The character table for the corepresentations of magnetic groups

J. D. Newmarch
Department of Physics, Universiti Pertanian Malaysia, Serdang, Selangor, Malaysia

R. M. Golding
Department of Physical Chemistry, University of New South Wales, P.O. Box 1, Kensington, New South Wales 2033, Australia

(Received 10 February 1981; accepted for publication 27 March 1981)

A square character table is shown to exist for all finite magnetic groups. The table possesses row and column orthogonality properties similar to the character table for linear groups.

PACS numbers: 02.20. + b, 11.30. − j

1. INTRODUCTION

In dealing with the problem of time reversal symmetry in a group theoretic way, Wigner[1] introduced the concept of a corepresentation of a group G of linear/antilinear operators analogous to a representation of a group of linear operators only. It was soon realized that this theory had a ready physical application in dealing with magnetic crystals, where both linear and antilinear operators commute with the Hamiltonian.[2] It is also likely that this theory can be applied to the study of elementary particles due to T or CPT invariance.[3]

Despite its usefulness though, the theory of corepresentations has unpleasant features as many results from the representation theory of groups over the complex numbers do not appear to hold. Let G be a group of linear/antilinear operators, and H the subgroup of linear operators. A corepresentation D of G is a set of matrices over the complex numbers

$$D = \{ D(u), D(a) : u \in H, a \in G - H \}$$

satisfying the following rules

$$\begin{aligned} D(u_1u_2) &= D(u_1)D(u_2), \\ D(ua) &= D(u)D(a), \\ D(au) &= D(a)D(u)^*, \\ D(a_1a_2) &= D(a_1)D(a_2)^*, \end{aligned} \tag{1}$$

where the asterisk denotes complex conjunction. Then

(a) if M is a matrix commuting with D in the sense

$$MD(u) = D(u)M \text{ and } MD(a) = D(a)M^*.$$

Then M is a scalar matrix if and only if it has a *real* eigenvalue.[4]

(b) if D is irreducible,

$$\sum_u D(u)_{ij} D(u)^*_{lk} + \sum_a D(a)_{ik} D(a)^*_{lj} = \frac{|G|}{f}\delta_{il}\delta_{jk},$$

where $|G|$ is the order of G and f the dimension of D.[5] Note how j and k are interchanged in the two sums.

(c) the character of the matrix of an antilinear operator is not invariant under a change of basis. This follows from the transformation rule[6]

$$D'(a) = P^{-1}D(a)P^*. \tag{2}$$

(d) the number of classes need not equal the number of irreducible corepresentations (ICR's). This and the next result can be verified from Cracknell[7] or Newmarch and Golding.[8]

(e) the sum of the squares of the dimensions of the ICR's need not equal the order of G.

After deriving (a) and (b) Dimmock[5] commented ".... further development of the representation theory of nonunitary groups (without using the representation theory of the linear subgroup) has so far proven untenable." He, and others following him, have then relied heavily on the representation theory of linear groups to obtain results about corepresentations (we are not excepted from this!). In particular, the reduction of direct products is usually performed through the intermediary of the irreducible representations of the linear subgroup.[6]

This inevitably gives the impression that corepresentation theory is a poor 'second cousin' to representation theory. In a recent book Cracknell[9] is forced to defend the use of corepresentation theory for magnetic materials against those who feel that ordinary representation theory is quite sufficient, and moreover, has better properties. The best theoretical argument against this view is a demonstration that all fundamental results in representation theory are mirrored by similar fundamental results in corepresentation theory, proved *without using any theorems on representations*. In this paper it is demonstrated that, with certain generalizations and additional concepts, a square character table exists for a finite magnetic group and that this table posseses row and column orthogonality.

All of the results contained here can in fact be derived in a simpler manner by use of representation theory (cf. the character test for the types of ICR). We do not adopt that course as we wish to show that corepresentation can stand independently of representation theory.

First, some preliminary results. From Eq. (1)

$$D(u)^{-1} = D(u^{-1}) \text{ and } D(a)^{-1} = D(a^{-1})^*.$$

Definition: Two corepresentations D_1 and D_2 are *equivalent* if there exists a matrix M such that

$$MD_1(u) = D_2(u)M \text{ and } MD_1(a) = D_2(a)M^*$$

for all $u, a \in G$. The matrix M is said to *intertwine* D_1 and D_2. If D_1 equals D_2, M *commutes* with D_1.

Theorem 1: Every corepresentation is equivalent to a corepresentation by unitary matrices. This has been shown by Dimmock.[5]

Definition: A corepresentation is reducible if it is equivalent to a corepresentation of the form

$$\begin{pmatrix} D_1 & D_2 \\ 0 & D_3 \end{pmatrix}.$$

Otherwise it is irreducible an (ICR).

Theorem 2 (Mashke): Every corepresentation is equivalent to a direct sum of irreducible corepresentations. This has been given before.[6]

2. SCHUR'S LEMMAS

An algebraist once remarked to us "but nothing *interesting* happens in ordinary representation theory!" To some extent we can now sympathize with this view, as what is lost in simplicity is here compensated for by variety, with four useful forms of Schur's lemmas.

Theorem 3 (Schur I): A matrix M intertwining two ICR's D_1 and D_2 is either nonsingular or is zero.

Theorem 4 (Schur II): If M is Hermitian and commutes with a unitary ICR D then M is a real constant matrix. Both of these have been shown by Dimmock.[5]

Theorem 5 (Schur III): If D is a unitary ICR, and M a matrix satisfying $MD(u) = D(u)M$ and $M^+D(a) = D(a)M^*$ for all u, $a \in G$ then M is a constant matrix.

Proof: From

$$\begin{aligned} D(a_1a_2)M &= MD(a_1a_2), \\ D(a_1)D(a_2)^*M &= MD(a_1)D(a_2)^*, \\ \text{or } D(a_1)M^{+*}D(a_2)^* &= MD(a_1)D(a_2)^*. \end{aligned}$$

Hence

$$D(a_1)M^{+*} = MD(a_1).$$

Similarly, from

$$\begin{aligned} D(ua)M^* &= M^+D(ua), \\ D(u)M^+ &= M^+D(u). \end{aligned}$$

Together with the assumptions

$$D(u)(M + M^+) = (M + M^+)D(u)$$

and

$$D(a)(M + M^+)^* = (M + M^+)D(a)$$

for all u, $a \in G$. By Schur II,

$$M + M^+ = \lambda I.$$

Next, from the linearity of u and antilinearity of a,

$$\begin{aligned} D(u)(iM) &= iMD(u), \\ D(a)(iM)^* &= -iM^+D(a), \\ D(u)(iM^+)^* &= iM^+D(u), \\ D(a)(iM^+)^* &= -iMD(a). \end{aligned}$$

So $iM - iM^+$ also satisfies Schur II and is a constant matrix. Hence M is constant as required.

The restriction imposed on M in Schur II that it be Hermitian is a very real one. If it is not, we have already shown[8] that M is nonconstant. It is not possible to say much about any single such matrix, but we can derive results about the *set* of commuting matrices:

$$\mathbf{m} = \{M{:}M \text{ commutes with } D\},$$

$\mathbf{m}$ is closed under matrix multiplication and addition; if $M \in \mathbf{m}$ then so is M^{-1}; it is also closed under scalar multiplication by $\mathbb{R}$, and finally if $M \neq 0$, $kM \neq 0$ for any integer k.

Hence $\mathbf{m}$ is a (skew) field of characteristic zero over $\mathbb{R}$. We can say more about $\mathbf{m}$. Any $M \in \mathbf{m}$ can be written as the sum of a Hermitian and a skew-Hermitian matrix, and it is simple to show that both these belong to $\mathbf{m}$. By Schur II, $\mathbf{m}$ can thus be written as a direct sum

$$\mathbf{m} = \{\lambda I{:}\lambda \in \mathbb{R}\} \oplus \mathbf{m}', \tag{3}$$

where $\mathbf{m}'$ contains only skew-Hermitian matrices. For any $M \in \mathbf{m}'$, M^2 is Hermitian, and since its eigenvalues are negative,

$$M^2 = -\mu^2 I, \quad \text{with } \mu \text{ real.} \tag{4}$$

As $\mathbf{m}'$ is closed under multiplication by $\mathbb{R}$, it follows that for $\mathbf{m}'$ nonempty we can find elements $M_1, M_2, M_3, \cdots$ such that

$$M_i^{-1} = M_i^+ = -M_i. \tag{5}$$

With these preliminaries out of the way, we now show

Theorem 6 (Schur IV): $\mathbf{m}$ is isomorphic to $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{Q}$.

Proof: If $\mathbf{m}'$ is empty, then by Schur II $\mathbf{m}$ is isomorphic to $\mathbb{R}$. Assume, then, that $\mathbf{m}'$ is nonempty. The proof is in two parts: First it is shown that $\mathbf{m}$ contains a multiplicative subgroup isomorphic to the group of $\mathbb{C}$ or the group of $\mathbb{Q}$. Then, it is shown that the algebra of this group over $\mathbb{R}$ equals $\mathbf{m}$.

Let G be a multiplicative subgroup of $\mathbf{m}$ consisting of elements

$$\begin{aligned} G = \{ &\pm I, \pm M_1, \ \pm M_2, \cdots{:}M_i \in \mathbf{m}', \\ &M_i^2 = -I, M_iM_j = -M_jM_i \text{ for all } i,j \neq i\}. \end{aligned}$$

As $M_iM_j = -M_jM_i, M_j \neq M_i$ for $i \neq j$. If such a subgroup only contains the four elements

$$\pm I, \pm M_1,$$

then it is isomorphic to the group of $\mathbb{C}$.

Suppose then it contains more. It cannot contain only six elements for this would mean that M_1M_2 is a multiple of I, M_1, or M_2, which gives a contradiction. Thus it will contain at least eight, and we show that this is the maximum. For consider any $M \in G$ which is not a multiple of I, M_1 or M_2. Then

$$M_1M_2M$$

is Hermitian as $M_i^+ = -M_i$ and all M_i anticommute.

Hence by Schur II

$$M_1M_2M = \lambda I \quad \text{with } \lambda \text{ real.}$$

As $M_i^2 = -I$, $\lambda = \pm 1$, so

$$M = \pm M_1M_2.$$

Therefore

$$G = \{\pm I, \pm M_1, \pm M_2, \pm M_1M_2\},$$

which is easily seen to be the quaternion group. Thus G is either the group of $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{Q}$. It is not hard to check that this is a property of $\mathbf{m}$ rather than the particular group, i.e., if one group is isomorphic to Q, then all are etc. and we can refer to *the* group of $\mathbf{m}$.

For the second part of the proof, we consider the case when the group of $\mathbf{m}$ is the quaternion group as the other two follow as special cases. Further, to show that any matrix in $\mathbf{m}$ belongs to the algebra over $\mathbb{R}$ of G, it is sufficient to show that any $M \in \mathbf{m}'$ is a real linear combination of M_1, M_2, and M_1M_2.

By Hermiticity and Schur II,

$$\begin{aligned} MM_1 &+ M_1M &= aI, \\ MM_2 &+ M_2M &= bI, \\ MM_1M_2 &+ M_1M_2M &= cI, \end{aligned}$$

with a, b, c real. Set

$$N = 2M + aM_1 + bM_2 + cM_1M_2.$$

Clearly $N \in \mathbf{m}'$.

It follows that N anticommutes with M_1, M_2, and M_1M_2.

Hence N is either zero or by normalization an element of the group of $\mathbf{m}$. As there are no other elements of this group, N equals zero and

$$M = -\tfrac{1}{2}(aM_1 + bM_2 + cM_1M_2) \tag{6}$$

as required.

Thus there are possibly three kinds of ICR according as $\mathbf{m}$ is isomorphic to $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{Q}$. That these three types actually occur is shown by our earlier work.[8] It is helpful to quantize this by introducing the intertwining number from the pure mathematicians' version of group theory.[10] Recall that any complex number may be written as an ordered pair of real numbers, and that any quaternion may be written as an ordered quadruple of real numbers. This leads to the following.

Definition: The intertwining number I of $\mathbf{m}$ is the dimension of $\mathbf{m}$ as an algebra over $\mathbb{R}$. An ICR is of type (a) if $\mathbf{m}$ is isomorphic to $\mathbb{R}$ in which case $I = 1$; of type (b) if $\mathbf{m}$ is isomorphic to $\mathbb{Q}$, when $I = 4$, and of type (c) if $\mathbf{m}$ is isomorphic to $\mathbb{C}$ with $I = 2$.

3. ORTHOGONALITY RELATIONS

The general forms of the orthogonality relations have previously been given by Dimmock.[5] They are

Theorem 7: If D_1 and D_2 are two inequivalent ICR's,

$$\sum_u D_1(u)_{ij} D_2(u)^*_{lk} = \sum_a D_1(a)_{ik} D_2(a)^*_{lj} = 0. \tag{7}$$

For D irreducible and unitary,

$$\sum_u D(u)_{ij} D(u)^*_{lk} + \sum_a D(a)_{ik} D(a)^*_{lj} = \delta_{il}\delta_{kj}|G|/f, \tag{8}$$

where f is the dimension of D.

We only remark that the last part of this theorem may be shown in a simpler manner as

$$M = \sum_u D(u) X D(u^{-1}) + \sum_a D(a) X^T D(a^{-1})^*$$

satisfies the conditions of Schur III and hence is diagonal.

This theorem does not take into account the different types of ICR and their properties. The following is proved for an ICR of type (b) and is specialized to types (a) and (c) later.

Theorem 8: If D is a unitary ICR of type (b) with the group of $\mathbf{m}$ generated by M_1 and M_2 then

$$\begin{aligned} &\sum_u D(u)_{ij} D(u)^*_{lk} \\ &= (|G|/2f)\delta_{il}\delta_{jk} - (|G|/2f)(M_1)_{kj}(M_1)_{il} \\ &- (|G|/2f)(M_2)_{kj}(M_2)_{il} - (|G|/2f)(M_1M_2)_{kj}(M_1M_2)_{il}. \end{aligned} \tag{9}$$

Proof: From Schur IV, the following matrix is in $\mathbf{m}$ and can be written

$$\begin{aligned} &\sum_u D(u) X D(u)^+ + \sum_a D(a) X^* D(a)^+ \\ &= \lambda I + \mu M_1 + \omega M_2 + \delta M_1 M_2 \end{aligned} \tag{10}$$

with λ, μ, ω, δ real. Taking Hermitian adjoints

$$\begin{aligned} &\sum_u D(u) X^+ D(u)^+ + \sum_a D(a) X^T D(a)^+ \\ &= \lambda I - \mu M_1 - \omega M_2 - \delta M_1 M_2. \end{aligned} \tag{11}$$

Adding and taking traces,

$$(|G|/2f)(\mathrm{tr}X + \mathrm{tr}X^*) = \lambda. \tag{12}$$

By pre- and post-multiplying these by M_1 we can isolate the term μI to give

$$\mu = -(|G|/2f)[\mathrm{tr}(XM_1) + \mathrm{tr}(XM_1)^*] \tag{13}$$

with similarly

$$\omega = -(|G|/2f)[\mathrm{tr}(XM_2) + \mathrm{tr}(XM_2)^*] \tag{14}$$

and $\delta = -(|G|/2f)[\mathrm{tr}(XM_1M_2) + \mathrm{tr}(XM_1M_2)^*]$. (15)

From Schur III we also have

$$\sum_u D(u) X D(u)^+ + \sum_a D(a) X^T D(a)^+ = zI, \tag{16}$$

with Hermitian adjoint

$$\sum_u D(u) X^+ D(u)^+ + \sum_a D(a) X^* D(a)^+ = z^* I, \tag{17}$$

where $z = (G/f)\mathrm{tr}X$. (18)

The sum over a may be eliminated from Eqs. (10) and (17) to give

$$\begin{aligned} &\sum_u D(u)(X - X^+) D(u)^+ \\ &= (\lambda - z^*)I + \mu M_1 + \omega M_2 + \delta M_1 M_2. \end{aligned}$$

By setting $X_{jk} = 1$ for some j, k and zero otherwise, and then setting $X_{jk} = i$ for the same j, k and zero otherwise, simple manipulations give the result.

These may be specialized to a type (c) ICR by setting $M_2 = 0$ and to a type (a) ICR by also setting $M_1 = 0$. We give a summary for each case, together with the character tests which follow directly with Eq. (8).

Type (a):

$$\sum_u D(u)_{ij} D(u)^*_{lk} = \sum_a D(a)_{ik} D(a)^*_{lj} = \frac{|G|}{2f}\delta_{il}\delta_{jk}, \tag{19}$$

$$\sum_u \chi(u)\chi(u)^* = \sum_a \chi(a^2) = \frac{|G|}{2}. \tag{20}$$

Type (b):

$$\begin{aligned} \sum_u D(u)_{ij} D(u)^*_{lk} &= \frac{|G|}{2f}\delta_{il}\delta_{jk} - \frac{|G|}{2f}(M_1)_{kj}(M_1)_{il} \\ &- \frac{|G|}{2f}(M_2)_{kj}(M_2)_{il} \\ &- \frac{|G|}{2f}(M_1M_2)_{kj}(M_1M_2)_{il}, \end{aligned} \tag{21}$$

$$\sum_a D(a)_{ik} D(a)^*_{lj} = \frac{|G|}{2f}\delta_{il}\delta_{jk} + \frac{|G|}{2f}(M_1)_{kj}(M_1)_{il} + \frac{|G|}{2f}(M_2)_{kj}(M_2)_{il} + \frac{|G|}{2f}(M_1M_2)_{kj}(M_1M_2)_{il}, \quad (22)$$

$$\sum_u \chi(u)\chi(u)^* = 2|G|, \quad (23)$$

$$\sum_a \chi(a^2) = -|G|. \quad (24)$$

Type (c):

$$\sum_u D(u)_{ij} D(u)^*_{lk} = \frac{|G|}{2f}\delta_{il}\delta_{jk} - \frac{|G|}{2f}(M_1)_{kj}(M_1)_{il}, \quad (25)$$

$$\sum_a D(a)_{ik} D(a)^*_{lj} = \frac{|G|}{2f}\delta_{il}\delta_{jk} + \frac{|G|}{2f}(M_1)_{kj}(M_1)_{il}, \quad (26)$$

$$\sum_u \chi(u)\chi(u)^* = |G|, \quad (27)$$

$$\sum_a \chi(a^2) = 0. \quad (28)$$

Equations (7), (20), (23), and (27), when combined with the intertwining number, are actually the row orthogonality relations of the character table. We defer the statement for a discussion of the class concept.

4. CLASSES IN COREPRESENTATION THEORY

It has already been remarked that the number of classes need not equal the number of ICR's (it is always equal to or larger). An examination of previously published tables[7] also shows that in many cases different classes have the same character for all ICR's. Clearly then, the definition of class must be extended for corepresentation theory.

Definition: Two elements u_1 and u_2 of the linear subgroup H are said to be in the same corepresentation class (C class) if *either* $u_1 = u\, u_2 u^{-1}$ for some $u \in H$ *or* $u_1 = a u_2^{-1} a^{-1}$ for some $a \in G - H$.

It is straightforward to check that this is an equivalence relation on H so that a C class may be labelled C_u where u is any element of the C class. This also follows easily:

Theorem 9: The character of a corepresentation is a C-class function on H.

The C class is here only defined over the linear subgroup H; it does not as yet appear useful to extend it to $G - H$.

Theorem 10 (Row Orthogonality): If D_i and D_j are two unitary ICR's with characters on H of χ_i and χ_j respectively, and the number of elements in C_u is n_u, then

$$\sum_{C_u} n_u \chi_i(u)\chi_j(u)^* = \delta_{ij} I_i |H|,$$

where the sum is over all C classes of H and I_i is the intertwining number of D_i.

This follows as stated at the end of the last section. Immediate results from this are

Corollary 1: If D is a corepresentation equivalent to a direct sum of ICR's D_i

$$D = \oplus \sum_i d_i D_i,$$

then

$$d_i = \frac{1}{I_i|H|}\sum_{C_u} n_u \chi(u)\chi_i(u)^*.$$

Corollary 2: If two corepresentations have the same character on H, they are equivalent. Returning to results on C classes,

Theorem 11: (a) $u' C_u u'^{-1} = C_u$ and $a C_u a^{-1} = C_u^{-1}$ for all $u, a \in G$. (b) C_u and C_u^{-1} are in one-to-one correspondence under the mapping $g \to g^{-1}$.

Theorem 12: Let D be a unitary ICR and

$$S_u = \sum_{u' \in C_u} D(u').$$

Then $S_u = zI$.

Proof: This follows by using the previous theorem to show that S_u satisfies the conditions of Schur III.

5. THE REGULAR COREPRESENTATION

The regular corepresentation D_R is useful in corepresentations for exactly the same reasons as the regular representation is; with the elements of G ordered in some arbitrary fixed order, define

$$D_R(g)_{ij} = 1 \text{ if } g = g_i g_j^{-1}$$
$$= 0 \text{ if } g \neq g_i g_j^{-1}.$$

Due to the reality of the matrices, this representation is also a corepresentation. The following are shown in exactly the same manner as in representation theory[11]—once the basic C-class results and row orthogonality are known, the methods of the two theories coincide.

Theorem 13: The number of times an ICR D_i is contained in D_R is

$$2f_i/I_i.$$

Theorem 14:

$$\sum_i \frac{f_i^2}{I_i} = |H|.$$

Theorem 15: If e is the identity of G and D_i is a unitary ICR

$$\sum_i \frac{\chi_i(e)\chi_i(u)^*}{I_i} = \delta(e,u)|H|.$$

Theorem 16 (Column Orthogonality): If D_i is a unitary ICR,

$$\sum_i \frac{\chi_i(u_1)\chi_i(u_2)^*}{I_i} = \delta(C_{u_1}, C_{u_2})\frac{|H|}{n_u}.$$

6. DIRECT PRODUCTS

The (inner) direct product is defined in the normal way by

$$D = D_1 \otimes D_2 \text{ if } D(g)_{ij,kl} = D_1(g)_{ik} D_2(g)_{jl}.$$

From the row orthogonality, this can be reduced directly without reference to the irreducible representations of H. We collect the interesting results in one theorem.

TABLE I. The character tables for the 58 magnetic point groups. The group or groups are given on the upper left of each table, with the ICR's beneath. In the upper middle is given the C classes with the character beneath. To the right is the intertwining number for each ICR.

(a)

$\bar{1}'$	$2'$	m'	E	I
A	A	A	1	1

(b)

	$2'/m$			E	σ_z	
		$2'/m'$		E	I	
$2/m'$			$22'2'$	E	C_{2z}	I
A	A'	A_g	A	1	1	1
B	A''	A_u	B	1	-1	1

(c)

$2m'm'$		E	C_{2z}	
	$2'm'm$	E	σ_y	I
A	A'	1	1	1
B	A''	1	-1	1

(d)

$4'$	$\bar{4}'$	E	C_{2z}	I
A	A	1	1	1
E	E	2	-2	4

(e)

$m'm'm'$			E	C_{2x}	C_{2y}	C_{2z}	
	mmm'		E	C_{2z}	σ_y	σ_x	
		$m'm'm$	E	C_{2z}	I	σ_z	I
A	A_1	A_g	1	1	1	1	1
B_3	A_2	A_u	1	1	-1	-1	1
B_2	B_1	B_g	1	-1	1	-1	1
B_1	B_2	B_u	1	-1	-1	1	1

(f)

$4'22'$	E	C_{2xz}, C_{2y}	C_{2z}	I
A	1	1	1	1
E	2	0	-2	2
B_1	1	-1	1	1

(g)

$4'/m$	E	C_{2z}	I	σ_2	I
A_g	1	1	1	1	1
A_u	1	1	-1	-1	1
E_g	2	-2	2	-2	4
E_u	2	-2	-2	2	4

(h)

	$\bar{4}'2m'$		E	C_{2x}, C_{2y}	C_{2z}	
$4'mm'$		$\bar{4}'2'm$	E	σ_x σ_y	C_{2z}	I
A_1	A	A_1	1	1	1	1
E	E	E	2	0	-2	2
A_2	B_1	A_2	1	-1	1	1

TABLE I (*Continued*).

(i)

$4^1/mmm$	E	C_{2x}, C_{2y}	C_{2z}	I	σ_x, σ_y	σ_z	I
A_g	1	1	1	1	1	1	1
E_g	2	0	-2	2	0	-2	2
B_{1g}	1	-1	1	1	-1	1	1
A_u	1	1	1	-1	-1	-1	1
E_u	2	0	-2	-2	0	2	2
B_{1u}	1	-1	1	-1	1	-1	1

(j)

42^12^1	E	C_2^z	C_{4z}^+	C_{4z}^-	I
A	1	1	1	1	1
B	1	-1	1	-1	1
${}^1E$	1	-1	i	$-i$	1
${}^2E$	1	-1	$-i$	i	1

(k)

$4/m^1$		E	C_{4z}^+, C_{4z}^-	C_{2z}	
	$4^1/m^1$	E	S_{4z}^+, S_{4z}^-	C_{2z}	I
A	A	1	1	1	1
B	B	1	-1	1	1
E	E	2	0	-2	2

(l)

$4m^1m^1$		E	C_{4z}^+	C_{2z}	C_{4z}^-	
	42^1m^1	E	S_{4z}^-	C_{2z}	S_{4z}^+	I
A	A	1	1	1	1	1
B	B	1	-1	1	-1	1
${}^1E$	${}^1E$	1	i	-1	$-i$	1
${}^2E$	${}^2E$	1	$-i$	-1	i	1

(m)

$4/mm^1m^1$	E	C_{4z}^+	C_{2z}	C_{4z}^-	I	S_{4z}^-	σ_z	S_{4z}^+	I
A_g	1	1	1	1	1	1	1	1	1
B_g	1	-1	1	-1	1	-1	1	-1	1
${}^1E_g$	1	i	-1	$-i$	1	i	-1	$-i$	1
${}^2E_g$	1	$-i$	-1	i	1	$-i$	-1	i	1
A_u	1	1	1	1	-1	-1	-1	-1	1
B_u	1	-1	1	-1	-1	1	-1	1	1
${}^1E_u$	1	i	-1	$-i$	-1	$-i$	1	i	1
${}^2E_u$	1	$-i$	-1	i	-1	i	1	$-i$	1

(n)

$4/m^1m^1m^1$			E	C_{2z}	$C_{4z}^\pm$	$C_{2x,y}$	$C_{2a,b}$	
	$4/m^1mm$		E	C_{2z}	$C_{4z}^\pm$	$\sigma_{x,y}$	$\sigma_{da,b}$	
		$4^1/m^1m^1m$	E	C_{2z}	$S_{4z}^\pm$	$C_{2x,y}$	$\sigma_{da,b}$	I
A_1	A_1	A_1	1	1	1	1	1	1
A_2	A_2	A_2	1	1	1	-1	-1	1
B_1	B_1	B_1	1	1	-1	1	-1	1
B_2	B_2	B_2	1	1	-1	-1	1	1
E	E	E	2	-2	0	0	0	1

TABLE I (*continued*).

(o)

$32'$	$3m'$	E	C_3^+	C_3^-	I
A	A	1	1	1	1
^1E	^1E	1	ω	ω^*	1
^2E	^2E	1	ω^*	ω	1

(p)

$\bar{6}'$	$6'$	$\bar{3}'$	E	$C_3^\pm$	I
A	A	A	1	1	1
E	E	E	2	-1	2

(q)

$\bar{3}m'$	E	C_3^+	C_3^-	I	S_6^-	S_6^+	I
A_g	1	1	1	1	1	1	1
^1E_g	1	ω	ω^*	1	ω	ω^*	1
^2E_g	1	ω^*	ω	1	ω^*	ω	1
A_u	1	1	1	-1	-1	-1	1
^1E_u	1	ω	ω^*	-1	$-\omega$	$-\omega^*$	1
^2E_u	1	ω^*	ω	-1	$-\omega^*$	$-\omega$	1

(r)

$6'/m'$	E	$C_3^\pm$	I	$S_6^\pm$	I
A_g	1	1	1	1	1
E_g	2	-1	2	-1	2
A_u	1	1	-1	-1	1
E_u	2	-1	-2	1	2

(s)

$\bar{6}'m2'$		$\bar{3}'m'$			$6'mm'$	E	$C_3^\pm$	$\sigma_{d\,1,2,3}$	
	$\bar{6}'m'2$		$\bar{3}'m'$	$6'22'$		E	$C_3^\pm$	$C'_{21,2,3}$	I
A_1	A_1	A_1	A_1	A_1	A_1	1	1	1	1
A_2	A_2	A_2	A_2	A_2	A_2	1	1	-1	1
E	E	E	E	E	E	2	-1	0	1

(t)

$6'/m'm'm$	E	$C_3^\pm$	$C'_{21,2,3}$	I	$S_6^\pm$	$\sigma_{d\,1,2,3}$	I
A_{1g}	1	1	1	1	1	1	1
A_{2g}	1	1	-1	1	1	-1	1
E_g	2	-1	0	2	-1	0	1
A_{1u}	1	1	1	-1	-1	-1	1
A_{2u}	1	1	-1	-1	-1	1	1
E_u	2	-1	0	-2	1	0	1

(u)

$\bar{6}m'2'$			E	S_3^-	C_3^+	σ_h	C_3^-	S_3^+	
	$62'2'$	$6m'm'$	E	C_6^+	C_3^+	C_2	C_3^-	C_6^-	I
A'	A	A	1	1	1	1	1	1	1
$^1E''$	$^1E_1$	$^1E_1$	1	$-\omega^*$	ω	-1	ω^*	$-\omega$	1
$^2E''$	$^2E_1$	$^2E_1$	1	$-\omega$	ω^*	-1	ω	$-\omega^*$	1
A''	B	B	1	-1	1	-1	1	-1	1
$^2E'$	$^2E_2$	$^2E_2$	1	ω	ω^*	1	ω	ω^*	1
$^1E'$	$^1E_2$	$^1E_2$	1	ω^*	ω	1	ω^*	ω	1

TABLE I (*Continued*).

(v)

$6^1/m$		E	$S_3^\pm$	σ_h	$C_3^\pm$	
$6/m^1$		E	$C_6^\pm$	C_2	$C_3^\pm$	I
A	A'	1	1	1	1	1
E_1	E''	2	1	-2	-1	2
B	A''	1	-1	-1	1	1
E_2	E'	2	1	2	-1	2

(w)

$6/mm^1m$	E	C_6^+	C_3^+	C_2	C_3^-	C_6^-	I	S_3^-	S_6^-	σ_h	S_6^+	S_3^-	I
A_g	1	1	1	1	1	1	1	1	1	1	1	1	1
${}^1E_{1g}$	1	$-\omega^*$	ω	-1	ω^*	$-\omega$	1	$-\omega^*$	ω	-1	ω^*	$-\omega$	1
${}^2E_{1g}$	1	$-\omega$	ω^*	-1	ω	$-\omega^*$	1	$-\omega$	ω^*	-1	ω	$-\omega^*$	1
B_g	1	-1	1	-1	1	1	1	-1	1	-1	1	-1	1
${}^2E_{2g}$	1	ω	ω^*	1	ω	ω^*	1	ω	ω^*	1	ω	ω^*	1
${}^1E_{2g}$	1	ω^*	ω	1	ω^*	ω	1	ω^*	ω	1	ω^*	ω	1
A_u	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	1
${}^1E_{1u}$	1	$-\omega^*$	ω	-1	ω^*	$-\omega$	-1	ω^*	$-\omega$	1	$-\omega^*$	ω	1
${}^2E_{1u}$	1	$-\omega$	ω^*	-1	ω	$-\omega^*$	-1	ω	$-\omega^*$	1	$-\omega$	ω^*	1
B_u	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	1
${}^2E_{2u}$	1	ω	ω^*	1	ω	ω^*	-1	$-\omega$	$-\omega^*$	-1	$-\omega$	$-\omega^*$	1
${}^1E_{2u}$	1	ω^*	ω	1	ω^*	ω	-1	$-\omega^*$	$-\omega$	-1	$-\omega^*$	$-\omega$	1

(x)

$6^1/mm^1m$			E	σ_h	$C_3^\pm$	$S_3^\pm$	C'_{2i}	σ_{vi}	
	$6/m^1m^1m$		E	C_{2z}	$C_3^\pm$	$C_6^\pm$	C'_{2i}	C''_{2i}	
		$6/m^1mm$	E	C_{2z}	C_3^+	$C_6^\pm$	σ_{di}	σ_{vi}	I
A'_1	A_1	A_1	1	1	1	1	1	1	1
A'_2	A_2	A_2	1	1	1	1	-1	-1	1
A''_1	B_1	B_2	1	-1	1	-1	1	-1	1
A''_2	B_2	B_1	1	-1	1	-1	-1	1	1
E''	E_1	E_1	2	-2	-1	1	0	0	1
E'	E_2	E_2	2	2	-1	-1	0	0	1

(y)

m^13	E	C_{2m}	$C_{3j}^\pm$	I
A	1	1	1	1
E	2	1	-1	2
T	3	-1	0	1

(z)

$\bar{4}^13m^1$	4^132^1	E	C_{2m}	C_{3j}^-	C_{3j}^+	I
A	A	1	1	1	1	1
${}^1E$	${}^1E$	1	1	ω	ω^*	1
${}^2E$	${}^2E$	1	1	ω^*	ω	1
T	T	3	-1	0	0	1

(aa)

$m3m^1$	E	C_{2m}	C_{3j}^-	C_{3j}^+	I	σ_m	S_{6j}^+	S_6	I
A_g	1	1	1	1	1	1	1	1	1
${}^1E_g$	1	1	ω	ω^*	1	1	ω	ω^*	1
${}^2E_g$	1	1	ω^*	ω	1	1	ω^*	ω	1
T_g	3	-1	0	0	3	-1	0	0	1
A_u	1	1	1	1	-1	-1	-1	-1	1
${}^1E_u$	1	1	ω	ω^*	-1	-1	$-\omega$	$-\omega^*$	1
${}^2E_u$	1	1	ω^*	ω	-1	-1	$-\omega^*$	$-\omega$	1
T_u	3	-1	0	0	-3	1	0	0	1

TABLE I (*Continued*).

(bb)

$m'3m'$		E	$C^{\pm}_{3j}$	C_{2m}	C_{2p}	$C^{\pm}_{4m}$	
	$m'3m$	E	$C^{\pm}_{3j}$	C_{2m}	σ_{dp}	$S^{\pm}_{4m}$	I
A_1	A_1	1	1	1	1	1	1
A_2	A_2	1	1	1	-1	-1	1
E	E	2	-1	2	0	0	1
T_1	T_1	3	0	-1	-1	1	1
T_2	T_2	3	0	-1	1	-1	1

Theorem 17:

(a) If D_i, D_j, and D_k are ICR's and

$$D_i \otimes D_j = \oplus \sum_k d^{\,k}_{ij} D_k.$$

Then

$$d^{\,k}_{ij} = \frac{1}{I_k |H|} \sum_{C_u} n_u \chi_i(u) \chi_j(u) \chi_k(u)^*.$$

(b) If $\mathbf{0}$ is the identity ICR,

$$d^{\,\mathbf{0}}_{ii^*} = I_i.$$

(c) If $d^{\,\mathbf{0}}_{ijk}$ is the multiplicity of $\mathbf{0}$ in $D_i \otimes D_j \otimes D_k$ then

$$d^{\,\mathbf{0}}_{ijk} = d^{\,k^*}_{ij} I_k.$$

This difference between the double and triple product is of great importance in developing a Racah algebra for such groups.[8]

Symmetrized and antisymmetrized squares are necessary in dealing with a number of fermions or bosons; symmetrized cubes are used in magnetic phase transitions[9]; symmetrized, antisymmetrized and mixed symmetry cubes separate out the permutation properties of the $3jm$ symbols. These can all be distinguished by character tests. For completeness we summarize them here. The notation used is $\{\lambda\}$, where $\{\lambda\}$ is a Young diagram of S_n.

(a) $\chi_{\{2\}}(u) = \frac{1}{2}([\chi(u)]^2 + \chi(u^2))$,

(b) $\chi_{\{1^2\}}(u) = \frac{1}{2}([\chi(u)]^2 - \chi(u^2))$,

(c) $\chi_{\{3\}}(u) = \frac{1}{6}([\chi(u)]^3 + 3\chi(u^2)\chi(u) + 2\chi(u^3))$,

(d) $\chi_{\{1^3\}}(u) = \frac{1}{6}([\chi(u)]^3 - 3\chi(u^2)\chi(u) + 2\chi(u^3))$,

(e) $\chi_{\{21\}}(u) = \frac{1}{3}([\chi(u)]^3 - \chi(u^3))$.

The row orthogonality now allows a direct reduction of these powers without use of tables relating these powers to the linear subgroup.

7. CONCLUSION

In this paper it has been shown that the powerful concept of a character table applies to finite magnetic groups as well as linear groups. The only added complexity is the simple intertwining number. The character table will expedite calculations as well as helping to show that corepresentation theory can stand upright without leaning on representation theory for most of its results.

To make the theory more concrete, the character tables for the fifty-eight magnetic point groups are given. They have been adapted from the tables of Cracknell[7] and all notations are the same as there.

APPENDIX: DESCENT IN SYMMETRY TO THE LINEAR SUBGROUP

The results obtained so far have been done without any reference to the representation theory of linear groups. It is known, however, that there are strong relations between the ICR's of G and the irreducible representations (IR's) of H.[1] These are generally shown by *ascent* in symmetry where the ICR's of G and their properties are determined by the IR's of H. From the methods developed earlier, we now reverse this and derive these relations by *descent* from G to H. No new results are demonstrated—rather the interplay between Schur's lemmas for linear and nonlinear groups is shown.

First we fix notation, and give the row orthogonality of an ICR of G subduced to H. Let D be a unitary ICR of G and Δ the possibly reducible representation of H obtained by descent to H. χ, the character of D on H, is also the character of Δ. Row orthogonality then gives

$$\sum_u \chi(u)\chi(u)^* = I\,|H|$$

with I the intertwining number of D. Each type of ICR is now considered in turn.

Type (a): Since I equals one, Δ is an IR. Setting

$$P = D(a_0)$$

for an arbitrary fixed element of $G - H$,

$$\Delta(a_0^2) = D(a_0) D(a_0)^* = PP^*$$

and

$$\Delta(a_0 u a_0^{-1}) = D(a_0 u a_0^{-1}) = P\Delta(u)^* P^+.$$

The ICR matrices are then given by

$$D(u) = \Delta(u) \text{ and } D(a) = D(a a_0^{-1} a_0) = \Delta(a a_0^{-1})P.$$

The IR satisfies the character test

$$\sum_a \chi_\Delta(a^2) = |H|.$$

For the other two types of ICR, Δ is reducible. To gain the results given by other authors[6] we consider the special case in which a unitary transformation has been applied to D so that Δ is in completely reduced form.

Type (b): As the intertwining number is four, Δ is reducible to either $\Delta_1 \oplus \Delta_1$ or $\Delta_1 \oplus \Delta_2 \oplus \Delta_3 \oplus \Delta_4$. This second possibility soon leads to a contradiction for by Schur's lemma for linear groups any $M \in \mathbf{m}$ must be

$$\begin{pmatrix} z_1 I & & & 0 \\ & z_2 I & & \\ & & z_3 I & \\ 0 & & & z_4 I \end{pmatrix}.$$

Schur IV shows then that $M \in \mathbf{m}'$ is purely imaginary so that $M_1, M_2, and\, M_1 M_2$ are purely imaginary. This is the required contradiction and so $\Delta = \Delta_1 \oplus \Delta_1$, i.e.,

$$D(u) = \Delta(u) = \begin{pmatrix} \Delta_1(u) & 0 \\ 0 & \Delta_1(u) \end{pmatrix}.$$

By Schur's lemma for linear groups applied to $M \in \mathbf{m}$

$$M = \begin{pmatrix} z_1 I & z_2 I \\ z_3 I & z_4 I \end{pmatrix}.$$

But by Schur II, $M + M^+ = \lambda I$ and $(M - M^+)^2 = -\mu^2 I$. Hence

$$M = x_1 \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} + x_2 \begin{pmatrix} iI & 0 \\ 0 & -iI \end{pmatrix} + x_3 \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 & iI \\ iI & 0 \end{pmatrix}.$$

This in turn imposes restrictions on $D(a)$ as $MD(a) = D(a)M^*$. Choosing an arbitrary element $a_0 \in G - H$ gives

$$D(a_0) = \begin{pmatrix} 0 & P \\ -P & 0 \end{pmatrix},$$

$$D(a) = \begin{pmatrix} 0 & \Delta_1(aa_0^{-1})P \\ -\Delta_1(aa_0^{-1})P & 0 \end{pmatrix},$$

$$\Delta_1(a_0^2) = -PP^*,$$

$$\Delta_1(a_0 u a_0^{-1}) = P\Delta_1(u)^* P^+,$$

and

$$\sum_a \chi_{\Delta_1}(a^2) = -|H|.$$

Type (c): The intertwining number is now two so Δ is equivalent to $\Delta_1 \oplus \Delta_2$ with $\Delta_1 \not\equiv \Delta_2$. If Δ is in completely reduced form

$$\Delta(u) = \begin{pmatrix} \Delta_1(u) & 0 \\ 0 & \Delta_2(u) \end{pmatrix},$$

the same reasoning as before gives $M \in \mathbf{m}$ as

$$M = x_1 \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} + x_2 \begin{pmatrix} iI & 0 \\ 0 & -iI \end{pmatrix}.$$

For fixed arbitrary $a_0 \in G - H$,

$$D(a_0) = \begin{pmatrix} 0 & P_1 \\ P_2 & 0 \end{pmatrix},$$

$$D(a) = \begin{pmatrix} 0 & \Delta_1(aa_0^{-1})P_1 \\ \Delta_2(aa_0^{-1})P_2 & 0 \end{pmatrix},$$

$$\Delta_1(a_0^2) = P_1 P_2^* \text{ and } \Delta_2(a_0^2) = P_2 P_1^*,$$

and

$$\Delta_1(a_0 u a_0^{-1}) = P_1 \Delta_2(u)^* P_1^+,$$

$$\Delta_1(a_0 u a_0^{-1}) = P_2 \Delta_2(u)^* P_2^+,$$

from which the character test follows:

$$\sum_u \chi_{\Delta_1}(a^2) = \sum_a \chi_{\Delta_2}(a^2) = 0.$$

[1] E. P. Wigner, *Group Theory and its Applications to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).
[2] B. A. Tavger and V. M. Zaitzev, Sov. Phys. JETP **3**, 430 (1956).
[3] J. McL. Emmerson, *Symmetry Principles in Particle Physics* (Clarendon, Oxford, England, 1972).
[4] P. Rudra, J. Math Phys. **15**, 2031 (1974).
[5] J. O. Dimmock, J. Math. Phys. **4**, 1307 (1963).
[6] C. J. Bradley and B. L. Davies, Rev. Mod. Phys. **40**, 359 (1968).
[7] A. P. Cracknell, Prog. Theor. Phys. **35**, 196 (1966); Erratum, Aust. J. Phys. **20**, 173 (1967).
[8] J. D. Newmarch and R. M. Golding, J. Math. Phys. **22**, 233 (1981).
[9] A. P. Cracknell, *Magnetism in Crystalline Materials* (Pergamon, New York, 1975).
[10] C. W. Curtiss and I. Reiner, *Representation Theory of Finite Groups and Associative Algebras* (Wiley, New York, 1962).
[11] L. Jansen and M. Boon, *Theory of Finite Groups. Applications in Physics* (North-Holland, Amsterdam, 1967).

Complex orthogonal and symplectic matrices depending on parameters [a)]

J. Patera
Centre de Recherche de Mathématiques Appliquées, Université de Montréal, Québec, Canada
C. Rousseau
Département de Mathématiques et de Statistique, Université de Montréal, Montréal, Québec, Canada

(Received 18 February 1981; accepted for publication 24 July 1981)

Versal deformations of elements of complex orthogonal and symplectic Lie algebras are studied. For a general element M of $o(n,\mathbb{C})$ or $sp(2n,\mathbb{C})$, a normal form M_A is found which, unlike the Jordan normal form M_J, depends holomorphically on M and on the similarity transformation $M_A = gMg^{-1}$ from the corresponding group. Orthogonal and symplectic cases are treated simultaneously in order to underline their close relation. Bundles of matrices of low codimension are listed and bifurcation diagrams of two-parameter families of orthogonal matrices are shown. Finally, versal deformations of all elements of $o(6,\mathbb{C})$ are explicitly shown.

PACS numbers: 02.20.Sv, 02.20.Qs

I. INTRODUCTION

In 1971, Arnold[1] pointed out that a complex matrix $M\in\mathbb{C}^{n\times n}$, whose matrix elements are known with only a finite precision, and with some multiple eigenvalues, may not always be transformed into its Jordan normal form M_J. A small change in M may result in a different structure of M_J. Thus the transformation $M_J = gMg^{-1}$, $g\in GL(n,\mathbb{C})$, is not stable: it does not depend in a continuous way on either M or g. This difficulty is not too serious, as long as one is interested in just one matrix M. Indeed, one can always find a matrix with all eigenvalues distinct and as close as necessary to M. If, however, we are interested in a whole family of matrices depending holomorphically on complex parameters, multiple eigenvalues cannot be avoided in some members of the family.

The result of Arnold[1] is a new normal form, say M_A, such that not only can M be transformed into M_A, but any family of matrices "close" to M. The normal form of Arnold, M_A, is computed as the versal deformation of M_J, using techniques of deformation theory. From this, Arnold describes the structure of any matrix belonging to a k-parameter family of matrices in general position. For small k he describes precisely the bifurcation diagrams: i.e., the set of parameters values which correspond to matrices with multiple eigenvalues.

The transformation of M to the Jordan normal form M_J or to the Arnold normal form M_A is a $GL(n,\mathbb{C})$-conjugacy transformation of the element $M\in gl(n,\mathbb{C})$ of the Lie algebra $gl(n,\mathbb{C})$ of the general linear group $GL(n,\mathbb{C})$. Since every M' close to M can be transformed into M_A, the normal form M_A "intersects" all $GL(n,\mathbb{C})$-orbits of elements of $gl(n,\mathbb{C})$ which are close to M. Therefore, M_A has to contain some parameters.

Normal forms of elements of Lie algebras or, equivalently, representatives of conjugacy classes were often studied.[2,3] Analogs of the Jordan normal form M_J in the case of $gl(n,\mathbb{C})$ were found for all real and complex Lie algebras.[3,4] The discontinuity of the dependence of the normal forms on the original matrix M and on the transforming matrix g was always disregarded. Thus the question concerning the description of the versal deformations M_A of M_J can be asked independently for matrices M belonging to any of these Lie algebras.

The problem is of particular importance in applications when the matrix M arises as a result of measurements, i.e., its matrix elements are given with errors. Besides the case $M\in gl(n,\mathbb{C})$ studied by Arnold,[1] the versal deformations were found also for $M\in gl(n,\mathbb{R})$ in Ref. 5, and for the symplectic Lie algebras $sp(2n,\mathbb{C})$ and $sp(2n,\mathbb{R})$ (cf. Ref. 6), by Galin. Investigation of the symplectic case was motivated by classical mechanics. The Arnold normal forms M_A of elements of the symplectic Lie algebras (Hamiltonian matrices) allow, for instance, the transformation of quadratic Hamiltonians[6,7] and Hamilton equations to a simple form depending continuously, and even smoothly, on any Hamiltonian known only approximately.

In the relativistic and quantum physics the orthogonal groups also play an important role. The purpose of this article is to find the Arnold normal forms of matrices M belonging to the Lie algebra $o(n,\mathbb{C})$. General results are illustrated on the case $o(6,\mathbb{C})$ as the potentially most interesting in physics. Indeed, among real forms of $o(6,\mathbb{C})$ and their subalgebras, one finds practically all the Lie algebras of relativistic physics.

Throughout the paper a Lie algebra (group) is identified with its natural representation in terms of matrices. In order to underline similarities and differences between the orthogonal and symplectic cases, they are treated simultaneously.

Section II contains a standard description of elements of the Lie algebras $o(n,\mathbb{C})$ and $sp(2n,\mathbb{C})$, and their "Jordan normal forms." In Sec. III versal deformations and their properties are recalled. The Arnold normal forms of orthogonal and symplectic matrices are found in Sec. IV. Bifurcation diagrams of two-parameter families of orthogonal matrices in general position are described in Sec. V. Section VI contains explicit description of minimal versal deformations of $o(6,\mathbb{C})$ matrices.

[a)] Work supported in part by the Natural Science and Engineering Research Council of Canada and by the Ministère de l'Education du Québec.

II. CONJUGACY CLASSES OF ELEMENTS OF COMPLEX LIE ALGEBRAS

Here we give representatives ("Jordan normal forms") of elements of conjugacy classes of the Lie algebras $o(2n+1,\mathbb{C})$, $o(2n,\mathbb{C})$, and $sp(2n,\mathbb{C})$ which are further used in Sec. IV. They are the analogs of the Jordan normal forms in the case of the algebra $gl(n,\mathbb{C})$. The results are well known, and our presentation follows Ref. 4.

Orthogonal and symplectic cases are treated simultaneously, $l(N,\mathbb{C})$ and $L(N,\mathbb{C})$ denoting the corresponding Lie algebra and Lie group, respectively. Unless otherwise stated the two cases differ only by the value of the parameter ϵ. $\epsilon = 1$ (resp., $= -1$) in the orthogonal (resp., symplectic) case. Below we often use the following $n \times n$ matrices.

$$I_n = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}, \quad J_n = \begin{pmatrix} 0 & 1 & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & 1 \\ & & & & 0 \end{pmatrix},$$

$$F_n = \begin{pmatrix} & & & & 1 \\ & & & -1 & \\ & & 1 & & \\ & \cdot^{\cdot^{\cdot}} & & & \\ \epsilon & & & & \end{pmatrix}. \tag{2.0}$$

The group $L(N,\mathbb{C})$ is the group of operators on $\mathbb{C}^n$ preserving a given nondegenerate bilinear form on $\mathbb{C}^n$. The form f is symmetric for $O(N,\mathbb{C})$ and skew-symmetric for $Sp(2n,\mathbb{C})$; f is unique up to isomorphism.

Each matrix K satisfying

$$K^T = \epsilon K, \quad \det K \neq 0. \tag{2.1}$$

specifies an invariant form $f(x,y) = x^T K y$. Two matrices K and K' both with the property (2.1) and the same value of ϵ are equivalent. That is, $K' = gKg^{-1}$ for some $g \in GL(N,\mathbb{C})$. $L(N,\mathbb{C})$ is then represented as the set of nonsingular matrices $g \in \mathbb{C}^{N\times N}$ such that

$$\forall x, y \quad f(x,y) = f(gx, gy) = x^T g^T K g y. \tag{2.2}$$

This implies

$$g^T K g = K. \tag{2.3}$$

A matrix $M \in \mathbb{C}^{N\times N}$ belongs to the Lie algebra $l(N,\mathbb{C})$ iff the "infinitesimal equivalent" of (2.2) is satisfied, i.e.,

$$\forall x, y \quad 0 = (Mx, y) + (x, My) = x^T M^T K y + x^T K M y, \tag{2.4}$$

which implies that

$$KM + M^T K = 0. \tag{2.5}$$

In order to specify an element of $l(N,\mathbb{C})$, one has to give two matrices M and K satisfying (2.1) and (2.5). Our matrix K can change from case to case in order to get M in the simplest possible form (cf. Ref. 4).

Two elements M and M' of $l(N,\mathbb{C})$ belong to the same $L(N,\mathbb{C})$-conjugacy class or to the same orbit [are $L(N,\mathbb{C})$-conjugate] iff

$$M' = gMg^{-1} \quad \text{for some } g \in L(N,\mathbb{C}). \tag{2.6}$$

Thus, the $l(N,\mathbb{C})$-orbit, Orb(M), containing the element M is defined as

$$\mathrm{Orb}(M) = \{gMg^{-1} | g \in L(N,\mathbb{C})\}. \tag{2.7}$$

It is convenient to split the set of orbits of $l(N,\mathbb{C})$-elements into three mutually exclusive sets. To the first set belong the completely indecomposable elements. The second consists of decomposable elements which are orthogonally indecomposable. The third set contains the decomposable elements. These are the other $l(N,\mathbb{C})$ matrices, which are diagonalizable or partially diagonalizable. Naturally this is the most "abundant" type of elements. Let us point out that nonzero eigenvalues of any element of $l(N,\mathbb{C})$ occur in pairs $\pm\alpha$.

We now list the normal forms. It is convenient to choose the matrix K for each case separately as in Ref. 4. This allows us to write the matrix M_J in a very simple way.

Elements of the completely indecomposable sets are $L(N,\mathbb{C})$-conjugate to

$$M_J = J_N, \quad K = F_N. \tag{2.8}$$

According to (2.1), $K^T = \epsilon K$. Therefore, (2.8) and (2.0) imply that

$$\epsilon = \begin{cases} +1 & \text{for } N \text{ odd} \\ -1 & \text{for } N \text{ even.} \end{cases} \tag{2.9}$$

Thus the normal form (2.8) occurs in $o(2n+1,\mathbb{C})$ and $sp(2n,\mathbb{C})$, but not in $o(2n,\mathbb{C})$. For simplicity we call (2.8) the normal form of case I.

The orthogonally indecomposable but otherwise decomposable elements appear only for N even, $N = 2n$. They are $L(N,\mathbb{C})$-conjugate to

$$M_J = \begin{pmatrix} A & 0 \\ 0 & -A^T \end{pmatrix}, \quad K = \begin{pmatrix} 0 & I_n \\ \epsilon I_n & 0 \end{pmatrix}, \tag{2.10}$$

where

$$A = \alpha I_n + J_n, \quad \alpha \in \mathbb{C}. \tag{2.11}$$

A is simply a Jordan block. We distinguish the value $\alpha = 0$ and call the corresponding normal form (2.10) case II (for $\epsilon = +1$ n is even, for $\epsilon = -1$ n is odd). When $\alpha \neq 0$, (2.10) is the normal form of case III.

The indecomposable Jordan normal forms are summarized in Table I.

The Jordan normal form of a decomposable element is a direct sum of normal forms $(M_J^{(i)}, K^{(i)})$ of types I, II, and III:

$$M_J = \bigoplus_{i=1}^{s} M_J^{(i)}, \quad K = \bigoplus_{i=1}^{s} K^{(i)}, \tag{2.12}$$

with the obvious restriction that orthogonal (symplectic) elements contain only orthogonal (symplectic) summands. Furthermore, two decomposable normal forms which differ only by permutation of terms of the direct sum are $L(N,\mathbb{C})$-conjugate.

III. VERSAL DEFORMATIONS

Here we consider deformations of elements of a complex Lie algebra $\mathscr{G}$, corresponding to a Lie group G. A deformation of an element $M_0 \in \mathscr{G}$ is a family $M(\lambda)$ depending holomorphically on complex parameters $\lambda_1, \ldots, \lambda_m$ such that

TABLE I. Indecomposable and orthogonally indecomposable Jordan normal forms of elements of the Lie algebras $\mathrm{gl}(n,\mathbb{C})$, $\mathrm{o}(2n+1,\mathbb{C})$, $\mathrm{o}(2n,\mathbb{C})$, and $\mathrm{sp}(2n,\mathbb{C})$. Matrices I_n, J_n, and F_n are defined in Eq. (2.0).

Algebra	Name	Case	M_J	K	Eigenvalues
$\mathrm{gl}(n,\mathbb{C})$	indecomposable		$\alpha I_n + J_n$	—	$\alpha\in\mathbb{C}$
$\mathrm{o}(2n+1,\mathbb{C})$ $\epsilon=+1$	completely indecomposable	I	J_{2n+1}	F_{2n+1}	0
$\mathrm{o}(2n,\mathbb{C})$ $\epsilon=+1$	decomposable but not orthogonally	II	$\begin{pmatrix} J_n & 0 \\ 0 & -J_n^T \end{pmatrix}$	$\begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix}$	0
		III	$\begin{pmatrix} \alpha I_n + J_n & 0 \\ 0 & -\alpha I_n - J_n^T \end{pmatrix}$	$\begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix}$	$\pm\alpha,\ \alpha\neq 0$
$\mathrm{sp}(2n,\mathbb{C})$ $\epsilon=-1$	completely indecomposable	I	J_{2n}	F_{2n}	0
	decomposable but not orthogonally	II	$\begin{pmatrix} J_n & 0 \\ 0 & -J_n^T \end{pmatrix}$	$\begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$	0
		III	$\begin{pmatrix} \alpha I_n + J_n & 0 \\ 0 & -\alpha I_n - J_n^T \end{pmatrix}$	$\begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$	$\pm\alpha,\ \alpha\neq 0$

$M(0)=M_0$. We are interested in versal deformations of M_0, i.e., deformations that "contain" representatives of orbits of all deformations of M_0 in $\mathscr{G}$.

Definition: $M(\lambda_1,\ldots,\lambda_m)$ is a versal deformation of M_0 iff for any deformation $M'(\mu_1,\ldots,\mu_k)$ of M_0, there exist holomorphic mappings

$$\varphi: V\subset\mathbb{C}^k\to\mathbb{C}^m,$$

$$g: V\to G$$

such that V is a neighborhood of 0 in $\mathbb{C}^k$

$$\varphi(0)=0,\quad g(0)=I,$$

$$M'(\mu)=g(\mu)M(\varphi(\mu))g^{-1}(\mu). \tag{3.1}$$

Intuitively $M(\lambda)$ is versal iff $M(\lambda)$ and the conjugates of its elements under the action of G fill up a neighborhood of M_0. Therefore, the image of $M(\lambda)$ and $\mathrm{Orb}(M_0)$ must intersect properly and the sum of their dimension must be at least the dimension of $\mathscr{G}$. Since we are interested in local properties, their precise formulation is in terms of tangent spaces, namely:

Lemma (*Arnold* [1]): $M(\lambda)$ is a versal deformation of M_0 iff $M(\lambda)$ is transversal to $\mathrm{Orb}(M_0)$ at M_0, i.e.,

$$M_* T_0\mathbb{C}^m + T_{M_0}\,\mathrm{Orb}\,(M_0) = T_{M_0}\mathscr{G}, \tag{3.2}$$

where $T_0\mathbb{C}^m$ is the tangent space of $\mathbb{C}^m$ at the point $0\in\mathbb{C}^m$, $T_{M_0}\,\mathrm{Orb}\,(M_0)$ is the tangent space of $\mathrm{Orb}(M_0)$ at M_0, $T_{M_0}\mathscr{G}$ is the tangent space of $\mathscr{G}$ at M_0, and M_* is the derivative of M at $0\in\mathbb{C}^m$.

Therefore, the minimal number of parameters of a versal deformation is the codimension of $\mathrm{Orb}(M_0)$.

The action of G on $\mathscr{G}$ gives us the map

$$\psi: G\to\mathscr{G},\quad \psi(g)=gM_0g^{-1}.$$

We remark that $\mathrm{Orb}M_0=\mathrm{Im}\psi$. The tangent space of $\mathrm{Orb}M_0$ is the image of the tangent space of G at the identity under the derivative map ψ_*:

$$\psi_*: T_eG=\mathscr{G}\to T_{M_0}\mathscr{G}=\mathscr{G},\quad \psi_*(C)=[C,M_0].$$

From this we get

$$T_{M_0}\mathrm{Orb}(M_0)=\{[C,M_0]\,|\,C\in\mathscr{G}\}. \tag{3.3}$$

The codimension of $\mathrm{Orb}M_0$ is the dimension of the centralizer of M_0 defined by

$$\mathrm{Cent}_{\mathscr{G}}M_0=\{C\in\mathscr{G}\,|\,[C,M_0]=0\}=\mathrm{Ker}\psi_*. \tag{3.4}$$

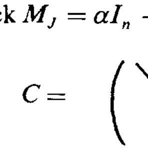
We specialize now to the case where elements of G and $\mathscr{G}$ are complex matrices. On $\mathscr{G}$ we have the following inner product $(A,B)=\mathrm{tr}(AB^*)$, where $B^*=\overline{B^T}$ is the adjoint of B (complex conjugate, transpose of B). Among all possible deformations of M_0, we choose one, $M(\lambda)$, which is orthogonal to $\mathrm{Orb}M_0$. We use therefore the following lemma.

Lemma (*Arnold* [1]): $B\in\mathscr{G}$ is orthogonal to $\mathrm{Orb}(M_0)$ iff $[B^*,M_0]=0$.

The problem is now reduced to the computation of the adjoint of the centralizer of M_0. Arnold did the calculation for the case $G=\mathrm{GL}(n,\mathbb{C})$, $\mathscr{G}=\mathrm{gl}(n,\mathbb{C})$, and $M_0=M_J$, the Jordan normal form. We recall here his results since we use them later.

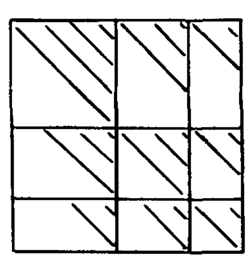
Centralizer of M_J in $\mathrm{gl}(n,\mathbb{C})$

(1) M_J is a Jordan block $M_J=\alpha I_n+J_n$ [cf. (2.0)]

$C\in\mathrm{Cent}_{\mathrm{gl}(n,\mathbb{C})}M_J$ iff $C=$  (3.5)

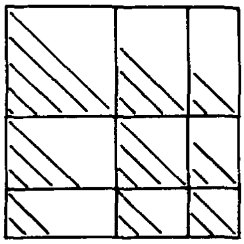
(each oblique segment denotes a sequence of identical numbers).

(2) M_J is a Jordan matrix with a single eigenvalue α and Jordan blocks $J_1,\ldots,J_r$. Then $C\in\mathrm{Cent}_{\mathrm{gl}(n,\mathbb{C})}M_J$ iff

$C=$  , (3.6) where each oblique segment denotes a sequence of identical numbers. Here the figure is done for $r = 3$.

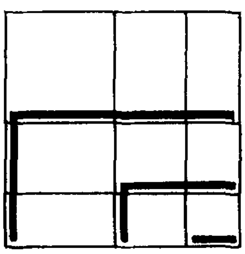
(3) If $M_J = \oplus_\alpha M_J^\alpha$, where the sum is over all distinct eigenvalues of M_J, then $C \in \mathrm{Cent}_{\mathrm{gl}(n,\mathbb{C})} M_J$ iff $C = \oplus_\alpha C^\alpha$ with $[C^\alpha, M_J^\alpha] = 0$.

Thus the versal deformation of the Jordan matrix M_J with a single eigenvalue α and three blocks ($r = 3$) which is orthogonal to the orbit $\mathrm{Orb}(M_J)$ at M_J and depends on minimal number of parameters is given by $M_J + M(\lambda)$, where

$M(\lambda) =$  . (3.7)

Each oblique segment denotes a sequence of identical numbers and to different segments correspond different parameters.

Often it is convenient to consider the minimal versal deformation of M_J with minimal number of nonzero entries in $M(\lambda)$. It can be chosen, for instance, as $M_J + M'(\lambda)$, where

$M'(\lambda) =$  . (3.8)

There is an independent parameter for each entry placed on a dark line. The number of nonzero entries of $M'(\lambda)$ equals the number of independent parameters of the deformation; the deformation is versal but not orthogonal to the orbit.

When M_J has several eigenvalues, the versal deformation of M_J is just the direct sum of the deformations described above for each eigenvalue.

IV. MINIMAL VERSAL DEFORMATIONS OF ELEMENTS OF o(N,$\mathbb{C}$) AND sp($2n$,$\mathbb{C}$)

In Sec. III, the construction of minimal versal deformations of the Jordan normal form of elements of $\mathrm{l}(N,\mathbb{C})$ was reduced to the computation of the centralizers

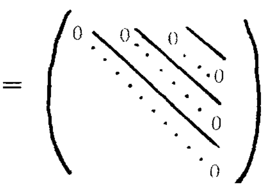
$$\mathrm{Cent}_{\mathrm{l}(N,\mathbb{C})} M_J = \{ C \mid [C, M_J] = 0 \quad \& \quad KC + C^T K = 0 \} \tag{4.1}$$

for each normal form specified by M_J and K. In this section we first find the centralizers and then describe the minimal versal deformations.

A. Centralizers of normal forms in l(N,$\mathbb{C}$)

The computation is straightforward but lengthy. It consists in imposing the two requirements in (4.1) on a general matrix C. Here we just give the results. We start with the orthogonally indecomposable cases: cases I, II, and III of Table I.

(*a*) *Case I*: $M_J = J_N$, $K = F_N$:
$N = 2n$ [for sp($2n$, $\mathbb{C}$)] or $N = 2n + 1$[for o($2n+1$)].

$C \in \mathrm{Cent}_{\mathrm{l}(N,\mathbb{C})} M_J$ iff $C =$  (4.2)

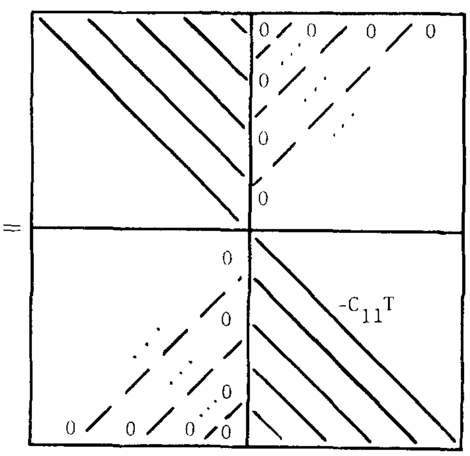
Each solid line indicates a unique complex parameter repeated on each row. The dimension of the centralizer is $d = n$.

(*b*) *Case II*:

$$M_J = \begin{pmatrix} J_n & 0 \\ 0 & -J_n^T \end{pmatrix}, \quad K = \begin{pmatrix} & I_n \\ \epsilon I_n & \end{pmatrix}, \quad N = 2n: \tag{4.3}$$

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \in \mathrm{l}(N,\mathbb{C}) \quad \text{iff} \quad \begin{cases} C_{22} = -C_{11}^T, \\ C_{12} = -\epsilon C_{12}^T, \\ C_{21} = -\epsilon C_{21}^T, \end{cases} \tag{4.4}$$

$C \in \mathrm{Cent}_{\mathrm{o}(N,\mathbb{C})} M_J$ iff

$C =$ 

. (4.5)

Each solid oblique line indicates a parameter repeated from row to row. Each dashed line indicates that the same parameter is repeated along each line but its sign varies from row to row. Depending on the parity of n, C_{12} and C_{21} have a dashed line or a line of zeros on the side diagonal.

If $\epsilon = -1$, C has a form similar to (4.5), but C_{12} and C_{21} start with a coefficient in the corner instead of a zero. The dimension of the centralizer is

$$d = \begin{cases} 2n - \epsilon & \text{if } n \text{ is odd}, \\ 2n & \text{if } n \text{ is even}. \end{cases}$$

(*c*) *Case III*:

$$M_J = \begin{pmatrix} \alpha I_n + J_n & 0 \\ 0 & -\alpha I_n - J_n^T \end{pmatrix},$$

$$K = \begin{pmatrix} & I_n \\ \epsilon I_n & \end{pmatrix}, \quad N = 2n. \tag{4.6}$$

Then

$$C=\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22}\end{pmatrix} \in \mathrm{Cent}_{1(N,\mathbb{C})} M_J$$

iff C is given as in (4.5) with $C_{12}=C_{21}=0$. Therefore, $d=n$.

Next, let us turn to centralizers of decomposable Jordan normal forms of $1(N,\mathbb{C})$-elements. M_J is now a direct sum of blocks of types I, II, and III. K is the direct sum of the corresponding small K's. The centralizer of the general case can be built up in an obvious way out of all possible pairs of blocks as a matrix of matrices. Thus it suffices to describe the centralizers of all possible M_J which are direct sums of two blocks of types I, II, or III.

Notation: Below we describe centralizers as matrices of triangular matrices. An oblique solid line indicates one complex parameter repeated along the line. A dashed line indicates one complex parameter changing sign from row to row.

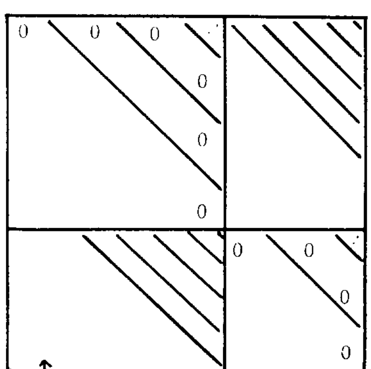
(*d*) *Cases I + III, II + III, III + III' with the eigenvalues of III' distinct from those of III*: In these cases the centralizer is just the direct sum of centralizers of each block.

(*e*) *Case I + I'*:

$$M_J=J_{N_1}\oplus J_{N_2},\quad K=F_{N_1}\oplus F_{N_2}, \tag{4.7}$$

where J_{N_i} and F_{N_i} are as in (2.0) and $N_1 \geqslant N_2$: N_1 and N_2 are simultaneously odd or even depending on $\epsilon=\pm 1$. Then

$$C=\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22}\end{pmatrix} \in \mathrm{Cent}_{1(N_1+N_2,\mathbb{C})} M_J \quad \text{iff}$$

$C=$  (4.8)

$C_{21}=-\epsilon K_2 C_{12}^T K_1$

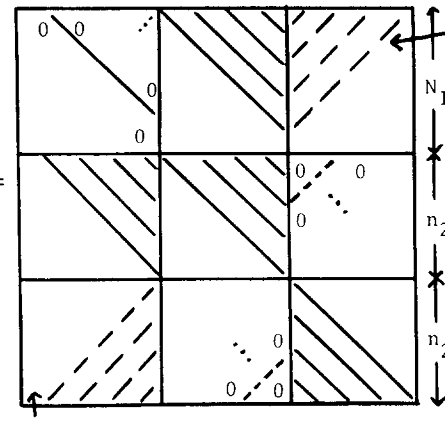
If the parameters appearing in C_{12} are denoted $\lambda_1,...,\lambda_{N_2}$, as they appear in the first row from right to left, then the corresponding parameters in C_{21} are $\lambda_1, -\lambda_2, \lambda_3, ..., \epsilon\lambda_{N_2}$ in the first row from right to left. $N_i=2n_i+1$ or $N_i=2n_i$ depending whether $\epsilon=\pm 1$. Then $d=n_1+n_2+N_2$.

(*f*) *Case I + II*:

$$M_J=J_{N_1}\oplus\begin{pmatrix} J_{n_2} & \\ & -J^T_{n_2}\end{pmatrix},\quad K=F_{N_1}\oplus\begin{pmatrix} 0 & I_{n_2} \\ \epsilon I_{n_2} & \end{pmatrix}, \tag{4.9}$$

$N_2=2n_2$, $N_1=2n_1$, or $N_1=2n_1+1$. Then $C\in\mathrm{Cent}_{1(N_1+N_2,\mathbb{C})}M_J$ iff

$$C=\begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33}\end{pmatrix}$$

$=$  (4.10)

(figure is done for $N_1 \geqslant n_2$ and $\epsilon=+1$). If paremeters appearing in C_{12} (C_{21}) are denoted by $\lambda_1,...,\lambda_{n_2}$ ($\mu_1,...,\mu_{n_2}$) as they appear in the first row from right to left, then the parameters of the last column (first row) of C_{31} (C_{13}) from down to up (left to right) are $-\lambda_1,..., -\lambda_{n_2}$ ($-\epsilon\mu_1,..., -\epsilon\mu_{n_2}$). We have

$$d=\begin{cases} n_1+2n_2+2\min(n_1,n_2) & \text{if } n_2 \text{ is even,}\\ n_1+2n_2-\epsilon+2\min(n_1,n_2) & \text{if } n_2 \text{ is odd.}\end{cases}$$

(*g*) *Case II + II'*:

$$M_J=\begin{pmatrix} J_{n_1} & \\ & -J^T_{n_1}\end{pmatrix}\oplus\begin{pmatrix} J_{n_2} & \\ & -J^T_{n_2}\end{pmatrix},$$

$$K=\begin{pmatrix} 0 & I_{n_1} \\ \epsilon I_{n_1} & 0\end{pmatrix}\oplus\begin{pmatrix} 0 & I_{n_2} \\ \epsilon I_{n_2} & 0\end{pmatrix}, \tag{4.11}$$

$C\in\mathrm{Cent}_{1(2n_1+2n_2,\mathbb{C})}M_J$ iff

$C=(C_{ij})_{i,j=1,...,4}$

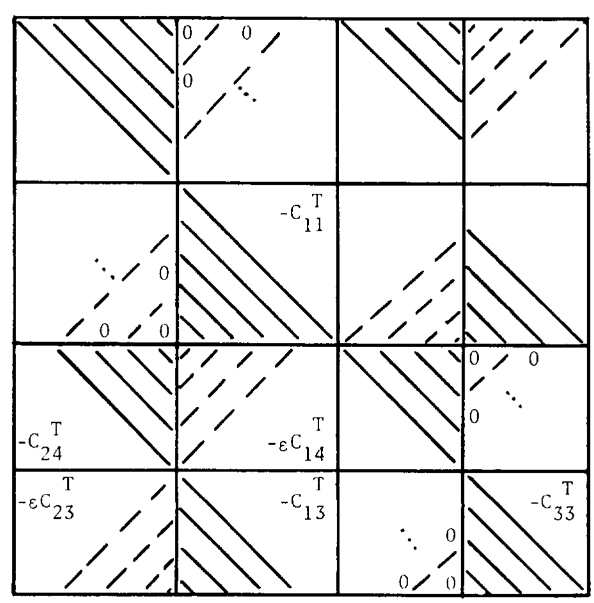
$=$  (4.12)

Figure is done for $\epsilon = +1$ [difference for $C_{12}, C_{21}, C_{34}, C_{43}$ if $\epsilon = -1$ as in (4.5)]. If $\epsilon = -1$, $C_{12}, C_{21}, C_{34}, C_{43}$ start with a parameter in the corner

$$d = \begin{cases} 2n_1 + 6n_2 & \text{if } n_1, n_2 \text{ are even,} \\ 2n_1 + 6n_2 - \epsilon & \text{if } n_1 + n_2 \text{ is odd,} \\ 2n_1 + 6n_2 - 2\epsilon & \text{if } n_1, n_2 \text{ are odd.} \end{cases}$$

(*h*) *Case III + III′ with same eigenvalues*:

$$M_J = \bigoplus_{i=1}^{2} \begin{pmatrix} \alpha I_{n_i} + J_{n_i} & 0 \\ 0 & -\alpha I_{n_i} - J^T_{n_i} \end{pmatrix},$$

$$K = \begin{pmatrix} 0 & I_{n_1} \\ \epsilon I_{n_1} & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & I_{n_2} \\ \epsilon I_{n_2} & 0 \end{pmatrix}, \tag{4.13}$$

$n_1 \geqslant n_2$. Then $C \in \mathrm{Cent}_{\mathrm{I}(2n_1+2n_2,\mathbb{C})} M_J$ iff C is given as in (4.12) with $C_{ij} = 0$ when $i + j$ is odd. Therefore, $d = n_1 + 3n_2$.

B. Minimal versal deformations of elements of $\mathrm{I}(N,\mathbb{C})$

As in Arnold's paper,[1] we give here two minimal versal deformations. The first one is orthogonal to the orbit. The second one has the minimal number of nonzero entries, but, unlike Arnold's case, the number of nonzero entries need not be equal to the number of independent parameters.

Theorem 1: Let $M_J \in \mathrm{I}(N,\mathbb{C})$ be in Jordan normal form. Then $M_A = M_J + C^*(\lambda)$ is a minimal versal deformation of M_J, where $C(\lambda)$ is a generic element of $\mathrm{Cent}_{\mathrm{I}(N,\mathbb{C})} M_J$ depending on d parameters. ($C^* = \overline{C^T}$ is the complex conjugate, transpose of C.) If M_J has s blocks of type I and of order $N_1 \geqslant N_2 \geqslant \cdots N_s$ ($N_i = 2n_i$ or $N_i = 2n_i + 1$), t blocks of type II and of order $2m_1 \geqslant 2m_2 \geqslant \cdots \geqslant 2m_t$, u of these with m_i odd, the rest with m_i even, r distinct pairs of eigenvalues ($\pm \alpha_j$), and for each pair ($\pm \alpha_j$) v_j blocks of type III and of orders $2m_1^j \geqslant 2m_2^j \geqslant \cdots \geqslant 2m_{v_j}^j$, then the dimension of $\mathrm{Cent}_{\mathrm{I}(N,\mathbb{C})} M_J$ is given by

$$d = \sum_{j=1}^{r} \sum_{i=1}^{v_j} (2i-1) m_i^j + 2 \sum_{i=1}^{t} (2i-1) m_i + \sum_{i=1}^{s} n_i + \sum_{i=2}^{s} (i-1) N_i + 2 \sum_{i=1}^{s} \sum_{j=1}^{t} \min(N_i, m_j) - \epsilon u. \tag{4.14}$$

To compute versal deformations with minimal number of nonzero entries, we find the general form of a tangent vector to $\mathrm{Orb}(M_J)$, i.e., a vector $[D, M_J]$, $D \in \mathrm{I}(N,\mathbb{C})$. The general form of an element $[D, M_J]$, is related to the general form of the adjoint C^* of an element $C \in \mathrm{Cent}(M_J)$. Corresponding to each full line (dashed line) of C^*, we have that the sum (alternated sum) of the elements on the line is zero for $[D, M_J]$.

We give here the minimal versal deformations M_A with minimum number of nonzero entries in each of the cases studied in part *A*.

(*a*) *Case I*: $M_J = J_N$, $K = F_N$: Then $M_A = M_J + M_1(\lambda)$ with

$$M_1(\lambda) = \begin{pmatrix} 0 & & & & & & \\ \lambda_1 & & & & & & \\ \vdots & & & & & & \\ 0 & & & & & & \\ \lambda_n & & & & & & \\ 0 & \lambda_n & 0 & \cdots & \lambda_1 & & 0 \end{pmatrix}, \tag{4.15}$$

$\epsilon = +1$, $N = 2n + 1$

$$M_1(\lambda) = \begin{pmatrix} 0 & & & & & \\ & 0 & & & & \\ & & 0 & & & \\ & & \lambda_1 & 0 & & \\ & \lambda_2 & & & & \\ \lambda_n & & & & & 0 \end{pmatrix}, \tag{4.16}$$

$\epsilon = -1$, $N = 2n$.

(*b*) *Case II*: M_J and K given by (4.3): $M_A = M_J + M_2(\lambda)$ with

$$M_2(\lambda) = \left(\begin{array}{cccc|cccc} & & & & & & & -\nu_k \\ & & & & & & & \vdots \\ & & & & & & & 0 \\ & & & & & & & -\nu_1 \\ \lambda_1 & \cdots\cdots & & \lambda_n & \nu_k & \cdots & 0 \quad \nu_1 & 0 \\ \hline 0 & \mu_1 & 0 \cdots & \mu_k & & & & -\lambda_1 \\ -\mu_1 & & & & & & & \\ 0 & & & & & & & \vdots \\ \vdots & & & & & & & \\ -\mu_k & & & & & & & -\lambda_n \end{array}\right), \tag{4.17}$$

$\epsilon = +1$,

$$M_2(\lambda) = \left(\begin{array}{ccc|ccc} & & & & \nu_k & \\ & & & & & \ddots \\ \lambda_1 & \cdots\cdots & \lambda_n & & & \nu_1 \\ \hline \mu_1 & & & & & -\lambda_1 \\ & \ddots & & & & \vdots \\ & & \mu_k & & & -\lambda_n \end{array}\right), \tag{4.18}$$

$\epsilon = -1$.

Depending on the parity of n, μ_k and ν_k are on diagonals or not.

(*c*) *Case III*: M_J and K given in (4.6): Then $M_A = M_J + M_3(\lambda)$ where $M_3(\lambda)$ is obtained from $M_2(\lambda)$ by taking $\nu_i = \xi_i = 0$.

(*d*) *Cases I + III, II + III, III + III' with the eigenvalues of III' distinct from those of III*: direct sums of the previous results.

(*e*) *Case I + I'*: M_J and K are given by (4.7): $M_A = M_J + M(\lambda)$ with

$$M(\lambda) = \left(\begin{array}{c|ccc} M_1(\lambda) & \alpha_1 & \cdots & \alpha_{n_2} \\ \hline -\epsilon\alpha_{n_2} & & & \\ \vdots & & M_1(\mu) & \\ \alpha_2 & & & \\ -\alpha_1 & & & \end{array}\right), \tag{4.19}$$

where $M_1(\lambda)$ and $M_1(\mu)$ are as in (4.15) or (4.16) depending on ϵ.

(*f*) *Case I + II*: M_J *and* K *given by* (*4.9*): $M_A = M_J + M(\lambda)$ with

$$M(\lambda) = \left(\begin{array}{c|c|c} M_1(\lambda) & \xi_1 \ \cdots \ \xi_{N_1} \ 0 \ldots 0 & 0 \ldots 0 \ -\nu_1 \ \cdots \ -\nu_{N_1} \\ \hline \begin{matrix} 0 \\ 0 \\ \nu_1 \\ \vdots \\ \nu_{N_1} \end{matrix} & \multicolumn{2}{c}{} \\ \cline{1-1} \begin{matrix} -\varepsilon\xi_1 \\ \vdots \\ -\varepsilon\xi_{N_1} \\ 0 \\ \vdots \\ 0 \end{matrix} & \multicolumn{2}{c}{M_2(\mu)} \end{array}\right). \tag{4.20}$$

The figure is done for $N_1 < n_2$ unlike (4.10). $M_1(\lambda)$ is as in (4.15) or (4.16); $M_2(\mu)$ is as in (4.17) or (4.18).

(*g*) *Case II + II'*: M_J *and* K *given by* (*4.11*): $M_A = M_J + M(\lambda)$ with

$$M(\lambda) = \left(\begin{array}{cc|c|c} \multicolumn{2}{c|}{} & -\gamma_1 \ \cdots \ -\gamma_{n_2} & -\varepsilon\delta_1 \ \cdots \ -\varepsilon\delta_{n_2} \\ \cline{3-4} \multicolumn{2}{c|}{M_2(\lambda)} & -\varepsilon\xi_1 \ \cdots \ -\varepsilon\xi_{n_2} & -\nu_1 \ \cdots \ -\nu_{n_2} \\ \hline \begin{matrix} \nu_1 \\ \vdots \\ \nu_{n_2} \end{matrix} & \begin{matrix} \delta_1 \\ \vdots \\ \delta_{n_2} \end{matrix} & \multicolumn{2}{c}{} \\ \cline{1-2} \begin{matrix} \xi_1 \\ \vdots \\ \xi_{n_2} \end{matrix} & \begin{matrix} \gamma_1 \\ \vdots \\ \gamma_{n_2} \end{matrix} & \multicolumn{2}{c}{M_2(\mu)} \end{array}\right) \tag{4.21}$$

$M_2(\lambda)$ and $M_2(\mu)$ as in (4.17) or (4.18).

(*h*) *Case III + III' with same eigenvalues*: M_J *and* K *as in* (*4.13*): Then $M_A = M_J + M(\lambda)$ with $M(\lambda)$ obtained from previous case by replacing $M_2(\lambda)$ and $M_2(\mu)$ by $M_3(\lambda)$ and $M_3(\mu)$ and by taking the ξ_i's and δ_i's to be zero.

Remark: As in Arnold,[1] this is one choice of versal deformation with minimal number of nonzero entries. There are other possible choices that the reader may be interested in considering.

Theorem 2: Let $M(\mu_1,\ldots,\mu_m)$ be a holomorphic family of elements of $\mathrm{l}(N, \mathbb{C})$ with $M(0) = M_0$, and let M_J be the normal form of M_0. Let $d = \operatorname{codim} \operatorname{orb}(M_J)$. Then there exist mappings

$$\varphi: \mathbb{C}^m \to \mathbb{C}^d, \quad \varphi(0) = 0, \quad \varphi \text{ holomorphic},$$

$$g: \mathbb{C}^m \to L(N, \mathbb{C}), \quad g \text{ holomorphic},$$

such that $M(\mu) = g(\mu) M_A(\varphi(\mu)) g^{-1}(\mu)$, where M_A is a versal deformation of M_J, either the one given in Theorem 1 or the one described above.

V. BIFURCATION DIAGRAMS

Let us consider families of elements of $\mathrm{l}(N,\mathbb{C})$ in general position (see below for a definition), investigate possible structures of the matrices of each family, and describe how the structure changes when we change the value of the parameters.

Definition: Two matrices of $\mathrm{l}(N,\mathbb{C})$ have the same *structure* or belong to the same *stratum* if their centralizers are conjugate under the action of $L(N,\mathbb{C})$.

It is easy to verify the following proposition:

Proposition: Two matrices of $\mathrm{l}(N,\mathbb{C})$ have the same structure iff their Jordan normal forms have: (i) the same number of pairs of eigenvalues $\pm\alpha_1,\ldots,\pm\alpha_r$ with the α_i distinct and either: $\pm\alpha_i$ are simple nonzero eigenvalues or: $\alpha_i = 0$

TABLE II. Strata in $o(N,\mathbb{C})$ of codimensions $c \leqslant 3$. A stratum is denoted by the product of determinants of irreducible blocks of one of its Jordan matrices.

c	0	1	2	3
	0 $(\pm\alpha)$	0^3 $(\pm\alpha)^2$	$(\pm\alpha)^3$, 0^5, $0^5 0$, $(\pm\alpha)^2(\pm\beta)^2$, $(\pm\alpha)^2 0^3$,	$(\pm\alpha)^3 0^3$, $(\pm\alpha)^2(\pm\beta)^2 0^3$, $(\pm\alpha)^3(\pm\beta)^2$, $(\pm\alpha)^2 0^5$, $(\pm\alpha)^2 0^3 0$, $(\pm\alpha)^2(\pm\beta)^2(\pm\gamma)^2$, 0^7, $0^5 0$, $(\pm\alpha)^4$, $(\pm\alpha)(\pm\alpha)$

and the eigenvalue 0 has multiplicity 2; (ii) the same number of multiple nonzero eigenvalues and for each of these pairs the same blocks of type III; (iii) the same blocks of types I and II when the eigenvalue 0 has multiplicity $\neq 2$.

Each stratum is a semialgebraic manifold (defined by equalities and inequalities). The splitting into strata is a finite semialgebraic stratification of $l(N,\mathbb{C})$. The transversality theorem gives us:

Corollary: The set of families of elements of $l(N,\mathbb{C})$ transversal to all the strata is dense (a countable intersection of dense open sets).

Definition: A family of matrices is in *general position* if it is transversal to all strata. The *bifurcation diagram* is the set of parameter values which correspond to matrices with multiple eigenvalues.

The bifurcation diagram is a finite union of smooth manifolds; each manifold corresponds to a set of matrices with the same structure. The codimension of a manifold is equal to the codimension of the corresponding stratum in $l(N,\mathbb{C})$. We have the property that a k-parameter family of elements of $l(N,\mathbb{C})$ in general position can have only singularities of codimension $\leqslant k$. We are interested to know which structures can occur in a k-parameter family in general position. The codimension of a stratum can be shown to be

$$c = d - \frac{1}{2}\sum_{\alpha\neq 0} 1 - \delta \begin{cases} \delta = 1 \text{ if } 0 \text{ has multiplicity } 2 \\ \text{and } l(N,\mathbb{C}) = o(N,\mathbb{C}) \\ \delta = 0 \text{ otherwise,} \end{cases} \tag{5.1}$$

where the summation extends over all nonzero eigenvalues α, and d is given by (4.14). Clearly a single pair $\pm\alpha$ of eigenvalues does not contribute to the codimension.

In order to describe bundles of low codimension contained in $o(N,\mathbb{C})$, $N = 2,3,\ldots$, we denote a Jordan matrix by the product of determinants of its irreducible blocks. In particular, the orthogonally indecomposable matrices of the types I, II, and III of Table I are denoted respectively by 0^n, $(\pm 0)^n$, and $(\pm\alpha)^n$.

Bundles of codimensions $c \leqslant 3$ in $o(N,\mathbb{C})$ are given in Table II; Table III contains bifurcation diagrams of two-parameter families in general position in $o(N,\mathbb{C})$. Similar results for symplectic algebras are found in Ref. 6.

VI. EXAMPLES

Here we give the Arnold normal forms of all elements of $o(6,\mathbb{C})$. In each case we calculate d, the codimensions of the orbit, and c, the codimension of the stratum. We list first all elements of $o(3,\mathbb{C})$, $o(4,\mathbb{C})$, and $o(5,\mathbb{C})$. One can find their Arnold normal forms as parts of the Arnold normal forms of elements of $o(6,\mathbb{C})$.

Remark: Parameters are called λ_i's when they appear both in the orbit and the stratum and μ_i's when they appear only in the stratum.

Table IV gives a list of all Jordan normal forms of elements of $o(3,\mathbb{C})$, $o(4,\mathbb{C})$, $o(5,\mathbb{C})$, $o(6,\mathbb{C})$.

Versal deformations of elements of o(6,$\mathbb{C}$)

In each case we give d the codimension of the orbit and c the codimension of the bundle.

(1) $0^3 0^3$:

$$\begin{pmatrix} 0 & 1 & 0 & & & \\ \lambda_1 & 0 & 1 & & & \\ 0 & \lambda_1 & 0 & \lambda_3 & \lambda_4 & \lambda_5 \\ -\lambda_5 & & & 0 & 1 & 0 \\ \lambda_4 & & & \lambda_2 & 0 & 1 \\ -\lambda_3 & & & 0 & \lambda_2 & 0 \end{pmatrix},$$

$d = c = 5.$

TABLE III. Bifurcation diagrams of 2-parameter families of matrices in general position in $o(N,\mathbb{C})$.

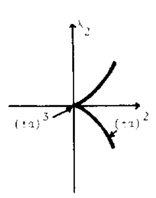
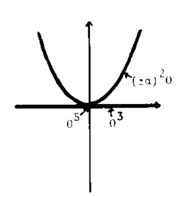
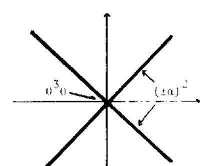
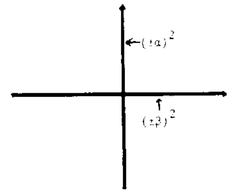
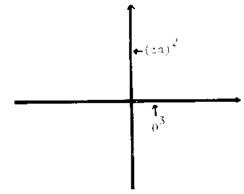
$(\pm\alpha)^3$ $4\lambda_1^3 = 27\lambda_2^2$	0^5 $\lambda_2(\lambda_1^2 - 2\lambda_2) = 0$	$0^3 0$ $(\lambda_1^2 - \lambda_2^2) = 0$	$(\pm\alpha)^2(\pm\beta)^2$ $\lambda_1\lambda_2 = 0$	$(\pm\alpha)^2 0^3$ $\lambda_1\lambda_2 = 0$
				

TABLE IV. Strata in o(3,C), o(4,C), o(5,C), o(6,C).

o(3,C)	o(4,C)	o(5,C)	o(6,C)
0^3, 000 $(\pm\alpha)$ 0	$0^3 0$, 0000 $(\pm\alpha)(\pm\beta)$ $(\pm\alpha)^2$, $(\pm 0)^2$, $(\pm\alpha)(\pm\alpha)$	$0^3 00$, 00000, $(\pm\alpha)000$, $(\pm\alpha)(\pm\beta)0$ $(\pm\alpha)^2 0$, $(\pm 0)^2 0$, 0^5, $(\pm\beta)0^3$, $(\pm\alpha)(\pm\alpha)0$	$0^3 0^3$, $0^3 000$, 000000, $(\pm\alpha)0000$, $(\pm\alpha)(\pm\beta)(\pm\gamma)$, $(\pm\alpha)^2(\pm\beta)$, $(\pm\alpha)(\pm\alpha)(\pm\beta)$ $(\pm\alpha)(\pm 0)^2$, $(\pm\alpha)0^3 0$, $0^5 0$, $(\pm\alpha)^3$, $(\pm\alpha)(\pm\alpha)(\pm\alpha)$ $(\pm\alpha)^2(\pm\alpha)$

The following Jordan normal forms belong to the same strata:
$(\pm\alpha)(\pm\beta)$ and $(\pm\alpha)0\,0$
$(\pm\alpha)(\pm\alpha)(\pm\beta)$ and $(\pm\alpha)(\pm\alpha)$
$(\pm\alpha)(\pm\beta)(\pm\gamma)$ and $(\pm\alpha)(\pm\beta)0\,0$
$(\pm\alpha)^2(\pm\beta)$ and $(\pm\alpha)^2 0\,0$

(2) $0^3 000$:

$$\begin{pmatrix} 0 & 1 & & & & \\ \lambda_1 & 0 & 1 & & & \\ 0 & \lambda_1 & 0 & \lambda_2 & \lambda_3 & \lambda_4 \\ -\lambda_2 & & & 0 & \lambda_5 & \lambda_6 \\ -\lambda_3 & & & -\lambda_5 & 0 & \lambda_7 \\ -\lambda_4 & & & -\lambda_6 & -\lambda_7 & 0 \end{pmatrix},$$

$d = c = 7.$

(3) 000000: We get an arbitrary antisymmetric matrix:

$d = c = 15.$

(4) $(\pm\alpha)0000$:

$$\begin{pmatrix} \alpha+\mu_1 & & & & & \\ & -\alpha-\mu_1 & & & & \\ & & 0 & \lambda_1 & \lambda_2 & \lambda_3 \\ & & -\lambda_1 & 0 & \lambda_4 & \lambda_5 \\ & & -\lambda_2 & -\lambda_4 & 0 & \lambda_6 \\ & & -\lambda_3 & -\lambda_5 & -\lambda_6 & 0 \end{pmatrix},$$

$d = 7, \quad c = 6.$

(5) $(\pm\alpha)(\pm\beta)(\pm\gamma)$:

$$\begin{pmatrix} \alpha+\mu_1 & & & & & \\ & -\alpha-\mu_1 & & & & \\ & & \beta+\mu_2 & & & \\ & & & -\beta-\mu_2 & & \\ & & & & \gamma+\mu_3 & \\ & & & & & -\gamma-\mu_3 \end{pmatrix},$$

$d = 3, \quad c = 0.$

(6) $(\pm\alpha)^2(\pm\beta)$:

$$\begin{pmatrix} \alpha & 1 & & & & \\ \lambda_1 & \alpha+\mu_1 & & & & \\ & & -\alpha & -\lambda_1 & & \\ & & -1 & -\alpha-\mu_1 & & \\ & & & & \beta+\mu_2 & \\ & & & & & -\beta-\mu_2 \end{pmatrix},$$

$d = 3, \quad c = 1.$

(7) $(\pm\alpha)(\pm\alpha)(\pm\beta)$:

$$\begin{pmatrix} \alpha+\mu_1 & & \lambda_2 & & & \\ & -\alpha-\mu_1 & & \lambda_3 & & \\ -\lambda_3 & & \alpha+\lambda_1 & & & \\ & -\lambda_2 & & -\alpha-\lambda_1 & & \\ & & & & \beta+\mu_2 & \\ & & & & & -\beta-\mu_2 \end{pmatrix},$$

$d = 5, \quad c = 3.$

(8) $(\pm\alpha)(\pm 0)^2$:

$$\begin{pmatrix} \alpha+\mu_1 & & & & & \\ & -\alpha-\mu_1 & & & & \\ & & 0 & 1 & 0 & -\lambda_3 \\ & & \lambda_2 & \lambda_1 & \lambda_3 & \\ & & & -\lambda_4 & 0 & -\lambda_2 \\ & & \lambda_4 & & -1 & -\lambda_1 \end{pmatrix},$$

$d = 5, \quad c = 4.$

(9) $(\pm\alpha)0^3 0$:

$$\begin{pmatrix} \alpha+\mu_1 & & & & & \\ & -\alpha-\mu_1 & & & & \\ & & 0 & 1 & & \\ & & \lambda_1 & 0 & 1 & \\ & & & \lambda_1 & 0 & \lambda_2 \\ & & -\lambda_2 & & & 0 \end{pmatrix},$$

$d = 3, \quad c = 2.$

(10) $(\pm 0)^2 00$:

$$\begin{pmatrix} 0 & 1 & 0 & -\lambda_3 & \lambda_8 & \lambda_9 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 & -\lambda_6 & -\lambda_7 \\ 0 & \lambda_4 & 0 & -\lambda_2 & 0 & 0 \\ -\lambda_4 & 0 & -1 & -\lambda_1 & 0 & 0 \\ \lambda_6 & 0 & 0 & -\lambda_8 & 0 & \lambda_5 \\ \lambda_7 & 0 & 0 & -\lambda_9 & -\lambda_5 & 0 \end{pmatrix},$$

$d = c = 9.$

(11) $0^5 0$:

$$\begin{pmatrix} 0 & 1 & & & & \\ \lambda_1 & 0 & 1 & & & \\ 0 & & 0 & 1 & & \\ \lambda_2 & & & 0 & 1 & \\ 0 & \lambda_2 & 0 & \lambda_1 & 0 & \lambda_3 \\ -\lambda_3 & & & & & 0 \end{pmatrix},$$

$d = c = 3.$

(12) $(\pm\alpha)^3$:

$$\begin{pmatrix} \alpha & 1 & & & & \\ & \alpha & 1 & & & \\ \lambda_2 & \lambda_1 & \alpha+\mu_1 & & & \\ & & & -\alpha & & -\lambda_2 \\ & & & -1 & -\alpha & -\lambda_1 \\ & & & & -1 & -\alpha-\mu_1 \end{pmatrix},$$

$d = 3, \quad c = 2.$

(13) $(\pm\alpha)(\pm\alpha)(\pm\alpha)$:

$$\begin{pmatrix} \alpha+\mu_1 & & \lambda_3 & & \lambda_5 & \\ & -\alpha-\mu_1 & & \lambda_4 & & \lambda_6 \\ -\lambda_4 & & \alpha+\lambda_1 & & \lambda_7 & \\ & -\lambda_3 & & -\alpha-\lambda_1 & & \lambda_8 \\ -\lambda_6 & & -\lambda_8 & & \alpha+\lambda_2 & \\ & -\lambda_5 & & -\lambda_7 & & -\alpha-\lambda_2 \end{pmatrix},$$

$d = 9, \quad c = 8.$

(14) $(\pm\alpha)^2(\pm\alpha)$:

$$\begin{pmatrix} \alpha & 1 & & & & \\ \lambda_1 & \alpha+\mu_1 & & & -\lambda_3 & \\ & & -\alpha & -\lambda_1 & & -\lambda_4 \\ & & -1 & -\alpha-\mu_1 & & \\ \lambda_4 & & & & \alpha+\lambda_2 & \\ & & & \lambda_3 & & -\alpha-\lambda_2 \end{pmatrix},$$

$d = 5, \quad c = 4.$

ACKNOWLEDGMENT

We are grateful to D. Djoković, D. Schlomiuk, and Y. Saint-Aubin for helpful discussions.

[1] V. I. Arnold, "Matrices depending on parameters," Russ. Math. Surv. **26**, 29–43 (1971).

[2] H. Zassenhaus, "On a normal form of the orthogonal transformation," Can. Math. Bull. **1**, 31–40, 101–111, 183–191 (1958); G. E. Wall, "On the conjugacy classes in the unitary, symplectic and orthogonal groups," J. Austr. Math. Soc. **3**, 1–62 (1963); I. K. Cikunov, "Structure of isometric transformations of a symplectic or orthogonal vector space," Ukr. Zh. **18**, 79–93, 122–127 (1966); T. A. Springer and R. Sternberg, *Conjugacy Classes*, Lecture Notes in Math. Vol. 131 (Springer, New York, 1970), pp. 167–266; J. Milnor, "On isometries of inner product spaces," Invent. Math. **8**, 83–97 (1969); N. Burgoyne and R. Cushman, "Normal forms for linear Hamiltonian systems," Celest. Mech. **8**, 435–443 (1974); J. Williamson, "On the algebraic problem concerning the normal forms of linear dynamical systems," Am. J. Math. **58**, 141–163 (1936); A. J. Laub and K. Meyer, Celest. Mech. **9**, 2134–238 (1974); J. G. F. Belifante and P. Winternitz, "One parameter subgroups of unitary groups with indefinite metric, and in particular of the conformal group," J. Math. Phys. **12**, 1041–1054 (1971).

[3] N. Burgoyne and R. Cushman, J. Alg. **44**, 339–362 (1977).

[4] D. Djoković, J. Patera, P. Winternitz, and H. Zassenhaus, "Normal forms of elements of classical real and complex Lie and Jordan algebras " (to be submitted to J. Math. Phys.).

[5] D. M. Galin, "On real matrices depending on parameters," Usp. Math. Nauk. **27**, 241–242 (1972) (in Russian).

[6] D. M. Galin, "Versal deformations of linear Hamiltonian systems," Trudy Semin. I. G. Petrovsky **1**, 63–73 (1975) (in Russian).

[7] M. Moshinsky and P. Winternitz, "Quadratic Hamiltonians in phase space and their eigenstates," J. Math. Phys. **21**, 1667 (1980).

[8] L. O'Raighfeartaigh, Rep. Prog. Phys. **42**, 160 (1979); L. Michel, Rev. Mod. Phys. **52**, 617 (1980); R. Slansky, Phys. Rep. 1981 (to be published).

A theorem on A-proper mappings and its application in scattering theory

Helmut Kröger
Institut für Theoretische Physik, Universität Giessen, D-6300 Giessen, West Germany
Rainer Perne
Gerhard Rohlfstr. 22, D-5320 Bad Godesberg, West Germany

(Received 3 December 1980; accepted for publication 18 March 1981)

We propose a projectionally complete scheme yielding an approximate solution of the functional equation $Bf = g$ in a Hilbert space. We prove that B is an A-proper mapping. The result is applied to an integral equation with a kernel appearing in multichannel scattering theory.

PACS numbers: 02.30.Ks, 02.30.Mv, 02.30.Rz, 11.80.Gw

I. INTRODUCTION

Multichannel scattering integral equations of the structure

$$f = g + Af, \tag{1.1}$$

defined in a Hilbert space $\mathscr{H}$, with $g \in \mathscr{H}$ and A being a linear mapping from $\mathscr{H}$ into $\mathscr{H}$, have been solved successfully using the moments method[1] and the Padé method.[2,3] The moments method solves the projector equation

$$P^{(n)}f^{(n)} = P^{(n)}g + P^{(n)}AP^{(n)}f^{(n)}, \tag{1.2}$$

where $P^{(n)}$ is the orthogonal projection from $\mathscr{H}$ on the subspace

$$T^{(n)}_{A,g} = \text{linear span of } \{A^k g \,|\, k = 0,1,...,n\}. \tag{1.3}$$

The solution $f^{(n)}$ converges strongly to f (Ref. 4) if A is compact, $1 \in \rho(A)$, and $f \in T_{A,g}$, where

$$T_{A,g} = \overline{\bigcup_{n=0,1,\dots} T^{(n)}_{A,g}}. \tag{1.4}$$

The Padé method calculates (h,f) with $h \in \mathscr{H}$ by solving a projector equation like Eq. (1.2), but with $P^{(n)}$ now given by

$$T^{(n)}_{A^+,h} = \text{linear span of } \{A^{+k} h \,|\, k = 0,1,...,n\}, \tag{1.5}$$

$$P^{(n)}|_{T^{(n)}_{A,g}} = 1, \tag{1.6}$$

$$P^{(n)+}|_{T^{(n)}_{A^+,h}} = 1. \tag{1.7}$$

Recently, integral equations for multichannel scattering theory have been given by Chandler and Gibson[5] and Kröger and Perne[6] with kernels of a simple structure in contrast to the Faddeev-type equations. But these kernels are not connected and hence not compact. Thus it is not guaranteed that they can be approximated separably. Nevertheless one can try to solve the equations approximately using projector schemes. There are methods, introduced by Petryshyn,[7] which solve functional equations with so called A-proper mappings, which are more general than compact mappings.

In Sec. II we propose a projectionally complete scheme to solve $Bf = g$. We take the linear space spanned by g, Bg, $B^2g, \cdots$ as projectional subspace. We prove that the projectional solution converges strongly to the solution f and that B is A-proper.

In Sec. III we investigate an integral equation with a nonconnected kernel of the type appearing in multichannel scattering equations and show that the previous results can be applied.

II. THEOREM ON A-PROPER MAPPINGS

Let us start with the definition of A-proper mappings and projectionally complete schemes given in Ref. 7. Let X, Y be Banach spaces, D a given subset of X, $T: D \subseteq X \to Y$ a possibly nonlinear mapping and $\Gamma = \{X_n, P_n; Y_n, Q_n\}$ a suitable approximation scheme for the equation

$$Tx = y, \ (x \in D, y \in Y). \tag{2.1}$$

The scheme Γ is called projectionally complete for (X,Y) provided that $\{X_n\} \subset X$ and $\{Y_n\} \subset Y$ are sequences of monotonically increasing finite-dimensional subspaces with $\dim X_n = \dim Y_n$ for each n and $P_n: X \to X_n$ and $Q_n: Y \to Y_n$ are linear projections such that $P_n x \to x$ and $Q_n y \to y$ for $x \in X$ and $y \in Y$. Here and in the following $\to$ denotes strong convergence. Let $D_n = D \cap X_n$, $T_n: D_n \to Y_n$, $T_n = Q_n T|_{D_n}$ and x_n be the solution of the approximate equation

$$T_n x_n = Q_n y. \tag{2.2}$$

The mapping T is said to be A-proper with respect to the projectionally complete scheme Γ if T_n is continuous for each n and if $\{u_n | u_n \in D_n\}$ is a bounded sequence such that $T_n u_n \to v$ for some v in Y, then there exists a subsequence $\{u_{n_\nu}\}$ and $u \in D$ such that $u_{n_\nu} \to u$ and $Tu = v$. Now we can prove the following result.

Theorem: Let $\mathscr{H}$ be a Hilbert space, let g be an element of $\mathscr{H}$, let $B: \mathscr{H} \to \mathscr{H}$ be a linear, bounded mapping, and assume the origin 0 to lie in the resolvent set $\rho(B)$. We define $T^{(n)}_{B,g}, T^{(n)}_{B,Bg}, T_{B,g}, T_{B,Bg}$ in analogy to Eqs. (1.3) and (1.4). Let $P^{(n)}$ be the orthogonal projection onto $T^{(n)}_{B,g}$ and let $Q^{(n)}$ be the orthogonal projection onto $T^{(n)}_{B,Bg}$. Let f be defined as the solution of

$$Bf = g, \tag{2.3}$$

and assume f to be an element of $T_{B,g}$. Let $f^{(n)}$ be defined as the solution of

$$Q^{(n)}BP^{(n)}f^{(n)} = Q^{(n)}g. \tag{2.4}$$

Then we claim that $P^{(n)}f^{(n)}$ converges strongly to f, the approximation scheme is projectionally complete, and B is an A-proper mapping.

Proof: Without loss of generality let $f^{(n)} \in T^{(n)}_{B,g}$, thus $Bf^{(n)} \in T^{(n)}_{B,Bg}$ and $f^{(n)}$ fulfills

$$Bf^{(n)} = Q^{(n)}g. \tag{2.5}$$

The definition of $Q^{(n)}$ implies the existence of $f^{(n)}$. The assumption $f \in T_{B,g}$ implies the existence of a sequence $\tilde{f}^{(n)} \in T^{(n)}_{B,g}$, such that

$$\tilde{f}^{(n)} \underset{n}{\rightarrow} f. \tag{2.6}$$

The boundedness of B and Eq. (2.3) imply

$$B\tilde{f}^{(n)} \underset{n}{\rightarrow} g. \tag{2.7}$$

From $B\tilde{f}^{(n)} \in T^{(n)}_{B,Bg}$ and $\|Q^{(n)}\| = 1$ it follows that

$$B\tilde{f}^{(n)} - Q^{(n)}g \underset{n}{\rightarrow} 0. \tag{2.8}$$

Thus Eqs. (2.7) and (2.8) yield

$$Q^{(n)}g \underset{n}{\rightarrow} g, \tag{2.9}$$

while Eqs. (2.3) and (2.5) yield

$$B(f - f^{(n)}) = g - Q^{(n)}g. \tag{2.10}$$

The property $0 \in \rho(B)$ implies the existence and boundedness of B^{-1}, thus follows from Eqs. (2.9) and (2.10)

$$f^{(n)} \underset{n}{\rightarrow} f. \tag{2.11}$$

It remains to show the projectional completeness of the approximation scheme and the A-properness of the mapping. First, we prove the following relation:

$$T_{B,g} = T_{B,Bg}. \tag{2.12}$$

The implication $\supseteq$ is trivial; now consider $\subseteq$: Eq. (2.7) implies the existence of a sequence $B\tilde{f}^{(n)} \in T^{(n)}_{B,Bg} \subseteq T_{B,Bg}$, which strongly converges to g.
$T_{B,Bg}$ is closed, hence $g \in T_{B,Bg}$. For each $u \in T_{B,g}$ there exists a sequence $u^{(n)} \in T^{(n)}_{B,g}$

$$u^{(n)} \rightarrow u, \tag{2.13}$$

which can be written as

$$u^{(n)} = u_0^{(n)}g + \sum_{i=1}^{n} u_i^{(n)} B^i g. \tag{2.14}$$

One has

$u_0^{(n)} g \in T_{B,Bg}$, $\sum_{i=1}^{n} u_i^{(n)} B^i_{\ g} \in T^{(n-1)}_{B,Bg} \subseteq T_{B,Bg}$, hence $u^{(n)} \in T_{B,Bg}$. Because of Eq. (2.13) and $T_{B,Bg}$ being closed, one has $u \in T_{B,Bg}$, which proves Eq. (2.12). $T_{B,g}$ is a Hilbert space.

In order to establish the correspondence to the general definition of a projectionally complete approximation scheme and an A-proper mapping we define

$$\begin{aligned} &X = Y = D = T_{B,g}, \quad T = B\,|_{T_{B,g}}, \\ &X_n = T^{(n)}_{B,g}, \quad P_n = P^{(n)}|_{T_{B,g}}, \\ &Y_n = T^{(n)}_{B,Bg}, \quad Q_n = Q^{(n)}|_{T_{B,g}}, \\ &D_n = T^{(n)}_{B,g} \quad T_n = Q^{(n)}|_{T_{B,g}} B\,|_{T^{(n)}_{B,g}}. \end{aligned} \tag{2.15}$$

Equation (2.3) reads

$$B\,|_{T_{B,g}} f = g, \tag{2.16}$$

and Eq. (2.4) reads

$$Q^{(n)}|_{T_{B,g}} B\,|_{T^{(n)}_{B,g}} f^{(n)} = Q^{(n)}|_{T_{B,g}} g. \tag{2.17}$$

$T^{(n)}_{B,g}$ and $T^{(n)}_{B,Bg}$ are finite-dimensional, monotonically increasing subspaces of the same dimension. Because of Eq. (2.12) one has for each $h \in T_{B,g}$

$$P^{(n)}h \underset{n}{\rightarrow} h, \; Q^{(n)}h \underset{n}{\rightarrow} h, \tag{2.18}$$

$$Q^{(n)}|_{T_{B,g}} B\,|_{T^{(n)}_{B,g}}$$

is continuous for each n, because $Q^{(n)}$ and B are bounded mappings.

Let $\{u^{(n)} | u^{(n)} \in T^{(n)}_{B,g}\} \in T^{(n)}_{B,g}$ be a bounded sequence such that

$$Q^{(n)}|_{T_{B,g}} B_{T^{(n)}_{B,g}} u^{(n)} \underset{n}{\rightarrow} v \tag{2.19}$$

for some $v \in T_{B,g}$. Equation (2.19) can be written as

$$Bu^{(n)} \underset{n}{\rightarrow} v. \tag{2.20}$$

Because $0 \in \rho(B)$, B^{-1} exists and is bounded,

$$u^{(n)} \underset{n}{\rightarrow} B^{-1}v = w, \tag{2.21}$$

with $w \in T_{B,g}$, hence

$$B\,|_{T_{B,g}} w = v, \tag{2.22}$$

which completes the proof of the theorem. This yields the following corollary.

Corollary 1: Let $\mathscr{H}$ be a Hilbert space, $g \in \mathscr{H}$, let A: $\mathscr{H} \rightarrow \mathscr{H}$ be a linear, bounded mapping with $z_0 \in \rho(A)$. We define

$$\begin{aligned} &T^{(n)}_{A,g} = T^{(n)}_{z_0 - A,g}, \; T_{A,g} = T_{z_0 - A,g}, \\ &T^{(n)}_{z_0 - A,(z_0 - A)g}, T_{z_0 - A,(z_0 - A)g} \end{aligned}$$

in analogy to Eqs. (1.3) and (1.4). Let $P^{(n)}, Q^{(n)}$ be the orthogonal projections onto $T^{(n)}_{A,g}$, $T^{(n)}_{z_0 - A,(z_0 - A)g}$, respectively. Let f, $f^{(n)}$ be defined as the solutions of

$$(z_0 - A)f = g, \tag{2.23}$$

and

$$Q^{(n)}(z_0 - A)P^{(n)}f^{(n)} = Q^{(n)}g, \tag{2.24}$$

respectively. We assume $f \in T_{A,g}$. Then $f^{(n)}$ converges to f in $\mathscr{H}$. The proof follows from $B = z_0 - A$ and the above theorem.

Corollary 2: We assume the conditions of Corollary 1, except $f \in T_{A,g}$. Instead we assume that the resolvent set $\rho(A)$ has a specific form, namely such that there is a tongue-shaped extension from the region outside the spectral radius $r(A)$ to the point z_0 (Fig. 1). Then the result of Corollary 1 holds.

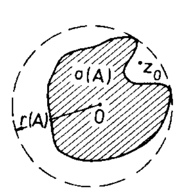


FIG. 1. Schematic plot of the spectrum $\sigma(A)$ with a tongue shaped extension from the region with $|z| > r(A)$ to the point z_0.

Proof: We will show $f \in T_{A,g}$, which completes the conditions of Corollary 1. The function $F(z) = (z - A)^{-1} g$ is an analytical function for $z \in \rho(A)$, i.e., in particular for $|z| > r(A)$. For $|z| > r(A)$ the Neumann series converges,

$$\sum_{k=0,1}^{n} \frac{1}{z}\left(\frac{A}{z}\right)^k g \underset{n}{\rightarrow} (z - A)^{-1} g. \tag{2.25}$$

Thus the property $F(z) \in T_{A,g}$ holds for $|z| > r(A)$, which implies $F(z)/T_{A,g} = 0$ for $|z| > r(A)$. $F(z)$ is analytical for $z \in \rho(A)$ and so is $F(z)/T_{A,g}$. Thus the analytical continuation of $F(z)/T_{A,g}$ from the region $|z| > r(A)$ to $z = z_0$ yields $F(z_0)/T_{A,g} = 0$ and thus $f = F(z_0) \in T_{A,g}$.

III. APPLICATION TO SCATTERING EQUATIONS

Transition operators $T^{(N)}$, which describe the amplitudes for N-nucleon scattering processes, obey Faddeev-type coupled integral equations of the form

$$T^{(N)} = I^{(N)} + K^{(N)}(T^{(N-1)}) T^{(N)}. \tag{3.1}$$

The kernel $K^{(N)}$ depends on $(N-1)$-nucleon transition operators $T^{(N-1)}$. The kernels $K^{(N)}$ are disconnected; that means, the integration is not performed in all variables. An example for an integral equation with a disconnected kernel is

$$f(x,y) = g(x,y) + \int_a^b dx'\, k(x,y,x') f(x',y). \tag{3.2}$$

This kernel leads to a noncompact operator. One aims to set up integral equations for $T^{(N)}$ with connected kernels, which is usually possible by iteration of Eq. (3.1). Connectedness of the kernels and adequate treatment of their inherent singularities leads to compact kernels, which allows an approximate solution of the integral equation by some standard techniques like separable approximation of the kernels, for example.

In Refs. 5 and 6 scattering integral equations have been given, which have the advantage of simply structured kernels, compared with the kernels of Faddeev-type equations. But these kernels are disconnected and cannot become connected by iteration.

We want to show now for the three-nucleon case that a disconnected kernel, which is a typical part of the kernel of integral equations of Ref. 6, is A-proper. That means that an integral equation with only that kernel is approximately solvable by projector methods. We introduce some notation. The center of mass motion is dropped. The index α denotes a particle and the two-body subsystem, which does not contain the particle α. $\mathbf{p}_\alpha$ is the relative momentum between particle α and the c.m. of subsystem α, $\mathbf{q}_\alpha$ is the relative momentum between the particles in the subsystem α, m_α is the reduced mass of particle α and the subsystem α, and μ_α is the reduced mass of the particles of subsystem α. $|\mathbf{p},\mathbf{q}\rangle_\alpha$ denotes a plane wave state.

$$H_0 = p_\alpha^2/2m_\alpha + q_\alpha^2/2\mu_\alpha \tag{3.3}$$

is the Hamiltonian of the free motion in momentum representation, $G_0(z) = (z - H_0)^{-1}$ is the corresponding Green's function and V_α is the two-body potential in subsytem α given by

$${}_\alpha\langle \mathbf{p}',\mathbf{q}'|V_\alpha|\mathbf{p},\mathbf{q}\rangle_\alpha = \delta(\mathbf{p}' - \mathbf{p})\langle \mathbf{q}'|V^{(2)}|\mathbf{q}\rangle, \tag{3.4}$$

where $V^{(2)}$ is a Hermitian potential in the two-body space, which we assume to be a rotational invariant.

The kernel we want to consider is

$$K = V_\alpha G_0(E + i0), \tag{3.5}$$

for a positive energy E. K, applied to a state f, reads in momentum-angular momentum representation

$$\begin{aligned}
&{}_\alpha\langle pLM,qlm|K|f\rangle_\alpha \\
&= \lim_{\epsilon\to+0}\int_0^\infty dp'\,p'^2\int_0^\infty dq'\,q'^2 \sum_{L'M'}\sum_{l'm'} \frac{\delta(p-p')}{p^2}\,\delta_{LM,L'M'}\,\delta_{lm,l'm'}\,\frac{\langle qlm|V^{(2)}|q'l'm'\rangle}{E + i\epsilon - \dfrac{p'^2}{2m_\alpha} - \dfrac{q'^2}{2\mu_\alpha}}\;{}_\alpha\langle p'L'M',q'l'm'|f\rangle_\alpha \\
&= \lim_{\epsilon\to+0}\int_0^\infty dq'q'^2 \frac{V_l(q,q')}{E + i\epsilon - p'^2/2m_\alpha - q'^2/2\mu_\alpha} f(pLM,q'lm),
\end{aligned} \tag{3.6}$$

where

$$V_l(q,q') = \langle qlm|V^{(2)}|q'lm\rangle \tag{3.7}$$

is independent of the quantum number m because of rotational invariance. In the following we omit the angular momentum quantum numbers and put $2m_\alpha = 2\mu_\alpha = 1$, without loss of generality and assume $0 < E < 1$. The singularity of the kernel can occur only for $0 \leqslant p, q' \leqslant 1$. For sake of technical simplicity we modify K by restricting the variables to $0 \leqslant p$, q, and $q' \leqslant 1$, but the same kind of result also holds without this restriction. Thus we finally consider the kernel

$$(Kf)(p,q) = \lim_{\epsilon\to+0}\int_0^1 dq'\,q'^2 \frac{V(q,q')}{E + i\epsilon - p^2 - q'^2} f(p,q'). \tag{3.8}$$

We want to show now that there is a Hilbert space $\mathscr{H}$, such that under some smoothness conditions on V, $K: \mathscr{H} \to \mathscr{H}$ becomes a linear, bounded operator, with the spectrum $\sigma(K)$ having the form required in the assumption of Corollary 2. Let $e(p) = E - p^2$, $s(p) = |e(p)|^{1/2}$ and define $\mathscr{H}$ by

$$L_2 = \left\{f \,\middle|\, \int_0^1 dp\,p^2 \int_0^1 dq\,q^2\,|f(p,q)|^2 \text{exists}\right\},\quad \mathscr{H} = \left\{f \,\middle|\, f(p,q) \in L_2, \phi^f(p,q) = \frac{f(p,q) - f(p,s(p))}{q^2 - e(p)} \in L_2\right\}. \tag{3.9}$$

A scalar product is introduced on $\mathscr{H}$.

$$(f,g)_{\mathscr{H}} = (f,g)_{L_2} + (\phi^f,\phi^g)_{L_2}. \tag{3.10}$$

We claim that $\mathscr{H},(.,.)_{\mathscr{H}}$ is a Hilbert space.

The completeness remains to be shown. Let $\{f_n\}$ be a Cauchy sequence in $\mathscr{H}$, then $\{f_n\}$ is a Cauchy sequence in L_2 and there is a limit element $f \in L_2$,

$$f_n \underset{n}{\rightarrow} f. \tag{3.11}$$

Also ϕ^{f_n} is a Cauchy sequence in L_2 and there is a limit element $g \in L_2$,

$$\phi^{f_n} \underset{n}{\rightarrow} g. \tag{3.12}$$

$\mathscr{H}$ is complete if $\phi^f = g$.

(i) First, we claim that $\{f_n(p,s(p))\}$ is a Cauchy sequence in

$$L_2(p) = \left\{ f \,\Big|\, \int_0^1 dp\, p^2\, |f(p)|^2 \text{ exists} \right\}, \tag{3.13}$$

because

$$\begin{aligned} \int_0^1 dp\, p^2 |f_n[p,s(p)] - f_m[p,s(p)]|^2 &= \frac{1}{3}\int_0^1 dp\, p^2 \int_0^1 dq\, q^2 |f_n[p,s(p)] - f_m[p,s(p)]|^2 \\ &= \frac{1}{3}\int_0^1 dp\, p^2 \int_0^1 dq\, q^2 |[\phi^{f_n}(p,q) - \phi^{f_m}(p,q)] \\ &\quad [e(p) - q^2] + f_n(p,q) - f_m(p,q)|^2 \end{aligned} \tag{3.14}$$

tends to zero with $m,n \to \infty$ as $e(p) - q^2$ is bounded, and because of Eqs. (3.11) and (3.12). $L_2(p)$ is complete, hence there exists a limit element $F \in L_2(p)$,

$$f_n(p,s(p)) \underset{n}{\overset{L_2(p)}{\longrightarrow}} F(p). \tag{3.15}$$

(ii) The set $\{(p,q) | q = s(p)\}$ is a subset of measure 0 of the integration domain $\{(p,q) | 0 \leqslant p,q \leqslant 1\}$. We modify f on this set of measure 0, putting

$$f[p,s(p)] = F(p), \tag{3.16}$$

which does not change $f \in L_2$.

(iii) Now we can prove the relation

$$f(p,q) - f[p,s(p)] - g(p,q)[q^2 - e(p)] = 0 \tag{3.17}$$

almost everywhere, which follows from

$$\begin{aligned} \int_0^1 dp\, p^2 \int_0^1 dq\, q^2\, |f(p,q) - f[p,s(p)] - g(p,q)[q^2 - e(p)]|^2 &= \int_0^1 dp p^2 \int_0^1 dq q^2 |f(p,q) - f_n(p,q) - \{f[p,s(p)] \\ &\quad - f_n[p,s(p)]\} + [\phi^{f_n}(p,q) - g(p,q)][q^2 - e(p)]|^2 \end{aligned} \tag{3.18}$$

which can be estimated by the triangle inequality. The first term tends to zero because of Eq. (3.11), the second term tends to zero because of Eqs (3.15) and (3.16), and the last term tends to zero because of Eq. (3.12). Equation (3.18) implies

$$\frac{f(p,q) - f[p,s(p)]}{q^2 - e(p)} = \phi^f(p,q) = g(p,q) \in L_2 \tag{3.19}$$

almost everywhere, thus proving the completeness of $\mathscr{H}$.

Now we want to show that K, defined by Eq. (3.8), becomes a linear, bounded operator mapping $\mathscr{H}$ into $\mathscr{H}$, provided that $V(q,q')$ is a sufficiently smooth function. Let $V(q,q')$ be a once continuously differentiable function in the domain $0 \leqslant q,q' \leqslant 1$, and let $f \in \mathscr{H}$. Thus one can write the right hand side of Eq. (3.8)

$$\begin{aligned} &\lim_{\epsilon \to +0} \int_0^1 dq'\, q'^2 \frac{V(q,q')}{e(p) + i\epsilon - q'^2} f(p,q') \\ &= \int_0^1 dq'\, q'^2 \frac{V(q,q') - V[q,s(p)]}{e(p) - q'^2} f(p,q') \\ &+ V[q,s(p)] \int_0^1 dq'\, q'^2 \frac{f(p,q') - f[p,s(p)]}{e(p) - q'^2} \\ &+ V[q,s(p)] f[p,s(p)] \lim_{\epsilon \to +0} \int_0^1 dq'\, q'^2 \frac{1}{e(p) + i\epsilon - q'^2}. \end{aligned} \tag{3.20}$$

We abbreviate

$$\phi^V(p,q,q') = \frac{V(q,q') - V[q,s(p)]}{q'^2 - e(p)}, \tag{3.21}$$

$$\psi^V(p,q,q') = \frac{V(q,q') - V[q,s(p)] - V[s(p),q'] + V[s(p),s(p)]}{[q^2 - e(p)][q'^2 - e(p)]}. \tag{3.22}$$

From the smoothness property of $V(q,q')$ there follows the existence of an upper bound M, such that for $0 \leqslant p,q$ and $q' \leqslant 1$,

$$|V(q,q')| \leqslant M, \tag{3.23}$$

$$|\phi^V(p,q,q')| \leqslant M, \tag{3.24}$$

$$|\psi^V(p,q,q')| \leqslant M. \tag{3.25}$$

Now the first term on the right-hand side of Eq. (3.20),

$$(K_1 f)(p,q) = -\int_0^1 dq'\, q'^2 \phi^V(p,q,q') f(p,q'), \tag{3.26}$$

can be estimated as

$$\begin{aligned} \|K_1 f\|_{L_2}^2 &= \int_0^1 dp\, p^2 \int_0^1 dq\, q^2 \left|\int dq'\, q'^2 \phi^V(p,q,q') f(p,q')\right|^2 \\ &\leqslant \int_0^1 dp\, p^2 \int_0^1 dq q^2 \left(\int_0^1 dq' q'^2 |\phi^V(p,q,q')|^2\right) \\ &\quad \times \left(\int_0^1 dq' q'^2 |f(p,q')|^2\right) \\ &\leqslant \frac{M^2}{9} \int_0^1 dp\, p^2 \int_0^1 dq\, q^2 |f(p,q)|^2 = \frac{M^2}{9} \|f\|^2_{L_2}. \end{aligned} \tag{3.27}$$

Similarly one has

$$\begin{aligned} \|\phi^{K_1 f}\|_{L_2}^2 &= \int_0^1 dp\, p^2 \int_0^1 dq\, q^2 \left| \int_0^1 dq' q'^2 \psi^V(p,q,q') f(p,q')\right|^2 \\ &\leqslant \frac{1}{9} M^2 \|f\|^2_{L_2} \end{aligned} \tag{3.28}$$

and thus

$$\|K_1 f\|_{\mathscr{H}}^2 = \|K_1 f\|_{L_2}^2 + \|\phi^{K_1 f}\|_{L_2}^2 \leqslant \frac{2M^2}{9} \|f\|_{L_2}^2 \leqslant \frac{2M^2}{9} \|f\|_{\mathscr{H}}^2. \tag{3.29}$$

The second term on the right-hand side of Eq. (3.20) reads

$$(K_2 f)(p,q) = -V[q,s(p)]\int_0^1 dq'\, q'^2 \phi^f(p,q'). \tag{3.30}$$

The following estimates hold

$$\|K_2 f\|^2_{L_2} = \int_0^1 dp\, p^2 \int_0^1 dq\, q^2 |V[q,s(p)]\int_0^1 dq'\, q'^2\phi^f(p,q')|^2 \leqslant \tfrac{1}{3}M^2\|\phi_f\|^2_{L_2}, \tag{3.31}$$

$$\|\phi^{K_2 f}\|^2_{L_2} = \int_0^1 dp\, p^2 \int_0^1 dq\, q^2 |-\phi^V[p,s(p),q]^* \int_0^1 dq'\, q'^2 \phi^f(p,q')|^2 \leqslant \tfrac{1}{3}M^2\|\phi^f\|^2_{L_2}, \tag{3.32}$$

and hence

$$\|K_2 f\|^2_{\mathscr{H}} \leqslant \tfrac{2}{3}M^2\|f\|^2_{\mathscr{H}}. \tag{3.33}$$

The last term on the right-hand side of Eq. (3.20) reads

$$(K_3 f)(p,q) = V[q,s(p)]f[p,s(p)]c(p), \tag{3.34}$$

where

$$c(p) = \lim_{\epsilon \to +0}\int_0^1 dq\, q^2 \frac{1}{e(p) + i\epsilon - q^2} \tag{3.35}$$

exists for all p in $0 \leqslant p \leqslant 1$ and $c(p)$ is bounded,

$$|c(p)| \leqslant N. \tag{3.36}$$

$K_3 f$ can be estimated as

$$\begin{aligned}\|K_3 f\|^2_{L_2} &= \int_0^1 dp\, p^2 \int_0^1 dq\, q^2 |V[q,s(p)]f[p,s(p)]c(p)|^2 \\ &\leqslant M^2N^2\int_0^1 dp\, p^2 \int_0^1 dq\, q^2 |f[p,s(p)]|^2 \\ &= M^2N^2\int_0^1 dp\, p^2 \int_0^1 dq\, q^2 |\phi^f(p,q)[e(p)-q^2] + f(p,q)|^2 \\ &\leqslant M^2N^2(2\|\phi^f\|_{L_2} + \|f\|_{L_2})^2 \leqslant 9M^2N^2\|f\|^2_{\mathscr{H}},\end{aligned} \tag{3.37}$$

where $|e(p) - q^2| \leqslant 2$ has been used. Analogously we have

$$\|\phi^{K_3 f}\|^2_{L_2} = \int_0^1 dp\, p^2 \int_0^1 dq\, q^2 |\phi^V[p,s(p),q]^* f[p,s(p)]c(p)|^2 \leqslant 9M^2N^2\|f\|^2_{\mathscr{H}}, \tag{3.38}$$

hence

$$\|K_3 f\|^2_{\mathscr{H}} \leqslant 18M^2N^2\|f\|^2_{\mathscr{H}}. \tag{3.39}$$

From Eqs. (3.29), (3.33), and (3.39) one concludes that range $(K) \subset \mathscr{H}$ and K is bounded.

Finally we investigate the spectrum $\sigma(K)$. We define for a fixed p, $0 \leqslant p \leqslant 1$,

$$\mathscr{H}_p = \left\{f \,\Big|\, \int_0^1 dq\, q^2|f(q)|^2 \right\} \text{exists},$$

$$\int_0^1 dq\, q^2 \left|\frac{f(q) - f[s(p)]}{q^2 - e(p)}\right|^2 \text{exists}, \tag{3.40}$$

and K_p is defined on $\mathscr{H}_p$

$$(K_p f)(q) = \lim_{\epsilon \to +0}\int_0^1 dq' q'^2 \frac{V(q,q')}{e(p) + i\epsilon - q'^2} f(q'). \tag{3.41}$$

It has been shown in Ref. 8 that $\mathscr{H}_p$ is a Hilbert space and K_p a compact mapping from $\mathscr{H}_p$ into $\mathscr{H}_p$. The spectrum of a compact operator consists only of the point spectrum and possibly the origin. Let λ_ν be the eigenvalues and φ_ν be the eigenvectors of K_p,

$$K_p \varphi_\nu = \lambda_\nu \varphi_\nu. \tag{3.42}$$

Of course λ_ν and φ_ν are dependent of p. Under variation of p the eigenvalues run along the so-called Weinberg trajectories[9] (Fig. 2). Thus the spectrum $\sigma(K)$ is given by

$$\sigma(K) = \{\lambda_\nu[e(p)] \,|\, \nu = 1,2,\ldots;\ 0 \leqslant p \leqslant 1\} \cup [0]. \tag{3.43}$$

Thus the spectrum $\sigma(K)$ has a shape, which allows a tongue-shaped extension from the region $|z| \geqslant r(K)$ into the region inside the spectral radius and fulfills for those values z_0 the assumption of Corallary 2, and hence K is A-proper.

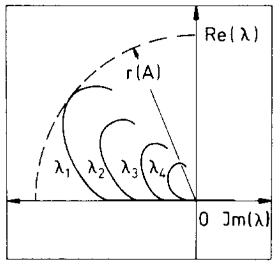


FIG. 2. Schematic plot of the eigenvalues $\lambda_\nu(p)$ [Weinberg trajectories given by Eq. (3.42) for an attractive potential].

[1]E. Harms, Phys. Lett. B **41**, 26 (1972).
[2]J. A. Tjon, Phys. Rev. D **1**, 2109 (1970).
[3]H. Kröger and W. Sandhas, Phys. Rev. Lett. **40**, 834 (1978).
[4]S. G. Mihlin, "On the convergence of Galerkin's method," Dokl. Akad. Nauk SSSR **61**, 197–199 (1948).
[5]C. Chandler and A. G. Gibson, "A two-Hilbert-space formulation of multichannel scattering theory," Proceedings of the Conference on Mathematical Methods and Applications of Scattering Theory, Washington, D.C., 1979.
[6]H. Kröger and R. Perne, Lett. Nuovo Cimento **27**, 406 (1980).
[7]W. V. Petryshyn, Bull. Am. Math. Soc. **81**, 223 (1975); *Proceedings Symposium Pure Mathematics* (Amer. Math. Soc., Providence, R. I., 1970), Vol. 18, part 1, p. 206.
[8]H. Kröger and C. Fenske, Reports Math. Phys. (in press).
[9]S. Weinberg, Phys. Rev. **131**, 440 (1963).

Bargmann transform, Zak transform, and coherent states

A. J. E. M. Janssen
California Institute of Technology, Pasadena, California 91125

(Received 7 November 1980; accepted for publication 11 March 1981)

It is well known that completeness properties of sets of coherent states associated with lattices in the phase plane can be proved by using the Bargmann representation or by using the kq representation which was introduced by J. Zak. In this paper both methods are considered, in particular, in connection with expansions of generalized functions in what are called Gabor series. The setting consists of two spaces of generalized functions (tempered distributions and elements of the class S^*) which appear in a natural way in the context of the Bargmann transform. Also, a thorough mathematical investigation of the Zak transform is given. This paper contains many comments and complements on existing literature; in particular, connections with the theory of interpolation of entire functions over the Gaussian integers are given.

PACS numbers: 02.30.Mv, 02.30.Lt

1. INTRODUCTION

If $x\in\mathbb{R}$, $y\in\mathbb{R}$, then $G(x,y)$ denotes the function

$$(G(x,y))(t) = 2^{\frac{1}{4}}\exp(-\pi(t-x)^2 + 2\pi iyt - \pi ixy) \quad (t\in\mathbb{R});$$

$G(x,y)$ is called a coherent state,[1,2] or also a Gabor function.[3,4] In the past ten years a number of papers[1,2,4–7] appeared about the completeness of the collection $\{G(n\alpha,m\beta)|n,m \text{ integers}\}$ where $\alpha>0$, $\beta>0$. These papers deal with the following question: if f is a (generalized) function and $(f,G(n\alpha,m\beta))=0$ for all integers n and m, then does it follow that $f\equiv 0$? As early as 1932, von Neumann[8] noticed (apparently without publishing a proof) that the answer is "yes" if $f\in L^2(\mathbb{R})$, $\alpha\beta=1$. Two proofs of this fact were given in 1970 by using the Bargmann transform,[9] and in 1975 a proof was given by using the kq representation. The most complete answer to the above question was probably given in 1979. It is shown[7] that $(f,G(n\alpha,m\beta))=0$ for all integers n and m implies $f\equiv 0$ for a very large class of generalized functions f whenever $\alpha\beta<1$. Also, in case $\alpha\beta=1$, a characterization of all tempered distributions f with $(f,G(n\alpha,m\beta))=0$ for all n and m is given. The main tools are a Phragmén–Lindelöf theorem and the Bargmann transform, although the latter is not explicitly mentioned.

A related question concerns expansion of (generalized) functions f in series of the form $\Sigma_{n,m}c_{nm}G(n\alpha,m\beta)$ with $\alpha\beta=1$ (Gabor series). In 1946 Gabor[10] suggested a simultaneous time-frequency analysis of signals based on these expansions. In 1979 existence and uniqueness theorems about Gabor series were given (cf. Ref. 6, where expressions for the coefficients c_{nm} are given, and Ref. 4, where existence of Gabor's expansions for tempered distributions is proved; in both papers the kq representation, although not explicitly mentioned, plays an important role).

We give a survey of the content of this paper. In Sec. 2 we consider the spaces S of smooth functions and S^* of generalized functions, and we show that, in connection with the Bargmann transform, these spaces arise in a natural way. In Sec. 3 the Zak transform T, which maps functions f of one real variable onto functions Tf defined on the unit square, is introduced and studied in detail. A peculiar property of the Zak transform[11] is the following one: if $f\in L^2(\mathbb{R})$ and Tf is continuous, then Tf has a zero in the unit square. In Sec. 4 a number of consequences of this property are given. One of the consequences is that one can improve the convergence of Gabor series (which, in general, converge not even in L^2 sense for elements of S) by shifting the lattice over a distance (a,b) with suitably chosen numbers a and b. Also, the results about completeness after deleting one or more coherent states[2,5,7] are completed and generalized, and a relation with classical results in interpolation theory is indicated. Although almost all results deal with square lattices of unit area with axes parallel to the x and y axis in the phase plane, some indications are given how to handle general lattices. Finally, the paper shows existence of Gabor's expansion for elements of S^*.

2. THE SPACES S AND S^* AND THE BARGMANN TRANSFORM

In 1961 Bargmann[12] constructed a unitary mapping of $L^2(\mathbb{R})$ onto the set F of all entire functions f of growth $\leqslant(2,\frac{1}{2})$ for which $\int_{\mathbb{C}}|f(z)|^2e^{-|z|^2}dz<\infty$. On the space F Fock's solution $\xi=\partial/\partial\eta$ of the commutation rule $[\xi,\eta]=1$ is realized. In 1967 Bargmann[13] described several spaces of test functions and generalized functions in terms of certain subsets of F and duals of these. In particular the spaces S and S′ (Schwartz's space of functions of rapid decrease and of tempered distributions respectively) were considered. In this section we shall investigate the relation between Bargmann transform and the spaces S and S^* (of smooth and generalized functions respectively) which were introduced in Ref. 14 and studied extensively in Refs. 3, 15, and 16.

2.1. The space S consists of all entire functions f for which there are $M>0$, $A>0$, $B>0$ such that

$$(*)\,|f(x+iy)|\leqslant M\exp(-\pi Ax^2+\pi By^2) \quad (x\in\mathbb{R},\, y\in\mathbb{R}).$$

A sequence $(f_n)_n$ in S converges to zero in S sense ($f_n \overset{S}{\to} 0$) if there are $M>0$, $A>0$, $B>0$ such that (*) holds for all f_n and such that $f_n\to 0$ pointwise. The space S^* consists of all continuous antilinear functionals defined on S. The action of $F\in S^*$ on $f\in S$ is denoted by (F,f). A sequence $(F_n)_n$ in S^*

 0022-2488/82/050720-12$02.50

converges to zero in S^* sense ($F_n \xrightarrow{S^*} 0$) if $(F_n, f) \to 0$ for all $f \in S$. Note that $G(x,y) \in S$ for all $x \in \mathbb{R}$, $y \in \mathbb{R}$.

2.2. For $n = 0,1,\cdots,\psi_n$ denotes the nth Hermite function. We choose our normalizations such that[17]

$$\exp(\pi x^2 - 2\pi(x-w)^2) = \sum_{n=0}^{\infty} c_n w^n \psi_n(x) \quad (x \in \mathbb{C},\ w \in \mathbb{C}),$$

where $c_n = 2^{-1/4}(4\pi)^{n/2}/(n!)^{1/2}$ for all n. We have $\psi_n \in S$.

There is a one-to-one correspondence between the space S and the space D of all complex sequences $(a_n)_n$ with $a_n = O(e^{-n\epsilon})$ for some $\epsilon > 0$; if $f \in S$, then $((f,\psi_n))_n \in D$, and if $(a_n)_n \in D$, then $\Sigma_n a_n \psi_n$ converges in S sense to an element of S. There is a similar correspondence between the space S^* and the space D^* of all complex sequences $(b_n)_n$ with $b_n = O(e^{n\epsilon})$ for all $\epsilon > 0$. It follows[18] that $S \subset \mathrm{S}$ and that $\mathrm{S}' \subset S^*$.

2.3. A different way to describe the space S is the following one: In Ref. 14 the space $S_\beta^\alpha (\alpha > 0, \beta > 0)$ is defined as the set of all functions f: $\mathbb{R} \to \mathbb{C}$ for which there exist $C > 0, A > 0, B > 0$ such that

$$(**) |x^k f^{(q)}(x)| \leqslant CA^k B^q k^{k\alpha} q^{q\beta},$$

for all $x \in \mathbb{R}$, $k = 0,1,\cdots,q = 0,1,\cdots$. Our space S can be identified with $S_{1/2}^{1/2}$ as follows: If $f \in S$, then the restriction of f to $\mathbb{R}$ satisfies inequalities as in (**), and if we have an f: $\mathbb{R} \to \mathbb{C}$ satisfying inequalities as in (**), then f can be extended to an entire function satisfying an inequality as in 2.1(*). Also, the notions of convergence in 2.1 for S and in Ref. 14 for $S_{1/2}^{1/2}$ can be shown to be equivalent.

We note some topological properties of the spaces S and S^*. If we consider S and S^* with the weak * topologies, i.e., with the linear topologies generated by all sets of the form $\{f \in S \mid (F,f) \in O\}$ (where $F \in S^*$, $O \subset \mathbb{C}$ open) and $\{F \in S^* \mid (F,f) \in O\}$ (where $f \in S$, $O \subset \mathbb{C}$ open), then the dual of S is S^* and the dual of S^* is S. The space S (S^*) is complete in the sense that if $f_n \in S$ ($F_n \in S^*$) and $\lim_{n\to\infty} (F, f_n)$ [$\lim_{n\to\infty} (F_n, f)$] exists for all $F \in S^*$ ($f \in S$), then there is an $f \in S$ ($F \in S^*$) such that $(F,f) = \lim_{n\to\infty} (F, f_n)$ [$(F,f) = \lim_{n\to\infty}(F_n, f)$] for all $F \in S^*$ ($f \in S$). More information can be found in Ref. 15.

To indicate how big the space S^* is, we observe that any measurable F:$\mathbb{R} \to \mathbb{C}$ for which $\int_{-\infty}^{\infty} \exp(-\epsilon t^2)|F(t)|\, dt < \infty$ for all $\epsilon > 0$ can be regarded as an element of S^* by putting $(F,f) := \int_{-\infty}^{\infty} F(t)\, \overline{f(\bar{t})}\, dt$ for $f \in S$.

2.4. We give a list of operators of S: if $f \in S$, $a \in \mathbb{C}$, $b \in \mathbb{C}$, $\alpha > 0$, then

$$(T_a f)(t) = f(t+a), \quad (R_b f)(t) = e^{-2\pi i b t} f(t),$$

$$(N_\alpha f)(t) = \left(\frac{1}{\sinh\alpha}\right)^{1/2} \times \int_{-\infty}^{\infty} \exp\left(\frac{-\pi}{\sinh\alpha}((t^2 + z^2)\cosh\alpha - 2tz)\right) f(z)\, dz,$$

$$(\mathscr{F} f)(t) = \int_{-\infty}^{\infty} e^{-2\pi i t z} f(z)\, dz, \ (\mathscr{F}^* f)(t) = (\mathscr{F} f)(-t),$$

$$(Pf)(t) = (1/2\pi i) f'(t), \quad (Qf)(t) = t f(t),$$

for $t \in \mathbb{C}$.[19] These operators are continuous and have adjoints that map S into S; they can therefore be extended to continuous linear operators of S^*.[20]

2.5. *Definition*: For $F \in S^*$ the Bargmann transform BF of F is defined by

$$(BF)(z) = e^{\frac{1}{2}\pi z^2}(T_z F, g) \quad (z \in \mathbb{C}).$$

Here $g(t) = 2^{1/4}\exp(-\pi t^2)$ for $t \in \mathbb{C}$.

2.6. The following formula (for $F \in \mathrm{S}'$) is due to Bargmann[12]; for the sake of completeness we give a proof.

Theorem: Let $F \in S^*$. Then

$$(BF)(z) = \sum_{n=0}^{\infty} \frac{(F,\psi_n)}{(n!)^{1/2}} (\pi^{1/2} z)^n \quad (z \in \mathbb{C}).$$

Proof: Put $h_z(t) = 2^{1/4}\exp(\frac{1}{2}\pi z^2 - \pi(t-z)^2)$ for $z \in \mathbb{C}$, $t \in \mathbb{C}$. Since $(BF)(z) = (f, h_{\bar{z}})$ and

$$h_z(t) = 2^{1/4}\exp(\pi t^2 - 2\pi(t - \tfrac{1}{2}z)^2) = \sum_{n=0}^{\infty} \frac{\psi_n(t)}{(n!)^{1/2}} (\pi^{1/2} z)^n,$$

with convergence in S sense for every $z \in \mathbb{C}$, we have

$$(BF)(z) = (F, h_{\bar{z}}) = \sum_{n=0}^{\infty} \frac{(F,\psi_n)}{(n!)^{1/2}} (\pi^{1/2} z)^n.$$

2.7. Let $\mathscr{E}$ be the space of all entire functions of growth $\leqslant (2,\pi/2)$, and let μ be the measure on $\mathbb{C}$ defined by $d\mu(z) = e^{-\pi|z|^2}\, dz$. If $\mathscr{H} = \mathscr{E} \cap L^2(\mathbb{C},\mu)$, then $\mathscr{H}$ is a Hilbert space for which $(\pi^{\frac{1}{2}n} z^n / \sqrt{n!})_n$ is a complete orthonormal system, and B maps $L^2(\mathbb{R})$ isometrically onto $\mathscr{H}$.[12] Also,[13]

$$B(\mathrm{S}) = \{ f \in \mathscr{E} \mid f(z)\exp(-\tfrac{1}{2}\pi|z|^2) = O((1+|z|)^{-N}) \quad \text{for all } N > 0 \},$$

$$B(\mathrm{S}) = \{ f \in \mathscr{E} \mid f(z)\exp(-\tfrac{1}{2}\pi|z|^2) = O((1+|z|)^{N}) \quad \text{for some } N > 0 \}.$$

Theorem: (i) $B(S) = \{ f \in \mathscr{E} \mid \text{growth of } f < (2,\pi/2\}$, (ii) $B(S^*) = \mathscr{E}$.

Proof: Let $F \in S$. There is an $\epsilon > 0$ such that $(F,\psi_n) = O(e^{-n\epsilon})$. Hence, by Stirling's formula

$$[(F,\psi_n)/(n!)^{1/2}]\pi^{n/2} = O(n^{-1/4}(\pi/n)^{n/2} e^{-n(\epsilon + \frac{1}{2})}).$$

It follows from Ref. 21, Theorem 2.2.2 that Bf has growth $< (2,\pi/2)$. Conversely, let $f \in \mathscr{E}$, growth of $f < (2,\pi/2)$. Writing $f(z) = \Sigma_n a_n z^n$ we know from Ref. 21, 2.2.10 that lim sup $n|a_n|^{2/n} < \pi e$. Hence $b_n := a_n \pi^{-n/2}(n!)^{1/2} = O(e^{-n\epsilon})$ for some $\epsilon > 0$. So if we put $F = \Sigma_n b_n \psi_n$, then $F \in S$ and $BF = f$.

The proof of (ii) is similar.

Remarks: (1) There are similar characterizations for the elements of $B(\mathscr{C})$ and $B(\mathscr{M})$ ($\mathscr{C}$ is the convolution class and ($\mathscr{M}$) is the multiplication class; cf. Ref. 16). It may be shown that

$$B(\mathscr{C}) = \{ f \in \mathscr{E} \mid \forall_{p > \frac{1}{2}} \exists_{q < \frac{1}{2}} [f(x+iy) = O(\exp(\pi q x^2 + \pi p y^2))] \},$$

$$B(\mathrm{M}) = \{ f \in \mathscr{E} \} \mid \forall_{p > \frac{1}{2}} \exists_{q < \frac{1}{2}} [f(x+iy) = O(\exp(\pi q y^2 + \pi p x^2))] \}.$$

(2) Theorem 2.7 shows that Theorem 2.8 in Ref. 7 is in some sense the best possible result that can be obtained by using the Bargmann transform.

2.8. In the list below we have $F \in S^*$, $a \in \mathbb{C}$, $b \in \mathbb{C}$, $\alpha > 0$, $z \in \mathbb{C}$.

(1) $(BT_a F)(z) = e^{-\frac{1}{2}\pi a^2 - \pi a z}(BF)(z+a)$,

(2) $(BR_b F)(z) = e^{-\frac{1}{2}\pi b^2 - \pi i b z}(BF)(z - ib)$,

(3) $(BN_\alpha F)(z) = e^{-\frac{1}{2}\alpha}(BF)(e^{-\alpha}z)$,

(4) $(B\mathscr{F}F)(z) = (BF)(iz)$,

(5) $(BPF)(z) = \frac{1}{2}iz(BF)(z) + (1/2\pi i)(BF)'(z)$,

(6) $(BQF)(z) = \frac{1}{2}z(BF)(z) + (1/2\pi)(BF)'(z)$,

(7) $(B(Q-iP)F)(z) = z(BF)(z)$,

(8) $(B(Q+iP)F)(z) = (1/\pi)(BF)'(z)$,

(9) $(B(Q^2+P^2)F)(z) = (1/2\pi)(BF)(z) + (z/\pi)(BF)'(z)$.

The proofs of these formulas are straightforward; compare also Ref. 12 where BU is calculated with U a canonical operator associated with a symplectic transformation of the phase plane, and Ref. 3, Sec. 27.3.

The obvious advantage of the space S^* over $\mathscr{S}'$ is that we can consider in (1) and (2) complex values of a and b. The obvious disadvantage is the fact that S^* is described in terms of entire functions so that its elements are hard to localize. Nevertheless, it appears that one can say at least something[22] about the carriers of the elements of S^* with the aid of the Bargmann transform and the theory of analytic functionals.

Some other useful formulas are

$$(BG(x,y))(z) = \exp(-\tfrac{1}{2}\pi(x^2+y^2) + \pi(x+iy)z) \quad (z\in\mathbb{C}),$$

$$(B\delta_0^{(k)})(z) = (-1)^k(k!)^{1/2}(2\pi)^{k/2}\psi_k(z/\sqrt{2}) \quad (z\in\mathbb{C})$$

for $k = 0,1,\cdots$. For $F\in S^*$, $f\in S$, $a\in\mathbb{R}$, $b\in\mathbb{R}$,

$$(F,G(a,b)) = \exp(-\tfrac{1}{2}\pi(a^2+b^2))(BF)(a-ib),$$

$$(F,f) = \int_{\mathbb{C}} e^{-\pi|z|^2}(BF)(z)\,\overline{(Bf)(z)}\,dz,$$

so that (integration over $\mathbb{R}^2$)

$$(F,f) = \int\int (F,G(a,b))(G(a,b),f)\,da\,db,$$

which agrees with the formula 27.12.1.5 in Ref. 3.

3. THE ZAK TRANSFORM

In this section we study the Zak transform which was introduced in 1967 by Zak to construct a quantum mechanical representation (kq representation) for the description of the motion of a Bloch electron in the presence of a magnetic or electric field.[23–25] This representation can also be used for the quantum mechanical description of angle and phase.[26] The Zak transform T maps functions f defined on $\mathbb{R}$ onto functions Tf of two variables as follows:

$$(Tf)(z,w) = \sum_{n=-\infty}^{\infty} f(z-n)e^{-2\pi inw}.$$

Zak denotes the first variable (quasiposition variable) by q and the second variable (quasimomentum variable) by k. We consider here T as a mapping of $L^2(\mathbb{R})$ into $L^2([0,1]^2)$, and also as a mapping of S^* into S^{2^*} [and of S' into $(\mathrm{S}^2)'$]. Although the Zak transform looks, at first sight, less interesting from the mathematical point of view than does the Bargmann transform, it pays (as we shall see in the next section) to investigate its properties systematically. A striking property is that Tf has a zero in $[0,1]^2$, provided that Tf is continuous. We further give a formula for the product $TF\cdot\overline{Tf}$ (in case this makes sense) which is very convenient when proving completeness properties, and we determine $T(S)$, $T(S^*)$, $T(\mathrm{S})$, and $T(\mathrm{S}')$.

3.1. *Definition*: Let $F\in S^*$. We define

$$TF := \sum_{n=-\infty}^{\infty} T_{-n}^{(1)}R_n^{(2)}(F\otimes H),$$

where $H\equiv 1$ [for the definition of the tensor product, cf. Ref. 15, Appendix 1, 1.17; we have $(F_1\otimes F_2, f_1\otimes f_2) = (F_1,f_1)(F_2,f_2)$ for $F_i\in S^*$, $f_i\in S$ $(i=1,2)$]. This definition makes sense, for if $F\in S^*$, $f_1\in S$, $f_2\in S$, then

$$\sum_{n=-\infty}^{\infty} (T_{-n}^{(1)}R_n^{(2)}(F\otimes H), f_1\otimes f_2)$$

$$= \sum_{n=-\infty}^{\infty} (T_{-n}F,f_1)(\mathscr{F}\bar{f}_2)(n)$$

converges absolutely by Ref. 16, Theorem 5.5. By Ref. 15, Appendix 1, Theorem 3.7, the series $\Sigma_n T_{-n}^{(1)}R_n^{(2)}(F\otimes H)$ converges unconditionally in S^{2^*} sense. It also follows from Ref. 15, Appendix 1, 4.14 that T is a continuous linear mapping from S^* into S^{2^*}, and we have $F=0$ if $TF=0$. Similar things hold if we consider T as a mapping from S' into $(\mathrm{S}^2)'$. In the case $f\in\mathrm{S}$, Tf can be identified with the function

$$\sum_{n=-\infty}^{\infty} e^{-2\pi inw}f(z-n) \quad [(z,w)\in\mathbb{R}^2].$$

3.2. Part of the following theorem is taken from Ref. 4; for the sake of completeness we include a proof. We also note that part (i) occurs in a more abstract version in the proof of Ref.27, Chap. 1, Sec. 5, Lemma 4.

Theorem: (i) T maps $L^2(\mathbb{R})$ isometrically onto $L^2([0,1]^2)$. (ii) Let $1\leqslant p<2$. Then T maps $L^p(\mathbb{R})$ into $L^p([0,1]^2)$, and the operator norm $\leqslant 1$; T is injective but not surjective.

Proof: (i) Let $f\in L^2(\mathbb{R})$. Since the functions $f(z-n)e^{-2\pi inw}$ are orthogonal over $[0,1]^2$ we see that

$$\int_0^1\int_0^1 |(Tf)(z,w)|^2\,dz\,dw$$

$$= \sum_{n=-\infty}^{\infty}\int_0^1\int_0^1 |f(z-n)e^{-2\pi inw}|^2\,dz\,dw$$

$$= \int_{-\infty}^{\infty} |f(z)|^2\,dz.$$

Hence T is well defined as a mapping of $L^2(\mathbb{R})$ into $L^2([0,1]^2)$, and it is norm-preserving.

Now let $g\in L^2([0,1]^2)$, and let

$$c_{nm} := \int_0^1\int_0^1 g(z,w)e^{2\pi imz+2\pi inw}\,dz\,dw$$

for integers n and m. Putting $f(z-n) := \Sigma_m c_{nm}e^{-2\pi inz}$ for $0\leqslant z<1$ and integer n, we easily see that $f\in L^2(\mathbb{R})$, and that $Tf=g$.

(ii) Let $f\in L^1(\mathbb{R})$. Then

$$\int_0^1\int_0^1 |(Tf)(z,w)|\,dz\,dw$$

$$\leqslant \int_0^1\int_0^1 \Sigma_n|f(z-n)|\,dz\,dw = \int_{-\infty}^{\infty}|f(z)|\,dz.$$

Hence $Tf\in L^1([0,1]^2)$ and $\|Tf\|_1\leqslant\|f\|_1$. It follows from convexity theory that T maps $L^p(\mathbb{R})$ into $L^p([0,1]^2)$, and that

$\|Tf\|_p \leqslant \|f\|_p$ for $f \in L^p(\mathbb{R})$, $1 \leqslant p \leqslant 2$.

To show injectivity, let $f \in L^p(\mathbb{R}), f \neq 0$. Let $f_n \in L^2(\mathbb{R})$ be such that $|f_n| \leqslant |f_{n+1}|$ and $f_n \to f$, $Tf_n \to Tf$ a.e. By Fatou's lemma and part (i),

$$\int_0^1 \int_0^1 |(Tf)(z,w)|^2 \, dz \, dw$$

$$\geqslant \limsup_{n\to\infty} \int_0^1 \int_0^1 |(Tf_n)(z,w)|^2 \, dz \, dw = \limsup_{n\to\infty} \|f_n\|_2^2 > 0.$$

It is trivial that T is not surjective if $1 \leqslant p < 2$; otherwise we would have $T(L^p(\mathbb{R})) \supset T(L^2(\mathbb{R}))$, whence $L^p(\mathbb{R}) \supset L^2(\mathbb{R})$.

Remarks: (1) There is no way to define T as a mapping of $L^p(\mathbb{R})$ into any $L^r([0,1]^2)$ if $p > 2$ (cf. Ref. 28, Chap. XII, 2, p. 102). (2) As a mapping of $L^p(\mathbb{R})$ into $L^p([0,1]^2)$ with $1 \leqslant p < 2$, T is not bounded below. To see this, put $f_c := \Sigma_n c_n \chi_{[n,n+1]}$ for $c = (c_n)_n \in l^p$. If T were bounded below there would be an $m > 0$ such that $\|Tf_c\| \geqslant m \|f_c\|_p = m \Sigma_n |c_n|^p)^{1/p}$ for $c \in l^p$. As $(Tf_c)(z,w) = \Sigma_n c_n e^{-2\pi i n w}$ for $c \in l^p$, this implies that $\{\Sigma_n c_n e^{-2\pi i n w} | c \in l^p\} = L^p([0,1])$. Contradiction.

3.3. In the list below we have $F \in S^*$, $a \in \mathbb{C}$, $b \in \mathbb{C}$, $\alpha > 0$,

(1) $T(T_a F) = T_a^{(1)}(TF)$,

(2) $T(R_b F) = R_b^{(1)} T_b^{(2)}(TF)$,

(3) $T_1^{(1)}(TF) = R_1^{(2)}(TF)$,

(4) $T_1^{(2)}(TF) = TF$,

(5) $T(N_\alpha F)$

$$= N_\alpha^{(1)}\left(\sum_{n=-\infty}^{\infty} e^{-\pi n^2 \cosh\alpha \sinh\alpha} (R_{in\sinh\alpha} T_{-n\cosh\alpha} F) \otimes H \right),$$

(6) $T\mathscr{F}F = e^{-2\pi i z w} UTF$,

(7) $T(PF) = P^{(1)}TF$,

(8) $T(QF) = (Q^{(1)} + P^{(2)})TF$.

Here U is (the extension of) the mapping that takes $f(z,w) \in S^2$ into $f(w, -z)$. All formulas except (6) follow directly from a computation. To prove (6) we first take an $f \in S$. We have by definition

$$(T\mathscr{F} f)(z,w) = \sum_{n=-\infty}^{\infty} (\mathscr{F} f)(z-n) e^{-2\pi i n w}.$$

Observing that $(\mathscr{F} f)(z-n)e^{-2\pi i n w} = e^{-2\pi i z w}(\mathscr{F} R_z T_w f)(-n)$, we get by the Poisson summation formula

$$(T\mathscr{F} f)(z,w) = e^{-2\pi i z w} \sum_{n=-\infty}^{\infty} e^{2\pi i n z} f(w-n)$$

$$= e^{-2\pi i z w}(Tf)(w, -z).$$

For the general case take a sequence $(f_n)_n$ in S that converges in S^* sense to F, and use continuity of T (cf. 3.1).

Remarks: (1) Formula (6) and Theorem 3.2 (i) give a quick proof of Plancherel's theorem since U maps $L^2([0,1]^2)$ unitarily onto $L^2([0,1] \times [-1,0])$. It is of course the Poisson summation formula that does the trick here. (2) If Tf is sufficiently well behaved we can recover f and $\mathscr{F} f$ by integration. We have

$$f(z) = \int_0^1 (Tf)(z,w) \, dw, \quad (\mathscr{F} f)(-w)$$
$$= \int_0^1 (Tf)(z,w) e^{2\pi i z w} \, dz,$$

for $z \in \mathbb{R}$, $w \in \mathbb{R}$.

3.4. We calculate $TG(x,y)$ and $T\psi_n$ for $x \in \mathbb{R}$, $y \in \mathbb{R}$, $n = 0,1,\cdots$. We have by the formulas of 3.3

$$(TG(x,y))(z,w) = (T(e^{-\pi i x y} R_{-y} T_{-x} g))(z,w)$$
$$= e^{\pi i x y + 2\pi i y z}(Tg)(z-x, w-y),$$

so that we need Tg. In general, we have by the generating function of the Hermite functions (cf. 2.2),

$$c_k(T\psi_k)(z,w)$$
$$= C_{t^k}\left[e^{\pi z^2 - 2\pi(z-t)^2} \sum_n e^{-\pi n^2 + 2\pi i n(w + iz - 2it)} \right]$$
$$= C_{t^k}\left[\theta_3(w + iz - 2it) \sum_{l=0}^{\infty} c_l t^l \psi_l(x) \right].$$

Here $\theta_3(z) = \Sigma_n \exp(-\pi n^2 + 2\pi i n z)$ is the 3rd theta function [in the notation of Ref. 29 we have $\theta_3(z) = \vartheta(\pi z, e^{-\pi})$]. By the Taylor expansion of θ_3 around the point $w + iz$ we get

$$(T\psi_k)(z,w) = \sum_{l=0}^{k} \frac{\psi_{k-l}(z)}{\sqrt{l!}} \binom{k}{l}^{1/2} (-\pi^{-1/2} i)^l \theta_3^{(l)}(w + iz).$$

In particular,

$$(TG(x,y))(z,w) = (G(x,y))(z) \theta_3(w + iz - y - ix),$$

and in case n and m are integers we get by 3.3,

$$(TG(n,m))(z,w) = (-1)^{nm} e^{2\pi i m z + 2\pi i n w} e^{-\pi z^2} \theta_3(w + iz).$$

As another example, let $e_a(t) := e^{-2\pi i a t}$ where $a \in \mathbb{R}$. We have

$$T(e_a) = \sum_{n=-\infty}^{\infty} e_a \otimes \delta_{n-a}.$$

Hence, if a is an integer, $T(e_a) = e_a \otimes \Sigma_n \delta_n$, and if $f \in S^*$ is periodic with period one, then $Tf = f \otimes \Sigma_n \delta_n$. Similarily, if f is a function of the form $f = \Sigma_n c_n \delta_n$, then $Tf = e^{-2\pi i z w} (\Sigma_n \delta_n) \otimes \mathscr{F} f$.

3.5. It is easy to see that Tf has a zero in $[0,1]^2$ if Tf is continuous and f is real-valued, or even, or odd, or a Gabor function. The following theorem shows this is general.

Theorem: Let $f \in L^2(\mathbb{R})$ be such that Tf is continuous. Then Tf has a zero in $[0,1]^2$.

Proof[30]: Assuming $(Tf)(z,w) \neq 0$ for $(z,w) \in [0,1]^2$ we can write

$$(Tf)(z,w) = e^{2\pi i \varphi(z,w)},$$

where $\varphi: \mathbb{R}^2 \to \mathbb{C}$ is continuous. Indeed, this follows at once from Ref. 31, Part VI, Sec. 1, Lemma 7.

We have by 3.3

$$(Tf)(z+1,w) = e^{2\pi i w}(Tf)(z,w),$$
$$(Tf)(z,w+1) = (Tf)(z,w),$$

for $(z,w) \in [0,1]^2$. Hence, for some integers k and l

$$\varphi(z+1,w) = \varphi(z,w) + w + k,$$
$$\varphi(z,w+1) = \varphi(z,w) + l,$$

for $(z,w) \in [0,1]^2$. Calculating $\varphi(1,1)$ in two different ways, we get $\varphi(0,0)+k+l=\varphi(1,1)=\varphi(0,0)+k+l+1$. Contradiction.

Remark: If $f=\chi_{(0,1]}$, then $|(Tf)(z,w)|=1$ for all $(z,w)\in\mathbb{R}^2$.

3.6. In the remainder of this section we determine $T(S)$, $T(S^*)$, $T(\mathrm{S})$, and $T(\mathrm{S}')$. We first give a formula which will also be used in Sec. 4 to answer questions about completeness (also cf. Ref. 6).

Theorem: Let (1) $F\in S^*$, $g\in S$ or (2) $F\in\mathrm{S}'$, $g\in\mathrm{S}$ or (3) $F\in L^2(\mathbb{R})$, $g\in L^2(\mathbb{R})$. Then

$$\sum_{n,m}(F,R_{-m}T_{-n}g)e^{2\pi inw+2\pi imz}=TF\cdot\overline{Tg},$$

where the identity is to be interpreted in S^{2^*} for case (1) and in $(\mathrm{S}^2)'$ sense for case (2). For case (3) the identity must be interpreted in the sense that the (mn)th Fourier coefficient of $TF\cdot\overline{Tg}$ equals $(F,R_{-m}T_{-n}g)$.

Proof: First take $F\in S$, $g\in S$. Noting that

$$(F,R_{-m}T_{-n}g)e^{2\pi imz}=\big(\mathscr{F}T_z(F\cdot\overline{T_{-n}g})\big)(m)$$

for all n and m we get by the Poisson summation formula (applied to the summation over m)

$$\sum_{n,m}(F,R_{-m}T_{-n}g)e^{2\pi inw+2\pi imz}$$

$$=\sum_{n,m}F(m+z)\,\overline{g(-n+m+z)\,e^{2\pi inw}}.$$

Now the formula easily follows by first summing over n and then over m.

For the general case (i.e., $F\in S^*$) take a sequence $(F_k)_k$ in S which converges to F in S^* sense. It follows as in the proof of Ref. 16, Lemma 5.2 that for every $\epsilon>0$ there are positive numbers M and β such that

$$|(F_k,R_{-m}T_{-n}g)|\leqslant M\,\|N_\beta F_k\|_2\exp(\pi\epsilon(n^2+m^2)),$$

for all n,m, and k (note that $N_\beta F_k\in S$ for $\beta>0$). Since $\|N_\beta F_k\|_2$ is bounded in k for every $\beta>0$ it is not hard to complete the proof of the theorem for case (1).

The proof for case (2) is similar to that for case (1). For the proof of case (3) we take F_k and g_k in S with $F_k\to F$, $g_k\to g$ in $L^2(\mathbb{R})$ sense. Now we note that $TF_k\cdot\overline{Tg_k}\to TF\cdot\overline{Tg}$ in $L^1([0,1]^2)$ and use the result already proved with F_k and g_k in the role of $F\in S$, $g\in S$. Hence, the (mn)th Fourier coefficient of $TF\cdot\overline{Tg}$ is given by $(F,R_{-m}T_{-n}g)$.

3.7. **Theorem**: $T(S)$ equals the set of all entire functions φ of two variables such that $\varphi(z+1,w)=e^{-2\pi iw}\varphi(z,w)$, $\varphi(z,w+1)=\varphi(z,w)$ for all $(z,w)\in\mathbb{C}^2$, and such that there are $M>0$, $A>0$, $B>0$ with

$$|\varphi(x+iy,u+iv)|\leqslant M\exp(2\pi xv+\pi Ay^2+\pi Bv^2).$$

Furthermore, $T(\mathrm{S})$ equals the set of all $\varphi\in C^\infty(\mathbb{R}^2)$ such that $\varphi(z+1,w)=e^{-2\pi iw}\varphi(z,w)$, $\varphi(z,w+1)=\varphi(z,w)$ for all $(z,w)\in\mathbb{C}^2$. Finally,

$T(S^*)=\{F\in S^{2^*}\,|\,T_1^{(1)}F=R_1^{(2)}F, T_1^{(2)}F=F\}$, and
$T(\mathrm{S}')=\{F\in(\mathrm{S}^2)'\,|\,T_1^{(1)}F=R_1^{(2)}F,\ T_1^{(2)}F=F\}$.

Proof: Let $f\in S$. It is clear that Tf is an entire function of two variables. Take $K>0$, $C>0$, and $D>0$ such that

$$|f(x+iy)|\leqslant K\exp(-\pi Cx^2+\pi Dy^2)\quad(x\in\mathbb{R},\ y\in\mathbb{R}).$$

Then for real x,y,u,v,

$$\begin{aligned}&|\Sigma_n f(x+iy-n)e^{-2\pi in(u+iv)}|\\&\quad\leqslant K\,\Sigma_n\exp(-\pi C(x-n)^2+\pi Dy^2+2\pi nv)\\&\quad=K\exp(2\pi xv+\pi v^2/C+\pi Dy^2)\Sigma_n\\&\quad\times\exp(-\pi C(x+v/C-n)^2),\end{aligned}$$

whence $T(S)$ is contained in the set mentioned in the theorem. Conversely, let φ be an entire function of two variables such that $\varphi(z+1,w)=e^{-2\pi iw}\varphi(z,w)$, $\varphi(z,w+1)=\varphi(z,w)$ for all $(z,w)\in\mathbb{C}^2$, and assume that $M>0$, $A>0$, $B>0$ are such that

$$|\varphi(x+iy,u+iv)|\leqslant M\exp(2\pi xv+\pi Ay^2+\pi Bv^2).$$

Put $\psi(z)=\int_0^1\varphi(z,w)\,dw$ for $z\in\mathbb{C}$. Then ψ is an entire function for which $|\psi(x+iy)|\leqslant M\exp(\pi Ay^2)$. Also,

$$\int_0^1\varphi(x,w)\,dw=\int_0^1\varphi(x-[x],w)e^{-2\pi i[x]w}\,dw.$$

Let $t\in[0,1]$, $n\in\mathbb{Z}$. We have by analyticity and periodicity of φ in its second variable and by the estimates on φ

$$\begin{aligned}&\left|\int_0^1\varphi(t,w)e^{-2\pi inw}\,dw\right|\\&\quad=\left|\int_{0+i\gamma}^{1+i\gamma}\varphi(t,w)e^{-2\pi inw}\,dw\right|\\&\quad\leqslant M\exp(2\pi t\gamma+\pi B\gamma^2+2\pi n\gamma)\end{aligned}$$

for all real γ. Minimizing with respect to γ gives

$$\left|\int_0^1\varphi(t,w)e^{2\pi inw}\,dw\right|\leqslant M\exp(-\pi B^{-1}(t+n)^2).$$

Hence $\psi(x)=O(\exp(-\pi B^{-1}x^2))$ $(x\in\mathbb{R})$. It follows easily from the Phragmén–Lindelöf theorem that $\varphi\in S$. It is trivial that $T\psi=\varphi$.

The proof for the S case is similar and will be omitted.

To prove the assertion about $T(S^*)$ let $F\in S^{2^*}$ satisfy $T_1^{(1)}F=R_1^{(2)}F$, $T_1^{(2)}F=F$. For any $\psi\in S$, $\overline{T\psi}$ is a multiplicator of S^{2^*} and it is easy to see that $F\cdot\overline{T\psi}$ is an element of S^{2^*} which is periodic in its both variables. Hence $F\cdot\overline{T\psi}$ has a Fourier series $\Sigma_{n,m}c_{nm}(\psi)e^{2\pi imz+2\pi inw}$ (cf. Ref. 3, 27.24.3). Define G by $(G,\psi):=c_{00}(\psi)$. Then $G\in S^*$, and by Theorem 3.6, the (nm)th Fourier coefficient of $TG\cdot\overline{T\psi}$ equals $(G,R_{-m}T_{-n}\psi)=c_{nm}(\psi)$. Hence $(F-TG)\cdot\overline{T\psi}=0$ for all $\psi\in S$. To show that this implies $F_1:=F-TG=0$, let $\psi\in S$, $\psi\neq0$. We see from the formula

$$T(R_{-b}T_{-a}\psi)=e^{2\pi ibz}(T\psi)(z-a,w-b)$$

that

$$F_1\cdot\overline{(T\psi)(\bar z-a,\bar w-b)}\,\exp(-\pi(z-a)^2-\pi(w-b)^2)=0$$

for all a and b. Putting $f(z,w)=\exp(-\pi z^2-\pi w^2)\,\overline{(T\psi)(\bar z,\bar w)}$, we have $f\in S^2$, $F_1\cdot T_a^{(1)}T_b^{(2)}f=0$ for all $a\in\mathbb{R}$, $b\in\mathbb{R}$. So, if $h\in S^2$, then (cf. Ref. 16, Sec. 5)

$$\begin{aligned}0&=(F_1\cdot T_a^{(1)}T_b^{(2)}f,h)=(R_{-a}^{(1)}R_{-b}^{(2)}\mathscr{F}f,\mathscr{F}(h\cdot\overline{F_1}))\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{2\pi iaz+2\pi ibw}\\&\quad\times(\mathscr{F}f)(z,w)\,\overline{(\mathscr{F}(h\cdot\overline{F_1}))(z,w)}\,dz\,dw\end{aligned}$$

for all real a and b. As $f\neq0$, this implies that $\mathscr{F}(h\cdot\overline{F_1})=0$.

We conclude that $F_1 = 0$.

3.8. *Example*: Let $(a,b) \in \mathbb{R}^2$, $k = 0,1,\cdots$, $l = 0,1,\cdots$. It will be of some interest to know which $F \in S'$ satisfies

$$TF = \Sigma_{n,m} T_n^{(1)} R_n^{(2)} T_m^{(2)} (\delta_a^{(k)} \otimes \delta_b^{(l)}).$$

[note that the right-hand side is indeed in $T(S')$]. For convenience take $(a,b) \in (0,1)^2$, and let $\varphi \in S$ have support in $(n,n+1)$ where n is an integer. Then $(T\varphi)(z,w) = \varphi(z-n)e^{-2\pi inw}$, and $TF \cdot \overline{T\psi}$ is the periodic distribution of two variables for which the restriction to $[0,1]^2$ equals $\overline{T_{-n}\varphi \cdot \delta_a^{(k)}} \otimes R_{-n}\delta_b^{(l)}$. The 00th Fourier coefficient of this function is given by

$$(\delta_a^{(k)}, T_{-n}\varphi) \cdot (-2\pi in)^l e^{-2\pi inb}$$
$$= (\delta_{a-n}^{(k)}, \varphi)(-2\pi in)^l e^{-2\pi inb}.$$

This suggests that $F = \Sigma_{n=-\infty}^{\infty} (-2\pi in)^l e^{-2\pi inb} \delta_{a-n}^{(k)}$, and indeed, it can be verified directly that this is the F we are looking for.

4. BARGMANN TRANSFORM, ZAK TRANSFORM, AND COMPLETENESS

This section contains material that completes and elucidates the results of Refs. 1–7 about completeness of coherent states and expansions of the Gabor type. The Zak transform gives rise to more general results than the Bargmann transform does in the sense that with the former completeness properties for other functions than Gauss functions can be proved. On the other hand, the Zak transform is only useful when lattices of which the cells have unit area are considered.

We start this section by drawing some conclusions from Theorems 3.5 and 3.6. In particular, L^2 convergence of certain expansions of the Gabor type for well-behaved functions is proved. We consider, of course, the case in which Gabor functions are taken as basic functions in detail. The results of Refs. 2 and 7 about completeness if one or more coherent states are deleted are improved. Also, a connection with a result of J. M. Whittaker about interpolation over the lattice points in the complex plane is made. We finally indicate how the results can be extended to the cases with general lattices with cell area equal to one in the phase plane, and we show existence of expansions of the Gabor type for elements of S^*.

4.1. Theorem 3.6 is useful for analyzing the mapping $f \to ((f, R_{-m}T_{-n}g))_{n,m}$, where g is some fixed function. If, e.g., Tg is continuous and $1 \leqslant p \leqslant 2$, then $((f, R_{-m}T_{-n}g))_{n,m} \in l^q$ for $f \in L^p(\mathbb{R})$ (q conjugate exponent), but the mapping is not bounded below (as a mapping from $L^p(\mathbb{R})$ into l^q) by Theorem 3.5.

This has an interesting signal-theoretic consequence. The function $S_g f$ given by $(S_g f)(x,y) := (f, R_{-y}T_{-x}g)$ is sometimes called a spectrogram of f. Hence, if we sample $S_g f$ over a lattice of which the cells have area one, then it may happen that the double sequence of sample values have small q norm while both f and Tf have large p norm.

In case $f = g$ we have

$$(f, R_{-y}T_{-x}f) = e^{-\pi ixy}\mathrm{Amb}(x,y;f),$$

where $\mathrm{Amb}(\cdot,\cdot;f)$ is the so-called ambiguity function[32] of f, defined by

$$\mathrm{Amb}(x,y;f) = \int_{-\infty}^{\infty} e^{-2\pi iyt} f(t+\tfrac{1}{2}x)\,\overline{f(t-\tfrac{1}{2}x)}\,dt.$$

We derive from Theorems 3.5 and 3.6 the following inequality: if $f \in L^2(\mathbb{R})$, Tf continuous, then $\mathrm{Amb}(0,0;f) \leqslant \Sigma_{(n,m)\neq(0,0)} |\mathrm{Amb}(n,m;f)|$, which is an inequality expressing the uncertainty principle.[32,33]

Note, however, that the assumption "Tf continuous" is essential, for if $f = \chi_{[0,1)}$, then $\mathrm{Amb}(0,0;f) = 1$, $\mathrm{Amb}(n,m;f) = 0$ $[(n,m) \neq (0,0)]$.

Taking $g = f_-$, where $f_-(t) = f(-t)$, we get

$$(f, R_{-y}T_{-x}f_-) = \tfrac{1}{2}e^{-\pi ixy}\, W(\tfrac{1}{2}x, \tfrac{1}{2}y; f, f),$$

where $W(\cdot,\cdot;f)$ is the Wigner distribution of f defined by

$$W(x,y;f) = \int_{-\infty}^{\infty} e^{-2\pi iyt} f(x+\tfrac{1}{2}t)\,\overline{f(x-\tfrac{1}{2}t)}\,dt.$$

Hence the Fourier coefficients of $Tf \cdot \overline{Tf_-}$ equal $\tfrac{1}{2}(-1)^{nm}W(\tfrac{1}{2}n, \tfrac{1}{2}m; f)$. As $(Tf_-)(z,w) = (Tf)(-z,-w)$, we see that it may well happen that $W(\tfrac{1}{2}n, \tfrac{1}{2}m; f) = 0$ for all integers n and m. This is seen from the Fig. This can happen even if $f \in S$, but not if $f \in S$ (cf. Theorem 3.7).

4.2. If we take a $g \in L^2(\mathbb{R})$ for which the set of zeros of Tg has measure zero, we get the completeness results obtained in Ref. 5: the set $(R_{-m}T_{-n}g)_{n,m}$ is complete in $L^2(\mathbb{R})$. If, in addition, $Tg \in L^q([0,1]^2)$, where $q \geqslant 2$, then $(R_{-m}T_{-n}g)_{n,m}$ is complete in $L^p(\mathbb{R})$, where $p = q/(q-1)$. This can be proved by using Theorem 3.2 and generalizing Theorem 3.6 properly. And if Tg is, e.g., continuously differentiable [which implies by Theorem 3.5 that $1/\overline{Tg} \notin L^2([0,1]^2)$, then $(R_{-m}T_{-n}g)_{(n,m)\neq(0,0)}$ is still complete in $L^2(\mathbb{R})$ [and probably also in $L^p(\mathbb{R})$ for $1 \leqslant p \leqslant 2$]. Finally, if $f \in L^2(\mathbb{R})$, $g \in L^2(\mathbb{R})$ and the set of zeros of Tg has positive measure, then either $(f, R_{-m}T_{-n}g) = 0$ for all n and m or $(f, R_{-m}T_{-n}g) \neq 0$ for infinitely many n and m.

Remark: Let $g \in S$. We can use Theorem 3.6 for describing the set of all $f \in S'$ such that $(f, R_{-m}T_{-n}g) = 0$ in case Tg has a finite number of zeros $(a_1,b_1),\ldots,(a_k,b_k)$ in the unit square. By Theorem 3.6 Tf is concentrated in $(a_1,b_1),\ldots,(a_k,b_k)$, and the translates of these points over distances (n,m) with integer n and m. Using Ref. 34, Chap. 24, Theorem 2.4.6 (the restriction to the unit square of) we see that Tf has the form

$$\sum_{l=1}^{k} \sum_{i=0}^{p_l} \sum_{j=0}^{q_l} c_{lij}\, \delta_{a_l}^{(i)} \otimes \delta_{b_l}^{(j)}.$$

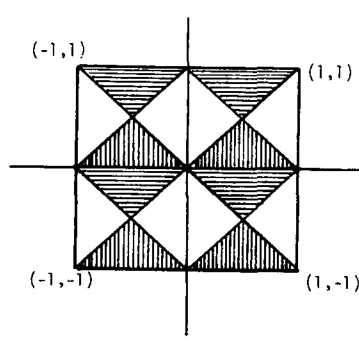


FIG. 1. If the support of Tf is in the ||||| region, then the support of Tf_- is in the $\equiv$ region.

Now Example 3.8 can be applied, and we see in particular that f is concentrated in the set $\{a_i + n | i = 1,\dots,k; n$ integer$\}$.

4.3. In Ref. 4 it was shown that for any $F \in S'$ there exists a (not unique) expansion in a Gabor series, with convergence in S' sense. If $F \in S$, say, then it may still be the case that the convergence of the series is not in L^2 sense. The next application of Theorems 3.5 and 3.6 deals with improving convergence of series of the Gabor type. Take $f \in L^2(\mathbb{R})$, $g \in L^2(\mathbb{R})$ and assume that $Tg \in L^\infty([0,1]^2)$. Let $(a,b) \in [0,1]^2$. We have for integers n and m

$$T(R_{-m-b}T_{-n-a}g) = e^{2\pi i b z}(Tg)(z-a, w-b)e^{2\pi i m z + 2\pi i n w}.$$

Hence, if $(c_{nm})_{n,m} \in l^2$, then $f = \Sigma_{n,m} c_{nm} R_{-m-b} T_{-n-a} g$ (in $L^2(\mathbb{R})$ sense) if and only if

$$(Tf)(z,w) = e^{2\pi i b z}(Tg)(z-a, w-b)\sum_{n,m} c_{nm} e^{2\pi i m z + 2\pi i n w}$$

[in $L^2([0,1]^2)$ sense]. Now, if Tf and Tg are continuous, then it is in general advisable to take a and b such that $(Tf)(z,w) = 0$ whenever $(Tg)(z-a, w-b) = 0$. By Theorem 3.5 this is always possible if Tg has only one zero in $[0,1]^2$. If a and b are such that $Tf/T(R_{-b}T_{-a}g) \in L^2([0,1]^2)$, then one can take for c_{nm} the (m,n)th Fourier coefficient of $Tf/T(R_{-b}T_{-a}g)$. We see in particular (cf. 3.4) that a considerable class of functions f have an expansion $f = \Sigma_{n,m} c_{nm} G(n+a, m+b)$ in an $L^2(\mathbb{R})$ convergent Gabor series for suitable values of a and b (this class contains all functions f for which Tf is Hölder continuous).

Note: We can use the Zak transform to prove existence of Gabor type expansions for tempered distributions in general. Let $g \in S$, $g \neq 0$ be such that Tg is real analytic and has no zeros on the edges of the unit squares (the latter assumption is probably superfluous, but convenient). According to Łojasiewicz's theorem[35] there exists a distribution Φ of two variables such that $\Phi \cdot Tg = Tf$. This Φ is, in general, not periodic, but by our assumption on g we may assume it is. Hence Φ has a Fourier series expansion $\Sigma_{n,m} c_{nm} e^{2\pi i n z + 2\pi i m w}$, and it follows that $f = \Sigma_{nm} c_{nm} R_{-n} T_{-m} g$. This generalizes Ref. 4, Theorem 4.7 where the Gabor case was considered [however, in the proof of the theorem quoted a method is given to determine the double sequence $(c_{nm})_{n,m}$].

4.4. We now turn to the Gabor case in detail. We know from Ref. 7, 2.14 that any regular tempered distribution f with $(f, G(n,m)) = 0$ for all integers n and m, $(n,m) \neq (0,0)$ is a multiple of the function f_0 given by

$$f_0(t) = 2^{-1/4}\exp(\pi t^2)\sum_{n-\frac12 \geq t}(-1)^n \exp(-\pi(n-\tfrac12)^2).$$

This function which can also be found in Ref. 6, is an interesting one as we shall see. We let $g(t) = 2^{1/4}\exp(-\pi t^2)$.

Theorem: For every p, $1 \leqslant p < \infty$ we have $f_0 \in L^\infty(\mathbb{R}) \setminus L^p(\mathbb{R})$. The Bargmann transforms $(Bf_0)(z)$ of f_0 equals $-\theta_1(z)e^{\frac12\pi z^2}/2\pi z$, where θ_1 is the first theta function [in the notation of Ref. 29 we have $\theta_1(z) = \vartheta_1(\pi z, e^{-\pi})$]. We further have $\mathscr{F} f_0 = f_0$, and $Tf_0 = d/\overline{Tg}$ where $d = -\frac12\vartheta_1'(0, e^{-\pi})$.

Proof: It was already observed in the proof of Ref. 7, Theorem 2.14 that f_0 is bounded. Now let $1 \leqslant p < \infty$. We have for $t \geqslant 0$

$$|f_0(t)| = 2^{-1/4}\exp(\pi t^2 - \pi\varphi^2(t)) + O(\exp(-2\pi t)),$$

where $\varphi(t) = [t + \frac12] - \frac12$. If $n = 1, 2, \cdots$, then

$$\begin{aligned}\int_{n-\frac12}^{n+\frac12} &\exp(\pi p t^2 - \pi p \varphi^2(t))\,dt \\ &= \int_{n-\frac12}^{n+\frac12}\exp(\pi p t^2 - \pi p(n+\tfrac12)^2)\,dt \\ &\geqslant \exp(-p\pi)[(n+\tfrac12) - ((n+\tfrac12)^2 - 1)^{1/2}] \\ &\geqslant \exp(-p\pi)/(2n+1).\end{aligned}$$

It follows that $f_0 \notin L^p(\mathbb{R})$.

To show that $(Bf_0)(z) = -\theta_1(z)e^{\frac12\pi z^2}/2\pi z$, we observe that

$$P(g\cdot f_0) = -\sum_{n=-\infty}^{\infty}(-1)^n e^{-\pi(n-\frac12)^2}\delta_{n-\frac12}/2\pi i.$$

Taking Fourier transforms and using that $\mathscr{F}P = Q\mathscr{F}$, we get[36]

$$-2\pi i z(\mathscr{F}(g\cdot f_0))(z) = \sum_{n=-\infty}^{\infty}(-1)^n e^{-\pi(n-\frac12)^2 - 2\pi i(n-\frac12)z}.$$

The right-hand side is equal to $i\vartheta_1(\pi z, e^{-\pi})$. Also

$$(\mathscr{F}(g\cdot f_0))(z) = (R_z f_0, g) = e^{-\pi z^2}(T_{-iz} f_0, g).$$

Now using Jacobi's identity $\vartheta_1(\pi z, e^{-\pi})$ $= i\exp(-\pi z^2)\vartheta_1(\pi i z, e^{-\pi})$ (Ref. 29, 21.51) we get

$$-2\pi i z e^{-\pi z^2}(T_{-iz} f_0, g) = e^{-\pi z^2}\vartheta_1(\pi i z, e^{-\pi}).$$

It follows from the oddness of ϑ_1 that

$$(T_z f_0, g) = -\vartheta_1(\pi z, e^{-\pi})/2\pi z.$$

Now use the definition of B.

We next show that $Tf_0 = d/\overline{Tg}$. As it stands, this relation must be considered in distributional sense since $f_0 \notin L^p(\mathbb{R})$ for $1 \leqslant p \leqslant 2$. We have $(f_0, G(n,m)) = 0$ for all integers n and m with $(n,m) \neq (0,0)$, and according to what we have found above,

$$-2\pi z(R_z f_0, g) = \vartheta_1(\pi z, e^{-\pi}),$$

so that (cf. Ref. 29, 21.41)

$$\begin{aligned}(f_0, g) &= \frac{-1}{2\pi}\frac{d}{dz}[\vartheta_1(\pi z, e^{-\pi})]|_{z=0} \\ &= -\tfrac12\vartheta_1'(0, e^{-\pi}) \\ &= -\tfrac12\vartheta_2(0, e^{-\pi})\vartheta_3(0, e^{-\pi})\vartheta_4(0, e^{-\pi}) \neq 0.\end{aligned}$$

Applying Theorem 3.6 we get

$$Tf_0\cdot\overline{Tg} = -\tfrac12\vartheta_1'(0, e^{-\pi}) = d.$$

Since $Tf_0 \in (S^2)'$ and Tg has a zero in $(\frac12, \frac12)$ and nowhere else in $[0,1]^2$, there is a $\varphi \in (S')$ concentrated in the points $(n+\frac12, m+\frac12)$ with integers n and m such that $Tf_0 = d/\overline{Tg} + \varphi$ [by Theorem 3.7 we see that $d/\overline{Tg}] \in T(S')$. It follows from the remark in Sec. 4.2 that $\varphi = Tf_1$, where f_1 is concentrated in the points $n + \frac12$, with n an integer. However, $d/\overline{Tg_1} = \overline{Tk}$, where k is the regular distribution given by

$$k(z) = \int_0^1 \frac{d\,dw}{\overline{(Tg_1)(z,w)}}.$$

Here we use that if $\psi: \mathbb{R}^2 \to \mathbb{C}$ is absolutely integrable over

$[0,1]^2$ and satisfies $T_1^{(1)}\psi = R_1^{(2)}\psi, T_1^{(2)}\psi = \psi$, then $\psi = T\psi_0$, where $\psi_0(z) = \int_0^1 \psi(z,w)\,dw$ (this follows from Theorem 3.6). Now the fact that f_0 and k are regular and f_1 is concentrated in the points $n + \frac{1}{2}$ with integer n leads to a contradiction, unless $\varphi \equiv 0$. This shows that $Tf_0 = d / \overline{Tg}$.

We finally show that $\mathscr{F} f_0 = f_0$. We therefore observe that $Tf_0 = d / \overline{Tg} = d / \overline{T\mathscr{F}^* g} = de^{2\pi i z w}/(\overline{Tg})(w, -z)$ by 3.3,(6). Hence $(T\mathscr{F} f_0)(z,w) = e^{-2\pi i z w}(Tf_0)(z,w)$ $= d/(\overline{Tg})(z,w) = (Tf_0)(z,w)$, and the result follows from 3.1

Remarks: (1) It is easy to see that $\Sigma_n f_0(z-n)e^{-2\pi i n w}$ converges uniformly and absolutely (to $1/\overline{Tg}$ on any compact set not containing points (z,w) for which $z - \frac{1}{2}$ is an integer. In case $z = 0$ we thus find the Fourier series for $1/\theta_3$:

$$\left(\sum_{n=-\infty}^{\infty} e^{-\pi n^2 + 2\pi i n w}\right)^{-1} = d^{-1} \sum_{n=-\infty}^{\infty} f_0(n) e^{2\pi i n w}$$

(compare Ref. 29, Chap. 21, Miscellaneous Examples, 14). Note also that $d^{-1} f_0(r-s)$ for integers r and s is the limit of the (rs)th element of the inverse of the matrix $(e^{-\pi(k-l)^2})_{k,l=-n,\dots,n}$ if $n \to \infty$; these matrices occur in the study of the inverses of discrete Gauss transforms.[37] (2) As a consequence of Theorem 4.4 we have that the set $(G(n,m))_{(n,m)\neq(0,0)}$ is complete in $L^p(\mathbb{R})$ if $1 \leqslant p < \infty$, and not complete in $L^\infty(\mathbb{R})$. This generalizes the result in Ref. 2 where the case $p = 2$ was treated. (3) Let $f \in L^1(\mathbb{R})$ so that for every integer m

$$c_{nm} := (-1)^{nm} d^{-1}(f, R_{-m} T_{-n} f_0) \to 0$$

if $n \to \infty$. Since for all integers k and l

$$d^{-1} \Sigma_{n,m} (-1)^{nm} (G(n,m), G(k,l)) R_{-m} T_{-n} f_0 = \overline{G(k,l)},$$

where the series converges boundedly, we have

$$\sum_{n,m} c_{nm} (G(n,m), G(k,l)) = (f, G(k,l)).$$

It follows from Ref. 4, Theorem 4.1 and 4.1, Remark 1 that $f = \Sigma_{n,m} c_{nm} G(n,m)$ with convergence in S′ sense.

4.5. In Ref. 2 it is shown that the set $G(n,m)$ with $(n,m) \neq (0,0)$, $(n,m) \neq (k_0, l_0)$ is not complete in $L^2(\mathbb{R})$ if $(k_0, l_0) \neq (0,0)$. We generalize this result fo $L^p(\mathbb{R})$ as follows. If k and l are integers and $F \in S'$, then $R_{-k}^{(1)} R_{-l}^{(2)} TF = T(R_{-k} T_{-l} F)$ by 3.3. So, if $P(z,w) = \Sigma_{k,l} c_{lk} e^{2\pi i k z + 2\pi i l w}$ is a trigonometric polynomial, then

$$T(\Sigma_{k,l} c_{lk} R_{-k} T_{-l} f_0) = P / \overline{Tg}.$$

Taking P such that $P(\frac{1}{2},\frac{1}{2}) = 0$, we get $P/\overline{Tg} \in L^\infty([0,1]^2)$. Since T maps $L^2(\mathbb{R})$ onto $L^2([0,1]^2) \supset L^\infty([0,1]^2)$, we see that there is an $f \in L^2(\mathbb{R})$ such that $Tf = P/\overline{Tg}$, i.e., an $f \in L^2(\mathbb{R})$ with $(f, G(k,l)) = (-1)^{kl} c_{kl}$. Taking $c_{00} = 1$, $c_{k_0 l_0} = (-1)^{k_0 + l_0 + 1}$, $c_{kl} = 0$ otherwise, we get an $f \in L^2(\mathbb{R})$ with $(f, G(n,m)) = 0$ for $(n,m) \neq (0,0)$, $(n,m) \neq (k_0, l_0)$. This f is given by

$$f = f_0 + (-1)^{k_0 + l_0 + 1} R_{-k_0} T_{-l_0} f_0,$$

and it can be shown that $f \in L^p(\mathbb{R})$ for all p with $1 \leqslant p \leqslant \infty$. Hence, the collection $G(n,m)$ with $(n,m) \neq (k_0, l_0)$ is not complete in $L^p(\mathbb{R})$ for $1 \leqslant p \leqslant \infty$. In case $l_0 = 0$ we can even show that $f(x) = ((-1)^{k_0+1} e^{2\pi i k_0 x} + 1) f_0(x) = O(1/|x|)(|x| \to \infty)$ and that f is continuous. The discontinuities of f_0 (which occur at the points x for which $x - \frac{1}{2}$ is an integer) are annihilated by the factor $(-1)^{k_0+1} e^{2\pi i k_0 x} + 1$ in this case. And if $k_0 = 0$ we get an f for which $(\mathscr{F} f)(x) = O(1/|x|)(|x| \to \infty)$ and $\mathscr{F} f$ is continuous. One can carry this process further, and it is easy to see that deletion of more coherent states gives rise to the existence of smoother functions perpendicular to all coherent states but the deleted ones. We note that the latter assertion is, to some extent, also true if one takes more general functions g than the Gauss function.

4.6. We now discuss an expansion which is in some sense dual to Gabor's expansion. Let A be a finite set of lattice points, and let $F \in S'$ satisfy $(F, G(n,m)) = 0$ for $(n,m) \notin A$. It is clear that F is of the form

$$F = G + d^{-1} \sum_{(n,m)\in A} (-1)^{nm} (F, G(n,m)) R_{-m} T_{-n} f_0,$$

where $(G, G(n,m)) = 0$ for all n and m [this is a consequence of the formula $(R_{-m} T_{-n} f_0, G(k,l)) = (-1)^{nm} d \delta_{nk} \delta_{ml}$.] In case $F \in S'$ is regular and satisfies $\Sigma_{n,m} |(F, G(n,m))| < \infty$, then we can take for A all of $\mathbb{Z} \times \mathbb{Z}$ and $G \equiv 0$:

$$(*) F = d\, \Sigma_{n,m} (-1)^{nm} (F, G(n,m)) R_{-m} T_{-n} f_0.$$

For the proof we use that either function in (*) is regular (cf Theorem 4.4) and Ref. 4, Theorem 4.1. Of course, (*) is false in general if $F \in S'$ is not regular.

We shall now give a connection with Whittaker's result[38] about interpolation over the Gaussian integers. Therefore, let $F \in S'$ be regular and assume that $\Sigma_{n,m} |(F, G(n,m))| < \infty$ so that the expansion in (*) holds for F. Applying the Bargmann transform to both sides and using that

$$(Bf_0)(z) = -\theta_1(z) e^{\frac{1}{2}\pi z^2} / 2\pi z,$$

$$B(R_{-m} T_{-n} f_0)(z) = (-1)^{nm} e^{-\frac{1}{2}\pi(n^2+m^2) + \pi(n+im)z} \times (Bf_0)(z + im - n),$$

$$\theta_1(z + im - n) e^{\frac{1}{2}\pi(z+im-n)^2} = (-1)^{n+m+nm} \theta_1(z) e^{\frac{1}{2}\pi z^2} e^{-\frac{1}{2}\pi(n^2+m^2) - \pi(n+im)z},$$

$$(F, G(n,m)) = e^{-\frac{1}{2}\pi(n^2+m^2)} (BF)(n - im),$$

we get

$$B(R_{-m} T_{-n} f_0)(z) = \frac{(-1)^{n+m+1} \theta_1(z) e^{\frac{1}{2}\pi z^2}}{2\pi(z + im - n)},$$

so that

$$(BF)(z) = -d\theta_1(z) e^{\frac{1}{2}\pi z^2} \sum_{n,m} (-1)^{n+m+nm} \times \frac{(BF)(n - im)}{z - n + im} e^{-\frac{1}{2}\pi(n^2+m^2)}.$$

This is a slight generalization of the result of Ref. 38, because there only functions $f = BF$ are admitted that have growth $< (2, \pi/2)$, while we admit certain functions with growth $= (2, \pi/2)$ (cf. 2.7).

The expansion discussed here can be used to generalize a result of Iyer and Pfluger[39] about entire functions of growth $< (2, \pi/2)$ which are bounded at the lattice points. Let $\varphi = Bf$, where $f \in S$, and assume that φ is bounded at the lattice points. We shall show that φ is constant, and to that end we suppose that φ has type $\pi/2$. Since $z^k\varphi(z) = (B(Q-iP)^k f)(z)\in B(\mathrm{S})$, we see that

$$z^k\varphi(z) = -d\theta_1(z)e^{\frac{1}{2}\pi z^2}\sum_{n,m}(-1)^{n+m+nm} \times \frac{(n+im)^k\varphi(n+im)}{z-n-im}e^{-\frac{1}{2}\pi(n^2+m^2)}$$

for all $k=0,1,\cdots$. Since $\varphi\in B(\mathrm{S})$ we know that $\Sigma_{n=0}^{\infty}|a_n||z|^n = O(\exp(\frac{1}{2}\pi|z|^2)(1+|z|)^{-K})$ for all $K>0$ [here we write $\varphi(z) = \Sigma_{n=0}^{\infty}a_n z^n$]. It follows easily that

$$\varphi^2(z) = \sum_{k=0}^{\infty}a_k z^k\varphi(z) = -d\theta_1(z)e^{\frac{1}{2}\pi z^2}\Sigma_{n,m}(-1)^{n+m+nm}\frac{\varphi^2(n+im)}{z-n-im} \times e^{-\frac{1}{2}\pi(n^2+m^2)}.$$

Now the right-hand side has growth $\leqslant(2,\pi/2)$, while the left-hand side has growth $=(2,\pi)$.[40] We conclude that φ has type $<\pi/2$, whence φ is constant by the result of Iyer and Pfluger.

A trivial extension of this theorem is: If $\varphi\in B(\mathrm{S})$ satisfies $\varphi(n+im) = O((n+im)^K)$ for some $K=0,1,\cdots$, then φ is a polynomial of degree $\leqslant K$. As a consequence we have: If $f\in\mathrm{S}$ and $(f,G(n,m)) = O((n^2+m^2)^{K/2}\exp(-\frac{1}{2}\pi(n^2+m^2))$, then f is of the form $\Sigma_{k=0}^{K}a_k\psi_k$. We observe (cf. 4.5) that the assumption "$f\in\mathrm{S}$" cannot be weakened to a condition in which boundedness of only finitely many of the functions Q^kP^lf is required.

With methods similar to the ones used above it can be shown that if $f\in\mathrm{S}$, $0\leqslant\epsilon_0<1$, $(f,G(n,m)) = O[\exp(-\frac{1}{2}\pi\epsilon(n^2+m^2))]$ for all ϵ, $0<\epsilon<\epsilon_0$, then $f\in N_\alpha(S^*)$, where $\alpha = -\frac{1}{2}\log(1-\epsilon_0)$.

4.7. Translating Ref. 4, Theorem 4.7 by using the Bargmann transform, we get the following theorem:

Theorem: Let $f\in B(\mathrm{S}')$, i.e., let f be an entire function such that

$$f(z)\exp(-\tfrac{1}{2}\pi|z|^2) = O((1+|z|)^N)$$

for some $N>0$. There exists a double sequence $(c_{nm})_{nm}$ satisfying $c_{nm} = O((1+n^2+m^2)^K)$ for some $K>0$ such that

$$f(z) = \Sigma_{n,m}c_{nm}\exp(-\tfrac{1}{2}\pi(n^2+m^2))e^{\pi(n+im)z}.$$

The convergence of the series is such that

$$\left| f(z) - \sum_{n^2+m^2<L} c_{nm}\exp(-\tfrac{1}{2}\pi(n^2+m^2)) \times e^{\pi(n+im)z}\right| \frac{\exp(-\frac{1}{2}\pi|z|^2)}{(1+|z|)^M} \to 0$$

uniformly in $z\in\mathbb{C}(L\to\infty)$ for some $M>0$. Write $f(z) = e^{\frac{1}{2}\pi|z|^2}\varphi(z)$. If $\varphi\in L^2(\mathbb{C})$ and

$$\int_{\mathbb{C}}\left|\frac{\varphi(z)}{z-w}\right| dz\to 0 \quad (|w|\to\infty),$$

then there exists a unique sequence $(c_{nm})_{nm}$ with $c_{nm}\to 0$ $(n^2+m^2\to\infty)$ such that the above expansion holds; these c_{nm}'s are given by

$$c_{nm} = -(-1)^{n+m+nm}d^{-1}\int_{\mathbb{C}} e^{\frac{1}{2}\pi(|z|^2-\bar{z}^2)}\frac{\theta_1(\bar{z})\varphi(z)}{2\pi(\bar{z}-n-im)}dz.$$

Proof: Take $h\in\mathrm{S}'$ such that $f=Bh$. This h has a Gabor expansion $h = \Sigma_{n,m}c_{nm}G(n,m)$ which converges in S$'$ sense. Now apply B to both sides and use 2.7 and 2.8. The convergence is in the sense indicated by Ref. 13, Sec. 4.1.

Assume in addition that $\varphi\in L^2(\mathbb{C})$ and that

$$\int_{\mathbb{C}}\left|\frac{\varphi(z)}{z-w}\right| dz\to 0 \quad (|w|\to\infty),$$

where φ is as in the theorem. Put $f_\alpha = BN_\alpha h$ for $\alpha>0$ (cf. 2.4). As $N_\alpha h\in S$ for $\alpha>0$ we have $N_\alpha h = \Sigma_{n,m}c_{nm}(\alpha)G(n,m)$, where

$$c_{nm}(\alpha) = (-1)^{nm}d^{-1}(N_\alpha h, R_{-n}T_{-m}f_0)$$

by 4.4, Remark 3. By 4.6 and 2.8 we can write this as

$$c_{nm}(\alpha) = -(-1)^{n+m+nm}d^{-1}\int_{\mathbb{C}} e^{-\pi|z|^2}f_\alpha(z)\overline{\frac{\theta_1(z)e^{\pi z^2/2}}{2\pi(z+im-n)}}dz.$$

Since $(BN_\alpha h)(z) = \exp(-\frac{1}{2}\alpha+\frac{1}{2}\pi|ze^{-\alpha}|^2)\varphi(ze^{-\alpha})$ by 2.8(3) and $\theta_1(z) = O(\exp(\pi(\mathrm{Im}z)^2))$, we get

$$\lim_{\alpha\downarrow 0}c_{nm}(\alpha) = -(-1)^{n+m+nm}d^{-1}\int_{\mathbb{C}} e^{\frac{1}{2}\pi(\bar{z}^2-|z|^2)}\frac{\theta_1(\bar{z})\varphi(z)}{2\pi(\bar{z}-n-im)}dz$$

boundedly in (n,m) as $\int_{\mathbb{C}}|\varphi(z)|/|z-w|\,dz$ is bounded in w. We also have $c_{nm}\to 0$ and

$$f(z) = \Sigma_{n,m}c_{nm}\exp(-\frac{\pi}{2}(n^2+m^2))e^{\pi(n+im)z}.$$

Uniqueness follows from Ref. 4, Theorem 4.1, for it follows from the assumptions about φ and 2.7 that $h\in L^2(\mathbb{R})$.

4.8. We give an explicit formula for the unique coefficients $c_{nm}(x,y)$ in the Gabor expansion $\Sigma_{n,m}c_{nm}(x,y)G(n,m)$ of $G(x,y)$, with $c_{nm}(x,y)\to 0(n^2+m^2\to\infty)$. Using the formulas

$$c_{nm}(x,y) = -(-1)^{nm}d^{-1}(G(x,y),R_{-m}T_{-n}f_0),$$

$$(T_aR_bG(x,y),f_0) = \exp(\pi i(ay-xb))(G(x-a,y-b),f_0),$$

$$(G(x-a,y-b),f_0) = \exp(-\tfrac{1}{2}\pi(x-a)^2-\tfrac{1}{2}\pi(y-b)^2)\,\overline{(Bf_0)(x-a-i(y-b))},$$

$$(Bf_0)(z) = -d\theta_1(z)e^{\frac{1}{2}\pi z^2}/2\pi z, \quad \overline{\theta_1(z)} = \theta_1(\bar{z}),$$

we get

$$(G(x,y),R_{-b}T_{-a}f_0) = -d\exp(-\pi(y-b)^2-\pi i(xy+ab)) \times\theta_1(x-a+i(y-b))/2\pi(x-a+i(y-b)).$$

Hence,

$$c_{nm}(x,y) = \exp(-\pi(y-m)^2-\pi ixy) \times\theta_1(x-n+i(y-m))/2\pi(x-n+i(y-m))$$

for all real x and y and all integers n and m.

If we denote for $f\in S$ by $c_{nm}(f)$ the unique Gabor coefficients in the expansion $f = \Sigma_{n,m}c_{nm}(f)G(n,m)$, then we see from Ref. 3, 27.12.1.5 that

$$c_{nm}(f) = \int\int(f,G(x,y))c_{nm}(x,y)\,dx\,dy.$$

In particular,

$$\iint (G(a,b),G(x,y))c_{nm}(x,y)\,dx\,dy = c_{nm}(a,b)$$

which shows that c_{nm} is an "eigenfunction" of the kernel operator with kernel $(G(a,b),G(x,y))$.

4.9. We next consider the completeness problem with general lattices in phase plane where the cells have area equal to one. Such a lattice can be described by six real numbers $a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}$, with $a_{11}a_{22} - a_{12}a_{21} = 1$, by putting

$$(x_{n,m},y_{n,m}) = (a_{11}n + a_{12}m + a_{13}, a_{21}n + a_{22}m + a_{23})$$

for integers n and m. We have according to Ref. 3, 27.12.2.1, for all lattice points $(a,b) = (x_{n,m},y_{n,m})$

$$(\Gamma_{(A,\psi)}f,R_{-b}T_{-a}(\Gamma_{(A,\psi)}g)_-)$$
$$= \tfrac{1}{2}e^{-\pi iab}W(\tfrac{1}{2}a,\tfrac{1}{2}b;\ \Gamma_{(A,\psi)}f,\Gamma_{(A,\psi)}g)$$
$$= \tfrac{1}{2}e^{-\pi iab}W(\tfrac{1}{2}n,\tfrac{1}{2}m;\ f,g) = (-1)^{nm}e^{-\pi iab}(f,R_{-m}T_{-n}g_-)$$

for $f\in S$, $g\in S$ (cf. 4.1 for the definition of W). Here $\Gamma_{(A,\psi)}$ is the special transform introduced in Ref. 3, 27.3.8 and 9 associated with the matrix

$$\begin{bmatrix} a_{22} & -ia_{21} & -ia_{23}/2 \\ ia_{21} & a_{11} & a_{13}/2 \\ 0 & 0 & 1 \end{bmatrix}.$$

The above formula can easily be generalized to the case that $f\in S^*$, $g\in S$ or $f\in S'$, $g\in S$ or $f\in L^2(\mathbb{R})$, $g\in L^2(\mathbb{R})$. Hence, characterizing the set of all $f\in S'$ for which $(f,R_{-b}T_{-a}g) = 0$ for all lattice points $(a,b) = (x_{n,m},y_{n,m})$ amounts to characterizing all $F\in S'$ for which $(F,R_{-m}T_{-n}G_-) = 0$ for all integers n and m, where $G = \Gamma^{-1}_{(A,\psi)}g$.

As an example take $g(t) = 2^{1/4}\exp(-\pi t^2)$ and $a_{13} = a_{23} = 0$. We have

$$(\Gamma^{-1}_{(A,\psi)}g)(z) = 2^{1/4}a_{22}^{-1/2}\exp(-\pi z^2(1 + ia_{21}a_{22})/a_{22}^2)$$

if $a_{12} = 0$, and

$$(\Gamma^{-1}_{(A,\psi)}g)(z)$$
$$= 2^{1/4}(a_{22} - ia_{12})^{-1/2}\exp\left(-\pi z^2\left[\frac{a_{12}}{a_{22}^2(a_{12} + ia_{22})} + \frac{ia_{11}}{a_{12}}\right]\right)$$

if $a_{12}\neq 0$ (the choice of the square root is determined by ψ; cf. Ref. 3, 27.3). Now if $\mathrm{Re}\gamma > 0$, $g_\gamma(t) = \exp(-\pi\gamma t^2)$, then a calculation shows that

$$(Tg_\gamma)(z,w) = \exp(-\pi\gamma z^2)\vartheta_3(\pi(w + i\gamma z);e^{-\pi\gamma}).$$

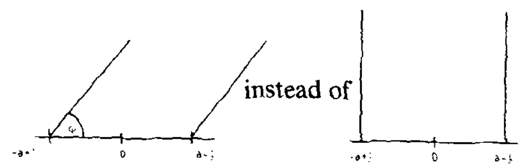
It follows that Tg_γ has zeros at the points $(n + \tfrac{1}{2}, m + \tfrac{1}{2})$ with integers n and m and nowhere else. Now the remark in Sec. 4.2 gives an indication of the general form of all $F\in S'$ with $(F,R_{-m}T_{-n}g_\gamma) = 0$ for all n and m. In this case, however, we can do better by using the Bargmann transform. Let $\gamma\neq 0$. We want to characterize all $F\in S'$ with $(F,G(n + \gamma m,m)) = 0$ for all integers n and m. That is, we must find all $F\in S'$ such that $(BF)(n + \gamma m - im) = 0$ for all integers n and m. Note that the function $\vartheta_1(\pi z,e^{-\pi i\gamma - \pi})$ has zeros at the points $n + \gamma m - im$, that there is an $M > 0$ such that

$$|\vartheta_1(\pi z,e^{-\pi i\gamma - \pi})| \leqslant M\exp(\pi(\mathrm{Im}z)^2) \quad (z\in\mathbb{C}),$$

and that there is a $K > 0$ such that

$$|\vartheta_1(\pi z,e^{-\pi i\gamma - \pi})| \geqslant K\exp(\pi v^2),$$

where $z = -\gamma v + iv + a - \tfrac{1}{2}$ with v real and a an even integer (compare Ref. 7, 2.6). Now we can proceed as in the proof of Ref. 7, 2.11 to conclude that $(BF)(z) = e^{\frac{1}{2}\pi z^2}\vartheta_1(\pi z,e^{-\pi i\gamma - \pi})P(z)$, where P is a polynomial. The only difference from the proof quoted is that we need a theorem of Phragmén–Lindelöf type for regions of the form ($\varphi = \arctan\gamma$):



With the aid of the Schwarz–Christoffel formula one can construct a conformal mapping τ (continuous at the boundaries) that maps the second region onto the first one such that $\tau(-a + \tfrac{1}{2}) = -a + \tfrac{1}{2}$, $\tau(a - \tfrac{1}{2}) = a - \tfrac{1}{2}$, $\tau(i\infty) = (-\gamma + i)\cdot\infty$, and such that $|\tau(x + iw)/w|$ tends to a finite limit $\neq 0$ uniformly in x, $|x| \leqslant a - \tfrac{1}{2}$ if $w\to\infty$. Hence, the Phragmén–Lindelöf theorem for the rectangular region can be modified in such a way that we get the required bounds on $e^{-\frac{1}{2}\pi z^2}(BF)(z)/\vartheta_1(\pi z,e^{-\pi i\gamma - \pi})$. From this we can derive a similar characterization as in Ref. 7, 2.12. We note that the more general problem of characterizing all $F\in S'$ with $(F,G(n\alpha + \gamma m,m\alpha^{-1})) = 0$ for all integers n and m can be handled similarly ($\alpha > 0$). Finally, if we have a general lattice $(a,b) = \alpha n(\cos\theta_1,\sin\theta_1) + \beta m(\cos\theta_2,\sin\theta_2)$ with $\alpha\beta\sin(\theta_2 - \theta_1) = 1$, and $(F,G(a,b)) = 0$, then we can use that

$$(BF)(z) = c(BN_{e^{-i\theta_1}}F)(e^{i\theta_1}z) \quad (z\in\mathbb{C})$$

for some c, $|c| = 1$ (cf. 2.8 and the references given there). We then get $(F,G(a,b)) = 0$ at the lattice points if and only if $(N_{e^{-i\theta_1}}F,G(n\alpha + \beta m\cos(\theta_2 - \theta_1), m\alpha^{-1})) = 0$ for all integers n and m.

4.10. We finally show that every $F\in S^*$ can be expanded in an S^* convergent Gabor series. Our proof consists of a suitable variation of the argument used in the proof of Ref. 4, Theorem 4.7, where the S'-case was considered.

We start with the observation that any $F\in S^*$ which can be written in the form

$$(*)F = \sum_{n=0}^{\infty} a_n(Q + iP)^n F_n,$$

where $F_n\in L^2(\mathbb{R})$, $\|F_n\| \leqslant 1$, $a_n = O(\exp(-\tfrac{1}{2}n\log n - Bn))$ for all $B > 0$ has a Gabor representation. Indeed, as $F_n\in L^2(\mathbb{R})$ we can find $(c_{kl}^{(n)})_{k,l}$ such that

$$F_n = \Sigma_{k,l}c_{kl}^{(n)}G(k,l)$$

(convergence in S' sense). Here we may assume in addition that

$$|c_{kl}^{(n)}| \leqslant C((\log|k|)^{1/2} + (\log|l|)^{1/2})$$

for some $C>0$ independent of n (cf. Ref. 4, Theorem 4.6). Hence, since $(Q+iP)G(x,y)=(x+iy)G(x,y)$,

$$F=\sum_{n=0}^{\infty} a_n\,(\Sigma_{k,l} c_{kl}^{(n)}(k+il)^n G(k,l)).$$

Now we note that

$$\sum_{n=0}^{\infty} |a_n c_{kl}^{(n)}(k+il)^n| \leqslant C\sum_{n=0}^{\infty} a_n\,|k+il|^{n+1}=f(|k+il|),$$

where $f(z):=\Sigma_n a_n z^{n+1}$. It follows from the assumptions on $(a_n)_n$ and Ref. 21, Theorem 2.2.10 that f has growth $\leqslant(2,0)$. Hence

$$\sum_{n=0}^{\infty} |a_n c_{kl}^{(n)}(k+il)^n| = O(\exp(\epsilon(k^2+l^2)))$$

for all $\epsilon>0$. We conclude that

$$F=\Sigma_{k,l}\Big(\sum_{n=0}^{\infty} a_n c_{kl}^{(n)}(k+il)^n\Big)G(k,l),$$

with convergence in S^* sense [this follows from the fact that for every $f\in S$ there is an $\epsilon>0$ such that $(f,G(k,l)) = O(\exp(-\epsilon k^2+l^2))$].

Our next aim is to show that any $F\in S^*$ can be written in the form (*) with $F_n\in L^2(\mathbb{R})$, $\|F_n\|\leqslant 1$, $a_n = O(\exp(-\frac{1}{2}n\log n-Bn))$ for all $B>0$. So, let $F=\Sigma_{k=0}^{\infty} c_k\psi_k\in S^*$, where $c_k=O(e^{k\epsilon})$ for all $\epsilon>0$. Let k_0, k_1, $k_2,\cdots$ be a sequence of integers with $0=k_0\leqslant k_1\leqslant k_2\leqslant\cdots$, and let

$$e_{k,l}=\pi^{\frac{1}{2}l}c_k((k+l)!/k!)^{-1/2}$$

for $k=k_l,k_l+1,\cdots,k_{l+1}-1$, $l=0,1,\cdots$. The definition of the $e_{k,l}$'s is such that

$$\sum_{k=k_l}^{k_{l+1}-1} c_k\psi_k=(Q+iP)^l\sum_{k=k_l}^{k_{l+1}-1} e_{k,l}\psi_{k+l}.$$

We have to choose the k_l's such that for all $B>0$,

$$\Big\|\sum_{k=k_l}^{k_{l+1}-1} e_{k,l}\psi_{k+l}\Big\|_2 = \Big(\sum_{k=k_l}^{k_{l+1}-1}|e_{k,l}|^2\Big)^{1/2}=O(\exp(-\tfrac{1}{2}l\log l-Bl)).$$

Equivalently, we want the k_l's such that for all $B>0$,

$$A_l:=\sum_{k=k_l}^{k_{l+1}-1}\frac{k!|c_k|^2}{(k+l)!}=O(\exp(-l\log l-Bl)).$$

Denote for $l=0,1,\cdots$ by $f(l)$ the integer with the property that $|c_k|\leqslant e^l(k=0,1,\cdots,f(l)-1)$, $|c_{f(l)}|>e^l$ [we may assume that $f(l)$ exists, otherwise $(c_k)_k$ is bounded and so $F\in S'$]. Now $f(l)/l\to\infty$, for otherwise we could find an $M>0$ and integers l_1, $l_2,\cdots$ with $l_k\to\infty$ such that $f(l_k)\leqslant Ml_k$, $|c_{f(l_k)}|>\exp(l_k)$ for all k, contradicting $c_l=O(\exp(l\epsilon))$ if $\epsilon<1/M$. Now put $k_{l+1}:=\min(l^2,f(l))$ for $l=0,1,2,\cdots$. We have

$$A_l\leqslant e^{2l}\sum_{k=k_l}^{k_{l+1}-1}\frac{k!}{(k+l)!},$$

and it follows easily from Stirling's formula that there is an $M>0$ such that

$$k!/(k+l)!\leqslant M\exp(l-l\log l-l\log(1+k_l/l)),$$

for $l=0,1,\cdots,k=k_l,\ldots,k_{l+1}-1$. Hence,

$$A_l\leqslant Me^{3l}(k_{l+1}-k_l)\exp(-l\log l-l\log(1+k_l/l)).$$

Since $k_l/l\to\infty, k_{l+1}-k_l=O(l^2)$, it follows that $A_l=O(\exp(-l\log l-Bl))$ for all $B>0$, and this completes the proof.

We mention a consequence: It follows as in 4.7 that for every entire f of growth $\leqslant(2,\pi/2)$ there exists a double sequence $(c_{nm})_{nm}$ with $c_{nm}=O[\exp(\epsilon(n^2+m^2))]$ for all $\epsilon>0$ such that

$$f(z)=\Sigma_{n,m}c_{nm}\exp(-\frac{\pi}{2}(n^2+m^2))e^{\pi(n+im)z}.$$

[1] V. Bargmann, P. Butera, L. Girardello, J. R. Klauder, Rep. Math. Phys. **2**, 221 (1971).
[2] A. M. Perelomov, Theor. Math. Phys. **6**, 156 (1971).
[3] N. G. de Bruijn, Nieuw Arch. Wisk. **21**, 205 (1973).
[4] A. J. E. M. Janssen, "Gabor representation of generalized functions," (to appear in J. Math. Anal. Appl.).
[5] H. Bacry, A. Grossman, and J. Zak, Phys. Rev. B **12**, 1118 (1975).
[6] M. J. Bastiaans, "Gabor's expansion of a signal into Gaussian elementary signals," (to appear).
[7] A. J. E. M. Janssen, "Weighted Wigner distributions vanishing on lattices," (to appear in J. Math. Anal. Appl.).
[8] J. von Neumann, "Mathematical Foundations of Quantum Mechanics" (Princeton U. P., Princeton, N.J., 1955), Chap. 5, Sec. 4.
[9] See Refs. 1 and 2; also the case that $f\in L^2(\mathbb{R})$, $\alpha\beta<1$ was considered in both references.
[10] D. Gabor, J. Inst. Elec. Eng. **93**, 429 (1946).
[11] After completing the manuscript the author was informed by J. Zak that the proof of this property has been announced before in J. Zak, Phys. Rev. B **12**, 3023 (1975), and that a proof (essentially the same as the one given by the author) will appear in J. Math. Phys. See also Ref. 5 for an important special case. Zak also pointed out that in Ref. 2 and in M. Boon and J. Zak, Phys. Rev. B **18**, 6744 (1978), expansions of the Gabor type were considered.
[12] V. Bargmann, Comm. Pure Appl. Math. **14**, 187 (1961).
[13] V. Bargmann, Comm. Pure Appl. Math. **20**, 1 (1967).
[14] I. M. Gel'fand, G. E. Shilov, *Generalized Functions*, Vol. 2 (Academic, New York, 1968), Chap. IV, Sec. 2.3.
[15] A. J. E. M. Janssen, "Applications of the Wigner distribution to harmonic analysis of generalized stochastic processes," MC-tract 114, Appendix 1 (Amsterdam, 1979).
[16] A. J. E. M. Janssen, Proc. K. Ned. Akad. Wet. A **82**, 283 (1979).
[17] See Ref. 3, 27.6.
[18] M. Reed and B. Simon, *Methods of modern mathematical physics*, Vol. 1 (Academic, New York, 1972), Appendix to Chap. V, Sec. 3. The Hermite coefficients of a tempered distribution have at most polynomial growth.
[19] See Ref. 3, Sec. 27.3 for generalizations.
[20] See Ref. 15, Sec. 4.
[21] R. P. Boas, *Entire Functions* (Academic, New York, 1954).
[22] A. J. E. M. Janssen,. "Bargmann transform, carriers of generalized functions, and the Wigner distribution," (to appear). See also Ref. 13, Sec. 6.
[23] J. Zak, Phys. Rev. Lett. **19**, 1385 (1967).
[24] J. Zak, Phys. Rev. **168**, 686 (1968).
[25] J. Zak, Solid State Phys. **27**, 1 (1972).
[26] J. Zak, Phys. Rev. **187**, 1803 (1969).
[27] J. Igusa, *Theta Functions* (Springer, New York, 1972).
[28] A. Zygmund, *Trigonometric Series*, Vol. II (Cambridge U. P., 1968).
[29] E. T. Whittaker, G. N. Watson, *A Course of Modern Analysis*, 4th ed. (Cambridge U. P., Cambridge, 1927).
[30] The author thanks Ralph Howard for proving this theorem; see also Ref. 11.
[31] T. Raeb and P. V. Reichelderfer, *Continuous Transformations in Analysis* (Springer, New York, 1955).
[32] R. Price and E. M. Hofstetter, IEEE Trans. Inf. Theory **11**, 207 (1965).
[33] N. G. de Bruijn, in *Inequalities*, edited by O. Shisha (Academic, New

[33]N. G. de Bruijn, in *Inequalities*, edited by O. Shisha (Academic, New York, 1967), p. 57.
[34]F. Trèves, *Topological vector spaces, distributions and kernels* (Academic, New York, 1972).
[35]S. Łojasiewicz, Studia Math. **18**, 87 (1959).
[36]This can be justified by using the theorems in Ref. 16, Sec. 5.
[37]I. J. Good, Applicable Anal. **9**, 205 (1979).
[38]J. M. Whittaker, *Interpolatory Function Theory* (Cambridge, U.P., Cambridge, 1935), Chap. V, Sec. 12, L_{501}.
[39]V. G. Iyer, J. London Math. Soc. **13**, 91 (1938); A. Pfluger, Proc. London Math. Soc. **42**, 305 (1939). Both authors prove (by different methods) that any entire function of growth $<(2,\pi/2)$ which is bounded at the lattice points $n + im$ must be a constant. The result we prove is more general in the sense that certain functions of growth $=(2,\pi/2)$ are allowed.
[40]The proof given here consists of a suitable adaptation of Iyer's proof of the result stated in Ref. 39.

Valleys and fall lines on a Riemannian manifold [a)]

D. J. Rowe and A. Ryman [b)]
Department of Physics, University of Toronto, Toronto, Ontario, M5S 1A7, Canada

(Received 28 September 1981; accepted for publication 11 December 1981)

The concepts of fall lines, valleys, ridges, and general stationary paths are defined for a potential energy function on a Riemannian manifold. Some theorems governing their properties and relationships are derived. These concepts are of interest in the classical mechanics of constrained systems and in the theory of collective motions in many-body quantum mechanics.

PACS numbers: 02.40.Ky, 03.20. + i, 03.65. − w

1. INTRODUCTION AND DEFINITIONS

Recent attempts[1,2] to extract a collective subdynamics from the dynamics of a many-body system make use of the concepts of valleys and fall lines. The objective of this paper is to examine some of their properties and to exhibit the geometrical structures upon which they are based. Some aspects of this subject, in the context of the Earth's surface, were first addressed by Cayley[3] and Maxwell,[4] who also gave the first expression of the well-known theorem relating the numbers of pits, peaks, and passes on a two-dimensional surface.

Consider a Riemannian manifold M with metric g and a potential energy function $v: M \rightarrow \mathbb{R}$. For example, M might be a configuration space for a dynamical system and g an inertial mass tensor such that the kinetic energy is proportional to the Laplace–Beltrami operator

$$\Delta = \frac{1}{\sqrt{\det(g)}} \frac{\partial}{\partial x^i} \sqrt{\det(g)}\, g^{ij} \frac{\partial}{\partial x^j}. \tag{1}$$

Introduce the potential gradient $dv^{\#}$ and its negative, the force field

$$F = -dv^{\#}. \tag{2}$$

In terms of a coordinate chart (x^i), F is a vector at each point with components $F^i = -g^{ij}\partial v/\partial x^j$.

Definition (fall line): A fall line is an integral curve of the potential gradient.

A fall line is thus a line of steepest descent. Some fall lines are illustrated, for example, in Fig. 1 for the potential landscape model of Ref. 5 in which the potential is given in terms of Cartesian coordinates on the Euclidean plane by

$$v = y^2 + \tfrac{1}{2}x^2 - x^2 y. \tag{3}$$

This figure also shows the equipotential surfaces of v.

The potential (3) contains a minimum at the origin and a saddle point marked S in Fig. 1. It evidently has a valley connecting the two, a maximal path (the opposite of a valley) perpendicular to the valley at the minimum, and a ridge perpendicular to the valley at the saddle point. We refer to all such paths as stationary paths. They are illustrated for the potential landscape model in Fig. 2. One of our objectives is to identify these paths precisely. For example, it is tempting to conclude from Fig. 1 that the valley might be the fall line connecting the saddle point to the minimum; i.e., the line called by Maxwell[4] the "watercourse". This we shall show to be false.

Definition (minimal and stationary paths): A point m in M is said to be on a minimal path if the square magnitude $g(F, F)$ of the potential gradient at m is a proper minimum on the equipotential surface through m. If $g(F, F)$ is merely stationary, m is said to be on a stationary path.

This definition is phrased mathematically as follows. Let X be a vector tangent to an equipotential surface at m (see Fig. 3); i.e., $g(X, F) = 0$. For m on a stationary path, the directional derivative $Xg(F, F)$ of $g(F, F)$ in the direction X mush vanish. Now $Xg(F, F) = 2g(\nabla_X F, F)$, where ∇ is the covariant derivative for the Riemannian metric. Thus m is on a stationary path if

$$g(X, F) = 0 \Rightarrow g(\nabla_X F, F) = 0. \tag{4}$$

From Definition (2) of F we have $g(X, F) = -Xv$ and $g(\nabla_X F, F) = -(\nabla_X F)v$. Thus Condition (4) can also be written

$$Xv = 0 \Rightarrow (\nabla_X F)v = 0. \tag{5}$$

As illustrated in Fig. 3 a minimal path can be either a valley or a ridge or, in a multidimensional space, something in between.

Definition (valleys and ridges): Let X be a vector tangent to the equipotential surface at a point on a minimal path. A minimal path is said to be a valley if

$$g(\nabla_X F, X) < 0 \tag{6}$$

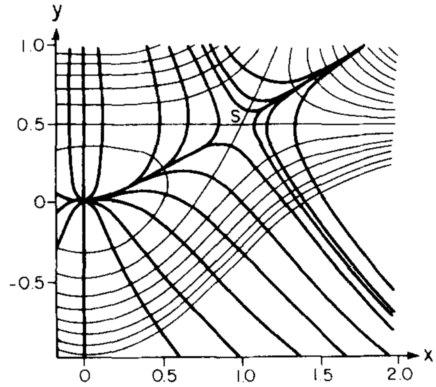


FIG. 1. Equipotential surfaces (light lines) and fall lines (heavy lines) for the potential energy function (3).

[a)]Work supported in part by the Natural Sciences and Engineering Research Council of Canada.

[b)]Present address: Ontario Hydro Dobson Research Laboratory, 800 Kipling Avenue, Toronto, Canada.

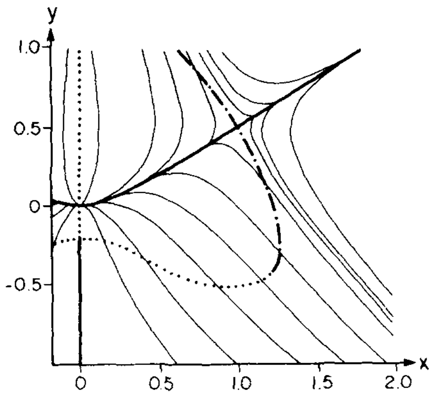


FIG. 2. Fall lines (light lines) and stationary paths for the potential energy function (2). Valleys are shown as heavy lines, a ridge as a dot-dashed line, and maximal paths as dotted lines.

and a ridge if

$$g(\nabla_X F, X) > 0 \tag{7}$$

for all such X.

Finally, we introduce the concept of a local normal mode vector. First we recall that the Hessian C is the bilinear form given by the second covariant derivative of the potential

$$C = \nabla\nabla v. \tag{8}$$

Definition (local normal mode vectors): Local normal mode vectors (η_α) are unit vectors for which $C(\eta_\alpha, \eta_\alpha)$ is stationary.

If (e_i) are orthonormal basis vectors at a point, i.e., $g(e_i, e_j) = \delta_{ij}$ and (C_{ij}) is the matrix $C_{ij} = C(e_i, e_j)$, the normal mode vectors at the point are obtained by an orthogonal transformation to new basis vectors which diagonalize (C_{ij}).

2. SOME THEOREMS AND RELATIONSHIPS

In examining the properties of fall lines and valleys and their relationships, we shall need to make use of the fact that F is the gradient of a scalar field and thus satisfies an integrability condition. Recall[6] that if X and Y are vectors and ω is a one-form then

$$(\nabla_X\omega)(Y) - (\nabla_Y\omega)(X) = d\omega(Y, X). \tag{9}$$

Now $d\omega = 0$ when $\omega = dv$. Hence we have

$$(\nabla_X dv)(Y) - (\nabla_Y dv)(X) = 0, \tag{10}$$

which implies

$$g(\nabla_X F, Y) = g(\nabla_Y F, X) \tag{11}$$

for any vectors X and Y.

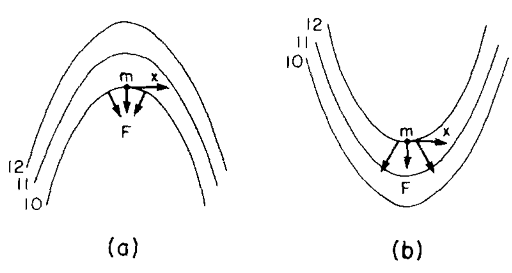


FIG. 3. A vector X tangent to the equipotential surface at a point m and hence orthogonal to the force F for (a) a valley and (b) a ridge.

Theorem 1: A point m of M is on a stationary path if and only if one of the local normal mode vectors is parallel to the potential gradient.

Proof: First observe that

$$C(X, F) = -g(\nabla_X F, F). \tag{12}$$

This can be seen, for example, by expanding both sides in terms of coordinates. From Eq. (4) it then follows that the point is on a stationary path if and only if

$$g(X, F) = 0 \Rightarrow C(X, F) = 0. \tag{13}$$

It now follows from the definition of local normal mode vectors that F parallel to a normal mode vector and X orthogonal to F implies $C(X, F) = 0$ and hence that the point is on a stationary path. Conversely, if $C(X, F)$ vanishes for all X orthogonal to F then, again by definition, F must be parallel to a normal mode vector. □

Theorem 2 (well known): The hyperplane spanned by the lowest frequency normal mode vectors at the potential minimum is tangent to all generic fall lines.

Proof: Let (x^i) be local normal coordinates at the potential minimum; i.e., the origin of the coordinate chart is the minimum and the tangents to the coordinate lines at the minimum are normal mode vectors. The Hessian matrix is then diagonal, $C_{ij} = C_i\delta_{ij}$, and the normal mode frequencies are given by $(\sqrt{C_i})$. In a neighborhood of the minimum we can make an expansion of the potential v in terms of these coordinates:

$$v = v(0) + \frac{1}{2}\sum_i C_i(x^i)^2 + \cdots. \tag{14}$$

Now let $(x^i(t))$ be the coordinates of a generic fall line passing through the minimum and parametrized by t such that it has a tangent at each point equal to the potential gradient $dv^\#$. Then

$$\frac{dx^i}{dt} = C_i x^i + O(x^2) \quad \text{(no summation)} \tag{15}$$

and hence

$$x^i(t) = \alpha_i e^{C_i t} + O(x^2). \tag{16}$$

It follows therefore that

$$\frac{dx^i}{dt}\bigg/\frac{dx^j}{dt} = \frac{\alpha_i C_i}{\alpha_j C_j} e^{(C_i - C_j)t}, \tag{17}$$

which vanishes at $t \to -\infty$, i.e., at $x^i(t) \to 0$, whenever $C_i > C_j$. This proves the theorem. □

We now prove two theorems concerning the curvatures of fall lines. First, since the force field F is tangent to the fall line at each point, observe that, provided F does not vanish, the condition that the fall line has vanishing curvature is that the covariant derivative $\nabla_F F$ of F in the direction F is parallel to F. Note that the component of $\nabla_F F$ parallel to F need not vanish, even if the fall line is straight, because its slope may be changing. In general one observes that, provided F does not vanish, the curvature of the fall line is in the direction of the component of $\nabla_F F$ orthogonal to F. A point at which F vanishes is called a critical point.

Let f: $M \to \mathbb{R}$ be the function

$$f = g(F, F). \tag{18}$$

Then for any vector X at a point in M, we have

$$Xf = 2g(\nabla_X F, F), \tag{19}$$

and, since $Xf = g(df^{\#}, X)$ and by Eq. (11) $g(\nabla_X F, F) = g(\nabla_F F, X)$, it follows that

$$df^{\#} = 2\nabla_F F. \tag{20}$$

Thus the curvature at a noncritical point is in the direction of the vector $df^{\#}_{\perp}$, where $df^{\#}_{\perp}$ is the component of $df^{\#}$ orthogonal to F.

Theorem 3: All fall lines have vanishing curvature at noncritical points on a stationary path. Conversely, if a fall line has vanishing curvature at a point, that point is on a stationary path.

Proof: On a stationary path, by definition, f is stationary with respect to variations orthogonal to F. We have from Eqs. (4) and (11),

$$g(X, F) = 0 \Rightarrow g(\nabla_X F, F) = g(\nabla_F F, X) = 0 \tag{21}$$

and thus $df^{\#}_{\perp} = 0$. Conversely, $g(X, F) = 0$ and $df^{\#}_{\perp} = 0$ implies that $g(\nabla_X F, F) = 0$ and hence that the point is on a stationary path. □

If a line has vanishing curvature at all points it is said to be geodesic. Similarly, if its curvature vanishes at a subset of points we say that it is locally geodesic at those points. Theorem 3 states then that fall lines are locally geodesic at noncritical points of intersection with a stationary path. The theorem leads directly to the following somewhat surprising lemma.

Lemma: If a stationary path is also a fall line it must be geodesic and conversely, if a fall line is geodesic it must also be a stationary path.

Theorem 4: In a neighborhood of a minimal path, the fall lines curve away from the minimal path.

Proof: Since, on a minimal path, the function f is minimum with respect to variations orthogonal to F, it follows that at a point in a neighborhood of the minimal path the vector $df^{\#}_{\perp}$ is directed away from the minimal path. Thus the curvature of the fall line at the point, being in the direction $df^{\#}_{\perp}$, is directed towards the minimal path. □

3. DISCUSSION

Fall lines and stationary paths, particularly valleys, are of special interest in the theory of collective motions for the following reason. One of the objectives in collective theory is to identify decoupled manifolds, whenever they exist. Consider a particle that moves in a potential v on a Riemannian configuration space M with kinetic energy proportional to the Laplace–Beltrami operator (1). If the initial velocity of the particle at a point in M is tangent to the fall line at the point, then subsequent motion of the particle will remain on the fall line if the fall line is geodesic. Thus geodesic fall lines, if they exist, are decoupled submanifolds. In general, of course, they do not exist and then the objective becomes to find maximally decoupled submanifolds in order to obtain an approximate collective subdynamics. Since stationary paths are sets of points at which the fall lines are locally geodesic, their identification is an important step in the theory. Two candidates that have been proposed[1,2] for an approximate decoupled collective configuration space, or collective path as it is called, are the valley through the potential minimum and the fall line from the lowest saddle point to the minimum.

From the potential landscape model and Figs. 1 and 2 we might have supposed these two paths to be identical. However, from Theorem 3 and its lemma we learn that this is not so. The only situation in which a valley is also a fall line is when it is geodesic, which is when it is a completely decoupled configuration space.

It is interesting to recall the observation of Maxwell[4] that "In the pure theory of surfaces there is no method of determining a line of watershed or of watercourse, except by first finding a pass or a bar and drawing the line of slope from that point." In contrast, we find that valleys and ridges, as opposed to watercourses and watersheds, are well defined locally. Indeed the above theorems give a simple mechanism for calculating stationary paths. Using Theorem 1, one observes that in the potential landscape model the equation for a stationary path is the line

$$C(RF, F) = 0, \tag{22}$$

where R is a rotation operator that rotates F through an angle $\pi/2$. The use of this equation readily enables one to construct the stationary paths shown in Fig. 2. Furthermore, by inspection of the equipotentials and the use of Theorem 4 one readily infers the character of the stationary path. For an n-dimensional manifold the equation of the stationary path becomes

$$C(R_i F, F) = 0, \quad i = 1,..., n-1, \tag{23}$$

where (R_i) is a set of rotations of F into linearly independent orthogonal vectors.

An alternative method was advanced in the theory of Rowe and Basserman.[7] First let (n_α) be the normal mode vectors at each point and suppose that they are smooth vector fields over some domain of the manifold which includes the potential minimum. The potential gradient can then be expanded,

$$dv^{\#} = k^{\alpha} n_{\alpha}. \tag{24}$$

By Theorem 1 it follows that on a stationary path

$$dv^{\#} = k^{1} n_1, \tag{25}$$

where n_1 is a particular normal mode vector field. For the valley, n_1 is simply the normal mode of lowest frequency at the potential minimum. Equation (25) is a practical equation for the construction of a stationary path. Recall that the critical points are the points at which $dv^{\#} = 0$ and that this equation can be solved iteratively by Hartree–Fock or Newton methods.[8] Thus, in a similar way, a stationary path is constructed by finding the points at which the constrained field $dv^{\#} - kn_1$ vanishes for varying values of k.

A more detailed discussion of the construction of collective paths using these concepts will be presented in a following paper.

ACKNOWLEDGMENT

The authors are pleased to acknowledge helpful discussions with G. Rosensteel.

[1] K. Goeke, P. G. Reinhard, and D. J. Rowe, Nucl. Phys. A **359**, 408 (1976).
[2] D. J. Rowe, "Coordinate independent theory of the collective path" (to be published).
[3] C. A. Cayley, Philos. Mag. **18**, 264 (1859).
[4] J. C. Maxwell, Philos. Mag. **40**, 421 (1870).
[5] D. J. Rowe, "Contribution to the Orsay-Saclay workshop on Time-Dependent Hartree-Fock method," 1979 (unpublished).
[6] N. J. Hicks, *Notes on Differential Geometry* (Van Nostrand, London, 1971).
[7] D. J. Rowe and R. Basserman, Can. J. Phys. **54**, 1941 (1976).
[8] D. J. Rowe, Nucl. Phys. A **107**, 99 (1968).

A mechanical model with constraints [a)]

A. Z. Capri and M. Kobayashi [b)]
Theoretical Physics Institute, Department of Physics, University of Alberta, Edmonton, Alberta T6G 2J1, Canada

(Received 29 September 1981; accepted for publication 23 December 1981)

We analyze a dynamical system with a finite number of degrees of freedom. A complete analysis is presented both for first class constraints as well as for second class constraints. The results are applicable to Yang–Mills fields as well as higher spin fields.

PACS numbers: 03.20. + i

I. INTRODUCTION

Constrained systems have been of interest in field theory for a long time. In the "modern"era, Pauli and Fierz[1] studied such systems by introducing further auxiliary fields into the Lagrangian. These auxiliary fields then served to eliminate unwanted spin components.

Presently, Yang–Mills fields provide a very interesting class of field equations with constraints. The Lagrangian for these fields is usually written as a quadratic form in the kinetic terms. We therefore construct a mechanical model (finite degrees of freedom) which has a Lagrangian with the same kinetic structure as the Lagrange density of Yang–Mills or higher-spin tensor fields.

Takahashi and co-workers[2] also studied constrained mechanical models using Lagrange multiplier fields. Their kinetic terms corresponded to systems of first-order differential equations. The results they obtained for their systems are equivalent to those of Dirac,[3] who studied constrained Hamiltonian systems in general.

Having chosen a mechanical model with appropriate kinetic terms, we apply Dirac's method to it. This serves as a preliminary to studying the quantization of higher-spin as well as Yang–Mills fields; a task we carry out elsewhere for higher-spin fields.[4]

In Sec. II we present the model and analyze the situation corresponding to the occurrence of only first-order constraints. Then in Sec. III we examine the situation corresponding to the occurrence of higher-class constraints and analyze it completely. In Sec. IV we state our conclusions.

II. THE MECHANICAL MODEL

Consider the system with $2N$ degrees of freedom $\phi_a, \phi_b^*, a, b = 1,...,N$, described by the Langrangian

$$L = \dot{\phi}_a^* M_{ab}\dot{\phi}_b + (\dot{\phi}_a^* c_{ab}\phi_b - \phi_a^* c_{ab}\dot{\phi}_b) - \phi_a^* r_{ab}\phi_b - V(\phi_a,\phi_b^*). \tag{1}$$

We assume summation over repeated indices.

The Lagrange density for Yang–Mills fields as well as for a large class of higher-spin fields can be cast in the form (1). The potential term V contains terms cubic as well as quadratic in the coordinates ϕ_a, ϕ_b^*. It plays no role in the constraint structure and will therefore be dropped from now on. The "kinetic mass matrix" M_{ab} as well as the matrix r_{ab} are both Hermitian. The matrix c_{ab} is anti-Hermitian.

Since the matrix M is Hermitian, it can always be diagonalized by a unitary transformation. We assume that this has been done and

$$M_{ab} = m_a\delta_{ab} \tag{2}$$

with *no* summation over the repeated index of the eigenvalue m_a now or later. This eigenvalue is called the kinetic mass.[2]

Dropping the potential term V, the Langrangian can now be written

$$L = m_a\dot{\phi}_a^*\delta_{ab}\dot{\phi}_b + (\dot{\phi}_a^* c_{ab}\phi_b - \phi_a^* c_{ab}\dot{\phi}_b) - \phi_a^* r_{ab}\phi_b. \tag{3}$$

The canonical momenta conjugate to ϕ_a^* and ϕ_b, respectively, are

$$\pi_a = \frac{\partial L}{\partial\dot{\phi}_a^*} = m_a\dot{\phi}_a + c_{ab}\phi_b, \tag{4}$$

$$\pi_b^* = \frac{\partial L}{\partial\dot{\phi}_b} = m_b\dot{\phi}_b^* - \phi_a^* c_{ab}. \tag{5}$$

If the matrix M is nonsingular or, what is the same thing, none of the kinetic masses m_a vanishes, then no constraints occur in the system and the canonical Hamiltonian

$$H_c = \dot{\phi}_a^*\pi_a + \pi_b^*\dot{\phi}_b - L \tag{6}$$

yields the equations of motion

$$\dot{\phi}_a = \{\phi_a,H_c\}, \quad \dot{\phi}_b^* = \{\phi_b^*,H_c\}, \tag{7}$$

$$\ddot{\phi}_a = \{\dot{\phi}_a,H_c\}, \quad \ddot{\phi}_b^* = \{\dot{\phi}_b^*,H_c\}, \tag{8}$$

where the Poisson brackets are defined by[5]

$$\{\pi_a^*,\phi_b\} = -\{\phi_b,\pi_a^*\} = \frac{\partial\pi_a^*}{\partial\phi_c}\frac{\partial\phi_b}{\partial\pi_c^*} - \frac{\partial\phi_b}{\partial\phi_c}\frac{\partial\pi_a^*}{\partial\pi_c^*} = -\delta_{ab}, \tag{9}$$

$$\{\pi_a,\phi_b^*\} = -\{\phi_b^*,\pi_a\} = \frac{\partial\pi_a}{\partial\phi_c^*}\frac{\partial\phi_b^*}{\partial\pi_c} - \frac{\partial\phi_b^*}{\partial\phi_c^*}\frac{\partial\pi_a}{\partial\pi_c} = -\delta_{ab}, \tag{10}$$

and all other Poisson brackets involving only $\phi_a, \phi_b^*, \dot{\phi}_a^*, \dot{\phi}_b$ vanish. Thus such a system is completely determined.

We now assume, therefore, that the matrix M is singular of rank $r < N$. By a suitable choice of the unitary transformation that diagonalized M we can arrange to have

[a)] Research supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

[b)] Permanent address: Department of Physics, Gifu University, Nagara, Gifu 502, Japan.

$$m_a = 0 \quad \text{for } a = i, j,k,... = 1,2,...,N-r,$$
$$m_a \neq 0 \quad \text{for } a = \alpha,\beta,\gamma,... = N-r+1, N-r+2,...,N. \tag{11}$$

We adhere to this notation of letting lower case Latin indices from the middle of the alphabet, namely i, $j,k,...$, run from 1 to $N-r$ and lower case Greek indices from the begining of the alphabet, namely $\alpha,\beta,\gamma,...$, run from $N-r+1$ to N.

The Euler–Lagrange equations corresponding to the Lagrangian (3) now split into equations of motion

$$m_\beta \ddot{\phi}_\beta + 2c_{\beta b}\dot{\phi}_b + r_{\beta b}\phi_b = 0,$$
$$m_\alpha \ddot{\phi}^*_\alpha + 2\dot{\phi}^*_a c_{a\alpha} + \phi^*_a r_{a\alpha} = 0, \tag{12}$$

and equations of constraint

$$2c_{ib}\dot{\phi}_b + r_{ib}\phi_b = 0,$$
$$2\dot{\phi}^*_b c_{bi} + \phi^*_b r_{bi} = 0. \tag{13}$$

In terms of the canonical momenta, these constraints may also be written

$$\chi_i \equiv \pi_i - c_{ib}\phi_b = 0,$$
$$\chi^*_j \equiv \pi^*_j - \phi^*_a c_{aj} = 0. \tag{14}$$

Here we have followed Dirac's terminology and introduced the "constraints" χ_i,χ^*_j.

For this particular case, the canonical Hamiltonian reads

$$H_c = (\pi^*_\alpha + \phi^*_a c_{a\alpha})\frac{1}{m_\alpha}\delta_{\alpha\beta}(\pi_\beta - c_{\beta b}\phi_b) + \phi^*_a r_{ab}\phi_b. \tag{15}$$

The total Hamiltonian [yielding the constraints (14)] is given by

$$H = H_c + u^*_i\chi_i + \chi^*_j u_j, \tag{16}$$

where we have introduced the Lagrange multiplier fields u^*_i, u_j.

The constraints [Eq. (14)] are not merely initial conditions but must hold for all times; thus they must be independent of time. This requires that both $\dot{\chi}_i$ and $\dot{\chi}^*_j$ vanish identically. Writing out these consistency conditions, we obtain

$$\begin{aligned}\dot{\chi}_i &= \frac{\partial \chi_i}{\partial t} + \{\chi_i, H\} = 0 \\ &= -c_{i\beta}\pi_\beta\frac{2}{m_\beta} - (\dot{c}_{ib} - c_{i\beta}\frac{2}{m_\beta}c_{\beta b} + r_{ib})\phi_b - 2c_{ij}u_j \\ &\equiv -2d_{i\beta}\pi_\beta - 2e_{ib}\phi_b - 2c_{ij}u_j\end{aligned} \tag{17}$$

and

$$\begin{aligned}\dot{\chi}^*_j &= \frac{\partial \chi^*_j}{\partial t} + \{\chi^*_j, H\} = 0 \\ &= \frac{2}{m_\alpha}\pi^*_\alpha c_{\alpha j} + \phi^*_a(\dot{c}_{aj} + c_{a\alpha}\frac{2}{m_\alpha}c_{\alpha j} - r_{aj}) + 2u^*_i c_{ij} \\ &\equiv 2\pi^*_a d_{aj} - 2\phi^*_a e_{aj} + 2u^*_i c_{ij},\end{aligned} \tag{18}$$

where we have defined the new matrices $d_{i\beta}, d_{\alpha j}$ and e_{ib}, e_{aj}.

The retention of the terms involving $\dot{c}$ corresponds to time-dependent external forces being present.

If the $N-r$ by $N-r$ submatrix c_{ij} is nonsingular, we can solve Eq. (17) and (18) for the Lagrange multiplier fields u_j, u^*_i and substitute the results back in the total Hamiltonian (16). The Hamiltonian so obtained involves the constrained variables. Thus, to obtain the true equations of motion, it is necessary that the Poisson brackets themselves take the constraints into account. Such generalized Poisson brackets were defined by Dirac.[3]

The only constraints involved here are the "first-class" constraints given by Eqs. (14). We now illustrate the handling of these constraints. As a first step we compute the "constraint matrix"

$$K = \begin{bmatrix} \{\chi_i,\chi_j\} & \{\chi_i,\chi^*_j\} \\ \{\chi^*_i,\chi_j\} & \{\chi^*_i,\chi^*_j\}\end{bmatrix} \tag{19}$$

$$= \begin{bmatrix} 0 & -2c_{ij} \\ 2\tilde{c}_{ij} & 0 \end{bmatrix}, \tag{20}$$

where the tilde means transpose. The inverse of K is easily found to be

$$K^{-1} = \begin{bmatrix} 0 & \frac{1}{2}\tilde{c}_{ij}^{-1} \\ -\frac{1}{2}c_{ij}^{-1} & 0 \end{bmatrix}. \tag{21}$$

The generalized Poisson bracket or Dirac bracket[3] for any two dynamical quantities F and G is defined by

$$\{F,G\}^* = \{F,G\} - \sum_\chi \{F,\chi\}K^{-1}\{\chi,G\}. \tag{22}$$

The ordinary Poisson bracket on the right is defined by Casalbuoni.[5] Writing this out, we get

$$\{F,G\}^* = \{F,G\} + \tfrac{1}{2}\{F,\chi^*_i\}c_{ij}^{-1}\{\chi_j,G\} - \tfrac{1}{2}\{F,\chi_i\}\tilde{c}_{ij}^{-1}\{\chi^*_j,G\}. \tag{23}$$

This allows us to compute the generalized Poisson brackets for any coordinates, velocities and momenta. In particular, we get

$$\{\dot{\phi}^*_\alpha,\phi_\beta\}^* = -\frac{1}{m_\alpha}\delta_{\alpha\beta}, \tag{24}$$

$$\{\phi^*_\alpha,\dot{\phi}_\beta\}^* = \frac{1}{m_\alpha}\delta_{\alpha\beta}. \tag{25}$$

Now using the fact that

$$\dot{\phi}_\alpha = \frac{1}{m_\alpha}(\pi_\alpha - c_{\alpha b}\phi_b), \tag{26}$$

we get the true equations of motion from

$$\begin{aligned}\ddot{\phi}_\alpha &= \{\dot{\phi}_\alpha, H\} = \{\dot{\phi}_\alpha, H\}^* - \frac{1}{2}\{\dot{\phi}_\alpha,\chi^*_i\}c_{ij}^{-1}\{\chi_j,H\} \\ &= \{\dot{\phi}_\alpha, H\}^* + \frac{1}{2}\{\dot{\phi}_\alpha,\chi^*_i\}c_{ij}^{-1}\frac{\partial \chi_j}{\partial t}.\end{aligned} \tag{27}$$

Here we used the consistency condition

$$\{\chi_j, H\} = -\frac{\partial \chi_j}{\partial t}. \tag{28}$$

Evaluating the various Dirac brackets we finally get the true equations of motion, namely

$$m_\alpha\delta_{\alpha\beta}\ddot{\phi}_\beta + (2c_{\alpha\beta} - c_{\alpha k}\tilde{c}_{kl}^{-1}c_{l\beta})\dot{\phi}_\beta + (r_{ab} - c_{aj}\dot{c}_{jb})\phi_b = 0. \tag{29}$$

This determines completely the system for the case of first class constraints only. Suppose, however, that c_{ij} is singular. In this case, further "secondary" constraints appear. We examine this situation in the next section.

III. HIGHER CLASS CONSTRAINTS

In this section we assume that the $N-r$ by $N-r$ submatrix c_{ij} is singular of rank r_1 so that further or second-class constraints occur. Since c_{ij} is anti-Hermitian, it can also be diagonalized by a unitary transformation S. This diagonalization does not change the diagonalization of the kinetic mass matrix M_{ab} since if either of the indices a, b, equals i, j (i.e., lies in the range $1,...,N-r$), the matrix elements $M_{a,b}$ vanish.

Thus we can write

$$S_{ik}^{-1}c_{kl}S_{lj}=c_i\delta_{ij}, \tag{30}$$

where no sum occurs over the index of the eigenvalues c_i. The eigenvalues are arranged so that

$$c_i=0 \quad \text{for } i=I,J,K,...=1,2,...,N-r-r_1, \tag{31}$$

$$c_i\neq 0 \quad \text{for } i=A,B,\Gamma,...=N-r-r_1+1,...,N-r. \tag{32}$$

This notation will again be adhered to with upper case Latin letters $I,J,K,...$, having the range 1 to $N-r-r_1$, and upper case Greek letters $A,B,\Gamma,...$, having the range $N-r-r_1+1$ to $N-r$.

Now operating on Eq. (17) with S^{-1} from the left we can cast it in the form

$$c_iS_{ik}^{-1}u_k=-S_{ik}^{-1}d_{k\beta}\pi_\beta-S_{ik}^{-1}e_{kb}\phi_b. \tag{33}$$

Now using (31) and (32) this becomes

$$u_B^{(1)}=-d_{B\beta}^{(1)}\pi_\beta-e_{Bb}^{(1)}\phi_b, \tag{34}$$

$$0=\chi_I^{(1)}\equiv-d_{I\beta}^{(1)}\pi_\beta-e_{Ib}^{(1)}\phi_b, \tag{35}$$

where

$$u_B^{(1)}=S_{Bi}^{-1}u_i,\quad u_I^{(1)}=S_{Ij}^{-1}u_j, \tag{36}$$

$$d_{B\beta}^{(1)}=\frac{1}{c_B}S_{Bi}^{-1}d_{i\beta},\quad d_{I\beta}^{(1)}=S_{Ii}^{-1}d_{i\beta}, \tag{37}$$

$$e_{Bb}^{(1)}=\frac{1}{c_B}S_{Bi}^{-1}e_{ib},\quad e_{Ib}^{(1)}=S_{Ii}^{-1}e_{ib}. \tag{38}$$

Similarly, by operating with S from the right on Eq. (18), we obtain

$$u_A^{(1)*}=-\pi_\alpha^*d_{\alpha A}^{(1)}+\phi_a^*e_{aA}^{(1)}, \tag{39}$$

$$0=\chi_J^{(1)*}\equiv-\pi_\alpha^*d_{\alpha J}^{(1)}+\phi_a^*e_{aJ}^{(1)}, \tag{40}$$

where

$$u_A^{(1)*}=u_i^*S_{iA},\quad u_I^{(1)*}=u_i^*S_{iI}, \tag{41}$$

$$d_{\alpha A}^{(1)}=d_{\alpha j}S_{jA}\frac{1}{c_A},\quad d_{\alpha J}^{(1)}=d_{\alpha j}S_{jJ}, \tag{42}$$

$$e_{aA}^{(1)}=e_{aj}S_{jA}\frac{1}{c_A},\quad e_{aJ}^{(1)}=e_{aj}S_{jJ}. \tag{43}$$

Also, as is easy to verify, we have

$$u_i^*\chi_i=u_A^{(1)}\chi_A^{(1)}+u_I^{(1)}\chi_I^{(1)} \tag{44}$$

and

$$\chi_j^*u_j=\chi_B^{(1)*}u_B^{(1)}+\chi_J^{(1)*}u_J^{(1)}, \tag{45}$$

where we have further defined

$$\chi_A^{(1)}=S_{Ai}^{-1}\chi_i=S_{Ai}^{-1}\pi_i-S_{Ai}^{-1}c_{ib}\phi_b, \tag{46}$$

$$\chi_B^{(1)*}=\chi_j^*S_{jB}=\pi_j^*S_{jB}+\phi_a^*c_{aj}S_{jB}. \tag{47}$$

The total Hamiltonian (14) is then given by

$$H=H_c+(-\pi_\alpha^*d_{\alpha A}^{(1)}+\phi_a^*e_{aA}^{(1)})\chi_A^{(1)}+\chi_B^{(1)*}(-d_{B\beta}^{(1)}\pi_\beta-e_{Bb}^{(1)}\phi_b)+u_I^{(1)*}\chi_I^{(1)}+\chi_J^{(1)*}u_J^{(1)}, \tag{48}$$

where we have eliminated the Lagrange multipliers $u_B^{(1)}$ and $u_A^{(1)*}$ by using Eqs. (34) and (39). Thus only $u_I^{(1)*}$ and $u_J^{(1)}$ remain undetermined. The corresponding constraints are $\chi_I^{(1)}$ and $\chi_J^{(1)*}$. Again consistency requires that these constraints be time-independent. Thus we get

$$\dot\chi_I^{(1)}=0=\dot\chi_J^{(1)*}. \tag{49}$$

Writing these out we obtain the equations

$$f_{IJ}^{(1)}u_J^{(1)}=-\dot S_{Ii}^{-1}\chi_i-S_{Ii}^{-1}p_{ij}S_{jA}\frac{1}{c_A}\chi_A^{(1)}+S_{Ii}^{-1}g_{i\beta}\pi_\beta+S_{Ii}^{-1}h_{ib}\phi_b \tag{50}$$

and

$$u_I^{(1)*}f_{IJ}^{(1)}=\chi_j^*\dot S_{jJ}-\chi_B^{(1)*}\frac{1}{c_B}S_{Bi}^{-1}p_{ij}S_{jJ}+\pi_\alpha^*g_{\alpha j}S_{jJ}+\phi_a^*h_{aj}S_{jJ}, \tag{51}$$

where we have introduced the notation

$$f_{ij}=d_{i\alpha}c_{\alpha j}-e_{ij}, \tag{52}$$

$$p_{ij}=e_{i\alpha}d_{\alpha j}+d_{i\beta}e_{\beta j}=-p_{ji}^*, \tag{53}$$

$$g_{i\beta}=\dot d_{i\beta}-(d_{i\gamma}c_{\gamma\beta}-e_{i\beta})\frac{1}{m_\beta}+(d_{i\gamma}c_{\gamma k}-e_{ik})\times S_{kB}\frac{1}{c_B}S_{Bl}^{-1}d_{l\beta}, \tag{54}$$

$$h_{ib}=\dot e_{ib}+(d_{i\gamma}c_{\gamma\beta}-e_{i\beta})d_{\beta b}-d_{i\beta}r_{\beta b}+(d_{i\gamma}c_{\gamma k}-e_{ik})\times S_{kB}\frac{1}{c_B}S_{Bl}^{-1}e_{lb}, \tag{55}$$

and all quantities with a superscript (1) are similarity transformed with S. Thus, for example,

$$f_{IJ}^{(1)}=S_{Ii}^{-1}f_{ij}S_{jJ},\quad f_{IA}^{(1)}=S_{Ii}^{-1}f_{ij}S_{jA},\ \text{etc.},\dots . \tag{56}$$

Now if $f_{IJ}^{(1)}$ is nonsingular we can solve Eq. (50) and (51) for the Lagrange multipliers $u_J^{(1)}$, $u_I^{(1)*}$ and replace the results in the Hamiltonian (48). This is the case of second-class constraints. If $f_{IJ}^{(1)}$ is singular such a solution is not possible and we must again diagonalize $f_{IJ}^{(1)}$ to obtain further higher-order constraints $\chi^{(2)}$, $\chi^{(2)*}$ with corresponding Lagrange multipliers. The situation is now the same as the transition from first-class to second-class constraints and thus may be treated in the same manner. Since we started with a finite matrix, this procedure must eventually exhaust all constraints.

This shows that higher-class constraints are treated in a manner similar to second-class constraints and involve no new concepts. For this reason we assume that $f_{IJ}^{(1)}$ is nonsingular. This allows us to solve (50) and (51) for the Lagrange multipliers $u_J^{(1)}$ and $u_I^{(1)*}$ and replace them in the total Hamiltonian. The Hamiltonian so obtained still involves constrained variables. This means that the evolution equations for all the dynamical variables can again only be obtained if the Poisson brackets themselves take the constraints into account. These are again the generalized Poisson brackets defined by Dirac in terms of the usual Poisson brackets and the constraints.

The constraints for this case are simply

$$\chi_A^{(1)} = 0, \tag{57}$$

$$\chi_B^{(1)*} = 0, \tag{58}$$

$$\chi_I^{(1)} = 0, \tag{59}$$

$$\chi_J^{(1)*} = 0. \tag{60}$$

The Poisson brackets between any two of these is easily computed. Thus, for example, using (9), (10) and (35), (40) we get

$$\begin{aligned}\{\chi_I^{(1)},\chi_J^{(1)*}\} &= -S_{Ii}^{-1}\{d_{i\beta}\pi_\beta + e_{ib}\phi_b, -\pi_\alpha^* d_{\alpha j} + \phi_a^* e_{aj}\}S_{jJ} \\ &= S_{Ii}^{-1}(d_{i\beta}e_{\beta j} + e_{i\alpha}d_{\alpha j})S_{jJ} = p_{IJ},\end{aligned}$$

with similar results for the other brackets. We can, therefore, obtain the complete "constraint matrix":

$$K = \begin{bmatrix} \{\chi_A^{(1)},\chi_B^{(1)}\} & \{\chi_A^{(1)},\chi_B^{(1)*}\} & \{\chi_A^{(1)},\chi_J^{(1)}\} & \{\chi_A^{(1)},\chi_J^{(1)*}\} \\ \{\chi_A^{(1)*},\chi_B^{(1)}\} & \{\chi_A^{(1)*},\chi_B^{(1)*}\} & \{\chi_A^{(1)*},\chi_J^{(1)}\} & \{\chi_A^{(1)*},\chi_J^{(1)*}\} \\ \{\chi_I^{(1)},\chi_B^{(1)}\} & \{\chi_I^{(1)},\chi_B^{(1)*}\} & \{\chi_I^{(1)},\chi_J^{(1)}\} & \{\chi_I^{(1)},\chi_J^{(1)*}\} \\ \{\chi_I^{(1)*},\chi_B^{(1)}\} & \{\chi_I^{(1)*},\chi_B^{(1)*}\} & \{\chi_I^{(1)*},\chi_J^{(1)}\} & \{\chi_I^{(1)*},\chi_J^{(1)*}\} \end{bmatrix} \tag{61}$$

$$= \begin{bmatrix} 0 & -2c_A\delta_{AB} & 0 & f_{AJ}^{(1)} \\ 2c_A\delta_{AB} & 0 & -\tilde{f}_{AJ}^{(1)} & 0 \\ 0 & f_{IB}^{(1)} & 0 & p_{IJ}^{(1)} \\ -\tilde{f}_{IB}^{(1)} & 0 & -p_{IJ}^{(1)} & 0 \end{bmatrix}. \tag{62}$$

The inverse of this matrix exists and is given by

$$K^{-1} = \begin{bmatrix} 0 & \frac{1}{2c_A}\left(\delta_{AB} - \frac{1}{2c_A}\tilde{f}_{AI}^{(1)}\tilde{\Delta}_{IJ}^{-1}\tilde{f}_{JB}^{(1)}\right) & 0 & -\frac{1}{2c_A}\tilde{f}_{AK}^{(1)}\tilde{\Delta}_{KJ}^{-1} \\ \frac{-1}{2c_A}\left(\delta_{AB} - \frac{1}{2c_A}f_{AI}^{(1)}\Delta_{IJ}^{-1}f_{JB}^{(1)}\right) & 0 & \frac{1}{2c_A}f_{AK}^{(1)}\Delta_{KJ}^{-1} & 0 \\ 0 & -\tilde{\Delta}_{IJ}^{-1}\tilde{f}_{JB}^{(1)}\frac{1}{2c_B} & 0 & -\tilde{\Delta}_{IJ}^{-1} \\ \Delta_{IJ}^{-1}f_{JB}^{(1)}\frac{1}{2c_B} & 0 & \Delta_{IJ}^{-1} & 0 \end{bmatrix}. \tag{63}$$

The notation used here is

$$\Delta_{IJ} = p_{IJ}^{(1)} + f_{IB}^{(1)}\frac{1}{2c_B}f_{BJ}^{(1)}, \tag{64}$$

$$\tilde{\Delta}_{IJ} = \tilde{p}_{IJ} + \tilde{f}_{IB}^{(1)}\frac{1}{2c_B}\tilde{f}_{BJ}^{(1)}. \tag{65}$$

Using the inverse given by Eq. (63) we obtain the generalized Poisson or Dirac bracket.

$$\begin{aligned}\{F,G\}^* = \{F,G\} &+ \{F,\chi_A^{(1)*}\}\frac{1}{2c_A}\left(\delta_{AB} - \frac{1}{2c_A}f_{AI}^{(1)}\Delta_{IJ}^{-1}f_{JB}^{(1)}\right)\{\chi_B^{(1)},G\} \\ &- \{F,\chi_I^{(1)*}\}\Delta_{IJ}^{-1}f_{JB}^{(1)}\frac{1}{2c_B}\{\chi_B^{(1)},G\} - \{F,\chi_A^{(1)*}\}\frac{1}{2c_A}f_{AI}^{(1)}\Delta_{IJ}^{-1}\{\chi_J^{(1)},G\} \\ &- \{F,\chi_I^{(1)*}\}\Delta_{IJ}^{-1}\{\chi_J^{(1)},G\} - \{F,\chi_A^{(1)}\}\frac{1}{2c_A}\left(\delta_{AB} - \frac{1}{2c_A}\tilde{f}_{AI}^{(1)}\Delta_{IJ}^{-1}\tilde{f}_{JB}^{(1)}\right)\{\chi_B^{(1)*},G\} \\ &+ \{F,\chi_I^{(1)}\}\tilde{\Delta}_{IJ}^{-1}\tilde{f}_{JB}^{(1)}\{\chi_B^{(1)*},G\} + \{F,\chi_A^{(1)}\}\frac{1}{2c_A}\tilde{f}_{AI}^{(1)}\tilde{\Delta}_{IJ}^{-1}\{\chi_J^{(1)*},G\} + \{F,\chi_I^{(1)}\}\tilde{\Delta}_{IJ}^{-1}\{\chi_J^{(1)},G\}.\end{aligned} \tag{66}$$

We have used this equation to obtain the generalized Poisson brackets for all the coordinates ϕ, ϕ^* and the momenta π^*, π. The results are tabulated in an appendix.

Of immediate utility to us are the following generalized Poisson brackets:

$$\{\phi_\alpha,\dot{\phi}_\beta^*\}^* = \delta_{\alpha\beta}\frac{1}{m_\beta} + 2d_{\alpha i}S_{iI}\Delta_{IJ}^{-1}f_{JB}^{(1)}\frac{1}{2c_B}d_{B\beta}^{(1)} - d_{\alpha i}S_{iI}\Delta_{IJ}^{-1}f_{J\beta}^{(1)}\frac{1}{m_\beta}, \tag{67}$$

$$\{\phi_\alpha^*,\dot{\phi}_\beta\}^* = \delta_{\alpha\beta}\frac{1}{m_\beta} - 2\tilde{d}_{\alpha i}\tilde{S}_{iI}^{-1}\tilde{\Delta}_{IJ}^{-1}\tilde{f}_{JB}^{(1)}\frac{1}{2c_B}\tilde{d}_{B\beta}^{(1)} - \tilde{d}_{\alpha i}\tilde{S}_{iI}^{-1}\tilde{\Delta}_{IJ}^{-1}\tilde{f}_{J\beta}^{(1)}\frac{1}{m_\beta}. \tag{68}$$

Using these results we can now obtain the *true* equations of motion from the total Hamiltonian. They are given by

$$\dot{\phi}_\alpha = \{\phi_\alpha, H\}, \tag{69}$$

$$\ddot{\phi}_\alpha = \{\dot{\phi}_\alpha, H\}, \tag{70}$$

where again the ordinary Poisson bracket on the right is obtained by using Eq. (56) and the consistency condition

$$\{\chi_J, H\} = -\frac{\partial \chi_J}{\partial t}. \tag{71}$$

After a rather tedious calculation involving the repeated evaluation of generalized Poisson brackets we arrive at the true equations of motion:

$$\begin{aligned} m_\alpha \delta_{\alpha\beta} \ddot{\phi}_\beta + 2\Big[& c_{\alpha\beta} + f_{\alpha j} S_{jJ} \Delta^{-1}_{IJ} f^{(1)}_{JB} \frac{1}{2c_B} S_{Bj} c_{j\beta} - 2c_{\alpha i} S_{iA} \frac{1}{2c_A} \Big(\delta_{AB} - \frac{1}{2c_A} f^{(1)}_{AI} \Delta^{-1}_{IJ} f^{(1)}_{JB}\Big) S^{-1}_{Bj} c_{j\beta} \\ & - c_{\alpha i} S_{iA} \frac{1}{2c_A} f^{(1)}_{AI} \Delta^{-1}_{IJ} S^{-1}_{Jj} (f_{j\beta} - c_{j\beta}) - \frac{1}{2} f_{\alpha i} S_{iI} \Delta^{-1}_{IJ} S^{-1}_{Jj} (f_{j\beta} - c_{j\beta}) \Big] \dot{\phi}_\beta \\ & + \Big[r_{ab} + f_{\alpha i} S_{iI} \Delta^{-1}_{IJ} f^{(1)}_{JB} \frac{1}{2c_B} S^{-1}_{Bj} (r_{jb} + \dot{c}_{jb}) - 2c_{\alpha i} S_{iA} \frac{1}{2c_A} \Big(\delta_{AB} - \frac{1}{2c_A} f^{(1)}_{AI} \Delta^{-1}_{IJ} f^{(1)}_{JB}\Big) S^{-1}_{Bj} (r_{jb} + \dot{c}_{jb}) \\ & - c_{\alpha i} S_{iA} \frac{1}{2c_A} f^{(1)}_{AI} \Delta^{-1}_{IJ} S^{-1}_{Jj} (d_{j\beta} r_{\beta b} - 2\dot{d}_{j\beta} c_{\beta b} - 2\dot{e}_{jb}) - \frac{1}{2} f_{\alpha i} S_{iI} \Delta^{-1}_{IJ} S^{-1}_{Jj} (d_{j\beta} r_{\beta b} - 2\dot{d}_{j\beta} c_{\beta b} - 2\dot{e}_{ib}) \Big] \phi_b = 0. \end{aligned} \tag{72}$$

CONCLUSION

We have completely analyzed a Lagrangian system both in the case of only first-order constraints as well as in the case of second-order constraints. Our analysis also makes it clear how to proceed for higher-order constraints. The form of the Lagrangian is such that our results are immediately applicable to a variety of constrained systems. Thus, both the Yang–Mills Lagrange density as well as many Lagrange densities for higher-spin fields can be cast into this form. The results obtained remain valid in the presence of external forces.

One final conjecture. The way the kinetic mass m_α appears as a multiplicative factor in the various Poisson brackets suggests that the indefiniteness associated with the quantization of higher-spin fields may be related to the relative signs of the kinetic masses. We propose to investigate this for some specific field theories.

APPENDIX

We list here the various generalized Poisson brackets for the case of second-class constraints.

$$\{\phi_i, \phi^*_j\}^* = -S_{iA} \frac{1}{2c_A} \Big(\delta_{AB} - \frac{1}{2c_A} f^{-1}_{AI} \Delta^{-1}_{IJ} f^{(1)}_{JB}\Big) S^{-1}_{Bj}, \tag{A1}$$

$$\{\phi_i, \phi^*_\beta\}^* = -S_{iA} \frac{1}{2c_A} f^{(1)}_{AI} \Delta^{-1}_{IJ} S^{-1}_{Jl} d_{l\beta}, \tag{A2}$$

$$\{\phi_i, \pi^*_j\}^* = \delta_{ij} - S_{iA} \frac{1}{2c_A} \Big(\delta_{AB} - \frac{1}{2c_A} f^{(1)}_{AI} \Delta^{-1}_{IJ} f^{(1)}_{JB}\Big) S^{-1}_{Bl} c_{lj} + S_{iA} \frac{1}{2c_A} f^{(1)}_{AI} \Delta^{-1}_{IJ} S^{-1}_{Jl} e_{lj}, \tag{A3}$$

$$\{\phi_i, \pi^*_\beta\}^* = -S_{iA} \frac{1}{2c_A} \Big(\delta_{AB} - \frac{1}{2c_A} f^{(1)}_{AI} \Delta^{-1}_{IJ} f^{(1)}_{JB}\Big) S^{-1}_{Bl} c_{l\beta} + S_{iA} \frac{1}{2c_A} f^{(1)}_{AI} \Delta^{-1}_{IJ} S^{-1}_{Jl} e_{l\beta}, \tag{A4}$$

$$\{\phi_\alpha, \phi^*_j\}^* = d_{\alpha k} S_{kI} \Delta^{-1}_{IJ} f^{(1)}_{JB} \frac{1}{2c_B} S^{-1}_{Bj}, \tag{A5}$$

$$\{\phi_\alpha, \phi^*_\beta\}^* = -d_{\alpha k} S_{kI} \Delta^{-1}_{IJ} S_{Jl} d_{l\beta}, \tag{A6}$$

$$\{\phi_\alpha, \pi^*_j\}^* = d_{\alpha k} S_{kI} \Delta^{-1}_{IJ} f^{(1)}_{JB} \frac{1}{2c_B} S^{-1}_{Bk} c_{kj} + d_{\alpha k} S_{kI} \Delta^{-1}_{IJ} S^{-1}_{Jk} e_{kj}, \tag{A7}$$

$$\{\phi_\alpha, \pi^*_\beta\}^* = \delta_{\alpha\beta} + d_{\alpha k} S_{kI} \Delta^{-1}_{IJ} f^{(1)}_{JB} \frac{1}{2c_B} S^{-1}_{Bl} c_{l\beta} + d_{\alpha k} S_{kI} \Delta^{-1}_{IJ} S^{-1}_{Jl} e_{l\beta}, \tag{A8}$$

$$\{\pi_i,\phi_j^*\}^* = -\delta_{ij} + c_{ik}S_{kA}\frac{1}{2c_A}(\delta_{AB} - \frac{1}{2c_A}f^{(1)}_{AI}\Delta^{-1}_{IJ}f^{(1)}_{JB})S^{-1}_{Bj} + e_{ik}S_{kI}\Delta^{-1}_{IJ}f^{(1)}_{JB}\frac{1}{2c_B}S^{-1}_{Bj}, \tag{A9}$$

$$\{\pi_i,\phi_\beta^*\}^* = c_{ik}S_{kA}\frac{1}{2c_A}f^{(1)}_{AI}\Delta^{-1}_{IJ}S^{-1}_{Jl}d_{l\beta} - e_{ik}S_{kI}\Delta^{-1}_{IJ}S^{-1}_{Jl}d_{l\beta}, \tag{A10}$$

$$\begin{aligned}\{\pi_i,\pi_j^*\}^* = {} & c_{ik}S_{kA}\frac{1}{2c_A}(\delta_{AB} - \frac{1}{2c_A}f^{(1)}_{AI}\Delta^{-1}_{IJ}f^{(1)}_{JB})S^{-1}_{Bl}c_{lj} + e_{ik}S_{kI}\Delta^{-1}_{IJ}f^{(1)}_{JB}\frac{1}{2c_B}S^{-1}_{Bl}c_{lj} \\ & - c_{ik}S_{kA}\frac{1}{2c_A}f^{(1)}_{AI}\Delta^{-1}_{IJ}S^{-1}_{Jl}e_{lj} + e_{ik}S_{kI}\Delta^{-1}_{IJ}S^{-1}_{Jl}e_{lj},\end{aligned} \tag{A11}$$

$$\begin{aligned}\{\pi_i,\pi_\beta^*\}^* = {} & c_{ik}S_{kA}\frac{1}{2c_A}(\delta_{AB} - \frac{1}{2c_A}f^{(1)}_{AI}\Delta^{-1}_{IJ}f^{(1)}_{JB})S^{-1}_{Bl}c_{l\beta} + e_{ik}S_{kI}\Delta^{-1}_{IJ}f^{(1)}_{JB}\frac{1}{2c_B}S^{-1}_{Bl}c_{l\beta} \\ & - c_{ik}S_{kA}\frac{1}{2c_A}f^{(1)}_{AI}\Delta^{-1}_{IJ}S^{-1}_{Jl}e_{l\beta} + e_{ik}S_{kI}\Delta^{-1}_{IJ}S^{-1}_{Jl}e_{l\beta},\end{aligned} \tag{A12}$$

$$\{\pi_\alpha,\phi_j^*\}^* = c_{\alpha k}S_{kA}\frac{1}{2c_A}(\delta_{AB} - \frac{1}{2c_A}f^{(1)}_{AI}\Delta^{-1}_{IJ}f^{(1)}_{JB})S^{-1}_{Bj} + e_{\alpha k}S_{kI}\Delta^{-1}_{IJ}f^{(1)}_{JB}\frac{1}{2c_B}S^{-1}_{Bj}, \tag{A13}$$

$$\{\pi_\alpha,\phi_\beta^*\}^* = -\delta_{\alpha\beta} + c_{\alpha k}S_{kA}\frac{1}{2c_A}f^{(1)}_{AI}\Delta^{-1}_{IJ}S^{-1}_{Jl}d_{l\beta} - e_{\alpha k}S_{kI}\Delta^{-1}_{IJ}S^{-1}_{Jl}d_{l\beta}, \tag{A14}$$

$$\begin{aligned}\{\pi_\alpha,\pi_j^*\}^* = {} & c_{\alpha k}S_{kA}\frac{1}{2c_A}(\delta_{AB} - \frac{1}{2c_A}f^{(1)}_{AI}\Delta^{-1}_{IJ}f^{(1)}_{JB})S^{-1}_{Bl}c_{lj} + e_{\alpha k}S_{kI}\Delta^{-1}_{IJ}f^{(1)}_{JB}\frac{1}{2c_B}S^{-1}_{Bl}c_{lj} \\ & - c_{\alpha k}S_{kA}\frac{1}{2c_A}f^{(1)}_{AI}\Delta^{-1}_{IJ}S^{-1}_{Jl}e_{lj} + e_{\alpha k}S_{kI}\Delta^{-1}_{IJ}S^{-1}_{Jl}e_{lj},\end{aligned} \tag{A15}$$

$$\begin{aligned}\{\pi_\alpha,\pi_\beta^*\}^* = {} & c_{\alpha k}S_{kA}\frac{1}{2c_A}(\delta_{AB} - \frac{1}{2c_A}f^{(1)}_{AI}\Delta^{-1}_{IJ}f^{(1)}_{JB})S^{-1}_{Bl}c_{l\beta} + e_{\alpha k}S_{kI}\Delta^{-1}_{IJ}f^{(1)}_{JB}\frac{1}{2c_B}S^{-1}_{Bl}c_{l\beta} \\ & - c_{\alpha k}S_{kA}\frac{1}{2c_A}f^{(1)}_{AI}\Delta^{-1}_{IJ}S^{-1}_{Jl}e_{l\beta} + e_{\alpha k}S_{kI}\Delta^{-1}_{IJ}S^{-1}_{Jl}e_{l\beta}.\end{aligned} \tag{A16}$$

[1]M. Fierz and W. Pauli, Proc. R. Soc. London Ser. A **155**, 447 (1939).
[2]Y. Takahshi, Physica (Utrecht) **31**, 205 (1965); A. Aurilia, M. Kobayashi, and Y. Takahashi, Phys. Rev. D **22**, 1368 (1980).
[3]P. A. M. Dirac, Can. J. Math. **2**, 129 (1950).
[4]A. Z. Capri, M. Kobayashi, and Y. Ohnuki, "The Second Rank Tensor Field, A Systematic Study," University of Alberta Preprint (Alberta-Thy-28-81) (unpublished).
[5]R. Casalbuoni, Nuovo Cimento A **33**, 115 (1976).

Hamiltonian perturbation theory in noncanonical coordinates

Robert G. Littlejohn [a)]
Center for Nonlinear Dynamics, La Jolla Institute, La Jolla, California 92038

(Received 22 April 1981; accepted for publication 25 November 1981)

The traditional methods of Hamiltonian perturbation theory in classical mechanics are first presented in a way which clearly displays their differential-geometric foundations. These are then generalized to the case of noncanonical in phase space. In the new method the Hamiltonian H is treated, not as a scalar in phase space, but as one component of the fundamental form $p\,dq - H dt$. The perturbation analysis is applied to this entire form, in all of its components.

PACS numbers: 03.20. + i, 02.40. + m

1. INTRODUCTION

It is becoming widely appreciated that a Hamiltonian system is better characterized by its abstract properties rather than by any necessary relation among q's and p's. For example, preservation of phase volume, regardless of the coordinates used to express it, has certain definite consequences, such as the impossibility of bounded attracting sets. The situation is somewhat analogous to the Pauli matrices in quantum mechanics: It is reassuring to have the standard representations, but all that is ever really called for is their formal, algebraic properties. Of course, Darboux's theorem always guarantees the existence of q's and p's, at least locally, so that it is never wrong to derive relations by using these coordinates.

Nevertheless, there are a number of practical reasons why noncanonical coordinates arise in applications. First, Hamiltonian systems are sometimes discovered in a context that is independent of Lagrangians or the Legendre transformation.[1] Canonical coordinates may or may not be offered up in such circumstances. Second, the use of canonical coordinates may demand the choice of a gauge with no physical significance; this is seen in ordinary electrodynamics,[2] and it is a pervasive feature of Hamiltonian formulations of fluid mechanics which employ the Clebsch potentials.[3] Third, the use of canonical coordinates often involves the representation of physically interesting quantities by means of awkward mathematical constructions. For example, the Lagrangian coordinates in fluid mechanics indicate where a fluid element was at some initial time. They are related to the actual, Eulerian position of a fluid element by the awkward and impractical device of integrating along streamlines; and yet they provide simple coordinates for fluid flow.[3] Similarly, the α and β coordinates[4] commonly used to represent magnetic fields are canonical coordinates of a kind (on the symplectic manifold consisting of magnetic field lines), and yet in order to know the values of α and β at a point $\mathbf{x}$ of space, one must follow a field line back to some reference surface. Likewise, a direct application of Darboux's theorem often yields canonical coordinates whose relation to physical coordinates is nonlocal in nature. Fourth, in multiple time-scale phenomena, the different p's and q's on the different time scales evolve in time relative to one another in some ratio ϵ; thus the Hamiltonian H must depend on some of the q's and p's through such combinations as ϵq, ϵp. At best, this is awkward; at worst, it can present serious obstacles to a Hamiltonian treatment of the problem, so that investigations are sometimes carried out with non-Hamiltonian methods.[5,6] These problems can often be obviated by using noncanonical coordinates.[2,7]

For all these reasons, noncanonical coordinates for Hamiltonian systems are receiving increasing attention.[8–10] This paper is a contribution to the techniques of performing calculations in noncanonical coordinates and, in particular, perturbation calculations. In some sense, the methods we propose in this paper are similar to the asymptotic methods of Kruskal,[11] but they are explicitly Hamiltonian throughout.

There are several reasons for preferring Hamiltonian perturbation methods, including the noncanonical methods we present here, to non-Hamiltonian methods. One is the essence of practicality: The sheer labor of perturbation calculation is vastly reduced with Hamiltonian methods. Another is the philosphy that if one is given a physical system that is Hamiltonian, then it is best to approximate it with other Hamiltonian systems. In this way conservation of energy, Liouville's theorem, and other consequences of Hamiltonian systems are all valid for each of the approximations.

In this paper we present the outlines of basic perturbation methods for noncanonical coordinates. We pay no attention to more sophisticated techniques such as superconvergent methods,[12] since we presume that these techniques can be adapted to noncanonical variables when the basic methods are understood.

In Secs. 2 and 3 we describe traditional perturbation methods. Section 2 is devoted to non-Hamiltonian systems, while Sec. 3 describes Hamiltonian systems. In these descriptions we call on the theory of differential forms, which is especially appropriate for Hamiltonian systems.[13–15] Our purpose in doing this is to clearly identify the differential-geometric underpinnings of these traditional methods, which are generalized in later sections. We do not intend these descriptions to provide a tutorial on the traditional methods, for which we recommend Nayfeh[16] and Cary.[17]

The language and notation of differential geometry and differential forms has achieved only an incomplete popularity in the physics community today, but it is one which is

[a)] Present address: Department of Physics, University of California, Los Angeles, California 90024.

 0022-2488/82/050742-06$02.50

growing, and one which we wish to endorse. Out of numerous references on this subject, we find Spivak[18] to be informative and entertaining. To make the presentation more readable, however, we have converted all the differential-geometric statements of importance into component language, so that they can be understood in terms of traditional tensor analysis.

In Sec. 4 we present an invariant formulation of Hamiltonian systems which describes how the fundamental form $p\,dq - H\,dt$ is related to the equations of motion, and how the two transform under arbitrary coordinate transformations. In Sec. 5, the new perturbation methods, based on this transformation theory, are described.

2. NON-HAMILTONIAN PERTURBATION METHODS

We consider an n-dimensional manifold M endowed with a vector field X. In some system of coordinates $\{x^i\}$, we assume that X has the representation

$$\frac{dx^i}{dt} = X^i(x) = \sum_{k=0}^{\infty} \epsilon^k X^i_k(x), \tag{1}$$

where ϵ is an ordering parameter. We treat all power series formally; as often in perturbation theory, convergence is of secondary concern to us.

By hypothesis, the leading term X_0 of Eq. (1) represents a solvable system, so that the integral curves of X are approximated by the known integral curves of X_0. These are often represented by straight lines in the coordinate system $\{x^i\}$. The basic strategy of perturbation theory is to seek a coordinate transformation to new coordinates $\{\bar{x}^i\}$, such that in the new coordinates the new equations of motion are simplified. Since the system (1) is solvable at lowest order, the coordinate transformation is the identity at lowest order, i.e., $\bar{x}^i = x^i + O(\epsilon)$.

In this point of view, we have two coordinate systems $\{x^i\}$ and $\{\bar{x}^i\}$, and a given point $x \in M$ is represented alternatively by its coordinates x^i or $\bar{x}^i$ in the two systems. The vector field X in the abstract does not change under the coordinate transformation, only its coordinate/component representation. We will call this the "passive" point of view.

In what follows, however, we adopt the "active" point of view. For this we consider mappings $T: M \to M$, and we write $Tx = \bar{x}$. That is, we will write x for a typical "old" point of M, and $\bar{x}$ for a typical "new" point. When coordinates are called for, we will write x^i and $\bar{x}^i$ for the coordinates of the old and new points, relative to a single coordinate system $\{x^i\}$.

In the active point of view, the vector field X is changed under the mapping into a new vector field $\bar{X}$ by the derivative or tangent map T_*:

$$\bar{X} = T_* X. \tag{2}$$

The coordinate representation of this is

$$\bar{X}^i(\bar{x}) = \frac{\partial \bar{x}^i}{\partial x^j} X^j(x), \tag{3}$$

in which, as throughout, the summation convention is employed, unless greater clarity is called for. The previously mentioned strategy for perturbation theory is stated in the active point of view by saying that we seek T such that $\bar{X}$ is easier to solve than X. The precise criterion depends on circumstances; sometimes one asks for $\bar{X}$ which has the same (straight) integral curves as X_0. In this case, one has "transformed away" the higher order terms of X.

One must be able to transform scalar fields under T as well as vector fields. It is best to treat scalars as 0-forms on M, and use the pullback T^*. Because the pullback T^* works in a direction opposite that of T, we define a transformed scalar $\bar{s}$ in terms of a given scalar $s: M \to \mathbb{R}$ by means of

$$\bar{s} = T^{*-1} s. \tag{4}$$

In terms of values, this can be written

$$(T^* \bar{s})(x) = s(x) = \bar{s}(Tx) = \bar{s}(\bar{x}), \tag{5}$$

where we have used $\bar{x} = Tx$.

Particular scalar fields of interest are the coordinate functions $I^i: U \subset M \to \mathbb{R}$, $i = 1,...,n$, which satisfy $x^i = I^i(x)$ whenever x^i are the coordinates of the point $x \in M$. The set $U \subset M$ is the range of the chart. Applying I^i to the point $\bar{x} = Tx$, we have

$$\bar{x}^i = I^i(\bar{x}) = I^i(Tx) = (T^* I^i)(x). \tag{6}$$

This is often written (imprecisely) as $\bar{x}^i = Tx^i$. The fact that points are pushed forward by T and differential forms (including scalars) are pulled back by T^* is the source of many sign errors and much confusion in practice.

In the so-called Lie transform method,[19] one uses transformations T which are represented as the exponential of some vector field, or rather compositions of such transformations. To begin, let us consider a vector field G, which is associated with the system of ordinary differential equations,

$$\frac{dx^i}{d\epsilon} = G^i(x). \tag{7}$$

We denote the advance map associated with these equations by T^ϵ, so that if x and $\bar{x}$ are initial and final points along an integral curve of (7), separated by elapsed parameter ϵ, then $\bar{x} = T^\epsilon x$. In the usual exponential representation for advance maps, we have $T^\epsilon = \exp(\epsilon G)$. We will call G the *generator* of the transformation T.

The tangent map T^ϵ_* can be represented as an exponential involving the Lie derivative L_G which, acting on vector fields X, is defined by

$$L_G X = \lim_{\epsilon \to 0} \frac{1}{\epsilon} (X - T^\epsilon_* X), \tag{8}$$

in which both X and $T_* X$ must be evaluated at the same point, say $x = T^{-\epsilon}\bar{x}$. The Lie derivative is simply the Lie bracket or commutator of the two vector fields:

$$L_G X = [G, X],$$

$$(L_G X)^i = G^j X^i{}_{,j} - X^j G^i{}_{,j}. \tag{9}$$

Then we have

$$T^\epsilon_* = \exp(-\epsilon L_G). \tag{10}$$

Thus, the relation $\bar{X} = T^\epsilon_* X$, as in Eqs. (2) and (3), can be written as a power series of commutators.

The pullback $T^{\epsilon *}$ can be represented similarly. The action of L_G on an arbitrary differential form α is given by

$$L_G \alpha = \lim_{\epsilon \to 0} \frac{1}{\epsilon} (T^{\epsilon *}\alpha - \alpha), \tag{11}$$

so that

$$T^{\epsilon *} = \exp(+\epsilon L_G). \tag{12}$$

In particular, if α is a scalar s, then

$$L_G s = Gs = G^i \frac{\partial s}{\partial x^i}. \tag{13}$$

In perturbation theory, one applies the tangent map to the vector field X [Eqs. (2) and (10)] in order to make the new vector field $\bar{X}$ easier to solve. This involves solving a simple differential equation for G based on the given desiderata for $\bar{X}$. Then the transformation of the coordinates themselves [Eqs. (5) and (12)] can be written out using the pullback.

However, it is generally only possible to meet the desiderata for the $O(\epsilon)$ term, X_1, of Eq. (1), with a single generator G in the form we have given. Therefore we introduce a sequence of generators[20] $G_1, G_2, \ldots$, Lie derivative operators $L_1 = L_{G_1}$, etc., and transformation operators, $T_1, T_2, \ldots$, defined by

$$T_n = \exp(\epsilon^n G_n). \tag{14}$$

The corresponding tangent maps and pullbacks are

$$T_{n*} = \exp(-\epsilon^n L_n),$$
$$T_n^* = \exp(+\epsilon^n L_n). \tag{15}$$

In effect, each successive T_n takes care of the vector field X of Eq. (1) at each successive order in ϵ; hence the ϵ^n factors in the exponent of Eqs. (14) and (15). Then one has the overall transformation operator T which satisfies

$$T = \cdots T_3 T_2 T_1, \quad T^{-1} = T_1^{-1} T_2^{-1} T_3^{-1} \cdots, \tag{16}$$

$$T_* = \cdots T_{3*} T_{2*} T_{1*} T_*^{-1} = T_{1*}^{-1} T_{2*}^{-1} T_{3*}^{-1} \cdots, \tag{17}$$

$$T^* = T_1^* T_2^* T_3^* \cdots, \quad T^{*-1} = \cdots T_3^{*-1} T_2^{*-1} T_1^{*-1}. \tag{18}$$

Note the reversal of factors in Eqs. (18).

An alternative to the use of a sequence of generators is to allow the one generator G of Eqs. (7) to depend on ϵ, giving a nonautonomous system of differential equations.[21] This method appears to be less efficient at higher orders, however.[17]

3. TRADITIONAL METHODS FOR HAMILTONIAN SYSTEMS

In Hamiltonian systems the manifold M is a $2N$-dimensional phase space, which we denote by Φ. Instead of the coordinates x^i, we use canonical coordinates (q_i, p_i), $i = 1, \ldots, N$. We write X_H for the vector field on Φ which corresponds to X in Eq. (1); X_H is determined from Hamilton's equations of motion and the Hamiltonian H:

$$X_H^i = \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i},$$
$$X_H^{i+N} = \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \quad i = 1, \ldots, N. \tag{19}$$

We assume that H is time independent. Time-dependent Hamiltonian systems are subsumed under the formalism of the next section.

The vector field X_H of Eq. (19) is linear in the 1-form dH. The general relation between a Hamiltonian vector field X_A and the scalar A can be expressed in the component form

$$(X_A)^i = \sum_{j=1}^{2N} J^{ij} \frac{\partial A}{\partial z^j}, \tag{20}$$

where $z^i = (q_1, \ldots, q_N, p_1, \ldots, p_N)$. The matrix J^{ij} is the component matrix of a second rank, contravariant tensor J. The components of J are the Poisson brackets of the coordinates among themselves, i.e., $J^{ij} = \{z^i, z^j\}$. We will also write Eq. (20) in the index-free form $X_A = J(dA)$.

The Hamiltonian H is assumed to have the power series representation

$$H = \sum_{n=0}^{\infty} \epsilon^n H_n, \tag{21}$$

in analogy with Eq. (1), and H_0 is assumed to be solvable. Instead of transforming the vector field X_H directly, one transforms the Hamiltonian H into a new Hamiltonian $\bar{H}$. H and $\bar{H}$ are treated as 0-forms, i.e., scalars, and the transformation equation is

$$\bar{H} = T^{*-1} H, \tag{22}$$

as indicated by Eq. (4).

In Hamiltonian perturbation theory the transformation T is usually required to be a canonical transformation (but see Sec. 5 below). Canonical transformations have the virtue that they preserve the form of Hamilton's equations of motion. Canonical transformations can be represented by mixed-variable generating functions, as in the Poincaré–von Zeipel method,[16] or by means of Lie transforms,[17] as in Sec. 2. In the latter method, the vector fields G_n of Eq. (14) are specified by scalar generating functions g_n, through Hamilton's equations, i.e., the tensor J:

$$G_n = -J(dg_n). \tag{23}$$

The minus sign is conventional.

The Lie derivative operators L_n, when acting on scalars, are now represented by Poisson brackets:

$$L_n s = \{g_n, s\}. \tag{24}$$

In particular, Eq. (22) is a power series involving Poisson brackets. In Sec. 5 we will consider the action of the L_n on differential forms of higher rank.

For Hamiltonian systems it is seldom necessary to apply the Lie derivative operators L_n to vector fields, because the only vector fields of interest are Hamiltonian vector fields. For these, one uses the relation

$$[X_A, X_B] = X_{\{B,A\}}, \tag{25}$$

in which vectors correspond to scalars through Eq. (20). In short, Hamiltonian vector fields can all be dealt with implicitly through their (scalar) Hamiltonian functions.

Typical transformation desiderata for Hamiltonian systems are the elimination of the dependence of H on one or more of the q's. These requirements produce simple differential equations for the generators g_n.

4. AN INVARIANT FORMULATION OF HAMILTONIAN MECHANICS

We consider the space $\Phi \times \mathbb{R} = E$, which we call the "extended phase space;" it is a $(2N + 1)$-dimensional space on which we employ the coordinates (q_i, p_i, t), $i = 1,\dots,N$. In this space we introduce the *fundamental 1-form* γ, given by

$$\gamma = p_i dq^i - H\,dt. \tag{26}$$

The Hamiltonian H is allowed to depend on time.

The exterior derivative of γ is a 2-form ω:

$$\omega = d\gamma = dp_i \wedge dq_i - dH \wedge dt. \tag{27}$$

In the coordinates (q_i, p_i, t), the component matrix ω_{ij} can be partitioned by splitting the coordinates into their $N + N + 1$ parts:

$$\omega_{ij} = \begin{pmatrix} 0 & -\mathbf{I} & -\partial H/\partial \mathbf{q} \\ +\mathbf{I} & 0 & -\partial H/\partial \mathbf{p} \\ +\partial H/\partial \mathbf{q} & +\partial H/\partial \mathbf{p} & 0 \end{pmatrix}. \tag{28}$$

Here $\mathbf{I}$ represents the $N \times N$ identity matrix. The upper $2N \times 2N$ diagonal submatrix is nonsingular, so ω_{ij}, which must have even rank, has a rank which is precisely $2N$. Therefore at each point $z \in E$, precisely one of the $2N + 1$ eigenvectors of ω_{ij} is null, i.e., has eigenvalue zero. Indeed, this real null eigenvector can be taken to be $(\partial H/\partial p_i, -\partial H/\partial q_i, 1)$ in the (q_i, p_i, t) coordinates, as is easily verified.

The null eigenvector of ω_{ij} establishes a one-dimensional subspace of TE_z for each $z \in E$, i.e., a one-dimensional distribution on E. The corresponding integral manifolds are curves in E which are nowhere parallel to the surfaces $t = \text{const}$. These are the "vortex lines" of the fundamental 1-form γ.

The vortex lines specify the motion corresponding to the given Hamiltonian H. This is easily seen by promoting the eigenvector field determined above into a system of ordinary differential equations:

$$\frac{dq_i}{ds} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{ds} = -\frac{\partial H}{\partial q_i}, \quad \frac{dt}{ds} = 1. \tag{29}$$

The parameter s is of no significance; when it is eliminated, we have Hamilton's equations of motion in the usual sense.

The preceding construction is a demonstration, in the coordinates (q_i, p_i, t), of the fact that the time evolution of a dynamical system is invariantly associated with the fundamental 1-form γ. More generally, let $\{z^i\}$, $i = 1,\dots,2N + 1$, represent an arbitrary coordinate system on E. Usually one would take $z^{2N+1} = t$, but this is not essential. From γ, we can derive the 2-form ω. The components ω_{ij} of ω can always be represented, by permuting the order of the coordinates z^i, if necessary, so that the upper $2N \times 2N$ diagonal submatrix is nonsingular. Let us call this submatrix $\hat{\omega}_{ij}$, and let J^{ij} be its inverse:

$$\sum_{k=1}^{2N} J^{ik}\hat{\omega}_{kj} = \delta^i_j. \tag{30}$$

If $z^{2N+1} = t$, then $J^{ij} = \{z^i, z^j\}$, as before. Then the null eigenvector X of ω is represented in components by

$$X^i = \sum_{j=1}^{2N} J^{ij}\left(\frac{\partial h}{\partial z^j} + \frac{\partial \gamma_j}{\partial \tau}\right), \quad i = 1,\dots,2N,$$
$$X^{2N+1} = 1, \tag{31}$$

where $h = -\gamma_{2N+1}$ and $\tau = z^{2N+1}$. Note that if $\tau = t$, then $h = H$. Finally, the equations of motion in the coordinates z^i are

$$\frac{dz^i}{d\tau} = X^i, \quad i = 1,\dots,2N. \tag{32}$$

The relation between γ, ω, the eigenvector field X, and the equations of motion has precisely the behavior under mappings that we expect and require for perturbation theory. To see this, let us characterize a null eigenvector X of ω at a point $z \in E$ as the nonzero vector which satisfies $\omega(Y,X) = 0$ for all $Y \in TE_z$. Let us now apply a diffeomorphism $T: E \to E$, according to the rules $\bar{\omega} = T^{*-1}\omega$, $\bar{X} = T_* X$, $\bar{Y} = T_* Y$, as indicated already by Eqs. (2) and (4). Then we have

$$\bar{\omega}(\bar{X},\bar{Y}) = \bar{\omega}(T_* X, T_* Y) = (T^* \bar{\omega})(X,Y) = \omega(X,Y) = 0, \tag{33}$$

which holds for all $\bar{Y} \in E_{\bar{z}}$, where $\bar{z} = Tz$, since we assume that $T: E \to E$ is a diffeomorphism. Thus, $\bar{X}$ is the null eigenvector of $\bar{\omega}$.

The upshot of this is that if we simply transform points by the mapping T and the 2-form ω by T^{*-1}, then the null eigenvector of the new $\bar{\omega}$ will automatically give the transformed equations of motion. Thus, we can concentrate our attention, as far as perturbation theory is concerned, on the 2-form ω, and not worry about the equations of motion.

Actually, it suffices to transform only the fundamental 1-form γ, since the pullback commutes with the exterior derivative. It is important to note that γ is subject to a "gauge transformation," i.e., we can write $\gamma' = \gamma + dS$, for any scalar S, and γ' serves equally as well as γ for finding ω, the null eigenvector X, and the equations of motion.

5. HAMILTONIAN PERTURBATION THEORY IN NONCANONICAL COORDINATES

We consider for a moment only time-independent systems. It is a simple observation of Hamiltonian mechanics in usual (q,p) language that the Hamiltonian H is determined only to within an additive constant. This is evidently a gauge transformation of a simple kind; although the equations of motion are determined only by the 1-form dH, it is more convenient in practice to work with the 0-form H. Indeed, the Hamiltonian itself transforms as a scalar under (time-independent) changes of coordinates. Traditionally, one works with canonical transformations, which can be characterized as those transformations $(q,p) \to (\bar{q},\bar{p})$ which satisfy $p_i dq_i = \bar{p}_i d\bar{q}_i + dS$, for some scalar S.

The discussion of the preceding sections provides another point of view. The Hamiltonian H is now seen to be the t component of the 1-form $\gamma = p_i dq_i - H\,dt$; the fact that it appears to transform as a scalar is due to the restricted class of transformations previously considered, namely time-independent canonical transformations. Most generally, H should transform as one component of a covariant vector.

The traditional strategy of Hamiltonian perturbation theory in canonical coordinates can now be seen in another light. In the traditional strategy, one seeks mappings (or coordinate transformations) such that the first $2N$ components of γ remain inviolate, i.e., the $p_i dq_i$ part, while the last component, i.e., the term $-H\,dt$, is modified according to will. Of course, one allows for a gauge transformation, $\gamma \rightarrow \gamma + dS$, in this process.

An obvious generalization of this process is to transform all $2N+1$ components of γ, and not to pick on the Hamiltonian alone. Of course, this must be done with regard to some final form which one wishes γ to assume; that final form might well be $p_i dq_i - H\,dt$ in which, say, H is a function only of the p_i, or it could be something else. But even in the former case, one is not compelled to use either canonical coordinates or canonical transformations in the intermediate steps. Indeed, if a problem is originally formulated in noncanonical coordinates, then by definition the first $2N$ components of γ will not be represented by $p_i dq_i$. Thus, transforming the Hamiltonian to the desired form and finding appropriate canonical variables can be considered to be part of the same process, that of transforming γ.

Let us now take up the transformation properties of γ. Quite generally, we will transform γ by the rule

$$\bar{\gamma} = T^{*-1}\gamma + dS, \tag{34}$$

where $T: E \rightarrow E$, $Tz = \bar{z}$, is a diffeomorphism of the extended phase space onto itself. In components, this is

$$\bar{\gamma}_i(\bar{z}) = \frac{\partial z^j}{\partial \bar{z}^i}\gamma_j(z) + \frac{\partial S}{\partial \bar{z}^i}, \tag{35}$$

in accordance with the usual rule for covariant vectors.

These laws are good for any transformation T. But as in Sec. 2, let us consider mappings which are composed of advance maps derived from a set of vector fields G_1, G_2, etc. Through second order, we have

$$\begin{aligned} T^{*-1} &= \cdots \exp(-\epsilon^2 L_2)\exp(-\epsilon L_1) \\ &= 1 - \epsilon L_1 + \epsilon^2(\tfrac{1}{2}L_1^2 - L_2) + \cdots. \end{aligned} \tag{36}$$

The transformation of the coordinates themselves, shown in Eq. (6), involves T^*, which can be obtained from T^{*-1} by substituting $-L_n$ for L_n everywhere in Eq. (36), and reversing the order of all Lie derivative operators.

Let us now assume that $\bar{\gamma}$, γ, and S in Eq. (34) are all expanded in powers of ϵ, i.e., $\bar{\gamma} = \Sigma\epsilon^n\bar{\gamma}_n$, $\gamma = \Sigma\epsilon^n\gamma_n$, $S = \Sigma\epsilon^n S_n$. Then substituting Eq. (36) and collecting terms gives

$$\begin{aligned} \bar{\gamma}_0 &= \gamma_0 + dS_0, \\ \bar{\gamma}_1 &= \gamma_1 - L_1\gamma_0 + dS_1, \\ \bar{\gamma}_2 &= \gamma_2 - L_1\gamma_1 + (\tfrac{1}{2}L_1^2 - L_2)\gamma_0 + dS_2, \end{aligned} \tag{37}$$

and so forth. These equations are to be solved order by order for the vector fields G_n, implicitly contained in the L_n, and the scalars S_n, so that the $\bar{\gamma}_n$ have the new form desired in a given application. Similar formulas arise in traditional Hamiltonian perturbation theory, except that there the Lie derivative operators act on the 0-form H instead of the 1-form γ.

Equation (11) gives the definition of the Lie derivative on an arbitrary differential form. However, for a practical application of Eq. (37), there is considerable advantage in calling on the homotopy formula

$$L_X\alpha = i_X d\alpha + d i_X\alpha, \tag{38}$$

where X is an arbitrary vector field [represented in Eq. (37) by the G_n], where i_X is the interior product and where α is an arbitrary differential form. The interior derivative is defined as follows. If β is a k-form, then $i_X\beta$ is a $(k-1)$-form which satisfies

$$(i_X\beta)(X_1,\ldots,X_{k-1}) = \beta(X,X_1,\ldots,X_{k-1}), \tag{39}$$

where $X_1,\ldots,X_{k-1}$ are arbitrary vectors. The advantage in using Eq. (38) is that the second term is an exact differential form, which can be absorbed into the terms dS_n in Eq. (37), and the first term of Eq. (38) is easier to compute in practice than the entire expression. In addition, note that the first term of Eq. (38) involves the vector field X only algebraically, while the second involves its derivatives. We have found in practice that the second term of Eq. (38) often contains gauge dependent quantities that one is not interested in, anyway.

Thus we can reinterpret Eq. (37) in the following way, with no loss of generality. The operators L_n are no longer the Lie derivatives L_{G_n}, but rather $i_{G_n}d$; and the scalars S_n are redefined so as to absorb all the exact forms produced by the second term of Eq. (38). Then with α representing any of the 1-forms which arise in Eqs. (37), we have the component relation

$$(L_n\alpha)_i = G_n^j\left(\frac{\partial\alpha_i}{\partial z^j} - \frac{\partial\alpha_j}{\partial z^i}\right). \tag{40}$$

The entire labor of the perturbation calculation reduces to computing terms of this form, just as it reduces to computing Poisson brackets in ordinary Hamiltonian perturbation theory.

The terms $L_n\gamma_0$ in Eq. (37) are used to solve algebraically for the vectors G_n, and the terms dS_n represent differential equations for the scalars S_n. The system is generally underdetermined, as may be expected from the nature of the end product. For example, if one ends with $\bar{\gamma} = p_i dq_i - H(p_i)\,dt$, then this result can be subjected to an arbitrary canonical transformation of the form $\bar{p}_i = \bar{p}_i(p)$, $\bar{q}_i = \bar{q}_i(q,p)$ (a point transformation in the momenta). Sometimes gauge invariance or other considerations enter to restrict the possibilities.

We have applied this perturbation theory to analyze the resonance structure of magnetic field lines in physical systems which are nearly axisymetric. This is an example of a Hamiltonian system for which canonical coordinates are not given in the problem specification, and must be found. We will report on this and other applications in future papers.

ACKNOWLEDGMENTS

I would like to acknowledge several fruitful discussions with T. Frankel. This work was supported by the Office of Naval Research under Contract No. N00014-79-C-0537 and the Office of Fusion Energy of the U.S. Department of Energy under Contract DOE-DE-AM03-76500010 PA 26, Task I.

[1]H. Aref and N. Pomphrey, Phys. Lett. A **78**, 297 (1980).
[2]R. G. Littlejohn, J. Math. Phys. **20**, 2445 (1979).
[3]Francis P. Bretherton, J. Fluid Mech. **44**, 19 (1970).
[4]D. P. Stern, Am. J. Phys. **38**, 494 (1970).
[5]T. G. Northrop, M. D. Kruskal, and C. S. Liu, Phys. Fluids **9**, 1503 (1966).
[6]T. G. Northrop and J. A. Rome, Phys. Fluids **21**, 384 (1978).
[7]R. G. Littlejohn, Phys. Fluids **24**, 1730 (1981).
[8]P. J. Morrison and J. M. Greene, Phys. Rev. Lett. **45**, 790 (1980).
[9]P. J. Morrison, Phys. Lett. A **80**, 383 (1980).
[10]P. J. Morrison, Princeton Plasma Physics Laboratory Report No. PPPL-1783, 1981 (unpublished).
[11]M. D. Kruskal, J. Math. Phys. **3**, 806 (1962).
[12]B. McNamara, J. Math. Phys. **19**, 2154 (1978).
[13]V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978).
[14]R. Abraham and J. E. Marsden, *Foundations of Mechanics*, 2nd ed. (Benjamin/Cummings, Reading, Mass., 1978).
[15]J. -M. Souriau, *Structure des Systèmes Dynamiques* (Dunod, Paris, 1970).
[16]A. H. Nayfeh, *Perturbation Methods* (Wiley, New York, 1973).
[17]J. R. Cary, Phys. Rep. **79**, 131 (1981).
[18]M. Spivak, *Differential Geometry* (Publish or Perish, Berkeley, CA, 1979).
[19]A. A. Kamel, Cel. Mech. **3**, 90 (1970).
[20]A. J. Dragt and J. M. Finn, J. Math. Phys. **17**, 2215 (1976).
[21]A. Deprit, Cel. Mech. **1**, 12 (1969).

Electromagnetic multipole propagation in a homogeneous conducting wholespace

G. J. Weir
Applied Mathematics Division, Department of Scientific and Industrial Research, PO Box 1335, Wellington, New Zealand

(Received 3 September 1980; accepted for publication 18 March 1981)

The electric and magnetic field components are expanded in terms of spin weighted spherical harmonics, thereby defining a multipole structure whose propagation through a homogeneous conducting wholespace is obtained. For current sources contained inside the unit sphere, the resultant exterior electromagnetic fields are uniquely associated with either currents from electrodes placed inside the unit sphere, or from insulated wires inside the unit sphere. The former fields have transverse magnetic fields, while the latter (with the exception of the magnetic monopole) have transverse electric fields. When displacement currents are ignored, the fields from step function current sources are described by the incomplete gamma functions, while the exact solutions containing the hyperbolic contributions are obtained from convolutions with integer order modified Bessel functions and functions dependent on the temporal behavior of the current source.

PACS numbers: 03.50.De

1. INTRODUCTION

The number of exact solutions to the idealized problems of geophysical prospecting is very small. In order to overcome mathematical difficulties posed by Maxwell's equations, many attempts to construct relevant solutions assume the quasistatic approximation, or equivalently; contributions from the hyperbolic wave terms are ignored. Even then, the number of known quasistatic solutions is small. Subject to the quasistatic approximation, solutions in the form of known tabulated functions have been given for finite electric and magnetic dipoles excited by step function current sources in a homogeneous conducting wholespace. To the author's knowledge these are the only published wholespace solutions excited by step function current sources. The aim of this paper is to present the general multipole solution to Maxwell's equations in a homogeneous conducting wholespace, where current sources are contained inside the unit sphere and where the sources have a step function time behavior.

Our study of Maxwell fields in a homogeneous conducting wholespace begins by utilizing the spin harmonics.[1] We have chosen to employ the nonrelativistically invariant electric and magnetic fields, since the presence of a nonzero conductor disturbs the elegance of Maxwell's equations. This can be seen by comparing the equations of this paper [see Eqs. (2.12)–(2.15)] with those of Goldberg *et al.*,[2] but this is not considered a disadvantage since in terrestrial experiments, for example, we often have effectively a unique time coordinate, and so there is no need to apply Lorentz transformations to the field.

In Sec. 2 the spin harmonics are defined, and then we introduce a spin base and derive a simple form of Maxwell's equations in a uniform conductor. These equations are solved in Sec. 3 for steady state fields. In Sec. 4 we consider a set of particular solutions whose magnetic field is transverse, while Sec. 5 considers fields with transverse electric fields. Section 6 contains the conclusion. We use SI units throughout this paper, and our summation convention (e.g., ΣY_{lm}) means summation over all possible values (of l and m). Only integral spin fields are considered.

This paper is an extension of an unpublished research report,[3] which determined the quasistatic approximations to the solutions of Sec. 4. We shall only consider fields which are bounded at spatial infinity, and are produced by step function current sources.

2. SPIN HARMONICS

In this section we shall collect results which will be needed later.

In a spherical polar coordinate system (r, θ, ϕ), the components $X_r, X_\theta + iX_\phi, X_\theta - iX_\phi$ have spin weight[4] 0, 1, -1, respectively, where $X = X_\mu \omega^\mu$ is a one-form.[5] The spin harmonics ${}_sY_{lm}$,

$$
\begin{aligned}
{}_sY_{lm} &= e^{im\phi}\sin^{2l}\left(\frac{\theta}{2}\right)\sum a_{slmn}\left(\cot\frac{\theta}{2}\right)^{2n+s-m},\\
a_{slmn} &= (-1)^{l-n-s}C^{l-s}_{n}C^{l+s}_{n+s-m}\\
&\quad\times\left(\frac{(l+m)!(l-m)!(2l+1)}{(l+s)!(l-s)!4\pi}\right)^{\frac{1}{2}},\\
{}_{-s}Y_{lm} &= (-1)^{s+m}{}_s\overline{Y}_{lm},\\
&\qquad |s|\leqslant l,\quad |m|\leqslant l;\quad m,s,l \text{ integers},
\end{aligned}
$$

provide a complete orthonormal set of functions for regular integral spin fields on the unit sphere, and can be used, for example, as a basis for regular electric fields,

$$E_0 = E_r = \sum {}_0\kappa_{lm}(r,t)\,{}_0Y_{lm}(\theta,\phi), \tag{2.1}$$

$$E_+ = E_\theta + iE_\phi = \sum {}_1\kappa_{lm}(r,t)\,{}_1Y_{lm}(\theta,\phi), \tag{2.2}$$

$$E_- = E_\theta - iE_\phi = \overline{E}_+, \tag{2.3}$$

where ${}_0Y_{lm}$ are the ordinary spherical harmonics and

 0022-2488/82/050748-05$02.50

${}_0\kappa_{lm}, {}_1\kappa_{lm}$ are fixed uniquely through Eqs. (2.1) and (2.2). For a spin field χ of spin weight s, the differential operators $\eth$ and $\bar{\eth}$ are defined as

$$\eth\chi = -(\partial_\theta + i\csc\theta\partial_\phi - s\cot\theta)\chi, \tag{2.4}$$

$$\bar{\eth}\chi = -(\partial_\theta - i\csc\theta\partial_\phi + s\cot\theta)\chi, \tag{2.5}$$

$$\partial_x = \frac{\partial}{\partial x},$$

and satisfy

$$\eth_s Y_{lm} = ((l-s)(l+s+1))^{\frac{1}{2}}{}_{s+1}Y_{lm}, \tag{2.6}$$

$$\bar{\eth}_s Y_{lm} = -((l+s)(l-s+1))^{\frac{1}{2}}{}_{s-1}Y_{lm}, \tag{2.7}$$

where $|s| \leqslant l$ and $|m| \leqslant l$.

In a homogeneous wholespace of conductivity σ, Maxwell's equations can be expressed as four linear first-order complex equations among the two real (E_0, B_0) and two complex (E_+, B_+) field variables. This is illustrated by the following sequence:

$$\begin{aligned}\eth B_0 &= -\partial_\theta B_r - i\csc\theta\partial_\phi B_r \\ &= r(\nabla\times B)_\phi - \partial_r(rB_\theta) - i(r(\nabla\times B)_\theta + \partial_r(rB_\phi)) \\ &= r(\mu\sigma E_\phi + \mu\epsilon\partial_t E_\phi) - ir(\mu\sigma E_\theta + \mu\epsilon\partial_t E_\theta) \\ &\quad -\partial_r r(B_\theta + iB_\phi),\end{aligned}$$

and so

$$\eth B_0 = -ir(\mu\sigma E_+ + \mu\epsilon\partial_t E_+) - \partial_r r B_+, \tag{2.8}$$

$$\eth E_0 = ir\partial_t B_+ - \partial_r r E_+, \tag{2.9}$$

$$\bar{\eth} r B_+ = \partial_r(r^2 B_0) - ir^2(\mu\sigma E_0 + \mu\epsilon\partial_t E_0), \tag{2.10}$$

$$\bar{\eth} r E_+ = \partial_r(r^2 E_0) + ir^2\partial_t B_0, \tag{2.11}$$

where the derivation of Eqs. (2.9)–(2.11) follows that for Eq. (2.8). It is straightforward to show that Maxwell's equations in a uniform conductor are equivalent to Eqs. (2.8)–(2.11); and since E_- and B_- do not appear in these equations, it is sufficient to consider only E_0, E_+, B_0, B_+ when describing electric and magnetic fields.

The second-order equations following from Eqs. (2.8)–(2.11) are

$$\bar{\eth}\eth E_0 + \partial_r^2 r^2 E_0 = \mu\sigma r^2\partial_t E_0 + \mu\epsilon r^2\partial_t^2 E_0, \tag{2.12}$$

$$\eth\bar{\eth} E_+ + r\partial_r^2 r E_+ = 2\eth E_0 + r^2\mu\sigma\partial_t E_+ + r^2\mu\epsilon\partial_t^2 E_+, \tag{2.13}$$

$$\bar{\eth}\eth B_0 + \partial_r^2 r^2 B_0 = \mu\sigma r^2\partial_t B_0 + \mu\epsilon r^2\partial_t^2 B_0, \tag{2.14}$$

$$\eth\bar{\eth} B_+ + \partial_r r^2\partial_r B_+ = 2\eth B_0 + \mu\sigma r^2\partial_t B_+ + \mu\epsilon r^2\partial_t^2 B_+ . \tag{2.15}$$

3. STEADY STATE FIELDS

In order to provide a simple example of the application of spin weighted spherical harmonics, as well as to obtain results which are needed later, we shall discuss those steady state fields which are regular outside the unit sphere. A steady state electric field in a homogeneous conducting medium satisfies $\nabla\times\mathbf{E} = 0 = \nabla\cdot\mathbf{E}$, and so

$$\mathbf{E} = \nabla e, \tag{3.1}$$

where $\nabla^2 e = 0$. Then Eqs. (3.1) and (2.4) allow us to write

$$\begin{aligned}\mathbf{E} &= (E_0, E_+) \\ &= (\partial_r e, (\partial_\theta + i\csc\theta\partial_\phi)(e/r)) \\ &= (\partial_r e, \eth(e/r)),\end{aligned} \tag{3.2}$$

and provided that e is regular on the unit sphere,

$$e = \sum e_{lm}(r)_0 Y_{lm}(\theta,\phi), \tag{3.3}$$

where $e_{lm}(r)$ are functions of the radial coordinate r. From Eqs. (3.2), (3.3), and (2.6),

$$\mathbf{E} = \sum[(\partial_r e_{lm})_0 Y_{lm}, -e_{lm}(l(l+1))^{\frac{1}{2}}{}_1Y_{lm}/r] \tag{3.4}$$

and from Eqs. (2.8)–(2.11), Maxwell's equations are satisfied providing $\eth E_0 = -\partial_r(rE_+)$ and $\bar{\eth}(rE_+) = \partial_r(r^2 E_0)$, or equivalently

$$\partial_r r^2\partial_r e_{lm} = l(l+1)e_{lm},$$

which implies (as it must, of course) that

$$e_{lm} = -\alpha_{lm} r^{-l-1},$$

$$\mathbf{E} = \sum\alpha_{lm} r^{-l-2}[(l+1)_0 Y_{lm}, (l(l+1))^{\frac{1}{2}}{}_1Y_{lm}], \tag{3.5}$$

where α_{lm} are constants, and only terms which are bounded at spatial infinity have been retained.

It remains to solve for the magnetic field components. These cannot be unique (unless the required boundary conditions are given), since the magnetic field in uniform conductors is defined by the electric field to within the gradient of a harmonic function. Since the analytic form of such regular functions is given explicitly in Eq. (3.5), it only remains to observe that the particular solution

$$\begin{aligned}\mathbf{B} &= (0, -i\mu\sigma\alpha_{00}\cot(\theta/2))/\sqrt{4\pi}r \\ &\quad + \sum(l>0)[0, i\mu\sigma\alpha_{lm} r^{-l-1}((l+1)/l)^{1/2}{}_1Y_{lm}]\end{aligned} \tag{3.6}$$

is satisfied by Eqs. (2.8)–(2.11), subject to Eq. (3.5), where $\Sigma(l>0)$ means summation of l from one to positive infinity, and m from $-l$ to l. The expression $\cot((\theta/2)/r)$ has been called a singular spin harmonic[6] and results presumably from the presence of a "conduction path" along which charge has passed to the coordinate origin. Clearly, the general asymptotically regular, steady state magnetic field is found by replacing α_{lm} in Eq. (3.5) by arbitrary constants, b_{lm} say, and adding the result to Eq. (3.6). Strictly, the term resulting from α_{00} in Eq. (3.6) is not regular at spatial infinity (for $\theta = 0$), but nevertheless, we shall retain it in what follows.

4. TRANSIENT TRANSVERSE MAGNETIC FIELDS

In this section we shall investigate the transient development of electromagnetic fields which result from a step function current source, and which tend with time to the steady state fields in Eqs. (3.5) and (3.6). Initially we shall only seek a particular solution to this problem, deferring the homogeneous solutions to the next section. First, note that the following canonical form for the magnetic field,

$$B_0 = 0,$$

$$B_+ = \frac{-i\mu\sigma\alpha_{00}\alpha_0(r,t)\cot(\theta/2)}{(4\pi r)^{1/2}} + \sum(l>0)i\mu\sigma\left(\frac{l+1}{l}\right)^{1/2}\frac{\alpha_{lm}\alpha_l(r,t)\,{}_1Y_{lm}}{r^{1/2}}, \quad (4.1)$$

will supply the required particular solution, provided we can solve Eq. (4.2),

$$(\partial_r^2 + r^{-1}\partial_r - (l+\tfrac{1}{2})^2 r^{-2} - \mu\sigma\partial_t - \mu\epsilon\partial_t^2)\alpha_l = 0, \quad (4.2)$$

subject to the following initial and boundary conditions,

$$\lim_{r\to 0}[r^{l+1/2}\alpha_l(r,t)] = 1, \quad t>0 \quad (4.3)$$

$$\alpha_l = 0 = \partial_t\alpha_l, \quad t=0 \quad (4.4)$$

$$\left(\lim_{r\to\infty}\alpha_l\right) \text{ is bounded,} \quad (4.5)$$

where Eq. (4.2) follows from Eqs. (4.1) and (2.15); and Eq. (4.3) states that the steady state *magnetic* field is established at the origin for all positive times. We have used only l as a subscript for α_l, since α_l is obviously independent of m.

Introducing the Laplace transform of α_l, $\mathfrak{L}\alpha_l$,

$$\mathfrak{L}\alpha_l = \int_0^\infty e^{-st}\alpha_l dt,$$

allows Eqs. (4.2) and (4.4) to be rewritten as

$$(\partial_r^2 + r^{-1}\partial_r - r^{-2}(l+\tfrac{1}{2})^2 - \mu\sigma s - \mu\epsilon s^2)\mathfrak{L}\alpha_l = 0,$$

or

$$\mathfrak{L}\alpha_l = c_1 I_{l+1/2}(p) + c_2 K_{l+1/2}(p), \quad (4.6)$$

$$p = (\mu\sigma s + \mu\epsilon s^2)^{1/2} r, \quad (4.7)$$

where I and K are modified (spherical) Bessel functions and c_1 and c_2 are independent of r. From Eqs. (4.5) and (4.6), $c_1 = 0$, while from Eqs. (4.6) and (4.3),

$$\mathfrak{L}\alpha_l = p^{l+1/2}K_{l+1/2}(p)/s\Gamma(l+\tfrac{1}{2})2^{l-1/2}r^{l+1/2}. \quad (4.8)$$

The inverse Laplace transform of Eq. (4.8) can be performed by deforming the contour integral into an anticlockwise loop around the segment of the real axis between $s = -\sigma/\epsilon$ and $s = 0$. There is a pole at $s = 0$, and a branch cut between $s = 0$ and $s = -\sigma/\epsilon$, since the multivalueness of $p^{l+1/2}$ can be absorbed into $K_{l+1/2}(p)$. A straightforward calculation yields

$$\alpha_l = r^{-(l+1/2)}\left[1 - \int_0^{\sigma/\epsilon}\frac{e^{-ut}q^{l+1/2}J_{l+1/2}(q)du}{u\Gamma(l+\tfrac{1}{2})2^{l+1/2}r^{l+1/2}}\right], \quad (4.9)$$

$$q = (\mu\sigma u - \mu\epsilon u^2)^{1/2} r,$$

where $J_{l+1/2}$ is an ordinary Bessel function of order $l+\tfrac{1}{2}$.

The corresponding expressions for the electric fields can be obtained directly. From Eqs. (2.10), (4.1), and (4.8),

$$\mathfrak{L}E_0 = \sum\frac{(l+1)\alpha_{lm}p^{l+1/2}K_{l+1/2}(p)\,{}_0Y_{lm}}{s(1+\epsilon s/\sigma)\Gamma(l+\tfrac{1}{2})r^{l+2}}, \quad (4.10)$$

while from Eqs. (2.8), (4.1), and (4.8), and setting $l>0$,

$$\mathfrak{L}E_+ = \sum(l>0)\frac{\alpha_{lm}((l+1)/l)^{1/2}\,{}_1Y_{lm}p^{l+1/2}}{s(1+\epsilon s/\sigma)\Gamma(l+\tfrac{1}{2})2^{l-1/2}r^{l+2}} \times[pK_{l+3/2}(p) - (l+1)K_{l+1/2}(p)], \quad (4.11)$$

where $\mathfrak{L}E_0$, $\mathfrak{L}E_+$ are the Laplace transforms of E_0, E_+, and so

$$E_0 = \sum\left[\alpha_{lm}(l+1)_0Y_{lm}(1-e^{-\sigma t/\epsilon}) - \frac{\alpha_{lm}(l+1)_0Y_{lm}}{r^{l+2}\Gamma(l+\tfrac{1}{2})2^{l+1/2}} \times\int_0^{\sigma/\epsilon}\frac{e^{-ut}q^{l+1/2}J_{l+1/2}(q)du}{u(1-\epsilon u/\sigma)}\right], \quad (4.12)$$

while for $l>0$,

$$E_+ = \sum(l>0)\left[\alpha_{lm}(l(l+1))^{1/2}\,{}_1Y_{lm}r^{-l-2}(1-e^{-\sigma t/\epsilon}) + \frac{\alpha_{lm}((l+1)/l)^{1/2}\,{}_1Y_{lm}}{\Gamma(l+1/2)2^{l+1/2}r^{l+2}}\int_0^{\sigma/\epsilon}\frac{e^{-ut}}{u(1-\epsilon u/\sigma)} \times[q^{l+1/2}J_{l+1/2}(q) - q^{l+3/2}J_{l+3/2}(q)]du\right]. \quad (4.13)$$

It is possible to rewrite the solutions for B_+, E_0, and E_+, given in Eqs. (4.1), (4.9), (4.12), (4.13), by utilizing the following relationship:

$$\Phi_l \overset{(defn)}{=} \frac{\alpha^{(l+1)}k^{2l+1}e^{-\alpha t}I_{l+1}(\alpha(t^2-k^2)^{1/2})}{(2l-1)!!(t^2-k^2)^{(l+1)/2}}, \quad t>k$$

$$= \mathfrak{L}^{-1}\left(\frac{p^{l+1/2}K_{l+1/2}(p)}{2^{l-1/2}\Gamma(l+\tfrac{1}{2})}\right), \quad (4.14)$$

where $\mathfrak{L}^{-1}$ is the inverse Laplace transform, and

$$\alpha = \sigma/2\epsilon, \quad k = (\sqrt{\mu\epsilon})r. \quad (4.15)$$

Equation (4.14) can be proved by induction by first observing that the $l = 0$ case is given in tables of inverse Laplace transforms[7], and finally, that the operator $k^{2l+1}\partial_k k^{-(2l+1)}$ maps the result for l into the corresponding result for $l+1$. We can now write the fields B_+, E_0, E_+ as a convolution of Φ_l with given functions of time; for example, from Eqs. (4.14) and (4.10),

$$E_0 = \sum\left(\frac{(l+1)\alpha_{lm\,0}Y_{lm}}{r^{l+2}} - \frac{\alpha_{lm}(l+1)_0Y_{lm}}{r^{3/2}}\int_t^\infty\Phi_l(t-\tau)(1-e^{-\sigma\tau/\epsilon})d\tau\right), \quad (4.16)$$

where we have written the solution in a form which avoids the light cone where $k = t$.

When displacement currents are ignored, or equivalently, ϵ is set to zero, the fields above take simple forms. By setting ϵ to zero in the expression for α_l in Eq. (4.9), we find that

$$\alpha_l = r^{-(l+1/2)} - \int_0^\infty\frac{p^{l-1/2}e^{-p^2t/\mu\sigma}}{2^{l-1/2}\Gamma(l+\tfrac{1}{2})}J_{l+1/2}(rp)dp$$

$$= \Gamma(l+\tfrac{1}{2}, \zeta^2)/r^{l+1/2}\Gamma(l+\tfrac{1}{2}),$$

$$\zeta^2 = \mu\sigma r^2/(4t),$$

$$\Gamma(\alpha,z) = \int_z^\infty e^{-u}u^{\alpha-1}du,$$

where we have utilized a well known Bessel transform.[8]

Consequently, *in the quasistatic approximation* ($\epsilon = 0$),

$$B_0 = 0,$$

$$B_+ = -\frac{i\mu\sigma\alpha_{00}\Gamma(\frac{1}{2},\zeta^2)\cot(\theta/2)}{\sqrt{4\pi}r\Gamma(\frac{1}{2})} + \sum(l>0)\Big(\frac{l+1}{l}\Big)^{1/2}\frac{i\mu\sigma\alpha_{lm}\Gamma(l+\frac{1}{2},\zeta^2)_1Y_{lm}}{r^{l+1}\Gamma(l+\frac{1}{2})},$$

$$E_0 = \sum\frac{(l+1)\alpha_{lm}\Gamma(l+\frac{1}{2},\zeta^2)_0Y_{lm}}{r^{l+2}\Gamma(l+\frac{1}{2})},$$

$$E_+ = 0 \quad \text{for } l=0,$$

$$= \sum(l>0)\frac{(l(l+1))^{1/2}\alpha_{lm\,1}Y_{lm}}{\Gamma(l+\frac{1}{2})r^{l+2}} \times\Big[\Gamma(l+\tfrac{1}{2},\zeta^2) + \frac{2}{l}\zeta^{2l+1}e^{-\zeta^2}\Big] \quad (l>0).$$

By integrating repeatedly by parts, it can be shown that

$$\frac{\Gamma(l+\frac{1}{2},\zeta^2)}{\Gamma(l+\frac{1}{2})} = \operatorname{erfc}\zeta + \frac{2}{\sqrt{\pi}}e^{-\zeta^2}\sum_{j=0}^{l-1}\frac{2^j\zeta^{2j+1}}{(2j+1)!!},$$

where the summation term on the right is zero for $l = 0$.

This completes our derivation of the transverse magnetic fields. Since these fields have nonzero steady state electric fields, it is natural to associate them with electrodes placed inside the unit sphere. Finally, the lowest order physically interesting solution is that of the infinitesimal electric dipole which follows by setting

$$\alpha_{10} = I\delta l/2\sigma(2\pi)^{1/2},$$

where δl is the "length" of the dipole. For this case, the solutions above are well known.[9] The only difference between our dipole results and those presented earlier are in Φ_1 in Eq. (4.14), where we have dropped the wave front delta functions, and the singular derivatives of the wave front delta functions, which were included in the earlier derivations.[10]

5. TRANSIENT TRANSVERSE ELECTRIC FIELDS

We shall begin this section by discussing the general transient electromagnetic fields which are bounded at spatial infinity, and which are established by a step function current. Such fields must tend with time to steady state solutions, and as we have found a particular set of fields which tend with time to the steady solutions in Eqs. (3.5) and (3.6), any additional fields must tend to the "complimentary" steady state fields mentioned in Sec. 3,

$$B_0 = \sum b_{lm}r^{-l-2}(l+1)_0Y_{lm},$$

$$B_+ = \sum b_{lm}r^{-l-2}(l(l+1))^{1/2}{}_1Y_{lm},$$

$$E_0 = 0 = E_+.$$

These fields are found by writing

$$B_0 = r^{-3/2}\sum b_{lm}\beta_l(r,t)(l+1)_0Y_{lm}, \tag{5.1}$$

$$B_+ = r^{-1/2}\sum b_{lm}\gamma_l(r,t)(l(l+1))^{1/2}{}_1Y_{lm}, \tag{5.2}$$

and solving Eqs. (2.14) and (2.15),

$$(\partial_r^2 + r^{-1}\partial_r - r^{-2}(l+\tfrac{1}{2})^2 - \mu\sigma\partial_t - \mu\epsilon\partial_t^2)\beta_l = 0,$$

$$(\partial_r^2 + r^{-1}\partial_r - r^{-2}(l+\tfrac{1}{2})^2 - \mu\sigma\partial_t - \mu\epsilon\partial_t^2)\gamma_l = 2r^{-3}(l(l+1))^{1/2}\beta_l,$$

to the initial and boundary conditions

$$\beta_l = 0 = \gamma_l, \quad t=0,$$

$$\partial_t\beta_l = 0 = \partial_t\gamma_l, \quad t=0,$$

$$\lim_{r\to0}(r^{l+1/2}\beta_l) = 1 = \lim_{r\to0}(r^{l+3/2}\gamma_l), \quad t>0,$$

$$\lim_{r\to\infty}\beta_l,\ \lim_{r\to\infty}\gamma_l \text{ bounded.}$$

Taking Laplace transforms, and using our earlier notation, the general solution to the above equations is

$$\mathfrak{L}\beta_l = \frac{p^{l+1/2}K_{l+1/2}(p)}{2^{l-1/2}\Gamma(l+\frac{1}{2})sr^{l+1/2}}, \tag{5.3}$$

$$\mathfrak{L}\gamma_0 = 0, \tag{5.4}$$

$$\mathfrak{L}\gamma_l = \frac{p^{l+3/2}K_{l+3/2}(p) - (l+1)p^{l+1/2}K_{l+1/2}(p)}{l\,2^{l-1/2}\Gamma(l+\frac{1}{2})sr^{l+3/2}} + P(s)K_{l+1/2}(p), \tag{5.5}$$

where $P(s)$ is an arbitrary function of s. Reference to the previous section shows that $P(s)$ generates the transverse magnetic fields found earlier, so we shall set $P(s)$ to zero,

$$P(s) = 0. \tag{5.6}$$

Then, from Eqs. (2.10), (2.8), and (5.1)–(5.6),

$$\mathfrak{L}E_0 = \frac{\sqrt{2}ib_{00\,0}Y_{00}}{(\mu\sigma+\mu\epsilon s)s\Gamma(\frac{1}{2})r^3}[p^{3/2}K_{3/2}(p) - p^{\frac{1}{2}}K_{\frac{1}{2}}(p)],$$

$$\mathfrak{L}E_+ = -\sum(l>0)i\Big(\frac{l+1}{l}\Big)^{1/2} \times\frac{b_{lm}p^{l+1/2}K_{l+1/2}(p)_1Y_{lm}}{r^{l+1}2^{l-1/2}\Gamma(l+\frac{1}{2})}, \tag{5.7}$$

and so from Eq. (4.14),

$$E_+ = -\sum(l>0)i\Big(\frac{l+1}{l}\Big)^{1/2}\frac{b_{lm}\Phi_{l\,1}Y_{lm}}{r^{l+1}}. \tag{5.8}$$

Consequently, with the exception of the magnetic monopole (b_{00}) term, the electric field is transverse and can be found in closed form. The latter fact is suggested by the work of Wait,[11] who has derived the explicit form (modulo wave front delta function terms) of Φ_1.

From the results in Sec. 4, the solutions for B_0, B_+, and E_0 may be written down using the Green's function solution based on Φ_l in Eq. (4.14), or using the half-integer order (spherical) ordinary Bessel functions; for example,

$$B_0 = \sum\frac{(l+1)b_{lm\,0}Y_{lm}}{r^{l+2}} \times\Big[1 - \int_0^{\sigma/\epsilon}\frac{e^{-ut}q^{l+1/2}J_{l+1/2}(q)du}{u\Gamma(l+\frac{1}{2})2^{l+1/2}}\Big].$$

We shall not pursue this further, however, but instead shall present the quasistatic solutions which follow by setting ϵ to zero. For $\epsilon = 0$,

$$B_0 = \sum \frac{b_{lm}(l+1)_0 Y_{lm}\Gamma(l+\frac{1}{2},\zeta^2)}{r^{l+2}\Gamma(l+\frac{1}{2})}, \tag{5.9}$$

$$B_+ = \sum(l>0)\frac{(l(l+1))^{1/2}b_{lm\,1}Y_{lm}}{\Gamma(l+\frac{1}{2})r^{l+2}} \times\left(\Gamma(l+\tfrac{1}{2},\zeta^2) + \frac{2}{l}\zeta^{2l+1}e^{-\zeta^2}\right), \tag{5.10}$$

$$E_0 = \frac{b_{00}i_0Y_{00}e^{-\zeta^2}}{2(\sqrt{\pi})rt}, \tag{5.11}$$

$$E_+ = -\sum(l>0)ib_{lm}\left(\frac{l+1}{l}\right)^{1/2}\frac{{}_1Y_{lm}\zeta^{2l+1}e^{-\zeta^2}}{t\Gamma(l+\frac{1}{2})r^{l+1}}. \tag{5.12}$$

By construction, the steady state electric fields in this section have all been zero, and so it is natural to associate the fields in this section as arising from insulated current carrying wires. The lowest order, physically interesting, example is that of the infinitesimal magnetic dipole, which follows by setting

$$b_{10} = \mu I dA/2(2\pi)^{1/2},$$

where dA is the dipole's "area". For this case, the solutions above agree with earlier results.[12]

6. CONCLUSION

We have constructed all Maxwell fields in a homogeneous wholespace which are established by a step function current source, and are regular outside of the unit sphere. With the exception of the (nonphysical) magnetic monopole, fields from insulated wires inside the unit sphere produce transverse electric fields, while fields from electrodes inside the unit sphere produce transverse magnetic fields. Such a decomposition into transverse electric, and magnetic, fields is unique, and as a current source with an arbitrary time dependence can be constructed from step function current sources, this decomposition is also quite general.

When displacement currents are ignored, the fields from step function current sources are described by the incomplete gamma functions, while the exact solutions containing the hyperbolic contributions are obtained from convolutions with integer order modified Bessel functions and functions dependent on the temporal behavior of the current source.

The spin weighted spherical harmonics played a central role in our solution technique. Similar approaches appear in the book by Gel'fand, Minlos, and Shapiro,[13] and in the paper by Goldberg *et al.*[14] These works considered empty space Maxwell solutions, and showed that in a spherical polar coordinate system, the angular dependence of such fields could be completely separated by using field components which were irreducible under the action of the three dimensional rotation group. Gel'fand *et al.* utilized the vector potential, whereas Goldberg *et al.* used the three components of the Maxwell spinor.[15] Both approaches, as well as a Hertz vector approach, could be used in the presence of finite conductivity, but the resulting equations contain, either implicitly or explicitly, components with spin weight negative one, since the natural gauge condition ($\nabla\cdot A = \mu\sigma\psi + \mu\epsilon\partial_t\psi$) reads

$$\partial_r(r^2A_0) - \mathrm{Re}(r\bar{\eth}A_+) = r^2(\mu\sigma\psi + \mu\epsilon\partial_t\psi),$$

where (A_0, A_+) is the vector potential, ψ the scalar potential, and Re denotes the real part of an expression. In order to avoid such reality conditions the physical field components (E_0, E_+) and (B_0, B_+) were used. (However, any such choice is simply a matter of personal preference, and each option has probably some merit.)

Finally, the only solutions known to the author which are directly analogous to those derived in this paper are the infinitesimal electric and magnetic dipoles, both of which are well known. The solutions for the higher multipoles appears to be new, and as an arbitrary bounded current source can be expanded in a multipole expansion, our expressions above can be used, for example, to estimate how closely a given source field is approximated by dipole fields.

ACKNOWLEDGMENT

The author benefited from a discussion with C. Earl on the derivation of Eq. (4.14).

[1]E. T. Newman and R. Penrose, J. Math. Phys. **7**, 862 (1966).
[2]J. N. Goldberg, A. J. MacFarlane, E. T. Newman, F. Rohrlich, and E. G. C. Sudarshan, J. Math. Phys. **8**, 2155 (1967).
[3]G. J. Weir, Technical Report No. 85, Applied Math. Div., Dept. Sci. Indust. Research, Wellington, N.Z. (1979).
[4]Reference 1.
[5]H. Flanders, *Differential Forms with Applications to the Physical Sciences* (Academic, New York, 1963).
[6]G. J. Weir, J. Math. Phys. **20**, 1648 (1979).
[7]G. E. Roberts and H. Kaufman, *Table of Laplace Transforms* (Saunders, Philadelphia, 1966).
[8]F. Oberhettinger, *Table of Bessel Transforms* (Springer-Verlag, Berlin, 1972).
[9]J. R. Wait, Appl. Sci. Res. B **8**, 213 (1960).
[10]Reference 9.
[11]Reference 9.
[12]Reference 9.
[13]I. M. Gel'fand, R. A. Milnos, and Z. Ya Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications* (McMillan, New York, 1963).
[14]Reference 2.
[15]R. Penrose, Ann. Phys. **10**, 171 (1960).

Nonrecurrence of the stochastic process for the hydrogen atom problem in stochastic electrodynamics

Pierre Claverie and Francisco Soto [a)]

Laboratoire de Chimie Quantique, Institut de Biologie Physico-Chimique 13, rue Pierre et Marie Curie 75005, Paris, France

(Received 4 December 1980; accepted for publication 6 March 1981)

It is shown, following a criterion borrowed from Khas'minskii, that the stochastic process associated with the (approximate) Fokker–Planck equation of the hydrogen atom problem in stochastic electrodynamics (SED) is nonrecurrent and therefore also nonergodic. The demonstration of this nonrecurrence property does not use any explicit solution. The property implies, among other things, that all the invariant measures of the process will be nonfinite. Some remarks concerning the consequences for SED are made.

PACS numbers: 03.50.Kk, 41.70. + t, 02.50.Ey

I. INTRODUCTION

Stochastic electrodynamics (SED) is classical electrodynamics (including Lorentz–Dirac radiation damping) supplemented by the assumption that there exists a stochastic electromagnetic field filling up the whole space. From various arguments (e.g., retrieving the ground state energy of the quantal harmonic oscillator,[1] or requiring relativistic invariance for the spectrum[2]; it may be found that this background field must have zero mean value and a spectral density[1–5]

$$s_\epsilon(\omega) = \frac{2\hbar}{3c^3}|\omega|^3, \tag{1}$$

i.e., we recognize the properties of the vacuum state of quantum electrodynamics.

The theory constitutes a well-defined problem of mathematical physics, but its solution is rather difficult, due to the nonwhite character of the random background field, and the nonstandard techniques required for the treatment of nonlinear systems have been worked out only recently [Refs. 6–9 and references therein].

For a single nonrelativistic charged particle the equation of motion in SED is the Braffort–Marshall equation.[1,3,5,10,11]

$$m\ddot{\mathbf{r}} = \mathbf{F}(\mathbf{r}) + m\tau\dddot{\mathbf{r}} + e\mathbf{E}(t), \tag{2}$$

where m is the mass of the particle, e its charge and $\tau = 2e^2/3mc^3$; $\mathbf{F}(\mathbf{r})$ is the external (known) force and $\mathbf{E}(t)$ the electric field due to the background radiation. This last term is written in the electric-dipole approximation, which neglects the magnetic force of the background field and the spatial dependence of the electric field. Starting with (2) and using a stochastic Liouville equation, a generalized Fokker–Planck equation (FPE) (of a rather complicated form) for the distribution in phase space may be derived[12,13]; the process obtained turns out to be definitely non Markoffian in phase space. The fact that the damping and stochastic force are small with respect to the "deterministic" Hamiltonian[5] makes it possible to use perturbation methods and find several approximations of the usual type (second-order partial) derivatives to the generalized FPE.[5,8] The lack of uniqueness of such equations is due to the non-Markoffian character mentioned above. These equations may be reduced to a unique FPE in terms of a smaller number of variables namely some relevant constants of notion corresponding to the unperturbed deterministic motion[6,8] (for the case of multiperiodic systems, such relevant constants of motion are the action variables).

This reduced FPE can also be obtained directly by calculating, through perturbation methods, the variations of these constants of motion under the effect of the damping and stochastic force and by averaging these variations to get the drift and diffusion coefficients.[6,9]

In the case of linear system, although the predictions of this FPE are not fully identical with those of quantum theory, they are rather satisfactory.[1,3–5,10] However, in the case of nonlinear systems, such as the anharmonic oscillator and the Kepler problem (Coulomb potential) the results obtained until now, from the reduced FPE are not in agreement with quantum theory.[8,14]

We show in this paper, using a criterion borrowed from Khas'minskii,[15] that the process associated with the Kepler problem FPE is nonrecurrent and therefore nonergodic. We examine the implications of this nonrecurrence property and its possible origins.

II. THE KEPLER SYSTEM

In the Kepler problem[6,7] we have the Coulomb potential

$$V(r) = -\kappa/r, \tag{3}$$

κ being a positive constant. The suitable variables for writing the reduced Fokker–Planck equation are the energy of the particle E, its total angular momentum M and the eccentricity ϵ given by

$$\epsilon = (1 + 2EM^2/m\kappa^2)^{1/2}. \tag{4}$$

or the variable $\eta = (1-\epsilon^2)^{1/2}$. We restrict our attention to bound states, thus $E<0$. In this case the reduced FPE is[6,8]

[a)]Research fellow from the UNAM (Universidad Nacional Autonoma de Mexico).

$$8\pi^2 MT\frac{\partial W}{\partial t}$$
$$=\frac{\partial}{\partial E}\Big[C^E W\Big]+\frac{\partial}{\partial M}\Big[C^M W\Big]$$
$$+\frac{\partial}{\partial E}\Big[C^{EE}\frac{\partial W}{\partial E}\Big]+\frac{\partial}{\partial E}\Big[C^{EM}\frac{\partial W}{\partial M}\Big]+\frac{\partial}{\partial M}\Big[C^{ME}\frac{\partial W}{\partial E}\Big]$$
$$+\frac{\partial}{\partial M}\Big[C^{MM}\frac{\partial W}{\partial M}\Big], \tag{5}$$

where

$$T=2\pi\kappa(m/|2E|^3)^{1/2} \tag{6}$$

denotes the orbital period. The coefficients are

$$C^E=16\pi^3\tau m\kappa^2\big(E/M^2+\tfrac{3}{2}m\kappa^2/M^4\big), \tag{7a}$$
$$C^M=16\pi^3\tau m\kappa^2(1/M), \tag{7b}$$
$$C^{EE}=16\pi^3\tau\hbar M(|2E|^4/m\kappa^2)\phi_3(\epsilon), \tag{7c}$$
$$C^{MM}=16\pi^3\tau\hbar M|2E|\phi_1(\epsilon), \tag{7d}$$
$$C^{ME}=C^{EM}=16\pi^3\tau\hbar M(|2E|^{5/2}/(m\kappa^2)^{1/2})\phi_2(\epsilon), \tag{7e}$$

where the functions $\phi_r(\epsilon)$ may be expressed as series in terms of Bessel functions J_n (so-called Kapteyn series)[6–8] (a thorough study of these functions ϕ_r, including the derivation of their asymptotic behavior near $\eta=0$, may be found in Ref. 16):

$$\phi_r(\epsilon)=\frac{1}{2}\sum_{n=-\infty}^{\infty} sg(n)n^r\Big[\frac{\eta}{\epsilon}J_n(n\epsilon)+J'_n(n\epsilon)\Big]^2, \tag{8}$$

and sg (n) denotes the sign of n.

The domain of definition of Eq. (5) is

$$-\infty<E\leqslant 0,$$
$$0\leqslant M\leqslant[m\kappa^2/|E|]^{1/2}. \tag{5a}$$

It is worth noting that, in this so-called "current" form, Eq. (5), of the FPE [see Eqs. (9)–(11) below], the drift coefficients depend only on the damping Lorentz force and the diffusion coefficients only on the stochastic force (see Ref. 6 for example). This property holds for any stochastic force which depends on time only (but not on the phase space point),[8] or, more generally[17] for any stochastic force with zero divergence (with respect to phase space variables).

The problem is to solve the FPE in the stationary case. The stationary FPE can be written as

$$\operatorname{div}(W_0\overline{C}+\overline{\overline{C}}\,\mathbf{grad}\,W_0)=0, \tag{9}$$

where

$$\overline{C}=\begin{pmatrix}C^E\\C^M\end{pmatrix}\quad\text{and}\quad\overline{\overline{C}}=\begin{pmatrix}C^{EE} & C^{EM}\\C^{ME} & C^{MM}\end{pmatrix}, \tag{10}$$

and

$$\mathbf{J}=W_0\overline{C}+\overline{\overline{C}}\,\mathbf{grad}\,W_0 \tag{11}$$

is called the probability current.

From Eqs. (7a) and (7b) we easily find div $\overline{C}=0$ and then we always have the trivial constant solution $W_0=\text{const.}$[18] To the authors' knowledge no other exact solution has been obtained; however, Claverie and Pesquera[9,17] have been looking for solutions under the form $W=\text{Cst}\exp(-\phi)$, where ϕ is expressed as a series expansion. We remark that the constant solution does not satisfy the "zero flux" boundary conditions, i.e., $\mathbf{J}\cdot\mathbf{n}=0$ which seems natural for a good solution.[8]

The difficulty in finding an exact nontrivial solution (which could be compared with the experiment and with quantum theoretical results) is at the origin of the present work. In actual fact, as we are going to see below, we do not need to know any solution of the stationary FPE, in order to prove the nonrecurrence of the associated process (and all its consequences). In order to get some idea about the behavior of another solution (supposing that it exists), different from the trivial one, we can proceed as follows:

We modify slightly the coefficients given by Eq. (7) in such a way as to satisfy the so-called detailed balance condition (DBC), i.e., the vanishing of the probability current[8]

$$W_0\overline{C}+\overline{\overline{C}}\text{grad }W_0=0, \tag{12}$$

where $\overline{C}$ and $\overline{\overline{C}}$ are given by (10). Equation (12) gives $\overline{C}+\overline{\overline{C}}\,\mathbf{grad}$ Log $W_0=0$, hence provided $\overline{\overline{C}}^{-1}$ exists (which is the case, except on some boundaries of our domain), the DBC explicitly reads:[8]

$\overline{\overline{C}}^{-1}\ \overline{C}$ must be a gradient field (grad Log W_0).

One among the possible modifications is then[8]

$$\phi_1^{\text{mod}}(\epsilon)=\frac{1}{2}\frac{1}{\eta^2}\quad\text{and}\quad\phi_2^{\text{mod}}(\epsilon)=\frac{1}{2}\frac{1+\epsilon^2/2}{\eta^5} \tag{13}$$

whatever $\phi_3(\epsilon)$ may be. These modified functions coincide with $\phi_1(\epsilon)$ and $\phi_2(\epsilon)$ respectively up to order ϵ^3 near 0 and diverge as $\eta\to 0$ with the same power of η; thus the exact problem satisfies the DBC along the boundary $\epsilon=0$ and the exact DB solution of the modified problem could provide an approximation to some stationary solution of it.

The exact DB stationary solution of the modified problem is[8]

$$W_0^{\text{mod}}(M)=\text{const }e^{-2M/\hbar} \tag{14}$$

and, on the other hand, since the drift coefficients are the same as for the exact problem, we still have the solution $W_0^{\text{mod}}=\text{const}$.

Both solutions, the constant and the approximate one, unavoidably lead to a divergent integral for the *complete* probability density W_0M^2/η^2. The volume element in the "reduced" phase space is $M^2/\eta^2\,dM\,d\eta$. It is due to the reduction from the complete six-dimensional phase space $(\mathbf{q},\mathbf{p})$, with volume element $\mathbf{dq}\,\mathbf{dp}$, to the space (M,η) (see Refs. 6–8). In other words, W_0 corresponds to a electron diffusion toward infinity (self-ionization process), in clear contradiction with experiment and with quantum theory. (See the end of the previous section.)

III. KHAS'MINSKII'S CRITERION[15]

Let $X(t)$ be a diffusion process on a σ-compact complete metric space (E,ρ) with Markov transition $P(t,x,A)$, $t\geqslant 0$, $x\in A$ and $A\in\mathbb{B}$ where $\mathbb{B}$ is the σ-algebra of the measurable sets generated by the open sets in the space (E,ρ).[15]

The definition of a recurrent diffusion process is the following:

Definition: If there exists a compactum K such that for all points $x \in E$,

$$P_x \{ \text{there is a } t \text{ such that } x_t \in K \} = 1, \tag{15}$$

the process is called recurrent.

In other words: for every point $x \in E$, every trajectory starting from x will cross the compactum K with probability 1.

It is easy to show that the trajectory of a recurrent diffusion process is everywhere dense in E with probability 1, then it is always ergodic.

We recall the following important properties proved by Khas'minskii[15] for general multidimensional processes:

(1) If a process is recurrent, it has a *unique* invariant measure, which may be finite (its integral is finite) or infinite.

(2) If a process has a finite invariant measure, then it is recurrent (and consequently this invariant measure is unique).

Thus, if a process is nonrecurrent, it cannot have a finite invariant measure; only infinite invariant measure (s) is (are) possible. From the physical point of view, a bound state obviously corresponds to the existence of a finite invariant measure, i.e., to a recurrent process (in S.E.D, examples of this situation are provided by the harmonic oscillator and the rigid rotator[1,3,5]). Conversely, a nonrecurrent process, having no finite invariant measure, may describe only an "unbound state." The importance of the distinction between recurrent and nonrecurrent processes is therefore quite clear from the physical point of view. Now, Khas'minskii[15] provides several sufficient criteria ensuring either recurrence or nonrecurrence of a process, based upon the knowledge of the drift and diffusion coefficients of the process. We proceed to the description of the one among these criteria that we used in the present work.

From the theory of Markov process it is well known that in the case of diffusion process in a domain E of N-dimensional Euclidean space $\mathbb{R}^N$ the Dynkin infinitesimal operator has the form[19–21]

$$Lu = \sum_{i,j=1}^{N} a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^{N} b_i \frac{\partial u}{\partial x_i}. \tag{16}$$

for twice differentiable functions.

We now assume that the operator (16) is given in all Euclidean space $\mathbb{R}^N$, and, moreover, that the coefficients a_{ij} and b_i are sufficiently smooth (for example, that a continuous third derivative exists for all the coefficients), and $\sum_{i,j=1}^{N} a_{ij}\lambda_i\lambda_j > 0$ everywhere in $\mathbb{R}^N$ if $\sum_{i=1} \lambda_i^2 \geqslant 0$.

Under these assumptions, and defining

$$\underline{B}_i(x_i) = \inf b_i(x_1,\dots,x_N)/a_{ii}(x_1,x_2,\dots,x_N), \tag{17a}$$

$$\overline{B}_i(x_i) = \sup b_i(x_1,x_2,\dots,x_n)/a_{ii}(x_1,x_2,\dots,x_N), \tag{17b}$$

where the infimum (supremum) is taken over all the coordinates except the ith, it is possible to derive the Khas'minskii criterion (see Ref. 15, supplement, Theorem II, p. 194).

Khas'minskii's criterion: In order that a diffusion process be nonrecurrent, it is sufficient that for some i, $1 \leqslant i \leqslant N$, one of the following two inequalities be satisfied:

$$\int_{x_0}^{\infty} \exp\left\{ - \int_{x_0}^{x} \underline{B}_i(y)\, dy \right\} dx < +\infty, \tag{18a}$$

$$\int_{-\infty}^{x_0} \exp\left\{ \int_{x}^{x_0} \overline{B}_i(y)\, dy \right\} dx < +\infty, \tag{18b}$$

whatever x_0 may be.

This criterion (and others also given by Khas'minskii[15]) may be considered as generalizations to the multidimentional case of well-known criteria corresponding to the case of one-dimensional diffusion processes.[22] A more abstract treatment for multidimensional processes based upon the martingale approach, may be found in the recent book by Strook and Varadhan.[23]

IV. THE STANDARD FORM OF THE FPE

In order to apply the Khas'minskii criterion to the Kepler problem, we have to put its FPE under a form which satisfies the criterion assumptions; i.e., the equation must be defined in the whole corresponding Euclidean space ($\mathbb{R}^2$ in this case) and it must be in the forward form which corresponds to the Dynkin infinitesimal operator [Eq. (16)]. [The continuity condition on the third derivates of the coefficients is satisfied for ϵ in the interval $(0,1)$]. We perform this transformation in several steps.

(i) We change form the E-M representation to the E-ϵ representation, with the idea to have a domain of definition more suitable for the extension to $\mathbb{R}^2$.

(ii) By a change of variables, we write the current form FPE in all $\mathbb{R}^2$.

(iii) In this step the original equation, which in current form describes the incomplete probability density, is changed into a equation for the complete probability density. This is carried out by introducing the volume element.

(iv) Finally, the mixed equation obtained in (iii) is transformed into the forward form.

Steps (i) and (ii) require us to know the behavior of the FPE coefficients under the effect of a change of variables.

We have the FPE (in current form)

$$C \frac{\partial W}{\partial t} = \sum_{i=1}^{2} \left\{ \frac{\partial}{\partial \xi_i} \left[D_i W + \sum_{j=1}^{2} D_{ij} \frac{\partial W}{\partial \xi_j} \right] \right\}, \tag{19}$$

where W is the density in the full (nonreduced) phase-space. We make the change of variables

$$\xi_i \to \xi'_i \quad (i = 1,2), \tag{20}$$

and we want the new coefficients D'_i and D'_{ij}, and the new volume element C'.

The result has been given by Lax (Ref. 24, Sec. 3: Behavior of the diffusion coefficient under transformation) for the case where the FPE is written in terms of the *complete* probability density $P = CW$, and the corresponding result for an equation of the form (19) will be found in Ref. 25.

$$D'_i = \left|\det\left(\frac{D(\xi)}{D(\xi')}\right)\right| \sum_{k=1}^{2} \frac{\partial \xi'_i}{\partial \xi_k} D_k \quad (i = 1,2), \tag{21a}$$

$$D'_{ij} = \left|\det\left(\frac{D(\xi)}{D(\xi')}\right)\right| \sum_{k,l=1}^{2} \frac{\partial \xi'_i}{\partial \xi_k} \frac{\partial \xi'_j}{\partial \xi_l} D_{kl} \quad (i,j = 1,2), \tag{21b}$$

and

$$C' = \left|\det\left(\frac{D(\xi)}{D(\xi')}\right)\right|C, \tag{21c}$$

where $\det(D(\xi)/D(\xi'))$ is the Jacobian of the transformation (20).

We proceed now to step (i):

(i) The domain of definition (5a) of our current form FPE (5) is of a rather complicated form, so we have to change of representation in order to obtain a suitable one for the extension to $\mathbb{R}^2$. Among the possible new representations there is the "energy-eccentricity" one, whose domain is

$$\begin{cases} -\infty < E < 0, \\ 0 < \epsilon < 1. \end{cases} \tag{22}$$

The change of variables is [see Eq. (20)]

$$\begin{Bmatrix} E \\ M \end{Bmatrix} \to \begin{cases} E \\ \epsilon = \left[1 - \dfrac{2M^2}{m\kappa^2}|E|\right]^{1/2}, \end{cases} \tag{23}$$

which, when used in Eq. (21), gives (without primes)

$$D_E = 8\pi^3\tau(m\kappa^2)^{1/2}|2E|^{3/2}(\epsilon/\eta^5)(3-\eta^2), \tag{24a}$$

$$D_\epsilon = 24\pi^3\tau(m\kappa^2)^{1/2}|2E|^{1/2}\epsilon^2/\eta^3, \tag{24b}$$

$$D_{EE} = 16\pi^3\tau\hbar|2E|^3\epsilon\phi_3(\epsilon), \tag{24c}$$

$$D_{\epsilon\epsilon} = 16\pi^3\tau\hbar|2E|\frac{\eta^2}{\epsilon}[\eta^2\phi_3(\epsilon) - 2\eta\phi_2(\epsilon) + \phi_1(\epsilon)], \tag{24d}$$

$$D_{E\epsilon} = D_{\epsilon E}$$

$$= 16\pi^3\tau\hbar|2E|^2\eta[\eta\phi_3(\epsilon) - \phi_2(\epsilon)], \tag{24e}$$

and

$$C = 16\pi^3(m\kappa^2)^{3/2}\epsilon/|2E|^{5/2}. \tag{24f}$$

(ii) The rectangular domain of definition (22) of the current form FPE in the $(E$–$\epsilon)$ representation is easily extended to all $\mathbb{R}^2$, by a new change of variables. One, among the several possibilities, is

$$\begin{Bmatrix} E \\ \epsilon \end{Bmatrix} \to \begin{cases} E' = \ln|E| \\ \epsilon' = \operatorname{Arctanh}(2\epsilon^2 - 1) = \dfrac{1}{2}\ln\dfrac{\epsilon^2}{1-\epsilon^2}, \end{cases} \tag{25}$$

which gives [applying again expressions (21) for the new coefficients and the new volume element]

$$D_{E'} = -\epsilon\eta^2 D_E, \tag{26a}$$

$$D_{\epsilon'} = |E|D_\epsilon, \tag{26b}$$

$$D_{E'E'} = \epsilon\eta^2/|E|D_{EE}, \tag{26c}$$

$$D_{\epsilon'\epsilon'} = (|E|/\epsilon\eta^2)D_{\epsilon\epsilon}, \tag{26d}$$

$$D_{E'\epsilon'} = D_{\epsilon'E'} = -D_{E\epsilon}, \tag{26e}$$

$$C' = \epsilon\eta^2|E|C. \tag{26f}$$

We now have a current form FPE [Eq. (19)] defined in all $\mathbb{R}^2$ [variables E', ϵ' given by (25)], and its coefficients are given by (26). In the two following steps we shall put it in the standard form, namely,

$$\frac{\partial P}{\partial t} = \sum_{i=1}^{2}\frac{\partial}{\partial\xi_i}(G_iP) + \sum_{i,j=1}^{2}\frac{\partial^2}{\partial\xi_i\partial\xi_j}(G_{ij}P), \tag{27}$$

where

$$P = CW \tag{28}$$

is the complete probability density.

(iii) From (19) it is evident that we can write (C is time independent)

$$\frac{\partial P}{\partial t} = \sum_{i=1}^{2}\left\{\frac{\partial}{\partial\xi_i}\left[\frac{D_i}{C}P + \sum_{j=1}^{2}D_{ij}\frac{\partial}{\partial\xi_j}\left(\frac{P}{C}\right)\right]\right\}.$$

or

$$\frac{\partial P}{\partial t} = \sum_{i=1}^{2}\frac{\partial}{\partial\xi_i}\left\{\left[\frac{D_i}{C} + \sum_{j=1}^{2}D_{ij}\frac{\partial}{\partial\xi_j}\left(\frac{1}{C}\right)\right]P + \sum_{j=1}^{2}\frac{D_{ij}}{C}\frac{\partial P}{\partial\xi_j}\right\}. \tag{29}$$

Then defining

$$G_i' = \frac{D_i}{C} + \sum_{j=1}^{2}\frac{D_{ij}}{C^2}\frac{\partial C}{\partial\xi_j} \tag{30a}$$

and

$$G_{ij}' = \frac{D_{ij}}{C}, \tag{30b}$$

we write our FPE in a current form

$$\frac{\partial P}{\partial t} = \sum_{i=1}^{2}\frac{\partial}{\partial\xi_i}\left\{G_i'P + \sum_{j=1}^{2}G_{ij}'\frac{\partial P}{\partial\xi_i}\right\}. \tag{31}$$

Applying Eqs. (30) to the coefficients (26) it is long, but easy, to find the new coefficients G_i' and G_{ij}':

$$G'_{E'} = \frac{1}{|E|C}\left[-D_E + \frac{3}{2}\frac{D_{EE}}{|E|} + \frac{2(1-2\epsilon^2)}{\epsilon\eta^2}D_{E\epsilon}\right], \tag{32a}$$

$$G'_{\epsilon'} = \frac{1}{\epsilon\eta^2 C}\left[+D_\epsilon - \frac{3}{2}\frac{D_{\epsilon E}}{|E|} - \frac{2(1-2\epsilon^2)}{\epsilon\eta^2}D_{\epsilon\epsilon}\right], \tag{32b}$$

$$G'_{E'E'} = \frac{1}{|E|^2}\frac{D_{EE}}{C}, \tag{32c}$$

$$G'_{\epsilon'\epsilon'} = \frac{1}{\epsilon^2\eta^4}\frac{D_{\epsilon\epsilon}}{C}, \tag{32d}$$

$$G'_{E'\epsilon'} = G'_{\epsilon'E'} = -\frac{D_{E\epsilon}}{|E|\epsilon\eta^2 C}. \tag{32e}$$

(iv) Finally, by comparing the standard form (27) and the current form (31), we get the desired relations

$$G_i = G_i' - \sum_{j=1}^{2}\frac{\partial G_{ij}'}{\partial\xi_j} \quad (i = 1,2), \tag{33a}$$

and

$$G_{ij} = G_{ij}' \quad (i,j = 1,2). \tag{33b}$$

The coefficients of our forward FPE for the complete probability density are found by substitution of (32) in (33) and are explicitly

$$G_{E'} = -g(E)\frac{\epsilon^2+2}{\eta^5} - 2f(E)\left[2\frac{2-\epsilon}{\epsilon}\phi_3(\epsilon) + \frac{1}{\eta}\phi_2(\epsilon) + \frac{\eta^2}{\epsilon}\phi_3'(\epsilon) - \frac{\eta}{\epsilon}\phi_2(\epsilon)\right], \tag{34a}$$

$$G_{\epsilon'} = \frac{3}{2} g(E) \frac{1}{\eta^5} + \frac{f(E)}{\epsilon^3}\Big[2\frac{1+2\epsilon^2}{\epsilon}\phi_3(\epsilon) - 2\frac{2+\epsilon^2}{\epsilon\eta}\phi_2(\epsilon) + \phi_1(\epsilon) - \eta^2\phi_3'(\epsilon) + 2\eta\phi_2'(\epsilon) - \epsilon\phi_1'(\epsilon)\Big], \tag{34b}$$

$$G_{E'E'} = 4f(E)\phi_3(\epsilon)/\epsilon, \tag{34c}$$

$$G_{\epsilon'\epsilon'} = (f(E)/\epsilon^3)[\phi_3(\epsilon)/\epsilon - 2\phi_2(\epsilon)/\epsilon\eta + \phi_1(\epsilon)/\eta^2], \tag{34d}$$

$$G_{E'\epsilon'} = G_{\epsilon'E'} = \frac{2f(E)}{\epsilon^2}\Big[\phi_3(\epsilon) - \frac{\phi_2(\epsilon)}{\eta}\Big], \tag{34e}$$

where

$$g(E) = \frac{\tau}{m\kappa^2}|2E|^3, \tag{35a}$$

$$f(E) = (\tau\hbar/(m\kappa^2)^{3/2})|2E|^{7/2}. \tag{35b}$$

V. APPLICATION OF THE KHAS'MINSKII CRITERION TO THE KEPLER SYSTEM

Now our FPE [Eq. (27) with coefficients (34)] is in the required form for the application of the Khas'minskii criterion and satisfies all his assumptions. In order to apply this criterion we must find the value of one of the expressions (17) for one of the two variables E', ϵ'; the simplest choice is $B_{\epsilon'}(E')$. We show in this section that $B_{\epsilon'}(E') > 0$ and then (18a) is satisfied, implying that the process is non-recurrent. Let us define

$$B_{\epsilon'}(E') = -G_{\epsilon'}(E,\epsilon)/G_{E'E'}(E,\epsilon). \tag{36}$$

Using the expressions (35) we find

$$B_{\epsilon'}(E') = \frac{g(E)}{4f(E)}\frac{\epsilon(3-\eta^2)}{\eta^5\phi_3(\epsilon)} + 2 + \frac{1}{2\phi_3(\epsilon)}\frac{d}{d\epsilon}[\eta^2\phi_3(\epsilon) - \eta\phi_2(\epsilon)]; \tag{37}$$

by using (39a) for expressing $\phi_3(\epsilon)$ in the bracket [ ], this can be transformed to

$$B_{\epsilon'}(E') = 2 + \frac{g(E)}{4f(E)}\frac{\epsilon(3-\eta^2)}{\eta^5\phi_3(\epsilon)} + \frac{\epsilon}{4\eta}\Big(1+\frac{1}{\eta^2}\Big)\frac{\phi_2(\epsilon)}{\phi_3(\epsilon)} + \frac{\eta}{4}\frac{\phi_2'(\epsilon)}{\phi_3(\epsilon)} + \frac{\epsilon\eta}{4}\frac{\phi_2''(\epsilon)}{\phi_3(\epsilon)}. \tag{38}$$

Now, the functions ϕ_2 obey the following recurrence relations, derived by Marshall:[16]

$$\phi_{2r+1}(\epsilon) = \hat{L}\phi_{2r}(\epsilon), \tag{39a}$$

$$r = 0,1,2,\ldots,$$

$$\phi_{2r+2}(\epsilon) = (\hat{L} + \hat{M})\phi_{2r+1}(\epsilon), \tag{39b}$$

where

$$\hat{L}f = -\frac{1}{2\eta}\frac{d}{d\eta}\Big(\frac{\epsilon^2}{\eta}f\Big) = \frac{1}{2\epsilon}\frac{d}{d\epsilon}\Big(\frac{\epsilon^2}{\eta}f\Big) \tag{40}$$

and

$$\hat{M}f = -\frac{1}{2\eta^2}\int_\eta^1 \frac{\epsilon'^2}{\eta'^2}f\,d\eta' = -\frac{1}{2\eta^2}\int_0^\epsilon \frac{\epsilon'^3}{\eta'^3}f\,d\epsilon'. \tag{41}$$

Using these relations, we shall now prove the following:

Theorem: All the functions $\phi_r(\epsilon)$ with $r \geqslant 1$ are power series of ϵ all of whose coefficients are nonnegative (and consequently the same property holds for all their derivatives with respect to ϵ).

(i) First we have (see Ref. 16)

$$\phi_1(\epsilon) = \frac{1}{\pi}\int_0^\infty \frac{u\sin^2 u\,du}{(u^2 - \epsilon^2\sin^2 u)^{1/2}} = \frac{1}{\pi}\int_0^{+\infty}\frac{\sin^2 u\ du}{(1-\epsilon^2(\sin u/u)^2)^{1/2}}. \tag{42}$$

Since the series expansion of $[1 - \epsilon^2\sin^2 u/u^2]^{-1/2}$ has all its coefficients positive, we immediately see that $\phi_1(\epsilon)$ will appear as a series of (even) powers of ϵ of all whose coefficients are positive (they essentially involve the integrals $\int_0^{+\infty}(\sin u/u)^{2n}\sin^2 u\,du$).

(ii) Recurrence relation (39a): Let us prove that if $f(\epsilon)$ is a series with nonnegative coefficients, then $\hat{L}f$ has the same property. Now, $1/\eta = (1-\epsilon^2)^{-1/2}$ is a series with positive coefficients (in actual fact this is true for any negative power of η), and consequently the same property holds for $(\epsilon^2/\eta)f$. Moreover, since the first term has exponent 2 (or larger), the derivative $(d/d\epsilon)[(\epsilon^2/\eta)f]$ has its first term with exponent 1 (or larger), hence dividing by ϵ still gives us a genuine power series of ϵ (nonnegative powers), and all coefficients are nonnegative. Q. E. D.

(iii) Recurrence relation (39b): We now want to prove that if $f(\epsilon)$ is a series with nonnegative coefficients, then $(\hat{L}+\hat{M})f$ has the same property. The proof is slightly more involved than the previous one. From the definitions (40) and (41) of the operators $\hat{L}$ and $\hat{M}$, we get, for any function $f(\epsilon)$

$$\frac{d}{d\epsilon}[\eta^2(\hat{L}+\hat{M})f] = \frac{\epsilon\eta}{2}f'' + \frac{3}{2}\eta f', \tag{43a}$$

and by direct differentiation,

$$\frac{d}{d\epsilon}[\eta^2(\hat{L}+\hat{M})f] = \eta^2\frac{d}{d\epsilon}[(\hat{L}+\hat{M})f] - 2\epsilon(\hat{L}+\hat{M})f. \tag{43b}$$

Therefore, upon replacing the left hand side of (43b) by its expression (43a),

$$\frac{d}{d\epsilon}[(\hat{L}+\hat{M})f] = \frac{2\epsilon}{\eta^2}(\hat{L}+\hat{M})f + \frac{3}{2}\frac{1}{\eta}f' + \frac{\epsilon}{2\eta}f''. \tag{43c}$$

Now, introducing the notation $g(\epsilon) = (\hat{L}+\hat{M})f(\epsilon)$, Eq. (43c) appears as a differential equation for g,

$$dg/d\epsilon = A(\epsilon)g + B(\epsilon), \tag{44}$$

where $A(\epsilon) = 2\epsilon/\eta^2$ and $B(\epsilon) = \frac{3}{2}f'/\eta + \epsilon f''/\eta^2$.

Now, η^{-1} and η^{-2} are series with nonnegative coefficients. Thus, if we assume that the power series of $f(\epsilon)$ has this property, the property also holds for f', f'' $A(\epsilon)$, and $B(\epsilon)$. On the other hand, the first coefficient of the power series of $g(\epsilon)$, namely $g(0)$, is nonnegative

$$g(0) = (\hat{L} + \hat{M})f(\epsilon)|_{\epsilon = 0},$$
$$= \left[\left(\frac{1}{\eta} + \frac{\epsilon^2}{2\eta^3}\right)f + \frac{\epsilon}{2\eta}f'\right]_{\epsilon = 0} - \frac{1}{2\eta^2}\int_0^{\epsilon = 0} \frac{\epsilon'^3}{\eta'^3} f d\epsilon',$$
$$= f(0) \geqslant 0. \tag{45}$$

Then Eq. (44) gives us $g'(0) = A(0)g(0) + B(0) \geqslant 0$, similarly, by successive differentiation of (44) we prove that all the derivatives of $g(\epsilon)$ take nonnegative values for $\epsilon = 0$; differentiating (44) n times through Leibnitz rule gives us $g^{(n+1)}(\epsilon)$ as $B^{(n)}(\epsilon)$ plus a linear combination of $\{g(\epsilon), g'(\epsilon), ..., g^{(n)}(\epsilon)\}$ with all coefficients being nonnegative: Thus, if $g(0), ..., g^{(n)}(0) \geqslant 0$, then $g^{(n+1)}(0) \geqslant 0$, and since $g(0) \geqslant 0$, the property holds for every n. Thus, if $f(\epsilon)$ is a power series with nonnegative coefficients, the same property holds for $(\hat{L} + \hat{M})f$.

It may be recalled that this procedure (namely obtaining the coefficients of the Taylor expansion of the solution of a differential equation through successive differentiation of the equation) is just the one used in the original existence proof by Cauchy (so-called method of limits), which uses "dominant series" with nonnegative coefficients (see, e.g., Ince, Ref. 26).

(iv) We may now trivially conclude: If the property of being a series of powers of ϵ with nonnegative coefficients holds for some function $\phi_{r_0}(\epsilon)$, since this property is conserved by both recurrence relations (39a) and (39b), it is obvious that, through the suitable application of (39a) and (39b) alternatively, the same property will also hold for all $\phi_r(\epsilon)$ with $r > r_0$. But, according to point (i) above, the property holds for $r = 1$, hence also for $r > 1$, which completes the proof.

Let us now return to the Eq. (38) for $B_{\epsilon'}(E')$. It is clear that $B_{\epsilon'}(E') \geqslant 2$, because all terms beyond 2 in the right-hand side are nonnegative (for $0 \leqslant \epsilon \leqslant 1$). More precisely, these terms are strictly positive for $\epsilon > 0$, and vanish for $\epsilon = 0$. This vanishing is obvious for the 2nd, 3rd, and 5th terms since ϵ is explicitly factored out, and as concerns the 4th term we have $\phi_2'(0) = 0$ because all series $\phi_r(\epsilon)$ involve only even powers of ϵ [this may be directly seen from their definition Eq. (8)], and therefore $\phi_r'(0) = 0$ whatever r may be.

We therefore get

$$\underline{B}_\epsilon'(E') = \mathrm{Inf}_{\epsilon'} B_{\epsilon'}(E') = 2, \tag{46}$$

since the lower bound 2 is actually reached for $\epsilon = 0$.

Inserting (46) into the left-hand side of (18a), we get

$$\int_{x_0}^{+\infty} \exp\left\{-\int_{x_0}^{x} 2\, dy\right\} dx$$
$$= \int_{x_0}^{+\infty} \exp[-2(x - x_0)]\, dx$$
$$= \int_0^{+\infty} \exp(-2z)\, dz = 1/2. \tag{47}$$

Thus the criterion (18a) is satisfied, and we may therefore conclude that the *stochastic process associated with the Fokker–Planck equation* (5) (of the hydrogen atom problem) *is nonrecurrent.*

VI. REMARKS

We finish this paper with some remarks:

(i) As mentioned in Sec. III, we know from Khas'minskii[15] that if a stochastic process is nonrecurrent there does not exist any finite invariant measure. Therefore, if another stationary solution different from the trivial one (W = const) exists, it will also present the problem of self-ionization.

(ii) Also from Khas'minskii[15] we have: If the invariant measure is not unique the associated process is nonrecurrent. Thus the knowledge of a positive stationary solution different from the trivial one (the constant) would be enough for proving the nonrecurrence property. In the Kepler system the existence of stationary (positive) solutions, different from the constant, has not yet been proved rigorously; however, since we have proved the nonrecurrence property, the existence of other stationary (positive) solutions (nonfinite) seems quite likely. A possibility could be to prove the existence of such a solution determined by suitable boundary conditions different from those corresponding to the trivial constant solution; prescribed positive nonconstant values along the boundaries, or zero-flux condition along some boundary (along which the constant solution has nonzero flux), or any mixture of both types. In any case, the main problem would not be to prove that the solution exists, but rather that it is *nonnegative* throughout the whole domain.

The modified Kepler problem [Sec. II, Eq. (13)] is an example where the nonuniqueness is proved [there is the constant solution and the solution given by Eq. (14)], and this is enough for ensuring the nonrecurrence property. Another example is the Kepler system in the Rayleigh–Jeans field [$S_E(\omega)$ = const ω^2] with Lorentz–Dirac damping force, where the constant and the Boltzmann–Gibbs solutions are known and both are divergent.[18]

(iii) The present results could have led to the idea of modifying the spectral density (1) while keeping unchanged the seemingly well-established Lorentz–Dirac damping force. However, it is this last one which is essentially responsible.[18] As we already said, the drift coefficients [we consider here the coefficients of the equation in *current* form (5) or (19), not those of the equation in standard form (27)] depend only on the damping force [see Eqs. (5a)] and the existence of the trivial constant solution is a consequence of them, because div $\overline{C} = 0$. Then in all problems with this damping force two possibilities arise:

(a) The solution is unique and is therefore the constant one, and it is nonfinite.

(b) The solution is nonunique; then the process is nonrecurrent, and only nonfinite invariant measures may exist. As we have seen in the preceding remark, possibility (b) is the most likely, but in any case the result is unsatisfactory since we do *not* have any *finite* invariant measure.

It is a rather remarkable and unexpected result that, for the Coulomb potential, if we assume the Lorentz–Dirac damping law and a position and velocity-independent spectrum for the stochastic force (or, even more generally[17] if we simply assume a zero divergence for this force), then no spectral density allows us to get a "reasonable" stationary state (stationary distribution with a finite integral).

We conclude that the problem of finding a better agreement between classical stochastic theories, such as SED, and quantum theory (e.g., finding some reasonable stationary state for the Kepler system) is deeper than previously thought. This paper suggest that the change of the spectral density only is not enough and more drastic modifications (e.g., introducing a position and/or velocity-dependent stochastic force, or a modified damping force, or both) would be necessary in the present version of S. E. D. (as described in Refs. 5–8).

[1]T. W. Marshall, Proc. Roy. Soc. London A **276**, 475 (1963).
[2]T. H. Boyer, Phys. Rev. **182**, 1374 (1969).
[3]E. Santos, Nuovo Cimento B **19**, 57 (1974).
[4]T. H. Boyer, Phys. Rev. D **11**, 790 (1975).
[5]P. Claverie and S. Diner, Intern. J. Quantum Chem. **12**, Suppl. 1, 41 (1977).
[6]T. W. Marshall and P. Claverie, J. Math. Phys. **21**, 1819 (1980).
[7]P. Claverie, L. De La Peña-Auerbach, and S. Diner, "Stochastic electrodynamics of non-linear systems II. Derivation of a reduced Fokker–Planck equation in terms of relevant constants of motion" (to be published).
[8]P. Claverie, in *Dynamical Systems and Microphysics*, edited by A. Blaquière, F. Fer, and A. Marzollo, CISM Courses and Lectures **261** (Springer, Wien, 1980), p. 111; P. Claverie, "Stochastic Electrodynamics versus Quantum Theory: recent advances in the study non-linear systems," contribution to the *Einstein Centennial Symposium on Fundamental Physics*, (Bogota, 1979) (in press).
[9]P. Claverie and L. Pesquera "Derivation of Fokker–Planck equations through response theory" (submitted to Z. Phys. B)
[10]L. De la Peña-Auerbach and A. M. Cetto, Rev. Mex. Fis. **25**, 1 (1976).
[11]L. De la Peña-Auerbach and A. M. Cetto, J. Math. Phys. **18**, 1612 (1977).
[12]L. De la Peña-Auerbach and A. M. Cetto, Ann. Fond. L. de Broglie **3**, 15 (1978).
[13]L. De la Peña-Auerbach and A. M. Cetto, Found. Phys. **8**, 191 (1978).
[14]L. Pesquera, in *Dynamical Systems and Microphysics*, edited by A. Blaquière, F. Fer, and A. Marzollo, CISM Courses and Lectures **261** (Springer, Wien, 1980), p. 145.
[15]R. Z. Khas'minskii, Theor. Prob. Appl. **5**, 179 (1960).
[16]T. W. Marshall, Z. A. M. P. (J. Appl. Math. Phys.) **30**, 1011 (1979).
[17]L. Pesquera, "Etude des Equations Différentielles stochastiques non Markoviennes et Application à l'Electrodynamique Stochastique," Thèse (Doctorat de 3e cycle), Paris (1980), or "Ecuaciones diferenciales estocásticas no-markovianas. Aplicación a la Electrodinámica Estocástica (E. D. E.)," Tesis, Santander (1980) (Chap. II, Sec. 1b).
[18]P. Claverie, L. Pesquera, and F. Soto, Phys. Lett. A. **80**, 113 (1980).
[19]E. B. Dynkin, *Theory of Markov Processes* (Prentice Hall, Englewood Cliffs, N. J., 1961).
[20]E. B. Dynkin, *Markov Processes* (Springer, Berlin, 1965), Vol. 1.
[21]J. Lamperti, *Stochastic Processes* (Springer, New York, 1977).
[22]V. Prohorov and Y. A. Rozanov, *Probability Theory: Basic Concepts, Limit Theorems, Random Processes* (Springer, Berlin, 1969).
[23]D. W. Stroock and S. R. S. Varadhan, *Multidimensional Diffusion Processes* (Springer, Berlin, 1979).
[24]M. Lax, Rev. Mod. Phys. **38**, 541 (1966).
[25]P. Claverie "Some remarks concerning the change of variables in Fokker Planck equations" (to be published).
[26]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956), Chap. XII.

Pseudoscalar interaction of coupled quantum-mechanical oscillators with independent Fermi systems [a)]

Guy A. Battle III
Mathematics Department, Texas A&M University, College Station, Texas 77843

(Received 27 January 1981; accepted for publication 27 March 1981)

Using the techniques of constructive quantum field theory we analyze the dynamics of a cubic lattice of quantum mechanical oscillators with nearest-neighbor coupling that interact with a corresponding lattice of finite-volume truncations of independent relativistic Fermi fields. Since the model is nonrelativistic, we rely on a nonrelativistic version of the Osterwalder–Schrader (OS) reconstruction theorem. Also, the absence of Nelson's symmetry in the (un)Euclidean picture is not serious because the transfer matrix in a given space direction is simple enough to make the verification of spatial OS positivity easy. After establishing for our model many of the basic results that hold for the more standard models, we give a proof of the Fortuin–Kasteleyn–Ginibre (FKG) inequality that is essentially independent of the dimension of the Fermi systems.

PACS numbers: 03.65. − w, 03.70. + k, 02.50.Sk, 02.30.Ap

1. INTRODUCTION

Recently the Fortuin–Kasteleyn–Ginibre (FKG) inequality was shown to hold for the scalar Yukawa_2 quantum field theory.[1] Whether the inequality holds for the pseudoscalar model is still an open question, however, and the method of Ref. 1 does not seem to suggest specifically how this question might be resolved (although the general strategy appears to be the best approach). Nevertheless, a special calculation was made in Ref. 1 which tended to suggest that the FKG inequality actually held for the pseudoscalar Yukawa_n model for arbitrary space-time dimension n.

In this paper we present further evidence by studying a nonrelativistic modification of the pseudoscalar Yukawa model in which the role of the fermions is reduced but not eliminated. More specifically, we consider a lattice of quantum-mechanical oscillators with nearest-neighbor coupling interacting with a lattice of Fermi systems *which do not exchange fermions*. We may take the lattice dimension to be arbitrary; as we will see, the singular behavior of the model is determined by the dimension of the Fermi system at each lattice site. We consider only one-dimensional Fermi systems (i.e., fermions that live in two space-time dimensions), but our proof of the FKG inequality for such a model is effectively independent of dimension.

Since our model is nonrelativistic, however, we must investigate to what extent the techniques and results of constructive quantum field theory are affected by such a modification, and much of the paper is devoted to the establishment of basic results for the model. Before describing our model in greater detail, we pause to make a few remarks about the axiomatics of nonrelativistic field theories.

Let $\{\mathscr{W}_n(x_1,\ldots,x_n)\}_{n=0}^{\infty}$ be scalar distributions on products of, say, $\mathbf{R}^\nu\times\mathbf{R}$. If these distributions are the vacuum expectations for the dynamics of a nonrelativistic field in ν space dimensions, then the strongest set of axioms we can generally expect the distributions to satisfy is the following:

Non RW1 (Temperedness): $\mathscr{W}_n$ is tempered; $\mathscr{W}_0=1$, and $\mathscr{W}_n(x_1,\ldots,x_n)=\mathscr{W}_n(x_n,\ldots,x_1)$.

Non RW2 (Translation Invariance in Space and Time): For $c\in\mathbf{R}^\nu\times\mathbf{R}$, $\mathscr{W}_n(x_1-c,\ldots,x_n-c)=\mathscr{W}_n(x_1,\ldots,x_n)$.

Non RW3 (Positive Definiteness): For $f_{jk}\in\mathscr{S}(\mathbf{R}^\nu\times\mathbf{R})$, $1\leqslant j\leqslant l$, $1\leqslant k\leqslant n_j$,

$$\sum_{j,j'=1}^{l}\int \mathscr{W}_{n_j+n_{j'}}(x_1,\ldots,x_{n_j},y_{n_{j'}},\ldots,y_1)\prod_{k=1}^{n_j}f_{jk}(x_k)$$
$$\times\prod_{k=1}^{n_{j'}}\overline{f_{j'k}(y_k)}\,d^{n_j}x\,d^{n_{j'}}y\geqslant 0.$$

Non RW4 (Non-Relativistic Spectrum Condition): For $n>0$, set $W_n(\xi_1,\ldots,\xi_{n-1})=\mathscr{W}_n\left(0,\xi_1,\xi_1+\xi_2,\ldots,\sum_{k=1}^{n-1}\xi_k\right)$. Then $\operatorname{supp}\widehat{W}_n\subset(\mathbf{R}^\nu\times\mathbf{R}^+)^{n-1}$.

Non RW5 (Time Cluster Property): Let γ^λ: $\{1,\ldots,m+n\}\rightarrow\{x_1,\ldots,x_m,y_1-(0,\lambda),\ldots,y_n-(0,\lambda)\}$ be a bijection. Then

$$\lim_{\lambda\to\pm\infty}\mathscr{W}_{m+n}(\gamma^\lambda(1),\ldots,\gamma^\lambda(m+n))$$
$$=\mathscr{W}_m(x_1,\ldots,x_m)\mathscr{W}_n(y_1,\ldots,y_n)$$

in the sense of distributions.

Non RW6 (Space Cluster Property): For nonzero $a\in\mathbf{R}^\nu$ let
γ^λ: $\{1,\ldots,m+n\}\rightarrow\{x_1,\ldots,x_m,y_1-(\lambda a,0),\ldots,y_n-(\lambda a,0)\}$ be a bijection. Then

$$\lim_{\lambda\to\infty}\mathscr{W}_{m+n}(\gamma^\lambda(1),\ldots,\gamma^\lambda(m+n))$$
$$=\mathscr{W}_m(x_1,\ldots,x_m)\mathscr{W}_n(y_1,\ldots,y_n)$$

in the sense of distributions.

For convenience we call these axioms the *nonrelativistic Wightman axioms*; they comprise the obvious nonrelativistic analog of the Wightman set of axioms.[2,3,4] We cannot expect locality to hold (i.e., symmetry of $\mathscr{W}_n$ with respect to arguments whose differences lie outside the light cone) because the propagation of an effect due to a nonrelativistic interaction is typically instantaneous. The space cluster property is the obvious analog of the relativistic cluster prop-

a)Research supported by NSF Grant MCS 8002499.

erty, but the time cluster property must now be stated explicitly and separately because it is now essential for establishing uniqueness of the vacuum when one applies the proof of the Wightman reconstruction theorem to construct a nonrelativistic Garding–Wightman theory. The point is that the absence of relativistic invariance severs the connection between the Hamiltonian and momentum operators, so the two cluster properties are in general independent.

The price we pay for this nonrelativistic modification of the Wightman axioms lies in the weakness of the spectrum condition. Since the distributions $\widehat{W}_n$ are required only to be supported in products of forward half-space-times rather than in products of forward light cones, the general theory of Laplace transforms[5] can no longer provide us with analytic continuations of the $\mathscr{W}_n$ to extended forward tubes.[2–5] In particular, we cannot prove a PCT theorem[2,4,5] for a nonrelativistic theory. What our spectrum condition does allow, however, is an analytic continuation of the $\mathscr{W}_n$ to products of complex half-planes with respect to successive differences in the time arguments, *where the space arguments are held fixed*. (Strictly speaking, we smear with respect to the space arguments and hold the test functions fixed.) This simple-minded analytic continuation certainly captures the Euclidean points with strictly ordered imaginary times so the nonrelativistic Wightman theory has Schwinger functions (Euclidean Green's functions[4,6,7]).

Before discussing what properties the Schwinger functions of a nonrelativistic theory should have in order to determine the theory, we proceed with a detailed description of the model we wish to study.

Consider a ν-dimensional cubic lattice indexed by $\mathbf{Z}^\nu$ and let $\Lambda \subset \mathbf{Z}^\nu$ have the form

$$\Lambda = \{ m \in \mathbf{Z}^\nu | -L_j < m_j \leqslant L_j \},$$

where $L_1,...,L_\nu \in \mathbf{Z}^+$. At each lattice site $m \in \Lambda$ we introduce a one-dimensional quantum-mechanical system; i.e., we let $\mathscr{H}_m$ be the mth copy of $L^2(\mathbf{R})$ and denote the usual momentum and position operators on this mth copy by P_m and Q_m, respectively. The Hilbert space for this system is $\otimes_{m\in\Lambda}\mathscr{H}_m$ and we wish to consider the Hamiltonian

$$H^\Lambda = \sum_{m\in\Lambda} (P_m^2 + \omega_0^2 Q_m^2) + \sum_{\langle m,m'\rangle \subset \Lambda} (Q_m - Q_{m'})^2, \tag{1.1}$$

where "$\langle m,m'\rangle \subset \Lambda$" means "$m,m' \in \Lambda$ and are nearest neighbors in the periodic sense." This interaction describes an array of harmonic oscillators with fundamental frequency ω_0 and nearest-neighbor coupling, and H_Λ is self-adjoint, bounded below, and has only discrete spectrum. As usual, the eigenvalues and eigenfunctions are explicitly computed by decomposing H_Λ into normal modes via the Fourier transform. One defines

$$a_\Lambda(k) = \omega(k)^{1/2} \sum_{m\in\Lambda} e^{im\cdot k} Q_m + \frac{i}{\omega(k)^{1/2}} \sum_{m\in\Lambda} e^{im\cdot k} P_m, k \in \Lambda^*, \tag{1.2}$$

where

$$\Lambda^* = \left\{ \left(\frac{\pi l_1}{L_1}, ..., \frac{\pi l_\nu}{L_\nu} \right) \middle| \quad -L_j < l_j \leqslant L_j, l_j \in \mathbf{Z} \right\},$$

$$\omega(k)^2 \equiv \omega_0^2 + \sum_{j=1}^{\nu} (1 - \cos k_j). \tag{1.3}$$

Hence

$$a_\Lambda(k)^\dagger = \omega(k)^{1/2} \sum_{m\in\Lambda} e^{-im\cdot k} Q_m - \frac{i}{\omega(k)^{1/2}} \sum_{m\in\Lambda} e^{-im\cdot k} P_m, \tag{1.4}$$

where $\dagger$ denotes operator adjoint; moreover,

$$Q_m = \frac{1}{|\Lambda|} \sum_{k\in\Lambda^*} e^{-im\cdot k} \frac{1}{\omega(k)^{1/2}} [a_\Lambda(k) + a_\Lambda(-k)^\dagger], \tag{1.5}$$

$$P_m = \frac{-i}{|\Lambda|} \sum_{k\in\Lambda^*} e^{-im\cdot k} \omega(k)^{1/2} [a_\Lambda(k) - a_\Lambda(-k)^\dagger], \tag{1.6}$$

$$H^\Lambda = \frac{1}{|\Lambda|} \sum_{k\in\Lambda^*} \omega(k) a_\Lambda(k)^\dagger a_\Lambda(k) + \sum_{k\in\Lambda^*} \omega(k). \tag{1.7}$$

Let Ω^Λ be the ground state and consider the corresponding expectations

$$\left(\prod_{\ell=1}^{n} (e^{it_\ell H^\Lambda} Q_{m_\ell} e^{-it_\ell H^\Lambda}) \Omega^\Lambda, \Omega^\Lambda \right), m_\ell \in \Lambda,$$

for our dynamical system. The limits of such expectations as $\Lambda \to \infty$ exist and can be computed explicitly, and it is well known that these vacuum expectation values for the resulting infinite system of coupled quantum-mechanical oscillators yield the discrete version of the free scalar quantum field theory[2–5] in the sense of Wightman reconstruction.[2] The (reconstructed) Hilbert space can be realized as the symmetric Fock space $\mathscr{F}_B$ over $l^2(\mathbf{Z}^\nu)$ or some Sobelev variation thereof, and the annihilation (creation) operator $a(k)(a(k)^\dagger)$ is defined as an operator-valued distribution over T^ν in periodic analogy to the scalar field case.[5] The (reconstructed) position operators Q_m form the "time-zero field" and they are related to the annihilation and creation operators via the formula

$$Q_m = \frac{1}{(2\pi)^\nu} \int_{T^\nu} d^\nu k \frac{e^{-im\cdot k}}{\omega(k)^{1/2}} [a(k) + a(-k)^\dagger]. \tag{1.8}$$

Similarly the (reconstructed) momentum operators P_m form the "time-zero conjugate field," and we have

$$P_m = \frac{-i}{(2\pi)^\nu} \int_{T^\nu} d^\nu k \, e^{-im\cdot k} \omega(k)^{1/2} [a(k) - a(-k)^\dagger]. \tag{1.9}$$

Our Hamiltonian is given by

$$H = \frac{1}{(2\pi)^\nu} \int_{T^\nu} d^\nu k \, \omega(k) a(k)^\dagger a(k). \tag{1.10}$$

[The constant term in (1.7) diverges as $\Lambda \to \infty$, but it is cancelled in the vacuum expectation expressions.] We also have the formula for H in configuration space

$$H = \sum_{m\in\mathbf{Z}^\nu} (:P_m^2: + \omega_0^2 :Q_m^2:) + \sum_{\langle m,m'\rangle \subset \Lambda} :(Q_m - Q'_m)^2: \tag{1.11}$$

where : : denotes the usual Wick ordering.[3,5] The expressions (1.10) and (1.11) are understood in the sense of quadratic forms[5,8] on the form domain for the self-adjoint operator H.

Following the customary choice of one-particle space in the scalar field case, we choose the Hilbert space

$$\left\{ f \in \mathbf{R}^{\mathbf{Z}^\nu} \,\middle|\, \|f\|^2 \equiv \int_{T^\nu} d^\nu k \frac{|\hat{f}(k)|^2}{\omega(k)} < \infty \right\} \tag{1.12}$$

as our "one-particle" space; the symmetric Fock space over this space will serve as our time-zero Hilbert space. As usual, it should be noted that as a consequence of this choice the *functorial* annihilation (creation) operator[3,5] is not $a(k)(a(k)^\dagger)$ but $\omega(k)^{-1/2}a(k)(\omega(k)^{-1/2}a(k)^\dagger)$. Finally, since the ground state representation of our infinite quantum-mechanical system is the "free boson" part of our model (where the fundamental frequency ω_0 is the "boson mass"), we will denote H by H_B in the sequel. Similarly, we denote the vacuum state by Ω_B.

The free Fermi part of our model is described as follows: at each lattice site we place an interval, say $[-\pi,\pi]$, in which the periodic truncation of the time-zero relativistic Fermi fields (in two space-time dimensions) with Fermi mass M is introduced. Moreover, we stipulate that these Fermi systems do not exchange fermions, so the Fermi fields at different lattice sites are independent. Thus we denote the Fermi fields at lattice site m by $\psi_\alpha^{(m)}(x)$, $\overline{\psi_\alpha^{(m)}}(x)$ for $\alpha = 0,1$ and $-\pi \leqslant x \leqslant \pi$, where, as usual, the latter fields are the adjoint spinor fields. We let $\mathscr{F}_F^{(m)}$ denote the antisymmetric Fock space over an mth copy of the periodic version of the usual one-particle Hilbert space for fermions in one space dimension,[9,10] and we denote the free fermion Hamiltonian by $H_F^{(m)}$. The smearing of $\overline{\psi_\alpha^{(m)}}, \psi_\alpha^{(m)}$ against appropriate test functions yields bounded operators on $\mathscr{F}_F^{(m)}$, and the operator $H_F^{(m)}$ is bounded below. Let $\Omega_F^{(m)}$ denote the vacuum state.

Since the analysis of our interacting system will be based entirely on the Matthews–Salam–Seiler strategy,[11–14] in which the fermions are "integrated out," we omit the explicit description of the Fermi fields and Fermi Hamiltonian and merely state the typical formulas we will need. First, for $s_1,\dots,s_n,t_1,\dots,t_n$ distinct we have the explicit formula

$$\begin{aligned}&\Big(T\Big[\prod_{\ell=1}^n \big(e^{-t_\ell H_F^{(m)}}\overline{\psi_{\alpha_\ell}^{(m)}}(x_\ell)e^{t_\ell H_F^{(m)}}\big)\prod_{\ell=1}^n \big(e^{-s_\ell H_F^{(m)}} \\ &\quad\times \psi_{\beta_\ell}^{(m)}(y_\ell)e^{s_\ell H_F^{(m)}}\big)\Big]\Omega_F^{(m)},\Omega_F^{(m)}\Big)_{\mathscr{F}_F^{(m)}} \\ &= \pm \det_{1\leqslant j,k\leqslant n} S((x_j,s_j),(y_k,t_k))\alpha_j,\beta_k,\end{aligned} \tag{1.13}$$

where $T[\ ]$ denotes the ordering of the product based on the one-to-one correspondence $\gamma:\{1,\dots,2n\}\to\{s_1,\dots,s_n,t_1,\dots,t_n\}$ defined by $\gamma(k)<\gamma(k+1)$ (i.e., time ordering),

$$S((x,s),(y,t)) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} e^{in(x-y)}\int_{-\infty}^{\infty} d\omega\, e^{i\omega(s-t)}\frac{i\beta_0 + i\beta_1 n + M}{\omega^2+n^2+M^2}, \tag{1.14}$$

$$\beta_0 = \begin{pmatrix} 0 & 1 \\ 1 & 0\end{pmatrix},\quad \beta_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1\end{pmatrix}, \tag{1.15}$$

and $\pm$ on the rhs of (1.13) depends on the number of transpositions required to change the time ordering of operators on the lhs to an ordering where the ψ's alternate with the $\overline{\psi}$'s ($+$ if the number is even and $-$ if the number is odd). S is the Euclidean Fermi propagator[11] with periodic boundary conditions in one direction and free boundary conditions in the other. As always, the structure of the matrix on the rhs of (1.13) ensures that the determinant is independent of the ordering of $\{1,\dots,n\}$. (1.13) can be derived from the original description of the free Fermi field theory in periodic analogy to the derivation[15] for the free infinite volume theory. Following rules for evaluating free expectations involving Wick-ordered fermions,[15,16] one also obtains formulas for currents. For example,

$$\begin{aligned}&\Big(T\Big[\prod_{\ell=1}^n\Big(e^{-s_\ell H_F^{(m)}}\sum_{\alpha,\alpha'=0}^{1} :\overline{\psi_\alpha^{(m)}}(x_\ell)\Gamma_{\alpha\alpha'}\psi_{\alpha'}^{(m)}(x_\ell): e^{s_\ell H_F^{(m)}}\Big)\Big]\Omega_F^{(m)},\Omega_F^{(m)}\Big)_{\mathscr{F}_F^{(m)}} \\ &= \sum_{\alpha_1,\dots,\alpha_n=0}^{1} \det_{1\leqslant j,k\leqslant n}[S((x_j,s_j),(x_k,s_k))\Gamma]_{\alpha_j\alpha_k}(1-\delta_{jk}),\end{aligned} \tag{1.16}$$

where

$$\Gamma = \begin{pmatrix} 0 & -1 \\ 1 & 0\end{pmatrix}. \tag{1.17}$$

We have singled out this particular matrix because it will be involved in our interacting theory.

If only a finite set Λ of lattice sites are occupied by these Fermi systems, the Hilbert space for the composite free theory is

$$\mathscr{H}_\Lambda \equiv \mathscr{F}_B \otimes \bigotimes_{m\in\Lambda} \mathscr{F}_F^{(m)}; \tag{1.18}$$

the Hamiltonian is $H_B + \Sigma_{m\in\Lambda} H_F^{(m)}$, and the vacuum is

$$\Omega_\Lambda \equiv \Omega_B \otimes \bigotimes_{m\in\Lambda} \Omega_F^{(m)}. \tag{1.19}$$

We propose to perturb the ground state of our system with an interaction that is "local." More specifically, we wish to construct and study the vacuum expectation values of the theory whose (unrenormalized) Hamiltonian for a finite volume Λ is given by

$$H_B + \sum_{m\in\Lambda} H_F^{(m)} + \lambda\sum_{m\in\Lambda} Q_m \int_{-\pi}^{\pi} dx \sum_{\alpha,\alpha'=0}^{1} :\overline{\psi_\alpha^{(m)}}(x)\Gamma_{\alpha\alpha'}\psi_{\alpha'}^{(m)}(x):,$$

where λ is the coupling constant. Thus, each quantum-mechanical oscillator interacts with its own Fermi microcosm, where the additional force acting on the oscillator is proportional to a relativistic current density averaged over the interval. The Fermi systems at different lattice sites interact only through the quantum-mechanical couplings (i.e., "bosonically"). As in the case of the Yukawa$_2$ quantum field theory,[10] the Hamiltonian has ultraviolet divergences that must be cancelled by an infinite energy counterterm and an infinite "boson mass" counterterm (i.e., an infinite shift in the spring constant ω_0^2). Thus, our Hamiltonian is given by

$$H_\Lambda \equiv H_B + \sum_{m\in\Lambda} H_F^{(m)} + \lambda \sum_{m\in\Lambda} Q_m \int_{-\pi}^{\pi} dx \sum_{\alpha,\alpha'=0}^{1} :\overline{\psi_\alpha^{(m)}}(x) \times \Gamma_{\alpha\alpha'} \psi_{\alpha'}^{(m)}(x): - \delta\omega^2(\lambda) \sum_{m\in\Lambda} :Q_m^2: - \delta E(\lambda),$$

where $\delta\omega^2(\lambda)$ and $\delta E(\lambda)$ will be determined when we rigorously define H_Λ by the removal of a standard ultraviolet cutoff in Sec. 3.

It should not be surprising that the renormalization problems are so similar to those of the Yukawa$_2$ (Y_2) model. The ultraviolet singularities are independent of the lattice dimension ν, and in the case $\nu = 1$ it is not hard to see that our model is mathematically equivalent to a nonrelativistic truncation of the Y_2 model that preserves the singular structure of the Fermi fields. (This fact will be even more transparent in the Matthews–Salam–Seiler formulation.) More generally, if the periodic Fermi system at each lattice site is chosen to be n-dimensional, we expect the renormalization problems to be as difficult for our model as they are for the Y_{n+1} model.

We begin our analysis in Sec. 2 by deriving a Matthews–Salam–Seiler formula with an ultraviolet cutoff imposed on the time-zero Fermi fields. We follow the "semi-Euclidean" approach of Seiler and Simon[14] (i.e., we avoid the use of Osterwalder–Schrader fields[15]). This strategy is based on the Phillips perturbation expansion for semigroups together with an application of free expectation formulas to each term. The free Euclidean measure in our case is the Gaussian measure $d\mu$ on $\mathscr{S}'(\mathbf{Z}^\nu \times \mathbf{R})$ with mean zero and covariance

$$\int \phi(m,s)\phi(m',s')\, d\mu(\phi) = \frac{1}{(2\pi)^{\nu+1}} \int_{-\infty}^{\infty} d\omega\, e^{i\omega(s-s')} \int_{T^\nu} d^\nu k \frac{e^{ik\cdot(m-m')}}{\omega^2 + \omega(k)^2}. \tag{1.20}$$

As in the case of the standard free boson measure,[3,17,18] we have checkerboard and hypercontractive inequalities available to us via the Markov property and Nelson's fundamental hypercontractive result.[19]

Although our derivation of a Matthews–Salam–Seiler formula is basically a mimicry of the Seiler–Simon derivation for the Y_2 model, we illustrate the proof for two reasons. First, there is an error in the Seiler–Simon paper which was pointed out by Lon Rosen and Barry Simon (private communication), and so we must supplement the proof with an argument due to Rosen. Second, the term-by-term identification of the Phillips perturbation expansion with the expansion for the corresponding Matthews–Salam–Seiler expression is complicated by the fact that different lattice sites exchange "bosons" but do not exchange fermions.

With regard to Secs. 3 and 4 there is a general point that must be stressed. Since our model is nonrelativistic, the analytic continuation of the model to imaginary time is unEuclidean. (We avoid the word "non-Euclidean" for obvious reasons). In particular, Nelson's symmetry does not hold. Since several of the basic results in constructive quantum field theory seem to depend on Nelson's symmetry, this appears to be a disaster until one realizes the implications of Ref. 14. It is well known but rarely emphasized[20] that the main function of Nelson's symmetry is to identify the transfer matrix for a spatial direction. As Seiler and Simon point out,[14] the *crucial* property is OS positivity in the spatial directions (i.e., self-adjointness of the transfer matrices). Thus the power of Nelson's symmetry is no more mysterious than that of any other symmety that eliminates the need for a possibly difficult calculation. In our case Nelson's symmetry is not needed because the transfer matrices for the spatial directions are very simple; their off-diagonal elements are merely those that arise from a nearest-neighbor ferromagnetic coupling.[18]

It is the absence of a Reeh–Schlieder theorem[4,14] that creates problem for us. The free vacuum Ω_Λ is certainly cyclic in $\mathscr{H}_\Lambda$ with respect to products of free propagations of Q_m and smeared $\psi_\alpha^{(m')}(x)$, $\overline{\psi_\alpha^{(m')}}(x)$, where $m\in\mathbf{Z}, m'\in\Lambda, -\pi \leqslant x \leqslant \pi, \alpha = 0,1$; by analytic continuation to imaginary time one can obtain a cyclicity statement for semigroup propagations as well. However, since the Fermi systems are independent Ω_Λ is clearly not cyclic with respect to products of free propagations supported in an arbitrary fixed rectangle in our space-time $\mathbf{Z}^\nu \times \mathbf{R}$. In the Seiler–Simon analysis of the Y_2 model such a property of Y_2 (known as a Reeh-Schlieder theorem) seems to be involved in three arguments:

(a) Proof of semiboundedness of H_Λ,

(b) Proof of vacuum overlap,

(c) Refinement of Fröhlich bounds to temperedness bounds for the Schwinger functions for *arbitrary* coupling strength λ. (In the case of weak coupling the temperedness bounds are derivable from the cluster expansion.[21,22])

In Sec. 3 we accomplish (a) and (b) for our model without using a Reeh–Schlieder theorem or even OS positivity in the spatial directions. Our proof is based on the Matthews–Salam–Seiler formula together with the fact that the interaction is pseudoscalar (i.e., that Γ is anti-Hermitian). We use modified Jost states involving truncations of Q_m, so we cannot extend our method to proving temperedness bounds; (c) is beyond our reach at the present time. The consequences of Sec. 3 are a Euclidean Gell-Mann–Low formula and a Matthews–Salam–Seiler formula without ultraviolet cutoffs. The interacting measure is

$$\left[\prod_{m\in\Lambda} \det{}_{\text{ren}}(1 - \lambda K_m)\right] d\mu, \tag{1.21}$$

where $\det_{\text{ren}}$ is a suitably renormalized determinant and K_m is an operator-valued random variable that is $\Sigma_{\{m\}\times[-t/2,t/2]}$-measurable (i.e., K_m depends only on $\chi\phi(m,\cdot)$, where χ is the characteristic function of the interval $[-t/2,t/2]$). $\Lambda \times [-t/2,t/2]$ is, of course, the given Euclidean finite-volume cutoff, and the Gell-Mann–Low formula involves the $t = \infty$ limit with Λ fixed. Although the only couplings of the interacting measure in the spatial directions are due to the free measure $d\mu$, the measure is highly nonlocal in the imaginary time direction because the fermions are coupled in that direction. Thus, on a scale of difficulty in controlling the infinite-volume theory, our model lies roughly between $P(\phi)_2$ and Y_2.

In Sec. 4 we establish OS positivity,[6,14] existence of the Euclidean pressure,[18,23] and Fröhlich bounds[14,24] for the model. Also, we point out that in the case of weak coupling the cluster expansion[21,22] for the Y_2 model adapted to our model yields convergence of the finite-volume Schwinger functions to infinite-volume Schwinger functions satisfying exponential clustering and the *unEuclidean Osterwalder–Schrader axioms* [i.e., the (EO') Osterwalder–Schrader axioms[6] modified by the explicit inclusion of space reflections and translations in the positivity and clustering axioms and by the replacement of the Euclidean group with the group of imaginary time translations and those space translations appropriate to the model in question]. This information yields the following basic results:

Theorem 1.1: Let E_Λ be the energy of the ground states of H_Λ and $\widetilde{\Omega}_\Lambda = \| P_\Lambda \Omega_\Lambda \|^{-1} P_\Lambda \Omega_\Lambda$, where P_Λ is the projection onto the eigenspace of E_Λ. Then the expectations

$$\Big(T\Big[\prod_{k=1}^{r} (e^{it_k H_\Lambda} Q_{m_k} e^{-it_k H_\Lambda}) \prod_{k=1}^{n} (e^{is_k H_\Lambda} \overline{\psi_{\alpha_k}^{(m_k')}}(x_k) e^{-is_k H_\Lambda})\Big) \times \prod_{k=1}^{n} (e^{is_{n+k} H_\Lambda} \psi_{\beta_k}^{(m_k')}(y_k) e^{-is_{n+k} H_\Lambda})\Big] \widetilde{\Omega}_\Lambda, \widetilde{\Omega}_\Lambda \Big)$$

converge in the sense of distributions as $\Lambda \to \infty$, provided that $\omega_0^{-1}|\lambda|$ and $M^{-1}|\lambda|$ are sufficiently small.

Theorem 1.2: The infinite volume expectations (properly indexed) satisfy all of the nonrelativistic Wightman axioms including a positive energy gap.

The axiomatic result that is used here is a nonrelativistic Osterwalder–Schrader reconstruction theorem. As in our discussion of the nonrelativistic Wightman axioms, we concentrate on scalar distributions on products of $\mathbf{R}^\nu \times \mathbf{R}$ with the understanding that our remarks extend to the more complicated objects that our model would involve.

Theorem 1.3: Let $\{S_n(x_1,\dots,x_n)\}_{n=0}^{\infty}$ be distributions on products of $\mathbf{R}^\nu \times \mathbf{R}$ such that S_n is supported by

$$(\mathbf{R}^\nu \times \mathbf{R})^n_{\neq} \equiv \{((\mathbf{x}_1,s_1),\dots,(\mathbf{x}_n,s_n)) \in (\mathbf{R}^\nu \times \mathbf{R})^n | s_1,\dots,s_n \text{ are distinct}\}$$

in the sense of distributions.[3–6] If the sequence $\{S_n(x_1,\dots,x_n)\}_{n=0}^{\infty}$ satisfies the un-Euclidean Osterwalder-Schrader axioms, then these distributions are the Schwinger functions of a nonrelativistic Wightman theory.

This theorem follows by inspection of the standard OS reconstruction proof.[6] The point is that the OS analytic continuation is the same analytic continuation mentioned in our discussion of the nonrelativistic axioms: time arguments play a distinguished role. Osterwalder and Schrader obtain the nonrelativistic spectrum condition on the Wightman distributions they construct before using Euclidean invariance to obtain relativistic invariance (from which the relativistic spectrum condition, analytic continuation of Wightman distributions to extended forward tubes, and locality follow); i.e., they prove relativistic properties "after the fact." The time cluster property of nonrelativistic Wightman distributions follows from uniqueness of the vacuum they construct, while the space cluster property follows from the space cluster property and temperedness bounds of the Schwinger functions together with the general theory of Laplace transforms and the Vitali convergence theorem.[5] Although space clustering is usually not proved this way in the case of relativistic theories, the argument is of a standard type in constructive field theory.[25]

Remark: The converse of this theorem does not hold. In particular, the Schwinger functions of an arbitrary nonrelativistic theory may not be symmetric in their arguments.

Sections 5 and 6 are devoted to proving that our model satisfies the FKG inequality[26]

Theorem 1.4: Let $\langle\ \rangle_{\Lambda,t}$ be the normalized expectation for the measure (1.21). Let F,G be increasing functions on $\mathbf{R}_n$ such that

$$f(\phi) \equiv F(\phi(m_1,t_1),\dots,\phi(m_n,t_n)),$$
$$g(\phi) \equiv G(\phi(m_1',t_1'),\dots,\phi(m_n',t_n')),$$

and fg are integrable with respect to (1.21). Then

$$\langle fg \rangle_{\Lambda,t} - \langle f \rangle_{\Lambda,t} \langle g \rangle_{\Lambda,t} \geqslant 0.$$

Our basic strategy is the same as in Ref. 1. In Sec. 5 we find a condition that implies Theorem 1.4; as in the case of the Yukawa model, we expect the corresponding sufficiency statement for higher-dimensional Fermi systems to hold because, the counterterms that cancel ultraviolet divergences must be local, and local terms are annihilated in the derivation of the sufficient condition.

In Sec. 6 we prove the sufficient condition for an arbitrary Fermi dimension κ. For our model the condition boils down to

$$\mathrm{tr}\widetilde{S}_m^{(n)}(s,t)\widetilde{S}_m^{(n)}(t,s) \geqslant 0, \quad m \in \mathbf{Z}^\nu, \ n \in \mathbf{Z}^\kappa, \tag{1.22}$$

where $\widetilde{S}_m^{(n)}(s,t)$ is a particular solution of the ordinary differential equation

$$\Big[-\beta_0 \frac{d}{ds} - i \sum_{\ell=1}^{\kappa} n_\ell \beta_\ell + M - \Gamma\phi(m,s) \Big] \widetilde{S}_m^{(n)}(s,t) = \Gamma\delta(s-t)$$

and $\beta_0,\dots,\beta_\kappa$ are Dirac matrices. As in Ref. 1, the two-point function of interest is not explicitly computable, nor do the symmetries of the equation make the desired positivity condition manifest. Roughly speaking, we prove (1.22) by writing the quantity in question as the sum of a *conserved quantity* and a quantity whose sign is manifest.

2. MATTHEWS–SALAM–SEILER PICTURE

As our starting point we take the obvious adaptation of the free boson Feynman–Kac–Nelson formula to our model.

Theorem 2.1: For a finite $\Lambda \subset \mathbf{Z}^\nu$ and $c > 0$, set $H = H_B + c\, \Sigma_{m\in\Lambda} Q_m^2$ and let $\sigma \in \Lambda^n$. Let $d\mu$ be the Gaussian measure on $\mathscr{S}'(\mathbf{Z}^\nu \times \mathbf{R})$ defined by (1.20) and let $F_1,\dots,F_n$ be arbitrary continuous, polynomially bounded functions on $\mathbf{R}$. Then for $s_0 \leqslant s_1 \leqslant \cdots \leqslant s_n \leqslant s_{n+1}$,

$$\Big(e^{-(s_1-s_0)H} \prod_{\ell=1}^{n} (F_\ell(Q_{\sigma(\ell)}) e^{-(s_{\ell+1}-s_\ell)H}) \Omega_B, \Omega_B \Big) = \int d\mu(\phi) \prod_{\ell=1}^{n} F_\ell(\phi(\sigma(\ell),s_\ell)) \times \exp\Big(-c \sum_{m\in\Lambda} \int_{s_0}^{s_{n+1}} \phi(m,s)^2\, ds \Big), \tag{2.1}$$

where the operator $F_\ell(Q_{\sigma(\ell)})$ is defined by the functional calculus. We omit the proof of this theorem because it differs in no significant way from the proof of the Feynman-Kac-Nelson formula for the standard free scalar model.[3,17,18] Our measure $d\mu$ is essentially a lattice of massive Wiener measures coupled by a nearest-neighbor interaction and the proof of (2.1) is based on the Markov property corresponding to such a measure.

Remarks: 1. As in Refs. 3, 17, and 18, we can translate the Gaussian definition of Wick ordering into the Euclidean analog of the definition for time-zero fields by introducing the Euclidean Fock space appropriate to our model—the symmetric Fock space over the Hilbert space

$$\Big\{ f \in \mathscr{S}'(\mathbf{Z}^\nu \times \mathbf{R}) \big| \; \|f\|^2 \equiv \int_{-\infty}^{\infty} d\omega \int_{T^\nu} d^\nu k \frac{|\hat{f}(k,\omega)|^2}{\omega^2 + \omega(k)^2} < \infty \Big\}. \tag{2.2}$$

2. $d\mu$ is supported by the Borel set of all ϕ such that $\phi(m,s)$ is α-Holder continuous in s, provided that $\alpha < \frac{1}{2}$.[27] Moreover, it follows from (1.20) that

$$\int d\mu(\phi)\phi(m,s)^2 < \infty. \tag{2.3}$$

3. It follows from (2.3) and hypercontractivity[3,17,18] that $\phi(m,s)$ lies in $L^p(d\mu)$ for $1 \leqslant p < \infty$, so our random field will not require smearing against test functions when we construct our interacting expectations.

Theorem 2.2: Let $m_1,\dots,m_n \in \mathbf{Z}^\nu$ and $\ell_1,\dots,\ell_n \in \mathbf{Z}$ such that the pairs (m_k,ℓ_k) are distinct. For each k, let F_k be an arbitrary Σ_{Λ_k}-measurable function on $\mathscr{S}'(\mathbf{Z}^\nu \times \mathbf{R})$, where $\Lambda_k = \{m_k\} \times [\ell_k, \ell_k + 1]$. There are constants $\beta,\gamma > 0$, dependent only on the fundamental frequency ω_0 (boson mass), such that if the minimum separation between sets Λ_k is greater than or equal to γ, then for $1 \leqslant p < \infty$

$$\Big(\int \prod_{k=1}^{n} |F_k|^p d\mu\Big)^{1/p} \leqslant \prod_{k=1}^{n} \Big(\int |F_k|^{\beta p} d\mu\Big)^{1/\beta p}. \tag{2.4}$$

This theorem represents an obvious adaptation of checkerboard estimates[18]; as in the case of the free scalar field theory, the basic reasoning for proving such a result involves the Markov property and Nelson's general hypercontractive result.[19] We omit the details, but emphasize that the explicit estimates needed to make the strategy work are

$$\begin{aligned}&\Big(\omega_0^2 + \sum_{\ell=1}^{\nu}(1-\cos k_\ell)\Big)^{1/2} \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{i\omega(s-s')} \times \frac{1}{\omega_0^2+\omega^2+\Sigma_\nu^{\ell=1}(1-\cos k_\ell)}\\ &\quad = \exp\Big[-\Big(\omega_0^2+\sum_{\ell=1}^{\nu}(1-\cos k_\ell)\Big)^{1/2}|s-s'|\Big],\end{aligned} \tag{2.5}$$

$$\begin{aligned}&\Big(\omega_0^2+\omega^2+\sum_{\substack{\ell=1\\ \ell\neq j}}^{\nu}(1-\cos k_\ell)\Big)^{1/2}\frac{1}{2\pi}\int_{-\pi}^{\pi} dk_j\, e^{ik_j(n-n')} \times \frac{1}{\omega_0^2+\omega^2+\Sigma_\nu^{\ell=1}(1-\cos k_\ell)}\\ &\quad \leqslant 2\omega_0^{-1}\Big[1+\frac{1}{2}\Big(\omega_0^2+\omega^2+\sum_{\substack{\ell=1\\ \ell\neq j}}^{\nu}(1-\cos k_\ell)\Big)\Big]^{-|n-n'|}.\end{aligned} \tag{2.6}$$

(2.5) is a standard calculation, while (2.6) follows from the contour-shifting proof of Lemma IV.5 in Ref. 18, where we are shifting from $\operatorname{Im} k_j = 0$ to

$$\operatorname{Im} k_j = \operatorname{sgn}(n-n') \times \ln\Big[1+\frac{1}{2}\Big(\omega_0^2+\omega^2+\sum_{\substack{\ell=1\\ \ell\neq j}}^{\nu}(1-\cos k_\ell)\Big)\Big].$$

We should also note that when one applies (2.6) to the GRS (Guerra, Rosen, and Simon) arguments,[18] one uses the coordinate-zero subspaces of the Euclidean Fock space instead of the time-zero subspace [i.e., the symmetric Fock space over the Hilbert space

$$\Big\{ f \in \mathscr{S}'(\mathbf{Z}^{\nu-1}\times\mathbf{R}) \big| \; \|f\|^2 \equiv \int_{-\pi}^{\pi} d\omega \int_{T^{\nu-1}} d^{\nu-1}k \frac{|\hat{f}(k,\omega)|^2}{(\omega_0^2+\omega^2+\Sigma_{\ell=1}^{\nu-1}(1-\cos k_\ell))^{1/2}} < \infty \Big\} \tag{2.7}$$

instead of the symmetric Fock space over the Hilbert space (1.12)]. The latter space is, of course, the Hilbert space used in applying (2.5). Notice also that the (*a priori*) necessity for a separation γ is made transparent by the estimate (2.6). If ω_0 is small, then the hypercontractivity argument is destroyed unless one chooses a separation like

$$\gamma > \ln 2\omega_0^{-1} / \ln(1+\tfrac{1}{2}\omega_0^2).$$

Because of the stated need for a separation γ, Theorem 2.2 is too weak for useful applications. However, given this theorem, we can immediately eliminate the need for γ.

Theorem 2.3: With the objects and assumptions of Theorem 2.2, drop the separation requirement (i.e., set $\gamma = 1$). There is a constant $\beta > 0$, dependent only on ω_0, such that (2.3) holds.

One proves this by merely decoupling the lattice into sublattices with spacing γ via the Holder inequality and then applying Theorem 2.2. Such a crude initial decoupling does not hurt us because the number of sublattices depends only on γ, which in turn depends only on ω_0.

Remark: It has been known for some time that checkerboard estimates for lattice Markov fields can be derived. (See the concluding remarks of Sec. IV.2 in Ref. 18.)

In our study of the interacting theory we impose an ultraviolet cutoff on the time-zero Fermi fields at each lattice site $m \in \mathbf{Z}$. More specifically, we let $\psi_{\alpha,N}^{(m)}(x)$($\overline{\psi_{\alpha,N}^{(m)}}(x)$) denote the regularization of $\psi_\alpha^{(m)}(x)$($\overline{\psi_\alpha^{(m)}}(x)$) that includes only those Fourier components with momenta $-N,\dots,0,\dots,N$. The cutoff versions of (1.13) and (1.16) are

$$\Big(T\Big[\prod_{\ell=1}^{n}\big(e^{-t_\ell H_F^{(m)}}\overline{\psi_{\alpha_\ell N}^{(m)}}(x_\ell)e^{t_\ell H_F^{(m)}}\big)\times\prod_{\ell=1}^{n}\big(e^{-s_\ell H_F^{(m)}}\psi_{\beta_\ell N}^{(m)}(y_\ell)e^{s_\ell H_F^{(m)}}\big)\Omega_F^{(m)},\Omega_F^{(m)}\Big]\Big)_{\mathcal{F}_F^{(m)}}=\det_{1\leqslant j,k\leqslant n}S_N((x_j,s_j),(y_k,t_k))_{\alpha_j\beta_k}\tag{2.8}$$

and

$$\Big(T\Big[\prod_{\ell=1}^{n}\Big(e^{-s_\ell H_F^{(m)}}\times\sum_{\alpha,\alpha'=0}^{1}:\overline{\psi_{\alpha,N}^{(m)}}(x_\ell)\Gamma_{\alpha\alpha'}\psi_{\alpha',N}^{(m)}(x_\ell):e^{s_\ell H_F^{(m)}}\Big)\Big]\Omega_F^{(m)},\Omega_F^{(m)}\Big)_{\mathcal{F}_F^{(m)}}=\sum_{\alpha_1,\dots,\alpha_n=0}^{1}\det_{1\leqslant j,k\leqslant n}[S_N((x_j,s_j),(x_k,s_k))\Gamma]_{\alpha_j\alpha_k}(1-\delta_{jk}),\tag{2.9}$$

respectively, where

$$S_N((x,s),(y,t))\equiv\frac{1}{2\pi}\sum_{n=-N}^{N}e^{in(x-y)}\times\int_{-\infty}^{\infty}d\omega\,e^{i\omega(s-t)}\frac{i\beta_0\omega+i\beta_1 n+M}{\omega^2+n^2+M^2}.\tag{2.10}$$

Following the semi-Euclidean strategy of Seiler-Simon,[14] we begin by stating the basic result for our regularized Hamiltonian.

Theorem 2.4: Fix $N\in\mathbf{Z}^+$, λ real, $\Lambda\subset\mathbf{Z}^\nu$ finite, $c_N>0$ (where the N-dependence will be specified in Sec. 3), and set

$$H_N=H_B+\sum_{m\in\Lambda}H_F^{(m)}+c_N\sum_{m\in\Lambda}Q_m^2,\tag{2.11}$$

$$V_N=\sum_{m\in\Lambda}Q_m\sum_{\alpha,\alpha'=0}^{1}\int_{-\pi}^{\pi}dx:\overline{\psi_{\alpha,N}^{(m)}}(x)\Gamma_{\alpha\alpha'}\psi_{\alpha',N}^{(m)}(x):.\tag{2.12}$$

Then $H_N+\lambda V_N$ is self-adjoint and bounded below on the domain of H_N.

This theorem follows immediately from the fact that V_N is a Phillips perturbation[14,28] with respect to H_N. This fact in turn follows from the relative operator bound

$$V_N^2\leqslant\text{const}(H_N+1),\tag{2.13}$$

which is a consequence of the operator inequality $\Sigma_{m\in\Lambda}Q_m^2\leqslant(1/c_N)H_N$ and the boundedness of $\int_{-\pi}^{\pi}dx:\overline{\psi_{\alpha,N}^{(m)}}(x)\Gamma_{\alpha\alpha'}\psi_{\alpha',N}^{(m)}(x):$.

Remark: Actually, the additional quadratic term in (2.11) is not necessary for (2.13) to hold[14]; we are introducing it at this early stage because we will need a quadratic counterterm when the ultraviolet cutoff is removed.

Given $f\in C_0^\infty(T^1\times\mathbf{R})\oplus C_0^\infty(T^1\times\mathbf{R})$, set

$$\Psi_N^{(m)}(f)=\sum_{\alpha=0}^{1}\int_{-\infty}^{\infty}ds\,e^{-s(H_N+\lambda V_N)}\times\int_{-\pi}^{\pi}dx\,f_\alpha(x,s)\psi_{\alpha,N}^{(m)}(x)e^{s(H_N+\lambda V_N)},\tag{2.14}$$

$$\overline{\Psi_N^{(m)}}(f)=\sum_{\alpha=0}^{1}\int_{-\infty}^{\infty}ds\,e^{-s(H_N+\lambda V_N)}\times\int_{-\pi}^{\pi}dx\,f_\alpha(x,s)\overline{\psi_{\alpha,N}^{(m)}}(x)e^{s(H_N+\lambda V_N)},\tag{2.15}$$

$$\Phi_{m,N}(t)=e^{-t(H_N+\lambda V_N)}Q_m e^{t(H_N+\lambda V_N)}.\tag{2.16}$$

Although these operators are densely defined as a result of Theorem 2.4 it is not clear whether an arbitrary product of them makes any sense. However, as in the Y_2 case the reordering of an arbitrary product by $T[\ ]$ is defined on a dense domain containing the free vacuum Ω_Λ, where $T[\ ]$ is understood to reorder the integrand of the multiple integral involved.

The Matthews–Salam–Seiler picture of our model is embodied in the following theorem:

Theorem 2.5: Let $\sigma\in\Lambda^n$, $\gamma\in\Lambda^r$, $t>0$, $-t/2\leqslant t_1,\dots,t_r\leqslant t/2$, and $f_1,\dots,f_n$, $g_1,\dots,g_n\in C_0^\infty(T^1\times\mathbf{R})\oplus C_0^\infty(T^1\times\mathbf{R})$ such that the projections of $\operatorname{supp}f_1,\dots,\operatorname{supp}f_n,\operatorname{supp}g_1,\dots,\operatorname{supp}g_n$ onto the second variable are disjoint, contain no t_k, and lie in the interval $[-t/2,t/2]$. Then

$$\Big(e^{-t/2(H_N+\lambda V_N)}T\Big[\prod_{k=1}^{r}\Phi_{\gamma(k),N}(t_k)\prod_{k=1}^{n}\overline{\Psi_N^{(\sigma(k))}}(f_k)\prod_{k=1}^{n}\Psi_N^{(\sigma(k))}(g_k)\Big]e^{-t/2(H_N+\lambda V_N)}\Omega_\Lambda,\Omega_\Lambda\Big)$$
$$=\pm\int d\mu(\phi)\prod_{m\in\Lambda}\Big\{\det_2(1-\lambda S_N\Gamma\chi\phi_m)\exp\Big(-c_N\int_{-t/2}^{t/2}\phi(m,s)^2ds\Big)\Big\}\prod_{\ell=1}^{r}\phi(\gamma(\ell),t_\ell)\prod_{m\in\Lambda}\det_{j,k\in\sigma^{-1}(m)}(\tilde S_{m,N}f_j,g_k),\tag{2.17}$$

where χ is the characteristic function of the interval $[-t/2,t/2]$, $\det_2$ is the regularized determinant[29] for perturbations that are Hilbert–Schmidt with respect to

$$\mathscr{H}_F\equiv\Big\{f\in L^2(T^1\times\mathbf{R})\oplus L^2(T^1\times\mathbf{R})\,\Big|\,\|f\|_{1/2}^2\equiv\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}d\omega\,(n^2+\omega^2+M^2)^{1/2}\sum_{\alpha=0}^{1}|\hat f_\alpha(n,\omega)|^2<\infty\Big\},\tag{2.18}$$

S_N is the operator on $\mathscr{H}_F$ whose kernel is given by (2.10),

$$\tilde S_{m,N}\equiv(1-\lambda S_N\Gamma\chi\phi_m)^{-1}S_N\Gamma,\tag{2.19}$$

$$\phi_m(s)\equiv\phi(m,s),\tag{2.20}$$

and (,) on the rhs of (2.17) denotes the inner product for $L^2(T^1\times\mathbf{R})\oplus L^2(T^1\times\mathbf{R})$.

(2.17) is the Matthews–Salam–Seiler formula[14] for our cutoff model. Since the Fermi systems do not exchange fermions, this formula is not surprising to anyone familiar with the corresponding formula for the Y_2 model. Moreover, the rigorous interpretation of the quantities on the rhs is exactly the same as in the case of Y_2 with a similar ultraviolet cutoff,[14] except that the mechanism[11] for cancellation of poles is not needed in our case; $(\widetilde{S}_{m,N} f_j,g_k)$ is actually a bounded random variable! More specifically,

Lemma 2.6: $(\widetilde{S}_{m,N} f_j,g_k)$ is a $\Sigma_{\{m\}\times[-t/2,t/2]}$-measurable function on $\mathscr{S}'(\mathbf{Z}^\nu\times\mathbf{R})$ (defined $d\mu$-almost-everywhere) and

$$|(\widetilde{S}_{m,N} f_j,g_k)|\leqslant M^{-1}\|f_j\|_{L^2\oplus L^2}\|g_k\|_{L^2\oplus L^2} \tag{2.21}$$

$d\mu$-almost-everywhere.

Proof: Since $d\mu$ is supported by the Borel subset of all ϕ such that $\phi(m,s)$ is α-Holder continuous in s, where we fix $\alpha<1/2$, we may confine our attention to such ϕ. By (2.19), we have

$$\widetilde{S}_{m,N}=(-\nabla_N+M-\lambda\Gamma\chi\phi_m)^{-1}, \tag{2.22}$$

where ∇_N denotes the ultraviolet regularization of

$$\nabla\equiv\beta_0\frac{\partial}{\partial x_0}+\beta 1\frac{\partial}{\partial x_1},$$

which affects the periodic variable x_1 only and includes the Fourier components $\{-N,...,0,...,N\}$.

(Thus $S_N=(-\nabla_N+M)^{-1}$.)

Now since β_0,β_1 are Hermitian matrices and Γ is an anti-Hermitian matrix, we know that $-i\nabla_N-i\lambda\Gamma\chi\phi_m$ is an (unbounded) self-adjoint operator on $L^2(T^1\times\mathbf{R})\oplus L^2(T^1\times\mathbf{R})$. ($\chi\phi_m$ is a bounded perturbation, since we have assumed ϕ_m to be continuous.) Hence

$$\|(-i\nabla_N+iM-i\lambda\Gamma\chi\phi_m)f\|_{L^2\oplus L^2}\geqslant M\|f\|_{L^2\oplus L^2}$$

and this imaginary translate of our self-adjoint operator has dense range. It follows from (2.22) that $\widetilde{S}_{m,N}$ is a bounded operator on our L^2 space and that (2.21) holds.

To show that $\phi\to(\widetilde{S}_{m,N} f_j,g_k)$ is a measurable function, we define an approximating sequence in the following way: for each positive integer l, partition the interval $[-t',t']$ into l equal subintervals and replace $\chi\phi_m$ with the corresponding step function based on evaluation of ϕ_m at, say, the right-hand endpoints of the subintervals. The operator $\widetilde{S}_{m,N,l}$ corresponding to such an approximation ϕ^l_m of ϕ_m clearly depends on ϕ continuously in operator norm (as an operator on our L^2 space) because the relative topology on the Borel set of α-Holder continuous functions is just the topology for pointwise convergence of nets (and therefore $\chi\phi^l_m$ depends on ϕ continuously in *sup norm* by the nature of the approximation). To prove pointwise convergence of $\widetilde{S}_{m,N,l}$ to $\widetilde{S}_{m,N}$ as $l\to\infty$, we need only note that

$$\begin{aligned}|(\widetilde{S}_{m,N,l} f_j,g_k)-(\widetilde{S}_{m,N} f_j,g_k)|&=|(\widetilde{S}_{m,N,l}(\lambda\chi\phi^l_m-\lambda\chi\phi_m)\widetilde{S}_{m,N} f_j,g_k)|\\&\leqslant M^{-1}|\lambda|\,\|(\chi\phi^l_m-\chi\phi_m)\widetilde{S}_{m,N} f_j\|_{L^2\oplus L^2}\|g_k\|_{L^2\oplus L^2}\end{aligned}$$

[since (2.21) clearly holds for $\widetilde{S}_{m,N,l}$ as well]. The uniform convergence of $\chi\phi^l_m$ to $\chi\phi_m$ for a fixed, arbitrary α-Holder continuous ϕ completes the proof.

Notice that the vital ingredients of the proof are that the interaction is pseudoscalar (i.e., Γ is an anti-Hermitian matrix) and that $d\mu$ is supported by continuous ϕ. This lemma will be the key to our vacuum overlap result as well as to our semiboundedness result.

Although the analysis involved in proving Theorem 2.5 is the same as that employed in the proof of the parallel result in Ref. 14, the combinatorial aspect of the proof is not quite trivial. Since this difficulty makes the proof of our theorem a notational nightmare, we will follow the expository spirit of Ref. 14 and merely illustrate the basic argument by proving the following (more modest) theorem.

Theorem 2.5': For $t>0,\lambda$ real,

$$\begin{aligned}(e^{-t(H_N+\lambda V_N)}\Omega_\Lambda,\Omega_\Lambda)&=\int d\mu(\phi)\prod_{m\in\Lambda}\{\det{}_2(1-\lambda S_N\Gamma\phi_m)\\&\quad\times\exp\left(-c_N\int_{-t/2}^{-t/2}\phi(m,s)^2ds\right)\Big\}.\end{aligned} \tag{2.23}$$

Proof: Since V_N is a Phillips perturbation, we know by Theorem 3.1 in Ref. 14 that the Phillips perturbation expansion of $e^{-t(H_N+\lambda V_N)}$ is convergent in operator norm. Now the nth order coefficient of the resulting power series in λ for the lhs of (2.23) is

$$(-1)^n\int_0^t dt_1\cdots\int_0^{t-t_1-\cdots-t_{n-1}}dt_n\,\Big(e^{-(t-t_1-\cdots-t_n)H_N}\times\prod_{\ell=1}^n(V_Ne^{-t_\ell H_N})\Omega_\Lambda,\Omega_\Lambda\Big),$$

which can be rewritten as follows:

$$(-1)^n\int_{-t/2\leqslant s_1\leqslant\cdots\leqslant s_n\leqslant t/2}d^ns\,\Big(e^{-(t/2+s_1)H_N}\times\Big(\prod_{\ell=1}^n(Ve^{-(s_{\ell+1}-s_\ell)H_N}\Big)\Omega_\Lambda,\Omega_\Lambda,$$

where $s_{n+1}\equiv t/2$ and the change in integration variables is given by

$$s_\ell=t/2-t_\ell-\cdots-t_n, 1\leqslant\ell\leqslant n.$$

Applying (1.10), (2.11), and (2.12), we get

$$\begin{aligned}&(-1)^n\int_{-t/2\leqslant s_1\cdots s_n\leqslant t/2}d^ns\\&\times\int_{|x_\ell|\leqslant\pi}d^nx\sum_{\sigma\in\Lambda^n}(e^{-(t/2+s_1)(H_B+c\Sigma_{m\in\Lambda}Q_m^2)}\\&\times\prod_{\ell=1}^n(Q_{\sigma(\ell)}e^{-(s_{\ell+1}-s_\ell)(H_B+c\Sigma_{m\in\Lambda}Q_m^2)})\Omega_B,\Omega_B)\\&\times\Big(e^{-(t/2+s_1)\Sigma_{m\in\Lambda}H_F^{(m)}}\prod_{\ell=1}^n\Big(\sum_{\alpha,\alpha'=0}^1:\overline{\psi_{\alpha,N}^{(\sigma(\ell))}}(x_\ell)\\&\times\Gamma_{\alpha\alpha'}\psi_{\alpha,N}^{(\sigma(\ell))}(x_\ell):e^{-(s_{\ell+1}-s_\ell)\Sigma_{m\in\Lambda}H_F^{(m)}}\Big)\\&\times\bigotimes_{m\in\Lambda}\Omega_F^{(m)},\bigotimes_{m\in\Lambda}\Omega_F^{(m)}\Big),\end{aligned} \tag{2.24}$$

where we temporarily drop the subscript of c_N for convenience. The fermion factor splits into

$$\prod_{m\in\Lambda}\Big(\prod_{\ell\in\sigma^{-1}(m)}\Big(\sum_{\alpha,\alpha'=0}^1:\psi_{\alpha,N}^{(m)}(x_\ell)\times\Gamma_{\alpha\alpha'}\psi_{\alpha',N}^{(m)}(x_\ell):e^{-(s_{\ell_m}-s_\ell)H_F^{(m)}}\Big)\Omega_F^{(m)},\Omega_F^{(m)}\Big),$$

where ℓ_m denotes the successor of ℓ in $\sigma^{-1}(m)$ and the operator product is taken in the order of increasing ℓ. $(e^{-(1/2t+s_1)\Sigma_{m\in\Lambda}H_F^{(m)}}$ has disappeared by application to the other side of the inner product, and we have used all available commutativity to telescope exponents.) By Theorem 2.1 the boson factor can be replaced by

$$\int d\mu(\phi)\prod_{\ell=1}^{n}\phi(\sigma(\ell),s_\ell)e^{-c\Sigma_{m\in\Lambda}\int_{-t/2}^{t/2}ds\,\phi(m,s)^2}$$

and by formula (2.9) the mth fermion factor can be replaced by

$$\sum_{\alpha:\sigma^{-1}(m)\to\{0,1\}}\det{}_{j,k\in\sigma^{-1}(m)}(1-\delta_{jk})$$
$$\times[S_N((x_j,s_j),(x_k,s_k))\Gamma]_{\alpha_j\alpha_k},$$

where the structure of the matrix makes the determinant independent of the ordering of $\sigma^{-1}(m)$. Thus (2.24) becomes

$$(-1)^n\int_{-t/2<s_1<\cdots<s_n<t/2}d^ns\int_{|x_\ell|<\pi}d^nx\sum_{\sigma\in\Lambda}\int$$
$$\times d\mu(\phi)\prod_{\ell=1}^{n}\phi(\sigma(\ell),s_\ell)e^{-c\Sigma_{m\in\Lambda}\int_{-t/2}^{t/2}\phi(m,s)^2ds}$$
$$\times\sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_n=0}^{1}\prod_{m\in\Lambda}\det{}_{j,k\in\sigma^{-1}(m)}(1-\delta_{jk})$$
$$\times[S_N((x_j,s_j),(x_k,s_k))\Gamma]_{\alpha_j\alpha_k}.$$

Now since the s-integrand involves a summation over all $\sigma\in\Lambda^n$, it follows via arbitrary permutations of variables in the x-integration and α-summation that the s-integrand is symmetric in $s_1,\ldots,s_n$, so we may replace

$$(-1)^n\int_{-t/2<s_1<\cdots<s_n<t/2}d^ns\ \text{ with }\ \frac{(-1)^n}{n!}\int_{|s_\ell|<t/2}d^ns.$$

Now for convenience set

$$A_n(\phi,\sigma)$$
$$=\int_{|s_\ell|<t/2}d^ns\int_{|x_\ell|<\pi}d^nx\sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_n=0}^{1}\prod_{\ell=1}^{n}\phi(\sigma(\ell),s_\ell)$$
$$\times e^{-c\Sigma_{m\in\Lambda}\int_{-t/2}^{t/2}\phi(m,s)^2ds}$$
$$\times\prod_{m\in\Lambda}\det{}_{j,k\in\sigma^{-1}(m)}(1-\delta_{jk})$$
$$\times[S_N((x_j,s_j),(x_k,s_k))\Gamma]_{\alpha_j\alpha_k}$$

so that our expression becomes $[(-1)^n/n!]\int d\mu(\phi)\,\Sigma_{\sigma\in\Lambda^n}A_n(\phi,\sigma)$. Consider the mapping $\tau:\Lambda^n\to\{0,\ldots,n\}^\Lambda$ defined by

$$\tau(\sigma)(m)=|\sigma^{-1}(m)|$$

and notice that if $\tau(\sigma)=\tau(\sigma')$ then σ can be transformed into σ' via a permutation of $\{1,\ldots,n\}$. But in the expression for $A_n(\phi,\sigma)$ such a transformation to σ' amounts to a permutation in the integration and summation variables. Thus

$$\tau(\sigma)=\tau(\sigma')\Rightarrow A_n(\phi,\sigma)=A_n(\phi,\sigma'),$$

so our expression may be rewritten as

$$\frac{(-1)^n}{n!}\int d\mu(\phi)\sum_{\kappa:\Lambda\to\{0,\ldots,n\}}|\tau^{-1}(\kappa)|A_n(\phi,\sigma_\kappa),$$

where σ_κ is a representative element of $\tau^{-1}(\kappa)$. Since

$$|\tau^{-1}(\kappa)|=\begin{cases}n!/\prod_{m\in\Lambda}\kappa(m)!, & \sum_{m\in\Lambda}\kappa(m)=n,\\ 0, & \text{otherwise,}\end{cases}$$

$$(2.24)=\int d\mu(\phi)\sum_{\substack{\kappa:\Lambda\to\{0,\ldots,n\}\\ \sum_{m\in\Lambda}\kappa(m)=n}}\prod_{m\in\Lambda}\frac{(-1)^{\kappa(m)}}{\kappa(m)!}A_n(\phi,\sigma_\kappa)$$
$$=\int d\mu(\phi)\sum_{\substack{k:\Lambda\to\{0,\ldots,n\}\\ \sum_{m\in\Lambda}\kappa(m)=n}}\left(\prod_{m\in\Lambda}\frac{(-1)^{\kappa(m)}}{\kappa(m)!}\right)\int_{|s_\ell|<t/2}d^ns$$
$$\times\int_{|x_\ell|<\pi}d^nx\sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_n=0}^{1}\prod_{m\in\Lambda}\Big\{\prod_{\ell\in\sigma_\kappa^{-1}(m)}\phi(m,s_\ell)$$
$$\times\det{}_{j,k\in\sigma_\kappa^{-1}(m)}(1-\delta_{jk})[S_N((s_j,s_j),(x_k,s_k))\Gamma]_{\alpha_j\alpha_k}\Big\}$$
$$\times e^{-c\Sigma_{m\in\Lambda}\int_{-t/2}^{t/2}ds\,\phi(m,s)^2}.$$

But the $\int d\mu(\phi)$-integrand of this last expression is just the nth order coefficient of

$$\prod_{m\in\Lambda}\Big(\sum_{n=0}^{\infty}\frac{(-\lambda)^n}{n!}\int_{|s_\ell|<t/2}d^ns\int_{|x_\ell|<\pi}d^nx\prod_{\ell=1}^{n}\phi(m,s)$$
$$\times\sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_n=0}^{1}\det{}_{1\le j,k\le n}(1-\delta_{jk})[S_N((x_j,s_j)(x_k,s_k))\Gamma]_{\alpha_j\alpha_k}\Big)$$
$$\times e^{-c\Sigma_{m\in\Lambda}\int_{-t/2}^{t/2}\phi(m,s)^2ds},$$

and the mth factor is the Fredholm expansion[29,30] of

$$\det{}_2(1-\lambda S_N\Gamma\chi\phi_m),\tag{2.25}$$

so we have now established the desired equation (2.23) *formally*. To establish it rigorously, we must show that the Fredholm expansion

$$\sum_{n=0}^{\infty}\frac{(-\lambda)^n}{n!}\int_{|s_\ell|<t/2}d^ns\int_{|x_\ell|<\pi}d^nx\prod_{\ell=1}^{n}\phi(m,s_\ell)$$
$$\times\sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_n=0}^{1}\det{}_{1\le j,k\le n}(1-\delta_{jk})$$
$$\times[S_N((x_j,s_j),(x_k,s_k))\Gamma]_{\alpha_j\alpha_k}\tag{2.26}$$

converges in $L^p(d\mu)$ for $1\leqslant p<\infty$. [Obviously this will also establish that (2.25) lies in $L^p(d\mu)$, so that the rhs of (2.23) makes sense.] At this point we must interrupt the proof of our theorem to discuss a serious gap in the proof of Theorem 3.2 in Ref. 14 and show how it can be filled.

By a straightforward calculation it can be shown that the Hilbert–Schmidt norm of

$$K_{m,N}\equiv S_N\Gamma\chi\phi_m\tag{2.27}$$

as an operator on $\mathscr{H}_F$ lies in $L^2(d\mu)$, so by the basic Fredholm theory (2.26) converges to (2.25) in the $d\mu$-almost everywhere sense. To obtain the stronger convergence one needs a stronger property in order to exploit the hypercontractive properties of $d\mu$; e.g., it would help if $K_{m,N}$ were trace-class. However, we do not know whether the trace norm of $K_{m,N}$ lies in any $L^p(d\mu)$, or even whether $K_{m,N}$ is trace-class in the $d\mu$-almost everywhere sense. Moreover, we do not see how to prove that the corresponding operator for Y_2 is trace-class as it is claimed to be in Theorem 3.2 of Ref. 14. The point is that the regularization on the time-zero Fermi fields induces a regularization on the Euclidean Fermi propagator *with respect to the space variable only* (see formula (2.10) in the case of our model); there is no ultraviolet cutoff with respect to imaginary time, and the method in Ref. 12 for proving that the appropriate operator is trace-class seems to depend on having an ultraviolet cutoff in *all* directions. Actually, the K operator that Seiler and Simon work with in Ref. 14 is trace-class anyway, but it is defined by the wrong Euclidean Fermi propagator (see Remark 1 after Theorem 3.2 in Ref. 14). Even if $K_{m,N}$ is not trace-class, however, the argument in Ref. 14 involving wholesale expansion in powers of λ is basically suited for our case as well as for theirs; one only needs more delicate estimates. We introduce these estimates within the context of our model, but they are virtually the same as in the Y_2 case.

Since $\phi_m(s)\in L^p(d\mu)$, $1\leqslant p<\infty$, an easy calculation shows that the Hilbert–Schmidt norm of $K_{m,N}$ with respect to $L^2(T^1\times\mathbf{R})\oplus L^2(T^1\times\mathbf{R})$ is also in $L^2(d\mu)$. Hence $\det_2(1-\lambda K_{m,N})$ is well defined $d\mu$-almost everywhere with respect to $L^2(T^1\times\mathbf{R})\oplus L^2(T^1\times\mathbf{R})$ and certainly coincides with its definition with respect to $\mathscr{H}_F$. Since it does not matter which Hilbert space we use, we choose to work with $L^2(T^1\times\mathbf{R})\oplus L^2(T^1\times\mathbf{R})$ because it is more convenient. Let D be the positive operator defined by

$$D^2\equiv-\frac{d^2}{dt^2}-\frac{d^2}{dx^2}+M^2,$$

and set

$$D(\omega,n)\equiv(\omega^2+n^2+M^2)^{1/2}.$$

For $1\leqslant p<\infty$ we denote the class $\mathscr{C}_p$ norm with respect to $L^2(T^1\times\mathbf{R})\oplus L^2(T^1\times\mathbf{R})$ by $\|\ \|_p$. We have:

Lemma 2.7: (Rosen) For $\epsilon>0,\epsilon'>\frac{1}{2}$,

$$\|D^{1/2-\epsilon}K_{m,N}\|_2^2=\text{const}\int_{-t/2}^{t/2}\phi_m(s)^2\,ds, \tag{2.28}$$

$$\|D^{-\epsilon'}K_{m,N}\|_1^2\leqslant\text{const}\int_{-t/2}^{t/2}\phi_m(s)^2\,ds. \tag{2.29}$$

Proof: For $\epsilon>0$

$$\|D^{1/2-\epsilon}K_{m,N}\|_2^2=\operatorname{Tr}\chi\phi_m D^{-1-2\epsilon}h_N\chi\phi_m,$$

where h_N is multiplication in momentum space by the characteristic function of $\{-N,\ldots,0,\ldots,N\}\times\mathbf{R}$. This follows from the relation

$$S_N^\dagger S_N=D^{-2}h_N, \tag{2.30}$$

where $\dagger$ denotes adjoint with respect to $L^2(T^1\times\mathbf{R})\oplus L^2(T^1\times\mathbf{R})$. The integral expression in momentum space is

$$\begin{aligned}\|D^{1/2-\epsilon}K_{m,N}\|_2^2&=2\sum_{n=-N}^{N}\sum_{n'=-\infty}^{\infty}\int_{-\infty}^{\infty}d\omega\\&\quad\times\int_{-\infty}^{\infty}d\omega'\,D(\omega,n)^{-1-2\epsilon}|\chi\phi_m(\omega'-\omega)|^2\delta_{n'n}\\&=2\sum_{n=-N}^{N}\int_{-\infty}^{\infty}d\omega D(\omega,n)^{-1-2\epsilon}\\&\quad\times\int_{-\infty}^{\infty}d\omega'\,|\widehat{\chi\phi}_m(\omega')|^2.\end{aligned}$$

Since the $\int_{-\infty}^{\infty}d\omega$-integration yields finite constants and $\int_{-\infty}^{\infty}d\omega'\,|\chi\phi_m(\omega')|^2=\int_{-t/2}^{t/2}ds\,\phi_m(s)^2$, we have

$$\|D^{1/2-\epsilon}K_{m,N}\|_2^2=\text{const}\int_{-t/2}^{t/2}ds\,\phi_m(s)^2.$$

To establish (2.29) we try to write $D^{-\epsilon'}K_{m,N}$ as a product of Hilbert–Schmidt operators. Let $\zeta\in C_0^\infty(\mathbf{R})$ such that $\zeta=1$ on $[-t/2,t/2]$ and set

$$J=D^{-\epsilon'}S_N\zeta D^a,$$

$$L=D^{-a}\Gamma\chi\phi_m,$$

where $a>0$ is to be chosen. Obviously $D^{-\epsilon'}K_{m,N}=JL$; moreover,

$$\|J\|_2^2=\operatorname{Tr}D^a\zeta D^{-2-2\epsilon'}h_N\zeta D^a$$

by (2.30), so we have

$$\begin{aligned}\|J\|_2^2&=2\sum_{n=-N}^{N}\sum_{n'=-\infty}^{\infty}\int_{-\infty}^{\infty}d\omega\int_{-\infty}^{\infty}d\omega'\,D(\omega',n')^{2a}\\&\quad\times D(\omega,n)^{-2-2\epsilon'}|\hat{\zeta}(\omega'-\omega)|^2\delta_{nn'}\\&=2\sum_{n=-N}^{N}\int_{-\infty}^{\infty}d\omega\,D(\omega,n)^{-2-2\epsilon'}\\&\quad\times\int_{-\infty}^{\infty}d\omega'\,D(\omega',n)^{2a}|\hat{\zeta}(\omega'-\omega)|^2.\end{aligned}$$

Since $\hat{\zeta}$ decays faster than any inverse polynomial, we know that

$$\int_{-\infty}^{\infty}d\omega'\,D(\omega',n)^{2a}|\hat{\zeta}(\omega'-\omega)|^2\leqslant\text{const}_n(\omega^2+M^2)^a$$

so

$$\|J\|_2^2\leqslant2\sum_{n=-N}^{N}\text{const}_n\int_{-\infty}^{\infty}d\omega(\omega^2+M^2)^{a-1-\epsilon'}$$

because $D(\omega,n)^2\geqslant\omega^2+M^2$. Obviously $\|J\|_2^2<\infty$ if $a<\frac{1}{2}+\epsilon'$. Examining L, we see that

$$\begin{aligned}\|L\|_2^2&=\operatorname{Tr}\chi\phi_m D^{-2a}\chi\phi_m\\&=2\sum_{n=-\infty}^{\infty}\sum_{n'=-\infty}^{\infty}\int_{-\infty}^{\infty}d\omega\int_{-\infty}^{\infty}d\omega' D(\omega,n)^{-2a}\\&\quad\times|\chi\phi_m(\omega'-\omega)|^2\delta_{n'n}\\&=2\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}d\omega\,D(\omega,n)^{-2a}\int d\omega'\,|\chi\phi_m(\omega')|^2\\&=2\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}d\omega\,D(\omega,n)^{-2a}\int_{-t/2}^{t/2}\phi_m(s)^2\,ds.\end{aligned}$$

The constant is finite if $a>1$. Such an a can be picked for J and L if and only if $\epsilon'>\frac{1}{2}$. Hence

$$\begin{aligned}\|D^{-\epsilon'}K_{m,N}\|_1^2&\leqslant\|J\|_2^2\|L\|_2^2\\&\leqslant\text{const}\int_{-t/2}^{t/2}\phi_m(s)^2\,ds\end{aligned}$$

for such ϵ'. □

Since $D^zK_{m,N}$ is an analytic family of operators, it follows from complex interpolation[12] that for $\epsilon\leqslant\frac{1}{2}$,

$$\|K_{m,N}\|_q\leqslant\|D^{-\epsilon'}K_{m,N}\|_1^{1-t}\|D^{1/2-\epsilon}K_{m,N}\|_2^t,$$

where $q^{-1}=1-\frac{1}{2}t$ and $t=2\epsilon'/(1-2\epsilon+2\epsilon')$. Thus, Lemma 2.7 implies that

$$\|K_{m,N}\|_q^2\leqslant\text{const}\int_{-t/2}^{t/2}\phi_m(s)^2\,ds\equiv F(\phi), \tag{2.31}$$

provided that $0<\epsilon\leqslant\frac{1}{2},\epsilon'>\frac{1}{2}$. From now on we fix such ϵ,ϵ'

(and therefore q) with $\epsilon < \frac{1}{2}$ (and therefore $q < 2$). In particular, $\|K_{m,N}\|_q < \infty$ $d\mu$-almost everywhere, so by the theory of regularized determinants[28] $\det_2(1 - \lambda K_{m,N})$ is entire in λ $d\mu$-almost everywhere and we have the bound

$$|\det_2(1 - \lambda K_{m,N})| \leqslant \exp(C|\lambda|^q \|K_{m,N}\|_q^q) \tag{2.32}$$

for some $C > 0$.

Corollary 2.8: (Rosen) (2.26) converges in $L^p(d\mu)$ for $1 \leqslant p < \infty$.

Proof: Since $\det_2(1 - \lambda K_{m,N})$ is entire in λ, we may write

$$\det_2(1 - \lambda K_{m,N}) = \sum_{n=0}^{\infty} a_n \lambda^n,$$

where

$$a_n = \frac{1}{n!} \frac{d^n}{d\lambda^n} \det_2(1 - \lambda K_{m,N})|_{\lambda = 0}.$$

The Cauchy formula yields

$$a_n = \frac{1}{2\pi i} \int_{|\lambda| = R} \frac{\det_2(1 - \lambda K_{m,N})}{\lambda^{n+1}} d\lambda.$$

By (2.32),

$$|a_n| \leqslant R^{-n} \exp(CR^q \|K_{m,N}\|_q^q),$$

so if we set $R = n^{1/q} \|K_{m,N}\|_q^{-1}$, we have

$$|a_n| \leqslant n^{-n/q} \|K_{m,N}\|_q^n e^{nC}.$$

By (2.31)

$$\|K_{m,N}\|_q^n \leqslant F(\phi)^{n/2},$$

so by hypercontractivity

$$\begin{aligned} \left(\int |a_n|^p d\mu \right)^{1/p} &\leqslant n^{-n/q} e^{Cn} \left(\int \|K_{m,N}\|_q^{pn} d\mu \right)^{1/p} \\ &\leqslant n^{-n/q} e^{Cn} \left(\int F^{pn/2} d\mu \right)^{1/p} \\ &\leqslant n^{-n/q} e^{Cn} \left[\left(\frac{pn}{2} - 1 \right) \left(\int F^2 d\mu \right)^{1/2} \right]^{n/2}, \quad p \geqslant 2, n \geqslant 2, \end{aligned}$$

because $F(\phi)$ involves the square of ϕ. Thus

$$\begin{aligned} \left(\int_{-\infty}^{\infty} |a_n|^p d\mu \right)^{1/p} &\leqslant \text{const}^n n^{-n/q} ((pn/2) - 1)^{n/2} \\ &\leqslant \text{const}^n n^{-n/q} (p/2)^{n/2} n^{n/2} \end{aligned}$$

for $p \geqslant 2, n \geqslant 2$. Since $q < 2$, this estimate yields an infinite radius of $L^p(d\mu)$-convergence of the power series for $p \geqslant 2$, and so we have such convergence for $p \geqslant 1$. But the series is just (2.26).

This completes the proof of Theorem 2.5′. □

3. SEMIBOUNDEDNESS AND VACUUM OVERLAP

We now proceed to show how the ultraviolet cutoff is removed. Our basic approach is very similar to the Seiler–Simon approach[14] to the Y_2 model, except that the absence of a Reeh–Schlieder theorem in our case compels us to use methods that are (apparently) peculiar to the pseudoscalar character of the interaction in our model.

Lemma 3.1: There is a dense subspace D of $\mathscr{H}_\Lambda$ with the property that for every $\eta \in D$ there exists a $C > 0$ such that

$$(e^{-t(H_N + \lambda V_N)}\eta, \eta) \leqslant C(e^{-t(H_N + \lambda V_N)}\Omega_\Lambda, \Omega_\Lambda)$$

for all $N \in \mathbf{Z}^+$, $t > 0$, where Ω_Λ is the free vacuum.

This lemma is the analog of Theorem 5.4 in Ref. 14. To prove it we must consider a modification of the type of states (Jost states) used in Ref. 14.

Consider vectors in $\mathscr{H}_\Lambda$ of the form

$$\begin{aligned} &\prod_{k=1}^{r} (e^{it_k H_0} Q_{\gamma(k)} e^{-it_k H_0}) \\ &\quad \times \prod_{k=1}^{n} (e^{it_{r+k} H_0} \overline{\psi^{(\sigma(k))}}(F_k) e^{-it_{r+k} H_0}) \\ &\quad \times \prod_{k=1}^{n'} (e^{it_{r+n+k} H_0} \psi^{(\sigma'(k))}(G_k) e^{-it_{r+n+k} H_0}) \Omega_\Lambda, \end{aligned}$$

where H_0 is the free Hamiltonian, $\sigma \in \Lambda^n$, $\sigma' \in \Lambda^{n'}$, $\gamma: \{1, \ldots, r\} \to \mathbf{Z}^\nu$, $F_1, \ldots, F_n, G_1, \ldots, G_{n'} \in C^\infty(T^1) \oplus C^\infty(T^1)$, and

$$\psi^{(m)}(G) \equiv \sum_{\alpha=0}^{1} \int_{-\pi}^{\pi} G_\alpha(x) \psi_\alpha^{(m)}(x)\, dx,$$

$$\overline{\psi^{(m)}}(F) \equiv \sum_{\alpha=0}^{1} \int_{-\pi}^{\pi} F_\alpha(x) \overline{\psi_\alpha^{(m)}}(x)\, dx.$$

By cyclicity of the vacuum for our free model, the set of linear combinations of such vectors is a dense subspace of $\mathscr{H}_\Lambda$. Clearly, if we consider arbitrary C_0^∞ functions $\zeta_1, \ldots, \zeta_r$ on $\mathbf{R}$, then the expression

$$\begin{aligned} &\prod_{k=1}^{r} (e^{it_k H_0} \zeta_k(Q_{\gamma(k)}) e^{-it_k H_0}) \prod_{k=1}^{n} \left(e^{it_{r+k} H_0} \overline{\psi^{(\sigma(k))}}(F_k) e^{-it_{r+k} H_0}\right) \\ &\quad \times \prod_{k=1}^{n'} (e^{-t_{n+r+k} H_0} \psi^{(\sigma'(k))}(G_k) e^{-it_{n+r+k} H_0}) e^{it_{r+n+n'} H_0} \end{aligned} \tag{3.1}$$

is a product of bounded operators only. ($\zeta_k(Q_{\gamma(k)})$ is understood in the sense of the functional calculus.) Moreover, arbitrary linear combinations of the application of such products to Ω_Λ from a dense subspace of $\mathscr{H}_\Lambda$ because $Q_{\gamma(k)}$ can be strongly approximated by a sequence of operators like $\zeta_k(Q_{\gamma(k)})$. Also, (3.1) remains bounded if $t_1, \ldots, t_{r+n+n'}$ are replaced by complex variables $z_1, \ldots, z_{r+n+n'}$ such that $\operatorname{Im} z_1 \geqslant 0$ and $\operatorname{Im} z_k \leqslant \operatorname{Im} z_{k+1}$, and this operator-valued function is strongly analytic in each variable on the interior of this domain; this follows from the fact that H_0 is bounded below.

Theorem 3.2: Consider all linear combinations of vectors of the form

$$\begin{aligned} &\prod_{k=1}^{r} (e^{-s_k H_0} \zeta_k(Q_{\gamma(k)}) e^{s_k H_0}) \prod_{k=1}^{n} \left(e^{-s_{k+r} H_0} \overline{\psi^{(\sigma(k))}}(F_k) e^{s_{k+r} H_0}\right) \\ &\quad \times \prod_{k=1}^{n'} (e^{-s_{n+r+k} H_0} \psi^{(\sigma'(k))}(G_k) e^{s_{n+r+k} H_0}) \Omega_\Lambda, \end{aligned} \tag{3.2}$$

where $s_k \leqslant s_{k+1}$. The linear space is dense in $\mathscr{H}_\Lambda$.

As in the case of Lemma 5.2 in Ref. 14, the proof consists of pointing out that the orthogonality statement for a candidate vector with respect to all the vectors (3.2) is just the statement that the analytic functions defined by the inner product vanish on determining sets; this implies that the candidate vector is orthogonal to the application of all the operators (3.1) to Ω_Λ, and so the candidate vector is zero by the density of linear combinations of such states.

Corollary 3.3: Let D be the set of all linear combinations of vectors of the form

$$\prod_{k=1}^{r}\left(e^{-s_k H_0}\zeta_k(Q_{\gamma(k)})e^{s_k H_0}\right)\int_{-\infty}^{\infty} ds_{r+1}\cdots\int_{-\infty}^{\infty} ds_{r+2n}$$
$$\times\prod_{k=1}^{n}\left(e^{-s_{r+k}H_0}\,\overline{\psi^{(\sigma(k))}}(f_k(\cdot,s_{r+k}))\right)$$
$$\times e^{s_{r+k}H_0}\prod_{k=1}^{n'}\left(e^{-s_{r+n+k}H_0}\psi^{(\sigma'(k))}\right.$$
$$\left.\times(g_k(\cdot,s_{r+n+k}))e^{s_{r+n+k}H_0}\right)\Omega_\Lambda, \tag{3.3}$$

where $\gamma:\{1,...,r\}\to\mathbf{Z}^\nu$, $\sigma\in\Lambda^n$, $\sigma'\in\Lambda^{n'}$, $\zeta_1,...,\zeta_r\in C_0^\infty(\mathbf{R})$, $f_1,...,f_n,g_1,...,g_{n'}\in C_0^\infty(T^1\times\mathbf{R})\oplus C_0^\infty(T^1\times\mathbf{R})$, $s_1\geqslant 0$, $s_k\leqslant s_{k+1}$ for $1\leqslant k<r$, and (denoting the projection of $T^1\times\mathbf{R}$ onto $\mathbf{R}$ by π_2)

$s_r\leqslant\pi_2\,\mathrm{supp}f_1\leqslant\cdots\leqslant\pi_2\,\mathrm{supp}f_n\leqslant\pi_2\,\mathrm{supp}g_1\leqslant\cdots\leqslant\pi_2\,\mathrm{supp}\,g_{n'}$.

Then D is dense in $\mathscr{H}_\Lambda$.

Vectors of this form are the "good Jost states"[14] that we have decided to use because we can show that the dense subspace D works for Lemma 3.1. We need a Matthews–Salam–Seiler formula for these states.

Theorem 3.4: With the objects introduced in Corollary 3.3, set $\eta=(3.3)$. Then for $t>0$

$$(e^{-t(H_N+\lambda V_N)}\eta,\eta)=\int d\mu(\phi)\prod_{m\in\Lambda}\Big\{\det{}_2(1-\lambda K_{m,N})$$
$$\times\exp\Big(-c_N\int_{-t/2}^{t/2}\phi_m(s)^2\,ds\Big)\Big\}$$
$$\times\prod_{k=1}^{r}\zeta_k(\phi_{\gamma(k)}(s_k+\tfrac{1}{2}t))$$
$$\times\prod_{k=1}^{r}\zeta_k(\phi_{\gamma(k)}(-s_k-\tfrac{1}{2}t))$$
$$\times\prod_{m\in\Lambda}\det{}_{j,k\in\sigma^{-1}(m)\vee\sigma'^{-1}(m)}(\tilde{S}_{m,N}\tilde{f}_j,\tilde{g}_k), \tag{3.4}$$

where $\vee$ denotes adjunction,

$$\tilde{f}_j=\begin{cases}f_j^+, & j\in\sigma^{-1}(m),\\ g_j^-, & j\in\sigma'^{-1}(m),\end{cases}\qquad \tilde{g}_k=\begin{cases}g_k^+, & k\in\sigma^{-1}(m),\\ f_k^-, & k\in\sigma'^{-1}(m),\end{cases}$$

$f^\pm(x,s)\equiv f(x,\pm(s+\tfrac{1}{2}t))\}$,

and the χ that occurs in the definitions (2.19) [*and* (2.27)] *of $\tilde{S}_{m,N}$ and $K_{m,N}$, respectively, is the characteristic function of* $[t/2,t/2]$.

Given Theorem 2.1, the proof of this theorem does not differ significantly from the proof of Theorem 2.5. The support of χ and the displacement of the test functions arise from the fact that the good Jost states are generated by *free* propagations: the relationship of the time-ordering to the Matthews–Salam–Seiler picture causes the test functions to be supported outside the region of interaction.

Proof of Lemma 3.1: By the Schwartz inequality we may assume that $\eta\in D$ is a good Jost state instead of an arbitrary linear combination of such states. Thus, with $\eta=(3.3)$ we apply (3.4) to obtain

$$(e^{-t(H_N+V_N)}\eta,\eta)\leqslant\prod_{k=1}^{r}\|\zeta_k\|_\infty^2\int d\mu(\phi)\prod_{m\in\Lambda}\Big\{\det{}_2(1-\lambda K_{m,N})\exp\Big(-c_N\int_{-t/2}^{t/2}\phi_m(s)^2\,ds\Big)\Big\}$$
$$\times\prod_{m\in\Lambda}|\det{}_{j,k\in\sigma^{-1}(m)\vee\sigma'^{-1}(m)}(\tilde{S}_{m,N}\tilde{f}_j,\tilde{g}_k)|,$$

where the positivity of $\det_2(1-\lambda K_{m,N})$ follows from Seiler's reasoning.[11] Applying Lemma 2.6 we get

$$(e^{-t(H_N+V_N)}\eta,\eta)\leqslant\prod_{m\in\Lambda}(2|\sigma^{-1}(m)|)!M^{-n}\prod_{k=1}^{r}\|\zeta_k\|_\infty^2\prod_{k=1}^{n}(\|f_k\|_{L^2\oplus L^2}\|g_k\|^2_{L^2\oplus L^2}$$
$$\times\int d\mu(\phi)\prod_{m\in\Lambda}\Big\{\det{}_2(1-\lambda K_{m,N})\exp\Big(-c_N\int_{-t/2}^{t/2}\phi_m(s)^2\,ds\Big)\Big\}=\mathrm{const}(e^{-t(H_N+\lambda V_N)}\Omega_\Lambda,\Omega_\Lambda).\qquad\square$$

Theorem 3.5. Let

$$H_{\Lambda,N}\equiv H_N+\lambda V_N-\frac{\lambda^2}{2}\sum_{m\in\Lambda}\int\mathrm{Tr}(K_{m,N}^2+K_{m,N}^\dagger K_{m,N})\,d\mu, \tag{3.5}$$

where Tr denotes the trace for operators that are trace-class with respect to $\mathscr{H}_F$, and set

$$c_N=\frac{\lambda^2}{4\pi}\sum_{n=-N}^{N}\int_{-\infty}^{\infty}d\omega\,\frac{1}{\omega^2+n^2+M^2} \tag{3.6}$$

in the definition (2.11) of H_N. Then there is a $C>0$, independent of Λ,t,N such that

$$(e^{-tH_{\Lambda,N}}\Omega_\Lambda,\Omega_\Lambda)\leqslant C^{t|\Lambda|}.$$

Given the Matthews–Salam–Seiler formula (2.23) for our model and the checkerboard estimates of Sec. 2, this theorem follows from the estimates of Seiler–Simon[13] adapted to our model.

Remark: Obviously $c_N\to\infty$ as $N\to\infty$, and this constant is the shift in the spring constant for each harmonic oscillator (i.e., shift in the square of the boson mass) that contributes to the ultraviolet renormalization of the model. Another counterterm is the last term in the definition of $H_{\Lambda,N}$; this constant is the energy shift that completes the renormalization, and it diverges as $N\to\infty$ because the Hilbert–Schmidt norm of

$$K_m\equiv S\Gamma\chi\phi_m \tag{3.7}$$

does not lie in $L^2(d\mu)$. We should also note that by the methods of Seiler–Simon[12] this cancellation of infinities can still be controlled if c_N is modified by the addition of a finite constant. (For the energy counterterm such a claim is, of course, trivial.)

Theorem 3.6: $H_{\Lambda,N}$ converges in the strong resolvent sense as $N\to\infty$ to a self-adjoint operator H_Λ bounded below.

Proof: Given Lemma 3.1, Theorem 3.4, and Theorem 3.5, this theorem follows from exactly the same reasoning as in the proof of the Y_2 result in Ref. 14. □

Corollary 3.7: For every $\eta \in D$ there is a $C > 0$ such that

$$(e^{-tH_\Lambda}\eta,\eta) \leqslant C(e^{-tH_\Lambda}\Omega_\Lambda,\Omega_\Lambda)$$

for all $t > 0$.

Proof: According to Lemma 3.1 the constant C is independent of N, so the corollary is immediate. □

Theorem 3.8: $E_\Lambda \equiv \inf \operatorname{spec} H_\Lambda$ is an isolated eigenvalue of H_Λ with finite multiplicity.

Proof: This result follows from the adaptation to our model of the regularizations, first-order operator inequalities, and operator resolvent arguments of Glimm and Jaffe in their treatment of Y_2.[10] We do not know whether E_Λ is simple. □

By the same reasoning used in the proof of Lemma 5.1 of Ref. 14, it follows from Corollary 3.7 that Ω_Λ is not orthogonal to the space of interacting ground states for a given volume Λ. More precisely,

Theorem 3.9: Let P_Λ denote the projection of $\mathscr{H}_\Lambda$ onto the eigenspace of E_Λ. Then $P_\Lambda\Omega_\Lambda \neq 0$.

This theorem is the desired vacuum overlap result.

We also have a Matthews–Salam–Seiler formula in the ultraviolet limit $N = \infty$.

Theorem 3.10: Let $\sigma \in \Lambda^n$, $\gamma \in \Lambda^r$, $t > 0$, $-\frac{1}{2}t \leqslant t_1,\dots,t_r \leqslant \frac{1}{2}t$, $f_1,\dots,f_n, g_1,\dots,g_n \in C_0^\infty(T^1\times\mathbf{R}) \oplus C_0^\infty(T^1\times\mathbf{R})$, and for arbitrary $f \in C_0^\infty(T^1\times\mathbf{R}) \oplus C_0^\infty(T^1\times\mathbf{R})$ set

$$\Psi^{(m)}(f) = \int_{-\infty}^{\infty} ds \int_{-\pi}^{\pi} dx \sum_{\alpha=0}^{1} f_\alpha(x,s) e^{-sH_\Lambda}\psi_\alpha^{(m)}(x) e^{sH_\Lambda}, \tag{3.8}$$

$$\overline{\Psi^{(m)}}(f) = \int_{-\infty}^{\infty} ds \int_{-\pi}^{\pi} dx \sum_{\alpha=0}^{1} f_\alpha(x,s) e^{-sH_\Lambda}\, \overline{\psi_\alpha^{(m)}}(x) e^{sH_\Lambda}, \tag{3.9}$$

$$\Phi_m(s) = e^{-sH_\Lambda} Q_m e^{sH_\Lambda}. \tag{3.10}$$

Assume that $\{t_1\},\dots,\{t_r\}$, $\pi_2 \operatorname{supp} f_1,\dots,\pi_2 \operatorname{supp} f_n, \pi_2 \operatorname{supp} g_1,\dots,\pi_2 \operatorname{supp} g_n$ are disjoint and contained in $[-t/2, t/2]$. Then with $\Phi_{m,N}$, $\Psi_N^{(m)}$, $\overline{\Psi_N^{(m)}}$, and c_N given by (2.16), (2.14), (2.15), and (3.6), respectively,

$$\Big(e^{-(t/2)H_{\Lambda,N}} T\Big[\prod_{k=1}^{r} \Phi_{\gamma(k),N}(t_k) \prod_{k=1}^{n} \overline{\Psi_N^{(\sigma(k))}}(f_k) \times \prod_{k=1}^{n} \Psi_N^{(\sigma(k))}(g_k)\Big] e^{-(t/2)H_{\Lambda,N}}\Omega_\Lambda,\Omega_\Lambda\Big) \tag{3.11}$$

converges to

$$\Big(e^{-(t/2)H_\Lambda} T\Big[\prod_{k=1}^{r} \Phi_{\gamma(k)}(t_k) \prod_{k=1}^{n} \overline{\Psi^{(\sigma(k))}}(f_k) \times \prod_{k=1}^{n} \Psi^{(\sigma(k))}(g_k)\Big] e^{-(t/2)H_\Lambda}\Omega_\Lambda,\Omega_\Lambda\Big) \tag{3.12}$$

as $N \to \infty$. Moreover

$$\exp\Big[\frac{\lambda^2}{2}\sum_{m\in\Lambda}\int \operatorname{Tr}(K_{m,N}^2 + K_{m,N}^\dagger K_{m,N}) d\mu\Big] \times \int d\mu(\phi) \prod_{m\in\Lambda}\Big\{\det_2(1-\lambda K_{m,N}) \times \exp\Big(-c_N \int_{-t/2}^{t/2} \phi_m(s)^2 ds\Big)\Big\} \prod_{k=1}^{r}\phi_{\gamma(k)}(t_k) \times \prod_{m\in\Lambda} \det_{j,k\in\sigma^{-1}(m)}(\tilde{S}_{m,N} f_j, g_k) \tag{3.13}$$

converges to

$$\int d\mu(\phi) \prod_{m\in\Lambda}\Big\{\det_3(1-\lambda K_m) \times \exp\Big[-\frac{\lambda^2}{2} : \operatorname{Tr}(K_m^2 + K_m^\dagger K_m):\Big]\Big\} \times \prod_{k=1}^{r}\phi_{\gamma(k)}(t_k) \prod_{m\in\Lambda} \det_{j,k\in\sigma^{-1}(m)}(\tilde{S}_m f_j, g_k) \tag{3.14}$$

as $N\to\infty$, where K_m is given by (3.7),

$$\tilde{S}_m \equiv (1-\lambda K_m)^{-1} S\Gamma, \tag{3.15}$$

$: \ :$ denotes Wick ordering with respect to $d\mu$, and $\det_3$ is the regularized determinant for operators that are class C_3 with respect to $\mathscr{H}_F$. Thus, by (2.17) we have (3.12) = (3.14).

The convergence of (3.11) to (3.12) follows from Theorem 3.6, while the convergence of (3.13) to (3.14) follows essentially from the estimates of Seiler [11] and of Seiler–Simon [12]. (It should be noted that

$$\frac{\lambda^2}{2}\operatorname{Tr}K_{m,N}^\dagger K_{m,N} = c_N \int_{-t/2}^{t/2} \phi_m(s)^2\, ds.)$$

It is worth mentioning that:

Lemma 3.11: $\{(\tilde{S}_{m,N} f_j, g_k)\}_{N=0}^\infty$ converges to $(\tilde{S}_m f_j, g_k)$ pointwise on the Borel set of α-Holder continuous ϕ. Thus $(\tilde{S}_m f_j, g_k)$ is a random variable and

$$|(\tilde{S}_m f_j, g_k)| \leqslant M^{-1}\|f\|_{L^2\oplus L^2}\|g\|_{L^2\oplus L^2} \tag{3.16}$$

$d\mu$-almost-everywhere.

Proof: Clearly

$$\begin{aligned}|(\tilde{S}_{m,N} f_j, g_k) - (\tilde{S}_m f_j, g_k)| &= |(\tilde{S}_{m,N}\Gamma(\nabla_N - \nabla)\tilde{S}_m f_j, g_k)| \\ &\leqslant M^{-1}\|(\nabla_N - \nabla)\tilde{S}_m f_j\|_{L^2\oplus L^2}\|g_k\|_{L^2\oplus L^2}\end{aligned}$$

and

$$\tilde{S}_m = (-\nabla + M - \Gamma\chi\phi_m)^{-1}\Gamma$$

is a bounded operator by the same reasoning as in the proof of Lemma 2.6. Since the range of $\tilde{S}_m$ is the domain of ∇ and $\nabla_N \to \nabla$ strongly on this domain, we have the desired convergence. □

By the dominated convergence theorem it follows that

Corollary 3.12: $\{(\tilde{S}_{m,N} f_j, g_k)\}_{N=0}^\infty$ converges to $(\tilde{S}_m f_j, g_k)$ in $L^p(d\mu)$ for $1 \leqslant p < \infty$.

The point of this corollary is that the $L_p(d\mu)$ convergence of the density

$$\exp\Big[\frac{\lambda^2}{2}\int \operatorname{Tr}(K_{m,N}^2 + K_{m,N}^\dagger K_{m,N})\, d\mu\Big] \times \prod_{m\in\Lambda}\Big\{\det_2(1-\lambda K_{m,N}) \times \exp\Big(-c_N\int_{-t/2}^{t/2} ds\, \phi_m(s)^2\Big)\Big\} \tag{3.17}$$

to the density

$$\prod_{m\in\Lambda}\det_{\text{ren}}(1-\lambda K_m) \equiv \prod_{m\in\Lambda}\Big\{\det_3(1-\lambda K_m) \times \exp\Big[-\frac{1}{2}\lambda^2 : \operatorname{Tr}(K_m^2 + K_m^\dagger K_m):\Big]\Big\} \tag{3.18}$$

immediately implies the convergence of (3.13) to (3.14), so some of the work usually involved in proving a result like Theorem 3.10 could be avoided in our case.

Remark: In view of Corollary 3.12, one might believe that our model is "purely bosonic" in some sense, but this is probably not the case in the infinite volume limit. We cannot prove that $(\widetilde{S}_m f_j, g_k)$ is a random variable in the $t = \infty$ limit. The point is that $d\mu$ is *not* supported by the set of continuous ϕ such that $\phi(m,s) \to 0$ as $s \to \pm\infty$.

Finally, it follows from Theorems 3.9 and 3.10 that:

Corollary 3.13: With the objects and assumptions of Theorem 3.10,

$$\Big(T\Big[\prod_{k=1}^{r}\Phi_{\gamma(k)}(t_k)\prod_{k=1}^{n}\overline{\Psi^{(\sigma(k))}(f_k)}\prod_{k=1}^{n}\Psi^{(\sigma(k))}(g_k)\Big]\widetilde{\Omega}_\Lambda,\widetilde{\Omega}_\Lambda\Big),$$

$$= \lim_{t\to\infty}(e^{-tH_\Lambda}\Omega_\Lambda,\Omega_\Lambda)^{-1}(3.12)$$

$$= \lim_{t\to\infty}\Big[\int\prod_{m\in\Lambda}\det{}_{\rm ren}(1-\lambda K_m)\,d\mu\Big]^{-1}(3.14), \qquad (3.19)$$

where $\widetilde{\Omega}_\Lambda \equiv \|P_\Lambda\Omega_\Lambda\|^{-1}P_\Lambda\Omega_\Lambda$.

Remark: That the partition function

$$Z_{\Lambda;t} \equiv \int\prod_{m\in\Lambda}\det{}_{\rm ren}(1-\lambda K_m)\,d\mu = (e^{-tH_\Lambda}\Omega_\Lambda,\Omega_\Lambda) \qquad (3.20)$$

does not vanish follows from Seiler's reasoning.[11] Equation (3.19) is, of course, the Euclidean Gell-Mann–Low formula,[18] which is a standard consequence of vacuum overlap. It expresses interacting vacuum expectations as thermodynamic quantities involving the free vacuum.

4. OS POSITIVITY, FRÖHLICH BOUNDS, AND NON-RELATIVISTIC WIGHTMAN AXIOMS

We now have the finite-volume noncoincident (with respect to imaginary time) Schwinger functions (Euclidean Green's functions) for our model, and we denote them in the following way:

Clearly, all of the results of the previous sections are preserved with respect to the replacement $-t/2 \to -t', t/2 \to t, t \to t+t'$. With the objects and assumptions of Theorem 3.10 and with this minor generalization in effect, we set

$$S_{\Lambda;-t',t}(f,g,\sigma;\tau,\gamma) \equiv Z_{\Lambda;t+t'}^{-1}(3.12), \qquad (4.1)$$

where f,g,τ denote the strings $(f_1,\ldots,f_n)$, $(g_1,\ldots,g_n)$, $(t_1,\ldots,t_r)$, respectively, and $\gamma(k)(\sigma(k))$ is the lattice site associated with t_k (f_k and g_k). It follows from our Matthews–Salam–Seiler formula that

$$S_{\Lambda;-t',t}(f,g,\sigma;\tau,\gamma) = Z_{\Lambda;t+t'}^{-1}(3.14). \qquad (4.2)$$

$S_{\Lambda;-t',t}(f,g,\sigma;\tau,\gamma)$ is the partially smeared Schwinger function for n fermion-antifermion pairs (at σ-selected sites) and r "bosons."

Theorem 4.1: Consider $2l$ strings of test functions:

$F_j \equiv (f_{j1},\ldots,f_{jn_j}), 1 \leqslant j \leqslant l, f_{jk} \in C_0^\infty(T^1\times\mathbf{R}) \oplus C_0^\infty(T^1\times\mathbf{R})$,

$G_j \equiv (g_{j1},\ldots g_{jn_j}), 1 \leqslant j \leqslant l, g_{jk} \in C_0^\infty(T^1\times\mathbf{R}) \oplus C_0^\infty(T^1\times\mathbf{R})$,

and l strings of real numbers:

$\tau_j \equiv (t_{j1},\ldots,t_{jr_j}), 1 \leqslant j \leqslant l,$

where the sets $\pi_2\,\mathrm{supp} f_{jk}, \pi_2\,\mathrm{supp}\, g_{jk}, \{t_{jk}\}$ are disjoint and contained in the interval $[-t',t]$. Let $\Lambda_1,\ldots,\Lambda_l$ be rectangular subsets of $\mathbf{Z}^\nu$ containing the origin differing only in a given direction, and denote the common coordinate-zero subrectangle by Λ'. Let

$$\sigma_j:\{1,\ldots,n_j\} \to \Lambda_j^+, \gamma_j:\{1,\ldots,r_j\} \to \Lambda_j^+,$$

where Λ_j^+ denotes the strictly positive part of Λ_j with respect to the given direction. Let θ be the reflection of $\mathbf{Z}^\nu$ through the plane that separates Λ' from the lattice sites that are positive with respect to that direction. Then

$$\sum_{l,j=1}^{l} Z_{\Lambda_l^+\cup\theta\Lambda_j^+;-t',t}S_{\Lambda_l^+\cup\theta\Lambda_j^+;-t',t}$$
$$(F_l \vee \overline{F}_j, G_l \vee \overline{G}_j, \sigma_l \vee \theta\sigma_j; \tau_l \vee \tau_j, \gamma_l \vee \theta\gamma_j) \geqslant 0,$$

where $\vee$ denotes adjunction of strings.

This theorem is essentially the statement that our model satisfies OS positivity[3,6,14] in the spatial directions, and the proof follows easily from (4.2). The point is that the only part of the interaction that couples lattice sites is $d\mu$, which is certainly OS positive.[31,32]

Remark: If $\Lambda_1 = \cdots = \Lambda_l \equiv \Lambda$, then $\Lambda_l^+ \cup \theta\Lambda_j^+ = \Lambda^+ \cup \theta\Lambda^+$ and we have a more familiar statement of OS positivity. We have stated this slight generalization because such a version is implicitly used in the Seiler–Simon derivation[14] of Fröhlich bounds and we intend to appeal to their reasoning.

Theorem 4.2. With the same objects that were introduced in Theorem 4.1, except that θ is no longer a lattice reflection but the reflection of $\mathbf{R}$, assume that

$\{t_{j1}\},\ldots,\{t_{jr_j}\}, \pi_2\,\mathrm{supp} f_{j1},\ldots,\pi_2\,\mathrm{supp} f_{jn_j},$

$\pi_2\,\mathrm{supp} g_{j1},\ldots,\pi_2\,\mathrm{supp} g_{nj_j}$

$\subset [0,t_j],$

where $t_1,\ldots,t_l > 0$. Let $\Lambda_1 = \cdots = \Lambda_l \equiv \Lambda$ and require only that the ranges of σ_j,γ_j lie in Λ. Then

$$\sum_{l,j=1}^{l} Z_{\Lambda;-t_l,t_j}S_{\Lambda;-t_l,t_j}(F_l \vee \overline{\theta F_j},$$
$$G_l \vee \overline{\theta G_j}, \sigma_l \vee \sigma_j;$$
$$\tau_l \vee -\tau_j, \gamma_1 \vee \gamma_j) \geqslant 0,$$

where $\theta F_j \equiv (\theta f_{j1},\ldots,\theta f_{jn_j})$ and $(\theta f_{jk})(x,s) \equiv f_{jk}(x,-s)$. This result states that our Schwinger functions satisfy OS positivity in the imaginary time direction, and it follows from elementary manipulations with the expression that (4.1) yields for such a combination.

A consequence of these OS positivity results is that the Euclidean pressure exists.

Theorem 4.3. $\alpha_\infty \equiv \lim_{t\to\infty, \Lambda\to\infty}(1/t|\Lambda|)\ln Z_{\Lambda;t}$ exists if $\Lambda \to \infty$ through rectangles.

The point is that $\ln Z_{\Lambda;t}$ is convex in t and in each dimension variable of the rectangle Λ by virtue of OS positivity. Thus, the theorem follows immediately from the general convexity lemmas in Ref. 14 (Lemmas 6.3 and 6.4) and the linear upper bound on $\ln Z_{\Lambda;t}$ furnished by Theorem 3.5.

Theorem 4.4. With the objects and assumptions of

Theorem 3.10, there are constants $C_1, C_2, C_3 > 0$ dependent only on the model parameters, such that

$$Z_{\Lambda;t'+t}|S_{\Lambda;-t',t}(f,g,\sigma;\tau,\gamma)|$$
$$\leqslant C_1^{|\Lambda|(t+t')} C_2^n C_3^r \prod_{m,\ell} (r_{m\ell}!)^{1/2}$$
$$\times \prod_{k=1}^{n} (\|f_k\|_{-1/2}\|g_k\|_{-1/2}), \tag{4.3}$$

where $r_{m\ell}$ denotes the number of k's such that $\ell \leqslant t_k < \ell + 1$ and $\gamma(k) = m$, and

$$\|f\|^2_{-\beta} \equiv \int_{-\infty}^{\infty} d\omega \sum_{\ell=-\infty}^{\infty} (\ell^2 + \omega^2 + M^2)^{-\beta}$$
$$\times \sum_{\alpha=0}^{1} |\hat{f}_\alpha(\ell,\omega)|^2. \tag{4.4}$$

This result is just the adaptation to our model of the volume-divergent bounds obtained by McBryan[33] and by Seiler–Simon[13] for the Y_2 model. Given the checkerboard estimates for the lattice directions (briefly discussed in Sec. 2), the proof of this theorem may be patterned after either the McBryan proof or the Seiler–Simon proof. It should be noted that although the lattice checkerboard estimates essentially reduce the problem to that of a single lattice site (because the fermions do not couple lattice sites) the actual work is no easier than in the case of the Y_2 model. The Euclidean boson field at a given lattice site depends only on imaginary time, but the Euclidean fermi propagator is still two-dimensional (albeit periodic in one direction) and the renormalization problems are unchanged.

Theorem 4.5: With the objects and assumptions of Theorem 4.4, let $\{t_k\}$, $\pi_2 \operatorname{supp} f_k$, $\pi_2 \operatorname{supp} g_k$ lie in the interval $[-t_0/2, t_0/2]$ and assume that the ranges of σ and γ lie in the box Λ_0. There is a constant $\widetilde{C}_1 > 0$, dependent only on the model parameters, such that

$$\overline{\lim_{t\to\infty}} |S_{\Lambda;-t/2,t/2}(f,g,\sigma;\tau,\gamma)| \leqslant \widetilde{C}_1^{|\Lambda_0|t_0} C_2^n C_3^r$$
$$\times \prod_{m,\ell} (r_{m\ell}!)^{1/2} \prod_{k=1}^{n} (\|f_k\|_{-1/2}\|g_k\|_{-1/2}), \tag{4.5}$$

where C_2, C_3 are the constants C_2, C_3 of Theorem 4.4.

These bounds are Fröhlich bounds.[14,24] As in the proof of such bounds for the Y_2 model,[14] the basic ingredients of the proof are OS positivity in both spatial and imaginary time directions, volume divergent bounds like (4.3), and control over the energy difference between volumes. At first glance it might appear that we are in trouble because Seiler and Simon have controlled this energy difference by an argument that seems to depend on Nelson's symmetry. However, the convexity lemmas (Lemmas 6.3 and 6.4 in Ref. 14) give us control over such a difference, and indeed Seiler and Simon argue precisely this way in the proof of a later theorem.

Equation (4.5) is not enough to prove that the noncoincident (with respect to imaginary time) distributions one obtains for the infinite volume theory are tempered. The usual argument involving a partition of unity is ruined by the factor $\widetilde{C}_1^{t_0|\Lambda_0|}$ if $\widetilde{C}_1 > 1$. In the Y_2 case Seiler and Simon show that the constant $\widetilde{C}_1 = 1$ works for the Fröhlich bounds in the $\Lambda = \mathbf{Z}^\nu$ limit by showing that the constant $C_1 = e^{\alpha_\infty}$ works for the volume-divergent bounds. However, their proof depends on the Euclidean Reeh–Schlieder theorem, so we cannot duplicate this refinement for our model.

Although we cannot prove that our infinite volume Schwinger functions will be tempered for *arbitrary* coupling strength λ, we can certainly prove both uniqueness and temperedness for the *weakly* coupled model.

Theorem 4.6: If $\omega_0^{-1}|\lambda|$ and $M^{-1}|\lambda|$ are sufficiently small, then with the objects and assumptions of Theorem 4.5 we have

(a) $\lim_{\substack{t\to\infty \\ \Lambda\to\infty}} S_{\Lambda;-t/2,t/2}(f,g,\sigma;\tau,\gamma) \equiv S(f,g,\sigma;\tau,\gamma)$ exists.

(b) There are constants $c_1, c_2 > 0$, dependent only on the model parameters such that

$$|S_{\Lambda;-t/2,t/2}(f,g,\sigma;\tau,\gamma)|$$
$$\leqslant c_1^n c_2^r \prod_{m,\ell} (r_{m\ell}!)^{1/2} \prod_{k=1}^{n} (\|f_k\|_{-1/2}\|g_k\|_{-1/2}).$$

(c) Let $\tilde{\sigma}:\{1,...,n\}\to\mathbf{Z}^\nu$, $\tilde{\gamma}:\{1,...,r\}\to\mathbf{Z}^\nu$, $\tilde{t}_1,...,\tilde{t}_{\tilde{r}} \in R$, and $\tilde{f}_1,...,\tilde{f}_{\tilde{n}}, \tilde{g}_1,...,\tilde{g}_{\tilde{n}} \in C_0^\infty(T^1\times\mathbf{R}) \oplus C_0^\infty(T^1\times\mathbf{R})$; then $|S(f \vee \tilde{f}^t, g \vee \tilde{g}^t \sigma \vee \tilde{\sigma}; \tau \vee \tilde{\tau}_t, \gamma \vee \tilde{\gamma}) - S(f,g,\sigma;\tau,\gamma) S(\tilde{f},\tilde{g},\tilde{\sigma};\tilde{\tau},\tilde{\gamma})| \leqslant ce^{-ct}$ for all $t > 0$, where $\tilde{f}^t_k(x,s) \equiv \tilde{f}_k(x,s-t)$ and $\tilde{\tau}_t \equiv (\tilde{t}_1 - t,...,\tilde{t}_{\tilde{r}} - t)$.

(d) If ϵ is a unit coordinate vector for $\mathbf{Z}^\nu$,

$$|S(f \vee \tilde{f}, g \vee \tilde{g}, \sigma \vee \tilde{\sigma}_j; \tau \vee \tilde{\tau}, \gamma \vee \tilde{\gamma}_j)$$
$$- S(f,g,\sigma;\tau,\gamma) S(\tilde{f},\tilde{g},\tilde{\sigma},\tilde{\tau},\tilde{\gamma})|$$
$$\leqslant ce^{-cj} \text{ for all } j \in \mathbf{Z}^+,$$

where $\tilde{\sigma}_j(k) \equiv \tilde{\sigma}(k) - \epsilon j$.

This theorem can be proven by an adaptation of the Y_2 cluster expansion[21,22] to our model. We omit the proof because it differs in no significant way from Refs. 21 and 22. Obviously the expansion would be modified in the spatial directions by the fact that the fermions are already decoupled and the bosonic part of the model is lattice-like with a nearest-neighbor coupling. Part (b) of the theorem, which implies that our infinite volume Schwinger functions are tempered, follows from estimates on the cluster expansion—at least if we pattern our expansion after Magnen and Sénéor.[22] Unlike Magnen and Sénéor, Cooper and Rosen[21] appealed to the Seiler–Simon and McBryan results instead of explicitly deriving temperedness estimates from their version of the Y_2 cluster expansion, but such estimates should follow from a detailed inspection of their convergence proof.

Corollary 4.7: If $\omega_0^{-1}|\lambda|$ and $M^{-1}|\lambda|$ are sufficiently small, then the infinite volume Schwinger functions satisfy all of the unEuclidean OS axioms including exponential clustering.

Combining parts (a) and (b) of Theorem 4.6 with Corollary 3.13, the general theory of Laplace transforms, and standard Vitali arguments,[3,25] we obtain Theorem 1.1. In view of Corollary 4.7 and Osterwalder–Schrader reconstruction, we see that Theorem 1.2 immediately follows.

5. FKG INEQUALITY: SUFFICIENT CONDITION

The free boson measure that arises in our model is given by the covariance (1.20). It is evident from the formula that a boson lattice approximation of the model will involve imaginary time only. Thus, we set

$$q_j(m) = \int_{-\infty}^{\infty} ds\, f_{\delta,j}(m,s) \quad \phi(m,s), \quad \delta > 0, \quad j\in\mathbf{Z},$$

$$f_{\delta,j}(m,s) = \frac{1}{(2\pi)^{\nu+1}} \int_{T^\nu} d^\nu k \times \int_{-\pi\delta^{-1}}^{\pi\delta^{-1}} d\omega \frac{e^{ik\cdot m + i\omega(s - j\delta)}(\omega(k)^2 + \omega^2)^{1/2}}{(2\delta^{-2}(1 - \cos\delta\omega) + \omega(k)^2)^{1/2}},$$

where $[s]$ denotes the element of $\delta\,\mathbf{Z}$ closest to s (with, say, the left-hand convention if s lies midway between two successive members of $\delta\,\mathbf{Z}$). This lattice approximation is completely analogous to the lattice approximation of the free boson measure for the standard quantum field theories.[1,3,18] It is easy to see that

$$\int q_j(m) q_{j'}(m') d\mu = \frac{1}{(2\pi)^{\nu+1}} \int_{T^\nu} d^\nu k \int_{-\pi\delta^{-1}}^{\pi\delta^{-1}} d\omega \times \frac{e^{ik\cdot(m - m') + i\omega(j - j')}}{2\delta^{-2}(1 - \cos\delta\omega) + \omega(k)^2},$$

so we have the expected covariance for the new random variables.

Now our strategy is to prove the FKG inequality for our model with both the fermion ultraviolet cutoff and this boson lattice cutoff in effect. In Sec. 3 we saw how the former cutoff can be removed, so our regularization problem is reduced to showing only that the lattice cutoff can be removed *with the fermion ultraviolet cutoff in effect*. The FKG inequality would obviously be preserved under successive removal of the cutoffs.

Since the FKG inequality is a bosonic result, we may concentrate on the Euclidean measure, which describes the boson subtheory of our model. Fix a fermion ultraviolet cutoff N and consider the corresponding interaction density:

$$\rho(\phi) \equiv \prod_{m\in\Lambda} \{\det_2(1 - \lambda S_N \Gamma\chi\phi_m) e^{-c_N \int_{-t/2}^{t/2} ds\, \phi(m,s)^2}\},$$

where the objects are those introduced in Sec. 2 and we are suppressing the dependence of the interaction density on N, Λ, and t. The problem is to show that if we define the interaction density for lattice cutoff δ as

$$\rho^\delta(\phi) \equiv \prod_{m\in\Lambda} \{\det_2(1 - \lambda S_N \Gamma\chi\phi^\delta_m) e^{-c_N \int_{-t/2}^{t/2} ds\, \phi^\delta(m,s)^2}\},$$

then there is a sequence $\{\delta_k\}$ such that $\delta_k \downarrow 0$ and $\rho^{\delta_k} \to \rho$ in $L^p(d\mu)$ for $1 \leqslant p < \infty$. Clearly, the following result suffices:

Theorem 5.1: For an arbitrary time cutoff t, fermion ultraviolet cutoff N, and lattice site $m\in\mathbf{Z}^\nu$, there is a sequence $\{\delta_k\}$ such that $\delta_k \downarrow 0$ and

$$e^{-c_N \int_{-t/2}^{t/2} ds\, \phi^{\delta_k}(m,s)^2} \to e^{-c_N \int_{-t/2}^{t/2} ds\, \phi(m,s)^2},$$

$$\det_2(1 - \lambda S_N \Gamma\chi\phi^{\delta_k}_m) \to \det_2(1 - \lambda S_N \Gamma\chi\phi_m)$$

in $L^p(d\mu)$ for $1 \leqslant p < \infty$.

We omit the proof of this theorem because the basic strategy is the same as that used in Ref. 1 for the lattice approximation of Y_2. Indeed, the proof of Theorem 5.1 is simpler because we are working with a fixed cutoff on Fermi momenta. Without this cutoff, each factor in the interaction density would be either zero or infinity and we would have to control the cancellation. As it is, each factor can be handled separately, and the momentum integrals to be estimated[1] in this case are essentially one-dimensional[34] and contain no ultraviolet singularities that might complicate the derivation of δ-independent bounds.

With this double approximation in effect, let $2J + 1$ be the (odd) number of random variables that we are effectively considering at a given lattice site $m\in\Lambda$ [i.e., $(2J + 1)\delta = t$; the variables themselves are labelled $q_{-J}(m),\ldots,q_0(m),\ldots,q_J(m)$]. The approximate measure is the measure $d\nu$ on $\mathbf{R}^{|\Lambda|(2J+1)}$ given by

$$d\nu(q) = \prod_{m\in\Lambda} \Big\{\det_2\Big(1 - \lambda S_N \Gamma \sum_{j=-J}^{J} \chi_j q_j(m)\Big) \times e^{-c_N \delta \Sigma_{j=-J}^{J} q_j(m)^2}\Big\} e^{-(Aq,q)} \prod_{m,j} dq_j(m),$$

where χ_j is the characteristic function for $[(j - \frac{1}{2})\delta, (j + \frac{1}{2})\delta]$, $(\ ,\ )$ denotes the inner product on $\mathbf{R}^{|\Lambda|(2J+1)}$, and A is the $|\Lambda|(2J + 1)\times|\Lambda|(2J + 1)$ matrix defined by

$$A = -\Delta_\delta + \omega_0^2, \tag{5.1}$$

where Δ_δ is the lattice analog of the $(\nu + 1)$-dimensional Laplacian[18] (with δ-spacing in the imaginary time variable and 1-spacing in the space variables) with free boundary conditions on $\Lambda \times \delta\{-J,\ldots,0,\ldots,J\}$. This section and the next will be devoted to proving the FKG inequality for $d\nu$. We must prove

Theorem 5.2: Let f,g be increasing functions on $\mathbf{R}^{|\Lambda|(2J+1)}$ (in the sense that $f(q) \leqslant f(q')$ if $q_j \leqslant q'_j$ for all j) such that f,g,fg are $d\nu$-integrable. Then

$$\langle fg\rangle - \langle f\rangle\langle g\rangle \geqslant 0,$$

where $\langle\ \rangle$ denotes the normalized expectation with respect to $d\nu$.

Remark: By the same argument as in Refs. 1 and 18 (except that our fields ϕ^δ, ϕ are not smeared), Theorem 1.4 follows from Theorem 5.2 via the successive limits $\delta\to 0, N\to\infty$. As in Ref. 1, our approach to proving Theorem 5.2 lies in Theorem 1.1 of Ref. 1, which we state here for convenience:

Theorem 5.3: Let $W\in C^2(\mathbf{R}^n)$ such that

$$\partial^2 W/\partial q_j \partial q_k \geqslant 0, \quad j\neq k.$$

If f,g are increasing functions on $\mathbf{R}^n$ such that f,g,fg are $e^{W(q)}d^n q$-integrable, then

$$\langle fg\rangle - \langle f\rangle\langle g\rangle \geqslant 0,$$

where $\langle\ \rangle$ denotes the normalized expectation with respect to $e^{W(q)}d^n q$.

In order to show that $d\nu$ can even be written in the form suitable for this theorem, we must show that $\det_2(1 - \lambda S_N \Gamma \Sigma_{j=-J}^{J} q_j(m)\chi_j)$ cannot vanish for any $q\in\mathbf{R}^{|\Lambda|(2J+1)}$, $m\in\Lambda$. Indeed we have

Theorem 5.4: $1 - \lambda S_N \Gamma \Sigma_{j=-J}^{J} q_j(m)\chi_j$ is invertible on $\mathscr{H}_F$ for all $m\in\Lambda$, $q\in\mathbf{R}^{|\Lambda|(2J+1)}$.

Proof: Let

$$\mathscr{H}_F^{(n)} = \Big\{v\in L^2(\mathbf{R}) \oplus L^2(\mathbf{R}) \,\Big|\, \|v\|^2_{(n)} \equiv \sum_{\alpha=0}^{1} \int_{-\infty}^{\infty} d\omega(\omega^2 + n^2 + M_F^2)^{1/2} |\hat v_\alpha(\omega)|^2 < \infty\Big\}$$

and $S^{(n)}$ be the bounded operator on $\mathscr{H}_F^{(n)}$ defined by the kernel

$$S^{(n)}(s,s') = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{i\omega(s-s')}\frac{i\beta_0\omega + i\beta_1 n + M}{\omega^2 + n^2 + M^2}. \quad (5.2)$$

Then

$$\mathscr{H}_F \cong \bigoplus_{n=-\infty}^{\infty} \mathscr{H}_F^{(n)},$$

$$S_N \cong \bigoplus_{n=-N}^{N} S^{(n)}$$

under the obvious isomorphism. Since the characteristic functions χ_j have no y-dependence, we also have

$$S_N\Gamma \sum_{j=-J}^{J} q_j(m)\chi_j \cong \bigoplus_{n=-N}^{N} S^{(n)}\Gamma \sum_{j=-J}^{J} q_j(m)\chi_j,$$

so it is enough to show that $1-\lambda S^{(n)}\Gamma\Sigma_{j=-J}^{J} q_j(m)\chi_j$ is invertible on $\mathscr{H}_F^{(n)}$. Now by the same calculation mentioned in Sec. 2 we see that $S^{(n)}\Gamma\Sigma_{j=-J}^{J} q_j(m)\chi_j$ is a Hilbert–Schmidt operator on $\mathscr{H}_F^{(n)}$, so by the Fredholm alternative it suffices to prove that $1-\lambda S^{(n)}\Gamma\Sigma_{j=-J}^{J} q_j(m)\chi_j$ has zero kernel. Since Γ is anti-Hermitian, the proof is similar to an argument of Seiler.[11] □

It is clear from our proof of this theorem that $\det_2(1-\lambda S_N\Gamma\Sigma_{j=-J}^{J} q_j(m)\chi_j)$ splits into the product $\Pi_{n=-N}^{N}\det_2^{(n)}(1-\lambda S^{(n)}\Gamma \sum_{j=-J}^{J} q_j(m)\chi_j)$, where $\det_2^{(n)}$ is the restriction of $\det_2$ to $\mathscr{H}_F^{(n)}$. Since these factors are nonzero and real (hence positive by continuity in λ), we may define

$$W(q) = \sum_{m\in\Lambda}\left\{\sum_{n=-N}^{N}\ln\det{}_2^{(n)}\left(1-\lambda S^{(n)}\Gamma\sum_{j=-J}^{J} q_j(m)\chi_j\right) - c_N\delta\sum_{j=-J}^{J} q_j(m)^2\right\} - (Aq,q),$$

so that $d\nu(q) = e^{W(q)}\Pi_{m,j}dq_j(m)$. We also have:

Lemma 5.5: $W_{m,n}(q)\equiv\ln\det_2^{(n)}(1-\lambda S^{(n)}\Gamma \times\Sigma_{j=-J}^{J} q_j(m)\chi_j)$ lies in $C^1(\mathbf{R}^{2J+1})$ and for $-J\leqslant k\leqslant J$,

$$\frac{\partial}{\partial q_k(m)}W_{m,n} = -\lambda\,\mathrm{Tr}^{(n)}\left(1-\lambda S^{(n)}\Gamma\sum_{j=-J}^{J} q_j(m)\chi_j\right)^{-1}\times S^{(n)}\Gamma\chi_k,$$

where $\mathrm{Tr}^{(n)}$ denotes an improper trace on $\mathscr{H}_F^{(n)}$ for which $\mathrm{Tr}^{(n)}S^{(n)}\Gamma\chi_k = 0$.

Proof: For q in a fixed compact set and for sufficiently small λ, a standard computation based on the Fredholm expansion yields

$$W_{m,n}(q) = -\sum_{\ell=2}^{\infty}\frac{\lambda^{\ell}}{\ell}\mathrm{Tr}^{(n)}\left[S^{(n)}\Gamma\sum_{j=-J}^{J} q_j(m)\chi_j\right]^{\ell},$$

where $\mathrm{Tr}^{(n)}$ is well-defined for each term because $S^{(n)}\Gamma\Sigma_{j=-J}^{J} q_j(m)\chi_j$ is Hilbert–Schmidt. Partial differentiation yields

$$\frac{\partial}{\partial q_k(m)}W_{m,n} = -\sum_{\ell=2}^{\infty}\lambda^{\ell}\mathrm{Tr}^{(n)}\left[S^{(n)}\Gamma\sum_{j=-J}^{J} q_j(m)\chi_j\right]^{\ell-1}S^{(n)}\Gamma\chi_k,$$

by cyclicity of the trace. But the rhs is just the power series expansion for

$$-\lambda\,\mathrm{Tr}^{(n)}\left[\left(1-\lambda S^{(n)}\Gamma\sum_{j=-J}^{J} q_j(m)\chi_j\right)^{-1}\times S^{(n)}\Gamma\chi_k - S^{(n)}\Gamma\chi_k\right].$$

By analyticity in a neighborhood of the real axis, the resulting equation holds for all real λ. Finally, although $S^{(n)}\Gamma\chi_k$ is not trace-class, the properties of the Dirac matrices together with (5.2) imply that there is a principal value of $\mathrm{Tr}^{(n)}$ for which $\mathrm{Tr}^{(n)}S^{(n)}\Gamma\chi_k = 0$ (in fact, any principal value whose cutoff preserves the matrix trace will do). □

It immediately follows that

Corollary 5.6: $W_{m,n}\in C^2(\mathbf{R}^{2J+1})$, and for $-J\leqslant k,k'\leqslant J$,

$$\partial^2 W_{m,n}/\partial q_k(m)\partial q_{k'}(m) = -\lambda^2\mathrm{Tr}^{(n)}\tilde{S}_m^{(n)}\chi_k\tilde{S}_m^{(n)}\chi_{k'}, \quad (5.3)$$

where

$$\tilde{S}_m^{(n)} \equiv \left(1-\lambda S^{(n)}\Gamma\sum_{j=-J}^{J} q_j(m)\chi_j\right)^{-1}S^{(n)}\Gamma,$$
$$= (1-\lambda S^{(n)}\Gamma\chi\phi_m^{\delta})^{-1}S^{(n)}\Gamma. \quad (5.4)$$

Of course, $\mathrm{Tr}^{(n)}$ is well-defined here. Collecting everything, we get:

Theorem 5.7: $W\in C^2(\mathbf{R}^{|\Lambda|(2J+1)})$, and for $-J\leqslant k,k'\leqslant J$, $l,l'\in\Lambda$,

$$\partial^2 W/\partial q_k(l)\partial q_{k'}(l') = -A_{(k,l)(k',l')}, \quad l'\neq l, \quad (5.5)$$

and

$$\frac{\partial^2 W}{\partial q_k(l)\partial q_{k'}(l)} = -A_{(k,l)(k',l)} - \lambda^2\sum_{n=-N}^{N}\mathrm{Tr}^{(n)}\tilde{S}_l^{(n)}\chi_k\tilde{S}_l^{(n)}\chi_{k'}. \quad (5.6)$$

Our aim is to show that these second partial derivatives are nonnegative for $(k,l)\neq(k',l')$. Now since the off-diagonal elements of the finite-difference Laplacian Δ^δ are positive or zero, it follows from (5.1) that the off-diagonal elements of A are negative or zero. Therefore, it suffices to show that

$$\mathrm{Tr}^{(n)}\tilde{S}_m^{(n)}\chi_k\tilde{S}_m^{(n)}\chi_{k'}\leqslant 0, \quad k\neq k'. \quad (5.7)$$

From now on we let Tr denote the trace on the Hilbert space $L^2(\mathbf{R})\oplus L^2(\mathbf{R})$. The arguments of the next section will show that $1-\lambda S^{(n)}\Gamma\Sigma_{j=-J}^{J} q_j(m)\chi_j$ is invertible on $L^2(\mathbf{R})\oplus L^2(\mathbf{R})$ and that $\tilde{S}_m^{(n)}$ is Hilbert–Schmidt on this space. Hence $\mathrm{Tr}\,\tilde{S}_m^{(n)}\chi_k\tilde{S}_m^{(n)}\chi_{k'}$ is defined and

$$\mathrm{Tr}^{(n)}\tilde{S}_m^{(n)}\chi_k\tilde{S}_m^{(n)}\chi_{k'} = \mathrm{Tr}\,\tilde{S}_m^{(n)}\chi_k\tilde{S}_m^{(n)}\chi_{k'}.$$

Thus, our object in the next section is to show that

$$\mathrm{Tr}\,\tilde{S}_m^{(n)}\chi_k\tilde{S}_m^{(n)}\chi_{k'}\leqslant 0, \quad k\neq k'. \quad (5.8)$$

6. FKG INEQUALITY: PROOF

Consider the more general situation:

$$S^{(n)}(s,s') = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\,\frac{i\beta_0\omega + i\not{n} + M}{\omega^2 + n^2 + M^2}e^{i\omega(s-s')}, \quad n\in\mathbf{Z}^\kappa, \quad (6.1)$$

where

$$\not{n} = \sum_{\ell=1}^{\kappa} n_\ell\beta_\ell$$

and the matrices $\beta_0,\ldots,\beta_\kappa$, Γ anticommute and satisfy the properties

$$\beta_\ell^* = \beta_\ell, \quad \beta_\ell^2 = 1,$$

$$\Gamma^* = -\Gamma, \quad \Gamma^2 = -1,$$

where * denotes matrix adjoint. The standard real-valued irreducible representation of this collection of Dirac matrices is $d_\kappa \equiv 2^{[(\kappa+2)/2]}$ dimensional, where [ ] denotes the least-integer function.

By implication we are considering the same basic model, where the Fermi system at each lattice site is now a κ-dimensional box in which the time-zero Fermi fields appropriate to that dimension satisfy periodic boundary conditions.

Throughout this section we replace $\chi\phi_m^\delta$ with an arbitrary bounded measurable, real-valued function h with compact support and the Hilbert space we use is

$$\mathscr{H}_\kappa \equiv \bigoplus_{\ell=1}^{d_\kappa} L^2(\mathbf{R}).$$

Theorem 6.1: $1 - S^{(n)}\Gamma h$ is invertible,

$$\widetilde{S}^{(n)} \equiv (1 - S^{(n)}\Gamma h)^{-1} S^{(n)}\Gamma = \left(-\beta_0 \frac{d}{ds} - i n\!\!\!/ + M - \Gamma h\right)^{-1}, \tag{6.2}$$

and $\widetilde{S}^{(n)}\chi$ is Hilbert–Schmidt for the characteristic function χ of any compact set.

Proof: By computation we have the explicit formula

$$S^{(n)}(s,s') = \text{const}\left((n^2+M^2)^{1/2}\beta_0 \frac{s'-s}{|s-s'|} + i n\!\!\!/ + M\right) \times e^{-(n^2+M^2)^{1/2}|s-s'|}, \tag{6.3}$$

so $S^{(n)}(s,s')\Gamma h(s')$ is L^2 on $\mathbf{R}\times\mathbf{R}$. Thus $S^{(n)}\Gamma h$ is Hilbert–Schmidt on $\mathscr{H}_\kappa$, so by the Fredholm alternative, the first claim is established if we can show that $1 - S^{(n)}\Gamma h$ has dense range. But

$$1 - S^{(n)}\Gamma h = S^{(n)}\left(-\beta_0 \frac{d}{ds} - i n\!\!\!/ + M - \Gamma h\right)$$

on the domain of self-adjointness for $i(d/ds)$. Since $i(-\beta_0(d/ds) - i n\!\!\!/ - \Gamma h)$ is self-adjoint on this domain and $M \neq 0$, we know that the unbounded operator $-\beta_0(d/ds) - i n\!\!\!/ + M - \Gamma h$ is onto.[5] $S^{(n)}$ is a bounded operator with dense range, so we have verified the first claim. The second claim is algebraically evident, and the third claim follows from the first claim together with the fact that $S^{(n)}\chi$ is Hilbert–Schmidt for every such χ. □

Our goal is to prove that $\operatorname{Tr} \widetilde{S}^{(n)}\chi_k \widetilde{S}^{(n)}\chi_{k'}$ is real and that

$$\operatorname{Tr} \widetilde{S}^{(n)}\chi_k \widetilde{S}^{(n)}\chi_{k'} \leqslant 0, \quad k \neq k',$$

where $\chi_k, \chi_{k'}$ are the characteristic functions of the preceding section and the trace is taken with respect to $\mathscr{H}_\kappa$. The quantity is well-defined by virtue of the preceding theorem; indeed, the fundamental theorem[28] for Hilbert–Schmidt operators on L^2 spaces yields:

Corollary 6.2: $\widetilde{S}^{(n)}$ has a measurable kernel $\widetilde{S}^{(n)}_{\ell,\ell'}(s,s')$ that is L^2 on $\mathbf{R}\times K$ for every compact $K \subset \mathbf{R}$, and

$$\operatorname{Tr} \widetilde{S}^{(n)}\chi_k \widetilde{S}^{(n)}\chi_{k'} = \int ds \int ds'\, \chi_k(s')\chi_{k'}(s) \operatorname{tr} \widetilde{S}^{(n)}(s,s')\widetilde{S}^{(n)}(s's),$$

where tr denotes matrix trace.

Therefore it is enough to show that

$$\operatorname{tr} \widetilde{S}^{(n)}(s,s')\widetilde{S}^{(n)}(s',s) \leqslant 0$$

for almost all $(s,s') \in \mathbf{R}\times\mathbf{R}$.

We will be concerned with symmetries that will involve the dependence of $\widetilde{S}^{(n)}$ on h and M, so we set

$$S_M^{(n)} \equiv S^{(n)},$$
$$\widetilde{S}_{M,h}^{(n)} \equiv \widetilde{S}^{(n)}.$$

Now notice that since

$$\widetilde{S}_{M,h}^{(n)} = (1 - S_M^{(n)}\Gamma h)^{-1} S_M^{(n)}\Gamma = S_M^{(n)}\Gamma(1 - hS_M^{(n)}\Gamma)^{-1},$$

we have

$$\widetilde{S}_{M,h}^{(n)\dagger} = -(1 + \Gamma S_M^{(n)\dagger} h)^{-1}\Gamma S_M^{(n)\dagger},$$

where † denotes operator adjoint on $\mathscr{H}_\kappa$. Since $A^\dagger(s,s') = A(s',s)^*$ almost everywhere for any bounded operator A that is densely defined by, say, a locally integrable kernel $A_{\ell\ell'}(s,s')$, it follows from (6.1) that

$$\Gamma S_M^{(n)\dagger} = S_M^{(n)}\Gamma,$$

so

$$\widetilde{S}_{M,h}^{(n)\dagger} = -(1 + S_M^{(n)}\Gamma h)^{-1} S_M^{(n)}\Gamma = -\widetilde{S}_{M,-h}^{(n)}.$$

Hence

$$\widetilde{S}_{M,h}^{(n)}(s',s)^* = -\widetilde{S}_{M,-h}^{(n)}(s,s'), \quad \text{a.a. } (s,s') \in \mathbf{R}\times\mathbf{R},$$

so the desired inequality becomes

$$\operatorname{tr} \widetilde{S}_{M,h}^{(n)}(s,s')\widetilde{S}_{M,-h}^{(n)}(s,s')^* \geqslant 0 \tag{6.4}$$

for almost all $(s,s') \in \mathbf{R}$. Obviously this holds in the special case $h \equiv 0$. For technical reasons we will now impose the condition $h \in C_0^\infty(\mathbf{R})$ and remove the regularization later. Note that

$$\left(-\beta_0 \frac{d}{ds} - i n\!\!\!/ + M - \Gamma h\right)\left(\beta_0 \frac{d}{ds} + i n\!\!\!/ + M + \Gamma h\right) = -\frac{d^2}{ds^2} + n^2 + M^2 + h^2 + \Gamma\beta_0 h'.$$

is a self-adjoint operator on $\mathscr{H}_\kappa$ which is bounded below by $n^2 + M^2$ because

$$-\frac{d^2}{ds^2} + h^2 + \Gamma\beta_0 h' = \left(-\beta_0 \frac{d}{ds} - \Gamma h\right)\left(\beta_0 \frac{d}{ds} + \Gamma h\right)$$

and

$$-\beta_0 \frac{d}{ds} - \Gamma h = \left(\beta_0 \frac{d}{ds} + \Gamma h\right)^\dagger \text{ on the domain of } -\frac{d^2}{ds^2}.$$

Let

$$C_h = \left(-\frac{d^2}{ds^2} + n^2 + M^2 + h^2 + \Gamma\beta_0 h'\right)^{-1}. \tag{6.5}$$

The point is that

$$\widetilde{S}_{M,h}^{(n)} = \left(\beta_0 \frac{d}{ds} + i n\!\!\!/ + M + \Gamma h\right) C_h \Gamma, \tag{6.6}$$

while the matrix structure of C_h involves only the self-adjoint matrix $\Gamma\beta_0$ and the identity matrix. Unfortunately, if we write down (6.6) in terms of kernels and substitute the result into (6.4), we still do not have a manifest inequality. However, part of our proof will involve the relation

$$\widetilde{S}^{(n)}_{M,h} - \widetilde{S}^{(-n)}_{-M,h} = 2(i\not{n} + M)C_h\Gamma. \tag{6.7}$$

Lemma 6.3: C_h has a measurable matrix kernel $C_h(s,s')$ which is L^2 on $\mathbf{R}\times K$ for every compact $K\subset\mathbf{R}$. Moreover, $C_h(s,s')$ is a positive matrix for almost all $(s,s')\in\mathbf{R}\times\mathbf{R}$.

Proof: The first claim follows from equation (6.7) and Corollary 6.2. Now since the second-order differential operator is a self-adjoint operator with a bounded inverse, basic spectral theory together with (6.5) gives us the formula

$$C_h = \int_0^\infty dt\, e^{-t(-(d^2/ds^2)+n^2+M^2+h^2+\Gamma\beta_0 h')},$$

where the integration is done in the strong operator sense. Hence, there exists an improper $\int_0^\infty dt$-integration with respect to which

$$C_h(s,s') = \int_0^\infty dt \times e^{-t(n^2+M^2)}e^{-t(-(d^2/ds^2)+h^2+\Gamma\beta_0 h')}(s,s')$$

for almost all $(s,s')\in\mathbf{R}\times\mathbf{R}$. Since there is a subsequence of the Trotter product approximations

$$(e^{(t/k)(d^2/ds^2)}e^{-(t/k)(h^2+\Gamma\beta_0 h')})^k(s,s') \tag{6.8}$$

converging to $e^{-t(-d^2/ds^2)+h^2+\Gamma\beta_0 h')}(s,s')$ for almost all $(s,s')\in\mathbf{R}\times\mathbf{R}$, the second claim is verified if we can show that (6.8) is a positive matrix. But this can be seen by writing (6.8) down explicitly as a multiple integral[5] because in the integrand there are no anti-commutation problems interfering with the combination of exponential factors into an exponentiated sum, and the resulting matrix exponential has a self-adjoint exponent (so, in fact, we can prove a Feynman–Kac formula for the semi-group, although we will not need such a high-powered result). □

Lemma 6.4: For each $s'\in\mathbf{R}$ and positive integer $\ell'\leqslant d_\kappa$ there is a unique solution $R_{\cdot,\ell'}(\cdot,s')$ in $\mathscr{H}_\kappa$ of the integral equation

$$R_{\ell,\ell'}(s,s') - \sum_{\ell''=1}^{d_\kappa}\int_{-\infty}^{\infty} ds''\,[S^{(n)}_M(s,s'')\Gamma]_{\ell,\ell''}h(s'')R_{\ell'',\ell'}(s'',s') = [S^{(n)}_M(s,s')]_{\ell,\ell'}, \tag{6.9}$$

$R(s,s') = \widetilde{S}^{(n)}_{M,h}(s,s')$ almost everywhere, and for each s', $R(s,s')$ is smooth in s for $s\neq s'$ and

$$-\beta_0\frac{d}{ds}R(s,s') - i\not{n}R(s,s') + MR(s,s') - \Gamma h(s)R(s,s') = 0, \quad s\neq s'. \tag{6.10}$$

Proof: By (6.3) we know that the rhs of (6.9) is L^2 for fixed s', so the invertibility of $1 - S^{(n)}_M\Gamma h$ establishes the first claim. Now apply the differential operator $-\beta_0(d/ds) - i\not{n} + M$ to (6.9) in the sense of distributions to obtain

$$-\left(\beta_0\frac{d}{ds} - i\not{n} + M - \Gamma h(s)\right)R(s,s') = \Gamma\delta(s-s').$$

Since h is smooth, $R(s,s')$ is smooth in s for $s\neq s'$[35] and we have (6.10). To establish that $R(s,s') = \widetilde{S}_{M,h}(s,s')$ almost everywhere, we note that $\widetilde{S}^{(n)}_{M,h}(s,s')$ satisfies the integral equation (6.9) in an almost everywhere sense because it is the matrix kernel of $(1 - S^{(n)}_M\Gamma h)^{-1}S^{(n)}_M\Gamma$. By Corollary 6.2 we know that $\widetilde{S}^{(n)}_{M,h}(\cdot,s')$ is L^2 for almost all s', and so for all such s', $\widetilde{S}^{(n)}_{M,h}(\cdot,s') = R(\cdot,s')$ almost everywhere, column for column, by virtue of uniqueness. □

Lemma 6.5: For $h\in C_0^\infty(\mathbf{R})$ real-valued,

$$\widetilde{S}^{(-n)}_{-M,-h}(s,s')^*\widetilde{S}^{(n)}_{M,h}(s,s') = 0, \quad \text{a.a. } (s,s')\in\mathbf{R}\times\mathbf{R}.$$

Proof: By the preceding lemma it suffices to show that for fixed s'

$$R^{(-n)}_{-M-h}(s,s')^*R^{(n)}_{M,h}(s,s') = 0, \quad s\neq s'.$$

Now by (6.10), we have the equations

$$\frac{d}{ds}R^{(n)}_{M,h}(s,s') = (-i\beta_0\not{n} + \beta_0 M - \beta_0\Gamma h(s))R^{(n)}_{M,h}(s,s'),$$

$$\frac{d}{ds}R^{(-n)}_{-M,-h}(s,s') = (i\beta_0\not{n} - \beta_0 M + \beta_0\Gamma h(s))R^{(-n)}_{-M,-h}(s,s').$$

The matrix adjoint of the second equation is

$$\begin{aligned}\frac{d}{ds}R^{(-n)}_{-M,-h}(s,s')^* &= R^{(-n)}_{-M,-h}(s,s')^*(-i\not{n}\beta_0 - \beta_0 M - \Gamma\beta_0 h(s))\\ &= R^{(-n)}_{-M,-h}(s,s')^*(i\beta_0\not{n} - \beta_0 M + \beta_0\Gamma h(s)).\end{aligned}$$

Hence

$$\frac{d}{ds}[R^{(-n)}_{-M,-h}(s,s')^*R^{(n)}_{M,h}(s,s')] = 0, \quad s\neq s',$$

so $R^{(-n)}_{-M,-h}(s,s')^*R^{(n)}_{M,h}(s,s')$ is a constant matrix on either side of the singularity. But the integral equation (6.9) implies exponential decay of this quantity as $s\to\pm\infty$. □

We are finally in a position to prove (6.4) for smooth h.

Theorem 6.6: For real-valued $h\in C_0^\infty(\mathbf{R})$,

$$\operatorname{tr}\widetilde{S}^{(n)}_{M,h}(s,s')\widetilde{S}^{(n)}_{M,-h}(s,s')^* = \operatorname{tr}\widetilde{S}^{(n)}_{M,-h}(s,s')^*\widetilde{S}^{(n)}_{M,h}(s,s')\geqslant 0$$

for almost all $(s,s')\in\mathbf{R}\times\mathbf{R}$.

Proof. For convenience we abuse notation by suppressing the arguments s,s'. Obviously

$$\begin{aligned}&[\widetilde{S}^{(n)}_{M,-h} - \widetilde{S}^{(-n)}_{-M,-h}]^*[\widetilde{S}^{(n)}_{M,h} - \widetilde{S}^{(-n)}_{-M,h}]\\ &\quad= \widetilde{S}^{(n)}_{M,-h}{}^*\widetilde{S}^{(n)}_{M,h} - \widetilde{S}^{(n)}_{M,-h}{}^*\widetilde{S}^{(-n)}_{-M,h}\\ &\qquad- \widetilde{S}^{(-n)}_{-M,-h}{}^*\widetilde{S}^{(n)}_{M,h} + \widetilde{S}^{(-n)}_{-M,-h}{}^*\widetilde{S}^{(-n)}_{-M,h},\end{aligned}$$

where the multiplication is understood to be pointwise matrix multiplication. Now the second and third terms on the rhs vanish by virtue of the preceding lemma; moreover, we can apply (6.7) to the lhs Hence

$$[2(i\not{n} + M)C_{-h}\Gamma]^*[2(i\not{n} + M)C_h\Gamma] = \widetilde{S}^{(n)}_{M,-h}{}^*\widetilde{S}^{(n)}_{M,h} + \widetilde{S}^{(-n)}_{-M,-h}{}^*\widetilde{S}^{(-n)}_{-M,h}$$

almost everywhere. Simplifying the lhs, we get

$$-4(n^2 + M^2)\Gamma C^*_{-h}C_h\Gamma = \widetilde{S}^{(n)}_{M,-h}{}^*\widetilde{S}^{(n)}_{M,h} - \Gamma\widetilde{S}^{(-n)*}_{M,h}\widetilde{S}^{(-n)}_{M,-h}\Gamma,$$

where we have applied the relation

$$\Gamma\widetilde{S}^{(n)}_{M,h} = -\widetilde{S}^{(n)}_{-M,-h}\Gamma \text{ a.e.}$$

to the second term on the rhs. Taking the trace of this equation, we have

$$4(n^2+M^2)\mathrm{tr}\, C^*_{-h} C_h$$
$$= \mathrm{tr}\, \tilde{S}^{(n)}_{M,-h}{}^* \tilde{S}^{(n)}_{M,h} + \mathrm{tr}\, \tilde{S}^{(-n)*}_{M,h} \tilde{S}^{(-n)}_{M,-h},$$
$$= \mathrm{tr}\, \tilde{S}^{(n)}_{M,-h}{}^* \tilde{S}^{(n)}_{M,h} + \overline{\mathrm{tr}\, \tilde{S}^{(-n)}_{M,-h}{}^* \tilde{S}^{(-n)}_{M,h}}.$$

But since the Dirac matrices have been chosen to be real-valued, we also have the relation

$$\overline{\tilde{S}^{(n)}_{M,h}} = \tilde{S}^{(-n)}_{M,h} \text{ a.e.}$$

Now for any matrices A,B,

$$\mathrm{tr}\bar{A} = \overline{\mathrm{tr}\, A}, \; \overline{A^*} = \bar{A}^*, \; \overline{AB} = \bar{A}\bar{B}$$

so

$$4(n^2+M^2)\mathrm{tr}\, C^*_{-h} C_h = 2\mathrm{tr}\, \tilde{S}^{(n)}_{M,h}{}^* \tilde{S}^{(n)}_{M,-h} \text{ a.e.}$$

By Lemma 6.3, C_h and C_{-h} are positive matrices, so

$$\mathrm{tr}\, C^*_{-h} C_h = \mathrm{tr}\, C_{-h} C_h \geqslant 0 \text{ a.e.},$$

and so the desired inequality is established. □

Corollary 6.7: For real-valued, bounded, measurable h with compact support,

$$\mathrm{tr}\, \tilde{S}^{(n)}_{M,h}(s,s')\tilde{S}^{(n)}_{M,-h}(s,s')^* \geqslant 0$$

for almost all $(s,s') \in \mathbf{R}\times\mathbf{R}$.

Proof: There is a sequence $\{h_k\}$ of C_0^∞ functions which are uniformly bounded and supported in a common bounded set such that

$$h_k \to h \text{ a.e.}, \quad k\to\infty.$$

By the preceding theorem we need only show that a subsequence of

$$\{S^{(n)}_{M,\pm h_k}(s,s')\}_{k=1}^{\infty}$$

converges to $\tilde{S}^{(n)}_{M,h}(s,s')$ for almost all (s,s'). Now by the dominated convergence theorem

$$S^{(n)}_M \Gamma h_k \to S^{(n)}_M \Gamma h, \quad k\to\infty,$$

in the Hilbert–Schmidt sense and therefore in the bounded operator sense. It follows from the Neumann expansion

$$(1-A_k)^{-1} - (1-A)^{-1}$$
$$= \sum_{k'=1}^{\infty} [(1-A)^{-1}(A_k - A)]^{k'}(1-A)^{-1}$$

[which certainly converges for $(1-A)^{-1}$ bounded and A_k sufficiently close to A in bounded norm] that

$$(1 \mp S^{(n)}_M \Gamma h_k)^{-1} \to (1 \mp S^{(n)}_M \Gamma h)^{-1}, \quad k\to\infty,$$

in bounded norm. Hence

$$\tilde{S}^{(n)}_{M,\pm h_k}$$
$$= (1\mp S^{(n)}_M \Gamma h_k)^{-1} S^{(n)}_M \Gamma \xrightarrow[k\to\infty]{} (1 \mp S^{(n)}_M \Gamma h)^{-1} S^{(n)}_M \Gamma$$
$$= \tilde{S}^{(n)}_{M,\pm h}$$

in Hilbert–Schmidt norm, and so we have convergence of kernels in the L^2 sense. □

ACKNOWLEDGMENTS

I wish to thank Lon Rosen for several useful conversations about the Yukawa$_2$ model and for providing an argument that completes the Seiler–Simon proof of the Mathews–Salam–Seiler formula. I am also grateful for being granted permission to state his argument in the context of my model. (See Lemma 2.7 and Corollary 2.8.) I would also like to thank Barry Simon for a couple of helpful comments.

[1]G. Battle and L. Rosen, J. Stat. Phys. **22**, 123 (1980).
[2]R. Streater and A. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964).
[3]B. Simon, *The P(φ)$_2$ Euclidean (Quantum) Field Theory* (Princeton U. P., Princeton, N. J., 1974).
[4]R. Jost, *The General Theory of Quantized Fields* (Am. Math. Soc., Providence, R. I., 1965).
[5]M. Reed and B. Simon, *Methods of Modern Mathematical Physics I, II, Functional Analysis* (Academic, New York, 1975); *ibid., Fourier Analysis, Self-Adjointness.*
[6]K. Osterwalder and R. Schrader, Commun. Math. Phys. **42**, 281 (1975).
[7]J. Schwinger, Phys. Rev. **115**, 721 (1959).
[8]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, N. J., 1971).
[9]J. Glimm, Commun. Math. Phys. **6**, 61 (1967).
[10]J. Glimm and A. Jaffe, "Quantum Field Theory Models," *Quantum Field Models in Statistical Mechanics and Quantum Field Theory*, Les Houches, 1970, edited by C. De Witt and R. Stora (Gordon and Breach, New York, 1971), p. 1.
[11]E. Seiler, Commun. Math. Phys. **42**, 163 (1975).
[12]E. Seiler and B. Simon, J. Math. Phys. **16**, 163 (1975).
[13]E. Seiler and B. Simon, Commun. Math. Phys. **45**, 99 (1975).
[14]E. Seiler and B. Simon, Ann. Phys. **97**, 470 (1976).
[15]K. Osterwalder and R. Schrader, Helv. Phys. Acta **46**, 277 (1973–1974).
[16]S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row, Peterson, Evanston, Ill., 1961).
[17]E. Nelson, J. Func. Anal. **12**, 97 (1973).
[18]F. Guerra, L. Rosen, and B. Simon, Ann. Math. **101**, 111 (1975).
[19]E. Nelson, J. Func. Anal. **12**, 211 (1973).
[20]B. Simon, "The Glimm–Jaffe φ-Bound: A Markov Proof," *Constructive Quantum Field Theory*, edited by G. Velo and A. Wightman (Springer-Verlag, Berlin and New York, 1973), p. 125.
[21]A. Cooper and L. Rosen, Trans. AMS **234**, 1 (1977).
[22]J. Magnen and R. Seneor, Commun. Math. Phys. **51**, 297 (1976).
[23]F. Guerra, Phys. Rev. Lett. **28**, 1213 (1972).
[24]J. Fröhlich, Helv. Phys. Acta **47**, 265 (1974).
[25]J. Glimm, A. Jaffe, and J. Spencer, Ann. Math. **100**, 585 (1974).
[26]C. Fortuin, P. Kastelyn, and J. Ginibre, Commun. Math. Phys. **22**, 89 (1971).
[27]S. Varadhan, *Stochastic Processes*, lecture notes from the Courant Institute, New York, 1968.
[28]E. Davies, J. London Math. Soc. **7**, 483 (1973).
[29]N. Dunford and J. Schwartz, *Linear Operators, Part II* (Interscience, New York, 1963).
[30]B. Simon, Adv. Math. **24**, 244 (1977).
[31]J. Fröhlich, Bull. AMS **84**, 165 (1978).
[32]J. Fröhlich, R. Israel, E. Lieb, and B. Simon, Commun. Math. Phys. **62**, 1 (1978).
[33]O. McBryan, Commun. Math. Phys. **45**, 279 (1975).
[34]Of course, $\int_{T^\nu} d^\nu k$ -integrations and $\Sigma_{n=-N}^{N}$-summations will also be involved, but they are δ-independent, and uniformity with respect to k is provided by $\omega(k) > \omega_0$.
[35]L. Hörmander, *Linear Partial Differential Operators* (Springer-Verlag, New York, 1969).

Solutions of the wave equation for superposed potentials with application to charmonium spectroscopy

L. K. Sharma, V. P. Iyer, Sheilly Mehta, S. V. Kulkarni, and Jyoti Choubey
Department of Applied Physics, Government Engineering College, Jabalpur-482 011, India

(Received 17 April 1981; accepted for publication 7 August 1981)

In this paper complete solutions of the Schrödinger equation for three different superposed potentials have been obtained. In particular high-energy asymptotic expansions of the bound-state eigenfunctions and eigenvalues are derived. Various properties of these expansions have been examined including the behavior of Regge trajectories. Finally the relevance of these investigations to various aspects of the spectroscopy of heavy quark composites is discussed in detail.

PACS numbers: 03.65.Ge

I. INTRODUCTION

After the discovery of J/ψ resonance[1] much theoretical effort has gone into understanding the spectrum and decay properties of the J/ψ family.[2,3] The charmonium model[4–7] has been quite successful in interpreting ψ and ψ' as charmed-quark–antiquark ($c\bar{c}$) bound states and other subsequently observed phenomena. The discovery[8] of charmed mesons, besides giving strong support to the theory, also provides a semiquantitative description of the vast body of new ideas accumulated in colliding-beam experiments in the 3–5 GeV energy range.[9]

The interaction between quarks is mediated by gluons. From the asymptotic-freedom arguments it is known that if the distance between quarks gets small (the interaction energy being very high), the strong coupling constant $\alpha_s < 1$ and one has a similar situation as in QED. One gluon exchange prevails and perturbation theory can be applied. At large distances, however, the interaction energy being small α_s gets large and this coupling prevents the liberation of quarks since the production of $q\bar{q}$ pairs is more likely.

The large masses of ψ resonances and charmed mesons led to the assumption that the charmed quarks are so heavy that they may be treated nonrelativistically.[4] No one has yet succeeded in calculating the effective form of the interquark forces from quantum chromodynamics even in the nonrelativistic limit.[10,11] Most of the attempts, therefore, at understanding the level spacing and the decay rates of heavy mesonic states in the ψ and Υ regions start from a nonrelativistic considerations of the bound-state problems for a vector (or scalar) interaction and utilize numerical techniques for solving the wave equation, while the spin-orbit coupling, tensor forces and hyperfine structure are treated as additional corrections. In spite of the fact that different models have been able to provide most of the important features of meson spectroscopy, a comparison of the results so obtained, however, shows that various parameters involved in these models differ appreciably in their values. This difference obviously indicates a certain range of uncertainty in their values. Although numerical methods have been found to be quite useful in such type of study it would be worthwhile to obtain theoretical expressions for obtaining good approximations to various features of the theory viz., the level spacings, leptonic decay width, etc. For the interpretation of various results, therefore, various potential models have been developed. The long-range binding between quarks has been "guessed" as $V(r) = ar$, i.e., $V(r)$ which has linear dependence on r and the constant a has to be determined experimentally. In some cases even a harmonic-oscillator potential[12] has been used. Müller-Kirsten *et al.*[13,14] have recently considered in detail the logarithmic and the superposition of arbitrary quark confining potential with a short-range gluon exchange Coulomb component. Eichten *et al.*[15] have also recently postulated that in the nonrelativistic limit many of the gross features of the potential between the charmed quarks can be simulated by the potential $V(r) = -k/r + r/a^2$. They have chosen this potential to give a simple interpolation between the known Coulomb-type force at short distance and a linear growth of the static potential suggested by some models of quark confinement.[16] In the following we consider in detail a perturbation-theoretical solution of the wave equation for the three potentials which are the superposition of

(I) a quark confining linear potential and a short-range gluon exchange modified Coulomb potential $-g_2/r \ln(r/r_0)$ as suggested by asymptotic-freedom arguments,[17]

(II) a logarithmic potential and a short-range gluon exchange modified Coulomb potential similar to that considered in (I) above, and

(III) an arbitrary power quark confining potential and a short-range gluon exchange centrifugal potential.

Besides the reasons mentioned in the beginning, the other motivations which led us to undertake such a study are that after the discovery of Υ states[18] doubts have been raised on the application of linear potential as a phenomenological *ansatz* for the quark confining interaction. In view of these difficulties with the linear potential, Quigg and Rosner[19,20] investigated the spectroscopy from a logarithmic and a power potential viewpoint. We hope, therefore, that our study made here will not only be useful in finding out the approximate form of the potential but will also pave the way for investigating the spectroscopy of heavy quark composites of various other new states which may be discovered in the future. In the present work we, however, confine ourselves to the study of charmonium spectroscopy only.

Since an exactly similar procedure has been adoped for solving all three potentials (I), (II), and (III), we show the

 0022-2488/82/050780-10$02.50

details of our calculations for potential (I) only. This has been done in Sec. II. We derive various types of solutions of the wavefunctions as well as asymptotic expansions for the energy eigenvalues and Regge trajectories. For the other two potentials, (II) and (III), we give in Sec. II only the final expressions for wavefunctions, energy eigenvalues, and Regge trajectories. In Sec. III then, we investigate the physical implication of these asymptotic expansions of the energy eigenvalues and Regge trajectories for all three potentials. In this section we also consider the spin-dependent corrections (spin-orbit, tensor, spin-spin) and calculate the splitting of the $3P_J$ levels of charmonium. The S-wave bound-state wave function at the origin and the leptonic decay rates have also been calculated. Finally, in Sec. IV, we give a brief discussion of our results.

II. ASYMPTOTIC EIGENSOLUTIONS

We consider the Schrödinger equation for the following potentials:

$$\text{(I) } V(r) = g_1 r - g_2/r \ln(r/r_0) - V_0, \tag{2.1}$$

$$\text{(II) } V(r) = g_1 \ln(r/r_0) - g_2/r \ln(r/r_0) - V_0, \tag{2.2}$$

where $g_1, g_2 > 0$, $g_1 \gg g_2$ and V_0 is constant, and

$$\text{(III) } V(r) = g_1 r^\lambda + g_2/r^2 - V_0, \tag{2.3}$$

where $\lambda \geqslant 1, g_1, g_2 > 0$, and V_0 is a constant. Separating off the motion of the center of mass in the usual way we obtain the radial wave equation for the relative motion of the two particles of masses m_1, m_2, i.e.,

$$\frac{d^2\psi}{dr^2} + \frac{2\mu}{\hbar^2}\left[E - \frac{l(l+1)\hbar^2}{2\mu r^2} - V(r)\right]\psi = 0, \tag{2.4}$$

where, as usual $\psi = (1/r)\psi(r) P_l^m(\cos\theta) e^{im\phi}$, $\mu = m_1 m_2/(m_1 + m_2)$ is the reduced mass of the two particles, and r is their separation.

Now we consider potential (2.1), on setting

$$\alpha = 2\mu(E + V_0)/\hbar^2,$$
$$\beta = 2\mu g_1/\hbar^2, \tag{2.5}$$
$$\delta = 2\mu g_2/\hbar^2,$$

and

$$\gamma = l(l+1),$$

Eq. (2.4) can be written as

$$\frac{d^2\psi}{dr^2} + \left[\alpha - \frac{\gamma}{r^2} - \beta r + \frac{\delta}{r\ln(r/r_0)}\right]\psi = 0. \tag{2.6}$$

Next setting

$$r = e^z \quad (-\infty < z < \infty), \quad \text{and} \qquad \psi = e^{z/2}\phi \tag{2.7}$$

we obtain the following basic equation:

$$\frac{d^2\phi}{dz^2} + (-L^2 + v(z))\phi = 0, \tag{2.8}$$

where

$$v(z) = \alpha e^{2z} - \beta e^{3z} + \delta\left(\frac{e^z}{z-c}\right),$$
$$L^2 = r + \tfrac{1}{4} \tag{2.9}$$

and

$$c = \ln r_0.$$

Now such value of z, say z_0, is determined for which $v(z)$ becomes maximal. In the vicinity of this maximum $[v(z) - L^2]$ may become positive, thereby making the solutions oscillatory as needed for the existence of eigenvalues. Hence setting $(dv/dz)_{z=z_0} = 0$, one gets on solving for z_0

$$z_0 = \ln\left[\frac{\alpha}{3\beta} \pm \frac{\alpha}{3\beta}\left\{1 + \frac{3\beta\delta}{\alpha^2}\left(\frac{\ln(2\alpha/3\beta r_0 e)}{\{\ln(2\alpha/3\beta r_0)\}^2}\right)\right\}^{1/2}\right] \tag{2.10}$$

for

$$\alpha > 0, \beta > 0.$$

Expanding $v(z)$ in the neighborhood of the maximum at z_0, we obtain

$$v(z) = v(z_0) + \sum_{i=2}^{\infty}\frac{(z-z_0)^i}{i!}v^{(i)}(z_0), \tag{2.11}$$

where for $i = 0, 1, 2, \cdots$,

$$v^{(i)}(z_0) = 2^i\alpha e^{2z_0} - 3^i\beta e^{3z_0} + \delta e^{z_0}[(z_0-c)^{-1} - i(z_0-c)^{-2} + i(i-1)(z_0-c)^{-3} - i(i-1)(i-2)(z_0-c)^{-4} + i(i-1)(i-2)(i-3)(z_0-c)^{-5} - \cdots]. \tag{2.12}$$

For $i = 0$ this expression is positive, for $i = 1$ it is zero, and for $i > 1$ it is negative [as required for a maximum of $v(z)$ at $z = z_0$ for $\alpha > 0$)]. We now set $h = \{-2v^{(2)}(z_0)\}^{1/4}$, i.e.,

$$h^2 = \frac{4\alpha^{3/2}}{3\beta} + \frac{5\delta K}{2\alpha^{1/2}} - \frac{3\beta\delta[a]}{4\alpha^{3/2}} + O(\beta^2), \tag{2.13}$$

where

$$K = \frac{\ln(2\alpha/3\beta r_0 e)}{[\ln(2\alpha/3\beta r_0)]^2}$$

and

$$[a] = [(z_0-c)^{-1} + 2(z_0-c)^{-3} - 2(z_0-c)^{-2}]. \tag{2.14}$$

On changing the independent variable in (2.8) to $\omega = h(z - z_0)$ one gets

$$\frac{d^2\phi}{d\omega^2} + \left[\frac{-L^2 + v(z_0)}{h^2} - \frac{\omega^2}{4}\right]\phi = \sum_{i=3}^{\infty}\left(\frac{v^{(i)}(z_0)}{2v^{(2)}(z_0)}\right)\frac{\omega^i}{i!h^{i-2}}\phi, \tag{2.15}$$

where

$$\frac{v^{(i)}(z_0)}{v^{(2)}(z_0)} = \frac{2^i\alpha e^{2z_0} - 3^i\beta e^{3z_0} + \delta e^{z_0}[d]}{2^2\alpha e^{2z_0} - 3^2\beta e^{3z_0} + \delta e^{z_0}[a]} \tag{2.16}$$

and

$$[d] = [(z_0-c)^{-1} - i(z_0-c)^{-2} + i(i-1)(z_0-c)^{-3} - i(i-1)(i-2)(z_0-c)^{-4} + \cdots]. \tag{2.17}$$

In particular, we have for $i = 3, 4$

$$\frac{v^{(3)}(z_0)}{v^{(2)}(z_0)} = \left[5 + \frac{15}{4}\frac{\beta\delta}{\alpha^2}[a] - \frac{3}{4}\frac{\beta\delta}{\alpha^2}[b] - \frac{9}{2}\frac{\beta\delta K}{\alpha^2}\right] \tag{2.18}$$

and

$$\frac{v^{(4)}(z_0)}{v^{(2)}(z_0)} = \left[19 + \frac{57\beta\delta[a]}{4\alpha^2} - \frac{3\beta\delta[c]}{4\alpha^2} - \frac{45}{2}\frac{\beta\delta K}{\alpha^2}\right], \tag{2.19}$$

where

$$[b] = [(z_0-c)^{-1} - 3(z_0-c)^{-2} + 6(z_0-c)^{-3} - 6(z_0-c)^{-4}] \tag{2.20}$$

and

$$[c] = [(z_0-c)^{-1} - 4(z_0-c)^{-2} + 12(z_0-c)^{-3} - 24(z_0-c)^{-4} + 24(z_0-c)^{-5}]. \tag{2.21}$$

For large values of h the right-hand side of (2.15) may, to a first approximation, be neglected. The corresponding behavior of the eigenvalues $(1/h^2)(-L^2+v(z_0))$ can then be determined by comparing the equation with the equation of parabolic cylinder functions. The solutions are square integrable only if

$$(1/h^2)[-L^2+v(z_0)] = \tfrac{1}{2}q,$$

where $q = 2n+1$, $n = 0, 1, 2, \ldots$ (provided the wavefunction is required to vanish at infinity; otherwise it is only approximately an odd integer). For the complete solution we get

$$(1/h^2)(-L^2+v(z_0)) = \tfrac{1}{2}q + \Delta/h. \tag{2.22}$$

The quantity Δ in (2.22) remains to be determined. Substituting (2.22) into (2.15) one gets an equation which can be written as

$$\mathscr{D}_q\phi = \frac{2\Delta}{h}\phi - \sum_{i=3}^{\infty}\left(\frac{v^{(i)}(z_0)}{v^{(2)}(z_0)}\right)\frac{\omega^i}{i!h^{i-2}}\phi, \tag{2.23}$$

where

$$\mathscr{D}_q = -2\frac{d^2}{d\omega^2} - q + \frac{1}{2}\omega^2. \tag{2.24}$$

Equation (2.23) is now in a form suitable for the application of our perturbation method. $\phi = \phi^{(0)}$ is simply a parabolic cylinder function $D_{(q-1)/2}(\omega)$, i.e.,

$$\phi^{(0)} = \phi_q = D_{(q-1)/2}(\omega), \qquad \mathscr{D}_q\phi_q = 0, \tag{2.25}$$

where

$$D_{(q-1)/2}(\omega) = 2^{(q-3)/4}e^{-\omega^2/4}\Psi\left(\frac{3-q}{4}, \frac{3}{2}; \frac{\omega^2}{2}\right),$$

Ψ being a confluent hypergeometric function. The function ϕ_q is well known to obey the recursion formula

$$\omega\phi_q = (q, q+2)\phi_{q+2} + (q, q-2)\phi_{q-2}, \tag{2.26}$$

where

$$(q, q+2) = 1, \quad (q, q-2) = \tfrac{1}{2}(q-1). \tag{2.27}$$

For higher powers we have

$$\omega^i\phi_q = \sum_{j=2i}^{-2i} S_i(q,j)\phi_{q+j} \tag{2.28}$$

and a recursion relation can be written down for the coefficients S_i. The first approximation $\phi = \phi^{(0)}$ then leaves uncompensated terms amounting to

$$R_q^{(0)} = \left\{\frac{2\Delta}{h} - \sum_{i=3}^{\infty}\left(\frac{v^{(i)}(z_0)}{v^{(2)}(z_0)}\right)\frac{\omega^i}{i!h^{i-2}}\right\}\phi_q(\omega)$$
$$= \frac{2\Delta}{h}\phi_q - \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{j=2i}^{-2i}\widetilde{S}_i(q,j)\phi_{q+j}(\omega), \tag{2.29}$$

where we have set

$$\widetilde{S}_i(q,j) = \frac{v^{(i)}(z_0)}{v^{(2)}(z_0)}\frac{1}{i!}S_i(q,j). \tag{2.30}$$

We rewrite (2.29) in the form

$$R_q^{(0)} = \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{j=2i}^{-2i}[q, q+j]_i\phi_{q+j}(\omega), \tag{2.31}$$

where

$$[q, q]_3 = 2\Delta - \widetilde{S}_3(q,j),$$

and for $j \neq 0$

$$[q, q+j]_3 = -\widetilde{S}_3(q,j), \tag{2.32}$$

and for

$$i > 3, \quad -2i \leqslant j \leqslant 2i,$$
$$[q, q+j]_i = -\widetilde{S}_i(q,j).$$

Since $\mathscr{D}_{q+j} = \mathscr{D}_q - j$, $\mathscr{D}_q\phi_{q+j} = j\phi_{q+j}$, a term $\mu\phi_{q+j}$ in $R_q^{(0)}$ can be removed by adding to $\phi^{(0)}$ the contribution $\mu\phi_{q+j}/j$ except, of course, when $j = 0$. Thus, the next-order contribution of ϕ becomes

$$\phi^{(1)} = \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{\substack{j=2i\\ j\neq 0}}^{-2i}\frac{[q, q+j]_i}{j}\phi_{q+j}(\omega). \tag{2.33}$$

In its turn this contribution leaves uncompensated

$$R_q^{(1)} = \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{\substack{j=2i\\ j\neq 0}}^{-2i}\frac{[q, q+j]_i}{j}R_{q+j}^{(0)}$$

and yields the next contribution of ϕ

$$\phi^{(2)} = \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{\substack{j=2i\\ j\neq 0}}^{-2i}\frac{[q, q+j]_i}{j}\sum_{i'=3}^{\infty}\frac{1}{h^{i'-2}}$$
$$\times\sum_{\substack{j'=2i'\\ j+j'\neq 0}}^{-2i'}\frac{[q+j, q+j+j']_{i'}}{j+j'}\phi_{q+j+j'}. \tag{2.34}$$

Proceeding this way one gets the solution

$$\phi = \phi^{(0)} + \phi^{(1)} + \phi^{(2)} + \cdots,$$

which is an asymptotic expansion in descending powers of h valid for

$$z - z_0 = O(1/h), \tag{2.35}$$

i.e., around $z = z_0$. Together with this solution we obtain an eigenvalue equation from which Δ in (2.22) follows. The latter is obtained by setting equal to zero the sum of the terms in ϕ_q in $R_q^{(0)}, R_q^{(1)}, \ldots$ which have been unaccounted for so far. Thus

$$0 = \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}[q, q]_i + \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{\substack{j=2i\\ j\neq 0}}^{-2i}\frac{[q, q+j]_i}{j}\sum_{i'=3}^{\infty}\frac{1}{h^{i'-2}}[q+j, q]_{i'} + \cdots, \tag{2.36}$$

or

$$0=\frac{1}{h}[q,q]_3+\frac{1}{h^2}\left\{[q,q]_4+\sum_{\substack{j=6\\ j\neq 0}}^{-6}\frac{[q,q+j]_3}{j}[q+j,q]_3\right\}+O(1/h^3). \tag{2.37}$$

This is the equation from which Δ and hence the eigenvalues are determined. Thus

$$2h\Delta=\{\widetilde{S}_4(q,0)-\tfrac{1}{6}\widetilde{S}_3(q,6)\widetilde{S}_3(q+6,\ -6)+\tfrac{1}{6}\widetilde{S}_3(q,\ -6)\widetilde{S}_3(q-6,6)-\tfrac{1}{2}\widetilde{S}_3(q,2)\widetilde{S}_3(q+2,\ -2)$$
$$+\tfrac{1}{2}\widetilde{S}_3(q,\ -2)\widetilde{S}_3(q-2,2)\}+O(1/h^2)$$
$$=\frac{q^2+1}{4}\left\{19-\frac{45\beta\delta K}{2\alpha^2}-\frac{3\beta\delta[c]}{4\alpha^2}+\frac{57\beta\delta[a]}{4\alpha^2}\right\}-\frac{(15q^2+7)}{2^4.3^2}\left\{25-\frac{45\beta\delta K}{\alpha^2}-\frac{15\beta\delta}{2\alpha^2}[b]+\frac{75\beta\delta[a]}{2\alpha^2}\right\}$$
$$+O(1/h^2).$$

Using (2.5), (2.18), (2.19), and (2.22) we obtain

$$(l+1/2)^2=\frac{1}{12}\left[\frac{16\alpha^3}{9\beta^2}+\frac{20\alpha\delta K}{3\beta}-2\delta[a]\right]-\frac{1}{2}q\left[\frac{4\alpha^{3/2}}{3\beta}+\frac{5\delta K}{2\alpha^{1/2}}-\frac{3\beta\delta[a]}{4\alpha^{3/2}}\right]+\frac{(105q^2+25)}{2^6}\frac{\beta\delta K}{\alpha^2}$$
$$+\frac{(51q^2+1)}{72}+\frac{(219q^2+59)}{3\times2^6}\frac{\beta\delta}{\alpha^2}[A]^{-1}-\frac{(93q^2+23)}{3\times2^4}\frac{\beta\delta}{\alpha^2}[A]^{-2}+\frac{(183q^2+23)}{3\times2^6}\frac{\beta\delta}{\alpha^2}[A]^{-3}$$
$$+\frac{(57q^2+17)}{2^5}\frac{\beta\delta}{\alpha^2}[A]^{-4}+\frac{9(q^2+1)}{2^4}\frac{\beta\delta}{\alpha^2}[A]^{-5}+\frac{q}{2^8}\left[\frac{4\alpha^{3/2}}{3\beta}+\frac{5\delta K}{2\alpha^{1/2}}-\frac{3\beta\delta[a]}{4\alpha^{3/2}}\right]^{-1}$$
$$\times[(-19.851q^2+0.2963)-(857.775q^2-451.11)\frac{\beta\delta K}{\alpha^2}+\frac{\beta\delta}{\alpha^2}\{(2932.444q^2+1999.11)[A]^{-1}$$
$$-(4380.443q^2+2695.111)[A]^{-2}+(697.777q^2-192.89)[A]^{-3}$$
$$+(4530.66q^2-5309.334)[A]^{-4}+(7296q^2+7616)[A]^{-5}$$
$$+(4640q^2+10\,400)[A]^{-6}+(960q^2+4800)[A]^{-7}\}]+O(1/h^3),$$

where

$$[A]=\left[\ln\left(\frac{2\alpha}{3\beta r_0}\right)\left(1+\frac{3\beta\delta K}{4\alpha^2}\right)\right]. \tag{2.38}$$

It may be of interest to note that for $\delta=0$ the expression (2.38) reduces to the eigenfunction expansion for the pure linear potential. This expansion is found to be in exact agreement up to $O(1/h^2)$ with that obtained by Müller-Kirsten *et al.*[14] Proceeding in a similar manner we can write the following large-h asymptotic expansions of eigenfunctions of the Schrödinger equation for potential (2.2)

$$(l+1/2)^2=\frac{1}{12}\left[4\beta r_0\exp2\left(\frac{E+V_0}{g_1}-\frac{1}{2}\right)-2r_0\exp\left(\frac{E+V_0}{g_1}-\frac{1}{2}\right)\delta\left(4\beta r_0\exp\left(\frac{E+V_0}{g_1}-\frac{1}{2}\right)+[a]\right)\right.$$
$$-\frac{1}{2}q\left[4\beta r_0\exp2\left(\frac{E+V_0}{g_1}-\frac{1}{2}\right)-2r_0\exp\left(\frac{E+V_0}{g_1}-\frac{1}{2}\right)\delta\left(4\beta r_0\exp\left(\frac{E+V_0}{g_1}-\frac{1}{2}\right)+[a]\right)\right]^{1/2}$$
$$+\frac{(33q^2+1)}{2^3 3^2}+\frac{(21q^2+5)}{2^2 3^2}\delta\rho-\frac{\delta r_0}{\beta}\exp\left[-\left(\frac{E+V_0}{g_1}-\frac{1}{2}\right)\right]\left\{-\frac{(357q^2+109}{2}[a]\right.$$
$$\left.-\frac{9}{8}(q^2+1)[c]+4(15q^2+7)[b]\right\}+\frac{q}{2^8h^2}\left[(-7.0125q^2+0.2962)+(212.148q^2+2967.25)\delta\rho\right.$$
$$+\delta r_0\exp\left\{-\left(\frac{E+V_0}{g_1}-\frac{1}{2}\right)\right\}\left\{(2554.96q^2+3843.85)[a]-(827.259q^2+926.814)[b]\right.$$
$$\left.\left.+(498q^2+822)[c]-(99.55q^2+270.22)[\alpha]+(5.55q^2+27.778)[e]\right\}\right]+O(1/h^3), \tag{2.39}$$

where the notations β and δ have the same values as given in (2.5) while other notations used in the above expansion have the following meaning:

$$\alpha = 2\mu(E + g_1 \ln r_0 + V_0), \quad \rho = \frac{e^{-\alpha/\beta + 1/2}(c' + 3/2)}{\beta(c' + 1/2)^2}, c' = -\frac{\alpha}{\beta} + \ln r_0,$$
$$h^4 = 4\beta e^{(2\alpha - \beta)/\beta} - 2e^{(2\alpha - \beta)/2\beta}[4\beta e^{(2\alpha - \beta)/2\beta} + [a]]\delta + O(\delta^2),$$
$$[a] = \left[\frac{1}{(z_0 - c')} - \frac{2}{(z_0 - c')^2} + \frac{2}{(z_0 - c')^3}\right],$$
$$[b] = \left[\frac{1}{(z_0 - c')} - \frac{3}{(z_0 - c')^2} + \frac{6}{(z_0 - c')^3} - \frac{6}{(z_0 - c')^4}\right],$$
$$[c] = \left[\frac{1}{(z_0 - c')} - \frac{4}{(z_0 - c')^2} + \frac{12}{(z_0 - c')^3} - \frac{24}{(z_0 - c')^4} + \frac{24}{(z_0 - c')^5}\right],$$
$$[d] = \left[\frac{1}{(z_0 - c')} - \frac{5}{(z_0 - c')^2} + \frac{20}{(z_0 - c')^3} - \frac{60}{(z_0 - c')^4} + \frac{120}{(z_0 - c')^5} - \frac{120}{(z_0 - c')^6}\right],$$
$$[f] = \left[\frac{1}{(z_0 - c')} - \frac{6}{(z_0 - c')^2} + \frac{30}{(z_0 - c')^3} - \frac{120}{(z_0 - c')^4} + \frac{360}{(z_0 - c')^5} - \frac{720}{(z_0 - c')^6} + \frac{720}{(z_0 - c')^7}\right]. \tag{2.40}$$

It may be of interest to note that for $\delta = 0$ the expression (2.39) reduces to a eigenfunction expansion for the pure logarithmic potential. This expansion is found to be in agreement up to $O(1/h^2)$ with that obtained by Müller-Kirsten *et al.*[13]

Similarly for potential (2.3), we can write the following large h-asymptotic expansion for the eigenfunction

$$\begin{aligned}(l + 1/2)^2 = {} & \delta + \alpha\rho^2\left(\frac{1}{\lambda + 2}\right) - q\rho(\alpha\lambda)^{1/2} + \frac{1}{2^5 3^2}[-3^2(q^2 + 1)(\lambda^2 + 6\lambda + 12) + (15q^2 + 7)(\lambda + 4)^2] \\ & - \frac{q}{2^9(\alpha\lambda)^{1/2}\rho}\left[\frac{(-320q^2 - 1600)}{720}\left\{\frac{(\lambda + 2)^5 - 2^5}{\lambda}\right\} + \frac{(2240q^2 + 6080)}{5!3!}\left\{\frac{(\lambda + 2)^4 - 2^4}{\lambda}\right\}\left\{\frac{(\lambda + 2)^2 - 2^2}{\lambda}\right\}\right. \\ & + \frac{(544q^2 + 2144)}{2^6 3^2}\left\{\frac{(\lambda + 2)^3 - 2^3}{\lambda}\right\}^2 - \frac{(7200q^2 + 14688)}{4!(3!)^2} \\ & \times\left\{\frac{(\lambda + 2)^3 - 2^3}{\lambda}\right\}\left\{\frac{(\lambda + 2)^2 - 2^2}{\lambda}\right\}^2 + \frac{(5640q^2 + 9240)}{(3!)^4}\left(\frac{(\lambda + 2)^2 - 2^2}{\lambda}\right)^4 + O(1/h^3),\end{aligned} \tag{2.41}$$

where δ, α, and β have the same values as defined in Eq. (2.5). Other notations used in (2.41) have the following meaning:

$$\rho = \left[\frac{2\alpha}{(2 + \lambda)\beta}\right]^{1/\lambda}, \quad h^2 = 2(\alpha\lambda)^{1/2}\rho. \tag{2.42}$$

Equation (2.40) for $\lambda = 1$ and $\delta = 0$ reduces to the eigenfunction expansion for the pure linear potential. This expansion is found to be in exact agreement up to $O(1/h^2)$ with that obtained by Müller-Kirsten *et al.*[14]

III. APPLICATIONS

A. Regge trajectories

We have utilized the square roots of the expansions (2.39) and (2.41) for plotting Regge trajectories. The Regge trajectories for potential (2.2) has been shown in Fig. 1, while for the potential (2.3) they have been plotted in Figs. 2 and 3. In Fig. 2, in particular, the first four Regge trajectories have been shown for $\lambda = 1$. Similarly in Fig. 3 they have been plotted for $\lambda = 2$. The values of quark mass m, δ, and β chosen for plotting these trajectories have been given in Table I. We observe that these trajectories are almost linear over the range of immediate interest. It is, therefore, of interest to conclude that Regge trajectories for the superposed potentials considered by us in this paper do not lose their linearity and their bahavior is similar to the ones obtained for pure logarithmic and linear potentials.

B. Calculation of bound quark-antiquark mass

Since the quark mass is twice the reduced mass μ, the mass M_q of a bound quark-antiquark pair in the state q is given by

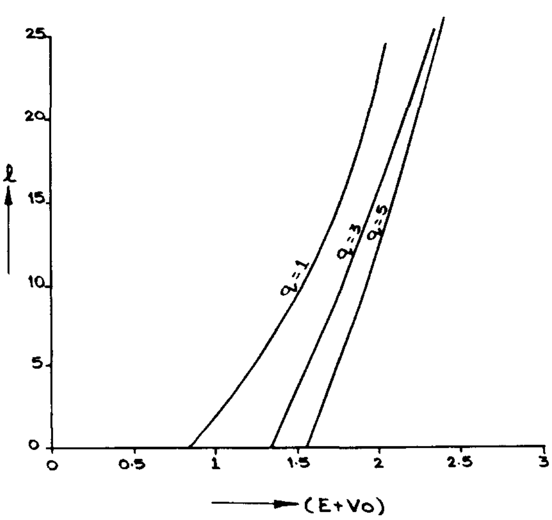


FIG. 1. First three Regge trajectories for the potential (2.2) with $m = 1.65$ GeV, $\beta = 2.37$ GeV^3, $\delta = 0.15$ GeV, and $r_0 = 1$ (GeV)$^{-1}$.

$$M_q = 4\mu + E. \tag{3.1}$$

In Table I, the calculated values of lowest S and P states obtained from (2.38), (2.39), (2.41), and (3.1) along with their corresponding experimental values are given. The values of the corresponding parameters used in the calculation (i.e., δ, β, r_0, and m) are also given in the same table. The values predicted for the higher S states agree with those stated by other authors (see, e.g., Refs. 21 and 22). The P states, of course, have to be corrected for tensor forces and spin-orbit and spin-spin coupling. This is done in the following Sec. III C.

C. Spin-dependent corrections

Spin-dependent corrections have been considered in many previous investigations[23–27] and are generally taken over from the corresponding work on positronium. To leading order in $(v/c)^2$, where v is the relative velocity of the quark and antiquark (each of mass m), the correction to be applied to the potential $V(r)$ is

$$V_c(r) = \frac{3}{2m^2}\frac{1}{r}\frac{dV(r)}{dr}\mathbf{L}\cdot\mathbf{s} + \frac{1}{6m^2}\nabla^2 V(r)\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2 - \frac{1}{12m^2}\left(\frac{d^2V(r)}{dr^2} - \frac{1}{r}\frac{dV(r)}{dr}\right)S_{12}. \tag{3.2}$$

Here $\mathbf{L}$ is the orbital angular momentum operator, $\mathbf{S} = \frac{1}{2}(\boldsymbol{\sigma}_1 + \boldsymbol{\sigma}_2)$, and S_{12} is the standard tensor operator, i.e.,

$$S_{12} = 3(\boldsymbol{\sigma}_1\cdot\hat{r})(\boldsymbol{\sigma}_2\cdot\hat{r}) - \boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2.$$

In the usual way we have for $S = 1$ and, respectively, $J = L-1, L, L+1$

$$\mathbf{L}\cdot\mathbf{S} = -(L+1),\ -1,\ L. \tag{3.3}$$

Similarly we have for $S = 1$ and, respectively,[28] $L = J-1, J, J+1$:

$$S_{12} = -\frac{2(J-1)}{2J+1},\ 2,\ -\frac{2(J+2)}{2J+1}. \tag{3.4}$$

We first discuss the correction to be applied to the potential (2.1). Substituting V into V_c we obtain terms which are singular at $r = 0$ (i.e., more divergent than $1/r^2$). Since there are no acceptable bound-state solutions for such singular potentials, we have to regularize the singularities by the introduction of cutoff parameters a, b, and c into the potential. We choose these parameters by using the following replacements in the singular terms[29]:

$$\left.\begin{aligned}
\frac{1}{r^3\ln(r/r_0)} &\to \frac{1}{r\ln(r/r_0)}\,\frac{1}{(r^2+a^2)},\\
\frac{1}{r^3(\ln(r/r_0))^2} &\to \frac{1}{r\ln(r/r_0)}\,\frac{1}{(r^2\ln(r/r_0)+b^2)},\\
\frac{1}{(r\ln(r/r_0))^3} &\to \frac{1}{r\ln(r/r_0)}\,\frac{1}{[(r\ln(r/r_0))^2+c^2]}.
\end{aligned}\right\} \tag{3.5}$$

Substituting V into V_c we then have

$$\begin{aligned}
V_c(r) = {} & \frac{3}{2m^2}\left[\frac{g_1}{r} + \frac{g_2}{r\ln(r/r_0)(r^2+a^2)} + \frac{g_2}{r\ln(r/r_0)\{r^2\ln(r/r_0)+b^2\}}\right]\mathbf{L}\cdot\mathbf{S}\\
& + \frac{1}{6m^2}\left[\frac{2g_1}{r} - \frac{g_2}{r\ln(r/r_0)\{r^2\ln(r/r_0)+b^2\}} - \frac{2g_2}{r\ln(r/r_0)\{(r\ln(r/r_0))^2+c^2\}}\right]\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2\\
& + \frac{1}{12m^2}\left[\frac{2g_2}{r\ln(r/r_0)(r^2+a^2)} + \frac{3g_2}{r\ln(r/r_0)\{r^2\ln(r/r_0)+b^2\}} + \frac{2g_2}{r\ln(r/r_0)\{(r\ln(r/r_0))^2+c^2\}}\right.\\
& \left. + \frac{g_1}{r} + \frac{g_2}{r\ln(r/r_0)(r^2+a^2)} + \frac{g_2}{r\ln(r/r_0)(r^2\ln(r/r_0)+b^2)}\right]S_{12}.
\end{aligned} \tag{3.6}$$

The Coulomb terms here can be treated as a further perturbation. We next assume that on the average the separation of quark and antiquark in the meson is such that $0 < r/a;\ [r^2\ln(r/r_0)]/b^2;\ [r\ln(r/r_0)]/c \ll 1$. In such a case we can expand the regularized denominators in rising powers of r and ignore all powers of r higher than the first since these would have to be treated as further perturbations of the linear confinement potential.

Under these conditions, we have to add to $V(r)$ the contribution

$$V_c(r) = \frac{d}{r\ln(r/r_0)},$$

where

$$d = \frac{3}{2m^2}\left(\frac{1}{a^2} + \frac{1}{b^2}\right)\mathbf{L}\cdot\mathbf{S} - \frac{1}{6m^2}\left(\frac{1}{b^2} + \frac{2}{c^2}\right)\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2 + \frac{1}{12m^2}\left(\frac{3}{a^2} + \frac{4}{b^2} + \frac{2}{c^2}\right)S_{12}. \tag{3.7}$$

This means that the coupling constant g_2 associated with the $(r\ln(r/r_0))^{-1}$ term in $V(r)$ has only to be replaced by $(g_2 + d)$ to obtain the well-known $1\,^{2s+1}P_J$ states.

Expressions similar to (3.7) can also be found for the spin-dependent corrections to the potentials (2.2) and (2.3). Thus, for the potential (2.2)

$$V_c(r) = \frac{d_1}{r\ln(r/r_0)} + \frac{d_2}{r^2},$$

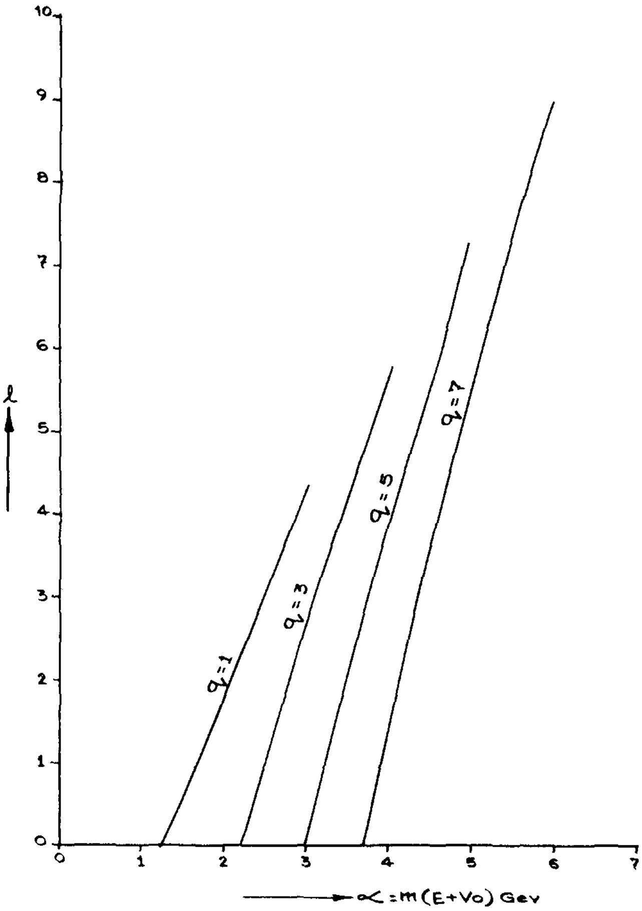


FIG. 2. First four Regge trajectories for the potential (2.3) with $\lambda = 1$, $m = 1.65$ GeV, $\beta = 0.35$ GeV^4, $\delta = 0.01$.

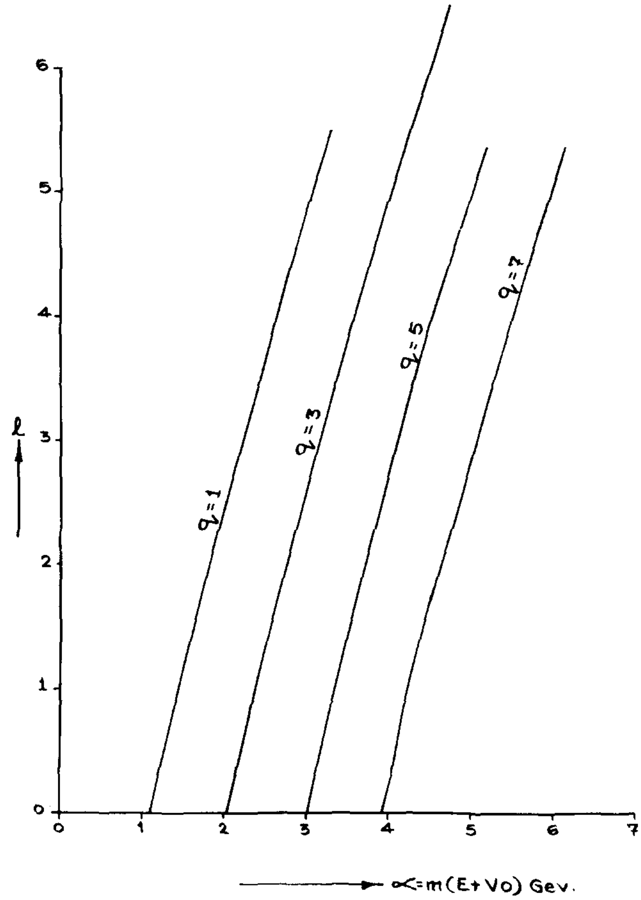


FIG. 3. First four Regge trajectories for the potential (2.3) with $\lambda = 2$, $m = 1.65$ GeV, $\beta = 0.055$ GeV^4, $\delta = 1.18$.

TABLE I. The lowest S and P states of charmonium obtained for the potentials (2.1), (2.2), and (2.3). The input value is underlined. The quark mass 'm' in each case has been chosen to be equal to 1.65 GeV.

States	Mass (GeV) calculated for potential (2.1)[a]	Mass (GeV) calculated for potential (2.2)[b]	Mass (GeV) calculated for potential (2.3)[c] with $\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$	Mass (GeV) (observed)
1^3S_1	<u>3.096</u>	<u>3.096</u>	<u>3.096</u>	<u>3.096</u>	<u>3.096</u>	<u>3.096</u>	<u>3.096</u>	<u>3.096</u>	<u>3.096</u>	<u>3.096</u>	3.096
1^3P	3.52	3.43	3.39	3.28	3.20	3.17	3.14	3.13	3.12	3.12	...
2^3S_1	3.64	3.69	3.66	3.66	3.67	3.68	3.66	3.68	3.68	3.77	3.684
2^3P	3.88	3.70	3.86	3.85	3.80	3.77	3.73	3.74	3.72	3.81	...
3^3S_1	4.06	3.88	4.14	4.23	4.22	4.23	4.16	4.17	4.16	4.29	4.16

[a]For this potential the input values used are $\beta = 0.3$ GeV^3, $\delta = 0.15$ GeV, $m = 1.65$ GeV, and $r_0 = 1$ GeV^{-1}.
[b]For this potential the input values used are $\beta = 2.37$ GeV^3, $\delta = 0.15$ GeV, $r_0 = 1$ GeV^{-1}.
[c]For this potential the input values used are

(i)	for $\lambda = 1$;	$\beta = 0.35$ GeV^4;	$\delta = +0.01$,
(ii)	for $\lambda = 2$;	$\beta = 0.055$ GeV^4;	$\delta = +1.18$,
(iii)	for $\lambda = 3$;	$\beta = 0.008$ GeV^4;	$\delta = 4.5$,
(iv)	for $\lambda = 4$;	$\beta = 0.003$ GeV^4;	$\delta = 10.5$,
(v)	for $\lambda = 5$;	$\beta = 9\times 10^{-5}$ GeV^4;	$\delta = 18.5$,
(vi)	for $\lambda = 6$;	$\beta = 8.5\times 10^{-6}$ GeV^4;	$\delta = 29.5$,
(vii)	for $\lambda = 7$;	$\beta = 7\times 10^{-7}$ GeV^4;	$\delta = 45.0$,
(viii)	for $\lambda = 8$;	$\beta = 1\times 10^{-7}$ GeV^4;	$\delta = 61.8$.

where

$$d_1 = \frac{g_2}{m^2}\left(\frac{1}{a^2} + \frac{1}{b^2}\right)\mathbf{L}\cdot\mathbf{S} - \frac{g_2}{6m^2}\left(\frac{1}{b^2} + \frac{2}{c^2}\right)\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2 + \frac{g_2}{12m^2}\left(\frac{3}{a^2} + \frac{4}{b^2} + \frac{2}{c^2}\right)S_{12} \tag{3.8}$$

and

$$d_2 = \frac{3g_1}{2m^2}\mathbf{L}\cdot\mathbf{S} + \frac{1}{6m^2}\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2 + \frac{1}{12m^2}S_{12}.$$

This means that, in this case, the coupling constant g_2 associated with the $(r\ln(r/r_0))^{-1}$ term in $V(r)$ has to be replaced by $(g_2 + d_1)$ and a term d_2 has to be added to the centrifugal term of the Schrödinger equation to obtain the well-known P states.

For the potential (2.3),

$$V_c(r) = \frac{d_3}{r} + d_4,$$

where

$$d_3 = \left[\frac{3}{2}\mathbf{L}\cdot\mathbf{S} + \frac{1}{3}\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2\right]\frac{g_1}{m^2}$$

and

$$d_4 = \left[3\mathbf{L}\cdot\mathbf{S} - \frac{1}{3}\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2 + \frac{1}{2}S_{12}\right]\frac{g_2}{a^4m^2}. \tag{3.9}$$

Thus a Coulomb term gets added to the potential $V(r)$. The solution given by Eq. (3.2) in Ref. 14 has therefore been used to calculate the P states.

In Table II the charmonium states $1\,{}^{2s+1}P_J$ for the three potentials (2.1), (2.2), and (2.3) along with the corresponding experimentally observed values[30] are given. It will be quite in order to state here that even with our simple (and perhaps crude) arguments, the agreement of our results with the experimentally observed values is quite good.

D. Calculations for leptonic decay widths

Decay widths are well known to play an important role in exploring the origin of a newly-found hadronic state. The leptonic decay widths of a vector quark–antiquark bound state such as ψ can be expressed in terms of the s-wave bound-state wavefunction at the origin. Thus

$$\Gamma(\psi\to l\bar{l}) = \frac{16\pi\alpha^2 e_Q^2}{M_\psi^2}|\Psi(0)|^2. \tag{3.10}$$

Here α is the fine structure constant [not to be confused with α of Eq. (2.5)]. e_Q is the charge of the constituent quark of ψ.

In the case of an s-wave bound state, the wavefunction is related to the potential V via the following expression:

$$|\Psi(0)|^2 = \frac{\mu}{2\pi}\left\langle\frac{dV}{dr}\right\rangle = \frac{\mu}{2\pi}\int\Psi^*(r)\frac{dV}{dr}\Psi(r)dr. \tag{3.11}$$

In order to calculate $\langle dV/dr\rangle$ and hence $|\Psi(0)|^2$, the WKB approximation method[31] is used.

Thus, for the potential (2.1)

$$|\Psi(0)|^2 = \frac{\mu}{2\pi} \times \frac{\int_{r_0}^{\bar{r}} dr\,[E - g_1r + g_2/r\ln(r/r_0) + V_0]^{-1/2}\,dV/dr}{\int_0^{\bar{r}} dr\,[E - g_1r + g_2/r\ln(r/r_0) + V_0]^{-1/2}}, \tag{3.12}$$

where $\bar{r}$ is the classical turning point, $r_0 = 1$, and $dV/dr = g_1 + g_2r^{-2}\{(\ln r)^{-1} + (\ln r)^{-2}\}$. Similarly for the potential (2.2), we have

$$|\Psi(0)|^2 = \frac{\mu}{2\pi} \times \frac{\int_{r_0}^{\bar{r}} dr[E - g_1\ln(r/r_0) + g_2/r\ln(r/r_0) + V_0]^{-1/2}\,dV/dr}{\int_0^{\bar{r}} dr\,[E - g_1\ln(r/r_0) + g_2/r\ln(r/r_0) + V_0]^{-1/2}}, \tag{3.13}$$

where $\bar{r}$ is the classical turning point, $r_0 = 1$, and

TABLE II. The charmonium states $1\,{}^{2s+1}P_J$ for the three potentials after corrections for spin interactions. All parameters are as given in Table I along with $a = b = c = 2.25$ GeV for potential (2.1); $a = 4.5$ GeV, $b = c = 2.25$ GeV for potential (2.2); and $a = 2.25$ GeV for potential (2.3).

States	Decay width (keV) calculated for potential (2.1)	Decay width (keV) calculated for potential (2.2)	Decay width (keV) calculated for potential (2.3) with $\lambda = 1$	$\lambda = 2$	Decay width (keV) (observed)
1^3S_1	4.8	4.8	3.36	3.8	4.8 ± 0.6[a]
2^3S_1	2.0	2.24	2.45	2.45	2.1 ± 0.3[a]
3^3S_1	1.02	1.25	1.9	1.7	0.77 ± 0.1[b]

[a]See Ref. 29.
[b]See Ref. 36.

TABLE III. Leptonic decay rates in keV for the three potentials. All parameters are as given in Table I.

States	Mass(GeV) calculated for potential (2.1)	Mass(GeV) calculated for potential (2.2)	Mass(GeV) calculated for potential (2.3) with $\lambda = 1$	Mass (GeV) (observed)[a]
1^3P_2	3.531	3.540	3.550	$3.561 \pm 7\times 10^{-3}$
1^3P_1	3.440	3.452	3.456	$3.511 \pm 7\times 10^{-3}$
1^3P_0	3.362	3.361	3.361	$3.413 \pm 9\times 10^{-3}$

[a]See Refs. 24 and 35.

$dV/dr = g_1/r + g_2 r^{-2}\{(\ln r)^{-1} + (\ln r)^{-2}\}$, and for the potential (2.3)

$$|\Psi(0)|^2 = \frac{\mu}{2\pi} \times \frac{\int_{r_1}^{r_2} dr\,[E - g_1 r^\lambda - g_2/r^2 + V_0]^{-1/2}\, dV/dr}{\int_{r_1}^{r_2} dr\,[E - g_1 r^\lambda - g_2/r^2 + V_0]^{-1/2}}, \tag{3.14}$$

where r_1 and r_2 are the classical turning points given by $\{E - V(r)\} = 0$. Equations (3.12) and (3.13) have been solved analytically to obtain $|\Psi(0)|^2$ while (3.14) has been evaluated numerically for $\lambda = 1$ and $\lambda = 2$.

The results obtained for the decay widths for all three potentials have been shown in Table III and compared with the experimentally-observed values.[32]

IV. DISCUSSION

The investigation made in this paper shows that simple perturbation methods can be formulated for solving the wave equation for the superposition of potentials. Our endeavor here in particular has been to show that besides the deployment of numerical integration techniques for obtaining the meson spectrum, decay widths, and evaluation of P states by considering the spin-orbit corrections, perturbation theory can also be successfully used in such studies.

In particular the asymptotic eigenexpansions and eigenvalues for the three superposed potentials (2.1), (2.2), and (2.3) have been derived. As in the earlier investigations by Dingle *et al.*[33] in connection with the asymptotic expansions of Mathieu functions and their eigenvalues it is found that expansions have successive terms which alternate in sign, thereby indicating the Borel summability of the expansion. In the expansions obtained by us a similar condition is satisfied.

In our study we have used perturbation terms up to $O(1/h^2)$ only. At first sight it might appear that the accuracy of our results could have been further improved had we evaluated perturbation terms beyond second order. We have, however, checked that the contribution due to second-order perturbation terms is found to be very small when compared to the first-order term. Therefore, the consideration of even this term does not appreciably change our results.

In potentials (2.1) and (2.2) we have superposed an improved form of the Coulomb potential—$g_2/[r \ln(r/r_0)]$ as suggested by asymptotic-freedom arguments[17] with linear and logarithmic confining potentials respectively. We have, however, assumed that this modified form of the Coulomb potential is sufficiently weak and is, therefore, treated as a perturbation of the linear and logarithmic confining potentials. Our results for the mass and leptonic decay widths for both these potentials are found to be quite satisfactory.

Potentials of the form $(\alpha r + \beta r^{-1})$ (or its invariants) have been quite fashionable in the ψ region although their causal connection with QCD (except perhaps in very general terms) may well be debated. Lack of interest in the harmonic oscillator (h.o.) like descriptions of the $c\bar{c}$ region, on the other hand, is due to an almost instinctive feeling that for c quarks which are heavy, only a nonrelativistic description makes sense, and that any relativisitic sophistication is simply not warranted. But a nonrelativistic h.o. model would necessarily yield equal mass spacings at the linear level, in definite disagreement with experiment, Joshi and Mitra[34] have, however, shown that the harmonic oscillator potential could offer some results on $c\bar{c}$ spectra provided it incorporates a short-range modification of the form $\sim r^{-2}$, as a sort of counterpart to the Coulomb term of a confining potential. Led by these considerations we have here studied a general power potential along with a short-range r^{-2} term as given in Eqs. (2.3). The main features of the results obtained for this may be summarized as follows:

1. The values of the predicted s states agree well with those stated by other authors.[21,35]

2. For the uncorrected P states it is observed that as the power λ is increased, the mass obtained for these states goes on decreasing.

3. As λ is increased the centrifugal term associated with the potential becomes more and more dominant, as is obvious from the increase in δ and decreases in β.

4. The decay width calculations do not give good result for the $1S$ state while for other states they are satisfactory.

It should also be noted that expansions (2.38), (2.39), and (2.41) are invariant under the joint interchange $q \to -q$ and $h^2 \to -h^2$, which obviously converts one solution into another.

ACKNOWLEDGMENT

The authors are indebted to Professor H. J. W. Müller-Kirsten for his valuable suggestions which resulted in the substantial improvement of this paper.

[1]J. J. Aubert *et al.*, Phys. Rev. Lett **33**, 1404 (1974); J. E. Augustin *et al.*, *ibid.* **33**, 1406 (1974).
[2]G. S. Abrams *et al.*, Phys. Rev. Lett. **33**, 1453 (1974).
[3]For a comprehensive review of the data, see G. J. Feldman and M. L. Perl, Phys. Rep. **33** C, 285 (1977).
[4]T. Appelquist and H. D. Politzer, Phys. Rev. Lett. **34**, 43 (1975); Phys. Rev. D **12**, 1404 (1975).
[5]T. Appelquist, A. De Rujula, H. D. Politzer, and S. L. Glashow, Phys. Rev. Lett. **34**, 365 (1975).
[6]E. Eichten, K. Gottfried, T. Kinoshita, K. Kogut, K. D. Lane, and T-M. Yan, Phys. Rev. Lett. **34**, 369 (1975).
[7]A. De Rujula and S. L. Glashow, Phys. Rev. Lett. **34**, 46 (1975); C. G. Callan, R. L. Kingsley, S. B. Treiman, F. Wilczek, and A. Zee, *ibid.* **34**, 52 (1975); B. J. Harrington, S. Y. Park, and A. Yildiz, *ibid.* **34**, 168, 706 (1975); J. S. Kang and H. J. Schnitzer, Phys. Rev. D **12**, 841 (1975); **12**, 2791 (1975).
[8]G. Goldhaber *et al.*, Phys. Rev. Lett. **37**, 255 (1976).
[9]E. Lohrman, in *Proceedings of the 1977 International Symposium on Lepton and Photon Interactions at High Energies, Hamburg*, edited by F. Gutbord (DESY, Hamburg), p. 641.
[10]See however, C. G. Callan, R. Dashen, and D. J. Gross, Inst. for Adv. Study Report No. C00-2220-115, 1977 (Aug. 1977, unpublished); F. L. Fienberg, Phys. Rev. Lett. **39**, 316 (1977); T. Appelquist, M. Dine, and I. Muzinich, Phys. Lett. B **69**, 231 (1977); W. Fischler, Nucl. Phys. B **129**, 157 (1977), E. C. Poggio, Phys. Lett. B **71**, 381 (1977).
[11]E. C. Poggio, Phys. Lett. B **71**, 381 (1977).
[12]M. Krammer and H. Krasemann, Nuovo Cimento A **32**, 394 (1976).
[13]H. J. W. Müller-Kirsten and S. K. Bose, J. Math. Phys. **20**, 2471 (1979).
[14]H. J. W. Müller-Kirsten, G. E. Hite, and S. K. Bose, J. Math. Phys. **20**, 1878 (1979).
[15]E. Eichten, K. Gottfried, T. Kinoshita, K. D. Lane, and T. M. Yan, Phys. Rev. D **17**, 3090 (1978).
[16]J. Kogut and L. Susskind, Phys. Rev. D **9**, 3501 (1974).
[17]F. L. Feinberg, Phys. Rev. Lett. **39**, 316 (1977); T. Appelquist, M. Dine, and I. Muzinich, Phys. Lett. B **69**, 231 (1977); W. Fischler, CERN-TM 2321, 1977 (unpublished); E. C. Poggio, Phys. Lett. B, **71** 381 (1977); E. Eichten, K. Gottfried, T. Kinoshita, K. D. Lane, and T. M. Yan, Cornell University preprint CINS-375, 1978 (unpublished).
[18]L. M. Lederman, *Proceedings of the 1977 International Symposium on Lepton and Photon Interactions at High Energies*, edited by F. Gutbrod (DESY, Hamburg), p. 567.
[19]C. Quigg and J. L. Rosner, Fermilab preprint 77/82-THY, 1977 (unpublished).
[20]C. Quigg and J. L. Rosner, Fermilab preprint 77/90-THY, 1977 (unpublished).
[21]T. Appelquist, R. M. Barnett, and K. Lane, SLAC-PUB-2100 (1978), to be puslished in Ann. Rev. Nucl. Particle Sci., Vol. 28.
[22]H. Schopper, DESY preprint 77/79 (1977), Lectures given at the Erice-School, Erice, Italy, 1977.
[23]R. Barbieri, R. Kögerler, Z. Kunszt, and R. Gatto, Nucl. Phys. B **105**, 125 (1976).
[24]D. Pignon and C. A Piketty, University of Paris preprints LPTENS 77/21, 78/2 (unpublished).
[25]J. D. Jackson, "Lectures on the New particles," Lawrence Berkeley report LBL-5500 (1976), also published in *The Proceedings of the 1976 Summer Institute on Particle Physics*, edited by M. C. Zipf (SLAC, Stanford, Calif., 1976).
[26]H. J. Schnitzer, Phys. Rev. D **13**, 74 (1975).
[27]H. J. Schnitzer, Phys. Rev. Lett. **35**, 1540 (1975).
[28]C. B. Dover, S. H. Kahana, and T. L. Trueman, Phys. Rev. D **16**, 799 (1977); W. W. Buck, C. B. Dover, and J. M. Richard, Brookhaven National Laboratory preprint BNL-23315, 1977 (unpublished).
[29]The same procedure was followed by Müller-Kirsten *et al.*, see Ref. 14; J. G. Wills, D. B. Lichtenberg, and J. T Kiehl, Phys. Rev. D **15**, 3358 (1977); D. B. Lichtenberg *et al.*, *ibid.* **35**, 1055 (1975).
[30]See, D. Pignon *et al.*, Ref. 24; H. F. W. Sadrozinski in the *Proceedings of the 1977 International Symposium on Lepton and Photon Interactions at High Energies*, edited by F. Gutbrod (Hamburg, DESY), p. 47.
[31]C. Quigg and J. L. Rosner, Fermilab-Pub-77/106-THY and Fermilab-Pub-78/19-THY, 1978 (unpublished).
[32]P. A. Rapidis *et al.*, Phys. Rev. Lett. **39**, 526 (1977).
[33]R. D. Dingle and H. J. W. Müller, J. Reine Angew. Math. **211**, 11 (1962).
[34]G. C. Joshi and A. N. Mitra, Hadronic J. **1**, 1591 (1978).
[35]H. F. W. Sadrozinski, in *Proceedings of the 1977 International Symposium on Lepton and Photon Interactions at High Energies*, edited by F. Gutbrod, (DESY, Hamburg), p. 47.
[36]A. Martin, CERN preprint TH. 2980 (1980).

Quantum mechanical scattering theory for potentials of the form $V(r) = (\sin r)/r^{\beta}$, $1/3 < \beta \leqslant 1/2$

Brian Bourgeois
Department of Mathematics, Texas A&M University, College Station, Texas 77843

(Received 23 December 1980; accepted for publication 27 March 1981)

A rigorous nonrelativistic time dependent quantum-mechanical scattering theory for a single particle is developed for potentials of the form $V(r) = (\sin r)/r^{\beta}$, $\frac{1}{3} < \beta \leqslant \frac{1}{2}$. The positive-energy solutions of the radial Schrödinger equation are used to construct modified wave operators which converge as $t \to \pm\infty$ on a dense set of states to the familiar time-independent formulas for the wave operators.

PACS numbers: 03.65.Nk

I. INTRODUCTION

Nonrelativistic quantum-mechanical scattering theory for short-range potentials and for long-range Coulomb-like potentials has been developed to a high degree of satisfaction in recent years (e.g., see Refs. 1–3). There remains incomplete the study of the so-called oscillatory potentials. Of these, the prototypical class of radial examples is given by

$$V(r) = (\sin r^{\alpha})/r^{\beta}, \beta > 0. \tag{1}$$

Existence of the Möller wave operators is known for (1) when $\alpha + \beta > 1$ and $\beta > \frac{1}{2}$ and also[4,5] when $2\beta + \alpha > 2$. The purpose of this paper is to establish a rigorous scattering theory for potentials of the form (1) for $\alpha = 1$ and $\frac{1}{3} < \beta \leqslant \frac{1}{2}$. This is achieved by making a detailed analysis of the positive-energy solutions of the radial Schrödinger equation. It is discovered that these solutions do not behave asymptotically like plane waves. Because of this fact, we expect that the usual Möller wave operators do not exist. We use the asymptotic form of the solutions to define a "modified" free propagator which is in turn employed to construct wave operators. These operators converge (in the time-dependent sense) on a dense set of states to the operators defined by the familiar time-independent formulas on partial wavespaces.

Let us now introduce some notation in order that we may state our main results. Since the potential $V(r)$ is spherically symmetric, we may separate radial and angular variables to reduce the problem from three to one dimension. The partial wavespaces

$$S_{lm} = \{\psi \in L^2(\mathbb{R}^3) \mid \psi(x) = R(r) Y_{lm}(\Omega)\},$$

when Y_{lm} is a spherical harmonic and $|\Omega| = 1$, are invariant subspaces for the Hamiltonian

$$H = -\Delta + V,$$

$$\mathscr{D}(H) = \{\psi \in L^2(\mathbb{R}^3) \mid \Delta\psi \in L^2(\mathbb{R}^3) \text{ in the sense of distributions}\}.$$

On S_{lm}, H is unitarily equivalent to the operator h on $L^2(\mathbb{R}^+)[\mathbb{R}^+ \equiv (0,\infty)]$ given by

$$h\phi = \left(-\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + V(r)\right)\phi,$$

$$\mathscr{D}(h) = \left\{\phi \in L^2(\mathbb{R}^+) \,\middle|\, \left(-\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + V(r)\right)\phi \in L^2(\mathbb{R}^+) \text{ in the sense of distributions and } \phi(r) \to 0 \text{ as } r \to 0\right\}.$$

(For a proof of this, see Ref. 6.) These facts allow us to focus our analysis (without loss of generality) on the operator h for fixed angular momentum l. For convenience and reference, we state our main results here.

Theorem 1: Let $V(r) = (\sin r)/r^{\beta}$, $\frac{1}{3} < \beta \leqslant \frac{1}{2}$. Suppose that $\phi(k,r)$ is a solution of the positive-energy radial Schrödinger equation, i.e., that ϕ satisfies

$$\left\{-\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + V(r)\right\}\phi(k,r) = k^2\phi(k,r),$$
$$(k \notin \{0, \pm\tfrac{1}{2}, \pm 1, \pm 2, \pm 3\}). \tag{2}$$

Then,

$$\phi(k,r) = c_+(k) e^{ikr + i\gamma(k)v(r)} + c_-(k) e^{-ikr - i\gamma(k)v(r)} + o(1) \text{ as } r \to \infty \tag{3}$$

where

$$v(r) = \int^r s^{-2\beta}\, ds \quad \text{and} \quad \gamma(k) = \frac{-1}{4k(4k^2-1)}. \tag{4}$$

Theorem 2: For $f \in L^2(\mathbb{R}^+)$, let

$$\tilde{f}(k) = \underset{R\to\infty}{\text{l.i.m.}} \int_0^R \phi_0(k,r) f(r)\, dr \tag{5}$$

[l.i.m.$\equiv$limit in mean, i.e., in $L^2(\mathbb{R}^+)$], where $\phi_0(k,r)$ is a solution of (2) with $V \equiv 0$, chosen so that $\phi_0(k,r) = O(r^{l+1})$ as $r \to 0$ and so that the map $f \to \tilde{f}$ is an isometry (see Ref. 7). Let

$$\Lambda = \{f \in L^2(\mathbb{R}^+) \mid \tilde{f} \in C_0^{\infty}(\mathbb{R}^+) \text{ and } \{\tfrac{1}{2}, 1, 2, 3, \not\subset \operatorname{supp} \tilde{f}\}. \tag{6}$$

Let us define the modified free propagator by

$$[u(t)f](r) = \int_0^{\infty} a(k,r) e^{-itk^2} \tilde{f}(k)\, dk, \quad f \in \Lambda, \tag{7}$$

where [see Eq. (4)]

$$a(k,r) = c_+(k) e^{ikr + i\gamma(k)v(r)} + c_-(k) e^{-ikr - i\gamma(k)v(r)}. \tag{8}$$

Then the limits

$$w^{\pm} f = \lim_{t \to \pm\infty} e^{ith} u(t) f \tag{9}$$

exist on the dense set $\Lambda \subset L^2(\mathbb{R}^+)$ and are given by

$$w^{\pm} f = \int_0^{\infty} \phi(k,r) \tilde{f}(k)\, dk. \tag{10}$$

The solutions $\phi(k,r)$ of (2) may be chosen so that the map $f \to w^{\pm} f$ is an isometry.[7]

The next two sections of the paper are devoted to proving these theorems. The idea for the definition of $u(t)$ is based on a technique of Zinnes and Mulherin.[8]

Product integral notation is used throughout to facilitate the study of the Schrödinger equation. Numerous references to product integration are available in the literature.[6,9,10] Frequent reference is made to a previous work[4] of the author which contains a list of useful properties of the product integral. This work will hereafter be labeled I, and the properties labeled I (i), I (ii), etc.

The reader may regard the product integral formalism merely as a useful notation for expressing solutions of ordinary differential equations. The situation is analogous to that of defining the antiderivative F of a function f as $F(x) = \int_a^x f(s)\,ds$ instead of saying that F satisfies $F'(x) = f(x)$. The integral notation is concise and lends itself to several quite natural properties and estimates, such as $\int f + g = \int f + \int g$ and $|\int f| < \int |f|$.

The product integral notation has analogous attributes. Much of our experience and intuition with integrals may be carried over to product integrals, provided that we keep in mind that the latter is derived from a product rather than a sum. Briefly, the product integral is the propagator of the solution of the differential equation.

2. ASYMPTOTIC FORM OF THE SOLUTIONS OF THE RADIAL SCHRÖDINGER EQUATION

This section is devoted to proving Theorem 1. In this proof, attention is focused on a fixed value of k, so that the implicit dependence of functions and constants on k is sometimes suppressed in the notation. The k dependence will become important later, in the proof of Theorem 2. For simplicity in the calculations, we shall take $l = 0$ in the proof of the theorem. The term $l(l+1)/r^2$ for $l \neq 0$ falls off at infinity sufficiently fast to have no effect on the result.

Before proving the theorem we state a slightly sharper version of a lemma proved in I by integration by parts.

Lemma 1: Let β, ρ, and c be constants with $\frac{1}{2} < \beta < 1$, $\frac{1}{2} < \rho < 1$, and $c = i\xi$, $\xi \in \mathbb{R}$, $\xi \notin \{0, \pm 1, \pm 2, \pm 3\}$. Let $\theta(r) = \int_r^\infty (\sin s)/s^\beta\,ds$. Then,

$$\begin{aligned}\int_r^\infty e^{cs}\frac{e^{-i\theta(s)/k}}{s^\rho}\,ds &= -\frac{e^{cr}}{c}\frac{e^{-i\theta(r)/k}}{r^\rho} + \frac{1}{2kc}\frac{e^{(c+i)r}}{c+i}\frac{e^{-i\theta(r)/k}}{r^{\rho+\beta}} \\ &\quad - \frac{1}{2kc}\frac{e^{(c-i)r}}{c-i}\frac{e^{-i\theta(r)/k}}{r^{\rho+\beta}} - \frac{\rho}{c}\frac{e^{cr}}{c}\frac{e^{-i\theta(r)/k}}{r^{\rho+1}} \\ &\quad - \frac{1}{4k^2c(c+i)(c+2i)}e^{(c+2i)r}\frac{e^{-i\theta(r)/k}}{r^{\rho+2\beta}} \\ &\quad + \frac{1}{2k^2c(c^2+1)}e^{cr}\frac{e^{-i\theta(r)/k}}{r^{\rho+2\beta}} \\ &\quad - \frac{1}{4k^2c(c-i)(c-2i)}e^{(c-2i)r}\frac{e^{-i\theta(r)/k}}{r^{\rho+2\beta}} \\ &\quad + O(r^{-\rho-\beta-1}).\end{aligned}$$

Proof of Theorem 1: Let

$$A(r) = \begin{pmatrix} 0 & 1 \\ V(r) - k^2 & 0 \end{pmatrix}, \quad \Phi(r) = \begin{pmatrix} \phi(r) \\ \phi'(r) \end{pmatrix}.$$

Then, Eq. (2) has the standard matrix form

$$\Phi'(r) = A(r)\Phi(r), \tag{11}$$

and its solution is given by

$$\Phi(r) = \prod_a^r e^{A(s)\,ds}\Phi_a, \tag{12}$$

where $\Phi_a = \Phi(a)$, for any $a \in \mathbb{R}^+$. $\Pi_a^r e^{A(s)\,ds}$ is the product integral of A from a to r and as such is the propagator of the solution of (11).

We may write

$$\phi(r) = P_1\prod_a^r e^{A(s)\,ds}\Phi_a, \tag{13}$$

where P_1 is the projection operator on the first component, that is,

$$P_1\begin{pmatrix} x \\ y \end{pmatrix} = x.$$

In order to analyze $\phi(r)$, we study $\Pi_a^r e^{A(s)\,ds}$.

Let

$$M = \begin{pmatrix} 1 & 1 \\ ik & -ik \end{pmatrix}.$$

Then, by the similarity rule I(ii) we can write

$$\begin{aligned}\prod_a^r e^{A(s)\,ds} &= M\prod_a^r e^{\{M^{-1}A(s)M\}\,ds}M^{-1} \\ &= M\prod_a^r e^{\{A_1(s)+A_2(s)\}\,ds}M^{-1},\end{aligned} \tag{14}$$

where

$$A_1(s) = \left(ik - \frac{iV(s)}{2k}\right)\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

$$A_2(s) = \frac{iV(s)}{2k}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

By the sum rule I (iii),

$$\prod_a^r e^{\{A_1(s)+A_2(s)\}\,ds} = Q(r)\prod_a^r e^{B(s)\,ds},$$

where

$$\begin{aligned}Q(r) &= \prod_a^r e^{A_1(s)\,ds} = e^{\int_a^r A_1(s)\,ds} \quad \text{[by I (vii)]} \\ &= \begin{pmatrix} e^{ik(r-a)}e^{-\frac{i}{2k}\int_a^r V(s)\,ds} & 0 \\ 0 & e^{-ik(r-a)}e^{\frac{i}{2k}\int_a^r V(s)\,ds} \end{pmatrix} \\ &= \begin{pmatrix} d_+e^{i\theta(r)/2k}e^{ikr} & 0 \\ 0 & d_-e^{-i\theta(r)/2k}e^{-ikr} \end{pmatrix},\end{aligned} \tag{15}$$

$$\begin{aligned}B(s) &= (Q^{-1}A_2Q)(s) \\ &= V(s)\begin{pmatrix} 0 & \gamma_-e^{-i\theta(s)/k}e^{-2iks} \\ \gamma_+e^{i\theta(s)/k}e^{2iks} & 0 \end{pmatrix},\end{aligned} \tag{16}$$

and where we have defined

$$\theta(r) = \int_r^\infty V(s)\,ds,$$

$$d_\pm = e^{\mp ika \mp \frac{i}{2k}\int_a^\infty V(s)\,ds},$$

and

$$\gamma_\pm = \pm(i/2k)d_\pm^2. \quad (17)$$

That $\int_r^\infty V(s)\,ds$ exists follows from an elementary integration by parts. Note that $\gamma_- = \bar{\gamma}_+$ and $|\gamma_\pm|^2 = 1/4k^2$.

Now let

$$f(s) = \gamma_- V(s) e^{-i\theta(s)/k} e^{-2iks}, \quad (18)$$

where $V(s) = (\sin s)/s^\beta$, $\frac{1}{3}<\beta\leqslant\frac{1}{2}$. Combination of (14), (15), (16), and (18) yields

$$\prod_a^r e^{A(s)\,ds} = MQ(r)\prod_a^r e^{B(s)\,ds}M^{-1}, \quad (19)$$

where

$$B(s) = \begin{pmatrix} 0 & f(s) \\ \bar{f}(s) & 0 \end{pmatrix}. \quad (20)$$

We see that the study of $\Pi_a^r e^{A(s)\,ds}$ reduces to the study of $\Pi_a^r e^{B(s)\,ds}$. The latter will be undertaken now, using the product integral identity

$$\prod_a^r e^{B(s)\,ds} = [I + H(r)]\prod_a^r e^{\{H(s)B(s)[I+H(s)]^{-1}\}\,ds}(I + H(a)), \quad (21)$$

for a and r sufficiently large and $H(r) = \int_r^\infty B(s)\,ds$. We shall exploit the fact that B has only off-diagonal elements.

In order that (21) be applicable, it must be shown that $H(r) = \int_r^\infty B(s)\,ds$ exists, where the integral may be taken componentwise, which is to say that

$$H(r) = \begin{pmatrix} 0 & \int_r^\infty f(s)\,ds \\ \int_r^\infty \bar{f}(s)\,ds & 0 \end{pmatrix} \quad (22)$$

exists. For this, it suffices to show that

$$\int_r^\infty f(s)\,ds = \gamma_- \int_r^\infty V(s)e^{-i\theta(s)/k}e^{-2iks}\,ds$$

exists. By writing

$$V(s) = (\sin s)/s^\beta = (1/s^\beta)\{(e^{is} - e^{-is})/2i\},$$

we may express

$$\int_r^\infty f(s)\,ds = \frac{\gamma_-}{2i}\int_r^\infty e^{-i(2k-1)s}\frac{e^{-i\theta(s)/k}}{s^\beta}\,ds - \frac{\gamma_-}{2i}\int_r^\infty e^{-i(2k+1)s}\frac{e^{-i\theta(s)/k}}{s^\beta}\,ds.$$

If we use Lemma 1 on each of these terms and then recombine exponentials to form sines and cosines, we obtain (after laborious yet elementary calculations)

$$\int_r^\infty f(s)\,ds = \gamma_-(k)e^{-2ikr}e^{-i\theta(r)/k}z(k,r) + z_1(k,r), \quad (23)$$

where

$$\begin{aligned} z(k,r) &= -(4k^2-1)^{-1}(1/r^\beta)[2ik\sin r + \cos r] \\ &\quad + [2k(4k^2-1)(4k^2-4)]^{-1}(1/r^{2\beta})\{-(4k^2+2)\sin 2r \\ &\quad + 6ik\cos 2r\} - i[4k^2(4k^2-1)]^{-1}(1/r^{2\beta}) \\ &\quad + \beta(4k^2-1)^{-2}(1/r^{\beta+1})\{4k\sin r - i(4k^2+1)\cos r\} \\ &\quad + [8k^3(4k^2-1)]^{-1}(1/r^{3\beta})\cos r \\ &\quad + [4k^2(4k^2-1)(4k^2-4)(4k^2-9)]^{-1} \\ &\quad \times(1/r^{3\beta})\{i(24k^2+6)\sin 3r + (8k^3+22k)\cos 3r\} \\ &\quad - [8k^3(k^2-1)(4k^2-1)]^{-1}(1/r^{3\beta})\{3ik\sin r \\ &\quad + (4k^2+1)\cos r\} \end{aligned} \quad (24)$$

and

$$z_1(k,r) = O(r^{-4\beta}) \quad \text{as } r\to\infty. \quad (25)$$

This shows that $H(r)$ exists, which means that we are now ready to apply (21). We have from (20) and (22) that

$$HB(r) = \begin{pmatrix} f(r)\int_r^\infty f(s)\,ds & 0 \\ 0 & f(r)\int_r^\infty \bar{f}(s)\,ds \end{pmatrix}. \quad (26)$$

We may write

$$HB(I+H)^{-1} = HB - HBH(I+H)^{-1}$$

and use the sum rule I (iii) to estimate $\Pi_a^r e^{HB(I+H)^{-1}\,ds}$. Since HB is diagonal, we may use I (vii) to calculate its product integral directly, i.e.,

$$\prod_a^r e^{HB\,ds} = e^{(\int_a^r HB\,ds)}.$$

We therefore turn our attention to the calculation of

$$\int_a^r \bar{f}(s)\left\{\int_s^\infty f(u)\,du\right\}ds.$$

Using (18) and (23)–(25), we may write

$$\bar{f}(k,s)\int_s^\infty f(k,u)\,du = \left(\frac{1}{4k^2}\right)[(\sin s)/s^\beta]z(k,s) + O(s^{-5\beta}) \quad (27)$$

as $s\to\infty$.

For our present purpose of calculating the asymptotic form of solutions of (2), we need only be concerned with the first term of $z(k,s)$ [see Eq. (24)] in (27), as this is the only term in (27) which is not conditionally integrable. This is true because $\sin^2 s = \frac{1}{2} - \frac{1}{2}\cos 2s$, and $\int_r^\infty s^{-2\beta}\,ds$ does not exist for $\frac{1}{3}<\beta\leqslant\frac{1}{2}$. In the proof of the existence of the modified wave operators in Sec. 3, we shall have to make a more precise estimate of (27).

With this in mind, let us define

$$\gamma(k) = -[4k(4k^2-1)]^{-1}, \quad (28)$$

$$v(r) = \int_a^r s^{-2\beta}\,ds, \quad (29)$$

$$Y = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (30)$$

$$C = i\gamma(k)r^{-2\beta}Y, \quad (31)$$

$$D = HB - C - HBH(I+H)^{-1}, \quad (32)$$

and write

$$HB(I+H)^{-1} = C + D. \quad (33)$$

Note that $D = D(k,s) = O(s^{-3\beta})$ as $s\to\infty$.

Using the sum rule

$$\prod_a^r e^{[HB(I+H)^{-1}]\,ds} = \prod_a^r e^{[C+D]\,ds}$$
$$= T(k,r)\prod_a^r e^{[T^{-1}DT]\,ds},$$

where

$$T(k,r) = \prod_a^r e^{C\,ds}. \tag{34}$$

By I (vii), we have

$$T(k,r) = e^{(\int_a^r C\,ds)}$$
$$= \begin{pmatrix} e^{i\gamma(k)v(r)} & 0 \\ 0 & e^{-i\gamma(k)v(r)} \end{pmatrix}.$$

Since T is bounded, we have that

$$(T^{-1}DT)(k,s) = O(s^{-3\beta}) \text{ as } s \to \infty.$$

Therefore, $T^{-1}DT \in L^1[(a,\infty)]$, which implies by I (iv) the existence of $\Pi_a^\infty e^{\{T^{-1}DT\}\,ds} \equiv \lim_{r\to\infty} \Pi_a^r e^{\{T^{-1}DT\}\,ds}$. Putting together (13), (19), (21), (33), and (34), we have

$$\phi(k,r)$$
$$= P_1 MQ(I+H)^{-1}T\prod_a^r e^{\{T^{-1}DT\}\,ds}(I+H(a))M^{-1}\Phi_a. \tag{35}$$

Let

$$\Upsilon = \Upsilon(k) = P_1 M, \tag{36}$$

$$G = G(k,r) = \prod_a^r e^{\{T^{-1}DT\}\,ds}, \tag{37}$$

$$\hat{G} = \hat{G}(k) = \prod_a^\infty e^{\{T^{-1}DT\}\,ds}, \tag{38}$$

$$L = L(k) = (I+H(a))M^{-1}\Phi_a. \tag{39}$$

We may now write

$$\begin{aligned}\phi(k,r) &= \Upsilon Q(I+H)^{-1}TGL \\ &= \Upsilon Q(I+H)^{-1}T\hat{G}L + \Upsilon Q(I+H)^{-1}T(G-\hat{G})L \\ &= \Upsilon QT\hat{G}L - \Upsilon QH(I+H)^{-1}T\hat{G}L \\ &\quad + \Upsilon Q(I+H)^{-1}T(G-\hat{G})L \\ &= \Upsilon PT\hat{G}L + \Upsilon(Q-P)T\hat{G}L \\ &\quad - \Upsilon QH(I+H)^{-1}T\hat{G}L \\ &\quad + \Upsilon Q(I+H)^{-1}T(G-\hat{G})L,\end{aligned} \tag{40}$$

where

$$P(k,r) = \begin{pmatrix} d_+(k)e^{ikr} & 0 \\ 0 & d_-(k)e^{-ikr} \end{pmatrix}$$

[see Eq. (15)]. (41)

If we let

$$\epsilon = \phi - \Upsilon PT\hat{G}L, \tag{42}$$

$$\epsilon_1 = \Upsilon Q(I+H)^{-1}T(G-\hat{G})L, \tag{43}$$

$$\epsilon_2 = \Upsilon(Q-P)T\hat{G}L, \tag{44}$$

$$\epsilon_3 = -\Upsilon QH(I+H)^{-1}T\hat{G}L, \tag{45}$$

then we have

$$\epsilon = \sum_{i=l}^{3} \epsilon_i. \tag{46}$$

Since

$$\Upsilon PT\hat{G}L = c_+(k)e^{ikr+i\gamma(k)v(r)} + c_-(k)e^{-ikr-i\gamma(k)v(r)}, \tag{47}$$

the proof of the theorem will be complete if we can show that $\epsilon \to 0$ as $r\to\infty$. But this follows easily if we note that each of the factors in each ϵ_i is bounded and each ϵ_i has a factor that tends to 0 as $r\to\infty$. A detailed analysis of all of these terms will be given in the next section, where we prove the existence of the modified wave operators.

3.THE WAVE OPERATORS

In this section, we prove Theorem 2. By construction, the wave operators $W^\pm$ defined by piecing together the various $w^\pm$ given in (9) and (10) for each partial wavespace S_{lm} satisfy the usual results of scattering theory. That is, the operators are norm-preserving [provided the solutions $\phi(k,r)$ of (2) are chosen properly] and they satisfy the intertwining relations. The proof, however, shows that these operators may be used to approximate the asymptotic behavior of a particle whose initial prepared state belongs only to a dense set in $L^2(\mathbb{R}^3)$ rather than to all of $L^2(\mathbb{R}^3)$. Physically, this drawback should have little effect, since any initial state can be approximated arbitrarily closely by a state in the dense set.

We note that, as in the proof of Theorem 1, we carry out the calculations for the case $l=0$. The results remain valid for $l\neq 0$. In addition, we focus attention on w^+ with the observation that the proof for w^- is entirely analogous.

Proof of Theorem 2: We wish to show that

$$\lim_{t\to\infty} \| [e^{ith}u(t)f](r) - \int_0^\infty \phi(k,r)\tilde{f}(k)\,dk \|_{L^2(\mathbb{R}^+)} = 0$$

[See (5)–(10)]. This is equivalent to showing that

$$\begin{aligned}&\lim_{t\to\infty} \| [u(t)f](r) - e^{-ith}\int_0^\infty \phi(k,r)\tilde{f}(k)\,dk \|_{L^2(\mathbb{R}^+)} \\ &= \lim_{t\to\infty} \| [u(t)f](r) - \int_0^\infty \phi(k,r)e^{-itk^2}\tilde{f}(k)\,dk \|_{L^2(\mathbb{R}^+)} \\ &= \lim_{t\to\infty} \| \int_0^\infty [\phi(k,r)-a(k,r)]e^{-itk^2}\tilde{f}(k)dk \|_{L^2(\mathbb{R}^+)} = 0.\end{aligned}$$

A proof that allows us to bring e^{-ith} through the integral sign in the first equality above is given in Ref. 6. If we let

$$\epsilon(k,r) = \phi(k,r) - a(k,r)$$

[see (8), (42), and (47)] and

$$g(r,t) = \int_0^\infty \epsilon(k,r)e^{-itk^2}\tilde{f}(k)\,dk, \tag{48}$$

then our goal of proving Theorem 2 will be accomplished by showing that

$$\lim_{t\to\infty} \|g(r,t)\|_{L^2(\mathbb{R}^+)} = 0. \tag{49}$$

Using (42)–(46) and letting

$$g_i(r,t) = \int_0^\infty \epsilon_i(k,r)e^{-itk^2}\tilde{f}(k)\,dk, \tag{50}$$

we see that we may write

$$g(r,t) = \sum_{i=1}^{3} g_i(r,t).$$

To complete the proof, we will show that

$$\lim_{t\to\infty} \|g_i(r,t)\|_{L^2(\mathbb{R}^+)} = 0, \quad i = 1, 2, 3.$$

First, we indicate some additional notation and state and prove a lemma found in I. For the remainder of the paper, we interpret the statement

$$p(k,r) = O(q(r))$$

to mean that there exist constants C and R in $\mathbb{R}^+$ such that $r > R$ implies

$$p(k,r) \leqslant Cq(r)$$

independent of $k \in \operatorname{supp} \tilde{f}$, where $f \in \Lambda$ is fixed. We introduce the symbol $L_a^p(dm)$ (or simply L_a^p) to denote $L^p([a,\infty); dm)$, for $a, p \in \mathbb{R}^+$.

Lemma 2: Let $P(r,t) = \int_0^\infty p(k,r) e^{-itk^2} \tilde{f}(k)\, dk, f \in \Lambda$, $|p(k,r)| < B$, $\forall r \in \mathbb{R}^+$, $\forall k \in \operatorname{supp} \tilde{f}$, $B = \text{const}$. Suppose that

$$\int_a^\infty |P(r,t)|^2\, dr < q(a),$$

where $\lim_{a\to\infty} q(a) = 0$. Then

$$\lim_{t\to\infty} \|P(r,t)\|_{L^2(\mathbb{R}^+)} = 0.$$

Proof of Lemma 2:

$$\|P(r,t)\|^2_{L^2(\mathbb{R}^+)} = \int_0^a |P(r,t)|^2\, dr + \int_a^\infty |P(r,t)|^2\, dr, \quad \forall a \in \mathbb{R}^+.$$

The latter term can be made arbitrarily small independent of t by taking a large enough. The former term tends to 0 as $t\to\infty$ by the Riemann–Lebesgue lemma and the Lebesgue dominated convergence theorem because, for each $r \in \mathbb{R}^+$, $P(r,t)$ is the Fourier transform at t of a function in $L^2(\mathbb{R}^+; dk)$, and $P(r,t)$ is bounded on $(0,a)$.

Remark: If $p(k,r) = O(q(r))$, where $q \in L_a^2(dr)$, and if p is bounded, then p satisfies the hypotheses of Lemma 3a.

We shall now prove (49) with three lemmas.

Lemma 3a: $\lim_{t\to\infty} \|g_1(r,t)\|_{L^2(\mathbb{R}^+)} = 0.$

Proof: We have

$$g_1(r,t) = \int_0^\infty \epsilon_1(k,r) e^{-itk^2} \tilde{f}(k)\, dk, \tag{51}$$

where $\epsilon_1(k,r) = \gamma Q (I+H)^{-1} T (G - \widehat{G}) L$. We use the definitions (37) and (38) and the property I (i) to write

$$G - \widehat{G} = \left\{ \prod_\infty^r e^{\{T^{-1}DT\}\, ds} - I \right\} \widehat{G}. \tag{52}$$

We shall write D [defined in (32)] as a sum of three terms, i.e.,

$$D = D_1 + D_2 + D_3, \tag{53}$$

where

$$D_1(k,r) = i\gamma_1(k) r^{-4\beta} Y \ \text{[see (30) and (66)]}, \tag{53a}$$

$$D_2 = U D_4 U^{-1}, \ \text{with} \int_r^\infty D_4(k,s)\, ds = O(r^{-3\beta}), \tag{53b}$$

$$U = TT_1 \ \text{and} \ T_1(k,r) = \prod_\infty^r e^{D_1\, ds},$$

$$D_3(k,r) = O(r^{-5\beta}). \tag{53c}$$

That (53a)–(53c) suffice to complete the proof of Lemma 3a is given as follows. Using (53a) and the sum rule I (iii), we may write

$$\begin{aligned} &\prod_\infty^r e^{\{T^{-1}DT\}\, ds} - I \\ &\quad = T_1 \prod_\infty^r e^{\{U^{-1}(D_2 + D_3)U\}\, ds} - I \\ &\quad = (T_1 - I) + T_1 \left(\prod_\infty^r e^{\{U^{-1}(D_2 + D_3)U\}\, ds} - I \right). \end{aligned} \tag{54}$$

Since D_1 is diagonal, we may use I (vii) to write

$$\begin{aligned} T_1(k,r) &= \prod_\infty^r e^{D_1\, ds} \\ &= e^{\int_\infty^r D_1\, ds} \\ &= \begin{pmatrix} e^{i\gamma_1(k) v_1(r)} & 0 \\ 0 & e^{-i\gamma_1(k) v_1(r)} \end{pmatrix}, \end{aligned} \tag{55}$$

where

$$v_1(r) = \int_\infty^r s^{-4\beta}\, ds. \tag{56}$$

Therefore,

$$T_1 - I = i\gamma_1(k) v_1(r) Y + W_1(k,r), \tag{57}$$

where

$$W_1(k,r) = O(v_1^2(r)) = O(r^{-8\beta+2}). \tag{58}$$

Now use the sum rule and (53b) on the second term of the rhs of (54) to get

$$\begin{aligned} &T_1 \left(\prod_\infty^r e^{\{U^{-1}(D_2 + D_3)U\}\, ds} - I \right) \\ &\quad = T_1 \left(\prod_\infty^r e^{\{D_4 + U^{-1} D_3 U\}\, ds} - I \right) \\ &\quad = T_1 \left(T_4 \prod_\infty^r e^{\{T_4^{-1} U^{-1} D_3 U T_4\}\, ds} - I \right) \\ &\quad = T_1 \left[T_4 - I + T_4 \left(\prod_\infty^r e^{\{T_4^{-1} U^{-1} D_3 U T_4\}\, ds} - I \right) \right], \end{aligned} \tag{59}$$

where $T_4(k,r) = \Pi_\infty^r e^{D_4\, ds}$. We may write T_4 as a "time-ordered exponential," i.e.,

$$T_4(k,r) = I - \int_r^\infty D_4(k,s_1)\, ds + \int_r^\infty D_4(k,s_1) \times \left\{ \int_{s_1}^\infty D_4(k,s_2)\, ds_2 \right\} ds_1 - \cdots.$$

From (53b), we have $\int_r^\infty D_4(k,s)\, ds = O(r^{-3\beta})$ and thus,

$$T_4(k,r) - I = O(r^{-3\beta}). \tag{60}$$

Finally, using (53c), I (vi), and the fact that U and T_4 are bounded, we have

$$\left|\left|\prod_{\infty}^{r} e^{\{T_4 U^{-1} D_3 U T_4\}\,ds} - I\right|\right| < e^{\int_r^\infty \|T_4 U^{-1} D_3 U T_4\| ds} - 1$$
$$= O(r^{-5\beta+1}). \qquad (61)$$

Putting (52), (54), (57), and (59) together, we may write

$$G - \widehat{G} = G_4 + G_5,$$

where

$$G_4 = i\gamma_1(k)v_1(r)Y\widehat{G}, \qquad (62)$$

$$G_5 = W_1\widehat{G} + T_1(T_4 - I)\widehat{G} + T_1T_4\left(\prod_{\infty}^{r} e^{\{T_4^{-1}U^{-1}D_3UT_4\}\,ds} - I\right)\widehat{G}. \qquad (63)$$

From (58), (60) and (61), it follows that

$$G_5(k,r) = O(r^{-2\beta}). \qquad (64)$$

We may now write (51) as

$$g_1 = g_4 + g_5,$$

where

$$g_4(r,t) = \int_0^\infty \epsilon_4(k,r)e^{-itk^2}\tilde{f}(k)\,dk$$

and

$$g_5(r,t) = \int_0^\infty \epsilon_5(k,r)e^{-itk^2}\tilde{f}(k)\,dk$$

with

$$\epsilon_4(k,r) = \Upsilon Q(I+H)^{-1}TG_4L$$

and

$$\epsilon_5(k,r) = \Upsilon Q(I+H)^{-1}TG_5L.$$

For Lemma 2 and (64), it follows that

$$\lim_{t\to\infty} \|g_5(r,t)\|_{L^2(\mathbf{R}^+)} = 0.$$

We must now show that

$$\lim_{t\to\infty} \|g_4(r,t)\|_{L^2(\mathbf{R}^+)} = 0.$$

Let us write

$$\epsilon_4 = \epsilon_6 + \epsilon_7,$$

where

$$\epsilon_6 = \Upsilon PTG_4L,$$

$$\epsilon_7 = -\Upsilon PH(I+H)^{-1}TG_4L + \Upsilon(Q-P)(I+H)^{-1}TG_4L.$$

From (62), (56), (22), (23), (15), and (41), it follows that

$$\epsilon_7(k,r) = O(r^{-2\beta}).$$

Therefore, $\lim_{t\to\infty}\left|\left|\int_0^\infty \epsilon_7(k,r)e^{-itk^2}\tilde{f}(k)\,dk\right|\right|_{L^2(\mathbf{R}^+)} = 0.$
We must show that

$$\lim_{t\to\infty} \|g_6(r,t)\|_{L^2(\mathbf{R}^+)} = 0,$$

where $g_6(r,t) = \int_0^\infty \epsilon_6(k,r)e^{-itk^2}\tilde{f}(k)\,dk$. We have that

$$g_6(r,t) = v_1(r)\int_0^\infty \Upsilon PTY\widehat{G}Le^{-itk^2}\{i\gamma_1(k)\tilde{f}(k)\}\,dk.$$

Now, $g_6(r,t)$ is a function which has the same form as $v_1(r)[u(t)q](r)$, where $\tilde{q}(k) = i\gamma_1(k)\tilde{f}(k)$ and $u(t)$ is defined in (7). In fact, $\Upsilon PTY\widehat{G}L$ and $\Upsilon PT\widehat{G}L$ differ only by the constant factor Y, i.e.,

$$\Upsilon PT\widehat{G}L = c_+(k)e^{ikr + i\gamma(k)v(r)} + c_-e^{-ikr - i\gamma(k)v(r)}.$$

and

$$\Upsilon PTY\widehat{G}L = c_+(k)e^{-ikr + i\gamma(k)v(r)} - c_-(k)e^{-ikr - i\gamma(k)v(r)}.$$

Therefore, it suffices to show that

$$\lim_{t\to\infty} \|v_1(r)[u(t)f](r)\|_{L^2(\mathbf{R}^+)} = 0, \quad \forall f \in \Lambda.$$

Using (7) and (48), we may write

$$[u(t)f](r) = \int_0^\infty \phi(k,r)e^{-itk^2}\tilde{f}(k)\,dk - g(r,t),$$

so that

$$v_1(r)[u(t)f](r) = v_1(r)\int_0^\infty \phi(k,r)e^{-itk^2}\tilde{f}(k)\,dk - v_1(r)g(r,t),$$

By Lemma 2,

$$\lim_{t\to\infty} \|v_1(r)\int_0^\infty \phi(k,r)e^{-itk^2}\tilde{f}(k)\,dk\|_{L^2(\mathbf{R}^+)} = 0,$$

because

$$\left|\left|\int_0^\infty \phi(k,r)e^{-itk^2}\tilde{f}(k)\,dk\right|\right|_{L^2(\mathbf{R}^+)} = \|f\|_{L^2(\mathbf{R}^+)},$$

for $\phi(k,r)$ chosen properly.[7] Also by Lemma 2,

$$\lim_{t\to\infty} \|v_1(r)g(r,t)\|_{L^2(\mathbf{R}^+)} = 0,$$

because

$$v_1(r)g(r,t) = \sum_{i=1}^{3} v_1(r)g_i(r,t),$$

and examination of (56), (50), and (43)–(45) reveals that

$$v_1(r) = O(r^{-\beta})$$

and

$$\epsilon_i(r,t) = O(r^{-\beta}), \quad i = 1,2,3. \qquad (65)$$

This shows that the statements (53a)–(53d) imply the proof of Lemma 3a. We must verify these statements.

Recall from (32) that

$$D = HB - C - HBH(I+H)^{-1}.$$

We use the $(\sin r)/r^{3\beta}$ term of $z(k,r)$ defined in (24) along with (23)–(27) to motivate the following definitions.

$$\gamma_1(k) = -3[64k^4(k^2-1)(4k^2-1)]^{-1},$$

$$D_1(k,r) = i\gamma_1(k)r^{-4\beta}Y,$$

$$Z_1(k,r) = (4k^2r^\beta)^{-1}\begin{pmatrix} z_1(k,r) & 0 \\ 0 & \bar{z}_1(k,r)\end{pmatrix} = O(r^{-5\beta}),$$

$$Z_2 = Z_3 - Z_4, \qquad (66)$$

$Z_3 =$ the sum of all terms of HBH which are $O(r^{-5\beta})$,

$Z_4 =$ the sum of all terms of HBH^2 which are $O(r^{-5\beta})$,

$$D_2 = HB - C - D_1 - Z_1 - HBH + HBH^2 + Z_2,$$
$$D_3 = Z_1 - Z_2 - HBH^3(I+H)^{-1}.$$

From these definitions, it follows immediately that

$$D = D_1 + D_2 + D_3,$$

and that $D_3(k,r) = O(r^{-5\beta})$. Thus, we have (53a), and (53c). For the sake of brevity, the proof of (53b) will not be given here. The proof is elementary, consisting of nothing more than several judicious integrations by parts of many of the terms in

$$(1/4k^2)[(\sin r)/r^\beta]z(k,r),$$
$$\gamma_-(k)e^{-2ikr}e^{-i\theta(r)/k},$$
$$(1/4k^2)[(\sin r)/r^\beta]z^2(k,r),$$

and

$$(1/4k^2)e^{-4ikr}e^{-2i\theta(r)/k}(1/4k^2)\,[(\sin r)/r^\beta]z^3(k,r).$$

The only thing that has to be checked carefully is that none of these integrals results in terms such as $v(r)$ and $v_1(r)$ defined in (4) and (56), respectively. This completes the proof of Lemma 3a.

Lemma 3b: $\lim_{t\to\infty} \|g_2(r,t)\|_{L^2(\mathbf{R}^+)} = 0.$

Proof: We have

$$g_2(r,t) = \int_0^\infty \epsilon_2(k,r)e^{-itk^2}\tilde{f}(k)\,dk,$$

where $\epsilon_2(k,r) = [\Upsilon(Q-P)T\hat{G}L](k,r)$. Using (15) and (41), we may write

$$Q - P = P\begin{pmatrix} e^{i\theta(r)/2k} - 1 & 0 \\ 0 & e^{-i\theta(r)/2k} - 1 \end{pmatrix}$$
$$= [i\theta(r)/2k]PY - O(\theta^2(r)),$$

which gives

$$\epsilon_2(k,r) = [i\theta(r)/2k]\,\Upsilon PTY\hat{G}L + O(r^{-2\beta}).$$

Therefore,

$$g_2(r,t) = \theta(r)\int_0^\infty \Upsilon PTY\hat{G}Le^{-itk^2}\left(\frac{i}{2k}\tilde{f}(k)\right)dk + \int_0^\infty O(r^{-2\beta})e^{-itk^2}\tilde{f}(k)\,dk.$$

By Lemma 2,

$$\lim_{t\to\infty}\left|\left|\int_0^\infty O(r^{-2\beta})e^{-itk^2}\tilde{f}(k)\,dk\right|\right|_{L^2(\mathbf{R}^+)} = 0.$$

The other term is similar to $g_6(r,t)$, discussed in the proof of Lemma 3a. The constant factor Y may be disregarded (without loss of generality), and we have

$$\theta(r)\int_0^\infty \Upsilon PT\hat{G}Le^{-itk^2}\left(\frac{i}{2k}\tilde{f}(k)\right)dk = \theta(r)[u(t)q](r),$$

where $\tilde{q}(k) = \dfrac{i}{2k}\tilde{f}(k)$ and $u(t)$ is defined in (7). As in the proof of Lemma 3a, we have that

$$\lim_{t\to\infty}\|\theta(r)[u(t)q](r)\|_{L^2(\mathbf{R}^+)} = 0.$$

This completes the proof of Lemma 3b.

Lemma 3c: $\lim_{t\to\infty}\|g_3(r,t)\|_{L^2(\mathbf{R}^+)} = 0.$

Proof: We have

$$g_3(r,t) = \int_0^\infty \epsilon_3(k,r)e^{-itk^2}\tilde{f}(k)\,dk,$$

where $\epsilon_3(k,r) = [-\Upsilon QH(I+H)^{-1}T\hat{G}L](k,r)$. We may write

$$\epsilon_3 = \epsilon_8 + \epsilon_9,$$

where

$$\epsilon_8 = -\Upsilon PHT\hat{G}L$$

and (67)

$$\epsilon_9 = \Upsilon PH^2(I+H)^{-1}T\hat{G}L - \Upsilon(Q-P)H(I+H)^{-1}T\hat{G}L.$$

Since $\epsilon_9(k,r) = O(r^{-2\beta})$, we have by Lemma 2 that

$$\lim_{t\to\infty}\left|\left|\int_0^\infty \epsilon_9(k,r)e^{-itk^2}\tilde{f}(k)\,dk\right|\right|_{L^2(\mathbf{R}^+)}.$$

We consider the term

$$g_8(r,t) = \int_0^\infty \epsilon_8(k,r)e^{-itk^2}\tilde{f}(k)\,dk.$$

From (41), (23), and (24), we see that

$$-PH = \begin{pmatrix} 0 & -d_+e^{ikr}\int_r^\infty f(k,s)\,ds \\ -d_-e^{-ikr}\int_r^\infty \overline{f}(k,s)\,ds & 0 \end{pmatrix}$$
$$= r^{-\beta}SP + O(r^{-2\beta}), \tag{68}$$

where

$$S(k,r) = \begin{pmatrix} 0 & \tau_-(k)(2ik\sin r + \cos r) \\ \tau_+(k)(-2ik\sin r + \cos r) & 0 \end{pmatrix}$$

and $\tau_\pm(k) = \pm[(4k^2-1)2ik]^{-1}$ [See Eq. (17)]. Putting (68) into (67) yields

$$\epsilon_8 = \epsilon_{10} + \epsilon_{11},$$

where

$$\epsilon_{10} = r^{-\beta}\Upsilon SPT\hat{G}L$$

and

$$\epsilon_{11} = O(r^{-2\beta}).$$

By Lemma 2,

$$\lim_{t\to\infty}\left|\left|\int_0^\infty \epsilon_{11}(k,r)e^{-itk^2}\tilde{f}(k)\,dk\right|\right|_{L^2(\mathbf{R}^+)} = 0.$$

Except for the factor $r^{-\beta}S$, the term ϵ_{10} is the same as

$$(\Upsilon PT\hat{G}L)(k,r)$$
$$= c_+(k)e^{ikr + i\gamma(k)v(r)} + c_-(k)e^{-ikr - i\gamma(k)v(r)}.$$

In fact,

$$r^{-\beta}\Upsilon SPT\hat{G}L$$
$$= r^{-\beta}\sin r\,\{a_+(k)e^{ikr+i\gamma(k)v(r)} + a_-(k)e^{-ikr-i\gamma(k)v(r)}\} + r^{-\beta}\cos r\{b_+(k)e^{ikr+i\gamma(k)v(r)} + b_-(k)e^{-ikr-i\gamma(k)v(r)}\},$$

where

$$a_\pm(k) = -(4k^2-1)^{-1}c_\pm(k)$$

and

$$b_\pm(k) = \pm[2ik(4k^2-1)]^{-1}c_\pm(k).$$

Just as in the proof in Lemma 3a, it can be easily seen using Lemma 2 that

$$\lim_{t\to\infty} \|r^{-\beta}\sin r[u(t)f](r)\|_{L^2(\mathbb{R}^+)} = 0$$

and

$$\lim_{t\to\infty} \|r^{-\beta}\cos r[u(t)f](r)\|_{L^2(\mathbb{R}^+)} = 0, \quad \forall f\in\Lambda.$$

This implies that

$$\lim_{t\to\infty} \left|\left|\int_0^\infty \epsilon_{10}(k,r)e^{-itk^2}\tilde{f}(\mathrm{k})\,dk\right|\right|_{L^2(\mathbb{R}^+)} = 0$$

and, hence, that

$$\lim_{t\to\infty} \|g_3(r,t)\|_{L^2(\mathbb{R}^+)} = 0,$$

which completes the proof of Lemma 3c and of Theorem 2.

Using (10), we may extend the definition of our modified wave operators (9) on each partial wavespace from the dense set $\Lambda \subset L^2(\mathbb{R}^+)$ to all of the $L^2(\mathbb{R}^+)$. Explicitly, for $f\in L^2(\mathbb{R}^+)$, we define

$$w^{\pm}f = \underset{R\to\infty}{\mathrm{l.i.m.}}\int_0^R \phi(k,r)\tilde{f}(k)\,dk. \tag{69}$$

We conjecture that

$$[u(t)f](r) = \underset{R\to\infty}{\mathrm{l.i.m.}}\int_0^R a(k,r)e^{-itk^2}\tilde{f}(k)\,dk$$

[see (7)] exists in $L^2(\mathbb{R}^+)$ and that

$$w^{\pm}f = \lim_{t\to\infty} e^{ith}u(t)f.$$

This does not appear to be an immediate consequence of Theorem 2, because the modified free propagator $u(t)$ is not a simple function of the usual free propagator as, for example, in the Coloumb case.[11]

With this extended definition of the modified wave operators, we conclude easily from (69) that the range of both operators is the absolutely continuous subspace of h. We may define the scattering operator $S = (w^+)^* w^-$ on each partial wavespace and conclude that S is unitary. We thus have completeness of the wave operators. The question of asymptotic completeness, i.e., absence of the essential spectrum, requires additional investigation.

ACKNOWLEDGMENTS

The author would like to thank J. D. Dollard and F. J. Narcowich for several valuable discussions.

[1] M. N. Hack, "On convergence to the Möller wave operators," Nuovo Cimento **9**, 731–733 (1958).

[2] V. S. Buslaev and V. B. Matveev, "Wave operators for the Schrödinger equation with slowly decreasing potential," Teoret. Mat. Fiz. **2**, 367–376 (1970).

[3] M. Reed and B. Simon, *Scattering Theory* (Academic, New York, 1979).

[4] B. Bourgeois, "On scattering theory for oscillatory potentials of slow decay," Ann. Phys. (N.Y.) **121**, 415–431 (1979). This reference is referred to as I.

[5] A. Devinatz, "The existence of wave operators for oscillating potentials," J. Math. Phys. **21**, 2406–2411 (1980).

[6] J. D. Dollard and C. N. Friedman, "Existence of the Möller wave operators for $V(r) = \lambda \sin \mu r^\alpha / r^\beta$," Ann. Phys. (N.Y.) **111**, 251–266, (1978).

[7] E. C. Titchmarsh, *Eigenfunction Expansions Associated with Second Order Differential Equations* (Clarendon, Oxford, 1946),Vol. 1.

[8] D. Muhlerin and I. Zinnes, "Coulomb scattering I," J. Math. Phys. **11**, 1402–1408 (1970).

[9] J. D. Dollard and C. N. Friedman, "On strong product integration," J. Func. Anal. **28**, 309–354 (1978).

[10] J. D. Dollard and C. N. Friedman, *Product Integration* (Addison-Wesley, Reading, Mass., 1979).

[11] J. D. Dollard, "Asymptotic convergence and the Coulomb interaction," J. Math. Phys. **5**, 729–738 (1964).

Generalized completeness relations in the theory of resonant scattering

O. Zohni[a)]
USDOE-Ames Laboratory and Department of Chemistry, Iowa State University, Ames, Iowa 50011 [b)]

(Received 15 October 1980; accepted for publication 7 January 1982)

Generalized completeness relations involving resonance states are constructed within the framework of analytically continued symmetrized scattering kernels into the unphysical sheet of the complex-energy plane. The bases states utilized are identified with complex-energy generalized eigenvectors over an extended or rigged Hilbert space. The resulting relations are uniquely defined and do not exhibit the usual divergence problems encountered with the regularization methods.

PACS numbers: 03.65.Nk

I. INTRODUCTION

Resonances occur in many places in nuclear and particle physics. Our main interest lies in the off-energy-shell continuation of the two-body t matrix in the presence of resonances (e.g., in π–N and N–N scattering). Some time ago, Haftel[1] extended the method of Baranger *et al.*[2] of continuing the off-shell t matrix when the uncoupled partial wave eigenchannel has only scattering states to the case when a bound state is also present. The same problem was considered by Sauer and Sevgen[3] in the presence of inelasticities. Working in the spirit of these approaches,[1–3] we need to have a completeness relation for resonance states which restrict the arbitrariness in the full off-shell t matrix to only the symmetric part of the off-shell function.[4,5]

Because of their violent behavior, resonance states are not usually used as a basis of eigenfunction expansion. It turns out that the main problem that arises when the scattering process is mediated by resonances (among other possible states) is the construction of appropriate completeness relations which bring about the presence of resonance states in an explicit way.

A number of studies[6–14] have been made to define the integral properties of resonance states and to construct appropriate completeness relations involving such states. Various restrictive conditions, however, are usually associated with such completeness relations (see, e.g., Ref. 10). This has consequently lead to nonuniqueness in establishing the orthogonality and completeness properties of resonance states. A brief outline of those attempts may clarify this point.

Berggren[6] made use of a regularization method due to Zel'dovich[7] to define the value of integrals over resonance states. Based on this regularization method, Berggren[6] also constructed a completeness relation in which a discrete set of bound (orthonormal) and resonant (biorthonormal) states is completed by a set of continuum states. The analytically continued complex energies and asymptotic wavenumbers of this continuum part are chosen along a specified integration contour which determines the properties of the expansion, and the complete set is not a simple extension of Hilbert space.

In another method, More and Gerjuoy[8] obtained anomalous normalizations of the resonance wavefunctions which are determined by the energy dependence of an effective Hamiltonian. Garcia-Calderon and Peierls[9] obtained normalization constants related to the residues of an outgoing Green function which defines the resonance states, and a completeness formula was derived[9,10] which involved a certain part of the continuum. Romo[11] used analytic continuation to develop a generalization of the inner-product integrals for resonances based on the S-matrix approach. Kim and Vasavada[12] worked also in the S-matrix formalism and treated the normalization of the resonant states differently. In the work of these authors,[12] a completeness relation in which all resonances are screened out from the continuum was obtained. Thus in the work of Refs. 6, 11, and 12 both resonance and bound states contributions to the completeness relation are treated on the same footing.

Underlying the nonuniqueness in defining the integral properties of resonance states are the divergence difficulties inherent in the coordinate-space representation of these states. This has to do in a sense with the fact that the resonance states considered[6–14] do not span the entire Hilbert space. Avoiding such nonuniqueness seems, therefore, to be intimately connected with the proper description of resonance states in Hilbert space (see Sec. IV).

In the present paper, an attempt is made to overcome the above limitations by utilizing biorthonormal resonance states defined in terms of symmetrized scattering kernels[15–18] analytically continued into the "unphysical sheet" of complex energies. As such, the resonance states are treated as generalized eigenfunctions in a complex-energy representation which lends itself to the avoidance of the divergence problems associated with coordinate-space calculations. The resonance states considered form complete bases states in an "extended" or rigged Hilbert space (see below).

In addressing the problem of constructing appropriate completeness relations for resonance states, the present work extends to the general treatment of unstable states in scattering theory and their generalized description in quantum mechanics.[19–30] Using functional analysis techniques,[31,32] resonances are described within such approaches[19–30] in terms of generalized eigenvectors with complex eigenvalues in a rigged Hilbert space formulation.[19,33] By analytically continuing the expansions of sym-

[a)]Present address: Rochester Institute of Technology, Rochester, New York 14623.
[b)]Affiliation 1978–79. USDOE Contract number W-7405-ENG-82.

metrized resolvent scattering kernels into the unphysical sheet of the complex-energy plane, the generalized bases states given for resonances in the present work are identified in a natural way with rigged Hilbert space structures.[22,31–33]

In Sec. II, a brief summary of the properties of symmetrized scattering kernels in Hilbert space is given. Extension to complex energies and analytic continuation into the unphysical sheet is considered in Sec. III. Generalized completeness relations involving resonance states are constructed in Sec. IV. Section V contains a conclusion.

II. SYMMETRIZED SCATTERING KERNELS IN HILBERT SPACE

In this section, results based on the symmetrized scattering kernels of Meetz,[15] Weinberg,[16,17] and Sasakawa[18] are briefly outlined, and their properties which have direct bearing on our results are discussed.

Let λV be an interaction and $G_0(E \pm i\epsilon) = (E - H_0 \pm i\epsilon)^{-1}$ the two-body Green's function, with H_0 being the kinetic energy operator. Defining the symmetrized scattering kernel,[17–19,24,25]

$$K^{\pm}(E) = |V|^{1/2} G_0^{\pm}(E) |V|^{1/2}, \tag{2.1}$$

the outgoing and ingoing solutions are given by a "modified" Lippman–Schwinger equation[35,36] (assuming no degeneracy) as

$$|V|^{1/2}\psi^{\pm}(E)|V|^{1/2} = \phi(E) + \lambda K^{\pm}(E)|V|^{1/2}\psi^{\pm}(E). \tag{2.2}$$

The properties of $K^{\pm}(E)$ depend upon the potential. We confine ourselves to the class of potentials which make these kernels square integrable,[34–39] i.e.,

$$\|K^{\pm}\|^2 = \int d\mathbf{r} \int d\mathbf{r}' |K^{\pm}(\mathbf{r},\mathbf{r}';E)|^2 < \infty. \tag{2.3}$$

In this case, $K^{\pm}(E)$ behaves as a Hilbert–Schmidt operator[26] belonging to the (complex) space L^2.

The kernel symmetrization in (2.1) applies for $V \geqslant 0$. For potentials which change sign, the "polar" form[37,38]

$$K^{\pm}(E) = |V|^{1/2} G_0^{\pm}(E) |V|^{1/2} (\mathrm{sgn} V) \tag{2.4}$$

should be used, where $(\mathrm{sgn} V)$ denotes a sign factor equal to $+1$ for positive V and -1 for negative V. The discussion in this paper is restricted, for simplicity, to the symmetrized kernels in (2.1). This will not affect our main results, and extension to the form in (2.4) can be done in a straightforward manner.[15]

The symmetrized kernel $K^{\pm}(E)$ is a two-sheeted function of E. The upper half E-plane is mapped onto the first (physical) sheet and the lower half is mapped onto the second (unphysical) sheet. $K^{\pm}(E)$ represent the boundary values of $K(E)$ on the real axis in the first sheet

$$K^{\pm}(E) = \lim_{\epsilon \to 0} K(E \pm i\epsilon). \tag{2.5}$$

Spectral decompositions of $K(E)$ can be obtained for all E by investigating its properties in the first and second sheets. On the first sheet including the real axis, $K(E)$ is symmetric and the completely continuous L^2 operator has pure point spectrum.[15–18,40,41] Its resolvent[40,41] $K(1 - \lambda K)^{-1}$ is a meromorphic function of the coupling constant λ with singular values at[15–17]

$$\lambda = \eta_n^{-1}(E), \tag{2.6}$$

where the $\eta_n(E)$ are the eigenvalues of $K(E)$, i.e.,

$$K(E)\chi_n(E) = \eta_n(E)\chi_n(E), \tag{2.7}$$

and the $\chi_n(E)$ the corresponding eigenelements, viz.,

$$\chi_n(E) = \eta_n^{-1}(E) K(E) \chi_n(E). \tag{2.8}$$

On the negative real axis, the set $\{\chi_n\}$ is orthonormal:

$$\langle \chi_n | \chi_m \rangle = \delta_{nm}. \tag{2.9}$$

Then, for $-\infty < E \leqslant 0$, $K(E)$ has the spectral decomposition[15–18,40,41]

$$K(E) = \sum_n \eta_n(E) \chi_n(E) \otimes \chi_n(E), \tag{2.10}$$

which is convergent in the mean.[40] Since elements of the form $K\chi$ are dense in L^2, the system $\{\chi_n\}$ is complete,[40] and the kernel $K(E)$ is normal, i.e., permutable with its adjoint:

$$KK^{\dagger} = K^{\dagger}K. \tag{2.11}$$

For positive energies, $K(E \pm i\epsilon)$ is not Hermitian and ceases to be normal so that its spectral decomposition into the orthonormal set $\{\chi_n\}$ is no longer possible. It is still possible, however, to construct an expansion in terms of biorthogonal eigenfunctions by analytic continuation from the negative real axis[15] which yields

$$K(E + i\epsilon) = \sum_n \eta_n(E + i\epsilon) \chi_n(E + i\epsilon) \cdot \chi_n(E + i\epsilon), \tag{2.12a}$$

$$\begin{aligned} K(E - i\epsilon) &= \tilde{K}(E + i\epsilon) \equiv K^*(E + i\epsilon) \\ &= \sum_n \tilde{\eta}_n(E + i\epsilon) \tilde{\chi}_n(E + i\epsilon) \cdot \tilde{\chi}_n(E + i\epsilon), \end{aligned} \tag{2.12b}$$

with

$$\begin{aligned} \tilde{\eta}_n(E + i\epsilon) &= \eta_n^*(E + i\epsilon) = \eta_n(E - i\epsilon), \\ \tilde{\chi}_n(E + i\epsilon) &= \chi_n^*(E + i\epsilon) = \chi_n(E - i\epsilon). \end{aligned} \tag{2.13}$$

Denoting $z = E \pm i\epsilon$, the eigenfunctions χ_n and $\tilde{\chi}_n$ satisfy

$$K(z)\chi_n(z) = \eta_n(z)\chi_n(z), \quad \tilde{K}(z)\tilde{\chi}_n(z) = \tilde{\eta}_n(z)\tilde{\chi}_n(z) \tag{2.14}$$

and the biorthogonality relation

$$\langle \chi_n | \tilde{\chi}_m \rangle = \delta_{nm}. \tag{2.15}$$

Comparing (2.12) with (2.10), it is seen that the tensorial product $\chi_n \otimes \chi_n$ is replaced with the product $\chi_n \cdot \chi_n$, which does not involve the complex conjugate. This results from the fact that the analytic continuation of $\|\chi_n(E \pm i\epsilon)\|^2$ is given by

$$\|\chi_n(E \pm i\epsilon)\|^2 = \langle \chi_n(E \pm i\epsilon) | \tilde{\chi}_n(E \pm i\epsilon) \rangle. \tag{2.16}$$

It also follows that in contrast to (2.10) the expansions in (2.12) have the property of being "complex normal,"[15] i.e.,

$$KK^{\dagger} = (K^{\dagger}K)^*. \tag{2.17}$$

Thus, the biorthogonal system $\{\chi_n, \tilde{\chi}_n\}$ allows us to extend the expansion (2.10) to positive energies. The resulting expansion in (2.12) converges in $E \pm i\epsilon$, implying that $\{\chi_n, \tilde{\chi}_n\}$ is a complete biorthogonal system[15] of L^2.

Other choices of symmetrized scattering kernels are also possible[16–18,34,42–46] as long as they lead to good convergence properties. In this regard, the symmetric kernel

$$K(E) = |V|^{1/2} P G_0(E) |V|^{1/2}, \tag{2.18}$$

where P denotes the principal value, has better convergence properties at positive energies[18] than the kernel in (2.1) and is therefore more practical to use. In particular, the eigenelements corresponding to the kernel (2.18) form a complete set, and spectral decomposition is applicable (which permits also use of the Weinberg quasiparticle approach in a straightforward manner[16,17]).

The above discussion is equivalent to applying the Fredholm theory[45,46] to the completely continuous kernel $K(E)$ for all E in the space L^2 and hence may always be approximated by a kernel of finite rank.[15–18,38–41] The above method, however, leads to expressions[15–18] which are easier to handle than the corresponding Fredholm determinants. In the next sections, we pursue and extend this discussion to the complex-energy plane and obtain generalized completeness relations for resonant scattering.

III. ANALYTIC SPECTRAL DECOMPOSITIONS

Our aim in this section is to obtain spectral decompositions which exhibit the resonance states explicitly. To this purpose, we extend the spectral definition of the operators considered. In general, given a bounded self-adjoint linear operator on a Hilbert space $\mathscr{H}$, we can decompose[35–39]

$$\mathscr{H} = \mathscr{H}_{\rm p} \oplus \mathscr{H}_{\rm c}, \tag{3.1}$$

where $\mathscr{H}_{\rm p}$ has a discrete point spectrum and $\mathscr{H}_{\rm c}$ has a continous spectrum. Making use of the resolvent kernel[40,41]

$$\begin{aligned} R(z) &= K(1-\lambda K)^{-1} \\ &= |V|^{1/2}(z-H)^{-1}|V|^{1/2}, \end{aligned} \tag{3.2}$$

one may, accordingly, write the spectral decomposition

$$|V|^{1/2}\frac{1}{z-H}|V|^{1/2} = \sum_n \frac{|\xi_n\rangle\langle\xi_n|}{z-z_n} + \int_c dE \sum_\alpha \frac{|\xi^\alpha\rangle\langle\xi^\alpha|}{z-E}, \tag{3.3}$$

where $|\xi_n\rangle$ specifies a set of discrete eigenstates for $z = z_n$, a real solution to an equation similar to (2.14), viz.,

$$K(z)|\xi_n\rangle = \eta_n(z)|\xi_n\rangle, \tag{3.4}$$

and $|\xi^\alpha\rangle \equiv |V|^{1/2}\psi^\pm(E,\alpha)$ specifies a set of continuum solutions to (2.2), with α indexing possible degeneracies. In (3.3), the integration is taken over a set of continuum eigenvalues denoted by c. Further, the set $\{\xi\}$ consisting of $\{\xi_n\}$ and $\{\xi^\alpha\}$ form a complete orthonormal set:

$$\begin{aligned} &\langle\xi_n|\xi_m\rangle = \delta_{nm}, \\ &\langle\xi^\alpha(E)|\xi^{\alpha'}(E')\rangle = \delta(E-E')\delta_{\alpha\alpha'}, \\ &\langle\xi_n|\xi^\alpha(E)\rangle = 0. \end{aligned} \tag{3.5}$$

Resonances are now treated through analytic continuation of the resolvent kernel in (3.3) into the unphysical sheet. In doing this, it is known that if the integration contour c (corresponding to the real cut) is deformed into a curve σ in an allowable fashion which preserves the analyticity of the integrand, the integral itself remains constant during deformation. Therefore, by deforming the integration contour c into the lower half-plane while keeping its end points fixed to a contour σ which crosses the resonance poles at the points $z_\nu = E - i\Gamma$ but avoids other singularities of the analytically continued integrand, one obtains, upon applying the Cauchy residue theorem and noticing (2.16), the result

$$|V|^{1/2}\frac{1}{z-H}|V|^{1/2} = \sum_n \frac{|\xi_n\rangle\langle\xi_n|}{z-z_n} + \int_\sigma dz' \sum_\alpha \frac{|\xi^\alpha\rangle\langle\tilde\xi^\alpha|}{z-z'} + \sum_\nu \operatorname{Res} R^-(z)\Big|_{z=z_\nu}, \tag{3.6}$$

where use has been made of

$$2\pi i \sum_\alpha |\xi^\alpha\rangle\langle\xi^\alpha| = R^-(z) - R^+(z), \tag{3.7}$$

which gives the discontinuity of $R(z)$ across the cut c as follows[20] from (3.3), with

$$R^\pm(z) = R(E \pm i\Gamma) = |V|^{1/2}(E \pm i\Gamma - H)^{-1}|V|^{1/2}. \tag{3.8}$$

In writing (3.6), only the resolvent term $R^-(z)$ appears in the last sum since it is this term which may have resonance poles in its continuation through the cut c into the unphysical sheet. The term $R^+(z)$ remains analytic for $\operatorname{Im} z > 0$ because it does not cross the real-axis cut.

Noting the association of the discrete (bound) states $\{\xi_n\}$ in (3.3) with residues from the first-order poles of the resolvent kernel in the physical branch, i.e.,

$$\operatorname{Res} R(z)|_{z=z_n} = \lim_{z\to z_n}(z-z_n)R(z) = |\xi_n\rangle\langle\xi_n| \tag{3.9}$$

and assuming that each point z_ν in (3.6) is a first-order pole (resonance) of $R(z)$, the residue terms in the last sum of (3.6) can be evaluated similarly.

Defining a set $\{\xi_n\}$ at points in the physical sheet other than those satisfying the bound-state solutions, then continuing these functions to resonance solutions at complex points z_ν in the unphysical sheet, one gets

$$\operatorname{Res} R^-(z) \equiv \operatorname{Res} R(z) = \lim_{z\to z_\nu}(z-z_\nu)R(z) = |\xi_\nu\rangle\langle\tilde\xi_\nu|. \tag{3.10}$$

By construction, the analytically continued eigenelements in (3.10) above form a biorthogonal system $\{\xi_\nu, \tilde\xi_\nu\}$ as follows from (2.16) and the discussion in Sec. II.

Consequently, we are able to write for the resolvent kernel the generalized spectral decomposition

$$|V|^{1/2}\frac{1}{z-H}|V|^{1/2} = \sum_n \frac{|\xi_n\rangle\langle\xi_n|}{z-z_n} + \sum_\nu \frac{|\xi_\nu\rangle\langle\tilde\xi_\nu|}{z-z_\nu} + \int_\sigma dz' \sum_\alpha \frac{|\xi^\alpha\rangle\langle\tilde\xi^\alpha|}{z-z'}. \tag{3.11}$$

In (3.11), the system $\{\xi, \tilde\xi\}$ consisting of $\{\xi_\nu, \tilde\xi_\nu\}$ and $\{\xi^\alpha, \tilde\xi^\alpha\}$ satisfies the biorthogonality relations:

$$\begin{aligned} &\langle\xi_\nu|\tilde\xi_\mu\rangle = \delta_{\nu\mu}, \\ &\langle\xi^\alpha(z)|\tilde\xi^{\alpha'}(z')\rangle = \delta(z-z')\delta_{\alpha\alpha'}, \\ &\langle\xi_\nu|\tilde\xi^\alpha\rangle = \langle\xi^\alpha|\tilde\xi_\nu\rangle = 0, \end{aligned} \tag{3.12}$$

which ensure its completeness in the space L^2.

IV. GENERALIZED COMPLETENESS RELATIONS

Let $\langle f\rangle$ be an arbitrary element of (complex) L^2. As a result of the completeness and orthonormality of the set

$\{\chi_n\}$ in (2.7)–(2.9), the Parseval identity

$$\|f\|^2 = \sum_n |\langle \chi_n | f \rangle|^2 \tag{4.1}$$

holds true in L^2.

In the following, a generalized Parseval's identity involving the contribution of complex points in the unphysical sheet as demonstrated in the set $\{\xi, \tilde{\xi}\}$ is obtained.

Spectral decomposition in terms of discrete and continuum eigenstates yields[37]

$$\|f\|^2 = \sum_n \langle f|\xi_n\rangle\langle \xi_n|f\rangle + \int_c dE \sum_\alpha \langle f|\xi^\alpha\rangle\langle\xi^\alpha|f\rangle. \tag{4.2}$$

Applying analytic continuation through the cut c into the unphysical sheet, as done in Sec. III, one gets the following result on deforming the integration contour and using Cauchy theorem, taking into account (3.9):

$$\|f\|^2 = \sum_n \langle f|\xi_n\rangle\langle \xi_n|f\rangle + \sum_\nu \langle f|\xi_\nu\rangle\langle\tilde{\xi}_\nu|f\rangle + \int_\sigma dz \sum_\alpha \langle f|\xi^\alpha\rangle\langle\tilde{\xi}^\alpha|f\rangle, \tag{4.3}$$

where the contour σ is taken as in (3.5).

The completeness relation in (4.3), in contrast to (4.1), is defined over an "extended" or rigged Hilbert space[31–33] of generalized eigenvectors (locally integrable distributions) corresponding to complex eigenvalues. In the terminology of functional analysis,[19,20,22–33] a rigged Hilbert space is represented by a Gel'fand triplet $\Phi \subset L^2 \subset \tilde{\Phi}$, where Φ is a dense subspace of L^2 with nuclear embedding and $\tilde{\Phi}$ is the dual space of Φ. Generalized eigenfunctions corresponding to complex eigenvalues given in (4.3) are then elements of the enlarged Hilbert space $\tilde{\Phi} \supset L^2$. The use of rigged Hilbert space structures provides a powerful tool for the study of unstable and resonance states in quantum mechanics,[19–30] whereby the restrictive nature of the usual Hilbert space is overcome.

In this connection, it is worth nothing that, in the absence of bound states, the last two terms in (4.3) imply the decomposition of the rigged Hilbert space $\Phi \subset L^2 \subset \tilde{\Phi}$ into both "discrete" and "continuous" parts corresponding to the discrete and continuous contributions of the "complete" set of generalized eigenfunctions.

Since relation (4.3) is defined over an entire rigged Hilbert space, restrictions associated with "undercompleteness" or "overcompleteness," which often give rise to nonuniqueness in constructing completeness relations for resonance states,[6,8–13] are avoided. Provided the resolvent kernel has a well-defined analytic continuation from $E \in c$ to $E \in \sigma$, the existence of such a generalized completeness relation allows the expansion of any integral property, e.g., norm or inner product, in terms of generalized bases involving resonance states. Thus, for two arbitrary functions $f, g \in \Phi$, one has

$$\langle f|g\rangle = \sum_n \langle f|\xi\rangle\langle \xi|g\rangle + \sum_\nu \langle f|\xi_\nu\rangle\langle\tilde{\xi}_\nu|g\rangle + \int_\sigma dz \sum_\alpha \langle f|\xi^\alpha\rangle\langle\tilde{\xi}^\alpha|g\rangle. \tag{4.4}$$

It should be noted that the generalized bases in (4.3) and (4.4) are evaluated in a spectral representation corresponding to complex-energy solutions according to (3.4). This avoids the use of coordinate-space representation of resonance states and the associated divergence problems.[47]

Further, in view of (4.3) and (4.4), it is justifiable to write the generalized closure relation

$$\sum_n |\xi_n\rangle\langle\xi_n| + \sum_\nu |\xi_\nu\rangle\langle\tilde{\xi}_\nu| + \int_\sigma dz \sum_\alpha |\xi^\alpha\rangle\langle\tilde{\xi}^\alpha| = 1, \tag{4.5}$$

which is uniquely defined over the entire rigged Hilbert space $\Phi \subset L^2 \subset \tilde{\Phi}$ as mentioned before.

V. CONCLUSION

The generalized completeness relations obtained in this work (Sec. IV) should prove of relevance in various applications where resonance scattering can take place. They provide a generalized scheme for eigenfunction expansions based on complex-energy generalized eigenfunctions which span an entire rigged Hilbert space.

In this regard, the present completeness relations can be applied, for example, to sum rules in nuclear physics,[6] leading to well-defined separation of resonance contributions as contrasted with other approaches.[6–14] Further, the expansion of the multichannel Green function, half-shell and off-shell scattering amplitudes, using these generalized completeness relations is within immediate reach. It should prove also useful to apply these relations to few-body problems involving two-body resonant t matrices.[4,5]

[1]M. I. Haftel, Phys. Rev. Lett. **25**, 120 (1970).
[2]M. Baranger *et al.*, Nucl. Phys. A **138**, 1 (1969).
[3]P. U. Sauer and A. Sevgen, Phys. Rev. C **13**, 720 (1976).
[4]O. Zohni, Bull. Am. Phys. Soc. **22**, 1030 (1977).
[5]A. Lagu and O. Zohni, in *Few-Body Systems and Nuclear Forces, I*, edited by H. Zingl, M. Haftel, and H. Zankel, *Lecture Notes in Physics*, Vol. 82 (Springer-Verlag, Berlin, 1978), p. 106.
[6]T. Berggren, Nucl. Phys. A **109**, 265 (1968); Phys. Lett. B **38**, 62 (1972); Phys. Lett. B **44**, 23 (1973).
[7]Ya. B. Zel'dovich, Zh. Eksp. Teor. Fiz. **39**, 776 (1961) [Sov. Phys. JETP **12**, 542 (1962)].
[8]R. M. More and Gerjuoy, Phys Rev. A **7**, 1288 (1973).
[9]G. Garcia-Calderon and R. Peierls, Nucl. Phys. A **265**, 443 (1976).
[10]G. Garcia-Calderon, Nucl. Phys. A **261**, 130 (1976).
[11]W. J. Romo, Nucl. Phys. A **116**, 618 (1968).
[12]Y. S. Kim and K. V. Vasavada, Technical Report No. 71-050, University of Maryland, College Park, Maryland, 1970.
[13]D. L. Huestis, J. Math. Phys. **16**, 2148 (1975).
[14]M. Wagner, Phys. Rev. **136**, B562 (1964).
[15]K. Meetz, J. Math. Phys. **3**, 690 (1962).
[16]S. Weinberg, Phys. Rev. **131**, 440 (1963).
[17]M. Scadron, S. Weinberg and J. Wright, Phys. Rev. **135**, B202 (1964).
[18]T. Sasakawa, Nucl. Phys. A **160**, 321 (1971).
[19]Parravicini, V. Gorini, and E. C. G. Sudarshan, J .Math. Phys. **21**, 2208 (1980).
[20]E. C. G. Sudarshan, C. B. Chin, and V. Gorini, Phys. Rev. D **18**, 2914 (1978).
[21]L. Fonda, G. C. Ghirardi, and A. Rimini, Rep. Prog. Phys. **41**, 587 (1978).

[22]A. Grossman, J. Math. Phys. **5**, 1025 (1964).
[23]A. Böhm, J. Math. Phys. **21**, 1040 (1980).
[24]O. Melsheimer, J. Math. Phys. **15**, 902, 917 (1974).
[25]H. Baumgärtel, Math. Nach. **75**, 133 (1976).
[26]J. S. Howland, J. Math. Anal. Appl. **50**, 415 (1975).
[27]L. P. Horowitz and J. P. Marchand, Rocky Mount. J. Math. **1**, 225 (1971).
[28]J. E. Roberts, Commun. Math. Phys. **3**, 98 (1966).
[29]J. -P. Antoine, J. Math. Phys. **10**, 53 (1969).
[30]W. G. Faris, Helv. Phys. Acta **44**, 930 (1971).
[31]I. M. Gel'fand, M. I. Graev, G. E. Shilov, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1964).
[32]K. Maurin, *General Eigenfunction Expansions and Unitary Representations of Topological Groups* (PWN, Warsaw, 1968).
[33]A. Böhm, *The Rigged Hilbert Space and Quantum Mechanics, Lecture Notes in Physics*, Vol. 178 (Springer, New York, 1978).
[34]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, N. J., 1971).
[35]M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol. III, Scattering Theory* (Academic, New York, 1979).
[36]W. O. Armien, J. M. Jauch, and K. B. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, New York, 1977).
[37]R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[38]P. R. Halmos, *Finite Dimensional Vector Spaces* (Van Nostrand, Princeton, N. J., 1958).
[39]G. Helmberg, *Introduction to Spectral Theory in Hilbert Space* (Elsevier, New York, 1961).
[40]F. Riesz and B. Az. Nagy, *Functional Analysis* (Unger, New York, 1966).
[41]F. Smithies, *Integral Equations* (Cambridge U. P., Cambridge, 1958).
[42]J. Schwinger, Proc. Natl. Acad. Sci. USA **47**, 122 (1961).
[43]F. Coester, Phys. Rev. **133**, B1516 (1964).
[44]J. S. Ball and D. Y. Wong, Phys. Rev. **169**, 1362 (1968).
[45]R. Jost and A. Pais, Phys. Rev. **82**, 840 (1951).
[46]N. N. Khuri, Phys. Rev. **107**, 1148 (1957).
[47]T. K. Bailey and W. C. Schieve, Nuovo Cimento A **47**, 231 (1978).

Eigenvalues of an anharmonic oscillator. II

R. N. Kesarwani
Department of Mathematics, University of Ottawa, Ottawa K1N 9B4, Canada
Y. P. Varshni
Department of Physics, University of Ottawa, Ottawa K1N 9B4, Canada

(Received 26 May 1981; accepted for publication 2 September 1981)

An expression is derived for the sixth nonzero term in the WKBJ approximation. The six-term WKBJ approximation is applied to calculate the eigenvalues for the potential $V(x) = \frac{1}{2}kx^2 + ax^4$, $k > 0$ and $a > 0$. At low values of λ [$= a\hbar/(\mu k^3)^{1/2}$], the calculated results are in excellent agreement, to 15 significant figures, with those of Banerjee *et al.* for all quantum numbers. At medium and high values of λ, the calculated results are poor at low quantum numbers, but improve rapidly as n increases. A 15-significant-figure accuracy is achieved at $n = 10$ for $\lambda = 0.05$, and at about $n = 15$ for $\lambda = 20\,000$. For $\lambda = 0.5$, eigenvalues are calculated to 20 significant figures and an argument is presented to show that by $n = 50$, the calculated value is correct to 19 significant figures. The accuracy further improves at higher quantum numbers.

PACS numbers: 03.65.Sq

I. INTRODUCTION

Recently, we have used[1] (hereafter referred to as Paper I) the five-term (or eighth order) WKBJ approximation to calculate the eigenvalues for the anharmonic oscillator potential,

$$V(x) = \tfrac{1}{2}kx^2 + ax^4, \quad k>0 \quad \text{and} \quad a>0. \tag{1}$$

Numerical results were compared with those of Hioe and Montroll.[2] It was found that at $n = 4$, a seven-significant-figure accuracy was achieved and at $n = 6$, a nine-significant-figure accuracy. References to previous work on this problem may be found in Paper I. Another recent investigation of the energy levels of potential (1) by the WKBJ approximation is that of Hioe *et al.*[3]

Banerjee and co-workers, in two recent papers,[4,5] have used an interesting method[6] to calculate the eigenvalues of potential (1) to 15-significant-figure accuracy. Their approach is similar to that of Biswas *et al.*,[7] but differs from it in one important respect—the introduction of a scaling parameter. Banerjee *et al.*[4,5] use an appropriately scaled basis for the expansion of each eigenfunction:

$$\psi_n(\lambda) = e^{-\alpha x^2} \sum_{m=0}^{\infty} a_m x^m, \tag{2}$$

where the scaling is introduced through the parameter α. For an effective expansion the scale of coordinates is chosen so that sufficient lower members of the basis functions $\{x^m e^{-\alpha x^2}\}$ in the expansion (2) have appreciable values in the region of oscillations of the actual eigenfunctions and decay outside the region of oscillation. Banerjee *et al.*[4] give arguments to show that the scaling parameter α will depend on n and λ according to the following formula:

$$\alpha(n,\lambda) = \tfrac{1}{2} + Bn^{1/3}\lambda^{1/3}, \tag{3}$$

where the constant B is adjusted empirically. The effect of using this scaled basis is remarkable. Banerjee *et al.*[4] are able to calculate the eigenvalues in any realm of n and λ to high accuracy.

This has prompted us to examine the improvement which may be obtained in the eigenvalues for potential (1) by taking into account the sixth nonzero term in the WKBJ approximation. In the present paper, we first derive an expression for the sixth nonzero term in the WKBJ approximation, and then apply the six-term (tenth order) WKBJ approximation to calculate the eigenvalues for potential (1). The results are compared with those of Banerjee *et al.*[4,5]

II. SIX-TERM WKBJ APPROXIMATION

Using Dunham's[8] approach, we have derived the fifth nonzero term in the WKBJ approximation in a previous paper.[9] The discrete energy eigenvalues are determined in the WKBJ approximation by the following condition:

$$n = \frac{1}{2\pi\hbar} \oint_C \sum_{s=0}^{\infty} \left(\frac{\hbar}{i}\right)^s y_s(x)\, dx, \tag{4}$$

where n is the quantum number, x is in a complex plane cut along the real axis between the classical turning points, and the integration is carried out along a contour C enclosing the two turning points but no other singularities of the integrand $y_s(x)$, and not crossing the cut.

Recursion formulas for determining y_s are given in Ref. 9, which also contains explicit expressions for y_s's up to and including y_{13}. As shown in that paper, the contribution to (4) due to $y_3, y_5, y_7, \cdots$ vanish, and it is only $y_0, y_1, y_2, y_4, \cdots$ that contribute nonzero values to the right-hand side of (4). Equation (4) may be rewritten as

$$n + \tfrac{1}{2} = I_1 + I_2 + I_3 + \cdots, \tag{5}$$

where $I_1, I_2, I_3, \cdots$ represent the contribution due to $y_0, y_2, y_4, \cdots$, respectively. Explicit expressions for I_1, I_2, I_3, I_4, and I_5 were obtained in Ref. 9. Here we derive an expression for I_6, the next nonzero term in the right-hand side of (4). The nth derivative ($n > 3$) of V shall be represented by $V^{(n)}$.

 0022-2488/82/050803-07$02.50

$$I_6 = \frac{-\hbar^9}{2\pi}\oint_C y_{10}(x)\,dx = -\frac{[\hbar/(2\mu)^{1/2}]^9}{2^{29}\pi}\oint_C \{350229325V'^{10}(\epsilon - V)^{-29/2}$$
$$+ 1400917300V'^8V''(\epsilon - V)^{-27/2} + 160[2350680V'^7V''' + 11893561V'^6V''^2](\epsilon - V)^{-25/2}$$
$$+ 128[525825V'^6V^{(4)} + 6755070V'^5V''V''' + 7840519V'^4V''^3](\epsilon - V)^{-23/2}$$
$$+ 256[27120V'^5V^{(5)} + 522510V'^4V''V^{(4)} + 393465V'^4V'''^2 + 1937784V'^3V''^2V''' + 696101V'^2V''^4]$$
$$\times(\epsilon - V)^{-21/2} + 1024[35160V'^3V'''V^{(4)} + 13560V'^3V''V^{(5)} + 58938V'^2V''V'''^2$$
$$+ 45738V'^2V''^2V^{(4)} + 80364V'V''^3V''' - 1087V''^5](\epsilon - V)^{-19/2}$$
$$+ 4096[900V'^2V'''V^{(5)} + 785V'^2(V^{(4)})^2 + 2552V'V''V'''V^{(4)}$$
$$+ 1224V'V''^2V^{(5)} + 2321V''^2V'''^2 - 114V''^3V^{(4)}](\epsilon - V)^{-17/2}$$
$$+ 16384[40V'V^{(4)}V^{(5)} + 68V''V'''V^{(5)} - 3V''(V^{(4)})^2](\epsilon - V)^{-15/2} + 32768(V^{(5)})^2(\epsilon - V)^{-13/2}\}\,dx. \tag{6}$$

We perform integration by parts repeatedly to simplify the above expression to a form in which V' is absent. This gives us

$$I_6 = \frac{[\hbar/(2\mu)^{1/2}]^9}{135\pi 2^{24}}\oint_C \{25409475V''^5(\epsilon - V)^{-19/2} + 780[2632V''^2V'''^2 + 23991V''^3V^{(4)}](\epsilon - V)^{-17/2}$$
$$+ 52[23606V''^2V^{(6)} + 9152V''V'''V^{(5)} + 46215V''(V^{(4)})^2 + 3760V'''^2V^{(4)}](\epsilon - V)^{-15/2}$$
$$+ 40[1001V''V^{(8)} + 352V'''V^{(7)} + 3146V^{(4)}V^{(6)} + 416(V^{(5)})^2](\epsilon - V)^{-13/2} + 560V^{(10)}(\epsilon - V)^{-11/2}\}\,dx. \tag{7}$$

The apparent nonintegrable singularities in the above expression can be removed, and the final expression may be written as

$$I_6 = -\frac{[\hbar/(2\mu)^{1/2}]^9}{48\pi 12!}\Big\{7665\frac{d^9}{d\epsilon^9}\int_{r_1}^{r_2} V''^5(\epsilon - V)^{-1/2}\,dx$$
$$-\frac{d^8}{d\epsilon^8}\int_{r_1}^{r_2}[5264V''^2V'''^2 + 47982V''^3V^{(4)}](\epsilon - V)^{-1/2}\,dx$$
$$+\frac{d^7}{d\epsilon^7}\int_{r_1}^{r_2}[23606V''^2V^{(6)} + 9152V''V'''V^{(5)} + 46215V''(V^{(4)})^2 + 3760V'''^2V^{(4)}](\epsilon - V)^{-1/2}dx$$
$$-5\frac{d^6}{d\epsilon^6}\int_{r_1}^{r_2}[1001V''V^{(8)} + 352V'''V^{(7)} + 3146V^{(4)}V^{(6)} + 416(V^{(5)})^2](\epsilon - V)^{-1/2}\,dx$$
$$+385\frac{d^5}{d\epsilon^5}\int_{r_1}^{r_2} V^{(10)}(\epsilon - V)^{-1/2}\,dx\Big\}. \tag{8}$$

Here r_1, r_2 are the classical turning points defined by $\epsilon - V(x) = 0$.

The six-term (or tenth order) WKBJ approximation may then be written as

$$n + \tfrac{1}{2} = I_1 + I_2 + I_3 + I_4 + I_5 + I_6. \tag{9}$$

A. A special case

In view of the fact that the derivation of I_6 is lengthy and cumbersome, it was felt desirable to check its correctness by any test which may be possible. One such test is provided by considering a special case,

$$V(x) = Cx^N, \quad N = 2,4,6,\dots. \tag{10}$$

For this potential, Bender *et al.*[10] have used the MACSYMA computer program at MIT to perform the algebraic manipulation required to calculate the first eight terms in the series in (5). For the potential (10), our expression for I_6 gives

$$I_6 = -\frac{[\hbar/(2\mu)^{1/2}]^9}{48\pi 12!}\Big[7665C^5N^5(N-1)^5\frac{d^9}{d\epsilon^9}\int_{r_1}^{r_2} x^{5N-10}(\epsilon - Cx^N)^{-1/2}\,dx$$

$$-C^4N^4(N-1)^4(N-2)(53246N-154474)\frac{d^8}{d\epsilon^8}\int_{r_1}^{r_2} x^{4N-10}(\epsilon - Cx^N)^{-1/2}\,dx$$

$$+C^3N^3(N-1)^3(N-2)(N-3)(82733N^2-513481N+837666)\frac{d^7}{d\epsilon^7}\int_{r_1}^{r_2} x^{3N-10}(\epsilon - Cx^N)^{-1/2}\,dx$$

$$-5C^2N^2(N-1)^2(N-2)(N-3)(N-4)[11(N-5)(N-6)(123N-701)$$

$$+26(N-2)(N-3)(137N-669)]\frac{d^6}{d\epsilon^6}\int_{r_1}^{r_2} x^{2N-10}(\epsilon - Cx^N)^{-1/2}\,dx$$

$$+385CN(N-1)(N-2)(N-3)(N-4)(N-5)(N-6)(N-7)(N-8)(N-9)$$

$$\times\frac{d^5}{d\epsilon^5}\int_{r_1}^{r_2} x^{N-10}(\epsilon - Cx^N)^{-1/2}\,dx\Big]. \tag{11}$$

Here $r_1 = -(\epsilon/C)^{1/N}$ and $r_2 = (\epsilon/C)^{1/N}$. Note that when m is any positive integer and N is any positive even integer,

$$\frac{d^{m+4}}{d\epsilon^{m+4}}\int_{r_1}^{r_2} x^{mN-10}(\epsilon - Cx^N)^{-1/2}\,dx = \frac{d^{m+4}}{d\epsilon^{m+4}}\left[\frac{2(\pi)^{1/2}\epsilon^{m-1/2-9/N}\Gamma(m-9/N)}{NC^{m-9/N}\Gamma(m+\frac{1}{2}-9/N)}\right]$$
$$= \frac{2(\pi)^{1/2}\epsilon^{-9/2-9/N}\Gamma(m-9/N)}{NC^{m-9/N}\Gamma(-7/2-9/N)}.$$

Therefore

$$I_6 = -\frac{[\hbar C^{1/N}/(2\mu)^{1/2}]^9\epsilon^{-9(N+2)/2N}}{24(\pi)^{1/2}12!}\frac{\Gamma(1-9/N)}{\Gamma(-7/2-9/N)}(N-1)(N-3)(N-9)$$

$$\times[3(N-1)^3(2N-9)\{7665(N-1)(4N-9)-(N-2)(53246N-154474)\}$$

$$+(N-1)^2(N-2)(2N-9)(82733N^2-513481N+837666)$$

$$-5(N-1)(N-2)(N-4)\{11(N-5)(N-6)(123N-701)+26(N-2)(N-3)(137N-669)\}$$

$$+385(N-2)(N-4)(N-5)(N-6)(N-7)(N-8)]. \tag{12}$$

On setting $\hbar/(2\mu)^{1/2} = C^{-1/N}$, this gives

$$I_6 = -\frac{\epsilon^{-9(N+2)/2N}\Gamma(1-9/N)}{24(\pi)^{1/2}12!\Gamma(-7/2-9/N)}(N-1)(N-3)(N-9)(2N+3)(320N^5-504N^4$$

$$-4854N^3-957N^2+14754N+12801). \tag{13}$$

The above expression is the same as the corresponding term obtained by Bender *et al.*[10]

III. APPLICATION TO THE ANHARMONIC OSCILLATOR

The evaluation of the terms I_1, I_2, I_3, I_4, and I_5 for the potential (1) has been carried out in paper I, and their values obtained in terms of $K(\omega)$ and $E(\omega)$, the complete elliptic integrals of the first and second kind. We shall find it convenient to express our results in terms of λ, ω, and η, which are defined by

$$\lambda = a\hbar/(\mu k^3)^{1/2}, \tag{14}$$

$$\omega = [\tfrac{1}{2} - \tfrac{1}{2}(1+16a\epsilon/k^2)^{-1/2}]^{1/2}, \tag{15}$$

and

$$\eta = 4\omega^2(1-\omega^2) \quad\text{or}\quad 1-2\omega^2 = (1-\eta)^{1/2}. \tag{16}$$

In Paper I, we had used z, which is equal to $\eta/4$. The use of η, instead of z, helps to reduce the size of many such numerical coefficients which are very large and which occur in the course of the evaluation of the integrals and in the final result, and facilitates the derivation. The evaluation of I_6 is similar to that of I_5, the evaluation of which is given in the Appendix of Paper I. Hence we only give the final expression for I_6. For the sake of completeness, we also include the expressions for I_1 to I_5 in terms of η.

$$I_1=\frac{1}{6\pi\lambda(1-\eta)^{3/4}}\{\eta K(\omega)-2(1-\eta)^{1/2}[E(\omega)-(1-\omega^2)K(\omega)]\},$$

$$I_2=\frac{2\lambda(1-\eta)^{3/4}}{3\pi}\{-(1-\eta)^{1/2}K(\omega)-(2+1/\eta)[E(\omega)-(1-\omega^2)K(\omega)]\},$$

$$I_3=\frac{\lambda^3(1-\eta)^{9/4}}{45\pi}\left\{\left[896\eta-708+\frac{33}{\eta}-\frac{56}{\eta^2}\right]K(\omega)\right.$$
$$\left.+(1-\eta)^{1/2}\left[-1792+\frac{72}{\eta}-\frac{138}{\eta^2}+\frac{448}{\eta^3}\right][E(\omega)-(1-\omega^2)K(\omega)]\right\},$$

$$I_4=\frac{16\lambda^5(1-\eta)^{15/4}}{315\pi}\left\{(1-\eta)^{1/2}\left(63488\eta-34464+\frac{277}{\eta}+\frac{587}{4\eta^2}-\frac{441}{\eta^3}+\frac{1984}{\eta^4}\right)K(\omega)\right.$$
$$\left.+\left(126976\eta-100672+\frac{2906}{\eta}-\frac{2135}{2\eta^2}-\frac{10189}{4\eta^3}+\frac{14936}{\eta^4}-\frac{15872}{\eta^5}\right)[E(\omega)-(1-\omega^2)K(\omega)]\right\},$$

$$I_5=\frac{16^2\lambda^7(1-\eta)^{21/4}}{315\pi}\left\{\left(-1560576\eta^2+2078848\eta-584840+\frac{20899}{16\eta}+\frac{32591}{32\eta^2}\right.\right.$$
$$\left.-\frac{7253}{256\eta^3}-\frac{55079}{16\eta^4}+\frac{23532}{\eta^5}-\frac{24384}{\eta^6}\right)K(\omega)$$
$$+(1-\eta)^{1/2}\left(3121152\eta-1816832+\frac{26512}{\eta}-\frac{79123}{8\eta^2}-\frac{87275}{16\eta^3}\right.$$
$$\left.\left.+\frac{316085}{128\eta^4}+\frac{15169}{\eta^5}-\frac{133392}{\eta^6}+\frac{195072}{\eta^7}\right)[E(\omega)-(1-\omega^2)K(\omega)]\right\},$$

$$I_6=-\frac{16^4\lambda^9(1-\eta)^{27/4}}{10395\pi}\left\{(1-\eta)^{1/2}\left(146513920\eta^2-165882304\eta+37772442-\frac{371691}{8\eta}-\frac{18681849}{1024\eta^2}\right.\right.$$
$$\left.+\frac{11535975}{2048\eta^3}+\frac{217913259}{16384\eta^4}+\frac{1595181}{2048\eta^5}-\frac{2828529}{32\eta^6}+\frac{1694833}{2\eta^7}-\frac{1144640}{\eta^8}\right)K(\omega)$$
$$+\left(293027840\eta^2-405021568\eta+124146836-\frac{6249915}{4\eta}+\frac{132448023}{512\eta^2}+\frac{86986521}{512\eta^3}+\frac{128838999}{8192\eta^4}\right.$$
$$\left.\left.-\frac{1865978121}{16384\eta^5}-\frac{43880145}{128\eta^6}+\frac{39272717}{8\eta^7}-\frac{13361012}{\eta^8}+\frac{9157120}{\eta^9}\right)[E(\omega)-(1-\omega^2)K(\omega)]\right\}. \quad (17)$$

IV. RESULTS AND DISCUSSION

The energy eigenvalues were calculated by solving (9). The procedure is explained in Paper I. The parameter λ and the energy E in the present paper (as well as in Paper I) are related to λ and E of Banerjee *et al.*[4,5] as follows:

$\lambda\,(\text{our}) = \frac{1}{2}\lambda\,(\text{Banerjee})$,

$E\,(\text{our}) = \frac{1}{2}E\,(\text{Banerjee})$.

Eigenvalues were calculated for certain values of λ and n for which results have been obtained by Banerjee *et al.*[4,5] We compare our results with those of Banerjee *et al.* in Tables I and II. We shall represent the eigenvalue obtained from a j-term WKBJ approximation by $E^{(j)}$. The series expansion in (4) is, in general, semiconvergent.[11,12] Consequently, if in any case $|I_{j+1}/I_j|$ is greater than 1, the series in (5) has to be truncated at I_j. For medium and large values of λ, such a situation is encountered at very low quantum numbers ($n=0,1$, or 2), where $|I_6/I_5|$ is greater than 1. Such cases are omitted from Tables I and II. Our calculated eigenvalues to 15 significant figures are shown in Table I along with those of Banerjee[5] (represented by E^B) in Table I. The difference between the two is shown as $[E^{(6)}-E^B]\times 10^{15-p}$, where p is the number of digits to the left of the decimal point in a calculated value, in column 5. This type of representation shows the difference in terms of the number of significant figures, without regard to the decimal position. Thus the extreme right hand digit of any number in column 5 of Table I corresponds to the 15th significant figure. Banerjee[5] estimates the accuracy of his results to be ± 1 in the 15th significant figure. The eigenvalues tabulated in column 4 of Table I are half of those given in the paper of Banerjee.[5] Thus one may consider the uncertainty to be $\pm\frac{1}{2}$ in the 15th significant figure, but because of the round-off problem, a more realistic estimate will be ± 1.

At low values of λ ($\leqslant 0.005$), our results are seen to be in excellent agreement with those of Banerjee[5] for all values of the quantum number n. At medium and high values of λ, our results are poor at very small quantum numbers, but there is a rapid improvement as n increases. By $n=10$ a 15-significant-figure accuracy is achieved for $\lambda=0.05$. As λ increases, the quantum number at which this accuracy is attained slowly inches upwards. For $\lambda = 20\,000$, this accuracy would occur at about $n=15$. The trend of errors in column 5 of Table I indicates that at higher quantum numbers, the accuracy is expected to be even better.

TABLE I. Calculated eigenvalues for the anharmonic oscillator from the six-term WKBJ approximation compared with those of Banerjee *et al.*, E^B. p is the number of digits to the left of the decimal point in the calculated value.

λ	n	$E^{(6)}$	E^B	$[E^{(6)}-E^B]\times 10^{15-p}$
0.00005	0	0.500037493440100	0.500037493440100	0
	1	1.50018744846806	1.50018744846806	0
	2	2.50048730796919	2.50048730796920	0
	3	3.50093700833383	3.50093700833383	0
	4	4.50153648602231	4.50153648602231	0
	5	5.50228567756483	5.50228567756485	-2
	6	6.50318451956137	6.50318451956135	2
	7	7.50423294868150	7.50423294868150	0
	8	8.50543090166433	8.50543090166435	-2
	9	9.50677831531834	9.50677831531835	-1
	10	10.5082751265213	10.5082751265213	0
	100	101.247039929621	101.247039929621	0
	1000	1067.12127261611	1067.12127261610	0
0.0005	0	0.500374346336593	0.500374346336595	0
	1	1.50186987408436	1.50186987408437	0
	2	2.50485593639405	2.50485593639406	0
	3	3.50932629602876	3.50932629602876	0
	4	4.51527478303736	4.51527478303736	0
	5	5.52269529358966	5.52269529358965	1
	6	6.53158178883924	6.53158178883925	-1
	7	7.54192829381299	7.54192829381300	-1
	8	8.55372889632674	8.55372889632675	-1
	9	9.56697774592613	9.56697774592615	-2 .
	10	10.5816690528519	10.5816690528519	0
	100	107.219226645926	107.219226645926	0
	1000	1409.24275880910	1409.24275880910	0
0.005	0	0.503686836040688	0.503686836040690	0
	1	1.51826265225667	1.51826265225668	0
	2	2.54696956637115	2.54696956637116	0
	3	3.58928659035025	3.58928659035025	0
	4	4.64473990815594	4.64473990815595	0
	5	5.71289632309315	5.71289632309315	0
	6	6.79335790079480	6.79335790079480	0
	7	7.88575754252129	7.88575754252130	-1
	8	8.98975529185561	8.98975529185560	1
	9	10.1050352263980	10.1050352263980	0
	10	11.2313028210831	11.2313028210831	0
	100	141.681535142538	141.681535142538	0
	1000	2509.87095366631	2509.87095366630	0
0.05	0	0.532635890218554	0.532642754771860	-686455331
	1	1.65343624348236	1.65343600657646	23690590
	2	2.87397963202150	2.87397963441678	-239528
	3	4.17633891279824	4.17633891289288	-9464
	4	5.54929781127881	5.54929781131650	-3769
	5	6.98496309886228	6.98496309887140	-912
	6	8.47739734306955	8.47739734307205	-250
	7	10.0219318020935	10.0219318020943	-8
	8	11.6147760899694	11.6147760899696	-3
	9	13.2527773762682	13.2527773762683	-1
	10	14.9332626173356	14.9332626173357	-0
	100	252.448468315049	252.448468315048	0
	1000	5147.03066134697	5147.03066134695	2
5	2	8.31763939452466	8.31796074620690	-32135168224
	3	12.9031534730702	12.9031381075278	153655424
	4	17.9425849292248	17.9425856111269	-6819021
	5	23.3645404826557	23.3645404504085	322471
	6	29.1206493687030	29.1206493698766	-11736
	7	35.1755259697227	35.1755259696173	1054
	8	41.5019335188017	41.5019335187926	90
	9	48.0781314906026	48.0781314905988	37
	10	54.8862854321677	54.8862854321665	12
	100	1103.21433253216	1103.21433253216	0
	1000	23569.3184776116	23569.3184776116	0
50	2	17.4360350985341	17.4369921309974	-9570324633
	3	27.1926934443284	27.1926457858015	476585268
	4	37.9384998188621	37.9385020143348	-21954728
	5	49.5164187636862	49.5164186577037	1059824
	6	61.8203488092129	61.8203488133390	-41261
	7	74.7728287219740	74.7728287216440	3300
	8	88.3143279788785	88.3143279788570	215
	9	102.397387256482	102.397387256473	10
	10	116.983112938121	116.983112938118	3
	100	2371.08045966044	2371.08045966044	0
	1000	50752.0565830495	50752.0565830495	0
500	2	37.3384925248442	37.3407021000824	-22095752382
	3	58.3017104648547	58.3015994686465	1109962082
	4	81.4011819434597	81.4011870984875	-51550278
	5	106.297091954837	106.297091704867	249970
	6	132.759975829273	132.759975839140	-9867
	7	160.622380137948	160.622380137178	771
	8	189.755655589411	189.755655589364	46
	9	220.057266116850	220.057266116828	22
	10	251.443199642365	251.443199642358	7
	100	5105.67997547800	5105.67997547800	0
	1000	109329.688980615	109329.688980615	0
5000	2	80.3381245572672	80.3429563058560	-48317485888
	3	125.475615120816	125.475371945856	243174960
	4	175.217936794075	175.217948107783	-11313708
	5	228.832288050985	228.832287501845	549140
	6	285.823895787974	285.823895809713	-21739
	7	345.831728819376	345.831728817686	1690
	8	408.578437484469	408.578437484368	100
	9	473.842980833087	473.842980833039	48
	10	541.444259001110	541.444259001095	15
	100	10998.6201372634	10998.6201372635	0
	1000	235537.964187979	235537.964187979	
20000	2	127.501149803157	127.508838644787	-7688841630
	3	199.145510566688	199.145123478029	387088658
	4	278.100219300728	278.100237315262	-18014534
	5	363.201844098700	363.201843224176	874523
	6	453.664874757558	453.664874792195	-34637
	7	548.916140662283	548.916140659590	2693
	8	648.515328513767	648.515328513610	157
	9	752.111522526102	752.111522526025	77
	10	859.417217943562	859.417217943540	22
	100	17458.8967468885	17458.8967468885	0
	1000	373891.710751417	373891.710751417	

Sometimes the degree of anharmonicity is represented by α, where α is related to λ by

$$\alpha = \lambda^{2/3}/(1+\lambda^{2/3}). \tag{18}$$

The two extreme values of α are 0 (harmonic oscillator) and 1 (quartic oscillator). Thus $\alpha = 0.5$ (which corresponds to $\lambda = 1$) may be considered to be a "midway" point between the two extremes. Banerjee *et al.*[4] have calculated eigenvalues for a greater number of quantum numbers for $\lambda = 0.5$ (which corresponds to $\alpha = 0.387$). We have investigated the

TABLE II. Calculated eigenvalues from five- and six-term WKBJ approximations for $\lambda = 0.5$.

λ	n	$E^{(5)}$	$E^{(6)}$	$[E^{(6)} - E^{(5)}]\times 10^{20-p}$	E^B	$[E^{(6)} - E^B]\times 10^{15-p}$
0.5	4	9.0287786535753370553	9.0287786709610363783	173856993230	9.02877871815165	-4719061
	5	11.648720724227815370	11.648720727565973504	3338158134	11.6487207256116	19544
	6	14.417669228861977967	14.417669229695024079	833046112	14.4176692297521	-571
	7	17.320424160311568038	17.320424160563087955	251519917	17.3204241605556	74
	8	20.345193040966719798	20.345193041054480161	87760363	20.3451930410532	13
	9	23.482504752803942995	23.482504752838246326	34303331	23.4825047528377	5
	10	26.724551069818122133	26.724551069832814932	14692799	26.7245510698326	2
	11	30.064761479572176306	30.064761479578961524	6785218	30.0647614795789	1
	12	33.497515000620279447	33.497515000623616717	3337270	33.4975150006236	0
	13	37.017937179549549272	37.017937179551280943	1731671	37.0179371795512	0
	14	40.621752525382643344	40.621752525383584213	940869	40.6217525253836	0
	15	44.305174400399051427	44.305174400399583493	532066	44.3051744003996	0
	16	48.064821022616716574	48.064821022617028203	311629	48.0648210226170	0
	17	51.897650161136117781	51.897650161136306052	188271	51.8976501611365	-2
	18	55.800907522586363033	55.800907522586479963	116930	55.8009075225865	0
	19	59.772085366525081533	59.772085366525155975	74442	59.7720853665250	2
	20	63.808888897677410953	63.808888897677459415	48462	63.8088888976775	0
	21	67.909208662805154631	67.909208662805186824	32193	67.9092086628050	2
	22	72.071097648199060180	72.071097648199081963	21783	72.0710976481990	1
	23	76.292752102786945857	76.292752102786960846	14989	76.2927521027870	0
	24	80.572495347256465501	80.572495347256475975	10474	80.5724953472565	0
	25	84.908764000797666703	84.908764000797674127	7424	84.9087640007975	2
	26	89.300096183437875283	89.300096183437880615	5332	89.3000961834380	-1
	27	93.745121346475157409	93.745121346475161285	3876	93.7451213464750	2
	28	98.242551455110218967	98.242551455110221817	2850	98.2425514551105	-3
	29	102.79117330221175725	102.79117330221175937	212	102.791173302212	0
	30	107.38984177458831113	107.38984177458831272	159	107.389841774589	0
	31	112.03747392630015194	112.03747392630015315	121	112.037473926300	0
	32	116.73304373968758154	116.73304373968758246	92	116.733043739688	0
	33	121.47557747557356105	121.47557747557356176	71	121.475577475574	0
	34	126.26414953074628379	126.26414953074628434	55	126.264149530746	0
	35	131.09787873425992305	131.09787873425992348	43	131.097878734260	0
	36	135.97592502500366384	135.97592502500366418	34	135.975925025003	0
	37	140.89748646190965299	140.89748646190965326	27	140.897486461910	0
	38	145.86179652550667572	145.86179652550667593	20	145.861796525507	0
	39	150.86812167559355163	150.86812167559355180	17	150.868121675594	0
	40	155.91575913485060894	155.91575913485060908	14	155.915759134851	0
	41	161.00403487242263695	161.00403487242263707	12	161.004034872422	0
	42	166.13230176504552823	166.13230176504552832	9	166.132301765045	0
	43	171.29993791627328985	171.29993791627328993	8	171.299937916273	0
	44	176.50634511688983864	176.50634511688983870	6	176.506345116890	0
	45	181.75094743173960891	181.75094743173960896	5	181.750947431739	0
	46	187.03318990004614057	187.03318990004614061	4	187.033189900046	0
	47	192.35253733786026966	192.35253733786026970	4	192.352537337860	0
	48	197.70847323263166370	197.70847323263166373	3	197.708473232632	0
	49	203.10049872106401575	203.10049872106401577	2	203.100498721064	0
	50	208.52813164242393733	208.52813164242393735	2	208.528131642424	0
	100	517.77209156945416761	517.77209156945416761	0	517.772091569455	-1
	1000	10966.391855333463404	10966.391855333463404	0	10966.3918553335	0
	10000	235551.88889540475932	235551.88889540475932	0	235551.888895404	0

question of the dependence of the accuracy of the WKBJ values on n for this value of λ. While there is no known method of obtaining exact error bounds for energy eigenvalues obtained by the WKBJ method, we have shown in Paper I that a study of the behavior of $|E^{(j+1)} - E^{(j)}|$ is helpful in assessing the accuracy.

In Table II, for $\lambda = 0.5$, we show $E^{(5)}$ and $E^{(6)}$ to 20 significant figures, together with Banerjee *et al.*'s values to 15 significant figures. Notice the very regular behavior of the difference $[E^{(6)} - E^{(5)}]\times 10^{20-p}$, shown in column 5. The difference $[E^{(7)} - E^{(6)}]$ is anticipated to be about an order of magnitude smaller than $[E^{(6)} - E^{(5)}]$. Thus if any number in column 5 has t digits, we can be reasonably confident that the corresponding $E^{(6)}$ value is correct to $(20 - t)$ significant figures. This indicates that a 20-significant-figure accuracy for $E^{(6)}$ will be achieved at about $n = 55$. Also, between $n = 5$ and $n = 50$, there appears to be an approximate linear relation between log (number of correct significant figures) and log n. If this relationship persists for higher quantum numbers, we can estimate that at $n = 100$, $E^{(6)}$ can be expected to have a 24 significant figure accuracy. The difference between $E^{(6)}$ and E^{B} is shown in column 7 of Table II. It may be noted that the figures in column 5 differ from the corresponding figures in column 7 by a factor of 10^5. An examination of column 7 appears to indicate that occasionally the uncertainty in the 15th significant figure of E^{B} is somewhat greater than ± 1.

The available evidence indicates that among the various methods which have been used for calculating the eigenvalues of the potential (1), for medium and high quantum numbers, the method of Banerjee *et al.*[4] and the WKBJ method used in the present paper are about the best. Computationally, the six-term WKBJ method is more efficient than the method of Banerjee *et al.*[4] The computer time required to calculate an eigenvalue increases with n for the method of Banerjee *et al.*; they quote a time of about 3 min for $n = 10\,000$ and $\lambda = 0.5$ on IBM 7044. In our case, above $n \sim 10$, the required computer time remains more or less the same for any n. For $n = 10\,000$ and $\lambda = 0.5$ the calculation of the eigenvalue by the six-term WKBJ method took less than 2 sec on CYBER 74.

ACKNOWLEDGMENT

This work was supported in part by research grants from the Natural Sciences and Engineering Research Council of Canada to the authors.

[1]R. N. Kesarwani and Y. P. Varshni, J. Math. Phys. **22**, 1983 (1981).
[2]F. T. Hioe and E. W. Montroll, J. Math. Phys. **16**, 1945 (1975).
[3]F. T. Hioe, E. W. Montroll, and M. Yamawaki, in *Perspectives in Statistical Physics*, edited by H. J. Raveche (North-Holland, Amsterdam, 1981).
[4]K. Banerjee, S. P. Bhatnagar, V. Choudhury, and S. S. Kanwal, Proc. R. Soc. London, Ser. A **360**, 575 (1978).
[5]K. Banerjee, Proc. R. Soc. London, Ser. A **364**, 265 (1978).
[6]K. Banerjee, Lett. Math. Phys. **1**, 323 (1976).
[7]S. N. Biswas, K. Dutta, R. P. Saxena, P. K. Srivastava, and V. S. Varma, J. Math. Phys. **14**, 1190 (1973).
[8]J. L. Dunham, Phys. Rev. **41**, 713, 721 (1932).
[9]R. N. Kesarwani and Y. P. Varshni, J. Math. Phys. **21**, 90 (1980); **21**, 2852 (1980).
[10]C. M. Bender, K. Olaussen, and P. S. Wang, Phys. Rev. D **16**, 1740 (1977).
[11]G. D. Birkoff, Bull. Am. Math. Soc. **39**, 696 (1933).
[12]E. C. Kemble, *The Fundamental Principles of Quantum Mechanics* (Dover, New York, 1958); this is a reprint of the 1937 edition published by McGraw-Hill.

Existence theorem for solutions of Witten's equation and nonnegativity of total mass [a)]

Oscar Reula
Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637

(Received 1 October 1981; accepted for publication 6 November 1981)

We prove that given an asymptotically flat (in a very weak sense) initial data set, there always exists a spinor field that satisfies Witten's equation and that becomes constant at infinity. Thus we fill a gap in Witten's arguments on the nonnegativity of the total mass of an isolated system, when measured at spatial infinity. We also include a review of Witten's argument.

PACS numbers: 04.20.Cv, 02.30. + g

I. INTRODUCTION

For almost fifteen years there has been in general relativity a certain conjecture about the sign of the total mass. It says that for any isolated system whose matter has nonnegative local mass density, its total mass, measured asymptotically, must be nonnegative. This conjecture has concerned many relativists, not only because of the important physical consequences that would arise should it fail, but also because of the general feeling that our understanding of the theory should be such that we can resolve it.

There are two distinct regimes where one can measure the mass (i.e., extract it from the asymptotic behavior of the field), and therefore two distinct conjectures. One regime, at null infinity, yields the Bondi mass[1]; the other, at spatial infinity, the Arnowitt–Deser–Misner (ADM) mass.[2] We are concerned here only with the second regime, i.e., with the issue of the nonnegativity of the ADM mass.

There have been established a number of special cases of this conjecture, in addition to several attempts to prove it in general.[3–13] Recently two general arguments for the conjecture have been given—by Schoen and Yau[14] and by Witten.[15] The validity of Witten's proof depends on the existence, on an asymptotically flat initial data set, of a spinor field which satisfies a certain first-order elliptic equation and which becomes constant at infinity. In this paper we give a proof of the existence and uniqueness of such a spinor field, thereby completing Witten's argument.

Consider an initial data set, (S,h_{ab},π_{ab}), for Einstein equations. It consists of S, a three-dimensional manifold without boundary, h_{ab}, a metric of signature (- - -), and π_{ab}, a second-rank symmetric tensor field. We assume our initial data set to be smooth (C^∞), although this condition can be weakened considerably. We say that initial data set (S,h_{ab},π_{ab}) satisfies the local mass condition if

$$\mu \geqslant |J_a J^a|^{1/2}. \tag{1.1}$$

Here, μ and J_a are the scalar and the vector fields defined by

$$-2\mu \equiv \pi^2 - \pi^{ab}\pi_{ab} - R, \tag{1.2}$$

$$J_a \equiv D^b(\pi_{ab} - \pi h_{ab}), \tag{1.3}$$

where R is the scalar curvature and D_b the covariant derivative of (S,h_{ab}), and $\pi \equiv \pi^a{}_a$. One thinks of S as a Cauchy surface of some space-time, h_{ab} as the induced metric, and π_{ab} as the extrinsic curvature of S. Then Eqs. (1.2) and (1.3) are the constraint equations of general relativity, and (μ,J_a) forms the energy momentum of the matter, as locally measured by an observer at rest with respect to S. Thus, (1.1) is the statement that the local energy momentum of the matter is timelike.

We remark that only the initial data set (S,h_{ab},π_{ab}) itself will be used in what follows while, of course, its physical interpretation lies in the full space-time.

Witten's argument further makes use of spinors on (S,h_{ab}). There follows a summary of their properties. It is well known that in a three-dimensional Riemannian manifold we can define two-component SU(2,C) spinor fields. The spinors at each point form a two-dimensional complex vector space and will be labeled by capital Latin indices, e.g., λ^A. Indices are raised and lowered by the antisymmetric symplectic spinor field ϵ^{AB} and its inverse ϵ_{AB}. The SU(2,-C) structure also gives rise to a Hermitian inner product. It is implemented by an adjoint operation—which maps spinor field λ^A into another spinor field denoted by $\lambda^{A\dagger}$—with the following properties:

$$((\lambda^A)^\dagger)^\dagger = -\lambda^A, \quad \lambda^{A\dagger}\lambda_A \geqslant 0 \quad (= \text{iff } \lambda_A = 0).$$

Three-dimensional complex vectors are represented by symmetric two-spinors, $S_a \rightarrow S_{AB} = S_{(AB)}$, and the metric h_{ab} by $h_{(AB)(CD)} = -\epsilon_{A(C}\epsilon_{D)B}$. The real vectors are those which satisfy

$$S_{AB} = -(S_{AB})^\dagger,$$

and so, for any spinor λ_A, $\lambda^\dagger_{(A}\lambda_{B)}$ is always a real vector. Thus, for example, for S_a real,

$$S_{AB}(S^{AB})^\dagger = -S_{AB}S^{AB} = -S_aS^a > 0.$$

We also introduce a derivative operator on spinors, $D_{AB} = D_{(AB)}$, which satisfies the usual rules of the covariant derivative and which, when acting on a real scalar, yields a real vector.

One can derive all these properties of SU(2,C) spinors by reducing the SL(2,C) spinors of space-time to SU(2,C) on a spacelike hypersurface S (see e.g., Sen[16–17].) Then the adjoint operation arises naturally as $\lambda^{\dagger A} = \sqrt{2}t^{AB'}\bar{\lambda}_{B'}$, where $t^{AB'}$ is the spinor form of the timelike vector orthonormal to S.

We are now ready to review Witten's argument.[15] Introduce the operator $\mathscr{D}_{AB}$ whose action on spinor fields is

[a)]Supported in part by the Aileen S. Andrew Foundation and by the NSF.

$$\mathscr{D}_{AB}\lambda_C = D_{AB}\lambda_C + \frac{1}{\sqrt{2}}\pi_{ABC}{}^D\lambda_D, \tag{1.4}$$

where $\pi_{ABCD} = \pi_{(AB)(CD)}$ is the spinor representation of the tensor field π_{ab}.[18] Consider now the identity,[19]

$$\begin{aligned}\int_S \lambda^{\dagger A}(\mathscr{D}_{AB}\mathscr{D}^B{}_C\lambda^C)\,dV &= \frac{1}{2}\int_S \{(\mathscr{D}^{AB}\lambda^C)^\dagger(\mathscr{D}_{AB}\lambda_C) \\ &\quad + \tfrac{1}{2}(\mu\epsilon^N{}_M - \sqrt{2}\,J^N{}_M)\lambda^{\dagger M}\lambda_N\}\,dV \\ &\quad + \frac{1}{2}\int_S D^a(\lambda^{\dagger C}\mathscr{D}_a\lambda_C)\,dV,\end{aligned} \tag{1.5}$$

obtained by integrating the left-hand side by parts, and then using the constraint equations (1.2) and (1.3), and the contracted Ricci Identity, $D^A_{(B}D_{|A|C)}\tau^C = -(R/8)\tau_B$.

Now, let β^C be any solution of

$$\mathscr{D}_{AB}\mathscr{D}^B{}_C\beta^C = 0, \tag{1.6}$$

with β^C going asymptotically to a constant spinor β_0^C.[20] Then, for this β^C the left-hand side of identity (1.5) is zero and therefore, using Gauss's theorem, we obtain

$$\begin{aligned}\lim\int_\Sigma [\beta^\dagger_C\mathscr{D}_a\beta^C]\,d\Sigma^a = \int_S\{(\mathscr{D}^{AB}\beta^C)^\dagger(\mathscr{D}_{AB}\beta_C) + \tfrac{1}{2}(\mu\epsilon^N{}_M - \sqrt{2}\,J^N{}_M)\beta^{\dagger M}\beta_N\}\,dV,\end{aligned}$$

where the limit is that in which Σ, a topologically S^2 surface, recedes to infinity and $d\Sigma^a$ is the surface element normal to Σ. Witten has argued that the limit of the surface integral in (1.7), provided the initial data set is asymptotically flat in a suitable sense, is given by

$$\lim\int_\Sigma [\beta^\dagger_C\mathscr{D}_a\beta^C]\,d\Sigma^a = \tfrac{1}{2}(E\epsilon^N{}_M - \sqrt{2}P^N{}_M)\beta_0^{\dagger M}\beta_{0N}, \tag{1.8}$$

where E and $P^N{}_M$ (regarded as an asymptotically constant vector field) are the ADM expressions for the energy and momentum, respectively, and β_0^N is the value of β^N at infinity. Thus, since for an initial data set satisfying the local mass condition the left-hand side of (1.8) is explicitly nonnegative, and since β_0^C is arbitrary, we conclude with Witten that

$$E \geqslant |P_aP^a|^{1/2}, \tag{1.9}$$

that is, that the total mass is nonnegative.

It immediately follows from (1.7) and (1.8) that the vanishing of E (and hence P_a) occurs only if both $\mu = 0$ (and hence $J_a = 0$), and $\mathscr{D}_{AB}\beta_C = 0$. It has been shown by Sen[17] that any space-time having an initial data set that admits a nonzero spinor field satisfying $\mathscr{D}_{AB}\beta_C = 0$, is at least of Petrov type [3,1]. It is very likely that the only asymptotically flat space-time of type [3,1] or more special is flat space-time. Thus, one expects that the only one having zero mass is Minkowski space-time.

To summarize, Witten's argument shows that the total mass is nonnegative, is the sense of (1.9), provided there exists a spinor field β^C satisfying Eq. (1.6).

Thus, to complete Witten's proof, it suffices to show the existence of such a solution.

Existence Theorem: Let (S,h_{ab},π_{ab}) be an initial data set that satisfies the local mass condition and is asymptotically flat (as defined in Sec. II). Then, for any β_0^C, a constant spinor field (in the sense of Sec. II), there exists a spinor field β^C, satisfying the equation

$$\mathscr{D}_{AB}\mathscr{D}^B_C\beta^C = 0,$$

with β^C approaching β_0^C at infinity.

Its proof is given in Sec. II. As suggested by O'Murchadha, it is based on the fact that Eq. (1.6) comes from a variational principle. We also show there that the solution is unique.

II. EXISTENCE AND UNIQUENESS

Before proving the theorem, we must define asymptotic flatness of an initial data set. Our definition, considerably weaker than those commonly used, requires only what we shall need in the proof.[21]

Definition: An initial data set (S,h_{ab},π_{ab}) is said to be *asymptotically flat* provided:

(1) There exists a flat metric η_{ab} on $S-K$, where K is a compact set, such that for some $C>0$ and any vector l^a,

$$C^{-1}\eta_{ab}l^al^b \leqslant h_{ab}l^al^b \leqslant C\eta_{ab}l^al^b,$$

and $(S-K,\ \eta_{ab})$ consists of a finite number of connected components, each of them isometric to Euclidean space minus a ball.

(2) $|\underset{\eta}{\nabla}_a h_{bc}|^2$, $\pi^{ab}\pi_{ab}$, μ, and $|J^a|$ are all integrable. Here $\underset{\eta}{\nabla}_a$ denotes the covariant derivative on (S,η_{ab}).

It will be immediate, from the proof of the theorem, that the above conditions are the weakest possible that make the integral in the right-hand side of (1.7), and hence the total mass, to be finite.

The plan of the proof is as follows: First write β^C and β_0^C plus a spinor field $\hat\beta^C$, which belongs to a certain Hilbert space H. Substitution into (1.6) results in an inhomogeneous equation on $\hat\beta^C$. We then define a certain complex-valued, continuous, linear functional on H. It gives rise, by the Riesz Theorem, to a spinor $\hat\beta^C$, solution of the weak (in the sense of distributions) form of the inhomogeneous equation. The elliptic character of the equation ensures that this solution is in fact smooth, and so a solution of the strong form. Finally we prove uniqueness.

We first define the Hilbert space H. Let H be the Hilbert space obtained as the completion of C_0^∞ (smooth and compact support) spinor fields under the norm

$$\|\sigma\|_H^2 \equiv \int_S (\mathscr{D}^A_B\sigma^B)^\dagger(\mathscr{D}_{AC}\sigma^C)\,dV. \tag{2.1}$$

Note that for C_0^∞ spinors (2.1) is positive definite [as follows from integrating the right-hand side of identity (1.5)]. Thus, (2.1) is indeed a norm and therefore H a Hilbert space.

Two important properties of the elements of H will be derived from the following Lemma.

Generalized Sobolev Lemma: Let (S,h_{ab},π_{ab}) be an asymptotically flat initial data set satisfying the local mass condition. Then H, defined as above, is a subset of L_6.[22] That is, for some $C>0$,

$$\|\lambda\|_H^3 \geqslant C\int_S (\lambda^{\dagger A}\lambda_A)^3\,dV$$

for any $\lambda^A \in H$.

This lemma is proved in Appendix A by generalizing the standard Sobolev lemma[23] on flat R^3, as suggested by Geroch. A similar lemma has been discussed by Schoen and Yau.[14]

It is immediate from the generalized Sobolev lemma that the elements on H are measurable spinor fields (and so are distributions). There also follows from the lemma the result that, given any C_0^∞ Cauchy sequence on H α_n^A, (hence, converging on L_6 to some α^A), $\mathscr{D}^B{}_A \alpha_n^A$ weakly converges to the distributional derivative $\mathscr{D}^B{}_A\alpha^A$ of α^A. To see this, consider any σ^A, smooth and of compact support. Then

$$\lim_{n\to\infty}\left|\int_S [\sigma^{\dagger A}\mathscr{D}_{AB}\alpha_n^B - (\mathscr{D}_{BA}\sigma^A)^\dagger\alpha^B]\,dV\right|$$

$$= \lim_{n\to\infty}\left|\int_S (\mathscr{D}_{BA}\sigma^A)^\dagger(\alpha_n^B - \alpha^B)\,dV\right|$$

$$\leqslant \lim_{n\to\infty} \|\mathscr{D}_{BA}\sigma^A\|_{L_{6/5}}\|\alpha_n^B - \alpha^B\|_{L_6}$$

$$= 0,$$

where we have used Hölder's inequality. From these results, in turn, it is clear that the expression for the norm, (2.1), not only makes sense for C_0^∞ spinors but for any spinor field in H. From now on, we shall consider all derivatives on elements of H as distributional derivatives (note that, by construction, the first derivatives are in L_2). This completes our discussion of H.

We now split β^A, the solution we are seeking, into two parts: a smooth spinor field β_0^A, which goes to constant at infinity (in the sense that $\nabla_a\beta_0^A = 0$ outside a compact region), $\hat{\beta}^A$, some spinor field. We now demand $\hat{\beta}^A$ to be in H, thus making precise the requirement that β^A approaches β_0^A asymptotically. Now, Eq. (1.6) becomes

$$\mathscr{D}_A{}^B\mathscr{D}_{BC}\hat{\beta}^C = \mathscr{D}_A{}^B\rho_B, \qquad (2.2)$$

where

$$\rho^A \equiv -\mathscr{D}^A{}_C\beta_0^C. \qquad (2.3)$$

From the definition of asymptotic flatness we see that ρ^A is square integrable ($\|\rho^A\|_{L_2} < \infty$). Thus the linear functional defined by

$$f(\sigma^A) \equiv \int_S (\mathscr{D}^A{}_B\sigma^B)^\dagger\rho_A\,dV \qquad (2.4)$$

for all $\sigma^A \in H$, satisfies

$$|f(\sigma^A)| \leqslant \|\sigma^A\|_H\|\rho\|_{L_2}$$

for all $\sigma^A \in H$. So, $f(\cdot)$ is a bounded linear functional on H, and therefore it is continuous. Then, by the Riesz theorem, there exists $\hat{\beta}^A$ in H such that

$$f(\sigma^A) = \int_S (\mathscr{D}^A{}_B\sigma^B)^\dagger(\mathscr{D}_{AC}\hat{\beta}^C)\,dV \qquad (2.5)$$

for all $\sigma^A \in H$. That is, by (2.4),

$$\int_S (\mathscr{D}^A{}_B\sigma^B)^\dagger\{\mathscr{D}_{AC}\hat{\beta}^C - \rho_A\}\,dV = 0 \qquad (2.6)$$

for all $\sigma^A \in H$.

Restricting now σ^A to be smooth and of compact support, and integrating (2.6) by parts, we get

$$\int_S \{(\mathscr{D}_{CA}\mathscr{D}^A{}_B\sigma^B)^\dagger\hat{\beta}^C - \sigma^{\dagger B}\mathscr{D}_{BL}\rho^L\}\,dV = 0, \qquad (2.7)$$

which is precisely the weak form of Eq. (2.2).

From standard theorems on elliptic equations,[24] we now conclude that the weak solution of (2.2) is a C^∞ solution. This completes the proof of existence.

There remains to prove uniqueness, that is, to show that given β_0^A, the spinor field $\hat{\beta}^A$ is determined uniquely on H. It suffices to show that Eq. (1.6) has no nontrivial solutions in H. But were $\alpha^A \in H$ a solution of (1.6) and α_n^A a C_0^∞ sequence converging to α^A in H, then, by integrating by parts the integral over S of $\alpha_n^{\dagger A}(\mathscr{D}_{AB}\mathscr{D}^B{}_C\alpha^C)$, we would obtain, in the limit $n\to\infty$, $\|\alpha\|_H = 0$ and so $\alpha^A = 0$. This concludes the proof of the theorem.

III. CONCLUSION

The existence theorem proved in Sec. II fills the gap in Witten's argument, thus making rigorous his proof of the nonnegativity of total mass.

It might be possible to prove a similar theorem, using the new methods for elliptical equations on manifolds Euclidean at infinity developed by Choquet–Bruhat and Christodoulou.[25] But it is very likely, even were this possible, that there would be required much stronger conditions on the asymptotic flatness of the initial data set than the ones we have used. There would result, then, a weaker version of the present existence theorem. But, as we shall see, the use of weaker boundary conditions becomes of interest due to the possibility of extending the definition of total mass to include those systems which are only asymptotically flat in the weak sense of our definition.

We have been using the second-order equation (1.6) instead of Witten's first order equation, $\mathscr{D}^A{}_B\beta^B = 0$, because the former is all that is needed to prove the nonnegativity of mass. On the other hand, the latter arises naturally in supergravity, and so is in some sense more physical. However, all our solutions of (1.6) also satisfy Witten's equation.[26] Thus, the two versions are equivalent.

To define the total energy momentum of an isolated system one generally starts with a very small class of initial data sets, for which it is easy to determine some expressions for E and P_a and to give physical arguments (such as the existence of conservation laws or of a canonical formalism) to justify that they are the right expressions. The definition is then corroborated in the weak field approximation and by applying it to exact solutions. After that, the definition is extended to all systems for which the expressions make sense (e.g., give finite values for E and P_a). Until Witten's paper[15] all such expressions were either surface or volume integrals with non-positive-definite integrands. Thus, because of possible cancellations in the integrals for E and P_a, it could not immediately be stated which are the weakest asymptotic conditions a system must satisfy in order to have a well-defined total energy-momentum. But now we have a volume integral with positive definite integrand that, according to Witten's arguments, under suitable boundary conditions yields the same answer for E and P_a, as do the ADM expressions[2] [via Eqs. (1.7) and (1.8)]. But by the present existence theorem, the volume integral makes sense under far weaker boundary conditions, namely those of Sec. II. This circumstance suggests that energy and momentum be defined in terms of that volume integral [the one of Eq. (1.7)]. Thus, with this definition we extend—in some sense maximally—the class of initial data sets admitting a finite total energy and linear momentum.[27]

This generalization has the added advantage that the integrand of the volume integral in (1.7) [where β^c is a solution of (1.6)], acquires essentially the meaning of an energy-momentum density. (Actually, it is the component along the null direction $\beta^A\bar{\beta}^{A'}$ of the local contribution to the total energy momentum.) This density, the integrand on the right-hand side of (1.7), contains not only contributions from the matter fields, the second term, but also those from the gravitational field, the first term. One expects physically that the local energy momentum of matter and gravity would contribute to the total energy momentum through a red shift factor, which will depend on the configuration of the entire system. In fact it is precisely the β^C's that play the role of red shift factors and the nonlocality is reflected in the β^C's satisfying an elliptic equation.

Further one obtains, in the following way, a variational principal for total mass. First introduce the normed vector space V of spinor fields β^A of the form $\beta^A = \beta^A_0 + \hat{\beta}^A$, where β^A_0 is any smooth spinor field that is asymptotically constant in the sense of Sec. II and $\hat{\beta}_C$ is any spinor field belonging to H. The norm is chosen to be the square root of the right-hand side of (1.7). Consider now the hyperplane in V obtained by fixing β^C_0. That spinor field in this hyperplane which minimizes the V norm exists and is precisely our solution of Eq. (1.6) corresponding to this particular fixed value of β^C_0 Thus the squared norm of this minimizing spinor field is just $\frac{1}{2}(E\epsilon_M{}^N - \sqrt{2}\,P_M{}^N)\beta_0^{\dagger M}\beta_{0N}$. So we obtain a variational principle, for which the spinor field β^A, used in the expression for E and P_a, is just the one which minimizes $\frac{1}{2}(E\epsilon_M{}^N - \sqrt{2}P_M{}^N) \times\beta_0^{\dagger M}\beta_{0N}$.

Thus the present definition of E and P_a,—as given in terms of the volume integral of a density—is suggested by the fact that it agrees with the ADM definition, when the latter is applicable, and strictly generalizes it. Further, this definition fits naturally into the framework of a variational principle. Can a case be made that this version is the more fundamental? Some insight might be gleaned from a study of the linear case.

Note added in proof: After the completion of this work, we learned of a simultaneous work of T. Parker and C. Taubes substantially overlapping ours.

ACKNOWLEDGMENTS

I am very grateful to Robert Geroch for inspiration, guidance, and patience through all this work. Without his advice this paper would not have been possible. I wish to thank also Robert Wald, Niall O'Murchadha, and Prasun Kundu for many discussions and suggestions.

APPENDIX: PROOF OF THE GENERALIZED SOBOLEV LEMMA

It suffices to prove that for any C^∞_0 functions f, the inequality

$$\int_S |D_a f|^2\, dV \geqslant c\left[\int_S |f|^6\, dV\right]^{1/3} \tag{A1}$$

holds on (S,h_{ab}). To see that this suffices, first note that we can use the form of the H norm given by the first integral in the right-hand side of identity (1.5), and that nothing is lost by dropping the second term in this integral, since it is always nonnegative. Second, note that we can pass to scalar functions, since the inequality

$$|D_a(\lambda^{\dagger C}\lambda_C)^{1/2}| = \tfrac{1}{2}|\lambda^{\dagger C}\mathscr{D}_a\lambda_C + \lambda_C\mathscr{D}_a\lambda^{\dagger C}|(\lambda^{\dagger M}\lambda_M)^{-1/2} \leqslant |\mathscr{D}_a\lambda_C|$$

holds for any λ_C, where we have used the symmetry of π_{ab}. Finally, note that once (A1) is proved for C^∞_0 functions, the lemma will then follow by Cauchy completion in H.

But it even suffices to prove that (A1) holds on every set of some finite cover of S. Indeed, following Schoen and Yau, let, for contradiction, (A1) fail for S. That is, let there exist a sequence of C^∞_0 functions, f_i, such that both

$$\text{(a)}\quad \int_S |D_a f_i|^2\, dV \to 0 \quad\text{and}\quad \text{(b)}\quad \int_S |f_i|^6\, dV = 1.$$

Now consider any finite cover of S. Then, for every set of the finite cover (a) holds, and for at least one of them the left side of (b) must be bounded away from zero. But this contradicts that (A1) holds for every set of some finite cover.

The proof is now completed by showing that there does exist such a cover. Inequality (A1) is well known to hold on flat R^3 (see, e.g., Ref. 23). It still holds on flat "$\frac{1}{2}R^3$", that is on the closure of one side of a plane. This can be shown by just performing the x-integration in Ref. 23 between $(-\infty, 0)$ instead of between $(-\infty, +\infty)$. And finally (A1) still holds when the flat metric on $\frac{1}{2}R^3$ is replaced by any curved metric bounded by it (in the sense of definition of asymptotic flat-

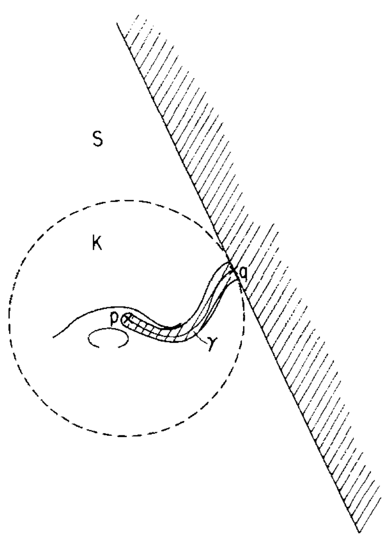


FIG. 1. The Riemannian three-manifold (S,h_{ab}) with all nontrivial topology inside the compact region K. To prove the generalized Sobolev lemma we show that (S,h_{ab}) can be covered by sets like the one shaded above.

ness). This follows since some multiple of each side of (A1) with the flat metric bounds the same side with the curved metric.

All that remains is to prove that a finite number of sets of this form covers S. Nothing is lost by considering K, the compact set (in the definition of asymptotic flatness) containing all the nontrivial topology of S, to have as its boundary a sphere [with respect to the flat background metric of $(S-K)$]. Now let **p** by any point of K, and **q** any point of ∂K (the boundary of K). By connectedness, there exists a smooth, non-self-intersecting curve γ in K, joining **p** to **q**, as in Fig. 1. Consider the subset T, formed by taking the closure of the union of (i) a tubular neighborhood of γ and (ii) that region lying in that side, not containing K, of the plane (with respect to the flat background metric) tangent to K at **q**. But, using the η_{ab} on $S-K$ whose existence is guaranteed by the definition of asymptotic flatness, this (T,h_{ab}) is of the required form.[28] Thus, inequality (A1) holds on this (T,h_{ab}). Since **p** is arbitrary, T sets cover K, and since **q** is arbitrary, they also cover S. But K is compact and so a finite number must cover S.

[1]H. Bondi, M. G. J. Van Der Berg, and A. W. K. Metzner, Proc. R. Soc. London Ser. A **269**, 21 (1962); R. Sachs, Proc. R. Soc. London Ser. A **270**, 193 (1962).
[2]R. Arnowitt, S. Deser, and C. W. Misner, Phys. Rev. **118**, 1100 (1960); **121**, 1556 (1961); **122**, 997 (1961); R. Geroch, J. Math. Phys. **13**, 956 (1972).
[3]D. Brill, Ann. Phys. **7**, 466 (1959).
[4]D. Brill and S. Deser, Am. Phys. **50**, 548 (1968).
[5]D. Brill, S. Deser, and L. Fadeev, Phys. Lett. A **26**, 538 (1968).
[6]R. Arnowitt, S. Deser, and C. W. Misner, Ann. Phys. (N.Y.) **11**, 116 (1960).
[7]C. Leibovitz and W. Israel, Phys. Rev. D **1**, 3226 (1970).
[8]C. W. Misner, *Astrophysics and General Relativity*, edited by M. Chretieu, S. Deser, and J. Goldstein (Gordon and Breach, New York, 1968).
[9]R. Geroch, Ann. N.Y. Acad. Sci. **224**, 108 (1973).
[10]P. S. Jang, J. Math. Phys. **17**, 141 (1976).
[11]N. O'Murchadha and J. W. York, Jr., Phys. Rev. D 10, 2345 (1974).
[12]M. Leite, Ph.D. Dissertation, M.I.T., 1977 (unpublished).
[13]P. S. Jang, J. Math. Phys. **19**, 1155 (1978).
[14]P. Schoen and S. T. Yau, Phys. Rev. Lett. **42**, 547 (1979).
[15]E. Witten, Commun. Math. Phys. **80**, 381 (1981).
[16]A. Sen, "A Quantum Theory of Spin 3/2 Field in Einstein Spaces," to appear in Int. J. Theor. Phys.
[17]A. Sen, J. Math. Phys. **22** (8), 1781 (1981).
[18]With the adjoint given by $\lambda^{\dagger A}=\sqrt{2}t^{AB'}\bar{\lambda}_{B'}$, one gets $\mathscr{D}_{AB}\lambda_c=\sqrt{2}t_{(A}{}^{B'}\nabla_{B)B'}\lambda_c$, where $\nabla_{BB'}$ is the spinor representation of the space-time covariant derivative. Thus $\mathscr{D}_{AB}\lambda_c$ is the neutrino zero mode equation or Witten's equation.
[19]This identity is the spinorial version of the Green identity for scalars.
[20]It is convenient for simplicity to use the second order equation (1.6) instead of Witten's equation. $\mathscr{D}_{AB}\beta^B=0$. In fact, they are equivalent, in the sense that every solution of (1.6) found in Sec. II satisfies Witten's equation.[25]
[21]In fact, our definition is so weak that it even admits initial data sets with metrics not "approaching" the metric η_{ab}, of the definition, in all directions. As an example consider the "spiked" space with metric $(1+\exp[-x^2y^2z^2])\eta_{ab}$.
[22]L_p denotes the Banach space of measurable functions on S such that $|f|^p$ is integrable. Its norm is given by $\|f\|_{L_p}\equiv(\int_s|f|^p\,dV)^{1/p}$.
[23]C. M. Morrey, *Multiple Integrals in the Calculus of Variations* (Springer, New York, 1966), p. 81.
[24]A. Friedman, *Partial Differential Equations* (Holt, Reinhart, and Winston, New York, 1969), p. 54.
[25]Y. Choquet-Bruhat and D. Cristodoulou, Acta Math. **146**, 129 (1981).
[26]To see this, note that for any of our solutions of (1.6), β^A, $\lambda^B=\mathscr{D}^B{}_A\beta^A$ is square integrable. Thus, it suffices to show that the equation $\mathscr{D}^C{}_B\lambda^B=0$ has no nontrivial solution in $L_2\cap C^\infty$. Assume, on the contrary, there exists one, λ^A. Then, using identity (1.5) for $\mathscr{D}^A{}_C\mathscr{D}^C{}_B\lambda^B=0$, and the local mass condition (1.9) we get

$$\int_{S(r)}(\mathscr{D}^{AB}\lambda^C)^\dagger(\mathscr{D}_{AB}\lambda_C)\,dV\leqslant\left|\int_{\Sigma(r)}(\lambda^{\dagger C}\mathscr{D}_a\lambda_C)\,d\Sigma^a\right|.$$

Here, $\Sigma(r)$, with $r\in[0,\infty)$, denotes a typical element of a one-parameter family of nested surfaces (topologically S^2), and $S(r)$ denotes the region enclosed by the $\Sigma(r)$. The surfaces and the parameters are to be chosen so that $dV\geqslant dr\,d\Sigma$. Calling the left-hand side of th expression above $g^2(r)$, integrating it over $r\in[0,x]$, and using the Hölder inequality, we get $\int_0^x g^2(r)\,dr\leqslant cg(x)$, where $c^2=\int_S(\lambda^{\dagger C}\lambda_C)\,dV$. But one can show there exists no function g which is positive, nondecreasing, and defined everywhere on $[0,\infty)$ and which satisfies the above inequality.
[27]It is immediate, from the form of β^C, that the volume integral in (1.7) is finite for any asymptotically flat initial data set, in the sense of Sec. II.
[28]To see this, first note that T is diffeomorphic to $\frac{1}{2}R^3$. The required flat metric on T is then chosen to agree with η_{ab} outside some compact set.

Linear transport in nonhomogeneous media. II [a)]

Edward W. Larsen
Theoretical Division, University of California, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

G. C. Pomraning and V. C. Badham
School of Engineering and Applied Science, University of California, Los Angeles, California 90024

(Received 26 August 1981; accepted for publication 8 December 1981)

Elementary continuum and discrete solutions are constructed for transport equations with scattering ratios which are bilinear functions of position. Numerical results are given for some albedo problems.

PACS numbers: 05.60. + w, 95.30.Jx, 42.68.Db

I. INTRODUCTION

Recently, a considerable amount of research in analytical neutron transport theory and radiative transfer has been directed toward problems of the form

$$\mu \frac{\partial}{\partial z} \psi(z,\mu) + \psi(z,\mu) = \frac{c(z)}{2} \int_{-1}^{1} \psi(z,\mu')d\mu', \quad 0<z<\infty, \tag{1.1}$$

$$\psi(0,\mu) = g(\mu), \quad 0<\mu\leqslant 1, \tag{1.2}$$

$$\lim_{z\to\infty} \psi(z,\mu) = 0, \tag{1.3}$$

where $c(z)$ is a continuous function of position.[1–13] (See Ref. 12 for a review, and Refs. 14 and 15 for a similar transport equation describing the flow of neutral particles through a plasma.) Most of the research of Refs. 1–13 has treated

$$c(z) = c_0 e^{-z/s} \tag{1.4}$$

or

$$c(z) = \int_0^\infty \alpha(t) e^{-z/t} dt, \tag{1.5a}$$

where

$$\int_0^\infty |t\alpha(t)|^2 dt < \infty. \tag{1.5b}$$

For these choices of c, elementary continuum (distributional) solutions of Eqs. (1.1) and (1.3) have been constructed which are complete on the half-range $0<\mu\leqslant 1$; thus, a certain unique linear superposition of these solutions will solve the full problem (1.1)–(1.3). For the case of a homogeneous medium with scattering,

$$c(z) = c_0, \quad 0\leqslant c_0 \leqslant 1, \tag{1.6}$$

a discrete solution must be added to the set of continuum solutions to obtain half-range completeness.[16] The discrete solution exists for Eq. (1.6) but not for Eqs. (1.4) or (1.5) because $c\to 0$ sufficiently fast as $z\to\infty$ for these latter functions c; thus for large z the medium is essentially a pure absorber for which there is no discrete solution.[16]

A nonhomogeneous medium which does not become a pure absorber as $z\to\infty$ has been discussed by Pomraning and Larsen[3]; it is defined by

$$c(z) = \frac{c_1 s + c_2 z}{s+z}, \quad 0\leqslant c_1, c_2 \leqslant 1, \quad 0<s<\infty. \tag{1.7}$$

This function varies monotonically from its $z=0$ value (c_1) to its $z=\infty$ value (c_2). Continuum ($-1\leqslant\nu\leqslant 1$), and [provided c_1, c_2, and s satisfy a certain condition] one discrete ($\nu=\nu_0$) solution are constructed in Ref. 3. The continuum solutions, however, contain the usual delta and principal-value functions *and* their first derivatives, and thus are not well suited for numerical computation.

In this article we reconsider the half-space characterized by Eq. (1.7) and derive more general and more useful results than in Ref. 3. In Sec. II we show how to construct, for all c_1 and c_2, a set of continuum solutions ($-1\leqslant\nu\leqslant 1$) which contain only the usual delta and principal-value distributions, and which are much more suitable for numerical calculations than those derived in Ref. 3. In Sec. III we construct, for all $c_1<c_2$ and all s, a discrete solution which decays as $z\to\infty$. In Sec. IV we construct a discrete solution which decays as $z\to\infty$ for all $c_1>c_2$ and for an infinite denumerable number of choices of s. Finally, in Sec. V, we use the solutions constructed in Secs. II and III to obtain numerical results for some albedo problems via the F_N method.[2,7,13,15] Unfortunately, we cannot at the present obtain analytic solutions; we discuss this also in Sec. V.

II. CONTINUUM SOLUTIONS

The general ansatz for continuum solutions of Eqs. (1.1) and (1.3), with $c(z)$ given by Eq. (1.7), is

$$\psi(z,\mu) = A(\mu)e^{-z/\mu} + \int_0^1 \frac{\nu B(\nu)}{\nu-\mu} e^{-z/\nu} d\nu, \tag{2.1a}$$

$$0 = A(\mu), \quad -1\leqslant\mu<0. \tag{2.1b}$$

Introducing Eq. (2.1) into Eq. (1.1) one can—after lengthy manipulations described below—obtain $A(\mu)$ explicitly as a function of $B(\mu)$. Thus for "any" $B(\mu)$, the resulting pair of functions (A, B), when introduced into Eq. (2.1), gives a solution of the transport equation. If in particular one chooses $B_{\nu'}(\nu) = \frac{1}{2}c_2\delta(\nu-\nu')$, then elementary solutions are obtained which are parametrized by ν'. [The factor $\frac{1}{2}c_2$ is a convenient normalization.]

An easier way to derive these elementary solutions, however, is to set $B_{\nu'}(\nu) = \frac{1}{2}c_2\delta(\nu'-\nu)$ at the outset. Doing

[a)] Work by the first author (E.W.L.) was performed under the auspices of the U. S. Department of Energy, and work by the second and third authors (G.C.P. and V.C.B.) was partly supported by the National Science Foundation under grant CPE-77-24992.

 0022-2488/82/050815-07$02.50

this, and then replacing ν' by ν, Eqs. (2.1) become

$$\psi_\nu(z,\mu) = A_\nu(\mu)e^{-z/\mu} + \frac{c_2\nu}{2}\frac{1}{\nu-\mu}e^{-z/\nu}, \quad 0\leqslant\nu\leqslant 1, \tag{2.2a}$$

$$0 = A_\mu(\mu), \quad -1\leqslant\mu<0. \tag{2.2b}$$

We now rewrite Eqs. (1.1) and (1.7) as

$$0 = (s+z)\left(\mu\frac{\partial\psi}{\partial z}+\psi\right) - (c_1 s + c_2 z)\left(\frac{1}{2}\int_{-1}^{1}\psi d\mu'\right). \tag{2.3}$$

Introducing Eqs. (2.2) into Eq. (2.3) and rearranging gives

$$0 = c_2[s\lambda_1(\nu)+z\lambda_2(\nu)]e^{-z/\nu} - [sc_1+zc_2]\int_0^1 A_\nu(\mu)e^{-z/\mu}d\mu, \tag{2.4}$$

where

$$\lambda_i(\nu) = 1 - \frac{c_i\nu}{2}\int_{-1}^{1}\frac{d\mu}{\nu-\mu}, \quad i=1,2 \tag{2.5}$$

are the usual dispersion functions for the homogeneous $c=c_i$ media.[16]

To solve Eq. (2.4), we take

$$A_\nu(\mu) = a_\nu(\mu) + b_\nu\delta(\nu-\mu), \tag{2.6a}$$

$$0 = a_\nu(\mu), \quad -1\leqslant\mu<0, \quad \nu<\mu\leqslant 1 \tag{2.6b}$$

and we use the integration by parts formula

$$z\int_0^\nu a_\nu(\mu)e^{-z/\mu}d\mu = \mu^2 a_\nu(\mu)e^{-z/\mu}\big|_0^\nu - \int_0^\nu\left[\frac{d}{d\mu}\mu^2 a_\nu(\mu)\right]e^{-z/\mu}d\mu. \tag{2.7}$$

Introducing Eqs. (2.6) and (2.7) into Eq. (2.4) now gives

$$\begin{aligned}0 = {} & [\lambda_2(\nu)-b_\nu]ze^{-z/\nu} \\ & + \left[s\lambda_1(\nu) - s\frac{c_1}{c_2}b_\nu - \nu^2 a_\nu(\nu)\right]e^{-z/\nu} \\ & + \int_0^\nu\left[\frac{d}{d\mu}\mu^2 a_\nu(\mu) - \frac{c_1}{c_2}sa_\nu(\mu)\right]e^{-z/\mu}d\mu \\ & + \left[\lim_{\mu\to 0}\mu^2 a_\nu(\mu)e^{-z/\mu}\right].\end{aligned} \tag{2.8}$$

This equation is satisfied if each term in brackets is zero. The first two such terms and Eq. (2.5) give

$$b_\nu = \lambda_2(\nu), \tag{2.9}$$

$$a_\nu(\nu) = \frac{s}{\nu^2}\left(1-\frac{c_1}{c_2}\right). \tag{2.10}$$

The third term and Eq. (2.10) give

$$a_\nu(\mu) = s\left(1-\frac{c_1}{c_2}\right)\frac{1}{\mu^2}e^{-(c_1 s/c_2)(1/\mu - 1/\nu)}, \tag{2.11}$$

which automatically causes the fourth term in Eq. (2.8) to vanish.

To summarize, we have constructed for all c_1 and c_2 the following family of elementary solutions of Eqs. (1.1), (1.3), and (1.7):

$$\psi_\nu(z,\mu) = \left[\lambda_2(\nu)\delta(\nu-\mu) + \frac{c_2\nu}{2}\frac{1}{\nu-\mu}\right]e^{-z/\nu} + \alpha_\nu(\mu)e^{-z/\mu}, \quad 0\leqslant\nu\leqslant 1 \tag{2.12}$$

where

$$\alpha_\nu(\mu) = \begin{cases}0, & -1\leqslant\mu\leqslant 0 \text{ and } \nu<\mu\leqslant 1, \\ s\left(1-\dfrac{c_1}{c_2}\right)\dfrac{1}{\mu^2}\exp\left[\dfrac{c_1 s}{c_2}\left(\dfrac{1}{\nu}-\dfrac{1}{\mu}\right)\right], & 0<\mu\leqslant\nu.\end{cases} \tag{2.13}$$

These solutions consist of the standard continuum solutions for the homogeneous $c=c_2$ medium[16] plus an extra term which is not a distribution.

To conclude this section, we note that a procedure completely analogous to the above can be used to generate continuum solutions, which correspond to $-1\leqslant\nu\leqslant 0$, and which grow rather than decay as $z\to\infty$. These extra solutions play no role in half-space problems, but they will enter into the solutions of finite slab problems.

III. DISCRETE MODE FOR $C_1 < C_2$

For $\nu>1$ let us define $\lambda_i(\nu)$ by Eqs. (2.5). As mentioned earlier, $\lambda_i(\mu)$ are the dispersion functions for the homogeneous $c=c_i$ media. It is well known[16] that on the interval $1<\nu<\infty$, $\lambda_i(\nu)$ has exactly one zero, at the point ν_i, and for $c_1<c_2$ we have $1\leqslant\nu_1\leqslant\nu_2\leqslant\infty$. The proper ansatz for the discrete solution of Eqs. (1.1), (1.3), and (1.7) can now be given as

$$\psi(z,\mu) = A(\mu)e^{-z/\mu} + \int_1^{\nu_2}\frac{\nu B(\nu)}{\nu-\mu}e^{-z/\nu}d\nu, \tag{3.1a}$$

$$0 = A(\mu), \quad -1\leqslant\mu<0. \tag{3.1b}$$

This ansatz differs from that for the continuum modes, Eqs. (2.1), in that the range of integration in the integral term is now $[1,\nu_2]$, whereas in Eqs. (2.1) it is $[0,1]$. The integral term in Eq. (3.1a) is thus not a principal value, and it has the interpretation as a superposition of constant c discrete modes for $0<c\leqslant c_2$. This "superposition" type of discrete mode has not appeared in any prior eigenfunction analysis of the transport equation.

Introducing Eqs. (3.1) into the transport equation (2.3) and rearranging, we obtain

$$0 = \int_1^{\nu_2}[s\lambda_1(\nu)+z\lambda_2(\nu)]B(\nu)e^{-z/\nu}d\nu - (sc_1+zc_2)\left(\frac{1}{2}\int_{-1}^{1}A(\nu)e^{-z/\nu}d\nu\right). \tag{3.2}$$

Integrations by parts give

$$\int_1^{\nu_2}z\lambda_2(\nu)B(\nu)e^{-z/\nu}d\nu = \nu^2\lambda_2(\nu)B(\nu)e^{-z/\nu}\big|_1^{\nu_2} - \int_1^{\nu_2}\left[\frac{d}{d\nu}\nu^2\lambda_2(\nu)B(\nu)\right]e^{-z/\nu}d\nu \tag{3.3}$$

and

$$\int_0^1 zA(\nu)e^{-z/\nu}d\nu = \nu^2 A(\nu)e^{-z/\nu}\big|_0^1 - \int_0^1\left[\frac{d}{d\nu}\nu^2 A(\nu)\right]e^{-z/\nu}d\nu, \tag{3.4}$$

and thus Eq. (3.2) can be written

$$
\begin{aligned}
0 = &\int_1^{\nu_2}\left[s\lambda_1(\nu)B(\nu) - \frac{d}{d\nu}\nu^2\lambda_2(\nu)B(\nu)\right]e^{-z/\nu}d\nu \\
&-\frac{1}{2}\int_0^1\left[c_1 sA(\nu) - c_2\frac{d}{d\nu}\nu^2 A(\nu)\right]e^{-z/\nu}d\nu \\
&+\left[\lim_{\nu\to\nu_2}\nu^2\lambda_2(\nu)B(\nu)\right]e^{-z/\nu_2} \\
&-\left[\lim_{\nu\to 1}\nu^2\lambda_2(\nu)B(\nu) + \frac{c_2}{2}A(1)\right]e^{-z} \\
&+\frac{c_2}{2}\left[\lim_{\nu\to 0}\nu^2 A(\nu)e^{-z/\nu}\right].
\end{aligned}
\tag{3.5}
$$

Each of the five bracketed terms in Eq. (3.5) must vanish. Setting the first and second to zero gives

$$\nu^2\lambda_2(\nu)B(\nu) = \gamma e^{s\theta(\nu)}, \tag{3.6a}$$

$$\theta(\nu) = \int_{\nu_1}^{\nu}\frac{\lambda_1(t)}{t^2\lambda_2(t)}dt, \tag{3.6b}$$

and

$$A(\mu) = A(1)\frac{1}{\mu^2}\exp\left[\frac{c_1 s}{c_2}\left(1 - \frac{1}{\mu}\right)\right], \tag{3.7}$$

where γ and $A(1)$ are constants.

On the interval $1 < \nu < \nu_2$, $\lambda_2(\nu)$ is negative, while $\lambda_1(\nu)$ is negative for $1 < \nu < \nu_1$ and positive for $\nu_1 < \nu < \nu_2$, and therefore $\theta(\nu) \leqslant 0$. It is easily shown that θ is continuous and bounded except for $\nu = \nu_2$, where we have

$$\theta(\nu) = \alpha\ln(\nu_2 - \nu) + O(1), \quad \nu \cong \nu_2, \tag{3.8a}$$

$$\alpha = \frac{\lambda_1(\nu_2)}{\nu_2^2\lambda_2'(\nu_2)} > 0. \tag{3.8b}$$

Hence,

$$e^{s\theta(\nu)} = O(1)(\nu_2 - \nu)^{s\alpha}, \quad \nu \cong \nu_2 \tag{3.9}$$

and so by Eq. (3.6a), the third bracketed term in Eq. (3.5) automatically vanishes. It can easily be shown that the results in this paragraph do not hold if $c_1 > c_2$; hence, for this analysis, the assumption $c_1 < c_2$ is essential.

Setting the fourth bracketed term in Eq. (3.5) to zero gives, with Eq. (3.6a),

$$A(1) = -\frac{2}{c_2}\gamma e^{s\theta(1)}, \tag{3.10}$$

and hence

$$A(\mu) = -\frac{2}{c_2}\gamma\exp\left[s\theta(1) + s\frac{c_1}{c_2}\left(1 - \frac{1}{\mu}\right)\right], \tag{3.11}$$

which automatically causes the fifth bracketed term in Eq. (3.5) to vanish.

By inspection, $A(\mu)$ is a continuous, bounded function of μ for $0 \leqslant \mu \leqslant 1$, if we define $A(0) = 0$. By Eqs. (3.6a) and (3.9) and the properties of $\lambda_2(\nu)$, the same is true for $B(\nu)$ on the interval $1 \leqslant \nu \leqslant \nu_2$, except possibly at the point ν_2, where

$$B(\nu) = O(1)(\nu_2 - \nu)^{s\alpha - 1}, \quad \nu \cong \nu_2. \tag{3.12}$$

For $0 < s < \alpha^{-1}$, $B(\nu)$ has a singularity at $\nu = \nu_2$, but this singularity is integrable. For $\alpha^{-1} < s \leqslant \infty$, $B(\nu)$ has no singularity at $\nu = \nu_2$, and in fact $B(\nu_2) = 0$. The "turnover" value of s, $s = \alpha^{-1}$, is the only value for which $B(\nu_2)$ is finite and nonzero.

Thus for all $0 \leqslant s < \infty$, $B(\nu)$ is integrable, and we can choose γ so that $B(\nu)$ is a weight function:

$$1 = \int_1^{\nu_2} B(\nu)d\nu; \tag{3.13}$$

the proper value of γ is

$$\gamma = \left[\int_1^{\nu_2}\frac{e^{s\theta(\nu)}}{\nu^2\lambda_2(\nu)}d\nu\right]^{-1}. \tag{3.14}$$

Our results are summarized as follows.

The discrete solution is given by Eqs. (3.1), with $A(\mu)$ defined by Eq. (3.11), $B(\nu)$ and $\theta(\nu)$ by Eqs. (3.6), and γ by Eq. (3.14). The function $A(\nu)$ is continuous and bounded, and $B(\nu)$ is continuous and bounded for $s < \alpha^{-1} < \infty$, where α is defined by Eq. (3.8b). However, for $0 < s < \alpha^{-1}$, $B(\nu)$ has an integrable singularity at $\nu = \nu_2$.

From Eq. (1.7), it is apparent that as s tends to 0 or ∞, $c(z)$ tends to the constants c_2 or c_1, respectively. For these cases, we have verified analytically that $B(\nu)$ converges to $\delta(\nu - \nu_2)$ or $\delta(\nu - \nu_1)$, respectively, and that $A(\mu)$ converges to zero. Thus, the discrete mode of Eq. (3.1) converges, within a normalization factor, to the usual discrete mode for the homogeneous $c = c_2$ or $c = c_1$ medium.

We have computed $B(\nu)$ numerically for $c_1 = 0.9$, $c_2 = 0.99$, and various values of s. (For these values of c_1 and c_2, $\nu_1 = 1.903\ 20$, $\nu_2 = 5.796\ 73$, and the "turnover" value of s is $s = 1.298\ 63$.) The results, plotted in Fig. 1, numerically and visually substantiate the above comments concerning the behavior of $B(\nu)$ as $s \to 0$ or $s \to \infty$.

To conclude this section, we remark that we have not been able to directly extend the above analysis to either the case $c_1 > c_2$ or to the construction of a discrete solution growing as $z \to \infty$. For $c_1 > c_2$ a discrete solution decaying as $z \to \infty$ can, however, be constructed, in certain cases, but by a different analysis which we present in the next section.

IV. DISCRETE MODE FOR $C_1 > C_2$

In this section we shall construct a discrete solution of Eqs. (1.1), (1.3), and (1.7), for constants c_1, c_2, and s satisfying

$$s(c_1 - c_2) = \frac{m}{2}(c_2\nu_2)^2\left[\ln\left(\frac{\nu_2 - 1}{\nu_2 + 1}\right) + \left(\frac{2\nu_2}{\nu_2^2 - 1}\right)\right], \tag{4.1}$$

where m is any positive integer and, as in Sec. III, ν_2 is the positive root of $\lambda_2(\nu_2) = 0$. It is shown below that the above

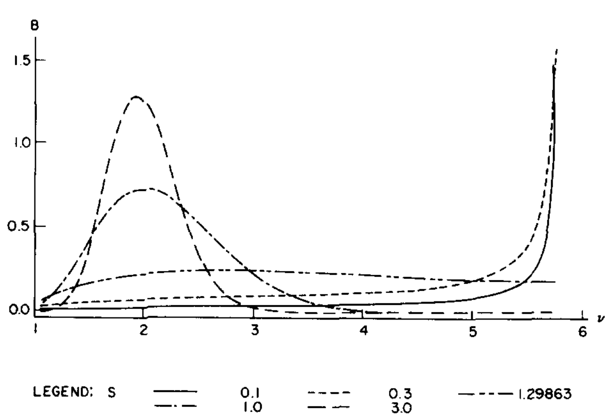


FIG. 1. $B(\nu)$ versus ν.

expression in brackets is positive; hence Eq. (4.1) implies $c_1 > c_2$. Thus if c_1 and c_2 are chosen, with $c_1 > c_2$, then Eq. (4.1) defines an infinite denumerable set of admissible values of s. We do not now know how to treat other values of s.

An ansatz which leads to the solution, and which is analogous to Eqs. (2.1) and (3.1), is

$$\psi(z,\mu) = \int_1^\infty \frac{\nu B(\nu)}{\nu - \mu} e^{-z/\nu} d\nu. \tag{4.2}$$

Introducing Eq. (4.2) into Eq. (2.3) and integrating by parts, as in the previous sections, we obtain

$$\frac{d}{d\nu} \nu^2 \lambda_2(\nu) B(\nu) + s\lambda_1(\nu) B(\nu) = 0, \tag{4.3a}$$

$$B(1) = B(\infty) = 0. \tag{4.3b}$$

It is easy to show that Eqs. (4.3) have no continuous solutions. However, if we formally introduce the distributional form into Eq. (4.3a),

$$B(\nu) = \sum_{n=0}^{m} b_n \delta^{(n)}(\nu - \nu_2) \tag{4.4}$$

(where $\delta^{(n)}$ is the nth derivative of the delta function and b_n are constants to be determined), then Eq. (4.3b) is satisfied and Eq. (4.3a) reduces to a homogeneous system of $m+1$ equations for the $m+1$ unknowns b_n. This system has a unique (up to a normalization) solution provided its determinant vanishes, and this condition is given exactly by Eq. (4.1).

The ansatz (4.2) and the form (4.4) for B imply that

$$\psi(z,\mu) = e^{-z/\nu_2} \sum_{n=0}^{m} \psi_n(\mu) z^n. \tag{4.5}$$

Since this equivalent form for ψ is algebraically much easier to manipulate, we shall use it as our basic ansatz.

Thus, we introduce Eq. (4.5) into Eq. (2.3) and rearrange to get

$$\begin{aligned} 0 = (s+z)\Bigg\{ &\sum_{n=0}^{m-1} (n+1)\mu\psi_{n+1}(\mu) z^n \\ &+ \sum_{n=0}^{m} \left[\left(1 - \frac{\mu}{\nu_2}\right)\psi_n(\mu) - \frac{c_2}{2}\int_{-1}^{1} \psi_n(\mu')d\mu'\right] z^n \Bigg\} \\ &+ s(c_2 - c_1) \sum_{n=0}^{m} \left(\frac{1}{2}\int_{-1}^{1} \psi_n(\mu')d\mu'\right) z^n. \end{aligned} \tag{4.6}$$

The following equation is an identity, which can be verified by comparing similar powers of z:

$$\begin{aligned} \sum_{n=0}^{m} \left[\frac{1}{2}\int_{-1}^{1} \psi_n(\mu')d\mu'\right] z^n &= \sum_{j=0}^{m} (-s)^j \left(\frac{1}{2}\int_{-1}^{1} \psi_j(\mu')d\mu'\right) \\ &+ (s+z) \sum_{n=0}^{m-1} \left[\sum_{j=1}^{m-n} (-s)^{j-1} \right. \\ &\left. \times \left(\frac{1}{2}\int_{-1}^{1} \psi_{n+j}(\mu')d\mu'\right)\right] z^n. \end{aligned} \tag{4.7}$$

Replacing the last term in Eq. (4.6) by Eq. (4.7) gives

$$\begin{aligned} 0 = (s+z)\Bigg\{ &\sum_{n=0}^{m} \left[\left(1 - \frac{\mu}{\nu_2}\right)\psi_n - \frac{c_2}{2}\int_{-1}^{1} \psi_n(\mu')d\mu'\right] z^n \\ &+ \sum_{n=0}^{m-1} \Big[(n+1)\mu\psi_{n+1}(\mu) - (c_2 - c_1) \\ &\times \sum_{j=1}^{m-n} (-s)^j \left(\frac{1}{2}\int_{-1}^{1} \psi_{n+j}(\mu')d\mu'\right)\Big] z^n \Bigg\} \\ &+ s(c_2 - c_1) \sum_{j=0}^{m} (-s)^j \left(\frac{1}{2}\int_{-1}^{1} \psi_j(\mu')d\mu'\right). \end{aligned} \tag{4.8}$$

By setting the last term, and the coefficient of each power of z in the braces equal to zero, we obtain the following set of equations:

$$\psi_m(\mu) - \phi_2(\mu)\int_{-1}^{1} \psi_m(\mu')d\mu' = 0, \tag{4.9}$$

$$\begin{aligned} \psi_n(\mu) - \phi_2(\mu)\int_{-1}^{1} \psi_n(\mu')d\mu' &= -(n+1)\frac{2}{c_2}\mu\phi_2(\mu)\,\psi_{n+1}(\mu) \\ &+ \phi_2(\mu)\left(1 - \frac{c_1}{c_2}\right) \sum_{j=1}^{m-n} (-s)^j \int_{-1}^{1} \psi_{n+j}(\mu')d\mu', \\ &\qquad 0 \leqslant n \leqslant m-1, \end{aligned} \tag{4.10}$$

$$0 = \sum_{j=0}^{m} (-s)^j \int_{-1}^{1} \psi_j(\mu')d\mu'. \tag{4.11}$$

Here we have defined

$$\phi_2(\mu) = \frac{c_2\nu_2}{2} \frac{1}{\nu_2 - \mu} \tag{4.12}$$

and we have

$$\int_{-1}^{1} \phi_2(\mu)d\mu = 1 - \lambda_2(\nu_2) = 1. \tag{4.13}$$

The solution of Eq. (4.9), up to a multiplicative constant, is

$$\psi_m(\mu) = \phi_2(\mu). \tag{4.14}$$

Equation (4.10) with $n = m-1$ now becomes

$$\begin{aligned} \psi_{m-1}(\mu) - \phi_2(\mu)\int_{-1}^{1} \psi_{m-1}(\mu')d\mu' \\ = -m\frac{2}{c_2}\mu\phi_2^2(\mu) - s\phi_2(\mu)\left(1 - \frac{c_1}{c_2}\right). \end{aligned} \tag{4.15}$$

Integrating over μ, the left side vanishes, and we obtain the condition

$$\begin{aligned} s(c_1 - c_2) &= 2m\int_{-1}^{1} \mu\phi_2^2(\mu)d\mu \\ &= 2mc_2^2\nu_2^3 \int_0^1 \left(\frac{\mu}{\nu_2^2 - \mu^2}\right)^2 d\mu. \end{aligned} \tag{4.16}$$

The right side of this equation is positive, and when the integrand is evaluated, the equation reduces to Eq. (4.1), which we have *a priori* assumed to hold.

Let us now denote the right side of Eq. (4.15) as $\eta_{m-1}(\mu)$. Then the general solution of Eq. (4.15) is

$$\psi_{m-1}(\mu) = a_{m-1}\phi_2(\mu) + \eta_{m-1}(\mu), \tag{4.17a}$$

and

$$\int_{-1}^{1} \psi_{m-1}(\mu) d\mu = a_{m-1}. \tag{4.17b}$$

To determine a_{m-1}, we introduce Eqs. (4.14) and (4.17) into the right side of Eq. (4.10) with $n = m - 2$ and (again) integrate over μ. The left side (again) vanishes, and we obtain a condition which explicitly determines a_{m-1}. With a_{m-1} thus determined, we denote the right side of the equation for ψ_{m-2} as η_{m-2}, and we obtain the general solution

$$\psi_{m-2}(\mu) = a_{m-2}\phi_2(\mu) + \eta_{m-2}(\mu), \tag{4.18a}$$

$$\int_{-1}^{1} \psi_{m-2}(\mu) d\mu = a_{m-2}. \tag{4.18b}$$

The remaining equations (4.10) can be solved explicitly and recursively in this manner, yielding the forms

$$\psi_n(\mu) = a_n \phi_2(\mu) + \eta_n(\mu), \quad 0 \leqslant n \leqslant m, \tag{4.19a}$$

$$\int_{-1}^{1} \psi_n(\mu) d\mu = a_n, \quad a_m = 1, \quad \eta_m = 0, \tag{4.19b}$$

and where every constant a_n is determined except for the last one, a_0. However, all of Eqs. (4.9) and (4.10) are satisfied, and the one remaining equation to be solved, Eq. (4.11), reduces to

$$0 = a_0 + \sum_{j=1}^{m} (-s)^j a_j. \tag{4.20}$$

This establishes the value of a_0, and thus the solution is now fully and uniquely determined.

To summarize, if m is a positive integer and c_1, c_2, s, and m satisfy Eq. (4.1), then a solution of the transport equation can be constructed having the form of Eqs. (4.5) and (4.19). (The special case $m = 1$ was treated earlier in Ref. 3.)

The essential ingredients in the above analysis are the constraint (4.1) and the ansatz (4.5). To satisfy Eq. (1.3) we have assumed $\nu_2 > 0$, but if one desires discrete solutions which grow as $z \to \infty$, then one can take ν_2 to be the negative (rather than the positive) root of the equation $\lambda_2(\nu_2) = 0$. The right side of Eq. (4.1) is an odd function of ν_2, and thus if ν_2 changes sign from plus to minus, then $c_1 - c_2$ must also change sign from plus to minus. The above analysis now follows directly. To summarize, we obtain for $c_1 < c_2$ and the infinite discrete set of values of s defined by Eq. (4.1), a discrete solution which grows as $z \to \infty$.

V. NUMERICAL RESULTS

We consider the solution of the albedo problem for Eqs. (1.1)–(1.3), where $c(z)$ is given by Eq. (1.7) with $c_1 < c_2$. If the eigenfunctions described in Secs. II and III are half-range complete, we may write the solution of this transport problem as

$$\psi(z, \mu) = A_+ \psi_+(z, \mu) + \int_0^1 A(\nu) \psi_\nu(z, \mu) d\nu, \tag{5.1}$$

where $\psi_\nu(z, \mu)$ are the continuum eigenfunctions obtained in Sec. II, $\psi_+(z, \mu)$ is the discrete eigenfunction derived in Sec. III, and A_+ and $A(\nu)$ are expansion coefficients. A complete and conceptually straightforward solution method for this problem would be to construct a half-range completeness proof for the eigenfunctions, thus justifying the writing of Eq. (5.1), coupled with an analytic solution technique for the singular integral equation which yields A_+ and $A(\nu)$. At this time, however, we have been unable to find a completeness proof and solution technique.

Instead, we derive from Eq. (5.1) a singular integral equation for the outgoing angular flux $\psi(0, -\mu)$, $\mu > 0$, and use the F_N method described by Siewert *et al.*[2,7,13,15] to obtain a solution for the outgoing flux. This procedure starts with Eq. (5.1) and hence presumes completeness of the eigenfunctions. If the final numerical results are correct (we compare our results with an independent numerical solution of the transport equation), we then have established numerical evidence for half-range completeness of the eigenfunctions. To obtain the above-mentioned integral equation, we need the relationship

$$\int_{-1}^{1} \mu \psi_1(z, \mu) \psi_2(z, -\mu) d\mu = 0, \tag{5.2}$$

where ψ_1 and ψ_2 are any two solutions of the transport equation (1.1) which vanish at $z = +\infty$. Equation (5.2) is easily derived by writing Eq. (1.1) for $\psi_1(z, \mu)$ as well as for $\psi_2(z, -\mu)$, cross multiplying, integrating over μ, and substracting the two results. Now, we multiply Eq. (5.1) successively by $\mu\psi_+(z, -\mu)$ and $\mu\psi_\nu(z, -\mu)$, integrate over μ, and evaluate the results at $z = 0$. Since Eq. (5.2) implies that the right-hand side of Eq. (5.1) vanishes upon integration, we obtain the two equations

$$\int_{-1}^{1} \mu \phi_+(-\mu) \psi(0, \mu) d\mu = 0, \tag{5.3}$$

$$\int_{-1}^{1} \mu \phi_\nu(-\mu) \psi(0, \mu) d\mu = 0, \quad 0 \leqslant \nu \leqslant 1, \tag{5.4}$$

where we have defined

$$\phi_+(\mu) = \psi_+(0, \mu), \quad \phi_\nu(\mu) = \psi_\nu(0, \mu). \tag{5.5}$$

Using the boundary condition, Eq. (1.2), in Eqs. (5.3) and (5.4) for $\psi(0, \mu)$ for positive values of μ, we obtain

$$\int_0^1 \mu \phi_+(\mu) \psi(0, -\mu) d\mu = S_+, \tag{5.6}$$

$$\int_0^1 \mu \phi_\nu(\mu) \psi(0, -\mu) d\mu = S(\nu), \quad 0 \leqslant \nu \leqslant 1, \tag{5.7}$$

with the "source" terms given by

$$S_+ = \int_0^1 \mu \phi_+(-\mu) g(\mu) d\mu, \tag{5.8}$$

$$S(\nu) = \int_0^1 \mu \phi_\nu(-\mu) g(\mu) d\mu, \quad 0 \leqslant \nu \leqslant 1. \tag{5.9}$$

For any incident distribution $g(\mu)$, these sources may be evaluated and Eq. (5.7) then represents a singular integral equation to be solved, subject to the constraint given by Eq. (5.6), for the outgoing angular distribution $\psi(0, -\mu)$, $\mu > 0$.

The F_N solution method represents $\psi(0, -\mu)$ as

$$\psi(0, -\mu) = \sum_{n=1}^{N} b_n \mu^{n-1}, \quad \mu > 0, \tag{5.10}$$

where the expansion coefficients b_n are to be determined. If we define

$$D_n = \int_0^1 \mu^n \phi_+(\mu) d\mu, \tag{5.11}$$

TABLE I. The albedo for an isotropic incident flux with $c_1 = 0.9$ and $c_2 = 0.99$.

$N\backslash s$	0.1	0.3	1.0	3.0	10.0	20.0
2	0.716 90	0.651 97	0.571 56	0.519 35	0.492 09	0.485 21
3	0.716 87	0.651 92	0.571 62	0.519 37	0.492 13	0.485 30
4	0.716 89	0.651 92	0.571 64	0.519 38	0.492 14	0.485 30
5	0.716 90	0.651 92	0.571 64	0.519 38	0.492 14	0.485 30
ANISN	0.718	0.653	0.572	0.520	0.493	0.486

$$C_n(\nu) = \int_0^1 \mu^n \phi_\nu(\mu) d\mu, \quad 0<\nu<1, \tag{5.12}$$

the use of Eq. (5.10) in Eqs. (5.6) and (5.7) gives

$$\sum_{n=1}^{N} D_n b_n = S_+, \tag{5.13}$$

$$\sum_{n=1}^{N} C_n(\nu) b_n = S(\nu), \quad 0<\nu<1. \tag{5.14}$$

To obtain the required N equations to determine the b_n, the F_N philosophy is to evaluate Eq. (5.14) at $N-1$ points, say ν_m, $m = 1,2,...,N-1$. Equation (5.13) provides the Nth equation. We used the simple choice of equally spaced collocation points, that is,

$$\nu_m = \frac{2m-1}{2N-2}, \quad m = 1,2,...,N-1. \tag{5.15}$$

Thus we obtain the $N \times N$ matrix problem

$$\sum_{n=1}^{N} C_n(\nu_m) b_n = S(\nu_m), \quad m = 1,2,...,N, \tag{5.16}$$

where for compactness of notation we have written

$$C_n(\nu_N) \equiv D_n, \quad S(\nu_N) \equiv S_+. \tag{5.17}$$

Once the b_n are obtained by applying any standard linear equation solver to Eq. (5.16), the albedo A, defined as the ratio of outgoing to incoming currents, is given by

$$A = \int_0^1 \mu\psi(0, -\mu) d\mu \Big/ \int_0^1 \mu g(\mu) d\mu = \sum_{n=1}^{N} \frac{b_n}{n+1}, \tag{5.18}$$

where in writing the last equality, we have assumed that the incoming current is normalized to unity.

Introducing the eigenfunctions explicitly into Eqs. (5.8) and (5.9), we find that the sources can be written

$$S_+ = \int_1^{\nu_2} \nu B(\nu) \int_0^1 \frac{\mu}{\nu+\mu} g(\mu)\, d\mu\, d\nu, \tag{5.19}$$

$$S(\nu) = \frac{c_2 \nu}{2} \int_0^1 \frac{\mu}{\nu+\mu} g(\mu) d\mu, \quad 0<\nu<1. \tag{5.20}$$

For the special case of isotropic incidence, $g(\mu) = 2$, Eqs. (5.19) and (5.20) reduce to

$$S_+ = 2\int_1^{\nu_2} \nu B(\nu)\left[1 - \nu \ln\left(\frac{1+\nu}{\nu}\right)\right] d\nu, \tag{5.21}$$

$$S(\nu) = c_2 \nu\left[1 - \nu \ln\left(\frac{1+\nu}{\nu}\right)\right], \quad 0<\nu<1. \tag{5.22}$$

Similarly, the matrix elements are given explicitly as

$$D_n = -\frac{2}{c_2} E_n(c_1 s/c_2) e^{s[\theta(1) + (c_1/c_2)]} + \int_1^{\nu_2} \nu B(\nu) K_n(\nu) d\nu, \quad n = 1,2,...,N \tag{5.23}$$

$$C_n(\nu) = s\left(1 - \frac{c_1}{c_2}\right)\nu^{n-1} E_n(c_1 s/c_2 \nu) e^{c_1 s/c_2 \nu} + \nu^n \lambda_2(\nu) + \frac{c_2 \nu}{2} K_n(\nu), \quad 0<\nu<1, \quad n = 1,2,...,N. \tag{5.24}$$

Here $E_n(z)$ is the usual nth order exponential integral,[16] and $K_n(\nu)$ is defined as

$$K_n(\nu) = \int_0^1 \frac{\mu^n}{\nu-\mu} d\mu, \tag{5.25}$$

which can be computed recursively from

$$K_n(\nu) = \nu K_{n-1}(\nu) - 1/n, \tag{5.26}$$

with

$$K_0(\nu) = \ln\left(\frac{\nu}{1-\nu}\right), \quad 0\leqslant\nu\leqslant 1. \tag{5.27}$$

As an application of the foregoing numerical method, we have obtained explicit results for $g(\mu) = 2$ (isotropic incidence) and various values of c_1, c_2, and s. Typical results from our computations are shown in Table I, which corresponds to $c_1 = 0.9$ and $c_2 = 0.99$. For these parameters, the turnover value of s is 1.298 63 and Table I gives results for s both smaller and larger than this turnover value. These results show that for this problem the F_N solution method converges extremely well as a function of N, and apparently converges to the correct answer as generated independently by the ANISN computer code.[17] (ANISN generates a direct numerical solution of the transport equation by the standard, well-tested method of discrete ordinates.)

The excellent agreement between F_N and ANISN calculations gives numerical evidence that the eigenfunctions constructed above are indeed complete on the half-range. In addition, our results show that the F_N method is a very effective tool for obtaining numerical albedo results for the class of problems considered in this article.

[1] B. J. Martin, "A New Singular Integral Equation in Radiative Transfer," SIAM J. Appl. Math. 20, 703 (1971).
[2] T. W. Mullikin and C. E. Siewert, "Radiative Transfer in Inhomogeneous Atmospheres," Ann. Nucl. Energy 7, 205 (1980).
[3] G. C. Pomraning and E. W. Larsen, "Elementary Solutions of the Linear Transport Equation for Continuously Varying Spatial Media," J. Math. Phys. 21, 1603 (1980).

[4]E. W. Larsen, G. C. Pomraning, and V. C. Badham, "On the Singular Eigenfunctions for Linear Transport in an Exponential Atmosphere," J. Math. Phys. **21**, 2448 (1980).
[5]T. W. Mullikin, "Some Singular Integral Equations in Linear Transport Theory," Sandia National Laboratories Report SAND 80-1069 (1980) (unpublished).
[6]E. W. Larsen and T. W. Mullikin, "Linear Transport in an Exponential Atmosphere," J. Math. Phys. **22**, 856 (1981).
[7]R. D. M. Garcia and C. E. Siewert, "Radiative Transfer in Inhomogeneous Atmospheres—Numerical Results," J. Quant. Spectrosc. Radiat. Transfer **25**, 277 (1981).
[8]C. T. Kelley and T. W. Mullikin, "Collocation Methods for Some Singular Integral Equations in Linear Transport Theory," J. Int. Eq. (to appear).
[9]G. C. Pomraning, "The Criticality Problem for an Exponential Atmosphere," J. Math. Phys. **22**, 1623 (1981).
[10]E. W. Larsen, "Linear Transport in Nonhomogeneous Media. I," J. Math. Phys. **22**, 1463 (1981).
[11]C. T. Kelley, "Approximation of Solutions of Some Quadratic Integral Equations in Transport Theory" (submitted for publication).
[12]E. W. Larsen, "Singular Eigenfunction Analysis of Transport Problems with Spatially Dependent Cross Sections—An Overview," Prog. Nucl. Energy (to appear).
[13]R. D. M. Garcia and C. E. Siewert, "Radiative Transfer in Finite, Inhomogeneous, Plane-Parallel Atmospheres," J. Quant. Spectrosc. Radiat. Transfer **27**, 141 (1982)
[14]G. C. Pomraning, "Neutral Transport through Plasmas," Prog. Nucl. Energy (to appear).
[15]R. D. M. Garcia, G. C. Pomraning, and C. E. Siewert, "A Kinetic Theory for the Transport of Neutral Hydrogen Atoms in a Hydrogen Plasma," Plasma Phys. (to appear).
[16]K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).
[17]W. W. Engle, Jr., The User's Manual for ANISN: A One-Dimensional Discrete Ordinates Code with Anisotropic Scattering, K-1693, Oak Ridge Gaseous Diffusion Plant (1967). Available from Radiation Shielding Information Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, as CCC-82/ANISN (March, 1967).

Unrenormalized Schwinger–Dyson equations and dynamical mass generation

Tetz Yoshimura
Department of Physics, University of Kaiserslautern, D-6750 Kaiserslautern, West Germany

(Received 24 February 1981; accepted for publication 6 November 1981)

In dynamical mass generation of fermion fields by fermion-spin-1-boson interactions, anomalous magnetic moment-like terms in three-point functions play an essential role, so that the Schwinger–Dyson equations do not admit a trivial solution, i.e., a solution without dynamically generated mass. Applicability of Altman's and Lika's versions of nonlinear operator theory to unrenormalized Schwinger–Dyson equations is discussed and algorithms for construction of approximate solutions are proposed.

PACS numbers: 11.10.Lm

1. INTRODUCTION

In early 1960's, interest in the problem of dynamically generated masses (DGM's) was revived in various contexts[1–15] after a dormancy due to the success of the renormalization theory. One aspect of the problem is concerned with DGM due to a breakdown of chiral, gauge, conformal and other invariances, and another is concerned with the existence and uniqueness of solution of nonlinear and/or singular integral equations. Here we discuss mainly the latter aspect of the problem and try to apply the modern nonlinear operator theory[16–21] to the equations arising in the problem of the dynamical mass generation (DMG).

Though we do not discuss much about the group theoretic aspect of DMG in this note, Poincaré group and spinor structure cannot be ignored. As is seen in Eq. (11), γ matrices play essential roles in the DMG of spin 1/2 fields. If the vertex part contains a magnetic moment-like term, the equation for the mass term of the self energy is coupled to the equation of the $\gamma{\cdot}p$ part, and does not have a trivial solution, even if the self energy of the vector field is ignored. In this case one cannot use the bare vertex even for the γ_μ part of the three-point function, so one must begin with form factors with nice asymptotic behaviors, because we cannot subtract divergent terms without throwing away the dynamically-generated mass we want.

In section 2, we formulate the physical equations. As we are mainly concerned with the operator theoretical aspect, we take the $U_V(1)$ and $U_A(1)$ models as the simplest examples in order to see the essential features of nonlinear integral equations in the theory of DMG. In Sec. 3 we interpret the equations proposed in Sec. 2 in terms of operator theory so that the operator theory will be applicable. In Sec. 4 we propose Lipschitz approximation to the nonlinear operator arising in the unrenormalized Schwinger–Dyson equation proposed in Sec. 2. Existence of a solution to systems with equations with constraint is also discussed. Section 4 contains concluding remarks. Relevant mathematical definitions and theorems are presented in the Appendix.

2. SCHWINGER–DYSON EQUATIONS TO BE CONSIDERED

In order to see the essential features of integral equations for the self-energy parts relevant to the problem of DMG, we consider the $U_V(1)$ and $U_A(1)$ models as the simplest examples. We begin with the following Lagrangian:

$$\mathscr{L} = \mathscr{L}_f + \mathscr{L}_g + \mathscr{L}_{\rm int}, \tag{1}$$

$$\mathscr{L}_f = \bar{\psi} i\gamma{\cdot}\partial\psi, \tag{2}$$

$$\mathscr{L}_g = \tfrac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial_\mu A_\nu - \partial_\nu A_\mu) + \text{gauge fixing term}, \tag{3}$$

$$\mathscr{L}_{\rm int} = -ig\bar{\psi}\hat{\gamma}_5\gamma_\mu\psi A_\mu, \tag{4}$$

As we are going to deal with unrenormalized Schwinger–Dyson equations, counterterms are not included in the above Lagrangian, where $\hat{\gamma}_5 = \gamma_5$ or 1.

Then the equations for the self-energy part reads

$$\Sigma(p) = \Xi(\Sigma;p) = :g\hat{\gamma}_5\gamma_\mu \int d^4k\, \{\gamma{\cdot}k + \Sigma(k)\}^{-1} \times F_\nu(k,p;-p-k)D_{\mu\nu}(\Sigma,p), \tag{5}$$

where

$$D_{\mu\nu}(\Sigma,p) = :\left(g_{\mu\nu} - \frac{p_\mu p_\nu}{p^2}\right) \times \left\{p^2 + \frac{1}{3}\pi^\lambda_\lambda(\Sigma,p) - \frac{1}{3}\frac{p_\lambda p_\kappa}{p^2}\pi_{\lambda\kappa}(\Sigma,p)\right\}^{-1} + \frac{p_\mu p_\nu}{p^2}\left\{\xi p^2 + \frac{p_\lambda p_\kappa}{p^2}\pi_{\lambda\kappa}(\Sigma;p)\right\}^{-1}, \tag{6}$$

$$\pi_{\mu\nu}(\Sigma,p) = :-g\int d^4k\, {\rm Tr}\,[\hat{\gamma}_5\gamma_\mu\{\gamma{\cdot}k + \Sigma(k)\}^{-1} \times F_\nu(k,-k-p;p)\{\gamma{\cdot}(p+k) + \Sigma(p+k)\}^{-1}]. \tag{7}$$

The only integral that may diverge is $\pi_{\mu\nu}(\Sigma,0)$, but it does not affect the convergence of the integral (5). As for the domain $D(\Xi)$ one can choose a suitable bounded set in the direct sum of Banach spaces $\mathfrak{B}_1 \supset \Sigma_1$ and $\mathfrak{B}_2 \supset \Sigma_2'$ with norms defined by (23) in Sec. 3. Σ_i may have "singularities" $\sim(k^2-\kappa)^\delta$, $\delta > 0$. For convenience let us set the gauge parameter $\xi = 0$. Then the longitudinal part of $D_{\mu\nu}(p)$ does not have a pole unless $\pi_{\lambda\kappa}(\Sigma,p) = 0$ for some p with $p^2 \neq 0$. If one ignores the form factor F and the momentum dependence of π, one gets the following equation for the mass part of the self-energy function:

$$\Sigma(q^2) = \lambda \int d^4k \frac{\Sigma(k)}{[k^2 + \Sigma(k^2)^2]\,[(k-q)^2 - \mu^2(\Sigma)]}, \tag{8}$$

where $\mu^2(\Sigma)$ is a functional of Σ.

Obviously, this equation has the trivial solution $\Sigma(p)\equiv 0$. Since 1964, the great majority of the papers[3–9] on DMG in models with boson-fermion interactions have been concerned with finding nontrivial solutions of this "homogeneous" equation. But the situation is not so simple if the form factors are taken into account. Suppose the vertex part (amputated three-point function) F_ν contains an "anomalous magnetic moment-like term" $\hat\gamma_5\gamma_{[\nu}\gamma_{\rho]}(p+q)_\rho f_2$

$$F_\nu(p,q;-p-q)=\hat\gamma_5\gamma_\nu f_1(p^2,q^2,p\cdot q)+\hat\gamma_5\gamma_{[\nu}\gamma_{\rho]}(p+q)_\rho f_2(p^2,q^2,p\cdot q)+\cdots. \tag{9}$$

Then even in the approximation with the "Feynman gauge" free propagator

$$D^0_{\mu\nu}(p)=ig_{\mu\nu}(p^2-\mu^2-i\epsilon)^{-1} \tag{10}$$

the equation for the $\gamma\cdot p$ part and the scalar part of the self-energy part remain coupled

$$\Sigma_1(p^2)=g\int d^4k\,\{-2\gamma\cdot k\gamma\cdot p(p^2)^{-1}(1+\Sigma_1(k^2))f_1(Q)+3\tau(1-\gamma\cdot k\gamma\cdot p/p^2)$$
$$\times\Sigma_2(k^2)f_2(Q)+\cdots\}\,\{(p-k)^2-\mu^2\}^{-1}\{k^2(1+\Sigma_1(k^2))^2-\Sigma_2(k^2)^2\}^{-1}, \tag{11a}$$

$$\Sigma_2(p^2)=g\int d^4k\,\{(-\gamma\cdot k\gamma\cdot p-2k\cdot p+3k^2)(1+\Sigma_1(k^2))f_2(Q)$$
$$+\Sigma_2(k^2)f_1(Q)+\cdots\}\,\{(p-k)^2-\mu^2\}^{-1}\{k^2(1+\Sigma_1(k^2))^2-\Sigma_2(k^2)^2\}^{-1}, \tag{11b}$$

where $\Sigma(p)=\gamma\cdot p\Sigma_1(p^2)+\Sigma_2(p^2)$, $f_i(Q)=f_i(k^2,p^2,k\cdot p)$, and $\tau=-1$ if $\hat\gamma_5=\gamma_5$ and $\tau=1$ if $\hat\gamma_5=1$. It can be easily seen that the system (11) does not have a solution with $\Sigma_2\equiv 0$ unless the form factors satisfy some complicated constraints involving Σ_1. In other words, the anomalous magnetic momentum-like term in the vertex part induces a DMG. Moreover, it is easily seen that $\Sigma_2(0)\neq 0$ unless the form factors satisfy further constraints.

Some comments are in order. Though the system (11) does not admit the trivial solution $\Sigma_2\equiv 0$, it is still of Urysohn type. However, if the vacuum polarization π is taken into account [see Eq. (5)], the mapping Ξ is no longer Urysohn but noncompact, so that many theories developed for compact mapping are not applicable. If one sets $\xi\neq 0$, the longitudinal part of $D_{\mu\nu}(p)$ has a pole at a different value of p^2 from that of the transversal part. This may cause additional difficulties in the handling of the equations.

An interesting model of the vertex part with a nice asymptotic behavior is

$$F_\mu(p,q;-p-q)\cong c_1\hat\gamma_5\gamma_\mu\frac{|p^2|^\alpha|q^2|^\alpha|(p+q)^2|^\beta}{\{|p^2|+|q^2|+|(p+q)^2|+m^2\}^{2\alpha+\beta}}$$
$$+c_2\hat\gamma_5\gamma_{[\mu}\gamma_{\nu]}(p+q)_\nu\frac{|p^2|^{\alpha'}|q^2|^{\alpha'}|(p+q)^2|^{\beta'}}{\{|p^2|+|q^2|+|(p+q)^2|+m^2\}^{2\alpha'+\beta'+1/2}},\quad \alpha,\alpha',\beta'>0,\beta>1, \tag{12}$$

where m is a parameter with the dimension of mass. This form factor reproduces self-consistently $\Sigma(p)\sim\gamma\cdot p$, $p\to\infty$, up to at most a logarithmic factor. If one begins with a more general vertex function

$$F'_\mu(p,q,-p-q)\cong c'_1\hat\gamma_5\gamma_\mu\frac{m^{2\gamma}|p^2|^\alpha|q^2|^\alpha|(p+q)^2|^\beta}{\{|p^2|+|q^2|+|(p+q)^2|+m^2\}^{2\alpha+\beta+\gamma}}$$
$$+c'_2\gamma_5\gamma_{[\mu}\gamma_{\nu]}(p+q)_\nu\frac{m^{2\gamma}|p^2|^{\alpha'}|q^2|^{\alpha'}|(p+q)^2|^{\beta'}}{\{|p^2|+|q^2|+|(p+q)^2|+m^2\}^{2\alpha'+\beta'+\gamma+\frac12}},\quad \alpha,\alpha',\beta',\gamma>0,\beta>1, \tag{13}$$

and indulges in naive power counting, one gets $\Sigma(p)\sim\gamma\cdot p(p^2/m^2)^{-\gamma}$. But let us curtail further discussion of this topic and return to the main issue.

Another question to be considered is whether there are any different features in DMG if the Lagrangian is not P-invariant. As an example of systems with parity-violating interactions, let us consider the system with the following Lagrangian:

$$\mathscr{L}=i\bar\psi\gamma\cdot\partial\psi-\tfrac14F_{\mu\nu}F^{\mu\nu}-\tfrac14G_{\mu\nu}G^{\mu\nu}-if\bar\psi\gamma_\mu\psi A^\mu-ig\bar\psi(1+\beta\gamma_5)\gamma_\mu\psi B_\mu, \tag{14}$$

$$F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu,\quad G_{\mu\nu}=\partial_\mu B_\nu-\partial_\nu B_\mu. \tag{15}$$

Then we get the following Schwinger–Dyson equations:

$$\Sigma(p)=\Omega(\Sigma,\Gamma,p)$$
$$=f\int d^4k\,\gamma_\mu\frac{1}{\gamma\cdot k-\Sigma(k)}\Gamma^A_\nu(k,p;-p-k)D^A_{\mu\nu}(p+k)$$
$$+f\int d^4k\,\gamma_\mu\frac{1}{\gamma\cdot k-\Sigma(k)}\Gamma^B_\nu(k,p;-p-k)D^{AB}_{\mu\nu}(p+k)$$
$$+g\int d^4k\,(1+\beta\gamma_5)\gamma_\mu\frac{1}{\gamma\cdot k-\Sigma(k)}\Gamma^B_\nu(k,p;-p-k)D^B_{\mu\nu}(p+k)$$
$$+g\int d^4k\,(1+\beta\gamma_5)\gamma_\mu\frac{1}{\gamma\cdot k-\Sigma(k)}\Gamma^A_\nu(k,p;-p-k)D^{BA}_{\mu\nu}(p+k), \tag{16}$$

$$\{k^2 g_{\lambda\mu} - k_\lambda k_\mu - \pi^A_{\lambda\mu}(k)\} D^A_{\mu\nu}(k) - \pi^{AB}_{\lambda\mu}(k) D^{BA}_{\mu\nu}(k) = i g_{\lambda\nu}, \tag{17a}$$

$$\{k^2 g_{\lambda\mu} - k_\lambda k_\mu - \pi^A_{\lambda\mu}(k)\} D^{AB}_{\mu\nu}(k) - \pi^{AB}_{\lambda\mu}(k) D^{B}_{\mu\nu}(k) = 0, \tag{17b}$$

$$\{k^2 g_{\lambda\mu} - k_\lambda k_\mu - \pi^B_{\lambda\mu}(k)\} D^B_{\mu\nu}(k) - \pi^{BA}_{\lambda\mu}(k) D^{AB}_{\mu\nu}(k) = i g_{\lambda\nu}, \tag{17c}$$

$$\{k^2 g_{\lambda\mu} - k_\lambda k_\mu - \pi^B_{\lambda\mu}(k)\} D^{BA}_{\mu\nu}(k) - \pi^{BA}_{\lambda\mu}(k) D^{A}_{\mu\nu}(k) = 0, \tag{17d}$$

$$\pi^A_{\lambda\mu}(k) = f \int d^4q \,\mathrm{Tr}\left[\gamma_\lambda \frac{1}{\gamma\cdot(q-\frac{1}{2}k) - \Sigma(q-\frac{1}{2}k)} \Gamma^A_\mu(q-\tfrac{1}{2}k, -q-\tfrac{1}{2}k; k) \frac{1}{\gamma\cdot(q+\frac{1}{2}k) - \Sigma(q+\frac{1}{2}k)}\right], \tag{18a}$$

$$\begin{aligned}\pi^{AB}_{\lambda\mu}(k) &= f \int d^4q \,\mathrm{Tr}\left[\gamma_\lambda \frac{1}{\gamma\cdot(q-\frac{1}{2}k) - \Sigma(q-\frac{1}{2}k)} \Gamma^B_\mu(q-\tfrac{1}{2}k, -q-\tfrac{1}{2}k; k) \frac{1}{\gamma\cdot(q+\frac{1}{2}k) - \Sigma(q+\frac{1}{2}k)}\right] \\ &= g \int d^4q \,\mathrm{Tr}\left[(1+\beta\gamma_5)\gamma_\mu \frac{1}{\gamma\cdot(q-\frac{1}{2}k) - \Sigma(q-\frac{1}{2}k)} \Gamma^A_\mu(q-\tfrac{1}{2}k, -q-\tfrac{1}{2}k; k) \frac{1}{\gamma\cdot(q+\frac{1}{2}k) - \Sigma(q+\frac{1}{2}k)}\right],\end{aligned} \tag{18b}$$

$$\pi^B_{\lambda\mu}(k) = g \int d^4q \,\mathrm{Tr}\left[(1+\beta\gamma_5)\gamma_\lambda \frac{1}{\gamma\cdot(q-\frac{1}{2}k) - \Sigma(q-\frac{1}{2}k)} [\Gamma^B_\mu(q-\tfrac{1}{2}k, -q-\tfrac{1}{2}k; k)] \frac{1}{\gamma\cdot(q+\frac{1}{2}k) + \Sigma(q+\frac{1}{2}k)}\right], \tag{18c}$$

where Γ^A and Γ^B are the vertex part. Now we assume that Σ, Γ^A, and Γ^B have the following forms:

$$\Sigma(p) = \gamma\cdot p\sigma_1(p^2) + \sigma_2(p^2) + \gamma_5\gamma\cdot p\sigma_3(p^2) + \gamma_5\sigma_4(p^2), \tag{19}$$

$$\Gamma^A_\mu(p,q;-p-q) = \gamma_\mu f_1(p^2,q^2,p\cdot q) + \gamma_{[\mu}\gamma_{\nu]} f_2(p^2,q^2,p\cdot q), \tag{20}$$

$$\Gamma^B_\mu(p,q;-p-q) = \gamma_\mu h_1(p^2,q^2,p\cdot q) + \gamma_{[\mu}\gamma_{\nu]} h_2(p^2,q^2,p\cdot q) + \gamma_5\gamma_\mu h_3(p^2,q^2,p\cdot q) + \gamma_5\gamma_{[\mu}\gamma_{\nu]} h_4(p^2,q^2,p\cdot q), \tag{21}$$

with σ's, f's, and h's with appropriate asymptotic behaviors. Of course, (20) and (21) are not the most general forms of Γ's but we take them as reasonable models of the vertex parts. If one begins with an arbitrary set of σ's, f's, and h's, one finds that $\Omega[\Sigma,\Gamma,p]$ generates the structure $\sim\gamma_5$ which violates the CP-invariance. For such a structure to be absent in Σ the input model vertex functions must satisfy certain constraints. In other words one cannot choose all the model functions f's and h's independently. For systems with Lagrangian containing boson–boson interaction together with fermion–boson interaction, for example a system with

$$\mathscr{L}_{\mathrm{int}} = -ig\bar{\psi}\gamma_5\psi\varphi - \lambda\varphi^4,$$

or non-abelian gauge theory, one cannot express boson self energies and boson propagators in terms of fermion propagators, so that one has to deal with a set of unknown functions $\{\Sigma_1, \Sigma_2, \pi\}$.

3. OPERATOR THEORETIC CONSIDERATIONS OF THE RELEVANT EQUATIONS

In this section we reformulate the problem of the unrenormalized Schwinger–Dyson equations and DMG in the language of modern, nonlinear operator theory (NLOT).

Let us write a nonlinear integral equation abstractly

$$\sigma - \Upsilon[\sigma] = :\Phi[\sigma] = 0. \tag{22}$$

As we are thinking of Eq. (5) and Eq. (11), we do not assume that the map Υ is Fréchet differentiable, because these equations contain the unknown function in the denominators. Though the Euclidian versions of these maps are Fréchet differentiable, we do not discuss the Euclidean version because one cannot Wick-rotate an approximate or model vertex part and approximate solutions. So we must apply theories of nonlinear maps without Fréchet differentiability developed by Altman[17] and Lika.[18]

As $D_{\mu\nu}$ can be expressed in terms of $\Sigma_{1,2}$ Eq. (5) can be rewritten as a system of equations for two unknown functions Σ_1 and Σ_2. So, Eq. (5) as well as Eq. (11) can be dealt with in the direct sum of the Banach space of candidates for Σ_1 and those for $\Sigma_2'(p^2) = :(|p^2| + \eta)^{-1}\Sigma_2(p^2)$, $\eta > 0$. As $\Sigma_2(p^2)$ may increase $\sim |p^2|^{1/2}$ as $|p^2| \to \infty$, we take Σ_2' as a normable function instead, and define the norms as follows:

$$\|\Sigma\| = \|\Sigma_1\| + \|\Sigma_2'\|, \tag{23a}$$

$$\|\Sigma_1\| = \sup|\Sigma_1(p^2)|, \tag{23b}$$

$$\|\Sigma_2'\| = \sup|\Sigma_2'(p^2)| + \sup\big||p^2|^{1/2}\Sigma_2'(p^2)\big|. \tag{23c}$$

Because of the presence of Σ in the denominators of Eq. (5) and Eq. (11), the domains of the maps under consideration are bounded, in other words, not vector spaces, so that global theories are not applicable. On the other hand, in terms of the propagator

$$G(p) = [\gamma\cdot p - \Sigma(p)]^{-1} \tag{24}$$

the Schwinger–Dyson equation is "polynomially" nonlinear, but the unknown function $G(p)$ has a pole at an unknown place so that we cannot construct a Banach space that contains $G(p)$.

As for the vertex part $F(p,q;p-q)$, we assume the following asymptotic bound so that the integrals in Eqs. (5) and (7) exist for finite values of p^2 if $D_{\mu\nu}(p) \underset{p\to\infty}{\sim} 1/p^2$:

$$G(p) \underset{p\to\infty}{\sim} \gamma\cdot p/p^2$$

$$\begin{aligned}F_\mu(p,q;-p-q) &= \hat{\gamma}_5\gamma_\mu f_1(p^2,q^2,p\cdot q) \\ &\quad + \hat{\gamma}_5\gamma_{[\mu}\gamma_{\nu]}(p-q)_\nu f_2(p^2,q^2,p\cdot q),\end{aligned} \tag{25}$$

$$|f_1(p^2,q^2,p\cdot q)| < c_1 \frac{|p^2|^\alpha |q^2|^\alpha |(p+q)^2|^\beta}{\{|p^2| + |q^2| + |(p+q)^2| + m^2\}^{2\alpha+\beta}}, \tag{26a}$$

$$|f_2(p^2,q^2,p\cdot q)| < c_2 \frac{|p^2|^{\alpha'}|q^2|^{\alpha'}|(p+q)^2|^{\beta'}}{\{|p^2|+|q^2|+|(p+q)^2|+m^2\}^{2\alpha'+\beta'+\frac{1}{2}}}, \tag{26b}$$

with

$$\alpha,\alpha', \quad \beta'>0, \quad \beta>1, \tag{27}$$

where m is a constant with the dimension of mass.

The asymptotic bounds (26)–(27) are a sort of asymptotic freedom, and seem reasonable especially in non-abelian gauge theories, though theories of asymptotic freedom based on "renormalization group" equations[19–21] do not say anything about asymptotic behaviors of many-point functions with some moments squared and some products of momenta remaining finite and others tending to ∞.

4. ALGORITHMS FOR CONSTRUCTION OF APPROXIMATE SOLUTIONS

In this section, we try to apply the algorithms for construction of approximate solutions to nonlinear operator equations, developed by Altman[17] and Lika[18] for the cases where the mappings are not Fréchet differentiable.

Now let us consider Eq. (5). One of the simplest zeroth approximation one can think of is of the form

$$\Sigma^0(p) = Z\gamma\cdot p + M(1 + b(\operatorname{sgn} p^2)|p^2|^{1/2}), \tag{28}$$

where $b:\{-1,+1\}\to\mathbb{C}$. Then fix the parameters Z and M so as to satisfy the condition

$$\Xi(\Sigma^0,p) \sim Z\gamma\cdot p + M \quad \text{as } p\to 0. \tag{29}$$

Or one may begin with a more complicated zeroth approximation $\Sigma^{0'}$ involving many parameters and fix those parameters so that $\Sigma^{0'}(p)$ and $\Xi(\Sigma^{0'},p)$ have the same asymptotic behaviors for $p^2\to 0$ or $p^2\to 0$ and $p^2\to\infty$ or they coincide at several fixed values of p^2.

In order to apply theorems 1–4 we take a Lipschitz approximation Θ to Ξ defined as follows:

$$\Theta(\Sigma,\alpha_1,\alpha_2,\omega_1,\omega_2;p) = :g\int d^4k\hat{\gamma}_5\gamma_\mu \tilde{G}(\Sigma,\alpha_1,\omega_1,k) \times F_\nu(k,p;-p-k)\tilde{D}_{\mu\nu}(\tilde{G}(\Sigma),\alpha_1,\alpha_2,\omega_1,\omega_2,p+k), \tag{30}$$

$$\tilde{G}(\Sigma,\alpha_1,\omega_1;p) = :\frac{\gamma\cdot p(1+\Sigma_1(p^2))-\Sigma_2(p^2)}{p^2(1+\Sigma_1(p^2))-\Sigma_2(p^2)^2}\{1-\theta_{\omega_1}(\alpha_1-|p^2-P|)\} + \frac{\gamma\cdot p(1+\Sigma_1(p^2))-\Sigma_2(p^2)}{p^2(1+\Xi_1(\Sigma^0,p^2))^2-\Xi_2(\Sigma^0,p^2)^2}\theta_{\omega_1}(\alpha_1-|p^2-P|), \tag{31}$$

$$\begin{aligned}\tilde{D}_{\mu\nu}(\tilde{G}(\Sigma),\alpha_1,\alpha_2,\omega_1,\omega_2;k) = :&\left(g_{\mu\nu}-\frac{k_\mu k_\nu}{k^2}\right)\Big\{k^2+\tfrac{1}{3}\tilde{\pi}^\lambda{}_\lambda(\tilde{G}(\Sigma,\alpha_1,\omega_1),\alpha_2,\omega_2;k) \\ &-\tfrac{1}{3}\frac{k_\lambda k_\kappa}{k^2}\tilde{\pi}_{\lambda\kappa}(\tilde{G}(\Sigma,\alpha_1,\omega_1),\alpha_2,\omega_2;k)\Big\}^{-1}\{1-\theta_{\omega_2}(\alpha_2-|k^2-K|)\} \\ &+\left(g_{\mu\nu}-\frac{k_\mu k_\nu}{k^2}\right)\Big\{k^2+\tfrac{1}{3}\pi^\lambda{}_\lambda(\Sigma^0,k)-\tfrac{1}{3}\frac{k_\lambda k_\kappa}{k^2}\pi_{\lambda\kappa}(\Sigma^0,k)\Big\}^{-1}\theta_{\omega_2}(\alpha_2,-|k^2-K|),\end{aligned} \tag{32}$$

$$\tilde{\pi}_{\lambda\kappa}(\tilde{G}(\Sigma,\alpha_1,\omega_1),\alpha_2,\omega_2;k) = :\pi_{\lambda\kappa}(\Sigma^0,k)\theta_{\omega_2}(\alpha_2-|k^2-K|) + \pi^{\blacktriangle}_{\lambda\kappa}(\tilde{G}(\Sigma,\alpha_1,\omega_1),k)\{1-\theta_{\omega_2}(\alpha_2-|k^2-K|)\}, \tag{33}$$

$$\pi^{\blacktriangle}_{\lambda\kappa}(\tilde{G}(\Sigma,\alpha_1,\omega_1),k) = :g\int d^4q \operatorname{Tr}[\hat{\gamma}_5\gamma_\lambda \tilde{G}(\Sigma,\alpha_1,\omega_1,q-\tfrac{1}{2}k))F_\kappa(q-\tfrac{1}{2}k,-q-\tfrac{1}{2}k;k)\tilde{G}(\Sigma,\alpha_1,\omega_1,q+\tfrac{1}{2}k)] \tag{34}$$

$$\theta_\omega(x) = \int_{-\infty}^{x}\varphi_\omega(x')dx' \tag{35}$$

$$\varphi_\omega(x) = \begin{cases}0, & |x|\geqslant\omega \\ \exp\{-\omega^2(\omega^2-x^2)^{-1}\}, & x<\omega\end{cases} \tag{36}$$

where P and K are real roots of the equations

$$\eta(1+\Xi_1(\Sigma^0,\eta))^2-\Xi_2(\Sigma^0,\eta)=0, \tag{37}$$

$$\Delta(\xi) = :\xi+\tfrac{1}{3}\pi^\lambda{}_\lambda(\Sigma_0,k)|_{k^2=\xi}-\tfrac{1}{3}\frac{k_\lambda k_\kappa}{k^2}\pi_{\lambda\kappa}(\Sigma_0,k)|_{k^2=\xi}=0 \tag{38}$$

with respect to η and ξ, respectively. K in Eq. (32), etc., should not be confused with K in Eq. (A1), etc. If one or both of these equations have two or more roots, the definition of the Lipschitz approximation (30)–(34) must be modified accordingly. Or one may try different input form factors and/or different zeroth approximations.

Now, the Fréchet derivative of the Lipschitz approximation Θ is

$$\begin{aligned}\Theta'(\Sigma,\alpha_1,\alpha_2,\omega_1,\omega_2;\Sigma';p) = -2g\int d^4k\,\hat{\gamma}_5\gamma_\mu &\frac{\gamma\cdot p(1+\Sigma_1(p^2))-\Sigma_2(p^2)^2}{[p^2(1+\Sigma_1(p^2))^2-\Sigma_2(p^2)^2]^2} \\ &\times[p^2(1+\Sigma_1(p^2))\Sigma'_1(p^2)-\Sigma_2(p^2)\Sigma'_2(p^2)]\{1-\theta_{\omega_1}(\alpha-|p^2-P|)\} \\ &\times F_\nu(k,p;-p-k)\tilde{D}_{\mu\nu}(\tilde{G}(\Sigma),\alpha_1,\alpha_2,\omega_1,\omega_2,p+k)+\cdots\end{aligned} \tag{39}$$

so that integration of terms with double factors in denominators are not carried out over neighborhoods of the poles. Now the problem is to check whether various conditions in the theorems presented in the Appendix are satisfied for various choices of parameters and input (model) form factors. If those conditions are satisfied for a certain choice of input form factors and parameters, one can proceed to construct an approximate solution.

For form factors with asymptotic behaviors, (26) and (27), and Σ 's in the Banach space with norm (23), the integrals on the right-hand sides of (5), (30), and (39) are convergent so that Lipschitz conditions are satisfied, and for small $\Sigma^{(\cdot)}$ we get the following estimates:

$$\left.\begin{aligned}&\|\Theta'(\Sigma^{(1)};\alpha_1,\alpha_2,\omega_1,\omega_2)-\Theta'(\Sigma^{(2)};\alpha_1,\alpha_2,\omega_1,\omega_2)\|\leqslant L\,\|\Sigma^{(1)}-\Sigma^{(2)}\|,\\ &\qquad\qquad L=O(\hat{c}g),\hat{c}=\mathrm{Max}\,(c_1,c_2),\end{aligned}\right\}\tag{40a}$$

$$\left.\begin{aligned}&\|\Xi(\Sigma^{(1)})-\Theta(\Sigma^{(1)};\alpha_1,\alpha_2,\omega_1,\omega_2)-\Xi(\Sigma^{(2)})+\Theta(\Sigma^{(2)};\alpha_1,\alpha_2,\omega_1,\omega_2)\|\leqslant L_1\|\Sigma^{(1)}-\Sigma^{(2)}\|,\\ &\qquad\qquad L_1=O(\hat{c}g).\end{aligned}\right\}\tag{40b}$$

Consequently, we get, e.g., the following estimate for constants B_0, η_0, r_0 and h_0 relevant to the applicability of Theorem 3:

$$\left.\begin{aligned}&B_0=(1-O(\hat{c}g))^{-1},\quad \eta_0=O\,[\mathrm{Max}(\|\Sigma^{(0)}\|,\hat{c}gs(\Sigma^{(0)})],\\ &r_0=O(\eta_0),\quad h_0=O\,\{B_0\,\mathrm{Max}\,[\hat{c}g\|\Sigma^{(0)}\|,(\hat{c}g)^2s(\Sigma^{(0)}]\},\end{aligned}\right\}\tag{41}$$

where $s(\Sigma^{(0)})$ is a functional of $\Sigma^{(0)}$. Therefore, the conditions for applicability of Theorem 3 can be satisfied if $\hat{c}g$ is sufficiently small and one chooses a zeroth approximation $\Sigma^{(0)}$ and the parameters α's and ω's appropriately. As for the local uniqueness of the solution, see Theorem 4.

If necessary or desirable, one can modify the definition of the norms in $\mathfrak{B}_1\oplus\mathfrak{B}_2$ by adding some seminorms, e.g.,

$$c^j\sup_{P-\alpha_1-\omega_1\leqslant k^2\leqslant P+\alpha_1+\omega_1}|\partial\Sigma_j(k^2)/\partial(k^2)|,$$

so that functions with undesirable behaviors are excluded from the domain $D(\Xi)$.

Now let us consider the applicability of Theorem 1. If one begins with $\Sigma^{(0)}$ with sufficiently small norm, say $O(\hat{c}g)<1$, one finds that K of (A2) is $O(\hat{c}g)$, so that for the inequality $KC<1$ to hold C need not be small but may be $O\,[(\hat{c}g)^{-1}]>1$. On the other hand, if $cg\ll 1$ then

$$r=O\,[C\exp(1-q)(1-q)^{-1}\,\mathrm{Max}\,(\|\Sigma^{(0)}\|,\hat{c}g)]\tag{42}$$

and

$$C\,\|Px\|=O\,[(\hat{c}g)^{-2}\exp(1-q)(1-q)^{-1}\|\Sigma^{(0)}\|],\tag{43}$$

so that (A8) is a very mild condition unless q is close to 1. Theorem 2 is useful when it is difficult to solve Eq. (A7) at each stage of the successive approximations.

It is also possible to define a new Lipschitz approximation by

$$\begin{aligned}\Theta^N(\Sigma;\alpha_1,\alpha_2,\omega_1,\omega_2;p)=:g\int d^4k\,\hat{\gamma}_5\gamma_\mu\tilde{G}^N(\Sigma;\alpha_1,\omega_1,k)\\ \times F_\nu(k,p;-p-k)\tilde{D}^N_{\mu\nu}(\tilde{G}^N(\Sigma),\alpha_1,\alpha_2,\omega_1,\omega_2,p+k),\end{aligned}\tag{44}$$

$$\begin{aligned}\tilde{G}^N(\Sigma;\alpha_1,\omega_1,p)=:&\frac{\gamma\cdot p(1+\Sigma_1(p^2))-\Sigma_2(p^2)}{p^2(1+\Sigma_1(p^2))^2-\Sigma_2(p^2)^2}\{1-\theta_{\omega_1}(\alpha_1-|p^2-P^N|)\}\\ &+\frac{\gamma\cdot p(1+\Sigma_1(p^2))-\Sigma_2(p^2)}{p^2(1+\Sigma_1^N(p^2))^2-\Sigma_2{}^N(p^2)^2}\theta_{\omega_1}(\alpha_1-|p^2-P^N|),\end{aligned}\tag{45}$$

etc., where Σ^N is the Nth approximation to Σ by the algorithm (A11), (A17), (A26), or (A31), P^N is the zero of the denominator of the second term on the r.h.s. of (41) with respect to p^2, etc., and proceed to higher approximations.

Similarly one can define a Lipschitz approximation and devise an algorithm for Eq. (11), but we do not write it down here, leaving the problem as an exercise for the reader.

Unfortunately, however, these algorithms are not applicable in the presence of constraints, but the existence theorem 5 has a significance to the problem of DMG by parity-violating interactions. To see this let us write the equation for σ in the following abstract form:

$$\sigma_i(p)=\Omega_i[\sigma_1,\sigma_2,\sigma_2,\sigma_4,f,h;p]\quad i=1,2,3,4.\tag{46}$$

If one requires that the γ_5 term be absent, the system (42) is reduced to

$$\sigma_i=\Omega_i[\sigma_1,\sigma_2,\sigma_3,0,f,h;p],\quad i=1,2,3,\tag{47}$$

$$0=\Omega_4[\sigma_1,\sigma_2,\sigma_3,0,f,h;p].\tag{48}$$

If all the f's and h 's are specified as model inputs, there are more equations than unknown functions. On the other hand, if one of h 's, say h_4, is left unspecified, Eq. (18c) can be regarded as a constraint upon h_4. Obviously, this constraint is nonlinear with respect to σ's and linear with respect to f. Moreover, h_4 and Ω_4 do not belong to the same Banach space. The system of equations (17) and (18) as a whole can be regarded as an equation of the form $\Xi(\sigma,h_4)=0$ with the map $\Xi:\mathfrak{D}\subset\mathfrak{B}=:B_1\oplus B_2\oplus B_3\oplus B_h\to\mathfrak{B}'=:\oplus_{i=1}^4 B_i$, where B_i stands for the Banach space of the candidates for σ_i and B_h for the Banach space of the candidates for h_4. As the σ's are functions of one variable while f is a function of three variables, $\mathfrak{B}$ and $\mathfrak{B}'$ are obviously different, and the system $\Xi(\sigma, f) = 0$ does not determine the σ's and f uniquely. But if the conditions of these theorems are satisfied, one can say that the model under consideration has solutions.

5. CONCLUDING REMARKS

As a problem of broken symmetry, the problem of DMG has two fairly distinct aspects. One is the group theoretic aspect and the other is the operator theoretic aspect. In this note we tried to apply the modern NLOT to the latter aspect of the problem, which is essentially concerned with nonlinear singular integral equations, i.e., the unrenormalized Schwinger–Dyson equations or simplified versions of those equations.

As we are mainly concerned with the NLOT aspect, in order to avoid unnecessary complications we have dealt with the $U_{V,A}(1)$ models, but generalizations of the methods and arguments to the SU(n) cases, etc. are rather straightforward, unless parity-violating interactions are involved.

So far we have presented some theorems concerning existence of solutions of unrenormalized Schwinger–Dyson equations with input (model) vertex parts with nice asymptotic behaviors, and algorithms for construction of approximate solutions. As has been seen in Sec.2, the anomalous magnetic moment term in the vertex part generates a mass unless the form factor satisfies some extra conditions. In other words, $\Sigma_2(p^2) = 0$ is not a solution in general, so that one need not look for an exotic (singular at $g = 0$) solution for mass generation. (The g dependence of Γ is irrelevant here.)

It may be helpful to add seminorms to the definition of norms in Banach spaces so that functions with undesirable behaviors will be excluded from the domains and the radii of the domains be increased.

Theorem 4 asserts local uniqueness of solution under certain conditions. On the other hand, the bifurcation theory[22–24] requires Fréchet differentiability, so that it is not applicable to our equations, unfortunately.

Though one may begin with input form factors that are complex above the thresholds of particle production, and Σ and π, also become complex above the thresholds, we do not use the complex analyticity of those functions but treat them as complex-valued functions of real variables, so that our method is applicable also when, for example, the transcendental equations (37) and/or (38) have complex roots. (We do not use Wick rotations.)

Though we dealt with DGM of fermion fields with massless Lagrangian in Sec. 2, the technique developed there can be applied to fermion fields with massive Lagrangians. For example, it may be worthwhile to solve equations for the proton–neutron mass difference with various models of the charge and anomalous magnetic moment form factors.

Though Wick rotation may be applicable to the exact solution of Eq. (5) with the exact three-point function, one cannot expect the applicability of Wick rotation to approximate solutions to equations with an approximate or model three-point function. Moreover, the theory of equations with non-Fréchet-differentiable maps may be useful in solid state physics where the Wick rotation is out of the question.

APPENDIX: RELEVANT MATHEMATICAL DEFINITIONS AND THEOREMS

In this Appendix, some relevant mathematical notions are defined and theorems are presented.

In order to formulate algorithms for construction of approximate solutions of nonlinear operator equations with maps that are not Fréchet differentiable, Altman[17] introduced the notion of Lipschitz approximation.

Definition 1: Let $P : D(p) \subset X \to Y$ be a nonlinear operator. Then an operator $T : D(p) \to Y$ is called a Lipschitz approximation to P if $P - T$ is Lipschitz continuous, i.e., there exists a positive constant K such that for any $x, \bar{x} \in D(p)$

$$\|Px - Tx - P\bar{x} + T\bar{x}\| \leqslant K\|x - \bar{x}\|. \tag{A1}$$

Definition 2: An operator is called a Lipschitz approximation in the narrow sense if there exists a functional $K : D(p) \to \mathbb{R}^+$ and element $h(x) \in X$, a sequence $\epsilon_n \downarrow 0$, and a number $\bar{q} \in (0,1)$ such that for any $\epsilon \in (0,1]$,

$$\|P(x + \epsilon h) - T(x + \epsilon h) - Px - Tx\| \leqslant \epsilon K(x)\|h\|, \tag{A2}$$

$$[T(x + \epsilon_n h) - Tx]/\epsilon_n \to \delta[x, h(x)], \quad n \to \infty, \tag{A3}$$

$$\|\delta(x, h(x)) + Px\| \leqslant \bar{q}\|Px\|, \tag{A4}$$

where $x + \epsilon_n h \in W$ for sufficiently small ϵ_n's.

If a Lipschitz approximation T or P is Fréchet differentiable, the Fréchet derivative $T'(x)$ is continuous in

$$U = :D(p) \cap \bar{S}(x_0, r) \tag{A5}$$

and for any $x \in U_0$ where

$$U_0 = :D(p) \cap S(x_0, r), \tag{A6}$$

there exists an element $h(x) \in X$ such that

$$T'(x)h(x) + Px = 0, \tag{A7}$$

$$\|h(x)\| \leqslant C\|Px\|, \quad KC < 1, \tag{A8}$$

then one can define an algorithm as follows:

Given $x_0 \in D(p)$, $KC/q < \beta < 1$, and $KC < q < 1$, suppose $x_1, \cdots, x_n, n \geqslant 0$ are already defined. Then put $\epsilon_n = 1$ if $\Phi(1, x_n, h_n) \leqslant q\|Px_n\|$, where

$$\Phi(\epsilon, x, h) = :\|P(x + \epsilon h) - (1 - \epsilon)Px\|/\epsilon. \tag{A9}$$

and h_n is the solution of Eq. (A7) with $x = x_n$.

If $\Phi(1, x_n, h_n) > q\|Px_n\|$, there exists a number $\epsilon_n \in (0,1)$ such that

$$\beta q\|Px_n\| \leqslant \Phi(\epsilon_n, x_n, h_n) \leqslant q\|Px_n\|. \tag{A10}$$

In either case, put

$$x_{n+1} = x_n + \epsilon_n h_n. \tag{A11}$$

In other words, the problem is reduced to the inversions of the linear operators $T'(x)$. Then we have the following theorem due to Altman.[17]

Theorem 1: If T is a Lipschitz approximation to P in U_0 and T' is continuous in U and satisfies the conditions (A7) and (A8) in U_0 with radius

$$r \geqslant (1 - q)^{-1} C \exp(1 - q)\|Px_0\|, \tag{A12}$$

then $x_n \in U_0$ and there exists at least one x^* such that $Px^* = 0$. The error estimate reads

$$\|x_n - x^*\| \leqslant (1 - q)^{-1} C b_n, \tag{A13}$$

where

$$b_n = \|Px_0\| \exp((1-q)(1-t_n)),$$
$$t_0 = 0, \quad t_n = \sum_{i=0}^{n-1} \epsilon_i. \tag{A14}$$

Alternatively we have the following theorem, also due to Altman.[17]

Theorem 2: If there exists $q \in (0,1)$ such that for any $x \in U_0$ there exists $h(x)$ such that

$$\|T'(x)h(x) + Px\| \leqslant \bar{q}\|Px\|,$$
$$\|h(x)\| \leqslant C\|Px\|, \tag{A15}$$
$$KC + \bar{q} < 1,$$

one can define an algorithm as follows. Given $x_0 \in D(p)$ and

$$KC + \bar{q} < q < 1, \ (KC + \bar{q})/q < \beta < 1. \tag{A16}$$

Suppose that $x_1,\ldots,x_n$ are already defined. Then put $\epsilon_n = 1$ or choose $\epsilon_n < 1$ in the same way as in Theorem 1 provided q and β are subject to the condition (A16), and define

$$x_{n+1} = x_n + \epsilon_n h_n. \tag{A17}$$

If the radius r of S is larger than $(1-q)^{-1}\exp(1-q)\|Px_0\|$ then $\{x_n\} \subset U$ and $x_n \to x^*$ as $n \to \infty$.

We have also the following theorems due to Lika[18]

Theorem 3: If the conditions

1) The Fréchet derivative $Q'(x)$ of an operator Q and an operator

$$F = :P - Q \tag{A18}$$

satisfy Lipschitz conditions in a domain M with constants L and L_1, respectively:

$$\|Q'(x) - Q'(x')\| \leqslant L\|x - x'\|, \tag{A19}$$
$$\|F(x) - F(x')\| \leqslant L_1\|x - x'\|; \tag{A20}$$

2) there exists

$$\Gamma_0 = :[Q'(x_0)]^{-1}, \tag{A21}$$

and

$$\|\Gamma_0\| \leqslant B_0, \quad \|\Gamma_0 P(x_0)\| \leqslant \eta;. \tag{A22}$$

$$3)\ B_0 L_1 < 1; \tag{A23}$$

$$4)\ h_0 = :B_0 L \eta_0 \leqslant \tfrac{1}{2}(1 - B_0 L_1)^2; \tag{A24}$$

5) ball $S(x_0, r_0)$ is constrained in M where

$$r_0 = :[1 - B_0 L_1 - \{(1 - B_0 L_1)^2 - 2h_0\}^{1/2}]\eta_0 h_0^{-1}$$
$$= :R(h_0)\eta_0; \tag{A25}$$

then the equation $P(x) = 0$ admits a solution $x^* \in S(x_0, r_0)$ and the sequence $\{x_n\}$ defined by

$$x_{n+1} = x_n - [Q'(x_n)]^{-1}P(x_n) \tag{A26}$$

converges to this solution. The rate of convergence is

$$\|x_n - x^*\| \leqslant R(h_n)\eta_n, \tag{A27}$$

where

$$\|\Gamma_n\| \leqslant B_n, \quad \|\Gamma_n P(x_n)\| \leqslant \eta_n, \quad h_n = :B_n L \eta_n \tag{A28}$$

$$R(h_n)\eta_n = [1 - B_n L_1 - \{(1 - B_n L_1)^2 - 2h_n\}^{1/2}] \times (B_n L)^{-1}.$$

Theorem 4: If the conditions 1)–3) of Theorem 3 and the condition

$$4')\ h_0 < 2N^{-2}\{(1 - B_0 L_1)N - 1\} \tag{A29}$$

for an N such that

$$(1 - B_0 L_1)^{-1} < N \leqslant 2(1 - B_0 L_1)^{-1}, \tag{A30}$$

are satisfied, then the sequence $\{x_n\}$ defined by

$$x_{n+1} = x_n - [Q'(x_0)]^{-1}P(x_n) \tag{A31}$$

converges to the only solution $x^* \in S(x_0, N\eta_0)$ of the equation $Px = 0$, i.e., the solution x^* is locally unique. The error estimate reads

$$\|x_n - x^*\| \leqslant (Nh_0 + B_0 L_1)^n \eta_0 (1 - Nh_0 - B_0 L_1)^{-1}. \tag{A32}$$

In order to see the existence of solution of problems with constraints Theorem 5 below, also due to Altman,[17] is useful.

Definition 3: The set $\Gamma_x(\Phi)$ of special contraction direction for a map $\Phi : D(\Phi) \to Y$ at $x \in D(\Phi)$ is defined as the set of elements $y \in Y$ such that there exist numbers $B, q \in (0,1)$, $\epsilon \in (0,1]$, and an element $\bar{x} \in X$ such that

$$\|\Phi\bar{x} - \Phi x - \epsilon y\| \leqslant q\epsilon\|y\|, \tag{A33}$$

where the distance between x and $\bar{x}$, $d(x, \bar{x}) \leqslant \epsilon\|y\|$.

If Φ is a closed operator, one can state the following theorem.

Theorem 5: Suppose

1) $\Phi : X_0 \to Y$ is closed on $U = :X_0 \cap \bar{S}$ where X_0 is a subset of X and $\bar{S}$ the closure in X of the ball

$$S(x_0, r) = :\{x \in X \mid d(x, x_0) < r\}; \tag{A34}$$

2) for any $x \in U_0 = :X_0 \cap S$ the set $\Gamma_x(\Phi)$ of special contractor directions is dense in some ball in Y with center 0;

$$3)\ r \geqslant B(1 - \bar{q})^{-1}\|\Phi x_0\|, \quad q < \bar{q} < 1; \tag{A35}$$

then the equation $\Phi x = 0$ has a solution $x^* \in U$. The main difficulty in application of this theorem is to verify the denseness of the special contractor directions in a ball in Y for a given map Φ.

[1] Y. Nambu and G. Jona-Lasinio, Phys. Rev. **122**, 345 (1961); **124**, 246 (1961).
[2] J. Schwinger, Phys. Rev. **125**, 397 (1962).
[3] Th. A. J. Maris, V. E. Herscovitz, and G. Jakob, Phys. Rev. Lett. **12**, 313 (1964).
[4] F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964).
[5] F. Englert, J. M. Frère, and P. Nicoletopoulos, Phys. Lett. B **49**, 77 (1974); B **52**, 433 (1975); B **59**, 346 (1975); Nucl. Phys. B **95**, 269 (1975).
[7] P. Budini, Acta Phys. Austr. Suppl. **15**, 499 (1976).
[8] S. Weinberg, Phys. Rev. D **13**, 974 (1976); D **19**, 1277 (1979).
[9] H. Pagels, Phys. Rev. D **19**, 3080 (1979); D **21**, 2336 (1980).
[10] D. Fryberger, Phys. Rev. D **20**, 952 (1979).
[11] K. Higashijima, T. K. Kuo, and P. G. Lauwers, Phys. Rev. D **20**, 1670 (1979).
[12] L. Susskind, Phys. Rev. D **20**, 2619 (1979).
[13] L. H. Ford and T. Yoshimura, Phys. Lett. A **70**, 89 (1979).
[14] T. Yoshimura, Phys. Lett. A **72**, 391 (1979).
[15] D. J. Toms, Phys. Rev. D **21**, 928 (1980); Phys. Lett. A **77**, 303 (1980).
[16] M. Altman, *Contractors, Contractor Directions, Theory and Applications* (Marcel Dekker, New York, 1977).
[17] M. Altman, Nonlinear Analysis, Theory, Methods, and Applications **4**, 347 (1980); **4**, 761 (1980).
[18] D. K. Lika, *Analiza funktsionale shi rezolvarya aproksimative a ekuatsiilor*

neliniare (Functional Analysis and Approximate Solutions of Nonlinear Equations) (Editura "Shtiintsa," Kishineu, 1975).
[19]D. J. Gross and F. Wilczek, Phys. Rev. D **8**, 3633 (1973); D **9**, 980 (1974).
[20]A. Hosoya and A. Sato, Phys. Lett. B **48**, 36 (1973).
[21]R. Oehme and W. Zimmermann, Phys. Rev. D **21**, 1661 (1980); D **22**, 2534 (1980).
[22]M. A. Krasnosel'skiĭ, G. M. Vainikko, P. P. Zabreiko, Ya. B. Rutitskiĭ, and V. Ya. Sitenko, *Approximate solution of operator equations* (Noordhof, Groningen, 1972).
[23]M. S. Berger, *Nonlinearity and Functional Analysis* (Academic, New York, 1977).
[24]D. H. Sattinger, *Group Theoretic Methods in Bifurcation Theory* (Springer, Berlin, 1979).

Remarks on spacetime symmetries and nonabelian gauge fields

Marc Henneaux [a]
Center for Theoretical Physics, The University of Texas at Austin, Austin, Texas 78712 [b]

(Received 13 January 1981; accepted for publication 18 March 1981)

Some properties of nonabelian gauge fields invariant under a given group of spacetime transformations are studied and applied to the case of spatial homogeneity and isotropy. All the $k = 0$ and $k = 1$ Robertson–Walker models filled with a SO (3) gauge field are derived.

PACS numbers: 11.10.Np, 11.30. − j, 04.20.Jb

1. INTRODUCTION

A gauge field of the Yang–Mills type is said to be invariant under a Lie group of spacetime diffeomorphisms G if and only if any element $g \in G$ induces a transformation of the potential that can be compensated for by a gauge transformation.[1]

The first purpose of this paper is to derive some properties of invariant Yang–Mills fields. This study is illustrated in the case of spatial homogeneity and isotropy. It is shown that there exists spatially homogeneous and isotropic nonabelian configurations which are not pure gauge (this is in sharp contrast with electromagnetism). We end up with the resolution of the coupled Einstein–Yang–Mills equations under these symmetry assumptions.

The gauge field is taken to be SO(3). The spacetime manifold and the Yang–Mills potential are, respectively, denoted by M and A ($= A^a_\mu T_a dx^\mu$ in local coordinates; $[T_a, T_b] = \epsilon_{abc} T_c$). $A(P)$ is the value of A at $P \in M$.

2. SYMMETRIC GAUGE FIELDS

When the gauge field A is G-invariant (or, as it is also said, "G-symmetric"), there exists a differentiable (up to the required order) application R from $M \times G$ to SO (3) such that

$$(g^*A)(P) = R(P,g)A(P)R^{-1}(P,g) + dR(P,g)R^{-1}(P,g) \tag{2.1}$$

and reciprocally. Here, g^*A is the transformed potential obtained from A by the usual pull back of 1-forms.

The connection 1-form $\omega(A)$ defined on the corresponding principal bundle—supposed to be the trivial bundle $M \times \mathrm{SO}(3)$ for simplicity[2]—is invariant by a group H of bundle automorphisms whose action on the base M is given by G. However, as we shall show, G is isomorphic to no subgroup of H in some cases. This leads to difficulties in the study of G-invariant gauge fields.

The compensating gauge transformations $R(P,g)$ are gauge-dependent. In a new gauge, the gauge transformations $R'(P,g)$ given by

$$R'(P,g) = U(g(P))R(P,g)U^{-1}(P) \tag{2.2}$$

satisfy (2.1), with $A' = UAU^{-1} + dU\,U^{-1}$ in place of A.

Let $\mathscr{L}$ be the group of gauge transformations that leave A unchanged. It is easy to prove that the gauge transformations $C(P;g_1,g_2)$ defined by

$$C(P;g_1,g_2) = R^{-1}(P,g_1)R^{-1}(g_1(P),g_2)R(P,g_2g_1) \tag{2.3}$$

belong to $\mathscr{L}$ (in the product g_2g_1, g_1 acts first). They are obviously modified when, instead of the application R, one uses the compensating application $R(P,g)L(P,g)$, $L(P,g) \in \mathscr{L}$. A necessary and sufficient condition for the existence of a set of compensating transformations $\bar{R}(P,g) = R(P,g)L(P,g)$ obeying the simple composition rule

$$\bar{R}(P;g_2g_1) = \bar{R}(g_1(P),g_2)\bar{R}(P,g_1) \tag{2.4}$$

is that $C(P;g_1,g_2) - I$ can be made to vanish by a suitable choice for $L(P,g)$. This is not always the case. Consider, for example, the flat space potential $A_\sigma(x^\mu)$ given in Minkowskian coordinates by $M\delta^0_\sigma x^1$, where M is a constant matrix belonging to the SO (3)-algebra. This potential is translation-invariant; with the 4-translation a^μ which maps x^μ on $x^\mu + a^\mu$, one can associate the compensating gauge transformations $R(x^\mu, a^\mu) = \exp[(a^1x^0 + f(a^\mu))M]$. Computing (2.3) for any two translations a^μ and b^μ, one finds

$$C(x^\mu;a^\mu,b^\mu) = \exp[(-b^1a^0 + f(a^\mu + b^\mu) - f(a^\mu) - f(b^\mu))M] \neq I.$$

Indeed, the exponent cannot vanish for all a^μ, b^μ since b^1a^0 is not symmetric under the exchange $a^\mu \leftrightarrow b^\mu$.

It is only when (2.4) holds or can be made to hold that the action of the group G on M can be "lifted" to the bundle $M \times \mathrm{SO}(3)$ in a manner that preserves the group structure. Indeed, let g be an element of G and $R(P,g)$, one of the corresponding compensating gauge transformations. They define a transformation of the section $M \times \{I\}$,

$$h_g: (P,I) \mapsto h_g(P,I) = (g(P), R^{-1}(P,g)), \tag{2.5}$$

that can be extended in a unique way to the whole of $M \times \mathrm{SO}(3)$ if one requires that h_g be a bundle automorphism [i.e., commute with any element of SO (3)]:

$$h_g: M \times \mathrm{SO}(3) \to M \times \mathrm{SO}(3): (P,\mathscr{R}) \mapsto (g(P), \mathscr{R}R^{-1}(P,g)). \tag{2.6}$$

By construction, the connection ω is h_g-invariant: $h^*_g\omega = \omega$ [use (2.1), (2.5), and (2.6)]. However, if (2.4) does not hold, $h_{g_1g_2} \neq h_{g_1}h_{g_2}$. The difference is a transformation along the fibers. Let H be the group generated by the h_g's. It is a subgroup of the group of bundle automorphisms that leave the connection 1-form ω invariant. It possesses a "normal" (or "invariant") subgroup $\mathscr{N}$, which contains all the elements of H that map each fiber on itself (i.e., the action of $\mathscr{N}$ on the base space M reduces to the identity). $\mathscr{N}$ is also a subgroup of $\mathscr{L}$. The group G is the quotient of H by $\mathscr{N}$. The existence of

[a] Aspirant du Fonds National Belge de la Recherche Scientifique.
[b] Permanent address: Interactions Fondamentales, Université Libre de Bruxelles, Campus Plaine C. P. 228, B-1050 Bruxelles, Belgium.

a group homomorphism $H \to G$ does not imply that G is isomorphic to some subgroup of H.

The classification of G-invariant gauge fields must be based on a classification of all possible groups H and $\mathcal{N}$ with the above properties. Two cases need be considered here: (i) $\mathcal{N} = \{I\}$; (ii) $\mathcal{N}$ contains a one-parameter subgroup, say $\{e^{\lambda\varphi(P)}\}$ ($\lambda \in [0,2\pi)$, $\varphi(P) \in$ SO(3)-algebra).

In the first case, H and G are isomorphic. The compensating gauge transformations $R(P,g)$ obey the composition law (2.4), from which the relations

$$R(P,e) = I, \tag{2.7a}$$

$$R(P,g^{-1}) = R^{-1}(g^{-1}(P),g) \tag{2.7b}$$

follow. The construction (up to a gauge transformation) of all G-invariant Yang–Mills fields may be found in the literature.[3] We shall not repeat it in this note; rather, we shall turn to the second case.

3. THE QUASIABELIAN CASE

If the gauge transformation $e^{\lambda\varphi(P)}$ leave the potential A invariant, the field $\varphi(P)$ is covariantly constant. Let us assume for simplicity that the spacetime manifold M is connected and has a zero second Betti number. In the SO(3) case, this implies the existence of a gauge in which φ is constant, $A(P) = a(P)\varphi$ ($A^a_\mu(x) = a_\mu(x)\varphi^a$), and $F(P) = f(P)\varphi$ ($F^a_{\mu\nu}(x) = f_{\mu\nu}(x)\varphi^a; f = da$).

Unless the field strengths all vanish, any gauge transformation $L(P)$ that leaves the potential unchanged belongs to the group $\{e^{\lambda\varphi}\}$ (use $F = LFL^{-1}$), which thus coincides with $\mathcal{N}$. Besides, any gauge transformation $R(P,g)$ satisfying (2.1) must be a (point-dependent in general) rotation about the direction in isospin space determined by φ. The SO(2) potential $a(P)$ is G-invariant ($g^*a - a$ is a gradient) and the problem is completely reduced to the abelian one.[4]

We shall merely analyze here the properties of G-invariant abelian potentials from a local point of view. For the sake of briefness, we shall also assume that the group G acts transitively on M. The discussion can readily be extended to more general situations.

In local coordinates, the invariance equations (2.1) read

$$\pounds_{\xi_A} a_\mu = -\partial_\mu \chi_A, \tag{3.1}$$

where ξ_A ($A = 1,...,n$) are the generators of G. The scalar functions $\chi_A(x)$ are submitted to the following conditions.

$$\partial_\mu(\pounds_{\xi_A}\chi_B - \pounds_{\xi_B}\chi_A - C^C_{AB}\chi_C) = 0, \tag{3.2a}$$

which one easily derives from (3.1). Here, the constants C^C_{AB} are the structure constants of G ($[\xi_A, \xi_B] = C^C_{AB}\xi_C$). From (3.2a), it follows that

$$\pounds_{\xi_A}\chi_B - \pounds_{\xi_B}\chi_A - C^C_{AB}\chi_C = K_{AB}, \tag{3.2b}$$

where K_{AB} are antisymmetric constants obeying

$$C^D_{AB}K_{DC} + C^D_{BC}K_{DA} + C^D_{CA}K_{DB} = 0 \tag{3.3}$$

[to get (3.3), take the Lie derivative of (3.2b) with respect to ξ_C, and add to the resulting expression the similar ones obtained by cyclic permutations of (ABC)]. Moreover, these constants are gauge invariant (i.e., invariant when the functions χ_A are replaced by $\chi_A + \pounds_{\xi_A} u$).

The geometrical meaning of the K_{AB}'s is clear: let $\bar{\xi}_A = (\xi^\mu_A(x), \chi_A(x))$ and $\bar{\eta} = (0,1)$ be the generators in $M \times$ SO(2) [local coordinates used: (x^μ, λ)] of the $n+1$ dimensional Lie group H (the lift of G on the bundle). The vector fields $\bar{\xi}_A$ commute with $\bar{\eta}$:

$$[\bar{\xi}_A, \bar{\eta}] = 0, \tag{3.4a}$$

because they generate bundle automorphisms. It is easy to compute the other structure constants of H:

$$[\bar{\xi}_A, \bar{\xi}_B] = C^C_{AB}\bar{\xi}_C + K_{AB}\bar{\eta}, \tag{3.4b}$$

$$[\bar{\eta}, \bar{\eta}] = 0. \tag{3.4c}$$

The relation (3.3) is nothing but (part of) the Jacobi identity for these structure constants.

The generators $\bar{\xi}_A$ are determined up to arbitrary multiples $\mu_A\bar{\eta}$ of the vector field $\bar{\eta}$. This amounts to adding arbitrary constants μ_A to the functions $\chi_A(x)$, and results in a change of the constants K_{AB} given by

$$K_{AB} \to K'_{AB} = K_{AB} - C^C_{AB}\mu_C. \tag{3.5}$$

Two sets K_{AB} and K'_{AB} that differ by such a combination are to be identified. One can use this arbitrariness to give some "canonical" values to K_{AB}. For example, one can assume $K_{AB} = 0$ when the group G of space-time diffeomorphisms is two-dimensional and nonabelian, or three-dimensional and of Bianchi type VIII or IX. The G-Lie algebra is isomorphic to an H-Lie subalgebra in those cases.

Let $S(P_0)$ be the "isotropy" (or "stability") subgroup of G at some reference point P_0, and let $\xi^*_a = d^A_a\xi_A$ ($a = 1,...,m = n-4$) be its generators ($[\xi^*_a, \xi^*_b] = C^c_{ab}\xi^*_c$). As is known, they vanish at P_0. However, the gauge invariant constants $k_a = d^A_a\chi_A(P_0)$ do not necessarily vanish and cannot, in general, be made to vanish by a transformation $\chi_A \to \chi_A + \mu_A$ satisfying $C^C_{AB}\mu_C = 0$. Thus, even when $K_{AB} = 0$, the relation (3.2b) does not imply that the functions χ_A are the components of a gradient (i.e., $\chi_A = \pounds_{\xi_A}\mu (= \xi^\mu_a u_\mu)$). The remaining arbitrariness mentioned above can nevertheless be used to bring the numbers k_a to some canonical values.

Consider two sets of functions $\chi_A(x)$ and $\chi'_A(x)$ defining the same constants K_{AB} and k_a. It is easily shown that they differ by a gradient:

$$\chi_A(x) - \chi'_A(x) = \pounds_{\xi_A}\mu(x).$$

Accordingly, the vector field $\bar{\xi}_A$ is mapped on the vector field $\bar{\xi}'_A$ by the gauge transformation $x'^\mu = x^\mu, \lambda' = \lambda + u(x)$.

It results from this analysis that the lifted groups H and their action on the bundle $M \times$ SO(2) can be classified according to the various canonical values of the pair (K_{AB}, k_b), subject to the conditions (3.3) and $d^A_a d^B_b K_{AB} = -C^c_{ab}k_c$. Given such a pair, one can integrate, at least locally, Eqs. (3.2b) for χ_A with the "initial" conditions $d^A_a\chi_A(P_0) = k_a$. One can then find all the corresponding G-invariant gauge fields a_μ by solving the invariance equations $\pounds_{\bar{\xi}_A}\omega = 0$ for the connection form $\omega = (a_\mu(x),1)$; this is a standard problem of differential geometry. In most cases, however, it is simpler to construct all the G-invariant field strengths $f_{\mu\nu}(\pounds_{\xi_A} f_{\mu\nu} = 0)$ and to require that they be closed.

4. SPATIAL HOMOGENEITY AND ISOTROPY

Suppose that M is a pseudo-Riemannian manifold of the Robertson–Walker type; its metric possesses a six-dimensional group G_6 of isometries acting on a one-parameter family of spacelike hypersurfaces. In standard coordinates, one has

$$ds^2 = -dt^2 + b^2(t)d\sigma^2, \tag{4.1}$$

where $d\sigma^2$ is the time-independent metric of a Riemannian three-space of constant curvature k. For simplicity, only the $k = 0$ (flat) and $k = 1$ (closed) models will be considered here. We intend to find all the SO(3) gauge fields that share the symmetry of the metric.

The quasiabelian case is not interesting, since the Yang–Mills field would be pure gauge. Indeed, the field strengths would vanish as a result of the symmetry requirements.

We shall thus turn to the case $\mathscr{N} = \{I\}$. In order to derive all the homogeneous and spherically symmetric Yang–Mills fields, we shall use the properties (2.4), (2.7a), and (2.7b) of the compensating gauge transformations $R(P,g)$.

As is known, the group G_6 of isometries has a three-parameter subgroup, called here the "translation" group, that acts simply transitively on the surfaces $t =$ const. This group is abelian when $k = 0$, and of type IX (according to the classification given by Bianchi) when $k = 1$. Let $\{dt, \omega^m = \omega_n^m(x^k)dx^n\}$ be a set of independent translation-invariant 1-forms. In this paragraph, latin indices take the "spatial" values 1,2,3. The 1-forms ω^m can be chosen so that $d\omega^m = 2^{-1}C^m_{pq}\omega^p \wedge \omega^q$ (with $C^m_{pq} = k\epsilon_{mpq}$) and $d\sigma^2 = \Sigma_m(\omega^m)^2$ (see the book by Ryan and Shepley[5]).

It is always possible to find a gauge in which $R(P,g) = I$ for all translations. Indeed, consider an arbitrary reference point P_0 and all the other points P'_0, P''_0...that lie on the same t-coordinate line (one such reference point per hypersurface of transitivity). By performing the gauge transformation $U(P) = R^{-1}(P_0(P),\bar{g}_P) = R(P,\bar{g}_P^{-1})$, where $\bar{g}_P$ is the unique translation that maps on P the reference point $P_0(P)$ belonging to the same hypersurface of transitivity, one gets new compensating gauge transformations $R'(P,g)$ which all reduce to the identity. In this gauge, the Yang–Mills field is strictly translation-invariant ($g^*A = A$ for all translations). Accordingly, one has

$$A = A_0(t)dt + A_m(t)\omega^m, \tag{4.2}$$

where the matrices A_0 and A_m depend only on t. Time-dependent gauge transformations leave the potential (4.2) form-invariant.

Let us now analyze the conditions resulting from the isotropy. The application R defines group homomorphisms from the isotropy subgroups $S(P)$ to SO(3) because it obeys the composition law (2.4). Since the groups $S(P)$ are all isomorphic to SO(3)—we exclude spatial reflections from G_6—their image by R can only be the trivial group $\{I\}$ or SO(3) itself. If two points belong to the same surface of homogeneity, their stability subgroups are mapped on the same SO(3) subgroup. This is a consequence of the relation

$$R(P,g) = R(P_0(P),u), \quad u = \bar{g}^{-1}_{g(P)} g \bar{g}_P \in S(P_0(P)), \tag{4.3}$$

which holds for any $g \in G_6$ in the gauges chosen above.

Suppose that the isotropy subgroups at all reference points—and thus everywhere in spacetime—are mapped on SO(3) (if the continuous application R maps these subgroups on $\{I\}$, the potential A, being strictly symmetric, is pure gauge). After an appropriate time-dependent gauge transformation is performed, any spatial rotation about the reference points is compensated by the "same" rotation in isotopic space. In that particular gauge, $R(P_0,g)$ is independent of the hypersurfaces of transitivity. Moreover, since the translation group is an invariant subgroup of G_6, one can infer from (4.3) that $R(P,g) = R(P_0,g)$, i.e., $dR = 0$. The invariance equations read

$$R^{\ n}_{m} A^{\ a}_{n} = R^{\ a}_{b} A^{\ b}_{m}, \quad A^{\ a}_{0} = R^{\ a}_{b} A^{\ b}_{0} \tag{4.4}$$

for every $R^{\ n}_{m} \in$ SO(3). This implies

$$A(t) = a(t)T_m\omega^m. \tag{4.5}$$

The field (4.5) is, up to a gauge transformation, the most general spatially homogeneous and spherically symmetric SO(3) Yang–Mills field.

By direct computation, one gets the field strengths

$$F(t) = \dot{a}T_m dt \wedge \omega^m + 2^{-1}(ka - a^2)T^m\epsilon_{mpq}\omega^p \wedge \omega^q. \tag{4.6}$$

In the flat model, they vanish at $a = 0$, whereas in the closed one, they vanish at both $a = 0$ and $a = 1$.

5. A SOLUTION TO EINSTEIN–YANG–MILLS EQUATIONS

On account of its symmetry, the energy–momentum tensor of the Yang–Mills field is a function of time alone and has the perfect fluid form. Moreover, the pressure it defines is equal to one-third of the energy density $\epsilon[2\epsilon = 3b^{-4}(b^2\dot{a}^2 + (ka - a^2)^2)]$, because its trace vanishes. The solution to Einsteins' equations for the metric b is thus well known [$b \sim t^{1/2}$ in the open model and $b \sim \sin\eta$, $t \sim (1 - \cos\eta)$ in the closed one]. As to the Yang–Mills field, it is determined by quadratures from the conservation law $\epsilon b^4 =$ const. This problem is equivalent to the integration of the one-dimensional motion in the positive quartic potential $(ka - a^2)^2$, which has two minima in the $k = 1$ case.

This example shows the existence of nontrivial, nonstationary, spherically symmetric and homogeneous real solutions to Yang–Mills equations (with finite energy when the spatial sections are closed). Such solutions have no analog in the electromagnetic case. Their detailed physical significance (if any) is, however, not known.

Note added in proof: The interested reader might have noticed the great similarities between the question studied in Sec. 3 and the problem of ray representations of Lie groups: V. Bargmann, "Ray Representations of Lie Groups," a series of six lectures delivered at the University of Texas on Feb. 23, 25, 27 and March 2, 4, 6, 1981 (unpublished); Ann.

Math. **59**, (1) (1954). In both cases, one lifts a group of transformations defined on the base space to the bundle space of an appropriate fiber bundle. After this paper was completed, the author became aware of the following works, which deal with similar topics: Gu Chaohao and Hu Hesheng, Commun. Math. Phys. **79**, 75 (1981) (spherically symmetric gauge fields) and A. V. Gaiduk, Theoret. Math. Phys. **44**, 795 (1980) (translation-invariant gauge fields).

ACKNOWLEDGMENTS

It is a pleasure to thank Professor J. A. Wheeler and the Center for Theoretical Physics of The University of Texas at Austin for their kind hospitality. Many discussions with Professor C. Teitelboim are also gratefully acknowledged.

[1]P. Cordero and C. Teitelboim, Ann. Phys. (N.Y.) **100**, 607 (1976); R. Benguria, P. Cordero, and C. Teitelboim, Nucl. Phys. B **122**, 61 (1977); V. N. Romanov, A. S. Schwarz, and Yu. S. Tyupkin, Nucl. Phys. B **130**, 209 (1977); see also P. G. Bergmann and E. J. Flaherty, J. Math. Phys. **19**, 212 (1978); J. Harnad, S. Shnider, and L. Vinet, *ibid.* **20**, 931 (1979); A Trautman, J. Phys. A **13**, L1 (1980).

[2]The notation $\omega[A]$ means that A derives from ω, i.e., $A = s^*\omega$, where s is in our case the cross section of $M \times SO(3)$ defined by $s{:}P \to s(P) = (P,I)$. See the recent introductions to fiber bundles for physicists: A. Trautman, "Fiber Bundles, Gauge Fields and Gravitation," in *General Relativity and Gravitation-One Hundred Years After the Birth of Albert Einstein*, edited by A. Held (Plenum, New York, 1980), Vol. 1, p. 287; M. Daniel and C. M. Viallet, Rev. Mod. Phys. **52**, 175 (1980).

[3]A. S. Schwarz, Commun. Math. Phys. **56**, 79 (1977); P. Forgács and N. S. Manton, *ibid.* **72**, 15 (1980).

[4]In fact, one could have $g^*f = -f$ (and thus $g^*a - a \neq dm$). Example: $A_t = A_y = A_z = 0$, $A_x = a_x(t)T_z$. This potential is invariant by the group $\{e,g\}$, where g is a 180° rotation about the z axis: g: $t' = t$, $x' = -x$, $y' = -y$, $z' = z$. One has $g^*A = RAR^{-1}(= -A)$, where R is, for example, a constant 180° rotation about the x axis in isotopic space. But $g^*a = -a$ does not differ from a by a gradient. The abelian potential a is not g-invariant. Such a possibility is ruled out, however, in the case of spacetime diffeomorphisms and compensating gauge transformations close to the identity.

[5]M. P. Ryan and L. C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U. P., Princeton, N.J., 1975), Chap 6.

SU(3) wave solutions [a)]

Eve Kovacs
Stanford Linear Accelerator Center, Stanford University, Stanford, California 94305

(Received 22 September 1981; accepted for publication 20 November 1981)

An *ansatz* yielding propagating wave solutions for pure SU(3) gauge theories is exhibited. The solutions are self-dual and have a superposition property like their SU(2) analogs. Possible generalizations of the *ansatz* which may be used to obtain additional irreducible SU(3) solutions are also suggested.

PACS numbers: 11.10.Np, 11.30.Jw

I. INTRODUCTION

SU(2) Yang–Mills theories provide the simplest examples of theories with a non-abelian gauge symmetry and hence have been the focus of most investigations. However, since the strong interactions may be mediated by an octet of color SU(3) gauge fields, it is also important to consider classical solutions of SU(3) gauge theories. Apart from the possibility that such solutions will be relevant to the quantum theory, the study of SU(3) solutions is attractive for a number of other reasons. Firstly, results obtained for SU(2) theories generalize readily to higher rank gauge groups since it is always possible to embed known SU(2) solutions. Secondly, because SU(3) has an inherently more complex structure than SU(2), it is possible that some nontrivial generalizations of SU(2) solutions exist, apart from the straightforward embeddings. Finally, unlike the SU(2) case, it is possible to construct a stable solution to the pure SU(3) gauge field equations without the introduction of explicit scalar fields.[1]

A large number of SU(3) solutions have now been discovered. These include generalizations of the 't Hooft–Polyakov monopole[1–6] and the Prasad–Sommerfield monopole,[7] as well as SU(3) dyons.[8] Models of SU(3) monopoles coupled to fermions have also been considered.[9]

SU(3) instantons with topological charges of ± 1 and ± 4, corresponding to the two inequivalent embeddings of the group SU(2) inside SU(3),[10,11] have been obtained. Irreducible SU(3) solutions have resulted from an O(3) symmetric *ansatz*[12,13] corresponding to an SU(3) generalization of Witten's cylindrically symmetric multi-instanton solution.[14] SU(3) versions of meron[15] and multimeron[16] configurations also exist and investigation of complex SL(3,C) self-dual fields[17] has yielded a number of interesting nontrivial solutions.

The Corrigan–Fairlie–'t Hooft–Wilczek (CFtHW) ansatz,[18] however, has not yet been generalized to SU(3) due to the difficulty of finding an analog of the 't Hooft tensor $\eta_{a\mu\nu}$. We have utilized a particularly simple version of this ansatz for the investigation of propagating wave solutions in SU(2) gauge theories.[19] In addition to the natural interest in SU(3) versions of these solutions, we might hope to gain some indication of possible generalizations of $\eta_{a\mu\nu}$ appropriate to SU(3).

In Sec. II we exhibit a generalization of the O(3) symmetric *ansätze* used by other authors[8,13] to find SU(3) solutions. This generalization is shown to yield the SU(3) version of the wave solutions mentioned above. The self-duality properties of the *ansatz* are considered in detail in Sec. III and possible generalizations which may be useful for finding the SU(3) analog of the CFtHW ansatz are discussed in Sec. IV.

II. SU(3) WAVE SOLUTIONS

We begin by writing the gauge potential A_μ and field strength $F_{\mu\nu}$ as matrices in the space of infinitesimal group generators

$$A_\mu = \frac{A_\mu{}^a T^a}{2i}, \tag{2.1a}$$

$$F_{\mu\nu} = \frac{F_{\mu\nu}{}^a T^a}{2i}, \tag{2.1b}$$

with

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]. \tag{2.1c}$$

The equation of motion may thus be written as

$$D^\nu F_{\mu\nu} = \partial^\nu F_{\mu\nu} + [A^\nu, F_{\mu\nu}] = 0. \tag{2.2}$$

For SU(2) gauge theories the matrices T^a are given by the 2×2 Pauli matrices σ^a, $a = 1,...,3$, whereas for SU(3) gauge theories they are chosen to be the usual 3×3 Gell-Mann matrices λ^a, $a = 1,...,8$.

As discussed in an earlier paper,[19] a suitable *ansatz* for wavelike solutions of SU(2) gauge theories characterized by a propagation vector k_μ is given by

$$A_\mu = i\sigma_{\mu\nu} k^\nu f(k\cdot x), \tag{2.3}$$

where $k\cdot x = k^\mu x_\mu$. Equation (2.3) is just a special case of the CFtHW *ansatz*. The antisymmetric matrices $\sigma_{\mu\nu}$ satisfy the O(4) commutation relations and are defined as usual by

$$\sigma_{ij} = \frac{1}{4i}[\sigma_i, \sigma_j]$$

and

$$\sigma_{i4} = \tfrac{1}{2}\sigma_i \tag{2.4}$$

or

$$\sigma_{\mu\nu} = \eta_{a\mu\nu}\sigma^a/2,$$

[a)]Work supported by the Department of Energy, contract DE-AC03-76SF00515.

with

$$\eta_{a\mu\nu} = \epsilon_{a\mu\nu}, \quad a,\mu\nu = 1,2,3$$
$$= \delta_{a\mu}, \quad \nu = 4. \tag{2.5}$$

The CFtHW *ansatz* may be extended to Minkowski space [our metric is $g^{\mu\nu} = \mathrm{diag}\,(+\ -\ -\ -)$] by defining

$$\sigma_{i0} = i\sigma_i/2.$$

The set of $\sigma_{\mu\nu}$ matrices thus obtained satisfy O(3,1) commutation relations and yield complex solutions in Minkowski space. In the particular case of *ansatz* (2.3) one class of self-dual solutions with a restricted superposition property is obtained provided that $k^2 = k_\mu k^\mu = 0$. Since the function $f(k\cdot x)$ remains completely arbitrary, these solutions may be regarded as non-abelian generalizations of electromagnetic plane waves. The Euclidean space version of these solutions with $k^2 = 0$ is of course trivial.

In order to obtain SU(3) versions of these SU(2) wave solutions, we begin by defining a generalization of the O(3) symmetric *ansätze* used by Horvath and Palla[8] and Bais and Weldon.[13] As in the SU(2) case it is possible to exhibit the ansatz in either Euclidean or Minkowski space.

A. Euclidean space version

We choose the SU(3) gauge potentials to be

$$A_i = i\epsilon_{ijk}k_jL_kH(v) + i\epsilon_{ijk}\frac{k_jk_p}{k}Q_{kp}G(v) + iL_ik_4D(v) + iQ_{ij}\frac{k_jk_4}{k}E(v) + ik_jL_jk_iA(v) + iQ_{rs}\frac{k_rk_sk_i}{k}B(v), \tag{2.6a}$$

$$A_4 = -iL_ak_aC(v) - iQ_{ab}\frac{k_ak_b}{k}F(v), \tag{2.6b}$$

where $v = k_\mu x_\mu = k_1x_1 + k_2x_2 + k_3x_3 + k_4x_4$,

$k = |\mathbf{k}|$,

$(L_a)_{ij} = i\epsilon_{iaj}$,

$(Q_{ab})_{ij} = \delta_{ai}\delta_{bj} + \delta_{aj}\delta_{bi} - \frac{2}{3}\delta_{ab}\delta_{ij}, \quad i,j,a,b = 1,3$

and H, G, D, E, A, B, C, F are unknown functions of v. It can be seen that ansatz (2.6) corresponds to the most general form of the potential which can be constructed from L_a, Q_{ab}, and the vector k_μ, and reduces to the simple SU(2) embedding if

$$G = D = E = A = B = F \equiv 0.$$

L_a and Q_{ab} satisfy the commutation relations

$$[L_a, L_b] = i\epsilon_{abc}L_c, \tag{2.7a}$$
$$[L_a, Q_{bc}] = i(\epsilon_{cna}Q_{bn} + \epsilon_{bna}Q_{cn}), \tag{2.7b}$$
$$[Q_{ab}, Q_{cd}] = i(\delta_{ad}\epsilon_{bcs} + \delta_{bd}\epsilon_{acs} + \delta_{bc}\epsilon_{ads} + \delta_{ac}\epsilon_{bds})L_s. \tag{2.7c}$$

The Lorentz condition $\partial_\mu A_\mu = 0$ is satisfied if

$$D' + \frac{\mathbf{k}^2A'}{k_4} = C', \tag{2.8a}$$
$$E' + \frac{\mathbf{k}^2B'}{k_4} = F', \tag{2.8b}$$

where the prime denotes differentiation with respect to $v = k_\mu x_\mu$.

It is straightforward to calculate the field strengths from Eq. (2.1c),

$$\begin{aligned} F_{ij} = i\Big\{&(\epsilon_{jrs}k_i - \epsilon_{irs}k_j)\Big[k_rL_s\Big(H' + \frac{k_4{}^2E^2}{k^2} + k_4AD + 2k_4BE\Big) + \frac{Q_{ps}k_pk_r}{k}(G' - HG + k_4AE + 2k_4BD)\Big] \\ &- \epsilon_{ijr}k_rk_sL_s(H^2 + G^2) - \epsilon_{ijp}L_pk_4{}^2(D^2 + E^2) - 2\epsilon_{ijn}\frac{Q_{mn}k_m}{k}k_4{}^2DE + (k_iL_j - k_jL_i) \\ &\times[k_4D' - k_4(HD + 2GE) - \mathbf{k}^2(AH + 2BG)] - 2\epsilon_{ijs}k_s\frac{Q_{np}k_nk_p}{k}HG \\ &+ \Big(\frac{Q_{jm}k_mk_i}{k} - \frac{Q_{im}k_mk_j}{k}\Big)[k_4E' - k_4(2HE + DG) - \mathbf{k}^2(AG + 2BH)] \\ &- (\epsilon_{irm}Q_{mj} - \epsilon_{jrm}Q_{mi})\frac{k_r}{k}(\mathbf{k}^2HG + k_4{}^2DE)\Big\}, \end{aligned} \tag{2.9a}$$

$$\begin{aligned} F_{4i} = i\Big\{&\epsilon_{ijk}k_jL_k(H' - CD - 2EF)k_4 + \epsilon_{ijk}k_jk_p\frac{Q_{kp}}{k}k_4(G' - CE - 2FD) + L_i[k_4{}^2D' + \mathbf{k}^2(HC + 2FG)] \\ &+ \frac{Q_{ij}k_j}{k}[k_4{}^2E' + \mathbf{k}^2(2HF + CG)] + k_i\Big[L_ak_a(C' + k_4A' - HC - 2FG) \\ &+ \frac{Q_{rs}k_rk_s}{k}(F' + k_4B' - 2HF - CG)\Big]\Big\}. \end{aligned} \tag{2.9b}$$

Equations (2.9) may now be inserted into the field equations (2.2). In order to illustrate the simplifications required to obtain the field equations in a suitable form it is convenient to conisder

$$\partial_iF_{4i} + [A_i, F_{4i}] = 0. \tag{2.10}$$

From Eq. (2.9b) it follows immediately that

$$\partial_iF_{4i} = i\{L_ik_i(k_4{}^2D'' + \mathbf{k}^2C'' + \mathbf{k}^2k_4A'') + (Q_{ab}k_ak_b/k)(k_4{}^2E'') + \mathbf{k}^2F'' + k_4\mathbf{k}^2B'')\}. \tag{2.11a}$$

The Lorentz condition (2.8) may be used to simplify this expression to

$$\partial_i F_{4i} = i\left\{L_i k_i k_\mu{}^2 C'' + \frac{Q_{ab} k_a k_b}{k} k_\mu{}^2 F''\right\}, \tag{2.11b}$$

where $k_\mu{}^2 = k_4{}^2 + \mathbf{k}^2$.

Similarly, the inhomogeneous term in Eq. (2.10) becomes

$$\begin{aligned}[A_i,F_{4i}] = i\{k_j L_j[&-k_4^2(2CD^2+2CE^2+8EFD)\\ &+2DH'-2HD'-\mathbf{k}^2(2CH^2+2CG^2\\ &+8HGF)+2EG'-2GE'] + (Q_{ij}k_ik_j/k)\\ &\times[-k_4^2(6FD^2+6FE^2+6ECD)\\ &+3DG'-3GD'-\mathbf{k}^2(6FH^2+6FG^2+6GHC)\\ &+3EH'-3HE']\}.\end{aligned} \tag{2.12a}$$

Upon choosing

$$D = \begin{Bmatrix} H \\ G \end{Bmatrix} \text{and } E = \begin{Bmatrix} G \\ H \end{Bmatrix},$$

Eq. (2.12a) simplifies to

$$\begin{aligned}[A_i,F_{4i}] &= i\Big\{-k_jL_jk_\mu{}^2(2CH^2+2CG^2+8HGF)\\ &\quad -\frac{Q_{ij}k_ik_j}{k}k_\mu{}^2(6FH^2+6FG^2+6CHG)\Big\}.\end{aligned} \tag{2.12b}$$

The important point to note in Eqs. (2.11b) and (2.12b) is that they are now multiplied by a factor $k_\mu{}^2$. When these solutions are continued to Minkowski space, the class of solutions with $k_\mu{}^2=0$ will automatically satisfy the equations of motion and represent the SU(3) analogs of the self-dual propagating SU(2) wave solutions obtained in earlier work.

The field equations for the field strengths F_{ij} may be simplified by the methods described above, and after much tedious algebra, the equations of motion finally become

$$\begin{aligned} ik_\mu{}^2\Big\{&k_jL_j(C''-2H^2C-2CG^2-8HGF)\\ &+\frac{Q_{ij}k_ik_j}{k}(F''-6FH^2-6FG^2-6CHG)\Big\}=0\end{aligned} \tag{2.13a}$$

and

$$\begin{aligned} ik_\mu{}^2\Bigg\{&\left(\epsilon_{jap}k_aL_p + \begin{Bmatrix} k_4L_j \\ Q_{jn}k_nk_4/k \end{Bmatrix}\right)\\ &\times[H''-H^3-7HG^2-4GFC-4HF^2-HC^2]\\ &+\left(\epsilon_{jan}\frac{k_ak_sQ_{sn}}{k} + \{Q_{jn}k_nk_4/k_4L_j\}\right)\\ &\times[G''-G^3-7H^2G-4CHF-4GF^2-GC^2]\\ &+\frac{k_4}{k^2}k_jL_ak_a\left[C''-\begin{Bmatrix} H'' \\ G'' \end{Bmatrix}-2H^2C-2G^2C-8HFG\right.\\ &\left.+\begin{Bmatrix} H^3+7HG^2+4GFC+4HF^2+HC^2 \\ G^3+7H^2G+4CHF+4GF^2+GC^2 \end{Bmatrix}\right]\\ &+\frac{k_jQ_{im}k_ik_mk_4}{k^3}\left[F''-\begin{Bmatrix} H'' \\ G'' \end{Bmatrix}-6FH^2-6FG^2-6GHC\right.\\ &\left.+\begin{Bmatrix} H^3+7HG^2+4GFC+4HF^2+HC^2 \\ G^3+7H^2G+4CHF+4GF^2+GC^2 \end{Bmatrix}\right]\Bigg\}=0,\end{aligned} \tag{2.13b}$$

where the terms in braces correspond to choosing

$$D = \begin{Bmatrix} H \\ G \end{Bmatrix},\ E = \begin{Bmatrix} G \\ H \end{Bmatrix}.$$

Hence solutions of Eq. (2.13) are obtained if

(i) $C''-(2CH^2+2CG^2+8HGF)=0,$ (2.14a)

(ii) $F''-6(CHG+FH^2+FG^2)=0,$ (2.14b)

(iii) $H''-(H^3+7HG^2+HC^2+4GCF+4HF^2)=0,$ (2.14c)

(iv) $G''-(G^3+7H^2G+GC^2+4HCF+4GF^2)=0,$ (2.14d)

or

$$k_\mu{}^2=0. \tag{2.15}$$

The trivial SU(2) embedding is recovered by setting

$H=C,$

$G=F=0.$

B. Minkowski space version

By analogy with the SU(2) case, the *ansatz* for the Minkowski space version of the above solutions may be written as

$$\begin{aligned}A_i &= i\epsilon_{ijk}k^jL_kH + i\epsilon_{ijk}\frac{k^jk^pQ_{kp}G}{k} - L_ik^oD - \frac{Q_{ij}k^jk^oE}{k}\\ &\quad + k^jL_jk_iA + \frac{Q_{rs}k^rk^sk_iB}{k},\end{aligned} \tag{2.16a}$$

$$A_o = L_ak^aC + Q_{ab}\frac{k^ak^b}{k}F. \tag{2.16b}$$

Inserting *ansatz* (2.16) into Eq. (2.1c) yields the field strengths:

$$\begin{aligned}F_{oi} = \Big\{&i\epsilon_{ijk}k^jL_k(H'+CD+2EF)k^o + i\epsilon_{ijk}\\ &\times\frac{k^jk^pQ_{kp}k^o}{k}(G'+CE+2FD)\\ &+L_i[-k_o{}^2D'-\mathbf{k}^2(HC+2FG)]\\ &+Q_{ij}\frac{k^j}{k}[-k_o{}^2E'-\mathbf{k}^2(CG+2HF)]\\ &+k_i\Big[L_jk^j(k_oA'-C'-HC-2FG)\\ &+\frac{Q_{ab}k^ak^b}{k}(k_oB'-F'-CG-2HF)\Big]\Big\}\end{aligned} \tag{2.17a}$$

and

$$\begin{aligned}F_{ij} = \Big\{&(i\epsilon_{jrs}k_i - i\epsilon_{irs}k_j)\Big[L_sk^r\Big(H'+\frac{k_o{}^2}{k^2}E^2\\ &+k^oAD+2k^oBE\Big)+Q_{ps}\frac{k^pk^r}{k}\end{aligned}$$

$$\times (G' + HG + k^oAE + 2k^oBD)\Big]$$
$$- i\epsilon_{ijr} k^r L_s k^s (H^2 + G^2) + i\epsilon_{ijp} L_p k_o^2 (D^2 + E^2)$$
$$+ 2i\epsilon_{ijn} \frac{Q^{mn} k_m}{k} k_o^2 DE - 2i\epsilon_{ijs} k^s Q_{np} \frac{k^n k^p}{k} HG$$
$$+ (k_i L_j - k_j L_i)[- k^o D' - k^o (HD + 2GE)$$
$$- \mathbf{k}^2 (AH + 2BG)] + \left(Q_{jm} \frac{k^m k_i}{k} - Q_{im} \frac{k^m k_j}{k}\right)$$
$$\times [- k^o E' - k^o (GD + 2HE) - \mathbf{k}^2 (AG + 2BH)]$$
$$- (\epsilon_{irm} Q_{mj} - \epsilon_{jrm} Q_{mi}) \frac{k^r}{k} (i\mathbf{k}^2 HG - ik_o^2 DE)\Big\}.$$

Substituting these expressions into the field equations it is found that just as in the SU(2) case, solutions are obtained if Eqs. (2.14) or (2.15) are satisfied, where now $k_\mu^2 = k_o^2 - \mathbf{k}^2$. Hence the fields (2.17), with the functions C, F, H, and G remaining arbitrary and depending on a propagation vector k_μ such that $k_\mu^2 = 0$, describe the SU(3) analog of SU(2) nonabelian plane waves. Clearly for these solutions just as in the SU(2) case, it is possible to superpose gauge fields of the form (2.16) and still obtain a solution of the field equatons.

As expected, the properties of SU(3) plane wave solutions do not differ significantly from those for SU(2) gauge theories.

The important result of this section is not the investigation of these properties, but rather the construction of the *ansatz* (2.6) or (2.16) in such a way as to ensure that the equations of motion reduce to expressions multiplied by an overall factor of k_μ^2.

III. SELF-DUALITY PROPERTIES OF THE SU(3) WAVE SOLUTIONS

It is interesting to examine the consequence of demanding that the field strengths (2.9) satisfy the self-duality condition. For convenience the Euclidean space solutions are considered; however, analogous results may be derived for the Minkowski space solutions.

To obtain the duals of the field strengths (2.9) it is much more convenient to express them in a covariant form. Accordingly, the gauge fields (2.6) are written as

$$A_\mu = i\tau_{\mu\nu} k_\nu H + i\xi_{4\mu\alpha\beta} k_\alpha k_\beta \frac{G}{k} + i\xi_{44\alpha\beta} \frac{k_\alpha k_\beta}{kk_4} K_\mu B$$
$$+ i\tau_{4\nu} k_\nu K_\mu \frac{A}{k_4}, \tag{3.1}$$

where

$$K_\mu = \delta_{\mu 4} k_\gamma k_\gamma - k_\mu k_4, \tag{3.2a}$$
$$k = |\mathbf{k}|, \tag{3.2b}$$
$$\tau_{\mu\nu} = \eta_{a\mu\nu} L_a, \quad a = 1,...,3, \tag{3.2c}$$
$$\xi_{\mu\nu\alpha\beta} = -\eta_{a\mu\alpha}\eta_{b\nu\beta} Q_{ab}, \quad a,b = 1,...,3 \tag{3.2d}$$

and $\eta_{a\mu\nu}$ is the 't Hooft tensor defined by Eq. (2.5).

The tensors $\tau_{\mu\nu}$ are seen to correspond to the embedding of the SU(2) tensors $\sigma_{\mu\nu}$ inside SU(3) which yields instantons with charges of $q = \pm 4$. The tensors $\xi_{\mu\nu\alpha\beta}$ have no analog in SU(2). With the aid of the somewhat cumbersome commutation relations given in Appendix A, the field strengths may be evaluated:

$$F_{\mu\nu} = i\Big\{(\tau_{\nu\rho} k_\mu - \tau_{\mu\rho} k_\nu) k_\rho (H' - H^2 - G^2)$$
$$+ \tau_{\nu\mu} k_\gamma k_\gamma (H^2 + G^2) + (\tau_{\nu\rho} K_\mu - \tau_{\mu\rho} K_\nu) \frac{K_\rho}{k_4}$$
$$\times \left(AH + 2GB + \frac{G^2 k_4}{\mathbf{k}^2}\right) + \frac{1}{k_4}(K_\nu k_\mu - k_\nu K_\mu)\tau_{4\rho} k_\rho$$
$$\times \left(A' - AH - 2GB - \frac{G^2 k_4}{\mathbf{k}^2}\right) + (\xi_{4\nu\alpha\beta} k_\mu - \xi_{4\mu\alpha\beta} k_\nu)$$
$$\times \frac{k_\alpha k_\beta}{k}(G' - 3GH) - 3\xi_{4\mu\alpha\nu} \frac{k_\alpha}{k} GHk_\gamma k_\gamma$$
$$+ (K_\nu k_\mu - K_\mu k_\nu)\xi_{44\alpha\beta} \frac{k_\alpha k_\beta}{kk_4}(B' - GA - 2HB)$$
$$+ (\xi_{4\nu\alpha\beta} K_\mu - \xi_{4\mu\alpha\beta} K_\nu) \frac{K_\beta k_\alpha}{kk_4}(GA + 2HB)\Big\}. \tag{3.3}$$

It is trivial to check that Eq. (3.3) is identical to the field strengths (2.9) for $D = H$ and $E = G$.

The dual of Eq. (3.3), ${}^*F_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta} F_{\alpha\beta}$, may now be calculated using the formulas given in Appendix B:

$${}^*F_{\mu\nu} = i\Big\{- (\tau_{\nu\alpha} k_\mu - \tau_{\mu\alpha} k_\nu) k_\alpha (H' - H^2 - G^2)$$
$$+ \tau_{\nu\mu} k_\gamma k_\gamma H' - (\tau_{\nu\alpha} K_\mu - \tau_{\mu\alpha} K_\nu) \frac{K_\alpha}{k_4}$$
$$\times \left(AH + 2GB + \frac{k_4 G^2}{\mathbf{k}^2}\right) - \tau_{\mu\nu} k_\gamma k_\gamma \frac{\mathbf{k}^2}{k_4}$$
$$\times \left(AH + 2GB + \frac{G^2 k_4}{\mathbf{k}^2}\right) + \epsilon_{\mu\nu\alpha 4} k_\alpha k_\gamma k_\gamma \frac{\tau_{4\rho} k_\rho}{k_4}$$
$$\times \left(A' - AH - 2GB - \frac{G^2 k_4}{\mathbf{k}^2}\right)$$
$$- (\xi_{4\nu\gamma\delta} k_\mu - \xi_{4\mu\gamma\delta} k_\nu)$$
$$\times \frac{k_\gamma k_\delta}{k}(G' - 3GH) - \xi_{4\mu\gamma\nu} \frac{k_\gamma G'}{k}$$
$$- \xi_{4\mu\gamma\nu} \frac{k_\gamma \mathbf{k}^2}{kk_4}(GA + 2HB) k_\alpha k_\alpha$$
$$- (\xi_{4\nu\gamma\delta} K_\mu - \xi_{4\mu\gamma\delta} K_\nu) \frac{K_\delta k_\gamma}{kk_4}(GA + 2HB)$$
$$+ \epsilon_{\mu\nu\alpha 4} k_\alpha k_\gamma k_\gamma \xi_{44\alpha\beta} \frac{k_\alpha k_\beta}{kk_4}(B' - GA - 2HB)\Big\}. \tag{3.4}$$

Comparing Eqs. (3.3) and (3.4) it is easy to see that the anti-self-duality condition ${}^*F_{\mu\nu} = -F_{\mu\nu}$, is satisfied if

$$k_\gamma k_\gamma = 0 \tag{3.5}$$

or

$$\text{(i)} \quad H' + H^2 + 2G^2 + \frac{\mathbf{k}^2}{k_4}(AH + 2GB) = 0, \tag{3.6a}$$
$$\text{(ii)} \quad A' - (AH + 2GB + G^2 k_4/\mathbf{k}^2) = 0, \tag{3.6b}$$
$$\text{(iii)} \quad G' + 3GH + \frac{\mathbf{k}^2}{k_4}(GA + 2HB) = 0, \tag{3.6c}$$
$$\text{(iv)} \quad B' - GA - 2HB = 0. \tag{3.6d}$$

It is not difficult to check that the self-duality equations (3.6) imply the equations of motion (2.14), by remembering the Lorentz condition for the ansatz (3.1) is satisfied if

$$H + \frac{\mathbf{k}^2 A}{k_4} = C,$$

$$G + \frac{\mathbf{k}^2 B}{k_4} = F.$$

Hence, as expected the $k_\mu{}^2 = 0$ solutions for SU(3) gauge theories are anti-self-dual, just like their SU(2) analogs. The system of Eqs. (3.6) gives a set of first order equations for wavelike SU(3) solutions, which although simpler than the corresponding equations of motion, are still not trivial to solve.

IV. REMARKS

(i) The form of the *ansatz* (3.1) is very suggestive. With a suitable choice of the functions G, B, and A it reduces to an embedding of the CFtHW *ansatz* inside SU(3). From a generalization of Eq. (3.1), a possible candidate for an SU(3) version of the CFtHW *ansatz* may be written as

$$A_\mu = i\tau_{\mu\nu}\partial_\nu \ln h + i\xi_{\mu\rho\alpha\beta}\left[(\partial_\rho \ln C)(\partial_{\alpha\beta}\ln f) - (\partial_\rho \ln f)(\partial_{\alpha\beta}\ln C)\right], \qquad (4.1)$$

where h, f, and C are some superpotentials. It is not difficult to show Eq. (4.1) satisfies the Lorentz condition.

The field strengths and their duals resulting from ansatz (4.1) have been evaluated. The algebra is rather involved and unfortunately applying the self-duality condition does not lead to the nice simplification which occurs for SU(2). However, it is still possible that (4.1) results in some simplification of the equations of motion, so further investigation of this ansatz is indicated.

(ii) We have recently obtained the most general self-dual SU(2) plane wave solutions[20] by the use of Yang's R-gauge equations.[21] Yang's formulation has also been extended to the gauge group SU(3).[22] Just as in the SU(2) case, it is not difficult to see that the most general self-dual SU(3) plane wave solutions may be obtained by simply requiring that the functions used in the R-gauge *ansatz* be dependent on the Lorentz scalar $k{\cdot}x$.

ACKNOWLEDGMENT

I would like to thank Professor B. H. J. McKellar for critical reading of the manuscript and to acknowledge the hospitality of the Theory Group at SLAC, where part of this work was completed. I am also grateful for the financial assistance of an Australian Postgraduate Research Award and a University of Melbourne Travelling Scholarship.

APPENDIX A

The commutation relations for the tensors $\tau_{\mu\nu}$ and $\xi_{\mu\nu\rho\sigma}$ as defined by Eqs. (3.2c)–(3.2d) are given by

$$[\tau_{\mu\nu}, \tau_{\rho\delta}] = i(\tau_{\mu\rho}\delta_{\nu\delta} - \tau_{\mu\delta}\delta_{\nu\rho} + \tau_{\nu\delta}\delta_{\mu\rho} - \tau_{\nu\rho}\delta_{\mu\delta}), \qquad (A1)$$

$$\begin{aligned}[\xi_{\mu\nu\alpha\beta}, \tau_{\rho\sigma}] &= i(\xi_{\mu\nu\alpha\rho}\delta_{\beta\sigma} - \xi_{\mu\nu\alpha\sigma}\delta_{\beta\rho} + \xi_{\mu\beta\alpha\sigma}\delta_{\nu\rho} - \xi_{\mu\beta\alpha\rho}\delta_{\nu\sigma} \\ &\quad + \xi_{\nu\mu\beta\rho}\delta_{\alpha\sigma} - \xi_{\nu\mu\beta\sigma}\delta_{\alpha\rho} + \xi_{\nu\alpha\beta\sigma}\delta_{\mu\rho} - \xi_{\nu\alpha\beta\rho}\delta_{\mu\sigma}), \qquad (A2)\end{aligned}$$

$$\begin{aligned}[\xi_{\mu\nu\alpha\beta}, \xi_{\rho\sigma\gamma\delta}] &= (\delta_{\mu\sigma}\delta_{\alpha\delta} - \delta_{\mu\delta}\delta_{\alpha\sigma})[\tau_{\nu\beta}, \tau_{\rho\gamma}] \\ &\quad + (\delta_{\nu\sigma}\delta_{\beta\delta} - \delta_{\nu\delta}\delta_{\beta\sigma})[\tau_{\mu\alpha}, \tau_{\rho\gamma}] \\ &\quad + (\delta_{\nu\rho}\delta_{\beta\gamma} - \delta_{\nu\gamma}\delta_{\beta\rho})[\tau_{\mu\alpha}, \tau_{\sigma\delta}] \\ &\quad + (\delta_{\mu\rho}\delta_{\alpha\gamma} - \delta_{\mu\gamma}\delta_{\alpha\rho})[\tau_{\nu\beta}, \tau_{\sigma\delta}]. \qquad (A3)\end{aligned}$$

APPENDIX B

Using the well-known identity

$$\epsilon_{\nu k\sigma\delta}\eta_{\alpha\mu k} = \eta_{\alpha\nu k}\delta_{\mu\sigma} + \eta_{\alpha\sigma\nu}\delta_{\mu k} + \eta_{\alpha k\sigma}\delta_{\mu\nu}, \qquad (B1)$$

the duals of the tensors $\tau_{\mu\nu}$ and $\xi_{\mu\nu\alpha\beta}$ may be evaluated:

$$\tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}\tau_{\beta\rho} = \tau_{\nu\mu}\delta_{\alpha\rho} + \tau_{\alpha\nu}\delta_{\rho\mu} + \tau_{\mu\alpha}\delta_{\rho\nu}, \qquad (B2)$$

$$\tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}\xi_{\beta\rho\gamma\delta} = -\tfrac{1}{2}(\xi_{\mu\rho\nu\delta}\delta_{\gamma\alpha} + \xi_{\alpha\rho\mu\delta}\delta_{\gamma\nu} + \xi_{\nu\rho\alpha\delta}\delta_{\gamma\mu}), \qquad (B3)$$

$$\tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}\xi_{\beta\rho\alpha\delta} = -\xi_{\mu\rho\nu\delta}, \qquad (B4)$$

$$\begin{aligned}\tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}\xi_{\alpha\sigma\delta\beta} &= -\tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}\xi_{\beta\sigma\delta\alpha} \\ &= -\tfrac{1}{4}\epsilon_{\mu\nu\alpha\beta}\xi_{\sigma\beta\delta\alpha} = \tfrac{1}{2}\xi_{\mu\sigma\nu\delta}. \qquad (B5)\end{aligned}$$

The following formulas are also useful:

$$\tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}(\tau_{\beta\rho}k_\alpha - \tau_{\alpha\rho}k_\beta)k_\rho = -(\tau_{\nu\alpha}k_\mu - \tau_{\mu\alpha}k_\nu)k_\alpha + \tau_{\nu\mu}k_\gamma k_\gamma, \qquad (B6)$$

$$\begin{aligned}&\tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}(\xi_{4\beta\gamma\delta}k_\alpha - \xi_{4\alpha\gamma\delta}k_\beta)k_\gamma k_\delta \\ &\quad = -(\xi_{4\nu\gamma\delta}k_\mu - \xi_{4\mu\gamma\delta}k_\nu)k_\gamma k_\delta - \xi_{4\mu\gamma\nu}k_\gamma k_\delta k_\delta. \qquad (B7)\end{aligned}$$

[1] W. Marciano and H. Pagels, Phys. Rev. D **12**, 1093–1095 (1975).
[2] A. C. T. Wu and T. T. Wu, J. Math. Phys. **15**, 53–54 (1974).
[3] A. Chakrabarti, Ann. Inst. Henri Poincaré, **23**, 235–249 (1975).
[4] E. Corrigan, D. I. Olive, D. B. Fairlie, and J. Nuyts, Nucl. Phys. B **106**, 475–492 (1976).
[5] A. Sinha, Phys. Rev. D **14**, 2016–2022 (1976).
[6] Z. Horvath and L. Palla, Nucl. Phys. B **116**, 500–524 (1976).
[7] R. Kerner, Phys. Lett. B **81**, 21–24 (1979); D Wilkinson and F. Alexander Bais, Phys. Rev. D **19**, 2410–2415 (1979); A. Czechowski, CERN-TH-2282, 1977 (unpublished).
[8] Z. Horvath and L. Palla, Phys. Rev. D **14**, 1711–1714 (1976).
[9] A. Chakrabarti, Nucl. Phys. B **101**, 159–172 (1975); L. Jacobs, Phys. Rev. D **14**, 1739–1746 (1976).
[10] F. Wilczek, Phys. Lett. B **65**, 160–162 (1976); C. W. Bernard, N. H. Christ, A. H. Guth, and E. J. Weinberg, Phys. Rev. D **16**, 2967–2977 (1977).
[11] W. Marciano, H. Pagels, and Z. Parsa, Phys. Rev. D **15**, 1044–1049 (1977).
[12] R. G. Yates, Nucl. Phys. B **123**, 333–338 (1977).
[13] F. Alexander Bais and H. Arthur Weldon, Phys. Rev. D **18**, 561–566 (1978).
[14] E. Witten, Phys. Rev. Lett. **38**, 121–124 (1977).
[15] J. Cervero, Z. Horvath, and L. Palla, University of Salamanca Preprint 54/79 (unpublished).
[16] J. Z. Imbrie, Lett. Math. Phys. **2**, 483–492 (1978).
[17] E. Elizade, Phys. Rev. D **19**, 1798–1801 (1979).
[18] E. Corrigan and D. B. Fairlie, Phys. Lett. B **67**, 69–71 (1977); G. 't Hooft (unpublished); F. Wilczek in *Quark Confinement and Field Theory*, edited by D. Stump and D. Weingarten (Wiley, New York, 1979).
[19] E. Kovacs and S. Y. Lo, Phys. Rev. D **19**, 3649–3652 (1979).
[20] S. Y. Lo, P. Desmond, and E. Kovacs, Phys. Lett. B **90**, 419–424 (1980).
[21] C. N. Yang, Phys. Rev. Lett. **38**, 1377–1379 (1977).
[22] Prasad, Phys. Rev. D **17**, 3243–3246 (1978).

Supermanifolds and BRS transformations

L. Bonora, P. Pasti, and M. Tonin
Istituto di Fisica dell'Università di Padova, Istituto Nazionale di Fisica Nucleare—Sezione di Padova, Padua, Italy

(Received 20 April 1981; accepted for publication 18 September 1981)

We use a supermanifold formalism to support a geometrical interpretation of a recently proposed superfield formulation of BRS (Becchi–Ronet–Stora) and anti-BRS transformations.

PACS numbers: 11.10.Np, 11.30.Pb

1. INTRODUCTION

Since the discovery of the Becchi–Ronet–Stora (BRS) transformations as symmetry transformations of a quantized gauge theory, it was clear that they could be interpreted somehow as supersymmetric transformations. In Ref. 1 a new superfield formulation of a quantized gauge theory has been proposed. It exploits both BRS and anti-BRS symmetry[2] and it looks compact and effective in supplying a prescription for constructing the most general Lagrangian. The key to the superfield construction is the transformation

$$\underline{A}_\mu dx^\mu \rightarrow U^{-1}(x,\theta,\bar\theta)\underline{A}_\mu(x)dx^\mu U(x,\theta,\bar\theta) + U^{-1}(x,\theta,\bar\theta)\underline{d}U(x,\theta,\bar\theta), \tag{1.1}$$

where θ and $\bar\theta$ are "anticommuting variables" ($\theta^2 = \bar\theta^2 = 0$, $\theta\bar\theta = -\bar\theta\theta$) and $\underline{d} = \frac{\partial}{\partial x_\mu}dx_\mu + \frac{\partial}{\partial\theta}d\theta + \frac{\partial}{\partial\bar\theta}d\bar\theta$.

Moreover

$$\begin{aligned} U(x,\theta,\bar\theta) &= \exp\{\theta\underline{\bar c}(x) + \bar\theta\underline{c}(x) + \theta\bar\theta\,[\underline{B}(x) + \tfrac12(\underline{c}(x)\underline{\bar c}(x) + \underline{\bar c}(x)\underline{c}(x))]\} \\ &= 1 + \theta\underline{\bar c}(x) + \bar\theta\underline{c}(x) + \theta\bar\theta\,[\underline{B}(x) + \underline{c}(x)\underline{\bar c}(x)]. \end{aligned} \tag{1.2}$$

Underlined quantities are matrix-valued fields: $\underline{A}_\mu(x) = A^i_\mu(x)\tau^i$, etc., where $\tau^i (i = 1,...,N)$ are the generators of the Lie algebra $\mathscr{G}$ of some classical Lie group G; $A^i_\mu(x)$ are the gauge potentials, $c^i(x)$ and $\bar c^i(x)$ the Faddeev–Popov (FP) fields, and $B^i(x)$ the auxiliary fields. The transformation (1.1) generates a one-form

$$\underline{\phi}_\mu(x,\theta,\bar\theta)dx^\mu + \underline{\eta}(x,\theta,\bar\theta)d\bar\theta + \underline{\bar\eta}(x,\theta,\bar\theta)d\theta \tag{1.3}$$

in the superspace Σ with local coordinates $(\{x_\mu\},\theta,\bar\theta)$. From the explicit expression of the superfields $\underline{\phi}_\mu$, $\underline{\eta}$, and $\underline{\bar\eta}$, the BRS and anti-BRS transformations appear as translations in $\bar\theta$ and θ, respectively.

In Ref. 3 a geometrical interpretation of the transformation (1.1) has been proposed, resorting to an unusual geometrical structure. Here we present a different and very simple geometrical interpretation. Equation (1.1) is understood just as it looks: a gauge transformation on a connection form defined in the superspace Σ. A difficulty arises when we look for the structure group of the principal fiber bundle whose base space should be this superspace. It cannot be G itself, because $U(x,\theta,\bar\theta)$ is never a pure real (complex) matrix when θ or $\bar\theta$ are different from zero.

The clue for solving this difficulty is provided by Rogers' supermanifold theory[4,5] (in this context see also Ref. 6). In Rogers' approach supermanifolds are spaces locally homeomorphic to "generalized Euclidean spaces," called Grassmann algebras, whose coordinates are both odd Grassmann variables (like θ and $\bar\theta$) and even Grassmann variables replacing the usual real coordinates. In our case if we interpret $\{x_\mu\}$ and the group parameters in G as even Grassmann variables, the above difficulty disappears. Then the underlying geometrical structure is a (super) principal fiber bundle whose base space is a supermanifold which, according to Rogers' approach, supplies a suitable refined version of the superspace Σ, and whose structure group is the Grassmann enlargement of G.

This interpretation automatically accounts for the anticommutativity of the fields $c^i(x)$ and $\bar c^i(x)$. It is interesting to observe that replacing real variables by even Grassmann variables, which looks at first sight like a useless complication from a practical point of view, allows us to set Eq. (1.1) and the BRS transformations in a very simple geometrical framework.

The paper is organized as follows. In Sec. 2 we derive the necessary mathematical results. Our approach is to link tightly real or complex geometrical objects (such as manifolds, groups,...) to their Grassmann enlargements, in such a way that operations defined on them (such as maps, products,...) work in exactly parallel ways. Our approach is constructive: for every real or complex geometrical structure we show that there is a Grassmann enlargement and build it. So the result need not be unique. In Sec. 3 we explain in detail the geometrical interpretation of Eq. (1.1). In Sec. 4 we deal with the analogous problem for the matter fields.

2. THE G FORMALISM

In this section we derive some results about G^∞ manifolds not contained in Ref. 4. First we recall a few fundamental concepts about the G^∞ formalism. For conventions and proofs see Ref. 4.

Consider a Grassmann algebra B_L over R^L.[7] We shall choose L finite, until Sec. 2E. In Sec. 2F we shall turn to $L = \infty$ (see also Sec. 3). B_L can be given a Banach space topology. Consider $B_L^{(0)}$ and $B_L^{(1)}$, the even and odd part of B_L, respectively. A parameter or variable taking its values in $B_L^{(0)}$ ($B_L^{(1)}$) will be called an even (odd) Grassmann parameter or variable. Now let us construct the Banach space

$$B_L^{m,n} = \underbrace{B_L^{(0)}\times\cdots\times B_L^{(0)}}_{m\ \text{times}}\times\underbrace{B_L^{(1)}\times\cdots\times B_L^{(1)}}_{n\ \text{times}}.$$

The "body" map $\epsilon: B_L^{m,n}\rightarrow R^m$ associates to each $(m+n)$-tuple $(\bar x_1,..., \bar x_m, \bar y_1,..., \bar y_n)\in B_L^{m,n}$ the m-tuple $(x_1, x_2,..., x_m)$, where x_i is the (real) coefficient of the trivial

 0022-2488/82/050839-07$02.50

Grassmann generator in $\bar{x}_i$. The "soul map" is defined by $s = id - \epsilon$. The definition of a G^∞ function is the fundamental definition in Ref. 4. G^∞ differentiability is more restrictive that C^∞ differentiability in the usual Banach sense and assures the existence and unicity of the Grassmann continuation for C^∞ functions.

Let U be an open set in $B_L^{m,n}$ and V be open in R^m with $V = \epsilon(U)$. The Z-continuation[4] or Grassmann continuation of a function $f \in C^\infty(V, B_L)$ is defined by

$$z(f)(\bar{x}_1,\ldots,\bar{x}_m,\bar{y}_1,\ldots,\bar{y}_n) = \sum_{k_1=\cdots=k_m=0}^{L} \frac{1}{k_1!\cdots k_m!} \times [\partial_1^{k_1}\cdots\partial_m^{k_m} f](x_1,\ldots,x_m) s(\bar{x}_1)^{k_1}\cdots s(\bar{x}_m)^{k_m}, \tag{2.1}$$

where, according to our definitions, $s(\bar{x}_i) = \bar{x}_i - \epsilon(\bar{x}_i) = \bar{x}_i - x_i$. Then $z(f) \in G^\infty(U, B_L)$. It is evident that the continuation (2.1) involves only even variables. Let $f, g \in C^\infty(V, B_L)$, then it is easy to show that

$$z(fg) = z(f)\, z(g). \tag{2.2}$$

Let f: open set in $R^m \to B_L^{k,0}$, be represented by the set of C^∞ functions $\{f_i(x_i,\ldots,x_m), i = 1,\ldots,m\}$. Define $z(f)$ as the set of functions $\{z(f_i)\}$. Now let $f: U \to V$, where U is open in R^m and V is open in R^n, and $g: V' \to B_L^{k,0}$, where $V' \subseteq V$, and both f and g are C^∞ functions. Then

$$z(g \circ f) = z(g) \circ z(f). \tag{2.3}$$

A. G^∞ Manifolds

A G^∞(m,n) manifold Y_s is a Hausdorff space locally homeomorphic to $B_L^{m,n}$. The exact definition[4] parallels the usual C^∞ manifold definition, with C^∞ functions replaced by G^∞ ones. Precisely if $\{\bar{U}_\alpha\}$ is an open covering of Y_s and $\{\bar{\psi}_\alpha : \bar{U}_\alpha \to$ open subset of $B_L^{m,n}, \alpha \in A\}$ is the relative family of homeomorphisms, then $\bar{\psi}_\alpha \circ \bar{\psi}_\beta^{-1} : \bar{\psi}_\beta(\bar{U}_\alpha \cap \bar{U}_\beta) \to \bar{\psi}_\alpha(\bar{U}_\alpha \cap \bar{U}_\beta)$ must be a G^∞ function for any $\alpha, \beta \in A$.

Given a $G^\infty(m,n)$ manifold Y_s consider the equivalence relation among points of $Y_s : \bar{x} \sim \bar{y}$ if and only if for some $\alpha \in A, \epsilon(\bar{\psi}_\alpha(\bar{x})) = \epsilon(\bar{\psi}_\alpha(\bar{y}))$. It has been shown in Ref. 4 that the corresponding quotient space is an m-dimensional C^∞ manifold. Let us call the quotient map, or "body" map, Φ_s.

A problem we are faced with, in this paper, is whether, given an m-dimensional C^∞ manifold M, we are able to define a $G^\infty(m,n)$ manifold $M_{(S)}$, such that the quotient manifold just defined is exactly M. Let us first consider the problem of how to define a $G^\infty(m,0)$ manifold $M_{(G)}$, starting with an m manifold M. A simple solution is the following one.

Suppose $\{U_\alpha, \psi_\alpha | \alpha \in A\}$ is an atlas for M. For every $\alpha \in A$ consider the subset $\bar{U}_\alpha$ of the Cartesian product $U_\alpha \times B_L^{m,0}$ defined by

$$\bar{U}_\alpha = \{(x,\bar{x}) : x \in U_\alpha, \bar{x} \in B_L^{m,0} \text{ and } \epsilon(\bar{x}) = \psi_\alpha(x)\} \tag{2.4}$$

and define $\bar{\psi}_\alpha : \bar{U}_\alpha \to B_L^{m,0}$ by $\bar{\psi}_\alpha(x,\bar{x}) = \bar{x}$ for $(x,\bar{x}) \in \bar{U}_\alpha$. It is easy to see that $\bar{\psi}_\alpha$ is a bijection and the image of $\bar{\psi}_\alpha$ is an open subset of $B_L^{m,0}$. So we can transfer the topology of this open subset into $\bar{U}_\alpha$ in such a way that $\bar{\psi}_\alpha$ is a homeomorphism.

Now construct the disjoint union $\bar{\mathfrak{M}} = \bigcup_{\alpha \in A} \bar{U}_\alpha$. It is of course a Hausdorff space. Define a relation in $\bar{\mathfrak{M}}$ by

$$(x,\bar{x}) \sim (x',\bar{x}') \text{ iff } (x,\bar{x}) \in \bar{U}_\alpha, (x',\bar{x}') \in \bar{U}_\beta \text{ and}$$
$$x = x', \quad \bar{x}' = z(\psi_\beta \circ \psi_\alpha^{-1})(\bar{x}). \tag{2.5}$$

Due to the property (2.3) and to the fact that $z(\psi_\beta \circ \psi_\alpha^{-1})$ is one-to-one the relation (2.5) is an equivalence relation. Moreover it matches the overlapping of the $\bar{U}_\alpha$'s.

Then consider the space $M_{(G)}$ equal to the space $\bar{\mathfrak{M}}$ modulo the above equilvalence relation. $M_{(G)}$ is a Hausdorff space as the canonical projection onto it is open and the relation (2.5) is closed in $\bar{\mathfrak{M}} \times \bar{\mathfrak{M}}$.[8] The $\bar{\psi}_\alpha$'s provide $M_{(G)}$ with a G^∞ differentiable structure, so that $M_{(G)}$ is a $G^\infty(m,0)$ manifold.

If we consider the "body" map Φ_M, we find that the image of $M_{(G)}$ can be identified with M. Locally, $\Phi_{M|\bar{U}_\alpha}(x,\bar{x}) = x$ for $(x,\bar{x}) \in \bar{U}_\alpha$. We find straightforwardly that

$$\Phi_M \circ \bar{\psi}_\alpha^{-1} = \psi_\alpha^{-1} \circ \epsilon \tag{2.6}$$

for $\bar{x} \in \bar{\psi}_\alpha(\bar{U}_\alpha)$.

From now on $M_{(G)}$ will be called the Grassmann[9] enlargement of M. It is evident that if $M = R^m$, $M_{(G)} = B_L^{m,0}$.

Now let $M(N)$ be a C^∞ $m(n)$ manifold, $\{U_\alpha, \psi_\alpha | \alpha \in A\}(\{V'_{\alpha'}, \rho'_{\alpha'} | , \alpha' \in A'\})$ an atlas for $M(N)$ and $M_{(G)}(N_{(G)})$ the relative enlargement defined by the atlas $\{\bar{U}_\alpha, \bar{\psi}_\alpha, \alpha \in A\}(\{\bar{V}_{\alpha'}, \bar{\rho}_{\alpha'}, \alpha' \in A'\})$. Let $\phi: M \to N$ be a C^∞ map. For any couple $(U_\alpha, V_{\alpha'})$ such that $\phi(U_\alpha) \cap V_{\alpha'} \neq \phi$ define

$$\bar{\phi}_{\alpha\alpha'} = \bar{\rho}_{\alpha'}^{-1} \circ z(\rho_{\alpha'} \circ \phi \circ \psi_\alpha^{-1}) \circ \bar{\psi}_\alpha, \tag{2.7}$$

which is defined in the set of pairs $(x,\bar{x}) \in \bar{U}_\alpha$ with $x \in U_\alpha \cap \phi^{-1}(V_{\alpha'})$. The set of $\bar{\phi}_{\alpha\alpha'}$ defines a unique function $\bar{\phi}: M_{(G)} \to N_{(G)}$, which we call the z extension of $\phi, z(\phi)$.

Indeed, for example, if $(x,\bar{x}) \sim (x,\bar{x}')$, where $(x,\bar{x}) \in \bar{U}_\alpha$, $(x,\bar{x}') \in \bar{U}_\beta$, and $x \in U_\alpha \cap \phi^{-1}(V_{\alpha'})$, then

$$\begin{aligned} \bar{\phi}_{\alpha\alpha'} x,\bar{x}) &= (\bar{\rho}_{\alpha'}^{-1} \circ z(\rho_{\alpha'} \circ \phi \circ \psi_\alpha^{-1}) \circ \bar{\psi}_\alpha)(x,\bar{x}) \\ &= (\bar{\rho}_{\alpha'}^{-1} \circ z(\rho_{\alpha'} \circ \phi \circ \psi_\alpha^{-1}))(\bar{x}) \\ &= (\bar{\rho}_{\alpha'}^{-1} \circ z(\rho_{\alpha'} \circ \phi \circ \psi_\alpha^{-1}))(z(\psi_\alpha \circ \psi_\beta^{-1})(\bar{x}')) \\ &= (\bar{\rho}_{\alpha'}^{-1} \circ z(\rho_{\alpha'} \circ \phi \circ \psi_\beta^{-1}))(\bar{x}') \\ &= (\bar{\rho}_{\alpha'}^{-1} \circ z(\rho_{\alpha'} \circ \phi \circ \psi_\beta^{-1}) \circ \bar{\psi}_\beta)(x,\bar{x}') = \bar{\phi}_{\alpha'\beta}(x,\bar{x}'), \end{aligned} \tag{2.8}$$

where $\epsilon(\bar{x}) = \psi_\alpha(x)$ and $\epsilon(\bar{x}') = \psi_\beta(x)$ due to Eq. (2.3).

An analogous matching property holds in the image space. For if $\bar{\phi}_{\alpha'\alpha}(x,\bar{x}) = (y,\bar{y})$ and $\bar{\phi}_{\beta'\alpha}(x,\bar{x}) = (y',\bar{y}')$ for $(x,\bar{x}) \in \bar{U}_\alpha$, where $(y,\bar{y}) \in \bar{V}_{\alpha'}$ and $(y',\bar{y}) \in \bar{V}_{\beta'}$, then in the same way one shows that $y = y'$ and $\bar{y}' = z(\rho_{\beta'} \circ \rho_{\alpha'}^{-1})(\bar{y})$. Likewise one can show that, given a C^∞ map $\phi: M \to N$ and a C^∞ map $\chi: N \to R$, where R is an r-dimensional C^∞ manifold, and considering the z extensions $z(\phi): M_{(G)} \to N_{(G)}$ and $z(\chi): N_{(G)} \to R_{(G)}$, the composition property

$$z(\chi \circ \phi) = z(\chi) \circ z(\phi), \tag{2.9}$$

holds, where essential use has been made of Eq. (2.3). We can also easily show that

$$z(\mathrm{id}_M) = \mathrm{id}_{M_{(G)}}, \tag{2.10}$$

where $\mathrm{id}_M(\mathrm{id}_{M_{(G)}})$ is the identify function on $M(M_{(G)})$. Now let $\phi: M \to N$ be a C^∞ diffeomorphism, then $z(\phi): M_{(q)} \to N_{(q)}$ is a G^∞ diffeomorphism. Indeed $z(\phi)$ and $z(\phi^{-1})$ are G^∞ maps and, by Eqs. (2.3) and (2.10), we have

$$z(\phi^{-1})\circ z(\phi) = \mathrm{id}_{M(G)}, \text{ etc.} \tag{2.11}$$

If $pr_{1(2)}: M \times N \to M(N)$ is the projection map onto the first (second) factor, then

$$z(pr_{1(2)}) = \overline{pr}_{1(2)}: M_{(G)} \times N_{(G)} \to M_{(G)}(N_{(G)}). \tag{2.12}$$

If $f_0: M \to N$ is the constant map $f_0(x) = y_0$ for $x \in M$ and $y_0 \in V_{\alpha'} \subset N$, where $(V_{\alpha'}, \rho_{\alpha'})$ is a chart in N, then $z(f_0)$ is the constant map from $M_{(G)}$ onto the point $(y_0, \rho_{\alpha'}(y_0)1) \in \overline{V}_{\alpha'}$, where 1 is the trivial generator in B_L.

Finally let $\Phi_M: M_{(G)} \to M$ and $\Phi_N: N_{(G)} \to N$ be the generalized "body" maps defined above. Then if $\chi: M \to N$ is any C^∞ map,

$$\phi_N \circ z(\chi) = \chi \circ \phi_M. \tag{2.13}$$

Therefore if $V \subset N$ is any open set, and we define: $V_{(G)}$ as $\phi_N^{-1}(V)$, we have

$$(\chi^{-1}(V))_{(G)} = z(\chi)^{-1}(V_{(G)}). \tag{2.14}$$

Indeed $(\chi^{-1}(V))_{(G)} = \{\bar{x} \in M_{(G)} : (\chi \circ \phi_{(M)})(\bar{x}) \in V\}$ and $z(\chi)^{-1}(V_{(G)}) = \{\bar{x} \in M_{(G)} : (\phi_N \circ z(\chi))(\bar{x}) \in V\}$, where $\bar{x}$ stands here for a generic point in $M_{(G)}$.

B. G^∞ groups

Let G be a C^∞ Lie group. Consider the enlargement $G_{(G)}$ of the manifold G, defined in the previous subsection. The question is whether $G_{(G)}$ is itself a group whose structure is tightly related to the old one. To this end consider the C^∞ product function $\mu: G \times G \to G$ and the C^∞ inverse function $i: G \to G$ and take their G^∞ z continuation $z(\mu): G_{(G)} \times G_{(G)} \to G_{(G)}$ and $z(i): G_{(G)} \to G_{(G)}$ as defined in Sec. 2A. Then we have associativity,

$$\begin{aligned} z(\mu)(z(\mu)(\bar{a},\bar{b}),\bar{c}) &= z(\mu)\circ(z(\mu)\times \mathrm{id}_{G_{(G)}})(\bar{a},\bar{b},\bar{c},) \\ &= z(\mu\circ(\mu\times\mathrm{id}_G))(\bar{a},\bar{b},\bar{c}) \\ &= z(\mu\circ(\mathrm{id}_G\times\mu))(\bar{a},\bar{b},\bar{c}) = z(\mu)(\bar{a}, z(\mu)(\bar{b},\bar{c})) \end{aligned} \tag{2.15}$$

by (2.9) and (2.10), for every triple $\bar{a},\bar{b},\bar{c},\in G_{(G)}$.

If e is the unit element in G and $\{U_\alpha, \psi_\alpha\}$ is a chart in G such that $e \in U_\alpha$, then given any element $\bar{b} \in G_{(G)}$,

$$z(\mu)|_{\overline{U}_\alpha \times G_{(G)}}((e,\psi_\alpha(e)1),\bar{b}) = z(\mu)\circ(z(f_e)\times\mathrm{id}_{G_{(G)}})(\bar{a},\bar{b}), \tag{2.16}$$

where $\bar{a}$ is an element of $G_{(G)}$. Here symbols and properties established below, Eq. (2.12), have been used. Therefore the element $\bar{e} \in G_{(G)}$, which in the chart $\overline{U}_\alpha$ has the representation $(e,\psi_\alpha(e)1) \in \overline{U}_\alpha$, is the unit element in $G_{(G)}$. Of course there is no loss of generality if we choose for a particular $\alpha, \psi_\alpha(e) = 0$. Finally we also have

$$\begin{aligned} &z(\mu)(\bar{a},z(i)(\bar{a})) \\ &\quad = z(\mu)\circ(\mathrm{id}_{G_{(G)}}\times z(i))\circ z(\Delta)(\bar{a}) \qquad (2.17) \\ &\quad = z(\mu\circ(\mathrm{id}_G\times i)\circ\Delta)(\bar{a}) = z(f_e)(\bar{a}) = \bar{e} = z(\mu)(z(i)(\bar{a}),\bar{a}), \end{aligned}$$

where $\Delta: G \to G \times G$ is the function $\Delta(a) = (a,a)$. Therefore $z(i)$ is the inverse function in $G_{(G)}$. In conclusion we may call $G_{(G)}$ a G^∞ Lie group.

Now let G be a classical matrix Lie group. G may be considered as a submanifold of R^{n^2} (if the matrices are $n \times n$), specified by a set of k ($k < n^2$) polynomial equations:

$$g_i(a_{jl}) = 0, \quad i = 1,\dots,k, \tag{2.18}$$

where a_{jl} are the matrix elements. It is natural to ask if the group $\overline{G}$ obtained by thinking of a_{jl} as Grassmann variables $\bar{a}_{jl}$ and by imposing the constraint equations

$$z(g_i)(\bar{a}_{jl}) = 0, \quad i = 1,\dots,k \tag{2.19}$$

is the same as the group $G_{(G)}$ defined above. The answer is yes. Indeed the matrix product and inverse function in $\overline{G}$ are exactly the z continuations of the corresponding functions in G, as the latters are rational functions. And this is just the reason why $\overline{G}$ specified by the conditions (2.19) is a group. We should also verify that the topological and differential structure are the same. This is done in Appendix A.

We remark that our definition of a Grassmann Lie group is based on G^∞ product and inverse functions, whereas in Ref. 5 a superanalytic product function is used to define a super-Lie group. However, if we start with an analytic Lie group G we end up with a Grassmann analytic product function. Indeed the z continuation of an analytic function is trivially a Grassmann analytic function. See Sec. 2F for further comments on this point.

C. Principal fiber bundle

Let $P(M,G)$ be a principal fiber bundle with structure group G. Let $\pi: P \to M$ be the projection. Let $\{U_\alpha\}$ be a covering of M formed by trivializing sets. That is, there exist diffeomorphisms $h_\alpha: U_\alpha \times G \to \pi^{-1}(U_\alpha)$ with the property

$$pb = \tau(p,b) = h_\alpha(x,a)b = h_\alpha(x,ab) \tag{2.20}$$

for every $b \in G$, provided that $p \in \pi^{-1}(U_\alpha)$. Here τ is the product function $\tau: P \times G \to P$.

We want to transform $P(M,G)$ into a similar mathematical object in which manifolds are replaced by G^∞ manifolds and C^∞ maps by G^∞ functions. To this end we enlarge P,M, and G as in Sec. 2A and get $P_{(G)}, M_{(G)}$, and $G_{(G)}$. As shown in the preceding subsection $G_{(G)}$ is a G^∞ Lie group. We extend the product function $\tau: P \times G \to G$ by means of $z(\tau): P_{(G)} \times G_{(G)} \to P_{(G)}$. We define also the continuation $z(\pi): P_{(G)} \to M_{(G)}$. Then by (2.9)

$$z(\pi)\circ z(\tau)(\bar{p},\bar{a}) = z(\pi)(\bar{p}) \tag{2.21}$$

for $\bar{p} \in P_{(G)}$ and $\bar{a} \in G_{(G)}$.

Let us define the continuations $z(h_\alpha): U_{\alpha(G)} \times G_{(G)} \to \pi^{-1}(U_\alpha)_{(G)}$. They are G^∞ diffeomorphisms by Eq. (2.11). Moreover by (2.12)

$$\overline{pr}_1 \circ z(h_\alpha^{-1}) = z(pr_1 \circ h_\alpha^{-1}) = z(\pi)|_{\pi^{-1}(U_\alpha)_{(G)}}. \tag{2.22}$$

By Eq. (2.14) we know that $\pi^{-1}(U_\alpha)_{(G)} = z(\pi^{-1})(U_{\alpha_{(G)}})$. It follows that $z(\pi)$ is onto, and therefore it is a G^∞ fibration. Continuing (2.20) we get,

$$\begin{aligned} z(\tau)|_{\pi^{-1}(U_{\alpha(G)})\times G_{(G)}}(\bar{p},\bar{b}) &= z(\tau)\circ(z(h_\alpha)\times\mathrm{id}_{G_{(G)}})(\bar{x},\bar{a},\bar{b}) \\ &= z(h_\alpha(\bar{x},z(\mu)(\bar{a},\bar{b}))). \end{aligned} \tag{2.23}$$

We see that the action of the group $G_{(G)}$ on $P_{(G)}$ is free. Therefore all conditions are satisfied and we may call $P_{(G)}(M_{(G)}, G_{(G)})$ a G^∞ principal fiber bundle.

D. Vector fields and forms

A common attitude in dealing with superspace is not to make a distinction between real coordinates in space-time and corresponding coordinates in the superspace. Rephrased in the language of G^∞ manifolds, we should be able to deal interchangeably with real coordinates and even Grassmann coordinates. We have seen that, with respect to enlarged manifolds $M_{(G)}$, this makes sense as for the differential structure. But of course we must verify this point also for vector fields and forms. Since $M_{(G)}$ has a far richer structure than M, some attention has to be paid to the correspondence between objects on M and objects on $M_{(G)}$. In Ref. 4, given an open set U in a G^∞ supermanifold Y_s, the graded commutative algebra $G^\infty(U)$, of G^∞ functions: $U \to B_L$ is defined. In analogy with the C^∞ case, one can define vector fields: they form a graded Lie left B_L module $D^1(U)$ and in local coordinates they have an expression which exactly parallels the usual one in the C^∞ case. Those considerations are easily extended to the dual $D_1(U)$ of $D^1(U)$, i.e., the module of forms.

When the G^∞ manifold is an $M_{(G)}$ the grading is trivial and the procedure is even easier. Let $\{\bar{U}_\alpha, \bar{\psi}_\alpha\}$ be a chart in $M_{(G)}$ and let $\{\bar{x}_i,...,\bar{x}_m\}$ be a system of Grassmann coordinates in $B_L^{m,0}$. Let (U_α, ψ_α) be the corresponding chart in M and $\{x_1,...,x_m\}$ the corresponding system of coordinates in R^m. We denote by X_i the vector field which in this chart corresponds to the partial derivatives $\partial/\partial x_i$. If $\bar{f} \in G^\infty(\bar{U}_\alpha)$ we write $z(X_i)$ to mean the vector fields defined by

$$z(X_i)\bar{f} = \left(\frac{\partial}{\partial \bar{x}_i}\bar{f}\circ\bar{\psi}_\alpha^{-1}\right)\circ\bar{\psi}_\alpha, \quad i = 1,...,m, \tag{2.24}$$

where $\partial/\partial\bar{x}_i$ is the ith G derivative in $B_L^{m,0}$ [remember that if $\bar{f} \in G^\infty(U_\alpha)$, then $\bar{f}\circ\bar{\psi}_\alpha^{-1}$ is G^∞].[4] It is interesting to observe that if $f \in C^\infty(U_\alpha)$, then

$$z(X_i)z(f) = z(X_i f). \tag{2.25}$$

To every vector field X, which in the chart (U_α, ψ_α) has the expression

$$X = \sum_{i=1}^{m} \zeta_i X_i$$

with $\zeta_i \in C^\infty(U_\alpha)$, we associate the vector field in $\bar{U}_\alpha$,

$$z(X) = \sum_{i=1}^{m} z(\zeta_i)z(X_i). \tag{2.26}$$

Conversely as $\{z(X_i)\}$ form a basis of the module $\{D^1(\bar{U}_\alpha)\}$,[4] we may write an arbitrary vector field $\bar{X} \in D^1(\bar{U}_\alpha)$ as

$$\bar{X} = \Sigma_i \bar{\zeta}_i z(X_i), \tag{2.27}$$

where $\bar{\zeta}_i \in G^\infty(\bar{U}_\alpha)$.

If U is an open set not coinciding with a chart, and $X \in D^1(U)$, we define $z(X)$ by (2.26) in every chart $\{U_\alpha, \psi_\alpha\}$ such that $U \cap U_\alpha \neq \phi$. This definition is correct provided that the coefficients $z(\zeta_i)$ satisfy the usual transformation property when we pass from one chart to another. This is an easy consequence of the transformation properties of the coefficients ζ_i. Let $\{\omega^i \in D_1(U)\}$ form a basis dual to $\{X_i\}$. In local coordinates $\{\bar{x}_i\}$ they are equal to the forms $d\bar{x}_i$. Therefore

$$z(\omega^i)z(X_j) = z(\omega^i(X_j)) = \delta^i_j. \tag{2.28}$$

Then every form $\bar{\omega} \in D_1(\bar{U})$ can be expressed locally as

$$\bar{\omega} = \Sigma_i \bar{\eta}_i z(\omega^i); \tag{2.29}$$

therefore to every form $\omega \in D_1(U)$ which in the chart (U_α, ψ_α) has the expression $\omega = \Sigma_i \eta_i \omega^i$, we associate in $U_{(G)}$ the form

$$z(\omega) = \Sigma_i z(\eta_i)z(\omega^i). \tag{2.30}$$

As a consequence of all these definitions we may express very simply the module corresponding to the Lie algebra $\mathscr{G}$ of G. If $A_i (i = 1,...,N)$ is a basis for the Lie algebra $\mathscr{G}$ of G, then $z(A_i)$ form a basis for the $B_L^{(0)}$ module $\mathscr{G}_{(G)}$ of $G_{(G)}$. The structure constants are the same.

Now let $P(M,G)$ be a principal fiber bundle (pfb) and $P_{(G)}(M_{(G)},G_{(G)})$ the corresponding G^∞ pfb. Suppose ω is as connection form in $P(M,G)$, then Eq. (2.30) defines a form in $P_{(G)}$. The question is whether $z(\omega)$ is a connection form for $P_{(G)}(M_{(G)}, G_{(G)})$. The answer is affirmative but the rather lengthy proof is deferred to Appendix B.

E. Odd Grassmann variables supermanifolds

Up to now we have been solving the problem of how to define a $G^\infty(m,0)$ manifold $M_{(G)}$, starting with an m manifold M. As explained in the Introduction, we must deal with a $G^\infty(m,2)$ manifold $M_{(s)}$. Locally it looks like $B_L^{m,2}$ with two odd Grassmann variables θ and $\bar{\theta}$. As there is no particular topological requirement with respect to these two variables, we choose the simplest solution and define $M_{(S)} = M_{(G)} \times B_L^{0,2}$. Then the "body" may $\Phi_s: M_{(S)} \to M$ is the composite map $\Phi_M \circ j$, where $j: M_{(S)} \to M_{(G)}$ is the projection onto the first factor. Unlike Φ_M, which is a C^∞ function but not a G^∞ one, j is G^∞. Therefore the problem of defining, for example, a new form in $M_{(S)}$ starting with a form in $M_{(G)}$ is easily solved by the inducing procedure.

F. Analytic Grassmann (G^ω) manifolds

As we shall see in the next section, in the specific case dealt with in this paper, as well as in other instances, a Grassmann algebra B_L, with L finite, is not enough to guarantee a suitable differentiable structure. For many purposes we need an infinite L. A Grassmann algebra B_L with $L = \infty$ has been studied in Ref. 4. Due to Proposition (3.1) (iii) of Ref. 6 we can easily define a z continuation of a function $f: V \to B_L$, where V is an open set in R^m, by means of Eq. (2.1) with $L = \infty$, provided that f is analytic in V. Then $z(f)$ is Grassmann analytic (G^ω) in the domain $\epsilon^{-1}(V)$, i.e., it can be expressed as the sum of an absolutely convergent power series in a neighborhood of any point of $\epsilon^{-1}(V)$. We call $z(f)$ the analytic z continuation of f. Equations (2.2) and (2.3) still hold.

Therefore we can repeat step by step what we have done from Secs. 2A–2D simply by replacing C^∞ functions by analytic functions, G^∞ manifolds by G^ω manifolds, and so on. It should be remarked that this implies we can get an analytic Grassmann enlargement of a manifold M only if M is itself analytic.

As a final remark we observe that it is not necessary to stick to the particular algebra B_L proposed in Ref. 4. Indeed in Ref. 6 a generalization of B_L has been proposed, which though preserving all the results of Ref. 4 assumes less stringent requirements. For example, the number of odd generators needs not be discrete. Of course we could as well have adopted the Banach–Grassmann algebras of Ref. 6.

3. THE GEOMETRICAL CONSTRUCTION

After the mathematical preliminaries of the previous section, we define a suitable geometrical framework in which the transformation (1.1) can be correctly interpreted as a gauge transformation. We start of course with a principal fiber bundle $P(M,G)$, in which the potentials $A_\mu(x)$ have a well established geometrical meaning and determine a connection form in P. M is the space-time manifold and G is a classical compact matrix Lie group, say SU(N).

We make a first step toward the super-principal fiber bundle we need, shifting from the real space-time coordinates $\{x_\mu\}$ in M to *even* Grassmann variables and from real parameters in the group G to even Grassmann parameters. There is a well defined way of doing it. As we have shown in Sec. 2, we can enlarge any m manifold M to a G^∞ or $G^\omega(m,0)$ manifold $M_{(G)}$ in which the local coordinates are m even Grassmann coordinates. In the same way we can enlarge the group G and get a corresponding G^α Lie group $G_{(G)}$ (where α means either ∞ or ω). In the particular case in which G is a matrix group, this enlargement amounts to replacing the real matrix elements by even Grassmann parameters.

Using these new objects we can define a G^α principal fiber bundle, in which manifolds are replaced by G^α manifolds, C^α maps by G^α ones, as is shown in Sec. 2C. In the present case we get a G^α principal fiber bundle $P_{(G)}(M_{(G)}, G_{(G)})$. As is shown in Sec. 2D, though $P_{(G)}$ has a richer structure than P, we have a well defined prescription for passing from a connection ω in $P(M,G)$ to a connection $z(\omega)$ in $P_{(G)}$.

Equation (1.1) requires all relevant objects to be defined with respect to a base manifold in which the local coordinates are the usual space-time coordinates $\{x_\mu\}$ plus two anticommuting variables θ and $\bar\theta$. The simplest way of doing this is to choose the supermanifold $M_{(S)} = M_{(G)} \times B_L^{(0,2)}$, so that $\{x_\mu\}$ are replaced by even Grassmann variables $\{\bar x_\mu\}$, and θ and $\bar\theta$ are to be interpreted as odd Grassmann variables. We can go back from $M_{(S)}$ to $M_{(G)}$ by a simple projection j, which is of course a G^α map. For example, when $M = R^m$, $M_{(S)} = B_L^{(m,2)}$.

Now it is straightforward to construct a super principal fiber bundle $P_{(S)}(M_{(S)}, G_{(G)})$ by the usual induction procedure.[10] Let $P_{(S)}$ coincide with the subset of $M_{(S)} \times P_{(G)}$ consisting of ordered pairs $(\bar y, \bar p)$ such that $j(\bar y) = z(\pi)(\bar p)$ for every $\bar y \in M_{(S)}$ and $\bar p \in P_{(G)}$, where $z(\pi)$ is the fibration map in $P_{(G)}(M_{(G)}, G_{(G)})$. The group action is defined by $(\bar y, \bar p)\bar a = (\bar y, \bar p \bar a)$ for $\bar a \in G_{(G)}$ and $(\bar y, \bar p) \in P_{(S)}$, and the fibration map is defined by $\bar\pi(\bar y, \bar p) = \bar y$. Since all the maps involved are G^α, $P_{(S)}(M_{(S)}, G_{(G)})$ is a G^α super principal fiber bundle.

The map j induces a map $\hat j: P_{(S)} \to P_{(G)}$ defined by $\hat j((\bar y, \bar p)) = \bar p$ for $(\bar y, \bar p) \in P_{(S)}$. Then to each connection $\bar\omega$ in $P_{(G)}$ there corresponds a unique connection $\bar\omega'$ in $P_{(S)}$ given by the pullback of $\bar\omega$ by $\hat j$:

$$\bar\omega' = \hat j^* \bar\omega. \tag{3.1}$$

If $\bar\sigma$ is a local cross section in P, then

$$\bar\sigma'(\bar y) = (\bar y, \bar\sigma(\bar x)), \tag{3.2}$$

where $j(\bar y) = \bar x \in M_{(G)}$ is a local cross section in $P_{(S)}$ and $\hat j \circ \bar\sigma' = \bar\sigma \circ j$. It is easy to verify that $\bar\sigma'^* \bar\omega' = j^* \sigma^* \bar\omega$ has a trivial dependence on θ and $\bar\theta$ in the corresponding chart. Indeed we expect by analogy with other papers[11] that a dependence on θ and $\bar\theta$ is going to appear as a result of a gauge transformation, by which we mean an automorphism χ of $P_{(S)}$, that is a diffeomorphism $\chi: P_{(S)} \to P_{(S)}$ such that $\chi(\bar p' \bar a) = \chi(\bar p')\bar a, \forall \bar a \in G_{(G)}$ and $\bar p' \in P_{(S)}$. It can be represented by a function $\gamma: P_{(S)} \to G_{(G)}$, with $\chi(\bar p') = \bar p' \gamma(\bar p')$. As a consequence of an automorphism operation

$$\bar\omega' \to \bar\omega'_\gamma = \chi^* \bar\omega' = ad(\gamma^{-1})\bar\omega' + \gamma^{-1} d\gamma. \tag{3.3}$$

Undergoing the gauge transformation represented by γ, $\bar\omega'$ acquires an explicit θ and $\bar\theta$ dependence, as we shall see in detail in a moment.

Let U be a trivializing open set in M, $h: U \times G \to \pi^{-1}(U)$ the trivializing diffeomorphism and $A_\mu(x)dx^\mu$ a local one-form in U defining a connection form ω in $\pi^{-1}(U)$. It is well known that if certain compatibility conditions in the passage from a trivializing set to another are satisfied, ω is globally defined. Let $\sigma(x) = h(x,e), x \in U, e$ the identity in G, be the preferred cross section in U. Consider $z(\omega)$, and $\bar\omega'$ obtained by $z(\omega)$ through the pullback (3.1), and $\bar\sigma'$ obtained by $z(\sigma)$ through Eq. (3.2) and defined in $U_{(G)} \times B_L^{0,2}$. Then

$$\bar\sigma'^* \bar\omega' = A_\mu(\bar x) d\bar x^\mu, \tag{3.4}$$

where $\{\bar x^k\}$ is a coordinate system in $U_{(G)}$, and the transformation (3.3) in local coordinates is given by

$$A_\mu(\bar x)d\bar x^\mu \to g^{-1}(\bar x,\theta,\bar\theta) A_\mu(\bar x) d\bar x^\mu g(\bar x,\theta,\bar\theta) + g^{-1}(\bar x,\theta,\bar\theta) \underset{\sim}{d} g(\bar x,\theta,\bar\theta), \tag{3.5}$$

where $g = \gamma \circ \bar\sigma'$ and

$$\underset{\sim}{d} = \frac{\partial}{\partial \bar x_\mu} d\bar x^\mu + \frac{\partial}{\partial \theta} d\theta + \frac{\partial}{\partial \bar\theta} d\bar\theta.$$

We expand $g(\bar x,\theta,\bar\theta)$ in power series of θ and $\bar\theta$[4]:

$$g(\bar x,\theta,\bar\theta) = g_0(\bar x) + \theta g_1(\bar x) + \bar\theta g_2(\bar x) + \theta\bar\theta g_3(\bar x). \tag{3.6}$$

$g_0(\bar x)$ represents a usual gauge transformation and it is irrelevant here. Then let us consider $h(\bar x,\theta,\bar\theta) = g_0^{-1}(\bar x) g(\bar x,\theta,\bar\theta)$ and rewrite it as

$$\begin{aligned} h(\bar x,\theta,\bar\theta) &= 1 + \theta \underset{\sim}{\bar c}(\bar x) + \bar\theta \underset{\sim}{c}(\bar x) + \theta\bar\theta [\underset{\sim}{B}(\bar x) + \underset{\sim}{c}(\bar x)\underset{\sim}{\bar c}(\bar x)] \\ &= \exp\{\bar\theta \underset{\sim}{c}(\bar x) + \theta \underset{\sim}{\bar c}(\bar x) + \theta\bar\theta \\ &\quad \times [\underset{\sim}{B}(\bar x) + \tfrac{1}{2}(\underset{\sim}{c}(\bar x)\underset{\sim}{\bar c}(\bar x) + \underset{\sim}{\bar c}(\bar x)\underset{\sim}{c}(\bar x))]\}. \end{aligned} \tag{3.7}$$

The argument of the exponential map belongs to the Lie module $\mathscr{G}_{(G)}$ of the Grassmann Lie group $G_{(G)}$. Therefore $\underset{\sim}{c}(\bar x) = c^i(\bar x) z(\tau^i), \underset{\sim}{\bar c}(\bar x) = \bar c^i(\bar x) z(\tau^i), \underset{\sim}{B}(\bar x) = B^i(\bar x) z(\tau^i)$, where $\tau^i (i = 1,...,N)$ are the generators of the Lie algebra $\mathscr{G}$ of G. An important fact is that, as $\theta \bar c^i$ and $\bar\theta c^i$ are even Grassmann numbers, the functions $c^i(\bar x)$ and $\bar c^i(\bar x)$ must take their values in the odd part $B_L^{[1]}$ of B_L, so that they have an anticommuting character.

Therefore we have succeeded in recovering formula (1.1) as a real gauge transformation in the principal fiber bundle $P_{(S)}(M_{(S)}, G_{(G)})$, provided that we interpret $\{x_\mu\}$ as even Grassmann variables $A^i_\mu(x)$ and $B^i(x)$ as even Grassmann fields, and τ^i as the basis of the Grassmann extension of the Lie algebra.

Now we are able to decide what kind of Grassmann algebra B_L is the most appropriate in our case. In general we have to deal with products of fields containing an arbitrary number of FP fields. As the latters are $B_L^{(1)}$-valued, the products vanish from a certain order on, if L is finite. So we must choose L infinite; but, in this case, the entire differential structure must be G^ω (see Sec. 2F).

A final remark concerns the curvature $\bar{\Omega}'$ of the connection $\bar{\omega}'$ in Eq. (3.4) and $\bar{\Omega}'_\gamma$ corresponding to its gauge transformed $\bar{\omega}'_\gamma$. It is obvious that $\bar{\sigma}'^*\bar{\Omega}'$ has no dependence on θ and $\bar{\theta}$. From the form of the transformation (3.5) it is clear that $\bar{\sigma}'^*\bar{\Omega}'_\gamma$ has a nontrivial dependence on θ and $\bar{\theta}$, but it is zero in the θ and $\bar{\theta}$ direction.[1]

4. MATTER FIELDS

In this section we study the geometrical meaning of the matter fields in the framework of the super principal fiber bundle $P_{(S)}$.

It is well known that in connection with a principal fiber bundle $P(M,G)$, a matter field $\psi = \{\psi_i | i = 1,...,s\}$, transforming according to a finite dimensional representation r of G, can be regarded as a suitable function $\not{f}:P\to F$, where F is the representation space. The map $\not{f}$ satisfies the relation

$$\not{f}(pa) = r(a^{-1})\not{f}(p) \tag{4.1}$$

for every $p\in P$, $a\in G$. The relation between $\not{f}$ and ψ is given by a cross section σ [let $P(M,G)$ be trivial so that σ is a global cross section]:

$$\psi = \not{f}\circ\sigma. \tag{4.2}$$

As a consequence of Eqs. (4.1) and (4.2) a gauge transformation γ (see above for conventions) transforms ψ in the following way:

$$\psi(x)\to r(g^{-1}(x))\psi(x), \tag{4.3}$$

where $g = \gamma\circ\sigma$.

Now consider the space $F_{(G)} = B_L^{(0)}\otimes F$. $F_{(G)}$ is a free $B_L^{(0)}$ module. Indeed if $\{X_i | i = 1,...,s\}$ is a base for F, $\{1\otimes X_i | i = 1,...,s\}$ is a base for $F_{(G)}$. Consider the Grassmann Lie group GL $(F_{(G)})$ (see Sec. 2B). Then a representation $r:G\to \mathrm{GL}(F)$ can be continued to $z(r):G_{(G)}\to \mathrm{GL}(F_{(G)})$. Moreover, let us continue $\not{f}$ to $z(\not{f}):P_{(G)}\to G_{(G)}$. Then Eqs. (4.1)–(4.3) hold for the z-continued quantities with p,a,x replaced by $\bar{p}\in P_{(G)}$, $\bar{a}\in G_{(G)}$, and $\bar{x}\in M_{(G)}$, respectively. Now define $\bar{\not{f}}':P_{(S)}\to F_{(G)}$ by

$$\bar{\not{f}}' = z(\not{f})\circ\hat{j}, \tag{4.4}$$

then $\bar{\not{f}}'\circ\bar{\sigma}' = z(\psi)\circ j = \bar{\psi}$, if $\bar{\sigma}'$ is the same as in Eq. (3.4). Therefore the new field has trivial dependence on θ and $\bar{\theta}$. But as a consequence of a gauge transformation we obtain

$$\bar{\psi}\to z(r)(h^{-1}(\bar{x},\theta,\bar{\theta}))\bar{\psi}(\bar{x})\equiv\bar{\psi}_\gamma(\bar{x},\theta,\bar{\theta}), \tag{4.5}$$

where $h(\bar{x},\theta,\bar{\theta})$ is given by Eq. (3.7). The transformation (4.5) generates the superfield $\bar{\psi}_\gamma$. Again, a translation in $\bar{\theta}$ and θ coordinates generates the BRS and anti-BRS transformations of the matter field.[12]

ACKNOWLEDGMENTS

One of us (L. B.) wishes to thank the Department of Physics of the University of Massachusetts at Boston for the hospitality extended to him and Fondazione "Angelo della Riccia" for financial support.

APPENDIX A

Let G be the matrix group introduced in Sec. 2B. As it is a submanifold of some R^{n^2} we can find[13] a covering $\{U_\alpha,\alpha\in A\}$ of G and a set of C^∞ functions $F_\alpha:V_\alpha\to R^q$, where $q = \dim G$ and $V_\alpha\subset R^{n^2}$ is an open set containing U_α, with the following property: the restrictions $\psi_\alpha = f_{\alpha|U_\alpha}$ are such that $\{U_\alpha,\psi_\alpha,\alpha\in A\}$ is an atlas for G.

This is based on the implicit function theorem. By repeating a parallel argument for z continuations, we reach the same conclusion when we consider in the space $B_L^{n^2,0}$ the subset satisfying the constraint equations (2.19)—let us call it $\hat{G}$. There will be an atlas $\{\hat{U}_\alpha,\hat{\psi}_\alpha,\alpha\in A\}$, for $\hat{G}$, where $\hat{\psi}_\alpha = z(f_\alpha)_{|\hat{U}_\alpha}$.

In other words if $a\in U_\alpha$ we shall write

$$\zeta = \psi_\alpha(a) = F_\alpha(\xi), \tag{A1}$$

where $\zeta\in R^q$, $\xi = \{a_{jl}\}$ is the point in R^{n^2} representing a, that is $g_i(\xi) = 0$ for $i = 1,...,k$. Likewise if $\hat{a}\in\hat{U}_\alpha$, we write

$$\bar{\zeta} = \hat{\psi}_\alpha(\hat{a}) = z(F_\alpha)(\bar{\xi}), \tag{A2}$$

where $\bar{\zeta}\in B_L^{q,0}$, $\bar{\xi}$ is the point in $B_L^{n^2,0}$ satisfying $z(g_i)(\bar{\xi}) = 0$ for $i = 1,...,k$.

Let us identify a point $\hat{a}\in\hat{U}_\alpha$ to a point $\bar{a} = ((\zeta,\bar{\zeta})\in\bar{U}_\alpha$, if and only if

$$\bar{\zeta} = z(F_\alpha)(\bar{\xi}) = \hat{\psi}_\alpha(\hat{a}). \tag{A3}$$

For this identification to be correct we must show that if $\bar{\zeta}' = z(F_\beta)(\bar{\xi})$ for some $\beta\in A$, then $(\zeta,\bar{\zeta})\sim(\zeta'\bar{\zeta}')$, see Sec. 2A. Indeed $\psi_\alpha(\zeta) = \epsilon(\bar{\zeta}) = \epsilon(z(F_\alpha)(\bar{\xi})) = F_\alpha(\xi)$ and likewise $\psi_\beta(\zeta') = F_\beta(\xi)$, so that $\zeta = \zeta'$.

Moreover $\bar{\zeta}' = z(F_\beta\circ F_\alpha^{-1})(\bar{\zeta}) = z(\psi_\beta\circ\psi_\alpha^{-1})(\bar{\zeta})$, therefore $(\zeta,\bar{\zeta})\sim(\zeta',\bar{\zeta}')$. The converse is also true. There is a bijection between $G_{(G)}$ and $\hat{G}$, which is trivially a diffeomorphism.

APPENDIX B

Let ω be a connection form in $P(M,G)$. For notations we refer to Sec. 2. If $\{A_i\}$ is a basis for the Lie algebra $\mathscr{G}$ of G, ω can be written as

$$\omega = \sum\omega_i A_i. \tag{B1}$$

We define $z(\omega)$ by

$$z(\omega) = \sum_i z(\omega_i)z(A_i), \tag{B2}$$

where $z(\omega_i)$ have been defined in Sec. 2D and $z(A_i) = z(\tilde{A}_i)(\bar{e})$. $\tilde{A}_i$ is the left invariant vector field such that $\tilde{A}_i(e) = A_i$, where e is the identity in G, and $\bar{e}$ is the identity in $G_{(G)}$. It is easy to show that if $a(t)$, with $a(0) = e$, is the generating curve of $A\in\mathscr{G}$, then $z(a)(\bar{t})$ is the generating curve of $z(A)$, $\bar{t}\in B_L^{(0)}$. A generic element $\bar{A}$ of the Lie module $\mathscr{G}_{(G)}$ of $G_{(G)}$ has the form

$$\bar{A} = \sum_i\bar{c}_i z(A_i), \quad \text{where } \bar{c}_i\in B_L^{(0)}. \tag{B3}$$

The vertical vector field $\hat{A}_i$ is defined by

$$\hat{A}_i(p)f = \frac{d}{dt}f(r(p,a_i(t))), \quad f\in C^\infty(P). \tag{B4}$$

The vertical vector field in $P_{(G)}$ corresponding to $z(A_i)$ is given by $z(\hat{A}_i)$. Indeed,

$$z(\hat{A}_i)(\bar{p})\bar{F} = \frac{d}{d\bar{t}}\bar{f}(z(\tau)(\bar{p},z(a^i)(\bar{t}))) = z(\hat{A}_i)(\bar{p})\bar{f} \qquad \text{(B5)}$$

because if the one-parameter group of transformations[10] $\tau(p,a^i(t))$ generates $\hat{A}_i$, the one-parameter group $z(\tau)(\bar{p},z(a^i)(\bar{t}))$ generates $z(\hat{A}_i)$. Therefore any vertical vector field $\hat{\bar{A}}$ derived from a unique element $\bar{A}\in\mathscr{G}_{(G)}$, will have the form

$$\hat{\bar{A}} = \sum_i \bar{c}_i z(\hat{A}_i). \qquad \text{(B6)}$$

Now it is elementary to see that

$$z(\omega)(\hat{\bar{A}}) = \bar{A}. \qquad \text{(B7)}$$

Moreover

$$(R^*_{\bar{a}} z(\omega))(\hat{\bar{A}}) = ad_{\bar{a}^{-1}}\bar{A}, \qquad \text{(B8)}$$

where $R_{\bar{a}}:\bar{p}\to\tau(\bar{p},\bar{a})$ for $\bar{a}\in G_{(G)}$ and $\bar{p}\in P_{(G)}$. And this is enough to conclude that

$$R^*_{\bar{a}} z(\omega) = ad_{\bar{a}^{-1}} z(\omega). \qquad \text{(B9)}$$

Therefore $z(\omega)$ is a G^∞ connection in $P_{(G)}(M_{(G)}, G_{(G)})$.

[1] L. Bonora and M. Tonin, Phys. Lett. B **98**, 48 (1981).
[2] G. Curci and R. Ferrari, Nuovo Cimento A **32**, 151 (1976); I. Ojima, Prog. Theor. Phys. **65**, 625 (1980).
[3] L. Bonora, P. Pasti, and M. Tonin, Nuovo Cimento A **63**, 353 (1981).
[4] A. Rogers, J. Math. Phys. **21**, 1352 (1980).
[5] A. Rogers, "Super Lie Groups: Global Topology and Local Structure," ICTP/79-80/15 (unpublished).
[6] A. Jadczyk and K. Pilch, Commun. Math. Phys. **78**, 373 (1981).
[7] We refer to R^L for notational simplicity, but the whole section is valid if we replace R^L by C^L with only obvious changes. This observation is particularly relevant in connection with the discussion about groups (Sec. 2B).
[8] J. Kelley, *General Topology* (Van Nostrand Reinhold, New York, 1970).
[9] We use the prefix "super" only for geometrical entities involving odd Grassmann variables.
[10] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1969).
[11] J. Thierry-Mieg, J. Math. Phys. **21**, 2834 (1980); M. Quiros, F. J. de Urries, J. Hoyos, M. L. Mazon, and E. Rodriguez, J. Math. Phys. **22**, 1767 (1981).
[12] L. Bonora, P. Pasti, and M. Tonin, Padova preprint IFPD 37/81, to appear in Nuovo Cimento A.
[13] Pham Mau Quam, *Introduction à la Geometrie des Varietés Differentielles* (Dunod, Paris, 1969).

Higher Legendre transforms and their relationship to Bethe–Salpeter kernels and r–field projectors [a)]

Alan Cooper, Joel Feldman, and Lon Rosen
Mathematics Department, University of British Columbia, Vancouver, British Columbia, V6T 1Y4, Canada

(Received 24 June 1981; accepted for publication 7 August 1981)

We analyze the structure of the higher Legendre transforms $\Gamma^{(r)}\{A\}$ $(r\geqslant 1)$ of the generating functional G of the connected Green's functions G_n in Euclidean boson field theories. In addition to the vertex functions, $\Gamma^{(r)}$ generates a variety of objects of interest for their r-irreducibility in certain channels, e.g., r-irreducible expectations, rth order Bethe–Salpeter kernels, and r-field projectors. Our analysis is independent of perturbation theory, our definition of r-irreducibility being based on Spencer's idea of t-lines. We derive formulas for $\partial_t^n\Gamma^{(r)}\{A;t\}$ (in terms of either $\delta_A^n\Gamma^{(r)}\{A;t\}$ or the G_n's) to be used as input in the proofs of r-irreducibility. For the case of the weakly coupled $P(\phi)_2$ model, we establish the existence of the moments $\delta_A^n\Gamma^{(r)}\{0;t\}$ and their regularity in t.

PACS numbers: 11.10.St, 02.30. + g, 05.20. − y

I. INTRODUCTION

This is the third in a series of papers dealing with higher Legendre transforms[1,2] in Euclidean quantum field theory. In our first two papers[3,4] (which we shall refer to as I and II), we studied the first two Legendre transforms $\Gamma^{(N)}$ $(N=1,2)$, giving elementary proofs of their irreducibility properties and establishing their connection with N-irreducible expectations and with (generalized) Bethe–Salpeter (BS) kernels. In the present paper we analyze the general structure of $\Gamma^{(N)}$ for $N\geqslant 1$. The structural results we obtain prepare us for the proofs[5] of various irreducibility properties of $\Gamma^{(N)}$ when $N>2$ and they exhibit $\Gamma^{(N)}$ as the generator of N-irreducible expectations, Nth order BS kernels, "N-particle projectors",... . It is, of course, well known[1,6] that $\Gamma^{(N)}$ generates vertex functions $\Gamma_n^{(N)}$ that are N-irreducible $(N\leqslant 4)$, but what we wish to call attention to here is the remarkable role $\Gamma^{(N)}$ plays in organizing together all of these other field-theoretic objects which are of interest for their N-irreducibility in certain "channels".

To define $\Gamma^{(N)}$ in a Euclidean boson theory whose expectation is denoted $\langle\cdot\rangle$, we first introduce source terms of order N,

$$U\{J\}=\sum_{i=0}^{N}J_i\phi^i=J_i\phi^i, \tag{1.1}$$

where $J=(J_0,J_1,\ldots,J_N)$, $J_0\in\mathbb{C}$, and, for $n\geqslant 1$, $J_n=J_n(x_1,\ldots,x_n)$ is a symmetric function of n variables $x_i\in\mathbb{R}^d$. [We use the summation–integration convention that repeated (but possibly suppressed!) variables are summed or integrated over; e.g., $J_2\phi^2=\int dx_1\,dx_2\,J_2(x_1,x_2)\,\phi(x_1)\,\phi(x_2)$.] Then

$$Z\{J\}=\langle e^{U\{J\}}\rangle-1 \tag{1.2}$$

is the generator of the Schwinger functions $S_n=(\delta Z/\delta J_n)\{0\}$. If we transform $J_0,\ldots,J_r$ $(r<N)$ to "Schwinger variables"

$$A_\alpha^S\{J\}(x_1,\ldots,x_\alpha)=\delta_{J_\alpha(x_1,\ldots,x_\alpha)}Z\{J\}, \tag{1.3}$$

where α runs from 0 to r, then the rth *partial Schwinger Legendre transform* ${}^S\Gamma^{(r/N)}$ of Z is defined by

$${}^S\Gamma^{(r/N)}\{A_0^S,\ldots,A_r^S,J_{r+1},\ldots,J_N\}=Z\{J\}-A_\alpha^S\cdot J_\alpha, \tag{1.4}$$

where, on the right side of (1.4), $J_\alpha=J_\alpha\{A_0^S,\ldots,A_r^S,J_{r+1},\ldots,J_N\}$ is obtained by inverting (1.3). When all the J_i's are transformed (i.e., $r=N$) we obtain the (*complete*) Nth *Schwinger Legendre transform* ${}^S\Gamma^{(N)}\{A^S\}\equiv{}^S\Gamma^{(N/N)}\{A^S\}$. It is not hard to check[2] that ${}^S\Gamma^{(r/N)}$ can be defined iteratively in the sense that ${}^S\Gamma^{(r+1/N)}$ is the Legendre transform of ${}^S\Gamma^{(r/N)}$ with respect to the variable J_{r+1}. As an example, it is trivial to compute that

$${}^S\Gamma^{(0/N)}\{A_0^S,J_1,\ldots,J_N\}=A_0^S\ln\Big\langle\exp\Big(\sum_{i=1}^{N}J_i\phi^i\Big)\Big\rangle+A_0^S(1-\ln A_0^S)-1.$$

When A_0^S takes its "physical value" $S_0=1$,

$$\begin{aligned}{}^S\Gamma^{(0/N)}\{1,J_1,\ldots,J_N\}&=\ln\Big\langle\exp\sum_{i=1}^{N}J_i\phi^i\Big\rangle\\&=G^{(N)}\{J_1,\ldots,J_N\},\end{aligned} \tag{1.5}$$

where $G=G^{(N)}$ is the generator of (generalized) connected Green's functions. In other words, G is the 0th partial Legendre transform of Z. For most of this paper we shall suppress this 0th iteration of the transform and define ${}^S\Gamma^{(r/N)}$ directly as the transform of $G\{J\}=G\{J_1,\ldots,J_N\}$.

While ${}^S\Gamma^{(N)}\{A^S\}$ is the simplest transform to describe and while it will prove to be very useful for computations, there is a more physically meaningful transform $\Gamma^{(N)}$ with better irreducibility properties. $\Gamma^{(N)}$ can be obtained from ${}^S\Gamma^{(N)}$ by making a change of variables to connected variables,[1,2] $A^S\to A$. Alternatively, we can define $\Gamma^{(N)}$ directly as the transform of the functional $G\{J\}$ of (1.5): Let

$$\begin{aligned}A_i\{J\}&=\delta_{J_i}^i\,G\{J\}-\delta_{J_i}^i\,G\{0\},\quad 1\leqslant i\leqslant N,\\&\equiv G_i\{J\}-A_i^0,\qquad 1\leqslant i\leqslant N,\end{aligned} \tag{1.6}$$

[a)]Research supported by the Natural Sciences and Engineering Research Council of Canada.

where $A_i^0 = G_i\{0\} = \delta^i_{J_i} G\{0\}$. The *(complete)N*th *Legendre transform* of G is then

$$\Gamma^{(N)}\{A\} = G\{J\{A\}\} - G_{J_i}\{J\{A\}\}\cdot J_i\{A\}, \quad (1.7)$$

where $J\{A\}$ is the inverse of the map $A\{J\}$ of (1.6). Similarly we can define the rth partial Legendre transform $\Gamma^{(r/N)}\{A_1,..., A_r, J_{r+1},..., J_N\}$ by transforming only $J_1,...,J_r$.

There is yet a third transform ${}^W\Gamma^{(N)}\{A\}$ which is defined as in (1.6)–(1.7) but with G replaced by G^W, where the W means that the sources are physically Wick ordered (see Sec. II). ${}^W\Gamma^{(N)}$ turns out to be more suitable for rigorous mathematical purposes. In Sec. II we establish the (simple) relationships among ${}^S\Gamma,\Gamma$ and ${}^W\Gamma$; in particular, we show that Γ and ${}^W\Gamma$ are the same functional. (Note that we shall often suppress the superscript N.)

The reader should note that the subtraction of A^0 in (1.6) does not occur in the definitions of the Legendre transform in Refs. 1 and 6. Because we make this subtraction, the "physical value" of A (i.e., the value corresponding to $J=0$) is $A=0$, and so it is natural to make power series expansions in J or A about the origin. In addition, the subtractions introduce a small but (from the point of view of spectral analysis) desirable change in our notion of irreducibility. With the variables (1.6) we are led to say that a graph is *r-reducible* if by cutting up to r lines it may be disconnected into two components *each of which contains an external vertex*; otherwise we say it is *r-irreducible* (see Sec. II of I).

The underlying theme of this series of papers is that r-irreducibility is best understood and analyzed through the interplay of two key ideas. The first is that the systematic use of generating functionals (= higher Legendre transforms) eliminates much of the combinatorial complexity of the subject. The second idea is to use Spencer's "t-lines" or "separating surfaces" to give an analytic definition of r-irreducibility *that does not rely on perturbation theory*. This definition thus eliminates the complexity and lack of rigor arising from the use of (divergent) perturbation series.

We briefly explain Spencer's approach (for more details see I). The free part of the Euclidean measure of a boson field theory is a Gaussian measure

$$d\mu_C = \text{const}\, e^{-1/2\int dx[(\nabla\phi)^2 + m_0^2\phi^2]} = \text{const}\, e^{-1/2(\phi,C^{-1}\phi)} \quad (1.8)$$

with covariance $C = (-\Delta + m_0^2)^{-1}$. Let $C_\sigma = (-\Delta_\sigma + m_0^2)^{-1}$, where Δ_σ is the Laplacian with zero Dirichlet data on a $(d-1)$-dimensional surface $\sigma\subset\mathbb{R}^d$ which separates $\mathbb{R}^d$ into two parts. (We shall generally think of σ as being a hyperplane.) Introduce the interpolating covariance

$$C(t)(x,y) = tC(x,y) + (1-t)\,C_\sigma(x,y), \quad 0\leqslant t\leqslant 1 \quad (1.9)$$

and corresponding Gaussian measure $d\mu_{C(t)}$ defined as in (1.8) but with C replaced by $C(t)$. The effect of using $d\mu_{C(t)}$ instead of $d\mu_C$ in perturbation theory is that graphs have a factor $C(t)(x,y)$ instead of $C(x,y)$ associated with each line joining vertices x and y. Lines crossing σ and hence perturbation theory graphs containing such lines are distinguished by the fact that they are zero when $C(t) = C_\sigma$ at $t=0$. [Note that $C_\sigma(x,y)>0$ if x and y are on the same side of σ.] That is, given a graph $\mathscr{G} = \mathscr{G}(t)$ with external vertices $x, y,...$, the property "vertex x is connected to vertex y" is equivalent to the property "$\mathscr{G}(0)=0$ if x and y are separated by σ". Unlike $C(t)(x,y)$,

$$\dot C(x,y) = \frac{\partial C(t)}{\partial t}(x,y) = C(x,y) - C_\sigma(x,y)$$

is not zero when σ separates x and y. This means that taking a t-derivative of a line corresponds to cutting the line. As a result, the property that "$\mathscr{G}$ is 1-irreducible between x and y" is equivalent to the property that "$\mathscr{G}(0) = (\partial\mathscr{G}/\partial t)(0) = 0$ if x and y are separated by σ", and similarly for higher irreducibility properties.

At the nonperturbative level, the effect of using $d\mu_{C(t)}$ instead of $d\mu_C$ to define expectations is that $G = G\{J;t\}$ acquires a t dependence (as do $\Gamma^{(N)}, \Gamma_n^{(N)},...$). Motivated by the above discussion of graphs in perturbation theory, we can define the irreducibility properties of an object like a BS kernel or vertex function in terms of the vanishing of its t-derivatives, without any reference to perturbation theory (see Sec. II of I). This is Spencer's approach, and it has already been successfully applied in a number of situations.[7–11]

As an illustration of how nicely the ideas of generating functionals and separating surfaces combine, consider the assertion that the Green's functions

$$G_n(x_1,...,x_n) = \prod_{i=1}^{n} \delta_{J_1(x_i)}\, G\{J\}\big|_{J=0}$$

are connected for $n = 2,3,...$. This assertion may be equivalently expressed in terms of the generating functional $G\{J;t\}$ as

$$\frac{\delta^2 G\{J_1,0,...,0;0\}}{\delta J_1(x_1)\,\delta J_1(x_2)} = 0 \quad \text{for all } J_1 \quad (1.10)$$

if x_1 and x_2 are on opposite sides of σ. We refer to (1.10) as the "connectedness of $G\{J;t\}$" (see Sec. II of I). But (1.10) is obvious: When $t=0$ $d\mu_{C(t)} = d\mu_{C_\sigma}$ factors across σ and, since the interaction part of the Euclidean action is local, $\langle\cdot\rangle$ also factors. Hence if $F_\pm$ are functions of ϕ depending only on ϕ on the $\pm$ side of σ we have

$$\langle F_+ F_-\rangle = \langle F_+\rangle\langle F_-\rangle \quad \text{at} \quad t=0. \quad (1.11)$$

In particular, taking $F_\pm = \exp(\chi_\pm J_1\cdot\phi)$, we obtain

$$G\{J_1,...,0;0\} = G\{\chi_+ J_1,...\} + G\{\chi_- J_1,...\}, \quad (1.12)$$

where $\chi_\pm(x)$ is the characteristic function of the $\pm$ side of σ. (1.10) follows at once from (1.12).

The basic differentiation formula for $\Gamma^{(N)}$ $(N>1)$ is quite simple:

$$\partial_t\,\Gamma\{A;t\} = \Gamma_{A_i}\{A;t\}\,\dot A_i^0$$

for suitable A (Theorems III.1 and VI.11). The evaluation of $\partial_t^r\Gamma$ thus reduces to the evaluation of functional derivatives $\Gamma_{A_iA_j\cdots}$ of Γ. The thrust of Sec. III is to express such functional derivatives in terms of the more familiar connected Green's functions G_n.

In Sec. IV we show how easily r-irreducible expectations[12] and r-particle projections[13] may be expressed in terms of the Legendre transform. A typical formula is the one for the r-irreducible expectation of ϕ^i and ϕ^j (i.e., r-

irreducible between the arguments of ϕ^i and the arguments of ϕ^j):

$$\langle \phi^i ; \phi^j \rangle^{(r)} = \Gamma^{(r/N)}_{J_i J_j}\{0\} \quad \text{for } r<i,j\leqslant N.$$

In Sec. V we show how $\Gamma^{(N)}$ generates the Nth order BS kernel $K^{(N)}$, the basic formula being

$$K^{(N)}(x,y) = -\text{connected part of } \Gamma^{(N)}_{A_N(x) A_N(y)}\{0\},$$

where $x, y \in \mathbb{R}^{Nd}$.

Although Spencer's idea liberates us from divergent perturbation theory, it does not solve all problems of mathematical rigor. As a matter of fact it introduces new problems: not only must we show that, say, $\Gamma^{(N)}\{A;t\}$ is well defined and differentiable with respect to A for A in some Banach space of functions, but we must also worry about whether it is differentiable in t and whether the resulting formulas make sense. We postpone such questions until Sec. VI. In particular Secs. II–V are written as though $G\{J\}$, $\Gamma^{(N)}\{A\}$,... were *bona fide* regular functionals on some Banach space and all the formulas made sense in this framework. However, as we explained in I, such assumptions are dubious for $N\geqslant 3$ and it is necessary to interpret the results of Secs. II–V in the framework of formal power series (fps) (see II). Thus when we provide the mathematical justification for these results in Sec. VI (such as the existence of G, $\Gamma^{(N)}$, their regularity in t, validity of the differentiation formulas, ...) it is in the context of fps.

In conclusion, $\Gamma^{(N)}$ and $\Gamma^{(r/N)}$ generate a variety of objects (in addition to the vertex functions) whose desired irreducibility properties may all be summarized by appropriate (channel) irreducibility properties of the Legendre transform. The proofs of some of these properties have already been given in I and others will be the subject of future papers in this series.[5,14]

II. THE DEFINITION(S) OF $\Gamma^{(N)}$

In this section we discuss the relationships among several Nth Legendre transforms (see Theorem II.3). We shall deal explicitly only with complete transforms although with obvious modifications the results are equally valid for partial transforms. We start by reminding the reader of the definitions of $\Gamma^{(N)}$ and ${}^S\Gamma^{(N)}$ and by defining ${}^W\Gamma^{(N)}$. Let $J=(J_0,\ldots,J_N)$, where $J_n = J_n(x_1,\ldots,x_n)$ is a symmetric function of n variables $x_i \in \mathbb{R}^d$. For notational convenience we allow a $J_0 \in \mathbb{C}$ component in J although generally $J_0=0$. Then the source terms are

$$U\{J\} = J_i \phi^i \tag{2.1a}$$

and

$$U^W\{J\} = J_i :\phi^i:, \tag{2.1b}$$

where the "physically Wick ordered" powers are defined by

$$:\phi^i: \; = \; :\phi(x_1)\cdots\phi(x_i): \; = \left(\prod_{k=1}^{i} \frac{\delta}{\delta f(x_k)}\right) \frac{e^{\phi f}}{\langle e^{\phi f}\rangle}\bigg|_{f=0}. \tag{2.2}$$

For example $:\phi: = \phi - \langle\phi\rangle$ and

$$\begin{aligned} :\phi^2: &= (:\phi:)^2 - \langle(:\phi:)^2\rangle \\ &= \phi^2 - 2\phi\langle\phi\rangle - \langle\phi^2\rangle + 2\langle\phi\rangle^2. \end{aligned}$$

Remark 1. It is obvious from the definition that $\langle :\phi^i: \rangle = 0$ for $i\geqslant 1$. However, Wick powers are not orthogonal with respect to $\langle\cdot\rangle$, i.e., $\langle :\phi^i: \, :\phi^j: \rangle \neq 0$ for $1\leqslant i<j$. See Sec. IV for the appropriate orthogonalization procedure.

The generators to be transformed are

$$G\{J\} = \ln\langle \exp U\{J\}\rangle \tag{2.3a}$$

and

$$G^W\{J\} = \ln\langle \exp U^W\{J\}\rangle \tag{2.3b}$$

with $J_0 = 0$. The conjugate variables we shall use are the connected variables

$$A_i\{J\} = G_i\{J\} - G_i\{0\} = G_i\{J\} - A_i^0, \quad 1\leqslant i\leqslant N, \tag{2.4a}$$

$$A_i^W\{J\} = G_i^W\{J\} - G_i^W\{0\}, \quad 1\leqslant i\leqslant N, \tag{2.4b}$$

where $G_i\{J\} = \delta^i_{J_i}\, G\{J\}$, $A_i^0 = G_i\{0\}$, and the Schwinger variables

$$A_i^S\{J\} = G_{J_i}\{J\}. \tag{2.4c}$$

The conjugate functions to the G's written in terms of these variables are the Legendre transforms of interest:

$$\Gamma^{(N)}\{A\} = G\{J\{A\}\} - G_{J_i}\{J\{A\}\}\, J_i\{A\}, \tag{2.5a}$$

$${}^S\Gamma^{(N)}\{A^S\} = G\{J^S\{A^S\}\} - G_{J_i}\{J^S\{A^S\}\}\, J_i^S\{A^S\}, \tag{2.5b}$$

$$\begin{aligned} &{}^W\Gamma^{(N)}\{A^W\} \\ &\quad = G^W\{J^W\{A^W\}\} - G^W_{J_i}\{J^W\{A^W\}\}\, J_i^W\{A^W\}, \end{aligned} \tag{2.5c}$$

where $J\{A\}$ is the inverse of the map $(0,J_1,\ldots,J_N)\mapsto A\{J\}$, etc. In Sec. IV we also consider transforms ${}^q\Gamma^{(r/N)}$, where $q_n = \phi^n + \cdots$ is a more general polynomial in the fields than ϕ^n or $:\phi^n:$.

We will shortly see (in Theorem II.3) that $\Gamma\{A\} = {}^W\Gamma\{A\}$. Nonetheless it is much easier to deal rigorously with ${}^W\Gamma\{A\}$ since $G^W\{J\}$ provides a natural domain for J, namely

$$\{J=(0,J_1,\ldots,J_N) \,|\, (J,\mathscr{C}J)<\infty, \quad J_i \text{ symmetric}\},$$

where $\mathscr{C}_{ij} = \delta_{ij} C^{\otimes i}$ and C is the covariance of the theory. For example, in $\epsilon P(\phi)_2$ [weakly coupled $P(\phi)_2$] we have by Theorem VI.2 that $|\langle (J_i :\phi^i:)^n\rangle| \leqslant \text{const}(J,\mathscr{C}J)^{n/2}$. The reason for this is that, thanks to the subtraction scheme in $: \; :$ every field in $:\phi^i:$ must be connected to to something outside its dots. To be precise suppose the expectation $\langle\cdot\rangle$ factors across σ, $g(\phi)$ is some polynomial in the field supported on the $+$ side of σ, and the set $I_- = \{k \,|\, \text{the argument } x_k \text{ of } :\phi^i: \text{ is on the } - \text{ side of } \sigma\}$ is nonempty. Then

$$\begin{aligned} &\langle :\phi^i: g(\phi)\rangle \\ &= \left(\prod_{k\in I_+} \frac{\delta}{\delta f(x_k)}\right) \frac{\langle e^{\phi f} g(\phi)\rangle}{\langle e^{\phi f}\rangle} \left(\prod_{k\in I_-} \frac{\delta}{\delta f(x_k)}\right) \frac{\langle e^{\phi f}\rangle}{\langle e^{\phi f}\rangle} \\ &= \langle :\phi^{I_+}: g(\phi)\rangle \, \langle :\phi^{I_-}:\rangle = 0, \end{aligned}$$

where $:\phi^{I_+}: \; = \; :\Pi_{k\in I_+}\phi(x_k):$. There is also a natural domain for A in ${}^W\Gamma\{A\}$ because of the subtraction in the definition (2.4b) of $A^W\{J\}$. This domain is

$$\{A \,|\, (A,\mathscr{C}^{-1}A)<\infty, \; A_i \text{ symmetric}\}.$$

See Theorem VI.1.

The virtue of ${}^S\Gamma$ is that it is easier to compute with than

Γ while at the same time being simply related to Γ (see Theorem II.3). Hence we may derive formulas for (the more physically meaningful) Γ by first obtaining the corresponding formulas for ${}^S\Gamma$.

The change of variables $A = {}^SF\{A + A^0\}$ (see Lemma II.2) relating Γ and ${}^S\Gamma$ is implemented by the functionals $F_i\{A\}$, where

$$\sum_{i=0}^{\infty} \frac{1}{i!} F_i f^i = \exp\left\{\sum_{i=1}^{\infty} \frac{1}{i!} A_i f^i\right\},$$

i.e., (2.6)

$$F_i\{A\}(x_1,\dots,x_i) = \left(\sum_{r=1}^{i} \frac{\delta}{\delta f(x_r)}\right) \exp\left\{\sum_{p=1}^{\infty} \frac{1}{p!} A_p f^p\right\}\Bigg|_{f=0}.$$

For example, $F_0 = 1$,

$$F_1(x_1) = A_1(x_1),$$

$$F_2(x_1, x_2) = A_2(x_1, x_2) + A_1(x_1) A_1(x_2),$$

$$\begin{aligned}F_3(x_1, x_2, x_3) &= A_3(x_1, x_2, x_3) + (A_2(x_1, x_2) A_1(x_3) + \text{perms}) \\ &\quad + A_1(x_1) A_1(x_2) A_1(x_3) \\ &= A_3 + 3A_2 A_1 + A_1^3.\end{aligned}$$

Remark 2. We must take into account in the definitions of $\delta/\delta A_i$ and $\delta/\delta J_i$ the fact that our generating functionals are defined for J_i and A_i symmetric. This is done in the obvious way. For example,

$$\begin{aligned}\frac{\delta}{\delta A_2(x, y)} \Gamma^{(2)} &= \Gamma^{(2)}_{A_2(x,y)} \\ &= \frac{d}{d\lambda} \Gamma^{(2)} \Big\{A_1, A_2 + \frac{\lambda}{2}\,[\delta(\cdot - x)\,\delta(\cdot - y) \\ &\quad + \delta(\cdot - y)\,\delta(\cdot - x)]\Big\}\Big|_{\lambda=0}.\end{aligned}$$

We shall also use the symmetrization convention that all formulas are to be interpreted as symmetrized so that J_i, A_i, $\delta/\delta J_i$, $\delta/\delta A_i$, F_i, etc., are symmetric under permutations of their arguments. With this convention it is often possible to write $\{f(x_1,\dots,x_m) + \text{perms}\}$ as $nf(x_1,\dots,x_m)$, where n is the number of permutations.

It follows immediately from the definition (2.6) that $F_i\{A\}$ depends only on $A_1,\dots,A_i$, that F obeys the addition formula

$$F_i\{A + B\} = \sum_{j=0}^{i} \binom{i}{j} F_{i-j}\{A\}\, F_j\{B\}, \tag{2.7}$$

and that for $j \leqslant i$ F obeys the differentiation formula

$$\begin{aligned}\frac{\delta F_i\{A\}(u_1,\dots,u_i)}{\delta A_j(x_1,\dots,x_j)} &= \left(\prod_{r=1}^{i} \frac{\delta}{\delta f(u_r)}\right) \frac{1}{j!} \prod_{k=1}^{j} f(x_k) \exp\left[\sum_{p=1}^{\infty} \frac{1}{p!} A_p f^p\right]\Bigg|_{f=0} \\ &= \binom{i}{j} \left(\prod_{k=1}^{j} \delta(u_k - x_k)\right) F_{i-j}\{A\}(u_{j+1},\dots,u_i).\end{aligned} \tag{2.8}$$

(Note the use of the symmetrization convention of Remark 2.)

These functionals are also intimately involved in the relationship between Wick-ordered and un-Wick-ordered products of fields. For example,

$$:\phi^2: = \phi^2 + 2F_1\{-A^0\}\,\phi + F_2\{-A^0\}.$$

In general

*Lemma II.*1.

$$\text{(a)}\quad :\phi^i: = \sum_{j=0}^{i} W^-_{ij}\, \phi^j, \tag{2.9}$$

$$\text{(b)}\quad \phi^i = \sum_{j=0}^{i} W^+_{ij} :\phi^j:, \tag{2.10}$$

$$\text{(c)}\quad U^W\{J\} = U\{W^{-t} J\}, \tag{2.11}$$

where

$$W^{\pm}_{ij} = \begin{cases}\binom{j}{i} F_{i-j}\{\pm A^0\} = \binom{i}{j}\left(\frac{\delta}{\delta f}\right)^{i-j} \langle e^{\phi f}\rangle^{\pm 1}\Big|_{f=0}, & 0 \leqslant j \leqslant i \\ 0, & \text{otherwise,}\end{cases}$$

$$W^{-\,t}_{ij}(x_1\cdots; y_1\cdots) = \begin{cases}\binom{j}{i}\left(\prod_{k=1}^{i} \delta(x_k - y_k)\right) F_{j-i}\{-A^0\}(y_{i+1},\dots,y_j), & i \leqslant j \leqslant N \\ 0, & \text{otherwise.}\end{cases}$$

Remark 3. Here the products are to be interpreted so that the number of variables and the symmetry properties match up properly. Hence $W^-_{ij}\,\phi^j$ is the symmetrized tensor product of W^-_{ij} and ϕ^j with the factor $\binom{i}{j}$ in the definition of W^- being the number of terms in the sum

$$(W^-_{ij}\,\phi^j)(x_1\cdots x_i) = \sum_{\substack{I \subset \{1,\dots,i\} \\ |I| = j}} F_{i-j}\{-A^0\}(x_{I'})\,\phi^I,$$

where $\phi^I = \Pi_{k \in I}\, \phi(x_k)$ and $x_{I'} = (x_k)_{k \notin I}$. On the other hand W^{-t} (the transpose of W^-) acts as the integral operator

$$\begin{aligned}(W^{-t} J)_i(x_1\cdots x_i) &= J_i(x_1\cdots x_i) + \sum_{j=i+1}^{N} \binom{j}{i} \int dy_{i+1}\cdots dy_j\, F_{j-i}\{-A^0\}(y) \\ &\quad \times J_j(y, x).\end{aligned}$$

Note that W^+, W^-, and W^{-t} are all triangular matrices with identity operators on the diagonal and hence are all invertible.

Proof. The lemma is an immediate consequence of the definitions

$$:\phi^i: = \sum_{j=0}^{i} \binom{i}{j}\left(\frac{\delta}{\delta f}\right)^{i-j} \langle e^{\phi f}\rangle^{-1}\Big|_{f=0} \phi^j \quad [\text{see (2.2)}],$$

$$\phi^i = \sum_{j=0}^{i} \binom{i}{j}\left(\frac{\delta}{\delta f}\right)^{i-j} \langle e^{\phi f}\rangle\Big|_{f=0} :\phi^j:,$$

$$\langle e^{\phi f}\rangle^{\pm 1} = \exp\left\{\pm \sum_{i=1}^{\infty} \frac{1}{i!} A^0_i f^i\right\} = \sum_{i=0}^{\infty} \frac{1}{i!} F_i(\pm A^0) f^i$$

[see (2.4) and (2.6)],

$$\begin{aligned}U^W\{J\} &= \sum_{j=0}^{N} J^j :\phi^j: = \sum_{j=0}^{N} \sum_{i=0}^{j} J_j(W^-_{ji}\,\phi^i) \\ &= \sum_{i=0}^{N} \sum_{j=i}^{N} (J_j(W^-_{ji}\,\phi^i) = U\{W^{-t} J\}.\end{aligned} \qquad \blacksquare$$

*Lemma II.*2.

(a) $A^S\{J\} = F\{A\{J\} + A^0\}$,

(b) $G^W_{J_i}\{J\} = F_i\{A^W\{J\}\}$,

(c) $A^W\{J\} = A\{W^{-t}J\}$.

Proof. (a) $A_i^S = \langle \phi^i e^{U\{J\}} \rangle \langle e^{U\{J\}} \rangle^{-1}$

$$= \left(\frac{\delta}{\delta f}\right)^i \langle e^{\phi f + U\{J\}} \rangle \langle e^{U\{J\}} \rangle^{-1} \Big|_{f=0}$$

$$= \left(\frac{\delta}{\delta f}\right)^i \exp\{G\{J+f\} - G\{J\}\}\Big|_{f=0}, \qquad (2.12)$$

where we identify f and $(0,f,0,0,...,0)$. But

$$G\{J+f\} - G\{J\} = \sum_{j=1}^{\infty} \frac{1}{j!} G_j\{J\} f^j = \sum_{j=1}^{\infty} \frac{1}{j!} (A_j + A_j^0) f^j$$

and the assertion follows from the definition (2.6) of F.

(b) From definition (2.3b)

$$G_{J_i}^W = \langle :\phi^i: e^{U^W} \rangle \langle e^{U^W} \rangle^{-1}$$
$$= \left(\frac{\delta}{\sigma f}\right)^i \langle e^{U^W} + \phi f \rangle \langle e^{U^W} \rangle^{-1} \langle e^{\phi f} \rangle^{-1} \Big|_{f=0}$$
$$= \left(\frac{\delta}{\delta f}\right)^i \exp\{G^W\{J+f\} - G^W\{J\} - G^W\{f\}\}\Big|_{f=0}$$
$$= \left(\frac{\delta}{\delta f}\right)^i \exp\left\{\sum_{j=1}^{\infty} \frac{1}{j!} (G_j^W\{J\} - G_j^W\{0\}) f^j\right\}\Big|_{f=0}$$
$$= F\{A^W\{J\}\} \quad \text{by (2.4b) and (2.6).}$$

(c) Note that in (2.11) the only components $(W^{-t}J)_i$ of $W^{-t}J$ having a dependence on J_1 are the $i = 0,1$ components and

$$(W^{-t}J)_0 = J_0 - A_1^0 J_1 + \text{terms in } J_2 \cdots J_N,$$
$$(W^{-t}J)_1 = J_1 + \text{terms in } J_2 \cdots J_N.$$

Hence

$$G_i^W\{J\} = G_i\{W^{-t}J\} - \delta_{i1} A_i^0$$

and the claim follows from definitions (2.4a) and (2.4b) of A and A^W. ∎

We are now in a position to present the (formal) relationships between Γ, ${}^W\Gamma$, and ${}^S\Gamma$.

Theorem II.3.

(a) $\Gamma\{A\} = {}^W\Gamma\{A\}$,

(b) $\Gamma\{A\} = {}^S\Gamma\{F\{A + A^0\}\}$.

Proof.

(a) Denote by $B_{ij}^{(N)}$ the restriction of B_{ij} to $1 < i,j < N$. Setting $J_0 = 0$ we have by (2.11)

$$G^W\{J\} - G_{J_i}^W\{J\} J_i$$
$$= G\{(W^{-t})^{(N)} J\} - W_{i0}^{-} J_i$$
$$- [(W^{-})_{ik}^{(N)} G_{J_k}\{(W^{-t})^{(N)}J\} J_i - W_{i0}^{-} J_i]$$
$$= G\{W^{-t})^{(N)}J\} - G_{J_k}\{(W^{-t})^{(N)}J\}\{(W^{-t})^{(N)}J\}_k.$$

However, both $A^W\{J\}$ and $A\{J\}$ are independent of J_0 so Lemma II.2(c) gives $(W^{-t})^{(N)} J^W\{A\} = J\{A\}$ and hence

$${}^W\Gamma\{A\} = G^W\{J^W\{A\}\} - G_{J_i}^W\{J^W\{A\}\} J_i^W\{A\}$$
$$= G\{J\{A\}\} - G_{J_k}\{J\{A\}\} J_k\{A\}$$
$$= \Gamma\{A\}.$$

(b) By Lemma II.2a $F\{A + A_0\{ = A^S\{J\{A\}\}$ so that from the definition (2.5b) of ${}^S\Gamma$

$${}^S\Gamma\{F\{A + A_0\}\} = {}^S\Gamma\{A^S\{J\{A\}\}\}$$
$$= G\{J\{A\}\} - G_{J_i}\{J\{A\}\} J_i\{A\}$$
$$= \Gamma\{A\}. \quad ∎$$

Theorem II.3 provides two special cases of a general result which states roughly that changes of variables in the source space (i.e., the J's) or in the maps defining the conjugate variables (i.e., $A\{J\}$) cause only a change of variables in the Legendre transform. More precisely, if Q is a real Banach space, G is a real-valued Fréchet differentiable function on Q and P is an invertible map from Q to its dual Q^* we define the P-Legendre transform by

$${}^P G^*(p) = G(P^{-1}(p)) - \langle \nabla G(P^{-1}(p)), P^{-1}(p)\rangle,$$

where $\nabla G: Q \to Q^*$ is the Fréchet derivative of G. It is not hard to see that if $G_1, G_2: Q \to \mathbb{R}$ are related by

$$G_2(q) = G_1(Tq) - \langle a,q \rangle,$$

where T is a bounded linear map on Q and $A \in Q^*$, then

$${}^{P_2}G_2^*(p) = {}^{P_1}G_1^*(P_1(TP_2^{-1}(p))). \qquad (2.13)$$

In Theorem II.3a we have $G_1 = G$, $G_2 = G^W$, $T = (W^{-t})^{(N)}$, $P_1 = A$, and $P_2 = A^W$. What makes these variables so special as far as (2.13) is concerned is that (by Lemma II.2c) $P_2 = P_1 T$ rather than $P_2 = T^* P_1 T$ as one would expect for the "vector" variable $P = \nabla G$. As a result in (2.13) $P_1 T P_2^{-1}$ is the identity operator and ${}^{P_2}G_2^* = {}^{P_1}G_1^*$.

III. DERIVATIVES of $\Gamma^{(N)}$

As we discussed in the Introduction we use the approach to irreducibility introduced by Spencer.[12] In this approach the covariance C is replaced by $C(t)$ [see (1.9)] and cutting a line corresponds to taking a t-derivative. Accordingly we modify definitions (2.3)–(2.5) in the obvious way to include a t dependence; thus, e.g., $G = G\{J;t\}$, $A^S = A^S\{J;t\}$ so that upon inversion $J^S = J\{A^S;t\}$, and

$${}^S\Gamma = {}^S\Gamma\{A^S;t\} = G\{J^S\{A^S;t\};t\} - A_i^S J_i^S\{A^S;t\}.$$

To discuss irreducibility properties of $\Gamma\{A;t\}$ we must be able to differentiate it with respect to t. Note that in such computations A and t are the independent variables. Hence, denoting $\partial/\partial t$ by , we have

$${}^S\dot\Gamma = \dot G + G_{J_i}\dot J_i^S - A_i^S \dot J_i^S = \dot G. \qquad (3.1)$$

In Theorem III.1 we derive the corresponding formula for $\dot\Gamma$ and express $\dot\Gamma$ in terms of functional derivatives Γ_{A_i}. Then in Theorems III.3 and III.7 we give formulas expressing functional derivatives of Γ in terms of the more familiar connected Green's functions G_k. Many of the properties (and in particular irreducibility properties) of Γ follow from an analysis of these formulas.[3,5] Again as in Sec. II we deal explicitly only with complete transforms although again with obvious modifications the results are valid for partial transforms too.

Theorem III.1. For $N \geqslant 2$

(a) ${}^S\dot\Gamma^{(N)}\{A;t\} = -\frac{1}{2}\dot C(t)^{-1}(A_2 - A_2^0 - A_1^0 A_1^0)$,

(b) $\dot\Gamma^{(N)}\{A;t\} = -\frac{1}{2}\dot C(t)^{-1}(A_2 + A_1 A_1 + 2A_1 A_1^0) + \Gamma_{A_i}^{(N)} \dot A_i^0$,

where $\dot C^{-1} A_2$ means

$$-\operatorname{tr} C^{-1}\dot{C}C^{-1}A_2 = -\int dx\,dy\,\dot{C}(x,y)(C^{-1}\otimes C^{-1}A_2)(x,y)\,.$$

Remark 1. We warn the reader that some care must be taken when manipulating $\dot{C}^{-1}$. See Sec. III of I.

Proof. (a) If $d\mu_{C(t)}$ is the Gaussian measure of mean zero and covariance $C(t)$ then[15]

$$\partial_t \int f(\phi)\,d\mu_{C(t)} = -\tfrac{1}{2}\int\left[\int :\phi(x)\,\dot{C}^{-1}\phi(y):\,dx\,dy\right] f(\phi)\,d\mu_{C(t)}\,,$$

where : : is Wick ordering with respect to $d\mu_{C(t)}$. Hence

$$\dot{G}\{J\} = -\tfrac{1}{2}\dot{C}^{-1}(G_{J_2}\{J\} - G_{J_2}\{0\}) \qquad (3.2)$$

and so by (3.1)

$${}^S\dot{\Gamma}\{A^S\} = -\tfrac{1}{2}\dot{C}^{-1}\{A_2^S - A_2^0 - A_1^0 A_1^0\}\,.$$

(b) From Theorem II.3b

$$\begin{aligned}\dot{\Gamma}\{A\} &= {}^S\dot{\Gamma}\{F\{A+A^0\}\} + {}^S\Gamma_{A_i}\frac{\delta F_i}{\delta A_j}\dot{A}_j^0 \\ &= -\tfrac{1}{2}\dot{C}^{-1}(F_2\{A+A^0\} - F_2\{A^0\}) + \Gamma_{A_j}\dot{A}_j^0 \\ &= -\tfrac{1}{2}\dot{C}^{-1}(A_2 + A_1A_1 + 2A_1A_1^0) + \Gamma_{A_j}\dot{A}_j^0\,. \blacksquare\end{aligned}$$

We now turn to the question of expressing $\Gamma_{A_{j_1}\cdots A_{j_m}}$ in terms of G_k. This will be done for $m\leqslant 2$ in Theorem III.3 and for $m>2$ in Theorem III.7. In Theorem III.3 we express $\Gamma^{(N)}_{A_iA_j}$ in terms of

$$S_{ij} = \left(\frac{\delta}{\delta f}\right)^i\left(\frac{\delta}{\delta g}\right)^j \mathscr{S}\{f,g\}\Big|_{f=g=0}\,, \qquad (3.3a)$$

where

$$\mathscr{S}\{f,g\} = \exp[G\{J+f+g\} - G\{J+f\} - G\{J+g\} + G\{J\}]\,. \qquad (3.3b)$$

Here f and g are functions of one variable and we identify f with $(0,f,0,0,\ldots)$. Each index i is an abbreviation for a set $x^{(i)} = \{x_1,\ldots,x_i\}$ of i points in $\mathbb{R}^d$ and $(\delta/\delta f)^i$ is an abbreviation for $\Pi_{m=1}^i(\delta/\delta f(x_m))$. Hence S_{ij} is a "function" on $\mathbb{R}^{id}\times\mathbb{R}^{jd}$ and S should be thought of as a matrix whose (i,j) entry is the operator with kernel S_{ij}:

$$S = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & G_2 & G_3 & \cdots \\ 0 & G_3 & G_4 + 2G_2^2 & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix},$$

where $G_n = G_{J_1^n}\{J\}$ and we have used the symmetrization convention of Remark 2 of Sec. II. For example,

$$S_{2,2}(x^{(2)},y^{(2)}) = G_4(x_1,x_2,y_1,y_2) + G_2(x_1,y_1)\,G_2(x_2,y_2) + G_2(x_1,y_2)\,G_2(x_2,y_1)\,.$$

The significant features of S are isolated in the following remarks.

Remark 2. As a consequence of

$$\mathscr{S}\{f,g\} = \exp\left[\sum_{i,j=1}^{\infty} G_{i+j}\frac{f^i}{i!}\frac{g^j}{j!}\right]$$

we have that $S_{ij} = S_{ji}$ is a polynomial in $\{G_k\,|\,k\subset i\cup j,\ k\cap i\neq\emptyset,\ k\cap j\neq\emptyset\}$. So every space-time point in i is connected to a space-time point in j and vice versa.

Remark 3. Note that S is independent of N, the order of the Legendre transform. In particular, S is a function of $(J_0, J_1, J_2,\ldots)$ rather than $(J_1,\ldots,J_N)$. Furthermore, we do *not* restrict the arguments of S_{ij} to the range $1\leqslant i,j\leqslant N$. When we wish to do so we will use $S^{(N)}$ to denote the resulting $N\times N$ matrix. In the following analysis, inasmuch as G generally enters in the combination $G\{J\{A\}\}$, S enters either in the combination $S\{J\{A\}\}$ or in the combination $S^{(N)}\{J\{A\}\}$. Hence the N dependence enters through the superscript (N) and through $J\{A\}$.

The reason $\Gamma^{(N)}_{A_iA_j}$ may be expressed in terms of S is provided in

Lemma III.2.
(a) For all $m\geqslant 0$, $1\leqslant n\leqslant N$

$$\frac{\delta A_m}{\delta J_n} = \sum_{k=0}^{n}\binom{n}{k}A_{n-k}^S S_{km} \qquad (3.4)$$

$$= \sum_{k=0}^{n}\frac{\delta F_n\{A+A^0\}}{\delta A_k}S_{km}\,, \qquad (3.5)$$

with the conventions that for *all* $m\geqslant 0$

$$A_m^S = \langle\phi^m e^{U\{J\}}\rangle\,\langle e^{U\{J\}}\rangle^{-1}\,,$$

$$A_m = G_m\{J\} - G_m\{0\}\,,$$

$$\frac{\delta F_n}{\delta A_0} = F_n\,.$$

(b) As $N\times N$ matrices

$$\frac{\delta A}{\delta J} = S^{(N)}\left(\frac{\delta F\{A+A^0\}}{\delta A}\right)^t.$$

Proof. (a)

$$\begin{aligned}\frac{\delta A_m}{\delta J_n} &= \left(\frac{\delta}{\delta J_1}\right)^m A_n^S\{J\} \\ &= \left(\frac{\delta}{\delta J_1}\right)^m\left(\frac{\delta}{\delta f}\right)^n \exp[G\{J+f\} - G\{J\}]\,|_{f=0}\end{aligned}$$

[by (2.12)]

$$\begin{aligned}&= \left(\frac{\delta}{\delta g}\right)^m\left(\frac{\delta}{\delta f}\right)^n \exp[G\{J+g+f\} - G\{J+g\}]\,|_{f=g=0} \\ &= \left(\frac{\delta}{\delta g}\right)^m\left(\frac{\delta}{\delta f}\right)^n \exp[G\{J+f\} - G\{J\}] \\ &\quad\times\exp[G\{J+g+f\} - G\{J+g\} - G\{J+f\} + G\{J\}]\,|_{f=g=0}\,,\end{aligned}$$

which yields (3.4) by the definition (3.3) of S. To derive (3.5) we need only apply Lemma II.2a and the differentiation formula (2.8) for F.

(b) When $1\leqslant m,n\leqslant N$, S_{km} is zero unless $k\geqslant 1$ and $\delta F_n/\delta A_k$ is zero unless $k\leqslant n\leqslant N$ so that (3.5) becomes

$$\begin{aligned}\frac{\delta A_m}{\delta J_n} &= \sum_{k=1}^{N} S_{mk}\frac{\delta F_n}{\delta A_k} \\ &= \sum_{k=1}^{N} S_{mk}\left(\frac{\delta F}{\delta A}\right)^t_{kn}\,. \qquad \blacksquare\end{aligned}$$

Theorem III.3. If $x^{(i)}$ denotes $(x_1,\ldots,x_i)\in\mathbb{R}^{id}$

$$\Gamma^{(N)}_{A_i(x^{(i)})} = -J_k\{A\}\frac{\delta F_k\{A+A^0\}}{\delta A_i(x^{(i)})}$$
$$= -\sum_{k=i}^{N}\binom{k}{i}J_k(x^{(i)},u^{(k-i)})(F_{k-i}\{A+A^0\})(u^{(k-i)}), \tag{3.6a}$$

$$\Gamma^{(N)}_{A_iA_j} = \binom{i+j}{i}\Gamma^{(N)}_{A_{i+j}} - (S^{(N)})^{-1}_{ij}, \tag{3.6b}$$

where the argument of $\Gamma^{(N)}_{A_{i+j}}$ is the union $(x^{(i)}, y^{(j)})$ of the arguments $x^{(i)}$ and $y^{(j)}$ of $\Gamma^{(N)}_{A_i}$ and $\Gamma^{(N)}_{A_j}$, respectively. Furthermore, $\Gamma^{(N)}_{A_{i+j}}\equiv 0$ if $i+j>N$.

Remark 4. Parts (a) and (b) are the analogs for Γ of the conjugate relation ${}^S\Gamma_A = -J$ and the Jacobian relation ${}^S\Gamma_{AA} = -G_{JJ}^{-1}$, respectively. In particular, when $J=0$, (3.6) reads $\Gamma^{(N)}_{A_iA_j} = -(S^{(N)})^{-1}_{ij}$.

Proof. (a) The first part is an immediate consequence of

$$\Gamma^{(N)}\{A\} = G\{J\{A\}\} - J_k\{A\}F_k\{A+A_0\}.$$

To prove the second part we observe that, by the differentiation formula (2.8) we must have $k\geqslant i$ and then

$$J_k\{A\}\frac{\delta F_k}{\delta A_i} = \binom{k}{i}J_k(u^{(k)})\Big(\prod_{r=1}^{i}\delta(u_r-x_r)\Big)F_{k-i}(u^{(k-i)})$$
$$= \binom{k}{i}J_k(x^{(i)},u^{(k-i)})F_{k-i}(u^{(k-i)}). \tag{3.7}$$

Note that the symmetrization convention implicit in (2.8) is irrelevant since we are integrating against a symmetric function J_k.

(b) Differentiating part (a) yields

$$\Gamma^{(N)}_{A_iA_j} = -\sum_{k=i}^{N}\binom{k}{i}J_k(x^{(i)},u^{(k-i)})\frac{\delta F_{k-i}(u^{(k-i)})}{\delta A_j(y^{(j)})}$$
$$-\frac{\delta F_k}{\delta A_i}\frac{\delta J_k\{A\}}{\delta A_j}$$
$$= -\sum_{k=i+j}^{N}\binom{k}{i}\binom{k-i}{j}J_k(x^{(i)},y^{(j)},u^{(k-i-j)})$$
$$\times F_{k-i-j}(u^{(k-i-j)})$$
$$-\Big(\frac{\delta F}{\delta A}\Big)^t_{ik}\frac{\delta J_k}{\delta A_j}$$

by the same calculation as in (3.7). That the sum is $\Gamma_{A_{i+j}}$ is the content of part (a). That $(\delta F/\delta A)^t_{ik}(\delta J_k/\delta A_j)$ is $S^{(N)^{-1}}_{ij}$ is a consequence of the Jacobian relationship $\delta J/\delta A = (\delta A/\delta J)^{-1}$ and Lemma III.2b. Note in particular that $\delta A/\delta J$ is invertible by hypothesis, $\delta F/\delta A$ is invertible since it is a triangular matrix with identities on the diagonal, and hence $S^{(N)}$ is invertible by Lemma III.2b. ■

By Theorems III.1, III.3, and III.7 we may reduce the study of the irreducibility properties of Γ (see I and Ref. 5) to the study of the connectedness properties of functionals like S_{ij} that are defined in terms of $G\{J\}$. The connectedness of G itself is easy to establish from its decomposition property [see (1.12)]. S inherits from G a similar decomposition property. It may be stated fairly simply if (when $t=0$) we view S, $S^{(N)}$, and $S^{(N)^{-1}}$ all as operators on spaces of the form

$$\mathscr{H} = \bigoplus_{i=0}^{\infty}\mathscr{H}^{(i)},$$

where $\mathscr{H}^{(0)} = \mathbb{C}$, $\mathscr{H}^{(1)} = \mathscr{H}^{(1)}_+ \oplus \mathscr{H}^{(1)}_-$ is a space of functions on $\mathbb{R}^d$, and $\mathscr{H}^{(i)}\equiv(\mathscr{H}^{(1)})^{\otimes_s^i}$ is a space of symmetric functions on $\mathbb{R}^{di}$. $\mathscr{H}^{(1)}$ is the space of functions from which J_1 or A_1 is chosen. We assume $\mathscr{H}^{(1)}$ is composed of the two pieces.

$\mathscr{H}^{(1)}_\pm$ = {those functions in $\mathscr{H}^{(1)}$ supported on $\mathbb{R}^d_\pm$ }, where $\mathbb{R}^d_\pm$ denotes the $\pm$ side of σ. Similarly $\mathscr{H}$, the space from which J or A is chosen, is naturally composed of four pieces:

$$\mathscr{H} = \mathbb{C}\oplus\mathscr{H}_+\oplus\mathscr{H}_-\oplus\mathscr{H}_0,$$

where

$$\mathscr{H}_\pm = \bigoplus_{i=1}^{\infty}(\mathscr{H}^{(1)}_\pm)^{\otimes_s^i}.$$

We denote by P_+, P_-, and P_0 the projectors on $\mathscr{H}_+$, $\mathscr{H}_-$, and $\mathscr{H}_0$, respectively.

When $t=0$ and $P_0J=0$, henceforth referred to as "at 0", the expectation $\langle\cdot\rangle$ factors [see (1.11)] yielding the decomposition property

$$G\{J\} = G\{P_+J\} + G\{P_-J\} \tag{3.8}$$

of G. Hence

$$G\{J+f\} = G\{J+f\chi_+\} + G\{J+f\chi_-\} - G\{J\}$$

so that

$$\mathscr{S}\{f,g\} = \mathscr{S}\{f\chi_+,g\chi_+\}\mathscr{S}\{f\chi_-,g\chi_-\} \tag{3.9}$$

and

$$S_{ij} = S_{i_+j_+}S_{i_-j_-} \tag{3.10a}$$
$$= \sum_{i'=0}^{i}\sum_{j'=0}^{j}\binom{i}{i'}\binom{j}{j'}\underset{+}{S}_{i'j'}\underset{-}{S}_{i-i',j-j'} \tag{3.10b}$$

by the product rule for differentiation. Here i_+ denotes those arguments of i that are in $\mathbb{R}^d_+$; $\underset{+}{S}$ denotes the restriction of S to $\mathbb{C}\oplus\mathscr{H}_+$. We impose the convention that $\underset{+}{S}_{ij}$ is defined for all i and j (regardless of which side of σ they are on) but is zero unless all the arguments of i and j are in $\mathbb{R}^d_+$. Hence all but one of the terms in the expanded symmetrized version of (3.10b) are zero.

The formulas (3.10) represent S as a tensorlike product of $\underset{+}{S}$ and $\underset{-}{S}$. It is only the symmetrization that prevents $\mathscr{H}$ from being the tensor product $(\mathbb{C}\oplus\mathscr{H}_+)\otimes(\mathbb{C}\oplus\mathscr{H}_-)$. Accordingly we identify each $f(x)\epsilon\mathscr{H}^{(i)}_+\otimes\mathscr{H}^{(j)}_-$ with its symmetric extension to

$$\bar{f}(x) = \binom{i+j}{i}(S_xf)(x)\epsilon\mathscr{H}^{(i+j)}$$

(where S_x denotes symmetrization with respect to x). Note that if $x^{(i)}\epsilon\mathbb{R}^d_+$ and $y^{(j)}\epsilon\mathbb{R}^d_-$ we have, as desired,

$$\bar{f}(x^{(i)},y^{(j)}) = \binom{i+j}{i}(Sf)(x^{(i)},y^{(j)})$$
$$= \binom{i+j}{i}\frac{1}{(i+j)!}\sum_{\pi\epsilon S_{i+j}}f(\pi(x^{(i)},y^{(j)}))$$
$$= f(x^{(i)},y^{(j)})$$

since there are precisely $i!j!$ nonzero $f(\pi(x^{(i)},y^{(j)}))$'s and each one equals $f(x^{(i)},y^{(j)})$. We extend this identification to all

$f\in(\mathbb{C}\oplus\mathscr{H}_+)\otimes(\mathbb{C}\oplus\mathscr{H}_-)$. It is then natural to define the "tensor" product $\underset{+}{A}\times\underset{-}{A}:\mathscr{H}\to\mathscr{H}$ of two operators

$\underset{\pm}{A}:\mathbb{C}\oplus\mathscr{H}_\pm\to\mathbb{C}\oplus\mathscr{H}_\pm$ by

$$\underset{+}{A}\times\underset{-}{A}\bar{f}=\overline{\underset{+}{A}\otimes\underset{-}{A}f}. \tag{3.11a}$$

The kernel of $\underset{+}{A}\times\underset{-}{A}$ is given by

$$\Big(\underset{+}{A}\times\underset{-}{A}\Big)_{mn}(x^{(m)},y^{(n)}) = S_xS_y\sum_{i=0}^{m}\sum_{j=0}^{n}\binom{m}{i}\underset{+}{A}_{ij}(x^{(i)},y^{(j)})\underset{-}{A}_{m-i,n-j}(x^{(m/i)},y^{(n/j)}), \tag{3.11b}$$

where $x^{(i)}=\{x_1,\dots,x_i\}$ and $x^{(m/i)}=\{x_{i+1},\dots,x_m\}$. As a result of (3.11a) this "tensor" product enjoys the algebraic properties one expects—namely linearity in both arguments and the product rule

$$\Big(\underset{+}{A}\times\underset{-}{A}\Big)\Big(\underset{+}{B}\times\underset{-}{B}\Big)=\Big(\underset{+}{A}\underset{+}{B}\Big)\times\Big(\underset{-}{A}\underset{-}{B}\Big). \tag{3.12}$$

The decomposition property (3.10) of S may be rewritten in terms of the tensor product as

$$S=\Big[\Big(\underset{+}{S}F^{-1}\Big)\times\Big(\underset{-}{S}F^{-1}\Big)\Big]F \text{ at } 0,$$

where F is the (infinite) diagonal matrix with $F_{ii}=i!$. From this we have immediately that

$$S^{-1}=F^{-1}\Big[\Big(F\underset{+}{S}^{-1}\Big)\times\Big(F\underset{-}{S}^{-1}\Big)\Big]. \tag{3.13}$$

However, (3.13) is of little value since it is $S^{(N)^{-1}}$ that we are primarily interested in and the projections that restrict S to $S^{(N)}$ do not commute with either taking tensor products or inverses. [We remind the reader that, to view $S^{(N)}$ and $(S^{(N)})^{-1}$ as operators on $\mathscr{H}$ we here impose the conventions $S^{(N)}_{ij}=(S^{(N)})^{-1}_{ij}=0$ unless $1\leqslant i,j\leqslant N$.] Nonetheless, $S^{(N)^{-1}}$ does have a simple decomposition property:

Theorem III. 4. At 0

$$S=\Big[\Big(\underset{+}{S}F^{-1}\Big)\times\Big(\underset{-}{S}F^{-1}\Big)\Big]F. \tag{3.14}$$

$$S^{(N)^{-1}}=F^{-1}\Big\{\sum_{m=0}^{N}\Big(F\underset{+}{S}^{(m)^{-1}}\Big)\times\Big(F\underset{-}{S}^{(N-m)^{-1}}\Big) - \sum_{m=1}^{N-2}\Big(F\underset{+}{S}^{(m)^{-1}}\Big)\times\Big(F\underset{-}{S}^{(N-m-1)^{-1}}\Big)\Big\}, \tag{3.15}$$

where the last sum is zero if $N\leqslant 2$ and $S^{(0)}_{ij}=(S^{(0)})^{-1}_{ij}=\delta_{i,0}\delta_{j,0}$.

Example.

$$P_0S^{(3)^{-1}}=F^{-1}\Big\{\underset{+}{S}^{(1)^{-1}}\times\Big(F\underset{-}{S}^{(2)^{-1}}\Big)+\Big(F\underset{+}{S}^{(2)^{-1}}\Big)\times\underset{-}{S}^{(1)^{-1}} - \underset{+}{S}^{(1)^{-1}}\times\underset{-}{S}^{(1)^{-1}}\Big\},$$

where by (3.3) $(S^{(1)})^{-1}_{ij}=G_2^{-1}$ if $i=j=1$ and zero otherwise, and

$$(S^{(2)})^{-1}_{ij}=\begin{cases}\begin{pmatrix}G_2 & G_3\\ G_3 & G_4+2G_2^2\end{pmatrix}^{-1}_{ij}, & \text{if } 1\leqslant i,j\leqslant 2\\ 0, & \text{otherwise.}\end{cases}$$

Proof. Let $\pi^{(m)}_\pm$ be the projection on $\mathscr{H}^{(m)}_\pm$, $P^{(m)}_\pm=\sum_{j=1}^{m}\pi^{(j)}_\pm$ be the projection on $\mathscr{H}^{(1)}_\pm\oplus\cdots\oplus\mathscr{H}^{(m)}_\pm$,

$$P^{(N)}=\sum_{\substack{m,n>0\\1\leqslant m+n\leqslant N}}\pi^{(m)}_+\times\pi^{(n)}_-$$

be the projection on $\mathscr{H}^{(1)}\oplus\cdots\oplus\mathscr{H}^{(N)}$, and $P^{(0)}=P^{(0)}_\pm=\pi^{(0)}_\pm$ be the projection on $\mathbb{C}$. We must verify that multiplying $P^{(N)}SP^{(N)}$ by the right hand side of (3.15) yields $P^{(N)}$. Now

$$\begin{aligned}P^{(N)}SP^{(N)}F^{-1}&\Big\{\sum_{m=0}^{N}\Big(F\underset{+}{S}^{(m)^{-1}}\Big)\times\Big(F\underset{-}{S}^{(N-m)^{-1}}\Big) - \sum_{m=1}^{N-2}\Big(F\underset{+}{S}^{(m)^{-1}}\Big)\times\Big(F\underset{-}{S}^{(N-m-1)^{-1}}\Big)\Big\}\\ &=P^{(N)}\Big\{\sum_{m=0}^{N}\Big(\underset{+}{S}\underset{+}{S}^{(m)^{-1}}\Big)\times\Big(\underset{-}{S}\underset{-}{S}^{(N-m)^{-1}}\Big) - \sum_{m=1}^{N-2}\Big(\underset{+}{S}\underset{+}{S}^{(m)^{-1}}\Big)\times\Big(\underset{-}{S}\underset{-}{S}^{(N-m-1)^{-1}}\Big)\Big\}\end{aligned} \tag{3.16}$$

by (3.12) and (3.14). The crux of this proof is the observation that while $\underset{\pm}{S}\underset{\pm}{S}^{(m)^{-1}}$ is not $P^{(m)}_\pm$ the error "lives"in $\mathscr{H}^{(m+1)}_\pm\oplus\mathscr{H}^{(m+2)}_\pm\oplus\cdots$. To be precise

$$\underset{\pm}{S}\underset{\pm}{S}^{(m)^{-1}}=P^{(m)}_\pm+E^{(m+1)}_\pm, \tag{3.17}$$

where the error satisfies

$$P^{(m)}_\pm E^{(m+1)}_\pm=P^{(m)}_\pm\Big(\underset{\pm}{S}-\underset{\pm}{S}^{(m)}\Big)\underset{\pm}{S}^{(m)^{-1}}=0.$$

Substituting (3.17) into the first sum in (3.16) gives

$$P^{(N)}\sum_{m=0}^{N}(P^{(m)}_+ + E^{(m+1)}_+)\times(P^{(N-m)}_- + E^{(N-m+1)}_-).$$

Many of the resulting terms in this sum are annihilated by $P^{(N)}$ because if $(P^{(0)}_+ + P^{(m-1)}_+)\underset{+}{A} = (P^{(0)}_- + P^{(N-m)}_-)\underset{-}{A}=0$ for some $m=1,\dots,N$ (so that $\underset{+}{A}\times\underset{-}{A}$ lives in $\mathscr{H}^{(N+1)}\oplus\mathscr{H}^{(N+2)}\oplus\cdots$) then $P^{(N)}\underset{+}{A}\times\underset{-}{A}=0$. We thus obtain

$$\sum_{m=0}^{N}P^{(m)}_+\times P^{(N-m)}_- + P^{(N)}\sum_{m=1}^{N-2}E^{(m+1)}_+\times P^{(N-m-1)}_- + P^{(N)}\sum_{m=2}^{N-1}P^{(m-1)}_+\times E^{(N-m+1)}_-, \tag{3.18a}$$

where we have dropped the terms involving $E^{(1)}_\pm=0$ and all those terms annihilated by $P^{(N)}$, namely $E^{(m+1)}_+\times P^{(N-m)}_-$ $(m=N-1,N)$, $P^{(m)}_+\times E^{(N-m+1)}_-$ $(m=0,1)$, and $E^{(m+1)}_+\times E^{(N-m+1)}_-$, $E^{(m+1)}_+\times\pi^{(N-m)}_-$, and $\pi^{(m)}_+\times E^{(N-m+1)}_-$ (all m). For the second sum in (3.16) we drop the $E_+\times E_-$ terms to obtain

$$\sum_{m=1}^{N-2}P^{(m)}_+\times P^{(N-m-1)}_- + P^{(N)}\sum_{m=1}^{N-2}E^{(m+1)}_+\times P^{(N-m-1)}_- + P^{(N)}\sum_{m=1}^{N-2}P^{(m)}_+\times E^{(N-m)}_-. \tag{3.18b}$$

The last two sums in (3.18a) cancel those in (3.18b) and so

$$(3.16) = \sum_{m=0}^{N} P_+^{(m)} \times P_-^{(N-m)} - \sum_{m=1}^{N-2} P_+^{(m)} \times P_-^{(N-m-1)}$$
$$= P_+^{(0)} \times P_-^{(N)} + \sum_{m=1}^{N-1} P_+^{(m)} \times \pi_-^{(N-m)} + P_+^{(N)} \times P_-^{(0)}$$
$$= P^{(N)}. \qquad \blacksquare$$

Remark 5. Theorem III.4 expresses $S^{(N)^{-1}}$ as a sum of terms each of which depends on the variables $\chi^i_+ A_i$ (or $\chi^i_- A_i$) only through $\underset{+}{S}^{(m)}$ ($\underset{-}{S}^{(m)}$) for some $m \leqslant N/2$. However, for $m \leqslant N/2$,

$$S_{ij}^{(m)} = \left(\frac{\delta}{\delta f}\right)^i \left(\frac{\delta}{\delta g}\right)^j \times \exp\left\{\sum_{h,k=1}^{m} (A_{h+k} + A^0_{h+k}) \frac{f^h}{h!} \frac{g^k}{k!}\right\}\Big|_{f=g=0}$$

is a polynomial in A_l for $l = 2,3,...,N$ and hence is independent of A_1. Hence each term in (3.15) is independent of one of $\chi_+ A_1$ and $\chi_- A_1$. This fact leads to the two-irreducibility of the vertex functions $\Gamma_n^{(N)}$ for $N \geqslant 2$.

We may also use S to give a convenient evaluation of $\dot{A}^0_n$.

Theorem III.5.

$$P_0 \dot{A}^0_n (x^{(n)})|_{t=0} = -\tfrac{1}{2} P_0 S_{n2}(x^{(n)}, y^{(2)}) \dot{C}^{-1}(y_1,y_2)|_{t=J=0}. \tag{3.19}$$

Remark. We remind the reader that the effect of the projection P_0 is to ensure that not all of $\{x_1,...,x_n\}$ are on the same side of σ.

Proof. Suppose $t = 0$ and not all of $\{x_1,...,x_n\}$ are on the same side of σ.

$$\begin{aligned}
\dot{A}^0_n &= \dot{G}_{J_1^n}\{J=0\} \\
&= -\tfrac{1}{2}\dot{C}^{-1}(y^{(2)}) G_{J_2}(y^{(2)})_{J_1^n} \quad \text{[by (3.2)]} \\
&= -\tfrac{1}{2}\dot{C}^{-1}\{G_2 + G_1 G_1\}_{J_1^n} \quad \text{(by Lemma II.2)} \\
&= -\tfrac{1}{2}\dot{C}^{-1}(y_1,y_2)\{G_1(y_1) G_1(y_2)\}_{J_1^n} \quad \text{(since } G \text{ is connected).}
\end{aligned}$$

When y_1 and y_2 are on the same side of σ the connectedness of G causes this as well as the right hand side of (3.19) to vanish. So to complete the proof we need only show that when y_1 and y_2 are separated by σ

$$\frac{\delta}{\delta f(y_1)} \frac{\delta}{\delta f(y_2)} \mathscr{S}\{f,g\}\Big|_{f=J=0} = \prod_{i=1}^{2} [G_1(y_1) - A^0_1(y_i)]\Big|_{J=(0,g,0,0,...)}.$$

(The A^0_1 terms are eliminated by the J_1 derivatives.) But when $J = 0$ (3.8) implies

$$\begin{aligned}
&\frac{\delta}{\delta f(y_1)} \frac{\delta}{\delta f(y_2)} \mathscr{S}\{f,g\}\Big|_{f=0} \\
&= \prod_{i=1}^{2} \left(\frac{\delta}{\delta f(y_i)}\right) \exp\{G\{f+g\} - G\{f\} - G\{g\}\}|_{f=0} \\
&= \prod_{i=1}^{2} (G_{J_1(y_i)}\{f+g\} - G_{J_1(y_i)}\{f\}) \\
&\quad \times \exp\{G\{f+g\} - G\{f\} - G\{g\}\}|_{f=0} \\
&= \prod_{i=1}^{2} [G_{J_1(y_i)}\{g\} - A^0_i(y_i)]. \qquad \blacksquare
\end{aligned}$$

We now continue with the project, started in Theorem III.3, of expressing the moments $\Gamma_{A_{i_1} \cdots A_{i_m}}$ of Γ in terms of G_k. While the case $m > 2$ is conceptually similar to that of Theorem III.3 it is characterized by a richness of obscuring detail. Since this material will not be used later in this paper we suggest that the reader omit it (at least on a first reading) and go directly to our analysis of the N-field projections and generalized Bethe–Salpeter kernels in Sec. IV and V, respectively.

In Theorem III.7 we will express $\Gamma_{A_{i_1} \cdots A_{i_m}}$ (for $m \geqslant 3$) in terms of tensors S_{ij} (of degree 2), $T^{(n)}_{i_1 \cdots i_n}$ (of degree $n \geqslant 3$), and $U^{(n,m)}_{i_1 \cdots i_{n+m}}$ (of degree $n + m$ with $n \geqslant 3$, $m \geqslant 0$). When we say, for example, that $T^{(n)}$ is of degree n we are referring only to the fact that $T^{(n)}$ has n indices $i_1 \cdots i_n$. As with S_{ij}, each index i_k in $T^{(n)}$ (and $U^{(n,m)}$) is an abbreviation for an i_k-tuple of points in $\mathbb{R}^d$ so that $T^{(n)}_{i_1 \cdots i_n}$ is a "function" on $\mathbb{R}^{i_1 d} \times \cdots \times \mathbb{R}^{i_n d}$. We now define $T^{(n)}$ and $U^{(n,m)}$. The most significant feature of these definitions (see Remark 6) is the fact that $T^{(n)}$ and $U^{(n,m)}$ are polynomials in the S_{ij}'s, which in turn are polynomials in G_k. The appropriate generating functionals are defined inductively by

$$\mathscr{T}^{(3)}\{f,g,h\} = [\mathscr{S}\{h,g+f\} - \mathscr{S}\{h,f\} - \mathscr{S}\{h,g\} + 1]\mathscr{S}\{f,g\}, \tag{3.20}$$

$$\mathscr{T}^{(n)}\{f_1,...,f_n\} = \mathscr{D}_{f_n} \mathscr{T}^{(n-1)}\{f_1,...,f_{n-1}\} \quad \text{for } n > 3, \tag{3.21}$$

$$\mathscr{U}^{(n,0)}\{f_1,...,f_n\} = -\prod_{i=1}^{n-1} \mathscr{S}\{f_i f_n\}, \tag{3.22}$$

$$\mathscr{U}^{(n,m)}\{f_1,...,f_{n+m}\} = \mathscr{D}_{f_{n+m}} \mathscr{U}^{(n,m-1)}\{f_1,...,f_{n+m-1}\} \text{ for } m > 0, \tag{3.23}$$

where the "derivation" $\mathscr{D}_h$ is defined by

$$\mathscr{D}_h \mathscr{A}(\{S_{ij}\} f_1 f_2 \cdots) = \sum_{i,j,k=0}^{\infty} \frac{\delta \mathscr{A}}{\delta S_{ij}} T^{(3)}_{ijk} \frac{h^k}{k!}. \tag{3.24}$$

The evaluation of $\mathscr{D}_h$ derivatives is straightforward if one uses the following properties:

$$\mathscr{D}_h \mathscr{S}\{f,g\} = \mathscr{T}^{(3)}\{f,g,h\}, \tag{3.25a}$$

$$\mathscr{D}_h(a\mathscr{A} + b\mathscr{B}) = a\mathscr{D}_h \mathscr{A} + b\mathscr{D}_h \mathscr{B}, \tag{3.25b}$$

$$\mathscr{D}_h(\mathscr{A}\mathscr{B}) = \mathscr{A}\mathscr{D}_h \mathscr{B} + \mathscr{B}\mathscr{D}_h \mathscr{A}. \tag{3.25c}$$

For example,

$$\mathscr{U}^{(n,1)}\{f_1,\cdots f_{n+1}\} = -\sum_{i=1}^{n-1} \Bigg(\prod_{\substack{1 \leqslant j \leqslant n-1 \\ j \neq i}} \mathscr{S}\{f_j f_n\}\Bigg) \mathscr{T}^{(3)}\{f_i f_n f_{n+1}\}.$$

Then $\mathscr{U}$ and $\mathscr{T}$ act as generating functions for U and T:

$$T^{(n)}_{i_1 \cdots i_n} = \left[\prod_{j=1}^{n} \left(\frac{\delta}{\delta f_j}\right)^{i_j}\right] \mathscr{T}^{(n)}\{f_1 \cdots f_n\}|_{f_i=0},$$

$$U^{(n,m)}_{i_1 \cdots i_{n+m}} = \left[\prod_{j=1}^{n+m} \left(\frac{\delta}{\delta f_j}\right)^{i_j}\right] \mathscr{U}^{(n,m)}\{f_1 \cdots f_{n+m}\}|_{f_i=0}.$$

For example, since

$$\mathscr{S}\{h,g+f\} - \mathscr{S}\{h,f\} - \mathscr{S}\{h,g\} + 1 = \sum_{\substack{i,j>1\\ k>0}} S_{i+j,k} \frac{f^i}{i!}\frac{g^j}{j!}\frac{h^k}{k!}$$

we have

$$T^{(3)}_{ijk} = \sum_{i'=1}^{i} \sum_{j'=1}^{j} \binom{i}{i'}\binom{j}{j'} S_{i-i',j-j'} S_{k,i'+j'} \tag{3.26}$$

and

$$U^{(n,0)}_{i_1\cdots i_n} = - \sum_{\substack{i'_1,\ldots,i'_{n-1}\\ i'_1+\cdots+i'_{n-1}=i_n}} \binom{i_n}{i'_1,\ldots,i'_{n-1}} S_{i_1,i'_1}\cdots S_{i_{n-1},i'_{n-1}}. \tag{3.27}$$

Remark 6. $T^{(n)}_{i_1\cdots i_{n+m}}$ is a linear combination of terms of the form $S_{j_1k_1}S_{j_2k_2}\cdots S_{j_{n-1},k_{n-1}}$ with each point of $i_1\cup\cdots\cup i_n$ occurring precisely once in $j_1\cup\cdots\cup j_{n-1}\cup k_1\cup\cdots\cup k_{n-1}$ and vice versa. (It is possible for some of the j's and k's to be empty.) The structure of $U^{(n,m)}$ is the same as that of $T^{(n)}$ but with degree of homogeneity (in S) $n+m-1$ instead of $n-1$. The basis for these definitions is

Lemma III.6. For $1\leqslant k\leqslant N$

(a) $\dfrac{\delta}{\delta A_k} S_{ij}\{J\{A\}\} = T^{(3)}_{ijk'} S^{(N)^{-1}}_{k'k}$,

(b) $\dfrac{\delta}{\delta A_k} T^{(n-1)}_{i_1\cdots i_{n-1}}\{J\{A\}\} = T^{(n)}_{i_1\cdots i_{n-1}k'} S^{(N)^{-1}}_{k'k}$ for $n\geqslant 4$,

(c) $\dfrac{\delta}{\delta A_k} U^{(n,m-1)}_{i_1\cdots i_{n+m-1}}\{J\{A\}\} = U^{(n,m)}_{i_1\cdots i_{n+m-1}k'} S^{(N)^{-1}}_{k'k}$

for $n\geqslant 3$, $m\geqslant 1$.

Proof. (a) Firstly,

$$\begin{aligned}\frac{\delta \mathscr{S}\{f,g\}}{\delta A_k} &= \mathscr{S}\{f,g\} \sum_{i',j'>1} \frac{\delta}{\delta A_k}\left[G_{i'+j'}\{J\{A\}\}\frac{f^{i'}}{i'!}\frac{g^{j'}}{j'!}\right]\\
&= \mathscr{S}\{f,g\} \sum_{\substack{i',j'>1\\ 1<n<N}} \frac{\delta G_{i'+j'}}{\delta J_n}\frac{\delta J_n}{\delta A_k}\frac{f^{i'}}{i'!}\frac{g^{j'}}{j'!}\\
&= \mathscr{S}\{f,g\} \sum_{\substack{i',j'>1\\ 1<n,k'<N}} S_{k',i'+j'} \frac{\delta F_n}{\delta A_{k'}}\frac{\delta J_n}{\delta A_k}\frac{f^{i'}}{i'!}\frac{g^{j'}}{j'!} \quad \text{[by (3.5)]}\\
&= \mathscr{S}\{f,g\}\sum S_{k',i'+j'} S^{(N)^{-1}}_{k',k}\frac{f^{i'}}{i'!}\frac{g^{j'}}{j'!}\end{aligned}$$

by the Jacobian relationship $\delta J/\delta A = (\delta A/\delta J)^{-1}$ and Lemma III.2b. Hence

$$\begin{aligned}\frac{\delta}{\delta A_k} S_{ij} &= \left(\frac{\delta}{\delta f}\right)^i\left(\frac{\delta}{\delta g}\right)^j \frac{\delta\mathscr{S}(f,g)}{\delta A_k}\Bigg|_{f=g=0}\\
&= \sum_{\substack{1<i'<i\\ 1<j'<j\\ 1<k'<N}} \binom{i}{i'}\binom{j}{j'} S_{i-i',j-j'} S_{k',i'+j'} S_{k',i'+j'} S^{(N)^{-1}}_{k'k}\end{aligned}$$

as desired.

(b) Think of $T^{(n-1)}_{i_1\cdots i_{n-1}}\{J\{A\}\}$ as a function of $\{S_{ij}\}$ each of which is in turn a function of $J\{A\}$. Then, by the chain rule

$$\begin{aligned}\frac{\delta}{\delta A_k} T^{(n-1)}_{i_1\cdots i_{n-1}} &= \frac{\delta}{\delta S_{ij}} T^{(n-1)}_{i_1\cdots i_{n-1}} \frac{\delta S_{ij}}{\delta A_k}\\
&= \frac{\delta}{\delta S_{ij}} T^{(n-1)}_{i_1\cdots i_{n-1}} T^{(3)}_{ijk'} S^{(N)^{-1}}_{k'k} \quad \text{[by part (a)]}\\
&= T^{(n)}_{i_1\cdots i_{n-1}k'} S^{(N)^{-1}}_{k'k} \quad \text{[by (3.21) and (3.24)]}.\end{aligned}$$

(c) The proof is virtually identical to that of part (b). ■

We are now in a position to complete the project commenced in Theorem III.3.

Theorem III.7.

$$\begin{aligned}\text{(a)}\ \binom{i_1+i_2+\cdots+i_{n-1}}{i_1,i_2,\ldots,i_{n-1}} \Gamma^{(N)}_{A_{i_1+\cdots+i_{n-1}}A_{i_n}} &= \binom{i_1+\cdots+i_n}{i_1,\ldots,i_n}\Gamma^{(N)}_{A_{i_1+\cdots+i_n}}\\
&\quad + \left(\overset{n}{\otimes} S^{(N)^{-1}} U^{(n,0)}\right)_{i_1\cdots i_n},\end{aligned}$$

where each $S^{(N)^{-1}}$ in $\overset{n}{\otimes} S^{(N)^{-1}}$ acts on the corresponding index if $U^{(n,0)}$.

$$\begin{aligned}\text{(b)}\ \Gamma^{(N)}_{A_iA_jA_k} &= \binom{i+j}{i}\Gamma^{(N)}_{A_{i+j}A_k} + \left(\overset{3}{\otimes} S^{(N)^{-1}} T^{(3)}\right)_{ijk}\\
&= \binom{i+j+k}{i,j,k}\Gamma^{(N)}_{A_{i+j+k}} + \left(\overset{3}{\otimes} S^{(N)^{-1}} T^{(3)}\right)_{ijk}\\
&\quad + \left(\overset{3}{\otimes} S^{(N)^{-1}} U^{(3,0)}\right)_{ijk}.\end{aligned}$$

(c) For $n\geqslant 3$

$$\begin{aligned}\Gamma^{(N)}_{A_{i_1}\cdots A_{i_n}} &= \binom{i_1+\cdots+i_n}{i_1,\ldots,i_n}\Gamma^{(N)}_{A_{i_1+\cdots+i_n}}\\
&\quad + \sum_{p=3}^{P(i_1\cdots i_n)} \left[\left(\overset{p}{\otimes} S^{(N)^{-1}} U^{(p,0)}\right)_{i_1\cdots i_p}\right]_{A_{i_{p+1}}\cdots A_{i_n}}\\
&\quad + \left[\left(\overset{3}{\otimes} S^{(N)^{-1}} T^{(3)}\right)_{i_1i_2i_3}\right]_{A_{i_4}\cdots A_{i_n}},\end{aligned}$$

where $P(i_1\cdots i_n) = \max\{p \mid i_1+\cdots+i_{p-1}\leqslant N, p\leqslant n\}$.

Remark 7. Evaluation of the remaining A_{i_α} derivatives in part (c) by means of Lemma III.6 results in a sum of terms each of which may be viewed as a tree graph with

(i) n external vertices each associated with a different i_α,

(ii) each line associated with $S^{(N)^{-1}}$,

(iii) each vertex having m legs (m is always at least 3) associated with either $T^{(m)}$ or $U^{(p,m-p)}$ (with $p\geqslant 3$, $m-p\geqslant 0$).

Proof. (a) By Theorem III.3b

$$\begin{aligned}&\left[\begin{matrix}i_1+i_2+\cdots+i_{n-1}\\ i_1,i_2,\ldots,i_{n-1}\end{matrix}\right]\Gamma^{(N)}_{A_{i_1+\cdots+i_{n-1}}A_{i_n}} - \binom{i_1+\cdots+i_n}{i_1,\ldots,i_n}\Gamma^{(N)}_{A_{i_1+\cdots+i_n}}\\
&= -\binom{i_1+\cdots+i_{n-1}}{i_1,\ldots,i_{n-1}} S^{(N)^{-1}}_{i_1+\cdots+i_{n-1},i_n}.\end{aligned}$$

Applying $\overset{n}{\otimes} S^{(N)}$ yields, for the right-hand side,

$$S^{(N)}_{j_1 i_1}\cdots S^{(N)}_{j_{n-1}i_{n-1}}S^{(N)}_{j_n i_n}$$
$$\times\left\{-\binom{i_1+\cdots+i_{n-1}}{i_1,\ldots,i_{n-1}}S^{(N)^{-1}}_{i_1+\cdots+i_{n-1},i_n}\right\}$$
$$=S^{(N)}_{j_1 i_1}\cdots S^{(N)}_{j_{n-1}i_{n-1}}$$
$$\times\left\{-\binom{i_1+\cdots+i_{n-1}}{i_1,\ldots,i_{n-1}}P^{(N)}_{i_1+\cdots+i_{n-1},j_n}\right\}$$
$$=-\sum\binom{j_n}{i_1\cdots i_{n-1}}S_{j_1 i_1}\cdots S_{j_{n-1}i_{n-1}}=U^{(n,0)}_{j_1\cdots j_n},$$

where the sum runs over $i_1,\ldots,i_{n-1}\in\{1,\ldots,N\}$ with $i_1+\cdots+i_{n-1}=j_n$. We remind the reader that $P^{(N)}$ is the projection (i.e., identity) on $H^{(1)}\oplus\cdots\oplus H^{(N)}$. It is legitimate to drop the superscript N from $S^{(N)}$ since we are considering only $j_1\cdots j_{n-1}\in\{1,\ldots,N\}$.

(b) This is an immediate consequence of Theorem III.3(b), followed by Lemma III.6a, followed by part (a).

(c) This follows immediately by induction from parts (a) and (b). About the only point worth noting is that if $i_1+\cdots+i_{p-1}>N$ then by part (a)

$$\left(\overset{p}{\otimes}S^{(N)^{-1}}U^{(p,0)}\right)_{i_1\cdots i_p}=\text{const}\,\Gamma_{A_{i_1+\cdots+i_{p-1}}A_{i_p}}$$
$$-\text{const}\,\Gamma_{A_{i_1+\cdots+i_p}}=0.$$

Remark 8. We observed in Remark 7 that Theorem III.7 could be used to produce a tree graph expansion for $\Gamma^{(N)}_{A_{i_1}\cdots A_{i_n}}\{0\}$ having $S^{(N)^{-1}}$'s for lines and $T^{(m)}$'s and $U^{(p,m-p)}$'s for vertices. Now $T^{(m)}$ and $U^{(p,m-p)}$ are polynomials in S_{ij}, which is in turn a polynomial in G_k (with $k\geqslant 2$). Hence knowing the perturbation theory expansion for $G_k\{J=0\}$ yields the perturbation theory expansion for $\Gamma^{(N)}_{A_{i_1}\cdots A_{i_n}}\{A=0\}$. In any given order of perturbation theory then $T^{(m)}$ and $U^{(p,m-p)}$ may be represented as a sum of Schwinger functionlike graphs. By this we mean that the graphs are *not amputated* and that they need not be connected. On the other hand, in the free theory $S^{(N)}(\lambda=0)=\overline{\mathscr{C}}$, where $\overline{\mathscr{C}}_{ij}=\delta_{ij}i!\mathscr{C}^{\otimes i}$, so that

$$S^{(N)^{-1}}=\overline{\mathscr{C}}^{-1}+\sum_{n=1}^{\infty}\overline{\mathscr{C}}^{-1}(I-S^{(N)}C^{-1})^n.$$

Each order of the perturbation theory of $S^{(N)^{-1}}$ is not a sum of conventional graphs. Instead it is a sum of terms each of which is a tensor product $G^{\text{amp}}\otimes C^{-1}\otimes\cdots\otimes C^{-1}$ of an *amputated graph* G^{amp} and a number (possibly zero) of C^{-1}'s. A careful analysis of what happens when these expansions for $T^{(m)}$, $U^{(p,m-p)}$, and $S^{(N)^{-1}}$ are substituted into the tree graph expansion of Remark 7 yields the result that, in perturbation theory, $\Gamma^{(N)}_{A_{i_1}\cdots A_{i_n}}\{0\}$ may also be represented as a sum of terms of the form $G^{\text{amp}}\otimes C^{-1}\otimes\cdots\otimes C^{-1}$. Here G^{amp} is a conventional amputated graph. Furthermore the arguments x_1 and x_2 of any $C^{-1}(x_1,x_2)$ must be arguments from two different A_{i_α}'s. Finally there can be *no* C^{-1}'s in any graph in the perturbation theory expansion of $(\delta/\delta A_1)^k\Gamma^{(N)}|_{A=0}$ if $k\neq 2$. This is no surprise since $\Gamma^{(N)}_{A_1\cdots A_1}\{0\}$ is connected and every C^{-1} occurs as a distinct connected component of a graph. Hence we expect $\Gamma^{(N)}_{A_1\cdots A_1}\{0\}$ to have better regularity properties than a general $\Gamma^{(N)}_{A_{i_1}\cdots A_{i_n}}$.

IV. THE r-FIELD PROJECTIONS

The r-field projections for a Euclidean scalar boson field theory, whose infinite volume measure and expectation are denoted $d\nu$ and $\langle\cdot\rangle$, respectively, are defined as follows. If

$$H=L^2_{\text{real}}(d\nu),$$
$$H_{<0}=\mathbb{C}\subset H,$$
$$H_{<r}=\overline{\text{span}}\,\{\phi(f_1)\cdots\phi(f_n)\,|\,n\leqslant r, f_j\in C^\infty_{0,\text{real}}(\mathbb{R}^d)\}$$

then $P_{<r}$, the projection onto r or fewer fields, is simply the orthogonal projection onto $H_{<r}$. The r-field projection P_r is then defined by

$$P_r=\begin{cases}P_{<0}, & \text{if } r=0,\\ P_{<r}-P_{<(r-1)}, & \text{if } r>0.\end{cases}\tag{4.1}$$

We also let $P_{>r}=I-P_{<r}$. These projections have the basic properties

$$P_0\psi=\langle\psi\rangle,\tag{4.2a}$$
$$P_r\phi(x_1)\cdots\phi(x_m)=0\quad\text{if } m<r,\tag{4.2b}$$
$$P_{<r}=\sum_{n=0}^{r}P_n,\tag{4.2c}$$
$$I=\sum_{r=0}^{\infty}P_r\ \ (\text{assuming the } H_{<r}\text{'s span } H),\tag{4.2d}$$
$$P_{<r}P_{<s}=P_{<\min(r,s)},\tag{4.2e}$$
$$P_rP_s=\delta_{r,s}P_r,\tag{4.2f}$$

all of which follow immediately from the definitions.

The projections P_r for $r\leqslant 2$ have long been used to analyze the low mass spectrum of quantum field theory models.[7–9,11,16,17] Of particular interest have been the r-irreducible expectations

$$\langle\phi^i;\phi^j\rangle^{(r)}=\langle\phi^iP_{>r}\phi^j\rangle$$

used by Spencer and Zirilli[7,12] (for $r\leqslant 2$) to study the low mass spectrum of (even) $\epsilon P(\phi)_2$ models. The higher order projections (i.e., $r>2$) were introduced by Glimm and Jaffe.[13] In particular, the inverse of the operator $R^{(r)}$ given by the kernel

$$R^{(r)}(x^{(r)},y^{(r)})=\langle\phi^r(x^{(r)})P_r\phi^r(y^{(r)})\rangle$$

played a central role in their definition of the higher order Bethe–Salpeter kernels (see Sec. V). The irreducibility properties of various of the above objects have been studied by Combescure and Dunlop[10] and Koch.[8]

It is our goal in this section to show how simply one may express the above objects in terms of partial Legendre transforms (thus generalizing the results for $r=1,2$ given in I). This provides a convenient path for the analysis of their irreducibility properties (see I).

We now explain why the projections $P_{<r}$ can be expressed in terms of Legendre transforms in a natural way. This can be done quite generally as follows. Let

$$q_m=\phi^m+\sum_{n=0}^{m-1}T_{mn}\phi^n,\quad m=0,1,2,\ldots.$$

be a sequence of real polynomials in ϕ. Two examples are

$$q_n(x_1,\ldots,x_n)=\prod_{k=1}^{n}\phi(x_k)\tag{4.3}$$

and

$$q_n(x_1,...,x_n) = :\prod_{k=1}^{n} \phi(x_k):. \tag{4.4}$$

Let $Q^{(0,r)}$ be the $(r+1)\times(r+1)$ matrix whose (α,β) entry is the operator with kernel $\langle q_\alpha q_\beta\rangle$ $\alpha,\beta = 0,1,...,r$. Then if $Q^{(0,r)}$ is invertible

$$P_{<r}\psi = q_\alpha (Q^{(0,r)})^{-1}_{\alpha\beta}\langle q_\beta \psi\rangle. \tag{4.5}$$

To see this we need only observe that the righ- hand side defines a self-adjoint, idempotent operator that leaves q_γ invariant for $\gamma = 0,1,...,r$.

Furthermore $Q^{(0,r)^{-1}}$ may be represented in terms of a Legendre transform of

$$Z^q\{J\} = \left\langle \exp\left[\sum_{i=0}^{N} q_i J_i\right]\right\rangle - 1 \tag{4.6}$$

where $N \geqslant r$. Let

$$A^q_\alpha\{J\} = Z^q_{J_\alpha}\{J\}, \quad 0\leqslant \alpha \leqslant r \tag{4.7}$$

and

$$^q\Gamma^{(r/N)}\{A^q_0,...,A^q_r,J_{r+1},...,J_N\} = Z^q\{J^q\} - A^q_\alpha J^q_\alpha, \tag{4.8}$$

where, on the right-hand side of (4.8)
$J^q_\alpha = J^q_\alpha\{A^q_0,...,A^q_r,J_{r+1},...,J_N\}$ is obtained by inverting (4.7) when $0\leqslant\alpha\leqslant r$ and $J^q_i = J_i$ when $r+1\leqslant i\leqslant N$. (We shall use the conventions that Greek indices take values up to and including r while Latin indices run from $r+1$ to N.) Note that unlike the Legendre transforms of Secs. II and III $^q\Gamma^{(r/N)}$ is a function of A_0. The Jacobian relation for $^q\Gamma^{(r/N)}$ now gives

$$^q\Gamma^{(r/N)}_{A_\alpha A_\beta} = -\left(\frac{\delta A^q}{\delta J}\right)^{-1}_{\alpha\beta}$$

so that if we set the arguments of $^q\Gamma$ equal to their physical values (i.e., $A^q_\alpha = A^q_\alpha\{0\}$ and $J_i = 0$)

$$^q\Gamma^{(r/N)}_{A_\alpha A_\beta} = -(Q^{(0,r)})^{-1}_{\alpha\beta}. \tag{4.9}$$

As a result (4.5) may be written, for any $N\geqslant r\geqslant 0$, as

$$(P_{<r}\Psi)(\phi) = \int d\nu(\phi')\, p_{<r}(\phi,\phi')\,\Psi(\phi'),$$

where

$$P_{<r}(\phi,\phi') = -q_\alpha(\phi)\,{}^q\Gamma^{(r/N)}_{A_\alpha A_\beta}\, q_\beta(\phi').$$

From the point of view of irreducibility and hence spectral analysis it is preferable to express $P_{<r}$ in terms of $\Gamma^{(r/N)}$ rather than $^q\Gamma^{(r/N)}$. This we do in

Theorem IV.1. At the physical values $A = J = 0$, we have

(a) $(P_{<r}\Psi)(\phi) = \int d\nu(\phi')\, p_{<r}(\phi,\phi')\,\Psi(\phi')$,

where $p_{<r}(\phi,\phi') = 1 - :\phi^\alpha:\, \Gamma^{(r/N)}_{A_\alpha A_\beta} :\phi'^\beta:$
for any $N\geqslant r\geqslant 1$,

(b) $\langle \phi^i P_{>r}\,\phi^k\rangle = \Gamma^{(r/N)}_{J_j J_k}$ for all $0\leqslant r < i,j\leqslant N$,

(c) $P_r\Psi = -P_{>r}\,\phi^r\,\Gamma^{(r)}_{A_r A_r}\,\langle\phi^r P_{>r}\,\Psi\rangle$ for all $r\geqslant 1$,

(d) $R^{(r)^{-1}} = -\Gamma^{(r)}_{A_r A_r} = (S^{(r)})^{-1}_{rr}$ for all $r\geqslant 1$.

Proof. (a) We remind the reader that *when* $A = J = 0$

$$\Gamma^{(r/N)}_{A_\alpha A_\beta} = \Gamma^{(r)}_{A_\alpha A_\beta} = -(S^{(r)})^{-1}_{\alpha\beta} \quad \text{(by Theorem III.3b)},$$

where

$$S_{\alpha\beta} = \left(\frac{\delta}{\delta f}\right)^\alpha\left(\frac{\delta}{\delta g}\right)^\beta \langle e^{\phi f+\phi g}\rangle\,\langle e^{\phi f}\rangle^{-1}\,\langle e^{\phi g}\rangle^{-1}\big|_{f=g=0} \quad \text{[by (3.3)]}$$

$$= \langle :\phi^\alpha::\phi^\beta:\rangle \quad \text{[by (2.2)]}$$

and that $S^{(r)}$ is the restriction of S whose indices run from 1 to r. Then choosing $q_\alpha = :\phi^\alpha:$ we have $Q = S$ and

$$(Q^{(0,r)})^{-1}_{\alpha\beta} = \begin{Bmatrix} 1, & \alpha=\beta=0 \\ 0, & \text{precisely one of } \alpha,\beta \text{ are zero} \\ (S^{(r)})^{-1}_{\alpha\beta}, & \alpha\geqslant 1,\ \beta\geqslant 1 \end{Bmatrix}$$

and part (b) follows from (4.5).

(b) By Lemma II.2 $G_{J_\beta}\{J\{A\}\} = F_\beta\{A + A_0\}$ is independent of J_i and so the usual cancellation gives (with Greek indices running from 1 to r)

$$\Gamma^{(r/N)}_{J_i} = G_{J_\beta}\frac{\delta J_\beta}{\delta J_i} + G_{J_i} - G_{J_\beta}\frac{\delta J_\beta}{\delta J_i}$$

$$= G_{J_i}.$$

Applying another derivative we get

$$\Gamma^{(r/N)}_{J_i J_k} = G_{J_i J_k} + G_{J_i J_\beta}\frac{\delta J_\beta}{\delta J_k}$$

$$= G_{J_i J_k} - G_{J_i J_\beta}(G^{(r)})^{-1}_{\beta\alpha}\, G_{J_\alpha J_k},$$

where $G^{(r)}_{\alpha\beta}$ is the restriction of $G_{J_\alpha J_\beta}$ to $1\leqslant\alpha,\beta\leqslant r$, since

$$0 = \frac{\delta}{\delta J_k}(G_{J_\alpha}\{J\{A\}\}) = G_{J_\alpha J_\beta}\frac{\delta J_\beta}{\delta J_k} + G_{J_\alpha J_k}.$$

On the other hand by part (a) of this theorem (since $A = J = 0$)

$$\langle\phi^i P_{>r}\,\phi^k\rangle$$
$$= \langle\phi^i\phi^k\rangle - \langle\phi^i\rangle\langle\phi^k\rangle - \langle\phi^i:\phi^\alpha:\rangle\,(S^{(r)})^{-1}_{\alpha\beta}\,\langle:\phi^\beta:\phi^k\rangle$$
$$= G_{J_i J_k} - G_{J_i J_\alpha}(G^{(r)})^{-1}_{\alpha\beta}\,G_{J_\beta J_k}$$

since

$$G_{J_\beta J_k} = W^+_{\beta\gamma}\,\langle:\phi^\gamma:\phi^k\rangle \quad \text{[by (2.10)]}$$

and

$$G^{(r)}_{J_\alpha J_\beta} = W^+_{\alpha\gamma}\langle:\phi^\gamma::\phi^\delta:\rangle(W^+)^t_{\delta\beta} \quad \text{[by (2.10)]}$$
$$= W^+_{\alpha\gamma}\,S^{(r)}_{\gamma\delta}(W^+)^t_{\delta\beta}.$$

(c) $P_r\Psi = P_{>r}\,P_{<r}\,P_{>r}\,\Psi$

$$= -P_{>r}:\phi^\alpha:\Gamma^{(r)}_{A_\alpha A_\beta}\langle:\phi^\beta:P_{>r}\,\Psi\rangle$$
$$= -P_{>r}\phi^r\,\Gamma^{(r)}_{A_r A_r}\langle\phi^r P_{>r}\,\Psi\rangle \quad \text{[by (4.2b)]}$$

(d) By part (c) with $\Psi = \phi^r$

$$R^{(r)} = -R^{(r)}\Gamma^{(r)}_{A_r A_r}R^{(r)}$$

and so the desired conclusion follows if $R^{(r)}$ is invertible. But by Lemma IV.2 (which follows), for $\alpha = 1,...,r$,

$$P_\alpha:\phi^\alpha: = (P_{<\alpha} - P_{<(\alpha-1)}):\phi^\alpha: = \sum_{\beta=1}^{\alpha} V_{\alpha\beta}:\phi^\beta:,$$

where V is a lower triangular $r\times r$ matrix with identities on the diagonal and is hence invertible. Now

$$(VSV^t)_{\alpha\beta} = \langle (P_\alpha :\phi^\alpha:)(P_\beta :\phi^\beta:)\rangle$$
$$= \langle (P_\alpha \phi^\alpha)(P_\beta \phi^\beta)\rangle \quad \text{[by (4.2b)]}$$
$$= \delta_{\alpha\beta} R^{(\alpha)} \quad \text{[by (4.2f)]}.$$

Therefore the invertibility of $R^{(1)},\ldots, R^{(r)}$ is equivalent to the invertibility of $S^{(r)}$, which we are assuming. ■

Remarks 1. The 1 in the formula for $P_{<r}$ of part (a) of the theorem may be absorbed into the sum over α and β by using a Legendre transform of $Z\{J_0,\ldots,J_N\}$.

2. The formula

$$\langle q_i P_{>r}\, q_k\rangle = {}^q\Gamma^{(r/N)}_{J_iJ_k} \text{ for all } 0 \leqslant r < i,j \leqslant N$$

may be proved by the same method that its analog for $\Gamma^{(r/N)}_{J_iJ_k}$ is proven in part (b).

Lemma IV.2.

$$\text{(a) } P_{<r}:\phi^m: = \sum_{\alpha=0}^{r} \mathscr{P}^{(r)}_{m,\alpha} :\phi^\alpha: \quad \text{for all } r,m \geqslant 0,$$

where

$$\mathscr{P}^{(r)}_{m,\alpha} = \begin{cases} \delta_{m,0}\delta_{\alpha,0} & \text{for } r = 0, \\ \delta_{m,0} + S_{m,\beta}\{0\}(S^{(r)}\{0\})^{-1}_{\beta,\alpha} & \text{for } r \geqslant 1. \end{cases}$$

$$\text{(b) } P_{<r}\,\phi^m = \sum_{\alpha=0}^{r} \overline{\mathscr{P}}^{(r)}_{m,\alpha}\,\phi^\alpha \quad \text{for all } r,m \geqslant 0,$$

where

$$\overline{\mathscr{P}}^{(r)}_{m,\alpha} = W^+_{mj}\,\mathscr{P}^{(r)}_{j,\beta} W^-_{\beta,\alpha}.$$

[Here we use the convention that $(S^{(r)})^{-1}_{\beta,\alpha}$ is defined for all α,β but is nonzero only for $1 \leqslant \alpha,\beta \leqslant r$.]

Proof. (a) For $r = 0$ the result is obvious. For $r \geqslant 1$

$$P_{<r}:\phi^m:$$
$$= \langle :\phi^m:\rangle - \sum_{\alpha,\beta=1}^{r} :\phi^\alpha: \Gamma^{(r)}_{A_\alpha A_\beta}\langle :\phi^\beta:\,:\phi^m:\rangle$$
$$= \delta_{m,0} - \sum_{\alpha,\beta=1}^{r} S_{m\beta}\Gamma^{(r)}_{A_\beta A_\alpha} :\phi^\alpha:.$$

$$\text{(b) } P_{<r}\phi^m = \sum_{j=0}^{m} W^+_{mj} P_{<r}\phi^j \quad \text{[by (2.10)]}$$
$$= \sum_{j=0}^{m}\sum_{\beta=0}^{r} W^+_{mj}\,\mathscr{P}^{(r)}_{j,\beta} :\phi^\beta: \quad \text{[by part (a)]}$$
$$= \sum_{j=0}^{m}\sum_{\beta=0}^{r}\sum_{\alpha=0}^{\beta} W^+_{mj}\,\mathscr{P}^{(r)}_{j,\beta} W^-_{\beta\alpha}\phi^\alpha \text{ [by (2.4)]}. \ ■$$

V. THE BETHE–SALPETER KERNELS

Glimm and Jaffe have given[13] an inductive definition of a higher order rth Bethe–Salpeter kernel that reads

$$K^{(r)}(x^{(r)},y^{(r)}) = R^{(r)^{-1}}(x^{(r)},y^{(r)}) - \sum_{\pi\in\mathscr{P}^r_2} n_\pi^{-1}(S_xS_y \underset{\alpha\in\pi}{\otimes} K^{(|\alpha|)})(x^{(r)},y^{(r)}), \tag{5.1}$$

where $x^{(r)} = (x_1,\ldots,x_r) \in \mathbb{R}^{rd}$, S_x is symmetrization in $x^{(r)}$,

$R^{(r)} = \langle \phi^r P_r \phi^r\rangle$ (see Sec IV),

$\mathscr{P}^r_2$ is the set of partitions of $\{1,\ldots r\}$ into two or more subsets,

$$n_\pi = \frac{r!}{\prod_{\alpha\in\pi}|\alpha|!}, \text{ where } |\alpha| \text{ is the cardinality of } \alpha,$$

and

$$K^{(1)} = -\Gamma^{(1)}_{A_1A_1}\{0\} = G^{-1}_{J_1J_1}\{0\}.$$

Thanks to the symmetrization all partitions $\pi = \{\alpha_1,\ldots,\alpha_k\}$ that have the same $\{|\alpha_1|,\ldots,|\alpha_k|\}$ will give the same $S_xS_y \underset{\alpha\in\pi}{\otimes} K^{(|\alpha|)}$. (For example $\pi = \{\{1,2\},\{3\}\}, \pi = \{\{1,3\},\{2\}\}$ and $\pi = \{\{2,3\},\{1\}\}$ all yield the same $S_xS_y \underset{\alpha\in\pi}{\otimes} K^{(|\alpha|)}$.) This redundancy in (5.1) may be removed to give the equivalent definition

$$K^{(r)} = R^{(r)^{-1}} - \sum_{p\in\mathscr{D}^r_2} n_p^{-1} K^p, \tag{5.2}$$

where $\mathscr{D}^r$ is the set of decompositions of r into positive integer summands without regard to order (e.g., $\mathscr{D}^3 = \{\{3\},\{2,1\},\{1,1,1\}\}$),

$$\mathscr{D}^r_2 = \mathscr{D}^r\setminus\{r\}, \quad (\text{e.g., } \mathscr{D}^3_2 = \{\{2,1\},\{1,1,1\}\}),$$

$$n_p = \prod_k n_p(k)!,$$

$n_p(k)$ is the number of k's occurring in p [e.g., $n_{\{1,1,1\}}(1) = 3$, $n_{\{2,1\}}(2) = 1$],

$$K^p = S_xS_y \underset{k\in p}{\otimes} K^{(k)}.$$

As we have shown in Theorem IV.1d, $R^{(r)^{-1}} = -\Gamma^{(r)}_{A_rA_r}\{0\}$ so that the definition (5.2) may be stated directly in terms of the Legendre transform as

$$K^{(r)} = -\Gamma^{(r)}_{A_rA_r}\{0\} - \sum_{p\in\mathscr{D}^r_2} n_p^{-1} K^p. \tag{5.3}$$

Note the principle that if we want a kernel $K(x_1,\ldots,x_j;\ldots)$ with r-irreducibility properties when the points $(x_1,\ldots,x_j) = x^{(j)}$ are in a cluster we take an $A_j(x^{(j)})$-derivative of $\Gamma^{(r)}$ instead of $j\,A_1$-derivatives (see I and Ref. 5 for more details). For example, for $r = 2$ and 3

$$K^2(x^{(2)},y^{(2)}) = S_xS_y\big(-\Gamma^{(2)}_{A_2A_2}\{0\} - \tfrac{1}{2}\Gamma^{(1)}_{A_1A_1}\{0\}\otimes\Gamma^{(1)}_{A_1A_1}\{0\}\big),$$

$$K^{(3)}(x^{(3)},y^{(3)}) = S_xS_y\big[-\Gamma^{(3)}_{A_3A_3}\{0\} - K^{(2)}\otimes K^{(1)} - 1/3!\,K^{(1)}\otimes K^{(1)}\otimes K^{(1)}\big].$$

Remark 1. Our $K^{(r)}$ is the negative of Glimm and Jaffe's.[13] We have chosen this sign convention so that positive kernels are associated with repulsive potentials. For example, in a ϕ^4 field theory

$$K^{(2)}(x_1,x_2,y_1,y_2) = 24\lambda\delta(x_1-x_2)\delta(x_1-y_1)\times\delta(x_1-y_2) + O(\lambda^2).$$

Remark 2. Combescure and Dunlop[10] have analyzed the cluster-irreducibility properties (with the $x^{(r)}$ as one cluster and $y^{(r)}$ as a second) of $K^{(r)}(x^{(r)},y^{(r)})$ without recourse to the Legendre transform. We have used Legendre transforms to analyze[3,5] the cases $r = 2$, 3, and 4. The use of the Legendre transform has two important consequences. Firstly, the combinatorics are handled automatically yielding simple proofs. Secondly, it is natural to consider not just the cluster irreducibility of $\Gamma^{(r)}_{A,A_r}\{0\}$ but more generally that of $\Gamma^{(r)}_{A_{i_1}A_{i_2}}\{A\}$. Taking further functional derivatives $\Gamma^{(r)}_{A_{i_1}\cdots A_{i_k}}$ yields irreducibility for arbitrary numbers of clusters.

Note that the subtracted terms $(1/n_p)K^p$ in (5.3) are themselves r-cluster irreducible if $K^{(k)}$ is k-cluster irreducible and so are irrelevant as far as the cluster irreducibility properties of $K^{(r)}$ are concerned. Their role is to "eliminate lower body scattering processes." In other words their role is to make $K^{(r)}(x^{(r)},y^{(r)})$ connected and not just cluster connected. When $t=0$ so that the covariance decouples across the surface σ we should have $K^{(r)}(x^{(r)},y^{(r)})=0$ whenever σ splits the $2r$ variables $x^{(r)},y^{(r)}$ and not just when σ separates all the variables in $x^{(r)}$ from all of the variables in $y^{(r)}$. This is indeed the case (Ref. 10, Corollary 2.1, case $d=0$). We now prove this result without the restriction $J=0$. Define the functional $K^{(r)}\{J\}$ by

$$K^{(1)}=(S^{(1)})^{-1}_{11},$$

$$K^{(r)}=(S^{(r)})^{-1}_{rr}-\sum_{p\in\mathscr{D}_2^r}n_p^{-1}K^p. \tag{5.4a}$$

Alternatively

$$(S^{(r)})^{-1}_{rr}=\sum_{p\in\mathscr{D}^r}n_p^{-1}K^p. \tag{5.4b}$$

Theorem V. 1. $K^{(r)}\{J\}$ is connected, i.e., at 0,

$$K^{(r)}(x^{(r)},y^{(r)})=0 \tag{5.5}$$

whenever σ splits the $2r$ variables $x_1,x_2,\ldots,y_r$.

Proof. The proof is by induction on r. The result is trivial for $r=1$ so assume that (5.5) has been proved for $r\leqslant M-1$. We use the notation preceding Theorem III. 4.

Since $S^{(M)}$, and hence $K^{(M)}$, commutes with the projections P_+, P_-, and P_0 it suffices to prove $P_0K^{(M)}=0$ or equivalently

$$Q_m^{(M)}K^{(M)}=0 \quad \text{for all } m=1,\ldots,M-1. \tag{5.6}$$

Here $Q_m^{(M)}=\pi_+^{(m)}\times\pi_-^{(M-m)}\restriction\mathscr{H}^{(M)}$ is the projection in $\mathscr{H}^{(M)}$ onto functions which vanish unless exactly m arguments are on the $+$ side of σ and the remaining $M-m$ arguments are on the $-$ side. By theorem III.4

$$Q_m^{(M)}(S^{(M)})^{-1}_{MM}=\binom{M}{m}^{-1}\left(S_+^{(m)}\right)^{-1}_{mm}\times\left(S_-^{(M-m)}\right)^{-1}_{M-mM-m}. \tag{5.7}$$

[The second term in (3.15) drops out.] By the inductive hypothesis each $K^{(k)}$ in $\sum_{p\in\mathscr{D}_2^M}n_p^{-1}K^p$ must be supported entirely on one side of σ. Furthermore, given any $p\in\mathscr{D}^M$, $p'\in\mathscr{D}^m$, and $p''\in\mathscr{D}^{M-m}$ such that $p=p'\cup p''$ there are precisely

$$\prod_k\binom{n_p(k)}{n_{p'}(k)}=\frac{n_p}{n_{p'}n_{p''}}$$

ways of assigning either a $+$ or a $-$ to each element of p in such a way that the elements assigned $+$ form p' and those assigned $-$ form p''. Hence

$$\begin{aligned}
&Q_m^{(M)}\sum_{p\in\mathscr{D}_2^M}n_p^{-1}K^p\\
&=S_xS_y\sum_{p'\in\mathscr{D}^m}\sum_{p''\in\mathscr{D}^{M-m}}\frac{1}{n_{p'}n_{p''}}K_+^{p'}\otimes K_-^{p''}\\
&=\sum_{p'\in\mathscr{D}^m}\sum_{p''\in\mathscr{D}^{M-m}}\frac{1}{n_{p'}n_{p''}}\binom{M}{m}^{-1}K_+^{p'}\times K_-^{p''}\quad\text{[see (3.14)]}\\
&=\binom{M}{m}\left(S_+^{(m)}\right)^{-1}_{mm}\times\left(S_-^{(M-m)}\right)^{-1}_{M-mM-m}\quad\text{[by (5.4b)]}.
\end{aligned}$$

Together with (5.7) this establishes (5.6) and hence the theorem.■

Remark 3. The connectivity of $K^{(r)}=K^{(r)}\{0\}$ implies that there are no C^{-1}'s in its perturbation series. As a result we expect $K^{(r)}$ to be more regular than $\Gamma^{(r)}_{A,A_r}\{0\}$. We have established this improved regularity for $r=1$ and 2 in I (see Sec. V) and II.

VI. THE EXISTENCE AND REGULARITY OF $\Gamma^{(N)}\{A\}$ IN WEAKLY COUPLED $P(\phi)_2$ MODELS

We now consider the rigorous construction of the generalized vertex functions of $\Gamma^{(N)}\{A\}$ in weakly coupled $P(\phi)_2$ models [$\epsilon P(\phi)_2$]. The main results of this section will be the existence of these vertex functions (Theorem VI.1), their regularity in t (Theorem VI.4), and the validity of the formula

$$\dot{\Gamma}^{(N)}\{A\}=\Gamma^{(N)}_{A_i}\{A\}\dot{A}_i^0$$

for suitable A (Theorem VI.11). Since for large n $\langle e^{J_n\phi^n}\rangle$ is not well defined even in the free theory, we view generating functionals like $G\{J\}$ not as ordinary complex valued functionals but as formal power series in J. This framework was considered in detail in II, so we shall only outline its essential features here. This formal power series (fps) is simply the set of all moments of the generating functional. If $\mathscr{B}_1$ and $\mathscr{B}_2$ are Banach spaces (or more generally Fréchet spaces) we define the space of all formal power series on $\mathscr{B}_1$ with values in $\mathscr{B}_2$ to be

$$\mathscr{F}(\mathscr{B}_1,\mathscr{B}_2)=\mathop{\times}_{n=0}^{\infty}\mathscr{M}_n(\mathscr{B}_1,\mathscr{B}_2),$$

where $\mathscr{M}_0(\mathscr{B}_1,\mathscr{B}_2)=\mathscr{B}_2$ and for $n\geqslant1$ and

$\mathscr{M}_n(\mathscr{B}_1,\mathscr{B}_2)=\{$symmetric, continuous, n-linear, $\mathscr{B}_2$-valued forms on $\mathscr{B}_1\}$.

A typical element F of $\mathscr{F}(\mathscr{B}_1,\mathscr{B}_2)$ is then a sequence $\{F_n\}$ of $\mathscr{B}_2$-valued, n-linear, continuous symmetric forms on $\mathscr{B}_1$. In the event that $\Sigma_{n=0}^{\infty}1/n!\|F_n\|x^n$ has a nonzero radius of convergence r the series $\Sigma_{n=0}^{\infty}1/n!F_n\{J,J,\ldots,J\}$ converges strongly to a function $F\{J\}$ which is analytic for $\|J\|_{\mathscr{B}_1}<r$. If so, we identify the fps F with the analytic function $F\{J\}$. Based on this identification we can define fps operations such as addition, scalar multiplication, other multiplications (when $\mathscr{B}_2$ has the appropriate additional structure, e.g., Hilbert space or C^*-algebra), composition, and differentiation. For example, the sum of the formal power series $F=\{F_n\}$ and $G=\{G_n\}$ is defined to be $F+G\equiv\{F_n+G_n\}$. We showed in Sec. II of II that these operations have all the usual algebraic properties.

As we found in II the choice of space $\mathscr{B}$ for a functional like $G\in\mathscr{F}(\mathscr{B},\mathbb{C})$ can be delicate with the "right" choice depending on the purpose of the analysis. In this section we shall first choose $\mathscr{B}$ to be $\mathscr{B}_t$ of (6.2a) when discussing the existence and regularity of G and Γ, and then we choose $\mathscr{B}$ to be the Fréchet space $\mathscr{D}$ of (6.19) when discussing formulas for $\partial_t^r\Gamma$.

To prove the existence of the (generalized) vertex functions of ${}^W\Gamma^{(N)}$, (we often suppress the N and W) using this framework we start with the (generalized) Schwinger functions.

$$S_n^W(J^{(1)},...,J^{(n)}) = \langle U^W\{J^{(1)}\}\cdots U^W\{J^{(n)}\}\rangle_t. \tag{6.1}$$

Here $\langle\cdot\rangle_t$ is the Euclidean expectation in an $\epsilon P(\phi)_2$ theory with covariance $C(t)$ [see (1.9)]. We remind the reader that the source terms in $U^W\{J\} = \Sigma_{i=1}^N :\phi^i:J_i$ are physically Wick ordered [see (2.2)] and that ${}^W\Gamma = \Gamma$ (Theorem II.3a). We choose to use U^W rather than U because the subtraction scheme implicit in the Wick ordering results in natural test function spaces:

$$\mathscr{B}_t = \{J = (J_1,...,J_N)|J_i \text{symmetric}, \|J\|_t^2 = \sum_i (J_i, \otimes^i C(t)J_i) = (\mathrm{J},\mathscr{C}J) < \infty\}, \tag{6.2a}$$

$$\mathscr{B}_t^* = \{A = (A_1,...,A_N)|A_i \text{symmetric}, \|A\|_t^{*2} = \sum_i (A_i, \otimes^i C(t)^{-1}A_i) = (A,\mathscr{C}^{-1}A) < \infty\}. \tag{6.2b}$$

We prove in Theorem VI.2 that S_n^W is indeed a continuous n-linear form on $\mathscr{B}_t$, or in other words that $\mathbf{Z}^W = \{S_n^W\}\in\mathscr{F}(\mathscr{B}_t,\mathbb{C})$. From $\mathbf{Z}^W$ we may define $G^W = \ln\mathbf{Z}^W\in\mathscr{F}(\mathscr{B}_t,\mathbb{C})$ by (fps) composition (note that $S_0^W = 1$) and $A^W\{J\}\in\mathscr{F}(\mathscr{B}_t,\mathscr{B}_t^*)$ by differentiation [see (2.4b)]. The crucial step in the construction of ${}^W\Gamma$ is taking the composition inverse of $A^W\{J\}$ to get $J^W\{A\}\in\mathscr{F}(\mathscr{B}_t^*,\mathscr{B}_t)$. According to Lemma II.10 of II this composition inverse exists provided $(\delta A^W/\partial J)\{J=0\}$, viewed as an operator from $\mathscr{B}_t$ to $\mathscr{B}_t^*$, has a bounded inverse. That it does is proven in Theorem VI.3. Then ${}^W\Gamma\{A\}\in\mathscr{F}(\mathscr{B}_t^*,\mathbb{C})$ may be defined by ${}^W\Gamma\{A\} = G^W\{J^W\{A\}\} - G_{J_i}^W\{J^W\{A\}\}\cdot J_i^W\{A\}$. Here the product $G_{J_i}^W\cdot J_i^W$ is the (fps) evaluation product mapping $\mathscr{B}_t^*\times\mathscr{B}_t$ into $\mathbb{C}$. Hence we have

Theorem VI.1. In $\epsilon P(\phi)_2$
$A^W\{J\}\in\mathscr{F}(\mathscr{B}_t,\mathscr{B}_t^*)$ is invertible and
${}^W\Gamma\{A\}\in\mathscr{F}(\mathscr{B}_t^*,\mathbb{C})$.

Remark 1. Note that this theorem provides a rigorous definition and proof of existence of the (generalized) vertex functions of ${}^W\Gamma^{(N)}$ without requiring the existence of ${}^W\Gamma^{(N)}\{A\}$ as a complex valued function on $\mathscr{B}_t^*$. When $N = 1,2$ ${}^W\Gamma^{(N)}\{A\}$ does indeed exist as a conventional function. This was proven for $N=1$ by Glimm and Jaffe[18] and for $N=2$ by us.[19]

Theorem VI.2. In $\epsilon P(\phi)_2$

$$\left|\left\langle\prod_{i=1}^n\left(\sum_{j=1}^N J_j^{(i)}:\phi^j:\right)\right\rangle_t\right| \leqslant c_{n,N}\prod_{i=1}^n\|J^{(i)}\|_t. \tag{6.3}$$

Remark 2. Henceforth we shall use c to denote various (positive) constants like $c_{n,N}$, which depend on n and N but are independent of J, etc.

Proof. By Hölder's inequality it suffices to prove

$$|\langle(M:\phi^k:)^n\rangle_t| \leqslant c\|M\|_t^n$$

for $M\in C_0^\infty(\mathbb{R}^{2k})$. Let

$$\{\zeta_\alpha\in C_0^\infty(\mathbb{R}^2)|\alpha\in\mathbf{Z}^2\}$$

be a partition of unity invariant under lattice translations. Given any $\boldsymbol{\alpha} = (\alpha^1,...,\alpha^k)\in(\mathbf{Z}^2)^k$ define the localization of M at $\boldsymbol{\alpha}$ to be

$$M_{\boldsymbol\alpha}(x_1,...,x_k) = M(x_1,...,x_k)\prod_{i=1}^k\zeta_{\alpha^i}(x_i).$$

Then

$$\langle(M:\phi^k:)^n\rangle_t = \sum_{\boldsymbol\alpha_1\cdots\boldsymbol\alpha_n}\left\langle\prod_{i=1}^n(M_{\boldsymbol\alpha_i}:\phi^k:)\right\rangle_t$$

and the proof goes in two steps.

In Step 1 we prove

$$\left|\left\langle\prod_{i=1}^n(M_{\boldsymbol\alpha_i}:\phi^k:)\right\rangle_t\right| \leqslant cD(\boldsymbol\alpha_1,...,\boldsymbol\alpha_n)\prod_{i=1}^n\|M_{\boldsymbol\alpha_i}\|_t. \tag{6.4a}$$

Here D is the decay factor

$$D(\boldsymbol\alpha_1,...,\boldsymbol\alpha_n) = \sum_{\sigma,\tau}\prod_{\substack{1\leqslant i\leqslant n\\ 1\leqslant j\leqslant k}}\exp(-c|\alpha_i^j - \alpha_{\sigma(i,j)}^{\tau(i,j)}|), \tag{6.4b}$$

where σ and τ run over all maps,

$$\tau:\{(i,j)\}\to\{1, 2, 3,...,k\},$$
$$\sigma:\{(i,j)\}\to\{1, 2, 3,...,n\}, \quad \sigma(i,j)\neq i.$$

This complicated looking decay factor says nothing more than that there is decay from every variable of every $M_{\boldsymbol\alpha_i}$ to some variable in a different $M_{\boldsymbol\alpha_j}(j\neq i)$. It arises because the subtraction scheme in $:\phi^k:$ forces every ϕ in every $:\phi^k:$ in (6.4a) to be connected to some ϕ in a different $:\phi^k:$. We have already remarked on this connectivity following the definition (2.5c) of ${}^W\Gamma$. It will surface again in the Integration by Parts Lemma VI.7.

In Step 2 we prove that the estimate (6.4a) implies (6.3).

Step 1. We apply the cluster expansion[20] to the finite volume approximations to $\left\langle\prod_{i=1}^n(M_{\boldsymbol\alpha_i}:\phi^k:)\right\rangle_t$. In doing so we must be careful not to destroy the cancellations built into $:\phi^k:$ or else we lose the crucial decay factor D of (6.4). Accordingly we implement the cancellations by Ginibre's duplicate copy trick. We use a large number of independent isomorphic copies of our $P(\phi)_2$ theory. To be precise we use a multifold theory containing nk copies (labelled by two indices $1\leqslant a\leqslant k$, $1\leqslant i\leqslant n$) in addition to the original. If $\langle\langle\cdot\rangle\rangle$ is the expectation with respect to the product measure of all the theories we have

$$\left\langle\left\langle f(\phi)\prod_{a=1}^k\prod_{i=1}^n f_{a,i}(\phi_{a,i})\right\rangle\right\rangle = \langle f(\phi)\rangle_t\prod_{a=1}^k\prod_{i=1}^n\langle f_{a,i}(\phi_{a,i})\rangle_t. \tag{6.5}$$

Consider the simple example

$$\langle:\phi^2: :\phi^2:\rangle_t = \langle(\phi^2 - 2\phi\langle\phi\rangle_t - \langle\phi^2\rangle_t + 2\langle\phi\rangle_t\langle\phi\rangle_t)(\phi^2 - 2\phi\langle\phi\rangle_t - \langle\phi^2\rangle_t + 2\langle\phi\rangle_t\langle\phi\rangle_t)\rangle_t.$$

Defining

$$[\phi^2]_i = \phi^2 - 2\phi\phi_{1,i} - \phi^2_{1,i} + 2\phi_{1,i}\phi_{2,i}$$

gives

$$\langle :\phi^2: \; :\phi^2: \rangle_t = \langle\langle [\phi^2]_1 [\phi^2]_2 \rangle\rangle.$$

In general

$$:\phi^k: = \left(\frac{\delta}{\delta f}\right)^k e^{\phi f} \langle e^{\phi f}\rangle_t^{-1}\Big|_{f=0}$$

$$= \sum_{I\subset\{1,2,\ldots,k\}} \sum_{\pi\in\mathscr{P}_{I^c}} (-1)^{|\pi|}|\pi|!\phi^I \prod_{l\in\pi}\langle\phi^l\rangle_t ,$$

where $\mathscr{P}_I c$ is the set of all partitions of $I^c = \{1,\ldots,k\}\setminus I$,

and $\phi^I = \prod_{j\in I}\phi(x_j)$.

Thus we define

$$[\phi^k]_i = \sum_{I\subset\{1,2,\ldots,k\}} \sum_{\pi\in\mathscr{P}_{I^c}} (-1)^{|\pi|}|\pi|!\phi^I \prod_{a=1}^{|\pi|}\phi^{l_a}_{a,i}, \qquad (6.6)$$

where $\pi = \{l_1,\ldots,l_{|\pi|}\}$, in order to give

$$\left\langle \prod_{i=1}^n (M_{\alpha_i} :\phi^k:)\right\rangle_t = \left\langle\!\left\langle \prod_{i=1}^n n(M_{\alpha_i}[\phi^k]_i)\right\rangle\!\right\rangle. \qquad (6.7)$$

We advise any reader not familiar with the cluster expansion to read Ref. 20 forthwith. For then it will be obvious that

$$\left\langle\!\left\langle \prod_{i=1}^n (M_{\alpha_i}[\phi^k]_i)\right\rangle\!\right\rangle_\Lambda = \sum_{X,\Gamma\in I_X}\int_0^{s(\Gamma)} d\sigma(\Gamma)$$

$$\times\partial^\Gamma\left\langle\!\left\langle \prod_{i=1}^n (M_{\alpha_i}\,[\phi^k]_i)\right\rangle\!\right\rangle^u_{X,\sigma(\Gamma)} \frac{Z_{\Lambda\setminus X,\partial X}}{Z_\Lambda}, \qquad (6.8)$$

where $\langle\langle\;\rangle\rangle_\Lambda$ is a finite volume approximation to $\langle\langle\;\;\rangle\rangle$,

X is any finite union of closed lattice squares containing the supports of $\{\zeta_{\alpha_i^j} | 1\leqslant j\leqslant k, 1\leqslant i\leqslant n\}$,

$\mathscr{B} = \{$lattice bonds that do not intersect the support of any $\zeta_{\alpha_i^j}\}$,

$I_X = \{\Gamma\subset\mathscr{B} | \Gamma$ is finite, $\Gamma\subset \operatorname{Int} X$, every connected component of $X\setminus(\mathscr{B}\setminus\Gamma)$ intersects the support of some $\zeta_{\alpha_i^j}\}$,

$$s(\Gamma) = \left(s(\Gamma)_b = \begin{cases}1 & b\in\Gamma\\ 0 & b\in\mathscr{B}\setminus\Gamma\end{cases}\right)$$

so that Γ is the set of coupling bonds and $\mathscr{B}\setminus\Gamma$ is the set of Dirichlet bonds,

$\sigma(\Gamma)$ is a vector having one component for each nonzero $s(\Gamma)_b$ [$\sigma(\Gamma)_b$ is a measure of the strength of coupling on the bond $b\in\Gamma$],

$$\partial^\Gamma = \prod_{b\in\Gamma}\partial/\partial\sigma(\Gamma)_b,$$

$\langle\langle\rangle\rangle^u_{X,\sigma(\Gamma)}$ is the unnormalized expectation in volume X all of whose covariances (remember the $nk+1$ copies) have boundary conditions given by $\sigma(\Gamma)$,

$$Z_\Lambda = \langle\langle 1\rangle\rangle^u_{\Lambda,\sigma(\Gamma)_{b=1}},$$

$$Z_{\Lambda\setminus X,\partial X} = \langle\langle 1\rangle\rangle^u_{\Lambda\setminus X,s(\Gamma)}.$$

The connectedness properties we have built into the subtraction scheme $:\phi^k:$ now imply that [for every nonzero term in (6.8)] every connected component of $X\setminus(\mathscr{B}\setminus\Gamma)$ must intersect the supports of $\zeta_{\alpha_i^j}$ for at least two different values of i. This is proven by applying the argument (modified appropriately for unnormalized expectations) which follows the definition (2.5) of ${}^W\Gamma$.

We now proceed along the usual cluster expansion road.[20] We use the Kirkwood–Salsburg equations to estimate $Z_{\Lambda\setminus X,\partial X}/Z_\Lambda$. We then evaluate the derivatives ∂^Γ. We use the Schwarz inequality (with respect to the Gaussian measure) to separate the interaction from the polynomial in the fields downstairs. The Gaussian integral containing the latter is evaluated and estimated as usual. It is important to observe that in this Gaussian integral two ϕ's belonging to the same factor M_{α_i} cannot contract directly to each other without intervening interaction vertices. This is a consequence of the fact that, if : : denotes Wick ordering with respect to the covariance $C(\sigma(\Gamma))$ of the *multifold* theory, then

$$\langle\langle[\phi^k]_i\psi\rangle\rangle^u_{X,\sigma(\Gamma)}$$

$$= \left(\frac{\delta}{\delta f}\right)^k \langle\langle e^{\phi f}(\langle\langle e^{\phi f}\rangle\rangle^u_{X,\sigma(\Gamma)})^{-1}\psi\rangle\rangle^u_{X,\sigma(\Gamma)}\big|_{f=0}\langle\langle 1\rangle\rangle^u_{X,\sigma(\Gamma)}$$

$$= \left(\frac{\delta}{\delta f}\right)^k \langle\langle :e^{\phi f}:(\langle\langle :e^{\phi f}:\rangle\rangle^u_{X,\sigma(\Gamma)})^{-1}\psi\rangle\rangle^u_{X,\sigma(\Gamma)}\big|_{f=0}\langle\langle 1\rangle\rangle^u_{X,\sigma(\Gamma)}$$

$$= \langle\langle :[\phi^k]_i:\psi\rangle\rangle^u_{X,\sigma(\Gamma)} \qquad (6.9)$$

for any ψ that is independent of $\{\phi_{ai} | 1\leqslant a\leqslant k\}$.

When this is done we get, as usual,

$$|\langle\prod_{i=1}^n (M_{\alpha_i} :\phi^k:)\rangle_t| \leqslant c\sum_{X,\Gamma}\prod_{i=1}^n \|M_{\alpha_i}\|_t e^{K'|X| - K(m_0)|\Gamma|},$$

where the constant $K(m_0)$ may be made arbitrarily large by increasing the bare mass m_0. Now we use Prop. 5.1 of Ref. 20 which states that the number of terms in the above sum having a given value $|X|$ is at most $e^{K_2|X|}$, and Eq. (5.1) of Ref. 20 which states $2|\Gamma|\geqslant|X| - nk$. We thus obtain

$$|\langle\prod_{i=1}^n (M_{\alpha_i} :\phi^k:)\rangle_t| \leqslant c\prod_{i=1}^n \|M_{\alpha_i}\|_t e^{-|X|_{\min}},$$

if m_0 is chosen large enough. Here $|X|_{\min}$ is the area of the smallest $|X|$ occurring in a nonzero term of (6.8).

Now consider any α_i^j. It is "in" (that is, in or at least adjacent to) some component $X_{p(i,j)}$ of X. Some other $\alpha^{\tau(i,j)}_{\sigma(i,j)}$ also must be "in" that component. Then

$$|\alpha_i^j - \alpha^{\tau(i,j)}_{\sigma(i,j)}| \leqslant c|X_{p(i,j)}|$$

implies

$$\sum_{i,j}|\alpha_i^j - \alpha^{\tau(i,j)}_{\sigma(i,j)}| \leqslant c\sum_{i,j}|X_{p(i,j)}| \leqslant nk\,c|X|$$

so

$$e^{-|X|_{\min}} \leqslant \exp\Big(-c\sum_{i,j}|\alpha_i^j - \alpha^{\tau(i,j)}_{\sigma(i,j)}|\Big) \leqslant D(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n).$$

This establishes the estimate (6.4) and so completes Step 1.

Step 2. We must now show that

$$\sum_{\alpha_1,\dots,\alpha_n} D(\alpha_1,\dots,\alpha_n)\prod_i \|M_{\alpha_i}\|_t \leqslant c\|M\|_t^n. \tag{6.10}$$

The proof is similar to the corresponding proof in Step 6 of Theorem VI. 1 of II. Since D contains a finite number (depending only on k,n) of terms we need only consider one of them (i.e., one fixed but arbitrary σ and τ). We can represent its contribution to (6.10),

$$\sum_{\alpha_1,\dots,\alpha_n}\prod_{i,j}\exp(-c|\alpha_i^j-\alpha_{\sigma(i,j)}^{\tau(i,j)}|)\prod_i\|M_{\alpha_i}\|_t,$$

by a graph in which

(1) Each $\|M_{\alpha_i}\|_t$ is represented as a box with k dots (the k components of α_i) on its boundary.

(2) These boxes are arranged in a line in their natural order.

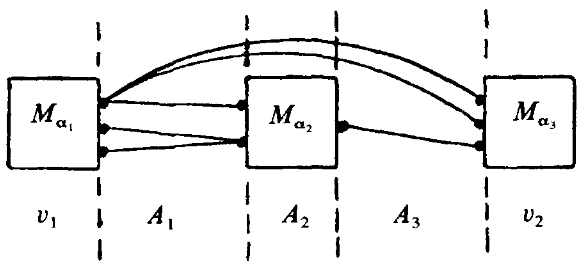
(3) Dots are joined in pairs by lines representing the decay factor $e^{-c|\alpha_i^j-\alpha_{i'}^{j'}|}$ of the pair of dots $(\alpha_i^j, \alpha_{i'}^{j'})$.

(4) Every dot has at least one line leaving it.

By use of the triangle inequality we can bound this graph by one which in addition satisfies

(5) All lines leaving a dot travel in the same direction.

We express such a graph in the form $\langle v_1, A_1\cdots A_p v_2\rangle$, where $v_i \in l^2(\mathbf{Z}^{2k})$ and the A_i's are operators between various l^2's, simply by drawing in vertical dotted lines such as



Then $v_1=\|M_{\alpha_1}\|_t$, $v_2=\|M_{\alpha_n}\|_t$, and A_i is the operator whose kernel contains everything between the ith pair of dotted lines. [When an edge crosses several dotted lines we rewrite it as $e^{-|\cdot-\cdot|}\,I\,I$, where I is the identity operator on $l^2(\mathbf{Z}^2)$.] In the example pictured

$$\begin{aligned}
&A_1(i_1,i_2,i_3;j_1,j_2,j_3,j_4)\\
&\quad=\exp(-c[|i_1-j_1|+|i_1-j_2|\\
&\qquad+|i_1-j_3|+|i_2-j_4|+|i_3-j_4|]),\\
&A_2(i_1,i_2,i_3,i_4;j_1,j_2,j_3)=\delta_{i_1j_1}\delta_{i_2j_2}\|M_{(i_3,i_4,j_3)}\|,\\
&A_3(i_1,i_2,i_3;j_1,j_2,j_3)=\delta_{i_1j_1}\delta_{i_2j_2}\exp(-c|i_3-j_3|).
\end{aligned}$$

The estimate is completed by invoking

$$\Big(\sum_\alpha\|M_\alpha\|_t^2\Big)^{1/2}\leqslant \text{const}\|M\|_t \quad \text{(Refs. 18 and 19)},$$

$$\|I\otimes\cdots\otimes I\otimes A\|\leqslant\|A\|_{H.S.},$$

$$\|A\|\leqslant\sup_{i_1\cdots}\sum_{j_1\cdots}|A(i_1\dots;j_1\dots)|$$

to estimate $\|\mathbf{v}_i\|$, A_i for i even, and A_i for i odd, respectively. ■

Theorem VI.3. In an $\epsilon P(\phi)_2$ model with covariance $C(t)$, $A_J^W\{J=0\}$ has a bounded inverse when viewed as an operator from $\mathscr{B}_t$ to $\mathscr{B}^*_t$.

Proof. The (k,k') entry of $A_J^W\{J=0\}$ is

$$\begin{aligned}
&\frac{\delta}{\delta J_1^k}\frac{\delta}{\delta J_{k'}}\ln\langle e^{J_m:\phi^m:}\rangle_t\Big|_{J=0}\\
&\quad=\frac{\delta}{\delta J_1^k}\langle:\phi^{k'}:e^{J_1:\phi:}\rangle_t\langle e^{J_1:\phi:}\rangle_t^{-1}\Big|_{J_1=0}\\
&\quad=\frac{\delta}{\delta J_1^k}\langle e^{J_1\phi}\langle e^{J_1\phi}\rangle_t^{-1}:\phi^{k'}:\rangle_t\Big|_{J_1=0}\\
&\quad=\langle:\phi^k:\,:\phi^{k'}:\rangle_t.
\end{aligned} \tag{6.11}$$

In a free theory $A_J^W\{J_k=0\}$ is trivially invertible since then $\langle:\phi^k:\,:\phi^{k'}:\rangle_t$ reduces to $\delta_{kk'}k!C(t)^{\otimes k}_s$. Hence it suffices to prove that $A_J^W\{J=0\}$ is norm differentiable in the coupling constant λ or in other words that

$$\Big|\frac{d}{d\lambda}\langle M:\phi^k:M':\phi^{k'}:\rangle_t\Big|\leqslant c\|M\|_t\|M'\|_t.$$

Now

$$\begin{aligned}
&\frac{d}{d\lambda}\langle M:\phi^k:M':\phi^{k'}:\rangle_t\\
&\quad=\sum_{\Delta,\alpha_i}\frac{\partial}{\partial\lambda_\Delta}\langle M_{\alpha_1}:\phi^k:M'_{\alpha_2}:\phi^{k'}:\rangle_t,
\end{aligned}$$

where Δ runs over lattice squares and

$$\int_\Lambda d^2x\,\lambda V(\phi(x))=\sum_\Delta\int_\Delta d^2x\,\lambda_\Delta V(\phi(x))\Big|_{\lambda_\Delta=\lambda}$$

is the interaction. So we wish to prove, in the notation of Theorem VI. 2, that

$$\begin{aligned}
&\Big|\frac{\partial}{\partial\lambda_\Delta}\langle M_{\alpha_1}:\phi^k:M'_{\alpha_2}:\phi^{k'}:\rangle_t\Big|\\
&\qquad\leqslant c\Big(\sum_{i,j}e^{-\epsilon\,\mathrm{dist}\,(\Delta,\alpha_i^j)}\Big)D(\alpha_1,\alpha_2)\|M_{\alpha_1}\|_t\|M'_{\alpha_2}\|_t.
\end{aligned}$$

That is, in addition to the decay factor $D(\alpha_1,\alpha_2)$ of Theorem VI.2, we must extract decay from Δ to some M or M' argument to enable us to perform the sum over Δ. This decay arises from the connectedness between the interaction vertex brought down by $\partial/\partial\lambda_\Delta$ and $M_{\alpha_1}:\phi^k:M'_{\alpha_2}:\phi^{k'}:$. We implement this connectedness by doubling the multifold theory of Theorem VI.2. Then every connected component of X in the cluster expansion that contains Δ must also intersect the support of at least one $\xi_{\alpha_i^j}$. The fact that every connected component of X must intersect the supports of $\xi_{\alpha_i^j}$ for at least two different values of i remains true even with the interaction term present: if $\langle[\phi^k]_1\psi\rangle^u_{X,\sigma(\Gamma)}$ is zero for all small enough λ_Δ then so is $(\partial/\partial\lambda_\Delta\langle[\phi^k]_1\psi\rangle^u_{X,\sigma(\Gamma)}$. ■

We now consider the regularity in t of $G^W\{J;t\}$ and $\Gamma\{A;t\}$ for $J\in\mathscr{B}_t$ and $A\in\mathscr{B}_t^*$. As defined in (6.2) $\mathscr{B}_t$ is the quadratic form domain of $\mathscr{C}(t)$, and $\mathscr{B}_t^*$ the form domain of $\mathscr{C}(t)^{-1}$. Since $\mathscr{C}(t)=t\mathscr{C}(1)+(1-t)\mathscr{C}(0)$ and $\mathscr{C}(0)\leqslant\mathscr{C}(1)\leqslant m_0^{-2}$ we have the natural inclusions (for $0<t\leqslant1$)

$$\mathscr{B}_0^*\subset\mathscr{B}_t^*=\mathscr{B}_1^*\subset\mathscr{L}^2\subset\mathscr{B}_1=\mathscr{B}_t\subset\mathscr{B}_0, \tag{6.12}$$

where $\mathscr{L}^2 = \bigoplus_{i=1}^{N} L^2((\mathbb{R})^i)_{\rm sym}$. By $\mathscr{B}_1 = \mathscr{B}_t$ we mean equality as sets although of course $\|f\|_t$ increases with t (and similarly for the equality $\mathscr{B}_t^* = \mathscr{B}_1^*$). Furthermore, all the norms $\|\cdot\|_t^*$ are independent of t when restricted to $\mathscr{B}_0^*$ so that $\mathscr{B}_0^*$ is a closed proper subspace of $\mathscr{B}_t^*$ for $t>0$. This follows from the fact (Lemma V.2b of II) that the forms $C(t)^{-1}$ restricted to the form domain of $C(0)^{-1}$ are independent of t. Note that the form domain of $C(0)^{-1}$ is just the Sobolev space[21]

$$W_0^{1,2}(\mathbb{R}^2\setminus\sigma) = \{f|f,\nabla f\in L^2, f\restriction\sigma\equiv 0\}.$$

To motivate the regularity results for G and Γ that we obtain in the following theorem, we consider the special case of $N=1$ for a free theory. Then

$$G = G_0\{J;t\} = \tfrac{1}{2}(J,C(t)J) = \tfrac{1}{2}\|J\|_t^2$$

and

$$\Gamma = \Gamma_0\{A;t\} = -\tfrac{1}{2}(A,C(t)^{-1}A) = -\tfrac{1}{2}\|A\|_t^{*2}.$$

Clearly, G_0 is defined for $J\in\mathscr{B}_t$ and we see from (6.12) that the largest possible domain on which G_0 can be differentiable in t (for fixed J) is $\mathscr{B}_1$. In fact, $\partial_t G_0\{J;t\} = \tfrac{1}{2}(J,(C(1)-C(0))J)$ on $\mathscr{B}_1$. On the other hand, Γ_0 is defined for $A\in\mathscr{B}_t^*$ and we see from the discussion following (6.12) that $\partial_t\Gamma_0\{A;t\} = 0$ on $\mathscr{B}_0^*$, the largest domain on which $\partial_t\Gamma_0\{A;0\}$ can exist. In those cases in which it is possible to modify $\Gamma^{(N)}$ so as to remove the C^{-1}'s explicitly (as in the case of $\Phi^{(1)}$ and $\Phi^{(2)}$ of I), we would anticipate that the modified $\Gamma^{(N)}$ would exist and be differentiable in t on a larger space than $\mathscr{B}_0^*$ (see Theorems V.5 and V.8 of II).

Theorem VI.4. For $\epsilon P(\phi)_2$ models,

(a) $G^W\{J;t\}\in\mathscr{F}(\mathscr{B}_t,\mathbb{C})$ and is C^∞ in t for $J\in\mathscr{B}_1$.

(b) ${}^W\Gamma\{A;t\}\in\mathscr{F}(\mathscr{B}_t^*,\mathbb{C})$ and is C^∞ in t for $A\in\mathscr{B}_0^*$.

Proof. (a) The proof that $\left\langle \prod_{i=1}^{n} J_{k_i}:\phi^{k_i}:\right\rangle_t$ is C^∞ in t when $J_{k_i}\in\mathscr{B}_1$ is based on the formula[15]

$$\partial_t\langle f(\phi)\rangle_{C(t)} = \tfrac{1}{2}\int dx\,dy\,\dot C(x,y)\left\langle \frac{\delta^2}{\delta\phi(x)\delta\phi(y)}f(\phi)\right\rangle_{C(t)}$$

$$= \sum_{\Delta_1,\Delta_2}\tfrac{1}{2}\int dx\,dy\,\dot C(x,y)\left\langle \chi_{\Delta_1}(x)\frac{\delta^2}{\delta\phi(x)\delta\phi(y)}\chi_{\Delta_2}(y)f(\phi)\right\rangle_{C(t)},$$

where $\langle\rangle_{C(t)}$ is the expectation with respect to the Gaussian measure of mean zero and covariance $C(t)$ and $\dot C = C(1)-C(0)\leqslant C(1)$. This formula is applied repeatedly to the finite volume approximations to $\left\langle \prod_{i=1}^{n} J_{k_i}:\phi^{k_i}:\right\rangle_t$ and the result is estimated just as in Theorems VI.2 and VI.3. Since the details are so similar to but easier than those of Theorem VI.1 of II. We shall not repeat them here.

(b) Define $\tilde A\{J;t\} = \mathscr{C}(t)^{-1}A^W\{J;t\}$ where, by an abuse of notation, $\mathscr{C}(t)^{-1}$ is the continuous extension of our old $\mathscr{C}(t)^{-1}$ to a unitary operator from $\mathscr{B}_t^*$ to $\mathscr{B}_t$. Clearly $\tilde A(t)\in\mathscr{F}(\mathscr{B}_t,\mathscr{B}_t)$ and $\tilde A^{-1} = A^{-1}\circ\mathscr{C}(t)\in\mathscr{F}(\mathscr{B}_t,\mathscr{B}_t)$. In addition by Lemma VI.5 (to follow)

$$\tilde A\{\cdot\,;t\}\in\mathscr{F}(\mathscr{B}_1,\mathscr{B}_1)\quad\text{and is } C^\infty \text{in } t,$$

$$\tilde A^{-1}\{\cdot\,;t\}\in\mathscr{F}(\mathscr{B}_1,\mathscr{B}_1)\quad\text{and is } C^\infty \text{in } t.$$

In essence what this says is that if J is restricted to lie in $\mathscr{B}_1$ then $\tilde A\{J;t\}$ is also in $\mathscr{B}_1$ and furthermore $\restriction\tilde A\mathscr{B}_1$ has an inverse defined on $\mathscr{B}_1$ (all in the sense of formal power series). This should not be surprising since in a free theory $\tilde A_J\{\cdot;t\}$ is the linear operator

$$F^{(N)} = \begin{bmatrix} I & & & \\ & 2I & 0 & \\ & & \ddots & \\ & 0 & & \\ & & & N!I \end{bmatrix}.$$

By the definition (2.5c) and Lemma II.2b we then have

$$\begin{aligned}{}^W\Gamma\{A;t\} &= G^W\{J\{A\};t\} - F_i\{A\}J_i\{A\}\\ &= G^W\{\tilde A^{-1}\{\mathscr{C}(t)^{-1}A;t\};t\} - F_i\{A\}\tilde A_i^{-1}\{\mathscr{C}(t)^{-1}A;t\}\\ &= G^W\{\tilde A^{-1}\{\mathscr{C}(1)^{-1}A;t\};t\} - F_i\{A\}\tilde A_i^{-1}\{\mathscr{C}(1)^{-1}A;t\}\end{aligned}$$

for all $A\in\mathscr{D}_0 = \{(A_1,\dots,A_N)|A_i$ symmetric,

$A_i\in D((-\Delta+m_0^2)^{\otimes i}_{\,s}), A_i(x_1,\dots,x_i) = 0$ if any $x_j\in\sigma\}$ by Lemma VI.6 (also to follow). Then the facts that $G^W\{\cdot\,;t\}\restriction\mathscr{B}_1$ is C^∞ in t and that $\tilde A^{-1}\{\cdot\,;t\}\restriction\mathscr{C}(1)^{-1}\mathscr{D}_0$ is C^∞ in t [note that $\mathscr{C}(1)^{-1}\mathscr{D}_0\subset\mathscr{B}_1$] imply that ${}^W\Gamma\{A;t\}$ is C^∞ in t whenever $A\in\mathscr{D}_0$. Furthermore the derivatives are continuous in the $\mathscr{B}_1^*$ topology. A uniformity argument allows us to use that continuity to extend the domain on which $\Gamma^W\{\cdot\,;t\}$ is C^∞ to $\overline{\mathscr{D}_0}^{\mathscr{B}_1^*} = \mathscr{B}_0^*$. ■

Lemma VI.5. In $\epsilon P(\phi)_2\tilde A\{\cdot;t\} = \mathscr{C}(t)^{-1}A\{\cdot;t\}$, upon restriction to $\mathscr{B}_1$, becomes a C^∞ (in t) in $\mathscr{F}(\mathscr{B}_1,\mathscr{B}_1)$ with C^∞ inverse.

Proof. The claim that $\tilde A\{J;t\}\in\mathscr{F}(\mathscr{B}_1,\mathscr{B}_1)$ is C^∞ amounts to the estimate

$$\begin{aligned}&\left|\frac{d^k}{dt^k}\langle J^{(1)},\mathscr{C}(t)^{-1}A_{n-1}(J^{(2)},\dots,J^{(n)};t)\rangle\right|\\ &\leqslant c\|J^{(1)}\|_1^*\|J^{(2)}\|_1\cdots\|J^{(n)}\|_1\\ &= c\|\mathscr{C}(1)^{-1}J^{(1)}\|_1\|J^{(2)}\|_1\cdots\|J^{(n)}\|_1.\end{aligned}$$

The proof of this estimate is the same as the proof that $G^W\{\cdot;t\}\restriction\mathscr{B}_1$ is PC^∞. We need only integrate by parts (see Lemma VI.7) to cancel the $\mathscr{C}(t)^{-1}$ and write $J^{(1)}$as $\mathscr{C}(1)\mathscr{C}(1)^{-1}J^{(1)}$. This in effect replaces the $C(t)$ propagators hooked to the first test function by $C(1)$ propagators and leaves $\mathscr{C}(1)^{-1}J^{(1)}$ as the first test function.

The invertibility of $\tilde A$ and its regularity in t follow from Lemma II.10 of II once we verify that the linear approximation to $\tilde A\{\cdot\,;t\}$ (viewed as a linear operator on $\mathscr{B}_1$) has a bounded inverse. But in a free theory the linear approximation (6.13) to $\tilde A\{\cdot\,;t\}$ is trivially invertible. By continuity in λ it is also invertible in all sufficiently weakly coupled $P(\phi)_2$ models. (See the proof of Theorem VI.3.) ■

Lemma VI.6. $\mathscr{C}(t)^{-1}\restriction\mathscr{D}_0$ is independent of t where

$\mathscr{D}_0 = \{(A_1,\dots,A_N)|A_i$ symmetric, $A_i\in D((-\Delta+m_0^2)^{\otimes i}_{\,s})$, $A_i(x_1,\dots,x_i) = 0$ if any $x_j\in\sigma\}$.

Proof. By Lemma C.2 of I $C(t)^{-1}\restriction\mathscr{D}_1$ is independent of t where

$$\begin{aligned}\mathscr{D}_1 &= D(-\Delta+m_0^2)\cap D(-\Delta_\sigma+m_0^2)\\ &= \{A_1|A_1\in D(-\Delta+m_0^2), A_1(x_1) = 0 \text{ if } x_1\in\sigma\}.\end{aligned}$$

Taking tensor products and direct sums yields the result. ■

While Theorem VI.4 assures us that Γ is C^∞ with respect to t it does not provide us with a means for calculating its t-derivatives. In particular the formula

$$\dot{\Gamma}\{A;t\} = -\tfrac{1}{2}\dot{C}(t)^{-1}(A_2 + A_1A_1 + 2A_1A_1^0) + \Gamma_{A_i}\dot{A}_i^0 \tag{6.14}$$

of Theorem III.1 cannot be true for all $A \in \mathscr{B}_0^*$. All of the $\dot{A}_i^0$'s are invariant under translations along σ. So every nonzero A_i^0 fails to have the decay required to be in $\mathscr{B}_0$. Furthermore $C(t)^{-1/2}\dot{A}_2^0C(t)^{-1/2}$, which in a free theory is $C(t)^{-1/2}$ $\dot{C}C(t)^{-1/2}$, is too singular near σ to be locally L^2. Hence in (6.14) we must restrict the class of allowed test functions A sufficiently to ensure that Γ_{A_i} is globally L^1 and locally sufficiently regular to allow smearing with $\dot{A}_2^0$. To make these requirements more precise consider the linear approximation $\Gamma_{A_iA_j}\{0\}$ to ΓA_i. By Theorem III.3b

$$-\Gamma_{A_iA_j}\{0\} = S^{-1}{}_{ij} = [\langle:\phi^\alpha::\phi^\beta:\rangle]^{-1}{}_{ij}.$$

In a free theory

$$-\Gamma_{A_iA_j}\{0\} = (j!)^{-1}\delta_{ij}(C(t)^{-1})^{\overset{i}{\otimes}}_s \equiv \overline{\mathscr{C}(t)^{-1}}{}_{ij}.$$

In an interacting theory, denoting $\langle:\phi^\alpha::\phi^\beta:\rangle - \overline{\mathscr{C}(t)}_{\alpha\beta}$ by $\Sigma_{\alpha\beta}$, we have

$$\begin{aligned}
-{}^W\Gamma_{A_iA_j}\{0\}A_j &= \Big[\mathscr{C}(t) + \Sigma\Big]^{-1}_{ij}A_j \\
&= \Big[1 + \mathscr{C}(t)^{-1}\Sigma\Big]^{-1}_{ik}\overline{\mathscr{C}(t)}{}^{-1}_{kj}A_j \\
&= \sum_{n=0}^{\infty}(-1)^n\Big[\Big(\overline{\mathscr{C}(t)}{}^{-1}\Sigma\Big)^n\overline{\mathscr{C}(t)}{}^{-1}A\Big]_i.
\end{aligned} \tag{6.15}$$

We can neutralize the final $\overline{\mathscr{C}(t)}{}^{-1}$ simply by requiring A to be in

$$\mathscr{N}_\sigma \equiv \bigoplus_{j=1}^{N} C_0^\infty(\mathbb{R}^d\setminus\sigma)^{\otimes j}_s,$$

for then $\overline{\mathscr{C}(t)}{}^{-1}A$ is independent of t and may be viewed as just another test function in $\mathscr{N}_\sigma$. The typical behavior of $\overline{\mathscr{C}(t)}{}^{-1}\Sigma$ may be determined by using integration by parts.

Lemma VI.7. (*Integration by parts with physical Wick ordering*). Let $\langle\cdot\rangle$ be the expectation in an $\epsilon P(\phi)_2$ model with covariance C. If $F, A_1,\ldots,A_n$ are polynominals in the field then

$$\begin{aligned}
&\Big\langle:(C^{-1}\phi)\prod_{i=1}^{n}A_i:F\Big\rangle \\
&= \sum_{k=1}^{n}\Big\langle:\frac{\delta A_k}{\delta\phi}\prod_{i\neq k}A_i:F\Big\rangle + \Big\langle:\prod_{i=1}^{n}A_i:\frac{\delta F}{\delta\phi}\Big\rangle \\
&\quad - \Big\langle:V'\prod_{i=1}^{n}A_i:F\Big\rangle,
\end{aligned} \tag{6.16}$$

where the meaning of the physical Wick dots depends explicitly on the indicated factorization, i.e.,

$$:B_1\cdots B_m: = \frac{\partial}{\partial\mu_1}\cdots\frac{\partial}{\partial\mu_m}\frac{e^{B\mu_i}}{\langle e^{B\mu_i}\rangle}\bigg|_{\mu_i=0}.$$

In particular if any B_i is independent of ϕ, $:B_1\cdots B_m: = 0$.

Remark 3. This lemma is very similar in spirit to Theorem 3.2.2 of Ref. 22 and its proof is a straightforward application of Theorem 3.1.1 of Ref. 22. Note the compatibility of physical Wick ordering with the operations of integration by parts and differentiation with respect to ϕ: for instance if in (6.16) $F = :\pi B_j:$ then we have

$$\frac{\delta}{\delta\phi}:\prod_{j=1}^{m}B_j: = \sum_{i=1}^{m}\frac{\delta B_i}{\delta\phi}:\prod_{j\neq i}B_j:.$$

Applying this lemma to $(\overline{\mathscr{C}}^{-1}\Sigma)_{21}$, for example, yields

$$\begin{aligned}
(\overline{\mathscr{C}}^{-1}\Sigma)_{21} &= \tfrac{1}{2}\langle:(C^{-1}\phi)(x_1)(C^{-1}\phi)(x_2)::\phi(y):\rangle \\
&= -\tfrac{1}{2}\delta(x_1-x_2)\langle:V''(x_1)::\phi(y):\rangle \\
&\quad + \tfrac{1}{2}\langle:V'(x_1)V'(x_2)::\phi(y):\rangle.
\end{aligned} \tag{6.17}$$

Equations (6.15) and (6.17) suggest that we choose the same space for the Γ_{A_i}'s (essentially the J_i's) and the $\mathscr{C}(t)^{-1}A_i$'s and that a typical "function" $f(x_1,\ldots,x_i)$ in that space be of the form $\Sigma_{\pi\in\mathscr{P}^i}\delta_\pi f_\pi(x_\pi)$, where $\mathscr{P}^i$ is the set of all partitions of $\{1,\ldots,i\}$, δ_π is a product of δ functions forcing all variables within each set of the partition π to be equal, x_π has one variable from each set of the partition π, and $f_\pi\in L^p(\mathbb{R}^{2|\pi|})$ for all $p<\infty$. For example if $\pi = \{\{1,2,3\}\{4\}\}$ then

$$\delta_\pi f_\pi(x_\pi) = \delta(x_1-x_2)\delta(x_1-x_3)f_\pi(x_1,x_4).$$

In order to separate local from global regularity we take advantage of the amalgam $l^q(L^p)$ spaces.[23] These spaces have the norms

$$\|f\|_{p;q} = \Big[\sum_{\Delta}\|\chi_\Delta f\|^q_{L^p}\Big]^{1/q}, \tag{6.18}$$

where the sum runs over a covering of $\mathbb{R}^2$ by unit lattice squares. (Make the obvious modification when $q=\infty$.) We require that each f_π be in the Fréchet space determined by the increasing family of norms $\{\|\cdot\|_{p;1}\,|p = 1,2,3,4,\ldots\}$. More precisely the space we will use for the J's and $\mathscr{C}(t)^{-1}A$'s is

$$\mathscr{D} = \{(J_1,\ldots,J_N|J_i\text{ symmetric},$$

$$J_i = \sum_{\pi\in\mathscr{P}^i}\delta_\pi J_{i,\pi}(x_\pi), |J|_{p;1}<\infty\;\forall p<\infty\}, \tag{6.19a}$$

where

$$|J|_p = \sum_{i=1}^{N}\sum_{\pi\in\mathscr{P}^i}\|J_{i,\pi}\|_{L^p}$$

and

$$|J|_{p;1} = \sum_{i=1}^{N}\sum_{\Delta}|\chi_\Delta J_i|_p$$

with Δ running over a covering of $\mathbb{R}^{2i}$ by unit lattice squares. Using the $l^q(L^p)$ mapping properties of C it is not hard to check that $\mathscr{D}\subset\mathscr{B}_1$ and that $(\dot{A}_i^0)$ is indeed in the dual space

$$\mathscr{D}^* = \{(A_1,\ldots,A_N)|A_i\text{ symmetric}, \exists p<\infty\ni|A|^*_{p';\infty}<\infty\}, \tag{6.19b}$$

where

$$|A|^*_{p'} = \sup_i\sup_{\pi\in\mathscr{P}^i}\|\pi A_i\|_{L^{p'}},$$

$$|A|^*_{p';\infty} = \sup_i\sup_{\Delta}|\chi_\Delta A_i|_{p'},$$

and πA_i is defined as follows.

Definition. Given any function A_i of i variables $x_j\in\mathbb{R}^2$ and any partition $\pi\in\mathscr{P}^i$ we define πA_i to be the function of

$|\pi|$ variables obtained from A_i be setting all variables in each set of π equal. For example if $\pi = \{\{1,2,3\},\{4\}\}$ then $(\pi A_4)(x,y) = A_4(x,x,x,y)$. We are using a somewhat schematic notation. $\mathscr{D}^*$ is really obtained by completion from C^∞ functions and consists of functions that have appropriate L^p properties with respect to measures like $\Sigma_{\pi\in\mathscr{P}^i}\delta_\pi dx_1\cdots dx_i$.

Remark. For $\Gamma^{(2)}$ it is possible to replace $\mathscr{D}$ by the Banach space

$$\mathscr{D}^{(2)} = \{(J_1,J_2) \mid \|J_1\|_{4/3;1} < \infty, J_2 \ \text{symmetric},$$
$$J_2 = \delta_{\pi_1} J_{2,\pi_1} + J_{2,\pi_2},$$
$$\|J_{2,\pi_1}\|_{2;1} < \infty, \|J_{2,\pi_2}\|_{4/3;1} < \infty\}, \tag{6.20}$$

where $\pi_1 = \{1,2\}$ and $\pi_2 = \{\{1\},\{2\}\}$. In the following proofs we shall identify those arguments that must be modified to effect this replacement.

We pause here to comment on the strategy involved in choosing different spaces in which J is to live (such as $\mathscr{B}_1$ and $\mathscr{D}$). In general, if $G\{J\} = G^W\{J\}$ is a functional (or fps) on the Fréchet space $\mathscr{X}$ then $\delta G/\delta J\{J\}\in\mathscr{X}^*$ so that we expect $A\{J\}$ to map into the dual space

$$J\in\mathscr{X}\to A\{J\} = \mathscr{C}J + F\{J\} \tag{6.21}$$

and, if this map is invertible, then the Legendre transform $\Gamma\{A\}$ of G is a functional (or fps) on $\mathscr{X}^*$. For certain purposes $\mathscr{X}^*$ may be too small (for example, it may not contain $\dot{A}^0$). We can then try cutting down $\mathscr{X}$ to $\mathscr{X}_0 \subset \mathscr{X}$ in order that $\mathscr{X}_0^* \supset \mathscr{X}^*$ be large enough (think of $\mathscr{X}$ as $\mathscr{B}_t$ and of $\mathscr{X}_0$ as $\mathscr{D}$). By itself such a reduction of the space is too simple-minded to work since the map (6.21) restricted to $\mathscr{X}_0$ will certainly not map onto $\mathscr{X}_0^*$ and we lose the crucial invertibility of $A\{J\}$. However, if we have the additional information that the perturbation F in (6.21) is sufficiently regular to absorb a $\mathscr{C}^{-1}$ and that $\mathscr{C}^{-1}F$ is "small," then A will map $\mathscr{X}_0$ onto $\mathscr{C}\mathscr{X}_0$. This is the point of introducing the variable $\tilde{A}\{J\} \equiv \mathscr{C}^{-1}A\{J\}$ which is then an invertible map of $\mathscr{X}_0$ onto itself (the proof of this is the crux of the argument!). As a result we obtain a Legendre transform $\Gamma\{A\}$ defined on $\mathscr{C}\mathscr{X}_0$, with (*a priori*) $\delta_A\Gamma\{A\}\in(\mathscr{C}\mathscr{X}_0)^*$. But $A\in\mathscr{C}\mathscr{X}_0 \Rightarrow \tilde{A}\in\mathscr{X}_0 \Rightarrow J\in\mathscr{X}_0 \Rightarrow \delta_A\Gamma\{A\}\in\mathscr{X}_0$ (since $\delta_A\Gamma$ is essentially J) and so $\delta_A\Gamma\{A\}\cdot\dot{A}^0$ is well defined if $\dot{A}^0\in\mathscr{X}_0^*$. $\delta_A\Gamma\{A\}\cdot\dot{A}^0$ is well defined if $\dot{A}^0\in\mathscr{X}_0^*$.

We shall henceforth denote the map $\mathscr{C}(t)^{-1}A^W\{J;t\}$ by $\tilde{A}\{J;t\}$ and its composition inverse (whose existence we prove in Theorem VI.11) by $\tilde{J}\{A;t\}$. Our battle plan is as follows. In Lemma VI.8 we prove that $\tilde{A}\{J;t\}\in\mathscr{F}(\mathscr{D},\mathscr{D})$ is C^∞ in t. In Lemma VI.10 we prove that the linear approximation $\tilde{A}_J\{0;t\}$ to $\tilde{A}$ has a continuous inverse as a map from $\mathscr{D}$ to $\mathscr{D}$. In Theorem VI.11 we combine the results of Lemmas VI.8 and VI.10 to prove the validity of the formula (6.14) for $\dot{\Gamma}\{A;t\}$.

Lemma VI.8. In $\epsilon P(\phi)_2$, $\tilde{A}\{J;t\}\in\mathscr{F}(\mathscr{D},\mathscr{D})$ is C^∞ in t.

Proof. The cornerstone of this proof is the estimate that for every n, N, and $p_1 < p_2 < \infty$ there exist constants c,ϵ (possibly depending on p_i, n, N) such that

$$|Z^W_{J_{i_1}J_{i_2}\cdots J_{i_n}}\{0\}(\mathscr{C}(t)^{-1}J^{(1)}_{i_1})J^{(2)}_{i_2}\cdots J^{(n)}_{i_n}|$$
$$= \left|\left\langle\left(\overset{i}{\otimes} C(t)^{-1}J^{(1)}_{i_1}{:}\phi^{i_1}{:}\right)(J^{(2)}_{i_2}{:}\phi^{i_2}{:})\cdots J^{(n)}_{i_n}{:}\phi^{i_n}{:}\right\rangle_t\right| \tag{6.22}$$

$$\leqslant c\sum_{\Delta_1,\ldots,\Delta_n}\max_{2<m<n} e^{-\epsilon d(\Delta_1,\Delta_M)}|\chi_{\Delta_1}J^{(1)}_{i_1}|^*_{p_1'}\prod_{m=2}^n|\chi_{\Delta_m}J^{(m)}_{I_m}|_{p_2},$$

where the c_m's are not summed over

$$\Delta_m = \Delta_{m,1}\times\cdots\times\Delta_{m,i_m}$$

with each $\Delta_{m,j}$ a unit square in $\mathbb{R}^2$,

$$d(\Delta_1,\Delta_m) = \sum_{j=1}^{i_1}\min_{1<j'<i_m}\operatorname{dist}(\Delta_{1,j};\Delta_{m,j'}).$$

The decay factor controls the sum over Δ_1 so that

$$|(\delta_J)^n Z^W\{0\}(\mathscr{C}(t)^{-1}J^{(1)})J^{(2)}\cdots J^{(n)}|$$
$$\leqslant c|J^{(1)}|^*_{p_1';\infty}\prod_{m=2}^n|J^{(m)}|_{p_2}.$$

This estimate ensures that $\mathscr{C}(t)^{-1}Z^W_J\{J;t\}\in\mathscr{F}(\mathscr{D},\mathscr{D})$. Since $\mathscr{C}(t){:}\mathscr{D}\subset\mathscr{B}_t\to\mathscr{B}^*_t\subset\mathscr{D}^*$, it also ensures that $Z^W_J\{J;t\}\in\mathscr{F}(\mathscr{D},\mathscr{D}^*)$ and hence that $Z^W\{J;t\}\in\mathscr{F}(\mathscr{D},\mathbb{C})$. The appropriate modification of Theorem VI.4 provides an estimate on $(\partial^k/\partial t^{k-1})\mathscr{C}(t)^{-1}Z^W_{J_{i_1}\cdots J_{i_n}}$ analogous to (6.22). It ensures that $\mathscr{C}(t)^{-1}Z^W_J\{J;t\}$ and $Z^W\{J;t\}$ are C^∞ in t as elements of $\mathscr{F}(\mathscr{D},\mathscr{D})$ and $\mathscr{F}(\mathscr{D},\mathbb{C})$, respectively. The conclusion of the lemma then follows from

$$\mathscr{C}(t)^{-1}G_J\{J;t\} = \mathscr{C}(t)^{-1}Z^W_J\{J;t\}/Z^W\{J;t\}$$

and

$$\tilde{A}_J\{J;t\} = F^{-1}\{\mathscr{C}(t)^{-1}G_J\{J;t\}.$$

[See the definition (2.6) of F and Lemma II.2.]

We now prove the estimate (6.22) (uniformly in the volume of a volume cutoff theory). Without loss of generality we may assume that each $J^{(m)}_{i_m}$ is nonzero for only one value of i_m and has support in a single hypercube Δ_m. The estimate takes four steps.

Step 1. (Implementation of the $\vdots\ \vdots$ subtractions by Ginibre's duplicate field trick) Just repeat the analogous step of Theorem VI.2.

Step 2. [Use of integration by parts to get rid of the $C(t)^{-1}$'s] We repeatedly apply (6.16) until all of the $C(t)^{-1}$'s of (6.22) are gone. This produces a sum of terms in which each argument of $J^{(1)}$ is connected by a δ function either to an argument of a $J^{(m)}$ with $m>1$ ($m\neq 1$ by Wick ordering) or to an interaction vertex.

Step 3. (Application of the cluster expansion) Apply the cluster expansion as in Theorem VI.2. [See (6.8).]

Step 4. (Estimation of the cluster expansion) The estimation procedure is again essentially the same as in Step 1 of Theorem VI.2. The only real difference is that we now wish to end up with a different norm on the $J^{(k)}$'s. Consider one term

$$\int_0^{s(\Gamma)} d\sigma(\Gamma)\partial^\Gamma\langle J^{(1)}_{i_1}\psi\rangle^u_{X,\sigma(\Gamma)}\frac{Z_{\Lambda\setminus X,\partial X}}{Z_\Lambda}$$

from the cluster expansion. Here ψ is a polynomial in the fields and contains

(i) $\prod_{m=2}^n J^{(m)}_{i_m}[\phi^{i_m'}]_m$ (recall that $[\phi^k]_m$ is just ${:}\phi^k{:}$ rewritten with duplicate fields; here $i_m'\leqslant i_m$ with strict inequality if some ϕ's have been differentiated in Step 2),

(ii) possibly some interaction vertices (produced in Step 2) and

(iii) a delta function propagator joining each argument of $J_{i_1}^{(1)}$ either to an interaction vertex or to an argument of a $J_{i_m}^{(m)}$ with $m > 1$.

It is possible that more than one argument of $J_{i_1}^{(1)}$ is connected to the same interaction vertex or $J_{i_m}^{(m)}$. As before, the Kirkwood–Salsburg equations are used to bound $Z_{\Lambda \setminus X, \partial X}/Z_\Lambda$ by $e^{O(1)|X|}$; the derivative ∂^Γ is evaluated; the Cauchy–Schwarz inequality is used to separate the interaction from the polynomial in the field downstairs and the Gaussian integral containing the latter is evaluated as a sum of graphs. (Note that as in Theorem VI.2 Wick ordering ensures that no two fields belonging to the same $J^{(k)}$ can contract directly to each other without intervening interaction vertices.)

Each graph in this sum is an integral whose integrand consists of our $J^{(k)}$'s, the δ function propagators of Step 2, some $\partial^\gamma C$ propagators from the evaluation of ∂^Γ, and some C propagators from the evaluation of the Gaussian integral. Do enough of the integral to dispose of all the delta functions including those built into the $J^{(m)}$'s with $m \geqslant 2$ by the definition of $\mathscr{D}$. In general this will result in terms with two or more arguments of $J_{i_1}^{(1)}$ set equal; we denote the resulting function by $\pi_1 J_{i_1}^{(1)}$ as in the definition (6.19b) of $\mathscr{D}^*$. So a typical integrand is of the form

$$(\overline{\pi_1 J_{i_1}^{(1)}})(\pi_1 J_{i_1}^{(1)}) \prod_{m=2}^{n} (\overline{J_{i_m,\pi_m}^{(m)}})(J_{i_m,\pi_m'}^{(m)})G,$$

where G is the value of a graph which contains all the C's and $\partial^\gamma C$'s and which has arguments of the $J^{(k)}$'s as external variables. (Do not forget that we have already been through a Schwarz inequality — that is why we now have $2n$ J's.) Each integration variable can appear simultaneously in

(i) one of $\overline{\pi_1 J_{i_1}^{(1)}}$ and $\pi_1 J_{i_1}^{(1)}$,

(ii) one of $\overline{J_{i_2,\pi_2}^{(2)}}, J_{i_2,\pi_2'}^{(2)} \cdots J_{i_n,\pi_n'}^{(n)}$,

(iii) G.

Any such integral may be estimated by the following generalization,[20,4] of Hölder's inequality.

Lemma VI.9. If $\int d\mu_k = 1$ for all k in the finite index set K, then

$$\left| \int_{k \in K} d\mu_k(x_k) \prod_m g_m(x^{K_m}) \right| \leqslant \|g_m\|_{L^{q_m}},$$

where $K_m \subset K, x^{K_m} = \{x_k | k \in K_m\}$ and for each k

$$\sum_{m \ni k \in K_m} \frac{1}{q_m} \leqslant 1.$$

We simply assign $q = p_1'$ to $\overline{\pi_1 J_{i_1}^{(1)}}$ and $\pi_1 J_{i_1}^{(1)}$, $q = p_2$ to all the $\overline{J_{i_m,\pi_m'}^{(m)}}$'s and $J_{i_m,\pi_m'}^{(m)}$'s with $m \geqslant 2$ and $q = p_3 \equiv (1/p_1 - 1/p_2)^{-1}$ to G (recall $p_2 > p_1$).

Since $\|G\|_{p_3} \leqslant c e^{O(1)|X| - K(m_0)|\Gamma|}$ by standard cluster expansion techniques we now have, all together,

$$\left| \int_0^{s(\Gamma)} d\sigma(\Gamma) \partial^\Gamma \langle \psi \rangle^u_{X,\sigma(\Gamma)} \frac{Z_{\Lambda \setminus X, \partial X}}{Z_\Lambda} \right|$$
$$\leqslant c e^{O(1)|X| - K(m_0)|\Gamma|} |\chi_{\Delta_1} J_{i_1}^{(1)}|^*_{p_1'} \prod_{m=2}^{n} |\chi_{\Delta_m} J_{i_m}^{(m)}|_{p_2}.$$

Continuing as in Theorem VI.2 we have

$$|Z^W_{J_{i_1} \cdots J_{i_n}}(\mathscr{C}(t)^{-1} J_{i_1}^{(1)}) J_{i_2}^{(2)} \cdots J_{i_n}^{(n)}|$$
$$\leqslant c \sum_{X,\Gamma} e^{O(1)|X| - K(m_0)|\Gamma|} |\chi_{\Delta_1} J_{i_1}^{(1)}|^*_{p_1'} \prod_{m=2}^{n} |\chi_{\Delta_m} J_{i_m}^{(m)}|_{p_2}$$
$$\leqslant c e^{-|X|_{\min}} |\chi_{\Delta_1} J_{i_1}^{(1)}|^*_{p_1'} \prod_{m=2}^{n} |\chi_{\Delta_m} J_{i_m}^{(m)}|_{p_2}.$$

Now for each j the component of X containing $\Delta_{1,j}$ must also contain a $\Delta_{m,j'}$ for some $m > 1$ so that for all j, dist $(\Delta_{1,j}, \Delta_{m,j'}) \leqslant c|X|_{\min}$. Hence $\min_m d(\Delta_1, \Delta_m) \leqslant c|X|_{\min}$ and (6.22) follows. ■

Remark 5. For the special case of $\Gamma^{(2)}$ it is always possible to assign, in Lemma VI.9, $q = 4$ to $G, J_2^{(1)}(x,y)$ and $J_1^{(1)}(x)$, $q = 2$ to $J_2^{(1)}(x,x)$ and $J_{2,\{1,2\}}^{(m)}$ for all $m \geqslant 2$, and $q = \frac{4}{3}$ to $J_1^{(m)}$ and $J_{2,\{\{1\},\{2\}\}}^{(m)}$ for all $m \geqslant 2$. This is because $J_2^{(1)}(x,x)$ occurs if and only if both arguments of $J_2^{(1)}$ are connected to the δ functions of a single $J_{2,\{1,2\}}^{(m)}$ with $m \geqslant 2$. Thus we may choose the space of J's to be the Banach space $\mathscr{D}^{(2)}$ of Remark 4. Moreover, it follows by continuity in the coupling constant that $\widetilde{A}_J\{0\}$ has a continuous inverse as a linear transformation on $\mathscr{D}^{(2)}$. For the general case $N > 2$ we establish the corresponding invertibility in

Lemma VI.10. $\widetilde{A}_J\{0\}: \mathscr{D} \to \mathscr{D}$ has a continuous inverse.

Proof. Now (see the proof of Theorem VI.3)

$$\widetilde{A}_J\{0\}_{\alpha\beta} = \langle (\mathscr{C}(t)^{-1} :\phi^\alpha:) :\phi^\beta: \rangle.$$

Integrate by parts (using Lemma VI.7) sufficiently often that $\mathscr{C}(t)^{-1}$ disappears. Denote by S those arguments of $:\phi^\alpha:$ that contract directly to arguments of $:\phi^\beta:$ the latter being denoted T (so $|S| = |T|$). The remaining arguments of $:\phi^\alpha:$ (if any) naturally partition themselves with all the arguments in each subset of the partition contracting to the same interaction vertex. Thus

$$\widetilde{A}_J\{0\}_{\alpha\beta} = \sum_{\substack{S \subset \{1 \cdots \alpha\} \\ T \subset \{1 \cdots \beta\}}} \sum_{\pi \in \mathscr{P}^{S'}} |S|! \delta_{S,T}$$
$$\left\langle :\prod_{\gamma \in \pi} (-1)|\gamma|! \delta_\gamma V^{(|\gamma|)}: \ :\phi^{T'}: \right\rangle, \tag{6.23}$$

where $S' = \{1 \cdots \alpha\} \setminus S$, $T' = \{1 \cdots \beta\} \setminus T$, $\delta_{S,T}$ is a symmetrized product of delta functions forcing each argument of S equal to a different argument of T, and δ_γ is a product of delta functions forcing all arguments of γ to equal the argument of $V^{(|\gamma|)}$, which is the $|\gamma|$th derivative of the interaction. Note that, thanks to the Wick ordering, if either S' or T' is nonempty they both must be nonempty. We can thus rewrite (6.23) as

$$\widetilde{A}_J\{0\}_{\alpha,\beta} = F^{(N)}_{\alpha,\beta} + \sum_{\substack{S \subsetneq \{1,\ldots,\alpha\} \\ F \subsetneq \{1,\ldots,\beta\}}} |S|! \delta_{S,T} \sum_{\pi \in \mathscr{P}^{S'}} \left(\prod_{\gamma \in \pi} \delta_\gamma \right) f_{\pi,T'},$$

where $F^{(N)}$ is the free theory value of $\widetilde{A}_J$ given by (6.13) and

$$f_{\pi,T'} = \left\langle :\prod_{\gamma \in \pi} (-1)|\gamma|! V^{(|\gamma|)}: \phi^{T'}: \right\rangle.$$

Now $f_{\pi,T'}$ is a regular function except for singularities at coinciding arguments that are at worst powers of logarithms and has exponential decay from each argument of T' to some argument of S' and vice versa. Furthermore, again thanks to the Wick ordering, there is no probelm in setting equal various arguments of T' to give $f_{\pi,T'}\pi'$ (with the π' acting on the arguments in T'), as is necessary when $\widetilde{A}_J\{0\}$ acts on an element of $\mathscr{D}$. Thus for any $p<\infty$, there exists a constant c_p (depending only on p and N) such that

$$\|\chi_{\Delta_\pi} f_{\pi,T'}\chi_{\Delta_{T'}}\pi'\|_{L^p} \leqslant \lambda c_p E_{\Delta_\pi \Delta_{T'}}, \tag{6.24a}$$

provided the coupling constant $\lambda \leqslant \lambda_0$. Here Δ_π and $\Delta_{T'}$ are hypercubes giving the localizations of incoming and outgoing variables of $f_{\pi,T'}$ and $E_{\Delta_\pi\Delta_{T'}}$ is an exponential decay factor obeying

$$\sup_{\Delta_{T'}} \sum_{\Delta_\pi} E_{\Delta_\pi \Delta_{T'}} \leqslant c < \infty. \tag{6.24b}$$

To prove that $\widetilde{A}_J\{0\}$ has a continuous inverse we use a Neumann expansion about the diagonal operator $F^{(N)}$:

$$J^{(1)}[\widetilde{A}_J\{0\}^{-1} - F^{(N)^{-1}}]J^{(2)}$$
$$= \sum_{n=1}^{\infty} (-1)^n J^{(1)}_{\alpha_1} \frac{1}{\alpha_1!} \prod_{i=1}^{n} \left\{ \sum_{\substack{S_i \subsetneq \{1\ldots\alpha_i\} \\ T_i \subsetneq \{1\ldots\alpha_{i+1}\}}} |S_i|!\delta_{S_i T_i} \times \sum_{\pi_i\in\mathscr{P}^{S_i'}} \left(\prod_{\gamma\in\pi_i}\delta_\gamma\right) f_{\pi_i T_i'} \frac{1}{\alpha_{i+1}!}\right\} \left\{\sum_{\pi\in\mathscr{P}^{\alpha_{n+1}}} \delta_\pi J^{(2)}_{\alpha_{n+1},\pi}\right\}, \tag{6.25}$$

where the α_i's are "integrated over" and where we have inserted the representation for $J^{(2)}$ as an element of $\mathscr{D}$.

It suffices to show that given any $p_1<\infty$ there exist constants $c, p_2<\infty$ (depending only on p_1, N) and λ_1 (depending only on N and *not on* p_1) such that the right hand side of (6.25) is bounded by $c|J^{(1)}|^*_{p_1'}|J^{(2)}|_{p_2}$ whenever $0\leqslant\lambda\leqslant\lambda_1$.

Pick any $n>0$. Corresponding to that n there are in (6.25) at most $N\{N2^N2^N|\mathscr{P}^N|\}^n\{|\mathscr{P}^N|\}\equiv c_1c_2^n$ terms, each consisting of a product of numerical factors (no larger than 1 in magnitude since $(1/\alpha_i!)|S_i|!\leqslant 1$) and an integral

$$J^{(1)}_{\alpha_1}\prod_{i=1}^n\left\{\delta_{S_iT_i}\left(\prod_{\gamma\in\pi_i}\delta_\gamma\right)f_{\pi_i T_i'}\right\}\{\delta_\pi J^{(2)}_{\alpha_{n+1},\pi}\}.$$

Do enough of this integral to get rid of all the δ functions. This leaves us with a true integral whose integrand consists of one $\pi_1' J^{(1)}_{\alpha_1}$, n $f_{\pi_i T_i'}\pi_i'$'s, and one $J^{(2)}_{\alpha_{n+1},\pi}$. The integration variables which appear in $\pi_1' J^{(1)}_{\alpha_1}$ may also appear in $J^{(2)}_{\alpha_{n+1},\pi}$ and up to N $f_{\pi_i T_i'}\pi_i'$'s in total. Any single integration variable which does not appear in $\pi_1' J^{(1)}_{\alpha_1}$ may appear in a total of $(N+1)$ kernels ($J^{(2)}_{\alpha_{n+1},\pi}$ and/or $f_{\pi_i T_i'}\pi_i'$'s). Hence we may now localize each integration variable by $1=\Sigma_\Delta\chi_\Delta$ and apply Lemma VI.9 with $q=p_1'$ assigned to $J^{(1)}$, $q=p_2\equiv(N+1)p_1$ assigned to $J^{(2)}$, and to those (up to N) $f_{\pi_i T_i'}\pi_i'$'s having variables in common with $J^{(1)}$ and $q=N+1$ assigned to all remaining $f_{\pi_i T_i'}\pi_i'$'s. The sum over localizations may be controlled by (6.24b) and we are left with

$$|J^{(1)}[\widetilde{A}_J\{0\}^{-1} - F^{(N)^{-1}}]J^{(2)}|$$
$$\leqslant \sum_{n=1}^{\infty} |J^{(1)}|^*_{p_1'} c_1 c_2^n \lambda^n c^N_{(N+1)p_1} c^{n-N}_{N+1} |J^{(2)}|_{(N+1)p_1}.$$

This does indeed converge to give the desired bound provided $\lambda<\lambda_1\equiv\min[\lambda_0, K_2 c_{N+1})^{-1}]$. Note that c_{N+1} and hence the condition on the coupling constant are independent of p_1. ∎

Theorem VI.11. In $\epsilon P(\phi)_2$ we have, for $A\in N_\sigma$ and $N\geqslant 2$,

$$\dot{\Gamma}\{A;t\} = \Gamma_{A_i}\{A;t\}\dot{A}^0_i. \tag{6.26}$$

Proof. Suppose $A\in\mathscr{N}_\sigma$. Then (6.26) is a consequence of three facts. Firstly, as in Theorem VI.4, with $\widetilde{J}=\widetilde{A}^{-1}$

$$\dot{\Gamma}\{A,t\}$$
$$=\frac{\partial}{\partial t}[G^W\{\widetilde{J}\}\mathscr{C}(t)^{-1}A;t\};t\} - F_i\{A\}\widetilde{J}_i\{\mathscr{C}(t)^{-1}A;t\}]$$
$$=\dot{G}^W\{\widetilde{J}\{\mathscr{C}(1)^{-1}A;t\};t\}$$

since $\mathscr{C}(t)^{-1}A=\mathscr{C}(1)^{-1}A$.

Secondly, by (2.11),

$$\dot{G}^W\{J\} = \partial_t G\{W^{-t}J\}$$
$$=\dot{G}\{W^{-t}J\}+G_{J_j}\{W^{-t}J\}(\dot{W}^{-t}J)_j. \tag{6.27}$$

We compute the first term using (3.2),

$$\dot{G}\{W^{-t}J\} = -\tfrac{1}{2}\dot{C}^{-1}(G_{J_2}\{W^{-t}J\}-G_{J_2}\{0\})$$
$$= -\tfrac{1}{2}\dot{C}^{-1}(A_2^W + A_1^W A_1^W + 2A_1^W A_1^0)\{J\}.$$

As for the second term, we have, by Lemma II.2b,

$$G_{J_j}\{W^{-t}J\} = \sum_{l=0}^N G^W_{J_l}\{J\}(W^{-t})^{-1}_{lj}$$
$$=\sum_{l=0}^N F_l\{A^W\{J\}\}(W^{-t})^{-1}_{lj},$$

and from the definition

$$W^{-t}_{jk} = \binom{k}{j}\delta_f^{k-j}\exp\left(-\sum_{n=1}^{\infty}A^0_n f^n/n!\right)\Big|_{f=0}$$

we have

$$\dot{W}^{-t}_{jk} = -\sum_{i=1}^{k-j}\binom{k}{i}W^{-t}_{j,k-i}\dot{A}^0_i.$$

Hence the second term in (6.27) is

$$-\binom{k}{i}F_l\{A^W\{J\}\}(W^{-t})^{-1}_{lj}W^{-t}_{j,k-i}\dot{A}^0_iJ_k$$
$$=-\binom{k}{i}F_{k-i}\{A^W\{J\}\}\dot{A}^0_iJ_k$$

and so

$$\dot{G}^W\{J\} = -\tfrac{1}{2}\dot{C}^{-1}(A_2^W + A_1^W A_1^W + 2A_1^W A_1^0)\{J\}$$
$$-\binom{k}{i}F_{k-i}\{A^W\{J\}\}\dot{A}^0_iJ_k.$$

This formula is valid for all $J\in\mathscr{D}$ by Lemma VI.8.

Thirdly, by Lemma II.10 of II the composition inverse $\widetilde{J}\{A\}$ to $\widetilde{A}\{J\}$ exists in $\mathscr{F}(\mathscr{D},\mathscr{D})$ since $\widetilde{A}\{J\}\in\mathscr{F}(\mathscr{D},\mathscr{D})$ (Lemma VI.8) and the linear approximation $\widetilde{A}_J\{0\}$ has a continuous inverse on $\mathscr{D}$ (Lemma VI.10).

Putting all three facts together we have that for $A\in\mathscr{N}_\sigma\subset\mathscr{D}$

$$\begin{aligned}\dot{\Gamma}\{A;t\} &= -\tfrac{1}{2}\dot{C}^{-1}[A_2 + A_1A_1 + 2A_1A_1^0] - \binom{k}{i}F_{k-i}\{A\}J_k\dot{A}_i^0\\ &= \Gamma_{A_i}\dot{A}_i^0,\end{aligned}$$

where we may drop the first term for $A\in\mathcal{N}_\sigma$ [see (3.17) of I] and where we have used (the analog for ${}^W\Gamma$ of) Theorem III.3a. ∎

We note that formula (6.26) is valid only for a very restricted class of test functions A. This class may be extended a bit beyond $\mathcal{N}_\sigma$ by continuity but certainly does not include $\dot{A}_i^0$. Hence we may not iterate this formula, as it stands, to evaluate higher derivatives. However, if we choose a sequence $\{\zeta_n\}$ of functions in $C_0^\infty(\mathbb{R}^d/\sigma)$ that are bounded by 1 and converge pointwise to 1 then

$$a_{i,n} \equiv \dot{A}_i^0(\zeta_n^{\otimes i}) \in \mathcal{N}_\sigma$$

and

$$\dot{\Gamma}\{A;t\} = \lim_{n\to\infty} a_{i,n}\Gamma_{A_i}\{A,t\} \tag{6.28}$$

may be iterated to evaluate higher derivatives since the appropriate limits are uniform in t. We shall exploit (6.28) and its iterates in our proofs[5] of the irreducibility properties of $\Gamma^{(3)}$ and $\Gamma^{(4)}$.

[1]C. De Dominicis and P. Martin, J. Math. Phys. **5**, 14 (1964); **5**, 31 (1964).

[2]A. N. Vasil'ev and A. K. Kazanskii, Theor. Math. Phys. **14**, 215 (1973).

[3]A. Cooper, J. Feldman, and L. Rosen, "Legendre Transforms and r-Particle Irreducibility in Quantum Field Theory. I. The Formalism for $r = 1$, 2," Ann. Phys. **137**, 146 (1981).

[4]Reference 3, "Legendre transforms and r-particle irreducibility in quantum field theory: The formal power series framework," Ann. Phys. **137**, 213 (1981).

[5]Reference 3, "Cluster Irreducibility of the third and fourth Legendre transforms in quantum field theory," Phys. Rev. D **25**, 1565 (1982).

[6]A. N. Vasil'ev and A. K. Kazanskii, Theor. Math. Phys. **12**, 875 (1972); Yu. M. Pis'mak, Theor. Math. Phys. **18**, 211 (1974); A. N. Vasil'ev, A. K. Kazanskii, and Yu. M. Pis'mak, Theor. Math. Phys. **20**, 181 (1974).

[7]T. Spencer and F. Zirilli, Commun. Math. Phys. **49**, 1 (1976).

[8]H. Koch, Ann. Inst. Henri Poincaré **31**, 173 (1979).

[9]J. Dimock and J.-P. Eckmann, Commun. Math. Phys. **51**, 41 (1976); Ann. Phys. **103**, 289 (1977).

[10]M. Combescure and F. Dunlop, Ann. Phys. **122**, 102 (1979).

[11]J. Imbrie, Commun. Math. Phys. **78**, 169 (1980).

[12]T. Spencer, Commun. Math. Phys. **44**, 143 (1975).

[13]J. Glimm and A. Jaffe, in *International Symposium on Mathematical Problems in Theoretical Physics*, edited by H. Araki, *Springer Lecture Notes in Physics*, Vol. 39 (Springer, Berlin, 1975).

[14]A. Cooper, J. Feldman, and L. Rosen, "Legendre transforms and r-particle irreducibility for the Yukawa model" (in preparation).

[15]J. Dimock and J. Glimm, Adv. Math. **12**, 58 (1974).

[16]J. Glimm, A. Jaffe, and T. Spencer, Ann. Math. **100**, 585 (1974).

[17]J. Glimm and A. Jaffe, Phys. Rev. D **11**, 2816 (1975).

[18]J. Glimm and A. Jaffe, Ann. Inst. Henri Poincaré A **21**, 1 (1975).

[19]A. Cooper, J. Feldman, and L. Rosen, "The Second Legendre Transform for the Weakly Coupled $P(\phi)_2$ Model" (in preparation).

[20]J. Glimm, A. Jaffe, and T. Spencer, in *Constructive Quantum Field Theory*, edited by G. Velo and A. Wightman, *Springer Lecture Notes in Physics*, Vol. 25 (Springer, Berlin, 1973), pp. 199–242.

[21]R. Adams, *Sobolev Spaces* (Academic, New York, 1975).

[22]J. Glimm and A. Jaffe, Commun. Math. Phys. **44**, 293 (1975).

[23]F. Holland, J. London Math. Soc. **10**, 295 (1975); M. Birman and M. Solomjak, Vest. LSU Math. **2**, 9 (1975).

The holonomy group and spontaneous symmetry breaking

D. K. Ross
Physics Department, Iowa State University, Ames, Iowa 50011

(Received 9 September 1981; accepted for publication 8 December 1981)

We show that the residual symmetry group after spontaneous symmetry breaking must either be the holonomy group H of the connection or contain H as a subgroup. Since H is a connected, normal Lie subgroup of the gauge group G, knowing the dimension of H is often sufficient to determine H itself. We develop an algorithm for determining dim H for any gauge theory given the structure constants of the gauge group. This algorithm is then applied to a fiber bundle with an SU_2 or SU_3 gauge group.

PACS numbers: 11.30.Qc, 02.20. + b

I. INTRODUCTION

Gauge theories play a key role in our understanding of all four of the basic interactions of physics. Electromagnetism is the prototype gauge theory with an abelian U_1 gauge group.[1] In more recent years the weak interactions have also been understood as a gauge theory through the work of Weinberg[2] and Salam,[3] unifying the weak and electromagnetic interactions through an $SU_2 \times U_1$ gauge group. The strong interactions also appear to be described as the color SU_3 gauge theory,[4] and even gravity is the gauge theory of T_4 or the Poincaré[1] group as shown originally by Utiyama[5] and more recently by others.[6,7]

The most natural and most elegant mathematical description of these ubiquitous gauge theories is through fiber bundles. A principal fiber bundle can be thought of as a generalized topological product of a base space such as four-dimensional Minkowski space and the structure group. In more detail, a gauge is a global section and a gauge field is identified with the connection 1-form in a principal fibration. The Yang–Mills field strengths become the coefficients in some basis of the curvature 2-form. This description goes back to Trautman.[8] Detailed definitions can be found in various texts.[9–11] Because they are such a natural description of gauge theories, fiber bundles are rapidly coming into physics literature. For example, Wu and Yang[12] have used fiber-bundle concepts in their treatment of magnetic monopoles, and Cho[7] has used fiber bundles in his treatment of gravity as a gauge theory.

Spontaneous symmetry breaking and the Higgs[13] mechanism for generating masses for gauge fields is a crucial ingredient in modern gauge theories. This plays a key role in the Weinberg–Salam unification of electromagnetism with the weak interactions. It also plays a key role in the grand unified theories[14] which promise to unite the electromagnetic, weak, and strong interactions under a single large group such as SU_5. Extended $N = 8$ supergravity[15] or some larger scheme holds the promise of perhaps uniting gravity with all of the other interactions. Here, we are faced with a large graded Lie algebra with probably sequential symmetry breakings first to something like SU_5 and then down to what we actually observe. Spontaneous symmetry breaking is usually put in by hand, and this suffices for small gauge groups. For the large complicated groups encountered in various unification schemes, however, this is quite unsatisfactory, and some guidance from the underlying fiber bundle structure would be quite useful. We hope to shed some light on this problem in the present paper.

Once we appreciate the importance of spontaneous symmetry breaking and the naturalness of a fiber bundle description of gauge theories, we are led to investigate spontaneous symmetry breaking in fiber bundles. No satisfactory theory has emerged yet, although preliminary work by Mayer[10,16] suggests that the problem involves a description of quantized gauge fields in terms of operator-valued distributions with the Higgs mechanism involving affine vector bundles. More recent work[17] has centered on dimensional reduction of higher-dimensional spaces. The problem of working out the details of spontaneous symmetry breaking in a fiber bundle with full mathematical rigor is clearly a very difficult one. It is certainly worth doing, however, since the underlying fiber-bundle structure is there whether we realize it or not. Properly taking into account this structure will greatly clarify spontaneous symmetry breaking and is certainly superior to putting this breaking in by hand. An example of a similar clarification is the realization that the "vector potential" A_μ in electromagnetism is really the coefficient of a connection form in a principal fiber bundle. This realization leads one directly to the peculiarities of a Bleuler–Gupta quantization scheme.

In this paper, we sidestep the major mathematical difficulties and address ourselves to the problem of what can be said about the observable effects of spontaneous symmetry breaking in a general gauge theory, taking into account the fiber-bundle structure but not doing the spontaneous symmetry breaking in the fiber-bundle in detail. We will need to know only that the fiber-bundle structure exists and that spontaneous symmetry breaking has taken place. We will show, in fact, that the residual symmetry group after spontaneous symmetry breaking must either be the holonomy group of the connection or contain that group. Mayer[10,16] mentioned this possibility. Loos[18] discussed the internal holonomy groups of Yang–Mills fields but did not consider their role in spontaneous symmetry breaking.

In Sec. II below, we define the holonomy group of the connection and discuss its role in spontaneous symmetry breaking. We calculate the holonomy group of the connection of a principal fiber bundle with SU_2 and SU_3 structure groups in Sec. III and discuss our results and conclusions in Sec. IV.

 0022-2488/82/050869-04$02.50

II. HOLONOMY GROUP

We can define the holonomy group as follows: Let $\lambda = (P,G,B,\pi)$ be a principal fibration with structure group the Lie group G, base space B, and projection π. Also let Γ be the connection and E an associated vector bundle. Any path $\gamma(t)$ in B (a piecewise smooth mapping of the interval $[0,1]$ into B) can be lifted to a horizontal path $Z(t)$ in P. The path γ and the connection Γ determine an isomorphism of the fiber $E_{\gamma(0)}$ onto the fiber $E_{\gamma(1)}$. If γ is a loop so that $\gamma(0) = (1)$, the two fibers coincide and the horizontal lift induces in P a fiber isomorphism of the fiber over $b = \gamma(0)$ onto itself and a linear isomorphism of $E_{\gamma(0)}$ onto itself. The set of all such isomorphisms is a group called the holonomy group of the connection at b. If b_1 and b_2 are two points in B which can be joined by a path, the holonomy groups at these points are isomorphic, and it makes sense to speak of the holonomy group of the fibration. The restricted holonomy group is the subset of the holonomy group obtained by considering only null-homotopic loops.

An important property of the holonomy group is contained in the following theorem from Kobayashi and Nomizu[11] and others.[19]

Theorem: Let $P(B,G,\pi)$ be a principal fiber bundle with projection π and structure group G whose base manifold B is connected and parcompact (admits locally finite open coverings). $H(u)$ and $H^0(u)$, $u \in p$, are the holonomy group and restricted holonomy group of a connection Γ with reference point u. Then (a) $H^0(u)$ is a connected Lie subgroup of G; (b) $H^0(u)$ is a normal subgroup of $H(u)$, and $H(u)/H^0(u)$ is countable.

In our work we will take the base manifold B to be Minkowski space, since we are interested in gauge theories defined over Minkowski space. Minkowski space is a metric space, and all metric spaces are paracompact[9] so that this theorem applies to our situation. Also, we need not distinguish between $H(u)$ and $H^0(u)$, and the holonomy group is a Lie subgroup of the gauge group G. In the following, we will also restrict our attention to semisimple G, since Loos has pointed out that for a nonsemisimple gauge group G, one may have G an invariant subgroup of H, rather than H a subgroup of G

We now want to show that the holonomy group H as defined above is the smallest possible residual symmetry group of the theory after spontaneous symmetry breaking. Let $\phi_i(r)$ be an n-component real field transforming as a representation $\mathbf{r}$ of the gauge group G. ϕ_i is the cross section of an associated vector bundle associated with the principal fibration via the representation $\mathbf{r}$. Assume that spontaneous symmetry breaking takes place (not necessarily by the Higgs mechanism) so that one or more components of ϕ_i develop a nonzero vacuum expectation value. The vacuum will then no longer be invariant under the full gauge group G, and we can define a residual symmetry group R which is a subgroup of G and which leaves $\langle 0|\phi_i|0\rangle$ invariant. We wish to get some handle on R.

Now using the connection, we can execute a parallel displacement of multiplets around a closed loop in event space (Minkowski space) from x^k back to x^k. Doing this for all closed loops in event space results in a group of linear multiplet transformations at x^k. This group is a representation of the holonomy group $H(x^k)$, from the above definitions. Let the specific multiplet, which is to be carried around the closed loop, be the $\phi_i(\mathbf{r})$ mentioned above. The resulting holonomy group structure has to do with the properties of the connection but has nothing whatever to do with whether or not ϕ_i has a vacuum expectation value. Thus the holonomy group will certainly leave $\langle 0|\phi_i|0\rangle$ invariant. In fact the physical residual symmetry group R of $\langle 0|\phi_i|0\rangle$ certainly must contain the group obtained as we carry the space-time dependent quantity $\langle 0|\phi_i|0\rangle$ around all possible closed loops in the event space. The latter, of course, is just the holonomy group of the connection H. R may have additional symmetries beyond those defined by the physical transport of $\langle 0|\phi_i|0\rangle$ around closed loops so that we have shown that $H \subseteq R$. From the above theorems, H is a Lie subgroup of the gauge group G go that we have finally

$$H \subseteq R \subseteq G. \tag{1}$$

Thus the residual symmetry group R after spontaneous symmetry breaking must be at least as big as H. These results are independent of the initial spontaneous symmetry breaking mechanism. In a specific gauge theory, if H turns out to be sufficiently large, then R is strongly constrained. We will see an example of this below for $G = SU_3$.

We are led by the above results to calculate the holonomy group of the connection for fiber bundles with various interesting structure (gauge) groups such as SU_3 or SU_5. The following theorem due to Ambrose and Singer (cf. Kobayashi and Nomizu[11]) will be useful.

Theorem: Let $P(B,G,\pi)$ be a principal fiber bundle, where B is connected and paracompact. Let Γ be a connection in P, Ω the curvature form, $H(u)$ the holonomy group with reference point $u \in P$, and $P(u)$ the holonomy bundle through u of Γ. Then the Lie algebra of $H(u)$ is equal to the subspace of $\mathfrak{g}$, Lie algebra of G, spanned by all elements of the form $\Omega_V(X,Y)$, where $V \in P(u)$ and X and Y are arbitrary horizontal vectors at V.

Using this theorem, we can calculate the Lie algebra of the holonomy group H and hence the holonomy group itself. Instead of doing this, we shall do a much simpler calculation and find the dimensionality of the Lie algebra of H. Finding dim H should suffice in most problems of physical interest since H is a Lie subgroup of G, so that knowing dim H is almost equivalent to finding H itself.

Another way of stating the above theorem[18,20] is that the Lie algebra of the holonomy group at u is generated by $F_{\mu\nu}(u)$ and all its covariant derivatives $D_\rho F_{\mu\nu}(u)$, $D_\rho D_\alpha F_{\mu\nu}(u), \ldots$. $F_{\mu\nu}$ is the coefficient of the curvature 2-form (the gauge fields). Now we can write

$$F_{\mu\nu} = \partial_\mu \Gamma_\nu - \partial_\nu \Gamma_\mu - [\Gamma_\mu, \Gamma_\nu], \tag{2}$$

where $F_{\mu\nu} \equiv F_{\mu\nu}{}^i l_i$; and $\Gamma_\mu \equiv \Gamma_\mu{}^i l_i$. These basis vectors l_i in the fiber bundle obey

$$[l_i, l_j] = f_{ij}{}^k l_k, \tag{3}$$

where f_{ij} are the structure constants of the structure group G. $i,j,k = 1,\ldots,n$, where n is the dimensionality of the structure group G. The covariant derivative of (2) is

$$D_\alpha F_{\mu\nu} = \partial_\alpha F_{\mu\nu} - [\Gamma_\alpha, F_{\mu\nu}]. \quad (4)$$

The second covariant derivative can be written as

$$D_\beta D_\alpha F_{\mu\nu} = \partial_\beta \partial_\alpha F_{\mu\nu} - [\partial_\beta \Gamma_\alpha, F_{\mu\nu}] - [\Gamma_\alpha, \partial_\beta F_{\mu\nu}] - [\Gamma_\beta, \partial_\alpha F_{\mu\nu}] + [\Gamma_\beta, [\Gamma_\alpha, F_{\mu\nu}]]. \quad (5)$$

Now in order to calculate dim H, we need to find the number of independent vectors generated by $F_{\mu\nu}$, $D_\alpha F_{\mu\nu}$, $D_\beta D_\alpha F_{\mu\nu}$,... in the n-dimensional space defined by the basis vectors l_i of G. We shall refer to this as "i space." Clearly dim $H \leqslant n$ since at most n independent vectors can exist in an n-dimensional space. Now taken as vectors in i space, (2), (4), and (5) determine different independent directions given by

$$\begin{aligned} &F_{\mu\nu} \equiv F_{\mu\nu}{}^i l_i, \\ &[\Gamma_\mu, F_{\kappa\lambda}] \equiv F_{\kappa\mu}{}^j \Gamma_\mu{}^i f_{ij}{}^k l_k, \\ &[\Gamma_\nu, [\Gamma_\mu, F_{\kappa\lambda}]] \equiv F_{\kappa\lambda}{}^j \Gamma_\mu{}^i f_{ij}{}^k \Gamma_\nu{}^1 f_{1k}{}^q l_q, \end{aligned} \quad (6)$$

since partial derivatives ∂_α do not rotate vectors in i space. Thus dim H is simply the number of independent vectors in the sequence in (6). This clearly depends strongly on the properties of the structure constants $f_{ij}{}^k$ of the gauge group G. Now it is important to note that for our purposes $F_{\mu\nu}{}^i$ represents a single vector in i space rather than a large number as μ,ν range over 1,...,4. The reason is clear from the work of Loos[18]: If we carry a field ψ, which transforms as some representation of the gauge group G, completely around an infinitesimal parallelogram, we get an infinitesimal gauge transformation

$$\begin{aligned} \Psi' &= [1 + \tfrac{1}{2} F_{\mu\nu}(dx^\mu d'x^\nu - dx^\nu d'x^\mu)]\Psi \\ &\equiv [1 + \tfrac{1}{2} F_{\mu\nu} df^{\mu\nu}]\Psi, \end{aligned} \quad (7)$$

where we suppressed the i indices. Thus $F_{\mu\nu}{}^i \, df^{\mu\nu}$ can be taken to be an infinitesimal generator of the holonomy group rather than $F_{\mu\nu}{}^i$ itself, i.e., $F_{\mu\nu}{}^i$ always appears contracted with $df^{\mu\nu}$. Appending handles on the infinitesimal parallelogram brings in the covariant derivative terms and similar remarks hold for $\Gamma_\mu{}^i$. Thus we can rewrite the sequence (6) as

$$F^i, F^j \Gamma^i f_{ij}^k, F^j \Gamma^i f_{ij}^k \Gamma^1 f_{1k}^q, \ldots, \quad (8)$$

where $F^i \equiv F^i{}_{\mu\nu} \, df^{\mu\nu}$ and $\Gamma^i = dx^\mu \Gamma_\mu{}^i$. Our job then reduces to finding the number of independent vectors in the sequence (8). This will be dim H. Note that $f_{ij}{}^k$ can be used to define a generalized cross product and (8) can also be written as

$$\vec{F}, \vec{F} \times \vec{\Gamma}, (\vec{F} \times \vec{\Gamma}) \times \vec{\Gamma}, \ldots . \quad (9)$$

The number of independent vectors in the sequence (8) can be found by considering the sequence

$$\Gamma^i f_{ij}^k \equiv \Gamma f, \quad \Gamma^i f_{ij}^k \Gamma^1 f_{1k}^q \equiv (\Gamma f)^2, \ldots \quad (10)$$

or in general $(\Gamma f)^m$ for progressively larger m. The first m for which $(\Gamma f)^m$ can be written as a linear combination of preceding members in the series defines the dimensionality of the holonomy group, where dim $H = m$. Thus the equations which must be checked for progressively larger m are

$$\begin{aligned} m &= 2 \quad (\Gamma f)^2 = A\Gamma^2 \delta, \\ m &= 3 \quad (\Gamma f)^3 = B(\Gamma f)^1 \Gamma^2, \\ m &= 4 \quad (\Gamma f)^4 = C\Gamma^2 (\Gamma f)^2 + D\Gamma^4 \delta, \\ m &= 5 \quad (\Gamma f)^5 = E\Gamma^2 (\Gamma f)^3 + F\Gamma^4 (\Gamma f)^1, \end{aligned} \quad (11)$$

where $\Gamma^2 \equiv \Gamma^i \Gamma_i$, $\delta \equiv \delta^i_j$, and $A,B,\ldots$ are constants. The above defines an algorithm from which dim H can be calculated for any gauge group G by plugging in the appropriate structure constants $f_{ij}{}^k$. We will now apply this to $G = SU_2$ and SU_3 in the following section.

III. dim H FOR SU_2 AND SU_3 FIBER BUNDLES

SU_2: For $G \equiv SU_2$ we have the structure constants

$$f_{ijk} = \epsilon_{ijk}, \quad (12)$$

which satisfy the identity

$$f_{ijk} f_{klq} \equiv \delta_{il}\delta_{jq} - \delta_{iq}\delta_{jl}. \quad (13)$$

Let us first look at the $m = 2$ equation from (11). This is

$$\Gamma^i f_{ijk} \Gamma^1 f_{lkq} = A\Gamma^2 \delta_{jq}. \quad (14)$$

The left hand side is

$$-\Gamma^i \Gamma^1 (\delta_{il}\delta_{jq} - \delta_{iq}\delta_{jl}) \equiv -\Gamma^2 \delta_{jq} + \Gamma_q \Gamma_j, \quad (15)$$

so that (14) is *not* satisfied. For $n = 3$, using (12) and (13), we can show

$$(\Gamma f)^3 \equiv -\Gamma^2 \Gamma^m f_{imp}. \quad (16)$$

Thus the $(\Gamma f)^3$ term is a multiple of the $(\Gamma f)^1$ term and the $m = 3$ equation of (11) *is satisfied* for SU_2. Thus dim $H = 3$ for $G = SU_2$. Since SU_2 itself has three generators, this is a reasonable result.

SU_3: The fully antisymmetric structure constants of SU_3 can be taken to be[21]

$$\begin{aligned} &f_{123} = 1, \quad f_{147} = f_{246} = f_{257} = f_{345} = \tfrac{1}{2}, \\ &f_{156} = f_{367} = -\tfrac{1}{2}, \quad f_{458} = f_{678} = (\sqrt{3}/2). \end{aligned} \quad (17)$$

The overall normalization of these is irrelevant since arbitrary constants appear in (11) in any case. We now have eight generators and the calculation is somewhat more difficult. We find that the $m = 2$–7 equations of (11) are *not* satisfied for the structure constants of SU_3. The $m = 8$ equation is satisfied as we might expect. Thus we have eight independent vectors in the sequence (8) and dim $H = 8$ for SU_3. [To show that dim $H = 8$ for SU_3, it is actually sufficient to show that the $m = 4$ equation in the sequence (8) *cannot* be satisfied. This implies that dim $H \geqslant 5$. Since H must be a connected, normal Lie subgroup of SU_3, and since the largest subgroup of SU_3, other than SU_3 itself, is $SU_2 \times U_1$ with dimension 4, H can only be SU_3 itself with dimension 8.] Note that for large groups and for large m in (11), the number of arbitrary constants $A,B,\ldots$ which must be found becomes large and the problem becomes much more tedious.

IV. DISCUSSION AND CONCLUSIONS

We found in the preceding section that dim $H = 3$ for an SU_2 gauge theory and dim $H = 8$ for an SU_3 gauge theory. The holonomy group H must be a connected, normal Lie subgroup of the gauge group G, and we have shown that the residual symmetry group must be H or contain H as a subgroup. The holonomy groups are then SU_2 in the SU_2 case and SU_3 in the SU_3 case, and these are also the residual symmetry groups. These are very puzzling results and suggest that spontaneous symmetry breaking does not take place in a pure SU_2 or SU_3 gauge theory. Symmetry breaking in an SU_3 gauge theory such as the eightfold way of Gell-Mann and Ne'eman[22] must take place some other way, perhaps dynamically.

We have given a well defined algorithm for uniquely finding the dimension of the holonomy group, given any gauge group G and its structure constants. dim H then provides valuable information about possible residual symmetry groups after spontaneous symmetry breaking in the gauge theory. This algorithm is such that it can readily be put on a computer. For gauge groups as large as SU_5 with 24 generators, this will be necessary. SU_5 and other large gauge groups will be the subject of a subsequent paper. Preliminary work suggests that the holonomy group of SU_5 is *not* SU_5 itself.

[1] H. Weyl, *Gruppentheorie and Quantenmechanik* (Hirzel, Leipzig, 1931).
[2] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).
[3] A. Salam, in *Elementary Particle Physics* (Nobel Sym. No. 8), edited by N. Svartholm (Almquist and Wiksell, Stockholm, 1968).
[4] For a good review of QCD see W. Marciano and H. Pagels, Phys. Rep. **36**, 139 (1978).
[5] R. Utiyama, Phys. Rev. **101**, 1597 (1956).
[6] T. W. B. Kibble, J. Math. Phys. **2**, 212 (1961); D. W. Sciama, *Recent Developments in General Relativity* (Pergamon, New York, 1962).
[7] Y. M. Cho, Phys. Rev. D **14**, 2521, 3335 (1976).
[8] A. Trautman, Rep. Math. Phys. **1**, 29 (1970).
[9] Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds, and Physics* (North-Holland, Amsterdam, 1977).
[10] M. E. Mayer, *Introducion to the Fiber-Bundle Approach to Gauge Theories, Lecture Notes in Physics* (Springer, Berlin, 1977).
[11] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.
[12] T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975).
[13] P. W. Higgs, Phys. Rev. **145**, 1156 (1966).
[14] H. Georgi and S. L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).
[15] D. Z. Freedman and P. van Nieuwenhuizen, Phys. Rev. D **13**, 3214 (1976); S. Deser and B. Zumino, Phys. Lett. B **62**, 335 (1976); D. Z. Freedman and B. De Wit, Nucl. Phys. B **130**, 105 (1977); E. Cremmer, J. Scherk, and J. Schwarz,, Phys. Lett. B **84**, 83 (1979).
[16] M. E. Mayer, "Gauge Fields as Quantized Connection Forms," *Proceedings of the Symposium on Differential-Geometrical Methods in Physics, Bonn 1975*, edited by K. Bleuler and A. Reetz, Lecture Notes in Mathematics (Springer, Berlin, 1977).
[17] N. S. Manton, Nucl. Phys. B **158**, 141 (1979); M. E. Mayer, Hadronic J. **4**, 108 (1981); A. S. Schwarz and Yu. S. Tyupkin, Nucl. Phys. B **187**, 321 (1981); M. E. Mayer and A. Trautman, Acta Phys. Aust. Suppl. **23**, 433 (1981).
[18] H. G. Loos, J. Math. Phys. **8**, 2114 (1967).
[19] A. Lichnerowicz, *Theorie Globale des Connexions et des Groupes d'Holonomie*, edited by Cremonese (Rome, 1955); G. Vrânceanu, *Leçons de Géometrie Différentielle*, edited by the Academy of the Peoples Republic of Romania (Gauthier-Villars, Paris, 1964).
[20] M. Daniel and C. M. Viallet, Rev. Mod. Phys. **52**, 175 (1980).
[21] J. Bernstein, *Elementary Particles and Their Currents* (Freeman, San Francisco, 1968), p. 213.
[22] M. Gell-Mann and Y. Ne'eman, *The Eightfold Way* (Benjamin, New York, 1964).

On the topological meaning of magnetic charge quantization

M. Quirós and J. Ramírez Mittelbrunn
Instituto de Estructura de la Materia, Serrano, 119, Madrid-6, Spain
Departmenta de Física, Universidad de Alcalá de Henares, Madrid, Spain

E. Rodríguez
Departmenta de Matemáticas, Universidad de Alcalá de Henares, Madrid, Spain

(Received 10 November 1981; accepted for publication 8 December 1981)

In this paper, all inequivalent principal fiber bundles with base space S^2 and fiber S^1, corresponding to Milnor classification, are constructed. Electromagnetism with magnetic monopoles is geometrically described in terms of these fiber bundles. In each of them, the coincidence of winding number and Chern number is established. The Dirac quantization condition is shown to have a topological origin. In fact, magnetic monopoles with charge $g = (\hbar c/2e)n$ correspond, in a topological sense, to the fiber bundle with winding number equal to n.

PACS numbers: 14.80.Hv, 02.40. + m

1. INTRODUCTION

Electromagnetism with magnetic monopoles was introduced by Dirac in 1931.[1] The most important consequences of the presence of magnetic monopoles, already pointed out in Dirac's paper, are the following (a) The electromagnetic vector potential **A** becomes singular along a string coming from infinity to the point where the monopole is fixed. (b) The wave function of an electron moving in the field of a magnetic monopole should vanish on the string. (c) From (b), the quantization condition $2eg/\hbar c = n \in Z$, which relates electric charge e with magnetic charge g, follows.

More recently, Wu and Yang,[2] were able to reformulate Dirac's theory, avoiding the singularity in the potential. The procedure employed by these authors was to split up the region surrounding the monopole in two overlapping regions, with a different vector potential defined in each one. The two vector potentials are then related by a U(1) gauge transfromation in the overlapping region. Wu and Yang observe in their paper that the suitable mathematical structure to describe gauge theories is that of principal fiber bundle. More precisely, gauge fields are to be identified with the components of a connection 1-form in some trivialization, a strength field tensor with the curvature 2-form, and gauge transformations with a fiber bundle homomorphism.[3] So, electromagnetism is geometrically described by a principal fiber bundle structure with base space S^2 and structure group U(1). Electromagnetism without monopoles is then described by the trivial bundle $S^2 \times S^1$, while electromagnetism with monopoles correspond to nontrivial principal fiber bundles.

On the other hand, a theorem of Milnor[4] provides the classification of all inequivalent fiber bundles with base space S^n and structure group G: "Let G be an arcwise connected group. Then the set of all classes of equivalent fiber bundles with group G and base S^n, is in one-to-one correspondence with $\Pi_{n-1}(G)$". In the case of electromagnetism this theorem leads to $\Pi_1(S^1) = Z$ inequivalent fiber bundles. For a non-abelian gauge theory, corresponding to fiber bundles with base S^4 and group SU(N), the Milnor theorem leads also to $\Pi_3(\mathrm{SU}(N)) = Z$ inequivalent fiber bundles.

Another way to discuss the classification of fiber bundles is given by the cohomology classes. Starting with connections in a principal fiber bundle $P(M, G)$, it is possible to derive a set of closed forms on M, of even degree, by means of the Weyl homomorphism.[5] It happens that cohomology classes determined by this set of closed forms do not depend on the particular choice of connection in $P(M, G)$. It is, therefore, possible to associate elements of $H^\star(M, R)$ with P. Furthermore, these elements verify the axioms for Chern classes: they belong to $H^\star(M, Z)$ and the fiber bundles $P(M, G)$ over M are classified by them.

In the case of bundles with base S^4 and group SU(N), only the second Chern class does not vanish and its integration over S^4 yields the so-called instanton number or second Chern number. This has been shown to coincide[6] with the "winding number" or element of $\Pi_3(\mathrm{SU}(N))$ in Milnor's classification. For electromagnetism, only the first Chern class is different from zero, and its integral over S^2 produces an integer number.

Guided by the results about instantons, it is possible at this point to formulate some questions concerning the mathematical structure of the Dirac monopole:

(i) Does Chern number (which might be called monopole number) coincide with the winding number, or element of $\Pi_1(S^1)$ in Milnor's classification?

(ii) Is there any relation between the integer $n = 2eg/\hbar c$ appearing in Dirac's quantization condition, the Chern number, and the winding number? In other words: Has magnetic charge quantization a topological meaning?

The aim of this work is to answer the preceding questions. A first step in this direction was recently given by Ryder[7] and Minami.[8] Both authors interpret Wu–Yang potentials as components of a connection 1-form in the Hopf bundle S^3, with base S^2, fiber S^1, and projection Π given by the Hopf map with invariant equal to 1. However, a misinterpretation of the Gauss–Bonnet–Chern theorem leads Ryder to Schwinger quantization condition $n = 2$. Let us illuminate this point. Take E_1 to be the canonical linear bundle over $P_1(C) = S^2$, with fiber C and structure group $C\star = C - \{0\}$, associated with the principal fiber bundle $C^2 - \{0\}(P_1(C), C\star)$ which, by reduction on the group, re-

 0022-2488/82/050873-05$02.50

sults in the Hopf bundle $S^3(S^2, \mathrm{U}(1))$. Let $C_1(E_1)$ be the first Chern class. Acording to the IV Chern axiom $C_1(E_1)$ is a generator of $H^2(P_1(C), Z)$, satisfying $\int_{S^2} C_1(E_1) = \pm 1$, where the sign $\pm$ depends on the orientation. This would lead to a quantization condition $n = \pm 1$, in contrast with the result of Ryder.

On the other hand, $\mathrm{SO}(3) = S^3/Z_2$ is another principal fiber bundle with base S^2 and fiber S^1. Since the "tangent sphere bundle of S^n is associated to the principal fiber bundle $\mathrm{SO}(n+1)$ over S^n with group $\mathrm{SO}(n)$",[9] in particular, the fiber bundle $\mathrm{SO}(3)(S^2, S^1)$ is associated with the tangent sphere bundle of S^2. The Gauss–Bonnet–Chern theorem says[5]: "If M is an oriented compact Riemannian manifold of dimension $2p$ and if E is the tangent bundle of M, then the closed $2p$-form γ integrated over M gives the Euler number of M", where γ is the pth Chern class of the bundle E. Therefore, since the Euler number of S^2 is 2, using the preceding theorems, it follows that $\int_{S^2} C_1(E) = 2$ for the fiber bundle SO(3).

The content of this work is as follows. In Sec. 2 the Z inequivalent fiber bundles with base S^2 and fiber S^1 of Milnor classification are constructed explicitly. We observe that the bundles corresponding to the integers m, and $-m$, are inequivalent. However, if instead of the group $\mathrm{U}(1) = \mathrm{SO}(2)$, we consider $\mathrm{O}(2) = \mathrm{SO}(2) \times Z_2$ as the structure group of the bundle, with $Z_2 = (I, C)$, we obtain that the bundles corresponding to m and $-m$ are equivalent. Matrix $C = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ is to be physically identified with magnetic charge conjugation.

As a particular case, we obtain for $n = 1$ the Hopf bundle S^3 used in the construction of Ryder and Minami, and for $n = 2$ the bundle SO(3) associated with the tangent bundle to the sphere S^2.

In Sec. 3 we prove that the winding number in Milnor's classification coincides with the Chern number and with the integer appearing in the Dirac quantization condition. For this purpose, we write down explicitly connections in each of the fiber bundles, and calculate its Chern classes. We note through the previously mentioned identifications that all connections correspond to the Wu–Yang potentials.

2. CONSTRUCTION OF THE FIBER BUNDLES (S^2, S^1)

As we said in Sec. 1, the fiber bundles with base S^2 and fiber S^1 are classified according to the first homotopy group of S^1, $\Pi_1(S^1) = Z$. Thus, we are going to construct Z inequivalent principal coordinate fiber bundles with base S^2 and fiber S^1: $P_n(S^2, S^1, \{U_\alpha\}, \{\psi^{(n)}_{\alpha\beta}\})$. Choosing as open covering for S^2 $\{U_1, U_2\}$, given by

$$U_1 = \{(\theta, \Phi),\quad 0 \leqslant \Phi < 2\pi,\quad 0 < \theta \leqslant \pi\}, \tag{1a}$$

$$U_2 = \{(\theta, \Phi),\quad 0 \leqslant \Phi < 2\pi,\quad 0 \leqslant \theta < \pi\}, \tag{1b}$$

where (θ, Φ) are the polar coordinates on the sphere S^2. The transition functions $\psi^{(n)}_{21}$: $U_1 \cap U_2 \rightarrow S_1$ are

$$\psi^{(n)}_{21}(\theta, \phi) = n\Phi, \tag{2}$$

where Φ is also the polar angle on S^1. Let us remark that the functions $\psi^{(n)}_{21}$ are homotopically inequivalent and thus the fiber bundles P_n constructed from them are also inequivalent. As we shall see in Sec. 3, the choice of transition function $\psi^{(n)}_{21}$ is essential for the geometrical interpretation of the Dirac quantization condition.

For $n = 0$, the transition functions are constantly equal to the identity in S^1 and the resulting fiber bundle is then the product bundle $S^2 \times S^1$, which describes electromagnetism without monopoles.

A. The case $n = 1$

Let us see that for $n = 1$ we obtain the Hopf bundle corresponding to the Hopf projection $S^3 \rightarrow S^2$.[9] We sum up concisely the structure of this bundle, which will be frequently used in this work. The elements of S^3 can be identified with pairs of complex numbers (Z_0, Z_1) satisfying $|Z_0|^2 + |Z_1|^2 = 1$, in such a way that there exists a one-to-one correspondence between elements of S^3 and 2×2 complex matrices $\mathrm{U} \in \mathrm{SU}(2)$. We shall parametrize these matrices by Euler angles:

$$U(\Phi, \theta, \psi) = \begin{pmatrix} Z_0 & iZ_1 \\ iZ_1^* & Z_0^* \end{pmatrix} = e^{i(\Phi/2)\sigma_3} e^{i(\theta/2)\sigma_2} e^{i(\psi/2)\sigma_3}, \tag{3}$$

where

$$0 \leqslant \Phi < 2\pi,\quad 0 \leqslant \theta < \pi,\quad 0 \leqslant \psi < 4\pi \tag{4}$$

and σ_i, $i = 1, 2, 3$, are Pauli matrices.

The Hopf projection is given in these coordinates by

$$\Pi_1(\Phi, \theta, \psi) = (\Phi, \theta). \tag{5}$$

We shall represent the elements of $\mathrm{U}(1) = S^1$, as a subgroup of SU(2), by $e^{i\sigma_3\alpha}$, $0 \leqslant \alpha < 2\pi$, α being the polar coordinate on S^1.

The right action R_α of U(1) over the bundle space is then defined as

$$\begin{aligned} R_\alpha \mathrm{U}(\Phi, \theta, \psi) &= e^{i(\Phi/2)\sigma_3} e^{i(\theta/2)\sigma_2} e^{i(\psi/2)\sigma_3} e^{i\alpha\sigma_3} \\ &\equiv \mathrm{U}(\Phi, \theta, \psi + 2\alpha); \end{aligned} \tag{6}$$

we choose the local sections

$$\sigma_1(\Phi, \theta) = \mathrm{U}(\Phi, \theta, \Phi), \tag{7a}$$

$$\sigma_2(\Phi, \theta) = \mathrm{U}(\Phi, \theta, -\Phi). \tag{7b}$$

Consequently, local trivializations ϕ_i: $\Pi_1^{-1}(U_i) \rightarrow \mathrm{U}(1)$, $i = 1, 2$, are given by

$$\phi_1(\Phi, \theta, \psi) = (\psi - \Phi)/2, \tag{8a}$$

$$\phi_2(\Phi, \theta, \psi) = (\psi + \Phi)/2 \tag{8b}$$

and the corresponding transition function ψ_{21}: $U_1 \cap U_2 \rightarrow \mathrm{U}(1)$ is

$$\psi_{21}(\Phi, \theta) = \Phi \tag{9}$$

which corresponds to the mapping $S^1 \rightarrow S^1$ with winding number $n = 1$. Q.E.D.

B. The case $n = 2$

Next we shall discuss the bundle $\mathrm{SO}(3)(S^2, S^1)$, associated with the tangent bundle to the sphere S^2, and we shall show that it corresponds to $n = 2$ in Milnor's classification.

We represent the points of S^2 in Cartesian coordinates $S^2 = \{x = (t_0, t_1, t_2) | \Sigma_i t_i^2 = 1\}$ and the elements of SO(3) by real 3×3 matrices. We define the projection p: $\mathrm{SO}(3) \rightarrow S^2$ by $P(R) = Rx_2$ with $R \in \mathrm{SO}(3)$, $x_2 = (0, 0, 1)$, and the local sections σ_i: $U_i \rightarrow SO(3)$ by

$$\sigma_2(x) = \begin{pmatrix} \delta_{\alpha\beta} - \dfrac{t_\alpha t_\beta}{1+t_2} & t_0 \\ & t_1 \\ -t_0 \quad -t_1 & t_2 \end{pmatrix} \quad (\alpha, \beta = 0, 1), \tag{10a}$$

$$\sigma_i(x) = \lambda\sigma_2(\lambda x); \quad \text{with } \lambda = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \tag{10b}$$

where matrix λ is a rotation of angle π around axis 1. Transition functions can be obtained as usual from local sections: ψ_{21}: $U_2 \cap U_1 \rightarrow SO(2)$

$$\psi_{21}(x) = \sigma_2^{-1}(x)\sigma_1(x) = \begin{pmatrix} \dfrac{-t_1^2+t_0^2}{t_1^2+t_0^2} & \dfrac{2t_0t_1}{t_1^2+t_0^2} & 0 \\ \dfrac{-2t_0t_1}{t_1^2+t_0^2} & \dfrac{-t_1^2+t_0^2}{t_1^2+t_0^2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{11}$$

which belongs to $SO(2) \subset SO(3)$.

Using polar coordinates on S^2, $x = (t_0, t_1, t_2) \equiv (\Phi, \theta)$ with

$$t_0 = \sin\theta\cos\Phi,$$
$$t_1 = \sin\theta\sin\Phi, \tag{12}$$
$$t_2 = \cos\theta.$$

Equation (11) is written down as

$$\psi_{21}(x) = \begin{pmatrix} \cos 2\Phi & \sin 2\Phi & 0 \\ -\sin 2\Phi & \cos 2\Phi & 0 \\ 0 & 0 & 1 \end{pmatrix} = e^{2i\Phi J_3}. \tag{13}$$

Therefore, transition functions are given by

$$\psi_{21}(\Phi, \theta) = 2\Phi, \tag{14}$$

which correspond to a mapping $S^1 \rightarrow S^1$ with "winding number" 2.

Let us now see how the fiber bundle (S_2, S_1) with $n = 2$ can be obtained from the fiber bundle (S_2, S_1) with $n = 1$. Taking Euler parametrization for the elements $R \in SO(3)$,

$$R(\Phi, \theta, \psi) = e^{i\Phi J_3} e^{i\theta J_2} e^{i\psi J_3}, \tag{15}$$

with $0 \leqslant \Phi < 2\pi$, $0 \leqslant \theta < \pi$, $0 \leqslant \psi < 2\pi$ and $\mathbf{J}$ being the infinitesimal generators of SO(3), the projection p can be expressed as

$$p(R) = R\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \sin\theta\cos\Phi \\ \sin\theta\sin\Phi \\ \cos\theta \end{pmatrix} \in S^2. \tag{16}$$

Representing the elements of $S^1 \equiv U(1) \subset SO(3)$ by $R_\alpha = e^{i\alpha J_3}$, the right action of S^1 on the bundle is given by the right matrix multiplication

$$R_\alpha R(\Phi, \theta, \psi) = R(\Phi, \theta, \psi + \alpha). \tag{17}$$

Finally, local sections are given in this parametrization by

$$\sigma_2(\Phi, \theta) = R(\Phi, \theta, -\Phi), \tag{18a}$$

$$\sigma_1(\Phi, \theta) = R(\Phi, \theta, \Phi). \tag{18b}$$

On the other hand, as is well known, the group SO(3) is isomorphic to the factor group $SU(2)/Z_2$. Using Euler parametrization for both groups, the isomorphism is expressed as

$$R(\Phi, \theta, \psi) \leftrightarrow \{U(\Phi, \theta, \psi), U(\Phi, \theta, \psi + 2\pi)\} \equiv U^{(2)}(\Phi, \theta, \psi), \tag{19}$$

where we have introduced the notation $U^{(2)}(\Phi, \theta, \psi)$ in the class belonging to $SU(2)/Z_2$ associated with the element $U(\Phi, \theta, \psi)$.

With the identification provided by this isomorphism, it is possible to construct the fiber bundle (S^2, S^1) with $n = 2$ from the fiber bundle (S_2, S_1) with $n = 1$, by means of the Hopf projection Π_1 and the canonical projection p_2 following the commutative diagram.

$$\begin{tikzcd} SU(2) \arrow[r, "\Pi_1"] \arrow[d, "p_2"'] & S^2 \\ SU(2)/Z_2 \arrow[ur, "\Pi_2"'] & \end{tikzcd} \tag{20}$$

This is possible due to the compatibility of Π and p_2: Two points in SU(2) which belong to the same class in $SU(2)/Z_2$ have the same Hopf projection on S^2.

In a similar way, local sections (18) in $SU(2)/Z_2$, which will be denoted by $\sigma_i^{(2)}$, are obtained from the local sections of SU(2) by the commutative diagram

$$\begin{tikzcd} SU(2) \arrow[d, "p_2"'] & U_i \arrow[l, "\sigma_i"'] \arrow[dl, "\sigma_i^{(2)}"] \\ SU(2)/Z_2 & \end{tikzcd} \tag{21}$$

C. The general case

Generalizing the former ideas we shall construct the Z fiber bundles with base S^2 and fiber S^1 corresponding to Milnor's classification.

(i) Positive n

Let Z_n be the cyclic subgroup of SU(2) given by

$$Z_n = \{e^{(i2\pi m/n)\sigma_3}, \quad m = 0, ..., n-1\} \tag{22}$$

and $SU(2)/Z_n$ the right-cosets space of SU(2).

Let us remark that, since Z_n is not an invariant subgroup of SU(2), right and left cosets do not coincide. Our choice of right cosets is appropriate with respect to the Hopf projection defined in Ref. 5. We shall denote the elements of $B_n = SU(2)/Z_n$ by

$$U^{(n)}(\Phi, \theta, \psi) = \{U(\Phi, \theta, \psi + 4\pi m/n), m = 0, ..., n-1\}. \tag{23}$$

Since all elements of the class $U^{(n)}(\Phi, \theta, \psi)$ have the same Hopf projection, we can define the projection Π_n: $B_n \rightarrow S^2$ by the following commutative diagram:

$$\begin{tikzcd} SU(2) \arrow[r, "\Pi_1"] \arrow[d, "p_n"'] & S^2 \\ SU(2)/Z_n, \arrow[ur, "\Pi_n"'] & \end{tikzcd} \tag{24}$$

where p_n is the canonical projection p_n: $SU(2) \rightarrow SU(2)/Z_n$.

Let us see that B_n is a fiber bundle with base S^2, fiber S^1, projection Π_n, and transition functions with winding number n. Let us first define the right action of the group $U(1) = S^1$ on the bundle B_n in the following way:

$$R_\alpha U^{(n)}(\Phi, \theta, \psi) = U^{(n)}(\Phi, \theta, \psi + 2\alpha/n). \tag{25}$$

This action verifies the property

$$p_n R_\alpha = R_{n\alpha} p_n, \tag{26}$$

which is equivalent to the commutativity of the diagram

$$\begin{tikzcd} \mathrm{SU(2)} \arrow[r, "R_\alpha"] \arrow[d, "p_n"'] & \mathrm{SU(2)} \arrow[d, "p_n"] \\ B_n \arrow[r, "R_{n\alpha}"] & B_n. \end{tikzcd} \tag{27}$$

The local sections in B_n are given by the commutative diagram

$$\begin{tikzcd} \mathrm{SU(2)} \arrow[dr, "p_n"'] & & U_i \arrow[ll, "\sigma_i"'] \arrow[dl, "\sigma_i^{(n)}"] \\ & B_n & \end{tikzcd} \tag{28}$$

Thence

$$\sigma_1^{(n)}(\Phi, \theta) = U^{(n)}(\Phi, \theta, \Phi),$$
$$\sigma_2^{(n)}(\Phi, \theta) = U^{(n)}(\Phi, \theta, -\Phi). \tag{29}$$

The trivializations $\phi_i^{(n)}: \Pi_n^{-1}(U_i) \to \mathrm{U}(1)$ can be derived from local sections (29) obtaining

$$\phi_1^{(n)}(U^{(n)}(\Phi, \theta, \psi)) = (n/2)(\psi - \Phi),$$
$$\phi_2^{(n)}(U^{(n)}(\Phi, \theta, \psi)) = (n/2)(\psi + \Phi). \tag{30}$$

Therefore, the transition function is

$$\psi_{21}^{(n)}(\Phi, \theta) = n\Phi, \tag{31}$$

which corresponds to a map $S^1 \to S^1$ with winding number n. Let us finally remark that this general construction includes as particular cases the bundles $B_1 = \mathrm{SU(2)}$ and $B_2 = \mathrm{SO(3)}$ previously discussed.

(ii) Negative n

Let us now show that it is possible to give a different fibration for SU(2) in such a way that the fiber bundle with $n = -1$ in the Milnor classification comes out. To show it, we take the same Hopf projection and change the right action[6] for

$$\bar{R}_\alpha U(\Phi, \theta, \psi) = U(\Phi, \theta, \psi - 2\alpha). \tag{32}$$

Then, using the same local sections as in Ref. 7, the following local trivializations $\bar{\phi}_i: \Pi_1^{-1}(U_i) \to \mathrm{U}(1)$ arise:

$$\bar{\phi}_1(\Phi, \theta, \psi) = -(\psi - \Phi)/2, \quad \bar{\phi}_2(\Phi, \theta, \psi) = -(\psi + \Phi)/2, \tag{33}$$

which result in the transition function $\bar{\phi}_{21}: U_2 \cap U_1 \to U_1$ given by

$$\bar{\psi}_{21}(\Phi, \theta) = -\Phi. \tag{34}$$

In a similar way we obtain the fiber bundle with winding number $-n$ $(n > 0)$, by means of a fibration of $\mathrm{SU(2)}/Z_n$ different from that used in section (i). We choose the same projection $\Pi_n: B_n \to S^2$ as in Sec. 2C, and define the right action of the group U(1) on B_n by

$$\bar{R}_\alpha U^{(n)}(\Phi, \theta, \psi) = U^{(n)}(\Phi, \theta, \psi - 2\alpha/n), \tag{35}$$

which satisfies

$$p_n \bar{R}_\alpha = \bar{R}_{n\alpha} p_n. \tag{36}$$

Using the last property and local sections (29), it is easy to find the trivializations

$$\bar{\phi}_1^{(n)}(U^{(n)}(\Phi, \theta, \psi)) = -(n/2)(\psi - \Phi),$$
$$\bar{\phi}_2^{(n)}(U^{(n)}(\Phi, \theta, \psi)) = -(n/2)(\psi + \Phi) \tag{37}$$

and the transition functions

$$\bar{\psi}_{21}^{(n)}(\Phi, \theta) = -n\Phi. \tag{38}$$

We shall denote the space $\mathrm{SU(2)}/Z_n$, fibered by the group action (32), by B_{-n}.

Let us finally note that the bundles B_n and B_{-n} are equivalent, provided that we consider O(2) instead of U(1) as a structure group. Actually, if we consider the continuous functions $\lambda_i: U_i \to \mathrm{O}(2)$ given by

$$\lambda_i(\Phi, \theta) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad i = 1, 2 \tag{39}$$

we obtain

$$\bar{\psi}_{21}(\Phi, \theta) = \lambda_2^{-1}(\Phi, \theta)\psi_{21}(\Phi, \theta)\lambda_1(\Phi, \theta) \tag{40}$$

which, by virtue of Lemma 2.10 of Ref. 9, proves the equivalence of B_n and B_{-n}. Nevertheless, as we have considered as the structure group $\mathrm{SO(2)} = \mathrm{U(1)}$ and the functions λ_i are not defined on it, the fiber bundles B_n and B_{-n} are not equivalent in our construction.

3. GEOMETRICAL INTERPRETATION OF THE DIRAC QUANTIZATION

In this section, we are going to interpret Wu–Yang potentials as local 1-forms of a connection in the fiber bundles B_n. We shall show that the integer n is the topological index corresponding to the first Chern class of these connections.

As is well known, a connection in a principal fiber bundle is given by a family of 1-forms ω_i defined on U_i, valued in the Lie algebra of the structure group of the bundle, and which satisfy the compatibility relations

$$\omega^i = \mathrm{ad}(\psi_{ji}^{-1})\omega^j + \psi_{ji}^{-1} d\psi_{ji}. \tag{41}$$

In our case, we shall choose the connection in B_n given by the family

$$\omega_{(n)}^1 = i(n/2)(1 + \cos\theta)\, d\Phi,$$
$$\omega_{(n)}^2 = -i(n/2)(1 - \cos\theta)\, d\Phi, \tag{42}$$

which satisfy condition (41), taking into account the transition functions given in (31) and (38). The choice of the factor in front of the connection 1-forms (42) is by no means arbitrary, but fixed by the following facts.

The Wu–Yang potentials $(A^i = A_\mu^i dx^\mu)$

$$A^1 = -g(1 + \cos\theta)\, d\Phi,$$
$$A^2 = g(1 - \cos\theta)\, d\Phi \tag{43}$$

are related by the gauge transformation

$$A^2 = A^1 - i\frac{\hbar c}{e} e^{-2ige/\hbar c} d\, e^{2ige/\hbar c} = 2g d\Phi \tag{44}$$

and with the local connection ω^i by the equation

$$\omega^i_{(n)} = (-ie/\hbar c)A^i. \tag{45}$$

Furthermore, relation (44) coincides with compatibility condition (41), in view of the Dirac quantization condition

$$2eg/\hbar c = n \in Z.$$

Let us now describe the topological interpretation of the Dirac quantization condition. For that purpose, we lift the local family $\{\omega^i_{(n)}\}$ in the usual way:

$$\omega = \mathrm{ad}(\phi_i^{-1})\Pi^*\omega_i + \phi_i^{-1}d\phi_i \tag{46}$$

which, in the particular case of the local forms given in Eq. (43), results in

$$\omega^{(n)} = i(n/2)(d\psi + \cos\theta\, d\Phi). \tag{47}$$

Thus, the local 1-forms associated with the Wu–Yang potentials are lifted by means of (45) and (46), producing the global 1-form

$$A^{(n)} = -g(d\psi + \cos\theta\, d\Phi). \tag{48}$$

Let us remark that no dependence on n appears in (48), because we have used the Dirac quantization condition and the 1-form (47) projects onto A^i by the local sections $\sigma^i_{(n)}$.

The curvature associated with the connection $\omega^{(n)}$ is

$$\Omega_{(n)} = d\omega_{(n)} = -i(n/2)\sin\theta\, d\theta\wedge d\Phi \tag{49}$$

and the first Chern class is

$$\Pi^*_n(\gamma^{(n)}_1) = \frac{-1}{2\pi i}\Omega_{(n)} = \frac{n}{4\pi}\sin\theta\, d\theta\wedge d\Phi, \tag{50}$$

so that the topological index associated with $\gamma^{(n)}_1$ is

$$\nu = \int_{S^2}\gamma^{(n)}_1 = n. \tag{51}$$

Finally, the magnetic field is given by the curvature 2-form associated with Wu–Yang potentials,

$$B_{(n)} = dA_{(n)} = g\sin\theta\, d\theta\wedge d\Phi, \tag{52}$$

which again does not depend on n, and whose flux is

$$\Phi_{(n)} = \int_{S^2} B_{(n)} = 4\pi g, \tag{53}$$

as it should be for a magnetic monopole.

4. CONCLUSION

The main result of this work is to prove that the Dirac quantization condition $2eg/\hbar c = n$ has a topological origin. To this end we explicitly build Z unequivalent fiber bundles over S^2 with fiber S^1, corresponding to the Milnor classification $\Pi_1(S^1) = Z$. From a connection in each of these fiber bundles the corresponding Chern classes are constructed and the equality between the Chern number and the winding number [element of $\Pi_1(S^1)$] is proved. Furthermore, it is proven that the former topological numbers coincide with the Dirac quantization number n. We finally conclude from the above construction that Z topologically unequivalent versions of the Dirac monopole do arise, although all of them give rise to the same monopole magnetic field and Wu–Yang potentials.

It should be interesting to apply a similar analysis to study fiber bundles with base space S^4 and fiber S^3. After Milnor construction there are $\Pi_3(S^3) = Z$ unequivalent fiber bundles whose first element is the Hopf fibration $S^7(S^4, S^3)$.[10] This analysis would be relevant for the description and classification of instantons and nonabelian [SU(2)] monopoles. Some work in this direction is in progress.

[1] P. A. M. Dirac, Proc. R. Soc. London, Ser. A **133**, 60 (1931).
[2] T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975).
[3] M. Daniel and C. M. Viallet, Rev. Mod. Phys. **52**, 175 (1980).
[4] D. Husemoller, *Fibre Bundles* (Springer, New York, 1966).
[5] S. Kobayshi and K. Nomizu, *Foundations of Differential Geometry* (Wiley-Interscience, New York, 1969), Vol. II.
[6] A. A. Belavin, A. M. Polyakov, A. S. Schwarz, and Yu. S. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[7] L. H. Ryder, J. Phys. A **13**, 437 (1980).
[8] M. Minami, Prog. Theor. Phys. **62**, 1128 (1979).
[9] N. Steenrod, *The Topology of Fibre Bundles* (Princeton U. P., Princeton, N. J., 1951).
[10] M. Minami, Prog. Theor. Phys. **63**, 303 (1980).

Dynamical group of microscopic collective states. I. One-dimensional case

J. Deenen and C. Quesne [a)]

Physique Théorique et Mathématique CP 229, Université Libre de Bruxelles, Boulevard du Triomphe, 1050 Brussels, Belgium

(Received 22 July 1981; accepted for publication 25 September 1981)

In the present series of papers it is intended to determine the nature and study various realizations of the dynamical group of microscopic collective states for an A-nucleon system, defined as those A-particle states invariant under the orthogonal group O(n) associated with the $n = A - 1$ Jacobi vectors. The present paper discusses the case of a hypothetical one-dimensional space. Simple invariance considerations show that the dynamical group of collective states is then the group $\mathscr{S}\!p_c(2, R)$, which is the restriction to the collective subspace of the group $\mathscr{S}\!p(2, R)$ of linear canonical transformations in n dimensions conserving the O(n) symmetry. In addition to the well-known realization of the dynamical group in the Schrödinger representation based upon the Dzublik–Zickendraht transformation, two new realizations are proposed. The first acts in a Barut Hilbert space, which is the subspace of a Bargmann Hilbert space of analytic functions left invariant by O(n). A unitary mapping is established between the ordinary Hilbert space of collective states and the Barut Hilbert space and coherent collective states are defined in the latter. The second is carried out in terms of one boson creation and one boson annihilation operator through a generalized Holstein–Primakoff representation. The generator of a U(1) group, which is the one-dimensional analog of the U(6) group of the interacting boson model (IBM), can then be expressed in terms of the generators of $\mathscr{S}\!p_c(2, R)$. Finally the generalization of the preceding analysis to a d-dimensional space is outlined in the cases where $d = 2$ or 3. The dynamical group of collective states becomes $\mathscr{S}\!p_c(2d, R)$.

PACS numbers: 21.60.Fw, 02.20. + b

1. INTRODUCTION

The construction of a microscopic collective Hamiltonian for an A-nucleon system, as well as the determination of its eigenstates and dynamical group is one of the major open problems in nuclear physics. In various recent works the microscopic Hamiltonian is projected out from the A-nucleon Hamiltonian by restricting the latter to a single irreducible representation (IR) of the orthogonal group O(n) associated with the $n = A - 1$ Jacobi vectors.[1,2] The simplest choice for the IR of O(n) corresponds to the scalar representation, as was suggested by Vanagas.[2] In that approach a basis for the microscopic collective states consists of all those A-nucleon states that are left invariant by the transformations of O(n).

In a recent publication, Chacón, Moshinsky, and Vanagas[3] construct the microscopic collective Hamiltonian for a system of three nucleons moving in two dimensions and interacting through harmonic-oscillator forces, as an introduction to the more difficult case of A particles moving in three dimensions and interacting through arbitrary forces. The generalization to an arbitrary number A of nucleons still moving in two dimensions and interacting through oscillator forces is carried out in another paper by Chacón and Moshinsky.[4] In both cases the eigenstates of the collective Hamiltonian are explicitly obtained and their dynamical group is proved to be SO(3, 2). The microscopic collective states are then put in one-to-one correspondence with macroscopic collective states classified according to IR's of a U(3) group, which is the two-dimensional analog of the U(6) group of the interacting boson model (IBM).[5] This step requires a canonical transformation, whose form in the classical picture is given in a paper by Moshinsky and Seligman[6] and whose unitary representation in quantum mechanics is contained in Ref. 3.

To explicitly project onto the collective subspace of the A-particle space, the authors quoted above make use of a coordinate transformation which was independently proposed by Dzublik *et al.*[7] and Zickendraht.[8] That transformation changes the $3A - 3$ Jacobi coordinates into a set of coordinates including $3A - 9$ noncollective ones and 6 collective ones, namely the three Euler angles that take us from the principal axis to the frame of reference fixed in space, and the three principal moments of inertia of the A-body system. Although it has a clear and appealing geometrical significance, the Dzublik–Zickendraht transformation is not easy to handle because the collective variables cannot be expressed in a simple way in terms of the Jacobi coordinates. Consequently the projection onto the collective subspace is difficult to work out. In Refs. 3 and 4 it was carried out step by step by assuming a simple Hamiltonian and first allowing the number of particles to be small, namely $A = 3$, and then generalizing the results to an aribtrary number of particles. This procedure tends to conceal some simple properties that would follow from general invariance considerations, independently of the number of particles.

It is one of the purposes of the present series of papers to show that general invariance principles allow one to construct a basis for the collective subspace and determine the dynamical group of collective states without preliminary use of the Dzublik–Zickendraht transformation. That program is most easily carried out in the formalism of the boson creation and annihilation operators associated with the Jacobi

[a)] Maître de recherches F.N.R.S.

 0022-2488/82/050878-12$02.50

coordinates. We are thus naturally led to look for the realization of the collective states and their dynamical group in a Bargmann Hilbert space[9] or, more exactly, in the subspace of the latter invariant under O(n). We shall refer to such a subspace as a Barut Hilbert space, in accordance with Kramer, Moshinsky, and Seligman,[10] who introduced that terminology to emphasize that such a space belongs to a class of spaces first considered by Barut and Girardello.[11] In the Barut Hilbert space, the collective variables assume a very simple form in terms of the variables of the A-particle system, so that the projection onto the collective subspace is easy to carry out.

Another purpose of the present series of papers is to investigate further the relation between the O(n) invariant microscopic collective model and the IBM that has already been studied by various authors.[2,3,6,12] In this connection, we want to show that it is interesting to consider a generalized Holstein–Primakoff representation[13] of the dynamical group of collective states.

In the present paper we restrict ourselves, for pedagogical convenience, to reviewing in detail the case of A particles in a hypothetical one-dimensional space and outlining the generalization of the analysis to a d-dimensional space. By taking d equal to 2 or 3, we can then recover the hypothetical two-dimensional space considered in Refs. 3 and 4 or the physical three-dimensional space, respectively. The details of the d-dimensional-space analysis will be given for $d > 1$ in subsequent publications.

In Sec. 2, we start to prove by means of invariance considerations that the dynamical group of A-particle collective states in one dimension is the group $\mathscr{Sp}_c(2, R)$, which is the restriction of the group $\mathscr{Sp}(2, R)$ of linear canonical transformations conserving the O(n) symmetry[14] to the collective subspace of the A-particle space. We then show in Sec. 3 that the realization of $\mathscr{Sp}_c(2, R)$ in the Hilbert space of collective states $\mathscr{H}_c$, which we obtain by using the Dzublik–Zickendraht transformation in one dimension, coincides with the dynamical group previously derived by Moshinsky and his collaborators.[15] In Sec. 4, we turn to the definition of collective variables in a Barut Hilbert space of collective states $\mathscr{F}_c$, and obtain a realization of $\mathscr{Sp}_c(2, R)$ in that space. In Sec. 5, we show that there exists a unitary mapping between the elements of the Hilbert spaces $\mathscr{H}_c$ and $\mathscr{F}_c$. Section 6 is devoted to the definition and study of coherent collective states in the sense of Barut and Girardello.[11] In Sec. 7 we then pass on to the generalized Holstein–Primakoff representation of $\mathscr{Sp}_c(2, R)$ and show that the generator of a U(1) group can be written in terms of the generators of $\mathscr{Sp}_c(2, R)$. Finally in Sec. 8, we prove that when shifting to d dimensions, the dynamical group of collective states becomes $\mathscr{Sp}_c(2d, R)$ and outline the generalization to that case of the analysis carried out in the preceding sections.

2. DYNAMICAL GROUP OF MICROSCOPIC COLLECTIVE STATES IN ONE DIMENSION

In this section, we consider a system of A particles in a one-dimensional space, construct a basis for its collective states, and obtain the dynamical group of the latter.

Let us denote by x_s, $s = 1,2,...,n$, the $n = A - 1$ Jacobi coordinates of the A-particle system, and by $p_s = -i\partial/\partial x_s$, $s = 1,2,...n$, the momenta canonically conjugate to x_s. As usual, we can introduce boson creation and annihilation operators associated with the Jacobi coordinates η_s and ξ_s, $s = 1,2,...,n$, defined by

$$\eta_s = (1/\sqrt{2})(x_s - ip_s), \quad \xi_s = (1/\sqrt{2})(x_s + ip_s) \tag{2.1}$$

and satisfying the following commutation relations:

$$[\eta_s, \eta_t] = [\xi_s, \xi_t] = 0, \quad [\xi_s, \eta_t] = \delta_{st}. \tag{2.2}$$

In Eq. (2.1), we use units in which $\hbar$, the mass m of the particles, and the frequency ω are equal to 1. In terms of the operators η_s, we can construct a basis for the translationally invariant A-particle states,

$$|\Psi_{\mathscr{N}_1 ... \mathscr{N}_n}\rangle = (\mathscr{N}_1! ... \mathscr{N}_n!)^{-1/2}\eta_1^{\mathscr{N}_1} \cdots \eta_n^{\mathscr{N}_n}|0\rangle, \quad \mathscr{N}_1, ... \mathscr{N}_n = 0, 1, ..., \tag{2.3}$$

where $|0\rangle$ is the boson vacuum state.

It is well known[14] that a dynamical group for the A-particle states (2.3) is the group of linear canonical transformations in n dimensions, the symplectic group $\mathbf{Sp}(2n, R)$, whose generators are the $n(2n + 1)$ bilinear operators

$$\mathbf{D}^\dagger_{st} = \eta_s\eta_t, \quad s \leqslant t = 1,...,n. \tag{2.4a}$$

$$\mathbf{D}_{st} = \xi_s\xi_t, \quad s \leqslant t = 1,...,n, \tag{2.4b}$$

$$\mathbf{E}_{st} = \tfrac{1}{2}(\eta_s\xi_t + \xi_t\eta_s) = C_{st} + \tfrac{1}{2}\delta_{st}, \quad s,t = 1,...,n, \tag{2.4c}$$

and satisfy the following commutation relations:

$$\begin{aligned}
&[\mathbf{E}_{st}, \mathbf{E}_{s't'}] = \delta_{ts'}\mathbf{E}_{st'} - \delta_{st'}\mathbf{E}_{s't},\\
&[\mathbf{E}_{st}, \mathbf{D}^\dagger_{s't'}] = \delta_{ts'}\mathbf{D}^\dagger_{st'} + \delta_{tt'}\mathbf{D}^\dagger_{ss'},\\
&[\mathbf{E}_{st}, \mathbf{D}_{s't'}] = -\delta_{ss'}\mathbf{D}_{tt'} - \delta_{st'}\mathbf{D}_{ts'},\\
&[\mathbf{D}^\dagger_{st}, \mathbf{D}^\dagger_{s't'}] = [\mathbf{D}_{st}, \mathbf{D}_{s't'}] = 0,\\
&[\mathbf{D}_{st}, \mathbf{D}^\dagger_{s't'}] = \delta_{ss'}\mathbf{E}_{t't} + \delta_{st'}\mathbf{E}_{s't} + \delta_{ts'}\mathbf{E}_{t's} + \delta_{tt'}\mathbf{E}_{s's}.
\end{aligned} \tag{2.5}$$

The n weight generators are the operators $\mathbf{E}_{ss}$, $s = 1,...,n$. In Eq. (2.4c), we introduced the operators

$$C_{st} = \eta_s\xi_t, \quad s, t = 1,...,n, \tag{2.6}$$

which are the generators of the U(n) subgroup of $\mathbf{Sp}(2n, R)$, and satisfy the same commutation relations as the $\mathbf{E}_{st}$ operators.

In the construction of collective A-particle states, we are interested in the (full) orthogonal subgroup O(n) of U(n), whose generators are the Hermitian operators

$$\Lambda_{st} = -i(C_{st} - C_{ts}), \quad s < t = 2,...,n, \tag{2.7}$$

and satisfy the following commutation relations:

$$[\Lambda_{st}, \Lambda_{s't'}] = i(\delta_{ss'}\Lambda_{tt'} + \delta_{tt'}\Lambda_{ss'} - \delta_{st'}\Lambda_{ts'} - \delta_{ts'}\Lambda_{st'}). \tag{2.8}$$

We have thus obtained the chain of groups

$$\mathbf{Sp}(2n, R) \supset \mathrm{U}(n), \tag{2.9a}$$

$$\mathrm{U}(n) \supset \mathrm{O}(n). \tag{2.9b}$$

The A-particle states (2.3) are classified according to the chain of groups (2.9a). They indeed belong to one of two IR's of $\mathbf{Sp}(2n, R)$, $\langle(\tfrac{1}{2})^n\rangle$ or $\langle(\tfrac{1}{2})^{n-1}\tfrac{3}{2}\rangle$ according to whether the total number of bosons $\mathscr{N} = \mathscr{N}_1 + \cdots + \mathscr{N}_n$ is even or odd.[14]

Those IR's of $\mathbf{Sp}(2n, R)$ are characterized by their minimum weights, corresponding to the state $|0\rangle$ or $\eta_n|0\rangle$, respectively. The states (2.3) also belong to a given IR of U(n), namely the totally symmetrical one characterized by $[\mathscr{N}]$. The remaining quantum numbers necessary to specify the states correspond to a classification according to the canonical chain of U(n), i.e., $\mathrm{U}(n)\supset\mathrm{U}(n-1)\supset\cdots\supset\mathrm{U}(1)$.

To get states classified according to the chain (2.9b), one would have to take appropriate linear combinations of the states (2.3). The IR's of O(n) contained in the IR $[\mathscr{N}]$ of U(n) can be entirely characterized by a single quantum number λ, which can take the values $\mathscr{N}, \mathscr{N}-2,\ldots,0(1)$ if $\mathscr{N}$ is even (odd).

In the present work, the A particle collective states are assumed to belong to the IR(0) of O(n). A basis for them can be easily constructed from the states (2.3). As the collective states must be invariant under O(n) and the vacuum state $|0\rangle$ obviously belongs to them, the remaining collective states are obtained from $|0\rangle$ by application of all the possible O(n) invariants that can be formed from the creation operators η_s, $s=1,\ldots,n$. It is well known that there is only one basic invariant (by an invariant we mean an absolute invariant of the theory of invariants) of O(n) that can be formed from the vector η_s, namely the scalar product $\Sigma_{s=1}^{n}\eta_s^2$.[16] Therefore a basis for the collective states is made of the states

$$|\phi_N\rangle \propto \left(\sum_{s=1}^{n}\eta_s^2\right)^N|0\rangle, \quad N=0,1,\ldots . \tag{2.10}$$

Incidentally, let us mention that we should have reached the same result if we had considered the collective states as those states invariant under the rotation group SO(n) instead of those invariant under the full orthogonal group O(n). It is indeed not possible to form a pseudoscalar from the vector η_s, $s=1,\ldots,n$, so that all the invariants under SO(n) are also invariants under O(n). In one dimension, one may therefore indifferently use O(n) or SO(n) to construct collective states.

To discuss the properties of the states $|\phi_N\rangle$, it is advantageous to consider a chain of subgroups of $\mathbf{Sp}(2n, R)$ other than (2.9), the chain

$$\mathbf{Sp}(2n, R)\supset \mathscr{S}\!\mathscr{p}(2, R)\times \mathrm{O}(n), \tag{2.11a}$$

$$\mathscr{S}\!\mathscr{p}(2, R)\supset \mathscr{U}(1). \tag{2.11b}$$

The $\mathscr{S}\!\mathscr{p}(2, R)$ group can be interpreted as the group of linear canonical transformations conserving the O(n) symmetry.[14] Its generators are obtained by contracting those of $\mathbf{Sp}(2n, R)$ with respect to the particle index s. They are given by

$$\begin{aligned}\mathscr{D}^\dagger &= \sum_{s=1}^{n}\eta_s^2,\\ \mathscr{D} &= \sum_{s=1}^{n}\xi_s^2,\\ \mathscr{E} &= \frac{1}{2}\sum_{s=1}^{n}(\eta_s\xi_s+\xi_s\eta_s) = \mathscr{C}+n/2,\end{aligned} \tag{2.12}$$

and their commutation relations by

$$[\mathscr{E},\mathscr{D}^\dagger]=2\mathscr{D}^\dagger,\quad [\mathscr{E},\mathscr{D}]=-2\mathscr{D},\quad [\mathscr{D},\mathscr{D}^\dagger]=4\mathscr{E}. \tag{2.13}$$

It can be easily checked that the operators $\mathscr{D}^\dagger$, $\mathscr{D}$, and $\mathscr{E}$ commute with the generators Λ_{st} of O(n). The $\mathscr{U}(1)$ subgroup of $\mathscr{S}\!\mathscr{p}(2, R)$ is generated by the single operator

$$\mathscr{C} = \sum_{s=1}^{n}\eta_s\xi_s. \tag{2.14}$$

Using definition (2.1), the basic collective states (2.1) can be rewritten as

$$|\phi_N\rangle = a_N(\mathscr{D}^\dagger)^N|0\rangle, \quad N=0,1,\ldots, \tag{2.15}$$

where a_N is a normalization coefficient to be determined by the condition

$$\langle\phi_{N'}|\phi_N\rangle = \delta_{N'N}. \tag{2.16}$$

From the commutation relations (2.13), one gets

$$a_N = 2^{-N}[N!(n/2)_N]^{-1/2}, \tag{2.17}$$

where $(n/2)_N = n/2\,(n/2+1)\cdots(n/2+N-1)$ is Pochhammer's symbol, and the phase of a_N has been chosen to be equal to $+1$. With this convention, the generators of $\mathscr{S}\!\mathscr{p}(2, R)$ act on the collective states as follows:

$$\mathscr{D}^\dagger|\phi_N\rangle = 2[(N+1)(N+n/2)]^{1/2}|\phi_{N+1}\rangle, \tag{2.18a}$$

$$\mathscr{D}|\phi_N\rangle = 2[N(N+n/2-1)]^{1/2}|\phi_{N-1}\rangle, \tag{2.18b}$$

$$\mathscr{E}|\phi_N\rangle = 2(N+n/4)|\phi_N\rangle. \tag{2.18c}$$

Equation (2.18) clearly shows that the basic collective states belong to a single IR of the group $\mathscr{S}\!\mathscr{p}(2, R)$, characterized by its minimum weight $\langle n/2\rangle$ corresponding to the vacuum state $|\phi_0\rangle = |0\rangle$. When acting upon the collective states, the Casimir operator of $\mathscr{S}\!\mathscr{p}(2, R)$, defined as

$$\mathscr{G} = \mathscr{E}^2 - \tfrac{1}{2}[\mathscr{D}^\dagger\mathscr{D} + \mathscr{D}\mathscr{D}^\dagger], \tag{2.19}$$

is diagonal and its eigenvalue is equal to

$$\langle\mathscr{G}\rangle = \frac{n}{2}\left(\frac{n}{2}-2\right). \tag{2.20}$$

The group $\mathscr{S}\!\mathscr{p}(2, R)$, however, also acts upon the noncollective A-particle states [with the restriction that states belonging to different IRs of O(n) are not mixed]. Therefore it is not the dynamical group of collective states that we are looking for. To get the latter we have to restrict $\mathscr{S}\!\mathscr{p}(2, R)$ to the IR(0) of O(n), or in other words, to project it onto the collective subspace of the A-particle space. We shall denote the dynamical group of collective states obtained in this way by $\mathscr{S}\!\mathscr{p}_c(2, R)$. The projection of any operator onto the collective subspace will be distinguished from the operator itself by an upper or lower index c. The generators of $\mathscr{S}\!\mathscr{p}_c(2, R)$ are then the operators

$$\begin{aligned}\mathscr{D}^{c\dagger} &= \mathscr{P}_c\mathscr{D}^\dagger\mathscr{P}_c,\\ \mathscr{D}^c &= \mathscr{P}_c\mathscr{D}\mathscr{P}_c,\end{aligned} \tag{2.21}$$

and

$$\mathscr{E}^c = \mathscr{P}_c\mathscr{E}\mathscr{P}_c,$$

where $\mathscr{P}_c$ denotes the projection operator onto the collective subspace. As the operators $\mathscr{D}^\dagger$, $\mathscr{D}$, and $\mathscr{E}$ do not connect collective states with noncollective ones, we may suppress one $\mathscr{P}_c$ operator in the definition of any of the generators (2.21), writing, for instance, $\mathscr{D}^{c\dagger} = \mathscr{P}_c\mathscr{D}^\dagger = \mathscr{D}^\dagger\mathscr{P}_c$. As a consequence, the operators (2.21) satisfy commutation relations similar to Eq. (2.13) so that the dynamical group is indeed a symplectic group, as mentioned previously. We may now replace the generators of $\mathscr{S}p(2, R)$ by those of $\mathscr{S}p_c(2, R)$ in Eqs. (2.15) and (2.18). By doing so in Eq. (2.19), we get the Casimir operator $\mathscr{G}_c$ of $\mathscr{S}p_c(2, R)$ which reduces to a constant equal to

$$\mathscr{G}_c = n/2(n/2 - 2). \tag{2.22}$$

For practical use of the projection procedure one has to find explicit ways of carrying it out. In the following sections we describe three different methods of projection leading to three different realizations of the dynamical group. In Sec. 3, we begin by applying the well-known procedure based upon the Dzublik–Zickendraht transformation.

3. THE GROUP $\mathscr{S}p_c(2, R)$ IN THE SCHRÖDINGER REPRESENTATION OF COLLECTIVE STATES

In the representation wherein the operators $x_1,..., x_n$ are diagonal an A-particle state $|\Psi\rangle$ is represented by the wave function $\Psi(x_1,..., x_n)$, which is an element of a Hilbert space $\mathscr{H}$. The wave functions $\Psi_{t_1,...,t_n}(x_1,...,x_n)$, representing the states $|\Psi_{t_1,...,t_n}\rangle$ defined in Eq. (2.3), form a basis of $\mathscr{H}$. Projection onto the collective subspace $\mathscr{H}_c$ of $\mathscr{H}$ is carried out by means of the Dzublik–Zickendraht transformation. Although the projection of the collective part of an arbitrary Hamiltonian has already been worked out and the dynamical group of collective states has been obtained in that way by Moshinsky and his collaborators,[15] we exhibit below the detailed application of the Dzublik–Zickendraht transformation for two purposes. First we wish to show that the collective states and their dynamical group derived in Sec. 2 coincide with those of Ref. 15. Second, we shall need their explicit form in Sec. 5 so as to establish a mapping between $\mathscr{H}_c$ and the Barut Hilbert space of collective states $\mathscr{F}_c$ to be introduced in Sec. 4.

For the one-dimensional case, the Dzublik–Zickendraht transformation has the form[2]

$$x_s = \rho D^1_{ns}(\alpha_1, ..., \alpha_{n-1}), \quad s = 1, ..., n, \tag{3.1}$$

where

$$\rho = \Big(\sum_{s=1}^n x_s^2\Big)^{1/2} \tag{3.2}$$

is the collective coordinate, and $\alpha_1...,\alpha_{n-1}$ are the remaining $n-1$ noncollective coordinates. In Eq. (3.1), $\|D^1_{ts}(\alpha)\|$ is an $n\times n$ matrix defining the IR of the SO(n) group characterized by 1. As we do not need the full matrix of the representation but just the row $t = n$, we have only $n-1$ angular coordinates α rather than the full complement of $\frac{1}{2}n(n-1)$. Note that to be consistent with the definition of collective states as O(n) invariant states, one should use a representation matrix of O(n) instead of that of SO(n) appearing in Eq. (3.1). The replacement of O(n) by SO(n) in the definition of collective states in one dimension was justified in Sec. 2.

The elements of the collective subspace $\mathscr{H}_c$ depend upon the single collective coordinate ρ. We shall denote by $\psi(\rho)$ the wave function of an arbitrary collective state $|\psi\rangle$. The functions $\phi_N(\rho)$, representing the states $|\phi_N\rangle$ defined in Eq. (21.5), form a basis of $\mathscr{H}_c$. We shall find their explicit form below. Before coming to that point, let us first consider the realizations of the groups $\mathscr{S}p(2, R)$ and $\mathscr{S}p_c(2, R)$ in the Hilbert spaces $\mathscr{H}$ and $\mathscr{H}_c$, respectively.

When replacing the boson creation and annihilation operators in Eq. (2.12) by their definition (2.1), we get a realization of the generators of $\mathscr{S}p(2, R)$ in $\mathscr{H}$ that we continue to denote by the same symbols:

$$\begin{aligned}
\mathscr{D}^\dagger &= \frac{1}{2}\sum_s(-p_s^2 + x_s^2 - 2ix_sp_s - 1),\\
\mathscr{D} &= \frac{1}{2}\sum_s(-p_s^2 + x_s^2 + 2ix_sp_s + 1),\\
\mathscr{E} &= \frac{1}{2}\sum_s(p_s^2 + x_s^2).
\end{aligned} \tag{3.3}$$

Carrying out the transformation (3.1) in Eq. (3.3), we can express the generators of $\mathscr{S}p(2, R)$ in terms of $\rho, \alpha_1,..., \alpha_{n-1}, \partial/\partial\rho, \partial/\partial\alpha_1,...,$ and $\partial/\partial\alpha_n$. If in those expressions we keep only the part depending upon ρ and $\partial/\partial\rho$, we get their projection on $\mathscr{H}_c$. In this way, we find that the generators of $\mathscr{S}p_c(2, R)$ are realized in $\mathscr{H}_c$ by the following operators:

$$\mathscr{D}^{c\dagger} = \frac{1}{2}\left[\frac{\partial^2}{\partial\rho^2} + \left(\frac{n-1}{\rho} - 2\rho\right)\frac{\partial}{\partial\rho} + \rho^2 - n\right], \tag{3.4a}$$

$$\mathscr{D}^c = \frac{1}{2}\left[\frac{\partial^2}{\partial\rho^2} + \left(\frac{n-1}{\rho} + 2\rho\right)\frac{\partial}{\partial\rho} + \rho^2 + n\right], \tag{3.4b}$$

$$\mathscr{E}^c = \frac{1}{2}\left[-\frac{\partial^2}{\partial\rho^2} - \frac{n-1}{\rho}\frac{\partial}{\partial\rho} + \rho^2\right]. \tag{3.4c}$$

One can check that the operators (3.4) satisfy commutation relations similar to Eq. (2.1) and that the Casimir operator $\mathscr{G}_c$ defined in terms of them fulfills Eq. (2.22), as it should.

As shown in Eq. (2.18c), the basic collective states $|\phi_N\rangle$ are the eigenvectors of the weight generator $\mathscr{E}^c$ of $\mathscr{S}p_c(2, R)$ corresponding to the eigenvalues $2(N + n/4)$. The functions $\phi_N(\rho)$ that represent them in $\mathscr{H}_c$ are therefore the solutions of the differential equation

$$\frac{1}{2}\left[-\frac{d^2}{d\rho^2} - \frac{n-1}{\rho}\frac{d}{d\rho} + \rho^2\right]\phi_N(\rho) = 2\left(N + \frac{n}{4}\right)\phi_N(\rho), \tag{3.5}$$

which is the radial equation of an n-dimensional harmonic oscillator. They are given by

$$\phi_N(\rho) = b_N\,\rho^{-(n-1)/2} f_N^{(n-2)/2}(\rho), \tag{3.6}$$

where b_N is a normalization coefficient, and $f_N^{(n-2)/2}(\rho)$ can be expressed in terms of an associated Laguerre polynomial,[17]

$$f_N^{(n-2)}(\rho) = [2(N)!]^{1/2}[\Gamma(N + n/2)]^{-1/2} \times \exp(-\tfrac{1}{2}\rho^2)\rho^{(n-1)/2} L_N^{(n-2)/2}(\rho^2). \tag{3.7}$$

The functions $f_N^{(n-2)/2}(\rho)$ are normalized in such a way that

$$\int_0^\infty d\rho\, f_{N'}^{(n-2)/2}(\rho) f_N^{(n-2)/2}(\rho) = \delta_{N'N}. \tag{3.8}$$

To calculate the normalization coefficient b_N in Eq. (3.6), we first have to determine the measure $d\tau(\rho)$ of the new Hilbert space $\mathscr{H}_c$. In $\mathscr{H}$ the scalar product of any two elements $\Phi(x_1,..., x_n)$ and $\Psi(x_1,..., x_n)$ is defined by

$$\langle\Phi|\Psi\rangle = \int_{-\infty}^{+\infty} dx_1 \cdots \int_{-\infty}^{+\infty} dx_n\, \Phi^*(x_1,..., x_n)\Psi(x_1,..., x_n). \tag{3.9}$$

We choose the measure $d\tau(\rho)$ of $\mathscr{H}_c$ in such a way that the scalar product of any two elements $\phi(\rho)$ and $\psi(\rho)$ of $\mathscr{H}_c$ is the same whether calculated in $\mathscr{H}_c$ or in $\mathscr{H}$, i.e.,

$$\langle\phi|\psi\rangle = \int_{-\infty}^{+\infty} dx_1 \cdots \int_{-\infty}^{+\infty} dx_n\, \phi^*(\rho)\psi(\rho)$$

$$= \int_0^\infty d\tau(\rho)\,\phi^*(\rho)\psi(\rho). \tag{3.10}$$

When expressed in terms of the variables $\rho, \alpha_1,\dots,\alpha_{n-1}$, the volume element $dx_1\cdots dx_n$, becomes[2]

$$dx_1\dots dx_n = \rho^{n-1}d\rho\, d\omega(\alpha_1,\dots,\alpha_{n-1}), \tag{3.11}$$

where $d\omega(\alpha_1,\dots,\alpha_{n-1})$ is the measure in the space of noncollective variables. Therefore

$$d\tau(\rho) = A\rho^{n-1}d\rho, \tag{3.12}$$

where

$$A = \int d\omega(\alpha_1,\dots,\alpha_{n-1}) \tag{3.13}$$

is a constant that remains to be determined. For that purpose, let us apply Eq. (3.1) to the wave function of the vacuum state by taking $\phi(\rho) = \psi(\rho) = \phi_0(\rho)$, where

$$\phi_0(\rho) = \Psi_{0,\dots,0}(x_1,\dots,x_n) = \pi^{-n/4}\exp(-\tfrac{1}{2}\rho^2). \tag{3.14}$$

One gets

$$d\tau(\rho) = 2\pi^{n/2}[\Gamma(n/2)]^{-1}\rho^{n-1}d\rho. \tag{3.15}$$

It can be easily checked that the realizations (3.4a) and (3.4b) of $\mathscr{D}^{c\dagger}$ and $\mathscr{D}^c$ are Hermitian conjugates of each other and that the realization (3.4c) of $\mathscr{E}^c$ is Hermitian with respect to the scalar product (3.10) and the measure (3.15).

In order that the wave functions $\phi_N(\rho)$ form an orthonormal set under that measure, i.e.,

$$\int d\tau(\rho)\phi^*_{N'}(\rho)\phi_N(\rho) = \delta_{N'N}, \tag{3.16}$$

b_N must satisfy the condition

$$|b_N|^2 = A^{-1}, \tag{3.17}$$

owing to Eqs. (3.6) and (3.8). Only the phase of b_N now remains to be determined. Up to now, we have only considered the realization in $\mathscr{H}_c$ of the last of Eqs. (2.18a)–(2.18c). When considering the first of these equations (the second follows from it by Hermitian conjugation), one can fix the relative phase of the b_N's. It was shown in Ref. 17 that the functions $f_N^{(n-2)/2}(\rho)$ satisfy the following equation:

$$\frac{1}{2}\Big[-\frac{d^2}{d\rho^2} + 2\rho\frac{d}{d\rho} + \frac{(n-1)(n-3)}{4\rho^2} - \rho^2 + 1\Big]f_N^{(n-2)/2}(\rho)$$

$$= 2[(N+1)(N+n/2)]^{1/2}f_{N+1}^{(n-2)/2}(\rho). \tag{3.18}$$

Therefore $\phi_N(\rho)$ will fulfill the condition

$$\frac{1}{2}\Big[\frac{d^2}{d\rho^2} + \Big(\frac{n-1}{\rho} - 2\rho\Big)\frac{d}{d\rho} + \rho^2 - n\Big]\phi_N(\rho)$$

$$= 2[(N+1)(N+n/2)]^{1/2}\phi_{N+1}(\rho) \tag{3.19}$$

if $b_{N+1} = -b_N$. By requiring that the expressions for $\phi_0(\rho)$ deduced from Eqs. (3.6) and (3.14) be the same, one gets $b_0 = 1$ so that the overall phase factor of the b_N's is determined to be $(-1)^N$. The final result for b_N is therefore

$$b_N = (-1)^N(1/\sqrt{2})\pi^{-n/4}[\Gamma(n/2)]^{1/2}. \tag{3.20}$$

This completes the derivation of the explicit form of the functions representing the collective states in $\mathscr{H}_c$.

In the next section we shall proceed to realize the collective states and their dynamical group in the Bargmann Hilbert space of analytic functions.

4. THE GROUP $\mathscr{S}\not{p}_c(2, R)$ IN THE BARUT REPRESENTATION OF COLLECTIVE STATES

In the Bargmann realization of quantum mechanics,[9] an arbitrary A-particle state $|\Psi\rangle$ is represented by an analytic function $\bar{\Psi}(z_1,\dots,z_n)$ of n complex variables z_s, $s = 1,\dots,n$. The space $\mathscr{F}$ spanned by those analytic functions is a Hilbert space, whose scalar product is defined by

$$\langle\Phi|\Psi\rangle = \int d\mu(z_1)\cdots d\mu(z_n)\,[\bar{\Phi}(z_1,\dots,z_n)]^*\bar{\Psi}(z_1,\dots,z_n), \tag{4.1}$$

where

$$d\mu(z_s) = \pi^{-1}\exp(-z_sz_s^*)\,d\,\mathrm{Re}\,z_s\,d\,\mathrm{Im}\,z_s, \quad s = 1,\dots,n. \tag{4.2}$$

The Bargmann representation is well adapted to harmonic oscillator problems as the boson operators (2.1) are represented by

$$\eta_s = z_s \quad \text{and} \quad \xi_s = \partial/\partial z_s. \tag{4.3a,b}$$

The functions $\bar{\Psi}_{\mathcal{N}_1,\dots,\mathcal{N}_n}(z_1,\dots,z_n)$, representing the basic states (2.3), are obtained from the ground state function $\bar{\Psi}_{0,\dots,0}(z_1,\dots,z_n) = 1$ by action of the creation operators (4.3a) and are therefore written as

$$\bar{\Psi}_{\mathcal{N}_1,\dots,\mathcal{N}_n}(z_1,\dots,z_n) = (\mathcal{N}_1!\dots\mathcal{N}_n!)^{-1/2}z_1^{\mathcal{N}_1}\cdots z_n^{\mathcal{N}_n}. \tag{4.4}$$

They span an orthonormal basis of the Bargmann Hilbert space $\mathscr{F}$. From Eq. (4.3), it is easy to show that the generators of the dynamical groups $\mathbf{Sp}(2n, R)$ and $\mathscr{S}\not{p}(2, R)$, defined in Sec. 2, are given by

$$\begin{aligned} \mathbf{D}^\dagger_{st} &= z_sz_t, & s\leqslant t &= 1,\dots,n, \\ \mathbf{D}_{st} &= \frac{\partial^2}{\partial z_s\partial z_t}, & s\leqslant t &= 1,\dots,n, \\ \mathbf{E}_{st} &= z_s\frac{\partial}{\partial z_t} + \tfrac{1}{2}\delta_{st}, & s, t &= 1,\dots,n, \end{aligned} \tag{4.5}$$

and

$$\mathscr{D}^\dagger = \sum_s z_s^2, \tag{4.6a}$$

$$\mathscr{D} = \sum_s \frac{\partial^2}{\partial z_s^2}, \tag{4.6b}$$

$$\mathscr{E} = \sum_s z_s\frac{\partial}{\partial z_s} + \frac{n}{2}, \tag{4.6c}$$

respectively.

In Sec. 2, we showed that the basic collective states are given by Eq. (2.15). In Bargmann representation, their wave functions can be written as

$$\bar{\phi}_N(z_1,...,z_n) = a_N\Big(\sum_s z_s^2\Big)^N, \tag{4.7}$$

by using the expression (4.6a) of the generator $\mathscr{D}^\dagger$. This expression of $\bar{\phi}_N$ suggests the introduction of the complex collective variable

$$w = \sum_s z_s^2. \tag{4.8}$$

As can be easily seen, w is the only scalar with respect to O(n) that can be built from the vector components z_s. The n complex variables z_s, $s = 1,...,n$, can therefore be replaced by the single complex collective variable w and $n-1$ complex noncollective variables, whose explicit expression will not be needed hereafter. The wave functions (4.7) of the basic collective states are rewritten as

$$\bar{\phi}_N(w) = 2^{-N}\Big[N!\Big(\frac{n}{2}\Big)_N\Big]^{-1/2} w^N, \tag{4.9}$$

and form an orthonormal set of functions in $\mathscr{F}$, i.e.,

$$\int d\mu(z_1)\cdots d\mu(z_n)\ [\bar{\phi}_{N'}(w)]^*\bar{\phi}_N(w) = \delta_{N'N}\,. \tag{4.10}$$

As any collective state $|\psi\rangle$ can be expanded in terms of the states $|\phi_N\rangle$, it is represented in $\mathscr{F}$ by an analytic function of w, $\bar{\psi}(w)$. All those functions span a subspace $\mathscr{F}_c$ of $\mathscr{F}$. We shall proceed to determine the nature of that subspace below.

Before coming to that point, let us first realize the dynamical group of collective states $\mathscr{S}\!\mathscr{p}_c(2, R)$ in $\mathscr{F}_c$. As the collective wave functions depend upon w, we have to retain in Eq. (4.6) only the terms containing w or $\partial/\partial w$. In this way we get

$$\mathscr{D}^{c\dagger} = w, \tag{4.11a}$$

$$\mathscr{D}^c = 4w\frac{\partial^2}{\partial w^2} + 2n\frac{\partial}{\partial w}, \tag{4.11b}$$

$$\mathscr{E}^c = 2w\frac{\partial}{\partial w} + \frac{n}{2}. \tag{4.11c}$$

In agreement with Sec. 2, the Casimir operator of the symplectic group becomes a multiple of the unit operator in this realization. It can be also checked that the operators (4.11) act on the functions $\bar{\phi}_N(w)$ in accordance with Eq. (2.18). In particular the functions $\bar{\phi}_N(w)$ are the eigenfunctions of $\mathscr{E}^c$ corresponding to the eigenvalues $2(N + n/4)$.

To make $\mathscr{F}_c$ a Hilbert space we have to equip it with a scalar product. That scalar product follows in a natural way from the scalar product $\langle\Phi|\Psi\rangle$ defined in $\mathscr{F}$ by restricting it to collective states. When replacing $\bar{\Psi}(z_1,...,z_n)$ and $\bar{\Phi}(z_1,...,z_n)$ by $\bar{\psi}(w)$ and $\bar{\phi}(w)$ respectively, Eq. (4.1) can indeed be put into the form

$$\langle\phi|\psi\rangle = \int d\mu(z_1)\cdots d\mu(z_n)\ [\bar{\phi}(w)]^*\bar{\psi}(w)$$

$$= \int d\sigma(w)\ [\bar{\phi}(w)]^*\bar{\psi}(w), \tag{4.12}$$

after performing the transformation from the variables $z_1,...,z_n$ to the variable w and $n-1$ noncollective variables and integrating over the latter. The measure

$$d\sigma(w) = f(w)\, d\,\mathrm{Re}\, w\, d\,\mathrm{Im}\, w \tag{4.13}$$

of $\mathscr{F}_c$ depends upon a weight function $f(w)$, independent of the functions $\bar{\psi}(w)$ and $\bar{\phi}(w)$, that remains to be determined.

Instead of explicitly performing the transformation from the variables $z_1,...,z_n$ to the collective and noncollective variables, it is much easier to determine the weight function $f(w)$ from the Hermiticity condition to be fulfilled by the realizations (4.11a) and (4.11b) of $\mathscr{D}^{c\dagger}$ and $\mathscr{D}^c$, i.e., $(\mathscr{D}^{c\dagger})^\dagger = \mathscr{D}^c$. For arbitrary collective wave functions $\bar{\psi}(w)$ and $\bar{\phi}(w)$, the following relation must hold

$$\int d\sigma(w)\ [w\bar{\phi}(w)]^*\bar{\psi}(w)$$

$$= \int d\sigma(w)\ [\bar{\phi}(w)]^*\Big[4w\frac{\partial^2}{\partial w^2} + 2n\frac{\partial}{\partial w}\Big]\bar{\psi}(w). \tag{4.14}$$

By integrating by parts in the right-hand side of Eq. (4.14), we obtain the differential equation

$$\Big[4w\frac{\partial^2}{\partial w^2} - 2(n-4)\frac{\partial}{\partial w} - w^*\Big]f(w) = 0. \tag{4.15}$$

When we multiply this equation by w, we obtain a differential equation invariant under the transformation $w \rightarrow w\exp(i\varphi)$. Let us therefore consider a new variable $u = (ww^*)^{1/2}$, in terms of which Eq. (4.15) becomes

$$\Big[u^2\frac{d^2}{du^2} - (n-3)u\frac{d}{du} - u^2\Big]f(u) = 0. \tag{4.16}$$

For the scalar product of $\bar{\psi}(w)$ and $\bar{\phi}(w)$ in $\mathscr{F}_c$ to be convergent, an asymptotic condition has to be imposed on $f(u)$ when $u \rightarrow \infty$. Taking this into account, the solution of Eq. (4.16) is given by

$$f(u) = \alpha u^{(n-2)/2}K_{(n-2)/2}(u), \tag{4.17}$$

where K is a modified Bessel function[18] and α is a yet undetermined constant. It can be checked that with the weight function (4.17), the realization (4.11c) of $\mathscr{E}^c$ is Hermitian as it should be.

Finally we determine the constant α by imposing Eq. (4.12) validity. For that purpose, we only have to consider two particular functions for $\bar{\psi}(w)$ and $\bar{\phi}(w)$, for instance $\bar{\psi}(w) = \bar{\phi}(w) = \bar{\phi}_0(w) = 1$. Equation (4.12) then reduces to $\int d\sigma(w) = 1$, and using formula (6.561.16) of p. 684 in Ref. 18, we finally get the measure

$$d\sigma(w) = [\pi 2^{n/2}\Gamma(n/2)]^{-1}(ww^*)^{1/4(n-2)}$$

$$\times K_{1/2(n-2)}(\sqrt{ww^*})d\,\mathrm{Re}\, w\, d\,\mathrm{Im}\, w. \tag{4.18}$$

As a consequence of Eq. (4.12), the set of functions $\bar{\phi}_N(w)$ remains orthonormal in $\mathscr{F}_c$, i.e.,

$$\int d\sigma(w)\ [\bar{\phi}_{N'}(w)]^*\bar{\phi}_N(w) = \delta_{N'N}\,. \tag{4.19}$$

A scalar product similar to the one defined in Eqs. (4.12) and (4.18) was obtained by Barut and Girardello[11] in considering the coherent states associated with the Lie algebra of the symplectic group Sp(2, R). Therefore the states left invariant by the transformations of the orthogonal group O(n) span a Barut Hilbert space. A similar result was obtained by Kramer *et al.*[10] for O(2) in another context.

In the next section we shall consider in detail the mapping between $\bar{\psi}(w)$, which we shall call the Barut representation for the above-mentioned reasons, and the Schrödinger representation $\psi(\rho)$ encountered in Sec. 3. This will achieve a link between the two realizations of the collective states considered above.

5. UNITARY MAPPING BETWEEN THE SCHRÖDINGER AND BARUT REPRESENTATIONS OF COLLECTIVE STATES

For the purpose of establishing a mapping between the collective subspaces $\mathscr{H}_c$ and $\mathscr{F}_c$, let us briefly recall the relation between the Schrödinger and Bargmann pictures in the full Hilbert space.[9] The representations $\Psi(x_1,...,x_n)$ and $\bar{\Psi}(z_1,...,z_n)$ of any state $|\Psi\rangle$ in $\mathscr{H}$ and $\mathscr{F}$, respectively, are related by the integral transform

$$\bar{\Psi}(z_1,...,z_n) = \int dx_1...dx_n\, A(z_1,x_1)...A(z_n,x_n)\Psi(x_1,...,x_n), \tag{5.1}$$

whose kernel is given by

$$A(z_s,x_s) = \pi^{-1/4}\exp[-\tfrac{1}{2}(z_s^2+x_s^2)+\sqrt{2}z_s x_s], \quad s=1,...,n. \tag{5.2}$$

That transformation is unitary, meaning that the scalar product $\langle\Phi|\Psi\rangle$, defined in Eqs. (3.9) and (4.1) for $\mathscr{H}$ and $\mathscr{F}$, respectively, is preserved.

Let us now show that the transformation (5.1) maps $\mathscr{H}_c$ onto $\mathscr{F}_c$. If $\Psi(x_1,...,x_n)$ is the Schrödinger picture $\psi(\rho)$ of a collective state, if follows from Eq. (5.1) that its Bargmann picture is given by

$$\bar{\Psi}(z_1,...,z_n) = \int dx_1...dx_n\, A(z_1,x_1)...A(z_n,x_n)\psi(\rho). \tag{5.3}$$

It can be easily seen that the right-hand side of Eq. (5.3) is invariant under the transformation $z_s \to \Sigma_t O_{st} z_t$ of O(n), so that the left-hand side depends only on the collective variable w and thus belongs to $\mathscr{F}_c$. Using Eqs. (3.11) and (3.12), the integral transform (5.3) takes the form

$$\bar{\psi}(w) = \int d\tau(\rho) B(w,\rho)\psi(\rho), \tag{5.4}$$

where the kernel $B(w,\rho)$ results from the integration over the noncollective variables. We shall, however, use another method to determine $B(w,\rho)$ because the direct method just outlined leads to very complicated calculations.

For that purpose, let us profit from the fact that the action of the generators of $\mathscr{S}p_c(2,R)$ must be the same in both the Bargmann and Schrödinger representations. This leads to the following relations:

$$w\bar{\psi}(w) = \int d\tau(\rho)\, B(w,\rho)\frac{1}{2}\left[\frac{\partial^2}{\partial\rho^2}+\left(\frac{n-1}{\rho}-2\rho\right)\frac{\partial}{\partial\rho} + \rho^2 - n\right]\psi(\rho), \tag{5.5a}$$

$$\left(4w\frac{\partial^2}{\partial w^2}+2n\frac{\partial}{\partial w}\right)\bar{\psi}(w) = \int d\tau(\rho)\, B(w,\rho)\frac{1}{2}\left[\frac{\partial^2}{\partial\rho^2}+\left(\frac{n-1}{\rho}+2\rho\right)\frac{\partial}{\partial\rho} + \rho^2 + n\right]\psi(\rho), \tag{5.5b}$$

$$\left(2w\frac{\partial}{\partial w}+\frac{n}{2}\right)\bar{\psi}(w) = \int d\tau(\rho)\, B(w,\rho)\frac{1}{2}\left[-\frac{\partial^2}{\partial\rho^2}-\frac{n-1}{\rho}\frac{\partial}{\partial\rho}+\rho^2\right]\psi(\rho), \tag{5.5c}$$

when Eqs. (3.4) and (4.11) are used. As these relations must be satisfied by any collective function $\psi(\rho)$, they are equivalent to the following system of partial differential equations:

$$\left[\frac{\partial^2}{\partial\rho^2}+\left(\frac{n-1}{\rho}+2\rho\right)\frac{\partial}{\partial\rho}+\rho^2+n-2w\right]B(w,\rho)=0, \tag{5.6a}$$

$$\left[\frac{\partial^2}{\partial\rho^2}+\left(\frac{n-1}{\rho}-2\rho\right)\frac{\partial}{\partial\rho}+\rho^2-n-8w\frac{\partial^2}{\partial w^2} - 4n\frac{\partial}{\partial w}\right]B(w,\rho)=0, \tag{5.6b}$$

$$\left[\frac{\partial^2}{\partial\rho^2}+\frac{n-1}{\rho}\frac{\partial}{\partial\rho}-\rho^2+4w\frac{\partial}{\partial w}+n\right]B(w,\rho)=0. \tag{5.6c}$$

To solve the latter it is convenient to first substract Eq. (5.6c) from Eq. (5.6a). In this way we get the first-order equation

$$\left[\rho\frac{\partial}{\partial\rho}+\rho^2\right]B(w,\rho) = \left[2w\frac{\partial}{\partial w}+w\right]B(w,\rho), \tag{5.7}$$

whose solution is

$$B(w,\rho) = \exp(-\tfrac{1}{2}\rho^2-\tfrac{1}{2}w)g(x), \tag{5.8}$$

where $g(x)$ is an arbitrary function of $x = w\rho^2$. Any of the three equations (5.6a)–(5.6c) then leads to the same ordinary differential equation for $g(x)$,

$$2x\frac{d^2g}{dx^2}+n\frac{dg}{dx}-g=0, \tag{5.9}$$

whose regular solution is an hypergeometric function ${}_0F_1(n/2;x/2)$. The kernel of the integral transform (5.4) is therefore given by

$$B(w,\rho) = \pi^{-n/4}\exp(-\tfrac{1}{2}\rho^2-\tfrac{1}{2}w)\,{}_0F_1(n/2;\tfrac{1}{2}w\rho^2), \tag{5.10a}$$

or

$$B(w,\rho) = \pi^{-n/4}\Gamma(n/2)\exp(-\tfrac{1}{2}\rho^2-\tfrac{1}{2}w) \times[\rho(w/2)^{1/2}]^{-(n-2)/2}I_{(n-2)/2}[\rho(2w)^{1/2}], \tag{5.10b}$$

where I is a modified Bessel function.[18] The normalization coefficient has been determined by the condition that the Schrödinger picture of the vacuum state $\phi_0(\rho) = \pi^{-n/4}\exp(-\tfrac{1}{2}\rho^2)$ be mapped on its Barut picture $\bar{\phi}_0(w) = 1$. In its calculation, formula (6.631.4) of p. 717 in Ref. 18 has been used.

From Secs. 3 and 4, we know that the scalar product of two collective states is the same whether calculated in $\mathscr{H}(\mathscr{F})$

or in $\mathscr{H}_c(\mathscr{F}_c)$. From the unitarity of the mapping between $\mathscr{H}$ and $\mathscr{F}$, it follows that the scalar product of two collective states is preserved when going from $\mathscr{H}_c$ to $\mathscr{F}_c$:

$$\int d\sigma(w)\,[\bar{\phi}(w)]^*\bar{\psi}(w) = \int d\tau(\rho)\,[\phi(\rho)]^*\psi(\rho). \qquad (5.11)$$

By introducing the integral transform (5.4) into the left-hand side of Eq. (5.11), we obtain the relation

$$\int d\sigma(w)\,d\tau(\rho)\,d\tau(\rho')\,[B(w,\rho)]^*B(w,\rho')[\phi(\rho)]^*\psi(\rho') = \int d\tau(\rho)\,[\phi(\rho)]^*\psi(\rho), \qquad (5.12)$$

which implies that

$$\int d\sigma(w)\,[B(w,\rho)]^*B(w,\rho') = \Gamma(n/2)[2\pi^{n/2}\rho^{n-1}]^{-1}\delta(\rho-\rho'). \qquad (5.13)$$

This expression is nothing else than the reproducing kernel of $\mathscr{H}_c$:

$$\int d\tau(\rho')\left\{\int d\sigma(w)\,[B(w,\rho)]^*B(w,\rho')\right\}\psi(\rho') = \psi(\rho). \qquad (5.14)$$

By permuting the order of integration in the left-hand side of Eq. (5.14) and using Eq. (5.4), we get the inverse transform

$$\psi(\rho) = \int d\sigma(w)\,[B(w,\rho)]^*\bar{\psi}(w). \qquad (5.15)$$

As the transformation is unitary, it is not surprising to find that the kernel of the inverse transform is the complex conjugate of that of the direct transform as was already the case for Bargmann's representation.

The last point we want to discuss in connection with the Barut space $\mathscr{F}_c$ is its reproducing kernel. Let us apply successively Eqs. (5.4) and (5.15). We obtain

$$\bar{\psi}(w') = \int d\sigma(w)\,d\tau(\rho)\,B(w',\rho)[B(w,\rho)]^*\bar{\psi}(w), \qquad (5.16)$$

which means that

$$K(w',w) = \int d\tau(\rho)\,B(w',\rho)[B(w,\rho)]^* \qquad (5.17)$$

is the reproducing kernel in $\mathscr{F}_c$, i.e.,

$$\bar{\psi}(w') = \int d\sigma(w)\,K(w',w)\bar{\psi}(w). \qquad (5.18)$$

A simple calculation, using formula (6.615) of p. 710 in Ref. 18, leads to the relation

$$K(w',w) = 2^{(n-2)/2}\Gamma(n/2)(w'w^*)^{-(n-2)/4}I_{1/2(n-2)}[(w'w^*)^{1/2}], \qquad (5.19a)$$

or

$$K(w',w) = {}_0F_1(n/2;\tfrac{1}{4}w'w^*). \qquad (5.19b)$$

The reproducing kernel $K(w',w)$ will play an important part in considering coherent collective states in the next section.

6. COHERENT COLLECTIVE STATES

As shown in Sec. 4, the collective states span a Barut space associated with a well-defined IR of $\mathscr{S}\!p_c(2,R)$. This makes it possible to define coherent collective states.

The reproducing kernel, defined in Eqs. (5.18) and (5.19), is an analytic function in both complex variables w' and w^*. Its complex conjugate $[K(w',w)]^*$ is therefore an analytic function of w and belongs to $\mathscr{F}_c$. Let us denote by $|w'\rangle$ the state whose Barut representation is given by $[K(w',w)]^*$, w' playing the role of a complex parameter.

The scalar product (4.12) of $|w'\rangle$ with an arbitrary state $|\psi\rangle$ is equal to

$$\langle w'|\psi\rangle = \int d\sigma(w)\,K(w',w)\bar{\psi}(w) = \bar{\psi}(w'), \qquad (6.1)$$

using the definition of the reproducing kernel. The Barut representation of a collective state is therefore given by a scalar product, as for any representation in quantum mechanics, provided the states $|w'\rangle$ form a basis. This will be the case if they give rise to a resolution of the unit operator I_c of the collective subspace $\mathscr{F}_c$, i.e., if

$$\int d\sigma(w)\,|w\rangle\langle w| = I_c. \qquad (6.2)$$

To prove Eq. (6.2), one only has to show that the matrix elements of both sides with respect to the basic collective states $|\phi_N\rangle$ are equal, or in other words that

$$\int d\sigma(w)\,\langle\phi_{N'}|w\rangle\langle w|\phi_N\rangle = \delta_{N'N}. \qquad (6.3)$$

From Eq. (6.1), one has $\langle w|\phi_N\rangle = \bar{\phi}_N(w)$, so that Eq. (6.3) reduces to the orthonormality condition (4.19) of the set of functions $\bar{\phi}_N(w)$. Therefore the closure relation (6.2) holds true and the states $|w\rangle$ span a basis of $\mathscr{F}_c$. It is interesting to note that when the left-hand side of Eq. (6.2) is considered as acting on the whole space $\mathscr{F}$, the operator I_c in the right-hand side must be replaced by the projection operator $\mathscr{P}_c$, defined in Sec. 2.

The overlap between the old basis states $|\phi_N\rangle$ and the new ones $|w\rangle$ follows from Eqs. (6.1) and (4.9),

$$\langle w|\phi_N\rangle = \bar{\phi}_N(w) = 2^{-N}[N!(n/2)_N]^{-1/2}w^N, \qquad (6.4)$$

and implies the following expansion:

$$|w\rangle = \sum_{N=0}^{\infty}|\phi_N\rangle\langle\phi_N|w\rangle = \sum_{N=0}^{\infty}2^{-N}[N!(n/2)_N]^{-1/2}(w^*)^N|\phi_N\rangle. \qquad (6.5)$$

This expression is very useful in proving the most important property of $|w\rangle$, namely that $|w\rangle$ is an eigenvector of the lowering operator $\mathscr{D}^c$ of $\mathscr{S}\!p_c(2,R)$ corresponding to the eigenvalue w^*. Acting with $\mathscr{D}^c$ on Eq. (6.5) and using Eq. (2.18b), we get immediately

$$\mathscr{D}^c|w\rangle = w^*|w\rangle, \qquad (6.6)$$

showing that $|w\rangle$ is a coherent state associated with the Lie algebra of $\mathscr{S}\!p_c(2,R)$ in the sense of Barut and Girardello.[11]

The coherent states $|w\rangle$ are overcomplete and do not form an orthonormal set. From Eq. (6.1), the overlap between any two coherent states is nothing else than the reproducing kernel

$$\langle w'|w\rangle = K(w',w). \qquad (6.7)$$

As a consequence of the closure relation (6.2), any collective state $|\psi\rangle$ can be expanded in terms of the coherent states $|w\rangle$:

$$|\psi\rangle = \int d\sigma(w)\, |w\rangle\, \langle w|\psi\rangle = \int d\sigma(w)\, \bar{\psi}(w) |w\rangle. \tag{6.8}$$

The coherent states introduced in the present section will play an important role, similar to that of the usual coherent sates associated with the Heisenberg algebra,[19,20] in the collective operators (or collective part of operators) representation and in their matrix elements evaluation.

7. GENERALIZED HOLSTEIN–PRIMAKOFF REPRESENTATION OF $\mathscr{S}\!\not{p}_c(2, R)$

In this section we present a third realization of the dynamical group of collective states that is suitable for investigating the relation between the microscopic model of collective states considered in the present paper and the macroscopic model known as the IBM.[5] In one dimension the latter makes use of a $\mathfrak{U}(1)$ group instead of the $\mathfrak{U}(6)$ one considered in three dimensions. Therefore we have to establish a connection between the basic collective states $|\phi_N\rangle$ and the eigenstates of a one-dimensional harmonic oscillator, whose symmetry group is $\mathfrak{U}(1)$. For such purpose, it is interesting to define new boson creation and annihilation operators.

Let us first introduce a creation operator $a^\dagger$ that is acting on $|\phi_N\rangle$ as the creation operator η acts on a one-dimensional harmonic oscillator state

$$a^\dagger|\phi_N\rangle = (N+1)^{1/2}|\phi_{N+1}\rangle. \tag{7.1a}$$

It follows that its Hermitian conjugate a acts as

$$a|\phi_N\rangle = \sqrt{N}\,|\phi_{N-1}\rangle, \tag{7.1b}$$

and that $a^\dagger a$ measures the number of quanta,

$$a^\dagger a|\phi_N\rangle = N\,|\phi_N\rangle. \tag{7.2}$$

Moreover the a and $a^\dagger$ operators so defined obey the commutation rule

$$[a, a^\dagger] = I. \tag{7.3}$$

Next let us realize the generators of $\mathscr{S}\!\not{p}_c(2, R)$ in terms of these boson operators. From Eq. (2.18a), we know that

$$\mathscr{D}^{c\dagger}|\phi_N\rangle = 2[(N+1)(N+n/2)]^{1/2}|\phi_{N+1}\rangle. \tag{7.4}$$

Using Eqs. (7.1a) and (7.2), this relation can be rewritten as

$$\mathscr{D}^{c\dagger}|\phi_N\rangle = 2a^\dagger[a^\dagger a + n/2]^{1/2}|\phi_N\rangle. \tag{7.5a}$$

A similar analysis leads to

$$\mathscr{D}^{c}|\phi_N\rangle = 2[a^\dagger a + n/2]^{1/2}a|\phi_N\rangle, \tag{7.5b}$$

and

$$\mathscr{E}^{c}|\phi_N\rangle = 2[a^\dagger a + n/4]|\phi_N\rangle. \tag{7.5c}$$

We therefore obtain the following realization of $\mathscr{S}\!\not{p}_c(2, R)$ in terms of boson operators

$$\mathscr{D}^{c\dagger} = 2a^\dagger[a^\dagger a + n/2]^{1/2}, \tag{7.6a}$$

$$\mathscr{D}^{c} = 2[a^\dagger a + n/2]^{1/2}a, \tag{7.6b}$$

$$\mathscr{E}^{c} = 2[a^\dagger a + n/4], \tag{7.6c}$$

which is simply the generalized Holstein–Primakoff representation[13] of $\mathscr{S}\!\not{p}_c(2, R)$, previously derived by Mlodinow and Papanicolaou[21] in another context.

Finally we are able to invert Eq. (7.6) and to express the $\mathfrak{U}(1)$ generator associated with the one-dimensional harmonic oscillator in terms of the generators of $\mathscr{S}\!\not{p}_c(2, R)$. From Eq. (7.6c), the generator of $\mathfrak{U}(1)$ is given by

$$\mathfrak{C} = a^\dagger a = \tfrac{1}{2}(\mathscr{E}^c - n/2). \tag{7.7}$$

Introducing Eq. (7.7) into Eqs. (7.6a) and (7.6b), we also get the boson creation and annihilation operators in terms of the generators of $\mathscr{S}\!\not{p}_c(2, R)$:

$$a^\dagger = \mathscr{D}^{c\dagger}[2\mathscr{E}^c + n]^{-1/2}, \tag{7.8a}$$

$$a = [2\mathscr{E}^c + n]^{-1/2}\mathscr{D}^c. \tag{7.8b}$$

The results of the present section will make it possible to express the collective part of any operator in terms of boson creation and annihilation operators and to study under which conditions it reduces to the predictions of the IBM.

8. OUTLINE OF THE GENERALIZATION TO d DIMENSIONS

In this section we wish to determine the dynamical group of collective states in the d-dimensional case (where $d = 2$ or 3) and show how to find realizations of that group similar to the ones described in the preceding sections for the one-dimensional case.

For a system of A particles in d dimensions, the Jacobi coordinates and their conjugate momenta are denoted respectively by x_{is} and $p_{is} = -i\partial/\partial x_{is}$, where $i = 1,...,d$, and $s = 1,..., n = A - 1$. Boson creation and annihilation operators, η_{is} and ξ_{is}, can be defined by relations similar to Eq. (2.1). A basis for the A-particle states is made of the states

$$|\Psi_{\mathscr{N}_{11},...\mathscr{N}_{dn}}\rangle = \prod_{i=1}^{d}\prod_{s=1}^{n}\,[(\mathscr{N}_{is}!)^{-1/2}\eta_{is}^{\mathscr{N}_{is}}]|0\rangle,$$
$$\mathscr{N}_{11},...\mathscr{N}_{dn} = 0, 1,... . \tag{8.1}$$

A dynamical group for the states (8.1) is the group of linear canonical transformations in dn dimensions $\mathbf{Sp}(2dn, R)$, whose generators are[14]

$$\begin{aligned}
\mathbf{D}^\dagger_{is,jt} &= \eta_{is}\eta_{jt}, && is \leqslant jt = 11,..., dn,\\
\mathbf{D}_{is,jt} &= \xi_{is}\xi_{jt}, && is \leqslant jt = 11,..., dn,\\
\mathbf{E}_{is,jt} &= \tfrac{1}{2}(\eta_{is}\xi_{jt} + \xi_{jt}\eta_{is})\\
&= \mathbf{C}_{is,jt} + \tfrac{1}{2}\delta_{ij}\delta_{st}, && is,jt = 11,..., dn,
\end{aligned} \tag{8.2}$$

where

$$\mathbf{C}_{is,jt} = \eta_{is}\xi_{jt} \tag{8.3}$$

are the generators of the $\mathbf{U}(dn)$ subgroup. The analog of the chain (2.9) is

$$\begin{gathered}
\mathbf{Sp}(2dn, R) \supset \mathbf{U}(dn) \supset \mathscr{U}(d) \times U(n),\\
\mathscr{U}(d) \supset \mathscr{S}\mathscr{O}(d), \quad U(n) \supset \mathrm{O}(n).
\end{gathered} \tag{8.4}$$

The generators of $\mathscr{U}(d)$ and $U(n)$ are the operators

$$\mathscr{C}_{ij} = \sum_{s=1}^{n} \mathbf{C}_{is,js}, \quad \text{and} \quad C_{st} = \sum_{i=1}^{d} \mathbf{C}_{is,it}, \tag{8.5}$$

respectively. Those of $\mathscr{S}\mathscr{O}(d)$ and $\mathrm{O}(n)$ are the angular momentum operators L_{ij} in the d-dimensional space and the operators Λ_{st}, defined in Eq. (2.7), respectively.

The states (8.1) belong to one of two IR's of $\mathbf{Sp}(2dn, R)$, $\langle \frac{1}{2}^{dn}\rangle$ or $\langle \frac{1}{2}^{dn-1}\frac{3}{2}\rangle$ according to whether the total number of bosons $\mathscr{N} = \Sigma_{i=1}^{d}\Sigma_{s=1}^{n}\mathscr{N}_{is}$ is even or odd, and to the IR$[\mathscr{N}]$ of U(dn). By taking appropriate linear combinations, they could be classified according to the other groups of the chain (8.4). The IR's of $\mathscr{U}(d)$ and U(n) contained in the IR $[\mathscr{N}]$ of $U(dn)$ are characterized by the same partition $[h_1...h_p]$ of $\mathscr{N}$, where $p = \min(d, n)$. Those of $\mathscr{S}\mathscr{O}(d)$ and O(n) are specified by L and λ respectively, where λ denotes a Young diagram $(\lambda_1...\lambda_{[n/2]})$. Some additional quantum numbers may now be necessary to complete the classification of the states.

The collective A-particle states belong to the IR $\lambda = (0)$ of O(n). To form invariants from the boson creation operators, we now have at our disposal d vectors η_{is}, $s = 1,..., n$, corresponding to $i = 1,...,d$. The basic invariants of O(n) that can be formed from them are their scalar products $\Sigma_{s=1}^{n}\eta_{is}\eta_{js}$, where $i\leqslant j = 1,...,d$.[16] A basis for the collective states therefore consists of the states

$$|\phi_{N_{11}...N_{dd}}\rangle \propto \prod_{i\leqslant j=1}^{d}\left(\sum_{s=1}^{n}\eta_{is}\eta_{js}\right)^{N_{ij}}|0\rangle, \tag{8.6}$$

depending upon $\frac{1}{2}d(d+1)$ quantum numbers N_{ij}, $i\leqslant j = 1, ...,d$, which can take the values 0, 1, ...

The properties of the states (8.6) are most easily discussed in the framework of the chain

$$\begin{aligned} &\mathbf{Sp}(2dn, R)\supset \mathscr{S}\not{p}(2d, R)\times \mathrm{O}(n),\\ &\mathscr{S}\not{p}(2d, R)\supset \mathscr{U}(d)\supset \mathscr{S}\mathscr{O}(d), \end{aligned} \tag{8.7}$$

where $\mathscr{S}\not{p}(2d, R)$ is the group of linear canonical transformations in dn dimensions conserving the O(n) symmetry.[14] Its generators are the operators

$$\begin{aligned} \mathscr{D}^\dagger_{ij} &= \sum_{s=1}^{n}\eta_{is}\eta_{js}, && i\leqslant j = 1,..., d,\\ \mathscr{D}_{ij} &= \sum_{s=1}^{n}\xi_{is}\xi_{js}, && i\leqslant j = 1,..., d,\\ \mathscr{E}_{ij} &= \frac{1}{2}\sum_{s=1}^{n}(\eta_{is}\xi_{js} + \xi_{js}\eta_{is})\\ &= \mathscr{C}_{ij} + \frac{n}{2}\delta_{ij}, && i,j = 1,...,d, \end{aligned} \tag{8.8}$$

and satisfy the following commutation relations:

$$\begin{aligned} &[\mathscr{E}_{ij}, \mathscr{E}_{i'j'}] = \delta_{ji'}\mathscr{E}_{ij'} - \delta_{ij'}\mathscr{E}_{i'j},\\ &[\mathscr{E}_{ij}, \mathscr{D}^\dagger_{i'j'}] = \delta_{ji'}\mathscr{D}^\dagger_{ij'} + \delta_{jj'}\mathscr{D}^\dagger_{ii'},\\ &[\mathscr{E}_{ij}, \mathscr{D}_{i'j'}] = -\delta_{ii'}\mathscr{D}_{jj'} - \delta_{ij'}\mathscr{D}_{ji'},\\ &[\mathscr{D}^\dagger_{ij}, \mathscr{D}^\dagger_{i'j'}] = [\mathscr{D}_{ij}, \mathscr{D}_{i'j'}] = 0,\\ &[\mathscr{D}_{ij}, \mathscr{D}^\dagger_{i'j'}] = \delta_{ii'}\mathscr{E}_{j'j} + \delta_{ij'}\mathscr{E}_{i'j}\\ &\qquad + \delta_{ji'}\mathscr{E}_{j'i} + \delta_{jj'}\mathscr{E}_{i'i}. \end{aligned} \tag{8.9}$$

The weight generators are the operators $\mathscr{E}_{ii}$, $i = 1, ...,d$. The basic collective states (8.6) can be rewritten in terms of the operators $\mathscr{D}^\dagger_{ij}$ as

$$|\phi_{N_{11}...N_{dd}}\rangle \propto \prod_{i\leqslant j=1}^{d}(\mathscr{D}^\dagger_{ij})^{N_{ij}}|0\rangle. \tag{8.10}$$

They belong therefore to a single IR of $\mathscr{S}\not{p}(2d, R)$, characterized by its minimum weight $\langle (n/2)^d\rangle$ corresponding to the vacuum state $|\phi_{0...0}\rangle = |0\rangle$.

The dynamical group of collective states is the restriction of $\mathscr{S}\not{p}(2d, R)$ to the IR(0) of O(n). It is generated by the operators

$$\begin{aligned} &\mathscr{D}^{c\dagger}_{ij} = \mathscr{P}_c\mathscr{D}^\dagger_{ij}\mathscr{P}_c, \quad i\leqslant j = 1,..., d,\\ &\mathscr{D}^{c}_{ij} = \mathscr{P}_c\mathscr{D}_{ij}\mathscr{P}_c, \quad i\leqslant j = 1,..., d,\\ \text{and}\\ &\mathscr{E}^{c}_{ij} = \mathscr{P}_c\mathscr{E}_{ij}\mathscr{P}_c, \quad i,j = 1,..., d, \end{aligned} \tag{8.11}$$

where as before $\mathscr{P}_c$ is the projection operator onto the collective subspace and we may drop one of the $\mathscr{P}_c$ operators in the definition of $\mathscr{D}^{c\dagger}_{ij}$, $\mathscr{D}^{c}_{ij}$, and $\mathscr{E}^{c}_{ij}$. As the operators (8.11) satisfy commutation relations similar to Eq. (8.9), the dynamical group of collective states is a symplectic group in $2d$ dimensions, which we denote by $\mathscr{S}\not{p}_c(2d, R)$. In the case of the hypothetical two-dimensional space considered in Refs. 3 and 4, we get here the group $\mathscr{S}\not{p}_c(4, R)$ which is locally isomorphic to the group SO(3, 2) found in those papers. Our general result shows that when going to the physical three-dimensional space, the dynamical groups becomes $\mathscr{S}\not{p}_c(6, R)$.

As in Refs. 3 and 4 the projection onto the collective subspace was partly discussed in terms of the rotation group SO(n), it is worthwhile to examine what happens when we replace O(n) by SO(n) in the preceding discussion. The basic SO(n) invariants that can be formed from the d vectors η_{is} $(s = 1, ..., n)$, $i = 1,...,d$, include the determinants

$$\Delta^\dagger_{i_1i_2...i_n} = \begin{vmatrix} \eta_{i_11} & \eta_{i_21} & \cdots & \eta_{i_n1}\\ \eta_{i_12} & \eta_{i_22} & \cdots & \eta_{i_n2}\\ \vdots & \vdots & & \vdots\\ \eta_{i_1n} & \eta_{i_2n} & \cdots & \eta_{i_nn} \end{vmatrix}, \quad 1\leqslant i_1 < i_2 < \cdots < i_n \leqslant d, \tag{8.12}$$

in addition to the scalar products $\Sigma_{s=1}^{n}\eta_{is}\eta_{js}$, $i\leqslant j = 1,...,d$.[16] As those determinants only exist when $d\geqslant n$, we conclude that all the SO(n) invariant states remain invariant under O(n) whenever the number of particles A is large enough to fulfill the condition $n = A - 1 > d$. In the latter case, the collective states and their dynamical group may be studied in terms of either O(n) or SO(n).

When $d\geqslant n$ however, one must be cautious in using SO(n) instead of O(n). Fortunately, the condition $d\geqslant n$ is only fulfilled in three borderline cases: those of three particles in two dimensions ($d = n = 2$), and of three or four particles in three dimensions ($d = 3$, $n = 2$ or 3). Let us construct a basis for the SO(n) invariant states in these three cases. By using the relation

$$\Delta^\dagger_{i_1...i_n}\Delta^\dagger_{j_1...j_n} = \begin{vmatrix} \mathscr{D}^\dagger_{i_1j_1} & \mathscr{D}^\dagger_{i_1j_2} & \cdots & \mathscr{D}^\dagger_{i_1j_n}\\ \mathscr{D}^\dagger_{i_2j_1} & \mathscr{D}^\dagger_{i_2j_2} & \cdots & \mathscr{D}^\dagger_{i_2j_n}\\ \vdots & \vdots & & \vdots\\ \mathscr{D}^\dagger_{i_nj_1} & \mathscr{D}^\dagger_{i_nj_2} & \cdots & \mathscr{D}^\dagger_{i_nj_n} \end{vmatrix}, \tag{8.13}$$

the product of any two determinants (8.12) can be reexpressed as a polynomial in the operators $\mathscr{D}_{ij}^{\dagger}$, so that the total degree of the basic SO(n) invariant states in the determinants $\Delta^{\dagger}_{i_1\ldots i_n}$ may be restricted to the values zero and one. In the former case, we get the states (8.10), and in the latter one the states

$$\Delta^{\dagger}_{k_1\ldots k_n} \prod_{i\leqslant j=1}^{d} (\mathscr{D}_{ij}^{\dagger})^{N_{ij}}|0\rangle, \text{ where } 1\leqslant k_1<\cdots<k_n\leqslant d,$$

$$\text{and } N_{ij}=0,1,\ldots, \quad i\leqslant j=1,\ldots d. \tag{8.14}$$

As seen before, the states (8.10) can be generated from the vacuum state by application of the generators of $\mathscr{S}\!p(2d, R)$ and belong to a single IR of the latter, namely $\langle 1^2\rangle$, $\langle 1^3\rangle$, or $\langle \frac{1}{2}^3\rangle$ according to whether $d=n=2$, $d=3$ and $n=2$, or $d=n=3$. In the same way, all the states (8.14) can be generated from one of them, which is of minimum weight with respect to $\mathscr{S}\!p(2d, R)$: $\Delta^{\dagger}_{12}|0\rangle$, $\Delta^{\dagger}_{23}|0\rangle$, or $\Delta^{\dagger}_{123}|0\rangle$ according to whether $d=n=2$, $d=3$ and $n=2$, or $d=n=3$. They therefore belong to a single IR of $\mathscr{S}\!p(2d, R)$, which is $\langle 2^2\rangle$, $\langle 12^2\rangle$, or $\langle \frac{5}{2}^3\rangle$, respectively. In conclusion, the group $\mathscr{S}\!p(2d, R)$ separates the SO(n) invariant states into two classes corresponding to two inequivalent IR's of $\mathscr{S}\!p(2d, R)$: the O(n) invariant states on one hand, and the pseudoscalar states on the other hand. Projection onto the collective subspace can be made in two different ways: either using O(n) directly, or in two steps, first using SO(n) and then discarding the pseudoscalar states as in Ref. 3. In any case, however, $\mathscr{S}\!p_c(2d, R)$ is the dynamical group of collective states as mentioned before.

As in the one-dimensional case, it is possible to find three explicit ways of carrying out the projection onto the collective subspace. First by considering the A-particle states in the Schrödinger representation, we get a Hilbert space $\mathscr{H}$, whose collective supspace is denoted by $\mathscr{H}_c$. The projection onto $\mathscr{H}_c$ is carried out by using the Dzublik–Zickendraht transformation in d dimensions, which can be written as[2]

$$x_{is}=\sum_{j=1}^{d}\rho_j D^{1}_{ji}(\theta_1,\ldots,\theta_q) D^{1}_{n-d+j,s}(\alpha_1,\ldots,\alpha_r), \tag{8.15}$$

where $\rho_1^2,\ldots,\rho_d^2$ are connected with the principal moments of inertia of the A-body system, $\theta_1,\ldots,\theta_q$ are the $q=\frac{1}{2}d(d-1)$ Euler angles taking us from the frame of reference fixed in the body to the one fixed in space, and $\alpha_1,\ldots,\alpha_r$ are the remaining $r=d[n-\frac{1}{2}(d+1)]$ noncollective variables. In Eq. (8.15), the matrices $\|D^{1}_{ji}(\theta_1,\ldots,\theta_q)\|$ and $\|D^{1}_{ts}(\alpha)\|$ are $d\times d$ and $n\times n$ matrices defining the IR characterized by 1 of the $\mathscr{S}\mathscr{O}(d)$ and SO(n) groups, respectively. In writing the transformations, we have implicitly assumed that we are in the general case where $n>d$, for which O(n) may be replaced by SO(n). For the three borderline cases for which $n\leqslant d$, one should use an IR matrix of O(n) instead of SO(n), and we refer the reader to Ref. 2 for the corresponding detailed expression of the Dzublik–Zickendraht transformation. When realized in $\mathscr{H}_c$, the basic collective states (8.10) are represented by square-integrable functions $\phi_{N_{11}\ldots N_{dd}}(\rho_1,\ldots,\rho_d;\theta_1,\ldots,\theta_q)$, and the generators of their dynamical group depend upon $\rho_1,\ldots,\rho_d$, $\theta_1,\ldots,\theta_q$, $\partial/\partial\rho_1,\ldots,\partial/\partial\rho_d$, $\partial/\partial\theta_1,\ldots,\partial/\partial\theta_q$. In the two-dimensional case, one should find the results of Refs. 3 and 4 again.

Second we can also represent the A-particle states in a Bargmann Hilbert space $\mathscr{F}$ of entire analytic functions in dn complex variables z_{is}, where i $=1,\ldots,d$, and $s=1,\ldots$,n. A basis of that space consists of the functions

$$\bar{\Psi}_{\mathscr{N}_{11}\ldots\mathscr{N}_{dn}}(z_{11},\ldots,z_{dn})=\prod_{i=1}^{d}\prod_{s=1}^{n}[(\mathscr{N}_{is}!)^{-1/2}z_{is}^{\mathscr{N}_{is}}],$$

$$\mathscr{N}_{11},\ldots,\mathscr{N}_{dn}=0,1,\ldots\,. \tag{8.16}$$

The boson creation and annihilation operators η_{is} and ξ_{is} are represented in $\mathscr{F}$ by z_{is} and $\partial/\partial z_{is}$, respectively. Consequently, the generators of $\mathscr{S}\!p(2d, R)$ are realized in $\mathscr{F}$ by the operators

$$\begin{aligned}
\mathscr{D}_{ij}^{\dagger}&=\sum_{s=1}^{n}z_{is}z_{js}, && i\leqslant j=1,\ldots,d,\\
\mathscr{D}_{ij}&=\sum_{s=1}^{n}\frac{\partial^2}{\partial z_{is}\partial z_{js}}, && i\leqslant j=1,\ldots,d,\\
\mathscr{E}_{ij}&=\sum_{s=1}^{n}z_{is}\frac{\partial}{\partial z_{js}}+\frac{n}{2}\delta_{ij}, && i,j=1,\ldots,d.
\end{aligned} \tag{8.17}$$

Complex collective variables can be easily constructed from the dn variables z_{is} by noting that the latter can be grouped into d vectors with respect to O(n), corresponding to $i=1,\ldots,d$. The complex collective variables are therefore the $\nu=\frac{1}{2}d(d+1)$ scalar products of those d vectors,

$$w_{ij}=\sum_{s=1}^{n}z_{is}z_{js}, \quad i\leqslant j=1,\ldots,d. \tag{8.18}$$

The analytic functions in these ν variables form the collective subspace $\mathscr{F}_c$ of $\mathscr{F}$, which may be equipped with a scalar product making it into a generalized Barut Hilbert space. The functions

$$\bar{\phi}_{N_{11}\ldots N_{dd}}(w_{11},\ldots,w_{dd})\propto\prod_{i\leqslant j=1}^{d}w_{ij}^{N_{ij}}, \tag{8.19}$$

representing the states (8.10), form a basis of $\mathscr{F}_c$.

The realization of the dynamical group $\mathscr{S}\!p_c(2d, R)$ in $\mathscr{F}_c$ is also very easy to obtain. For that purpose, we have to retain in Eq. (8.17) only the part depending upon $w_{11},\ldots,w_{dd}$, $\partial/\partial w_{11},\ldots,\partial/\partial w_{dd}$. By proceeding in that way, we straightforwardly get

$$\mathscr{D}_{ij}^{c\dagger}=w_{ij},$$

$$\mathscr{D}_{ij}^{c}=\sum_{kl}(1+\delta_{ki})(1+\delta_{lj})w_{kl}\frac{\partial^2}{\partial w_{ik}\partial w_{jl}}+n(1+\delta_{ij})\frac{\partial}{\partial w_{ij}}, \tag{8.20}$$

and

$$\mathscr{E}_{ij}^{c}=\sum_{k}(1+\delta_{kj})w_{ik}\frac{\partial}{\partial w_{jk}}+\frac{n}{2}\delta_{ij}.$$

As in the one-dimensional case, we could establish a unitary mapping between $\mathscr{H}_c$ and $\mathscr{F}_c$ and define coherent collective states in $\mathscr{F}_c$.

Finally, we can also implement the projection onto the space of collective states by considering a generalized Holstein–Primakoff representation of their dynamical group. That representation establishes a connection between the basic collective states $|\phi_{N_{11}\ldots N_{dd}}\rangle$ and the eigenstates of a ν-dimensional harmonic oscillator, where $\nu=\frac{1}{2}d(d+1)$, by introducing ν boson creation and annihilation operators, $a_\alpha^{\dagger}$ and a_α, $\alpha=1,\ldots,\nu$, in terms of which the generators of

$\mathscr{S}\!\!/\!\!\!h_c(2d, R)$ can be expressed. The $a_\alpha^\dagger$ and a_α operators can be used to define the generators

$$\mathfrak{C}_{\alpha\beta} = a_\alpha^\dagger a_\beta, \quad \alpha, \beta = 1,...\nu, \tag{8.21}$$

of the $\mathfrak{U}(\nu)$ symmetry group of the ν-dimensional harmonic oscillator. Inversion of the generalized Holstein–Primakoff representation leads to the expression of the operators $\mathfrak{C}_{\alpha\beta}$ in terms of the generators of the $\mathscr{S}\!\!/\!\!\!h_c(2d, R)$ group. Actually the generalized Holstein–Primakoff representation of the group Sp(4, R) has been explicitly worked out by Mlodinow and Papanicolaou in another context.[22] Their results can be directly used to study the link between $\mathscr{S}\!\!/\!\!\!h_c(4, R)$ and the $\mathfrak{U}(3)$ group of the IBM in two dimensions. In a similar way we could study the relation between $\mathscr{S}\!\!/\!\!\!h_c(6, R)$ and the $\mathfrak{U}(6)$ group of the IBM in three dimensions.

We plan to implement in subsequent publications the step briefly described in the present section. Another point of interest to be considered later is the projection of the collective Hamiltonian from an arbitrary A-nucleon Hamiltonian. As was shown here for the generators of the dynamical group of collective states, that projection might be easier to carry out in the Barut representation than in the usual Schrödinger representation.

As a final point, let us mention that the dynamical group of collective states, as derived in the present paper, could be used to establish some connections between the O(n) invariant collective model and the symplectic shell model of collective motion developed by Rosensteel and Rowe.[23] The latter is indeed based upon the $\mathscr{S}\!\!/\!\!\!h(6, R)$ group [Note that the authors of Ref. 23 use the notation Sp(3, R) instead of Sp(6, R)], from which the dynamical group $\mathscr{S}\!\!/\!\!\!h_c(6, R)$ is projected out.

Note added in proof: After completion of the present work, A. O. Barut and M. Moshinsky pointed out to us that the Holstein-Primakoff representation of Sp(2, R) was already implicitly contained in the following two references: A. O. Barut, Phys. Rev. B **139**, 1433 (1965); P. A. Mello and M. Moshinsky, J. Math. Phys. **16**, 2017 (1975).

[1] G. F. Filippov, Fiz. Elem. Chastits At. Yadra **4**, 992 (1973), [Sov. J. Part. Nucl. **4**, 405 (1974)].
[2] V. Vanagas, The microscopic nuclear theory within the framework of the restricted dynamics in *Lecture Notes in Physics* (Univ. Toronto Press, Toronto, 1977).
[3] E. Chacón, M. Moshinsky, and V. Vanagas, J. Math. Phys. **22**, 605 (1981).
[4] E. Chacón and M. Moshinsky, Kinam **3**, 3 (1981).
[5] A. Arima and F. Iachello, Ann. Phys. (N.Y.) **99**, 253 (1976); **111**, 201 (1978); **123**, 468 (1979).
[6] M. Moshinsky and T. H. Seligman, J. Math. Phys. **22**, 1526 (1981).
[7] A. Ya. Dzublik, V. I. Ovcharenko, A. I. Steshenko, and G. Filippov, Yad. Fiz. **15**, 869 (1972) [Sov. J. Nucl. Phys. **15**, 487 (1972)].
[8] W. Zickendraht, J. Math. Phys. **12**, 1663 (1971).
[9] V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961).
[10] P. Kramer, M. Moshinsky, and T. H. Seligman, *Group Theory and Its Applications*, edited by E. M. Loebl (Academic, New York, 1975), Vol. 3, p. 250.
[11] A. O. Barut and L. Girardello, Commun. Math. Phys. **21**, 41 (1971).
[12] V. Vanagas, "Relationship between the generalized Bohr–Mottelson and the U(6) scheme models."
[13] T. Holstein and H. Primakoff, Phys. Rev. **58**, 1098 (1940).
[14] M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971).
[15] M. Moshinsky, (private communication); see also Ref. 4.
[16] H. Weyl, *The Classical Groups* (Princeton U. P., Princeton, NJ, 1946), p. 52.
[17] M. Moshinsky, T. H. Seligman, and K. B. Wolf, J. Math. Phys. **13**, 901 (1972).
[18] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1980).
[19] K. E. Cahill and R. J. Glauber, Phys. Rev. **177**, 1857 (1969).
[20] G. S. Agarwal and E. Wolf, Phys. Rev. D **2**, 2161 (1970).
[21] L. D. Mlodinow and N. Papanicolaou, Ann. Phys. (N.Y.) **128**, 314 (1980).
[22] L. D. Mlodinow and N. Papanicolaou, Ann. Phys. (N.Y.) **131**, 1 (1981).
[23] G. Rosensteel and D. J. Rowe, Ann. Phys. (N.Y.) **126**, 343 (1980).

The eikonal equation: some results applicable to computer vision

Anna R. Bruss
Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139[a]

(Received 15 September 1981; accepted for publication 20 November 1981)

In this paper we investigate certain first order partial differential equations which formulate the relationship between the light reflected from a surface and its shape. Particular emphasis is given to eikonal equations. Two results are presented. First, we prove that a special type of eikonal equation has only one convex and positive C^2 solution in some neighborhood of a singular point. Using this result, we show that a restricted form of this equation has exactly two solutions. These results have application in scanning electron microscopy.

PACS numbers: 42.10.Fa, 42.30.Va, 42.30.Di

1. MOTIVATION

How much information about the shape of an object can be inferred from its image? We are interested in a special aspect of this question: the *reconstruction problem*, which is to determine the shape of an object from measurements of the light reflected from its surface. Our work is based on Horn's thesis.[1] He formulated a set of conditions (which are discussed in the next section) which lead to a relation between the perceived brightness of a small patch of a surface and its normal vector. This relation, the *image irradiance equation*, is a first order partial differential equation (abbreviated in the following by FOPDE) and each of its solutions determines the shape of an object. The problem of finding solutions to the image irradiance equation is referred to in the literature as the *shape from shading problem*.

We will take two approaches towards finding a solution to the shape from shading problem termed as the *local* and the *global* approach. By the local approach we mean that only a small patch of an image is used to determine the shape of a surface. To the contrary, in the global approach we examine images in which a silhouette can be detected (here we refer to the outline of an image as a silhouette).

Intuitively, it seems clear that by looking at an image in which a silhouette can be identified we should be able to conclude more about the shape of a surface whose image we are analyzing than by just looking at a little patch. We will show that from certain images which contain a silhouette we can uniquely infer the shape of the surface which gives rise to that image. Unfortunately, the global approach is not always satisfactory; there are also many images containing silhouettes which could be the images of infinitely many different surfaces. There are also infinitely many surfaces which locally look the same. So we will determine conditions under which the global approach is better than the local approach. Notwithstanding, one can sometimes draw interesting conclusions about the shape of surfaces which give rise to the same image by just looking at a small patch of this image.

The local approach is taken to an extreme when we pose the following question: What can be deduced about the shape of a surface from so-called *singular points* of an image irradiance equation? At these points the surface normal to all solutions to such an equation is uniquely determined by the brightness there. We investigate the above stated question for a certain class of image irradiance equations, the so-called eikonal equations, which describe a variety of physical phenomena. For instance, experimental data suggest that the flux of secondary electrons in a scanning electron microscope can be described by an eikonal equation.[2] By using these secondary electrons to modulate the appropriate devices, an image of a surface is created by the microscope. Such an image exhibits shading (Ref. 1) and therefore to determine the shape of a surface from its image one effectively has to solve an eikonal equation. In the case of eikonal equations, we show that the absolute value of the Gaussian curvature at a singular point of all surfaces which give rise to a particular image, is the same. Furthermore, assuming that the surface is convex at a singular point, we show that its shape can be uniquely determined in some neighborhood of such a point from the image intensities alone.

The other aspect of the shape from shading problem which we explore is its solution when the image contains a *b-silhouette* (which is defined below). In this case a global approach is taken. Let us first define the *bounding contour* of a surface: a point P is on the bounding contour if the line connecting the viewer and P *grazes* the surface (i.e., if this line lies in the tangent plane of P). Furthermore we assume that no two parts of a surface obscure each other, i.e., we assume that the bounding contour is not an occluding contour. The image (assuming orthographic projection) of a bounding contour will be called the *b-silhouette*. The surface normal at a point on a bounding contour is parallel to the normal vector to the *b*-silhouette and both vectors lie in the same plane. Thus, some or all of the first order partial derivatives of the function defining the surface are infinite for points on the bounding contour (we will say that some components of the surface gradient are *singular* along a curve).

For example, the bounding contour of a hemisphere lying on a plane is a circle. Consider a Lambertian surface, which has the property that each surface patch appears equally bright from all viewing directions. If we look at a Lambertian hemisphere such that the viewer and the light source are at the same point, its *b*-silhouette can be determined from its image. In this case the image irradiance equation describing the imaging situation is singular and *all* the surfaces which satisfy such an equation have a bounding

[a] Present address: IBM T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598.

 0022-2488/82/050890-07$02.50

contour. (In Sec. 4 we extend the notion of a singular function to equations.)

As the image irradiance equation is a FOPDE, it does not, in general, provide enough information to solve the reconstruction problem uniquely. It remained an open question whether there are *any* imaging situations for which every surface gives rise to a different image. As indicated before, taking the global approach towards finding a solution to the shape from shading problem allows us to answer this question affirmatively by proving a uniqueness result for a class of eikonal equations.

2. THE SHAPE FROM SHADING PROBLEM

There are basically three components to the shape from shading problem which must be taken into account. They are the light source, the object, and the camera as depicted in Fig. 1, which is taken from Ref. 3, p. 32, and are termed an *imaging configuration*. Henceforth, we will assume that an image of a surface is produced by a camera. The shading of such an image can be explained as follows. The exposure of film in a camera (for fixed shutter speed) is proportional to *image irradiance*, the light flux per unit area falling on the image plane. Similarly, grey levels measured in an electronic imaging device are quantized measurements of image irradiance. It can be shown that image irradiance in turn is proportional to *scene radiance*, the light flux emitted by the object per unit projected surface area per unit solid angle.[4] The factor of proportionality depends on details of the optical system, including the effective f-number. Scene radiance depends on the

- surface material and its microstructure,
- the incident light flux, and
- the orientation of the surface.

Now we want to relate the shape of a surface to the shading of its image. Consider a viewer-oriented coordinate system with the viewer located far above the surface on the z-axis. If the objects imaged are small compared to their distance from the viewer, one can approximate the imaging situation by an orthographic projection:

$$\tilde{x} = xf/z_0, \quad \tilde{y} = yf/z_0, \tag{2.1}$$

where $(\tilde{x},\tilde{y})$ are the coordinates of the image of a point (x,y,z) made with a system of effective focal length f, and the viewer

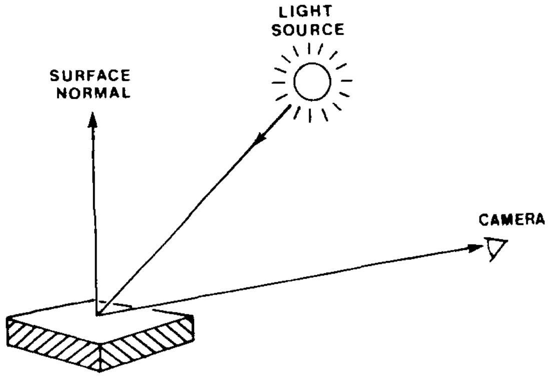


FIG. 1. Imaging configuration.

is at a distance z_0 above the origin. We assume that $(x^2+y^2+z^2) \ll z_0^2$. For simplicity and without loss of generality it is also assumed that the viewing direction coincides with the z-axis.

The orientation of a patch of a surface can be specified by its gradient $(p,q,-1)$, where p and q are the first order partial derivatives of z with respect to x and y. For a given surface material and known incident light flux, scene radiance will depend only on surface gradient. The function which describes this dependence, $R(p,q)$, is called the reflectance map.

Recall that image irradiance and scene radiance are proportional and that we assume orthographic projection. If $E(x,y)$ is the observed image irradiance at the point $(\tilde{x},\tilde{y})$ in the image, then

$$R(p,q) = E(x,y), \tag{2.2}$$

where (p,q) are two components of the gradient at the corresponding point on the object being imaged. This equation is called the *image irradiance equation*. Unless otherwise stated, we will assume that the functions $R(p,q)$ and $E(x,y)$ are C^1. We will refer to image irradiance equations of the form

$$p^2+q^2 = E(x,y) \tag{2.3}$$

as *eikonal equations*.

In general, at a point (x,y) in the image plane, the gradient is constrained by an image irradiance equation to a one-parameter manifold. Only at so-called singular points does the measured image intensity uniquely define the surface normal there and we will investigate such points further. We now give some relevant definitions.

Definition: Let $R(p,q) = E(x,y)$ be an image irradiance equation. Then a point (x_0,y_0,p_0,q_0) is a *critical element* if (x_0,y_0) is a stationary point of $E(x,y)$ and (p_0,q_0) is a stationary point of $R(p,q)$.
The point (x_0,y_0,p_0,q_0) is a *critical point* of the image irradiance equation if it is a critical element and if the values (x_0,y_0,p_0,q_0) satisfy the image irradiance equation.
The point (x_0,y_0,p_0,q_0) is a *singular point* if it is a critical point for which the values (p_0,q_0) are uniquely determined by the values (x_0,y_0).
A point P is an *isolated* critical element if in some neighborhood of it, it is the only stationary point of $E(x,y)$ and $R(p,q)$. Isolated critical (singular) points can be defined similarly.

3. SINGULAR POINTS

The question addressed in this section is: How much information about the shape of a surface can one obtain from a singular point of an eikonal equation? To obtain our results we will impose some technical conditions upon $E(x,y)$ and thus define *constrained* eikonal equations.

Definition: An eikonal equation $p^2+q^2 = E(x,y)$ is *constrained* if $E(x,y)$ is a C^3 function satisfying the following conditions in some neighborhood of the point (x_0,y_0):

$$\begin{aligned}
&(1)\ (x_0,y_0) \text{ is a stationary point of } E(x,y),\\
&(2)\ E(x_0,y_0) = 0,\\
&(3)\ E(x,y) > 0 \text{ for } (x,y) \neq (x_0,y_0),\\
&(4)\ E(x,y) \text{ vanishes precisely to second order at } (x_0,y_0).
\end{aligned} \tag{3.1}$$

Let us discuss these conditions a bit further. Since the reflectance map of an eikonal equation is $R(p,q) = p^2 + q^2$, the point $P = (x, y, p,q) = (x_0, y_0,0,0)$ is a critical point of a constrained eikonal equation, whence it follows from conditions 2 and 3 that P is an *isolated* singular point. By using a suitable linear transformation we may assume, without loss of generality, that the point (0,0) is the stationary point of $E(x,y)$. We will denote the (limited) Taylor series expansion of $E(x,y)$ as

$$E(x,y) = \alpha x^2 + \beta xy + \gamma y^2 + 0((|x| + |y|)^2). \tag{3.2}$$

Since $E(x,y)$ is assumed to be positive near the origin,

$$\alpha x^2 + \beta xy + \gamma y^2 > 0 \quad \text{for} \quad (x,y) \neq (0,0) \tag{3.3}$$

defines a positive bilinear form. Thus the subsequent inequality (Ref. 5, p. 182) holds:

$$\alpha\gamma - \beta^2/4 > 0. \tag{3.4}$$

Moreover, α and γ must be positive.

The first result which we prove in this section is formulated in the following theorem.

Theorem: Let $p^2 + q^2 = E(x,y)$ be a constrained eikonal equation. Then there exists a unique locally convex solution in some neighborhood of the singular point.

The theorem can be expressed in other words as: If $z = z(x,y)$ defines one locally convex solution, then $\tilde{z} = -z(x,y)$ defines the other. Hence this result can be viewed as a uniqueness result modulo the concave/convex ambiguity. To simplify subsequent discussions, we will say that $z(x,y)$ is a locally convex *solution* to a constrained eikonal equation if it satisfies the following *positivity* conditions in some neighborhood of the origin:

$$\begin{aligned} &(1)\ z(0,0) = 0, \\ &(2)\ z(x,y) \in C^2, \\ &(3)\ z(x,y) \geqslant 0. \end{aligned} \tag{3.5}$$

Before proving this theorem we introduce some relevant concepts and show a lemma.

Definition: Let $p^2 + q^2 = E(x,y)$ be a constrained eikonal equation and let ξ denote the four-tuple (x, y, p,q). In some neighborhood of the singular point, the characteristic equations can be written as

$$\frac{d\xi}{dt} = A\xi + G(\xi), \tag{3.6}$$

where A is the four by four matrix

$$A = \begin{pmatrix} 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \\ 2\alpha & \beta & 0 & 0 \\ \beta & 2\gamma & 0 & 0 \end{pmatrix} \tag{3.7}$$

and where G has the following properties:

$$\begin{aligned} &(1)\ G(\xi) \in C^2, \\ &(2)\ G(0) = 0, \\ &(3)\ \frac{\partial G}{\partial \xi}(0) = 0. \end{aligned} \tag{3.8}$$

Every solution $\xi = \xi(t)$ to (3.6) is called an *orbit*. The equation

$$\frac{d\xi}{dt} = A\xi \tag{3.9}$$

is called the *linearized* characteristic equation. An orbit $\xi = \xi(t)$ is *quasiradial* if

$$\lim_{t \to \pm\infty} \xi(t) = 0. \tag{3.10}$$

Lemma: Let $p^2 + q^2 = E(x,y)$ be a constrained eikonal equation. If a locally convex solution exists in some neighborhood of the singular point, then it is swept out by quasiradial characteristic curves.

Proof: Suppose a locally convex solution $z = z(x,y)$ exists. As $z = z(x,y)$ is assumed to be C^2, we can write p and q in some neighborhood of the singular point as

$$\begin{aligned} p &= a_{11}x + a_{12}y + o(|x| + |y|), \\ q &= a_{12}x + a_{22}y + o(|x| + |y|), \end{aligned} \tag{3.11}$$

where the a_{ij}'s, for $i, j = 1,2$, are constants. Since the origin is a singular point, p and q have no constant terms. Note that the Gaussian curvature K of $z = z(x,y)$ at the origin is given by

$$K = a_{11}a_{22} - a_{12}^2. \tag{3.12}$$

Substituting the expressions (3.11) into the first two characteristic equations of a constrained eikonal equation gives

$$\begin{aligned} \frac{dx}{dt} &= 2(a_{11}x + a_{12}y) + o(|x| + |y|), \\ \frac{dy}{dt} &= 2(a_{12}x + a_{22}y) + o(|x| + |y|). \end{aligned} \tag{3.13}$$

Using a standard theorem from the theory of ordinary differential equations, the so-called node theorem, we deduce that the characteristic curves are quasiradial if and only if both eigenvalues of the linearized equations have the same sign (Ref. 6, p. 213). A simple calculation shows that this is the case only when $K > 0$. Assuming that $K > 0$, the sign of the eigenvalues is the same as the sign of a_{11} or equivalently of a_{22}. ■

Proof of theorem: We have to show that a locally convex solution to a constrained eikonal equation exists and is unique. This is achieved by showing that the unstable manifold is the locally convex solution. It follows from the previous lemma that if such a solution exists, it is swept out by quasiradial characteristic curves. To prove its existence, we investigate the linearized characteristic equations (3.9) of a constrained eikonal equation. An easy calculation shows that the matrix A has two positive real eigenvalues and two negative real eigenvalues. Thus, we can apply the stable manifold theorem which states that there exist exactly two C^2 manifolds, each of which is swept out by quasiradial characteristic curves (Ref. 7, p. 527 and Ref. 6, p. 242). Hence we can deduce that a locally convex solution exists and is unique. From the node theorem, we get that the solution $z = z(x,y)$ satisfying the positivity condition is the unstable manifold, whereas the stable manifold is the surface defined by $\tilde{z} = -z(x,y)$. ■

In the case where $E(x,y)$ is C^∞ we can compute the locally convex formal power series solution as shown in the following lemma.

Lemma: Let $p^2 + q^2 = E(x,y)$ be a constrained eikonal equation where $E(x,y)$ is C^∞. Then there exists a unique, locally convex formal power series solution to this equation in some neighborhood of the singular point.

Proof: (outline) Equating the appropriate terms in of a formal power solution we obtain

- an equation for the quadratic terms and
- a recurrence relation for each of the higher order terms.

First we prove that there is a unique solution to the equation for the quadratic terms if we impose the constraints that the formal power series solution be positive and convex. The next step is to determine the higher order terms which is done by inductively solving the recurrence relation. If the quadratic terms have been determined such that the formal power series solution is convex, each step of this induction can be carried out uniquely. The details of this proof can be found in Ref. 8.■

In the case where $E(x,y)$ is analytic we can show the following lemma.

Lemma: Let $p^2 + q^2 = E(x,y)$ be a constrained eikonal equation where $E(x,y)$ is analytic. Then its formal power series solution is the solution to the equation.

Proof: A version of the stable manifold theorem proves that if $E(x,y)$ is analytic, then the stable (unstable) manifold is analytic (Ref. 9, p. 330). The lemma follows.■

The second main result of this section is formulated in the following theorem.

Theorem: Let $p^2 + q^2 = E(x,y)$ be a constrained eikonal equation. Then at the singular point, the Gaussian curvature of each integral surface has the same absolute value and is determined by the (limited) Taylor series expansion of $E(x,y)$ at that singular point.

Proof: Recall that the curvature at the origin, denoted by K, is

$$K = a_{11}a_{22} - a_{12}^2. \tag{3.14}$$

Using Eqs. (3.11) in an eikonal equation, we derive an expression for this curvature in terms of α, β, and γ:

$$|K| = \tfrac{1}{2}\,|(4\alpha\gamma - \beta^2)^{1/2}|. \tag{3.15}$$

The details of these calculations can be found in Ref. 8.■

4. *B*-SILHOUETTES

As discussed, our goal is to find sufficient constraints such that an image can be interpreted in a unique way when its image irradiance equation is known. We now investigate whether the knowledge of *b*-silhouettes can be used to interpret an image and thus study *singular* image irradiance equations:

Definition: Let $R(p,q) = E(x,y)$ be an image irradiance equation. It is called singular if there exist finite values for x and y denoted by x_0 and y_0 such that

$$\lim_{\substack{x\to x_0\\ y\to y_0}} E(x,y) = \pm\infty. \tag{4.1}$$

Note that the *b*-silhouette consists of such points (x_0, y_0).

In this section we will identify three constraints upon an image irradiance equation, one upon the reflectance map, one upon the *b*-silhouette, and one upon the function $E(x,y)$. If these constraints hold for some image irradiance equation, exactly one surface defined by a C^2 function which satisfies the equation exists.

4.1. Uniqueness theorem

Let

$$R(p,q) = E(x,y) \tag{4.2}$$

be a singular image irradiance equation. Consider the following constraints upon this equation:

(C1) $R(p,q) = p^2 + q^2$.

(C2) The *b*-silhouette defined by $w(x,y) = 0$ is a closed, smooth curve in the x–y plane. Furthermore, the points (x,y) at which the image irradiance equation is defined lie in the region bounded by this *b*-silhouette.

(C3) The function $E(x,y)$ has exactly one stationary point (x_0,y_0) and satisfies the following conditions in some neighborhood of (x_0,y_0): $E(x_0,y_0) = 0$, $E(x,y) > 0$ for $(x,y) \neq (x_0,y_0)$ and $E(x,y)$ vanishes precisely to second order at (x_0,y_0).

Uniqueness Theorem: Let $R(p,q) = E(x,y)$ be an image irradiance equation for which constraints C1, C2, and C3 hold and suppose a C^2 integral surface defined by $z = z(x,y)$ of this equation exists. Then the only other solution to the equation is $\tilde{z} = -z(x,y)$.

Proof: Let $R(p,q) = E(x,y)$ be a fixed image irradiance equation for which constraints C1, C2, and C3 hold. First note that the point $P = (x, y, p, q) = (x_0, y_0, 0, 0)$ is an isolated singular point of the image irradiance equation. There are then two observations which allow us to prove the theorem. First, as the *b*-silhouette is a closed curve, an integral surface of the equation has to be compact. Second, from constraints C1 and C3 we can deduce that such a surface is convex at the singular point, which allows us to apply results of the previous section.

Suppose $z = z(x,y)$ defines a C^2 integral surface of an image irradiance equation. Then from C2 we may infer that z defines a compact surface. Note also that $z(x,y)$ is defined at every point (x,y) which lies within or on the *b*-silhouette and therefore has a bounding contour. Thus there exists a point $\tilde{P}$ at which z has an extremum from which it follows that the tangent plane at $\tilde{P}$ is parallel to the x–y plane. From condition C3 we can deduce that there exists exactly one such point (x_0,y_0,z_0), i.e., $\tilde{P} = (x_0,y_0,z_0)$. By the assumption on $E(x,y)$, z is either convex at $\tilde{P}$ or has a saddle point. Since $\tilde{P}$ is the point where the surface has maximal (minimal) height, z must be convex there.

In the previous section we proved that if an image irradiance equation satisfies C1 and C3, there exists exactly one positive and one negative convex solution denoted by z_p and $z_n = -z_p$, respectively. Thus there are exactly two integral surfaces $z = z(x,y)$ and $\tilde{z} = -z(x,y)$.■

By using transformation methods, we can enlarge the class of singular image irradiance equations for which the uniqueness theorem holds. Let

$$f(Ap^2 + 2Bpq + Cq^2 + 2Dp + 2Eq) = E(x,y) \tag{4.3}$$

by a singular image irradiance equation where f is a bijection and A, B, C, D, and E are real constants such that $\delta > 0$ and $\Delta S < 0$ where δ, Δ, and S are defined in the following equations:

$$\delta = AC - B^2,$$

$$\Delta = \begin{vmatrix} A & B & D \\ B & C & E \\ D & E & 0 \end{vmatrix}, \tag{4.4}$$

$$S = A + C.$$

The constraints upon the constants A, B, C, D, and E in Eq. (4.3) assure that the curves $R(p,q) = c$, for any constant c, are closed. Let the b-silhouette of the equation be a closed and smooth curve. Then (4.3) can be transformed into an image irradiance equation of the form (4.2) for which C2 holds. If, after the transformation, the function $E(x,y)$ satisfies C3, then the uniqueness theorem holds for (4.3).

The next corollary follows directly from the uniqueness theorem in this section. We will abbreviate $(x^2 + y^2)^{1/2}$ by r.

Corollary: Let $p^2 + q^2 = E(r)$ be an image irradiance equation where $E(r)$ satisfies constraints C2 and C3. Suppose a C^2 integral surface of this image irradiance equation exists. Then it is *rotationally symmetric* and can be obtained by integrating $E(r)$. In this case the b-silhouette is a circle.

Proof: First we write the eikonal equation in polar coordinates:

$$z_r^2 + (1/r^2)z_\vartheta^2 = E(r). \tag{4.5}$$

Let $\tilde{z} = \tilde{z}(r)$ define the rotationally symmetric integral surface of the above eikonal equation. Thus $\tilde{z}_\vartheta(r) = 0$ and we can compute both rotationally symmetric solutions by integrating $\pm\sqrt{E(r)}$. It follows from the uniqueness theorem that the image irradiance equation has no other solutions. ■

Note that the above corollary is false if we omit the condition that the image contains a b-silhouette. The integral surfaces of a continuous rotationally symmetric eikonal equation are not themselves necessarily rotationally symmetric.

4.2. Counterexamples

In the previous section we discussed sufficient constraints under which the solution to a singular image irradiance equation is unique. Are these constraints necessary? Although we are not able to answer this question completely, we now shed some light upon it. In particular, we will try to find the class of image irradiance equations for which most likely there is no set of constraints that assure the existence of only one global solution.

Image irradiance equations satisfying the constraints of the uniqueness theorem have closed iso-brightness curves, i.e., the curves $R(p,q) = c$ are closed. Let us now examine singular image irradiance equations whose iso-brightness curves are not closed. One such image irradiance equation is given by

$$p + q = -(x+y)/(1-(x^2+y^2))^{1/2}. \tag{4.6}$$

While constraint C2 holds for (4.6), an image irradiance equation whose reflectance map is $R(p,q) = p + q$ never has a singular point. The general solution to (4.6) is

$$z(x,y) = (1-(x^2+y^2))^{1/2} + w(y-x), \tag{4.7}$$

where w is any C^1 function. Figures 2 and 3 illustrate some solutions to Eq. (4.6).

Another example of an image irradiance equation whose iso-brightness curves are not closed is

$$pq = xy/(1-(x^2+y^2)). \tag{4.8}$$

This equation satisfies constraint C2, i.e., its b-silhouette is a closed and smooth curve. Furthermore the origin is the singular point of (4.8) and $E(x,y)$ vanishes precisely to second order there although $E(x,y)$ is not positive in the neighborhood of (0,0). One of the solutions to (4.8) is the sphere

$$z(x,y) = (1-(x^2+y^2))^{1/2}, \tag{4.9}$$

whereas another surface satisfying (4.8) is

$$z(x,y) = f(t) + x^2 - y^2,$$

where

$$t = 1 - (x^2 + y^2)$$

and

$$f(t) = t(1/4t + 1)^{1/2} + \tfrac{1}{8}[\ln((1/4t+1)^{1/2}+1) - \ln((1/4t+1)^{1/2}-1)]. \tag{4.10}$$

Recall that constraint C2 expressed the fact that the b-silhouette is a closed and smooth curve. We now demonstrate that if the b-silhouette does not obey C2, our uniqueness result does not hold. An example of an equation for which C1 holds, but whose b-silhouette is not a closed curve, is

$$p^2 + q^2 = 1/4x + 1. \tag{4.11}$$

Equation (4.11) does not have a singular point. A solution to this equation is

$$z(x,y) = \sqrt{x} + y, \tag{4.12}$$

which is shown in Fig. 4. Two other solutions to (4.11) are

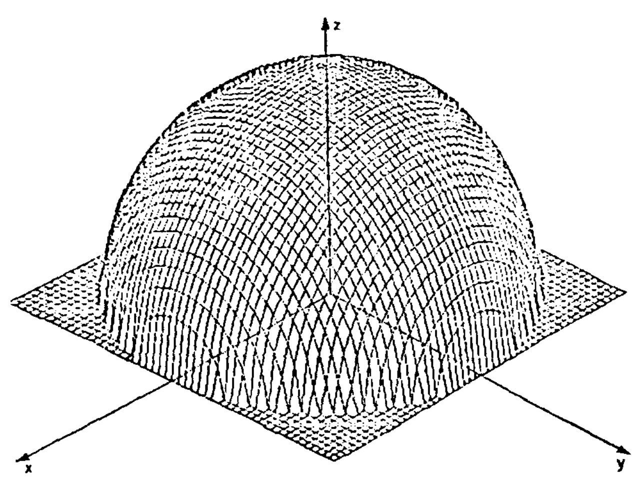


FIG. 2. $z(x,y) = (1-(x^2+y^2))^{1/2}$.

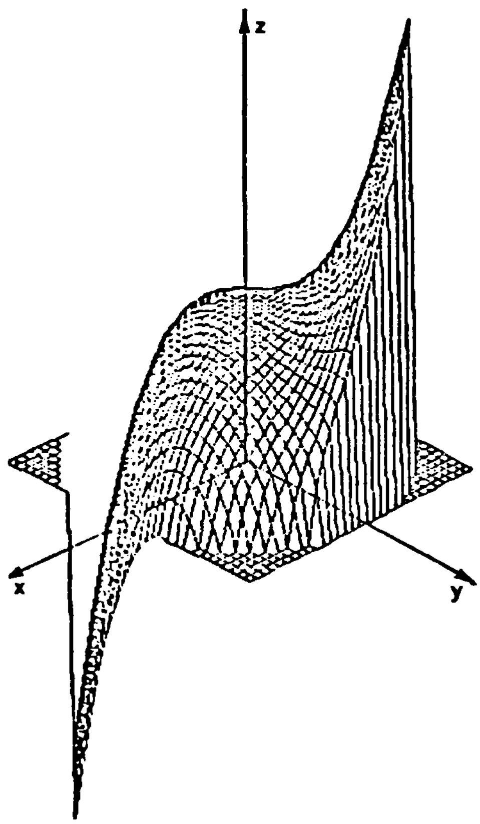


FIG. 3. $z(x,y) = (1 - (x^2 + y^2)^{1/2}) + (y - x)^3$.

$$z(x,y) = x\left(\left(\frac{1}{4x} + 1\right)^{1/2}\right) + \tfrac{1}{8}\left[\ln\left(\left(\frac{1}{4x} + 1\right)^{1/2} + 1\right)\right.$$
$$\left. - \ln\left(\left(\frac{1}{4x} + 1\right)^{1/2} - 1\right)\right],$$
$$z(x,y) = \frac{(x(1 - 8x))^{1/2}}{2} - \frac{1}{4\sqrt{2}} \arctan\left(\left(\frac{1}{8x} - 1\right)^{1/2}\right)$$
$$+ (3)^{1/2} y. \tag{4.13}$$

Only when the b-silhouette is a closed curve can we deduce that a surface is convex at the singular point, an observation which allows us to prove the uniqueness theorem. For the following eikonal equation we give two solutions only one of which is convex:

$$p^2 + q^2 = \frac{x^2 + y^2}{1 - x^2 y^2}. \tag{4.14}$$

Equation (4.14) has a singular point and it satisfies constraints C1 and C3. Two solutions to this equation are

$$z(x,y) = \arcsin(xy),$$
$$z(x,y) = (1 - x^2 y^2)^{1/2} + (x^2 - y^2)/2. \tag{4.15}$$

5. CONCLUSION

In this paper, we have investigated the question of how much information concerning the shape of an object can be deduced from its shaded image. Even assuming that adequate data is available to derive an image irradiance equation is insufficient to solve the reconstruction problem uniquely; in general, for a fixed imaging configuration there are many surfaces which have the same shaded image. Thus our goal has been to identify constraints by which the reconstruction problem can be solved uniquely.

We discussed how singular points of an eikonal equation constrain its possible solutions. In particular, we proved that for any eikonal equation of a certain type there exists a unique (up to translation in the z direction) positive convex surface which satisfies the equation in some neighborhood of a singular point.

However, our ultimate ambition was to answer the following question.

Is there a set of constraints which assure that if an image irradiance equation has a solution, it is unique?

We answered this question affirmatively in Sec. 4. It was shown there that if three constraints on an image irradiance equation are known to hold, the information about the imaging situation and the surface as captured by this equation allow one to reconstruct the shape of the surface uniquely. Furthermore, one can easily check whether an image irradiance equation satisfies these constraints. It is surprising that our uniqueness theorem holds only when the b-silhouette is a closed curve (constraint C2).

In order to evaluate the usefulness of our uniqueness theorem, we need to know which of the commonly arising image irradiance equations actually obey the above mentioned restrictions. In his paper on hill-shading,[10] Horn discusses eighteen different reflectance maps which are applied to solve that problem. Constraint C1 holds for five of those reflectance maps. These equations are of the form

$$R(p,q) = f(p^2 + q^2), \tag{5.1}$$

where f is a bijection. A reflectance map of the form (5.1) describes, for instance, the situation where the object is Lambertian and the light source and the viewer have the same position. In addition, eikonal equations can be used to automatically analyze images taken by a scanning electron microscope.

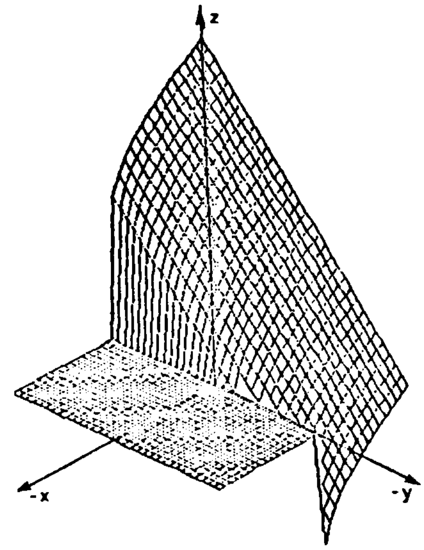


FIG. 4. $z(x,y) = \sqrt{x} + y$.

ACKNOWLEDGMENTS

I would like to thank B. K. P. Horn and R. Melrose for their valuable suggestions and S. M. Trilling for proofreading this paper. This research was supported in part by the Advanced Research Projects Agency of the Department of Defense under contract number N00014-80-C-0505 and in part by the Office of Naval Research under contract number N00014-77-C-0389.

[1]B. K. P. Horn, "Obtaining Shape from Shading Information," in *The Psychology of Computer Vision,* edited by P. H. Winston (McGraw-Hill, New York, 1975), pp. 115–155.

[2]A. B. Laponsky and N. R. Whetten, "Dependence of Secondary Emission from MgO Single Crystals on Angle of Incidence," Phys. Rev. **120**, 3 (1980).

[3]R. J. Woodham, "Reflectance Map Techniques for Analyzing Surface Defects in Metal Casting," TR-457, Artificial Intelligence Laboratory, M.I.T., 1978 (unpublished).

[4]B. K. P. Horn and R. W. Sjoberg, "Calculating the Reflectance Map," Appl. Opt. **18**, 1770–1779 (1979).

[5]I. N. Bronstein and K. A. Semendjajew, *Taschenbuch der Mathematik* (Teubner, Leipzig, 1975).

[6]P. Hartman, *Ordinary Differential Equations* (Wiley, New York, 1964).

[7]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin–Cummings, New York, 1980).

[8]A. R. Bruss, *The Image Irradiance Equation: Its Solution and Application*, A. I. Technical Report 623, Artificial Intelligence Laboratory, M.I.T., 1981 (unpublished).

[9]E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955), pp. 371–377.

[10]B. K. P. Horn, "Hill-Shading and the Reflectance Map," in *Proceedings: Image Understanding Workshop*, April, 1979.

Biconical receiving antenna

Dale M. Grimes
Electrical Engineering Department, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 3 September 1980; accepted for publication 30 October 1981)

The modal solution to a biconical receiving antenna with arbitrary arm angles and lengths is presented in the form of sums over special functions; sums which are exact in the sense that no electromagnetic approximations are made. The antenna load is confined within a sphere of small radius centered at the apices of the cones. The analytical formulation is presented, along with solutions for ψ very small and very large, and certain selected numerical data. The general solution permits calculation of all detailed fields near and far, both receiving and transmitting current modes on the antenna arms and caps and the power and the momentum absorbed by the antenna from an incoming plane wave. It is shown that the receiving current modes are necessary for electromagnetic momentum to be conserved during power reception. Detailed calculation of comparative receiving and transmitting admittances confirms that they are identical, as predicted by the reciprocity theorem. The radiation patterns, however, for retransmission during reception and for transmission differ.

PACS numbers: 84.40.Af, 02.30.Lt

I. INTRODUCTION

The receiving antenna parameters of radiation impedance and radiation pattern commonly are found by finding, first, the properties the antenna would have were it transmitting and, second, by using reciprocity to show that these properties carry over to receiving. Although there is no question of the validity of this approach, a detailed, modal, and mechanistic view of the functioning of a receiving antenna may still be of interest. For example, although integral expressions give quite accurate results for linear, cylindrical receiving antennas,[1–6] still, exact solutions are available only for vanishingly small or limitlessly large ones, in contrast with the solution presented here which is electromagnetically exact for a biconical receiving antenna of arbitrary arm length and cone angle. Results are given as sums over an infinite series of special spherical functions, with certain ones selected for numerical analysis.

This work is a direct extension of previous work on biconical transmitting antennas.[6–8] For the first time it provides, for a receiving antenna of arbitrary size, expressions for all fields near and far, surface currents on the antenna arms and caps, power and momentum transfer between field and antenna, and a direct confirmation of the reciprocity theorem. As a specific example of its potential use, even vanishingly small dipolar antennas, when properly loaded, absorb power from a fully directed, correctly polarized wave, and presumably time reversal of all currents would result in the same dipole transmitting a fully directed wave. The solution presented here permits the calculation of the currents needed for such a result to be obtained.

The antenna analyzed is centered at the origin of coordinates, and is spherically capped, symmetric, and biconical with arms of length a and half-angle ψ, see Fig. 1. The center load is a uniform, spherical one centered on the origin and of radius b, which is a small fraction of a wavelength, i.e., $kb \ll 1$.

Space surrounding the antenna is divided into three regions, see Fig. 1. Region I is the spherical load, region II is a concentric spherical volume of radius a, but excluding region I and the antenna arms. Region III has radius $r > a$, and is empty.

It is most convenient to consider an incident plane wave of unit electric field intensity, polarized with its electric field collinear with the z axis and the axis of the antenna to be analyzed. We pick its magnetic field to be collinear with the x axis. The wave travels in the positive y direction and is scattered by the antenna. We seek to find the absorbed power, the surface currents, the scattered field, and the force on

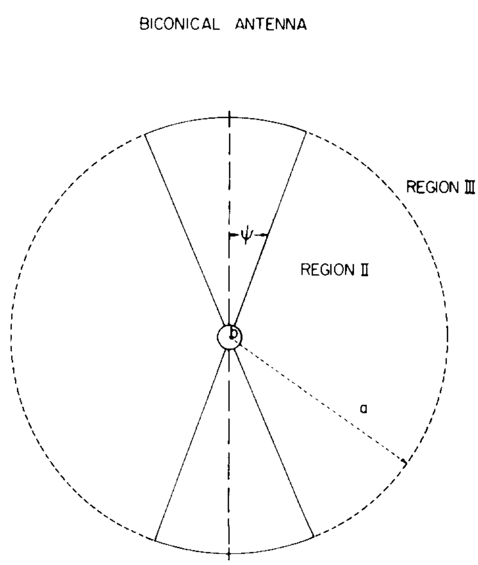


FIG. 1. Representation of a biconical receiving antenna with a small, spherical load of radius b, perfectly conducting arms of length a, and half-angle ψ. Region II contains the open area with radius $< a$, Region III is all space with radius $> a$.

 0022-2488/82/050897-18$02.50

the antenna as functions of the configurative parameters a and ψ and the load impedance.

II. THE FIELDS

A. Plane wave

The incoming wave has phase $\exp[i(\omega t - ky)]$. In spherical coordinates and with suppressed time dependence the spatial phase dependence is $\exp(-i\sigma\sin\theta\sin\phi)$, where $\sigma = kr$, $\theta =$ polar angle measured from the z axis, and $\phi =$ azimuth angle measured from the x axis. Expanding the phase factor in terms of spherical Bessel functions of integer order $j_l(\sigma)$, and associated Legendre polynomials $P_l^m(\cos\theta)$, it is shown in Appendix A [see (A1)–(A14)] that

$$\begin{aligned}\exp(-i\sigma\sin\theta\sin\phi) &= \sum_{\substack{2\\ le}}^{\infty}\sum_{\substack{0\\ me}}^{l} l(l+1)\frac{C_{lm}}{m} j_l(\sigma) P_l^m \cos m\phi \\ &\quad - i\sum_{\substack{1\\ lo}}^{\infty}\sum_{\substack{1\\ mo}}^{l} l(l+1)\frac{C_{lm}}{m} j_l(\sigma) P_l^m \sin m\phi,\end{aligned} \tag{1}$$

where

$$C_{lm} = \frac{2m(2l+1)(l-m)!U(m)\delta(l+m,2q)}{2^l l(l+1)[l+m/2]![l-m/2]!}, \tag{2}$$

where $U(m)$ is the step function

$$\begin{aligned} U(m) &= 0 \quad \text{for } m<0,\\ &= \tfrac{1}{2} \quad \text{for } m=0,\\ &= 1 \quad \text{for } m>0,\end{aligned} \tag{3}$$

where δ is Kronecker's delta function as defined in (4), and q is any integer,

$$\begin{aligned}\delta(l+m,2q) &= 1 \quad \text{for } l+m \text{ even},\\ &= 0 \quad \text{for } l+m \text{ odd},\end{aligned} \tag{4}$$

and subscripts o and e below the summation indicate, respectively, only odd or even values are to be taken for the summation index indicated. This redundant notation assists in later integral evaluations. Equations (1)–(4) lead to the radial components

$$\begin{aligned} E_r &= \sum_{\substack{1\\ lo}}^{\infty}\sum_{\substack{0\\ me}}^{l-1} l(l+1)D_{lm}\frac{j_l(\sigma)}{\sigma}P_l^m\cos m\phi \\ &\quad - i\sum_{\substack{2\\ le}}^{\infty}\sum_{\substack{1\\ mo}}^{l-1} l(l+1)D_{lm}\frac{j_l(\sigma)}{\sigma}P_l^m\sin m\phi \end{aligned} \tag{5}$$

$$\begin{aligned} \eta H_r &= \sum_{\substack{1\\ lo}}^{\infty}\sum_{\substack{1\\ mo}}^{l} l(l+1)C_{lm}\frac{j_l(\sigma)}{\sigma}P_l^m\cos m\phi \\ &\quad - i\sum_{\substack{2\\ le}}^{\infty}\sum_{\substack{2\\ me}}^{l} l(l+1)C_{lm}\frac{j_l(\sigma)}{\sigma}P_l^m\sin m\phi \end{aligned} \tag{6}$$

where $\eta =$ wave impedance and it is shown in Appendix A (A15)–(A34), that

$$D_{lm} = \frac{4(2l+1)}{2^l l(l+1)}\,\frac{(l-m)!U(m)\ \delta(l+m,2q+1)}{[(l-1+m)/2]![(l-1-m)/2]!} \tag{7}$$

and

$$\begin{aligned}\delta(l+m,2q+1) &= 1 \quad \text{for } l+m \text{ odd},\\ &= 0 \quad \text{for } l+m \text{ even}.\end{aligned} \tag{8}$$

The angular fields follow directly from the radial ones, and complete the set for a y-directed, z-polarized plane wave.

B. Scattered field

The formal description of the scattered wave has two differences from that of the plane wave: the spherical Bessel functions $j_l(\sigma)$ are replaced by spherical Hankel functions of the second kind $h_l(\sigma)$, where

$$h_l(\sigma) = j_l(\sigma) - iy_l(\sigma), \tag{9}$$

where $y_l(\sigma)$ are spherical Neumann functions and the constant field parameters may be different from those of the plane wave. It is convenient for what follows to define two sets of constants α_{lm} and β_{lm}. In these terms, the radial components of the scattered field are described by (10) and (11).

$$\begin{aligned} E_r &= \sum_{\substack{1\\ lo}}^{\infty}\sum_{\substack{0\\ me}}^{l-1} l(l+1)D_{lm}\beta_{lm}\frac{h_l(\sigma)}{\sigma}P_l^m\cos m\phi \\ &\quad - i\sum_{\substack{2\\ le}}^{\infty}\sum_{\substack{1\\ mo}}^{l-1} l(l+1)D_{lm}\beta_{lm}\frac{h_l(\sigma)}{\sigma}P_l^m\sin m\phi, \end{aligned} \tag{10}$$

$$\begin{aligned} \eta H_r &= \sum_{\substack{1\\ lo}}^{\infty}\sum_{\substack{1\\ mo}}^{l} l(l+1)C_{lm}\alpha_{lm}\frac{h_l(\sigma)}{\sigma}P_l^m\cos m\phi \\ &\quad - i\sum_{\substack{2\\ le}}^{\infty}\sum_{\substack{2\\ me}}^{l} l(l+1)C_{lm}\alpha_{lm}\frac{h_l(\sigma)}{\sigma}P_l^m\sin m\phi. \end{aligned} \tag{11}$$

C. Total field, region III

For the total field, the radial components are the sum of (5) and (10) and of (6) and (11). For the angular components, it is convenient to define two additional sets of radial functions $J_l(\sigma)$ and $H_l(\sigma)$, where

$$J_l(\sigma) = \frac{1}{\sigma}\frac{d}{d\sigma}[\sigma j_l(\sigma)] \tag{12}$$

and

$$H_l(\sigma) = \frac{1}{\sigma}\frac{d}{d\sigma}[\sigma h_l(\sigma)]. \tag{13}$$

In these terms, the total fields in region III are described by the following equations.

$$E_\theta = \sum_{\substack{1\\ lo}}^{\infty}\sum_{\substack{0\\ me}}^{l-1} D_{lm}(J_l + \beta_{lm}H_l)\frac{dP_l^m}{d\theta}\cos m\phi - i\sum_{\substack{2\\ le}}^{\infty}\sum_{\substack{1\\ mo}}^{l-1} D_{lm}(J_l + \beta_{lm}H_l)\frac{dP_l^m}{d\theta}\sin m\phi \tag{14}$$

$$-\sum_{\substack{2\\le}}^{\infty}\sum_{\substack{2\\me}}^{l}C_{lm}(j_l+\alpha_{lm}h_l)\frac{mP_l^m}{\sin\theta}\cos m\phi+i\sum_{\substack{1\\lo}}^{\infty}\sum_{\substack{1\\mo}}^{l}C_{lm}(j_l+\alpha_{lm}h_l)\frac{mP_l^m}{\sin\theta}\sin m\phi,$$

$$\eta H_\phi=-i\sum_{\substack{1\\lo}}^{\infty}\sum_{\substack{0\\me}}^{l-1}D_{lm}(j_l+\beta_{lm}h_l)\frac{dP_l^m}{d\theta}\cos m\phi-\sum_{\substack{2\\le}}^{\infty}\sum_{\substack{1\\mo}}^{l-1}D_{lm}(j_l+\beta_{lm}h_l)\frac{dP_l^m}{d\theta}\sin m\phi \tag{15}$$

$$-i\sum_{\substack{2\\le}}^{\infty}\sum_{\substack{2\\me}}^{l}C_{lm}(J_l+\alpha_{lm}H_l)\frac{mP_l^m}{\sin\theta}\cos m\phi-\sum_{\substack{1\\lo}}^{\infty}\sum_{\substack{1\\mo}}^{l}C_{lm}(J_l+\alpha_{lm}H_l)\frac{mP_l^m}{\sin\theta}\sin m\phi,$$

$$-E_\phi=\sum_{\substack{1\\lo}}^{\infty}\sum_{\substack{2\\me}}^{l-1}D_{lm}(J_l+\beta_{lm}H_l)\frac{mP_l^m}{\sin\theta}\sin m\phi+i\sum_{\substack{2\\le}}^{\infty}\sum_{\substack{1\\mo}}^{l-1}D_{lm}(J_l+\beta_{lm}H_l)\frac{mP_l^m}{\sin\theta}\cos m\phi \tag{16}$$

$$-\sum_{\substack{2\\le}}^{\infty}\sum_{\substack{2\\me}}^{l}C_{lm}(j_l+\alpha_{lm}h_l)\frac{dP_l^m}{d\theta}\sin m\phi-i\sum_{\substack{1\\lo}}^{\infty}\sum_{\substack{1\\mo}}^{l}C_{lm}(j_l+\alpha_{lm}h_l)\frac{dP_l^m}{d\theta}\cos m\phi,$$

$$\eta H_\theta=-i\sum_{\substack{1\\lo}}^{\infty}\sum_{\substack{2\\me}}^{l-1}D_{lm}(j_l+\beta_{lm}h_l)\frac{mP_l^m}{\sin\theta}\sin m\phi+\sum_{\substack{2\\le}}^{\infty}\sum_{\substack{1\\mo}}^{l-1}D_{lm}(j_l+\beta_{lm}h_l)\frac{mP_l^m}{\sin\theta}\cos m\phi \tag{17}$$

$$-i\sum_{\substack{2\\le}}^{\infty}\sum_{\substack{2\\me}}^{l}C_{lm}(J_l+\alpha_{lm}H_l)\frac{dP_l^m}{d\theta}\sin m\phi+\sum_{\substack{1\\lo}}^{\infty}\sum_{\substack{1\\mo}}^{l}C_{lm}(J_l+\alpha_{lm}H_l)\frac{dP_l^m}{d\theta}\cos m\phi.$$

D. Total field, region II (References 6–9)

Since the antenna arms exclude all fields within angle ψ of the z axis, the appropriate mathematical functions to describe the fields are associated Legendre ones of integer order m and noninteger degree ν,

$$P_\nu^m(\cos\theta)\quad\text{and}\quad P_\nu^m(-\cos\theta), \tag{18}$$

multiplied by spherical Bessel functions of the first and second kind and of fractional order ν,

$$j_\nu(\sigma)\quad\text{and}\quad j_{-\nu}(\sigma), \tag{19}$$

where the order of each Bessel function is equal to the degree of the corresponding Lengendre function. For the special case where b, the radius of the source region, is much less than λ, the wavelength, the multiplicative coefficient of $j_{-\nu}(\sigma)$ is small and, for purposes of this paper, we put it equal to zero.

Exclusion of fields from the z axis by the antenna arms permits the presence of a TEM mode within region II. For this case, $\nu=0$ and the zenith angle function is the zero-degree associated Legendre function of the second kind, $Q_0(\cos\theta)$, multiplied by zero order spherical Bessel and Neumann functions, $j_0(\sigma)$ and $y_0(\sigma)$:

$$Q_0(\cos\theta)(C_0j_0(\sigma)+D_0y_0(\sigma)), \tag{20}$$

where

$$Q_0(\cos\theta)=\ln\cot(\theta/2). \tag{21}$$

Since spherical Bessel functions of the first kind of order greater than zero satisfy

$$\lim_{\sigma\to0}j_\nu(\sigma)=0, \tag{22}$$

it follows that only the product functions (20) describe real power carried to or from the source region for a physically small load located at the center of the antenna.

Turning to the polar angle solutions (18), since $P_\nu^m(\pm\cos\theta)$ are neither simple odd nor even functions of $\cos\theta$, we construct functions of definite parity

$$L_\lambda^m(\cos\theta)=\tfrac12\,[P_\lambda^m(\cos\theta)+P_\lambda^m(-\cos\theta)], \tag{23}$$

and

$$M_\nu^m(\cos\theta)=\tfrac12\,[P_\nu^m(\cos\theta)-P_\nu^m(-\cos\theta)], \tag{24}$$

where the dummy index λ of (23) is named for future convenience. An antenna loaded at the center and driven by the plane wave described in region III will have field symmetries in region II similar to those in region III. Combination of all the above shows the TE and TM fields in region II to be

$$E_r=\sum_{\nu>0}\sum_{\substack{0\\me}}^{\infty}\nu(\nu+1)\Upsilon_{\nu m}\frac{j_\nu(\sigma)}{\sigma}M_\nu^m\cos m\phi-i\sum_{\nu>0}\sum_{\substack{1\\mo}}^{\infty}\nu(\nu+1)\Upsilon_{\nu m}\frac{j_\nu(\sigma)}{\sigma}M_\nu^m\sin m\phi, \tag{25}$$

$$\eta H_r = \sum_{\substack{\lambda>0\\ mo}}\sum_{1}^{\infty}\lambda(\lambda+1)\zeta_{\lambda m}\frac{j_\lambda(\sigma)}{\sigma}L_\lambda^m\cos m\phi - i\sum_{\substack{\lambda>0\\ me}}\sum_{2}^{\infty}\lambda(\lambda+1)\zeta_{\lambda m}\frac{j_\lambda(\sigma)}{\sigma}L_\lambda^m\sin m\phi, \tag{26}$$

$$E_\theta = \sum_{\substack{\nu>0\\ me}}\sum_{0}^{\infty}\Upsilon_{\nu m}J_\nu(\sigma)\frac{dM_\nu^m}{d\theta}\cos m\phi - i\sum_{\substack{\nu>0\\ mo}}\sum_{1}^{\infty}\Upsilon_{\nu m}J_\nu(\sigma)\frac{dM_\nu^m}{d\theta}\sin m\phi \tag{27}$$

$$- \sum_{\substack{\lambda>0\\ me}}\sum_{2}^{\infty}\zeta_{\lambda m}j_\lambda(\sigma)\frac{mL_\lambda^m}{\sin\theta}\cos m\phi + i\sum_{\substack{\lambda>0\\ mo}}\sum_{1}^{\infty}\zeta_{\lambda m}j_\lambda(\sigma)\frac{mL_\lambda^m}{\sin\theta}\sin m\phi,$$

$$\eta H_\phi = -i\sum_{\substack{\nu>0\\ me}}\sum_{0}^{\infty}\Upsilon_{\nu m}j_\nu(\sigma)\frac{dM_\nu^m}{d\theta}\cos m\phi - \sum_{\substack{\nu>0\\ mo}}\sum_{1}^{\infty}\Upsilon_{\nu m}j_\nu(\sigma)\frac{dM_\nu^m}{d\theta}\sin m\phi \tag{28}$$

$$- i\sum_{\substack{\lambda>0\\ me}}\sum_{2}^{\infty}\zeta_{\lambda m}J_\lambda(\sigma)\frac{mL_\lambda^m}{\sin\theta}\cos m\phi - \sum_{\substack{\lambda>0\\ mo}}\sum_{1}^{\infty}\zeta_{\lambda m}J_\lambda(\sigma)\frac{mL_\lambda^m}{\sin\theta}\sin m\phi,$$

$$E_\phi = -i\sum_{\substack{\nu>0\\ mo}}\sum_{1}^{\infty}\Upsilon_{\nu m}J_\nu(\sigma)\frac{mM_\nu^m}{\sin\theta}\cos m\phi - \sum_{\substack{\nu>0\\ me}}\sum_{2}^{\infty}\Upsilon_{\nu m}J_\nu(\sigma)\frac{mM_\nu^m}{\sin\theta}\sin m\phi \tag{29}$$

$$+ i\sum_{\substack{\lambda>0\\ mo}}\sum_{1}^{\infty}\zeta_{\lambda m}j_\lambda(\sigma)\frac{dL_\lambda^m}{d\theta}\cos m\phi + \sum_{\substack{\lambda>0\\ me}}\sum_{2}^{\infty}\zeta_{\lambda m}j_\lambda(\sigma)\frac{dL_\lambda^m}{d\theta}\sin m\phi,$$

$$\eta H_\theta = \sum_{\substack{\nu>0\\ mo}}\sum_{1}^{\infty}\Upsilon_{\nu m}j_\nu(\sigma)\frac{mM_\nu^m}{\sin\theta}\cos m\phi - i\sum_{\substack{\nu>0\\ me}}\sum_{2}^{\infty}\Upsilon_{\nu m}j_\nu(\sigma)\frac{mM_\nu^m}{\sin\theta}\sin m\phi \tag{30}$$

$$+ \sum_{\substack{\lambda>0\\ mo}}\sum_{1}^{\infty}\zeta_{\lambda m}J_\lambda(\sigma)\frac{dL_\lambda^m}{d\theta}\cos m\phi - i\sum_{\substack{\lambda>0\\ me}}\sum_{2}^{\infty}\zeta_{\lambda m}J_\lambda(\sigma)\frac{dL_\lambda^m}{d\theta}\sin m\phi.$$

Off-centered loads would result in additional field terms with other symmetry. All fields are subject to the boundary conditions[6,8]

$$E_r(\cos\psi) = 0 \tag{31}$$

and

$$\frac{dH_r}{d\theta}\Big|_{\theta=\psi} = 0. \tag{32}$$

Equations (31) and (32) affect the solutions in two ways. First they restrict m to be less than azimuth indices ν or λ and, second, they require that

$$M_\nu^m(\cos\psi) = 0 \tag{33}$$

and

$$\frac{dL_\lambda^m}{d\theta}(\cos\psi)\Big|_{\theta=\psi} = 0. \tag{34}$$

Figure 2 shows values of ν and λ for which (33) and (34) are satisfied as a function of ψ, for $m = 0$ and 1. Appendix C contains numerical values of ν at roots of $M_\nu(\cos\psi)$ obtained by numerical techniques as shown in Appendix B, while Appendix D contains the functions $M_\nu(\cos\theta)$ at ν values which are roots of $M_\nu(\cos 1°)$.

In terms of the functions shown in (20) and (21), the TEM fields are

$$E_\theta = \frac{1}{i\sigma\sin\theta}(C_0\cos\sigma + D_0\sin\sigma) \tag{35}$$

and

$$H_\phi = \frac{1}{\eta\,\sigma\sin\theta}(-C_0\sin\sigma + D_0\cos\sigma). \tag{36}$$

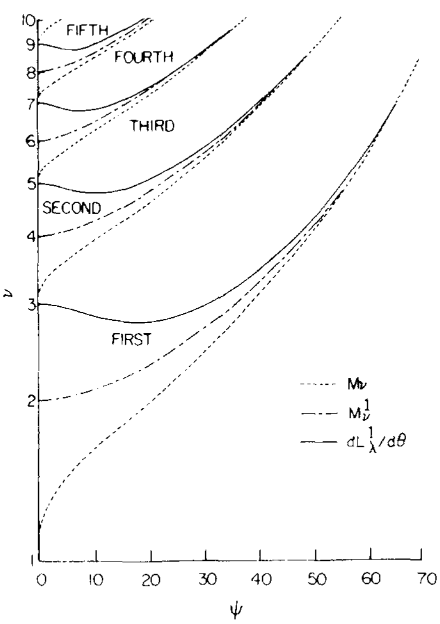


FIG. 2. Root values ν or λ as a function of antenna half-angle ψ for the three functions M_ν, M_ν^1, and $dL_\lambda^1/d\theta$.

It is convenient to define circuit parameters in terms of the TEM modes. Although the fields are singular across the load, the circuit parameters are not.[6,8] The potential $V(r)$ between antenna arms is found using the equation

$$V(r) = \frac{\sigma}{k}\int_{\psi}^{\pi-\psi} E_\theta d\theta. \tag{37}$$

The characteristic admittance G of a biconical transmission line for the TEM mode is

$$G = \frac{\pi}{\eta \ln \cot(\psi/2)}. \tag{38}$$

It follows from (37) and (38) that the TEM electric field of (35) can be written

$$E_\theta = \frac{\eta\, G\, V(r)}{2\pi\, r \sin\theta}. \tag{39}$$

The radial current $I(r)$ is calculated from the equation

$$I(r) = \frac{\sigma}{k}\int_0^{2\pi} H_\phi \sin\theta\, d\phi \,\bigg|_{\theta=\psi}. \tag{40}$$

It follows from (40) that the TEM radial current of (36) can be written

$$H_\phi = \frac{I(r)}{2\pi\, r \sin\theta}. \tag{41}$$

It is also convenient for what lies ahead to express $V(r)$ and $I(r)$ in terms of their values at the antenna termination where $r = a$. In those terms, the circuit quantities are

$$V(r) = V(a)\{\cos k(a-r) + i[Y(a)/G]\sin k(a-r)\} \tag{42}$$

and

$$I(r) = V(a)[Y(a)\cos k(a-r) + i\,G \sin k(a-r)], \tag{43}$$

where $Y(a)$ is the termination TEM admittance. The load admittance Y_L presented between the conical apices of the antenna is known or can be measured, and is equal to the ratio

$$Y_L = I(0)/V(0). \tag{44}$$

In terms of known parameters, the TEM admittance at all other radii is

$$\frac{Y(r)}{G} = \frac{Y_L - i\,G \tan\sigma}{G - i\,Y_L \tan\sigma}. \tag{45}$$

Comparison with Schekunoff[6] shows identical admittance and voltage dependencies for the transmitting and receiving cases.

The surface current density is related to the magnetic field adjacent to it as

$$\hat{n}\times\mathbf{H} = \mathbf{I}, \tag{46}$$

where $\mathbf{I}$ is the surface current density in amperes per meter and $\hat{n}$ is a unit vector normal to the conductor. Knowledge of the field coefficients and (46) permits the calculation of all antenna currents. Those due to the $m > 0$ modes are the receiver modes, and the $m = 0$ modes are the transmitter modes.[4]

III. BOUNDARY CONDITIONS

The boundary conditions yet to be applied are that the angular fields are continuous at radius a. Each of the four field components, separately evaluated in regions II and III, are equal to their own value across the interface within the aperture angle $\psi < \theta < \pi - \psi$, and the angular electric fields are zero in region III immediately adjacent to the spherical cap, i.e., for $0 \leqslant \theta < \psi$ and for $\pi - \psi < \theta \leqslant \pi$. Table I contains a list of integrals useful to establish the needed equalities, while Table II shows useful integral groupings. To evaluate all field constants another set of integrals is needed similar to (II-2)–(II-5) except sin $m'\phi$ is exchanged by cos $m'\phi$, and cos m' by $-$ sin $m'\phi$. Evaluation of these integrals and placing each equal to itself across the interface, including the pole caps for angular electric fields in Region III, yields the set of equations contained in Table III. Table III is sufficient to permit the evaluation of all field constants. Appendix E contains numerical evaluations of some integrals of Table I, for the $\psi = 1°$, $m = 0$ case.

Once the parameter set β_{lo} is known, (III-1) may be used to determine $V(a)$. Equation (45) evaluated at $r = a$ determines $Y(a)$ and since $I(a) = Y(a)V(a)$, (42) and (43) may be used to determine $V(r)$ and $I(r)$ for all radii. $V(0)$ and $I(0)$ are, of course, the load voltage and load current and sufficient to describe the receiving antenna as a circuit element. Equation (III-3) may be used to determine the parameter set $\Upsilon_{\nu 0}$. This completes the solution for all $m = 0$ modes.

IV. THE FIELD CONSTANTS

A. Spheroidal biconical antennas

The spheroidal biconical antenna is the special case of a biconical antenna with ψ approaching $\pi/2$. As was shown by Schelkunoff,[6] the complementary waves within the antenna region may be neglected when the gap between arm tips is much less than $\lambda/2$. For this case and for $m = 0$, (III-1) and (III-2) lead to the following equation for β_{l0}, where σ is the interfacial value:

$$\beta_{l0} = -\frac{J_l(\sigma)}{H_l(\sigma)} - \frac{\eta G^2}{\pi\sigma^2}\frac{P_l(\cos\psi)}{D_{l0}H_l(\sigma)Y(a)} \times\left(\frac{2l+1}{l(l+1)}\right)\sum_{l'}\frac{D_{l'0}P_{l'}(\cos\psi)}{H_{l'}(\sigma)}$$

and since $P_l(\cos\psi)$ decreases toward zero as ψ approaches $\pi/2$, it follows that

$$\beta_{l0} \cong J_l(\sigma)/H_l(\sigma). \tag{47}$$

In a similar way, for $m > 0$, the equations of Appendix C lead to the result that

$$\beta_{lm} = -J_l(\sigma)/H_l(\sigma) \tag{48}$$

and

$$\alpha_{lm} = -j_l(\sigma)/H_l(\sigma). \tag{49}$$

Equations (48) and (49), of course, are descriptive parameters for scattering from a conducting, spherical scatterer.

TABLE I. A table of integrals of associated Legendre functions. $m = 0$ values are found in Ref. 9.

1.	$\int_{\psi}^{\pi-\psi} P_l^m L_\lambda^m \sin\theta\, d\theta = \frac{2\sin\psi L_\lambda^m (dP_l^m/d\psi)}{l(l+1)-\lambda(\lambda+1)}\delta(l+m,2q) = I_{l\lambda}$ $= \frac{2L_\lambda^m(\cos\psi)}{l(l+1)-\lambda(\lambda+1)}[(l-m+1)P_{l+1}^m - (l+1)\cos\psi P_l^m]\delta(l+m,2q).$
2.	$\int_{\psi}^{\pi-\psi} P_l^m M_\nu^m \sin\theta\, d\theta = \frac{2\sin\psi P_l^m (dM_\nu^m/d\psi)}{\nu(\nu+1)-l(l+1)}\delta(l+m,2q+1) = I_{l\nu}$ $= \frac{-2P_l^m(\cos\psi)}{l(l+1)-\nu(\nu+1)}(\nu-m+1)L_{\nu+1}^m(\cos\psi)\delta(l+m,2q+1).$
3.	$\int_{\psi}^{\pi-\psi} L_\nu^m L_\lambda^m \sin\theta\, d\theta = \frac{2\sin\psi}{2\lambda+1} L_\lambda^m \frac{d^2 L_\lambda^m}{d\lambda d\psi}\delta(\nu,\lambda) = I_{\lambda\lambda}.$
4.	$\int_{\psi}^{\pi-\psi} M_\nu^m M_\lambda^m \sin\theta\, d\theta = -\frac{2\sin\psi}{2\nu+1}\frac{dM_\nu^m}{d\nu}\frac{dM_\nu^m}{d\psi}\delta(\nu,\lambda) = I_{\nu\nu}.$
5.	$\int_0^{\pi} P_l^m P_{l'}^m \sin\theta\, d\theta = \frac{2}{2l+1}\frac{(l+m)!}{(l-m)!}\delta(l,l') = I_{ll}.$
6.	$\int_{\psi}^{\pi-\psi} \sin\theta\, d\theta \frac{dP_l^m}{d\theta}\frac{dL_\lambda^m}{d\theta} + \frac{m^2 P_l^m L_\lambda^m}{\sin^2\theta} = \lambda(\lambda+1)I_{l\lambda}.$
7.	$\int_{\pi}^{\pi-\psi} \sin\theta\, d\theta \frac{dL_\lambda^m}{d\theta}\frac{dL_\nu^m}{d\theta}\frac{m^2 L_\lambda^m L_\nu^m}{\sin^2\theta} = \lambda(\lambda+1)I_{\lambda\nu}.$
8.	$\int_{\pi}^{\pi-\psi} \sin\theta\, d\theta \frac{dP_l^m}{d\theta}\frac{dM_\nu^m}{d\theta} + \frac{m^2 M_\nu^m P_l^m}{\sin^2\theta} = l(l+1)I_{l\nu}.$
9.	$\int_{\psi}^{\pi-\psi} \sin\theta\, d\theta \frac{dM_\nu^m}{d\theta}\frac{dM_\lambda^m}{d\theta} + \frac{m^2 M_\nu^m M_\lambda^m}{\sin^2\theta} = \nu(\nu+1)I_{\nu\nu}.$
10.	$\int_{\psi}^{\pi-\psi} \sin\theta\, d\theta \frac{dM_\nu^m}{d\theta}\frac{dL_\lambda^m}{d\theta} + \frac{m^2 M_\nu^m L_\lambda^m}{\sin^2\theta} = 0.$
11.	$\int_0^{\pi} \sin\theta\, d\theta \frac{dP_l^m}{d\theta}\frac{dP_{l'}^m}{d\theta} + \frac{m^2 P_l^m P_{l'}^m}{\sin^2\theta} = l(l+1)I_{ll}.$
12.	$m\int_{\psi}^{\pi-\psi} d\theta \frac{d}{d\theta}(P_l^m M_\nu^m) = 0.$
13.	$m\int_{\psi}^{\pi-\psi} d\theta \frac{d}{d\theta}(L_\lambda^m M_\nu^m) = 0.$
14.	$m\int_{\psi}^{\pi-\psi} d\theta \frac{d}{d\theta}(L_\lambda^m P_l^m) = -2L_\lambda^m P_l^m \delta(l+m,2q+1).$
15.	$m\int_0^{\pi} d\theta \frac{d}{d\theta}(P_l^m P_{l'}^m) = 0.$

B. General solution. $m = 0$

The set of field constants for $m = 0$ follow from (III-1)–(III-3). Solving them for β_{l0} gives

$$\beta_{l0} = A_l + \sum_{l'}{}'' \gamma_{l'}\beta_{l'0}/\beta_l, \tag{50}$$

TABLE II. Aperture expressions for establishing the equations.

1.	$m = 0 \int d\theta\, H_\phi.$
2.	$\int d\Omega\left(E_\theta \frac{dP_{l'}^{m'}}{d\theta}\cos m'\phi - E_\phi \frac{m' P_{l'}^{m'}}{\sin\theta}\sin m'\phi\right).$
3.	$\int d\Omega\left(H_\phi \frac{dM_{\nu'}^{m'}}{d\theta}\cos m'\phi + H_\theta \frac{m' M_{\nu'}^{m'}}{\sin\theta}\sin m'\phi\right).$
4.	$\int d\Omega\left(E_\theta \frac{m' P_{l'}^{m'}}{\sin\theta}\cos m'\phi - E_\phi \frac{dP_{l'}^{m'}}{d\theta}\sin m'\phi\right).$
5.	$\int d\Omega\left(H_\phi \frac{m' L_{\lambda'}^{m'}}{\sin\theta}\cos m'\phi + H_\theta \frac{dL_{\lambda'}^{m'}}{d\theta}\sin m'\phi\right).$

where the symbol ″ over the summation indicates that the $l' = l$ term is not to be included in the sum, and where

$$A_l = -l(l+1)D_{l0}J_l I_{ll} + \frac{2i\eta G^2}{\pi Y(a)}P_l(\cos\psi)\sum_{l'} D_{l'0}j_{l'}P_{l'}(\cos\psi) + l(l+1)\sum_\nu \frac{J_\nu I_{l\nu}}{\nu(\nu+1)j_\nu I_{\nu\nu}}\sum_{l'} l'(l'+1)D_{l'0}j_{l'}I_{l'\nu},$$

$$B_l = D_{l0}\left(l(l+1)H_l I_{ll} - \frac{2i\eta G^2}{\pi Y(a)}[P_l(\cos\psi)]^2 h_l - l^2(l+1)^2 h_l \sum_\nu \frac{J_\nu (I_{l\nu})^2}{\nu(\nu+1)j_\nu I_{\nu\nu}}\right),$$

$$\gamma_{l'} = D_{l'0}\left(\frac{2i\eta G^2}{\pi Y(a)}P_l(\cos\psi)h_{l'}P_{l'}(\cos\psi) + l(l+1)l'(l'+1)h_{l'}\sum_\nu \frac{J_\nu I_{l\nu}I_{l'\nu}}{\nu(\nu+1)j_\nu I_{\nu\nu}}\right).$$

Equation (50) may be solved to obtain each of the coefficient sets β_{l0} to whatever accuracy is warranted.

TABLE III. Equations for establishing values of coefficients.

1. $\dfrac{V(a)}{a} = \dfrac{2iG}{Y(a)} \sum_l D_{l0}(j_l + \beta_{l0} h_l) P_l(\cos\psi).$

2. $l(l+1) D_{lm}(J_l + \beta_{lm} H_l) I_{ll} = l(l+1) \sum_\nu \Upsilon_{\nu m} J_\nu I_{l\nu} - \dfrac{V(a)}{a} \dfrac{P_l(\cos\psi)}{\ln\cot\psi/2} \delta(m,0) + 2mP_l^m(\cos\psi) \sum_\lambda \zeta_{\lambda m} j_\lambda L_\lambda^m(\cos\psi).$

3. $\nu(\nu+1) \Upsilon_{\nu m} j_\nu I_{\nu\nu} = \sum_l l(l+1) D_{lm}(j_l + \beta_{lm} h_l) I_{l\nu}.$

4. $l(l+1) C_{lm}(j_l + \alpha_{lm} h_l) I_{ll} = \sum_\lambda \lambda(\lambda+1) \zeta_{\lambda m} j_\lambda I_{l\lambda}.$

5. $\lambda(\lambda+1) \zeta_{\lambda m} J_\lambda I_{\lambda\lambda} = \lambda(\lambda+1) \sum_l C_{lm}(J_l + \alpha_{lm} H_l) I_{l\lambda} - 2mL_\lambda^m(\cos\psi) \sum_l D_{lm}(j_l + \beta_{lm} h_l) P_l^m(\cos\psi).$

C. $m > 0$

Field coefficients for the $m > 0$ modes follow from (III-2)–(III-5). Combinations of (III-2), (III-3), and (III-5) give an equation for β_{lm} of the form

$$\beta_{lm} = \frac{1}{F'_{lm}} \left(\sum_{l'} g'_{l'} \beta_{l'm} + B'_{lm} + \sum_{l'} f'_{l'} \alpha_{l'm} \right), \tag{51}$$

while (III-4) and (III-5) give an equation for α_{lm} of the form

$$\alpha_{lm} = \frac{1}{F''_{lm}} \left(\sum_{l'} g''_{l'} \alpha_{l'm} + B''_{lm} + \sum_{l'} f''_{l'} \beta_{l'm} \right). \tag{52}$$

α_{lm} and β_{lm} may be found concurrently to whatever accuracy is deemed worthy of being obtained. Once α_{lm} and β_{lm} are known, (III-3) and (III-5) permit the calculation of $\Upsilon_{\nu m}$ and $\zeta_{\lambda m}$. Knowledge of these parameters give the total solution to the external field problem around a receiving antenna. Parameter values from (51) and (52) are given in Table IV. Appendix F contains some functional values of $J_\nu/\nu(\nu+1) j_\nu I_{\nu\nu}$ at roots of M_ν (cos 1°). Each electric multipole interacts with other electric multipoles and all magnetic multipoles.

All $m = 0$ modes are damped by the antenna load. However, for ψ small and approaching zero, the parameters ν approach integer values only slowly, as may be seen from Appendix C and Fig. 2. Therefore the integrals $I_{l\nu}$ remain significantly large even for ψ quite near zero, as may be seen in the values of Appendix E. The $m > 0$ modes, on the other hand, are undamped by a physically small and centered load, and their scattering is independent of it.

D. Thin biconical antenna

The thin biconical antenna is the special case of a biconical antenna with ψ approaching zero. For this case ν approaches an odd integer, λ approaches an even integer, P_l ($\cos\psi$) approaches unity, and $P_l^m(\cos\psi)$ approaches zero as $\sin^m\psi$.

E. $m = 0$

For this case, using Schelkunoff's results,[6] Sec. 2.10, and solving (III-1)–(III-3) we obtain, for very small values of ψ,

$$\beta_{l0} = \frac{j_l}{h_l} \left(\frac{\Upsilon_{l0} - D_{l0}}{D_{l0}} \right), \tag{53}$$

TABLE IV. Parameter values for (47) and (48).

$$F'_{lm} = l(l+1) D_{lm} H_l I_{ll} - l^2(l+1)^2 D_{lm} h_l \sum_\nu \frac{J_\nu (I_{l\nu})^2}{\nu(\nu+1) j_\nu I_{\nu\nu}} + 4m^2 D_{lm} h_l (P_l^m)^2 \sum_\lambda \frac{j_\lambda (L_\lambda^m)^2}{\lambda(\lambda+1) J_\lambda I_{\lambda\lambda}},$$

$$B'_{lm} = -l(l+1) D_{lm} J_l I_{ll} + l(l+1) \sum_{l'} l'(l'+1) D_{l'm} j_{l'} \sum_\nu \frac{J_\nu I_{l\nu} I_{l'\nu}}{\nu(\nu+1) j_\nu I_{\nu\nu}} - 4m^2 P_l^m \sum_{l'} D_{l'm} j_{l'} P_{l'}^m \sum_\lambda \frac{j_\lambda (L_\lambda^m)^2}{\lambda(\lambda+1) J_\lambda I_{\lambda\lambda}} + 2mP_l^m \sum_{l'} C_{l'm} J_{l'} \sum_\lambda \frac{j_\lambda L_\lambda^m I_{l'\lambda}}{J_\lambda I_{\lambda\lambda}},$$

$$f'_{l'm} = 2mP_l^m C_{l'm} H_{l'} \sum_\lambda (j_\lambda L_\lambda^m I_{l'\lambda}) / J_\lambda I_{\lambda\lambda}),$$

$$g'_{l'm} = l(l+1) l'(l'+1) D_{l'm} h_{l'} \sum_\nu (J_\nu I_{l\nu} I_{l'\nu}) / (\nu(\nu+1) j_\nu I_{\nu\nu}) - 4m^2 P_l^m D_{l'm} h_{l'} P_{l'}^m \sum_\nu [j_\lambda (L_\lambda^m)^2] / [\lambda(\lambda+1) J_\lambda I_{\lambda\lambda}],$$

$$F''_{lm} = l(l+1) C_{lm} h_l I_{ll} - C_{lm} H_l \sum_\lambda [\lambda(\lambda+1) j_\lambda (I_{l\lambda})^2] / (J_\lambda I_{\lambda\lambda}),$$

$$B''_{lm} = -l(l+1) C_{lm} j_l I_{ll} + \sum_{l'} C_{l'm} J_{l'} \sum_\lambda [\lambda(\lambda+1) j_\lambda I_{l\lambda} I_{l'\lambda}] / (J_\lambda I_{\lambda\lambda}) - 2m \sum_{l'} D_{l'm} j_{l'} P_{l'}^m \sum_\lambda (j_\lambda I_{l\lambda} L_\lambda^m) / (J_\lambda I_{\lambda\lambda}),$$

$$f''_{l'm} = -2m D_{l'}^m P_{l'}^m h_{l'} \sum_\lambda (J_\lambda I_{l\lambda} L_\lambda^m) / (J_\lambda I_{\lambda\lambda}),$$

$$g''_{l'm} = C_{l'm} H_{l'} \sum_\lambda [\lambda(\lambda+1) j_\lambda I_{l\lambda} I_{l'\lambda}] / (J_\lambda I_{\lambda\lambda}).$$

$$\frac{V(a)}{a} = \frac{2iG}{Y(a)} \sum_{\substack{1\\lo}}^{\infty} j_l \Upsilon_{lo}, \tag{54}$$

$$\Upsilon_{lo} = \left[D_{lo} - \frac{\eta G^2 \sigma^2 h_l}{Y(a)} \left(\frac{2l+1}{l(l+1)} \right) \sum_{\substack{1\\l'o}}^{\infty} \Upsilon_{l'o} j_{l'} \right] \times \left[1 + \frac{\eta G^2 \sigma^2 h_l j_l}{\pi Y(a)} \left(\frac{2l+1}{l(l+1)} \right) \right]^{-1}. \tag{55}$$

Since G approaches zero as the logarithm of ψ, (53) and (55) lead to

$$\Upsilon_{lo} = D_{lo} \quad \text{and} \quad \beta_{lo} = 0, \tag{56}$$

while (45), (54), and (56) combine to give

$$\frac{V(a)}{a} = 2 \tan \sigma \sum_{\substack{1\\lo}}^{\infty} D_{lo}\, j_l(\sigma). \tag{57}$$

F. $m > 0$

The coefficients for $m > 0$ may be evaluated using (51) and (52) for β_{lm} and α_{lm}, respectively, combined with the defining equations of Table IV. The results are

$$\beta_{lm} = \frac{im(2l+1)}{D_{lm} l(l+1)} \sigma^2 j_l(\sigma) P_l^m(\cos\psi) \sum_{l'} C_{l'm}\, j_{l'} P_{l'}^m(\cos\psi), \tag{58}$$

where, in (58), $(l+m)$ and $(l'+m)$ are odd and even integers, respectively.

$$\alpha_{lm} = \frac{im(2l+1)}{C_{lm} l(l+1)} \sigma^2 j_l(\sigma) P_l^m(\cos\psi) \sum_{l'} D_{l'm}\, j_{l'} P_{l'}^m(\cos\psi), \tag{59}$$

where, in (59), $(l+m)$ and $(l'+m)$ are even and odd integers, respectively. Equations (58) and (59) combine with the equations of Table III to show that, in the limit as $\psi \to 0$,

$$\zeta_{lm} - C_{lm} = \alpha_{lm} C_{lm} (h_l / j_l), \tag{60}$$

$$\Upsilon_{lm} - D_{lm} = \beta_{lm} D_{lm} (h_l / j_l). \tag{61}$$

It is, perhaps, of interest to note that in this thin antenna case, the scattered electric multipolar field has its origin in the driving magnetic multipolar field, and vice versa. Equations (58)–(61) show that α_{lm}, β_{lm}, $\zeta_{lm} - C_{lm}$, and $\Upsilon_{Lm} - D_{lm}$ all approach zero as $\sin^{2m}\psi$.

V. RECIPROCITY

The electromagnetically exact solution for an arbitrarily sized biconical receiving antenna permits direct confirmation of the reciprocity theorem for antennas.

The equality of admittance in an antenna when used for either reception or transmission may be seen by use of the following argument. Using (III-1), the terminal admittance is of the form

$$Y(a) = \frac{2iGa}{V(a)} \sum_{l} D_{lo} x_l P_l(\cos\psi), \tag{62}$$

while (III-2) may be written

$$l(l+1) D_{lo} X_l I_{ll} = -\frac{\eta G V(a)}{\pi a} P_l(\cos\psi) + l(l+1) \sum_{v} \sum_{l'} \frac{J_l I_{ll} I_{l'} l'(l'+1) D_{l'} x_{l'}}{v(v+1) j_v I_{vv}}. \tag{63}$$

For a receiving antenna

$$x_l = j_l(\sigma) + \beta_{lo} h_l(\sigma)$$

and for a transmitting antenna

$$x_l = \beta_{lo} h_l(\sigma).$$

In both cases

$$X_l = \frac{1}{\sigma} \frac{d}{d\sigma} x_l.$$

Since solutions are obtained by elimination of x_l from (62) and (63), it follows that ultimate solutions for $V(a)$ and $Y(a)$ are identical in the two cases.

Radiation patterns produced by the transmitting modes for receiving and transmitting are not the same. The radiation pattern for the $m = 0$ mode follows from Poynting's theorem and the field equations (14)–(17), and is

$$N_r = \sum_{\substack{1\\lo}}^{\infty} \frac{D_{lo}^2}{2\eta\sigma^2} \left(\frac{dP_l}{d\theta} \right)^2 Z_l, \tag{64}$$

where N_r is the radial component of the Poynting vector, and where for reception

$$Z_l = \mathrm{Re}\, \beta_{lo} + \beta_{lo} \beta_{lo}^*. \tag{65}$$

Although the formal result is the same in the two cases, the parameters are not, and as shown in (G8),

$$(\beta_t'^2 + \beta_t''^2) - (\beta_r'^2 + \beta_r''^2) = (2\beta_t' - 1)\sin^2\gamma + \beta_t'' \sin\gamma\cos\gamma, \tag{66}$$

where notation is defined in Appendix G. Since the difference in βs is l-dependent, different values of l will be weighted differently in the radiation transmitted during reception or transmission, and therefore the radiated power patterns will differ.

VI. POWER FLOWS

The power absorbed by an antenna may be calculated by the use of Poynting's theorem. Writing **N** for the Poynting vector and N_r for its radial component,

$$N_r = (E_\theta H_\phi^* - E_\phi H_\theta^*)/2,$$

where the asterisk indicates complex conjugate, and introducing differential solid angle $d\Omega$, where

$$d\Omega = \sin\theta\, d\theta\, d\phi,$$

it follows that the power flowing in through a Gaussian surface of large radius surrounding the antenna is

$$P = -\oint r^2 N_r\, d\Omega \tag{67}$$

and using conservation of power, P is also the power absorbed by the antenna. Evaluation of (67) using the external field equations when r is very large shows that

$$P = \frac{-4\pi}{\eta k^2}\left\{\left(\sum_{\substack{1\\ lo}}^{\infty}\sum_{\substack{0\\ me}}^{l-1} + \sum_{\substack{1\\ le}}^{\infty}\sum_{\substack{1\\ mo}}^{l-1}\right)\right.$$

$$\times \frac{4U(m)(2l+1)(l-m)!(l+m)!(\mathrm{Re}\,\beta_{lm} + \beta_{lm}\beta^*_{lm})}{2^{2l}l(l+1)[(l-1-m)/2]!^2[(l-1+m)/2]!^2}$$

$$+\left(\sum_{\substack{1\\ lo}}^{\infty}\sum_{\substack{1\\ mo}}^{l} + \sum_{\substack{2\\ le}}^{\infty}\sum_{\substack{2\\ me}}^{l}\right)$$

$$\left.\times \frac{m^2(2l+1)(l-m)!(l+m)!(\mathrm{Re}\,\alpha_{lm} + \alpha_{lm}\alpha^*_{lm})}{2^{2l}l(l+1)[(l-m)/2]!^2[(l+m)/2]!^2}\right\}. \tag{68}$$

The plane wave also carries momentum, and mometum will also be transferred to the antenna. According to Einstein,[10] even though it is very small, it still must be conserved. The momentum density associated with an electromagnetic wave is $\mathbf{N}/c^2$. Since we are considering an incoming y-directed plane wave and, by symmetry, there can be no other directions of momentum transferred to the antenna, we confine our attention to y-directed values. It is convenient to work with directed power, which is c times the rate of change of momentum. The directed power passing through a Gaussian surface of radius r is

$$P_{y0} = \oint r^2 N_r \sin\theta \sin\phi \, d\Omega. \tag{69}$$

In the absence of any scatterers, a plane wave with Poynting vector magnitude N_0 will show directed power

$$2\pi r^2 N_0$$

passing through the region. The difference between this value and (69) evaluated with the antenna in position is the momentum extracted from the beam. Conservation of momentum requires that difference also to be the momentum transferred to the antenna. c times the rate at which momentum is received by the antenna is, therefore, given by P_y, where

$$P_y = -\oint{}' r^2 N_r \sin\theta \sin\phi \, d\Omega, \tag{70}$$

where the prime over the integral indicates that only those terms proportional to α_{lm}, β_{lm}, or products thereof are to be retained.

Evaluation of (70) yields the following expression for P_y, after rather lengthy algebra.

$$P_y = \frac{-2\pi}{\eta k^2}\left\{\left(\sum_{\substack{1\\ lo}}^{\infty}\sum_{\substack{0\\ me}}^{l-1} + \sum_{\substack{2\\ le}}^{\infty}\sum_{\substack{1\\ mo}}^{l-1}\right)\frac{4U(m)(l+m)!(l-m)!}{2^{2l}[(l-1+m)/2]!^2[(l-1-m)/2]!^2}\right.$$

$$\times \mathrm{Re}\left[\frac{(2l+1)}{l(l+1)}\beta_{lm} + (1+2\beta^*_{lm})\left(\frac{(l+m+2)}{(l+1)^2}\beta_{l+1,m+1} + \frac{(l-m-1)}{l^2}\beta_{l-1,m+1} + \frac{(m+1)(2l+1)}{l^2(l+1)^2}\alpha_{l,m+1}\right)\right]$$

$$+\left(\sum_{\substack{1\\ lo}}^{\infty}\sum_{\substack{1\\ mo}}^{l} + \sum_{\substack{2\\ le}}^{\infty}\sum_{\substack{2\\ me}}^{l}\right)\frac{(m)(l+1)!(l-m)!}{2^{2l}\left(\frac{l+m}{2}\right)!^2\left(\frac{l-m}{2}\right)!^2}\mathrm{Re}\left[\frac{m(2l+1)}{l(l+1)}\alpha_{lm} + (1+2\alpha^*_{lm})\left(\frac{(m+1)(l+m+1)}{(l+1)^2}\alpha_{l+1,m+1}\right.\right.$$

$$\left.\left.\left.+\frac{(m+1)(l-m)}{l^2}\alpha_{l-1,m+1} - \frac{(l-m)(l+m+1)(2l+1)}{l^2(l+1)^2}\beta_{l,m+1}\right)\right]\right\}. \tag{71}$$

Appendix C and Fig. 2 show that for the $m=0$ case the parameters ν approach integer values quite slowly with decreasing ψ. Therefore the integrals $I_{l\nu}$ remain significantly large for values of ν in the neighborhood of l even when ψ is quite small, as may be seen in Appendix D. The currents pass through the load and, generally speaking, are damped. The situation with the $m=1$ modes is quite different. ν and λ approach integer values much more rapidly with decreasing ψ, and the currents are undamped and unaffected by the load. Nonetheless these currents are essential to the process of power reception, as we shall see.

Since transmitting antennas, generally speaking, do not have infinite gain, the ratio of radiated directed-to-total power satisfies

$$P_y/P \leqslant 1. \tag{72}$$

For example, in the case of a center-loaded biconical transmitting antenna, the ratio (72) is equal to zero. A receiving antenna, on the other hand, will gather power from the wave and, generally speaking, some of it will be dissipated in the load and the remainder reradiated away in the manner of a transmitting antenna. The received directed-to-total power ratio satisfies

$$P_y/P \geqslant 1. \tag{73}$$

Let us apply these concepts to (68) and (71). For the $m=0$ case, these equations become, respectively,

$$P = \frac{-8\pi}{\eta k^2}\sum_{\substack{1\\ lo}}^{\infty}\frac{(2l+1)}{l(l+1)}\frac{l!^2}{2^{2l}\left(\frac{l-1}{2}\right)!^4}(\mathrm{Re}\,\beta_{l0} + \beta_{l0}\beta^*_{l0}) \tag{74}$$

and

$$P_y = \frac{-4\pi}{\eta k^2}\sum_{\substack{1\\ lo}}^{\infty}\frac{(2l+1)}{l(l+1)}\frac{l!^2}{2^{2l}\left(\frac{l-1}{2}\right)!^4}\mathrm{Re}\,\beta_{l0}. \tag{75}$$

Power absorption occurs when $\mathrm{Re}\,\beta_{l0} < 0$. For that case, the term proportional to $8\,\mathrm{Re}\,\beta_{l0}$ in (74) represents power extracted from the beam while the $8\ \beta_{l0}\beta^*_{l0}$ term represents power radiated away with no net momentum. Therefore, directed power in proportion to $8\,\mathrm{Re}\,\beta_{l0}$ is extracted from the y-directed beam. Yet by (75) the momentum of the beam is reduced just half of that amount. We conclude, therefore, that the $m = 0$ terms cannot by and of themselves describe a satisfactory solution to the receiver problem. The β_{l0} terms are insufficient to account for the conservation of momentum and, therefore, the receiver modes and the momentum transfer associate with them are essential for power reception to occur.

VII. CONCLUSIONS

This paper contains the modal solution of a biconical receiving antenna. The solution permits the calculation of antenna directivity and impedance, momentum and power flow to the antenna, all current modes on the arms and caps, and all electromagnetic fields including near and scattered ones, all in response to a driving plane wave. It is shown that although the $m > 0$ modes contribute no useful load power, they must necessarily be present for momentum to be conserved in the act of power absorption. The special cases of very thin and nearly spherical antennas are treated, and certain selected numerical results are shown. Detailed calculation of comparative receiving and transmitting admittances confirms that they are identical, as predicted by the reciprocity theorem. The radiation patterns, however, for retransmission during reception and for transmission differ.

ACKNOWLEDGMENTS

The author is grateful to Clifford E. Melhiem and Frederick David of the University of Texas at El Paso for assistance with the numerical calculations.

APPENDIX A

It is desired to make a spherical harmonic expansion of e^{-iky}, as

$$e^{-i\sigma\sin\theta\sin\phi}$$
$$= \sum_{l=0}^{\infty}\sum_{m=0}^{l} C^c_{lm}\frac{l(l+1)}{m} j_l(\sigma)P^m_l(\cos\theta)\cos m\phi - i\sum_{l=0}^{\infty}\sum_{m=0}^{l} C^s_{lm}\frac{l(l+1)}{m} j_l(\sigma)P^m_l(\cos\theta)\sin m\phi, \tag{A1}$$

from which it is necessary to evaluate the coefficients C^c_{lm} and C^s_{lm}. Multiplying (A1) by $P^{m'}_{l'}(\cos\theta)\cos m'\phi$ and integrating over the full range of solid angle gives

$$\int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta\, d\theta\, P^m_l(\cos\theta)\cos m\phi\, e^{-i\sigma\sin\theta\sin\phi}$$
$$= 2\pi\frac{(l+m)!}{(l-m)!}\frac{l(l+1)}{2l+1}\frac{C^c_{lm}}{m} j_l(\sigma)[1+\delta(0,m)], \tag{A2}$$

while doing similarly with $P^{m'}_{l'}(\cos\theta)\sin m'\phi$ results in

$$\int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta\, d\theta\, P^m_l(\cos\theta)\sin m\phi\, e^{-i\sigma\sin\theta\sin\phi}$$
$$= 2\pi\frac{(l+m)!}{(l-m)!}\frac{l(l+1)}{2l+1}\frac{C^s_{lm}}{m} j_l(\sigma). \tag{A3}$$

Using the limiting value of $j_l(\sigma)$ that

$$\lim_{\sigma\to 0} j_l(\sigma) = \frac{2^l\sigma^l l!}{(2l+1)!}$$

and taking the l th derivative of both sides of (A2) with respect to σ, then letting σ become vanishingly small results in (A2) becoming equal to

$$\int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta\, d\theta\, P^m_l(\cos\theta)\cos m\phi\,(-i\sin\theta\sin\phi)^l$$
$$= 2\pi\frac{(l+m)!}{(l-m)!}\frac{l(l+1)}{2l+1}\frac{2^l l!^2}{(2l+1)!}\frac{C^c_{lm}}{m}[1+\delta(0,m)] \tag{A4}$$

and (A3) becoming equal to

$$\int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta\, d\theta P^m_l(\cos\theta)\sin m\phi\,(-i\sin\theta\sin\phi)^l$$
$$= 2\pi\frac{(l+m)!}{(l-m)!}\frac{l(l+1)}{2l+1}\frac{2^l l!^2}{(2l+1)!}\frac{C^s_{lm}}{m}. \tag{A5}$$

To evaluate the left sides of (A4) and (A5) we note from symmetry that

$$\int_0^{2\pi} d\phi\,\sin^l\phi\,\cos m\phi = 0 \quad \text{unless } (l,m) \text{ are both even integers,} \tag{A6}$$

$$\int^{\pi} d\phi\,\sin^l\phi\,\sin m\phi = 0 \quad \text{unless } (l,m) \text{ are both odd integers,} \tag{A7}$$

$$\int_0^{\pi} \sin\theta\, d\theta\, \sin^l\theta P^m_l(\cos\theta) = 0 \quad \text{for } (l+m) \text{ odd,}$$
$$= (-1)^{(m-l)/2}\frac{2^{l+1}l!(l+m)!}{(2l+1)!}, \quad (l+m) \text{ even.} \tag{A8}$$

Since C^c_{lm} and C^s_{lm} exist only for different sets of (l,m), it is no longer necessary to keep track of them separately, and the exponential may be written

$$\exp(-i\sigma\sin\theta\sin\phi) = \left(\sum_{\substack{0\\le}}^{\infty}\sum_{\substack{0\\me}}^{l}\cos m\phi - i\sum_{\substack{1\\lo}}^{\infty}\sum_{\substack{1\\mo}}^{l}\sin m\phi\right) \times \frac{l(l+1)}{m} C_{lm}\, j_l(\sigma)P^m_l(\cos\theta). \tag{A9}$$

By use of the identities

$$\sin^{2l}\phi = \frac{1}{2^{2l}}\left[\sum_{k=0}^{l-1}(-1)^{l-k}\, 2\binom{2l}{k}\cos 2(l-k)\phi + \binom{2l}{k}\right],$$
$$m \text{ even,} \tag{A10}$$

$$\sin^{2l-1}\phi = \frac{1}{2^{2l-2}}\sum_{k=0}^{l-1}(-1)^{l+k-1}\binom{2l-1}{k}\sin(2l-2k-1),$$
$$m \text{ odd,} \tag{A11}$$

we obtain

$$\int_0^{2\pi} d\phi\,\sin^l\phi\,\cos m\phi = \frac{2\pi}{2^l}(-1)^{m/2}\binom{l}{(l+m)/2},$$
$$m \text{ even,} \tag{A12}$$

$$\int_0^{2\pi} d\phi \sin^l\phi \sin m\phi = \frac{2\pi}{2^{l-1}}(-1)^{(m-1)/2}\binom{l}{(l+m)/2},$$
$$m \text{ odd.} \tag{A13}$$

Combination of (A5), (A8), (A9), (A12), and (A13) results in the equality

$$C_{lm} = \frac{2m(2l+1)}{2^l l(l+1)}\binom{l-m}{(l+m)/2}\delta(l+m,2q), \tag{A14}$$

where q is an integer.

Turning to the field equations, we consider a y-directed plane wave with field values

$$\mathbf{E} = \hat{z} \quad \text{and} \quad \mathbf{H} = (1/\eta)\hat{x}, \tag{A15}$$

where $\hat{x}$ and $\hat{z}$ represent unit vectors in the direction indicated. Using (A15), we obtain

$$H_r = \frac{1}{\eta}\sin\theta\cos\phi\, e^{-i\sigma\sin\theta\sin\phi}$$

or

$$H_r = \frac{1}{\eta}\sin\theta\cos\phi\left(\sum_{\substack{0\\ le}}^{\infty}\sum_{\substack{0\\ me}}^{l}\cos m\phi - i\sum_{\substack{1\\ lo}}^{\infty}\sum_{\substack{1\\ mo}}^{l}\sin m\phi\right)$$
$$\times\frac{l(l+1)}{m}C_{lm}\,j_l(\sigma)P_l^m(\cos\theta). \tag{A16}$$

Since

$$\frac{i}{\sigma}\frac{\partial}{\partial\phi}e^{-i\sigma\sin\theta\sin\phi} = \sin\theta\cos\phi\, e^{-i\sigma\sin\theta\sin\phi}, \tag{A17}$$

it follows that

$$H_r = \frac{1}{\eta}\left(\sum_{\substack{1\\ lo}}^{\infty}\sum_{\substack{1\\ mo}}^{l}\cos m\phi - i\sum_{\substack{2\\ le}}^{\infty}\sum_{\substack{2\\ me}}^{l}\sin m\phi\right)$$
$$\times l(l+1)C_{lm}\frac{j_l(\sigma)}{\sigma}P_l^m(\cos\theta). \tag{A18}$$

Next we seek an expression for the radial component of the electric field. We know that

$$E_r = \cos\theta\, e^{-i\sigma\sin\theta\sin\phi} \tag{A19}$$

and we seek a solution of the form

$$E_r = \left(\sum_{\substack{0\\ l}}^{\infty}\sum_{\substack{0\\ m}}^{l}D_{lm}^{c}\cos m\phi - i\sum_{\substack{0\\ l}}^{\infty}\sum_{\substack{0\\ m}}^{l}D_{lm}^{s}\sin m\phi\right)$$
$$\times l(l+1)\frac{j_l(\sigma)}{\sigma}P_l^m(\cos\theta) \tag{A20}$$

from which it is necessary to evaluate the coefficients D_{lm}^c and D_{lm}^s.

To begin we note that

$$\frac{i}{\sigma\sin\phi}\frac{\partial}{\partial\theta}e^{-i\sigma\sin\theta\sin\phi} = \cos\theta\, e^{-i\sigma\sin\theta\sin\phi} \tag{A21}$$

and

$$\frac{1}{\sin\phi} = 2\sum_{s=0}^{\infty}\sin(2s+1)\phi. \tag{A22}$$

Combination of (A9), (A19), (A20), and (A21) gives

$$E_r = \sum_{s=0}^{\infty}\left(\sum_{\substack{1\\ lo}}^{\infty}\sum_{\substack{1\\ mo}}^{\infty}\sin m\phi + i\sum_{\substack{0\\ le}}^{\infty}\sum_{\substack{0\\ me}}^{l}\cos m\phi\right)$$
$$\times 2\sin(2s+1)\phi\,\frac{l(l+1)}{m}C_{lm}$$
$$\times\frac{j_l(\sigma)}{\sigma}P_l^m(\cos\theta). \tag{A23}$$

We work in detail the case of l odd, and note that the solution for l even follows in a parallel way. For l odd we equate (A20) and (A23), then differentiate each side $(l-1)$ times with respect to σ, then let σ become vanishingly small, to obtain

$$\left(\sum_{m=0}^{l}D_{lm}^{c}\cos m\phi - i\sum_{m=0}^{l}D_{lm}^{s}\sin m\phi\right)P_l^m(\cos\theta)$$
$$= \sum_{\substack{1\\ mo}}^{l}\sum_{s=0}^{\infty}\sin m\phi\sin(2s+1)\phi\,\frac{2C_{lm}}{m}\frac{dP_l^m(\cos\theta)}{d\theta}. \tag{A24}$$

Since multiplying (A24) by $\sin m\phi$ and integrating on ϕ from 0 to 2π shows that $D_{lm}^s = 0$, we are justified in dropping the superscript on D_{lm}^c. Multiplying (A24) by $\cos n\phi\, d\phi$ and integrating on ϕ from 0 to 2π results in

$$D_{ln}[1+\delta(0,n)]P_l^n(\cos\theta)$$
$$= \sum_{s=0}^{\infty}\frac{C_{lm}}{m}\frac{dP_l^m(\cos\theta)}{d\theta}$$
$$\times(1+\delta(0,n))[\delta(n,|m-2s-1|)$$
$$-\delta(n,m+2s+1)] \tag{A25}$$

and the right-hand side (rhs) of (A25) may be written

$$\text{rhs} = \sum_{s=0}^{(l-n-1)/2}\frac{C_{l,2s+1+n}}{2s+1+n}\frac{dP_l^{2s+1+n}}{d\theta}$$
$$+\sum_{s=n/2}^{(l+n-1)/2}\frac{C_{l,2s+1-n}}{2s+1-n}\frac{dP_l^{2s+1-n}}{d\theta}$$
$$-\sum_{s=0}^{(n-2)/2}\frac{C_{l,n-2s-1}}{n-2s-1}\frac{dP_l^{n-2s-1}}{d\theta}. \tag{A26}$$

Substituting into the last term of (A26) that $s' = n-s-1$ changes it to

$$-\sum_{s=n/2}^{n-1}\frac{C_{l,2s-1-n}}{2s+1-n}\frac{dP_l^{2s+1-n}}{d\theta}$$

which, when added to the center term, gives

$$\sum_{s=n}^{(l+n-1)/2}\frac{C_{l,2s+1-n}}{2s+1-n}\frac{dP_l^{2s+1-n}}{d\theta}$$

into which the substitution $s'+n=s$ shows that it is equal to the first term, or

$$\text{rhs} = 2\sum_{s=0}^{(l-n-1)/2}\frac{C_{l,2s+1+n}}{2s+1+n}\frac{dP_l^{2s+1+n}}{d\theta}. \tag{A27}$$

Use of the identity from the theory of associate Legendre functions that

$$\frac{dP_l^m}{d\theta} = \tfrac{1}{2}[(l+m)(l-m+1)P_l^{m-1} - P_l^{m+1}] \tag{A28}$$

and combination of (A25), (A27), and (A28) gives

$$\begin{aligned} D_{lm}[1+\delta(0,m)]P_l^m(\cos\theta) &= \sum_{s=0}^{(l-m-1)/2} \frac{C_{l,2s+1+m}}{2s+1+m} \\ &\times [(l+2s+1+m)(l-2s-m)P_l^{m+2s} \\ &- P_l^{m+2s+2}]. \end{aligned} \tag{A29}$$

It may be shown using the theory of associated Legendre polynomials that

$$\int_0^\pi \sin\theta\, d\theta\, P_l^m(\cos\theta)P_l^{m+2s}(\cos\theta) = \frac{2(-1)^s}{(2l+1)}\frac{(l+m)!}{(l-m-2s)!}. \tag{A30}$$

Multiplying (A29) by $P_l^m(\cos\theta)\sin\theta\, d\theta$, then integrating over θ from 0 to π and using (A30) results in the equality

$$[1+\delta(0,m)]D_{lm} = 2l(l-m)! \sum_{s=0}^{(l-m-1)/2} \frac{C_{l,2s+1+m}}{2s+1+m}\frac{(-1)^s}{(l-2s-1-m)!}. \tag{A31}$$

Equations (A14) and (A31) combine to give

$$D_{lm} = \frac{4(2l+1)(l-m)!U(m)}{2^l l(l+1)} \sum_{s=0}^{(l-m-1)/2}\left[l(-1)^s\Big/\left(\frac{l+2s+1+m}{2}\right)!\left(\frac{l-2s-1-m}{2}\right)!\right] \tag{A32}$$

which, in turn, may be summed over s to obtain

$$D_{lm} = 4U(m)(2l+1)(l-m)!/2^l l(l+1) \times \left(\frac{l-1+m}{2}\right)!\left(\frac{l-1-m}{2}\right)! \tag{A33}$$

for m even.

By a parallel argument, for l even $D_{lm}^c = 0$ and D_{lm}^s may be written without a superscript. It too satisfies (A33). These results may be summarized as

$$D_{lm} = 4(2l+1)(l-m)!U(m)\delta(l+m,2q+1)/2^l l(l+1) \times \left(\frac{l-1+m}{2}\right)!\left(\frac{l-1-m}{2}\right)!, \tag{A34}$$

where q is an integer.

APPENDIX B

Numerical values of the functions may be obtained by using the expansion of the function $P_\nu^m(\cos\theta)$, valid for $0<\theta<\pi$.[8]

$$P_\nu^m(\cos\theta) = \left(\frac{\sin\theta}{2}\right)^m \sum_{s=0}^{\infty}\frac{\Gamma(\nu+1+m+s)}{\Gamma(\nu+1-m-s)} \times \frac{(-1)^s}{s!(s+m)!}\left(\frac{1-\cos\theta}{2}\right)^s. \tag{B1}$$

We first programmed (B1) on an HP-35C hand calculator to obtain results accurate to several places. Knowing $P_\nu^m(\cos\theta)$, the functions $L_\lambda^m(\cos\theta)$ and $M_\nu^m(\cos\theta)$ may be calculated directly by using (23) and (24), and $dL_\lambda^m(\cos\theta)/d\theta$ may be calculated using the identity

$$\frac{dL_\lambda^m}{d\theta} = -(\lambda+1)\frac{L_\lambda^m}{\tan\theta} + (\lambda-m+1)\frac{M_{\lambda+1}^m}{\sin\theta}. \tag{B2}$$

Useful recursion formulas are

$$\begin{matrix} M_\nu^{m+1} \\ L_\nu^{m+1}\end{matrix} = 2m\cot\theta \begin{matrix} L_\nu^m \\ M_\nu^m\end{matrix} - (\nu+m)(\nu-m+1)\begin{matrix} M_\nu^{m-1} \\ L_\nu^{m-1}\end{matrix} \tag{B3}$$

and

$$\begin{matrix} M_{\nu+1}^m \\ L_{\nu+1}^m\end{matrix} = \frac{1}{(\nu-m+1)} \times \left[(2\nu+1)\cos\theta \begin{matrix} L_\nu^m \\ M_\nu^m\end{matrix} - (\nu+m)\begin{matrix} M_{\nu-1}^m \\ L_{\nu-1}^m\end{matrix}\right]. \tag{B4}$$

Initially values of $P_\nu^m(\cos\theta)$ were obtained for $[\nu:(1+m)(0.1)10]$ and $[\theta:1°(1°)89°]$. Root values of M_ν^m and $dL_\nu^m/d\theta$ for $m=0$ and $m=1$ are shown in Fig. 2. The procedure was then repeated using FORTRAN G language on The University of Texas at El Paso's IBM-360 to ten-place accuracy, and all points confirmed with values given in Ref. 11.

The University of Texas at El Paso program was then combined with a root locator algorithm to evaluate roots to ten place accuracy at the 19 cone angles ψ shown in the primary lists of Appendix C. The numerical approach $\psi\to 0$ shown was obtained using the root equation

$$\tan\frac{\pi\nu}{2} = \frac{\mathscr{Y}_0(\theta(l(l+1))^{1/2})}{\mathscr{J}_0(\theta(l(l+1))^{1/2})}, \tag{B5}$$

where $\mathscr{J}_0$ and $\mathscr{Y}_0$ represent cylindrical Bessel and Neumann functions, respectively, and using the calculator.

Root values of ν so obtained were then placed into (B1) and the IBM-360 program used to obtain the results shown in Appendix D, for the special case of $\psi = 1°$. Values of $I_{\nu l}$ in Appendix E were obtained directly from (I-2), without numerical integration. Values of $I_{\nu\nu}$, however, did require use of a numerical integration algorithm (B1), and (4) of Table I.

Appendix F was obtained by using $I_{\nu\nu}$ in Appendix E and obtaining $j_\nu(\sigma)$ from the expansion

$$'j_\nu(\sigma) = \frac{\sqrt{\pi}}{2}\left(\frac{\sigma}{2}\right)^\nu \sum_{s=0}^{\infty}\frac{(-1)^s(\sigma/2)^{2s}}{s!\Gamma(\nu+\frac{3}{2}+s)} \tag{B6}$$

and then finding $J_\nu(\sigma)$ from (12).

All numerical work has since been confirmed using the Fortran H language on The Pennsylvania State University Computing Center's IBM-370.

The HP-34C program utilized to evaluate (B1) follows:

$m=0.$	$(\nu$,STO 1)	$(\theta$,STO 2)	
1. 1	11. RCL 1	21. RCL 1	31. RCL 2
2. RCL 2	12. h LBL A	22. $-$	32. RCL 5
3. h cos	13. 1	23. RCL 4	33. hy^x
4. $-$	14. $+$	24. $\times$	34. $\times$
5. 2	15. STO 4	25. RCL 5	35. RCL 3
6. $\div$	16. RCL 1	26. gx^2	36. $+$
7. STO 2	17. $-$	27. $\div$	37. STO 3
8. 1	18. STO 5	28. RCL 7	38. h PSE
9. STO 3	19. 1	29. $\times$	39. RCL 4
10. STO 7	20. $-$	30. STO 7	40. GTO A

Results were read on RCL 3.

APPENDIX C: ROOT VALUES OF ν AND λ, ψ IN DEGREES

First Root

ψ	$M_\nu(\cos\psi)$	$M^1_\nu(\cos\psi)$	$\dfrac{dL^1_\lambda}{d\psi}(\cos\psi)$
*0°	1.0	2.000 000 000	3.000 000 000
1	1.262 954 024	2.000 911 197	2.998 168 483
2	1.319 814 347	2.003 622 267	2.992 665 960
3	1.365 553 298	2.008 084 527	2.983 548 116
4	1.406 369 289	2.014 240 156	2.971 011 775
5	1.444 484 008	2.022 029 558	2.955 427 410
6	1.480 985 851	2.031 395 310	2.937 339 167
7	1.516 503 611	2.042 284 342	2.917 433 609
8	1.551 441 338	2.054 649 074	2.896 484 410
10	1.620 624 163	2.083 645 601	2.854 594 866
12	1.690 041 573	2.118 126 236	2.817 293 581
14	1.760 654 396	2.157 927 288	2.788 554 311
16	1.833 170 003	2.202 973 048	2.770 595 146
18	1.908 170 247	2.253 268 671	2.764 320 834
21	2.026 463 290	2.338 751 346	2.776 954 627
24	2.153 206 357	2.436 817 899	2.815 071 616
27	2.290 146 521	2.548 404 204	2.877 385 093
30	2.439 211 866	2.674 829 833	2.963 078 409
33	2.602 622 864	2.817 847 128	3.072 097 215
36	2.783 012 956	2.979 721 955	3.205 260 472

* ψ	$M_\nu(\cos\psi)$
10^{-1°	1.1628
10^{-2}	1.1185
10^{-3}	1.0932
10^{-4}	1.0767
10^{-5}	1.0652
10^{-6}	1.0567
10^{-8}	1.0450

Second Root

ψ	$M_\nu(\cos\psi)$			$M_\nu^1(\cos\psi)$			$\frac{dL_\lambda^1}{d\psi}(\cos\psi)$		
0°	3.0			4.000	000	000	5.000	000	000
1	3.323	635	023	4.003	018	842	4.995	412	764
2	3.408	792	057	4.011	853	585	4.981	677	947
3	3.480	410	488	4.026	054	557	4.959	404	322
4	3.546	353	691	4.045	132	592	4.930	263	571
5	3.609	447	465	4.068	627	165	4.896	983	164
6	3.671	078	494	4.096	131	586	4.862	897	552
7	3.732	050	336	4.127	299	086	4.831	274	760
8	3.792	882	288	4.161	840	460	4.804	752	716
10	3.915	488	842	4.240	140	918	4.773	179	415
12	4.040	892	484	4.329	641	388	4.773	224	272
14	4.170	419	778	4.429	491	154	4.802	816	920
16	4.305	101	966	4.539	238	182	4.857	824	606
18	4.445	838	186	4.658	731	348	4.934	396	643
21	4.670	155	684	4.856	487	210	5.083	590	406
24	4.912	930	502	5.077	505	745	5.268	995	091
27	5.177	336	316	5.323	737	373	5.488	160	381
30	5.466	996	648	5.597	956	269	5.741	125	621
33	5.786	177	372	5.903	821	160	6.029	812	247
36	6.140	009	840	6.246	108	497	6.357	771	283

Third Root

ψ	$M_\nu(\cos\psi)$			$M_\nu^1(\cos\psi)$			$\frac{dL_\lambda^1}{d\psi}(\cos\psi)$		
0°	5.0			6.000	000	000	7.000	000	000
1	5.370	105	197	6.006	286	581	6.991	425	007
2	5.480	449	956	6.024	299	328	6.966	025	869
3	5.576	003	415	6.052	455	008	6.926	687	356
4	5.665	740	078	6.089	212	072	6.880	075	073
5	5.752	872	182	6.133	271	729	6.834	977	834
6	5.838	969	859	6.183	600	616	6.799	147	416
7	5.924	941	074	6.239	398	091	6.777	224	978
8	6.011	373	196	6.300	052	314	6.770	720	621
10	6.187	171	603	6.434	193	311	6.800	868	046
12	6.368	665	718	6.583	557	048	6.877	844	394
14	6.557	438	143	6.746	844	503	6.989	952	845
16	6.754	772	476	6.923	496	022	7.129	303	339
18	6.961	836	378	7.113	448	485	7.291	169	285
21	7.293	192	695	7.424	042	290	7.570	807	920
24	7.653	099	721	7.767	442	393	7.891	793	298
27	8.046	109	253	8.146	987	690	8.254	315	347
30	8.477	509	679	8.567	166	410	8.661	058	944
33	8.953	593	052	9.033	722	907	9.116	675	708
36	9.481	976	769	9.553	887	870	9.627	700	794

Fourth Root

ψ	$M_\nu(\cos\psi)$			$M_\nu^1(\cos\psi)$			$\frac{dL_\lambda^1}{d\psi}(\cos\psi)$		
0°	7.0			8.000	000	000	9.000	000	000
1	7.410	226	873	8.010	667	627	8.986	211	284
2	7.544	461	080	8.040	509	575	8.946	210	761
3	7.663	241	993	8.085	815	898	8.888	900	491
4	7.776	369	960	8.143	364	235	8.831	226	255
5	7.887	327	150	8.210	719	660	8.789	441	405
6	7.997	804	564	8.286	124	006	8.771	759	942
7	8.108	776	292	8.368	326	238	8.778	817	933
8	8.220	872	389	8.456	439	665	8.807	573	992
10	8.450	112	907	8.648	077	529	8.915	298	914
12	8.688	042	984	8.857	062	784	9.071	013	712
14	8.936	451	746	9.084	704	568	9.260	834	113
16	9.916	846	707	9.327	878	843	9.477	515	817
18	9.470	647	436	9.587	669	951	9.717	448	016
21	9.909	637	608		> 10			> 10	

Fifth Root

ψ	$M_\nu(\cos\psi)$			$M^1(\cos\psi)$		
0°	9.0			10.000	000	000
1	9.446	643	017		> 10	
2	9.604	085	220			
3	9.745	782	962			
4	9.882	175	226			
5	10.016	936	968			

APPENDIX D: VALUES OF $M_\nu(\cos\theta)$ FOR $\psi = 1°$

$\psi = 1°$

θ	$\nu = 1.262$	954	0243	$\nu = 3.323$	635	0233
1°		$< 10^{-10}$			$\sim 10^{-10}$	
2	0.162	260	5584	0.187	714	7334
3	0.257	081	6503	0.296	877	0784
4	0.324	185	6071	0.373	231	2456
5	0.376	004	6792	0.431	022	1462
6	0.418	066	9846	0.476	544	9264
7	0.453	315	1285	0.513	128	0784
8	0.483	501	1912	0.542	740	8592
10	0.532	834	8235	0.585	692	7923
12	0.571	532	6264	0.611	513	5449
14	0.602	488	5992	0.623	496	7863
16	0.627	417	0994	0.623	702	2477
18	0.647	415	9295	0.613	591	6298
21	0.669	739	6172	0.581	569	1255
24	0.684	284	3657	0.532	233	5250

APPENDIX D (continued)

27	0.692	141	1019	0.468	625	7760
30	0.694	067	0231	0.393	725	7040
33	0.690	621	0462	0.310	530	8851
36	0.682	239	5539	0.222	062	7049
40	0.663	999	3163	0.101	101	8973
45	0.630	839	0439	− 0.045	224	3570
50	0.587	304	3781	− 0.174	570	3302
55	0.534	524	4815	− 0.276	463	2005
60	0.473	596	2113	− 0.343	164	7370
65	0.405	618	4787	− 0.370	230	3843
70	0.331	709	2602	− 0.356	813	0941
75	0.253	012	3164	− 0.305	677	6002
80	0.170	697	5391	− 0.222	919	1141
85	0.085	957	2579	− 0.117	407	2380

APPENDIX D (continued)

$\psi = 1°$

θ	$\nu = 5.370$	105	1974	$\nu = 7.410$	226	8727
1°		$< 10^{-10}$			$< 10^{-10}$	
2	0.202	713	9504	0.212	273	1240
3	0.319	606	7561	0.333	207	6818
4	0.399	696	4972	0.413	586	6081
5	0.458	129	0225	0.468	960	0659
6	0.501	546	8402	0.506	121	2166
7	0.533	447	3023	0.528	684	1391
8	0.555	907	1645	0.538	872	8638
10	0.577	513	0415	0.527	993	1063
12	0.573	311	3518	0.482	727	5712
14	0.547	697	0835	0.410	629	7953
16	0.504	149	8180	0.318	992	3601
18	0.445	847	9211	0.215	179	2871
21	0.337	554	7858	0.052	661	4461
24	0.213	681	9963	− 0.097	348	8006
27	0.084	800	2001	− 0.215	140	9168
30	− 0.038	969	7475	− 0.286	883	9044
33	− 0.148	508	1400	− 0.306	115	3199
36	− 0.236	246	5808	− 0.274	243	7632
40	− 0.310	064	0053	− 0.168	154	9007
45	− 0.325	091	9397	0.014	787	2856
50	− 0.259	316	3344	0.175	911	7239
55	− 0.134	330	5362	0.248	094	6657
60	0.016	502	4759	0.206	014	1167
65	0.155	796	9813	0.074	216	2516
70	0.250	550	1384	− 0.085	536	4145
75	0.279	384	2210	− 0.201	827	9133
80	0.237	066	8715	− 0.224	313	5343
85	0.135	352	9673	− 0.144	791	5271

APPENDIX D (continued)

$\psi = 1°$			
θ	$\nu = 9.446$	643	0.169
1°		$< 10^{-10}$	
2	0.218	106	0884
3	0.340	412	7266
4	0.418	405	9502
5	0.467	725	7562
6	0.495	254	4730
7	0.504	803	9872
8	0.498	927	8909
10	0.448	721	5053
12	0.358	868	1314
14	0.243	148	9032
16	0.115	525	8239
18	− 0.010	394	5959
21	− 0.169	676	2038
24	− 0.266	456	8403
27	− 0.285	795	6745
30	− 0.231	251	6238
33	− 0.122	872	8504
36	0.008	362	3998
40	0.160	474	8434
45	0.229	416	4579
50	0.137	188	2283
55	− 0.042	042	7157
60	− 0.181	790	8100
65	− 0.189	354	1079
70	− 0.065	488	6098
75	0.100	044	1961
80	0.191	876	0055
85	0.147	773	6022

APPENDIX E: EXAMPLE EVALUATION OF INTEGRALS

$\psi = 1°$			$m = 0$						
$\nu =$	1.262	954	0243	3.323	635	0233	5.370	105	1974
$I_{\nu\nu} =$	0.458	631	1361	0.194	573	0789	0.116	495	1668
$I_{\nu 1} =$	0.545	603	0742	0.043	806	5007	0.018	188	0377
$I_{\nu 3} =$	− 0.051	167	0902	0.228	452	8684	0.026	357	4559
$I_{\nu 5} =$	− 0.017	210	1421	− 0.034	595	5027	0.138	905	9110
$I_{\nu 7}$	− 0.008	772	2773	− 0.012	962	6301	− 0.026	769	5043
$I_{\nu 9} =$	− 0.005	335	4918	− 0.007	116	3367	− 0.010	428	3607
$I_{\nu 11} =$	− 0.003	588	4906	− 0.004	560	4705	− 0.005	930	1007
$\nu =$	7.410	226	8727	9.446	643	0169			
$I_{\nu\nu} =$	0.079	294	7911	0.057	645	2754			
$I_{\nu 1} =$	0.010	184	2647	0.006	540	9851			
$I_{\nu 3} =$	0.012	198	6631	0.007	289	9134			

$I_{\nu 5} =$	0.018	965	6803	0.009	187	5305
$I_{\nu 7} =$	0.096	772	9489	0.014	753	9302
$I_{\nu 9} =$	−0.022	044	4934	0.072	316	6208
$I_{\nu 11} =$	−0.008	728	1148	−0.018	792	8094
$I_{\nu 13} =$	−0.005	061	790	−0.007	485	2265

APPENDIX F: VALUES OF $J_\nu(\sigma)/\nu(\nu+1)j_\nu(\sigma)I_{\nu\nu}$ AT THE FIVE ROOTS OF $M_\nu(\cos 1°)$, FOR SIX SELECTED VALUES OF $\sigma = ka$. $m = 0$.

ka	1.262 954	0243	3.323 635	0233	5.370 105	1974	7.410 226	8727	9.446 643	0169
0.1	17.250 488	57	15.459 654	41	15.983 050	65	17.017 464	67	18.362 852	3
0.8	2.045 846	653	1.903 091	756	1.983 455	853	2.118 224	302	2.289 025	68
1.6	0.843 044	447	0.905 737	954	0.969 480	875	1.045 360	643	1.134 818	31
2.4	0.328 900	228	0.550 329	109	0.620 979	924	0.681 397	636	0.745 665	65
3.1	−0.030 323	860	0.372 344	643	0.456 292	318	0.512 777	777	0.567 012	09
4.0	−0.730 655	518	0.210 073	547	0.320 328	783	0.377 798	951	0.425 932	87

APPENDIX G: RECIPROCITY OF ADMITTANCE

For $m = 0$, the boundary conditions which led to (III-1)–(III-3) may be put in the form of (G1) and (G2):

$$\frac{I(a)}{a} = 2iG\sum_l D_{l0}x_l P_l(\cos\psi), \tag{G1}$$

$$\frac{V(a)}{a} = \frac{\pi l(l+1)}{\eta G P_l(\cos\psi)}\left[\sum_\nu\sum_{l'}\frac{J_\nu I_{l\nu}I_{l'\nu}l'(l'+1)D_{l'0}x_{l'}}{\nu(\nu+1)j_\nu I_{\nu\nu}} - D_{l0}X_l I_{ll}\right], \tag{G2}$$

where

$$x_l = j_l + \beta_{l0}h_l, \qquad X_l = J_l + \beta_{l0}H_l. \tag{G3}$$

Equations (G1) and (G2) also describe biconical transmitting antennas, where

$$x_l = \beta_{l0}h_l, \qquad X_l = \beta_{l0}H_l. \tag{G4}$$

An iterative solution to (G2) is used to find x_l where, for example, we may put $V(a) = 1$ as a normalization procedure. Since the correct complex number of set x_l is a unique solution to the equation, the same set applies both to cases (G3) and (G4). Therefore $I(a)$, as calculated from (G1), is identical in the two cases and, therefore, so is $Y(a)$. This verifies the equality of admittance during reception and transmission.

The radiative coefficients β_{l0} in the two cases are quite different. Dropping the $(l,0)$ coefficients, writing subscript r and t for reception and transmission, respectively, and ′ and ″ for real and imaginary parts, respectively, shows that, for each value of l,

$$\beta'_r = \beta'_t - \sin^2\gamma \tag{G5}$$

and

$$\beta''_r = \beta''_t - \sin\gamma\cos\gamma, \tag{G6}$$

where

$$\tan\gamma = j(\sigma)/y(\sigma). \tag{G7}$$

Since the power radiated in the two cases is proportional to the sum of the squares of the real and imaginary parts of the β coefficients, we note that

$$(\beta_t'^2 + \beta_t''^2) - (\beta_r'^2 + \beta_r''^2) = (2\beta'_t - 1)\sin^2\gamma + \beta''_t\sin\gamma\cos\gamma. \tag{G8}$$

[1]E. Hallen, Nova Acta Regiae Soc. Sci. Ups. Serv. IV 2, 3 (1938).
[2]R. W. P. King, "Cylindrical Antennas and Arrays," in *Antenna Theory*, Part I, edited by R. E. Collin and F. J. Zucker (McGraw-Hill, New York, 1969), Chap. 9.
[3]R. W. P. King and C. W. Harrison, Jr., Proc. IRE **32**, 18 (1944).
[4]T. Morita, Proc. IRE **38**, 898 (1950).
[5]R. E. Collin, "Receiving Antenna," in Ref. 2, Chap. 4.
[6]S. Schelkunoff, *Advanced Antenna Theory* (Wiley, New York, 1952), pp. 32–49.
[7]C. T. Tai, J. Appl. Phys. **20**, 1076 (1949).
[8]J. R. Wait, "Electromagnetic Radiation from Conical Structures," in Ref. 2, Chap. 12.
[9]S. Schelkunoff, *Applied Mathematics for Engineers and Scientists*, 2nd ed. (Van Nostrand, New York, 1965).
[10]A. Einstein, "The Quantum Theory of Radiation," in *The World of the Atom*, edited by H. A. Boorse and L. Motz (Basic Books, New York, 1966), Vol. 2, p. 901—last paragraph. [Originally Z. Phys. **18**, 121 (1917)].
[11]"Tables Numeriques des Fonctions Associes de Legendre," Collection Technique et Scientifique du C.N.E.T., 1959 (unpublished).

ERRATA

Erratum: Linear wave conversion in a warm, unmagnetized, collisionless plasma [J. Math. Phys. 22, 2692 (1981)]

G. A. Kriegsmann
Department of Engineering Science and Applied Mathematics, Northwestern University, Evanston, Illinois 60201

(Received 21 January 1982; accepted for publication 29 January 1982)

PACS numbers: 52.35. − g, 99.10. + g

The acknowledgments to the following government agencies were inadvertently omitted: DOE DE-AC02-78ER0-4650 and ONR N00014-80-C-0196.

Erratum: Nonuniqueness in the inverse source problem in acoustics and electromagnetics [J. Math. Phys. 18, 194 (1977)]

Norman Bleistein and Jack K. Cohen
Department of Mathematics, University of Denver, Denver, Colorado 80210

(Received 14 December 1981; accepted for publication 23 December 1981)

PACS numbers: 02.30.Rz, 43.90. + v, 41.90. + l, 99.10. + g

Reference 1 (N. Bleistein and N. N. Bojarski, "Recently developed Formulations of the Inverse Problem in Acoustics and Electromagnetics," DRI Report #MS-R-7501, NTIS #AD/A-003, 1974) should be replaced by N. N. Bojarski, "Inverse Scattering," Naval Air Systems Command Report to Contract N00019-73-C-0312, October 1973, Sec. 11, pp. 3–6.

