of $f(\mathbf{v},t)$ for any $t > 0$ by accounting for the three following binary events, that are supposed to take place between the particles of the mixture under examination. Whereas t.p. are removed, say, through absorption, by f.p., we assume instead that the t.p. can interact with each other through either scattering or removal (by absorption), the latter event being introduced to generalize the mathematical effects due to removal rather than having a strict physical meaning. We have thus to face a nonlinear evolution problem for the distribution function of the t.p. considered.

The physical situation sketched above has been recently the object of several investigations aimed at focusing the mathematical problems connected with the existence, uniqueness, and structure of its solution. This has been essentially done on the basis of two main hypotheses concerning the cross sections and the scattering probability, respectively. More precisely, the $1/|\mathbf{v}|$ approximation for the cross section and the model of isotropic scattering between rigid spheres for the scattering probability have been systematically exploited. In this context, we recall the works of Krook and Wu[1,2] and Bobylev,[3] who first obtained independently the exact solution for the isotropic distribution function in the absence of any removal. Successively, and for the same physical situation, a series of papers by Ernst[4] and by Barnsley and Turchetti[5–9] have contributed significant progress to the knowledge and understanding of the problem.

In this paper, still referring to the $1/|\mathbf{v}|$ approximation for the cross section and to the case without removal, but leaving the scattering probability unspecified, we succeed not only in establishing some general results concerning the existence and the uniqueness of the solution of the problem, but also in defining an iterative constructive scheme leading to explicit iterated solutions of some "practical" interest. This is achieved by starting with the so-called "scattering kernel" formulation of the relevant nonlinear integro-differential Boltzmann equation, and then by reformulating it in the equivalent integral form.

In Sec. 1 the general theory of the problem is expounded. In particular, we prove—for a general scattering probability and for a general velocity distribution $S(\mathbf{v})$ of the t.p., emitted by the external source—the existence and the uniqueness of the solution to the nonlinear integral evolution problem to be dealt with on the basis of a simple application of the contracting mapping principle. The contraction properties of the operator associated with the problem under consideration are guaranteed only up to a finite "critical" time T, that is shown to be a function of the scattering collision frequency C_S between the t.p., the intensity Q_0 of the source, and the functional properties of the scattering probability. In Sec. 2 we present, instead, on the basis of the theory of the approximate methods for the solution of operator equations,[10] the results for the sequence of the iterated solutions $f_0(\mathbf{v},t), f_1(\mathbf{v},t), \cdots$ of the problem, and give the explicit recursive representation for the general $f_n(\mathbf{v},t)$. By choosing $f_0(\mathbf{v},t) = Q_0 M(\mathbf{v})$, where $M(\mathbf{v})$ is the Maxwellian normalized to unity (with the physical parameters determined by the initial distribution $S(\mathbf{v})$ and by the conservation laws of the scattering mechanism), we are also able to show that all the iterated solutions so obtained not only satisfy the initial condition at $t = 0$, but also as $t \to \infty$ tend to the correct limit just given by the Maxwellian $M(\mathbf{v})$. This circumstance can be interpreted as showing that the "critical" time T can actually be extended much farther than the value estimated here.

I. THEORY

A. Statement of the problem

The physical situation sketched in the Introduction is adequately described by the nonlinear integro-differential Boltzmann equation, that in the frame of the "scattering kernel" formulation and for $t \geqslant 0$ read as[11,12]

$$\frac{\partial f}{\partial t} + [\hat{C}_R N + Cn(t)] f(\mathbf{v},t) = I_s(\mathbf{v},t)$$

$$\equiv C_S \int_{\mathbb{R}_3}\int_{\mathbb{R}_3} d\mathbf{v}'\, d\mathbf{v}'' f(\mathbf{v}',t) f(\mathbf{v}'',t) \pi(\mathbf{v}',\mathbf{v}'' \to \mathbf{v}),$$

$$\mathbf{v} \in \mathbb{R}_3, \quad t \in [0,\infty), \quad (1a)$$

 0022-2488/82/101859-06$02.50

and is to be integrated upon the initial condition

$$f(\mathbf{v},0) = Q_0 S(\mathbf{v}). \qquad (1b)$$

In Eq. (1a)

$$\hat{\nu}_R = \hat{C}_R N, \qquad (2a)$$

where N is the assigned total density of the f.p., is the constant collision frequency related to the removal of the t.p. by the f.p. in the $1/|\mathbf{v}|$ approximation for the relevant cross section. The appropriate real positive constant $\hat{C}_R$ is indeed equal to $|\mathbf{v}|\hat{\sigma}_R(|\mathbf{v}|)$. Analogously,

$$\nu(t) = Cn(t) = (C_S + C_R)n(t), \qquad (2b)$$

where

$$n(t) = \int_{\mathbb{R}_3} d\mathbf{v} f(\mathbf{v},t), \qquad (3)$$

is the unknown total density of the t.p., is the total collision frequency, scattering plus removal, of the t.p. among themselves, still in the $1/|\mathbf{v}|$ approximation for the cross sections. In this case Eq. (2b) follows from

$$\nu_\alpha(\mathbf{v},t) = \int_{\mathbb{R}_3} d\mathbf{v}'' |\mathbf{v}-\mathbf{v}''| \sigma_\alpha(|\mathbf{v}-\mathbf{v}''|) f(\mathbf{v}'',t), \quad \alpha = S,R, \qquad (4a)$$

with

$$|\mathbf{v}-\mathbf{v}''|\sigma_\alpha(|\mathbf{v}-\mathbf{v}''|) = C_\alpha, \qquad (4b)$$

C_α being an appropriate real positive constant. [Compare also C_S in Eq. (1a)].

We recall also that the scattering probability $\pi(\mathbf{v}',\mathbf{v}''\to\mathbf{v})$ obeys, by definition, the normalization condition.

$$\int_{\mathbb{R}_3} d\mathbf{v}\,\pi(\mathbf{v}',\mathbf{v}''\to\mathbf{v}) = 1, \qquad (5)$$

whereas for the scattering-in integral we have, in general,

$$\int_{\mathbb{R}_3} d\mathbf{v}\, I_S(\mathbf{v},t) = \int_{\mathbb{R}_3} d\mathbf{v}'\, \nu_S(\mathbf{v}',t) f(\mathbf{v}',t). \qquad (6a)$$

In the present context, it is easily verified that

$$\int_{\mathbb{R}_3} d\mathbf{v} I_S(\mathbf{v},t) = C_S n^2(t). \qquad (6b)$$

B. The continuity equation

Equation (1) is made fully explicit once we know $n(t)$. An autonomous equation for $n(t)$ is, indeed, obtained by just integrating both sides of Eq. (1a) itself over the domain of the velocity $\mathbf{v}$. Assuming that the exchange of the order of integration (over $\mathbf{v}$) and differentiation (with respect to t) is permissible, we get in fact for $n(t)$ the following nonlinear ordinary first-order differential equation of Riccati's type

$$n'(t) = -\hat{C}_R N n(t) - C_R n^2(t), \qquad (7a)$$

which is the continuity equation holding for any π and S. The general solution to Eq. (7a) satisfying the initial condition

$$n(0) = Q_0, \qquad (7b)$$

is[11,12]

$$n(t) = Q_0 \hat{C}_R N \left[(\hat{C}_R N + Q_0 C_R) e^{\hat{C}_R N t} - Q_0 C_R \right]^{-1}. \qquad (8)$$

We observe that, for large t, $n(t)$ behaves exponentially like $\exp(-\hat{C}_R N t)$, as physically expected.

C. The integral formulation of the problem

Equation (1) can be now integrated along the trajectory of the general t.p. to yield

$$f(\mathbf{v},t) = Q_0 S(\mathbf{v}) T_0(t) + C_S \int_{\mathbb{R}_3}\int_{\mathbb{R}_3}\int_0^t d\mathbf{v}'\, d\mathbf{v}''\, du \times T(t,u)\pi(\mathbf{v}',\mathbf{v}''\to\mathbf{v}) f(\mathbf{v}',u) f(\mathbf{v}'',u), \qquad (9)$$

where we set

$$T(t,u) = \exp\left[-\hat{C}_R N t - C \textstyle\int_u^t n(u')du' \right], \qquad (10a)$$

with

$$T_0(t) = T(t,0). \qquad (10b)$$

Equation (9)—which is of Volterra's type with respect to time—is a nonlinear integral equation for $f(\mathbf{v},t)$, and describes the evolution problem following the application of the pulsed source $Q_0 S(\mathbf{v})\delta(t)$. For any fixed π, the kernel of Eq. (9) becomes fully explicit once we introduce in its time-dependent factor $T(t,u)$ the expression of $n(t)$, Eq. (8). There results that

$$T(t,u) = \theta(t)\theta^{-1}(u), \qquad (11a)$$

with

$$\theta(t) = e^{-\hat{C}_R N t} \left[(\hat{C}_R N + Q_0 C_R) - Q_0 C_R e^{-\hat{C}_R N t} \right]^{-C/C_R}. \qquad (11b)$$

D. The case without removal

We shall consider hereafter Eq. (9) in the limiting case when both $\hat{C}_R$ and C_R vanish, that is when no removal is present. In this case, Eq. (8) for $n(t)$ reduces to

$$n(t) = Q_0, \qquad (12)$$

which is the simple conservation principle, holding now, for the t.p., whereas Eqs. (11a) and (11b) give

$$T(t,u) = \theta(t-u) = e^{-C_S Q_0 (t-u)}, \qquad (13)$$

that is, $T(t,u)$ not only is separable, but it is also of displacement type.

Equation (9) can be then rewritten in an operational form as

$$f = Af, \qquad (14)$$

where A is the nonlinear inhomogeneous operator defined by

$$Af \equiv Q_0 S(\mathbf{v}) e^{-C_S Q_0 t} + C_S \int_{\mathbb{R}_3}\int_{\mathbb{R}_3}\int_0^t d\mathbf{v}'\, d\mathbf{v}''\, du\, e^{-C_S Q_0 (t-u)} \times \pi(\mathbf{v}',\mathbf{v}''\to\mathbf{v}) f(\mathbf{v}',u) f(\mathbf{v}'',u). \qquad (15)$$

We shall briefly study next Eq. (14) from the point of view of functional analysis.

E. The operator A as a contraction operator

The proper Banach space to work with is the space E of the functions $\varphi(\mathbf{v},t)$, defined on $\mathbb{R}_3 \otimes [0,T]$, which are continuous in t for almost every $\mathbf{v}\in\mathbb{R}_3$, and summable in $\mathbf{v}$ for any $t\in[0,T]$. If

$$\|\varphi\|_t = \int_{\mathbf{R}_3} d\mathbf{v}|\varphi(\mathbf{v},t)|, \tag{16a}$$

denotes the L_1 norm with respect to $\mathbf{v}$, the norm in E is defined as

$$|||\varphi||| = \max_{t\in[0,T]} \|\varphi\|_t. \tag{16b}$$

We know that π is a nonnegative function in $\mathbb{R}_3\otimes\mathbb{R}_3\otimes\mathbb{R}_3$, summable with respect to $\mathbf{v}$. Let B denote the closed ball of E centered at the origin with radius Q_0 (in other words $|||\varphi|||<Q_0$ if $\varphi\in B$). It is easy to check that A maps B into itself, namely, $AB\subset B$. As we may exchange the relevant integration orders, we have in fact for $\varphi\in B$

$$\|A\varphi\|_t \leqslant Q_0 e^{-C_S Q_0 t} + C_S\int_0^t du\, e^{-C_S Q_0(t-u)}\|\varphi\|_u^2$$
$$\leqslant \frac{|||\varphi|||^2}{Q_0} + \left(Q_0 - \frac{|||\varphi|||^2}{Q_0}\right)e^{-C_S Q_0 t}, \tag{17}$$

and consequently

$$|||A\varphi||| \leqslant Q_0, \tag{18}$$

as required

We shall show now that it is possible to choose $T>0$ in such a way that A is a contraction operator on B. By accounting for the symmetry of π with respect to the velocities before collision, namely

$$\pi(\mathbf{v}',\mathbf{v}''\to\mathbf{v}) = \pi(\mathbf{v}'',\mathbf{v}'\to\mathbf{v}), \tag{19}$$

we verify successively that

$$A\varphi - A\psi$$
$$= C_S\int_0^t du\, e^{-C_S Q_0(t-u)}\int_{\mathbf{R}_3}\int_{\mathbf{R}_3} d\mathbf{v}'\, d\mathbf{v}''\pi(\mathbf{v}',\mathbf{v}''\to\mathbf{v})$$
$$\times[\varphi(\mathbf{v}',u)+\psi(\mathbf{v}',u)][\varphi(\mathbf{v}'',u)-\psi(\mathbf{v}'',u)], \tag{20a}$$

$$\|A\varphi - A\psi\|_t \leqslant C_S\int_0^t du\, e^{-C_S Q_0(t-u)}\|\varphi+\psi\|_u\|\varphi-\psi\|_u, \tag{20b}$$

$$|||A\varphi - A\psi||| \leqslant 2(1-e^{-C_S Q_0 t})|||\varphi-\psi|||. \tag{20c}$$

The opertor A satisfies thus a Lipschitz condition, and a sufficient condition for it to be a contraction is

$$T < \ln 2/C_S Q_0. \tag{21}$$

Another estimate for T can be obtained if we assume that the linear integral operator generated by the kernel $k(\mathbf{v}',\mathbf{v}') = \pi(\mathbf{v}',\mathbf{v}''\to\mathbf{v})$ (depending on the parameter $\mathbf{v}$) is a continuous mapping of $L_1(\mathbb{R}_3)$ into $L_\infty(\mathbb{R}_3)$. Let $N(\mathbf{v})$ denote its norm. If we make the further assumption that $N\in L_1(\mathbb{R}_3)$ [a sufficient condition for both the previous assumptions to be true is that there exist $\chi\in L_1(\mathbb{R}_3)$ such that $\pi(\mathbf{v}',\mathbf{v}''\to\mathbf{v})\leqslant$ constant $\chi(\mathbf{v})$], we may write

$$\left|\int_{\mathbf{R}_3} d\mathbf{v}''\ \pi(\mathbf{v}',\mathbf{v}''\to\mathbf{v})[\varphi(\mathbf{v}'',u)-\psi(\mathbf{v}'',u)]\right| \leqslant N(\mathbf{v})\|\varphi-\psi\|_u, \tag{22a}$$

and

$$|A\varphi - A\psi| \leqslant C_S N(\mathbf{v})\int_0^t du\, e^{-C_S Q_0(t-u)}\|\varphi-\psi\|_u\|\varphi+\psi\|_u, \tag{22b}$$

from which there follow that

$$|||A\varphi - A\psi||| \leqslant 2\|N\|(1-e^{-C_S Q_0 t})|||\varphi-\psi|||, \tag{23a}$$

and then the new estimate

$$T < \frac{1}{C_S Q_0}\ln\frac{2\|N\|}{2\|N\|-1}. \tag{23b}$$

We realize thus that Eq. (23b) actually enlarges the previous condition for T, Eq. (21), when $\|N\|\leqslant 1$. A sufficient condition for the latter inequality to occur is that π is bounded, and there exists at least one pair $\mathbf{v}'_0$, $\mathbf{v}''_0$ for which

$$\sup_{(\mathbf{v}',\mathbf{v}'')\in\mathbf{R}_3\otimes\mathbf{R}_3} \pi(\mathbf{v}',\mathbf{v}''\to\mathbf{v})$$
$$= \max_{(\mathbf{v}',\mathbf{v}'')\in\mathbf{R}_3\otimes\mathbf{R}_3} \pi(\mathbf{v}',\mathbf{v}''\to\mathbf{v}) = \pi(\mathbf{v}_0',\mathbf{v}_0''\to\mathbf{v}), \tag{24a}$$

so that

$$\|N\| \leqslant \int_{\mathbf{R}_3} d\mathbf{v}\ \sup_{\mathbf{v}',\mathbf{v}''}\pi(\mathbf{v}',\mathbf{v}''\to\mathbf{v}) = \int_{\mathbf{R}_3} d\mathbf{v}\ \pi(\mathbf{v}_0',\mathbf{v}_0''\to\mathbf{v}) = 1. \tag{24b}$$

We notice that no restriction on T would appear in the case $\|N\|\leqslant\frac{1}{2}$. Now, if one of the above restrictions for T is in order, A satisfies all the requirements of the contracting mapping principle on B, and therefore there exists in B a unique solution to Eq. (14). This is in agreement with the physical expectation that the solution $f(\mathbf{v},t)$ is a nonnegative function with $\int_{\mathbf{R}_3} d\mathbf{v} f(\mathbf{v},t) = Q_0$.

II. ITERATED SOLUTIONS

A. Statement of the problem

In studying the successive approximations scheme for the actual solution to Eq. (14) we recall that,[10] in the hypothesis of the previous section, if $f_0(\mathbf{v},t)$ is arbitrarily chosen in B, then the sequence $\{f_n\}$ of the iterated solution

$$f_n = Af_{n-1} \quad (n=1,2,\ldots), \tag{25a}$$

converges in the norm of E to the unique solution of Eq. (14) belonging to E, and very simple *a priori* estimates for the approximate solutions can be given, namely [compare Eq. (21)]

$$|||f - f_n||| \leqslant \frac{2^n(1-e^{-C_S Q_0 T})^n}{2e^{-C_S Q_0 T}-1}|||f_0 - Af_0|||. \tag{25b}$$

The main shortcoming is that the time variable is restricted to a finite interval $[0,T]$, where T cannot exceed a critical value [see Eqs. (21) and (23b)]. We shall construct in the sequel a practical iterated solution by means of a suitable choice of the starting point f_0, for which the proper behavior for $t\to\infty$ is reproduced to any approximation order.

We refer to the physical case in which the stationary version of Eq. (14) has a solution $Q_0 M(\mathbf{v})$, with $\int_{\mathbf{R}_3} d\mathbf{v}\, M(\mathbf{v}) = 1$, and $M(\mathbf{v})$ is a solution to the nonlinear integral equation

$$M(\mathbf{v}) = \int_{\mathbf{R}_3}\int_{\mathbf{R}_3} d\mathbf{v}'\, d\mathbf{v}''\ \pi(\mathbf{v}',\mathbf{v}''\to\mathbf{v})M(\mathbf{v}')M(\mathbf{v}''), \tag{26}$$

expressing equilibrium of the collision term. [In principle, $M(\mathbf{v})$ might be even different from a Maxwellian distribution.] We expect then that

$$\lim_{t\to\infty} f(\mathbf{v},t) = Q_0 M(\mathbf{v}), \tag{27}$$

as it necessarily occurs when an H-theorem exists.

Let us now try to choose $f_0 \in B$ in such a way that f_1 satisfies such a requirement. Confining ourselves, for the sake of simplicity, to a stationary f_0, we take $f_0 = Q_0 \psi(\mathbf{v})$, with $\|\psi\| \leqslant 1$, and get

$$f_1(\mathbf{v},t) = Q_0 S(\mathbf{v}) e^{-C_S Q_0 t} + Q_0(1 - e^{-C_S Q_0 t}) \times \int_{\mathbf{R}_3} \int_{\mathbf{R}_3} d\mathbf{v}' d\mathbf{v}'' \pi(\mathbf{v}',\mathbf{v}'' \to \mathbf{v}) \psi(\mathbf{v}') \psi(\mathbf{v}''). \tag{28}$$

The limiting condition, Eq. (27), is thus fulfilled if and only if

$$\int_{\mathbf{R}_3} \int_{\mathbf{R}_3} d\mathbf{v}'\, d\mathbf{v}'' \pi(\mathbf{v}',\mathbf{v}'' \to \mathbf{v}) \psi(\mathbf{v}') \psi(\mathbf{v}'') = M(\mathbf{v}). \tag{29a}$$

Combining this with the integral equation for M yields

$$\int_{\mathbf{R}_3} \int_{\mathbf{R}_3} d\mathbf{v}'\, d\mathbf{v}'' \pi(\mathbf{v}',\mathbf{v}'' \to \mathbf{v}) [M(\mathbf{v}') + \psi(\mathbf{v}')] [M(\mathbf{v}'') - \psi(\mathbf{v}'')] = 0, \tag{29b}$$

which, upon integration with respect to $\mathbf{v}$, gives $\|M + \psi\| \, \|M - \psi\| = 0$, namely $\psi(\mathbf{v}) = \pm M(\mathbf{v})$. We are led thus to the unique possible choice $f_0(\mathbf{v},t) = Q_0 M(\mathbf{v})$. The behavior of the higher iterated solution remains, however, to be investigated. In any case, the following decomposition is proposed:

$$f_n(\mathbf{v},t) = Q_0 M(\mathbf{v}) + g_{n-1}(\mathbf{v},t), \tag{30}$$

and the iteration scheme for the g_n's is

$$\begin{aligned} g_n(\mathbf{v},t) = {} & Q_0[S(\mathbf{v}) - M(\mathbf{v})] e^{-C_S Q_0 t} \\ & + 2 C_S Q_0 \int_0^t du\, e^{-C_S Q_0 (t-u)} \int_{\mathbf{R}_3} \int_{\mathbf{R}_3} d\mathbf{v}'\, d\mathbf{v}'' \times \pi(\mathbf{v}',\mathbf{v}'' \to \mathbf{v}) M(\mathbf{v}') g_{n-1}(\mathbf{v}'',u) \\ & + C_S \int_0^t du\, e^{-C_S Q_0 (t-u)} \int_{\mathbf{R}_3} \int_{\mathbf{R}_3} d\mathbf{v}'\, d\mathbf{v}'' \times \pi(\mathbf{v}',\mathbf{v}'' \to \mathbf{v}) g_{n-1}(\mathbf{v}',u) g_{n-1}(\mathbf{v}'',u) \end{aligned} \tag{31a}$$

starting from

$$g_0(\mathbf{v},t) = Q_0[S(\mathbf{v}) - M(\mathbf{v})] e^{-C_S Q_0 t}. \tag{31b}$$

We now prove by induction that

$$g_n(\mathbf{v},t) = Q_0 \sum_{l=1}^{2^n} e^{-l C_S Q_0 t} \sum_{j=0}^{d(n,l)} F_{l,j}^{(n)}(\mathbf{v}) (C_S Q_0 t)^j, \tag{32a}$$

where

$$d(n,l) = (n - 1 - r) l + 2^r \quad (l = 1,2,\ldots,2^n), \tag{32b}$$

$$r = 0 \text{ for } l = 1; \; r = 1 + [lg_2(l-1)] \text{ for } l = 2,\ldots,2^n. \tag{32c}$$

It is easily verified that $d(n,1) = n$, and that the maximum value of $d(n,l)$ is 2^{n-1} for $2^{n-2} \leqslant l \leqslant 2^{n-1}$.

B. Proof of Eq. (32)

Equation (32) is trivially true for $n = 0$ with $F_{1,0}^{(0)}(\mathbf{v}) = S(\mathbf{v}) - M(\mathbf{v})$. Let us assume that it is true for the index $n - 1$, and evaluate $g_n(\mathbf{v},t)$ by means by Eq. (31). We get

$$\begin{aligned} g_n(\mathbf{v},t) = {} & Q_0[S(\mathbf{v}) - M(\mathbf{v})] e^{-C_S Q_0 t} \\ & + 2Q_0 \sum_{l=1}^{2^n-1} \sum_{j=0}^{d(n-1,l)} V(F_{l,j}^{(n-1)}) e^{-C_S Q_0 t} C_S Q_0 \int_0^t du (C_S Q_0 u)^j e^{-(l-1) C_S Q_0 u} \\ & + Q_0 \sum_{l=1}^{2^n-1} \sum_{m=1}^{2^n-1} \sum_{j=0}^{d(n-1,l)} \sum_{i=0}^{d(n-1,m)} W(F_{l,j}^{(n-1)}, F_{m,i}^{(n-1)}) C_S Q_0 \int_0^t du (C_S Q_0 u)^{i+j} e^{-(l+m-1) C_S Q_0 u - C_S Q_0 t}, \end{aligned} \tag{33a}$$

where we use the position

$$V(\varphi) = \int_{\mathbf{R}_3} \int_{\mathbf{R}_3} d\mathbf{v}'\, d\mathbf{v}''\; \pi(\mathbf{v}',\mathbf{v}'' \to \mathbf{v}) M(\mathbf{v}'') \varphi(\mathbf{v}') = W(M,\varphi), \tag{33b}$$

$$W(\varphi,\psi) = W(\psi,\varphi) = \int_{\mathbf{R}_3} \int_{\mathbf{R}_3} d\mathbf{v}'\, d\mathbf{v}''\; \pi(\mathbf{v}',\mathbf{v}'' \to \mathbf{v}) \varphi(\mathbf{v}') \psi(\mathbf{v}''). \tag{33c}$$

After evaluating the integrals with respect to u, we end up with

$$\begin{aligned} g_n(\mathbf{v},t) = {} & Q_0 e^{-C_S Q_0 t} \sum_{j=0}^{d(n,1)} F_{1,j}^{(n)}(\mathbf{v}) (C_S Q_0 t)^j \\ & - 2Q_0 \sum_{l=2}^{2^n-1} e^{-l C_S Q_0 t} \sum_{j=0}^{d(n-1,l)} \frac{(l-1)^j}{j!} (C_S Q_0 t)^j \sum_{k=j}^{d(n-1,l)} \frac{k!}{(l-1)^{k+1}} V(F_{l,k}^{(n-1)}) \\ & - Q_0 \sum_{p=2}^{2^n-1} e^{-p C_S Q_0 t} \sum_{l=1}^{p-1} \sum_{j=0}^{d(n-1,l)} \sum_{i=0}^{d(n-1,p-l)} \frac{(i+j)!}{(p-1)^{i+j+1}} W(F_{l,j}^{(n-1)}, F_{p-l,i}^{(n-1)}) \sum_{k=0}^{i+j} \frac{(p-1)^k}{k!} (C_S Q_0 t)^k \\ & - Q_0 \sum_{p=2^{n-1}+1}^{2^n} e^{-p C_S Q_0 t} \sum_{l=p-2^{n-1}}^{2^n-1} \sum_{j=0}^{d(n-1,l)} \sum_{i=0}^{d(n-1,p-l)} \frac{(i+j)!}{(p-1)^{i+j+1}} W(F_{l,j}^{(n-1)}, F_{p-l,i}^{(n-1)}) \sum_{k=0}^{i+j} \frac{(p-1)^k}{k!} (C_S Q_0 t)^k, \end{aligned} \tag{34a}$$

where

$$F_{1,0}^{(n)}(\mathbf{v}) = S(\mathbf{v}) - M(\mathbf{v}) + 2\sum_{l=2}^{2^{n-1}}\sum_{j=0}^{d(n-1,l)}\frac{j!}{(l-1)^{j+1}}V(F_{l,j}^{(n-1)})$$

$$+\sum_{l=1}^{2^{n-1}}\sum_{m=1}^{2^{n-1}}\sum_{j=0}^{d(n-1,l)}\sum_{i=0}^{d(n-1,m)}\frac{(i+j)!}{(l+m-1)^{i+j+1}}W(F_{l,j}^{(n-1)},F_{m,i}^{(n-1)}), \tag{34b}$$

$$F_{1,j}^{(n)} = \frac{2}{j}V(F_{1,j-1}^{(n-1)}), \qquad j = 1,2\ldots,d(n,1) = n. \tag{34c}$$

Rearranging the summation orders yields then

$$g_n(\mathbf{v},t) = Q_0 e^{-C_S Q_0 t}\sum_{j=0}^{d(n,1)} F_{1,j}^{(n)}(\mathbf{v})(C_S Q_0 t)^j$$

$$-Q_0\sum_{p=2}^{2^{n-1}} e^{-pC_S Q_0 t}\Bigg[2\sum_{j=0}^{d(n-1,p)}(C_S Q_0 t)^j\frac{(p-1)^j}{j!}\sum_{k=j}^{d(n-1,p)}\frac{k!}{(p-1)^{k+1}}V(F_{p,k}^{(n-1)})$$

$$+\sum_{l=1}^{p-1}\sum_{k=0}^{d(n-1,l)+d(n-1,p-l)}(C_S Q_0 t)^k\frac{(p-1)^k}{k!}\sum_{j=\max[0,k-d(n-1,p-l)]}^{d(n-1,l)}$$

$$\times\sum_{i=\max(0,k-j)}^{d(n-1,p-l)}\frac{(i+j)!}{(p-1)^{i+j+1}}W(F_{l,j}^{(n-1)},F_{p-l,i}^{(n-1)})\Bigg]$$

$$-Q_0\sum_{p=2^{n-1}+1}^{2^n} e^{-pC_S Q_0 t}\sum_{l=p-2^{n-1}}^{2^{n-1}}\sum_{k=0}^{d(n-1,l)+d(n-1,p-l)}(C_S Q_0 t)^k\frac{(p-1)^k}{k!}$$

$$\times\sum_{j=\max[0,k-d(n-1,p-l)]}^{d(n-1,l)}\sum_{i=\max(0,k-j)}^{d(n-1,p-l)}\frac{(i+j)!}{(p-1)^{i+j+1}}W(F_{l,j}^{(n-1)},F_{p-l,i}^{(n-1)}). \tag{35}$$

Now, in order to check whether or not Eq. (32) is recovered, we must interchange the summation order between l and k. For this purpose we have to study $d(n-1,l) + d(n-1, p-l)$ versus l. It can be verified that the trend is always symmetric with respect to the midpoint $l = p/2$, and is monotonic in each of the half ranges with a maximum at $l = [p/2]$. To evaluate this maximum we have to prove the following result:

$$d(n-1,l) + d(n-1,m) \leqslant d(n,l+m) \qquad (l \leqslant m \leqslant 2^{n-1}), \tag{36}$$

the equality sign holding if and only if $m = 1$ or $l \geqslant 2^{[lg_2(m-1)]}$. The case $m = 1$ is trivial. When $m > 1$ and $l > 2^{[lg_2(m-1)]}$, then, putting $s - 1 = [lg_2(m-1)] = [lg_2(l-1)]$ we have $[lg_2(l+m-1)] = s$, and thus

$$d(n-1,l) + d(n-1,m) = [n-1-(s+1)](l+m) + 2^{s+1} = d(n,l+m). \tag{37a}$$

When $m > 2$ and $l = 2^{[lg_2(m-1)]}$, we have again $[lg_2(l+m-1)] = s$ and

$$d(n,l+m) - d(n-1,l) - d(n-1,m) = 2^{s-1} - l = 0. \tag{37b}$$

When $m = 2$ and $l = 1$, we have immediately

$$d(n-1,1) + d(n-1,2) = d(n,3). \tag{37c}$$

There remains thus to check only that the sign $<$ holds in all cases excluded so far. Setting now $r - 1 = [lg_2(l-1)]$, it is easily realized that there are only two alternatives:

(i) $[lg_2(l+m-1)] = s-1$, (ii) $[lg_2(l+m-1)] = s$.

In the case (i) we may write

$$d(n,l+m) - d(n-1,l) - d(n-1,m) = m - (s-1-r)l - 2^r = m - d(s,l) \geqslant m - 2^{s-1} > 0, \tag{38a}$$

whereas in the case (ii) we get

$$d(n,l+m) - d(n-1,l) - d(n-1,m) = 2^s - (s-r)l - 2^r = 2^s - d(s+1,l) > 0, \tag{38b}$$

as follows now since $l < 2^{s-1}$ and, consequently, $d(s+1,l) < 2^s$. This completes the proof of Eq. (36).

Let us now go back to $d(n-1,l) + d(n-1,p-l)$, whose maximum is reached at $l = [p/2]$. Since $[p/2]$ and $p - [p/2]$ either coincide or are adjacent, the previous lemma applies, and we have

$$d(n-1,l) + d(n-1,p-l) \leqslant d(n-1,[p/2]) + d(n-1,p-[p/2]) = d(n,p). \tag{39}$$

If we denote by $h = h(n-1,p,k_0) \leqslant [p/2]$ [with $p = 2,3,\ldots 2^n$ and $k_0 \leqslant d(n,p)$] the smallest intersection of the straight line $k = k_0$ with the stepwise function $k = d(n-1,l) + d(n-1,p-l)$, when n and p are fixed, and l is running along its domain depending on the chosen values of n and p [compare Eq. (35)], the final inversion of the relevant summation orders can be performed, to recover just Eq. (32a) with

$$F^{n}_{l,j}(\mathbf{v}) = -2U(\tfrac{1}{2} + 2^{n-1} - l)\frac{(l-1)^j}{j!} \times \sum_{k=j}^{d(n-1,l)} \frac{k!}{(l-1)^{k+1}} V(F^{(n-1)}_{l,k}) - \frac{(l-1)^j}{j!} \sum_{p=h(n-1,l,j)}^{l-h(n-1,l,j)} \sum_{k=\max[0,j-d(n-1,l-p)]}^{d(n-1,p)} \sum_{i=\max(0,k-j)}^{d(n-1,l-p)} \frac{(i+k)!}{(l-1)^{i+k+1}} W(F^{(n-1)}_{p,k}, F^{(n-1)}_{l-p,i}),$$

$$l = 2,3,\ldots 2^n, \qquad j = 0,1,\ldots,d(n,l), \tag{40}$$

U denoting the unit step function.

Equation (40) allows us then to construct one by one all the coefficients $F_{l,j}^{(n)}(\mathbf{v})$. The index $h(n-1,l,j)$ can be determined easily by using the following recipe, which is equivalent to its definition.

If the equation

$$(n-1) - (n-3-[lg_2(l-2)])(l-1) - 2\cdot 2^{[lg_2(l-2)]} = j \tag{41a}$$

is satisfied, then we have $h(n-1,l,j) = 1$. Otherwise, one puts successively $h = 2,3,\ldots,[l/2]$ in the equation

$$(n-3-[lg_2(h-1)])h - (n-3-[lg_2(l-h-1)])(l-h) + 2(2^{[lg_2(h-1)]} - 2^{[lg_2(l-h-1)]}) = j \tag{41b}$$

until it is satisfied; such a value is $h(n-1,l,j)$. There always results

$$\begin{aligned} &h(n-1,2,j) = 1, \quad && n\geqslant 1, 0\leqslant j\leqslant 2n-2,\\ &h(n-1,2^n,0) = 2^{n-1}, \quad && n\geqslant 1. \end{aligned} \tag{41c}$$

C. Conclusions

Inserting Eq. (32a) in Eq. (30), we have thus that

$$f_n(\mathbf{v},t) = Q_0 M(\mathbf{v}) + Q_0 \sum_{l=1}^{2^n-1} F^{(n-1)}_{l,0}(\mathbf{v}) e^{-lC_S Q_0 t} + Q_0 \sum_{l=1}^{2^n-1} e^{-lC_S Q_0 t} \sum_{j=1}^{d(n-1,l)} F^{(n-1)}_{l,j}(\mathbf{v})(C_S Q_0 t)^j,$$

$$n = 0,1,2,\ldots \tag{42}$$

is the nth iterated solution to the nonlinear integral evolution problem described by Eq. (14).

About this general approximate solution of order n we may comment as follows:

(i) according to Eq. (25b), its convergence in the norm of E is guaranteed even if only for a finite "critical" value T of the time, as estimated in Eq. (21) or Eq. (23b);

(ii) as $t \to 0$, Eq. (42) tends to the correct limit $Q_0 S(\mathbf{v})$, Eq. (1b). It is, in fact, readily verified that

$$\sum_{l=1}^{2^n-1} F^{(n-1)}_{l,0}(\mathbf{v}) = S(\mathbf{v}) - M(\mathbf{v}); \tag{43}$$

(iii) as $t \to \infty$, Eq. (42) tends also to the correct limit $Q_0 M(\mathbf{v})$, Eq. (27). This circumstance amounts physically to an increase of the estimated critical T.

In this respect, we have not attempted, indeed, to extend the value of T by considering $f(\mathbf{v},t_0)$ with $t_0 \leqslant T$ as the initial datum of a new evolution problem; it seems, however, very likely that this task may actually be accomplished, and the value of T may be extended up to infinity, following the line proposed in Ref. 9.

To conclude, let us examine briefly two particular cases for which an exact analytical solution to Eq. (14) is easily obtained. The first case is characterized by setting $\pi(\mathbf{v}',\mathbf{v}'' \to \mathbf{v}) = M(\mathbf{v})$. In this case $W(\varphi, S-M) = 0$, and $V(S-M) = W(M,S-M) = 0$ so that for any n all the $F^{(n)}_{l,j}(\mathbf{v})$'s for $l \geqslant 1, j \geqslant 0$ vanish except the $F^{(n)}_{1,0}(\mathbf{v})$'s that are equal to $S(\mathbf{v}) - M(\mathbf{v})$. Consequently, for any n we get

$$f_n(\mathbf{v},t) = Q_0\{M(\mathbf{v}) + [S(\mathbf{v}) - M(\mathbf{v})]e^{-C_S Q_0 t}\}, \tag{44a}$$

that is just the exact analytical solution to Eq. (14) in this case.

In the second case we take instead $S(\mathbf{v}) = M(\mathbf{v})$. As $W(\varphi,\psi) = 0$ if $\varphi = 0$ or $\psi = 0$, and $V(\varphi) = 0$ if $\varphi = 0$ we observe that all the $F^{(n)}_{l,j}(\mathbf{v})$'s for any n,l,j are zero so that for any n we get

$$f_n(\mathbf{v},t) = Q_0 M(\mathbf{v}), \tag{44b}$$

coinciding with the exact analytical solution to Eq. (14) that is now in order.

[1] M. Krook and T. T. Wu, Phys. Rev. Lett. **36**, 1107 (1976).
[2] M. Krook and T. T. Wu, Phys. Fluids **20**, 1589 (1977).
[3] A. V. Bobylev. Sov. Phys. Dokl. **20**, 822 (1976).
[4] M. H. Ernst, Phys. Lett. A**69**, 390 (1979).
[5] M. Barnsley and G. Turchetti, Phys. Lett. A**72**, 417 (1979).
[6] M. Barnsley and G. Turchetti, Lett. Nuovo Cimento, **26**, 188 (1979).
[7] M. Barnsley and G. Turchetti, in *Bifurcation Phenomena in Mathematical Physics and Related Topics*, edited by C. Bardos and D. Bessis (Reidel, Doortrecht, 1980), p. 351.
[8] M. Barnsley and G. Turchetti, Lett. Nuovo Cimento **30**, 359 (1981).
[9] M. Barnsley and G. Turchetti, "Calculation of Some Solutions of the Bobylev-Krook-Wu and Tjon-Wu Equations", Nuovo Cimento (in press).
[10] M. A. Krasnosel'skii *et al.*, *Approximate Solution of Operator Equations* (Wolters-Noordhoff, Groningen 1972).
[11] V. C. Boffi and V. G. Molinari, Nuovo Cimento B **65**, 29 (1981).
[12] V. C. Boffi, Nuovo Cimento B **67**, 127 (1982).

On the inconsistency of a photon creation mechanism in an expanding universe

H. Ceccatto, A. Foussats, H. Giacomini, and O. Zandron
Instituto de Física Rosario, Conicet, UNR, Av. Pellegrini 250, 2000 Rosario, Argentina

(Received 14 October 1981; accepted for publication 23 December 1981)

We show here that if the quantum equivalence principle (QEP), as it was formulated in previous papers, is applied to the massless vector field, an inconsistent unphysical photon creation is found. This timelike and longitudinal photon creation is obtained when the 4-potential A^μ is quantized in a covariant way.

PACS numbers: 11.10. − z, 04.60. + n

1. INTRODUCTION

In previous papers (cf. Refs. 1 and 2) we developed the quantum field theory in curved space-time for massive and massless vector fields. In these papers we only quantified the material field, while the gravitational field was introduced as an unquantized external field through the curved space-time metric. For this purpose we used the Green's functions general theory (cf. Ref. 3), which consists in generalizing to curved space-time the tensor kernels $\Delta^{\mu\nu'}(x,x')$ and $\Delta_1^{\mu\nu'}(x,x')$ of the flat space-time, constructed from the Pauli–Jordan function.

We showed that (unlike the flat space-time), for generalizing the bivectorial kernel $G_1^{\mu\nu'}(x,x')$ [curved space-time generalization of the kernel $\Delta_1^{\mu\nu'}(x,x', m=0)$] it is not sufficient to know the biscalar kernel $G_1(x,x')$, curved space-time generalization of the kernel $\Delta_1(x,x')$. We also showed that the kernel $G_1^{\mu\nu'}(x, x')$ is not unique. That is to say, the formal properties that the kernel $G_1^{\mu\nu'}(x,x')$ must verify do not determine it uniquely. Therefore, there is no unique way to define the positive- and negative-frequency parts of the vector field. This problem was pointed out in Ref. 1.

In the scalar case, a similar difficulty occurs for the biscalar kernel $G_1(x,x')$. In Refs. 4 and 5, and by using the so-called quantum equivalence principle (QEP) for scalar fields, we found an adequate kernel $G_1^{(\Sigma)}(x,x')$ on each hypersurface Σ of the curved space-time (that we suppose globally hyperbolic). The formulation based on this idea (cf. Refs. 6 and 7) leads to the existence of particle creation at the expense of the gravitational field.

If we extend this idea to the massless vector field—giving on each hypersurface Σ of the curved space-time adequate Cauchy date for the bivectorial kernel $G_1^{(\Sigma)\mu\nu'}(x,x')$—a photon creation is obtained. This fact is in disagreement with the commonly accepted result that, in a metric conformal to the flat space-time one, there is no creation of massless particles (cf. Ref. 8). For the electromagnetic field, by quantizing the equation for the stress tensor $F^{\mu\nu}$, the conformal invariance of such an equation allows one to choose, in a natural way, a unique base for all V_4. Here we consider the equation for the 4-potential A^μ in a particular gauge and, therefore, such a choice cannot be made. Consequently, a creation of particles is possible.

In this paper we analyze such a creation by finding the time evolution of the creation and annihilation operators and we show that there is an inconsistency because unphysical (timelike and longitudinal) photons are created. The appearance of such "photons" in the intermediate steps of the argument, connected with the nonobservable 4-potential A^μ, was made in order that the theory be relativistically symmetric and covariant.

2. QUANTIZATION OF THE VECTOR FIELD

In this section we summarize the most relevant aspects of the quantization of the massless vector field in curved space-time.

We work in the particular case of a spatially-flat Robertson–Walker metric, i.e.,

$$g^{\mu\nu}=0,\quad \mu\neq\nu,\quad g^{00}=1,\quad g^{ii}=-\frac{1}{a^2(t)},\quad i=1,2,3 \tag{2.1}$$

where $a(t)$ is an arbitrary function of time t. In this metric the scalar curvature R is

$$R=-6\Big(\frac{\ddot a}{a}+H^2\Big), \tag{2.2}$$

with $\dot a = da/dt$ and $H=\dot a/a$ (Hubble coefficient).

The components of the contracted curvature tensor $R_{\mu\nu\rho}{}^{\mu}$ are

$$R_{00}=-3\frac{\ddot a}{a},\quad R_{ii}=2\dot a^2+a\ddot a,\quad R_{\mu\nu}=0,\quad \mu\neq\nu. \tag{2.3}$$

Starting from an adequate action integral (cf. Ref. 1) the following field equations are obtained:

$$\partial_0^2 A^j-\frac{1}{a^2}\nabla^2A^j+5H\partial_0A^j+\Big(2H^2-\frac{R}{3}\Big)A^j =\frac{2H}{a^2}\partial_j A^0, \tag{2.4a}$$

$$\partial_0^2 A^0-\frac{1}{a^2}\nabla^2A^0+3H\partial_0A^0-\Big(\frac{R}{2}+6H^2\Big)A^0 =2H\sum_{j=1}^{3}\partial_j A^j, \tag{2.4b}$$

$$\partial_0A^0+\sum_{j=1}^{3}\partial_jA^j+3HA^0=-B, \tag{2.4c}$$

where B is an unphysical auxiliary scalar field.

The Lorentz condition is established as a condition on the physical states, i.e.,

$$(\nabla_\mu A^\mu)^-|\text{phys}\rangle = B^-|\text{phys}\rangle = 0\,, \tag{2.5}$$

which ensures that the mean values of $\nabla\cdot A$ and B vanish. In (2.5), $(\nabla_\mu A^\mu)^-$ and B^- are the negative-frequency parts of the scalar fields $\nabla\cdot A$ and B, respectively.

We define the inner product in the vector field case as follows (see Refs. 1 and 2):

$$\langle \varphi^\mu;\psi_\mu\rangle = -i\int\Big[(\nabla_\nu \overset{*}{\varphi}{}^\mu)\psi_\mu - \overset{*}{\varphi}{}^\mu(\nabla_\nu\psi_\mu)\Big]d\tau^\nu\,. \tag{2.6}$$

Using the definition (2.6) of the inner product the positive-and negative-frequency parts of the field $A^\mu(x)$ are defined by

$$\overset{\oplus}{A}{}^\mu(x) = -i\Big\langle \overset{\ominus}{G}{}^{\mu\nu'}(x,x',m=0);A_{\nu'}(x')\Big\rangle\,, \tag{2.7a}$$

and

$$\overset{\ominus}{A}{}^\mu(x) = -i\Big\langle \overset{\oplus}{G}{}^{\mu\nu'}(x,x',m=0);A_{\nu'}(x')\Big\rangle\,, \tag{2.7b}$$

where

$$\overset{\ominus}{G}{}^{\mu\nu'}(x,x',m=0) = \tfrac{1}{2}\big[G^{\mu\nu'}(x,x',m=0) + iG_1^{\mu\nu'}(x,x',m=0)\big]\,, \tag{2.8a}$$

$$\overset{\oplus}{G}{}^{\mu\nu'}(x,x',m=0) = \tfrac{1}{2}\big[G^{\mu\nu'}(x,x',m=0) - iG_1^{\mu\nu'}(x,x',m=0)\big]\,. \tag{2.8b}$$

In Eqs. (2.8) the bivectorial kernel $G^{\mu\nu'}(x,x',m=0)$ is the solution propagator of Eq. (2.4a), i.e.,

$$A^\mu(x) = -i\langle G^{\mu\nu'}(x,x',m=0);A_{\nu'}(x')\rangle\,. \tag{2.9}$$

Such a kernel is the curved space-time generalization of the kernel $\Delta^{\mu\nu'}(x,x',m=0)$ of the flat space-time. It gives the field commutator

$$[A^\mu(x),A^{\nu'}(x')] = iG^{\mu\nu'}(x,x',m=0)\,. \tag{2.10}$$

The kernel $G_1^{\mu\nu'}(x,x')$ is the curved space-time generalization of the kernel $\Delta_1^{\mu\nu}(x,x')$ and must satisfy the following conditions[9] (Lichnerowicz conditions generalization[1]):

$$G_1^{\mu\nu'}(x,x') = \overset{*}{G}{}_1^{\mu\nu'}(x,x')\,, \tag{2.11a}$$

$$G_1^{\mu\nu'}(x,x') = G_1^{\nu'\mu}(x',x)\,, \tag{2.11b}$$

$$\Delta_x G_1^{\mu\nu'}(x,x') = \Delta_{x'}G_1^{\mu\nu'}(x,x') = 0\,, \tag{2.11c}$$

$$G^{\mu\nu'}(x,x') = i\langle G_1^{\mu\rho''};G_{1\rho''}^{\nu'}\rangle\,, \tag{2.11d}$$

$$\langle\text{phys}|(A^\mu;A_\mu)|\text{phys}\rangle_G \equiv = \langle\text{phys}|\langle A^\mu;A_{1\mu}\rangle|\text{phys}\rangle_G \geqslant 0\,, \tag{2.11e}$$

where

$$A_1^\mu(x) = i\langle G_1^{\mu\nu'};A_{\nu'}\rangle\,.$$

Besides, the bivectorial kernel $G_1^{\mu\nu'}(x,x')$ must fulfill the following equation (see Ref. 1, Sec. 7):

$$\nabla_\mu G_1^{\mu\nu'}(x,x') = -\nabla^{\nu'}G_1(x,x')\,, \tag{2.12}$$

where the kernel $G_1(x,x')$ is the curved space-time generalization of the kernel $\Delta_1(x,x',m=0)$.

Let $\{\phi^\mu_{\vec{k}s}\}\cup\{\overset{*}{\phi}{}^\mu_{\vec{k}s}\}$ be a base of complex solutions of Eq. (2.4a), orthonormalized according to (Ref. 10)

$$\langle\phi^\mu_{\vec{k}s};\phi_{\vec{k}'s'\mu}\rangle = \eta_{ss'}\,\delta(\vec{k}-\vec{k}')\,, \tag{2.13a}$$

$$\langle\overset{*}{\phi}{}^\mu_{\vec{k}s};\phi_{\vec{k}'s'\mu}\rangle = 0\,. \tag{2.13b}$$

In this base, the field $A^\mu(x)$ can be expanded as follows.

$$A^\mu(x) = \int d^3k\sum_{s=0}^{3}\Big\{a_{\vec{k}s}\phi^\mu_{\vec{k}s}(x) + a^\dagger_{\vec{k}s}\overset{*}{\phi}{}^\mu_{\vec{k}s}(x)\Big\} \tag{2.14}$$

with

$$a_{\vec{k}s} = \eta_{ss}\langle\phi^\mu_{\vec{k}s}(x);A_\mu(x)\rangle \tag{2.15a}$$

and

$$a^\dagger_{\vec{k}s} = -\eta_{ss}\langle\overset{*}{\phi}_{\vec{k}s}(x);A_\mu(x)\rangle\,. \tag{2.15b}$$

The operators $a^\dagger_{\vec{k}s}$ and $a_{\vec{k}s}$ satisfy the commutation relations

$$\big[a_{\vec{k}s},a^\dagger_{\vec{k}'s'}\big] = -\eta_{ss'}\delta(\vec{k}-\vec{k}')\,, \tag{2.16a}$$

$$\big[a_{\vec{k}s},a_{\vec{k}'s'}\big] = \big[a^\dagger_{\vec{k}s},a^\dagger_{\vec{k}'s'}\big] = 0\,, \tag{2.16b}$$

and can be interpreted as the creation and annihilation particle operators, respectively.

The kernel $G^{\mu\nu'}(x,x',m=0)$ as a function of the base has the expansion

$$G^{\mu\nu'}(x,x') = i\int d^3k\sum_{s=0}^{3}\eta_{ss}\Big\{\phi^\mu_{\vec{k}s}(x)\overset{*}{\phi}{}^{\nu'}_{\vec{k}s}(x') - \overset{*}{\phi}{}^\mu_{\vec{k}s}(x)\phi^{\nu'}_{\vec{k}s}(x')\Big\}\,. \tag{2.17}$$

This expansion is invariant under a general base transformation which preserves the orthonormality conditions (2.13) (Bogoliubov transformations). This fact ensures the uniqueness of the kernel $G^{\mu\nu'}(x,x')$.

The following expansion for the kernel $G_1^{\mu\nu'}(x,x')$,

$$G_1^{\mu\nu'}(x,x') = \int d^3k\sum_{s=0}^{3}\eta_{ss}\Big\{\phi^\mu_{\vec{k}s}(x)\overset{*}{\phi}{}^{\nu'}_{\vec{k}s}(x') + \overset{*}{\phi}{}^\mu_{\vec{k}s}(x)\phi^{\nu'}_{\vec{k}s}(x')\Big\}\,, \tag{2.18}$$

satisfies the conditions (2.11). The condition (2.12) is satisfied taking into account that the base $\{\phi^\mu_{\vec{k}s}\}\cup\{\overset{*}{\phi}{}^\mu_{\vec{k}s}\}$ satisfies

$$\nabla^\nu\phi_{\vec{k}} = C^*_{\vec{k}}(\phi^\nu_{\vec{k}0} + \phi^\nu_{\vec{k}3})\,, \tag{2.19a}$$

where $\{\phi_{\vec{k}}\}\cup\{\phi^\mu_{\vec{k}}\}$ is a base of complex solutions of the motion equation for the scalar field taking $m=0$, and $C_{\vec{k}}$ depends only on $\vec{k}$. Besides, it can be proved that

$$\nabla_\mu\phi^\mu_{\vec{k}s} = -C_{\vec{k}}(\eta_{0s}+\eta_{3s})\phi_{\vec{k}}\,. \tag{2.19b}$$

Equations (2.19) ensure that the generalized transversality conditions (see Ref. 1)

$$\langle \nabla^\mu \phi_{\underset{\to}{k}} ; \phi_{\underset{\to}{k}'s\mu} \rangle = C_{\underset{\to}{k}} (\eta_{0s} + \eta_{3s}) \delta(\underset{\to}{k} - \underset{\to}{k}'), \qquad (2.20)$$

are fulfilled.[11]

We have shown in Ref. 1 that Eq. (2.18) is not invariant under the Bogoliubov transformation which preserves the condition (2.19). This fact proves the nonuniqueness of the kernel $G_1^{\mu\nu}(x,x')$.

The particle-number operator is defined by

$$\widehat{N} = \tfrac{1}{2}:(A^\mu;A_\mu):\equiv\tfrac{1}{2}:\langle A^\mu;A_{1\mu}\rangle: . \qquad (2.21)$$

Using (2.14) and (2.18), it is easy to obtain

$$\widehat{N} = -\int d^3k \sum_{s=0}^{3} \eta_{ss} a^\dagger_{\underset{\to}{k}s} a_{\underset{\to}{k}s} . \qquad (2.22)$$

The condition (2.5) enables us to show that

$$\langle \text{phys}|\widehat{N}|\text{phys}\rangle_G = \int d^3k \, \langle \text{phys}| a^\dagger_{\underset{\to}{k}1} a_{\underset{\to}{k}1} + a^\dagger_{\underset{\to}{k}2} a_{\underset{\to}{k}2} |\text{phys}\rangle_G, \qquad (2.23)$$

i.e., the contributions of the pseudophotons (timelike and logitudinal) are mutually cancelled.

The definition (2.21) clearly shows that the particle number operator is $G_1^{\mu\nu}(x,x')$-dependent. That is, the operator $\widehat{N}$ is not unique.

In the flat space-time this difficulty is overcome by requiring Lorentz invariance for the theory. This requirement leads one to choose the plane wave as a base in which the expansion (2.18) must be done. In curved space-time, there is no analogous symmetry group. Hence, extra conditions to (2.11) and (2.12) must be introduced in the formulation for determining the kernel $G_1^{\mu\nu}(x,x')$ and the corresponding particle-number operator $\widehat{N}$.

A similar difficulty appears in the scalar case. This problem has been treated in Ref. 4 and 5 by using the so-called QEP. In the next section we will generalize this idea to the vector case.

3. THE QUANTUM EQUIVALENCE PRINCIPLE FOR THE MASSLESS VECTOR FIELD

In the scalar case, the QEP proposes giving up the idea of determining a unique $G_1(x,x')$ for all the curved space-time. Instead of this, it proposes the existence of a different $G_1^{(\Sigma)}(x,x')$ on each hypersurface Σ of curved space-time (supposed globally hyperbolic).

For determining the kernel $G_1^{(\Sigma)}(x,x')$ the strong equivalence principle and simplicity arguments are taken into account. Hence, the following Cauchy data on the hypersurface Σ, $t=$ const are given[12]:

$$G_1^{(\Sigma)}(x,x')|_\Sigma = \Delta_1(s), \qquad (3.1a)$$

$$\nabla_0 G_1^{(\Sigma)}(x,x')|_\Sigma = \nabla_{0'} G_1^{(\Sigma)}(x,x')|_\Sigma = \nabla_0 \Delta_1(s), \qquad (3.1b)$$

where $S(x,x')$ is the length of the geodesic arc between x and x'.

The most important consequence of these assumptions is that they lead to a particle creation at the expense of the gravitational field. We have proved besides that the density of created particles is finite (cf. Ref. 5).

To generalize these ideas to the vector case, it is necessary to find on the hypersurface Σ a unique kernel $G_1^{(\Sigma)\mu\nu}(x,x')$, which satisfy the conditions (2.11) and (2.12). Since the kernels $G^{\mu\nu}(x,x')$ and $G_1^{\mu\nu}(x,x')$ are solutions in both variables of the field equations (2.4), the condition (2.11d) is equivalent to the following relations on the hypersurface Σ, $t=\tau$:

$$\int_\Sigma d^3x'' \{ \nabla_{0''} G_1^{\mu\rho''}(x,x'') G^{\nu}_{1\rho''}(x',x'') - G_1^{\mu\rho''}(x,x'') \nabla_{0''} G^{\nu}_{1\rho''}(x',x'') \} = 0, \qquad (3.2a)$$

$$a^3(\tau) \int_\Sigma d^3x'' \{ \nabla_0 \nabla_{0''} G_1^{\mu\rho''}(x,x'') G^{\nu}_{1\rho''}(x',x'') - \nabla_0 G_1^{\mu\rho''}(x,x'') \nabla_{0''} G^{\nu}_{1\rho''}(x',x'') \} = \delta_s^{\mu\nu}(x,x') = \frac{g^{\mu\nu}(\tau)}{a^3(\tau)} \cdot \delta(\vec{x} - \vec{x}'), \qquad (3.2b)$$

$$\int_\Sigma d^3x'' \{ \nabla_0 \nabla_{0''} G_1^{\mu\rho''}(x,x'') \nabla_{0'} G^{\nu}_{1\rho''}(x',x'') - \nabla_0 G_1^{\mu\rho''}(x,x'') \nabla_{0'} \nabla_{0''} G^{\nu}_{1\rho''}(x',x'') \} = 0, \qquad (3.2c)$$

and the condition (2.12) is equivalent to

$$[\nabla_\mu G_1^{\mu\nu}(x,x')]|_\Sigma = -\nabla^\nu G_1(x,x')|_\Sigma, \qquad (3.3a)$$

$$\nabla_0 [\nabla_\mu G_1^{\mu\nu}(x,x')]|_\Sigma = -\nabla_0 \nabla^\nu G_1(x,x')|_\Sigma, \qquad (3.3b)$$

$$\nabla_{0'} [\nabla_\mu G_1^{\mu\nu}]|_\Sigma = -\nabla_{0'} \nabla^\nu G_1(x,x')|_\Sigma, \qquad (3.3c)$$

$$\nabla_{0'} \nabla_0 [\nabla_\mu G_1^{\mu\nu}(x,x')]|_\Sigma = -\nabla_{0'} \nabla_0 \nabla^\nu G_1(x,x')|_\Sigma. \qquad (3.3d)$$

To deal with the relations (3.2) it is convenient to write them by using the Fourier integral representation. Owing to the symmetry properties of the metric (2.1) the kernel $G_1^{\mu\nu}(x,x')$ must be a function of $\vec{r} = (\vec{x} - \vec{x}')$; therefore, we have

$$G_1^{\mu\nu}(x,x')|_\Sigma = \frac{1}{(2\pi)^{3/2}} \int d^3k \, \mathscr{F}_{\underset{\to}{k}} [G_1^{\mu\nu}(x,x')|_\Sigma] e^{i\underset{\to}{k}(\vec{x} - \vec{x}')}$$

and analogous expressions for $\nabla_0 G_1^{\mu\nu}|_\Sigma$ and $\nabla_{0'} \nabla_0 G_1^{\mu\nu}|_\Sigma$.

Using these expressions in Eq. (3.2), we obtain the relations

$$\mathscr{F}_{\underset{\to}{k}} (\nabla_{0''} G_1^{\mu\rho''}|_\Sigma) \mathscr{F}_{-\underset{\to}{k}} (G^{\nu}_{1\rho''}|_\Sigma) - \mathscr{F}_{\underset{\to}{k}} (G_1^{\mu\rho''}|_\Sigma) \mathscr{F}_{-\underset{\to}{k}} (\nabla_{0''} G^{\nu}_{1\rho''}|_\Sigma) = 0, \qquad (3.4a)$$

$$\mathscr{F}_{\underset{\to}{k}} (\nabla_0 \nabla_{0''} G_1^{\mu\rho''}|_\Sigma) \mathscr{F}_{-\underset{\to}{k}} (G^{\nu}_{1\rho''}|_\Sigma) - \mathscr{F}_{\underset{\to}{k}} (\nabla_0 G_1^{\mu\rho''}|_\Sigma) \mathscr{F}_{-\underset{\to}{k}} (\nabla_{0''} G^{\nu}_{1\rho''}|_\Sigma) = \frac{g^{\mu\nu}(\tau)}{(2\pi)^3 a^6(\tau)}, \qquad (3.4b)$$

$$\mathscr{F}_{\underset{\to}{k}} (\nabla_0 \nabla_{0''} G_1^{\mu\rho''}|_\Sigma) \mathscr{F}_{-\underset{\to}{k}} (\nabla_0'' G^{\nu}_{1\rho''}|_\Sigma - \mathscr{F}_{\underset{\to}{k}} (\nabla_0 G_1^{\mu\rho''}|_\Sigma) \mathscr{F}_{-\underset{\to}{k}} (\nabla_0 \nabla_{0''} G^{\nu}_{1\rho''}|_\Sigma) = 0. \qquad (3.4c)$$

Then, we see from Eq. (3.4b) that the mixing data $\nabla_0 \nabla_{0'} G_1^{\mu\nu}|_\Sigma$ is fixed when the data $G_1^{\mu\nu}|_\Sigma$, $\nabla_0 G_1^{\mu\nu}|_\Sigma$, and

$\nabla_{0'} G_1^{\mu\nu'}|_\Sigma$ are given.

In the following we assume that the Cauchy data $G_1^{\mu\nu'}|_\Sigma$, $\nabla_0 G_1^{\mu\nu'}|_\Sigma$, and $\nabla_{0'} G_1^{\mu\nu'}|_\Sigma$, which satisfy Eqs. (3.3) and (3.4), are known. That is to say, we have a unique kernel $G_1^{(\Sigma)\mu\nu'}(x, x')$ on each hypersurface Σ of the curved space-time. Hence, a photon creation takes place when going from a hypersurface Σ, $t = 0$ to another $t = \tau$. It is possible to show (see Sec. 6) that the created unphysical photons are not mutually cancelled.

4. EQUATIONS FOR THE BASE OF SOLUTIONS

In this section, and according to the above assumption, we are going to determine on the hypersurface Σ the base in which the decomposition of the vector field $A^\mu(x)$ in its positive- and negative-frequency parts must be done.

Let $\{\phi^\mu_{\underset{\cdot}{k} s}\} \cup \{\overset{*}{\phi}{}^\mu_{\underset{\cdot}{k} s}\}$ be the "good" base of complex solutions of Eqs. (2.4a) and (2.4b) in which the kernel $G_1^{\mu\nu'}(x, x')$ has the expansion (2.18). Using Eq. (2.18) and the orthonormality conditions (2.13) we get

$$\langle \overset{*}{\phi}{}^\mu_{\underset{\cdot}{k} s}(x); G^{\nu'}_{1\cdot\mu}(x', x)\rangle = -\phi^{\nu'}_{\underset{\cdot}{k} s}(x'), \tag{4.1a}$$

$$\langle \overset{*}{\phi}{}^\mu_{\underset{\cdot}{k} s}(x); \nabla_{0'} G^{\nu'}_{1\cdot\mu}(x', x)\rangle = -\nabla_{0'}\phi^{\nu'}_{\underset{\cdot}{k} s}(x'). \tag{4.1b}$$

Taking into account the expression (2.6) for the inner product, Eqs. (4.1) on the hypersurface Σ, $t = \tau$ result:

$$ia^3(\tau)\int_\Sigma d^3x\Big[\nabla_0\phi^\mu_{\underset{\cdot}{k} s}(x) G^{\nu'}_{1\cdot\mu}(x', x) - \phi^\mu_{\underset{\cdot}{k} s}(x)\nabla_0 G^{\nu'}_{1\cdot\mu}(x', x)\Big] = \phi^{\nu'}_{\underset{\cdot}{k} s}(x'), \tag{4.2a}$$

$$ia^3(\tau)\int_\Sigma d^3x\Big[\nabla_0\phi^\mu_{\underset{\cdot}{k} s}(x)\nabla_{0'} G^{\nu'}_{1\cdot\mu}(x', x) - \phi^\mu_{\underset{\cdot}{k} s}(x)\nabla_0\nabla_{0'} G^{\nu'}_{1\cdot\mu}(x', x)\Big] = \nabla_{0'}\phi^{\nu'}_{\underset{\cdot}{k} s}(x'). \tag{4.2b}$$

If we know $G_1^{\mu\nu'}(x, x')|_\Sigma$, $\nabla_0 G_1^{\mu\nu'}(x, x')|_\Sigma$, and $\nabla_{0'} G_1^{\mu\nu'}(x, x')|_\Sigma$ Eqs. (2.13), (2.19), and (4.2) enable us to determine the adequate initial conditions for $\phi^\mu_{\underset{\cdot}{k} s}$ on the hypersurface Σ, $t = \tau$.[13]

According to the symmetry properties of the metric (2.1) we propose the following form for the base[14]:

$$\phi^\mu_{\underset{\cdot}{k} s}(x) = f^\mu_{\underset{\cdot}{k} s}(t) e^{-i\vec{k}\vec{x}}. \tag{4.3}$$

Taking into account Eqs. (2.4a), (2.4b), (2.13), and (2.19) it is possible to show that $f^\mu_{\underset{\cdot}{k} s}$ must satisfy

$$\ddot{f}^0_{\underset{\cdot}{k} s}(t) + 3H\dot{f}^0_{\underset{\cdot}{k} s}(t) + \Big(\frac{k^2}{a^2} - \frac{R}{2} - 6H^2\Big) f^0_{\underset{\cdot}{k} s}(t) = -2iH\sum_{j=1}^3 k_j f^j_{\underset{\cdot}{k} s}(t), \tag{4.4a}$$

$$\ddot{f}^i_{\underset{\cdot}{k} s}(t) + 5H\dot{f}^i_{\underset{\cdot}{k} s}(t) + \Big(\frac{k^2}{a^2} - \frac{R}{3} + 2H^2\Big) f^i_{\underset{\cdot}{k} s}(t) = 2iHk^i f^0_{\underset{\cdot}{k} s}(t), \tag{4.4b}$$

$$(f^\mu_{\underset{\cdot}{k} s}(t); f_{\underset{\cdot}{k} s'\mu}(t)) = \eta_{ss'}, \tag{4.5a}$$

$$(f^\mu_{\underset{\cdot}{k} s}(t); \overset{*}{f}_{-\underset{\cdot}{k} s'\mu}(t)) = 0, \tag{4.5b}$$

and

$$\dot{f}_{\underset{\cdot}{k}}(t) = \overset{*}{C}_{\underset{\cdot}{k}}\Big[f^0_{\underset{\cdot}{k} 0}(t) + f^0_{\underset{\cdot}{k} 3}(t)\Big], \tag{4.6a}$$

$$-ik^j f_{\underset{\cdot}{k}}(t) = \overset{*}{C}_{\underset{\cdot}{k}}\Big[f^j_{\underset{\cdot}{k} 0}(t) + f^j_{\underset{\cdot}{k} 3}(t)\Big], \tag{4.6b}$$

where

$$(g^\mu(t); h_\mu(t)) = \frac{(2\pi)^3}{i} a^3(t)\Big\{\Big(\nabla_0 \overset{*}{g}{}^\mu(t)\Big) h_\mu(t) - \overset{*}{g}{}^\mu(t)(\nabla_0 h_\mu(t))\Big\}, \tag{4.7}$$

and we have called $f_{\underset{\cdot}{k}}(t)$ the time-dependent part of the scalar base $\phi_{\underset{\cdot}{k}}(x) = f_{\underset{\cdot}{k}}(t)\exp(-i\vec{k}\vec{x})$.

Equations (4.2) for $f^\mu_{\underset{\cdot}{k} s}(t = \tau)$ and $\nabla_0 f^\mu_{\underset{\cdot}{k} s}(t = \tau)$ result in

$$(2\pi)^{3/2}a^3(\tau)i\Big\{\Big(\nabla_0 f^\mu_{\underset{\cdot}{k} s}(\tau)\Big) A^{\nu'}_{\underset{\cdot}{k}\cdot\mu}(\tau) - f^\mu_{\underset{\cdot}{k} s}(\tau) B^{\nu'}_{\underset{\cdot}{k}\cdot\mu}(\tau)\Big\} = f^{\nu'}_{\underset{\cdot}{k} s}(\tau), \tag{4.8a}$$

$$(2\pi)^{3/2}a^3(\tau)i\Big\{\Big(\nabla_0 f^\mu_{\underset{\cdot}{k} s}(\tau)\Big)\tilde{B}^{\nu'}_{\underset{\cdot}{k}\cdot\mu}(\tau) - f^\mu_{\underset{\cdot}{k} s}(\tau) C^{\nu'}_{\underset{\cdot}{k}\cdot\mu}(\tau)\Big\} = \nabla_{0'} f^{\nu'}_{\underset{\cdot}{k} s}(\tau), \tag{4.8b}$$

where

$$A^{\nu'}_{\underset{\cdot}{k}\cdot\mu}(\tau) = \mathscr{F}_{\underset{\cdot}{k}}(G^{\nu'}_{1\cdot\mu}|_\Sigma), \quad B^{\nu'}_{\underset{\cdot}{k}\cdot\mu}(\tau) = \mathscr{F}_{\underset{\cdot}{k}}(\nabla_0 G^{\nu'}_{1\cdot\mu}|_\Sigma),$$
$$\tilde{B}^{\nu'}_{\underset{\cdot}{k}\cdot\mu}(\tau) = \mathscr{F}_{\underset{\cdot}{k}}(\nabla_{0'} G^{\nu'}_{1\cdot\mu}|_\Sigma), \quad C^{\nu'}_{\underset{\cdot}{k}\cdot\mu}(\tau) = \mathscr{F}_{\underset{\cdot}{k}}(\nabla_{0'}\nabla_0 G^{\nu'}_{1\cdot\mu}|_\Sigma). \tag{4.9}$$

Owing to the symmetry property (2.11b),

$$A^{\nu'}_{\underset{\cdot}{k}\cdot\mu} = A_{-\underset{\cdot}{k}\mu\cdot}{}^{\nu'}, \quad B^{\nu'}_{\underset{\cdot}{k}\cdot\mu} = \tilde{B}_{-\underset{\cdot}{k}\mu\cdot}{}^{\nu'}, \quad C^{\nu'}_{\underset{\cdot}{k}\cdot\mu} = C_{-\underset{\cdot}{k}\mu\cdot}{}^{\nu'}. \tag{4.10}$$

Taking into account (4.10), Eqs. (3.4) and (4.8) result in

$$\mathbb{A}_{\underset{\cdot}{k}}(\tau)\tilde{\mathbb{B}}_{\underset{\cdot}{k}}(\tau) - \mathbb{B}_{\underset{\cdot}{k}}(\tau)\mathbb{A}_{\underset{\cdot}{k}}(\tau) = 0, \tag{4.11a}$$

$$\mathbb{A}_{\underset{\cdot}{k}}(\tau)\mathbb{C}_{\underset{\cdot}{k}}(\tau) - \mathbb{B}^2_{\underset{\cdot}{k}}(\tau) = \alpha^2(\tau)\mathbb{I}, \tag{4.11b}$$

$$\mathbb{B}_{\underset{\cdot}{k}}(\tau)\mathbb{C}_{\underset{\cdot}{k}}(\tau) - \mathbb{C}_{\underset{\cdot}{k}}(\tau)\tilde{\mathbb{B}}_{\underset{\cdot}{k}}(\tau) = 0, \tag{4.11c}$$

$$\mathbb{A}_{\underset{\cdot}{k}}(\tau) y_{\underset{\cdot}{k} s}(\tau) - \mathbb{B}_{\underset{\cdot}{k}}(\tau) x_{\underset{\cdot}{k} s}(\tau) = -i\alpha(\tau) x_{\underset{\cdot}{k} s}(\tau), \tag{4.12}$$

$$\tilde{\mathbb{B}}_{\underset{\cdot}{k}}(\tau) y_{\underset{\cdot}{k} s}(\tau) - \mathbb{C}_{\underset{\cdot}{k}}(\tau) x_{\underset{\cdot}{k} s}(\tau) = -i\alpha(\tau) y_{\underset{\cdot}{k} s}(\tau), \tag{4.13}$$

where we have defined the matrices $\mathbb{A}_{\underset{\cdot}{k}}$, $\mathbb{B}_{\underset{\cdot}{k}}$, $\tilde{\mathbb{B}}_{\underset{\cdot}{k}}$, $\mathbb{C}_{\underset{\cdot}{k}}$, and $\mathbb{I}$, the elements of which are $A^{\nu'}_{\underset{\cdot}{k}\cdot\mu}$, $B^{\nu'}_{\underset{\cdot}{k}\cdot\mu}$, $\tilde{B}^{\nu'}_{\underset{\cdot}{k}\cdot\mu}$, $C^{\nu'}_{\underset{\cdot}{k}\cdot\mu}$, and $\delta^{\nu'}_\mu$, respectively. Besides, we have set

$$\alpha = (2\pi)^{-3/2}a^{-3}, \quad x^\mu_{\underset{\cdot}{k} s} \equiv f^\mu_{\underset{\cdot}{k} s}, \quad y^\mu_{\underset{\cdot}{k} s} \equiv \nabla_0 f^\mu_{\underset{\cdot}{k} s}. \tag{4.14}$$

Taking into account (4.11a) and (4.11b) it is possible to show that Eq. (4.13) is dependent on (4.12). An analogous fact occurs in the scalar case (see Ref. 4).

From Eq. (4.12) we obtain

$$y_{\vec{k}s}(\tau) = \mathbb{D}_{\vec{k}}(\tau) x_{\vec{k}s}(\tau), \tag{4.15}$$

where

$$\mathbb{D}_{\vec{k}}(\tau) = \mathbb{A}_{\vec{k}}^{-1}(\tau)\Big[\mathbb{B}_{\vec{k}}(\tau) - i\alpha(\tau)\mathbb{I}\Big]. \tag{4.16}$$

In the matrix notation, for $t = \tau$, the orthonormality conditions (4.5) are

$$y_{\vec{k}s}(\tau)\overset{*}{x}_{\vec{k}s'}(\tau) - \overset{*}{y}_{\vec{k}s'}(\tau)x_{\vec{k}s}(\tau) = -i\eta_{ss'}/[2\pi a(\tau)]^3, \tag{4.17a}$$

$$y_{\vec{k}s}(\tau)x_{-\vec{k}s'}(\tau) - y_{-\vec{k}s'}(\tau)x_{\vec{k}s}(\tau) = 0. \tag{4.17b}$$

The transversality conditions (4.6) for $t = \tau$ are equivalent to the equations

$$x_{\vec{k}0}(\tau) + x_{\vec{k}3}(\tau) = \gamma_{\vec{k}}(\tau), \tag{4.18a}$$

$$y_{\vec{k}0}(\tau) + y_{\vec{k}3}(\tau) = \epsilon_{\vec{k}}(\tau), \tag{4.18b}$$

where

$$\gamma^0_{\vec{k}} = \frac{1}{C^*_{\vec{k}}}\nabla_0 f_{\vec{k}}, \quad \gamma^j_{\vec{k}} = \frac{i}{C_{\vec{k}}} k^j f_{\vec{k}},$$

$$\beta^0_{\vec{k}} = \frac{1}{C_{\vec{k}}}\left(\frac{k^2}{a^2} f_{\vec{k}} + 3H\nabla_0 f_{\vec{k}}\right),$$

$$\beta^j_{\vec{k}} = \frac{i}{C_{\vec{k}}} k^j(\nabla_0 f_{\vec{k}} - H f_{\vec{k}}), \tag{4.19}$$

(for obtaining $\beta^0_{\vec{k}}$, the motion equation for the scalar case has been used).

The system (4.15)–(4.18) is compatible (and dependent) if

$$\mathbb{D}_{\vec{k}}(\tau)\gamma_{\vec{k}}(\tau) = \beta_{\vec{k}}(\tau), \tag{4.20}$$

which can be proved by straightforward calculations by using Eqs. (3.2), (3.3), (4.9), (4.16), and (4.19). The fulfillment of relation (4.20) is a consequence of Eq. (2.12).

To solve Eqs. (4.15)–(4.18) it is necessary to generalize the usual choice of the polarization vectors in the flat space-time. These equations, which define the initial conditions for the base at $t = \tau$, are analogous to the following equations of the flat space-time:

$$\dot{f}^\mu_{\vec{k}s}(\tau) = -ikf^\mu_{\vec{k}s}(\tau),$$

$$\epsilon^\mu_{\vec{k}s}\epsilon_{\vec{k}s'\mu} = \eta_{ss'},$$

and

$$\epsilon^\mu_{\vec{k}0} + \epsilon^\mu_{\vec{k}3} = \frac{k^\mu}{k},$$

with

$$f^\mu_{\vec{k}s}(\tau) = \epsilon^\mu_{\vec{k}s}\frac{1}{(2\pi)^{3/2}}\frac{e^{-ik\tau}}{(2k)^{1/2}}.$$

According to Eq. (2.19) (cf. Ref. 11), the adequate generalization to curved space-time of the k^μ vector is

$$\frac{k^\mu}{k}\frac{e^{-ik_\mu x^\mu}}{(2\pi)^{3/2}(2k)^{1/2}} \to \frac{\nabla_\mu\phi_k}{C^*_{\vec{k}}} \equiv \gamma_\mu e^{-i\vec{k}\vec{x}}. \tag{4.21}$$

In the flat space-time the polarization vectors $\epsilon^\mu_{\vec{k}3}$ and $\epsilon^\mu_{\vec{k}0}$ are chosen as follows:

$$\epsilon^\mu_{\vec{k}3} = (0,\vec{k}/k), \quad \epsilon^\mu_{\vec{k}0} = (1,\vec{0}). \tag{4.22}$$

Consequently, in curved space-time, and according to (4.18), (4.21), and (4.22), we choose

$$x^\mu_{\vec{k}3} = (0,\vec{\gamma}_{\vec{k}}), \quad x^\mu_{\vec{k}0} = (\gamma^0_{\vec{k}},\vec{0}). \tag{4.23}$$

By replacing in (4.17), and taking into account (4.15), it is easy to verify that $x^\mu_{\vec{k}3}$ and $x^\mu_{\vec{k}0}$ are correctly orthonormalized for the value $C_{\vec{k}} = ik/a(\tau)$. This value of $C_{\vec{k}}$ can also be obtained by requiring that Eq. (2.19) is satisfied in the flat case.

For determining $x^\mu_{\vec{k}1}$ and $x^\mu_{\vec{k}2}$ we must solve the remaining equations (4.17), i.e.,

$$y_{\vec{k}r}(\tau)\overset{*}{x}_{\vec{k}r'}(\tau) - \overset{*}{y}_{\vec{k}r'}(\tau)x_{\vec{k}r}(\tau) = -i\frac{\eta_{rr'}}{[2\pi a(\tau)]^3},$$

$$y_{\vec{k}r}(\tau)x_{-\vec{k}r'}(\tau) - y_{-\vec{k}r'}(\tau)x_{\vec{k}r}(\tau) = 0, \quad r,r' = 1,2 \tag{4.24}$$

and

$$\beta^*_{\vec{k}}(\tau)x_{\vec{k}r}(\tau) - \beta_{\vec{k}}(\tau)\overset{*}{x}_{\vec{k}r}(\tau) = 0,$$

$$y^*_{\vec{k}r}(\tau)\gamma_{\vec{k}}(\tau) - y_{\vec{k}r}(\tau)\gamma^*_{\vec{k}}(\tau) = 0, \quad r = 1,2. \tag{4.25}$$

Equations (4.25) are the orthonormality conditions between $x_{\vec{k}r}$ $(r = 1,2)$ and $x_{\vec{k}0}$, $x_{\vec{k}3}$. They can be obtained from (4.17) taking into account (4.15), (4.20), and the following identity:

$$\beta^*_{\vec{k}}(\tau)\gamma_{\vec{k}}(\tau) - \beta_{\vec{k}}(\tau)\gamma^*_{\vec{k}}(\tau) = 0. \tag{4.26}$$

5. TIME EVOLUTION OF THE PARTICLE NUMBER OPERATOR

The present formulation of quantum field theory in curved space-time leads to particle creation at the expense of the gravitational field. As is well known (cf. Ref. 8), the Bogoliubov transformation can be interpreted as the mechanism to analyze such a creation.

Let $\left\{\phi^{(\tau)\mu}_{\vec{k}s}\right\} \cup \left\{\phi^{(\tau)*\mu}_{\vec{k}s}\right\}$ be a base of the space of complex solutions of Eq. (2.4a) and (2.4b) that satisfy the initial conditions (4.22) and (4.26) at the time $t=\tau$ and $\{\overset{(\tau')\mu}{\phi}_{\vec{k}s}\}\cup\{\overset{(\tau')*\mu}{\phi}_{\vec{k}s}\}$ the base that satisfies analogous initial conditions at the time $t=\tau'$, both orthonormalized according to (2.13). The most general transformation (Bogoliubov transformation) which links two bases is

$$\overset{(\tau)\mu}{\phi}_{\vec{k}s} = \int d^3k' \sum_{s'=0}^{3}\left(\alpha_{\vec{k}\vec{k}'ss'}\overset{(\tau')\mu}{\phi}_{\vec{k}'s'} + \beta_{\vec{k}\vec{k}'ss'}\overset{(\tau')*\mu}{\phi}_{\vec{k}'s'}\right). \tag{5.1}$$

Owing to the metric used we are going to take as a base of solutions those with well defined momentum $\vec{k}$. Thus, it is easy to notice that [using (4.3)] the expression (5.1) is reduced to

$$\overset{(\tau)\mu}{f}_{\vec{k}s} = \sum_{s'=0}^{3}\left(\alpha_{\vec{k}ss'}\overset{(\tau')\mu}{f}_{\vec{k}s'} + \beta_{-\vec{k}ss'}\overset{(\tau')*\mu}{f}_{-\vec{k}s'}\right), \tag{5.2a}$$

and

$$\overset{(\tau)*\mu}{f}_{\vec{k}s} = \sum_{s'=0}^{3}\left(\alpha^*_{\vec{k}ss'}\overset{(\tau')*\mu}{f}_{\vec{k}s'} + \beta^*_{-\vec{k}ss'}\overset{(\tau')\mu}{f}_{-\vec{k}s'}\right). \tag{5.2b}$$

From this relationship between the bases and taking into account the expression (2.14), the following relation between the creation and annihilation operators can be obtained:

$$\overset{(\tau')}{a}_{\vec{k}s'} = \sum_{s=0}^{3}\left(\alpha_{\vec{k}ss'}a^{(\tau)}_{\vec{k}s} + \beta^*_{\vec{k}ss'}\overset{(\tau)\dagger}{a}_{-\vec{k}s}\right), \tag{5.3a}$$

$$\overset{(\tau')\dagger}{a}_{\vec{k}s'} = \sum_{s=0}^{3}\left(\alpha^*_{\vec{k}ss'}\overset{(\tau)\dagger}{a}_{\vec{k}s} + \beta_{\vec{k}ss'}\overset{(\tau)}{a}_{-\vec{k}s}\right). \tag{5.3b}$$

Making use of Eq. (5.2) and the orthonormality conditions (2.13) in τ and in τ', the following conditions for the coefficients $\alpha_{\vec{k}ss'}$ and $\beta_{\vec{k}ss'}$ result:

$$\sum_{s'=0}^{3}\left(\alpha_{\vec{k}ss'}\alpha^*_{\vec{k}s''s'} - \beta_{-\vec{k}ss'}\beta^*_{-\vec{k}s''s'}\right) = -\eta_{ss''}, \tag{5.4a}$$

$$\sum_{s'=0}^{3}\left(\beta_{-\vec{k}ss'}\alpha_{-\vec{k}s''s'} - \alpha_{\vec{k}ss'}\beta_{\vec{k}s''s'}\right) = 0. \tag{5.4b}$$

Using (5.3) it is possible to calculate the mean value of the density of particles in $t=\tau$ if we know such a value in $t=0$.

If we suppose that the initial state in $t=0$ be the vacuum state $|0\rangle_\tau$, we get

$$\int d^3k\, {}_\tau\langle 0|N_{\vec{k}}(\tau)|0\rangle_\tau = \int d^3k \sum_{s,s'=0}^{3}|\beta_{\vec{k}ss'}(\tau,\tau')|^2. \tag{5.5}$$

Now we will discuss the particle creation in an alternative way, taking into consideration the time evolution of the creation and annihilation operators.

We define

$$\psi^\mu_{\vec{k}s}(\vec{x},\tau) = \overset{(\tau)\mu}{\phi}_{\vec{k}s}(\vec{x},\tau) = \overset{(\tau)\mu}{f}_{\vec{k}s}(\tau)e^{-i\vec{k}\vec{x}}, \tag{5.6}$$

$$F^\mu_{\vec{k}s}(\tau) = \overset{(\tau)\mu}{f}_{\vec{k}s}(\tau),$$

$$\nabla_0\psi^\mu_{\vec{k}s}(\vec{x},\tau) = \nabla_0\overset{(\tau)\mu}{\phi}_{\vec{k}s}(\vec{x},\tau) = \nabla_0\overset{(\tau)\mu}{f}_{\vec{k}s}(\tau)e^{-i\vec{k}\vec{x}}, \tag{5.7}$$

$$\nabla_0 F^\mu_{\vec{k}s}(\tau) = \nabla_0\overset{(\tau)\mu}{f}_{\vec{k}s}(\tau).$$

The functions $\psi^\mu_{\vec{k}s}(\vec{x},\tau)$ and $\nabla_0\psi^\mu_{\vec{k}s}(\vec{x},\tau)$ are defined like this throughout the space-time and they constitute the collection of the initial conditions for the basis.

By developing the field $A^\mu(x)$ in the base $\{\overset{(\tau)\mu}{\phi}_{\vec{k}s}\}\cup\{\overset{(\tau)*\mu}{\phi}_{\vec{k}s}\}$, we get

$$A^\mu(x) = \int d^3k \sum_{s=0}^{3}\left[\overset{(\tau)}{a}_{\vec{k}s}\overset{(\tau)\mu}{\phi}_{\vec{k}s}(x) + \overset{(\tau)\dagger}{a}_{\vec{k}s}\overset{(\tau)*\mu}{\phi}_{\vec{k}s}(x)\right]. \tag{5.8}$$

For $x=(\vec{x},\tau)$, according to (5.6) and (5.7),

$$A^\mu(\vec{x},\tau) = \int d^3k \sum_{s=0}^{3}\left[a_{\vec{k}s}(\tau)F^\mu_{\vec{k}s}(\tau) + a^\dagger_{-\vec{k}s}(\tau)F^{*\mu}_{-\vec{k}s}(\tau)\right] \times e^{-i\vec{k}\vec{x}}, \tag{5.9}$$

where we have used $a^{(\tau)}_{\vec{k}s} = a_{\vec{k}s}(\tau)$ and $\overset{(\tau)\dagger}{a}_{\vec{k}s} = a^\dagger_{\vec{k}s}(\tau)$.

Similarly,

$$\nabla_0 A^\mu(\vec{x},\tau) = \int d^3k \sum_{s=0}^{3}\left[a_{\vec{k}s}(\tau)\nabla_0F^\mu_{\vec{k}s}(\tau) + a^\dagger_{-\vec{k}s}(\tau)\nabla_0\overset{*}{F}{}^\mu_{-\vec{k}s}(\tau)\right] \times e^{-i\vec{k}\vec{x}}. \tag{5.10}$$

Since (5.9) and (5.10) are valid for every τ, the field equations (2.4a) and (2.4b) require that the initial conditions $F^\mu_{\vec{k}s}$ and $\nabla_0F^\mu_{\vec{k}s}$ and the $a_{\vec{k}s}$ and $a^\dagger_{\vec{k}s}$ operators change in such a way that the quantity

$$\sum_{s=0}^{3}\left[a_{\vec{k}s}(\tau)F^\mu_{\vec{k}s}(\tau) + a^\dagger_{-\vec{k}s}(\tau)\overset{*}{F}{}^\mu_{-\vec{k}s}(\tau)\right].$$

satisfies Eq. (4.4) and the equation

$$\frac{d}{d\tau}\left\{\sum_{s=0}^{3}\left[a_{\vec{k}s}(\tau)F^\mu_{\vec{k}s}(\tau) + a^\dagger_{-\vec{k}s}(\tau)\overset{*}{F}{}^\mu_{-\vec{k}s}(\tau)\right]\right\} = \sum_{s=0}^{3}\left[a_{\vec{k}s}(\tau)\nabla_0 F^\mu_{\vec{k}s}(\tau) + a^\dagger_{-\vec{k}s}(\tau)\nabla_0\overset{*}{F}{}^\mu_{-\vec{k}s}(\tau)\right]. \tag{5.11}$$

If we call $\overset{(1)}{h}{}^\mu_{\vec{k}s}$ and $\overset{(2)}{h}{}^\mu_{\vec{k}s}$ a base of solutions of Eqs. (4.4) normalized so that

$$\overset{(1)}{h}{}^\mu_{\vec{k}s}(\tau) = \delta^\mu_s, \quad \nabla_0\overset{(1)}{h}{}^\mu_{\vec{k}s}(\tau) = 0,$$

$$\overset{(2)}{h}{}^\mu_{\vec{k}s}(\tau) = 0, \quad \nabla_0\overset{(2)}{h}{}^\mu_{\vec{k}s}(\tau) = \delta^\mu_s, \tag{5.12}$$

the general solution of these equations is

$$h^{\mu}_{\vec{k}}(t) = \sum_{s=0}^{3}\left[C_{\vec{k}s}\overset{(1)}{h}{}^{\mu}_{\vec{k}s}(t) + D_{\vec{k}s}\overset{(2)}{h}{}^{\mu}_{\vec{k}s}(t)\right]. \quad (5.13)$$

We must impose on solution (5.13) the following initial conditions:

$$h^{\mu}_{\vec{k}}(\tau) = \sum_{s=0}^{3}\left[a_{\vec{k}s}(\tau)F^{\mu}_{\vec{k}s}(\tau) + a^{\dagger}_{-\vec{k}s}(\tau)\overset{*}{F}{}^{\mu}_{-\vec{k}s}(\tau)\right], \quad (5.14)$$

and

$$\frac{d}{dt}h^{\mu}_{\vec{k}}\bigg|_{t=\tau} = \sum_{s=0}^{3}\Big[a_{\vec{k}s}(\tau)\nabla_0 F^{\mu}_{\vec{k}s}(\tau) + a^{\dagger}_{-\vec{k}s}(\tau)\nabla_0\overset{*}{F}{}^{\mu}_{-\vec{k}s}(\tau)\Big]. \quad (5.15)$$

From Eqs. (5.14) and (5.15), we get the following expression for $C_{\vec{k}s}$ and $D_{\vec{k}s}$

$$C_{\vec{k}s} = \sum_{s'=0}^{3}\left[a_{\vec{k}s'}(\tau)F^{s}_{\vec{k}s'}(\tau) + a^{\dagger}_{-\vec{k}s'}(\tau)\overset{*}{F}{}^{s}_{-\vec{k}s'}(\tau)\right], \quad (5.16a)$$

$$D_{\vec{k}s} = \sum_{s=0}^{3}\left[a_{\vec{k}s'}(\tau)\nabla_0 F^{s}_{\vec{k}s'}(\tau) + a^{\dagger}_{-\vec{k}s'}(\tau)\nabla_0\overset{*}{F}{}^{s}_{-\vec{k}s'}(\tau)\right]. \quad (5.16b)$$

Taking into account (5.13) and (5.16) we obtain

$$h^{\mu}_{\vec{k}}(t) = \sum_{s=0}^{3}\left[a_{\vec{k}s}(t)F^{\mu}_{\vec{k}s}(t) + a^{\dagger}_{-\vec{k}s}(t)\overset{*}{F}{}^{\mu}_{-\vec{k}s}(t)\right]$$
$$= \sum_{s=0}^{3}\left[a_{\vec{k}s}(\tau)\overset{(\tau)}{f}{}^{\mu}_{\vec{k}s}(t) + a^{\dagger}_{-\vec{k}s}(\tau)\overset{(\tau)}{f}{}^{*\mu}_{-\vec{k}s}(t)\right], \quad (5.17)$$

where

$$\overset{(\tau)}{f}{}^{\mu}_{\vec{k}s}(t) = \sum_{r=0}^{3}\left[F^{r}_{\vec{k}s}(\tau)\overset{(1)}{h}{}^{\mu}_{\vec{k}r}(t) + \nabla_0 F^{r}_{\vec{k}s}(\tau)\overset{(2)}{h}{}^{\mu}_{\vec{k}r}(t)\right], \quad (5.18)$$

is the base orthonormalized in $t=\tau$.

Similarly, as a result of Eqs. (5.11), results

$$\sum_{s=0}^{3}\left[a_{\vec{k}s}(t)\nabla_0 F^{\mu}_{\vec{k}s}(t) + a^{\dagger}_{-\vec{k}s}(t)\nabla_0\overset{*}{F}{}^{\mu}_{-\vec{k}s}(t)\right]$$
$$= \sum_{s=0}^{3}\left[a_{\vec{k}s}(\tau)\nabla_0\overset{(\tau)}{f}{}^{\mu}_{\vec{k}s}(t) + a^{\dagger}_{-\vec{k}s}(\tau)\nabla_0\overset{(\tau)}{f}{}^{*\mu}_{-\vec{k}s}(t)\right], \quad (5.19)$$

where

$$\nabla_0\overset{(\tau)}{f}{}^{\mu}_{\vec{k}s}(t) = \sum_{s=0}^{3}\left[F^{r}_{\vec{k}s}(\tau)\nabla_0\overset{(1)}{h}{}^{\mu}_{\vec{k}r}(t) + \nabla_0 F^{r}_{\vec{k}s}(\tau)\nabla_0\overset{(2)}{h}{}^{\mu}_{\vec{k}r}(t)\right]. \quad (5.20)$$

From Eqs. (5.17) and (5.19), and taking into account the orthonormality conditions (4.5) and the definition (4.8), we find

$$a_{\vec{k}s}(\tau') = \eta_{ss}\sum_{s'=0}^{3}\left[\left(\overset{(\tau')}{f}{}^{\mu}_{\vec{k}s};\overset{(\tau)}{f}_{\vec{k}s'\mu}\right)a_{\vec{k}s'}(\tau) + \left(\overset{(\tau')}{f}{}^{\mu}_{\vec{k}s};\overset{(\tau)}{f}{}^{*}_{-\vec{k}s'\mu}\right)a^{\dagger}_{-\vec{k}s'}(\tau)\right], \quad (5.21a)$$

and

$$a^{\dagger}_{-\vec{k}s}(\tau') = -\eta_{ss}\sum_{s'=0}^{3}\left[\left(\overset{(\tau')}{f}{}^{*\mu}_{-\vec{k}s};\overset{(\tau)}{f}_{\vec{k}s'\mu}\right)a_{\vec{k}s'}(\tau) + \left(\overset{(\tau')}{f}{}^{*\mu}_{-\vec{k}s};\overset{(\tau)}{f}{}^{*}_{-\vec{k}s'\mu}\right)a^{\dagger}_{-\vec{k}s'}(\tau)\right]. \quad (5.21b)$$

Using Eq. (5.21), and if $|0\rangle_\tau$ represents the vacuum state in $t=\tau$, the mean value $N_{\vec{k}s}(\tau,\tau')$ of the density of the $\vec{k},s$ type created particle from $t=\tau$ to $t=\tau'$ is

$$N_{\vec{k}s}(\tau,\tau') = {}_{\tau}\langle 0|a^{\dagger}_{\vec{k}s}(\tau')a_{\vec{k}s}(\tau')|0\rangle_{\tau}$$
$$= -\sum_{r=0}^{3}\eta_{rr}\left|\left(\overset{(\tau')}{f}{}^{\mu}_{\vec{k}s};\overset{(\tau)}{f}{}^{*}_{-\vec{k}r\mu}\right)\right|^2. \quad (5.22)$$

From (2.22) the mean value of the created particle density is

$$N(\tau,\tau') = \int d^3k\sum_{s=0}^{3}\eta_{ss}N_{\vec{k}s}(\tau,\tau'). \quad (5.23)$$

6. THE UNPHYSICAL PHOTON CREATION

We will prove now that the created timelike and longitudinal photons are not mutually canceled for all τ, i.e.,

$$N_{\vec{k}3}(\tau,\tau') - N_{\vec{k}0}(\tau,\tau') \neq 0.$$

Using Eq. (2.19) we obtain

$$\nabla\cdot A = \int d^3k\left[\frac{ik}{a}\left(a_{\vec{k}3} - a_{\vec{k}0}\right)f_{\vec{k}} - \frac{ik}{a}\left(a^{\dagger}_{-\vec{k}3} - a^{\dagger}_{-\vec{k}0}\right)f^{*}_{-\vec{k}}\right]e^{-i\vec{k}\vec{x}}, \quad (6.1)$$

where the function $f_{\vec{k}}(t)$ verifies the equations (see Ref. 4)

$$\ddot{f}_{\vec{k}} + 3H\dot{f}_{\vec{k}} + \frac{k^2}{a^2}f_{\vec{k}} = 0, \quad (6.2)$$

$$\left(f_{\vec{k}};f_{\vec{k}'}\right) = \frac{(2\pi a)^3}{i}\left[\nabla_0 f^{*}_{\vec{k}}f_{\vec{k}'} - f^{*}_{\vec{k}}\nabla_0 f_{\vec{k}'}\right]$$
$$= \delta(\vec{k}-\vec{k}'), \quad (6.3a)$$

and

$$\left(f_{\vec{k}};\overset{*}{f}_{-\vec{k}'}\right) = 0. \quad (6.3b)$$

By expanding the scalar field $\nabla\cdot A$, according to Eq. (6.1), in two bases with initial conditions on $t=\tau$ and $t=\tau'$, respectively, we obtain

$$b_{\vec{k}}(\tau')f^{(\tau')}_{\vec{k}}(t) + b^{\dagger}_{-\vec{k}}(\tau')\overset{*}{f}{}^{(\tau')}_{-\vec{k}}(t) = b_{\vec{k}}(\tau)f^{(\tau)}_{\vec{k}}(t) + b^{\dagger}_{-\vec{k}}(\tau)\times\overset{*}{f}{}^{(\tau)}_{-\vec{k}}(t), \quad (6.4)$$

where we have set

$$b_{\vec{k}} = (a_{\vec{k}3} - a_{\vec{k}0})i\frac{k}{a},$$
$$b^{\dagger}_{\vec{k}} = (a^{\dagger}_{\vec{k}3} - a^{\dagger}_{\vec{k}0})\left(-i\frac{k}{a}\right). \quad (6.5)$$

Using Eq. (6.3), we find from Eq. (6.4)

$$b_{\vec{k}}(\tau') = \left(f^{(\tau')}_{\vec{k}}; f^{(\tau)}_{\vec{k}}\right) b_{\vec{k}}(\tau) + \left(f^{(\tau')}_{\vec{k}}; f^{(\tau)*}_{-\vec{k}}\right) b^{\dagger}_{-\vec{k}}(\tau), \qquad (6.6a)$$

$$b^{\dagger}_{\vec{k}}(\tau') = \left(f^{(\tau')}_{\vec{k}}; f^{(\tau)}_{\vec{k}}\right)^* b^{\dagger}_{\vec{k}}(\tau) + \left(f^{(\tau')}_{\vec{k}}; f^{(\tau)*}_{-\vec{k}}\right)^* b_{-\vec{k}}(\tau). \qquad (6.6b)$$

Taking into account the commutation relations (2.16) for $a_{\vec{k}s}$ and $a^{\dagger}_{\vec{k}s}$, the subsidiary conditions (2.5) on $t=\tau$ and Eqs. (5.21), (6.5), and (6.6), it is easy to prove that

$$\left[b_{\vec{k}}(\tau), b^{\dagger}_{\vec{k}'}(\tau')\right] = 0, \qquad (6.7)$$

$$_{\tau}\langle \text{phys}|b^{\dagger}_{\vec{k}}(\tau') b_{\vec{k}}(\tau')|\text{phys}\rangle_{\tau} = 0, \qquad (6.8)$$

$$_{\tau}\langle \text{phys}|\nabla A(\tau')|\text{phys}\rangle_{\tau} = 0, \qquad (6.9)$$

$$\left[b_{\vec{k}}(\tau), a_{\vec{k}'s}(\tau')\right] = \eta_{ss}\delta(\vec{k}+\vec{k}')\frac{ik}{a(\tau)}\left[\left(\overset{(\tau')}{f^{\mu}_{\vec{k}'s}}; \overset{(\tau)}{f^{*}_{-\vec{k}'0\mu}}\right) + \left(\overset{(\tau')}{f^{\mu}_{\vec{k}'s}}; \overset{(\tau)}{f^{*}_{-\vec{k}'3\mu}}\right)\right], \qquad (6.10a)$$

$$\left[b_{\vec{k}}(\tau), a^{\dagger}_{\vec{k}'s}(\tau')\right] = \eta_{ss}\delta(\vec{k}-\vec{k}')\frac{ik}{a(\tau)}\left[\left(\overset{(\tau')}{f^{\mu}_{\vec{k}'s}}; \overset{(\tau)}{f_{\vec{k}'0\mu}}\right)^* + \left(\overset{(\tau')}{f^{\mu}_{\vec{k}'s}}; \overset{(\tau)}{f_{\vec{k}'3\mu}}\right)^*\right]. \qquad (6.10b)$$

From Eq. (6.5) it follows that

$$\frac{k^2}{a^2}\left[a^{\dagger}_{\vec{k}3}a_{\vec{k}3} - a^{\dagger}_{\vec{k}0}a_{\vec{k}0}\right] = \frac{ik}{a}\left[b^{\dagger}_{\vec{k}}a_{\vec{k}0} - a^{\dagger}_{\vec{k}0}b_{\vec{k}}\right] + b^{\dagger}_{\vec{k}}b_{\vec{k}}. \qquad (6.11)$$

If $|0\rangle_{\tau}$ is the vacuum state in $t=\tau$, defined by $a_{\vec{k}s}(\tau)|0\rangle_{\tau} = 0$, then

$$_{\tau}\langle 0|a^{\dagger}_{\vec{k}3}(\tau')a_{\vec{k}3}(\tau') - a^{\dagger}_{\vec{k}0}(\tau')a_{\vec{k}0}(\tau')|0\rangle_{\tau}\frac{k^2}{a^2(\tau')}$$
$$= \frac{ik}{a(\tau')}\,{}_{\tau}\langle 0|b^{\dagger}_{\vec{k}}(\tau')a_{\vec{k}0}(\tau') - a^{\dagger}_{\vec{k}0}(\tau')b_{\vec{k}}(\tau')|0\rangle_{\tau}$$
$$= -\frac{k^2}{a(\tau)a(\tau')}\left(\overset{(\tau')}{f_{\vec{k}}}; \overset{(\tau)}{f^{*}_{-\vec{k}}}\right)\left(\overset{(\tau')}{f^{\mu}_{\vec{k}0}}; \overset{(\tau)}{f^{*}_{-\vec{k}3\mu}} + \overset{(\tau)}{f^{*}_{-\vec{k}0\mu}}\right)^* + \text{C.C.} \qquad (6.12)$$

According to Eq. (4.6), (4.19), and (4.23) we have

$$\left(\overset{(\tau')}{f^{\mu}_{\vec{k}0}}; \overset{(\tau)}{f^{*}_{-\vec{k}0\mu}} + \overset{(\tau)}{f^{*}_{-\vec{k}3\mu}}\right) = -\frac{a(\tau)a(\tau')}{k^2}\left(\overset{(\tau')}{f_{\vec{k}}}; \overset{(\tau)}{f^{*}_{-\vec{k}}}\right)_{t=\tau'},$$

and by using Eq. (6.2) we obtain

$$\left(\overset{(\tau')}{f_{\vec{k}}}; \overset{(\tau)}{f^{*}_{-\vec{k}}}\right)_{t=\tau'} = \frac{k^2}{a^2(\tau')}\left(\overset{(\tau')}{f_{\vec{k}}}; \overset{(\tau)}{f^{*}_{-\vec{k}}}\right). \qquad (6.13)$$

Hence the mean value of the density of created unphysical photons at $t=\tau'$ is[15]

$$\left\langle N_{\vec{k}3} - N_{\vec{k}0}\right\rangle_{t=\tau'} = {}_{\tau}\langle 0|a^{\dagger}_{\vec{k}3}(\tau')a_{\vec{k}3}(\tau') - a^{\dagger}_{\vec{k}0}(\tau')a_{\vec{k}0}(\tau')|0\rangle_{\tau}$$
$$= 2\left|\left(\overset{(\tau')}{f_{\vec{k}}}; \overset{(\tau)}{f^{*}_{-\vec{k}}}\right)\right|^2. \qquad (6.14)$$

We remark that the right-hand side of (6.14) is just twice as large as the limit for $m\to 0$ of the density of created particles for a massive scalar field (see Ref. 5.). As was proved in that paper, such a creation (at least for $k\to\infty$) does not depend on m. Consequently,

$$\left\langle N_{\vec{k}3} - N_{\vec{k}0}\right\rangle_{t=\tau'} \neq 0.$$

Finally, it can be proved that

$$N_{\vec{k}1} = N_{\vec{k}2} > 0.$$

CONCLUSION

We have shown that, independently of the chosen Cauchy data for the bivectorial kernel $G_1^{\mu\nu'}(x,x')$, an unphysical photon creation takes place and it is connected with the particle creation for a massive scalar field in the limit $m\to 0$, as can be seen from Eq. (6.14).

The chosen Cauchy data for the biscalar kernel $G_1(x,x')$ [see Eq. (3.1)], which are the simplest that allow one to formulate a model for the massive scalar field where the density of the created particle is finite, lead to a mass-independent particle creation (at least for $k\to\infty$). Hence, the creation of unphysical photons is non vanishing. To avoid this difficulty it seems to be necessary to choose other Cauchy data for $G_1(x,x',m)$ such that the scalar particle creation vanishes when $m\to 0$.

[1]H. Ceccatto, A. Foussats, H. Giacomini, and O. Zandron, "Quantum Field Theory in curved space-time. Massive and massless vector fields," Phys. Rev. D **24**, 2576 (1981).

[2]A. Foussats, R. Laura, and O. Zandron, J. Math. Phys. **22**, 357 (1981).

[3]A. Lichnerowicz, Inst. Haut. Et. Sci. Publ. Math. No. 10 p. 293 (1961); Ann. Inst. Henri Poincaré **3**, 223 (1964); Bull. Soc. Math. Fr. **92**, 11 (1964); A Lichnerowicz, in *Conférence Internationale sur les Théories relativistes de la gravitation, Warszawa et Jablona* (Gauthier-Villas, Paris, 1964), p. 177.

[4]H. Ceccatto, A. Foussats, H. Giacomini, and O. Zandron, "The kernel $G_1(x,x')$ and the quantum equivalence principle," Nuovo Cimento B **65**, 94 (1981).

[5]H. Ceccatto, A. Foussats, H. Giacomini, and O. Zandron, "Finite creation of particles in an expanding universe," Nuovo Cimento B **65**, 233 (1981).

[6]M. Castagnino and R. Weder, J. Math. Phys. **22**, 1 (1981).

[7]M. Castagnino, A. Foussats, R. Laura, and O. Zandron, Nuovo Cimento A **60**, 138 (1980).

[8]L. Parker, "The creation of particles in an expanding universe," Thesis, Harvard University, 1966 (unpublished). L. Parker, Phys. Rev. Lett. **21**, 562 (1968); Phys. Rev. **183**, 1057 (1969); Phys. Rev. D **3**, 346 (1971).

[9]The scalar product $\langle\text{phys}|\ |\text{phys}\rangle_G$ is the Gupta scalar product defined by $\langle\text{phys}|A|\text{phys}\rangle_G = \langle\text{phys}|\hat{\xi}A|\text{phys}\rangle$, where the operator $\hat{\xi}$ verifies $\hat{\xi}A_i = A_i\hat{\xi}$, $i=1,2,3$, $\hat{\xi}A_0 = -A_0\hat{\xi}$, $\hat{\xi}^2 = 1$, $\hat{\xi}|0\rangle = |0\rangle$.

[10]We use $\eta_{ss'} = 0$ if $s\neq s'$; $\eta_{00} = 1$, $\eta_{ii} = -1$, $i=1,2,3$.

[11]Equations (2.20) correspond to the transversality conditions: $k_\mu\epsilon^{\mu}_{(1)} = k_\mu\epsilon^{\mu}_{(2)} = 0$, $k_\mu\epsilon^{\mu}_{(3)} = -|\vec{k}|$, and $k_\mu\epsilon^{\mu}_{(0)} = |\vec{k}|$ of the flat space-time.

[12]The mixing data $\nabla_0\nabla_{0'}G_1^{(\Sigma)}(xx')|_{\Sigma}$ remain determined by knowing the data (3.1), because an analogous relation to (2.11d) for the kernels $G(x,x')$ and $G_1(x,x')$ exists (see Ref. 4).

[13]We note that the system (2.13), (2.19), and (4.2) does not determine univocally the initial conditions for the base. This fact is connected with the invariance of (2.18) under particular base transformation (see Ref. 4).

[14]We define $\vec{k}\cdot\vec{x} = (k_1x^1 + k_2x^2 + k_3x^3)$ and $k^2 = (k_1^2 + k_2^2 + k_3^2)$. In the flat space-time the function $f^{\mu}_{\vec{k}s} \sim \epsilon^{\mu}_{\vec{k}s}e^{-ikt}$, where $\epsilon^{\mu}_{\vec{k}s}$ are the polarization vectors.

[15]Let us note that the result (6.14) is a consequence of the assumption made in Sec. 3, that is, to consider a different $\overset{(\Sigma)}{G}{}_1^{\mu\nu'}(x,x')$ on each hypersurface Σ of V_4. On the other hand, the result (2.23) was obtaining by supposing the existence of a unique kernel $G_1^{\mu\nu'}(x,x')$ for all V_4.

On mapping approaches in axiomatic quantum field theory

Te Hai Yao
Mathematics Department, Hong Kong University, Hong Kong

(Received 3 December 1980; accepted for publication 24 July 1981)

We here state a collection of results in axiomatic quantum field theory obtained under the general philosophy of "mapping approaches." It is hoped that these results will stimulate further investigations along this direction, especially in connection with the problem of the existence and the construction of nontrivial four-dimensional quantum field theories.

PACS numbers: 11.10.Cd

I. INTRODUCTION

This paper presents some substantial results in axiomatic quantum field theory[1,2] under the general philosophy of mapping approaches. The precise formulation of various mapping approaches studied here will be clear from theorems to be stated. From a general point of view, a mapping approach, in the restricted sense studied here, consists of a linear map V from some function space $Q(\mathbb{R}^d)$ on $\mathbb{R}^d$ $(d \geqslant 2)$ to some function space $Q'(\mathbb{R}^{d'})$ on $\mathbb{R}^{d'}$ $(d' \geqslant 2)$ and of studying the properties of the set $\{K_n, n = 1,2,\cdots\}$ of multilinear functionals K_n $(f_1, f_2, \ldots, f_n)$ over $Q(\mathbb{R}^d) \times Q(\mathbb{R}^d) \times \cdots \times Q(\mathbb{R}^d)$ (n copies), defined by

$$K_n(f_1, f_2, \ldots, f_n) = K'_n(Vf_1, Vf_2, \ldots, Vf_n),$$
$$f_i \in Q(\mathbb{R}^d), \quad Vf_i \in Q'(\mathbb{R}^{d'}), \quad i = 1,2,\ldots,n$$

where K'_n is a multilinear functional over $Q'(\mathbb{R}^{d'}) \times Q'(\mathbb{R}^{d'}) \times \cdots \times Q'(\mathbb{R}^{d'})$ (n copies) for each $n = 1,2,\cdots$, with the set $\{K'_n, n = 1,2,\cdots\}$ possessing some additional structures. $\{K'_n, n = 1,2,\cdots\}$ might be a set of Wightman distributions,[1,2] in which case $Q'(\mathbb{R}^{d'})$ is identified with $\mathscr{S}(\mathbb{R}^{d'})$, the Schwartz space on $\mathbb{R}^{d'}$, or it might be a set of expectations of some random variables occurring in some random field[3–5] possessing some specific properties, in which case $Q'(\mathbb{R}^{d'})$ might be identified with $\mathscr{S}(\mathbb{R}^{d'})$, $\mathscr{D}(\mathbb{R}^{d'})$ (the subspace of $\mathscr{S}(\mathbb{R}^{d'})$ consisting of functions of compact support), or the Sobolev space $\mathscr{H}^{-1}(\mathbb{R}^{d'})$ on $\mathbb{R}^{d'}$.

The results we have obtained include, among others, an improvement of Nelson's method[3–5] of constructing d-dimensional quantum field theories $(d \geqslant 2)$, and we hope that further studies of various mapping approaches would lead to further developments in solving the problem of the existence and the construction of nontrivial four-dimensional quantum field theories.

We now present in the following an overall view of the results in this paper.

On Sec. II

In this section we present a mapping method of constructing quantum field theories of a lower space dimension from any given quantum field theory of a higher space dimension. We obtain generalized free fields from a free field, By means of a further limiting procedure, we can obtain a free field from a free field, with the masses equal.

On Sec. III

In this section the following problem is studied: Given a random field ϕ over[3] $\mathscr{H}^{-1}(\mathbb{R}^d)$ $(d = 3$ or $4)$ on some probability space (Ω, Σ, μ) satisfying certain properties, what can be said about the multilinear functional Γ_n over $\mathscr{S}(\mathbb{R}^4) \times \mathscr{S}(\mathbb{R}^4) \times \cdots \times \mathscr{S}(\mathbb{R}^4)$ (n copies) defined by

$$\Gamma_n(f_1, f_2, \ldots, f_n) = E\{\phi(h_1) \times \phi(h_2) \times \cdots \times \phi(h_n)\}$$
$$f_i \in \mathscr{S}(\mathbb{R}^4), \quad h_i \in \mathscr{H}^{-1}(\mathbb{R}^4), \quad h_i = Mf_i, \quad i = 1,2,\ldots,n,$$

where E denotes expectation and M is some real continuous linear map from $\mathscr{S}(\mathbb{R}^4)$ to $\mathscr{H}^{-1}(\mathbb{R}^d)$?

For $d = 4$, we find that if M satisfies certain simple properties then the set $\{\mathbf{\Gamma}_n, n = 1,2,\cdots\}$ defines the Euclidean Green's functions[6–8] of a four-dimensional quantum field theory if each $\mathbf{\Gamma}_n$ is Euclidean invariant, where $\mathbf{\Gamma}_n$ is the unique continuous linear functional on $\mathscr{S}(\mathbb{R}^{4n})$ obtained from Γ_n according to the Schwartz kernel theorem. This result is an improvement of Nelson's work,[5] where the map M is restricted to be the identity map (Nelson treated the case $Q' = \mathscr{S}$ is Ref. 5). (See *Note* at the end of subsection IIIB1.) Our results for $d = 4$ introduce an extra degree of freedom into Nelson's program.

The case $d = 3$ is concened with the deep problem of generating four-dimensional quantum field theories from three-dimensional probabilistic structures (including three-dimensional Euclidean Markov fields[3–5]) via some mapping approach.

On Sec. IV

In this section a theorem on the generation of quantum field theories from certain four-dimensional Euclidean invariant probabilistic structures over $\mathscr{H}^{-1}(\mathbb{R}^4)$ [including Euclidean–Markov fields[3,4] over $\mathscr{H}^{-1}(\mathbb{R}^4)$] is presented as a simple application of a theorem in Sec. III.

A remark

Theorems 1 and 2 in Sec. III and the theorem in Sec. IV can be generalized to any dimension d satisfying $d \geqslant 2$.

II. FROM HIGHER DIMENSIONAL QUANTUM FIELD THEORIES TO LOWER DIMENSIONAL ONES

A. The theorem and sketch of proof

We first state the central theorem in this section and then we give a sketch of the proof.

Theorem: suppose we are given a set of Hermitian scalar

 0022-2488/82/101873-08$02.50

Wightman destributions $\{\mathscr{W}_n^{M+N+1},\ n=1,2,\cdots\}$ in $(M+N+1)$-dimensional space–time ($M>0$, $N>0$, the time dimension is always one and is the first dimension), and suppose we define a distribution $\mathscr{W}_n^{M+1}$ for each $n=1,2,\cdots$, by

$$\mathscr{W}_n^{M+1}(f_1,f_2,\ldots,f_n) = \mathscr{W}_n^{M+N+1}(f_1\otimes g, f_2\otimes g\ldots,f_n\otimes g),$$

where

$$f_i\in\mathscr{S}(\mathbb{R}^{M+1}),\quad i=1,2,\ldots,n,$$

$$g\in\mathscr{S}(\mathbb{R}^N) \text{ and real.}$$

Then the set $\{\mathscr{W}_n^{M+1},\ n=1,2,\cdots\}$ is a set of Hermitian scalar Wightman distributions in $(M+1)$-dimensional space–time.

Sketch of proof of theorem: It is easy to show that each $\mathscr{W}_n^{M+1}$ defines a continuous linear functional over $\mathscr{S}(\mathbb{R}^{n(M+1)})$. We have also

(i) Proof of relativistic invariance: obvious.

(ii) Proof of local commutativity: Let $x=(x_0,x_1,\ldots,x_M)$ and $y=(y_0,y_1,\ldots y_M)$ be spacelike separated vectors in $(M+1)$-dimensional space–time; then $\underset{\sim}{x}=(x_0,x_1\ldots x_M,x_{M+1},\ldots,x_{M+N})$ and $\underset{\sim}{y}=(y_0,y_1,\ldots y_M,y_{M+1},\ldots,y_{M+N})$ are spacelike separated vectors in $(M+N+1)$-dimensional space–time, for any $x_i\in\mathbb{R},y_i\in\mathbb{R},\ i=M+1,M+2,\ldots,M+N$.

(iii) Proof of positive definiteness: We use the reality of g and the expansion of any $F_j\in\mathscr{S}(\mathbb{R}^{j(M+1)})$ as a linear combination of $h_{i_1}\otimes h_{i_2}\otimes\cdots\otimes h_{i_j}$ [in general, an infinite linear combination, and convergence is a convergence in the Schwartz topology of $\mathscr{S}(\mathbb{R}^{j(M+1)})$], where $\{h_k,\ k=0,1,2,\cdots\}$ forms a complete linearly independent basis of $\mathscr{S}(\mathbb{R}^{(M+1)})$.

(iv) Proof of hermiticity: obvious.

(v) Proof of spectral condition: We have

$$\mathscr{W}_n^{M+1}(u_1,u_2,\ldots,u_n) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} dv_1\,dv_2\cdots dv_n \times\mathscr{W}_n^{M+N+1}(u_1,v_1;u_2,v_2;\ldots u_n,v_n) \times g(v_1)\times g(v_2)\times\cdots\times g(v_n),$$

where $u_i\in\mathbb{R}^{M+1},\ v_i\exists\mathbb{R}^N,\ i=1,2,\ldots,n$; we then consider the Fourier transforms $(W_n^{M+1})^\sim$ of the diffenence variable distributions $W_n^{M+1}(\xi_1,\xi_2,\ldots\xi_{n-1})=\mathscr{W}_n^{M+1}(u_1,u_2,\ldots,u_n)$ $n=1,2,\ldots$, with $\xi_j=u_j-u_{j+1}$, $j=1,2,\ldots,n-1$, and we can then show that the joint energy–momentum spectrum of the lower dimensional structure is contained in the projection of the joint energy–momentum spectrum of the higher dimensional theory onto the zero hyperplane of the extra momentum variables. (*Note*: we use ˜ to denote Fourier Transform throughout.)

(vi) Proof of cluster decomposition property: let η be a spacelike vector in $(M+1)$-dimensional space–time; then $(\eta,0,0,\ldots0)$ with N zeroes following η is a spacelike vector in $M+N+1$-dimensional space–time.

B. An example, a proposition, and a remark

(I) *Example*: It can be easily shown that a Hermitian scalar free field in $(M+N+1)$-dimensional space–time gives a Hermitian scalar generalized free field in $(M+1)$-dimensional space–time.

(II) We now state the following important proposition, the proof of which is obvious.

Proposition: Let $\{\mathscr{W}_{n,(m)}^{M+N+1,(F)},\ n=1,2,\cdots\}$ be the set of Wightman distributions corresponding to the Hermitian scalar free field of mass m in $(M+N+1)$-dimensional space–time and we define, for each $n=1,2,\cdots$:

$$\mathscr{W}_n^{M+1,k,c}(f_1,f_2,\ldots,f_n) = \mathscr{W}_{n,(m)}^{M+N+1,(F)}(f_1\otimes g_k^{(c)},f_2\otimes g_k^{(c)},\ldots,f_n\otimes g_k^{(c)}),$$

$$f_i\in\mathscr{S}(\mathbb{R}^{(M+1)}),\quad i=1,2,\ldots,n,$$

$$g_k^{(c)}\in\mathscr{S}(\mathbb{R}^N),\quad g_k^{(c)}\text{real},\quad c>0,k=1,2,\cdots,$$

where $g_k^{(c)}$ are chosen such that $(g_k^{(c)})^\sim$ are real and satisfy

$$[(g_k^{(c)})^\sim(p)]^2 \underset{k\to\infty}{\to} c\times\delta(p),\quad p\in\mathbb{R}^N.$$

If we let

$$\mathscr{W}_n^{M+1,\infty,c}(f_1,f_2,\ldots,f_n) = \lim_{k\to\infty}\mathscr{W}_n^{M+1,k,c}(f_1,f_2,\ldots,f_n),$$

then $\{\mathscr{W}_n^{M+1,\infty,c},\ n=1,2,\cdots\}$ defines the Hermitian scalar free field of mass m in $(M+1)$-dimensional space–time for some suitable c.

Thus we can, by the above mapping method and a suitable limiting procedure, arrive at a free field of a lower space dimension from a free field of a higher space dimension, the masses being equal.

(III) *A Remark*: It is a natural question to ask whether, given an arbitrary higher-dimensional field theory, there exists a lower-dimensional field theory such that the joint energy–momentum spectrum of the lower-dimensional field theory is the same as the restriction of the joint energy–momentum spectrum of the higher-dimensional field theory to the zero hyperplane of the extra momentum variables. The problem is open. We hope that the above mapping method and a suitable limiting procedure, or some variations, might be of help in studying this problem.

III. ON THE CONSTRUCTION OF FOUR-DIMENSIONAL QUANTUM FIELD THEORIES FROM PROBABILISTIC STRUCTURES

A. Some definitions

1. Definition of a d-dimensional probabilistic X-structure (ϕ, (Ω, Σ, μ)) over $\mathscr{H}^{-1}(\mathbb{R}^d)$, for d>2

A d-dimensional probabilistic X-structure over $\mathscr{H}^{-1}(\mathbb{R}^d)$ is a real random field ϕ over $\mathscr{H}^{-1}(\mathbb{R}^d)$ on some probability space (Ω,Σ,μ) such that:

(i) $\phi(g)\in L^p(\Omega,\Sigma,\mu)$ for $g\in\mathscr{H}^{-1}(\mathbb{R}^d)$, $p=1,2,\cdots$, and such that $E\{\phi(g_1)\times\phi(g_2)\times\cdots\times\phi(g_n)\}$ is separately continuous in each g_i, for $i=1,2,\ldots,n$, $n=1,2,\cdots$;

(ii) There is reflection covariance in the x_0 variable in connection with a representation of the reflection group of the x_0 variable on (Ω,Σ,μ), where x_0 is the first coordinate of the point $(x_0,x_1,\ldots,x_{d-1})$ in $\mathbb{R}^d$;

(iii) The following "weak Markov property" holds:

$$E\{u|\,\Sigma(\Lambda^c)\} = E\{u|\,\Sigma(\partial\Lambda)\},$$

where u is an integrable random variable belonging to $\Sigma(\Lambda)$, with

$$\Lambda = \{x_0 > 0,\ (x_1, x_2, \ldots, x_{d-1}) \in \mathbb{R}^{d-1}\},$$
$$\Lambda^c = \{x_0 < 0,\ (x_1, x_2, \ldots, x_{d-1}) \in \mathbb{R}^{d-1}\},$$
$$\partial\Lambda = \{x_0 = 0,\ (x_1, x_2, \ldots, x_{d-1}) \in \mathbb{R}^{d-1}\}.$$

(See *Notes* at the end of this subsection for explanations on notations and terminology in the above definition.)

2. Definition of a map M_0 from $\mathscr{S}(\mathbb{R}^4)$ to $\mathscr{H}^{-1}(\mathbb{R}^3)$

(a) Some preliminaries: Let $f \in \mathscr{S}(\mathbb{R}^4)$. The $\tilde{f} \in \mathscr{S}(\mathbb{R}^4)$. We can expand [$L^2(\mathbb{R}^4)$ expansion]:

$$\tilde{f}(p) = \sum_{\substack{i,\nu,j \\ j \text{ even}}} C^{(1)}_{i\nu j} \phi_i(p_0) \frac{\phi_j(|\mathbf{p}|)}{|\mathbf{p}|} Y_\nu^{(1)}(\theta, \varphi) + \sum_{\substack{i,\nu,j \\ j \text{ even}}} C^{(2)}_{i\nu j} \phi_i(p_0) \frac{\phi_j(|\mathbf{p}|)}{|\mathbf{p}|} Y_\nu^{(2)}(\theta, \varphi)$$

where $p = (p_0, p_1, p_2, p_3)$ and is conjugate to (x_0, x_1, x_2, x_3), (x_0, x_1, x_2, x_3) being a point in $\mathbb{R}^4$ on which f is defined, $\mathbf{p} = (p_1, p_2, p_3)$, $|\mathbf{p}| = +(p_1^2 + p_2^2 + p_3^2)^{1/2}$, $\mathbf{p} = (|\mathbf{p}|, \theta, \varphi)$ in spherical polar coordinates, and

$$Y_\nu^{(1)}(\theta, \varphi) = d^{(1)}_{lm} P_l^m(\cos\theta)\cos m\phi,$$
$$Y_\nu^{(2)}(\theta, \varphi) = d^{(2)}_{lm} P_l^m(\cos\theta)\sin m\phi,$$

ν being an indexing of (l, m) $(l = 0,1,2,\cdots;$ $m = -l, -l+1, \ldots, l-1, l)$ such that ν is even if $(l+m)$ is even and ν is odd if $(l+m)$ is odd $(\nu = 0,1,2,\cdots)$, $d^{(1)}_{lm}$ and $d^{(2)}_{lm}$ being normalization constants and $\{\phi_k,\ k = 0,1,2,\cdots\}$ being the set of Hermite functions. We may choose $d^{(1)}_{lm}$ and $d^{(2)}_{lm}$ such that

$$\iint Y_{\nu_1}^{(1)}(\theta, \varphi) Y_{\nu_2}^{(1)}(\theta, \varphi) d\cos\theta\, d\phi = \delta_{\nu_1, \nu_2},$$
$$\iint Y_{\nu_1}^{(2)}(\theta, \phi) Y_{\nu_2}^{(2)}(\theta, \phi) d\cos\theta\, d\phi = \delta_{\nu_1, \nu_2}.$$

We also have

$$\iint Y_{\nu_1}^{(1)}(\theta, \varphi) Y_{\nu_2}^{(2)}(\theta, \varphi) d\cos\theta\, d\phi = 0, \quad \text{for all } \nu_1 \text{ and } \nu_2.$$

(b) The map M_0: The map M_0 from $\mathscr{S}(\mathbb{R}^4)$ to $\mathscr{H}^{-1}(\mathbb{R}^3)$ is now defined by

$$\tilde{h}(\underline{p}) = \sum_{\substack{i,\nu,j \\ j \text{ even}}} C^{(1)}_{i\nu j} \phi_i(p_0) \frac{\phi_j(|\underline{\mathbf{p}}|)}{|\underline{\mathbf{p}}|^{1/2}} \frac{\cos\nu\phi}{(2\pi)^{1/2}} + \sum_{\substack{i,\nu,j \\ j \text{ even}}} C^{(2)}_{i\nu j} \phi_i(p_0) \frac{\phi_j(|\underline{\mathbf{p}}|)}{|\underline{\mathbf{p}}|^{1/2}} \frac{\sin\nu\phi}{(2\pi)^{1/2}}$$

$([\mathscr{H}^{-1}(\mathbb{R}^3)]^\sim$ convergence)

where $h = M_0 f \in \mathscr{H}^{-1}(\mathbb{R}^3)$, given $f \in \mathscr{S}(\mathbb{R}^4)$ with the above expansion for $\tilde{f}$, and where $\underline{p} = (p_0, p_1, p_2)$, $\underline{\mathbf{p}} = (p_1, p_2)$, $|\underline{\mathbf{p}}| = +\sqrt{p_1^2 + p_2^2}$, and $\underline{\mathbf{p}} = (|\underline{\mathbf{p}}|, \phi)$ in polar coordinates.

Notes on (1): (i) A random field ϕ over $\mathscr{H}^{-1}(\mathbb{R}^d)$ is a stochastic process indexed by $\mathscr{H}^{-1}(\mathbb{R}^d)$, which is linear and such that if $g_\alpha \to g$ in $\mathscr{H}^{-1}(\mathbb{R}^d)$, then $\phi(g_\alpha) \mapsto \phi(g)$ in measure.

(ii) (a) A representation of the reflection group of the x_0 variable, denoted by G, is a homomorphism $\eta \to T_\eta$ of G into the group of measure-preserving transformations on the probability space (Ω, Σ, μ) on which the process is defined.

(b) Reflection covariance in the x_0 variable means:

$$T_\eta\{\phi(g)\} = \varphi(g \circ \eta^{-1}), \quad \eta \in G.$$

(iii) If $B \subset \mathbb{R}^d$, then $\Sigma(B)$ denotes the σ-algebra generated by the $\phi(g)$ with g in $\mathscr{H}^{-1}(\mathbb{R}^d)$ and supp $g \subset B$. $\Sigma(B)$ also denotes the set of all random variables which are measurable with respect to the σ-algebra $\Sigma(B)$.

(iv) $E\{\cdot | \Sigma(B)\}$ denoted conditional expectation with respect to $\Sigma(B)$.

B. Statement of theorems

1. Construction of four-dimensional quantum field theories from four-dimensional probabilistic structures

Theorem 1: (a) Let $(\phi, (\Omega, \Sigma, \mu))$ be a four-dimensional probabilistic X-structure over $\mathscr{H}^{-1}(\mathbb{R}^4)$. Let also $f \in \mathscr{S}(\mathbb{R}^4)$, and $f \xrightarrow{M} h = Mf$, $h \in \mathscr{H}^{-1}(\mathbb{R}^4)$, be a real continuous linear map from $\mathscr{S}(\mathbb{R}^4)$ to $\mathscr{H}^{-1}(\mathbb{R}^4)$ satisfying the following assumptions:

(1) Support of $f \subset \mathbb{R}^4_>$, implies that
support of $h \subset \mathbb{R}^4_>$,
$\mathbb{R}^4_> = \mathbb{R}_> \times \mathbb{R}^3$,
$\mathbb{R}_> = [0, \infty)$.

(2) Θf is mapped under M to Θh, for f with support of $f \subset \mathbb{R}^4_>$, where

$$(\Theta f)(x_0, x_1, x_2, x_3) = f(-x_0, x_1, x_2, x_3),$$
$$(\Theta h)(x_0, x_1, x_2, x_3) = h(-x_0, x_1, x_2, x_3).$$

(3) $\Gamma_n(f_1, f_2, \ldots, f_n)$, for each $n = 1,2,\cdots$, is invariant under the simultaneous action of any element of ISO(4) on all f_i, $i = 1,2,\ldots n$, where Γ_n is the multilinear functional on $\mathscr{S}(\mathbb{R}^4) \times \mathscr{S}(\mathbb{R}^4) \times \cdots \times \mathscr{S}(\mathbb{R}^4)$ (n copies) defined by

$$\Gamma_n(f_1, f_2 \cdots f_n) = E\{\phi(h_1) \times \phi(h_2) \times \cdots \times \phi(h_n)\},$$
$$f_i \in \mathscr{S}(\mathbb{R}^4), \quad h_i \in \mathscr{H}^{-1}(\mathbb{R}^4), \quad h_i = Mf_i \quad i = 1,2,\ldots n.$$

Then, for each n there is a unique tempered distribution $\mathbf{\Gamma}_n$ on $\mathbb{R}^{4n}$ satisfying $\mathbf{\Gamma}_n(f_1 \otimes f_2 \otimes \cdots \otimes f_n) = \Gamma_n(f_1, f_2, \ldots, f_n)$ such that the collection $\{\mathbf{\Gamma}_n, n = 0,1,2,\cdots\}$, with $\mathbf{\Gamma}_0 = 1$, satisfies all the Osterwalder–Schrader axioms[6–8] for Euclidean Green's functions in four-dimesions except cluster decomposition (for a four-dimensional Hermitian scalar quantum field theory).

(b) Further, if the set $\{E\{\phi(g_{n_1}) \times \phi(g_{n_2}) \times \cdots \times \phi(g_{n_n})\},\ n = 1,2,\cdots\}$ satisfies cluster decomposition for arbitrary $g_{n_i} \in \mathscr{H}^{-1}(\mathbb{R}^4)$, $i = 1,2,\ldots,n$, $n = 2,3,\cdots$, then the set $\{\mathbf{\Gamma}_n,\ n = 0,1,2,\cdots\}$ also satisfies cluster decomposition.

Theorem 2: Let N be a continuous linear operator in $L^2(\mathbb{R}^4)$ satisfying

(i) Nk is real for k real, with $k \in L^2(\mathbb{R}^4)$.

(ii) $Nk \in L^2(\mathbb{R}^4_>)$, for $k \in L^2(\mathbb{R}^4_>)$, where $L^2(\mathbb{R}^4_>)$ is the subspace of $L^2(\mathbb{R}^4)$ consisting of square integrable functions on $\mathbb{R}^4$ with support in $\mathbb{R}^4_>$.

(iii) (a) $(Nk)(x)$ is even in x_0 if $k(x)$ is even in x_0,

(b) $(Nk)(x)$ is odd in x_0 if $k(x)$ is odd in x_0, for $k \in L^2(\mathbb{R}^4)$, $x = (x_0, x_1, x_2, x_3) \in \mathbb{R}^4$

[this is equivalent to the statement

$(\Theta N)u = (N\Theta)u$ for $u \in L^2(\mathbb{R}^4)$,
with

$(\Theta v)(x_0,x_1,x_2,x_3) = v(-x_0,x_1,x_2,x_3)$ for $v \in L^2(\mathbb{R}^4)$.]

Then the map $N: f \xrightarrow{N} h = Nf$ for $f \in \mathscr{S}(\mathbb{R}^4) \subset L^2(\mathbb{R}^4)$ is a real continuous linear map from $\mathscr{S}(\mathbb{R}^4)$ to $\mathscr{H}^{-1}(\mathbb{R}^4)$ satisfying assumptions (1) and (2) of Theorem 1.

Note: Theorem 1, supported by Theorem 2, forms an improvement of Nelson's work[5] [Nelson treated the case where $\mathscr{H}^{-1}(\mathbb{R}^4)$ is replaced by $\mathscr{S}(\mathbb{R}^4)$; Theorem 1 still holds with $\mathscr{H}^{-1}(\mathbb{R}^4)$ replaced by $\mathscr{S}(\mathbb{R}^4)$ and with the "reflection property" postulated[5]]. The improvement is that whereas, given a certain probabilistic X-structure $(\phi,(\Omega, \Sigma, \mu))$ over $\mathscr{H}^{-1}(\mathbb{R}^4)$, Euclidean invariance of $E\{\phi(f_{n_1})\times\phi(f_{n_2})\times\cdots\times\phi(f_{n_n})\}$ for arbitrary $f_{n_i} \in \mathscr{S}(\mathbb{R}^4)$, $i = 1,2,\ldots,n$, $n = 1,2,\cdots$, is required in Nelson's work, in our case only Euclidean invariance of $E\{\phi(Mf_{n_1})\times\phi(Mf_{n_2})\times\cdots\times\phi(Mf_{n_n})\}$ for just one map M belonging to a certain wide class is required.

2. Construction of four-dimensional quantum field theories from three-dimensional probabilistic structures

Theorem 3: (a) Let $(\phi,(\Omega, \Sigma, \mu))$ be a three-dimensional probabilistic X-structure over $\mathscr{H}^{-1}(\mathbb{R}^3)$. Let also $f \in \mathscr{S}(\mathbb{R}^4)$, and $f \xrightarrow{M} h = Mf$, $h \in \mathscr{H}^{-1}(\mathbb{R}^3)$, be a real continuous linear map from $\mathscr{S}(\mathbb{R}^4)$ to $\mathscr{H}^{-1}(\mathbb{R}^3)$ satisfying the following assumptions:

(1) support of $f \subset \mathbb{R}^4_>$ implies that support of $h \subset \mathbb{R}^3_>$,
$\mathbb{R}^3_> = \mathbb{R}_> \times \mathbb{R}^2$;

(2) Θf is mapped under M to Θh, for f with support of $f \subset \mathbb{R}^4_>$,
where

$(\Theta f)(x_0,x_1,x_2,x_3) = f(-x_0,x_1,x_2,x_3)$,

$(\Theta h)(x_0,x_1,x_2) = h(-x_0,x_1,x_2)$.

(3) $\Gamma_n(f_1,f_2,\ldots,f_n)$, for each $n = 1,2,\cdots$, is invariant under the simultaneous action of any element of ISO(4) on all f_i, $i = 1,2,\ldots n$, where Γ_n is the multilinear functional on $\mathscr{S}(\mathbb{R}^4)\times\mathscr{S}(\mathbb{R}^4)\times\cdots\times\mathscr{S}(\mathbb{R}^4)$ (n copies) defined by

$$\Gamma_n(f_1,f_2,\ldots,f_n) = E\{\phi(h_1)\times\phi(h_2)\times\cdots\times\phi(h_n)\},$$
$$f_i \in \mathscr{S}(\mathbb{R}^4),\quad h_i \in \mathscr{H}^{-1}(\mathbb{R}^3),\quad h_i = Mf_i \quad i = 1,2,\ldots n.$$

Then, for each n there is a unique tempered distribution $\mathbf{\Gamma}_n$ on $\mathbb{R}^{4n}$ satisfying $\mathbf{\Gamma}_n(f_1 \otimes f_2 \otimes \cdots \otimes f_n) = \Gamma_n(f_1,f_2,\ldots,f_n)$ such that the collection $\{\mathbf{\Gamma}_n, n = 0,1,2,\cdots\}$, with $\mathbf{\Gamma}_0 = 1$, satisfies all the Osterwalder–Schrader axioms[6–8] for Euclidean Green's functions in four-dimesions except cluster decomposition (for a four-dimensional Hermitian scalar quantum field theory).

(b) Further, if the set $\{E\{\phi(g_{n_1})\times\phi(g_{n_2})\times\cdots\times\phi(g_{n_n})\}, n = 1,2,\cdots\}$ satisfies cluster decomposition for arbitrary $g_{n_i} \in \mathscr{H}^{-1}(\mathbb{R}^3)$, $i = 1,2,\ldots,n$, $n = 2,3,\cdots$, then the set $\{\mathbf{\Gamma}_n, n = 0,1,2,\cdots\}$ also satisfies cluster decomposition.

Theorem 4: Let M_0 be defined as in subsection IIIA. Let N be a continuous linear operator in $L_2(\mathbb{R}^3)$ satisfying

(i) Nk is real for k real, with $k \in L^2(\mathbb{R}^3)$.

(ii) $Nk \in L^2(\mathbb{R}^3_>)$, for $k \in L^2(\mathbb{R}^3_>)$, where $L^2(\mathbb{R}^3_>)$ is the subspace of $L^2(\mathbb{R}^3)$ consisting of square integrable functions on $\mathbb{R}^3$ with support in $\mathbb{R}^3_>$.

(iii) (a) $(Nk)(x)$ is even in x_0 if $k(x)$ is even in x_0,

(b) $(Nk)(x)$ is odd in x_0 if $k(x)$ is odd in x_0, for $k \in L^2(\mathbb{R}^3)$, $x = (x_0,x_1,x_2) \in \mathbb{R}^3$

[this is equivalent to the statement

$(\Theta N)u = (N\Theta)u$ for $u \in L^2(\mathbb{R}^3)$,
with

$(\Theta v)(x_0,x_1,x_2) = v(-x_0,x_1,x_2)$ for $v \in L^2(\mathbb{R}^3)$.)

Then the map NM_0 is a real continuous linear map from $\mathscr{S}(\mathbb{R}^4)$ to $\mathscr{H}^{-1}(\mathbb{R}^3)$ satisfying assumptions (1) and (2) of Theorem 3.

Theorem 5: Let M_0 be defined as in subsection IIIA. Further, let $(\phi,(\Omega, \Sigma, \mu))$ be the free Euclidean Markov field over $\mathscr{H}^{-1}(\mathbb{R}^3)$.[4] Then $\{\mathbf{\Gamma}_n, n = 0,1,2,\cdots\}$ is the set of Euclidean Green's functions for the free Hermitian scalar quantum field in one time dimension and three space dimensions, $\mathbf{\Gamma}_n$ being defined in terms of ϕ as in Theorem 3 via the map M_0, $n = 1,2,\cdots$. The masses in these fields are the same.

C. Proof of theorems

Here we present proof of Theorems 1 and 4. Proof of Theorem 2 and 5 is obvious. The proof of Theorem 3 is parallel to that of Theorem 1.

1. Proof of Theorem 1

We proceed to show that the set $\{\mathbf{\Gamma}_n, n = 0,1,2,\cdots\}$ satisfies Osterwalder–Schrader axioms[6–8] for Euclidean Green's functions provided that assumptions (1)–(3) are fulfilled and that cluster decomposition property for the random field ϕ holds.

(*a*) *Proof of distribution property*: The set of tempered distributions $\{\mathbf{\Gamma}_n, n = 1,2,\cdots\}$ defines, by restriction, continuous linear functionals on $\mathscr{S}^{(0)}(\mathbb{R}^{4n})$, $n = 1,2,\cdots$. $\mathbf{\Gamma}_n$ *defines, for each* $n = 2,3,\cdots$, the difference variable distributions S_{n-1} which, by restriction, defines a continuous linear functional $S^{(+)}_{n-1}$ on $\mathscr{S}(\mathbb{R}^{4(n-1)}_+)$, which is also continuous with respect to some $|\ |'_m$ norm. (*Note*: In the above, the notations $\mathscr{S}^{(0)}(\mathbb{R}^{4n})$, $\mathscr{S}(\mathbb{R}^{4(n-1)}_+)$, and $|\ |'_m$ norm follow Ref. 8).

(*b*) *Proof of Euclidean invariance*: by assumption (3).

(*c*) *Proof of positivity*: Let $F_0,F_1,\ldots,F_A$ be given $(A < \infty)$: $F_0 \in \mathbb{C}$, $F_m \in \mathscr{S}(\mathbb{R}^{4m})$ $(1 \leqslant m \leqslant A)$, and

support of $F_m(x^{[1]},x^{[2]},\ldots,x^{[m]})$
$\subset \mathbb{R}^4_> \times \mathbb{R}^4_> \times \cdots \times \mathbb{R}^4_>$ (m copies),
$x^{[i]} \in \mathbb{R}^4$, $i = 1,2,\ldots m$.

Positivity flows from the inequality

$$\sum_{\substack{m=0\\n=0}}^{A} \mathbf{\Gamma}_{m+n}(\Theta \overline{F}_m \otimes F_n) \geqslant 0,$$

$$\text{for any } F_0, F_1,\ldots F_A,\quad A = 0,1,2,\cdots, \tag{Y}$$

which we shall prove, where

$(\Theta F_m(x^{[1]},x^{[2]},\ldots,x^{[m]}) = F_m(\theta x^{[1]},\theta x^{[2]},\ldots,\theta x^{[m]})$,

$\theta x^{[i]} = (-x_0^{[i]},x_1^{[i]},x_2^{[i]},x_3^{[i]})$,

for $x^{[i]} = (x_0^{[i]},x_1^{[i]},x_2^{[i]},x_3^{[i]})$,

$i = 1,2,\ldots m$,

and the overbar denotes complex conjugation.

Now there exist sequences $F_{0j}, F_{1j}, \ldots, F_{Aj}, j = 1,2,\cdots$, such that $F_{0j} \in \mathbb{C}$, $F_{0j} \underset{j\to\infty}{\longrightarrow} F_0$, $F_{mj} \in \mathscr{S}(\mathbb{R}^{4m})$ $(1 \leqslant m \leqslant A)$ with

$$\begin{aligned}&\text{support of } F_{mj}(x^{[1]}, x^{[2]}, \ldots, x^{[m]})\\ &\subset \mathbb{R}^4_> \times \mathbb{R}^4_> \times \cdots \times \mathbb{R}^4_> \quad (m \text{ copies})\\ &\mathbb{R}^4_> = \mathbb{R}_> \times \mathbb{R}^3\\ &\mathbb{R}_> = (0, \infty)\end{aligned}$$

and $F_{mj} \underset{j\to\infty}{\longrightarrow} F_m$ in the Schwartz topology of $\mathscr{S}(\mathbb{R}^{4m})$.

We have

$$\begin{aligned}\sum_{\substack{m=0\\n=0}}^{A} \Gamma_{m+n}(\Theta \overline{F}_m \otimes F_n) &= \sum_{\substack{m=0\\n=0}}^{A} \left\{ \lim_{j\to\infty} \Gamma_{m+n}(\Theta \overline{F}_{mj} \otimes F_{nj}) \right\}\\ &= \lim_{j\to\infty} \sum_{\substack{m=0\\n=0}}^{A} \Gamma_{m+n}(\Theta \overline{F}_{mj} \otimes F_{nj})\\ &\geqslant 0\end{aligned}$$

if we can prove

$$\sum_{\substack{m=0\\n=0}}^{A} \Gamma_{m+n}(\Theta \overline{F}_{mj} \otimes F_{nj}) \geqslant 0, \quad j = 1,2,\cdots, \tag{Y'}$$

where ΘF_{mj} is defined as for ΘF_m.

We now prove statement (Y'): We can write

$$F_{mj}(x^{[1]}, x^{[2]}, \ldots, x^{[m]}) = \sum_{\substack{i_1 i_2 \cdots i_m\\ i_\delta^{(0)} \text{ even}}} \lambda_{j, i_1 i_2 \cdots i_m} \times u_{i_1}(x^{[1]}) \times u_{i_2}(x^{[2]}) \times \cdots \times u_{i_m}(x^{[m]})$$

$[\mathscr{S}(\mathbb{R}^{4m})$ expansion$]$

where

$$\begin{aligned}&u_{i_\delta}(x^{[\delta]}) = e^{-1/(x_0^{[\delta]})^2} \psi_{i_\delta}(x^{[\delta]}) \quad \text{for } x_0^{[\delta]} > 0, \quad \delta = 1,2,\ldots m,\\ &u_{i_\delta}(x^{[\delta]}) = 0 \quad \text{for } x_0^{[\delta]} \leqslant 0, \quad \delta = 1,2,\ldots m\end{aligned}$$

with $\{\psi_{i_\delta}, i_\delta = (i_\delta^{(0)}, i_\delta^{(1)}, i_\delta^{(2)}, i_\delta^{(3)}); i_\delta^{(0)}, i_\delta^{(1)}, i_\delta^{(2)}, i_\delta^{(3)} = 0,1,2,\cdots\}$ being the set of Hermite functions on $\mathbb{R}^4$ [i.e., $\psi_{i_\delta}(x^{[\delta]}) = \phi_{i_\delta^{(0)}}(x_0^{[\delta]}) \phi_{i_\delta^{(1)}}(x_1^{[\delta]}) \times \phi_{i_\delta^{(2)}}(x_2^{[\delta]}) \times \phi_{i_\delta^{(3)}}(x_3^{[\delta]})$ with $\{\phi_k, k = 0,1,2,\cdots\}$ being the set of Hermite functions on $\mathbb{R}$]. We note that we can so arrange things such that only even $i_\delta^{(0)}, \delta = 1,2,\ldots,m$, occur in the summation: This is indicated by the wording $i_\delta^{(0)}$ even. We can show that

$$|\lambda_{j, i_1 i_2 \cdots i_m}| \leqslant \frac{Q_{j,s}}{\prod_{\kappa=0}^{3} \{(1 + i_1^{(\kappa)})^s \times (1 + i_2^{(\kappa)})^s \times \cdots \times (1 + i_m^{(\kappa)})^s\}}$$

for $s = 1,2,\cdots$, where $Q_{j,s}$ depends only on j and s for fixed F_{mj}. This is seen as follows: For any $f \in \mathscr{S}(\mathbb{R})$, the expansion coefficients in the expansion

$$f = \sum_{k=0}^{\infty} C_k \phi_k \quad [\mathscr{S}(\mathbb{R}) \text{ expansion}],$$

where $\{\phi_k, k = 0,1,2,\cdots\}$ is the set of Hermite functions on $\mathbb{R}$, satisfy[9]

$$|C_k| \leqslant Q_s/(1 + k)^s,$$

where Q_s depends only on s for fixed f; here $s = 1,2,\cdots$.

The function $u_l \in \mathscr{S}(\mathbb{R}^4)$ is mapped under M to an element $v_l \in \mathscr{H}^{-1}(\mathbb{R}^4)$. Since M is real, continuous, and linear, we can say that v_l is real and further

$$\|v_l\|_{-1} \leqslant K \|u_l\|_{\beta,r} \quad \begin{cases} l = 0,1,2,\cdots,\\ \text{some } \beta, r,\end{cases}$$

where $K > 0$ and is independent of l, $\| \ \|_{-1}$ denotes the $\mathscr{H}^{-1}(\mathbb{R}^4)$ norm, and $\| \ \|_{\beta,r}$ is defined, for β and r being nonnegative integers, by

$$\begin{aligned}\|u_l\|_{\beta,r}{}^2 = &\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} dx_0 dx_1 dx_2 dx_3\\ &\times \sum_{0 \leqslant \rho \leqslant \beta} \sum_{\substack{0 \leqslant q_0 + q_1 + q_2 + q_3 \leqslant r\\ q_0, q_1, q_2, q_3 \geqslant 0}} (1 + x^2)^\rho\\ &\times \left| \frac{d^{q_0}}{dx_0{}^{q_0}} \frac{d^{q_1}}{dx_1{}^{q_1}} \frac{d^{q_2}}{dx_2{}^{q_2}} \frac{d^{q_3}}{dx_3{}^{q_3}} u_l(x) \right|^2\\ &x = (x_0, x_1, x_2, x_3) \in \mathbb{R}^4.\end{aligned}$$

We have

$$\begin{aligned}&\sum_{\substack{i_1 i_2 \cdots i_m\\ i_\delta^{(0)} \text{ even}}} |\lambda_{j, i_1 i_2 \cdots i_m}| \times E\{|\phi(v_{i_1}) \times \phi(v_{i_2}) \times \cdots \times \phi(v_{i_m})|\}\\ &\leqslant \sum_{\substack{i_1 i_2 \cdots i_m\\ i_\delta^{(0)} \text{ even}}} |\lambda_{j, i_1 i_2 \cdots i_m}| \times G_m \times \|v_{i_1}\|_{-1} \times \|v_{i_2}\|_{-1} \times \cdots \times \|v_{i_m}\|_{-1}\end{aligned}$$

with G_m depending on m only [since $\phi(g) \in L^p(\Omega, \Sigma, \mu)$, $p = 1,2,\cdots$, for any $g \in \mathscr{H}^{-1}(\mathbb{R}^4)$, and since $E\{\phi(v_{i_1}) \times \phi(v_{i_2}) \times \cdots \times \phi(v_{i_m})\}$ is separately continuous in $v_{i_1}, v_{i_2}, \ldots, v_{i_m}$ with respect to $\mathscr{H}^{-1}(\mathbb{R}^4)$ topology];

$$\begin{aligned}&\leqslant \sum_{\substack{i_1 i_2 \cdots i_m\\ i_\delta^{(0)} \text{ even}}} |\lambda_{j, i_1 i_2 \cdots i_m}| \times G_m \times K^m \times \|u_{i_1}\|_{\beta,r} \times \|u_{i_2}\|_{\beta,r}\\ &\quad \times \cdots \times \|u_{i_m}\|_{\beta,r}\\ &\leqslant B^m \cdot G_m \cdot K^m \cdot \sum_{\substack{i_1 i_2 \cdots i_m\\ i_\delta^{(0)} \text{ even}}} |\lambda_{j, i_1 i_2 \cdots i_m}|\\ &\quad \times \prod_{\kappa=0}^{3} \{(1 + i_1^{(\kappa)})^t \times (1 + i_2^{(\kappa)})^t \times \cdots \times (1 + i_m^{(\kappa)})^t\}\\ &< \infty\end{aligned}$$

since

$$\|u_l\|_{\beta,r} \leqslant B \times \prod_{\kappa=0}^{3} (1 + l^{(\kappa)})^t, \quad l = 0,1,2,\cdots$$

for some $t > 0$, B being a constant for fixed β and r, and since

$$|\lambda_{j, i_1 i_2 \cdots i_m}| \leqslant \frac{Q_{j,s}}{\Pi_{\kappa=0}^3 \{(1 + i_1^{(\kappa)})^s \times (1 + i_2^{(\kappa)})^s \times \cdots \times (1 + i_m^{(\kappa)})^s\}}$$

for any $s = 1,2,\cdots$.

We now let

$$\alpha_j = \sum_{m=0}^{A} \sum_{\substack{i_1 i_2 \cdots i_m\\ i_\delta^{(0)} \text{ even}}} \lambda_{j, i_1 i_2 \cdots i_m} \times \phi(v_{i_1}) \times \phi(v_{i_2}) \times \cdots \times \phi(v_{i_m}).$$

Then $\alpha_j \in L^1(\Omega, \Sigma, \mu)$.

Further, we have

support of $v_l \subset \mathbb{R}^4_>$

since

support of $u_l \subset \mathbb{R}^4_{\geqslant}$.

Hence we can construct a sequence $v_{l,\tau}$, $\tau = 1,2,\cdots$, for each l, such that

$$v_{l,\tau} \underset{\tau\to\infty}{\longrightarrow} v_l$$

in $\mathscr{H}^{-1}(\mathbb{R}^4)$ topology and such that

support of $v_{l,\tau} \subset \mathbb{R}^4_{>}$.

This follows from the fact that the sequence $\{T_{1/n}, n = 1,2,\cdots\}$ of translation operators (where $T_{1/n}$ corresponds to translation through $1/n$ in the positive x_0 direction) in $\mathscr{H}^{-1}(\mathbb{R}^4)$ converges strongly in $\mathscr{H}^{-1}(\mathbb{R}^4)$ since the set $\{T_{1/n}, n = 1,2,\cdots\}$ is uniformly bounded in norm (by 1) and since the sequence converges weakly on the dense subset of $\mathscr{H}^{-1}(\mathbb{R}^4)$ consisting of continuous functions on $\mathbb{R}^4$ with bounded support (*note*: this argument is due to the referee and replaces a previous longer proof). Then $\phi(v_{l,\tau}) \in \Sigma(\Lambda)$ and $\phi(v_{l,\tau}) \underset{\tau\to\infty}{\longrightarrow} \phi(v_l)$ in measure, where $\Lambda = \mathbb{R}^4_{>}$. Consequently, we have

$$\phi(v_l) \in \Sigma(\Lambda).$$

The proof of this goes as follows: There exists a subsequence τ_σ, $\sigma = 1,2,\cdots$, for each l, such that

$$\phi(v_{l,\tau_\sigma}) \underset{\sigma\to\infty}{\longrightarrow} \phi(v_l)$$

almost everywhere in $(\Omega, \Sigma(\Lambda), \mu)$ if we choose the convention that any sub-σ-algebra of (Ω, Σ, μ) contains all elements of measure zero in the σ-algebra (Ω, Σ, μ). Therefore, $\phi(v_l) \in \Sigma(\Lambda)$ since $\phi(v_{l,\tau_\sigma}) \in \Sigma(\Lambda)$.

Thus we also have $\phi(v_{i_1}) \times \phi(v_{i_2}) \times \cdots \times \phi(v_{i_m}) \in \Sigma(\Lambda)$ for all $i_1, i_2, \ldots, i_m$ ($i_\delta^{(0)}$ being even, $\delta = 1,2,\ldots,m$). Therefore,

$$\sum_{\substack{m=0\\ n=0}}^{A} \sum_{\substack{i_1 i_2 \cdots i_m = \langle 0\rangle \langle 0\rangle \cdots \langle 0\rangle\\ i_\delta^{(0)}\ \text{even}}}^{I_1 I_2 \cdots I_m} \lambda_{j,i_1 i_2 \cdots i_m} \times \phi(v_{i_1}) \times \phi(v_{i_2}) \times \cdots \times \phi(v_{i_m}),$$

$$\langle 0 \rangle = (0,0,0,0).$$

belongs to $\Sigma(\Lambda)$ and $L^1(\Omega, \Sigma(\Lambda), \mu)$, where $I_\delta = (I_\delta^{(0)}, I_\delta^{(1)}, I_\delta^{(2)}, I_\delta^{(3)})$ and $I_\delta^{(\nu)} = 0,1,2,\cdots$ for $\nu = 1,2,3$, and $I_\delta^{(0)} = 0,2,4,\cdots$. Since $L^1(\Omega, \Sigma(\Lambda), \mu)$ is complete and since

$$E\left\{\left|\sum_{\substack{m=0\\ n=0}}^{A} \sum_{\substack{i_1 i_2 \cdots i_m = L_1 L_2 \cdots L_m\\ i_\delta^{(0)}\ \text{even}}}^{\hat{L}_1 \hat{L}_2 \cdots \hat{L}_m} \lambda_{j,i_1 i_2 \cdots i_m} \times \phi(v_{i_1}) \times \phi(v_{i_2}) \times \cdots \times \phi(v_{i_m})\right|\right\} < \epsilon$$

for $L_\delta = (L_\delta^{(0)}, L_\delta^{(1)} L_\delta^{(2)} L_\delta^{(3)})$, $\hat{L}_\delta = (\hat{L}_\delta^{(0)}, \hat{L}_\delta^{(1)}, \hat{L}_\delta^{(2)}, \hat{L}_\delta^{(3)})$ with $L_\delta^{(0)}$ and $\hat{L}_\delta^{(0)}$ even and $L_\delta^{(\nu)}, \hat{L}_\delta^{(\nu)} >$ some $N(\epsilon)$ for $\nu = 0,1,2,3$, given any ϵ, we conclude that

$$\alpha_j \in L^1(\Omega, \Sigma(\Lambda), \mu).$$

Let T_ρ be the reflection operator corresponding to the reflection ρ: $x_0 \to -x_0$, x_1, x_2, x_3 kept fixed (see *Notes* in Sec. IIIA). Then we have, with Λ^c being the complement of Λ in $\mathbb{R}^4$ and $\partial\Lambda$ being the boundary of Λ in $\mathbb{R}^4$,

$$\begin{aligned}
&E[(T_\rho \bar{\alpha}_j)\alpha_j]\\
&\underset{\substack{\text{conditional}\\ \text{expectation}}}{=} E[(T_\rho \bar{\alpha}_j) E\{\alpha_j \mid \Sigma(\Lambda^c)\}]\\
&\underset{\substack{\text{weak Markov}\\ \text{property}}}{=} E[(T_\rho \bar{\alpha}_j) E\{\alpha_j \mid \Sigma(\partial\Lambda)\}]\\
&\underset{\substack{\text{reflection}\\ \text{property}}}{=} E[(\bar{\alpha}_j) E\{\alpha_j \mid \Sigma(\partial\Lambda)\}]\\
&\underset{\substack{\text{conditional}\\ \text{expectation}}}{=} E[E\{\bar{\alpha}_j \mid \Sigma(\partial\Lambda)\} E\{\alpha_j \mid \Sigma(\partial\Lambda)\}]\\
&= E[|E\{\alpha_j \mid \Sigma(\partial\Lambda)\}|^2]\\
&\geqslant 0.
\end{aligned}$$

[This proof of $E[(T_\rho \bar{\alpha}_j)\alpha_j] \geqslant 0$ is analogous to that of Rosen's for Euclidean Markov fields over $\mathscr{S}(\mathbb{R}^d)$.[5]]

On the other hand, we have, by assumption (2) and reflection covariance,

$$\sum_{\substack{m=0\\ n=0}}^{A} \mathbf{\Gamma}_{m+n}(\Theta \bar{F}_{mj} \otimes F_{nj}) = E[(T_\rho \bar{\alpha}_j)\alpha_j].$$

Hence we have proved statement (Y'); hence statement (Y), and hence positivity, follow.

(*d*) *Proof of symmetry*: Obvious.

(*e*) *Proof of cluster decomposition*:

Lemma: Let $\eta \in \mathbb{R}^4$. Then a rotation (proper) $\mathscr{R}$ in $\mathbb{R}^4$ exists such that $\mathscr{R}^{-1}\eta = \zeta$, where $\zeta = (\zeta_0, 0, 0, 0)$ and $\mathscr{R}^{-1}$ denotes the inverse of $\mathscr{R}$. Also let $f \in \mathscr{S}(\mathbb{R}^4)$. Then

$$f_{\lambda\eta} = \mathscr{R}(\mathscr{R}^{-1}f)_{\lambda\zeta}, \quad \lambda > 0,$$

where

$$f_{\lambda\eta}(x) = f(x - \lambda\eta),$$
$$(\mathscr{R}^{-1}f)_{\lambda\zeta}(x) = (\mathscr{R}^{-1}f)(x - \lambda\zeta), \quad x \in \mathbb{R}^4.$$

[*Note*: we define $(\mathscr{A}\mathring{f})(x) = \mathring{f}(\mathscr{A}^{-1}x)$ for any rotation (proper) $\mathscr{A}$ in $\mathbb{R}^4$ and for any $\mathring{f} \in \mathscr{S}(\mathbb{R}^4)$.]

Proof of lemma: Let $k = \mathscr{R}^{-1}f$. Then we can prove

$$(\mathscr{R}k)_{\lambda\eta} = \mathscr{R}(k_{\lambda\zeta})$$

where

$$(\mathscr{R}k)_{\lambda\eta}(x) = (\mathscr{R}k)(x - \lambda\eta),$$
$$k_{\lambda\zeta}(x) = k(x - \lambda\zeta).$$

Cluster decomposition for the set $\{\mathbf{\Gamma}_n, n = 0,1,2,\cdots\}$ then follows, using the above lemma, from rotational invariance of the set and cluster decomposition property of $E\{\phi(g_{n_1}) \times \phi(g_{n_2}) \times \cdots \times \phi(g_{n_n})\}$ for all $g_{n_i} \in \mathscr{H}^{-1}(\mathbb{R}^4)$, $i = 1,2,\ldots,n$, $n = 2,3,\cdots$.

2. Proof of Theorem 4

Let $f \in \mathscr{S}(\mathbb{R}^4)$ Then $M_0 f \in L^2(\mathbb{R}^3) \subset \mathscr{H}^{-1}(\mathbb{R}^3)$, and NM_0 is a linear map from $\mathscr{S}(\mathbb{R}^4)$ to $L^2(\mathbb{R}^3) \subset \mathscr{H}^{-1}(\mathbb{R}^3)$ and hence a linear map from $\mathscr{S}(\mathbb{R}^4)$ to $\mathscr{H}^{-1}(\mathbb{R}^3)$. Futher, let $f_i \underset{i\to\infty}{\longrightarrow} f$ in Schwartz topology, where $f_i \in \mathscr{S}(\mathbb{R}^4)$, $i = 1,2,\cdots$, and $f \in \mathscr{S}(\mathbb{R}^4)$, then $M_0 f_i$ converges, as $i \to \infty$, to $M_0 f$ in $L^2(\mathbb{R}^3)$ topology, and hence $(NM_0)f_i$ converges, as $i \to \infty$, to $(NM_0)f$

in $L^2(\mathbb{R}^3)$ topology and hence in $\mathscr{H}^{-1}(\mathbb{R}^3)$ topology. Hence NM_0 is a continuous linear map from $\mathscr{S}(\mathbb{R}^4)$ to $\mathscr{H}^{-1}(\mathbb{R}^3)$.

It remains now to show only, as a consequence of the properties of the map N, that the map M_0 is real and satisfies assumptions (1) and (2) of Theorem 3. We proceed as follows:

(*a*) *Proof that M_0 is a real map*: Let $f \in \mathscr{S}(\mathbb{R}^4)$ and f real. Let $u = M_0 f \in \mathscr{H}^{-1}(\mathbb{R}^3)$. Let further $\tilde{u}(\underline{p}) = v_1(\underline{p}) + iv_2(\underline{p})$, where v_1 and v_2 are real. Then the map M_0 is real (i.e., $u = M_0 f$ is real for f real) if and only if

$$v_1(-\underline{p}) = v_1(\underline{p}),$$
$$v_2(-\underline{p}) = -v_2(\underline{p}).$$

This condition is satisfied by the map M_0 since

$$\left.\begin{array}{l} Y^{(1)}_{lm}(\pi-\theta,\pi+\phi) = (-1)^{l+m} Y^{(1)}_{lm}(\theta,\varphi) \\ Y^{(2)}_{lm}(\pi-\theta,\pi+\phi) = (-1)^{l+m} Y^{(2)}_{lm}(\theta,\varphi) \end{array}\right\},$$
$$\begin{cases} \pi/2 \geqslant \theta \geqslant 0, & 2\pi > \phi \geqslant 0, \\ l = 0,1,2,\cdots, & m = 0, \pm 1,\ldots, \pm l. \end{cases}$$
$$\left.\begin{array}{l} \cos \nu(\pi+\phi) = (-1)^\nu \cos \nu\phi \\ \sin \nu(\pi+\phi) = (-1)^\nu \sin \nu\phi \end{array}\right\}, \quad \begin{array}{l} 2\pi > \phi \geqslant 0, \\ \nu = 0,1,2,\cdots. \end{array}$$

(*b*) *Proof that M_0 satisfies assumption* (1) *of Theorem* 3: Let

$$\Psi^{(1)}_{j\nu}(p_1,p_2,p_3) = \frac{\phi_j(|\mathbf{p}|)}{|\mathbf{p}|} Y^{(1)}_{\nu}(\theta,\varphi)$$
$$\Psi^{(2)}_{j\nu}(p_1,p_2,p_3) = \frac{\phi_j(|\mathbf{p}|)}{|\mathbf{p}|} Y^{(2)}_{\nu}(\theta,\varphi)$$
$$\underline{\Psi}^{(1)}_{j\nu}(p_1,p_2) = \frac{\phi_j(|\underline{\mathbf{p}}|)}{|\underline{\mathbf{p}}|^{1/2}} \frac{\cos \nu\varphi}{\sqrt{2\pi}}$$
$$\underline{\Psi}^{(2)}_{j\nu}(p_1,p_2) = \frac{\phi_j(|\underline{\mathbf{p}}|)}{|\underline{\mathbf{p}}|^{1/2}} \frac{\sin \nu\varphi}{\sqrt{2\pi}}$$

Then $(\widetilde{\Psi}^{(1)}_{j\nu}, \widetilde{\Psi}^{(2)}_{j\nu}, j = 0,2,\cdots, \nu = 0,1,2,\cdots)$ forms a complete orthonormal basis in $L^2(\mathbb{R}^3)$ and also $(\widetilde{\underline{\Psi}}^{(1)}_{j\nu}, \widetilde{\underline{\Psi}}^{(2)}_{j\nu}, j = 0,2,\cdots, \nu = 0,1,2,\cdots)$ forms a complete orthonormal basis in $L^2(\mathbb{R}^2)$.

We know that $f \in \mathscr{S}(\mathbb{R}^4) \subset L^2(\mathbb{R}^4)$ and $u = M_0 f \in L^2(\mathbb{R}^3) \subset \mathscr{H}^{-1}(\mathbb{R}^3)$. We have, letting $\chi_{[a,b]}$ be the characteristic function for the closed interval $[a,b]$, $0 > b > a > -\infty$:

$$\begin{aligned} &\langle \chi_{[a,b]} \otimes \widetilde{\underline{\Psi}}^{(1)}_{j\nu}, u \rangle_{L^2(\mathbb{R}^3)} \\ &= \langle \widetilde{\chi}_{[a,b]} \otimes \underline{\Psi}^{(1)}_{j\nu}, \tilde{u} \rangle_{L^2(\mathbb{R}^3)} \\ &= \langle \widetilde{\chi}_{[a,b]} \otimes \Psi^{(1)}_{j\nu}, \tilde{f} \rangle_{L^2(\mathbb{R}^4)} \\ &= \langle \widetilde{\chi}_{[a,b]} \otimes \widetilde{\Psi}^{(1)}_{j\nu}, f \rangle_{L^2(\mathbb{R}^4)} \\ &= 0 \end{aligned}$$

for all j (even) and all ν, if

support of $f \subset \mathbb{R}^4_>$

and similarly

$$\langle \chi_{[a,b]} \widetilde{\underline{\Psi}}^{(2)}_{j\nu}, u \rangle_{L^2(\mathbb{R}^3)} = 0$$

for all j (even) and all ν, if

support of $f \subset \mathbb{R}^4_>$.

Consequently,

support of $u \subset \mathbb{R}^3_>$

if

support of $f \subset \mathbb{R}^4_>$

Hence M_0 satisfies assumption (1) of Theorem 3.

(*c*) *Proof that M_0 satisfies assumption* (2) *of Theorem* 3: The proof is obvious if we notice that

$$\phi_i(-p_0) = (-1)^i \phi_i(p_0) \qquad \text{for all } i = 0,1,2,\cdots.$$

IV. GENERATION OF FOUR-DIMENSIONAL QUANTUM FIELD THEORIES FROM FOUR-DIMENSIONAL EUCLIDEAN MARKOV FIELDS

In this section we present a theorem on the generation of quantum field theories in four-dimensional space–time from four-dimensional probabilistic X-structures over $\mathscr{H}^{-1}(\mathbb{R}^4)$ which are also "Euclidean invariant." Such structures include Euclidean–Markov fields over $\mathscr{H}^{-1}(\mathbb{R}^4)$ studied by Nelson in Refs. 3 and 4.

Theorem: Let $(\phi_E, (\Omega, \Sigma, \mu))$ be a four-dimensional probabilistic X-structure over $\mathscr{H}^{-1}(\mathbb{R}^4)$ satisfying

$$\begin{aligned} &E\left\{(T_s\phi_E)(g_{n_1}) \times (T_s\phi_E)(g_{n_2}) \times \cdots \times (T_s\phi_E)(g_{n_n})\right\} \\ &\quad = E\left\{\phi_E(g_{n_1}) \times \phi_E(g_{n_2}) \times \cdots \times \phi_E(g_{n_n})\right\}, \end{aligned} \tag{O}$$

where $g_{n_i} \in \mathscr{H}^{-1}(\mathbb{R}^4)$ and $(T_s\phi_E)(g_{n_i}) = \phi_E(g_{n_i} \circ s^{-1})$, $i = 1,2,\ldots,n$, $n = 1,2,\cdots$, for any $s \in \mathrm{ISO}(4)$. Let $P(\Delta)$ be any real polynomial in the four-dimesional Laplacian and define the multilinear functional $\Gamma_n^{(E,P(\Delta))}$ over $\mathscr{S}(\mathbb{R}^4) \times \mathscr{S}(\mathbb{R}^4) \times \cdots \times \mathscr{S}(\mathbb{R}^4)$ (n copies) by

$$\begin{aligned} &\Gamma_n^{(E,P(\Delta))}(f_1,f_2,\ldots,f_n) \\ &\quad = E\{\phi_E(P(\Delta)f_1) \times \phi_E(P(\Delta)f_2) \times \cdots \times \phi_E(P(\Delta)f_n)\}, \\ &\qquad f_i \in \mathscr{S}(\mathbb{R}^4), \quad i = 1,2,\ldots,n. \end{aligned}$$

Then $\Gamma_n^{(E,P(\Delta))}$ defines a certain tempered distribution $\mathbf{\Gamma}_n^{(E,P(\Delta))}$ on $\mathbb{R}^{4n}$ for each $n = 1,2,\cdots$, and the set $\{\mathbf{\Gamma}_n^{(E,P(\Delta))}, \; n = 0,1,2,\cdots\}$ with $\mathbf{\Gamma}_0^{(E,P(\Delta))} = 1$ is a set of Euclidean Green's Functions for a four-dimensional Hermitian scalar quantum field theory not including the property of uniqueness of vacuum.

Proof: This theorem is an immediate consequence of Theorem 1 in Sec. III since $P(\Delta)$ is a real continuous linear map from $\mathscr{S}(\mathbb{R}^4)$ to $\mathscr{H}^{-1}(\mathbb{R}^4)$ satisfying assumptions (1) and (2) of that theorem and further since, for each $n = 1,2,\cdots$, $\Gamma_n^{(E,P(\Delta))}$ is jointly invariant under any element of ISO(4) as a consequence of condition (O) on $(\phi_E, (\Omega, \Sigma, \mu))$ and the fact that $P(\Delta)$ commutes with the action of any element of ISO(4) on any element of $\mathscr{S}(\mathbb{R}^4)$.

We now present a remark on the above theorem.

Remark: The theorem can be obtained by noting that the probabilistic X-structure $(\phi_E, (\Omega, \Sigma, \mu))$ gives rise to a Hermitian scalar quantum field theory in four-dimensional space–time with field θ_E and that $P(\Box)\theta_E$, together with a restriction of the same unitary representation of the Poincare group and the same vacuum, also form a four-dimensional Hermitian scalar quantum field theory, where $\Box$ is the d'Alembertian in four-dimensional space–time. What we want to emphasize here is that the theorem is also an extremely simple consequence of Theorem 1 in Sec. III.

ACKNOWLEDGMENTS

The author here thanks R. F. Streater for conversations after the establishment of the main results in this paper and to C. W. Chan for discussions on some mathematical points. He also wishes to thank Y. M. Park, S. Ruijsenaars, and A. S. Wightman for conversations. Further, a correspondence with A. Jaffe and a remark by K. Osterwalder are gladly acknowledgd. Lastly, he is happy to acknowledge the kind hospitality of the Dublin Institute for Advanced Study (where the " mapping idea" first arose), the Physics Department of Princeton University, and the Mathematics Department of Hong Kong University.

[1]R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).
[2]R. Jost, *The General Theory of Quantized Fields* (Amer. Math. Soc. Publ., Providence, R. I. 1965).
[3]E. Nelson, J. Funct. Anal. **12**, 97 (1973).
[4]E. Nelson, J. Funct. Anal. **12**, 211 (1973).
[5]E. Nelson, in *Constructive Quantum Field Theory*, edited by G. Velo and A. S. Wightman (Springer, New York, 1973), p. 94.
[6]K. Osterwalder and R. Schrader, Commun. Math. Phys. **31**, 83 (1973).
[7]K. Osterwalder and R. Schrader, Commun. Math. Phys. **42**, 281 (1975).
[8]K. Osterwalder, in *Constructive Quantum Field Theory*, edited by G. Velo and A. S. Wightman (Springer-Verlag, New York, 1973), p. 71.
[9]M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, (Academic, New York, 1972), Vol. 1.

Maxwell's equations in axiomatic quantum field theory. II. Covariant and noncovariant gauges

P. J. M. Bongaarts
Instituut Lorentz, Nieuwsteeg 18, 2311 SB Leiden, The Netherlands

(Received 15 June 1981; accepted for publication 7 August 1981)

The outlines of a general formalism for the description of the Maxwell field in the framework of axiomatic quantum field theory were given in a preceding paper. It was based on the use of locally convex topological spaces connected in a natural way with the distribution properties of the n-point functions of both the $A_\mu(x)$ and $F_{\mu\nu}(x)$ fields. In this paper this approach is developed further. We discuss in particular aspects of the reconstruction theorem, gauges and their equivalence, and symmetries and gauge transformations. Finally a systematic and unified derivation is given of the various Lorentz covariant and noncovariant free field gauges together with the properties of the associated $A_\mu(x)$ operator field theories.

PACS numbers: 11.10.Cd, 03.70. + k, 02.30. + g

I. INTRODUCTION

A general, mathematically rigorous formalism for the quantized electromagnetic field, in the spirit of Wightman's axiomatic approach to quantum field theory, was developed in Ref. 1. The special problems of the Maxwell quantum field such as the difficulties connected with the simultaneous occurrence of the two fields $A_\mu(x)$ and $F_{\mu\nu}(x)$ and the incompatibility of manifest Lorentz covariance with a positive-definite metric in the space of state vectors were met by extending the standard Wightman framework in a suitable manner, facilitated by the use of algebraic concepts due to Borchers.

The leading principles in this generalization of Wightman theory are, in the first place, the importance attached to the reconstruction theorem which states that a quantum field theory is completely determined by its system of n-point functions, i.e., vacuum expectation values of products of field operators, and, in the second place, the fact that in the light of the reconstruction theorem the fundamental mathematical structure of the Wightman formalism is not that given by the Hilbert space nature of the state space but the underlying structure of locally convex topological spaces connected with the distribution properties of the n-point functions.

Borchers' algebraic version of Wightman theory is particularly well suited to this point of view because it treats systems of n-point functions as continuous linear functionals on an involutive topological algebra constructed as tensor algebra from the basic space of test functions used for smearing the field operators. As a consequence, the reconstruction theorem takes the form of a special case of a well-known general theorem on the representation of algebras.

In its original form the Borchers formalism uses only positive linear functionals and is then completely equivalent to standard Wightman theory. The essential mathematical ingredient in the reconstruction theorem is, however, not positivity but continuity. It is therefore natural to extend the formalism to arbitrary continuous linear functionals leading to representations in topological vector spaces more general than Hilbert space, with inner products that are not necessarily positive definite. This is exactly what is needed for a general theory of the quantized electromagnetic field on the principles just indicated. In this extended form the Borchers formalism gives a natural generalization of Wightman theory in which the Hilbert space property is no longer a general requirement for the spaces in which the field operators are defined but appears as an additional property only at places where this is desirable for physical interpretation.

The fundamental difference between our approach and work such as that of Strocchi *et al.* (see Refs. 2 and 3 and, for further developments Ref. 4) is that we do not employ auxiliary, noninvariant Hilbert space structures, which are mathematically very awkward and, moreover, have no physical meaning, but make instead a consequent use of the locally convex spaces that are given in a natural way by the properties of the n-point functions. This involves us in mathematical methods that are slightly less familiar than Hilbert space theory, but this is more than compensated for by a considerable gain in coherence and transparency of the resulting formalism.

For the description of the $A_\mu(x)$ and $F_{\mu\nu}(x)$ fields we start with two distinct Borchers' algebras. There is an algebra $\mathscr{A}^A = \Sigma_{n=0}^\infty \oplus (\hat{\otimes}^n \mathscr{S}^{(3)})$, with $\mathscr{S}^{(3)}$ a space of test functions $f^\mu(x)$, suggested by the heuristic expression $A(f) = \int A_\mu(x) f^\mu(x)\, d^4x$ for the potential operator, and a second algebra $\mathscr{A}^F = \Sigma_{n=0}^\infty \oplus (\hat{\otimes}^n \mathscr{S}^{(2)})$, where $\mathscr{S}^{(2)}$ consist of antisymmetric functions $\psi^{\mu\nu}(x)$, because of the expression $F(\psi) = \int F_{\mu\nu}(x)\psi^{\mu\nu}(x)\, d^4x$. The algebras are related by an algebraic homomorphism θ_d, generated by a linear map d (or d_{32}): $\mathscr{S}^{(2)} \rightarrow \mathscr{S}^{(3)}$, defined as $(d\psi)^\mu = 2\partial_\nu \psi^{\mu\nu}$. The continuous linear functionals on the algebras, i.e., possible systems of n-point functions for the $A_\mu(x)$ and $F_{\mu\nu}(x)$ fields are then connected by the transposed map θ'_d: $\mathscr{A}'^A \rightarrow \mathscr{A}'^F$, and this takes the place of the classical relation $\partial_\mu A_\nu - \partial_\nu A_\mu = F_{\mu\nu}$. All this can be found in detail in Ref. 1, where also the first of the two basic results on which our formalism is built is given as Theorem 3. This states that the image of the transposed map θ'_d consists of all systems of F-field n-point functions that correspond, through the reconstruction theorem, to operator field theories in which the first Maxwell equation $\partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu} = 0$ holds as an operator equation. This allowed us to introduce the concept of gauge as a system of A-field n-point functions in the inverse image under θ'_d of a given F-field system, and also as the corresponding $A_\mu(x)$ operator field theory. Although the state spaces of F and A theories are in this way *a priori* distinct, we showed that they are connected by a canonical "partial" isometry which in special cases can be used as an identification map.

In this paper the formalism is developed further and is finally applied to the case of the free field, where a rigorous and systematic derivation is given of the various Lorentz covariant and noncovariant gauges together with their properties as operator field theories. In Sec. II special mathematical aspects of our version of the Borchers formalism necessary for the subsequent sections are discussed. Particular attention is paid to aspects of the reconstruction theorem and to the subject of transformations and symmetries. In Sec. III there is first a brief review of some results from Ref. 1, and then a definition of the concept of gauge equivalence. A characterization of this in terms of n-point functions is given in Theorem 3.1, the second result that is basic for the formalism. Remarks on different gauge conditions and a discussion of various possible meanings of the term gauge transformation make up Sec. IV. In Sec. V the systematic treatment of the free field gauges is started and the general form of the translation invariant two-point function for the free A_μ field derived. To facilitate the subsequent investigation of the operator field properties of the various free field gauges, we study in Sec. VI in a general setting what we call "Gaussian" states, i.e., systems of n-point functions determined by a two-point function in the way that is typical for free boson field theories. The main characteristic of such a system is shown to be that the associated operator field theories have canonical realizations in terms of systems of creation and annihilation operators in Fock or "many-particle' spaces. The Fock space structures involved are of a more general nature than those occurring in standard Hilbert space free field theories. They are based on "natural" locally convex topologies, in accordance with the spirit of our approach. In this respect this is different from and in fact more transparent than recent work on indefinite metric second quantization with auxiliary Hilbert space structures such as e.g., Ref. 5, although in Ref. 6 these additional structures have become already less important. In Sec. VII the discussion of the free field gauges is resumed. The general form of the Lorentz invariant two-point function for the free A_μ field is derived, and it is shown that a Lorentz invariant free field gauge always leads to an indefinite metric space. Two distinct important classes of Lorentz invariant free field gauges are discussed, the generalized Landau gauges and the generalized Gupta–Bleuler gauges, characterized respectively by the conditions $\partial^\mu A_\mu = 0$ and $\partial^\nu \partial_\nu A_\mu = 0$, as operator equations for the field A_μ. In Sec. VIII the Coulomb gauge is investigated, as the most important and typical example of noncovariant but positive metric gauge. Its properties as an operator field theory are worked out in some detail, in particular the precise meaning of noncovariance under Lorentz transformations. This provides a characteristic example of the interplay of locally convex aspects and of a physical Hilbert space structure arising in this case in a natural way.

The main reference for this paper is Ref. 1, which will be denoted as I. It contains among other things a brief review of the Borchers formalism as needed for this work; see, however, for more details Refs. 7–12. The references for Wightman axiomatic field theory are 13–15, and for the theory of topological vector spaces 16–19. There is an unfortunate but unavoidable double use of the terms homomorphism and isomorphism in this paper, denoting on one hand maps that preserve algebraic, i.e., multiplicative structure, and on the other hand continuous linear maps between topological vector spaces having the important additional property of being relatively open, see, e.g., Ref. 16, Chap. 17, or Ref. 19, Chap. III, Sec. 1. Hopefully the context together with additional remarks will prevent confusion. We use heuristic "generalized function" language for distributions at some places, in particular in Secs. V, VII, and VIII, where this makes for easier reading and where a reformulation in the correct test function language is obvious and straightforward. Finally we consider in this paper only Lorentz transformations belonging to the connected part of the Lorentz group.

II. REMARKS ON GENERAL MATHEMATICAL ASPECTS OF ALGEBRAIC WIGHTMAN THEORY

The principal theorem in standard axiomatic field theory is the reconstruction theorem, which says that an operator field theory is completely characterized, up to unitary equivalence, by the vacuum expectation values of products of field operators, from which it can in fact be recovered by an explicit construction. In Borchers' algebraic version of Wightman theory this theorem has a simple form; it is a special case of the general relation between cyclic representations of an involutive algebra and its positive linear forms. This relation is, of course, well known from C^*-algebra theory and its applications in statistical mechanics and field theory. The term GNS (Gel'fand–Naimark–Segal) representation which we shall occasionally use has its origin there. An advantage of the special context of topological algebras constructed as tensor algebras of test function spaces compared to that of C^*-algebras is that in the relation between linear functionals and representations the role of positivity is less important; only continuity is essential. Therefore, the reconstruction theorem remains true, after minor modifications, in our generalized approach to algebraic Wightman theory, in which, with an eye to application to the Maxwell field, positivity has been dropped as a general requirement.

For a more explicit discussion of this and related aspects of this approach we consider the general situation where we have a Borchers algebra $\mathscr{A} = \Sigma_{n=0}^{\infty} \oplus (\widehat{\otimes}^n \mathscr{V})$ constructed as a topological direct sum of completed tensor products of a basic complex vector space $\mathscr{V}$ which we suppose to be a nuclear Fréchet space with continuous conjugation. (The spaces $\mathscr{V}$ that we use are the spaces of multi component test functions on $\mathbb{R}^4$ that will be listed in the next section.)

We introduce some convenient terminology: By a *representation* of $\mathscr{A}$ we shall mean a homorphism π of $\mathscr{A}$ into the algebra of linear operators in a complex vector space $\mathscr{H}$, equipped with a nondegenerate Hermitian form $(\cdot, \cdot)$ (not necessarily positive definite!), such that $(\Omega_1, \pi(a)\Omega_2) = (\pi(a^*)\Omega_1, \Omega_2)$, $\forall a \in \mathscr{A}$, $\forall\ \Omega_1, \Omega_2 \in \mathscr{H}$. [The linear operators $\pi(a)$ are of course defined everywhere in $\mathscr{H}$.] A vector Ω in $\mathscr{H}$ will be called *strictly cyclic* if the map from $\mathscr{A}$ to $\mathscr{H}$ defined by $a\mapsto\pi(a)\Omega$ is surjective. The representation π for which such a vector exists is also called strictly cyclic. Two representations π_1 and π_2, in $\mathscr{H}_1$ and $\mathscr{H}_2$, are called *equivalent* if there exist a linear bijection $W: \mathscr{H}_1\to\mathscr{H}_2$, isometric, i.e., with $(W\Omega, W\Omega') = (\Omega, \Omega')$, $\forall\, \Omega, \Omega'\in\mathscr{H}_1$, and such that $W\pi_1(a) = \pi_2(a)W$, $\forall a\in\mathscr{A}$.

Lemma 2.1: Two strictly cyclic representations π_1 and π_2, in $\mathscr{H}_1$ and $\mathscr{H}_2$, are equivalent if and only if for some pair of strictly cyclic vectors Ω_1 and Ω_2 the equality $(\Omega_1, \pi_1(a)\Omega_1) = (\Omega_2, \pi_2(a)\Omega_2)$ holds for all $a\in\mathscr{A}$.

Proof: On one hand, if π_1 and π_2 are equivalent and Ω_1 strictly cyclic in $\mathscr{H}_1$, then $\Omega_2 = W\Omega_1$ is strictly cyclic in $\mathscr{H}_2$ and the equality holds: on the other hand, if for strictly cyclic Ω_1 and Ω_2, one has $(\Omega_1, \pi_1(a)\Omega_1) = (\Omega_2, \pi_2(a)\Omega_2)$, $\forall a\in\mathscr{A}$, then $\pi_1(a)\Omega_1\mapsto\pi_2(a)\Omega_2$ defines a linear map W with all the properties required, because $\pi_1(a)\Omega_1 = 0 \Leftrightarrow (\Omega_1, \pi_1(b^*a)\Omega_1) = 0$, $\forall b\in\mathscr{A} \Leftrightarrow (\Omega_2, \pi_2(b^*a)\Omega_2) = 0$, $\forall b\in\mathscr{A} \Leftrightarrow \pi_2(a)\Omega_2 = 0$.

A representation will be called *continuous* if $\mathscr{H}$ is a locally convex Hausdorff topological vector space, for which the Hermitian form $(\cdot,\cdot)$ is separately continuous (it will then be called an *inner product* on $\mathscr{H}$) and if the bilinear map from $\mathscr{A}\times\mathscr{H}$ to $\mathscr{H}$ defined by $(a, \Omega)\mapsto\pi(a)\Omega$ is separately continuous. Two continuous representations are called *topologically equivalent* if they are equivalent and the isometry W is a topological isomorphism.

For continuous strictly cyclic representations one cannot, from the equality of "expectation values" $(\Omega_1, \pi_1(a)\Omega_1) = (\Omega_2, \pi_2(a)\Omega_2)$, conclude that the representations are topologically equivalent. A slightly modified statement holds, however. To obtain it, one observes the following: If for a strictly cyclic, not necessarily continuous representation π, with strictly cyclic vector Ω, the expectation value $(\Omega, \pi(a)\Omega)$ happens to be a continuous function of a, then the surjective map $\chi: \mathscr{A}\to\mathscr{H}$, defined as $\chi(a) = \pi(a)\Omega$, can be used to induce a locally convex Hausdorff topology on $\mathscr{H}$, for which the Hermitian form $(\cdot,\cdot)$ is separately continuous and the representation π continuous. This topology is, of course, just the quotient topology on the quotient space $\mathscr{A}/\mathrm{Ker}\,\chi$ transferred to $\mathscr{H}$; it makes χ into a surjective topological homomorphism, giving the continuity properties and also the fact that we have in this way the strongest possible topology in $\mathscr{H}$ for which π is continuous. This implies that a continuous, strictly cyclic representation has on its representation space a "natural" strongest topology, the topology induced by the continuity of $(\Omega, \pi(a)\Omega)$.

Lemma 2.2: Two continuous, strictly cyclic representations with representation spaces equipped with the "natural" topologies are topologically equivalent if and only if for some pair of strictly cyclic vectors Ω_1, Ω_2 the equality $(\Omega_1, \pi_1(a)\Omega_1) = (\Omega_2, \pi_2(a)\Omega_2)$ holds for all a in $\mathscr{A}$.

Proof: The only point worth mentioning is that the isometry W is a topological isomorphism because $\chi_1: \mathscr{A}\to\mathscr{H}_1$ and $\chi_2: \mathscr{A}\to\mathscr{H}_2$ are topological homomorphisms.

Let ω be an element from the topological dual $\mathscr{A}'$ that is real in the sense of $\omega(a^*) = \overline{\omega(a)}$, $\forall a\in\mathscr{A}$. Define $\mathscr{I}_\omega = \{a\in\mathscr{A}\,|\,\omega(ba) = 0, \forall b\in\mathscr{A}\}$. This is a closed left ideal in $\mathscr{A}$. (For $\omega\geqslant 0$, i.e., $\omega(a^*a)\geqslant 0$, $\forall a\in\mathscr{A}$, this coincides with $\mathscr{I}_\omega = \{a\in\mathscr{A}\,|\,\omega(a^*a) = 0\}$). Let χ be the canonical surjection $\mathscr{A}\to\mathscr{A}/\mathscr{I}_\omega$. Then ω defines a continuous, strictly cyclic representation of $\mathscr{A}$, the GNS representation:

Lemma 2.3: The representation of $\mathscr{A}$ defined by taking $\mathscr{H} = \mathscr{A}/\mathscr{I}_\omega$ as representation space with inner product $(\chi(a), \chi(b)) = \omega(a^*b)$ and as operators $\pi(a)$ the operators given by $\pi(a)\chi(b) := \chi(ab)$, is a continuous, strictly cyclic representation, with strictly cyclic vector $\Omega_0 := \chi(e)$, e unit element of $\mathscr{A}$. The representation space $\mathscr{H}$ carries the "natural" topology and one has $\omega(a) = (\Omega_0, \pi(a)\Omega_0)$, $\forall a\in\mathscr{A}$.

The proof of this lemma is also very simple and follows from the preceding remarks. It is therefore omitted.

Lemma 2.2 and 2.3 together constitute the *reconstruction theorem* in the present version of algebraic Wightman theory.

We shall mainly use elements ω in $\mathscr{A}'$ that are not only real but have also the normalization $\omega(e) = 1$. Such elements will be called *states* on $\mathscr{A}$. The strictly cyclic vectors Ω corresponding to such ω through the reconstruction theorem have then unit length.

If a state is positive, i.e., if $\omega(a^*a)\geqslant 0$, $\forall a\in\mathscr{A}$, the GNS representation space $\mathscr{H}$ is a pre-Hilbert space. After completion of $\mathscr{H}$ to a Hilbert space, we are back in the situation of standard Wightman theory. $\mathscr{H}$ then appears as the invariant dense domain on which the in general unbounded operators $\pi(a)$ are defined and the continuity properties of the representation become continuity properties of expectation values with respect to vectors from this domain.

Transformations and *symmetries* are an important aspect of any field theoretic formalism. In the Borchers formalism as it is employed in this work it is natural to describe transformations by means of algebraic automorphism of $\mathscr{A}$. Such automorphisms should commute with the conjugation $*$ and as linear maps should be topological isomorphisms.

The most important and at the same time simplest of such automorphisms are those generated by real (i.e., conjugation preserving) linear topological isometries T in the basic test function space $\mathscr{V}$, according to $\alpha_T(f_1\otimes\cdots\otimes f_n) = (Tf_1)\otimes\cdots\otimes(Tf_n)$, $f_j\in\mathscr{V}$. This is the way Lorentz transformations appear in the theory. One has, for example, for the $A_\mu(x)$ field theory, where $\mathscr{V} = \mathscr{S}^{(3)}$, $\mathscr{A} = \mathscr{A}^A$, for every inhomogeneous Lorentz transformation (u, Λ) a topological isomorphism $T_{(u,\Lambda)}: \mathscr{S}^{(3)}\to\mathscr{S}^{(3)}$, given by $(T_{(u,\Lambda)}f)^\mu(x) = \Lambda^\mu{}_\nu f^\nu(\Lambda^{-1}(x-u))$, which in turn defines an automorphism $\alpha_{(u,\Lambda)}: \mathscr{A}^A\to\mathscr{A}^A$.

There is a second class of simple explicitly given automorphisms. These depend on real elements λ of the topological dual $\mathscr{V}'$ and are defined by extension of $\alpha_\lambda e = e$, $\alpha_\lambda f = f+\lambda(f)e$, $\alpha_\lambda(f_1\otimes f_2) = (f_1+\lambda(f_1)e)\otimes(f_2+\lambda(f_2)e) = f_1\otimes f_2+\lambda(f_1)f_2+\lambda(f_2)f_1+\lambda(f_1)\lambda(f_2)e$, $\forall f, f_1, f_2\in\mathscr{V}$, etc. These α_λ shift expectations values of single field operators. They will not be used in this paper, except as basis for a single remark on gauge transformations in Sec. IV, but are of importance in further developments connected with the definition of scattering states and appear in particular in the discussion of "displaced Fock representations" in the terminology of Ref. 20, or noncentered Gaussian states in the more general terminology of Sec. VI.

An automorphism α determines a *symmetry* if a state ω is invariant under α, i.e., $\omega(\alpha(a)) = \omega(a)$, $\forall a \in \mathscr{A}$. As one verifies easily the relation $\pi(a)\Omega_0 \rightarrow \pi(\alpha(a))\Omega_0$ defines then a linear map U of the GNS representation space $\mathscr{H}$ onto itself, which is a topological isomorphism, isometric with respect to the inner product in $\mathscr{H}$, leaves Ω_0 invariant and has the property $\pi(\alpha(a)) = U\pi(a)U^{-1}$, $\forall a \in \mathscr{A}$. If the state ω is positive, U can be extended to a unitary operator in the Hilbert-space completion of $\mathscr{H}$; we are then back in the situation of symmetry in standard quantum theory. In the example of $A_\mu(x)$ field theory the GNS space of a Lorentz-invariant state carries a representation of the inhomogeneous Lorentz group by isometric linear isomorphisms, transforming the field operators $A(f)$ as $U_{(u,\Lambda)}A(f)U^{-1}_{(u,\Lambda)} = A(T_{(u,\Lambda)}f)$, which corresponds to the heuristic relation $U_{(u,\Lambda)}A_\mu(x)U^{-1}_{(u,\Lambda)} = \Lambda^{\nu}{}_{\mu}A_\nu(\Lambda x + u)$, the conventional transformation rule for a Lorentz covariant vector field operator. It is a characteristic feature of the Borchers formalism as we use it that the representation $U_{(u,\Lambda)}$ as a representation in terms of isometric operators has strong smoothness properties, independent of a possible positivity of the invariant state. [In fact, for the free $A_\mu(x)$ photon field there are no positive Lorentz invariant states, as will be proved in Sec. VII.]

It is useful to consider transformations in a somewhat more general setting.

Let $\mathscr{A}_1$ and $\mathscr{A}_2$ be two Borchers algebras. Let α be a $*$-preserving, continuous algebraic homomorphism of $\mathscr{A}_1$ into $\mathscr{A}_2$ and ω_1, ω_2 two states on $\mathscr{A}_1$, $\mathscr{A}_2$ respectively, such that ω_1 is the image of ω_2 under the transposed map α': $\mathscr{A}'_2 \rightarrow \mathscr{A}'_1$, i.e., $\omega_1(a) = \omega_2(\alpha(a))$, $\forall\ a \in \mathscr{A}_1$.

Corresponding to α there is a canonical "partial" isometry between the GNS spaces of ω_1 and ω_2. Let the GNS representations be π_1 and π_2, in the spaces $\mathscr{H}_1$ and $\mathscr{H}_2$, $\pi_2(\alpha(a_1))\Omega_2 \mapsto \pi_1(a_1)\Omega_1$, $\forall\ a \in \mathscr{A}_1$, defines a surjective linear map W from a subspace $\mathscr{H}_2^{(\alpha)}$ of $\mathscr{H}_2$ onto $\mathscr{H}_1$, because $\omega_2(b_2\alpha(a_1)) = 0$, $\forall\ b_2 \in \mathscr{A}_2$, implies $\omega_1(b_1 a_1) = 0$, $\forall\ b_1 \in \mathscr{A}_1$. One verifies that this map W is isometric, maps Ω_2 onto Ω_1, and has the "intertwining" property $W\pi_2(\alpha(a_1)) = \pi_1(a_1)W$, $\forall\ a_1 \in \mathscr{A}_1$. It also agrees with α' in the sense that if a vector $\psi_2 \in \mathscr{H}_2^{(\alpha)} \subset \mathscr{H}_2$ corresponds to the state ν_2 on $\mathscr{A}_2$ according to $\nu_2(a_2) = (\psi_2, \pi_2(a_2)\psi_2)$, $\forall\ a_2 \in \mathscr{A}_2$, then the vector $W\psi_2 \in \mathscr{H}_1$ corresponds in the same way to a state ν_1 on $\mathscr{A}_1$ which is just $\alpha'\nu_2$. In general, W is not continuous, and neither is $\mathscr{H}_2^{(\alpha)}$ closed. If $\operatorname{Ker} W = 0$, as is the case for a positive state ω_2, then W^{-1} exists also and can then be used to identify $\mathscr{H}_1$ allbegraically with the subspace $\mathscr{H}_2^{(\alpha)}$. If α is also a homomorphism in the sense of topological vector spaces, then W can be shown to be an open map, and, therefore, W^{-1}, when it exists, a continuous map.

Symmetry as defined above is a special case of this situation where $\mathscr{A}_1 = \mathscr{A}_2$, $\omega_1 = \omega_2$, and α with continuous inverse. The isometry U, implementing the automorphism α is just W^{-1}, defined in this way to have agreement with conventional transformation formulas.

A less trivial case is the general connection between $A_\mu(x)$ and $F_{\mu\nu}(x)$ theories for the photon field, a basic feature of our formalism. In this situation $\mathscr{A}_1 = \mathscr{A}^F$, $\mathscr{A}_2 = \mathscr{A}^A$, and α is the algebraic and topological homomorphism θ_d. The state ω_1 is a state of the $F_{\mu\nu}(x)$ field, ω_2 a possible gauge for ω_1. The isometry W the canonical map from $\mathscr{H}^A_{ph}$ onto $\mathscr{H}^F$.

The case where $\mathscr{A}_1 = \mathscr{A}_2 = \mathscr{A}^A$, ω_1 and ω_2 gauge equivalent (to be defined in the next section) and α not necessarily invertible will be used in the discussion of gauge transformations.

III. GAUGE EQUIVALENCE

In classical electromagnetism two different potentials A_μ are equivalent when giving the same field strength $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. In that case the difference $A'_\mu - A_\mu$ can be written as a gradient $\partial_\mu \phi$.

In I we showed that in a Wightman formalism for the quantized electromagnetic field the relation between potentials and field strength is fundamentally a relation between systems of n-point functions. From the n-point functions for the potential operator $A_\mu(x)$ one obtains the n-point functions for the field strength operator $F_{\mu\nu}(x)$ by means of the map θ'_d, defined in I and corresponding to the heuristic formula

$$\omega^F_{\mu_1\nu_1\cdots\mu_n\nu_n}(x_1, \ldots, x_n) = \epsilon^{\rho_1\tau_1}_{\mu_1\nu_1}\cdots\epsilon^{\rho_n\tau_n}_{\mu_n\nu_n}\,\partial^1_{\rho_1}\cdots\partial^n_{\rho_n}\,\omega^A_{\tau_1\cdots\tau_n}(x_1, \ldots, x_n) \tag{3.1}$$

with $\partial^j_{\rho_j} = \partial/\partial x_j^{\rho_j}$, the usual summation convention, and $\epsilon^{\rho\tau}_{\mu\nu}$ the permutation symbol in two indices, nonzero only when ρ, τ is a permutation of μ, ν and then equal to ± 1 according to the sign of that permutation.

A system of n-point functions $\omega^F_{\mu_1\nu_1\cdots\mu_n\nu_n}$ associated with a $F_{\mu\nu}(x)$ theory satisfying the ordinary Wightman axioms and in which $\partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu} = 0$ holds as an operator relation can always be obtained in this manner from a system of functions $\omega^A_{\mu_1\cdots\mu_n}$. This was proved in I, as a statement on the image of the map θ'_d (Theorem 3). Such a system $\omega^A_{\mu_1\cdots\mu_n}$ together with the $A_\mu(x)$ operator field theory obtained from it by the reconstruction theorem as discussed in the preceding section was called in I a *gauge* for the given $F_{\mu\nu}(x)$ theory. There is an obvious notion of equivalence for two different gauges:

Two $A_\mu(x)$ field theories will be called *gauge equivalent* if their n-point functions $\omega^A_{\mu_1\cdots\mu_n}$ give in formula (3.1) the same n-point functions $\omega^F_{\mu_1\nu_1\cdots\mu_n\nu_n}$. The main result of this section is that the differences of such $A_\mu(x)$ n-point functions can also, in a certain sense, be written in terms of gradients. This will be formulated in a precise and compact way in Theorem 3.1. To see that we have indeed a rather natural generalization of the classical situation, it may be helpful to state the theorem first in more heuristic form.

Theorem 3.1 (heuristic formulation): Two $A_\mu(x)$ theories are gauge equivalent if and only if there exist (tempered) distributions $\phi^{(n,l)}_{\mu_1\cdots\mu_n}(x_1, \ldots, x_n)$, $n = 1, 2, \cdots$, $l = 1, 2, \ldots, n$, such that the differences of the n-point functions for the $A_\mu(x)$ fields can be written as sums:

$$\omega'^A_{\mu_1\cdots\mu_n} - \omega^A_{\mu_1\cdots\mu_n} = \sum_{l=1}^{n} \partial^l_{\mu_l}\,\phi^{(n,l)}_{\mu_1\cdots\mu_{l-1}\mu_{l+1}\cdots\mu_n}. \tag{3.2}$$

This theorem provides an explicit characterization of all possible $A_\mu(x)$ field theories gauge equivalent with a given $A_\mu(x)$ theory. Such an equivalence class is very large as it involves the choice of an infinite sequence of arbitrary distributions $\phi^{(n,1)}_{\mu_1\cdots\mu_{n-1}}$. Only a small part of these possibilities have practical value and will occur in the standard theory of the quantized Maxwell field. On the other hand, the notion of gauge equivalence based on the formula $A'_\mu = A_\mu + \partial_\mu\phi$, as operator relation, with $\partial_\mu\phi$ the gradient of some operator field, an idea which also might suggest itself as a generalization of the classical situation, is definitely too narrow. This will be argued in the next section and will be evident from the explicit discussion of the free field in Secs. VI and VII where the more general idea of equivalence through the n-point functions will be needed to connect some of the well-known free field gauges.

To prepare for the rigorous version of Theorem 3.1, we collect some material, part of which was already introduced in I.

The *basic test function spaces* $\mathscr{S}_{(j)}, j = 1, 2, 3, 4$, that we employ are defined as spaces of complex valued, multicomponent $\mathscr{S}(\mathbb{R}^4)$ functions; all antisymmetric tensor functions $\chi^{\mu\nu\rho}(x)$ for $j = 1$, $\psi^{\mu\nu}(x)$ for $j = 2$, all vector functions $f^\mu(x)$ for $j = 3$, and scalar functions $\phi(x)$ for $j = 4$. The spaces $\mathscr{S}^{(j)}$ are obviously nuclear Fréchet spaces.

We consider *linear subspaces* $\mathscr{S}^{(j)}_0 \subset \mathscr{S}^{(j)}, j = 2, 3, 4$, defined as

$$\mathscr{S}^{(2)}_0 = \{\psi\in\mathscr{S}^{(2)} | \psi^{\mu\nu} = \partial_\rho\chi^{\mu\nu\rho};\ \chi\in\mathscr{S}^{(1)}\},$$

$$\mathscr{S}^{(3)}_0 = \{f\in\mathscr{S}^{(3)} | f^\mu = \partial_\nu\psi^{\mu\nu};\ \psi\in\mathscr{S}^{(2)}\},$$

$$\mathscr{S}^{(4)}_0 = \{\phi\in\mathscr{S}^{(4)} | \phi = \partial_\mu f^\mu;\ f\in\mathscr{S}^{(3)}\}.$$

The $\mathscr{S}^{(j)}_0$ can be defined equivalently as

$$\mathscr{S}^{(2)}_0 = \{\psi\in\mathscr{S}^{(2)} | \partial_\mu\psi^{\mu\nu} = 0\},$$

$$\mathscr{S}^{(3)}_0 = \{f\in\mathscr{S}^{(3)} | \partial_\mu f^\mu = 0\},$$

$$\mathscr{S}^{(4)}_0 = \{\phi\in\mathscr{S}^{(4)} \Big| \int_{-\infty}^{+\infty} \phi(x) d^4x = 0\}.$$

For $j = 2, 3$, this was proved in I (Theorem 1). For $j = 4$ we give a proof along the same lines, using again the division property of test functions. After Fourier transformation one has to prove that for $\phi\in\mathscr{S}(\mathbb{R}^4)$ with $\phi(0) = 0$, there exists $\mathscr{S}(\mathbb{R}^4)$ functions such that $\phi = k_\mu f^\mu$. Choose a $\mathscr{S}(\mathbb{R}^1)$ function $\rho(u)$ with $\rho(0) = 1$. Define $f^0(k) = k_0^{-1}\phi(k_0, 0, 0, 0)\rho(k_1)\rho(k_2)\rho(k_3)$. Because $\phi(k_0, k_1, 0, 0) - k_0 f^0(k_0, k_1, 0, 0) = 0$ for $k_1 = 0$, one can define $f^1(k) = k_1^{-1}\{\phi(k_0, k_1, 0, 0) - k_0 f^0(k_0, k_1, 0, 0\}\rho(k_2)\rho(k_3)$. Now $\phi(k_0, k_1, k_2, 0) - k_0 f^0(k_0, k_1, k_2, 0) - k_1 f^1(k_0, k_1, k_2, 0) = 0$ for $k_2 = 0$, so one defines $f^2(k) = k_2^{-1}\{\phi(k_0, k_1, k_2, 0) - k_0 f^0(k_0, k_1, k_2, 0) - k_1 f^1(k_0, k_1, k_2, 0)\}\rho(k_3)$. Finally because $\phi(k) - k_0 f^0(k) - k_1 f^1(k) - k_2 f^2(k) = 0$, for $k_3 = 0$ one defines $f^3(k) = k_3^{-1}(\phi - k_0 f^0 - k_1 f^1 - k_2 f^2)$. Q.E.D.

Note that this implies also that $\mathscr{S}^{(4)}_0$, just at $\mathscr{S}^{(2)}_0$ and $\mathscr{S}^{(3)}_0$ is a closed subspace.

We define *linear maps* between the spaces $\mathscr{S}^{(j)}$; d_{21}: $\mathscr{S}^{(1)}\to\mathscr{S}^{(2)}$ by $\chi^{\mu\nu\rho}\mapsto -3\partial_\rho\chi^{\mu\nu\rho}$, d_{32}: $\mathscr{S}^{(2)}\to\mathscr{S}^{(3)}$ by $\psi^{\mu\nu}\mapsto 2\partial_\nu\psi^{\mu\nu}$, and d_{43}: $\mathscr{S}^{(3)}\to\mathscr{S}^{(4)}$ by $f^\mu\mapsto -\partial_\mu f^\mu$. These maps are continuous and, moreover, homomorphisms in the sense of topological vector spaces, because the $\mathscr{S}^{(j)}$ are Fréchet spaces, the $\mathscr{S}^{(j)}_0$ closed subspaces. We have in fact $\mathscr{S}^{(2)}_0 = \mathrm{Im}d_{21} = \mathrm{Ker}d_{32}$, $\mathscr{S}^{(3)}_0 = \mathrm{Im}d_{32} = \mathrm{Ker}d_{43}$, and $\mathscr{S}^{(4)}_0 = \mathrm{Im}d_{43}$. Observe that d_{32} was denoted as d in I and that the definition of the d_{jk} is such that the transposed maps d'_{jk}: $\mathscr{S}'^{(j)}\to\mathscr{S}'^{(k)}$ correspond to exterior differentiation of distributions written heuristically as $d'_{21}: F_{\mu\nu}\mapsto\partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu}$, $d'_{32}: A_\mu\mapsto\partial_\mu A_\nu - \partial_\nu A_\mu$, and $d'_{43}: \phi\mapsto\partial_\mu\phi$.

We shall make frequent use of the following lemmas containing more or less standard results from the theory of topological vector spaces.

Lemma 3.1: Let $\mathscr{V}_1, ..., \mathscr{V}_n$ be closed subspaces of the nuclear spaces $\mathscr{W}_1, ..., \mathscr{W}_n$. The space $\mathscr{V}_1 \hat{\otimes}\cdots\hat{\otimes} \mathscr{V}_n$ can then be identified algebraically and topolocially with a closed subspace of $\mathscr{W}_1 \hat{\otimes}\cdots\hat{\otimes} \mathscr{W}_n$.

This is a property of the ϵ-tensor product topology; see Ref. 16, the corollary of Proposition 43.7.

Lemma 3.2: Let $\mathscr{V}_1, ..., \mathscr{V}_n$ be nuclear Fréchet spaces; then we have a canonical identification $(\mathscr{V}_1 \hat{\otimes}\cdots\hat{\otimes} \mathscr{V}_n)' \cong \mathscr{V}'_1 \hat{\otimes}\cdots\hat{\otimes} \mathscr{V}'_n$, with respect to strong dual topologies. See Ref. 16, Proposition 50.7.

Lemma 3.3: Let $\mathscr{V}$ and $\mathscr{W}$ be Fréchet spaces and T: $\mathscr{V}\to\mathscr{W}$ a continuous linear map with closed image. The T is a homomorphism, $\mathrm{Ker}T' = (\mathrm{Im}T)^\perp$, and $\mathrm{Im}T' = (\mathrm{Ker}T)^\perp$. (with T' the dual map $\mathscr{W}'\to\mathscr{V}'$).

For a proof see Ref. 19, 7.7 and for the last part of the statement, I, the proof of Theorem 3.

Lemma 3.4: Let $\mathscr{V}_1, ..., \mathscr{V}_n$ and $\mathscr{W}_1, ..., \mathscr{W}_n$ be nuclear Fréchet spaces and $T_1, ..., T_n$ continuous linear maps; T_j: $\mathscr{V}_j\to\mathscr{W}_j$. If the T_j are, moreover, homomorphisms, then $T_1 \hat{\otimes}\cdots\hat{\otimes} T_n: \mathscr{V}_1 \hat{\otimes}\cdots\hat{\otimes}\mathscr{V}_n \to \mathscr{W}_1 \hat{\otimes}\cdots\hat{\otimes}\mathscr{W}_n$ is a homomorphism with $\mathrm{Im}(T_1 \hat{\otimes}\cdots\hat{\otimes} T_n) = (\mathrm{Im}T_1)\hat{\otimes}\cdots\hat{\otimes}(\mathrm{Im}T_n)$ and $\mathrm{Ker}(T_1 \hat{\otimes}\cdots\hat{\otimes} T_n)$ the closed linear subspace spanned by the subspaces

$$\mathscr{V}_1 \hat{\otimes}\cdots\hat{\otimes}\mathscr{V}_{j-1}\hat{\otimes}(\mathrm{Ker}T_j)\hat{\otimes}\mathscr{V}_{j+1}\hat{\otimes}\cdots\hat{\otimes}\mathscr{V}_n, \quad j = 1, 2, ..., n.$$

Proof: This can be assembled from Ref. 16, Proposition 50.1, sub f (for the equality of $\hat{\otimes}_\pi = \hat{\otimes}_\epsilon = \hat{\otimes}$), Propositions 43.6, 43.7, 43.9, and exercise 43.2 (for the properties of tensor product maps) and, of course, with the associativity of tensor products, just as for the preceding lemmas. Because the result in Ref. 16, Exercise 43.2, is rather crucial and no proof is given, we provide one for the special case considered here:

Denote $\mathscr{V}_1 \hat{\otimes}\cdots\hat{\otimes}\mathscr{V}_j\hat{\otimes}(\mathrm{Ker}T_j)\hat{\otimes}\mathscr{V}_{j+1}\hat{\otimes}\cdots\hat{\otimes}\mathscr{V}_n$ by $\mathscr{U}_j$; Let $F\in(\mathscr{U}_1 + \cdots + \mathscr{U}_n)^\perp \subset (\mathscr{V}_1\hat{\otimes}\cdots\hat{\otimes}\mathscr{V}_n)'$; then $\phi_F(T_1 v_1, ..., T_n v_n) = F(v_1\otimes\cdots\otimes v_n)$ defines an n-linear map ϕ_F: $\mathrm{Im}T_1\times\cdots\times\mathrm{Im}T_n\to\mathbb{C}$. Because of the continuity of F there exist neighborhoods of zero $\mathscr{B}_j\subset\mathscr{V}_j$ such that $F(v_1\otimes\cdots\otimes v_n)| \leqslant 1$ for $v_j\in\mathscr{B}_j, j = 1, ..., n$. The T_j are homomorphisms, i.e., relatively open, so there exist neighborhoods of zero $\mathscr{C}_j\subset\mathscr{W}_j$, such that $(\mathscr{C}_j\cap T_j(\mathscr{V}_j))\subset T_j(\mathscr{B}_j)$; then one has $|\phi_F(w_1, ..., w_n)|\leqslant 1$ for $w_j\in\mathscr{C}_j\cap T_j(\mathscr{V}_j), j = 1, ..., n$, which proves that ϕ_F is continuous. With ϕ_F there corresponds a continuous linear map from $(\mathrm{Im}T_1)\hat{\otimes}\cdots\hat{\otimes}(\mathrm{Im}T_n)$ to which after identification of $(\mathrm{Im}T_1)\hat{\otimes}\cdots\hat{\otimes}(\mathrm{Im}T_n)$ with a closed subspace of $\mathscr{W}_1\hat{\otimes}\cdots\hat{\otimes}\mathscr{W}_n$ can be extended to an element G of $(\mathscr{W}_1\hat{\otimes}\cdots\hat{\otimes}\mathscr{W}_n)'$, because of the Hahn–Banach theorem. One verifies that $(T_1\hat{\otimes}\cdots\hat{\otimes}T_n)'G = F$, so we have $(\mathscr{U}_1 + \cdots + \mathscr{U}_n)^\perp\subset\mathrm{Im}(T_1\hat{\otimes}\cdots\hat{\otimes}T_n)'$. The inclusion in the other direction is obvious. From $(\mathscr{U}_1 + \cdots + \mathscr{U}_n)^\perp$

$= \mathrm{Im}(T_1 \hat{\otimes} \cdots \hat{\otimes} T_n)'$ one then obtains finally $\mathrm{Ker}(T_1 \hat{\otimes} \cdots \hat{\otimes} T_n) = (\mathscr{U}_1 + \cdots + \mathscr{U}_n)$. Q.E.D.

A set of n-point functions for the $A_\mu(x)$ field is a state ω^A on the Borchers algebra $\mathscr{A}^A$, a set for the corresponding $F_{\mu\nu}(x)$ field is a state ω^F on $\mathscr{A}^F$, with $\omega^F = \theta'_d \omega^A$. We can reformulate the definition of gauge equivalence as follows: Two states ω_1^A and ω_2^A on $\mathscr{A}^A$ are *gauge equivalent* whenever $\omega_1^A - \omega_2^A \in \mathrm{Ker}\theta'_d$. With this definition and the observation that the operation $\phi^{(n,j)}_{\mu_1\cdots\mu_n}(x_1\cdots x_n) \to \partial^j_{\mu_j} \phi^{(n,j)}_{\mu_1\cdots\mu_j,\mu_j,\ldots\mu_n}(x_1\cdots x_n)$ can be written rigorously as the map

$$((\hat{\otimes}^{j-1} 1_{\mathscr{S}^{(3)}}) \hat{\otimes} d_{43} \hat{\otimes} (\hat{\otimes}^{n-j} 1_{\mathscr{S}^{(3)}}))':$$
$$((\hat{\otimes}^{j-1} \mathscr{S}^{(3)}) \hat{\otimes} \mathscr{S}^{(4)} \hat{\otimes} (\hat{\otimes}^{n-j} \mathscr{S}^{(3)}))' \to (\hat{\otimes}^n \mathscr{S}^{(3)})'$$

we formulate

Theorem 3.1 (rigorous version): Two states ω_1^A and ω_2^A on $\mathscr{A}^A$ are gauge equivalent if and only if, for every $n = 1, 2, \cdots$, the restriction of $\omega_1^A - \omega_2^A$ to $(\hat{\otimes}^n \mathscr{S}^{(3)})'$, can be written as a finite sum of vectors from the subspaces $\mathrm{Im}(\hat{\otimes}^{j-1} 1_{\mathscr{S}^{(3)}}) \hat{\otimes} d_{43} \hat{\otimes} (\hat{\otimes}^{n-j} 1_{\mathscr{S}^{(3)}}))'$, $j = 1,2,\ldots,n$. ($1_{\mathscr{S}^{(3)}}$ denotes the identity map on $\mathscr{S}^{(3)}$.)

Proof: (a) $\omega_1^A - \omega_2^A \in \mathrm{Ker}\theta'_d \Leftrightarrow \forall\ n = 1, 2, \cdots$; the restriction of $\omega_1^A - \omega_2^A$ to $\hat{\otimes}^n \mathscr{S}^{(3)}$ is in

$$\mathrm{Ker}(\hat{\otimes}^n d_{32})' = (\mathrm{Im}(\hat{\otimes}^n d_{32}))^\perp = (\hat{\otimes}^n \mathrm{Im} d_{32})^\perp = (\hat{\otimes}^n \mathscr{S}_0^{(3)})^\perp.$$

$$\mathrm{Im}((\hat{\otimes}^{j-1} 1_{\mathscr{S}^{(3)}}) \hat{\otimes} d_{43} \hat{\otimes} (\hat{\otimes}^{n-j} 1_{\mathscr{S}^{(3)}}))'$$
$$= (\mathrm{Ker}((\hat{\otimes}^{j-1} 1_{\mathscr{S}^{(3)}}) \hat{\otimes} d_{43} \hat{\otimes} (\hat{\otimes}^{n-j} 1_{\mathscr{S}^{(3)}})))^\perp$$
$$= ((\hat{\otimes}^{j-1} \mathscr{S}^{(3)} \hat{\otimes} \mathrm{Ker}\, d_{43} \hat{\otimes} (\hat{\otimes}^{n-j} \mathscr{S}^{(3)}))^\perp$$
$$= ((\hat{\otimes}^{j-1} \mathscr{S}^{(3)}) \hat{\otimes} \mathscr{S}_0^{(3)} \hat{\otimes} (\hat{\otimes}^{n-j} \mathscr{S}^{(3)}))^\perp.$$

So one has to prove, for $n = 2, 3, \cdots$, that

$$(\hat{\otimes}^n \mathscr{S}^{(3)})^\perp = \sum_{j=1}^{n} ((\hat{\otimes}^{j-1} \mathscr{S}^{(3)}) \hat{\otimes} \mathscr{S}_0^{(3)} \hat{\otimes} (\hat{\otimes}^{n-1} \mathscr{S}^{(3)}))^\perp,$$

as a subspace of $(\hat{\otimes}^n \mathscr{S}^{(3)})'$; the summation on the right-hand side means considering finite sums.

(b) We consider the case $n = 2$ and show first that

$$\mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)} = (\mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}^{(3)}) \cap (\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)}).$$

From

$$\mathscr{S}_0^{(3)} \otimes \mathscr{S}_0^{(3)} = (\mathscr{S}_0^{(3)} \otimes \mathscr{S}^{(3)}) \cap (\mathscr{S}^{(3)} \otimes \mathscr{S}_0^{(3)})$$

one obtains

$$\mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)} \subset (\mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}^{(3)}) \cap (\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)}).$$

For the inclusion in the other direction one considers an element

$$\psi_{33} \in (\mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}^{(3)}) \cap (\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)});$$
$$\psi_{33} \in \mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}^{(3)} = \mathrm{Im}(d_{32} \hat{\otimes} 1_{\mathscr{S}^{(3)}});$$
$$\Rightarrow \exists h_{23} \in \mathscr{S}^{(2)} \hat{\otimes} \mathscr{S}^{(3)},$$

such that $\psi_{33} = (d_{32} \hat{\otimes} 1_{\mathscr{S}^{(3)}}) h_{23}$. Because also

$$\psi_{33} \in \mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)} = \mathrm{Ker}(1_{\mathscr{S}^{(3)}} \hat{\otimes} d_{43}),$$

one has

$$(d_{32} \hat{\otimes} d_{43}) h_{23} = (1_{\mathscr{S}^{(3)}} \hat{\otimes} d_{43})(d_{32} \hat{\otimes} 1_{\mathscr{S}^{(3)}}) \psi_{33} = 0,$$

so that

$$h_{23} \in \overline{\mathscr{S}_0^{(2)} \hat{\otimes} \mathscr{S}^{(3)} + \mathscr{S}^{(2)} \hat{\otimes} \mathscr{S}_0^{(3)}}.$$

From $(d_{32} \hat{\otimes} 1_{\mathscr{S}^{(3)}})(\mathscr{S}_0^{(2)} \hat{\otimes} \mathscr{S}^{(3)}) = 0$ one has

$$(d_{32} \hat{\otimes} 1_{\mathscr{S}^{(3)}})(\mathscr{S}_0^{(2)} \hat{\otimes} \mathscr{S}^{(3)} + \mathscr{S}^{(2)} \hat{\otimes} \mathscr{S}_0^{(3)})$$
$$= (d_{32} \hat{\otimes} 1_{\mathscr{S}^{(3)}})(\mathscr{S}^{(2)} \hat{\otimes}\, {}_0^{(3)})$$
$$= \mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)},$$

and, because this is closed in $\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}^{(3)}$,

$$(d_{32} \hat{\otimes} 1_{\mathscr{S}^{(3)}})\ \overline{(\mathscr{S}_0^{(2)} \hat{\otimes} \mathscr{S}^{(3)} + \mathscr{S}^{(2)} \hat{\otimes} \mathscr{S}_0^{(3)})}$$
$$= \mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)},$$

and therefore

$$\psi_{33} = (d_{32} \hat{\otimes} 1_{\mathscr{S}^{(3)}})(h_{23} \in \mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)}).$$ Q.E.D

(c) One has in general for closed subspaces $\mathscr{W}_1$, $\mathscr{W}_2$ of a locally convex topological vector space $\mathscr{W}$ that

$$\mathscr{W}_1 \cap \mathscr{W}_2 = (\mathscr{W}_1^\perp)^\perp \cap (\mathscr{W}_2^\perp)^\perp = (\mathscr{W}_1 + \mathscr{W}_2)^\perp,$$

and therefore

$$(\mathscr{W}_1 \cap \mathscr{W}_2)^\perp = \overline{(\mathscr{W}_1^\perp + \mathscr{W}_2^\perp)}$$

(for properties of polars, see Ref. 18, Corollary 1 of Theorem 4, Chap. II). This gives us

$$(\mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)})^\perp = \overline{((\mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}^{(3)}) \cap (\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)}))^\perp}$$
$$= \overline{((\mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}^{(3)})^\perp + (\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)})^\perp)}.$$

(d) We have to show next that $(\mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}^{(3)})^\perp + (\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)})^\perp$ is a closed subspace in $(\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}^{(3)})'$ (in weak and strong dual topologies), and therefore we show first that $\mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}^{(3)} + \mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)}$ is closed in $\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}^{(3)}$. The map $d_{43} \hat{\otimes} 1_{\mathscr{S}^{(3)}} : \mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}^{(3)} \to \mathscr{S}^{(4)} \hat{\otimes} \mathscr{S}^{(3)}$ maps $\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)}$ into the closed subspace $\mathscr{S}_0^{(4)} \hat{\otimes} \mathscr{S}_0^{(3)}$, because by restricting $d_{43} \hat{\otimes} 1_{\mathscr{S}^{(3)}}$ to $\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}^{(3)}$ and using the identification properties of Lemma 3.1 one is allowed to apply Lemma 3.4. One has $\mathrm{Ker}(d_{43} \hat{\otimes} 1_{\mathscr{S}^{(3)}}) = \mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}^{(3)}$, and therefore

$$\mathscr{S}_0^{(3)} \otimes \mathscr{S}^{(3)} + \mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)}$$
$$= \mathrm{Ker}(d_{43} \hat{\otimes} 1_{\mathscr{S}^{(3)}}) + \mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)}$$
$$= (d_{43} \hat{\otimes} 1_{\mathscr{S}^{(3)}})^{-1}(d_{43} \hat{\otimes} 1_{\mathscr{S}^{(3)}})(\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)})$$
$$= (d_{43} \hat{\otimes} 1_{\mathscr{S}^{(3)}})^{-1}(\mathscr{S}_0^{(4)} \hat{\otimes} \mathscr{S}_0^{(3)}),$$

and this is a closed subspace of $\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}^{(3)}$. Consider next the map $(d_{32} \hat{\otimes} 1_{\mathscr{S}^{(3)}})' : (\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}^{(3)}) = \to (\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}^{(3)})'$. Because $\mathrm{Ker}(d_{32} \hat{\otimes} 1_{\mathscr{S}^{(3)}})' = (\mathrm{Im}(d_{32} \hat{\otimes} 1_{\mathscr{S}^{(3)}}))^\perp = (\mathscr{S}_0^{(3)} \hat{\otimes}\, {}^{(3)})^\perp$ one has $(\mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}^{(3)})^\perp + (\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)})^\perp = [(d_{32} \hat{\otimes} 1_{\mathscr{S}^{(3)}})']^{-1} \times (d_{32} \hat{\otimes} 1_{\mathscr{S}^{(3)}})' [(\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)})^\perp]$. It is therefore sufficient to show that $(d_{32} \otimes 1_{\mathscr{S}^{(3)}})' [(\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)\perp}]$ is closed in $(\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}^{(3)})'$. The map $d_{32} \hat{\otimes} 1_{\mathscr{S}^{(3)}} : \mathscr{S}^{(2)} \hat{\otimes} \mathscr{S}^{(3)} \to \mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}^{(3)}$ induces a continuous linear map T from the quotient space $(\mathscr{S}^{(2)} \hat{\otimes} \mathscr{S}^{(3)})/(\mathscr{S}^{(2)} \hat{\otimes} \mathscr{S}_0^{(3)})$ to the quotient space $(\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}^{(3)})/(\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)})$. $\mathrm{Im}\, T = (\mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}^{(3)})/(\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)})$ viewed as a subspace of $(\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}^{(3)})/(\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)})$; it is the same as $(\mathscr{S}_0^{(3)} \hat{\otimes} \mathscr{S}^{(3)} + \mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)})/(\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)})$ and is therefore closed. Consequently, T is a homomorphism between Fréchet spaces. Going over to duals, we know that $T' : ((\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}^{(3)})/(\mathscr{S}^{(3)} \hat{\otimes} \mathscr{S}_0^{(3)}))' \to ((\mathscr{S}^{(2)} \hat{\otimes} \mathscr{S}^{(3)})/(\mathscr{S}^{(2)} \hat{\otimes} \mathscr{S}_0^{(3)}))'$ has closed range (weakly and strongly). We have for $j = 2,3$ the canonical isomorphisms

$$\{(\mathscr{S}^{(j)}\hat{\otimes}\mathscr{S}^{(3)})/(\mathscr{S}^{(j)}\hat{\otimes}\mathscr{S}_0^{(3)})\}'$$
$$=(\mathscr{S}^{(j)}\hat{\otimes}\mathscr{S}_0^{(3)})^{\perp}\subset(\mathscr{S}^{(j)}\hat{\otimes}\mathscr{S}^{(3)})',$$

which are topological with respect to the weak duals [see Ref. 17, Sec. 17.14 (ii)]. This can be used to identify T' with the restriction of $(d_{32}\hat{\otimes}1_{\mathscr{S}^{(3)}})'$ to $(\mathscr{S}^{(3)}\hat{\otimes}\mathscr{S}_0^{(3)})^{\perp}$ and from this one sees that $(d_{32}\hat{\otimes}1_{\mathscr{S}^{(3)}})'\,(\mathscr{S}^{(3)}\hat{\otimes}\mathscr{S}_0^{(3)})^{\perp})$ is closed in $(\mathscr{S}^{(2)}\hat{\otimes}\mathscr{S}^{(3)})'$.

(e) From (c) and (d) we have $(\mathscr{S}_0^{(3)}\hat{\otimes}\mathscr{S}_0^{(3)})^{\perp}$ $=(\mathscr{S}_0^{(3)}\hat{\otimes}\mathscr{S}^{(3)})^{\perp}+(\mathscr{S}^{(3)}\hat{\otimes}\mathscr{S}_0^{(3)})^{\perp}$. In a completely analogous way one shows that

$$(\mathscr{V}_1\hat{\otimes}\mathscr{S}_0^{(3)}\hat{\otimes}\mathscr{V}_2\hat{\otimes}\mathscr{S}_0^{(3)}\hat{\otimes}\mathscr{V}_3)^{\perp}$$
$$=(\mathscr{V}_1\hat{\otimes}\mathscr{S}_0^{(3)}\hat{\otimes}\mathscr{V}_2\hat{\otimes}\mathscr{S}^{(3)}\hat{\otimes}\mathscr{V}_3)^{\perp}$$
$$+(\mathscr{V}_1\hat{\otimes}\mathscr{S}^{(3)}\hat{\otimes}\mathscr{V}_2\hat{\otimes}\mathscr{S}_0^{(3)}\hat{\otimes}\mathscr{V}_3)^{\perp}$$

for arbitrary nuclear Fréchet spaces $\mathscr{V}_j, j=1,2,3$. This can be used to transform $(\hat{\otimes}\mathscr{S}_0^{(3)})^{\perp}$ into

$$\sum_{j=1}^{n}((\hat{\otimes}^{j-1}\mathscr{S}^{(3)})\hat{\otimes}\mathscr{S}_0^{(3)}\hat{\otimes}(\hat{\otimes}^{n-j}\mathscr{S}^{(3)}))^{\perp},$$

in $n-1$ steps:

$$(\hat{\otimes}^n\mathscr{S}_0^{(3)})^{\perp}=(\mathscr{S}_0^{(3)}\hat{\otimes}\mathscr{S}_0^{(3)}\hat{\otimes}(\hat{\otimes}^{n-2}\mathscr{S}_0^{(3)}))^{\perp}$$
$$=(\mathscr{S}_0^{(3)}\hat{\otimes}\mathscr{S}^{(3)}\hat{\otimes}(\hat{\otimes}^{n-2}\mathscr{S}_0^{(3)}))^{\perp}$$
$$+(\mathscr{S}^{(3)}\hat{\otimes}\mathscr{S}_0^{(3)}\hat{\otimes}(\hat{\otimes}^{n-2}\mathscr{S}_0^{(3)}))^{\perp}$$

etc., for general n, and this completes, according to (a), the proof of the theorem.

IV. REMARKS ON GAUGES AND GAUGE TRANSFORMATIONS

All the different gauges for a given $F_{\mu\nu}(x)$ theory are physically completely equivalent. Choosing a gauge is a matter of mathematical convenience only. The variety of gauges allowed by Theorem 3.1 is very large, but only a small number characterized by simple general conditions will be used in practice.

A gauge for a Lorentz covariant $F_{\mu\nu}(x)$ theory is not necessarily itself Lorentz-covariant. It is therefore obvious to require this as an additional simplifying property. According to the discussion at the end of Sec. II this means Lorentz invariance of the state ω^A or equivalently the existence in the GNS state space of a representation of the Lorentz group by isometric operators, transforming the field operator $A_\mu(x)$ in the proper manner.

The positive-definite metric of the $F_{\mu\nu}(x)$ theory does not imply positivity of its gauges, and again it is natural, in order to obtain a Hilbert space for the $A_\mu(x)$ field, to impose this as an extra condition.

The fundamental difficulty which complicates the field-theoretic description of photons is an incompatibility between these two conditions. At the level of generality where we are at this point and where we have not yet used anything corresponding to the second Maxwell Equation $\partial^\nu F_{\mu\nu}=J_\mu$, this does not show up. In Sec. VII we shall give a rigorous proof of this incompatibility for the case of the free field, $\partial^\nu F_{\mu\nu}=0$.

In classical theory one uses as a convenient restriction the Lorentz gauge condition $\partial^\mu A_\mu=0$. It simplifies the second Maxwell equation for the potentials to the wave equation $\partial^\mu\partial_\mu A_\nu=J_\nu$. The analog for the quantized photon field is $\partial^\mu A_\mu=0$, as an operator relation for the field operator $A_\mu(x)$, or equivalently a condition for the state ω^A that can be written in terms of n-point functions as $\partial^{\mu_j}\omega^A_{\mu_1\cdots\mu_n}(x_1,\ldots,x_n)=0$, $\forall\, n=1,2,\ldots,\forall\, j=1,2,\ldots,n$. The resulting situation is, however, more complicated than in the classical case. The wave equation does not necessarily hold for the field operator $A_\mu(x)$, as we shall show in Sec. VII for the free field. In fact in that case the condition $\partial^\mu\partial_\mu A_\nu=0$ gives rise to a different important set of Lorentz-covariant gauges.

We have a precise and satisfactory definition of gauge and of equivalence between gauges. No such clearcut meaning can be given to the term *gauge transformation*. There are several distinct concepts, all having to do with the transformation of gauges into equivalent ones and playing a role somewhere in the formalism. We shall briefly discuss some of the possibilities.

(a) From our algebraic point of view it is natural to define a gauge transformation as a continuous $*$-automorphism α of $\mathscr{A}^A$ onto $\mathscr{A}^A$, with the inverse having the same properties, and such that every state ω^A on $\mathscr{A}^A$ is mapped by α', the dual of α, onto a state $\alpha'\omega^A$ that is gauge equivalent to ω^A. A simple characterization for this is $\alpha\theta_d=\theta_d$. (θ_d is the basic homomorphism from $\mathscr{A}^F$ into $\mathscr{A}^A$, generated by d_{32}.)

The scheme developed at the end of Sec. II allows us to identify the GNS state spaces of equivalent gauges connected in this manner and reduce the action of a gauge transformation in this sense to the introduction of a transformed field operator $A_\mu^{(\alpha)}(x)$ in the state space of the given $A_\mu(x)$ operator. Take $\mathscr{A}_1=\mathscr{A}_2=\mathscr{A}^A$, $\omega_2=\omega^A$, $\omega_1=\alpha'\omega^A$, $\mathscr{H}_2=\mathscr{H}^A$, the GNS space of ω^A and $\mathscr{H}_1=\mathscr{H}^{A(\alpha)}$, the GNS space of $\alpha'\omega^A$. The "partial isometry" W is a topological isomorphism from $\mathscr{H}^A$ into $\mathscr{H}^{A(\alpha)}$ and can be used to identify $\mathscr{H}^{A(\alpha)}$ with $\mathscr{H}^A$, and $\Omega^{A(\alpha)}$ with Ω^A. In $\mathscr{H}^A$ we have the representation π of $\mathscr{A}^A$, connected with ω^A, but also the "gauge transformed" representation $\pi^{(\alpha)}$, coming from $\mathscr{H}^{A(\alpha)}$. Due to the intertwining property of W, it can be written as $\pi^{(\alpha)}(a)=\pi(\alpha(a))$, $\forall\, a\in\mathscr{A}^A$. This means in particular for the field operator: $A^{(\alpha)}(f)=\pi(\alpha(f))$, $\forall\, f\in\mathscr{S}^{(3)}\subset\mathscr{A}^A$.

The transformed field operator $A^{\alpha}(f)$ may be written as $A^{(\alpha)}(f)=A(f)+D(f)$. The difference term $D(f)=A^{(\alpha)}(f)-A(f)$ is in fact a gradient. In heuristic language where $D(f)=\int D_\mu(x)f^\mu(x)\,dx$, this means that $D_\mu(x)=\partial_\mu\phi(x)$, for some operator field $\phi(x)$. To prove this, one must show, in a more rigorous formulation, the existence, $\forall\, g\in\mathscr{S}^{(4)}$, of a continuous linear operator $\phi(g)$: $\mathscr{H}^A\to\mathscr{H}^A$, with a continuous linear dependence on g, and such that $D(f)=\phi(d_{43}f)$, $\forall\, f\in\mathscr{S}^{(3)}$. Note first that $\mathscr{S}_0^{(4)}=\mathrm{Im}\,d_{43}$ has finite codimension (see Sec. III). There exists therefore a continuous projection P_0 in $\mathscr{S}^{(4)}$ on $\mathscr{S}_0^{(4)}$ (see Ref. 16, proposition 9.3). Note also that, $\forall\, f\in\mathscr{S}_0^{(3)}=\mathrm{Ker}\,d_{43}=\mathrm{Im}\,d_{32}$, $D(f)=0$, because then $D(f)=\pi^{(\alpha)}(f)-\pi(f)=\pi((\alpha-1)f)=\pi((\alpha-1)d_{32}\psi)$, for some $\psi\in\mathscr{S}^{(2)}$, and this is zero because $(\alpha-1)\theta_d=0$, the characteristic property of a gauge automorphism. These two results together imply that, $\forall\, g\in\mathscr{S}^{(4)}$, $\phi(g):=D(d_{43}^{-1}P_0 g)$ is a well-defined continuous operator in $\mathscr{H}^A$. Of course, one has then $\phi(d_{43}f) = D(g$. Moreover, ϕ depends continuously on $g\in\mathscr{S}^{(4)}$. To see this, consider for each fixed $\Omega\in\mathscr{H}^A$ the continuous map $\mathscr{S}^{(3)}\to\mathscr{H}^A$ defined by $f\mapsto D(f)\Omega$. Because $D(f) = 0$ on $\mathscr{S}_0^{(3)}$, this gives rise to a continuous linear map $\mathscr{S}^{(3)}/\mathscr{S}_0^{(3)}\to\mathscr{H}^A$. The map d_{43} is a topological homomorphism; therefore, $\mathscr{S}^{(3)}/\mathscr{S}_0^{(3)}$ is topologically isomorphic with $\mathscr{S}_0^{(4)}$ provided with its relative topology as a subspace of $\mathscr{S}^{(4)}$. This gives a continuous map $\mathscr{S}_0^{(4)}\to\mathscr{H}^A$. By composition with the projection P_0 one obtains finally a continuous linear map $\mathscr{S}^{(4)}\to\mathscr{H}^A$, which is just $g\mapsto\phi(g)\Omega$. In this way we have shown that gauge automorphisms can be realized as field operator transformations of the special "classical" form: $A'_\mu(x) = A_\mu(x) + \partial_\mu\phi(x)$, with $\phi(x)$ an operator field.

Examples of gauge automorphisms are readily available. Let T be a $*$-preserving, linear topological isomorphism of $\mathscr{S}^{(3)}$ onto itself with $(T-1)d_{32} = 0$. The $*$-automorphism α_T, as defined in Sec. II, is then a gauge automorphism. For an explicit case take T, in terms of Fourier transforms $((Tf)\hat{\ })^\mu(k) = \hat{f}^\mu(k) + C^\mu(k)K_\nu\hat{f}^\nu(k)$, with $C^\mu(k)$ a polynomially bounded vectorial C^∞ function, satisfying $k_\mu C^\mu(k) = 0$, $\overline{C^\mu(k)} = -C^\mu(-k)$. [The inverse is then $((T^{-1})\hat{\ }f)^\mu(k) = \hat{f}^\mu(k) - C^\mu(k)k_\nu\hat{f}^\nu(k)$.] A second class of examples consist of special "shift" automorphisms α_λ (see Sec. II) with $\lambda\in\mathrm{Im}d'_{43}\subset\mathscr{S}^{(3)\prime}$. This means that λ is a vectorial distribution of the form $\partial_\mu\phi(x)$, with $\phi(x)$ a real scalar tempered distribution. A gauge automorphism of this type is in terms of transformation of the field operator $A_\mu(x)$ just addition of a "c-number" gradient term $\partial_\mu\phi(x)$.

(b) A wider concept of algebraic gauge transformation is obtained by dropping the invertibility of α. A gauge transformation is then a continuous $*$-preserving algebraic homomorphism α of $\mathscr{A}^A$ into itself with the property $\alpha\theta_d = \theta_d$. As an example one may consider an α_T as given in (a), but without the condition $k_\mu C^\mu(k) = 0$ for $C^\mu(k)$. There is again a certain linear correspondence between the GNS space $\mathscr{H}^A$ of a gauge ω^A and the GNS space $\mathscr{H}^{A(\alpha)}$ of an equivalent, transformed state $\alpha'\omega^A$. However, in general the linear isometry W is not injective, not defined on all of $\mathscr{H}^A$, and therefore cannot be used to identify the spaces $\mathscr{H}^A$ and $\mathscr{H}^{A(\alpha)}$. The field operator $A_\mu(x)$ and the transformed operator $A_\mu^{(\alpha)}(x)$ remain in different spaces, and it does not even make sense to ask whether $A_\mu^{(\alpha)}(x) - A_\mu(x)$ can be written as a gradient. [One might be tempted to define in $\mathscr{H}^A$ as transformed field $A'(f) = \pi(\alpha f)$. However, unless W is a topological isomorphism of $\mathscr{H}^A$ onto $\mathscr{H}^{A(\alpha)}$, the corresponding representation is *not* the GNS representation associated with the gauge transformed state $\alpha'\omega^A$.]

It may happen, as a special case, that a gauge homomorphism α gives an adjoint α' that is invertible on a subset of states only. If there are appropriate continuity properties, this will again lead to the situation of (a), i.e., topological isomorphisms W between representation spaces and, after suitable identifications, gauge transformations as transformations of the field operator in one representation space, but all this only for a restricted set of gauges. An example of this will occur in Sec. VII where gauge homomorphisms α_T, with T given by $C^\mu(k) = \lambda k^\mu$, connect the so-called generalized Gupta–Bleuler gauges.

(c) Several other ways of connecting equivalent gauges can be detected in the literature, some of which are essentially further generalizations of case (b); see in particular Ref. 2. They all have in common that they cannot be realized by addition of a gradient term to the field operator $A_\mu(x)$ and involve distinct representation spaces that cannot be identified.

V. THE FREE FIELD

A conspicuous feature of the field-theoretic description of free photons is the lack of uniqueness for the field operators $A_\mu(x)$. This is in strong contrast to the situation for other particles. In the case of massive spinless particles, for instance, it can be proved that the standard free field theory is the only one that meets the requirements of the Wightman axioms combined with the free Klein–Gordon equation for the field operator $\phi(x)$.

The photon case follows this general pattern as long as one considers only the "physical" field $F_{\mu\nu}(x)$. There exists a system of Wightman functions $\omega^F_{\mu_1\nu_1\cdots\mu_n\nu_n}(x_1,\ldots,x_n)$, made up in the usual way from a two-point function which is

$$\omega^F_{\mu_1\nu_1\mu_2\nu_2}(x_1,x_2) = \epsilon^{\rho_1\tau_1}_{\mu_1\nu_1}\,\epsilon^{\rho_2\tau_2}_{\mu_2\nu_2}\,\partial^1_{\rho_1}\partial^2_{\rho_2}(-ig_{\tau_1\tau_2}D^{(+)}(x_1-x_2)) \tag{5.1}$$

with

$$D^{(+)}(x_1-x_2) = \Delta^{(+)}(x_1-x_2, m=0) = -\frac{i}{(2\pi)^3}\int_{k^0=|\mathbf{k}|}\frac{d\mathbf{k}}{2k^0}e^{-ik(x_1-x_2)}$$

and with

$$|\mathbf{k}| = \Big(\sum_{j=1}^{3}|k^j|^2\Big)^{1/2},\quad kx = g_{\mu\nu}k^\mu x^\nu,$$

$$g_{00} = 1,\quad g_{jj} = -1\quad \text{for } j = 1,2,3.$$

The corresponding operator field theory satisfies all standard Wightman axioms (including positivity for the inner product of the state space) together with the free Maxwell equations $\partial^\mu F_{\mu\nu} = 0$ and $\partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu} = 0$ as operator equations. It is the only $F_{\mu\nu}(x)$ theory known that has these properties and can therfore be regarded as the proper description of free photons by means of a $F_{\mu\nu}(x)$ field.

The discussion of the variety of different descriptions in terms of $A_\mu(x)$ fields will be based on this unique free $F_{\mu\nu}(x)$ theory. This amounts to an investigation of the possible gauges for this $F_{\mu\nu}(x)$ theory, in the sense discussed in the preceding sections. In the language of states and algebras it means studying the solutions of the inhomogeneous linear equation $\theta'_d\omega^A = \omega^F$, for ω^F given explicitly, essentially by formula (5.1).

There is an obvious particular solution, that is, the state ω^A associated with the standard Gupta–Bleuler formalism, usually called the Feynman gauge and consisting of n-point functions determined by the two-point function

$$\omega^A_{\mu\nu}(x_1,x_2) = -ig_{\mu\nu}D^{(+)}(x_1-x_2). \tag{5.2}$$

According to Theorem 3.1 all other gauges for the free photon field can be obtained from these n-point functions by adding gradient terms

$$\sum_{j=1}^{n} \partial^{j}_{\mu_j} \phi^{(n,j)}_{\mu_1\cdots\mu_{j-1}\mu_{j+1}\cdots\mu_n}(x_1,\dots,x_n),$$

with $\phi^{(n,j)}_{\mu_1\cdots\mu_{j-1}\mu_{j+1}\cdots\mu_n}$ tempered distributions, arbitrary except for a reality requirement following from the condition $\overline{\omega(a^*)} = \omega(a)$, $\forall\, a \in \mathscr{A}^A$.

In this way one has, in principle, a complete description of all possible gauges for the free $A_\mu(x)$ field as sets of n-point functions and consequently by the reconstruction theorem as operator field theories. As argued before, only a limited number of cases from this vast collection have practical value. These are selected by the application of additional requirements not necessary in themselves but leading to mathematically convenient formulations. Some of the obvious requirements are mutually incompatible, so choices have to be made.

A first simplification can be obtained from the requirement that the n-point functions have just as those of the free $F_{\mu\nu}(x)$ theory the free field form, i.e., are based on a two-point function according to

$$\omega^A_{\mu_1\cdots\mu_n}(x_1,\dots,x_n) = \sum \omega^A_{\mu_{j_1}\mu_{j_2}}(x_{j_1},x_{j_2})\cdots\omega^A_{\mu_{j_{n-1}}\mu_{j_n}}(x_{j_{n-1}},x_{j_n}) \tag{5.3}$$

for n even and with summation over all permutations $j_1,\dots,j_n$ of $1,\dots,n$ with $j_1<j_3<\cdots<j_{n-1}$, $j_1<j_2,\dots,j_{n-1}<j_n$, and

$$\omega^A_{\mu_1\cdots\mu_n}(x_1,\dots,x_n) = 0 \tag{5.4}$$

for n odd. Field theories for which the n-point functions have this general form will be called *Gaussian*. By this restriction the discussion of free field gauges is reduced to a discussion of two-point functions, which according to Theorem 3.1 have the form

$$\omega^A_{\mu\nu}(x_1,x_2) = -ig_{\mu\nu}D^{(+)}(x_1-x_2) + \partial^1_\mu \phi^{(1)}_\nu(x_1,x_2) + \partial^2_\nu \phi^{(2)}_\mu(x_1,x_2). \tag{5.5}$$

A second simplification can be obtained from the obvious requirement of translation invariance. Theorem 3.1 does not imply that in that case $\omega^A_{\mu\nu}(x_1-x_2)$ can be written in the form (5.5) with translation invariant distributions $\phi^{(1)}_\mu$, $\phi^{(2)}_\mu$; the situation is slightly more complicated.

Theorem 5.1: The translation-invariant two-point functions for the free $A_\mu(x)$ photon field have the form

$$\omega^A_{\mu\nu}(x_1-x_2) = -ig_{\mu\nu}D^{(+)}(x_1-x_2) + (\partial_\mu\phi_\nu)(x_1-x_2) + \overline{(\partial_\nu\phi_\mu)(x_2-x_1)} + C_{\mu\nu\rho}(x_1^\rho - x_2^\rho) \tag{5.6}$$

with the ϕ_μ arbitrary tempered distributions, the $C_{\mu\nu\rho}$ arbitrary real constants, totally antisymmetric in μ, ν, ρ. For a given $\omega^A_{\mu\nu}(x_1-x_2)$, the $C_{\mu\nu\rho}$ are uniquely determined, the ϕ_μ up to transformation $\phi_\mu \to \phi_\mu + ia_\mu + b_{\mu\nu}x^\nu + \partial_\mu\phi$, with a_μ, $b_{\mu\nu}$ real constants, $b_{\mu\nu} = -b_{\nu\mu}$, and ϕ a tempered distribution with $\overline{\phi(-x)} = -\phi(x)$.

Proof: We give this theorem and its proof in "generalized function" language. In this case rewriting everything with test functions would be elementary and not at all enlightening.

Suppose $\omega^A_{\mu\nu}(x) = \omega^A_{\mu\nu}(x_1-x_2)$ to be a translation invariant two-point function for the free $A_\mu(x)$ field. The expression $f_{\mu\nu}(x) = \omega^A_{\mu\nu}(\hat{x}) + ig_{\mu\nu}D^{(+)}(x)$ then satisfies

$$\partial_\mu(\partial_\rho f_{\nu\sigma} - \partial_\sigma f_{\nu\rho}) - \partial_\nu(\partial_\rho f_{\mu\sigma} - \partial_\sigma f_{\mu\rho}) = 0.$$

Because of our earlier result $\operatorname{Ker} d'_{32} = (\operatorname{Im} d_{32})^\perp = (\operatorname{Ker} d_{43})^\perp = \operatorname{Im} d'_{43}$ this implies, for each pair ρ, σ, the existence of a tempered distribution $\phi_{\rho\sigma}$, such that $\phi_{\rho\sigma} = -\phi_{\sigma\rho}$ and $\partial_\rho f_{\nu\sigma} - \partial_\sigma f_{\nu\rho} = \partial_\nu \phi_{\rho\sigma}$. One has

$$\begin{aligned}\partial_\sigma(\partial_\mu\phi_{\nu\rho} + \partial_\nu\phi_{\rho\mu} + \partial_\rho\phi_{\mu\nu}) &= \partial_\mu(\partial_\nu f_{\sigma\rho} - \partial_\rho f_{\sigma\nu}) + \partial_\nu(\partial_\rho f_{\sigma\mu} - \partial_\mu f_{\sigma\rho}) \\ &\quad + \partial_\rho(\partial_\mu f_{\sigma\nu} - \partial_\nu f_{\sigma\mu}) = 0,\end{aligned}$$

so that there are constants $\alpha_{\mu\nu\rho}$, totally antisymmetric in μ, ν, ρ, such that $\partial_\mu\phi_{\nu\rho} + \partial_\nu\phi_{\rho\mu} + \partial_\rho\phi_{\mu\nu} = \alpha_{\mu\nu\rho}$ or

$$\partial_\mu(\phi_{\nu\rho} - \tfrac{1}{3}\alpha_{\tau\nu\rho}x^\tau) + \partial_\nu(\phi_{\rho\mu} - \tfrac{1}{3}\alpha_{\tau\rho\mu}x^\tau) + \partial_\rho(\phi_{\mu\nu} - \tfrac{1}{3}\alpha_{\tau\mu\nu}x^\tau) = 0,$$

which with $\operatorname{Ker} d'_{21} = (\operatorname{Im} d_{21})^\perp = (\operatorname{Ker} d_{32})^\perp = \operatorname{Im} d'_{32}$ implies the existence of tempered distributions $\phi^{(1)}_\mu$ such that

$$\phi_{\mu\nu} - \tfrac{1}{3}\alpha_{\tau\mu\nu}x^\tau = \partial_\mu\phi^{(1)}_\nu - \partial_\nu\phi^{(1)}_\mu.$$

From this follows

$$\partial_\mu f_{\nu\rho} - \partial_\rho f_{\nu\mu} = \partial_\nu\phi_{\mu\rho} = \partial_\nu\partial_\mu\phi^{(1)}_\rho - \partial_\nu\partial_\rho\phi^{(1)}_\mu + \tfrac{1}{3}\alpha_{\nu\mu\rho}$$

or

$$\partial_\mu(f_{\nu\rho} - \partial_\nu\phi^{(1)}_\rho + \tfrac{1}{6}\alpha_{\nu\rho\tau}x^\tau) - \partial_\rho(f_{\nu\mu} - \partial_\nu\phi^{(1)}_\mu + \tfrac{1}{6}\alpha_{\nu\mu\tau}x^\tau) = 0,$$

which again with $\operatorname{Ker} d'_{21} = \operatorname{Im} d'_{32}$ implies, for each ν, the existence of a tempered distribution $\phi^{(2)}_\nu$ such that $f_{\nu\rho} - \partial_\nu\phi^{(1)}_\rho + \tfrac{1}{6}\alpha_{\nu\rho\tau}x^\tau = \partial_\rho\phi^{(2)}_\nu$, which means that $f_{\mu\nu}$ can be written as $f_{\mu\nu} = \partial_\mu\phi^{(1)}_\nu + \partial_\nu\phi^{(2)}_\mu + \tfrac{1}{6}\alpha_{\mu\nu\rho}x^\rho$. The reality condition for the two-point function implies $\overline{f_{\mu\nu}(x)} = f_{\nu\mu}(-x)$ or $f_{\mu\nu}(x) = \tfrac{1}{2}\big[f_{\mu\nu}(x) + \overline{f_{\nu\mu}(-x)}\big]$. This gives $f_{\mu\nu}(x) = (\partial_\mu\phi_\nu)(x) + \overline{(\partial_\nu\phi_\mu)(-x)} + C_{\mu\nu\rho}x^\rho$ with $\phi_\mu(x) = \tfrac{1}{2}\big[\phi^{(1)}_\mu(x) - \overline{\phi^{(2)}_\mu(-x)}\big]$ and $C_{\mu\nu\rho} = \tfrac{1}{6}\operatorname{Re}\alpha_{\mu\nu\rho}$. This proves that $\omega^A_{\mu\nu}(x)$ has the desired form.

To prove the uniqueness statements, suppose $(\partial_\mu\phi_\nu)(x) + \overline{(\partial_\nu\phi_\mu)(-x)} + C_{\mu\nu\rho}x^\rho = 0$. This implies $\partial_\mu\psi_\nu - \partial_\nu\psi_\mu + 2C_{\mu\nu\rho}x^\rho = 0$ for $\psi_\nu(x) = \phi_\nu(x) + \overline{\phi_\nu(-x)}$, and this gives $\partial_\rho\partial_\mu\psi_\nu - \partial_\rho\partial_\nu\psi_\mu + 2C_{\mu\nu\rho} = 0$. Adding to this the two relations obtained by cyclic permutation of ρ, μ, ν, one obtains $C_{\mu\nu\rho} = 0$ and this shows that the constants $C_{\mu\nu\rho}$ in (5.6) are uniquely determined by $\omega^A_{\mu\nu}(x)$. From $(\partial_\mu\phi_\nu)(x) + \overline{(\partial_\nu\phi_\mu)(-x)} = 0$ one obtains then by using the antisymmetric part, $\partial_\mu\psi_\nu - \partial_\nu\psi_\mu = 0$, which implies the existence of a tempered distribution $\phi(x)$ such that $\psi_\mu(x) = \phi_\mu(x) + \overline{\phi_\mu(-x)} = 2\partial_\mu\phi(x)$ and with $\overline{\phi(x)} = -\phi(x)$. The symmetric part gives $\partial_\mu\big[\phi_\nu(x) - \overline{\phi_\nu(-x)}\big] + \partial_\nu\big[\phi_\mu(x) - \overline{\phi_\mu(-x)}\big] = 0$ and therefore $\phi_\mu(x) - \overline{\phi_\mu(-x)} = 2ia_\mu + 2b_{\mu\rho}x^\rho$, with a_μ, $b_{\mu\rho}$ real and $b_{\mu\rho} = -b_{\rho\mu}$. Together this makes $\phi_\mu(x) = ia_\mu + b_{\mu\rho}x^\rho + \partial_\mu\phi(x)$, which proves the statement in the theorem.

Note that a term $C_{\mu\nu\rho}(x_1^\rho - x_2^\rho)$ with antisymmetric $C_{\mu\nu\rho}$ in the two-point function can still be written in the general form $\partial^1_\mu \phi^{(1)}_\nu + \partial^2_\nu \phi^{(2)}_\mu$ [take, e.g., $\phi^{(1)}_\mu(x_1, x_2) = C_{\mu\nu\rho} x_1^\nu x_2^\rho$, $\phi^{(2)}_\mu(x_1, x_2) = C_{\mu\nu\rho} x_1^\rho x_2^\nu$], but according to this theorem not with translation-invariant $\phi^{(j)}_\mu(x_1 - x_2)$.

The collection of two-point functions given by Theorem 5.1 must be analyzed further and the associated Gaussian operator field theories investigated. For this it is useful to have general information on Guassian states and their representations. This will be provided in the next section.

VI. GAUSSIAN STATES AND THEIR REPRESENTATIONS

We consider again the general situation of Sec. II, a (complex) nuclear Fréchet space $\mathscr{V}$ with continuous conjugation $*$, $\mathscr{A} = \Sigma^{n=0}_{\infty} \oplus (\hat{\otimes}^n \mathscr{V})$ the Borchers algebra over $\mathscr{V}$.

For every continuous bilinear form $b(\cdot,\cdot)$ on $\mathscr{V}$ with $\overline{b(f,g)} = b(g^*, f^*)$, $\forall f,g \in \mathscr{V}$, there is a state ω on $\mathscr{A}$ defined by extension of $\omega(e) = 1$, $\omega(f_1 \otimes \cdots \otimes f_n) = 0$, for n odd, $f_j \in \mathscr{V}$, and $\omega(f_1 \otimes \cdots \otimes f_n) = \Sigma b(f_{j_1}, f_{j_2}) \cdots b(f_{j_{n-1}}, f_{j_n})$ for n even and with the same summation over permutations $j_1 \cdots j_n$ as in formula (5.3).

Such a state will be called *Gaussian*. In the special case where b has the additional properties $b(f,g) = b(g,f)$ and $b(f, f) \geqslant 0$, $\forall f, g \in \mathscr{V}$, ω is indeed the system of moments of a Gaussian (generalized) stochastic process (up to a trivial complexification). The definition contains the essential algebraic elements of what is called a generalized free boson field in standard field theory. It is of course also related to the concept of quasifree state on a CCR algebra in C^*-algebra theory, in fact, the term Gaussian state has recently been used in a C^*-algebra formalism of classical systems. See Ref. 21.

Strictly speaking, the states ω just defined should be called boson Gaussian states with zero mean. There is an obvious modification for the fermion case which we do not need here. In this paper we also shall not use Gaussian states with nonzero expectation for the fields. These can be obtained from the mean-zero states by means of the "shift" automorphisms α_λ defined in Sec. II. They will play a role in the further development of our formalism.

Particular properties of the form b are reflected in properties of the Gaussian state defined by b, for instance, if b is invariant under a linear (topological) isomorphism T of $\mathscr{V}$ that commutes with the conjugation, then the Gaussian state is invariant under the automorphism α_T of $\mathscr{A}$. A Gaussian state is positive if and only if the Hermitian form h associated with b by $h(f,g) = b(f^*, g)$ is positive definite. This will be obvious later (Corollary of Theorem 6.1).

A Gaussian state is characterized by the property that its GNS representation is a Fock space representation with creation and annihilation operators. To show this, we have to set up a Fock space formalism in which the role of Hilbert space is taken over by topological inner product spaces:

Let there be given, as a "one-particle space," a (complex) nuclear Fréchet space $\mathscr{H}^{(1)}$ with a (continuous) inner product $(\cdot,\cdot)^{(1)}$. The Fock space over $\mathscr{H}^{(1)}$ is then defined as the topological direct sum $\mathscr{H} = \Sigma^\infty_{n=0} \oplus \mathscr{H}^{(n)}$, in which $\mathscr{H}^{(0)} = \mathbb{C}$, and $\mathscr{H}^{(n)} = (\hat{\otimes}^n \mathscr{H}^{(1)})_s$ (for $n = 1, 2, 3, \cdots$), the symmetrization of the n-fold tensor product completed in the projective tensor product topology (which because of nuclearity coincides with the ϵ-topology). Every $\mathscr{H}^{(n)}$ is again a nuclear Fréchet space and $\mathscr{H}$ is therefore a nuclear LF space. The inner product $(\cdot,\cdot)^{(1)}$ in $\mathscr{H}^{(1)}$ defines an inner product in each "n-particle space" $\mathscr{H}^{(n)}$ by extension of $((u_1 \otimes \cdots \otimes u_n)_s, (v_1 \otimes \cdots \otimes v_n)_s)^{(n)} = (n!)^{-1} \Sigma (u_{\sigma(1)}, v_1)^{(1)} \cdots (u_{\sigma(n)}, v_n)^{(1)}$ (sum over all permutations σ of $1, \ldots, n$), and subsequently in $\mathscr{H}$ by $(\phi, \psi) = \Sigma^\infty_{n=0} (\phi^{(n)}, \psi^{(n)})^{(n)}$. It is positive definite if and only if the given inner product $(\cdot,\cdot)^{(1)}$ in $\mathscr{H}^{(1)}$ is positive definite. The unit vector $\Omega_0 = 1 \in \mathbb{C} = \mathscr{H}^{(0)}$ is called the vacuum vector. For each $u \in \mathscr{H}^{(1)}$ there is an operator $C(u)$ in $\mathscr{H}$, called creation operator, defined by linear extension of $C(u)\Omega_0 = u$, $C(u)(u_1 \otimes \cdots \otimes u_n)_s = \sqrt{(n+1)}\,(u \otimes u_1 \otimes \cdots \otimes u_n)_s$ $(n = 1, 2, \cdots)$, and an operator $A(u)$ in $\mathscr{H}$, an annihilation operator, defined by extension of $A(u)\Omega_0 = 0$, $A(u)(u_1 \otimes \cdots \otimes u_n)_s = n^{-1/2} \Sigma^n_{j=1} (u, u_j)^{(1)} (u_1 \otimes \cdots \otimes u_{j-1} \otimes u_{j+1} \otimes \cdots \otimes u_n)_s$. The $C(u)$ and $A(u)$ are continuous linear operators, and the dependence of $C(u)$, respectively $A(u)$, on $u \in \mathscr{H}^{(1)}$ is linear, respectively antilinear; this dependence is continuous in the sense that $u \to C(u)\psi$ and $u \to A(u)\psi$ define continuous maps from $\mathscr{H}^{(1)}$ into $\mathscr{H}$, for each fixed ψ in $\mathscr{H}$. Finally on has the relations $[C(u), C(v)] = [A(u), A(v)] = 0$, $[A(u), C(v)] = (u, v)^{(1)} 1_{\mathscr{H}}$, and $(C(u)\psi_1, \psi_2) = (\psi_1, A(u)\psi_2)$, $\forall\, u, v \in \mathscr{H}^{(1)}$, $\forall\, \psi_1, \psi_2 \in \mathscr{H}$.

We shall refrain from writing out detailed proofs of all these statements. The following may, however, be observed:

Part of this many-particle structure is quite general and can, for instance, be built on an arbitrary locally convex Hausdorff topological vector space $\mathscr{H}^{(1)}$ with (separately) continuous inner product $(\cdot,\cdot)^{(1)}$. In that case it is not hard to show, using standard properties of multilinear maps and tensor products, suitably modified for antilinearity at places, that the expression given above for $(\cdot,\cdot)$ defines a (separately continuous) inner product on $\Sigma^\infty_{n=0} \oplus (\otimes^n_\pi \mathscr{H}^{(1)})_s$, the locally convex direct sum of the spaces $(\otimes^n_\pi \mathscr{H}^{(1)})_s$, which carry the projective tensor product topologies and are not completed. One also verifies that the operators $C(u)$ and $A(u)$ are well defined in this space and have all the properties mentioned. If one adds as an extra assumption joint continuity of $(\cdot,\cdot)^{(1)}$, then it is not hard to prove by the same methods that $C(u)$ and $A(u)$ have unique extensions to the completed space $\mathscr{H} = \Sigma^\infty_{n=0} \oplus (\hat{\otimes}^n \mathscr{H}^{(1)})_s$, with the same properties, and, moreover, that the inner product extends to a separately continuous Hermitian form on $\mathscr{H}$. All this is left to the reader.

One cannot, however, in this general setting prove that this Hermitian form on $\mathscr{H}$ is indeed an inner product, i.e., remains nondegenerate after the completion. For this, additional assumptions on $\mathscr{H}^{(1)}$ are needed.

In our application of this generalized Fock space formalism $\mathscr{H}^{(1)}$ will be a quotient space of a nuclear Fréchet space, that is, of $\mathscr{S}^{(2)}$ or $\mathscr{S}^{(3)}$, and therefore itself a nuclear Fréchet space. This takes care of the joint continuity of $(\cdot,\cdot)^{(1)}$ and is, moreover, sufficient to prove nondegeneracy. This proof runs as follows:

Consider the antilinear map τ from $\mathscr{H}^{(1)}$ into its dual $\mathscr{H}^{(1)\prime}$, defined by $\langle \tau(u), \cdot \rangle = (u, \cdot)^{(1)}$. It is continuous from the weak topology $\sigma(\mathscr{H}^{(1)}, \mathscr{H}^{(1)\prime})$ to the weak topology $\sigma(\mathscr{H}^{(1)\prime}, \mathscr{H}^{(1)})$. Because of the Hermitian symmetry of $(\cdot,\cdot)^{(1)}$, τ can be

for a given Lorentz transformation Λ (not a pure rotation) a new dense domain $\mathscr{H}^{A,\Lambda}$ as the image of $\mathscr{H}^{A}$ under $U(\Lambda)^{-1}$ identified with its adjoint τ' (the adjoint of an antilinear map T is defined by $\langle T'F, u\rangle = \overline{\langle F, Tu\rangle}$. Therefore, τ is also continuous with respect to the strong topologies. (See Ref. 19, Chap. IV, 7.4.) The proof of the theorem there rests on the fact that a continuous linear map carries weakly bounded sets into weakly bounded sets. This remains true for continuous antilinear maps.) One has because the nondegeneracy of $(\cdot,\cdot)^{(1)}$, $\{0\} = \{v\in\mathscr{H}^{(1)} | (u, v)^{(1)} = 0, \forall u\in\mathscr{H}^{(1)}\} = (\mathrm{Im} T)^{\perp}$, so the image of τ is weakly dense in $\mathscr{H}^{(1)\prime}$. A complete nuclear space is semi-reflective (see Ref. 19), Chap. IV, 5.5 and III, 7.2, Corollary 2), a Fréchet space is barreled (Ref. 19), Chap. II, 7.1, Corollary), so the nuclear Fréchet space $\mathscr{H}^{(1)}$ is reflexive and its strong topology is therefore the given topology (Ref. 19), Chap. IV, 5.5 and 5.6). The strong topology on the dual is the Mackey topology; this is compatible with the duality $(\mathscr{H}^{(1)}, \mathscr{H}^{(1)\prime})$ and hence the image of τ is also strongly dense in $\mathscr{H}^{(1)\prime}$. We have the continuous antilinear map $\hat{\otimes}^n\tau$ from the nuclear Fréchet space $\hat{\otimes}^n\mathscr{H}^{(1)}$ to the nuclear space $\hat{\otimes}^n(\mathscr{H}^{(1)\prime})$; because of the canonical topological isomorphism between $\hat{\otimes}^n(\mathscr{H}^{(1)\prime})$ and $(\hat{\otimes}^n\mathscr{H}^{(1)})'$ with respect to strong dual topologies (see Lemma 3.2), $\hat{\otimes}^n\tau$ is, in fact, a map from $\hat{\otimes}^n\mathscr{H}^{(1)}$ to $(\hat{\otimes}^n\mathscr{H}^{(1)})'$ and its image is strongly dense and fortiori weakly dense in $(\hat{\otimes}^n\mathscr{H}^{(1)})'$. the map $\hat{\otimes}^n\tau$ defines a continuous Hermitian form $(\cdot,\cdot)^{(n)}$ on $\hat{\otimes}^n\mathscr{H}^{(1)}$ by $(\psi_1, \psi_2)^{(n)} = \langle(\hat{\otimes}^n\tau)\psi_1, \psi_2\rangle$ and by restriction a Hermitian form on $\mathscr{H}^{(n)} = (\hat{\otimes}^n\mathscr{H}^{(1)})_s$, which, of course, corresponds to the extension of $(\cdot,\cdot)$. The symmetrization projection P_s: $u_1\otimes\cdots\otimes u_n \to (n!)^{-1}\Sigma_\sigma u_{\sigma(1)}\otimes\cdots\otimes u_{\sigma(n)}$ is a topological homomorphism with the property $(P_s\psi_1, \psi_2) = (\psi_1, P_s\psi_2)$, $\forall \psi_1, \psi_2\in\hat{\otimes}^n\mathscr{H}^{(1)}$. Therefore, $\{\psi_2\in\mathscr{H}^{(n)} | (\psi_1, \psi_2)^{(n)} = 0, \forall \psi_1\in\mathscr{H}^{(n)}\}$
$= \{\psi_2\in\mathscr{H}^{(n)} | (P_s\psi_1, \psi_2)^{(n)} = 0, \forall \psi_1\in\hat{\otimes}^n\mathscr{H}^{(1)}\}$
$= \{\psi_2\in\mathscr{H}^{(n)} | (\psi_1, P_s\psi_2)^{(n)} = 0, \forall \psi_1\in\hat{\otimes}^n\mathscr{H}^{(1)}\}$
$= \{\psi_2\in\mathscr{H}^{(n)} | (\psi_1, \psi_2)^{(n)} = 0, \forall\psi_1\in\hat{\otimes}^n\mathscr{H}^{-1)}\}$
$= \{\psi_2\in\mathscr{H}^{(n)} | \langle(\hat{\otimes}^n\tau)\psi_1, \psi_2\rangle = 0, \forall\psi_1\in\otimes^n\mathscr{H}^{(n)}\}$
$= (\mathrm{Im}(\hat{\otimes}^n\tau))^{\perp} = \{0\}$. This proves the nondegeneracy of $(\cdot,\cdot)^{(n)}$ on $\mathscr{H}^{(n)}$ and this gives immediately the nondegeneracy of $(\cdot,\cdot)$ on $\mathscr{H}$.

We are now in a position to state the basic theorem on Gaussian states:

Theorem 6.1: Let $\mathscr{V}$ be a (complex) nuclear Fréchet space with continuous conjugation $*$; $b(\cdot,\cdot)$ a continuous bilinear form on $\mathscr{V}$, with $\overline{b(f,g)} = b(g^*, f^*)$, $\forall f, g\in\mathscr{V}$; ω the Gaussian state on $\mathscr{A}$ determined by b. Let $\mathscr{H}^{(1)}$ be the quotient space $\mathscr{V}/\mathscr{V}_L$, $\mathscr{V}_L := \{f\in\mathscr{V} | b(g,f) = 0, \forall g\in\mathscr{V}\}$, and $(\chi f, \chi g)^{(1)} := b(f^*, g)$ the inner product on $\mathscr{H}^{(1)}$, where χ is the canonical surjection of $\mathscr{V}$ onto $\mathscr{V}/\mathscr{V}_L$.

Then the GNS representation of $\mathscr{A}$, associated with ω is algebraically and topologically equivalent to the representation in the Fock space $\mathscr{H}$ over $\mathscr{H}^{(1)}$ generated by the "field operators" $\pi(f) := C(\chi f) + A(\chi f^*)$, $\forall f\in\mathscr{V}$.

Proof: $\mathscr{V}_L$ is a closed subspace of the nuclear Fréchet space $\mathscr{V}$; therefore, $\mathscr{H}^{(1)} = \mathscr{V}/\mathscr{V}_L$ is also a nuclear Fréchet space and with the inner product as defined by $(\chi f, \chi g)^{(1)} = b(f^*, g)$ is a suitable one-particle space over which a Fock space structure can be constructed. The "field operators" $\pi(f)$ defined in the Fock space $\mathscr{H}$ by $C(\chi f) + A(\chi f^*)$ depend linearly on f in $\mathscr{V}$, the function $f \to \pi(f)\psi$ is continuous for every fixed $\psi\in\mathscr{H}$ and one has $(\pi(f)\psi_1, \psi_2) = (\psi_1, \pi(f^*)\psi_2)$, $\forall f\in\mathscr{V}$, $\forall \psi_1, \psi_2\in\mathscr{H}$. A continuous representation π of $\mathscr{A}$ is obtained by extension of $\pi(f_1\otimes\cdots\otimes f_n)$: $= \pi(f_1)\cdots\pi(f_n)$. To show this, one notes that, $\forall n, k$ the $n+1$-linear map from $(X^n\mathscr{V})\times\mathscr{H}^{(k)} \mapsto \mathscr{H}$ given by $(f_1,\ldots, f_n, \psi_k) \to \pi(f_1)\cdots\pi(f_n)\psi_k$ is not only separately continuous but also (jointly) continuous because $\mathscr{V}$ and $\mathscr{H}^{(k)}$ are Fréchet spaces, see Ref. 16, Corollary of Theorem 34.1. It therefore defines a continuous bilinear map from $(\hat{\otimes}^n\mathscr{V})\times\mathscr{H}^{(k)}$ into $\mathscr{H}$ and because of the direct sum properties of $\mathscr{H}$ and $\mathscr{A}$ a separately continuous bilinear map $\mathscr{A}\times\mathscr{H}\to\mathscr{H}$ that can be written as $(a, \psi) \mapsto \pi(a)\psi$.

We next show that the representation π is strictly cyclic, with respect to the vacuum vector Ω_0, i.e., that the continuous linear map ν: $\mathscr{A}\to\mathscr{H}$ defined by $a\to\pi(a)\Omega_0$ is surjective. For this we consider a second continuous map ρ: $\mathscr{A}\to\mathscr{H}$, defined by $f_1\otimes\cdots\otimes f_n \mapsto C(\chi f_1)\cdots C(\chi f_n)\Omega_0$ (together with $e\to\Omega_0$). This map is not only continuous but also a surjective (topological) homomorphism. For every n one has $C(\chi f_1)\cdots C(\chi f_n)\Omega_0 = \sqrt{n!}((\chi f_1)\otimes\cdots\otimes(\chi f_n))_s$, so $f_1\otimes\cdots\otimes f_n \mapsto C(\chi f_1)\cdots C(\chi f_n)\Omega_0$ defines a map $\rho^{(n)}$ from $\hat{\otimes}^n\mathscr{V}$ to $\mathscr{H}^{(n)}$, which is apart from the factor $\sqrt{n!}$ a composition of the tensor product map $(\hat{\otimes}^n\chi)$: $(\hat{\otimes}^n\mathscr{V}) \to (\hat{\otimes}^n\mathscr{H}^{(1)})$ a surjective homomorphism (Lemma 3.4), and the symmetrization projection P_s: $\hat{\otimes}^n\mathscr{H}^{(1)} \to \mathscr{H}^{(n)}$, also a surjective homomorphism. Again by the properties of direct sums, ρ is therefore a surjective homomorphism from $\mathscr{A}$ to $\mathscr{H}$. To connect ν and ρ we consider a linear map α: $\mathscr{A}\to\mathscr{A}$ defined by extension of

$$\alpha e = e,$$

$$\alpha(f_1\otimes\cdots\otimes f_n) = f_1\otimes\cdots\otimes f_n + \sum_{\substack{\text{perm. of } 1\cdots n:\\ j_1<j_2\\ j_3<j_4<\cdots<j_n}} b(f_{j_1}, f_{j_2}) f_{j_3}\otimes\cdots\otimes f_{j_n} + (-1)^2 \sum_{\substack{\text{perm. of } 1\cdots n:\\ j_1<j_2, j_3<j_4; j_1<j_3\\ j_5<j_6<\cdots<j_n}} b(f_{j_1}, f_{j_2}) b(f_{j_3}, f_{j_4}) f_{j_5}\otimes\cdots\otimes f_{j_n}$$

$$\begin{cases} +(-1)^{n/2} \displaystyle\sum_{\substack{j_1<j_2, \cdots j_{n-1}<j_n\\ j_1<j_3<\cdots<j_{n-1}}} b(f_{j_1}, f_{j_2})\cdots b(f_{j_{n-1}}, f_{j_n}) e & \text{for } n \text{ even},\\ +(-1)^{(n-1)/2} \displaystyle\sum_{\substack{j_1<j_2, \cdots j_{n-2}<j_{n-1}\\ j_1<j_3\cdots<j_{n-2}}} b(f_{j_1}, f_{j_2})\cdots b(f_{j_{n-2}}, f_{j_{n-1}}) f_{j_n} & \text{for } n \text{ odd}. \end{cases} \tag{6.1}$$

The map α is continuous by the same arguments as before. One can also define a map β by the same expressions but without the factors $(-1)^j$ in front of the summations. By straightforward and tedius manipulations one then verifies that $(\alpha\beta)(f_1 \otimes \cdots \otimes f_n) = (\beta\alpha)(f_1 \otimes \cdots \otimes f_n) = f_1 \otimes \cdots \otimes f_n$, i.e., α has a continuous inverse, it is a topological linear isomorphism. In the same spirit one obtains $\rho(f_1 \otimes \cdots \otimes f_n) = C(\chi f_1) \cdots C(\chi f_n)\Omega_0 = \pi(\alpha(f_1 \otimes \cdots \otimes f_n))$ $\Omega_0 = (\nu\alpha)(f_1 \otimes \cdots \otimes f_n)$. This means $\rho = \nu\alpha$ or $\nu = \rho\alpha^{-1}$, and consequently ν is a surjective topological homomorphism from $\mathscr{A}$ onto $\mathscr{H}$. This proves the strict cyclicity of the representation π of $\mathscr{A}$ in $\mathscr{H}$. One checks, by using $\pi(f) = C(\chi f) + A(\chi f^*)$, the commutation relations for the creation and annihilation operators and the property $A(u)\Omega_0 = 0$, $\forall\, u \in \mathscr{H}^{(1)}$, that the vacuum expectation values $(\Omega_0, \pi(f_1) \cdots \pi(f_n)\Omega_0)$ are indeed the n-point functions $\omega(f_1 \otimes \cdots \otimes f_n)$ of the given Gaussian state ω. Therefore, by Lemma 2.1 the representation π is algebraically equivalent, and because ν is a homomorphism also topologically equivalent. Q.E.D.

Corollary: A Gaussian state ω defined by the continuous linear form b is positive if and only if the Hermitian form $h(f,g) = b(f^*, g)$ is positive definite.

A Gaussian state is invariant under an automorphism α_T whenever $b(\cdot,\cdot)$ is invariant under T. Because $T* = *T$, also $h(\cdot,\cdot)$ is then invariant and T induces an isometric topological linear isomorphism $U^{(1)}$ in the quotient space $\mathscr{H}^{(1)} = \mathscr{V}/\mathscr{V}_L$. Extension of $(u_1 \otimes \cdots \otimes u_n)_s \to (U^{(1)}u_1 \otimes \cdots \otimes U^{(1)}u_n)_s$ leads to an operator U in the Fock space $\mathscr{H}$ with the same properties, and this U is, of course, the operator which according to the discussion of Sec. II implements the automorphism α_T.

For a positive state ω the representation space $\mathscr{H}$ may be completed with respect to the inner product norm. We are then back in the Hilbert space formulation of standard Wightman free field theory. Creation and annihilation operators, together with the field operators, become unbounded with $\mathscr{H}$ as a common invariant dense domain. Using Nelson's analytic vector theorem, it can be shown that in this case the field operators $\pi(f)$, for $f = f^*$, are essentially self-adjoint on $\mathscr{H}$ (see Ref. 22, Chap. X.7).

All the cases that will be considered in the next sections, the unique free $F_{\mu\nu}$ field and the various free A_μ theories, are GNS representations of Gaussian states on $\mathscr{A}^F$ or $\mathscr{A}^A$. They therefore all have the same simple mathematical structure: a "many-particle" space with field operators as sums of a creation and an annihilation operator. The basic element in this structure is the "one-particle" space $\mathscr{H}^{(1)} = \mathscr{V}/\mathscr{V}_L$, consisting of equivalence classes of space–time test functions from $\mathscr{S}^{(2)}$ or $\mathscr{S}^{(3)}$. In practice these equivalence classes are always represented by suitably chosen multicomponent momentum "wavefunctions." Such representations are in general not unique, as will become clear in the next sections; moreover, their use result, in complications that tend to obscure the underlying simple general structure.

VII. THE FREE FIELD: LORENTZ COVARIANT GAUGES

In the terminology developed in the preceding section the standard free Maxwell $F_{\mu\nu}(x)$ quantum field theory is the GNS representation of a Gaussian state ω^F on the Borchers algebra $\mathscr{A}^F$, determined by a bilinear form b^F on $\mathscr{V} = \mathscr{S}^{(2)}$, written in generalized function language as

$$b^F(\psi_1, \psi_2) = \int \omega^F_{\mu_1\nu_1\mu_2\nu_2}(x_1, x_2)\psi_1^{\mu_1\nu_1}(x_1)\psi_2^{\mu_2\nu_2}(x_2)\, dx_1 dx_2 \tag{7.1}$$

for $\psi_1, \psi_2 \in \mathscr{S}^{(2)}$ and $\omega^F_{\mu_1\nu_1\mu_2\nu_2}$ the two-point function of formula (5.1).

Using Fourier transformed test functions $\hat{\psi}^{\mu\nu}(k) = (2\pi)^{-2}\int \psi^{\mu\nu}(x)e^{ikx}dx$, one can write b^F in the more rigorous form

$$b^F(\psi_1, \psi_2)$$

$$= -8\pi \int_{k^0 = |\mathbf{k}|} g_{\mu_1\mu_2} k_{\nu_1} \hat{\psi}_1^{\mu_1\nu_1}(-k) k_{\nu_2} \hat{\psi}^{\mu_2\nu_2}(k)\, d\mathbf{k}/2k^0. \tag{7.2}$$

The associated Hermitian form $h^F(\psi_1, \psi_2) = b^F(\psi_1^*, \psi_2)$ is positive definite; therefore the representation space $\mathscr{H}^F$ is indeed a pre-Hilbert space. (The positivity of h^F can be seen from writing h^F as

$$h^F(\psi_1, \psi_2) = 8\pi \int_{k^0 = |\mathbf{k}|} \sum_{j,l=1}^{3} P_{jl}(k)\, \overline{\phi_1^j(k)}\phi_2^l(k)\, d\mathbf{k}/2k^0,$$

with $\phi^j(k) = k_\nu \hat{\psi}^{j\nu}(k)$, $P_{jl}(k) = \delta_{jl} - k^j k^l/k_0^2$. For $k^2 = 0$, $k^0 > 0$, the matrix $P_{jl}(k)$ is Hermitian and idempotent and therefore positive definite.)

Lorentz invariance of the two-point function and therefore of ω^F is obvious. According to Sec. II there is then a representation of the inhomogeneous Lorentz group that extends to a strongly continuous unitary representation in the Hilbert space completion $\overline{\mathscr{H}^F}$ of $\mathscr{H}^F$ and transforming the field operators in the proper way. $\overline{\mathscr{H}^F}$ is a many-particle space of usual Hilbert space type. It can be described explicitly in terms of momentum amplitudes, involving a choice of two polarization vectors for each momentum. This is well known and will not concern us further.

We now resume the discussion of the realizations or gauges of the free $A_\mu(x)$ field based on the free $F_{\mu\nu}(x)$ field given by (5.1) and (7.1), (7.2). In Sec. V we already imposed the conditions of Gaussian form and translational invariance. This led to a general form for the two-point function $\omega^A_{\mu\nu}(x_1 - x_2)$ given by Theorem 5.1, which can be written more conveniently in momentum variables as

$$\hat{\omega}^A_{\mu\nu}(k) = -2\pi\big[g_{\mu\nu}\delta_+(k^2) + k_\mu\phi_\nu(k) + k_\nu\overline{\phi_\mu(k)} + iC_{\mu\nu\rho}\partial^\rho\delta(k)\big], \tag{7.3}$$

using Fourier transforms $f^\mu(x) = (2\pi)^{-2}\int \hat{f}^\mu(k)e^{-ikx}d^4k$ and $\omega^A_{\mu\nu}(x_1 - x_2) = (2\pi)^{-4}\int \hat{\omega}^A_{\mu\nu}(k)e^{-ik(x_1 - x_2)}d^4k$ and with $\phi_\mu(k)$ an arbitrary tempered vectorial distribution, $C_{\mu\nu\rho}$ real constants, antisymmetric in μ, ν, ρ.

The most natural additional requirement for a gauge is, of course, Lorentz covariance. Imposing Lorentz invariance on (7.3), we obtain:

Theorem 7.1: The Fourier transform of the two-point function of a Lorentz covariant free Maxwell field $A_\mu(x)$ has the general form

$$\hat{\omega}^A_{\mu\nu}(k) = -2\pi[g_{\mu\nu}\delta_+(k^2) + k_\mu k_\nu \phi(k)] \tag{7.4}$$

with $\phi(k)$ an arbitrary real Lorentz invariant tempered distribution.

Formula (7.4) is, of course, not very surprising. It follows easily from the general relation between $\omega^A_{\mu\nu}$ and the two-point function $\omega^F_{\mu_1\nu_1\mu_2\nu_2}$ of the free $F_{\mu\nu}(x)$ field together with the assumption that the general form of a Lorentz invariant tensor $\phi_{\mu\nu}(k)$ is $g_{\mu\nu}\phi_1(k) + k_\mu k_\nu \phi_2(k)$. For this last assumption, however, although used almost universally, no straightforward and rigorous proof that takes into account the distribution aspects, especially those connected with the behavior in $k=0$, is known to us (see for a characteristic remark Ref. 23, Sec. 3). It seems probable that a proper proof can as a special case be extracted from the very general work on invariant two-point functions of Oksak and Todorov,[24] but we prefer to give a proof of (7.4) independent of all this, based on our formula (7.3).

We use the following well-known and easy to derive necessary and sufficient conditions for Lorentz invariance of distributions:

$$(k_\alpha\partial_\beta - k_\beta\partial_\alpha)\phi(k) = 0, \tag{7.5}$$

$$(k_\alpha\partial_\beta - k_\beta\partial_\alpha)\phi_\mu(k) = g_{\mu\beta}\phi_\alpha(k) - g_{\mu\alpha}\phi_\beta(k), \tag{7.6}$$

$$(k_\alpha\partial_\beta - k_\beta\partial_\alpha)\phi_{\mu\nu}(k) = g_{\mu\beta}\phi_{\alpha\nu}(k) - g_{\mu\alpha}\phi_{\beta\nu}(k) + g_{\nu\beta}\phi_{\mu\alpha}(k) - g_{\nu\alpha}\phi_{\mu\beta}(k) \tag{7.7}$$

respectively for scalar, vectorial, and tensorial distributions $\phi(k)$, $\phi_\mu(k)$, and $\phi_{\mu\nu}(k)$ and with $\partial_\alpha = \partial/\partial k^\alpha$, etc. We have the following lemmas:

Lemma 7.1: If $\phi_\mu(k)$ is a tempered Lorentz invariant vectorial distribution, then there exists a tempered Lorentz invariant scalar distribution $\phi(k)$ such that $\phi_\mu = k_\mu\phi$.

Proof: From (7.6) one obtains $k_\nu(k_\alpha\partial_\beta - k_\beta\partial_\alpha)\phi_\mu = g_{\mu\beta}k_\nu\phi_\alpha - g_{\mu\alpha}k_\nu\phi_\beta$. Adding to this the two expressions obtained by cylic permutation of ν, α, β and taking in the result α and β arbitrary, $\mu=\nu$, but $\mu\neq\alpha$, $\mu\neq\beta$ gives $k_\alpha\phi_\beta - k_\beta\phi_\alpha = 0$. Using an earlier result, Kerd'_{32} = Imd'_{43}, together with Fourier transformation and proper test function formulation, one sees that this implies the existence of a tempered scalar distribution $\phi(k)$, such that $\phi_\mu = k_\mu\phi$. Substituting this in (7.6) gives $(k_\alpha\partial_\beta - k_\beta\partial_\alpha)k_\mu\phi(k) = 0$, which implies $(k_\alpha\partial_\beta - k_\beta\partial_\alpha)\phi(k) = a_{\alpha\beta}\delta(k)$, with constants $a_{\alpha\beta}$, $a_{\alpha\beta} = -a_{\beta\alpha}$, and subsequently $\partial_\mu(k_\alpha\partial_\beta - k_\beta\partial_\alpha)\phi = a_{\alpha\beta}\partial_\mu\delta$. Adding to this the two expressions obtained by cyclic permutation of α, β, μ one gets the following for α, β, μ all different: $a_{\alpha\beta}\partial_\mu\delta + a_{\beta\mu}\partial_\alpha\delta + a_{\mu\alpha}\partial_\beta\delta = 0$. Because of the linear independence of the derivatives of the δ functions this implies $a_{\alpha\beta} = 0$, $\forall\, \alpha, \beta$. This proves that ϕ satisfies (7.5) and is therefore Lorentz invariant. Q.E.D.

Lemma 7.2: If a tempered tensorial distribution $\phi_{\mu\nu}(k)$ is Lorentz invariant and has the form $\phi_{\mu\nu} = k_\mu\phi_\nu - k_\nu\phi_\mu$, then it vanishes identically.

Proof: (a) $k^\mu\phi_{\mu\nu}$ is a Lorentz-invaraint vectorial distribution; therefore, there exists (Lemma 7.1) a Lorentz invariant ϕ such that $k^\mu\phi_{\mu\nu} = k_\nu\phi$, or $k^2\phi_\nu = k_\nu k^\mu\phi_\mu + k_\nu\phi$. This gives $k^2\phi_{\mu\nu} = k_\mu k^2\phi_\nu - k_\nu k^2\phi_\mu = 0$, so $\phi_{\mu\nu}$ vanishes on $\{k\in\mathbb{R}^4 | k^2\neq 0\}$.

(b) We show next that $\phi_{\mu\nu}$ vanishes on $\{k\in\mathbb{R}^4 | k\neq 0\}$. Consider an arbitrary point q in $\{k\in\mathbb{R}^4 | k^2 = 0, k\neq 0\}$, for example with $q^0 > 0$. Choose a sufficiently small open neighborhood U of q. Introduce new coordinates s, u^j ($j=1,2,3$) on U by $s = k^2$, $u^j = k^j$. This can be seen as a diffeomorphism of $U\subset\mathbb{R}^4$ onto $U'\subset\mathbb{R}^4$, an open neighborhood of the point $s=0$, $u^j = q^j$. There is a 1–1 correspondence between the distributions on U and those on U', given by the symbolic relation $\int F(k)f(k)\,d^4k = \int \tilde{F}(s,\mathbf{u})\tilde{f}(s,\mathbf{u})2^{-1}(s+\mathbf{u}^2)^{-1/2}\,ds\,d\mathbf{u}$, with $\tilde{f}(s,\mathbf{u}) = f(k)$. Using this with special test functions $f(k)$, such that $\tilde{f}(s,\mathbf{u}) = \chi(s)g(\mathbf{u})$, with $\chi\in\mathscr{D}(\mathbb{R}^1)$, supp $\chi\in[-\epsilon, +\epsilon]$, $\chi(0) = 1$, $g\in\mathscr{D}(U'')$, $U''\subset\mathbb{R}^3$, $[-\epsilon,+\epsilon]\times U''\subset U'$, one deduces from $k^2\phi_{\mu\nu}(k) = 0$ that $\tilde{\phi}_{\mu\nu}(x,\mathbf{u})$ has the form $f_{\mu\nu}(\mathbf{u})\delta(s)$, for a $f_{\mu\nu}\in\mathscr{D}'(U'')$. The invariance condition (7.7) for $\alpha = l$, $u = j$ ($l, j = 1,2,3$), $\beta = \nu = 0$, $(k_l\partial_0 - k_0\partial_l)\phi_{j0} = \phi_j$ gives, when written in the variables s, $\mathbf{u}$, the relation $f_{jl}(\mathbf{u}) = -|\mathbf{u}|(\partial/\partial u^l)f_{j0}(\mathbf{u})$. Using the antisymmetry of $f_{\mu\nu}$, one has $\partial/\partial u^l)f_{j0} + (\partial/\partial u^j)f_{l0} = 0$, so $f_{j0}(\mathbf{u}) = a_j + \Sigma^3_{r=1} b_{jr}u^2$, $f_{jl}(\mathbf{u}) = -|u|b_{jl}$, with a_j, b_{jr} constants, $b_{jr} = -b_{rj}$. Combining this result with the invariance condition (7.7) for $\alpha = l$, $\beta = \mu = j$, $\nu = 0$, $j\neq l$, $(k_l\partial_j - k_j\partial_l)\phi_{j0} = -\phi_{l0}$, results in $u^j b_{jl} = -a_l - \Sigma^3_{r=1} b_{lr}u^r$ ($j\neq l$, no summation over j), or $a_l = b_{lj}u^j$, $l\neq j$, so $a_j = 0$, $b_{jl} = 0$, $\forall\, j, l = 1, 2, 3$. This proves that $\phi_{\mu\nu} = 0$ on a neighborhood of q.

(c) We are left with a $\phi_{\mu\nu}$ having $\{0\}$ as possible support. Then $\phi_{\mu\nu}(k)$, or more conveniently $\phi^{\mu\nu}(k)$, has the form

$$\phi^{\mu\nu}(k) = \sum_{p=0}^{n}\sum_{\rho_1\cdots\rho_p} a_p^{\mu\nu\rho_1\cdots\rho_p}\partial_{\rho_1}\cdots\partial_{\rho_p}\delta(k)$$

with, $\forall\, p = 0, 1, \ldots, n$, $a_p^{\mu\nu\rho_1\cdots\rho_p}$ constants, antisymmetric in μ, ν and symmetric in $\rho_1\cdots\rho_p$. For each p these constitute a Lorentz-invariant tensor. The finite-dimensional irreducible representations of the Lorentz group are characterized by pairs (j_1, j_2), $j_s = 0, \frac{1}{2}, 1\frac{1}{2}, \ldots$, such that $j_1 + j_2$ are integer. A tensor product $(j_1', j_2')\otimes(j_1'', j_2'')$ can be reduced according to the formula $(j_1', j_2')\otimes(j_1'', j_2'') = \Sigma\oplus(j_1, j_2)$, with $j_1 = |j_1' - j_1''|, |j_1' - j_1''| + 1, \ldots, j_1' + j_1''$, $j_2 = |j_2' - j_2''|, |j_2' - j_2''| + 1, \ldots, j_2' + j_2''$. See, e.g., Ref. 25, Chap. 7. Using this, one checks that the space of tensors $a^{\mu\nu\rho_1\cdots\rho_p}$, antisymmetric in μ, ν and symmetric in $\rho_1\cdots\rho_p$, which forms the representation $((1,0)\oplus(0,1))\otimes((p/2, p/2)\oplus(p/2) - 1, p/2 - 1)\oplus\cdots)$, does not contain, after reduction, the representation (0,0). This completes the proof that $\phi_{\mu\nu}(k)$ vanishes on all of $\mathbb{R}^4$.

Lemma 7.3: If a tensorial tempered distribution $\phi_{\mu\nu}(k)$ is Lorentz-invariant and has the form $\phi_{\mu\nu} = k_\mu\phi_\nu + k_\nu\phi_\mu$, then there exists a tempered Lorentz-invariant scalar distribution $\phi(k)$ such that $\phi_{\mu\nu}(k) = k_\mu k_\nu\phi(k)$.

Proof: The condition (7.7) for Lorentz invariance gives

$$k_\mu\{(k_\alpha\partial_\beta - k_\beta\partial_\alpha)\phi_\nu - (g_{\nu\beta}\phi_\alpha - g_{\nu\alpha}\phi_\beta)\} + k_\nu\{(k_\alpha\partial_\beta - k_\beta\partial_\alpha)\phi_\mu - (g_{\mu\beta}\phi_\alpha - g_{\mu\alpha}\phi_\beta)\} = 0.$$

This implies the existence of constants $a_{\alpha\beta\mu}$ and $b_{\alpha\beta\mu\rho}$, with $a_{\alpha\beta\mu} = -a_{\beta\alpha\mu}$ and $b_{\alpha\beta\mu\rho} = -b_{\beta\alpha\mu\rho} = -b_{\alpha\beta\rho\mu}$, such that $(k_\alpha\partial_\beta - k_\beta\partial_\alpha)\phi_\mu - (g_{\mu\beta}\phi_\alpha - g_{\mu\alpha}\phi_\beta) = a_{\alpha\beta\mu}\delta(k)$

$+ b_{\alpha\beta\mu\rho}\partial^\rho\delta(k)$. Multiplying, as in the proof of Lemma 7.1, with k_ν and combining the result with the two expressions obtained by permuting ν, α, β cyclically, one gets $g_{\mu\beta}(k_\alpha\phi_\beta - k_\beta\phi_\alpha) + g_{\mu\alpha}(k_\beta\phi_\nu - k_\nu\phi_\beta) + g_{\mu\beta}(k_\nu\phi_\alpha - k_\alpha\phi_\nu) = -(b_{\alpha\beta\mu\nu} + b_{\beta\nu\mu\alpha} + b_{\nu\alpha\mu\beta})\delta(k)$. Taking, as in Lemma 7.1 α, β arbitrary and $\mu = \nu, \mu \neq \alpha, \mu \neq \beta$, one obtains $g_{\nu\nu}(k_\alpha\phi_\beta - k_\mu\phi_\alpha) = -(b_{\beta\nu\nu\alpha} + b_{\nu\alpha\nu\beta})\delta(k)$ or $(k_\alpha\phi_\beta - k_\beta\phi_\alpha) = C_{\alpha\beta}\delta(k)$ with constants $C_{\alpha\beta}, C_{\alpha\beta} = -C_{\beta\alpha}$. This can be written as $k_\alpha(\phi_\beta - \frac{1}{2}C_{\beta\rho}\partial^\rho\delta) - k_\beta(\phi_\alpha - \frac{1}{2}C_{\alpha\rho}\partial^\rho\delta) = 0$, which shows the existence of a $\phi(k)$ such that $\phi_\beta = \frac{1}{2}k_\rho\phi + \frac{1}{2}C_{\beta\rho}\partial^\rho\delta$. Then $k_\mu\phi_\nu + k_\nu\phi_\mu = k_\mu k_\nu\phi$. The invariance condition (7.7) gives $k_\mu k_\nu(k_\alpha\partial_\beta - k_\beta\partial_\alpha)\phi = 0$. This implies the existence of constants $D_{\alpha\beta}, E_{\alpha\beta}{}^\rho$ (antisymmetric in α, β) such that $(k_\alpha\partial_\beta - k_\beta\partial_\alpha)\phi = D_{\alpha\beta}\delta + E_{\alpha\beta}{}^\rho\partial_\rho\delta$. Differentiating this with respect to k_μ and adding the two expressions obtained by cyclic permutation of α, β, μ, one gets

$$D_{\alpha\beta}\partial_\mu\delta + D_{\beta\mu}\partial_\alpha\delta + D_{\mu\alpha}\partial_\beta\delta + E_{\alpha\beta}{}^\rho\partial_\rho\partial_\mu\delta + E_{\beta\mu}{}^\rho\partial_\rho\partial_\alpha\delta + E_{\mu\alpha}{}^\rho\partial_\rho\partial_\beta\delta = 0.$$

Using linear independence of the derivatives of δ, one obtains $D_{\alpha\beta} = 0, E_{\alpha\beta}{}^\alpha = E_{\mu\beta}{}^\mu, E_{\alpha\beta}{}^\mu = 0, \forall\,\alpha, \beta, \mu$ all different (and no summation). Define $a_\beta = E_{\alpha\beta}{}^\alpha$ (no summation over α); then $(k_\alpha\partial_\beta - k_\beta\partial_\alpha)\phi = (a_\beta\partial_\alpha - a_\alpha\partial_\beta)\delta$, and with $a_\beta\partial_\alpha\delta = -k_\beta\partial_\alpha(a_\rho\partial^\rho\delta) - g_{\alpha\beta}a_\rho\partial^\rho\delta$ one obtains $(k_\alpha\partial_\beta - k_\beta\partial_\alpha)\phi = (k_\alpha\partial_\beta - k_\beta\partial_\alpha)(a_\rho\partial^\rho\delta)$ or $(k_\alpha\partial_\beta - k_\beta\partial_\alpha)(\phi - a_\rho\partial^\rho\delta) = 0$, so $\phi' = \phi - a_\rho\partial^\rho\delta$ is a Lorentz invariant tempered distribution with $k_\mu k_\nu\phi' = k_\mu k_\nu\phi = k_\mu\phi_\nu + k_\nu\phi_\mu$. Q.E.D.

The proof of Theorem 7.1 can now easily be given. The general translation invariant two-point function is given by $\hat{\omega}^A_{\mu\nu}(k)$ in (7.3). It consists of three terms that transform separately under Lorentz transformations. The first term, $-2\pi g_{\mu\nu}\delta_+(k^2)$, is already Lorentz-invariant. The last part, $-2\pi i C_{\mu\nu\rho}\partial^\rho\delta(k)$, must be zero because the constants $C_{\mu\nu\rho}$ form a Lorentz invariant antisymmetric tensor of rank 3 which vanishes because of the same properties of representations as mentioned in the proof of Lemma 7.2. The remaining term, $-2\pi(k_\mu\phi_\nu + k_\nu\bar{\phi}_\mu)$, has independently transforming real and imaginary parts on which Lemmas 7.2 and 7.3 can be applied separately. This gives the result (7.4) and finishes the proof of Theorem 7.1.

No particular gauge has been generally adopted as the standard one for the description of the quantized photon field. The reason for this is the incompatibility between some of the properties that one usually considers to be natural and desirable for quantum field theory. The main incompatibility is that between manifest Lorentz covariance of the $A_\mu(x)$ field and positivity of the metric of the state space. This has been known since the early development of the Gupta–Bleuler formalism and has since then been discussed in various degrees of generality and mathematical rigor by many authors. We have the following general and precise result:

Theorem 7.2: A Lorentz covariant $A_\mu(x)$ theory for the free photon field has indefinite metric.

Proof: (a) Consider for a given Lorentz covariant gauge the Hermitian form $h^A(\cdot,\cdot)$ on $\mathscr{S}^{(3)}$, given as $h^A(f_1,f_2) = b^A(f_1^*,f_2) = (A(f_1)\Omega^A, A(f_2)\Omega^A)$. According to Theorem 7.1, this form can be written as

$$h^A(f_1,f_2) = -2\pi\int_{k^0=|\mathbf{k}|} g_{\mu\nu}\,\overline{\hat{f}_1^\mu(k)}\hat{f}_2^\nu(k)\,d\mathbf{k}/dk^0 - 2\pi\int\phi(k)k_\mu\,\overline{\hat{f}_1^\mu(k)}k_\nu\hat{f}_2^\nu(k)\,d^4k.$$

(The first term is a rigorous expression; the second term is symbolical but has an obvious rigorous meaning.) It is sufficient to prove that $h^A(\cdot,\cdot)$ is indefinite.

(b) For $f\in\mathscr{S}_0^{(3)}$, $\exists\psi\in\mathscr{S}^{(2)}$: $f = d_{32}\psi$, and then $(A(f)\Omega^A, A(f)\Omega^A) = (A(d_{32}\psi)\Omega^A, A(d_{32}\psi)\Omega^A) = (F(\psi)\Omega^F, F(\psi)\Omega^F)$. This is $\geqslant 0$ (and even >0 for suitably chosen $f\in\mathscr{S}_0^{(3)}$), because of the positivity property of the free $F_{\mu\nu}(x)$ field.

(c) The slightly more difficult part is to show that, for an arbitrary Lorentz invariant $\phi(k)$ in the second term in $h^A(\cdot,\cdot)$, there exists an $f\in\mathscr{S}^{(3)}$ such that $h^A(f,f) < 0$. For this we shall consider test functions f^μ of the special form $(f^0, 0, 0, 0)$. For these one has

$$h^A(f,f) = -2\pi\int_{k^0=|\mathbf{k}|}|f^0(k)|^2\,d\mathbf{k}/2k^0 - 2\pi\int\phi(k)(k^0)^2|f^0(k)|^2\,d^4k.$$

(d) If for given $\phi(k)$ there exists a test function $f^0(k)$ such that $\int\phi(k)(k^0)^2|f^0(k)|^2d^4k > 0$, then $h^A(f,f)$ is for this test function <0 and we have finished. Suppose therefore that $\int\phi(k)(k^0)^2|f^0(k)|^2d^4k \leqslant 0$, for all $f^0(k)$. By considering first functions f^0 with supports not containing $k^0 = 0$ and then using Lorentz invariance of $\phi(k)$ one shows that then $\int\phi(k)|g(k)|^2\,d^4k \leqslant 0$, $\forall\,g$ with $\operatorname{supp} g \subset \mathbb{R}^4 - \{0\}$. Take an arbitrary $h(k)$ with $\operatorname{supp} h \subset \mathbb{R}^4 - \{0\}$ and $h(k) \geqslant 0$. Given this, one can find an open set U with $\operatorname{supp} h \subset U \subset \mathbb{R}^4 - \{0\}$ and a test function $g(k) \geqslant 0$, with $U \subset \operatorname{supp} g \subset \mathbb{R}^4 - \{0\}$. Then the test function $h_\lambda(k) = h(k) + \lambda g(k)^2$ is the square of a test function, $\forall\lambda > 0$, and converges to $h(k)$ for $\lambda\downarrow 0$. Therefore, $\int\phi(k)h(k)\,d^4k = \lim_{\lambda\downarrow 0}\int\phi(k)h_\lambda(k)\,d^4k \leqslant 0$. From $\int\phi(k)h(k)\,d^4k \leqslant 0$ for all $h \geqslant 0$ one obtains the well-known fact that $\phi(k)$ is a Radon measure. This implies that the one-dimensional distribution $\tilde{\phi}(s)$ determined by the restriction of the Lorentz-invariant distribution $\phi(k)$ to $\mathbb{R}^4 - \{k\in\mathbb{R}^4 | k^2 \geqslant 0, k^0 \leqslant 0\}$ (see Ref. 26), and for $k^0 > 0$ given by the symbolic expression $\int\phi(k)f(k)\,d^4k = \int\tilde{\phi}(s)\tilde{f}(s,\mathbf{u})2^{-1}(s + \mathbf{u}^2)^{-1/2}$, with $\tilde{f}(s,\mathbf{u}) = f(k)$, $s = k^2$, $u^j = k^j$ $(j = 1, 2, 3)$ is also a Radon measure.

(e) To show that a Radon measure $\phi(k)$ leads to an indefinite $h^A(f,f)$, it is sufficient to shown that there exist a $f^0(k)$ such that

(1) $\displaystyle\int_{k^0=|\mathbf{k}|}|f^0(k)|^2\,d\mathbf{k}/2k^0 > 0$

and

(2) $\displaystyle\left|\int\phi(k)(k^0)^2|f^0(k)|^2\,d^4k\right| < \int_{k^0=|\mathbf{k}|}|f^0(k)|^2\,d\mathbf{k}/2k^0.$

Consider functions $f^0(k)$ having the special form $f^0(k) = \tilde{f}^0(s,\mathbf{u}) = \chi(s)g(\mathbf{u})$, with the supports of χ and g such that $\operatorname{supp} f^0 C\{k\in\mathbb{R}^4 | k^0 > 0\}$. Then $|\int\phi(k)(k^0)^2|f^0(k)|^2\,d^4k = |\int\tilde{\phi}(s)[|\chi(s)|^2\int\frac{1}{2}|g(\mathbf{u})|(s + \mathbf{u}^2)^{1/2}\,d\mathbf{u}]\,ds|$. Because $\tilde{\phi}(s)$ is a measure, there exists an $A > 0$ such that this expression $< A \operatorname{supp}|\chi(s)|^2 \int \frac{1}{2}|g(\mathbf{u})|^2 \times (s+\mathbf{u}^2)^{1/2} d\mathbf{u}$, for all χ with $\operatorname{supp}\chi \subset \{s\in\mathbb{R}^1 |\ |s|<1\}$. Choose a point $(s_0, \mathbf{u}_0)$: $s_0 = 0$, $u_0^1 = 2^{-1}A^{-1/2}$, $u_0^2 = u_0^3 = 0$. The value of the continuous function $(s+\mathbf{u}^2)\mathbf{u}^2$ in $(s_0, \mathbf{u}_0)$ is $(4A)^{-2}$. There exists then an open neighborhood around $(s_0, \mathbf{u}_0)$ such that $2(8A)^{-2} < (\mathbf{u}^2+s)\mathbf{u}^2 < 6(8A)^{-2}$, having the form $|s|<\delta_1<1$, $|\mathbf{u}-\mathbf{u}_0|<\delta_2<2(8A)^{-2}$. On this neighborhood U one has $A(\mathbf{u}^2+s)^{1/2} < 1/2|\mathbf{u}|$. Take $\chi(s)$ with $\operatorname{supp}\chi \subset \{s\in\mathbb{R}^1 |\ |s|<\delta_1\}$, and $\chi(0) = \sup|\chi(s)| = 1$ and take $g(\mathbf{u})$ with $\operatorname{supp} g \subset \{\mathbf{u}\in\mathbb{R}^3 |\ |\mathbf{u}-\mathbf{u}_0|<\delta_2\}$. One then obtains $|\int \phi(k)(k^0)^2 |f^0(k)|^2 d^4k|$

$$< A \sup_{s\in(-\delta_1,+\delta_1)} \tfrac{1}{2}\int |g(\mathbf{u})|^2 (s+\mathbf{u}^2)^{1/2} d\mathbf{u}$$

$$\leqslant \int \frac{|g(\mathbf{u})|^2}{2|\mathbf{u}|} d\mathbf{u} = \int_{k^0=|\mathbf{k}|} |f^0(k)|^2 \frac{d\mathbf{k}}{2k^0}.$$

This proves (2) and (1) holds, of course, whenever g is not identically 0. Q.E.D.

The collection of Lorentz covariant gauges as characterized by Theorem 7.1 can be restricted further by two different additional gauge conditions leading, as we shall see, to two disjoint classes of covariant gauges.

From the point of view of classical electromagnetism a natural gauge condition is the Lorentz gauge condition, as discussed in Sec. IV.

Theorem 7.3: The Lorentz covariant free field gauges that satisfy the Lorentz gauge condition have two-point functions, given by

$$\hat{\omega}^A_{\mu\nu}(k) = -2\pi\{g_{\mu\nu}\delta_+(k^2) + k_\mu k_\nu[\delta_+^{(1)}(k^2) + \lambda\delta_+(k^2) + \lambda_-\delta_-(k^2) + \alpha\partial^\rho\partial_\rho\delta(k)]\} \tag{7.8}$$

with λ, λ_-, α arbitrary real constants.

{The Lorentz invariant distribution $\delta_+^{(1)}(k^2)$ is most conveniently defined as $\lim_{m^2\downarrow 0}[\delta_+^{(1)}(k^2-m^2) + \frac{1}{2}\pi(\log m^2)\delta(k)]$, where $\delta_+^{(1)}(k^2-m^2)$, for $m^2>0$, the "derivative" of $\delta_+(k^2-m^2)$, is given by the convergent integral

$$\tfrac{1}{2}\int_{k^0=(\mathbf{k}^2+m^2)^{1/2}} \left(\frac{f(k)}{(k^0)^2} - \frac{1}{k^0}\frac{\partial f(k)}{\partial k^0}\right)\frac{d\mathbf{k}}{2k^0},$$

suggested by the symbolic expression

$$\int \delta_+^{(1)}(k^2-m^2) f(k)\, d^4k$$

$$= \int \delta^{(1)}(s-m^2)\hat{f}(s,\mathbf{u}) \frac{ds\,d\mathbf{u}}{2(s+\mathbf{u}^2)^{1/2}}$$

$$= -\int \delta(s-m^2)\frac{\partial}{\partial s}\left(\frac{\hat{f}(s,\mathbf{u})}{2(s+\mathbf{u}^2)^{1/2}}\right) d\mathbf{u}\,ds.$$

It is a solution of the equation $(k^2-m^2)\delta_+^{(1)}(k^2-m^2) = -\delta_+(k^2-m^2)$. In the limit the integral diverges but can be regularized by a term $\frac{1}{2}\pi(\log m^2)\delta(k)$. The regularized limit satisfies $k^2\delta_+^{(1)}(k^2) = -\delta_+(k^2)$. See Ref. 26, where $\delta_+(k^2-m^2)$ and $\delta_+^{(1)}(k^2-m^2)$ are called H_{m^2} and $H^1_{m^2}$.}

Proof: The Lorentz gauge condition on $\hat{\omega}^A_{\mu\nu}(k)$ is $k^\mu\hat{\omega}^A_{\mu\nu}(k) = 0$, so for a Lorentz covariant gauge $k_\nu[\delta_+(k^2) + k^2\phi(k)] = 0$. This is equivalent to the existence of a constant a such that $k^2\phi(k) = -\delta_+(k^2) + a\delta(k)$. The distribution $\delta_+^{(1)}(k^2)$ is a particular solution of $k^2\phi_1(k) = -\delta_+(k^2)$; a particular solution of $k^2\phi_2(k) = a\delta(k)$ is $\phi_2(k) = \frac{1}{8}a\partial^\alpha\partial_\alpha\delta(k)$ and the general solution of the inhomogeneous equation $k^2\phi_3(k) = 0$ is $\phi_3(k) = \lambda\delta_+(k^2) + \lambda_-\delta_-(k^2) + \lambda_0\delta(k)$; $\lambda, \lambda_-, \lambda_0$ real constants. All this together gives as general solution of $k^2\phi = -\delta_+(k^2) + a\delta(k)$ the distribution $\phi(k) = \delta_+^{(1)}(k^2) + \lambda\delta_+(k^2) + \lambda_-\delta_-(k^2) + \lambda_0\delta(k) + \frac{1}{8}a\partial^\alpha\partial_\alpha\delta(k)$, which gives the required form (7.9).

In (7.8) the term with $\delta_-(k^2)$ represents an irrelevant negative part in the spectrum of time translations. The term with $\partial^\rho\partial_\rho\delta(k)$ gives a set of extra unphysical Lorentz invariant state vectors [note that $k_\mu k_\nu\partial^\rho\partial_\rho\delta(k) = 2g_{\mu\nu}\delta(k)$]. Both terms may be dropped. The Gaussian states, built on the remaining two-point function,

$$\hat{\omega}^A_{\mu\nu}(k) = -2\pi\{g_{\mu\nu}\delta_+(k^2) + k_\mu k_\nu[\delta_+^{(1)}(k^2) + \lambda\delta_+(k^2)]\} \tag{7.9}$$

for arbitrary real λ, can be called *generalized Landau gauges*. The case $\lambda = 0$ is the standard Landau gauge. Because $\delta_+^{(1)}(k^2)$ satisfies $k^2\delta_+^{(1)}(k^2) = -\delta_+(k^2)$ it is often given by the rather ambiguous expression

$$\hat{\omega}^A_{\mu\nu}(k) = -2\pi(g_{\mu\nu} - k_\mu k_\nu/k^2)\delta_+(k^2). \tag{7.10}$$

The term $k_\mu k_\nu\delta_+^{(1)}(k^2)$ can also be written in a more convenient and at the same time unambiguous way as $\frac{1}{2}k_\mu\partial_\nu\delta_+(k^2)$, which because of the Lorentz invariance of $\delta_+(k^2)$ is also equal to $\frac{1}{4}(k_\mu\partial_\nu + k_\nu\partial_\mu)\delta_+(k^2)$. {To see this, note that

$$\int k_\mu\delta_+^{(1)}(k^2) f(k)\, d^4k$$

$$= \lim_{m^2\downarrow 0}\int \delta_+^{(1)}(k^2-m^2) k_\mu f(k)\, d^4k$$

$$= -\tfrac{1}{2}\lim_{m^2\downarrow 0}\int \frac{\partial}{\partial k^0}\left(\frac{k_\mu f}{k^0}\right)\frac{d\mathbf{k}}{2k^0},$$

with $k^0 = (\mathbf{k}^2+m^2)^{1/2}$. Using, for $m^2>0$, the support properties of $\delta_+(k^2-m^2)$ and its Lorentz invariance, which implies $k_\mu\partial_0\delta_+(k^2-m^2) = k_0\partial_\mu\delta_+(k^2-m^2)$, one writes this integral as

$$-\tfrac{1}{2}\lim_{m^2\downarrow 0}\int \frac{\partial f}{\partial k^\mu}\frac{d\mathbf{k}}{2k^0}$$

$$= \tfrac{1}{2}\lim_{m^2\downarrow 0}\int \partial_\mu\delta_+(k^2-m^2) f(k)\, d^4k$$

$$= \tfrac{1}{2}\int \partial_\mu\delta_+(k^2) f(k)\, d^4k,$$

which proves $k_\mu\delta_+^{(1)}(k^2) = \frac{1}{2}\partial_\mu\delta_+(k^2)$.}

The representation of a Landau gauge as operator theory is in principle quite simple and straightforward. As a Gaussian state it has the Fock space sturcture generated in a unique way from a one-particle space $\mathscr{H}^{A(1)}$ that is the quotient of the basic test function space $\mathscr{S}^{(3)}$ over the degeneration subspace of the Hermitian form $h^A(\cdot,\cdot)$, symbolically written as

$$h^A(f,g) = \int \hat{\omega}^A_{\mu\nu}(k)\,\overline{\hat{f}^\mu(k)}\,\hat{g}^\mu(k)\, d^4k.$$

The realization of the spaces $\mathscr{H}^{A(1)}$, $\mathscr{H}^A$, with creation, annihilation, and field operators becomes, however, quite complicated in terms of the momentum functions on the forward light cone that can be chosen to represent the equivalence classes appearing as elements of $\mathscr{H}^{A(1)}$. Because derivatives are involved in $\hat{\omega}^A_{\mu\nu}(k)$, these momentum functions have at least eight components. See for a description of such realizations Refs. 2, 3, 27, and 28. One should in particular note in Ref. 3 the formidable complications due to the insistence on an additional, and from our point of view superfluous, Hilbert space structure in the locally convex state space $\mathscr{H}^A$.

The two properties that characterize the generalized Landau gauges are Lorentz covariance and the Lorentz condition $\partial_\mu A^\mu = 0$ for the field operator. In these gauges the free wave equation $\partial^\mu \partial_\mu A_\nu = 0$ does not hold. This would be equivalent to $k^2 \hat{\omega}^A_{\mu\nu}(k) = 0$. From (7.8) one obtains immediately

$$k^2 \hat{\omega}^A_{\mu\nu}(k) = k^2 k_\mu k_\nu \delta^{(1)}_+(k^2) = -k_\mu k_\nu \delta_+(k^2) \neq 0.$$

In fact, the operator equation $\partial^\mu \partial_\mu A_\nu = 0$ determines a second class of important Lorentz covariant free field gauges.

Theorem 7.4: The Lorentz-covariant free field gauges for which the field operator $A_\mu(x)$ satisfies the free wave equation, as an operator relation, have two-point functions given by

$$\hat{\omega}^A_{\mu\nu}(k) = -2\pi\{g_{\mu\nu}\delta_+(k^2) + k_\mu k_\nu[\lambda\delta_+(k^2) + \lambda_-\delta_-(k^2) + \alpha\partial^\rho\partial_\rho\delta(k)]\} \quad (7.11)$$

with λ, λ_-, α arbitrary real constants.

Proof: The operator relation $\partial^\mu \partial_\mu A_\nu = 0$ implies (and is, for a translational invariant Gausian state, equivalent to) $k^2 \hat{\omega}^A_{\mu\nu}(k) = 0$. Combined with (7.4) this gives $k_\mu k_\nu k^2 \phi(k) = 0$. This is equivalent to the existence of constants a, b_μ such that $k^2\phi(k) = a\delta(k) + b_\mu \partial^\mu \delta(k)$. Because of Lorentz invaraince of $k^2\phi(k)$ the b_μ must vanish. One has $k^2(\partial^\rho\partial_\rho\delta) = 8\delta(k)$; therefore, the general Lorentz invariant solution of $k^2\phi(k) = a\delta(k)$ is

$$\phi(k) = \lambda\delta_+(k^2) + \lambda_-\delta_-(k^2) + \lambda_0\delta(k) + \tfrac{1}{8}a\partial^\rho\partial_\rho\delta(k),$$

which proves (7.11).

The terms with $\delta_-(k^2)$ and $\partial^\rho\partial_\rho\delta(k)$ can be dropped for the same reasons as in the Landau gauges. The Gaussian states constructed from the remaining two-point function

$$\hat{\omega}^A_{\mu\nu}(k) = -2\pi(g_{\mu\nu} + \lambda k_\mu k_\nu)\delta_+(k^2) \quad (7.12)$$

for arbitrary $\lambda \in \mathbb{R}^1$, may be called *generalized Gupta–Bleuler* gauges. The case $\lambda = 0$ is known as *Feynman gauge*. The corresponding operator field theory is the rigorous form of the standard Gupta–Bleuler formalism, of course without an auxiliary noninvariant Hilbert space structure. In this case, contrary to that of the Landau gauges, it would be easy to provide such a structure; however, it would again serve no useful purpose.

It is not hard to verify that the different generalized Gupta–Bleuler gauges, to be denoted as $\omega^A(\lambda)$, are connected by gauge homomorphisms α_T, defined as in Sec. IV by a linear operator $T_\lambda: \mathscr{S}^{(3)} \to \mathscr{S}^{(3)}$, given by $(T_\lambda f)^\mu(k) = \hat{f}^\mu(k) + \tfrac{1}{2}\lambda k^\mu k_\nu \hat{f}^\nu(k)$. One has, in fact, $\alpha'_{T_\lambda}\omega^A(\lambda_0) = \omega^A(\lambda_0 + \lambda)$, $\forall\, \lambda, \lambda_0 \in \mathbb{R}^1$. The linear maps W_λ between the representation spaces of the $\omega^A(\lambda)$ are therefore 1–1 isometries, with $W_\lambda^{-1} = W_{-\lambda}$. One proves easily that the W_λ are continuous and therefore also topological isomorphisms. All the representation spaces can be identified with a single space, e.g., that of the Feynman gauge. In this space there is then a single vacuum vector Ω^A, and for each generalized Gupta–Bleuler gauge $\omega^A(\lambda)$ different field operators $A^{(\lambda)}_\mu(x)$, which can be obtained from the Feynman gauge field operator $A^{(0)}_\mu(x)$ by $A^{(\lambda)}_\mu(x) = A^{(0)}_\mu(x) - \tfrac{1}{2}\lambda\partial_\mu\partial^\nu A^{(0)}_\nu(x)$.

The Feynman gauge is the simplest gauge in terms of realization by momentum functions. Because $h^A(f,g) = -2\pi\int_{k^0=k} \overline{\hat{f}^\mu(k)}\hat{g}_\mu(k)\, d\mathbf{k}/2k^0$, vectors in the "one-particle" space $\mathscr{H}^{A(1)}$, which are equivalence classes of test functions f from $\mathscr{S}^{(3)}$, have natural representations as functions $\phi^\mu(k)$ on $\mathbb{R}^3$, obtained by restriction of the Fourier transforms $\hat{f}^\mu(k)$ to the forward light cone, according to $\phi^\mu(\mathbf{k}) = \sqrt{2\pi}\hat{f}^\mu(|\mathbf{k}|, \mathbf{k})$.

The Landau and Gupta–Bleuler gauges do not, of course, exhaust the possibilities for Lorentz-covariant free field gauges. Starting from formulas (7.8) and (7.10), for instance, one may obtain others {such as the Yennie–Fried gauge: $\hat{\omega}^A_{\mu\nu}(k) = -2\pi[g_{\mu\nu}\delta(k^2) - 2\delta^{(1)}_+(k^2)] =$ "$-2\pi(g_{\mu\nu} + 2k_\mu k_\nu/k^2)\delta_+(k^2)$".} Still other, more general, gauges can be chosen by specifying various Lorentz invariant distributions $\phi(k)$ in (7.4).

It should finally be observed that in a Lorentz-covariant gauges the free Maxwell equation $\partial^\mu\partial_\mu A_\nu - \partial_\nu\partial^\mu A_\mu = 0$ never holds as an operator relation. This remarkable fact was noticed by Strocchi at an early stage.[29] According to our analysis of the relation between the $A_\mu(x)$ and $F_{\mu\nu}(x)$ operator fields, this is not in contradiction with the operator equation $\partial^\mu F_{\mu\nu} = 0$. Moreover, it follows immediately from our general expression for the invariant two-point function, (7.4), because $(k^2\delta^\mu{}_\nu - k_\nu k^\mu)[g_{\mu\alpha}\delta_+(k^2) + k_\mu k_\alpha \phi(k)] = -k_\nu k_\alpha\delta_+(k^2) \neq 0$.

VIII. THE FREE FIELD: THE COULOMB GAUGE

According to Theorem 7.2 insistence on a Hilbert space theory means giving up Lorentz covariance of the field operator $A_\mu(x)$. Covariance under rotations (and, of course, space–time translations) can be retained. In fact the combination of this with the requirement that the component $A_0(x)$ of the field operator should vanish identically leads to the *Coulomb gauge*: The operator condition $A_0(x) = 0$ is for a Gaussian theory equivalent to $\hat{\omega}^A_{00}(k) = \hat{\omega}^A_{0j}(k) = \hat{\omega}^A_{0j}(k) = 0$ ($j = 1, 2, 3$). Applying this to the formula (7.3), the expressionf or the general translation invariant two-point function, and dropping the constants $C_{\mu\nu\rho}$, one obtains for $\phi_\mu(k)$ the equations $k_0[\phi_0(k) + \overline{\phi_0(k)}] = -\delta_+(k^2)$ and $k_0\phi_j(k) + k_j\overline{\phi_0(k)} = 0$. These have the obvious rotation invariant solutions

$$\phi_0(k) = -\delta_+(k^2)/2k^0,$$

$$\phi_j(k) = -(k_j/k^0)\phi_0(k) = [k_j/2(k^0)^2]\delta_+(k^2) \quad (8.1)$$
$$(j = 1, 2, 3).$$

Both these distributions are well defined as convergent integrals over $\mathbf{k}$. Substituting these $\phi_\mu(k)$ in the general formula (7.3), one obtains the well-known standard form for the Fourier transform of the Coulomb gauge two-point function:

$$\hat{\omega}^A_{00}(k) = \hat{\omega}^A_{0j}(k) = \hat{\omega}^A_{j0}(k) = 0,$$
$$\hat{\omega}_{jl}(k) = 2\pi[\delta_{jl} - k_j k_l/(k^0)^2]\delta_+(k^2). \tag{8.2}$$

The positivity of the Coulomb gauge follows immediately from the positivity of the Hermitian form $h^A(\cdot,\cdot)$ associated with this two-point function, and which is very similar to that of the free field, connected with formula (7.2).

From this two-point function it follows also quite easily that in this case both the free wave equation $\partial^\alpha\partial_\alpha A_\mu = 0$ and the Lorentz condition $\partial^\mu A = 0$ hold as operator relations.

The most important consequence of positivity for the Coulomb gauge is the fact that the relation between potential and tensor field is essentially the classical one, that is, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, with $F_{\mu\nu}$ and A_μ operators in the same state space, after identification of the representation spaces $\mathscr{H}^F$ and $\mathscr{H}^A$ or rather their Hilbert space completions.

To establish the precise properties of this identification, we need the following lemma:

Lemma 8.1: For the Coulomb gauge the physical subspace $\mathscr{H}^A_{\rm ph}$ is strictly smaller than $\mathscr{H}^A$; however, the Hilbert space completions of $\mathscr{H}^A_{\rm ph}$ and $\mathscr{H}^A$ are the same.

Proof: Let $\mathscr{I}_\omega$ be the left null-ideal of the Coulomb gauge as state ω on $\mathscr{A}^A$, $\mathscr{A}^A_{\rm ph}$ the closed subalgebra of $\mathscr{A}^A$ generated by $\mathscr{S}^{(3)}_0 \subset \mathscr{S}^{(3)}$. Let ν be the linear map $f_1 \otimes \cdots \otimes f_n \mapsto A(f_1)\cdots A(f_n)\Omega^A$, (see the proof of Theorem 6.1), the canonical surjection from $\mathscr{A}^A$ onto $\mathscr{H}^A = \mathscr{A}^A/\mathscr{I}_\omega$, with $\mathscr{H}^A = \nu(\mathscr{A}^A_{\rm ph})$. To show that $\mathscr{H}^A_{\rm ph} \neq \mathscr{H}^A$ one must prove that $\mathscr{I}_\omega + \mathscr{A}^A_{\rm ph} \neq \mathscr{A}^A$. Take instead of $\mathscr{S}^{(3)}_0$ the larger and more convenient closed subspace called (in this proof only) $\mathscr{S}^{(3)}_1$ and defined as $\mathscr{S}^{(3)}_1 = \{ f \in \mathscr{S}^{(3)} | \hat{f}^\mu(0) = 0\}$. Let $\mathscr{A}^A_1$ be the closed subalgebra of $\mathscr{A}^A$ generated by $\mathscr{S}^{(3)}_1$. It is then sufficient to prove that $\mathscr{I}_\omega + \mathscr{A}^A_1 = \mathscr{A}^A$. Use the toplogical linear isomorphism α from the proof of Theorem 6.1. It leaves $\mathscr{A}^A_1$ invariant and maps $\mathscr{I}_\omega$ onto the kernel of the map ρ (also introduced in the proof of Theorem 6.1). The map ρ is a direct sum $\Sigma^\infty_{n=0} \oplus \rho^{(n)}$, $\rho^{(n)}$: $\mathscr{A}^{A(n)} = (\hat{\otimes}^n \mathscr{S}^{(3)}) \to \mathscr{H}^{A(n)}$; so $\text{Ker}\rho = \Sigma^\infty_{n=0} \oplus \text{Ker}\rho^{(n)}$. The subalgebra $\mathscr{A}^A_1$ can also be written as a direct sum $\Sigma^\infty_{n=0} \oplus \mathscr{A}^{A(n)}_1$, $\mathscr{A}^{A(n)}_1 = \mathscr{A}^A_1 \cap \mathscr{A}^{A(n)}$. Therefore, $\text{Ker}\rho + \mathscr{A}^A_1 = \Sigma^\infty_{n=0} \oplus (\text{Ker}\rho^{(n)} + \mathscr{A}^{A(n)}_1)$. It is then sufficient to show that $\text{Ker}\rho^{(1)} + \mathscr{S}^{(3)}_1 \neq \mathscr{S}^{(3)}$. Now

$$\text{Ker}\rho^{(1)} = \mathscr{V}_L = \{ f \in \mathscr{S}^{(3)} | h^A(g,f) = 0, \forall\, g \in \mathscr{S}^{(3)} \}$$
$$= \Big\{ f \in \mathscr{S}^{(3)} | (k^0)^2 \hat{f}^j(k) = k^j \sum_{l=1}^{3} k^l \hat{f}^l(k)$$
$$\text{for} \quad k^2 = 0,\, k^0 \geqslant 0,\, j = 1,2,3 \Big\},$$

so $f \in \text{Ker}\rho^{(1)}$ implies $\hat{f}^1(0) = 0$, this also holds for $f \in \mathscr{S}^{(3)}_1$ and therefore $\text{Ker}\rho^{(1)} + \mathscr{S}^{(3)}_1 \neq \mathscr{S}^{(3)}$, which proves that $\mathscr{H}^A_{\rm ph}$ is strictly smaller than $\mathscr{H}^A$. [Using the property that $\mathscr{S}^{(3)}_1$ has finite codimension and that $\mathscr{S}^{(3)}$ therefore can be written as a topological direct sum $\mathscr{S}^{(3)} = \mathscr{S}^{(3)}_1 \oplus \mathscr{S}^{(3)}_2$ for some finite dimensional subspace $\mathscr{S}^{(3)}_2$ (see Ref. 16, Proposition 9.3), one can in fact show that $\mathscr{H}^A_{\rm ph}$ is not even dense in $\mathscr{H}^A$, in the natural locally convex topology of $\mathscr{H}^A$.]

We prove next that the Hilbert space completions of $\mathscr{H}^A$ and $\mathscr{H}^A_{\rm ph}$ coincide. Because of the Fock space structure of $\mathscr{H}^A$ it is enough to show this for the one particle space, i.e., one should prove that $\mathscr{H}^{A(1)}_{\rm ph} = \alpha(\mathscr{S}^{(3)}_0)$ and $\mathscr{H}^{A(1)} = \alpha(\mathscr{S}^{(3)})$ have the same Hilbert space completions. Let $\mathscr{K}$ be the Hilbert space of functions $\phi^j(k)$, $j = 1, 2, 3$, with components in $L_2(\mathbb{R}^3, d\,\mathbf{k}/2|\mathbf{k}|)$ and inner product $(\phi_1, \phi_2) = \int \Sigma^3_{j=1} \overline{\phi^j_1(\mathbf{k})}\phi^j_2(\mathbf{k})\, d\,\mathbf{k}/2|\mathbf{k}|$. Let P_1 be the orthogonal projection operator in $\mathscr{K}$ defined by $(P_1\phi)^j(\mathbf{k}) = \phi^j(\mathbf{k}) - (k^j/|\mathbf{k}|^2)\Sigma^3_{l=1} k^l\phi^l(\mathbf{k})$ (a.e.). The corresponding subspace $\mathscr{K}_1$ consists of $\phi \in \mathscr{K}$ such that $\Sigma^3_{j=1} k^j\phi^j(\mathbf{k}) = 0$ (a.e.). By associating with each $f \in \mathscr{S}^{(3)}$ the functions $\phi^j(\mathbf{k}) = (\sqrt{2\pi})(f^j(k) - [k_j/(k^0)^2]\Sigma^3_{l=1} k^l f^l(k))_{k^0 = |\mathbf{k}|}$ one defines a linear map from $\mathscr{S}^{(3)}$ into $\mathscr{K}_1$. Its kernel is just $\mathscr{V}_L$, so it induces an injective map from $\mathscr{H}^{A(1)} = \mathscr{S}^{(3)}/\mathscr{V}_L$ into $\mathscr{K}_1$, which is obviously isometric. This means that we have realized the one-particle space $\mathscr{H}^{A(1)}$, a space of equivalence classes, as a subspace of the function space $\mathscr{K}_1 \subset \mathscr{K}$. We have to prove that the (Hilbert space) closure of the image of $\mathscr{H}^{A(1)}$ and of its subspace $\mathscr{H}^{A(1)}_{\rm ph}$ are the same. For this it is enough to show that the image under the map $\mathscr{S}^{(3)} \to \mathscr{K}$ of a subspace smaller than $\mathscr{S}^{(3)}_0$ is dense in $\mathscr{K}_1$. $\mathscr{K}_1$ is then the common completion of $\mathscr{H}^{A(1)}$ and $\mathscr{H}^{A(1)}_{\rm ph}$, represented as functions $\phi^j(\mathbf{k})$. Define (again for this proof only) $\mathscr{D}^{(3)}_1 = \{ f \in \mathscr{S}^{(3)}_0 \text{supp} \hat{f}^\mu(k)$ compact and not containing $k = 0\}$. The image of $\mathscr{D}^3_1$ in $\mathscr{K}_1$ consist of all vector functions $\phi^j(\mathbf{k})$ with each component C^∞, with compact support not containing $\mathbf{k} = 0$, and satisfying $\Sigma^3_{j=1} k^j\phi^j(\mathbf{k}) = 0$, because for each such $\phi^j(\mathbf{k})$ one has an $f \in \mathscr{D}^{(3)}_1$, e.g., by defining $\hat{f}^0(k) = 0$, $\hat{f}^j(k) = (2\pi)^{-1/2}\chi(k^2)\phi^j(\mathbf{k})$, $\chi \in \mathscr{D}(\mathbb{R}')$ with $\chi(0) = 1$. The subspace of these ϕ^j is dense in $\mathscr{K}_1$, because without the condition $\Sigma^3_{j=1} k^j\phi^j(\mathbf{k})$ they would be dense in $\mathscr{K}$. (An arbitrary $\phi \in \mathscr{K}$ can be approximated by a sequence ϕ_n from the collection; if ϕ happens to be in $\mathscr{K}_1$, then it is approximated by the sequence $P_1\phi_n$, by continuity of P_1.) Q.E.D.

After identification of $\mathscr{H}^F$ with $\mathscr{H}^A_{\rm ph} \subset \mathscr{H}^A$, Ω^F with Ω^A, and Hilbert space completion, the Coulomb gauge leads to the following situation: There is a single state space, the Hilbert space $\overline{\mathscr{H}}$, in which there is a dense domain $\mathscr{H}^F$ on which field operators $F_{\mu\nu}(x)$ are defined [or $F(\psi)$, $\forall\, \psi \in \mathscr{L}^{(2)}$, in rigorous language]. A unitary representation $U(u, \Lambda)$, of the inhomogeneous Lorentz group acts in $\overline{\mathscr{H}}$, leaves $\mathscr{H}^F$ and the vacuum vector $\Omega \in \mathscr{H}^F$ invariant and transforms the field $F_{\mu\nu}(x)$ in tensorial way. There is a second dense domain $\mathscr{H}^A$, containing $\mathscr{H}^F$, on which the potential field operators $A_\mu(x)$ [or $A(f)$, $\forall f \in \mathscr{L}^{(3)}$] are defined. Between these and the $F_{\mu\nu}(x)$ one has the classical relation $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ [or $F(\psi) = A(d_{32}\psi)$, $\forall\, \psi \in \mathscr{L}^{(2)}$]. The noncovariance of the $A_\mu(x)$ field in the Coulomb gauge is now in the first place the fact that the domain $\mathscr{H}^A$ is not invariant under the unitary operators $U(\Lambda)$ (unless Λ is a pure rotation) and that consequently the expressions $U(\Lambda)A(f)U^{-1}(\Lambda)$ and $A(T_\Lambda f)$ are defined on different domains and are therefore not equal. We may use the Lorentz transformations to obtain new gauges. Define, for a given Lorentz transformation Λ (not a pure rotation) a new dense doman $\mathscr{H}^{A,\Lambda}$ as the image of $\mathscr{H}^{A}$ under $U(\Lambda)^{-1}$ and, on this, new potential field operators $A(f)$ by $A^{\Lambda}(f) = U(\Lambda)^{-1} A(T_\Lambda f) U(\Lambda)$. Antisymmetric differentiation of A_{μ}^{Λ} gives the same field $F_{\mu\nu}$, because $F^{\Lambda}(\psi) = A^{\Lambda}(d_{32}\psi) = U(\Lambda)^{-1}A(T_\Lambda d_{32}\psi)U(\Lambda) = U(\Lambda)^{-1}A \times (d_{32}T_\Lambda \psi)U(\Lambda) = U(\Lambda)^{-1}F(T_\Lambda \psi)U(\Lambda) = F(\psi), \forall \psi \in \mathscr{L}^{(2)}$. One can show by working through the formalism developed in Secs. II and IV and making the necessary identifications, that the field $A^{\Lambda}(f)$ is indeed the GNS representation associated with the gauge transformed state $\alpha'_\Lambda \omega^A$, obtained from the Coulomb gauge state ω^A by the transpose of the Lorentz automorphism α_Λ, which in this special case has the effect of something that may be called a gauge transformation, again one more example, besides the ones given in Sec. IV, of the use of this term. Note that we have here a situation where the field operators of two equivalent gauges act in the same state space and are nevertheless not connected by a gradient term. This is because there is only a single state space after Hilbert space completion, the domains on which the field operators are defined are different, and the expression $A_{\mu}^{\Lambda}(x) - A_\mu(x)$ is therefore not defined.

This ends our discussion of the Coulomb gauge, the most important and typical positive-metric, noncovariant gauge. Other such gauges are known (see Ref. 2), but this should be sufficient to demonstrate how the operator field properties of such gauges are determined by the general formalism.

IX. CONCLUDING REMARKS

Quantum electrodynamics, up till now the most successful theory in elementary particle physics, does not fit in the axiomatic scheme of standard Wightman theory as it was developed in the fifties and early sixties. In this respect it cannot be seen as an unfortunate but isolated exceptional case. On the contrary, its typical features appear in more complicated form in general nonabelian gauge theories, the new field theories that have become dominant in particle physics in recent years. It seems therefore that standard Wightman axiomatic field theory is, as a framework in which the basic concepts of quantum field theory can be discussed, in need of extension.

In this paper and in a preceding one we have given a rigorous axiomatic formalism for the photon quantum field in which we derived from a few basic principles, and in a systematic way, the often heuristic and unrelated results on the subject that can be found scattered in the literature.

The coherence and essential simplicity of the formalism was obtained by giving up the Hilbert space of standard Wightman theory as general background and relying instead on the under lying mathematical structures of locally convex spaces, associated with the distribution properties of the n-point functions.

Although a wide gap separates the Maxwell field from nonabelian gauge fields, or even from full quantum electrodynamics, this may suggest a possible direction in which a further development of Wightman theory may take place.

ACKNOWLEDGMENTS

The author is grateful to Professor D. van Dulst for a valuable suggestion used in the proof of Theorem 3.1. This investigation is part of the research program of the Stichting voor Fundamental Onderzoek der Materie (FOM), which is financially supported by the Organisatie voor Zuiver Wetenschappelijk Onderzoek (ZWO).

[1]P. J. M. Bongaarts, J. Math. Phys. **18**, 1510 (1977).
[2]F. Strocchi and A. S. Wightman, J. Math. Phys. **15**, 2198 (1974).
[3]F. Strocchi and A. S. Wightman, J. Math. Phys. **17**, 1930 (1976).
[4]J. Fröhlich, G. Morchio, and F. Strocchi, Ann. Phys. (N.Y.) **119**, 241 (1979).
[5]M. Mintchev, J. Phys. A **13** 1841 (1980).
[6]M. Mintchev and E. d'Emilio, J. Math. Phys. (to be published).
[7]H. J. Borchers, Nuovo Cimento **24**, 214 (1962).
[8]H. J. Borchers, lecture in *Europhysics Conference on Statistical Mechanics and Field Theory, Haifa, 1971* (Israel U.P., Jerusalem, 1972).
[9]H. J. Borchers, lecture in *International Symposium on Mathematical Problems in Theoretical Physics, Kyoto, 1975* (Springer-Verlag, Berlin, 1975).
[10]K. Maurin, Bull. Acad. Pol. Sci. **XI**, 115 (1963).
[11]K. Maurin, Bull. Acad. Pol. Sci. **XI**, 121 (1963).
[12]W. Wyss, in *Lectures in Theoretical Physics, Vol. XI-D (Boulder, 1968)* Gordon and Breach, New York, 1969).
[13]R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).
[14]N. N. Bogolubov, A. A. Logunov, and I. T. Todorov, *Introduction to Axiomatic Quantum Field Theory* (Benjamin, New York, 1975).
[15]R. F. Streater, Rep. Progr. Phys. **38**, 771 (1975).
[16]F. Treves, *Topological Vector Spaces, Distributions and Kernels* (Academic, New York, 1967).
[17]J. L. Kelley and I. Namioka, *Linear Toplogical Spaces* (Van Nostrand, Princeton, N.J., 1963).
[18]A. P. Robertson and W. J. Robertson, *Topological Vector Spaces* (Cambridge U.P., Cambridge, 1964).
[19]H. H. Schaefer, *Topological Vector Spaces* (MacMillan, New York, 1966).
[20]P. Basarab-Horwath, R. F. Streater, and J. Wright, Commun. Math. Phys. **68**, 195 (1979).
[21]J. Naudts, J. Math. Phys. **22**, 221 (1981).
[22]M. Reed and B. Simon, *Methods of Modern Mathematical Physics. II: Fourier Analysis, Self-adjointness* (Academic, New York, 1975).
[23]H. Reeh, Commun. Math. Phys. **14**, 315 (1969).
[24]A. I. Oksak and I. T. Todorov, Commun. Math. Phys. 14, 271 (1969).
[25]H. Bacry, *Lecons sur la théorie des groupes et les symétries des particules elémentaires* (Gordon and Breach, Paris, 1967).
[26]P. Methée, Comment. Math. Helv. **28**, 225 (1954).
[27]G. Rideau, Lett Math. Phys. **1**, 17 (1975).
[28]A. L. Carey and C. A. Hurst, Lett. Math. Phys. **2**, 227 (1978).
[29]F. Strocchi, Phys. Rev. **162**, 1429 (1967).

The second Legendre transform for the weakly coupled $P(\phi)_2$ model[a)]

Alan Cooper, Joel Feldman, and Lon Rosen
Mathematics Department, University of British Columbia, Vancouver, British Columbia, V6T 1Y4 Canada

(Received 15 February 1982; accepted for publication 12 March 1982)

We prove the existence and analyticity of the second Legendre transform of the generating functional for Euclidean Green's functions in the weakly coupled $P(\phi)_2$ model. The proof involves a bound on the partition function with a nonlocal quadratic source term. This bound also implies bounds on the Schwinger functions $S_n(f_1,...,f_n)$ that are optimal with respect to both topology and n-dependence.

PACS numbers: 03.70. + k, 11.10.Mn, 02.30. + g

I. INTRODUCTION

This paper complements our series[1-4] on the higher Legendre transforms $\Gamma^{(r)}\{\mathbf{A}\}$ $(r\geqslant 1)$ of the generating functional $G\{\mathbf{J}\}$ of connected Green's functions in Euclidean quantum field theory. The main focus of the program in Refs. 1–4 is on the role of $\Gamma^{(r)}$ in generating a variety of field-theoretic objects with r-cluster-irreducibility properties, such as (generalized) vertex functions, Bethe–Salpeter kernels, and r-irreducible expectations, and on its role in unifying and simplifying the proofs of the irreducibility properties of these objects. Given these goals we found it convenient (and in general necessary!) to view $\Gamma^{(r)}\{\mathbf{A}\}$ as a formal power series in $\mathbf{A}$ with coefficients its generalized vertex functions. It was in this framework of formal power series that we provided[1] a rigorous justification of our analysis of $\Gamma^{(r)}$.

For the case $r=1$ and for the weakly coupled $P(\phi)_2$ model [hereafter denoted $\epsilon P(\phi)_2$], Glimm and Jaffe[5] have proved the existence of $\Gamma^{(1)}\{A\}$ as a genuine (analytic) functional of A, for small A in a suitable Banach space. Our contribution in this paper is to extend the Glimm–Jaffe results to the second Legendre transform $\Gamma^{(2)}$ and thus to provide a more complete justification for the formalism of Ref. 1 for $r=2$. Basically we do so by squeezing the cluster expansion[6] a little harder. However, we must confess that a law of diminishing returns seems to be operative here. Our (rather technical) proofs are unlikely to extend beyond $r=2$ or to models involving fermions; on the other hand, the (easier) method of formal power series certainly does[2,4] and, insofar as the generalized vertex functions are concerned, provides all the required information.

We now describe our results in more detail. All results stated in this paper apply to the (Euclidean) $\epsilon P(\phi)_2$ model whose expectation we denote by $\langle\cdot\rangle$. Let $Z\{J\}$ be the Schwinger generating functional

$$Z\{J\}=\langle e^{\hat{\mathscr{J}}}\rangle, \tag{1.1}$$

where

$$\hat{\mathscr{J}}=\int\hat{\phi}(x)J(x)\,dx\equiv\int(\phi(x)-\langle\phi(x)\rangle)J(x)\,dx. \tag{1.2}$$

For the case $r=2$, we let

$$Z\{J,L\}=\langle e^{\hat{\mathscr{J}}+\hat{\mathscr{L}}}\rangle, \tag{1.3}$$

where

$$\begin{aligned}\hat{\mathscr{L}}&=\iint{:}\phi(x)\phi(y){:}\,L(x,y)\,dx\,dy\\&\equiv\iint(\hat{\phi}(x)\hat{\phi}(y)-\langle\hat{\phi}(x)\hat{\phi}(y)\rangle)L(x,y)\,dx\,dy\end{aligned} \tag{1.4}$$

and $L(x,y)=L(y,x)$. It is important that we use physical Wick powers $:\phi^r:$ (see Ref. 2) in the source terms in order that J and L end up in the "right" Banach spaces. These spaces are

$$\mathscr{H}_{-1}(\mathbb{R}^2)=\{J\mid\|J\|_{-1}=\|(-\Delta+1)^{-1/2}J\|_{L^2}<\infty\}$$

and

$$\begin{aligned}\mathscr{H}^s_{-1}(\mathbb{R}^4)=\{L\mid&\|L\|_{-1}\\&=\|(-\Delta+1)^{-1/2}\otimes(-\Delta+1)^{-1/2}L\|_{L_2}<\infty;\\&L(x,y)=L(y,x)\}.\end{aligned} \tag{1.5}$$

We shall often omit the superscript s and shall also write $\mathscr{H}_{-1}=\mathscr{H}_{-1}(\mathbb{R}^2)\oplus\mathscr{H}_{-1}(\mathbb{R}^4)$. Now the key ingredient in the Glimm–Jaffe analysis[5] of $\Gamma^{(1)}$ is a bound on $Z\{J\}$ for $J\in\mathscr{H}_{-1,\delta}\equiv\{J\in\mathscr{H}_{-1}\mid\|J\|_{-1}<\delta\}$ for some $\delta>0$. In Sec. II we extend their bound by removing the restriction that J be small.

Theorem II.1: There exists a constant K such that

$$|Z\{J\}|\leqslant e^{K\|J\|^2_{-1}} \tag{1.6}$$

for all $J\in\mathscr{H}_{-1}$.

As a spinoff from this theorem, we obtain "optimal" bounds on the $\epsilon P(\phi)_2$ Schwinger functions of the form (see Corollary II.7)

$$\left|\left\langle\prod_{i=1}^n\hat{\phi}(J_i)\right\rangle\right|\leqslant c^n(n!)^{1/2}\prod_{i=1}^n\|J_i\|_{-1}. \tag{1.7}$$

We regard (1.7) as optimal since one can do no better in the free theory. Note that the bound (1.7) is global in the sense that we make no assumption about the support of $J_i\in\mathscr{H}_{-1}$. Previous bounds with the optimal $\mathscr{H}_{-1}$ norm involved an $n!$ dependence,[5] whereas bounds with the optimal $(n!)^{1/2}$ dependence involved norms on J_i which were globally L^1.[7-9]

Actually Theorem II.1 is a warmup for the corresponding result including quadratic source terms.

Theorem III.1: There are constants K, $\bar{K}$ and $\delta>0$ such that

$$|Z\{J,L\}|\leqslant\exp(K\|J\|^2_{-1}+\bar{K}\|L\|^2_{-1}), \tag{1.8}$$

[a)] Research partially supported by the Natural Sciences and Engineering Research Council of Canada.

for all $J \in \mathscr{H}_{-1}$ and $L \in \mathscr{H}^s_{-1,\delta}$.

The proof of Theorem III.1 differs from that of Theorem II.1 in one important respect: the *nonlocal* quadratic term $\hat{\mathscr{L}}$ in (1.3) prevents an immediate application of the cluster expansion. We are obliged to expand that part of $e^{\hat{\mathscr{L}}}$ which couples across the "decoupling" Dirichlet contour in the cluster expansion. The resulting terms can be controlled by an analyticity argument and can be summed up for small L since, by our optimal bounds, $(1/n!)\hat{L}^n$ contributes a factor $(1/n!)\, c^n (2n)!^{1/2} \|L\|^n_{-1}$ (roughly speaking). Such an argument clearly fails for source terms of degree $i>2$ since the number singularity $(1/n!)\, c^n (in)!^{1/2}$ is not summable.

Given the bound (1.8) it is then a simple matter to construct $\Gamma^{(r)}$ as a genuine functional. $\Gamma^{(2)}$ is defined as follows.

$$G\{J,L\} = \ln Z\{J,L\},$$

$$A\{J,L\}(x) = \frac{\delta}{\delta J(x)} G\{J,L\}, \tag{1.9a}$$

$$B\{J,L\}(x,y) = \frac{\delta^2}{\delta J(x)\,\delta J(y)} G\{J,L\} - \frac{\delta^2}{\delta J(x)\,\delta J(y)} G\{0,0\}, \tag{1.9b}$$

$$\Gamma^{(2)}\{A,B\} = G\{J,L\} - \int A(x)J(x)\,dx - \int\!\!\int (B(x,y) + A(x)A(y))\,L(x,y)\,dx\,dy, \tag{1.10}$$

where in (1.10) $J = J\{A,B\}$ and $L = L\{A,B\}$ are obtained by inverting (1.9). (A,B) lives in the space dual to $\mathscr{H}_{-1}$, namely $\mathscr{H}_1 = \mathscr{H}_1(\mathbb{R}^2) \oplus \mathscr{H}^s_1(\mathbb{R}^4)$. Our main result is

Theorem IV.3: There is a $\delta > 0$ such that the map $(J,L) \mapsto (A,B)$ of (1.9) has an inverse defined and analytic on $\mathscr{H}_{1,\delta} = \mathscr{H}_{1,\delta}(\mathbb{R}^2) \oplus \mathscr{H}^s_{1,\delta}(\mathbb{R}^4)$. $\Gamma^{(2)}\{A,B\}$ is defined and analytic on $\mathscr{H}_{1,\delta}$.

Remark: Burnap has informed us that he has obtained similar results by somewhat different techniques.[10]

The various properties of analytic functions on Banach spaces that we use are collected in Appendix A, and various estimates involving covariances with Dirichlet boundary conditions are collected in Appendix B.

II. GLOBAL BOUNDS ON SCHWINGER FUNCTIONS

The main result of this section is

Theorem II.1: In the $\epsilon P(\phi)_2$ model there exists a constant K such that

$$|\langle e^{\hat{\mathscr{J}}}\rangle| \leqslant e^{K\|J\|^2_{-1}} \tag{2.1}$$

for all $J \in \mathscr{H}_{-1}$.

Remark: This estimate may be re-expressed as global bounds on Schwinger functions. See Corollary II.7.

Proof: It suffices to prove (2.1) for the case $J \in C_0^\infty$, since then the analyticity of the finite volume approximations $\langle e^{\hat{\mathscr{J}}}\rangle_\Lambda$ and Vitali's theorem imply the analyticity and hence [by Theorem A.2(b)] continuity of $\langle e^{\hat{\mathscr{J}}}\rangle$ in the $\mathscr{H}_{-1}$ topology. It also suffices to consider real J since

$$|\langle e^{\hat{\mathscr{J}}}\rangle| \leqslant \langle e^{Re\,\hat{\mathscr{J}}}\rangle \leqslant e^{K\|\mathrm{Re}\,J\|^2_{-1}} \leqslant e^{K\|J\|^2_{-1}}.$$

Our strategy, following Glimm and Jaffe,[5] is to take an arbitrary but fixed $J \in C_0^\infty$, decompose it into localized pieces $J = \Sigma_{\alpha\in\mathbb{Z}^2} J_\alpha$ and incorporate one J_α at a time into $\langle e^{\hat{\mathscr{J}}}\rangle$. Here $J_\alpha = \zeta_\alpha J$ and $\{\zeta_\alpha \in C_0^\infty(\mathbb{R}^2) \mid \alpha \in \mathbb{Z}^2\}$ is a partition of unity invariant under lattice translations. Because we use the cluster expansion to estimate the effect of incorporating each J_α we are obliged to consider expectations with preassigned Dirichlet data on a general finite closed contour l of lattice bonds. We let $\langle\cdot\rangle_l$ be a finite volume approximation (in region Λ) to $\langle\cdot\rangle$, which has zero Dirichlet data on l, and we let

$$\mathscr{J} = \phi(J),$$
$$\mathscr{J}^l = \phi(J) - \langle\phi(J)\rangle_l, \tag{2.2}$$

and

$$\|J\|_{-1,l} = \|(-\Delta_l + 1)^{-1/2} J\|_{L^2},$$

where $-\Delta_l$ is the Laplacian with zero Dirichlet data on l. We also let $C_l = (-\Delta_l + m_0^2)^{-1}$.

Suppose we have already estimated $\langle e^{\mathscr{J}^l_{(B)}}\rangle_l$, where $\mathscr{J}^l_{(B)}$ is defined as in (2.2) with

$$J_{(B)} = \sum_{\beta\in B} J_\beta,$$

where B is a finite subset of $\mathbb{Z}^2$, and we wish to incorporate a new J_α (i.e., $\alpha \notin B$) to get $J_{(B^+)}$, where $B^+ = B \cup \{\alpha\}$. We simply define $J(t) = J_{(B)} + tJ_\alpha$ and use

$$\langle e^{\mathscr{J}^l_{(B^+)}}\rangle_l = \langle e^{\mathscr{J}^l_{(B)}}\rangle + \int_0^1 dt\, \langle \mathscr{J}^l_\alpha\, e^{\mathscr{J}(t)^l}\rangle_l$$
$$= \langle e^{\mathscr{J}^l_{(B)}}\rangle_l + \int_0^1 dt\, \langle \mathscr{J}^l_\alpha\, e^{\mathscr{J}(t)}\rangle_l\, e^{-\langle\mathscr{J}(t)\rangle_l}. \tag{2.3}$$

We will shortly estimate the second term by a cluster expansion. To exploit the subtraction in $\mathscr{J}^l_\alpha$ [see (2.2)] in that expansion we introduce two independent identical $\langle\cdot\rangle_l$ theories—one red and one white. From now on $\langle \mathscr{J}^l_\alpha\, e^{\mathscr{J}(t)}\rangle_l$ will stand for its representation

$$\langle\langle (\mathscr{J}_{\alpha,\mathrm{red}} - \mathscr{J}_{\alpha,\mathrm{white}})\, e^{\mathscr{J}(t)\mathrm{red}}\rangle_{l,\mathrm{red}}\rangle_{l,\mathrm{white}}$$

in terms of these duplicate theories. The cluster expansion (CE) of Ref. 6 says

$$\langle \mathscr{J}^l_\alpha\, e^{\mathscr{J}(t)^l}\rangle_l = \sum_{X,\Gamma\in I_X} \int_0^{s(\Gamma)} d\sigma(\Gamma)\, \partial^\Gamma \langle \mathscr{J}^l_\alpha\, e^{\mathscr{J}(t)X}\rangle^u_{X,\sigma(\Gamma)}\, \langle e^{\mathscr{J}^{\bar{l}}_{X^c}}\rangle_{\bar{l}} \times \frac{Z_{\Lambda\setminus X,\bar{l}}}{Z_{\Lambda,l}}\, e^{\langle \mathscr{J}_{X^c}\rangle_{\bar{l}} - \langle\mathscr{J}(t)\rangle_l}, \tag{2.4}$$

where X is any finite union of closed lattice squares that is connected and contains the support of J_α (of course X/l need not be connected), $\mathscr{B} = \{$lattice bonds in $(\mathbb{Z}^2)^*$ that do not intersect the support of $\zeta_\alpha\}$, $I_X = \{\Gamma \subset \mathscr{B} \mid \Gamma$ finite, $\Gamma \subset \mathrm{Int}\, X,\, X\setminus(\mathscr{B}\setminus\Gamma)$ is connected$\}$,

$$s(\Gamma) = \left(s(\Gamma)_b = \begin{cases} 0 & b\in\mathscr{B}\setminus\Gamma \quad\text{or}\quad b\in l \\ 1 & b\in\Gamma\setminus l \end{cases} \right)$$

so $l\cup B\setminus\Gamma$ is the set of Dirichlet bonds and $\Gamma\setminus l$ is the set of coupling bonds, $\sigma(\Gamma)$ is a vector having one component $\sigma(\Gamma)_b$ for each nonzero $s(\Gamma)_b$ [$\sigma(\Gamma)_b$ is a measure of the strength of coupling across the bond $b\in\Gamma$],

$$\partial^\Gamma = \prod_{b\in\Gamma} \partial/\partial\sigma(\Gamma)_b ,$$

$$\bar{l} = l\cup \partial X,$$

$$\langle \cdot \rangle^u_{X,\sigma(\Gamma)} = \int \cdot\, e^{-V_X}\, d\mu_{C_{\bar l}(\sigma(\Gamma))}$$

is the un-normalized expectation whose Gaussian measure $d\mu_{C_{\bar l}(\sigma(\Gamma))}$ has covariance $C_{\bar l}(\sigma(\Gamma))$ with boundary conditions given by $\sigma(\Gamma)$, including zero Dirichlet data on $\bar l$ and whose interaction V_X is in the volume X and has Wick ordering matched to $C_{\bar l}(\sigma(\Gamma))$,

$$J_X = J\chi_X \quad (\text{note that } \mathscr{J}_{X^c} \text{ is independent of } t),$$

$$Z_{\Lambda,l} = \int e^{-V_\Lambda}\, d\mu_{C_l} .$$

The bulk of the work in estimating each of the terms in (2.4) is placed in the following lemmas. We will use c_i and e_i to denote universal constants.

Lemma II.2:

$$\left|\partial^\Gamma \langle J^l_\alpha\, e^{\mathscr{J}(t)_X}\rangle^u_{X,\sigma(\Gamma)}\right|$$
$$\leqslant e^{c_1|X| - K(m_0)|\Gamma|}\left\{\sum_{\beta\in B^+} e(\alpha,\beta)\| J_\beta\|^2_{-1,l}\right\}$$
$$\times e^{c_2\| J(t)_X\|^2_{-1,\bar l}},$$

where $e(\alpha,\beta) = e_1\, e^{-|\alpha-\beta|}$ and $K(m_0)$ can be made arbitrarily large by choosing m_0 large.

Remarks: The significant features of this lemma are

(1) The bound involves $\| J_\beta\|^2_{-1,l}$ and not just $\| J_\beta\|_{-1,l}$. This is a consequence of $\langle \mathscr{J}^l_\alpha\rangle_{X,\sigma(\Gamma)} = 0$, which is true for all $\sigma(\Gamma)$ thanks to the implementation of the subtraction of (2.2) by means of duplicate theories.

(2) Note that there is no subtraction in $\mathscr{J}(t)_X$, i.e., no factor such as $e^{-\langle \mathscr{J}(t)_X\rangle_{\bar l}}$ in $\langle \mathscr{J}^l_\alpha\, e^{\mathscr{J}(t)_X}\rangle^u_{X,\sigma(\Gamma)}$. Such a factor is unnecessary because

$$\begin{aligned}|\langle J(t)_X\rangle_{\bar l}| &\leqslant \|\chi_X\langle\phi(x)\rangle_{\bar l}\|_{+1,\bar l}\| J(t)_X\|_{-1,\bar l}\\ &\leqslant c_2|X|^{1/2}\| J(t)_X\|_{-1,\bar l}\\ &\leqslant c_3|X| + c_3\| J(t)_X\|^2_{-1,\bar l} .\end{aligned} \tag{2.5}$$

(3) Suppose, following Ref. 5, we define the seminorm:

Definition II.3:

$$\| J\|^2_{1,B,l} = \sum_B e(\alpha,\beta)\| J_\beta\|^2_{-1,l},$$

where Σ_B means that α and β are summed over B. Then the sum in Lemma II.2 can be absorbed in this seminorm since

$$\sum_{\beta\in B^+} e(\alpha,\beta)\| J_\beta\|^2_{-1,l} \leqslant \| J\|^2_{1,B^+,l} - \| J\|^2_{1,B,l} . \tag{2.6}$$

Lemma II.4:

$$|\langle \mathscr{J}_{X^c}\rangle_{\bar l} - \langle \mathscr{J}(t)\rangle_l|$$
$$\leqslant c_4|X| + c_5\sum_{Y\in\mathscr{C}(l)}\sum_{\alpha\in Y\cap X}\sum_{\beta\in B^+} e(\alpha,\beta)\|\chi_Y J(t)_\beta\|^2_{-1,l},$$

where $\mathscr{C}(l)$ is the set of connected components of $\mathbb{R}^2\setminus l$.

Remark: This estimate is basically a consequence of

$$\begin{aligned}|\langle \mathscr{J}_\beta\rangle_{\bar l} - \langle \mathscr{J}_\beta\rangle_l| &\leqslant c_6\, e^{-d(\beta,\partial X)}\| J_\beta\|_{-1,l}\\ &\leqslant c_6\, e^{-d(\beta,X)}\| J_\beta\|_{-1,l} .\end{aligned}$$

The significant feature of this estimate is that if we define a second seminorm by

Definition II.5:

$$\| J\|^2_{2,B,l} = \sum_{Y\in\mathscr{C}(l)}\sum_{\alpha\in Y}\sum_{\beta\in B} e(\alpha,\beta)\|\chi_Y J_\beta\|^2_{-1,l},$$

then the sum in Lemma II.4 can be absorbed in this seminorm since

$$\begin{aligned}&\sum_Y\sum_{\alpha\in Y\cap X}\sum_{\beta\in B^+} e(\alpha,\beta)\|\chi_Y J(t)_\beta\|^2_{-1,l}\\ &= \| J(t)\|^2_{2,B^+,l} - \sum_Y\sum_{\alpha\in Y\cap X^c}\sum_{\beta\in B^+} e(\alpha,\beta)\|\chi_Y J(t)_\beta\|^2_{-1,l}\\ &\leqslant \| J\|^2_{2,B^+,l} - \| J\chi_{X^c}\|^2_{2,B,\bar l} .\end{aligned} \tag{2.7}$$

[By an abuse of notation $J(t)_\beta$ means not $J(t)\zeta_\beta$ but rather J_β if $\beta\in B$ and tJ_α if $\beta=\alpha$. Furthermore $\| J(t)\|_{2,B^+,l}$ is defined with the abused $J(t)_\beta$.] The last inequality is a consequence of

$$\begin{aligned}&\| J\chi_{X^c}\|^2_{2,B,\bar l}\\ &= \sum_{Y\in\mathscr{C}(l)}\sum_{\substack{Z\in\mathscr{C}(\bar l)\\ Z\subset Y}}\sum_{\alpha\in Z}\sum_{\beta\in B} e(\alpha,\beta)\|\chi_Z\chi_{X^c} J_\beta\|^2_{-1,\bar l}\\ &\leqslant \sum_{Y\in\mathscr{C}(l)}\sum_{\substack{Z\in\mathscr{C}(\bar l)\\ Z\subset Y\cap X^c}}\sum_{\alpha\in Y\cap X^c}\sum_{\beta\in B} e(\alpha,\beta)\|\chi_Z\chi_{X^c} J_\beta\|^2_{-1,\bar l}\\ &= \sum_{Y\in\mathscr{C}(l)}\sum_{\alpha\in Y\cap X^c}\sum_{\beta\in B} e(\alpha,\beta)\|\chi_Y\chi_{X^c} J_\beta\|^2_{-1,\bar l},\end{aligned}$$

where we have used the decoupling property

$$\sum_{Z\in\mathscr{C}(\bar l)}\|\chi_Z J\|^2_{-1,\bar l} = \| J\|^2_{-1,\bar l} \tag{2.8}$$

of $(-\Delta_{\bar l}+1)^{-1}$.

We now continue with the proof of Theorem II.1 using Lemmas II.2 and II.4, which we prove shortly. The proof will be by induction on the size of B with the *Induction hypothesis*: "There exists a constant K (independent of B) such that (for all finite closed contours l of lattice bonds)

$$\langle e^{\mathscr{J}^l(B)}\rangle_l \leqslant e^{K\| J\|^2_{B,l}}, \tag{2.9}$$

where

$$\| J\|^2_{B,l} \equiv \| J\|^2_{1,B,l} + \| J\|^2_{2,B,l} ." \tag{2.10}$$

Theorem II.1 will then follow thanks to the inequality

$$\| J\|^2_{B,l} \leqslant e_3\| J\|^2_{-1,l} . \tag{2.11}$$

This inequality is one of a number of useful properties of the seminorms $\|\cdot\|_{1,B,l}$, $\|\cdot\|_{2,B,l}$, and $\|\cdot\|_{B,l}$ that are proven in Lemma B.5.

The inductive hypothesis is trivially true when $B=\phi$. Assume it is true for some B. To prove that it is true for $B^+ = B\cup\{\alpha\}$ with $\alpha\notin B$ apply (2.3) followed by the cluster expansion (2.4). To bound the right-hand side of (2.4) we apply Lemma II.2 and (2.6) to $\left|\partial^{\Gamma}\langle \mathscr{J}_\alpha^l e^{\mathscr{J}(t)X}\rangle^u_{X,\sigma(\Gamma)}\right|$, the standard CE estimate (Ref. 6, Proposition 5.2) to $Z_{\Lambda\setminus X,\bar l}/Z_{\Lambda,l}$ and Lemma II.4 and (2.7) to $e^{\langle \mathscr{J}_{X^c}\rangle_{\bar l}-\langle \mathscr{J}(t)\rangle_l}$. Furthermore since $X^c\cap B^+\subset B$ (see Fig. 1) the factor $\langle e^{\mathscr{J}^{\bar l}_{X^c}}\rangle_{\bar l}$ can be bounded by the induction hypothesis. This gives

$$\left|\langle \mathscr{J}_\alpha^l e^{\mathscr{J}(t)^l}\rangle_l\right| \leqslant \sum_{X,\Gamma} e^{c_7|X|-K(m_0)|\Gamma|}\{\|J\|^2_{1,B^+,l}-\|J\|^2_{1,B,l}\}$$
$$\times\exp\{c_2\|J(t)_X\|^2_{1,\bar l}+K\|J_{X^c}\|^2_{B,\bar l}$$
$$+c_5\big[\|J\|^2_{2,B^+,l}-\|J_{X^c}\|^2_{2,B,\bar l}\big]\}.$$

Lemma II.6 below is now used to bound the exponent by

$$K\|J\|^2_{1,B,l}+tK\left[\|J\|^2_{1,B^+,l}-\|J\|^2_{1,B,l}\right]+K\|J\|^2_{2,B^+,l}$$

(assuming $K\geqslant c_5$) and standard CE estimates (Ref. 6, Proposition 5.1) are used to bound

$$\sum_{X,\Gamma} e^{c_7|X|-K(m_0)|\Gamma|}\leqslant c_8.$$

So far we have

$$\left|\langle \mathscr{J}_\alpha^l e^{\mathscr{J}(t)^l}\rangle_l\right| \leqslant \frac{d}{dt}\exp\{K\|J\|^2_{1,B,l}$$
$$+tK\left[\|J\|^2_{1,B^+,l}-\|J\|^2_{1,B,l}\right]$$
$$+K\|J\|^2_{2,B^+,l}\}. \tag{2.12}$$

It is of course crucial here that the universal constants e_i, c_i are independent of K so that we may choose a K for the inductive hypothesis that is larger than $\max\{c_5,c_8,2c_2e_2,2c_2e_1^{-1}\}$. Substituting (2.12) into (2.3) gives, by the inductive hypothesis (2.9),

$$\langle e^{\mathscr{J}^l(B^+)}\rangle_l\leqslant e^{K\|J\|^2_{B,l}}+\int_0^1 dt\,\frac{d}{dt}\exp\{K\|J\|^2_{1,B,l}$$
$$+tK\left[\|J\|^2_{1,B^+,l}-\|J\|^2_{1,B,l}\right]+K\|J\|^2_{2,B^+,l}\}$$
$$\leqslant e^{K\|J\|^2_{B,l}}+e^{K\|J\|^2_{B^+,l}}-e^{K\|J\|^2_{1,B,l}+K\|J\|^2_{2,B^+,l}}$$
$$\leqslant e^{K\|J\|^2_{B^+,l}}.$$

This concludes the proof of Theorem II.1. ■

Lemma II.6:

$$c_2\|J(t)_X\|^2_{-1,\bar l}+K\|J_{X^c}\|^2_{1,B,\bar l}$$
$$\leqslant K\|J\|^2_{1,B,l}+tK\left[\|J\|^2_{1,B^+,l}-\|J\|^2_{1,B,l}\right]$$

if $K\geqslant\max\{2c_2e_2,2c_2e_1^{-1}\}$ and $0\leqslant t\leqslant 1$.

Proof: Since $J(t)_X=tJ_\alpha+J\chi_X$,

$$c_2\|J(t)_X\|^2_{-1,\bar l}+K\|J_{X^c}\|^2_{1,B,\bar l}$$
$$\leqslant 2c_2t^2\|J_\alpha\|^2_{-1,\bar l}+2c_2\|J\chi_X\|^2_{-1,\bar l}$$
$$+K\|\chi_{X^c}J\|^2_{1,B,\bar l}$$
$$\leqslant 2c_2t^2\|J_\alpha\|^2_{-1,\bar l}+K\|J\chi_X\|^2_{1,B,\bar l}$$
$$+K\|\chi_{X^c}J\|^2_{1,B,\bar l}$$

(using $K\geqslant 2c_2e_2$ and Lemma B.5c)

$$\leqslant 2c_2t^2\|J_\alpha\|^2_{-1,\bar l}+K\|J\|^2_{1,B,\bar l}$$

[by (2.8)]

$$\leqslant tK\left[\|J\|^2_{1,B^+,l}-\|J\|^2_{1,B,l}\right]+K\|J\|^2_{1,B,\bar l},$$

since $t^2\leqslant t$ and $e_1\|J_\alpha\|^2_{-1,l}$ is one term in [ ]. ■

We now restate and prove Lemmas II.2 and II.4.

Lemma II.2: For m_0 sufficiently large

$$\left|\partial^{\Gamma}\langle \mathscr{J}_\alpha^l e^{\mathscr{J}(t)X}\rangle^u_{X,\sigma(\Gamma)}\right|$$
$$\leqslant e^{c_1|X|-K(m_0)|\Gamma|}\Big\{\sum_{\beta\in B^+}e(\alpha,\beta)\|J_\beta\|^2_{-1,l}\Big\}e^{c_2\|J(t)_X\|^2_{-1,\bar l}},$$

where $K(m_0)$ tends to infinity as m_0 does.

Proof: We expand the exponential $e^{\mathscr{J}(t)X}$,

$$\partial^{\Gamma}\langle \mathscr{J}_\alpha^l e^{\mathscr{J}(t)X}\rangle^u_{X,\sigma(\Gamma)}=\sum_{n=1}^{\infty}\frac{1}{n!}\partial^{\Gamma}\langle \mathscr{J}_\alpha^l \mathscr{J}(t)^n_X\rangle^u_{X,\sigma(\Gamma)},$$

observing that the $n=0$ term is absent since

$$\langle\langle \mathscr{J}_{\alpha,\mathrm{red}}-\mathscr{J}_{\alpha,\mathrm{white}}\rangle^u_{\mathrm{red}}\rangle^u_{\mathrm{white}}=0.$$

For notational convenience we shall only consider the $\mathscr{J}_{\alpha,\mathrm{red}}$ term. The $\mathscr{J}_{\alpha,\mathrm{white}}$ term is estimated similarly (but more easily) and at worst just doubles the number of terms. By the usual CE formula [see (8.3) of Ref. 6]

$$\partial^{\Gamma}\langle \mathscr{J}_\alpha \mathscr{J}^n_X\rangle_{X,\sigma(\Gamma)}$$
$$=\sum_{\pi\in\mathscr{P}(\Gamma)}\int\Big(\prod_{\gamma\in\pi}\tfrac12\partial^{\gamma}C\cdot\delta^2_\phi\Big)\mathscr{J}_\alpha\mathscr{J}^n_X e^{-V_X}\,d\mu_{C_{\bar l}(\sigma(\Gamma))},\tag{2.13}$$

where $\mathscr{P}(\Gamma)$ is the set of partitions of Γ.

Let the δ_ϕ's act on the $\mathscr{J}$'s and the V_X in (2.13). We classify the terms produced according to which δ_ϕ's attack an interaction V_X and according to which $\mathscr{J}$'s are attacked. Altogether there are

$$2^{2|\pi|}\cdot 2^{n+1}\leqslant e^{2(|\Gamma|+n)}\tag{2.14}$$

such classes of terms and, since the factor (2.14) may be harmlessly absorbed into the overall bound, we may restrict our attention to one such subclass from now on.

We next localize the δ_ϕ's and all the $n_1\leqslant n+1$ *attacked* $\mathscr{J}$'s as follows. If a δ_ϕ hits a V_X it is localized by

$$\frac{\delta}{\delta\phi(x)}=\sum_i\chi_{\Delta_i}(x)\frac{\delta}{\delta\phi(x)}\equiv\sum_i\chi^l_i(x)\frac{\delta}{\delta\phi(x)}\tag{2.15a}$$

and if it hits a $\mathscr{J}$ it is localized by

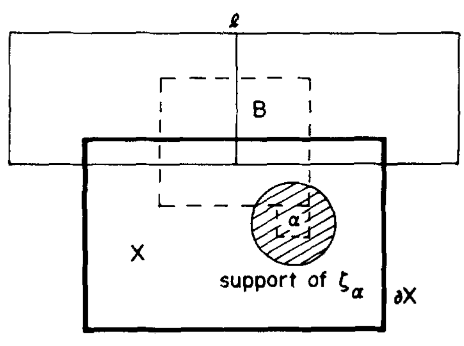


FIG. 1. The cluster expansion geometry.

$$\frac{\delta}{\delta\phi(x)} = \sum_j \int dy\,(C_{\bar{l}}^{-1/2}\chi_{\Delta_j}\,C_{\bar{l}}^{1/2})(x,y)\,\frac{\delta}{\delta\phi(y)}$$
$$\equiv \sum_j \int dy\,\chi_j^J(x,y)\,\frac{\delta}{\delta\phi(y)}. \qquad (2.15b)$$

The $\mathscr{J}$'s are localized by

$$J = \sum_j \int dy\,\chi_j^J(x,y)\,J(y) \equiv \sum_j J_j. \qquad (2.15c)$$

For each attacked $\mathscr{J}$ we insert the sum (2.15c) into (2.13) and for each $\partial^\gamma C$ we insert the sum

$$\partial^\gamma C(x,y) = \sum_{j_\gamma} \partial^\gamma C_j(x,y)$$
$$\equiv \sum_{j_\gamma} (\chi_{j_{\gamma,1}}^{Y_1}\,\partial^\gamma C\,\chi_{j_{\gamma,2}}^{Y_2})(x,y), \qquad (2.16)$$

where $j_\gamma = (j_{\gamma,1}, j_{\gamma,2})$ runs over $\mathbb{Z}^4$ and where $Y_i = I$ or J according to whether the corresponding δ_ϕ is localized by (2.15a) or (2.15b).

Now consider an arbitrary but fixed set of localizations $\{j_\gamma\}$, $\{j_i\}$. We have arranged in (2.15) that a localized δ_ϕ can hit only a V or $\mathscr{J}$ with the *same* localization. Hence if

$v =$ degree of interaction,

$M_I(\Delta_i) =$ number of δ_ϕ's localized by χ_i^I,

$M_J(\Delta_j) =$ number of δ_ϕ's localized by χ_j^J,

$=$ number of $\mathscr{J}$'s localized by χ_j^J,

and $M(\Delta) = M_I(\Delta) + M_J(\Delta)$, then the maximum number of terms (in our subclass) that have the given localizations is

$$\prod_\Delta (vM_I(\Delta))!M_J(\Delta)! \leqslant \prod_\Delta v^{vM_I(\Delta)}M(\Delta)!^v$$
$$\leqslant \prod_\Delta e^{c_9 M(\Delta)}M(\Delta)!^v \qquad (2.17)$$

where we have used the inequality $(ab)! \leqslant a^{ab}b!^a$. This bound is of standard CE form and will be controlled accordingly (see Ref. 6, Lemma 10.2).

Each of these terms is the integral with respect to $d\mu_{\bar{C}_{\bar{l}}(\sigma(\Gamma))}$ of a number of factors.

(i) First, there are the constant factors in which the same $\partial^\gamma C_j$ connects two J's. These factors are estimated by Lemma B.2 (recall that $\partial^\gamma C$ has Dirichlet data on $\bar{l}$):

$$|J_j\,\partial^\gamma C J_k| = |(\chi_{\Delta_j}C_{\bar{l}}^{1/2}J)(C_{\bar{l}}^{-1/2}\partial^\gamma C C_{\bar{l}}^{-1/2})(\chi_{\Delta_k}C_{\bar{l}}^{1/2}J)|$$
$$\leqslant \|\chi_{\Delta_j}C_{\bar{l}}^{1/2}J\|_{L^2}\,m_0^{c_{10}-|\gamma|/4}K_6(\gamma)\,e^{-m_0 d(j,k,\gamma)/2}$$
$$\times \|\chi_{\Delta_k}C_{\bar{l}}^{1/2}J\|_{L^2}.$$

Here

$$d(j,k,\gamma) = \sup_{b\in\gamma}\{d(\Delta_j,b) + d(\Delta_k,b)\}.$$

(ii) Second, there are the J's that have been connected by $\partial^\gamma C$'s to V's. We have the structure

$$\int dx\,F(x_1,x_2,\cdots)\prod_i \chi_\Delta(x_i)\,\partial^{\gamma_i}CJ_{j_i}(x_i)$$
$$= \sum_{k_i}\int dx\,F\prod_i \chi_{\Delta_i}(x_i)\,C_{\bar{l}}^{1/2}\chi_{\Delta_{k_i}}$$
$$\times (C_{\bar{l}}^{-1/2}\partial^{\gamma_i}CC_{\bar{l}}^{-1/2})(\chi_{\Delta_{J_i}}C_{\bar{l}}^{1/2}J), \qquad (2.18)$$

where F is a function of ϕ (which we have held fixed) as well as the arguments x_i of some $\partial^\gamma C$'s. We view $\Pi_i\,C_{\bar{l}}^{-1/2}\partial^{\gamma_i}CC_{\bar{l}}^{-1/2}$ as a tensor product of $C_{\bar{l}}^{-1/2}\partial^{\gamma_i}CC_{\bar{l}}^{-1/2}$'s and apply Lemma B.2 to bound the right-hand side of (2.18) by

$$\sum_{k_i} dx\,dx'\,F(x_1,x_2,\cdots)$$
$$\times\Big[\prod_i (\chi_{\Delta_i}C_{\bar{l}}^{1/2}\chi_{\Delta_{k_i}}C_{\bar{l}}^{1/2}\chi_{\Delta_i})(x_i,x_i')\Big]\,F(x_1',x_2',\cdots)$$
$$\times\prod_i m_0^{c_{10}-|\gamma_i|/4}K_6(\gamma_i)\,e^{-m_0 d(k_i,j_i,\gamma_i)}\|\chi_{\Delta_{j_i}}C_{\bar{l}}^{1/2}J\|_{L^2}.$$

(iii) Third, there is everything else—namely the uncontracted $\mathscr{J}$'s and the interaction. We use Holder's inequality to separate the interaction (and bound it by $e^{c_{11}|X|}$), the F factor (which is essentially of standard cluster expansion form and is estimated as such), and the $(n-n_1+1)$ uncontracted $\mathscr{J}$'s. These are estimated by

$$\Big\{\int \mathscr{J}_X^{4(n-n_1+1)}\,d\mu_{C_{\bar{l}}(\sigma(\Gamma))}\Big\}^{1/4}$$
$$\leqslant e^{c_{12}(n-n_1+1)}(n-n_1+1)!^{1/2}\,\|J_X\|_{-1,\bar{l}}^{n-n_1+1}$$

simply by evaluation of the Gaussian integral. (One of the J_X's could of course be J_α.)

We have now bounded each term in the series for $\partial^\Gamma\langle\mathscr{J}_\alpha\,\mathscr{J}_X^n\rangle_{X,\sigma(\Gamma)}$ that results from (2.13) when the δ_ϕ's are applied and $\{j_\gamma\}$ $\{j_i\}$ localizations are introduced. We have also shown how to control all the sums except $\Sigma_{\{j_i\}}\,\Sigma_{\{j_\gamma\}}$ and Σ_π. The sum $\Sigma_{\{j_i\}}$ over localizations of the n_1 attacked J's is controlled by

$$\sum_{j\in X}\|\chi_{\Delta_j}C_{\bar{l}}^{1/2}J\|_{L^2} \leqslant |X|^{1/2}\,\|J_X\|_{-1,\bar{l}}. \qquad (2.19)$$

The sums $\Sigma_\pi\,\Sigma_{\{j_\gamma\}}$ are controlled as in the standard cluster expansion (see Ref. 6, Proposition 8.1 and Lemma 10.2).

Gathering together all these estimates and recalling in particular that when a term has n_1 attacked $\mathscr{J}$'s each such $\mathscr{J}$ contributes a factor of $|X|^{1/2}\|J_X\|_{-1,\bar{l}}$ while the $(n-n_1+1)$ uncontracted $\mathscr{J}$'s contribute a factor of $e^{c_{12}(n-n_1+1)}(n-n_1+1)!^{1/2}\|J_X\|_{-1,\bar{l}}^{n-n_1+1}$ (or $e^{c_{12}(n-n_1+1)}(n-n_1+1)!^{1/2}\|J_X\|_{-1,\bar{l}}^{n-n_1}\,\|J_\alpha\|_{-1,\bar{l}}$), we have

$$|\partial^\Gamma\langle\mathscr{J}_\alpha^l\,e^{\mathscr{J}(t)X}\rangle^u_{X,\sigma(\Gamma)}|$$
$$\leqslant \sum_{n=1}^\infty \frac{1}{n!}e^{c_{13}|X|+c_{14}n-2K(m_0)|\Gamma|}\|J_\alpha\|_{-1,\bar{l}}\,\|J(t)_X\|_{-1,\bar{l}}^n$$
$$\times \max_{0<n_1<n+1}|X|^{n_1/2}(n-n_1+1)!^{1/2}$$
$$\leqslant \sum_{n=1}^\infty e^{c_{15}|X|+c_{14}n-2K(m_0)|\Gamma|}\|J_\alpha\|_{-1,\bar{l}}$$
$$\times\{\|J_\alpha\|_{-1,\bar{l}} + \|J_{(B),X}\|_{-1,\bar{l}}\}$$
$$\times\frac{\|J(t)_X\|_{-1,\bar{l}}^{n-1}}{(n-1)!^{1/2}}.$$

In the last inequality we used

$$|X|^{n_1/2}(n-n_1+1)!^{1/2}$$
$$\leqslant (e^{|X|}n_1!)^{1/2}(n-n_1+1)!^{1/2}$$
$$\leqslant e^{(1/2)|X|}(n+1)!^{1/2}.$$

Now using

$$\sum_{k=0}^{\infty}\frac{a^k}{(k!)^{1/2}}=\sum_{k=0}^{\infty}\frac{(\sqrt{2}a)^k}{(k!)^{1/2}}\frac{1}{(\sqrt{2})^k}$$

$$<\left(\sum_{k=0}^{\infty}\frac{(2a^2)^k}{k!}\right)^{1/2}\left(\sum_{k=0}^{\infty}\frac{1}{2^k}\right)^{1/2}<\sqrt{2}e^{a^2}, \qquad (2.20)$$

with $k=n-1$ and $a=e^{c_{14}}\|J(t)_X\|_{-1,\bar{l}}$, we do the sum over n. This yields

$$\partial^\Gamma\langle \mathscr{J}_\alpha^l e^{\mathscr{J}(t)X}\rangle^u_{X,\sigma(\Gamma)}|$$
$$\leqslant e^{c_{16}|X|-2K(m_0)|\Gamma|}\{\|J_\alpha\|^2_{-1,\bar{l}}+\|J_{(B)X}\|^2_{-1,\bar{l}}\}$$
$$\times e^{c_2\|J(t)_X\|^2_{-1,\bar{l}}}$$
$$\leqslant e^{c_{17}|X|-2K(m_0)|\Gamma|}\left\{\sum_{\beta\in B^+}\|J_\beta\chi_X\|^2_{-1,\bar{l}}\right\}e^{c_2\|J(t)_X\|^2_{-1,\bar{l}}}$$

by Lemma B.5c. (We have absorbed a number of constant factors into $e^{\text{const}|X|}$.) Finally observe that, since there are nine squares in the support of ζ_α,

$$|\Gamma|\geqslant\tfrac{1}{2}(|X|-9)\quad\text{by Ref. 6, Eq. (5.1)}$$
$$\geqslant c_{18}|\alpha-\beta|-c_{19}\ \forall\ \alpha,\beta\in|X|,$$

so that for m_0 sufficiently large

$$|\partial^\Gamma\langle\mathscr{J}_l^\alpha e^{\mathscr{J}(t)X}\rangle^u_{X,\sigma(\Gamma)}|$$
$$\leqslant e^{c_1|X|-K(m_0)|\Gamma|}$$
$$\times\left\{\sum_{\beta\in B^+}e(\alpha,\beta)\|J_\beta\|^2_{-1,l}\right\}e^{c_2\|J(t)_X\|^2_{-1,\bar{l}}}.$$

■

Lemma II.4: For m_0 sufficiently large

$$|\langle\mathscr{J}_{X^c}\rangle_{\bar{l}}-\langle\mathscr{J}(t)\rangle_l|$$
$$\leqslant c_4|X|+c_5\sum_{Y\in\mathscr{C}(l)}\sum_{\alpha\in Y\cap X}\sum_{\beta\in B^+}e(\alpha,\beta)\|\chi_Y J(t)_\beta\|^2_{-1,l}.$$

Proof:

$$|\langle\mathscr{J}_{X^c}\rangle_{\bar{l}}-\langle\mathscr{J}(t)\rangle_l|\leqslant|\langle\mathscr{J}(t)\rangle_l-\langle\mathscr{J}(t)\rangle_{\bar{l}}|+\langle\mathscr{J}(t)_X\rangle_{\bar{l}}.$$

The second term is bounded in (2.5) by

$$c_3|X|+c_3\|J(t)_X\|^2_{-1,\bar{l}}$$
$$\leqslant c_3|X|+\frac{c_5}{2}\sum_{Y\in\mathscr{C}(l)}\sum_{\beta\in B^+}\|\chi_Y J(t)_\beta\|^2_{-1,\bar{l}}\quad\text{(Lemma B.5c).}$$

To estimate the first term we use the interpolation

$$|\langle\mathscr{J}(t)\rangle_l-\langle\mathscr{J}(t)\rangle_{\bar{l}}|\leqslant\sum_Y\int_0^1 d\sigma\left|\frac{d}{d\sigma}\langle\mathscr{J}(t)\chi_Y\rangle_\sigma\right|,$$

where $\langle\cdot\rangle_\sigma$ denotes the expectation whose covariance is $\sigma C_{\bar{l}}+(1-\sigma)C_l$. Then the change of covariance formula [Ref. 6, Eq. (1.7)] gives

$$|\langle\mathscr{J}(t)\rangle_l-\langle\mathscr{J}(t)\rangle_{\bar{l}}|$$
$$\leqslant\sum_{\substack{Y,\beta,\Delta\\ \Delta\subset Y}}|\langle V'\chi_\Delta\rangle\,\partial C(\chi_Y J(t)_\beta)|$$
$$+\sum_{\substack{Y,\beta,\Delta,\Delta'\\ \Delta,\Delta'\subset Y}}|\langle\chi_Y\mathscr{J}(t)_\beta;\ (V'\chi_\Delta)\,\partial C(\chi_\Delta V')\rangle_\sigma|,$$

where $\partial C=C_{\bar{l}}-C_l=C_{l\cup\partial X}-C_l$ and $\langle\Phi;\Psi\rangle$ is the connected expectation $\langle\Phi\Psi\rangle-\langle\Phi\rangle\langle\Psi\rangle$. Now $\partial CX_Y=(C_{l\cup(\partial X\cap Y)}-C_l)\chi_Y$ can supply decay between its arguments and from each argument to $\partial X\cap Y$ and $\langle\Phi;\Psi\rangle$ can supply decay between the supports of Φ and Ψ so

$$|\langle\mathscr{J}(t)\rangle_l-\langle\mathscr{J}(t)\rangle_{\bar{l}}|$$
$$\leqslant c_{20}\sum_{Y,\beta,\Delta}e^{-d(\Delta,\beta)}e^{-d(\beta,\partial X\cap Y)}\|\chi_Y J(t)_\beta\|_{-1,l}$$
$$+c_{20}\sum_{Y,\beta,\Delta,\Delta'}e^{-\min\{d(\Delta,\beta),d(\Delta',\beta)\}}e^{-d(\Delta,\Delta')}$$
$$\times e^{-d(\Delta,\partial X\cap Y)}\|\chi_Y J(t)_\beta\|_{-1,l}.$$

[It is no problem to simultaneously get decay and a $\|\chi_Y J(t)_\beta\|_{-1,l}$ norm from $\partial C(\chi_Y J(t)_\beta)$—see Corollary B.3.] Finally, we perform the Σ_Δ and $\Sigma_{\Delta,\Delta'}$ sums and observe that

$$e^{-d(\beta,\partial X\cap Y)}\leqslant e^{-d(\beta,X\cap Y)}\leqslant\sum_{\alpha\in X\cap Y}e^{-|\alpha-\beta|+1}$$

to arrive at

$$|\langle\mathscr{J}(t)\rangle_l-\langle\mathscr{J}(t)\rangle_{\bar{l}}|$$
$$\leqslant c_5\sum_Y\sum_{\alpha\in Y\cap X}\sum_{\beta\in B^+}e(\alpha,\beta)\|\chi_Y J(t)_\beta\|_{-1,l})$$
$$\leqslant c_5\sum_Y\sum_{\alpha\in Y\cap X}\sum_{\beta\in B^+}e(\alpha,\beta)(\tfrac{1}{2}+\tfrac{1}{2}\|\chi_Y J(t)_\beta\|^2_{-1,l})$$
$$\leqslant c_5|X|+\frac{c_5}{2}\sum_Y\sum_{\alpha\in Y\cap X}\sum_{\beta\in B^+}e(\alpha,\beta)\|\chi_Y J(t)_\beta\|^2_{-1,1}.$$

■

Theorem II.1 amounts to global bounds on the Schwinger functions as we show in

Corollary II.7: In $\epsilon P(\phi)_2$

(a) $|\langle\hat{\phi}(f_1)\cdots\hat{\phi}(f_n)\rangle|\leqslant c_{21}^n(n!)^{1/2}\prod_{i=1}\|f_i\|_{-1},$

(b) $|\langle\hat{\phi}(f_1)\cdots\hat{\phi}(f_n)e^{\hat{\phi}(f)}\rangle|$

$$\leqslant c_{21}^n(n!)^{1/2}\left(1+\frac{K^{1/2}\|f\|_{-1}}{n^{1/2}}\right)^n$$
$$\times\prod_{i=1}\|f_i\|_{-1}e^{K\|f\|^2_{-1}},$$

where K is any constant for which (2.1) holds.

Remarks: (1) The numerical coefficient $c_{21}^n(n!)^{1/2}$ may be replaced by $(e^6Kn)^{n/2}$.

(2) The factor $(1+K^{1/2}\|f\|_{-1}/n^{1/2})^n$ in part (b) can be thought of as a contribution from the contractions of the $\hat{\phi}(f_i)$'s to the exponential $e^{\hat{\phi}(f)}$. For example, in a free theory of mass 1

$$\langle\phi(f_1)\cdots\phi(f_n)e^{\phi(f)}\rangle$$
$$=e^{(1/2)\|f\|^2_{-1}}\sum_{I\subset\{1,\dots,n\}}\left\langle\prod_{i\in I}\phi(f_i)\right\rangle\prod_{i\in I'}(f_i,Cf)$$
$$<e^{(1/2)\|f\|^2_{-1}}\sum_{\substack{j=0\\ j\text{ even}}}^{n}\binom{n}{j}(j-1)!!\,\|f\|^{n-j}_{-1}\prod_{i=1}^{n}\|f_i\|_{-1}$$
$$<e^{(1/2)\|f\|^2_{-1}}\prod_{i=1}^{n}\|f_i\|_{-1}\,n^{n/2}\sum_{j=0}^{n}\binom{n}{j}n^{(-n-j)/2}\|f\|^{n-j}_{-1}$$
$$=e^{(1/2)\|f\|^2_{-1}}\prod_{i=1}^{n}\|f_i\|_{-1}\,n^{n/2}\left(1+\frac{\|f\|_{-1}}{n^{1/2}}\right)^n.$$

Proof: Analyticity of its finite volume approximations imply the analyticity of $\langle e^{\hat\phi(f)}\rangle$ via Vitali's theorem. Hence the corollary is a direct consequence of Theorems A.1 and II.1. ∎

III. BOUNDS ON $Z\{J,L\}$

The principal result of this section is [see (1.3) for definitions]

Theorem III.1: There are constants K, $\bar K$ and $\delta > 0$ such that

$$|Z\{J,L\}| \leqslant e^{K\|J\|^2_{-1} + \bar K \|L\|^2_{-1}}$$

for all $J\in\mathscr{H}_{-1}$ and $L\in\mathscr{H}^s_{-1,\delta}$.

Proof: The proof is an extension of that of Theorem II.1. Again it suffices to consider J, L real and C_0^∞. In addition to the seminorms $\|J\|_{1,B,l}$, $\|J\|_{2,B,l}$, and $\|J\|_{B,l}$ of Theorem II.1 we use

$$\|L\|^2_{1,B,l} = \sum_B [e(\alpha,\gamma) + e(\beta,\gamma)]\, \|L_{\alpha,\beta}\|^2_{-1,l},$$

where $\Sigma_B \equiv \Sigma_{\alpha,\beta,\gamma\in B}$ and $L_{\alpha,\beta} = \zeta_\alpha L \zeta_\beta$,

$$\|L\|^2_{2,B,l} = \sum_{Y\in\mathscr{C}(l)} \sum_{\alpha\in Y} \sum_{\beta,\gamma\in B} e(\alpha,\beta) \times\{\|\chi_Y L_{\beta,\gamma}\|^2_{-1,l} + \|L_{\gamma,\beta}\chi_Y\|^2_{-1,l}\},$$

and

$$\|L\|^2_{B,l} = \|L\|^2_{1,B,l} + \|L\|^2_{2,B,l}.$$

(As in Theorem II.1 the $\|\cdot\|_{2,B,l}$ norm is used only to control boundary condition adjustments to the ˆ and : : subtractions.)

Pick any fixed C_0^∞ J and L with $\|L\|_{B,l} \leqslant \epsilon$ (for all B and l). Let $L_{(B)} = \Sigma_{\alpha,\beta\in B} L_{\alpha,\beta}$. Again we prove by induction on the size of B that

$$\langle e^{\hat{\mathscr{J}}(B) + \hat{\mathscr{L}}(B)}\rangle_l \leqslant e^{K\|J\|^2_{B,L} + \bar K(1+\|J\|^2_{B,l})\|L\|^2_{B,l}} \tag{3.1}$$

for all l. (K is chosen larger than certain universal constants that appear in the proof. $\bar K$ is chosen larger than some constant which depends on K. ϵ is chosen such that $\epsilon^2\bar K$ and $\epsilon^2 K$ are both smaller than some universal constant.) The inductive hypothesis is trivial for $B=\phi$ so we assume it true for all B's of at most some given size and prove it for $B^+ \equiv B\cup\{\alpha\}$ with $\alpha\notin B$. If $J(t) = tJ_\alpha + J_{(B)}$, $L(t) = t\delta L + L_{(B)}$, $\delta L = L_{(B^+)} - L_{(B)}$ then

$$\langle e^{\hat{\mathscr{J}}(B^+) + \hat{\mathscr{L}}(B^+)}\rangle_l = \langle e^{\hat{\mathscr{J}}(B) + \hat{\mathscr{L}}(B)}\rangle_l + \int_0^1 dt\, \langle[\hat{\mathscr{J}}_\alpha + \delta\hat{\mathscr{L}}]\, e^{\hat{\mathscr{J}}(t) + \hat{\mathscr{L}}(t)}\rangle_l. \tag{3.2}$$

Again the subtractions inherent in $[\hat{\mathscr{J}}_\alpha + \delta\hat{\mathscr{L}}]$ are implemented through the use of duplicate fields. We introduce five independent copies of the $P(\phi)_2$ theory labelled $\phi_r, \phi_y, \phi_g, \phi_b, \phi_v$, respectively. Then

$$\begin{aligned}\langle[\hat{\mathscr{J}}_\alpha + \delta\hat L]\, f(\phi)\rangle_l &= \langle J(\phi_r - \phi_y) f_r\rangle_l + \tfrac12\langle\{(\phi_r - \phi_y)\,\delta L\,(\phi_r - \phi_g)\\ &\quad - (\phi_b - \phi_y)\,\delta L\,(\phi_b - \phi_g)\}\,(f_r - f_b)\rangle_l,\end{aligned} \tag{3.3}$$

where $J\phi \equiv \int J(x)\,\phi(x)\,dx$, etc. The net effect of these subtractions in the cluster expansion is to require in the first term that $\hat{\mathscr{J}}$ be connected to f and in the second term that both legs of $\delta\hat{\mathscr{L}}$ be connected to f. (We will continue using the notation $\langle[\hat{\mathscr{J}}_\alpha + \delta\hat{\mathscr{L}}]f\rangle$). By the cluster expansion [using the same notation as (2.4)]

$$\begin{aligned}\langle[\hat{\mathscr{J}}_\alpha + \delta\hat L]\, e^{\hat{\mathscr{J}}(t) + \hat{\mathscr{L}}(t)}\rangle_l &= \sum_X \sum_{\substack{\Gamma_1\subset \Gamma_X\\ \Gamma_2\subset\mathscr{B}\setminus X\\ \Gamma_2 \text{ finite}}} \int_0^{S(\Gamma_1\cup\Gamma_2)} \frac{1}{Z_{\Lambda,l}}\\ &\quad\times \partial^{\Gamma_1\cup\Gamma_2}\langle[\hat{\mathscr{J}}_\alpha + \delta\hat L]\, e^{\mathscr{J}(t)^l + \mathscr{L}(t)^l}\rangle^u_{\Lambda,\sigma(\Gamma_1\cup\Gamma_2)}\\ &\quad\times d\sigma(\Gamma_1\cup\Gamma_2).\end{aligned} \tag{3.4}$$

In the $\mathscr{J}^l$ and $\mathscr{L}^l$ terms of the exponential, we do not make the ˆ subtractions $\sigma(\Gamma)$-dependent ($\Gamma = \Gamma_1\cup\Gamma_2$), but we do introduce a $\sigma(\Gamma)$ dependence in the bare Wick ordering in $\mathscr{L}$, i.e.,

$$\mathscr{J}^l = \mathscr{J} - J\langle\phi\rangle_l = J(\phi - \langle\phi\rangle_l) \equiv J\phi^l,$$

$$\mathscr{L} = :\phi L\phi:_{C(\sigma(0\Gamma))},$$

$$\mathscr{L}^l = \mathscr{L} - 2\langle\phi\rangle_l L\phi + \langle\phi\rangle_l L\langle\phi\rangle_l - \operatorname{tr} LS_l,$$

where

$$S_l = \langle\phi;\phi\rangle_l - C_l.$$

The fact that the ˆ subtractions are appropriate to the expectation $\langle\ \rangle_l$ rather than $\langle\ \rangle_{\bar l}$ causes no problem for those parts of $\mathscr{J}^l$ and $\mathscr{L}^l$ supported in X: we easily absorb the errors which are proportional to $|X|$. However, as we shall see, we must have the correct (i.e., $\langle\ \rangle_{\bar l}$) subtractions in X^c in order to apply the inductive hypothesis there. Accordingly, we rewrite $\mathscr{J}^l$ and $\mathscr{L}^l$ as follows. First we decompose L as a sum of three pieces with arguments in $X\times X$, $X\times X^c\cup X^c\times X$, and $X^c\times X^c$, respectively,

$$L_X(x,y) = \chi_X(x)\, L(x,y)\, \chi_X(y),$$

$$L_M(x,y) = \chi_X(x)\, L(x,y)\, \chi_{X^c}(y) + \chi_{X^c}(x)\, L(x,y)\, \chi_X(y),$$

$$L_{X^c}(x,y) = \chi_{X^c}(x)\, L(x,y)\, \chi_{X^c}(y),$$

and similarly for δL_X and δL_M. We then replace each ϕ^l in X^c by

$$\phi^l = \phi^{\bar l} - (\langle\phi\rangle_l - \langle\phi\rangle_{\bar l}) \equiv \phi^l - \delta\langle\phi\rangle.$$

The resulting straightforward calculation yields

$$\mathscr{J}^l + \mathscr{L}^l = \bar{\mathscr{J}}^{\bar l}_{X^c} + \mathscr{L}^{\bar l}_{X^c} + \phi L^*_M \phi^{\bar l} + \bar{\mathscr{J}}_X + \mathscr{L}_X + c(J,L),$$

where

$$\bar J_{X^c}(x) = J_{X^c}(x) - \int dy\, [2L_{X^c}(x,y)\delta\langle\phi(y)\rangle + L^*_M(x,y)\langle\phi(y)\rangle_l],$$

$$\bar J_X(x) = J_X(x) - \int dy\, [2L_X(x,y)\langle\phi(y)\rangle_l + L^*_M(x,y)\,\delta\langle\phi(y)\rangle],$$

$$\begin{aligned}c(J,L) &= \langle\mathscr{J}_{X^c}\rangle_{\bar l} - \langle\mathscr{J}\rangle_l + \delta\langle\phi\rangle\, L_{X^c}\,\delta\langle\phi\rangle\\ &\quad - \langle\phi\rangle_l L^*_M\,\delta\langle\phi\rangle - \langle\phi\rangle_l L_X\langle\phi\rangle_l\\ &\quad + \operatorname{tr}[L_X S_{\bar l} + L(S_{\bar l} - S_l)],\end{aligned}$$

and

$$L^{*}_{M}(x,y) = 2\chi_X(x)\, L(x,y)\, \chi_{X^c}(y)\,.$$

The constant terms are bounded in

Lemma III.2:

$$|c(J,L)| \leqslant c_{22}|X| + c_{23}\{\| J\|^2_{2,B^+,l} - \| J\chi_{X^c}\|^2_{2,B,\bar{l}}\}\,.$$

For future reference we also observe that $\bar{\mathscr{J}}^{\bar{l}}_{X^c}$, $\mathscr{L}(t)^{\bar{l}}_{X^c}$, and $\phi L^{*}_{M}\phi^{\bar{l}}$ have precisely the right subtractions in X^c and obey

Lemma III.3:

$$\begin{aligned}
&\|\bar{J}_{X^c}\|^2_{B,\bar{l}}\\
&\quad\leqslant \| J_{X^c}\|^2_{B,\bar{l}} + c_{24}K \sum_{Y\in\mathscr{C}(l)} \sum_{\gamma\in Y\cap X} \sum_{\beta_i\in B^+} e(\gamma,\beta_1)\\
&\qquad\times \|J_{X^c}\|^2_{B,\bar{l}} \|\chi_Y L_{\beta_1,\beta_2}\|^2_{-1,l} + c_{25}\frac{|X|}{K}\,,\\
&\|\bar{J}_X\|_{-1,\bar{l}}\\
&\quad\leqslant \| J_X\|_{-1,\bar{l}} + c_{26}|X|^{1/2}\{\|C^{1/2}_{\bar{l}}\chi_X L(t)\, C^{1/2}_{l}\|_{L^2}\\
&\qquad + \| L(t)_X\|_{-1,\bar{l}}\}\,.
\end{aligned}$$

We are now confronted with the one substantial difference between Theorems II.1 and III.1. We cannot, at this stage, apply the cluster expansion (2.4) since the integral $\langle\cdot\rangle^u_{A,\sigma(\Gamma_1\cup\Gamma_2)}$ does not factor even though the covariance $C_{\bar{l}}(\sigma(\Gamma_1\cup\Gamma_2))$ decouples X and X^c. The culprit is $e^{\phi L^{*}_{M}\phi^{\bar{l}}}$, which does couple X and X^c. We will expand $e^{\phi L^{*}_{M}\phi^{\bar{l}}}$, for then each resulting term does factor. However, the factor in X^c will not be just a partition function as in Ref. 6, but will contain a product of fields as well—the fields of $(\phi L^{*}_{M}\phi^{\bar{l}})^n$ that live in X^c. This factor will be estimated by applying the analyticity argument of Theorem A.1 to the bound (3.1) of the inductive hypothesis much as was done in Corollary II.7.

Let us start by expanding the exponentials $e^{\phi L^{*}_{M}\phi^{\bar{l}}}$, $e^{\bar{\mathscr{J}}X}$, and $e^{\mathscr{L}(t)X}$. Each resulting term factors and the $\delta\hat{\mathscr{L}}_M$ contribution to (3.4) becomes

$$\begin{aligned}
&\left\langle \delta\hat{\mathscr{L}}_M\, e^{\mathscr{J}(t)^l + \mathscr{L}(t)^l}\right\rangle_l\\
&\quad= \sum_{X,\Gamma\in I_X}\sum_{n,p,k}\frac{e^{c(J,L)}}{n!\,p!\,k!}\int_0^{s(\Gamma)} d\sigma(\Gamma)\\
&\qquad\times\partial^\Gamma\left\langle \bar{\mathscr{J}}^n_X\,\mathscr{L}(t)^p_X\prod_{i=1}^{k+1}\phi(x_i)\right\rangle^u_{X,\sigma(\Gamma)} \delta L^{*}_{M}(x_1,y_1)\\
&\qquad\times\prod_{i=2}^{k+1} L^{*}_{M}(x_i,y_i)\left\langle\prod_{i=1}^{k+1}\phi^{\bar{l}}(y_i)\, e^{\bar{\mathscr{J}}^{\bar{l}}_{X^c} + \mathscr{L}^{\bar{l}}_{X^c}}\right\rangle_{A\setminus X,\bar{l}}\\
&\qquad\times\frac{Z_{A\setminus X,\bar{l}}}{Z_{A,l}}\,. \qquad (3.5)
\end{aligned}$$

The contributions from $\hat{\mathscr{J}}_\alpha$ and $\delta\hat{\mathscr{L}}_X$ are similar (but easier to handle) and will not be considered explicitly. Thanks to the (color coded) subtractions in $[\hat{\mathscr{J}}_\alpha + \delta\hat{\mathscr{L}}]$ the $n = p = k = 0$ term is zero. We have suppressed the integral signs $\int\Pi\, dx_i\, dy_i$ under the convention that repeated arguments are integrated over.

The expansion (3.5) is controlled in several stages. In stage 1 the ∂^Γ is evaluated and localizations are introduced. We then consider a single resulting term. In stage 2 analyticity and the inductive hypothesis are used to bound that portion of the term that lives in $A\setminus X$. In stage 3 those J's and L's whose ϕ's have been destroyed by $\delta/\delta\phi$'s are separated off and estimated. In stage 4 we finish bounding that portion of the term that lives in X. Finally, in stage 5 the remaining sums are controlled.

Stage 1. As in Lemma II.2 we apply $\partial^\Gamma = \Sigma_{\pi\in\mathscr{P}(\Gamma)}\Sigma_{\gamma\in\pi}\frac{1}{2}\partial^\gamma C\delta^2_\phi$, we classify the resulting terms according to which δ_ϕ's attack interaction vertices and according to which $\mathscr{J}$'s and $\mathscr{L}$ legs are attacked (the two ϕ's in each $\mathscr{L}$ are referred to as $\mathscr{L}$ legs), and we localize these δ_ϕ's, $\mathscr{J}$'s, and $\mathscr{L}$ legs. Let

n_1 = number of $\mathscr{J}$'s attacked,

p_j = number of $\mathscr{L}_X$'s that have had j legs attacked,

k_1 = number of L^{*}_{M}'s that had their one X leg attacked, and

$M(\Delta)$ = number of interaction legs, $\mathscr{J}$'s, and $\mathscr{L}$ legs localized in Δ.

Then the number of classes is at most $e^{3(|\Gamma|+n+p+k)}$ and the number of terms in a class that have any given fixed set of localizations is $\Pi_\Delta e^{c_9 M(\Delta)}M(\Delta)!^\upsilon$. These bounds are controlled as in Lemma II.2. The sum over localizations will be controlled later and until then we generally suppress the notation specifying the localization.

Stage 2. We now estimate any one of these terms. We first focus our attention on the L^{*}_{M}'s. We have the following structure:

$$\begin{aligned}
&\int d\mu(\phi)_{C_{\bar{l}}(\sigma(\Gamma))}\cdots\prod^{k_1}\partial^\gamma C(\cdot,x_i)\, L^{*}_{M}(x_i,y_i)\\
&\quad\times\prod^{k-k_1+1}\phi(x_i)\, L^{*}_{M}(x_i,y_i)\left\langle\prod_{i=1}^{k+1}\phi^{\bar{l}}(y_i)e^{\bar{\mathscr{J}}^{\bar{l}}_{X^c}+\mathscr{L}^{\bar{l}}_{X^c}}\right\rangle_{A\setminus X,\bar{l}}\\
&\quad\times e^{c(J,L)}\frac{Z_{A\setminus X,\bar{l}}}{Z_{A,l}}\,,
\end{aligned}$$

so that each L^{*}_{M} must occur in one of the following configurations:

$$\begin{aligned}
&\int dx\,\phi(x)\, L^{*}_{M}(x,y),\\
&\int dx\,dx'\,\chi_{\Delta_\alpha}(C^{-1/2}_{\bar{l}}\partial^\gamma CC^{-1/2}_{\bar{l}})(x,x')\,\chi_{\Delta_\beta}\\
&\quad\times(C^{1/2}_{\bar{l}}L^{*}_{M})(x,y)(C^{1/2}_{\bar{l}}L^{*}_{M})(x',y'),\\
&\int dx(\partial^\gamma CC^{-1/2}_{\bar{l}})(\cdot,x)\,\chi_{\Delta_\alpha}(C^{1/2}_{\bar{l}}L^{*}_{M})(x,y)\,,
\end{aligned}$$

where in the last case the other end of the $\partial^\gamma C$ is not hooked to an L^{*}_{M}.

Apply analyticity in $\mathscr{J}_{X^c}$ and the inductive hypothesis in the form of the following lemma (which will be proven later).

Lemma III.4: For the constants K and $\bar K$ in the inductive hypothesis (3.1)

$$\left|\left\langle \prod_{i=1}^{n} \phi^{\bar l}(\chi_{X^c} f_{i,B})\, e^{\bar{\mathscr{J}}^{\bar l}_{X^c} + \mathscr{L}^{\bar l}_{X^c}} \right\rangle_{A\setminus X,\bar l}\right| e^{c(J,L)} \frac{Z_{A\setminus X,\bar l}}{Z_{A,l}}$$

$$\leqslant (n!)^{1/2} K'^{n}\left(1 + \frac{\|J_{X^c}\|_{-1,\bar l}}{n^{1/2}}\right)^{n} \prod_{i=1}^{n} \|\chi_{X^c} f_{i,B}\|_{-1,\bar l}$$

$$\times \exp\big\{c_{27}|X| + K(\|J_{X^c}\|^2_{1,B,\bar l} + \|J\|^2_{2,B^+,l})$$

$$+ \bar K(1 + \|J_{X^c}\|^2_{B,\bar l})(\|L_{X^c}\|^2_{1,B,\bar l} + \|L\|^2_{2,B^+,l})\big\},$$

where K' depends on K *but not on* $\bar K$.

This estimate is applied with the f_i's being the $L^*_M(x,y)$'s or $C_{\bar l}^{1/2} L^*_M(x,y)$'s with x [and the ϕ in the $d\mu_{C_{\bar l}(\sigma(\Gamma))}$ integral] held fixed. The $\|\chi_{X^c} f_{i,B}\|_{-1,\bar l}$ norms that result for our three configurations are

$$\|L^*_M(\phi,\cdot)\|_{-1,\bar l} = \left\|\int dx\, \phi(x) L^*_M(x,\cdot)\right\|_{-1,\bar l},$$

$$\int dx\, dx' \|(C_{\bar l}^{1/2} L^*_M)(x,\cdot)\|_{-1,\bar l}$$

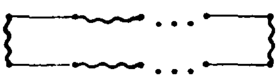
$$\times (\chi_{\Delta_\alpha} C_{\bar l}^{-1/2} \partial^\gamma C C_{\bar l}^{-1/2} \chi_{\Delta_\beta})(x,x')$$

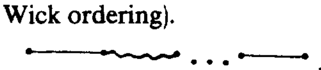
$$\times \|(C_{\bar l}^{1/2} L^*_M)(x,\cdot)\|_{-1,\bar l},$$

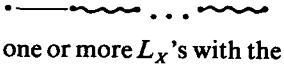
$$\int dx\, (\partial^\gamma C C_{\bar l}^{-1/2} \chi_{\Delta_\alpha})(\cdot,x)\, \|(C_{\bar l}^{1/2} L^*_M)(x,\cdot)\|_{-1,\bar l}.$$

The notation $\|f(x,\cdot)\|_{-1,\bar l}$ indicates that the $\|\ \|_{-1,\bar l}$ norm has been applied to $f(x,y)$ viewed as a function of the variable y with x held fixed.

Stage 3. Now that we have estimated the L^*_M's we analyze the other factors in the term we started to consider in stage 2. It is an integral with respect to $d\mu_{C_{\bar l}(\sigma(\Gamma))}$ of the interaction $e^{-V(\phi)}$ multiplied by a function of ϕ, the J's, L_X's, L^*_M's, and the $\partial^\gamma C$'s. For notational convenience think of this function as a graph having one vertex for each position space integration variable x_i or u_i (the u's being the position of the interaction vertices downstairs) and one edge for each δL_X, L_X (denoted ∿∿) or $\partial^\gamma C$ (denoted ———). In addition, some of the vertices may be multiplied by a $\bar J_X(x)$ or $:\phi^n(x):$. We are now primarily interested in the dependence on J and L. Each connected component of the graph containing a J or L must be of the following types (just classify all strings of L_X's according to what their ends look like):

Type I:  (there must be at least 2 L_X's by Wick ordering).

Type II: f_1  f_2 (there must be at least one $\partial^\gamma C$), where $f_i(x)$ is one of $J_\alpha(x)$, $\bar J_X(x)$, $C_{\bar l}^{-1/2}\|(C_{\bar l}^{1/2} L^*_M)(x,\cdot)\|_{-1,\bar l}$, and $C_{\bar l}^{-1/2}\|(C_{\bar l}^{1/2}\delta L^*_M)(x,\cdot)\|_{-1,\bar l}$.

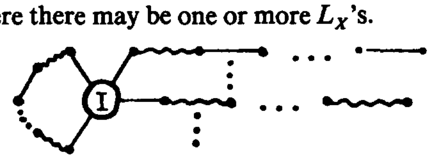
Type III (*a*): $f(x)$  ϕ.

Here there are one or more L_X's with the ϕ belonging to the last one. $f(x)$ is as in Type II.

Type III (*b*): $\bar{\mathscr{J}}_X$, $\|L^*_M(\phi,\cdot)\|_{-1,\bar l}$, or $\|\phi L^*_M(\phi,\cdot)\|_{-1,\bar l}$.

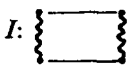
Type IV: ϕ  ϕ.

Here there may be one or more L_X's.

Type V:  f, ϕ,

where I contains interaction vertices and $\partial^\gamma C$'s only and f is as in Type II (i.e., a J or L_M vertex).

They are dealt with as follows. We emphasize that this is all done inside the $d\mu(\phi)$ integral so that ϕ is fixed. We lump all the usual cluster expansion decay factors into a factor that we denote D. Also recall that we have suppresed localization notations.

Type I: 

$$= \mathrm{tr}(L_X \partial^\gamma C)^2 = \mathrm{tr}(C_{\bar l}^{1/2} L_X C_{\bar l}^{1/2} C_{\bar l}^{-1/2} \partial^\gamma C C_{\bar l}^{-1/2})^2$$

$$\leqslant D\|L_X\|^2_{-1,\bar l} \quad \text{(by Lemma B.2)}.$$

Longer loops are handled similarly using

$$\|C_{\bar l}^{1/2} L C_{\bar l}^{1/2}\|_{L^2\to L^2} \leqslant \|C_{\bar l}^{1/2} L C_{\bar l}^{1/2}\|_{\mathrm{H.S.}} \leqslant \|L\|_{-1,\bar l}.$$

Type II: $(f_1, \partial^\gamma C L_X \partial^\gamma C \cdots \partial^\gamma C f_2) \leqslant D\|C_{\bar l}^{1/2} f_1\|_{L^2}\|L_X\|_{-1,\bar l} \cdots \|C_{\bar l}^{1/2} f_2\|_{L^2}$. When $f(x) = C_{\bar l}^{-1/2}\|C_{\bar l}^{1/2} L^*_M(x,\cdot)\|_{-1,\bar l}$ we have $\|C_{\bar l}^{1/2} f\|_{L^2} \leqslant \|L^*_M\|_{-1,\bar l} \leqslant 2\|L_M\|_{-1,\bar l}$; when $f = \bar J_X$ $\|C_{\bar l}^{1/2} f\|_{L^2} \leqslant \|\bar J_X\|_{-1,\bar l}$.

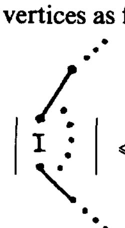
Type III (*a*): $|f\partial^\gamma C L_X \cdots \partial^\gamma C L_X \phi|$

$$\leqslant D\|C^{1/2} f\|_{L^2}\, \|L_X\|_{-1,\bar l} \cdots \|L_X(\phi,\cdot)\|_{-1,\bar l}.$$

Type III (*b*): These are, for the time being, left in peace.

Type IV: Except for $\phi L_X \phi$, which is left in peace, type IV components are estimated in the same manner as type III(b) components.

Type V: Each connected component of type V is dealt with by using Cauchy–Schwarz to separate off the interaction vertices as follows:



$$\leqslant \left|\sum_{\alpha_i} \textcircled{I} \begin{matrix} C_{\bar l}^{1/2}\chi_{\Delta_{\alpha_1}} C_{\bar l}^{-1/2} \text{—} \cdots \\ \vdots \\ C_{\bar l}^{1/2}\chi_{\Delta_{\alpha_n}} C_{\bar l}^{-1/2} \text{—} \cdots \end{matrix}\right|$$

$$\leqslant \sum_{\alpha_i} \left[\textcircled{I} \begin{matrix} C_{\bar l}^{1/2}\chi_{\Delta_{\alpha_1}} C_{\bar l}^{1/2} \\ \vdots \\ C_{\bar l}^{1/2}\chi_{\Delta_{\alpha_n}} C_{\bar l}^{1/2} \end{matrix} \textcircled{I}\right]^{1/2}$$

$$\times \left[\begin{matrix} \cdots \text{—} C_{\bar l}^{1/2}\chi_{\Delta_{\alpha_1}} C_{\bar l}^{-1/2} \text{—} \cdots \\ \vdots \\ \cdots \text{—} C_{\bar l}^{-1/2}\chi_{\Delta_{\alpha_n}} C_{\bar l}^{-1/2} \text{—} \cdots \end{matrix}\right]^{1/2}.$$

Those connected components in the second set of brackets that do not have any ϕ dependence have the structure of type I or II components and are estimated as such. All other components in the second set of brackets are of the form

$$[\phi L_X \partial^\gamma C \cdots L_X \phi]^{1/2}$$

$$\leqslant D\|L_X(\phi,\cdot)\|_{-1,\bar l} \prod \|L_X\|_{-1,\bar l},$$

where the Π contains precisely one factor for each L_X in the interior of the original type V string.

Stage 4. Let us now stop to reorient ourselves. We first used the inductive hypothesis and analyticity to bound that part of our term localized in $\Lambda \setminus X$. This yielded an exponential factor

$$e^{c_{28}|X|}\exp\Big[K(\|J_{X^c}\|^2_{1,B,\bar l}+\|J\|^2_{2,B+,l})+\bar K(1+\|J_{X^c}\|^2_{B,\bar l})(\|L_{X^c}\|^2_{1,B,\bar l}+\|L\|^2_{2,B+,\bar l})\Big],$$

a combinatoric factor $(k+1)!^{1/2}$, and a factor of $K'(1+\|J_{X^c}\|_{-1,\bar l}/k^{1/2})$ for each factor L^*_M and δL^*_M. The $d\mu(\phi)_{C_{\mathcal{I}\sigma(\Gamma)}}$ integrand then consisted of the interaction, some interaction vertices, some $\bar{\mathscr{J}}_X$ vertices, some $\|L^*_{\bar l t}(\phi,\cdot)\|_{-1,\bar l}$ [and $\|\delta L^*_M(\phi,\cdot)\|_{-1,\bar l}$] vertices, some $\mathscr{L}_X$'s as well as $\partial^\gamma C$'s, J's, and L's whose ϕ's had been attacked by $\delta/\delta\phi$'s. Some of the latter have now been estimated out of existence. Each $\partial^\gamma C$ that has been estimated has been replaced by its usual cluster expansion decay factors. Each J and L that has been estimated has been replaced by $\|J\|_{-1,\bar l}$ or $\|L\|_{-1,\bar l}$. The rest of the term has been bounded by

$$\int d\mu(\phi)\,|F_1(\phi)|\,|F_2(\phi)|,$$

where $F_1(\phi)$ contains the interaction e^{-V_X}, interaction vertices, and $C_{\bar l}^{1/2}\chi_{\Delta_\alpha}C_{\bar l}^{1/2}$ and $\partial^\gamma C$ propagators but no L's, J's, or ϕ's belonging to L's or J's (it is what is left of the type V components) and $F_2(\phi)$ contains

$\|L_X(\phi,\cdot)\|_{-1,\bar l}$'s	(from type III(a), IV, and V components),
$\bar J_X\phi$'s	(type III(b) components),
$\|(\delta)L^*_M(\phi,\cdot)\|_{-1,\bar l}$'s	(type III(b) components),
$:\phi L_X\phi:$	(type IV components).

We simply use the Schwarz inequality to separate F_1 and F_2. The F_1 factor is of classic CE form and is estimated as such. The F_2 factor is simply integrated as the Gaussian integral of a polynomial.

It may be bounded by

$$e^{c_{29}(p+k+n)}\frac{(p!)(k!)^{1/2}(n!)^{1/2}}{(p_1+2p_2+k_1+n_1)!^{1/2}}$$

times a product containing one $\|J\|_{-1,\bar l}$ for each $J\phi$ in F_2 and one $\|L\|_{-1,\bar l}$ for each $\|L_X(\phi,\cdot)\|_{-1,\bar l}$, $\|(\delta)L^*_M(\phi,\cdot)\|_{-1,\bar l}$, and $:\phi L\phi:$ in F_2. Note that we obtain L^2-type norms on L because Wick ordering rules out graphs containing the factor Tr CL.

Stage 5. We have now bounded each term in the series that results when the ∂^Γ of (3.4) is evaluated and localizations are introduced. We have also shown how to control all sums with the exceptions of those over X,Γ,n,p,k,π and over the $\partial^\gamma C$, $\mathscr{J}$, and $\mathscr{L}$ localizations. The sums over π and $\partial^\gamma C$ localizations are controlled as in the standard CE (see Ref. 6, Proposition 8.1 and Lemma 10.2). The sum over $\mathscr{J}$ and $\mathscr{L}$ localizations are controlled by

$$\sum_{j\in X}\|\chi_{\Delta_j}C_{\bar l}^{1/2}J\|_{L^2}\leqslant|X|^{1/2}\|J_X\|_{-1,\bar l},$$

$$\sum_{j\in X}\|\chi_{\Delta_j}C_{\bar l}^{1/2}LC_{\bar l}^{1/2}\|_{\rm H.S.}\leqslant|X|^{1/2}\|\chi_X L\|_{-1,\bar l},$$

$$\sum_{i,j\in X}\|\chi_{\Delta_i}C_{\bar l}^{1/2}LC_{\bar l}^{1/2}\chi_{\Delta_j}\|_{\rm H.S.}\leqslant|X|\,\|L_X\|_{-1,\bar l}.$$

In effect we get a factor of $|X|^{1/2}$ for each of the $n_1+k_1+p_1+2p_2$ $\mathscr{J}$ and $\mathscr{L}$ localizations we introduced.

Altogether we have

$$\Big|\Big\langle[\hat{\mathscr{J}}_\alpha+\delta\hat{\mathscr{L}}]\,e^{\mathscr{J}(t)^l+\mathscr{L}(t)^l}\Big\rangle_l\Big|$$

$$\leqslant\sum_{X,\Gamma}\sum_{n+p+k\neq0}\frac{1}{n!\,p!\,k!}e^{c_{30}(|\Gamma|+|X|+n+p+k)-K(m_0)|\Gamma|}$$
$$\times\frac{p!k!^{1/2}n!^{1/2}}{(n_1+k_1+p_1+2p_2)!^{1/2}}|X|^{(1/2)(n_1+k_1+p_1+2p_2)}$$
$$\times\Big\{\|J_\alpha\|_{-1,\bar l}+\|\delta L_X\|_{-1,\bar l}+\|\delta L_M\|_{-1,\bar l}$$
$$\times K'\Big(1+\frac{\|J_{X^c}\|_{-1,\bar l}}{(1+k)^{1/2}}\Big)\Big\}$$
$$\times(k!)^{1/2}\|\bar J_X\|^n_{-1,\bar l}\,\|L(t)_X\|^p_{-1,\bar l}$$
$$\times\Big\{\|L_M(t)\|_{-1,\bar l}K'\Big(1+\frac{\|J_{X^c}\|_{-1,\bar l}}{k^{1/2}}\Big)\Big\}^k$$
$$\times\exp\Big\{K(\|J_{X^c}\|^2_{1,B,\bar l}+\|J\|^2_{2,B+,l})$$
$$+\bar K(1+\|J_{X^c}\|^2_{B,\bar l})(\|L_{X^c}\|^2_{1,B,\bar l}+\|L\|^2_{2,B+,l})\Big\}$$
$$\leqslant\sum_{X,\Gamma}e^{c_{31}|X|-1/2K(m_0)|\Gamma|}\{\|J_\alpha\|_{-1,\bar l}+\|\delta L_X\|_{-1,\bar l}$$
$$+K'\|\delta L_M\|_{-1,\bar l}(1+\|J_{X^c}\|_{-1,\bar l})\}$$
$$\times\{\|\bar J_X\|_{-1,\bar l}+\|L(t)_X\|_{-1,\bar l}+K'\|L_M(t)\|_{-1,\bar l}$$
$$\times(1+\|J_{X^c}\|_{-1,\bar l})\}$$
$$\times\Big\{\sum_{n=0}^\infty\frac{1}{(n!)^{1/2}}(A_1\|\bar J_X\|_{-1,\bar l})^n\Big\}$$
$$\times\Big\{\sum_{p=0}^\infty(A_2\|L(t)_X\|_{-1,\bar l})^p\Big\}$$
$$\times\Big\{\sum_{k=0}^\infty\Big[A_3K'\|L_M(t)\|_{-1,\bar l}\Big(1+\frac{\|J_{X^c}\|_{-1,\bar l}}{k^{1/2}}\Big)\Big]^k\Big\}$$
$$\times\exp\{\cdots\}$$
$$\leqslant\sum_{X,\Gamma}e^{c_{32}|X|-1/2K(m_0)|\Gamma|}\{\|J_\alpha\|_{-1,\bar l}+\|\delta L_X\|_{-1,\bar l}$$
$$+K'\|\delta L_M\|_{-1,\bar l}(1+\|_{X^c}\|_{-1,\bar l})\}$$
$$\times\{\|J(t)_X\|_{-1,\bar l}+\|C_{\bar l}^{1/2}\chi_X L(t)C_{\bar l}^{1/2}\|_{L^2}$$
$$+\|L(t)_X\|_{-1,l}+K'\|L(t)_M\|_{-1,\bar l}(1+\|J_{X^c}\|_{-1,\bar l})\}$$
$$\times\exp\Big\{A_1^2\|J(t)_X\|^2_{-1,\bar l}$$
$$+4A_3^2K'^2\|L(t)_M\|^2_{-1,\bar l}\|J_{X^c}\|^2_{-1,\bar l}$$
$$+K(\|J_{X^c}\|^2_{1,B,\bar l}+\|J\|^2_{2,B+,l})$$
$$+\bar K(1+\|J_{X^c}\|^2_{B,\bar l})(\|L_{X^c}\|^2_{1,B,\bar l}+\|L\|^2_{2,B+,l})\Big\},$$

where we have used

Lemma III.3(b) and $|X|^{1/2}\leqslant e^{1/2|X|}$ on the first $\|\bar J_X\|_{-1,\bar l}$,

Lemma III.3(b) and $|L|<\epsilon$ on the other $\|\bar J_X\|_{-1,\bar l}$,

(2.20) to perform the sum over n,

$A_2\,\epsilon<\frac12$ to perform the sum over p,

Lemma III.5 and $A_3K'\,\epsilon<\frac14$ to perform the sum over p,

$$<\sum_{X,\Gamma} e^{c_{33}|X|-(1/2)K(m_0)|\Gamma|}\{\|J_\alpha\|^2_{-1,\bar l}+\|J(t)_X\|^2_{-1,\bar l}$$

$$+\|\delta L_X\|^2_{-1,\bar l}+\|L(t)_X\|^2_{-1,\bar l}$$

$$+\|C_{\bar l}^{1/2}\chi_X L(t)C_l^{1/2}\|^2_{L^2}+K'^2(1+\|J_{X^c}\|^2_{-1,\bar l})$$

$$\times(\|\delta L_M\|^2_{-1,\bar l}+\|L(t)_M\|^2_{-1,\bar l})\}$$

$$\times\exp\{K[\|J\|^2_{1,B,l}+t(\|J\|^2_{1,B^+,l}-\|J\|^2_{1,B,l})$$

$$+\|J\|^2_{2,B^+,l}]$$

$$+\bar K(1+\|J_{X^c}\|^2_{B,\bar l})[\|L\|^2_{1,B,l}+t(\|L\|^2_{1,B^+,l}$$

$$-\|L\|^2_{1,B,l})$$

$$+\|L\|^2_{2,B^+,l}]\},$$

where we have used

Lemma II.6 since we require

$K>\max\{2A_1^2 e_2, 2A_1^2 e_1^{-1}\}$,

Lemma III.6 since we require $\bar K>2e_2^4(4A_3^2 K'^2)$,

$$<\sum_{X,\Gamma} e^{c_{34}|X|-(1/2)K(m_0)|\Gamma|}\Big\{\sum_{\beta\in B^+\cap X}\|J_\beta\|^2_{-1,l}$$

$$+\sum_{\beta_i\in B^+\cap X}\|L_{\beta_1,\beta_2}\|^2_{-1,l}$$

$$+K'^2(1+\|J_{X^c}\|^2_{-1,\bar l})\sum_{\substack{\beta_1\text{ or }\beta_2\text{ in}\\ B^+\cap X}}\|L_{\beta_1,\beta_2}\|^2_{-1,l}\Big\}$$

$$\times\exp\{\cdots\}$$

$$<\sum_{X,\Gamma} e^{c_{35}|X|-(1/4)K(m_0)|\Gamma|}\Big\{\sum_{\beta\in B^+} e(\alpha,\beta)\|J_\beta\|^2_{-1,l}$$

$$+K'^2(1+\|J\|^2_{B,l})\sum_{\beta_i\in B^+}(e(\beta_1,\alpha)$$

$$+e(\alpha,\beta_2))\|L_{\beta_1,\beta_2}\|^2_{-1,l}\Big\}$$

$$\times\exp\{\cdots\}$$

as in Lemma II.2,

$$<\frac{d}{dt}\exp\{K[\|J\|^2_{1,B,l}+t(\|J\|^2_{1,B^+,l}-\|J\|^2_{1,B,l})$$

$$+\|J\|^2_{2,B^+,l}]$$

$$+\bar K(1+\|J\|^2_{B,l})[\|L\|^2_{1,B,l}+t(\|L\|^2_{1,B^+,l}$$

$$-\|L\|^2_{1,B,l})+\|L\|^2_{2,B^+,l}]\} \tag{3.6}$$

as in (2.6) if $K>\sum_{X,\Gamma} e^{c_{35}|X|-(1/4)K(m_0)|\Gamma|}$ and $\bar K>K'^2\sum_{X,\Gamma} e^{c_{35}|X|-(1/4)K(m_0)|\Gamma|}$.

Substituting (3.6) into (3.2) and applying the inductive hypothesis (3.1) gives

$$|\langle e^{\mathscr{J}(B^+)+\hat{\mathscr{L}}(B^+)}\rangle_l|$$

$$\leqslant\exp[K\|J\|^2_{B,l}+\bar K(1+\|J\|^2_{B,l})\|L\|^2_{B,l}]$$

$$+\int_0^1 dt\frac{d}{dt}\exp\{K[\|J\|^2_{1,B,l}+t(\|J\|^2_{1,B^+,l}$$

$$-\|J\|^2_{1,B,l})+\|J\|^2_{2,B^+,l}]$$

$$+\bar K(1+\|J\|^2_{B,l})[\|L\|^2_{1,B,l}+t(\|L\|^2_{1,B^+,l}$$

$$-\|L\|^2_{1,B,l})+\|L\|^2_{2,B^+,l}]\}$$

$$\leqslant\exp[K\|J\|^2_{B^+,l}+\bar K(1+\|J\|^2_{B^+,l})\|L\|^2_{B^+,l}]$$

$$=\exp[(K+\bar K\|L\|^2_{B^+,l})\|J\|^2_{B^+,l}+\bar K\|L\|^2_{B^+,l}].$$

The theorem follows. ■

Lemma III.2:

$$|c(J,L)|\leqslant c_{22}|X|+c_{23}\big[\|J\|^2_{2,B^+,l}-\|J_{X^c}\|^2_{2,B,\bar l}\big].$$

Proof: We already know

$$|\langle\mathscr{J}(t)_{X^c}\rangle_{\bar l}-\langle\mathscr{J}(t)\rangle_l|$$

$$\leqslant c_4|X|+c_5\big[\|J\|^2_{2,B^+,l}-\|J_{X^c}\|^2_{2,B,\bar l}\big]$$

by (2.7) and Lemma II.4. The terms $\delta\langle\phi\rangle L(t)_{X^c}\delta\langle\phi\rangle$, $\langle\phi\rangle_l L^*_M\delta\langle\phi\rangle$, and $\langle\phi\rangle_l L(t)_X\langle\phi\rangle_l$ are all the form $\int dx\,dy\,L(t)(x,y)f_1(x)f_2(y)$ with

$$\|f_i\|_{+1}\leqslant c_{36}^{1/2}|X|^{1/2}$$

so that

$$\Big|\int dx\,dy\,L(t)(x,y)f_1(x)f_2(y)\Big|\leqslant c_{36}\|L(t)\|_{-1}|X|\leqslant c_{37}|X|.$$

The remaining terms are

$$\operatorname{tr}L(t)_X S_{\bar l}$$

$$=\langle L(t);\chi_X S_{\bar l}\chi_X\rangle_{L^2(\mathbf{R}^4)}\leqslant\|L(t)\|_{-1,\bar l}\|\chi_X S_{\bar l}\chi_X\|_{+1,\bar l}$$

$$\leqslant\|L(t)\|_{-1,\bar l}c_{38}|X|^{1/2}\quad\text{(by Corollary B.4b)}$$

$$\leqslant c_{39}|X|,$$

$$\operatorname{tr}L(t)(S_{\bar l}-S_l)=\langle L(t),S_{\bar l}-S_l\rangle_{L^2(\mathbf{R}^4)}$$

$$\leqslant\|L(t)\|_{-1,l}\|S_{\bar l}-S_l\|_{+1,l}$$

$$\leqslant c_{39}|X|\quad\text{(by Corollary B.4c).}$$ ■

Lemma III.3:

(a) $\|\bar J_{X^c}\|^2_{B,\bar l}$

$$\leqslant \|J_{X^c}\|^2_{B,\bar l} + c_{24} K \sum_{Y\in\mathscr{C}(l)} \sum_{\gamma\in Y\cap X} \sum_{B_l\in B^+} e(\gamma,\beta_1)$$

$$\times \| J_{X^c}\|^2_{B,\bar l} \| \chi_Y L_{\beta_1,\beta_2}\|^2_{-1,l} + c_{25}|X|/K,$$

(b) $\|\bar J_X\|_{-1,\bar l} \leqslant \| J_X(t)\|_{-1,\bar l} + c_{26}|X|^{1/2}$

$$\times \{\| C_{\bar l}^{1/2} \chi_X L(t) C_l^{1/2}\|_{L^2} + \| L(t)_X\|_{-1,\bar l}\}.$$

Proof: (a) $\bar J_{X^c}(x) = J_{X^c}(x) - l(x)$,

where

$$l(x) = \int dy\, 2\delta\langle\phi(y)\rangle L_{X^c}(y,x) + \langle\phi(y)\rangle_l L^*_M(y,x)$$

$$= \int dy\, 2\delta\langle\phi(y)\rangle L(y,x)\chi_{X^c} + \langle\phi(y)\rangle_{\bar l} L^*_M(y,x),$$

so

$$\| J_{X^c}\|^2_{B,\bar l}$$

$$\leqslant \| J_{X^c}\|^2_{B,\bar l} + 2\| J_{X^c}\|_{B,\bar l} \| l\|_{B,\bar l} + \| l\|^2_{B,\bar l}$$

by Lemma B.5(a).

But applying

$$\left|\int dy\, \delta\langle\phi(y)\rangle f(y)\right|$$

$$\leqslant c_{40} \sum_{Y\in\mathscr{C}(l)} \sum_{\beta_2\in B^+} e^{-d(\beta_2,\partial X\cap Y)} \| \chi_Y f_{\beta_2}\|_{-1,l}$$

(see Lemma II.4) with $f(y) = L(y,x) C_{\bar l}^{1/2}$ gives

$$\left\| \int dy\, \delta\langle\phi(y)\rangle L(x,y)\chi_{X^c}\right\|_{B,\bar l}$$

$$\leqslant e_3 \left\| \int dy\, \delta\langle\phi\, y)\rangle L(y,x)\chi_{X^c}\right\|_{-1,\bar l}$$

$$\leqslant e_3 \left\| \int dy\, \delta\langle\phi(y)\rangle L(y,x) C_l^{1/2}\right\|_{L^2}$$

$$\leqslant c_{41} \sum_Y \sum_{\beta_2\in B^+} e^{-d(\beta_2,\partial X\cap Y)} \| \chi_Y L_{\beta_2,B^+}\|_{-1,l}$$

$$\leqslant c_{42} \sum_Y \sum_{\beta_1\in Y\cap X} \sum_{\beta_2\in B^+} e(\beta_1,\beta_2) \| \chi_Y L_{\beta_2,B^+}\|_{-1,l}\,.$$

Similarly

$$\left|\left| \int dy\, \chi_{X^c} L(x,y)\,\delta(y)\rangle\right|\right|_{B,\bar l}$$

$$\leqslant c_{42} \sum_Y \sum_{\beta_1\in Y\cap X} \sum_{\beta_2\in B^+} e(\beta_1,\beta_2) \| L_{B^+,\beta_2}\chi_Y\|_{-1,l}$$

and

$$\left|\left|\int_X dy\, \langle\phi(y)\rangle_{\bar l} L^*_M(y,x)\right|\right|_{B,\bar l}$$

$$\leqslant e_3 \left|\left| \int_X dy\, \langle\phi(y)\rangle_{\bar l} L^*_M(y,x)\right|\right|_{-1,\bar l}$$

$$\leqslant e_3 \left|\left| \int_X dy\, dz \langle P'(\phi(z))\rangle_{\bar l} C_{\bar l}(z,y) L^*_M(y,z)\right|\right|_{-1,\bar l}$$

$$\leqslant c_{43} \sum_Y \sum_{\beta_1\in Y\cap X} \sum_{\beta_2\in B^+} e(\beta_1,\beta_2) \| \chi_Y L^*_{M\beta_2,B^+}\|_{-1,\bar l}$$

$$\leqslant c_{43} \sum_Y \sum_{\beta_1\in Y\cap X} \sum_{\beta_2\in B^+} e(\beta_1,\beta_2) \| \chi_Y L^*_{M\beta_2,B^+}\|_{-1,\bar l}$$

This gives

$$2\| J_{X^c}\|_{B,\bar l} \| l\|_{B,\bar l}$$

$$\leqslant c_{44} \sum_{Y,\beta_i} e(\beta_1,\beta_2) \| J_{X^c}\|_{B,\bar l} \| \chi_Y L_{\beta_2,B^+}\|_{-1,l}$$

$$\leqslant c_{44} K \sum_{Y,\beta_1,\beta_2} e(\beta_1,\beta_2) \| \chi_Y L_{\beta_2,B^+}\|^2_{-1,l} \| J_{X^c}\|^2_{B,\bar l}$$

$$+ c_{45}|X|/K.$$

Similarly, by Corollary B.4 and (2.5) $\| l\|_{B,\bar l} \leqslant c_{46}|X|^{1/2}/K^{1/2}$ since we choose $\| L\| \leqslant \epsilon < 1/K^{1/2}$.

(b) $\bar J_X = J_X - l(x)$,

where

$$l(x) = \int dy\, \{L^*_M(x,y)\,\delta\langle\phi(y)\rangle + 2\langle\phi(y)\rangle_{\bar l} L(t)_X(x,y)\}$$

$$= \int dy\, \{ L^*_M(x,y) + 2L(t)_X(x,y)\}\, \delta\langle\phi(y)\rangle$$

$$+ 2\langle\phi(y)\rangle_{\bar l} L(t)_X(x,y)$$

so

$$\| \bar J_X\|_{-1,\bar l}$$

$$\leqslant \|J_X\|_{-1,\bar l} + \| l\|_{-1,\bar l}$$

$$\leqslant \| J_X\|_{-1,\bar l} + c_{26}|X|^{1/2}$$

$$\{\| C_{\bar l}^{1/2} \chi_X L(t) C_l^{1/2}\|_{L^2} + \| L(t)_X\|_{-1,\bar l}\}$$

by Corollary B.4 and (2.5).

Lemma III.4:

$$\left|\left\langle \prod_{i=1}^n \phi^{\bar l}(\chi_{X^c} f_{i,B})\, e^{\mathscr{T}^{\bar l}_{X^c} + \mathscr{L}^{\bar l}_{X^c}} \right\rangle_{\Lambda\setminus X,\bar l}\right| e^{c(J,L)} \frac{Z_{\Lambda\setminus X,\bar l}}{Z_{\Lambda,l}}$$

$$\leqslant n^{n/2} K'^n \left(1 + \frac{\| J_{X^c}\|_{-1,\bar l}}{n^{1/2}}\right)^n \prod_{i=1}^n \| \chi_{X^c} f_{i,B}\|_{-1,\bar l}$$

$$\times \exp\{c_{27}|X| + K(\| J_{X^c}\|^2_{1,B,\bar l} + \| J\|^2_{2,B^+,l})$$

$$+ \bar K(1 + \| J_{X^c}\|^2_{B,\bar l})(\| L_{X^c}\|^2_{1,B,\bar l} + \| L\|^2_{2,B^+,l})\},$$

where K' depends on K but not on $\bar K$.

Proof: By the inductive hypothesis (3.1)

$$\left|\left\langle e^{\mathscr{T}^{\bar l}_{X^c} + \mathscr{L}^{\bar l}_{X^c}}\right\rangle_{\Lambda\setminus X,\bar l}\right|$$

$$\leqslant \exp[K \| \bar J_{X^c}\|^2_{B,\bar l} + \bar K(1 + \| \bar J_{X^c}\|^2_{B,\bar l}) \| L_{X^c}\|^2_{B,\bar l}]$$

$$\leqslant \exp\{(K + \bar K \| L_{X^c}\|^2_{B,\bar l})\, [1 + c_{24} K(\| L\|^2_{2,B^+,\bar l}$$

$$- \| L_{X^c}\|^2_{2,B,\bar l})] \| J_{X^c}\|^2_{B,\bar l} + \bar K \| L_{X^c}\|^2_{B,\bar l} + 2c_{25}|X|\}$$

by Lemma III.3 and (2.6) since we choose ϵ such that $\epsilon^2\bar{K}<1(<K)$,

$$\leqslant \exp\{(K+\bar{K}\|L_{X^c}\|^2_{1,B,\bar{I}}+\bar{K}\|L\|^2_{2,B^+,\bar{I}})\|J_{X^c}\|^2_{B,\bar{I}}$$

$$+\bar{K}\|L_{X^c}\|^2_{B,\bar{I}}+2c_{25}|X|\}$$

since we choose $\bar{K}$ such that $(K+1)Kc_{24}<\bar{K}$. Hence by Vitali's theorem $\langle e^{\bar{\mathscr{J}}_{X^c}+\bar{\mathscr{L}}_{X^c}}\rangle$ is analytic in $\chi_{X^c}J_{(B)}$ and the lemma follows from Theorem A.1 [with $K'=e^3(K+2)e_3^2$ depending on K but not on $\bar{K}$], $Z_{A\setminus X,\bar{I}}/Z_{A,I}\leqslant e^{c_{47}|X|}$ and the estimate on $c(J,L)$ given in Lemma II.2. ■

Lemma III.5:

$$\sum_{k=0}^{\infty}a^k\left(1+\frac{b}{k^{1/2}}\right)^k\leqslant\frac{\sqrt{2}}{1-2a}e^{4a^2b^2}\quad\text{if } a<\tfrac{1}{2}.$$

Proof:

$$\sum_{k=0}^{\infty}a^k\left(1+\frac{b}{k^{1/2}}\right)^k=\sum_{k=0}^{\infty}\sum_{p=0}^{k}a^k\binom{k}{p}\left(\frac{b}{k^{1/2}}\right)^p$$

(binomial expansion)

$$\leqslant\sum_{k=0}^{\infty}\sum_{p=0}^{k}a^k2^k\frac{b^p}{k^{p/2}}$$

$$\left[\text{since } (1+1)^k\geqslant\binom{k}{p}1^p1^{k-p}\right]$$

$$\leqslant\sum_{k=0}^{\infty}\sum_{p=0}^{k}(2a)^k\frac{b^p}{(p!)^{1/2}}$$

$$=\sum_{n=0}^{\infty}\sum_{p=0}^{\infty}(2a)^n\frac{(2ab)^p}{(p!)^{1/2}},\quad\text{where } k=n+p,$$

$$=\frac{1}{1-2a}\sum_{p=0}^{\infty}\frac{(2ab)^p}{(p!)^{1/2}}2^{p/2}\frac{1}{2^{p/2}}$$

$$\leqslant\frac{1}{1-2a}\left(\sum_{p=0}^{\infty}\frac{(8a^2b^2)^p}{p!}\right)^{1/2}\left(\sum_{p=0}^{\infty}\frac{1}{2^p}\right)^{1/2}$$

$$=\frac{\sqrt{2}}{1-2a}e^{4a^2b^2}.$$ ■

Lemma III.6: $\bar{K}\|L_{X^c}\|^2_{1,B,\bar{I}}+A_4\|L(t)_M\|^2_{-1,\bar{I}}$
$\leqslant\bar{K}[\|L\|^2_{1,B,I}+t(\|L\|^2_{1,B^+,I}-\|L\|^2_{1,B,I})]$
if $\bar{K}\geqslant 2A_4e_2^4$.

Proof: $A_4\|L(t)_M\|^2_{-1,\bar{I}}+\bar{K}\|L_{X^c}\|^2_{1,B,\bar{I}}$

$$\leqslant 2A_4(t^2\|\delta L_M\|^2_{-1,\bar{I}}+\|L_{(B),M}\|^2_{-1,\bar{I}})+\bar{K}\|L_{X^c}\|^2_{1,B,\bar{I}}$$

$$\leqslant 2A_4e_2^4\left(t\sum_{\text{at least one }\beta_i=\alpha}\|L_{\beta_1,\beta_2}\|^2_{-1,\bar{I}}+\|L_M\|^2_{1,B,\bar{I}}\right)$$

$$+\bar{K}\|L_{X^c}\|^2_{1,B,\bar{I}}$$

(as in Lemma B.5c)

$$\leqslant\bar{K}\left(t\sum_{\text{at least one }\beta_i=\alpha}(e(\beta_1,\gamma)+e(\gamma,\beta_2))\|L_{\beta_1,\beta_2}\|^2_{-1,\bar{I}}\right)$$

$$+\bar{K}\|L\|^2_{1,B,\bar{I}}$$

(if $\bar{K}\geqslant 2A_4e_2^4$)

$$\leqslant\bar{K}\{\|L\|^2_{1,B,\bar{I}}+t[\|L\|^2_{1,B^+,I}-\|L\|^2_{1,B,I}]\}.$$

IV. EXISTENCE OF THE SECOND LEGENDRE TRANSFORM $\Gamma^{(2)}$

Our principal concern in this section is the inversion of the map

$$\binom{J}{L}\to\binom{A}{B}$$

defined in (1.9). We have

$$\mathscr{A}\{J,L\}\equiv\begin{bmatrix}A\{J,L\}(x)\\B\{J,L\}(y,z)\end{bmatrix}$$

$$=\begin{bmatrix}\langle\hat{\phi}(x)\rangle_{J,L}\\\langle:\phi(y)\phi(z):\rangle_{J,L}-\langle\hat{\phi}(y)\rangle_{J,L}\langle\hat{\phi}(z)\rangle_{J,L}\end{bmatrix},\tag{4.1}$$

where $\langle\cdot\rangle_{J,L}\equiv\langle\cdot e^{\mathscr{J}+\mathscr{L}}\rangle/\langle e^{\mathscr{J}+\mathscr{L}}\rangle$. The linear approximation to $\mathscr{A}$ near $J=L=0$ is (note that $\mathscr{A}\{0,0\}=[0,0]$)

$$\bar{\mathscr{A}}\{J,L\}=\begin{bmatrix}\langle\hat{\phi}(\hat{\mathscr{J}}+\hat{\mathscr{L}})\rangle\\\langle:\phi\phi:(\hat{\mathscr{J}}+\hat{\mathscr{L}})\rangle\end{bmatrix}.\tag{4.2}$$

It follows from the definition (1.4) of : : that the kernel of the integral operator mapping (J,L) to $\bar{\mathscr{A}}\{J,L\}$ is

$$\bar{\mathscr{A}}(x,y,z;x',y',z')$$

$$=\begin{bmatrix}\langle\phi(x);\phi(x')\rangle & \langle\phi(x);\phi(y')\phi(z')\rangle\\\langle\phi(y);\phi(z);\phi(x')\rangle & \langle\hat{\phi}(y)\hat{\phi}(z);\hat{\phi}(y')\hat{\phi}(z')\rangle\end{bmatrix}.\tag{4.3}$$

(Here the semicolons refer to connected expectations.) We will prove in Lemma IV.1 that (4.3) defines a bounded linear operator from $\mathscr{H}_{-1}\equiv\mathscr{H}_{-1}(\mathbb{R}^2)\oplus\mathscr{H}^s_{-1}(\mathbb{R}^4)$ onto $\mathscr{H}_{+1}=\mathscr{H}_1(\mathbb{R}^2)\oplus\mathscr{H}^s_1(\mathbb{R}^4)$ with bounded inverse and in Lemma IV.2 that $\mathscr{A}$ maps $\mathscr{H}_{-1,\delta}$ into $\mathscr{H}_{+1}$ and that $\mathscr{A}$ is analytic in J and L. In Theorem IV.3 we use the Contraction Mapping Theorem to prove the invertibility of $\mathscr{A}$ and the existence of $\Gamma^{(2)}$.

Lemma IV.1: In $\epsilon P(\phi)_2$, $\bar{\mathscr{A}}$ of (4.3) defines a bounded linear operator from $\mathscr{H}_{-1}$ to $\mathscr{H}_{+1}$ with a bounded inverse.

Proof: Since

$$\binom{J'}{L'}\cdot\bar{\mathscr{A}}\{J,L\}=\langle(\hat{\mathscr{J}}'+\hat{\mathscr{L}}')(\hat{\mathscr{J}}+\hat{\mathscr{L}})\rangle$$

it is a trivial consequence of Theorems III.1 and A.1 that $\bar{\mathscr{A}}$ is bounded as an operator from $\mathscr{H}_{-1}$ to $\mathscr{H}_{+1}$. When the coupling constant $\lambda = 0$,

$$\bar{\mathscr{A}}^{-1} = \begin{bmatrix} C^{-1} & 0 \\ 0 & 2C^{-1} \otimes C^{-1} \end{bmatrix}$$

is a bounded operator from $\mathscr{H}_{+1}$ to $\mathscr{H}_{-1}$. Hence, to prove the boundedness of $\bar{\mathscr{A}}^{-1}$ it suffices to prove that the matrix elements of $(d/d\lambda)\bar{\mathscr{A}}$, viewed as bilinear forms on $\mathscr{H}_{-1}$, are bounded uniformly in λ for λ sufficiently small. If $\{\zeta_\alpha\}$ is the partition of unity of Sec. II the cluster expansion implies

$$\left|\frac{d}{d\lambda}\langle \hat{\mathscr{L}}' \hat{\mathscr{L}}\rangle\right|$$

$$\leqslant \sum_{\alpha_i \beta_i \in Z} c_{48} \| L'_{\alpha_1 \alpha_2}\|_{-1} \times \{e^{-d(\alpha_1,\beta_1)-d(\alpha_2,\beta_2)} + e^{-d(\alpha_1,\beta_2)-d(\alpha_2,\beta_1)}\} \| L_{\beta_1,\beta_2}\|_{-1}$$

$$\leqslant c_{49} \| L'\|_{-1} \| L\|_{-1},$$

since $e^{-d(\alpha_1,\beta_1)-d(\alpha_2,\beta_2)} + e^{-d(\alpha_1,\beta_2)-d(\alpha_2,\beta_1)}$ defines a bounded operator on $l^2(\mathbf{Z}^4)$. The other matrix elements are handled similarly. ■

Lemma IV.2: In $\epsilon P(\phi)_2$, $\mathscr{A}\{J,L\}$ of (4.1) is defined on $\mathscr{H}_{-1,\delta} \equiv \mathscr{H}_{-1,\delta}(\mathbf{R}^2) \oplus \mathscr{H}^s_{-1,\delta}(\mathbf{R}^4)$ (for $\delta > 0$ sufficiently small) with range in $\mathscr{H}_{+1}$. $\mathscr{A}$ is analytic in J and L.

Proof: This is a simple consequence of Theorem III.1. The bounds

$$|\langle \hat{\mathscr{J}}_1 e^{\hat{\mathscr{J}}+\hat{\mathscr{L}}}\rangle| \leqslant \langle \hat{\mathscr{J}}_1 \bar{\hat{\mathscr{J}}}_1\rangle^{1/2} \langle e^{\hat{\mathscr{J}}+\hat{\mathscr{L}}+\bar{\hat{\mathscr{J}}}+\bar{\hat{\mathscr{L}}}}\rangle^{1/2} \leqslant c_{50} \| J_1\|_{-1} e^{K\|J\|^2_{-1} + \bar{K}\|L\|^2_{-1}}$$

and

$$|\langle \hat{\mathscr{L}}_1 e^{\hat{\mathscr{J}}+\hat{\mathscr{L}}}\rangle| \leqslant \langle \hat{\mathscr{L}}_1 \bar{\hat{\mathscr{L}}}_1\rangle^{1/2} \langle e^{\hat{\mathscr{J}}+\hat{\mathscr{L}}+\bar{\hat{\mathscr{J}}}+\bar{\hat{\mathscr{L}}}}\rangle^{1/2} \leqslant c_{50} \| L_1\|_{-1} e^{K\|J\|^2_{-1} + \bar{K}\|L\|^2_{-1}}$$

imply that $\langle \hat{\phi}(x) e^{\hat{\mathscr{J}}+\hat{\mathscr{L}}}\rangle$ is defined on $\mathscr{H}_{-1}(\mathbf{R}^2) \oplus \mathscr{H}^s_{-1,\delta}(\mathbf{R}^4)$ with image in $\mathscr{H}_{+1}(\mathbf{R}^2)$ and that $\langle :\phi(y)\phi(z): e^{\hat{\mathscr{J}}+\hat{\mathscr{L}}}\rangle$ is defined on $\mathscr{H}_{-1}(\mathbf{R}^2) \oplus \mathscr{H}^s_{-1,\delta}(\mathbf{R}^4)$ with image in $\mathscr{H}^s_{+1}(\mathbf{R}^4)$. Since $\langle e^{\hat{\mathscr{J}}+\hat{\mathscr{L}}}\rangle$ is continuous in J and L it is bounded away from zero on $\mathscr{H}_{-1,\delta}$. Hence $\mathscr{A}(J,L)$ is defined on $\mathscr{H}_{-1,\delta}$ with image a bounded subset of $\mathscr{H}_{+1}$. Hence analyticity in a suitably cutoff theory implies the analyticity of $\mathscr{A}$ by Vitali's theorem. ■

Theorem IV.3: In $\epsilon P(\phi)_2$, $\mathscr{A}$ has an inverse defined and analytic on $\mathscr{H}_{+1,\delta'}$ for some $\delta' > 0$. $\Gamma^{(2)}\{A,B\}$ is defined and analytic on $\mathscr{H}_{+1,\delta'}$.

Proof: This a direct application of Theorem A.4 with

$$V_1 = \mathscr{H}_{-1},$$
$$V_2 = \mathscr{H}_{+1},$$
$$\mathscr{F} = \mathscr{A}\{J,L\},$$
$$F = \bar{\mathscr{A}}\{J,L\}.$$
■

Properties of G, $\mathscr{A}$, and $\Gamma^{(2)}$ established in Ref. 1 in the the sense of formal power series can now be simply established in the sense of genuine functionals. For example, it is no problem to introduce Dirichlet boundary conditions and prove preservation of decoupling for quadratic sources [see Eq. (4.3) of the first reference in Ref. 1 and Lemma II.11 of the second].

Theorem IV.4: In an $\epsilon P(\phi)_2$ theory with zero Dirichlet data on ∂X (where X is a union of lattice squares in $\mathbf{R}^2$) $\mathscr{A}$ maps $V_{S,\delta}$ into W_S and $\mathscr{A}^{-1}$ maps $W_{S,\delta'}$ into V_S for some $\delta, \delta' > 0$ where

$$V_S = \{(J,L) \in H_{-1,\partial X}(\mathbf{R}^2) \oplus H^s_{-1,\partial X}(\mathbf{R}^4) | \text{ supt } L \subset \bar{X} \times \bar{X} \cup \bar{X}^c \times \bar{X}^c\},$$
$$W_S = \{(A,B) \in H_{+1,\partial X}(\mathbf{R}^2) \oplus H^s_{+1,\partial X}(\mathbf{R}^4) | \text{ supt } B \subset \bar{X} \times \bar{X} \cup \bar{X}^c \times \bar{X}^c\}.$$

(Here supt means support in the sense of distributions.)

Proof: This a direct application of Theorem A.5 with

$$V_M = \{(J,L) \in H_{-1,\partial X}(\mathbf{R}^2) \oplus H^s_{-1,\partial X}(\mathbf{R}^4) | J = 0, \text{ supt } L \subset \bar{X} \times \bar{X}^c \cup \bar{X}^c \times \bar{X}\},$$
$$W_M = \{(A,B) \in H_{+1,\partial X}(\mathbf{R}^2) \oplus H^s_{+1,\partial X}(\mathbf{R}^4) | A = 0, \text{ supt } B \subset \bar{X} \times \bar{X}^c \cup \bar{X}^c \times \bar{X}\}.$$ ■

APPENDIX A: ANALYTICITY BOUNDS

We gather together here those properties of analytic functions on Banach spaces that we use. The basic tool is the Cauchy integral formula, which is valid in a Banach space setting. (See Ref. 11).

Theorem A.1: Let $\mathscr{F}$ be a complex valued function on a vector space V and let $\|\cdot\|$ be a seminorm on V. If

(a) $\mathscr{F}$ is analytic in the sense that $\mathscr{F}(f + \zeta g)$ is analytic in ζ for any $f, g \in V$ and

(b) $|\mathscr{F}(f)| \leqslant e^{K\|f\|^2}$ for all $f \in V$, then

$$\left|\prod_{i=1}^{n} \frac{d}{d\alpha_i} \mathscr{F}\left(f + \sum_{i=1}^{n} \alpha_i f_i\right)\Big|_{\alpha_i=0}\right| \leqslant e^{3n}(K^{1/2} n^{1/2} + K\|f\|)^n e^{K\|f\|^2} \prod_{i=1}^{n} \|f_i\|.$$

Proof: Integrating α_i over the circle $|\alpha_i| = \rho_i$ (ρ_i to be determined later) we have

$$\left|\prod_{i=1}^{n} \frac{d}{d\alpha_i} \mathscr{F}\left(f + \sum_{i=1}^{n} \alpha_i f_i\right)\Big|_{\alpha_i=0}\right|$$
$$= \frac{1}{(2\pi)^n}\left|\int_{|\alpha_i|=\rho_i} \prod_{i=1}^{n} \frac{d\alpha_i}{\alpha_i^2} \mathscr{F}\left(f + \sum_{i=1}^{n} \alpha_i f_i\right)\right|$$
$$\leqslant \left[\prod_{i=1}^{n} \rho_i\right]^{-1} \sup_{|\alpha_i|=\rho_i} \exp\left(K\|f + \sum_{i=1}^{n} \alpha_i f_i\|^2\right)$$
$$\leqslant \left(\prod_{i=1}^{n} \|f_i\|\right) \rho^{-n} \exp[K(\|f\|^2 + 2n\rho\|f\| + n^2\rho^2)] \quad \left(\text{choosing } \rho_i = \frac{\rho}{\|f_i\|}\right)$$
$$\leqslant \left(\prod_{i=1}^{n} \|f_i\|\right) [(Kn)^{1/2} + K\|f\|]^n e^{K\|f\|^2 + 3n} \quad (\text{choosing } \rho = [(Kn)^{1/2} + K\|f\|]^{-1}).$$
■

Theorem A.2: Suppose V_i are Banach spaces with norms $\| \ \|_i$ and $V_{i,\delta} = \{v \in V_i | \|v\|_i < \delta\}$. Suppose (a) $\mathscr{F}$: $V_{1,\delta} \to V_2$ is analytic in the sense that $\mathscr{F}(v + zw)$ is analytic in z for all $v + zw \in V_{1,\delta}$.

(b) $\mathscr{F}$ is uniformly bounded on $V_{1,\delta}$ with $\|\mathscr{F}\| \equiv \sup_{v \in V_{1,\delta}} \|\mathscr{F}(v)\|_2 < \infty$. Let

$$\mathscr{F}_n(v) = \mathscr{F}(v) - \sum_{j=0}^{n-1} \frac{1}{j!} \frac{d^j}{dz^j} \mathscr{F}(zv)\Big|_{z=0} \quad \text{for } n = 1,2,\cdots.$$

Then there exists a constant c_δ such that

(a) $\| \mathscr{F}_n(v)\|_2 \leqslant c_\delta^n \| \mathscr{F}\| \, \| v\|_1^n$, for all $v \in V_{1,\delta/3}$,

(b) $\| \mathscr{F}(v) - \mathscr{F}(w)\|_2 \leqslant c_\delta \| \mathscr{F}\| \, \| v - w\|_1$, for all $v,w \in V_{1,\delta/9}$,

(c) $\| \mathscr{F}_n(v) - \mathscr{F}_n(w)\|_2 \leqslant c_\delta^n \| \mathscr{F}\|(\|v\|_1 + \|w\|_1)^{n-1} \|v - w\|_1$, for all $v,w \in V_{1,\delta/9}$ and $n = 1,2,\cdots$.

Proof: For any r with $0 < r < \delta/\| v\|_1$

$$\mathscr{F}(v) = -\frac{1}{2\pi i}\int_{|\xi| = r} \frac{\mathscr{F}(\xi v)}{\xi - 1} d\xi ,$$

$$\frac{d^j}{dz^j}\mathscr{F}(zv)|_{z=0} = \frac{j!}{2\pi i}\int_{|\xi| = r} \frac{\mathscr{F}(\xi v)}{\xi^{j+1}} d\xi ,$$

$$\mathscr{F}_n(v) = \frac{1}{2\pi i}\int_{|\xi| = r} \mathscr{F}(\xi v)\left[\frac{1}{\xi - 1} - \sum_{j=0}^{n-1}\frac{1}{\xi^{j+1}}\right] d\xi$$

$$= \frac{1}{2\pi i}\int_{|\xi| = r} \frac{\mathscr{F}(\xi v)}{\xi^n(\xi - 1)} d\xi .$$

Choosing $r = \frac{2}{3}\delta/\| v\|_1$, we have (since $\| v_1\| \leqslant \delta/3$)

$$\| \mathscr{F}_n(v)\|_2 \leqslant \frac{1}{2\pi} 2\pi r \frac{\| \mathscr{F}\|}{r^n(r - 1)} = \frac{1}{1 - 1/r}\frac{\| \mathscr{F}\|}{r^n}$$

$$\leqslant 2\left(\frac{3}{2\delta}\right)^n \| \mathscr{F}\| \, \|v\|_1^n$$

establishing (a). Similarly, part (b) follows from the representation

$$\mathscr{F}(v) - \mathscr{F}(w) = \int_0^1 d\alpha \frac{d}{d\alpha}\mathscr{F}(w + \alpha(v - w))$$

$$= \frac{1}{2\pi i}\int_0^1 d\alpha \int_{|\xi| = r} d\xi \frac{\mathscr{F}(w + \xi(v - w))}{(\xi - \alpha)^2}$$

with $r = \frac{1}{2}\delta/\|v - w\|_1$, and part (c) follows from part (b) and the representation

$$\mathscr{F}_n(v) - \mathscr{F}_n(w) = \frac{1}{2\pi i}\int_{|\xi| = r} \frac{\mathscr{F}(\xi v) - \mathscr{F}(\xi w)}{\xi^n(\xi - 1)} d\xi$$

with $r = \frac{1}{8}\delta/\max(\|v\|_1, \|w\|_1)$. ■

Remark: It follows from Theorem A.2 that if $\mathscr{F}$ satisfies the hypotheses of that theorem then the linear approximation defined by

$$Fv = \frac{d}{dz}\mathscr{F}(zv)|_{z=0} \tag{A1}$$

is a bounded linear operator from V_1 to V_2.

Corollary A.3: Under the hypotheses of Theorem A.2, $\mathscr{F}$ is also analytic in the sense of composition, i.e., if $b(z)$: $D \to V_{1,\delta}$ is analytic (where D is an open subset of C) then $\mathscr{F}(b(z))$: $D \to V_2$ is analytic.

Proof: Fix any $z_0 \in D$.

$$\frac{1}{\Delta z}\{ \mathscr{F}(b(z_0 + \Delta z)) - \mathscr{F}(b(z_0))\}$$

$$= \frac{1}{\Delta z}\{ \mathscr{F}(b(z_0) + \Delta z b'(z_0)) - \mathscr{F}(b(z_0))\}$$

$$+ \frac{1}{\Delta z}\{ \mathscr{F}(b(z_0 + \Delta z)) - \mathscr{F}(b(z_0) + \Delta z b'(z_0))\} .$$

The first term converges as $\Delta z \to 0$ by hypothesis so it suffices to prove that the second converges to zero. However, by Theorem A.2(b) the second term is bounded (for Δz small enough) by

$$O(1)\frac{1}{|\Delta z|} \| b(z_0 + \Delta z) - b(z_0) - \Delta z b'(z_0)\|_1 .$$

However, $b(z_0 + \Delta z) - b(z_0) - \Delta z b'(z_0) = \mathscr{G}_2(\Delta z)$ for $\mathscr{G}(z) = b(z_0 + z)$ so that $\|b(z_0 + \Delta) + b(z_0) - \Delta z b'(z_0)\|_1 \leqslant O(1)|\Delta z|^2$ by Theorem A.2(c).

Theorem A.4: Let $\mathscr{F}$, V_1, V_2 satisfy the hypotheses of Theorem A.2 and define F by (A1). If F has a bounded inverse from V_2 to V_1 then $\mathscr{F}$ has an inverse defined on $\mathscr{F}(0) + V_{2,\epsilon}$ for some $\epsilon > 0$ and this inverse is analytic and uniformly bounded.

Proof: We may without loss of generality consider the case $\mathscr{F}(0) = 0$. To reformulate this problem as a standard Contraction Mapping Theorem problem we write

$$w = \mathscr{F}(v) = Fv + \mathscr{F}_2(v)$$

and then

$$v = F^{-1}w - F^{-1}\mathscr{F}_2(v) \equiv \mathscr{G}_w(v). \tag{A2}$$

We wish to show that, for a suitable choice of δ' and ϵ, Eq. (A2) has, for each $w \in W_2$, a unique solution $v \in \overline{V}_{1,\delta'}$. We first note that $\mathscr{G}_w$ maps $\overline{V}_{1,\delta'}$ into itself if $\epsilon \leqslant \delta'/2\| F^{-1}\|$ and

$$\delta' \leqslant \min\left\{\frac{\delta}{4}, \frac{1}{2c_\delta^2\| \mathscr{F}\| \, \| F^{-1}\|}\right\},$$

since then

$$\| \mathscr{G}_w(v)\|_1 \leqslant \| F^{-1}\| \, [\|w\|_2 + c_\delta^2\| \mathscr{F}\| \, \|v\|_1^2] \leqslant \delta'$$

by Theorem A.2. To verify that $\mathscr{G}_w(v)$ is a strict contraction we observe that for any $v,v' \in \overline{V}_{1,\delta'}$

$$\| \mathscr{G}_w(v) - \mathscr{G}_w(v')\|_2$$
$$= \| F^{-1}(\mathscr{F}_2(v) - \mathscr{F}_2(v'))\|_2$$
$$\leqslant \| F^{-1}\| \, c_\delta^2\| \mathscr{F}\|(\|v\|_1 + \|v'\|_1)\|v - v'\|_1$$
$$\leqslant \tfrac{1}{2}\|v - v'\|_1 ,$$

provided we choose

$$\delta' \leqslant \min\left\{\frac{\delta}{10}, \frac{1}{4c_\delta^2\| \mathscr{F}\| \, F^{-1}\|}\right\}.$$

The Contraction Mapping Theorem now implies that the sequence of functions

$$I^{(0)}(w) = F^{-1}w,$$

$$I^{(n)}(w) = \mathscr{G}_w(I^{(n-1)}(w))$$

converges (uniformly for $w \in V_{2,\epsilon}$) as n tends to ∞ and that the limit is an inverse for $\mathscr{F}$. Since $I^{(0)}$ and $\mathscr{G}_w$ are both analytic (in the sense of Corollary A.3) so are all the $I^{(n)}$'s and hence, by the Weierstrass Limit Theorem, so is the limit. Since the range of $\mathscr{G}_w$ is contained in $\overline{V}_{1,\delta'}$, the inverse is clearly uniformly bounded. ■

The final theorem of this section determines conditions which ensure that if a mapping between Hilbert spaces maps one subspace into a second subspace then the inverse maps the second back into the first. We apply this theorem to the question of decoupling quadratic sources in Theorem IV.4. This decoupling is crucial for the analysis of irreducibility properties. See Refs. 1 and 3.

Theorem A.5: Let V_S, V_M, W_S, and W_M be Hilbert spaces. If

(a) $\mathscr{F}: (V_S \oplus V_M)_\delta \to W_S \oplus W_M$ is analytic,

(b) $\mathscr{F}$ is uniformly bounded on $(V_S \oplus V_M)_\delta$ and F defined by (A1) has a bounded inverse $F^{-1}: W_S \oplus W_M \to V_S \oplus V_M$,

(c) $\mathscr{F}: V_{S,\delta} \to W_S$ and $F \restriction V_S$ has a bounded inverse from W_S to V_S, then the inverse function $\mathscr{F}^{-1}$ maps $\mathscr{F}(0) + W_{S,\epsilon}$ into V_S for some $\epsilon > 0$.

Proof: Choose δ' sufficiently small that $F^{-1}\mathscr{F}_2 \restriction (V_S \oplus V_M)_{\delta'}$ is a strict contraction (see Theorem A.4). Then by Theorem A.4 $\mathscr{G} = \mathscr{F} \restriction V_{S,\delta'}$ has an inverse on $\mathscr{F}(0) + W_{S,\epsilon_1}$ with range in V_S and $\mathscr{F} \restriction (V_S \oplus V_M)_{\delta'}$ has an inverse on $\mathscr{F}(0) + (W_S \oplus W_M)_{\epsilon_2}$. Choose $\epsilon = \min(\epsilon_1, \epsilon_2)$. Then $\mathscr{G}^{-1} \restriction (\mathscr{F}(0) + W_{S,\epsilon}) = \mathscr{F}^{-1} \restriction (\mathscr{F}(0) + W_{S,\epsilon})$ since $\mathscr{F}$ is one-to-one on $(V_S \oplus V_M)_{\delta'}$. $[\mathscr{F}(v) = \mathscr{F}(0) + F(v + F^{-1}\mathscr{F}_2(v)$ with $F^{-1}\mathscr{F}_2$ a strict contraction.] Hence $\mathscr{F}^{-1}$ maps $\mathscr{F}(0) + W_{S,\epsilon}$ into V_S. ■

APPENDIX B: ESTIMATES INVOLVING DIRICHLET B.C.

In this Appendix we collect together various facts and estimates about Dirichlet B.C. that we have used in the main part of the paper. As in Secs. II and III we consider the Laplacian Δ_l with Dirichlet B.C. on a finite set l of lattice bonds and we let $C_l = (-\Delta_l + m_0^2)^{-1}$ be the corresponding covariance and $\|f\|_{\pm 1,l} = \|(-\Delta_l + 1)^{\pm 1/2} f\|_{L^2}$. By definition (see Ref. 12), $\langle f, C_l^{-1} g\rangle$ is the form closure of the form

$$\langle f, C^{-1} g\rangle = \int [(\overline{\nabla f})(\nabla g) + m_0^2 \bar{f} g]\, dx$$

restricted to $C_0^\infty(\mathbb{R}^2 \setminus l)$. We state this as

Lemma B.1: If f and g belong to the form domain $Q(C_l^{-1})$ then $f, g \in Q(C^{-1})$ and $\langle f, C_l^{-1} g\rangle = \langle f, C^{-1} g\rangle$.

Let $\partial^\gamma C$ be a multiple difference covariance arising in the cluster expansion. [See (8.1) of Ref. 6; here γ is a finite set of bonds.] $\partial^\gamma C$ can be written as a convex combination of C_Γ's. If all of these Γ's contain l, we say that $\partial^\gamma C$ has Dirichlet B.C. on l. In Lemma 2.3 of Ref. 5 Glimm and Jaffe established the operator bound

$$\| C^{-1/2} \partial^\gamma C\, C^{-1/2}\|_{L^2(\Delta_\alpha) \to L^2(\Delta_\beta)} \leqslant m_0^{c_{10} - |\gamma|/4} K_6(\gamma)\, e^{-(1/2) m_0 d(\alpha,\beta,\gamma)}, \quad \text{(B1)}$$

where Δ_α and Δ_β are two unit squares, $d(\alpha,\beta,\gamma) = \sup_{b\in\gamma} \{d(\Delta_\alpha, b) + d(b, \Delta_\beta)\}$, and $K_6(\gamma)$ is the usual CE quantity defined in Proposition 8.1 of Ref. 6. We now extend this bound to the case of Dirichlet B.C.:

Lemma B.2: If $\partial^\gamma C$ has Dirichlet B.C. on l, then

$$\| C_l^{1/2} \partial^\gamma C\, C_l^{1/2}\|_{L^2(\Delta_\alpha) \to L^2(\Delta_\beta)} \leqslant m_0^{c_{10} - |\gamma|/4} K_6(\gamma)\, e^{-(1/4) m_0 d(\alpha,\beta,\gamma)}.$$

Proof: Since $C_l^{-1/2} \partial^\gamma C\, C_l^{-1/2}$ is a bounded operator, $\partial^\gamma C\, C_l^{-1/2}$ maps L^2 into $Q(C_l^{-1})$ and so by Lemma B.1

$$\langle f, C_l^{-1/2} \partial^\gamma C\, C_l^{-1/2} g\rangle = \langle C_l^{1/2} f, C_l^{-1} \partial^\gamma C\, C_l^{-1/2} g\rangle = \langle C_l^{1/2} f, C^{-1} \partial^\gamma C\, C_l^{-1/2} g\rangle .$$

Similarly $(C^{-1/2} \partial^\gamma C)^*$ maps L^2 into $Q(C_l^{-1})$ so that

$$\langle f, C_l^{-1/2} \partial^\gamma C\, C_l^{-1/2} g\rangle = \langle C^{-1/2} C_l^{1/2} f, (C^{-1/2} \partial^\gamma C\, C^{-1/2})(C^{-1/2} C_l^{1/2}) g\rangle . \quad \text{(B2)}$$

The lemma now follows from (B1), (B2), and the bound

$$\| C^{-1/2} C_l^{1/2}\|_{L^2(\Delta_\alpha) \to L^2(\Delta_\beta)} \leqslant c_{51}\, e^{-3 m_0 |\alpha - \beta|/4} . \quad \text{(B3)}$$

To establish (B3) it suffices to consider α and β far apart since $C^{-1/2} C_l^{1/2}$ is a bounded operator on L_l^2. Now for $x \neq y$

$$\begin{aligned}(C^{-1/2} C_l^{1/2})(x,y) &= (C^{1/2}(-\Delta + m_0^2)\, C_l^{1/2})(x,y) \\ &= m_0^2\, (C^{1/2} C_l^{1/2})(x,y) - \left(\frac{\partial}{\partial x_i} C^{1/2}\right)\left(\frac{\partial}{\partial x_i} C_l^{1/2}\right)(x,y) .\end{aligned}$$

(B3) now follows from the representation

$$C_l^{1/2} = \int_0^\infty (-\Delta_l + m_0^2 + \zeta)^{-1} \zeta^{-1/2} d\zeta \Big/ \int_0^\infty (1+\zeta)^{-1} \zeta^{-1/2} d\zeta$$

and kernel estimates such as (Ref. 13 Lemma VI.8)

$$\left\| \frac{\partial}{\partial x_i} (-\Delta_l + m_0^2 + \zeta)^{-1} \right\|_{L^2(\Delta_\alpha \times \Delta_{\alpha'})} \leqslant c_{52}\, e^{-(1/2)(m_0^2 + \zeta)^{1/2} |\alpha - \alpha'|}$$

■

(valid when Δ_α and $\Delta_{\alpha'}$ are not adjacent.)

Corollary B.3: If supt $f_i \subset \Delta_i$ and $f_i \in D(C_l^{1/2})$

$$\langle f_1, C_l f_2\rangle \leqslant c_{53}\, e^{-(1/2) d(\Delta_1, \Delta_2)} \|f_1\|_{-1,l} \|f_2\|_{-1,l} .$$

Proof: It suffices to consider $f_i \in C_0^\infty(\Delta_i)$ with $d(\Delta_1, \Delta_2) \geqslant 1$. Then

$$\begin{aligned}\langle f_1, C_l f_2\rangle &= \langle f_1, \{ C_l - C_{l\cup\partial\Delta_1} - C_{l\cup\partial\Delta_2} + C_{l\cup\partial\Delta_1\cup\partial\Delta_2}\} f_2\rangle \\ &= \langle C_l^{1/2} f_1, C_l^{-1/2} \partial^2 C\, C_l^{-1/2} C_l^{1/2} f_2\rangle ,\end{aligned}$$

where $\partial^2 C \equiv C_l - C_{l\cup\partial\Delta_1} - C_{l\cup\partial\Delta_2} + C_{l\cup\partial\Delta_1\cup\partial\Delta_2}$. In the path integral representation of $\partial^2 C$ the integral is over paths that avoid l but pass through both $\partial\Delta_1$ and $\partial\Delta_2$. Hence as in Lemma B.2 $C_l^{-1/2} \partial^2 C\, C_l^{-1/2}$ is a bounded operator from L^2 to L^2 with norm at most $c_{53}\, e^{-(1/2) d(\partial\Delta_1, \partial\Delta_2)}$. ■

Corollary B.4: In the notation of Sec. II (in particular $\bar{l} = l\cup\partial X$)

(a) $\|\langle \phi(x)\rangle_l - \langle \phi(x)\rangle_{\bar{l}}\|_{+1,l} \leqslant c_{38} |X|^{1/2}$;

(b) $\| \chi_Y(x)(\langle \phi(x); \phi(y)\rangle_l) - C_l(x,y))\|_{+1,l} \leqslant c_{38} |Y|^{1/2}$

if Y is a finite union of lattice squares with $\partial Y \subset l$;

(c) $\|\langle \phi(x); \phi(y)\rangle_l - C_l(x,y) - \langle \phi(x); \phi(y)\rangle_{\bar{l}} + C_{\bar{l}}(x,y)\|_{+1,l} \leqslant c_{38} |X|^{1/2}$.

Proof: Let $C(\sigma) = \sigma C_l + (1-\sigma) C_{\bar{l}}$ and $\partial C \equiv (d/d\sigma) C(\sigma) = C_l - C_{\bar{l}}$. Let $\langle \cdot \rangle_\sigma$ denote the corresponding interpolating expectation [with the Wick ordering of the interaction matched to $C(\sigma)$].

(a) As in Lemma II.4

$$\langle\phi(x)\rangle_l - \langle\phi(x)\rangle_{\bar{l}}$$
$$= \int_0^1 d\sigma \frac{d}{d\sigma} \langle\phi(x)\rangle_\sigma$$
$$= \int_0^1 d\sigma \frac{d}{d\sigma} C(\sigma)(x,y)\langle V'(y)\rangle_\sigma$$

(repeated arguments are integrated over)

$$= \int_0^1 d\sigma\, \partial C(x,y)\langle V'(y)\rangle_\sigma + C(\sigma)(x,y)\langle V''(y)\,\partial C(y,z)\,V'(z)\rangle_\sigma + C(\sigma)(x,y)\langle V'(y); V'(z)\,\partial C(z,z')\,V'(z')\rangle_\sigma .$$

Now

$$\left|\left| \int \partial C(\cdot,y)\,\langle V'(y)\rangle_\sigma\,dy \right|\right|^2_{+1,l}$$
$$= \langle V'\rangle_\sigma\, \partial C\, C_l^{-1}\, \partial C\, \langle V'\rangle_\sigma$$
$$\leqslant \sum_{\Delta_1,\Delta_2,\Delta_3} \| \langle V'(\cdot)\rangle_\sigma \|_{L^2(\Delta_1)} \times \| \partial C\, C_l^{-1/2} \|_{L^2(\Delta_1)\to L^2(\Delta_2)} \| C_l^{-1/2}\,\partial C \|_{L^2(\Delta_2)\to L^2(\Delta_3)} \times \| \langle V'(\cdot)\rangle_\sigma \|_{L^2(\Delta_3)}$$
$$\leqslant c_{20} \sum_{\Delta_1,\Delta_2,\Delta_3} e^{-d(\Delta_1,\Delta_2)}\, e^{-d(\Delta_2,\Delta_3)}\, e^{-d(\Delta_3,\partial X)}$$
$$\leqslant c_{54} \sum_{\Delta_3,\Delta\subset X} e^{-d(\Delta_3,\Delta)} \leqslant c_{55} \sum_{\Delta\subset X} 1 = c_{55}|X| .$$

The remaining terms are estimated similarly. In the last term observe that the cluster expansion implies a decay in min$(d(y,z), d(y,z'))$.

(b) The crucial point here is that the singular term in the perturbation theory expansion of $\langle\phi(x);\phi(y)\rangle_l$ [namely $C_l(x,y)$, which is not locally in H^{+1}] is cancelled explicitly: By integration by parts

$$\langle\phi(x);\phi(y)\rangle_l - C_l(x,y)$$
$$= C_l(x,z)\langle V''(z)\rangle_l\, C_l(z,y) + C_l(x,z)\langle V'(z); V'(z')\rangle_l\, C_l(z',y) ,$$

which obeys the desired bound by standard arguments.

(c) The proof is similar to that of (a) and (b) and is based on the formula

$$\langle\phi(x);\phi(y)\rangle_l - C_l(x,y) - \langle\phi(x);\phi(y)\rangle_{\bar{l}} + C_{\bar{l}}(x,y)$$
$$= \int_0^1 d\sigma \frac{d}{d\sigma} \{\langle\phi(x);\phi(y)\rangle_\sigma - C(\sigma)\}$$
$$= \int_0^1 d\sigma \frac{d}{d\sigma} \{C(\sigma)(x,z)\langle V''(z)\rangle_\sigma\, C(\sigma)(z,y) + C(\sigma)(x,z)\,\langle V'(z); V'(z')\rangle_\sigma\, C(\sigma)(z',y)\} .$$

■

As in Sec. II we let $\{\zeta_\beta \mid \beta\in\mathbb{Z}^2\}$ be a C_0^∞ partition of unity invariant under lattice translations, $J_\beta = \zeta_\beta J$, $J_B = \Sigma_{\beta\in B} J_\beta$, $e(\alpha,\beta) = e_1\, e^{-|\alpha-\beta|}$; we define $\| J \|_{1,B,l}$, $\| J \|_{2,B,l}$, and $\| J \|_{B,l}$ by Definitions II.3, II.5, and (2.10), respectively.

Lemma B.5: Let $|J|_{B,l}$ by any one of $\| J \|_{1,\beta,l}$, $\| J \|_{2,B,l}$, or $\| J \|_{B,l}$. Then

(a) $|\ \ |_{B,l}$ is a seminorm,

(b) $|J_B|_{B,l} \leqslant |J_{B'}|_{B',l}$ if $B\subset B'$, $|J_B|_{B,l} \leqslant |J_B|_{B,l'}$ if $l\supset l'$,

(c) $\| J_B \|_{-1,l} \leqslant e_2 |J_B|_{B,l}$ for some universal constant $e_2 \geqslant 1$,

(d) $|J_B|_{B,l} \leqslant e_3 \| J \|_{-1,l}$ for some universal constant e_3.

Proof: (a) is obvious.

(b) The first inequality is obvious since the sum defining the right-hand side contains the sum defining the left-hand side. The second inequality is a consequence of

$$\| J_\beta \|_{-1,l} \leqslant \| J_\beta \|_{-1,l'}$$

and the decoupling property (2.8).

(c) $\| J_B \|^2_{-1,l}$

$$= \Big\| \sum_{\beta\in B} J_\beta \Big\|^2_{-1,l} = \sum_{\beta,\beta'\in B} \langle J_\beta, C_l J_{\beta'}\rangle$$

(with m_0 replaced by 1)

$$\leqslant c_{56} \sum_{\beta,\beta'\in B} \| J_\beta \|_{-1,l}\, e^{-(1/2)d(\beta,\beta')}\, \| J_{\beta'} \|_{-1,l} \quad \text{(Corollary B.3)}$$
$$\leqslant c_{57} \sum_{\beta\in B} \| J_\beta \|^2_{-1,l}$$

[since $e^{-(1/2)d(\beta,\beta')}$ is a bounded operation on $l^2(\mathbb{Z}^2)$]

$$\leqslant e_2^2\, |J_B|^2_{B,l} .$$

(d) The inequality $|J_B|^2_{B,l} \leqslant c_{58}\, \Sigma_\beta \| J_\beta \|^2_{-1,l}$ follows directly from the definitions. We now claim that (see Ref. 5)

$$\sum_\beta \| J_\beta \|^2_{-1,l} \leqslant c_{59} \| J \|^2_{-1,l} .$$

We offer a more complete proof than that given for the case $l = \phi$ in Ref. 5.

$$\sum_\beta \| J_\beta \|^2_{-1,l}$$
$$= \sum_\beta \langle \zeta_\beta J,\, C_l\, \zeta_\beta J\rangle \quad \text{(with } m_0 \text{ replaced by 1)}$$
$$= \sum_\beta \langle C_l^{1/2} J,\, C_l^{-1/2} \zeta_\beta\, C_l\, \zeta_\beta\, C_l^{-1/2} C_l^{1/2} J\rangle$$
$$\leqslant \sum_{\substack{\Delta,\Delta',\Delta''\\ \beta}} \| \chi_\Delta C_l^{1/2} J \|_{L^2}\, \| (\chi_\Delta C_l^{-1/2} \zeta_\beta C_l^{1/2} \chi_{\Delta'}) \times (\chi_{\Delta'} C_l^{1/2} \zeta_\beta C_l^{-1/2} \chi_{\Delta''}) \|_{\text{op}} \times \| \chi_{\Delta''} C_l^{1/2} J \|_{L^2} .$$

It suffices to show

$$\| \chi_\Delta C_l^{-1/2} \zeta_\beta C_l^{1/2} \chi_{\Delta'} \|_{\text{op}} \leqslant c_{60}\, e^{-d(\Delta,\Delta')} (e^{-d(\Delta,\beta)} + e^{-d(\beta,\Delta')}) .$$

But

$$\chi_\Delta C_l^{-1/2} \zeta_\beta C_l^{1/2} \chi_{\Delta'} = \chi_\Delta C_l^{1/2} C_l^{-1} \zeta_\beta C_l^{1/2} \chi_{\Delta'}$$
$$= \chi_\Delta C_l^{1/2} (-\Delta + 1) \zeta_\beta C_l^{1/2} \chi_{\Delta'}$$

(by Lemma B.1)

$$= \chi_\Delta C_l^{1/2} \zeta_\beta C_l^{1/2} \chi_{\Delta'}$$

$$+ \sum_{i=0}^{1} \chi_\Delta \, C_l^{1/2} \frac{\partial}{\partial y_i} \zeta_B \frac{\partial}{\partial y_i} C_l^{1/2} \chi_{\Delta'}$$

$$+ \sum_{i=0}^{1} \chi_\Delta \, C_l^{1/2} \left[\zeta_B, \frac{\partial}{\partial y_i} \right] \frac{\partial}{\partial y_i} C_l^{1/2} \chi_{\Delta'} .$$

The desired bound now follows from standard estimates on $C_l^{1/2}$ and $(\partial/\partial y_i)\, C_l^{1/2}(x, y)$ (for x and y separated). (See Lemma B.2.) ■

The same results hold for the analogous seminorms on $L(x, y)$:

Lemma B.6: Let $| \, L \, |_{B,l}$ by any one of $\| L \|_{1,B,l}$, $\| L \|_{2,B,l}$, or $\| L \|_{B,l}$ as defined in the first paragraph of Theorem III.1. Then

(a) $| \quad |_{B,l}$ is a seminorm,

(b) $|L_B|_{B,l} \leqslant |L_{B'}|_{B',l}$ if $B \subset B'$,
$|L_B|_{B,l} \leqslant |L_B|_{B,l'}$ if $l \supset l'$,

(c) $\| L_B \|_{-1,l} \leqslant e_2^2 \, |L_B|_{B,l}$,

(d) $|L_B|_{B,l} \leqslant e_3 \| L_B \|_{-1,l}$.

[1] A. Cooper, J. Feldman, and L. Rosen, "Legendre Transforms and r-Particle Irreducibility in Quantum Field Theory: The Formalism for $r = 1,2$," Ann. Phys. (NY) **137**, 146 (1981); "The Formal Power Series Framework," *ibid.* **137**, 213 (1981).

[2] A. Cooper, J. Feldman, and L. Rosen, "Higher Legendre Transforms and their Relationship to Bethe-Salpeter Kernels and r-field Projectors," J. Math. Phys. **23**, 846 (1982).

[3] A. Cooper, J. Feldman, and L. Rosen, "Cluster Irreducibility of the Third and Fourth Legendre Transforms in Quantum Field Theory," Phys. Rev. D **25**, 1565 (1982).

[4] A. Cooper, J. Feldman, and L. Rosen, "Legendre Transforms and r-Particle Irreducibility in Quantum Field Theory: The Formalism for Fermions for $r = 1,2$" (in preparation).

[5] J. Glimm and A. Jaffe, "The Entropy Principle for Vertex Functions in Quantum Field Models," Ann. Inst. Henri Poincaré A **22**, 97 (1975).

[6] J. Glimm, A. Jaffe, and T. Spencer, *Constructive Quantum Field Theory, Lecture Notes in Physics*, Vol. 25, edited by G. Velo and A. Wightman (Springer, Berlin, 1973).

[7] J. Fröhlich, "Schwinger Functions and their Generating Functionals. I," Helv. Phys. Acta **47**, 265 (1974).

[8] O. McBryan, Contribution to International Colloquium on Mathematical Methods of Quantum Field Theory, Marseille (1975).

[9] E. Seiler and B. Simon, "Nelson's Symmetry and All That in the Yukawa$_2$ and $(\phi^4)_3$ Field Theories," Ann. Phys. (NY) **97**, 470 (1976).

[10] C. Burnap, "The Second Legendre Transform of the $P(\phi)_2$ Generating Functional," preprint.

[11] M. Berger, *Nonlinearity and Functional Analysis* (Academic, New York, 1977).

[12] M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. II, Fourier Analysis, Self-Adjointness (Academic, New York, 1974).

[13] A. Cooper and L. Rosen, "The Weakly Coupled Yukawa$_2$ Field Theory: Cluster Expansion and Wightman Axioms," Trans. Am. Math. Soc. **234**, 1 (1977).

Nonperturbative confinement in quantum chromodynamics. II. Mandelstam's gluon propagator

D. Atkinson, P. W. Johnson,[a)] and K. Stam
Institute for Theoretical Physics, University of Groningen, P. O. Box 800, 9700 AV Groningen, The Netherlands

(Received 15 January 1982; accepted for publication 5 March 1982)

It is shown that Mandelstam's approximate equation for the gluon propagator has a solution with very singular infrared behavior. At the origin in the squared momentum variable there are a double pole, a branch-point, and an accumulation of complex first-sheet branch-points. Although the double pole is suggestive of confinement, the existence of acausal complex singularities indicates a deficiency in this first step of the approximation scheme.

PACS numbers: 11.10.Np

1. INTRODUCTION

This paper is an extension of a previous study[1] of nonperturbative confinement in quarkless quantum chromodynamics, to which we shall refer as I. We continue to explore the hypotheses that (1) it is an indication of confinement for the gluon propagator to be more singular than k^{-3} at small k^2, where k is the gluon four-momentum, and (2) its infrared singularity structure can be properly understood in truncated Dyson–Schwinger (DS) equations. In I we considered a truncated DS equation for the gluon propagator proposed by Mandelstam.[2] Mandelstam worked in the Landau gauge, ignored four-gluon coupling altogether, and moreover he replaced the three-gluon vertex and one of the two internal gluon propagators by bare values. He asserted that the propagator from such a truncated system would behave as k^{-4} at small k^2. We analyzed a somewhat simplified version of Mandelstam's equation and demonstrated (1) that the gluon propagator did have that infrared behavior, and (2) that it also acquired branch-points at complex k^2 in the vicinity of the origin. In fact, such complex branch-points are inconsistent with causality, and causality was used to justify Wick rotation of the internal momentum variable in the truncated DS equation.

It was not clear from I whether the occurrence of unphysical branch-points in the simplified Mandelstam equation was an artifact of additional, somewhat unmotivated assumptions, or whether the full Mandelstam equation [Eq. (2.1) below] would have similar behavior. Here it is shown that solutions of the full Mandelstam equation (however without ghost propagators) have both features of the approximate equation. Namely, the gluon propagator behaves as k^{-4} at asymptotically small k^2, except near the negative real axis, along which complex branch-points seem to accumulate.

Mandelstam justified replacement of the three-gluon vertex function, $\Gamma(p,q,r)$ with $p+q+r=0$, and one gluon propagator, $\Delta(q)$, by their bare values through the Slavnov–Taylor identity for the longitudinal part of the triple-gluon vertex. However, that identity does not require the longitudinal part of Γ to vanish as q and r separately approach zero,[3] so that the cancellation described by Mandelstam is incomplete. Since exact cancellation of propagator and vertex function does not follow from basic principles, the equation obtained by Mandelstam might be expected to be somewhat unphysical.

In contrast to the situation in quantum electrodynamics, the vacuum polarization tensor in quantum chromodynamics is a gauge-dependent entity. Consequently, the behavior of the gluon propagator at small k^2 does not provide direct evidence of confinement. Indeed, a second-order pole in the gluon propagator can be removed by a singular gauge transformation. Our expectation is that the gauge transformation, while removing the pole, will preserve the general feature that propagation of low-frequency modes of the gluon field is suppressed, as is indicative of confinement.

An alternative treatment of DS esquations in QCD has been proposed and examined by Baker *et al.*,[4] and further simplified by Schoenmaker.[5] In this work, an axial gauge is used, so that ghost fields are uncoupled, and may thus be neglected. The basic idea is an *ansatz* for the longitudinal part of the three-gluon vertex, in terms of the full propagator, such that the vertex Slavnov–Taylor identity is satisfied. Within this framework, it is possible to project out the four-gluon terms, so that a closed equation for the propagator results. This has a more complicated nonlinear structure than that of Mandelstam's equation; but there is some reason to hope that the approximation of Baker *et al.* is better than that of Mandelstam.

Baker *et al.* demonstrate that a double pole is a consistent infrared *ansatz*; and they obtain an approximate numerical solution at all energies. However, this work by no means demonstrates that a solution actually exists, much less that it has the required infrared behavior. The point is not merely academic, for Delbourgo has shown that his elegant spectral *ansatz* yields a *nonconfining* infrared behavior,[6] a result that has been confirmed by Khelashvili.[7] Delbourgo also used an axial gauge, and the spectral *ansatz* for the three-gluon vertex is motivated by means of the Slavnov–Taylor (ST) identity. Since it is not expected that a transverse

[a)] Permanent address: Illinois Institute of Technology, Chicago, Illinois 60616.

part of the gluon vertex dominates the infrared, the conflicting claims regarding the behavior of the two approximations, which have the same longitudinal part (in the sense that they are both consistent with the vertex ST identity), is suspect. A careful mathematical treatment of both equations is required, and we hope to provide this in the future.

In Sec. 2 we describe a consistent regularization procedure for Mandelstam's equation [Eq. (2.1) below]. It is reduced to a nonlinear integral equation suitable for analysis [Eq. (2.16)]. The existence of a solution of (2.16), which is analytic in k^2 in a heart-shaped region not including the negative real axis, is established in Sec. 3. A numerical solution for the gluon propagator and procedures for stable analytic continuation are described in Sec. 4. In particular, the existence of unphysical complex branch-points is established, and they are located with precision. The numerical work includes an expansion of the gluon propagator at small spacelike momenta, which is described in Sec. 4 and shown in the Appendix to be an asymptotic expansion.

2. MANDELSTAM'S GLUON EQUATION

In I, we sketched Mandelstam's derivation of an integral equation for the unknown function, $F_1(x)$. Now Eq. (2.9) of I, with the pole term removed, can be rewritten

$$\frac{x}{A+xF_1(x)} = 1 - C + Dx + g^2 \times \int_0^x dy\left\{25\left(1-\frac{y^2}{x^2}\right)-\frac{7}{2}\left(\frac{x}{y}-\frac{y^3}{x^3}\right)\right\}\frac{F_1(y)}{y}, \tag{2.1}$$

where g is proportional to the SU(3) coupling constant, and where

$$C = 25g^2\int_0^\infty \frac{dy}{y}F_1(y), \tag{2.2}$$

and

$$D = \frac{7}{2}g^2\int_0^\infty \frac{dy}{y^2}F_1(y). \tag{2.3}$$

In I, the further approximation was made of dropping the $\frac{7}{2}$ terms above, and it was possible then to prove the existence of a solution, $F_1(x)$, that behaves like x as $x\to 0$ (except along the negative real direction). In this paper we improve the treatment by retaining all the above terms.

The first constraint is that, for consistency, C must be equal to unity; but the integral in (2.2) is ultraviolet divergent, and we may regard $C=1$ as part of the renormalization prescription, as we did in I. The *ansatz* $F_1(x)\sim x$ as $x\to 0$ is no longer consistent, because of the $\frac{7}{2}$ terms, and must be replaced by $F_1(x)\sim x^\alpha$, $\alpha>1$. However, the left-hand side of (2.1) still goes linearly to zero, and this imposes the constraint $D=1/A$. In fact, having removed $1-C$, we can also scale A and g away by the transformations

$$x\to Agx,\quad y\to Agy,\quad F_1(x)\to g^{-1}F_1(x), \tag{2.4}$$

so that

$$G(x) = -\frac{1}{x^3}\int_0^x dy\left\{25\left(1-\frac{y^2}{x^2}\right)-\frac{7}{2}\left(\frac{x}{y}-\frac{y^3}{x^3}\right)\right\}\frac{F_1(y)}{y}, \tag{2.5}$$

where

$$G(x) = \frac{F_1(x)}{x+x^2F_1(x)}, \tag{2.6}$$

is a new unknown function, as in Sec. 3 of I. To this equation must be added the global constraint, corresponding to $D=1/A$, viz.,

$$\frac{7}{2}\int_0^\infty \frac{dy}{y^2}F_1(y) = 1. \tag{2.7}$$

It is remarkable that this integral will turn out to be ultraviolet and infrared convergent. This is a constraint that was missing in the more approximate equation of I; but we shall find that it can be met.

For infrared convergence in (2.5), we would like

$$F_1(x)\sim\gamma x^{\beta-1}, \tag{2.8}$$

as $x\to 0$, with $\beta>2$. Then $G(x)$, on the left-hand side of (2.5), behaves like $x^{\beta-2}$, while the right-hand side behaves in general like $x^{\beta-4}+O(x^{\beta-2})$. This is inconsistent unless the coefficient of $x^{\beta-4}$ vanishes; fortunately this happens if

$$\beta = (31/6)^{1/2}\approx 2.273\ldots, \tag{2.9}$$

a result found by Mandelstam. The value of the coefficient γ in (2.8) can only be obtained numerically, with the help of the global condition (2.7), as we shall see in Sec. 4.

The integral equation (2.5) can be reduced to the nonlinear differential equation

$$6x^2F_1'' + 18xF_1' - 25F_1 = -\frac{1}{6x}[x^5(x^3G)'']'', \tag{2.10}$$

with

$$F_1(x) = \frac{xG(x)}{1-x^2G(x)}. \tag{2.11}$$

The independent solutions of the homogeneous equation (the left-hand side equal to zero), are $x^{-1\pm\beta}$; so (2.10) can be resolved in terms of them by the method of variation of parameters. The result is

$$F_1(x) = \gamma x^{\beta-1} - \frac{1}{72\beta x}\times\int_0^x \frac{dy}{y}\left[\left(\frac{x}{y}\right)^\beta-\left(\frac{y}{x}\right)^\beta\right][y^5(y^3G(y))'']'', \tag{2.12}$$

where the correct boundary condition (2.8) is assured by the first term. The differentiations under the integral in (2.12) can be removed by four partial integrations, and we find

$$x^4G'' + 9x^3G' + (36+\tfrac{63}{4}x^2)G = \Sigma, \tag{2.13}$$

where

$$\Sigma(x) = 36\gamma x^{\beta-2} - \frac{36x^2G^2(x)}{1-x^2G(x)} - \frac{5}{12}x^2G(x) - \frac{175}{72\beta x^2}\int_0^x dy\left[\left(\frac{x}{y}\right)^\beta-\left(\frac{y}{x}\right)^\beta\right]y^3G(y). \tag{2.14}$$

Here F_1 has been eliminated in favor of G, by means of (2.11); this gives rise to the nonlinear term in (2.14). The left-hand side of (2.13) comes from the boundary terms in the partial integrations, except that part of the term proportional to x^2G has been transferred to the right-hand side [namely the term

$-\frac{5}{12}x^2G(x)$ in (2.14)]. The reason for this transposition is as in I, namely that (2.13) can now be resolved in terms of elementary functions, and the linear term in (2.14) will cause no trouble for small x, thanks to the factor x^2.

The homogeneous equation (2.13) (i.e., with the right-hand side equal to zero) is solved by the functions

$$x^{-7/2}\exp[\pm 6i/x]; \tag{2.15}$$

so (2.13) can be resolved by variation of parameters, the result being

$$G(x) = -\tfrac{1}{6}x^{-7/2}\int_0^x dy\, y^{3/2}\sin\left(\frac{6}{x}-\frac{6}{y}\right)\Sigma(y). \tag{2.16}$$

No homogeneous terms may be added. In the next section, we will show that a locally unique solution of (2.16) exists, that is analytic in a certain domain of the x plane, much as in I.

3. EXISTENCE PROOF

To show that Eq. (2.16) has a solution, it is convenient to make these transformations of variables:

$$\xi = \frac{6}{x};\ \tilde{G}(\xi) = G(x);\ \zeta = \frac{6}{y} - \frac{6}{x}. \tag{3.1}$$

Equation (2.16) takes the form

$$\tilde{G}(\xi) = P(\tilde{G},\xi) \equiv f(\xi) - \xi^{7/2}\int_0^\infty d\zeta \frac{\sin\zeta}{(\xi+\zeta)^{11/2}}\tilde{\Omega}(\xi+\zeta), \tag{3.2}$$

where

$$f(\xi) = \xi^{7/2}6^{\beta-4}\gamma\int_0^\infty d\zeta\frac{\sin\zeta}{(\xi+\zeta)^{\beta+3/2}}, \tag{3.3}$$

and

$$\tilde{\Omega}(\zeta) = -\frac{5}{12}\tilde{G}(\zeta) - \frac{36\tilde{G}^2(\zeta)}{1-36\zeta^{-2}\tilde{G}(\zeta)} + \frac{175}{72\beta}\zeta^4\int_\zeta^\infty\frac{d\kappa}{\kappa^5}\left[\left(\frac{\zeta}{\kappa}\right)^\beta - \left(\frac{\kappa}{\zeta}\right)^\beta\right]\tilde{G}(\kappa). \tag{3.4}$$

We shall establish that (3.2) has a solution $\tilde{G}(\xi)$ which is analytic in ξ in the domain $\mathscr{D}$, where

$$\mathscr{D}(\rho,\delta) = \left\{\begin{matrix}|\xi| > \rho^{-1}, & \operatorname{Re}\xi \geqslant 0\\ \dfrac{|\operatorname{Im}\xi| - \rho^{-1}}{|\operatorname{Re}\xi|} > \tan\delta, & \operatorname{Re}\xi < 0\end{matrix}\right\}. \tag{3.5}$$

The positive parameters ρ and δ are to be fixed later. The domain $\mathscr{D}$ is the same as that considered in I in connection with proof of existence of a solution of an equation very similar to (3.2). The analysis here is quite parallel to that presented in I.

Let $\mathscr{B}$ be the Banach space of functions analytic in $\mathscr{D}$, with the supremum norm

$$\|f\| = \sup_{\xi\in\mathscr{D}}|f(\xi)|. \tag{3.6}$$

Define the ball $\mathscr{S}$ in the Banach space $\mathscr{B}$ by

$$\mathscr{S} = \{\tilde{G}\,|\,\tilde{G}\in\mathscr{B} \text{ and } \|\tilde{G}\| \leqslant b\}. \tag{3.7}$$

The domain $\mathscr{D}$ has the feature that if ξ lies in $\mathscr{D}$, then so does $\xi+\zeta$, for $\zeta \geqslant 0$. Furthermore, if the constraint

$$36\rho^2 b < 1 \tag{3.8}$$

is satisfied, the function $\tilde{\Omega}(\xi+\zeta)$ is analytic in ζ throughout $\mathscr{D}$ when $\zeta \geqslant 0$, and the integral in (3.2) converges uniformly to a function analytic in $\mathscr{D}$. The inhomogeneous term in (3.2), $f(\xi)$, can be shown by an analysis similar to that of Appendix B of I to be analytic in the ξ plane cut along the negative real axis. Consequently, $P(\tilde{G},\xi)$ is analytic for ξ in $\mathscr{D}$ if (3.8) is met.

We shall show that P maps the ball into itself and is a contraction mapping, if suitable constraints are placed upon ρ, δ, and b. The Banach contraction mapping theorem may then be applied to give a solution of the equation

$$\tilde{G}(\xi) = P(\tilde{G},\xi), \tag{3.9}$$

which is unique in the ball $\mathscr{S}$ of $\mathscr{B}$.

By using condition (3.8), one obtains the following bound upon $\Omega(\zeta)$ for $\zeta\in\mathscr{D}$:

$$|\Omega(\zeta)| \leqslant \left(\frac{5}{12} + \frac{175}{36(16-\beta^2)}\right)b + \frac{36b^2}{1-36\rho^2 b} \equiv J(\rho,b). \tag{3.10}$$

One may obtain the following bound directly from (3.3):

$$|f(\xi)| \leqslant \frac{\gamma}{36}(6\rho)^{\beta-2}\left\{1 + \left(\beta+\frac{3}{2}\right)\int_0^\infty\frac{d\omega}{|\omega - e^{i\epsilon}|^{\beta+5/2}}\right\} \equiv C_\epsilon, \tag{3.11}$$

where $|\arg\xi| \leqslant \pi - \epsilon$. Using (3.10) and (3.11) in (3.2), we obtain

$$|P(\tilde{G},\xi)| \leqslant C_\epsilon + \rho J(b,\rho)D_\epsilon, \tag{3.12}$$

where

$$D_\epsilon = \int_0^\infty\frac{d\omega}{|\omega - e^{i\epsilon}|^{11/2}}. \tag{3.13}$$

Consequently, the ball $\mathscr{S}$ is mapped into itself by P if

$$C_\epsilon < b, \tag{3.14}$$

and

$$\rho \leqslant \frac{b - C_\epsilon}{J(b,\rho)D_\epsilon}. \tag{3.15}$$

The contractivity condition is

$$\|P(G_1) - P(G_2)\| \leqslant K\,\|G_1 - G_2\|, \tag{3.16}$$

with K less than 1, for any functions G_1 and G_2 in the ball $\mathscr{S}$.

To obtain an estimate on the difference of the nonlinear terms in P, it is convenient to introduce

$$\Sigma(\tilde{G},\zeta) = \frac{36\tilde{G}^2(\zeta)}{1-(36/\zeta^2)\tilde{G}(\zeta)}. \tag{3.17}$$

The derivative of this algebraic function with respect to $\tilde{G}$ is well-defined, and for $\tilde{G}$ in $\mathscr{S}$ and ζ in $\mathscr{D}$ it is subject to the bound

$$\left|\frac{d\Sigma}{d\tilde{G}}\right| \leqslant \frac{108b}{(1-36b\rho^2)^2} \equiv L(b,\rho). \tag{3.18}$$

One may then use this constraint, along with the mean value theorem, to obtain

$$|\Sigma(G_1,\zeta) - \Sigma(G_2,\zeta)| \leqslant L(b,\rho)|G_1(\zeta) - G_2(\zeta)|. \tag{3.19}$$

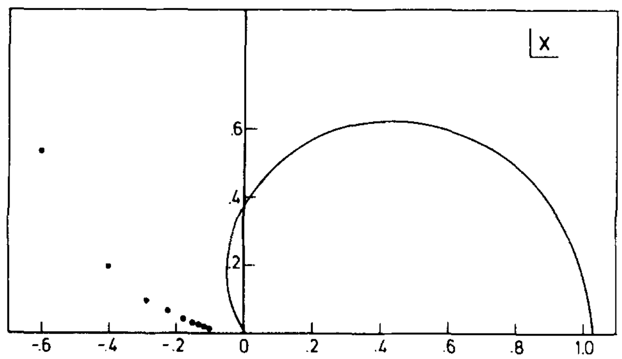


FIG. 1. The cardioid region to which the Banach theorem applies. The points outside this region are the locations of branch-points determined as described in Sec. 4.

Using (3.19) and making more direct estimates of the other terms in P, one obtains an estimate of the form (3.16), with

$$K=\rho\left[L(b,\rho)+\frac{5}{12}+\frac{175}{36(16-\beta^2)}\right]D_\epsilon. \tag{3.20}$$

Consequently, the mapping is contractive if

$$K<1. \tag{3.21}$$

The conditions for a contraction mapping, (3.8), (3.15), and (3.21), may simultaneously be met for any number ϵ between 0 and $\pi/2$. Because the integrals C_ϵ and D_ϵ depend upon ϵ, the maximal values of the parameters ρ, δ, and b also depend upon it. The function $\tilde{G}(\xi)$, obtained as the locally unique fixed point of Eq. (3.2) in each of the domains $\mathscr{D}(\rho,\epsilon)$, is analytic in ξ in the union of these domains. We have extended this fixed-point proof to a set of domains in the right-half x plane, which are sectors of varying radius and opening angle that are symmetric about the real axis. The full domain of analyticity in the variable x, which is obtained numerically as the union of the regions in which conditions (3.8), (3.15), and (3.21) are met, is shown in Fig. 1. The parameter γ is chosen so that condition (2.7) is met [see Eq. (4.10) below].

4. NUMERICAL ANALYSIS

We have shown that the integral equation (2.16) has a solution $G(x)$ which is bounded and analytic in the heart-shaped domain $\mathscr{D}$, with the asymptote $\gamma x^{\beta-2}$ as x approaches zero within $\mathscr{D}$. We wish to obtain this solution numerically, and thereby determine the behavior of $G(x)$ outside the domain $\mathscr{D}$. Equation (2.16) is a well-behaved functional equation for G—at least so long as x is in $\mathscr{D}$—but it seems impractical to attempt a direct global solution of (2.16). Instead, we have chosen to determine $G(x)$ in some domain of small x from (2.16), and then to get G elsewhere by solving a differential equation such as (2.10), which is equivalent to (2.16).

We obtain $G(x)$ at small x within $\mathscr{D}$ by developing an asymptotic series for x in that region. Although we justify the asymptotic series by analysis of Eq. (2.16), the series itself is most easily developed from the integro-differential system (2.13) and (2.14). One may make a consistent expansion in powers of both x^2 and x^β as follows:

$$G(x)=\sum_{j=1}^{\infty}\sum_{n=1}^{\infty}a_{jn}x^{j\beta+2n-4}. \tag{4.1}$$

The coefficient of the leading term is $a_{11}=\gamma$, and the higher-order terms may be determined recursively from these formulas:

$$a_{j,1}=-\gamma a_{j-1,1}\quad \text{for } j>1; \tag{4.2}$$

$$a_{1,n}=-c_{1,n-1}\quad \text{for } n>1; \tag{4.3}$$

$$a_{j,n}=-\gamma a_{j-1,n}-c_{j,n-1}+\sum_{j'=1}^{j-1}\sum_{n'=1}^{n-1}a_{j'n'}c_{j-j',n-n'}\quad \text{for } j,n>1. \tag{4.4}$$

We have used the array c_{jn}:

$$c_{jn}=\frac{a_{jn}}{36}\left[(2n+j\beta)^2+\frac{1}{6}+\frac{175}{36}\frac{1}{(2n+j\beta)^2-\beta^2}\right]. \tag{4.5}$$

We show in the Appendix that (4.1) is indeed an asymptotic series for $G(x)$ by truncating it to include only powers of x not greater than M. Our estimate for the difference between $G(x)$ and the truncated series depends upon M, as well as the location of the point x (its amplitude *and* phase) in the domain $\mathscr{D}$. In practice, for a given x_0 we truncate the series (4.1) so that the computed values of $G(x_0)$ and its first four derivatives give least discrepancy in the fourth-order differential equation (2.10). We can achieve single-precision accuracy for $G(x)$ (order 10^{-12}) on the CDC Cyber 160/170 computer in Groningen at small x in $\mathscr{D}$ with M of order 20; for $\gamma=0.0608$ we can use the series on the real x axis out to about 0.13, and less far in complex directions. The values of G and its first three derivatives are used as a starting point for solution of (2.10).

Let us consider the solution of the fourth-order nonlinear differential equation (2.10) from starting values of G and its first three derivatives at a point $x_0\neq 0$. If the values are such that $x_0^2G(x_0)\neq 1$, the fourth derivative can be determined from (2.10). Furthermore, from the general theory of differential equations involving analytic functions of both the dependent and independent variables,[8,9] one expects there to be a locally unique solution $G(x)$ corresponding to these initial data, which is analytic in x in some neighborhood of x_0. Of course, the solutions that develop from different initial data bear no simple relation to one another, because of the nonlinearity in G. The singularities of a solution of (2.10) may be of two types: (1) "fixed singularities" at $x=0$ and $x=\infty$, and (2) "movable singularities" at points for which

$$x^2G(x)=1. \tag{4.6}$$

The point $x=0$ is an irregular singular point of the differential equation, and one expects $G(x)$ to have an essential singularity at that point, with possibly nontrivial Riemann sheet structure as well. The locations of the movable singularities depend upon the initial data. There is no simple prescription to determine the locations of these movable singularities from the initial data; in general one must resort to numerical analysis.

It is consistent with the integro-differential system (2.13) and (2.14), and therefore with (2.10), for $G(x)$ to have the following asymptotic form near a branch point at $x = d$, at which (4.6) is satisfied:

$$G(x) \sim \frac{1}{d^2} \pm \frac{6}{2^{1/2} d^4}(x-d)\left[\ln\frac{(x-d)}{d_0}\right]^{1/2}, \qquad (4.7)$$

where d_0 is a constant. With this asymptotic form, for which $G'(x)$ diverges logarithmically as x approaches d, the most singular terms in the system cancel near $x = d$. This divergence of G' and the higher derivatives in the vicinity of the branch points makes it difficult to locate them numerically by direct solution of (4.6).

The solution of (2.16) described in Sec. 3 is one of an infinite number of solutions of the differential equation (2.10). Furthermore, we expect from the general theory of analytic differential equations that it is the only solution of (2.10) with the asymptote $\gamma x^{\beta-2}$ at small positive x, so that all other solutions are so singular as to be inconsistent with the original integral equation (2.5) in that region. In the fixed-point proof for existence of a solution $G(x)$, analytic in $\mathscr{D}$, it was important to ensure that condition (4.6) was not met anywhere in $\mathscr{D}$, so that the movable singularities are avoided in that domain.

We shall construct the function $G(x)$ and effect its analytic continuation outside $\mathscr{D}$ by numerical means. One would hope for physical reasons that $G(x)$, being related to the full gluon propagator in Mandelstam's truncation of Dyson–Schwinger equations in quantum chromodynamics, would turn out to be analytic on the physical sheet of the cut x plane, with a branch-cut lying only along the negative real x axis, and bounded at infinity in that plane. However, we have no analytical control over the behavior of G outside $\mathscr{D}$, and must resort to numerical procedures to determine its analytic structure. The real constant γ must be chosen so that the integral condition (2.7) is met by $F_1(x,\gamma,G)$. Strictly speaking, since G is not guaranteed by our analysis to have a continuation to the full positive real axis, the integral (2.7) need not even exist. Our procedure for choosing γ requires numerical work for its justification.

With initial data obtained from the asymptotic series (4.1), the differential equation (2.10) is integrated from a starting point x_0 by an explicit fourth-order Runge–Kutta routine, in which it is considered as four coupled first-order differential equations for G, G', G'', and G'''. For a discussion of this standard procedure, see Refs. 10 and 11. The step length Δx is changed with changing x to maintain accuracy. In particular, it is necessary to take rather small steps when x is small, or when $x^2 G(x)$ is close to $+1$. When one is near $x = 0$, or near a movable singularity, or both, instabilities are apt to creep in. There may be no immediate suggestion of inaccuracy, since cumulative errors are equivalent to changes in the values of G and its first three derivatives at the starting point. We have tested the integration routine to be certain that the values of $G(x)$ are indeed path-independent and stable away from the fixed and movable singularities.

The integral (2.7) is computed over small x, $0<x<0.1$, by using the asymptotic series (4.1). For $x>0.1$ we determine the integral

$$I(x) = \frac{7}{2}\int_0^x \frac{dy}{y}\,\frac{G(y)}{1-y^2G(y)}, \qquad (4.8)$$

by solving the equivalent differential equation

$$\frac{dI}{dx} = \frac{7}{2x}\,\frac{G(x)}{1-x^2G(x)}. \qquad (4.9)$$

The differential equation (4.9) for $I(x)$ is incorporated in the Runge–Kutta integration procedure to determine $G(x)$. The constraint (2.7), $I(\infty) = 1$, is satisfied by choosing the parameter γ to be

$$\gamma = 0.060\,870\,966\,1 \pm 0.000\,000\,000\,1. \qquad (4.10)$$

As an independent check of the accuracy of this result, we have verified that the ratio of the change in $I(\infty)$ to the change in γ,

$$\Delta I(\infty)/\Delta\gamma \approx 20.17, \qquad (4.11)$$

is numerically stable down to $\Delta\gamma$ of 10^{-10}. It is important for the asymptotic series to give an accurate representations of $I(x)$ at small x, since more than 40% of the integral comes from x below 0.1.

With the choice (4.10) for γ, the function $x^2G(x)$ is analytic in the right half x plane, approaches $+1$ at infinity in the right half-plane, and is monotonically increasing in x for real positive x. The behavior of the corresponding function $F_1(x)$ is shown in Fig. 2. This function has the following asymptote at large real x:

$$F_1(x) = \frac{1}{[50\ln(x/x_0)]^{1/2}}, \qquad (4.12)$$

as required for consistency with (2.5).

For exploring the behavior of $G(x)$ in the left half x plane, expecially at small x, it is quite useful to be able to integrate the differential equation (2.10) along implicitly defined contours that are determined as we go along. For example, to keep the magnitude of $G(x)$ constant to first order in step size Δx, one must require

$$\Delta\,[G^*(x)G(x)] = O(\Delta x)^2 \qquad (4.13)$$

or

$$\mathrm{Re}[G^*(x)G'(x)\Delta x] = 0. \qquad (4.14)$$

At each step of the Runge–Kutta routine we choose the *phase* of Δx so that (4.14) is met, using values of G and G' at the current position. Actually, it is advantageous to keep

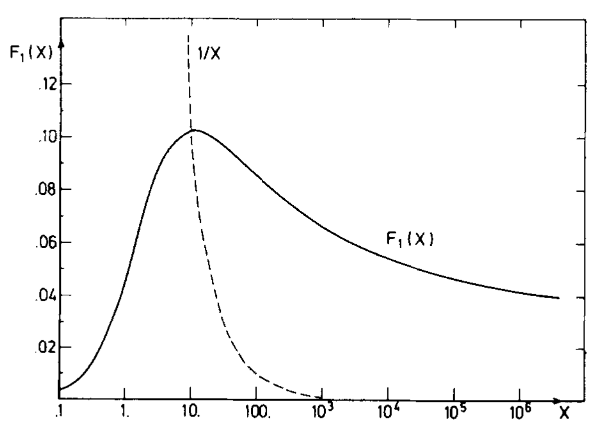


FIG. 2. Graph of $F_1(x)$ versus x. The other term in (4.21), $1/x$, is shown for reference.

$|G(x)|$ constant only to first order in step size, since the slight, gradual changes in G provide a good monitor on the level of accuracy with which the function $G(x)$ is being determined. It is even more useful to integrate along a contour on which $|x^2G(x)|$ is held roughly constant, to keep a safe distance away from the movable singularities of (2.10). The corresponding condition on the phase of Δx is

$$\mathrm{Re}\left\{G^*(x)\left[G'(x)+\frac{2}{x}G(x)\right]\Delta x\right\}=0. \qquad (4.15)$$

With thorough analysis and testing, we have made a stable extrapolation of $G(x)$ into the left half x plane. We find that, when γ is given by (4.10), there are branch-points at locations given in Table 1.

It is consistent to suppose that there is an infinite number of branch-points on the physical sheet, accumulating at $x=0$ near the negative real axis, but such a hypothesis cannot be tested numerically. Of course, it is reasonable to expect that $x^2G(x)$ takes on the value $+1$ at an infinite number of points near the essential singularity at $x=0$, but we have found no general argument to indicate that such points must lie on the physical Riemann sheet. We have no information on the asymptotic behavior of $G(x)$ as x approaches zero, except when x is in $\mathscr{D}$.

Since it is essentially a numerical problem to prove the existence of branch-points of $G(x)$ and to locate them, it is appropriate to give the following information concerning the accuracy with which G is determined:

1). At $x_0=(-0.5,0.75)$, the function $G(x)$ is reliably determined to be (0.324 361 862 88, 0.142 874 479 38), with the error in the last digit.

2). When Eq. (2.10) is started from x_0 and integrated counterclockwise around a square contour with sides -0.25, and $0.25\,i$, respectively, the total change in the real and imaginary parts of G is less than 10^{-11}.

3). By contrast, when Eq. (2.10) is started from x_0 and integrated counterclockwise around a square contour of sides -0.25 and $-0.25\,i$, respectively, the new value of G is $(+6.145\,867\,784\,1, -0.386\,126\,067\,6)$, with the error in the last digit.

4). The results in 2). and 3). are valid for 1000, 2000, and 4000 steps per side in the Runge–Kutta integration.

This information is our basis for concluding that a branch-point lies inside the second square, but not in the first; see Table I.

TABLE I. Location of first nine branch-points of $G(x)$.

n	Rex	Imx
1	− 0.601 22	± 0.535 25
2	− 0.403 17	± 0.191 20
3	− 0.289 55	± 0.098 45
4	− 0.224 28	± 0.060 43
5	− 0.182 57	± 0.041 08
6	− 0.153 76	± 0.029 86
7	− 0.132 71	± 0.022 74
8	− 0.116 71	± 0.017 94
9	− 0.104 11	± 0.014 54

It is a nontrivial numerical problem to maintain accuracy while getting close enough to branch-points to be able to find and isolate them, especially at small x, where the branch-points themselves are close together and other singularities are nearby. We have found it rather efficient to integrate (2.10) along a curve for which $|x^2G(x)|$ is fixed at a value somewhat less than 1. The phase of $x^2G(x)$ changes continuously along such a curve, and one is fairly close to a branch-point whenever $x^2G(x)$ becomes real and positive. The branch-points are located more precisely by integrating along closed paths enclosing successively smaller regions. The branch-points can be determined quite accurately by using steps determined by solving (4.6) through Newton iteration. Even though $G'(x)$ diverges logarithmically at the branch-point, according to (4.7), the method works rather well.

A direct numerical solution of (2.10) is subject to criticism on the grounds that is has solutions which are very singular at small x, but reasonably well-behaved elsewhere, and cumulative errors will, in effect, switch us over to one of the unacceptable solutions as we change x. We avoid this problem to a great extent by starting at small x in $\mathscr{D}$ using the asymptotic series (4.1), thereby assuring that at the outset there is very little contamination of the solution. Correspondingly, we expect a substantial loss in precision when we attempt to integrate from large to small $|x|$.

An alternate procedure is to solve the integro-differential equations (2.13) and (2.14). We can write them as a coupled system of equations for $G(x)$, $\Omega_1(x)$, and $\Omega_2(x)$; Ω_1 and Ω_2 being defined as

$$\Omega_1(x)=\frac{1}{x^2}\int_0^x dy\,y^3\left(\frac{y}{x}\right)^\beta G(y), \qquad (4.16)$$

$$\Omega_2(x)=\frac{1}{x^2}\int_0^x dy\,y^3\left(\frac{x}{y}\right)^\beta G(y). \qquad (4.17)$$

The coupled system is

$$\begin{aligned}G''(x)=&-\frac{9}{x}G'(x)-\frac{97}{6x^2}G(x)\\&+\frac{1}{x^4}\left[\frac{175}{72\beta}(\Omega_1(x)-\Omega_2(x))\right.\\&\left.+36\left(\gamma x^{\beta-2}-\frac{G(x)}{1-x^2G(x)}\right)\right],\end{aligned} \qquad (4.18)$$

$$\Omega_1'(x)=xG(x)-\frac{2+\beta}{x}\Omega_1(x), \qquad (4.19)$$

$$\Omega_2'(x)=xG(x)-\frac{2-\beta}{x}\Omega_2(x). \qquad (4.20)$$

The leading asymptotic term for $G(x)$,$\gamma x^{\beta-2}$, appears explicitly in this system of equations.

We have used a fourth-order Runge–Kutta routine to solve the system (4.18)–(4.20), which we treat as coupled first-order equations for G, G', Ω_1, and Ω_2. The results are virtually identical with those obtained by solving (2.10) for x not near zero, and the coupled system has virtually the same degree of instability at small x as (2.10). Although there is one solution of this system of equations which is well-behaved in $\mathscr{D}$, there is an infinite class of other solutions that are not, and cumulative uncertainties will surely lead to nu-

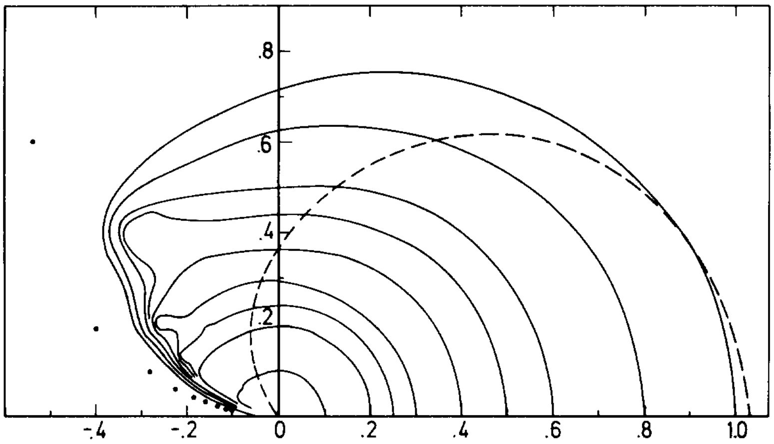


FIG. 3. Contours of constant magnitude of $x^2G(x)$ are shown. The values of $|x^2G|$ on successively larger contours are: 0.0003, 0.0015, 0.0025, 0.0034, 0.0070, 0.0112, 0.0161, 0.0282, 0.0423. The points give the locations of branch-points, at which $x^2G(x) = 1$.

merical instabilities here, just as they did with (2.10). In fact, one might expect that any replacement of (2.16) by a system of differential equations would behave in a similar fashion.

In Fig. 3 we have shown the contours in the upper half x plane along which $x^2G(x)$ is of constant magnitude, with γ given by (4.10). These contours are determined numerically from points that begin on the positive real axis. The contours become closer together in the vicinity of the branch-points in the second quadrant, and they all seem to approach the origin from the negative real direction. The large region between contours near $(-0.3,0.4)$ occurs because the derivative of $x^2G(x)$ has a zero in that region. The contours in Fig. 3 are numerically stable.

In Fig. 2 the function $F_1(x)$, which is given in terms of $G(x)$ by (2.11), is plotted for real x. The function has the asymptote (2.8) at small x, and the asymptote (4.12) at large x. The function

$$F(q^2) = q^{-2} + F_1(q^2) \tag{4.21}$$

is the factor multiplying the free-gluon propagator to give the full propagator in Mandelstam's equation. The physical scale for the momentum q^2 cannot be determined from the DS equation itself, but must be fixed by additional information, such as locations of gluonium states.

The solution of the full Mandelstam equation (2.1) is seen to have behavior similar to that obtained in I for the approximate case, and to suffer from the same deficiency, namely the appearance of branch-points at complex q^2. They must be regarded as an intrinsic deficiency of the Mandelstam equation, which one would hope to be able to eliminate by making a less drastic truncation of Dyson–Schwinger equations.

ACKNOWLEDGMENTS

We would like to thank W. J. Schoenmaker for helpful discussions. P. W. Johnson and K. Stam have carried out this work as scientific members of the Stichting F. O. M. (Foundation for Fundamental Research of Matter), which is financially supported by Z. W. O. (Netherlands Organization for Pure Scientific Research).

APPENDIX

Let us truncate the series (4.1) in such a way that only those terms are included for which the powers of x are less than, say, M; we call this truncated expression $G_M(x)$. We shall show that

$$\lim_{|x|\to 0} |x|^{-M}|G(x) - G_M(x)| = 0 \tag{A1}$$

for $x\in\mathscr{D}(\rho,\epsilon)$. This is the natural generalization of the concept of an asymptotic series[12] to the case in which nonintegral powers occur. Set

$$R_M(x) = x^4G''_M + 9x^3G'_M + (36 + \tfrac{63}{4})G_M - \Sigma(x,G_M), \tag{A2}$$

where Σ was defined in (2.14). For a given M, $G_M(x)$ is bounded for $x\in\mathscr{D}(\rho,\delta)$, and we can certainly find a subdomain, $\mathscr{F}_M \subset \mathscr{D}(\rho,\delta)$, for which say,

$$|x^2G_M(x)| \leqslant \tfrac{1}{2}. \tag{A3}$$

Now $[1 - x^2G_M(x)]R_M$ can be written as a finite number of terms, involving powers between x^M and x^{2M+2} and hence, in view of (A3),

$$|R_M(x)| \leqslant K_M|x|^M \tag{A4}$$

for $x\in\mathscr{F}_M$, where K_M depends on M. One may integrate (A2) to obtain an equation for G_M which is similar to (2.16), with an extra inhomogeneous term from R_M:

$$G_M(x) = -\tfrac{1}{6}x^{-7/2}\int_0^x dy\, y^{3/2} \times \sin\left(\frac{6}{x} - \frac{6}{y}\right)(\Sigma(y,G_M) + R_M(y)) . \tag{A5}$$

Let us subtract (A5) from (2.16), and express the result in terms of the function

$$h(x) = G(x) - G_M(x) \tag{A6}$$

as

$$h(x) = -\tfrac{1}{6}x^{-7/2}\int_0^x dy\, y^{3/2}\sin\left(\frac{6}{x} - \frac{6}{y}\right) \times [\Sigma(y,G) - \Sigma(y,G-h) + R_M(y)]. \tag{A7}$$

Equation (A7) is treated as a nonlinear integral equation for h, with the function G taken as the solution of (2.16) described in Sec. 3. The term $36\gamma x^{\beta-2}$ cancels out of (A7), so that R_M provides the only inhomogeneity. By an analysis similar to that described in Sec. 3, it is a simple exercise to establish the existence of a solution, $h(x)$, which is analytic in the domain $\mathscr{F}_M$, and in that region subject to the bound

$$|h(x)|<K'_M|x|^M, \tag{A8}$$

with the constant K'_M dependent upon M. The result, which may also be written as

$$|G(x)-G_M(x)|<K'_M|x|^M, \tag{A9}$$

guarantees that (4.1) is indeed an asymptotic series for G. For the simplified case considered in I, the corresponding series was not strongly asymptotic, and one would not expect that property of (4.1).

[1]D. Atkinson, J. K. Drohm, P. W. Johnson, and K. Stam, J. Math. Phys. **22**, 2074 (1981).

[2]S. Mandelstam, Phys. Rev. D **20**, 3223 (1979).

[3]S. K. Kim and M. Baker, Nucl. Phys. B **164**, 152 (1980); see their Eq. (1.1).

[4]For a summary on recent work, see M. Baker, "The Running Coupling Constant in Yang–Mills Theory", Lectures at Zakopane Summer School, May 1980, CERN TH-2905.

[5]W. J. Schoenmaker, Nucl. Phys. B **194**, 535 (1982).

[6]R. Delbourgo, J. Phys. G **5**, 603 (1979).

[7]A. A. Khelashvili, Theor. Math. Phys. **46**, 149 (1981) [Teoret. Mat. Fiz. Nauk SSSR **46**, 225 (1981)].

[8]K. O. Friedrichs, *Advanced Ordinary Differential Equations* (Gordon and Breach, New York, 1965).

[9]I. Stefanescu, "Irregular Singularities of Differential Equations", Groningen internal Report 167, December 1980, unpublished.

[10]L. Lapidus and J. Seinfeld, *Numerical Solution of Ordinary Differential Equations* (Academic, New York, 1971).

[11]L. Collatz, *Numerische Behandlungen von Differentialgleichungen* (Springer, Berlin, 1955).

[12]For definitions of an asymptotic power series, see M. Reed and B. Simon, *Methods of Mathematical Physics IV. Analysis of Operators* (Academic, New York, 1981) p. 40.

Local conformal-invariance of the wave equation for finite-component fields. I. The conditions for invariance, and fully-reducible fields

A. J. Bracken[a)]
Department of Physics, University of Colorado, Boulder, Colorado 80309

Barry Jessup[b)]
Department of Mathematics, University of Queensland, St. Lucia, Queensland, Australia 4067

(Received 17 March 1981; accepted for publication 20 July 1981)

The conditions for local conformal-invariance of the wave equation are obtained for finite-component fields, of Types Ia and Ib [in the terminology of Mack and Salam, Ann. Phys. **53**, 174 (1969).] These conditions generate a set of locally invariant free massless field equations and restrict the relevant representation of the Lie algebra $[(k_4 \oplus d) \oplus' sl(2,C)]$ in the index space of the field to belong to a certain class. Those fully-reducible representations which are in this class are described in full. The corresponding Type Ia field equations include only the massless scalar field equation, neutrino equations, Maxwell's equations, and the Bargmann–Wigner equations for massless fields of arbitrary helicity, and no others. In particular, it is confirmed [Bracken, Lett. Nuovo Cimento **2**, 574 (1971)] that not all Poincaré-invariant sets of massless Type Ia field equations are conformal-invariant, contrary to some often-quoted results of McLennan [Nuovo Cimento **3**, 1360 (1956)], which are shown to be invalid. It is also shown that in the case of a potential, the wave equation is never conformal-invariant in the strong sense (excluding gauge transformations).

PACS numbers: 11.10.Qr, 11.30.Ly

1. INTRODUCTION

Much has been written on the theory and possible applications to particle physics of the conformal group of space-time transformations: for reviews, see Kastrup,[1] Fulton, Rohrlich, and Witten,[2] Barut,[3] Ferrara, Gatto, and Grillo,[4] and Bayen.[5] These ideas were largely stimulated by observations that the wave equations satisfied by certain free, massless fields are locally[6] conformal-invariant.

Bateman[7] and Cunningham[8] (see also Dirac[9]) showed that this is so for the free-field Maxwell equations; and according to Cunningham, Bateman knew then of the invariance of the wave equation

$$\Box\psi(x) = 0$$
$$x = (ct,\mathbf{x}) = (x^\mu) \quad \mu = 0,1,2,3 \tag{1.1}$$

in the case of a scalar field ψ. We do not know who first proved the invariance of the two- and four-component neutrino equations. (See, however, Schouten and Haantjes,[10] Pauli,[11] and Bludman.[12]) McLennan[13] claimed to prove the invariance of each of Gårding's[14] "irreducible sets" of wave equations for massless multi-spinor fields (at least, of each set which admits plane-wave solutions, the remainder being unsuitable as free-field equations.) These sets of first-order equations are rather general and include ones described earlier by Dirac[15] and Fierz.[16] Gross[17] showed the invariance of the Bargmann-Wigner[18] equations for massless fields of arbitrary spin. The invariance of particular sets of massless field equations has also been argued by Lomont,[19] Penrose,[20] Kursunoglu,[21] Mack and Todorov,[22] Bayen,[23] Barut and Haugen,[24] Lopuszanski and Oziewicz,[25] Post,[26] Fegan,[27] Jakobsen and Vergne,[28] and Budini.[29] Kotecky and Niederle[30] have found the conditions for conformal invariance of a Lorentz-invariant equation of the form

$$L_\mu \partial^\mu \psi(x) = 0,$$
$$\partial^\mu = \partial/\partial x_\mu, \tag{1.2}$$

where the L_μ are matrices (not necessarily square), and ψ is a multicomponent field. However, they did not specifically require that ψ be massless in the sense of Eq. (1.1).

It is clear that a body of opinion has developed to the effect that wave equations for free, massless fields are conformal-invariant in all possible cases [at least, in all cases where the fields have (manifestly) Lorentz-invariant helicity[31]—there are known subtleties in the case of equations satisfied by potentials[22,32,33]]. In the introductory remarks to many papers on the conformal group and its applications, one can find passing reference to "the well known fact that massless wave equations are conformal-invariant."

This opinion has no doubt been reinforced by the observation[22,34,35] that every zero-mass, discrete spin, unitary, irreducible representation of the Poincaré group ISL(2,C) can be extended to a unitary, irreducible representation of SU(2,2), a group locally isomorphic to the conformal group. Given a consistent set of field equations for a free, massless, classical field with Lorentz-invariant helicity, one should be able to exhibit a Hilbert space of solutions carrying a representation of ISL(2,C) of this type. This solution space will then be invariant under the action of an SU(2,2) group.

One might be forgiven for thinking that there is little more to be said on this subject, at least in the case of fields having Lorentz-invariant helicity. On the other hand, it is

[a)]Permanent address: Department of Mathematics, University of Queensland, St. Lucia, 4067, Queensland, Australia.
[b)]Present address: Department of Mathematics, University of Toronto, Toronto, Ontario, M5S 1A7 Canada.

 0022-2488/82/101925-22$02.50

clear that the conformal invariance of the wave equation (1.1), which is evidently scale- and Poincaré-invariant, will in general require further, non-trivial, conditions to be satisfied when $\psi(x)$ is a multicomponent field. After all, the Poincaré group extended by dilatations is a *proper* subgroup of the conformal group, and we recall that in the case of Lagrangian field equations,[32,36–40] scale- and Poincaré-invariance does not guarantee conformal-invariance. We assert that, contrary to the body of opinion mentioned above, the wave equations satisfied by free massless fields are not in general locally conformal-invariant, even for fields having Lorentz-invariant helicity.

Some years ago, one of us showed[41] that if the index space of a field $\psi(x)$ carries an irreducible, finite-dimensional representation of sl(2,C) labeled (m,n) (in the familiar scheme, where $2m$ and $2n$ are non-negative integers, as described in the next section), then if $mn \neq 0$ the wave equation (1.1) is not locally conformal-invariant. If this be so, then some of the results of McLennan[13] in particular must be false. Indeed, it is not immediately clear that this result of Ref. 41 can be reconciled with the invariance of the Bargmann–Wigner equations,[18] though it turns out that there is no contradiction there, as we show in Sec. 4, where we discuss the results of earlier works in relation to ours. There also we point out some errors in McLennan's work, invalidating some of his conclusions.

What of the second argument suggested above, concerning the extendability of massless representations of ISL(2,C) to representations of SU(2,2)? The reconciliation of this fact with the noninvariance (in some cases) of the equation (1.1), has been discussed earlier.[41] Essentially, the point is that the group SU(2,2) which arises in this way cannot always be associated even locally with the conformal group. Suppose, for example, we construct a realization of the zero-mass, discrete spin, helicity λ, positive energy, unitary representation of ISL(2,C) in a Hilbert space of multicomponent fields $\psi(x)$, which have Lorentz-invariant helicity and whose index space carries a single representation (m,n) of sl(2,C). According to a result of Weinberg,[42] (see also Lemma 3.2 below), it must be true that $m - n = \lambda$, though not necessarily that $mn = 0$. According to the results of Ref. 22, we can find in addition to the ISL(2,C) generators P_μ and $M_{\mu\nu}$, operators D' and K'_μ acting on this space. Together these operators generate a unitary irreducible representation of SU(2,2) in the so-called "ladder series." Now what happens is this: If $mn \neq 0$, then K'_μ can *not* be identified with the generators of special conformal transformations of the fields $\psi(x)$. Those generators have rather specific forms, as described by Mack and Salam.[43] (See Sec. 2.) In particular, the operators K'_μ are not local in space–time when $mn \neq 0$. What is more, in those cases they only satisfy the appropriate commutation relations within the representation space—this is, only weakly on the fields, as a consequence of the free-field equations. In contrast, in the cases when $mn = 0$, K'_μ (and D') *are* identifiable with generators of conformal transformations.[22,35] They are local operators, and they can be defined on all (sufficiently smooth) fields of the given type, in such a way that the appropriate commutation relations are satisfied, whether or not the fields satisfy the free-field equations. These properties are crucial if one is to be able to talk meaningfully about conformal-invariance being preserved in the presence of interactions, when the free-field equations cease to hold.

In short, when $mn = 0$, the conformal group is a space–time symmetry group of the field equations, while when $mn \neq 0$, SU(2,2) is only a dynamical symmetry group of the one-particular Hilbert space. The difference between these two concepts is quite fundamental, but in the present context it has not generally been fully appreciated.

In view of the fact that not all possible Poincaré-invariant massless field theories are conformal-invariant, the invariance of the equations governing the electromagnetic and neutrino fields assumes, perhaps, a greater significance. Unfortunately, Ref. 41 seems to have been largely unnoticed,[44–46] and passing remarks persist to "the well known fact that...." Indeed, after this work[41] appeared, a proof of the conformal-invariance of the field equations in the cases $mn \neq 0$ was presented by Post.[26] This proof is deficient, as we show in Sec. 4, and Post's conclusions in this regard are false.

Recently there has been renewed interest in massless, higher-spin fields,[47–50] and fields of spin $\frac{3}{2}$ and $\frac{5}{2}$ in particular have been discussed in connection with "supergravity." The question now arises as to whether or not the theories proposed are conformal-invariant. While we do not examine this question specifically, it seems timely to investigate in detail the conditions under which the wave equation (1.1) is locally conformal-invariant when ψ is a finite-component field, and that is our main object here. We do not restrict ourselves to the cases where the index space carries an irreducible representation of sl(2,C), but rather consider the most general possible situation, according to Mack *et al.*,[43] where the field may be of Type Ib in their notation. (See Sec. 2.) Such fields have received comparatively little attention in the literature.[9,24,25,29,43,51–53] As free fields, their main interest lies in the possibility that one might be able to use them to describe spin multiplets of massless particles.[43] There are discouraging difficulties in attempting to describe such fields in any generality, because of the nature of the finite-dimensional index-space representations of the Lie algebra $\mathscr{W}$,

$$\mathscr{W} = (k_4 \oplus d) \oplus \mathrm{sl}(2,C), \tag{1.3}$$

which are involved. (See Sec. 2.) These representations are not in general completely reducible, and no classification of them is available. However, we find that only a certain class of representations is directly involved in the case of free massless fields obeying conformal-invariant equations.

Our main results are summarized in Theorems 3.1, 3.2, 3.3, 2.1, 3.4, 3.5, 4.1, and 4.2 below. In particular, we find that when Eq. (1.1) is locally conformal-invariant, then the field ψ must satisfy certain other equations. For example, if ψ is an antisymmetric tensor field $F_{\mu\nu}(x)$, then conformal-invariance of Eq. (1.1) requires that $F_{\mu\nu}$ satisfy *all* of Maxwell's free-field equations. Thus the imposition of conformal-invariance of the "mass condition" (1.1) can be a means of defining complete sets of conformal-invariant free-field equations. This fact leads us not only to well-known sets of wave equations, but also to new sets of locally conformal-invariant equations for massless fields of Type Ib with arbitrary helicity.

In general the extra equations which ψ must satisfy place severe restrictions on the representation of $\mathscr{W}$ carried by the index space of ψ. Furthermore, they imply that in every case ψ is a direct sum of fields having Lorentz-invariant helicity. Thus when ψ is a potential, the wave equation (1.1) is never conformal-invariant in the strong sense (i.e., excluding the possibility of gauge transformations to supplement the conformal transformations). This generalizes a well-known result[22,32,38] for the electromagnetic potential $A_\mu(x)$. We do not address the problem of classifying for potentials those equations which are conformal-invariant in the weak sense, i.e., up to a change of gauge.

Notation: We adopt the diagonal metric tensor $g_{\mu\nu} = g^{\mu\nu}$, with $g_{00} = -g_{11} = -g_{22} = -g_{33} = 1$. The alternating tensor $\epsilon_{\mu\nu\rho\sigma}$ is defined with $\epsilon^{0123} = -\epsilon_{0123} = 1$.

2. Preliminaries. Index space representations of sl(2,*C*) and $\mathscr{W}$

Consider infinitesmal conformal transformations of space–time,

$$x'^\mu = x^\mu + \epsilon x^\mu + \epsilon^\mu + \epsilon^{\nu\mu} x_\nu + (2\theta^\nu x_\nu x^\mu - \theta^\mu x^\nu x_\nu)$$

(symbolically,

$$x' = x + \delta g x), \tag{2.1}$$

where ϵ, ϵ^μ, $\epsilon^{\nu\mu}(= -\epsilon^{\mu\nu})$ and θ^ν are real infinitesimal parameters characterizing dilatations, translations, homogeneous Lorentz transformations, and special conformal transformations, respectively. Suppose we are given classical fields $\psi(x)$, with a fixed finite number of complex-valued components $\psi_a(x)$, and a cotransformation law of the general form

$$\psi'_a(x') = \psi_a(x) + \sum_b \delta S(\delta g, x)_{ab} \psi_b(x). \tag{2.2}$$

Mack *et al.*[43] (see also Flato *et al.*[38] and Kotecky and Niederle[54]) have shown that there is no loss of generality if the following statements are assumed to follow:

(1) The index space of the fields carries a finite-dimensional representation of the 11-dimensional Lie algebra[55] $\mathscr{W}$ of Eq. (1.3), with basis $\Sigma_{\mu\nu}(= -\Sigma_{\nu\mu})$, Δ and κ_μ, satisfying

$$i[\Sigma_{\mu\nu}, \Sigma_{\rho\sigma}] = g_{\mu\rho}\Sigma_{\nu\sigma} + g_{\nu\sigma}\Sigma_{\mu\rho} - g_{\nu\rho}\Sigma_{\mu\sigma} - g_{\mu\sigma}\Sigma_{\nu\rho}, \tag{2.3a}$$

$$i[\kappa_\mu, \Sigma_{\nu\rho}] = g_{\mu\rho}\kappa_\nu - g_{\mu\nu}\kappa_\rho, \tag{2.3b}$$

$$[\kappa_\mu, \kappa_\nu] = 0, \tag{2.3c}$$

$$[\Delta, \Sigma_{\mu\nu}] = 0, \tag{2.3d}$$

$$i[\kappa_\mu, \Delta] = \kappa_\mu. \tag{2.3e}$$

(2) The infinitesimal field transformation (2.2) corresponding to (2.1) can be written in the form

$$\psi'(x) = \psi(x) + i[\epsilon D + \epsilon^\mu P_\mu + \tfrac{1}{2}\epsilon^{\mu\nu} M_{\mu\nu} + \theta^\mu K_\mu]\psi(x), \tag{2.4}$$

where

$$P_\mu = i\partial/\partial x^\mu, \quad M_{\mu\nu} = x_\mu P_\nu - x_\nu P_\mu + \Sigma_{\mu\nu},$$
$$D = x^\mu P_\mu + \Delta, \quad K_\mu = 2x_\mu D - x^\nu x_\nu P_\mu + 2\Sigma_{\mu\nu} x^\nu + \kappa_\mu. \tag{2.5}$$

These statements (1) and (2) form the starting point of our analysis.

When we refer to "the field $\psi(x)$" we always have in mind the general element of the complex vector space $\mathscr{D}$ of smooth fields of a given type, i.e., fields whose components have partial derivatives of all orders, and which correspond to a given finite-dimensional representation of $\mathscr{W}$. This space $\mathscr{D}$ is the tensor product of the index space, with operators $\Sigma_{\mu\nu}$, Δ, etc., and the space of smooth functions $f(x)$, with operators x^μ, ∂_μ etc. In Eqs. (2.5) the operators Δ and x^μ, for example, really denote the extensions in the obvious way to the tensor product space, of the index-space operator Δ and the function-space operator x^μ. We abuse the notation in this way and rely on context to make precise what we mean in any given case. We remark also that a complex numerical multiple of the identity operator on any of these spaces will be denoted by the appropriate complex number; again we rely on context to make the meaning precise.

It can be seen that $\mathscr{D}$ is a common, invariant domain for the operators P_μ, $M_{\mu\nu}$, D, and K_μ. On this space, the following commutation relations hold[43]:

$$i[D, P_\mu] = P_\mu, \tag{2.6a}$$

$$i[K_\mu, D] = K_\mu, \tag{2.6b}$$

$$[D, M_{\mu\nu}] = 0, \tag{2.6c}$$

$$i[P_\mu, M_{\nu\rho}] = g_{\mu\rho}P_\nu - g_{\mu\nu}P_\rho, \tag{2.6d}$$

$$i[K_\mu, M_{\nu\rho}] = g_{\mu\rho}K_\nu - g_{\mu\nu}K_\rho, \tag{2.6e}$$

$$i[M_{\mu\nu}, M_{\rho\sigma}] = g_{\mu\rho}M_{\nu\sigma} + g_{\nu\sigma}M_{\mu\rho} - g_{\nu\rho}M_{\mu\sigma} - g_{\mu\sigma}M_{\nu\rho}, \tag{2.6f}$$

$$[P_\mu, P_\nu] = 0, \tag{2.6g}$$

$$[K_\mu, K_\nu] = 0, \tag{2.6h}$$

$$i[P_\mu, K_\nu] = 2M_{\mu\nu} - 2g_{\mu\nu}D. \tag{2.6i}$$

It follows that the operators D, P_μ, K_μ, and $M_{\mu\nu}$ (of which 15 are linearly independent) span a Lie algebra $\mathscr{A}$, which provides a representation in $\mathscr{D}$ of the Lie algebra of the conformal group.

The representation of $\mathscr{W}$ in the index space may also be regarded as a representation of the sl(2,*C*) subalgebra of $\mathscr{W}$, with basis $\Sigma_{\mu\nu}$. As such it will not in general be irreducible, but like any other finite-dimensional representation of sl(2,*C*) it will be fully reducible to a direct sum of irreducible representations. In any representation of sl(2,*C*), with basis $\Sigma_{\mu\nu}$, we can introduce the two Casimir operators[56]

$$C_1 = \tfrac{1}{2}\Sigma_{\mu\nu}\Sigma^{\mu\nu}$$
$$C_2 = \tfrac{1}{8}i\epsilon_{\mu\nu\rho\sigma}\Sigma^{\mu\nu}\Sigma^{\rho\sigma}. \tag{2.7}$$

Let (m,n) denote the irreducible representation, of dimension $(2m+1)(2n+1)$, in which these Casimir operators have the form

$$C_1 = 2m(m+1) + 2n(n+1)$$
$$C_2 = m(m+1) - n(n+1). \tag{2.8}$$

Here $2m$ and $2n$ are non-negative integers. Any given finite-dimensional representation $\mathscr{R}$ of sl(2,*C*), with representation space $\mathscr{V}$, will be a direct sum of such irreducible representations, for various *distinct* ordered pairs (m,n) in a finite set S

determined by $\mathscr{R}$, and with various positive integral multiplicites r_{mn}, determined by $\mathscr{R}$. Symbolically,

$$\mathscr{R} = \sum_{(m,n)\in S}^{\oplus} r_{mn}(m,n) \tag{2.9}$$

Let P_{mn} denote the projector onto that subspace $\mathscr{V}_{mn}$ of $\mathscr{V}$ which carries all the r_{mn} multiples of the irreducible representation (m,n). Then

$$\sum_{(m,n)\in S} P_{mn} = 1,$$
$$P_{mn}P_{kl} = P_{mn}\delta_{mk}\delta_{nl}, \quad (m,n),(k,l)\in S$$
$$[P_{mn},\Sigma_{\mu\nu}] = 0. \tag{2.10}$$

The space $\mathscr{V}$ is a direct sum of the subspaces $\mathscr{V}_{mn}$. Now define the operators

$$M = \sum_{(m,n)\in S} mP_{mn}, N = \sum_{(m,n)\in S} nP_{mn}, \tag{2.11}$$

and note from Eqs. (2.10) that

$$[M,\Sigma_{\mu\nu}] = 0 = [N,\Sigma_{\mu\nu}]$$
$$[M,N] = 0. \tag{2.12}$$

Thus M and N are commuting sl(2,C) scalars. We note also from Eqs. (2.8) and (2.11) that on all of $\mathscr{V}$,

$$C_1 = 2M(M+1) + 2N(N+1),$$
$$C_2 = M(M+1) - N(N+1). \tag{2.13}$$

These operators M and N are more convenient than C_1 and C_2 as labeling operators for the subspace $\mathscr{V}_{mn}$ of $\mathscr{V}$. While M and N by definition are functions of the projectors P_{mn}, it is important to see that, because the eigenvalues (m,n) of the pair (M,N) distinguish the subspaces $\mathscr{V}_{mn}$ onto which the P_{mn} project, these projectors can be regarded as functions of M and N. Any operator which commutes with M and N must commute with all the P_{mn}, and *vice versa*. A basis in $\mathscr{V}$ can be adopted, in which (the matrices of) all the operators $\Sigma_{\mu\nu}$ have the same block diagonal structure, each block corresponding to an irreducible representation of sl(2,C). In such a basis, the operators M, N, and P_{mn} are diagonal. Within any one of the blocks mentioned, M and N are multiples of the identity by the approproriate m and n values. We shall call such a basis an sl(2,C) basis, although it must be noted that M and N do not form a complete set of commuting operators on $\mathscr{V}$ if some of the r_{mn} are greater than unity.

In the case of interest, where $\mathscr{V}$ is the index space of the field ψ, and we have therein a representation of $\mathscr{W}$ which is being regarded as a representation $\mathscr{R}$ of sl(2,C), we see from Eqs. (2.3d) and (2.13) that

$$[\Delta,M(M+1)] = 0 = [\Delta,N(N+1)]. \tag{2.14}$$

It follows that Δ commutes with the positive, diagonalizable operators $(M+\frac{1}{2})^2$ and $(N+\frac{1}{2})^2$. But if a matrix A commutes with a diagonal, positive matrix B, then A commutes also with the positive, diagonal, square root of B. Thus Δ commutes with $(M+\frac{1}{2})$ and $(N+\frac{1}{2})$, and we have

$$[\Delta,M] = 0 = [\Delta,N], \tag{2.15}$$

and hence

$$[\Delta,P_{mn}] = 0. \tag{2.16}$$

It is *not* possible to prove that Δ can be taken to be diagonal in an sl(2,C) basis, as Mack *et al.*[43] claim to do in their Lemma 1, using Schur's lemma. The possible occurrence of repeated irreducible represenations of sl(2,C) causes the difficulty. A simple example counter to their result is provided by the representation of $\mathscr{W}$ on two-component fields ψ with[57]

$$\Sigma_{\mu\nu} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \kappa_\mu = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \Delta = \begin{pmatrix} 0 & 0 \\ \delta & 0 \end{pmatrix}, \tag{2.17}$$

which shows, indeed, that we cannot *a priori* assume the diagonalizability of Δ, and also that representations of $\mathscr{W}$ exist, more complicated than those described in Ref. 43.

A complete description of all finite-dimensional representations of $\mathscr{W}$ is not available. However, we shall see that only a subclass of such representations arises in connection with massless fields obeying locally conformal-invariant field equations. In particular, only representations[58] of Class $\mathscr{Q}$ (though not even all representations of this class) will arise:

Definition 2.1: A representation of $\mathscr{W}$ will be called of Class $\mathscr{Q}$ if it is finite-dimensional and its basis operators κ_μ, Δ and $\Sigma_{\mu\nu}$ satisfy

$$\kappa_\mu \kappa^\mu = 0, \tag{2.18a}$$
$$\Sigma_{\mu\nu}\kappa^\nu = (\Delta + i)\kappa_\mu \tag{2.18b}$$
$$\Delta^4 + (C_1+1)\Delta^2 + (C_2)^2 = 0 \tag{2.18c}$$

where C_1 and C_2 are the sl(2,C) invariants defined in terms of the $\Sigma_{\mu\nu}$ as in Eqs. (2.7). □

The representation defined by Eqs. (2.17) provides a rather simple example of a Class $\mathscr{Q}$ representation, although it is not one which arises in connection with locally conformal-invariant massless field equations, as we shall see.

It is important to show that this definition is a sensible one, to the extent that Eqs. (2.18) form a $\mathscr{W}$-invariant set. These equations are evidently invariant under transformations generated by $\Sigma_{\mu\nu}$ and Δ. For transformations generated by κ_μ, the invariance of Eq. (2.18a) follows because $[\kappa_\mu,\kappa_\nu] = 0$. Consider Eq. (2.18b) and the commutator

$$\begin{aligned}[\kappa_\mu,\Sigma_{\nu\rho}\kappa^\rho - (\Delta + i)\kappa_\nu] &= i(g_{\mu\nu}\kappa_\rho - g_{\mu\rho}\kappa_\nu)\kappa^\rho + i\kappa_\mu\kappa_\nu \\ &= ig_{\mu\nu}\kappa_\rho\kappa^\rho. \end{aligned} \tag{2.19}$$

When Eq. (2.18a) holds, this commutator vanishes, and the invariance of Eq. (2.18b) follows. Now consider Eq. (2.18c) and the commutator

$$[\kappa_\mu,\Delta^4 + (C_1+1)\Delta^2 + (C_2)^2]. \tag{2.20}$$

It can be deduced, using the commutation relations (2.3), that

$$[\kappa_\mu,\Delta^4] = (-4i\Delta^3 - 6\Delta^2 + 4i\Delta + 1)\kappa_\mu, \tag{2.21}$$

$$\begin{aligned}[\kappa_\mu,(C_1+1)\Delta^2] &= (\Delta - i)^2[\kappa_\mu,C_1+1] + (C_1+1)[\kappa_\mu,\Delta^2] \\ &= (\Delta - i)^2(2i\Sigma_{\mu\nu}\kappa^\nu + 3\kappa_\mu) + (C_1+1)(-2i\Delta - 1)\kappa_\mu, \end{aligned}$$

$$[\kappa_\mu,(C_2)^2] = 2i\Sigma_{\mu\nu}\Sigma^{\nu\rho}\Sigma_{\rho\tau}\kappa^\tau + 7\Sigma_{\mu\nu}\Sigma^{\nu\rho}\kappa_\rho \tag{2.22}$$
$$+ (2iC_1 - 6i)\Sigma_{\mu\nu}\kappa^\nu + 3C_1\kappa_\mu. \tag{2.23}$$

(In deriving the last of these equations, we found it helpful to use the identity

$$(C_2)^2 = \tfrac{1}{2}C_1(C_1+1) - \tfrac{1}{4}\Sigma_{\mu\nu}\Sigma^{\nu\rho}\Sigma_{\rho\tau}\Sigma^{\tau\mu}, \tag{2.24}$$

which follows from Eqs. (2.72) and (2.74) in Lemma 2.5 below.) When Eq. (2.18b) holds, Eqs. (2.22) and (2.23) reduce to

$$[\kappa_\mu,(C_1+1)\Delta^2] = \{(\Delta - i)^2(2i\Delta+1) - (C_1+1)(2i\Delta+1)\}\kappa_\mu \tag{2.22'}$$

$$[\kappa_\mu(C_2)^2] = \{2i(\Delta+i)^3 + 7(\Delta+i)^2 + (2iC_1 - 6i)(\Delta+i) + 3C_1\}\kappa_\mu, \tag{2.23'}$$

and, when combined with Eq. (2.21), enable us to see that the commutator (2.20) vanishes, so that Eq. (2.18c) is indeed $\mathscr{W}$-invariant.

Let us investigate something of the structure of an arbitrary representation $\mathscr{T}$ of class $\mathscr{D}$, with representation space $\mathscr{V}$. Regarded as a representation $\mathscr{R}$ of sl(2,C), it will have the form (2.9), for a finite set S and positive integers r_{mn} determined by $\mathscr{T}$. We introduce the projectors P_{mn} and the operators M and N as in the general discussion above. We first use the P_{mn} to write $\mathscr{V}$ as a direct sum of subspaces in two different ways:

(1) Let S_1 denote the set of *distinct* values θ of $|m-n|$ obtained as (m,n) runs over S. Every such number θ is a non-negative integer or semi-integer. For each θ in S_1, define the projector

$$P_{1\theta} = \sum_{\substack{(m,n)\in S\\ |m-n|=\theta}} P_{mn}. \tag{2.25}$$

It follows that

$$\sum_{\theta\in S_1} P_{1\theta} = 1, \quad P_{1\theta}P_{1\theta'} = P_{1\theta}\delta_{\theta\theta'}, \quad \theta,\theta'\in S_1. \tag{2.26}$$

Then $\mathscr{V}$ is a direct sum of the subspaces $\mathscr{V}_{1\theta}, \theta\in S_1$, where

$$\mathscr{V}_{1\theta} = P_{1\theta}\mathscr{V}. \tag{2.27}$$

It can be seen that on $\mathscr{V}_{1\theta}$ the operator $(M-N)^2$ has the value θ^2.

(2) Let S_2 denote the set of *distinct* values ν of $(m+n+1)$ obtained as (m,n) runs over S. Every such number ν is an integer or semi-integer, greater than or equal to 1. For each ν in S_2, define the projector

$$P_{2\nu} = \sum_{\substack{(m,n)\in S\\ m+n+1=\nu}} P_{mn}. \tag{2.28}$$

It follows that

$$\sum_{\nu\in S_2} P_{2\nu} = 1, \quad P_{2\nu}P_{2\nu'} = P_{2\nu}\delta_{\nu\nu'}, \quad \nu,\nu'\in S_2. \tag{2.29}$$

Then $\mathscr{V}$ is a direct sum of the subspaces $\mathscr{V}_{2\nu}, \nu\in S_2$, where

$$\mathscr{V}_{2\nu} = P_{2\nu}\mathscr{V}. \tag{2.30}$$

On $\mathscr{V}_{2\nu}$ the operator $(M+N+1)$ has the value ν.

We are not concerned with possible orthogonality or identity relations among the various $P_{1\theta}$ and $P_{2\nu}$. However, we note that as functions of the P_{mn}, they all commute with each other, and with Δ, $\Sigma_{\mu\nu}$, M, and N. Now consider Eq. (2.18c) which by definition holds in $\mathscr{T}$. Since the operators M and N satisfy Eqs. (2.13), we can write Eq. (2.18c) as

$$[\Delta^2 + (M-N)^2][\Delta^2 + (M+N+1)^2] = 0. \tag{2.31}$$

Define the operators

$$P_a = -[\Delta^2 + (M-N)^2][4MN + 2M + 2N + 1]^{-1}$$
$$P_b = [\Delta^2 + (M+N+1)^2][4MN + 2M + 2N + 1]^{-1}, \tag{2.32}$$

noting that $(4MN + 2M + 2N + 1)$ $[\equiv (M+N+1)^2 - (M-N)^2]$ has a well-defined inverse because M and N are commuting and non-negative. It follows from Eq. (2.31) that

$$P_a + P_b = 1,\ P_aP_b = P_bP_a = 0,$$
$$P_aP_a = P_a,\ P_bP_b = P_b. \tag{2.33}$$

Thus P_a and P_b are projectors, and with their help we can write $\mathscr{V}$ as a direct sum of two subspaces $\mathscr{V}_a$ and $\mathscr{V}_b$, where

$$\mathscr{V}_a = P_a\mathscr{V}, \mathscr{V}_b = P_b\mathscr{V}. \tag{2.34}$$

It follows from Eq. (2.31) that on $\mathscr{V}_a$, $[\Delta^2 + (M+N+1)^2]$ vanishes, while on $\mathscr{V}_b$, $[\Delta^2 + (M-N)^2]$ vanishes. We note that P_a and P_b as defined commute with all P_{mn}, and hence with Δ, $\Sigma_{\mu\nu}$, M, N, $P_{1\theta}$, and $P_{2\nu}$.

Finally, we define the projectors

$$P_{a\theta} = P_aP_{1\theta} = P_{1\theta}P_a, \quad \theta\in S_1$$
$$P_{b\nu} = P_bP_{2\nu} = P_{2\nu}P_b, \quad \nu\in S_2. \tag{2.35}$$

Since it follows that

$$\sum_{\theta\in S_1} P_{a\theta} = P_a, \quad \sum_{\nu\in S_2} P_{b\nu} = P_b, \tag{2.36}$$

we have

$$\sum_{\substack{\theta\in S_1\\ \nu\in S_2}} (P_{a\theta} + P_{b\nu}) = 1. \tag{2.37}$$

Furthermore, it is easily seen that

$$P_{a\theta}P_{a\theta'} = P_{a\theta}\delta_{\theta\theta'}, \quad \theta,\theta'\in S_1$$
$$P_{b\nu}P_{b\nu'} = P_{b\nu}\delta_{\nu\nu'}, \quad \nu,\nu'\in S_2$$
$$P_{a\theta}P_{b\nu} = P_{b\nu}P_{a\theta} = 0, \quad \theta\in S_1, \nu\in S_2. \tag{2.38}$$

We can therefore write $\mathscr{V}$ as a direct sum of subspaces $\mathscr{V}_{a\theta}, \mathscr{V}_{b\nu} (\theta\in S_1, \nu\in S_2)$ with

$$\mathscr{V}_{a\theta} = P_{a\theta}\mathscr{V}, \mathscr{V}_{b\nu} = P_{b\nu}\mathscr{V}. \tag{2.39}$$

Note that some of the projectors $P_{a\theta}, P_{b\nu}$ could vanish identically. (Indeed, this could even be true of P_a or P_b.) Then the corresponding $\mathscr{V}_{a\theta}$ or $\mathscr{V}_{b\nu}$ is the trivial subspace of $\mathscr{V}$.

It follows from what we have said above that on any vector in $\mathscr{V}_{a\theta}$,

$$[\Delta^2 + (M+N+1)^2] = 0 \tag{2.40a}$$

$$(M-N)^2 = \theta^2 \tag{2.40b}$$

and hence, in particular,

$$(\theta^2 - 1) = [\Delta^2 + 2M(M+1) + 2N(N+1)] = (\Delta^2 + C_1). \tag{2.41}$$

Similarly, on any vector in $\mathscr{V}_{b\nu}$

$$[\Delta^2 + (M-N)^2] = 0, \tag{2.42a}$$

$$(M+N+1) = \nu, \tag{2.42b}$$

and so

$$(\Delta^2 + C_1) = (\nu^2 - 1). \tag{2.43}$$

We shall now show that each of the subspaces $\mathscr{V}_{a\theta}, \mathscr{V}_{b\nu}$ is $\mathscr{W}$-invariant. The operator $(\Delta^2 + C_1)$ commutes with Δ and $\Sigma_{\mu\nu}$. Consider the commutator

$$[\Delta^2 + C_1, \kappa_\rho] = 2i(\Delta + i)\kappa_\rho - 2i\Sigma_{\rho\nu}\kappa^\nu. \tag{2.44}$$

In the representation $\mathscr{T}$, the right-hand side vanishes by virtue of Eq. (2.18b). It follows that in $\mathscr{T}$, the operator $(\Delta^2 + C_1)$ is a $\mathscr{W}$-scalar. Since the subspaces $\mathscr{V}_{a\theta}$ correspond to distinct eigenvalues of this operator, they are not mixed together under the action of $\mathscr{W}$. Similarly, the subspaces $\mathscr{V}_{b\nu}$ are not mixed together, nor are the subspaces $\mathscr{V}_{a\theta}$ and $\mathscr{V}_{b\nu}$, with $\theta \neq \nu$. It remains to show that in a case with $\theta = \nu = \rho$, say, the subspaces $\mathscr{V}_{a\rho}$ and $\mathscr{V}_{b\rho}$ are not mixed together. Now on $\mathscr{V}_{a\rho}$ we have $(M - N)^2 = \rho^2$, so that any $v_a \in \mathscr{V}_{a\rho}$ can only have components belonging to irreducible representations (m,n) of sl(2,C) with $|m - n| = \rho$, i.e., the representations $(\rho,0)$, $(\rho + \frac{1}{2},\frac{1}{2})$,$\cdots$ and $(0,\rho)$, $(\frac{1}{2},\rho + \frac{1}{2})$,$\cdots$. Similarly, any $v_b \in \mathscr{V}_{b\rho}$ can only have components in representations (m,n) with $(m + n + 1) = \rho$, i.e., the representations $(\rho - 1,0)$, $(\rho - \frac{3}{2},\frac{1}{2})$,..., $(0,\rho - 1)$. But these two sets of sl(2,C) representations are disjoint, and moreover cannot be linked by the operators Δ, $\Sigma_{\mu\nu}$ and κ_μ: the operators Δ and $\Sigma_{\mu\nu}$ cannot link inequivalent representations of sl(2,C) since they commute with M and N; and the four-vector operator κ_μ can link[56] a representation (m,n) only with $(m + \frac{1}{2}, n + \frac{1}{2})$, $(m + \frac{1}{2}, n - \frac{1}{2})$, $(m - \frac{1}{2}, n + \frac{1}{2})$ and $(m - \frac{1}{2}, n - \frac{1}{2})$. Thus, Δ, $\Sigma_{\mu\nu}$ and κ_μ cannot link the subspaces $\mathscr{V}_{a\rho}$ and $\mathscr{V}_{b\rho}$ which are therefore separately invariant under the action of $\mathscr{W}$. Thus we see that the decomposition

$$\mathscr{V} = \sum_{\theta \in S_1}^{\oplus} \mathscr{V}_{a\theta} \oplus \sum_{\nu \in S_2}^{\oplus} \mathscr{V}_{b\nu} \tag{2.45}$$

is a decomposition of $\mathscr{V}$ into $\mathscr{W}$-invariant subspaces. It defines a decompositions of $\mathscr{T}$ into a direct sum of subrepresentations of $\mathscr{W}$.

It follows that if the given representation $\mathscr{T}$ is indecomposable, only one of the subspaces $\mathscr{V}_{a\theta}, \mathscr{V}_{b\nu}$ is nontrivial.

Definition 2.2: A representation of $\mathscr{W}$ of Class $\mathscr{Q}$ will be called a $\langle \theta \rangle$-representation, where θ is a non-negative integer or semi-integer, if its basis operators Δ, κ_μ, $\Sigma_{\mu\nu}$ and the non-negative operators M,N defined by Eqs. (2.11), satisfy Eqs. (2.40). It will be called a $\{\nu\}$-representation, where $\nu (\nu \geqslant 1)$ is an integer or semi-integer, if Eqs. (2.42) are satisfied. □

Then we have proved the following:

Lemma 2.1: Any indecomposable representation of $\mathscr{W}$ of Class $\mathscr{Q}$ is either a $\langle \theta \rangle$-representation for some θ, or a $\{\nu\}$-representation for some ν. □

In the context of this work, we find that $\{\nu\}$-representations are not of interest. This is fortunate, because we shall see that in every $\langle \theta \rangle$-representation Δ is diagonalizable, while the same cannot be said of every $\{\nu\}$-representation, as the example of a $\{1\}$-representation defined by Eqs. (2.17) shows. The structure of $\langle \theta \rangle$-representations is comparatively simple. Let us look at this structure a little more closely, for an arbitrary $\langle \theta \rangle$-representation $\mathscr{T}$, with representation space $\mathscr{V}$. Noting that Eq. (2.40a) holds by definition, we define the projectors

$$P_+ = -\tfrac{1}{2}i[\Delta + i(M + N + 1)][M + N + 1]^{-1},$$
$$P_- = +\tfrac{1}{2}i[\Delta - i(M + N + 1)][M + N + 1]^{-1}, \tag{2.46}$$

which satisfy

$$P_+ + P_- = 1,\ P_+P_+ = P_+,\ P_-P_- = P_-,$$
$$P_+P_- = P_-P_+ = 0. \tag{2.47}$$

Then $\mathscr{V}$ is a direct sum of the corresponding subspaces $\mathscr{V}_+$ and $\mathscr{V}_-$,

$$\mathscr{V}_+ = P_+\mathscr{V}, \quad \mathscr{V}_- = P_-\mathscr{V}. \tag{2.48}$$

On $\mathscr{V}_+$ we have

$$\Delta = +i(M + N + 1), \tag{2.49}$$

and on $\mathscr{V}_-$ we have

$$\Delta = -i(M + N + 1). \tag{2.50}$$

Because P_+ and P_- commute with $\Sigma_{\mu\nu}$, the subspaces $\mathscr{V}_+$ and $\mathscr{V}_-$ are separately sl(2,C)-invariant. It follows that we can choose bases in these subspaces such M and N, and hence Δ, are diagonal. This justifies our assertion above that Δ is always diagonalizable in a $\langle \theta \rangle$-representation. Now on $\mathscr{V}$ we also have, by definition of a $\langle \theta \rangle$-representation,

$$(M - N)^2 = \theta^2, \tag{2.51}$$

and it then follows from Eq. (2.49) that on $\mathscr{V}_+$, $-i\Delta$ has eigenvalues belonging to the series $(\theta + 1)$, $(\theta + 2)$, $(\theta + 3)$,..., while on $\mathscr{V}_-$ it has eigenvalues belonging to the series $-(\theta + 1)$, $-(\theta + 2)$, $-(\theta + 3)$,$\cdots$. Consider the effect of Δ, $\Sigma_{\mu\nu}$, and κ_μ on a basis vector in $\mathscr{V}_-$. Since Δ and $\Sigma_{\mu\nu}$ commute with $-i\Delta$, and so cannot change its eigenvalue, they must carry such a vector back into $\mathscr{V}_-$. Now Eq. (2.3c) says that κ_μ converts an eigenvector of $-i\Delta$ with eigenvalue δ, into one with eigenvalue $(\delta + 1)$. Since any eigenvalue from the first series above is greater by at least two units than any eigenvalue from the second series, it follows that κ_μ carries no basis vector from $\mathscr{V}_-$ into $\mathscr{V}_+$. In this way we see that $\mathscr{V}_-$ is invariant under the action of the operators of $\mathscr{W}$. By a similar argument we deduce that $\mathscr{V}_+$ is $\mathscr{W}$-invariant, and we conclude that the decomposition (2.48) defines a decomposition of $\mathscr{T}$ into a direct sum of subrepresentations. If $\mathscr{T}$ is indecomposable, one or the other of $\mathscr{V}_+, \mathscr{V}_-$ must be trivial.

Definition 2.3: A $\langle \theta \rangle$-representation of $\mathscr{W}$ will be called a $\langle \theta, + \rangle$-representation [respectively, a $\langle \theta, - \rangle$-representation] if, with the same notation as before,

$$\Delta = +i(M + N + 1) \tag{2.52}$$

[respectively

$$\Delta = -i(M + N + 1)]. \tag{2.53}$$

□

Then we have proved:

Lemma 2.2: Any indecomposable $\langle \theta \rangle$-representation is either a $\langle \theta, + \rangle$-representation or a $\langle \theta, - \rangle$-representation. □

Comment: A similar analysis cannot be performed for an arbitrary $\{\nu\}$-representation. In place of the operator $(M + N + 1)$ in Eq. (2.46) above we would have $(M - N)$, which is not always invertible. [See again the example defined by Eqs. (2.17), for which $M = N = 0$.] □

We can carry our investigation of $\langle\theta\rangle$-representations still further. Consider a $\langle\theta,+\rangle$-representation $\mathscr{T}$, with representation space $\mathscr{V}$, which has $\theta>0$ but is otherwise arbitrary. As for the general case of a Class $\mathscr{Q}$ representation described above, introduce the projectors P_{mn}, $(m,n)\in S$. In the present case, $(m-n)^2=\theta^2$ for every $(mn)\in S$. Let us split the set S of ordered pairs (m,n) into two subsets S_α and S_β according as $(m-n)=+\theta$ or $-\theta$, and define the corresponding projectors

$$P_\alpha=\sum_{(m,n)\in S_\alpha}P_{mn},P_\beta=\sum_{(m,n)\in S_\beta}P_{mn}. \tag{2.54}$$

Then

$$P_\alpha+P_\beta=1,\quad P_\alpha P_\alpha=P_\alpha,\quad P_\beta P_\beta=P_\beta,$$
$$P_\alpha P_\beta=P_\beta P_\alpha=0, \tag{2.55}$$

and we can write $\mathscr{V}$ as a direct sum of the corresponding subspaces $\mathscr{V}_\alpha$ and $\mathscr{V}_\beta$,

$$\mathscr{V}_\alpha=P_\alpha\mathscr{V},\quad \mathscr{V}_\beta=P_\beta\mathscr{V}. \tag{2.56}$$

Then, on $\mathscr{V}_\alpha$

$$M-N=+\theta, \tag{2.57}$$

while on $\mathscr{V}_\beta$,

$$M-N=-\theta. \tag{2.58}$$

It follows that vectors in $\mathscr{V}_\alpha$ belong to certain representations (m,n) of sl$(2,C)$ from the series $(\theta,0)$, $(\theta+\frac{1}{2},\frac{1}{2})$, $(\theta+1,1),\cdots$, while those in $\mathscr{V}_\beta$ belong to certain representations (m,n) from the series $(0,\theta)$, $(\frac{1}{2},\theta+\frac{1}{2})$, $(1,\theta+1)\cdots$. It is at once clear that Δ and $\Sigma_{\mu\nu}$, which commute with M and N, leave the two subspaces $\mathscr{V}_\alpha$ and $\mathscr{V}_\beta$ separately invariant. As we remarked before, κ_μ can only link the representation (m,n) with $(m\pm\frac{1}{2},n+\frac{1}{2})$ and $(m\pm\frac{1}{2},n-\frac{1}{2})$. Then it follows at once that, at least for $\theta>\frac{1}{2}$, κ_μ leaves $\mathscr{V}_\alpha$ and $\mathscr{V}_\beta$ separately invariant. In the case $\theta=\frac{1}{2}$, it is at first glance conceivable that κ_μ could link a vector in $\mathscr{V}_\alpha$ belonging to $(\frac{1}{2},0)$ with one in $\mathscr{V}_\beta$ belonging to $(0,\frac{1}{2})$, and one in $\mathscr{V}_\alpha$ belonging to $(1,\frac{1}{2})$ with one in $\mathscr{V}_\beta$ belonging to $(\frac{1}{2},1)$ etc. However, we recall that on $\mathscr{V}$, by definition of a $\langle\theta,+\rangle$-representation,

$$\Delta=i(M+N+1) \tag{2.59}$$

so that Δ has the same value $3i/2$ on the first two vectors mentioned, and the same value $5i/2$ on the second two, etc. But Eq. (2.3e) shows that κ_μ cannot transform one eigenvector of Δ into another with the same eigenvalue. In this way we see that for every θ, $\theta>0$, the two subspaces $\mathscr{V}_\alpha,\mathscr{V}_\beta$ are separately $\mathscr{W}$-invariant, and the decomposition of $\mathscr{V}$ defines a decomposition of $\mathscr{T}$ into a direct sum of subrepresentation. If $\mathscr{T}$ is indecomposable, one of $\mathscr{V}_\alpha,\mathscr{V}_\beta$ must be trivial. (The case $\theta=0$ is special: there is just one subspace, on which $M=N$.) A completely analogous analysis can be given in the case of a $\langle\theta,-\rangle$-representation, with $\theta>0$.

Definition 2.4: A $\langle\theta,+\rangle$-representation of $\mathscr{W}$ (with $\theta>0$ or $\theta=0$) will be called a $[+\theta,+]$-representation if, with the same notation as before, Eq. (2.57) holds:

$$M-N=+\theta.$$

It will be called a $[-\theta,+]$-representation if Eq. (2.58) holds:

$$M-N=-\theta.$$

Similarly, we define $[+\theta,-]$-representation as a $\langle\theta,-\rangle$-representation in which Eq. (2.57) holds; and a $[-\theta,-]$-representation as one in which Eq. (2.58) holds. □

Then we have proved

Lemma 2.3: Any indecomposable $\langle\theta,+\rangle$-representation is either a $[+\theta,+]$-representation, or a $[-\theta,+]$-representation. □

Rather than refer $[+\theta,+]$-, $[0,+]$-, and $[-\theta,+]$-representations, where 2θ is a positive integer, we can henceforth refer simply to $[\lambda,+]$-representations, with 2λ an integer, positive, negative or zero. Such a representation is characterized by Eqs. (2.18), and in addition[59]

$$M-N=\lambda, \tag{2.60a}$$
$$\Delta=+i(M+N+1). \tag{2.60b}$$

Similarly, a $[\lambda,-]$-representation is characterized by Eqs. (2.18) and

$$M-N=\lambda, \tag{2.61a}$$
$$\Delta=-i(M+N+1). \tag{2.61b}$$

We shall give one further result concerning the structure of such representations. Recall that Δ is diagonalizable in these cases.

Definition 2.5: A $[\lambda,+]$-representation will be called a $[\lambda,+;l,u]$-representation, where l and u are non-negative integers with $u\geqslant l$, if the eigenvalues of $(-i\Delta)$ are

$$|\lambda|+l+1,|\lambda|+l+2,\ldots,|\lambda|+u+1. \tag{2.62}$$

Similarly, a $[\lambda,-]$-representation will be called a $[\lambda,-;l,u]$-representation if the eigenvalues of $(-i\Delta)$ are

$$-(|\lambda|+l+1),-(|\lambda|+l+2),\ldots,$$
$$-(|\lambda|+u+1). \tag{2.63}$$

□

Lemma 2.4: An indecomposable $[\lambda,+]$-representation is a $[\lambda,+;l,u]$-representation for some l and u; and an indecomposable $[\lambda,-]$-representation is a $[\lambda,-;l,u]$-representation for some l and u.

Proof: Consider an indecomposable $[\lambda,+]$-representation, with representation space $\mathscr{V}$. Then Eqs. (2.60) hold, so that

$$-i\Delta=(2M+1-\lambda)=(2N+1+\lambda). \tag{2.64}$$

Because $2M$ and $2N$ have non-negative integral eigenvalues, we see that every eigenvalue δ of $(-i\Delta)$ in this representation is of the form

$$\delta=|\lambda|+t+1 \tag{2.65}$$

with t a non-negative integer. If there is only one such t, we set $l=t=u$ and the proof is complete. If there are more than one, we order them thus:

$$0\leqslant l=t_1<t_2<\cdots<t_n=u. \tag{2.66}$$

Then we have to show that $t_1,t_2\cdots t_n$ comprise all the integers from l to u. Suppose this is not the case, so that for some integral value of i between 1 and $n-1$,

$$t_{i+1}>t_i+1. \tag{2.67}$$

Since $(-i\Delta)$ is diagonalizable, $\mathscr{V}$ is the direct sum of the eigenspaces of $(-i\Delta)$. Let $\mathscr{V}_i$ be the direct sum of the eigenspaces corresponding to values of t not greater than t_i, and

$\mathscr{V}'_i$ the direct sum of those corresponding to values of t not less than t_{i+1}. Then

$$\mathscr{V} = \mathscr{V}_i \oplus \mathscr{V}'_i. \tag{2.68}$$

Since Δ and $\Sigma_{\mu\nu}$ commute with $(-i\Delta)$, they leave $\mathscr{V}_i$ and $\mathscr{V}'_i$ separately invariant. According to Eq. (2.3e),

$$\Delta\kappa_\mu = \kappa_\mu(\Delta + i),$$

so the action of κ_μ is to increase the eigenvalue of $(-i\Delta)$ by one unit. Because of the inequality (2.67), it follows then that κ_μ cannot carry a vector from $\mathscr{V}_i$ into $\mathscr{V}'_i$, nor from $\mathscr{V}'_i$ into $\mathscr{V}_i$: these spaces are also separately invariant under the action of κ_μ. In this way we see that Eq. (2.68) defines a decomposition of $\mathscr{V}$ into a direct sum of $\mathscr{W}$-invariant subspaces. Since the given representation is indecomposable, we have a contradication, and the inequality (2.67) cannot hold. An analogous proof applies in the case of an indecomposable $[\lambda, -]$-representation. □

Combining Lemmas 2.1, 2.2, 2.3, and 2.4 we have

Theorem 2.1: An indecomposable representation of $\mathscr{W}$ of Class $\mathscr{D}$ must be one of the following types:

(i) $[\lambda, +; l,u]$ or $[\lambda, -; l,u]$, for some integer or semi-integer λ (positive, negative or zero) and some non-negative integers l and u ($u \geqslant l$).

(ii) $\{\nu\}$, for some integer or semi-integer $\nu (\nu > 1)$. □

Comments:

1. We are not concerned at this stage with proving the existence of any of these representation types. The only Class $\mathscr{D}$ representation we have exhibited so far is the $\{1\}$-representation defined by Eqs. (2.17).

2. It is, of course, not true that a given representation of any one of these types need be indecomposable. Moreover, we have not proved that any two given representations of the same type (for example, any two $[\lambda, +; l,u]$-representations having the same values of λ, l, and u) are necessarily equivalent, even if they are both indecomposable.

3. We shall refer to $\psi(x)$ as an (indecomposable) Class $\mathscr{D}$ field if its index space carries an (indecomposable) representation of $\mathscr{W}$ of Class $\mathscr{D}$. Similarly, we shall refer to (indecomposable) $[\lambda, +; l,u]$-fields, $\{\nu\}$-fields, etc. □

We complete this section by presenting some results valid for any representation of the Lie algebra sl(2,C) (whether or not finite-dimensional, and whether or not contained in a representation of $\mathscr{W}$). These results will be required below.

Lemma 2.5: Let $\Sigma_{\mu\nu}$ $(= -\Sigma_{\mu\nu})$ be linear operators defined everywhere on a vector space, and satisfying there the commutation relations (2.3a) of sl(2,C). Define the Casimir operators C_1 and C_2 as in Eqs. (2.7). Then the following identities hold on that vector space:

$$\text{(i) } \tilde{\Sigma}_{\mu\nu}\Sigma^{\nu\lambda} = \Sigma_{\mu\nu}\tilde{\Sigma}^{\nu\lambda} = iC_2\delta_\mu{}^\lambda + i\tilde{\Sigma}_\mu{}^\lambda, \tag{2.69}$$

where

$$\tilde{\Sigma}_{\mu\nu} = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}\Sigma^{\rho\sigma}; \tag{2.70}$$

$$\text{(ii) } \tilde{\Sigma}_{\mu\nu}\tilde{\Sigma}^{\nu\lambda} = C_1\delta_\mu{}^\lambda + \Sigma_{\mu\nu}\Sigma^{\nu\lambda} - 2i\Sigma_\mu{}^\lambda; \tag{2.71}$$

$$\text{(iii) } \Sigma_{\mu\nu}\Sigma^{\nu\rho}\Sigma_{\rho\sigma}\Sigma^{\sigma\tau} - 4i\Sigma_{\mu\nu}\Sigma^{\nu\rho}\Sigma_\rho{}^\tau + (C_1 - 5)\Sigma_{\mu\nu}\Sigma^{\nu\tau} - 2i(C_1 - 1)\Sigma_\mu{}^\tau - [C_1 - (C_2)^2]\delta_\mu{}^\tau = 0; \tag{2.72}$$

$$\text{(iv) } \tilde{\Sigma}_{\mu\nu}\tilde{\Sigma}^{\nu\rho}\tilde{\Sigma}_{\rho\sigma}\tilde{\Sigma}^{\sigma\tau} - (C_1 + 1)\tilde{\Sigma}_{\mu\nu}\tilde{\Sigma}^{\nu\tau} + (C_2)^2\delta_\mu{}^\tau = 0; \tag{2.73}$$

$$\text{(v) } \Sigma_{\mu\nu}\Sigma^{\nu\rho}\Sigma_\rho{}^\mu = -2iC_1. \tag{2.74}$$

If, in addition, the vector space is finite-dimensional, so that the operators M and N can be introduced as in Eqs. (2.11) and (2.13) above, then the following identities also hold:

$$\begin{aligned}\text{(vi) } &[-i\Sigma_{\mu\nu} - (M - N + 1)g_{\mu\nu}]\\ &\times[-i\Sigma^{\nu\rho} + (M - N - 1)g^{\nu\rho}]\\ &\times[-i\Sigma_{\rho\sigma} - (M + N + 2)g_{\rho\sigma}]\\ &\times[-i\Sigma^{\sigma\tau} + (M + N)g^{\sigma\tau}] = 0;\end{aligned} \tag{2.75}$$

$$\begin{aligned}\text{(vii) } &[\tilde{\Sigma}_{\mu\nu} - (M - N)g_{\mu\nu}]\\ &\times[\tilde{\Sigma}^{\nu\rho} + (M - N)g^{\nu\rho}]\\ &\times[\tilde{\Sigma}_{\rho\sigma} - (M + N + 1)g_{\rho\sigma}]\\ &\times[\tilde{\Sigma}^{\sigma\tau} + (M + N + 1)g^{\sigma\tau}] = 0.\end{aligned} \tag{2.76}$$

Proof. (i) This result is obtained by substitution of various values for μ and λ, and use of the commutation relations (2.3a). For example, with $\mu = 0$, $\lambda = 1$ we have

$$\begin{aligned}\tilde{\Sigma}_{\mu\nu}\Sigma^{\nu\lambda} &= \tilde{\Sigma}_{02}\Sigma^{21} + \tilde{\Sigma}_{03}\Sigma^{31},\\ &= -\Sigma^{31}\Sigma^{21} + \Sigma^{21}\Sigma^{31},\\ &= i\Sigma^{23},\\ &= i\tilde{\Sigma}_0{}^1,\end{aligned} \tag{2.77}$$

as required.

(ii) We note that

$$\begin{aligned}\epsilon_{\mu\nu\rho\sigma}\epsilon^{\nu\lambda\alpha\beta} = &\begin{pmatrix}\lambda\alpha\beta\\ \mu\rho\sigma\end{pmatrix} - \begin{pmatrix}\lambda\alpha\beta\\ \mu\sigma\rho\end{pmatrix} + \begin{pmatrix}\lambda\alpha\beta\\ \rho\sigma\mu\end{pmatrix}\\ &- \begin{pmatrix}\lambda\alpha\beta\\ \rho\mu\sigma\end{pmatrix} + \begin{pmatrix}\lambda\alpha\beta\\ \sigma\mu\rho\end{pmatrix} - \begin{pmatrix}\lambda\alpha\beta\\ \sigma\rho\mu\end{pmatrix},\end{aligned} \tag{2.78}$$

where, for example,

$$\begin{pmatrix}\lambda\alpha\beta\\ \mu\rho\sigma\end{pmatrix} = \delta_\mu{}^\lambda\delta_\rho{}^\alpha\delta_\sigma{}^\beta. \tag{2.79}$$

Then we have

$$\begin{aligned}4\tilde{\Sigma}_{\mu\nu}\tilde{\Sigma}^{\nu\lambda} &= \epsilon_{\mu\nu\rho\sigma}\epsilon^{\nu\lambda\alpha\beta}\Sigma^{\rho\sigma}\Sigma_{\alpha\beta},\\ &= \delta_\mu{}^\lambda(\Sigma^{\alpha\beta}\Sigma_{\alpha\beta} - \Sigma^{\beta\alpha}\Sigma_{\alpha\beta})\\ &\quad + \Sigma^{\lambda\alpha}\Sigma_{\alpha\mu} - \Sigma^{\lambda\beta}\Sigma_{\mu\beta}\\ &\quad + \Sigma^{\beta\lambda}\Sigma_{\mu\beta} - \Sigma^{\alpha\lambda}\Sigma_{\alpha\mu},\end{aligned} \tag{2.80}$$

which yields the result (2.71), with the help of the commutation relations (2.3a).

(iii) Define

$$\begin{aligned}A_\mu{}^\nu &= -i\Sigma_\mu{}^\nu - \delta_\mu{}^\nu,\\ B_\mu{}^\nu &= \tilde{\Sigma}_\mu{}^\nu,\end{aligned} \tag{2.81}$$

and, suppressing tensor indices for the moment, write

$$\begin{aligned}&A \text{ for } A_\mu{}^\nu,\ 1 \text{ for } \delta_\mu{}^\nu,\ B \text{ for } B_\mu{}^\nu,\\ &A\circ A \text{ for } A_\mu{}^\lambda A_\lambda{}^\nu,\ A\circ B \text{ for } A_\mu{}^\lambda B_\lambda{}^\nu,\end{aligned} \tag{2.82}$$

and so forth. Then Eqs. (2.69) and (2.71), respectively, read as

$$A\circ B = B\circ A = C_2, \tag{2.83}$$

$$B\circ B = -A\circ A + C_1 + 1. \tag{2.84}$$

It follows from the second of these, multiplying on the left or right by $A\circ A$, that

$$A\circ A\circ A\circ A = -A\circ A\circ B\circ B + (C_1 + 1)A\circ A. \tag{2.85}$$

Using Eq. (2.83) twice in succession we see that Eq. (2.85) reduces to

$$A\circ A\circ A\circ A-(C_1+1)A\circ A+(C_2)^2=0, \quad (2.86)$$

which is equivalent to Eq. (2.72).

(vi) On substituting for C_1 and C_2 in Eq. (2.72), in terms of M and N from Eqs. (2.13), we get

$$[A-(M-N)]\circ[A+(M-N)]\circ$$
$$[A-(M+N+1)]\circ[A+(M+N+1)]=0, \quad (2.87)$$

which is equivalent to Eq. (2.75).

(iv) Multiplying Eq. (2.84) on the left or right by $B\circ B$, we get

$$B\circ B\circ B\circ B=-B\circ B\circ A\circ A+(C_1+1)B\circ B. \quad (2.88)$$

Again using Eq. (2.83) twice, we get

$$B\circ B\circ B\circ B-(C_1+1)B\circ B+(C_2)^2=0, \quad (2.89)$$

which is equivalent to Eq. (2.73).

(vii) On substituting for C_1 and C_2 in Eq. (2.73) in terms of M and N from Eqs. (2.13), we get

$$[B-(M-N)]\circ[B+(M-N)]\circ$$
$$[B-(M+N+1)]\circ[B+(M+N+1)]=0, \quad (2.90)$$

which is equivalent to Eq. (2.76).

(v) Using the commutation relations (2.3a), it is straightforward to show that if

$$\Gamma_{\mu\nu}=\Sigma_{\mu\alpha}\Sigma^{\alpha\beta}\Sigma_{\beta\nu}-3i\Sigma_\mu{}^\alpha\Sigma_{\alpha\nu}-iC_1g_{\mu\nu}, \quad (2.91)$$

then

$$\Gamma_{\mu\nu}=-\Gamma_{\nu\mu}. \quad (2.92)$$

It follows that $\Gamma_\mu{}^\mu=0$, whence

$$\Sigma_{\mu\alpha}\Sigma^{\alpha\beta}\Sigma_\beta{}^\mu-3i\Sigma_{\mu\alpha}\Sigma^{\alpha\mu}-4iC_1=0, \quad (2.93)$$

which is equivalent to Eq. (2.74). □

Comment:

1. Some of the identities given here were presented earlier by Bracken and Green[60] in the general context of identities for the generators of representations of SO(n). □

3. CONDITIONS TO BE SATISFIED FOR LOCAL CONFORMAL-INVARIANCE OF THE WAVE EQUATION

We are concerned with massless fields, and we shall take that to mean that they satisfy[61] the wave equation

$$\Box\psi=-P^\mu P_\mu\psi=0. \quad (3.1)$$

Definition 3.1: This equation will be said to be locally conformal-invariant on a vector space $\mathscr{U}\ (\subseteq\mathscr{D})$ consisting of solutions, if $\mathscr{U}$ is $\mathscr{A}$-invariant; that is to say, if $\psi\in\mathscr{U}$ implies $X\psi\in\mathscr{U}$, where X is any element of the Lie algebra $\mathscr{A}$ spanned by D, P_μ, K_μ, and $M_{\mu\nu}$. □

Comments:

1. We do not require that $\mathscr{U}$ must consist of all the solutions of Eq. (3.1) which lie in $\mathscr{D}$. Nor do we require that if $\psi\in\mathscr{D}$ is a solution, then so is $X\psi$, where X is any element of $\mathscr{A}$. As we shall see, such requirements would rule out of further consideration such interesting cases as the free electromagnetic field $F_{\mu\nu}(x)$, where conformal invariance of the wave equation holds *not* on the space of all smooth solutions of that equation, but only on the subspace of fields satisfying certain extra equations, viz., Maxwell's equations.

2. If ψ is to be a *potential* for a massless field χ of a different type [e.g., with index space carrying a different finite-dimensional representation of sl(2,C)], then it may not be appropriate to require that ψ satisfy the wave equation; nor, when it does, to require local conformal-invariance of this equation in the manner defined. One might only expect these requirements to be met, roughly speaking, "up to a change of gauge" of ψ. Our results are relevant to a potential ψ only in the restricted situation where one chooses a gauge such that ψ satisfies the wave equation, and asks if this equation is locally conformal-invariant when ψ transforms as in Eqs. (2.4) and (2.5), supplementary gauge transformations being suppressed. It is known that in the case of the four-vector potential of the free electromagnetic field, the equation $\Box A_\mu=0$ is not conformal-invariant in this sense.[22,32,38] We shall see that this result generalizes to all potentials. The only fields for which the wave equation is locally conformal-invariant are fields "having invariant helicity." □

In order to prove our first result, we exploit the isomorphism of $\mathscr{A}$ and the Lie algebra so(4,2). Following Mack *et al.*,[43] we define $J_{AB}(=-J_{BA})$, $A,B=0,1,2,3,5,6$ by

$$J_{\mu\nu}=M_{\mu\nu},\qquad J_{65}=D,$$
$$J_{5\mu}=\tfrac12(P_\mu-K_\mu),\quad J_{6\mu}=\tfrac12(P_\mu+K_\mu). \quad (3.2)$$

Then the commutation relations (2.6) can be written as

$$i[J_{AB},J_{CD}]=g_{AC}J_{BD}+g_{BD}J_{AC}-g_{BC}J_{AD}-g_{AD}J_{BC}, \quad (3.3)$$

where the extended metric tensor is diagonal, with $g_{55}=-1$, $g_{66}=+1$.

Theorem 3.1: (1) The wave equation (3.1) is locally conformal invariant on a vector space $\mathscr{U}\subseteq\mathscr{D}$ if and only if $\mathscr{U}$ is $\mathscr{A}$ invariant and every field ψ in $\mathscr{U}$ satisfies

$$W_{AB}\psi=0,\quad A,B=0,1,2,3,5,6, \quad (3.4)$$

where

$$W_{AB}=J_{AC}J^C{}_B+J_{BC}J^C{}_A+\tfrac13 g_{AB}J_{CD}J^{CD}. \quad (3.5)$$

(2) Any one solution in $\mathscr{D}$ of Eqs. (3.4) generates under the action of $\mathscr{A}$ an $\mathscr{A}$-invariant space of such solutions, on which the wave equation is locally conformal invariant.

Proof: (1) Suppose that the wave equation is locally conformal invariant on $\mathscr{U}$, and $\psi\in\mathscr{U}$. It follows from Definition 3.1 that

$$[\cdots[[P^\mu P_\mu,X_1],X_2],\ldots,X_n]\psi=0 \quad (3.6)$$

for any finite set of operators $X_1, X_2,\ldots,X_n$ in $\mathscr{A}$. Now from Eqs. (3.2),

$$\begin{aligned}P^\mu P_\mu&=(J_{5\mu}+J_{6\mu})(J_5{}^\mu+J_6{}^\mu)\\&=-J_{5A}J^A{}_5-J_{6A}J^A{}_6-J_{5A}J^A{}_6-J_{6A}J^A{}_5\\&=-\tfrac12W_{55}-\tfrac12W_{66}-W_{56}.\end{aligned} \quad (3.7)$$

Since W_{AB} by construction is an so(4.2)-tensor operator, we have (on $\mathscr{D}$)

$$i[W_{AB},J_{CD}]=g_{AD}W_{CB}-g_{AC}W_{DB}+g_{BD}W_{AC}-g_{BC}W_{AD}, \quad (3.8)$$

and so

$$i[P_\mu P^\mu,J_{5\nu}]=-W_{5\nu}-W_{6\nu}$$
$$[[P_\mu P^\mu,J_{5\nu}],J_{5\rho}]=W_{\rho\nu}+g_{\nu\rho}W_{55}+g_{\nu\rho}W_{65}. \quad (3.9)$$

It then follows from Eq. (3.6) that

$$W_{\rho\nu}\psi = 0, \quad \rho \neq \nu \tag{3.10}$$

and

$$\begin{aligned}(W_{00}+W_{55}+W_{65})\psi &= 0,\\ (W_{11}-W_{55}-W_{65})\psi &= 0,\\ (W_{22}-W_{55}-W_{65})\psi &= 0,\\ (W_{33}-W_{55}-W_{65})\psi &= 0.\end{aligned} \tag{3.11}$$

Similarly, from the commutator

$$[[P_\mu P^\mu, J_{50}], J_{60}] = W_{66} - W_{00} + W_{65} \tag{3.12}$$

we deduce that

$$(W_{66} - W_{00} + W_{65})\psi = 0. \tag{3.13}$$

From Eq. (3.9) we also have, provided $\rho \neq \nu$,

$$i[[[P_\mu P^\mu J_{5\nu}], J_{5\rho}], J_{5\sigma}] = g_{\rho\sigma} W_{5\nu} + g_{\nu\sigma} W_{5\rho}, \tag{3.14}$$

from which we deduce (taking $\sigma = \rho \neq \nu$) that

$$W_{5\nu}\psi = 0. \tag{3.15}$$

Similarly, from the commutator ($\rho \neq \nu$ here)

$$i[[[P_\mu P^\mu, J_{5\nu}], J_{5\rho}], J_{6\sigma}] = g_{\rho\sigma} W_{6\nu} + g_{\nu\sigma} W_{6\rho}, \tag{3.16}$$

we deduce that

$$W_{6\nu}\psi = 0. \tag{3.17}$$

Finally, from the commutator

$$[[[[P_\mu P^\mu, J_{50}], J_{51}], J_{51}], J_{60}] = W_{65}$$

we have

$$W_{65}\psi = 0. \tag{3.18}$$

Noting from the definition (3.5) that

$$W_A{}^A = W_{00} - W_{11} - W_{22} - W_{33} - W_{55} + W_{66} = 0, \tag{3.19}$$

we can readily see from Eqs. (3.10–3.13, 3.15, and 3.17–3.19) that all of Eqs. (3.4) hold.

Conversely, suppose that every ψ in a vector space $\mathscr{U}(\subseteq \mathscr{D})$ satisfies Eqs. (3.4). Then by Eq. (3.7) every ψ in $\mathscr{U}$ satisfies the wave equation. If in addition $\mathscr{U}$ is $\mathscr{A}$-invariant, then the wave equation is by definition locally conformal-invariant on $\mathscr{U}$.

(2) Suppose $\psi(\in\mathscr{D})$ satisfies Eqs. (3.4). Then it is obvious from the relations (3.8) that $\psi' = X_1X_2\cdots X_n\psi$ also satisfies these equations, where $X_1X_2,\ldots,X_n$ is any finite set of elements of $\mathscr{A}$. Let $\mathscr{U}_\psi$ be the vector subspace of $\mathscr{D}$ consisting of all finite linear combinations of all such ψ'. Then $\mathscr{U}_\psi$ is an $\mathscr{A}$-invariant space of solutions in $\mathscr{D}$ of Eqs. (3.4), and so by the first part of this theorem, is a space on which the wave equation is locally conformal-invariant. □

Comments:

1. This theorem enables us to replace the problem of finding for which field types there exist $\mathscr{A}$-invariant spaces of solutions of the wave equation with the simpler problem of finding for which field types there exist *any* solutions of the Equations (3.4). This is the advantage of having found an irreducible $\mathscr{A}$-tensor set of equations.

2. There is an obvious generalization to any situation where one has a representation $\mathscr{A}'$, on a vector space $\mathscr{D}'$, of the so(4,2) Lie algebra, with basis P'_μ, K'_μ, D', and $M'_{\mu\nu}$. The equation

$$P'_\mu P'^\mu\psi' = 0, \quad \psi'\in\mathscr{D}' \tag{3.20}$$

will hold on an $\mathscr{A}'$-invariant subspace $\mathscr{U}'$ of $\mathscr{D}'$ if and only if every vector ψ' in $\mathscr{U}'$ satisfies

$$W_{AB}'\psi' = 0, \tag{3.21}$$

where W_{AB}' and J_{AB}' are defined in terms of P'_μ, etc., as in Eqs. (3.2) and (3.5). And any one vector in $\mathscr{D}'$ satisfying Eqs. (3.21) will generate under the action of $\mathscr{A}'$ an $\mathscr{A}'$-invariant space of such vectors.

3. Barut and Böhm[62] have shown that the self-adjoint generators $j_{AB}(=-j_{BA})$ of any irreducible unitary representation of SU(2,2), in the ladder series, satisfy (on a suitable domain)

$$j_{AC}j^C{}_B + j_{BC}j^C{}_A + \tfrac{1}{3}g_{AB}j_{CD}J^{CD} = 0. \tag{3.22}$$

These representations are associated with the mass-zero representations of ISL(2,C), as remarked in the Introduction, and this result can be seen to be a corollary to Theorem 3.1—or rather, to its generalization described in Comment 2. However, we emphasize that we do not assume the representation (unitary or otherwise) of any *group* on the fields ψ and we are not concerned with any Hilbert space structure for such fields. □

We proceed to investigate the content of the (20 linearly independent) equations (3.4), writing them out in SO(3,1)-tensor form. We have:

$$\begin{aligned}A=\mu,\ B=\nu&:(M_{\mu\rho}M^\rho{}_\nu + M_{\nu\rho}M^\rho{}_\mu - \tfrac{1}{2}K_\mu P_\nu\\ &- \tfrac{1}{2}K_\nu P_\mu - \tfrac{1}{2}P_\mu K_\nu - \tfrac{1}{2}P_\nu K_\mu)\psi\\ &= -\tfrac{1}{3}g_{\mu\nu}(J_{CD}J^{CD})\psi.\end{aligned} \tag{3.23a}$$

$$\begin{aligned}A=\mu,\ B=5&:[-M_{\mu\nu}(P^\nu - K^\nu) + (P^\nu - K^\nu)M_{\nu\mu}\\ &-(P_\mu + K_\mu)D - D(P_\mu + K_\mu)]\psi = 0.\end{aligned} \tag{3.23b}$$

$$\begin{aligned}A=\mu,\ B=6&:[-M_{\mu\nu}(P^\nu + K^\nu) + (P^\nu + K^\nu)M_{\nu\mu}\\ &-(P_\mu - K_\mu)D - D(P_\mu - K_\mu)]\psi = 0.\end{aligned} \tag{3.23c}$$

$$\begin{aligned}A=5,\ B=5&:[D^2 + \tfrac{1}{4}(P^\mu P_\mu - K^\mu P_\mu - P^\mu K_\mu + K^\mu K_\mu)]\psi\\ &= -\tfrac{1}{6}(J_{CD}J^{CD})\psi.\end{aligned} \tag{3.23d}$$

$$A=5, B=6: P^\mu P_\mu\psi = K^\mu K_\mu\psi. \tag{3.23e}$$

$$\begin{aligned}A=6,B=6&:[D^2 - \tfrac{1}{4}(P^\mu P_\mu + K^\mu P_\mu + P^\mu K_\mu + K^\mu K_\mu)]\psi\\ &= -\tfrac{1}{6}(J_{CD}J^{CD})\psi.\end{aligned} \tag{3.23f}$$

Also, we note that

$$\begin{aligned}J_{CD}J^{CD} &= M_{\mu\nu}M^{\mu\nu} + K^\mu P_\mu + K^\mu P_\mu - 2D^2,\\ &= M_{\mu\nu}M^{\mu\nu} + 2K^\mu P_\mu + 8iD - 2D^2,\end{aligned} \tag{3.24}$$

using Eqs. (2.6). A set of equations equivalent to Eqs. (3.23) and more convenient than them is obtained by taking certain linear combinations and using the commutation relations (2.6) to reorder factors in some products. We get

$$P^\mu P_\mu\psi = 0, \tag{3.25a}$$

$$K^\mu K_\mu\psi = 0, \tag{3.25b}$$

$$M_{\mu\nu}P^\nu\psi = (i - D)P_\mu\psi, \tag{3.25c}$$

$$M_{\mu\nu}K^\nu\psi = (i + D)K_\mu\psi, \tag{3.25d}$$

$$\begin{aligned}(M_{\mu\rho}M^\rho{}_\nu + M_{\nu\rho}M^\rho{}_\mu - K_\mu P_\nu - K_\nu P_\mu)\psi\\ = -g_{\mu\nu}(M_{\rho\sigma}M^{\rho\sigma} - 2iD + 2D^2)\psi,\end{aligned} \tag{3.25e}$$

and

$$K_\mu P^\mu \psi = (M_{\mu\nu} M^{\mu\nu} - 4iD + 4D^2)\psi. \qquad (3.26)$$

We note also that when Eq. (3.26) holds, we have from Eq. (3.24)

$$J_{CD} J^{CD} \psi = (3M_{\mu\nu} M^{\mu\nu} + 6D^2)\psi. \qquad (3.27)$$

Finally we note that Eq. (3.26) is redundant, as it follows from Eq. (3.25e) by contraction. We therefore drop it from the set, leaving again $(1 + 1 + 4 + 4 + 10 =)20$ equations to be satisfied by ψ.

We now obtain an equivalent set of 20 equations involving the generators $\Sigma_{\mu\nu}, \Delta$, and κ_μ, by substituting into Eqs. (3.25) the expressions (2.5) for $M_{\mu\nu}, D$ and K_μ. At first sight it appears that the resulting equations will be very complicated, but great simplifications occur. For example, consider the third equation. We have from Eqs. (2.5)

$$M_{\mu\nu} P^\nu = x_\mu (P_\nu P^\nu) - (x_\nu P^\nu) P_\mu + \Sigma_{\mu\nu} P^\nu \qquad (3.28)$$

and

$$(i - D) P_\mu = (i - x_\nu P^\nu - \Delta) P_\mu, \qquad (3.29)$$

and so

$$M_{\mu\nu} P^\nu \psi = (i - D) P_\mu \psi \Rightarrow [x_\mu (P_\nu P^\nu) + \Sigma_{\mu\nu} P^\nu] \psi$$
$$= (i - \Delta) P_\mu \psi. \qquad (3.30)$$

Since we shall retain $P_\nu P^\nu \psi = 0$ as one equation in our set, Eq. (3.30) reduces to

$$\Sigma_{\mu\nu} P^\nu \psi = (i - \Delta) P_\mu \Psi. \qquad (3.31)$$

It is no surprise that all x-dependent terms disappear in the transition from Eq. (3.25c) to Eq. (3.31): as Eqs. (3.25) are locally conformal-invariant, they are locally translation-invariant. This can be exploited in the reduction of the remaining equations in the set (3.25). We obtain

Theorem 3.2: Equations (3.4) are equivalent to Eqs. (3.25). For fields on which the generators of infinitesimal conformal transformations have the form (2.5), they are also equivalent to the following:

$$P_\mu P^\mu \psi = 0, \qquad (3.32a)$$

$$\kappa_\mu \kappa^\mu \psi = 0 \qquad (3.32b)$$

$$\Sigma_{\mu\nu} P^\nu \psi = (i - \Delta) P_\mu \psi, \qquad (3.32c)$$

$$\Sigma_{\mu\nu} \kappa^\nu \psi = (i + \Delta) \kappa_\mu \psi, \qquad (3.32d)$$

$$(\Sigma_{\mu\rho} \Sigma^\rho{}_\nu + \Sigma_{\nu\rho} \Sigma^\rho{}_\mu - \kappa_\mu P_\nu - \kappa_\nu P_\mu)\psi$$
$$= -g_{\mu\nu} (\Sigma_{\rho\sigma} \Sigma^{\rho\sigma} + 2\Delta^2 - 2i\Delta)\psi. \qquad (3.32e)$$

Proof: Suppose Eqs. (3.25) hold, and consider Eq. (3.25d)

$$M_{\mu\nu} K^\nu \psi = (i + D) K_\mu \psi,$$
$$= K_\mu (2i + D)\psi, \qquad (3.33)$$

using Eq. (2.6b). Noting the forms (2.5) of $M_{\mu\nu}$ and K_μ, we proceed to simplfy the left-hand side. We have

$$(x_\mu P_\nu - x_\nu P_\mu) K^\nu \psi$$
$$= [2x_\mu (x_\nu P^\nu)^2 + 8ix_\mu (x_\nu P^\nu)$$
$$- 2ix_\mu (x_\nu P^\nu) + 2x_\mu (x_\nu P^\nu)\Delta + 8ix_\mu \Delta$$
$$+ 2x_\mu x_\nu (\Delta - i) P^\nu + x_\mu (P_\nu \kappa^\nu) - 2(x_\nu x^\nu) P_\mu (x_\rho P^\rho)$$
$$- 2ix_\mu (x_\nu P^\nu) + (x_\nu x^\nu)(x_\rho P^\rho) P_\mu + 2ix_\mu (x_\nu P^\nu)$$
$$- 2(x_\nu x^\nu) P_\mu \Delta - 2ix_\mu \Delta + 2x_\nu P_\mu x^\rho \Sigma_\rho{}^\nu - (x_\nu \kappa^\nu) P_\mu]\psi, \qquad (3.34)$$

and

$$\Sigma_{\mu\nu} K^\nu \psi = [2x^\nu (x_\rho P^\rho) \Sigma_{\mu\nu} - (x_\nu x^\nu)(i - \Delta) P_\mu$$
$$+ 2x^\nu \Sigma_{\mu\nu} \Delta + 2x_\rho \Sigma_{\mu\nu} \Sigma^{\nu\rho} + \Sigma_{\mu\nu} \kappa^\nu]\psi, \qquad (3.35)$$

noting Eqs. (3.32a) and (3.32c), already seen to follow from Eqs. (3.25). The right-hand side of Eq. (3.33) is

$$[2x_\mu (x_\nu P^\nu)^2 + 2x_\mu (x_\nu P^\nu)\Delta + 4ix_\mu (x_\nu P^\nu) - (x_\nu x^\nu) P_\mu (x_\rho P^\rho)$$
$$- (x_\nu x^\nu) P_\mu \Delta - 2i(x_\nu x^\nu) P_\mu + 2x_\mu (x_\nu P^\nu)\Delta$$
$$+ 2x_\mu \Delta^2 + 4ix_\mu \Delta$$
$$+ 2x^\rho (x_\nu P^\nu) \Sigma_{\mu\rho} + 2x^\rho \Sigma_{\mu\rho} \Delta + 4ix^\rho \Sigma_{\mu\rho}$$
$$+ (x_\nu P^\nu) \kappa_\mu + \kappa_\mu \Delta + 2i\kappa_\mu]\psi.$$

Combining these results, we see that Eq. (3.33) is

$$[x_\mu (P_\nu \kappa^\nu) + 2ix_\mu \Delta - (x_\nu \kappa^\nu) P_\mu + 2x_\rho \Sigma_{\mu\nu} \Sigma^{\nu\rho}$$
$$+ \Sigma_{\mu\nu} \kappa^\nu - 2x_\mu \Delta^2 - 2ix^\rho \Sigma_{\mu\rho}$$
$$- (x_\nu P^\nu) \kappa_\mu - \kappa_\mu \Delta - 2i\kappa_\mu]\psi = 0, \qquad (3.36)$$

which we write as $A\psi = 0$.

Now we note that since Eqs. (3.25) form an $\mathscr{A}$-invariant set, $P_\lambda \psi$ satisfies those equations whenever ψ does. Therefore it is also true that

$$AP_\lambda \psi = 0, \qquad (3.37)$$

and hence that

$$[A, P_\lambda]\psi = 0. \qquad (3.38)$$

Evaluating the commutator appearing here, we then get from Eq. (3.36)

$$[g_{\lambda\mu} (\kappa_\nu P^\nu) + 2ig_{\lambda\mu} \Delta - \kappa_\lambda P_\mu + 2\Sigma_{\mu\nu} \Sigma^\nu{}_\lambda$$
$$- 2g_{\lambda\mu} \Delta^2 - 2i\Sigma_{\mu\lambda} - \kappa_\mu P_\lambda]\psi = 0, \qquad (3.39)$$

and Eq. (3.36) then implies further that

$$[\Sigma_{\mu\nu} \kappa^\nu - \kappa_\mu (2i + \Delta)]\psi = 0, \qquad (3.40)$$

which is Eq. (3.32d). Contracting Eq. (3.39) with $g^{\lambda\mu}$ we get

$$\kappa_\nu P^\nu \psi = (\Sigma_{\nu\rho} \Sigma^{\nu\rho} + 4\Delta^2 - 4i\Delta)\psi, \qquad (3.41)$$

and combining this with Eq. (3.39) we get

$$(2\Sigma_{\mu\nu} \Sigma^\nu{}_\lambda - 2i\Sigma_{\mu\lambda} - \kappa_\mu P_\lambda - \kappa_\lambda P_\mu)\psi$$
$$= -g_{\lambda\mu} (\Sigma_{\nu\rho} \Sigma^{\nu\rho} + 2\Delta^2 - 2i\Delta)\psi, \qquad (3.42)$$

which we see is equivalent to Eq. (3.32e), using the commutation relations (2.3a). Note that Eq. (3.42) is equivalent to Eq. (3.39), as it also implies Eq. (3.41) on contraction with $g^{\lambda\mu}$.

In a similar way we first reduce Eq. (3.25b) to

$$[4x^\rho x_\sigma \Sigma_{\rho\nu} \Sigma^{\nu\sigma} + 4i(x_\nu x^\nu)\Delta - 4(x_\nu x^\nu)\Delta^2$$
$$+ 2(x_\nu x^\nu)(P_\rho \kappa^\rho) - 4(x_\nu P^\nu)(x_\rho \kappa^\rho)]\psi = \kappa_\nu \kappa^\nu \psi. \qquad (3.43)$$

But the left-hand side of this vanishes, as is seen by contracting Eq. (3.39) with $x^\lambda x^\mu$. Therefore Eq. (3.32b) holds. Equation (3.25e) yields no equations not included in Eqs. (3.32).

To complete the proof, we need to show that Eqs. (3.32) imply Eqs. (3.25). It is easy to see that this is so for Eqs.

(3.25a)–(3.25d), essentially by reversing the arguments above. In order to prove it so for Eq. (3.25e), we can proceed in the same way, or, more simply, as follows:

If ψ satisfies Eqs. (3.32), then so does $P_\lambda \psi$. But Eqs. (3.32) imply Eq. (3.25d), and hence

$$[M_{\mu\nu}K^{\nu} - (i+D)K_\mu]P_\lambda \psi = 0 \tag{3.44}$$

as well. But then it follows that

$$[M_{\mu\nu}K^{\nu} - (i+D)K_\mu, P_\lambda]\psi = 0, \tag{3.45}$$

or, using Eqs. (2.6),

$$\begin{aligned}&[2M_{\mu\nu}M^{\nu}{}_{\lambda} - 2iM_{\mu\lambda} + g_{\lambda\mu}(K_\nu P^\nu)\\ &\quad + 2ig_{\lambda\mu}D - 2g_{\lambda\mu}D^2 - K_\lambda P_\mu - K_\mu P_\lambda]\psi = 0.\end{aligned} \tag{3.46}$$

Contracting with $g^{\lambda\mu}$ we get Eq. (3.26), and substituting this back in Eq. (3.46), and noting the relations (2.6), we get Eq. (3.25e) as required. □

Note: We also find that for fields satisfying Eqs. (3.32), Eq. (3.24) reduces to

$$J_{CD}J^{CD}\psi = 6(C_1 + \Delta^2)\psi. \tag{3.47}$$

Comments:

1. In view of Theorem 3.1 (2), any one (smooth) solution of Eqs. (3.32) generates an $\mathscr{A}$-invariant vector space of such solutions. Our main problem is to find for which field types, i.e., for which finite-dimensional representations of $\mathscr{W}$ with basis operators $\Sigma_{\mu\nu}, \kappa_\mu$, and Δ, there exist *any* solutions of Eqs. (3.32).

2. Any finite-dimensional representation of $\mathscr{W}$ can be reduced to a direct sum of indecomposable representations, not necessarily irreducible, and correspondingly, any field ψ can be written as a direct sum of $\mathscr{W}$-indecomposable fields. Now as far as the index space of the field ψ is concerned, Eqs. (3.32) involve only the $\mathscr{W}$ operators. It follows that when these equations hold, they hold separately on each $\mathscr{W}$-indecomposable component field in the direct sum decomposition of ψ. In addition to this, consider the above-mentioned $\mathscr{A}$-invariant space $\mathscr{U}_\psi$ of solutions of Eqs. (3.32), generated by one solution ψ in the manner described in the proof of Theorem 3.1. The operators in $\mathscr{A}$, as far as their action on the index space of ψ is concerned, only involve the $\mathscr{W}$-operators, according to their definitions (2.5). Therefore $\mathscr{U}_\psi$ is the direct sum of the $\mathscr{A}$-invariant spaces generated by the $\mathscr{W}$-indecomposable components of ψ. For these reasons it is sufficient at the outset to consider fields ψ which are $\mathscr{W}$-indecomposable, i.e., whose index space carries an indecomposable representation of $\mathscr{W}$. □

In examining the implications of Eqs. (3.32), we begin with (3.32e), which we write in the form

$$\tau_{\mu\nu}\psi = r_{\mu\nu}\psi, \tag{3.48}$$

with

$$\begin{aligned}\tau_{\mu\nu} &= \Sigma_{\mu\rho}\Sigma^{\rho}{}_{\nu} + \Sigma_{\nu\rho}\Sigma^{\rho}{}_{\mu} + g_{\mu\nu}C_1,\\ &= 2\Sigma_{\mu\rho}\Sigma^{\rho}{}_{\nu} - 2i\Sigma_{\mu\nu} + g_{\mu\nu}C_1,\end{aligned} \tag{3.49}$$

$$r_{\mu\nu} = \kappa_\mu P_\nu + \kappa_\nu P_\mu - g_{\mu\nu}G, \tag{3.50}$$

$$G = C_1 + 2\Delta^2 - 2i\Delta, \tag{3.51}$$

and C_1 as in Eqs. (2.7).

We note that

$$\begin{aligned}&\tau_{\mu\nu} = \tau_{\nu\mu}, \quad \tau_\mu{}^\mu = 0,\\ &r_{\mu\nu} = r_{\nu\mu}.\end{aligned} \tag{3.52}$$

Then Eq. (3.48) implies that

$$r_\mu{}^\mu\psi = 0, \tag{3.53}$$

or equivalently,

$$\kappa_\mu P^\mu\psi = 2G\psi. \tag{3.54}$$

Equation (3.48) also implies that

$$\tau^{\mu\nu}\tau_{\mu\nu}\psi = \tau^{\mu\nu}r_{\mu\nu}\psi. \tag{3.55}$$

Using Eqs. (3.52) we see that

$$\begin{aligned}\tau^{\mu\nu}\tau_{\mu\nu} &= 2\Sigma^{\mu\alpha}\Sigma_\alpha{}^\nu(2\Sigma_{\nu\rho}\Sigma^\rho{}_\mu - 2i\Sigma_{\nu\mu} + g_{\nu\mu}C_1)\\ &= 4\Sigma^{\mu\alpha}\Sigma_{\alpha\nu}\Sigma^{\nu\rho}\Sigma_{\rho\mu} - 4i\Sigma^{\mu\alpha}\Sigma_{\alpha\nu}\Sigma^\nu{}_\mu - 4(C_1)^2\\ &= 4(C_1)^2 - 16(C_2)^2,\end{aligned} \tag{3.56}$$

using Eqs. (2.72) and (2.74) of Lemma 2.5. Now consider

$$\begin{aligned}\Sigma_\mu{}^\nu r_{\nu\rho}\psi &= (\Sigma_{\mu\nu}\kappa^\nu P_\rho + \Sigma_{\mu\nu}\kappa_\rho P^\nu - \Sigma_{\mu\rho}G)\psi\\ &= (P_\rho\Sigma_{\mu\nu}\kappa^\nu + \kappa_\rho\Sigma_{\mu\nu}P^\nu - ig_{\mu\rho}\kappa_\nu P^\nu\\ &\quad + i\kappa_\mu P_\rho - \Sigma_{\mu\rho}G)\psi\end{aligned}$$

[using Eqs. (2.3b)]

$$\begin{aligned}&= [P_\rho(i+\Delta)\kappa_\mu + \kappa_\rho(i-\Delta)P_\mu + i\kappa_\mu P_\rho\\ &\quad - (\Sigma_{\mu\rho} + 2ig_{\mu\rho})G]\psi\end{aligned}$$

[using Eqs. (3.32) and (3.54)]

$$\begin{aligned}&= [\Delta(\kappa_\mu P_\rho - \kappa_\rho P_\mu) + 2i(\kappa_\mu P_\rho + \kappa_\rho P_\mu)\\ &\quad - (\Sigma_{\mu\rho} + 2ig_{\mu\rho})G]\psi\end{aligned} \tag{3.57}$$

[using Eq. (2.3e)]. Then

$$\begin{aligned}\Sigma^{\rho\mu}\Sigma_\mu{}^\nu r_{\nu\rho}\psi &= [\Delta\Sigma^{\rho\mu}(\kappa_\mu P_\rho - \kappa_\rho P_\mu) - \Sigma^{\rho\mu}\Sigma_{\mu\rho}G]\psi\\ &= [2\Delta(i+\Delta)\kappa_\rho P^\rho + 2C_1G]\psi\end{aligned}$$

[using Eq. (3.32d)]

$$\begin{aligned}&= 2(2\Delta^2 + 2i\Delta + C_1)G\psi\\ &= 2[4\Delta^4 + 4(C_1+1)\Delta^2 + (C_1)^2]\psi.\end{aligned} \tag{3.58}$$

Now

$$\begin{aligned}\tau^{\nu\rho}r_{\nu\rho}\psi &= \tau^{\rho\nu}r_{\nu\rho}\psi\\ &= 2\Sigma^{\rho\mu}\Sigma_\mu{}^\nu r_{\nu\rho}\psi,\end{aligned} \tag{3.59}$$

using the definition (3.49) and noting Eqs. (3.52) and (3.53). Combining Eqs. (3.55), (3.56), (3.58), and (3.59) we get

$$[4(C_1)^2 - 16(C_2)^2]\psi = 4[4\Delta^4 + 4(C_1+1)\Delta^2 + (C_1)^2]\psi,$$

i.e.,

$$[\Delta^4 + (C_1+1)\Delta^2 + (C_2)^2]\psi = 0. \tag{3.60}$$

Now consider this equation, together with Eqs. (3.32b) and (3.32d). Any field satisfying Eqs. (3.32) must satisfy these three equations in particular. In Sec. 2 we have shown that this set of equations is $\mathscr{W}$-invariant, and in fact characterizes what we have called a representation of $\mathscr{W}$ of Class $\mathscr{D}$. Therefore we have

Theorem 3.3: The nonzero components of any field ψ satisfying Eqs. (3.32), belong to a representations of $\mathscr{W}$ of Class $\mathscr{D}$. □

Comment:

1. In the context of free, massless fields satisfying locally conformal-invariant equations it follows that we can, without significant loss of generality, limit ourselves at the outset to fields whose index spaces carry indecomposable representations of $\mathscr{W}$ of Class $\mathscr{Q}$. Then Eqs. (3.32b), (3.32d), and (3.60) hold identically. However, we must bear in mind that such an indecomposable Class $\mathscr{Q}$ field *may* represent only some of the components of a given field, whose index space carries a larger indecomposable representation of $\mathscr{W}$; and whose extra components, though set to zero by Eqs. (3.32b), (3.32d), and (3.60) when the field is free and massless, could become operative when the field is "in interaction." Such a possibility exists because the algebra $\mathscr{W}$ has representations which are not fully reducible. A classification of all such possibilities would require a classification of all indecomposable representations of $\mathscr{W}$ which "contain" a representation of Class $\mathscr{Q}$. Such a classification will not be attempted here, and we restrict our attention henceforth to indecomposable Class $\mathscr{Q}$ fields. □

We know that an indecomposable representation of $\mathscr{W}$ of Class $\mathscr{Q}$ is of one of the types listed in Theorem 2.1. We shall show that if Eqs. (3.32) are required to admit plane wave solutions, then representations of all types except $[\lambda, +;0,u]$ are eliminated. The existence of plane wave solutions is essential if the associated fields are to be able to describe free, massless particles (at the many-particle or one-particle level, according as the fields are quantized or not).

Definition 3.2: A massless plane wave is a field $\psi(x)$ of the form

$$\psi(x) = \psi_0 \exp(-ik^\mu x_\mu), \tag{3.61}$$

where ψ_0 is a constant nonzero field and the k^μ are real constants, not all zero, satisfying

$$k^\mu k_\mu = 0. \tag{3.62}$$

□

Lemma 3.1: Let $\psi(x)$ be a field whose index space carries the irreducible representation (m,n) of sl$(2,C)$, with basis operators $\Sigma_{\mu\nu}$. If the equations

$$\Sigma_{\mu\nu}\partial^\nu\psi = i(\alpha+1)\partial_\mu\psi, \tag{3.63}$$

where α is a constant, admit a massless plane wave solution, then

$$\alpha = -(m+n+1). \tag{3.64}$$

Proof: In the notation used in the proof of Lemma 2.5, Eq. (3.63) reads as

$$A\circ\partial\psi = \alpha\partial\psi. \tag{3.65}$$

Suppose that these equations admit a solution in the form of a massless plane wave (3.61). Then it follows that

$$A\circ k\psi_0 = \alpha k\psi_0. \tag{3.66}$$

Now in the representation (m,n), according to Lemma 2.5, Eq. (2.87),

$$[A-(m-n)]\circ[A+(m-n)]\circ[A-(m+n+1)]\circ [A+(m+n+1)] = 0. \tag{3.67}$$

Applying the operator on the left-hand side of this identity to $k\psi_0$, we get from Eq. (3.66)

$$[\alpha-(m-n)][\alpha+(m-n)][\alpha-(m+n+1)] [\alpha+(m+n+1)]k_\mu\psi_0 = 0 \tag{3.68}$$

Since $k_\mu\psi_0$ by assumption does not vanish for all μ, it follows that

$$\alpha\epsilon\{m-n, n-m, m+n+1, -(m+n+1)\}. \tag{3.69}$$

Case (1): $\alpha\neq 0$.

Multiply Eq. (3.66) on the left by B (again in the notation of Lemma 2.5). Then we get

$$B\circ A\circ k\psi_0 = \alpha B\circ k\psi_0 \tag{3.70}$$

whence, with the help of Lemma 2.5, Eq. (2.83) we have

$$B\circ k\psi_0 = \alpha^{-1}C_2 k\psi_0 \tag{3.71}$$

or, in view of Eqs. (2.8),

$$\tilde{\Sigma}_{\mu\nu}k^\nu\psi_0 = \alpha^{-1}(m-n)(m+n+1)k_\mu\psi_0,$$

i.e.,

$$\tilde{\Sigma}_{\mu\nu}k^\nu\psi_0 = \beta k_\mu\psi_0, \tag{3.72}$$

where

$$\beta = \alpha^{-1}(m-n)(m+n+1). \tag{3.73}$$

In view of Eq. (2.90), we then have in addition

$$\beta\in\{m-n, n-m, m+n+1, -(m+n+1)\}. \tag{3.74}$$

Consider the $\mu=0$ component of Eq. (3.72):

$$\tilde{\Sigma}_{0i}k^i\psi_0 = \beta k_0\psi_0$$

i.e.,

$$\mathbf{S}\cdot\mathbf{k}\psi_0 = -\beta k_0\psi_0 \tag{3.75}$$

where

$$\mathbf{S} = (\tilde{\Sigma}_{10},\tilde{\Sigma}_{20},\tilde{\Sigma}_{30}) = (\Sigma_{23},\Sigma_{31},\Sigma_{12})$$
$$\mathbf{k} = (k^1,k^2,k^3). \tag{3.76}$$

Let (s) denote the $(2s+1)$-dimensional irreducible representation of su(2). It is known that the representation (m,n) of sl$(2,C)$, when regarded as a representation of su(2) with basis operators $\mathbf{S}$, is a direct sum of those irreducible representations (s) with

$$s\in\{m+n, m+n-1,\ldots,|m-n|\}, \tag{3.77}$$

each such representation occurring once. It is also known that if $\mathbf{n}$ is a real unit vector, then in the representation (s), the operator $\mathbf{S}\cdot\mathbf{n}$ has eigenvalues $s, s-1,\ldots,-s$. It follows that in the representation (m,n) of sl$(2,C)$, $\mathbf{S}\cdot\mathbf{n}$ has eigenvalues $m+n, m+n-1,\ldots,-(m+n)$; in particular, the largest eigenvalue of $(\mathbf{S}\cdot\mathbf{n})^2$ equals $(m+n)^2$. Now Eq. (3.75) implies

$$(\mathbf{S}\cdot\mathbf{k})^2\psi_0 = \beta^2\mathbf{k}\cdot\mathbf{k}\psi_0, \tag{3.78}$$

since, by assumption, $(k_0)^2 = \mathbf{k}\cdot\mathbf{k}$. Thus on ψ_0, $(\mathbf{S}\cdot\mathbf{n})^2$ has the eigenvalue β^2, where

$$\mathbf{n} = \mathbf{k}/|\mathbf{k}|. \tag{3.79}$$

It then follows that

$$\beta^2 \leqslant (m+n)^2. \tag{3.80}$$

Next consider the $\mu=0$ component of Eq. (3.63),

$$\Sigma_{0i}k^i\psi_0 = i(\alpha+1)k_0\psi_0$$

i.e.,

$$\mathbf{T}\cdot\mathbf{k}\psi_0 = i(\alpha+1)k_0\psi_0, \tag{3.81}$$

where

$$\mathbf{T} = (\Sigma_{01},\Sigma_{02},\Sigma_{03}). \tag{3.82}$$

Let us define the operators

$$\mathbf{S}_\pm = \tfrac{1}{2}(\mathbf{S} \pm i\mathbf{T}). \tag{3.83}$$

Then it is easily checked from Eqs. (2.3a) that the S_{+i} (and likewise the S_{-i}) satisfy the su(2) commutation relations. Moreover, the S_{+i} commute with the S_{-i}, and

$$\begin{aligned}\mathbf{S}_+\cdot\mathbf{S}_+ &= \tfrac{1}{4}(\mathbf{S}\cdot\mathbf{S} - \mathbf{T}\cdot\mathbf{T} + 2i\mathbf{S}\cdot\mathbf{T})\\ &= \tfrac{1}{8}\Sigma_{\mu\nu}\Sigma^{\mu\nu} + \tfrac{1}{8}i\tilde{\Sigma}_{\mu\nu}\Sigma^{\mu\nu}\\ &= \tfrac{1}{4}(C_1 + 2C_2)\\ &= m(m+1)\end{aligned} \tag{3.84}$$

in the representation (m,n). Similarly,

$$\mathbf{S}_-\cdot\mathbf{S}_- = n(n+1) \tag{3.85}$$

in this representation. We can regard $\mathbf{S}_+$ as the basis operators of a representation (m) of su(2), and $\mathbf{S}_-$ as the basis operators of a representation (n) of su(2). Then, by the argument employed above for the operators $\mathbf{S}$, we can deduce that if $\mathbf{n}$ is any real unit vector, the maximum eigenvalue of $(\mathbf{S}_+\cdot\mathbf{n})^2$ is m^2 in the representation (m,n); and the maximum eigenvalue of $(\mathbf{S}_-\cdot\mathbf{n})^2$ is n^2. But according to Eqs. (3.75) and (3.81) we have

$$\mathbf{S}_\pm\cdot\mathbf{k}\psi_0 = -\tfrac{1}{2}[\beta \pm (\alpha+1)]k_0\psi_0, \tag{3.86}$$

whence

$$(\mathbf{S}_\pm\cdot\mathbf{n})^2\psi_0 = \tfrac{1}{4}[\beta \pm (\alpha+1)]^2\psi_0, \tag{3.87}$$

with $\mathbf{n}$ as in Eqs. (3.79). From this we can conclude that

$$\begin{aligned}&\tfrac{1}{4}(\beta+\alpha+1)^2 \leqslant m^2\\ &\tfrac{1}{4}(\beta-\alpha-1)^2 \leqslant n^2.\end{aligned} \tag{3.88}$$

The only pair of numbers α,β satisfying the conditions (3.69), (3.74), (3.80), and (3.88) is

$$\alpha = -(m+n+1),\ \beta = n-m \tag{3.89}$$

Case (2): $\alpha = 0$

According to the Lemma to be proved, there should be no massless plane wave solutions of Eq. (3.63) in this case, since $(m+n+1)$ is never zero. Suppose on the contrary that such a solution does exist. From Eq. (3.81) we have

$$\mathbf{T}\cdot\mathbf{k}\psi_0 = ik_0\psi_0, \tag{3.90}$$

while from the $\mu = (1,2,3)$ components of Eq. (3.63) we get

$$(\Sigma_{ij}k^j + \Sigma_{i0}k^0)\psi_0 = ik_i\psi_0,$$

i.e.,

$$(\mathbf{k}\wedge\mathbf{S} - k_0\mathbf{T})\psi_0 = -i\mathbf{k}\psi_0. \tag{3.91}$$

Now take the dot product of Eq. (3.91) on the left with $\mathbf{T}$, noting Eq. (3.90), to obtain

$$(\mathbf{T}\cdot\mathbf{k}\wedge\mathbf{s} - k_0\mathbf{T}\cdot\mathbf{T})\psi_0 = k_0\psi_0. \tag{3.92}$$

Next take the cross product of Eq. (3.91) on the left with $\mathbf{k}$ to obtain

$$[(\mathbf{S}\cdot\mathbf{k})\mathbf{k} - (\mathbf{k}\cdot\mathbf{k})\mathbf{S} - k_0(\mathbf{k}\wedge\mathbf{T})]\psi_0 = 0. \tag{3.93}$$

Noting that $(\mathbf{k}\cdot\mathbf{k}) = (k_0)^2$, and that

$$\begin{aligned}\mathbf{S}\cdot(\mathbf{k}\wedge\mathbf{T})\psi_0 &= [-\mathbf{T}\cdot(\mathbf{k}\wedge\mathbf{S}) - 2i\mathbf{T}\cdot\mathbf{k}]\psi_0\\ &= -\mathbf{T}\cdot(\mathbf{k}\wedge\mathbf{S})\psi_0 + 2k_0\psi_0,\end{aligned} \tag{3.94}$$

we take the dot product of Eq. (3.93) on the left with $\mathbf{S}$ to obtain

$$[(\mathbf{S}\cdot\mathbf{k})^2 - (k_0)^2\mathbf{S}\cdot\mathbf{S} + k_0\mathbf{T}\cdot(\mathbf{k}\wedge\mathbf{S}) - 2(k_0)^2]\psi_0 = 0. \tag{3.95}$$

Combining this equation with Eq. (3.92), we get

$$(\mathbf{S}\cdot\mathbf{k})^2\psi_0 = (k_0)^2[1 + \mathbf{S}\cdot\mathbf{S} - \mathbf{T}\cdot\mathbf{T}]\psi_0. \tag{3.96}$$

But if $\alpha = 0$, it follows from Eq. (3.69) that $m = n = r$, say. In the representation (r,r) of sl(2,C),

$$C_1 = \tfrac{1}{2}\Sigma_{\mu\nu}\Sigma^{\mu\nu} = \mathbf{S}\cdot\mathbf{S} - \mathbf{T}\cdot\mathbf{T} = 4r(r+1), \tag{3.97}$$

so that we have from Eq. (3.96)

$$(\mathbf{S}\cdot\mathbf{n})^2\psi_0 = (2r+1)^2\psi_0, \tag{3.98}$$

with $\mathbf{n}$ again as in Eq. (3.79). However, as argued above, the maximum eigenvalue of $(\mathbf{S}\cdot\mathbf{n})^2$ in the representation (r,r) is $(2r)^2$. Thus we have a contradiction, and there is no massless plane wave solution if $\alpha = 0$. □

A closely related result is

Lemma 3.2 (Weinberg's Lemma): Let $\psi(x)$, $\Sigma_{\mu\nu}$ be as in Lemma 3.1. If the equations

$$\tilde{\Sigma}_{\mu\nu}\partial^\nu\psi = \beta\partial_\mu\psi, \tag{3.99}$$

where β is a constant, admit a massless plane wave solution, then

$$\beta = n - m. \tag{3.100}$$

Proof: On a massless plane wave solution, Eqs. (3.99) reduce to

$$B\circ k\psi_0 = \beta k\psi_0 \tag{3.101}$$

(again in the notation used in the proof of Lemma 3.1). Multiplying on the left with A and using Eq. (2.83) we have

$$\begin{aligned}\beta A\circ k\psi_0 &= C_2 k\psi_0\\ &= (m-n)(m+n+1)k\psi_0.\end{aligned} \tag{3.102}$$

Suppose $\beta = 0$. Then Eq. (3.102) implies $m = n$, so that $\beta = n - m$ as required. Suppose $\beta \neq 0$. Then Eq. (3.102) becomes

$$A\circ k\psi_0 = \beta^{-1}(m-n)(m+n+1)k\psi_0 \tag{3.103}$$

and by Lemma 3.1,

$$\beta^{-1}(m-n)(m+n+1) = -(m+n+1) \tag{3.104}$$

whence $\beta = n - m$ as required. □

Comment:

1. Weinberg[42] considered free, *quantized*, positive-energy, massless fields, belonging to the irreducible representation (m,n) of sl(2,C). He showed that if such a field has (Lorentz-invariant) helicity h then, in our notation, $h = m - n$. Now the covariant statement that the field has invariant helicity h is Eq. (3.99), with $\beta = -h$ [as Eq. (3.75) shows when $k_0 = |\mathbf{k}| > 0$]. Furthermore, the possibility of quantizing a field ψ which satisfies Eqs. (3.99) is, in the usual formulations, dependent upon the existence of plane-wave solutions of those equations. For these reasons it seems appropriate to call Lemma 3.2 "Weinberg's Lemma", as we have done here. □

Theorem 3.4: If ψ is an indecomposable Class $\mathscr{D}$ field, and Eqs. (3.32) admit a massless plane wave solution, then ψ is an indecomposable $[\lambda, + ;0,u]$ field, for some integer or semi-integer λ, and some non-negative integer u.

Proof: In view of Theorem 2.1, it suffices to show that ψ cannot be (1) a $[\lambda, - ;l,u]$ field, (2) a $[\lambda, + ;l,u]$ field, where $l>0$, or (3) a $\{\nu\}$ field.

1. Suppose that ψ is both a $[\lambda, - ;l,u]$ field and a solution of Eqs. (3.32). Then Eqs. (2.61) hold, and Eq. (3.32c) yields

$$\Sigma_{\lambda\mu} P^{\mu}\psi = i(M+N+2)P_{\lambda}\psi. \tag{3.105}$$

Let P_{mni} denote the projector onto the ith one (in some ordering) of the r_{mn} multiples of the irreducible representation (m,n) of sl(2,C) carried by the index space of ψ. (cf. Sec. 2). Then P_{mni} commutes with $\Sigma_{\lambda\mu}$, M, and N, so that

$$\Sigma_{\lambda\mu} P^{\mu}\psi_{mni} = i(m+n+2)P_{\lambda}\psi_{mni}, \tag{3.106}$$

where

$$\psi_{mni} = P_{mni}\psi. \tag{3.107}$$

Now if ψ is a massless plane wave, so is ψ_{mni}, if it does not vanish. But Lemma 3.1 shows that there are no massless plane wave solutions of Eqs. (3.106). Thus ψ_{mni} vanishes, for every i and every possible (m,n). But then ψ vanishes, and we have a contradiction. Thus ψ cannot be a massless plane wave.

2. Suppose that ψ is both a $[\lambda, + ;l,u]$ field (with $l>0$), and a solution of Eqs. (3.32). Then by Definition 2.5 the smallest eigenvalue of $-i\Delta$ is $(|\lambda| + 1 + l)$, which is greater than $(|\lambda| + 1)$. Since Eqs. (2.60) hold here, it follows that the least eigenvalue of $(M+N)$ is greater than $|\lambda|$, and that the least eigenvalues of M and N are both greater than 0. Hence the operator (MN) is invertible. Now let P_t be the projector onto that subspace of the index space associated with the eigenvalue $(|\lambda| + 1 + t)$ of $-i\Delta$ $(l<t<u)$. Since ψ satisfies Eq. (3.32e), it satisfies (by contraction)

$$\kappa_{\mu} P^{\mu}\psi = (\Sigma_{\mu\nu}\Sigma^{\mu\nu} + 4\Delta^2 - 4i\Delta)\psi, \tag{3.108a}$$
$$= -8MN\psi. \tag{3.108b}$$

If ψ is nontrivial, then not every P_t can annihilate ψ. Of those P_t satisfying

$$P_t\psi \neq 0 \tag{3.109}$$

let P be the one having the smallest value of t. Since Eq. (2.3e) implies that κ_{μ} shifts the eigenvalue of $-i\Delta$ (and hence the value of t) upward by one unit, it then follows that

$$P\kappa_{\mu}P^{\mu}\psi = 0. \tag{3.110}$$

Eqs. (3.108) and (3.110) together imply

$$P(MN)\psi = 0. \tag{3.111}$$

But the projectors P_t evidently all commute with M and N, so that

$$(MN)P\psi = 0, \tag{3.112}$$

and since (MN) is invertible, we have

$$P\psi = 0, \tag{3.113}$$

contradicting the definition of P. Thus ψ cannot be a nontrivial solution of Eqs. (3.32). [Note that we did not need to assume ψ to be a massless plane wave solution. There are *no* nontrivial solutions to Eqs. (3.32) if ψ is a $[\lambda, + ;l,u]$ field with $l>0$.]

3. Suppose that ψ is both a $\{\nu\}$ field and a massless plane wave solution (3.61) of Eqs. (3.32). Contracting Eq. (3.32c) on the left with $\tilde{\Sigma}^{\rho\mu}$, using Lemma 2.5, Eq. (2.69), and noting Eqs. (2.13), we obtain

$$\Delta\tilde{\Sigma}_{\rho\nu}k^{\nu}\psi_0 = (M-N)(M+N+1)k_{\rho}\psi_0. \tag{3.114}$$

The $\rho=0$ component of this equation is

$$\Delta\, \mathbf{S}\cdot\mathbf{k}\psi_0 = (M-N)(M+N+1)k_0\psi_0 \tag{3.115}$$

where $\mathbf{S}$ and $\mathbf{k}$ are defined as in Eqs. (3.76). Since the operators Δ, $\mathbf{S}\cdot\mathbf{k}$, M and N all commute, we get from Eq. (3.115)

$$\Delta^2(\mathbf{S}\cdot\mathbf{k})^2\psi_0 = (M-N)^2(M+N+1)^2(k_0)^2\psi_0. \tag{3.116}$$

Since Eq. (2.42a) holds in a $\{\nu\}$ representation, we then have [introducing $\mathbf{n}$ as in Eq. (3.79)]

$$(M-N)^2(\mathbf{S}\cdot\mathbf{n})^2\psi_0 = (M-N)^2(M+N+1)^2\psi_0. \tag{3.117}$$

Now introduce, as in Sec. 2, the projector P_{mn} onto that subspace of the index space associated with the totality of representations (m,n) of sl(2,C) that are contained in the given $\{\nu\}$ representation of $\mathscr{W}$. Recalling that, for each chosen m and n, this projector commutes with Δ, M, N and $\Sigma_{\mu\nu}$, we get from Eq. (3.117)

$$(m-n)^2(\mathbf{S}\cdot\mathbf{n})^2\psi_{mn} = (m-n)^2(m+n+1)^2\psi_{mn}, \tag{3.118}$$

where

$$\psi_{mn} = P_{mn}\psi_0. \tag{3.119}$$

If $m\neq n$ we have then

$$(\mathbf{S}\cdot\mathbf{n})^2\psi_{mn} = (m+n+1)^2\psi_{mn}. \tag{3.120}$$

But, as remarked in the proof of Lemma 3.1, the largest eigenvalue of $(\mathbf{S}\cdot\mathbf{n})^2$ in the representation (m,n) of sl(2,C) is equal to $(m+n)^2$. It follows that

$$\psi_{mn} = 0, \quad m\neq n. \tag{3.121}$$

Now in a $\{\nu\}$ representation of $\mathscr{W}$, Eq. (2.42b) holds, and we see that the only representation (m,m) of sl(2,C) which can occur have

$$m = n = \tfrac{1}{2}(\nu - 1). \tag{3.122}$$

Thus we have

$$P_{rr}\psi_0 = \psi_0, \quad r = \tfrac{1}{2}(\nu - 1), \tag{3.123}$$

whence

$$M\psi_0 = N\psi_0 = r\psi_0. \tag{3.124}$$

We recall again that the four-vector operator κ_{μ} can link a representation (m,n) of sl(2,C) only with $(m\pm\frac{1}{2}, n+\frac{1}{2})$ and $(m\pm\frac{1}{2}, n-\frac{1}{2})$. It follows that

$$P_{rr}\kappa_{\mu}P_{rr} = 0. \tag{3.125}$$

Now ψ satisfies Eq. (3.108a), so that

$$\kappa_{\mu}k^{\mu}\psi_0 = 4[M(M+1) + N(N+1) + \Delta^2 - i\Delta]\psi_0,$$
$$= 4[2MN + M + N - i\Delta]\psi_0, \tag{3.126}$$

using Eq. (2.42a). Multiplying on the left by P_{rr}, using Eqs. (3.123) and (3.125), and noting that P_{rr} commutes with M, N, and Δ, we get

$$(2MN + M + N - i\Delta)\psi_0 = 0. \tag{3.127}$$

Then Eqs. (3.124) imply

$$\Delta\psi_0 = -2ir(r+1)\psi_0. \quad (3.128)$$

Now, since Eqs. (2.42a) and (3.124) hold, we have

$$\Delta^2\psi_0 = 0. \quad (3.129)$$

Consistency of Eqs. (3.128) and (3.129) requires

$$r = 0(\Rightarrow \nu = 1) \quad (3.130)$$

and

$$\Delta\psi_0 = 0. \quad (3.131)$$

Now consider Eq. (3.32c), which is supposed to be satisfied by ψ. On a plane wave solution, we have

$$\Sigma_{\mu\nu}k^\nu\psi_0 = (i-\Delta)k_\mu\psi_0, \quad (3.132)$$

so that Eq. (3.131) implies

$$\Sigma_{\mu\nu}k^\nu\psi_0 = ik_\mu\psi_0. \quad (3.133)$$

But Eqs. (3.124) and (3.130) imply

$$P_{00}\psi_0 = \psi_0, \quad (3.134)$$

so that we have

$$\Sigma_{\mu\nu}k^\nu P_{00}\psi_0 = ik_\mu\psi_0. \quad (3.135)$$

Since (0,0) is the trivial representation of sl(2,C),

$$\Sigma_{\mu\nu}P_{00} = 0, \quad (3.136)$$

and Eq. (3.135) yields

$$k_\mu\psi_0 = 0, \quad (3.137)$$

providing a contradiction. Thus ψ cannot be both a $\{\nu\}$ field and a massless plane wave solution of Eqs. (3.32). □

Comment:

1. We have yet to show that indecomposable $[\lambda, + ;0,u]$ representations exist, and that plane wave solutions of Eqs. (3.32) exist if ψ is a $[\lambda, + ;0,u]$ field. These questions will be examined in full in subsequent papers. In the next section we shall see that well-known sets of conformal-invariant free-field equations do provide illustrative examples, but all corresponding to cases with $u = 0$. □

Now if ψ is a $[\lambda, + ;0,u]$ field, then in particular,

$$\Delta = i(M+N+1), \quad (3.138a)$$

$$M - N = \lambda, \quad (3.138b)$$

and Eq. (3.32c) becomes

$$\Sigma_{\mu\nu}P^\nu\psi = i(M+N)P_\mu\psi. \quad (3.139)$$

Contracting on the left with $\widetilde{\Sigma}_\rho{}^\mu$ and using Lemma 2.5, Eq. (2.69), we get

$$iC_2P_\rho\psi = -i(M+N+1)\widetilde{\Sigma}_{\rho\mu}P^\mu\psi. \quad (3.140)$$

Using Eqs. (2.13) and (3.138b), and noting that $(M+N+1)$ has a well-defined inverse, we then obtain

$$\widetilde{\Sigma}_{\mu\nu}P^\nu\psi = -\lambda P_\mu\psi. \quad (3.141)$$

If this ψ is a positive-energy (resp., negative-energy) plane wave, Eq. (3.141) is a covariant statement that ψ has helicity λ (resp., $-\lambda$.) (cf. Comment 1 following Lemma 3.2). Noting Theorems 3.1, 3.2, 3.3, 2.1, and 3.4, we therefore have

Theorem 3.5: If the wave equation (1.1) is locally conformal-invariant on a vector space $\mathscr{U} \subseteq \mathscr{D}$, then the nonzero components of any plane-wave solution $\psi \in \mathscr{U}$ belong to a direct sum of indecomposable $[\lambda, + ;0,u]$ representations of $\mathscr{W}$, for various values of λ and u. Moreover, if ψ_λ is a direct-summand of such a plane wave solution, corresponding to the representation $[\lambda, + ;0,u]$ for some u, then ψ_λ has Lorentz-invariant helicity λ or $-\lambda$ according as the plane wave has positive or negative energy. □

Comment:

1. In this sense we justify our assertion in the Introduction that Eq. (1.1) is not locally conformal-invariant when ψ is a potential, since such finite-component fields do not have (manifestly) Lorentz-invariant helicity,[31] i.e., they do not satisfy equations of the general form of Eq. (3.141). □

4. CONNECTION WITH EARLIER WORK

Most earlier works on the conformal-invariance of massless field equations have been concerned with fields corresponding to representations of $\mathscr{W}$ of Type Ia, in the notation of Mack *et al.*[43], i.e., representations in which the $\kappa_\mu = 0$. In the light of Theorem 3.5, the following result is significant for such fields:

Theorem 4.1: An indecomposable $[\lambda, + ;0,u]$-representation of $\mathscr{W}$ is of Type Ia if and only if $u = 0$. For each integral and semi-integral λ, there exists exactly one (up to equivalence) indecomposable $[\lambda, + ;0,u]$-representation. It is in fact irreducible, and remains so when restricted to sl(2,C), the sl(2,C) content being $(\lambda,0)$ when $\lambda \geqslant 0$, and $(0,-\lambda)$ when $\lambda < 0$. In either case, the basis operator Δ satisfies

$$\Delta = i(|\lambda| + 1). \quad (4.1)$$

Proof: In an indecomposable $[\lambda, + ;0,u]$ representation, the eigenvalues of $-i\Delta$ are, according to (2.62),

$$|\lambda| + 1, |\lambda| + 2, \cdots |\lambda| + u + 1.$$

Since $-i\Delta$ is diagonalizable, the representation space is a direct sum of the corresponding eigenspaces. But if $\kappa_\mu = 0$, Eqs. (2.3) show that these eigenspaces are separately invariant under the action of the $\mathscr{W}$ algebra, contradicting the assumed indecomposability unless $u = 0$.

Conversely, when $u = 0$ the representation space consists of the single eigenspace corresponding to the eigenvalue $(|\lambda| + 1)$ of $-i\Delta$. Since the action of κ_μ is to increase the eigenvalue of $-i\Delta$ by one unit, it follows that in such a representation

$$\kappa_\mu = 0, \quad (4.2a)$$

$$\Delta = i(|\lambda| + 1). \quad (4.2b)$$

In view of the defining relations (3.138) of such a representation, we have then

$$M + N = |\lambda|, \quad M - N = \lambda \quad (4.3)$$

so that if $\lambda \geqslant 0$,

$$M = \lambda, \quad N = 0, \quad (4.4a)$$

and if $\lambda < 0$,

$$M = 0, \quad N = -\lambda. \quad (4.4b)$$

It follows from the meaning of M and N that if $\lambda \geqslant 0$, the representation $[\lambda, + ;0,0]$, regarded as a representation of sl(2,C), is a direct sum of replicas of $(\lambda,0)$; while if $\lambda < 0$, it is a direct sum of replicas of $(0,-\lambda)$. But when Eqs. (4.2) hold, the corresponding irreducible sl(2,C) subspaces are also $\mathscr{W}$-invariant, so that if the given representation of $\mathscr{W}$ is indecomposable, it must consist of a single irreducible representation $(\lambda,0)$ or $(0,-\lambda)$ of sl(2,C).

It can now be seen that there exists exactly one (up to equivalence) indecomposable representation of $\mathscr{W}$ satisfying all these conditions for a given value of λ. It consists of the representation $(\lambda,0)$ of sl(2,C) [or $(0,-\lambda)$, if $\lambda<0$], extended to a representation of $\mathscr{W}$ by defining κ_μ and Δ via Eqs. (4.2). It is evidently irreducible. □

For an irreducible $[\lambda,+;0,0]$ field then, Eq. (4.2b) holds and it can be seen from the cotransformation law (2.4) and (2.5) for the field under changes of scale in particular, that such a field has the length dimension $-(|\lambda|+1)$. This is the "canonical" dimension of a field corresponding to a representation $(|\lambda|,0)$ or $(0,|\lambda|)$ of sl(2,C).

Combining Theorems 3.5 and 4.1, we have (cf. Bracken[41]):

Theorem 4.2: If ψ is a field of Type Ia, and the wave equation (1.1) is locally conformal-invariant on a vector space $\mathscr{U}\subseteq\mathscr{D}$, then the non zero components of any positive-energy (respectively, negative-energy) plane wave solution in $\mathscr{U}$ belong to a direct sum of irreducible representations of sl(2,C) of the type $(m,0)$ or $(0,n)$, with the corresponding length dimensions $(-m-1)$ and $(-n-1)$, and corresponding Lorentz-invariant helicities m and $-n$ [respectively, $-m$ and n]. □

What is the content of the critical Eqs. (3.32) for irreducible $[\lambda,+;0,0]$ fields, or direct sums of such fields for various values of λ? Equations (3.32b) and (3.32d) are satisfied identically. We note that since the only representations (m,n) of sl(2,C) involved here have $mn=0$, then

$$MN=0, \tag{4.5}$$

and Eq. (3.32e) can be written with the help of Eqs. (4.2a), (3.138a), (4.5), and (2.13) as

$$\tau_{\mu\nu}\psi=0, \tag{4.6}$$

with $\tau_{\mu\nu}$ as in Eq. (3.49). But in a representation of the type under consideration, $\tau_{\mu\nu}$ vanishes identically because of the following:

Lemma 4.1: Let $\Sigma_{\mu\nu}$ be basis operator of a finite-dimensional representation $(m,0)$ or $(0,n)$ of sl(2,C). Then the tensor $\tau_{\mu\nu}$, defined as in Eq. (3.49), vanishes identically.

Proof: In the representation $(m,0)$ we have [cf. Eqs. (3.83) and (3.85)]

$$\tilde{\Sigma}_{\mu\nu}=-i\Sigma_{\mu\nu} \tag{4.7}$$

and Eq. (2.69) of Lemma 2.5 becomes

$$\begin{aligned}-i\Sigma_{\mu\nu}\Sigma^{\nu}{}_{\lambda}-\Sigma_{\mu\lambda}&=iC_2g_{\mu\lambda}\\&=im(m+1)g_{\mu\lambda}\\&=\tfrac{1}{2}iC_1g_{\mu\lambda},\end{aligned}$$

i.e.,

$$\tau_{\mu\nu}=0.$$

The argument is similar for the representation $(0,n)$. □

It follws that for fields which correspond to a direct sum of irreducible $[\lambda,+;0,0]$ representations of $\mathscr{W}$, Eqs. (3.32) reduce to (3.32a) and (3.32c), i.e.,

$$P_\mu P^\mu\psi=0, \tag{4.8a}$$

$$\Sigma_{\mu\nu}P^\nu\psi=(i-\Delta)P_\mu\psi\equiv-i(M+N)P_\mu\psi. \tag{4.8b}$$

And furthermore, if the direct sum of fields contains no summand ψ_λ with $\lambda=0$, then Eq. (4.8a) is implied by Eq. (4.8b), since

$$(M+N)\psi_\lambda=(-i\Delta-1)\psi_\lambda=|\lambda|\psi_\lambda, \tag{4.9}$$

and contracting Eq. (4.8b) on the left with P^μ gives

$$-i(M+N)P^\mu P_\mu\psi_\lambda=0. \tag{4.10}$$

We now consider the results of earlier investigations in relation to ours.

A. The scalar field

The index space is one-dimensional in this case, and carries the trivial representation (0,0) of sl(2,C). This can be extended to the nontrivial representation $[0,+;0,0]$ of $\mathscr{W}$, by taking $\kappa_\mu=0$ and $\Delta=+i$. The dimension of the field is then (-1). Eq. (4.8b) is trivial in this case as $\Sigma_{\mu\nu}=0=M=N$. We are left with the single Eq. (4.8a), i.e., the wave equation, in our locally conformal-invariant set.

B. The two- and four-component neutrino equations

Consider the two-component neutrino field χ, with index space carrying the representation $(\frac{1}{2},0)$ of sl(2,C) with basis operators

$$\mathbf{S}=\tfrac{1}{2}\boldsymbol{\sigma},\ \mathbf{T}=-\tfrac{1}{2}i\boldsymbol{\sigma} \tag{4.11}$$

in the notation of Lemma 3.1. Here $\boldsymbol{\sigma}$ are the Pauli matrices. This representation can be extended to the representation $[\frac{1}{2},+;0,0]$ of $\mathscr{W}$, by taking $\kappa_\mu=0$ and $\Delta=3i/2$. Then χ has dimension $(-3/2)$. A locally conformal-invariant set of equations (implying $\Box\chi=0$) is then Eq. (4.8b), which is (since $M=\frac{1}{2}$, $N=0$ here)

$$\Sigma_{\mu\nu}\partial^\nu\chi=-\tfrac{1}{2}i\partial_\mu\chi, \tag{4.12}$$

or equivalently

$$\boldsymbol{\sigma}\cdot\nabla\chi=-\partial_0\chi, \tag{4.13a}$$

$$(\boldsymbol{\sigma}\wedge\nabla+i\boldsymbol{\sigma}\partial_0)\chi=-i\nabla\chi, \tag{4.13b}$$

where

$$\nabla=(\partial_1,\partial_2,\partial_3). \tag{4.14}$$

Eq. (4.13b) is implied by Eq. (4.13a), so we can consider Eq. (4.13a) alone, the Weyl equation, as a locally conformal-invariant equation. It implies that a positive energy field has helicity $(+\frac{1}{2})$.

The case of a two-component field corresponding to the representation $(0,\frac{1}{2})$ of sl(2,C), and $[-\frac{1}{2},+;0,0]$ of $\mathscr{W}$, is similar. Again the field has dimension $(-\frac{3}{2})$. The four-component (Dirac bispinor) neutrino field ψ is the direct sum of these two two-component fields. The appropriate representation of $\mathscr{W}$ is $[\frac{1}{2},+;0,0]\oplus[-\frac{1}{2},+;0,0]$, with basis operators

$$\Sigma_{\mu\nu}=\tfrac{1}{4}i[\gamma_\mu,\gamma_\nu],\quad \kappa_\mu=0,\quad \Delta=(\tfrac{3}{2})i, \tag{4.15}$$

where γ_μ are the Dirac matrices, satisfying

$$\gamma_\mu\gamma_\nu+\gamma_\nu\gamma_\mu=2g_{\mu\nu}. \tag{4.16}$$

Equation (4.8b) reads

$$\tfrac{1}{4}i[\gamma_\mu,\gamma_\nu]\partial^\nu\psi = -\tfrac{1}{2}i\partial_\mu\psi, \tag{4.17}$$

which, with the help of Eq. (4.16) can be reduced to

$$\gamma_\mu(\gamma_\nu\partial^\nu)\psi = 0,$$

or equivalently,

$$\gamma^\nu\partial^\nu\psi = 0. \tag{4.18}$$

The two-component fields are recovered with the use of the projectors

$$P_\pm = \tfrac{1}{2}(1 \pm \gamma_5), \tag{4.19}$$

where

$$\gamma_5 = -i\gamma_0\gamma_1\gamma_2\gamma_3. \tag{4.20}$$

Thus if

$$\psi_\pm = P_\pm\psi, \tag{4.21}$$

then $\psi_\pm$ satisfies

$$\gamma_5\psi_\pm = \pm\psi_\pm. \tag{4.22}$$

and corresponds to the representation $[\pm\frac{1}{2}, + ;0,0]$ of $\mathscr{W}$. This can be seen by evaluating

$$C_1 = \tfrac{1}{2}\Sigma_{\mu\nu}\Sigma^{\mu\nu} = \tfrac{3}{2} \tag{4.23a}$$

$$C_2 = \tfrac{1}{4}i\tilde{\Sigma}_{\mu\nu}\Sigma^{\mu\nu} = \tfrac{1}{4}i(-i\gamma_5\Sigma_{\mu\nu})\Sigma^{\mu\nu} = (\tfrac{3}{4})\gamma_5. \tag{4.23b}$$

A comparison with Eq. (2.13) shows that on $\psi_\pm$, $M = \frac{1}{2}$ and $N = 0$, while on ψ_-, $N = \frac{1}{2}$ and $M = 0$. The fields $\psi_\pm$ have helicity $\pm\frac{1}{2}$ (for positive energy) in accordance with Theorems 3.5 and 4.2.

C. Maxwell's equations for the free electromagnetic field

The index space of the electromagnetic field $F_{\mu\nu}(x)[= -F_{\nu\mu}(x)]$ carries the representation $(1,0)\oplus(0,1)$ of sl(2,C). We can extend this to a representation $[1, + ;0,0] \oplus [-1, + ;0,0]$ of $\mathscr{W}$, by taking $\kappa_\mu = 0$ and $\Delta = 2i$. Then $F_{\mu\nu}$ has dimension (-2). Since $(M+N) = 1$ here, a locally conformal-invariant set of equations (implying $\Box F = 0$) is, from Eq. (4.8b),

$$\Sigma_{\mu\nu}\partial^\nu F = -i\partial_\mu F. \tag{4.24}$$

The sl(2,C) operators act on $F_{\alpha\beta}$ as

$$(\Sigma_{\mu\nu}F)_{\alpha\beta} \equiv (\Sigma_{\mu\nu})_{\alpha\beta}{}^{\rho\sigma}F_{\rho\sigma}, \tag{4.25}$$

where

$$\begin{aligned}-2i(\Sigma_{\mu\nu})_{\alpha\beta}{}^{\rho\sigma} = &(g_{\mu\alpha}\delta_\nu{}^\rho - g_{\nu\alpha}\delta_\mu{}^\rho)\delta_\beta{}^\sigma \\ &+ \delta_\alpha{}^\rho(g_{\mu\beta}\delta_\nu{}^\sigma - g_{\nu\beta}\delta_\mu{}^\sigma) \\ &- (g_{\mu\beta}\delta_\nu{}^\rho - g_{\nu\beta}\delta_\mu{}^\rho)\delta_\alpha{}^\sigma \\ &- \delta_\beta{}^\rho(g_{\mu\alpha}\delta_\nu{}^\sigma - g_{\nu\alpha}\delta_\mu{}^\sigma),\end{aligned} \tag{4.26}$$

and on substituting this expression in Eq. (4.24) and using the antisymmetry of $F_{\mu\nu}$, we get

$$g_{\mu\alpha}\partial^\rho F_{\rho\beta} - g_{\mu\beta}\partial^\rho F_{\rho\alpha} = -\partial_\mu F_{\alpha\beta} - \partial_\alpha F_{\beta\mu} - \partial_\beta F_{\mu\alpha}. \tag{4.27}$$

Contracting both sides with $g_{\mu\alpha}$ we find

$$\partial^\rho F_{\rho\alpha} = 0, \tag{4.28}$$

and Eq. (4.27) then implies also

$$\partial_\mu F_{\alpha\beta} + \partial_\beta F_{\mu\alpha} + \partial_\alpha F_{\beta\mu} = 0. \tag{4.29}$$

Eqs. (4.28) and (4.29) are the free-field Maxwell's equations. They are written in compact form in Eq. (4.24) [or Eq. (4.27)]. Note that by Theorems 3.5 and 4.2, the $[\pm 1, + ;0,0]$ component $F_{\mu\nu}{}^{(\pm)}$ of $F_{\mu\nu}$ satisfies also

$$\tilde{\Sigma}_{\mu\nu}\partial^\nu F^{(\pm)} = \mp\partial_\mu F^{(\pm)}, \tag{4.30}$$

an equation which is also locally conformal-invariant, and which states that the invariant helicity of (positive energy) fields $F_{\mu\nu}{}^{(\pm)}$ is ± 1. It is easily checked that

$$F_{\mu\nu}{}^{(\pm)} = \tfrac{1}{2}(F_{\mu\nu} \mp i\tilde{F}_{\mu\nu}), \tag{4.31}$$

where

$$\tilde{F}_{\mu\nu} = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}. \tag{4.32}$$

Thus $F^{(+)}$ (respectively, $F^{(-)}$) is the right (respectively, left) circularly polarized component of F.

D. The Bargmann–Wigner equations

The index space of the fields ξ used by Bargmann *et al.*[18] to describe massive and massless particles with spin $s(>0)$ is the *symmetrized* tensor product of $2s$ identical, four-dimensional Dirac bispinor spaces, which we may label with $\alpha = 1,2,...,2s$. Let $\gamma_\mu{}^{(\alpha)}$ be the Dirac matrices acting on the αth four-dimensional space. Then for each α, the relations (4.16) are satisfied, and $\gamma_\mu{}^{(\alpha)}$ commutes with $\gamma_\mu{}^{(\beta)}$ if $\alpha \neq \beta$. Introduce also $\gamma_5{}^{(\alpha)}$, $\alpha = 1,2,...,2s$, by analogy with Eq. (4.20).

For massless particles with helicity $+s$, Bargmann *et al.* further required that $\xi\ (=\xi_+$ now) satisfies

$$\gamma_5{}^{(\alpha)}\xi_+ = \xi_+, \quad \alpha = 1,2,...,2s. \tag{4.33}$$

Since the eigenvalues $+1$ and -1, respectively, of $\gamma_5{}^{(\alpha)}$ label the representations $(\frac{1}{2},0)$ and $(0,\frac{1}{2})$ of sl(2,C) carried by the αth factor space, it follows that the index space of ξ_+ carries the symmetrized tensor product of the representation $(\frac{1}{2},0)$ with itself ($2s$) times. This is the representation $(s,0)$. Similarly we may introduce ξ_- satisfying

$$\gamma_5{}^{(\alpha)}\xi_-, \quad \alpha = 1,2,...,2s, \tag{4.34}$$

and associated with the representation $(0,s)$ of sl(2,C).

The sl(2,C) basis operators in both cases are (restrictions of)

$$\begin{aligned}\Sigma_{\mu\nu} &= \frac{1}{4}i\sum_{\alpha=1}^{2s}[\gamma_\mu{}^{(\alpha)},\gamma_\nu{}^{(\alpha)}] \\ &= -isg_{\mu\nu} + \frac{1}{2}i\sum_{\alpha=1}^{2s}\gamma_\mu{}^{(\alpha)}\gamma_\nu{}^{(\alpha)},\end{aligned} \tag{4.35}$$

so that

$$C_1 = \tfrac{1}{2}\Sigma_{\mu\nu}\Sigma^{\mu\nu} = 2s(s+1) - \frac{1}{4}\sum_{\alpha<\beta=1}^{2s}(\gamma_\mu{}^{(\alpha)}\gamma^{\mu(\beta)})^2. \tag{4.36}$$

According to Eqs. (2.8), on the representation $(s,0)$ or $(0,s)$,

$$C_1 = 2s(s+1). \tag{4.37}$$

It follows that

$$\sum_{\alpha<\beta=1}^{2s}(\gamma_\mu{}^{(\alpha)}\gamma^{\mu(\beta)})^2\xi_\pm = 0. \tag{4.38}$$

Now $\gamma_0^{(\alpha)}$ and $i\gamma_j^{(\alpha)}$ ($j = 1,2,3$) can be taken to be Hermitian, without loss of generality, for each value of α. Thus $(\gamma_\mu^{(\alpha)}\gamma^{\mu(\beta)})$ is Hermitian. It then follows from Eq. (4.38) that

$$(\gamma_\mu^{(\alpha)}\gamma^{\mu(\beta)})\xi_\pm = 0 \quad , \quad \alpha \neq \beta. \tag{4.39}$$

Conversely, it can easily be seen that if ξ satisfies Eqs. (4.39), then it belongs to that part of the tensor-product space assoicated with the representation $(s,0) \oplus (0,s)$ of sl(2,C). Eqs. (4.39), and equivalently the symmetrization conditions and Eqs. (4.33) and (4.34) of Bargmann *et al*, are not to be thought of as dynamical conditions, but rather as statements defining the index space of the fields to be used to describe massless particles. One could, of course, start with $2(2s + 1)$-component fields corresponding to this representation of sl(2,C), but the advantage of the approach used by Bargmann *et al.*—introducing redundant components and then imposing conditions which set them to zero—is simply that one can employ the familiar algebra of the Dirac matrices.

The representation $(s,0) \oplus (0,s)$ can be extended to the representation $[s, + ;0,0] \oplus [-s, + ;0,0]$ of $\mathscr{W}$, by setting $\kappa_\mu = 0$ and $\Delta = +i(s+1)$. Then ξ has the canonical dimension $(-s-1)$. Since $(M + N) = s$ here, a locally conformal-invariant set of equations for ξ (implying $\Box\xi = 0$) is then, from Eq. (4.8b),

$$\Sigma_{\mu\nu}\partial^\nu\xi = -is\xi. \tag{4.40}$$

Substituting for $\Sigma_{\mu\nu}$ from Eq. (4.35), we get

$$\sum_{\alpha=1}^{2s} \gamma_\mu^{(\alpha)}(\gamma_\nu^{(\alpha)}\partial^\nu)\xi = 0. \tag{4.41}$$

Contracting on the left with $\gamma^{\mu(\beta)}$, using the commutation and anticommutation relations between the $\gamma_\rho^{(\delta)}$, and noting Eqs. (4.39), we get

$$(\gamma_\mu^{(\beta)}\partial^\mu)\xi = 0, \quad \beta = 1,2,...,2s. \tag{4.42}$$

Conversely, if Eqs. (4.42) hold, then so do Eqs. (4.41) and hence Eqs. (4.40). Thus the locally conformal-invariant Eq. (4.8b) is in this case equivalent to Eqs. (4.42), which are the Bargmann–Wigner equations.[18] The component $\xi_\pm$ corresponding to the representation $[\pm s, + ;0,0]$ of $\mathscr{W}$ can be obtained as

$$\xi_\pm = \prod_{\alpha=1}^{2s} [\tfrac{1}{2}(1 \pm \gamma_5^{(\alpha)})]\xi. \tag{4.43}$$

It satisfies the locally conformal-invariant equation (3.141) with $\lambda = \pm s$ [and also Eqs. (4.33) or (4.34)], and so (for positive energy) has helicity $\pm s$ as expected.

The Eqs. (4.40), where the index space of ξ carries the representation $(s,0)$, $(0,s)$, or $(s,0) \oplus (0,s)$ of sl(2,C), were considered before the work of Bargmann *et al.* by Dirac,[15] Fierz,[16] and Gårding,[14] using the dotted–undotted spinor formalism. The complete equivalence of these different ways of writing *the same equations* must be emphasized. Those sections in other works,[19–23,26] concerned only with showing the conformal invariance of these equations, were repeating in different formalisms part of the work of Gross.[17] McLennan[13] had previously shown the local invariance of these same equations.[63] Note also that for $s = \frac{1}{2}$ and $s = 1$ the Bargmann–Wigner equations are completely equivalent to the neutrino equations, and Maxwell's equations, respectively, as can be seen from our discussion above.

When we write all these equations in the forms (4.12), (4.17), (4.24), and (4.40), we see most clearly that they belong to one family—*the* family of conformal-invariant equations for Type Ia fields.

E. Errors in the work of McLennan and Post

McLennan[13] claimed to prove the conformal invariance of certain sets of field equations described by Gårding.[14] In these papers the dotted-undotted spinor formalism is used. The index space of a field with p undotted indices and q dotted ones (p and q are non-negative integers), separately symmetric in each set, carries the irreducible representation $(\frac{1}{2}p,\frac{1}{2}q)$ of sl(2,C) in our notation. In particular, fields ψ corresponding to the representation $(\frac{1}{2}p,\frac{1}{2}q) \oplus (\frac{1}{2}q,\frac{1}{2}p)$ [with $p \neq q$] are considered, together with first-order field equations [McLennan's Eqs. (3.19)] which imply that the wave equation (1.1) is satisfied. According to our results above, these equations can not be locally conformal invariant unless $pq = 0$. This contradicts a claim made by McLennan, but it is easy to find an error in his analysis. He supposes [see his Eq. (6.4)] that under a special conformal transformation, a component of the field corresponding to the representation $(\frac{1}{2}p, \frac{1}{2}q)$ transforms in such a way that its p undotted indices are not affected. Similarly, for a component corresponding to $(\frac{1}{2}q, \frac{1}{2}p)$, the p dotted indices are not affected. But such transformation laws are not consistent with the structure of the Lie algebra of the conformal group, for an infinitesimal translation does not affect spinor indices, but the commutator of our infinitesimal special conformal transformation along one spatial axis, and an infinitesimal translation along another, is an infinitesimal rotation about the third [cf. Eq. (2.6i)], and so affects all dotted and undotted indices. Therefore, an infinitesimal special conformal transformation must in general also affect all dotted and undotted indices. McLennan's proposed transformation law is not consistent if $p \neq 0$.

In claiming to deduce the conformal invariance of equations satisfied by fields with $p = q$ and zero helicity (such fields can also be thought of as symmetric, traceless, tensor fields $\varphi_{\mu\nu...\rho}$ with p indices), McLennan merely remarked that such sets of equations "are equivalent to the scalar or pseudo-scalar wave equation" (1.1), which is conformal invariant. In fact one can show that[42]

$$\varphi_{\mu\nu\cdots\rho} = \partial_\mu\partial_\nu\cdots\partial_\rho\varphi. \tag{4.44}$$

where φ satisfies Eq. (1.1). However, the conformal invariance of Eq. (1.1) for φ does not ensure the invariance of the equations satisfied by $\varphi_{\mu\nu\cdots\rho}$ defined as in Eq. (4.44), and in fact our results imply that these equations are not invariant. The index-space representation of sl(2,C) associated with this tensor field is $(\frac{1}{2}p, \frac{1}{2}p)$. This can be extended to a representation of $\mathscr{W}$ only by setting $\kappa_\mu = 0$ and Δ equal to a constant, so that the field is in particular of Type Ia. But then Theorem 4.2 shows that the wave equation is not locally invariant on such a field, if $p \neq 0$. The reason for this breakdown of conformal invariance in the passage from φ to $\varphi_{\mu\nu\cdots\rho}$ is easily seen—the operators $\partial_\mu, \partial_\nu \cdots$ in Eq. (4.44) are Lorentz-covariant but not conformal-covariant objects.

More recently Post[26] considered free, massless, positive-energy fields $\psi^{(m,n)}(x)$ whose index space carries an irreducible representation (m,n) of sl(2,C), and which have Lorentz-invariant helicity $\lambda = (m - n)$ [cf. Lemma 3.2]. He claimed to prove that the equations satisfied by such fields, including the wave equation (1.1), are conformal-invariant, even if $mn \neq 0$. This contradicts our results, and indeed, the result given by one of us[41] before Post's work appeared. His proof is incorrect, and depends crucially on a result attributed to Mack *et al.*[22] [See the paragraph following Post's Eq. (5.11).] This result, which is in fact invalid, was not proved in Ref. 22, though its validity was implied there. The result in question can be described as follows.

A Hilbert space of the fields $\psi^{(m,n)}$ can be defined, carrying the unitary, irreducible representation of ISL (2,C) appropriate to a massless "particle" with positive energy and helicity $(m - n)$. This representation extends to a unitary irreducible representation of SU(2,2), with self-adjoint generators $P'_\lambda, K'_\lambda, D'$, and $M'_{\mu\nu}$ satisfying, on a suitable domain, the commutation relations (2.6). Then these operators can be identified on the Hilbert space with the generators (2.5) of conformal transformations for these fields, after appropriate choices for κ_μ and Δ are made.

Mack and Todorov showed that this is so if $mn = 0$, but they did not consider directly the cases with $mn \neq 0$. Instead they quoted a result of Weinberg,[42] who showed that if a free, massless positive-energy field χ corresponds to an irreducible index space representation (m,n) of sl(2,C) with $m - n = \lambda$, and has Lorentz-invariant helicity λ, then χ is a linear combination of the rth partial derivatives with respect to the variables x^μ, of a field ξ which also has invariant helicity λ. If $\lambda \geqslant 0$, then ξ corresponds to an index-space representation $(\lambda,0)$, and $r = 2n$. If $\lambda < 0$, then ξ corresponds to $(0, -\lambda)$, and $r = 2m$. On this basis, Mack and Todorov concluded that they could restrict their attention to the cases with $mn = 0$, in order to prove the desired result for the operators $P'_\lambda, K'_\lambda, D'$, and $M'_{\mu\nu}$. However, as remarked in the Introduction, and as implied by Theorem 4.2, the result in question is not valid if $mn \neq 0$. In fact one finds that the operators K'_λ in these cases, unlike the K_λ of Eqs. (2.5), are nonlocal. The reason for this breakdown of conformal invariance, in the context of Weinberg's result, is again that the operator ∂_μ relating massless fields with $mn = 0$ to ones with $mn \neq 0$ [cf. Eq. (4.44)] is not conformal covariant. Essentially the same misunderstanding of this point led McLennan into error, as noted above.

F. Other related works

Several authors[36–40,43] have considered the conditions to be satisfied if classical field equations derivable from an action principle are to be conformal invariant. However, they have not been concerned with the specific situation where the wave equation (1.1) is required to be one of the field equations obtained. The conditions obtained are accordingly much less specific than ours. (In another sense, they are more specific, since it is not clear which of the sets of field equations we have described are derivable from an action principle.) Furthermore, these works have concentrated on fields of Type Ia.

The conformal invariance (in a weaker sense) of wave equations for massive particles has been considered by other authors.[5,11,24,54,64] Because the taking of the zero mass limit is a nontrivial matter, particularly in the context of conformal invariance,[65] it is not clear how the results obtained in these works relate to ours.

The conditions under which Lorentz-invariant equations of the form (1.2) are also conformal invariant have been analyzed by Kotecky *et al.*[30] But again, because they did not specifically require that Eq. (1.1) should follow from Eq. (1.2), their results are not easily related to ours. They did relate their results to some extent with those of McLennan,[13] but did not detect any errors in that work. Only fields of Type Ia appear in the results of Kotecky *et al.* One reason for this is easily seen. If fields of Type Ib are involved, then one has a four-vector operator κ_μ acting on the index space, and having scale dimension $(+1)$. Then as well as equations of the forms (1.2), field equations of the form

$$L_\mu \partial^\mu \psi = A\psi \tag{4.45}$$

must also be considered, where A is a dimensionless matrix. Equation (3.54) provides an example. Massless wave equations of the general form (4.45) have appeared in a more general context in the work of Wightman.[66] Let us remark also that for field equations of the form (1.2), (4.45) where L_μ is rectangular, an important and nontrivial constraint [cf. Theorem 3.4], not considered by Kotecky *et al.*, is that the equation should admit plane wave solutions.

Fields of Type Ib have received comparatively little attention in the literature. Ciccariello and Sartori[52] (see also Ferrara *et al.*,[53] and Lopuszanski and Oziewicz[25]) considered fields of Type Ib and associated conformal-invariant wave equations, but once again, their aims were different from ours, and their results and ours are not easily related. Lopuszanski *et al.* did note the appearance of conformal-invariant equations of the form (4.40) for fields of Type Ia, as one of us had done earlier.[41] (See also Seetharaman.[46])

Since the Lie algebra $\mathscr{W}$ is a subalgebra of su(2,2), any finite-dimensional representation of the latter defines a representation of the former. Mack *et al.*[43] have considered fields of Type Ib generated in this way. But it must be emphasized that only a limited class of representations of $\mathscr{W}$, and consequently, only a limited class of field types, can be obtained in this way. There is a countable number of inequivalent, finite-dimensional representations for su(2,2), but an uncountable number for $\mathscr{W}$ and representations of $\mathscr{W}$ in which Δ is not diagonalizable [cf. Eqs. (2.17)] are not contained in representations of su(2,2).[67]

Dirac[9] and Hepner[51] (see also Mack *et al.*[43] and Budini[29]) have considered the particular case of Dirac spinors $\psi(x)$ and the associated four-dimensional representations of su(2,2) with [cf. Eqs. (2.5) and (2.6)]

$$\begin{aligned} p_\mu &= \tfrac{1}{2}\kappa^{-1}(1 \pm \gamma_5)\gamma_\mu, \quad m_{\mu\nu} = \tfrac{1}{4}i[\gamma_\mu, \gamma_\nu], \\ d &= \mp \tfrac{1}{2}i\gamma_5, \quad k_\mu = \tfrac{1}{2}\kappa(1 \mp \gamma_5)\gamma_\mu. \end{aligned} \tag{4.46}$$

Here the Dirac matrices are as in Sec. 4.2, and κ is a nonzero constant with dimensions of length. (Representations with different values of κ are equivalent, so this value has no physical significance.) Then one may take for the generators of conformal transformations of $\psi(x)$

$$P_\mu = i\partial_\mu + p_\mu, \quad M_{\mu\nu} = x_\mu P_\nu - x_\nu P_\mu + m_{\mu\nu},$$
$$D = x^\mu P_\mu + in + d, \quad K_\mu = 2x_\mu(x^\nu P_\nu + in) - x^\nu x_\nu P_\mu + k_\mu, \qquad (4.47)$$

where n is a constant. These operators satisfy the relations (2.6), but are not of the form (2.5). However, by a similarity transformation[43]

$$\psi(x) \to \exp(-ix^\nu p_\nu)\psi(x),$$
$$P_\mu \to \exp(-ix^\nu p_\nu)\cdot P_\mu \cdot \exp(ix^\nu p_\nu), \qquad (4.48)$$

etc., one can bring them to the form (2.5), with

$$\kappa_\mu = \tfrac{1}{2}\kappa(1 \mp \gamma_5), \quad \Delta = in \mp \tfrac{1}{2}i\gamma_5$$
$$\Sigma_{\mu\nu} = \tfrac{1}{4}i[\gamma_\mu, \gamma_\nu]. \qquad (4.49)$$

These operators (4.49) span a representation, $D_{n\mp}$ say, of $\mathscr{W}$, which is *not* a $[\lambda, +;0,u]$ representation for any λ, u. However, the representation D_{1+}, for example, is indecomposable but not irreducible, and contains the representation $[\frac{1}{2}, +;0,0]$ as an invariant subrepresentation, associated with the subspace of spinors on which $\gamma_5 = +1$. Accordingly, the equation (1.1) is then locally conformal invariant provided Eqs. (3.32) hold, and here they reduce to

$$(\gamma_\mu P^\mu)\psi = 0, \qquad (4.50a)$$
$$\gamma_5\psi = \psi. \qquad (4.50b)$$

This is an example of the type of behavior whose possibility was indicated in Comment 1, following Theorem 3.3. In the present example, so long as we are concerned only with free massless fields, there is no real loss of generality if we restrict our attention to spinors for which Eq. (4.50b) is satisfied identically—i.e., essentially two-component spinors corresponding to the representation $[\frac{1}{2}, +;0,0]$ of $\mathscr{W}$ [cf. Sec. 4.2].

On the other hand, the equations

$$\gamma^\mu P_\mu \psi = 0 \qquad (4.51a)$$

and

$$\gamma_5\psi = -\psi \qquad (4.51b)$$

are *not* conformal invariant if we adopt the representation D_{1+} for ψ, since they are not consistent with Eqs. (3.32). [The roles of the equations (4.50b) and (4.51b) are interchanged if we consider instead the representation D_{1-} for ψ.] The situation here is to be contrasted with that in Sec. 4.2, where the representation $[\frac{1}{2}, +;0,0] \oplus [-\frac{1}{2}, +;0,0]$ of $\mathscr{W}$ was adopted for ψ, and both sets of equations, (4.50) *and* (4.51) are conformal invariant. When we vary the relevant representation of $\mathscr{W}$ on Dirac spinors, we are really changing the field type, and when we talk about conformal invariance or noninvariance of equations like (4.50) of (4.51) we must be clear as to what type of fields we are considering. Failure to do so seems to have led to some confusion in the literature.[68,69] In particular, we should not confuse the results described here for spinors corresponding to the representations D_{1+}, D_{1-}, or $([\frac{1}{2}, +;0,0] \oplus [-\frac{1}{2}, +;0,0])$ of $\mathscr{W}$ with the result implied by Dirac[9] (see also Budini[29] and Castell[68]) that the equation

$$(1 \pm \gamma_5)\gamma^\mu P_\mu \psi = 0 \qquad (4.52)$$

is conformal invariant if ψ corresponds to the representation $D_{2\pm}$ of $\mathscr{W}$. Eq. (4.52) does not imply Eq. (1.1), so $\psi(x)$ is not a massless field according to our definition, and our general results are not directly relevant to this case.

5. CONCLUDING REMARKS

We have derived the conditions under which the wave equation (1.1) is locally conformal invariant, and have seen as a result that although some well-known sets of massless wave equations for fields of Type Ia are invariant, many others are not. Indeed, it is fair to say that most massless wave equations for fields of this type are not conformal invariant. In particular,[41] Eq. (1.1) is not invariant if the index space of ψ carries an irreducible representation (m,n) of sl$(2,C)$ with $mn \neq 0$.

Most generally, we have shown that only $[\lambda, +;0,u]$ fields are of direct interest in the discussion of locally-invariant wave equations, and that these always carry Lorentz-invariant helicity λ (for positive-energy plane waves). For $u > 0$, these fields are of Type Ib. In subsequent papers, we shall describe the representations $[\lambda, +;0,u]$ of $\mathscr{W}$ completely, and also examine in detail the consequences of Eqs. (3.32) for such fields, thus completing our analysis.

ACKNOWLEDGMENTS

It is a pleasure to thank A. O. Barut, M. Eastwood, J. Harnad, A. S. Wightman, and B. Xu for helpful conversations.

[1] H. A. Kastrup, Ann. Phys. (Leipzig) **9**, 388-428 (1962).
[2] T. Fulton, F. Rohrlich, and L. Witten, Rev. Mod. Phys. **34**, 442-457 (1962).
[3] A. O. Barut, "Introduction to de Sitter and Conformal Groups" in *Lectures in Theoretical Physics*, Vol. 13, edited by A. O. Barut and W. E. Brittin (Colorado Assoc. U. P., Boulder, Colorado, 1971).
[4] S. Ferrara, R. Gatto, and A. F. Grillo, "Conformal Algebra in Space-Time and Operator Product Expansion," *Springer Tracts in Modern Physics*, Vol. 67 (Springer, Berlin, 1973).
[5] F. Bayen "Conformal Invariance in Physics," in *Differential Geometry and Relativity*, edited by M. Cahen and M. Flato (Reidel, Dordrecht, 1976).
[6] Some of the works listed deal also with global invariance. In the present work, only local properties of the conformal group are involved, and all arguments are purely algebraic.
[7] H. Bateman, Proc. Lond. Math. Soc. (2) **8**, 223-264 (1910).
[8] E. Cunningham, *ibid.* **8**, 77-98 (1910).
[9] P. A. M. Dirac, Ann. of Math. **37**, 429-442 (1936).
[10] J. Schouten and J. Haantjes, Proc. Kon. Akad. Wet. **39**, 1059-1065 (1936).
[11] W. Pauli, Helv. Phys. Acta **13**, 204-208 (1940).
[12] S. A. Bludman, Phys. Rev. **107**, 1163-1168 (1957).
[13] J. A. McLennan, Nuovo Cimento (10) **3**, 1360-1379 (1956).
[14] L. Gårding, Proc. Cambridge Phil. Soc. **41**, 49-56 (1945).
[15] P. A. M. Dirac, Proc. R. Soc. London A **155**, 447-459 (1936).
[16] M. Fierz, Helv. Phys. Acta **12**, 3-37 (1939).
[17] L. Gross, J. Math. Phys. **5**, 687-695 (1964).
[18] V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. (USA) **34**, 211-223 (1946).
[19] J. S. Lomont, Nuovo Cimento (10) **22**, 673-679 (1961).
[20] R. Penrose, Proc. R. Soc. London A **284**, 159-203 (1965).
[21] B. Kursunoglu, J. Math. Phys. **8**, 1694-1699 (1967).
[22] G. Mack and I. T. Todorov, J. Math. Phys. **10**, 2078-2085 (1969).
[23] F. Bayen, Nuovo Cimento (10) **68 A**, 39-48 (1970).

[24]A. O. Barut and R. B. Haugen, Nuovo Cimento (11) **18A**, 511-531 (1973).
[25]J. T. Lopuszanski and Z. Oziewicz, Rep. Math. Phys. **8**, 241-254 (1975).
[26]G. Post, J. Math. Phys. **17**, 24-32 (1976).
[27]H. D. Fegan, Q. J. Math. Oxford (2) **27**, 371-378 (1976).
[28]H. P. Jakobsen and M. Vergne, J. Funct. Anal. **24**, 52-106 (1977).
[29]P. Budini, Czech. J. Phys. **B29**, 6-21 (1979).
[30]R. Kotecky and J. Niederle, Rep. Math. Phys. **12**, 237-249 (1977); see also J. Mickelsson and J. Niederle, Ann. Inst. H. Poincaré Sec. A **23**, 277-295 (1975).
[31]The helicity of a potential is not defined by the imposition of manifestly covariant equations. For a general discussion, see A. McKerrell, Ann. Phys. **40**, 237-267 (1966).
[32]J. Wess, Nuovo Cimento (10) **18**, 1086-1107 (1960).
[33]R. Jackiw, Phys. Rev. Letts. **41**, 1635-1638 (1978).
[34]S. K. Bose and R. Parker, J. Math. Phys. **10**, 812-813 (1969).
[35]W. Ruhl, Commun. Math. Phys. **34**, 149-166 (1973).
[36]D. J. Gross and J. Wess, Phys. Rev. D **2**, 753-764 (1970).
[37]C. G. Callan, S. Coleman, and R. Jackiw, Ann. Phys. **59**, 42-73 (1970).
[38]M. Flato, J. Simon, and D. Sternheimer, Ann. Phys. **61**, 78-97 (1970).
[39]S. Browne, Proc. R. Ir. Acad. A **74**, 49-59 (1974).
[40]P. Budini, P. Furlan, and R. Raczka, Nuovo Cimento (11) **52** A, 191-246 (1979).
[41]A. J. Bracken, Lett. Nuovo Cimento (2) **2**, 574-576 (1971).
[42]S. Weinberg, Phys. Rev. **138**, B988-1002 (1965); also, in *Lectures on Particles and Field Theory*, Vol. 2, edited by S. Deser and K. W. Ford (Prentice-Hall, New Jersey, 1965).
[43]G. Mack and A. Salam, Ann. Phys. **53**, 174-202 (1969).
[44]K. H. Mariwalla, Lett. Nuovo Cimento (2) **4**, 295-297 (1972).
[45]J. Muggli, Helv. Phys. Acta **46**, 253-272 (1973).
[46]M. Seetharaman, Prog. Theor. Phys. **49**, 1685-1688 (1973).
[47]C. Fronsdal, Phys. Rev. D **18**, 3624-3629 (1978).
[48]J. Fang and C. Fronsdal, Phys. Rev. D **18**, 3630-3633 (1978).
[49]F. A. Berends, J. W. van Holten, B. de Wit, and P. Van Nieuwenhuizen, J. Phys. A **13**, 1643-1649 (1980).
[50]T. Curtright, Phys. Lett. B **85**, 219-224 (1979).
[51]W. A. Hepner, Nuovo Cimento (10) **26**, 351-368 (1962).
[52]S. Ciccariello and G. Sartori, *ibid.* (11) **19** A, 470-500 (1974).
[53]S. Ferrara, R. Gatto, A. F. Grillo, and G. Parisi, "General Consequences of Conformal Algebra," in *Scale and Conformal Symmetry in Hadron Physics*, edited by R. Gatto (Wiley, New York, 1973).
[54]R. Kotecky and J. Niederle, Czech. J. Phys. B **25**, 123-149 (1975).
[55]This is the Lie algebra associated with those of the transformations (2.1) leaving invariant a fixed point in Minkowski space. Representations and corresponding fields will be called of Type Ia if $\kappa_\mu = 0$, and Type Ib otherwise.[43]
[56]I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon, London, 1963).
[57]G. F. Dell'Antonio, Nuovo Cimento (11) **12** A, 756-762 (1972).
[58]Here $\mathscr{Q}$ stands for quartic, a reference to Eq. (2.18c).
[59]Note that Eq. (2.18c) is now redundant, being implied by Eq. (2.60b).
[60]A. J. Bracken and H. S. Green, J. Math. Phys. **12**, 2099-2106 (1971).
[61]One could consider weaker equations such as $\partial_\mu \Box\psi = 0$, $\Box\Box\psi = 0,\ldots$. The conditions for local conformal invariance may be expected to be weaker in such cases.
[62]A. O. Barut and A. Bohm, J. Math. Phys. **11**, 2938-2945 (1970).
[63]We detected no errors in this part of McLennan's work.
[64]H. A. Buchdal, Nuovo Cimento (10) **11**, 496-506 (1959).
[65]Y. Takahashi, Phys. Rev. D **3**, 622-625 (1971).
[66]A. S. Wightman, "Invariant Wave Equations; General Theory and Applications to the External Field Problem," in *Lecture Notes in Physics*, Vol. 73 (Springer, Berlin, 1978).
[67]We have not considered the possible imbedding of the representations $[\lambda, + ;0,u]$ of $\mathscr{W}$ in representations of su(2,2).
[68]L. Castell, Nuovo Cimento (11) **42** A, 160-164 (1977).
[69]A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (Polish Sci. Publ., Warsaw, 1977), p. 419.

Local conformal invariance of the wave equation for finite-component fields. II. Classification of relevant indecomposable fields

A. J. Bracken [a)]
Department of Physics, University of Colorado, Boulder, Colorado 80302
Barry Jessup [b)]
Department of Mathematics, University of Queensland, St. Lucia, 4067, Queensland, Australia

(Received 28 April 1981; accepted for publication 20 July 1981)

It has been shown in Part I that the requirement of local conformal invariance of the wave equation for finite-component fields focuses attention on fields whose index spaces carry a certain type of finite-dimensional, indecomposable representation of the nonsemisimple Lie algebra $((k_4 \oplus d) \oplus \mathrm{sl}(2, C))$. All representations of this type are here described in complete detail, in each case in an sl(2, C) basis. Although indecomposable, these representations are in general not fully reducible.

PACS numbers: 11.10.Qn, 11.30.Ly, 02.20. + b

I. INTRODUCTION

In an earlier work[1] (henceforth referred to as BJ1), we have considered the conditions for local conformal invariance of the wave equation

$$\Box\psi(x) = 0, \quad x = (x^\mu), \quad \mu = 0, 1, 2, 3, \tag{1.1}$$

where ψ is a classical field with some fixed, finite number of complex-valued components. The index space of this field is assumed to carry a corresponding finite-dimensional representation of the Lie algebra

$$\mathscr{W} = ((k_4 \oplus d) \oplus \mathrm{sl}(2, C)), \tag{1.2}$$

with basis operators κ_μ, Δ, and $\Sigma_{\mu\nu} (= -\Sigma_{\nu\mu})$ satisfying the commutation relations

$$i[\Sigma_{\mu\nu}, \Sigma_{\rho\sigma}] = g_{\mu\rho}\Sigma_{\nu\sigma} + g_{\nu\sigma}\Sigma_{\mu\rho} - g_{\nu\rho}\Sigma_{\mu\sigma} - g_{\mu\sigma}\Sigma_{\nu\rho}, \tag{1.3a}$$

$$i[\kappa_\mu, \Sigma_{\nu\rho}] = g_{\mu\rho}\kappa_\nu - g_{\mu\nu}\kappa_\rho, \tag{1.3b}$$

$$[\kappa_\mu, \kappa_\nu] = 0, \tag{1.3c}$$

$$[\Sigma_{\mu\nu}, \Delta] = 0, \tag{1.3d}$$

$$i[\kappa_\mu, \Delta] = \kappa_\mu. \tag{1.3e}$$

Only if this assumption is made[2] can one define, for an arbitrary infinitesimal conformal transformation of space-time, an appropriate cotransformation law for the field ψ. The generators of infinitesimal conformal transformations of ψ then take the forms

homogeneous Lorentz transformations:	$i(x_\mu\partial_\nu - x_\nu\partial_\mu) + \Sigma_{\mu\nu}$,	(1.4a)
space-time translations:	$i\partial_\mu$,	(1.4b)
dilations:	$ix^\mu\partial_\mu + \Delta$,	(1.4c)
special conformal transformations:	$2ix_\mu x^\nu\partial_\nu + 2x_\mu\Delta - ix^\nu x_\nu\partial_\mu + 2\Sigma_{\mu\nu}x^\nu + \kappa_\mu$,	(1.4d)

and satisfy on suitably smooth ψ the commutation relations appropriate to the Lie algebra of the conformal group. The significance of $\mathscr{W}$ in this connection stems from the fact that it is the Lie subalgebra associated with those conformal transformations which leave invariant the point $x = 0$, *viz.* those composed of homogeneous Lorentz transformations, dilations, and special conformal transformations. (An isomorphic subalgebra is associate with the dilation group, composed of homogeneous Lorentz transformations, dilations, and space-time translations, and is of independent interest. The dilation group, like the conformal group itself, has been discussed as a possible approximate space-time symmetry group in particle physics. In that context, however, the main interest is in infinite-dimensional representations[3] of $\mathscr{W}$.)

The problem of classifying all finite-component field types having inequivalent cotransformation laws for infinitesimal conformal transformations, is seen to correspond to the problem of classifying all inequivalent finite-dimensional representations of $\mathscr{W}$. Such representations have been called[2] of type I, as distinct from infinite-dimensional (type II) representations. More particularly, a finite-dimensional representation and corresponding field is called of type Ia if the associated basis operators κ_μ vanish identically, and of type Ib otherwise. The Lie algebra $\mathscr{W}$ is not semisimple, and its representations of type I or II are not in general fully reducible. The problem of classifying all inequivalent representations of type Ib in particular seems quite beyond our present powers.

In BJ1, we have defined the wave equation (1.1) to be locally conformal-invariant on a vector space $\mathscr{U}$ of smooth solutions, if $\mathscr{U}$ is invariant under the action of the conformal algebra (1.4). Then we have shown that the non-zero components of any ψ in such a $\mathscr{U}$ must belong to a representation of $\mathscr{W}$ from a certain class $\mathscr{Q}$, characterized by the property that the basis operators of any representation from this class satisfy the $\mathscr{W}$-invariant set of equations

$$\kappa_\mu\kappa^\mu = 0, \tag{1.5a}$$

$$\Sigma_{\mu\nu}\kappa^\nu = (\Delta + i)\kappa_\mu, \tag{1.5b}$$

$$\Delta^4 + (C_1 + 1)\Delta^2 + (C_2)^2 = 0, \tag{1.5c}$$

[a)]Present address: Department of Mathematics, University of Queensland, St. Lucia 4067, Queensland, Australia.
[b)]Present address: Department of Mathematics, University of Toronto, Toronto, Ontario, M5S1A7, Canada

where

$$C_1 = \tfrac{1}{2}\Sigma_{\mu\nu}\Sigma^{\mu\nu}, \tag{1.6a}$$

$$C_2 = (1/8)i\epsilon^{\mu\nu\rho\sigma}\Sigma_{\mu\nu}\Sigma_{\rho\sigma}. \tag{1.6b}$$

(The metric tensor is diagonal with $g_{00} = -g_{11} = -g_{22} = -g_{33} = +1$, and the alternating tensor has $\epsilon^{0123} = +1$.) Accordingly, we have restricted our attention to indecomposable (but not necessarily irreducible) representations of $\mathscr{W}$ in Class $\mathscr{D}$, and associated indecomposable Class $\mathscr{D}$ fields.

We have shown furthermore that if (a) ψ is an indecomposable Class $\mathscr{D}$ field, (b) a locally conformal-invariant vector space $\mathscr{U}$ of solutions of the wave equation (1.1) does exist, and (c) at least one of the solutions in $\mathscr{U}$ is a plane wave, then the associated indecomposable Class $\mathscr{D}$ representation of $\mathscr{W}$ must be, for some integer 2λ and non-negative integer u, a representation of the type we have labeled $[\lambda, +; 0, u]$. Since we are interested primarily in the possibility of using locally conformal-invariant spaces of solutions of Eq. (1.1) in the description of free massless particles, the condition (c) is important, and our attention has therefore been limited further, to indecomposable $[\lambda, +; 0, u]$-representations and fields.

The basis operators of such a representation satisfy, by definition, certain conditions additional to (1.5). In order to be able to describe these conditions, we must first recall that every finite-dimensional representation of $\mathscr{W}$ must be fully reducible when regarded as a representation of the sl(2, C) subalgebra associated with the basis operators $\Sigma_{\mu\nu}$. Let (m, n) label the $(2m+1)(2n+1)$-dimensional irreducible representation[4] of sl(2, C), where $2m$ and $2n$ are non-negative integers, associated with eigenvalues $2[m(m+1) + n(n+1)]$ and $[m(m+1) - n(n+1)]$ of the sl(2, C)-invariants C_1 and C_2, respectively, of Eqs. (1.6). An arbitrary finite-dimensional representation of $\mathscr{W}$ must decompose into a direct sum of such representations (m, n), with various values of m and n, and various multiplicities. The operators C_1 and C_2 in such a representation of $\mathscr{W}$ will therefore have the form

$$C_1 = 2M(M+1) + 2N(N+1), \tag{1.7a}$$

$$C_2 = M(M+1) - N(N+1), \tag{1.7b}$$

where M and N are non-negative, simultaneously diagonalizable, sl(2, C) scalar operators whose eigenvalues are non-negative integers or semi-integers. On that subspace of the representation space for $\mathscr{W}$ which is associated with the totality of irreducible representations (m, n) of sl(2, C) for fixed m and n, M and N have the eigenvalues m and n, respectively.

The additional defining properties of a $[\lambda, +; 0, u]$-representation of $\mathscr{W}$ are then

$$\text{(i) } \Delta = i(M + N + 1), \tag{1.8a}$$

implying in particular that Δ is diagonalizable,

$$\text{(ii) } M - N = \lambda, \tag{1.8b}$$

(iii) the eigenvalues of $-i\Delta$ are exactly the set of numbers

$$\{|\lambda| + 1,\quad |\lambda| + 2, \cdots,\quad |\lambda| + u + 1\}. \tag{1.8c}$$

The conditions (i) and (ii) taken together are stronger than, and imply condition (1.5c), as can be seen with the help of Eqs. (1.7). Thus the independent conditions characterizing a $[\lambda, +; 0, u]$-representation are Eqs. (1.5a), (1.5b) and conditions (i)–(iii) above.

In BJ1 we have shown that a $[\lambda, +; 0, u]$-representation is of type Ia if and only if $u = 0$. We have shown also that for each integer 2λ there exists, up to equivalence, exactly one indecomposable $[\lambda, +; 0, 0]$-representation. It is in fact irreducible, and remains so when restricted to sl(2, C), being then labeled $(\lambda, 0)$ if $\lambda \geqslant 0$ and $(0, -\lambda)$ if $\lambda < 0$. The basis operator Δ is constant, having the value $i(|\lambda| + 1)$, and of course $\kappa_\mu = 0$.

We have shown also that if ψ is an indecomposable $[\lambda, +; 0, 0]$ field, and lies in a locally conformal-invariant vector space $\mathscr{U}$ of solutions of Eq. (1.1), then ψ actually satisfies a set of equations including (1.1). These equations are equivalent to the scalar wave equation if $\lambda = 0$; to two-component neutrino equations if $|\lambda| = \frac{1}{2}$; to Maxwell's free field equations if $|\lambda| = 1$; and in general to the Bargmann–Wigner equations for massless fields of helicity λ. The conformal invariance of these sets of equations is well known.[5] In order to find new conformal-invariant free massless field theories, possibly of interest to physics, it is therefore necessary to look at what are in effect, the only remaining possibilities, indecomposable $[\lambda, +; 0, u]$ fields with $u > 0$. These are fields of type Ib, and the corresponding representations, although indecomposable, are not irreducible.

We have not attempted a complete description of these representations in BJ1. Indeed, we have not even proved their existence for arbitrary integers 2λ and $u > 0$. It is the purpose of this work to fill these gaps. That we are able to achieve this object completely, as the ensuing Theorem 2.1 shows, is remarkable, given the apparent intractability of the corresponding task for the totality of representations of type Ib, or even those of Class $\mathscr{D}$. Our success depends upon the diagonalizability of Δ in $[\lambda, +; 0, u]$-representations, and the availability of Gabriel's theorem,[6,7] whose substance should not be underestimated. We were able to derive our own proof of the latter from "first principles" for the special case of interest to us (i.e., for the quiver corresponding to the Dynkin diagram for A_{u+1} - see Sec. 2) but this proof runs to several pages.

In subsequent work we shall describe the structure of the new sets of locally conformal-invariant massless field equations obtained for indecomposable $[\lambda, +; 0, u]$ fields with $u > 0$.

II. STRUCTURE OF THE RELEVANT REPRESENTATIONS OF $\mathscr{W}$

Theorem 2.1: Up to equivalence, there is exactly one indecomposable $[\lambda, +; 0, u]$-representation of $\mathscr{W}$ for each integer or semi-integer λ and each non-negative integer u. When regarded as a representation of sl(2, C), this representation of $\mathscr{W}$ has the decomposition

$$(\lambda, 0) \oplus (\lambda + \tfrac{1}{2}, \tfrac{1}{2}) \oplus \ldots \oplus (\lambda + \tfrac{1}{2}u, \tfrac{1}{2}u) \tag{2.1}$$

if $\lambda \geqslant 0$, and

$$(0, -\lambda) \oplus (\tfrac{1}{2}, \tfrac{1}{2} - \lambda) \oplus \ldots \oplus (\tfrac{1}{2}u, \tfrac{1}{2}u - \lambda) \tag{2.2}$$

if $\lambda < 0$. The dimension of the representation is

$$d = \tfrac{1}{6}(u+1)(u+2)(2u+3+6|\lambda|). \quad (2.3)$$

A basis consisting of vectors

$$|\delta, s, s_3\rangle,$$
$$\delta \in \{|\lambda|+1, |\lambda|+2, ..., |\lambda|+1+u\},$$
$$s \in \{|\lambda|, |\lambda|+1, ..., \delta-1\},$$
$$s_3 \in \{s, s-1, ..., -s\}, \quad (2.4)$$

can be introduced, on which the operators Δ, $\Sigma_{\mu\nu}$, κ_μ, and the related operators M, N, C_1, and C_2 of Eqs. (1.6) and (1.7) act as follows [we write

$$(\Sigma_{23}, \Sigma_{31}, \Sigma_{12}) = \mathbf{S}, \quad (2.5a)$$
$$(\Sigma_{01}, \Sigma_{02}, \Sigma_{03}) = \mathbf{T}]: \quad (2.5b)$$
$$\Delta\,|\delta, s, s_3\rangle = i\delta\,|\delta, s, s_3\rangle, \quad (2.6a)$$
$$\mathbf{S}\cdot\mathbf{S}|\delta, s, s_3\rangle = s(s+1)|\delta, s, s_3\rangle, \quad (2.6b)$$
$$S_3|\delta, s, s_3\rangle = s_3|\delta, s, s_3\rangle, \quad (2.6c)$$
$$C_1|\delta, s, s_3\rangle = (\lambda^2+\delta^2-1)|\delta, s, s_3\rangle, \quad (2.6d)$$
$$C_2|\delta, s, s_3\rangle = \lambda\delta\,|\delta, s, s_3\rangle, \quad (2.6e)$$
$$M\,|\delta, s, s_3\rangle = \tfrac{1}{2}(\delta+\lambda-1)|\delta, s, s_3\rangle, \quad (2.6f)$$
$$N\,|\delta, s, s_3\rangle = \tfrac{1}{2}(\delta-\lambda-1)|\delta, s, s_3\rangle, \quad (2.6g)$$
$$(S_1 \pm iS_2)|\delta, s, s_3\rangle = [(s \pm s_3+1)(s \mp s_3)]^{1/2}|\delta, s, s_3 \pm 1\rangle, \quad (2.7)$$

$$\begin{aligned}T_3|\delta, s, s_3\rangle &= D(s)[(\delta-s)(\delta+s)(s-s_3)(s+s_3)]^{1/2} \\ &\quad\times|\delta, s-1, s_3\rangle + s_3\delta E(s)|\delta, s, s_3\rangle \\ &\quad - D(s+1)[(\delta-s-1)(\delta+s+1) \\ &\quad\times(s-s_3+1)(s+s_3+1)]^{1/2}|\delta, s+1, s_3\rangle, \end{aligned} \quad (2.8)$$

$$\begin{aligned}(T_1 \pm iT_2)|\delta, s, s_3\rangle &= \pm D(s)[(\delta-s)(\delta+s)(s\mp s_3)(s\mp s_3-1)]^{1/2} \\ &\quad\times|\delta, s-1, s_3\pm 1\rangle \\ &\quad+\delta E(s)[(s\mp s_3)(s\pm s_3+1)]^{1/2}|\delta, s, s_3\pm 1\rangle \\ &\quad\pm D(s+1)[(\delta-s-1)(\delta+s+1) \\ &\quad\times(s\pm s_3+1)(s\pm s_3+2)]^{1/2} \\ &\quad\times|\delta, s+1, s_3\pm 1\rangle, \end{aligned} \quad (2.9)$$

$$\kappa_0|\delta, s, s_3\rangle = \kappa[(\delta-s)(\delta+s+1)]^{1/2}|\delta+1, s, s_3\rangle, \quad (2.10)$$

$$\begin{aligned}\kappa_3|\delta, s, s_3\rangle &= i\kappa D(s)[(\delta-s)(\delta-s+1)(s-s_3)(s+s_3)]^{1/2} \\ &\quad\times|\delta+1, s-1, s_3\rangle \\ &\quad+i\kappa s_3E(s)[(\delta-s)(\delta+s+1)]^{1/2}|\delta+1, s, s_3\rangle \\ &\quad-i\kappa D(s+1)[(\delta+s+1)(\delta+s+2)(s-s_3+1) \\ &\quad\times(s+s_3+1)]^{1/2}|\delta+1, s+1, s_3\rangle, \end{aligned} \quad (2.11)$$

$$\begin{aligned}(\kappa_1 \pm i\kappa_2)|\delta, s, s_3\rangle &= \pm i\kappa D(s)[(\delta-s)(\delta-s+1)(s\mp s_3) \\ &\quad\times(s\mp s_3-1)]^{1/2}|\delta+1, s-1, s_3\pm 1\rangle \\ &\quad+i\kappa E(s)[(\delta-s)(\delta+s+1)(s\mp s_3) \\ &\quad\times(s\pm s_3+1)]^{1/2}|\delta+1, s, s_3\pm 1\rangle \\ &\quad\pm i\kappa D(s+1)[(\delta+s+1)(\delta+s+2) \\ &\quad\times(s\pm s_3+1)(s\pm s_3+2)]^{1/2} \\ &\quad\times|\delta+1, s+1, s_3\pm 1\rangle. \end{aligned} \quad (2.12)$$

where

$$D(s) = (s^2-\lambda^2)^{1/2}/s(4s^2-1)^{1/2},$$
$$E(s) = -i\lambda/s(s+1), \quad (2.13)$$

and κ is a nonzero constant. The nonzero value of κ is immaterial, representations which differ only in this value being equivalent.[8] The formulas (2.10, 2.11) for κ_μ applied to $|\delta, s, s_3\rangle$ are only valid for $\delta < (|\lambda|+1+u)$, and

$$\kappa_\mu|(|\lambda|+1+u), s, s_3\rangle = 0. \quad (2.14)$$

The operators κ_μ are nilpotent, and the product $(\kappa_\mu\kappa_\nu...\kappa_\rho)$ is not identically zero only if it does not contain more than u factors. (In particular, if $u=0$ then $\kappa_\mu = 0$.)

Proof: We know that in such an indecomposable representation of $\mathscr{W}$, $(-i\Delta)$ has the eigenvalues (1.8c)

$$\delta = |\lambda|+1, |\lambda|+2, ..., |\lambda|+1+u.$$

Since Eqs. (1.8a, b) hold, it follows that, if $\lambda \geqslant 0$, the pair (M, N) has eigenvalue pairs

$$(m, n) = (\lambda, 0), (\lambda+\tfrac{1}{2}, \tfrac{1}{2}), ..., (\lambda+\tfrac{1}{2}u, \tfrac{1}{2}u), \quad (2.15)$$

while if $\lambda < 0$, it has eigenvalue pairs

$$(m, n) = (0, -\lambda), (\tfrac{1}{2}, \tfrac{1}{2}-\lambda), ..., (\tfrac{1}{2}u, \tfrac{1}{2}u-\lambda). \quad (2.16)$$

Accordingly, this representation of $\mathscr{W}$, when regarded as a representation of sl(2, C), has the general form

$$r_0(\lambda, 0) \oplus r_1(\lambda+\tfrac{1}{2}, \tfrac{1}{2}) \oplus \cdots \oplus r_u(\lambda+\tfrac{1}{2}u, \tfrac{1}{2}u), \quad (2.17)$$

for $\lambda \geqslant 0$, or

$$r_0(0, -\lambda) \oplus r_1(\tfrac{1}{2}, \tfrac{1}{2}-\lambda) \oplus \cdots \oplus r_u(\tfrac{1}{2}u, \tfrac{1}{2}u-\lambda), \quad (2.18)$$

for $\lambda < 0$, where $r_0, r_1, ..., r_u$ are certain *positive* integers. It is convenient at this stage to go from the (m, n) to the $[k_0, c]$ labeling scheme[4] for the finite-dimensional irreducible representations of sl(2, C), where

$$k_0 = m-n,$$
$$c = m+n+1. \quad (2.19)$$

In the case at hand, because Eq. (1.8b) holds, we get only representations with $k_0 = \lambda$, and the decompositions (2.15) and (2.16) have the common form

$$r_0[\lambda, |\lambda|+1] \oplus r_1[\lambda, |\lambda|+2] \oplus \cdots \oplus r_u[\lambda, |\lambda|+u]. \quad (2.20)$$

We see that the eigenspace $\mathscr{V}_\delta$, associated with the eigenvalue δ of $(-i\Delta)$, carries the direct sum of r_τ copies of the representation $[\lambda, \delta]$ of sl(2, C), where $\tau = \delta - |\lambda| - 1$. We imagine these copies ordered in some definite way, and labeled by an index α taking values 1, 2, ..., r_τ. Now each representation $[\lambda, \delta]$ of sl(2, C), when regarded as a representation of its su(2) subalgebra spanned by the operators $\mathbf{S}$, is a direct sum of $(2s+1)$-dimensional irreducible representations (s) of su(2), for $s = |\lambda|, |\lambda|+1, ..., \delta-1$ (each occurring once). And each representation (s) of su(2), when regarded as a representation of its u(1) subalgebra spanned by the operator S_3, is a direct sum of one-dimensional representations $[s_3]$ of u(1), for $s_3 = s, s-1, ..., -s$ (each occurring once.) Accordingly, we can introduce a set of basis vectors for the whole carrier space of the given representation of $\mathscr{W}$, labeled

$$|\delta, \alpha, s, s_3\rangle, \tag{2.21}$$

where δ runs over the eigenvalues of $(-i\Delta)$ as in Eqs. (2.4); for each δ $(=\tau+|\lambda|+1)$, α runs over the values $1, 2, ..., r_\tau$, and independently s runs over the values $|\lambda|, |\lambda|+1, ..., \delta-1$; and for each s, s_3 runs over the values $s, s-1, ..., -s$. On the basis vector (2.21), the operators Δ, $\mathbf{S}\cdot\mathbf{S}$ and S_3 will have the eigenvalues $i\delta$, $s(s+1)$, and s_3, respectively. Moreover, in view of Eqs. (1.6) and (1.7), the operators M, N, C_1, and C_2 will have the eigenvalues $\frac{1}{2}(\delta+\lambda-1)$, $\frac{1}{2}(\delta-\lambda-1)$, $(\lambda^2+\delta^2-1)$, and $\lambda\delta$, respectively. The action of the sl(2, C) operators in an su(2)$\geqslant$u(1) basis of an irreducible representation $[k_0, c]$ is well known.[4] We get Eqs. (2.7), (2.8), and (2.9) with $|\delta, s, s_3\rangle$ replaced by $|\delta, \alpha, s, s_3\rangle$ throughout. (These operators do not "see" the label α.)

We now turn to the action of the operators κ_μ. In view of the commutation relation (1.3e) and the fact that κ_0 commutes with $\mathbf{S}$, we must have

$$\kappa_0|\delta, \alpha, s, s_3\rangle = \sum_\beta A_{\beta\alpha}|\delta+1, \beta, s, s_3\rangle, \tag{2.22}$$

for some complex numbers $A_{\beta\alpha}$, which *á priori* could depend on δ and s (but not on s_3). The sum is over the $r_{\tau+1}$ values of β (with $\tau=\delta-|\lambda|-1$). Equation (2.22) can only hold for $\delta<\delta_{\max}=(|\lambda|+1+u)$, and we must have also

$$\kappa_0|(|\lambda|+1+u), \alpha, s, s_3\rangle = 0. \tag{2.23}$$

According to Eq. (1.3b), κ_μ is a four-vector operator. The most general structure possible for such operators within a finite-dimensional representation of sl(2, C) is well known.[4] We can apply these known general results to the particular situation at hand, or determine the structure directly, noting that a necessary and sufficient condition for κ_0 as in Eqs. (2.22) and (2.23) to be the fourth component of a four-vector is that

$$[[\kappa_0, T_3], T_3] = -\kappa_0. \tag{2.24}$$

[The remaining components of κ_μ can then be defined by

$$i\kappa_i = [\kappa_0, T_i], \tag{2.25}$$

and the commutation relations (1.3b) will then be satisfied.] We get, in place of Eq. (2.22),

$$\kappa_0|\delta, \alpha, s, s_3\rangle = \sum_\beta B^{(\tau)}_{\beta\alpha}[(\delta-s)(\delta+s+1)]^{1/2}|\delta+1, \beta, s, s_3\rangle, \tag{2.26}$$

where the $B^{(\tau)}_{\beta\alpha}$, $\tau \equiv (\delta-|\lambda|-1) = 0, 1, ..., u-1$, are complex numbers which do not depend on s or s_3, but are otherwise not restricted by Eq. (2.24). For each value of τ, we may regard them as the elements of an $(r_{\tau+1}\times r_\tau)$ matrix $B^{(\tau)}$. We might expect these matrices to be restricted in form by the relations (1.5a), (1.5b) and (1.3c) which are required of a $[\lambda, +; 0, u]$ representation, but in fact this is not the case. These relations place no restrictions whatsoever on the $B^{(\tau)}$ but are satisfied once κ_0 and κ_i have the forms determined by Eqs. (2.26), (2.23), and (2.25). We see this most simply as follows.

The operators κ_μ as defined so far are shift operators for $(-i\Delta)$, M, and N, and in fact we have

$$M\kappa_\mu = \kappa_\mu(M+\tfrac{1}{2}), \quad N\kappa_\mu = \kappa_\mu(N+\tfrac{1}{2}). \tag{2.27}$$

It follows that

$$\begin{aligned} M[\kappa_\mu, \kappa_\nu] &= [\kappa_\mu, \kappa_\nu](M+1), \\ N[\kappa_\mu, \kappa_\nu] &= [\kappa_\mu, \kappa_\nu](N+1). \end{aligned} \tag{2.28}$$

Thus $[\kappa_\mu, \kappa_\nu]$ shifts any vector from a repesentation subspace of sl(2, C) labeled (m, n) to one labeled $(m+1, n+1)$. But, just as a four-vector operator [transforming according to the representation $(\frac{1}{2}, \frac{1}{2})$ itself] can only link[4] (m, n) with $(m+\frac{1}{2}, n\pm\frac{1}{2})$ and $(m-\frac{1}{2}, n\pm\frac{1}{2})$, so any antisymmetric tensor operator like $[\kappa_\mu, \kappa_\nu]$ [transforming according to the representation $(1,0)\oplus(0,1)$] can only link (m, n) with $(m\pm1, n)$, (m, n), and $(m, n\pm1)$. It cannot link (m, n) with $(m+1, n+1)$—and to avoid a contradiction it must be true that

$$[\kappa_\mu, \kappa_\nu] = 0. \tag{2.29}$$

Similarly, we have

$$\begin{aligned} M(\kappa_\mu\kappa^\mu) &= (\kappa_\mu\kappa^\mu)(M+1), \\ N(\kappa_\mu\kappa^\mu) &= (\kappa_\mu\kappa^\mu)(N+1). \end{aligned} \tag{2.30}$$

But a scalar operator like $(\kappa_\mu\kappa^\mu)$ cannot link (m, n) with $(m+1, n+1)$, and so

$$\kappa_\mu\kappa^\mu = 0. \tag{2.31}$$

Consider the commutator

$$\begin{aligned} [\kappa_\mu, C_1] &= [\kappa_\mu, \tfrac{1}{2}\Sigma_{\nu\rho}\Sigma^{\nu\rho}] \\ &= 2i\Sigma_{\mu\rho}\kappa^\rho + 3\kappa_\mu \end{aligned} \tag{2.32}$$

using the relations (1.3b), already established. In view of Eqs. (1.7) and (2.27) we then have

$$\begin{aligned} i\Sigma_{\mu\rho}\kappa^\rho + \tfrac{3}{2}\kappa_\mu &= [\kappa_\mu, M(M+1)] + [\kappa_\mu, N(N+1)] \\ &= \kappa_\mu M(M+1) - M(M+1)\kappa_\mu \\ &\quad + \kappa_\mu N(N+1) - N(N+1)\kappa_\mu \\ &= \kappa_\mu M(M+1) - \kappa_\mu(M+\tfrac{1}{2})(M+\tfrac{3}{2}) \\ &\quad + \kappa_\mu N(N+1) - \kappa_\mu(N+\tfrac{1}{2})(N+\tfrac{3}{2}) \\ &= -\kappa_\mu(M+N+\tfrac{3}{2}), \end{aligned} \tag{2.33}$$

so that

$$\begin{aligned} \Sigma_{\mu\rho}\kappa^\rho &= i\kappa_\mu(M+N+3) \\ &= i(M+N+2)\kappa_\mu = (i+\Delta)\kappa_\mu, \end{aligned} \tag{2.34}$$

as required. Thus we see that Eqs. (1.3c), (1.5a), and (1.5b) are all satisfied.

How then are the matrices $B^{(\tau)}$ restricted? It is easy to see that for no τ can $B^{(\tau)}$ be identically zero; otherwise the representation space splits into the direct sum of nontrivial $\mathscr{W}$-invariant subspaces, contradicting the assumed indecomposability of the given representation. But the indecomposability restricts them much more than this. Consider the effect of a change of basis, of the special form

$$|\delta, \alpha, s, s_3\rangle' = \sum_\beta S^{(\tau)}_{\beta\alpha}|\delta, \beta, s, s_3\rangle, \tag{2.35}$$

where, for each δ as in Eqs. (2.4) and corresponding $\tau=\delta-|\lambda|-1$, the $(r_\tau\times r_\tau)$ matrix $S^{(\tau)}$ with complex elements $S^{(\tau)}_{\beta\alpha}$ is nonsingular. Then

$$|\delta, \alpha, s, s_3\rangle = \sum_{\beta} S^{(\tau)-1}_{\beta\alpha} |\delta, \beta, s, s_3\rangle', \tag{2.36}$$

where $S^{(\tau)-1}$ is the inverse of $S^{(\tau)}$, and so, from Eq. (2.26),

$$\kappa_0 \sum_{\beta} S^{(\tau)-1}_{\beta\alpha} |\delta, \beta, s, s_3\rangle'$$

$$= \sum_{\gamma} \sum_{\beta} B^{(\tau)}_{\beta\alpha} [(\delta - s)(\delta + s + 1)]^{1/2} \times S^{(\tau+1)-1}_{\gamma\beta} |\delta + 1, \gamma, s, s_3\rangle',$$

i.e.,

$$\kappa_0 |\delta, \alpha, s, s_3\rangle' = \sum_{\beta} B^{(\tau)'}_{\beta\alpha} [(\delta - s)(\delta + s + 1)]^{1/2} |\delta, \beta, s, s_3\rangle', \tag{2.37}$$

where

$$B^{(\tau)'}_{\beta\alpha} = \sum_{\gamma} \sum_{\sigma} S^{(\tau+1)-1}_{\beta\gamma} B^{(\tau)}_{\gamma\sigma} S^{(\tau)}_{\sigma\alpha}. \tag{2.38}$$

In short,

$$B^{(\tau)'} = S^{(\tau+1)-1} B^{(\tau)} S^{(\tau)}, \quad \tau = 0, 1, ..., u - 1. \tag{2.39}$$

Since we are only interested in the structure of the representation $[\lambda, +; 0, u]$ up to equivalence, we may look for a canonical form of the matrices $B^{(\tau)}$ with respect to transformations of the form (2.39).

Consider a sequence of $(u + 1)$ complex vector spaces Y_τ, $\tau = 0, 1, ..., u$ of dimension $r_0, r_1, ..., r_u$, respectively. The matrices $B^{(\tau)}$ define a sequence of linear mappings between the spaces Y_τ, shown diagrammatically thus:

$$Y_0 \xrightarrow{B^{(0)}} Y_1 \xrightarrow{B^{(1)}} Y_2 \quad \cdots \quad Y_{u-1} \xrightarrow{B^{(u-1)}} Y_u \tag{2.40}$$

Now consider in abstraction the oriented, connected graph appearing in that diagram,

$$\circ \longrightarrow \circ \longrightarrow \circ \cdots \circ \longrightarrow \circ . \tag{2.41}$$

Such a graph, and more generally, any finite, oriented, connected graph, is called a *quiver*. If with each vertex of the quiver (2.41) is associated a finite-dimensional vector space, and with each directed edge a linear mapping in the appropriate direction, as in the diagram (2.40), then one has a *representation* (Y, B) of the quiver. The *direct sum* of two such representations (Y, B), (Y', B') is the representation (Y'', B''), where for each τ,

$$Y_\tau'' = Y_\tau \oplus Y_\tau', \quad B^{(\tau)''} = B^{(\tau)} \oplus B^{(\tau)'}. \tag{2.42}$$

A representation (Y, B) is *indecomposable* if it cannot be represented as a direct sum of two nontrivial representations. Two representations (Y, B), (Y', B') are *equivalent* if there exist invertible mappings $S^{(\tau)}$

$$S^{(\tau)}: Y_\tau' \longrightarrow Y_\tau \tag{2.43}$$

such that

$$B^{(\tau)'} = S^{(\tau+1)-1} B^{(\tau)} S^{(\tau)} \tag{2.44}$$

for $\tau = 0, 1, ..., u - 1$. It can be seen that an indecomposable $[\lambda, +; 0, u]$ representation of $\mathscr{W}$ defines an indecomposable representation of the quiver (2.41), and that any indecomposable representation of the quiver in which none of the $B^{(\tau)}$ is identically zero, defines a $[\lambda, +; 0, u]$ representation of $\mathscr{W}$. Moreover, equivalent representations of the quiver define equivalent representations of $\mathscr{W}$. The problem now arises of classifying the equivalence classes of indecomposable representations of the quiver (2.41). The notion of a representation, and of the indecomposability and equivalence thereof, can be defined for any quiver. Gabriel[6] (see also Bernstein *et al.*[7]) has posed and answered the following question: for which quivers are there finitely many equivalence classes of indecomposable representations? He found that a necessary and sufficient condition is that the graph, when unoriented (i.e., with the arrows removed from the edges) must coincide with the Dynkin diagram for one of the simple Lie algebras[7] $A_2, A_3, ..., D_4, D_5, ..., E_6, E_7$, or E_8. What is more remarkable is that in every such case there is a one-to-one correspondence between the equivalence classes and the positive (integral[7]) roots associated with the corresponding Lie algebra. In the case at hand, we have the Dynkin diagram of A_{u+1}, and the result is that, if the positive root is $(r_0, r_1, ..., r_u)$, then the dimension of Y_τ is r_τ in any representation (Y, B) from the corresponding class. There are $\frac{1}{2}(u + 1)(u + 2)$ positive roots of A_{u+1}, *viz.*[7]

$$\begin{array}{l} (1, 0, 0, ..., 0), (0, 1, 0, ..., 0), ..., (0, 0, 0, ..., 1) \\ (1, 1, 0, ..., 0), (0, 1, 1, 0, ..., 0), ..., (0, 0, 0, ..., 0, 1, 1) \\ \vdots \\ (1, 1, 1, ..., 1) \end{array} \tag{2.45}$$

But we are only interested in the situation where all $B^{(\tau)}$ are nontrivial, as already remarked, so only the last root is of relevance. (The others correspond to representations $[\lambda, +; l, v]$ of $\mathscr{W}$ with $l > 0$ or $v < u$.) Accordingly, each of the spaces Y_τ is one-dimensional, and

$$r_0 = r_1 = \cdots = r_u = 1. \tag{2.46}$$

We may now drop the unnecessary label α from the basis vectors (2.21). Each matrix $B^{(\tau)}$ reduces to a nonzero constant—and furthermore, since there is just one equivalence class corresponding to the last of the roots (2.45), we can without loss of generality take all these constants equal, to κ say. Thus we arrive at the form (2.10) for the action of κ_0 on the basis vector $|\delta, s, s_3\rangle$, and the forms (2.11) for the remaining components are simply obtained from Eq. (2.25). The dimension d of the representation of $\mathscr{W}$ is now obtained by adding the dimensions of the irreducible representations $[\lambda, |\lambda| + 1], [\lambda, |\lambda| + 2], ..., [\lambda, |\lambda| + u]$ of sl(2, C), as

$$d = \sum_{\tau=0}^{u} (\tau + 1)(2|\lambda| + \tau + 1), \tag{2.47}$$

yielding the result (2.3). That the product $(\kappa_\mu \kappa_\nu \cdots \kappa_\rho)$ is not identically zero only if it does not contain more than u factors, follows at once from the action of κ_μ as defined by Eqs. (2.10), (2.11), (2.12), and (2.14). □

III. AN ILLUSTRATIVE EXAMPLE

Consider a $[0, +; 0, 1]$ field. It has five components, and the sl(2, C) content of the index space representation is

$$(0,0)\oplus(\tfrac{1}{2},\tfrac{1}{2}). \tag{3.1}$$

The basis vectors $|\delta, s, s_3\rangle$ of Theorem 2.1 run over $|1,0,0\rangle$, $|2,0,0\rangle$, $|2,1,1\rangle$, $|2,1,0\rangle$, and $|2,1,-1\rangle$. Represent them by column vectors $(1\,0\,0\,0\,0)^T$, $(0\,1\,0\,0\,0)^T$, etc. Let E_{RS} ($R, S\in\{1,2,3,4,5\}$) denote the 5×5 matrix with a 1 in the R th row and S th column, and zeros elsewhere. Then according to Eqs. (2.6)–(2.14), the matrix representations of the $\mathscr{W}$ operators are

$$\begin{aligned}
&S_3 = E_{33} - E_{55}, \quad T_3 = E_{24} - E_{42},\\
&S_1 + iS_2 = (\sqrt{2})(E_{34} + E_{45}),\\
&S_1 - iS_2 = (\sqrt{2})(E_{43} + E_{54}),\\
&T_1 + iT_2 = (\sqrt{2})(E_{25} + E_{32}),\\
&T_1 - iT_2 = -(\sqrt{2})(E_{52} + E_{23}),\\
&\kappa_0 = \kappa(\sqrt{2})E_{21}, \quad \kappa_3 = -i\kappa(\sqrt{2})E_{41},\\
&\kappa_1 + i\kappa_2 = 2i\kappa E_{31}, \quad \kappa_1 - i\kappa_2 = -2i\kappa E_{51},\\
&\Delta = iE_{11} + 2i(E_{22} + E_{33} + E_{44} + E_{55}).
\end{aligned} \tag{3.2}$$

Now make a unitary transformation

$$A \rightarrow UAU^{\dagger} \tag{3.3}$$

of each of the $\mathscr{W}$ operators A, where

$$\begin{aligned}U &= E_{11} + E_{22} - iE_{54}\\ &\quad + (1/\sqrt{2})(iE_{33} - iE_{35} - E_{43} - E_{45}).\end{aligned} \tag{3.4}$$

This corresponds to a change from the su(2)$\supset$u(1) basis to a tensor basis. An arbitrary $[0, +; 0, 1]$ field then takes the form

$$\psi(x) = \begin{pmatrix}\varphi(x)\\ A_\mu(x)\end{pmatrix}, \tag{3.5}$$

where φ is an sl(2, C) scalar field, and A_μ a four-vector field. The action of the $\mathscr{W}$ operators is then found to be

$$\Sigma_{\mu\nu}\begin{pmatrix}\varphi\\ A_\rho\end{pmatrix} = i\begin{pmatrix}0\\ g_{\mu\rho}A_\nu - g_{\nu\rho}A_\mu\end{pmatrix}, \tag{3.6a}$$

$$\Delta\begin{pmatrix}\varphi\\ A_\mu\end{pmatrix} = i\begin{pmatrix}\varphi\\ 2A_\mu\end{pmatrix}, \tag{3.6b}$$

$$\kappa_\mu\begin{pmatrix}\varphi\\ A_\nu\end{pmatrix} = \kappa'\begin{pmatrix}0\\ g_{\mu\nu}\varphi\end{pmatrix}, \quad \kappa' = \kappa\sqrt{2}. \tag{3.6c}$$

Consider an infinitesimal scale transformation

$$x'^\mu = (1+\epsilon)x^\mu \tag{3.7}$$

and the corresponding transformation of ψ, as generated by the operators (1.4c),

$$\begin{aligned}\psi'(x) &= \psi(x) + i\epsilon(ix^\mu\partial_\mu + \Delta)\psi(x)\\ &= (1 + i\epsilon\Delta)\psi((1-\epsilon)x),\end{aligned}$$

i.e.,

$$\psi'(x') = (1 + i\epsilon\Delta)\psi(x). \tag{3.8}$$

Then

$$\varphi'(x') = (1-\epsilon)\varphi(x), \tag{3.9a}$$

$$A_\mu'(x') = (1-2\epsilon)A_\mu(x), \tag{3.9b}$$

so $\varphi(x)$ has length dimension (-1) and $A_\mu(x)$ has dimension (-2). [Note that the four-vector potential of the electromagnetic field has dimension (-1).] Now consider an infinitesimal special conformal transformation

$$x'^\mu = x^\mu + 2\theta^\nu x_\nu x^\mu - \theta^\mu x_\nu x^\nu \tag{3.10}$$

and the corresponding field transformation

$$\begin{aligned}\psi'(x) &= \psi(x) + i\theta^\mu(2ix_\mu x^\nu\partial_\nu + 2x_\mu\Delta\\ &\qquad - ix^\nu x_\nu\partial_\mu + 2\Sigma_{\mu\nu}x^\nu + \kappa_\mu)\psi(x)\\ &= (1 + 2i\theta^\mu x_\mu\Delta + 2i\theta^\mu\Sigma_{\mu\nu}x^\nu + i\theta^\mu\kappa_\mu)\\ &\qquad \times\psi(x^\mu - 2\theta^\nu x_\nu x^\mu + \theta^\mu x_\nu x^\nu),\end{aligned}$$

i.e.,

$$\begin{aligned}\psi'(x') &= (1 + 2i\theta^\mu x_\mu\Delta\\ &\qquad + 2i\theta^\mu\Sigma_{\mu\nu}x^\nu + i\theta^\mu\kappa_\mu)\psi(x).\end{aligned} \tag{3.11}$$

Then

$$\varphi'(x') = (1 - 2\theta^\mu x_\mu)\varphi(x), \tag{3.12}$$

the usual transformation law for a scalar field, while

$$\begin{aligned}A'_\mu(x') &= A_\mu(x) - 4\theta^\nu x_\nu A_\mu(x) - 2\theta_\mu x^\nu A_\nu(x)\\ &\qquad + 2x_\mu\theta^\nu A_\nu(x) + i\kappa'\theta_\mu\varphi(x).\end{aligned} \tag{3.13}$$

Here we see the novel feature of Type Ib fields—under the action of the conformal group, fields belonging to different index-space irreducible representations of sl(2, C) are mixed together.

Note that the subspace of fields having $\varphi(x) = 0$ is invariant under this action. This corresponds to the fact that although the representation $[0, +; 0, 1]$ of $\mathscr{W}$ is indecomposable, it is not irreducible, and it "contains" the indecomposable (and irreducible) representation $[0, +; 1, 1]$ as an invariant subrepresentation. More generally, we can see that $[\lambda, +; 0, u]$ contains $[\lambda, +; 1, u]$, which contains $[\lambda, +; 2, u]$, etc. Of the representations $[\lambda, +; 0, u]$, only those with $u = 0$ are irreducible.

ACKNOWLEDGMENT

We are indebted to Dr. B. D. Jones for drawing our attention to Gabriel's theorem and for related discussions.

[1] A. J. Bracken and Barry Jessup, "Local conformal invariance of the wave equation for finite-component fields. I. The conditions for invariance, and fully-reducible fields," J. Math. Phys. **23**, 1925 (1982).

[2] G. Mack and A. Salam, Ann. Phys. **53**, 174 (1969).

[3] See, for example, J. Mickelsson and J. Niederle, J. Math. Phys. **13**, 23 (1972).

[4] I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications* (Pergamon, London, 1963).

[5] H. Bateman, Proc. Lond. Math. Soc. **8**(2), 223 (1910); E. Cunningham, *ibid.* **8**(2), 77 (1910); J. A. McLennan, Nuovo Cimento **3**(10), 1360 (1956); L. Gross, J. Math. Phys. **5**, 687 (1964).

[6] P. Gabriel, Manuscr. Math. **6**, 71 (1971).

[7] I. N. Bernstein, I. M. Gel'fand, and V. A. Ponomarev, Usp. Mat. Nauk **28** (2), 19 (1973). [Russ. Math. Surv. **28**(2), 17–32 (1973)].

[8] We could set $\kappa = 1$, but this would obscure the fact that κ (and κ_μ) has dimensions of a length.

Explicit construction of the exceptional superalgebras $F(4)$ and $G(3)$

Bryce S. DeWitt
Department of Physics, University of Texas at Austin, Austin, Texas 78712
Institute for Theoretical Physics, University of California at Santa Barbara, Santa Barbara, California 93106
Peter van Nieuwenhuizen
Department of Physics, State University of New York at Stony Brook, Stony Brook, New York 11794

(Received 24 August 1981; accepted for publication 16 October 1981)

A general method for constructing the extension of an ordinary Lie algebra $\mathscr{A}_0$ to a superalgebra $\mathscr{A}_0 \oplus \mathscr{A}_1$ is given, once one knows in which representation of $\mathscr{A}_0$ the odd generators $\mathscr{A}_1$ are. Explicit matrix representations for the superalgebras $F(4)$ and $G(3)$, and for ordinary algebras E_8, F_4, and G_2 are presented.

PACS numbers: 11.30.Pb

1. INTRODUCTION

Superalgebras play an increasingly important part in physics, since they are the mathematical foundation of supergravity theories. In mathematics, a complete classification of all simple superalgebras has been given by Kac.[1] A superalgebra is per definition a finite set of generators which can be divided into even (= bosonic) and odd (= fermionic) elements (a Z_2 grading), such that the bracket relation between any two generators is antisymmetric, except that for two odd generators it is symmetric. Moreover, the super-Jacobi identities are satisfied. They differ from the usual Jacobi identities by signs and an easy way to derive them is to require that they become identities if one defines the bracket relations by commutators and (in the case of two odd generators) by anticommutators. However, for the specific matrix representations which we consider below, the bracket relations are not simply given by *super*commutators. (They are, of course, always supercommutators for the adjoint representation).

The simple superalgebras consist of two main families: $\mathrm{Osp}(n/m)$, with $\mathrm{O}(n)$ and $\mathrm{Sp}(m)$ in the even sector, and $\mathrm{SU}(n/m)$, with $\mathrm{SU}(n)\times\mathrm{SU}(m)\times\mathrm{U}(1)$ in the even sector, except that for $n = m$ the U(1) is omitted. [One may always consider only generators with vanishing graded trace, since they form an ideal. For $n = m$ the U(1) is generated by the unit matrix which has vanishing graded trace but which forms an abelian ideal for $n = m$.] Moreover, there are three kinds of exceptional superalgebras:

(i) the algebras $F(4)$, $G(3)$, and $D(2, 1, \alpha)$;

(ii) the algebras $P(n)$ and $Q(n)$; and

(iii) the algebras W, S, $\widetilde{S}$, and H which can be represented as general coordinate transformations or canonical transformations of anticommuting variables only.

In this article we shall give matrix representations for $F(4)$ and $G(3)$ as well as for the purely bosonic exceptional algebras G_2, F_4, E_8. [We stick to V. Kac's name $F(4)$ for the superalgebra, although it should not be confused with the ordinalry algebra F_4 from Cartan's classification.]

Most of the known applications are based on $\mathrm{Osp}(N/4)$, which yields N-extended Poincaré supergravity with cosmological constant,[2] and $\mathrm{SU}(4/N)$, which yields N-extended conformal supergravity[3] (at least, this is known to be the case for $1\leqslant N\leqslant 3$). The reason is that Sp(4) and SU(4) [or, rather SU(2,2)] are the space-time algebras for $d = 4$ dimensions, namely the de-Sitter algebra SO(3, 2) and the conformal algebra, respectively. It is known that if one goes to higher dimensions, supergravity theories can go only as high as $d = 11$ dimensions,[4] not higher, since otherwise spins exceeding 2 enter, and it has been shown that for such spins no consistent coupling to gravity exists.[5] Gauge supersymmetry[6] is based on $\mathrm{Osp}(3,1/4N)$, which leaves the line element in superspace $\Sigma (x^i)^2 + \Sigma (\bar\theta^i\theta^i)$ invariant, where x^i are the coordinates of 4-dimensional Minkovski space-time.

The question arises which superalgebras yield supergravity theories in dimensions $d > 4$. It is known that simple Poincaré supergravity in $d = 5$ is described by $\mathrm{SU}(4/1)$,[7] where $\mathrm{SU}(4)\sim\mathrm{SO}(6)$ is the de-Sitter algebra for $d = 5$. It seems possible that the superalgebra $F(4)$ (the number 4 denotes the rank of the bosonic part as usual) whose even sector consists of spin $(7)\times\mathrm{SL}(2, R)$, will give rise to Poincaré supergravity in $d = 6$ dimensions, since O(7) is the de-Sitter algebra for $d = 6$, while its spinor representation spin (7) is 8-dimensional, which is indeed the dimension of the Clifford algebra in $d = 6$. Another possibility for $F(4)$[8] would be that it leads to conformal supergravity in $d = 5$, since O(7) can also be viewed as the conformal group in $d = 5$ (the generators Σ_{i6} and Σ_{i7} yield then P_i and K_i, with $i = 1, 5$ while Σ_{67} yields the dilaton generator D). Since one needs in $d = 5$ an even number of spinors (in order to be able to define Majorana spinors) pairs of spinors would then fill up the 8-dimensional columns and rows in the off-diagonal parts. Quite generally, the matrix representations have in the even sector the spinor representations of the bosonic groups, since the odd charges must transform under the bosonic symmetries as spinors (due to the spin-statistics theorem). The superalgebra $G(3)$ has no space-time group in its bosonic sector, but it may play a role in grand unified schemes, where it may combine several internal symmetries. The problems here will be the occurrence of ghosts, just as they occur in $\mathrm{SU}(2/1)$.[9]

We have added explicit matrix representations for the ordinary exceptional Lie algebras E_8, F_4, and G_2. After the demystification of the exceptional Lie algebra E_7 by Cremmer and Julia,[10] and of E_6 by Cremmer, Scherk, and Schwarz,[11] this seemed not only possible, but even desirable. We thank E. Cremmer and B. Julia for discussions about these matters. It may be that our explicit matrix representations will be helpful for the gauging of super $F(4)$, and for phenomenological applications based on the ordinary F_4 and G_2. Since E_7 and E_6 arises as global symmetries of the largest supergravity theory in $d=4$ [10] and $d=5$,[11] respectively, after dimensionally reducing the only supergravity model in $d=11$, one may expect E_8 to emerge in $d=3$.

2. STRUCTURE OF LIE SUPERALGEBRAS

We recall that a Lie superalgebra $\mathscr{A}$ is a vector space (real or complex) that decomposes into two vector spaces $\mathscr{A}_0$ and $\mathscr{A}_1$ ($\mathscr{A}=\mathscr{A}_0\oplus\mathscr{A}_1$) and is endowed with a binary bracket operation [,]: $\mathscr{A}\times\mathscr{A}\to\mathscr{A}$ satisfying the laws

$$[\mathbf{X},\mathbf{Y}] = -(-1)^{\mathbf{XY}}[\mathbf{Y},\mathbf{X}], \tag{2.1}$$

$$[\mathbf{X},[\mathbf{Y},\mathbf{Z}]] + (-1)^{\mathbf{X}(\mathbf{Y}+\mathbf{Z})}[\mathbf{Y},[\mathbf{Z},\mathbf{X}]] + (-1)^{\mathbf{Z}(\mathbf{X}+\mathbf{Y})}[\mathbf{Z},[\mathbf{X},\mathbf{Y}]]=0, \tag{2.2}$$

for all $\mathbf{X},\mathbf{Y},\mathbf{Z}$ in $\mathscr{A}$. Here a symbol appearing in an exponent of (-1) is understood to have the value 0 or 1 according to whether the corresponding element lies in $\mathscr{A}_0$ or $\mathscr{A}_1$. Such elements are called *pure*. Equations (2.1) and (2.2) are extended to the impure case (i.e., to arbitrary elements of $\mathscr{A}$) by application of the linear laws

$$[\mathbf{X},a\mathbf{Y}]=a[\mathbf{X},\mathbf{Y}], \tag{2.3}$$

$$[\mathbf{X},\mathbf{Y}+\mathbf{Z}]=[\mathbf{X},\mathbf{Y}]+[\mathbf{X},\mathbf{Z}], \tag{2.4}$$

a being any number (real or complex).

Equations (2.1) and (2.2) are nontrivially self consistent only if the bracket of two pure elements is itself pure and if $(-1)^{[\mathbf{X},\mathbf{Y}]}=(-1)^{\mathbf{X}+\mathbf{Y}}$. The problem of classifying all Lie superalgebras is equivalent to finding all possible solutions of these equations. Kac[1] has given the complete classification of the *simple* Lie superalgebras, i.e., those that possess no nontrivial subalgebras invariant under the bracket operation. Most of these are easily described in terms of specific representations of low order. A few, notably the exceptional superalgebras $D(2,1,\alpha)$, $F(4)$, and $G(3)$, appear to have no matrix representation of order lower than that of adjoint representation. In this respect they are similar to the exceptional ordinary Lie algebra E_8.

The adjoint representation *ad* $\mathscr{A}$ is obtained by assembling the structure constants into matrices. The structure constants themselves are defined relative to a pure basis $\{\mathbf{e}_i\}$

$$[\mathbf{e}_i,\mathbf{e}_j] \stackrel{\text{def}}{=} \mathbf{e}_k c^k{}_{ij}. \tag{2.5}$$

Since any element of $\mathscr{A}$ is expandable in terms of the $\mathbf{e}_i$ the structure constants determine the superalgebra. Equations (2.1) and (2.2) may be replaced by

$$c^i{}_{jk} = -(-1)^{jk}c^i{}_{kj}, \tag{2.6}$$

$$c^i{}_{jm}c^m{}_{kl} + (-1)^{j(k+l)}c^i{}_{km}c^m{}_{lj} + (-1)^{l(j+k)}c^i{}_{lm}c^m{}_{jk}=0. \tag{2.7}$$

where now an index appearing in an exponent of (-1) takes the value 0 or 1 according to whether the associated basis vector lies in $\mathscr{A}_0$ or $\mathscr{A}_1$.

Let $\mathbf{D}_i$ denote the matrix $(c^j{}_{ik})$. Then Eq. (2.7) is equivalent to

$$[\mathbf{D}_i,\mathbf{D}_j]=\mathbf{D}_k c^k{}_{ij}, \tag{2.8}$$

where the bracket symbol now denotes the *supercommutator*:

$$[\mathbf{D}_i,\mathbf{D}_j] \stackrel{\text{def}}{=} \mathbf{D}_i\mathbf{D}_j - (-1)^{ij}\mathbf{D}_j\mathbf{D}_i. \tag{2.9}$$

The $\mathbf{D}_i$ are the generators of *ad* $\mathscr{A}$. [The symbol "[, }" is often used for the supercommutator. This notation is deplorable in two respects: (1) It leaves the *anti*supercommutator $\mathbf{D}_i\mathbf{D}_j+(-1)^{ij}\mathbf{D}_j\mathbf{D}_i$ in limbo. (2) It fails to emphasize that the supercommutator is the analog of the commutator of ordinary Lie algebra theory, *not* of the anticommutator.]

It will be convenient to use Greek indices from the first of the alphabet to designate basis vectors lying in $\mathscr{A}_1$ and Greek indices from the middle of the alphabet to designate basis vectors lying in $\mathscr{A}_0$. Such indices are often referred to as *fermionic* and *bosonic*, respectively. Because of the rule $(-1)^{[\mathbf{e}_i,\mathbf{e}_j]}=(-1)^{i+j}$, structure constants bearing an odd number of fermionic indices vanish, and the adjoint generators $\mathbf{D}_i$ have the block structure

$$\mathbf{D}_\mu=\begin{pmatrix} c^\nu{}_{\mu\sigma} & 0 \\ 0 & c^\beta{}_{\mu\gamma}\end{pmatrix},\quad \mathbf{D}_\alpha=\begin{pmatrix} 0 & c^\nu{}_{\alpha\gamma} \\ c^\beta{}_{\alpha\sigma} & 0\end{pmatrix}. \tag{2.10}$$

Equations (2.7) decompose into the following four equations:

$$c^\mu{}_{\nu\lambda}c^\lambda{}_{\sigma\tau} + c^\mu{}_{\sigma\lambda}c^\lambda{}_{\tau\nu} + c^\mu{}_{\tau\lambda}c^\lambda{}_{\nu\sigma}=0, \tag{2.11}$$

$$c^\alpha{}_{\mu\gamma}c^\gamma{}_{\nu\beta} + c^\alpha{}_{\nu\gamma}c^\gamma{}_{\beta\mu} + c^\alpha{}_{\beta\sigma}c^\sigma{}_{\mu\nu}=0, \tag{2.12}$$

$$c^\mu{}_{\alpha\gamma}c^\gamma{}_{\beta\nu} - c^\mu{}_{\beta\gamma}c^\gamma{}_{\nu\alpha} + c^\mu{}_{\nu\sigma}c^\sigma{}_{\alpha\beta}=0, \tag{2.13}$$

$$c^\alpha{}_{\beta\mu}c^\mu{}_{\gamma\delta} + c^\alpha{}_{\gamma\mu}c^\mu{}_{\delta\beta} + c^\alpha{}_{\delta\mu}c^\mu{}_{\beta\gamma}=0. \tag{2.14}$$

Equation (2.11) shows that the c's with purely bosonic indices are the structure constants of an ordinary Lie algebra. This Lie algebra is just $\mathscr{A}_0$, which is a subalgebra of $\mathscr{A}$. Equation (2.11) and (2.12) together are equivalent to

$$[\mathbf{D}_\mu,\mathbf{D}_\nu]=\mathbf{D}_\sigma c^\sigma{}_{\mu\nu} \tag{2.15}$$

which shows that the $\mathbf{D}_\mu$ generate a representation of $\mathscr{A}_0$, namely the direct sum of *ad* $\mathscr{A}_0$ and another representation generated by the matrices

$$\mathbf{G}_\mu \stackrel{\text{def}}{=} (c^\alpha{}_{\mu\beta}). \tag{2.16}$$

The latter representation, denoted by $\mathscr{A}_0$: $\mathscr{A}_1$, will be called the *extending representation* because its existence is what makes possible the extension of $\mathscr{A}_0$ to a superalgebra $\mathscr{A}$. It determines the action of $\mathscr{A}_0$ on $\mathscr{A}_1$: Let $\mathbf{X}$ be an element of $\mathscr{A}_0$ and $\mathbf{Y}$ an element of $\mathscr{A}_1$. Then $[X,Y]=X^\mu\mathbf{G}_\mu\mathbf{Y}$, where X^μ are the components of $\mathbf{X}$ relative to the basis $\{\mathbf{e}_\mu\}$.

The extending representation cannot be just any representation of $\mathscr{A}_0$ for it must also be compatible with the structure equations (2.13) and (2.14). Checking this compatibility is the hardest part of the problem of constructing and classifying Lie superalgebras. We outline here a method that works in many cases:

First, choose $\mathscr{A}_0$ to be a semisimple Lie algebra. For such an algebra there always exists a nonsingular symmetric matrix $\overset{s}{\boldsymbol{\eta}}=(\eta_{\mu\nu})$ such that $\eta_{\mu\tau}c^\tau{}_{\nu\sigma}$ is completely antisymmetric in the indices μ, ν, and σ. One such matrix is the Cartan–Killing matrix $(-c^\sigma{}_{\mu\tau}c^\tau{}_{\nu\sigma})$. However, if $\mathscr{A}_0$ is not simple this is not the only possibililty. In an appropriate basis $\overset{s}{\boldsymbol{\eta}}$ may be built out of blocks, each block being the Cartan–Killing matrix of one of the simple invariant subalgebras of $\mathscr{A}_0$ and each block carrying its own arbitrary scale factor. The arbitrariness of the independent scale factors is important.

Next, choose for $\mathscr{A}_0$: $\mathscr{A}_1$ a representation of $\mathscr{A}_0$ for which there exists a nonsingular *anti*symmetric matrix $\overset{A}{\boldsymbol{\eta}} = (\eta_{\alpha\beta})$ such that the matrices $\overset{A}{\boldsymbol{\eta}}\mathbf{G}_\mu$ are all symmetric. The matrices $\overset{s}{\boldsymbol{\eta}}$ and $\overset{A}{\boldsymbol{\eta}}$, together with their inverses $\overset{s}{\boldsymbol{\eta}}{}^{-1} = (\eta^{\mu\nu})$ and $\overset{A}{\boldsymbol{\eta}}{}^{-1} = (\eta^{\alpha\beta})$ can be used to lower and raise bosonic and fermionic indices, respectively. If they are combined into a single matrix

$$\boldsymbol{\eta} \overset{\text{def}}{=} \begin{pmatrix} \overset{s}{\boldsymbol{\eta}} & 0 \\ 0 & \overset{A}{\boldsymbol{\eta}} \end{pmatrix} = (\eta_{ij}), \tag{2.17}$$

then the matrices $\boldsymbol{\eta}\mathbf{D}_\mu$ are *antisupersymmetric*:

$$(\boldsymbol{\eta}\mathbf{D}_\mu)_{ij} = \eta_{ik}c^k{}_{\mu j} \overset{\text{def}}{=} c_{i\mu j} = -(-1)^{ij}c_{j\mu i}. \tag{2.18}$$

Finally check whether the identity

$$\mathbf{G}_\mu \mathbf{S}\overset{A}{\boldsymbol{\eta}}\mathbf{G}^\mu \equiv -\frac{1}{2}\mathbf{G}_\mu \operatorname{tr}\Big(\mathbf{S}\overset{A}{\boldsymbol{\eta}}\,\mathbf{G}^\mu\Big),$$

$$\mathbf{G}^\mu \overset{\text{def}}{=} \eta^{\mu\nu}\mathbf{G}_\nu \tag{2.19}$$

holds, where **S** is *any* symmetric matrix. This is the crucial identity. It turns out that in order to satisfy it $\mathscr{A}_0$: $\mathscr{A}_1$ usually has to be a reducible representation of $\mathscr{A}_0$. If $\mathscr{A}_0$ is chosen to be simple $\mathscr{A}_0$: $\mathscr{A}_1$ can be irreducible only when $\mathscr{A}_0 = \mathrm{SL}(n)$ [or SU(n)]$\mathscr{A}_0$: $\mathscr{A}_1$ is then the adjoint representation. [This leads to the superalgebra $Q(n-1)$ (see Ref. 1).] Among the so-called classical Lie superalgebras[1] the only other examples for which $\mathscr{A}_0$: $\mathscr{A}_1$ is irreducible are $D(2, 1, \alpha)$, $F(4)$, and $G(3)$. For each of these superalgebras $\mathscr{A}_0$ has SL(2) as an invariant subalgebra, and the independent scale factors multiplying the blocks of which $(\eta_{\mu\nu})$ is therefore composed have to be chosen in a special way to make the identity (2.19) hold.

Give a Lie algebra $\mathscr{A}_0$ and an extending representation satisfying the requisite conditions, one immediately has half of the structure constants of the Lie superalgebra $\mathscr{A}_0 \oplus \mathscr{A}_1$, namely, $c^\mu{}_{\nu\sigma}$ and $c^\alpha{}_{\mu\beta}$. The remaining structure constants are obtained by defining

$$c^\alpha{}_{\beta\mu} \overset{\text{def}}{=} -c^\alpha{}_{\mu\beta}, \tag{2.20}$$

$$c^\mu{}_{\alpha\beta} \overset{\text{def}}{=} -\eta^{\mu\nu}\eta_{\alpha\gamma}c^\gamma{}_{\nu\beta} = c^\mu{}_{\beta\alpha}. \tag{2.21}$$

It is easy to see that Eq. (2.13) is then equivalent to (2.12), and Eq. (2.14) is an alternative version of (2.19). Moreover, the matrices $\boldsymbol{\eta}\mathbf{D}_\alpha$ are symmetric, and $\eta_{il}c^l{}_{jk}$ is completely antisupersymmetric in the indices i, j, and k.

We have not actually used the nonsingularity of $\overset{A}{\boldsymbol{\eta}}$ in the above. However, it is not difficult to show that if $\overset{A}{\boldsymbol{\eta}}$ is singular then $\mathscr{A}$ is not simple, nor even semisimple. If $\mathscr{A}$ is simple then $\boldsymbol{\eta}$ is necessarily nonsingular and the Cartan–Killing matrix is necessarily a multiple of it. That is,

$$-\operatorname{str}(\mathbf{D}_i\mathbf{D}_j) = -(-1)^k c^k{}_{il}c^l{}_{jk} = \lambda\eta_{ij}, \tag{2.22}$$

or, equivalently,

$$-c^\sigma{}_{\mu\tau}c^\tau{}_{\nu\sigma} + c^\alpha{}_{\mu\beta}c^\beta{}_{\nu\alpha} = \lambda\eta_{\mu\nu}, \tag{2.23}$$

$$-c^\mu{}_{\alpha\gamma}c^\gamma{}_{\beta\mu} + c^\gamma{}_{\alpha\mu}c^\mu{}_{\beta\gamma} = \lambda\eta_{\alpha\beta}, \tag{2.24}$$

for some constant λ. Because it is the supertrace (str) that appears here, nonsingularity of the Cartan–Killing matrix is not a necessary condition for a Lie superalgebra to be simple, as it is for ordinary Lie algebras. For example, λ vanishes in the case of $D(2, 1\alpha)$, although not in the case of $F(4)$ and $G(3)$.[1]

3. STRUCTURE OF $F(4)$

For this superalgebra $\mathscr{A}_0$ is SO(7) ⊕ SL(2) and $\mathscr{A}_0$: $\mathscr{A}_1$ is spin (7)×sl(2), spin (7) being the 8-dimensional spin representation of SO(7) and sl(2) the 2-dimensional fundamental (or defining) representations of SL(2). $\mathscr{A}_0$ has dimension 24 and rank 4 [the latter number explaining the "4" in "$F(4)$"] and $\mathscr{A}_1$ has 16 dimensions.

Spin (7) is most easily described in terms of γ matrices satisfying

$$\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu}\mathbf{1}, \quad \mu, \nu = 1, \ldots, 7. \tag{3.1}$$

There are only two inequivalent irreducible faithful representations for the γ's, each of dimension 8 and each the negative of the other. It does not matter which representation is chosen, for the generators of spin (7) involve the γ's only in the bilinear combinations

$$\mathbf{G}_{\mu\nu} \overset{\text{def}}{=} \tfrac{1}{4}[\gamma_\mu, \gamma_\nu]. \tag{3.2}$$

It turns out that the γ matrices for spin (7) can be chosen unitary, antisymmetric and pure imaginary. For example,

$$\begin{aligned}
\gamma_1 &= \mathbf{1}\times\sigma_3\times\sigma_2,\\
\gamma_2 &= \mathbf{1}\times\sigma_1\times\sigma_2,\\
\gamma_3 &= \sigma_2\times\mathbf{1}\times\sigma_3,\\
\gamma_4 &= \sigma_2\times\mathbf{1}\times\sigma_1,\\
\gamma_5 &= \sigma_3\times\sigma_2\times\mathbf{1},\\
\gamma_6 &= \sigma_1\times\sigma_2\times\mathbf{1},\\
\gamma_7 &= \sigma_2\times\sigma_2\times\sigma_2,
\end{aligned} \tag{3.3}$$

where the σ's are the Pauli matrices. With such a choice, which will be assumed from now on, the 21 matrices $\mathbf{G}_{\mu\nu}$ are antisymmetric and real while the 35 matrices

$$\mathbf{G}_{\mu\nu\sigma} \overset{\text{def}}{=} \tfrac{1}{3}(\gamma_\mu\mathbf{G}_{\nu\sigma} + \gamma_\nu\mathbf{G}_{\sigma\mu} + \gamma_\sigma\mathbf{G}_{\nu\mu}), \tag{3.4}$$

are symmetric and imaginary. Any antisymmetric 8×8 matrix **A** can be decomposed in the form

$$\mathbf{A} = a_\mu\gamma_\mu + \tfrac{1}{2}a_{\mu\nu}\mathbf{G}_{\mu\nu}, \tag{3.5}$$

and any symmetric 8×8 matrix **S** can be expressed as

$$\mathbf{S} = s\mathbf{1} + \tfrac{1}{6}s_{\mu\nu\sigma}\mathbf{G}_{\mu\nu\sigma}, \tag{3.6}$$

where

$$a_\mu = \tfrac{1}{8}\,\mathrm{tr}(\gamma_\mu \mathbf{A}), \quad a_{\mu\nu} = -\tfrac{1}{2}\,\mathrm{tr}(\mathbf{G}_{\mu\nu}\mathbf{A}), \tag{3.7}$$

$$s = \tfrac{1}{8}\,\mathrm{tr}\,\mathbf{S}, \quad s_{\mu\nu\sigma} = -\tfrac{1}{2}\,\mathrm{tr}(\mathbf{G}_{\mu\nu\sigma}\mathbf{S}). \tag{3.8}$$

We shall need the trace relations

$$\mathrm{tr}\,\gamma_\mu = 0, \tag{3.9}$$

$$\mathrm{tr}(\gamma_\mu\gamma_\nu\gamma_\sigma) = 0, \tag{3.10}$$

$$\mathrm{tr}(\gamma_\mu\gamma_\nu\gamma_\sigma\gamma_\tau\gamma_\rho) = 0, \tag{3.11}$$

$$\mathrm{tr}(\mathbf{G}_{\mu\nu}\mathbf{G}_{\sigma\tau}) = -2(\delta_{\mu\sigma}\delta_{\nu\tau} - \delta_{\mu\tau}\delta_{\nu\sigma}), \tag{3.12}$$

as well as the following identities:

$$\mathbf{G}_{\mu\nu}\mathbf{G}_{\mu\nu} = -\tfrac{21}{2}\mathbf{1}, \tag{3.13}$$

$$\mathbf{G}_{\mu\nu}\gamma_\sigma\mathbf{G}_{\mu\nu} = -\tfrac{9}{2}\gamma_\sigma, \tag{3.14}$$

$$\mathbf{G}_{\mu\nu}\mathbf{G}_{\sigma\tau}\mathbf{G}_{\mu\nu} = -\tfrac{1}{2}\mathbf{G}_{\sigma\tau}, \tag{3.15}$$

$$\mathbf{G}_{\mu\nu}\mathbf{G}_{\rho\sigma\tau}\mathbf{G}_{\mu\nu} = \tfrac{3}{2}G_{\rho\sigma\tau}, \tag{3.16}$$

$$[\mathbf{G}_{\mu\nu}, \mathbf{G}_{\sigma\tau}] = -\delta_{\mu\sigma}\mathbf{G}_{\nu\tau} + \delta_{\mu\tau}\mathbf{G}_{\nu\sigma} - \delta_{\nu\tau}\mathbf{G}_{\mu\sigma} + \delta_{\nu\sigma}\mathbf{G}_{\mu\tau}. \tag{3.17}$$

Equation (3.17) can be rewritten in the form

$$[\mathbf{G}_{\mu\nu}, \mathbf{G}_{\sigma\tau}] = \tfrac{1}{2}G_{\rho\lambda}c_{\rho\lambda\,\mu\nu\,\sigma\tau}, \tag{3.18}$$

where the c's are the structure constants of SO(7):

$$\begin{aligned} c_{\rho\lambda\,\mu\nu\,\sigma\tau} = {} & \delta_{\rho\tau}\delta_{\lambda\nu}\delta_{\mu\sigma} - \delta_{\lambda\tau}\delta_{\rho\nu}\delta_{\mu\sigma} \\ & + \delta_{\rho\nu}\delta_{\lambda\sigma}\delta_{\mu\tau} - \delta_{\lambda\nu}\delta_{\rho\sigma}\delta_{\mu\tau} \\ & + \delta_{\rho\sigma}\delta_{\lambda\mu}\delta_{\nu\tau} - \delta_{\lambda\sigma}\delta_{\rho\mu}\delta_{\nu\tau} \\ & + \delta_{\rho\mu}\delta_{\lambda\tau}\delta_{\nu\sigma} - \delta_{\lambda\mu}\delta_{\rho\tau}\delta_{\nu\sigma}. \end{aligned} \tag{3.19}$$

The Cartan–Killing matrix is, up to a factor, just the unit matrix in the 21-dimensional vector space of 7×7 antisymmetric matrices:

$$-\tfrac{1}{4}c_{\iota\kappa\,\mu\nu\,\rho\lambda}c_{\rho\lambda\,\sigma\tau\,\iota\kappa} = 10\delta_{\mu\nu\,\sigma\tau}, \tag{3.20}$$

$$\delta_{\mu\nu\,\sigma\tau} \stackrel{\text{def}}{=} \delta_{\mu\sigma}\delta_{\nu\tau} - \delta_{\mu\tau}\delta_{\nu\sigma}. \tag{3.21}$$

Turn now to SL(2). This is the 3-dimensional algebra of all traceless real 2×2 matrices, the bracket operation being the commutator. Any such matrix $\mathbf{B}$ can be expressed in the form

$$\mathbf{B} = B^a{}_b\mathbf{G}_a{}^b, \quad \text{with } B^a{}_a = 0,\ a, b = 1, 2, \tag{3.22}$$

$$\mathbf{G}_a{}^b \stackrel{\text{def}}{=} \mathbf{M}_a{}^b - \tfrac{1}{2}\delta_a{}^b\mathbf{1}, \tag{3.23}$$

where $M_a{}^b$ is the 2×2 matrix that has 1 in the ath row and b th column and zeros elsewhere. The $\mathbf{G}_a{}^b$ are the generators of sl(2). They are traceless and satisfy

$$[\mathbf{G}_a{}^b, \mathbf{G}_c{}^d] = \mathbf{G}_e{}^f c^e{}_{fa}{}^b{}_c{}^d, \tag{3.24}$$

where the c's are the structure constants of SL(2):

$$c^e{}_{fa}{}^b{}_c{}^d = \delta^e{}_a\,\delta_f{}^d\,\delta^b{}_c - \delta^e{}_c\,\delta_f{}^b\,\delta^d{}_a. \tag{3.25}$$

The Cartan–Killing matrix in this case is given by

$$-c^e{}_{fa}{}^b{}_g{}^h c^g{}_{hc}{}^d{}_e{}^f = -4(\delta_a{}^d\delta^b{}_c - \tfrac{1}{2}\delta_a{}^b\delta_c{}^d). \tag{3.26}$$

Expressions (3.20) and (3.26), with appropriate scale factors thrown in, are the blocks of the matrix $\overset{s}{\boldsymbol{\eta}}$ discussed in Sec. 2. It turns out that we shall succeed in securing the identity (2.19) if we choose for $\boldsymbol{\eta}$ the 24×24 matrix with elements [introduction of an overall scale factor changes the construction only trivially.]

$$\eta_{\mu\nu\,\sigma\tau} = \delta_{\mu\nu\,\sigma\tau} = \delta_{\mu\sigma}\delta_{\nu\tau} - \delta_{\mu\tau}\delta_{\nu\sigma}, \tag{3.27}$$

$$\eta_a{}^b{}_c{}^d = \tfrac{2}{3}(\delta_a{}^d\,\delta_c{}^b - \tfrac{1}{2}\delta_a{}^b\,\delta_c{}^d) = \eta_c{}^d{}_a{}^b, \tag{3.28}$$

$$\eta_{\mu\nu\,a}{}^b = \eta_a{}^b{}_{\mu\nu} = 0. \tag{3.29}$$

The elements of the inverse matrix $\overset{s}{\boldsymbol{\eta}}{}^{-1}$ are given by

$$\eta_{\mu\nu\,\sigma\tau} = \delta_{\mu\nu\,\sigma\tau}, \tag{3.30}$$

$$\eta^a{}_b{}^c{}_d = \tfrac{3}{2}(\delta^a{}_d\,\delta^c{}_b - \tfrac{1}{2}\delta^a{}_b\,\delta^c{}_d), \tag{3.31}$$

$$\eta_{\mu\nu}{}^a{}_b = \eta^a{}_{b\,\mu\nu} = 0, \tag{3.32}$$

and satisfy

$$\eta_{\mu\nu\,\rho\lambda}\eta_{\rho\lambda\,\sigma\tau} = \delta_{\mu\nu\,\sigma\tau}, \tag{3.33}$$

$$\eta^a{}_b{}^e{}_f\,\eta_e{}^f{}_c{}^d = \delta^a{}_{b\,c}{}^d, \tag{3.34}$$

where $\delta^a{}_{b\,c}{}^d$ is the unit matrix in the 3-dimensional vector space of traceless 2×2 matrices:

$$\delta^a{}_{b\,c}{}^d \stackrel{\text{def}}{=} \delta^a{}_c\,\delta_b{}^d - \tfrac{1}{2}\delta^a{}_b\,\delta_c{}^d. \tag{3.35}$$

In addition to the structure constants (3.19) and (3.25) of SO(7) ⊕ SL(2) we have, for $F(4)$, also the structure constants given by the matrix elements of the generators $\mathbf{G}_{\mu\nu}\times\mathbf{1}$ and $\mathbf{1}\times\mathbf{G}_a{}^b$ of spin $(7)\times$sl(2). Using indices from the first part of the Greek alphabet to denote components in the 8-dimensional spin space, we have

$$c_\alpha{}^a{}_{\mu\nu\,\beta b} = (\mathbf{G}_{\mu\nu}\times\mathbf{1})_\alpha{}^a{}_{\beta b} = \tfrac{1}{4}[\gamma_\mu, \gamma_\nu]_{\alpha\beta}\delta^a{}_b, \tag{3.36}$$

$$\begin{aligned} c_\alpha{}^a{}_c{}^d{}_{\beta b} &= (\mathbf{1}\times\mathbf{G}_c{}^d)_\alpha{}^a{}_{\beta b} \\ &= \delta_{\alpha\beta}(\delta_c{}^a\delta^d{}_b - \tfrac{1}{2}\delta_c{}^d\delta^a{}_b). \end{aligned} \tag{3.37}$$

We come now to the crucial role played by SL(2) in the structure of $F(4)$ [as well as $G(3)$, see below]. It consists in the following elementary fact: Every traceless 2×2 matrix is converted into a symmetric matrix through mulitplication by

$$\boldsymbol{\epsilon} \stackrel{\text{def}}{=} (\epsilon_{ab}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{3.38}$$

This allows us to choose (there is, in fact, no choice in the matter) for the matrix $\overset{A}{\boldsymbol{\eta}}$ in the present case

$$\eta_{\alpha a\,\beta b} = \delta_{\alpha\beta}\epsilon_{ab} = -\eta_{\beta b\,\alpha a}. \tag{3.39}$$

It is then straightforward to compute the remaining structure constants of $F(4)$:

$$c_\alpha{}^a{}_{\beta b\,\mu\nu} = -c_\alpha{}^a{}_{\mu\nu\,\beta b} = -\tfrac{1}{4}[\gamma_\mu, \gamma_\nu]_{\alpha\beta}\delta^a{}_b, \tag{3.40}$$

$$\begin{aligned} c_\alpha{}^a{}_{\beta b\,c}{}^d &= -c_\alpha{}^a{}_c{}^d{}_{\beta b} \\ &= -\delta_{\alpha\beta}(\delta_c{}^a\delta_b{}^d - \tfrac{1}{2}\delta_c{}^d\delta_b{}^a), \end{aligned} \tag{3.41}$$

$$\begin{aligned} c_{\mu\nu\,\alpha a\,\beta b} &= -\eta_{\mu\nu\,\sigma\tau}\eta_{\alpha a\,\gamma c}\,c_\gamma{}^c{}_{\sigma\tau\,\beta b} \\ &= -\tfrac{1}{4}[\gamma_\mu, \gamma_\nu]_{\alpha\beta}\,\epsilon_{ab} = c_{\mu\nu\,\beta b\,\alpha a}, \end{aligned} \tag{3.42}$$

$$\begin{aligned} c^c{}_{d\,\alpha a\,\beta b} &= \eta^c{}_d{}^e{}_f\eta_{\alpha a\gamma g}\,c_\gamma{}^g{}_e{}^f{}_{\beta b} \\ &= -\tfrac{3}{2}\delta_{\alpha\beta}(\epsilon_{ad}\delta^c{}_b - \tfrac{1}{2}\epsilon_{ab}\delta^c{}_d) \\ &= -\tfrac{3}{4}\delta_{\alpha\beta}(\delta^c{}_a\epsilon_{bd} + \delta^c{}_b\epsilon_{ad}) \\ &= c^c{}_{d\,\beta b\,\alpha a}. \end{aligned} \tag{3.43}$$

In order to verify the identity (2.19) form the combination

$$\tfrac{1}{4}\eta_{\mu\nu\,\sigma\tau}c_{\mu\nu\,a\alpha\,\beta b}c_{\sigma\tau\,\gamma c\,\delta d} + \eta_e{}^f{}_g{}^h c^e{}_{f\,a\alpha\,\beta b}c^g{}_{h\,\gamma c\,\delta d}$$
$$= \tfrac{1}{2}(\mathbf{G}_{\mu\nu})_{\alpha\beta}(\mathbf{G}_{\mu\nu})_{\gamma\delta}(\delta_{ac}\delta_{bd} - \delta_{ad}\delta_{bc}) + \tfrac{3}{4}(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc} - 2\delta_{ab}\delta_{cd}). \tag{3.44}$$

Next note that any 16×16 symmetric matrix $\mathbf{S}$ can be decomposed in the form

$$S_{a\alpha\,\beta b} = \delta_{\alpha\beta}S_{ab} + \tfrac{1}{6}(\mathbf{G}_{\mu\nu\sigma})_{\alpha\beta}T_{\mu\nu\sigma\,ab} + U_\mu(\gamma_\mu)_{\gamma\beta}\epsilon_{ab} + \tfrac{1}{2}V_{\mu\nu}(\mathbf{G}_{\mu\nu})_{\alpha\beta}\epsilon_{ab}, \tag{3.45}$$

where the S's, T's, U's, and V's are appropriate coefficients. A straightforward computation, which makes use of the identities (3.9)–(3.16), then leads to

$$(\tfrac{1}{4}\eta_{\mu\nu\,\sigma\tau}\, c_{\mu\nu\,a\alpha\,c\gamma}\, c_{\sigma\tau\,d\delta\,b\beta} + \eta_e{}^f{}_g{}^h\, c^e{}_{f\,a\alpha\,c\gamma}\, c^g{}_{h\,d\delta\,b\beta})S_{c\gamma\,d\delta}$$
$$= -4V_{\mu\nu}(\mathbf{G}_{\mu\nu})_{\alpha\beta}\epsilon_{ab} - 6\delta_{\alpha\beta}(S_{ab} - \delta_{ab}S_{cc}), \tag{3.46}$$

$$-\tfrac{1}{2}(\tfrac{1}{4}\eta_{\mu\nu\,\sigma\tau}c_{\mu\nu\,a\alpha\,\beta b}c_{\sigma\tau\,c\gamma\,d\delta} + \eta_e{}^f{}_g{}^h\, c^e{}_{f\,a\alpha\,\beta b}\, c^g{}_{h\,c\gamma\,d\delta})S_{c\gamma\,d\delta}$$
$$= -4V_{\mu\nu}(\mathbf{G}_{\mu\nu})_{\alpha\beta}\epsilon_{ab} - 6\delta_{\alpha\beta}(S_{ab} - \delta_{ab}S_{cc}). \tag{3.47}$$

The equality of the right sides of these equations shows that (2.19) is indeed satisfied. The reader who goes through the algebra will discover an intricate cancellation of terms, which comes about only because of the special choice that has been made for the matrix $\boldsymbol{\eta}$ [Eqs. (3.27), (3.28) and (3.29)].

Every element $\mathbf{X}$ of $F(4)$ can be associated with a 10×10 matrix:

$$\mathbf{X}\leftrightarrow\begin{pmatrix} \tfrac{1}{2}A_{\mu\nu}\mathbf{G}_{\mu\nu} & \mathbf{C}\boldsymbol{\epsilon} \\ \mathbf{C}^T & \mathbf{B}\end{pmatrix}. \tag{3.48}$$

$\mathbf{A}[=(A_{\mu\nu})]$ is an antisymmetric real 7×7 matrix, $\mathbf{B}$ is a traceless real 2×2 matrix, and $\mathbf{C}$ is a real 8×2 matrix. The elements of $\mathbf{A}$ and $\mathbf{B}$ are the components of $\mathbf{X}$ that lie in $\mathscr{A}_0$, and the elements of $\mathbf{C}$ are the components that lie in $\mathscr{A}_1$. The obvious relation of (3.48) to spin $(7)\times$sl(2) suggests that $F(4)$ has a relatively simple 10×10 matrix representation. Unfortunately, this is not so. The bracket relation for the superalgebra is not faithfully reproduced by the supercommutator of matrices of the form (3.48). That is why we are forced to work directly with the structure constants, i.e., with the 40×40 adjoint representation, in constructing $F(4)$.

Actually, the supercommutator does not fail by much. Denote by $\mathbf{X}_{12}$ the bracket of $\mathbf{X}_1$ and $\mathbf{X}_2$, and by $\mathbf{A}_{12}$, $\mathbf{A}_1$, $\mathbf{A}_2$, etc., the associated components. Then using the structure constants that have been constructed above, one finds

$$(\mathbf{A}_{12})_{\mu\nu} = [\mathbf{A}_1, \mathbf{A}_2]_{\mu\nu} - \tfrac{1}{2}\mathrm{tr}[\mathbf{G}_{\mu\nu}(\mathbf{C}_1\boldsymbol{\epsilon}\mathbf{C}_2^T + \mathbf{C}_2\boldsymbol{\epsilon}\mathbf{C}_1^T)], \tag{3.49}$$

$$\mathbf{B}_{12} = [\mathbf{B}_1, \mathbf{B}_2] + \tfrac{3}{4}(\mathbf{C}_1^T\mathbf{C}_2 + \mathbf{C}_2^T\mathbf{C}_1)\boldsymbol{\epsilon}, \tag{3.50}$$

$$\mathbf{C}_{12} = \tfrac{1}{2}(A_{1\,\mu\nu}\mathbf{G}_{\mu\nu}\mathbf{C}_2 - A_{2\,\mu\nu}\mathbf{G}_{\mu\nu}\mathbf{C}_1) - (\mathbf{C}_1\mathbf{B}_2^T - \mathbf{C}_2\mathbf{B}_1^T). \tag{3.51}$$

By virtue of (3.5), Eq. (3.49) may be rewritten in the form

$$\tfrac{1}{2}A_{12\,\mu\nu}\mathbf{G}_{\mu\nu} = \tfrac{1}{2}[\mathbf{A}_1, \mathbf{A}_2]_{\mu\nu}\mathbf{G}_{\mu\nu} + \mathbf{C}_1\boldsymbol{\epsilon}\mathbf{C}_2^T + \mathbf{C}_2^T\boldsymbol{\epsilon}\mathbf{C}_1 - \tfrac{1}{8}\gamma_\mu\mathrm{tr}[\gamma_\mu(\mathbf{C}_1\boldsymbol{\epsilon}\mathbf{C}_2^T + \mathbf{C}_2^T\boldsymbol{\epsilon}\mathbf{C}_1]. \tag{3.52}$$

Equations (3.49)–(3.51) show that the bracket operation for the elements of $F(4)$ may be recovered by taking the supercommutator of the matrices (3.48) and then subtracting the block matrix

$$\begin{pmatrix} \tfrac{1}{8}\gamma_\mu\mathrm{tr}[\gamma_\mu(\mathbf{C}_1\boldsymbol{\epsilon}\mathbf{C}_2^T + \mathbf{C}_2\boldsymbol{\epsilon}\mathbf{C}_1^T)] & 0 \\ 0 & \tfrac{1}{4}(\mathbf{C}_1^T\mathbf{C}_2 + \mathbf{C}_2^T\mathbf{C}_1)\boldsymbol{\epsilon}\end{pmatrix}. \tag{3.53}$$

Because the elements of $F(4)$ are so naturally assembled into the matrix (3.48) and because the supercommutator of this matrix does give correctly *some* of the bracket relations for the superalgebra, we propose to give the name "matrix *pseudo*representation" to what we have constructed here. In Sec. 6 an alternative, partly heuristic, derivation of this pseudorepresentation is given, in slightly different notation.

4. STRUCTURE OF $G(3)$

For the $G(3)$ superalgebra $\mathscr{A}_0$ is $G_2\oplus$ SL(2) and $\mathscr{A}_0$: $\mathscr{A}_1$ is $g_2\times$sl(2), where g_2 is the 7-dimensional fundamental representation (representation of lowest order) of the exceptional Lie algebra G_2. $\mathscr{A}_0$ has dimension 17 and rank 3, and the dimensionality of $\mathscr{A}_1$ is 14.

G_2 is often defined as the subalgebra of SO(7) generated by matrices of the form $\tfrac{1}{2}A_{\mu\nu}\mathbf{G}_{\mu\nu}$ having zeros in the eighth row and column. The matrices themselves constitute g_2. This way of defining G_2 and g_2 turns out to be rather cumbersome for applications. We give here a simpler description. (For additional details see Sec. 5.)

Let $\mathbf{X}$ be an element of G_2. $\mathbf{X}$ has 14 real components. In an appropriate basis 8 of these components may be assembled into the real and imaginary parts of the elements of a traceless anti-Hermitian 3×3 matrix $\mathbf{A}$, and the remaining 6 components may be assembled into the real and imaginary parts of a complex 3×1 matrix or 3-vector $\mathbf{V}$. If $\mathbf{A}$ and $\mathbf{V}$ themselves are assembled into the traceless anti-Hermitian 7×7 matrix

$$D(\mathbf{X}) = \begin{pmatrix} \mathbf{A} & (1/\sqrt{2})\bar{\boldsymbol{\epsilon}}\cdot\mathbf{V}^* & \mathbf{V} \\ (1/\sqrt{2})\boldsymbol{\epsilon}\cdot\mathbf{V} & \mathbf{A}^* & \mathbf{V}^* \\ -\mathbf{V}^\dagger & -\mathbf{V}^T & 0\end{pmatrix}, \tag{4.1}$$

then

$$[D(\mathbf{X}_1), D(\mathbf{X}_2)] = D([\mathbf{X}_1, \mathbf{X}_2]), \tag{4.2}$$

where $[\mathbf{X}_1, \mathbf{X}_2]$ is the bracket operation for G_2. Here $\bar{\epsilon}$ and ϵ denote the antisymmetric tensors $\epsilon^{\mu\nu\sigma}$ and $\epsilon_{\mu\nu\sigma}$, respectively. $\epsilon\cdot V$ and $\epsilon\cdot V^*$ stand for the 3×3 matrices $(\epsilon_{\mu\sigma\nu}V^\sigma)$ and $(\epsilon^{\mu\sigma\nu}V^*{}_\sigma)$. The asterisk denotes complex conjugation and effects a raising or lowering of indices according to the rules

$$V^{\mu *} = V^*{}_\mu, \quad A^\mu{}_\nu{}^* = A^*{}_\mu{}^\nu = -A^\nu{}_\mu. \tag{4.3}$$

The matrices (4.1) constitute an explicit realization of g_2, which allows a direct computation of the structure constants of G_2. Denote by $\mathbf{X}_{12}$ the bracket of $\mathbf{X}_1$ and $\mathbf{X}_2$ and by $\mathbf{A}_{12}$, $\mathbf{V}_{12}$, $\mathbf{A}_1$, $\mathbf{V}_1$, etc., the associated components. Then Eq. (4.2) yields

$$\mathbf{A}_{12} = [\mathbf{A}_1, \mathbf{A}_2] + \tfrac{1}{2}(\bar{\epsilon}\cdot\mathbf{V}_1^*\epsilon\cdot\mathbf{V}_2 - \bar{\epsilon}\cdot\mathbf{V}_2^*\epsilon\cdot\mathbf{V}_1) - (\mathbf{V}_1\mathbf{V}_2^\dagger - \mathbf{V}_2\mathbf{V}_1^\dagger), \tag{4.4}$$

$$\mathbf{V}_{12} = \mathbf{A}_1\mathbf{V}_2 - \mathbf{A}_2\mathbf{V}_1 + \sqrt{2}\bar{\epsilon}\cdot\mathbf{V}_1^*\mathbf{V}_2^*, \tag{4.5}$$

$$\mathbf{V}^*_{12} = \mathbf{A}_1^*\mathbf{V}_2^* - \mathbf{A}_2^*\mathbf{V}_1^* + \sqrt{2}\epsilon\cdot\mathbf{V}_1\mathbf{V}_2$$
$$= -\mathbf{A}_1^T\mathbf{V}_2^* + \mathbf{A}_2^T\mathbf{V}_1^* + \sqrt{2}\epsilon\cdot\mathbf{V}_1\mathbf{V}_2. \tag{4.6}$$

Use of the identity

$$\epsilon^{\mu\nu\rho}\epsilon_{\rho\sigma\tau} = \delta^\mu{}_\sigma\delta^\nu{}_\tau - \delta^\mu{}_\tau\delta^\nu{}_\sigma, \tag{4.7}$$

permits Eq. (4.4) to be rewritten in the manifestly traceless form

$$\mathbf{A}_{12} = [\mathbf{A}_1, \mathbf{A}_2] - \tfrac{3}{2}(\mathbf{V}_1\mathbf{V}_2^\dagger - \mathbf{V}_2\mathbf{V}_1^\dagger)$$
$$- \tfrac{1}{2}\mathbf{1}(\mathbf{V}_1^\dagger\mathbf{V}_2 - \mathbf{V}_2^\dagger\mathbf{V}_1). \tag{4.8}$$

Multiplication of the identity

$$0 \equiv \epsilon_{\mu\sigma\tau}A^\tau{}_\nu - \epsilon_{\sigma\tau\nu}A^\tau{}_\mu + \epsilon_{\tau\nu\mu}A^\tau{}_\sigma - \epsilon_{\nu\mu\sigma}A^\tau{}_\tau$$
$$= \epsilon_{\mu\sigma\tau}A^\tau{}_\nu - A^*{}_\mu{}^\tau\epsilon_{\tau\sigma\nu} + \epsilon_{\mu\tau\nu}A^\tau{}_\sigma, \tag{4.9}$$

by V^σ yields the relation

$$(\epsilon\cdot\mathbf{V})\mathbf{A} - \mathbf{A}^*(\epsilon\cdot\mathbf{V}) = -\epsilon\cdot(\mathbf{AV}), \tag{4.10}$$

which allows the entry in the second row and first column of $D(\mathbf{X}_{12})$ to be expressed in the form

$$\frac{1}{\sqrt{2}}\epsilon\cdot(\mathbf{A}_1\mathbf{V}_2 - \mathbf{A}_2\mathbf{V}_1) - (\mathbf{V}_1^*\mathbf{V}_2^\dagger - \mathbf{V}_2^*\mathbf{V}_1^\dagger) = \frac{1}{\sqrt{2}}\epsilon\cdot\mathbf{V}_{12}, \tag{4.11}$$

as is necessary for consistency.

The structure constants of G_2 are defined by

$$X_{12}{}^A = c^A{}_{BC}X_1{}^BX_2{}^C. \tag{4.12}$$

Replacing X^A by $A^\mu{}_\nu$, V^μ and $V^*{}_\mu$, and keeping only the nonvanishing structure constants, we have

$$A_{12}{}^\mu{}_\nu = c^\mu{}_{\nu\sigma}{}^\tau{}_\rho{}^\lambda A_1{}^\sigma{}_\tau A_2{}^\rho{}_\lambda$$
$$+ c^\mu{}_{\nu\sigma}{}^\tau V_1{}^\sigma V^*_{2\tau} + c^\mu{}_\nu{}^\sigma{}_\tau V^*_{1\sigma}V_2{}^\tau, \tag{4.13}$$

$$V_{12}{}^\mu = c^\mu{}_\nu{}^\sigma{}_\tau A_1{}^\nu{}_\sigma V_2{}^\tau + c^\mu{}_{\nu\sigma}{}^\tau V_1{}^\nu A_2{}^\sigma{}_\tau$$
$$+ c^{\mu\nu\sigma}V_1{}^*{}_\nu V_2{}^*{}_\sigma, \tag{4.14}$$

$$V_{12}{}^*{}_\mu = c_{\mu\nu}{}^{\sigma\tau}A_1{}^\nu{}_\sigma V_2{}^*{}_\tau + c_\mu{}^\nu{}_\sigma{}^\tau V_1{}^*{}_\nu A_2{}^\sigma{}_\tau$$
$$+ c_{\mu\nu\sigma}V_1{}^\nu V_2{}^\sigma, \tag{4.15}$$

which, on comparison with Eqs. (4.5), (4.6), and (4.8), yield

$$c^\mu{}_{\nu\sigma}{}^\tau{}_\rho{}^\lambda = \delta^\mu{}_\sigma\delta_\nu{}^\lambda\delta^\tau{}_\rho - \delta^\mu{}_\rho\delta_\nu{}^\sigma\delta^\lambda{}_\sigma, \tag{4.16}$$

$$c^\mu{}_{\nu\sigma}{}^\tau = -c^\mu{}_\nu{}^\tau{}_\sigma = -\tfrac{3}{2}(\delta^\mu{}_\sigma\delta^\tau{}_\nu - \tfrac{1}{3}\delta^\mu{}_\nu\delta_\sigma{}^\tau), \tag{4.17}$$

$$c^\mu{}_\nu{}^\sigma{}_\tau = -c^\mu{}_{\tau\nu}{}^\sigma = \delta^\mu{}_\nu\delta^\sigma{}_\tau - \tfrac{1}{3}\delta^\mu{}_\tau\delta_\nu{}^\sigma, \tag{4.18}$$

$$c_{\mu\nu}{}^{\sigma\tau} = -c_\mu{}^\tau{}_\nu{}^\sigma = -\delta_\mu{}^\sigma\delta_\nu{}^\tau + \tfrac{1}{3}\delta_\mu{}^\tau\delta_\nu{}^\sigma, \tag{4.19}$$

$$c^{\mu\nu\sigma} = \sqrt{2}\epsilon^{\mu\nu\sigma}, \tag{4.20}$$

$$c_{\mu\nu\sigma} = \sqrt{2}\epsilon_{\mu\nu\sigma}. \tag{4.21}$$

The terms involving the factor $\frac{1}{3}$ in expressions (4.18) and (4.19) arise from application of the projection operator onto the subspace of traceless matrices in the space of 3×3 matrices.

The Cartan–Killing matrix for G_2 is readily computed:

$$c^\iota{}_{\kappa\mu}{}^\nu{}_\rho{}^\lambda c^\rho{}_{\lambda\sigma}{}^\tau{}_\iota{}^\kappa - c^\rho{}_\mu{}^\nu{}_\lambda c^\lambda{}_\sigma{}^\tau{}_\rho - c_{\rho\mu}{}^{\nu\lambda}c_{\lambda\sigma}{}^{\tau\rho}$$
$$= -8(\delta_\mu{}^\tau\delta^\nu{}_\sigma - \tfrac{1}{3}\delta_\mu{}^\nu\delta_\sigma{}^\tau), \tag{4.22}$$

$$-c^\sigma{}_{\tau\mu}{}^\rho c_\rho{}^\nu{}_\sigma{}^\tau - c^\sigma{}_{\mu\tau}{}^\rho c^\tau{}_\rho{}^\nu{}_\sigma - c_{\sigma\mu\tau}c^{\tau\nu\sigma} = 12\delta_\mu{}^\nu, \tag{4.23}$$

$$-c^\sigma{}_\tau{}^\mu{}_\rho c^\rho{}_{\nu\sigma}{}^\tau - c_\sigma{}^\mu{}_\tau{}^\rho c^\tau{}_{\rho\nu}{}^\sigma - c^{\sigma\mu\tau}c_{\tau\nu\sigma} = 12\delta^\mu{}_\nu. \tag{4.24}$$

When this matrix is multiplied by a scale factor, expressions (4.22), (4.23), and (4.24) must all scale together.

The structure constants and Cartan–Killing matrix for SL(2) are given in Sec. 3, Eqs. (3.25) and (3.26). It turns out that in order to satisfy the identity (2.14) we shall need, for $G(3)$, a matrix $\overset{s}{\boldsymbol{\eta}}$ that contains the Cartan–Killing matrices of G_2 and SL(2) scaled by factors $\frac{1}{8}$ and $-\frac{3}{16}$ respectively:

$$\eta_\mu{}^\nu{}_\sigma{}^\tau = -(\delta_\mu{}^\tau\delta_\sigma{}^\nu - \tfrac{1}{3}\delta_\mu{}^\nu\delta_\sigma{}^\tau), \tag{4.25}$$

$$\eta_\mu{}^\nu = \tfrac{3}{2}\delta_\mu{}^\nu, \tag{4.26}$$

$$\eta^\mu{}_\nu = \tfrac{3}{2}\delta^\mu{}_\nu, \tag{4.27}$$

$$\eta_a{}^b{}_c{}^d = \tfrac{3}{4}(\delta_a{}^d\delta_c{}^b - \tfrac{1}{2}\delta_a{}^b\delta_c{}^d). \tag{4.28}$$

The elements of the inverse matrix $\overset{s}{\boldsymbol{\eta}}{}^{-1}$ are

$$\eta^\mu{}_\nu{}^\sigma{}_\tau = -(\delta^\mu{}_\tau\delta^\sigma{}_\nu - \tfrac{1}{3}\delta^\mu{}_\nu\delta^\sigma{}_\tau), \tag{4.29}$$

$$\eta^{-1\,\mu}{}_\nu = \tfrac{2}{3}\delta^\mu{}_\nu, \tag{4.30}$$

$$\eta^{-1}{}_\mu{}^\nu = \tfrac{2}{3}\delta_\mu{}^\nu, \tag{4.31}$$

$$\eta^a{}_b{}^c{}_d = \tfrac{4}{3}(\delta^a{}_d\delta^c{}_b - \tfrac{1}{2}\delta^a{}_b\delta^c{}_d). \tag{4.32}$$

Let the indices ${}^\mu$, ${}_\mu$, and 0 label, respectively, the first three rows, the second three rows and the last row of the matrix (4.1), and let the indices ${}_\nu$, ${}^\nu$, and 0 label, respectively, the first three columns, the second three columns and the last column. We can read off directly from Eqs. (3.37) and (4.1) the nonvanishing matrix elements of the generators of the extending representation $g_2\times sl(2)$:

$$c^{\mu a}{}_\sigma{}^\tau{}_{\nu b} = (\delta^\mu{}_\sigma\delta^\tau{}_\nu - \tfrac{1}{3}\delta^\mu{}_\nu\delta_\sigma{}^\tau)\delta^a{}_b = -c^{\mu a}{}_{\nu b\sigma}{}^\tau, \tag{4.33}$$

$$c_\mu{}^a{}_\sigma{}^{\tau\nu}{}_b = -(\delta_\mu{}^\tau\delta_\sigma{}^\nu - \tfrac{1}{3}\delta_\mu{}^\nu\delta_\sigma{}^\tau)\delta^a{}_b = -c_\mu{}^{a\nu}{}_{b\sigma}{}^\tau, \tag{4.34}$$

$$c^{\mu a\sigma\nu}{}_b = (1/\sqrt{2})\epsilon^{\mu\sigma\nu}\delta^a{}_b = -c^{\mu a\nu}{}_b{}^\sigma, \tag{4.35}$$

$$c_\mu{}^{a\sigma}{}_{0b} = \delta_\mu{}^\sigma\delta^a{}_b = -c_\mu{}^a{}_{0b}{}^\sigma, \tag{4.36}$$

$$c_0{}^{a\sigma}{}_{\nu b} = -\delta^\sigma{}_\nu\delta^a{}_b = -c_0{}^a{}_{\nu b}{}^\sigma, \tag{4.37}$$

$$c^{\mu a}{}_{\sigma 0b} = \delta^\mu{}_\sigma\delta^a{}_b = -c^{\mu a}{}_{0b\sigma}, \tag{4.38}$$

$$c_\mu{}^a{}_{\sigma\nu b} = (1/\sqrt{2})\epsilon_{\mu\sigma\nu}\delta^a{}_b = -c_\mu{}^a{}_{\nu b\sigma}, \tag{4.39}$$

$$c_0{}^a{}_\sigma{}^\nu{}_b = -\delta_\sigma{}^\nu\delta^a{}_b = -c_0{}^{a\nu}{}_{b\sigma}, \tag{4.40}$$

$$c^{\mu a}{}_c{}^d{}_{\nu b} = \delta^\mu{}_\nu(\delta^a{}_c\delta^d{}_b - \tfrac{1}{2}\delta^a{}_b\delta_c{}^d)$$
$$= -c^{\mu a}{}_{\nu bc}{}^d, \tag{4.41}$$

$$c_\mu{}^a{}_c{}^{d\nu}{}_b = \delta_\mu{}^\nu(\delta^a{}_c\delta^d{}_b - \tfrac{1}{2}\delta^a{}_b\delta_c{}^d)$$
$$= -c_\mu{}^{a\nu}{}_{bc}{}^d, \tag{4.42}$$

$$c_0{}^a{}_c{}^d{}_{0b} = \delta^a{}_c\delta^d{}_b - \tfrac{1}{2}\delta^a{}_b\delta_c{}^d$$
$$= -c_0{}^a{}_{0bc}{}^d. \tag{4.43}$$

Expressions (3.25), (4.16)–(4.21), and (4.33)–(4.43) constitute for $G(3)$, the structure constants that were called $c^\mu{}_{\nu\sigma}$, $c^\alpha{}_{\mu\beta}$, $c^\alpha{}_{\beta\mu}$ in Sec. 2. To obtain the remaining structure constants, denoted by $c^\mu{}_{\alpha\beta}$ in Sec. 2, we need to apply the matrix $\overset{s}{\boldsymbol{\eta}}{}^{-1}$ [Eqs. (4.29)–(4.32)] as well as the matrix $\overset{A}{\boldsymbol{\eta}}$, which in the present case we choose to be [cf. Eq. (3.39)]

$$\eta^\mu{}_{a\nu b} = \delta^\mu{}_\nu\epsilon_{ab} = -\eta_{\nu b}{}^\mu{}_a, \tag{4.44}$$

$$\eta_{0a0b} = \epsilon_{ab}. \tag{4.45}$$

This yields

$$
\begin{aligned}
c^{\mu}{}_{\nu}{}^{\sigma}{}_{a\,\tau b} &= c^{\mu}{}_{\nu\,\tau b}{}^{\sigma}{}_{a} \\
&= -\eta^{\mu}{}_{\nu}{}^{\rho}{}_{\lambda}\,\eta^{\sigma}{}_{a\,\kappa c}c^{\kappa c}{}_{\rho}{}^{\lambda}{}_{\tau b} \\
&= (\delta^{\mu}{}_{\tau}\delta_{\nu}{}^{\sigma} - \tfrac{1}{3}\delta^{\mu}{}_{\nu}\delta^{\sigma}{}_{\tau})\epsilon_{ab},
\end{aligned} \tag{4.46}
$$

$$
\begin{aligned}
c^{\mu\nu}{}_{a}{}^{\sigma}{}_{b} &= -\eta^{-1\mu}{}_{\tau\eta}{}^{\nu}{}_{a\,\rho c}c^{\rho c\,\tau\sigma}{}_{b} \\
&= (\sqrt{2/3})\epsilon^{\mu\nu\sigma}\epsilon_{ab},
\end{aligned} \tag{4.47}
$$

$$
\begin{aligned}
c_{\mu\,\nu a\,\sigma b} &= -\eta^{-1}{}_{\mu}{}^{\tau}\eta_{\nu a}{}^{\rho}{}_{c}\,c_{\rho}{}^{c}{}_{\tau\sigma b} \\
&= (\sqrt{2/3})\epsilon_{\mu\nu\sigma}\epsilon_{ab},
\end{aligned} \tag{4.48}
$$

$$
\begin{aligned}
c^{\mu}{}_{\nu a\,0b} &= c^{\mu}{}_{0b\,\nu a} = -\eta^{-1\mu}{}_{\tau}\,\eta_{\nu a}{}^{\sigma}{}_{c}c_{\sigma}{}^{c\tau}{}_{0b} \\
&= -\tfrac{2}{3}\,\delta^{\mu}{}_{\nu}\epsilon_{ab},
\end{aligned} \tag{4.49}
$$

$$
\begin{aligned}
c_{\mu}{}^{\nu}{}_{a\,0b} &= c_{\mu\,0b}{}^{\nu}{}_{a} = -\eta^{-1}{}_{\mu}{}^{\tau}\eta^{\nu}{}_{a\,\sigma c}\,c^{\sigma c}{}_{\tau\,0b} \\
&= -\tfrac{2}{3}\delta_{\mu}{}^{\nu}\epsilon_{ab},
\end{aligned} \tag{4.50}
$$

$$
\begin{aligned}
c^{a}{}_{b}{}^{\mu}{}_{c\,\nu d} &= c^{a}{}_{b\,\nu d}{}^{\mu}{}_{c} \\
&= -\eta^{a}{}_{b}{}^{e}{}_{f}\eta^{\mu}{}_{c\,\sigma g}c^{\sigma g}{}_{e}{}^{f}{}_{\nu d} \\
&= -\tfrac{2}{3}\delta^{\mu}{}_{\nu}(\delta^{a}{}_{c}\epsilon_{db} + \delta^{a}{}_{d}\epsilon_{cb}),
\end{aligned} \tag{4.51}
$$

$$
\begin{aligned}
c^{a}{}_{b\,0c\,0d} &= -\eta^{a}{}_{b}{}^{e}{}_{f}\eta_{0c\,0g}c_{0}{}^{g}{}_{e}{}^{f}{}_{\sigma d} \\
&= -\tfrac{2}{3}(\delta^{a}{}_{c}\epsilon_{db} + \delta^{a}{}_{d}\epsilon_{cb}).
\end{aligned} \tag{4.52}
$$

It is now a straightforward computation to check that the identity (2.14) holds:

$$
\begin{aligned}
&\eta_{\rho}{}^{\lambda}{}_{\iota}{}^{\kappa}c^{\rho}{}_{\lambda}{}^{\mu}{}_{a\,\nu b}c^{\iota}{}_{\kappa}{}^{\sigma}{}_{c\,\tau d} + \eta_{\rho}{}^{\lambda}c^{\rho\mu}{}_{a\,\nu b}c_{\lambda}{}^{\sigma}{}_{c\,\tau\delta} \\
&\quad + \eta_{e}{}^{f}{}_{g}{}^{h}c^{e}{}_{f}{}^{\mu}{}_{a\,\nu b}c^{g}{}_{h}{}^{\sigma}{}_{c\,\tau d} + \mathrm{cyc}\,({}_{\nu b},{}^{\sigma}{}_{c},{}_{\tau d}) \\
&= \tfrac{1}{3}(\delta^{\mu}{}_{\nu}\delta^{\sigma}{}_{\tau} - \delta^{\mu}{}_{\tau}\delta^{\sigma}{}_{\nu})(\epsilon_{ab}\epsilon_{cd} + \epsilon_{ac}\epsilon_{db} + \epsilon_{ad}\epsilon_{bc}) = 0,
\end{aligned} \tag{4.53}
$$

$$
\begin{aligned}
&\eta_{\tau}{}^{\rho}c^{\tau}{}_{\mu a\,0b}c_{\rho}{}^{\nu}{}_{c\,0d} + \eta_{e}{}^{f}{}_{g}{}^{h}c^{e}{}_{f\,\mu a\,0b}c^{g}{}_{h}{}^{\nu}{}_{c\,0d} \\
&\quad + \mathrm{cyc}({}_{0b},{}^{\nu}{}_{c},{}_{0d}) = 0,
\end{aligned} \tag{4.54}
$$

$$
\begin{aligned}
&\eta_{\tau}{}^{\rho}c^{\tau}{}_{0a\,\mu b}c_{\rho}{}^{\nu}{}_{c\,0d} + \eta^{\tau}{}_{\rho}c_{\tau\,0a\,\mu b}c^{\rho\nu}{}_{c\,0d} \\
&\quad + \eta_{e}{}^{f}{}_{g}{}^{h}c^{e}{}_{f\,0a\,\mu b}c^{g}{}_{h}{}^{\nu}{}_{c\,0d} + \mathrm{cyc}({}_{\mu b},{}^{\nu}{}_{c},{}_{0d}) = 0,
\end{aligned} \tag{4.55}
$$

$$
\eta_{e}{}^{f}{}_{g}{}^{h}c^{e}{}_{f\,0a\,0b}c^{g}{}_{h\,0c\,0d} + \mathrm{cyc}({}_{0b,0c,0d}) = 0. \tag{4.56}
$$

It is also straightforward to verify that Eq. (2.22) holds with $\lambda = 4$.

Just as every element of $F(4)$ can be associated with a 10×10 matrix so every element **X** of $G(3)$ can be associated with a 9×9 matrix

$$
\mathbf{X} \leftrightarrow \begin{pmatrix} \mathbf{A} & (1/\sqrt{2})\bar{\epsilon}\mathbf{V}^* & \mathbf{V} & \mathbf{C}\epsilon \\ \frac{1}{\sqrt{2}}\epsilon\cdot\mathbf{V} & \mathbf{A}^* & \mathbf{V}^* & \mathbf{C}^*\epsilon \\ -\mathbf{V}^{\dagger} & -\mathbf{V}^{T} & 0 & \mathbf{D}\epsilon \\ \mathbf{C}^{\dagger} & \mathbf{C}^{T} & \mathbf{D}^{T} & \mathbf{B} \end{pmatrix}. \tag{4.57}
$$

A and **V** are the matrices appearing in Eq. (4.1). **B** is a traceless real 2×2 matrix, **C** is a complex 3×2 matrix and **D** is a real 1×2 matrix. The 17 independent real and imaginary parts of the elements of **A**, **V**, and **B** are the components of **X** that lie in $\mathscr{A}_0$. The 14 independent real and imaginary parts of the elements of **C** and **D** are the components of **X** that lie in $\mathscr{A}_1$.

Using indices ${}^{\mu}$, ${}_{\mu}$, 0, and ^{a} to label the rows of the matrix (4.5) and indices ${}_{\nu}$, ${}^{\nu}$, 0, and ${}_{b}$ to label the columns, one can read off from the structure constants the bracket relations for $F(4)$:

$$
\begin{aligned}
\mathbf{A}_{12} &= [\mathbf{A}_1, \mathbf{A}_2] - \tfrac{3}{2}(\mathbf{V}_1\mathbf{V}_2^{\dagger} - \mathbf{V}_2\mathbf{V}_1^{\dagger}) - \tfrac{1}{2}\mathbf{1}(\mathbf{V}_1^{\dagger}\cdot\mathbf{V}_2 - \mathbf{V}_2^{\dagger}\cdot\mathbf{V}_1) \\
&\quad + \mathbf{C}_1\epsilon\mathbf{C}_2^{\dagger} + \mathbf{C}_2\epsilon\mathbf{C}_1^{\dagger} - \tfrac{1}{3}\mathbf{1}(\mathbf{C}_1\epsilon\mathbf{C}_2^{\dagger} + \mathbf{C}_2\epsilon\mathbf{C}_1^{\dagger}),
\end{aligned} \tag{4.58}
$$

$$
\begin{aligned}
\mathbf{V}_{12} &= \mathbf{A}_1\mathbf{V}_2 - \mathbf{A}_2\mathbf{V}_1 + \sqrt{2}\bar{\epsilon}:\mathbf{V}_1^*\mathbf{V}_2^{\dagger} + \tfrac{\sqrt{2}}{3}\bar{\epsilon}:\mathbf{C}_1^*\epsilon\mathbf{C}_2{}^{\dagger} \\
&\quad + \tfrac{2}{3}(\mathbf{C}_1\epsilon\mathbf{D}_2^{T} + \mathbf{C}_2\epsilon\mathbf{D}_1^{T}),
\end{aligned} \tag{4.59}
$$

$$
\begin{aligned}
\mathbf{B}_{12} &= [\mathbf{B}_1, \mathbf{B}_2] + \tfrac{2}{3}(\mathbf{C}_1^{\dagger}\mathbf{C}_2 + \mathbf{C}_2^{\dagger}\mathbf{C}_1 + \mathbf{C}_1^{T}\mathbf{C}_2^* \\
&\quad + \mathbf{C}_2^{T}\mathbf{C}_1^* + \mathbf{D}_1^{T}\mathbf{D}_2 + \mathbf{D}_2^{T}\mathbf{D}_1)\epsilon,
\end{aligned} \tag{4.60}
$$

$$
\begin{aligned}
\mathbf{C}_{12} &= [\mathbf{A}_1, \mathbf{C}_2] - [\mathbf{A}_2, \mathbf{C}_1] + (1/\sqrt{2})\bar{\epsilon}:(\mathbf{V}_1^*\mathbf{C}_2^* - \mathbf{V}_2^*\mathbf{C}_1^*) \\
&\quad + \mathbf{V}_1\mathbf{D}_2 - \mathbf{V}_2\mathbf{D}_1 - \mathbf{C}_1\mathbf{B}_2^{T} + \mathbf{C}_2\mathbf{B}_1^{T},
\end{aligned} \tag{4.61}
$$

$$
\mathbf{D}_{12} = -\mathbf{V}_1^{\dagger}\mathbf{C}_2 + \mathbf{V}_2^{\dagger}\mathbf{C}_1 - \mathbf{V}_1^{T}\mathbf{C}_2^* + \mathbf{V}_2^{T}\mathbf{C}_1 - \mathbf{D}_1\mathbf{B}_2^{T} + \mathbf{D}_2\mathbf{B}_1^{T}. \tag{4.62}
$$

The double dot appearing in some of the terms above indicates that contractions are to be performed over two pairs of dummy indices. The reader can easily compute by how much these bracket relations fail to be given by the supercommutator itself. The matrices (4.57) do *not* yield a matrix representation of $G(3)$, but rather, once again, a pseudorepresentation.

5. MATRIX REPRESENTATIONS FOR E_8, F_4, and G_2

In this section we use the work of Cremmer and Julia,[10] who gave a simple matrix representation for E_7. Their work was extended to E_6 in Ref. 11, and we here extend their work to E_8, F_4, and G_2 so that one now has simple matrix representations of all the exceptional groups, in a form suitable for physical applications.

Suppose one wants to find a matrix form for a linear representation R of an algebra A which acts on coordinates x^i. One first selects a maximal subalgebra H so that $A = H + K$, and determines how the x^i decompose into irreducible multiplets of H. For example, in the case of G_2, one maximal subgroup is $H = \mathrm{SU}(3)$ and x^i ($i = 1, 7$) decompose into $3 + \bar{3} + 1$. (See the preceding section.) Usually there are only a few decompositions possible, and out of these the correct one is always obvious. Since H is a maximal subalgebra, $A = H + K$ is a Cartan decomposition which means that $[H, K] \subset K$ and $[K, K] \subset H$. The former means also that the parameters of K are representations of H. In the example of G_2, there are 8 parameters for SU(3) while the remaining 6 can only be in $3 + \bar{3}$ since the 7 of G_2 is real (as are all its representations). The $[K, K] \subset H$ relation is a test on the correctness of the choices made so far.

Different choices for the maximal subalgebra H lead to matrices which look quite different, and only one particular form may be useful for a physical application, but all these matrices are equivalent, of course.

Although the separation of A into $H + K$ looks like a coset approach, it is not. For example, G_2 has 14 generators and $H = \mathrm{SU}(3)$, so that there exists a 6-dimensional *nonlinear* representation of G on the coset space $A - H$. Only H acts linearly on the coset space, but by adding one extra dimension in this example, one finds a 7-dimensional linear representation of the full G_2. We now consider E_8, F_4, and G_2.

E_8

The exceptional Lie algebra E_8 has 248 generators, its lowest dimensional representation is the adjoint representation, and is thus real. O(16) is a maximal compact subgroup.

Since spin (16) is 128 dimensional, one expects a representation of E_8 on an antisymmetric tensor $\phi^{ij} = -\phi^{ji}$ ($i, j = 1, ..., 16$) which has 120 components, and on a spinor ψ^α ($\alpha = 1, ..., 256$). Since in $d = 16$ dimensions the Dirac matrices $\Gamma^j(256\times 256)$ with $j = 1, ..., 16$ can be chosen real, and one can impose in even dimensions the Weyl condition on spinors

$$\psi^\alpha = (\Gamma^1\Gamma^2...\Gamma^{16})^\alpha{}_\beta \psi^\beta{}_\alpha \equiv \Gamma^{17\alpha}{}_\beta \psi^\beta, \tag{5.1}$$

it follows that ψ^α has 128 independent components. Thus we expect that ϕ^{ij} and ψ^α represent the fundamental representation. As parameters one has, of course, the 120 parameters Λ^{ij} of SO(16), plus 128 other parameters. Clearly, these could form a real spinor λ^α which satisfies again the Weyl condition. Therefore we write

$$\delta\phi^{ij} = (\Lambda^i{}_k \phi^{kj} + \Lambda^j{}_k \phi^{ik}) + (\lambda^\alpha \Gamma^{ij}_{\alpha\beta}\psi^\beta),$$
$$\delta\psi^\alpha = (\tfrac{1}{4}\Lambda^{ij}\Gamma_{ij}{}^\alpha{}_\beta\psi^\beta) + (\phi^{ij}\Gamma_{ij}{}^\alpha{}_\beta\lambda^\beta). \tag{5.2}$$

Clearly, the Weyl condition on λ is the same as on ψ. Since the Dirac matrices are real and Hermitian, they are symmetric, so that the charge conjugation matrix C, defined by $C\gamma_\mu C^{-1} = \pm\gamma_\mu^T$, is unity if one chooses the positive sign.[12] (For the negative sign choose Γ^{17}.) Thus one needs no Dirac (or rather Majorana) bar on the spinors and this proves the covariance of the transformation rules under O(16).

The nontrivial commutators are those with one or two λ parameters. For the former one easily finds that λ is indeed a spinorial parameter, as shown by the following commutator

$$[\delta(\lambda), \delta(\Lambda)] = \delta[\lambda' = \tfrac{1}{4}\Lambda_{ij}\Gamma^{ij}\lambda], \tag{5.3}$$

which holds both when acting on ϕ^{ij} and when acting on ψ^α.

The (λ, λ) commutator on ϕ^{ij} is easy to evaluate. Suppressing spinor indices and defining $\phi\cdot\Gamma = \phi_{ij}\Gamma^{ij}$ one has

$$[\delta(\lambda_2), \delta(\lambda_1)]\phi^{ij} = \lambda_1\Gamma^{ij}\phi\cdot\Gamma\lambda_2 - 1\leftrightarrow 2. \tag{5.4}$$

Since the Γ^{ij} are antisymmetric, one finds easily (assuming that λ_1 and λ_2 commute)

$$[\delta(\lambda_2), \delta(\lambda_1)]\phi^{ij} = \delta[\Lambda'_{ij} = 4\lambda_2\Gamma_{ij}\lambda_1]\phi^{ij}. \tag{5.5}$$

On ψ^α, however, the $\lambda\lambda$ commutator is more interesting to evaluate. One starts from its definition

$$[\delta(\lambda_2), \delta(\lambda_1)]\psi = (\lambda_2\Gamma^{ij}\psi)(\Gamma^{ij}\lambda_1) - 1\leftrightarrow 2. \tag{5.6}$$

The product of all the Γ^i forms an orthogonal and complete set, given by

$$O^I = \{1, \Gamma^i, \Gamma^{i_1i_2}, ..., \Gamma^{i_1...i_8}, \Gamma^{17}, \Gamma^{17}\Gamma^i, ..., \Gamma^{17}\Gamma^{i_1...i_7}\}. \tag{5.7}$$

This is the analog of the set of 16 Dirac matrices in four dimensions. Due to the completeness of the O^I, one can easily derive the following Fierz rearrangement identity:

$$(\lambda_2 M\psi)(N\lambda_1)$$
$$= \sum_{I,J}(\lambda_2 O^I\lambda_1)(NO^J M\psi)(\text{tr } O^I O^J)^{-1}. \tag{5.8}$$

From the Weyl condition satisfied by λ_1, λ_2, and ψ it is clear that the O^I containing Γ^{17} double or cancel the contributions of the corresponding O^I without Γ^{17}. In fact, the O^I with an odd number of Γ's cancel. This leaves

$$O^I = \{I, \Gamma^{ij}, \Gamma^{ijkl}, \Gamma^{i_1...i_6}, \Gamma^{i_1...i_8}\}. \tag{5.9}$$

It is easy to show that $\lambda_1\lambda_2$, $\lambda_1\Gamma^{ijkl}\lambda_2$ and $\lambda_1\Gamma^{i_1...i_8}\lambda_2$ are symmetric in (12), and hence also these do not contribute to the commutator. Finally,

$$\Gamma^{kl}\Gamma^{i_1...i_6}\Gamma^{kl} = 0 \quad (k, l = 1, 16) \tag{5.10}$$

as one may verify

$$\left(\binom{6}{2} + \binom{10}{2} = 6\times 10\right).$$

Hence

$$[\delta(\lambda_2), \delta(\lambda_1)]\psi$$
$$= 4(\text{tr }\Gamma^{12}\Gamma^{12})^{-1}(\lambda_2\Gamma^{ij}\lambda_1)(\Gamma^{kl}\Gamma^{ij}\Gamma^{kl})\psi. \tag{5.11}$$

From $\Gamma^{kl}\Gamma^{ij}\Gamma^{kl} = -128\Gamma^{ij}$ (note that one sums over $k>l$ and $k<l$ in $\delta\psi^\alpha$ but that in O^I the Γ^{ij} are counted once, so that in O^I one has $i<j$), one finds again the same result as for the (λ, λ) commutator on ϕ^{ij}.

Clearly, the O(16) generators and the rest of E_8 form a Cartan decomposition, as expected since O(16) is a maximal subgroup.

F_4

The exceptional algebra F_4 has 52 generators, a 26-dimensional real representation, and O(9) as a maximal subalgebra. Since spin (9) is 16 dimensional, one expects a representation in terms of a vector v^i ($i = 1, 9$), a spinor ψ^α ($\alpha = 1, 16$), and a scalar s. Of the 52 parameters, 36 parameters $\Lambda^i{}_j$ come from O(9) while the remaining 16 parameters λ^α clearly form an O(9) spinor. Thus we put

$$\delta v^i = (\Lambda^i{}_j v^j) + (\lambda^\alpha\Gamma^i{}_{\alpha\beta}\psi^\beta),$$
$$\delta\psi^\alpha = (\tfrac{1}{4}\Lambda^{ij}\Gamma_{ij}{}^\alpha{}_\beta\psi^\beta)$$
$$+ (s\lambda^\alpha + v^i\Gamma^{i,\alpha}{}_\beta\lambda^\beta),$$
$$\delta s = a(\lambda^\alpha\psi_\alpha), \tag{5.12}$$

where a is a free parameter which cannot be scaled away. The (Λ, Λ) and (Λ, λ) commutators are guaranteed to be uniform for v^i, ψ^α, and s, due to the manifest O(9) covariance. Let us therefore only check the (λ, λ) commutator.

On v^i one finds

$$[\delta(\lambda_2), \delta(\lambda_1)] = \delta(\Lambda' = 2\lambda_1\Gamma_{ij}\lambda_2), \tag{5.13}$$

using the fact that in $d = 9$ dimension the Dirac matrices can be taken real and Hermitian, thus symmetric. On s the same result is found (namely zero). On ψ^α, finally, one finds after a Fierz rearrangement

$$[\delta(\lambda_2), \delta(\lambda_1)]\psi$$
$$= \tfrac{1}{16}(\lambda_2 O^I\lambda_1 - \lambda_1 O^I\lambda_2)(aO^I + \Gamma^k O^I\Gamma^k)\psi, \tag{5.14}$$

where the factor $\frac{1}{16}$ is due to the fact that the Dirac matrices are 16×16 matrices in $d = 9$. The complete set O^I is given by

$$O^I = \{1, \Gamma^i, \Gamma^{i_1i_2}, ..., \Gamma^{i_1...i_4}\} \tag{5.15}$$

and 1, Γ_i, $\Gamma_{i_1...i_4}$ are symmetric. Hence we only need to consider $O^I = \Gamma^{i_1 i_2 i_3}$ and $O^I = \Gamma^{i_1 i_2}$. The former yield zero if $a = 3$, and with this result the latter yield the desired commutator.

The scalar s is thus needed because $\Gamma^k \Gamma^{i_1 i_2 i_3} \Gamma^k$ does not vanish. In the case of E_8 no scalar is needed, because in 16 dimensions the identity (5.10) holds.

G_2

The exceptional algebra G_2 has 14 generators, its lowest representation is 7 dimensional and real, and SU_3 is a maximal subgroup. Clearly, one can only have the following SU_3 reduction

$$\underset{\sim}{7} = \underset{\sim}{3} + \underset{\sim}{\bar{3}} + \underset{\sim}{1}. \tag{5.16}$$

Thus we expect a representation on an SU_3 triplet x^α, an antitriplet $\bar{x}_\alpha = (x^\alpha)^*$ and a real scalar x^0. The parameters consist of $8SU_3$ parameters $\lambda^\alpha{}_\beta$ and six other parameters, which can only be a triplet σ^α and an antitriplet $\bar{\sigma}_\alpha = (\sigma^\alpha)^*$. Thus we expect the following matrix representation for G_2:

$$\begin{aligned}
\delta x^\alpha &= \lambda^\alpha{}_\beta x^\beta + \tfrac{1}{\sqrt{2}}\epsilon^{\alpha\beta\gamma}\bar{\sigma}_\beta \bar{x}_\gamma + \sigma^\alpha x^0, \\
\delta \bar{x}_\alpha &= \tfrac{1}{\sqrt{2}}\epsilon_{\alpha\beta\gamma}\sigma^\beta x^\gamma + \bar{\lambda}_\alpha{}^\beta \bar{x}_\beta + \bar{\sigma}_\alpha x^0, \\
\delta x^0 &= a\bar{\sigma}_\alpha x^\alpha + \bar{a}\sigma^\alpha \bar{x}_\alpha,
\end{aligned} \tag{5.17}$$

where a is a free constant to be fixed below. Again we only need verify the (σ, σ) commutator since all transformation laws are manifestly SU_3 covariant.

Acting on x^α one has

$$\begin{aligned}
[\delta(\sigma_2), \delta(\sigma_1)]x^\alpha = {} & (a\sigma_1^\alpha \bar{\sigma}_{2,\beta})x^\beta + a\sigma_1^\alpha \sigma_2^\beta \bar{x}_\beta \\
& + \tfrac{1}{2}(\epsilon^{\alpha\beta\gamma}\bar{\sigma}_{1,\beta}\epsilon_{\gamma\delta\epsilon}\sigma_2^\delta)x^\epsilon \\
& + (1/\sqrt{2})(\epsilon^{\alpha\beta\gamma}\bar{\sigma}_{1,\beta}\bar{\sigma}_{2,\gamma})x^0 - 1\leftrightarrow 2,
\end{aligned} \tag{5.18}$$

The composite parameters are thus

$$\begin{aligned}
(\lambda')^\alpha{}_\beta &= (a - \tfrac{1}{2})(\sigma_1{}^\alpha \bar{\sigma}_{2,\beta} - \sigma_2^\alpha \bar{\sigma}_{1,\beta}) - \tfrac{1}{2}\delta^\alpha_\beta(\bar{\sigma}_1\cdot\sigma_2 - \bar{\sigma}_2\cdot\sigma_1), \\
(\sigma')^\alpha &= \sqrt{2}\epsilon^{\alpha\beta\gamma}\bar{\sigma}_{1,\beta}\bar{\sigma}_{2,\gamma}, \\
(\bar{\sigma}')_\alpha &= -a\sqrt{2}\epsilon_{\alpha\beta\gamma}\sigma_1{}^\beta \sigma_2^\gamma.
\end{aligned} \tag{5.19}$$

Clearly

$$a = -1, \tag{5.20}$$

and with this value λ' is traceless, as it must be, while also the (σ, σ) commutator is consistent. This completes the 7×7 matrix representation of G_2. In terms of $x + \bar{x}$, $(x - \bar{x})/i$ and x^0 it is real, but we prefer in this paper to use the basis x^α, $\bar{x}_\alpha$, and x^0.

6. ALTERNATIVE CONSTRUCTION OF THE MATRIX PSEUDOREPRESENTATION FOR $G(3)$

We now extend the methods of Sec. 5 from ordinary Lie algebras to superalgebras, and show how one can derive the 9×9 matrix pseudorepresentation for $G(3)$. The result coincides of course with the results obtained by the general method used in Sec. 3, but it may be that the simplicity of the arguments used below makes them useful also for non semisimple superalgebras.

Since the odd generators must be in the same representation as that used for G_2, the odd matrix elements are

$$(q^{\alpha i}, \bar{q}_\alpha{}^i, d^i), \tag{6.1}$$

with $\bar{q}_\alpha{}^i$ transforming under SU(3) as $\bar{3}$ and with d^i real. Thus we consider the matrix

$$M = \begin{pmatrix}
\lambda^\alpha{}_\beta & \tfrac{1}{\sqrt{2}}(\epsilon\cdot\bar{\sigma})^{\alpha\gamma} & \sigma^\alpha & q^{\alpha i} \\
\tfrac{1}{\sqrt{2}}(\bar{\epsilon}\cdot\sigma)_{\alpha\gamma} & \bar{\lambda}_\alpha{}^\beta & \bar{\sigma}_\alpha & \bar{q}_\alpha{}^i \\
-\bar{\sigma}_\alpha & -\sigma^\alpha & 0 & d^i \\
\bar{q}_{\alpha i} & q^\alpha{}_i & \alpha d_i & c_i{}^j
\end{pmatrix}, \tag{6.2}$$

where $(\epsilon\cdot\bar{\sigma})^{\alpha\gamma} = \epsilon^{\alpha\beta\gamma}\bar{\sigma}_\beta$ and $(\bar{\epsilon}\cdot\sigma)_{\alpha\gamma} = \epsilon_{\alpha\beta\gamma}\sigma^\beta$. We shall now discuss the elements in the last row. Since d^i is real, and this property must be preserved, $c_i{}^j$ must be real, and one can take c traceless (its unit matrix forms an ideal). Since under ordinary matrix multiplication the elements of c contain terms $\bar{q}_{1,\alpha i}q_2{}^{\alpha j} + q^\alpha_{1,i}\bar{q}_{2,\alpha}{}^j + \alpha d_{1,i}d_2^j + 1\leftrightarrow 2$, reality of c requires

$$\bar{q}_{1,\alpha i} = (q^\alpha_{1,i})^*. \tag{6.3}$$

If c is to be traceless, $d_{1,i}d_2{}^i + d_{2,i}d_1{}^i = 0$. This can only hold if $i = 1, 2$ and if $d_{1,i} = \epsilon_{ij}d_1{}^j$. Thus we have found the elements of the last row of M

$$\bar{q}_{\alpha,i} = \epsilon_{ij}(q^{\alpha j})^*, \quad q_i^\alpha = \epsilon_{ij}q^{\alpha j}, \quad d_i = \epsilon_{ij}d^j. \tag{6.4}$$

It also follows that the matrix c is an element of SL(2, R). Thus the bosonic algebra is $G_2\times$SL(2, R).

One must be careful with these arguments and we only present them as a heuristic guide to guessing the correct bosonic group. In fact, as we shall see, the matrix composition rule is not given by ordinary (anti)commutators, but in the even sector extra terms are needed. Hence, the above arguments are simply not applicable, since we assumed that the odd–odd elements of c were obtained by anticommuting the corresponding generators. Nevertheless, one arrives at the correct starting point and such arguments may also be useful for other cases.

One can rephrase the form of M as a set of transformation rules on coordinates $(x^\alpha, \bar{x}_\alpha, x^0)$ and (y_i):

$$\begin{aligned}
\delta x^\alpha &= \lambda^\alpha{}_\beta x^\beta + \tfrac{1}{\sqrt{2}}\epsilon^{\alpha\beta\gamma}\bar{\sigma}_\beta\bar{x}_\gamma + \sigma^\alpha x^0 + q^{\alpha i}y_i, \\
\delta x^0 &= -\bar{\sigma}_\alpha x^\alpha - \sigma^\alpha \bar{x}_\alpha + d^i y_i, \\
\delta y_i &= \bar{q}_{\alpha i}x^\alpha + q_i{}^\alpha \bar{x}_\alpha + \alpha d_i x^0 + c_i{}^j y_j,
\end{aligned} \tag{6.5}$$

and one easily convinces oneself that the only free parameter is α.

If one takes the commutator or anticommutator of two matrices M_1 and M_2, the result must be again a matrix of the same form. In particular, the composite entry $q^{\alpha i}_{12}$ must be related to the composite entry $\bar{q}_{12,\alpha i}$ by $\bar{q}_{12,\alpha i} = \epsilon_{ij}(q_{12}{}^{\alpha j})^*$. This means that

$$\begin{aligned}
&\epsilon_{ij}(\lambda_1 q_2^j + \tfrac{1}{\sqrt{2}}\epsilon\cdot\bar{\sigma}_1\bar{q}_2^j + \sigma_1 d_2^j + q_1{}^k c_{2,k}{}^j)^* - 1\leftrightarrow 2 \\
&\quad = \bar{q}_{1,i}\lambda_2 + \tfrac{1}{\sqrt{2}}q_{1,i}\bar{\epsilon}\cdot\sigma_2 - \alpha d_{1,i}\bar{\sigma}_2 + c_{1,i}{}^j\bar{q}_{2,j} - 1\leftrightarrow 2.
\end{aligned} \tag{6.6}$$

All terms match if one fixes α to be given by

$$\alpha = +1. \tag{6.7}$$

With this value for α, the composition rule of the parameters for the last row and column of M is consistent with ordinary matrix commutators or anticommutators.

In the even–even sectors, the composition rule cannot be given by ordinary matrix muliplication, as we already stressed. For example, the composite $\lambda^{\alpha}{}_{\beta}$ is no longer traceless, though it is still anti-Hermitian. By accident, c is still traceless, but we take now the most general composition rule for the elements of the bosonic sector, and only assume that the terms with purely bosonic parameters are as usual. The requirements of SU(3) and SL(2, R) covariance severely restrict the possible form of the terms bilinear in q and d.

$$\lambda^{\alpha}_{12,\beta} = \ldots + A\left[q_1^{\alpha i}\bar{q}_{2,\beta i} - \tfrac{1}{3}\delta^{\alpha}_{\beta}(q_1^{\gamma i}\bar{q}_{2,\gamma i}) + 1\leftrightarrow 2\right],$$
$$\sigma^{\alpha}_{12} = \ldots + \frac{B}{\sqrt{2}}[\epsilon^{\alpha\beta\gamma}\bar{q}^{i}_{1,\beta}\bar{q}_{2,\gamma i}] + C\,[q_1^{\alpha i}d_{2,i} + 1\leftrightarrow 2],$$
$$c_{12,i}{}^{j} = \ldots + D\,[(\bar{q}_{1,\alpha i}q_2{}^{\alpha j} + \text{c.c}) + 1\leftrightarrow 2] + E\,[d_{1,i}d_2^{j} + 1\leftrightarrow 2]. \qquad (6.8)$$

The terms denoted by ... in λ_{12} and σ_{12} are bilinear in λ_i and $\sigma_i (i = 1, 2)$, while the terms denoted by ... in $c_{12,i}{}^{j}$ are given by $[c_1, c_2]_i{}^{j}$.

In order to fix the five free parameters A, B, C, D, E we consider the Jacobi identities. As an example, consider the three matrices $M(\bar{q}_1)$, $M(d_2)$, and $M(\sigma_3)$. Since $M(\bar{q}_1)$ and $M(d_2)$ are odd generators, while $M(\sigma_3)$ is even, the Jacobi identity reads

$$[\{M(\bar{q}_1), M(d_2)\}, M(\sigma_3)] - \{[M(d_2), M(\sigma_3)], M(\bar{q}_1)\} + \{[M(\sigma_3), M(\bar{q}_1)], M(d_2)\} = 0. \qquad (6.9)$$

The signs are easily understood: pulling $M(\bar{q}_1)$ in the first term to the right, it passes the odd generator $M(d_2)$ which results in an extra minus sign, but permuting $M(\sigma_3)$ to the left, no minus signs can appear since $M(\sigma_3)$ is bosonic.

Consider now the (1, 1) entry in the Jacobi identity (the entry which in M is called $\lambda^{\alpha}{}_{\beta}$). The composite parameter λ, due to combining two matrices $M_{\rm I}$ and $M_{\rm II}$, is given by

$$\lambda^{\alpha}_{\rm I\,II,\beta} = (\lambda_{\rm I}\lambda_{\rm II})^{\alpha}{}_{\beta} + \tfrac{1}{2}(\epsilon\cdot\bar{\sigma}_{\rm I}\bar{\epsilon}\cdot\sigma_{\rm II})^{\alpha}{}_{\beta} - \sigma^{\alpha}_{\rm I}\bar{\sigma}_{\rm II,\beta} - \rm I\leftrightarrow II + A\left[q_{\rm I}^{\alpha i}\bar{q}_{\rm II,\beta i} - \tfrac{1}{3}\delta^{\alpha}{}_{\beta}q_{\rm I}{}^{\gamma i}\bar{q}_{\rm II,\gamma i} + \rm I\leftrightarrow II\right]. \qquad (6.10)$$

It is clear that the only contributions of the form $\bar{q}_1 d_2\sigma_3$ in the (1, 1) entry from the three terms in the Jacobi identity are given by

from term 1: $\tfrac{1}{2}(\epsilon\cdot\bar{\sigma}_{12}\bar{\epsilon}\cdot\sigma_3)^{\alpha}{}_{\beta} + (\sigma_3\bar{\sigma}_{12})^{\alpha}{}_{\beta}$,

from term 2: $-A\,(q^{\alpha i}_{23}\bar{q}_{1,\beta i} - \tfrac{1}{3}\delta^{\alpha}_{\beta}q^{\gamma}_{23}\bar{q}_{1,\gamma i})$, (6.11)

from term 3: none.

The only terms in $\bar{\sigma}_{12}$ of the form $\bar{q}_1 d_2$ and in q_{23} of the form $d_2\sigma_3$ are

$$\bar{\sigma}_{12,\alpha} = \overline{C}\bar{q}^{i}_{1,\alpha}d_{2,i}, \quad q^{\alpha i}_{23} = -\sigma^{\alpha}_3 d^{\,i}_2. \qquad (6.12)$$

To make life even simpler, one may consider only the terms with $\delta^{\alpha}{}_{\beta}$. There are only two such terms, and one arrives at (using $q^i d_i = -q_i d^i$).

$$\overline{C} = +\tfrac{2}{3}A. \qquad (6.13)$$

In the odd sector of M one finds the relation between A and B, C on the one hand, and between D, E and B,C on the other hand. Consider for example a relation $C\sim E$. Since C corresponds to qd terms and E to dd terms, we consider $d_1d_2q_3$ terms. Since the C terms come from $\sigma_{\rm I\,II}$ and the E terms from $c_{\rm I\,II}$, we must consider an entry whose composition rule contains terms like σd and qc. This is clearly the (1, 4) entry (the one called $q^{\alpha i}$ in M). Thus we consider the Jacobi identity for three matrices M, containing parameters q_3, d_1, and d_3, respectively,

$$[\{M(q_3), M(d_1)\}, M(d_2)] + [\{M(d_1), M(d_2)\}, M(q_3)] + [\{M(d_2), M(q_3)\}, M(d_1)] = 0. \qquad (6.14)$$

The composition rule of the parameter $q^{\alpha i}$ reads

$$q^{\alpha i}_{\rm I\,II} = \lambda_{\rm I}q^{i}_{\rm II} - \frac{1}{\sqrt{2}}\epsilon\cdot\bar{\sigma}_{\rm I}q^{i}_{\rm II} + \sigma_{\rm I}d^{\,i}_{\rm II} + q^{j}_{\rm I}c_{{\rm II},j}{}^{i} - \rm I\leftrightarrow II. \qquad (6.15)$$

Each of the terms in the Jacobi identity can contribute $q_3d_1d_2$ terms to the (1, 4) entry. The first term in the Jacobi identity corresponds to $q_{\rm I\,II}$ with I = (31) and II = 2, and thus we must consider those terms in $q_{\rm I\,II}$ which contain $d_{\rm II}$. Hence one receives a contribution from term 1 given by $\sigma_{31}d^{\,i}_2$. Similarly, the second term in the Jacobi identity can only contribute if there is a $q_{\rm II}$ term in $q_{\rm I\,II}$. Finally, the last term in the Jacobi identity is equal to the first if one interchanges the indices 1 and 2.

The sum of all contributions of the form $q_3d_1d_2$ to the Jacobi identity for the (1, 4) entry of M is given by

$$(\sigma_{31}d_2{}^{i}) + (-q_3{}^{j}c_{12,j}{}^{i}) + (\sigma_{23}d_1{}^{i}) = (Cq_3{}^{j}d_{1,j})d_2{}^{i} - q_3{}^{j}E\,(d_{1,j}d_2{}^{i} + d_{2,j}d_1{}^{i}) + (q_3{}^{j}d_{2,j})d_{1,i} = 0, \qquad (6.16)$$

from which one easily deduces that

$$E = C. \qquad (6.17)$$

The other parameters are fixed in a similar way. The $\bar{q}_1d_2\sigma_3$ terms in the (4, 4) entry of M yield $E = D$, while the $\bar{q}_1\bar{q}_2\bar{\sigma}_3$ in the (1, 1) entry yield $B = +\tfrac{4}{3}A$. Since there is always a free scale in the (odd, odd) bracket relations, our final result contains still one free parameter which we fix by $A = 1$. Thus

$$A = 1, \quad B = +\tfrac{4}{3}, \quad C = E = D = \tfrac{2}{3}. \qquad (6.18)$$

The completes our discussion of $G(3)$. The matrix representation is (6.5) with $\alpha = 1$. The bracket relation of two matrices is given by the usual commutators or anticommutators, except that for the bracket between two odd elements one must use the formulae (6.8) and (6.18).

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the hospitality of the Institute for Theoretical Physics at Santa Barbara where this work was done. This work was supported in part by the National Science Foundation Grant No. PHY-79-06376A01.

[1]V. Kac, Adv. Math. **26**, 8 (1977); Commun. Math. Phys. **53**, 31 (1977); Adv. Math. **26**, 9 (1977).

[2]S. MacDowell and F. Mansouri, Phys. Rev. Lett. **38**, 739 (1977); P. K. Townsend and P. van Nieuwenhuizen, Phys. Lett. B **67**, 439 (1977); A. Chemseddine and P. C. West, Nucl. Phys. B **129**, 39 (1977); P. Fré, Caltech Preprint CALT-68-797.

[3]M. Kaku, P. K. Townsend, and P. van Nieuwenhuizen, Phys. Rev. D **17**,

3179 (1978); S. Ferrara, M. Kaku, P. K. Townsend, and P. van Nieuwenhuizen, Nucl. Phys. B **129**, 125 (1977); E. Bergshoeff, M. de Roo, and B. de Wit, Nucl. Phys. B (to be published).
[4]W. Nahm, Nucl. Phys. B **135**, 149 (1978).
[5]P. van Nieuwenhuizen, Sanibel meeting, ITP-SB-79-13 (February 1979); F. A. Berends, B. deWit, J. W. van Holten, and P. van Nieuwenhuizen, J. Phys. A **13**, 164 (1980); C. Aragone and S. Deser, Phys. Lett. B **86**, 161 (1979).
[6]R. Arnowitt and P. Nath, Phys. Lett. B **56**, 177 (1975).
[7]R. D'Auria, P. Fré, and T. Regge, in *Unification of the Fundamental Particle Interactions*, edited by J. Ellis, S. Ferrara and P. van Nieuwenhuizen (Plenum, New York, 1980).
[8]J. H. Schwarz, Caltech preprint CALT-68-817.
[9]Y. Ne'eman, Phys. Lett. B **82**, 190 (1979).
[10]E. Cremmer and B. Julia, Nucl. Phys. B **159**, 141 (1979).
[11]E. Cremmer, J. Scherk, and J. H. Schwarz, Phys. Lett. B **84**, 83 (1979).
[12]For properties of Dirac matrices and the charge conjugation in higher dimensions, see P. van Nieuwenhuizen, Proc. Trieste Supergravity School, 1981.

Study of superluminal electromagnetic fields

O. P. S. Negi and B. S. Rajput[a)]
Department of Physics, Kumaun University, Naini Tal, 263 002 India.

(Received 19 March 1981; accepted for publication 30 July 1981)

Behavior of Maxwell's equations under imaginary and real superluminal Lorentz transformations respectively has been reexamined and the expression of a Lorentz force acting on an electric charge interacting with superluminal electromagnetic fields has been derived. It has been shown that in four-dimensional space-time as well as in six-dimensional space-time the electrically charged tachyon interacting with electromagnetic fields behaves neither as a purely electric charge nor as a purely magnetic monopole.

PACS numbers: 14.80.Kx, 12.20. – m

1. INTRODUCTION

Diverse approaches have been developed to carry out the field theory of tachyons, probably because fundamental quantal properties of these particles are not yet very well known. In the classical theory of tachyons, two main approaches have been followed by different authors. In the first one, adopted by Recami *et al.*,[1] Corben,[2] Teli *et al.*[3] and others,[4] the components of a four-vector in the directions perpendicular to the relative motion become imaginary on passing from the subluminal to superluminal realm, while in the second approach adopted by Antippa,[5] Antippa and Everett,[6] Gonzalez-Gascon[7] and Lemke,[8] the real superluminal Lorentz transformations are used.

We[9] have recently derived the transformations of electromagnetic fields, four-potential, and four-current density under both types of superluminal Lorentz transformations and showed that the Maxwell's equations are not invariant under any type of transformation in general, while the components of superluminal electromagnetic fields transverse with respect to the direction of the relative motion appear to satisfy field equations similar to Maxwell's equations under the superluminal Lorentz transformations of the second type. In the present paper it has been shown that the chronotopical mapping $(3,1)\to(1,3)$ in the first type of superluminal Lorentz transformations does not retain the invariance of Maxwell's equations, and to retain such an invariance we must include an extra negative sign in the transformations of four-current source density. The expansion for the superluminal magnetic field under the first type of transformations and mapping $(3,1)\to(1,3)$ is then inconsistent and it becomes isotropic, having the same strength in all directions. It has also been shown that under these transformations the nature of the Lorentz force is changed and an electrically charged tachyon interacting with superluminal electromagnetic fields does not behave exactly as expected of either electric charge or magnetic monopole, in contradiction with the results of Recami–Mignani,[10] Vysin,[11] and Dattoli–Mignani.[12] It has also been shown that in six-dimensional space-time formalism originated by Demers,[13] the consistency of the magnetic field may be retained under the first type of superluminal Lorentz transformations and the mapping then becomes one to one. It has also been shown that in six-dimensional space-time formalism, though the expression for the Lorentz force on the charge tachyon in the superluminal electromagnetic field is similar to that derived by Dattoli–Mignani,[12] it cannot lead to their conclusion that an infinite-speed tachyon behaves as a magnetic charge at rest. Rather, it has been shown to be similar neither to the force acting on a purely electric charge nor to that on a purely magnetic monopole.

We also reexamine in this paper the question of invariance of Maxwell's equations under real superluminal Lorentz transformations by using the reduced expansion of electromagnetic fields and potentials in terms of standard helicity representations of the Poincaré group and it has been shown that no suitable charge and current source densities satisfying equation of continuity can be defined for the Maxwell's equations to be satisfied even for transverse superluminal fields.

2. SUPERLUMINAL ELECTROMAGNETIC FIELDS UNDER SUPERLUMINAL LORENTZ TRANSFORMATIONS OF THE FIRST TYPE

Let us start with two parallel frames K and K' in relative motion with velocity $v > c$ along the Z direction such that their origins coincide at time $t = t' = 0$. Transformation equations for space and time coordinate in these frames may be written in the following form[1] in the natural units $c = \hbar = 1$;

$$\begin{aligned} x_j' &= \pm i x_j \quad (j = 1,2), \\ x_3' &= \pm \gamma(x_3 - vt), \\ t' &= \pm \gamma(t - v x_3), \end{aligned} \tag{1}$$

where $x_1 = x^1 = x$, $x_2 = x^2 = y$, $x_3 = x^3 = z$, and γ is given by

$$\gamma = (v^2 - 1)^{-1/2}.$$

From these equations we get

$$t^2 - \sum_{l=1}^{3} x_l^2 = \sum_{l=1}^{3} x_l'^2 - t'^2, \tag{2}$$

which shows that the reference metric $(+1, -1, -1, -1)$ in frame K is transformed to the metric $(-1, +1, +1, +1)$ and the transformations (1) and the roles of space and time get interchanged. In other words, the transformations (1) are incorporated with the chronotopical mapping $(3,1)\to(1,3)$, or

[a)]On leave from Physics Department, Garhwal University, Srinagar (Garhwal), U. P., India.

$$(x, y, z, it) \to (t', ix', iy', iz'), \tag{3}$$

from which we get

$$\Box = -\Box' \tag{4}$$

and the mapping

$$\left(\vec{\nabla}, i\frac{\partial}{\partial t}\right) \to \left(\frac{\partial}{\partial t'}, i\vec{\nabla}'\right). \tag{5}$$

Similar superluminal Lorentz transformations have been derived[14] for the components of the electromagnetic four-potential $\{A_\mu\}$ and it has been shown that

$$|A_\mu'|^2 = -|A_\mu|^2, \tag{6}$$

with the corresponding $(3,1) \to (1,3)$ mapping

$$(A_1, A_2, A_3, i\phi) \to (\phi', iA_1', iA_2', iA_3'), \tag{7}$$

where $\phi = -iA_0$. Using relation (2) and the mappings (5) and (7) and the similar mapping for the components of four-current density $\{J_\mu\}$, we may transform the Maxwell's field equations

$$\Box A_\mu = J_\mu \tag{8}$$

in the frame K to the following equation in frame K':

$$\Box' A_\mu' = -J_\mu', \tag{9}$$

which are the equations according to which the superluminal electromagnetic field is coupled to the tachyons (which may be considered as bradyons in superluminal frame K' in view of tachyon–bradyon reciprocity). These equations are not similar to Maxwell's field equations (8) for the superluminal electromagnetic fields. In other words, the field equations are not invariant under superluminal Lorentz transformations (1). In order to retain these field equations under the transformations (2) and mappings (5) and (7), we must include an extra negative sign in the transformations of four-current density, i.e., we must consider the mapping

$$(i\rho, J_x, J_y, J_z) \to -(\rho', iJ_x', iJ_y', iJ_z'). \tag{10}$$

In spite of the change in sign in this mapping, the real and imaginary components of the four-current lead to the corresponding real and imaginary components of the four-potential. Change in sign leaves the total charge and current densities invariant because the volume element also changes sign under the transformations (1). With this change in sign in the mapping (10) of the components of four-current density, Maxwell's field equations may be treated as invariant upon passing from the subluminal to superluminal realm and vice versa through the transformations (1).

Under the mappings (5) and (7) the usual equation for electric and magnetic fields for the superluminal case transform to the following equations for superluminal fields:

$$\vec{E}' = -\text{grad}'\phi' - \frac{\partial \vec{A}'}{\partial t},$$
$$\vec{H}' = \frac{\partial}{\partial t'}\phi'\hat{n}, \tag{11}$$

where $\hat{n}$ is the unit vector in the direction of the magnetic field. These equations are neither similar to those of fields produced by an electric charge source nor to those produced by a magnetic charge source. It may therefore be concluded that an electrically charged tachyon interacting with superluminal electromagnetic fields observed in frame K does not behave exactly as expected of either an electric charge or a magnetic monopole. This result is in contradiction with those of Recami,[10] Vysin[11] and Dattoli–Mignani.[12]

A similar conclusion can be drawn by transforming the usual expansion of the Lorentz force

$$\vec{F} = e\frac{\partial \vec{A}}{\partial t} + e\vec{\nabla}\phi + e\vec{\omega}\times(\vec{\nabla}\times\vec{A}) \tag{12}$$

under the mappings (5) and (7) into the following form:

$$\vec{F}' = e\vec{\nabla}'\phi' + e\frac{\partial \vec{A}'}{\partial t'} + e\overleftarrow{\omega}\frac{\partial \phi'}{\partial t'}, \tag{13}$$

where $\overleftarrow{\omega}$ is the inverse velocity defined by $dt/d\vec{x}$. This is similar neither to the Lorentz force acting on an electrically charged particle nor to the corresponding force acting on a magnetic monopole. This equation shows that a subluminal electric charge interacting with superluminal electromagnetic fields or an electrically charged tachyon interacting with a subluminal electromagnetic field behaves neither as a purely electric charge nor as a pure magnetic monopole. However, according to Eq. (13) an electrically charged transcedent tachyon (moving with infinite velocity i.e., $\overleftarrow{\omega} = 0$) will behave like a pure electric charge interacting with a subluminal pure electric field, while a subluminal electric charge moving with velocity close to the velocity of light and interacting with a purely superluminal magnetic field will behave like a magnetic monopole with magnetic charge $g = -e$.

Equation (11) for a superluminal magnetic field derived by using transformations (1) and mappings (5) and (7) is not a consistent one since it gives an isotropic magnetic field having the same strength in all directions. Moreover, under transformations (1), the components of the position-vector become imaginary in the directions perpendicular to the direction of relative superluminal motion between the frames K and K'. Similarly, the electromagnetic fields, potentials, and currents become imaginary in the direction perpendicular to the relative motion under the superluminal transformations (1). Consequently, a particle exposed to superluminal electromagnetic fields will not have real momentum and energy. In other words, the transformations (1) lead to the conclusion that once we are prepared to consider the tachyons, we must give up the idea that dynamical variables in relativistic classical mechanics are always real. Furthermore, we have already shown in our earlier papers[9,15] that by using the superluminal Lorentz transformations (1) (of the first kind) it is not possible to derive the reduction of fields associated with spin-1 and spin-$\frac{1}{2}$ tachyons in terms of the standard helicity representation of the inhomogeneous Lorentz group.

To overcome these problems of the superluminal Lorentz transformation we may adopt the formalism[13], where space and time play a symmetrical role and where the time t is considered as a vector in six-dimensional space-time with three spatial and three temporal coordinates. Recently a number of authors have supported the idea that the theory of relativity should involve the use of three dimensions of time. Ziino[16] argued in its favor on the grounds that light-speed invariance is not fully consistent with the standard relativity, while Pappas[17] has introduced the notion of parallelism and the desirability of space–time symmetry as a justification.

Following Demers' formalism,[13] in which the three components of the time vector are coupled together giving $|t| = (t_x^2 + t_y^2 + t_z^2)^{1/2}$ as measurable and all the individual space coordinates are measurable in the subluminal frame, while on passing from bradyon to tachyons via superluminal Lorentz transformations, all the components of time become measurable and the space components couple together giving only $\mathrm{mod}|r| = (x^2 + y^2 + z^2)^{1/2}$ as measurable, we get the following mappings under superluminal Lorentz transformation:

$$\{\vec{r}, it \equiv i(t_x^2 + t_y^2 + t_z^2)^{1/2}\} \to \{\vec{t}, ir \equiv i(x^2 + y^2 + z^2)^{1/2}\}, \tag{14}$$

$$\left(\vec{\nabla}_r, i\frac{\partial}{\partial t}\right) \to \left(\vec{\nabla}_t, i\frac{\partial}{\partial r}\right), \tag{15}$$

where $\vec{\nabla}_r$ and $\vec{\nabla}_t$ are del operators in three spatial and three temporal coordinates, respectively.

In a similar manner the corresponding mapping may be written for the components of electromagnetic potential in six-dimensional space

$$(\vec{A}, i\phi) \to (\vec{\phi}, iA), \tag{16}$$

where

$$\phi \equiv (\phi_x^2 + \phi_y^2 + \phi_z^2)^{1/2}, \quad A \equiv (A_x^2 + A_y^2 + A_z^2)^{1/2}; \tag{17}$$

then in place of equation (11), we get

$$\vec{E}' = -e\vec{\nabla}_t A - \frac{\partial \vec{\phi}}{\partial r}, \quad \vec{H}' = -\vec{\nabla}_t \times \vec{\phi}, \tag{18}$$

which is similar neither to the fields produced by an electric charge source nor to those produced by a magnetic charge source in, contradiction with result of Dattoli–Mignani.[12]

Similarly, the Eq. (12) for the Lorentz force is mapped into the following equation:

$$\vec{F}' = -e\vec{\nabla}_t A + e\frac{\partial \vec{\phi}}{\partial r} + e\vec{\varpi} \times (\nabla_t \times \phi), \tag{19}$$

where $\vec{\varpi}$ is the inverse velocity defined by

$$\vec{\varpi} = d\vec{t}/dr. \tag{20}$$

Though the expression (19) is identical to that derived earlier by Dattoli–Mignani,[12] it cannot lead to their conclusion that an infinite-speed tachyon behaves as a magnetic charge at rest. Rather it will behave as a purely electric charge interacting with a pure electric field.

In a forthcoming paper we shall undertake the study of classification of bradyons and tachyons in six-dimensional space-time, where it will be demonstrated how observable speeds are related to nine observables $\partial r_i/\partial t_j$ $(i, j = 1,2,3)$.

3. SUPERLUMINAL ELECTROMAGNETIC FIELDS UNDER THE SUPERLUMINAL LORENTZ TRANSFORMATIONS OF THE SECOND TYPE

Let us consider two frames of reference K and K' in relative motion with superluminal velocity $\vec{v}(v > 1)$ along an arbitrary direction of the tachyon corridor. Real superluminal Lorentz transformations of the vector and scalar parts of the space-time four-vector $\{X_\mu\}$, with $x_0 = it$ may be written as follows[14] in natural units $c = \hbar = 1$:

$$\vec{x} = \vec{x}' + (\mu\gamma - 1)\frac{\vec{v}(\vec{v}\cdot\vec{x}')}{v^2} - i\mu\phi\vec{v}x_0', \quad x_0 = \mu\gamma[x_0' + i(v\cdot x')], \tag{21}$$

where

$$\mu = \frac{v}{|\vec{v}|} \quad \text{and} \quad \gamma = (v^2 - 1)^{-1/2}.$$

Using these transformations and the techniques of Moses[18] for the reduction of the wave function in terms of the standard helicity representation of the Poincaré group, we[9] get the following reduced expansions of the scalar and vector electromagnetic potential under the Lorentz condition $f^0(k,\vec{p},0) = f^{0*}(k,\vec{p},0)$ with the usual definition for $\nabla = (\hat{i}\partial/\partial x_1, \hat{j}\partial/\partial x_2, \hat{k}\partial/\partial x_3)$ and $\nabla_0 = i\partial/\partial t$:

$$\phi = \frac{1}{4\pi^{3/2}}\int \frac{d\vec{p}}{k}\{f(k,\vec{p},0)\exp[i(\vec{p}\cdot\vec{x} - \omega T)] + f^*(k,\vec{p},0)\exp[-i(\vec{p}\cdot\vec{x} - \omega T)]\} \tag{22}$$

and

$$\vec{B} = \frac{1}{4\pi^{3/2}}\int \frac{d\vec{p}}{\omega(k,\vec{p})}\left(\frac{\vec{p}}{k}\right)\{f(k,\vec{p},0)\exp[i(\vec{p}\cdot\vec{x} - \omega T)] + f^*(k,\vec{p},0)\exp[-i(\vec{p}\cdot\hat{x} - \omega T)]\} + \frac{1}{4\pi^{3/2}}\sum_{\lambda=\pm 1}\frac{\lambda}{2^{1/2}}\int\frac{d\vec{p}}{\omega(k,\vec{p})}\{f(k,\vec{p},\lambda)\times\vec{\sigma}_1(k,\vec{p},\lambda)\exp[i(\vec{p}\cdot\vec{x} - \omega T)] + f^*(k,\vec{p},\lambda)\vec{\sigma}_1^*(k,\vec{p},\lambda)\exp[-i(\vec{p}\cdot\vec{x} - \omega T)]\}, \tag{23}$$

where

$$\vec{p}\cdot\vec{x} = p_1x_1 + p_2x_2 + p_3t, \quad \omega T = \omega x_3, \tag{24}$$

and $f(k,\vec{p},0)$ and $f(k,\vec{p},\lambda)$ are the complex functions depending upon mass k and momentum $\vec{p}$ of tachyons of spin-1 with helicity $\lambda = 0$ and $\lambda = \pm 1$, respectively. Vector $\vec{\sigma}_1(k,\vec{p},\lambda)$ in Eq. (23) is given by

$$\vec{\sigma}_1(k,\vec{p},\lambda) = \begin{pmatrix} -1 \\ -i\lambda \\ 0 \end{pmatrix}. \tag{25}$$

Reduced expansion of electromagnetic fields may be given as follows:

$$\vec{E}^L = -\frac{i}{4\pi^{3/2}}\int\frac{d\vec{p}}{\omega(k,\vec{p})}\left(k\frac{\vec{p}}{p}\right)\{f(k,\vec{p},0)\exp[i(\vec{p}\cdot\vec{x} - \omega T)] - f^*(k,\vec{p},0)\exp[-i(\vec{p}\cdot\vec{x} - \omega T)]\}, \tag{26a}$$

$$\vec{E}^T = -\frac{ip_3}{4\pi^{3/2}}\sum_{\lambda=\pm 1}\frac{\lambda}{2^{1/2}}\int\frac{d\vec{p}}{\omega(k,\vec{p})}\times\{f(k,\vec{p},\lambda)\vec{\sigma}_1(k,\vec{p},\lambda)\exp[i(\vec{p}\cdot\vec{x} - \omega T)] - f^*(k,\vec{p},\lambda)\vec{\sigma}_1^*(k,\vec{p},\lambda)\exp[-i(\vec{p}\cdot\vec{x} - \omega T)]\}, \tag{26b}$$

and

$$\vec{H}^T = -\frac{1}{4\pi^{3/2}} \sum_{\lambda=\pm 1} \frac{1}{2^{1/2}} \int dp \, \{ f(k,\vec{p},\lambda) \vec{\sigma}_1(k,\vec{p},\lambda) \times \exp[i(\vec{p}\cdot\vec{x} - \omega T)] + f^*(k,\vec{p},\lambda) \vec{\sigma}_1^*(k,\vec{p},\lambda) \exp[-i(\vec{p}\cdot\vec{x} - \omega T)] \}, \tag{27}$$

where $\vec{E}^L$ and $\vec{E}^T$ are, respectively, the longitudinal and transverse parts of electric fields while the magnetic field is a purely transverse one. If these reduced expansions of electric and magnetic fields are required to satisfy usual Maxwell's equations then the charge and current source densities can not both be made to vanish simultaneously and they must be given by the following reduced forms:

$$\rho = -\frac{k}{4\pi^{3/2}} \int d\vec{p} \, \{ f(k,\vec{p},0) \exp[i(\vec{p}\cdot\vec{x} - \omega T)] + f^*(k,\vec{p},0) \exp[-i(\vec{p}\cdot\hat{x} - \omega T)] \} \tag{28}$$

and

$$\vec{J} = \frac{k}{4\pi^{3/2}} \int d\vec{p} \left(\frac{\vec{p}}{p}\right) \{ f(k,\vec{p},0) \exp[i(\vec{p}\cdot\vec{x} - \omega T)] + f^*(k,\vec{p},0) \exp[-i(\vec{p}\cdot\vec{x} - \omega T)] \} - \frac{k^2}{4\pi^{3/2}} \sum_{\lambda=\pm 1} \frac{\lambda}{2^{1/2}} \int \frac{d\vec{p}}{\omega(k,\vec{p})} \times \{ f(k,\vec{p},\lambda) \vec{\sigma}_1(k,\vec{p},\lambda) \exp[i(\vec{p}\cdot\vec{x} - \omega T)] + f^*(k,\vec{p},\lambda) \sigma_1^*(k,\vec{p},\lambda) \exp[-i(\vec{p}\cdot\vec{x} - \omega T)] \}, \tag{29}$$

which do not satisfy the equation of continuity and, therefore, can not be considered as source densities for electromagnetic fields satisfying Maxwell's equations. As such, even the transverse superluminal electromagnetic fields do not satisfy ordinary Maxwell's equations. The expression for a Lorentz force acting on an electrically charged tachyon interacting with superluminal electromagnetic fields has already been derived in our recent earlier paper.[19]

ACKNOWLEDGMENT

One of the authors, O. P. S. Negi, is thankful to University Grants Commission New Delhi, for providing financial assistance.

[1] E. Recami and R. Mignani, Rev. Nuovo Cimento **4**, 209 (1974).
[2] H. C. Corben, Nuovo Cimento A **29**, 415 (1975).
[3] M. T. Teli and V. K. Sutar, Lett. Nuovo Cimento **21**, 127 (1978).
[4] K. H. Mariwalla, Am. J. Phys. **37**, 1281 (1969).
[5] A. F. Antippa, Phys. Rev. D **11**, 724 (1975).
[6] A. F. Antippa and A. E. Everett, Phys. Rev. D **8**, 2352 (1973).
[7] F. Gonzalez-Gascon, Phys. Lett. A **56**, 249 (1976).
[8] H. Lemke, Nuovo Cimento A **27**, 141 (1975) and references therein.
[9] B. S. Rajput, K. D. Purohit, and O. P. S. Negi, "Reduction of four vector field in terms of standard helicity representation of unhomogeneous Lorentz group for imaginary mass system" Ind. J. Pure Appl. Phys.
[10] E. Recami and R. Mignani, Lett. Nuovo Cimento **9**, 367 (1973).
[11] V. Vysin, Lett. Nuovo Cimento **22**, 76 (1978).
[12] G. Dattoli and R. Mignani, Lett. Nuovo Cimento **22**, 65 (1978).
[13] P. Demers, Can. J. Phys. **53**, 1687 (1975).
[14] B. S. Rajput and K. D. Purohit, "Invariance of Maxwell's equations under real superluminal Lorentz transformations," Pramana, 1980.
[15] B. S. Rajput, K. D. Purohit, and O. P. S. Negi, "Study of fields associated with spin-½ tachyons" Prog. Theor. Phys.
[16] G. Ziino, Lett. Nuovo Cimento **24**, 171 (1979).
[17] P. T. Pappas, Lett. Nuovo Cimento **22**, 601 (1978).
[18] H. E. Moses, J. Math. Phys. **9** 2039 (1968).
[19] O. P. S. Negi, H. C. Chandola, K. D. Purohit, and B. S. Rajput, "Behavior of Maxwell's equations under real superluminal Lorentz transformations," Phys. Lett. B.

Finite-size effects on the relativistic electrons in a Coulomb field. I

H. Ogata
Department of Physics, University of Windsor, Windsor, Ontario, N9B 3P4, Canada
J. Asai
Linear Accelerator Laboratory, University of Saskatchewan, Saskatoon, Saskatchewan, S7N OWO, Canada

(Received 19 June 1980; accepted for publication 30 April 1982)

The influence of the finite nuclear size and screening effects on the evaluation of the interaction matrix elements of the Dirac electron in the nuclear Coulomb field is manipulated into a closed analytic form. An application to the evaluation of the Bremsstrahlung cross section is discussed.

PACS numbers: 21.10.Sf, 21.10.Ft, 02.30. + g

I. INTRODUCTION

In the high energy photonuclear reactions, it is customary to employ a photon radiator to convert the primary incident beam of electrons into the photon beam according to the Bremsstrahlung process. While it is so essential to have a precise knowledge of the Bremsstrahlung spectrum as to interpret the succeeding photonuclear reactions, there has not been a single reliable experimental determination of the spectral distribution. This makes it very important to have an accurate theoretical estimate of the photon intensity available for the entire photon spectrum and for various incident electron energies. There have been many such calculations, which were extensively compiled and categorized by Koch and Motz in their review.[1] Most of these calculations are essentially based on the Born approximation supplemented by various kinds of corrections, such as the screening corrections, or by the extreme relativistic approximations and so on. Since the electron energy is getting so high, there is no doubt that the fully relativistic treatment is mandatory.

In order to evaluate the interaction matrix elements, one needs the Coulomb wavefunctions by solving the Dirac equation for the electron in the nuclear field. The radial wavefunctions have been known in terms of the confluent hypergeometric functions, which made the evaluation of the matrix elements a considerably difficult task, because one had to perform an integration involving a product of two such functions, which required a summation of infinite series one way or the other. Earlier calculations involve certain kinds of extreme approximations, such as the Sommerfeld-Maue approximation,[2,3] to make the problem within the reach of the computational capability.

There is an elaborate work of Gargaro and Onley,[4] who have succeeded in expressing the integral of this kind in terms of the generalized hypergeometric functions for more than a couple of cases. Their radial integrals are extended from zero to infinity by assuming that the nucleus might be regarded as a point charge rather than having a finite size. There has been a considerable interest as to whether the finite nuclear size might affect this type of evaluation as the available energy of the primary incident electron is getting higher and higher.

In this paper we will report that these authors' method can be extended to finite integrals, some of which are direct extensions of their results and the rest are partly along the same line, but are more convenient for practical uses. When such finite integrals are considered, we bear in mind that either the finite size of the nucleus may affect the evaluation or the screening of the nuclear field by the atomic electron cloud would shield out the Coulomb field outside the Thomas-Fermi radius, or both. As far as the latter is concerned, there have been many attempts to include the idea of the finite range of the effective interaction into the theory,[5–7] but nothing has been seriously considered for the former. So far, all formulations simply ignore the finite nuclear size effect probably because of the mathematical complexity and of the optimistic anticipation that such an effect would be very small. However, the contribution arising from the integral around the origin may be neither negligible nor strongly divergent, but there is no doubt that the Coulomb field is moderately divergent near the origin if the point nucleus picture is used.

In the following, we will consider a finite integral of the type

$$\int_{r_1}^{r_2} dr\, r^{\alpha-1} e^{-K'r}\, {}_1F_1(a;b;K'r)\, {}_1F_1(\bar{a};\bar{b};Kr).$$

(Of course, one finite limit suffices for our discussion.)

When the finite-size nucleus is considered, we have assumed that no photon emission took place within the nuclear volume and that the finite nuclear charge distribution altered the pure Coulomb field solutions of the Dirac equation. In order to find the new wavefunctions, we have used a modified trapezoidal nuclear charge distribution[8] and calculated the radial wavefunctions outside, as well as inside, the nuclear radius. Since the nuclear potential range, other than the Coulomb potential, is finite, the effect of the nuclear potential is represented by an extra phase shift,[9] which can be evaluated at the nuclear surface together with the mixing amplitudes of the regular and irregular solutions. One can use the outside solutions thus obtained to evaluate the above integral with appropriate upper and lower bounds. We, however, use the approximate phase shifted regular solutions to evaluate the integral, which may be expressed as a difference of two improper integrals, to see if we can find a feasible result. It is further necessary to convert this phase shift into an appropriate form of the scattering phase shift,[10] because of the phase ambiguity of the Coulomb phase shift η by a multiple of π, which will be explained in Sec. III. The evaluation using the exact solutions is currently studied and may be

 0022-2488/82/101968-11$02.50

discussed in our future publication.

As for the screening effect, there has been a convincing criterion for the cutoff radius,[5–7] which seemed to be consistent with our preliminary estimate and will not be discussed any further in this context.

In the next section we will derive the closed expressions for the finite integral, which will be applied to the evaluation of the cross section of the Bremsstrahlung in Sec. III. A brief discussion is given in Sec. IV.

II. FINITE INTEGRALS

The integral we are considering is of the form

$$I_\rho = \int_0^\rho dr\, r^{\alpha-1} e^{-k'r}\,{}_1F_1(a;b;k'r)\,{}_1F_1(\bar a;\bar b;kr). \tag{1}$$

The limiting case of this integral when the upper limit goes to infinity has been studied by Gargaro and Onley.[4] Our evaluation method of the finite integral (1) closely resembles the method used by these authors and, therefore, we follow their procedure as nearly as possible at the beginning.

Case1: $|k|>|k'|$, $\mathrm{Re}(k)<0<\mathrm{Re}(k')$,

$\mathrm{Re}(b)>\mathrm{Re}(a)>0, \mathrm{Re}(\bar b)>\mathrm{Re}(\bar a)>0$, and $\mathrm{Re}(\alpha)>0$:

$$\begin{aligned} I_\rho =\; & \rho^\alpha \sum_{n=0}^{\infty} \Gamma\begin{bmatrix}\alpha+n\\ \alpha+1,n+1\end{bmatrix} \\ & \times {}_1F_1(\alpha+n;1+\alpha;-\rho)\sum_{m=0}^{n}\binom{n}{m}(k)^{-(\alpha+m)} \\ & \times\Big\{\Gamma\begin{bmatrix}\bar b,\bar a-(\alpha+m)\\ \bar a,\bar b-(\alpha+m)\end{bmatrix} e^{i\pi\alpha} \\ & \times {}_3F_2(b-a,\alpha+m,1+(\alpha+m)-\bar b; \\ & b,1+(\alpha+m)-\bar a;k'/k) \\ & +\Big(\frac{k'}{k}\Big)^{\bar a-(\alpha+m)} \\ & \times\Gamma\begin{bmatrix}b,\bar b,(\alpha+m)-\bar a,b-a-(\alpha+m)+\bar a\\ b-a,\bar b-\bar a,\alpha+m,b-(\alpha+m)+\bar a\end{bmatrix} \\ & \times e^{i\pi(\bar a-m)} \\ & \times {}_3F_2(\bar a,1+\bar a-\bar b,b-a-(\alpha+m)+\bar a; \\ & 1+\bar a-(\alpha+m),b+\bar a-(\alpha+m);k'/k)\Big\}, \end{aligned} \tag{2}$$

where

$$\Gamma\begin{bmatrix}a,b,\cdots\\ p,q,\cdots\end{bmatrix} = \frac{\Gamma(a)\Gamma(b)\cdots}{\Gamma(p)\Gamma(q)\cdots} \tag{3}$$

and ${}_3F_2$ are the generalized hypergeometric functions.

In order to prove this equation, the integrand ${}_1F_1$ is converted into an integral representation form by making use of Eq. (A6), after using Kummer's first theorem (A2) as given in the Appendix:

$$\begin{aligned} I_\rho = \;& \Gamma\begin{bmatrix}b,\bar b\\ a,\bar a,b-a,\bar b-\bar a\end{bmatrix}\int_0^1 du\, u^{b-a-1}(1-u)^{a-1} \\ & \times \int_0^1 dv\, v^{\bar a-1}(1-v)^{\bar b-\bar a-1}\int_0^\rho dr\, r^{\alpha-1}e^{-[k'u-kv]r}. \end{aligned} \tag{4}$$

The integral over r gives nothing but the incomplete gamma function[11]:

$$\int_0^\rho dr\, r^{\alpha-1}e^{-[k'u-kv]r} = [k'u-kv]^{-\alpha}\gamma(\alpha,[k'u-kv]\rho), \tag{5}$$

where ρ is the nuclear radius and $\gamma(\alpha,x)$ is

$$\begin{aligned} \gamma(\alpha,x) &= \int_0^x t^{\alpha-1}e^{-t}\,dt \quad [\mathrm{Re}(\alpha)>0] \\ &= \alpha^{-1}x^\alpha e^{-x}\,{}_1F_1(1;\alpha+1;x). \end{aligned} \tag{6}$$

Thus we can write the result (5) as

$$\alpha^{-1}\rho^\alpha e^{-[k'u-kv]\rho}\,{}_1F_1(1;\alpha+1;[k'u-kv]\rho).$$

Then we apply the multiplication theorem for ${}_1F_1$, (A12), and we get

$$\begin{aligned} & \alpha^{-1}\rho^\alpha e^{-[k'u-kv]\rho}\,{}_1F_1(1;\alpha+1;[k'u-kv]\rho) \\ & = \alpha^{-1}\rho^\alpha e^{-\rho}[k'u-kv]^{-\alpha}\sum_{n=0}^{\infty}\frac{(\alpha)_n}{n!} \\ & \times(1-[k'u-kv]^{-1})^n\,{}_1F_1(1-n;1+\alpha;\rho) \\ & = \rho^\alpha\sum_{n=0}^{\infty}\Gamma\begin{bmatrix}\alpha+n\\ \alpha+1,n+1\end{bmatrix}{}_1F_1(\alpha+n;1+\alpha;-\rho) \\ & \times\sum_{m=0}^{n}(-1)^m\binom{n}{m}[k'u-kv]^{-(\alpha+m)}, \end{aligned} \tag{7}$$

where $(\alpha)_n = \Gamma\left[\begin{smallmatrix}\alpha+n\\ \alpha\end{smallmatrix}\right]$, $\binom{n}{m}$ are the binomial coefficients, and use has been made of the Kummer's first theorem (A2) and the binomial theorem.

Substituting (7) into (4),

$$\begin{aligned} I_\rho = \;& \Gamma\begin{bmatrix}b,\bar b\\ a,\bar a,b-a,\bar b-\bar a\end{bmatrix}\int_0^1 du\, u^{b-a-1} \\ & \times(1-u)^{a-1}\int_0^1 dv\, v^{\bar a-1}(1-v)^{\bar b-\bar a-1} \\ & \times\rho^\alpha\sum_{n=0}^{\infty}\begin{bmatrix}\alpha+n\\ \alpha+1,n+1\end{bmatrix} \\ & \times{}_1F_1(\alpha+n;1+\alpha;-\rho)\sum_{m=0}^{n}(-1)^m \\ & \times\binom{n}{m}(k'u)^{-(\alpha+m)}\Big[1-\Big(\frac{k}{k'u}\Big)v\Big]^{-(\alpha+m)} \end{aligned}$$

Integration over v is carried out, using (A7), and we get

$$\begin{aligned} I_\rho = \;& \Gamma\begin{bmatrix}b\\ a,b-a\end{bmatrix}\rho^\alpha\sum_{n=0}^{\infty}\Gamma\begin{bmatrix}\alpha+n\\ \alpha+1,n+1\end{bmatrix} \\ & \times{}_1F_1(\alpha+n;1+\alpha;-\rho)\sum_{m=0}^{n}(-1)^m \\ & \times\binom{n}{m}(k')^{-(\alpha+m)} \\ & \times\int_0^1 du\, u^{b-a-(\alpha+m)-1}(1-u)^{a-1} \\ & \times{}_2F_1\Big(\alpha+m,\bar a;\bar b;\frac{k}{k'u}\Big). \end{aligned} \tag{8}$$

Since the absolute value of the argument of ${}_2F_1$, $|k/k'u|$, is greater than unity, one has to use the analytic continuation formula (A11) to convert the argument into its inverse, and express (8) as

$$I_\rho = \Gamma\begin{bmatrix} b \\ a,b-a \end{bmatrix}\rho^\alpha \sum_{n=0}^{\infty}\Gamma\begin{bmatrix} \alpha+n \\ \alpha+1,n+1 \end{bmatrix}$$
$$\times {}_1F_1(\alpha+n;1+\alpha;-\rho)\sum_{m=0}^{n}(-1)^m\binom{n}{m}(k')^{-(\alpha+m)}$$
$$\times\Big\{\int_0^1 du\, u^{b-a-1}(1-u)^{a-1}$$
$$\times\Gamma\begin{bmatrix} \bar b,\bar a-(\alpha+m) \\ \bar a,\bar b-(\alpha+m) \end{bmatrix}(-1)^{-(\alpha+m)}\Big(\frac{k}{k'}\Big)^{-(\alpha+m)}$$
$$\times {}_2F_1\Big(\alpha+m,1+(\alpha+m)-\bar b;$$
$$1+(\alpha+m)-\bar a;\frac{k'}{k}u\Big)$$
$$+\int_0^1 du\, u^{b-a-(\alpha+m)+\bar a-1}(1-u)^{a-1}$$
$$\times\Gamma\begin{bmatrix} \bar b,(\alpha+m)-\bar a \\ \alpha+m,\bar b-\bar a \end{bmatrix}(-1)^{-\bar a}\Big(\frac{k}{k'}\Big)^{-\bar a}$$
$$\times {}_2F_1\Big(\bar a,1+\bar a-\bar b;1+\bar a-(\alpha+m);\frac{k'}{k}u\Big)\Big\}. \quad (9)$$

Finally, applying (A8), one gets Eq. (2).

Case 2: $|k'|>|k|$, $\mathrm{Re}(k)<0<\mathrm{Re}(k')$, $\mathrm{Re}(b)>\mathrm{Re}(a)>0$, $\mathrm{Re}(\otimes(b))>\mathrm{Re}(\bar a)>0$, and $\mathrm{Re}(\alpha)>0$.

$$I_\rho = \rho^\alpha\sum_{n=0}^{\infty}\Gamma\begin{bmatrix} \alpha+n \\ \alpha+1,n+1 \end{bmatrix}$$
$$\times {}_1\mathrm{F}_1(\alpha+n;1+\alpha;-\rho)\sum_{m=0}^{n}\binom{n}{m}(k')^{-(\alpha+m)}e^{i\pi m}$$
$$\times\Big\{\Gamma\begin{bmatrix} b,b-a-(\alpha+m) \\ b-a,b-(\alpha+m) \end{bmatrix}$$
$$\times {}_3F_2(\bar a,\alpha+m,1+(\alpha+m)-b;$$
$$\bar b,1+(\alpha+m)+a-b;\frac{k}{k'}\Big)$$
$$+\Big(\frac{k'}{k}\Big)^{a-b+(\alpha+m)}$$
$$\times\Gamma\begin{bmatrix} b,\bar b,(\alpha+m)-b+a,\bar a-(\alpha+m)+b-a \\ \bar a,a,\alpha+m,\bar b-(\alpha+m)+b-a \end{bmatrix}$$
$$\times e^{i\pi(a-b+\alpha+m)}$$
$$\times {}_3F_2(1-a,b-a,\bar a-(\alpha+m)+b-a;$$
$$1+b-a-(\alpha+m),\bar b+b-a-(\alpha+m);$$
$$\frac{k}{k'}\Big)\Big\}. \quad (10)$$

The proof of this result is similar to the Case 1. After integrating over r, we get

$$I_\rho = \Gamma\begin{bmatrix} b,\bar b \\ a,\bar a,b-a,\bar b-\bar a \end{bmatrix}\int_0^1 du\, u^{b-a-1}(1-u)^{a-1}$$
$$\times\int_0^1 dv\, v^{\bar a-1}(1-v)^{\bar b-\bar a-1}\rho^\alpha\sum_{n=0}^{\infty}\Gamma\begin{bmatrix} \alpha+n \\ \alpha+1,n+1 \end{bmatrix}$$
$$\times {}_1F_1(\alpha+n;1+\alpha;-\rho)$$
$$\times\sum_{m=0}^{n}\binom{n}{m}(-1)^{-\alpha}(kv)^{-(\alpha+m)}$$
$$\times\Big[1-\Big(\frac{k'}{kv}\Big)u\Big]^{-(\alpha+m)}. \quad (11)$$

Integration over u can be done by using (A7)

$$I_\rho = \Gamma\begin{bmatrix} \bar b \\ \bar a,\bar b-\bar a \end{bmatrix}(-1)^\alpha\rho^\alpha\sum_{n=0}^{\infty}\Gamma\begin{bmatrix} \alpha+n \\ \alpha+1,n+1 \end{bmatrix}$$
$$\times {}_1F_1(\alpha+n;1+\alpha;-\rho)\sum_{m=0}^{n}\binom{n}{m}(k)^{-(\alpha+m)}$$
$$\times\int_0^1 dv\, v^{\bar a-(\alpha+m)-1}(1-v)^{\bar b-\bar a-1}$$
$$\times {}_2F_1\Big(\alpha+m,b-a;b;\frac{k'}{kv}\Big). \quad (12)$$

Using the analytic continuation formulas (A11) and (A8), integration over v can be done, and one gets (10).

Case 3: $k=k'$, $\mathrm{Re}(k)=0$, $\mathrm{Re}(b)>\mathrm{Re}(a)>0$, $\mathrm{Re}[\bar b-\bar a+a-(\alpha+m)]>0$, for all values of $(\alpha+m)-b+1$ except zero or negative integers:

$$I_\rho = \rho^\alpha\sum_{n=0}^{\infty}\Gamma\begin{bmatrix} \alpha+n \\ \alpha+1,n+1 \end{bmatrix}{}_1F_1(\alpha+n;1+\alpha;-\rho)$$
$$\times\sum_{m=0}^{n}\binom{n}{m}(k)^{-(\alpha+m)}$$
$$\times\{e^{i\pi(\alpha+\bar a-\bar b)}$$
$$\times\Gamma\begin{bmatrix} \bar b,1+(\alpha+m)-\bar b \\ \bar a,1+(\alpha+m)-\bar a \end{bmatrix}$$
$$\times {}_3F_2(b-a,\alpha+m,1+(\alpha+m)-\bar b;$$
$$b,1+(\alpha+m)-\bar a;1)$$
$$+e^{i\pi(\bar a-m)}\Gamma\begin{bmatrix} \bar b,1+(\alpha+m)-\bar b \\ \bar b-\bar a,1+(\alpha+m)+\bar a-\bar b \end{bmatrix}$$
$$\times {}_3F_2(a,\alpha+m,1+(\alpha+m)-\bar b;$$
$$b,1+(\alpha+m)+\bar a-\bar b;1)\}. \quad (13)$$

We start with the recurrence relation[12] for the unit argument (A10).

$$\Gamma\begin{bmatrix} a_1,1+a_2+a_3-b_2,1-b_2 \\ b_1,1+a_2-b_2,1+a_3-b_2 \end{bmatrix}{}_3F_2(a_1,a_2,a_3;b_1b_2;1)$$
$$=\Gamma\begin{bmatrix} a_1 \\ b_1 \end{bmatrix}{}_3F_2(b_1-a_1,a_2,a_3;b_1,1+a_2+a_3-b_2;1)$$
$$-\Gamma\begin{bmatrix} 1+a_1-b_2,1+a_2+a_3-b_2,b_2-1 \\ 1+b_1-b_2,a_2,a_3 \end{bmatrix}$$
$$\times {}_3F_2(1+a_1-b_2,1+a_2-b_2,1+a_3-b_2;$$
$$1+b_1-b_2,2-b_2;1). \quad (14)$$

Choosing $a_1 = b - a,\ a_2 = \alpha + m,$

$a_3 = 1 + (\alpha + m) - \bar{b},\ b_1 = b,$ and $b_2 = 1 + (\alpha + m) - \bar{a},$ the identity (14) gives

$$\Gamma\begin{bmatrix} b,\bar{b},(\alpha+m)-\bar{a},b-a-(\alpha+m)+\bar{a} \\ b-a,\bar{b}-\bar{a},\alpha+m,b-(\alpha+m)+\bar{a} \end{bmatrix}$$
$$\times {}_3F_2(\bar{a},1+\bar{a}-\bar{b},b-a-(\alpha+m)+\bar{a};$$
$$1+\bar{a}-(\alpha+m),b+\bar{a}-(\alpha+m);1)$$
$$=\Gamma\begin{bmatrix} \bar{b},1+(\alpha+m)-\bar{b} \\ \bar{b}-\bar{a},1+\bar{a}-\bar{b}+(\alpha+m) \end{bmatrix}$$
$$\times {}_3F_2(a,\alpha+m,1+(\alpha+m)-\bar{b};b,1+\bar{a}-\bar{b}+(\alpha+m);1)$$
$$-\Gamma\begin{bmatrix} \bar{b},1+(\alpha+m)-\bar{b},\bar{a}-(\alpha+m) \\ \bar{b}-\bar{a},\bar{a},1+\bar{a}-\bar{b} \end{bmatrix}$$
$$\times {}_3F_2(b-a,\alpha+m,1+(\alpha+m)-\bar{b};b,1+(\alpha+m)-\bar{a};1). \tag{15}$$

Setting (k'/k) in Eq. (2) equal to unity and substituting (15), we get

$$I_\rho = \rho^\alpha \sum_{n=0}^{\infty} \Gamma\begin{bmatrix} \alpha+n \\ \alpha+1,n+1 \end{bmatrix} {}_1F_1(\alpha+n;1+\alpha;-\rho)$$
$$\times \sum_{m=0}^{n} \binom{n}{m} (k)^{-(\alpha+m)}$$
$$\times\left\{\left(e^{i\pi\alpha}\Gamma\begin{bmatrix} \bar{b},\bar{a}-(\alpha+m) \\ \bar{a},\bar{b}-(\alpha+m) \end{bmatrix}\right.\right.$$
$$-e^{i\pi(\bar{a}-m)}$$
$$\left.\times\Gamma\begin{bmatrix} \bar{b},\bar{a}-(\alpha+m),1+(\alpha+m)-\bar{b} \\ \bar{a},\bar{b}-\bar{a},1+\bar{a}-\bar{b} \end{bmatrix}\right)$$
$$\times {}_3F_2(b-a,\alpha+m,1+(\alpha+m)-\bar{b};$$
$$b,1+(\alpha+m)-\bar{a};1)$$
$$+e^{i\pi(\bar{a}-m)}\Gamma\begin{bmatrix} \bar{b},1+(\alpha+m)-\bar{b} \\ \bar{b}-\bar{a},1+\bar{a}-\bar{b}+(\alpha+m) \end{bmatrix}$$
$$\times {}_3F_2(a,\alpha+m,1+(\alpha+m)-\bar{b};$$
$$\left. b,1+\bar{a}-\bar{b}+(\alpha+m);1)\right\}.$$

Using the relation $\Gamma(c)\Gamma(1-c) = \pi/\sin\pi c$, the factor of the first term in the braces can be expressed in a single form, and Eq. (13) follows.

These results contain infinite series ${}_3F_2$ to be evaluated. We can also show that there are another form of these results expressed in terms of finite series, by making use of another formula of the multiplication theorem (A13).

Case 1': $|k| > |k'|$ and other conditions are the same as for Case 1:

$$I_\rho = \sum_{n=0}^{\infty} \frac{\gamma(\alpha+n;\rho)}{n!} \sum_{m=0}^{n} \binom{n}{m} (k)^m \Gamma\begin{bmatrix} \bar{b},\bar{a}+m \\ \bar{a},\bar{b}+m \end{bmatrix}$$
$$\times {}_3F_2\left(b-a,-m,1-m-\bar{b};b,1-m-\bar{a};\frac{k'}{k}\right). \tag{16}$$

After integrating over r, we select the alternative choice of (A13):

$$\alpha^{-1}\rho^\alpha\ {}_1F_1(\alpha;\alpha+1;-[k'u-kv]\rho)$$
$$=\rho^\alpha \times \sum_{n=0}^{\infty} \frac{(\alpha)_n(-\rho)^n}{\alpha(\alpha+1)_n n!}$$
$$\times([k'u-kv]-1)^n$$
$$\times {}_1F_1(\alpha+n;1+\alpha+n;-\rho)$$
$$=\sum_{n=0}^{\infty} \frac{\gamma(\alpha+n,\rho)}{n!}$$
$$\times \sum_{m=0}^{n} (-1)^m \binom{n}{m} (k'u)^m$$
$$\times\left[1-\left(\frac{k}{k'u}\right)v\right]^m. \tag{17}$$

Integrating over v, we have

$$I_\rho = \Gamma\begin{bmatrix} b \\ a,b-a \end{bmatrix} \sum_{n=0}^{\infty} \frac{\gamma(\alpha+n,\rho)}{n!} \sum_{m=0}^{n} (-1)^m \binom{n}{m} (k')^m$$
$$\times \int_0^1 du\, u^{b-a+m-1}(1-u)^{a-1}\ {}_2F_1\left(-m,\bar{a};\bar{b};\frac{k}{k'u}\right). \tag{18}$$

As we have done before, the analytic continuation formula must be used in order to make the absolute value of the argument of ${}_2F_1$ less than unity. If we formally apply (A11), we get

$$I_\rho = \Gamma\begin{bmatrix} b \\ a,b-a \end{bmatrix}$$
$$\times \sum_{n=0}^{\infty} \frac{\gamma(\alpha+n,\rho)}{n!} \sum_{m=0}^{n} (-1)^m \binom{n}{m} (k')^m$$
$$\times \int_0^1 du\, u^{b-a+m-1}(1-u)^{a-1}$$
$$\times\left\{\Gamma\begin{bmatrix} \bar{b},\bar{a}+m \\ \bar{a},\bar{b}+m \end{bmatrix}\left(-\frac{k}{k'}\right)^m\right.$$
$$\times u^{-m}{}_2F_1\left(-m,1-m-\bar{b};1-m-\bar{a};\frac{k'}{k}u\right)$$
$$+\Gamma\begin{bmatrix} \bar{b},-\bar{a}-m \\ -m,\bar{b}-\bar{a} \end{bmatrix}\left(-\frac{k}{k'}\right)^{-\bar{a}}$$
$$\left.\times u^{\bar{a}}\ {}_2F_1\left(\bar{a},1+\bar{a}-\bar{b};1+\bar{a}+m;\frac{k'}{k}u\right)\right\}. \tag{19}$$

The second term in the curly bracket has an infinitely large value of $\Gamma(-m)$ in its denominator and, hence, it should vanish. As a matter of fact, the ${}_2F_1$ in Eq. (18) is nothing but the Jacobi polynomial, and its analytic continuation formula lacks the second term of (19). This is shown in the Appendix. Dropping the second term and integrating over u, one gets (16).

Case 2': $|k'| > |k|$ and other conditions are the same as for Case 2:

$$I_\rho = \sum_{n=0}^{\infty} \frac{\gamma(\alpha+n,\rho)}{n!} \sum_{m=0}^{n} \binom{n}{m} (-k')^m \Gamma\begin{bmatrix} b,b-a+m \\ b-a,b+m \end{bmatrix}$$
$$\times {}_3F_2\left(\bar{b}-\bar{a},-m,1-m-b;\bar{b},a-b-m+1;\frac{k}{k'}\right). \tag{20}$$

Again, we start from Eq. (4) and integrate over r. After applying the multiplication theorem (A13), we get

$$\alpha^{-1}\rho^{\alpha}\,{}_1F_1(\alpha;\alpha+1;-[k'u-kv]\rho) = \sum_{n=0}^{\infty}\frac{\gamma(\alpha+n,\rho)}{n!}\sum_{m=0}^{n}\binom{n}{m}(kv)^m\left[1-\left(\frac{k'}{kv}\right)u\right]^m. \quad (21)$$

Integrating over u, we again have a Jacobi polynomial:

$$I_\rho = \Gamma\left[\frac{\bar b}{\bar a,\bar b-\bar a}\right]\sum_{n=0}^{\infty}\frac{\gamma(\alpha+n,\rho)}{n!}\sum_{m=0}^{n}\binom{n}{m}(k)^m \times\int_0^1 dv\, v^{\bar a+m-1}(1-v)^{\bar b-\bar a-1}\,{}_2F_1\left(-m,b-a;b;\frac{k'}{kv}\right). \quad (22)$$

Formal application of the analytic continuation formula results in a vanishing term as before, thus proving (20).

III. APPLICATION TO BREMSSTRAHLUNG

As was discussed in the first section, we apply the result of the preceding section to the Bremsstrahlung. The photon beam in the photonuclear reactions is generated by the Bremsstrahlung from the incident electrons interacting with the nuclear Coulomb field of a radiator. The cross section can be expressed as

$$d^3\sigma = \alpha\frac{W_1}{p_1}\frac{p_2W_2k\,dk\,d\Omega_k}{(2\pi)^4}\sum_{\epsilon}d\Omega_2\sum_{\zeta_2}|M_{if}|^2, \quad (23)$$

where

$$|M_{if}|^2 = \left|\int \Psi_2^\dagger\boldsymbol{\epsilon}\cdot\boldsymbol{\alpha}e^{-i\mathbf{k}\cdot\mathbf{r}}\Psi_1\,d^3r\right|^2. \quad (24)$$

The quantities p, W, and ζ are the momentum, energy, and spin of the electron, respectively, and $\boldsymbol{\epsilon}$ is the unit polarization vector of the outgoing photon of momentum $\mathbf{k}$. Ψ's are the electron wavefunctions and the subscripts 1 and 2 refer to the initial and final states, respectively. $\boldsymbol{\alpha}$ is the Dirac matrix, and $\boldsymbol{\alpha} = \begin{pmatrix}0 & \sigma\\ \sigma & 0\end{pmatrix}$, where $\boldsymbol{\sigma}$ is the Pauli spin matrix. We use the natural unit: $\hbar = c = m = 1$. In calculating the Bremsstrahlung intensity spectrum, the most crucial part arises from the radial integration in Eq. (24). We intend to proceed as strictly as possible and the following plans are adopted unless otherwise stated:

(i) The electron mass is kept in the entire calculation.

(ii) The finite size of the radiator nucleus is taken into account by considering the uniformly charged sphere of radius ρ.

(iii) The electron wavefunctions are expanded in partial waves with the appropriate phase shifts.

(iv) The nuclear recoil momentum is kept in our kinematics.

The Dirac equation for the electron in the nuclear Coulomb field is

$$[\boldsymbol{\alpha}\cdot\mathbf{p}+\beta+V]\psi = W\psi, \quad (25)$$

where

$$\beta = \begin{pmatrix}I & 0\\ 0 & -I\end{pmatrix}.$$

The potential V assumes the form

$$V = -\alpha Z/r \quad \text{for } r>\rho, \quad (26)$$

$$V = -(\alpha Z/2\rho)(3-r^2/\rho^2) \quad \text{for } r<\rho. \quad (27)$$

Here α is the fine structure constant. The potential form is in accord with our plan (ii). Z is the nuclear charge of the radiator. The solutions for the Dirac equation (25) are written in the form, for a given angular momentum,

$$\psi_\kappa^\mu = \begin{pmatrix} g_\kappa(r) & \chi_\kappa^\mu(\hat r)\\ if_\kappa(r) & \chi_{-\kappa}^\mu(\hat r)\end{pmatrix}, \quad (28)$$

where $\hat r$ is the coordinate unit vector of the electron and χ_κ^μ are the spin-angular functions, which are written[13]

$$\chi_\kappa^\mu = \sum_m (l\,\mu-m\;s\;m|ls\;j\mu)Y_{l,\mu-m}(\hat r)\chi^m, \quad (29)$$

where $(l\,\mu-m\;sm|ls\;j\mu)$ are the Clebsch–Gordan coefficients, $Y_{l,\mu-m}$ are the spherical harmonics, and χ^m are the spin functions. These spin-angular functions χ_κ^μ and $\chi_{-\kappa}^\mu$ are the eigenfunctions of the operator $K=(1+\boldsymbol{\sigma}\cdot\mathbf{L})$. Angular momentum quantum numbers l, l', and j are defined according to

$$j = |\kappa|-\tfrac{1}{2}, \quad (30)$$

$$l_\kappa = l = \begin{cases}\kappa & \text{if } \kappa>0,\\ -\kappa-1 & \text{if } \kappa<0,\end{cases}$$

$$l_{-\kappa} = l' = \begin{cases}\kappa-1 & \text{if } \kappa>0,\\ -\kappa & \text{if } \kappa<0.\end{cases} \quad (31)$$

When a point nucleus is considered, only the solutions regular at the origin are needed, and they are expressed as

$$\left.\begin{matrix} rf_\kappa^C(r)\\ rg_\kappa^C(r)\end{matrix}\right\} = \left\{\begin{matrix} i(1-W)^{1/2}\\ (1+W)^{1/2}\end{matrix}\right. \frac{(2pr)^\gamma e^{(\pi/2)y}|\Gamma(\gamma+iy)|}{(\pi p)^{1/2}\Gamma(2\gamma+1)} \times\left\{\begin{matrix}\mathrm{Im}\\ \mathrm{Re}\end{matrix}\right\}[(\gamma+iy)e^{-ipr}e^{i\eta} \times{}_1F_1(\gamma+1+iy;2\gamma+1;2ipr)], \quad (32)$$

where the superscript C denotes the pure Coulomb field and

$$\gamma^2 = \kappa^2-(\alpha Z)^2,$$
$$y = \alpha ZW/p, \quad (33)$$
$$e^{2i\eta} = -(\kappa-iy/W)/(\gamma+iy).$$

As far as the electron wavefunctions are concerned, we find a very close resemblance between the Bremsstrahlung and the beta-decay process for which the conditions (i), (ii), and (iii) are all met. Since we have a very reliable code for the beta-decay analysis, as was demonstrated in our earlier publication,[8] we have extended our code for the application to the Bremsstrahlung. We should, however, note that the phase convention employed in the beta-decay analysis is not the same as that employed in the electron scattering formalism, that is, the asymptotic form of the radial wavefunction, regular at the origin, say, in the electron scattering case (designated as YRW[14]) is

$$rg_\kappa \sim \sin(pr+y\ln 2pr-\tfrac{1}{2}(j-\tfrac{1}{2})\pi+\Delta_\kappa), \quad (34)$$

while in the beta-decay case (designated as BR[15])

$$rg_\kappa \sim \sin(pr+y\ln 2pr+\Delta_\kappa). \quad (35)$$

A method transforming the BR to the YRW scheme is described elsewhere.[10] It should be mentioned that the electron mass is set equal to zero in the YRW method. Table I shows our phase shifts obtained via BR together with the YRW

TABLE I. Values of phase shifts for gold.[a]

	YRW[b] ($\kappa > 0$)		Ours ($\kappa > 0$)		Ours ($\kappa < 0$)	
$\lvert\kappa\rvert$	δ^C	$\Delta - \delta^C$	δ^C	$\Delta - \delta^C$	δ^C	$\Delta - \delta^C$
1	0.407 36	− 0.858 20	0.406 04	− 0.857 14	0.408 65	− 0.859 58
2	− 0.237 97	− 0.271 43	− 0.238 62	− 0.271 59	− 0.237 13	− 0.272 58
3	− 0.533 03	− 0.076 33	− 0.533 47	− 0.75 98	− 0.532 60	− 0.076 36
4	− 0.726 59	− 0.014 94	− 0.726 91	− 0.014 91	− 0.726 26	− 0.015 01
5	− 0.870 98	− 0.001 99	− 0.871 23	− 0.001 96	− 0.870 71	− 0.001 98
6	− 0.986 23	− 0.000 17	− 0.986 44	− 0.000 19	− 0.986 00	− 0.000 19
7	− 1.082 18	− 0.000 01	− 1.082 36	− 0.000 01	− 1.081 98	− 0.000 01
8	− 1.164 38	− 0.000 00	− 1.164 52	− 0.000 00	− 1.164 20	− 0.000 00
9	− 1.236 28	− 0.000 00	− 1.236 41	− 0.000 00	− 1.236 12	− 0.000 00

[a] The electron mass is neglected here, and the incident energy of the electron is taken to be 113.12 MeV in our calculation.
[b] See Ref. 14.

results. For the sake of completeness we also list in Table I the Coulomb phase shifts δ_κ^C due to a point nucleus. As can be seen from Table I, our phase shifts are in excellent agreement with those of YRW. Table II shows our results when the electron mass is not taken to be zero.

For a point nucleus the irregular solution does not play a role in the radial integration involved in Eq. (24), but even the regular solution becomes mildly divergent at the origin; e.g., for $|\kappa| = 1$, $g_\kappa^C \approx f_\kappa^C \approx r^{\gamma-1}$. The divergence associated with this behavior of the wavefunction is inevitable for the point nucleus approximation. For an extended nucleus, on the other hand, the divergence would not occur if we integrate from the nuclear surface to the cutoff radius of the screened potential. It is then necessary to find the external electron wavefunction expressed as a linear combination of the regular and irregular solutions, which should be matched to the internal solution at the nuclear surface. We, however, in this work, only consider the regular solution whose asymptotic form is given by Eq. (34). If there is indeed an appreciable difference between a point nucleus approach and our approximation method, we may attribute the difference to the nuclear finite-size effect. More rigorous method in which the irregular solution is also taken into account will be presented in our forthcoming paper.

Once this approximation is adopted, the electron wavefunctions for the initial and final states are of the form

$$\Psi^{(\mathrm{o,i})} = 4\pi\left(\frac{\pi}{2Wp}\right)^{1/2} \sum_{\kappa\mu} S_\kappa (\chi_\kappa^{\mu}(\hat p)^\dagger v)\psi_\kappa^{\mu(\mathrm{o,i})}(\mathbf{r},\zeta), \tag{36}$$

TABLE II. Values of $(\Delta - \delta_j^C)$ for gold.[a]

$\lvert\kappa\rvert$	$\kappa > 0$	$\kappa < 0$
1	− 0.856 80	− 0.859 25
2	− 0.271 09	− 0.272 09
3	− 0.075 94	− 0.076 31
4	− 0.014 84	− 0.014 93
5	− 0.001 96	− 0.001 97
6	− 0.000 19	− 0.000 19
7	− 0.000 01	− 0.000 01
8	− 0.000 00	− 0.000 00

[a] The electron mass is explicitly included in this calculation.

where v is the "large component" of the Dirac spinor and S_κ is the amplitude of the κ-component eigenfunction. Since the initial and final states are asymptotically described by a plane wave plus outgoing and ingoing spherical waves, respectively, the superscripts (o) and (i) distinguish these spherical waves.

$$\psi_\kappa^{\mu(\mathrm{o})} \propto e^{i\Delta_\kappa}(i)^l \psi_\kappa^\mu \tag{37}$$

and ψ_κ^μ is defined in Eq. (28). For $\psi_\kappa^{\mu(\mathrm{i})}$ a phase factor $e^{i\Delta_\kappa}$ should be replaced by $e^{-i\Delta_\kappa}$.

Having obtained the electron wavefunctions for the initial and final states, we rewrite Eqs. (23) and (24) in a calculable form. The photon plane wave[16] is expanded in terms of the spherical harmonics:

$$e^{-i\mathbf{k}\cdot\mathbf{r}} = 4\pi\sum_{lm}(i)^{-l}\tfrac{1}{2}[h_l^{(1)}(kr) + h_l^{(2)}(kr)] \times Y_{lm}(\hat k)Y_{lm}^*(\hat r), \tag{38}$$

where $h_l^{(i)}(kr)$ are the spherical Hankel functions of the ith kind. In the following we will consider the outgoing component, $h_l^{(1)}(kr)$, only. After summing over polarization and the final state spin variables and integrating over the final electron angles, we have[17]

$$d^3\sigma = \frac{\alpha}{(2\pi)^4}\frac{p_2W_1W_2}{p_1}k\,dk\,d\Omega_k \times\left[\sum_{L=0}^{\infty}A_LP_L(\cos\theta) + \hat n\cdot\hat\zeta_1\sum_{L=1}^{\infty}B_LP_{L,1}(\cos\theta)\right], \tag{39}$$

where θ is the angle between $\mathbf{p}_1$ and $\mathbf{k}$, $\hat n$ is a unit vector in the direction of $\mathbf{k}\times\mathbf{p}_1$, and $P_{L,M}(\cos\theta)$ are the associated Legendre polynomials. In practice, we are interested in the photon intensity averaged out for the photon angles; thus only the first term in the expression (39) is the most important. The coefficient A_L is expressed, in general, as

$$A_L = 24(4\pi)^3\sum_{\kappa_1\kappa_2\bar\kappa_1}\sum_{l\bar l}\lambda_1\lambda_2\bar\lambda_1\hat{\hat{l}}\hat{\hat{\bar l}}(\hat l_1\hat{\bar l}_1)^{1/2}(-1)^{\bar l+\bar l_1} \times(l0\bar l0|l\bar lL0)\sum_{jj'}T_{\kappa_1\kappa_2\bar\kappa_1,l\bar l;jj'L}\{DI\}, \tag{40}$$

$$T_{\kappa_1\kappa_2\bar\kappa_1 l\bar l;f\bar f L} = \tfrac{2}{3}\delta_{f\bar f}\hat f(l_1 0\bar l_1 0|l_1\bar l_1 L\,0) \times W(L\bar l j\bar f;\bar l j_1) W(l_1\tfrac12 L\bar j_1;j_1\bar l_1) + 5^{1/2}(1010|1120)\hat f\hat{\bar f}W(f j_2 21;1\bar f) \times \sum_{L'}(\hat L')^{1/2}(l_1 0\bar l_1 0|l_1\bar l_1 L'0) \times(L'020|L'2L\,0)W(l_1\tfrac12 L\bar j_1;j_1\bar l_1) \times\begin{Bmatrix}2 & L' & L\\ f & j_1 & l\\ \bar f & \bar j_1 & \bar l\end{Bmatrix}, \tag{41}$$

$$\{DI\} = D_{\kappa_2-\kappa_1 l}{}^f I_{\kappa_1\kappa_2 l}(D_{\bar\kappa_2-\bar\kappa_1\bar l}{}^{\bar f} I_{\bar\kappa_1\bar\kappa_2\bar l})^* + D_{-\kappa_2\kappa_1 l}{}^f J_{\kappa_1\kappa_2 l}(D_{-\bar\kappa_2\bar\kappa_1\bar l}{}^{\bar f} J_{\bar\kappa_1\bar k_2\bar l})^* - D_{\kappa_2-\kappa_1 l}{}^f I_{\kappa_1\kappa_2 l}(D_{-\bar\kappa_2\bar\kappa_1\bar l}{}^{\bar f} J_{\bar\kappa_1\bar\kappa_2\bar l})^* - D_{-\kappa_2\kappa_1 l}{}^f J_{\kappa_1\kappa_2 l}(D_{\bar\kappa_2-\bar\kappa_1\bar l}{}^{\bar f} I_{\bar\kappa_1\bar\kappa_2\bar l})^*, \tag{42}$$

and

$$D_{\kappa_1\kappa_2 l}{}^f = (-1)^{l_1}(\hat l_1)^{1/2}(l\,0l_1 0|l l_1 l_2 0) \times W(l_1 j_1 l_2 f;\tfrac12 l)W(1\tfrac12 j_2 l_2;\tfrac12 f), \tag{43}$$

where angular momentum numbers with a hat $(\hat l, \hat f$, etc.) are $\hat l = 2l+1$, $\lambda = |\kappa| = j+\frac12$, $W(abcd;ef)$ are the Racah coefficients and

$$\begin{Bmatrix}a & b & e\\ c & d & e'\\ f & f' & g\end{Bmatrix}$$

are the 9-J symbols. The radial integrals are

$$I_{\kappa_1\kappa_2 l} = \int_\rho^\infty r^2\,dr\,S_{\kappa_1}S_{\kappa_2}(i)^{1-l}f_{\kappa_1}{}^{(o)}g_{\kappa_2}{}^{(i)*}h_l^{(1)}(kr), \tag{44}$$

$$J_{\kappa_1\kappa_2 l} = \int_\rho^\infty r^2\,dr\,S_{\kappa_1}S_{\kappa_2}(i)^{1-l}g_{\kappa_1}{}^{(o)}f_{\kappa_2}{}^{(i)*}h_l^{(1)}(kr). \tag{45}$$

After integrating over the photon angles Ω_k,

$$d\sigma = \frac{4\pi\alpha}{(2\pi)^4}\frac{p_2W_1W_2}{p_1}k\,dk\,A_0, \tag{46}$$

where A_0 is given by

$$A_0 = 12(4\pi)^3\sum_{\kappa_1\kappa_2\bar\kappa_1}\lambda_2\bar\lambda_1\sum_l \hat l\sum_{f\bar f}\{DI\} \times\big[\tfrac23\delta_{f\bar f}\,\delta_{l_1\bar l_1}\,\delta_{j_1\bar j_1} + 5^{1/2}(1010|1120)\bar f\hat j_1(\hat l_1)^{1/2}(l_1 020|l_1 2\bar l_1 0) \times W(f j_2 21;1\bar f)W(l_1\tfrac12 2\bar j_1;j_1\bar l_1)W(\bar f 2 l j_1;f\bar j_1)\big]. \tag{47}$$

The spherical Hankel functions in the integrands are now expanded[18]:

$$h_l^{(1)}(kr) = e^{ikr}\sum_{n=1}^{l+1}\Gamma\begin{bmatrix}l+n\\ n,2+l-n\end{bmatrix}(i)^{n-l-2}(2kr)^{-n}. \tag{48}$$

Substituting (48) into (44), we get

$$I_{\kappa_1\kappa_2 l} = S_{\kappa_1}S_{\kappa_2}(i)^{2+l_1-l_2-l}e^{i(\Delta_{\kappa_1}+\Delta_{\kappa_2})}|C_g(2)C_f(1)| \times\Big(\Big\{e^{i(\eta_1'+\eta_2')}(\gamma_2+iy_2)(\gamma_1+iy_1) \times\sum_{n=1}^{l+1}\Gamma\begin{bmatrix}l+n\\ n,2+l-n\end{bmatrix}(2k)^{-n}(i)^{n-l-2} \times\int_\rho^\infty dr\,r^{\gamma_1+\gamma_2-n}e^{-i(p_1+p_2-k)r} \times{}_1F_1(1+\gamma_2+iy_2;2\gamma_2+1;2ip_2r) \times{}_1F_1(1+\gamma_1+iy_1;2\gamma_1+1;2ip_1r) - \text{c.c.}\Big\} + \Big\{e^{i(\eta_1'-\eta_2')}(\gamma_2-iy_2)(\gamma_1+iy_1) \times\sum_{n=1}^{l+1}\Gamma\begin{bmatrix}l+n\\ n,2+l-n\end{bmatrix}(2k)^{-n}(i)^{n-l-2} \times\int_\rho^\infty dr\,r^{\gamma_1+\gamma_2-n}e^{-i(p_1+p_2-k)r} \times{}_1F_1(\gamma_2+iy_2;2\gamma_2+1;2ip_2r) \times{}_1F_1(1+\gamma_1+iy_1;2\gamma_1+1;2ip_1r) - \text{c.c.}\Big\}\Big), \tag{49}$$

where C_f and C_g are the normalization factors shown in Eq. (32) multiplied by the spinor function normalization factor in Eq. (36). The η_i' are defined by the equation

$$\eta_i' = \eta_i + \bar\delta_{\kappa_i}, \tag{50}$$

and the $\bar\delta_{\kappa_i}$ are the difference in phases between the extended and the point nuclei,

$$\bar\delta_{\kappa_i} = \Delta_{\kappa_i} - \delta^C_{\kappa_i}. \tag{51}$$

The presence of the exponential factors $e^{i\eta_i'}$ in Eq. (49) as compared to $e^{i\eta_i}$ for the point nucleus [Eq. (32)] guarantees the correct asymptotic form of the radial wavefunctions, as was exemplified in Eq. (34). It should be noted also that the appearance of the exponential $e^{-i(p_1+p_2-k)r}$ in the integrands of Eq. (49) is very crucial. The method of evaluating the integral of Eq. (1) developed in Sec. II can formally be applied only when the nuclear recoil momentum $q = p_1 - p_2 - k$ is totally neglected since, under such an approximation,

$$e^{-i(p_1+p_2-k)r} \approx e^{-i2p_2r}, \tag{52}$$

hence reducing the integrals to the form of Eq. (1). This is exactly what Gargaro and Onley[4] initially suggested to make the evaluation of such integrals possible. The discrepancy arising from the q-dependent terms in the integrands, however, can be removed with the aid of the addition theorem (see Appendix) through which we can fully take into account of the q dependence and yet are still able to make use of the results obtained in Sec. II. In the following we express Eq. (49) as

$$I_{\kappa_1\kappa_2 l} = S_{\kappa_1}S_{\kappa_2}(i)^{2+l_1-l_2-l}e^{i(\Delta_{\kappa_1}+\Delta_{\kappa_2})}|C_g(2)\,C_f(1)| \times[(M_1-\text{c.c.})+(M_2-\text{c.c.})], \tag{53}$$

where

$$M_1 = e^{i(\eta_1'+\eta_2')}(\gamma_1+iy_1)(\gamma_2+iy_2) \times\sum_{m=0}^\infty\frac{(1+\gamma_2+iy_2)_m}{(2\gamma_2+1)_m}\frac{(-iq)^m}{m!} \times\sum_{n=1}^{l+1}\Gamma\begin{bmatrix}l+n\\ n,l+2-n\end{bmatrix}(i)^{n-2-l}(2k)^{-n}$$

$$\times \mathscr{J}(\alpha;1+\gamma_2+iy_2+m,2\gamma_2+1+m,i(2p_2+q);$$
$$1+\gamma_1+iy_1,2\gamma_1+1;i2p_1), \tag{54}$$

$$M_2 = e^{i(\eta_1' - \eta_2')}(\gamma_1 + iy_1)(\gamma_2 - iy_2)$$
$$\times \sum_{m=0}^{\infty} \frac{(\gamma_2+iy_2)_m}{(2\gamma_2+1)_m} \frac{(-iq)^m}{m!}$$
$$\times \sum_{n=1}^{l+1} \Gamma\begin{bmatrix} l+n \\ n,l+2-n \end{bmatrix} (i)^{n-2-l}(2k)^{-n}$$
$$\times \mathscr{J}(\alpha;\gamma_2+iy_2+m,2\gamma_2+1+m,i(2p_2+q);$$
$$1+\gamma_1+iy_1,2\gamma_1+1;i2p_1), \tag{55}$$

with

$$\mathscr{J}(\alpha;a,b,K';\bar{a},\bar{b},K) = \int_\rho^\infty dr\, r^{\alpha-1} e^{-K'r}\, {}_1F_1(a,b;K'r)\, {}_1F_1(\bar{a},\bar{b};Kr) \tag{56}$$

and

$$\alpha = \gamma_1 + \gamma_2 - n + m + 1. \tag{57}$$

$(a)_m$ is the Pochhammer's symbol defined in Sec. II. Similarly we arrive at the expression for $J_{\kappa_1\kappa_2 l}$:

$$J_{\kappa_1\kappa_2 l} = S_{\kappa_1} S_{\kappa_2} (i)^{2+l_1-l_2-l} e^{i(\Delta_{\kappa_1}+\Delta_{\kappa_2})} |C_f(2)\, C_g(1)|$$
$$\times [(M_1 - \text{c.c.}) - (M_2 - \text{c.c.})]. \tag{58}$$

IV. REMARKS AND CONCLUSIONS

It may be seen that the finite size integrals of the Cases 1, 2, and 3 [Eqs. (2), (10), and (13)] reduce to those of the infinite integrals of Ref. 4 if the upper limit of the integrals approaches infinity. These formulae, however, involve certain difficulty when they are used to evaluate the cross section, mainly because of the unfortunate combinations of the parameters appearing in the generalized hypergeometric function ${}_3F_2$. The alternative Cases 1′ and 2′ [Eqs. (16) and (20)] are free of such difficulty and also require less computation time to evaluate the finite series of the ${}_3F_2$. Even though these formulae contain extra infinite summations as compared with the infinite integrals, the computation time is found to be reasonably manageable for our IBM 3031 computer.

The main reason why the above difficulty was encountered is because of the violation of a couple of conditions for the convergence of the integral representation formulae (A6)–(A8). When the finite size correction was introduced through the multiplication theorem formula (A12), it left the possibility of violating a condition for Eq. (A8) inevitable, while the use of formula (A13) eradicated the root of the trouble. A similar situation also exists for the recoil momentum correction by means of the addition theorem formulae (A19)–(A21). Use of either Eq. (A20) or (A21) causes violation of the conditions for (A6) and (A8) in Case 1′ and those for (A6) and (A7) in Case 2′, whereas the use of formula (A19) does not cause any violation at all.

As was mentioned in Sec. I, the experimental data are very scarce, and this makes it difficult to test the theoretical calculations. On the other hand, the theoretical plane wave intensity spectra with various corrections are readily available in Ref. 1. Among those calculations, the formulations after Schiff and Bethe–Heitler are commonly used, but there are considerable discrepancies between those spectra, particularly in the end point region of the photon energy. We also note here a remark made by Stoler *et al.*[19] that the end point region of the Bremsstrahlung spectrum is not experimentally well verified. Notably, Matthews and Owens[20] have stated that the spectrum obtained by using the Schiff formula was about 10% too high over most of the photon energy region than that from the Bethe–Heitler formula and also that it had incorrect shape in the end point photon energy region.

Since the shape of the spectrum obtained from our method was expected to be in gross agreement with those of Ref. 1, we started with the calculation for the incident electron energy of 90 MeV and the atomic number Z of the radiator nucleus being 78. This provided a first check of our computer program, and served as an indication as to whether our method would provide a means for detailed comparison between the theoretical calculations, particularly in the end point region of the photon energy.

Because of the computational complexity, we first calculated, in our initial phase of the computation, the cross section without including any of the corrections described in the preceding sections. Although the result was far from reality, we could have allowed the electron angular momentum variable in Eq. (36) to go up as high as to 20. When those corrections were introduced, the calculated intensity spectrum approached much closer to the ones found in Ref. 1. It became, however, inevitable to reduce the upper limits of the multiple summation variables in Eq. (47). Since we had observed during the trial runs that the calculated values had been practically the same when the electron angular momentum summation variables had gone beyond 8 and up to 20, we set, most of the time, the upper limit of the variable at around 8, whereas the range of the photon angular momentum was solely determined by Eq. (43) and the l summation was always extended within the full range of all possible values.

Ahrens *et al.*[21] measured the intensity spectrum with an end point energy of 140 MeV. To compare their result with our calculation, we expect, however, much higher partial waves are yet to be taken into account than those discussed above. It turned out, according to our realistic estimate, that the CPU time for the summation was proportional to the fourth power of the maximum value of the electron angular momentum variable and that a single run for the maximum value of 10 would require a continuous run of a whole day. Taking into account a further delay of the actual turn around time, we had to conclude that any further attempt to find a real criterion for the convergence of the series could be far beyond what one could afford. For such huge calculations, a direct integration of the integrals of Eqs. (44) and (45) is much more recommended. We are now in the process of converting the analytic calculation into the numerical integration. A practical limit for the summation associated with the recoil momentum correction is observed to be not more than a few couples of terms.

It should be pointed out here that the accuracy of the computed values is solely dependent on the software of the available computer. In other words, the accuracy is limited by that of the library subprograms of the individual computer, and, in particular, for this type of calculation a substantial inaccuracy[22] associated with the gamma function subprogram and its logarithm subprogram limits the meaningful summation of the terms in the series expansion formalism, even though we have overwritten some of these commercially available subprograms with our own more accurate algorithm.

In our current version of algorithm, most of the series are terminated when the magnitude of the last added term becomes less than a preset limit which is controlled by an input data set and the value of the limit is usually comparable to that of the truncation error due to a chosen length specification.

The nuclear radius parameter was tested for $\rho = 1.2A^{1/3}$ and for $\rho = 1.4A^{1/3}$ (in Fermi units), where A is the mass number of the radiator nucleus. Once a reasonable agreement was reached, the parameter was varied within a range around these values. It turned out that if the value of the parameter was away from these values by about several percent, the calculated intensity spectrum deviated from the experimental spectrum significantly. This indicates that the calculation is really sensitive to the choice of the radius parameter, but that the best choice of the parameter is consistent with the value determined from other experiments such as the electron scattering, μ-mesic atom, etc.

It is important to point out that the condition $\mathrm{Re}(\alpha) > 0$ appears to fail in a few cases. This happens when $l = l_1 + l_2'$ or $l_1' + l_2$ and the summation variable in Eq. (48) takes its highest value. This prohibits one from calculating the difference of the two improper integrals of the type of Eq. (1). One would have to overcome this difficulty by calculating directly the difference between the incomplete gamma functions appearing in Eq. (16) or Eq. (20) by setting the value of the argument of the function equal to the nuclear radius and the cutoff radius, say, in order to estimate the contribution arising from these cases.

One of the most interesting facts is that if one replaces the outgoing photon spherical wave in Eq. (38) with the photon standing wave, the extra ingoing component introduces an infrared divergent type spectrum, which destroys the whole calculation.

ACKNOWLEDGMENT

We would like to thank Professor Y. M. Shin for calling the authors' attention to this problem concerning the Bremsstrahlung intensity spectrum. Numerical computations were mainly carried out on the IBM 3031 computer at the University of Windsor.

APPENDIX

We give here all relevant formulae used in the main text. The confluent hypergeometric function, called Kummer's function, is defined as [23]

$$
\begin{aligned}
{}_1F_1(a;c;z) &= \sum_{n=0}^{\infty} \frac{(a)_n}{(c)_n} \frac{z^n}{n!}, \quad c \neq 0, -1, -2, \cdots, \\
&= \frac{\Gamma(c)}{\Gamma(a)} \sum_{n=0}^{\infty} \frac{\Gamma(a+n)}{\Gamma(c+n)} \frac{z^n}{n!}. \qquad \text{(A1)}
\end{aligned}
$$

This function satisfies the Kummer's first theorem

$$
{}_1F_1(a;c;z) = e^z\, {}_1F_1(c-a;c;-z). \qquad \text{(A2)}
$$

More generalized hypergeometric functions may be given in terms of integral representation[24]:

$$
\begin{aligned}
&\int_0^t x^{\alpha-1}(t-x)^{\beta-1}\, {}_pF_q\left[\begin{matrix} a_1,\cdots,a_p; \\ b_1,\cdots,b_q; \end{matrix} zx^k(t-x)^s\right] dx \\
&= \mathrm{B}(\alpha,\beta)\, t^{\alpha+\beta-1}\, {}_{p+k+s}F_{q+k+s}\left[\begin{matrix} a_1,\cdots,a_p; \dfrac{\alpha}{k}, \dfrac{\alpha+1}{k}, \cdots, \dfrac{\alpha+k-1}{k}, \dfrac{\beta}{s}, \dfrac{\beta+1}{s}, \cdots, \dfrac{\beta+s-1}{s}; \\ b_1,\cdots,b_q; \dfrac{\alpha+\beta}{k+s}, \dfrac{\alpha+\beta+1}{k+s}, \cdots, \dfrac{\alpha+\beta+k+s-1}{k+s}; \end{matrix} \; \frac{k^k s^s z t^{k+s}}{(k+s)^{k+s}}\right], \qquad \text{(A3)}
\end{aligned}
$$

where ${}_pF_q$ are generalized hypergeometric functions and $\mathrm{B}(\alpha,\beta)$ are Beta functions

$$
\mathrm{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \equiv \Gamma\left[\begin{matrix} \alpha,\beta \\ \alpha+\beta \end{matrix}\right]. \qquad \text{(A4)}
$$

Equation (A3) is valid if $\mathrm{Re}(\alpha) > 0$, $\mathrm{Re}(\beta) > 0$, k and s are nonnegative integers, but not both zero, and no b_i is to be zero or a negative integer. In particular, if $k = 1$, $s = 0$, and $t = 1$,

$$
\begin{aligned}
&{}_{p+1}F_{q+1}\left[\begin{matrix} a_1,\cdots,a_p;\alpha; \\ b_1,\cdots,b_q;\alpha+\beta; \end{matrix} z\right] \\
&= \Gamma\left[\begin{matrix} \alpha+\beta \\ \alpha,\beta \end{matrix}\right] \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, {}_pF_q\left[\begin{matrix} a_1,\cdots,a_p; \\ b_1,\cdots,b_q; \end{matrix} zx\right] dx, \qquad \text{(A5)}
\end{aligned}
$$

from which we have the following relations for $\mathrm{Re}(a_1) > 0$, $\mathrm{Re}(b_1) > 0$, and $|z| < 1$ for the last equation:

$$
{}_1F_1(a_1;a_1+b_1;z) = \Gamma\left[\begin{matrix} a_1+b_1 \\ a_1,b_1 \end{matrix}\right] \times \int_0^1 t^{a_1-1}(1-t)^{b_1-1} e^{zt}\, dt, \qquad \text{(A6)}
$$

$$
\begin{aligned}
{}_2F_1(a_1,a_2;a_1+b_1;z) &= \Gamma\left[\begin{matrix} a_1+b_1 \\ a_1,b_1 \end{matrix}\right] \times \int_0^1 t^{a_1-1}(1-t)^{b_1-1} \\
&\quad \times (1-tz)^{-a_2}\, dt, \qquad \text{(A7)}
\end{aligned}
$$

$$
{}_3F_2(a_1,a_2,a_3;a_1+b_1,b_2;z)
= \Gamma\begin{bmatrix} a_1+b_1 \\ a_1,b_1 \end{bmatrix} \int_0^1 t^{a_1-1}(1-t)^{b_1-1}
\times {}_2F_1(a_2,a_3;b_2;zt)\,dt. \quad \text{(A8)}
$$

A linear transformation formula for ${}_2F_1$ is[23]

$$
{}_2F_1(a,b;c;z) = \Gamma\begin{bmatrix} c,c-a-b \\ c-b,c-a \end{bmatrix} {}_2F_1(a,b;a+b-c+1;1-z)
+ (1-z)^{c-a-b}\Gamma\begin{bmatrix} c,a+b-c \\ a,b \end{bmatrix}
\times {}_2F_1(c-a,c-b;c-a-b+1;1-z),
$$

$$
|\arg(1-z)|<\pi,\ c-a-b\neq\pm m,\ m=0,1,2,3,\cdots. \quad \text{(A9)}
$$

A combination of (A8) and (A9) gives for $|z|=1$, provided that the series converge:

$$
{}_3F_2(a_1,a_2,a_3;b_1,b_2;1)
= \Gamma\begin{bmatrix} a_2-b_2+1,a_3-b_2+1 \\ a_2+a_3-b_2+1,1-b_2 \end{bmatrix}
\times {}_3F_2(b_1-a_1,a_2,a_3;b_1,a_2+a_3-b_2+1;1)
- \Gamma\begin{bmatrix} b_1,a_1-b_2+1,a_2-b_2+1,a_3-b_2+1,b_2-1 \\ a_1,b_1-b_2+1,a_2,a_3 \end{bmatrix}
{}_3F_2(a_1-b_2+1,a_2-b_2+1,a_3-b_2+1;
b_1-b_2+1,2-b_2;1). \quad \text{(A10)}
$$

Another linear transformation formula is useful for the analytic continuation of the function whose argument lies outside of the circle of convergence:

$$
{}_2F_1(a,b;c;z)
= \Gamma\begin{bmatrix} c,b-a \\ b,c-a \end{bmatrix}(-z)^{-a}\,{}_2F_1(a,a-c+1;a-b+1;1/z)
+ \Gamma\begin{bmatrix} c,a-b \\ a,c-b \end{bmatrix}(-z)^{-b}\,{}_2F_1(b,b-c+1;b-a+1;1/z),
$$

$$
|\arg(-z)|<\pi,\quad a-b\neq\pm m,\quad m=0,1,2,3,\cdots. \quad \text{(A11)}
$$

When the argument has a linear coefficient, the following multiplication theorem is most helpful:

$$
{}_1F_1(a;c;zz') = e^{z'(z-1)}z^{a-c}\sum_{n=0}^{\infty}\frac{(c-a)_n}{n!}
\times\left(1-\frac{1}{z}\right)^n {}_1F_1(a-n;c;z') \quad \text{(A12)}
$$

$$
= \sum_{n=0}^{\infty}\frac{(a)_n z'^n}{(c)_n n!}(z-1)^n {}_1F_1(a+n;c+n;z'). \quad \text{(A13)}
$$

Next we will show that the analytic continuation formula (A11) may be used even if the parameter a is a negative integer. Consider a function ${}_2F_1(-m,b;c;z)$, where m is a positive integer. Formal application of (A11) yields

$$
{}_2F_1(-m,b;c;z)
= \Gamma\begin{bmatrix} c,b+m \\ b,c+m \end{bmatrix}(-z)^m\,{}_2F_1(-m,1-c-m;1-b-m;1/z)
+ \Gamma\begin{bmatrix} c,-m-b \\ -m,c-b \end{bmatrix}(-z)^{-b}
\times {}_2F_1(b,b-c+1;b+m+1;1/z). \quad \text{(A14)}
$$

This function, however, is nothing but the Jacobi polynomial, which is defined in terms of the hypergeometric function ${}_2F_1(-n,n+\alpha+\beta+1;1+\alpha;z)$ as[23]

$$
P_n^{(\alpha,\beta)}(1-2z)
= \frac{(1+\alpha)_n}{n!}\,{}_2F_1(-n,n+\alpha+\beta+1;1+\alpha;z). \quad \text{(A15)}
$$

Therefore,[23]

$$
P_n^{(\alpha,\beta)}(x) = \binom{n+\alpha}{n}
\times {}_2F_1\left(-n,\alpha+\beta+n+1;\alpha+1;\frac{1-x}{2}\right)
= \binom{\alpha+\beta+2n}{n}\left(\frac{x-1}{2}\right)^n
\times {}_2F_1\left(-n,-\alpha-n;-2n-\alpha-\beta;\frac{2}{1-x}\right). \quad \text{(A16)}
$$

Setting $\alpha=c-1,\ \beta=b-c-n,\ 2/(1-x)=z^{-1}$,

$$
\binom{n+c-1}{n}\,{}_2F_1(-n,b;c;z)
= \binom{b+n-1}{n}(-z)^n\,{}_2F_1\left(-n,1-c-n;1-b-n;\frac{1}{z}\right);
$$

therefore,

$$
{}_2F_1(-n,b;c;z) = \Gamma\begin{bmatrix} c,b+n \\ b,c+n \end{bmatrix}
\times(-z)^n\,{}_2F_1\left(-n,1-c-n;1-b-n;\frac{1}{z}\right). \quad \text{(A17)}
$$

This shows that the second term of the formal equation (A14) should vanish for the Jacobi polynomial.

Finally we give a brief description of the addition and the multiplication theorems of Kummer's function. These are based on Taylor's theorem applied to Kummer's function, which is analytic in the whole domain of its complex variable.[25] If an analytic function is convergent for $|x|<\rho$,[26]

$$
f(x+y) = \sum_{n=0}^{\infty} f^{(n)}(x)\frac{y^n}{n!} \quad \text{for } |y|<\rho. \quad \text{(A18)}
$$

Here ρ is the radius of the circle of convergence which is infinity for Kummer's function. A few examples of the addition theorem are[25]

$$
{}_1F_1(a;c;z+z')
= \sum_{m=0}^{\infty}\frac{(a)_m z'^m}{(c)_m m!}\,{}_1F_1(a+m;c+m;z) \quad \text{(A19)}
$$

$$
= \left(\frac{z}{z+z'}\right)^{c-1}\sum_{m=0}^{\infty}\frac{(-1)^m(1-c)_m}{m!}\left(\frac{z'}{z}\right)^m {}_1F_1(a;c-m;z) \quad \text{(A20)}
$$

$$= \left(\frac{z}{z+z'}\right)^a \sum_{m=0}^{\infty} \frac{(a)_m}{m!} \times \left(\frac{z'}{z+z'}\right)^m {}_1F_1(a+m;c;z). \quad \text{(A21)}$$

By replacing y with $(y-1)x$, one gets from (A18)[25]

$${}_1F_1(xy) = \sum_{n=0}^{\infty} \frac{(y-1)^n x^n}{n!} \times \left(\frac{d}{dx}\right)^n {}_1F_1(x) \quad \text{for } |(y-1)x| < \infty. \quad \text{(A22)}$$

A few applications of this formula have been shown as (A12) and (A13).

[1]H. W. Koch and J. W. Motz, Rev. Mod. Phys. **31**, 920 (1959).
[2]H. Olsen, L. C. Maximon, and H. Wergeland, Phys. Rev. **106**, 27 (1957).
[3]H. Olsen and L. C. Maximon, Phys. Rev. **114**, 887 (1959).
[4]W. W. Gargaro and D. S. Onley, J. Math. Phys. **11**, 1191 (1970).
[5]H. A. Bethe and W. Heitler, Proc. R. Soc. London Ser. A **146**, 83 (1934).
[6]W. Heitler, *Quantum Theory of Radiation* (Oxford U. P., London, 1957).
[7]A. T. Nelms and I. Oppenheim, J. Res. Natl. Bur. Std. **55**, 53 (1955).
[8]J. Asai and H. Ogata, Nucl. Phys. A **233**, 55 (1974).
[9]See, e.g., D. F. Jackson, *Nuclear Reactions* (Methuen, London, 1970), p. 39.
[10]J. Asai, Ph.D. thesis, University of Windsor, 1974 (unpublished).
[11]A. Erdélyi, W. Magnus, F. Oberhettinger, and F. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 2.
[12]Reference 11, Vol. 1.
[13]M. E. Rose, *Relativistic Electron Theory* (Wiley, New York, 1961).
[14]D. R. Yennie, D. G. Ravenhall, and R. N. Wilson, Phys. Rev. **95**, 500 (1954).
[15]C. P. Bhalla and M. E. Rose, Oak Ridge National Laboratory Reports ORNL-2954 (1960), ORNL-3207 (1961).
[16]M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957).
[17]J. D. Rozics and W. R. Johnson, Phys. Rev. **135**, B56 (1964).
[18]M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables* (U. S. Dept. of Commerce, Natl. Bur. Std., Washington, D. C., 1968).
[19]P. Stoler, E. J. Winhold, F. O'Brien, P. F. Yergin, D. Rowley, K. Min, J. LeRose, A. M. Bernstein, K. I. Blomqvist, G. Franklin, N. Paras, and M. Pauli, Phys. Rev. C. **22**, 911 (1980).
[20]J. L. Matthews and R. O. Owens, Nucl. Instr. Meth. **111**, 157 (1973).
[21]J. Ahrens, H. Berchert, A. Zieger, and B. Ziegler, Nucl. Instr. Meth. **108**, 517 (1973).
[22]IBM System/360, FORTRAN IV Library Subprograms, File Number S360-25; Form C28-6596-3.
[23]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, New York, 1966).
[24]E. D. Rainville, *Special Functions* (MacMillan, New York, 1960).
[25]L. J. Slater, *Confluent Hypergeometric Functions* (Cambridge U. P., Cambridge, 1960).
[26]E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 4th ed. (Cambridge U. P., Cambridge, 1969).

On the uniqueness of the energy–momentum tensor for electromagnetism

D. B. Kerrighan
Department of Mathematics, Royal Roads Military College, Victoria, British Columbia, VOS 1B0, Canada

(Received 2 March 1982; accepted for publication 10 May 1982)

For the case of a space–time manifold we show that the metric tensor and the well-known energy–momentum tensor for electromagnetism are the only symmetric, rank 2 tensor concomitants of an arbitrary bivector, and the metric whose divergences vanish whenever the bivector satisfies the source-free Maxwell equations.

PACS numbers: 41.10. – j, 04.20. – q

1. INTRODUCTION

In the general theory of relativity the energy–momentum tensor for electromagnetism is taken to be[1]

$$T^{ij} = F^{ia}F^{j}{}_{a} - \tfrac{1}{4} g^{ij}F^{ab}F_{ab}, \tag{1.1}$$

where F_{ab} is a bivector which represents the electromagnetic field and g^{ij} is the inverse of the metric tensor g_{ij}. The question of the uniqueness of (1.1) naturally arises. On the one hand, Huggins[2] has proposed a possible modification to T^{ij} for the case of a Minkowski space–time. (A suitable generalization for curved space–times does not exist.[3]) On the other hand, the uniqueness of T^{ij} among various classes of tensors has been proved.[4] In particular, Lovelock[5] has shown the following:

Theorem: The only tensors, C^{ij}, which are symmetric, i.e.,

$$C^{ij} = C^{ji}, \tag{1.2}$$

and are tensor concomitants of g_{ab} and F_{ab}, i.e.,

$$C^{ij} = C^{ij}(g_{ab},F_{ab}), \tag{1.3}$$

and which satisfy the divergence condition[6]

$$C^{ij}{}_{|j} = \alpha^{i}{}_{h}\, F^{hj}{}_{|j} + \beta^{i}{}_{h}\, \eta^{hjab}F_{ab\,|j}, \tag{1.4}$$

where $\alpha^{i}h$ and $\beta^{i}h$ are tensor concomitants of g_{ab} and F_{ab}, and η^{hjab} is the Levi–Civita tensor, are of the form

$$C^{ij} = \alpha T^{ij} + \beta g^{ij}, \tag{1.5}$$

where α and β are constants.

The purpose of (1.4) is to give C^{ij} zero divergence whenever F^{ij} satisfies the source-free Maxwell equations. These equations may be written in the form

$$F^{ij}{}_{|j} = 0 \tag{1.6a}$$

and

$$\eta^{ijab}F_{ab\,|j} = F_{ab\,|j} + F_{ja|b} + F_{bj|a} = 0. \tag{1.6b}$$

However, assumption (1.4) leaves open the possibility that a C^{ij} exists which does not satisfy (1.4) or (1.5) and yet does satisfy (1.2) and (1.3) identically, and also

$$C^{ij}{}_{|j} = 0 \tag{1.7}$$

whenever Eqs. (1.6a) and (1.6b) hold. We shall show that this is not the case. Specifically, we shall prove the following theorem.

Theorem: Let C^{ij} satisfy (1.2) and (1.3). If, on a space–time manifold, C^{ij} satisfies (1.7) whenever (1.6a) and (1.6b) hold, then C^{ij} is given by (1.5). The proof relies on the explicit construction of all C^{ij} which satisfy (1.2). This has only recently been accomplished.[7]

2. PROOF OF THE THEOREM

For a space–time manifold any tensor which satisfies Eqs. (1.2) and (1.3) must be of the form[7]

$$C^{ij} = Ag^{ij} + BF^{ia}F^{j}{}_{a}, \tag{2.1}$$

where

$$A = A(g_{ab};F_{ab}) \tag{2.2a}$$

and

$$B = B(g_{ab};F_{ab}). \tag{2.2b}$$

The divergence of C^{ij} is

$$C^{ij}{}_{|j} = A_{|j}\, g^{ij} + B_{|j}\, F^{ia}F^{j}{}_{a} + BF^{ia}{}_{|j}\, F^{j}{}_{a} + BF^{ia}F^{j}{}_{a|j}. \tag{2.3}$$

We also have[8]

$$A_{|j} = \frac{\partial A}{\partial F_{ab}} F_{ab\,|j} \tag{2.4a}$$

and

$$B_{|j} = \frac{\partial B}{\partial F_{ab}} F_{ab\,|j}. \tag{2.4b}$$

Furthermore,[8] $\partial A/\partial F_{ab}$ and $\partial B/\partial F_{ab}$ are bivector concomitants of g_{ab} and F_{ab}. But all bivectors of this form are known[7]:

$$\frac{\partial A}{\partial F_{ab}} = A_1F^{ab} + A_2\eta^{abcd}F_{cd} \tag{2.5a}$$

and

$$\frac{\partial B}{\partial F_{ab}} = B_1F^{ab} + B_2\eta^{abcd}F_{cd}, \tag{2.5b}$$

where A_1, A_2, B_1, B_2 are scalar concomitants of g_{ab} and F_{ab}, and η^{abcd} is the Levi–Civita tensor. By using Eqs. (2.4) and (2.5) *and the Maxwell equations* [(1.6a) and (1.6b)] Eq. (2.3) can be written as

$$C^{ij}{}_{|j} = A^{ijab}F_{ab\,|j}, \tag{2.6}$$

where

$$A^{ijab} = \{(A_1 + \tfrac{1}{2}B)F^{ab} + A_2\eta^{abcd}F_{cd}\}\, g^{ij} + \{B_1F^{ab} + B_2\eta^{abcd}F_{cd}\}F^{ie}F^{j}{}_{e}. \tag{2.7}$$

Note that

$$A^{ijab} = A^{ijab}(g_{cd};F_{cd}). \tag{2.8}$$

The divergence requirement [i.e., (1.7) with (1.6a) and (1.6b)] can now be written as

$$A^{ijab}F_{ab|ij}=0, \tag{2.9a}$$

whenever

$$g^{jk}F_{ij|k}=0 \tag{2.9b}$$

and

$$F_{ij|k}+F_{ki|j}+F_{jk|i}=0. \tag{2.9c}$$

For an arbitrary bivector F_{ab}, there are 24 independent components of $F_{ab|j}$. A bivector which satisfies Eqs. (2.9b) and (2.9c) has only 16 independent components of $F_{ab|j}$ but these are more than sufficient to guarantee that

$$A^{ijab}=0. \tag{2.10}$$

(This assertion is verified in the Appendix.) Now, even if F_{ab} satisfied Eqs. (2.9b) and (2.9c), it still has the maximum number of independent components, namely, six. Therefore (2.10) holds for a bivector F_{ab} which is essentially arbitrary in the sense that F_{12}, F_{13}, F_{14}, F_{23}, F_{24}, and F_{34} may vary independently. One can easily establish that this situation obtains only if

$$A_1+\tfrac{1}{2}B=0 \tag{2.11a}$$

and

$$A_2=B_1=B_2=0. \tag{2.11b}$$

[By contracting with quantities of the form $F_{ab}\,g_{ij}$ one obtains a system of four equations in the unknowns $(A_1+\frac{1}{2}B)$, A_2, B_1, and B_2. This system has a nonzero determinant in general.]

Equations (2.5b) and (2.11b) imply

$$B=B(g_{ab}), \tag{2.12}$$

but the only scalar satisfying (2.12) is a constant,[9] i.e.,

$$B=\text{const}\equiv\beta. \tag{2.13}$$

Now Eq. (2.5a) becomes

$$\frac{\partial A}{\partial F_{ab}}=-\frac{1}{2}\beta F^{ab}.$$

One easily sees that

$$\frac{\partial(F^{cd}F_{cd})}{\partial F_{ab}}=2F^{ab};$$

therefore,

$$A=-\tfrac{1}{4}\beta F^{cd}F_{cd}+\alpha, \tag{2.14}$$

where α is a constant.[9]

Now Eqs. (2.1) and (2.14) together imply

$$C^{ij}=\alpha g^{ij}+\beta T^{ij}.$$

ACKNOWLEDGMENT

This work was supported by the Chief of Research and Development, Department of National Defence, Canada.

APPENDIX

We wish to establish that Eqs. (2.9) imply Eq. (2.10). First we choose any point p in our space–time manifold. Then we transform coordinates[10] so that at p,

$$g^{ij}=\eta^{ij}\equiv\text{diag}[1,1,1,-1].$$

Hence,

$$F_{ij|k}=F_{ij,k}$$

and

$$\Gamma^{i}_{\ ij}=0.$$

Then, from Eq. (2.9b) we have (at p),

$$\begin{aligned}F_{14,4}&=F_{12,2}+F_{13,3},\\ F_{24,4}&=F_{23,3}-F_{12,1},\\ F_{34,4}&=-F_{13,1}-F_{23,2},\\ F_{34,3}&=-F_{14,1}-F_{24,2}.\end{aligned} \tag{A1}$$

From Eq. (2.9c) we obtain

$$\begin{aligned}F_{12,3}&=F_{13,2}-F_{23,1},\\ F_{12,4}&=F_{14,2}-F_{24,1},\\ F_{13,4}&=F_{14,3}-F_{34,1},\\ F_{23,4}&=F_{24,3}-F_{34,2}.\end{aligned} \tag{A2}$$

Thus, eight of the 24 quantities $F_{ab,c}$ are dependent on the other 16. We can differentiate Eq. (2.9a) with respect to any one of these 16 independent quantities provided we use Eqs. (A1) and (A2). Among the 12 sets of equations we have (after suitable simplifications) the following eight:

$$\begin{aligned}A^{i213}&=-A^{i312}=-A^{i123},\\ A^{i212}&=-A^{i414}=A^{i313},\\ A^{i314}&=-A^{i413}=-A^{i134},\\ A^{i334}&=A^{i224}=A^{i224},\end{aligned} \tag{A3}$$

where i can take on any one of the values 1, 2, 3, or 4 in each of the equations.

By using Eqs. (A3) and (2.7) and the fact that

$$g^{ij}=\eta^{ij}$$

at p, we easily find that Eq. (2.10) holds at p. But p is arbitrary so (2.10) holds everywhere. This concludes the proof.

[1] Latin indices run from 1 to 4. The summation convention is used throughout. Indices are lowered and raised using the metric and its inverse.

[2] E. R. Huggins "Quantum Mechanics of the Interaction of Gravity with Electrons: Theory of a Spin-two Field Coupled to Gravity," Ph.D. thesis, California Institute of Technology (1962).

[3] D. Lovelock, Lett. Nuovo Cimento **14**, 581 (1974); B. Kerrighan, Gen. Relativ. Gravit. **13**, 283 (1981).

[4] D. Lovelock, Proc. R. Soc. London Ser. A **341**, 285 (1974); D. Lovelock, J. Math. Phys. **19**, 586 (1978).

[5] D. Lovelock, J. Math. Phys. **18**, 1491 (1977).

[6] The vertical bar denotes covariant differentiation.

[7] B. Kerrighan, Gen. Relativ. Gravit. **13**, 19 (1981).

[8] B. Kerrighan, "The Uniqueness of the Einstein–Maxwell Equations," PhD thesis, University of Waterloo, Canada (1978).

[9] D. Lovelock, Arch. Rational Mech. Anal. **33**, 54 (1969).

[10] These are the so-called normal coordinates at P; η^{ij} is the Minkowski metric; $\Gamma^{i}_{\ jk}$ is the Christoffel symbol; partial differentiation is denoted by a comma.

Quantum tachyon in a Friedmann universe

S. K. Srivastava
Government College, Port Blair 744104, India

(Received 31 December 1980; accepted for publication 30 July 1981)

The scalar wave equation of a primordial tachyon is investigated in a Friedmann universe of positive curvature containing different perfect fluids. Also the energy of a tachyon as well as the rate of emission of energy from a primordial tachyon in different models has been discussed.

PACS numbers: 98.80. − k

1. INTRODUCTION

It is now generally accepted that the existence of tachyons does not violate the theory of relativity, though so far experiments to detect them have yielded null results. Recently Narlikar and Sudarshan[1] have published a paper in which they have assumed that tachyons were produced at or just after the event of the big bang along with other particles of ordinary matter or bradyons. This gives rise to a question whether primordial tachyons survive up to the present epoch and if they survive, why experiments fail to detect them. Narlikar and Sudarshan have tackled this problem to some extent in the context of a flat Friedmann universe.

In the present paper, we have tackled this problem in Friedmann models of positive curvature containing different perfect fluids such as dust, radiation, and superdense matter. We have considered the Robertson–Walker line element

$$ds^2 = dt^2 - S^2(t)\left[\frac{dr^2}{1-r^2} + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2)\right]. \tag{1.1}$$

Under coordinate transformations

$$\tau = \int \frac{dt}{S(t)} \text{ and } \sigma = \int \frac{dr}{(1-r^2)^{1/2}} = \sin^{-1} r, \tag{1.2}$$

this line element is rewritten as

$$ds^2 = \Omega^2(\tau)[d\tau^2 - d\sigma^2 - \sin^2\sigma(d\theta^2 + \sin^2\theta\, d\varphi^2)], \tag{1.3}$$

where

$$\Omega(\tau) = S(t). \tag{1.4}$$

Here, for simplicity, we shall discuss the case of spinless tachyons with the help of relativistic quantum mechanics. The tachyon scalar wavefunction satisfies the Klein–Gordon equation

$$(\Box^2 - M^2)\psi = 0, \tag{1.5}$$

where we have chosen $c = 1, \hbar = 1$ ($\hbar$ is Planck's constant divided by 2π), $\psi(r,\theta,\phi,t)$ is the scalar wavefunction of a tachyon, and M is the metamass[2] of a tachyon which is given by

$$m_0 = iM, \tag{1.6}$$

(m_0 is the rest mass of a tachyon). The operator $\Box^2$ is given by

$$\Box^2 = \frac{1}{(-g)^{1/2}} \frac{\partial}{\partial x^2}\left((-g)^{1/2} g^{ij} \frac{\partial}{\partial x^j}\right). \tag{1.7}$$

The plane wave solutions of (1.5) are of the form $\exp(i\bar{k}\cdot\bar{r} - i\nu t)$, with $\bar{k}^2 - \nu^2 = m^2$ where $M = 2\pi m$ and $\bar{k}$ and ν are wavenumber and frequency, respectively. Therefore the group velocity $d\nu/d\bar{k} = \bar{k}/\nu$ is greater than unity. The curvature of space–time is incorporated in the $\Box^2$ operator.

2. SOLUTION OF THE KLEIN–GORDON EQUATION

Under coordinate transformations (1.2) the Klein–Gordon equation can be written as

$$(\Box^2 - M^2)\bar{\psi} = 0, \tag{2.1}$$

where

$$\bar{\psi} = \Omega(\tau)\psi. \tag{2.2}$$

On substituting g^{ij}'s from (1.4) in (2.1) we have

$$\frac{\partial^2\bar{\psi}}{\partial\tau^2} + \frac{2}{\Omega}\frac{\partial\Omega}{\partial\tau}\frac{\partial\bar{\psi}}{\partial\tau} - \frac{\partial^2\bar{\psi}}{\partial\sigma^2} - 2\cot\sigma\frac{\partial\bar{\psi}}{\partial\sigma} - \frac{1}{\sin^2\sigma}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\bar{\psi}}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2\bar{\psi}}{\partial\varphi^2}\right] - M^2\Omega^2(\tau)\bar{\psi} = 0. \tag{2.3}$$

$\bar{\psi}$ can be expanded in terms of a complete set of eigenfunctions of the angular momentum operator. Hence, setting

$$\bar{\psi} = \sum_{l=0}^{\infty} \bar{\psi}_l p_l(\cos\theta) \tag{2.4}$$

we have a partial differential equation of $\bar{\psi}_l(\sigma,\tau)$

$$\frac{\partial^2\bar{\psi}_l}{\partial\tau^2} + \frac{2}{\Omega}\frac{\partial\Omega}{\partial\tau}\frac{\partial\bar{\psi}_l}{\partial\tau} - \frac{\partial^2\bar{\psi}_l}{\partial\sigma^2} - 2\cot\sigma\frac{\partial\bar{\psi}_l}{\partial\sigma} + \left[\frac{l(l+1)}{\sin^2\sigma} - M^2\Omega^2(\tau)\right]\bar{\psi}_l = 0. \tag{2.5}$$

Also we have

$$(1/\Omega)(\partial\Omega/\partial\tau) = H\Omega = HS, \tag{2.6}$$

where H is Hubble's constant. After the epoch of the big bang, we assume that at time $t = t_q$ (say) the entire energy of the universe comes into thermal equilibrium. Now approximating HS near $t = t_q$ we have

$$HS = H_q S_q + (t - t_q)\left[H_q\left(\frac{\partial S}{\partial t}\right)_{t_q} + \left(\frac{\partial H}{\partial t}\right)_{t_q} S_q\right].$$

But S_q may be approximated by zero, because up to the event $t = t_q$ expansion would have been very small; hence

$$HS \approx 0. \tag{2.7}$$

Moreover, the velocity of a tachyon is

$$v = \frac{S(t)}{(1-r^2)^{1/2}}\frac{dr}{dt} = S(t)\frac{d\sigma}{dt} > 1, \tag{2.8}$$

hence

$$\sigma > \int \frac{dt}{S(t)}. \tag{2.9}$$

This shows that σ will increase rapidly. Hence there is no harm is taking

$$\cot\sigma = 0 \quad \text{and} \quad \sin\sigma = 1. \tag{2.10}$$

Substituting (2.7) and (2.10) in (2.5) we get

$$\frac{\partial^2 \overline{\psi}_l}{\partial \tau^2} - \frac{\partial^2 \overline{\psi}_l}{\partial \sigma^2} + [l(l+1) - M^2 \Omega^2(\tau)]\overline{\psi}_l = 0. \tag{2.11}$$

The plane wave solution of this equation can be written as

$$\overline{\psi}_l = \phi(\sigma)\exp(-iv\tau), \tag{2.12}$$

where v is a constant.

On substituting (2.12) in (2.11) we get

$$\frac{d^2\phi}{d\sigma^2} + [v^2 + M^2\Omega^2(\tau) - l(l+1)]\phi = 0. \tag{2.13}$$

This equation yields the solution

$$\phi = A\exp[\pm i\sigma(v^2 + M^2\Omega^2(\tau) - l(l+1))^{1/2}]. \tag{2.14}$$

Hence

$$\overline{\psi}_l = A\exp[-iv\tau \pm i\sigma(v^2 + M^2\Omega^2(\tau) - l(l+1))^{1/2}]. \tag{2.15}$$

This gives the solution of the Klein–Gordon equation (2.1) corresponding to the orbital angular momentum l.

3. SOLUTIONS OF THE KLEIN–GORDON EQUATION IN WORLD MODELS CONTAINING DIFFERENT PERFECT FLUIDS

A. Dust model of the Friedmann universe

If we consider the dust model of the universe we find that the Einstein field equations yield[3]

$$\Omega(\tau) = S(t) \approx t^{2/3}. \tag{3.1}$$

Connecting Eq. (3.1) with Eqs. (1.2) and (2.8) we have

$$\tau = \int \frac{dt}{t^{2/3}} = 3t^{1/3}. \tag{3.2}$$

We are considering free tachyons. Hence their velocity v will be constant. This yields

$$\sigma = 3vt^{1/3}. \tag{3.3}$$

Substituting Ω, τ, and σ from Eqs. (3.1), (3.2), and (3.3) in Eq. (2.15) we have

$$\overline{\psi}_l = A\exp[-3ivt^{1/3} \pm 3ivt^{1/3}(v^2 + M^2\Omega^2(\tau) - l(l+1))^{1/2}]. \tag{3.4}$$

Hence on substituting $\Omega(\tau)$ from Eq. (3.1) in (2.2) we have

$$\psi_l = At^{-2/3}\exp[-3ivt^{1/3} \pm 3ivt^{1/3}(v^2 + M^2\Omega^2(\tau) - l(l+r))^{1/2}]. \tag{3.5}$$

This shows damping of the scalar tachyon wave with time. The magnitude of energy associated with this wave is given by

$$E = \left| \frac{d}{dt}[3vt^{1/3} \pm 3vt^{1/3}(v^2 + M^2t^{4/3} - l(l+1))^{1/2}] \right|$$

$$= t^{-2/3}\left[v + v(v^2 + M^2t^{4/3} - l(l+1))^{1/2} + \frac{2vM^2t^{4/3}}{(v^2 + M^2t^{4/3} - l(l+1))^{1/2}}\right], \tag{3.6}$$

where v and l are constants. Hence there is no harm in taking

$$v^2 = l(l+1). \tag{3.7}$$

Now

$$E = t^{-2/3}[l(l+1)^{1/2} + 3Mvt^{2/3}]. \tag{3.8}$$

This expression of energy shows that the energy of the tachyon is decreasing with time. This means that the tachyon is emitting energy and the emitted energy is being absorbed in the surroundings. We have the rate of emission of energy as

$$-\frac{dE}{dt} = \frac{2}{3}\frac{(l(l+1))^{1/2}}{t^{5/3}}. \tag{3.9}$$

B. Radiation model of the Friedmann universe

From the Einstein field equations,[3] we find that

$$\Omega(\tau) = S(t) \approx t^{1/2}. \tag{3.10}$$

Equations (3.10) and (1.2) yield

$$\tau = \int \frac{dt}{t^{1/2}} = 2t^{1/2}. \tag{3.11}$$

Also we have

$$\sigma = 2vt^{1/2}. \tag{3.12}$$

Substituting τ and σ from Eqs. (3.11) and (3.12) in Eq. (2.15) we have

$$\overline{\psi}_l = A\exp[-2ivt^{1/2} \pm 2ivt^{1/2}(v^2 + M^2t - l(l+1))^{1/2}]. \tag{3.13}$$

Hence substituting $\Omega(\tau)$ from Eq. (3.10) in (2.2) we have

$$\psi_l = At^{-1/2}\exp[-2ivt^{1/2} \pm 2ivt^{1/2}(v^2 + M^2t - l(l+1))^{1/2}]. \tag{3.14}$$

This equation again shows damping of scalar tachyon waves with time. The magnitude of energy associated with this wave is given by

$$E = \left| \frac{d}{dt}[2vt^{1/2} \pm 2vt^{1/2}(v^2 + M^2t - l(l+1))^{1/2}] \right|$$

$$= t^{-1/2}\left[v + v(v^2 + M^2t - l(l+1))^{1/2} + \frac{2M^2vt}{(v^2 + M^2t - l(l+1))^{1/2}}\right]. \tag{3.15}$$

Connecting Eqs. (3.7) and (3.15) we have

$$E = t^{-1/2}(v + 3vMt^{1/2}). \tag{3.16}$$

This equation also shows that energy of the tachyon is decreasing with time. The rate of emission of energy in this case is given by

$$-\frac{dE}{dt} = \frac{1}{2}\frac{(l(l+1))^{1/2}}{t^{3/2}}. \tag{3.17}$$

C. Superdense model of the Friedmann universe

When the Friedmann model is in a superdense state, from the Einstein field equation[3] we have

$$\Omega(\tau) = S(t) = t^{1/3}. \tag{3.18}$$

Connecting Eqs. (3.18) and (1.2) we have

$$\tau = \int \frac{dt}{t^{1/3}} = \frac{3}{2} t^{2/3}. \tag{3.19}$$

Also

$$\sigma = \tfrac{3}{2} v t^{2/3}. \tag{3.20}$$

Substituting τ and σ from Eqs. (3.19) and (3.20) we have

$$\overline{\psi}_l = A \exp\left[-\tfrac{3}{2} i v t^{2/3} \pm \tfrac{3}{2} i v t^{2/3} (v^2 + M^2 t^{2/3} - l(l+1))^{1/2}\right]. \tag{3.21}$$

Now substituting $\Omega(\tau)$ from Eq. (3.18) in Eq. (2.2) we have

$$\psi_l = A t^{-1/3} \exp\left[-\tfrac{3}{2} i v t^{2/3} \pm \tfrac{3}{2} i v t^{2/3} (v^2 + M^2 t^{2/3} - l(l+1))^{1/2}\right]. \tag{3.22}$$

This equation also shows damping of scalar tachyon waves with time. The magnitude of energy associated with this wave is given by

$$E = \left| \frac{d}{dt}\left[\frac{3}{2} v t^{2/3} \pm \frac{3}{2} v t^{2/3} (v^2 + M^2 t^{2/3} - l(l+1))^{1/2} \right] \right|$$

$$= t^{-1/3}\left[v + v(v^2 + M^2 t^{2/3} - l(l+1))^{1/2} + \frac{1}{2} \frac{M^2 v t^{2/3}}{(v^2 + M^2 t^{2/3} - l(l+1))^{1/2}} \right]. \tag{3.23}$$

Connecting Eqs. (3.7) and (3.23) we have

$$E = t^{-1/3}\left[v + \tfrac{3}{2} v M t^{1/3} \right]. \tag{3.24}$$

This equation also shows that the energy of the tachyon is decreasing with time. The rate of emission of energy in this case is given by

$$-\frac{dE}{dt} = \frac{1}{2} \frac{(l(l+1))^{1/2}}{t^{4/3}}. \tag{3.25}$$

4. DISCUSSION

From Eqs. (3.5), (3.14), and (3.22) we note that the scalar wave of a tachyon is damped with time. But in the case of the dust model damping is fastest, and in the case of the superdense model damping is slowest. We have the order of damping as

$$D_d > D_r > D_s,$$

where D stands for damping and the subscripts d, r, and s denote dust, radiation, and superdense models, respectively.

Also, the graph sketched in Fig. 1 for rate of emission of energy against time shows that, at a particular time, the rate of emission of energy is fastest in the superdense model and slowest in the radiation model. But as time increases the graphical curves in the cases of the radiation and dust models come closer and closer faster than in the case of the superdense model.

However, in all cases, we find that the primordial tachyon loses its energy with time. Hence from the above investigations we are in a position to say that there would be a very narrow possibility of survival of a primordial tachyon up to the present epoch. Even if they survive, owing to very low energy, their experimental detection seems difficult.

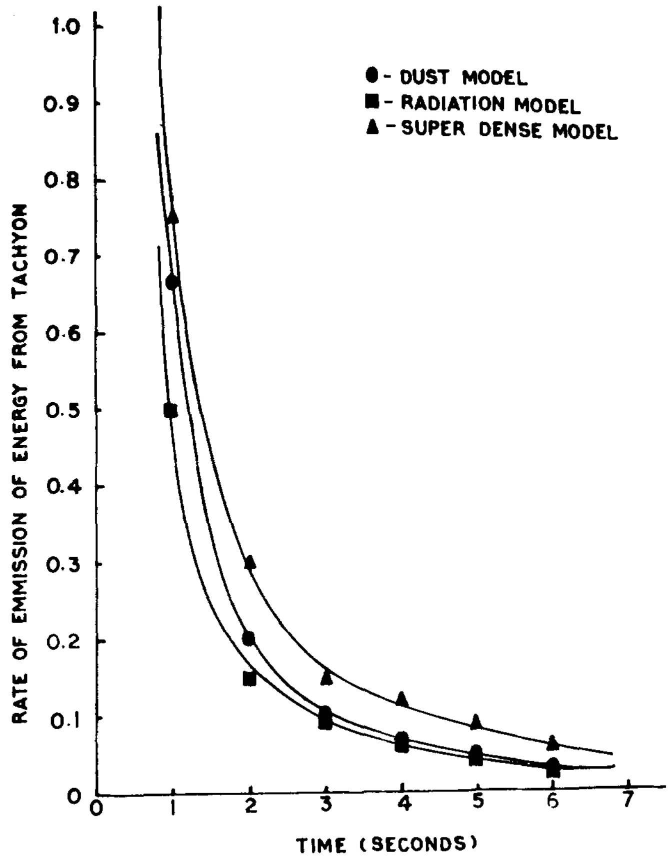


FIG. 1. Rate of emission of energy from a tachyon against time.

ACKNOWLEDGMENT

I am extremely grateful to Dr. P. K. Koranne for stimulating discussions and helpful suggestions.

[1]J. V. Narlikar and E. C. G. Sudarshan, Mon. Not. R. Qstr. Soc. **175**, 105 (1976).

[2]O. M. P. Bilanuik, V. K. Deshpande, and E. C. G. Sudarshan, Am. J. Phys. **30**, 718 (1962); Phys. Today **22**, 43 (1969).

[3]M. P. Pathak, Ph.D. thesis, Gorakhpur University, Gorakhpur, 1975.

Initial data and wave propagation in one-dimensional inhomogeneous cosmologies[a)]

Richard A. Matzner[b)]
Department of Astrophysics, South Parks Road, Oxford OX1 3RQ, England
M. Rosenbaum and M. P. Ryan, Jr.
Centro de Estudios Nucleares, Universidad Nacional Autonoma de Mexico, Mexico, D. F.

(Received 28 January 1982; accepted for publication 9 April 1982)

We consider the initial value problem and the short-time evolution of a system consisting of two Euclidean-homogeneous (Bianchi type I) cosmologies, each containing a homogeneous scalar field φ or a homogeneous electric field, abutted at the plane $z = 0$, We show that a matching is possible, and discuss the evolution for a vacuum ($\varphi = 0$) cosmology abutted to one with nonzero scalar field φ.

PACS numbers: 98.80. – k, 04.20. – q

I. INTRODUCTION

Centrella[1,2] has developed a technique for explicitly solving the initial data problem of one-dimensionally inhomogeneous cosmologies. Her technique consisted of abutting three-dimensional slices of two different anisotropic type I cosmologies:

$$ds^2 = -dt^2 + \left(\frac{t}{t_0}\right)^{2p_1} dx^2 + \left(\frac{t}{t_0}\right)^{2p_2} dy^2 + \left(\frac{t}{t_0}\right)^{2p_3} dz^2, \tag{1.1}$$

at a 2-plane (taken to be $z = 0$). We label the constants t_0 and p_i in the two different regions by A and B, i.e., t_{0A}, p_{1B}, etc. The Hamiltonian constraint[3] contains no derivatives for vacuum or for perfect fluid sources [since (1.1) can be made explicitly flat at any particular time $\hat{t}$], hence the abutted half-slices automatically satisfy the Hamiltonian constraint. The momentum constraint[3], $\pi^{ij}{}_{,j} = 0$, reduces in the vacuum or fluid case to

$$(1 - p_3)/t_0 = \text{const.} \tag{1.2}$$

Centrella's matching procedure is carried out by fixing t_{0A}, p_{3A}, and t_{0B} (say) at the initial time and then using the constraint (1.2) to determine p_{3B}. In vacuum we need not specify p_1 and p_2 because the p_i obey

$$p_1 + p_2 + p_3 = 1 \tag{1.3}$$

and

$$p_1^2 + p_2^2 + p_3^2 = 1, \tag{1.4}$$

but these equations are modified if matter is present. Thus for a vacuum, the matching as described completely determines the models on the two sides of the 2-surface $z = 0$. When matter is present, there are still other parameters needed to describe the solution on each side: schematically, the deviation of the sum of the p_i and/or of the sum of the squares of the p_i from unity.

Once initial data have been set, the Einstein equations can be solved either by numerical computation, or in the vacuum case—and the scalar field case—by analytic techniques. The models that result are examples (perhaps nonvacuum) of Gowdy[4] inhomogeneous cosmologies. The discontinuous data just described are of substantial interest from the viewpoint of the evolution of shocks in the early universe. For short times away from the "initial" data slice an approach based on a Green's function is the most reasonable, because the short-time Green function is especially simple. For longer-time evolution Fourier–Bessel decomposition[4–6] of the dynamical variable [cf. Eq. (6.2) below], although complicated, does allow analytical evolution of this data for large times. Of course the numerical approach allows solution of a wider class of problems than can be treated analytically.

In this paper we consider an extension of the Centrella initial-value technique to the case of inhomogeneous plane-symmetric cosmologies whose only matter content is a massless scalar field or an electromagnetic field sharing the plane symmetry, and we use the analytical techniques just outlined to describe the evolution of data corresponding to an instantaneously homogeneous universe containing an inhomogeneous (discontinuous) plane-symmetric scalar field.

II. SCALAR-FIELD INITIAL DATA

The massless scalar field obeys

$$\phi^{;\mu}{}_{;\mu} = 0. \tag{2.1}$$

The stress tensor associated with this field is

$$T_{\mu\nu} = \phi_{,\mu}\phi_{,\nu} - \tfrac{1}{2} g_{\mu\nu}\phi_{,\alpha}\phi^{,\alpha}. \tag{2.2}$$

Spatially homogeneous solutions of the scalar-field Einstein equations

$$G_{\mu\nu} = T_{\mu\nu} \tag{2.3}$$

may be written in the metric form (1.1), with a scalar field[7]

$$\varphi = r \ln(t/t_0) + \varphi_0(r, \varphi_0 \text{ const}) \tag{2.4}$$

and with constant exponents p_i satisfying (1.3) but with (1.4) replaced by

$$p_1^2 + p_2^2 + p_3^2 = 1 - r^2. \tag{2.5}$$

In carrying out the Centrella procedure, one first uses

[a)] Supported in part by NSF grants PHY 77-07619, PHY 81-07381, and INT 78-22583 and by Grant No. 955 from CONACyT de Mexico, one from the SERC, and one from the National Geographic Society.

[b)] Permanent address: Center for Relativity, University of Texas, Austin, Texas 78712.

coordinate freedom to adjust t_0 on the two sides of the join so that $t_A = t_{0A}$, $t_B = t_{0B}$, and the 3-metric on each side of the join is thus explicitly δ_{ij}. One might anticipate difficulty because the stress tensor involves gradients of φ. *A priori* this contradicts the assumption we want to make in solving the Hamiltonian constraint, that no spatial gradients enter. However, Eq. (2.4) shows that if $\varphi_0 = 0$ then φ vanishes on the slice $t = t_0$ and we henceforth make that choice. Then the argument, that the Hamiltonian constraint is a pointwise algebraic identity, holds on the initial slice, and by picking half slices that are scalar wave cosmologies with $\varphi_0 = 0$ on the t_0 slice, the Hamiltonian constraint is identically satisfied everywhere.

Additionally, the initial data require the satisfaction of the momentum constraint

$$T^{i0} = -(K^{im}{}_{|m} - K^{m}{}_{m}{}^{|i}) = g^{-1/2}\pi^{im}{}_{|m}, \tag{2.6}$$

where K^{im} is the extrinsic curvature of the surfaces being matched, g is the 3-space metric determinant, and π^{ij} is the ADM (Arnowitt–Deser–Misner) momentum. The symbol $|m$ denotes the spatial covariant derivative, in the m direction, which is here equal to the ordinary derivative since the 3-surfaces are explicitly flat. Further, from the form of the stress tensor (2.2), the energy current vanishes in each half-slice. We thus suppose $T^{i0}|_{z=0} = 0$, and the match condition becomes

$$\frac{1 - p_3(z)}{t_0(z)} = \text{const}, \tag{2.7}$$

as in the vacuum case.

Valid inhomogeneous data on a 3-slice then consist of the choices (which may be made smoothly, i.e., not necessarily discontinuously)

$$t_0(z), \tag{2.8a}$$

$$p_3(z), \tag{2.8b}$$

$$r(z), \tag{2.8c}$$

with Eqs. (2.5) and (2.7). Because of the presence of the scalar field, there is greater freedom to choose the variables than in the vacuum case. The number of variables is increased by one with the presence of φ, but the number of constraints remains unity, cf. Eq. (2.7).

III. ELECTROMAGNETIC INITIAL DATA

The scalar-field initial data set is a very simple generalization of the vacuum one, in part because there is no constraint (i.e., no initial-value) problem for the scalar field. However, for the electromagnetic situation there is a constraint equation, namely[3]

$$\mathscr{E}^i{}_{,i} = 0, \tag{3.1}$$

where

$$\mathscr{E}^i = \tfrac{1}{4}[ijk\,][jk\mu\nu](-g)^{1/2}g^{\mu\alpha}g^{\nu\beta}F_{\alpha\beta} \tag{3.2}$$

with $[ijk\,]$ and $[\alpha\beta\gamma\delta\,]$ the alternating symbols defined by

$$[123] = [0123] = 1.$$

Jacobs[8] has given some electric field-containing homogeneous cosmologies of Bianchi type I. These models have the electric field along one of the principal axes of the metric.

His solution for a field along the 1-direction is

$$ds^2 = -dt^2 + e^{2\alpha}e^{2\beta}{}_{ij}dx^i dx^j, \tag{3.3}$$

with α, β functions of time [$\beta_{ij}(t)$ is a diagonal traceless matrix] satisfying

$$3\dot\alpha^2 = \tfrac{1}{2}\dot\beta_{ij}\,\dot\beta_{ij} + \beta e^{-2\alpha}e^{-2\beta}{}_{22}e^{-2\beta}{}_{33}, \tag{3.4}$$

$$(\dot\beta_{22} - \dot\beta_{33})e^{3\alpha} = \text{const}, \tag{3.5}$$

$$\ddot\beta_{11} + 3\ddot\alpha\,\beta_{11} = \tfrac{2}{3}\beta e^{-2\alpha}e^{-2\beta}{}_{22}e^{-2\beta}{}_{33}, \tag{3.6}$$

$$\ddot\beta_{22} + 3\dot\alpha\,\dot\beta_{22} = \tfrac{1}{3}\beta e^{-2\alpha}e^{-2\beta}{}_{22}e^{-2\beta}{}_{33}. \tag{3.7}$$

The parametrization in terms of the p_i, $i = 1,2,3$, is not especially useful now because they are not time independent. The metric can, however, be made explicitly flat at any particular time.

Matching conditions for abutted homogeneous cosmologies, so far as the electromagnetic field is concerned, require from (3.1) only that

$$\mathscr{E}^3 = \text{const} \tag{3.8}$$

or, since as before, the metric may be set instantaneously equal to δ_{ij}:

$$F_{03} = \text{const.} \tag{3.9}$$

Bianchi type I cosmologies demand $T_{0i} = 0$. This means that the gravitational matching conditions also are easily carried out; since $T^{0i} = 0$ in each homogeneous cosmology, we have[1,3] only

$$\pi^{33}{}_{,3} = 0; \tag{3.10}$$

i.e., again, $\pi^{33} = \text{const}$.

In the metric form of Eq. (3.3) this is

$$(2\dot\alpha - \dot\beta_{zz})_{,z} = 0. \tag{3.11}$$

Furthermore, since the electromagnetic stress tensor contains no derivatives of the field, the Hamiltonian constraint is, as before, pointwise satisfied.

Hence, consistent data may be set by abutting slices from solutions of the Jacobs class that

(a) have a nonvanishing 3-component of the electric field [in which case the electric field is in fact homogeneous via (3.8) and the homogeneity is contained entirely in the geometrical variables]—or

(b) have a nonvanishing, and z-variable, electric field always lying along the x axis (or always lying along the y axis), truly inhomogeneous data—or

(c) have abutted half a 3-slice with a nonzero x-directed electric field against half a 3-slice with a nonzero y-directed electric field—step discontinuous data which cannot be continuously connected to homogeneous solutions within the class of Jacobs solutions.

Subsequent evolution of the electrovac field equations in general must be done numerically, since the electromagnetic field equations do not decouple in the simple way that the scalar ones do (see Sec. IV).

Work in progress by Waller[9] concerns homogeneous type I cosmologies in which the electric field is not constrained to lie along a principal-axis direction. In that case, data which combine features of (a) and (b) with a constant z

field but variable x or y field, will be feasible. Also, data in which a field rotates from the x to the y direction over a finite range in z will be possible, smoothing out the jump in case (c) above.

IV. CANONICAL COORDINATE SYSTEMS

The "canonical" Gowdy representation of a plane-symmetric cosmology is a modification of the Rosen[10] form

$$ds^2 = e^{\gamma - \psi}(-dT^2 + dZ^2) + (e^{\psi}dX^2 + T^2 e^{-\psi}dY^2), \tag{4.1}$$

where γ and ψ are functions of T and Z. For the remainder of this paper we will concentrate on the scalar-field case. We do point out, however, that the form (4.1) which has $g_{xx}g_{yy} = T^2$ can hold only if[11] $T^0_0 - T^Z_Z = 0$, which means $E_Z = B_Z = 0$. Hence data of type (a) in the preceding section are excluded by this form of the metric; said otherwise, Sec. III describes the statement of data that lead to more general plane-symmetric cosmologies than the usually considered canonical ones.

Given (4.1), $\phi^{;\mu}{}_{;\mu} = 0$ reduces to

$$\ddot{\varphi} + \dot{\varphi}/T - \varphi'' = 0 \tag{4.2}$$

(here $\cdot \equiv \partial/\partial T$ and $' \equiv \partial/\partial Z$), while the Einstein equation $G_{\mu\nu} = T_{\mu\nu}$ becomes

$$\ddot{\psi} + \dot{\psi}/T - \psi'' = 0, \tag{4.3}$$

$$\dot{\gamma} = T[\dot{\varphi}^2 + \varphi'^2 + \tfrac{1}{2}(\dot{\psi}^2 + \psi'^2)], \tag{4.4}$$

$$\gamma' = T[2\dot{\varphi}\varphi' + \dot{\psi}\psi']. \tag{4.5}$$

If we rewrite the homogeneous solutions of Sec. II we have[7]

$$\psi = (1 + \alpha_0)\ln(T/T_0), \tag{4.6}$$

$$\varphi = \beta_0 \ln(T/T_0), \tag{4.7}$$

$$e^{\gamma - \psi} = (T/T_0)^{(\beta_0 + \alpha_0^2/2 - 1/2)}. \tag{4.8}$$

For the homogeneous cosmologies, the relation between the coordinates of Eq. (1.1) and the canonical coordinates of (4.1) is fixed by

$$e^{\gamma - \psi} = (t/t_0)^{2p_3}; \tag{4.9}$$

hence

$$T = \frac{t}{1 - p_3}\left(\frac{t_0}{t}\right)^{p_3} \tag{4.10}$$

and

$$T_0 \equiv T(t_0(z)) = \frac{t_0}{1 - p_3} = \text{const}, \tag{4.11}$$

where the constancy holds by the matching conditions of Sec. II.

The complete correspondence with the p_i form of Eq. (1.1) becomes

$$p_1 = \frac{1 + \alpha_0}{2 + (\beta_0^2 + \frac{1}{2}\alpha_0^2 - \frac{1}{2})}, \tag{4.12}$$

$$p_2 = \frac{1 - \alpha_0}{2 + (\beta_0^2 + \frac{1}{2}\alpha_0^2 - \frac{1}{2})}, \tag{4.13}$$

$$p_3 = \frac{\beta_0^2 + \frac{1}{2}\alpha_0^2 - \frac{1}{2}}{2 + (\beta_0^2 + \frac{1}{2}\alpha_0^2 - \frac{1}{2})}, \tag{4.14}$$

and

$$r = \frac{2\beta_0}{2 + (\beta_0^2 + \frac{1}{2}\alpha_0^2 - \frac{1}{2})}. \tag{4.15}$$

The entire content of the matching procedure of Sec. II is contained in the relationship between Eqs. (2.7) and (4.11). The "constants" α_0 and β_0 may be varied arbitrarily as functions of Z so long as they are set on a 3-surface $T = T_0 = \text{const}$. Notice that the correspondence between the two formulations requires

$$\gamma(T = T_0) = \psi(T = T_0) = \varphi(T = T_0) = 0.$$

In order to see why in general the electromagnetic evolution equations must be integrated numerically, we will give them in a form due to Charach. For the metric (4.1), and the condition $E_Z = B_Z = 0$, the only two nonvanishing components of the electromagnetic potential are $A_X \equiv \chi$, $A_Y \equiv \omega$, and the field equations are

$$\dot{\omega}\dot{\chi} = \omega'\chi', \tag{4.16}$$

$$\ddot{\psi} + (1/T)\dot{\psi} - \psi'' = (e^{2\psi}/T^2)(\dot{\chi}^2 - \chi'^2) - e^{-2\psi}(\dot{\omega}^2 - \omega'^2), \tag{4.17}$$

$$\ddot{\omega} + (1/T)\dot{\omega} - \omega'' = 2(\dot{\omega}\dot{\psi} - \omega'\psi'), \tag{4.18}$$

$$\ddot{\chi} - (1/T)\dot{\chi} - \chi'' = -2(\dot{\chi}\dot{\psi} - \chi'\psi'), \tag{4.19}$$

$$\dot{\gamma} = T(\dot{\psi}^2 + \psi'^2) + Te^{-2\psi}(\dot{\omega}^2 + \omega'^2) + (1/T)e^{2\psi}(\dot{\chi}^2 + \chi'^2), \tag{4.20}$$

$$\gamma' = 2T\dot{\psi}\psi' + 2Te^{-2\psi}\dot{\omega}\omega' + (2/T)e^{2\psi}\dot{\chi}\chi'. \tag{4.21}$$

As before, Eqs. (4.16)–(4.19) determine the solution, and (4.20) and (4.21) can be solved by quadratures. It is easy to see that only in special cases will the equations decouple sufficiently to yield easy analytical solutions.

V. THE GREEN-FUNCTION FORMULATION

The dynamical equations (4.2) and (4.3) admit a Green function. For instance, if a general field point is labeled (Z, T) then, since $\psi = 0$ at T_0, we have

$$\psi(\tilde{Z}, \tilde{T}) = \int_{D_3}^{D_4} \tfrac{1}{2} R\dot{\psi}\, dZ, \tag{5.1}$$

where R is the Green function[12]

$$R = (T_0/\tilde{T})^{1/2} F(\tfrac{1}{2}, \tfrac{1}{2}; 1; -q), \tag{5.2}$$

where F is a hypergeometric function,

$$q = -\frac{(Z - D_4)(Z - D_3)}{4T_0\tilde{T}} > 0, \tag{5.3}$$

and the integration is over the portion of the $T = T_0$ surface within the intersection of the planes $T \pm Z = \tilde{T} \pm \tilde{Z}$ with the $T = T_0$ surface (see Fig. 1). From Eq. (4.6)

$$\dot{\psi}|_{T_0} = (1 + \alpha_0)/T_0 \tag{5.4}$$

and from Eq. (4.7)

$$\dot{\varphi}|_{T_0} = \beta_0/T_0. \tag{5.5}$$

The free variables φ, ψ determine the longitudinal part γ of the metric, but otherwise do not interact.

The situation with $\varphi \equiv 0$ (i.e., $\dot{\varphi} = 0$ everywhere on the initial slice) corresponds to the vacuum Gowdy[4] model, and has been extensively studied by Centrella and Matzner.[13] We will first consider the complementary case in which the data

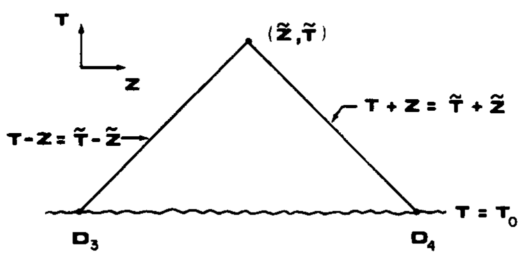


FIG. 1. The field at the point $(\tilde{Z}, \tilde{T})$ is evaluated by the Green-function integral (5.1) over the region of the initial data surface $T = T_0$ between the intersections D_3 and D_4 with the null planes passing through $(\tilde{Z}, \tilde{T})$.

for $\dot{\psi}$ is homogeneous, but that for $\dot{\varphi}$ has a step at $Z = 0$ being denoted φ_A for $Z < 0$, and φ_B fore $Z > 0$. [Note that $\dot{\psi} \equiv 0$ corresponds to a particular axisymmetric model, and there is a slight generalization by taking a homogeneous $\dot{\psi}|_{T_0} = (1 + \alpha_0)/T_0$, corresponding via (4.6) to a general anisotropic cosmology. The behavior of ψ then persists as $(1 + \alpha_0)\ln(T/T_0)$.]

The behavior of the Green function has been elaborated by Centrella and Matzner.[14] For small time $\Delta T = T - T_0$, $R \sim 1$. For larger ratios $|\Delta T/T_0|$ the contribution to the integral involves the more complicated behavior of the Green function. Let us first concentrate on a small interval centered on T_0, i.e., $|\Delta T/T_0| \ll 1$. We assume $T_0 > 0$ always. We take the appropriate expansion from Ref. 14 and find $R = 1$ in the integration interval. Then

$$\varphi = \Delta T\left\{\frac{\beta_{0B}}{T_0}\theta(\tilde{Z} - \Delta T) + \frac{\beta_{0A}}{T_0}\theta(-(\tilde{Z} + \Delta T)) + \frac{\beta_{0A}}{2T_0}\left(1 - \frac{\tilde{Z}}{\Delta T}\right) + \frac{\beta_{0B}}{2T_0}\left(1 + \frac{\tilde{Z}}{\Delta T}\right) \times \theta(\tilde{Z} + \Delta T)\theta(\Delta T - \tilde{Z})\right\} \quad (1 \gg \Delta T/T_0 > 0) \tag{5.6}$$

and

$$\varphi = \Delta T\left\{\frac{\beta_{0B}}{T_0}(\tilde{Z} + \Delta T) + \frac{\beta_{0A}}{T_0}(\Delta T - \tilde{Z}) + \frac{\beta_{0A}}{2T_0}\left(1 + \frac{\tilde{Z}}{\Delta T}\right) + \frac{\beta_{0B}}{2T_0}\left(1 - \frac{\tilde{Z}}{\Delta T}\right) \times \theta(Z - \Delta T)\theta(-(Z + \Delta T))\right\} \quad (-1 \ll \Delta T/T_0 < 0). \tag{5.7}$$

The stress tensor associated with the φ field is probably the most physically appropriate object to consider. From Eq. (2.2) we have

$$T_{TT} = T_{ZZ} = \tfrac{1}{2}(\dot{\varphi} + \varphi'^2) = \frac{1}{2}\left[\left(\frac{\beta_{0B}}{T_0}\right)^2\theta(\tilde{Z} - \Delta T) + \left(\frac{\beta_{0A}}{T_0}\right)^2\theta(-(\tilde{Z} + \Delta T)) + \left\{\left(\frac{\beta_{0B}}{2T_0}\right)^2 + \left(\frac{\beta_{0A}}{2T_0}\right)^2\right\}\theta(\tilde{Z} + \Delta T)\theta(\Delta T - \tilde{Z})\right], \tag{5.8}$$

$$T_{TZ} = \dot{\varphi}\varphi' = \left[\left(\frac{\beta_{0B}}{2T_0}\right)^2 - \left(\frac{\beta_{0A}}{2T_0}\right)^2\right]\theta(\tilde{Z} + \Delta T)\theta(\Delta T - \tilde{Z}), \tag{5.9}$$

$$T^X_X = T^Y_Y = -\tfrac{1}{2}g^{00}(\dot{\varphi}^2 - \varphi'^2) = \frac{1}{2}\left[\left(\frac{\beta_{0B}}{T_0}\right)^2\theta(\tilde{Z} - \Delta T) + \left(\frac{\beta_{0A}}{T_0}\right)^2\theta(-(\tilde{Z} + \Delta T)) + \frac{\beta_{0A}\beta_{0B}}{T_0^2}\theta(\tilde{Z} + \Delta T)\theta(\Delta T - \tilde{Z})\right]. \tag{5.10}$$

These expressions are written for $T_0 \gg \Delta T > 0$; similar equations hold for negative ΔT. Outside the causal past and future of the $T = T_0$ discontinuity, the expressions for the scalar field and for the stress energy take on their homogeneous-cosmology form.

As a specific example, suppose we take $\beta_{0B} = 0$. Then the solution represents at the "initial" time T_0 a region containing the homogeneous scalar field β_{0A} up to the boundary, with its subsequent "expansion" into the vacuum region.

It is interesting to note that the value of T_{TZ} is homogeneous in the "interaction" region (with the signature used here, our T_{TZ} indicates that the flux is directed toward the vacuum region B). In fact all the orthonormal frame components of the stress tensor are also homogeneous, to the order we are considering, in this interaction region, suffering jumps at the null surfaces $\Delta T \pm \tilde{Z} = 0$ to their homogeneous-cosmology values. It may be verified that for $\Delta T < 0$, T_{TZ} has a similar form but the flux is in the opposite direction. The particular initial data set at $T = T_0$ is thus achieved by the expected process of having a flux toward region A, up to the initial instant T_0 whereupon the flux reverses and the scalar field re-expands into the vacuum region.

VI. SEPARABLE SOLUTIONS

As was pointed out in Ref. 13, Eqs. (4.2) and (4.3) are easily soluble by separation of variables. For example, (4.2) has solutions of the form

$$\varphi \propto e^{ikZ}\mathscr{Z}_0(kT), \tag{6.1}$$

where $\mathscr{Z}_0$ is a zero-order Bessel function. The major problem with this approach is that our data set demands $\varphi(T = T_0) = 0$. This implies that φ must have the form

$$\varphi = \int_0^\infty dk\,[A(k)\sin kZ + B(k)\cos kZ] \times [J_0(kT_0)N_0(kT) - N_0(kT_0)J_0(kT)] + C\ln(T/T_0). \tag{6.2}$$

An integral such as this, a product of three Bessel functions [since $\sin(x) \sim J_{1/2}(x)$] times a weighting function such as $A(k)$, is at the very limit of known analytic integrals for all but the simplest $A(k)$. There are, however, a number of devices for reducing expressions such as (6.2) to a more tractable form. As an example we will consider the initial conditions of Sec. V with $\varphi(T_0) = 0$, $\dot{\varphi}|_{T_0} = \beta_0/T_0$ with a step at $Z = 0$. The exact solution becomes[13]

$$\varphi = \tfrac{1}{2}\bar{\beta}_0 \ln(T/T_0) + \tfrac{1}{2}(\beta_{0A} - \beta_{0B}) \int_0^\infty \frac{\sin kZ}{k} \times [N_0(kT_0)J_0(kT) - J_0(kT_0)N_0(kT)]\, dk, \quad (6.3)$$

where $\bar{\beta}_0 \equiv \beta_{0A} + \beta_{0B}$. First note that we need only consider $Z > 0$, since the solution for $Z < 0$ can be obtained by symmetry. For simplicity we will take $T > T_0$. The opposite case follows analogously.

In the particular case of (6.3) we can use the identity $N_0(x) = (2/\pi)\lim_{\lambda\to 0} \partial J_\lambda(x)/\partial\lambda$, the fact that $\sin kZ = (\pi kZ/2)^{1/2} J_{1/2}(kZ)$, and the integral (Gradshteyn and Rhyzik[15])

$$\int_0^\infty x^{\rho-1} J_\gamma(ax) J_\mu(bx) J_\nu(cx)\, dx = \frac{2^{\rho-1} a^\gamma b^\mu c^{-\gamma-\mu-\rho} \Gamma(\frac{1}{2}(\gamma+\mu+\nu+\rho))}{\Gamma(\gamma+1)\Gamma(\mu+1)\Gamma(1-\frac{1}{2}(\gamma+\mu-\nu+\rho))} \times F_4\left(\frac{\gamma+\mu-\nu+\rho}{2}, \frac{\gamma+\mu+\nu+\rho}{2}; \gamma+1, \mu+1; \frac{a^2}{c^2}, \frac{b^2}{c^2}\right), \quad (6.4)$$

where F_4 is one of Appell's hypergeometric functions of two variables,[15,16] to find

$$\varphi(Z < T) = \tfrac{1}{2}\bar{\beta}_0 \ln(T/T_0) + \tfrac{1}{2}(\beta_{0A} - \beta_{0B}) \frac{Z}{T}\frac{2}{\pi} \times \left[\ln(T/4T_0) F_4\left(\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{3}{2}, 1; \frac{Z^2}{T^2}, \frac{T_0^2}{T^2}\right) + \tilde{F}_4\left(\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{3}{2}, 1; \frac{Z^2}{T^2}, \frac{T_0^2}{T^2}\right)\right], \quad (6.5)$$

$$\varphi(Z > T) = \tfrac{1}{2}\bar{\beta}_0 \ln(T/T_0) + \tfrac{1}{2}(\beta_{0A} - \beta_{0B}) \left[-\ln(T/T_0) + \lim_{\lambda\to 0} \frac{\partial}{\partial\lambda} F_4\left(\frac{\lambda}{2}, \frac{1+\lambda}{2}; 1, 1+\lambda; \frac{T^2}{Z^2}, \frac{T_0^2}{Z^2}\right) - \lim_{\lambda\to 0} \frac{\partial}{\partial\lambda} F_4\left(\frac{\lambda}{2}, \frac{1+\lambda}{2}; 1+\lambda, 1; \frac{T^2}{Z^2}, \frac{T_0^2}{Z^2}\right)\right], \quad (6.6)$$

where

$$\tilde{F}_4\left(\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{3}{2}, 1; \frac{Z^2}{T^2}, \frac{T_0^2}{T^2}\right) \equiv \lim_{\lambda\to 0} \frac{\partial}{\partial\lambda}\left[F_4\left(\frac{1+\lambda}{2}, \frac{1+\lambda}{2}; \tfrac{3}{2}, 1+\lambda; \frac{Z^2}{T^2}, \frac{T_0^2}{T^2}\right) - F_4\left(\frac{1-\lambda}{2}, \frac{1+\lambda}{2}; \tfrac{3}{2}, 1; \frac{Z^2}{T^2}, \frac{T_0^2}{T^2}\right)\right].$$

The difficulties with this solution lie in the definition of F_4. This function is defined as a convergent series,

$$F_4(\alpha, \beta; \gamma, \gamma'; x, y) = \sum_{m=0} \sum_{n=0} \frac{(\alpha)_{m+n}(\beta)_{m+n}}{(\gamma)_m (\gamma')_n m! n!} x^m y^n, \quad (6.7)$$

only for $|\sqrt{x}| + |\sqrt{y}| < 1$, that is, for $Z < T - T_0$. Note, however, that for $Z > T - T_0$ the Green function discussion of Sec. V means that for $Z > T - T_0$, (6.6) reduces to

$$\varphi(Z > T - T_0) = \beta_{0B} \ln(T/T_0),$$

because in that case the point Z, T is causally connected only to the B-homogeneous region of the solution. Hence (6.7) provides the complete solution of the problem. [In Eq. (6.7), the notation $(\gamma)_m$ is Pochhammer's symbol[15]

$$(\gamma)_m = \Gamma(a+m)/\Gamma(a),$$

where

$$\Gamma(n) = (n-1)!].$$

Because explicit behavior of (6.7) is difficult to extract from the power series, we now discuss another method that can be used for handling integrals of the type of (6.3): the method of convolutions. This method may be more fruitful in cases that are less simple than (6.3), as it involves integrating a weighting function against trigonometric functions and one Bessel function, and there is ample literature on this type of integral. In our case we make use of[17]

$$\mathscr{F}_S^{-1}(F(k)G(k)) = \frac{1}{(2\pi)^{1/2}} \int_0^\infty g(\xi)[\tilde{f}(Z-\xi) - \tilde{f}(Z+\xi)]\, d\xi, \quad (6.8)$$

where

$$\mathscr{F}_S^{-1}(F(k)) \equiv \frac{\sqrt{2}}{\pi} \int_0^\infty F(k) \sin kZ\, dk \quad (6.9)$$

and

$$g(\xi, a) \equiv \frac{\sqrt{2}}{\pi} \int_0^\infty G(k, a) \sin k\xi\, dk, \quad (6.10)$$

$$\tilde{f}(\xi, a) \equiv \frac{\sqrt{2}}{\pi} \int_0^\infty F(k, a) \cos k\xi\, dk. \quad (6.11)$$

For (6.3) we have

$$F(k, a) = N_0(ka), \quad G(k, a) = J_0(ka)/k, \quad (6.12)$$

so (6.3) becomes

$$\varphi = \tfrac{1}{2}\bar{\beta}_0 \ln(T/T_0) + \tfrac{1}{2}(\beta_{0A} - \beta_{0B}) \times \int_0^\infty \{g(\xi, T)[\tilde{f}(Z-\xi, T_0) - \tilde{f}(Z+\xi, T_0)] - g(\xi, T_0)[\tilde{f}(Z-\xi, T) - \tilde{f}(Z+\xi, T)]\}\, d\xi. \quad (6.13)$$

Since

$$\tilde{f}(\xi, a) = \sqrt{\frac{2}{\pi}} \begin{cases} 0, & 0 < \xi < a \\ \dfrac{1}{(\xi^2 - a^2)^{1/2}}, & 0 < a < \xi \end{cases}, \quad (6.14)$$

$$g(\xi, a) = \sqrt{\frac{2}{\pi}} \begin{cases} \pi/2, & \xi > a \\ \arcsin(\xi/a), & \xi \leqslant a \end{cases}, \quad (6.15)$$

we find that there are six regions of Z (we again have $T > T_0$, and again we take $Z > 0$ and can find $Z < 0$ by symmetry), each with a different form of φ. They are R1: $0 < Z < T_0$, $Z < T - T_0$; R2: $Z < T_0 < T$, $Z > T - T_0$; R3: $T_0 < Z < T$, $Z < T - T_0$; R4: $T_0 < Z < T$, $Z > T - T_0$; R5: $T_0 < T < Z$, $Z < T + T_0$; R6: $T_0 < T < Z$, $Z > T + T_0$. The functions involved in the integral (6.13) turn out to be relatively simple in each of these regions, and are explicitly written out for all six regions in the Appendix. However, it is easy to verify that only regions R1 and R3 lie within the region causally con-

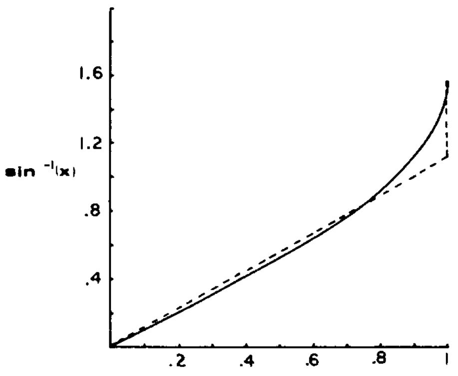


FIG. 2. The solid line is arcsin(x), while the dashed line is $(\pi - 2)x$.

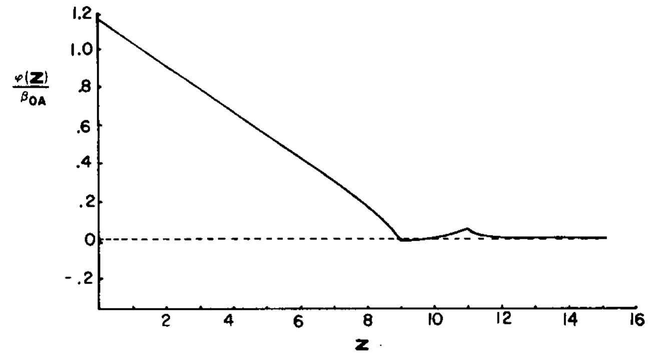


FIG. 3. The graph of the function $\varphi(Z, T)$ via Eq. (6.17), which uses the linear approximation to arcsin(x). Here we have taken $T_0 = 1$, $T = 10$, $\beta_{0B} = 0$, and β_{0A} arbitrary.

nected to the inhomogeneous data. Hence they are the only ones we need to consider. As a paradigm we will take region R1, and we find

$$\varphi_{R1} = \tfrac{1}{2}\bar{\beta}_0 \ln(T/T_0) + \tfrac{1}{4}(\beta_{0A} - \beta_{0B}) \times\Bigg[-\ln\left(\frac{[(Z+T)^2 - T_0^2]^{1/2} + T + Z}{[(T-Z)^2 - T_0^2]^{1/2} + T - Z}\right) + \frac{2}{\pi}\int_{T_0 - Z}^{T} \arcsin\left(\frac{\xi}{T}\right)\frac{d\xi}{[(Z+\xi)^2 - T_0^2]^{1/2}} - \frac{2}{\pi} \times \int_{Z+T_0}^{T} \arcsin\left(\frac{\xi}{T}\right)\frac{d\xi}{[(Z-\xi)^2 - T_0^2]^{1/2}}\Bigg]. \tag{6.16}$$

Unfortunately, the last two integrals above do not seem to have a representation in terms of known functions. However, since they are over a finite range, they could be done numerically or approximated by some analytic scheme. For example, there is a polynomial approximation to arcsin(x) given in Ref. 18 which might be useful. For the purposes of this paper we will present only the crudest possible approximation, $\arcsin(x) \approx (\pi - 2)x$ (see Fig. 2), where the constant has been chosen to give the same area under the two curves between zero and one. We calculate φ_{R1}; to this order of approximation (at least) the only other relevant form of φ, φ_{R3}, has exactly the same form. Hence, for the region $T > T_0$, $Z > 0$, of the spacetime causally connected to the inhomogeneous data,

$$\varphi_{R1} \approx \tfrac{1}{2}\bar{\beta}_0 \ln(T/T_0) + \tfrac{1}{4}(\beta_{0A} - \beta_{0B}) \times \Bigg\{\left(\frac{2\pi - 4}{\pi}\right)\frac{Z}{T}[\ln(T_0/2T)] - \tfrac{1}{2}\ln\left(\left[\tfrac{1}{2}[(1 - Z/T)^2 - T_0^2/T^2]^{1/2} + \frac{1}{2} - \frac{Z}{2T}\right]\times\left[\tfrac{1}{2}[(1+Z/T)^2 - T_0^2/T^2]^{1/2} + \frac{1}{2} + \frac{Z}{2T}\right]\right) - \tfrac{1}{2}\ln\left[\frac{[(1+Z/T)^2 - T_0^2/T^2]^{1/2} + 1 + Z/T}{[(1-Z/T)^2 - T_0^2/T^2]^{1/2} + 1 - Z/T}\right] + \frac{1}{2}\left(\frac{2\pi - 4}{\pi}\right)\left([(1+Z/T)^2 - T_0^2/T^2]^{1/2}[(1-Z/T)^2 - T_0^2/T^2]^{1/2}\right)\Bigg\}$$

$$(T > T_0;\ 0 < Z < T - T_0). \tag{6.17}$$

The leading linear term here is $(Z/T)(2\pi - 4/\pi)[\ln(T_0/T) - \ln(2) + 1] - 1$, while the most obvious leading term in (6.5) is $(Z/T)(2/\pi)[\ln(T_0/T) - \ln(4)]$, which differ by about 20%, but there are other terms in (6.5) proportional to sum of $(T_0/T)^{2n}$ that would have to be evaluated before the two leading terms could be directly compared.

In Fig. 3 we present a graph of φ for $T = 10$, $T_0 = 1$, and $\beta_{0B} = 0$, where we have included the approximation for the acausal region. Note that even in our crude approximation the curve reflects the features of the numerical integration reported in Fig. 3 of Ref. 13.

The main deviation from accuracy occurs at and beyond the causal boundary, where $\varphi \to 0$. While the scale of the graph is too small to show it there is, for example, infinite slope at $Z = 9$, presumably due to our approximation of arcsin(x). The numerical results of Ref. 13 do not show this effect, and the Green-function analysis of Ref. 13 shows that φ should be linear as it approaches zero. Although this error is small in the potential φ, it leads to substantial errors in the stress tensor [based on derivatives of φ; cf. Eqs. (5.8)–(5.10)] and more accurate approximations are being sought.

VIII. ON THE RELATION TO GOWDY MODELS

If $\varphi = 0$, these models correspond to the Gowdy[4] model, except that we have not imposed any closure on the model. Had we done so, the usual restriction would arise: Only certain periodic functions of Z are permitted, and the net Z momentum due to gravitational waves (and other waves, if the model is in fact nonvacuum) must vanish, in order for the nonlinear Eqs. (4.4) and (4.5) defining γ to admit a periodic solution.[7] Our models require no such restriction so, for instance, there is a net momentum flux in the example of Sec. IV. This poses absolutely no problem since, as we now show, our initial data are easily fitted into a periodic-in-Z data set.

Consider any data which for $Z<Z_A$ are homogeneous-cosmological data (cosmology A), for some interval $Z_A<Z<Z_B$ is inhomogeneous, and become homogeneous again (a new cosmology, B) for $Z>Z_B$. Our example of Sec. IV is a special case of such data. Now suppose $Z_C \gg Z_B$. We can begin another inhomogeneous region, in the interval $Z_D>Z>Z_C$ such that for $Z>Z_D$ the data are again those for the homogeneous cosmology A. Then any plane $Z<Z_A$ can be chosen to be identified with a plane $Z>Z_D$, and the result is periodic data, which *a priori* will yield periodic solutions for all the metric variables.

The periodicity can be chosen so that each homogeneous region extends over much more than a horizon size at the time T_0. Hence the different regions, with their apparent net momentum, can evolve freely (as our Sec. IV model does) for a long time until eventually the horizons overlap and it becomes locally more apparent that the total momentum of the solution vanishes. This also shows that it is dangerous to apply arguments based on linearization stability (the demand for periodicity of γ is the full-field version of a linearization stability argument) to categorize models for the real universe. Such arguments are global, but there is hopefully most of the universe we have not yet seen, and it is thus invalid to invoke global arguments to restrict our local geometry. Linearization stability does not, for instance, prevent a local flux of gravitational radiation in a closed universe; it only demands that the net momentum be somehow balanced elsewhere.

ACKNOWLEDGMENT

Richard Matzner thanks Professor C. W. Misner for stimulating discussions which prompted considerations of some of the points in this paper.

APPENDIX

In this Appendix we present the convolution solution to (6.3) in the regions of Z mentioned in the text. In the approximation $\arcsin(x) \approx (\pi-2)x$ the expressions reduce to ones similar to that given for φ_{R1} in Sec. VI. The expression for φ_{R1} is given in Sec. VI. The others are

$$\varphi_{R2} = \tfrac{1}{2}\bar{\beta}_0 \ln(T/T_0) + \tfrac{1}{4}(\beta_{0A}-\beta_{0B})\bigg[-\ln\bigg(\frac{[(Z+T)^2-T_0^2]^{1/2}+T+Z}{[(Z+T_0)^2-T^2]^{1/2}+T_0+Z}\bigg) - \ln(T/T_0)$$
$$+\frac{2}{\pi}\int_{T_0-Z}^{T} \arcsin\Big(\frac{\xi}{T}\Big)\frac{d\xi}{[(Z+\xi)^2-T_0^2]^{1/2}} - \frac{2}{\pi}\int_{T-Z}^{T_0} \arcsin\Big(\frac{\xi}{T_0}\Big)\frac{d\xi}{[(Z+\xi)^2-T^2]^{1/2}}\bigg],$$

$$\varphi_{R3} = \tfrac{1}{2}\bar{\beta}_0 \ln(T/T_0) + \tfrac{1}{4}(\beta_{0A}-\beta_{0B})\bigg[-\ln\bigg(\frac{[(Z+T)^2-T_0^2]^{1/2}+T+Z}{[(T-Z)^2-T_0^2]^{1/2}+T-Z}\bigg)$$
$$-\frac{2}{\pi}\int_{0}^{Z-T_0} \arcsin\Big(\frac{\xi}{T}\Big)\frac{d\xi}{[(Z-\xi)^2-T_0^2]^{1/2}} - \frac{2}{\pi}\int_{Z+T_0}^{T} \arcsin\Big(\frac{\xi}{T}\Big)\frac{d\xi}{[(Z-\xi)^2-T_0^2]^{1/2}}$$
$$+\frac{2}{\pi}\int_{0}^{T} \arcsin\Big(\frac{\xi}{T}\Big)\frac{d\xi}{[(Z+\xi)^2-T_0^2]^{1/2}}\bigg],$$

$$\varphi_{R4} = \tfrac{1}{2}\bar{\beta}_0 \ln(T/T_0) + \tfrac{1}{4}(\beta_{0A}-\beta_{0B})\bigg[-\ln(T/T_0) - \ln\bigg(\frac{[(T+Z)^2-T_0^2]^{1/2}+T+Z}{[(T_0+Z)^2-T^2]^{1/2}+T_0+Z}\bigg)$$
$$+\frac{2}{\pi}\int_{0}^{T} \arcsin\Big(\frac{\xi}{T}\Big)\frac{d\xi}{[(Z+\xi)^2-T_0^2]^{1/2}} - \frac{2}{\pi}\int_{0}^{Z-T_0} \arcsin\Big(\frac{\xi}{T}\Big)\frac{d\xi}{[(Z-\xi)^2-T_0^2]^{1/2}}$$
$$-\frac{2}{\pi}\int_{T-Z}^{T_0} \arcsin\Big(\frac{\xi}{T_0}\Big)\frac{d\xi}{[(Z+\xi)^2-T^2]^{1/2}}\bigg],$$

$$\varphi_{R5} = \tfrac{1}{2}\bar{\beta}_0 \ln(T/T_0) + \tfrac{1}{4}(\beta_{0A}-\beta_{0B})\bigg[-\ln(T/T_0) - \ln\bigg(\frac{[(Z+T)^2-T_0^2]^{1/2}+Z+T}{[(Z+T_0)^2-T^2]^{1/2}+Z+T_0}\bigg)$$
$$+\frac{2}{\pi}\int_{0}^{T} \arcsin\Big(\frac{\xi}{T}\Big)\frac{d\xi}{[(Z+\xi)^2-T_0^2]^{1/2}} - \frac{2}{\pi}\int_{0}^{T_0} \arcsin\Big(\frac{\xi}{T_0}\Big)\frac{d\xi}{[(Z+\xi)^2-T^2]^{1/2}}$$
$$+\frac{2}{\pi}\int_{0}^{Z-T} \arcsin\Big(\frac{\xi}{T_0}\Big)\frac{d\xi}{[(Z-\xi)^2-T^2]^{1/2}} - \frac{2}{\pi}\int_{0}^{Z-T_0} \arcsin\Big(\frac{\xi}{T}\Big)\frac{d\xi}{[(Z-\xi)^2-T_0^2]^{1/2}}\bigg],$$

$$\varphi_{R6} = \tfrac{1}{2}\bar{\beta}_0 \ln(T/T_0) + \tfrac{1}{4}(\beta_{0A} - \beta_{0B})\left[-\ln\left(\frac{[Z - T_0 - ((Z-T_0)^2 - T^2)^{1/2}][((Z+T)^2 - T_0^2)^{1/2} + Z + T]}{[Z - T - ((Z-T)^2 - T_0^2)^{1/2}][((Z+T_0)^2 - T^2)^{1/2} + T_0 + Z]} \right) \right.$$

$$+ \frac{2}{\pi}\int_0^T \arcsin\left(\frac{\xi}{T}\right)\left\{\frac{d\xi}{[(Z+\xi)^2 - T_0^2]^{1/2}} - \frac{d\xi}{[(Z-\xi)^2 - T_0^2]^{1/2}}\right\}$$

$$\left. - \frac{2}{\pi}\int_0^{T_0} \arcsin\left(\frac{\xi}{T_0}\right)\left\{\frac{d\xi}{[(Z+\xi)^2 - T^2]^{1/2}} - \frac{d\xi}{[(Z-\xi)^2 - T^2]^{1/2}}\right\}\right].$$

[1] J. Centrella, Ph.D. Thesis, Cambridge University (1978).
[2] J. Centrella, Astrophys. J. **241**, 875 (1980).
[3] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[4] R. Gowdy, Ann. Phys. (NY) **83**, 203 (1974).
[5] C. W. Misner, Phys. Rev. D **8**, 3271 (1973).
[6] B. K. Berger, Ann. Phys. (NY) **83**, 458 (1974).
[7] Ch. Charach, Phys. Rev. D **19**, 3516 (1979); **19**, 1058 (1979).
[8] K. C. Jacobs, Astrophys. J. **155**, 379 (1969).
[9] S. Waller, work in progress (1981).
[10] N. Rosen, Phys. Z. Sowjetunion **12**, 366 (1937).
[11] A. Lapedes, Phys. Rev. D **15**, 946 (1977).
[12] P. Szekeres, J. Math. Phys. **13**, 286 (1972).
[13] J. Centrella and R. Matzner, Astrophys. J. **230**, 311 (1979).
[14] J. Centrella and R. Matzner, Phys. Rev. D **25**, 930 (1982).
[15] I. Gradshteyn and I. Rizhik, *Table of Integrals, Series and Products* (Academic, New York, 1965).
[16] See, for example, W. Bailey, Proc. London Math. Soc. **40**, 37 (1936).
[17] See, for example, E. Butkov, *Mathematical Physics* (Addison-Wesley, Reading, Mass., 1968).
[18] M. Aramowitz and I. Stegun, *Handbook of Mathematical Functions* (U.S. G.P.O., Washington, D.C., 1965).

Erratum: Finite subgroups of SU(3) [J. Math. Phys. 22, 1543 (1981)]

A. Bovier, M. Lüling, and D. Wyler
Physikalisches Institut, Universität Bonn, D-5300 Bonn 1, West Germany

(Received 2 February 1982; accepted for publication 7 May 1982)

PACS numbers: 02.20.Df, 02.20.Rt, 99.10. + g

In the theorem on page 1544 we claimed that the groups $G(3^i PQ,3,a)$ are finite subgroups of SU(3) for all positive integers i. As has been pointed out by W. M. Fairbairn and T. Fulton (University of Lancaster preprint, 1981), this statement is in fact only true for $i = 0$ and $i = 1$. The reason why our proof is erroneous in the cases $i > 1$ is that the representation $[0,a-1]$, used in Lemma 3 of the appendix, is not faithful.

While this reduces the number of new finite SU(3) subgroups, the main result of Sec. II of our paper, namely, the existence of a series of "trihedral" groups as subgroups of SU(3), analogous to the dihedral groups in SU(2), still remains valid.

Section III is unafflicted by this error.

We wish to thank Professor W. M. Fairbairn for communicating his results to us prior to publication.