A generalization of the construction of Ilamed and Salingaros

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

(Received 28 January 1982; accepted for publication 30 September 1982)

Ilamed and Salingaros construct the real and complex algebras with three anticommuting elements which can arise in physics. It is shown here that the "algebra of color" can be similarly constructed with six anticommuting elements. As a consequence of this construction, these algebras are all simple, quadratic algebras.

PACS numbers: 02.10.Tq

1. INTRODUCTION

Ilamed and Salingaros[1] have determined the structure of all possible algebras with three anticommuting elements over the real and complex fields $\mathbb{R}$ and $\mathbb{C}$ that can arise in physical descriptions. In Sec. 2 we continue the study of algebras with anticommuting elements by looking at all algebras with at least three anticommuting elements and determine some properties of those algebras. In Sec. 3 we show that the "algebra of color" fits into this larger class of algebras that generalizes and contains those algebras constructed by Ilamed and Salingaros.

2. THE CONSTRUCTION

Consider an algebra A over $\mathbb{R}$ or $\mathbb{C}$ with a basis $1, e_1, e_2, \ldots, e_n$ where 1 is the scalar unit. These elements are defined to anticommute. Define a scalar square for each element,

$$e_i e_i = a_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where the a_i's are each equal to $+1$ or -1.

Ilamed and Salingaros,[1] considering only the case $n = 3$, construct all algebras with three anticommuting elements which can arise in physics. They are (i) the quaternions, (ii) the dihedral Clifford algebras N_1, which is related to the real 2-spinors, and (iii) the algebra of Pauli matrices S_1, which is related to the complex 2-spinors.

Theorem 1: Each algebra A constructed above is quadratic, that is, if $x \in A$, then $\{1, x, x^2\}$ is a linearly dependent set.

Proof: Let

$$x = \alpha_0 1 + \alpha_\gamma e_\gamma \tag{2}$$

with $1 \leqslant \gamma \leqslant n$; summation over repeated Greek indices is understood.

$$x^2 = \alpha_0^2 1 + 2\alpha_0(\alpha_\gamma e_\gamma) + \alpha_\gamma^2 a_\gamma 1, \tag{3}$$

$$x^2 = 2\alpha_0(\alpha_0 1 + \alpha_\gamma e_\gamma) + (\alpha_\gamma^2 a_\gamma - \alpha_0^2)1, \tag{4}$$

$$x^2 = 2\alpha_0 x + (\alpha_\gamma^2 a_\gamma - \alpha_0^2)1. \tag{5}$$

We can rewrite (5) as

$$x^2 - 2t(x)x + q(x)1 = 0, \tag{6}$$

where $t(x) = \alpha_0$ and $q(x) = \alpha_\gamma^2 a_\gamma - \alpha_0^2$ are scalars. The quantities $t(x)$ and $q(x)$ are called the trace and norm of x, respectively. The trace is a linear functional on A.[2] The norm $q(x)$ defines a symmetric bilinear form $q(x,y)$ on A via

$$q(x,y) = q(x + y) - q(x) - q(y). \tag{7}$$

Any quadratic algebra A is power associative, i.e., the subalgebra generated by an element $x \in A$ is associative; all quadratic algebras are Jordan admissible. The classical treatment of quadratic algebras is given by Braun and Koecher.[3]

Let x be any element of an algebra A. The right multiplication by x, R_x is defined by

$$R_x : a \to ax \quad \text{for all} \quad a \in A. \tag{8}$$

Similarly, the left multiplication L_x is defined by

$$L_x : a \to xa \quad \text{for all} \quad a \in A. \tag{9}$$

Each of R_x and L_x is a linear operator on A for all $x \in A$. Let $M(A)$ denote the (associative) algebra consisting of all finite sums of products of right and left multiplications of A; $M(A)$ is often called the associative multiplication algebra of A.

A subset of B of A is called an ideal of A if it is an invariant subspace under $M(A)$. An algebra A is called simple in case 0 and A itself are the only deals of A. Thus A is a simple algebra if and only if $M(A) \neq 0$ is an irreducible set of linear operators. Knowing that an algebra is simple often helps to determine other aspects of its structure. For example, if A is a simple, finite-dimensional associative algebra, then there is a division ring D and a positive integer n such that A is isomorphic to the ring of $n \times n$ matrices over D.[4]

Theorem 2: Each algebra A, $n \geqslant 2$, constructed above, is simple.

Proof: Suppose A is not simple and $x \in A$ is contained in a proper ideal of A. Then

$$x = \alpha_0 1 + \alpha_\gamma e_\gamma. \tag{10}$$

For some $k = 1, 2, \ldots, n$, $\alpha_k \neq 0$. Then

$$e_k x + x e_k = 2\alpha_k a_k 1 + 2\alpha_0 e_k = z \tag{11}$$

is an element of the ideal. Hence $\alpha_0 \neq 0$. Pick some $e_l, l \neq k$. Then

$$e_l z + z e_l = 4\alpha_k a_k e_l \tag{12}$$

is in the ideal. Hence $e_l e_l$ is in the ideal. But then 1 is the ideal.

A simple, finite-dimensional alternative algebra is either associative or an octonion algebra.[5] Since an octonion algebra is eight-dimensional over its center, all four-dimensional alternative algebras constructed in this manner must be associative. Hence the algebras of Ilamed and Salingaros must be associative.

 0022-2488/83/020221-03$02.50

3. THE ALGEBRA OF COLOR

Domokos and Kövesi-Domokos[6] have constructed an algebra of dynamical variables which describe the color properties of quarks and leptons. We will review this "algebra of color," give its multiplication table, and, finally, show that one can choose a basis for the algebra of color such that the algebra is a generalization of the construction of Ilamed and Salingaros.

As is usual, it is assumed that quarks are triplets (3), antiquarks are antitriplets ($\bar{3}$), and leptons are singlets under the color group $SU(3)_c$.[7] The multiplication rules should be such that mesons and baryons, being observables, are both singlets under $SU(3)_c$. Multiplication must then obey the triality rule, which can be symbolically written as

$$(3)\times(\bar{3})\sim(1), \quad (3)\times(3)\sim(\bar{3}). \tag{13}$$

In addition, Domokos and Kövesi-Domokos introduced an exact superselection rule between hadrons and leptons.

Define the fundamental dynamical variable ψ by

$$\psi = u_0 l + u_\alpha q^\alpha \tag{14}$$

with $1 \leqslant \alpha \leqslant 3$; summation over repeated Greek (color) indices understood. The coefficients l and q^α are to represent leptonic and quark variables, respectively. These are anticommuting Fermi variables with space-time and flavor labels suppressed.[8] All the color properties of the algebra must be realized in the basis elements u_0 and u_α.

1. Quarks and antiquarks (leptons and antileptons) are distinct. Both leptons and antileptons transform as $\sim(1)$ under $SU(3)_c$. Quarks (antiquarks) transform as $\sim(3)$ [$\sim(\bar{3})$]. Hence we have a vector space with basis elements u_α spanning (3) and $\bar{u}_\alpha$ spanning $\bar{3}$ of $SU(3)_c$. Write the conjugate of ψ as

$$\bar{\psi} = \bar{u}_0 \bar{l} + \bar{u}_\alpha \bar{q}_\alpha, \tag{15}$$

where both u_0 and $\bar{u}_0$ are singlets under $SU(3)_c$.

2. We assume that there are exactly two superselection sectors, i.e., hadrons and leptons, and that every observable Ω may be written as a direct sum

$$\Omega = L\Omega L + H\Omega H, \tag{16}$$

such that

$$L^2 = L, \quad H^2 = H, \tag{17}$$

$$L^+ = L, \quad H^+ = H, \tag{18}$$

$$LH = HL = 0, \tag{19}$$

where the projections L and H are singlets under $SU(3)_c$. Since we want to identify l with a leptonic variable, u_0 may be identified with the projection L, provided that $\bar{u}_0 = u_0$; this is permissible since both u_0 and $\bar{u}_0$ are color singlets. Hence we can write

$$E = L + H, \tag{20}$$

where E is the unit element of the color algebra. We have the obvious

$$L = u_0 = \bar{u}_0, \tag{21}$$

and

$$Lu_\alpha = u_\alpha L = L\bar{u}_\alpha = \bar{u}_\alpha L = 0. \tag{22}$$

3. The triality rule requires that

$$u_\alpha \bar{u}_\beta = \bar{u}_\alpha u_\beta = \delta_{\alpha\beta} H, \tag{23}$$

where $\delta_{\alpha\beta}$ is the Kronecker delta symbol. Likewise, baryons (antibaryons) are observables, and keeping (23) in mind, we define

$$\bar{u}_\alpha \bar{u}_\beta = \epsilon_{\alpha\beta\gamma} u_\gamma; \; u_\alpha u_\beta = \epsilon_{\alpha\beta\gamma} \bar{u}_\gamma, \tag{24}$$

where $\epsilon_{\alpha\beta\gamma}$ is the totally antisymmetric unit tensor.

4. Any observable and a quark (antiquark) must carry the same color as the quark (antiquark) does. Hence

$$Hu_\alpha = u_\alpha H = u_\alpha \tag{25}$$

and

$$H\bar{u}_\alpha = \bar{u}_\alpha H = \bar{u}_\alpha. \tag{26}$$

In summary we give multiplication Table I.

Theorem 3: The algebra in Table I has a basis $\{H, e_1, e_2, e_3, e_4, e_5, e_6\}$ with the e_i's anticommuting and $e_i^2 = \pm 1$ for each i.

Proof: Let

$$e_\alpha = (1/\sqrt{2})(u_\alpha + \bar{u}_\alpha), \tag{27}$$

$$e_{\alpha+3} = (1/\sqrt{2})(u_\alpha - \bar{u}_\alpha), \tag{28}$$

$\alpha = 1, 2, 3$. Direct computation verifies that

$$e_\alpha^2 = H, \quad \alpha = 1,2,3, \tag{29}$$

$$e_\alpha^2 = -H, \quad \alpha = 4,5,6, \tag{30}$$

$$e_\alpha e_\beta = -e_\beta e_\alpha \quad \text{if } \alpha \neq \beta. \tag{31}$$

All of the four-dimensional algebras and the algebra of Table I constructed in this manner are flexible; that is, $(xy)x = x(yx)$ for all x and y in the algebra. However, not all algebras constructed in this manner are flexible; Braun and Koecher provide the example. Take as a basis $\{1, e_1, e_2, e_3\}$ and multiplication

$$e_1^2 = e_2^2 = e_3^2 = 1, \tag{32}$$

$$e_1 e_2 = e_3 = -e_2 e_1, \tag{33}$$

$$e_1 e_3 = e_2 e_3 = e_3 e_2 = e_3 e_1 = 0. \tag{34}$$

For a further discussion of the algebra of color see Domokos and Kövesi-Domokos[9] and Wene.[10]

4. CONCLUSION

This paper, along with Ref. 1, shows that each of the quanterions, the dihedral Clifford algebra N_1, the algebra of Pauli matrices S, and the algebra of color can be constructed as an algebra over $\mathbb{R}$ or $\mathbb{C}$ with a basis $1, e_1, \ldots, e_n$ where the e_i's anticommute and $e_i^2 = a_i$, $i = 1, 2, \ldots, n$ (no summation intended) and each a_i is equal to ± 1.

Salingaros[11] extends the construction of algebras with three anticommuting elements and a unit given in Ref. 1 to

TABLE I.

	H	u_β	$\bar{u}_\beta$
H	H	u_β	$\bar{u}_\beta$
u_α	u_α	$\epsilon_{\alpha\beta\nu}\bar{u}_\nu$	$\delta_{\alpha\beta} H$
$\bar{u}_\alpha$	$\bar{u}_\alpha$	$\delta_{\alpha\beta} H$	$\epsilon_{\alpha\beta\nu} u_\nu$

the case where the underlying field is the ring Ω (that is, the Clifford algebra generated over $\mathbb{R}$ by the elements $\{1, \omega\}$ with $\omega^2 = +1$).

[1]Y. Ilamed and N. Salingaros, "Algebras with three anticommuting elements. I," J. Math. Phys. **22**, 2091 (1981).

[2]R. D. Schafer, *Introduction to Nonassociative Algebras* (Academic, New York, 1966), p. 49.

[3]H. Braun and M. Koecher, *Jordan-algebren* (Springer-Verlag, Berlin, 1966).

[4]N. Jacobson, *Structure of Rings* (Amer. Math. Soc., Colloquium Publications XXXVII, Providence, RI, 1968), p. 39.

[5]E. Kleinfeld, "Simple Alternative Rings," Ann. Math. **58**, 544–547 (1953).

[6]G. Domokos and S. Kövesi-Domokos, "The algebra of color," J. Math. Phys. **19**, 1477–1481 (1978); M. Günaydin and F. Gürsey, "Quark Statistics and Octonions," Phys. Rev. D **9**, 3387 (1974), present another possible "algebra of color," one based on the octonions.

[7]We assume that the color group is isomorphic to SU(3). For other possibilities, see S. Okubo, "Constraint on Color Gauge Groups," Phys. Rev. D **16**, 3535 (1977) and "Gauge Groups Without Triangular Anomaly," Phys. Rev. D **16**, 3528 (1977).

[8]This definition is based on the exactness of the color group, which, in the color-quark scheme of Gell-Mann, is equivalent to a theory of para-Fermi quarks. See M. Günaydin and F. Gürsey, Ref. 6.

[9]G. Domokos and S. Kövesi-Domokos, "Towards an Algebraic Quantum Chromodynamics," Phys. Rev. D **19**, 2984 (1979).

[10]G. P. Wene, "An Example of a Flexible, Jordan-admissible Algebra of Current use in Hadron Physics," Hadronic J. **1**, 944–954 (1978).

[11]N. Salingaros, "Algebras with three anticommuting elements. II. Two algebras over a singular field," J. Math. Phys. **22**, 2096 (1981). Salingaros constructs two algebra over Ω, the Clifford algebras Ω_1 and Ω_2, and discusses the use of these algebras in descriptions of vector fields.

Representations of the groups Sp(n,R) and Sp(n) in a U(n) basis

Anatoli U. Klimyk
Institute for Theoretical Physics, Kiev-130, USSR

(Received 20 August 1981; accepted for publication 30 October 1981)

The explicit expressions for the infinitesimal operators and the (finite) matrix elements with respect to a U(n) basis are obtained for the representations of the most degenerate series of Sp(n,R) and for the irreducible unitary representations of Sp(n) with the highest weights (M,0,...,0).

PACS numbers: 02.20.Qs

I. INTRODUCTION

The groups Sp(n, R) and Sp(n) have found wide applications in physics.[1-6] The infinitesimal operators and the finite matrix elements of the group representations are of great importance for physical applications. We need different bases for different problems. In this article we consider the representations of Sp(n, R) and Sp(n) in a U(n) basis. Here n denotes a rank of the groups Sp(n, R) and Sp(n). In some papers these groups are denoted by Sp(2n, R) and Sp(2n), respectively.

The formulas for the infinitesimal operators derived here correspond to any reductions
$\mathrm{Sp}(n,R)\supset \mathrm{U}(n)\supset G'\supset G''\supset\cdots$ and
$\mathrm{Sp}(n)\supset \mathrm{U}(n)\supset G'\supset G''\supset\cdots$. The formulas contain the Clebsch–Gordan coefficients of a tensor product of the representations of U(n) with the highest weights $(m_1, 0,...,0, m_2)$ and (2, 0,...,0) for the reduction $\mathrm{U}(n)\supset G'\supset G''\supset\cdots$. These Clebsch–Gordan coefficients are well known[7,8] for the Gel'fand–Zetlin basis.

The matrix elements of the representations of Sp(n, R) and Sp(n) are found in the basis corresponding to the reduction

$$\mathrm{U}(n)\supset \mathrm{U}(n-1)\supset\cdots\supset \mathrm{U}(1).$$

In order to find them in other bases we have to use the matrix elements of the representations of U(n) with the highest weights $(m_1, 0,...,0, m_2)$ for this other basis. The matrix elements of some representations of U(n) in a $\mathrm{U}(n-p)\otimes \mathrm{U}(p)$ basis are obtained in Ref. 9, and in a SO(n) basis in Ref. 10.

To find the infinitesimal operators of the representations of Sp(n, R) in a U(n) basis we use Lemma 5.2 of Ref. 11. The infinitesimal operators are used to calculate the intertwining operators for the representations. In turn, they are used to find the infinitesimal operators of the representations of Sp(n). The method of calculation of the infinitesimal operators and the matrix elements of the unitary irreducible representations of a compact Lie group with help of infinite dimensional representations of a corresponding noncompact Lie group are described in Refs. 12–14.

II. THE MOST DEGENERATE SERIES REPRESENTATIONS OF THE GROUP Sp(n, R)

The group Sp(n, R) consists of all matrices of GL(2n, R) which leave invariant the form

$$x_1\wedge x_{n+1}+x_2\wedge x_{n+2}+\cdots+x_n\wedge x_{2n}.$$

The Lie algebra sp(n, R) of Sp(n, R) consists of all matrices

$$\begin{pmatrix} X_1 & X_2 \\ X_3 & -X_1^{\mathrm{T}}\end{pmatrix},$$

where X_1, X_2, X_3 are real $n\times n$ matrices such that X_2 and X_3 are symmetric. Here T denotes a transposition.

A maximal compact subgroup of Sp(n,R) is U(n). This subgroup is imbedded into Sp(n,R) in the following manner. If $A+iB\in \mathrm{U}(n)$, A and B are real, then

$$A+iB\to\begin{pmatrix} A & B \\ -B & A\end{pmatrix}\in \mathrm{Sp}(n,R). \tag{1}$$

If $\mathrm{sp}(n,R)=\mathrm{u}(n)+\mathfrak{p}$ is a Cartan decomposition of sp(n, R) then $\mathfrak{p}$ consists of the matrices[15]

$$\begin{pmatrix} Z_1 & Z_2 \\ Z_2 & -Z_1\end{pmatrix}, \tag{2}$$

where $iZ_1\in \mathrm{u}(n)$ and is a pure imaginary $n\times n$ matrix and Z_2 is a symmetric real $n\times n$ matrix.

Sometimes it is convenient to consider the algebra sp$'$(n, R) which is isomorphic to sp(n, R). It is obtained by the transformation φ: $g\to\sigma g\sigma^{-1}$, $g\in\mathrm{sp}(n,R)$, where

$$\sigma=2^{-1/2}\begin{pmatrix} E_n & iE_n \\ E_n & -iE_n\end{pmatrix}$$

and E_n is a unit $n\times n$ matrix. Under φ the matrices (1) transform into the matrices

$$\begin{pmatrix} A-iB & 0 \\ 0 & A+iB\end{pmatrix} \tag{3}$$

and the matrices (2) into

$$\begin{pmatrix} 0 & Z_1+iZ_2 \\ Z_1-iZ_2 & 0\end{pmatrix}. \tag{4}$$

An Iwasawa decomposition for Sp(n, R) can be taken in the form $\mathrm{Sp}(n,R)=ANK$, where $K=\mathrm{U}(n)$, $A=\exp\mathfrak{a}$ and $\mathfrak{a}$ consists of the matrices

$$\mathrm{diag}(\omega_1,...,\omega_n,\ -\omega_1,...,-\omega_n),\quad \omega_j\in R. \tag{5}$$

The notation diag (...) is used for diagonal matrices.

Let us consider the subgroup $P=AN\cdot \mathrm{U}(n-1)$ of Sp(n, R), where U($n-1$) is imbedded into Sp(n, R) as

$$\mathrm{U}(n-1)\to\mathrm{diag}(\mathrm{U}(n-1),1)$$

using the imbedding (1). The subgroup P differs from the maximal parabolic subgroup of Sp(n, R) by the discrete subgroup Z_2 consisting of two elements. The subgroup P can also be represented as $P=A_1N_1\cdot \mathrm{Sp}(n-1,R)$, where

 0022-2488/83/020224-09$02.50

$\mathrm{Sp}(n-1, R)$ is obtained from $\mathrm{Sp}(n, R)$ by a deletion of the nth and $2n$th rows and columns. The subgroup A_1 is defined as $\exp \mathfrak{a}_1$, where $\mathfrak{a}_1 \subset \mathrm{sp}(n, R)$ consists of the matrices

$$\mathrm{diag}(0,..,0, a, 0,..., -a), \quad a \in R. \tag{6}$$

It is clear that the matrices of A_1 have the form

$$\mathrm{diag}(1,...,1, t, 1,...,1, t^{-1}), \quad 0 \neq t \in R. \tag{7}$$

The subgroup $N_1 \subset N$ is generated by root vectors corresponding to positive roots of the pair[16] $(\mathrm{sp}(n, R), \mathfrak{a}_1)$.

For $h_1 n_1 m \in A_1 N_1$. $\mathrm{Sp}(n-1, R) \equiv P$ the correspondence

$$h_1 n_1 m \to \exp(\lambda (\ln h_1)), \tag{8}$$

where λ is a linear form on $\mathfrak{a}_1$, defines a one dimensional linear representation of P. It is clear that λ is given by one complex number μ. If h_1 is of the form (7) then

$$\exp(\lambda (\ln h_1)) = t^{\mu}.$$

The representation (8) of P induces the representation of $\mathrm{Sp}(n, R)$. We denote it by π_μ. It acts on the Hilbert space $L^2_0(K)$, $K = \mathrm{U}(n)$, which consists of all functions from $L^2(K)$ satisfying the condition

$$f(mk) = f(k), \quad m \in \mathrm{U}(n-1). \tag{9}$$

The operators $\pi_\mu(g)$, $g \in \mathrm{Sp}(n, R)$ are given by

$$\pi_\mu(g) f(k) = \exp(\lambda (\ln h_1)) f(k_g), \tag{10}$$

where h_1 and k_g are defined in the following manner. k_g is determined by the Iwasawa decomposition $kg = h' n k_g$, $h' \in A$, $n \in N$, $k_g \in K$; h_1 is defined by the decomposition $h' = h_2 h_1$, $h_2 \in A_2$, $h_1 \in A_1$ [here $A_2 = \exp \mathfrak{a}_2$ is a subgroup of A such that $\mathfrak{a}_2$ consists of the matrices (5) with $\omega_n = 0$].

The representations π_μ are reducible. Every π_μ is decomposed into two representations of $\mathrm{Sp}(n, R)$, which are induced by representations of the maximal parabolic subgroup. This decomposition will be given in the following section.

III. INFINITESIMAL OPERATORS OF THE REPRESENTATIONS π_μ

Let $B(.,.)$ be a Cartan–Killing form on $\mathrm{sp}(n, R)$ and θ a Cartan involution. Then $\langle x, y \rangle = -cB(x, \theta y)$, $c > 0$ and fixed, is a scalar product on $\mathrm{sp}(n, R)$. The adjoint representation of $\mathrm{Sp}(n, R)$ in $\mathrm{sp}(n, R)$ will be denoted by Ad. In order to evaluate the infinitesimal operators of the representations π_μ of $\mathrm{Sp}(n, R)$ we use Lemma 5.2 of Ref. 11. For our case it can be formulated as follows.

Lemma. The infinitesimal operators $d\pi_\mu(Y)$, $Y \in \mathfrak{p}_c$ ($\mathfrak{p}_c$ is the complexification of $\mathfrak{p}$), of the representations π_μ of $\mathrm{Sp}(n, R)$ act upon the infinitely differentiable functions of $L^2_0(K)$ as

$$\begin{aligned} d\pi_\mu(Y) f(k) &= \langle (\mathrm{Ad}\, k) Y, H \rangle \lambda(H) f(k) - \langle (\mathrm{Ad}\, k) Y, \rho \rangle f(k) \\ &\quad + \tfrac{1}{2} [Q, \langle (\mathrm{Ad}\, k) Y, h \rangle] f(k), \end{aligned} \tag{11}$$

where H is a normalized element of $\mathfrak{a}_1$, h is an element of $\mathfrak{a}_1$ such that $\alpha(h) = 1$ [α is a simple restricted root[16] of the pair $(\mathrm{sp}(n, R), \mathfrak{a}_1)$] Q is identical to the operator Q_1 of the formula (5) in Ref. 17, and ρ is half the sum of the positive restricted roots of the pair $(\mathrm{sp}(n, R), \mathfrak{a}_1)$ represented as an element of $\mathfrak{a}_1$.

We need an orthonormal basis of $L^2_0(K)$. According to the Peter–Weyl theorem the matrix elements of the irreducible unitary representations of $\mathrm{U}(n)$ with the highest weights $(m_1, 0,...,0, m_2)$, $m_1 \geqslant 0$, $m_2 \leqslant 0$, which are left invariant with respect to $\mathrm{U}(n-1)$, can be taken as a basis of $L^2_0(K)$. We denote these representations of $\mathrm{U}(n)$ by $[m_1, 0,...,0, m_2] \equiv [m_1, m_2] \equiv D^{m_1, m_2}$. In the space of the representation $[m_1, m_2]$ we choose two orthonormal bases: the Gel'fand–Zetlin basis, i.e., the basis corresponding to the reduction $\mathrm{U}(n) \supset \mathrm{U}(n-1) \supset \cdots \supset \mathrm{U}(1)$, and an arbitrary orthonormal basis. The elements of this latter basis will be denoted by $|\Sigma\rangle$. The Gel'fand–Zetlin basis element which corresponds to the Gel'fand–Zetlin pattern

$$\begin{bmatrix} m_1 & & 0 & \cdots & 0 & & m_2 \\ & 0 & & \cdots & & 0 & \\ & & & \cdots\cdots & & & \\ & & & 0 & & & \end{bmatrix}, \tag{12}$$

where the first row is a highest weight of the representation of $\mathrm{U}(n)$, will be denoted by $|\Omega\rangle$. It is clear that $|\Omega\rangle$ is invariant with respect to $\mathrm{U}(n-1)$. The functions

$$\{\dim[m_1, m_2]\}^{1/2} \langle \Omega | D^{m_1, m_2}(k) | \Sigma \rangle \tag{13}$$

for all m_1, m_2, and Σ, constitute an orthonormal basis for $L^2_0(K)$. The basis elements (13) will be denoted by $|m_1, m_2, \Sigma\rangle$.

We shall find the infinitesimal operators $d\pi_\mu(Y)$ in the basis $|m_1, m_2, \Sigma\rangle$. The derivation is similar to the one given in Ref. 17 for the representations of the group $\mathrm{U}(p, q)$. Therefore, we omit here the details.

The scalar product $\langle .,. \rangle$ on $\mathfrak{p}$ can be given as

$$\langle X, Y \rangle = \tfrac{1}{2} \mathrm{Tr}\, X Y^{\mathrm{T}}. \tag{14}$$

Therefore, the matrices h and H of the lemma are

$$h = H = e_{nn} - e_{2n, 2n}. \tag{15}$$

Here e_{ij} is a matrix with matrix elements $(e_{ij})_{st} = \delta_{is} \delta_{jt}$. A direct evaluation shows that $\rho = nh$, where h is defined by (15). This evaluation can be done with a help of the Araki diagram (see pp. 30–32 in Ref. 16). Now we can write

$$\begin{aligned} &\langle (\mathrm{Ad}\, k) Y, H \rangle \lambda(H) - \langle (\mathrm{Ad}\, k) Y, \rho \rangle \\ &\quad = (\mu - n) \langle (\mathrm{Ad}\, k) Y, h \rangle. \end{aligned} \tag{16}$$

Since we consider the degenerate series of representations the chain (2) of subgroups of Ref. 17 (see also Chap. 5 in Ref. 11) reduces to

$$\mathrm{U}(n) = K \equiv K_1 \supset K_2 = \mathrm{U}(n-1).$$

Moreover, between K_1 and K_2 there is the subgroup K^1_2 [see the chain (3) in Ref. 17], and

$$K^1_2 = \mathrm{diag}(\mathrm{U}(n-1), \mathrm{U}(1)).$$

This information is utilized to find eigenvalues of the operator Q. The operator Q acts upon the states $|m_1, m_2, \Sigma\rangle$ as

$$Q |m_1, m_2, \Sigma\rangle = q(m) |m_1, m_2, \Sigma\rangle, \tag{17}$$

where $q(m) \equiv q(m_1, m_2)$ is a number.

From (16) and (17) we have

$$d\pi_\mu(Y)|m_1, m_2, \Sigma\rangle = (\mu - n + \tfrac{1}{2}Q - \tfrac{1}{2}q(m)) \times \langle (\mathrm{Ad}\, k)Y, h\rangle |m_1, m_2, \Sigma\rangle. \quad (18)$$

Now we consider the expression $\langle (\mathrm{Ad}\, k)Y, h\rangle |m_1, m_2, \Sigma\rangle$. The space $\mathfrak{p}_c$ is a carrier space of the representation of $\mathfrak{u}(n)$ with respect to the action ad b, $b \in \mathfrak{u}(n)$. This representation is a direct sum of two irreducible representations of $\mathfrak{u}(n)$ with the highest weights (2, 0,...,0) and (0,...,0, -2). In order to prove it we use the algebra $\mathrm{sp}'(n, R)$ for which $\mathrm{U}(n)$ and $\mathfrak{p}$ are realized by the matrices (3) and (4), respectively. The matrices

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

are transformed under the representation (2, 0,...,0), and the matrices

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

under the representation (0,...,0, -2). We shall find the infinitesimal operators of the representations π_μ for the matrices $e_{2n,n}$ and $e_{n,2n}$. If we consider the realization (2) for $\mathfrak{p}$ then these matrices correspond to

$$e_{2n,n} \to e_{nn} - e_{2n,2n} + i(e_{n,2n} + e_{2n,n}) \equiv Y_+,$$
$$e_{n,2n} \to e_{nn} - e_{2n,2n} - i(e_{n,2n} + e_{2n,n}) \equiv Y_-.$$

Thus, we shall obtain the operators $d\pi_\mu(Y_+)$ and $d\pi_\mu(Y_-)$. The other noncompact operators can be obtained by commutation of $d\pi_\mu(Y_+)$ and $d\pi_\mu(Y_-)$ with compact infinitesimal operators. We can also find them by utilizing the Clebsch–Gordan coefficients for $\mathrm{U}(n)$ (see Ref. 17).

The matrices Y_+ and Y_- are elements of the carrier spaces of the representations of $\mathrm{U}(n)$. They correspond to the Gel'fand–Zetlin patterns:

$$Y_+ \to \begin{bmatrix} 2 & 0 & \cdots & 0 \\ & 0 & \cdots & 0 \\ & & \cdots & \\ & & 0 & \end{bmatrix}, \quad Y_- \to \begin{bmatrix} 0 & 0 & \cdots & -2 \\ & 0 & \cdots & 0 \\ & & \cdots & \\ & & 0 & \end{bmatrix}.$$

These patterns will be denoted by (Y_+), (Y_-), respectively.

The expressions

$$\langle (\mathrm{Ad}\, k)Y_+, h\rangle = \langle (\mathrm{Ad}\, k)Y_+, Y_+\rangle,$$
$$\langle (\mathrm{Ad}\, k)Y_-, h\rangle = \langle (\mathrm{Ad}\, k)Y_-, Y_-\rangle,$$

are matrix elements of the representations of $\mathrm{U}(n)$ with the highest weights (2, 0,...,0) and (0,...,0, -2), respectively. The basis elements $|m_1, m_2, \Sigma\rangle$ are the functions (13). Therefore,

$$\langle (\mathrm{Ad}\, k)Y_+, h\rangle |m_1, m_2, \Sigma\rangle = \sum_{m_1', m_2'} \left(\frac{\dim[m_1, m_2]}{\dim[m_1', m_2']} \right)^{1/2} \langle m_1, m_2, \Omega; (Y_+)|m_1', m_2', \Omega\rangle \langle m_1', m_2', \Sigma | m_1, m_2, \Sigma; (Y_+)\rangle |m_1', m_2', \Sigma\rangle, \quad (19)$$

$$\langle (\mathrm{Ad}\, k)Y_-, h\rangle |m_1, m_2, \Sigma\rangle = \sum_{m_1', m_2'} \left(\frac{\dim[m_1, m_2]}{\dim[m_1', m_2']} \right)^{1/2} \langle m_1, m_2, \Omega; (Y_-)|m_1', m_2', \Omega\rangle \langle m_1', m_2', \Sigma | m_1, m_2, \Sigma; (Y_-)\rangle |m_1', m_2', \Sigma\rangle, \quad (20)$$

where $\langle \cdots | \cdots \rangle$ are Clebsch–Gordan coefficients of the tensor products

$$[m_1, 0,...,0, m_2] \otimes [2, 0,...,0],$$
$$[m_1, 0,...,0, m_2] \otimes [0,...,0, -2]$$

of the representations of $\mathrm{U}(n)$. Decomposing these tensor products we find that in the relation (19)

$$(m_1', m_2') = (m_1 + 2, m_2),\ (m_1, m_2 + 2),\ (m_1 + 1, m_2 + 1)$$

and in (20)

$$(m_1', m_2') = (m_1 - 2, m_2),\ (m_1, m_2 - 2),\ (m_1 - 1, m_2 - 1).$$

This information is used for evaluation of eigenvalues of the operator $Q - q(m)$ from (18). We evaluate them utilizing the considerations given in Sec. 4 in Ref. 17. Substituting the relations (19) and (20) into (18) and taking into account the formulas for eigenvalues of $Q - q(m)$ we obtain

$$\begin{aligned} d\pi_\mu(Y_+)|m_1, m_2, \Sigma\rangle = {} & (\mu + m_1 - m_2)\left(\frac{\dim[m_1, m_2]}{\dim[m_1 + 2, m_2]} \right)^{1/2} \langle m_1, m_2, \Omega; (Y_+)|m_1 + 2, m_2, \Omega\rangle \\ & \times \langle m_1 + 2, m_2, \Sigma | m_1, m_2, \Sigma; (Y_+)\rangle |m_1 + 2, m_2, \Sigma\rangle + (\mu - m_1 + m_2 - 2n + 2) \\ & \times \left(\frac{\dim[m_1, m_2]}{\dim[m_1, m_2 + 2]} \right)^{1/2} \langle m_1, m_2, \Omega; (Y_+)|m_1, m_2 + 2, \Omega\rangle \\ & \times \langle m_1, m_2 + 2, \Sigma | m_1, m_2, \Sigma; (Y_+)\rangle |m_1, m_2 + 2, \Sigma\rangle \\ & + (\mu - n)\left(\frac{\dim[m_1, m_2]}{\dim[m_1 + 1, m_2 + 1]} \right)^{1/2} \langle m_1, m_2, \Omega; (Y_+)|m_1 + 1, m_2 + 1, \Omega\rangle \\ & \times \langle m_1 + 1, m_2 + 1, \Sigma | m_1, m_2, \Sigma; (Y_+)\rangle |m_1 + 1, m_2 + 1, \Sigma\rangle, \end{aligned} \quad (21)$$

$$d\pi_\mu(Y_-)|m_1, m_2, \Sigma\rangle = (\mu - m_1 + m_2 - 2n + 2)\left(\frac{\dim[m_1, m_2]}{\dim[m_1 - 2, m_2]}\right)^{1/2}\langle m_1, m_2, \Omega;(Y_-)|m_1 - 2, m_2, \Omega\rangle$$
$$\times\langle m_1 - 2, m_2, \Sigma\,|m_1, m_2, \Sigma;(Y_-)\rangle|m_1 - 2, m_2, \Sigma\rangle$$
$$+ (\mu + m_1 - m_2)\left(\frac{\dim[m_1, m_2]}{\dim[m_1, m_2 - 2]}\right)^{1/2}\langle m_1, m_2, \Omega;(Y_-)|m_1, m_2 - 2, \Omega\rangle$$
$$\times\langle m_1, m_2 - 2, \Sigma\,|m_1, m_2, \Sigma;(Y_-)\rangle|m_1, m_2 - 2, \Sigma\rangle$$
$$+ (\mu - n)\left(\frac{\dim[m_1, m_2]}{\dim[m_1 - 1, m_2 - 1]}\right)^{1/2}\langle m_1, m_2, \Omega;(Y_-)|m_1 - 1, m_2 - 1, \Omega\rangle$$
$$\times\langle m_1 - 1, m_2 - 1, \Sigma\,|m_1, m_2, \Sigma;(Y_-)\rangle|m_1 - 1, m_2 - 1, \Sigma\rangle \tag{22}$$

In this formula Clebsch–Gordan coefficients with Ω are known.[7,8] They correspond to Gel'fand–Zetlin bases. Clebsch–Gordan coefficients with Σ are known if $|m_1, m_2, \Sigma\rangle$ are the Gel'fand–Zetlin basis elements.

The infinitesimal operators $d\pi_\mu(Y_+)$, $d\pi_\mu(Y_-)$ (and, therefore, other noncompact infinitesimal operators) change the number $m_1 + m_2$ by ± 2. Thus, the representations π_μ of Sp(n, R) are decomposed into a direct sum of two representations π_μ^+ and π_μ^-. For π_μ^+ the numbers $m_1 + m_2$ are even, for π_μ^- they are odd. It can be shown that the representations π_μ^+ and π_μ^- of Sp(n, R) are induced by one dimensional representations of the maximal parabolic subgroup related with the subgroup P.

The representations π_μ^+ and π_μ^-, for which $\mu - n$ are pure imaginary, are unitary. They constitute the principal most degenerate unitary series of Sp(n, R).

IV. STRUCTURE OF THE REPRESENTATIONS $\pi_\mu^\pm$

The multiplicity of the irreducible representations of U(n) in π_μ^+ and π_μ^- does not exceed 1. Thus we can determine the set of irreducible representations in the set of representations π_μ^+, π_μ^-, μ a complex number. We can also investigate the structure (the composition series) of the reducible representations π_μ^+, π_μ^-. The proofs are the same as in the case of the groups U(n, 1) and $SO_0(n, 1)$ in Ref. 11. Therefore, we list the theorems without giving proofs.

Theorem 1: The representation π_μ^+ is completely irreducible if and only if $\mu \neq 0, -2, -4, -6,...$ and $\mu \neq 2n, 2n + 2, 2n + 4,...$. The representation π_μ^- is completely irreducible if and only if $\mu \neq -1, -3, -5,...$ and $\mu \neq 2n + 1, 2n + 3, 2n + 5,...$.

Theorem 2: If $\mu = p; p = 0, -2, -4, -6,...$, then $\pi_{\mu=p}^+$ contains two completely irreducible representations of Sp(n, R), namely, the finite dimensional representation D_{-p} with the highest weight $(-p, 0,...,0)$ and the infinite dimensional representation, denoted by D^{d}_{-p+2}. These representations $\pi_{\mu=p}^+$ are indecomposable and the finite dimensional representations are realized in an invariant subspace. If $\mu = p; p = 2n, 2n + 2, 2n + 4,...$, then $\pi_{\mu=p}^+$ is indecomposable and contains two completely irreducible representations of Sp(n, R), namely, the finite dimensional representation D_{p-2n} and the representation D^{d}_{p-2n-2}. The latter one is realized in an invariant subspace.

The representations π_p^+, π_{2n-p}^+, $p = 0, -2, -4, -6,...$ contain the same completely irreducible representations of Sp(n, R).

Theorem 3: If $p = -1, -3, -5,...$, then the representation $\pi_{\mu=p}^-$ and the representation π_{2n-p}^- contain two completely irreducible representations of Sp(n, R), namely, the finite dimensional representation D_{-p} with the highest weight $(-p, 0,...,0)$ and the infinite dimensional representation, denoted by D^{d}_{-p+2}. The representations π_p^- and π_{2n-p}^- are indecomposable. For π_p^- the representation D_{-p} and for π_{2n-p}^- the representation D^{d}_{-p+2} are realized in invariant subspaces.

Using the infinitesimal operators (21) and (22), we can define which irreducible representations of U(n) are included in the representations D_{-p} and D^{d}_{-p+2} of Sp(n, R). We saw that the representations π_μ^+ and π_μ^- of Sp(n, R) under restriction upon U(n) are decomposed onto the irreducible representations of U(n) with the highest weights $(m_1, 0,...,0, m_2)$ for which $m_1 + m_2$ are even and odd, respectively. The finite dimensional irreducible representation D_M of Sp(n, R) with highest weight $(M = -p, 0,...,0)$ under restriction upon U(n) are decomposed onto those and only those irreducible representations of U(n) which have the highest weights $(m_1, 0,...,0, m_2)$ for which $m_1 - m_2 \leqslant M$ and $m_1 + m_2$ have the same parity (evenness) as M does. Now we can define irreducible representations of U(n) which are contained in the representations D^{d}_{-p+2} of Sp(n, R).

V. INFINITESIMAL OPERATORS OF THE UNITARY REPRESENTATIONS OF Sp(n) IN A U(n) BASIS

Let us consider the finite dimensional subrepresentations of the representations $\pi_\mu^\pm$ of Sp(n, R) (see Theorems 2 and 3). The Lie algebra of Sp(n, R) has the Cartan decomposition $\mathrm{sp}(n, R) = \mathrm{u}(n) + \mathfrak{p}$. The corresponding compact Lie algebra sp(n) has the decomposition

$$\mathrm{sp}(n) = \mathrm{u}(n) + i\mathfrak{p}. \tag{23}$$

Therefore, if we multiply infinitesimal operators $Y \in \mathfrak{p}$ for finite dimensional representations of Sp(n, R) by i then we obtain them for finite dimensional representations of Sp(n). Let

$$J_+ = iY_+, \quad J_- = iY_-,$$

where Y_+ and Y_- are taken from (21) and (22). The infinitesimal operators J_+ and J_- for the finite dimensional representations D_M, $M = -p$, of Theorems 2 and 3, given by the formulas (21) and (22), do not satisfy the unitarity condition

$J_+^* = -J_-$. This condition can be satisfied for the new basis. This basis can be found with the help of the intertwining operators Π_μ for the representations π_μ and $\pi_{-\mu+2n}$ of Sp(n, R):

$$\Pi_\mu \pi_\mu = \pi_{-\mu+2n}\Pi_\mu \tag{24}$$

(see Chap. 5 in Ref. 11). In the basis $|m_1, m_2, \Sigma\rangle$ the operators Π_μ are diagonal and their matrix elements $a_{m_1m_2}$ do not depend[11] on Σ. Let us consider the relation (24) for $d\pi_\mu(Y_+)$ and $d\pi_\mu(Y_-)$. Then taking matrix elements of both sides one obtains the relations for $a_{m_1m_2}$:

$$a_{m_1+2,m_2}(\mu + m_1 - m_2) = a_{m_1m_2}(-\mu + m_1 - m_2 + 2n),$$
$$a_{m_1,m_2+2}(\mu - m_1 + m_2 - 2m + 2) = a_{m_1m_2}(-\mu - m_1 + m_2 + 2),$$
$$-a_{m_1+1,m_2+1} = a_{m_1m_2}.$$

Therefore,

$$a_{m_1+2k,m_2} = a_{m_1m_2}\prod_{j=0}^{k-1}\frac{-\mu + m_1 - m_2 + 2n + 2j}{\mu + m_1 - m_2 + 2j}, \tag{25}$$

$$a_{m_1,m_2+2k} = a_{m_1m_2}\prod_{j=1}^{k}\frac{-\mu - m_1 + m_2 + 2j}{\mu - m_1 + m_2 - 2n + 2j}, \tag{26}$$

$$a_{m_1+k,m_2+k} = (-1)^k a_{m_1m_2}. \tag{27}$$

Now we can fix some initial value of $a_{m_1m_2}$ and find all matrix elements $a_{m_1m_2}$ of Π_μ. The operators J_+, J_- satisfy the unitarity condition for the basis

$$|m_1, m_2, \Sigma\rangle' = \Pi_\mu^{-1/2}|m_1, m_2, \Sigma\rangle \tag{28}$$

(see Sec. 5 in Ref. 11). Taking into account the relations (25)–(28) we obtain the formulas for the infinitesimal operators of the unitary irreducible representations of Sp(n) with highest weights (M, 0,...,0):

$$J_+|m_1, m_2, \Sigma\rangle' = [(-M + m_1 - m_2)(M + m_1 - m_2 + 2n)]^{1/2}K^{m_1m_2}_{m_1+2,m_2}|m_1 + 2, m_2, \Sigma\rangle'$$
$$+ [(-M - m_1 + m_2 - 2n + 2)(M - m_1 + m_2 + 2)]^{1/2}K^{m_1m_2}_{m_1m_2+2}|m_1, m_2 + 2, \Sigma\rangle'$$
$$- i(M + n)K^{m_1m_2}_{m_1+1,m_2+1}|m_1 + 1, m_2 + 1, \Sigma\rangle', \tag{29}$$

$$J_-|m_1, m_2, \Sigma\rangle' = -[(M + m_1 - m_2 + 2n - 2)(-M + m_1 - m_2 - 2)]^{1/2}K^{m_1m_2}_{m_1-2,m_2}|m_1 - 2, m_2, \Sigma\rangle'$$
$$- [(M + m_1 - m_2 + 2n)(-M + m_1 - m_2)]^{1/2}K^{m_1m_2}_{m_1m_2-2}|m_1, m_2 - 2, \Sigma\rangle'$$
$$- i(M + n)K^{m_1m_2}_{m_1-1,m_2-1}|m_1 - 1, m_2 - 1, \Sigma\rangle'. \tag{30}$$

Here $K^{\cdots}_{\cdots}$ denote the products of Clebsch–Gordan coefficients (multiplied by dimensionality multiplier) of Eqs. (21) and (22). Other infinitesimal operators can be obtained by commutation of the operators J_+ and J_- with infinitesimal operators corresponding to the subgroup U(n).

VI. MATRIX ELEMENTS OF THE REPRESENTATIONS π_μ IN A U(n) BASIS

Now we shall find the matrix elements of the representations $\pi_\mu^\pm$ in a U(n) basis. The Gel'fand–Zetlin basis will be taken as a U(n) basis. We shall use the formula (10) for the representations π_μ and the formula

$$\langle f_1, \pi_\mu(g)f_2\rangle = \int [f_1(k)]^*\pi_\mu(g)f_2(k)dk \tag{31}$$

for the scalar product in $L_0^2(K)$, $K = \mathrm{U}(n)$.

Any element $g\in\mathrm{Sp}(n, R)$ can be represented as

$$g = khk', \quad k', k\in\mathrm{U}(n), \quad h\in A. \tag{32}$$

The element $h\in A$ is represented as the product of the matrices (7), taken for the subgroups Sp(j, R), $j = 1, 2,...,n$. Since the matrix elements of the representations of U(n) in the Gel'fand–Zetlin basis are known, then we have to find the matrix elements of the operators $\pi_\mu(h_1)$, corresponding to the elements $h_1\in A_1$ given by Eq. (7).

According to Eq. (9) elements of $L_0^2(K)$ can be considered as functions on the complex sphere $S^{n-1} = \mathrm{U}(n-1)\backslash\mathrm{U}(n)$. Let us introduce a parametrization on S^{n-1}. For this aim we consider the decomposition of the elements k of U(n) in the form [see Eq. (7.39) in Ref. 11]

$$k = h'\beta_n(\theta_n)a_n(\varphi_n)\tilde{h}, \quad h'\in\mathrm{U}(n-1),$$
$$a_n(\varphi_n) = \mathrm{diag}(1,...,1, e^{-\varphi_n}),$$
$$\beta_n(\theta_n) = \mathrm{diag}(1,...,1, R(\theta_n)),$$
$$\tilde{h} = \left[\prod_{r=1}^{n-2}\beta_{n-r}(\theta_{n-r})a_{n-r}(\varphi_{n-r})\right]a_1(\varphi_1),$$
$$0\leqslant\varphi_i<2\pi, \quad 0\leqslant\theta_i\leqslant\pi/2.$$

Here $R(\theta)$ denotes a 2×2 matrix with the matrix elements $a_{11} = a_{22} = \cos\theta, a_{12} = -a_{21} = -\sin\theta$. Thus elements of $L_0^2(K)$ can be considered as functions of $\theta_n, \varphi_n, \theta_{n-1}, \varphi_{n-1},...,\theta_2, \varphi_2, \varphi_1$ or as functions of elements of U(n) of the form

$$k' = \left[\prod_{r=0}^{n-2}\beta_{n-r}(\theta_{n-r})a_{n-r}(\varphi_{n-r})\right]a_1(\varphi_1). \tag{33}$$

This parametrization of S^{n-1} corresponds to the reduction

$$\mathrm{U}(n)\supset\mathrm{U}(n-1)\supset\cdots\supset\mathrm{U}(1).$$

Now we find the action of $\pi_\mu(h_1)$ on the functions $f(\theta_n, \varphi_n,...,\theta_2, \varphi_2, \varphi_1)$. The formula (10) has to be used. Therefore, we need the explicit form of the matrices $n\in N$. Since $h_1\in A_1$, then the Iwasawa decomposition $kg = h'nk_g$, $g = h_1$, contains elements n of the subgroup N_1 of N. Let us construct this subgroup.

The linear forms on α [see Eq. (5)],

$$\pm\omega_i \pm \omega_j, \quad i \neq j, \quad \pm 2\omega_i, \quad i = 1, 2, \ldots, n,$$

are restricted roots[16] of the pair $(\mathrm{sp}(n, R), \mathfrak{a})$. The roots $\omega_i + \omega_j$, $i > j$, $2\omega_i$ can be taken as positive roots. All roots have the multiplicity 1 (see pp. 30–32 in Ref. 16 and Table 3 in Ref. 11). Let us construct root vectors corresponding to the positive roots $\omega_n \pm \omega_{n-1}$, $2\omega_{n-1}$, $2\omega_n$. A direct verification shows that they coincide with the matrices

$$E_{\omega_n + \omega_{n-1}} = e_{1,2n} + e_{2,2n-1}, \quad E_{2\omega_{n-1}} = e_{n-1,2n-1},$$
$$E_{\omega_n - \omega_{n-1}} = e_{n,n-1} - e_{n+1,2n-1}, \quad E_{2\omega_n} = e_{n,2n}.$$

The following one-parameter subgroups of $\mathrm{Sp}(n, R)$ correspond to them:

$$N_{\omega_n + \omega_{n-1}} = e + s(e_{1,2n} + e_{2,2n-1}),$$
$$N_{\omega_n - \omega_{n-1}} = e + t(e_{n,n-1} - e_{n+1,2n-1}),$$
$$N_{2\omega_{n-1}} = e + pe_{n-1,2n-1},$$
$$N_{2\omega_n} = e + qe_{n,2n},$$

where e is a unit matrix and s, t, p, q are real numbers. A product of this subgroups coincides with the subgroup N_1. Its elements have the form

$$\begin{bmatrix} 1 & & & & & & & & & \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & 1 & 0 & & & & q & s \\ & & & t & 1 & & & & tq+s & ts+p \\ & & & & & 1 & & & & \\ & & & & & & \ddots & & & \\ & & & & & & & 1 & & \\ & & & & & & & & 1 & -t \\ & & & & & & & & 0 & 1 \end{bmatrix},$$

where other matrix elements are equal to 0.

If k' is of the form (33) and $h_1 \in A_1$ then

$$k'h_1 = \beta_n(\theta_n)a_n(\varphi_n)h_1\left[\prod_{r=1}^{n-1}\beta_{n-r}(\theta_{n-r})a_{n-r}(\varphi_{n-r})\right]a_1(\varphi_1).$$

For the product $\beta_n(\theta_n)a_n(\varphi_n)h_1$ we have

$$\beta_n(\theta_n)a_n(\varphi_n)h_1 = dn_1\beta_n(\theta'_n)a_n(\varphi'_n), \quad n_1 \in N_1,$$
$$d = \mathrm{diag}(1, \ldots, 1, t'', t'), \quad t'', t' \in R.$$

Comparing the matrix elements for both sides of this relation we obtain

$$\cos\varphi' = \cos\varphi(\cos^2\varphi + t^4\sin^2\varphi)^{-1/2}, \tag{34}$$
$$\cos\theta' = \cos\theta\,[\cos^2\theta + t^2(\cos^2\varphi + t^4\sin^2\varphi)^{-1}\sin^2\theta\,]^{-1/2}, \tag{35}$$
$$\sin\theta' = t(\cos^2\varphi + t^4\sin^2\varphi)^{-1/2}\sin\theta\cos^{-1}\theta\cos\theta', \tag{36}$$
$$t' = t(\cos^2\varphi + t^4\sin^2\varphi)^{-1/2} \times [\cos^2\theta + t^2(\cos^2\varphi + t^4\sin^2\varphi)^{-1}\sin^2\theta\,]^{-1/2}. \tag{37}$$

We have omitted the index n at φ and θ in Eqs. (34)–(37). It follows that

$$\pi_\mu(h_1)f(\theta_n, \varphi_n, \theta_{n-1}, \varphi_{n-1}, \ldots, \varphi_1) = t'^{\mu}f(\theta'_n, \varphi'_n, \theta_{n-1}, \varphi_{n-1}, \ldots, \varphi_1), \tag{38}$$

where $t', \theta'_n, \varphi'_n$ are defined by (34)–(37). We shall use this formula.

Elements of the Gel'fand–Zetlin basis of the representation of $U(n)$ with highest weight $(m_1, 0, \ldots, 0, m_2)$ are denoted by the patterns

$$\alpha = \begin{bmatrix} m_1 & & 0 & & \cdots & & 0 & & m_2 \\ & m'_1 & & 0 & \cdots & 0 & & m'_2 & \\ & & & & \cdots\cdots & & & & \\ & & & & m_1^{(n-1)} & & & & \end{bmatrix}. \tag{39}$$

Let α_0 denote the pattern α with

$$m'_1 = m'_2 = \cdots = m_1^{(n-1)} = 0.$$

All the matrix elements (multiplied by a dimensionality multiplier)

$$(\dim[\,m_1, m_2\,])^{1/2}D^{m_1,m_2}_{\alpha_0\alpha}(k) \equiv |m_1, m_2, \alpha\rangle \tag{40}$$

of all the representations of $U(n)$ with highest weights $(m_1, 0, \ldots, 0, m_2)$ are the orthonormal basis of $L^2_0(K)$. For the matrix elements (40) we have

$$D^{m_1m_2}_{\alpha_0\alpha}(\beta_n(\theta)a_n(\varphi)) \equiv D^{m_1m_2}_{0,(m'_1m'_2)}(\theta, \varphi) = e^{i\varphi(-m_1-m_2+m'_1+m'_2)}d^{m_1m_2}_{0,(m'_1m'_2)}(\theta), \tag{41}$$

and there are two expressions for $d(\theta)$:

$$d^{m_1m_2}_{0,(m'_1m'_2)}(\theta) = (\cos\theta)^{m'_1+m'_2-m_1-m_2}\sum_{k=m'_1}^{m_1} N(m_1, m_2, m'_1, m'_2, k)(\sin\theta)^{2k-m'_1-m'_2}, \tag{42}$$

$$d^{m_1m_2}_{0,(m'_1m'_2)}(\theta) = (\cos\theta)^{m_1+m_2-m'_1-m'_2}\sum_{k=m_2}^{m'_2} N'(m_1, m_2, m'_1, m'_2, k)(\sin\theta)^{m'_1+m'_2-2k}. \tag{43}$$

The explicit expressions for N can be taken from Eq. (46) of Ref. 13 and for N' from Eq. (47) of Ref. 13 (see also Ref. 18).

The matrix elements of the operators $\pi_\mu(h_1)$ [and therefore, of $\pi^{\pm}_\mu(h_1)$] are defined by the formula

$$\langle m_1, m_2, \alpha|\pi_\mu(h_1)|\tilde{m}_1, \tilde{m}_2, \tilde{\alpha}\rangle = d^{\mu}_{(m_1m_2)(\tilde{m}_1\tilde{m}_2)(m'_1m'_2)}(t) = \frac{(n-1)}{\pi}\frac{(\dim[\,m_1, m_2\,]\cdot\dim[\,\tilde{m}_1, \tilde{m}_2\,])^{1/2}}{\dim[m'_1, m'_2\,]}\int_0^{2\pi}\int_0^{\pi/2} d\varphi d\theta \sin^{2n-3}\theta \times\cos\theta\,[D^{m_1m_2}_{0,(m'_1m'_2)}(\theta, \varphi)]^{*}t'^{\mu}D^{\tilde{m}_1\tilde{m}_2}_{0,(m'_1m'_2)}(\theta', \varphi'), \tag{44}$$

where t', θ', φ' have to be taken from Eqs. (34)–(37). The expressions (42) and (43) can be used for D functions in (44). If Eq. (42) is taken then we have that

$$d^{\mu}_{(m_1m_2)(\tilde m_1\tilde m_2)(m_1'm_2')}(t)=Q\sum_{k=m_1'}^{m_1}\sum_{k'=m_1'}^{m_1}N(m_1,m_2,m_1',m_2',k)N(\tilde m_1,\tilde m_2,m_1',m_2',k')t^{\mu-m_1'-m_2'+2k'}$$
$$\times\int_0^{2\pi}\int_0^{\pi/2}d\varphi\,d\theta\,e^{i\varphi(m_1+m_2-m_1'-m_2')}(\cos^2\varphi+t^4\sin^2\varphi)^{(-\mu+\tilde m_1+\tilde m_2-2k')/2}(\cos\varphi+it^2\sin\varphi)^{-\tilde m_1-\tilde m_2+m_1'+m_2'}$$
$$\times(\sin\theta)^{2(k+k'-m_1'-m_2'+n-1)-1}(\cos\theta)^{2(m_1'+m_2')-m_1-m_2-\tilde m_1-\tilde m_2+1}$$
$$\times[\cos^2\theta+t^2(\cos^2\varphi+t^4\sin^2\varphi)^{-1}\sin^2\theta]^{(-\mu+\tilde m_1+\tilde m_2-2k')/2}. \tag{45}$$

Here we have denoted the expression preceding the integral of Eq. (44) by Q. There is the multiplier

$$(\cos\varphi+it^2\sin\varphi)^{-\tilde m_1-\tilde m_2+m_1'+m_2'}$$

on the right hand side of Eq. (45). It appears because of the decomposition $\exp i\varphi'=\cos\varphi'+i\sin\varphi'$ and Eq. (34). The number $-\tilde m_1-\tilde m_2+m_1'+m_2'$ can be negative. Then this multiplier is not convenient to work with. We have to apply Eq. (34) to the decomposition $\exp(-i\varphi')=\cos\varphi'-i\sin\varphi'$. As a result we obtain a somewhat different expression in Eq. (45). We do not write it down here. The reader can easily obtain it.

If the expression (43) is taken for D function (41) then we have that

$$d^{\mu}_{(m_1m_2)(\tilde m_1\tilde m_2)(m_1'm_2')}(t)=Q\sum_{k=m_2}^{m_2'}\sum_{k'=\tilde m_2}^{m_2'}N'(m_1,m_2,m_1',m_2',k)N'(\tilde m_1,\tilde m_2,m_1',m_2',k')t^{\mu+m_1'+m_2'-2k'}$$
$$\times\int_0^{2\pi}\int_0^{\pi/2}d\varphi\,d\theta\,e^{i\varphi(m_1+m_2-m_1'-m_2')}(\cos^2\varphi+t^4\sin^2\varphi)^{k'-m_1'-m_2'+(-\mu+\tilde m_1+\tilde m_2)/2}$$
$$\times[\cos^2\theta+t^2(\cos^2\varphi+t^4\sin^2\varphi)^{-1}\sin^2\theta]^{(-\mu-\tilde m_1-\tilde m_2+2k')/2}(\cos\varphi+it^2\sin\varphi)^{-\tilde m_1-\tilde m_2+m_1'+m_2'}$$
$$\times(\sin\theta)^{2(k+k'+m_1'+m_2'+n-1)-1}(\cos\theta)^{m_1+m_2+\tilde m_1+\tilde m_2-2(m_1'+m_2')+1}. \tag{46}$$

If in Eq. (44) the expression (42) is taken for $D(\theta,\varphi)$ and the expression (43) for $D(\theta',\varphi')$ then we have

$$d^{\mu}_{(m_1m_2)(\tilde m_1\tilde m_2)(m_1'm_2')}(t)=Q\sum_{k=m_1'}^{m_1}\sum_{k'=\tilde m_2}^{m_2'}N(m_1,m_2,m_1',m_2',k)N'(\tilde m_1,\tilde m_2,m_1',m_2',k')t^{\mu+m_1'+m_2'-2k'}$$
$$\times\int_0^{2\pi}\int_0^{\pi/2}d\varphi\,d\theta\,e^{i\varphi(m_1+m_2-m_1'-m_2')}(\cos^2\varphi+t^4\sin^2\varphi)^{k'-m_1'-m_2'+(-\mu+\tilde m_1+\tilde m_2)/2}$$
$$\times[\cos^2\theta+t^2(\cos^2\varphi+t^4\sin^2\varphi)^{-1}\sin^2\theta]^{k'+(-\mu-\tilde m_1-\tilde m_2)/2}(\cos\varphi+it^2\sin\varphi)^{-\tilde m_1-\tilde m_2+m_1'+m_2'}$$
$$\times(\sin\theta)^{2(n+k-k'-1)-1}(\cos\theta)^{\tilde m_1+\tilde m_2-m_1-m_2+1}. \tag{47}$$

If the expression (43) is taken for $D(\theta,\varphi)$ and the expression (42) for $D(\theta',\varphi')$ then

$$d^{\mu}_{(m_1m_2)(\tilde m_1\tilde m_2)(m_1'm_2')}(t)=Q\sum_{k=m_2}^{m_2'}\sum_{k'=m_1'}^{\tilde m_1}N'(m_1,m_2,m_1',m_2',k)N(\tilde m_1,\tilde m_2,m_1',m_2',k')t^{\mu-m_1-m_2+2k'}$$
$$\times\int_0^{2\pi}\int_0^{\pi/2}d\varphi d\theta\,e^{i\varphi(m_1+m_2-m_1'-m_2')}(\cos^2\varphi+t^4\sin^2\varphi)^{-k'+(-\mu+\tilde m_1+\tilde m_2)/2}(\cos\varphi+it^2\sin\varphi)^{-\tilde m_1-\tilde m_2+m_1'+m_2'}$$
$$\times[\cos^2\theta+t^2(\cos^2\varphi+t^4\sin^2\varphi)^{-1}\sin^2\theta]^{-k'+(-\mu+\tilde m_1+\tilde m_2)/2}(\sin\theta)^{2(n-k+k'-1)-1}(\cos\theta)^{m_1+m_2-\tilde m_1-\tilde m_2+1}. \tag{48}$$

The remark which was made for Eq. (45) is valid for Eqs. (46)–(48).

We do not calculate here the integrals of Eqs. (45)–(48). They can be calculated (at least for partial values of t) with the help of the formulas 3.681(1) and 3.682 of Ref. 19. Let us show it for the integral of Eq. (48). It can be represented as

$$I=t^{-2k'+(-\mu+\tilde m_1+\tilde m_2)/2}\int_0^{2\pi}\int_0^{\pi/2}d\varphi\,d\theta\,e^{i\varphi(m_1+m_2-m_1'-m_2')}(\cos\varphi+it^2\sin\varphi)^{-\tilde m_1-\tilde m_2+m_1'+m_2'}$$
$$\times(\sin\theta)^{2(n-k+k'-1)-1}(\cos\theta)^{m_1+m_2-\tilde m_1-\tilde m_2+1}(t^{-2}(\cos^2\varphi+t^4\sin^2\varphi)\cos^2\theta+\sin^2\theta)^{-k'+(-\mu+\tilde m_1+\tilde m_2)/2}. \tag{49}$$

If $m_1+m_2-\tilde m_1-\tilde m_2+1\geqslant 0$, and for all φ,t satisfies the condition

$$t^{-2}(\cos^2\varphi+t^4\sin^2\varphi)-1<1,$$

i.e., $\frac{1}{2}<t^2<2$, then the integral over θ can be calculated using the formulas of Ref. 19, mentioned above. Now we have to expand the hypergeometric function into the hypergeometric series, then to invert the order of summation and integration over φ and to integrate summands.

VII. THE MATRIX ELEMENTS OF THE UNITARY REPRESENTATIONS OF Sp(n) IN A U(n) BASIS

The matrix elements of the unitary representations of Sp(n) with highest weights (M, 0,...,0) can be found by the method discribed in Ref. 14. This method presupposes utilization of the integrals (45)–(48). The matrix elements will correspond to the reduction

$$\mathrm{Sp}(n)\supset\mathrm{U}(n)\supset\mathrm{U}(n-1)\supset\cdots\supset\mathrm{U}(1).$$

According to this method we find the matrix elements of the finite dimensional irreducible representations of Sp(n, R) with highest weights (M, 0,...,0), which are subrepresentations of the representations π_{-M} of Sp(n, R). We saw that these finite dimensional representations of Sp(n,R) are decomposed onto irreducible representations of U(n) with highest weights (m_1, 0,...,0, m_2) for which $m_1 - m_2 \leqslant M$ and $m_1 - m_2$ has the same parity as M does.

For the matrix elements $D^{M}_{(m_1m_2)(\tilde{m}_1\tilde{m}_2)(m_1'm_2')}(t)$ of the finite dimensional representations of Sp(n, R) with highest weights (M, 0,...,0) we have

$$D^{M}_{(m_1m_2)(\tilde{m}_1\tilde{m}_2)(m_1'm_2')}(t) = d^{\mu=-M}_{(m_1m_2)(\tilde{m}_1\tilde{m}_2)(m_1'm_2')}(t). \tag{50}$$

If

$$m_1 + m_2 - \tilde{m}_1 - \tilde{m}_2 \geqslant 0, \quad m_1' + m_2' - \tilde{m}_1 - \tilde{m}_2 \geqslant 0 \tag{51}$$

then the formula (48) can be used for calculation of the matrix elements (50). The integral in (48) can be represented in the form (49). The powers $-k' + (-\mu + \tilde{m}_1 + \tilde{m}_2)/2$ in (49) are a non-negative integer for the matrix elements (50). Therefore,

$$[t^{-2}(\cos^2\varphi + t^4\sin^2\varphi)\cos^2\theta + \sin^2\theta]^{-k'+a}$$
$$= \sum_{p=0}^{-k'+a}\sum_{q=0}^{p}\binom{-k'+a}{p}\binom{p}{q}t^{4q-2p}(\sin\theta)^{2(-k'+a-p)}(\cos\theta)^{2p}(\sin\varphi)^{2q}(\cos\varphi)^{2(p-q)},$$

$$(\cos\varphi + it^2\sin\varphi)^s = \sum_{r=0}^{s}\binom{s}{r}i^r t^{2r}\sin^r\varphi\cos^{s-r}\varphi,$$

where the notations

$$(M + \tilde{m}_1 + \tilde{m}_2)/2 = a, \quad m_1' + m_2' - \tilde{m}_1 - \tilde{m}_2 = s \tag{52}$$

are introduced. Therefore, for the integral in (49) we have

$$I' = \sum_{p=0}^{-k'+a}\sum_{q=0}^{p}\sum_{r=0}^{s} t^{2r+4q-2p}\binom{-k'+a}{p}\binom{p}{q}\binom{s}{r}i^r$$
$$\times\int_0^{2\pi} d\varphi\, e^{i\varphi(m_1+m_2-m_1'-m_2')}(\sin\varphi)^{2q+r}(\cos\varphi)^{2p-2q+s-r}$$
$$\times\int_0^{\pi/2} d\theta\,(\sin\theta)^{2(n-k+a-p-1)-1}(\cos\theta)^{m_1+m_2-\tilde{m}_1-\tilde{m}_2+2p+1}. \tag{53}$$

This integral can be calculated with the help of the formulas 3.621(2) and 3.892(4) of Ref. 19:

$$\int_0^{\pi/2}\sin^{\mu-1}\theta\cos^{\nu-1}\theta\, d\theta = \tfrac{1}{2}B(\mu/2, \nu/2), \quad \mathrm{Re}\,\mu > 0, \quad \mathrm{Re}\,\nu > 0, \tag{54}$$

$$\int_0^{\pi} e^{2i\beta\varphi}\sin^{2\mu}\varphi\cos^{2\nu}\varphi\, d\varphi = \frac{\pi\exp[i\pi(\beta-\nu)]\,{}_2F_1(-2\nu, \beta-\mu-\nu; 1+\beta+\mu-\nu; -1)}{4^{\mu+\nu}(2\mu+1)B(1-\beta+\mu+\nu, 1+\beta+\mu-\nu)}. \tag{55}$$

If the conditions (51) are not satisfied we have to use other integral formulas for the matrix elements $d^{\mu}(t)$. If the relation (45) is used then at $\mu = -M$ for the integral in (45) we have that

$$I = \sum_{p=0}^{a-k'}\sum_{q=0}^{p}\sum_{r=0}^{s} t^{4q-2p+2r}\binom{a-k'}{p}\binom{p}{q}\binom{s}{r}i^r\int_0^{2\pi} d\varphi\, e^{i\varphi(m_1+m_2-m_1'-m_2')}(\sin\varphi)^{2q+r}(\cos\varphi)^{2p-2q+s-r}$$
$$\times\int_0^{\pi/2} d\theta\,(\sin\theta)^{2(k-m_1'-m_2'+a-p+n-1)-1}(\cos\theta)^{2(m_1'+m_2'+p)-m_1-m_2-\tilde{m}_1-\tilde{m}_2+1} t^{-\mu+m_1'+m_2'-2k'}, \tag{56}$$

where the notations (52) are introduced. This formula can be used for evaluation of the matrix elements (50) by Eqs. (54) and (55), if the conditions

$$\tilde{m}_1 + \tilde{m}_2 - m_1 - m_2 \geqslant 0, \quad m_1' + m_2' - \tilde{m}_1 - \tilde{m}_2 \geqslant 0$$

are satisfied.

At the conditions

$$\tilde{m}_1 + \tilde{m}_2 - m_1 - m_2 \geqslant 0, \quad -m_1' - m_2' + \tilde{m}_1 + \tilde{m}_2 \geqslant 0$$

the formula (47) is used for evaluation of the matrix elements (50). For the integral in (47) we have

$$I = \sum_{p=0}^{a+k'}\sum_{q=0}^{p}\sum_{r=0}^{s} t^{4q-2p+2r-\mu-m_1'-m_2'+2k'}\binom{a+k'}{p}\binom{p}{q}\binom{s}{r}(-i)^r$$
$$\times\int_0^{2\pi} d\varphi\, e^{i\varphi(m_1+m_2-m_1'-m_2')}(\sin\varphi)^{2q+r}(\cos\varphi)^{2(p-q)+s-r}$$
$$\times\int_0^{\pi/2} d\theta\,(\sin\theta)^{2(k+a-p+n-1)-1}(\cos\theta)^{\tilde{m}_1+\tilde{m}_2-m_1-m_2+2p+1}, \tag{57}$$

where the notations

$$(M-\tilde{m}_1-\tilde{m}_2)/2=a,\quad \tilde{m}_1+\tilde{m}_2-m_1'-m_2'=s \tag{58}$$

are introduced. Now we have to use Eqs. (54) and (55).

At the conditions

$$m_1+m_2-\tilde{m}_1-\tilde{m}_2\geqslant 0,\quad \tilde{m}_1+\tilde{m}_2-m_1'-m_2'\geqslant 0$$

the formula (46) is used. For the integral in (46) we have

$$I=\sum_{p=0}^{a+k'}\sum_{q=0}^{p}\sum_{r=0}^{s} t^{4q-2p+2r-\mu-m_1'-m_2'+2k'}\binom{a+k'}{p}\binom{p}{q}\binom{s}{r}(-i)^r\int_0^{2\pi} d\varphi\, e^{i\varphi(m_1+m_2-m_1'-m_2')}(\sin\varphi)^{2q+r}(\cos\varphi)^{2(p-q)+s-r}$$
$$\times\int_0^{\pi/2} d\theta\,(\sin\theta)^{2(m_1'+m_2'+k+a-p+n-1)-1}(\cos\theta)^{m_1+m_2+\tilde{m}_1+\tilde{m}_2-2(m_1'+m_2')+2p+1}, \tag{59}$$

where the notations (58) are introduced. Now we have to use Eqs. (54) and (55).

The integrals (53), (56), (57), and (59) define all matrix elements (50). Let us note that in these integrals integration over φ leads to a hypergeometric series which is a finite sum [see Eq. (55)].

We have obtained the matrix elements (50) of the finite dimensional irreducible representations of Sp(n, R). Analytic continuation of t to $e^{i\varphi}$, $0<\varphi<2\pi$, leads to the matrix elements for finite dimensional representations of the compact group Sp(n) with highest weights (M, 0,...,0). They correspond to the following elements of Sp(n):

$$\mathrm{diag}(1,\ldots,1,e^{i\varphi},1,\ldots,1,e^{-i\varphi}). \tag{60}$$

The representation matrices, which are obtained are, however, not unitary. In order to make them unitary, we have to change the basis $|m_1, m_2, \alpha\rangle$ to the basis $|m_1, m_2, \alpha\rangle'$ with the help of Eqs. (25)–(28). As a result we have the matrix elements of the unitary representations of Sp(n) with highest weights (M, 0,...,0) in the U(n) basis:

$$\tilde{D}^{M}_{(m_1,m_2)(\tilde{m}_1,\tilde{m}_2)(m_1'm_2')}(e^{i\varphi})=(a_{m_1m_2}/a_{\tilde{m}_1\tilde{m}_2})^{1/2}\times D^{M}_{(m_1,m_2)(\tilde{m}_1,\tilde{m}_2)(m_1'm_2')}(e^{i\varphi}),$$

where D^m is defined by Eq. (50).

These matrix elements define matrix elements of an operator, corresponding to any matrix of Sp(n). In reality, any matrix of Sp(n) can be represented as a product of elements of the subgroup U(n) and elements (60), corresponding to the groups Sp(k), $k=1,2,\ldots,n$.

[1] B. G. Wybourne, *Symmetry Principles in Atomic Spectroscopy* (Wiley, New York, 1970).
[2] N. J. Gunther, J. Math. Phys. **20**, 1539 (1979).
[3] K. T. Hecht and D. Braunschweig, Nucl. Phys. A **295**, 34 (1978).
[4] G. F. Filippov, V. I. Ovcharenko, and O. L. Weaver, Preprint ITP-80-140E, Institute for Theoretical Physics, Kiev, 1980 (unpublished).
[5] G. Rosensteel and D. J. Rowe, Ann. Phys. (N.Y.) **126**, 343 (1980).
[6] B. Gruber, F. A. Matsen, and M. S. Thomas, "Symmetry Chains for Atomic Shell Model", *Symmetries in Science*, edited by B. Gruber and R. Millman (Plenum, New York, 1980).
[7] E. Chacon, M. Ciftan, and L. C. Biedenharn, J. Math. Phys. **13**, 577 (1972).
[8] S. J. Alisauskas, A. A.-A. Jucys, and A. P. Jucys, J. Math. Phys. **13**, 1329 (1972).
[9] B. Gruber and A. U. Klimyk, J. Math. Phys. **23**, 1399 (1982).
[10] B. Gruber and A. U. Klimyk, J. Math. Phys. **22**, 2762 (1981).
[11] A. U. Klimyk, *Matrix elements and Clebsch–Gordan Coefficients of Group Representations* (Naukova Dumka, Kiev, 1979) (in Russian).
[12] A. U. Klimyk, "New Approach to Matrix Elements and Clebsch–Gordan Coefficients for Compact and Non-Compact Groups" in *Symmetries in Science*, edited by B. Gruber and R. Millman (Plenum, New York, 1980).
[13] A. U. Klimyk and A. M. Gavrilik, J. Math. Phys. **20**, 1624 (1979).
[14] A. U. Klimyk, Lett. Math. Phys. **4**, 399 (1980).
[15] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[16] G. Warner, *Harmonic Analysis on Semisimple Lie Groups* (Springer, Berlin, 1972), Vol. 1.
[17] A. U. Klimyk and B. Gruber, J. Math. Phys. **20**, 1995 (1979).
[18] N. Ja. Vilenkin, Sb. Nauchnych Trudov Mosk. Gos. Zaoch. Ped. Inst. Moscow **39**, 77 (1974) (in Russian).
[19] I. S. Gradshtein and I. M. Ryzhik, *Tables of Integrals, Sums, Series and Products* (Academic, New York, 1965).

Reduction of inner-product representations of unitary groups

Department of Physics, Indian Institute of Technology, Bombay-400 076, India

(Received 21 May 1981; accepted for publication 20 November 1981)

A direct method for the reduction of inner products of irreducible representations (irreps) of unitary groups has been proposed using the duality between the permutation and unitary groups. A canonical tensor basis set has been used to obtain a closed expression for the Clebsch–Gordan coefficients of U(n). This expression involves the subduction coefficients arising in the outer-product reduction of $S_{N_1} \otimes S_{N_2} \rightarrow S_{N_1+N_2}$ of the permutation groups, the symmetrization coefficients of U(n), and matrix elements of the standard representation of S_N. The expression holds good for an inner-product reduction of irreps of U(n), and is independent of n. The method has been illustrated with examples.

PACS numbers: 02.20.Qs

I. INTRODUCTION

The reduction of Kronecker products of group representations is a problem of considerable importance in many physical applications. For finite dimensional irreducible representations (irreps) of unitary groups, this problem has been extensively studied by a number of workers.[1] The Clebsch–Gordan coefficients (CGC) occurring in this reduction find many applications in both particle[2] and nuclear[3,4] physics problems. The standard methods for determining the CGC consist of obtaining Isoscalar Factors (ISF) for a given group–subgroup chain and the CGC for the subgroups to generate the CGC for the group.[5,6] These results on subgroup CGC and ISF combined with Racah's factorization lemma[7] lead to a direct determination of the CGC for the group. If a canonical subgroup chain $U(n) \supset U(n-1) \supset \cdots \supset U(1)$ is used for the unitary groups, the ISF follow from matrix elements of the generators of U(n) [cf. Eq. (2.62), Ref. 1, for these matrix elements]. This was basically the method used for obtaining the ISF for SU(4)/SU(3)[8] and SU(3)/SU(2)[9] chains. Though programs are available for obtaining these matrix elements,[10] a drawback of these methods is that computational complexity increases rapidly with n so far as ISF determination is concerned.

In view of the above difficulty, an alternative would be to consider a decomposition of the N th rank tensor basis of U(n) into tensors of ranks $N-1$,1 and diagonalize the permutation operator,

$$P^N = \sum_{i<j}^{N} (i,j),$$

over a decoupled inner product basis. Such an approach was successfully used recently for the CGC in the subgroup chain SU(6)/SU(3) ⊗ SU(2).[11] Though the procedure is highly recursive, the computational effort increases more with the rank of the tensors than with the dimensionality of the fundamental representation space. In a more recent note, Chen[12] demonstrated that the ISF resulting in the reduction of inner-product representations of the permutation group adapted to the chain $S_N \supset S_{N_1} \otimes S_{N_2}$ are the same as those for the chain SU(nm) ⊃ Su(n) ⊗ Su(m). This permits a direct determination of ISF using permutation-group-based techniques independent of n and m. There are, however, some limitations in this approach which are worth noting. Firstly, the complexity of the "eigenequations" to be solved [cf. Eqs. (20) and (22), Ref. 12] is considerable except when $N_1 = N-1$ and $N_2 = 1$. In this case the procedure becomes highly recursive as in So and Strottman's approach.[11] Secondly, the inner–outer dualism[13,14] between permutation and unitary groups leads to complete identity between the CGC and subduction coefficients of S_N and the subduction and CG coefficients of U(n), respectively [cf. Eqs. (28) and (29) and the discussions preceding them in Ref. 12], only for special Gel'fand[15] states of U(n) which have all weights equal to unity. This, in turn, implies that either generator or some other algebra has to be used for obtaining the CGC for the other basis states in either a canonical or noncanonical subgroup adapted structure of U(n). This aspect of the dualism has been the subject of recent studies by some of us.[16,17] A nonrecursive procedure[18] for determining the CGC of S_N was successfully combined with symmetrization methods[16] for generating the canonical basis for U(n) to obtain the subduction coefficients occurring in the subgroup adaptation $U(nm) \supset U(n) \otimes U(m)$.[17]

In line with the recent studies of unitary groups,[16,17] we have now attempted to use the outer–inner dualism between the product representations of S_N and U(n) to obtain an explicit realization of the CGC for the canonical basis spanning the irreps of U(n) in terms of easily determined subduction coefficients[19–21] of S_N and the symmetrization coefficients of U(n).[16] We demonstrate that the compact list of subduction coefficients for S_N can easily replace the ISF in determining the CGC of U(n). The present scheme is outlined in Sec. 2 and a number of illustrative examples are considered in Sec. 3. A brief discussion is presented in Sec. 4.

2. CLEBSCH–GORDAN COEFFICIENTS FOR THE UNITARY GROUP

Let $\{u_i \mid i = 1,\ldots,n\}$ define an ordered orthonormal set of basis functions spanning the fundamental representation space V_n of the unitary group U(n). Using these basis functions we can define sets of ordered tensor monomials of ranks $N,N,\ldots$ spanning the tensor spaces $V_n \otimes^{N'}, V_n \otimes^{N''}, \ldots$, respectively. Since the ranks of these tensors are distinct,

 0022-2488/83/020233-06$02.50

each of the sets is stable under the transformations induced by the generators $\{E_{ij} | i,j = 1,2,...,n\}$ of U(n), which act as shift operators in each of the tensor spaces. The reduction of a given space $V_n \otimes^N$ yields a set of subspaces corresponding to irreps of U(n),

$$\langle m \rangle = \langle m_{1n}\ m_{2n} \cdots m_{nn} \rangle \tag{1}$$

such that

$$M_{1n} \geqslant m_{2n} \geqslant \cdots \geqslant m_{nn} \geqslant 0 \quad \sum_{i=1}^{n} m_{in} = N .$$

The above irrep is characterized by a Young diagram (YD) with m_{1n} boxes in the first row, m_{2n} in the second row, etc. Corresponding to each irrep $\langle m \rangle$ we can generate f_N^m sets of symmetry adapted tensors, each set spanned by f_N^m basis functions. The f_N^m sets correspond to different coupling schemes used in generating the basis functions and are not related by any transformations induced by the generators E_{ij} of U(n). Using the index r to distinguish between the various coupling schemes and a Weyl tableau index (p) to distinguish among the orthonormal basis states spanning a given set, we can represent the tensor basis states as[16]

$$|\langle m \rangle r(p); (N_1\ N_2 \cdots N_n)\rangle ,$$

where $(N_1\ N_2 \cdots N_n)$ represents a primitive monomial belonging to $V_n \otimes^N$, and the index (p) characterizes an allowed distribution of N_1 entries 1, N_2 entries 2, etc., in the Young diagram $[m]$.

The above tensor basis can be readily generated using nonstandard elements $e^m_{r(p)}$, of the algebra of S_N, defined as[16,17]

$$e^m_{r(p)} = \sum_{s \in (p)} a^m_{s(p)}\ e^m_{rs} , \tag{2}$$

where the summation on the right is over all standard Young tableaux s corresponding to a given Weyl tableau (p) (in the sense used by Patterson and Harter[22]) and e^m_{rs} are standard Wigner operators of S_N defined as[23]

$$e^m_{rs} = \left(\frac{f^m_N}{N!}\right)^{1/2} \sum_{P \in S_N} [P]^m_{rs}\, P \tag{3}$$

with $[P]^m_{rs}$ being the Young representation matrix element. The symmetrization coefficients $a^m_{s(p)}$ are determined using the right invariance of $e^m_{r(p)}$ under the elementary transpositions belonging to the subgroup $S_{N_1} \otimes S_{N_2} \otimes \cdots \otimes S_{N_n} \subset S_N$.[16,17] Using the operators defined in Eq. (2), the normalized tensor basis of U(n) can be obtained as

$$\begin{aligned} &|\langle m \rangle r(p); (N_1 \cdots N_n)\rangle \\ &\quad = \left(\prod_{i=1}^{n} N_i!\right)^{-1/2} e^m_{r(p)} |(N_1 \cdots N_n)\rangle \\ &\quad = \left(\prod N_i!\right)^{-1/2} (a^m_{s(p)})^{-1}\, e^m_{rs} |(N_1 \cdots N_n)\rangle , \end{aligned} \tag{4}$$

where the last step follows as in Eq. (6) of Ref. 17.

Consider now the inner-product mapping U(n)$\times$U(n) $\rightarrow$U(n). If $\langle m' \rangle$ and $\langle m'' \rangle$ are two irreps of U(n) defined over the tensors of ranks N' and N'', respectively, the reduction of their inner-product representation is given by the Clebsch–Gordan series

$$\langle m' \rangle \times \langle m'' \rangle = \sum_m b^m_{m'm''} \langle m \rangle , \tag{5}$$

where $b^m_{m'm''}$ is the multiplicity of $\langle m \rangle$ in the reduction $\langle m' \rangle \times \langle m'' \rangle$ and is determined using Littlewood's rules.[2] Equations (4) and (5) now yield

$$\begin{aligned} &|\langle m' \rangle r'(p'); (N'_1 \cdots N'_n)\rangle\, |\langle m'' \rangle r''(p''); (N''_1 \cdots N''_n)\rangle \\ &\quad = \left(\prod_{i=1}^{n} N'_i! \prod_{j=1}^{n} N''_j!\right)^{-1/2} (a^{m'}_{s'(p')}\, a^{m''}_{s''(p'')})^{-1} \\ &\qquad \times e^{m'}_{r's'} \otimes e^{m''}_{r''s''}\, |(N'_1 \cdots N'_n)(N''_1 \cdots N''_n)\rangle , \end{aligned} \tag{6}$$

where $\otimes$ represents the outer multiplication for the Wigner operators of $S_{N'} \otimes S_{N''} \subset S_{N'+N''}$. Using Eq. (3) for the Wigner operators, we have

$$\begin{aligned} e^{m'}_{r's'} \otimes e^{m''}_{r''s''} = &\left(\frac{f^{m'}_{N'} f^{m''}_{N''}}{N'!N''!}\right)^{1/2} \sum_{P' \in S_{N'}} \sum_{P'' \in S_{N''}} [P']^{m'}_{r's'} \\ &\times [P'']^{m''}_{r''s''}\, P'P'' . \end{aligned} \tag{7}$$

Since $P'P''$ is also an element of S_N, we have[23]

$$P'P'' = \sum_m \sum_{r,s=1}^{f^m_N} \left(\frac{f^m_N}{N!}\right)^{1/2} [P'P'']^m_{rs}\, e^m_{rs} . \tag{8}$$

Using the result from Eq. (8) on the right-hand side of Eq. (7) we obtain

$$\begin{aligned} e^{m'}_{r's'} \otimes e^{m''}_{r''s''} = &\left(\frac{f^{m'}_{N'} f^{m''}_{N''}}{N'!N''!}\right)^{1/2} \sum_m \left(\frac{f^m_N}{N!}\right)^{1/2} \\ &\times \sum_{P',P''} \sum_{r,s=1}^{f^m_N} [P']^{m'}_{r's'} [P'']^{m''}_{r''s''}\, [P'P'']^m_{rs}\, e^m_{rs} . \end{aligned} \tag{9}$$

In order to evaluate the matrix elements $[P'P'']^m_{rs}$, it is convenient to use the nonstandard basis of S_N adapted to the irreps of $S_{N'} \otimes S_{n''}$. The transformation between the standard Young basis and the nonstandard ones is given in terms of a unitary matrix called "subduction coefficient" matrix[19,20,23]

$$\begin{aligned} |[m]r\rangle^N = &\sum_{t'=1}^{f^{k''}_{N''}} \sum_{\tau_{k''}} S\begin{pmatrix} m & k' & k'' \\ r & t' & t'' \end{pmatrix} \Big| \tau_{k''} \Big) |[k']t'\rangle^{N'} \\ &\times |[k'']t''\rangle^{N''}, \end{aligned} \tag{10}$$

where

$$S\begin{pmatrix} m & k' & k'' \\ r & t' & t'' \end{pmatrix} \Big| \tau_{k''} \Big)$$

is a subduction coefficient. The index $\tau_{k''}$ distinguishes multiply occurring irreps $[k'']$ in the reduction of the skew representation $[m] - [k']$ over the last N'' entries.[24] Using the result of Eq. (10) on the right of Eq. (9) and using the orthogonality of the matrix representations of S_N, viz.,

$$\sum_{P \in S_N} [P]^m_{rs}\, [P]^k_{tu} = \left(\frac{N}{f^m_N}\right) \delta_{rt}\, \delta_{su}\, \delta_{mk}, \tag{11}$$

we obtain the result

$$\begin{aligned} e^{m'}_{r's'} \times e^{m''}_{r''s''} = &\left(\frac{N'!N''!}{f^{m'}_{N'} f^{m''}_{N''}}\right)^{1/2} \sum_m \left(\frac{f^m_N}{N!}\right)^{1/2} \\ &\times \sum_{\tau_{m''}} \sum_{r,s=1}^{f^m_N} S\begin{pmatrix} m & m' & m'' \\ r & r' & r'' \end{pmatrix} \Big| \tau_{m''} \Big) \\ &\times S\begin{pmatrix} m & m' & m'' \\ s & s' & s'' \end{pmatrix} \Big| \tau_{m''} \Big) e^m_{rs} . \end{aligned} \tag{12}$$

Using the above result on the right of Eq. (6), we have

$$|\langle m'\rangle r'(p');(N'_1\cdots N'_n)\rangle|\langle m''\rangle r''(p'');(N''_1\cdots N''_n)\rangle$$
$$=\left(\prod_{i=1}^{n}N'_i!\prod_{j=1}^{n}N''_j!\right)^{-1/2}(a^{m'}_{s'(p')}\,a^{m''}_{s''(p'')})^{-1}$$
$$\times\left(\frac{N'!N''!}{f^{m'}_{N'}f^{m''}_{N''}}\right)^{1/2}\sum_{m}\sum_{\tau_{m''}}\sum_{r,s=1}^{f^m_N}\left(\frac{f^m_N}{N}\right)^{1/2}$$
$$\times S\begin{pmatrix}m & m' & m''\\ r & r' & r''\end{pmatrix}\Big|\tau_{m''}\Big)\,S\begin{pmatrix}m & m' & m''\\ s & s' & s''\end{pmatrix}\Big|\tau_{m''}\Big)$$
$$\times e^m_{rs}|(N'_1\cdots N'_n)(N''_1\cdots N''_n)\rangle\,. \tag{13}$$

Before proceeding further with Eq. (13) we note that the monomial on the right is not in proper form as required by the ordering in $(N_1N_2\cdots N_n)\in V_n\otimes^N$. Let P be the permutation which reorders the monomial. Therefore, using the definition of e^m_{rs}, we get the result

$$|\langle m'\rangle r'(p);(N'_1\cdots N'_n)\rangle|\langle m''\rangle r''(p'');(N''_1\cdots N''_n)\rangle$$
$$=\left(\frac{N'!N''!}{f^{m'}_{N'}f^{m''}_{N''}\prod_{i=1}^{n}N'_i!\prod_{j=1}^{n}N''_j!}\right)^{1/2}(a^{m'}_{s'(p')}\,a^{m''}_{s''(p'')})^{-1}$$
$$\times\sum_{m}\sum_{\tau_{m''}}\sum_{r,s,t=1}^{f^m_N}\left(\frac{f^m_N}{N}\right)^{1/2}S\begin{pmatrix}m & m' & m''\\ r & r' & r''\end{pmatrix}\Big|\tau_{m''}\Big)$$
$$\times S\begin{pmatrix}m & m' & m''\\ s & s' & s''\end{pmatrix}\Big|\tau_{m''}\Big)[P]^m_{ts}\,e^m_{rt}|(N_1\cdots N_n)\rangle\,. \tag{14}$$

Using Eq. (4), we can re-express the above result in terms of the tensor basis of U(n) as

$$|\langle m'\rangle r'(p');(N'_1\cdots N'_n)\rangle|\langle m''\rangle r''(p'');\ (N''_1\cdots N''_n)\rangle$$
$$=\left[\frac{N'!N''!}{f^{m'}_{N'}f^{m''}_{N''}\prod_{i=1}^{n}N'_i!\prod_{j=1}^{n}N''_j!}\right]^{1/2}[a^{m'}_{s'(p')}\,a^{m''}_{s''(p'')}]^{-1}$$
$$\times\sum_{m}\sum_{\tau_{m''}}\sum_{r,s,=1}^{f^m_N}\left(\frac{f^m_N}{N}\right)^{1/2}\left(\prod_{k=1}^{n}N_k!\right)^{1/2}a^m_{t(p)}[P]^m_{ts}$$
$$\times S\begin{pmatrix}m & m' & m''\\ r & r' & r''\end{pmatrix}\Big|\tau_{m''}\Big)\times S\begin{pmatrix}m & m' & m''\\ s & s' & s''\end{pmatrix}\Big|\tau_{m''}\Big)$$
$$\times|\langle m\rangle r(p);(N_1\cdots N_n)\rangle\,. \tag{15}$$

However, since the transformations induced by E_{ij} on the tensor basis is independent of the coupling scheme index r, we can choose the tensor basis as

$$|\langle m\rangle\tau_{m''}r'r''(p);(N_1\cdots N_n)\rangle$$
$$=\sum_{r=1}^{f^m_N}S\begin{pmatrix}m & m' & m''\\ r & r' & r''\end{pmatrix}\Big|\tau_{m''}\Big)$$
$$\times|\langle m\rangle r(p);(N_1\cdots N_n)\rangle\,. \tag{16}$$

It is also worth noting that the summation over the index t on the right of Eq. (15) can be factored into summation over all t defining a given (p) and summation over all possible distinct Weyl tableaux (p) occurring for each irrep $\langle m\rangle$. Using this result and Eq. (6) on the right of Eq. (15) we finally obtain the result

$$|\langle m'\rangle r'(p');(N'_1\cdots N'_n)\rangle|\langle m''\rangle r''(p'');(N''_1\cdots N''_n)\rangle$$
$$=\sum_{m}\sum_{\tau_{m''}}\sum_{(p)}\begin{pmatrix}m & m' & m''\\ p & p' & p''\end{pmatrix}\Big|\tau_{m''}\Big)$$
$$\times|\langle m\rangle\tau_{m''}\,r'r''(p);(N_1\cdots N_n)\rangle\,, \tag{17}$$

where the required CGC of U(n) is given by

$$\begin{pmatrix}m & m' & m''\\ p & p' & p''\end{pmatrix}\Big|\tau_{m'}\Big)=\left(\frac{\prod_{k=1}^{n}N_k!N'!N''!f^m_N}{\prod_{i=1}^{n}N'_i!\prod_{j=1}^{n}N''_j!N!f^{m'}_{N'}f^{m''}_{N''}}\right)^{1/2}$$
$$\times\sum_{s=1}^{f^m_N}\sum_{t\in(p)}\left(\frac{a^m_{t(p)}}{a^{m'}_{s'(p')}a^{m''}_{s''(p'')}}\right)[P]^m_{ts}\,S\begin{pmatrix}m & m' & m''\\ s & s' & s''\end{pmatrix}\Big|\tau_{m''}\Big)\,. \tag{18}$$

The right-hand side of Eq. (18) is determinable in a straightforward manner using simple algorithms for symmetrization coefficients[16] and subduction coefficients for S^{20}_N developed recently. In all cases the matching permutation is a product of simple cycles, and as has recently been shown,[25] each such cyclic permutation can be handled relatively easily. Thus the determination of CGC using Eq. (18) is not as formidable as might appear at first sight. In the next section we illustrate the use of Eq. (18) in determining the reduction of the product basis of the irreps of U(n). Some of the advantages of the present method will be illustrated using special cases.

3. SOME ILLUSTRATIVE EXAMPLES

We first consider a simple case of Eq. (18) for which the first N' particles occupy the first n' single particle orbitals and the last $N''=N-N'$ occupy the last $n''=n-n'$ orbitals. Here we note that the monomial $(N'_1\cdots N'_{n'})(N''_{n'_1+1}\cdots N''_{n''})$ is already in properly ordered form, so that the reordering permutation reduces to the identity. Further, the reducible tensor space of primitive tensor monomials is now a subspace $(V'_n\otimes^{N'})\otimes(V_{n-n'}\otimes^{N''})$ of $(V_n\otimes^N)$ and product representations arise from the restriction U(n)sis to that of the group have recently been suggested using the semimaximal weight concept and the lowering operators of the subgroup U(n') ⊗ U(n'').[26,27] We now examine what Eq. (18) reduces to in this case. Letting $P=e$ in that equation, we readily obtain the result

$$\begin{pmatrix}m & m' & m''\\ p & p' & p''\end{pmatrix}\Big|\tau_{m''}\Big)=\left(\frac{N'!N''!f^m_N}{N!f^{m'}_{N'}f^{m''}_{N''}}\right)^{1/2}$$
$$\times(a^{m'}_{s'(p')}\,a^{m''}_{s''(p'')})^{-1}\sum_{s\in(p)}a^m_{s(p)}$$
$$\times S\begin{pmatrix}m & m' & m''\\ s & s' & s''\end{pmatrix}\Big|\tau_{m''}\Big)\,. \tag{19}$$

As an illustration of the use of the above equation consider the restriction $\langle 5,2,0^4\rangle\downarrow\langle 2,1,0\rangle\otimes\langle 3,1,0\rangle$ of U(6) to U(3) ⊗ U(3). Let V'_3 be spanned by the three orthornormal basis states u_1,u_2,u_3 and V''_3 be spanned by u_4,u_5,u_6. The Weyl tableau for maximal weight states of the two groups U(3) are $\begin{smallmatrix}1&1\\2\end{smallmatrix}$ and $\begin{smallmatrix}4&4&4\\5\end{smallmatrix}$, respectively. The semimaximal weight states of U(6) for the given irrep are $\begin{smallmatrix}1&1&4&4&4\\2&5&5&5\end{smallmatrix}$ and $\begin{smallmatrix}1&1&4&4&5\\2&5&5&5\end{smallmatrix}$. Using the Yamanouchi notation for the standard Young tableaux of the irrep [5,2] of S_7, the set of the Young tableaux corresponding to the above two Weyl tableaux (in the sense defined in Sec. 2) are listed below:

$$\begin{smallmatrix}1&1&4&4&4\\2&5&5&5\end{smallmatrix}\equiv(1121112)\,,$$

$$\begin{matrix}11445\\24\end{matrix} \equiv (1121121),(1121211),(1122111)\,.$$

The basis transformation for the restriction $[5,2]\downarrow[2,1]\otimes[3,1]$ of S_7 to $S_3\otimes S_4$ can be readily worked out using the techniques developed in an earlier paper[20] and is

$$\begin{aligned}(112)\times(1112) = \tfrac{1}{2}[&-3\sqrt{15}(1121112) + (1121121)\\ &+\sqrt{2}(1121211)\\ &+\sqrt{6}(1122111)]\,. \end{aligned} \tag{20}$$

Consider first the states $\begin{matrix}11\\2\end{matrix}\times\begin{matrix}444\\5\end{matrix}$ defining the product state of $U(3)\times U(3)$. We observe that each of the Weyl states corresponds uniquely to one standard Young tableau so that the required symmetrization coefficients are unity,

$$a^{\langle 2,1,0\rangle}_{(112)(\begin{smallmatrix}11\\2\end{smallmatrix})} = 1, \quad a^{\langle 3,1,0\rangle}_{(1112)(\begin{smallmatrix}444\\5\end{smallmatrix})} = 1\,. \tag{21}$$

Similar is the case with $\begin{matrix}11444\\25\end{matrix}$ so that

$$a^{\langle 5,2,0^4\rangle}_{(1121112)(\begin{smallmatrix}11444\\25\end{smallmatrix})} = 1\,. \tag{22}$$

To the Weyl tableau $\begin{matrix}11445\\24\end{matrix}$ correspond three standard Young tableaux, as indicated earlier, over which symmetrization is required. This symmetrization requires essentially that we determine the subduction coefficients for the restriction $[4,2]\downarrow[2,1]\otimes[3]$. This determination leads to the symmetrization coefficients

$$a^{\langle 5,2,0^4\rangle}_{(1121121)(\begin{smallmatrix}11445\\24\end{smallmatrix})} = \tfrac{1}{3}\,,$$

$$a^{\langle 5,2,0^4\rangle}_{(1121211)(\begin{smallmatrix}11445\\24\end{smallmatrix})} = \tfrac{1}{3}\sqrt{2}\,, \tag{23}$$

$$a^{\langle 5,2,0^4\rangle}_{(1122111)(\begin{smallmatrix}11445\\24\end{smallmatrix})} = \tfrac{1}{3}\sqrt{6}\,.$$

Using the results of Eqs. (20) and (21) in Eq. (19) and noting that $f^{[2,1]}_3 = f^{[3,1]}_4 = 3$ and $f^{[5,2]}_7 = 14$, we readily obtain

$$\begin{pmatrix}\langle 5,2,0^4\rangle & \langle 2,1,0\rangle & \langle 3,1,0\rangle\\ \begin{smallmatrix}11444\\25\end{smallmatrix} & \begin{smallmatrix}11\\2\end{smallmatrix} & \begin{smallmatrix}444\\5\end{smallmatrix}\end{pmatrix} = -\frac{1}{4}\,.$$

Similarly using Eqs. (22) and (20) in Eq. (19), we get

$$\begin{aligned}&\begin{pmatrix}\langle 5,2,0^4\rangle & \langle 2,1,0\rangle & \langle 3,1,0\rangle\\ \begin{smallmatrix}11445\\24\end{smallmatrix} & \begin{smallmatrix}11\\2\end{smallmatrix} & \begin{smallmatrix}444\\5\end{smallmatrix}\end{pmatrix}\\ &= \frac{1}{\sqrt{15}}\left[\frac{1}{3}\cdot\frac{1}{12} + \frac{\sqrt{2}}{3}\cdot\frac{\sqrt{2}}{12}\right.\\ &\quad\left. + \frac{\sqrt{6}}{3}\cdot\frac{\sqrt{6}}{12}\right] = \frac{1}{4\sqrt{15}}\,.\end{aligned}$$

Thus the reduction of the product representation yields the result

$$\left|\begin{matrix}11\\2\end{matrix}\right\rangle\times\left|\begin{matrix}444\\5\end{matrix}\right\rangle = \left(-\frac{1}{4}\right)\left|\begin{matrix}11444\\25\end{matrix}\right\rangle + \left(\frac{1}{4\sqrt{15}}\right)\left|\begin{matrix}11445\\24\end{matrix}\right\rangle\,. \tag{24}$$

A similar result,

$$\left|\begin{matrix}2&&1&&0\\&2&&1\\&&2\end{matrix}\right\rangle\times\left|\begin{matrix}3&&1&&0\\&3&&1\\&&3\end{matrix}\right\rangle$$

$$\begin{aligned}&= -\frac{\sqrt{15}}{4}\left|\begin{matrix}5&2&0&0&0&0\\&5&2&0&0&0\\&&5&1&0&0\\&&&2&1&0\\&&&&2&1\\&&&&&2\end{matrix}\right\rangle\\ &\quad + \left(\frac{1}{4}\right)\left|\begin{matrix}5&2&0&0&0&0\\&5&2&0&0&0\\&&4&2&0&0\\&&&2&1&0\\&&&&2&1\\&&&&&2\end{matrix}\right\rangle\,,\end{aligned} \tag{25}$$

follows from the restriction $\langle 5,2,0^4\rangle\downarrow\langle 2,1,0\rangle\otimes\langle 3,1,0\rangle$ of $U(6)$ to $U(3)\otimes U(3)$ on applying the weight raising generator E_{45} to an arbitrary linear combination of the Gel'fand basis on the right of the above equation and equating the result to zero. This relates one coefficient to the other and the final unknown can be determined by normalizing the result. The results of Eqs. (24) and (25) differ only by an overall multiplicative factor 15. This can be accounted for using the fact that CGC of Eq. (24) have to be normalized over the entire set of basis functions defining the outer-product series, instead of two functions as in Eq. (25).

At first sight it might appear that Eq. (25) follows more easily than the corresponding result of Eq. (24). While it is true of this particular semimaximal state it is not in general so. To illustrate this fact consider the product state $|\begin{smallmatrix}12\\3\end{smallmatrix}\rangle \times |\begin{smallmatrix}456\\7\end{smallmatrix}\rangle$ of $U(3)\otimes U(4)\subset U(7)$ under the same restriction as above. Here we find that we would have to apply a number of lowering generators E_{ij} $(i>j; i=1,2,3$ and $4,5,6,7)$ to Eq. (25) and their matrix elements in order to obtain the states of

TABLE I. Subduction coefficients for the reduction $[2,1]\times[2,1] = [4,2] + [4,1^2] + [3^2] + 2[3,2,1] + [3,1^3] + [3^3] + [2^2,1^2]$ (the Young bases are represented by lattice permutation symbols).

		$[21]\times[21]$ (112)(112)	(112)(121)
[42]	(112112)	$-\frac{1}{3}\sqrt{8}$	0
	(112121)	$\frac{1}{6}$	$-\sqrt{3}/2$
	(112211)	$1/\sqrt{12}$	$+\frac{1}{2}$
$[41^2]$	(112113)	$-2/\sqrt{5}$	0
	(112131)	$\sqrt{3}/\sqrt{40}$	$-\sqrt{5}/\sqrt{8}$
	(112311)	$-1/\sqrt{8}$	$-\sqrt{3}/\sqrt{8}$
$[3^2]$	(112122)	$-\frac{1}{2}$	$-\sqrt{3}/2$
	(112212)	$-\frac{1}{2}\sqrt{3}$	$\frac{1}{2}$
$[321]_1$	(112123)	$-\sqrt{5}/\sqrt{24}$	0
	(112132)	0	$-\sqrt{5}/\sqrt{24}$
	(112213)	$-\sqrt{15}/\sqrt{24}$	0
	(112231)	$1/\sqrt{24}$	$-2/\sqrt{8}$
	(112312)	0	$1/\sqrt{8}$
	(112321)	$\sqrt{3}/\sqrt{24}$	$2/\sqrt{24}$
$[321]_2$	(112123)	$1/\sqrt{96}$	$-3/\sqrt{32}$
	(112132)	$-3/\sqrt{32}$	$-5/\sqrt{96}$
	(112213)	$1/\sqrt{32}$	$\sqrt{3}/\sqrt{32}$
	(112231)	$\sqrt{5}/\sqrt{96}$	$\sqrt{5}/\sqrt{32}$
	(112312)	$-\sqrt{15}/\sqrt{32}$	$\sqrt{5}/\sqrt{32}$
	(112321)	$\sqrt{5}/\sqrt{32}$	$-\sqrt{5}/\sqrt{96}$
$[31^3]$	(112134)	$\sqrt{3}/\sqrt{8}$	$-1/\sqrt{8}$
	(112314)	$\sqrt{5}/\sqrt{8}$	$\sqrt{3}/\sqrt{40}$
	(112341)	0	$2/\sqrt{5}$
$[2^3]$	(112233)	$\frac{1}{2}$	$\frac{1}{2}\sqrt{3}$
	(112323)	$\frac{1}{2}\sqrt{3}$	$-\frac{1}{2}$
$[2^21^2]$	(112234)	$\frac{1}{2}$	$-1/\sqrt{12}$
	(112324)	$\frac{1}{2}\sqrt{3}$	$\frac{1}{6}$
	(112342)	0	$\frac{1}{3}\sqrt{8}$

$\langle 5,2,0^5\rangle$ induced from the product state under consideration. On the other hand, in the present scheme, once the subduction coefficients have been determined, the only other quantities required are the symmetrization coefficients in order to obtain the linear combination. In the present example, all orbitals are singly occupied so that no symmetrization is required and the result follows readily as

$$\left|\begin{matrix}12\\3\end{matrix}\right\rangle \times \left|\begin{matrix}456\\7\end{matrix}\right\rangle = \left(-\frac{1}{4}\right)\left|\begin{matrix}12456\\37\end{matrix}\right\rangle + \frac{1}{12\sqrt{15}}\left|\begin{matrix}12457\\36\end{matrix}\right\rangle + \left(\frac{\sqrt{2}}{12\sqrt{15}}\right)\left|\begin{matrix}12467\\35\end{matrix}\right\rangle + \left(\frac{\sqrt{6}}{12\sqrt{15}}\right)\left|\begin{matrix}12567\\34\end{matrix}\right\rangle, \tag{26}$$

where it is to be noted that the above are Weyl and not Young tableaux.

The above examples are only special cases and we now illustrate the application of the present method to the irreps $\langle 2,1,0^{n-2}\rangle \times \langle 2,1,0^{n-2}\rangle$ of U(n). For this purpose consider the reduction of outer products of the irreps $[2,1]\otimes[2,1]$ of $S_3 \otimes S_3$ yielding the irreps of S_6, viz.,

$$[2,1]\times[2,1] = [4,2] + [4,1^2] + [3^2] + 2[3,2,1] + [3,1^3] + [2^3] + [2^2,1^2].$$

The subduction coefficients for the above restriction, resulting from procedures outlined in an earlier note,[20] are listed in Table I. A certain amount of arbitrariness is unavoidable in the choice of these coefficients for the doubly occurring irrep [3,2,1]. Using these subduction coefficients we now determine some of the CGC for the reduction of the product representations $\langle 2,1.0^{n-2}\rangle \times \langle 2,1,0^{n-2}\rangle$ of U(n) for $n \geqslant 3$. Consider, for example, the product $|{}^{11}_{2}\rangle \times |{}^{12}_{3}\rangle$ which gives six Weyl states:

1112	1113	1112	111	111	111
23	22	2	223	22	22
		3		$3_{(1)}$	$3_{(2)}$

The last two states belong to the irrep $\langle 3,2,1,0^{n-3}\rangle$ occurring twice in the reduction, distinguished by suffixes (1) and (2). The monomial in the product state is $u_1\,u_1\,u_2\,u_1\,u_2\,u_3$, needing a permutation $P = (3,4)$ to bring it into standard six rank tensor form $u_1\,u_1\,u_1\,u_2\,u_2\,u_3$. We now illustrate in detail the method for determining the CGC for $|{}^{11}_{2}\rangle \times |{}^{12}_{3}\rangle \rightarrow |{}^{1112}_{23}\rangle$. From Eq. (18) we obtain

$$\begin{pmatrix} \langle 4,2,0^{n-2}\rangle & \langle 2,1,0^{n-2}\rangle & \langle 2,1,0^{n-2}\rangle \\ {\scriptstyle 1112\atop 23} & {\scriptstyle 11\atop 2} & {\scriptstyle 12\atop 3} \end{pmatrix} = \left(\frac{3!2!1!3!3!9!}{2!1!1!1!1!6!2!2!}\right)^{1/2} \times \sum_{s=1}^{9} \sum_{t\in\left(\begin{smallmatrix}1112\\23\end{smallmatrix}\right)} a^{\langle 4,2,0^{n-2}\rangle}_{s\left(\begin{smallmatrix}1112\\23\end{smallmatrix}\right)} a^{\langle 2,1,0^{n-2}\rangle}_{3^{12}\left(\begin{smallmatrix}11\\2\end{smallmatrix}\right)} a^{\langle 2,1,0^{n-2}\rangle}_{3^{12}\left(\begin{smallmatrix}12\\3\end{smallmatrix}\right)} \times [(3,4)]^{[4,2]}_{st}\, S\left(\begin{smallmatrix}[4,2]\\st\end{smallmatrix}\ \begin{smallmatrix}[3,1]\\(112)\end{smallmatrix}\ \begin{smallmatrix}[2,1]\\(112)\end{smallmatrix}\right). \tag{27}$$

The possible standard Young tableaux t leading to the Weyl tableaux ${}^{1112}_{23}$ are (111122) and (111212). Under $P = (3,4)$ these can only be linked to $s = (112112)$ having a nonzero subduction coefficient (cf. Table I). The required subduction coefficient is $-\frac{1}{3}\sqrt{8}$. Thus the right-hand side of Eq. (27) involves only $s = 112112$ and $t = 1111212$. The symmetrization coefficient $a^{\langle 4,2,0^{n-2}\rangle}_{(111212)\left(\begin{smallmatrix}1112\\23\end{smallmatrix}\right)} = (\frac{5}{8})^{1/2}$. Using these results in Eq. (27) we get

$$\begin{pmatrix} \langle 4,2,0^{n-2}\rangle & \langle 2,1,0^{n-2}\rangle & \langle 2,1,0^{n-2}\rangle \\ {\scriptstyle 1112\atop 23} & {\scriptstyle 11\atop 2} & {\scriptstyle 12\atop 3} \end{pmatrix} = \frac{3\sqrt{3}}{2\sqrt{10}} \cdot \frac{\sqrt{5}/\sqrt{8}}{1\times 1} \cdot \frac{\sqrt{8}}{3}\left(-\frac{\sqrt{8}}{3}\right) = -\frac{1}{3}.$$

Similar calculations of other CGC leads to the reduction

$$\left|\begin{matrix}11\\2\end{matrix}\right\rangle \times \left|\begin{matrix}12\\3\end{matrix}\right\rangle = -\frac{1}{\sqrt{3}}\left|\begin{matrix}1112\\23\end{matrix}\right\rangle + \frac{1}{2\sqrt{15}}\left|\begin{matrix}1113\\22\end{matrix}\right\rangle - \frac{1}{\sqrt{3}}\left|\begin{matrix}1112\\2\\3\end{matrix}\right\rangle - \frac{1}{2\sqrt{3}}\left|\begin{matrix}111\\223\end{matrix}\right\rangle - \frac{\sqrt{2}}{3}\left|\begin{matrix}111\\22\\3\end{matrix}\right\rangle_{(1)} + \frac{1}{3\sqrt{10}}\left|\begin{matrix}111\\22\\3\end{matrix}\right\rangle_{(2)}.$$

The procedure for calculating the CGC is thus evident. We first determine the possible $\{t\,|t\in(p)\}$ and determine the set $\{s\}$ to which these can be linked through reordering permutation P. From the table of subduction coefficients we then find the subset of this set which yields nonzero coefficients. Knowing the required symmetrization coefficients and the matrix elements $[P]^m_{st}$ of S_N, we can determine the CGC using Eq. (18). As a final example, we consider the product state ${}^{12}_{2}\times{}^{11}_{3}$ of U(n). Here the reordering permutation is $P = (3,4)(4,5)(2,3)(3,4)$. Using the same procedure as before we get

$$\left|\begin{matrix}12\\2\end{matrix}\right\rangle \times \left|\begin{matrix}11\\3\end{matrix}\right\rangle = -\frac{1}{\sqrt{6}}\left|\begin{matrix}1112\\23\end{matrix}\right\rangle - \frac{1}{\sqrt{6}}\left|\begin{matrix}1112\\2\\3\end{matrix}\right\rangle - \frac{1}{\sqrt{30}}\left|\begin{matrix}1113\\22\end{matrix}\right\rangle - \frac{1}{\sqrt{6}}\left|\begin{matrix}111\\223\end{matrix}\right\rangle + \frac{2}{3}\left|\begin{matrix}111\\22\\3\end{matrix}\right\rangle_{(1)} - \frac{1}{3\sqrt{5}}\left|\begin{matrix}111\\22\\3\end{matrix}\right\rangle_{(2)}.$$

The correctness of these results can be checked using tables of CGC for SU(3).[2]

4. DISCUSSION

The procedure outlined in Sec. 2 and illustrated extensively using examples in Sec. 3 is a relatively straightforward method for determining the CGC for canonical basis states spanning the irreps of U(n). To the best of our knowledge Eq. (18) is the first quantitative statement of the dualism between the product representations in $S_{N_1+N_2} \rightarrow S_{N_1} \otimes S_{N_2}$ and $\mathrm{U}(n)\otimes\mathrm{U}(n) \rightarrow \mathrm{U}(n)$. A recent statement of this dualism by Chen is at best formal [cf. Eq. (29), Ref. 12] since, as shown by Eq. (18), the CGC involves a number of factors in addition to the subduction coefficients of S_N. The correctness of Eq. (18) has been verified by determining a large number of CGC listed by Lichtenberg.[2]

In this context it is worth noting that Eq. (28) of Ref. 12 is also formal and the exact relationship between the CGC of S_N and the subduction coefficients of $\mathrm{U}(nm)\downarrow\mathrm{U}(n)\otimes\mathrm{U}(m)$

also involve a number of additional factors [cf. Eq. (26), Ref. 17].

For canonical basis states of the irreps of U(n), it might appear that the ISF determination for the U(n)/U($n-1$) chain would be simpler than using Eq. (18). Even if this is true for low n, the procedure for going from the ISF to the CGC of U(n) would require a complete knowledge of all the CGC for U($n-1$). This is likely to become a tedious recursive procedure for sufficiently large n. Alternatively Eq. (18) involves factors which are easily determinable except for the matrix representation of a single matching permutation for each CGC. Though this is a stumbling block the fact that useful techniques exist for determining the representations of both transpositions[28] and cylic permutations[25] makes this task feasible. Finally, once the subduction coefficients have been determined and listed, the CGC follow readily for any n.

ACKNOWLEDGMENTS

The authors are extremely grateful to Shri G. G. Sahasrabudhe for all the useful discussions they had with him. One of us (K. D. V.) is grateful to Council of Scientific and Industrial Research (India) for providing financial support.

[1]J. D. Louck, Am. J. Phys. **80**, 3 (1970).
[2]D. B. Lichtenberg, *Unitary Symmetry in Elementary Particles* (Academic, New York, 1970).
[3]K. T. Hecht and S. C. Pang, J. Math. Phys. **10**, 1571 (1969).
[4]M. Moshinsky, *Group Theory and the Many-Body Problem* (Gordon and Breach, New York, 1968).
[5]J. C. Carter and J. J. Coyne, J. Math. Phys. **10**, 1204 (1969).
[6]M. Machacek and Y. Tomazowa, J. Math. Phys. **17**, 458 (1975).
[7]G. Racah, *Group Theory and Spectroscopy*, Princeton Lecture notes (1951).
[8]R. Anderson and G. C. Joshi, J. Math. Phys. **20**, 1015 (1979).
[9]J. J. de Swart, Rev. Mod. Phys. **35**, 916 (1963).
[10]C. R. Sarma and S. Rettrup, J. Phys. A **13**, 2267 (1980).
[11]S. I. So and D. Strottman, J. Math. Phys. **20**, 153 (1979).
[12]Jin Quan-Chen, J. Math. Phys. **22**, 1 (1981).
[13]A. Bohr and B. R. Mottelson, *Nuclear Structure* (Benjamin, New York, (1969), Vol. I.
[14]H. Weyl, *Theory of Groups and Quantum Mechanics* (Dover, New York, 1956).
[15]I. M. Gel'fand and M. Zetlin, Dok. Akad. Nauk SSSR **71**, 825 (1950).
[16]C. R. Sarma and G. G. Sahasradbudhe, J. Math. Phys. **21**, 638 (1980).
[17]G. G. Sahasradbudhe, K. V. Dinesha, and C. R. Sarma, J. Phys. A **14**, 2189 (1981).
[18]G. G. Sahasradbudhe, K. V. Dinesha, and C. R. Sarma, J. Phys. A **14**, 85 (1981).
[19]C. R. Sarma and K. V. Dinesha, Int. J. Quantum Chem. **16**, 1195 (1979).
[20]C. R. Sarma, J. Phys. A **14**, 565 (1981).
[21]W. G. Harter and C. W. Patterson, Phys. Rev. A **13**, 1667 (1976).
[22]C. W. Patterson and W. G. Harter, J. Math. Phys. **17**, 1125 (1976).
[23]I. G. Kaplan, *Symmetry of Many-Electron Systems* (Academic, New York, 1975).
[24]G. De B. Robinson, *Representation Theory of the Symmetric Group* (Toronto U. P., Toronto, 1961), p. 48.
[25]G. G. Sahasradbudhe, K. V. Dinesha, and C. R. Sarma, Theor. Chim. Acta **54**, 333 (1980).
[26]J. Michelson, J. Math. Phys. **12**, 2378 (1971).
[27]C. R. Sarma, "Determination of Basis for Irreducible Representations of the Unitary Group for $U(p+q) \downarrow U(p) \times U(q)$," J. Math. Phys. **23**, 1235 (1982).
[28]S. Rettrup, Chem. Phys. Lett. **47**, 59 (1977).

An integral transform related to quantization. II. Some mathematical properties

I. Daubechies[a),b)]
Physics Department, Princeton University, Princeton, New Jersey 08544

A. Grossmann
CPT2-CNRS, Centre de Luminy, Route Leon Lachamp 70, F 13288 Marseille Cedex 2, France

J. Reignier
Theoretische Natuurkunde, Vrije Universiteit Brussel, Pleinlaan 2, B 1050 Brussels, Belgium

(Received 28 August 1981; accepted for publication 8 December 1981)

We study in more detail the mathematical properties of the integral transform relating the matrix elements between coherent states of a quantum operator to the corresponding classical function. Explicit families of Hilbert spaces are constructed between which the integral transform is an isomorphism.

PACS numbers: 02.30.Qy

1. INTRODUCTION

In a preceding paper,[1] two of us studied an integral transform giving a direct correspondence between a classical function on the one hand and matrix elements of the corresponding quantum operator between coherent states on the other hand:

$$(\Omega^a, Qf\Omega^b) = \int_E dv\, f(v)\{a,b\,|v\}. \tag{1.1}$$

Here E is the phase space (i.e., a $2n$-dimensional real vector space, where n is the number of degrees of freedom), and the Ω^a are the usual coherent states, labeled by phase space points (they can be considered as states centered round the phase space point a labelling them, and they minimize the uncertainty inequalities[2]).

Formula (1.1) was obtained in Ref. 1 from the correspondence formula

$$Qf = 2^r \int_E dv\, f(v) W(2v)\Pi \tag{1.2}$$

where the $W(v)$ are the Weyl operators (see Ref. 1) and Π is the parity operator. [This formula is not the original Weyl formula[3]; it gives a more direct correspondence $f \leftrightarrow Qf$ than the usual expression, since no Fourier analysis step is needed. It was shown in Ref. 4 that (1.2) is equivalent to the Weyl quantization formula.]

The integral kernel $\{a,b\,|v\}$ in (1.1) is then defined as

$$\{a,b\,|v\} = 2^n(\Omega^a, W(2v)\Pi\,\Omega^b). \tag{1.3}$$

This function was computed explicitly in Ref. 1, where we also gave some properties of both the function and the integral transform defined by it, together with some examples. A deeper mathematical study of the integral transform was, however, not intended in Ref. 1; we propose to fill this gap at least partially with the present article.

Ultimately our aim is to use the results of the mathematical study of the integral transform (1.1) to derive properties of the Weyl quantization procedure. One can indeed use the well-known "resolution of the identity" property of the coherent states,[2]

$$\int da\,|\Omega^a)(\Omega^a| = \mathbf{1} \tag{1.4}$$

to see that, at least formally, any operator A is characterized by its coherent state matrix elements

$$A = \int_E da \int_E db\,|\Omega^a)(\Omega^b|(\Omega^a, A\Omega^b). \tag{1.5}$$

A detailed knowledge of the properties of the integral transform with kernel $\{a,b\,|v\}$ might therefore be useful for

(1) giving a sense to the Weyl quantization formula for rather large classes of functions (essentially, once a precise sense is given to the integral transform on a certain class of functions, one can try to define the corresponding operators from their matrix elements between coherent states),

(2) deriving properties of the quantum operator Qf directly from properties of the corresponding function f (and vice versa).

As we shall show, the inverse of the integral transform (1.1) is given again by using the same kernel

$$f(v) = \int_E \int_E da\, db\, Qf(a,b)\{b,a|v\}. \tag{1.6}$$

(Actually, this integral does not converge absolutely in most cases, and some limiting procedure has to be introduced.) Therefore we shall also be able to use the results of our study of the integral transforms associated with the kernel function $\{a,b\,|v\}$ to obtain information on the "dequantization procedure" [i.e. the inverse map of the "quantization procedure" as defined by (1.2)]. Note that this dequantization procedure is actually the same map as the one associating to each density matrix the corresponding Wigner function[5] extended, however, to a much larger class of operators. These applications shall be further developed in a following paper (a first application was given in Ref. 6); in the present article we restrict ourselves to a study of the integral transforms

[a)] Research fellow at the Interuniversitair Instituut voor Kernwetenschappen, Belgium.
[b)] On leave of absence from Dienst voor Theoretische Natuurkunde, Vrije Universiteit Brussel, Belgium.

 0022-2488/83/020239-16$02.50

$$(If)(a,b) = \int_E dv\, f(v)\{a,b\,|v\}, \tag{1.7}$$

$$(\tilde{I}\varphi)(v) = \int_E \int_E da\, db\, \varphi(a,b)\{b,a|v\}. \tag{1.8}$$

We shall see that though f may be locally quite singular (one can even consider some classes of nontempered distributions), its image If will always be very gentle locally, with analyticity properties in both its arguments. It therefore makes sense to study not only the function $If(a,b)$, but also the coefficients $If_{m,n}$ of the Taylor series for $If(a,b)$. It turns out that one can construct a family of functions giving directly the link between $f(v)$ and $If_{m,n}$,

$$If_{m,n} = \int_E dv\, f(v) h_{m,n}(v). \tag{1.9}$$

Actually these $h_{m,n}$ are just the functions occurring in the bilinear expansion of $\{a,b\,|v\}$ in Ref. 1: formally (1.9) can be seen as the result of commuting in (1.7) the integral and the series expansion for $\{a,b\,|v\}$. However, (1.9) holds true for many functions for which such an interchanging of summation and integral would be *a priori* pure heresy. The functions $h_{m,n}$ have lots of beautiful properties, most of which are a consequence of the fact that they form a complete orthonormal set of eigenfunctions for the "harmonic oscillator" $x^2 + p^2 - \frac{1}{4}\Delta_x - \frac{1}{4}\Delta_p$ on $L^2(E)$, i.e., on phase space, where we consider an explicit decomposition of the phase space into x space p space: $E \cong \mathbb{R}^{2n} \cong \mathbb{R}^n + \mathbb{R}^n = x$ space $+ p$ space (see also Sec. 2); in the context of Weyl quantization the $h_{m,n}$ can be seen as the classical functions corresponding to the dyadics $|n\rangle\langle m|$, where $|n\rangle$ are the harmonic oscillator eigenstates (see Refs. 1, 7, and 8). Note that the $h_{m,n}$ are not the usual set of Hermite functions (though they can of course be written as linear combinations of Hermite functions); they are related to the Laguerre polynomials.[7,8]

One can then derive all kinds of results relating the growth of $If(a,b)$ or $If_{m,n}$ to the behavior of f, and analogously for $\tilde{I}\phi$ and ϕ. The derivation of such results amounts to the construction of suitable Banach or Hilbert spaces between which the integral transforms $I,\tilde{I}$ become continuous linear maps or even isomorphisms.

Our main tool for the study of $I,\tilde{I}$ will be the link between the integral transform I and the Bargmann integral transform as defined in Ref. 9 (see Sec. 6 in Ref. 1). Using this link we shall be able to translate bounds obtained in Ref. 9 to our present context, and to obtain other bounds (for other families of spaces) by similar techniques. As in Ref. 9, we can give a complete characterization of the images of $\mathcal{S}(E)$, $\mathcal{S}'(E)$ under I; by a suitable generalization we shall even go beyond the tempered distributions. (Related results, but in a completely different context, and concerning quantization restricted to functions with certain holomorphicity properties, can be found in Ref. 10.)

The paper is organized as follows. In Sec. 2 we give a survey of our notations and some properties of the kernel $\{a,b\,|v\}$, in Sec. 3 we reintroduce the h_{mn} and state some related results, in Sec. 4 we show how bounds on If can be obtained starting from bounds on f, and vice versa: in Sec. 4A we review the Banach space approach found in Bargmann[9b]; in Sec. 4B we go over to Hilbert spaces, which are better suited to our purpose, and in Sec. 4C we generalize the construction of Sec. 4B, which enables us to treat certain Hilbert spaces of distributions "of type S" which are larger than $\mathcal{S}'(E)$. In Sec. 5 we shortly discuss the integral transform I when restricted to functions on phase space which can be split up into a product of a function depending only on x with a function depending only on p. Essentially the same types of statements can be formulated, and a short survey of results is given. In Sec. 6 we give some concluding remarks.

2. NOTATIONS AND BASIC PROPERTIES OF $\{a,b/v\}$

In Ref. 1 we worked with an intrinsic coordinate-free notation system using a symplectic structure on the phase space (basically this is the bilinear structure underlying the Poisson brackets), and a complex structure yielding a Euclidean form on the phase space, compatible with the symplectic structure. By choosing a suitable basis, this could be seen to lead to a decomposition of the phase space into a direct sum of two canonically conjugate subspaces. This decomposition is not unique: for a given symplectic structure, several compatible complex structures can be constructed; different complex structures correspond then to different decompositions of phase space. This freedom in the choice of the splitting up of the phase space is particularly useful whenever (linear) canonical transformations are discussed[11] or used (as, e.g., in the presence of a constant magnetic field). Here we shall not need to use simultaneously different decomposition possibilities for the phase space, and we shall therefore fix the decomposition once and for all. We shall use this decomposition from the very start to introduce our notations in a way that is less intrinsic but probably more familiar to most readers. It goes without saying that the results we shall obtain are independent of this approach, and that they could as well be obtained in the more intrinsic setting of Ref. 1 (see Ref. 8).

The phase space E is a $2n$-dimensional real vector space, which we shall consider as a direct sum of two n-dimensional subspaces

$$E = x \text{ space} \oplus p \text{ space}, \qquad E \ni v = (x,p). \tag{2.1}$$

The x and p need not be the conventional position and momentum variables: any set of canonically conjugate coordinates which are linear combinations of position and momentum are equally good candidates for these x and p. On E we have a symplectic structure

$$\sigma(v,v') = \sigma((x,p),(x',p')) = \tfrac{1}{2}(p\cdot x' - x\cdot p') \tag{2.2}$$

and a Euclidean structure

$$s(v,v') = s((x,p),(x',p')) = \tfrac{1}{2}(x\cdot x' + p\cdot p') \tag{2.3}$$

[this is the Euclidean structure corresponding with the σ-compatible complex structure $J((x,p)) = (p, -x)$—see Ref. 1]. For further convenience we introduce a Gaussian in the phase space variables,

$$\omega(v) = \exp[-\tfrac{1}{2}s(v,v)] = \exp[-\tfrac{1}{4}(x^2 + p^2)], \tag{2.4}$$

and a family of analytic functions

$$h^{[m]}(v) = \prod_{j=1}^{n} \left(\frac{p_j + ix_j}{\sqrt{2}}\right)^{m_j}, \tag{2.5}$$

where we have used the multi-index notation $[m] = (m_1, \ldots, m_n)$.

Note: Whenever we use the term "analyticity" when speaking of a function for phase space, this means that $f(v) = f(x,p)$ is analytic in the variable $p + ix$; if f is analytic in the variable $p - ix$, we say that f is antianalytic on phase space. For a definition of these concepts without using an *a priori* decomposition of the phase space, see Ref. 1.

We shall often need the set of functions which can be written as a product of the Gaussian ω, (2.4), with an analytic function on E. We call these functions "modified holomorphic," and denote their set by $Z(E)$ or Z:

$$Z(E) = \{\phi: E \to \mathbb{C}; \phi = f\cdot\omega, \text{ with } f \text{ analytic on } E\}. \tag{2.6}$$

Note that the pointwise product of two modified holomorphic functions is not modified holomorphic, having a factor ω too many.

The square integrable modified holomorphic functions form a closed subspace of $L^2(E)$ (see Refs. 1 and 9); we shall denote this Hilbert space by $\mathcal{L}_0$:

$$\mathcal{L}_0(E) = \{\phi \in Z(E); \int dv |\phi(v)|^2 < \infty\}. \tag{2.7}$$

The measure on E used here is just the usual translationally invariant measure on E, with normalization fixed by the requirement

$$\int dv\, \omega^2(v) = \int dv \exp[-s(v,v)] = 1,$$

i.e.,

$$dv = \frac{1}{(2\pi)^n} d^n x d^n p. \tag{2.8}$$

For any function $\phi = f\cdot\omega$ in Z one can, of course, decompose the analytic function f into its Taylor series, which gives

$$\phi(v) = \sum_{[m]} a_{\phi,[m]}\, h^{[m]}(v)\omega(v), \tag{2.9}$$

where the convergence is uniform on compact sets. One can prove (see Ref. 9) that for $\phi \in Z$ one has

$$\phi \in \mathcal{L}_0 \Leftrightarrow \sum_{[m]} |a_{\phi,[m]}|^2 [m!] < \infty,$$

$$\phi, \psi \in \mathcal{L}_0 \Rightarrow (\phi,\psi) = \int dv\, \overline{\phi(v)}\, \psi(v) = \sum_{[m]} \overline{a_{\phi,[m]}}\, a_{\psi,[m]}\, [m!], \tag{2.10}$$

where

$$[m!] = \prod_{j=1}^{n} (m_j!).$$

Equation (2.10) implies that the set of functions $u_{[m]}$,

$$u_{[m]}(v) = \frac{1}{\sqrt{[m!]}} h^{[m]}(v)\omega(v), \tag{2.11}$$

constitutes an orthonormal base in $\mathcal{L}_0$, and that the series (2.9) converges not only uniformly on compact sets, but also in L^2 as long as $\phi \in \mathcal{L}_0$. We shall often rewrite (2.9) and (2.10) in the following way.

$\forall \phi \in Z$: we write a "modified Taylor expansion"

$$\phi(v) = \sum_{[m]} \phi_{[m]} u_{[m]}(v), \tag{2.12}$$

with uniform convergence on compact sets,

$$\forall \phi, \psi \in \mathcal{L}_0: (\phi,\psi) = \sum_{[m]} \overline{\phi_{[m]}}\, \psi_{[m]}, \tag{2.13}$$

in particular

$$\int dv\, \overline{u_{[m]}(v)}\, \phi(v) = (u_{[m]}, \phi) = \phi_{[m]}. \tag{2.14}$$

The same construction can be made in the space $E \times E$. Most functions on $E \times E$ in this study will have the property that they are *modified holomorphic in one variable and modified antiholomorphic in the other one. We denote the set of functions having this property by* $Z(E_2)$ (or shorter Z_2):

$$Z(E_2) = \{\phi: E \times E \to \mathbb{C}; \phi(v,v') = f(v,v')\omega(v)\omega(v'), \text{ with } f(v,v') \text{ analytic in } (p + ix, p' - ix')\}. \tag{2.15}$$

Again we can restrict ourselves to the square integrable functions in $Z(E_2)$:

$$\mathcal{L}_0(E_2) = Z(E_2) \cap L^2(E \times E); \tag{2.16}$$

again this is a closed subspace of $L^2(E \times E)$, with orthonormal basis

$$u_{[m_1,m_2]}(v_1,v_2) = u_{[m_1]}(v_1)\, \overline{u_{[m_2]}(v_2)}. \tag{2.17}$$

The analogs of (2.12) and (2.13) are now

$$\forall \phi \in Z(E_2): \phi(\zeta) = \sum_{[m_1],[m_2]} \phi_{[m_1,m_2]} u_{[m_1,m_2]}(\zeta), \tag{2.18}$$

with uniform convergence on compact sets,

$$\forall \phi, \psi \in \mathcal{L}_0(E_2): (\phi,\psi) = \sum_{[m],[m_2]} \overline{\phi_{[m_1,m_2]}}\, \psi_{[m_1,m_2]}, \tag{2.19}$$

in particular

$$\phi_{[m_1,m_2]} = \int d\zeta\, \overline{u_{[m_1,m_2]}(\zeta)}\, \phi(\zeta), \tag{2.20}$$

where we have used the notation $\zeta = (v_1,v_2)$ (in general, the Greek letters ζ,ξ will denote elements of $E \times E$).

Both the spaces $\mathcal{L}_0(E)$ and $\mathcal{L}_0(E_2)$ have "reproducing vectors" (this is a common feature for Hilbert spaces of analytic functions[13]):

$$\forall a \in E, \quad \forall \zeta = (a_1,a_2) \in E \times E,$$
$$\exists \omega^a \in \mathcal{L}_0(E), \quad \exists \omega^\zeta \in \mathcal{L}_0(E_2),$$

such that

$$\forall \phi \in \mathcal{L}_0(E): (\omega^a, \phi) = \phi(a),$$
$$\forall \phi \in \mathcal{L}_0(E_2): (\omega^\zeta, \phi) = \phi(\zeta). \tag{2.21}$$

These ω^a, ω^ζ are given explicitly by

$$\omega^a(v) = e^{i\sigma(a,v)}\, \omega(v - a) = \sum_{[m]} \overline{u_{[m]}(a)}\, u_{[m]}(v), \tag{2.22}$$

$$\omega^{\zeta}(\xi) = \omega^{a_1,a_2}(b_1,b_2) = \omega^{a_1}(b_1)\,\overline{\omega^{a_2}(b_2)}$$
$$= \sum_{[k],[l]} \overline{u_{[k,l]}(\zeta)}\, u_{[k,l]}(\xi). \tag{2.23}$$

The series in (2.22) and (2.23) converge uniformly on compact sets, but also in $L^2(E)$ (separately in a and v), respectively, $L^2(E\times E)$ (separately in ζ and ξ). Note that the reproducing properties (2.21), (2.14), and (2.20) can be proved for much more general classes of ϕ than only $\mathscr{L}_0$ (see below).

The integral kernel $\{a,b/v\}$

For any three points $a,b,v,\in E$ we define the function $\{a,b\,|v\}$ as follows (see Ref. 1):

$$\{a,b\,|v\} = 2^n e^{i[\sigma(a,b)+2\sigma(b,v)+2\sigma(v,a)]}\,\omega(2v-a-b)$$
$$= 2^n \exp\Big[i(\tfrac{1}{2}p_a x_b - \tfrac{1}{2}p_b x_a + p_b x_v - p_v x_b + p_v x_a - p_a x_v) - \Big(x_v - \frac{x_a+x_b}{2}\Big)^2 - \Big(p_v - \frac{p_a+p_b}{2}\Big)^2\Big]$$
$$= 2^n \exp[(p_v - ix_v)(p_a + ix_a) - \tfrac{1}{2}(p_b - ix_b) \times (p_a + ix_a) + (p_b - ix_b)(p_v + ix_v)] \times \omega(a)\omega(b)\omega(2v). \tag{2.24}$$

From the last expression in (2.24) it is obvious that $\{a,b\,|v\}$ is modified holomorphic in a, modified antiholomorphic in b, i.e.,

$$\forall v\in E: \{\cdot,\cdot\,|v\}\in Z_2. \tag{2.25}$$

Moreover, one can easily check (see Refs. 1 and 8) that $\{a,b\,|v\}$ has the following properties:

$$|\{a,b\,|v\}| < 2^n,$$
$$\int dv \{a,b\,|v\}\ \{c,d\,|v\} = \omega^a(c)\omega^d(b). \tag{2.26}$$

This function $\{a,b\,|v\}$ will be used to define two integral transforms

$$(If)(\zeta) = \int dv\, f(v)\{\zeta\,|v\}, \tag{2.27a}$$
$$(\tilde{I}\varphi)(v) = \int d\zeta\, \phi(\zeta)\ \overline{\{\zeta\,|v\}}. \tag{2.27b}$$

It is our purpose here to investigate some of the properties of these integral transforms and their extensions.

3. BILINEAR EXPANSION OF $\{a,b/v\}$—THE FUNCTIONS $h_{[k,l]}$

A. Bilinear expansion of $\{a,b/v\}$

As elements of $Z(E_2)$, the functions $\{\cdot\,|v\}$ can be developed in a series with respect to the $u_{[k,l]}$ [see (2.13)]:

$$\{a,b\,|v\} = \{\zeta\,|v\} = \sum_{[k],[l]} u_{[k,l]}(\zeta) h_{[k,l]}(v). \tag{3.1}$$

The $h_{[k,l]}$ are defined, up to a factor $\sqrt{[k!]\,[l!]}$, as derivatives of the function $\{\zeta\,|v\}\,(\omega_2(\zeta))^{-1}$ in $\zeta = 0$:

$$h_{[k,l]}(v) + 2^n\sqrt{[k!][l!]}\ 2^{(|k|+|l|)/2} \times \left[\frac{d^{[k]}}{dp_a^{[k]}}\frac{d^{[l]}}{dp_a^{[l]}}\{a,b\,|v\}\omega(a)^{-1}\omega(b)^{-1}\right]_{a=b=0}. \tag{3.2}$$

Because of the explicit form (2.24) of $\{\zeta\,|v\}$ it is obvious that every $h_{[k,l]}$ is a polynomial in v multiplied by the Gaussian $\omega(2v) = \exp[-(x_v^2 + p_v^2)]$. This automatically implies that all the $h_{[k,l]}$ are elements of $\mathscr{S}(E)$, the Schwartz space of C^∞ function which decrease faster than any negative power of (x,p).

B. Orthonormality of the $h_{[k,l]}$

On the other hand, we have [see (2.26)]

$$\int dv\ \overline{\{a,b\,|v\}}\ \{c,d\,|v\} = \omega^a(c)\omega^d(b).$$

Multiplying both sides with $\omega(a)^{-1}\,\omega(b)^{-1}$, and computing derivatives with respect to p_a, p_b, we obtain (it is obvious from the explicit form of $\{a,b\,|v\}$ that these derivatives can be commuted with the integral in the left hand side)

$$\int dv\ \overline{h_{[k,l]}(v)}\ \{c,d\,|v\} = u_{[k]}(c)\ \overline{u_{[l]}(d)}. \tag{3.3}$$

Repeating the same operations in the variables c and d, we obtain

$$\int dv\ \overline{h_{[k,l]}(v)}\ h_{[k',l']}(v) = \delta_{[k],[k']}\,\delta_{[l][l']}, \tag{3.4}$$

implying that the $h_{[k,l]}$ form an orthonormal set in $L^2(E)$.

C. Completeness of the $h_{[k,l]}$

From (3.3) we see that the coefficients with respect to the orthonormal set $h_{[k,l]}$ of the orthogonal projection of any $\{\zeta\,|\}$ (ζ fixed) onto the closed linear span of the $h_{[k,l]}$ are exactly the $u_{[k,l]}(\zeta)$; comparing this with (3.1) we conclude that for any ζ the function $\{\zeta\,|\cdot\}$ is an element of the closed span of the $h_{[k,l]}$; (3.1) can now be considered to be the composition in L^2 of $\{\zeta\,|\cdot\}$ with respect to the orthonormal set $h_{[k,l]}$. From this it is now easy to see that the closed span of the $h_{[k,l]}$ is all of $L^2(E)$. Indeed, let ψ be orthogonal to all $h_{[k,l]}$:

$$\forall[k],[l]: (\psi, h_{[k,l]}) = 0.$$

Then

$$\forall a,b: (\psi, \{a,b\,|\cdot\}) = 0$$
$$\Rightarrow \forall c: \int dv\, \psi(v)\, e^{i\sigma(c,v)}\,\omega(2v) = 0$$
$$\Rightarrow \psi(v)\cdot\omega(2v) = 0 \text{ a.e. } \Rightarrow \psi = 0$$

[ω is bounded, and the Fourier transform is unitary on $L^2(E)$]. Hence the $h_{[k,l]}$ constitute an orthornormal base for $L^2(E)$.

Note: The properties in Secs. 3B and 3C were already stated in Ref. 1, in more generality (valid also for the coefficient functions of other bilinear expansions of $\{a,b\,|v\}$), without proof. It is possible to prove them (see Ref. 8) using Godement's theorem on irreducible square integrable representations of unimodular locally compact groups. In

the special case of the $h_{[k,l]}$, however, one can also prove them with very simple arguments, as shown here.

D. Unitarity of the integral transform I

It is now easy to show that the integral transform I, as defined by (2.27), defines a unitary operator from $L^2(E)$ onto $\mathscr{L}_0(E_2)$.

Proposition 3.1: The integral transform I:

$$If(\zeta) = \int dv \{\zeta | v\} f(v) \tag{3.5}$$

with $\{\zeta | v\} = \{a,b | v\}$ as defined by (2.24), defines a unitary operator from $L^2(E)$ to $\mathscr{L}_0(E_2)$; in particular

$$Ih_{[k,l]} = u_{[l,k]}. \tag{3.6}$$

Proof: We start by defining a linear operator on the span of the $h_{[k,l]}$ by putting

$$Uh_{[k,l]} = u_{[l,k]}.$$

Since the $h_{[k,l]}$, $u_{[k,l]}$ constitute orthonormal bases in $L^2(E)$, $\mathscr{L}_0(E_2)$, respectively, this U can be extended to a unitary operator from $L^2(E)$ onto $\mathscr{L}_0(E_2)$. In particular,

$$\begin{aligned} U(\{b,a|\cdot\}) &= U\Big(\sum_{[k][l]} u_{[k,l]}(b,a) h_{[k,l]}\Big) \\ &= \sum_{[k][l]} \overline{u_{[l,k]}(a,b)}\, u_{[l,k]} = \omega^{(a,b)}, \end{aligned}$$

where we have used (2.23). Take now any ϕ in $L^2(E)$. Then $U\phi \in \mathscr{L}_0(E_2)$, and its value at any point is given by the reproducing property (2.21),

$$\begin{aligned} (U\varphi)(\zeta) &= (\omega^\zeta, U\varphi) = (U^* \omega^\zeta, \varphi) \\ &= (\{\bar{\zeta} | \cdot\}, \varphi) \\ &= \int dv\, \overline{\{\bar{\zeta} | v\}} \varphi(v) \\ &\quad [\text{for } \zeta = (a,b), \text{ we denote } (b,a) \text{ by } \bar{\zeta}\,]. \end{aligned}$$

Hence $(U\phi)(\zeta) = (I\phi)(\zeta)$, which proves the proposition.■

Remarks:

1. A different proof of the unitarity of I between $L^2(E)$ and $\mathscr{L}_0(E_2)$ was given in Sec. 4E in Ref. 1 (the argument given there is not completely rigorous, but it can easily be transformed into a rigorous one).

2. The integral in (3.5) converges absolutely for any $f \in L^2(E)$, since $\{\zeta | \cdot\}$ is in $L^2(E)$ for each fixed ζ. The situation is different if one tries to apply $\tilde{I}$ to $\mathscr{L}_0(E_2)$: since $\{\cdot | v\}$ is not square integrable on $E \times E$, the integral transform (2.27b) cannot be defined on all of $\mathscr{L}_0(E_2)$. One has, however,

$$\int d\zeta\, \overline{\{\zeta | v\}}\, u_{[k,l]}(\zeta) = h_{[l,k]}(\zeta), \tag{3.7}$$

where the integral converges absolutely because $u_{[l,k]}$ is absolutely integrable. So

$$\tilde{I}Ih_{[k,l]} = h_{[k,l]}, \tag{3.8}$$

which leads one to believe that $\tilde{I}$ is the inverse of I. Indeed, if one tries to circumvent the problem of possible divergence of the integral by taking limiting procedures, one finds (as in Ref. 9a), e.g.,

$$\begin{aligned} \forall \varphi \in \mathscr{L}_0(E_2) : I^{-1}\varphi &= L^2 - \lim_{R\to\infty} \tilde{I}(\chi_R \varphi) \\ &= L^2 - \lim_{\alpha\to 0} \tilde{I}(\omega(\alpha\cdot)\varphi), \end{aligned} \tag{3.9}$$

where

$$\chi_R(\zeta) = \begin{cases} 1, & |\zeta| < R \\ 0, & |\zeta| > R \end{cases} \quad \begin{aligned} &[\text{here } |\zeta|^2 = |a|^2 + |b|^2 \\ &\text{and } |a|^2 = s(a,a)]. \end{aligned}$$

The same is true for any other reasonable limiting procedure.

E. Other properties of the $h_{[k,l]}$

One can show (see Refs. 1 and 8) that

$$|h_{[k,l]}(v)| \leq 1. \tag{3.10}$$

Explicit calculation of the $h_{[k,l]}$ yields (see Refs. 7 or 6)

$$\begin{aligned} h_{[k,l]}(x,p) = 2^n e^{-x^2-p^2} \sum_{[s]=0}^{\min([k],[l])} \Bigg[(-2)^{-|s|} 2^{(|k|+|l|)/2} \\ \frac{\sqrt{[k!][l!]}}{[s!][(l-s)!][(k-s)!]} (p+ix)^{[l-s]} (p-ix)^{[k-s]} \Bigg]. \end{aligned} \tag{3.11}$$

One can check (by direct calculation) that these $h_{[k,l]}$ are the eigenfunctions of a dilated harmonic-oscillator-type operator on phase space E:

$$(-\tfrac{1}{4}(\Delta_x + \Delta_p) + x^2 + p^2) h_{[k,l]} = (|k| + |l| + m) h_{[k,l]}. \tag{3.12}$$

As a consequence of this, the $h_{[k,l]}$ are linear combinations of products of Hermite functions:

$$h_{[k,l]}(x,p) = \sum_{\substack{[r][s] \\ |r|+|s| = |k|+|l|}} \alpha_{[k,l],[r,s]} H_{[r,s]}(x,p), \tag{3.13}$$

with

$$H_{[r,s]}(x,p) = 2^n H_{[r]}(\sqrt{2}x) H_{[s]}(\sqrt{2}p), \tag{3.14}$$

$$\sum_{[r],[s]} |\alpha_{[k,l],[r,s]}|^2 = 1 \tag{3.15}$$

and where $H_{[r]}$ is the $[r]$th order Hermite function.

There exists also a relationship between the $h_{[k,l]}$ and the Laguerre polynomials (see Ref. 7). For $n = 1$, $k = l$ one has for instance

$$h_{kk}(x,p) = 2(-1)^k e^{-(x^2+p^2)} L_k(2x^2 + 2p^2), \tag{3.16}$$

where L_k is the Laguerre polynomial of order k.

One can also prove the following recurrence relation for the $h_{[k,l]}$:

$$\begin{aligned} (|k| + |l|)\, h_{[k,l]} &= \sqrt{|k|} \sum_j k_j (p_j - ix_j) h_{[k-\delta_j,l]} \\ &\quad + \sqrt{|l|} \sum_j l_j (p_j + ix_j) h_{[k,l-\delta_j]} \\ &\quad - \sqrt{|k|\,|l|} \sum_j h_{[k-\delta_j,l-\delta_j]} k_j l_j, \end{aligned} \tag{3.17}$$

where $[\delta_j]$ is the multi-index $(\delta_j)_m = \delta_{jm}$.

4. THE INTEGRAL TRANSFORM I AS A CONTINUOUS MAP BETWEEN SUITABLE FAMILIES OF BANACH SPACES AND HILBERT SPACES

As was already mentioned in the Introduction, there exists a link between our integral transform I and the Bargmann integral transform as defined in Ref. 9. Since we shall use this link to derive some of our results, we shall first show here what exactly is the connection between the two integral transforms. For $z\in\mathbb{C}^n$, $q\in\mathbb{R}^n$, the Bargmann integral kernel $A(z,q)$ is given by[9]

$$A(z,q)=\pi^{-n/4}\exp[-\tfrac{1}{2}(z^2+q^2)+\sqrt{2}\,z\cdot q]. \tag{4.1}$$

Identifying z with $(1/\sqrt{2})(x-ip)$ (which makes of the multiplication by z—on a suitable Hilbert space of analytic functions—a representation of the harmonic oscillator creation operator: see Ref. 9a), we can rewrite (4.1) as

$$A(x,p;q)\cdot e^{-1/4(x^2+p^2)}=\pi^{-n/4}\,e^{(1/2)xp}\,e^{-ipq}\,e^{-(1/2)(x-q)^2}. \tag{4.2}$$

Comparing this with (2.24) we see that

$$\{a,b\,|v\}=2^n\pi^{n/2}A\left(\frac{x_a+x_b}{\sqrt{2}},\frac{p_b-p_a}{\sqrt{2}};\sqrt{2}x_v\right)$$
$$A\left(\frac{p_a+p_b}{\sqrt{2}},\frac{x_a-x_b}{\sqrt{2}};\sqrt{2}\,pv\right)$$
$$e^{-(1/4)(x_a^2+x_b^2+p_a^2+p_b^2)} \tag{4.3}$$

or

$$\{a,b\,|v\}=2^n r^{n/2}A(c_{ab};\sqrt{2}x_v)\,A(d_{ab};\sqrt{2}p_v)\omega(c_{ab})\cdot\omega(d_{ab}),$$

with

$$c_{ab}=\frac{1}{\sqrt{2}}(x_a+x_b,p_b-p_a), \tag{4.4}$$
$$d_{ab}=\frac{1}{\sqrt{2}}(p_a+p_b,x_a-x_b).$$

Actually, (4.3) implies that we can consider the integral transform I as a $2n$-dimensional Bargmann transform. The explicit Gaussian factors $\omega(c_{ab})$, $\omega(d_{ab})$ just compensate for the difference in definition between our Hilbert space $\mathcal{L}_0(E_2)$ and the Bargmann Hilbert space [we absorb the Gaussian in the functions in $\mathcal{L}_0(E_2)$, whereas in Ref. 9 it is always displayed as a weight function in the definition of the inner product]. The constant factors $2^n\pi^{n/2}$ account for the difference in normalization in the measure, and for the dilation in x_v,p_v. Moreover, one can easily check that analyticity in c_{ab},d_{ab} is equivalent to analyticity in a, antianalyticity in b. So, from a mathematical point of view, I can be assimilated with a $2n$-dimensional Bargmann transform. Physically however, the two integral transforms have a different meaning: I gives a correspondence between classical and quantum aspects, while the Bargmann transform gives the unitary transformation between two different but equivalent realizations (for a short discussion, see Sec. 6 in Ref. 1).

The remarks above will enable us to translate various results obtained by Bargmann in Ref. 9b to the present setting. An example of this is the following (we keep the same notations as in Ref. 9, even though the functions considered here are in fact modified holomorphic instead of holomorphic):

Define

$$\forall\varphi\in Z(E_2):|\varphi|_\rho=\sup_\zeta|(1+|\zeta|^2)^\rho\varphi(\zeta)|,$$
$$\mathfrak{E}=\{\varphi\in Z(E_2);\quad\forall\rho\in\mathbb{R};|\varphi|_\rho<\infty\},$$
$$\mathfrak{E}'=\{\varphi\in Z(E_2);\quad\exists\rho\in\mathbb{R}\text{ such that }|\varphi|_\rho<\infty\}.$$

The spaces $\mathfrak{E}$, $\mathfrak{E}'$ can be equipped with very natural locally convex topologies by means of the norms $\|\,\|_\rho$, and one has then the following result.

I defines an isomorphism between $\mathcal{S}(E)$ and $\mathfrak{E}$, with

$$\forall f\in\mathcal{S}(E):(If)(\zeta)=\int dv\{\zeta\,|v\}f(v);$$

by duality, I defines also an isomorphism between $\mathcal{S}'(E)$ and $\mathfrak{E}'$, with

$$\forall T\in\mathcal{S}'(E):(IT)(\zeta)=\overline{T(\{\bar{\zeta}\,|\cdot\})}.$$

The results in Ref. 9b also concern two families of Banach spaces interpolating between $\mathcal{S}$ and $\mathcal{S}'$, $\mathfrak{E}$ and $\mathfrak{E}'$, respectively, and between which the integral transform I or its inverse are continuous. We give a survey of these results, translated to our present setting, in Sec. 4A.

The chains of Banach spaces presented in Sec. 4A display, however, several inconveniences. As already mentioned in Ref. 9b the $\mathfrak{E}^\rho$ spaces are not separable, and the little space $\mathfrak{E}$ is not dense in any $\mathfrak{E}^\rho$. Moreover, in relation to the present setting, it turns out that though one can always choose suitably matched spaces in the two ladders to make either I or its inverse continuous, it is impossible to choose them in such a way that I is an isomorphism. None of these problems arises when one uses a suitable interpolating chain of Hilbert spaces instead of Banach spaces (see Sec. 4B). The resulting bounds on I are much more precise between these Hilbert spaces, and therefore more useful for applications to quantization than the results of Sec. 4A.

Generalizing the construction of the Hilbert spaces in Sec. 4B, one can obtain even larger families containing spaces smaller than $\mathcal{S}$ (or $\mathfrak{E}$) or larger than $\mathcal{S}'(\mathfrak{E}')$, on which the integral transform I can still be defined and has continuity properties. The results in Secs. 4B and 4C can be considered as extensions of the bounds in Ref. 9b (Sec. 4B uses some estimates made in Ref. 9b). Other results on the Bargmann transform can, of course, easily be translated to the present context and be useful in a Weyl quantization setting (see, e.g., Ref. 1, where a characterization of the images under the Bargmann transform of the Gel'fand–Shilov spaces S and S^* are given; in a sense this can be considered as complementary to our results in Sec. 4C).

A. The Banach spaces $\mathcal{S}^k$, $\mathfrak{E}^\rho$ and related results on the integral transform I

For any C^k function f on E, we define (this norm is the same as in Ref. 9b, up to a dilation:

$$|f|_k^S = ||f\left(\frac{\cdot}{\sqrt{2}}\right)||_{k,\text{Bargmann}}^S)$$

$$|f|_k^S = \max_{\substack{[m_1],[m_2] \\ |m_1|+|m_2|\leq k}} \sup_{x,p} |2^{-(|m_1|+|m_2|)/2} (1+2x^2+2p^2)^{(k-|m_1|-|m_2|)/2} (\nabla_x^{[m_1]} \nabla_p^{[m_2]} f)(x,p)|. \tag{4.5}$$

The Banach space $\mathcal{S}^k$ is then defined as

$$\mathcal{S}^k = \{ f: E \to \mathbb{C}; f \text{ is } C^k, |f|_k^S < \infty \}. \tag{4.6}$$

On the other hand, we define, $\forall \rho \in \mathbb{R}$, the following subspaces $\mathfrak{E}^\rho$ of $Z(E_2)$:

$$\mathfrak{E}^\rho = \{ \varphi \in Z(E_2); |\varphi|_\rho = \sup_\zeta |(1+|\zeta|^2)^\rho \varphi(\zeta)| < \infty \}. \tag{4.7}$$

The following theorem was proved in Ref. 9b.

Theorem 4.1:

1. $\forall f \in \mathcal{S}^k$, the function

$$If(\zeta) = \int dv \{\zeta|v\} f(v)$$

is well defined and an element of $Z(E_2)$. Moreover,

$$If \in \mathfrak{E}^k$$

and

$$|If|_k \leq b_k |f|_k^s, \tag{4.8}$$

with

$$b_k = \frac{3e}{2} 2^{n/2} (16n)^{k/2} \begin{cases} 1, & k \leq 2 \\ e^{-k} k^k, & k \geq 3 \end{cases}. \tag{4.9}$$

2. $\forall \varphi \in \mathfrak{E}^\mu$, with $\mu > 2n$, the function

$$\tilde{I}\varphi(v) = \int dv \{\xi|v\} \varphi(\zeta)$$

is well defined on E. Moreover,

$$\forall k \in \mathbf{N},$$

with

$$k < \mu - 2n : \tilde{I}\varphi \in \mathcal{S}^k$$

and

$$|\tilde{I}\varphi|_k^s \leq b'_{k,\mu} |\varphi|_k, \tag{4.10}$$

with

$$b'_{k,\mu} = 2^{n+k} \left(\frac{2}{3}\right)^{k/2} \Gamma\left(\frac{k}{2}+m+1\right) \Gamma\left(\frac{\mu-k}{2}-m\right) \times \Gamma\left(\frac{\mu-k}{2}\right)^{-1} \int_E dv\, e^{-2|v|^2} (1+|v|^2)^k. \tag{4.11}$$

$$3.\ f \in \bigcup_{k>2n+1} \mathcal{S}^k \Rightarrow \tilde{I}If = f, \tag{4.12}$$

$$\varphi \in \bigcup_{\mu>2n} \mathfrak{E}^\mu \Rightarrow I\tilde{I}\varphi = \varphi.$$

Note: This theorem was used in Ref. 6 to derive some restrictions in the class of distributions corresponding to bounded operators.

It is obvious from the definition (4.6) of the $\mathcal{S}^k$ spaces that $\mathcal{S}(E) = \bigcap_{k\in\mathbf{N}} \mathcal{S}^k$, and that the locally convex topology on $\mathcal{S}$ defined by the $\|\|_k^S$-norms coincides with the usual Schwartz topology. Defining, on the other hand,

$$\mathfrak{E} = \{ \varphi \in Z(E_2); \forall \rho : |\varphi|_\rho < \infty \}, \tag{4.13}$$

and equipping this space with the locally convex topology induced by the norms $\|\|_\rho$, we have immediately the following corollary to Theorem 4.1.

Corollary 4.2: The integral transform I, restricted to $\mathcal{S}$, defines an isomorphism from $\mathcal{S}$ onto $\mathfrak{E}$, with inverse $\tilde{I}$ (restricted to $\mathfrak{E}$).

In Ref. 9b it was shown that $\mathfrak{E}'$, the dual of $\mathfrak{E}$, can be identified with $\cup_{\rho\in\mathbf{R}} \mathfrak{E}^\rho$ in the following way:

$$\forall L \in \mathfrak{E}' : g_L(\zeta) = \overline{L(\omega^\zeta)} \Rightarrow g_L \in \bigcup_{\rho\in\mathbf{R}} \mathfrak{E}^\rho,$$

$$\forall \varphi \in \bigcup_{\rho\in\mathbf{R}} \mathfrak{E}^\rho : L_\varphi(\psi) = \int d\zeta\, \overline{\varphi(\zeta)} \psi(\zeta) \Rightarrow L_\varphi \in \mathfrak{E}', \tag{4.14}$$

with

$$g_{L_\varphi} = \varphi, \quad L_{g_L} = L.$$

The topology on $\mathfrak{E}'$ corresponds with the natural topology on $\cup_{\rho\in\mathbf{R}} \mathfrak{E}^\rho$ induced by the norms $||_\rho$. In what follows, we shall always identify $\mathfrak{E}'$ with $\cup_{\rho\in\mathbf{R}} \mathfrak{E}^\rho$ and implicitly use (4.14).

Since I is an isomorphism between $\mathcal{S}$ and $\mathfrak{E}$, it is obvious that by duality I also defines an isomorphism between $\mathcal{S}'$ and $\mathfrak{E}'$:

$$\forall T \in \mathcal{S}' : \text{we define } (IT)(\varphi) = T(\tilde{I}\varphi), \quad \forall \varphi \in \mathfrak{E}. \tag{4.15}$$

By means of the identification $\mathfrak{E}' = \cup_{\rho\in\mathbf{R}} \mathfrak{E}^\rho$, we define the function $IT(\zeta)$ as

$$IT(\zeta) = \overline{IT(\omega^\zeta)} = \overline{T(\tilde{I}\omega^\zeta)} = \overline{T(\{\tilde{\zeta}|\cdot\})}.$$

One can easily check that for $f \in \mathcal{S}$, this new definition of If coincides with the old If defined as an integral transform. We have now immediately

Theorem 4.3: $\forall T \in \mathcal{S}'$, the function $IT(\zeta) = T(\{\tilde{\zeta}|\cdot\})$ is a well-defined function on $E \times E$, with $IT \in \mathfrak{E}'$. This map $I : \mathcal{S}' \to \mathfrak{E}'$ is an isomorphism extending the isomorphism in Corollary 4.2.

Remarks: 1. The inverse map of $I : \mathcal{S}' \to \mathfrak{E}'$, with I defined as in (4.15), can again be constructed by combining $\tilde{I}$ and a limiting procedure. For instance,

$$\forall \varphi \in \mathfrak{E}' : I^{-1}\varphi = s' - \lim_{k\to\infty} \tilde{I}(\varphi \cdot \chi_R). \tag{4.16}$$

2. One can enter in some detail into a discussion of I as an isomorphism between $\mathcal{S}'$ and $\mathfrak{E}'$, and compute explicit bounds on $|IT|_\rho$ for T in $(\mathcal{S}^k)'$, using the bounds in Theorem 4.1 (see Refs. 9b or 8).

So finally I defines an isomorphism between $\mathcal{S}$ and $\mathfrak{E}$ and between $\mathcal{S}'$ and $\mathfrak{E}'$. Moreover, we have two sets of interpolating spaces: the $\mathcal{S}^k$ $(\mathcal{S}^k)'$ between $\mathcal{S}$ and $\mathcal{S}'$ and the $\mathfrak{E}^\rho$ between $\mathfrak{E}$ and $\mathfrak{E}'$, and we have at hand continuity statements and bounds for I between elements of these two interpolating chains, giving more detailed information on the action of I. Except for the two ends of the chain we have, however, no bicontinuity of I, considered as a map from $\mathcal{S}^k$ [or $(\mathcal{S}^k)'$] to a suitably chosen $\mathfrak{E}^\rho$. This problem will

not occur with the chains of Hilbert spaces in the next subsection.

B. The Hilbert spaces $\mathcal{F}^\rho, W^\rho$, and related results concerning I

The Hilbert spaces $\mathcal{F}^\rho, W^\rho$ we define below constitute again two chains interpolating $\mathfrak{E}$ with $\mathfrak{E}'$, $\mathcal{S}$ with $\mathcal{S}'$ respectively. Actually the $\mathcal{F}^\rho$ spaces were already introduced in Ref. 9b as a tool for studying $\mathfrak{E}'$; they are weighted L^2 spaces of modified holomorphic functions. Their inverse images under the Bargmann integral transform were not displayed in Ref. 9b; we call these spaces W^ρ spaces; essentially they are the Hilbert spaces associated to the N-representation of $\mathcal{S}(E)$, $\mathcal{S}'(E)$ with respect to the harmonic oscillator-type operator $x^2+p^2-\frac{1}{4}\Delta_x-\frac{1}{4}\Delta_p$ (see, e.g., Ref. 14).

The $\mathcal{F}^\rho$ spaces

The $\mathcal{F}^\rho$ spaces are defined as ($\rho\in\mathbb{R}$)

$$\mathcal{F}^\rho=\left\{\varphi\in Z(E_2);\|\varphi\|_\rho^2=\int d\zeta\,(1+|\zeta|^2)^\rho|\varphi(\zeta)|^2<\infty\right\}, \tag{4.17}$$

with associated inner product:

$$(\varphi,\psi)_\rho=\int d\zeta\,(1+|\zeta|^2)^\rho\,\overline{\varphi(\zeta)}\,\psi(\zeta). \tag{4.18}$$

The $\mathcal{F}^\rho$ spaces are Hilbert spaces [$\mathcal{F}^0=\mathcal{L}_0(E_2)$]; one can check (see Appendix A or Ref. 9b) that the $u_{[k,l]}$ are orthogonal elements of the $\mathcal{F}^\rho$:

$$(u_{[k,l]},u_{[k',l']})_\rho=\delta_{[k][k']}\delta_{[l][l']}\tau(\rho;|k|+|l|) \tag{4.19}$$

with

$$\tau(\rho;m)=\Gamma(m+2n)^{-1}\int_0^\infty dx\,x^{m+2n-1}e^{-x}(1+x)^\rho.$$

Moreover, for any $\phi\in\mathcal{F}^\rho$ with series expansion (2.18) one has

$$\|\phi\|_\rho^2=\sum_{[k][l]}|\phi_{[k,l]}|^2\tau(\rho;|k|+|l|), \tag{4.20}$$

and

$$\phi_{[k,l]}=\int d\zeta\;\overline{u_{[k,l]}(\zeta)}\,\phi(\zeta). \tag{4.21}$$

Equations (4.19) and (4.20) imply that the $\tau(\rho;|k|+|l|)^{-1/2}u_{[k,l]}$ constitute an orthonormal base of $\mathcal{F}^\rho$.

The following estimates for $\tau(\rho;m)$ were computed in Ref. 9b:

$$c'_\rho\leqslant\tau(\rho;m)(m+2n)^{-\rho}\leqslant c''_\rho, \tag{4.22}$$

with

$$\left.\begin{aligned}c'_\rho&=\left(1+\frac{\rho}{2n}\right)^{-1}\\ c''_\rho&=\left(1+\frac{\rho}{2n}\right)^{\rho+2n}e^{1-\rho}\end{aligned}\right\}\rho\geqslant 0,$$
$$\left.\begin{aligned}c'_\rho&=\left(1-\frac{\rho}{2n}\right)^{-2n+\rho}e^{-1-\rho}\\ c''_\rho&=e^{-\rho}\end{aligned}\right\}\rho\leqslant 0. \tag{4.23}$$

The W^ρ-spaces

We put $\forall f\in\mathcal{S}(E),\forall\rho\in\mathbb{R}$:

$$(\|f\|_\rho^s)^2=(f,(x^2+p^2-\tfrac{1}{4}\Delta_x-\tfrac{1}{4}\Delta_p+n)^\rho f). \tag{4.24}$$

Note: Actually, the operator $x^2+p^2-\frac{1}{4}\Delta_x-\frac{1}{4}\Delta_p$ has spectrum $\mathbb{N}\cap[n,\infty]$, which implies we could drop the extra term n in (4.24): the resulting topology on W would be exactly the same. We nevertheless introduce the extra term n in order to obtain the sharpest possible estimates on the integral transform I: to obtain these estimates, we shall use (4.22), where this extra n is already present.

We define then W^ρ as the closure of $\mathcal{S}(E)$ with respect to the norm $\|\|_\rho^s$; equipped with this norm, W^ρ is a Hilbert space.

The renormalized Hermite functions $(|r|+|s|+2n)^{-\rho/2}H_{[r,s]}$ [see (3.14)] constitute an orthonormal base in W^ρ; one has

$$\forall T\in\mathcal{S}'(E):T\in W^\rho\Leftrightarrow\sum_{[r][s]}|T(H_{[r,s]})|^2(|r|+|s|+2n)^\rho<\infty \tag{4.25}$$

and

$$T\in W^\rho\Rightarrow(\|T\|_\rho^s)^2=\sum_{[r][s]}|T(H_{[r,s]})|^2(|r|+|s|+2n)^\rho$$

(see, e.g., Sec. V.5 in Ref. 13).

Because of (3.13) and (3.15), we can rewrite (4.25) in terms of the $h_{[k,l]}$:

$$\forall T\in\mathcal{S}'(E):\text{define}\quad T_{[k,l]}=\overline{\overline{T(\,h_{[k,l]})}}=\overline{T(h_{[l,k]})}. \tag{4.26}$$

Then

$$\forall T\in W^\rho:(\|T\|_\rho^s)^2=\sum_{[k][l]}|T_{[k,l]}|^2(|k|+|l|+2n)^\rho. \tag{4.27}$$

The integral transform I as a map from W^ρ onto $\mathcal{F}^\rho$

From the definitions of $W^\rho,\mathcal{F}^\rho$ one can check that

$$\bigcap_{\rho\in\mathbb{R}}W^\rho=\mathcal{S}(E),\quad\bigcup_{\rho\in\mathbb{R}}W^\rho=\mathcal{S}'(E),$$

$$\bigcap_{\rho\in\mathbb{R}}\mathcal{F}^\rho=\mathfrak{E},\quad\bigcup_{\rho\in\mathbb{R}}\mathcal{F}^\rho=\mathfrak{E}'.$$

The extended definition (4.15) of the integral transform I can therefore be applied to all W^ρ; for any $T\in W^\rho$, the resulting IT will be in $\mathfrak{E}'$ and have series expansion

$$IT(\zeta)=\sum_{[m][n]}(IT)_{[m,n]}u_{[m,n]}(\zeta),$$

with

$$\begin{aligned}(IT)_{[m,n]}&=\int d\zeta\;\overline{u_{[m,n]}(\zeta)}\,IT(\zeta)\quad[\text{use (4.21)}]\\ &=\overline{IT(u_{[m,n]})}\quad[\text{use (4.14)}]\\ &=\overline{T(h_{[n,m]})}\quad[\text{use (4.15) and Proposition 3.1}]\\ &=T_{[m,n]}.\end{aligned} \tag{4.28}$$

Using the definitions of the norms $\|\|_\rho$ and $\|\|_\rho^s$, and the estimates (4.22) we see now that

$$T\in W^\rho\Leftrightarrow IT\in\mathcal{F}^\rho \tag{4.29}$$

and

$$c'_\rho(\|T\|^s_\rho)^2 \le \|IT\|^2_\rho \le c''_\rho(\|T\|^s_\rho)^2. \tag{4.30}$$

Hence the following theorem.

Theorem 4.4: The map $I: IT(\zeta) = \overline{T(\{\zeta\,|\,\})}$ defines an isomorphism from W^ρ onto $\mathcal{F}^\rho$, and this $\rho \in \mathbb{R}$. Estimates on the norms of this isomorphism and its inverse are given by

$$\|I\|_{W^\rho \to \mathcal{F}^\rho} \le c''^{1/2}_\rho, \|I^{-1}\|_{\mathcal{F}^\rho \to W^\rho} \le c'^{-1/2}_\rho, \tag{4.31}$$

where C'_ρ and C''_ρ are defined by (4.23).

Remarks: As we announced before, the restriction of I to a W^ρ is a bijection onto $\mathcal{F}^\rho$, which means we have no qualitative loss of information when mapping to and fro (this was not the case for the $\mathcal{S}^k, \mathcal{E}^\rho$). Due to the fact that the product of the estimates on the norms in (4.31) is larger than 1, we have, however, still a "quantitative" loss of information, which gets worse for large $|\rho|$.

Up to now, we have considered the spaces $\mathcal{S}^k \mathcal{E}^\mu$ and later $\mathcal{F}^\rho, W^\rho$, in order to obtain some fine structure in the study of I as an isomorphism from $\mathcal{S}'$ to $\mathcal{E}'$; it turns out that the Hilbert spaces $\mathcal{F}^\rho, W^\rho$ are better suited to this end than the Banach spaces $\mathcal{S}^k, \mathcal{E}^\rho$. Our ultimate aim is to use these results to derive properties of the Weyl quantization procedure, using the fact, mentioned in the Introduction, that the integral transform I constitutes the link between a classical function and the coherent state matrix elements of its quantal counterpart. Theorems 4.1 and 4.4 can then be used to translate restrictions on a tempered distribution to growth restrictions on the coherent state matrix elements of the corresponding operator. A first application of Theorem 4.1 was given in Ref. 6, where it was also noted that stronger results could be obtained by means of Theorem 4.4. Other applications shall be given in Ref. 15.

C. The Hilbert spaces $\mathcal{F}^G, W^\phi$

We shall here generalize the structures of both $\mathcal{F}^\rho, W^\rho$ to obtain Hilbert spaces larger than $\mathcal{S}'$, and which can still be handled by I.

The Hilbert spaces $\mathcal{F}^G$

$\mathcal{F}^\rho$ was constructed as a weighted L^2 space of $Z(E_2)$ functions, with the special weight $(1 + |\zeta|^2)^\rho$. To generalize this construction, we consider now more general weights.

Let G be a function from $\mathbb{R}^+$ to $\mathbb{R}^+$. We define

$$\mathcal{F}^G = \{\phi \in Z(E_2); \|\phi\|^2_G = \int d\zeta\, |\phi(\zeta)|^2 G(|\zeta|^2) < \infty\} \tag{4.32}$$

and we equip this space with the norm $\|\ \|_G$. Since one has to be careful with Hilbert spaces of analytic functions, we shall first investigate the conditions to impose on G to ensure that $\mathcal{F}^G$ is an infinitely dimensional Hilbert space (see also Ref. 16).

Proposition 4.5:

1. If $\forall r \in \mathbb{R}^+: \operatorname{ess\,inf}_{x<r} G(x) > 0, \operatorname{ess\,sup}_{x<r} G(x) < \infty$, (4.33)

then $\mathcal{F}^G$ is complete.

2. Define

$$\lambda^G_m = \frac{1}{\Gamma(m+2n)} \int_0^\infty dx\, x^{m+2n-1} e^{-x} G(x). \tag{4.34}$$

A necessary and sufficient condition for $\mathcal{F}^G$ to be infinitely dimensional is

$$\forall m: \lambda^G_m < \infty. \tag{4.35}$$

If this condition is satisfied, then

$$\forall [k],[l]: u_{[k,l]} \in \mathcal{F}^G$$

and

$$\forall \phi = \sum_{[k][l]} \phi_{[k,l]} u_{[k,l]} \in \mathcal{F}^G: \|\phi\|^2_G = \sum_{[k][l]} |\phi_{[k,l]}|^2 \lambda^G_{|k|+|l|}. \tag{4.36}$$

The following three conditions involve not only G, but also $1/G$:

3. If

$$\forall m: \lambda^{1/G}_m < \infty, \text{ then } \forall \phi \in \mathcal{F}^G: \phi_{[k,l]} = \int d\zeta\, \overline{u_{[k,l]}(\zeta)} \phi(\zeta) \tag{4.37}$$

[i.e., (2.20) holds for all ϕ in $\mathcal{F}^G$].

4. If

$$\lim_{m\to\infty} \lambda^{1/G}_m (m \lambda^{1/G}_{m-1})^{-1} = 0, \tag{4.38}$$

then

$$\forall \phi \in \mathcal{F}^G: \phi(\zeta) = \int d\xi\, \overline{\omega^\zeta(\xi)} \phi(\xi) \tag{4.39}$$

[i.e., the reconstruction property (2.21) still holds for $\mathcal{F}^G$].

5. If

$$\exists K'_G, K''_G > 0 \text{ such that } \forall m: K'_G \le \lambda^G_m \lambda^{1/G}_m \le K''_G, \tag{4.40}$$

then $\mathcal{F}^{1/G}$ can be identified with the dual, $(\mathcal{F}^G)'$, of $\mathcal{F}^G$, by means of the map

$$\mathcal{F}^{1/G} \to (\mathcal{F}^G)',$$

$$\psi \mapsto L_\psi \text{ with } L_\psi(\phi) = \int d\zeta\, \overline{\psi(\zeta)} \phi(\zeta) \tag{4.41}$$

$$= \sum_{[k][l]} \overline{\psi_{[k,l]}} \phi_{[k,l]}. \tag{4.42}$$

Proof:

1. Using

$$\phi(\zeta) = (\pi r^2)^{-2n} \omega_2(\zeta) \int_{|\zeta' - \zeta| < r} d\zeta'\, \phi(\zeta') \omega_2(\zeta')^{-1},$$

and (4.33) one can check that

$$\forall R: \exists K_{G,R} \text{ such that } \forall \zeta, |\zeta| < R: |\phi(\zeta)| < K_{G,R} \|\phi\|_G. \tag{4.43}$$

Hence, convergence with respect to $\|\ \|_G$ automatically entails uniform convergence on compact sets. Therefore, any Cauchy sequence in $\mathcal{F}^G$ has a limit in $\mathcal{F}^G$, and $\mathcal{F}^G$ is complete.

2. Proposition A1 in Appendix A proves that $\forall \phi \in Z(E_2)$:

$$\int d\zeta\, |\phi(\zeta)|^2 G(|\zeta|^2) = \sum_{[k][l]} |\phi_{[k,l]}|^2 \lambda^G_{|k|+|l|}, \tag{4.44}$$

where these expressions can be finite or infinite. If $\lambda_m^G < \infty$ for all m, then (4.44) shows that all the $u_{[k,l]} \in \mathcal{F}^G$; (4.36) is proved in Appendix A. If $\lambda_m^G = \infty$, then $\forall m' \geqslant m: \lambda_{m'}^G = \infty$ [use (4.33)]. Hence, $\forall \phi \in \mathcal{F}^G: \phi_{[k,l]} = 0$ if $|k| + |l| \geqslant m$ and $\mathcal{F}^G$ is finite-dimensional.
($\mathcal{F}^G \subset \text{span}\{u_{[k,l]}; |k| + |l| < m\}$).

3. If $\forall m: \lambda_m^{1/G} < \infty$, then

$$\forall [k],[l]: u_{[k,l]} \in \mathcal{F}^{1/G} \subset L^2\left(E \times E; \frac{d\zeta}{G(|\zeta|^2)}\right),$$

(4.37) is then a consequence of Proposition A2 in Appendix A.

4. Equation (4.38) implies
$\forall \zeta: \omega^\zeta \in \mathcal{F}^{1/G} \subset (\mathcal{F}^G)'$

$$\left(\text{use} \sum_{\substack{[k][l] \\ |k|+|l|=m}} |u_{[k,l]}(\zeta)|^2 = e^{-|\zeta|^2} \frac{|\zeta|^{2m}}{m!}\right),$$

apply Proposition A2.

5. See Proposition 4.3 in Ref. 16.

Note that once (4.40) is satisfied, (4.39) holds automatically;

$(4.40) \Rightarrow (\mathcal{F}^G)' \cong \mathcal{F}^{1/G}$.

Because of (4.43), $\exists \tilde{\omega}^\xi \in \mathcal{F}^{1/G}$ such that

$$\forall \phi \in \mathcal{F}^G: \int \overline{\tilde{\omega}^\xi(\zeta)} \phi(\zeta) = \phi(\xi).$$

In particular

$$\int d\zeta\ \overline{\tilde{\omega}^\xi(\zeta)} u_{[k,l]}(\zeta) = u_{[k,l]}(\xi) \Rightarrow \tilde{\omega}^\xi = \omega^\xi. \qquad \blacksquare$$

Examples:

1. Take $G(x) = (1+x)^\rho$. This weight satisfies all the conditions in Proposition 4.5; the corresponding $\mathcal{F}^G$ spaces are of course exactly the $\mathcal{F}^\rho$ of Sec. 4B.

2. Another possibility is $G(x) = e^{\beta x}$, with $|\beta| < 1$. This choice for G satisfies (4.33), (4.35), and (4.38); one has $\lambda_m^G = (1-\beta)^{-m-2n}$, from which one clearly sees that the duality condition (4.40) is not satisfied.

3. $G(x) = e^{\beta\sqrt{x}}$. This corresponds to a simple exponential weight for $\mathcal{F}^G$:

$$\|\phi\|_G^2 = \int d\zeta\, |\phi(\zeta)|^2 e^{\beta|\zeta|}.$$

This choice also satisfies all the conditions in Proposition 4.5 (see below).

4. A rather general class of interesting weight functions is given by taking $G = F_\rho^{q,\tau}$,
with

$$F_\rho^{q,\tau}(x) = (1+x)^\rho e^{\tau x^q} \rho \in \mathbb{R}, \tau \in \mathbb{R} \setminus \{0\}, q \in (0,1). \tag{4.45}$$

For all the values of the parameters indicated above, $F_\rho^{q,\tau}$ satisfies (4.33), (4.35), and (4.38).

A detailed analysis of the asymptotic behavior of $\lambda_m^{F_\rho^{q,\tau}}$ yields (see Appendix B)

$$\forall q \in \left(1 - \frac{1}{n}, 1 - \frac{1}{n+1}\right), \quad n = 1,2,\ldots,$$

$$\lambda_m^{F_\rho^{q,\tau}} \underset{m\to\infty}{\sim} \text{const} \times m^\rho \exp[\tau m^q + A_1 m^{2q-1} + \cdots + A_{n-2} m^{nq-(n-1)}](1 + O(m^\alpha)), \tag{4.46}$$

with

$$\alpha = \max(q-1; q + n(q-1)).$$

For $q \leqslant \frac{1}{2}$ this specializes to

$$\lambda_m^{F_\rho^{q,\tau}} \sim \text{const} \times m^\rho e^{\tau m^q}[1 + O(m^\alpha)], \tag{4.47}$$

which implies the duality condition (4.40) is satisfied for $q \leqslant \frac{1}{2}$; for $q > \frac{1}{2}$ it is not.

The Hilbert spaces W^ϕ.

The $\mathcal{F}^G$ spaces were constructed on the same principle as the $\mathcal{F}^\rho$ spaces, with more general weight functions. We shall likewise generalize the construction of the W^ρ spaces. Let $\phi(m)$ be any sequence of strictly positive real numbers. We define

$$(\|f\|_\phi^s)^2 = \sum_{[r][s]} |(f, H_{[r,s]})|^2 \phi(|r| + |s| + n) = \sum_{[k][l]} |(f, h_{[k,l]})|^2 \phi(|k| + |l| + n). \tag{4.48}$$

The set of all functions f in $\mathcal{S}$ for which $\|f\|_\phi$ is finite we call $\mathcal{S}_\phi$; W^ϕ is then the completion of $\mathcal{S}_\phi$ with respect to $\|\ \|_\phi$.

We can, of course, as in (4.24), consider W^ϕ as the natural domain of $\phi(x^2 + p^2 - \frac{1}{4}\Delta_x - \frac{1}{4}\Delta_p)^{1/2}$, and put $\|f\|_\phi^s = \|\phi(x^2 + p^2 - \frac{1}{4}\Delta_x - \frac{1}{4}\Delta_p)^{1/2} f\|$.

Examples:

1. Taking $\phi_\rho(l) = (l+n)^\rho$, one has $\mathcal{S}_{\phi_\rho} = \mathcal{S}$ (it is only for ϕ increasing faster than polynomials that $\mathcal{S}_\phi$ may become a proper subset of $\mathcal{S}$), and $W^{\phi_\rho} = W^\rho$.

2. The $H(\alpha, A)$, $H(\bar{\alpha}, \bar{A})$ spaces, introduced in Ref. 17, are a special case of a W^ϕ-structure. For the detailed definition we refer to Ref. 17; a survey is given in Appendix C (our definitions are slightly adjusted to deal with the dilation in $x^2 + p^2 - \frac{1}{4}\Delta_x - \frac{1}{4}\Delta_p$ with respect to a normal harmonic oscillator). Essentially the $H(\alpha, A)$ form a scale of spaces of test functions "of type S" and their duals $H(\bar{\alpha}, \bar{A})$ a scale of Hilbert spaces of distributions or generalized functions of type S.[18] They are defined (see Appendix C) as W^ϕ spaces with:
for $H(\alpha, A)$:

$$\phi(k+n) = \gamma_k^{-2}(\alpha, A) = \sum_{m=0}^{\infty} \frac{a(m;k)}{A^{2m} \Gamma^2(\alpha m)}, \tag{4.49}$$

for $H(\bar{\alpha}, \bar{A})$:

$$\phi(k+n) = \gamma_k^2(\alpha, A),$$

where the $a(m;k)$ are numbers satisfying

$$2^{-m} \frac{\Gamma(k+2n+m)}{\Gamma(k+2n)} \leqslant a(m;k) \leqslant \frac{\Gamma(k+n+m)}{\Gamma(k+n)} \tag{4.50}$$

[for the exact definition of $a(m;k)$, see Appendix C]. For all (α, A) with $\alpha < \frac{1}{2}$, A arbitrary, or $\alpha = \frac{1}{2}$, $A > \sqrt{2}$, $H(\alpha, A)$ is an infinitely dimensional Hilbert space, with orthonormal basis $\gamma_{|k|+|l|}(\alpha, A) h_{[k,l]}$; $H(\bar{\alpha}, \bar{A})$ is its dual: for any $f \in H(\alpha, A)$, the action of $T \in H(\bar{\alpha}, \bar{A})$ on f is simply the natural extension of the action of elements of $\mathcal{S}'$ on $H(\alpha, A)$:

$$T(f) = \sum_{[k][l]} T(h_{[k,l]})(h_{[k,l]}, f). \tag{4.51}$$

In Ref. 17 it was shown that $\forall(\alpha, A)$ satisfying the restriction above, $\exists C(\alpha, A)$ such that $\forall f \in H(\alpha, A)$:

$$|f(x,p)| \leq C(\alpha, A)\|f\|_{\alpha,A} \prod_{j=1}^{n} (x_j p_j)^{n+2} \times \exp\left[-\frac{1}{2n(2A)^{1/\alpha}} \sum_{j=1}^{n} (x_j^{1/\alpha} + p_j^{1/\alpha}) \right]. \quad (4.52)$$

In the extreme case $\alpha = \frac{1}{2}$, $A > \sqrt{2}$ this becomes

$$|f(x,p)| \leq C(\tfrac{1}{2}, A)\|f\|_{1/2,A} \prod_{j=1}^{n} (x_j p_j)^{n+2} \exp\left[-\frac{x^2 + p^2}{8nA^2} \right], \quad (4.53)$$

i.e., f has a Gaussian-like behavior at infinity.

Another property proved in Ref. 17 is the following.

$\forall \alpha \in (\frac{1}{2}, 1), \forall f \in H(\alpha, A)$:

f is the restriction to the reals of an entire analytic function of order $\rho < (1-\alpha)^{-1}$.

The integral transform I as a map from W^{ϕ} to $\mathcal{F}^{G}$ and vice versa

Looking back at the arguments leading to the formulation of Theorem 4.4, we see that the estimates (4.22) played a crucial role in the proof of the bijectivity of I between W^{ρ} and $\mathcal{F}^{\rho}$. In the case of a general W^{ϕ}–$\mathcal{F}^{G}$ pair, we shall use again such estimates.

Theorem 4.5: Let $\mathcal{F}^{G}$, W^{ϕ} be two Hilbert spaces as defined above [with G satisfying (4.33), (4.35)]:

(1) If $\exists K_1 > 0$ such that $\forall m \in \mathbf{N}: K_1 \phi(m+n) \geq \lambda_m^{G}$, (4.54)

then I can be considered as a bounded linear map from W^{ϕ} to $\mathcal{F}^{G}$, with

$$\forall T \in W^{\phi}: IT(\zeta) = \sum_{[k][l]} \overline{T(h_{[l,k]})}\, u_{[k,l]}(\zeta), \quad (4.55)$$

where the series converges uniformly on compact sets. Moreover,

$$\forall T \in W^{\phi}: \|IT\|_G \leq K_1^{1/2}\|T\|_{\phi}. \quad (4.56)$$

(2) *If* $\exists K_2 > 0$ such that $\forall m \in \mathbf{N}: K_2 \phi(m+n) \leq \lambda_m^{G}$, (4.57)

then $\tilde{I}$ can be extended to a bounded linear map from $\mathcal{F}^{G}$ to W^{ϕ} with

$$\forall \Phi = \sum_{[k][l]} \Phi_{[k,l]} u_{[k,l]} \in \mathcal{F}^{G}: \tilde{I}\Phi = W^{\phi} - \lim_{m \to \infty} \sum_{\substack{[k][l] \\ |k|+|l|<m}} \Phi_{[k,l]} h_{[l,k]}. \quad (4.58)$$

One has

$$\|\tilde{I}\Phi\|_{\phi} \leq K_2^{-1/2}\|\Phi\|_G.$$

(3) If $\exists K_1, K_2 > 0$ such that

$$\forall m: K_1 \phi(m+n) \leq \lambda_m^{G} \leq K_2 \phi(m+n), \quad (4.59)$$

then I as defined by 4.55) is an isomorphism from W^{ϕ} onto $\mathcal{F}^{G}$, with inverse $\tilde{I}$ [as defined by (4.58)]

Proof: $I, \tilde{I}$ are already defined on the finite linear combinations of the $h_{[k,l]}, u_{[k,l]}$, respectively. The bounds (4.54), (4.57) ensure that $I, \tilde{I}$ can be extended as indicated. Formula (4.55) is a consequence of the fact that $\|\ \|_G$ convergence implies uniform and absolute convergence on compact sets. ■

Note: Of course, one can always define $\mathcal{F}^{G}$ first, and then take $\phi(m+n) = \lambda_m^{G}$; for this particular W^{ϕ} space, the theorem becomes trivial. A W^{ϕ} space defined in this way would, however, be rather useless because of its too intrinsic definition: we are more interested in the situation where $\mathcal{F}^{G}$ and W^{ϕ} are defined separately, but where nevertheless a link can be established via I or $\tilde{I}$. This was the case for the $\mathcal{F}^{\rho}$ and the W^{ρ} spaces: the W^{ρ} spaces made sense as ladder spaces in the N-representation of $\mathcal{S}'(E)$, and the $\mathcal{F}^{\rho}$ spaces were polynomially weighted L^2 spaces of modified holomorphic functions. Results such as Theorem 4.4 (or Theorem 4.5 with explicit $\mathcal{F}^{G}$–W^{ϕ} pairs) can then be used to characterize the behavior of the coherent state matrix elements of an operator by means of the properties of the corresponding classical function (or distribution) or vice versa.

An example of corresponding W^{ϕ}–$\mathcal{F}^{G}$ pairs different from the W^{ρ}–$\mathcal{F}^{\rho}$ pairs in Sec. 4B is given in the following subsection.

The action of the integral transform I on Hilbert spaces of distributions of type S

We shall study in this subsection the action of I on the Hilbert spaces $H(\alpha, A)$, $H(\bar{\alpha}, \bar{A})$ defined above. In order to be able to apply Theorem 4.5, we have to find suitable weight functions ${}_jG_{\alpha,A}, {}_j\bar{G}_{\alpha,A}$ such that

$$K_2' \lambda_m^{{}_2G_{\alpha,A}} \leq \gamma_m^{-2}(\alpha, A) \leq K_1' \lambda_m^{{}_1G_{\alpha,A}},$$
$$\bar{K}_2' \lambda_m^{{}_2\bar{G}_{\alpha,A}} \leq \gamma_m^{2}(\alpha, A) \leq \bar{K}_1' \lambda_m^{{}_1\bar{G}_{\alpha,A}} \quad (4.60)$$

for some $K_1', K_2', \bar{K}_1', \bar{K}_2' > 0$.

Using the bounds (4.50) we can easily construct the ${}_jG_{\alpha,A}$ functions. Indeed we have

$$\sum_{l=0}^{\infty} \frac{2^{-l}}{A^{2l}\Gamma^2(\alpha l)} \frac{\Gamma(m+2n+l)}{\Gamma(m+2n)} \leq \gamma_m^{-2}(\alpha, A) \leq \sum_{l=0}^{\infty} \frac{1}{A^{2l}\Gamma^2(\alpha l)} \frac{\Gamma(m+n+l)}{\Gamma(m+n)}.$$

Since it is clear from (4.34) that for

$$\hat{G}_{\beta,B}(x) = \sum_{j=0}^{\infty} \frac{x^j}{\Gamma^2(\beta_j)B^{2j}},$$

the corresponding λ_m^{G} are given by

$$\lambda_m^{\hat{G}_{\beta,B}} = \sum_{j=0}^{\infty} \frac{1}{B^{2j}\Gamma^2(2j)} \frac{\Gamma(m+2n+j)}{\Gamma(m+2n)},$$

we immediately have

$$\lambda_m^{\hat{G}_{\alpha,\sqrt{2}A}} \leq \gamma_m^{-2}(\alpha, A) \leq \lambda_m^{\hat{G}_{\alpha,A}}. \quad (4.61)$$

To find candidates for the functions ${}_j\bar{G}_{\alpha,A}$, we have to do a little more work. We shall study the asymptotic behavior of the $\lambda_m^{G_{\beta,B}}$, then invert (4.61) and try to find suitable ${}_j\bar{G}_{\alpha,A}$.

$\hat{G}_{\beta,B}$ as defined above is typically an entire function of finite order. Computing its order and type we find[19]

$$\rho(\beta, B) = \overline{\lim_{n \to \infty}} \frac{m \ln n}{\ln(\Gamma^2(\beta n)B^{2n})} = \frac{1}{2\beta},$$

$$\tau(\beta, B) = \frac{1}{2\rho(\beta, B)} \overline{\lim_{n \to \infty}}\, n\, [\Gamma^2(\beta n)B^{2n}]^{-1/2\beta n} = 2B^{-1/\beta}.$$

So $\hat{G}_{\beta,B}$ is an entire function of growth $\leqslant (1/2\beta, 2B^{-1/\beta})$. Since the positive real axis is the direction of fastest growth for $\hat{G}_{\beta,B}$, this implies that

$$\forall \tau' < \tau(\beta,B),\ \forall \tau'' > \tau(\beta,B),\ \exists K',K'' > 0$$

such that

$$\forall x \in \mathbb{R}^+ : K' F^{(2\beta)^{-1},\tau'}(x) \leqslant \hat{G}_{\beta,B}(x) \leqslant K'' F^{(2\beta)^{-1},\tau''}(x)$$

[we define $F^{q,\tau}(x) = F_0^{q,\tau}(x) = \exp(\tau x^q)$]

and hence

$$K' \lambda_m^{F^{(2\beta)^{-1},\tau'}} \leqslant \lambda_m^{\hat{G}_{\beta,B}} \leqslant K'' \lambda_m^{F^{(2\beta)^{-1},\tau''}}.$$

Using now the estimates (4.46), inverting (4.61), these inequalities imply that

$$\forall \alpha > \tfrac{1}{2}; \forall \tau_2 > \tau(\alpha,A) = 2A^{-1/2}, \forall \tau_1 < \tau(\alpha,\sqrt{2}A) = 2(\sqrt{2}A)^{-1/2}: \exists K_1,K_2 > 0$$

such that

$$K_2 \lambda_m^{F^{(2\alpha)^{-1},-\tau_2}} \leqslant \gamma_m^2(\alpha,A) \leqslant K_1 \lambda_m^{F^{(2\alpha)^{-1},-\tau_1}}. \tag{4.62}$$

Since this inequality has exactly the right form of (4.60), we are now in a position to apply Theorem 4.5; we get the following results.

Theorem 4.6: For any $q \in (0,1), \tau \in \mathbb{R}$, we define $F^{q,\tau}(x) = e^{\tau x^q} (x \in \mathbb{R}_+)$; for any (β,B) with $\beta > \frac{1}{2}$ or $\beta = \frac{1}{2}$, $B > \sqrt{2}$, we define

$$\hat{G}_{\beta,B}(x) = \sum_{n=0}^{\infty} \frac{x^n}{\Gamma^2(\beta n) B^{2n}}.$$

Take any $[\alpha, A)$ with $\alpha > \frac{1}{2}$ or $\alpha = \frac{1}{2}, A > \sqrt{2}$. Then

(1) The integral transform I defines a continuous linear map from the Hilbert space $H(\alpha, A)$ of functions of type S to the weighted L^2 space of holomorphic functions $\mathcal{F}^{\hat{G}_{\alpha,\sqrt{2}A}}$: we have

$$\forall f \in H(\alpha, A): If(\zeta) = \int dv\, \{\zeta | v\} f(v)$$

and

$$\|If\|^2_{\hat{G}_{\alpha,\sqrt{2}A}} = \int d\zeta\, |If(\zeta)|^2 \hat{G}_{\alpha,\sqrt{2}A}(|\zeta|^2) \leqslant \|f\|^2_{\alpha,A}. \tag{4.63}$$

(2) The integral transform $\tilde{I}$ defines a continuous linear map from $\mathcal{F}^{\hat{G}_{\alpha,A}}$ to

$$H(\alpha, A): \forall \phi \in \mathcal{F}^{\hat{G}_{\alpha,A}}: (\tilde{I}\phi)(v) = \int d\zeta\, \{\bar{\zeta} | v\} \phi(\zeta)$$

and

$$\|\tilde{I}\phi\|^2_{\alpha,A} \leqslant \int d\zeta\, |\phi(\zeta)|^2 \hat{G}_{\alpha,A}(|\zeta|^2) = \|\phi\|^2_{\hat{G}_{\alpha,A}}. \tag{4.64}$$

For the next two results, we restrict ourselves to the case $\alpha > \frac{1}{2}$.

(3) $\forall \tau_2 > \tau(\alpha, A) = 2A^{-1/2}$, the integral transform I extends to a continuous linear map from the Hilbert space $H(\bar{\alpha}, \bar{A})$ of distributions of type S to the weighted L^2 space of holomorphic functions $\mathcal{F}^{F^{(2\alpha)^{-1},-\tau_2}}$:

$$\forall T \in H(\bar{\alpha}, \bar{A}): IT(\zeta) = \sum_{[k][l]} \overline{T(h_{[l,k]})} u_{[k,l]}(\zeta) = \overline{T(\{\bar{\xi} | \cdot\})} \tag{4.65}$$

and

$$\exists K \text{ such that } \|IT\|^2_{F^{(2\alpha)^{-1},-\tau_2}} = \int d\zeta\, |IT(\zeta)|^2 e^{-\tau_2 |\zeta|^{1/2}} \leqslant K \|T\|^2_{\bar{\alpha},\bar{A}}.$$

(4) $\forall \tau_1 < \tau(\alpha,\sqrt{2}A) = 2(\sqrt{2}A)^{-1/\alpha}$, $\tilde{I}$ extends to a continuous linear map from $\mathcal{F}^{F^{(2\alpha)^{-1},-\tau_1}}$ to $H(\bar{\alpha},\bar{A})$:

$$\forall \phi \in \mathcal{F}^{F^{(2\alpha)^{-1},-\tau_1}}: \tilde{I}\phi = H(\bar{\alpha},\bar{A}) - \lim_{R \to \infty} \int_{|\zeta| < R} d\zeta\, \{\bar{\xi} \cdot\} \phi(\zeta) \tag{4.66}$$

and

$$\exists K' > 0 \text{ such that } K' \|\tilde{I}\phi\|^2_{\bar{\alpha},\bar{A}} \leqslant \int d\zeta\, |\phi(\zeta)|^2 e^{-\tau_1 |\zeta|^{1/2}} = \|\phi\|^2_{F^{(2\alpha)^{-1},-\tau_1}}.$$

For $\alpha = \frac{1}{2}, A > \sqrt{2}$ we have

(5) $\forall y > (A^2/2 - 1)^{-1}$, the integral transform I extends to a continuous linear map from $H(\bar{\frac{1}{2}}, \bar{A})$ to the weighted L^2 space $\mathcal{F}^{G_{-y}}$, where

$$G_a(x) = e^{ax} \quad (a < 1).$$

We have $\forall T \in H(\bar{\frac{1}{2}}, \bar{A}): IT(\zeta) = \overline{T(\{\bar{\xi} | \cdot\})}$ and

$$\|IT\|^2_{G_{-y}} = \int d\zeta\, |IT(\zeta)|^2 e^{-y|\zeta|^2} \leqslant (1+y)^{-n+1} K_{\alpha,A}(y^{-1})^{-1} \|T\|^2_{\bar{1/2},\bar{A}},$$

with

$$K_{\alpha,A}(z) = \sum_{m=0}^{\infty} \frac{m!}{\Gamma^2(\alpha m) A^{2m}} (z+2)^m. \tag{4.67}$$

Proof: (1)–(4) were essentially proven above. Since $\forall (\alpha, A): \{\zeta | \cdot\} \in H(\alpha, A)$, we can always write $IT(\zeta)$ as $\overline{T(\{\bar{\xi} | \cdot\})}$. For (5) we use the estimate $\gamma_m^{-2}(\frac{1}{2}, A) \leqslant (1+y)^{m+n-1} K_{\alpha,A}(y^{-1})$ (see Appendix C). Since $\lambda_m^{G_a} = (1-a)^{-(m+2n)}$, (4.67) follows. ■

Remarks: As we already mentioned previously, our motivation for this detailed study of the integral transforms $I, \tilde{I}$ is their relation with the Weyl quantization procedure [see (1.1) and (1.6)]. Possible applications of Theorem 4.6 in this quantization context are, e.g., the following.

1. In Ref. 17 it was shown that for $\alpha > 1$, the functions in $H(\alpha, A)$ are the restrictions to the real line of entire functions of order $(1-\alpha)^{-1}$. On these $H(\alpha, A)$ one can therefore define the complex δ functions $\delta_{v+iw} :\rightarrow (v + iw)$ as continuous linear forms, i.e., as elements of $H(\bar{\alpha},\bar{A})$ (see Ref. 17). By means of the integral transform I, and applying Theorem 4.6, one can therefore quantize these δ functions with complex argument. The same can be done for the real exponentials e^{ax+bp}; the quantal operators corresponding to both these functions are actually complex translation operators, and can therefore be useful in the study of certain resonance problems. Complex dilations also can be obtained as quantizations of $H(\bar{\alpha}_{\frac{1}{2}}, \bar{A})$-objects (at least for the dilation parameter θ in some strip of the complex plane).

2. Using the $\tilde{I}$ results, the statements in Theorem 4.6 enable one to dequantize certain families of operators with coherent state matrix elements with fast growth (up to Gaus-

sian-like growth) in the coherent state labels, and to derive properties of the corresponding classical functions.

5. THE INTEGRAL TRANSFORM I ACTING ON FUNCTIONS FACTORIZING INTO A PRODUCT OF A FUNCTION DEPENDING ON x WITH A FUNCTION DEPENDING ON p

Whereas for "dequantization procedures" it may be useful to know how to treat an operator in which the x and p parts cannot be disentangled, for quantization purposes one is mostly interested in functions depending only on x or on p or linear combinations of such functions. We shall therefore indicate here how the additional information that a given function is factorizable, $f(x,p) = f_1(x) f_2(p)$ or, depending only on x or on p, $f(x,p) = f_1(x)$, $f(x,p) = f_2(p)$, can be used to sharpen the results derived in the preceding section. To achieve this, we shall use the decomposition (4.3) of the integral kernel $\{a,b\,|v\}$:

$$\{a,b\,|v\} = K_B(c_{ab};x_v)K_B(d_{ab};p_v), \tag{5.1}$$

where

$$K_B(c;y) = K_B((c_1,c_2);y) = 2^{n/2}\pi^{n/4}A\,(c_1,c_2;\sqrt{2}y)e^{-(1/4)(c_1^2+c_2^2)} \tag{5.2}$$

($y\in\mathbb{R}^n; c\in\mathbb{R}^{2n} = \mathbb{R}^n \oplus \mathbb{R}^n$), with A given by (4.1), and where

$$c_{ab} = \frac{1}{\sqrt{2}}(x_a + x_b, p_b - p_a),$$

$$d_{ab} = \frac{1}{\sqrt{2}}(p_a + p_b, x_a - x_b). \tag{5.3}$$

One immediately sees from (5.1) that the integral transform I, when applied to a factorizable function $f(x,p) = f_1(x)f_2(p)$, splits into two pieces:

$$If(a,b) = I_B f_1(c_{ab}) I_B f_2(d_{ab}), \tag{5.4}$$

with $\forall g$ function on $\mathbb{R}^n$, $\forall d = (d_1, d_2)\in\mathbb{R}^{2n}$:

$$(I_B\, g)(d_1, d_2) = (2\pi)^{-n/2}\int_{\mathbb{R}^n} d^n y K_B(d;y)\, g(y). \tag{5.5}$$

It is not difficult to check that the integral transform I_B with kernel K_B has exactly the same properties as the integral transform I, except that all the dimensions have to be halved. Since the exact value of the dimension n plays no role whatever in the results derived up until now, we see that all the results for I hold also for I_B, provided we replace each n by $n/2$.

We give below a list of bounds on $I(f_1,f_2)$ which can be obtained in this way. For all the cases where the images $I_B f_1$, $I_B f_2$ cannot be defined directly (i.e., $f_1, f_2 \notin L^\infty + L^2$), we define I_B as a continuous extension of the integral transform with kernel K_B (just as we did for I).

In the case where $f_2 = 1$, i.e., where the function f depends only on x (the case $f_1 = 1$ is completely similar), one of the factors in (5.4) can be calculated explicitly:

$$I(f_1\cdot 1)(a,b) = I_B f_1(c_{ab}) \times \exp\left[-\frac{(x_a - x_b)^2}{4} + \frac{i(p_a + p_b)\cdot(x_a - x_b)}{4}\right]. \tag{5.6}$$

We also give some bounds for this special case.

Examples:

1. Define $\mathcal{J}^k = \{\, g: \mathbb{R}^n \to \mathbb{C};\ g \text{ is } C^k \text{ and}$

$$|g|_k^s = \max_{|m|<k} \sup |2^{-|m|/2}(1 + 2|y|^2)^{(k-|m|)/2} \nabla^{[m]} g(y)| < \infty\,\}.$$

Take $f_1\in\mathcal{J}^{k_1}$, $f_2\in\mathcal{J}^{k_2}$. Then

$$|I(f_1,f_2)(a,b)| \leqslant \tilde{b}_{k_1}\tilde{b}_{k_2}|f_1|_{k_1}^s |f_2|_{k_2}^s \times\left(1 + \frac{|x_a + x_b|^2 + |p_a - p_b|^2}{2}\right)^{-k_1/2} \times\left(1 + \frac{|x_a - x_b|^2 + |p_a + p_b|^2}{2}\right)^{-k_1/2},$$

with

$$\tilde{b}_k = \frac{3e}{2}2^{n/4}(8n)^{k/2}\begin{cases}1 & \text{if } k\leqslant 2,\\ e^{-k}k^k & \text{if } k\geqslant 3.\end{cases}$$

2. Define $\tilde{W}^\rho$ as the closure of $\mathcal{S}(\mathbb{R}^n)$ with respect to $\|\ \|_\rho^s$, with $(\|g\|_\rho^s)^2 = (g,(\tilde{H} + n/2)^\rho g)$ and $\tilde{H} = y^2 - \frac{1}{4}\Delta_y$. Take $f_1\in\tilde{W}^{\rho_1}$, $f_2\in\tilde{W}^{\rho_2}$. Then

$$\int\int da\, db\, |I(f_1,f_2)(a,b)|^2 \times\left(1 + \frac{|x_a + x_b|^2 + |p_a - p_b|^2}{2}\right)^{\rho_1} \times\left(1 + \frac{|x_a - x_b|^2 + |p_a + p_b|^2}{2}\right)^{\rho_2} \leqslant \tilde{c}''_{\rho_1}\cdot\tilde{c}''_{\rho_2}(\|f_1\|_{\rho_1}^s\cdot\|f_2\|_{\rho_2}^s)^2,$$

with

$$c''_\rho = \begin{cases} e^{1-\rho}\left(1 + \dfrac{\rho}{n}\right)^{\rho+n}, & \rho > 0\\ e^{-\rho}, & \rho\leqslant 0.\end{cases}$$

3. Take $f_1\in\mathcal{S}^{-k}$. Then

$$|I(f_1\cdot 1)(a,b)| \leqslant \tilde{b}_k e^{-(1/4)|x_a - x_b|^2} \times |f_1|_k^s\left(1 + \frac{|x_a + x_b|^2 + |p_a - p_b|^2}{2}\right)^{-k/2}.$$

4. Take $T_1\in\mathcal{S}'(\mathbb{R}^n)$, with $\forall g\in\mathcal{S}(\mathbb{R})^n$: $|T_1(g)| \leqslant K_{T_1}|g|_k^s$. Then $\forall \mu > k + n$:

$$|I(T_1\cdot 1)(a,b)| \leqslant 2c_\mu \tilde{b}'_{k,\mu} K_{T_1} e^{-(1/4)|x_a - x_b|^2} \times\left(1 + \frac{|x_a + x_b|^2 + |p_a - p_b|^2}{2}\right)^{\mu/2},$$

with

$$c_\mu = \mu^{\mu/2}\exp[-\tfrac{1}{2}(\mu - 1)],$$

$$\tilde{b}'_{k,\mu} = \pi^{-n/2}\left(\frac{8}{3}\right)^{k/2}\Gamma\left(\frac{k+n}{2} + 1\right)\Gamma\left(\frac{\mu - k - n}{2}\right)\Gamma\left(\frac{\mu - k}{2}\right)^{-1} \times\int_{\mathbb{R}^n} d^n y\, e^{-2y^2}(1 + y^2)^k.$$

5. Take $f_1\in\tilde{W}^\rho$; let g be any function in $L^2(\mathbb{R}^n)$. Then

$$\int\int da\,db\,|I(f_1\cdot 1)(a,b)|^2\left|g\left(\frac{p_a+p_b}{\sqrt{2}}\right)\right|^2$$
$$\times\left(1+\frac{|x_a+x_b|^2+|p_a-p_b|^2}{2}\right)^{\rho}<\tilde{c}_\rho''\,\|g\|_2^2(\|f_1\|_\rho^s)\,.$$

By taking other combinations for f_1, f_2 one can easily derive other bounds of this kind.

6. CONCLUDING REMARKS

Using intensively the analyticity properties of the function $\{a,b\,|v\}$ we have derived several families of bounds on the action of I and $\tilde{I}$. These results will be used in Ref. 15 to derive properties of the Weyl quantization procedure. Examples of such results are

$\left.\begin{array}{l}\forall f\in W^{\nu+\epsilon}\colon Qf \text{ is trace-class}\\ \forall A\in\mathscr{B}(\mathscr{H})\colon Q^{-1}A\in W^{-\nu-\epsilon}\end{array}\right\}$ see also Ref. 5,

$\forall f_1,f_2\in\tilde{W}^{\nu+\epsilon}\colon Qf_1Qf_2$ trace-class,

$f_1,f_2\in L^2(\mathbf{R}^n)\Rightarrow Qf_1Qf_2$ Hilbert–Schmidt,

$\forall T\in W^{\mu}\colon QT$ is a quadratic form, relatively form-bounded with respect to a power of the harmonic oscillator Hamiltonian QH.

$\forall f\in W^{\rho},g\in W^{-\rho}$: the twisted product $f\circ g$ is defined, and $\in W^{-2\rho}$.

The bounds derived here can also be used to show that all the operations in, e.g., Ref. 11 were well defined.

Because of the link of our integral transform I with the Bargmann integral transform in Ref. 9 any result on the Bargmann integral transform (such as, e.g., in Ref. 1) can be translated to give properties of I, and hence of the Weyl quantization procedure. Note also that analogous bounds can be obtained if one starts not with the coherent state family $\{\Omega^a\}$, but with any other overcomplete family depending analytically on its label, and having the reproducing property (1.4); an example of such a family would be given by[20] $\{\Omega^a_{[m]}\}$, where $\Omega^a_{[m]}=W(a)u_{[m]}$ [$W(a)$ are the Weyl operators; to obtain the usual coherent states one takes $[m]=[0]$]. This would give rise to another integral kernel $\{a,b\,|v\}_{[m]}=2^n(\Omega^a_{[m]},\Pi(v)\Omega^b_{[m]})$ but essentially the same theorems could be derived (with some adjustments). Finally, it is important to note that the integral transform I has the following invariance property with respect to the symplectic Fourier transform (see also Ref. 1).

Define

$$(F_\alpha f)(v)=2^{-n}|\alpha|^n\int dv'\,e^{i\alpha\sigma(v,v')}f(v')\,,$$

then

$$F_4(\{a,b\,|\cdot\})(v)=\{a,-b\,|v\}$$

and hence

$$\forall T\colon I(F_{-4}T)(a,b)=IT(a,-b)$$

and this for T in any of the classes considered above (all the spaces we have introduced are invariant under the Fourier transform). This leads to the property $Q(F_{-4}T)=QT\cdot\Pi$ (Π is the parity operator) for the Weyl quantization procedure, but it also implies that the same Fourier invariance will turn up in all results, thereby weakening some of them (e.g., the result of the trace-class properties of Qf_1Qf_2: see Ref. 15).

APPENDIX A

We prove some results on $\mathscr{F}^G$ spaces.

Proposition A.1: Let G be a function $\mathbf{R}^+\to\mathbf{R}^+$, such that

$$\forall m\colon\ \lambda^G_m=\frac{\int_0^\infty dx\,x^{m+2n-1}e^{-x}G(x)}{\Gamma(m+2n)}$$

is finite. Then

$$\forall\phi=\sum_{[k][l]}\phi_{[k,l]}u_{[k,l]}\in Z_2\,,$$

$$\int d\zeta\,|\phi(\zeta)|^2G(|\zeta|^2)<\infty\Leftrightarrow\sum_{[k][l]}|\phi_{[k,l]}|^2\lambda^G_{|k|+|l|}<\infty\,.$$

If one of these expressions is finite, they are equal.

Proof: From

$$\int d\zeta\,\overline{u_{[k,l]}(\zeta)}u_{[k',l']}(\zeta)=\delta_{[k][k']}\delta_{[l][l']}$$

one sees that (see Ref. 9b)

$$\int_{|\zeta|=1}d\zeta\,\overline{u_{[k,l]}(\zeta)}u_{[k',l']}(\zeta)$$
$$=\frac{2}{\Gamma(|k|+|l|+2n)}\delta_{[k][k']}\delta_{[l][l']}\,.$$

Hence

$$\int_{|\zeta|<R}d\zeta\,G(|\zeta|^2)\,\overline{u_{[k,l]}(\zeta)}u_{[k',l']}(\zeta)$$
$$=\delta_{[k][k']}\delta_{[l][l']}A_G(|k|+|l|;R)\,,$$

with

$$A_G(m;R)=\frac{1}{\Gamma(m+2n)}\int_0^{R^2}dy\,e^{-y}y^{m+2n-1}G(y)\underset{R\to\infty}{\nearrow}\lambda^G_m$$

Then

$$\int d\zeta\,|\phi(\zeta)|^2G(|\zeta|^2)$$
$$=\lim_{R\to\infty}\int_{|\zeta|<R}d\zeta\,|\phi(\zeta)|^2G(|\zeta|^2)$$
$$=\lim_{R\to\infty}\sum_{\substack{[k][l]\\[k'][l']}}\overline{\phi_{[k,l]}}\phi_{[k',l']}\delta_{[k][k']}\delta_{[l][l']}$$
$$\times A_G(|k|+|l|;R)$$
$$=\lim_{R\to\infty}\sum_{[k][l]}|\phi_{[k,l]}|^2A_G(|k|+|l|;R)$$
$$=\sum_{[k][l]}|\phi_{[k,l]}|^2\lambda^G_{|k|+|l|}\,.$$

Proposition A.2: Let $\Sigma_{[k][l]}\phi_{[k,l]}u_{[k,l]}$ be an element of $Z(E_2)$.

1. If $\int d\zeta\,|u_{[k,l]}(\zeta)\phi(\zeta)|<\infty$, then

$$\int d\zeta\,\overline{u_{[k,l]}(\zeta)}\phi(\zeta)=\phi_{[k,l]}\,.$$

2. If $\int d\zeta\,|\omega^\xi(\zeta)\phi(\zeta)|<\infty$, then

$$\int d\zeta\,\overline{\omega^\xi(\zeta)}\delta(\zeta)=\phi(\xi)\,.$$

Proof:

1. We have

$$\int d\zeta\ \overline{u_{[k,l]}(\zeta)}\,\phi(\zeta)$$

$$= \lim_{R\to\infty} \int_{|\zeta|<R} d\zeta\ \overline{u_{[k,l]}(\zeta)}\,\phi(\zeta)$$

$$= \lim_{R\to\infty} \sum_{[k'][l']} \phi_{[k',l']}\delta_{[k][k']}\delta_{[l][l']}\,A_1(|k|+|l|;R)$$

$$= \lim_{R\to\infty} \phi_{[k,l]}\,A_1(|k|+|l|;R) = \phi_{[k,l]}\,.$$

2. Analogously,

$$\int d\zeta\ \overline{\omega^{\xi}(\zeta)}\phi(\zeta) = \lim_{R\to\infty} \int_{|\zeta|<R} d\zeta\ \overline{\omega^{\xi}(\zeta)}\phi(\zeta)$$

$$= \lim_{R\to\infty} \sum_{\substack{[k][l]\\[k'][l']}} \phi_{[k,l]}u_{[k',l']}(\xi)\delta_{[k][k']}\delta_{[l][l']} \times A_1(|k|+|l|;R)$$

$$= \lim_{R\to\infty} \sum_{[k][l]} \phi_{[k,l]}u_{[k,l]}(\xi)\,A_1(|k|+|l|;R)$$

$$= \phi(\xi)\,.$$

APPENDIX B

We compute the asymptotic behavior of

$$\lambda_m^{F_\rho^{q,\tau}} = \frac{1}{\Gamma(m+2n)} \int_0^\infty dx\, x^{m+2n-1}e^{-x}(1+x)^\rho e^{\tau x^q} \tag{B1}$$

for $m\to\infty$. To estimate the asymptotic behavior in m of the integral

$$I_{\tau,\rho,q,m} = \int_0^\infty dx \exp[-x+\tau x^q + (m+2n-1)\ln x + \rho\ln(x+1)]\,, \tag{B2}$$

we shall use a stationary point method. The exponent

$$X(x) = -x+\tau x^q + (m+2n-1)\ln x + \rho\ln(x+1)$$

has a unique maximum in $x_0 = m(1+\tau q m^{q-1} + O(m^{q-1}))$. One can use this to estimate that

$$I_{\tau,\rho,q,m} = \tilde{I}_{\tau,\rho,q,m}(1+O(m^{-1}))\,, \tag{B3}$$

where

$$\tilde{I}_{\tau,\rho,q,m} = \int_0^\infty dx \exp[-x+\tau x^q + (m+2n+p-1)\ln x]\,. \tag{B4}$$

We shall therefore restrict ourselves to this last integral. The exponent in (B4),

$$Y(x) = -x+\tau x^q + (m+2n+\rho-1)\ln x\,, \tag{B5}$$

has a unique maximum defined by the equation

$$x = -(m+2n+\rho-1) = \tau q x^q\,. \tag{B6}$$

The solution to this equation can be computed using perturbation techniques:

$$\tilde{x} = m(1+\tau q m^{q-1})(1+O(m^{\max(-1,2q-2)}))\,. \tag{B7}$$

One has then

$$Y''(\tilde{x}) = \frac{1}{m}(1+O(m^{\max(-1,2q-2)}))\,, \tag{B8}$$

$$\forall j>2:\ Y^{(j)}(\tilde{x}) = O(m^{1-j})\,, \tag{B9}$$

$$Y(\tilde{x}) = (-m+(m+2n-\tfrac{1}{2})\ln m) + (\rho-\tfrac{1}{2})\ln m + m\tau m^{q-1} + O(m^{\max(0,2q-1)})\,,$$

hence

$$-\ln\Gamma(m+2n) + Y(\tilde{x}) = (\rho-\tfrac{1}{2})\ln m + \tau m^q + O(m^{\max(0,2q-1)})\,. \tag{B10}$$

Collecting all these results, we see now that

$$\frac{1}{\Gamma(m+2n)} I_{\tau,\rho,q,m} \underset{m\to\infty}{\sim} \frac{1}{\sqrt{m^{-1}}} \times\exp[(\rho-\tfrac{1}{2})\ln m + \tau m^q + O(m^{\max(0,2q-1)})] \times(1+O(m^{\max(-1/2,2q-2)})) \tag{B11}$$

[the higher derivatives contribute only a factor $(1+O(m^{-q/2}))$ because of the estimate (B9)—see Ref. 21].

If $q\leqslant\frac{1}{2}$, we can rewrite (B11) as [being a little more careful in estimating $\tilde{x}$ in (B7)]

$$\lambda_m^{F_\rho^{q,\tau}} \underset{m\to\infty}{\sim} \text{const}\, x m^\rho e^{\tau m^q}(1+O(m^{\max(-1/2,2q-1)}))\,. \tag{B12}$$

For $q>\frac{1}{2}$, the estimate (B7) is too coarse. The next term in the perturbation gives

$$\tilde{x} = m(1+\tau q m^{q-1} + \tau^2 q^3 m^{2q-2})(1+O(m^{\max(-1,3q-3)}))$$

yielding, for $\frac{1}{2}<q\leqslant\frac{2}{3}$,

$$\lambda_m^{F_\rho^{q,\tau}} \underset{m\to\infty}{\sim} \text{const}\, x m^\rho e^{\tau m^q + (1/2)\tau^2 q^2 m^{2q-1}}(1+O(m^{\max(q-1,3q-2)}))\,. \tag{B13}$$

It is easy to see that for $q\in[1+1/n, 1-1/(n+1)]$, n extra terms have to be introduced in the perturbation series for $\tilde{x}$, and that finally

$$1-\frac{1}{n}<q<1-\frac{1}{n+1}\quad (n\geqslant 2)$$

$$\Rightarrow \lambda_m^{F_\rho^{q,\tau}} \underset{m\to\infty}{\sim} \text{const}\, x m^\rho \times\exp[\tau m^q + A_1 m^{2q-1} + A_2 m^{3q-2} + \cdots + A_{n-1}m^{nq-(n-1)}] \times[1+O(m^{\max[q-1,q+n(q-1)]})]\,, \tag{B14}$$

where $A_1 = \frac{1}{2}\tau^2 q^2$ [as in (B13)].

APPENDIX C

We indicate here how the definitions of Ref. 17 have been adjusted to fit the $h_{[k,l]}$.

Define on $\mathcal{S}(E)$ two sequences of operators by the following recursion:

$$M_m = \sum_\delta \Big(x_j M_{m-1} x_j + p_j M_{m-1} p_j - \frac{1}{4}\frac{\partial}{\partial x_j} M_{m-1} \frac{\partial}{\partial x_j} - \frac{1}{4}\frac{\partial}{\partial p_j} M_{m-1} \frac{\partial}{\partial p_j} \Big)\,,$$

$$\tilde{M}_m = \sum_j \left(\tfrac{1}{4} x_j \tilde{M}_{m-1} x_j + \tfrac{1}{4} p_j \tilde{M}_{m-1} p_j - \frac{\partial}{\partial x_j} \tilde{M}_{m-1} \frac{\partial}{\partial x_j} - \frac{\partial}{\partial p_j} \tilde{M}_{m-1} \frac{\partial}{\partial p_j} \right),$$

$$M_0 = \tilde{M}_0 = \mathbf{1}.$$

Both M_m and $\tilde{M}_m$ are positive; one can easily check that $\forall f, g \in \mathcal{S}(E)$: $(\tilde{f}, \tilde{M}_m \tilde{g}) = (f, M_m g)$, where $\sim$ denotes the Fourier transform F_1 (see Sec. 6). The M_m conserve the orthogonality of the $h_{[k,l]}$ (see Ref. 17),

$$(h_{[k',l']}, M_m h_{[k,l]}) = \delta_{[k][k']}\delta_{[l][l']} a(m; |k| + |l|);$$

the $a(m;k)$ satisfy the following relations:

$$\forall k: \; a(0;k) = 1,$$

$$m \geqslant 1 \Rightarrow a(m;k) = \frac{k+2n}{2} a(m-1;k+1) + \frac{k}{2} a(m-1;k-1).$$

This last recursion relation implies

$$2^{-m} \frac{\Gamma(k+2n+m)}{\Gamma(k+2n)} \leqslant a(m;k) \leqslant \frac{\Gamma(k+n+m)}{\Gamma(k+n)}.$$

The Hilbert space $H(\alpha, A)$ is then defined as

$$H(\alpha, A) = \left\{ f \in \mathcal{S}(E); \; \|f\|^2_{\alpha,A} = \sum_m \frac{1}{A^{2m}\Gamma^2(\alpha m)} (f, M_m f) < \infty \right\}.$$

Since

$$\|h_{[k,l]}\|^2_{\alpha,A} = \sum_m \frac{1}{A^{2m}\Gamma^2(\alpha m)} a(m; |k| + |l|) \leqslant (1+y)^{|k|+|l|+n-1} \sum_m \frac{(1+y^{-1})^m}{A^{2m}\Gamma^2(\alpha m)} \Gamma(m+1)$$

and this $\forall y > 0$, we see that if $\exists z > 0$ such that

$$k_{\alpha,A}(z) = \sum_m \frac{(1+z)^m m!}{A^{2m}\Gamma^2(\alpha m)}$$

converges, then $h_{[k,l]} \in H(\alpha, A) \forall [k], [l]$. This convergence is guaranteed for any A if $\alpha > \frac{1}{2}$, for $A > (2(1+z))^{1/2}$ if $\alpha = \frac{1}{2}$. Hence $H(\alpha, A)$ is an infinitely dimensional Hilbert space with orthonormal basis $\|h_{[k,l]}\|^{-1}_{\alpha,A} h_{[k,l]}$ if $\alpha > \frac{1}{2}$, or $\alpha = \frac{1}{2}$, $A > \sqrt{2}$. One can check that the topology on $H(\alpha, A)$ defined by the norm $\| \; \|_{\alpha,A}$ is really stronger than the topology on $\mathcal{S}(E)$, and that $H(\alpha, A)$ is a proper subset of $\mathcal{S}(E)$ (see Ref. 17).

The norm $\| \; \|_{\alpha,A}$ on $H(\alpha, A)$ can also be written as

$$\|f\|^2_{\alpha,A} = \sum_{[k][l]} |(f, h_{[k,l]})|^2 \gamma^{-2}_{|k|+|l|}(\alpha, A),$$

with

$$\gamma_k^{-2}(\alpha, A) = \sum_{m=0}^{\infty} \frac{a(m;k)}{\Gamma^2(\alpha m) A^{2m}}.$$

The Hilbert space $H(\bar{\alpha}, \bar{A})$ is then defined as the dual of $H(\alpha, A)$ with respect to the normal action of $\mathcal{S}'$ on $\mathcal{S}$. It can be constructed as the closure of $\mathcal{S}'(E)$ for the norm $\| \; \|_{\bar{\alpha},\bar{A}}$:

$$\|T\|^2_{\bar{\alpha},\bar{A}} = \sum_{[k][l]} |T(h_{[k,l]})|^2 \gamma^2_{|k|+|l|}(\alpha, A).$$

[1] I. Daubechies and A. Grossmann, J. Math. Phys. **21**, 2080 (1980).
[2] J. R. Klauder and E. C. G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968).
[3] H. Weyl, *Theory of Groups and Quantum Mechanics* (Dover, New York, 1950, English translation).
[4] A. Grossmann, Commun. Math. Phys. **48**, 191 (1976).
[5] E. P. Wigner, Phys. Rev. **40**, 743 (1932).
[6] I. Daubechies, Commun. Math. Phys. **75**, 229 (1980).
[7] A. M. Nachin, Thèse de 3e cycle, Université d'Aix-Marseille, 1972 (unpublished).
[8] I. Daubechies, Ph.D. Thesis, Vrije Universiteit Brussel, 1980 (unpublished).
[9] V. Bargmann, (a) Commun. Pure Appl. Math. **14**, 187 (1961); (b) Commun. Pure Appl. Math. **20**, 1 (1967).
[10] P. Krée and R. Racszka, Ann. Inst. Henri Poincaré A **28**, 41 (1978).
[11] I. Daubechies, J. Math. Phys. **21**, 1377 (1980).
[12] A. J. E. M. Janssen, "Bargmann transform, Zak transform and coherent states," preprint, Caltech, 1980 (unpublished).
[13] N. Aronszajn, Trans. Am. Math. Soc. **68**, 337 (1950).
[14] M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol. I, Functional Analysis* (Academic, New York, 1972).
[15] I. Daubechies, "Continuity statements and counterintuitive examples in connection with Weyl quantization," preprint Princeton University, submitted for publication.
[16] J. P. Antoine and M. Vause, Ann. Inst. Henri Poincaré A **35**, 195 (1981).
[17] A. Grossmann, J. Math. Phys. **6**, 54 (1965).
[18] I. M. Gel'fand and G. E. Shilov, *Generalized Functions, Vol. 2, Spaces of Fundamental and Generalized Functions* (Academic, New York, 1968, English translation).
[19] R. P. Boas, *Entire Functions* (Academic, New York, 1954).
[20] J. R. Klauder, J. Math. Phys. **4**, 1058 (1963).
[21] R. B. Dingle, *Asymptotic Expansions: their derivation and interpretation* (Academic, New York, 1973).

On the Laplace asymptotic expansion of conditional Wiener integrals and the Bender–Wu formula for x^{2N}-anharmonic oscillators

Ian Davies[a] and Aubrey Truman[b,c]
Mathematics Department, Heriot–Watt University, Riccarton, Currie, Edinburgh, EH14 4AS, Scotland

(Received 19 August 1981; accepted for publication 11 December 1981)

Rigorous results on the Laplace expansions of conditional Wiener integrals with functional integrands having a finite number of global maxima are established. Applications are given to the Bender–Wu formula for the x^{2N}-anharmonic oscillator.

PACS numbers: 02.30.Rz, 03.65.Db

1. INTRODUCTION

In a previous paper we showed how it is possible to extend Schilder's rigorous results for the Laplace asymptotic expansions of Wiener integrals, with integrands having unique nondegenerate global maxima, to conditional Wiener integrals.[1] There we gave applications of this result to the derivation of generalized Mehler kernel formulas. In this paper we treat the case of the Laplace expansion of a conditional Wiener integral with a functional integrand having a finite number of nondegenerate global maxima—a situation which arises frequently in theoretical physics.[2] Here we obtain equivalent results in this multiple maxima situation and give some applications to the Bender–Wu formula[3] for the x^{2N}-anharmonic oscillator, $N \geqslant 2$.

The Bender–Wu formula for the large order behavior of the perturbation series for the ground-state energy of the x^{2N}-anharmonic oscillator has been obtained formally from a function space integral, with a manifold of maximum points, by a number of authors.[4] Modulo the interchange of two limits, using a very clever argument, the behavior of the leading term for the large order behavior of the ground state of the x^4-anharmonic oscillator has been obtained rigorously from a function space integral by Simon.[5] The fact that the functional integrand in these treatments has a manifold of maximum points makes it difficult to use function space arguments to take the calculations beyond the leading term and difficult to extract more detailed information about even the leading term.

Here we give a functional integral realization of the ground-state energy for the x^{2N}-anharmonic oscillator so that (modulo the interchange of virtually the same two limits) the large order behavior is given rigorously by the Laplace expansion of a conditional Wiener integral with an integrand having exactly two nondegenerate global maxima. This offers the possibility of taking the function-space calculation to a higher order and yields more detailed information about the leading term.

At first sight the more detailed rigorous results which this realization gives seem somewhat disappointing in that they differ, in the fine detail, from the behavior given by the numerical Bender–Wu formula. (Presumably the problem here is the interchange of the two limits.) Nevertheless, as will be seen below, the method does lead to results for the x^{2N}-anharmonic oscillator ($N \geqslant 2$) parallel to Simon's for the x^4-anharmonic oscillator. To this extent then the interchange of the two limits does appear to be justified. Moreover, since we are dealing with the conditional Wiener integral, as opposed to a general Gaussian, our method is easier to apply. For instance, because the necessary bounds on the functional integrands in our Schilder type results can be expressed in terms of the sup-norm on the path-space, rather than the L^2-norm, these bounds are easy to check and do not require the clever technical estimates given by Simon. This seems to us to be the main practical advantage of applying the Schilder type results over using other treatments.[6]

We use substantially the same notation as in our previous paper. We recapitulate our main conventions here. $C_0[0, T]$ is the Banach space of continuous functions z: $[0, T] \to \mathbb{R}$ with $z(0) = z(T) = 0$, equipped with sup-norm $\|z\| = \sup_{\tau \in [0,T]} |z(\tau)|$. $C_0[0, T]$ supports the conditional Wiener measure, with covariance

$$\int_{C_0[0,T]} z(s)z(t)\, d\mu_{0,0;0,T}(z) = (2\pi T)^{-1/2} s(1 - t/T),$$

$0 \leqslant s < t \leqslant T$, with mean zero $\int_{C_0[0,T]} z(s)\, d\mu_{0,0;0,T}(z) = 0$, $0 \leqslant s \leqslant T$. For the associated probability measure $\mu^{-1}_{0,0;0,T}\{C_0[0,T]\}\, d\mu_{0,0;0,T}(z) = (2\pi T)^{1/2}\, d\mu_{0,0;0,T}(z)$, we use the notation

$$(2\pi T)^{1/2} \int_{C_0[0,T]} F(z)\, d\mu_{0,0;0,T}(z) = \mathbb{E}_z^T\{F(z)\},$$

for suitable functionals F. Abusing notation, for measurable sets A, we shall sometimes write

$$\mathbb{E}_z^T\{\chi_A(z)\} = \mathbb{E}_z^T\{A\},$$

where χ_A is the characteristic function of the set A. $C_0^*[0, T]$ is the reproducing kernel Sobolev space associated with $C_0[0, T]$; $z \in C_0^*[0, T]$, if z is absolutely continuous with derivative $\dot{z}(\cdot)$ in $L^2[0, T]$, $\int_0^T [\dot{z}(\tau)]^2 d\tau < \infty$.

We are now ready to state our basic theorem. This theorem deals with the case in which the functional integrand has two global maxima but the method of proof easily extends to the situation in which there are a finite number of such maxima.

[a] Supported by an SRC research studentship.
[b] Centre de Physique théorique, CNRS, Marseille, France.
[c] Supported in part by CNRS–ATP Internationale research grant: "Les intégrales stochastiques et leurs applications en mécanique statistique rigoreuse, quantification à la Feynman et théorie des champs." ATP no. 055.

 0022-2488/83/020255-12$02.50

Theorem 1: Let $F(z)$ be a real-valued continuous functional defined on $C_0[0, T]$ and suppose that the functional $\{F(z) - 2^{-1}\int_0^T[\dot{z}(\tau)]^2\,d\tau\}$ has exactly two distinct global maxima $x_1, x_2 \in C_0^*[0, T]$, with $\{F(x_i) - 2^{-1}\int_0^T[\dot{x}_i(\tau)]^2\,d\tau\} = b$, for $i = 1, 2$. If F satisfies conditions 1–6 below, then

$$\exp\{-b\lambda^{-2}\}\mathbb{E}_z^T\{\exp\{\lambda^{-2}F(\lambda z)\}\}$$
$$= \Gamma_0 + \lambda\Gamma_1 + \lambda^2\Gamma_2 + \cdots + \lambda^{m-3}\Gamma_{m-3} + O(\lambda^{m-2}),$$

as $\lambda \to 0$, where the Γ_j are functional integrals depending only on the functional F and its Frechet derivatives evaluated at x_1 and x_2.

(1) $F(z)$ is measurable.

(2) $F(z) \leqslant (b + L_1) + L_2\|z\|^2$, $\mu_{0,0;0,T}$ almost everywhere, L_1 and L_2 being positive numbers, with $L_2 < \min\{\gamma/2T, 1/4T\}$, γ being the constant in Lemma 6.

(3) $F(z)$ is continuous for $\|z\| \leqslant \max\{(L_1 + 1)^{1/2}/|L_2 - 1/2T|^{1/2}, [2T(L_1 + 1)/\gamma]^{1/2}\}$ and upper semicontinuous on $C_0[0, T]$. We do not preclude the possibility that, for some z_0, $F(z_0) = -\infty$, but then $F(z) \to -\infty$ as $z \to z_0$.

(4) $F(z)$ has $m \geqslant 3$ continuous Frechet derivatives in a ball of radius δ, $\delta > 0$, centered at x_1 and x_2 in $C_0[0, T]$. We further assume that $D^jF(x_i + \eta)z^j = O(\|z\|^j)$, if $\|\eta\| < \delta$, for $i = 1, 2$.

(5) For some $\epsilon > 0$, for $\|\eta\| < \delta$, $\mathbb{E}_z^T\{\exp\{(1 + \epsilon)D^2F(x_i + \eta)z^2/2\}\}$ is uniformly bounded for $i = 1, 2$.

(6) $\dot{x}_1(\cdot)$ and $\dot{x}_2(\cdot)$ are of bounded variation on $[0, T]$.

We need only prove the above result for $b = 0$. The result for $b \neq 0$ follows easily by considering the functional $\{F(z) - b\}$. We defer the actual proof until Sec. 3 of this paper. This proof is virtually the same as the proof we gave in our earlier treatment, save for the fact that we now have to divide up the function space into two disjoint pieces containing x_1 and x_2 and treat these separately. The burden of the proof is to show that it is possible to do this in a manner consistent with the earlier treatment.

In our earlier treatment, after Simon, we obtained two of our crucial estimates (Lemmas 5 and 7) by simply exploiting the Gaussian nature of the conditional Wiener integral and the underlying idea in the proof of Kolmogorov's lemma. In this paper, following a suggestion of Baxendale,[7] we present proofs of improved estimates for the conditional Wiener integral (e.g., Lemma 6 gives the best possible value for γ, $\gamma = 2$) by making use of the reflection principle for Brownian motion. This gives a simple proof of our basic estimate and, incidentally, a nice application of the strict Markov property of the Wiener process. Other proofs of this best possible estimate are available, but our results here do not merit these more abstract treatments. The improved estimates can be used to extend our results on the Mehler kernel formula, as we will discuss in a later paper.

We begin in the next section by establishing our application to the Bender–Wu formula for the x^{2N}-anharmonic oscillator. Our main result in this direction is contained in Theorem 2. For the statement of this theorem, let $N \geqslant 2$ and let $\beta > 0$, $H(\beta) = [2^{-1}(-d^2/dx^2 + x^2) + \beta x^{2N}]$ be the self-adjoint anharmonic oscillator Hamiltonian defined on some suitable domain in $L^2(\mathbb{R})$, with eigenfunctions ϕ_n and corresponding eigenvalues $E_n(\beta)$, arranged in order of increasing magnitude, $n = 0,1,2,\ldots$. Then we shall prove:

Theorem 2: Define the functions $g_n(T)$, for $n = 0,1,2,\ldots$ by

$$g(T, \beta) = \sum_{n=0}^{\infty} e^{-TE_n(\beta)}|\phi_n(0)|^2 = \sum_{n=0}^{\infty} g_n(T)\beta^n, \quad (\beta > 0).$$

Then, as $n \to \infty$,

$$\left[\frac{(-1)^n n! g_n(T)}{n^{Nn}}\right]^{1/n} \to e^{-\alpha(T)},$$

where, as $T \to \infty$,

$$\alpha(T) \to N + (N-1)\ln\{2^{1/N-1}\Gamma^2(N/(N-1))/(N-1)\Gamma(2N/(N-1))\}.$$

This result, which is formally consistent with the numerical Bender–Wu formula, is the analog for the x^{2N}-anharmonic oscillator of Simon's Theorem 18.3, which gives the large order behavior of $\{\Sigma_{n=0}^{\infty} e^{-TE_n(\beta)}\}$ for the x^4-anharmonic oscillator. As will be seen below and in the next section our results actually yield more detailed information. This leads to a slightly different result from the numerical Bender–Wu formula as we now explain.

The connection with the Bender–Wu formula comes about by observing that

$$E_0(\beta) = \sum_{n=0}^{\infty} E_n\beta^n = \lim_{T\to\infty} -T^{-1}\ln g(T, \beta), \tag{0}$$

since $|\phi_0(0)|^2 \neq 0$.

After previous authors, formally commuting the T and n limits, gives for the leading term $E_n \to \lim_{T\to\infty}\{-T^{-1}[g_0(T)]^{-1}\lim_{n\to\infty} g_n(T)\}$, as $n \to \infty$. Using the actual result established in this paper,

$$g_n(T) \to \gamma(T)\frac{n^{\beta(T)}e^{-n\alpha(T)}}{n!}n^{Nn}(-1)^n\left(1 + O\left(\frac{1}{n}\right)\right),$$

and Stirling's formula yields, as $n \to \infty$,

$$E_n \to [n(N-1)]!\left\{-\frac{1}{2}[\Gamma(2N/(N-1))/\Gamma^2(N/(N-1))]^{N-1}\right\}^n \lim_{T\to\infty}\left\{\frac{-T^{-1}\gamma(T)n^{\beta(T)-1}}{2\pi g_0(T)(N-1)^{1/2}}\right\}. \tag{1}$$

Our method (in contradistinction to previous treatments) gives rigorously explicit values for the above $\gamma(T)$ and $\beta(T)$ [as well as $\alpha(T)$] and enables all higher-order terms to be computed. These results lead to $\beta(T) \equiv 0$. To obtain exact agreement with the numerical Bender–Wu formula would require $\beta(T) \to \frac{1}{2}$. Thus, although the method does lead to the correct rapidly varying factors, as given by the first two terms, the relatively slowly varying third term does not agree. Hence, the T and n limits only seem to be interchangeable for $\lim_{n\to\infty} E_n^{1/n}$, E_n being defined by Eq. (0). This then is the extent to which the limits commute. We discuss this further in the Conclusion.

2. THE PROOF OF THEOREM 2

We begin with a proposition which may be of some independent interest. We fix the integer $N>2$ in what follows.

Proposition 1: For $\beta>0$, let $H(\beta)=2^{-1}(-d^2/dx^2+x^2)+\beta x^{2N}$ be the quantum mechanical Hamiltonian for the x^{2N}-anharmonic oscillator. Then, on a suitable domain in $L^2(\mathbf{R})$, $H(\beta)$ is self-adjoint, with a discrete spectrum $\{E_0(\beta), E_1(\beta), E_2(\beta),\cdots\}$, $0<E_0(\beta)<E_1(\beta)<\cdots$, with corresponding non-degenerate orthonormal eigenfunctions $\{\phi_0, \phi_1,\cdots\}$, $\phi_j\in C^\infty(\mathbf{R})$ and $\sup_{x\in\mathbf{R}}|\phi_j(x)|<\infty$, each j. Moreover, defining $G(x, y, T)$ by

$$G(x,y,T)=\int d\mu_{x,0;y,T}(z)\exp\left\{-2^{-1}\int_0^T z^2(s)\,ds - \beta\int_0^T z^{2N}(s)\,ds\right\},\quad T>0,$$

$\mu_{x,0;y,T}$ being the conditional Wiener measure,

$$G(x,y,T)=\sum_{n=0}^{\infty} e^{-E_n(\beta)T}\phi_n^*(x)\phi_n(y),\quad x,y\in\mathbf{R}$$

pointwise.

Proof: See the Appendix. □

The above proposition leads easily to the desired functional integral representation for $g_n(T)$. Setting $x=y=0$, $G(0,0,T)=g(T,\beta)=\Sigma_{n=0}^{\infty}g_n(T)\beta^n$, and so we obtain

$$\frac{(-1)^n g_n(T)n!}{n^{Nn}}=\int d\mu_{0,0;0,T}(z)\left\{\int_0^T\left[\frac{z(s)}{n^{1/2}}\right]^{2N}ds\right\}^n \times\exp\left\{-2^{-1}\int_0^T z^2(s)\,ds\right\}.$$

Equivalently

$$\frac{(-1)^n g_n(T)n!}{n^{Nn}}=(2\pi T)^{-1/2}\mathbf{E}_z^T\{\exp\{nF(n^{-1/2}z)\}\},$$

where

$$F(z)=-2^{-1}\int_0^T z^2(s)\,ds+\ln\left[\int_0^T z^{2N}(s)\,ds\right],\quad z\neq 0,$$

and $F(0)=-\infty$. Hence, to apply Theorem 1 we must prove that

$$G(z)=\left\{2^{-1}\int_0^T \dot z^2(s)\,ds+2^{-1}\int_0^T z^2(s)\,ds-\ln\int_0^T z^{2N}(s)\,ds\right\}$$

has exactly two global minima in $C_0^*[0,T]$. This we now establish.

Lemma 1: The functional $G(z)=2^{-1}\int_0^T\dot z^2(s)\,ds+2^{-1}\int_0^T z^2(s)\,ds-\ln\int_0^T z^{2N}(s)\,ds$ $(N>2)$ attains its global minimum $\alpha(T)$ at at least one path $X\in C_0^*[0,T]$; X satisfies the "Euler Lagrange" equation

$$-\ddot X(t)+X(t)-2NX^{2N-1}(t)/\int_0^T X^{2N}(s)\,ds=0,\quad t\in[0,T].$$

Proof: The proof is only a slight extension of well-known results in the direct methods of the calculus of variations.[8] Firstly the existence of the minimizing sequence follows from the inequalities

$$\int_0^T z^{2N}(s)\,ds\leq\|z\|^{2N-2}\int_0^T z^2(s)\,ds,$$

$$\|z\|=\left|\int_0^t\dot z(s)\,ds\right|\leq T^{1/2}\left\{\int_0^T\dot z^2(s)\,ds\right\}^{1/2},\quad \text{some } t\in[0,T],$$

where $\|\ \|$ denotes sup-norm. We can now see that, for $N>2$,

$$G(z)\geq 4^{-1}\int_0^T\dot z^2(s)\,ds+\left\{\frac{\|z\|^2}{4T}-(2N-2)\ln\|z\| + 2^{-1}\int_0^T z^2(s)\,ds-\ln\int_0^T z^2(s)\,ds\right\}\geq 4^{-1}\int_0^T\dot z^2(s)\,ds+c,$$

for some constant c. Hence, $\exists$ a minimizing sequence $\{z_n\}$ with $\lim_{n\to\infty}G(z_n)=\alpha(T)$. Without loss we can assume that $G(z_1)\geq G(z_n)$, $n=2,3,\cdots$ and so by the Cauchy–Schwarz inequality

$$|z_n(s)-z_n(t)|=\left|\int_s^t\frac{dz_n}{d\tau}(\tau)\,d\tau\right|\leq|s-t|^{1/2}2[G(z_1)-c]^{1/2},\quad s,t\in[0,T].$$

Thus $\{z_n\}$ is a family of uniformly bounded equicontinuous functions, which by Ascoli's theorem[9] has a subsequence uniformly convergent to some $X\in C_0^*[0,T]$. The lower semicontinuity of G now guarantees that $G(X)=\alpha(T)$. The result that X satisfies the above equation follows because X must be a local minimum as well as a global minimum. □

We now scale according to $X=kY$, where k is given by $k^{2N-2}=\int_0^T X^{2N}(s)\,ds$. Then Y satisfies the "instanton" equation

$$-\ddot Y(t)+Y(t)-2NY^{2N-1}(t)=0,\quad Y(0)=Y(T)=0, \tag{2}$$

$k^2=[\int_0^T Y^{2N}(s)\,ds]^{-1}$. Also, multiplying the last equation by $Y(t)$ and integrating by parts leads to

$$\int_0^T\dot Y^2(s)\,ds+\int_0^T Y^2(s)\,ds=2N\int_0^T Y^{2N}(s)\,ds,$$

from which it follows that

$$G(X)=N+(N-1)\ln\left\{\int_0^T Y^{2N}(s)\,ds\right\},$$

where $X=kY$, Y satisfying Eq. (2).

As expected Eq. (2) governs classical motion in the potential $V(y)=-2^{-1}y+y^{2N}$, leading to the usual "instanton" interpretation of Y. The potential V has turning points at $y=0$, or $y=\pm(2N)^{-1/(2N-2)}$, V having a double-well shape. Each instanton solution has a constant energy $E=2^{-1}(\dot Y^2(s)-Y^2(s))+Y^{2N}(s)$. For the instanton to satisfy $Y(0)=Y(T)=0$, the corresponding energy $E>0$. Hence the solution Y is periodic, oscillating between the extremes $\pm y_1(E)$, $y_1(E)$ being a positive solution of $E=-2^{-1}y^2+y^{2N}$. Trivially $y_1(E)\geq 2^{-1/2N-2}>(2N)^{-1/2N-2}$, $N>1$, and so, for $E\geq 0$, $\partial E/\partial y_1=-y_1+2Ny_1^{2N-1}\geq 0$. For $y_1=y_1(E)$, $E\geq 0$, the periodic time $t(y_1)$ is given by

$$t(y_1) = 2\sqrt{2}\int_0^{y_1} \left\{(y_1^2 - y^2)\left(-\frac{1}{2} + \frac{y_1^{2N} - y^{2N}}{y_1^2 - y^2}\right)\right\}^{-1/2} dy$$

$$= 2\sqrt{2}\int_0^{\pi/2} \{-\tfrac{1}{2} + y_1^{2N-2}\sec^2\theta(1 - \sin^{2N}\theta)\}^{-1/2}\, d\theta,$$

where we have set $y = y_1 \sin\theta$.

It follows easily from above expression that, for $N > 1$, $\partial t(y_1)/\partial y_1 < 0$ and $t(y_1)$ is a monotonic decreasing function of y_1. Hence, fixing the periodic time T', say, uniquely fixes a corresponding positive value of the energy $E = E(T')$, with $E(T') \nearrow \infty$ as $T' \searrow 0$.

To summarize then: The solutions Y of Eq. (2) can be labeled in 1-1 fashion by their periodic times T', which give rise to corresponding energies $E = E(T') \geqslant 0$. Clearly, to satisfy the boundary condition $Y(0) = Y(T) = 0$ we must demand that $T' = T/n$, or $T' = T/(m + \frac{1}{2})$; $n = 1,2,...$, $m = 0,1,2,...$, to within a sign, Y consisting of n or $(m + \frac{1}{2})$ cycles of the periodic orbit starting from 0. At least one solution Y must give the global minimizer X of G.

Hence, according to the above identity for G, minimizing $G(z)$ over $C_0^*[0, T]$ is equivalent to minimizing $\{nS[T/n], (m + \frac{1}{2})S[T/(m + \frac{1}{2})] | n = 1,2,..., m = 0,1,2,...\} = \{2^{-1}nS[2T/n] | n = 1,2,...\}$, where

$$S[T'] = 2\sqrt{2}\int_0^{y_1} y^{2N}[E(T') - V(y)]^{-1/2}\, dy, \tag{3}$$

$V(y) = -2^{-1}y^2 + y^{2N}$ and $E(T') = V(y_1)$. It is easy to prove that $S[T']$ as defined above is a continuous function of $T' \in (0, \infty)$ with

$$S[\infty] = \lim_{T' \to \infty} S[T']$$

$$= 2^2 2^{-N/(N-1)} \int_0^1 x^{2N-1}(1 - x^{2N-2})^{-1/2}\, dx.$$

Here the last identity follows by observing that, as $T' \to \infty$, $y_1 \to 2^{-1/2N-2}$ and by changing integration variables using $y = 2^{-1/2N-2}\sin\theta = 2^{-1/2N-2}x$. We now need an elementary lemma.

Lemma 2: Let $S[T']$ be a continuous function, for $T' \in (0, \infty)$, with $S[T'] \to S[\infty]$ as $T' \to \infty$. Then, if $S[T'] > 2^{-1}S[\infty] > 0$, for $T' \in (0, \infty)$, we can deduce that, for all sufficiently large T,

$$\min_{n = 1,2,\cdots} \left\{2^{-1}nS\left[\frac{2T}{n}\right]\right\} = 2^{-1}S[2T].$$

Proof: For each T_0, consider the sequence $\{nS[T_0/n]\}$. Then, since $nS[T_0/n] \to \infty$, as $n \to \infty$, $\exists$ a finite integer $n(T_0)$ such that

$$n(T_0)S\left[\frac{T_0}{n(T_0)}\right] = \min_{n = 1,2,\cdots} \left\{nS\left[\frac{T}{n}\right]\right\}.$$

Either for all sufficiently large T_0, $n(T_0) = 1$, or $\exists$ a sequence $\{T_r\}$, $T_r \to \infty$, as $r \to \infty$, with $n(T_r) \geqslant 2$. In the first case there is nothing further to prove. In the second case either $T_r/n(T_r) < K$, for some finite K, for all r, or $\exists$ a subsequence $\{T_{r_s}/n(T_{r_s})\}$, with $T_{r_s}/n(T_{r_s}) \to \infty$ as $s \to \infty$, where in both cases $S[T_r/n(T_r)] - n(T_r)^{-1}S[T_r] < 0$, $n(T_r) \geqslant 2$. In the first subcase, letting $r \to \infty$, we see that $\inf_{t < K} S[t] < 2^{-1}S(\infty)$, which is not so. In the second subcase, letting $s \to \infty$, gives $S[\infty] \leqslant 2^{-1}S[\infty]$, which is not so, since $S[\infty] > 0$. □

Lemma 3: Let $S[T']$ be defined as in Eq. (3). Then, for each integer $N \geqslant 2$, $S[T'] > 2^{-1}S[\infty]$, $T' \in (0, \infty)$.

Proof: Arguing as above,

$$S[T'] = 2^{3/2}\int_0^{\pi/2} y_1^{2N}\sin^{2N}\theta\, \{-\tfrac{1}{2} + y_1^{2N-2}$$

$$\times \sec^2\theta(1 - \sin^{2N}\theta)\}^{-1/2}\, d\theta.$$

Using the obvious inequality, for $0 \leqslant \theta \leqslant \pi/2$,

$$\sec^2\theta(1 - \sin^{2N}\theta) = 1 + \sin^2\theta + \cdots + \sin^{2N-2}\theta \leqslant N,$$

we obtain

$$S[T'] \geqslant 2^{3/2}y_1^{2N}\{-\tfrac{1}{2} + Ny_1^{2N-2}\}^{-1/2}\int_0^{\pi/2} \sin^{2N}\theta\, d\theta,$$

where by inspection $y_1 \geqslant 2^{-1/2N-2}$, $E \geqslant 0$. A simple computation now shows that

$$\min_{y_1 \geqslant 2^{-1/2N-2}} [y_1^{2N}\{-\tfrac{1}{2} + Ny_1^{2N-2}\}^{-1/2}]$$

$$= 2^{-(N+1)/(2N-2)}(N-1)^{-1/2},$$

the minimum being attained at $y_1 = 2^{-1/2N-2}$. Hence, we have shown that

$$S[T'] \geqslant 2^{-1/N-1}(N-1)^{-1/2}\frac{(2N-1)!!\pi}{2^N N!},$$

where we have used the well-known result $\int_0^{\pi/2}\sin^{2N}\theta\, d\theta = (2N-1)!!\pi/2^{N+1}N!$, $(2N-1)!! = (2N-1)(2N-3)\cdots 1$. Integrating by parts

$$S[\infty] = 2^2 2^{-N/(N-1)}\int_0^1 x^{2N-1}(1 - x^{2N-2})^{-1/2}\, dx$$

$$= 2^2 2^{-N/(N-1)}(N-1)^{-1}\int_0^1 (1 - x^{2N-2})^{1/2} 2x\, dx.$$

Observing that the final integral is a monotonic increasing function of N, we see that

$$S[\infty] < 2^2 2^{-N/(N-1)}(N-1)^{-1}.$$

Comparing the last inequality with the inequality for $S[T']$ we see that all we must prove is that

$$\pi \geqslant 2^N N!(N-1)^{-1/2}[(2N-1)!!]^{-1},$$

for each integer $N \geqslant 2$. But denoting by I_n, $I_n = \int_0^{\pi/2}\sin^n\theta\, d\theta$, as is well known

$$I_{2N} = (2N-1)!!\pi/2^{N+1}N!$$

and

$$I_{2N+1} = 2^N N!/(2N+1)!!$$

and since

$$I_{2N} > I_{2N+1}, \quad \pi^{1/2} > 2^N N![(2N-1)!!]^{-1}(N + \tfrac{1}{2})^{-1/2}.$$

The final result now follows from the simple observation that, for $N \geqslant 2$, $\pi^{1/2} > (N + \frac{1}{2})^{1/2}(N-1)^{-1/2}$. □

The last two lemmas lead to the corollary.

Corollary: For each fixed integer $N \geqslant 2$, for sufficiently large T, $F(z) = -2^{-1}\int_0^T [\dot{z}(s)]^2\, ds - 2^{-1}\int_0^T z^2(s)\, ds + \ln\{\int_0^T z^{2N}(s)\, ds\}$, achieves its global maximum $(-\alpha(T))$ at exactly two paths X in $C_0^*[0, T]$, $X(s)$

$= \pm Y_0(s)[\int_0^T Y_0^{2N}(t)dt\,]^{-1/2}$, $s\in[0,T]$, where $Y_0(s)$ is the unique instanton solution of $\ddot{Y}(s) = Y(s) - 2NY^{2N-1}(s)$, $s\in[0,T]$, with $Y(0) = Y(T) = 0$, with time period $2T$, and corresponding energy $E(2T)\geqslant 0$. In addition

$$\alpha(T)\to N + (N-1)\ln\left\{\frac{2^{1/N-1}\Gamma^2(N/(N-1))}{(N-1)\Gamma(2N/(N-1))}\right\},$$

as $T\to\infty$.

Proof: From the above we see that $\alpha(T)\to N + (N-1)\ln\{2^{-1}S[\infty]\}$, where

$$S[\infty] = 2^2 2^{-N/(N-1)}(N-1)^{-1}\int_0^1 (1-x^{2N-2})^{1/2}2x\,dx.$$

The proof is completed by the following identity, which is a simple exercise on Γ and B functions using the duplication formula,

$$\int_0^1 (1-x^{2N-2})^{1/2}2x\,dx = (N-1)^{-1}B\left(\frac{3}{2},\frac{1}{N-1}\right)$$
$$= 2^{2/N-1}\frac{\Gamma^2(N/(N-1))}{\Gamma(2N/(N-1))}.$$ □

It is a routine exercise now to check the conditions 1–6 in Theorem 1 for the functional F defined above $F: C_0^*[0,T]\to\mathbf{R}$. In this connection notice in particular how easy it is to establish condition (1) because $\|\ \|$ here refers to the sup-norm and not the L^2-norm. Finally, recall that $H = C_0^*[0,T]$ is a reproducing kernel Sobolev space when equipped with the inner product

$$\langle g, h\rangle = \int_0^T \dot{g}(s)\dot{h}(s)\,ds,$$

$g, h\in H$. We require a last lemma to complete the proof.

Lemma 4: The functional $F: H = C_0^*[0,T]\to\mathbf{R}$ has a trace-class second Frechet derivative at X, $L = D^2F(X)\in L(H,H)$, $X = \pm Y_0[\int_0^T Y_0^{2N}(t)dt\,]^{-1/2}$ being either global minimizer of F. If det denotes the Fredholm determinant defined with respect to H,

$$\det[1+L\,] = -4\pi E\frac{\partial T}{\partial E}(N-1),$$

where E is the energy of Y_0, and (with above conditions)

$$T(E) = \sqrt{2}\int_0^{y_1(E)}\frac{dy}{[E-V(y)]^{1/2}}.$$

Proof: A straightforward calculation yields

$$\left(h,\left[-\frac{d^2}{ds^2} + D^2F(X)\right]h\right)_{L^2} = \langle h,(1+L)h\rangle,$$

where $L\in L(H,H)$ is defined by $L = (L_0 + 4N^2I^{-1}\pi)$, $I = \int_0^T X^{2N}(s)\,ds$,

$$-\frac{d^2}{ds^2}(L_0h)(s) = h(s) - 2N(2N-1)I^{-1}X^{2N-2}(s)h(s),$$

and

$$-\frac{d^2}{ds^2}(\pi h)(s) = X^{2N-1}(s)\int_0^T X^{2N-1}(t)h(t)\,dt.$$

The above L_0 can be expressed as the product of two Hilbert–Schmidt operators in exactly the same way as in Corollary 2 of Sec. 3 of Ref. 1a. Hence L_0 is trace-class. We now show that $(1+L_0)$: $H\to H$ is a bijection, i.e., $(1+L_0)v = 0\Rightarrow v = 0\in H$.

To see this from above observe that v must satisfy

$$\ddot{v}(s) = v(s) - 2N(2N-1)I^{-1}X^{2N-2}(s)v(s),$$

with $v(0) = v(T) = 0$. In terms of the scaled variable Y_0, v satisfies

$$\ddot{v}(s) = v(s) - 2N(2N-1)Y_0^{2N-2}(s)v(s), \tag{4}$$

with $v(0) = v(T) = 0$. However, Y_0 satisfies the instanton equation

$$\ddot{Y}_0(s) = Y_0(s) - 2NY_0^{2N-1}(s),$$

with $Y_0(0) = Y_0(T) = 0$. The Jacobi fields along Y_0 give rise to the two linearly independent solutions of Eq. (4). Let p_0 denote the initial momentum of the instanton solution Y_0. Then the Jacobi field $v_0 = \partial Y_0/\partial p_0$ spans the solution space of Eq. (4) with $v_0(0) = 0$. We now show that $v_0(T)\neq 0$, $N\geqslant 2$.

To see this let $y(\xi, p_0, s)$ be the solution of the instanton equation, Eq. (2), with $y(\xi,p_0,0) = \xi$, $\dot{y}(\xi,p_0,0) = p_0$. If p_0 is chosen so that $p_0^2 = 2E$, E being the instanton energy, $Y_0(s) = y(0,p_0,s)$ and $v_0(s) = (\partial y/\partial p_0)(0,p_0,s)|_{p_0=\sqrt{2E}}$ satisfies Eq. (4) with $v_0(0) = 0$, $\dot{v}_0(0) = 1$, the remaining Jacobi field $u_0(s) = (\partial y/\partial\xi)(0,p_0,s)|_{p_0=\sqrt{2E}}$ satisfying Eq. (4), with $u_0(0) = 1$, $\dot{u}_0(0) = 0$. By definition of $T(E) = T(p_0)$,

$$y(0,p_0,T(p_0)) = 0.$$

Partially differentiating with respect to p_0 yields $v_0(T) = p_0\,\partial T/\partial p_0\neq 0 (N\neq 1)$. Hence, $(1+L_0)$: $H\to H$ is a bijection. Arguing as in Ref. 10 we actually obtain, for $N\neq 1$,

$$\det(1+L_0) = 2\pi p_0\frac{\partial T}{\partial p_0} = 4\pi E\frac{\partial T}{\partial E}\neq 0, \tag{5}$$

confirming that $(1+L_0)$ is a bijection.

We now turn our attention to π. By definition of π, for $g, h\in H$,

$$\langle g,\pi h\rangle = (\dot{g},(\pi h)^{\cdot})_{L^2} = -(g,(\pi h)^{\cdot\cdot})_{L^2}$$
$$= (g,X^{2N-1})_{L^2}(X^{2N-1},h)_{L^2} = \langle g,W\rangle\langle W,h\rangle,$$

the last step following by definition of W, $\ddot{W}: = X^{2N-1}$, $W\in H$. Thus π is just a multiple of a projection in H.

Since we know that $(1+L) = (1+L_0)[1 + 4N^2I^{-2}\times(1+L_0)^{-1}\pi]$, we can now deduce that

$$\det(1+L)$$
$$= \det(1+L_0)[1 + 4N^2I^{-2}\langle W,(1+L_0)^{-1}\pi W\rangle\langle W,W\rangle^{-1}].$$

It only remains to calculate $Z = (1+L_0)^{-1}W\in H$. Using $W = (1+L_0)Z$ gives $\ddot{W} = \ddot{Z} + (L_0\ddot{Z})$, i.e.,

$$X^{2N-1} = \ddot{Z} - Z + 2N(N-1)I^{-1}X^{2N-2}Z,$$

where $\ddot{X} = X - 2NI^{-1}X^{2N-1}$. Trying $Z = kX$ gives the (unique) solution for $k = I/2N(2N-2)$, $N\geqslant 2$. The final expression for $\det(1+L)$ follows by combining Eq. (5), the definition of π and the last expression for $Z = (1+L_0)^{-1}W$. □

We can summarize the above: let

$$\sum_{n=0}^{\infty} e^{-TE_n(\beta)}|\phi_n(0)|^2 = \sum_{n=0}^{\infty} g_n(T)\beta^n,$$

$E_n(\beta)$, ϕ_n being the eigenvalues and eigenfunctions of $H(\beta) = 2^{-1}(-d^2/dx^2 + x^2) + \beta x^{2N}$, $\beta > 0$, then, as $n\to\infty$,

$$g_n(T) \to \gamma(T) e^{-n\alpha(T)} \frac{n^{Nn}}{n!} (-1)^n \left[1 + O\left(\frac{1}{n}\right)\right] \quad (N \geqslant 2),$$

where $\gamma(T) = 2[\det(1+L)]^{-1/2} = (-\pi E\, \partial T/\partial E)^{-1/2} \times (N-1)^{1/2}$, E being the instanton energy, the instanton Y_0 having time period $2T$. Finally, we recall that

$$\alpha(T) \to N + (N-1)\ln\left\{2^{1/N-1}\Gamma^2\left(\frac{N}{N-1}\right)\Big/ (N-1)\Gamma\left(\frac{2N}{N-1}\right)\right\}$$

as $T \to \infty$.

This completes the proof of Theorem 2.

3. PRELIMINARY LEMMAS AND THE PROOF OF THEOREM 1

A. Preliminary lemmas

The proof which we give here is a simple extension of the proof which we gave in Ref. 1a. For Wiener integrals the basic argument was first elucidated by Schilder in Ref. 1b. We shall frequently refer to these papers in what follows. Following a suggestion of Baxendale we first use the reflection principle to establish some improved estimates. The corresponding estimates in our earlier treatment were established by using the underlying idea in Kolmogorov's lemma. The treatment below seems more transparent. Since the reflection principle is not well known to physicists we include a brief account of it below.[11]

Lemma 5: Let $x(s)$ be Brownian motion starting at 0. Let $\tau(a)$ be the first hitting time of a, i.e., $\tau(a) = \inf\{s > 0 | x(s) = a\}$. Let B be any Borel subset of $(-\infty, 0)$. Then, if $\mathscr{W}_0$ denotes expectation with respect to Wiener measure $d\mu^W(x)$,

$$\mathscr{W}_0\{x(t) \in (a+B),\ \tau(a) < t\} = \mathscr{W}_0\{x(t) \in (a-B),\ \tau(a) < t\}.$$

Remark: There is a simple geometrical interpretation of the reflection principle: to every Brownian path $x(\cdot)$ with $\tau(a) < t$ arriving in the set $(a+B)$ at time t there is an equally likely reflected Brownian path $r(a)x(\cdot)$, reflected in the line $x = a$ for times $s > \tau(a)$, arriving in the set $(a-B)$ at time t (see Fig. 1).

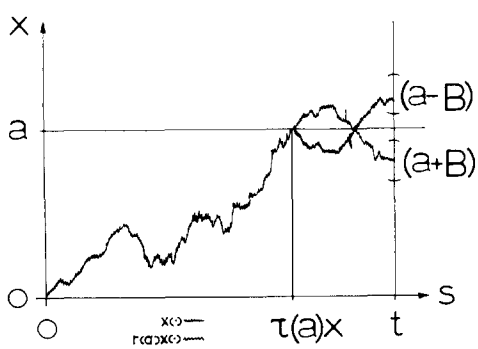

FIG. 1. The reflection principle.

Proof: Let $f = \chi(a+B)$ be the characteristic function of the set $(a+B)$. For each $\alpha > 0$,

$$\int_0^\infty e^{-\alpha t} \mathscr{W}_0[x(t) \in (a+B),\ \tau(a) < t]\, dt$$
$$= \mathscr{W}_0\left[\int_{\tau(a)}^\infty e^{-\alpha t} f(x(t))\, dt\right]$$
$$= \mathscr{W}_0\left[e^{-\alpha\tau(a)} \int_0^\infty e^{-\alpha s} f(x(s+\tau(a)))\, ds\right].$$

For the stopping time $\tau(a)$, let $B_{\tau(a)+}$ be defined as on p. 67 of Simon.[5a] Taking conditional expectations $\mathscr{W}_0(\cdot | B_{\tau(a)+})$, using the fact that $e^{-\alpha\tau(a)}$ is $B_{\tau(a)+}$ measurable, gives

$$\text{rhs} = \mathscr{W}_0\left[e^{-\alpha\tau(a)} \mathscr{W}_0\left(\int_0^\infty e^{-\alpha s} f(x(s+\tau(a)))\, ds \,|\, B_{\tau(a)+}\right)\right].$$

Using the strict Markov property of Dynkin–Hunt,[12] we see that

$$\text{rhs} = \mathscr{W}_0\left[e^{-\alpha\tau(a)} \int_0^\infty e^{-\alpha s} \mathscr{W}_a[f(x(s))]\, ds\right],$$

$\mathscr{W}_a$ being the expectation with respect to $d\mu^W(a+x)$. Hence, setting $F_B(s) = \int_{b \in (a+B)} e^{-(a-b)^2/2s} (2\pi s)^{-1/2}\, db$ gives, for $\alpha > 0$,

$$\int_0^\infty e^{-\alpha t} \mathscr{W}_0[x(t) \in (a+B),\ \tau(a) < t]\, dt$$
$$= \mathscr{W}_0\left[e^{-\alpha\tau(a)} \int_0^\infty e^{-\alpha s} F_B(s)\, ds\right]$$
$$= \mathscr{W}_0\left[e^{-\alpha\tau(a)} \int_0^\infty e^{-\alpha s} F_{-B}(s)\, ds\right]$$
$$= \int_0^\infty e^{-\alpha t} \mathscr{W}_0[x(t) \in (a-B),\ \tau(a) < t]\, dt,$$

proving the result. □

To obtain the usual form of André's reflection principle, set $B = (-\infty, 0)$ in above to give

$$\mathscr{W}_0[x(t) > a | \tau(a) < t] = 2^{-1}.$$

By continuity of sample paths then observe

$$\mathscr{W}_0[x(t) > a] = \mathscr{W}_0[x(t) > a,\ \tau(a) < t]$$
$$= \mathscr{W}_0[\tau(a) < t]\, \mathscr{W}_0[x(t) > a | \tau(a) < t]$$
$$= 2^{-1} \mathscr{W}_0[\tau(a) < t].$$

We now come to one of our basic estimates.

Lemma 6: For some fixed constants $C, \gamma > 0$

$$\mathbb{E}_z^T\{\|z\| > a'\} < C \exp\{-\gamma a'^2/T\}.$$

The best possible value of γ is 2.

Proof: Set $B = (-a + A)$, where $A \subset (-\infty, a)$, some $a > 0$, in above lemma. We obtain

$$\mathscr{W}_0[x(T) \in A,\ \tau(a) < T]$$
$$= \mathscr{W}_0[x(T) \in (2a - A),\ \tau(a) < T]$$
$$= \mathscr{W}_0[x(T) \in (2a - A)],$$

since $x(T) \in (2a - A) \Rightarrow \tau(a) < T$ by continuity of sample paths. Hence, we have shown that

$$\mathscr{W}_0\{\tau(a) < T | x(T) \in A\}$$
$$= \int_{(2a-A)} (2\pi T)^{-1/2} e^{-b^2/2T}\, db \Big/ \int_A e^{-b^2/2T} (2\pi T)^{-1/2}\, db.$$

Choosing $A = (-\delta, \delta)$ and letting $\delta \to 0$ gives for the conditional Wiener process z

$$\mathbb{E}_z^T\left\{\max_{s<T} z(s) > a\right\} = e^{-2a^2/T}.$$

Finally then

$$e^{-2a^2/T} < \mathbb{E}_z^T\left\{\max_{s<T} |z(s)| > a\right\} < \mathbb{E}_z^T\left\{\max_{s<T} z(s) > a\right\} + \mathbb{E}_z^T\left\{\max_{s<T} -z(s) > a\right\} = 2e^{-2a^2/T},$$

which proves the lemma. □

Let $z^n(\cdot)$ denote the polygonalization of the sample path z.

$$z^n(s) = z\left(\frac{jT}{n}\right) + \left(s - \frac{jT}{n}\right)\left[z\left(\frac{(j+1)T}{n}\right) - z\left(\frac{jT}{n}\right)\right]\frac{n}{T}, \qquad \frac{jT}{n} \leqslant s \leqslant \frac{(j+1)T}{n}$$

for $j = 0,1,2,\ldots,(n-1)$. Then the following lemma is a trivial consequence of the triangle inequality.

Lemma 7 (Schilder): If

$$\max_{0<j<n-1}\left[\frac{jT}{n} \leqslant \overset{\sup}{s} \leqslant \frac{(j+1)T}{n}\left|z(s) - z\left(\frac{jT}{n}\right)\right|\right] < \frac{\delta}{2},$$

then $\|z - z^n(\cdot)\| < \delta$.

Proof: See Lemma 6 in Ref. 1b. □

We also need the basic estimate:

Lemma 8:

$$\mathbb{E}_z^T\{\|z - z^m(\cdot)\| \geqslant \delta'\} \leqslant \frac{32}{(2\pi)^{1/2}}\delta'^{-1}(mT)^{1/2}\exp\left\{-\frac{m\delta'^2}{8T}\right\}.$$

Proof: Unless otherwise stated $\overset{\sup}{s}$ denotes $\overset{\sup}{s \in \left[\frac{jT}{m}, \frac{(j+1)T}{m}\right]}$ in what follows and initially j can vary so that $j = 1,2,\ldots,(m-2)$. Then

$$\mathscr{W}_0\left\{\overset{\sup}{s}\left(x(s) - x\left(\frac{jT}{m}\right)\right) > b, x(T) \in (-\delta, \delta)\right\} = \int \mathscr{W}_0\left\{\overset{\sup}{s}\left(x(s) - x\left(\frac{jT}{m}\right)\right) > b, x\left(\frac{(j+1)T}{m}\right) \in da, x(T) \in (-\delta, \delta)\right\}.$$

Let

$$A_j = \left\{\overset{\sup}{s}\left(x(s) - x\left(\frac{jT}{m}\right)\right) > b, x\left(\frac{(j+1)T}{m}\right) \in da\right\}$$

and observe that $A_j \in B_{(j+1)T/m}$, B_τ being the σ-field generated by $\{x(s) | s \leqslant \tau\}$. Then, taking conditional expectations $\mathscr{W}_0(\cdot | B_{(j+1)T/m})$, we obtain

$$\text{rhs} = \int \mathscr{W}_0\left\{\chi_{A_j}\,\mathscr{W}_{x((j+1)T/m)}\left\{x\left[T - \frac{(j+1)T}{m}\right] \in (-\delta, \delta)\right\}\right\} = \int \mathscr{W}_0\left\{\overset{\sup}{s}\left(x(s) - x\left(\frac{jT}{m}\right)\right) > b, x\left(\frac{(j+1)T}{m}\right) \in da\right\}\mathscr{W}_a\left\{x\left[T - \frac{(j+1)T}{m}\right] \in (-\delta, \delta)\right\}.$$

Repeating the above argument for the random variable $x(jT/m)$, we obtain

$$\mathscr{W}_0\left\{\overset{\sup}{s}\left(x(s) - x\left(\frac{jT}{m}\right)\right) > b, x\left(\frac{(j+1)T}{m}\right) \in da\right\} = \int \mathscr{W}_0\left\{x\left(\frac{jT}{m}\right) \in dc\right\}\mathscr{W}_0\left\{\overset{\sup}{s}\left(x(s) - x\left(\frac{jT}{m}\right)\right) > b, x\left(\frac{(j+1)T}{m}\right) - x\left(\frac{jT}{m}\right) \in d(a-c) | B_{jT/m}\right\}.$$

Using the Markov property

$$\text{rhs} = \int dc\, e^{-mc^2/2jT}\left(\frac{2\pi jT}{m}\right)^{-1/2} \times \mathscr{W}_0\left\{\sup_{0<s<T/m} x(s) > b, x\left(\frac{T}{m}\right) \in d(a-c)\right\}$$

and using the reflection principle

$$\mathscr{W}_0\left\{\overset{\sup}{s}\left(x(s) - x\left(\frac{jT}{m}\right)\right) > b, x\left(\frac{(j+1)T}{m}\right) \in da\right\} = \int dc\, e^{-mc^2/2jT}\left(\frac{2\pi jT}{m}\right)^{-1/2}\{\theta(a-c-b)e^{-m(a-c)^2/2T} + \theta(b-a+c)e^{-m(2b-a+c)^2/2T}\}\left(\frac{2\pi T}{m}\right)^{-1/2} da,$$

where θ is the Heaviside function.

This gives for $j = 1,2,\ldots,(m-2)$

$$\mathbb{E}_z^T\left\{\overset{\sup}{s}\left(z(s) - z\left(\frac{jT}{m}\right)\right) > b\right\} = \int\int dc\, da\, e^{-mc^2/2jT}\left(\frac{2\pi jT}{m}\right)^{-1/2} \times \{e^{-m(a-c)^2/2T}\theta(a-c-b) + e^{-m(2b-a+c)^2/2T}\theta(b-a+c)\} \times \left(\frac{2\pi T}{m}\right)^{-1/2} e^{-ma^2/2(m-j-1)T}\left(\frac{m}{m-j-1}\right)^{1/2}.$$

Using the inequality $\int_{\lambda>b} e^{-\lambda^2/2t_0}\,d\lambda < (t_0/b)e^{-b^2/2t_0}$, then yields for $j = 1,2,\ldots,(m-2)$

$$\mathbb{E}_z^T\left\{\overset{\sup}{s}\left|z(s) - z\left(\frac{jT}{m}\right)\right| > b\right\} \leqslant \frac{4T^{1/2}}{(2\pi)^{1/2}}\frac{1}{b}\frac{1}{(m-j-1)^{1/2}}e^{-mb^2/2T}.$$

Similarly, it can be shown that the above inequality is valid for $j = 0$ and, setting $j = 0$ in rhs for $j = (m-1)$. Now let

$$Q_j^m = \left\{z \in C_0[0,T] \,\middle|\, \overset{\sup}{s}\left|z(s) - z\left(\frac{jT}{m}\right)\right| > \frac{b}{2}\right\}.$$

From above

$$\mathbb{E}_z^T\left\{\bigcup_{j=0}^{m-1} Q_j^m\right\} \leqslant \frac{16T^{1/2}}{b(2\pi)^{1/2}} e^{-mb^2/8T} \sum_{j=1}^{m-1} \frac{1}{j^{1/2}} \leqslant \frac{32}{(2\pi)^{1/2}} \frac{(mT)^{1/2}}{b} e^{-mb^2/8T},$$

the last step following by adding the inequalities $j^{-1/2} < \int_{(j-1)}^{j} x^{-1/2}\,dx$, for $j = 2,\ldots,(m-1)$. But, if $x \notin \bigcup_{j=0}^{m-1} Q_j^m$, $\sup_{jT/m \leqslant s \leqslant (j+1)T/m} |z(s) - z(jT/m)| < \frac{1}{2}b$, for $j = 0,1,2,\ldots,(m-1)$, and so by the last lemma $\|z - z^m(\cdot)\| < b$ and so

$$\mathbb{E}_z^T\{\|z - z^m(\cdot)\| \geqslant b\} < \frac{32}{(2\pi)^{1/2}} \frac{(mT)^{1/2}}{b} e^{-mb^2/8T}. \quad \square$$

Lemma 9: For $k, l \in \mathbb{R}$, $\mathbb{E}_z^T\{\exp\{k\|z\|^2 + l\|z\|\}\} < \infty$, for $k < \gamma/T$, γ being as given in Lemma 6.

Proof: See the proof of Lemma 9 in Ref. 1b. $\square$

As has already been stated, for each $z \in C_0[0,T]$, $z^n(\cdot)$ denotes its polygonalization.

$$z^n(s) = z\left(\frac{jT}{n}\right) + \left(s - \frac{jT}{n}\right)\left[z\left(\frac{\overline{j+1}T}{n}\right) - z\left(\frac{jT}{n}\right)\right]\frac{n}{T}, \qquad \frac{jT}{n} \leqslant s \leqslant \frac{(j+1)T}{n},$$

$j = 0,1,2,\ldots,(n-1)$. We write the associated vector z^n, where $z^n = (z(T/n), z(2T/n),\ldots,z(T))$ and in the following $\|z^n\| = \sup_{j=1,2,\ldots,n} |z_j^n|$, z_j^n being the components of z^n. The next four lemmas are due to Schilder.[1]

Lemma 10 (Schilder): Let A_n be the following $(n \times n)$ tridiagonal matrix

$$n/T \begin{bmatrix} 2 & -1 & 0 & 0 & . & . & . \\ -1 & 2 & -1 & 0 & . & . & . \\ 0 & -1 & 2 & -1 & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & 0 & -1 & 2 & -1 \\ . & . & . & 0 & 0 & -1 & 1 \end{bmatrix}.$$

Then, for s^n an n-dimensional vector, $s^n A_n s^n = \Sigma_{j=1}^n (n/T)(s_j^n - s_{j-1}^n)^2$, $s_0^n = 0$,

$$\int_0^T \left[\frac{ds^n}{d\tau}(\tau)\right]^2 d\tau = s^n A_n s^n \leqslant \int_0^T [s'(\tau)]^2\,d\tau,$$

s being any trajectory in $C_0^*[0,T]$ with polygonalization $s^n(\cdot)$ and associated vector s^n.

Proof: See the proof of Lemma 4 in Ref. 1b. $\square$

Lemma 11 (Schilder): If $z \in C_0^*[0,T]$ then, for $\tau_2 > \tau_1$,

$$\sup_{\tau_1 \leqslant \tau \leqslant \tau_2} [z(\tau) - z(\tau_1)]^2 \leqslant (\tau_2 - \tau_1)\int_{\tau_1}^{\tau_2} [z'(\tau)]^2\,d\tau.$$

Proof: See the proof of Lemma 5 in Ref. 1b. $\square$

Recall that x_1 and x_2 are the global maxima and set $S_1 = \{z \in C_0[0,T] \mid \|z^n(\cdot) - x_2\| \geqslant \|z^n(\cdot) - x_1\|\}$, $S_2 = \{z \in C_0[0,T] \mid \|z^n(\cdot) - x_1\| > \|z^n(\cdot) - x_2\|\}$, so that, for sufficiently large n, $x_1 \in S_1$, $x_2 \in S_2$, and $\chi_{S_1} + \chi_{S_2} = 1$, χ being the characteristic function. We temporarily defer specifying n, although we shall assume in the following that n is so large that $x_1 \in S_1$ and $x_2 \in S_2$. We refer to this as the condition $*$ on n. Let $A_{i\delta} = \{z \in S_i \mid \|z - x_i\| \geqslant \delta\}$, for $i = 1,2$, and let

$$A_i(\delta) = \sup_{z \in A_{i\delta}} \left\{F(z) - 2^{-1}\int_0^T [\dot{z}(\tau)]^2\,d\tau\right\},$$

$\delta > 0$. Then the following lemma is valid.

Lemma 12: Suppose $F(z)$ satisfies the conditions given in Theorem 1 then $A_i(\delta) < 0$, for $\delta > 0$, $i = 1,2$.

Proof: The proof is an easy generalization of Schilder's original Lemma 6 in Ref. 1b. We only prove the result in the case $i = 1$. Suppose the result is not true; then there is a sequence $\{z_m\} \subset A_{i\delta}$ such that

$$\lim_{m \to 0} \left\{F(z_m) - 2^{-1}\int_0^T [\dot{z}_m(\tau)]^2\,d\tau\right\} = 0.$$

We show that under the above assumptions the sequence $\{z_m\}$ has a subsequence $\{z_m^*\}$, uniformly convergent to x_1^*, in $C_0^*[0,T]$, $x_1^* \neq x_1$ or x_2, and yet $F(x_1^*) - 2^{-1}\int_0^T [\dot{x}_1^*(\tau)]^2\,d\tau \geqslant 0$. This is contrary to the hypothesis of Theorem 1.

It can clearly be assumed that, for all m,

$$-1 \leqslant F(z_m) - 2^{-1}\int_0^T [\dot{z}_m(\tau)]^2\,d\tau.$$

From the last lemma and the second condition on $F(z)$ it follows that

$$\int_0^T [\dot{z}_m(\tau)]^2 d\tau \leqslant 4(L_1 + 1).$$

From the last lemma again

$$\sup_{\tau_1 \leqslant \tau \leqslant \tau_2} |z_m(\tau) - z_m(\tau_1)| \leqslant \left[(\tau_2 - \tau_1)\int_0^T [\dot{z}_m(\tau)]^2\,d\tau\right]^{1/2} \leqslant 2[(\tau_2 - \tau_1)(L_1 + 1)]^{1/2}.$$

Therefore, the sequence $\{z_m\}$ is equicontinuous and bounded. By Ascoli's theorem there exists $x_1^* \in C_0^*[0,T]$ and a subsequence $\{z_m^*\}$ such that x_1^* is the uniform limit of $\{z_m^*\}$. Moreover $x_1^* \in \overline{A}_{1\delta}$ entails $x_1^* \neq x_1$ or x_2. From the inequality

$$-1 \leqslant F(z_m) - 2^{-1}\int_0^T [\dot{z}_m(\tau)]^2\,d\tau$$

and Conditions 2 and 3 on F it follows that F is continuous at x_1^* and $F(x_1^*) \neq -\infty$. Since $\int_0^T [\dot{z}(\tau)]^2 d\tau$ is lower semicontinuous, it follows that

$$F(x_1^*) - 2^{-1}\int_0^T [\dot{x}_1^*(\tau)]^2 d\tau \geqslant \liminf\left\{F(z_m^*) - 2^{-1}\int_0^T [\dot{z}_m^*(\tau)]^2\,d\tau\right\} \geqslant 0. \quad \square$$

Lemma 13: If $s_n(\cdot) \in S_i$ and $\|s^n(\cdot) - x_i^n(\ )\| \geqslant \omega$ and $\omega - \rho_n^i > 0$, then $F(s^n(\cdot)) - 2^{-1}s^n A_n s^n \leqslant A_i(\omega - \rho_n^i)$, where $\rho_n^i = \|x_i - x_i^n(\cdot)\|$, $x_i^n(\cdot)$ being the polygonalization of the unique maximizing paths x_i of $\{F(z) - 2^{-1}\int_0^T [\dot{z}(\tau)]^2\,d\tau\}$; for $i = 1,2$.

Proof: The proof follows easily from the last lemma and the proof of Lemma 12 in Ref. 1a. $\square$

Lemma 14: Let A_n be the matrix defined in Lemma 10. Then if w^n is any vector in $\mathbb{R}^n$,

$$w^n A_n w^n \geqslant T^{-1}\|w^n\|^2,$$

$\|\ \|$ being the sup-norm.

Proof: See the proof of Lemma 13 in Ref. 1a. $\square$

We now come to one of our most important lemmas,

which is based on Lemma 14 of Ref. 1a.

Lemma 15: Let F satisfy the conditions in Theorem 1 and let $\delta > 0$. Then, for sufficiently small λ,

$$I^{(i)}(\lambda) = \mathbf{E}_z^T\left\{\left[1 - \chi\left(\frac{\delta}{\lambda}, \frac{x_i}{\lambda}, z\right)\right]\chi_{S_i}(\lambda z)\exp\{\lambda^{-2}F(\lambda z)\}\right\} = O(\exp(\alpha_i \lambda^{-2})),$$

for some $\alpha_i < 0$, for $i = 1,2$.

Proof: Since $x_i(\tau)$ is continuous on $[0, T]$ we have $\lim_{n\to\infty} \|x_i - x_i^n(\cdot)\| = 0$ and from Lemma 10 we have

$$\lim_{n\to\infty} x_i^n A_n x_i^n \leqslant \int_0^T \left[\frac{dx_i}{d\tau}(\tau)\right]^2 d\tau,$$

for $i = 1, 2$. Therefore, both sequences $\{\|x_i^n(\cdot)\|\}$ and $\{x_i^n A_n x_i^n\}$ are bounded for $i = 1,2$. Hence $\exists$ a positive constant c sufficiently large so that, for all n,

$$L_1/c + 2L_2\|x_i^n\|(T/c)^{1/2} + L_2\|x_i^n\|^2/c + (x_i^n A_n x_i^n/c)^{1/2} < \tfrac{1}{8}$$

and, simultaneously,

$$-A_i(\delta)/c < \tfrac{1}{8}$$

for $i = 1, 2$.

From the continuity assumptions on F we can find an $\eta < \delta/4$ such that, for $\|z\| < a = [T(L_1 + 1)]^{1/2}$ and $\|z - y\| < \eta$,

$$F(z) - F(y) \leqslant D_i = -2\left(\frac{\delta}{2}\right)^2 A_i\left(\frac{\delta}{2}\right)\Big/Tc > 0, \quad \text{for } i = 1,2.$$

Finally, we choose n so large that condition * holds, that

$$L_1 - n\eta^2/16T < -1$$

and so large that $\rho_n^i = \|x_i - x_i^n(\cdot)\| < \delta/8$, for $i = 1,2$. In what follows we keep these choices of n, η, and c fixed. We only prove the result for $I^{(1)}(\lambda)$; the proof of the result for $I^{(2)}(\lambda)$ is similar. We observe that $I^{(1)}(\lambda) = I_2(\lambda) + I_3(\lambda) + I_4(\lambda)$, where

$$I_2(\lambda) = \mathbf{E}_z^T\left\{\left[1 - \chi\left(\frac{\delta}{\lambda}, \frac{x_1}{\lambda}, z\right)\right]\left[1 - H\left(\frac{\eta}{\lambda}, n, z\right)\right] \times \chi_{S_i}(\lambda z)\exp\{\lambda^{-2}F(\lambda z)\}\right\},$$

$$I_3(\lambda) = \mathbf{E}_z^T\left\{\left[1 - \chi\left(\frac{\delta}{\lambda}, \frac{x_1}{\lambda}, z\right)\right]H\left(\frac{\eta}{\lambda}, n, z\right)\chi\left(\frac{a}{\lambda}, 0, z\right) \times \chi_{S_i}(\lambda z)\exp\{\lambda^{-2}F(\lambda z)\}\right\},$$

$$I_4(\lambda) = \mathbf{E}_z^T\left\{\left[1 - \chi\left(\frac{\delta}{\lambda}, \frac{x_1}{\lambda}, z\right)\right]H\left(\frac{\eta}{\lambda}, n, z\right) \left[1 - \chi\left(\frac{a}{\lambda}, 0, z\right)\right]\chi_{S_i}(\lambda z)\exp\{\lambda^{-2}F(\lambda z)\}\right\},$$

$H(\eta/\lambda, n, z)$ is the characteristic function of the set $\{z \in C_0[0,T] \mid \|z - z^n(\cdot)\| \leqslant \eta/\lambda\}$. We show that $I_2(\lambda), I_3(\lambda)$, and $I_4(\lambda)$ are each $O(\exp\{\alpha_1\lambda^{-2}\})$, for some $\alpha_1 < 0$, just as in Lemma 14 of Ref. 1a.

In $I_2(\lambda)$ [and $I_4(\lambda)$] we simply observe that $\chi_{S_i}(\lambda z) \leqslant 1$ and $[1 - \chi(\delta/\lambda, x_1/\lambda, z)] \leqslant 1$ [and $H(\eta/\lambda, n, z) \leqslant 1$] so that

$$I_2(\lambda) \leqslant \mathbf{E}_z^T\{[1 - H(\eta/\lambda, n, z)]\exp\{\lambda^{-2}F(\lambda z)\}\}$$

and

$$I_4(\lambda) \leqslant \mathbf{E}_z^T\{[1 - \chi(a/\lambda, 0, z)]\exp\{\lambda^{-2}F(\lambda z)\}\}.$$

We can now argue in exactly the same way as we did in Lemma 14 of Ref. 1a. Using the Cauchy–Schwarz inequality and the improved estimates of Lemmas 6 and 8, together with the facts that here $(L_1 - n\eta^2/16T) < -1$ and $a = [T(L_1 + 1)]^{1/2}$, yields $I_2(\lambda) = O(\exp(-\lambda^{-2}))$ and $I_4(\lambda) = O(\exp(-\lambda^{-2}))$.

The tricky term is $I_3(\lambda)$. It is here that we require the presence of the χ_{S_i} term. Repeating the argument of Lemma 14 we obtain, using Lemmas 12 and 13 and the presence of the χ_{S_i} term,

$$I_3(\lambda) \leqslant v^{-(n-1)}\exp(D_1\lambda^{-2})\mathbf{E}_z^T \times\{[1 - \chi((\delta - \eta - \rho_n^1)\lambda^{-1}v, 0, z)]\},$$

where $v = [-2A_1(\delta - \eta - 2\rho_n^1)/c]^{1/2}$, $\rho_n^1 = \|x_1 - x_1^n(\cdot)\|$. We now use Lemma 6 to deduce that, for an absolute constant K, for sufficiently small λ,

$$\mathbf{E}_z^T\{1 - \chi((\delta - \eta - \rho_n^1)\lambda^{-1}v, 0, z)\} \leqslant K\exp\{-2(\delta - \eta - \rho_n^1)^2 v^2/\lambda^2 T\}.$$

Finally then, since $\eta < \delta/4$, $\rho_n^1 < \delta/8$, and $A_1(\cdot)$ is a negative decreasing function, for K' an absolute constant,

$$I_3(\lambda) \leqslant K'\exp\{\lambda^{-2}[D_1 + 4(\delta - \eta - \rho_n^1)^2 A_1(\delta - \eta - \rho_n^1)/Tc]\} \leqslant K'\exp\{\lambda^{-2}[D_1 + 4(\delta/2)^2 A_1(\delta/2)/Tc]\} = K'\exp\{\lambda^{-2}\delta^2 A_1(\delta/2)/2Tc\}.$$

This completes the proof if we take $\alpha_1 = \max\{-1, \delta^2 A_1(\delta/2)/2Tc\}$. □

Lemma 16: If $\dot{x}_i(\cdot)$ is of bounded absolute variation on $[0,T]$ and if $\int_0^T f(\tau)y(\tau)d\tau - \int_0^T \dot{x}_i(\tau)\,dy(\tau) = 0$, $f \in L^2[0,T]$, for all $y \in C_0^*[0, T]$, then

$$\int_0^T f(\tau)y(\tau)\,d\tau = \int_0^T \dot{x}_i(\tau)\,dy(\tau)$$

for a.e. $y \in C_0[0,T]$; $i = 1,2$.

Proof: See the proof of Lemma 15 in Ref. 1a. □

B. Proof of Theorem 1

Choose a $\delta > 0$, $\delta < \|x_1 - x_2\|/4$, such that the hypotheses on F hold. Then, following the argument given in Ref. 1a,

$$\mathbf{E}_z^T\{\exp\{\lambda^{-2}F(\lambda z)\}\} = h_1(\lambda) + h_2(\lambda),$$

where

$$h_1(\lambda) = \mathbf{E}_z^T\{\chi(\delta/\lambda, x_1/\lambda, z)\chi_{S_1}(\lambda z)\exp\{\lambda^{-2}F(z)\}\} + \mathbf{E}_z^T\{\chi(\delta/\lambda, x_2/\lambda, z)\chi_{S_2}(\lambda z)\exp\{\lambda^{-2}F(z)\}\},$$

$$h_2(\lambda) = \mathbf{E}_z^T\{[1 - \chi(\delta/\lambda, x_1/\lambda, z)]\chi_{S_1}(\lambda z)\exp\{\lambda^{-2}F(\lambda z)\}\} + \mathbf{E}_z^T\{[1 - \chi(\delta/\lambda, x_2/\lambda, z)]\chi_{S_2}(\lambda z)\exp\{\lambda^{-2}F(\lambda z)\}\}.$$

From the penultimate lemma we deduce that $h_2(\lambda) = O(\exp\{\alpha\lambda^{-2}\})$, $\alpha = \max\{\alpha_1, \alpha_2\} < 0$ and so for any integer m, $h_2(\lambda) = O(\lambda^{m-2})$.

We will now consider the first term in $h_1(\lambda)$. The necessary results for the second term follow similarly. Let

$h_1^1(\lambda) = \mathbb{E}_z^T\{\chi(\delta/\lambda, x_1/\lambda, z)\chi_{S_1}(\lambda z)\exp\{\lambda^{-2}F(\lambda z)\}\}$,
then using a Cameron–Martin transformation we arrive at

$$h_1^1(\lambda) = \exp\left\{-2^{-1}\lambda^{-2}\int_0^T [x_1'(\tau)]^2\,d\tau\right\} \times \mathbb{E}_y^T\left\{\chi(\delta/\lambda, 0, y)\chi_{S_1}(\lambda y + x_1) \times \exp\left\{-\lambda^{-1}\int_0^T x_1'(\tau)\,dy(\tau) + \lambda^{-2}F(\lambda y + x_1)\right\}\right\}.$$

From Taylor's theorem for functionals we may write

$$\begin{aligned} F(\lambda y + x_1) &= F(x_1) + \lambda DF(x_1)y + \lambda^2 D^2F(x_1)(y, y)/2 + k_3^1(\lambda y) \\ &= f_0^1(0) + \lambda f_1^1(0)y + \lambda^2 f_2^1(0)y^2 + k_3^1(\lambda y), \end{aligned}$$

say, where $|k_3^1(\lambda y)| = O(\lambda^3)$ if $\|\lambda y\| < \delta$. Therefore,

$$h_1^1(\lambda) = \exp\left\{-2^{-1}\lambda^{-2}\int_0^T [x_1'(\tau)]^2\,d\tau + \lambda^{-2}f_0^1(0)\right\} \times \mathbb{E}_y^T\{\chi(\delta/\lambda, 0, y)\chi_{S_1}(\lambda y + x_1)\} \times \exp\left\{-\lambda^{-1}\int_0^T x_1'(\tau)\,dy(\tau) + \lambda^{-1}f_1^1(0)y\right\} \times \exp\{f_2^1(0)y^2 + \lambda^{-2}k_3^1(\lambda y)\}\Big\}.$$

By hypothesis $\{F(z) - 2^{-1}\int_0^T [z'(\tau)]^2 d\tau\}$ has a maximum of 0 at x_1 over $C_0^*[0, T]$, so it and its first Frechet derivative are zero at x_1. Thus,

$$f_1^1(0)y - \int_0^T x_1'(\tau)\,dy(\tau) = 0$$

for $y \in C_0^*[0, T]$ and so, by the last lemma,

$$h_1^1(\lambda) = \mathbb{E}_z^T\{\chi(\delta/\lambda, 0, z)\chi_{S_1}(\lambda y + x_1)\} \times \exp\{f_2^1(0)y^2 + \lambda^{-2}k_3^1(\lambda y)\}.$$

The Taylor expansion of $\exp\{x\}$ reads

$$\exp\{x\} = \sum_{i=0}^{n-1} x^i/i! + R_n(x),$$

where

$$R_n(x) \leqslant \begin{cases} x^n/n!\exp\{x\} & \text{if } x \geqslant 0, \\ |x|^n/n! & \text{if } x < 0. \end{cases}$$

We may now write $h_1^1(\lambda)$ in the form

$$h_1^1(\lambda) = \sum_{i=0}^{m-3} (i!)^{-1}\mathbb{E}_z^T\{\chi(\delta/\lambda, 0, z)\chi_{S_1}(\lambda z + x_1) \times \exp\{f_2^1(0)z^2\}[\lambda^{-2}k_3^1(\lambda z)]^i\} + J_{m-2}(\lambda)$$

where, denoting the characteristic function of the set $\{z \in C_0[0,T] | k_3^1(\lambda z) > 0\}$ by $B^1(\lambda, z)$,

$$\begin{aligned} |J_{m-2}(\lambda)| \leqslant{}& ((m-2)!)^{-1}\mathbb{E}_z^T\{\chi(\delta/\lambda, 0, z)|\lambda^{-2}k_3^1(\lambda z)|^{m-2} \\ &\times \chi_{S_1}(\lambda z + x_1)\exp\{f_2^1(0)z^2 + \lambda^{-2}k_3^1(\lambda z)\}B^1(\lambda, z)\} \\ &+ ((m-2)!)^{-1}\mathbb{E}_z^T\{\chi(\delta/\lambda, 0, z)|\lambda^{-2}k_3^1(\lambda z)|^{m-2} \\ &\times \chi_{S_1}(\lambda z + x_1)\exp\{f_2^1(0)z^2\}[1 - B^1(\lambda, z)]\}. \end{aligned}$$

From Taylor's theorem for functionals it follows that, if $\|\lambda y\| < \delta$, then

$$\begin{aligned} &\lambda^2 f_2^1(0)y^2 + k_3^1(\lambda y) \\ &\quad = k_2^1(\lambda y) = \lambda^2 f_2^1(\eta)y^2; \\ &\qquad\qquad \eta \in C_0[0, T] \text{ with } \|\eta\| < \delta, \end{aligned}$$

where by hypothesis $|k_3^1(\lambda y)| \leqslant C_3\lambda^3\|y\|^3$, C_3 being a constant. Thus,

$$\begin{aligned} &|J_{m-2}(\lambda)| \\ &\quad \leqslant ((m-2)!)^{-1}\mathbb{E}_z^T\{\chi(\delta/\lambda, 0, z)(C_3\lambda)^{m-2}\|z\|^{3(m-2)} \\ &\qquad \times \chi_{S_1}(\lambda z + x_1)\exp\{f_2^1(\eta)z^2\}B^1(\lambda, z)\} \\ &\qquad + ((m-2)!)^{-1}\mathbb{E}_z^T\{\chi(\delta/\lambda, 0, z)(C_3\lambda)^{m-2}\|z\|^{3(m-2)} \\ &\qquad \times \chi_{S_1}(\lambda z + x_1)\exp\{f_2^1(0)z^2\}[1 - B^1(\lambda, z)]\}. \end{aligned}$$

By using the Cauchy–Schwarz inequality and Condition 5 of Theorem 1 we have $J_{m-2}(\lambda) = O(\lambda^{m-2})$. We have now proved that

$$h_1^1(\lambda) = \sum_{i=0}^{m-3} (i!)^{-1}\mathbb{E}_z^T\{\chi(\delta/\lambda, 0, z) \times \exp\{f_2^1(0)z^2\}[\lambda^{-2}k_3^1(\lambda z)]^i\} + O(\lambda^{m-2}),$$

using $\chi_{S_1}(\lambda z + x_1) = 1$, if $\|\lambda z\| \leqslant \delta$, for sufficiently large n.

However, $k_3^1(\lambda z) = \lambda^3 f_3^1(0)z^3 + \cdots + \lambda^{m-1}f_{m-1}^1(0)z^{m-1} + k_m^1(\lambda z)$, where $\lambda^{-2}k_n^1(\lambda z) = O(\lambda^{n-2}\|z\|^n)$, for $\|\lambda z\| < \delta$; expanding by the binomial theorem, therefore, gives, using Condition 5 of Theorem 1 and Hölder's inequality,

$$h_1^1(\lambda) = \sum_{i=0}^{m-3} (i!)^{-1}\mathbb{E}_z^T\{\chi(\delta/\lambda, 0, z)\exp\{f_2^1(0)z^2\}[\lambda f_3^1(0)z^3 + \cdots + \lambda^{m-3}f_{m-1}^1(0)z^{m-1}]^i\} = 0(\lambda^{m-2}).$$

It can be seen from the Hölder inequality, Lemmas 6 and 9, and Condition 5 of Theorem 1, that for sufficiently small λ,

$$\begin{aligned} &\sum_{i=0}^{m-3} (i!)^{-1}\mathbb{E}_z^T\{[1 - \chi(\delta/\lambda, 0, z)] \\ &\quad \times \exp\{f_2^1(0)z^2\}[\lambda f_3^1(0)z^3 + \cdots \\ &\quad + \lambda^{m-3}f_{m-1}^1(0)z^{m-1}]^i\} \\ &\qquad = O(P(\lambda)\exp\{\beta\lambda^{-2}\}) = O(\lambda^{m-2}), \end{aligned}$$

where P is a polynomial and β is a negative constant. Replacing χ by $[1 - (1 - \chi)]$, we finally obtain

$$h_1^1(\lambda) = \sum_{i=0}^{m-3} (i!)^{-1}\mathbb{E}_z^T\{\exp\{f_2^1(0)z^2\}[\lambda f_3^1(0)z^3 + \cdots + \lambda^{m-3}f_{m-1}^1(0)z^{m-1}]^i\} + O(\lambda^{m-2}),$$

so that

$$h_1^1(\lambda) = \Gamma_0^1 + \lambda\Gamma_1^1 + \cdots + \lambda^{m-3}\Gamma_{m-3}^1 + O(\lambda^{m-2}),$$

where the Γ_i^1 depend only on F and its Frechet derivatives at

x_1, for $i = 1,2,...,m-3$.

Repeating the above argument for $[h_1(\lambda) - h_1^1(\lambda)]$ we may finally write

$$h_1(\lambda) = \Gamma_0 + \lambda\Gamma_1 + \cdots + \lambda^{m-3}\Gamma_{m-3} + O(\lambda^{m-2}),$$

as $\lambda \to 0$, where the Γ_i depend only on F and its Frechet derivatives evaluated at x_1 and x_2. □

Corollary: For $\lambda > 0$ and F satisfying the hypotheses of Theorem 1, with $D^2F(x_1) \equiv D^2F(x_2)$,

$$\mathbb{E}_z^T\{\exp\{\lambda^{-2}F(\lambda z)\}\} = 2\mathbb{E}_z^T\{\exp\{D^2F(x_1)(z,z)/2\}\} + O(\lambda).$$

Proof: By observation. □

4. CONCLUSION

We have seen how Schilder's rigorous results on the Laplace expansions of Wiener integrals can be extended to conditional Wiener integrals with functional integrands having a finite number of nondegenerate global maxima. We have applied these results to the derivation of Bender–Wu type formulas for the large order behavior of the perturbation series of the x^{2N}-anharmonic oscillator ($N \geqslant 2$), in a scheme which (modulo the commutativity of two limits) leads to detailed information about the asymptotic behavior. The rapidly varying terms in this large order behavior agree exactly with Bender and Wu's remarkable formula. The slowly varying term, however, differs slightly from the predicted behavior of the Bender–Wu formula. (The problem here presumably is the noncommutativity of the two limits). Nevertheless, our results do give an analog of Simon's rigorous results for the x^{2N}-anharmonic oscillator, for $N \geqslant 2$. A Feynman graph analysis of higher order terms in the asymptotic expansion is possible. This is currently being attempted to compare with the formal Feynman graph analysis carried out with collective coordinate methods. Generalizations to field theory are also currently being investigated.

ACKNOWLEDGMENTS

It is a pleasure to thank Peter Baxendale for suggesting the simple method of proof of Lemma 6 and David Elworthy for a helpful suggestion. One of us (A.T.) is deeply indebted to CNRS–CPT, Marseille, for hospitality during a one month stay in July 1981, where this work was completed. We are also very grateful to Professors M. Sirugue, M. Sirugue-Collin, of CPT Marseille, for stimulating discussions of the above material which resulted in some improvements. Finally, it is a pleasure to acknowledge the influence on our work of the earlier results of Professors E. Brezin *et al.* and B. Simon.

APPENDIX: THE PROOF OF PROPOSITION 1

The self-adjointness and discreteness of the spectrum are well known.[13a,b] We now show by a standard argument[5b] that the eigenfunctions are not degenerate and bounded. We consider the eigenvalue equation

$$\left[2^{-1}\left(-\frac{d^2}{dx^2} + x^2\right) + \beta x^{2N}\right]u(x) = \left[-2^{-1}\frac{d^2}{dx^2} + V(x)\right]u(x) = Eu(x),$$

$E \in \sigma[H(\beta)], H(\beta) = 2^{-1}(-d^2/dx^2 + x^2) + \beta x^{2N}$. Without loss we assume that u is real valued. Firstly, observe that if $u \in L^2(\mathbb{R})$ is a solution of above equation necessarily $u' \in L^2(\mathbb{R})$. To see this we multiply the above equation by $u(x)$ and integrate by parts from $x = a$ to $x = b$, giving

$$u(b)u'(b) - u(a)u'(a) = \int_a^b [u'(x)]^2\,dx + 2\int_a^b [V(x) - E]u^2(x)\,dx \geqslant \int_a^b [u'(x)]^2\,dx - 2E\int_a^b u^2(x)\,dx, \qquad u'(x) = \frac{du}{dx}(x).$$

For $u \in L^2(\mathbb{R})$, letting $b \to \infty$, we see that the divergence of $\int_a^\infty [u'(x)]^2\,dx$ would imply that, for all sufficiently large b, $(d/db)(2^{-1}u^2(b)) > 0$, contradicting the hypothesis that $u \in L^2(\mathbb{R})$, so $u' \in L^2(\mathbb{R})$. Now let $u, v \in L^2(\mathbb{R})$ be two linearly independent solutions of the above eigenvalue equation, real valued and normalized so that $W(u,v)(x) = u(x)v'(x) - u'(x)v(x) = 1$. Then the Cauchy–Schwarz inequality shows that the divergence of $\int_{-\infty}^{\infty} W(u,v)(x)dx$ is not consistent with the hypotheses $u, v, u', v' \in L^2(\mathbb{R})$. Hence at most one of u or $v \in L^2(\mathbb{R})$. What is more the inequality

$$|u^2(b) - u^2(0)| = \left|2\int_0^b u(x)u'(x)\,dx\right| \leqslant 2\|u\|_2\|u'\|_2,$$

proves that any L^2-eigenfunction satisfies $\sup_{x\in\mathbb{R}}|u(x)| < \infty$. We denote orthonormal eigenfunctions and eigenvalues of $H(\beta)$ by ϕ_n and $E_n(\beta)$, $n = 0,1,2,\ldots$.

We now establish that $G(0,0,T) = \Sigma_{n=0}^{\infty} e^{-E_n(\beta)T}|\phi_n(0)|^2$; the more general result involving x's and y's follows in almost exactly the same way. For the Brownian bridge α[5a]

$$G(x,y,T) = (2\pi T)^{-1/2}\exp\{-|x-y|^2/2T\} \times \mathbb{E}_\alpha\left[\exp\left\{-\int_0^T V\left[\left(1 - \frac{s}{T}\right)x + \frac{s}{T}y + \sqrt{T}\alpha\left(\frac{s}{T}\right)\right]ds\right\}\right].$$

Evidently then, for $T > 0$, $|G(x,y,T)| \leqslant (2\pi T)^{-1/2}$ and by dominated convergence, we see $G(\delta x, \delta y, T) \to G(0,0,T)$, for each $x, y \in \mathbb{R}$, as $\delta \to 0$. Let $\theta \in \mathscr{S}(\mathbb{R})$ be such that $\theta(x) \geqslant 0$ with $\int\theta(x)\,dx = 1$ and define $\theta_\delta(x) = \delta^{-1}\theta(\delta^{-1}x)$, $x \in \mathbb{R}$, $\theta_\delta \in \mathscr{S}(\mathbb{R})$. Changing integration variables, using dominated convergence for Lebesque measure, for $T > 0$, gives

$$\int G(x,y,T)\theta_\delta(x)\theta_\delta(y)\,dx\,dy = \int G(\delta x, \delta y, T)\theta(x)\theta(y)\,dx\,dy \to G(0,0,T), \quad \text{as } \delta \to 0.$$

The proof will therefore be completed if we prove that for

$T>0$

$$\int G(x,y,T)\theta_\delta(x)\theta_\delta(y)\,dx\,dy$$

$$\to \sum_{n=0}^{\infty} e^{-E_n(\beta)T}|\phi_n(0)|^2, \quad \text{as } \delta\to 0,$$

ϕ_n being orthonormal eigenfunctions of $H(\beta)$ with eigenvalues $E_n(\beta)$. This we now prove by another application of the dominated convergence theorem. Let δ_n be a Dirac measure concentrated at $x=n$, $n=0,1,2,3,\ldots$. Consider $\mu=\Sigma_{n=0}^{\infty}e^{-TE_n(\beta)}\delta_n$, for $T>0$. Then, since for $T>0, \beta>0$, $e^{-TH(\beta)}\leqslant e^{-TH(0)}$, $e^{-TH(\beta)}$ is trace-class and so μ is a measure on $\mathbb{Z}$ (with $2^{\mathbb{Z}}$ as σ-field), with $\int_{\mathbb{Z}} f(x)d\mu(x)=\Sigma_{n=0}^{\infty} f(n)e^{-TE_n(\beta)}$, for bounded continuous $f\colon \mathbb{Z}\to\mathbb{R}$. We now consider the particular function $f^\delta\colon \mathbb{Z}\to\mathbb{R}$ defined by $f^\delta(x)=|(\phi_n,\theta_\delta)|^2$, when $x=n$, $n=0,1,2,\ldots$.

Then, using the above inequality

$$|f^\delta(n)|\leqslant\|\phi_n\|_\infty^2\leqslant 2\|\phi_n\|_2\|\phi_n'\|_2\leqslant 2\sqrt{2}E_n^{1/2}(\beta),$$

where in the last step we are using

$$\|\phi_n'\|^2=\int \phi_n'^2(x)\,dx\leqslant 2E_n(\beta).$$

Define now $g(\cdot)$ by $g(x)=2\sqrt{2}E_n^{1/2}(\beta)$, for $x=n$, $n=0,1,2,\ldots$. Then, since $E_n\to\infty$ as $n\to\infty$ implies $E_n^{1/2}<e^{\epsilon E_n}$, for sufficiently large n, arbitrary $\epsilon>0$,

$$\int_{\mathbb{Z}} g(x)d\mu(x)=2\sqrt{2}\sum_0^{\infty}E_n^{1/2}(\beta)e^{-TE_n(\beta)}<\infty, \quad T>0.$$

But $f^\delta(n)\to|\phi_n(0)|^2=f^0(n)$, as $\delta\to 0$, $n=0,1,2,\ldots$, since by dominated convergence $(\theta_\delta,\phi_n)=\int\phi_n(\delta x)\theta(x)\,dx\to\phi_n(0)$, as $\delta\to 0$. Hence by dominated convergence for the measure μ on $\mathbb{Z}$

$$\int f^\delta(x)d\mu(x)\to\int f^0(x)d\mu(x)$$

$$=\sum_{n=0}^{\infty}|\phi_n(0)|^2e^{-E_n(\beta)T}, \quad T>0$$

i.e.,

$$\sum_{n=0}^{\infty}e^{-TE_n(\beta)}|(\theta_\delta,\phi_n)|^2\to\sum_{n=0}^{\infty}e^{-TE_n(\beta)}|\phi_n(0)|^2, \quad T>0$$

as required. □

[1]a) I. Davies and A. Truman, "Laplace asymptotic expansions of conditional Wiener integrals and generalized Mehler kernel formulas," J. Math. Phys. **23**, 2059 (1982); b) M. Schilder, Trans. Am. Math. Soc. **125**, 63–85 (1965).

[2]For general Gaussian measures see also R. S. Ellis and J. R. Rosen, Bull. Am. Math. Soc. **3**, 705–9 (1980); "Asymptotic analysis of Gaussian integrals, I: Isolated minimal points," to appear in Trans. Am. Math. Soc.; "Asymptotic analysis of Gaussian integrals, II; Manifold of minimal points," Commun. Math. Phys. **82**, 153–81 (1981).

[3]C. M. Bender and T. T. Wu, "Anharmonic Oscillator," Phys. Rev. **184**, 1231–60 (1969); "Large order behavior of Perturbation Theory," Phys. Rev. Lett. **27**, 461–5 (1971); "Anharmonic Oscillator II. A study of Perturbation Theory in large order," Phys. Rev. D **7** 1620–36 (1973).

[4]L. N. Lipatov, JETP Lett. **25**, 104–7 (1977). E. Brezin *et al.*, Phys. Rev. D **15**, 1544–57, 1558–64 (1977); G. Auberson and G. Mennessier, Nuovo Cimento **A48**, 1–23 (1978); J. C. Collins and D. E. Soper, Ann. Phys. (NY) **112**, 209–34 (1978); V. R. Figerou, Commun. Math. Phys. **79**, 401–33 (1981).

[5]a) B. Simon, *Functional Integration and Quantum Physics* (Academic, New York, 1979). See pp. 181–98 in particular. b) See also T. Spencer, Commun. Math. Phys. **74**, 273–80 (1980).

[6]Compare with the treatment of Ellis and Rosen in Ref. 2.

[7]P. Baxendale, private communication.

[8]N. I. Akhiezer, *The Calculus of Variations* (Blaisdell, New York, 1962). See Chap. 4.

[9]M. Reed and B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis* (Academic, New York, 1972). See Theorem 1.28.

[10]A. Truman, "The polygonal path formulation of the Feynman Path Integral", *Lecture notes in Physics* **106**, (Springer, Berlin, 1979), pp. 73–102.

[11]K. Ito and H. P. McKean, *Diffusion processes and their sample paths* (Springer, Berlin, 1965); D. Williams *Diffusion, Markov processes, and martingales V. I: Foundations* (Wiley, New York, 1979).

[12]See Theorem 7.9 of Ref. 5a.

[13]a) E. C. Titchmarsh, *Eigenfunction Expansions associated with Second Order differential equations; Part One* (Oxford U.P., Oxford, 1962).
b) J. Glimm and A. Jaffe, *Quantum Physics, a functional integral point of view* (Springer, Berlin, 1981).

Generalized Fokker–Planck equation for non-Markovian processes

V. S. Volkov
Institute of Petrochemical Synthesis, USSR Academy of Sciences, Moscow, 117912, USSR

V. N. Pokrovsky
Altai Polytechnical Institute, Barnaul, 656099, USSR

(Received 24 March 1981; accepted for publication 25 November 1981)

This paper deals with a system of linear non-Markovian–Langevin equations with memory functions that are not constant in time and a nonzero initial instant of time. A set of statistical means, based on the application of a generalized Furutsu–Novikov formula, was used to derive a generalized Fokker–Planck equation corresponding to this system and holding for both long and short instants of time. Considered as an example is the Brownian motion of a particle in a viscoelastic fluid with a particular relaxation time.

PACS numbers: 02.50.Ey

1. INTRODUCTION

At present, problems occur in various branches of physics, that are reduced to Langevin equations with memory, descriptive of a non-Markovian stochastic process. In the linear case, its statistical characteristics can be determined directly by presenting the solution of the initial equations in an explicit form. However, in many cases, a more convenient probabilistic description gives an equation for the distribution function. This is why it is desirable to establish a clearly defined relationship between the Langevin equations with memory and the corresponding equation for the distribution function, similar to the Fokker–Planck (FP) equation for a Markovian process.

In doing so we shall follow the Novikov–Klyatskin–Tatarsky method[1] with a modification enabling its generalization, whereby it becomes applicable to non-Markovian processes as well. This method permits deriving a generalized FP equation directly from non-Markovian–Langevin equations. A specific feature of the derivation that follows is that it is based on a generalized Furutsu–Novikov formula[2] that enables separation of the mean of the product of two functionals from a Gaussian stochastic process. Thus, it becomes possible to establish the necessary fluctuation–dissipation relations without resorting to any additional assumptions that are not contained in the formulations of the initial stochastic equations.

2. A SYSTEM OF STOCHASTIC INTEGRO-DIFFERENTIAL EQUATIONS

Let us consider a system of linear stochastic integro-differential equations for a set of variables $a_\alpha(t)$, $\alpha = 1,2,\ldots,n$, determining the state of a physical system

$$\frac{d}{dt}a_\alpha(t) = -\int_{t_0}^{t}\lambda_{\alpha\beta}(t,s)\cdot a_\beta(s)\,ds + \Phi_\alpha(t), \tag{1}$$

where $\lambda_{\alpha\beta}(t,s)$ are predetermined nonrandom memory functions characterizing the dissipative properties of the system.

Random forces $\Phi_\alpha(t)$ are assumed to be Gaussian with a mean value equal to zero. In this case, their statistical properties are fully determined by correlation functions

$$K_{\alpha\beta}(t,s) = \langle \Phi_\alpha(t)\Phi_\beta(s)\rangle.$$

The angle brackets indicate averaging over an assembly of realizations of the random forces.

The system of equations (1) together with the above assumptions as to the statistical properties of random forces determines a multidimensional nonstationary Gaussian non-Markovian process starting at instant t_0. Its explicit solution satisfying initial condition $\mathbf{a}(t_0) = \mathbf{a}_0$ is essentially a linear functional of random forces

$$a_\alpha(t) = \chi_{\alpha\beta}(t,t_0)\cdot a_{0\beta} + \int_{t_0}^{t}\chi_{\alpha\beta}(t,s)\Phi_\beta(s)\,ds. \tag{2}$$

Green's functions $\chi_{\alpha\beta}(t,t_0)$ are defined by a system of equations

$$\frac{\partial}{\partial t}\chi_{\alpha\beta}(t,t_0) = -\int_{t_0}^{t}\lambda_{\alpha\zeta}(t,s)\,\chi_{\zeta\beta}(s,t_0)\,ds,$$
$$\chi_{\alpha\beta}(t_0,t_0) = \delta_{\alpha\beta}. \tag{3}$$

In the case of real physical systems the causality principle must come into play, according to which the reaction cannot precede an action. In the case under consideration, this is equivalent to condition

$$\chi_{\alpha\beta}(t,t') = 0 \qquad \text{if} \quad t' < t_0, t' > t.$$

If the system under investigation is asymptotically stable, Green's functions also meet condition

$$\chi_{\alpha\beta}(t,t') \to 0 \qquad \text{at} \quad t \to \infty.$$

For a complete formulation of the system under consideration in the statistical sense one must establish the corresponding fluctuation–dissipation relations that express the correlation functions of random forces in terms of memory functions and one-time cumulant functions

$$\varphi_{\alpha\gamma}(t,t_0) = \langle a_\alpha(t)a_\gamma(t)\rangle - \langle a_\alpha(t)\rangle\langle a_\gamma(t)\rangle,$$

which are determined by a particular physical situation. The averaging is accomplished, provided at instant t_0 the stochastic process $\mathbf{a}(t)$ takes the value $\mathbf{a}_0$. Proceeding from Eq. (2) we obtain the desired relations

$$\varphi_{\alpha\gamma}(t,t_0) = \int\int_{t_0}^{t}\chi_{\alpha\beta}(t,t_1)\chi_{\gamma\zeta}(t,t_2)\,K_{\beta\zeta}(t_1,t_2)\,dt_1\,dt_2, \tag{4}$$

as well as the necessary conditional means of the type

 0022-2488/83/020267-04$02.50

$$\psi_{\alpha\gamma}(t,t_0) = \langle \dot{a}_\alpha(t) a_\gamma(t)\rangle$$
$$= \int\int_{t_0}^{t} \frac{\partial}{\partial t}\left[\chi_{\alpha\eta}(t,t_1)H(t-t_1)\right] \times \chi_{\gamma\zeta}(t,t_2)K_{\eta\zeta}(t_1,t_2)\,dt_1\,dt_2, \quad (5)$$

where

$$H(t) = \begin{cases} 1 & \text{at } t \geqslant 0 \\ 0 & \text{at } t < 0. \end{cases}$$

Relations (4) and (5) also follow from the generalized Furutsu–Novikov formula[2] that permits separation of the correlations of two linear functionals from a stochastic process with a zero mean since in this case we are dealing with functional derivatives

$$\delta a_\alpha(t)/\delta\Phi_\beta(s) = \chi_{\alpha\beta}(t,s)\,H(t-s).$$

Note that in deriving the fluctuation–dissipation theorem for non-Markovian–Langevin equations with memory functions of the $\lambda_{\alpha\beta}(t,s) = \lambda_{\alpha\beta}(t-s)$ type, Kubo[3] and Henery[4] proceeded from an erroneous assumption that the solution is noncorrelated with the subsequent values of force. In fact, the correlations are in this case different from zero within intervals of the force correlation time order. This assumption holds only when the random force is a δ-correlated noise (has a zero correlation time).

3. NON-MARKOVIAN FOKKER–PLANCK EQUATION

Let us introduce conditional probability density value x of stochastic process $a(t)$ at time t,

$$P(\mathbf{x},t/\mathbf{x}_0,t_0) = \langle \delta(\mathbf{x} - a(t,t_0,\mathbf{x}_0))\rangle, \quad (6)$$

where $a(t,t_0,\mathbf{x}_0)$ is a solution of system (1) at a given value $\mathbf{x}_0$ at time $t = t_0$, corresponding to a definite realization $\Phi(t)$. The averaging is done over the set of all realizations $\{\Phi(t)\}$.

Differentiating Eq. (6) with respect to t, we obtain the following equation for the conditional probability density:

$$\frac{\partial}{\partial t}P(\mathbf{x},t/\mathbf{x}_0,t_0) = -\frac{\partial}{\partial x_\alpha}\left\langle \frac{da_\alpha(t)}{dt}\delta(\mathbf{x}-a(t))\right\rangle. \quad (7)$$

With the aid of the generalized Furutsu–Novikov formula[2] for the mean of the product of the linear $P[\Phi]$ and nonlinear $R[\Phi]$ functionals of the Gaussian stochastic process $\Phi(t)$ with a zero mean value

$$\langle P[\Phi]R[\Phi]\rangle = \langle P[\Phi]\rangle\langle R[\Phi]\rangle + \int\int \left\langle \frac{\delta P[\Phi]}{\delta\Phi(t_1)}\right\rangle \left\langle \frac{\delta R[\Phi]}{\delta\Phi(t_2)}\right\rangle \langle\Phi(t_1)\Phi(t_2)\rangle\,dt_1\,dt_2,$$

we determine

$$\langle \dot{a}_\alpha(t)\delta(\mathbf{x}-a(t))\rangle = \langle \dot{a}_\alpha(t)\rangle P(\mathbf{x},t/\mathbf{x}_0,t_0) + \int\int_{t_0}^{t}\left\langle \frac{\delta\dot{a}_\alpha(t)}{\delta\Phi_\zeta(t_1)}\right\rangle \left\langle \frac{\delta}{\delta\Phi_\eta(t_2)}[\delta(\mathbf{x}-a(t))]\right\rangle \times\langle\Phi_\zeta(t_1)\Phi_\eta(t_2)\rangle\,dt_1\,dt_2, \quad (8)$$

where we have used Eq. (6).

The limits of integration in Eq. (8) are arranged in accordance with the causality condition. Determination of Eq. (6) gives

$$\left\langle \frac{\delta}{\delta\Phi_\eta(t_2)}[\delta(\mathbf{x}-a(t))]\right\rangle = -\frac{\partial}{\partial x_\gamma}\frac{\delta a_\gamma(t)}{\delta\Phi_\eta(t_2)}P(\mathbf{x},t/\mathbf{x}_0,t_0).$$

Given Eq. (2), $\delta a_\gamma(t)/\delta\Phi_\eta(s) = \chi_{\gamma\eta}(t,s)\,H(t-s)$ may be interpreted as the linear response function to the external force $\Phi_\eta(t)$.

Taking the last relations into account, Eq. (7) can be written as follows:

$$\frac{\partial}{\partial t}P(\mathbf{x},t/\mathbf{x}_0,t_0) = -\frac{\partial}{\partial x_\alpha}\left[\frac{\partial m_\alpha(t,t_0)}{\partial t}P(\mathbf{x},t/\mathbf{x}_0,t_0)\right] + \frac{\partial}{\partial x_\alpha}\left[\int\int_{t_0}^{t}\frac{\partial}{\partial t}\left[\chi_{\alpha\zeta}(t,t_1)H(t-t_1)\right] \times\chi_{\gamma\eta}(t,t_2)K_{\zeta\eta}(t_1,t_2)\,dt_1\,dt_2\right] \times\frac{\partial}{\partial x_\gamma}P(\mathbf{x},t/\mathbf{x}_0,t_0). \quad (9)$$

Here, $m_\alpha(t,t_0)$ is the mean value of stochastic process $a_\alpha(t)$, provided at instant t_0 it assumes the value $a_{0\alpha}$. It can be easily seen from Eqs. (1) and (2) that

$$\frac{\partial}{\partial t}m_\alpha(t,t_0) = -\beta_{\alpha\beta}(t,t_0)m_\beta(t,t_0),$$

where

$$\beta_{\alpha\beta}(t,t_0) = -\frac{\partial\chi_{\alpha s}(t,t_0)}{\partial t}\bar{\chi}_{s\beta}^{-1}(t,t_0).$$

In addition, according to the generalized Furutsu–Novikov formula we have

$$m_\beta(t,t_0)P(\mathbf{x},t/\mathbf{x}_0,t_0) = x_\beta P(\mathbf{x},t/\mathbf{x}_0,t_0) + \int\int_{t_0}^{t}\chi_{\beta\zeta}(t,t_1)\chi_{\gamma\eta}(t,t_2)K_{\zeta\eta}(t_1,t_2)\,dt_1\,dt_2 \times\frac{\partial}{\partial x_\gamma}P(\mathbf{x},t/\mathbf{x}_0,t_0).$$

Hence, Eq. (9) can be written as

$$\frac{\partial}{\partial t}P_{t/t_0} = \frac{\partial}{\partial x_\alpha}[\beta_{\alpha\beta}(t,t_0)x_\beta P_{t/t_0}] + \beta_{\alpha\beta}(t,t_0)\int\int_{t_0}^{t}\chi_{\beta\zeta}(t,t_1)\chi_{\gamma\eta}(t,t_2)K_{\zeta\eta}(t_1,t_2)\,dt_1\,dt_2\frac{\partial^2}{\partial x_\alpha x_\gamma}P_{t/t_0} + \int\int_{t_0}^{t}\frac{\partial\chi_{\alpha\zeta}(t,t_1)}{\partial t}\chi_{\gamma\eta}(t,t_2)K_{\zeta\eta}(t_1,t_2)\,dt_1\,dt_2\frac{\partial^2}{\partial x_\alpha\partial x_\gamma}P_{t/t_0} + \int_{t_0}^{t}\chi_{\gamma\eta}(t,t_2)K_{\alpha\eta}(t,t_2)\,dt_2\frac{\partial^2}{\partial x_\alpha\partial x_\gamma}P_{t/t_0}. \quad (10)$$

In the case of a one-dimensional stochastic motion starting at instant $t_0 = 0$ and involving an aftereffect characterized by memory function $\lambda(t,s) = \lambda(t-s)$, Eq. (10) leads to Hänggi's result[5] obtained by a known method.[1]

Finally, using fluctuation–dissipation relations (4) and

(5), we derive a generalized FP equation for the conditional probability density of a multidimensional nonstationary non-Markovian process determined by a system of equations (1), in the form

$$\frac{\partial}{\partial t} P(\mathbf{x},t/\mathbf{x}_0,t_0) = \frac{\partial}{\partial x_\alpha} [\beta_{\alpha e}(t,t_0)x_e\, P(\mathbf{x},t/\mathbf{x}_0,t_0)] + \frac{\partial^2}{\partial x_\alpha \partial x_\gamma} [D_{\alpha\gamma}(t,t_0)\, P(\mathbf{x},t/\mathbf{x}_0,t_0)], \quad (11)$$

where kinetic coefficients $D_{\alpha\beta}(t,t_0)$ are determined, with due account for their subscript symmetry, as

$$2D_{\alpha\gamma}(t,t_0) = \frac{\partial}{\partial t}\varphi_{\alpha\gamma}(t,t_0) + \beta_{\alpha e}(t,t_0)\varphi_{e\gamma}(t,t_0) + \beta_{\gamma e}(t,t_0)\varphi_{e\alpha}(t,t_0).$$

Unlike the usual FP equation, the kinetic coefficients of the non-Markovian FP equation depend on both the current and initial instants of time.

The solution of Eq. (11) must be positive, normalized to unity, and satisfy the condition

$$P(\mathbf{x},t_0/\mathbf{x}_0,t_0) = \delta(\mathbf{x}-\mathbf{x}_0).$$

Since two-time distribution function $\mathcal{f}(\mathbf{x}_0,t_0;\mathbf{x},t)$ is expressed in terms of the conditional probability density and initial distribution $\mathcal{f}(\mathbf{x}_0,t_0)$ as follows,

$$\mathcal{f}(\mathbf{x}_0,t_0;\mathbf{x},t) = \mathcal{f}(\mathbf{x}_0,t_0)\, P(\mathbf{x},t/\mathbf{x}_0,t_0),$$

multiplication of Eq. (11) by $\mathcal{f}(\mathbf{x}_0,t_0)$ gives

$$\frac{\partial}{\partial t}\mathcal{f}(\mathbf{x}_0,t_0;\mathbf{x},t) = \frac{\partial}{\partial x_\alpha} [\beta_{\alpha\beta}(t,t_0)\, x_\beta\, \mathcal{f}(\mathbf{x}_0,t_0;\mathbf{x},t)] + \frac{\partial^2}{\partial x_\alpha \partial x_\gamma} [D_{\alpha\gamma}(t,t_0)\, \mathcal{f}(\mathbf{x}_0,t_0;\mathbf{x},t)].$$

Thus, for the two-time distribution function fully descriptive of the non-Markovian stochastic process under consideration, we derive the same kinetic equation.

In conclusion of this section we should like to point out that the generalized FP equation for non-Markovian–Langevin equations with memory functions $\lambda_{\alpha\beta}(t,s) = \lambda(t-s)\delta_{\alpha\beta}$, derived on the basis of the conditional probability density by Adelman[6] for the case of $t_0 = 0$ and by Fox[7] for an arbitrary t_0, is not exact. This is so because the derivation was based essentially on fluctuation–dissipation relations for a system with initial moment t_0 related to $-\infty$, rather than a system of stochastic equations with finite t_0. Their approximate equation follows from Eq. (11) if in the expression for kinetic coefficients $D_{\alpha\gamma}(t,t_0)$, one substitutes $\varphi_{\alpha\gamma}(0)$ for $\varphi_{\alpha\gamma}(t,t_0)$.

4. BROWNIAN MOTION OF A PARTICLE IN A VISCOELASTIC FLUID

The Brownian motion of a particle both in a viscoelastic fluid and in a viscous one should be considered as a non-Markovian stochastic process because in both cases the force acting upon the moving particle is defined by a formula including aftereffects.[8,9] If the fluid in which the particle moves is viscoelastic, and generally speaking, all fluids are in fact viscoelastic, the Brownian motion of the particle is a non-Markovian process even if the inertial effects in the fluid are ignored.[10] The interpretation of the Brownian motion as a non-Markovian process is corroborated by comparison of the results of a study based on numerical computation of simple models.[11]

The Brownian motion with due account for the hydrodynamic aftereffect described by Boussinesq's formula was for the first time investigated by Vladimirsky and Terletsky.[9] They derived the mean square of displacement of the Brownian particle. However, the method they used taking the aftereffect into consideration is essentially based on the assumption of the random force being noncorrelated.[12]

Consider by way of an example the Brownian motion of a spherical particle of radius a and mass m in an unbounded viscoelastic fluid with a particular relaxation time τ, described by a system of generalized Langevin equations

$$\frac{d}{dt} u_i(t) = -\frac{\zeta}{m\tau}\int_{t_0}^{t} e^{-(t-s)/\tau} u_i(s)\, ds + \frac{1}{m}\Phi_i(t). \quad (12)$$

Here, $u_i, i = 1,2,3$ stands for particle velocity components, while $\zeta = 6\pi a\eta$ is the Stokes frictional coefficient of the particle in a fluid having viscosity η. For random force Φ acting upon the Brownian particle all assumptions mentioned in Sec. 2 are true.

System (12) with finite t_0 describes the motion of a Brownian particle at random instants of time and determines a nonstationary Gaussian non-Markovian process. At the initial instant related to $-\infty$, system (12) describes the Brownian motion of the particle at instants infinitely remote from the initial instant and determines stationary process $\mathbf{u}_\infty(t)$. According to (12), the equation for $v_i(t) = u_i(t) - u_{\infty i}(t)$ takes the form

$$\frac{d}{dt} v_i(t) = -\frac{\zeta}{m\tau}\int_{t_0}^{t} e^{-(t-s)/\tau} v_i(s)\, ds + \frac{\zeta}{m\tau} e^{-(t-t_0)/\tau} C_i(t_0), \quad (13)$$

where

$$C_i(t_0) = -\int_{-\infty}^{0} e^{s/\tau} u_{\infty i}(s+t_0)\, ds.$$

Solution of Eqs. (13) establishes that

$$u_i(t) = u_{\infty i}(t) + v_i(t_0)\chi(t-t_0) - C_i(t_0)\dot{\chi}(t-t_0), \quad (14)$$

where $\chi(t) = (\tau^+ - \tau^-)^{-1}[\tau^+ e^{-t/\tau^+} - \tau^- e^{-t/\tau^-}]$ and τ^+ and τ^- are relaxation times defined by formulas

$$\tau^\pm = \frac{1 \pm (1 - 4\tau\zeta/m)^{1/2}}{2\zeta/m}.$$

At $t - t_0 \to +\infty$, $\mathbf{u}(t) \to \mathbf{u}_\infty(t)$; that is, process $\mathbf{u}(t)$ is stationary but asymptotically.

Now, using the results obtained in Sec. 2, we find the following equation for the conditional probability density corresponding to system (12):

$$\frac{\partial}{\partial t} P(\mathbf{u},t-t_0/\mathbf{u}_0) = \beta(t-t_0)\frac{\partial}{\partial u_i}[u_i\, P(\mathbf{u},t-t_0/\mathbf{u}_0)] + D(t-t_0)\frac{\partial^2}{\partial u_i^2} P(\mathbf{u},t-t_0/\mathbf{u}_0), \quad (15)$$

where $\beta(t-t_0) = -\dot{\chi}(t-t_0)/\chi(t-t_0)$,

$$D(t-t_0) = \frac{1}{2}\frac{\partial}{\partial t}\varphi(t-t_0) + \beta(t-t_0)\varphi(t-t_0).$$

Function $\varphi(t-t_0)$ results from multiplication of $u_i(t)$ by $u_j(t)$, defined in accordance with Eq. (14), with subsequent averaging at preset $\mathbf{u}(t_0)=\mathbf{u}_0$. The associated moments $\langle u_{\infty i}(t)u_{\infty j}(t)\rangle=(kT/m)\delta_{ij}$ are calculated using the equilibrium Maxwell distribution, while moments $\langle u_{\infty i}(t)u_{\infty j}(t')\rangle=(kT/m)\chi(t'-t)\delta_{ij}$ are calculated using Eq. (15) at $t-t_0\to\infty$. In this limiting case, $\varphi(t-t_0)=kT/m$,

$$\frac{\partial}{\partial t}\varphi(t-t_0)=0,$$

and Eq. (15) is completely defined. As a result, we have

$$\langle u_i(t),u_j(t)\rangle=\varphi(t-t_0)\delta_{ij},$$
$$\varphi(t-t_0)=(kT/m)\{1-\chi^2(t-t_0)+\tau^+\tau^-[2+\tau^+\tau^-/(2\tau^++\tau^-)(2\tau^-+\tau^+)]\times\dot{\chi}^2(t-t_0)\}. \quad (16)$$

The averaging bracket in Eq. (16), which involves a comma separating the stochastic processes, is essentially a cumulant bracket. In the case of a Brownian motion of a particle in a viscous fluid, when $\tau^+=m/\zeta$ and $\tau^-=0$, Eqs. (16) and, consequently, (15) are reduced to the known classical result. Multiplying Eq. (15) by $\mathscr{f}(\mathbf{u}_0,t)$, we find that the two-time distribution function of this process $\mathscr{f}(\mathbf{u}_0,t_0;\mathbf{u},t)$ satisfies the same equation as the conditional probability density.

The Brownian motion theory under consideration should be characterized as a phenomenological theory as opposed to the statistical theory of Brownian motion,[13–15] which provides a more detailed explanation of the effect of the fluid on Brownian particles, proceeding from point particles with a Brownian particle among them. Averaging of microscopic equations of motion has given[13,14] a generalized FP equation with a memory kernel. This equation, however, is approximate[16] and exactly defines only first-order moments.[6]

[1]V. I. Klyatskin and V. I. Tatarsky, Usp. Fiz. Nauk **110**, 499 (1973) [Sov. Phys. Usp. **16**, 494 (1974)].
[2]G. N. Bochkov, A. A. Dubkov, and A. N. Malakhov, Izv. Vysš. Učebn. Zaved. Radiofizika **20**, 406 (1977).
[3]R. Kubo, Rep. Prog. Phys. **29**, 255 (1966).
[4]R. J. Henery, J. Phys. A **4**, 685 (1971).
[5]P. Hänggi, Z. Phys. B **31**, 407 (1978).
[6]S. A. Adelman, J. Chem. Phys. **64**, 124 (1976).
[7]R. F. Fox, J. Math. Phys. **18**, 2331 (1977).
[8]T. S. Chow and J. J. Hermans, J. Chem. Phys. **56**, 3150 (1972).
[9]V. Vladimirsky and Ya. Terletsky, Zh. Eksp. Teor. Fiz. **15**, 258 (1945).
[10]T. S. Chow and J. J. Hermans, J. Chem. Phys. **59**, 1283 (1973).
[11]G. Subramanian and H. T. Davis, Phys. Rev. A **11**, 1430 (1975).
[12]P. I. Kuznetsov, R. L. Stratonovich, and V. I. Tikhonov, Zh. Eksp. Teor. Fiz. **26**, 189 (1954).
[13]J. L. Lebowitz and E. Rubin, Phys. Rev. **131**, 2381 (1963).
[14]R. M. Maso, J. Stat. Phys. **1**, 101 (1969).
[15]P. Mazur and I. Oppenheim, Physica (Utrecht) **50**, 241 (1970).
[16]V. S. Volkov, V. N. Pokrovsky, and V. A. Rozhnev, Izv. Akad. Nauk SSSR Meh. Zhidk. Gaza **6**, 3 (1976).

Practical use of the Hamilton–Jacobi equation

Alan Chodos and Charles M. Sommerfield
Physics Department, Yale University, P.O. Box 6666, New Haven, Connecticut 06511

(Received 3 June 1982; accepted for publication 24 September 1982)

We show by means of several examples how the Hamilton–Jacobi equation can be used to solve nonlinear ordinary differential equations whose direct integration is otherwise difficult.

PACS numbers: 03.20. + i, 02.30.Hg

I. INTRODUCTION

The Hamilton–Jacobi formalism has not played a central role either in the solution of classical mechanical problems or in the subsequent development of quantum mechanics (although for a brief period action-angle variables appeared to be the fundamental link between classical and quantum ideas). It is probably fair to say that to most physicists the Hamilton–Jacobi equation is something which was once learned (and/or taught) as part of an advanced course on classical mechanics,[1] and which has since been seldom if ever used.

It is the purpose of this note to point out that the Hamilton–Jacobi equation can be of direct practical utility in integrating certain nonlinear ordinary differential equations. We shall illustrate this contention by means of a number of examples, dealing with equations that have been of interest in recent years: the spherically symmetric Yang–Mills–Higgs monopole[2,3]; the superconducting vortex[4,5]; the axially symmetric Liouville equation[6]; and the spherically symmetric Einstein equations in five-dimensional space–time.[7]

These examples fall into two classes. In the first, only a particular solution to the Hamilton–Jacobi equation can be readily found, containing no free parameters. In this case, one obtains a set of first-order equations which imply the original second-order equations that one started with. (For instance, the Yang–Mills–Higgs equations are reduced to the Bogomol'ny equations). Whether these first-order equations generate solutions of interest will depend upon the individual problem.

In the second class, one is able to find not necessarily the most general solution to the Hamilton–Jacobi equations, but at least a solution depending upon a number of independent free parameters equal to the number of degrees of freedom. In this case, one is able, without further integration, to deduce the most general solution to the original equations. The Liouville equation and the general relativity example fall into this category.

We stress that the use of the Hamilton–Jacobi equation is not guaranteed to lead to a solution. Rather, it maps the original (presumably nonlinear) equation onto another nonlinear equation which, in some cases at least, is more tractable.

In Sec. II, we briefly review the Hamilton–Jacobi formalism. In Sec. III, we discuss the monopole and vortex examples, and in Sec. IV, the Liouville equation and the Einstein equation. Section V is devoted to some conclusions.

II. THE HAMILTON–JACOBI FORMALISM

Consider a dynamical system whose degrees of freedom are a number of coordinates $q_i(t)$, $i = 1,...,n$, and which is governed by a Hamiltonian $\mathscr{H}(q_i, p_i, t)$. We wish to perform a canonical transformation to a new set of variables

$$Q_i = Q_i(q_j, p_j, t),$$
$$P_i = P_i(q_j, p_j, t).$$

One way to do this[1] is to specify a function

$$S(q_i, P_i, t)$$

and to let

$$Q_i = \frac{\partial S}{\partial P_i}, \tag{1}$$

$$p_i = \frac{\partial S}{\partial q_i}. \tag{2}$$

Then the dynamics will be the same provided the new Hamiltonian $K(Q_i, P_i, t)$ is given by

$$K(Q_i, P_i, t) = \mathscr{H}(q_i, p_i, t) + \frac{\partial S}{\partial t}.$$

We observe that if K vanishes, then the dynamics has been rendered trivial:

$$Q_i = Q_{i0} = \text{const},$$
$$P_i = P_{i0} = \text{const}.$$

The Hamilton–Jacobi equation is simply the requirement that $K = 0$:

$$\mathscr{H}\left(q_i, \frac{\partial S}{\partial q_i}, t\right) + \frac{\partial S}{\partial t} = 0. \tag{3}$$

We now explain the two variants of the Hamilton–Jacobi method that we shall encounter below. (i) If we have a particular solution to Eq. (3), then we can apply Eq. (2). The canonical momentum p_i is found in terms of $\{q_j\}$ and $\{\dot{q}_j\}$ either from Hamilton's equation

$$\dot{q}_i = \frac{\partial \mathscr{H}}{\partial p_i}$$

or directly from the Lagrangian:

$$p_i = \frac{\partial L}{\partial \dot{q}_i},$$

if L is known. Then Eq. (2) becomes a first-order equation for $q_i(t)$. Of course, since we have started from a particular S, we can at most obtain particular solutions to the original equa-

 0022-2488/83/020271-05$02.50

tions of motion. Sometimes these will be the solution of interest, and sometimes not. We shall experience both possibilities in the next section. (ii) If we have a family of solutions to Eq. (3), containing a number of parameters γ_i, $i = 1,...,n$, then in general the γ_i will turn out to be functions of the new momenta P_i and, therefore, by Eq. (1) we have

$$\frac{\partial S}{\partial \gamma_i} = \frac{\partial S}{\partial P_j}\frac{\partial P_j}{\partial \gamma_i} = Q_{f0}\frac{\partial P_j}{\partial \gamma_i} = \text{const.}$$

That is, the set of n equations

$$\frac{\partial S}{\partial \gamma_i} = \alpha_i,$$

where the α_i are constant, allow us to solve algebraically for the unknown functions $q_i(t)$. Furthermore, this must be the general solution, because we have enough parameters (the n γ_i and the n α_i) to specify both q_i and $\dot{q}_i$ at some initial time. We shall explore two applications of this method in Sec. IV.

III. THE MONOPOLE AND VORTEX EXAMPLES

A. The monopole

We consider the classical dynamics of the Yang–Mills field $A^a_\mu(x)$ coupled to a Higgs field $\Phi^a(x)$ in the adjoint representation, in the limit of vanishing Higgs self-coupling. By making the ansatz[2]

$$A^a_i = \epsilon_{aij}\hat{r}_j[1 - K(r)]/er,$$
$$A^a_0 = 0,$$
$$\Phi^a = \hat{r}_a H(r)/er,$$

we find two coupled second-order equations for the functions $H(r)$ and $K(r)$:

$$r^2K'' = (H^2 + K^2 - 1)K \tag{4}$$

and

$$r^2K'' = 2K^2H. \tag{5}$$

The first step in applying the Hamilton–Jacobi formalism is to find a Lagrangian for this system of equations where the variable r plays the role of time. This is easily done. The result is

$$L = 2K'^2 + H'^2 + \frac{1}{r^2}(K^4 + 2H^2K^2 - 2K^2). \tag{6}$$

Consequently, the canonical momenta are

$$P_K = 4K', \quad P_H = 2H',$$

and the Hamiltonian is

$$\mathscr{H} = \tfrac{1}{8}P_K^2 + \tfrac{1}{4}P_H^2 - (1/r^2)(K^4 + 2H^2K^2 - 2K^2). \tag{7}$$

Note that the potential is unbounded below. Since we are using the Hamilton–Jacobi formalism merely as a technical device, this is of no particular concern.

From $\mathscr{H}$ we immediately obtain the Hamilton–Jacobi equation

$$\frac{\partial S}{\partial r} + \frac{1}{8}\left(\frac{\partial S}{\partial K}\right)^2 + \frac{1}{4}\left(\frac{\partial S}{\partial H}\right)^2 = \frac{1}{r^2}(K^4 + 2H^2K^2 - 2K^2). \tag{8}$$

On first, and subsequent, glances this equation appears too difficult to solve generally. The best we can do is to guess a plausible form for a particular solution. First we eliminate r from the equation by making the substitution

$$S = U(H,K)/r,$$

which results in

$$-U + \tfrac{1}{8}U_K^2 + \tfrac{1}{4}U_H^2 = K^4 + 2H^2K^2 - 2K^2. \tag{9}$$

Because the right-hand side is a polynomial in H and K, it seems natural to try a polynomial in H and K as an ansatz for U. One finds the solution

$$U(H,K) = 2HK^2 + H^2 - 2H + 1. \tag{10}$$

Not only are there no free parameters in the solution, it was actually fortunate that the ansatz worked at all, because there was one more equation for the coefficients of the polynomial than there were unknowns.

Armed with this solution, we can apply Eq. (2). We have

$$P_K = 4K' = \frac{\partial S}{\partial K} = \frac{1}{r}(4HK),$$

$$P_H = 2H' = \frac{\partial S}{\partial H} = \frac{1}{r}(2K^2 + 2H - 2).$$

These are the Bogomol'ny equations[3]

$$rK' = HK, \tag{11}$$

$$rH' = K^2 + H - 1. \tag{12}$$

One can eliminate H from these equations. In terms of $W = \ln K$, one has

$$W'' = (1/r^2)(e^{2W} - 1). \tag{13}$$

The best way to proceed at this point is to make the substitution

$$W = X(r) + \ln r, \tag{14}$$

in which case

$$X'' = e^{2X},$$

which can be integrated straightaway. However, it is instructive to continue with the Hamilton–Jacobi method. A Lagrangian for Eq. (13) is

$$L = \tfrac{1}{2}(W')^2 + (1/2r^2)(e^{2W} - 2W),$$

from which follows

$$P_W = W', \quad \mathscr{H} = \tfrac{1}{2}P_W^2 + (1/2r^2)(2W - e^{2W})$$

and

$$\frac{1}{2}\left(\frac{\partial S}{\partial W}\right)^2 + \frac{\partial S}{\partial r} = \frac{1}{2r^2}(e^{2W} - 2W). \tag{15}$$

Once again, it is tempting to eliminate the r dependence from this equation by letting $S = U/r$. Then

$$U_W^2 - 2U = e^{2W} - 2W.$$

Furthermore, if one is guided by the form of the right-hand side, it is natural to try

$$U = \alpha e^W + \beta W + \gamma,$$

and one finds the solution

$$U = \eta e^W + W + \tfrac{1}{2}, \quad \eta = \pm 1.$$

From this, we get the first-order equation

$$W' = (1/r)(\eta e^W + 1)$$

with solution

$$K(r) = e^W = \eta c r(1 - cr)^{-1}, \quad c = \text{const.}$$

For the monopole solutions, however, we want

$$K(0) = 1, \quad K(\infty) = 0$$

so these solutions have the wrong boundary conditions.

To obtain the desired solution (in fact, the most general solution) requires a more complicated ansatz. We try

$$S = W/r + m(r) + F(t), \tag{16}$$

where t is the variable

$$t = e^{2W}/r^2. \tag{17}$$

The motivation for this guess is as follows: the W/r term, when inserted into S_r, will cancel the $-W/r^2$ term on the right-hand side of Eq. (15). Then the rest of the right-hand side is simply $\frac{1}{2}t$, so it is reasonable to think that the remainder of S should be a function of t. It is necessary, however, to add a function of r alone to deal with the term $1/r$ in S_W induced by the W/r term in S. That the complete ansatz, Eq. (16), actually works is due to the fortunate cancellation of the cross term

$$\frac{1}{r}\frac{dt}{dW}F'$$

in $\frac{1}{2}S_W^2$, with the term

$$\frac{dt}{dr}F'$$

in S_r on the left-hand side of Eq. (15).

Putting Eq. (16) into Eq. (15) and choosing

$$m(r) = \frac{1}{2r} - \frac{\alpha^2}{2}r,$$

where α is arbitrary, we find the equation

$$\frac{dF}{dt} = \frac{\eta}{2}\frac{(t+\alpha^2)^{1/2}}{t}, \tag{18}$$

where $\eta = \pm 1$. This equation is easily integrated, and we find

$$S = \frac{W}{r} + \frac{1}{2r} - \frac{\alpha^2}{2}r + \eta\left[v + \frac{\alpha}{2}\ln\left(\frac{v-\alpha}{v+\alpha}\right)\right], \tag{19}$$

where $v \equiv (t+\alpha^2)^{1/2}$.

The presence of an abitrary parameter in this solution enables us to use the second method outlined in Sec. II. We differentiate S with respect to α, and equate the result to the constant value $-\alpha r_0$. After some straightforward algebra, we obtain

$$e^{2W} = K^2(r) = \alpha^2 r^2/\sinh^2\alpha(r - r_0).$$

The desired boundary conditions lead to the choices $\alpha = 1$, $r_0 = 0$, with the expected result

$$K(r) = r/\sinh r.$$

B. The vortex

In this example as well, the Hamilton–Jacobi technique will be useful in reducing the original second-order equations to a system of first-order equations, but in this case we shall not be able to proceed further and solve the problem completely.

The superconducting vortex is a special case of a charged scalar field ϕ coupled to the electromagnetic field. Physically, it describes the magnetic lines of flux which are expelled from the interior of a superconducting medium.[5] We assume axial symmetry, and let the distance from the axis be ρ, and the angle about the axis to be θ. We make the ansatz

$$\mathbf{A} = V(\rho)\hat{\theta}, \quad A_0 = 0, \quad \phi = \phi(\rho) \quad (\phi \text{ real}).$$

The Maxwell–Klein–Gordon equations then reduce to

$$V'' + (1/\rho)V' - V/\rho^2 = 2e^2\phi^2 V,$$

$$\phi'' + (1/\rho)\phi' = e^2V^2\phi + \tfrac{1}{2}\lambda(\phi^2 - 1)\phi,$$

where we have included a term

$$-\tfrac{1}{4}\lambda(\phi^2 - 1)^2$$

in the original Lagrangian.

As before, we seek a Lagrangian for this system in which ρ plays the role of time. We find

$$L = \rho(\phi')^2 + \tfrac{1}{2}\rho(V')^2 + e^2\rho\phi^2V^2 + \tfrac{1}{4}\lambda\rho(\phi^2 - 1)^2 + V^2/2\rho.$$

Thus

$$P_\phi = 2\phi'\rho, \quad P_V = \rho V',$$

and the Hamilton–Jacobi equation is

$$\frac{1}{4\rho}(S_\phi)^2 + \frac{1}{2\rho}(S_V)^2 - e^2\rho\phi^2V^2 - \frac{\lambda}{4}\rho(\phi^2 - 1)^2 - \frac{V^2}{2\rho} + \frac{\partial S}{\partial\rho} = 0. \tag{20}$$

To eliminate ρ, we try

$$S = R(\phi, V) + \rho T(\phi, V).$$

Comparing powers of ρ, we find the three equations

$$\tfrac{1}{2}R_\phi^2 + R_V^2 = V^2,$$

$$\tfrac{1}{2}R_\phi T_\phi + R_V T_V + T = 0,$$

$$\tfrac{1}{4}T_\phi^2 + \tfrac{1}{2}T_V^2 = e^2\phi^2V^2 + \tfrac{1}{4}\lambda(\phi^2 - 1)^2.$$

Although these are three equations in two unknown functions, the choice

$$T = e(\phi^2 - 1)V, \quad R = -\tfrac{1}{2}V^2$$

is a solution, provided we also demand that the self-coupling constant λ take the critical value

$$\lambda = 2e^2.$$

When λ takes on this value, the force between two vortices vanishes. Thus $\lambda = 2e^2$ demarcates the boundary between the type I superconductor ($\lambda < 2e^2$) in which vortices attract, and the type II superconductor ($\lambda > 2e^2$) in which they repel.[5] Then

$$S = e(\phi^2 - 1)V\rho - \tfrac{1}{2}V^2,$$

and, therefore, from Eq. (2),

$$\phi' = eV\phi,$$

$$V' + V/\rho = e(\phi^2 - 1).$$

Letting $W = \ln\phi$, we have $W' = eV$ and hence

$$W'' + W'/\rho = e^2(e^{2W} - 1). \tag{21}$$

If we let $e\rho = \bar{\rho}$, this is the axially symmetric version of the inhomogeneous Liouville equation

$$\nabla^2 W = e^{2W} - 1. \tag{22}$$

We can continue with the Hamilton–Jacobi method, much as we did for the monopole. A Lagrangian for Eq. (21) is

$$L = \bar{\rho}(W')^2 + \bar{\rho}(e^{2W} - 2W) \tag{23}$$

and thus

$$P = 2\bar{\rho}W'$$

and

$$\frac{1}{4\bar{\rho}}\left(\frac{\partial S}{\partial W}\right)^2 + \frac{\partial S}{\partial \bar{\rho}} = \bar{\rho}(e^{2W} - 2W). \tag{24}$$

Unfortunately, this equation is much less amenable to solution than the corresponding monopole equation (15). The situation changes dramatically, however, if we drop the last term on the right-hand side, that is, if we consider the homogeneous axially symmetric Liouville equation. This is no longer relevant to the vortex, of couse, but it is of interest in its own right and it is a problem to which we now turn.

IV. THE LIOUVILLE AND EINSTEIN EQUATIONS

A. The axially symmetric Liouville equation

The general solution to the two-dimensional Liouville equation

$$\nabla^2 W = e^{2W}$$

is known; therefore, in particular, all solutions to

$$W'' + (1/\rho)W' = e^{2W} \tag{25}$$

have been found (we drop the distinction between ρ and $\bar{\rho}$ introduced above). However, it is not trivial to integrate Eq. (25), and it is of interest to see how the application of the Hamilton–Jacobi technique successfully confronts this problem. As noted in Sec. III, the equation we wish to solve is the following truncated version of Eq. (24):

$$\frac{1}{4\rho}\left(\frac{\partial S}{\partial W}\right)^2 + \frac{\partial S}{\partial \rho} = \rho e^{2W}. \tag{26}$$

Multiplying by ρ, we have

$$\tfrac{1}{4}S_W^2 + S_\sigma = e^{2(W+\sigma)}, \tag{27}$$

where $\sigma = \ln\rho$.

The essential simplifying feature of this equation is that the right-hand side depends only on the combination $W + \sigma$. Note that this happy circumstance would no longer obtain had we retained the extra term in Eq. (24).

We take advantage of the situation by making the ansatz

$$S(W,\sigma) = -2W - (\lambda^2 + 1)\sigma + F(t),$$

where t is the variable $e^{2(W+\sigma)}$. This is essentially the same variable we defined earlier for the monopole problem in Eq. (17). Then Eq. (27) becomes

$$t^2[F'(t)]^2 = t + \lambda^2$$

and

$$F'(t) = \eta(t + \lambda^2)^{1/2}/t,$$

where $\eta = \pm 1$.

This equation is immediately integrable; the result is

$$S(W,\sigma) = -2W - (\lambda^2 + 1)\sigma + \eta\left[2\sqrt{t+\lambda^2} + \lambda \ln\left(\frac{\sqrt{t+\lambda^2} - \lambda}{\sqrt{t+\lambda^2} + \lambda}\right)\right]. \tag{28}$$

We now apply the second variant of the Hamilton–Jacobi method. Letting β be an arbitrary constant, we have

$$\frac{\partial S}{\partial \lambda} = -2\lambda\sigma + \eta \ln\left(\frac{\sqrt{t+\lambda^2} - \lambda}{\sqrt{t+\lambda^2} + \lambda}\right) = 2\beta,$$

which leads to

$$e^{2W} = (4\lambda^2/\rho^2)(\rho^\lambda e^\beta - \rho^{-\lambda} e^{-\beta})^2, \tag{29}$$

which is the general solution.

B. An equation from general relativity

Finding the spherically symmetric solutions to five-dimensional general relativity[7] amounts to solving Einstein's equations for a metric of the form

$$ds^2 = -e^\nu dt^2 + e^\lambda dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2) + 2A\, dx_5 dt + \Phi^2 (dx_5)^2,$$

where ν, λ, A, and Φ are functions only of r. One finds that the function

$$\Omega(r) = e^\nu \Phi^2 + A^2$$

obeys the equation

$$r\left(\frac{\Omega''}{\Omega'} - \frac{\Omega'}{\Omega}\right) + 1 + \frac{1}{\mu^2} r^4 (\Omega')^2 = 0. \tag{30}$$

Here μ is a constant of integration. We want the solution to obey the boundary condition

$$\lim_{r\to\infty} \Omega(r) = 1.$$

As usual, we must first find a Lagrangian for Eq. (30); it is

$$L = (1/r^2)\Omega/\Omega' - (r/\mu^2)\Omega^2, \tag{31}$$

from which follows

$$P_\Omega = (1/r^2)\Omega/\Omega'^2$$

and

$$\mathcal{H} = (-2/r)(-\Omega P_\Omega)^{1/2} + (r/\mu^2)\Omega^2.$$

The Hamilton–Jacobi equation is

$$S_\sigma - 2(-S_\kappa)^{1/2} + e^{2(\kappa+\sigma)} = 0, \tag{32}$$

where $\kappa = \ln(\Omega/\mu)$, and $\sigma = \ln r$. Defining

$$T = S + \tfrac{1}{2}e^{2(\kappa+\sigma)},$$

we find

$$\tfrac{1}{4}T_\sigma^2 + T_\kappa = e^{2(\kappa+\sigma)}. \tag{33}$$

Observe that this is identical to the Hamilton–Jacobi equation for the Liouville problem, Eq. (27), although the roles of the dependent and independent variables have been interchanged. This enables us immediately to set up a dictionary to transcribe our previous results

$$\begin{aligned} &S_{\text{rel}} \to S_{\text{Liou}} - \tfrac{1}{2}e^{2(\kappa+\sigma)}, \\ &\ln(\Omega/\mu) = \kappa \to \ln\rho, \\ &\ln r \to W. \end{aligned} \tag{34}$$

Letting t now stand for

$$t = r^2\Omega^2/\mu^2,$$

we find that the equation

$$\frac{\partial S_{\text{rel}}}{\partial \lambda} = \alpha$$

becomes

$$\ln\left(\frac{\sqrt{t+\lambda^2}-\lambda}{\sqrt{t+\lambda^2}+\lambda}\right) = 2\lambda\eta \ln \Omega + \gamma, \tag{35}$$

where $\gamma = (\alpha - 2\lambda \ln \mu)\eta$, and $\eta \pm 1$.

This is somewhat harder to solve explicitly for Ω as a function of r than was the corresponding Liouville case, because $\ln \Omega$ appears on the right-hand side as well. It is convenient to define a new variable R by the equation

$$r\Omega = (R^2 - B^2)/R, \quad \text{where } B^2 = \mu^2\lambda^2/4. \tag{36}$$

Then we have from Eq. (35)

$$\Omega = e^{-\gamma}\left(\frac{R-B}{R+B}\right)^{\mu/2B}, \tag{37}$$

where the choice of η has been absorbed into the choice of sign of μ. The relationship between r and R is then given explicitly as

$$r = \frac{R^2 - B^2}{R}\frac{1}{\Omega} = \left(\frac{R^2-B^2}{R}\right)\left(\frac{R+B}{R-B}\right)^{\mu/2B} \tag{38}$$

and we see that the choice $\gamma = 0$ insures that $\Omega \to 1$ as $r \to \infty$.

Note also that we obtain real solutions for Ω either when B is real or when B is purely imaginary.

V. CONCLUSIONS

We hope to have convinced the reader that the Hamilton–Jacobi equation has been undeservedly languishing in obscurity. It is true that in the paper we have not solved any equations whose solutions were not previously known; nevertheless, in the examples we have discussed straightforward integration of the equations [except for the monopole equation (13)] is difficult if not impossible. We have sought to illustrate what types of equations are likely to yield to the Hamilton–Jacobi technique, and to give the reader some intuition for the kinds of Ansätze that will solve the relevant Hamilton–Jacobi equations.

ACKNOWLEDGMENT

This work was supported in part by the U. S. Department of Energy Contract No. DE-AC 02-76 ERO 3075.

[1]See, for example, H. Goldstein, *Classical Mechanics*, 2nd ed. (Addison-Wesley, Reading, Mass., 1980), Chap. 10.
[2]M. K. Prasad and C. M. Sommerfield, Phys. Rev. Lett. **35**, 760 (1975).
[3]E. B. Bogomol'ny, Sov. J. Nucl. Phys. **24**, 449 (1976).
[4]V. L. Ginzburg and L. D. Landau, Zh. Eksp. Theor. Fiz. **20**, 1064 (1950).
[5]A. Jaffe and C. Taubes, *Vortices and Monopoles* (Birkhauser, Boston, 1980), esp. Secs. I.5 and III.3.
[6]J. Liouville, J. Math. Pure Appl. **18**, 71 (1853).
[7]A. Chodos and S. Detweiler, Gen. Rel. Grav. **14**, 879 (1982).

Covariance and geometrical invariance in *quantization

D. Arnal, J. C. Cortet, P. Molin, and G. Pinczon
Laboratoire de Physique-Mathématique, Université de Dijon, BP 138, 21004-Dijon Cedex, France

(Received 21 July 1981; accepted for publication 6 November 1981)

Several notions of invariance and covariance for * products with respect to Lie algebras and Lie groups are investigated. Some examples, including the Poincaré group, are given. The passage from the Lie-algebra invariance to the Lie-group covariance is performed. The compact and nilpotent cases are treated.

PACS numbers: 03.65.Bz, 02.20.Qs, 02.20.Sv

I. INTRODUCTION

The Hamiltonian formulation of classical mechanics is based on the symplectic structure of the phase space W of the system here considered. The space N of observables (C^∞ functions on W) is given a Lie-algebra structure by the Poisson bracket.

A detailed study of deformations of this Poisson algebra and of the usual associative structure on N was carried out in Ref. 1. Such a study allows a new approach to quantum mechanics as a deformation of classical mechanics which generalizes the Weyl–Wigner's quantization.

The phase-space W is supposed to be a symplectic, connected, paracompact manifold. We denote by $\{u,v\}$ the Poisson bracket of two elements u and v of N. Quantization of the associated dynamical system is given by a deformation of the usual associative product on N (a so-called * product):

Definition 1: Let $\mathscr{E}$ be the space of formal series in λ, with coefficients in N. A * product on W is a bilinear map:

$$(u,v)\in N\times N\to u*v=\sum_{r\geq 0}\lambda^r C^r(u,v)\in\mathscr{E},$$

where $C^r (r\geqslant 1)$ is a bidifferential operator on $N\times N$, vanishing on constants and such that, if $u,v,w\in N$

(i) $C^0(u,v)=u\cdot v$,

(ii) $$\sum_{r+s=t} C^r[C^s(u,v),w]=\sum_{r+s=t} C^r[u,C^s(v,w)],$$

(iii) $C^1(u,v)=\{u,v\}$,

(iv) $C^r(u,v)=(-1)^r C^r(v,u)$.

Each * product on W admits a natural extension to $\mathscr{E}\times\mathscr{E}$. Thanks to properties (i) and (ii), a * product is a deformation of the associative product on N; properties (iii) and (iv) ensure that the bracket defined on $N\times N$ by

$$[u,v]_*=(1/2\lambda)(u*v-v*u)$$

is a deformation (with parameter λ^2) of the Lie-algebra structure of N. Existence conditions of such deformations are given in Refs. 2 and 3.

In any quantization procedure, it is generally supposed that some particular Lie algebra of observables is preserved. Therefore the following notion is quite natural:

Definition 2: Let $\mathfrak{g}_W$ be a-finite dimensional Lie algebra of observables. A * product is $\mathfrak{g}_W$-relative quantization if

$$[u,v]_*=\{u,v\} \text{ for all } u \text{ and } v \text{ in } \mathfrak{g}_W. \quad (\mathfrak{g}_W\cdot Q)$$

All physical examples, previously treated in the * formalism, admit $\mathfrak{g}$-relative quantization with respect to suitable Lie algebras. These Lie algebras are usually obtained from natural geometrical symmetry groups acting by symplectomorphisms on phase space.

If G is a connected Lie group of symplectomorphisms of W, we denote the action of G on N (and its canonical extension to $\mathscr{E}$) by

$$(x\cdot u)(\xi)=u(x^{-1}\cdot\xi) \text{ for all } u\in N,\ x\in G,\ \xi\in W.$$

By differentiation, each element X of the Lie algebra $\mathfrak{g}$ of G is represented by a vector field X^- on W.

Since it is the case in physical examples, we shall assume that all vector fields generated by $\mathfrak{g}$ are globally Hamiltonian, and that the associated functions satisfy the commutation rules of $\mathfrak{g}$, i.e.,

Assumption (H):

If $X\in\mathfrak{g}$, there exists $u_X\in N$ such that $X^- v=\{u_X,v\}$ for all $v\in N$.

If $X,Y\in\mathfrak{g}$, $u_{[X,Y]}=\{u_X,u_Y\}$.

Assumption (H) means that $\mathfrak{g}$ is represented by a Lie algebra $\mathfrak{g}_W=\{u_X;\ X\in\mathfrak{g}\}$ of observables. Assumption (H) is satisfied, for instance, on the orbits of the coadjoint representation of G in the dual space $\mathfrak{g}^*$ of $\mathfrak{g}$.

Let us define some invariance properties of * products with respect to a Lie group of symplectomorphisms of W.

Definition 3: Under Assumption (H), a * product is called geometrically G invariant (G.I.) if:

$$x\cdot(u*v)=(x\cdot u)*(x\cdot v) \text{ for all } x\in G,\ u,v\in N. \quad \text{(G.I.)}$$

Definition 4: Under Assumption (H), a * product is called strongly G invariant (S.I.) if it satisfies (G.I.) and $(\mathfrak{g}_W\cdot Q)$ with $\mathfrak{g}_W=\{u_X,\ X\in\mathfrak{g},\ \mathfrak{g}$ the Lie algebra of $G\}$.

Strongly invariant * products were introduced in Ref. 1, where their physical meaning is discussed.

Unfortunately, except in the case of the cotangent bundle of a Stiefel manifold,[1,4] or in the case of homogeneous symmetric spaces,[5] we do not know if such * products exist in general. Moreover, we suspect that there exist a lot of symplectic manifolds [among which are natural phase spaces of some relativistic systems (see the examples in Sec. III)] on which there are no geometrically invariant * products. This negative observation is not so surprising, since, in fact, quantum mechanics does not impose conservation of all geometrical properties of the underlying classical system. For instance, the notion of trajectories being generally meaningless at the quantum level, there is no reason to require a

 0022-2488/83/020276-08$02.50

geometrical realization of the invariance group. Therefore it is convenient to generalize Definitions 3 and 4:

Definition 5: Under Assumption (H), a $*$ product is called G covariant if there exists a representation τ of G by automorphisms of the algebra $(\mathscr{E},*)$, such that

$$\tau_x(u) = \left(\mathrm{Id}_N + \sum_{s\geq 1} \lambda^s \tau_x^s\right)(x\cdot u), \quad \text{for all } x\in G,\ u\in N,$$

where τ_x^s is a differential operator on W.

Note that Definition 5 means that τ is a deformation of the geometrical action: $u\to x\cdot u$ of G.

The main result of this paper is that each quantization relative to a Lie algebra $\mathfrak{g}$, is a G covariant $*$ product with respect to the connected, simply connected Lie group G with Lie algebra $\mathfrak{g}$ (see Theorem 1 and Corollary).

In all physical papers,[1,6] the $*$ products introduced are $\mathfrak{g}_W$-relative quantizations with respect to some suitable Lie algebras $\mathfrak{g}$. From our result, this assumption seems physically quite natural, since it means that some covariance properties with respect to symmetry groups of the system are preserved after quantization, even if geometrical invariance might be lost. We give in this paper some examples of nilpotent Lie groups G, where no invariant $*$ product is known on the orbits of the coadjoint representation, though relative quantization can be constructed by natural generalization of the Moyal techniques.

On the other hand, so many formulations of local invariance were introduced that it seems necessary to establish relations between all these notions. In fact, we shall show in this paper that they are all more or less equivalent. For instance, under Assumption (H) the following local invariance properties of a $*$ product with respect to a Lie algebra $\mathfrak{g}$ were introduced in Ref. 1:

$$(\mathrm{IP}_1)\ \{u_X, v*w\} = \{u_X,v\}*w + v*\{u_X,w\}$$

for all X in $\mathfrak{g}$ and v,w in N.

$$(\mathrm{IP}'_1)\ \{u_X, v*w\} = \{u_X,v\}*w + v*\{u_X,w\}$$

for all X in $\mathfrak{g}$ and v,w elements of the associative $*$ algebra generated by $\mathfrak{g}_W$.

$$(\mathrm{IP}_2)\ \{u_X,v\} = [u_X,v]_* \text{ for all } X \text{ in } \mathfrak{g},\ v \text{ in } N.$$

In this paper, we prove the following:

1. If G is connected, (IP_1) is equivalent to the geometrical G invariance of the $*$ product.

2. If, moreover, $\mathfrak{g}_W$ is "sufficiently large" (i.e., if vector fields X^- for X in $\mathfrak{g}$ generate the tangent space at each point of W), then the following notions of invariance are equivalent:

a) $\mathfrak{g}_W$-relative quantization satisfying (IP_1),
b) $\mathfrak{g}_W$-relative quantization satisfying (IP'_1),
c) strong invariant $*$ product,
d) $*$ product satisfying (IP_2).

The natural notion of equivalence of $*$ products is given by

Definition 6: Two $*$ products $*$ and $*'$ on W are equivalent if there exists an operator T of the form

$$\mathrm{T} = \mathrm{Id}_N + \sum_{s\geq 1} \lambda^s T_s,$$

where the T_s's are differential operators, null on constants, such that

$$T(u*'v) = (Tu)*(Tv) \text{ for all } u \text{ and } v \text{ in } N.$$

In the compact semisimple case, we prove that, in the equivalence class of any $\mathfrak{g}$-relative quantization, there exists a strongly G invariant $*$ product. This means that geometrical properties are unchanged by quantization when the symmetry group of the system is a semisimple compact group.

In Refs. 1 and 6 a useful $*$ exponential map is introduced under the implicit assumption that the $*$ product series converges (in some sense) for particular value of the parameter λ. Is it possible to introduce this $*$ exponential without these assumptions of convergence, in fact, directly at the formal level? We give a partial answer to this question: In the case of a nilpotent Lie group G, we show that it is possible to define a $*$ exponential mapping from G into $(\mathscr{E},*)$ for each orbit of the coadjoint representation (see Theorem 2).

II. RELATIONS BETWEEN INVARIANCE NOTIONS OF $*$ PRODUCTS

Let $\mathfrak{g}_W$ be a finite-dimensional Lie algebra of observables on W.

Proposition 1: If $\mathfrak{g}_W$ is sufficiently large, the following are equivalent:

(i) $*$ is a $\mathfrak{g}_W$-relative quantization satisfying (IP_1).
(ii) $*$ is a $*$ product satisfying (IP_2).

Proof: Let us write

$$u*v = u\cdot v + \lambda\{u,v\} + \sum_{n\geq 2} \lambda^n C^n(u,v); \quad u,v\in N.$$

Then:

$$\text{(i)}\Leftrightarrow \begin{cases} C^{2r+1}(u_X,u_Y) = 0, \\ \{u_X, C^r(v,w)\} = C^r(\{u_X,v\},w) + C^r(v,\{u_X,w\}), \end{cases}$$
$$\text{(ii)}\Leftrightarrow C^{2r+1}(u_X,v) = 0 \qquad (1)$$
$$(\forall r\geq 1;\ \forall u_X,u_Y\in\mathfrak{g}_W;\ \forall v,w\in N).$$

(ii) implies (i) because (IP_2) implies trivially $\mathfrak{g}_W\cdot Q$ and

$$\begin{aligned}\{u_X, v*w\} &= [u_X, v*w]_* \\ &= (2\lambda)^{-1}((u_X*v - v*u_X)*w \\ &\quad + v*(u_X*w - w*u_X)) \\ &= \{u_X,v\}*w + v*\{u_X,w\}.\end{aligned}$$

If (i) is satisfied, we shall prove (1) by induction. Using the associativity of $*$, we obtain the following expression for the Hochschild coboundary of C^3 (see Ref. 2 for the definition):

$$\begin{aligned}\partial C^3(u,v,w) &= C^2(\{u,v\},w) - C^2(u,\{v,w\}) \\ &\quad + \{C^2(u,v),w\} - \{u, C^2(v,w)\}.\end{aligned}$$

Then, computing

$$\partial C^3(u_X,v,w) - \partial C^3(v,u_X,w) + \partial C^3(v,w,u_X) \quad \text{for } u_X\in\mathfrak{g}_W,$$

we obtain, using (1),

$$\begin{aligned}&C^3(u_X,vw) - wC^3(u_X,v) - vC^3(u_X,w) \\ &\quad = C^2(\{u_X,v\},w) + C^2(v,\{u_X,w\}) - \{u_X, C^2(v,w)\} = 0 \qquad (2)\end{aligned}$$

and, for $v = u_Y$ and $w = u_Z$ in $\mathfrak{g}_W$,

$$C^3(u_X,u_Y\cdot u_Z) = 0 \quad \forall u_X,u_Y,u_Z\in\mathfrak{g}_W.$$

By induction, the differential operator $C^3(u_X,\cdot)$ is null on the polynomial elements on $\mathfrak{g}_W$ and, since $\mathfrak{g}_W$ is sufficiently large, null identically. Now suppose (1) is satisfied for $r \leqslant l-1$. By a computation similar to (2), we obtain, using the induction hypothesis,

$$2C^{2l+1}(u_X,vw) - 2wC^{2l+1}(u_X,v) - 2vC^{2l+1}(u_X,w) = (a) + (b) + (c),$$

where

$$(a) = 2C^{2l}(\{u_X,v\},w) + 2C^{2l}(v,\{u_X,w\}) - 2\{u_X,C^{2l}(v,w)\},$$

$$(b) = \sum_{\substack{r+s=2l+1 \\ r,s>2 \\ r \text{ even}}} [-C^r(u_X,C^s(v,w)) + C^r(C^s(v,w),u_X)],$$

$$(c) = \sum_{\substack{r+s=2l+1 \\ r,s>2 \\ r \text{ odd}}} [C^r(C^s(u_X,v),w) + C^r(v,C^s(u_X,w)) - C^r(C^s(v,u_X),w) - C^r(v,C^s(w,u_X))].$$

The symmetry properties of the C^n imply that (b) = (c) = 0, and (i) that (a) = 0. We now have the analog of Eq. (2), and the end of the proof is similar to that for the case $l=1$.

Now suppose G is a connected Lie group acting on W under Assumption (H), and $\mathfrak{g}_W$ is the space of functions u_X. By differentiation of (G.I.), we immediately obtain

Proposition 2: A $*$ product is geometrically G invariant if and only if it satisfies (IP_1).

Corollary: A $*$ product is strongly G-invariant if and only if it is a $\mathfrak{g}_W$-relative quantization satisfying (IP_1).

Moreover, if $\mathfrak{g}_W$ is sufficiently large, strong G invariance is equivalent to (IP_2).

A priori, (IP'_1) is weaker then (IP_1). However,

Proposition 3: If $\mathfrak{g}_W$ is sufficiently large,

(i) $(IP'_1) \Leftrightarrow (IP_1)$.

(ii) A $\mathfrak{g}_W$-relative quantization satisfying (IP'_1) is strongly G invariant.

Proof: Let $(u_{X_i})_{i=1,2,\ldots,n}$ be elements of $\mathfrak{g}_W$. We write

$$u_{X_1} * u_{X_2} * \cdots * u_{X_n} = \sum_{k>0} \lambda^k S^k_n(X_1,\ldots,X_n).$$

From (IP'_1) we have

$$\{u_X,S^k_n(X_1,\ldots,X_n)\} = \sum_{i=1}^{n} S^k_n(X_1,\ldots,X_{i-1},[X,X_i],\ldots,X_n).$$

Let us write, with obvious notation,

$$\{u_X,S^k_n\} = \sum_{i=1}^{n} S^k_{n,i}.$$

The associativity of the $*$ product implies:

$$S^p_n = \sum_{j=0}^{p} \sum_{r+s=p-j} C^j(S^r_k,S^s_{n-k}), \quad \forall k = 0,\ldots,n.$$

Suppose (IP_1) is satisfied up to order $(p-1)$; then

$$\{u_X,S^p_n\} = \sum_{j=0}^{p-1} \sum_{r+s=p-j} C^j(\{u_X,S^r_k\},S^s_{n-k}) + C^j(S^r_k,\{u_X,S^s_{n-k}\}) + \{u_X,C^p(S^0_k,S^0_{n-k})\}$$

$$= \sum_{i=1}^{n} S^p_{n,i}$$

$$= \sum_{j=0}^{p} \sum_{r+s=p-j} C^j\left(\sum_{i=1}^{k} S^r_{k,i},S^s_{n-k}\right) + C^j\left(S^r_k,\sum_{i=k+1}^{n} S^s_{n-k,i}\right)$$

$$= \sum_{j=0}^{p} \sum_{r+s=p-j} C^j(\{u_X,S^r_k\},S^s_{n-k}) + C^j(S^r_k,\{u_X,S^s_{n-k}\}).$$

Then

$$\{u_X,C^p(S^0_k,S^0_{n-k})\} = C^p(\{u_X,S^0_k\},S^0_{n-k}) + C^p(S^0_k,\{u_X,S^0_{n-k}\}).$$

By induction on n and k, and using the sufficient largeness of $\mathfrak{g}_W$, we obtain (IP_1).

III. VEY $*^n$-PRODUCTS

Let Γ be a symplectic connection on W. The rth power of the Poisson bracket is the bidifferential operator defined in any chart by

$$P^r_\Gamma(u,v) = \Lambda^{i_1 j_1}\cdots\Lambda^{i_r j_r}\nabla_{i_1\cdots i_r}u\nabla_{j_1\cdots j_r}v, \quad \forall u,v \in N,$$

where Λ^{ij} are the coefficients of the structure 2-tensor Λ.

If $W = \mathbb{R}^{2n}$, the Weyl–Wigner quantization procedure is associated with the well-known Moyal $*$ product:

$$u*v = u\cdot v + \sum_{r\geqslant 1} \frac{\lambda^r}{r!} P^r_\Gamma(u,v), \quad \forall u,v \in N,$$

(the connection being the Riemannian flat one).

Flatness of the connection is essential and, in fact, if there is curvature, Moyal's formula no longer defines an associative product. Following Ref. 7 it seems quite natural to generalize Moyal's notion for any symplectic manifold W by introducing Vey $*^n$-products:

Definition 7: A $*$ product on W is called a Vey $*^n$-product if the principal symbol[5] of the bidifferential operators C^r and $(1/r!)P^r_\Gamma$ $(r \leqslant n)$ coincide, $n = 2,3,\ldots,\infty$.

Looking for Vey $*^n$-products is technically easier, and supported by the fact that each equivalence class of $*$ products contains a Vey $*^2$-product.[3] [*Note added in proof*: A. Lichnerowicz proved in a recent work; Deformations d'algèbres associées à une variété symplectique (les $*_\nu$-produits) to appear in Annales de l'Institut Fourier (Grenoble), that each equivalence class of $*$ product contains a Vey $*^\infty$-product. Then Sec. III could be rewritten, considering only Vey $*^\infty$-products.] Nevertheless, when geometrical invariance is required, this problem happens to have no solution, as shown by the examples that we are going to give in this section. The examples give some justification to our feeling that geometrical invariance after quantization is too much to ask. Let us remark that in each of these examples we can construct a Vey $*^\infty$-product which is a $\mathfrak{g}_W$-relative quantization and therefore (as will be shown in Section IV) we obtain covariance after quantization.

Our first example is a massless orbit of the coadjoint representation of the Poincaré group P.

Proposition 4: There is no geometrically P invariant Vey $*^2$-product on the orbit with $m^2 = 0$ and null helicity.

Proof: Consider a Vey $*^2$-product; the third term in its expansion is $\frac{1}{2}P_\Gamma^2 + \partial H$, where H is a differential operator with order less than three.[7] Then if this $*$ product is geometrically invariant, Γ is an invariant linear connection.[7]

We shall prove that the considered orbit does not carry any $SO_0(3,1)$ invariant connection. Let us suppose that Γ is an invariant connection; then the relation

$$\nabla_{[Y^-,X]}Z + \nabla_X[Y^-,Z] = [Y^-,\nabla_X Z]$$

holds for any Y in $so(3,1)$ (Lie algebra of the Lorentz group) and any vector fields X and Z on W (Ref. 5).

Here W is the cotangent bundle of the half light cone in $\mathbf{R}^8$ with equations

$$p_1^2 + p_2^2 + p_3^2 - p_4^2 = 0 \quad (p_4 > 0),$$
$$p_1q_1 + p_2q_2 + p_3q_3 - p_4q_4 = 0.$$

We choose a global chart $(p_1, p_2, p_3, q_1, q_2, q_3)$ on W and $X = Z = \partial/\partial p_1$ and successively, $Y^- = M_{23}^-$, M_{24}^-, and M_{34}^-.

Since in our chart

$$M_{23}^- = p_3\frac{\partial}{\partial p_2} - p_2\frac{\partial}{\partial p_3} + q_3\frac{\partial}{\partial q_2} - q_2\frac{\partial}{\partial q_3},$$
$$M_{24}^- = |p|\frac{\partial}{\partial p_2} + \frac{p\cdot q}{|p|}\frac{\partial}{\partial q_2},$$
$$M_{34}^- = |p|\frac{\partial}{\partial p_3} + \frac{p\cdot q}{|p|}\frac{\partial}{\partial q_3},$$

writing

$$\nabla_{\partial/\partial p_1}\frac{\partial}{\partial p_1} = \sum_{i=1}^{3}\alpha_i\frac{\partial}{\partial p_i} + \sum_{i=1}^{3}\beta_i\frac{\partial}{\partial q_i},$$

we obtain with the invariance relation for Γ at the point $(0, p_2, p_3, 0, 0, 0)$ $(p_2 \cdot p_3 \neq 0)$

for $Y^- = M_{23}^-$,

$$p_3\frac{\partial\alpha_2}{\partial p_2} - p_2\frac{\partial\alpha_2}{\partial p_3} - \alpha_3 = 0,$$
$$p_3\frac{\partial\alpha_3}{\partial p_2} - p_2\frac{\partial\alpha_3}{\partial p_3} + \alpha_2 = 0;$$

for $Y^- = M_{i4}^-$ $(i = 2,3)$

$$-\frac{1}{|p|} = |p|\frac{\partial\alpha_i}{\partial p_i} - \frac{\alpha_2 p_2 + \alpha_3 p_3}{|p|} \quad (i = 2,3),$$
$$\frac{\partial\alpha_3}{\partial p_2} = \frac{\partial\alpha_2}{\partial p_3} = 0.$$

In this linear system with respect to α_2, α_3, $\partial\alpha_i/\partial p_j$ $(i,j = 2,3)$, it is possible to eliminate the derivatives $\partial\alpha_i/\partial p_j$ and the remaining system

$$-1 = \frac{p_3^2}{p_2}\alpha_2 - p_3\alpha_3,$$
$$-1 = -p_2\alpha_2 + \frac{p_2^2}{p_3}\alpha_3,$$

does not have any solution.

Remarks:

1. On the orbit considered here, there exists a Vey $*^\infty$-product which is a quantization relative to the Lie algebra of Poincaré group.[8]

2. On the orbit with nonzero mass (without or with spin) there exists an invariant connection and it is possible to prove existence for a strongly invariant Vey $*^2$-product with respect to the Poincaré group.[9]

We shall now exhibit a group such that "almost" all orbits of the coadjoint representation do not admit any geometrically G invariant Vey $*^2$-product. We are indebted to M. Flato for suggesting this example.

Proposition 5: Let $\mathfrak{g}$ be the (nilpotent) Lie algebra with basis $(X_0, X_1, \ldots, X_k, Y;\ k > 3)$ and commutation relations

$$[Y, X_i] = X_{i-1}\ (i \geqslant 1); \quad [Y, X_0] = 0;$$
$$[X_i, X_j] = 0, \quad i,j = 0,1,\ldots,k,$$

and G be the corresponding connected and simply connected Lie group. The nontrivial orbits of the coadjoint representation are two-dimensional. The coordinate function ξ_0 defined on $\mathfrak{g}^*$ by X_0 is invariant. On the orbits satisfying $\xi_0 \neq 0$, there is no geometrically G invariant Vey $*^2$-product).

Proof: Let $(\xi_0, \xi_1, \ldots, \xi_k, \eta)$ be the coordinate functions associated with the basis $(X_0, X_1, \ldots, X_k, Y)$. Then the orbit of the point $(\xi_0, \xi_1, \ldots, \xi_k, \eta)$ is given by

$$\left\{\left(\xi_0, \xi_1 + s\xi_0, \ldots, \xi_k + s\xi_{k-1} + \cdots + \frac{s^k}{k!}\xi_0, \eta - t_k\xi_{k-1} - \cdots - \left(\sum_{j=1}^{k}\frac{t_j s^{j-1}}{j!}\right)\xi_0\right);\ \ s,t_j \in \mathbb{R},\ \ j = 1,\ldots,k\right\}.$$

The generic orbits $(\xi_0 \neq 0)$ are characterized by k invariant rational functions ρ_i defined by (see Ref. 10 for definitions):

$$\rho_0 = \xi_0, \rho_2 = \xi_2 - \frac{1}{2}\left(\frac{\xi_1}{\xi_0}\right)^2\xi_0, \ldots, \rho_k = \xi_k - \sum_{i=2}^{k-1}\frac{1}{2(k-i)}\left(\frac{\xi_1}{\xi_0}\right)^{k-i}\rho_i - \frac{1}{k!}\left(\frac{\xi_1}{\xi_0}\right)^k\xi_0.$$

These orbits are two-dimensional (they are of maximal dimension) and parametrized by ξ_1 and η.

From these expressions we immediately see that the subgroup of G acting trivially on these orbits is the one-parameter group generated by X_0. On the other hand, the group of automorphisms of a given linear connection Γ on W is at most four-dimensional. Therefore if k is larger than 3, there is no G invariant connection and then no geometrically G invariant Vey $*^2$-product on W.

Remarks:

1) We can use the above argument to prove that if k is sufficiently large, there does not exist a geometrically G invariant Vey $*^2$-product on the cotangent bundle of W.

2) We shall prove in Sec. V the existence of Vey $*^\infty$-products which are $\mathfrak{g}$-relative quantizations on each orbit of the coadjoint representation of any nilpotent Lie group.

IV. RELATIVE QUANTIZATION AND COVARIANCE

Definition 7:

(i) We denote by $L(\mathscr{E})$ the space of linear maps $T=\Sigma_{s>0}\lambda^s T_s$ from $\mathscr{E}$ into itself, and by $\mathrm{GL}(\mathscr{E})$ the group of invertible elements in $L(\mathscr{E})$. $L(\mathscr{E})$ is endowed with the Lie algebra structure associated to its usual associative algebra structure.

(ii) Let $*$ be a given $*$ product on W. $D=\Sigma_{s>0}\lambda^s D_s \in L(\mathscr{E})$ is called a $*$-derivation of $\mathscr{E}$ if

$$D(u*v)=D(u)*v+u*D(v);u,v\in N.$$

We denote by $\mathrm{Der}(\mathscr{E})$ the space of $*$ derivations of $\mathscr{E}$. $\mathrm{Der}(\mathscr{E})$ is a Lie subalgebra of $L(\mathscr{E})$.

(iii) We denote by $\mathrm{Aut}(\mathscr{E})$ the subgroup in $\mathrm{GL}(\mathscr{E})$ of automorphisms of $(\mathscr{E},*)$.

Theorem 1:

(i) Let $\Delta=\Sigma_{s>0}\lambda^s\Delta_s$ be an element of $L(\mathscr{E})$ such that

Δ_0 is a complete vector field X^- on W,

$\Delta_s (s\geqslant 1)$ is a differential operator.

Then, there exists a unique one-parameter-group: $t\to A(t)$ solution of the equation

$$\frac{dA}{dt}(t)=\Delta\circ A(t) \text{ with } A(0)=\mathrm{Id}_{\mathscr{E}}. \tag{3}$$

$A(t)=\Sigma_{s>0}\lambda^s A_s(t)$ commutes with Δ, and $a_s(t)=A_s(t)\circ\exp(-tX^-)$ is a differential operator with uniformly bounded order with respect to t, vanishing on constants if that is the case for Δ_s.

(ii) If, moreover, $\Delta\in\mathrm{Der}(\mathscr{E})$ for some $*$ product on W, then $A(t)\in\mathrm{Aut}(\mathscr{E}),\forall t\in\mathbb{R}$.

(iii) Let $\mathfrak{g}$ be a Lie algebra. If $D:\mathfrak{g}\to L(\mathscr{E})$ is a morphism of Lie algebras such that, for all $X\in\mathfrak{g}$, $D(X)$ satisfies the properties assumed for Δ in (i), then there exists an unique morphism A from the connected and simply connected Lie group $\tilde{G}$ with Lie algebra $\mathfrak{g}$, into $\mathrm{GL}(\mathscr{E})$ such that, for all $X\in\mathfrak{g}$,

$$\frac{d}{dt}A(\exp tX)=D(X)\circ A(\exp tX). \tag{4}$$

If $x\in\tilde{G}$, then $A(x)=\Sigma_{s>0}\lambda^s A_s(x)$, where $A_0(x)$ is the action of $\tilde{G}$ on W obtained by exponentiation of the morphism D_0 from $\mathfrak{g}$ into the vector fields on W, and $A_s(x)\circ A_0(x^{-1})(s\geqslant 1)$ is a differential operator.

Proof: The proof of (i) lies in Ref. 11. The unique solution of

$$\frac{dA_0}{dt}(t)=X^-\circ A_0(t) \text{ with } A_0(0)=Id_N \text{ is given by}$$

$$(A_0(t).u)(\xi)=u(\exp(-tX^-).\xi);\forall u\in N,\forall\xi\in W. \tag{5}$$

Equation (3) is solved by induction on the order s in λ. The unique solution of the equation of order s is

$$A_s(t)=\sum_{p=0}^{s-1}\int_0^t \exp(t-\tau)X^-\circ\Delta_{s-p}\circ A_p(\tau)\,d\tau.$$

$A_0(t)$ being a diffeomorphism of W, $A(t)$ is invertible. On the other hand, $B(t)=A(t')^{-1}\circ A(t+t')$ satisfies

$$\frac{dB}{dt}(t)=\Delta\circ B(t) \text{ with } B(0)=\mathrm{Id}_N.$$

Therefore $A(t)$ is a one-parameter group, and it is easy to prove that $A(t)$ is the unique solution of equation

$$\frac{dA}{dt}(t)=A(t)\circ\Delta, \text{ with } A(0)=\mathrm{Id}_N. \tag{6}$$

Let us define an action of $\mathbb{R}$ on the space $\mathscr{D}$ of differential operators on W by

$$(\exp tX.T)(u)=(\exp tX^-\circ T\circ\exp -tX^-)(u);$$

$$\forall T\in\mathscr{D},\ \forall t\in\mathbb{R},\ \forall u\in N. \tag{7}$$

Then

$$a_1(t)=A_1(t)\circ\exp(-tX^-)=\int_0^t \exp\tau X.\Delta_1 d\tau.$$

$a_1(t)$ is a differential operator, with bounded order on $\mathbb{R}$, vanishing on constants if that is the case for Δ_1. By induction, $a_s(t)$ is a differential operator with the same properties.

In order to prove (ii), if $u,v\in N$, we put

$$C(t)=A(t)(u*v)-A(t)(u)*A(t)(v).$$

$\frac{dC}{dt}(t)=\Delta\circ C$, with $C(0)=0$, and therefore $A(t)\in\mathrm{Aut}(\mathscr{E})$.

From (i) and (ii), we deduce that D can be exponentiated to a neighborhood $\mathscr{V}$ of the neutral element e of $\tilde{G}$ by elements $A(x)\in\mathrm{Aut}(\mathscr{E})(x\in\mathscr{V})$. Moreover, A_0 can be extended to a morphism on $\tilde{G}$.[12]

The proof of (iii) is adapted from Ref. 13. Let us suppose that A is a morphism on $\tilde{G}$ up to order $(n-1)$ in λ, i.e.,

$$A_p(xy)=\sum_{q<p}A_q(x)\circ A_{p-q}(y);\quad \forall x,y\in\tilde{G},\forall p\leqslant n-1,$$

or equivalently,

$$a_p(xy)=\sum_{q<p}a_q(x)\circ x\circ a_{p-q}(y);\quad \forall x,y\in\tilde{G},\forall p\leqslant n-1. \tag{8}$$

For any element $T=(T_0,T_1,\ldots,T_{n-1})\in\mathscr{D}^n$ and any $x\in\tilde{G}$, we put

$$(\pi(x)T)_p=\sum_{q=0}^{p}a_{p-q}(x)\circ(x\cdot T_q);\quad p=0,1,\ldots,n-1.$$

Using (8), it is easy to prove that π is a representation of $\tilde{G}$ on $\mathscr{D}^n$. We identify $\mathscr{D}^{n-1}\times N\times W$ with a separating subset of the dual of $\mathscr{D}^n$ by

$$\langle f|T\rangle=(T_0u+S_1T_1u+\cdots+S_{n-1}T_{n-1}u)(\xi),$$

where

$$f=(S_1,S_2,\ldots,S_{n-1},u,\xi)\in\mathscr{D}^{n-1}\times N\times W,$$

$$T=(T_0,\ldots,T_{n-1})\in\mathscr{D}^n.$$

$\tilde{G}$ acts on $\mathscr{D}^{n-1}\times N\times W$ by

$$\langle x.f|T\rangle=\langle f|\pi(x^{-1})T\rangle;\forall T\in\mathscr{D}^n,\forall x\in\tilde{G},$$

$$\forall f\in\mathscr{D}^{n-1}\times N\times W. \tag{9}$$

For each f in $\mathscr{D}^{n-1}\times N\times W$, we define a 1-form ω^f on $\tilde{G}$ by

$$(\omega^f(X))_x=\langle f|\pi(x)(D_{p+1}(X))_{p=0,\ldots,n-1}\rangle;\quad \forall x\in\tilde{G},\forall X\in\mathfrak{g}$$

($\mathfrak{g}$ being identified with the Lie algebra of left-invariant vector fields on $\tilde{G}$).

D being a morphism, $(D_{i+1}(X))$ is a cocycle for the cohomology of $\mathfrak{g}$ with coefficients in the $\mathfrak{g}$-module $\mathscr{D}^n$ obtained by differentiation of π. From this fact we deduce that ω^f is closed. Therefore, there exists a function $I(x,f)$ on $\tilde{G}$ such that $\omega^f_x=dI(x,f)$. We fix I such that $I(e,f)=0$. Computing I on $\mathscr{V}$, we find

$$I(\exp X, f) = \langle f | (a_{p+1}(\exp X))_p \rangle,$$

$$I(\exp X.\exp Y, f) = I(\exp X, f) + I(\exp Y, (\exp(-X), f).$$

Since $\mathscr{D}^{n-1} \times N \times W$ separates the points of $\mathscr{D}^n$, we deduce

$$(a_{p+1}(x.x'))_p = (a_{p+1}(x))_p + \pi(x)(a_{p+1}(x'))_p.$$

Thus

$$A_n(x.x') = \sum_{q=0}^{n} A_{n-q}(x) \circ A_q(x').$$

Then A is a morphism on $\tilde{G}$ up to order n (Q.E.D.).

Corollary: Each $\mathfrak{g}$-relative quantization, for a Lie algebra $\mathfrak{g}$ of observables, obtained from a group action with Assumption H, is G covariant with respect to the connected and simply connected Lie group with Lie algebra $\mathfrak{g}$.

Proof: Indeed, for $X \in \mathfrak{g}$ and $v \in N$, we put

$$D(X)v = \frac{1}{2\lambda}(u_X * v - v * u_X) = [u_X, v]_*.$$

D is a morphism from $\mathfrak{g}$ into Der $(\mathscr{E})$ which satisfies the hypothesis of Theorem 1.

We now examine the case of compact semi-simple groups.

Theorem 2: Let G be a connected compact semi-simple Lie group, acting under Assumption (H) on a symplectic manifold W. Then, each $\mathfrak{g}$-relative quantization (where $\mathfrak{g}$ is the Lie algebra of G) is equivalent to a strongly G invariant $*$ product.

Proof: With the same notations as in Theorem 1 above and the corollary, we write: $D(X) = \Sigma_{s>0} \lambda^s D_s(X)$.

Let us denote by $\tilde{G}$ the universal covering group of G and define

$$T = \int_{\tilde{G}} A(x) \circ A_0(x)^{-1} dx = \sum_{s>0} \lambda^s \int_{\tilde{G}} a_s(x)\, dx$$

$$= \sum_{s>0} \lambda^s T_s \quad (dx = \text{Haar measure on } G).$$

From the relation

$$a_s(\exp X) = \int_0^s \exp(tX^-) \circ D_s(X) \circ \exp(-tX^-)\, dt$$

we deduce that, for $s \geqslant 0$, T_s is a differential operator on W.

On the other hand, by definition of T, we have

$$A(x) \circ T = T \circ A_0(x), \quad \forall x \in \tilde{G}. \tag{10}$$

But since

$$D(X)(u_Y) = X^- u_Y, \quad \forall X, Y \in \mathfrak{g},$$

we have the relations

$$A(\exp X)(u_Y) = \exp X^-.u_Y, \quad \forall X, Y \in \mathfrak{g}$$

and

$$T(u_Y) = u_Y, \quad \forall Y \in \mathfrak{g}.$$

By differentiation of (10), we obtain

$$D(X)(Tv) = T(X^- v), \quad \forall X \in \mathfrak{g}, \forall v \in N,$$

which proves that the $*$ product

$$u *' v = T^{-1}(Tu * Tv)$$

satisfies IP_2, i.e., is strongly G invariant.

V. THE NILPOTENT CASE

Let G be a connected, simply connected nilpotent Lie group, $\mathfrak{g}$ its Lie algebra, and W an orbit of the coadjoint representation ρ of G.

Proposition 6: There exists on W a global chart $\xi \to (p,q) \in \mathbb{R}^{2k}$ such that

1. The canonical symplectic form on W is $\Sigma_{i=1}^{k} dp_i \wedge dq_i$.

2. Each function $u_X (X \in \mathfrak{g})$ defined by

$$u_X(\xi) = \xi(X), \quad \xi \in W,$$

has the form

$$u_X(p,q) = \sum_{i=1}^{k} \alpha_i(q)\, p_i + \beta(q),$$

where α_i and β are polynomial functions and $\partial \alpha_i / \partial q_j \equiv 0$, $\forall j = 1,2,\ldots,i$.

Proof: Using essentially the methods of Ref. 10, we prove the proposition by induction on dim$\mathfrak{g}$. If dim $\mathfrak{g} = 1$ or 2, dim $W = 0$ and the result is trivial. Let $\mathfrak{z}$ be the center of $\mathfrak{g}$, the restriction of the form ξ to $\mathfrak{z}$ is constant on W. Let f be this restriction.

First case: Ker $f \neq 0$.

If $\mathfrak{g}_1$ is the Lie algebra $\mathfrak{g}/\text{Ker}\, f$, $\mathfrak{g}_1^*$ is canonically injected in $\mathfrak{g}^*$ by the transposition map π^* of $\pi: \mathfrak{g} \to \mathfrak{g}_1$. By construction, $W \subset \pi^*(\mathfrak{g}_1^*)$. In fact, it is easy to prove that W is the image by π^* of one orbit W_1 of the coadjoint representation in $\mathfrak{g}_1^*$. Moreover, by definition, π^* is a symplectomorphism from W_1 to W and the conclusions of Proposition 6 follow immediately.

Second case: Ker $f = 0$.

Then dim$\mathfrak{z} = 1$ and $\mathfrak{g}$ can be decomposed in:

$$\mathfrak{g} = \mathbf{R}X \oplus \mathbf{R}Y \oplus \mathbf{R}Z \oplus \hat{\mathfrak{g}},$$

with $\mathfrak{z} = \mathbf{R}Z$, $[X, Y] = Z$ and $\mathfrak{g}_1 = \mathbf{R}Y \oplus \mathbf{R}Z \oplus \hat{\mathfrak{g}} = \{X_1 \in \mathfrak{g}$ such that $[X_1, Y] = 0\}$; $\mathfrak{g}_1$ is an ideal of $\mathfrak{g}$. Let ξ_X be the element of $\mathfrak{g}^*$ defined by $\xi_X(X) = 1$, $\xi_X(X_1) = 0$ if $X_1 \in \mathfrak{g}_1$. We identify $\mathfrak{g}_1^*$ with the orthogonal space of X in $\mathfrak{g}^*$. Let π be the projection from $\mathfrak{g}^*$ to $\mathfrak{g}_1^*$ with kernel $\mathbf{R}\xi_X$, G_1 the connected and simply connected subgroup of G with Lie algebra $\mathfrak{g}_1$ and ρ_1 the coadjoint representation of G_1. We have

$$\rho_1(\exp X_1) \circ \pi = \pi \circ \rho(\exp X_1), \quad \forall X_1 \in \mathfrak{g}_1. \tag{11}$$

Thus $\pi(W)$ is a family of ρ_1 orbits in $\mathfrak{g}_1^*$. Now let ξ_1 be in $\pi(W)$ and λ be a real number such that $\xi_1 + \lambda \xi_X$ belongs to W.

An easy calculation shows that

$$\xi_1 = \rho\left(\exp -\frac{\lambda}{f(Z)} Y\right)(\xi_1 + \lambda \xi_X).$$

$\pi(W)$ is a subset of W. ξ_1 being fixed, let W_1 be its ρ_1 orbit in $\mathfrak{g}_1^*$. Let us define

$$\varphi: W_1 \times \mathbf{R}^2 \to \mathfrak{g}^*$$

by

$$\varphi(\xi_1, t, s) = \rho(\exp tX)\, \xi_1 + s\xi_X \quad \xi_1 \in W_1, \ (t,s) \in \mathbf{R}^2.$$

First,

$$\varphi(\xi_1, t, s) = \rho\left(\exp tX \cdot \exp \frac{s}{f(Z)} Y\right) \xi_1 \in W.$$

Moreover, the relation $\varphi(\xi_1, t, s) = \varphi(\xi_1', t', s')$ implies $s = s'$.

The value of this form at Y is $\zeta_1(Y) - tf(Z) = \zeta'_1(Y) - t'f(Z)$, and since Y is in the center of $\mathfrak{g}_1$, we have $\zeta_1(Y) = \zeta'_1(Y)$, thus $t = t'$. This proves that φ is one-to-one. The surjectivity of φ onto W follows from the fact that G is the semidirect product of $\mathbb{R}$ by G_1 (Ref. 10); for each ξ in W

$$\begin{aligned}\xi &= \rho(\exp tX\cdot x_1)(\xi_1), \quad \text{with } x_1 \in G_1,\\ &= \rho(\exp tX)(\rho_1(x_1)\xi_1 + s\xi_X) \quad [\text{see formula (11)}]\\ &= \varphi(\rho_1(x_1)\xi_1, t, s).\end{aligned}$$

Clearly, φ is C^∞. Now, if we identify W_1 with $\mathbb{R}^{2(k-1)}$ by our induction hypothesis, putting $p_k = s$, $q_k = t$, we define a global chart on W. In this chart $u_X(\xi) = p_k$; moreover,

$$\begin{aligned}u_{X_i}(\xi) &= \langle X_1 | \rho(\exp q_k X)(\zeta_1)\rangle\\ &= \sum_l \frac{(-q_k)^l}{l!}\langle \mathrm{ad}^l X(X_1) | \zeta_1\rangle\\ &= \sum_l (-q_k)^l/l!\left(\sum_{i=1}^{k-1}\alpha_{li}(q)p_i + \beta_l(q)\right),\end{aligned}$$

where $\alpha_{li}(q)$ and $\beta_l(q)$ are polynomial functions of $q_1,\ldots,q_{k-1}$ and $\partial\alpha_{li}/\partial q_j \equiv 0$, $\forall j = 1,\ldots,i$. Thus, the second part of our proposition is proved. Finally let us consider two elements $T = T_1 + \lambda X, S = S_1 + \mu X$ of $\mathfrak{g}$ with T_1 and S_1 in $\mathfrak{g}_1$. If $\{\ ,\ \}$ is the Poisson bracket for the form $\Sigma_{i=1}^k dp_i \wedge dq_i$ on W, the above computations allow us to write

$$\frac{\partial u_{T_1}}{\partial q_k} = u_{[X,T_1]},\quad u_{[T_1,S_1]} = \{u_{T_1}, u_{S_1}\}.$$

Then

$$u_{[T,S]} = \{u_T, u_S\}.$$

This relation finishes the proof of Proposition 6.

Corollary: On W, there exist $\mathfrak{g}$-relative quantizations.

Indeed, let us consider on $\mathbb{R}^{2k}$ the $*$ product of Moyal.[1] With Proposition 6, we can write

$$u_X * u_Y = u_X u_Y + \lambda\{u_X, u_Y\} + \tfrac{1}{2}\lambda^2 P^2(u_X, u_Y), \qquad \forall X, Y \in \mathfrak{g}.$$

Then

$$[u_X, u_Y]_* = \{u_X, u_Y\}, \quad \forall X, Y \in \mathfrak{g}.$$

This $*$ product is G covariant (Theorem 1). In fact, it is possible to define directly a morphism from G to $(\mathscr{E},*)$ in this case. An easy computation allows us to define this morphism first at the Lie algebra level.

Lemma: Let $\mathfrak{g}_W$ be a finite-dimensional Lie algebra of functions on $W = \mathbb{R}^{2k}$ of the form

$$u_X = \sum_{i=1}^{k+1}\alpha_{n_i,n_{i+1},\ldots,n_{k+1}} q_k^{n_k} q_{k-1}^{n_{k-1}}\cdots q_i^{n_i} p_{i-1},$$

where $p_0 = 1$, $\alpha_{n_i,\ldots,n_{k+1}} \in R$, the bracket being the Poisson bracket associated to the form $\Sigma_{i=1}^k dp_i \wedge dq_i$. Let us define

$$r_{k+1} = 0, \quad r_j = \sup_{\{n_i\}}\sum_{i=j}^{k} n_i(r_{i+1}+1)\ j = k, k-1, \ldots 1,$$

and φ_λ from $\mathfrak{g}_W$ to $\mathscr{E}$ by

$$\varphi_\lambda(u_X) = \sum_i \alpha_{n_i,\ldots,n_{k+1}}(2\lambda)^{r_i - \Sigma_{j=i}^k n_j(r_{j+1}+1)} q_k^{n_k}\cdots q_i^{n_i} p_{i-1}.$$

Then

$$2\lambda\{\varphi_\lambda(u_X), \varphi_\lambda(u_Y)\} = \varphi_\lambda(\{u_X, u_Y\}), \quad \forall u_X, u_Y \in \mathfrak{g}_W.$$

The map φ: $\mathfrak{g} \to \mathscr{E}$ defined by $\varphi(X) = \varphi_\lambda(u_X)$ is a morphism from $\mathfrak{g}$ to $\mathscr{E}$ endowed with the bracket $u*v - v*u$. Now, we exponentiate φ to the corresponding Lie group.

Theorem 3: Let G be a connected, simply connected nilpotent Lie group and W an orbit of its coadjoint representation. If we identify W with $\mathbb{R}^{2k}$ as in Proposition 6, and define the $*$ product of Moyal on W, then there exists a morphism Φ from G into $(\mathscr{E},*)$ ($*$ exponential).

Proof: First, we solve the equation in $\mathscr{E}$:

$$\frac{d}{dt}A(t) = \varphi(X) * A(t), \quad \text{with } A(0) = 1. \tag{12}$$

If $A(t) = \Sigma_{s\geq 0}\lambda^s A_s(t)$ and $\varphi(X)*v = \Sigma_{k\geq 0}\lambda^k D_k(X)v$ with $D_k(X) \in \mathscr{D}$ (the space of differential operators on W) since $D_0(X)$ is a function $\varphi_0(X)$, we have

$$A_0(t) = e^{t\varphi_0(X)}$$

and

$$\frac{dA_n}{dt} = \varphi_0(X)A_n(t) + \sum_{k=1}^{n} D_k(X)A_{n-k}(t).$$

Putting $a_k(t) = A_k(t)\exp(-t\varphi_0(X))$, we can solve equation (12) by induction as in the proof of Theorem 1. We find

$$a_n(t) = \int_0^t \sum_{k=1}^{n} (e^{-s\varphi_0(X)} \circ D_k(X) \circ e^{s\varphi_0(X)})(a_{n-k}(s))\, ds.$$

The element

$$A(t) = \sum_{n\geq 0}\lambda^n a_n(t) e^{t\varphi_0(X)}$$

is invertible in $\mathscr{E}$ (Ref. 11) and the solution of Eq. (12). It is easy to prove that $A(t)$ is a one-parameter group in $\mathscr{E}$ and also the solution of

$$\frac{dA}{dt}(t) = A(t) * \varphi(X), \quad \text{with } A(0) = 1.$$

Let us define

$$\Phi(\exp X) = A(1), \quad X \in \mathfrak{g},$$

and suppose that Φ is a morphism up to order $n-1$, i.e., that the relation

$$\Phi(\exp X) * \Phi(\exp Y) = \Phi(\exp X \cdot \exp Y)$$

holds for the terms in λ^k, $k = 0,1,\ldots,n-1$. Putting $\Phi_0(\exp X) = e^{\varphi_0(X)}$, $a_n(\exp X) = A_n(1)\Phi_0(\exp X)$, we define on $\mathscr{D}^n$ a representation of G by

$$(\pi(x)T)_p = \sum_{q=0}^{p} a_{p-q}(x)\,\Phi_0(x) \circ T_q \circ \Phi_0^{-1}(x),$$

where $x \in G$, $T = (T_0, T_1, \ldots T_{n-1}) \in \mathscr{D}^n$; $p = 0,1,\ldots,n-1$. By duality, G acts on $\mathscr{D}^{n-1} \times N \times W$ (see the proof of Theorem 1)

$$\langle \pi^*(x^{-1})f | T\rangle = \langle f | \pi(x)T\rangle.$$

Or,

$$\begin{aligned}&\pi^*(x)(S_1, S_2, \ldots, S_{n-1}, u, \xi)\\ &\quad = \left[\left(\sum_{p=q}^{n-1} S_p a_{p-q}(x^{-1})\,\Phi_0(x^{-1})\right)_q,\ \Phi_0(x)u, \xi\right].\end{aligned}$$

φ being a morphism of Lie algebras, we define a closed form

ω^f on G by

$$\omega^f(X)_x = \langle f | \pi(x)(D_{p+1}(X))_{p=0,\ldots,n-1} \rangle .$$

We now define $I(x,f)$ as in the proof of Theorem 1. We have

$$I(x,f) = \langle f | (a_{p+1}(x))_p \rangle ;$$
$$I(xy,f) = I(x,f) + I(y,\pi^*(x^{-1})f) .$$

This proves the relation at order n:

$$a_n(xy) = \sum_{q=0}^{n} a_{n-q}(x)\, \Phi_0(x)\, a_q(y)\, \Phi_0(y^{-1}) ,$$

and Φ is a morphism up to order n. Q.E.D.

ACKNOWLEDGMENTS

We are indebted to M. Flato for many fruitful suggestions. We want to thank especially S. Gutt and M. Cahen for stimulating discussions and critical remarks during the development of this paper. We are also grateful to C. Fronsdal for useful information.

[1] F. Bayen *et al.*, Ann Phys. (N.Y.) **111**, 61 and 111 (1978).
[2] J. Vey, Comment. Math. Helvet. **50**, 421 (1975).
[3] O. M. Neroslavski and A. T. Vlassov, C. R. Acad. Sci. Paris, Série I **292**, 71 (1981).
[4] A. Lichnerowicz, Lett. Math. Phys. **1**, 505 (1977).
[5] M. Cahen and S. Gutt, C. R. Acad. Sci. Paris, Série A **291**, 543 (1980); S. Gutt: Mém. Acad. Sci. Bruxelles (in press).
[6] C. Fronsdal, Rep. Math. Phys. **15**, 111 (1979).
[7] A. Lichnerowicz, C. R. Acad. Sc. Paris, Série A **289**, 349 (1979).
[8] D. Arnal, J. C. Cortet, and P. Molin, C. R. Acad. Sci. Paris, Série A **291**, 327 (1980).
[9] P. Molin, Thèse 3° cycle (Université de Dijon, France, 1981).
[10] L. Pukanszky, *Leçons sur les représentations des groupes* (Dunod, Paris, 1967).
[11] A. Lichnerowicz, C. R. Acad. Sci. Paris Série A **287**, 1029 (1979).
[12] R. S. Palais, Memoirs Am. Math. Soc. **22** (1957).
[13] H. Stetkaer, Trans. Am. Math. Soc. **214**, 349 (1975).

Interaction-set scattering equations

K. L. Kowalski and A. Picklesimer[a)]
Department of Physics, Case Western Reserve University, Cleveland, Ohio 44106

(Received 19 May 1982; accepted for publication 20 August 1982)

The generalization of the pair-labeled Rosenberg equations for many-particle scattering are found in the case where there are arbitrary multiparticle interactions. These are called interaction-set equations because they involve auxiliary transition operators which are labeled by the same set of partitions which characterizes the various connectivities of the interparticle interactions. The technique which we employ also provides the analogous extension of a recently proposed set of connected-kernel multiple scattering equations for the Watson-type transition operators. Further, the structure of the interaction-set equations leads to the identification of an entire class of interaction-set connected-kernel scattering integral equations, each of which is based upon a distinct choice of unperturbed Green's function and its associated connectivity structure. The generalized Rosenberg equations and the connected-kernel Watson-type multiple scattering equations, which are limiting members of this class, correspond to the choice of the N-free-particle and two-cluster-channel unperturbed Green's functions, respectively.

PACS numbers: 03.65.Nk, 03.80. + r

I. INTRODUCTION

The scattering integral equations proposed by Rosenberg[1] were among the first generalizations to more than three particles of Faddeev's[2] connected-kernel approach to scattering theory. These equations have played an important, although not well-recognized, role in the development of multiparticle scattering theory in the last decade.[3–9] The distinguishing feature of the Rosenberg equations is that they are posed in terms of auxiliary transition operators that are indexed by the same partitions used to label the interactions. The physical amplitudes are then recovered by selective summations over these indices. Some possible advantages of the Rosenberg approach for nuclear reaction theory are discussed in Refs. 9 and 10.

In all previous work[1,3,5,6,9] involving the Rosenberg equations, only pair interactions are considered so that the transition operators are labeled by the $N(N-1)/2$ pair partitions i', where N is the total number of particles. The derivations of the Rosenberg equations given in Refs. 1, 6, and 9 do not generalize in any obvious fashion to include multiparticle interactions. The reason for this is basically graph-theoretical; namely, these derivations depend crucially upon the particularly simple form of the so-called *almost-connected* graphs,[1,5,11,12] which is realized when one has only pair interactions. With multiparticle interactions the possibilities for "almost connectedness" are much more numerous, and this complicates matters considerably. In the present work we avoid these difficulties by means of an entirely different approach, which enables us to obtain the generalization of the Rosenberg equations to include multiparticle interactions. We refer to these equations as the *interaction-set equations* because they are posed in terms of auxiliary transition operators which are indexed by the same set of partitions which characterize the interparticle interactions.

As emphasized in Refs. 13 and 14, truncation schemes in nuclear reaction theory generally lead to few-body effective interactions which, from a microscopic point of view, are of a multiparticle form. Our primary motivation for taking up the generalization of the Rosenberg equations is derived from this occurrence of multiparticle interactions rather than merely to include the possibility of fundamental multiparticle forces, although the latter may prove important in themselves. In fact, as discussed at length in Ref. 14, the fundamental multiparticle interactions are actually a very special case of the types of interaction we wish to include. The special advantages of interactions-set-type equations as proposed in Refs. 9 and 10 can then be considered in the context of realistic approximation schemes to nuclear reactions.

The similarity of the Watson[15] and Rosenberg approaches when there are only pair interactions has been known for a long time and provided the original motivation in Ref. 1. Recently this structural similarity was employed to obtain connected-kernel forms of the original Watson integral equations.[9] The generalizations of the Watson formalism to the elastic scattering of two composite fragments with arbitrary interactions have also been found.[14,16] In this article we establish the structural similarity between these generalized Watson equations and the interaction–set equations derived herein and then use this relationship to find connected-kernel integral equations for the generalized Watson transition operators, thus extending the connected-kernel multiple scattering equations of Ref. 9. Finally, we show that the interaction-set equations and the connected-kernel multiple scattering equations are simply limiting cases corresponding to N and two fragments, respectively, of a whole class of interaction-set-type equations. Each of these equations is shown to be identical in terms of its associated connectivity structure.[17]

When we refer to the *connectivity* of operators, we refer to the translational invariance properties of these operators with respect to various subgroups of particles.[17] This is known as *string connectivity* $\mathscr{C}_0$. In brief, subgroup invariance (noninvariance) is correlated with disconnectedness (connectedness). The elaboration and extension of this con-

a) Present address: Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742.

 0022-2488/83/020284-07$02.50

ception of connectivity is reviewed in Appendix B.

Many of the general expressions in this paper are rather condensed. Therefore, for illustrative purposes we explicate in Appendix C some features of the interaction-set equations in the particular case of four particles.

II. INTERACTION-SET EQUATIONS

Our development makes extensive use of the results obtained and the notation[18] employed in Refs. 14 and 17 (see also Appendix A[19]).

The *interaction set* is the set[18] $\hat{A}$ of all partitions b of the N-particle system such that $[V]_b \neq 0$, where $[V]_b$ is the b-connected part of the total interaction V:

$$V = \sum_{b \in \hat{A}} [V]_b \,. \tag{2.1}$$

The ("prior") transition operators

$$T^{a,b} = V^a(1 + GV^b) \tag{2.2}$$

are expressed in terms of the external interactions

$$V^a = \sum_{b \in \hat{A}} \bar{\Delta}_{a,b} [V]_b \,, \tag{2.3}$$

where

$$\begin{aligned} \bar{\Delta}_{a,b} &= 1, \quad b \not\subset a \\ &= 0, \quad b \subseteq a. \end{aligned} \tag{2.4}$$

The notation $b \subseteq a$ a means that b is *contained in* (or equal to) a while $b \not\subset a$ signifies that b is *not contained in a.*[17]

The operators $T^{a,b}$ satisfy the generalized BRS equations[14,19]

$$T = (M + \mathscr{M})\, G_0 \{ Q_1 \mathscr{S} \hat{G}^{-1} + (P_2 + \Gamma)\, T \} \,, \tag{2.5}$$

where we have used a matrix notation in the partition indices for the sake of brevity. For instance, T represents the matrix $(T^{a,b})$, while $\mathscr{S}$ is the matrix with elements $(\mathscr{S})_{a,b} = 1$. Also

$$M_{a,b} = [V^a G G_0^{-1}]_b \, \bar{\delta}_{1,b} \,, \tag{2.6}$$

where $[\cdots]_b$ denotes the b-connected part of the bracketed operator, $\bar{\delta}_{a,b} = 1 - \delta_{a,b}$, and[19]

$$\mathscr{M} = \bar{\Delta} P_1 [V]_1 \mathscr{S} Q_1 [\hat{G}]\, G_0^{-1} \,, \tag{2.7}$$

$$\Gamma = Q_2 \Delta^t C Q_1 \,. \tag{2.8}$$

Here P_i is the projector onto the i-cluster partitions, $Q_i = I - P_i$ and $([\hat{G}])_{a,b} = [G]_a \, \delta_{a,b}$. The partition which contains only one cluster is denoted by 1 [cf. (2.6)]. The diagonal matrix C has elements

$$C_a = (-1)^{n_a} (n_a - 1)! = -(\Delta^{-1})_{1,a} \,, \tag{2.9}$$

n_a is the number of clusters contained in partition a, while $(\hat{G})_{a,b} = G_a \, \delta_{a,b}$. The superscript t represents the transpose operation while

$$\Delta = \mathscr{S} - \bar{\Delta}. \tag{2.10}$$

The full and free Green's functions are denoted by G and G_0, respectively. The channel Green's functions are $G_b = (E + i0 - H_b)^{-1}$, where H_b is the channel Hamiltonian $H_0 + V_b$, the N-particle kinetic energy operator is H_0 and $V_b = V - V^b$. Obviously, $G = G_1$ and $G_0 = (E + i0 - H_0)^{-1}$, where we use 0 to denote the N-cluster partition.

Our strategy for obtaining the generalized Rosenberg equations is the reverse of that employed in Refs. 1, 6, and 9 for pair interactions. We exploit the fact that T is proportional to $\bar{\Delta}$, namely,

$$T^{a,b} = \sum_c \bar{\Delta}_{a,c} \, T_c G_0 \, G_b^{-1} \,, \tag{2.11}$$

in order to derive integral equations for the interaction-set operators

$$T_c \equiv [V]_c [1 + GV] \,. \tag{2.12}$$

This is done by using the inversion theorems derived in Ref. 17 for $\bar{\Delta}$.[20,21] In this regard we note that the sum in (2.11) effectively excludes the partition 0 since $[V]_0 = 0$ and, consequently, $T_0 = 0$. We note that when there are only pair interactions, corresponding to the partitions i', then $[V]_c = 0$ except for $[V]_{i'} = V_{i'}$ and then the $T_{i'}$ are essentially the Rosenberg operators.

The original[21] inversion theorem for $\bar{\Delta}$ is equivalent to the statement that the restriction[17]

$$\tilde{\Delta} \equiv Q_{0,1} \, \bar{\Delta} Q_{0,1} \,, \tag{2.13}$$

$$Q_{0,1} \equiv I - P_1 - P_N \,, \tag{2.14}$$

of $\bar{\Delta}$ to the (0,1)-excluded space possesses an inverse:

$$\tilde{\Delta}^{-1} = [Q_{0,1} \, \bar{\Delta} Q_{0,1}]^{-1} \,. \tag{2.15}$$

If we note that $T^{1,b} = 0$, then, for $a \neq 0,1$, we find from (2.11) that

$$\sum_{d \neq 0,1} (\tilde{\Delta}^{-1})_{a,d} \, T^{d,b} = \left(T_a + \left\{ \sum_{d \neq 0,1} (\tilde{\Delta}^{-1})_{a,d} \right\} T_1 \right) G_0 \, G_b^{-1} \,. \tag{2.16}$$

A somewhat neater inversion algorithm follows if we exploit the fact that[17]

$$\hat{\Delta}^{-1} = -Q_0 \Delta^{-1} Q_1, \tag{2.17}$$

where

$$\hat{\Delta} \equiv Q_1 \, \bar{\Delta} \, Q_0. \tag{2.18}$$

We have then from (2.11) that (recall $T_0 = 0$):

$$\sum_{d \neq 1} (\hat{\Delta}^{-1})_{a,d} \, T^{d,b} = T_a G_0 \, G_b^{-1} \,. \tag{2.19}$$

[If one uses Eqs. (3.33b), (3.49a), and (3.50a) of Ref. 17, it is easy to show that (2.16) and (2.19) agree for $a \neq 1$.]

In matrix notation (2.5) becomes

$$Q_1 T Q_1 = \hat{\Delta} B Q_1 G_0 \{ Q_1 \mathscr{S} \hat{G}^{-1} Q_1 + (P_2 + \Gamma)\, Q_1 T Q_1 \} \,, \tag{2.20}$$

where

$$B_{a,b} = \{ [[V]_a \, G G_0^{-1}]_b + \delta_{a,1} [V]_1 \, [G G_0^{-1}]_b \} \, \bar{\delta}_{1,b} \,. \tag{2.21}$$

The representation

$$[T_a]_b = [[V]_a \, G G_0^{-1}]_b = \{ ([\hat{V}] \cdot [\hat{G}])\, G_0^{-1} \}_{a,b} \,, \tag{2.22}$$

where

$$([\hat{V}]\cdot[\hat{G}])_{a,b} \equiv \sum_c \delta_{b,a\cup c}\,[V]_a\,[G]_c\,, \tag{2.23}$$

will prove to be useful in Sec. III. Here $a\cup c$ is the partition with the largest number of clusters which satisfies $a\subseteq a\cup c$ and $c\subseteq a\cup c$. Evidently (2.23) is a sum of b-connected terms.

If we apply the inversion formulae (2.17)–(2.19) to (2.20), then we obtain the *interaction-set scattering integral equations*:

$$T_a = \sum_b B_{a,b} + \sum_b (K_{\rm I})_{a,b}\,T_b\,, \tag{2.24}$$

where

$$(K_{\rm I})_{a,b} = \sum_{c,d\neq 1} B_{a,c}\,(\delta_{n_c,2}\,\delta_{c,d} + \Gamma_{c,d})\,G_0\,\hat{\Delta}_{d,b}\,. \tag{2.25}$$

We observe that if we call K the kernel of (2.5) or (2.20), then

$$K = \hat{\Delta} K_{\rm I}\,\hat{\Delta}^{-1}\,. \tag{2.26}$$

Hence

$$K_{\rm I}^2 = \hat{\Delta}^{-1} K^2 \hat{\Delta}\,. \tag{2.27}$$

It is shown in Ref. 14 that K^2 is a connected operator and so it is obvious from (2.27) that $K_{\rm I}^2$ is connected as well. Thus (2.24) constitute a set of connected-kernel equations, and they represent the generalization of the Rosenberg equations to include arbitrary multiparticle forces.

It is easy to recover the standard results when there are only pair interactions. First, we note the identity [Ref. 17, Eq. (3.61a)]

$$\Delta' Q_1 C \bar{\Delta} P_{N-1} = P_2\,\bar{\Delta} P_{N-1}\,, \tag{2.28}$$

where P_{N-1} is the projector on the pair partitions. Then if $[V]_a = 0$ except for $[V_{i'}]_{i'} = V_{i'}$ we obtain

$$T_{i'} = \sum_{b\neq 1} [V_{i'}(1+GV)]_b + \sum_{\gamma,j} [V_{i'}(1+GV)]_\gamma\,\bar{\Delta}_{\gamma,j}\,G_0\,T_j\,, \tag{2.29}$$

where the sums in the kernel of (2.29) are over all pairs j' and all two-cluster partitions γ. In that we begin with (2.20), the result (2.29) represents still another derivation of the Rosenberg equations.

The passage from the interaction-set equations (2.24) to the BRS equations or other sets of channel-labeled equations depends entirely on the properties of the matrix $\bar{\Delta}$ and its various submatrices square or nonsquare.[9,10] The matrix $\bar{\Delta}$, which was introduced in Ref. 22, and its various restrictions, such as $P_2\bar{\Delta}P_{N-1}$,[23] possess a host of remarkable properties, particularly in regard to inversion. Some of these properties have been explored in Ref. 17. A more extensive analysis of the inversions of the restrictions of $\bar{\Delta}$ is undertaken in Ref. 10.

III. WATSON EQUATIONS

As we pointed out in Sec. I, the Watson[15] multiple scattering equations are the prototype interaction-set equations. They provided some of the motivation for Rosenberg's original work. In this section we show how the techniques of Ref. 14 and Sec. II can be adapted to the problem of obtaining the generalization to arbitrary multibody interactions of the connected-kernel multiple scattering equations of Ref. 9.

The Watson formalism is specifically designed to deal with the elastic multiple scattering of a particle from a bound target. The generalization to the case of two complex fragments is straightforward.[16] A significant aspect of this formalism is that it singles out the two-cluster partition α, which corresponds to the relevant asymptotic channel. With this in mind, we introduce the α-biased counterparts of the transition operators $T^{a,b}$:

$$T^{a,b}(\alpha) \equiv V^{a,\alpha} + V^{a,\alpha} G V^{b,\alpha}\,, \tag{3.1}$$

where

$$V^{a,b} = \sum_c \bar{\Delta}_{a,c}\,\bar{\Delta}_{b,c}\,[V]_c\,. \tag{3.2}$$

The only operator among (3.1) which (in general) corresponds to a physical transition operator is

$$T^{\alpha,\alpha}(\alpha) = T^{\alpha,\alpha}\,, \tag{3.3}$$

which is the object of primary interest. We note that

$$T^{a,0}(\alpha) = \sum_b \bar{\Delta}_{a,b}\,T_b(\alpha)\,, \tag{3.4}$$

where

$$T_b(\alpha) \equiv [V^\alpha]_b(1+GV^\alpha) \tag{3.5}$$

and

$$[V^\alpha]_b \equiv \bar{\Delta}_{\alpha,b}\,[V]_b\,. \tag{3.6}$$

It is shown in Ref. 14 that the $T_a(\alpha)$ satisfy the (Watson-type) multiple scattering equations

$$T_a(\alpha) = t_a(\alpha)\left[1 + G_\alpha \sum_b \bar{\delta}_{a,b}\,T_b(\alpha)\right], \tag{3.7}$$

where

$$t_a(\alpha) = [V^\alpha]_a + [V^\alpha]_a G_\alpha t_a(\alpha)\,. \tag{3.8}$$

We note that

$$t_a(\alpha) = [V^\alpha]_a + [V^\alpha]_a G_\alpha [a][V^\alpha]_a\,, \tag{3.9}$$

where

$$G_\alpha[a] = \{G_\alpha^{-1} - [V^\alpha]_a\}^{-1}\,. \tag{3.10}$$

Our strategy for deriving connected-kernel interaction-set equations for $T_a(\alpha)$ is the same as in Sec. II. Namely, we use the BRS-type equations for the operators $T^{a,b}(\alpha)$, which are derived in Appendix A, and the inversion formulae (2.17)–(2.19). The repetition of that same procedure, as well as Eqs. (3.1)–(3.10), suggests a further generalization which makes the structural similarity of the Rosenberg and Watson equations far less mysterious. It should be clear that both equations are merely extreme cases of an entire class of interaction-set equations which are biased with respect to an arbitrary partition b. The relevant interaction-set operators are then

$$T_a(b) = [V^b]_a(1+GV^b)\,. \tag{3.11}$$

We obtain Rosenberg-type operators for $b = 0$ and Watson-type operators when $b = \alpha$ is a two-cluster partition. Since

$[V^1]_a = 0$, the case $b = 1$ is trivial. The generalization (3.11) may prove useful, for example, in few-body treatments of breakup and rearrangement scattering, that single out certain multicluster asymptotic channels for special consideration.

We now return to the specific example of the Watson-type operators, although all of our analysis will carry over to the operators (3.11). The full Green's functions in (3.1) and (3.5) are handled in an α-biased fashion. We note that

$$H = H_\alpha + V_a^\alpha + V^{\alpha,a}, \tag{3.12}$$

where

$$V_a^\alpha = \sum_b \Delta_{a,b}\,[V^\alpha]_b\,. \tag{3.13}$$

Thus, if we call $G_\alpha(a) \equiv (G_\alpha^{-1} - V_a^\alpha)^{-1} = G_a(\alpha)$, then

$$G = G_a(a) + G_a(a)\,V^{a,\alpha}\,G\,. \tag{3.14}$$

Equations (3.1)–(3.14) make evident the fact that the analysis of Sec. IV of Ref. 14 and Sec. II of the present paper carries over to the Watson case with the replacements

$$G_0 \to G_\alpha = G_\alpha(0) = G_0(\alpha), \tag{3.15}$$

$$G_a \to G_\alpha(a) = G_a(\alpha)\,, \tag{3.16}$$

$$[V]_a \to [V^\alpha]_a\,. \tag{3.17}$$

Specifically, it is shown in Appendix A that one then finds as the counterpart of (2.20):

$$Q_1 T(\alpha)\,Q_1 = \hat{\Delta} B(\alpha)\,Q_1 G_\alpha\,\{\mathscr{S}\,\hat{G}(\alpha)^{-1}\,Q_1 + (P_2 + \Gamma)\,Q_1 T(\alpha)\,Q_1\}\,. \tag{3.18}$$

We have again employed a matrix notation in the partition indices [e.g., $(\hat{G}(\alpha))_{a,b} = G_a(\alpha)\,\delta_{a,b}$] and, in this notation,

$$B(\alpha) \equiv ([V^\alpha]\cdot\{\hat{G}(\alpha)\})\,G_\alpha^{-1}\,Q_1 + P_1[V]_1\,\mathscr{S}\,Q_1\,\{\hat{G}(\alpha)\}\,G_\alpha^{-1}\,. \tag{3.19}$$

Here, corresponding to (2.23), we have

$$([V^\alpha]\cdot\{\hat{G}(\alpha)\})_{a,b}\,Q_1 = \bar{\delta}_{1,b}\sum_c \delta_{b,a\cup c}\,[V^\alpha]_a\,\{G(\alpha)\}_c\,, \tag{3.20}$$

where

$$\{G(\alpha)\}_b \equiv \sum_a (\Delta^{-1})_{b,a}\,G_a(\alpha) \tag{3.21}$$

is the counterpart of $[G]_b$, the b-connected part of G, and $(\{\hat{G}(\alpha)\})_{a,b} = \{G(\alpha)\}_a\,\delta_{a,b}$. We discuss the connectivity properties of $\{G(\alpha)\}_b$ shortly. From (3.18) and the inversion formulae (2.17)–(2.19), we obtain the generalization of the multiple scattering equations of Ref. 9 [cf. (2.24) and (2.25)]:

$$T_a(\alpha) = \sum_b B_{a,b}(\alpha) + \sum_b \{K_1(\alpha)\}_{a,b}\,T_b(\alpha)\,, \tag{3.22}$$

where

$$\{K_1(\alpha)\}_{a,b} = \sum_{c,d\neq 1} B_{a,c}(\alpha) \times (\delta_{n_c,2}\,\delta_{c,d} + \Gamma_{c,d})\,G_\alpha\,\hat{\Delta}_{d,b}\,. \tag{3.23}$$

The connectedness structure of $\{G(\alpha)\}_b$ is of crucial importance in determining whether (3.18) and (3.22) are connected-kernel equations or not. First we note that

$$G_a(\alpha) = \sum_b \Delta_{a,b}\,\{G(\alpha)\}_b\,, \tag{3.24}$$

so that, in particular,

$$G = G_1(\alpha) = \sum_b \{G(\alpha)\}_b \tag{3.25}$$

and

$$G_\alpha = G_0(\alpha) = \{G(\alpha)\}_0\,. \tag{3.26}$$

From the resolvent identity

$$G_a(\alpha) = G_\alpha + G_\alpha\,V_a^\alpha\,G_a(\alpha)\,, \tag{3.27}$$

(3.24) and the inversion properties of $\Delta_{a,b}$,[17] it is easy to show that

$$\{G(\alpha)\}_a = \delta_{a,0}\,G_\alpha + G_\alpha \sum_{b,c} \delta_{a,b\cup c}\,[V]_b^\alpha\,\{G(\alpha)\}_c\,, \tag{3.28}$$

which can be compared with

$$[G]_a = \delta_{a,0}\,G_0 + G_0 \sum_{b,c} \delta_{a,b\cup c}\,[V]_b\,[G]_c\,. \tag{3.29}$$

It is shown in Appendix B that $\{G(\alpha)\}_a$ consists of a sum of operators of connectivity c, $c \supseteq a$:

$$\{G(\alpha)\}_a = \sum_c \Delta_{c,a}\,[\{G(\alpha)\}_a]_c\,. \tag{3.30}$$

We also note that since $[A]_a$ is a-connected, then

$$[A]_a\,\{G(\alpha)\}_b = \sum_c \Delta_{c,b}\,[A]_a\,[\{G(\alpha)\}_b]_c \tag{3.31}$$

consists of a sum of operators of connectivity $a\cup c$, where $a\cup c \supseteq a\cup b$.

Equations (3.24)–(3.31) show that, in the terminology of Ref. 17, we can define an α-biased *connectivity structure*, $\mathscr{C}_\alpha$. This is the analog of the usual structural conception of connectivity in scattering theory which is called *string connectivity* and is identical with $\mathscr{C}_0$, except that the role of G_0 is now taken over by G_α [in going from (3.29) to (3.28), for example].[24] The substantiation of this observation, as well as some of the comments to follow, is provided in Appendix B.

As suggested by (3.26), G_α is completely disconnected in $\mathscr{C}_\alpha$. Also, as illustrated by (3.28), the product rules for the $\mathscr{C}_\alpha$, or $\{\ \}$, connectivity structure are the same as for the usual string connectivity. Thus the product of $\{A\}_a$ and $\{B\}_b$ has $\mathscr{C}_\alpha$-connectivity $a\cup b$. By definition, the $\mathscr{C}_\alpha$ properties of an operator are determined solely by the interactions external to the partition α (cf. Appendix B).[9,24] In particular, $\{A\}_a$ is that part of A which has $\mathscr{C}_0$ connectivity a when the interactions internal to partition α are treated as if they were completely disconnected operators. For example,

$$\{V_\alpha^a\}_b = V_\alpha^a\,\delta_{b,0} \tag{3.32}$$

and

$$\{V^\alpha\}_b = [V^\alpha]_b\,. \tag{3.33}$$

Since the product rules in $\mathscr{C}_\alpha$ for operators of different connectivities are the same as in $\mathscr{C}_0$, the same proof (given in Appendix A of Ref. 14) that (2.5) is a connected-kernel equa-

tion establishes (3.18) as a connected-kernel equation in $\mathscr{C}_\alpha$ *and*, most importantly, in $\mathscr{C}_0$ [cf. Eq. (B13)]. Clearly (3.22) shares these properties with (3.18), and we note that (3.19) and (3.20) can be expressed as

$$(B(\alpha))_{a,b} = [\,\bar{\delta}_{1,a}\,\{\{V^\alpha\}_a\, GG_\alpha^{-1}\}_b + \delta_{1,a}\,[V]_1\,\{GG_\alpha^{-1}\}_b\,]\,\bar{\delta}_{1,b}\,. \tag{3.34}$$

Equation (3.34) is to be compared with (2.21). In the case of only pair interactions, it is straightforward to combine (3.34) with (3.22) and (3.23) to obtain the analog of (2.29):

$$T_{i'}(\alpha) = \sum_{b\neq 1}\{\{V^\alpha\}_{i'}\, GG_\alpha^{-1}\}_b + \sum_{\gamma,j'}\{\{V^\alpha\}_{i'}\, GG_\alpha^{-1}\}_\gamma\,\bar{\Delta}_{\gamma,j'}\,G_\alpha\,T_{j'}(\alpha)\,. \tag{3.35}$$

This provides an alternative derivation of the result of Ref. 9.

The development of this section is a consequence of the fact that the unperturbed Green's function G_0 plays no role in the connectivity analysis of Sec. II and of Ref. 14. Thus, if we deal with equations with the same formal structure except for the replacements (3.15)–(3.17), then it is obvious that analogous connected-kernel properties will realize for the new equations. The same considerations apply to (3.11). For example, if we take as the counterparts of (3.15)–(3.17)

$$G_0 \to G_b = G_b(0)\,, \tag{3.36}$$

$$G_a \to G_b(a) = G_a(b) = (G_a^{-1} - V_b^a)^{-1}\,, \tag{3.37}$$

$$[V]_a \to [V^b]_a = \bar{\Delta}_{b,a}\,[V]_a\,, \tag{3.38}$$

then the relevant connectivity structure is $\mathscr{C}_b$, where the role of G_0 in string connectivity is taken over by G_b (cf. Appendix B). The counterparts of (3.18) and (3.22) are obvious, as is the fact that they constitute connected-kernel equations in $\mathscr{C}_b$ and $\mathscr{C}_0$. Since nothing we have done is predicated upon α being a two-cluster partition, all that is required is the identification $\alpha \to b$, $\mathscr{C}_\alpha \to \mathscr{C}_b$.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grants PHY78-26595 and PHY81-219561.

APPENDIX A

Here we sketch the derivation of (3.18). With the use of (2.9), (3.13), and the sum rules

$$\sum_a C_a\,\Delta_{a,b}\,\bar{\delta}_{a,1} = \bar{\delta}_{b,1} \tag{A1}$$

and

$$\sum_a C_a\,\bar{\delta}_{a,1} = 1, \tag{A2}$$

it is easy to show that

$$V^{\alpha,a} = \sum_b C_b\,(V_b^{\alpha,a} + \bar{\Delta}_{a,1}\,[V^\alpha]_1)\,\bar{\delta}_{b,1} \equiv \sum_b (\hat{\mathscr{V}}(\alpha))_{a,b}\,. \tag{A3}$$

With (A3) and the resolvent identities (B5) we find that (3.1) can be placed in the form

$$T^{a,b}(\alpha) = \sum_c (\hat{\mathscr{V}}(\alpha))_{a,c}\,G_c(\alpha)\,[G_b(\alpha)^{-1} + T^{c,b}(\alpha)]\,. \tag{A4}$$

In a matrix notation (A4) can be written as

$$T(\alpha) = \hat{\mathscr{V}}(\alpha)\,\hat{G}(\alpha)\,[\mathscr{S}\hat{G}(\alpha)^{-1} + T(\alpha)]\,, \tag{A5}$$

where, e.g., $(\hat{G}(\alpha))_{a,b} = G_a(\alpha)\,\delta_{a,b}$. We note that $\hat{\mathscr{V}}(\alpha)$ can be expressed alternatively as

$$\hat{\mathscr{V}}(\alpha) = (V_M(\alpha) + \bar{\Delta}P_1\,[V]_1\,\mathscr{S}Q_1)\,CQ_1 \equiv \hat{V}_M(\alpha)\,CQ_1\,, \tag{A6}$$

where

$$(V_M(\alpha))_{a,b} \equiv V_b^{a,\alpha}\,\bar{\delta}_{1,b}\,, \tag{A7}$$

so that (A5) becomes

$$T(\alpha) = \hat{V}_M(\alpha)\,CQ_1\,\hat{G}(\alpha)\,[\mathscr{S}\hat{G}(\alpha)^{-1} + T(\alpha)]\,. \tag{A8}$$

It is useful to express $\hat{V}_M(\alpha)\,Q_1$ as (note $[Q_1, C] = 0$)

$$\hat{V}_M(\alpha)\,Q_1 = \bar{\Delta}\,([V^\alpha] + P_1[V]_1\,\mathscr{S}P_0)\,\Delta' Q_1 \equiv \bar{\Delta}\hat{V}^\alpha\Delta' Q_1\,, \tag{A9}$$

where the matrix $[V^\alpha]$ is

$$([V^\alpha])_{a,b} = \bar{\delta}_{a,1}\,[V^\alpha]_a\,\delta_{a,b}\,. \tag{A10}$$

In obtaining (A9) we have used (2.3), (3.13), and the identity $\mathscr{S}P_0\Delta' = \mathscr{S}$.

Now from (3.21)

$$\{G(\alpha)\}_b \equiv \sum_a (\Delta^{-1})_{b,a}\,G_a(\alpha)\,, \tag{A11}$$

so that

$$G_a(\alpha) = \sum_b \Delta_{a,b}\,\{G(\alpha)\}_b\,. \tag{A12}$$

Then, if we employ (A12), (A9), and the identity $\Delta_{a,d}\,\Delta_{a,e} = \Delta_{a,d\cup e}$, we can show that

$$\hat{V}_M(\alpha)\,Q_1\,\hat{G}(\alpha) = \bar{\Delta}\,[\hat{V}^\alpha\cdot\{\hat{G}(\alpha)\}]\,Q_1\Delta' Q_1\,, \tag{A13}$$

where $(\{\hat{G}(\alpha)\})_{a,b} \equiv \{G(\alpha)\}_b\,\delta_{a,b}$, and

$$([\hat{V}^\alpha\cdot\{\hat{G}(\alpha)\}])_{a,b} \equiv \sum_{d,e}\delta_{b,e\cup d}\,(\hat{V}^\alpha)_{a,d}\,\{G(\alpha)\}_e\,. \tag{A14}$$

Thus (A8) may be written as

$$T(\alpha) = \bar{\Delta}\,[\hat{V}^\alpha\cdot\{\hat{G}(\alpha)\}]\,Q_1\Delta' Q_1 C\,[\mathscr{S}\,\hat{G}(\alpha)^{-1} + T(\alpha)]\,, \tag{A15}$$

or, upon noting that $\Delta' Q_1 C\mathscr{S} = Q_1\mathscr{S}$, as

$$T(\alpha) = \bar{\Delta}\,[\hat{V}^\alpha\cdot\{\hat{G}(\alpha)\}]\,Q_1\,[\mathscr{S}\hat{G}(\alpha)^{-1} + (P_2 + \Gamma)\,T(\alpha)]\,, \tag{A16}$$

where we have made use of the identity $\Delta' CQ_1 = P_2 + \Gamma$ and $\Gamma \equiv Q_2\Delta' CQ_1$. Equations (3.18)–(3.20) follow immediately upon noting that

$$[\hat{V}^\alpha\cdot\{\hat{G}(\alpha)\}]\,Q_1 = [[V^\alpha]\cdot\{\hat{G}(\alpha)\}]\,Q_1 + P_1[V]_1\,\mathscr{S}Q_1\,\{\hat{G}(\alpha)\}\,, \tag{A17}$$

while (2.5)–(2.7) follow analogously by setting $\alpha = 0$, in which case $\{\hat{G}(\alpha)\} \to [\hat{G}]$.

APPENDIX B

In this appendix we explicate the basis of the $\mathscr{C}_\alpha$ connectivity structure and, in particular, establish Eq. (3.30). We define

$$A = \sum_a [A]_a^\alpha , \tag{B1}$$

where $[A]_a^\alpha$ is defined to be that part of A which has $\mathscr{C}_0$ connectivity *a in the interactions external to partition* α. That is, in the assignment of connectivities with respect to the cluster decomposition (B1), the interactions internal to partition α are treated, for the purpose of classification of terms, as if they were completely disconnected operators. This defines a connectivity structure $\mathscr{C}_\alpha$.[17]

The sum over a in (B1) may be restricted further depending upon the structure of A. An example of this which is of particular importance is the Green's function $G_b(\alpha)$ in which the interactions external to α are also internal to b; thus

$$G_b(\alpha) = \sum_a \Delta_{b,a} \, [G_b(\alpha)]_a^\alpha . \tag{B2}$$

The resolvent identity

$$G_b(\alpha) = G_a(\alpha) + G_a(\alpha) \, V_b^a \, G_b(\alpha) , \tag{B3}$$

holds if $a \subseteq b$. It follows from (B3) that

$$[G_b(\alpha)]_a^\alpha = [G_a(\alpha)]_a^\alpha , \quad a \subseteq b. \tag{B4}$$

In a similar manner one infers from

$$G = G_a(\alpha) + G_a(\alpha) \, V^{a,\alpha} \, G, \tag{B5}$$

which holds for arbitrary a, that

$$[G]_a^\alpha = [G_a(\alpha)]_a^\alpha . \tag{B6}$$

We conclude from (B2)–(B6) that

$$G_b(\alpha) = \sum_a \Delta_{b,a} \, [G]_a^\alpha , \tag{B7}$$

and therefore

$$[G]_a^\alpha = \sum_b (\Delta^{-1})_{a,b} \, G_b(\alpha) . \tag{B8}$$

The comparison of (B7) and (B8) with (3.24) and (3.21), respectively, yields the identification

$$\{G(\alpha)\}_a = [G]_a^\alpha . \tag{B9}$$

Since it is evident from the definition of $[A]_a^\alpha$ that

$$[[A]_a^\alpha]_c = 0, \quad a \not\subset c, \tag{B10}$$

we infer, in the case of the Green's function, the string connectivity content for $\{G(\alpha)\}_a$ indicated in (3.30).

None of the preceding discussion depends upon the two-cluster nature of α. Consequently, all of our arguments generalize in an obvious fashion to a connectivity structure $\mathscr{C}_b$ defined with respect to an arbitrary partition b. In this context we encounter connectivity brackets $[\ ]_a^b$. We note that the usual string connectivity corresponds to

$$[\ ]_a^0 = [\ ]_a . \tag{B11}$$

The counterpart of (B10) is

$$[[A]_a^b]_c = 0, \quad a \not\subset c, \tag{B12}$$

restricting the string connectivity content of operators which are defined in terms of other connectivity structures. We note, in particular, that if $[A]_1^b \neq 0$, then

$$[[A]_1^b]_1 \neq 0. \tag{B13}$$

Finally, we remark that in the text we have chosen, for the sake of contrast and emphasis, to use the $\{\ \}$ rather than the $[\ ]^\alpha$ notation, where the connection is provided by (B9), for example. In general, however, the $[A]_a^b$ designation of the a-connected part of an operator in $\mathscr{C}_b$ is perhaps the more consistent notation.

APPENDIX C

The interaction-set equations (2.24),

$$T_a = B_a + \sum_b (K_I)_{a,b} \, T_b , \tag{C1}$$

where

$$B_a \equiv \sum_b B_{a,b} \tag{C2}$$

and where $B_{a,b}$ and $(K_I)_{a,b}$ are given by (2.21) and (2.25), respectively, possess a deceptively compact appearance. In order to illustrate the detailed substructure of these equations, we consider the case of $N = 4$ with pairwise, three-body, and four-body forces.

It will be necessary to classify the partitions $a,b,\cdots$ only on the basis of the number of clusters which they contain. As in the text, the one-, two-, three-, and four-cluster partitions are denoted as 1, α, i', and 0, respectively.

The full interaction for $N = 4$ is

$$V = \sum_{i'} V_{i'} + \sum_\alpha [V]_\alpha + [V]_1 . \tag{C3}$$

We stress that, for the two-cluster partitions α which have two particles in each cluster, the α-connected operators $[V]_\alpha$ do not correspond to any of the usual microscopic interactions. Such seemingly unphysical operators do appear very naturally in truncated nuclear reaction theories.[13,14]

We recall that the full Green's function G possesses the cluster decomposition

$$G = G_{\rm DIS} + [G]_1 , \tag{C4}$$

where

$$G_{\rm DIS} \equiv G_0 + \sum_{i'} [G]_{i'} + \sum_\alpha [G]_\alpha \tag{C5}$$

and where we denote the disconnected part of an operator $\mathcal{O}$ as $\mathcal{O}_{\rm DIS}$. The following linear combinations of the variously connected components of G are also of interest:

$$g_{i'} \equiv G_0 + [G]_{i'} + \sum_{f,\alpha} \delta_{\alpha,i' \cup f} \, [G]_f + \sum_\alpha \Delta_{\alpha,i'} \, [G]_\alpha , \tag{C6}$$

$$g_\alpha \equiv G_0 + \sum_{i'} \Delta_{\alpha,i'} \, [G]_{i'} + [G]_\alpha . \tag{C7}$$

It will suffice to tabulate the nonzero components of B_a

and $(K_I)_{a,b}$. One finds, for the Born terms,

$$B_{i'} = V_{i'}\, g_{i'}\, G_0^{-1} = (V_{i'} G)_{\rm DIS}\, G_0^{-1}\,, \tag{C8a}$$

$$B_\alpha = [V]_\alpha\, g_\alpha\, G_0^{-1} = [\,[V]_\alpha\, G\,]_\alpha\, G_0^{-1}\,, \tag{C8b}$$

$$B_1 = [V]_1\, G_{\rm DIS}\, G_0^{-1}\,. \tag{C8c}$$

The nine components of the kernel are

$$(K_I)_{i',j'} = \sum_\gamma [V_{i'}\, G\,]_\gamma\, \bar{\Delta}_{\gamma,j'}\,, \tag{C9a}$$

$$(K_I)_{i',\beta} = (V_{i'}\, G)_{\rm DIS} - \Delta_{\beta,i'}([V_{i'}\, G\,]_\beta + [V_{i'}\, G\,]_{i'})\,, \tag{C9b}$$

$$(K_I)_{i',1} = V_{i'}\, g_{i'} = (V_{i'}\, G)_{\rm DIS}\,, \tag{C9c}$$

$$(K_I)_{\alpha,\beta} = [V]_\alpha\, g_\alpha\, \bar{\delta}_{\alpha,\beta}\,, \tag{C9d}$$

$$(K_I)_{\alpha,j'} = [V]_\alpha\, g_\alpha\, \bar{\Delta}_{\alpha,j'}\,, \tag{C9e}$$

$$(K_I)_{\alpha,1} = [V]_\alpha\, g_\alpha\,, \tag{C9f}$$

$$(K_I)_{1,1} = [V]_1\, G_{\rm DIS}\,, \tag{C9g}$$

$$(K_I)_{1,\beta} = [V]_1 (G_{\rm DIS} - g_\beta)\,, \tag{C9h}$$

$$(K_I)_{1,j'} = [V]_1 \sum_\gamma [G]_\gamma\, \bar{\Delta}_{\gamma,j'}\,. \tag{C9i}$$

[1] L. Rosenberg, Phys. Rev. **140**, B217 (1965).

[2] L. D. Faddeev, Z. Eksp. Teor. Fiz. **39**, 1459 (1960) [Sov. Phys. JETP **12**, 1014 (1961)].

[3] References 4 and 5 represent recent reviews of the subject. The significance of the Rosenberg equations is discussed in Ref. 5. We remark in particular that the derivation in Ref. 6 by Redish of what have become known as the Bencze–Redish–Sloan (BRS) equations (Refs. 6–8) begins from what are essentially the Rosenberg equations. The BRS equations are derived from the Rosenberg equations in a similar manner in Ref. 9.

[4] F. S. Levin, Nucl. Phys. A **353**, 143C (1981).

[5] I. M. Narodetsky and O. A. Yakubovsky, ITEP-1, Moscow, 1980 (unpublished).

[6] E. F. Redish, Nucl. Phys. A **225**, 16 (1974).

[7] Gy. Bencze, Nucl. Phys. A **210**, 568 (1973).

[8] I. H. Sloan, Phys. Rev. C **6**, 1945 (1972).

[9] A. Picklesimer, P. C. Tandy, and R. M. Thaler, Ann. Phys. (N.Y.) (to be published); Phys. Rev. C **26**, 315 (1982).

[10] K. L. Kowalski and A. Picklesimer, Phys. Rev. C **26**, 1835 (1982).

[11] S. Weinberg, Phys. Rev. **133**, B232 (1964).

[12] K. L. Kowalski, Phys. Rev. C (to be published).

[13] W. N. Polyzou and E. F. Redish, Ann. Phys. (N.Y.) **119**, 1 (1979).

[14] R. Goldflam and K. L. Kowalski, Phys. Rev. C **21**, 483 (1980)

[15] K. M. Watson, Phys. Rev. **89**, 575 (1953).

[16] K. L. Kowalski, Ann. Phys. (N.Y.) **120**, 328 (1979).

[17] K. L. Kowalski, W. N. Polyzou, and E. F. Redish, J. Math. Phys. **22**, 1965 (1981).

[18] The script capital letters employed in Ref. 14 are exceptions to our rule of notational consistency. Some of the script capitals in Ref. 14 are replaced by their Latin counterparts in the present work.

[19] The result (2.5) was first derived in Ref. 14, although the expression given there for $\mathscr{M}$, which involves the fully-connected interaction $[V]_1$, is incorrect. The correct form for $\mathscr{M}$ is given by (2.7). The analysis of Appendix A of Ref. 14 contains several additional minor errors, which are corrected as part of the somewhat more general analysis in Appendix A of the present paper. For example, both (2.5) and (2.7) can be easily inferred from the development in our Appendix A.

[20] One of these theorems was first derived in Ref. 21.

[21] W. Polyzou, J. Math. Phys. **21**, 506 (1980).

[22] K. L. Kowalski, Phys. Rev. C **16**, 7 (1977).

[23] This restriction is called the selector matrix in Ref. 9.

[24] This is pointed out in the case of pair interactions, via an entirely different line of reasoning, in Ref. 9.

Improvement of the Froissart–Martin bound for complex scattering angles

Masud Chaichian
Department of High Energy Physics, University of Helsinki, Helsinki, Finland

Jan Fischer
Institute of Physics, Czechoslovak Academy of Sciences, Prague, Czechoslovakia

N. F. Nelipa
Institute of Nuclear Research, Moscow State University, Moscow, USSR

Ivo Vrkoč
Mathematical Institute, Czechoslovak Academy of Sciences, Prague, Czechoslovakia

(Received 5 August 1981; accepted for publication 6 November 1981)

Extension of the Froissart–Martin bound for complex scattering angles is improved using the solution of the Dirichlet boundary value problem for doubly connected domains. The Froissart–Martin bound for physical scattering angles is used as input value on one of the two boundaries. The obtained bound is valid in an ellipse smaller than the Lehmann–Martin one. Possibilities for further improvements and applications are discussed.

PACS numbers: 03.80. + r

1. INTRODUCTION

In a previous paper,[1] we obtained an extension of the two Froissart–Martin high-energy bounds[2] on the elastic scattering amplitude $f(s,\cos\theta)$,

$$|f(s,\cos\theta)| < C_1 s^{3/4}\ln^{3/2}s \Big/ |\sin\theta|^{1/2}, \quad \text{for } 0\neq 0\neq\pi \tag{1.1}$$

and

$$|f(s,\cos\theta)| < C_2 s\ln^2 s, \quad \text{for } \theta = 0 \text{ or } \pi \tag{1.2}$$

to unphysical scattering angles. Here, s is the center-of-mass (c.m.) total energy squared and θ is the c.m. scattering angle. The extended bound is valid in a complex neighborhood of the physical interval $-1\leqslant\cos\theta\leqslant 1$ inside the Lehmann–Martin ellipse, whose semimajor axis ρ is equal to $1 + a/s$ with some $a > 0$ and whose foci are at $\cos\theta = \pm 1$. If we represent a general point z of the complex plane in the form

$$z = \gamma\cos\theta + i(\gamma^2 - 1)^{1/2}\sin\theta \tag{1.3}$$

(with $0\leqslant\theta < 2\pi$ and $\gamma\geqslant 1$) and if we introduce the following two functions of s,

$$\delta = (2a)^{1/2}s^{-1/2}\ln^{-1}s,$$
$$\Lambda = 1 + 1/\ln s, \tag{1.4}$$

we can write the new bound[1] in the form

$$|f(s,x)| < C\frac{\delta - (\gamma - 1)\Lambda}{[\delta - (\gamma^+ - 1)\Lambda]^{3/2}[\delta - (\gamma^- - 1)\Lambda]^{3/2}}, \tag{1.5}$$

where $\gamma^{\pm} = \gamma \pm (\gamma^2 - 1).^{1/2}$ The bound is valid for all θ inside an ellipse with $\gamma = \gamma_2$, where

$$\gamma_2 = 1 + a's^{-1}\ln^{-2}s \tag{1.6}$$

for any positive $a' < a$. Comparing (1.6) with the Lehmann–Martin ellipse, $\gamma_L = 1 + as^{-1}$, we see that the validity domain of (1.5) shrinks faster than it by $\ln^2 s$ with increasing energy.

Setting $\gamma = 1$ on the right-hand side of (1.5), we easily check that the formula (1.5) reproduces the original Froissart–Martin bound (1.2) for forward and backward scattering, but it does not reduce to (1.1) for the other scattering angles, giving a bound that rises faster with increasing energy than the right-hand side of (1.1) at fixed θ. This suggests that an improvement of (1.5) should be possible.

In the present paper we show that such an improvement of (1.5) is indeed possible, by solving the Dirichlet boundary value problem for a ring. As boundary values of the solution on the inner and the outer circles of the ring, we take the right-hand side of (1.1) and the right-hand side of (1.5), respectively. As a result, we obtain an improvement of the asymptotic bound (1.5) everywhere inside the ellipse (1.6), which is a conformal map of the ring, with the exception of some neighborhoods of the foci. After a brief survey of methods and results of Ref. 1 in Sec. 2, we find in Sec. 3 the explicit form of the solution to the Dirichlet problem for a ring (Theorem 1) and show that the solution yields a new bound on $f(s,z)$ that is lower than (1.5) (Theorem 2). Proofs are deferred to Appendix A. Then, in Sec. 4, we use the general formula (3.2) to derive the bound. As the solution (3.2) has the form of an infinite series, we show in Appendix B that the order of the summation and of the high-energy limit $s\to\infty$ can be interchanged. This allows us to express the result in terms of hypergeometric functions [see formula (4.19)]. In Sec. 5, we consider various possibilities leading to a further improvement of the bound obtained. Concluding remarks are given and possible applications of the result are discussed in Sec. 6.

2. BOUNDS ON SCATTERING AMPLITUDE FOR PHYSICAL AND UNPHYSICAL ANGLES

The amplitude $f(s,z)$ describing the scattering of two spinless particles can be expanded in the Legendre series

$$f(s,z) = \frac{(s)^{1/2}}{2k}\sum_{l=0}^{\infty}(2l + 1)a_l(s)P_l(z), \tag{2.1}$$

where k is the c.m. momentum of the particle. The unitarity condition implies that

$$|a_l(s)| < 1 \tag{2.2}$$

at all energies and $l = 0,1,2,\ldots$.

Further, we consider the auxiliary function $g(s,w)$ de-

 0022-2488/83/020291-09$02.50

fined in terms of the partial amplitudes $a_l(s)$,

$$g(s,w) = \frac{(s)^{1/2}}{2k} \sum_{l=0}^{\infty} (2l+1) a_l(s) w^l. \tag{2.3}$$

Using the well-known representation of the general Legendre polynomial $P_l(z)$,

$$P_l(z) = \frac{1}{\pi i} \int_\Gamma w^l (w^2 - 2wz + 1)^{-1/2} \, dw,$$

where Γ is a curve connecting the points $z - (z^2-1)^{1/2} = \gamma^- e^{-i\theta}$ and $z + (z^2-1)^{1/2} = \gamma^+ e^{i\theta}$, we can relate $f(s,z)$ to $g(s,w)$ by the formula

$$f(s,z) = \frac{1}{\pi i} \int_\Gamma g(s,w) \frac{dw}{(w^2 - 2wz + 1)^{1/2}}, \tag{2.4}$$

where Γ is chosen so that the points $\gamma^{\pm} e^{\pm i\theta}$ and the curve Γ connecting them lie inside the convergence circle $C_{\tilde{R}}$ of (2.3) with $\tilde{R} > R = \rho + (\rho^2 - 1)^{1/2}$. Then, following the approach of Kinoshita, Loeffel, and Martin,[3] we made in Ref. 1 the additional assumption that $f(s,z)$ is bounded, at sufficiently high energies, by a polynomial in energy everywhere inside the Lehmann–Martin ellipse,

$$|f(s,z)| < s^N, \quad s > s_0, \tag{2.5}$$

and obtained[1] the following high-energy bound on $g(s,w)$:

$$|g(s,w)| < 4(2a')^{-1/4} \pi^{-1/2} s^{N+1/4} (R - |w|)^{-2} \tag{2.6}$$

for any $a' < a$ and $|w| < R$.

Besides the polynomial bound (2.6), an s-independent upper bound on $g(s,w)$ holds inside the unit circle. It follows from the unitarity condition (2.2) and from the Taylor expansion (2.3), and has the form

$$|g(s,w)| < C_1 (1 - |w|)^{-2}, \quad \text{for } |w| < 1. \tag{2.7}$$

While (2.7) cannot be used at $|w| \geqslant 1$, its influence nevertheless extends to the ring $1 \leqslant |w| < R$ thanks to the bound (2.6). This is shown in Ref. 1 with the help of the three-arc theorem. The result is

$$|g(s,w)| < C_2 \left(1 - |w| + \frac{R - |w|}{\ln s}\right)^2, \quad 1 \leqslant |w| \leqslant \tilde{r}, \tag{2.8}$$

where

$$\tilde{r} = 1 + \frac{1}{1 + \ln s} \left(\frac{2a}{s}\right)^{1/2}.$$

As the first step, we derived in Ref. 1 a bound that unifies (1.1) and (1.2) in one compact formula. To this end, introducing the notation

$$K = w^2 - 2wz + 1 \tag{2.9}$$

and expressing K as

$$K = (w - \gamma^+ e^{i\theta})(w - \gamma^- e^{-i\theta}), \quad \gamma^{\pm} = \gamma \pm (\gamma^2 - 1)^{1/2},$$

we estimate K from below

$$|K| \geqslant |\gamma^+ - |w|| \, |\gamma^- - |w|, \tag{2.10}$$

while the integration curve Γ is chosen according to Fig. 1. The resulting bound is

$$|f(s,\cos\theta)| < \frac{C_3}{2} \left(\frac{1}{\delta^{3/2} |\sin\theta|^{1/2}} \times \arctan \frac{|\sin\theta|^{1/2}}{\delta^{1/2}} + \frac{1}{\delta(\delta + |\sin\theta|)} \right). \tag{2.11}$$

One can easily check that it reproduces, apart from constant factors, the bound (1.1) and (1.2) for $\theta \gg \delta$ and $\theta \ll \delta$, respectively.

Both the bound (2.8) on $g(s,w)$ and the formula (2.4) can be used outside the unit circle in th w plane, i.e., in a complex neighborhood, $\gamma > 1$, of the interval $[-1,1]$ in the z plane. This means that, replacing $g(s,w)$ in (2.4) by its bound (2.8), we obtain a bound on the amplitude $f(s,z)$ for unphysical scattering angles. Its form depends on the choice of the curve Γ in the w plane. In Ref. 1, we obtained the bound (1.5) by choosing Γ to be the straight line connecting the fixed end points $\gamma^- e^{-i\theta}$ and $\gamma^+ e^{i\theta}$ (see Fig. 2).

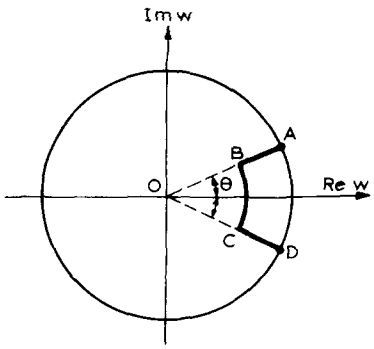

FIG. 1. Integration curve ABCD for the estimate of the right-hand side of Eq. (2.4) in the case of physical scattering angles θ. The resulting bound is (2.11).

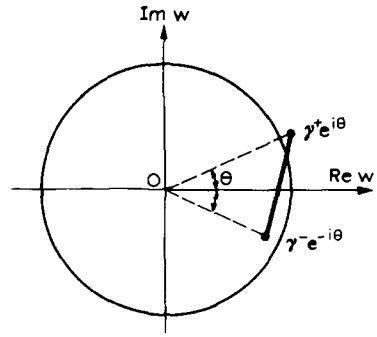

FIG. 2. In the case of complex scattering angles, the integration curve Γ connecting the points $\gamma^+ e^{i\theta}$ and $\gamma^- e^{-i\theta}$ is chosen to be the straight line. The resulting bound is (1.5).

3. THE BOUNDARY VALUE PROBLEM FOR HARMONIC FUNCTIONS IN A RING

We shall need in the next section the solution of the Dirichlet boundary value problem for a ring. Theorems on the existence and uniqueness of the solution to this problem are well known (see e.g., Ref. 4), as well as the explicit form of the solution[5] (see, e.g., Ref. 6). We nevertheless derive in Theorem 1 the solution in a different form, one that is more appropriate for our purpose. A sketch of the proof is given in Appendix A. In Remark 2, we generalize the result to the case of unbounded integrable functions $g_1(\varphi)$ and $g_2(\varphi)$. For the case of a general doubly connected domain, the solution can be obtained from (3.2) by the conformal mapping trans-

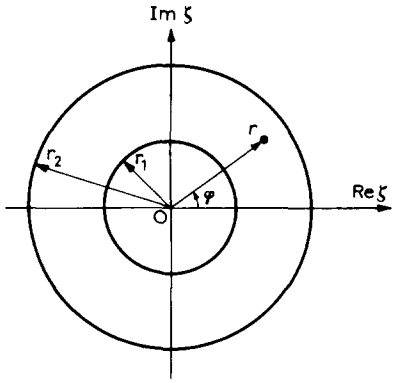

FIG. 3. The domain G: a ring in the complex ζ plane, $\zeta = r\,e^{i\varphi}$.

forming the ring to the domain considered.

Let C be the set of all complex numbers ζ. Denote by G the domain $\{\zeta\in C: r_1 < |\zeta| < r_2\}$, where r_1 and r_2 are positive real numbers (see Fig. 3). If a function $u(\zeta)$ is defined in G we shall use the notation $u(r,\varphi)$ for it where $u(r,\varphi) = u(\zeta)$ for $\zeta = re^{i\varphi}$.

Definition 1: A real continuous function $u(\zeta)$ defined on G is called harmonic if it has continuous partial derivatives of the second order and satisfies Laplace's equation.

Theorem 1: Let $g_1(\varphi)$ and $g_2(\varphi)$ be real continuous functions given for $0\leqslant\varphi\leqslant 2\pi$, and let $g_1(0) = g_1(2\pi)$ and $g_2(0) = g_2(2\pi)$. Then there exists a harmonic function $u(r,\varphi)$ on G continuous on the closure $\overline{G}$ and such that

$$u(r_1,\varphi) = g_1(\varphi) \quad \text{and} \quad u(r_2,\varphi) = g_2(\varphi) \quad \text{for all } \varphi. \tag{3.1}$$

The function u is unique and is given by

$$\begin{aligned} u(r,\varphi) = {} & \frac{1}{2\pi\ln(r_2/r_1)} \\ & \times\left[\ln\frac{r_2}{r}\int_0^{2\pi} g_1(\psi)\,d\psi + \ln\frac{r}{r_1}\int_0^{2\pi} g_2(\psi)\,d\psi\right] \\ & + \frac{1}{\pi}\int_0^{2\pi}\left[g_1(\psi)Q\left(\frac{r_1}{r},\frac{r}{r_2},\varphi-\psi\right)\right. \\ & \left. + g_2(\psi)Q\left(\frac{r}{r_2},\frac{r_1}{r},\varphi-\psi\right)\right]d\psi, \end{aligned} \tag{3.2}$$

where

$$Q(a,b,\varphi) = \sum_{n=1}^{\infty} a^n(b^{2n}-1)[(ab)^{2n}-1]^{-1}\cos n\varphi. \tag{3.3}$$

Proof of the theorem is given in Appendix A.

Remark 1: Certainly $Q(a,b,\varphi) = Q(1/a,1/b,\varphi)$ for $a\neq 0$, $b\neq 0$, and $|Q(a,b,\varphi)| < \infty$ for $|a| < 1, |ab^2| < 1$. The last statement follows from the inequality $|ab| = (|a|)^{1/2}(|ab^2|)^{1/2} < 1$.

Remark 2: Assume that the function $g_1(\varphi)$ is not continuous at $\varphi = 0$ but is integrable. Then $u(r,\varphi)$ given by the formula (3.2) is a harmonic function fulfilling $u(r,\varphi) = g_1(\varphi)$ for $\varphi\neq 0$. Define $f_n(\varphi) = g_1(\varphi)$ for $|g_1(\varphi)|\leqslant n$, $f_n(\varphi) = n$ for $g_1(\varphi) > n$ and $f_n(\varphi) = -n$ for $g_1(\varphi) < -n$, and $u_n(r,\varphi)$ the corresponding solutions given by the formula (3.2). Then $u(r,\varphi) = \lim_{n\to\infty} u_n(r,\varphi)$ with the exception of the point $r = r_1$, $\varphi = 0$.

Let C again denote the set of all complex numbers, and B a bounded region in C. Denote by $B_r(\zeta)$ the domain $\{w\in C: |w-\zeta|\leqslant r\}$, i.e., the disk of radius r around ζ.

Definition 2: A real continuous function $u(\zeta)$ defined on B is called subharmonic if $u(\zeta)\leqslant(1/2\pi)\int_0^{2\pi}u(\zeta + re^{i\varphi})\,d\varphi$ for $\zeta\in B$ and $r > 0$ such that $B_r(\zeta)B$.

We remind the reader that if $f(\zeta)$ is a holomorphic function on B then $|f(\zeta)|$ is a subharmonic function in $\subset B$.

Lemma (Principle of the maximum): Let $u(\zeta)$ be a subharmonic function in B. If a point $\zeta_0\in B$ exists such that $u(\zeta_0) = \max\{u(\zeta): \zeta\in\overline{B}\}$ then u is a constant, $u(\zeta) = u(\zeta_0)$.

This theorem can be found in standard textbooks; we therefore give only a brief sketch of the proof. Let $\zeta_0\in B$ be such a point. Then there exists $r_0 > 0$ such that $B_{r_0}(\zeta_0)\subset B$. Assume that a point $\zeta_1\in B_{r_0}(\zeta_0)$ exists such that $u(\zeta_1) < u(\zeta_0)$. The continuity implies that a $d > 0$ exists such that $u(\zeta) < \frac{1}{2}[u(\zeta_1) + u(\zeta_0)]$ for $\zeta\in B_d(\zeta_1)$. Setting $r_1 = |\zeta_1 - \zeta_0|$ we have $u(\zeta)\leqslant u(\zeta_0)$ for $\zeta\in B$ and $u(\zeta) < \frac{1}{2}[u(\zeta_1) + u(\zeta_0)]$ for $\zeta\in B_d(\zeta_1)$. This implies that

$$\frac{1}{2\pi}\int_0^{2\pi} u(\zeta_0 + r_1e^{i\varphi})\,d\varphi < u(\zeta_0),$$

which contradicts $u(\zeta_0)\leqslant(1/2\pi)\int_0^{2\pi}u(\zeta_0 + r_1e^{i\varphi})\,d\varphi$. Thus, it follows that $u(\zeta)$ is constant on $B_{r_0}(\zeta_0)$. Repeating this procedure, we successively prove that u is constant on the whole B.

Theorem 2: Let the function u be subharmonic in the ring G and continuous on the closure $\overline{G}$. Let v be harmonic in G and continuous on $\overline{G}$. If $u(\zeta)\leqslant v(\zeta)$ on the boundary of G, then $u(\zeta)\leqslant v(\zeta)$ in G.

As in Theorem 1, we defer the proof to Appendix A.

This is the mathematical background of our approach. In the following section, we use it to improve the bound (1.5) everywhere inside the ellipse given by (1.6). Theorem 2 is used to show that the new bound, which is a harmonic function in a doubly connected domain, is better than the old one, (1.5). Theorem 1 gives the explicit form of the harmonic function.

4. DERIVATION OF THE BOUND

The solution (3.2) to the Dirichlet boundary value problem for the ring G can be applied to any doubly connected domain D, provided that the corresponding conformal mapping between G and D has been carried out. In our case, the domain D which the interval of the ellipse $\gamma_2\cos\theta + i(\gamma_2^2 - 1)^{1/2}\sin\theta$ from which the interval $[-1,1]$ has been removed. We denote this domain by E; it is the set of points that is parametrized by (1.3) with $0\leqslant\theta < 2\pi$ and $1 < \gamma < \gamma_2$. Its inner and outer boundaries, (1.3) with $\gamma = 1$ and (1.3) with $\gamma = \gamma_2$, are denoted by E_1 and E_{γ_2}, respectively (see Fig. 4).

The ellipse E is mapped onto the ring G in the complex ζ plane, $\zeta = |\zeta|e^{i\varphi}$, $1 < |\zeta| < r_2$ (see Fig. 3) so that

$$\theta = \varphi,$$
$$\gamma = \tfrac{1}{2}(|\zeta| + |\zeta|^{-1}). \tag{4.1}$$

On E_1, the boundary value of the harmonic function is equal to the right-hand side of (1.1), i.e.,

$$g_1(\varphi) = A_1/|\sin\varphi|^{1/2}, \tag{4.2}$$

where

$$A_1 = C_1 s^{3/4}\ln^{3/2} s. \tag{4.3}$$

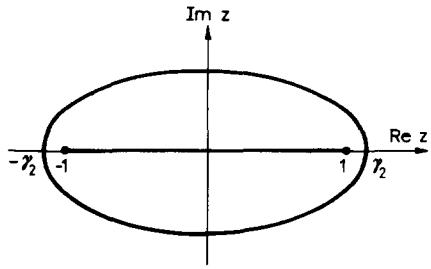

FIG. 4. The domain E: the set of points of the form $z = \gamma\cos\theta + i(\gamma^2 - 1)\sin\theta$ with $1 < \gamma < \gamma_2$. E is obtained from G (see Fig. 3) by the conformal mapping (4.1).

On E_{γ_2}, the function $g_2(\varphi)$ is the right-hand side of (1.5), i.e.,

$$g_2(\varphi) = M(\gamma_2), \tag{4.4}$$

where

$$M(\gamma_2) = C\frac{\delta - (\gamma_2 - 1)\Lambda}{[\delta - (\gamma_2^+ - 1)\Lambda]^{3/2}[\delta - (\gamma_2^- - 1)\Lambda]^{3/2}} \tag{4.5}$$

is independent of φ.

Since $f(s,z)$ is bounded, at sufficiently high s, by $g_1(\varphi)$ and $g_2(\varphi)$ on E_1 and E_{γ_2}, respectively, it must be bounded, being a subharmonic function, by the harmonic function $u(|\zeta|,\varphi)$ everywhere in the domain E. This is the content of Theorem 2. It applies to our situation at $s\to\infty$. Indeed, $g_2(\varphi) = M(\gamma_2)$ on E_{γ_2} and $g_1(\varphi) < M(1) = C(2a)^{-1}s\ln^2 s$ on E_1 except for asymptotically small neighborhoods of the foci. Thus, $u(|\zeta|,\varphi) < M(\gamma_2)$ holds everywhere except these neighborhoods. This proves that the new bound is asymptotically better than the original one.

Let us turn now to technical aspects of the derivation of the bound. The right-hand side of (3.2) is a linear combination of four integrals, which we denote by I_1, I_2, J_1, and J_2, respectively. Three of them can be easily calculated:

$$\begin{aligned} I_1 &= A_1(2/\pi)^{1/2}[\Gamma(\tfrac{1}{4})]^2, \\ I_2 &= 2\pi M(\gamma_2), \\ J_2 &= 0. \end{aligned} \tag{4.6}$$

The integral J_1 has the form of an infinite series

$$J_1(\varphi) = A_1\sum_{n=1}^{\infty}\alpha^n(\beta^{2n} - 1)[(\alpha\beta)^{2n} - 1]^{-1} \times\int_0^{2\pi}\cos[n(\varphi - \psi)]\frac{d\psi}{|\sin\psi|^{1/2}}, \tag{4.7}$$

where $\alpha = 1/r$ and $\beta = r/r_2$.

In this notation, the new bound on $|f(s,z)|$ can be written in the form

$$|f(s,z)| < u(r,\varphi), \tag{4.8}$$

where

$$u(r,\varphi) = \frac{1}{2\pi}\frac{1}{\ln r_2}\left[\left(\ln\frac{r_2}{r}\right)I_1 + (\ln r)I_2 + 2J_1(\varphi)\right], \tag{4.9}$$

where z is given by (1.3), $r = |\zeta|$, while γ and γ_2 are related to r and r_2, respectively, by the conformal mapping (4.1). The bound (4.8) is valid at all $1 \leqslant \gamma \leqslant \gamma_2$, where γ_2 is given by (1.6). Thus, γ_2 tends to 1 with increasing energy and, as a consequence, γ_2, r_2, and r also tend to 1. In other words, the ellipse E shrinks to E_1 and the ring G shrinks to the unit circle with increasing energy.

Thus, the main aim of this paper has been achieved. We have found a new bound, (4.8), which is better than the previous one, (1.5). On the other hand, one should spend some effort to simplify the expression (4.7) for J_1, in order to make the bound more transparent. The remaining part of this section is devoted to this task.

Evaluating the integrals on the right-hand side of (4.7) and performing some elementary operations, we can give Eq. (4.7) the form

$$J_1(\varphi) = 4\pi\sqrt{2}A_1\frac{\Gamma(\frac{3}{2})}{[\Gamma(\frac{3}{4})]^2}\sum_{k=1}^{\infty}\frac{\alpha^{2k}(1-\beta^{4k})}{1-(\alpha\beta)^{4k}}[\cos(2k\varphi)]\frac{(\frac{1}{4})_k}{(\frac{3}{4})_k}, \tag{4.10}$$

where we adopt the notation

$$(a)_k = \Gamma(a+k)/\Gamma(a).$$

In order to simplify further the expression for J_1, let us consider only its high-energy dominant part. Denoting

$$b_k(s) = (1-\beta^{4k})/[1-(\alpha\beta)^{4k}], \tag{4.11}$$

we notice that $\lim_{s\to\infty} b_k(s)$ is independent of k. Indeed, taking into account that $\lim_{s\to\infty}\alpha(s) = \lim_{s\to\infty}\beta(s) = 1$ and using the l'Hôpital rule, we obtain

$$\tilde{b} := \lim_{s\to\infty} b_k(s) = \beta'(s)/[\alpha(s)\beta'(s) + \alpha'(s)\beta(s)]. \tag{4.12}$$

The infinite sum in (4.10) can now be simplified as follows. Introducing the notation

$$\begin{aligned} a_k &= \alpha^{2k}, \\ b_k &= b_k(s), \end{aligned} \tag{4.13}$$

and

$$c_k = [\cos(2k\varphi)]\frac{(\frac{1}{4})_k}{(\frac{3}{4})_k},$$

we write (4.10) as

$$J_1(\varphi) = 4\pi\sqrt{2}A_1\frac{\Gamma(\frac{3}{2})}{[\Gamma(\frac{3}{4})]^2}\sum_{k=1}^{\infty}a_k b_k c_k. \tag{4.14}$$

Then, we use the following relations, which are proved in detail in Appendix B;

$$\lim_{s\to\infty}\left\{\sum_{k=1}^{\infty}a_k b_k c_k - \tilde{b}\sum_{k=1}^{\infty}a_k c_k\right\} = 0 \tag{4.15}$$

for $\varphi \neq 0, \pi$, and

$$\lim_{s\to\infty}\left\{\sum_{k=1}^{\infty}a_k b_k c_k\Big/\left(\tilde{b}\sum_{k=1}^{\infty}a_k c_k\right)\right\} = 1 \tag{4.16}$$

for $\tilde{b} = 1$ and $\varphi = 0$ or π.

This allows us to express the high-energy approximation to J_1 in terms of the hypergeometric function ${}_2F_1$. Introducing the symbol $\sigma = (r_2 - r)(r_2 - 1)^{-1}$,

$$J_1(\varphi) = 2\pi\sqrt{2}\sigma A_1\frac{\Gamma(\frac{3}{2})}{[\Gamma(\frac{3}{4})]^2}[{}_2F_1(1,\tfrac{1}{4};\tfrac{3}{4};\chi^2) + {}_2F_1(1,\tfrac{1}{4};\tfrac{3}{4};\chi^{*2})], \tag{4.17}$$

where $\chi = (1/r)e^{i\varphi} \equiv \alpha e^{i\varphi}$ and χ^* denotes the complex conjugate to χ. We take φ to be an arbitrary fixed, s-independent real number. As s tends to ∞, v approaches the unit circle

from inside. The functions ${}_2F_1(1,\frac{1}{4};\frac{3}{4};\omega)$ for $\omega=\chi^2$ and $\omega=\chi^{*2}$ have singularities on the circle, as can be seen from the following well-known formula[7]

$$ {}_2F_1(1,\tfrac{1}{4};\tfrac{3}{4};\omega)=(1-\omega)^{-1/2}\,{}_2F_1(-\tfrac{1}{4},\tfrac{1}{2};\tfrac{3}{4};\omega). \tag{4.18}$$

Note that the series ${}_2F_1(-\frac{1}{4},\frac{1}{2};\frac{3}{4};\omega)$ converges in the whole disk including the boundary circle and, in particular, also the point $\omega=1$.

Using (4.18) in (4.17) and inserting J_1 into (4.9), we obtain as $s\to\infty$

$$ u(r,\varphi)=(r_2-1)^{-1}\Big\{(r-1)M(\gamma_2) + \frac{[\Gamma(\frac{1}{4})]^2}{\pi(2\pi)^{1/2}}\,[1+F(r,\varphi)](r_2-r)A_1\Big\}, \tag{4.19}$$

where $F(r,\varphi)$ is shorthand notation for

$$ F(r,\varphi)={}_2F_1(-\tfrac{1}{4},\tfrac{1}{2};\tfrac{3}{4};e^{2i\varphi})(1-\alpha^2e^{2i\varphi})^{-1/2} + {}_2F_1(-\tfrac{1}{4},\tfrac{1}{2};\tfrac{3}{4};e^{-2i\varphi})(1-\alpha^2e^{-2i\varphi})^{-1/2}$$

and is finite at $r=1$. The original bounds (1.1) and (4.4) are obtained from (4.19) by setting $r=1$ and $r=r_2$, respectively, in which case the first and the second terms in (4.19) vanish, respectively.

We have already mentioned that both γ_2 and r_2 approach 1 from above with increasing s. As a consequence, γ and r must also tend to 1. Let us now consider this problem quantitatively; this will give us the high-energy behavior of the new bound $u(r,\varphi)$.

The high-energy behavior of γ_2 is given by (1.6). Using (4.1), we get the behavior of r_2:

$$ r_2\cong 1+(2a')^{1/2}s^{-1/2}\ln^{-1}s. \tag{4.20}$$

The high-energy behavior of $u(r,\varphi)$ is determined by r and γ. These two quantities, being smaller than r_2 and γ_2, respectively,

$$ 1\leqslant r\leqslant r_2, \qquad 1\leqslant\gamma\leqslant\gamma_2, \tag{4.21}$$

cannot approach unity slower than r_2 and γ_2, respectively, but may do so faster. This is up to our choice, and we can use this fact to make the bound $u(r,\varphi)$ asymptotically lower than the original $M(\gamma)$ given by formula (1.5). It is clear that $r-1$ must vanish faster than r_2-1 at $s\to\infty$; indeed, if $r-1$ behaves like r_2-1 then it follows from (4.19) that $u(r,\varphi)$ has the same behavior as $M(\gamma)$ and represents no improvement. Thus, assuming now that $r-1$ vanishes faster than r_2-1,

$$ r-1\cong\eta s^{-\kappa}\ln^{-\lambda}s, \tag{4.22}$$

with $\kappa>\frac{1}{2}$, $\lambda>1$, we obtain the following asymptotic form of $u(r,\varphi)$:

$$ u(r,\varphi)\cong A\eta s^{3/2-\kappa}\ln^{3-\lambda}s+B(\varphi)s^{3/4}\ln^{3/2}s, \tag{4.23}$$

where, for $\varphi\neq 0$ or π,

$$ A=\frac{1}{2\sqrt{2}}\,a^{1/2}a'^{-1/2}(a-a')^{-3/2}c, $$

$$ B(\varphi)=[\Gamma(\tfrac{1}{4})]^2 2^{-1/2}\pi^{-3/2}[1+F(1,\varphi)]C_1, \tag{4.24}$$

and η is an arbitrary positive constant.

For $\kappa=\lambda=0$, (4.23) represents no improvement in comparison with (1.5). For $0<\kappa<\frac{3}{4}$ and $0<\lambda<\frac{3}{2}$, the first term in (4.23) is dominant and $u(r,\varphi)$ gets asymptotically lower with increasing κ and/or λ. At the same time, however, the ellipse of validity of the improved bound (4.23) shrinks faster with increasing κ and/or λ, because of (4.22). The corresponding γ which determines the semimajor axis of the ellipse is obtained from (4.22) and (4.1)

$$ \gamma\cong 1+\tfrac{1}{2}\eta^2 s^{-2\kappa}\ln^{-2\lambda}s. \tag{4.25}$$

Consequently, by choosing the values of κ, λ, and η, we choose either a tight bound (4.23) with a narrow ellipse of validity ($\kappa=\frac{3}{4}$, $\lambda=\frac{3}{2}$), or a loose bound with a wider ellipse of validity (κ and λ close to zero). The slowest increase of the new bound (4.8), (4.23) is reached for $\kappa=\frac{3}{4}$ and $\lambda=\frac{3}{2}$. Then the bound has the form

$$ |f(s,z)|<[A\eta+B(\theta)]s^{3/4}\ln^{3/2}s, \quad \theta\neq 0,\pi, \tag{4.26}$$

with A and $B(\theta)$ given by (4.24). Its energy dependence is the same as that of (1.1); thus, the bound (4.26) represents an extension of the Froissart bound (1.1) to complex scattering angles.

The ellipse of validity of (4.26) shrinks as

$$ \gamma\cong 1+\tfrac{1}{2}\eta^2-s^{-3/2}\ln^{-3}s. \tag{4.27}$$

For a comparison see Table I.

5. DISCUSSION OF PERSPECTIVES FOR FURTHER IMPROVEMENT OF THE BOUND

Our bound (4.26) can be further improved by lowering the boundary value functions $g_1(\varphi)$ and $g_2(\varphi)$. Concerning $g_1(\varphi)$, let us note that we have used the right-hand side of (1.1) [call it $B_1(\varphi)$; note that $\theta=0$ according to (4.1)] for $g_1(\varphi)$,

TABLE I. Survey of ellipses in the complex z plane.

	Semimajor axis	Radius of the circle according to (4.1)
Lehmann–Martin ellipse [validity domain of partial wave expansion (2.1)]	$\gamma_L=1+\frac{a}{s}$	$r_L=1+(2a)^{1/2}/s$
Interior of E_{γ_2} [validity domain of (1.5)]	$\gamma_2=1+a's^{-1}\ln^{-2}s$	$r_2=1+(2a')^{1/2}s^{-1/2}\ln^{-1}s$
Interior of E_γ [validity domain of (4.26)]	$\gamma=1+\frac{1}{2}\eta^2 s^{-3/2}\ln^{-3}s$	$r=1+\eta s^{-3/4}\ln^{-3/2}s$

although for any given energy $s^{1/2}$ there is a small interval of angles, $\epsilon(s)$, around the forward and backward directions ($\sin\theta = 0$) for which the right-hand side of (1.2), B_2, becomes smaller than $B_1(\theta)$. One should therefore take for $g_1(\varphi)$ the function

$$B_1(\varphi)\Theta(B_2 - B_2(\varphi)) + B_2\Theta(B_1(\varphi) - B_2), \tag{5.1}$$

where Θ is the step function. As (5.1) is always smaller than or equal to $B_1(\varphi)$, we reach by this an improvement of the bound (4.26) for every s.

The function $g_2(\varphi)$, which is the right-hand side of (4.4), offers several possibilities of improvement. To see this, let us notice that, in deriving (1.5), Eqs. (2.1)–(2.4) play the decisive role. Formally, we can derive a bound on $f(s,z)$ immediately from (2.4):

$$|f(s,z)| \leqslant \frac{1}{\pi}\int_\Gamma |g(s,w)|\ |w^2 - 2wz + 1|^{-1/2}\, dw. \tag{5.2}$$

This suggests that the value of the right-hand side depends on three factors: on the estimate of $|g(s,w)|$, on that of $|w^2 - 2wz + 1|^{-1/2}$ and on the choice of the integration curve Γ, which has only its end points fixed. We discuss these three possibilities.

Let us start with the estimate of $g(s,w)$. Following the approach of Ref. 3, we derived[1] a bound on $g(s,w)$ by making only the assumption of the polynomial boundedness (2.5) for $f(s,z)$ in the Lehmann–Martin ellipse. Our result (2.6) is to be compared with $|g(s,w)| < Cs^{N+3}$ of Ref. 3.

Then, by using (2.6), we transform the right-hand side of (5.2) to

$$\frac{C_2}{\pi}\int_{-\pi/2}^{\pi/2} [b - \Lambda(b_0 + b_1\sin x + b_2\sin^2 x)^{1/2}]^{-2}\, dx, \tag{5.3}$$

where

$$b = 1 + R/\ln s,$$
$$b_0 = \gamma^2 - \sin^2\theta$$
$$b_1 = 2\gamma(\gamma^2 - 1)^{1/2},$$
$$b_2 = \gamma^2 - \cos^2\theta,$$

and where the integration curve Γ was chosen to be the straight line as depicted in Fig. 2. Finally, we majorize (5.3) by another integral,

$$\frac{C_2}{\pi}\int_{-\pi/2}^{\pi/2} [b - \Lambda(b_0 + (\gamma^2 - 1)^{1/2}\sin x)]^{-2} dx. \tag{5.4}$$

This integral, being evaluated analytically, leads to formula (1.5), on which our final result (4.26) is based.

As we have already pointed out there are two other ways leading to an improvement of our bound (4.26): to choose a more approriate integration curve Γ, and to make a better estimate of the integrand. The former method leads to a variational problem of choosing the optimal curve Γ_{opt} lying in a domain and having its end points fixed. We do not have much to say concerning this problem, although we believe that, if solved, it may lead to a considerable improvement of the result.

Finally, we want to show how a better estimate of the integral (5.3) can be found, assuming that the curve Γ has been chosen to be the straight line (Fig. 2). It can be proved (by exploiting the properties of a second order polynomial) that the expression

$$A + B\sin x + C\sin^2 x, \tag{5.5}$$

with

$$A = (b - \Lambda\gamma)^2 + d,$$
$$B = -2\Lambda(b - \Lambda\gamma)(\gamma^2 - 1)^{1/2},$$
$$C = \Lambda^2(\gamma^2 - 1) - d,$$

where

$$0 \leqslant d \leqslant b(\gamma^- - \Lambda)\sin^2\theta,$$

is bounded from above and from below by the denominator of the integrand of (5.3) and (5.4), respectively. As a consequence, the integral

$$\frac{C_2}{\pi}\int_{-\pi/2}^{\pi/2} (A + B\sin x + C\sin^2 x)^{-1}\, dx \tag{5.6}$$

leads to a better bound than (5.4). Explicit calculation leads to the expression

$$|f(s,z)| < M(\gamma_2)(1 + p|\sin\theta|)^{-1/2}, \tag{5.7}$$

with

$$p = \delta^2/[\delta - (\gamma^- - 1)\Lambda].$$

Thanks to positivity of p, (5.7) represents a lower bound than (4.5) and, consequently, would also lead to a better estimate than (4.26). Needless to say, a further improvement can be achieved by means of explicit numerical evaluation of (5.3) and then invoking the Dirichlet problem according to Sec. 3.

The case of particles with spin can be analogously treated and the same improvement by use of the Dirichlet problem, as performed in the present paper, can be carried our for each helicity amplitude as described in Ref. 1.

6. CONCLUDING REMARKS

Rigorous bounds on scattering amplitude have been important as general, model-independent consequences of the principles of local field theory. From the practical point of view, they have served as guidance for model building, not to be violated in different calculational schemes, and the like. They can be also used to stabilize the extrapolation of experimental information to regions which are not experimentally accessible.

The fixed-angle Froissart bound (1.1) is rather low compared with other rigorous bounds. It is therefore interesting that its extension (4.26) to the ellipse (4.27), although smaller than the original one, around the physical interval $[-1.1]$, is at all possible. Of course, we obtained this result thanks to the assumption (2.5) that the scattering amplitude is polynomially bounded at high energies everywhere in the Lehmann–Martin ellipse, a feature that has not been proved from axiomatic field theory. Similar assumptions, however, have been frequently made in the past, to obtain results that were later proved on a more rigorous basis.

Let us also mention that, on the other hand, it is not excluded that an iteration procedure might help in finding better and better bounds.

A serious obstacle to applications might be the rather fast shrinkage, (4.27), of the ellipse of validity of our bound.

Indeed, as it is seen from Table I, it is only the Lehmann–Martin ellipse E_{γ_L} that shrinks to E_1 at the same rate as do the singularities (poles and cuts) in the t-plane. It is therefore desirable to try to obtain a bound valid in a larger domain than E_γ, which is given by (4.27).

The authors are convinced that, using the methods discussed in Sec. 5, it is possible to obtain a still better bound $g_2(\varphi)$ at the points of E_{γ_2} and, thereby, to achieve further improvement of the bound (4.26) as well as an extension of its ellipse of validity (4.27).

ACKNOWLEDGMENT

It is our pleasure to thank V. Glaser for useful suggestions and stimulating discussions.

APPENDIX A: PROOFS OF THEOREMS 1 AND 2 OF SEC. 3

Proof of Theorem 1: The functions $g_1(\varphi)$ and $g_2(\varphi)$ can be expressed, in the sense of convergence in L_2, in the form

$$g_i(\varphi) = \sum_{n=0}^{\infty} (u_i^{(n)} \cos n\varphi + v_i^{(n)} \sin n\varphi), \quad i = 1,2.$$

We seek the solution in the form

$$u(r,\varphi) = \sum_{n=0}^{\infty} u_n(r,\varphi),$$

where $u_n(r,\varphi)$ are harmonic functions in G fulfilling

$$u_n(r_i,\varphi) = u_i^{(n)} \cos n\varphi + v_i^{(n)} \sin n\varphi, \quad i = 1,2, \tag{A1}$$

for $n = 0,1,2\cdots$.

Using the fact that the real part of a holomorphic function is harmonic we put

$$u_0(r,\varphi) = \mathrm{Re}(a_0 + b_0 \ln z),$$
$$u_n(r,\varphi) = \mathrm{Re}(a_n z^n + b_n z^{-n}), \quad n \geqslant 1. \tag{A2}$$

The relations (A1) imply a set of linear equations for a_n and b_n that allow us to give (A2) the form

$$u_0(r,\varphi) = \frac{1}{2\pi} \frac{1}{\ln(r_2/r_1)} \times \left(\ln\frac{r}{r_1} \int_0^{2\pi} g_2(\varphi)\, d\varphi + \ln\frac{r_2}{r} \int_0^{2\pi} g_1(\varphi)\, d\varphi \right),$$

$$u_n(r,\varphi) = \frac{1}{\pi} \int_0^{2\pi} g_1(\psi) \left(\frac{r_1}{r}\right)^n \left[\left(\frac{r}{r_2}\right)^{2n} - 1\right] \times \left[\left(\frac{r_1}{r_2}\right)^{2n} - 1\right]^{-1} \cos n(\varphi - \psi)\, d\psi + \frac{1}{\pi} \int_0^{2\pi} g_2(\psi) \left(\frac{r}{r_2}\right)^n \left[\left(\frac{r_1}{r}\right)^{2n} - 1\right] \times \left[\left(\frac{r_1}{r_2}\right)^{2n} - 1\right]^{-1} \cos n(\varphi - \psi)\, d\psi.$$

First we prove that $\Sigma_{n=0}^{\infty} u_n(r,\varphi)$ exists and fulfills our assumptions under the condition that the functions $g_1(\varphi)$ and $g_2(\varphi)$ have continuous derivatives. In this case, $nu_1^{(n)},\ldots,nv_2^{(n)}$ are the Fourier coefficients of the derivatives $g_1'(\varphi)$ and $g_2'(\varphi)$ and, because of the Parseval identity

$$\frac{1}{2\pi} \int_0^{2\pi} [g_1'(\varphi)]^2\, d\varphi = \sum_{n=0}^{\infty} n^2 [(u_1^{(n)})^2 + (v_1^{(n)})^2],$$
$$\frac{1}{2\pi} \int_0^{2\pi} [g_2'(\varphi)]^2\, d\varphi = \sum_{n=0}^{\infty} n^2 [(u_2^{(n)})^2 + (v_2^{(n)})^2].$$

We can show now that the functions $\Sigma_{n=0}^{N} u_n(r_1,\varphi)$ and $\Sigma_{n=0}^{N} u_n(r_2,\varphi)$ converge, respectively, for $N \to \infty$ to $g_1(\varphi)$ and to $g_2(\varphi)$ uniformly with respect to φ. The convergence is proved with the help of the following inequalities:

$$\left| \sum_{n=0}^{N} u_n(r_i,\varphi) - g_i(\varphi) \right| < \sum_{N+1}^{\infty} |u_i^{(n)} \cos n\varphi + v_i^{(n)} \sin n\varphi|$$
$$< \sum_{N+1}^{\infty} ((u_i^{(n)})^2 + (v_i^{(n)})^2)^{1/2} = \sum_{N+1}^{\infty} n((u_i^{(n)})^2 + (v_i^{(n)})^2)^{1/2} \frac{1}{n}$$
$$< \left(\sum_{N+1}^{\infty} n^2 [(u_i^{(n)})^2 + (v_i^{(n)})^2] \right)^{1/2} \left(\sum_{N+1}^{\infty} \frac{1}{n^2} \right)^{1/2}, i = 1,2,$$

where the last expression converges to zero with $N \to \infty$.

The principle of the maximum for harmonic functions implies that $\Sigma_{n=0}^{N} u_n(r,\varphi)$ converges to $u(r,\varphi) = \Sigma_{n=0}^{\infty} u_n(r,\varphi)$ uniformly on G so that $u(r,\varphi)$ is the harmonic function in G continuous on $\bar{G}$ and fulfilling the given boundary conditions.

In the general case we can approximate the continuous functions $g_1(\varphi)$ and $g_2(\varphi)$ by a sequence of functions $g_1^{(s)}(\varphi)$ and $g_2^{(s)}(\varphi)$, respectively, such that

$$\lim_{s \to \infty} \sup_{\varphi} |g_i(\varphi) - g_i^{(s)}(\varphi)| = 0, \quad i = 1,2, \tag{A3}$$

where $g_1^{(s)}(\varphi)$ and $g_2^{(s)}(\varphi)$ have continuous derivatives. Denote by $u^{(s)}(r,\varphi)$ the corresponding solution of the boundary value problem. The principle of the maximum together with the relation (A3) implies again that $u^{(s)}(r,\varphi)$ converges to a harmonic function $u(r,\varphi)$ fulfilling the boundary conditions with $g_1(\varphi)$ and $g_2(\varphi)$. Relation (A3) and Remark 1 yield that

$$\int_0^{2\pi} |g_i^{(s)}(\psi) - g_i(\psi)|\, d\psi \to 0$$

and

$$\int_0^{2p} |g_i^{(s)}(\psi) - g_i(\psi)|\ |Q(a,b,\varphi - \psi)|\, d\psi \to 0$$

for $s \to \infty$ if $|a| < 1$, $|b| < 1$. This implies that formula (3.2) is valid in the general case, too. This completes the proof of Theorem 1.

Proof of Theorem 2: The function $w(z) = u(z) - v(z)$ is subharmonic in G, continuous on $\bar{G}$, and satisfies

$$w(z) \leqslant 0 \tag{A4}$$

on the boundary of G. Assume there exists a point z_0 such that $w(z_0) > 0$. Then, $\max\{w(z): z \in \bar{G}\} > 0$, and a point $z_1 \in \bar{G}$ exists such that $w(z_1) = \max\{w(z): x \in \bar{G}\}$. Because of (A4) the point z_1 certainly cannot be on the boundary of G such that $z_1 \in G$. We can apply the principle of the maximum and obtain a constant $w(z)$, $w(z) = w(z_1) > 0$, which contradicts (1).

APPENDIX B: PROOF OF RELATIONS (4.15) AND (4.16)

Let $\alpha(s)$ and $\beta(s)$ be functions defined for $s > 0$ and having continuous derivatives. Assume that

$$\lim_{s\to\infty} \beta'(s)/[\alpha'(s)+\beta'(s)] = \tilde{b}$$

is finite, and that

$$0\leqslant\alpha(s)\leqslant 1,$$
$$0\leqslant\beta(s)\leqslant 1,$$

$\beta(s)$ is nondecreasing, and

$$\lim_{s\to\infty} \alpha(s) = \lim_{s\to\infty} \beta(s) = 1.$$

Theorem 3: For $\alpha(s)$ and $\beta(s)$ defined above, the relation (4.15) holds for $\sin\psi\neq 0$, where a_k, b_k, and c_k are defined by (4.13).

Proof: First of all, we notice the following asymptotic formula for $k\to\infty$:

$$\left|\frac{(\frac{1}{4})_k}{(\frac{3}{4})_k} - \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}\frac{1}{\sqrt{k}}\right| < \frac{1}{64}k^{-5/2} \tag{B1}$$

(see Ref. 8). This formula indicates that the series $\Sigma_{k=0}^{\infty} c_k$ is not absolutely convergent.

In the following, we shall need the following three statements on infinite series:

Statement 1:

$$\sum_{i=k}^{n} A_i B_i = S_n B_n - \sum_{i=k}^{n-1} S_i(B_{i+1} - B_i),$$

where $S_l = \Sigma_{i=k}^{l} A_i$.

Statement 2: Let $\Sigma_{i=1}^{n} A_i$ be uniformly bounded and $\{B_i\}$ a monotonic sequence with $\lim_{i\to\infty} B_i = 0$. Then $\Sigma_{i=1}^{\infty} A_i B_i$ is convergent.

Statement 3: Let $\Sigma_{i=1}^{\infty} A_i$ be convergent and $\{B_i\}$ a monotonic sequence with a finite $\lim_{i\to\infty} B_i$. Then $\Sigma_{i=1}^{\infty} A_i B_i$ is also convergent.

Lemma 1: The series $\Sigma_{k=1}^{\infty} c_k$ and $\Sigma_{k=1}^{\infty} a_k c_k$ are convergent if $\sin\psi\neq 0$.

Proof: Certainly,

$$C_k = \frac{(\frac{1}{4})_k}{(\frac{3}{4})_k}\cos(2k\psi) = H_k + L_k,$$

where

$$H_k = \left[\frac{(\frac{1}{4})_k}{(\frac{3}{4})_k} - \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}\frac{1}{(k)^{1/2}}\right]\cos(2k\psi),$$

$$L_k = \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}\frac{1}{(k)^{1/2}}\cos(2k\psi).$$

The sum $\Sigma_{k=1}^{\infty} H_k$ is absolutely convergent because of (B1). The sum $\Sigma_{k=1}^{\infty} L_k$ is a convergent series because $k^{-1/2}$ is a monotonic sequence and

$$\sum_{k=1}^{n}\cos(2k\psi) = [\sin(2n+1)\psi - \sin\psi](2\sin\psi)^{-1}$$

are bounded for $\sin\psi\neq 0$. So, using Statement 2 we obtain that $\Sigma_{\infty}^{k=1} c_k$ is convergent. Then, using Statement 3 we obtain that $\Sigma_{k=1}^{\infty} a_k c_k$ is convergent since $\{a_k\}$ is a monotonic sequence [see (4.13)].

Define

$$S_k = \sum_{i=1}^{k} a_i c_i,$$

$$S = \sum_{i=1}^{\infty} a_i c_i.$$

Lemma 2: There exist four constants d_1, d_2, d_3, and d_4 such that

$$|S_n - S_k| \leqslant \frac{d_1}{k+1} + \frac{d_2\alpha^{2k+2}}{(k+1)^{1/2}|\sin\psi|} \quad \text{for} \quad 1\leqslant k\leqslant n, \tag{B2}$$

and

$$|S_k| \leqslant d_3 + d_4/|\sin\psi|. \tag{B3}$$

Proof:

$$S_n - S_k = \sum_{j=k+1}^{n} a_j c_j = \sum_{j=k+1}^{n} a_j\left[\frac{(\frac{1}{4})_j}{(\frac{3}{4})_j} - \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}\frac{1}{(j)^{1/2}}\right]\cos(2j\psi) + \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}\sum_{j=k+1}^{n}\frac{a_j}{(j)^{1/2}}\cos(2j\psi).$$

Using relation (B1) we can estimate the first term on the right hand side of (B2) by $\Sigma_{j=k}^{n}\alpha^{2j}cj^{-5/2}$. Denote

$$Z_k = \sum_{j=1}^{k}\cos(2j\psi) = (\sin[(2k+1)\psi] - \sin\psi)/(2\sin\psi).$$

Applying Statement 1 we obtain

$$\sum_{j=k+1}^{n} a_j j^{-1/2}\cos(2j\psi) = (Z_n - Z_k)\frac{a_n}{(n)^{1/2}} - \sum_{j=k}^{n-1}(Z_j - Z_k)\left(\frac{a_{j+1}}{(j+1)^{1/2}} - \frac{a_j}{(j)^{1/2}}\right)$$
$$= \frac{\sin[(2n+1)\psi] - \sin[(2k+1)\psi]}{2\sin\psi}\frac{a_n}{(n)^{1/2}} - \sum_{j=k}^{n-1}\frac{\sin[(2j+1)\psi] - \sin[(2k+1)\psi]}{2\sin\psi} \times\left(\frac{a_{j+1}}{(j+1)^{1/2}} - \frac{a_j}{(j)^{1/2}}\right).$$

Since the sequence $\{a_j/(j)^{1/2}\}$ is monotonic we obtain

$$\left|\sum_{j=k+1}^{n}\frac{a_j}{(j)^{1/2}}\cos(2j\psi)\right| \leqslant c\alpha^{2k+2}(k+1)^{-1/2}(\sin\psi)^{-1}.$$

By this, the inequality (B2) is proved. The inequality (B3) is proved analogously. We have

$$S_k = \sum_{j=1}^{k} a_j\left[\frac{(\frac{1}{4})_j}{(\frac{3}{4})_j} - \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})(j)^{1/2}}\right]\cos(2j\psi) + \left[\frac{\sin[(2k+1)\psi] - \sin\psi}{2\sin\psi}\frac{a_k}{\sqrt{k}} - \sum_{j=1}^{k-1}\frac{\sin[(2j+1)\psi] - \sin\psi}{2\sin\psi} \times\left(\frac{a_{j+1}}{(j+1)^{1/2}} - \frac{a_j}{(j)^{1/2}}\right)\right]\frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}.$$

Since $|a_j| = |a^{2j}| < 1$, relation (B3) is proved, too.

Lemma 3: The sequence $\{b_k(s)\}$ is monotonically in-

creasing in k for any s. Further, we have

$$\lim_{k\to\infty} b_k(s) = 1 \tag{B4}$$

and

$$\lim_{s\to\infty} b_k(s) = \tilde{b}. \tag{B5}$$

Proof: Relations (B4) follows from the fact that $0<\beta(s)<1$. (B5) follows from the l'Hôpital rule. The fact that $b_k(s)$ is monotonic,

$$b_k(s)<b_{k+1}(s) \tag{B6}$$

is equivalent to

$$(1-\beta^{4k})(1-\beta^{4k+4})^{-1}$$
$$<[1-(\alpha\beta)^{4k}][1-(\alpha\beta)^{4k+4}]^{-1}.$$

As $0<\alpha<1$, it is sufficient to prove that

$$h(x) = (1-x^k)/(1-x^{k+1})$$

is decreasing in x for all $x\in(0.1)$. To prove this, it is sufficient to notice that

$$h'(x) = x^{k-1}(1-x^{k+1})^{-2}(1-x)$$
$$\times(x+x^2+\cdots+x^k-k)<0.$$

Thus, (B6) holds. Lemma 3 is proved.

We are now in a position to prove Theorem 3. Because of Statement 1, we can write

$$\sum_{k=1}^{n} a_k c_k(b_k-\tilde{b})$$
$$= S_n(s)[b_n(s)-\tilde{b}] - \sum_{k=1}^{n-1} S_k(s)[b_{k+1}(s)-b_k(s)].$$

This equality can be rewritten in the following way:

$$\sum_{k=1}^{n} a_k c_k(b_k-\tilde{b}) = S_n(s)[b_n(s)-\tilde{b}]$$
$$-\sum_{k=1}^{n_0-1} S_k(s)[b_{k+1}(s)-b_k(s)]$$
$$-\sum_{k=n_0}^{n-1} [S_k(s)-S_n(s)][b_{k+1}(s)-b_k(s)]$$
$$-S_n(s)[b_n(s)-b_{n_0}(s)],$$

where $1<n_0<n-1$. Finally,

$$\sum_{k=1}^{n} a_k c_k(b_k-\tilde{b})$$
$$= S_n(s)[b_{n_0}(s)-\tilde{b}] - \sum_{k=1}^{n_0-1} S_k(s)[b_{k+1}(s)-b_k(s)]$$
$$-\sum_{k=n_0}^{n-1} [S_k(s)-S_n(s)][b_{k+1}(s)-b_k(s)].$$

Because of Lemma 2, the values of $S_k(s)$ are uniformly bounded for $\sin\psi\neq 0$. We have

$$\left|\sum_{k=1}^{n} a_k c_k(b_k-\tilde{b})\right| < (d_3+d_4/|\sin\psi|)|b_{n_0}(s)-\tilde{b}|$$
$$+\sum_{k=1}^{n_0-1}\left|d_3+\frac{d_4}{|\sin\psi|}\right|\cdot|b_{k+1}(s)-b_k(s)|$$
$$+\sum_{k=n_0}^{n-1}\left|\frac{d_1}{n_0+1}+\frac{d_2\alpha^{2n_0+2}}{(n_0+1)^{1/2}|\sin\psi|}\right|\cdot|b_{k+1}(s)-b_k(s)|.$$

Since $\{b_k(s)\}$ is a monotonic sequence we obtain

$$\left|\sum_{k=1}^{n} a_k c_k(b_k-\tilde{b})\right| \leqslant \left(d_3+\frac{d_4}{|\sin\psi|}\right)|b_{n_0}(s)-\tilde{b}|$$
$$+\left(d_3+\frac{d_4}{|\sin\psi|}\right)[b_{n_0}(s)-b_1(s)]$$
$$+\left(\frac{d_1}{n_0+1}+\frac{d_2\alpha^{2n_0+2}}{(n_0+1)^{1/2}|\sin\psi|}\right)[b_n(s)-b_{n_0}(s)].$$

Applying the Statements 2 and 3 we see that $\Sigma_{k=1}^{\infty} a_k c_k(b_k-\tilde{b})$ is a convergent series. We have

$$\left|\sum_{k=1}^{\infty} a_k c_k(b_k-\tilde{b})\right| \leqslant \left(d_3+\frac{d_4}{|\sin\psi|}\right)$$
$$\times[|b_{n_0}(s)-\tilde{b}|+b_{n_0}(s)-b_1(s)]$$
$$+\left(\frac{d_1}{n_0+1}+\frac{d_2\alpha^{2n_0+2}}{(n_0+1)^{1/2}|\sin\psi|}\right)[1-b_{n_0}(s)].$$

Let $s\to\infty$ and n_0 be fixed; then

$$\limsup_{s\to\infty}\left|\sum_{k=1}^{\infty} a_k c_k(b_k-\tilde{b})\right|$$
$$\leqslant\left(\frac{d_1}{n_0+1}+\frac{d_2\alpha^{2n_0+2}}{(n_0+1)^{1/2}|\sin\psi|}\right)(1-\tilde{b}).$$

Since the number n_0 can be arbitrary, the relation (4.15) is proved. By this, the first part of the Theorem 3 is proved.

It is now easy to prove also the second part, i.e., relation (4.16). Because of Lemma 3, we have

$$0<b_1(s)\leqslant b_k(s)\leqslant 1$$

and

$$b_1(s)\leqslant\left[\sum_{k=1}^{\infty}\alpha^{2k}b_k\left(\tfrac{1}{4}\right)_k/\left(\tfrac{3}{4}\right)_k\right]\left[\sum_{k=1}^{\infty}\alpha^{2k}\left(\tfrac{1}{4}\right)_k/\left(\tfrac{3}{4}\right)_k\right]^{-1}\leqslant 1.$$

Since $\lim_{s\to\infty} b_1(s)=\tilde{b}=1$, relation (4.16) is proved. This completes the proof of the Theorem 3.

[1] M. Chaichian, J. Fischer, S. M. Molodensky, N. F. Nelipa, and I. Vrkoč, J. Math. Phys. **22**, 380 (1981).

[2] M. Froissart, Phys. Rev. **123**, 1053 (1961); A. Martin, Phys. Rev. **129**, 1432 (1963).

[3] T. Kinoshita, J. J. Loeffel, and A. Martin, Phys. Rev. Lett. **10**, 460 (1964).

[4] R. Courant, *Dirichlet's Principle, Conformal Mapping and Minimal Surfaces* (Interscience, New York, 1950).

[5] We are indebted to C. Cronström for drawing our attention to Ref. 6 and for useful discussions on this point.

[6] N. I. Akhiezer, *Elementy Teorii Ellipticheskikh Funktsii* (in Russian) (Nauka, Moscow, 1970).

[7] H. Bateman, *Higher Transcedental Functions*, (McGraw-Hill, New York, 1953), Vol. I; I. S. Gradshteyn and I. W. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965).

[8] M. Abramowitz and Irene A. Stegun (Eds.), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Applied Mathematics Series -55, (National Bureau of Standards, Washington, DC, 1964).

Conditions of existence for a class of self-dual solutions of the Einstein equations

N. Caderni
Institute for Theoretical Physics, University of California, Santa Barbara, California 93111

M. Martellini
Istituto Nazionale di Fisica Nucleare, Sezione di Pavia, Pavia I-27100, Italy and Istituto di Fisica dell'Università, Milano I-20133, Italy

F. Salmistraro
Istituto Nazionale di Fisica Nucleare, Sezione di Pavia, and Istituto di Fisica dell'Università, Pavia I-27100, Italy

(Received 17 December 1979; accepted for publication 17 September 1982)

We study the integrability of the Einstein equations for a class of empty homogeneous space-times (Bianchi class A), once the self-duality constraints are imposed on the space-time itself. This system is integrable in the case of Bianchi type I and is a subset of Bianchi types VI_0 and VII_0. Bianchi type II space-times do not admit self-dual solutions and for the case of Bianchi VIII and IX we were unable to find a general integrability condition.

PACS numbers: 04.20.Jb

Until very recently relatively little work has been done in vacuum homogeneous anisotropic solutions of the Einstein equations. Although we know that these solutions exist for any Bianchi types,[1] they were regarded for their mathematical significance only. The interest in the field has been raised in connection with the Euclidean approach to quantum gravity[2] where, analogous to the Yang–Mills case, the main contribution to the vacuum-to-vacuum transition amplitudes is supposed to come from the vacuum field configurations which are stationary phase points for the classical action. The analogy with Yang–Mills instantons is then carried further and brings special importance to metrics whose curvature is self-dual. As a matter of fact it has been shown that self-dual (or anti-self-dual) metrics are local minima in the action amongst vacuum metrics.

The aim of this paper is to study the constraints imposed by the self-duality conditions over the vacuum Bianchi class A solutions of the Einstein equations. To this purpose we write down the field equations and the self-duality conditions separately and we investigate the integrability of the Einstein equations under these particular constraints.

In what follows we start with a pseudo-Riemann space-time which admits a simply transitive, three-parameter group of motion G_3 acting on spacelike hypersurfaces. There exists an orthonormal tetrad $\{\mathbf{e}_a\}$, where $\mathbf{e}_0$, which is normal to the spacelike hypersurface, can be identified with the four velocity $\mathbf{U}$. We can therefore define the "electric" and "magnetic" parts of the Weyl tensor C^a_{bcd}, E_{ab}, and H_{ab}, respectively, according to

$$E_{ac} \equiv C_{abcd} U^b U^d, \quad E_{ab} = E_{ba},$$
$$E^a_a = 0, \quad E_{ab} U^b = 0, \tag{1a}$$
$$H_{ac} \equiv {}^*C_{abcd} U^b U^d, \quad H_{ab} = H_{ba},$$
$${}^*C_{abcd} \equiv \tfrac{1}{2}\sqrt{-g}\,\epsilon_{ablm} C_{cd}{}^{lm},$$
$$H_a{}^a = 0, \quad H_{ab} U^b = 0, \tag{1b}$$

$$C_{ab}{}^{cd} = 2U_{(a}E_{b)}{}^{(c}U^{d)} + \delta^{(c}_{(a}E^{d)}_{b)} - \sqrt{-g}(\epsilon_{abmn} U^m H^{n(c} U^{d)} + \bar{\epsilon}^{cdmn} U_m H_{n(a} U_{b)}),$$

where g is the metric and $(\,)$ denotes antisymmetrization. We can perform a dual transformation on the structure constants of the group of motion C^a_{bc}, obtaining the quantities C^{ad} defined by $C^c_{ab} = \epsilon_{abd} C^{dc}$, where ϵ_{abd} is the three-dimensional Levi–Civita tensor.

We can also profit from the "tensor" properties of C^{ad}, dividing it in its symmetric and antisymmetric parts: $C^{ad} = n^{ad} + \epsilon^{adc} a_c$. The following canonical choices are always allowed:

(a) to take the vector $\{\mathbf{e}_a\}$ as eigenvectors of the symmetric tensor $n_{\alpha\beta}$;

(b) to reduce n^{ad} to the principal axes, i.e., $\mathrm{diag}(n^{ad}) = (n_1, n_2, n_3)$ and to assume $a_c = (a,0,0)$, with $an_1 = 0$. The Bianchi solutions in which $\mathbf{a} = 0$ are known as Bianchi–Behr class A solutions and in the following discussion we shall restrict ourselves to this case. (There are the Bianchi types I, II, VI_0, VII_0, VIII, and IX.)

At this point we make use of the following relation, valid for pseudo-Riemannian manifolds:

$$(C + i{}^*C)_{abcd} = (g_{abpq} + i\eta_{abpq})(g_{cdrs} + i\eta_{cdrs}) \times U^p U^r (E + iH)^{qs}, \tag{2}$$

where

$$\eta_{abcd} \equiv \sqrt{-g}\,\epsilon_{abcd},$$
$$g_{abcd} \equiv (g_{ac}g_{db} - g_{ad}g_{bc}). \tag{3}$$

The Weyl tensor is self-dual, i.e., $-\mathbf{C} = i{}^*\mathbf{C}$, if $\mathbf{E} = -i\mathbf{H}$. On Euclidean section this condition therefore becomes $\mathbf{E} = \mathbf{H}$. We point out that in the following these equations are to be regarded as integrability conditions on the existence of the Euclidean solution of the field equations which are taken in the same form of the pseudo-Riemannian case. For

 0022-2488/83/020300-03$02.50

present purposes, we will not investigate the problem of the existence and form of such solutions.

The Einstein equations for vacuum Bianchi class A space-times,[3] when written in the orthonormal basis, read[4] (here and henceforth we use geometrical units and we set the cosmological constant $\Lambda = 0$)

$$\begin{aligned} \dot{n}_1 &= -(\theta_2 + \theta_3 - \theta_1)n_1, \\ \dot{n}_2 &= -(\theta_1 + \theta_3 - \theta_2)n_2, \\ \dot{n}_3 &= -(\theta_1 + \theta_2 - \theta_3)n_3, \end{aligned} \tag{4}$$

$$\begin{aligned} \dot{\theta}_1 &= -\theta\theta_1 - \tfrac{1}{2}n_1^2 + \tfrac{1}{2}(n_2 - n_3)^2, \\ \dot{\theta}_2 &= -\theta\theta_2 - \tfrac{1}{2}n_2^2 + \tfrac{1}{2}(n_1 - n_3)^2, \\ \dot{\theta}_3 &= -\theta\theta_3 - \tfrac{1}{2}n_3^2 + \tfrac{1}{2}(n_2 - n_1)^2, \end{aligned} \tag{5}$$

where the dot represents differentiation with respect to the Euclidean time; θ_{ab} is the expansion, $\theta = \theta_a{}^a$, and $\mathrm{diag}(\theta_{ab}) = (\theta_1,\theta_2,\theta_3)$. We shall search for the integrability conditions of Eqs. (4) and (5) under the following constraints.

The first one is the condition

$$\theta_1\theta_2 + \theta_2\theta_3 + \theta_3\theta_1 = \tfrac{1}{4}(n_1^2 + n_2^2 + n_3^2 - 2n_1n_2 - 2n_2n_3 - 2n_3n_1), \tag{6}$$

which represents the Hamiltonian constraint.

The Euclidean self-duality constraints read

$$-\tfrac{1}{2}n_1^2 + \tfrac{1}{2}(n_2 - n_3)^2 - \theta\theta_1 + \theta_1^2 = -\tfrac{3}{2}n_1\sigma_{11} + \tfrac{1}{2}(n_2 - n_3)(\theta_2 - \theta_3), \tag{7a}$$

$$-\tfrac{1}{2}n_2^2 + \tfrac{1}{2}(n_3 - n_1)^2 - \theta\theta_2 + \theta_2^2 = -\tfrac{3}{2}n_2\sigma_{22} + \tfrac{1}{2}(n_3 - n_1)(\theta_3 - \theta_1), \tag{7b}$$

$$-\tfrac{1}{2}n_3^2 + \tfrac{1}{2}(n_1 - n_2)^2 - \theta\theta_3 + \theta_3^2 = -\tfrac{3}{2}n_3\sigma_{33} + \tfrac{1}{2}(n_1 - n_2)(\theta_1 - \theta_2), \tag{7c}$$

where σ_{ab} denotes the shear tensor. Equations (7) are an explicit writing of the conditions $E_{aa} = H_{aa}$. Since the tetrad vectors satisfying the condition $n^{ab}a_b = 0$ can be chosen as eigenvectors of the shear tensor σ_{ab}, we obtain $\sigma_{ab} = 0$ whenever $a \neq b$, and $\sigma_{ii} = \theta_i - \frac{1}{3}\theta$. We observe also that $H_{ab} = E_{ab} = 0$ whenever $a \neq b$ and, since both **H** and **E** are trace-free tensors, only two out of the three constraints (7) are independent [e.g., (7a) and (7b)].

To study the integrability conditions of the system consisting of Eqs. (4)–(7) we were unable to find a general algorithm and the following discussion considers separately all the different cases.

TYPE I

The eigenvectors n_a are $n_1 = n_2 = n_3 = 0$.

The self-duality conditions read

$$\theta_1(\theta_2 + \theta_3) = 0, \tag{8a}$$

$$\theta_2(\theta_3 + \theta_1) = 0, \tag{8b}$$

Notice that in this simple case the Hamiltonian constraint (6) coincides with the sum of Eqs. (8). In order to prove the integrability of the system consisting of Eqs. (4), (5), (6), and (8), we differentiate Eq. (8) with respect to time and we obtain

$$\begin{aligned} \dot{\theta}_1(\theta_2 + \theta_3) + \theta_1(\dot{\theta}_2 + \dot{\theta}_3) &= 0, \\ \dot{\theta}_2(\theta_3 + \theta_1) + \theta_2(\dot{\theta}_3 + \dot{\theta}_1) &= 0, \end{aligned} \tag{9}$$

Using the field equations (5) which now read

$$\begin{aligned} \dot{\theta}_1 &= -\theta\theta_1, \\ \dot{\theta}_2 &= -\theta\theta_2, \end{aligned}$$

and substituting into Eqs. (9) we reproduce the self-duality constraints (8). The self-dual solutions are then found as the solution of the system formed by Eqs. (8) together with the first integral (6) and they are

$$\theta_a \neq 0, \quad \theta_b = \theta_c = 0, \tag{10}$$

so that there is only one of the expansion factors different from zero. The nonzero expansion factor is found by direct integration of the corresponding equation (5).

TYPE II

The eigenvectors are $n_1 \neq 0$, $n_2 = n_3 = 0$. The independent self-duality conditions read

$$-\tfrac{1}{2}n_1^2 - \theta_1(\theta_2 + \theta_3) = -n_1\theta_1 + \tfrac{1}{2}n_1(\theta_2 + \theta_3), \tag{11a}$$

$$\tfrac{1}{2}n_1^2 - \theta_2(\theta_3 + \theta_1) = -\tfrac{1}{2}n_1(\theta_3 - \theta_1). \tag{11b}$$

Let us define α as

$$\theta_1\theta_2 + \theta_2\theta_3 + \theta_3\theta_1 \equiv \alpha = \tfrac{1}{4}n_1^2. \tag{12}$$

Using Eq. (12) we can eliminate n_1^2 from Eqs. (11a) and (11b), obtaining

$$2\alpha + \theta_1(\theta_2 + \theta_3) = n_1[\theta_1 - \tfrac{1}{2}(\theta_2 + \theta_3)], \tag{13a}$$

$$2\alpha - \theta_2(\theta_1 + \theta_3) = \tfrac{1}{2}n_1(\theta_1 - \theta_3). \tag{13b}$$

We can demonstrate that, in order for Eqs. (13) to be compatible with the field equations (4) and (5), the matrix of the coefficients of n_1 (formally considered as unknown) has to be of rank 1. In particular,

$$(\theta_1 - \theta_3) \neq 0$$

or otherwise we would get $\alpha < 0$.

The conditions to be satisfied in order for Eqs. (13) to be solved are given by equating to zero the matrix of the coefficients and the known terms; that is, after some manipulations

$$(\theta_2 - \theta_3)(\theta_1^2 + \alpha) = 0,$$

which implies $(\theta_2 - \theta_3) = 0$;

that is,

$$\alpha = 2\theta_1\theta_3 + \theta_3^2. \tag{14a}$$

Putting Eq. (14) into Eqs. (13), we obtain

$$n_1 = \frac{2(3\theta_1\theta_3 + \theta_3^2)}{(\theta_1 - \theta_3)}. \tag{14b}$$

At this juncture, as in the former Bianchi type I case, in order to prove integrability differentiate the constraints (14b) with respect to time and use the field equations (4) and (5). We thus get the new constraints

$$(\theta_1 + \theta_3) = 0, \tag{15}$$

and substituting Eq. (14a) we arrive at the expression

$$\alpha = -\theta_3^2 < 0,$$

which contradicts Eq. (12). This incompatibility implies that there are no self-dual type II solutions.

TYPE VI_0 AND VII_0

The eigenvectors n_a are $n_3 = 0$, $n_1, n_2 \neq 0$. The independent self-duality constraints are

$$-\tfrac{1}{2}n_1^2 + \tfrac{1}{2}n_2^2 - \theta\theta_1 + \theta_1^2 = -\tfrac{3}{2}n_1\theta_1 + \tfrac{1}{2}n_1\theta + \tfrac{1}{2}n_2(\theta_2 - \theta_3), \tag{16a}$$

$$-\tfrac{1}{2}n_2^2 + \tfrac{1}{2}n_1^2 - \theta\theta_2 + \theta_2^2 = -\tfrac{3}{2}n_2\theta_2 + \tfrac{1}{2}n_2\theta + \tfrac{1}{2}n_1(\theta_3 - \theta_1). \tag{16b}$$

Let us define α as

$$\theta_1\theta_2 + \theta_2\theta_3 + \theta_3\theta_1 \equiv \alpha = \tfrac{1}{4}(n_1 - n_2)^2. \tag{17}$$

Adding together Eqs. (16a) and (16b), we get

$$\alpha + \theta_1\theta_2 = \tfrac{1}{2}(\theta_2 - \theta_1)(n_2 - n_1) \tag{18}$$

and squaring, we obtain an algebraic constraint that depends on θ_i only and reads

$$(\alpha + \theta_1\theta_2)^2 = \alpha(\theta_2 - \theta_1)^2. \tag{19}$$

By use of Eq. (18) and the field equation (4) it is possible to notice that the following properties are valid:

(p1) $\alpha = 0 \Leftrightarrow \theta_1 = \theta_2 = 0$,

(p2) $\alpha \neq 0 \Leftrightarrow (\theta_1 - \theta_2) \neq 0$ or $(\alpha + \theta_1\theta_2) \neq 0$.

Moreover, in the (p1) case the Hamiltonian constraint (17) reads $n_1 = n_2 \neq 0$ and in the (p2) case $(n_1 - n_2) \neq 0$.

The existence of these two properties provides a natural way to classify the several possibilities.

Case (p1): $n_2 = n_1 \neq 0$; $\theta_1 = \theta_2 = 0$. These last conditions satisfy identically the self-duality constraints (16).

The integrability conditions are obtained again by time differentiation of Eqs. (16), substitution by Eqs. (4) and (5), together with the new constraint $\theta_1 = \theta_2 = 0$. We find that the system so obtained is completely integrable, by imposition of the new Hamiltonian constraint which under the property (p1) reads $n_1 = n_2$. In this case both the Hamiltonian constraints $n_2 = n_1 \neq 0$ and $\theta_1 = \theta_2 = 0$ and the self-duality conditions are first integrals of Eqs. (4) and (5) and define a family of self-dual solutions. These solutions are explicitly obtainable by integration of the equation for n_1, n_2, and θ_3.

Case (p2): $n_1 - n_2 \neq 0$, $(\theta_2 - \theta_1) \neq 0$, $\alpha > 0$. From Eq. (18) we get

$$n_2 = n_1 + \frac{2(\alpha + \theta_1\theta_2)}{\theta_2 - \theta_1}.$$

Substituting this expression for n_2 into Eq. (16a) we get

$$n_1[2(\alpha + \theta_1\theta_2) - (\theta_2 - \theta_1)^2] = (\alpha + \theta_2^2)(\theta_1 - \theta_3). \tag{20}$$

By time differentiation of Eq. (18) we get

$$2n_1(\theta_2 - \theta_1)F = -2(\alpha + \theta_1\theta_2)F - \theta_3(\theta_1 + \theta_2)^2(\theta_2 - \theta_1), \tag{21}$$

where F is defined by

$$F \equiv 2(\alpha + \theta_1\theta_2) + 2\alpha - \tfrac{1}{2}(\theta_2 - \theta_1)^2.$$

By use of Eq. (19) it is possible to show that $2(\alpha + \theta_1\theta_2) - (\theta_2 - \theta_1)^2 \neq 0$. Therefore, we can express n_1, with the help of Eq. (20), as

$$n_1 = \frac{(\alpha + \theta_2^2)(\theta_1 - \theta_3)}{2(\alpha + \theta_1\theta_2) - (\theta_2 - \theta_1)^2}.$$

From the explicit expressions for n_1 and n_2 and from the hypothesis (p2), we derive that all the terms $(\theta_1 - \theta_3)$, $(\theta_1 + \theta_2)$, etc., are different from zero.

Substituting the last expression for n_1 into Eq. (21) we get

$$2F(\theta_1\theta_2 - \theta_3^2) = -\theta_3(\theta_1 + \theta_2)[2(\alpha + \theta_1\theta_2) - (\theta_2 - \theta_1)^2]. \tag{22}$$

Comparing Eq. (22) with Eq. (19) we observe that the new constraint is given by

$$4\theta_3^2 - (\alpha + \theta_1\theta_2) = 0. \tag{23}$$

Now, the integrability condition should be given by differentiation of Eq. (23) and use of Eqs. (4), (5), (18), and (19). If we do so we obtain $\alpha = 0$, which contradicts the hypothesis $\alpha > 0$. This incongruence shows that there are no self-dual solutions in the (p2) case.

TYPE VIII AND IX

The eigenvectors are $n_1, n_2, n_3 \neq 0$. The self-duality conditions are the complete Eqs. (7).

Our purpose here, to find the integrability conditions of the system (4)–(7), could not be attained by any of the manipulations we tried. The manipulations of this work allowed us to identify all of the self-dual solutions and only those for any of the empty homogeneous space times here considered. In the Bianchi type VIII and IX case we could not find such an existence theorem. Our procedure was of an essentially algebraic nature. However, in this last case our failure to find the integrability does not mean that such solutions do not exist. As a matter of fact we know of a self-dual Bianchi type IX class of solutions by Belinsky *et al.*,[5] whose existence, however, has been demonstrated by direct construction.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the referee for valuable comments and suggestions.

[1] A. H. Taub, Ann. Math. **53**, 472 (1951).

[2] S. W. Hawking, *Cargèse Lecture in Physics, 1978*, edited by M. Levy and S. Deser (Gordon and Breach, New York, 1978).

[3] G. G. Behr, F. B. Estabrook, and H. D. Wahlquist, J. Math. Phys. **9**, 497 (1968).

[4] G. F. R. Ellis and M. A. H. McCallum, Commun. Math. Phys. **12**, 108 (1969).

[5] V. A. Belinskii, G. W. Gibbons, D. N. Page, and C. N. Pope, Phys. Lett. B **76**, 433 (1978).

On Riemannian spaces with conformal symmetries or a tool for the study of generalized Kaluza–Klein theories

Osvaldo M. Moreschi[a] and George A. J. Sparling
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 24 June 1982; accepted for publication 30 September 1982)

The Riemann geometry of a space with conformal symmetries is written in terms of intrinsic objects defined from the action of the symmetries. Its application in the study of generalized Kaluza–Klein theories is discussed.

PACS numbers: 04.40.Ky

1. INTRODUCTION

Sixty one years ago Kaluza suggested the unification of gravitation and electromagnetism by means of considering a five-dimensional Riemannian space.[1] The idea of eliminating the dependence of the fifth dimension by imposing a symmetry on it was of considerable interest in the following years.[2]

If one extends the Kaluza–Klein theories to spaces of higher dimensions, then one naturally incorporates nonabelian gauge fields.[3]

One starts with a Riemannian manifold of $4+q$ dimensions which is assumed to be the cross product of the four-dimensional space-time with a q-dimensional compact manifold. A Lie Group acts on the last space transitively usually as a group of actual symmetries. In general, the dimension of the Lie group has been considered to be the same as that of the compact manifold, but lately the suggestion of having a group of greater dimension has been raised.[4]

The introduction of extra dimensions has caught much attention among the supersymmetry theorist, as for example in the construction of the $N=8$ supergravity theory by Cremmer, Julia, and Scherk.[5]

With these sort of ideas in mind one was led to look for suitable expressions related to the geometrical ideas just mentioned. More concretely the expressions of the Riemann tensor for a space with conformal symmetries will be discussed below, where linear dependence of the conformal Killing vectors will be allowed.

In spite of the somewhat specialized motivations of this work, all the expressions are of a geometrical character only, in particular they are developed around the skeleton of the Riemann geometry. So for example the expressions of the Ricci tensor calculated here, offer a useful tool for the study of space-times with symmetries in general relativity.

In Sec. 2 the action of the conformal Lie group G on the manifold P with metric g is used to separate the tangent bundle of P as a direct sum of two vector bundles, one (usually called vertical) is spanned by the fundamental vector fields[6] corresponding to the symmetries, and the other (called horizontal) is the orthogonal complement of the former. This split of the tangent space could be thought of as providing a generalized connection, analogous to the Yang–Mills connection. And correspondingly the "generalized" curvature is introduced here (the abstract index notation[7] is extensively used in the horizontal space).

In Sec. 3 the definition of objects introduced before is extended. Also a linear connection is presented mainly with the purpose of abbreviating the subsequent expressions.

The Riemann connection, Riemann and Ricci tensors, and the Ricci scalar are expressed in Sec. 4 in terms of the intrinsic geometrical language presented before.

An extensive list of appendices is added mainly for the purpose of completeness for future reference, and to give also a more "coordinate approach" of the notation used in this work.

2. LIE TRANSFORMATION GROUP ON A MANIFOLD WITH METRIC

This section is intended mainly to describe the notations that are going to be used in this work. Let P be a C^∞-differentiable manifold of dimension p, G be a Lie transformation group on P; denote by $\hat{g}$ its q'-dimensional Lie algebra, and by $\hat{\nabla}_{\alpha'}$ the elements of a base of $\hat{g}$. Use $\nabla_{\alpha'}$ to represent the fundamental vector field on P corresponding to $\hat{\nabla}_{\alpha'}$. Assume that the space V spanned by the $\nabla_{\alpha'}$ at the point $u\in P$ has dimension q, $\forall u\in P$. Then the assignment $u\to V_u$ is an involutive distribution[8] denoted by G_0.

The local Frobenius theorem[9] tells us that we can find an open set $U\subset P$ isomorphic to $A\times B$, where A is an integral manifold of G_0; that is given $u=(u_1,u_2)\in U$, $u_1\in A$, $u_2\in B$ we have $V_u=T_{u_1}(A)$, and B is a $(p-q)$-dimensional submanifold of P.

Define π: $U\to B$ by $\pi(u)=u_2$.

Let g be a nondegenerate metric on P, with $g_{\alpha'\beta'}\equiv g(\nabla_{\alpha'},\nabla_{\beta'})$ being a matrix of rank q, $\forall u\in P$; then

Theorem 1: Given a vector field $\mathbf{v}$ at $\pi(u)\in B$, $\mathbf{f}\in C^\infty(B)$; $\exists$ a unique $\nabla_{\mathbf{v}}\in T_u(P)$ such that

(a) $\nabla_{\mathbf{v}}(\pi^*\mathbf{f})=\pi^*(\mathbf{v}(\mathbf{f}))$,

(b) $g(\nabla_{\alpha'},\nabla_{\mathbf{v}})=0$.

One calls $\nabla_{\mathbf{v}}$ the lift of $\mathbf{v}$.

Proof: in Appendix A.

In this way, for the case in which G is a conformal Lie group, one is providing a "generalized" connection in the sense that for any $u\in U\subset P$ one has defined a subspace $H_u\subset T_u(P)$ such that

(a) $T_u(P)=V_u\oplus H_u$.

[a] On leave of absence from Instituto de Matemáticas Astronomía y Física, Universidad Nacional de Córdoba, (5000) Córdoba, Argentina.

 0022-2488/83/020303-08$02.50

(b) If h is an element of the component of G connected to the identity $H_{uh} = (R_h)_* H_u$, where R_h is the transformation of P induced by $h \in G$, $R_h u = uh$,

(c) H_u depends differentiably on u.

Later (b) will be shown to be obvious.

In particular, if G were acting freely on P, there would be cases in which P is a principal fiber bundle over B and where one would have defined a connection in P.

If $v \in T_u(P)$ one can write

$$v = v_V + v_H = v_V + \nabla_{\pi_*(v)} . \tag{2.1}$$

Let $a,b,c,\ldots$ be abstract indices in B. Define v^a by

$$v^a \equiv \pi_*(v)^a , \tag{2.2}$$

then it is natural to define ∇_a by

$$v^a \nabla_a \equiv \nabla_{\pi_*(v)} . \tag{2.3}$$

Similarly if ω is a 1-form in B, define θ^a: $T^*_{\pi(u)}(B) \to T^*_u(P)$ by

$$\omega_a \theta^a \equiv \theta_\omega \equiv \pi^*(\omega) . \tag{2.4}$$

In general, any tensor in B will be denoted by T; that is, $\mathbf{v}^a \in T(B)$, $\omega_a \in T^*(B)$, $\mathbf{f} \in C^\infty(B)$. See Appendix B for more discussion involving this notation.

Note that

$$0 = \mathcal{L}_{\nabla_{a'}} \pi^*(\omega) = \omega_a \mathcal{L}_{\nabla_{a'}} \theta^a \quad \forall \omega_a ,$$

so

$$\mathcal{L}_{\nabla_{a'}} \theta^a = 0 , \tag{2.5}$$

where $\mathcal{L}$ denoted Lie derivative.

Now consider G to be a Lie group of conformal transformation; that is,

$$[\nabla_{\alpha'} , \nabla_{\beta'}] = C_{\alpha'\beta'}{}^{\sigma'} \nabla_{\sigma'} , \tag{2.6}$$

$$\mathcal{L}_{\nabla_{a'}} g = \lambda_{a'} g . \tag{2.7}$$

Then

$$0 = \mathcal{L}_{\nabla_{a'}} g(\nabla_{\beta'} , \nabla_{\mathbf{v}}) = g(\nabla_{\beta'} , \mathcal{L}_{\nabla_{a'}} \nabla_{\mathbf{v}})$$

so

$$\mathcal{L}_{\nabla_{a'}} \nabla_{\mathbf{v}} = w^a \nabla_a \quad \text{for some } w^a ,$$

but

$$0 = \mathcal{L}_{\nabla_{a'}} (\nabla_{\mathbf{v}} (\pi^* \mathbf{f})) = (\mathcal{L}_{\nabla_{a'}} \nabla_{\mathbf{v}})(\pi^* \mathbf{f}) = w^a \nabla_a (\pi^* \mathbf{f}) , \quad \forall \mathbf{f}$$

so $w^a = 0$,

and

$$0 = \mathcal{L}_{\nabla_{a'}} \nabla_{\mathbf{v}} = \nabla_{a'} \mathbf{v}^a \nabla_a - \mathbf{v}^a \nabla_a \nabla_{a'} = \mathbf{v}^a [\nabla_{a'} , \nabla_a] \quad \forall \mathbf{v}^a$$

so

$$[\nabla_{a'} , \nabla_a] = 0 . \tag{2.8}$$

This is equivalent to last property (b).

By taking $U \subset P$ small enough it is possible to express

$$\nabla_{a'} = b_{a'}{}^{\alpha} \nabla_\alpha , \quad \alpha' = 1,2,\ldots,q' \tag{2.9a}$$

$$b_\alpha{}^\beta = \delta_\alpha{}^\beta , \quad \alpha = 1,2,\ldots,q . \tag{2.9b}$$

Given the vector fields $\mathbf{v}^a$ and $\mathbf{w}^b$ at $\pi(u)$, define at u the "curvature"

$$R_{\mathbf{v},\mathbf{w}} \equiv [\nabla_{\mathbf{v}} , \nabla_{\mathbf{w}}] - \nabla_{[\mathbf{v},\mathbf{w}]} . \tag{2.10}$$

From the discussion in Appendix C, one gets that $R_{\mathbf{v},\mathbf{w}}$ is of the form

$$R_{\mathbf{v},\mathbf{w}} = \mathbf{v}^a \mathbf{w}^b R_{ab}{}^{\alpha} \nabla_\alpha . \tag{2.11}$$

3. EXTENDED DEFINITION OF ∇_a

In this section the action of ∇_a is defined to act on quantities with lower case latin indices.

One requires the following properties:

(a) $\mathbf{v}^a \mathbf{w}^b [\nabla_a , \nabla_b](f) = R_{\mathbf{v},\mathbf{w}}(f)$,

where $\mathbf{v}$ and $\mathbf{w}$ are vector fields in B, $f \in C^\infty(P)$.

(b) $\nabla_a g_{bc} = 0$,

where

$$g_{bc} \equiv g(\nabla_b , \nabla_c) .$$

Theorem 2: There exist a unique ∇_a satisfying (a) and (b).

Proof: Here one uses a construction that is going to be useful later.

Let a torsion-free connection $\mathbf{D}_a$ be given on B. Then if $\mathbf{x}$ is a vector field and $\mathbf{T}$ a tensor in B, one denotes the covariant derivative of $\mathbf{T}$ with respect to $\mathbf{x}$ by

$$\mathbf{D}_{\mathbf{x}} \mathbf{T} ,$$

where

$$\mathbf{D}_{\mathbf{x}} \equiv \mathbf{x}^a \mathbf{D}_a .$$

Now define $\hat{D}_a$ such that if v is a vector field in P and $f \in C^\infty(P)$ one has

(I) $\hat{D}_a(f) \equiv v^a \hat{D}_a(f) = v^a \nabla_a(f)$,

(II) $\hat{D}_a(v^{\mathscr{A}} \theta_{\omega\mathscr{A}}) = \hat{D}_a(v^b) \omega_b + v^b \mathbf{D}_a(\omega_b)$,

where $\mathscr{A}, \mathscr{B}, \mathscr{C}$ denote abstract indices in P. Note that using the notation of Appendix B one can easily calculate $\hat{D}_a(v^b)$. Now define $\hat{\gamma}_a{}^b{}_c$ by

$$\nabla_a(v^b) \equiv \hat{D}_a(v^b) + \hat{\gamma}_a{}^b{}_c v^c . \tag{3.1}$$

From Appendix D one can see that with this definition one has

$$R_{\mathbf{v},\mathbf{w}}(f) = \mathbf{v}^a \mathbf{w}^a \{ [\nabla_a , \nabla_b] + (\hat{\gamma}_a{}^c{}_b - \hat{\gamma}_b{}^c{}_a) \nabla_c \}(f) ,$$

so from condition (a) one gets

$$\hat{\gamma}_a{}^c{}_b = \hat{\gamma}_b{}^c{}_a . \tag{3.2}$$

And condition (b) gives

$$\nabla_a g_{bc} = \hat{D}_a g_{bc} - \hat{\gamma}_a{}^e{}_b g_{ec} - \hat{\gamma}_a{}^e{}_c g_{be} = 0 .$$

Defining $\hat{\gamma}_{abc} \equiv \hat{\gamma}_a{}^e{}_c g_{eb}$ one can easily see that the solution is

$$\hat{\gamma}_{acb} = \tfrac{1}{2} (\hat{D}_a g_{bc} + \hat{D}_b g_{ac} - \hat{D}_c g_{ab}) . \tag{3.3}$$

Similarly one extends the definition of the $\nabla_{\alpha'}$'s to act on quantities with latin small indices by requiring

$$\mathcal{L}_{\nabla_a}(v^a \nabla_a) = \nabla_\alpha(v^a) \nabla_a . \tag{3.4}$$

Let the indices $A,B,C,\ldots = \alpha,\beta,\sigma,\ldots; a,b,c$; that is, if $v \in T_u(P)$,

$$v = v^A \nabla_A = v^\alpha \nabla_\alpha + v^a \nabla_a . \tag{3.5}$$

Now define the linear connection D' by

$$D'_A(\nabla_B) = 0 , \tag{3.6a}$$

$$D'_A(v^B) = \nabla_A(v^B) , \tag{3.6b}$$

$$D'_A(f) = \nabla_A(f), \quad f \in C^\infty(P). \tag{3.6c}$$

Note that up to now the index that denote the tensor character of ∇_A has been omitted; that is, if v is a vector field in P, one would write

$$v^{\mathscr{A}} = v^A \nabla_A{}^{\mathscr{A}} = v^\alpha \nabla_\alpha{}^{\mathscr{A}} + v^a \nabla_a{}^{\mathscr{A}}.$$

From the discussion in Appendix E and G, one gets

$$[D'_a, D'_b] = R_{ab}{}^\alpha D'_\alpha + \underset{\sim}{R}_{abC}{}^D I^C{}_D, \tag{3.7a}$$

$$[D'_a, D'_\beta] = \underset{\sim}{R}_{\alpha\beta C}{}^D I^C{}_D, \tag{3.7b}$$

$$[D'_\alpha, D'_\beta] = \mathbb{C}_{\alpha\beta}{}^\sigma D'_\sigma, \tag{3.7c}$$

where the derivation I^C_D is defined by

$$I^C_D(v^{\mathscr{A}}) = v^C \nabla_D{}^{\mathscr{A}}, \quad I^C_D(v^E) = \delta_D{}^E v^C \tag{3.8a}$$

$$I^C_D(f) = 0 \quad \text{for } f \in C^\infty(P), \tag{3.8b}$$

and $\underset{\sim}{R}_{ABC}{}^D$ is the curvature of the connection just defined and

$$\mathbb{C}_{\alpha'\beta'}{}^{\sigma} \equiv C_{\alpha'\beta'}{}^{\delta'} b_{\delta'}{}^{\sigma}. \tag{3.9}$$

Defining

$$C_{AB}{}^C = \mathbb{C}_{\alpha\beta}{}^\sigma \quad \text{if } A,B,C \text{ are greek indices,} \tag{3.10a}$$

$$C_{AB}{}^C = R_{ab}{}^\sigma \quad \text{if } A \text{ and } B \text{ are latin and } C \text{ is greek,} \tag{3.10b}$$

and

$$C_{AB}{}^C = 0 \quad \text{otherwise} \tag{3.10c}$$

one can write

$$[D'_A, D'_B] = C_{AB}{}^C D'_C + \underset{\sim}{R}_{ABC}{}^D I^C_D. \tag{3.11}$$

4. RIEMANNIAN CONNECTION IN P

In this section a Riemann connection D is introduced. One first defines θ^α by

$$\theta^\alpha{}_{\mathscr{A}} \nabla^{\mathscr{A}}_\beta = \delta^\alpha{}_\beta, \tag{4.1a}$$

$$\theta^\alpha{}_{\mathscr{A}} \nabla^{\mathscr{A}}_a = 0, \tag{4.1b}$$

which implies

$$\mathscr{L}_{\nabla_a} \theta^\beta = \theta^\sigma \mathbb{C}_{\sigma\alpha}{}^\beta. \tag{4.2}$$

Now one can write

$$g = g_{\alpha\beta} \theta^\alpha \otimes \theta^\beta + g_{ab} \theta^a \otimes \theta^b \equiv g_{AB} \theta^A \otimes \theta^B. \tag{4.3}$$

As usual one uses g to raise and lower indices. It can easily be found that

$$\nabla_a g_{\beta\sigma} = \lambda_a g_{\beta\sigma} + \mathbb{C}_{a\beta}{}^\delta g_{\delta\sigma} + \mathbb{C}_{a\sigma}{}^\delta g_{\delta\beta}, \tag{4.4a}$$

$$\nabla_a g_{ab} = \lambda_a g_{ab}, \tag{4.4b}$$

$$\nabla_c g_{ab} = 0. \tag{4.4c}$$

These and other relations are discussed in Appendix F.

Let v and w be vector fields in P, then one will express the covariant derivative of w respect to v by

$$D_v w = v^A D_A w.$$

In particular, one wants to extend the definition of D_A to act on quantities with small latin index. This is done by defining

$$D_A \equiv D'_A + \gamma_A{}^B{}_C I^C{}_B, \tag{4.5}$$

where $\gamma_A{}^B{}_C$ is uniquely determined by the requirement of D_A to be a Riemann connection.

The equations to be satisfied are (see Appendix G)

$$\gamma_A{}^C{}_B - \gamma_B{}^C{}_A = C_{AB}{}^C, \tag{4.6a}$$

$$g_{DA}\gamma_C{}^D{}_B + g_{DB}\gamma_C{}^D{}_A = \nabla_C g_{AB}, \tag{4.6b}$$

with solution

$$\gamma_{ACB} = \tfrac{1}{2}(\nabla_A g_{BC} + \nabla_B g_{AC} - \nabla_C g_{AB} + C_{ABC} + C_{CBA} + C_{CAB}). \tag{4.7}$$

Explicitly one has

$$\gamma_\alpha{}^\sigma{}_\beta = \tfrac{1}{2}(\mathbb{C}_{\alpha\beta}{}^\sigma + \mathbb{C}_\alpha{}^\sigma{}_\beta + \mathbb{C}_\beta{}^\sigma{}_\alpha + \lambda_\alpha \delta_\beta{}^\sigma + \lambda_\beta \delta_\alpha{}^\sigma - \lambda^\sigma g_{\alpha\beta}), \tag{4.8a}$$

$$\gamma_\alpha{}^a{}_\beta = -(g^{ab}/2)\nabla_b g_{\alpha\beta}, \tag{4.8b}$$

$$\gamma_a{}^\alpha{}_\beta = (g^{\alpha\sigma}/2)\nabla_a g_{\sigma\beta}, \tag{4.8c}$$

$$\gamma_\alpha{}^\beta{}_a = (g^{\beta\sigma}/2)\nabla_a g_{\alpha\sigma}, \tag{4.8d}$$

$$\gamma_a{}^\alpha{}_b = \tfrac{1}{2}(-\lambda^\alpha g_{ab} + R_{ab}{}^\alpha), \tag{4.8e}$$

$$\gamma_\alpha{}^a{}_b = \tfrac{1}{2}(\lambda_\alpha \delta^a{}_b + R^a{}_{b\alpha}), \tag{4.8f}$$

$$\gamma_a{}^b{}_\alpha = \tfrac{1}{2}(\lambda_\alpha \delta_a{}^b - R_a{}^b{}_\alpha), \tag{4.8g}$$

$$\gamma_a{}^b{}_c = 0. \tag{4.8h}$$

Let u, v, and w be vector fields in P. Then one expresses the curvature tensor by

$$R(u,v)w \equiv \{[D_u, D_v] - D_{[u,v]}\} w = R_{ABC}{}^D u^A v^B w^C \nabla_D, \tag{4.9}$$

where (see Appendix G),

$$R_{AB\sigma}{}^D = \nabla_A(\gamma_B{}^D{}_C) - \nabla_B(\gamma_A{}^D{}_C) + \gamma_B{}^E{}_C \gamma_A{}^D{}_E - \gamma_A{}^E{}_C \gamma_B{}^D{}_E - C_{AB}{}^E \gamma_E{}^D{}_C + \underset{\sim}{R}_{ABC}{}^D. \tag{4.10}$$

Explicitly, one has

$$\begin{aligned} R_{\alpha\beta\sigma}{}^\delta = {} & \tfrac{1}{2} g^{eb} g^{\rho\delta} \nabla_e g_{\sigma[\alpha} \nabla_b g_{\beta]\rho} \\ & + \tfrac{1}{2}(\lambda^\epsilon \lambda_\epsilon \delta^\delta{}_{[\beta} g_{\alpha]\sigma} + \lambda_\sigma \lambda_{[\beta} \delta_{\alpha]}{}^\delta + \lambda^\delta \lambda_{[\alpha} g_{\beta]\sigma}) \\ & + \delta^\delta{}_{[\beta} \nabla_{\alpha]} \lambda_\sigma - g_{\sigma[\beta} \nabla_{\alpha]} (\lambda_\rho) g^{\rho\delta} \\ & + (\lambda\rho/2) g_{\sigma[\beta} \mathbb{C}_{\alpha]}{}^{\rho\delta} + (\lambda\rho/2) g_{\sigma[\beta} \mathbb{C}_{\alpha]}{}^{\delta\rho} \\ & - (\lambda\rho/2) \mathbb{C}^{\rho\delta}{}_{[\alpha} g_{\beta]\sigma} + (\lambda\rho/2) \delta_{[\alpha}{}^\delta \mathbb{C}_{\beta]\sigma}{}^\rho \\ & + (\lambda\rho/2) \delta_{[\alpha}{}^\delta \mathbb{C}_{\beta]}{}^\rho{}_\sigma + (\lambda\rho/2) \mathbb{C}_\sigma{}^\rho{}_{[\beta} \delta_{\alpha]}{}^\delta \\ & + \tfrac{1}{2} \mathbb{C}_{\sigma[\beta}{}^\epsilon \mathbb{C}_{\alpha]\epsilon}{}^\delta - \tfrac{1}{2} \mathbb{C}^\delta{}_{[\beta}{}^\epsilon \mathbb{C}_{\alpha]\epsilon\sigma} \\ & - \mathbb{C}_\sigma{}^{\delta\epsilon'} \mathbb{C}_{\epsilon'[\alpha\beta]} + \tfrac{1}{2} \mathbb{C}_\sigma{}^{\delta\epsilon} \mathbb{C}_{\alpha\beta\epsilon} \\ & + \tfrac{1}{2} \mathbb{C}_{[\beta}{}^{\delta\epsilon} \mathbb{C}_{\alpha]\sigma\epsilon} - \tfrac{1}{2} \mathbb{C}_{\sigma\epsilon[\beta} \mathbb{C}_{\alpha]}{}^{\epsilon\delta} \\ & + \tfrac{1}{2} \mathbb{C}^\delta{}_{\epsilon[\beta} \mathbb{C}_{\alpha]}{}^\epsilon{}_\sigma - \tfrac{1}{2} \mathbb{C}^\epsilon{}_{\sigma[\beta} \mathbb{C}_{|\epsilon|}{}^\delta{}_{\alpha]} \\ & - \tfrac{1}{2} \mathbb{C}_{[\beta|\epsilon\sigma|} \mathbb{C}_{\alpha]}{}^{\epsilon\delta} + \tfrac{1}{2} \mathbb{C}_\epsilon{}^\delta{}_{[\alpha} \mathbb{C}_{\beta]\sigma}{}^\epsilon \\ & - \tfrac{1}{2} \mathbb{C}_{\epsilon\sigma[\alpha} \mathbb{C}_{\beta]}{}^{\delta\epsilon}, \end{aligned} \tag{4.11a}$$

$$\begin{aligned} R_{a\beta\sigma}{}^\delta = {} & (\delta_\beta{}^\delta/2) \nabla_a \lambda_\sigma - (g^{\epsilon\delta}/2) g_{\beta\sigma} \nabla_a \lambda_\epsilon \\ & + (\lambda^\epsilon/4) g^{\rho\delta} g_{\beta\sigma} \nabla_a g_{\epsilon\rho} - (\lambda^\epsilon/4) \delta_\beta{}^\delta \nabla_a g_{\epsilon\sigma} \\ & + (\mathbb{C}_\sigma{}^{\delta\epsilon}/2) \nabla_a g_{\epsilon\beta} + (\mathbb{C}/4)_\beta{}^{\epsilon\delta} \nabla_a g_{\epsilon\sigma} \\ & - (\mathbb{C}/4)_\beta{}^\epsilon{}_\sigma g^{\delta\rho} \nabla_a g_{\epsilon\rho} + (\mathbb{C}/4)_\beta{}^{\delta\epsilon} \nabla_a g_{\epsilon\sigma} \\ & - (\mathbb{C}/4)_{\beta\sigma}{}^\epsilon g^{\rho\delta} \nabla_a g_{\epsilon\rho} + (\mathbb{C}^{\delta\epsilon}{}_\beta/4) \nabla_a g_{\epsilon\sigma} \\ & - (\mathbb{C}/4)_\sigma{}^\epsilon{}_\beta g^{\rho\delta} \nabla_a g_{\epsilon\rho} \\ & + (R_a{}^\epsilon{}_\sigma/4) g^{\epsilon\delta} \nabla_e g_{\epsilon\beta} - (R_a{}^{e\delta}/4) \nabla_e g_{\beta\sigma}, \end{aligned} \tag{4.11b}$$

$$R_{a\beta c}{}^{\delta} = (g^{\epsilon\delta}/2)\nabla_{(a}\nabla_{c)}g_{\epsilon\beta} - (g^{\epsilon\rho}/4)g^{\gamma\delta}\nabla_a(g_{\rho\gamma})\nabla_c(g_{\epsilon\beta}) + (g_{ac}/2)g^{\epsilon\delta}\nabla_{(\beta}\lambda_{\epsilon)} + (\lambda^{\epsilon}/4)\lambda_{\epsilon}g_{ac}\delta_{\beta}{}^{\delta} - (g_{ac}/4)\lambda_{\beta}\lambda^{\delta} + (g_{ac}/4)\lambda_{\epsilon}\mathbb{C}^{\epsilon}{}_{\beta}{}^{\delta} + (g_{ac}/4)\lambda_{\epsilon}\mathbb{C}^{\epsilon\delta}{}_{\beta} + (R_{ae}{}^{\delta}/4)R^{e}{}_{c\beta} - (R_{ac}{}^{\epsilon}/4)\mathbb{C}_{\beta}{}^{\delta}{}_{\epsilon}\,, \tag{4.11c}$$

$$R_{abc}{}^{d} = (\delta_b{}^{d}/2)\nabla_c\lambda_{\alpha} - (g_{bc}/2)g^{de}\nabla_e\lambda_{\alpha} + (g_{bc}/4)g^{de}\lambda^{\epsilon}\nabla_e g_{\alpha\epsilon} - (\delta_b{}^{d}/4)\lambda^{\epsilon}\nabla_c g_{\alpha\epsilon} + (g^{ed}/2)g_{\alpha\epsilon}\nabla_b(R_{ce}{}^{\epsilon}) + (R_b{}^{d\epsilon}/4)\nabla_c g_{\alpha\epsilon} - (R_{bc}{}^{\epsilon}/4)g^{de}\nabla_e g_{\alpha\epsilon} + (R_c{}^{d\epsilon}/2)\nabla_b g_{\alpha\epsilon}\,, \tag{4.11d}$$

$$R_{abc}{}^{d} = (\lambda^{\epsilon}/4)\lambda_{\epsilon}g_{ac}\delta_b{}^{d} - (\lambda^{\epsilon}/4)\lambda_{\epsilon}\delta_a{}^{d}g_{bc} + (g_{bc}/4)R_a{}^{d\epsilon}\lambda_{\epsilon} - (\delta_b{}^{d}/4)R_{ac}{}^{\epsilon}\lambda_{\epsilon} + (\delta_a{}^{d}/4)R_{bc}{}^{\epsilon}\lambda_{\epsilon} - (g_{ac}/4)R_b{}^{d\epsilon}\lambda_{\epsilon} + (R_{ac}{}^{\epsilon}/4)R_b{}^{d}{}_{\epsilon} - (R_a{}^{d}{}_{\epsilon}/4)R_{bc}{}^{\epsilon} + (R_{ab}{}^{\epsilon}/2)R_c{}^{d}{}_{\epsilon} - (\delta_c{}^{d}/2)R_{ab}{}^{\epsilon}\lambda_{\epsilon} + \underline{R}_{abc}{}^{d}\,. \tag{4.11e}$$

One is usually interested also in the components of the Ricci tensor, $R_{AC} \equiv R_{ABC}{}^{B}$.

One has

$$R_{\alpha\sigma} = (g^{ab}/2)\nabla_a\nabla_b g_{\alpha\sigma} - (g^{ab}/2)g^{\rho\epsilon}\nabla_a g_{\alpha\rho}\nabla_b g_{\sigma\epsilon} + (g^{ab}/4)g^{\rho\epsilon}\nabla_a(g_{\alpha\sigma})\nabla_b(g_{\rho\epsilon}) + ((p-2)/4)(\lambda^{\epsilon}\lambda_{\epsilon}g_{\alpha\sigma} - \lambda_{\alpha}\lambda_{\sigma}) + ((p-2/2)(\nabla_{(\alpha}\lambda_{\sigma)} + \lambda_{\rho}\mathbb{C}^{\rho}{}_{(\alpha\sigma)}) + (g_{\alpha\sigma}/2)(g^{\beta\epsilon}\nabla_{\beta}\lambda_{\epsilon} + \lambda_{\rho}\mathbb{C}^{\rho}{}_{\beta}{}^{\beta}) + \mathbb{C}_{\epsilon'\beta(\alpha}C_{\sigma)}{}^{\beta\epsilon'} + \tfrac{1}{2}\mathbb{C}_{\alpha\epsilon}{}^{\beta}\mathbb{C}_{\sigma\beta}{}^{\epsilon} + \mathbb{C}_{\epsilon\beta(\alpha}\mathbb{C}^{\beta}{}_{\sigma)}{}^{\epsilon} + \tfrac{1}{2}\mathbb{C}_{\sigma}{}^{\beta\epsilon}\mathbb{C}_{\alpha\beta\epsilon} + \mathbb{C}_{(\alpha}{}^{\epsilon}{}_{\sigma)}\mathbb{C}_{\beta\epsilon}{}^{\beta} - \tfrac{1}{4}\mathbb{C}_{\beta\epsilon\alpha}\mathbb{C}^{\beta\epsilon}{}_{\sigma} - \tfrac{1}{4}R_{ab\alpha}R^{ab}{}_{\sigma}\,, \tag{4.12a}$$

$$R_{a\sigma} = ((p-2)/2)(\nabla_a\lambda_{\sigma} - (\lambda^{\epsilon}/2)\nabla_a g_{\epsilon\sigma}) + (\mathbb{C}_{\sigma}{}^{\beta\epsilon}/2)\nabla_a g_{\beta\epsilon} + (\mathbb{C}_{\beta}{}^{\epsilon\beta}/2)\nabla_a g_{\epsilon\sigma} + (g^{eb}/2)g_{\sigma\epsilon}\nabla_b(R_{ac}{}^{\epsilon}) + (R_a{}^{e}{}_{\sigma}/4)g^{\beta\epsilon}\nabla_e g_{\beta\epsilon} + (R_a{}^{e\epsilon}/2)\nabla_e g_{\epsilon\sigma}\,, \tag{4.12b}$$

$$R_{ac} = (g^{\beta\epsilon}/2)\nabla_{(a}\nabla_{c)}g_{\beta\epsilon} - (g^{\epsilon\rho}/4)g^{\gamma\beta}\nabla_a g_{\rho\gamma}\nabla_c g_{\beta\epsilon} + (g_{ac}/2)g^{\beta\epsilon}\nabla_{\beta}\lambda_{\epsilon} + ((p-2)/4)\lambda^{\epsilon}\lambda_{\epsilon}g_{ac} + (g_{ac}/2)\lambda_{\epsilon}\mathbb{C}^{\epsilon}{}_{\beta}{}^{\beta} + (R_{ab}{}^{\epsilon}/2)R_c{}^{b}{}_{\epsilon} + \underline{R}_{(ac)}\,. \tag{4.12c}$$

The Ricci scalar $R = R_{AB}g^{AB}$ is

$$R = g^{ab}g^{\alpha\beta}\nabla_a\nabla_b g_{\alpha\beta} + \tfrac{3}{4}g^{ab}\nabla_a(g^{\alpha\beta})\nabla_b(g_{\alpha\beta}) + (g^{ab}/4)g^{\alpha\beta}\nabla_a(g_{\alpha\beta})g^{\sigma\delta}\nabla_b(g_{\sigma\delta}) + ((p-2)(p-1)/4)\lambda^{\epsilon}\lambda_{\epsilon} + (p-1)\times(g^{\alpha\beta}\nabla_{\alpha}\lambda_{\beta} + \lambda_{\alpha}\mathbb{C}^{\alpha}{}_{\beta}{}^{\beta}) + C_{\alpha}{}^{\beta\epsilon'}\mathbb{C}_{\epsilon'\beta}{}^{\alpha} - \tfrac{1}{2}\mathbb{C}_{\alpha\beta}{}^{\sigma}\mathbb{C}^{\alpha}{}_{\sigma}{}^{\beta} + \tfrac{1}{4}\mathbb{C}_{\alpha\beta\sigma}\mathbb{C}^{\alpha\beta\sigma} + \mathbb{C}_{\alpha\sigma}{}^{\alpha}\mathbb{C}_{\beta}{}^{\alpha\beta} + \tfrac{1}{4}R_{ab\alpha}R^{ab\alpha} + \underline{R}\,. \tag{4.13}$$

5. COMMENTS

One could think of other possibilities of organizing these ideas, but the splitting of the tangent space introduced in Sec. 1 seems to be very convenient for calculations. In particular, the use of abstract indices for the horizontal space produces very compact expressions.

The "components" written in Sec. 3 generalize the traditional expressions of the Kaluza–Klein theories[10] in several ways; by allowing conformal symmetries in contrast to actual symmetries, by considering linear dependence among the fundamental vector fields corresponding to the symmetries, or by not demanding from the beginning the structure of a principal fiber bundle.

The expressions derived are trivially reduced for the case of Killing symmetries ($\lambda_{\alpha} \to 0$) and/or for the linear independent symmetries case ($\mathbb{C}_{\alpha\beta}{}^{\sigma} \to C_{\alpha\beta}{}^{\sigma}$).

It is particularly interesting to observe the form of the Ricci scalar in which the Yang–Mills term, a G term, a conformal term, and a dynamical term can be clearly separated.

From the discussion of Appendix H one also notes that the cross terms of the metric can be thought of as been given by the Yang–Mills gauge potentials[11] of the Yang–Mills fields $R_{ij}{}^{\alpha}$ [see also Eqs. (F4) and (F5)].

From another perspective the expressions given here seem to offer a useful tool for the study of solutions of the field equations in general relativity; in particular, as a sort of check the metrics for a highly symmetric low-dimensional manifold and for a well-known solution of the Einstein vacuum field equation were derived using the expressions of this work.

APPENDIX A

Proof of Theorem 1: Consider first the case in which $g_{\alpha'\alpha'} = 0$ for some α'. The question is, can $g_{\alpha'\beta'} = 0$, $\forall\beta' \neq \alpha'$? Because if it were so, and

$$g(\nabla_{\beta'}, \nabla_{\mathbf{v}}) = 0 \quad \forall\beta'\,,$$

then

$$\nabla'_{\mathbf{v}} = \nabla_{\mathbf{v}} + f\nabla_{\alpha'}\,, \quad f \in C^{\infty}(P)$$

will also satisfy

$$g(\nabla_{\beta'}, \nabla'_{\mathbf{v}}) = 0 \quad \forall\beta'\,.$$

By carefully choosing U it is possible to express

$$\nabla_{\alpha'} = b_{\alpha'}{}^{\alpha}\nabla_{\alpha}\,, \quad \alpha' = 1,2,...,q'\,,$$

$$b_{\alpha}{}^{\beta} = \delta_{\alpha}{}^{\beta}\,, \quad \alpha = 1,2,...,q\,,$$

so

$$0 = g_{\alpha'\beta'} \Rightarrow 0 = g_{\alpha'\beta} \Rightarrow b_{\alpha'}{}^{\alpha}g_{\alpha\beta} = 0\,,$$

but $g_{\alpha\beta}$ is a nondegenerate matrix so the only solution to $b_{\alpha'}{}^{\alpha}g_{\alpha\beta} = 0$ is $b_{\alpha'}{}^{\alpha} = 0$. So there is no $\nabla_{\alpha'}$ such that $g_{\alpha'\beta'} = 0\ \forall\beta' \neq \alpha'$.

By the local Frobenius theorem it is possible to choose $U \subset P$ with a coordinate system $(y_1,...,y_q,x_1,...,x_{p-q})$ with $|y_i| < \xi$, $|x_j| < \xi$, $\xi \in \mathbb{R}$, so that

$$\nabla_{\alpha'} = a_{\alpha'}{}^{i}\frac{\partial}{\partial y^i} \quad \text{for } \nabla_{\alpha'} \in T_u(U)\,.$$

By condition (a), $\nabla_{\mathbf{v}}$ must be of the form

$$\nabla_{\mathbf{v}} = \mathbf{v}^{\mathbf{j}}\frac{\partial}{\partial x^{\mathbf{j}}} + v^{i}\frac{\partial}{\partial y^{i}}\,,$$

where $\pi^{*}(\mathbf{v}^{\mathbf{j}})$ is identified with $\mathbf{v}^{\mathbf{j}}$.

Let

$$g_{ij} \equiv g\left(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j}\right), \quad g_{i\mathbf{j}} \equiv g\left(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial x^{\mathbf{j}}}\right), \quad g_{\mathbf{ij}} \equiv g\left(\frac{\partial}{\partial x^{\mathbf{i}}}, \frac{\partial}{\partial x^{\mathbf{j}}}\right).$$

Then condition (b) gives

$$0 = g(\nabla_{\alpha'}, \nabla_{\mathbf{v}}) = g_{ij} a_{\alpha'}{}^{i} v^{j} + g_{ij} a_{\alpha'}{}^{i} \mathbf{v}^{\mathbf{j}} \quad \forall \alpha' ,$$

so

$$v^{j} = -g^{ji} g_{ij} \mathbf{v}^{\mathbf{j}} .$$

APPENDIX B

This Appendix contains comments on the notation.

If the vector fields $e_{\mathbf{i}}$, $\mathbf{i} = 1,\ldots,p-q$ form a basis of $T_{\pi(u)}(B) \quad \forall u \in U$, then at u $\nabla_{\mathbf{i}} \equiv e_{\mathbf{i}}^{a} \nabla_a$ form a basis of a subspace $H_u \subset T_u(P)$; in particular, any vector $v \in T_u(P)$ can be expressed by

$$v = v^{\alpha} \nabla_{\alpha} + v^{\mathbf{i}} \nabla_{\mathbf{i}} = v^{\alpha} \nabla_{\alpha} + v^{a} \nabla_a ,$$

where $v^{a} = v^{\mathbf{i}} e_{\mathbf{i}}^{a}$ and so $\nabla_{\mathbf{v}} = \mathbf{v}^{a} \nabla_a = \mathbf{v}^{\mathbf{i}} \nabla_{\mathbf{i}}$. Actually one should write $\nabla_{\mathbf{v}} = \pi^{*}(v^{\mathbf{i}}) \nabla_{\mathbf{i}}$, but all the π^{*} and π_{*} are deliberately omitted for the sake of simplicity in the notation.

Similarly, let $d^{\mathbf{i}} \in T^{*}(B)$ be such that $d^{\mathbf{i}}{}_{a} e^{a}{}_{\mathbf{j}} = \delta^{\mathbf{i}}{}_{\mathbf{j}}$; then one defines

$$\theta^{\mathbf{i}} \equiv d^{\mathbf{i}}{}_{a} \theta^{a} .$$

Denote with $\mathscr{A}, \mathscr{B}, \mathscr{C}, \cdots$ abstract indices in P, then

$$\theta_{\omega,\mathscr{A}} v^{\mathscr{A}} = \pi^{*}(\omega)_{\mathscr{A}} v^{\mathscr{A}} = \omega_a \pi_{*}(v)^{a} = \omega_a v^{a} \quad \forall \omega \in T^{*}(B) ,$$

so

$$\theta^{a}{}_{\mathscr{A}} \nabla_{\alpha}{}^{\mathscr{A}} = 0 ,$$

$$\theta^{a}{}_{\mathscr{A}} \nabla_{b}{}^{\mathscr{A}} = \delta^{a}{}_{b} , \quad \theta^{\mathbf{i}}{}_{\mathscr{A}} \nabla_{\mathbf{j}}{}^{\mathscr{A}} = \delta^{\mathbf{i}}{}_{\mathbf{j}} .$$

If ω is a one-form in P, define

$$\omega_{\mathbf{i}} \equiv \omega_{\mathscr{A}} \nabla_{\mathbf{i}}{}^{\mathscr{A}} , \quad \omega_a \equiv \omega_{\mathscr{A}} \nabla_a{}^{\mathscr{A}} ,$$

then one easily gets the following relations

$$\theta^{a} = e^{a}{}_{\mathbf{i}} \otimes \theta^{\mathbf{i}} , \quad \omega_a = d^{\mathbf{i}}{}_{a} \omega_{\mathbf{i}} ,$$

$$\theta^{\mathbf{i}} \equiv d^{\mathbf{i}}{}_{a} \theta^{a} , \quad \omega_{\mathbf{i}} = e^{a}{}_{\mathbf{i}} \omega_a ,$$

$$\nabla_a = d^{\mathbf{i}}{}_{a} \otimes \nabla_{\mathbf{i}} , \quad v^{a} = v^{\mathbf{i}} e^{a}{}_{\mathbf{i}} ,$$

$$\nabla_{\mathbf{i}} \equiv e^{a}{}_{\mathbf{i}} \nabla_a , \quad v^{\mathbf{i}} = v^{a} d^{\mathbf{i}}{}_{a} .$$

APPENDIX C

Using the notation of Appendix B, one writes

$$[\mathbf{v},\mathbf{w}] = [\mathbf{v}(\mathbf{w}^{\mathbf{i}}) - \mathbf{w}(\mathbf{v}^{\mathbf{i}})] e_{\mathbf{i}} + \mathbf{v}^{\mathbf{i}} \mathbf{w}^{\mathbf{j}} [e_{\mathbf{i}}, e_{\mathbf{j}}] .$$

Let $f \in C^{\infty}(P)$, then

$$\begin{aligned} \nabla_{\mathbf{v}} \nabla_{\mathbf{w}} f &= \mathbf{v}^{\mathbf{i}} \nabla_{\mathbf{i}} (\mathbf{w}^{\mathbf{j}} \nabla_{\mathbf{j}} (f)) \\ &= \mathbf{v}^{\mathbf{i}} \nabla_{\mathbf{i}} (w^{\mathbf{j}}) \nabla_{\mathbf{j}} (f) + \mathbf{v}^{\mathbf{i}} \mathbf{w}^{\mathbf{j}} \nabla_{\mathbf{i}} \nabla_{\mathbf{j}} (f) , \end{aligned}$$

so

$$\begin{aligned} R_{\mathbf{v},\mathbf{w}} &= \mathbf{v}(w^{\mathbf{j}}) \nabla_{\mathbf{j}} - \mathbf{w}(\mathbf{v}^{\mathbf{j}}) \nabla_{j} + \mathbf{v}^{\mathbf{i}} \mathbf{w}^{\mathbf{j}} \\ &\quad \times [\nabla_{\mathbf{i}}, \nabla_{\mathbf{j}}] - [\mathbf{v}(\mathbf{w}^{\mathbf{j}}) - \mathbf{w}(\mathbf{v}^{\mathbf{i}})] \nabla_{\mathbf{j}} \\ &\quad - \mathbf{v}^{\mathbf{i}} \mathbf{w}^{\mathbf{j}} [e_{\mathbf{i}}, e_{\mathbf{j}}]^{\mathbf{k}} \nabla_{\mathbf{k}} \\ &= \mathbf{v}^{\mathbf{i}} \mathbf{w}^{\mathbf{j}} ([\nabla_{\mathbf{i}}, \nabla_{\mathbf{j}}] - [e_{\mathbf{i}}, e_{\mathbf{j}}]^{\mathbf{k}} \nabla_{\mathbf{k}}) . \end{aligned}$$

If $\mathbf{f} \in C^{\infty}(B)$ one easily observes that

$$R_{\mathbf{v},\mathbf{w}}(\mathbf{f}) = 0$$

so that it must be that

$$R_{\mathbf{v},\mathbf{w}} = \mathbf{v}^{\mathbf{i}} \mathbf{w}^{\mathbf{j}} R_{\mathbf{ij}}{}^{\alpha} \nabla_{\alpha} = \mathbf{v}^{a} \mathbf{w}^{b} R_{ab}{}^{\alpha} \nabla_{\alpha} .$$

APPENDIX D

$$\begin{aligned} R_{\mathbf{v},\mathbf{w}}(f) &\equiv ([\nabla_{\mathbf{v}}, \nabla_{\mathbf{w}}] - \nabla_{[\mathbf{v},\mathbf{w}]})(f) \\ &= \mathbf{v}^{a} \nabla_a \mathbf{w}^{b} \nabla_b f - \mathbf{w}^{b} \nabla_b \mathbf{v}^{a} \nabla_a f - [\mathbf{v},\mathbf{w}]^{a} \nabla_a f \\ &= \mathbf{v}^{a} (\mathbf{D}_a \mathbf{w}^{b} + \hat{\gamma}_a{}^{b}{}_{c} \mathbf{w}^{c}) \nabla_b f + \mathbf{v}^{a} \mathbf{w}^{b} \nabla_a \nabla_b f \\ &\quad - \mathbf{w}^{b} (\mathbf{D}_b \mathbf{v}^{a} + \hat{\gamma}_b{}^{a}{}_{c} \mathbf{v}^{c}) \nabla_a f \\ &\quad - \mathbf{w}^{b} \mathbf{v}^{a} \nabla_b \nabla_a f - [\mathbf{v},\mathbf{w}]^{a} \nabla_a (f) \\ &= [\mathbf{D}_{\mathbf{v}}(\mathbf{w}^{a}) - \mathbf{D}_{\mathbf{w}}(\mathbf{v}^{a}) - [\mathbf{v},\mathbf{w}]^{a}] \nabla_a f \\ &\quad + \mathbf{v}^{a} \mathbf{w}^{b} \{ [\nabla_a, \nabla_b] + (\hat{\gamma}_a{}^{c}{}_{b} - \hat{\gamma}_b{}^{c}{}_{a}) \nabla_c \} (f) \\ &= \mathbf{v}^{a} \mathbf{w}^{b} \{ [\nabla_a, \nabla_b] + (\hat{\gamma}_a{}^{c}{}_{b} - \hat{\gamma}_b{}^{c}{}_{a}) \nabla_c \} (f) , \end{aligned}$$

because $\mathbf{D}_a$ is a torsion-free connection.

APPENDIX E

Let z be a vector field in P; then

$$\begin{aligned} \nabla_a (z^{c}) &= \nabla_a (z^{\mathbf{i}} e^{c}{}_{\mathbf{i}}) = \nabla_a (z^{\mathbf{i}}) e^{c}_{\mathbf{i}} + \mathbf{z}^{\mathbf{i}} \nabla_a (e^{c}_{\mathbf{i}}) \\ &= \nabla_a (z^{\mathbf{i}}) e^{c}{}_{\mathbf{i}} + z^{\mathbf{i}} (\gamma_a{}^{c}{}_{\mathbf{i}} + \hat{\gamma}_a{}^{c}{}_{\mathbf{i}}) \\ &= \nabla_a (z^{\mathbf{i}}) e^{c}{}_{\mathbf{i}} + (\gamma_a{}^{c}{}_{d} + \hat{\gamma}_a{}^{c}{}_{d}) z^{d} . \end{aligned}$$

$$\begin{aligned} \nabla_b \nabla_a (z^{c}) &= \nabla_b (\nabla_a (z^{\mathbf{i}})) e^{c}{}_{\mathbf{i}} + \nabla_a (z^{\mathbf{i}}) \gamma_b{}^{c}{}_{\mathbf{i}} \\ &\quad + \nabla_b (\gamma_a{}^{c}{}_{\mathbf{i}}) z^{\mathbf{i}} + \gamma_a{}^{c}{}_{\mathbf{i}} \nabla_b (z^{\mathbf{i}}) , \end{aligned}$$

where

$$\mathbf{D}_a (e^{c}{}_{\mathbf{i}}) = \gamma_a{}^{c}{}_{\mathbf{i}} ,$$

and

$$\underline{\gamma}_a{}^{b}{}_{c} \equiv \gamma_a{}^{b}{}_{c} + \hat{\gamma}_a{}^{b}{}_{c} ,$$

$$\begin{aligned} (\nabla_a \nabla_b - \nabla_b \nabla_a)(z^{c}) &= (\nabla_a \nabla_b - \nabla_b \nabla_a)(z^{\mathbf{i}}) e^{c}{}_{\mathbf{i}} \\ &\quad + z^{\mathbf{i}} (\nabla_a \nabla_b - \nabla_b \nabla_a) e^{c}{}_{\mathbf{i}} \\ &= R_{ab}{}^{\alpha} \nabla_{\alpha} (z^{\mathbf{i}}) e^{c}{}_{\mathbf{i}} + \underline{R}_{abd}{}^{c} z^{d} , \end{aligned}$$

where

$$\begin{aligned} \underline{R}_{abc}{}^{d} &\equiv \nabla_a (\underline{\gamma}_b{}^{d}{}_{c}) - \nabla_b (\underline{\gamma}_a{}^{d}{}_{c}) + \underline{\gamma}_a{}^{d}{}_{e} \underline{\gamma}_b{}^{e}{}_{c} \\ &\quad - \underline{\gamma}_b{}^{d}{}_{e} \underline{\gamma}_a{}^{e}{}_{c} - (\underline{\gamma}_a{}^{e}{}_{b} - \underline{\gamma}_b{}^{e}{}_{a}) \gamma_e{}^{d}{}_{c} \end{aligned}$$

also

$$\underline{R}_{abd}{}^{c} = \mathbf{R}_{abd}{}^{c} + \hat{R}_{abd}{}^{c} ,$$

where $\mathbf{R}_{abd}{}^{c}$ is the curvature of the connection $\mathbf{D}$ and

$$\hat{R}_{abd}{}^{c} \equiv \hat{D}_a (\hat{\gamma}_b{}^{c}{}_{d}) - D_b (\hat{\gamma}_a{}^{c}{}_{d}) + \hat{\gamma}_b{}^{e}{}_{d} \hat{\gamma}_a{}^{c}{}_{e} - \hat{\gamma}_a{}^{e}{}_{d} \hat{\gamma}_b{}^{c}{}_{e} ,$$

so

$$[\nabla_a, \nabla_b] z^{d} = R_{ab}{}^{\alpha} \nabla_{\alpha} (z^{d}) + \underline{R}_{abc}{}^{d} z^{c}$$

$$\begin{aligned} [\nabla_a, \nabla_b](\omega_d z^{d}) &= R_{ab}{}^{\alpha} \nabla_{\alpha} (\omega_d z^{d}) \\ &= R_{ab}{}^{\alpha} \nabla_{\alpha} (\omega_d) z^{d} + \omega_d R_{ab}{}^{\alpha} \nabla_{\alpha} (z^{d}) \\ &= [\nabla_a, \nabla_b](\omega_d) z^{d} + \omega_d [\nabla_a, \nabla_b](z^{d}) \\ &= [\nabla_a, \nabla_b](\omega_d) z^{d} + \omega_d R_{ab}{}^{\alpha} \nabla_{\alpha} (z^{d}) \\ &\quad + \omega_d \underline{R}_{abc}{}^{d} z^{c} , \end{aligned}$$

so

$$[\nabla_a, \nabla_b](\omega_c) = R_{ab}{}^{\alpha} \nabla_{\alpha} (\omega_c) - \underline{R}_{abc}{}^{d} \omega_d .$$

Let the derivation $I^{c}{}_{d}$ be defined by

$$I^{c}{}_{d}(z) = z^{c} \nabla_d ,$$

$$I^{c}{}_{d}(f) = 0 \quad \text{if } f \in C^{\infty}(P) .$$

Then

$$0 = I^c{}_d(\omega_{\mathscr{A}} z^{\mathscr{A}}) = I^c{}_d(\omega_{\mathscr{A}}) z^{\mathscr{A}} + \omega_{\mathscr{A}} I^c{}_d(z^{\mathscr{A}})$$
$$= I^c{}_d(\omega_{\mathscr{A}}) z^{\mathscr{A}} + \omega_{\mathscr{A}} z^c \nabla_d{}^{\mathscr{A}}$$
$$= I^c{}_d(\omega_{\mathscr{A}}) z^{\mathscr{A}} + \omega_d z^c$$
$$= I^c{}_d(\omega_{\mathscr{A}})(z^\alpha \nabla_\alpha{}^{\mathscr{A}} + z^a \nabla_a{}^{\mathscr{A}}) + \omega_d z^c ,$$

so

$$I^c{}_d(\omega_{\mathscr{A}}) \nabla_\alpha{}^{\mathscr{A}} = 0 ,$$
$$I^c{}_d(\omega_{\mathscr{A}}) \nabla_a{}^{\mathscr{A}} = -\omega_d \delta^c{}_a \Rightarrow I^c{}_d(\omega_{\mathscr{A}}) = -\theta^c{}_{\mathscr{A}} \omega_d .$$

With this definition, one writes

$$[D'_a, D'_b](v^e \nabla_e) = (R_{ab}{}^\alpha D'_\alpha + \underline{R}_{abc}{}^d I^c{}_d)(v^e \nabla_e) .$$

Now

$$(\nabla_a \nabla_\beta - \nabla_\beta \nabla_a) z^c = \nabla_a \nabla_\beta (z^c) - \nabla_\beta (\hat{D}_a z^c + \hat{\gamma}_a{}^c{}_b z^b)$$
$$= \hat{D}_a \nabla_\beta (z^c) + \hat{\gamma}_a{}^c{}_b \nabla_\beta (z^b) - \nabla_\beta \hat{D}_a (z^c)$$
$$- \nabla_\beta (\hat{\gamma}_a{}^c{}_b) z^b - \hat{\gamma}_a{}^c{}_b \nabla_\beta (z^b)$$
$$= -\nabla_\beta (\hat{\gamma}_a{}^c{}_b) z^b ,$$

so

$$[D'_\alpha, D'_\beta](z) = \underline{R}_{a\beta c}{}^d z^c \nabla_d = \underline{R}_{a\beta c}{}^d I^c{}_d(z)$$

and

$$[\nabla_\alpha, \nabla_\beta] = C_{\alpha\beta}{}^{\delta'} \nabla_{\delta'} = C_{\alpha\beta}{}^{\delta'} b_{\delta'}{}^\sigma \nabla_\sigma .$$

Defining $\mathbb{C}_{\alpha\beta}{}^\sigma \equiv C_{\alpha\beta}{}^{\delta'} b_{\delta'}{}^\sigma$ one has

$$[\nabla_\alpha, \nabla_\beta] = \mathbb{C}_{\alpha\beta}{}^\sigma \nabla_\sigma$$

and

$$[D'_\alpha, D'_\beta] = \mathbb{C}_{\alpha\beta}{}^\sigma D'_\sigma .$$

APPENDIX F

$$0 = \mathscr{L}_{\nabla_\alpha}(\theta^\beta{}_{\mathscr{A}} \nabla_\sigma{}^{\mathscr{A}}) = \mathscr{L}_{\nabla_\alpha}(\theta^\beta{}_{\mathscr{A}}) \nabla^{\mathscr{A}}{}_\sigma + \theta^\beta{}_{\mathscr{A}} \mathbb{C}_{\alpha\sigma}{}^\delta \nabla_\delta{}^{\mathscr{A}}$$

so

$$\mathscr{L}_{\nabla_\alpha}(\theta^\beta{}_{\mathscr{A}}) = -\theta_{\mathscr{A}}{}^\sigma \mathbb{C}_{\alpha\sigma}{}^\beta + \mathscr{C}_\alpha{}^\beta{}_a \theta^a{}_{\mathscr{A}} ,$$

but

$$\theta^\alpha{}_{\mathscr{A}} \nabla_a{}^{\mathscr{A}} = 0 \Rightarrow \mathscr{C}_\alpha{}^\beta{}_a = 0 .$$

One has

$$\mathscr{L}_{\nabla_a} g = \nabla_\alpha (g_{\beta\sigma}) \theta^\beta \otimes \theta^\sigma + g_{\beta\sigma} \mathbb{C}_{\delta\alpha}{}^\beta \theta^\delta \otimes \theta^\sigma$$
$$+ g_{\beta\sigma} \mathbb{C}_{\delta\alpha}{}^\sigma \theta^\beta \otimes \theta^\delta + \nabla_\alpha (g_{ab}) \theta^a \otimes \theta^b = \lambda_\alpha g ,$$

so

$$\nabla_\alpha g_{\beta\sigma} = \lambda_\alpha g_{\beta\sigma} + \mathbb{C}_{\alpha\beta}{}^\delta g_{\delta\sigma} + \mathbb{C}_{\alpha\sigma}{}^\delta g_{\delta\beta} \tag{F1}$$

and

$$\nabla_\alpha g_{ab} = \lambda_\alpha g_{ab} . \tag{F2}$$

$$[D'_a, D'_b] g$$
$$= [\nabla_a, \nabla_b](g_{\alpha\beta}) \theta^\alpha \otimes \theta^\beta = R_{ab}{}^\sigma \nabla_\sigma (g_{\alpha\beta}) \theta^\alpha \otimes \theta^\beta$$
$$= R_{ab}{}^\sigma \nabla_\sigma g - \underline{R}_{abc}{}^d g_{de} \theta^c \otimes \theta^e - \underline{R}_{abe}{}^d g_{cd} \theta^c \otimes \theta^e$$
$$= R_{ab}{}^\sigma \nabla_\sigma (g_{\alpha\beta}) \theta^\alpha \otimes \theta^\beta + R_{ab}{}^\sigma \nabla_\sigma (g_{cd}) \theta^c \otimes \theta^d$$
$$- (\underline{R}_{abce} + \underline{R}_{abec}) \theta^c \otimes \theta^e$$

so

$$R_{ab}{}^\sigma \lambda_\sigma g_{cd} = 2 \underline{R}_{ab(cd)} . \tag{F3}$$

From

$$\{[\nabla_a, [\nabla_b, \nabla_c]] + [\nabla_b, [\nabla_c, \nabla_a]]$$
$$+ [\nabla_c, [\nabla_a, \nabla_b]]\}(F) = 0 ,$$

one gets

$$\nabla_{[a} R_{bc]}{}^\alpha = 0 . \tag{F4}$$

$$0 = \mathscr{L}_{\nabla_\alpha}(\mathbf{v}^a \mathbf{w}^b [\nabla_a, \nabla_b])$$
$$= \mathscr{L}_{\nabla_\alpha}(R_{\mathbf{v},\mathbf{w}}) = \mathbf{v}^a \mathbf{w}^b \mathscr{L}_{\nabla_\alpha}(R_{ab}{}^\sigma \nabla_\sigma)$$

so

$$\nabla_\alpha (R_{ab}{}^\beta) = R_{ab}{}^\sigma \mathbb{C}_{\sigma\alpha}{}^\beta . \tag{F5}$$

$$[D'_\alpha, D'_\beta] g = D'_\alpha (\nabla_\beta (g_{\sigma\delta}) \theta^\sigma \otimes \theta^\delta + \lambda_\beta g_{ab} \theta^a \otimes \theta^b)$$
$$- D'_\beta (\nabla_\alpha (g_{\sigma\delta}) \theta^\sigma \otimes \theta^\delta + \lambda_\alpha g_{ab} \theta^a \otimes \theta^b)$$
$$= [\nabla_\alpha, \nabla_\beta](g_{\sigma\delta}) \theta^\sigma \otimes \theta^\delta$$
$$+ (\nabla_\alpha \lambda_\beta - \nabla_\beta \lambda_\alpha) g_{ab} \theta^a \otimes \theta^b$$
$$= \mathbb{C}_{\alpha\beta}{}^\epsilon \nabla_\epsilon (g_{\sigma\delta}) \theta^\sigma \otimes \theta^\delta + \mathbb{C}_{\alpha\beta}{}^\sigma \lambda_\sigma g_{ab} \theta^a \otimes \theta^b$$

so

$$\nabla_\alpha \lambda_\beta - \nabla_\beta \lambda_\alpha = \mathbb{C}_{\alpha\beta}{}^\sigma \lambda_\sigma . \tag{F6}$$

$$0 = \mathscr{L}_{\nabla_{\alpha'}} \nabla_{\mathbf{v}} = \nabla_{\alpha'} (\mathbf{v}^i) \nabla_i - \nabla_{\mathbf{v}} (b_{\alpha'}{}^\alpha) \nabla_\alpha + b_{\alpha'}{}^\alpha \mathbf{v}^a [\nabla_\alpha, \nabla_a]$$

so

$$\nabla_a (b_{\alpha'}{}^\alpha) = 0$$

or

$$\nabla_a (\mathbb{C}_{\alpha\beta}{}^\sigma) = 0 . \tag{F7}$$

$$\mathscr{L}_{\nabla_\alpha} \nabla_{\alpha'} = \nabla_\alpha (b_{\alpha'}{}^\beta) \nabla_\beta + b_{\alpha'}{}^\beta [\nabla_\alpha, \nabla_\beta]$$
$$= (\nabla_\alpha (b_{\alpha'}{}^\beta) + b_{\alpha'}{}^\sigma \mathbb{C}_{\alpha\sigma}{}^\beta) \nabla_\beta$$
$$= C_{\alpha\alpha'}{}^{\delta'} \nabla_{\delta'} = \mathbb{C}_{\alpha\alpha'}{}^\beta \nabla_\beta .$$

So

$$\nabla_\alpha (b_{\alpha'}{}^\beta) = \mathbb{C}_{\alpha\alpha'}{}^\beta + b_{\alpha'}{}^\sigma \mathbb{C}_{\sigma\alpha}{}^\beta$$

or

$$\nabla_\alpha (\mathbb{C}_{\sigma\delta}{}^\beta) = C_{\sigma\delta}{}^{\alpha'} \mathbb{C}_{\alpha\alpha'}{}^\beta + \mathbb{C}_{\sigma\delta}{}^\epsilon \mathbb{C}_{\epsilon\alpha}{}^\beta . \tag{F8}$$

$$\mathscr{L}_{\nabla_{\alpha'}} \nabla_{\beta'} = b_{\alpha'}{}^\alpha \nabla_\alpha (b_{\beta'}{}^\beta) \nabla_\beta$$
$$- b_{\beta'}{}^\beta \nabla_\beta (b_{\alpha'}{}^\alpha) \nabla_\alpha + b_{\alpha'}{}^\alpha b_{\beta'}{}^\beta [\nabla_\alpha, \nabla_\beta]$$
$$= [b_{\alpha'}{}^\alpha (\mathbb{C}_{\alpha\beta'}{}^\delta + b_{\beta'}{}^\sigma \mathbb{C}_{\sigma\alpha}{}^\delta)$$
$$- b_{\beta'}{}^\beta (\mathbb{C}_{\beta\alpha'}{}^\delta + b_{\alpha'}{}^\sigma \mathbb{C}_{\sigma\beta}{}^\delta)] \nabla_\delta$$
$$+ b_{\alpha'}{}^\alpha b_{\beta'}{}^\beta \mathbb{C}_{\alpha\beta}{}^\delta \nabla_\delta$$
$$= (b_{\alpha'}{}^\alpha \mathbb{C}_{\alpha\beta'}{}^\delta + b_{\beta'}{}^\beta \mathbb{C}_{\alpha'\beta}{}^\delta - b_{\alpha'}{}^\alpha b_{\beta'}{}^\beta \mathbb{C}_{\alpha\beta}{}^\delta) \nabla_\delta$$
$$= C_{\alpha'\beta'}{}^{\delta'} \nabla_{\delta'} = \mathbb{C}_{\alpha'\beta'}{}^\delta \nabla_\delta ,$$

so

$$\mathbb{C}_{\alpha'\beta'}{}^\delta = b_{\alpha'}{}^\alpha \mathbb{C}_{\alpha\beta'}{}^\delta + b_{\beta'}{}^\beta \mathbb{C}_{\alpha'\beta}{}^\delta - b_{\alpha'}{}^\alpha b_{\beta'}{}^\beta \mathbb{C}_{\alpha\beta}{}^\delta . \tag{F9}$$

$$\nabla_{\alpha'} g_{ab} = \lambda_{\alpha'} g_{ab} = b_{\alpha'}{}^\alpha \nabla_\alpha g_{ab} = b_{\alpha'}{}^\alpha \lambda_\alpha g_{ab} ,$$
$$\lambda_{\alpha'} = b_{\alpha'}{}^\alpha \lambda_\alpha , \tag{F10}$$

$$[\nabla_{\alpha'}, \nabla_{\beta'}] g_{\sigma\delta} = C_{\alpha'\beta'}{}^{\delta'} \nabla_{\delta'} g_{\sigma\delta}$$
$$= C_{\alpha'\beta'}{}^{\delta'} (\lambda_{\delta'} g_{\sigma\delta} + \mathbb{C}_{\delta'\sigma}{}^\rho g_{\rho\delta} + \mathbb{C}_{\delta'\delta}{}^\rho g_{\sigma\rho})$$
$$= \mathbb{C}_{\alpha'\beta'}{}^\epsilon (\lambda_\epsilon g_{\sigma\delta} + \mathbb{C}_{\epsilon\sigma}{}^\rho g_{\rho\delta} + \mathbb{C}_{\epsilon\delta}{}^\rho g_{\sigma\rho}) ,$$

so

$$C_{\alpha'\beta'}{}^{\epsilon'} \mathbb{C}_{\epsilon'(\sigma\delta)} = \mathbb{C}_{\alpha'\beta'}{}^\epsilon \mathbb{C}_{\epsilon(\sigma\delta)} . \tag{F11}$$

The Jacobi identity gives

$$0 = C_{\alpha'\beta'}{}^{\epsilon'}\mathbb{C}_{\sigma'\epsilon'}{}^{\delta} + C_{\beta'\sigma'}{}^{\epsilon'}\mathbb{C}_{\alpha'\epsilon'}{}^{\delta} + C_{\sigma'\alpha'}{}^{\epsilon'}\mathbb{C}_{\beta'\epsilon'}{}^{\delta}. \tag{F12}$$

APPENDIX G

Evaluation of $\gamma_A{}^C{}_B$, $R_{ABC}{}^D$, R_{AC}, R. Let v and w be vector fields in P, then the torsion is given by

$$\begin{aligned}T(v,w) &\equiv D_v w - D_w v - [v,w]\\ &= v^A D_A(w^B\nabla_B) - w^B D_B(v^A\nabla_A) - v^A\nabla_A(w^B)\nabla_B\\ &\quad + w^B\nabla_B(v^A)\nabla_A - v^A w^B[\nabla_A,\nabla_B]\\ &= v^A D_A(w^B)\nabla_B - w^B D_B(v^A)\nabla_A\\ &\quad - v^A\nabla_A(w^B)\nabla_B + w^B\nabla_B(v^A)\nabla_A\\ &\quad + v^A w^B\{D_A\nabla_B - D_B\nabla_A - [\nabla_A,\nabla_B]\}\\ &= v^A w^B(\gamma_A{}^C{}_B - \gamma_B{}^C{}_A - C_{AB}{}^C)\nabla_C;\end{aligned}$$

so torsion-free means

$$\gamma_A{}^C{}_B - \gamma_B{}^C{}_A = C_{AB}{}^C.$$

One also has

$$\begin{aligned}0 &= D_C g = D_C(g_{AB})\theta^A\otimes\theta^B + g_{AB}D_C(\theta^A)\otimes\theta^B\\ &\quad + g_{AB}\theta^A\otimes D_C\theta^B\\ &= (\nabla_C(g_{AB}) - g_{DB}\gamma_C{}^D{}_A - g_{AD}\gamma_C{}^D{}_B)\theta^A\otimes\theta^B\end{aligned}$$

so

$$\nabla_C g_{AB} = g_{AD}\gamma_C{}^D{}_B + g_{DB}\gamma_C{}^D{}_A,$$

$$\begin{aligned}\gamma_{ACB} &\equiv \tfrac{1}{2}[\gamma_{ABC} + \gamma_{ACB} + \gamma_{BAC} + \gamma_{BCA} - \gamma_{CAB} - \gamma_{CBA}\\ &\quad + \gamma_{ACB} - \gamma_{BCA} + \gamma_{CAB} - \gamma_{BAC} + \gamma_{CBA} - \gamma_{ABC}]\\ &= \tfrac{1}{2}[\nabla_A g_{BC} + \nabla_B g_{AC} - \nabla_C g_{AB}\\ &\quad + C_{ABC} + C_{CBA} + C_{CAB}].\end{aligned}$$

Then

$$\begin{aligned}\gamma_{\alpha\sigma\beta} &= \tfrac{1}{2}(\mathbb{C}_{\alpha\beta\sigma} + \mathbb{C}_{\alpha\sigma\beta} + \mathbb{C}_{\beta\sigma\alpha}\\ &\quad + \lambda_\alpha g_{\beta\sigma} + \lambda_\beta g_{\alpha\sigma} - \lambda_\sigma g_{\alpha\beta}),\\ \gamma_{\alpha\alpha\beta} &= -\tfrac{1}{2}\nabla_a g_{\alpha\beta},\\ \gamma_{a\alpha\beta} &= \tfrac{1}{2}\nabla_a g_{\alpha\beta},\\ \gamma_{\alpha\beta a} &= \tfrac{1}{2}\nabla_a g_{\alpha\beta},\\ \gamma_{\alpha a b} &= \tfrac{1}{2}(-\lambda_\alpha g_{ab} + R_{ab\alpha}),\\ \gamma_{a\alpha b} &= \tfrac{1}{2}(\lambda_\alpha g_{ab} + R_{ab\alpha}),\\ \gamma_{ab\alpha} &= \tfrac{1}{2}(\lambda_\alpha g_{ab} - R_{ab\alpha}),\\ \gamma_{abc} &= 0.\end{aligned}$$

Let u, v, and w be vector fields in P, then

$$\begin{aligned}R(u,v)w &\equiv \{[D_u,D_v] - D_{[u,v]}\}w\\ &= \{u^A D_A v^B D_B - v^B D_B u^A D_A - [u,v]^C D_C\}w\\ &= \{u^A D_A(v^B)D_B - v^B D_B(u^A)D_A\\ &\quad + u^A v^B[D_A,D_B] - u^A\nabla_A(v^C)D_C\\ &\quad + v^B\nabla_B(u^C)D_C - u^A v^B C_{AB}{}^C D_C\}w\\ &= u^A v^B\{D_A D_B w - D_B D_A w - C_{AB}{}^C D_C w\}\\ &= u^A v^B\{D_A(D'_B(w^C)\nabla_C + w^D\gamma_B{}^C{}_D\nabla_C)\\ &\quad - D_B(D'_A(w^C)\nabla_C + w^D\gamma_A{}^C{}_D\nabla_C\\ &\quad - C_{AB}{}^C(D'_C(w^D)\nabla_D + w^E\gamma_C{}^D{}_E\nabla_d)\}\\ &= u^A v^B\{[D'_A,D'_B](w^D)\nabla_D + D'_B(w^C)\gamma_A{}^D{}_C\nabla_D\\ &\quad - D'_A(w^C)\gamma_B{}^D{}_C\nabla_D + D'_A(w^D)\gamma_B{}^C{}_D\nabla_C\\ &\quad - D'_B(w^D)\gamma_A{}^C{}_D\nabla_C + w^D[D'_A(\gamma_B{}^C{}_D)\\ &\quad - D'_B(\gamma_A{}^C{}_D)]\nabla_C + w^C\gamma_B{}^E{}_C\gamma_A{}^D{}_E\nabla_D\\ &\quad - w^C\gamma_A{}^E{}_C\gamma_B{}^D{}_E\nabla_D - C_{AB}{}^C D'_C(w^D)\nabla_D\\ &\quad - w^C C_{AB}{}^E\gamma_E{}^D{}_C\nabla_D\}\\ &= u^A u^B w^C\{D'_A(\gamma_B{}^D{}_C) - D'_B(\gamma_A{}^D{}_C) + \gamma_B{}^E{}_C\gamma_A{}^D{}_E\\ &\quad - \gamma_A{}^E{}_C\gamma_B{}^D{}_E - C_{AB}{}^E\gamma_E{}^D{}_C + \underline{R}_{ABC}{}^D\}\nabla_D\\ &\equiv R_{ABC}{}^D u^A v^B w^C\nabla_D.\end{aligned}$$

Note that the only nontrivial components of $\underline{R}_{ABC}{}^D$ are $\underline{R}_{abc}{}^d$ and $\underline{R}_{abc}{}^d = -\underline{R}_{bac}{}^d$.

APPENDIX H

Let the vector fields e_i, $i = 1,\ldots,p-q$ be now a coordinate basis, and define $A_i{}^\alpha$ by

$$\nabla_i = \frac{\partial}{\partial x^i} + A_i{}^\alpha\nabla_\alpha. \tag{H1}$$

Then one sees that

$$[\nabla_i,\nabla_j] = R_{ij}{}^\alpha\nabla_\alpha$$

and so

$$R_{ij}{}^\alpha = \frac{\partial}{\partial x^i}(A_j{}^\alpha) - \frac{\partial}{\partial x^j}(A_i{}^\alpha) + A_i{}^\beta A_j{}^\sigma\mathbb{C}_{\beta\sigma}{}^\alpha. \tag{H2}$$

And from

$$[\nabla_\alpha,\nabla_i] = 0,$$

one gets

$$\nabla_\alpha(A_i{}^\beta) = A_i{}^\sigma\mathbb{C}_{\sigma\alpha}{}^\beta. \tag{H3}$$

It is also easy to see that

$$A_i{}^\alpha = -g_{ik}g^{kj}a_j{}^\alpha, \quad \text{where } \frac{\partial}{\partial y^j} = a_j{}^\alpha\nabla_\alpha \tag{H4a}$$

or

$$g_{ik} = -A_i{}^\alpha a_\alpha{}^j g_{jk}. \tag{H4b}$$

And defining

$$g_{i\alpha} \equiv g(\nabla_i,\nabla_\alpha),$$

one has

$$A_i{}^\alpha = -g_{i\beta}g^{\beta\alpha} \tag{H5a}$$

or

$$g_{i\alpha} = -A_i{}^\beta g_{\beta\alpha}. \tag{H5b}$$

[1] T. Kaluza, Sitzungsberichte der Preussischen Akademie der Wissenschaften, 966 (1921).

[2] See, for example, O. Klein, Z. Phys. **37**, 895 (1926); P. G. Bergmann, *Introduction to the Theory of Relativity* (Prentice-Hall, New York, 1942); A. Lichnerowicz, *Theories relativistes de la gravitation et de l'electromagnetisme* (Masson, Paris, 1955); W. Thirring, in *11th Schladming Conference*, edited by P. Urban (Springer-Verlag, New York, 1972); M. W. Kalinowski, Int. J. Theor. Phys. **20(8)**, 563 (1981).

[3] B. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965); R. Kerner, Ann. Inst. H. Poincaré **9**, 143 (1968); A. Trautman, Rep. Math. Phys. **1**, 29 (1970); Y. M. Cho, J. Math. Phys. **16**, 2029 (1975); Y. M. Cho and P. G. O. Freund, Phys. Rev. D **12**, 1711 (1975); E. Witten, Nucl. Phys. B **186**, 412 (1981); S. Tanaka, Prog. Theor. Phys. **66(4)**, 1477 (1981).

[4] J. Scherk and J. H. Schwarz, Phys. Lett. B **57**, 463 (1975); E. Cremmer and J. Scherk, Nucl. Phys. B **108**, 409 (1976); J. F. Luciani, Nucl. Phys. B **135**,

111 (1978).
[5]E. Cremmer, B. Julia, and J. Scherk, Phys. Lett. B **76**, 409 (1978); E. Cremmer and B. Julia, Phys. Lett. B **80**, 48 (1978); Nucl. Phys. B **159**, 141 (1979). For a review of supergravity, see P. van Nieuwenhuizen, Phys. Rep. **68(4)**, 191 (1981), where an extensive list of references are provided.
[6]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.
[7]R. Penrose, in *Battelle Rencontres* (Benjamin, New York, 1968).
[8]See, for example, C. Chevalley, *Theory of Lie Groups* (Princeton U. P., Princeton, New Jersey, 1946).
[9]See, for example, Serge Lang, *Introduction to Differentiable Manifolds* (Interscience, New York, 1962).
[10]See, for example, Y. M. Cho (1975) in Ref. 3.
[11]C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).

Convergence of multitime correlation functions in the weak and singular coupling limits[a)]

R. Dümcke
Fachbereich Physik, Universität München, Munich, Federal Republic of Germany

(Received 18 August 1981; accepted for publication 11 December 1981)

For a system coupled to a thermal bath we prove the convergence of the multitime correlation functions of system observables in the weak and singular coupling limits. The limiting correlation functions are given by the quantum regression law. Therefore, our result implies that in the limit the dynamics of the system are governed by a quantum stochastic process in the sense of Lindblad.

PACS numbers: 05.30. − d, 02.50. + s

1. INTRODUCTION

We consider a quantum mechanical system coupled to a quasifree heat bath. The total Hamiltonian is given by

$$H = H_S + H_B + H_I. \tag{1}$$

The system Hamiltonian H_S is a self-adjoint operator on $\mathscr{H}_S$, the system Hilbert space. The bath Hamiltonian, acting on the bath Hilbert space $\mathscr{H}_B$, is formally given by $H_B = \int dk\, \omega(k)a^+(k)a(k)$ and the interaction is $H_I = Q \otimes F$, where Q is bounded and self-adjoint on $\mathscr{H}_S$ and $F = \int dk\, \lambda(k)(a^+(k) + a(k))$. We denote by $\mathscr{T}(\mathscr{H})$ the Banach space of trace class operators with the trace norm $\|\cdot\|_1$ and by $\mathscr{B}(\mathscr{H})$ the Banach space of bounded operators on $\mathscr{H}$. The initial state of the system is specified by the state operator $W = \rho \otimes \omega_\beta \in \mathscr{T}(\mathscr{H}_S \otimes \mathscr{H}_B)$, where $\rho \in \mathscr{T}(\mathscr{H}_S)$ is an arbitrary state operator of the system and ω_β is the thermal equilibrium state of the reservoir at temperature β^{-1}. The time evolution of density operators of the joint system is given by $U(t)W = e^{-iHt}We^{iHt}$. We are interested in the reduced dynamics of the system defined by

$$T(t)\rho = \mathrm{Tr}_B[U(t)\rho \otimes \omega_\beta], \tag{2}$$

where Tr_B denotes the partial trace over the bath Hilbert space.

The reduced dynamics is governed by the Nakajima–Zwanzig generalized master equation.[1-3] This equation contains memory terms which make the evolution non-Markovian. But if the decay time of the reservoir time correlation function $g(t) = \mathrm{Tr}[e^{iH_Bt}Fe^{-iH_Bt}F\omega_\beta]$ becomes short compared to the typical relaxation time of the system, then $T(t)$ may be approximated by a dynamical semigroup. There are two different approximation procedures. In the weak coupling limit H_I becomes weak of order ϵ. To obtain a nontrivial effect one has to observe the system up to times of order ϵ^{-2}. In this case Davies[4] proved that $T(t)$ converges to a dynamical semigroup. In the singular coupling limit $g(t)$ converges to a delta function. This may be achieved by scaling the interaction as ϵ^{-1} and speeding up the free reservoir motion as ϵ^{-2} (Ref. 5). The convergence of the reduced dynamics in this limit was first proved by Hepp and Lieb[6] for quasifree systems and studied in more detail in Refs. 7 and 8.

It was pointed out by Lindblad[9] that a semigroup law is not sufficient for the system dynamics to be Markovian. As in the classical case the Markov character can be established only by considering all higher order time correlation functions. The purpose of this paper is to show that in the weak and singular coupling limits the dynamics of the system is indeed governed by a Markovian quantum stochastic process in the sense of Lindblad.[10] Technically we prove the convergence of all multitime correlation functions and show that in the limit the structure is given by the quantum regression law.[11-16]

We briefly indicate the precise result and its physical interpretation.

Let $X_i, Y_i \in \mathscr{B}(\mathscr{H}_S)$ be system operators and define the map

$$E_i W = X_i \otimes \mathbf{1} W Y_i \otimes \mathbf{1}. \tag{3}$$

Furthermore, we define

$$T(E_n,t_n;\ldots;E_1,t_1)\rho = \mathrm{Tr}_B E_n U(t_n - t_{n-1})E_{n-1}\cdots E_1 U(t_1)\rho \otimes \omega_\beta \tag{4}$$

as a map on $\mathscr{T}(\mathscr{H}_S)$. Tracing over the system results in the multitime correlation function for the system in the form

$$\mathrm{Tr}_S T(E_n,t_n;\ldots;E_1,t_1)\rho = \mathrm{Tr}[(Y_1 \otimes \mathbf{1})(t_1)\cdots(Y_n \otimes \mathbf{1})(t_n)(X_n \otimes \mathbf{1})(t_n) \cdots (X_1 \otimes \mathbf{1})(t_1)\rho \otimes \omega_\beta], \tag{5}$$

where $(A \otimes \mathbf{1})(t) = e^{iHt}A \otimes \mathbf{1}e^{-iHt}$.

In Secs. 2 and 3 we prove the convergence of $T(E_n,t_n;\ldots;E_1,t_1)\rho$ in the weak and singular coupling limits, respectively. If $T_0(t)$ denotes the asymptotic semigroup, then $T(E_n,t_n;\ldots;E_1,t_1)\rho$ converges to

$$\tilde{E}_n T_0(t_n - t_{n-1})\tilde{E}_{n-1}\cdots\tilde{E}_1 T_0(t_1)\rho, \tag{6}$$

where $\tilde{E}_i\rho = X_i\rho Y_i$. Equation (6) is a quantum mechanical analog of the formula for the finite dimensional probability distributions for classical Markov processes.

$T(E_n,t_n;\ldots;E_1,t_1)\rho$ has the following physical interpretation. Let $A_i = \Sigma_j c_i^{(j)} P_i^{(j)}$ be observables of the system, $c_i^{(j)}$ being the real eigenvalues and $P_i^{(j)}$ the corresponding eigenprojections. We set $X_i = Y_i = P_i^{(j_i)}$. Then $T(E_n,t_n;\ldots;E_1,t_1)\rho$ is the reduced state of the system at time t_n subjected to measurements of A_1 at time $t_1,\ldots,A_n$ at time t_n. At time t_i the value $c_i^{(j_i)}$ is observed and at each measurement the state is

[a)]This work is part of the author's doctoral dissertation submitted to Fakultät für Physik, Universität München.

 0022-2488/83/020311-05$02.50

reduced according to the von Neumann projection postulate.

2. THE WEAK COUPLING LIMIT

The Hamiltonian is scaled as

$$H^\epsilon = \epsilon^{-2}H_S + \epsilon^{-2}H_B + \epsilon^{-1}H_I. \tag{7}$$

This Hamiltonian describes the time evolution on the rescaled time scale. On the original time scale the Hamiltonian reads $H_S + H_B + \epsilon H_I$, where the interaction is weak of order ϵ.

The bath is assumed to be an infinite quasifree fermion system. It is well known that thermal equilibrium states of infinite systems cannot be represented by state operators on Fock space, in general. In the algebraical framework of quantum statistical mechanics the GNS construction associates to each state a Hilbert space, and the state is represented by a state vector.[17] Let $\mathscr{H}_B$ be the Hilbert space associated with the initial state ω_β of the bath. The representation of the algebra of field operators in $\mathscr{H}_B$ satisfies the canonical anticommutation relations in the form

$$\phi(f)\phi(g) + \phi(g)\phi(f) = 2\mathrm{Re}(f,g). \tag{8}$$

The initial state ω_β of the bath is completely specified by the two point function

$$\omega_\beta(\phi(f)\phi(g)) = (g,(1+e^{\beta(h-\mu)})^{-1}f) + (f,(1+e^{-\beta(h-\mu)})^{-1}g), \tag{9}$$

where h is the one particle Hamiltonian, β is the inverse temperature, and μ is the chemical potential. Bath correlation functions of odd order vanish and even order correlation functions are given by

$$\omega_\beta(\phi(f_1)\cdots\phi(f_{2n})) = \sum_{p\in P(2n)} \mathrm{sgn}\, p \prod_{k=1}^{n} \omega_\beta(\phi(f_{p(2k-1)})\phi(f_{p(2k)})), \tag{10}$$

where the sum is over all ordered pairings of $1,\ldots,2n$.[18]

We put $F(t) = e^{iH_B t}Fe^{-iH_B t} = \phi(e^{iht}\lambda)$, where λ is a fixed test function. On $g(t) = \omega_\beta(F(t)F)$ we impose the condition

$$\int_{-\infty}^{\infty} dt\, |g(t)| < \infty. \tag{11}$$

This condition is satisfied for $h = -\Delta$ on $\mathbb{L}^2(\mathbb{R}^\nu)$ and $\nu \geqslant 3$.

We define the Liouville operators

$L_S W = -i[H_S, W]$, $L_B W = -i[H_B, W]$, $L_I W = -i[H_I, W]$ and put $U^\epsilon(t) = \exp(\epsilon^{-2}L_S + \epsilon^{-2}L_B + \epsilon^{-1}L_I)t$, $U_0^\epsilon(t) = \exp(\epsilon^{-2}L_S + \epsilon^{-2}L_B)t$.

Furthermore, we define the semigroup

$$T_0^\epsilon(t) = \exp(\epsilon^{-2}L_S + K)t, \tag{12}$$

where

$$K\rho = \int_0^\infty dt\, \mathrm{Tr}_B e^{-(L_S+L_B)t}L_I e^{(L_S+L_B)t}L_I \rho\otimes\omega_\beta. \tag{13}$$

$T^\epsilon(E_n,t_n;\ldots;E_1,t_1)$ is obtained from (4) with $U(t)$ replaced by $U^\epsilon(t)$.

Theorem 1: Assume (11). Then for all $t>0$ and all $\rho\in\mathscr{T}(\mathscr{H}_S)$

$$\lim_{\epsilon\to 0}\sup_{0<t_1<\cdots<t_n<t} \| T^\epsilon(E_n,t_n;\ldots;E_1,t_1)\rho - \widetilde{E}_n T_0^\epsilon(t_n - t_{n-1})\widetilde{E}_{n-1}\cdots\widetilde{E}_1 T_0^\epsilon(t_1)\rho\|_1 = 0.$$

Proof: We prove the case $n = 2$. The proof for arbitrary n is analogous. The steps of the proof are as follows. Considering L_I as a perturbation $U^\epsilon(t)$ is expanded in a Dyson series,

$$U^\epsilon(t) = \sum_{k=0}^{\infty}\epsilon^{-k}\int_{0<s_1<\cdots<s_k<t} ds_k\cdots ds_1 U_0^\epsilon(t-t_k)L_I\cdots L_I U_0^\epsilon(t_1). \tag{14}$$

This expansion is inserted in $T^\epsilon(E_2,t_2;E_1,t_1)\rho$. The resulting series is majorized by an absolutely convergent series uniformly in ϵ. Finally, we show that most of the terms in the series vanish in the limit and that the remaining terms converge to corresponding terms in the series expansion of $E_2T_0^\epsilon(t_2-t_1)E_1T_0^\epsilon(t_1)\rho$.

In the proof we rely on technical results by Davies.[4]

To simplify the notation of the integrations we introduce $\Delta(t_1,k,t_2) = \{(s_1,\ldots,s_k)\in\mathbb{R}^k \mid t_1\leqslant s_1\leqslant\cdots\leqslant s_k\leqslant t_2\}$ and $\Delta(t_1,k,t_2,l,t_3) = \{(s_1,\ldots,s_k,s_{k+1},\ldots,s_{k+l})\in R^{k+l} \mid t_1\leqslant s_1\leqslant\cdots\leqslant s_k\leqslant t_2\leqslant s_{k+1}\leqslant\cdots\leqslant s_{k+l}\leqslant t_3\}$.

(1) The uniform estimate

Using (14) we obtain

$$T^\epsilon(E_2,t_2;E_1,t_1)\rho = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} R_{kl}\rho, \tag{15}$$

where

$$R_{kl}\rho = \epsilon^{-(k+l)}\int_{\Delta(0,k,t_1,l,t_2)}\underline{ds}\,\mathrm{Tr}_B\,[E_2U_0^\epsilon(t_2-s_{k+l})L_I \cdots L_I U_0^\epsilon(s_{k+1}-t_1)E_1 \times U_0^\epsilon(t_1-s_k)L_I\cdots L_I U_0^\epsilon(s_1)\rho\otimes\omega_\beta]. \tag{16}$$

We have

$$L_I W = -i[C_l B_l W - C_r B_r W], \tag{17}$$

where

$$C_l W = Q\otimes \mathbf{1}W, \quad C_r W = WQ\otimes\mathbf{1},$$
$$B_l W = \mathbf{1}\otimes FW, \quad B_r W = W\mathbf{1}\otimes F.$$

Let $\Pi(n)$ denote the set of functions $\{1,\ldots,n\}\to\{r,l\}$. For $\pi\in\Pi(n)$ we define $\mathrm{sgn}\,\pi = (-1)^{\mathrm{card}|\pi^{-1}(\{l\})|}$.

Using (17), we obtain from (16)

$$R_{kl}\rho = (i\epsilon)^{-(k+l)}\int_{\Delta(0,k,t_1,l,t_2)}\underline{ds}\sum_{\pi\in\Pi(k+l)}\mathrm{sgn}\,\pi \times\widetilde{E}_2 e^{\epsilon^{-2}L_S t_2}C_{\pi(k+l)}(s_{k+l})\cdots C_{\pi(k+1)}(s_{k+1})e^{-\epsilon^{-2}L_S t_1} \times\widetilde{E}_1 e^{\epsilon^{-2}L_S t_1}C_{\pi(k)}(s_k)\cdots C_{\pi(1)}(s_1)\rho \times\mathrm{Tr}[B_{\pi(k+l)}(s_{k+l})\cdots B_{\pi(k+1)}(s_{k+1}) \times B_{\pi(k)}(s_k)\cdots B_{\pi(1)}(s_1)\omega_\beta], \tag{18}$$

where

$$C_i(t) = e^{i\epsilon^{-2}H_S t} C_i e^{-i\epsilon^{-2}H_S t},$$
$$B_i(t) = e^{i\epsilon^{-2}H_B t} B_i e^{-i\epsilon^{-2}H_B t}, \quad i = r,l.$$

Now we estimate

$$\|R_{kl}\| \leqslant \epsilon^{-(k+l)} \|X_1\| \; \|Y_1\| \; \|X_2\| \; \|Y_2\| \; \|Q\|^{k+l} \times \sum_{\pi \in \Pi(k+l)} \int_{\Delta(0,k,t_1,l,t_2)} \underline{ds} \left| \operatorname{Tr} B_{\pi(k+l)}(s_{k+l}) \cdots B_{\pi(1)}(s_1)\omega_\beta \right|. \tag{19}$$

To each $\pi \in \Pi(n)$ we associate a permutation $\hat{\pi}$: $\{1,\ldots,n\} \to \{1,\ldots,n\}$ defined in the following way. Set $j_0 = \text{card}\,|\pi^{-1}(\{l\})|$. We put $\hat{\pi}(1) = \min\{j | \pi(j) = l\}$, $\hat{\pi}(i) = \min\{j > \hat{\pi}(i-1) | \pi(j) = l\}$, $i = 2,\ldots,j_0$, $\hat{\pi}(j_0+1) = \max\{j | \pi(j) = r\}$, $\hat{\pi}(i) = \max\{j < \hat{\pi}(i-1) | \pi(j) = r\}$, $i = j_0+2,\ldots,n$. Using this definition we write

$$\operatorname{Tr} B_{\pi(k+l)}(s_{k+l}) \cdots B_{\pi(1)}(s_1)\omega_\beta = \operatorname{Tr} F(\epsilon^{-2}s_{\hat{\pi}(k+l)}) \cdots F(\epsilon^{-2}s_{\hat{\pi}(1)})\omega_\beta,$$

where all field operators are commuted to the left of ω_β.

For odd $k+l$ R_{kl} vanishes. For even $k+l = 2n$ we obtain from (10)

$$\operatorname{Tr} F(\epsilon^{-2}s_{\hat{\pi}(2n)}) \cdots F(\epsilon^{-2}s_{\hat{\pi}(1)})\omega_\beta = \sum_{p \in P(2n)} \operatorname{sgn} p \prod_{j=1}^{n} g^\epsilon(s_{\sigma(2j)} - s_{\sigma(2j-1)}), \tag{20}$$

with $\sigma = \hat{\pi} \circ p$ and $g^\epsilon(t) = g(\epsilon^{-2}t)$. To estimate the time integral of the correlation function in (19) the domain of integration is enlarged,

$$\int_{\Delta(0,k,t_1,l,t_2)} \underline{ds} |\cdots| \leqslant \int_{\Delta(0,2n,t)} \underline{ds} |\cdots|,$$

and with the estimates from the proof of Theorem 3.4 of Ref. 4 and the symmetrization trick of Lemma 3.3 of Ref. 4 one obtains

$$\int_{\Delta(0,2n,t)} \underline{ds} |\operatorname{Tr} B_{\pi(2n)}(s_{2n}) \cdots B_{\pi(1)}(s_1)\omega_\beta| \leqslant \epsilon^{2n} \frac{(\|g\|_1 t)^n}{n! 2^n}.$$

This proves that the series (15) is majorized by

$$\sum_{\mu=0}^{\infty} \sum_{\nu=0}^{2\mu} \|R_{2\mu-\nu,\nu}\rho\|_1 \leqslant \|X_1\| \; \|Y_1\| \; \|X_2\| \; \|Y_2\| \; \|\rho\|_1 \times \sum_{\mu=0}^{\infty} \frac{2\mu}{\mu!} (2\|g\|_1 \|Q\|^2 t)^\mu,$$

which is absolutely convergent for all t and independent of ϵ.

(2) Term by term estimates: vanishing terms

Inserting (20) into (18) one gets

$$R_{kl}\rho = (i\epsilon)^{-2n} \sum_{\pi \in \Pi(2n)} \operatorname{sgn} \pi \times \sum_{p \in P(2n)} \operatorname{sgn} p \int_{\Delta(0,k,t_1,l,t_2)} \underline{ds} \times \widetilde{E}_2 e^{\epsilon^{-2}L_S t_2} C_{\pi(2n)}(s_{2n}) \cdots \widetilde{E}_1 \cdots C_{\pi(1)}(s_1)\rho \times \prod_{j=1}^{n} g^\epsilon(s_{\sigma(2j)} - s_{\sigma(2j-1)}) = \sum_{\pi \in \Pi(2n)} \sum_{p \in P(2n)} R^\sigma_{kl}\rho, \tag{21}$$

where R^σ_{kl} is defined by the last equation.

First we show that in the limit $\epsilon \to 0$ all R^σ_{kl} vanish for which there is j_0 with $\sigma_1 = \min(\sigma(2j_0-1), \sigma(2j_0)) < k$ and $\sigma_2 = \max(\sigma(2j_0-1), \sigma(2j_0)) > k$. We estimate

$$\int_{\Delta(0,k,t_1,l,t_2)} \underline{ds} \left| \prod_{j=1}^{n} g^\epsilon(s_{\sigma(2j)} - s_{\sigma(2j-1)}) \right| < \int_{t_1}^{t_2} ds_{2n} \cdots \int_{t_1}^{t_2} ds_{k+1} \int_0^{t_1} ds_k \cdots \int_0^{t_1} ds_1 \left| \prod_{j=1}^{n} g^\epsilon(s_{\sigma(2j)} - s_{\sigma(2j-1)}) \right| < \epsilon^{2n-2} t^{n-1} \|g\|_1^{n-1} \int_{0<\sigma_1<t_1<\sigma_2<t_2} ds_{\sigma_1} ds_{\sigma_2} \; |g^\epsilon(s_{\sigma_2} - s_{\sigma_1})| = \epsilon^{2n} t^{n-1} \|g\|_1^{n-1} G(\epsilon,t_1,t_2),$$

where $G(\epsilon,t_1,t_2) = \epsilon^{-2} \int_{\Delta(0,1,t_1,1,t_2)} \underline{ds} |g^\epsilon(s_2 - s_1)|$. In Lemma A1(i) of the Appendix we show $\lim_{\epsilon \to 0} G(\epsilon,t_1,t_2) = 0$.

For the remaining terms the integrals $\int_{\Delta(0,k,t_1)} \underline{ds}$ and $\int_{\Delta(t_1,l,t_2)} \underline{ds}$ factorize. Therefore, it is sufficient to discuss one of them, say the first one. We show that in the limit all R^σ_{kl} vanish for which there is j_0 such that $|\sigma(2j_0-1) - \sigma(2j_0)| > 1$. With σ_1,σ_2 defined as in the preceding paragraph we get

$$\int_{\Delta(0,k,t_1)} \underline{ds} \prod_{j=1}^{k/2} |g^\epsilon(s_{\sigma(2j)} - s_{\sigma(2j-1)})| \leqslant \int_{0<s_{\sigma_1}<s_{\sigma_1+1}<s_{\sigma_2}<t} ds_{\sigma_2} ds_{\sigma_1+1} ds_{\sigma_1} \int_0^t \cdots \int_0^t ds_1 \cdots' \cdots ds_k \times \prod_{j=1}^{k/2} |g^\epsilon(s_{\sigma(2j)} - s_{\sigma(2j-1)})| < \epsilon^k t^{k/2-2} \|g\|_1^{n-1} \int_0^t ds \; G(\epsilon,s,t),$$

where $ds_1 \cdots' \cdots ds_k$ indicates that the variables $s_{\sigma_1}, s_{\sigma_1+1}, s_{\sigma_2}$ are lacking. According to Lemma A1(ii) $\int_0^t ds \; G(\epsilon,s,t)$ converges to zero for $\epsilon \to 0$. So only terms with $|\sigma(2j-1) - \sigma(2j)| = 1$ contribute in the limit.

(3) Term by term estimates: contributing terms

For the terms contributing in the limit we obtain the series

$$\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \int_{\Delta(0,k,t_1,l,t_2)} \underline{ds} \; \widetilde{E}_2 e^{\epsilon^{-2}L_S(t_2 - s_{k+l})} \times K_\epsilon(t_2 - s_{k+l}) e^{\epsilon^{-2}L_S(s_{k+l} - s_{k+l-1})} \times K_\epsilon(s_{k+l} - s_{k+l-1}) \cdots \times e^{\epsilon^{-2}L_S(s_{k+1} - t_1)} \widetilde{E}_1 e^{\epsilon^{-2}L_S(t_1 - s_k)} \times K_\epsilon(t_1 - s_k) \cdots K_\epsilon(s_2 - s_1) e^{\epsilon^{-2}L_S s_1} \rho,$$

where

$$K_\epsilon(t)\rho = \int_0^{\epsilon^{-2}t} ds\, e^{-L_S s}\mathrm{Tr}_B L_I e^{(L_S+L_B)s} L_I \rho\otimes\omega_\beta.$$

On the other hand we have

$$\begin{aligned}\widetilde{E}_2 T_0^\epsilon(t_2-t_1)\widetilde{E}_1 T_0^\epsilon(t_1)\rho &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\int_{\Delta(0,k,t_1,L,t_2)} ds \widetilde{E}_2 e^{\epsilon^{-2}L_S(t_2-s_{k+l})}\\ &\times K\cdots K e^{\epsilon^{-2}L_S(s_{k+1}-t_1)}\\ &\times \widetilde{E}_1 e^{\epsilon^{-2}L_S(t_1-s_k)}K\cdots K e^{\epsilon^{-2}L_S s_1}\rho.\end{aligned}$$

Applying the inequality $\|K_\epsilon(t)-K\|\leqslant 4\|Q\|^2 \times\int_{\mathbf{R}\setminus[-\epsilon^{-2}t,\epsilon^{-2}t]} ds|g(s)|$ completes the proof of the theorem. ∎

The semigroup T_0^ϵ does not preserve positivity, in general,[19] and consequently the quantities $\widetilde{E}_n T_0^\epsilon(t_n-t_{n-1})\widetilde{E}_{n-1}\cdots\widetilde{E}_2 T_0^\epsilon(t_2-t_1)\widetilde{E}_1 T_0^\epsilon(t_1)\rho$ do not define a quantum stochastic process. To obtain a quantum dynamical semigroup T_0^ϵ should be replaced by $T_w^\epsilon(t) = \exp(\epsilon^{-2}L_S + K^\#)t$ constructed from the averaged generator

$$K^\# = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} dt\, e^{-L_S t}Ke^{L_S t}. \tag{22}$$

We assume that H_S has a purely discrete spectrum. Then the limit exists in the strong sense and $T_w^\epsilon(t)$ is a completely positive dynamical semigroup. Under assumption (11) Davies[20] proves for all $\rho\in\mathcal{T}(\mathcal{H}_S)$ and all $t>0$

$$\lim_{\epsilon\to 0}\sup_{0<s<t}\|T_0^\epsilon(t)\rho - T_w^\epsilon(t)\rho\|_1 = 0.$$

This result can be extended to prove the convergence of all multitime correlation functions.

Theorem 2: If H_S has a purely discrete spectrum and if (11) holds

$$\begin{aligned}\lim_{\epsilon\to 0}\sup_{0<t_1<\cdots<t_n<t}\|T^\epsilon(E_n,t_n;\ldots;E_1,t_1)\rho\\ -\widetilde{E}_n T_w^\epsilon(t_n-t_{n-1})\widetilde{E}_{n-1}\cdots\widetilde{E}_1 T_w^\epsilon(t_1)\rho\|_1 = 0\end{aligned}$$

for all $t>0$ and all $\rho\in\mathcal{T}(\mathcal{H}_S)$.

To prove the theorem we need

Lemma 3: For every n and every $\delta>0$ there exists a compact set K_δ and a decomposition $\widetilde{E}_n T_w^\epsilon(t_n-t_{n-1})\cdots\widetilde{E}_1 T_w^\epsilon(t_1)\rho = \rho_1^\epsilon(t_1,\ldots,t_n)+\rho_2^\epsilon(t_1,\ldots,t_n)$ satisfying

$$\rho_1^\epsilon(t_1,\ldots,t_n)\in K_\delta,$$

$$\|\rho_2^\epsilon(t_1,\ldots,t_n)\|_1<\delta,$$

for all $\epsilon>0$ and all $(t_1,\ldots,t_n)\in\Delta(0,n,t)$.

Proof: The proof is by induction over n.

(1) Let $H_S=\Sigma_i\omega_i P_i$ be the spectral decomposition of H_S and put $Q_k=\Sigma_{i=1}^k P_i$. Define $\mathcal{P}_k:\mathcal{T}(\mathcal{H}_S)\to\mathcal{T}(\mathcal{H}_S)$, $\rho\mapsto Q_k\rho Q_k$. Clearly $s\text{-}\lim_{k\to\infty}\mathcal{P}_k=\mathbf{1}$. $\{e^{L_S t}\mathcal{P}_k\rho\,|\,t\in\mathbf{R}\}$ is compact for $k\in\mathbf{N}$. As L_S and $K^\#$ commute $T_w^\epsilon(t)=e^{K^\# t}e^{\epsilon^{-2}L_S t}$. From the continuity of $(t,\rho)\mapsto e^{K^\# t}\rho$ we conclude that $\overline{\{E_1 T_w^\epsilon(t_1)\mathcal{P}_k\rho\,|\,0\leqslant t_1\leqslant t,\epsilon>0\}}$ is also compact. The estimate

$$\|\widetilde{E}_1 T_w^\epsilon(t_1)(\mathbf{1}-\mathcal{P}_k)\rho\|_1\leqslant\|E_1\|\|(\mathbf{1}-\mathcal{P}_k)\rho\|_1$$

shows that for some $k=k_0$ the right-hand side is less than δ. Choose $K_\delta=\overline{\{\widetilde{E}_1 T_w^\epsilon(t_1)\mathcal{P}_{k_0}\rho\,|\,0\leqslant t_1\leqslant t,\epsilon>0\}}$ and $\rho_1^\epsilon(t_1)=\widetilde{E}_1 T_w^\epsilon(t_1)\mathcal{P}_{k_0}\rho$, $\rho_2^\epsilon(t_1)=\widetilde{E}_1 T_w^\epsilon(t_1)(\mathbf{1}-\mathcal{P}_{k_0})\rho$.

(2) Suppose the lemma holds for $n=n_0$. Then there is a decomposition $\widetilde{E}_{n_0}T_w^\epsilon(t_{n_0}-t_{n_0-1})\cdots\widetilde{E}_1 T_w^\epsilon(t_1)\rho = \rho_1^\epsilon(t_1,\ldots,t_{n_0})+\rho_2^\epsilon(t_1,\ldots,t_{n_0})$ satisfying the assumptions of the lemma. As $(\mathbf{1}-\mathcal{P}_k)$ converges to zero for $k\to\infty$ uniformly on compact sets there is a k_0 such that $\|(\mathbf{1}-\mathcal{P}_{k_0})\rho_1^\epsilon(t_1,\ldots,t_{n_0})\|_1<\delta$. We put

$$\begin{aligned}\tilde{\rho}_1^\epsilon(t_1,\ldots,t_{n_0+1}) &= \widetilde{E}_{n_0+1}T_w^\epsilon(t_{n_0+1}-t_{n_0})\mathcal{P}_{k_0}\rho_1^\epsilon(t_1,\ldots,t_{n_0}),\\ \tilde{\rho}_2^\epsilon(t_1,\ldots,t_{n_0+1}) &= \widetilde{E}_{n_0+1}T_w^\epsilon(t_{n_0+1}-t_{n_0})(\mathbf{1}-\mathcal{P}_{k_0})\rho_1^\epsilon(t_1,\ldots,t_{n_0})\\ &\quad+\widetilde{E}_{n_0+1}T_w^\epsilon(t_{n_0+1}-t_{n_0})\rho_2^\epsilon(t_1,\ldots,t_{n_0}).\end{aligned}$$

We obtain the estimate $\|\tilde{\rho}_2^\epsilon(t_1,\ldots,t_{n_0+1})\|_1\leqslant 2\|\widetilde{E}_{n_0+1}\|\delta$. We put $\widetilde{K}_{2\|\widetilde{E}_{n_0+1}\|\delta}=\overline{\{\widetilde{E}_{n_0+1}T_w^\epsilon(s)\mathcal{P}_{k_0}\rho\,|\,0\leqslant s\leqslant t,\epsilon>0,\rho\in K_\delta\}}$. $\widetilde{K}_{2\|\widetilde{E}_{n_0+1}\|\delta}$ is compact and $\tilde{\rho}_1^\epsilon(t_1,\ldots,t_{n_0+1})\in\widetilde{K}_{2\|\widetilde{E}_{n_0+1}\|\delta}$. As δ is arbitrary the lemma is proved.

Proof of Theorem 2: It is sufficient to prove

$$\begin{aligned}\lim_{\epsilon\to 0}\sup_{0<t_1<\cdots<t_n<t}\|\widetilde{E}_n T_0^\epsilon(t_n-t_{n-1})\\ \cdots\widetilde{E}_1 T_0^\epsilon(t_1)\rho - \widetilde{E}_n T_w^\epsilon(t_n-t_{n-1})\\ \cdots\widetilde{E}_1 T_w^\epsilon(t_1)\rho\|_1 = 0.\end{aligned}$$

The proof is by induction over n. For $n=1$ Davies's result applies. Suppose the theorem holds for $n=n_0$. Then we estimate

$$\begin{aligned}&\|\widetilde{E}_{n_0+1}T_0^\epsilon(t_{n_0+1}-t_{n_0})\widetilde{E}_{n_0}\cdots\widetilde{E}_1 T_0^\epsilon(t_1)\rho\\ &\quad-\widetilde{E}_{n_0+1}T_w^\epsilon(t_{n_0+1}-t_{n_0})\widetilde{E}_{n_0}\cdots\widetilde{E}_1 T_w^\epsilon(t_1)\rho\|_1\\ &\leqslant\|\widetilde{E}_{n_0+1}T_0^\epsilon(t_{n_0+1}-t_{n_0})\\ &\quad\times[\widetilde{E}_{n_0}\cdots\widetilde{E}_1 T_0^\epsilon(t_1)\rho-\widetilde{E}_{n_0}\cdots\widetilde{E}_1 T_w^\epsilon(t_1)\rho]\|_1\\ &\quad+\|\widetilde{E}_{n_0+1}[T_0^\epsilon(t_{n_0+1}-t_{n_0})-T_w^\epsilon(t_{n_0+1}-t_{n_0})]\\ &\quad\times\rho_1^\epsilon(t_1,\ldots,t_{n_0})\|_1+2\|\widetilde{E}_{n_0+1}\|\|\rho_2^\epsilon(t_1,\ldots,t_{n_0})\|_1,\end{aligned}$$

where we used the decomposition of Lemma 3. By appropriately choosing δ the last term becomes abritrarily small. For $\epsilon\to 0$ the first term converges to zero by assumption and the fact that $\widetilde{E}_{n_0+1}T_0^\epsilon(t_{n_0+1}-t_{n_0})$ is bounded. The second term converges to zero because $\rho_1^\epsilon(t_1,\ldots,t_{n_0})$ belongs to a compact set. This concludes the proof of the theorem.

The result may also be formulated in the interaction picture. There one considers the dynamics relative to the free motion of the system. Using the commutativity of L_S and $K^\#$ we obtain $s\text{-}\lim_{\epsilon\to 0}e^{-\epsilon^{-2}L_S t}T_0^\epsilon(t)=e^{K^\# t}$. Instead of (4) we consider

$$\begin{aligned}&\mathrm{Tr}_B E_n(-t_n)U^\epsilon(t_n-t_{n-1})E_{n-1}(-t_{n-1})\\ &\quad\cdots E_1(-t_1)U^\epsilon(t_1)\rho\otimes\omega_\beta,\end{aligned}$$

where

$$\begin{aligned}&E_k(t)W\\ &\quad=(e^{i\epsilon^{-2}H_S t}X_k e^{-i\epsilon^{-2}H_S t}\otimes\mathbf{1})W(e^{i\epsilon^{-2}H_S t}Y_k e^{-i\epsilon^{-2}H_S t}\otimes\mathbf{1}).\end{aligned}$$

Slightly modifying the proof of Theorem 1 one obtains

$$\lim_{\epsilon\to 0}\sup_{0<t_1<\cdots<t_n<t}\|\mathrm{Tr}_B E_n(-t_n)U^\epsilon(t_n-t_{n-1})$$

$$\cdots E_1(-t_1)U^\epsilon(t_1)\rho\otimes\omega_\beta-\tilde{E}_n e^{K^\#(t_n-t_{n-1})}\tilde{E}_{n-1}$$

$$\cdots\tilde{E}_1 e^{K^\# t_1}\rho\|_1=0.$$

Let $H_S=\Sigma_{i\in I}\omega_i P_i$ be the spectral decomposition of the system Hamiltonian. For simplicity we assume that all eigenvalues are nondegenerate. $T_w^\epsilon(t)$ leaves invariant the set of density matrices commuting with H_S. $T_w^\epsilon(t)$ induces a classical Markov process $X(t)$ on the state space I with the initial distribution $P(X(0)=i)=\mathrm{Tr}\,P_i\rho$ and the transition probability

$$P(X(t_2)=i|X(t_1)=j)=\mathrm{Tr}\,P_i T_w^\epsilon(t_2-t_1)P_j.$$

The joint probabilities are given by

$$P(X(t_n)=i_n,\ldots,X(t_1)=i_1)$$

$$=\sum_{i_0\in I}P(X(t_n)=i_n|X(t_{n-1})=i_{n-1})$$

$$P(X(t_{n-1})=i_{n-1}|X(t_{n-2})=i_{n-2})$$

$$\cdots P(X(t_1)=i_1|X(0)=i_0)P(X(0)=i_0)$$

$$=\mathrm{Tr}\,\tilde{E}_{i_n}T_w^\epsilon(t_n-t_{n-1})\cdots\tilde{E}_{i_1}T_w^\epsilon(t_1)\rho,$$

where

$$\tilde{E}_i=P_i\rho P_i.$$

However, only in the limit a classical stochastic process is imbedded. The corresponding quantities

$$P^\epsilon(X(t_n)=i_n,\cdots,X(t_1)=i_1)$$

$$=\mathrm{Tr}\,E_{i_n}U^\epsilon(t_n-t_{n-1})$$

$$\cdots E_{i_1}U^\epsilon(t_1)\rho\otimes\omega_\beta$$

for finite ϵ do not form a consistent set of probability measures, in general.

3. THE SINGULAR COUPLING LIMIT

According to Palmer[5] the singular coupling problem can be transformed to a weak coupling problem. One scales the Hamiltonian as

$$H_{sg}^\epsilon=H_S+\epsilon^{-2}H_B+\epsilon^{-1}H_I.$$

We put $U_{sg}^\epsilon(t)W=e^{-iH_{sg}^\epsilon t}We^{iH_{sg}^\epsilon t}$. In this case the dynamics of the system and the dissipation are on the same time scale. The generator does not depend on H_S,

$$K_{sg}\rho=\int_0^\infty ds\,\mathrm{Tr}_B L_I e^{L_B s}L_I\rho\otimes\omega_\beta.$$

The semigroup $T_{sg}(t)=\exp(L_S+K_{sg})t$ is already a completely positive dynamical semigroup and no averaging is necessary.

Theorem 4: If (11) holds

$$\lim_{\epsilon\to 0}\sup_{0<t_1<\cdots<t_n<t}\|\mathrm{Tr}_B E_n U_{sg}^\epsilon(t_n-t_{n-1})$$

$$\times E_{n-1}\cdots E_1 U_{sg}^\epsilon(t_1)\rho\otimes\omega_\beta$$

$$-\tilde{E}_n T_{sg}(t_n-t_{n-1})\tilde{E}_{n-1}$$

$$\cdots\tilde{E}_1 T_{sg}(t_1)\rho\|_1=0$$

for all $t\geqslant 0$ and all $\rho\in\mathcal{T}(\mathcal{H}_S)$.

The proof is similar to the proof of Theorem 1. Only minor changes are necessary to account for the different scaling of the system Hamiltonian.

ACKNOWLEDGMENT

The author would like to thank H. Spohn for suggesting the problem.

APPENDIX

Lemma A 1: Put $G(\epsilon,t_1,t_2)=\epsilon^{-2}\int_{0<s_1<t_1<s_2<t_2}ds_2\,ds_1\,|g^\epsilon(s_2-s_1)|$ for $0\leqslant t_1\leqslant t_2$. Then

(i) $\lim_{\epsilon\to 0}G(\epsilon,t_1,t_2)=0\quad$ for $t_1<t_2$

and

(ii) $\lim_{\epsilon\to 0}\int_0^t ds\,G(\epsilon,s,t)=0.$

Proof: (i) Set $f(x)=\int_x^\infty dy\,|g(y)|$. Then $|f(x)|\leqslant\|g\|_1$ for all x and $\lim_{x\to\infty}f(x)=0$. Using the substitution $s_1'=\epsilon^{-2}(s_2-s_1)$ in the definition of G and extending the upper bound of the resulting integral to infinity one obtains

$$G(\epsilon,t_1,t_2)\leqslant\int_{t_1}^{t_2}ds_2 f(\epsilon^{-2}(s_2-t_1)).\tag{A1}$$

The conclusion follows by dominated convergence.

(ii) $G(\epsilon,s,t)$ converges pointwise for $s<t$ and from (A1) follows the bound $|G(\epsilon,s,t)|\leqslant t\,\|g\|_1$. Application of the Lebesgue theorem completes the proof.

[1] S. Nakajima, Prog. Theor. Phys. **20**, 948 (1958).
[2] R. Zwanzig, J. Chem. Phys. **33**, 1338 (1960); Lect. Theoret. Phys. **3**, 106 (1960).
[3] I. Prigogine and P. Resibois, Physica (Utrecht) **27**, 629 (1961).
[4] E. B. Davies, Commun. Math. Phys. **39**, 91 (1974).
[5] P. F. Palmer, J. Math. Phys. **18**, 527 (1977).
[6] K. Hepp and E. H. Lieb, Helv. Phys. Acta **46**, 573 (1973).
[7] V. Gorini and A. Kossakowski, J. Math. Phys. **17**, 1298 (1976).
[8] A. Frigerio and V. Gorini, J. Math. Phys. **17**, 2123 (1976).
[9] G. Lindblad, "Response of Markovian and non-Markovian quantum stochastic systems to time-dependent forces," preprint (1979) (unpublished).
[10] G. Lindblad, Commun. Math. Phys. **65**, 281 (1979).
[11] H. Haken and W. Weidlich, Z. Phys. **205**, 96 (1967).
[12] R. Bonifacio and F. Haake, Z. Phys. **200**, 526 (1967).
[13] M. Lax, Phys. Rev. **172**, 350 (1968).
[14] R. Graham, F. Haake, H. Haken, and W. Weidlich, Z. Phys. **213**, 21 (1968).
[15] U. Gnutzmann, Z. Phys. **225**, 416 (1969).
[16] F. Haake, Phys. Rev. A **3**, 1723 (1971).
[17] G. G. Emch, *Algebraic methods in statistical mechanics and quantum field theory* (Wiley, New York, 1972).
[18] E. Balslev and A. Verbeure, Commun. Math. Phys. **7**, 55 (1968).
[19] R. Dümcke and H. Spohn, Z. Phys. B **34**, 419 (1979).
[20] E. B. Davies, Math. Ann. **219**, 147 (1976).

Rational von Neumann lattices

M. Boon
Institute for Theoretical Physics, University of Nijmegen, Nijmegen, The Netherlands
J. Zak and I. J. Zucker[a)]
Department of Physics, Technion, Israel Institute of Technology, Haifa, Israel

(Received 11 February 1982; accepted for publication 11 June 1982)

A rational von Neumann lattice is defined as a lattice in phase space with the constants a and b in the x and p directions given by a ratio of integers. Zeros of harmonic oscillator functions in the kq representation on such lattices are found. It is shown that the number of zeros of the kq function determines the number of states by which a set on a von Neumann lattice is overcomplete. Interesting relations between theta functions are derived on the basis of their connection with the harmonic oscillator states in the kq representation.

PACS numbers: 05.30.Ch

I. INTRODUCTION

The concept of a set of states on a lattice in phase space was first introduced by von Neumann in the early thirties.[1] This lattice has an underlying unit cell of area h, the Planck constant, which shows that the fundamental commutation relation of x and p is explicitly contained in its construction. Because of this striking feature such a lattice in phase space becomes physically very attractive. Thus, von Neumann stated[1] that a set of coherent states[2,3] on this lattice is complete. Later this statement was proved independently by Perelomov[4] and by Bargmann *et al.*[5] It was turned out that completeness also holds for general sets of square integrable states on a von Neumann lattice[6] (phase space lattice with a unit cell of area h). This leads one to a very elegant way of constructing complete sets of states in quantum mechanics. There remains, however, a very intriguing feature of sets of states on a von Neumann lattice which is connected with their nonorthogonality. It was first noticed by Perelomov[4] that a set of coherent states on a von Neumann lattice is overcomplete by exactly one state. This is a rather strange feature whose physical meaning is not clear. It turns out that the overcompleteness by one state is connected with the fact that coherent states in the kq representation have exactly one zero in each unit cell of the von Neumann lattice.[6] The existence of this zero influences strongly the nature of the expansion in the discrete set of coherent sets.[7] In particular, if a kq function $C(k,q)$ has any zeros, the von Neumann set built out of $C(k,q)$ cannot be orthogonal on different sites of the lattice, because orthogonality requires $|C(k,q)| = 1$.[6] This feature of von Neumann sets was recently used in proving what is called the strong uncertainty principle in quantum mechanics.[8]

In this paper we show that the von Neumann set built out of the wave function $C(kq)$ is overcomplete to an extent that depends on the zeros of $C(kq)$ in number and kind. Notably, if $C(kq)$ has r isolated simple zeros (zeros of order one) then overcompleteness is by exactly r states. It was recently proved that any continuous wave function $C(k,q)$ has at least one zero.[9] This means that the von Neumann set built out of any continuous $C(k,q)$ will be overcomplete by at least one state. In particular, we investigate zeros of the harmonic oscillator states $C_N(k,q)$ in the kq representation. As is well known, the harmonic oscillator state depends on the parameter λ ($\lambda^2 = \hbar/m\omega$) while the kq representation is defined by using a lattice[10] (we put the constant a of this lattice as a superscript on the function)

$$C^{(a)}(k,q) = \left(\frac{a}{2\pi}\right)^{1/2} \sum_{n=-\infty}^{\infty} \exp(ikan)\psi(q - na), \qquad (1)$$

where $\psi(x)$ is the wave function in the x representation. The constants λ and a appear also in the definition of the von Neumann lattice[7]

$$\alpha_{mn} = \frac{1}{\lambda\sqrt{2}}(na + imb), \quad b = \frac{2\pi}{a}\lambda^2. \qquad (2)$$

In the particular case, when

$$b/a = s/l \qquad (3)$$

where s and l are integers we shall say that the set of points in Eq. (2) forms a *rational* von Neumann lattice. Thus, a special case of a rational von Neumann lattice is a square lattice with $b = a$. In this paper we find sets of zeros of harmonic oscillator states for rational von Neumann lattices. In particular, we consider the lattices $b = sa$ with $s = 1,2,3$, and 4. For these particular rational lattices we find zeros for infinite sets of harmonic oscillator states in the kq representation. These zeros are located at symmetric positions in the unit cell of the von Neumann lattice. For finding the zeros we use a relation connecting kq functions for a lattice with a constant a and a superlattice with a constant sa, s being any integer. This relation together with a canonical transformation gives zeros of harmonic oscillator states on rational von Neumann lattices. Although it is at present uncertain whether there are more zeros than the ones we have found, we gain a great deal of insight into the properties of harmonic oscillator lattices. These results are contained in Secs. I–III. In Sec. IV we discuss the relationship between the zeros and the degree of overcompleteness of the von Neumann lattice set. Finally, Sec. V presents some interesting relations between theta functions that come out as a by-product of the search for zeros of harmonic oscillator states. That there is a connection between coherent states on a von Neumann lattice and theta functions has been known for some time.[4,11] It turns out that this connection is a very far reaching one and it is extended in this paper to higher harmonic oscillator states.

[a)] On leave from Department of Physics, University of Surrey, Guildford, England.

 0022-2488/83/020316-08$02.50

II. RATIONAL VON NEUMANN LATTICES

The Hamiltonian of a harmonic oscillator

$$\hat{H}=\frac{1}{2m}\left(\hat{p}^2+\frac{\hbar^2}{\lambda^4}\hat{x}^2\right), \tag{4}$$

with $\lambda^2=\hbar/m\omega$, is invariant under the following canonical transformation:

$$T:\hat{x}\to-\frac{\lambda^2}{\hbar}\hat{p},\quad \hat{p}\to\frac{\hbar}{\lambda^2}\hat{x}. \tag{5}$$

In the x representation the transformation T assumes the following form:

$$\begin{aligned}T\psi(x) &= \frac{1}{\lambda(2\pi)^{1/2}}\int_{-\infty}^{\infty}\exp\left(\frac{i}{\lambda^2}xy\right)\psi(y)dy\\ &= \frac{\hbar^{1/2}}{\lambda}F\left(-\frac{\hbar x}{\lambda^2}\right),\end{aligned} \tag{6}$$

where $F(p)$ is the Fourier transform of $\psi(x)$

$$F(p)=\frac{1}{(2\pi\hbar)^{1/2}}\int_{-\infty}^{\infty}\exp\left(-\frac{i}{\hbar}px\right)\psi(x)dx. \tag{7}$$

When $\psi(x)$ is the Nth harmonic oscillator state $\psi_N(x)$ its Fourier transform $F_N(p)$ is well known[12]

$$F_N(p)=\frac{(\lambda)^{1/2}}{i^N(2^N N!\hbar\sqrt{\pi})^{1/2}}\exp\left(-\frac{\lambda^2p^2}{2\hbar^2}\right)H_N\left(\frac{\lambda p}{\hbar}\right), \tag{8}$$

where $H_N(x)$ are the Hermite polynomials. From Eqs. (6) and (8), and from the explicit form for $\psi_N(x)$ [see Ref. (12)] we find

$$T\psi_N(x)=i^N\psi_N(x). \tag{9}$$

This means that the canonical transformation (5) when applied to the harmonic oscillator states multiplies the latter by i^N. Such a behavior of the states $|N\rangle$ is in agreement with the fact that they are nondegenerate.

Let us now find Eq. (6) (the first half of it) in the kq representation. For defining the kq function $C(k,q)$ one can also use the Fourier transform $F(p)$ of the wave function[10]

$$\begin{aligned}C^{(a)}(k,q) &= \left(\frac{\hbar}{a}\right)^{1/2}\exp(ikq)\\ &\times\sum_{n=-\infty}^{\infty}\exp\left(iq\frac{2\pi}{a}n\right)F\left(\hbar k+n\frac{2\pi}{a}\hbar\right).\end{aligned} \tag{10}$$

From Eqs. (1), (6), and (10) it follows that

$$TC^{(a)}(k,q)=\exp(ikq)C^{(b)}\left(-\frac{q}{\lambda^2},\lambda^2k\right);\quad b=\frac{2\pi}{a}\lambda^2, \tag{11}$$

where as in Eq. (1) the superscripts denote the lattice constant. Note that the same relationship between b and a occurs as in Eq. (2). Equation (11) expresses the canonical transformation (5) in the kq representation: a kq function $C^{(a)}(k,q)$ is transformed by T into a kq function $C^{(b)}(-q/\lambda^2,\lambda^2k)$ for the constant b, multiplied by the phase $\exp(ikq)$. It is interesting that while in the x representation the canonical transformation (5) is given by an integral [Eq. (6)], in the kq representation one just has to replace in the function $k\to-q/\lambda^2$, $q\to\lambda^2k$, multiply it by the phase $\exp(ikq)$ and change the constant from a to b. This replacement corresponds to the orignal transformation (5) if one remembers that in the kq representation p is replaced by $\hbar k$ and x by q.

For rational von Neumann lattices [see the definition (3)] the right hand side of Eq. (11) can be expressed as a linear combination of kq functions for the constant a. This can be achieved by using formulas expressing kq functions $C^{(sa)}(k,q)$ for constant sa (s an integer) via $C^{(a)}(k,q)$, and vice versa. Such formulas are very easy to obtain and a short derivation of them is given in the Appendix. [Formulas (A4) and (A5)]. By using these formulas, the canonical transformation (11) can be written in terms of kq functions for the constant a. Thus, when $b=sa$ we have

$$TC^{(a)}(k,q)=\frac{\exp(ikq)}{(s)^{1/2}}\sum_{r=0}^{s-1}C^{(a)}\left(-\frac{q}{\lambda^2}+r\frac{2\pi}{sa},\lambda^2k\right). \tag{12}$$

For a general rational von Neumann lattice $b=sa/l$ [Eq. (3)], we can combine the formulas (A4) and (A5) in the Appendix and express $C^{(b)}(k,q)$ as a combination of the functions $C^{(a)}(k,q)$. We have

$$\begin{aligned}C^{(b)}(k,q) &= \frac{1}{(s)^{1/2}}\sum_{r=0}^{s-1}C^{(a/l)}\left(k+r\frac{2\pi}{sa},q\right)\\ &= \frac{1}{(sl)^{1/2}}\sum_{r=0}^{s-1}\sum_{r'=0}^{l-1}C^{(a)}\left(k+r\frac{2\pi}{s(a/l)},q-r\frac{a}{l}\right)\\ &\times\exp(ikr'a/l).\end{aligned} \tag{13}$$

It is easy to see that for $l=1$ ($b=sa$) this formula goes over into (A4), while for $s=1$ ($b=a/l$) it is equivalent to (A5). Correspondingly, the formula (13) can be used in order to express the canonical transformation (11) in terms of kq functions for the parameter a only.

It is interesting that in the general case the canonical transformation (11) connects the function $C^{(a)}(k,q)$ for the constant a with the kq function for the constant b. These two constants a and b are related to one another by Eq. (2) and they appear in the definition of a von Neumann lattice. That is, the canonical transformation (11) connects functions for the particular constants a and b belonging to a von Neumann lattice, which therefore appears quite naturally in this context. Other lattices, where $ab\neq2\pi\lambda^2$ (cell area $\neq h$) lead either to incompleteness ($ab>2\pi\lambda^2$) or to infinite overcompleteness ($ab<2\pi\lambda^2$).[4,5] The von Neumann lattice, on the other hand, will permit exact completeness or overcompleteness by a finite number of states, depending on the function $C^{(a)}(k,q)$.

For harmonic oscillator states $C_N(k,q)$ we can use Eq. (9) and obtain an interesting formula from the canonical transformation (11)

$$i^NC_N^{(a)}(k,q)=\exp(ikq)C_N^{(b)}(-q/\lambda^2,\lambda^2k);\quad b=\frac{2\pi}{a}\lambda^2. \tag{14}$$

This formula connects harmonic oscillator functions for the constants a and b on a von Neumann lattice. As we show in Sec. V it is equivalent to the Jacobi imaginary transformation of elliptic functions[13] or a derivative thereof. When combined with Eq. (13) for a rational von Neumann lattice formula (14) leads to the following results [for $b=(s/l)a$]:

$$i^N C_N^{(a)}(k,q) = \frac{\exp(ikq)}{(sl)^{1/2}} \sum_{r=0}^{s-1} \sum_{r'=0}^{l-1} C_N^{(a)} \times\left(-\frac{q}{\lambda^2} + r\frac{2\pi}{s(a/l)}, \lambda^2 k - r'\frac{a}{l}\right) \times\exp\left(ikr'\frac{a}{l}\right). \tag{15}$$

This formula connects harmonic oscillator functions at different points in the unit cell of the rational von Neumann lattice with $b = (s/l)a$. Since $2\pi\lambda^2 = ab$ [from (14)] we also have

$$\frac{a^2}{\lambda^2} = 2\pi\frac{l}{s}. \tag{16}$$

The unit cell in formula (15) is therefore given by $a = \lambda(2\pi l/s)^{1/2}$ and $b = \lambda(2\pi s/l)^{1/2}$. Formula (15) has a number of interesting consequences, which are discussed in the Secs. III and V of this paper.

III. ZEROS OF HARMONIC OSCILLATOR STATES IN THE kq REPRESENTATION

As was mentioned in the Introduction, every continuous kq function has at least one zero in the unit cell of the von Neumann lattice.[9] From some general considerations it is easy to show that any $C_{\text{even}}^{(a)}(k,q)$ which is built from an even function $\psi_{\text{even}}(x)$ in the x representation, $\psi_{\text{even}}(-x) = \psi_{\text{even}}(x)$ has a zero at $k = \pi/a$, $q = a/2$

$$C_{\text{even}}^{(a)}\left(\frac{\pi}{a}, \frac{a}{2}\right) = 0. \tag{17}$$

Similarly, one can show that $C_{\text{odd}}^{(a)}(k,q)$ which is built from an odd function $\psi_{\text{odd}}(-x) = -\psi_{\text{odd}}(x)$, has zeros at the following three points in the unit cell of the von Neumann lattice:

$$C_{\text{odd}}^{(a)}(0,0) = C_{\text{odd}}^{(a)}(0,a/2) = C_{\text{odd}}^{(a)}(\pi/a,0) = 0. \tag{18}$$

For showing this we use the boundary conditions satisfied by any kq function[10]

$$C^{(a)}(k + 2\pi/a, q) = C^{(a)}(k,q), \tag{19}$$

$$C^{(a)}(k, q + a) = \exp(ika)C^{(a)}(k,q), \tag{20}$$

and the fact which follows from the definition (1) that $C_{\text{even}}^{(a)}(-k,-q) = C_{\text{even}}^{(a)}(k,q)$ and $C_{\text{odd}}^{(a)}(-k,-q) = -C_{\text{odd}}^{(a)}(k,q)$. Thus, for an even function $C_{\text{even}}^{(a)}(k,q)$ we have from the boundary conditions (19) and (20)

$$C_{\text{even}}^{(a)}(\pi/a, a/2) = C_{\text{even}}^{(a)}(-\pi/a, -a/2) = -C_{\text{even}}^{(a)}(\pi/a, a/2). \tag{21}$$

From here the zero in (17) follows. In a similar way we verify the zeros of an odd function $C_{\text{odd}}^{(a)}(k,q)$ as given in Eq. (18). It immediately follows that for harmonic oscillator states for any M we have

$$C_{2M}^{(a)}(\pi/a, a/2) = 0 \tag{22}$$

$$C_{2M+1}^{(a)}(0,0) = C_{2M+1}^{(a)}(0,a/2) = C_{2M+1}^{(a)}(\pi/a,0) = 0. \tag{23}$$

These zeros appear for an arbitrary constant a and are connected with the fact that harmonic oscillator states are even for $N = 2M$ and odd for $N = 2M + 1$.

From Eq. (14) we can also obtain for the harmonic oscillator states the following general results about their zeros. Let us assume that there is a zero of $C_N^{(a)}(k,q)$ at $k = \alpha\pi/a$ and $q = \beta a/2$ with $|\alpha| \leqslant 1$ and $|\beta| \leqslant 1$. From Eq. (14) it then follows that $C_N^{(b)}(-\beta\pi/b, \alpha b/2) = 0$. Since the harmonic oscillator functions $\psi_N(x)$ are real, it is easy to see that $C_N^*(k,q) = C_N(-k,q)$. We arrive therefore at the following general results if

$$C_N^{(a)}\left(\alpha\frac{\pi}{a}, \beta\frac{a}{2}\right) = 0$$

then

$$C_N^{(a)}\left(\pm\alpha\frac{\pi}{a}, \pm\beta\frac{a}{2}\right) = 0, \qquad C_N^{(b)}\left(\pm\beta\frac{\pi}{b}, \pm\alpha\frac{b}{2}\right) = 0, \tag{24}$$

where the constants a and b are connected by Eq. (2). Equation (24) does not tell us, however, where the zeros actually are. In order to look for them we now use the results of the previous section for rational von Neumann lattices.

Let us start with the simplest case of a square von Neumann lattice, $b = a$. In this case, Eq. (14) becomes ($\lambda^2 = a^2/2\pi$),

$$i^N C_N^{(a)}(k,q) = \exp(ikq)C_N^{(a)}\left(-\frac{2\pi q}{a^2}, \frac{a^2}{2\pi}k\right). \tag{25}$$

From this it follows that for $N \neq 4n$ (n is any integer), $C_N^{(a)}(0,0) = 0$. For an odd N this is not a new result and it is contained in Eq. (18). However, when $N = 2 + 4n$, this is a new zero and we have

$$C_{2+4n}^{(a)}(0,0) = 0, \quad n \text{ arbitrary}, \quad \lambda^2 = a^2/2\pi. \tag{26}$$

This relation shows that the even harmonic oscillator functions in the series $N = 2 + 4n$ have a zero in the kq representation at the origin of the square von Neumann lattice unit cell.

From Eq. (25) we find also the following new zero for odd harmonic oscillator states. It is easy to check that

$$C_{3+4n}^{(a)}\left(\frac{\pi}{a}, \frac{a}{2}\right) = 0, \quad n \text{ arbitrary}, \quad \lambda^2 = \frac{a^2}{2\pi}. \tag{27}$$

It is to be remarked from (17) that for general von Neumann lattices the point $k = \pi/a, q = a/2$ is a zero only for even harmonic oscillator states. It is therefore interesting that for square lattices Eq. (27) holds for the odd states of the series $N = 3 + 4n$ (we shall see in Sec. IV that (26) and (27) are actually zeros of order two).

Next we consider the rational case $b = 2a$. By using Eq. (15) for $s = 2, l = 1$ (or Eq. (A4) in the Appendix for $s = 2$) we have ($\lambda^2 = a^2/\pi$),

$$i^N C_N^{(a)}(k,q) = \frac{\exp(ikq)}{\sqrt{2}}\left[C_N^{(a)}\left(-\frac{q\pi}{a^2}, \frac{a^2}{\pi}k\right) + C_N^{(a)}\left(-\frac{q\pi}{a^2} + \frac{\pi}{a}, \frac{a^2}{\pi}k\right)\right]. \tag{28}$$

Consider Eq. (28) at the point $k = \pi/2a$, $q = a/2$. We have

$$i^N C_N^{(a)}\left(\frac{\pi}{2a}, \frac{a}{2}\right) = \frac{\exp(i\pi/4)}{\sqrt{2}}\left[C_N^{(a)}\left(-\frac{\pi}{2a}, \frac{a}{2}\right) + C_N^{(a)}\left(\frac{\pi}{2a}, \frac{a}{2}\right)\right]. \tag{29}$$

However, since

$$C_N^{(a)}\left(-\frac{\pi}{2a},\frac{a}{2}\right)=(-1)^N\exp\left(-i\frac{\pi}{2}\right)C_N^{(a)}\left(\frac{\pi}{2a},\frac{a}{2}\right),$$

we find

$$C_{2+4n}^{(a)}\left(\frac{\pi}{2a},\frac{a}{2}\right)=C_{3+4n}^{(a)}\left(\frac{\pi}{2a},\frac{a}{2}\right)=0,\quad n \text{ arbitrary},\ \lambda^2=\frac{2a^2}{\pi}. \tag{30}$$

This shows that the point $k=\pi/2a$, $q=a/2$ is a new zero for both the above series of even and odd states when $b=2a$.

In a similar way, one can carry out the zero-searching process for a rational von Neumann lattice with $b=3a$. In this case the zeros for the functions $C_{2+4n}^{(a)}$ are at $k=\pi/3a$, $q=a/2$ while for the functions $C_{3+4n}^{(a)}$ they are at $k=2\pi/3a$, $q=0$. We have

$$C_{2+4n}^{(a)}(\pi/3a,a/2)=C_{3+4n}^{(a)}(2\pi/3a,0)=0,\quad n \text{ arbitrary},\ \lambda^2=3a^2/2\pi. \tag{31}$$

We have also looked for the zeros in the case of $b=4a$. We found no new zeros for the kq functions of even harmonic oscillator states. For odd states we have

$$C_{3+4n}^{(a)}(\pi/2a,0)=0,\quad n \text{ arbitrary},\ \lambda^2=2a^2/\pi. \tag{32}$$

The results of the zeros for harmonic oscillator states in the kq representations are summarized in Table I and Fig. 1. In Table I we list the zeros of the functions $C_{2+4n}^{(a)}(k,q)$ and $C_{3+4n}^{(a)}(k,q)$ [correspondingly, for $C_{2+4n}^{(b)}(k,q)$ and $C_{3+4n}^{(b)}(k,q)$; see Eq. (24)] for rational lattices $b=sa$, $s=1,2,3,4$. In this Table we summarize also the zeros for even $C_{2M}(k,q)$ and odd $C_{2M+1}(k,q)$ harmonic oscillator functions. Figure 1 shows the zeros in the unit cell of the rational von Neumann lattices for the series $N=2+4n$ and $N=3+4n$.

Knowledge about the zeros of the harmonic oscillator functions is pertinent for an examination of the completeness of the corresponding von Neumann lattice sets. This is discussed in the next Section. It also turns out that the new zeros [Eqs. (26), (27), and (30)–(32)] lead to some interesting relations for theta functions. This is described in Sec. V.

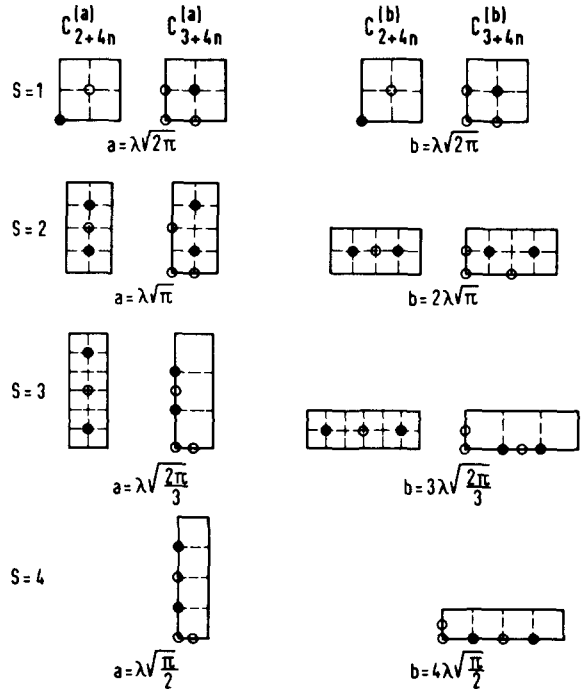

FIG. 1. Zeros of harmonic oscillator functions in the kq representation for rational von Neumann Lattices. a and b are the constants of the rational lattice, $b=sa$. The q-coordinate is on the horizontal axis and k is on the vertical one. The numerical values for the zeros are given in Table I. Filled circles show zeros derived from the rationality condition; open circles show zeros present for arbitrary a.

IV. ZEROS AND COMPLETENESS

Zeros of the kq wave function $C(k,q)$ play an important role for the completeness properties of the set of states gener-

TABLE I. Zeros of harmonic oscillator functions. In the upper part zeros are listed for rational von Neumann lattices. The lower part contains the zeros of even and odd functions for an arbitrary constant a. n and M are arbitrary nonnegative integers.

	$C_{2+4n}^{(a)}$	$C_{3+4n}^{(a)}$	$C_{2+4n}^{(b)}$	$C_{3+4n}^{(b)}$
$s=1$: $a=\lambda\sqrt{2\pi}$ $(b=\lambda\sqrt{2\pi})$	$(0,0)$	$\left(\frac{\pi}{a},\frac{a}{2}\right)$	$(0,0)$	$\left(\frac{\pi}{b},\frac{b}{2}\right)$
$s=2$: $a=\lambda\sqrt{\pi}$ $(b=2\lambda\sqrt{\pi})$	$\left(\pm\frac{\pi}{2a},\frac{a}{2}\right)$	$\left(\pm\frac{\pi}{2a},\frac{a}{2}\right)$	$\left(\frac{\pi}{b},\pm\frac{b}{4}\right)$	$\left(\frac{\pi}{b},\pm\frac{b}{4}\right)$
$s=3$: $a=\lambda\sqrt{\frac{2\pi}{3}}$ $\left(b=3\lambda\sqrt{\frac{2\pi}{3}}\right)$	$\left(\pm\frac{\pi}{3a},\frac{a}{2}\right)$	$\left(\pm\frac{2\pi}{3a},0\right)$	$\left(\frac{\pi}{b},\pm\frac{b}{6}\right)$	$\left(0,\pm\frac{b}{3}\right)$
$s=4$: $a=\lambda\sqrt{\frac{\pi}{2}}$ $\left(b=4\lambda\sqrt{\frac{\pi}{2}}\right)$		$\left(\pm\frac{\pi}{2a},0\right)$		$\left(0,\pm\frac{b}{4}\right)$
	$C_{2M}^{(a)}$		$C_{2M+1}^{(a)}$	
	$\left(\frac{\pi}{a},\frac{a}{2}\right)$		$(0,0)$	
			$\left(0,\frac{a}{2}\right)$	
			$\left(\frac{\pi}{a},0\right)$	

ated from $C^{(a)}(k,q)$ on the von Neumann lattice. It was shown in Ref. 9 that this set has the simple form

$$C^{(a)}_{mn}(k,q) = \exp\left(i\frac{2\pi}{a}qm - iakn\right)C^{(a)}(k,q) \tag{33}$$

in the kq representation. Physically, this corresponds to a set of functions generated by shifts in position and momentum corresponding to the points α_{mn} of the lattice (2). The set is complete when

$$\int_0^{\frac{2\pi}{a}} dk \int_0^a dq\, f^*(k,q)C^{(a)}_{mn}(k,q) = 0 \tag{34}$$

for all pairs (m,n) implies that $f(k,q)\equiv 0$ (the zero function) for any square-integrable state $f(k,q)$.[14] It follows[9] from (33) and Fourier-sum theory that completeness holds if and only if

$$f^*(k,q)C^{(a)}(k,q)\equiv 0 \Rightarrow f(k,q)\equiv 0.$$

In the case of the harmonic oscillator state $C^{(a)}_N(k,q)$ completeness certainly holds because $C^{(a)}_N(k,q)$ does not vanish on a finite area of kq coordinates. In fact, the $C^{(a)}_N(k,q)$ are analytic functions of k and q. To begin with, $C^{(a)}_0(k,q)$ has been shown[6] to be related to a theta function[13]

$$C^{(a)}_0(k,q) = \left(\frac{a}{2\pi}\frac{1}{\lambda(\pi)^{1/2}}\right)^{1/2}\exp\left(-\frac{q^2}{2\lambda^2}\right) \times\Theta_3\left(\frac{ka}{2} - i\frac{qa}{2\lambda^2}\,\middle|\,\frac{ia^2}{2\pi\lambda^2}\right) \tag{35}$$

whereas for the higher functions

$$C^{(a)}_N(k,q) = (N!)^{-1/2}\left[\frac{1}{\lambda\sqrt{2}}\left(q + i\frac{\partial}{\partial k} - \lambda^2\frac{\partial}{\partial q}\right)\right]^N C^{(a)}_0(k,q). \tag{36}$$

Here the operator in brackets is the creation operator

$$\hat{a}^\dagger = \frac{1}{\lambda\sqrt{2}}\left(\hat{x} - i\frac{\lambda^2}{\hbar^2}\hat{p}\right)$$

in the kq representation.[10] We know from the properties of theta functions[13] that $C^{(a)}_0(k,q)$ vanishes only at the one point $k=\pi/a$, $q=a/2$, and it is clear from (35) and (36) that all $C^{(a)}_N(k,q)$ are analytic functions and vanish at most on sets of zero measure.

We saw in Sec. III that $C^{(a)}_N(k,q)$ always has at least one zero when N is even and at least three when N is odd. This holds for the general case, i.e., arbitrary a. For values of a corresponding to rational von Neumann lattices we discovered further series of zeros (Table I and Fig. 1) when $N = 2 + 4n$ and $N = 3 + 4n$. In addition we can show that these new zeros (indicated by filled circles in Fig. 1) are actually of order two for the square lattice $s = 1$, i.e. their first derivatives also vanish. This is seen from the recurrence relations

$$\frac{1}{\lambda\sqrt{2}}\left(q + i\frac{\partial}{\partial k} - \lambda^2\frac{\partial}{\partial q}\right)C^{(a)}_N(k,q) = (N+1)^{1/2}C^{(a)}_{N+1}(k,q),$$

$$\frac{1}{\lambda\sqrt{2}}\left(q + i\frac{\partial}{\partial k} + \lambda^2\frac{\partial}{\partial q}\right)C^{(a)}_N(k,q) = N^{1/2}C^{(a)}_{N-1}(k,q),$$

obtained by operation with creation and annihilation operators. Evidently, if $C^{(a)}_N(k,q)$ has a zero at (k,q) then it is of order two there if and only if both $C^{(a)}_{N+1}(k,q)$ and $C^{(a)}_{N-1}(k,q)$ have zeros at the same point. A glance at Fig. 1 then shows that for $s = 1$ the zero at (0,0) for $N = 2 + 4n$ is of order two, and the zero at $(\pi/a, a/2)$ for $N = 3 + 4n$ is also of order two, where we also use (18).

Although we have not yet complete information on the zeros of $C^{(a)}_N(k,q)$, the discovery of new zeros and of zeros of order higher than one, prompts the question of the relation between the completeness properties of the von Neumann set generated from $C^{(a)}_N$ as in (33) and the zeros themselves. It was pointed out in Ref. (9) that in the $N = 0$ case, which gives the von Neumann set of coherent states, the single zero (of first order) of $C^{(a)}_0(k,q)$ has the consequence that the set is overcomplete by just one member. We can prove the following:

If $C^{(a)}(k,q)$ has r isolated zeros of order one, the corresponding von Neumann set is overcomplete by just r members. (There is no unique choice of members.) We assume $C^{(a)}(k,q)$ to be at least differentiable.

Thus we show that we can always find r members whose removal leaves the set still complete. Let us remove the members labeled by the r points of the lattice

$$\{(jL, jL')\,|\, j = 0,1,\ldots,r-1\},\quad (L,L' \text{ integers})$$

on a line through the origin and (L,L'). From (13), (34), and Fourier theory any $f(k,q)$ orthogonal to the remaining set of states satisfies

$$f^*(k,q)C^{(a)}(k,q) = \sum_{j=0}^{r-1} a_j \exp j\left(i\frac{2\pi}{a}qL - iakL'\right) \tag{37}$$

for some set of coefficients a_j. Since $C^{(a)}(k,q)$ has zeros of order one, for $f(k,q)$ to be square integrable it is necessary and sufficient that the right-hand side of (37) vanish at these zeros. Suppose they are at (k_l,q_l), $l = 1,2,\ldots,r$. Then we have only to choose the lattice point (L,L') such that the r quantities

$$\exp\left(i\frac{2\pi}{a}q_lL - iak_lL'\right)j,\quad l = 1,2,\ldots,r$$

are all different, and this is always possible. With such a (nonunique) choice, Eq. (37) for the arguments (k_l,q_l) gives a set of homogeneous linear equations in the a_j with a nonvanishing determinant. The last assertion holds because the determinant is a simple alternant of different arguments.[15] Consequently, all $a_j = 0$ and thus $f(k,q)\equiv 0$.

To complete the proof we point out that the removal of any $(r+1)$ members always leaves an incomplete set. This is because we will arrive at r homogeneous linear equations in $(r+1)$ unknown coefficients with, consequently, a nontrivial solution for the a_j. There exist then nontrivial (square integrable) $f(k,q)$ orthogonal to the remaining set.

The extension of the above proposition to the case when there are zeros of order higher than one goes along similar lines but is considerably more tedious. For each zero at which higher derivatives vanish we can remove just so many members as the number of vanishing derivatives in addition to the original member, and retain completeness. Thus for each zero of order two, we can remove three members. We omit the proof, except to remark that it is similar to the above but involves so-called confluent alternants.[15]

The above results applied to von Neumann lattices of

harmonic oscillator states mean that for the general case (arbitrary a) overcompleteness is by at least one member when N is even and by at least 3 members when N is odd. (For $N=0$ overcompleteness is by precisely one[6] from our earlier discussion, a fact that was originally proved in a totally different manner[4].) For rational lattices with $s=1,2,3,4$ we learn from Fig. 1 that for the series $N=2+4n$ overcompleteness is by at least 3 and in the square ($s=1$) case by at least 4; likewise for the series $N=3+4n$ overcompleteness is by at least 5 and in the square case by at least 6. (Here we recall that in the square case one zero is of order two.) The qualification "at least" is, of course, inserted because for $N>0$ we are not yet sure to have made an inventory of all the zeros.

V. HARMONIC OSCILLATOR STATES AND THETA FUNCTION RELATIONS

The results which hold amongst kq representations of harmonic oscillator states lead to quite a variety of relations among Θ functions. Many of these relations appear to be new. The connection between $C_0^{(a)}(k,q)$ and the Θ_3 function is given in (35). We note that Eq. (14) for $N=0$ amounts to the Jacobi imaginary transformation,[13] whereas for $N>0$ it gives derivatives of that transformation. When (35) is put into the superlattice transformation formula (A4) for $N=0$ it leads to the following additive decomposition result for Θ_3:

$$s\Theta_3(sz|s\tau)=\sum_{r=0}^{s-1}\Theta_3\left(z+\frac{r\pi}{s}\middle|\frac{\tau}{s}\right), \tag{38}$$

writing $z=\frac{1}{2}(ka-iqa/\lambda^2)$.

A decomposition of $\Theta_4(sz|s\tau)$ in product form is known, namely[16]

$$\Theta_4(sz|s\tau)=\frac{Q_0(x^s)}{Q_0^s(x)}\prod_{r=0}^{s-1}\Theta_4\left(z+\frac{r\pi}{s}\middle|\tau\right), \tag{39}$$

where $Q_0(x)=\Pi_1^\infty(1-x^{2n})$ with $x=\exp(i\pi\tau)$. (We use x and not the usual q to avoid confusion with the symbol used in the kq representation.) A product decomposition of Θ_3 may be found from (39) using the result $\Theta_3(z|\tau)=\Theta_4(z+\pi/a|\tau)$. However, (38) is new to us and we have not found it anywhere. Equation (38) is directly obtainable from the definition of Θ_3 as

$$\Theta_3(z|\tau)=\sum_{-\infty}^{\infty}\exp(i\pi n^2\tau)\exp(2inz), \tag{40}$$

for then from (40)

$$\begin{aligned}\sum_{r=0}^{s-1}&\Theta_3\left(z+\frac{r\pi}{s}\middle|\frac{\tau}{s}\right)\\ &=\sum_{n=-\infty}^{\infty}\exp i\left(\pi n^2\frac{\tau}{s}+2nz\right)\sum_{r=0}^{s-1}\exp i\left(2\pi\frac{r}{s}n\right)\\ &=\sum_{n=-\infty}^{\infty}\exp i\left(\pi n^2\frac{\tau}{s}+2nz\right)\left(s\sum_{m=-\infty}^{\infty}\delta_{n,ms}\right)\\ &=s\sum_{m=-\infty}^{\infty}\exp i(\pi m^2\tau s+2msz)\\ &=s\Theta_3(sz|s\tau).\end{aligned}$$

If instead of using $N=0$ in (A4) we use greater values of N we obtain the derivatives of (38).

Equation (36) can be conveniently expressed in the form

$$C_N^{(a)}(k,q)=\frac{1}{(2^N N!)^{1/2}}H_N\left(\frac{i\frac{\partial}{\partial k}+q}{\lambda}\right)C_0^{(a)}(k,q), \tag{41}$$

where H_N is the Nth Hermite polynomial. Formula (41) can be obtained in the following way. In the x representation the Nth harmonic oscillator state is

$$\psi_N(x)=\frac{1}{(2^N N!)^{1/2}}H_N\left(\frac{x}{\lambda}\right)\psi_0(x),$$

where $\psi_0(x)$ is the ground state. In the kq representation this will assume the form (41), if we look at $H_N(x/\lambda)$ as an operator applied to the ground state [see Eq. (35)] and remember that $\hat{x}=q+i\partial/\partial k$ in the kq representation.[10] An alternative expression of harmonic oscillator states in the kq representation was given by Janssen.[9]

In Sec. III zeros of $C_N^{(a)}(k,q)$ were found for various s and (k,q). These, when expressed in terms of Θ functions (Table II), lead to such a plethora of identities as to be almost embarrassing. Consider for example the square lattice for which $C_{4n+2}^{(a)}(0,0)=0$. Using T.1 of Table II it is almost trivial to derive such results as

$$\pi\Theta_3''(0|i)+\Theta_3(0|i)=0, \tag{42a}$$

$$\pi^3\Theta_3^{\rm vi}(0|i)+15\pi^2\Theta_3^{\rm iv}(0|i)-30\Theta_3(0|i)=0, \tag{42b}$$

which are the first two members of T.1.

However, a direct proof from properties of Θ functions is by no means obvious, and we sketch one here to show the difficulties. It is known that[13]

$$\frac{\Theta_3'(z|\tau)}{\Theta_3(z|\tau)}=f(z|\tau)=4\sum_{r=1}^{\infty}\frac{(-1)^r x^r}{1-x^{2r}}\sin 2rz. \tag{43}$$

For $\tau=i$, $x=e^{-\pi}$; then

$$f(z|i)=2\sum_{r=1}^{\infty}(-1)^r\,\mathrm{csch}(r\pi)\sin 2rz,\quad f(0|i)=0.$$

TABLE II. Some θ-function identities derived from Rels. (35) and (41) and the new zeros. Here $n\geq 0$ and $z=\frac{1}{2}(ak-iaq/\lambda^2)$.

$s=1$ (Square) lattice: $a=\lambda(2\pi)^{1/2}=b$

$$C_{4n+2}(0,0)=0\Rightarrow H_{4n+2}\left(i\sqrt{\frac{\pi}{2}}\frac{\partial}{\partial z}\right)\Theta_3(z|i)\Big|_{z=0}=0 \qquad \text{T.1}$$

$$C_{4n+3}\left(\frac{\pi}{a},\frac{a}{2}\right)=0\Rightarrow H_{4n+3}\left(i\sqrt{\frac{\pi}{2}}\frac{\partial}{\partial z}\right)\Theta_1(z|i)\Big|_{z=0}=0 \qquad \text{T.2}$$

$s=2$ lattice: $a=\lambda\sqrt{\pi}=b/2$

$$C_{4n+2}\left(\frac{\pi}{2a},\frac{a}{2}\right)=0\Rightarrow H_{4n+2}\left(i\sqrt{\frac{\pi}{4}}\frac{\partial}{\partial z}\right)\Theta_2\left(z\middle|\frac{i}{2}\right)\Big|_{z=\pi/4}=0 \qquad \text{T.3}$$

$$C_{4n+3}\left(\frac{\pi}{2a},\frac{a}{2}\right)=0\Rightarrow H_{4n+3}\left(i\sqrt{\frac{\pi}{4}}\frac{\partial}{\partial z}\right)\Theta_2\left(z\middle|\frac{i}{2}\right)\Big|_{z=\pi/4}=0 \qquad \text{T.4}$$

$s=3$ lattice: $a=\lambda(2\pi/3)^{1/2}=b/3$

$$C_{4n+2}\left(\frac{\pi}{3a},\frac{a}{2}\right)=0\Rightarrow H_{4n+2}\left(i\sqrt{\frac{\pi}{6}}\frac{\partial}{\partial z}\right)\Theta_2\left(z\middle|\frac{i}{3}\right)\Big|_{z=\pi/6}=0 \qquad \text{T.5}$$

$$C_{4n+3}\left(\frac{2\pi}{3a},0\right)=0\Rightarrow H_{4n+3}\left(i\sqrt{\frac{\pi}{6}}\frac{\partial}{\partial z}\right)\Theta_3\left(z\middle|\frac{i}{3}\right)\Big|_{z=\pi/3}=0 \qquad \text{T.6}$$

$s=4$ lattice: $a=\lambda(\pi/2)^{1/2}=b/4$

$$C_{4n+3}\left(\frac{\pi}{2a},0\right)=0\Rightarrow H_{4n+3}\left(i\sqrt{\frac{\pi}{8}}\frac{\partial}{\partial z}\right)\Theta_3\left(z\middle|\frac{i}{4}\right)\Big|_{z=\pi/4}=0 \qquad \text{T.7}$$

By continuous differentiation with respect to z we have

$$f'(z|i) = 4\sum_{r=1}^{\infty}(-1)^r r\,\mathrm{csch}(r\pi)\cos 2rz,$$

$$f''(z|i) = -8\sum_{r=1}^{\infty}(-1)^r r^2\,\mathrm{csch}(r\pi)\sin 2rz.$$

It is evident that $f^{2n}(0|i) = 0$ for all n. Now from (43)

$$\Theta_3''(z|i) = \Theta_3'(z|i)f(z|i) + \Theta_3(z|i)f'(z|i),$$

hence

$$\Theta_3''(0|i) = \Theta_3(0|i)\times 4\sum_{r=1}^{\infty}(-1)^r r\,\mathrm{csch}(r\pi). \tag{44}$$

The sum $\Sigma_{r=1}^{\infty}(-1)^r r\,\mathrm{csch}(r\pi)$ has a long history[17] and it was first evaluated by Cauchy[18] who showed that it is equal to $-1/4\pi$. Equation (42a) follows immediately from (44). By further differentiation (42b) may also be proved with the aid of the unexpected result that in addition to the even derivatives $f^{2n}(0|i) = 0$, also $f^{v}(0|i) = \Sigma_{r=1}^{\infty}(-1)^r r^5 \times \mathrm{csch}(r\pi) = 0$. This result also goes back to Cauchy[18] and is part of a general formula, namely

$$\sum_{r=1}^{\infty}(-1)^r r^{4n+1}\,\mathrm{csch}(r\pi) = 0, \quad n > 1. \tag{45}$$

The similarity between T.1 of Table II and (45) is evident but the connection not obvious.

The second result for the square lattice (Table II, T.2) gives for its first two members

$$\pi\Theta_1'''(0|i) + 3\Theta_1'(0|i) = 0, \tag{46a}$$

$$\pi^3\theta_1^{vii}(0|i) + 21\pi^2\Theta_1^{v}(0|i) - 210\Theta_1'(0|i) = 0. \tag{46b}$$

Again, these results may be proved directly with some difficulty in the same manner as for (42) using the result

$$\frac{\Theta_1'(z|\tau)}{\Theta_1(z|\tau)} = \cot z + 4\sum_{r=1}^{\infty}\frac{x^{2r}}{1-x^{2r}}\sin 2rz.$$

Equation (46a) can be proved knowing that

$$\sum_{r=1}^{\infty}\frac{r}{e^{2\pi r}-1} = \frac{1}{4}\sum_{r=1}^{\infty}\mathrm{csch}^2(r\pi) = \frac{1}{24} - \frac{1}{8\pi}. \tag{47}$$

To prove (46b) and higher analogs requires knowing that

$$\sum_{r=1}^{\infty}\frac{r^{4n+1}}{e^{2\pi r}-1} = \frac{B_{4n+2}}{4(2n+1)}, \quad n > 1, \tag{48}$$

where B_n is the nth Bernoulli number. Equations (47) and (48) are closely connected with Ramanujan's work and are discussed by Berndt.[19] Again there appears a tantalizing connection between (48) and T.2 of Table II. Direct proofs of results obtained from the rectangular lattices are even more difficult to furnish since the Θ functions concerned and their derivatives have to be evaluated at finite z. Thus, for example, the lattice $b = 3a$ (T.5, Table II) yields with little labor

$$\pi\Theta_2''\left(\frac{\pi}{6}\middle|\frac{i}{3}\right) + 3\Theta_2\left(\frac{\pi}{6}\middle|\frac{i}{3}\right) = 0, \tag{49a}$$

$$\pi^3\Theta_2^{vi}\left(\frac{\pi}{6}\middle|\frac{i}{3}\right) + 45\pi^2\Theta_2^{iv}\left(\frac{\pi}{6}\middle|\frac{i}{3}\right) - 810\Theta_2\left(\frac{\pi}{6}\middle|\frac{i}{3}\right) = 0, \tag{49b}$$

for $n = 0$ and 1. An attempt to prove (49a) directly following the lines already indicated was successful, but is very long and elaborate. Another method, shown here, makes use of the infinite product representation[13] of Θ_2:

$$\Theta_2(z|\tau) = 2x^{1/4}Q_0(x)\prod_{r=1}^{\infty}(1 + 2x^{2r}\cos 2z + x^{4r}). \tag{50}$$

We also have for all Θ functions the relation[13] $\Theta''(z|\tau) = -4x\,\partial\Theta/\partial x$. Hence, differentiating (50) logarithmically with respect to x and multiplying by $-4x$ we have

$$\frac{\Theta_2''(z|\tau)}{\Theta_2(z|\tau)} = \sum_{r=1}^{\infty}\frac{8rx^{2r}}{1-x^{2r}} - 1 - 16\sum_{r=1}^{\infty}\frac{r\cos 2z x^{2r} + rx^{4r}}{1 + 2x^{2r}\cos 2z + x^{4r}}. \tag{51}$$

Putting $z = \pi/6$ and taking just the last term of (51), by successive rearrangement we have

$$-8\sum_{r=1}^{\infty}\frac{rx^{2r} + 2rx^{4r}}{1 + x^{2r} + x^{4r}} = -8\sum_{r=1}^{\infty}\frac{rx^{2r} + rx^{4r} - 2rx^{6r}}{1 - x^{6r}} = -8\sum_{r=1}^{\infty}\frac{rx^{2r}}{1-x^{2r}} + 24\sum_{r=1}^{\infty}\frac{rx^{6r}}{1-x^{6r}}.$$

Hence

$$\frac{\Theta_2''\left(\frac{\pi}{6}\middle|\tau\right)}{\Theta_2\left(\frac{\pi}{6}\middle|\tau\right)} = -1 + 24\sum_{r=1}^{\infty}\frac{rx^{6r}}{1-x^{6r}}. \tag{52}$$

If we now put $\tau = i/3$ and $x = e^{-\pi/3}$, (52) becomes

$$\frac{\Theta_2''\left(\frac{\pi}{6}\middle|\frac{i}{3}\right)}{\Theta_2\left(\frac{\pi}{6}\middle|\frac{i}{3}\right)} = -1 + 24\sum_{r=1}^{\infty}\frac{r}{e^{2\pi r}-1}.$$

Then, using (47), Eq. (49a) is proved. No attempt has been made at a direct proof of (49b) and it would seem a formidable proposition even if it could be carried out.

To summarize, it seems that each zero of a $C_N^{(a)}(k,q)$ function yields a nontrivial relation between Θ functions and their derivatives. For given s, infinite sets of relations are found. In each case it is comparatively simple to write down a Θ-function relation from the given zero, whereas direct proofs of these relations from definitions of Θ functions are laborious. It is remarkable that these relations are so easily formulated with the aid of the kq representation.

VI. DISCUSSION

The interest in von Neumann lattices stems in large measure from the desire to study at a fundamental level the relationship between classical quantities and their quantum mechanical counterparts. For this two concepts are basic: phase space and the elementary area h. From there it is natural to give particular attention to quantum states whose properties can also be given an approximate classical interpretation in phase-space terms: coherent states and higher harmonic-oscillator states as well. The question of the completeness of lattices of these states, one per Planck cell, is

essentially a density-of-states question and it is not surprising that it can be verified that they are indeed complete.[4–6] What is surprising is that they are overcomplete, albeit by a finite number only. This paper carries further the investigation of overcompleteness, and of its strong relationship with the zeros of the wave function in the kq representation. The relationship is mathematically well determined but remains physically mysterious. The paper also announces the discovery of new zeros for the higher harmonic oscillator states in the case of certain rational relationships between the sides of the unit cell of the von Neumann lattice and discusses some of the extraordinary wealth of mathematical relations which follow from the existence of these zeros, especially those among theta functions. Other such relations, involving Laguerre polynomials, will also automatically follow from formulas derived in Ref. 9 (Sec. IV). Finally it should be mentioned that lattices of coherent states have interesting applications in communications theory[8,14] and the theory of a Bloch electron in a magnetic field.[20]

In connection with the last example there is a striking feature of the rational von Neumann lattices. Equation (16) reminds one very much of the definition of rational magnetic fields in the problem of the dynamics of a Bloch electron.[21,22] In the Bloch electron system $\lambda^2 = \hbar c/eH$ and (16) then becomes

$$\frac{Ha^2}{(hc/e)} = \frac{l}{s}.$$

This is just the rationality condition for a Bloch electron in a magnetic field.

ACKNOWLEDGMENTS

One of us (I.J.Z.) wishes to thank the Technion, Haifa, for its hospitality and the Royal Society for its support while this work was being done.

APPENDIX

Let $C^{(a)}(k,q)$ be a function defined on the constant a and, correspondingly, $C^{(sa)}(k,q)$ for the constant sa. We shall be looking for a transformation connecting these two functions. We use the definitions[10]

$$\langle x|kq\rangle^{(a)} = \left(\frac{a}{2\pi}\right)^{1/2} \sum_n \exp(ikan)\delta(x - q - na), \quad \text{(A1)}$$

$$\langle x|kq\rangle^{(sa)} = \left(\frac{sa}{2\pi}\right)^{1/2} \sum_n \exp(iksan)\delta(x - q - nsa). \quad \text{(A2)}$$

The transformation matrix is ${}^{(sa)}\langle k'q'|kq\rangle^{(a)}$. The latter can be found from (A1) and (A2)

$${}^{(sa)}\langle k'q'|kq\rangle^{(a)} = \frac{1}{(s)^{1/2}} \sum_{n=-\infty}^{\infty} \delta\left(k - k' - n\frac{2\pi}{sa}\right) \times \sum_{\substack{l=-\infty \\ m=0,\ldots,s-1}}^{\infty} \exp(ikam + ik'sal) \times \delta(q' - q - ma - lsa). \quad \text{(A3)}$$

By using the matrix (A3) we find, for $0 \leqslant q' < a$; $0 \leqslant k' < 2\pi/sa$,

$$C^{(sa)}(k',q') = \int {}^{(sa)}\langle k'q'|kq\rangle^{(a)} C^{(a)}(k,q)\, dk\, dq = \frac{1}{(s)^{1/2}} \sum_{r=0}^{s-1} C^{(a)}\left(k' + r\frac{2\pi}{sa}, q'\right) \quad \text{(A4)}$$

integrated over the kq cell for the constant sa. Correspondingly, for the same range of k' and q'

$$C^{(a)}(k',q') = \int {}^{(a)}\langle k'q'|kq\rangle^{(sa)} C^{(sa)}(k,q)dkdq = \frac{1}{(s)^{1/2}} \sum_{r=0}^{s-1} C^{(sa)}(k',q' - ra)\exp(ik'ar). \quad \text{(A5)}$$

[1] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U. P., Princeton, 1955).
[2] E. Schrödinger, Naturwissenschaften **14**, 664 (1926).
[3] R. J. Glauber, Phys. Rev. **131**, 2766 (1963).
[4] A. M. Perelomov, Teor. Mat. Fiz. **6**, 213 (1971) [Theor. Math. Phys. **6**, 156 (1971)].
[5] V. Bargmann, P. Butera, L. Girardello, and J. R. Klauder, Rep. Math. Phys. **2**, 221 (1971).
[6] H. Bacry, A. Grossmann, and J. Zak, Phys. Rev. B **12**, 1118 (1975).
[7] M. Boon and J. Zak, Phys. Rev. B **18**, 6744 (1978).
[8] R. Balian, "A strong uncertainty principle in signal theory or in quantum mechanics" (Preprint).
[9] M. Boon and J. Zak, J. Math. Phys. **22**, 1090 (1981). An independent proof was given by A. Janssen, J. Math. Phys. **23**, 720 (1982).
[10] J. Zak, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1972), Vol. 27.
[11] Pierre Cartier, Am. Math. Soc. Symp. Pure Math. **9**, 361 (1966).
[12] J. L. Powell and B. Crasemann, *Quantum Mechanics* (Addison-Wesley, Reading, MA, 1965).
[13] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U. P., Cambridge, England, 1950).
[14] Conditions for completeness in case f is a distribution have been investigated by A. Janssen, J. Math. Anal. Appl. **80**, 156 (1981) and Ref. 9 above.
[15] A. C. Aitken, *Determinants and Matrices* (Oliver and Boyd, Edinburgh, 1951), Chap. VI.
[16] J. Tannery and J. Molk, *Elements de la Théorie des Fonctions Elliptiques*, 2nd. Ed. (Chelsea, New York, 1972), Vol. 2, Sec. 258.
[17] B. C. Berndt, J. Reine Angew. Math. **303/304**, 147–164 (1978).
[18] A. Cauchy, Oeuvres, Serie II, Vol. VII (Paris, 1889).
[19] B. C. Berndt, *Theory and Applications of Special Functions* (Academic, New York, 1975), pp. 143–189.
[20] M. H. Boon, Helv. Phys. Acta **48**, 551 (1975); *Lecture Notes in Physics* (Springer, Berlin, 1976), Vol. 50, p. 282.
[21] E. Brown, Phys. Rev. **133**, A1038 (1964).
[22] J. Zak, Phys. Rev. **134**, A1602, A1607 (1964).

A geometrical theory of energy trajectories in quantum mechanics

Richard L. Hall

Department of Mathematics, Concordia University, 1455 de Maisonneuve Blvd. W., Montreal, Quebec, Canada

(Received 20 July 1982; accepted for publication 1 October 1982)

Suppose $f(r)$ is an attractive central potential of the form $f(r) = \Sigma_{i=1}^{k} g^{(i)}(f^{(i)}(r))$, where $\{f^{(i)}\}$ is a set of *basis potentials* (powers, log, Hulthén, sech^2) and $\{g^{(i)}\}$ is a set of smooth increasing transformations which, for a given f, are either all convex or all concave. Formulas are derived for bounds on the energy trajectories $E_{nl} = F_{nl}(v)$ of the Hamiltonian $H = -\Delta + vf(r)$, where v is a coupling constant. The transform $\Lambda(f) = F$ is carried out in two steps: $f \to \bar{f} \to F$, where $\bar{f}(s)$ is called the *kinetic potential* of f and is defined by $\bar{f}(s) = \inf\{\psi, f, \psi\}$ subject to $\psi \in \mathscr{D} \subset L^2(R^3)$, where $\mathscr{D}$ is the domain of H, $\|\psi\| = 1$, and $(\psi, -\Delta\psi) = s$. A table is presented of the basis kinetic potentials $\{\bar{f}^{(i)}(s)\}$; the general trajectory bounds $F_*(v)$ are then shown to be given by a Legendre transformation of the form $(s, \bar{f}_*(s)) \to (v, F_*(v))$, where $\bar{f}_*(s) = \Sigma_{i=1}^{k} g^{(i)}(\bar{f}^{(i)}(s))$ and $F_*(v) = \min_{s>0}\{s + v\bar{f}_*(s)\}$. With the aid of this potential construction set (a kind of Schrödinger Lego), ground-state trajectory bounds are derived for a variety of translation-invariant N-boson and N-fermion problems together with some excited-state trajectory bounds in the special case $N = 2$. This article combines into a single simplified and more general theory the earlier "potential envelope method" and the "method for linear combinations of elementary potentials."

PACS numbers: 05.30.Fk, 05.30.Jp, 03.65.Ge

I. INTRODUCTION

We consider Schrödinger Hamiltonians of the form

$$H = -\Delta + vf(r), \quad v > 0, \tag{1.1}$$

where f is a central potential ($r = |r|$) and v is a positive coupling constant. We suppose that f has the form

$$f(r) = \sum_{i=1}^{k} g^{(i)}(f^{(i)}(r)), \tag{1.2}$$

in which the *basis potentials* $\{f^{(i)}\}$ are certain well-known attractive "soluble potentials" [powers $\operatorname{sgn}(p) r^p$, $p \geqslant -1$, $p \neq 0$, $\log(r)$, $\operatorname{sech}^2(r)$, and Hulthén; see Table I in Sec. III], and the transformation functions $\{g^{(i)}\}$ are smooth (C_2), increasing, and either all convex or all concave on the ranges of the corresponding $f^{(i)}$.

Formula (1.2) therefore generates a large variety of smooth increasing potentials. The purpose of the present article is to provide *recipes* for upper and lower bounds on the eigenvalues of H defined in some suitable domain $\mathscr{D} \subset L^2(R^3)$.

Fortunately, the functional analysis of nonrelativistic quantum mechanics is now readily accessible in text books (for example, Prugovečki,[1] Reed and Simon,[2] and Thirring[3]); we refer the reader particularly to Chap. XIII of Ref. 2. Let us assume that, by appropriate control of the transformation functions $\{g^{(i)}\}$, the potentials we generate by Eq. (1.2) also satisfy

(i) $f'(r) > 0, \quad r > 0,$

(ii) $\lim_{r \downarrow 0} |r^2 f(r)| = 0,$

(iii) $\lim_{r \uparrow \infty} \{f(r)\exp(-\alpha r)\} = 0 \quad \text{for some } \alpha > 0. \tag{1.3}$

With these restrictions and v sufficiently large, the bottom E of the spectrum of $H = -\Delta + vf(r)$ will always be a nondegenerate discrete eigenvalue; assumption (iii) conveniently allows us to use exponential and Gaussian trial functions. These assumptions can, of course, be weakened and, moreover, since we shall only use variational arguments, our results in any case apply to the bottom of the spectrum of H, whether or not it it is an eigenvalue. We shall discuss higher eigenvalues in Sec. VI.

We call the curve which describes how E depends on v the *energy trajectory* of f and write this

$$F = \Lambda(f) \quad \text{and} \quad E = F(v). \tag{1.4}$$

We know the exact "component" trajectories $F^{(i)} = \Lambda(f^{(i)})$, and we wish to relate $F = \Lambda(f)$ to these for a potential f generated by Eq. (1.2). In earlier articles,[4-6] which we shall henceforth call Papers I, II, and III, we dealt with convex (or concave) transformations of soluble potentials (the method of potential envelopes[4,5]) and with linear combinations of basis potentials.[6] In the present paper we combine these results into a single theory, which is at once simpler and more general. The principal new idea can be thought of as a factoring of the Λ transform so that the trajectory F is reached from the potential f in two stages: $f \to \bar{f} \to F$. The new curves $\bar{f}$ are called *kinetic potentials* (short for "minimum mean isokinetic potentials") which, for the bottom of the spectrum, are defined by

$$\bar{f}(s) = \inf_{\substack{\psi \in \mathscr{D} \\ \|\psi\| = 1 \\ (\psi, -\Delta\psi) = s}} (\psi, f\psi). \tag{1.5}$$

In terms of the kinetic potential we have, for the second stage of minimization,

$$F(v) = \min_{s>0}(s + v\bar{f}(s)), \tag{1.6}$$

or

$$v^{-1} = -\bar{f}'(s)$$

 0022-2488/83/020324-12$02.50

and

$$F(v) = s + v\bar{f}(s). \tag{1.7}$$

Equations (1.7) are parametric equations for the energy trajectory in terms of the parameter $s > 0$ which is equal to the mean kinetic energy. We note that the θ function that was introduced in Paper III is related to the corresponding kinetic potential by the equation $\theta(s) = -\{\bar{f}'(s)\}^{-1}$; we shall not need to use these θ functions in the present article, which is self-contained.

A quantity like $\bar{f}$ will be useful generally when one is discussing a sum of operators having essentially common domains. We shall see later that $\bar{f}(s) = F'(v)$ in the present problem; in a more general situation, objects like $\bar{f}(s)$ would be related to the partial derivatives of the eigenvalue with respect to the coefficients of the corresponding operator terms. However, in this paper we are mainly concerned with describing a *constructive* approximation theory for a specific class of Schrödinger operators. We shall first find the kinetic potentials $\{\bar{f}^{(i)}\}$ and $\{\bar{f}_{\phi}^{(i)}\}$ (where ϕ is the *shape* of a trial function) corresponding to the basis potentials $\{f^{(i)}\}$. Our approximation F_A to the energy trajectory $F = \Lambda(\Sigma_{i=1}^{k} g^{(i)}(f^{(i)}))$ is then expressed in terms of the $\{\bar{f}^{(i)}\}$ or the $\{\bar{f}_{\phi}^{(i)}\}$ by general parametric equations having the form

$$\begin{aligned} F_A(v) &= s + v\sum_{i=1}^{k} g^{(i)}(\bar{f}^{(i)}(s)), \\ v^{-1} &= -\frac{d}{ds}\sum_{i=1}^{k} g^{(i)}(\bar{f}^{(i)}(s)). \end{aligned} \tag{1.8}$$

If the $\{g^{(i)}\}$ are all *convex* and the $\{\bar{f}^{(i)}\}$ are all exact, then F_A is a *lower* bound to F; if the $\{g_i\}$ are all *concave* and $\{\bar{f}^{(i)}\} = \{\bar{f}_{\phi}^{(i)}\}$, $i = 1,2,\ldots,k$, then F_A is an *upper* bound to F (see Theorem 4). The table of $\{f^{(i)}\}$ and $\{\bar{f}^{(i)}\}$ and the general recipe (1.8) provide us with a potential construction set for which the in-house name has become "Schrödinger Lego."

The main motivation for this work remains our interest in the N-body problem whose energy is intimately related to the energy trajectories of the corresponding two-body system[4,7–9]; we discuss N-boson system in Sec. V, N-fermion systems in Sec. VII, and we present some new examples in Sec. VIII.

Even for the two-body system itself with, say, a linear combination of powers and the log potential, it is very useful to have a *recipe* [e.g., Sec. IV, Eq. (4.10)] for the energy as a function of the potential parameters; no computer output can as yet compete with this is terms of visible information content. In Sec. VI we extend some of these results to higher eigenvalues. The purpose of our geometrical theory is to provide a construction set for potentials whose energy trajectories are then automatically estimated by the bounds (1.8), very often to within a few percent, as functions of the potential parameters.

II. TRAJECTORY FUNCTIONS AND KINETIC POTENTIALS

In this section we present some fundamental convexity, monotonicity, scaling, and ordering properties of the trajectory functions and kinetic potentials. Higher eigenvalues will be discussed in Sec. VI.

An energy trajectory $F(v)$ tells us how the energy E depends on the coupling constant v. However, our parametric equations (1.8) for the bounds on the energy trajectories (yet to be established) suggest that it may be more natural to work with the quantities E/v and $1/v$. We therefore define for $v > 0$

$$\begin{aligned} &u = 1/v, \\ &\hat{G} = H/v = -u\Delta + f(r), \\ &G(u) = F(v)/v. \end{aligned} \tag{2.1}$$

Consequently,

$$G(u) = \inf_{\substack{\psi\in\mathscr{D} \\ \|\psi\|=1}} (\psi,\hat{G}\psi) = uK(u) + P(u), \tag{2.2}$$

where $K(u) = (\psi, -\Delta,\psi) > 0$ and $P(u) = (\psi, f, \psi)$ are, for a given value of u, the expectation values we get *after* the minimization. Meanwhile, with the kinetic potential $\bar{f}(s)$ given by

$$\bar{f}(s) = \inf_{\substack{\psi\in\mathscr{D} \\ \|\psi\|=1 \\ (\psi,-\Delta\psi)=s}} (\psi, f\psi), \tag{2.3}$$

we have

$$G(u) = \min_{s>0}(us + \bar{f}(s)). \tag{2.4}$$

The properties we need to establish may conveniently be listed together; they are:

Theorem 2:

(a) $G(u)$ is monotone increasing and concave,

(b) $F(v)$ is concave,

(c) $\bar{f}(s)$ is monotone decreasing and convex,

(d) $f(r) = A + Bf_1(r/b)$ with $B > 0$ and $b > 0$ implies $\bar{f}(s) = A + B\bar{f}_1(b^2 s)$,

(e) $f_1 < f_2 \Rightarrow \bar{f}_1 < \bar{f}_2 \Rightarrow G_1 < G_2 \Rightarrow F_1 < F_2$.

Proof: We use variational arguments based on the premise that for u sufficiently small the bottom of the spectrum of $\hat{G}$ is the discrete eigenvalue. From (2.2) we have

$$G(u) = uK(u) + P(u) < uK(u^*) + P(u^*), \quad u^* \neq u.$$

Hence, $G(u) < G(u^*) + (u - u^*)K(u^*)$, $u^* \neq u$. Since $K \geqslant 0$, Theorem 2(a) follows (See Feller,[10] p. 153). Since $G(u) = uF(1/u)$, Theorem 2(b) immediately follows from 2(a). From (2.4) we have for a *given* u (sufficiently small)

$$G(u) = \min_{t>0}(ut + \bar{f}(t)) = us + \bar{f}(s) \leqslant us^* + \bar{f}(s^*), \quad s^* \neq s.$$

Hence $-\bar{f}(s) \geqslant -\bar{f}(s^*) + (s - s^*)u$, $s^* \neq s$. Since $u > 0$, 2(c) follows. Theorem 2(d) is derived with the aid of a change of variables in definition (2.3) and the fact that $b^2\Delta_r = \Delta_{r/b}$. The ordering Theorem 2(e) follows directly from (2.3) and (2.4) by simple variational arguments. Of course, it is understood that G_1 is compared to G_2 only over the common domain, and similarly for F_1 and F_2.

A theorem which is essentially Theorem 2(b) may be found in the book by Thirring (Ref. 3, p. 153). The consistency of the inequalities in the ordering Theorem 2(e) explains why we have chosen to work with $\bar{f}$ rather than with $-\bar{f}$: The latter alternative, however, would have allowed us di-

TABLE I. Some basis potentials and their ground-state kinetic potentials. The basis potentials have been selected according to their usefulness and the simplicity of the corresponding kinetic potentials defined by Eqs. (2.3) or (3.8). The subscripts g and e correspond, respectively, to Gaussian and exponential trial functions ϕ in Eq. (3.8). The coefficients for the power-law and log potentials are given in Table II.

Potential	$f(r)$	$\bar{f}(s)$	$\bar{f}_g(s)$	$\bar{f}_e(s)$
Power	$\mathrm{sgn}(p)r^p$, $p\geqslant -1, p\neq 0$	$2G^{(p)}(ps^{p/2})^{-1}$	$2G_g^{(p)}(ps^{p/2})^{-1}$	$2G_e^{(p)}(ps^{p/2})^{-1}$
Log	$\ln r$	$\frac{1}{2}\ln(v_{10}/2es)$	$\frac{1}{2}\ln(v_g/2es)$	$\frac{1}{2}\ln(v_e/2es)$
Hulthén	$-(e^r-1)^{-1}$	$-\frac{1}{2}[(1+4s)^{1/2}-1]$		
sech^2	$-\mathrm{sech}^2(r)$	$-2\{[(s+2)^2+s]^{1/2}-(s+2)\}$		
Exponential	$-e^{-r}$			$-8s^{3/2}(1+2s^{1/2})^{-3}$ $s>1/16\Rightarrow E<0$
Gauss	$-e^{-r^2}$		$-(1+3/2s)^{-3/2}$ $S>3/4\Rightarrow E<0$	
Yukawa	$-e^{-r}/r$			$-4s^{3/2}(1+2s^{1/2})^{-2}$ $s>1/4\Rightarrow E<0$

rectly to use the notion of the "Legendre transformation" (e.g., Ref. 11, p. 71), which underlies the relation between $-\bar{f}$ and G.

III. THE BASIS POTENTIALS AND THEIR ASSOCIATED KINETIC POTENTIALS

There are two types of kinetic potential which we shall need: exact kinetic potentials and kinetic potentials labelled by the *shape* ϕ of a trial function. We consider only the lowest eigenvalue and discuss the exact case first.

Since the basis potentials, by hypothesis, yield *soluble* eigenproblems, we do not use the general definition (2.3) to find $\bar{f}$ but rather the following procedure. From (2.4) we have

$$G(u) = us + \bar{f}(s), \qquad u = -\bar{f}'(s). \tag{3.1}$$

It follows that

$$s = G'(u) = F(v) - vF'(v)$$

and

$$\bar{f}(s) = G(u) - uG'(u) = F'(v). \tag{3.2}$$

In practice, therefore, we solve Eqs. (3.2) for $\bar{f}(s)$. In Paper III we used the term "elementary potential" to label the situations where this is possible. If it is hard to solve for $\bar{f}(s)$ explicitly, we can instead use either u or v as the trajectory parameter. Since the basis potentials have been discussed already in Papers I and III, we simply list the results here in Tables I and II. It is interesting that all the exact kinetic potentials are defined for all $s>0$ so that the question of the various domains of the $G(u)$ functions is automatically looked after by the uniform constraint $s>0$.

Now we consider real central trial functions $\phi(r)$ and suppose

$$\int \phi^2(r)\, d^3r = 1 \tag{3.3}$$

and define

$$K(\phi) = \int \phi(r)(-\Delta)\phi(r)\, d^3r, \tag{3.4}$$

where the integrations are over all of R^3. Then, for a given ϕ, we consider the domain $\mathscr{D}_\phi \subset \mathscr{D}$ given by

$$\mathscr{D}_\phi = \{\psi \mid \psi(r) = C\phi(r/\sigma),\ \sigma>0,\ C\in R\}. \tag{3.5}$$

It follows that

$$(\psi, -\Delta\psi)/\|\psi\|^2 = K(\phi)/\sigma^2, \quad \psi\in\mathscr{D}_\phi. \tag{3.6}$$

Consequently, we *define*

$$\bar{f}_\phi(s) = (\psi, f\psi), \quad \psi\in\mathscr{D}_\phi, \quad \|\psi\| = 1, \quad (\psi, -\Delta\psi) = s \tag{3.7}$$

TABLE II. Coefficients for power-law and log kinetic potentials. These coefficients are required for the kinetic potentials of Table I. The numerical values have been rounded up or down so as to preserve the validity of the trajectory bounds. For the power-law potentials, if $E^{(p)} = \inf\{\psi, (-\Delta + \mathrm{sgn}(p)r^p)\psi\}$, $\|\psi\| = 1$, then $G^{(p)} = [|p|E^{(p)}|/(p+1)]^{|p+2)/2}$ [see Paper III, Eq. (5.3)]. The value $E^{(4)} = 3.799\,67$ was taken from Hioe and Montroll[17] [Eq. (IV.16)] and is the energy of the first excited state of the problem in one dimension. Similarly, we have obtained $E^{(p)}$ for $p = 6$ and 8 from Hioe, MacMillen, and Montroll[17] [Eqs. (III.7) and (III.8)] by using the correspondence $E^{(p)} = 2^{p/(p+2)}\epsilon_n$ and $n = 1$.

$f(r)$	G	G_g	G_e
$-1/r$	0.5	0.460 659	0.5
r	0.688 041	0.690 988	0.75
r^2	2.25	2.25	3
r^4	16.254 17	16.875	45
r^6	112.11	132.9	
r^8	797.21	1196.1	

$\mathrm{sgn}(p)r^p$ $p\geqslant -1$ $p\neq 0$	$G_g^{(p)} = \lvert p\rvert 3^{p/2}[\Gamma((p+3)/2)/\Gamma(3/2)]2^{-(p+2)/2}$ $G_e^{(p)} = \lvert p\rvert\Gamma(p+3)2^{-(p+2)}$
$\ln r$	$v_{10} = 8.07$ [$n=1$ and $l=0$; see Eq. (8.24)] $v_g = \frac{3}{4}\exp(3-\gamma) = 8.457\,92$ $v_e = \frac{1}{2}\exp(4-2\gamma) = 8.605\,68$ γ = Euler's constant

and find

$$\bar{f}_\phi(s) = \int \phi^2(\xi) f([K(\phi)/s]^{1/2}\xi)\, d^3\xi. \tag{3.8}$$

Therefore, by the variational principle,

$$G(u) \leqslant G_\phi(u) = \inf_{s>0} (us + \bar{f}_\phi(s)) \tag{3.9}$$

and

$$\bar{f}(s) \leqslant \bar{f}_\phi(s). \tag{3.10}$$

In Tables I and II we list some kinetic potentials labelled by the exponential and Gaussian trial functions:

$$\begin{aligned} &e\text{: } \phi(r) = C_1 e^{-r/2}, \quad K(\phi) = \tfrac{1}{4}, \\ &g\text{: } \phi(r) = C_2 e^{-r^2/4}, \quad K(\phi) = \tfrac{3}{4}, \end{aligned} \tag{3.11}$$

where C_1 and C_2 are normalization constants.

In this way the upper bounds are described in the same framework as the lower bounds: we simply restrict $\mathscr{D}$ to $\mathscr{D}_\phi$, that is to say, we use $\bar{f}_\phi$ instead of $\bar{f}$.

Of course, it is straightforward to find kinetic potentials for problems in one dimension (see Papers I and III) or to project the three-dimensional problem into an angular momentum subspace: The corresponding trajectory bounds (1.8) remain essentially the same (see Sec. VI).

IV. THE TRAJECTORY BOUNDS

In this section we establish the trajectory bounds which are a principal result of the paper. These bounds are given in terms of the kinetic potentials which are functions of the kinetic-energy parameter $s > 0$. The kinetic potentials obey some basic inequalities from which the trajectory bounds (Theorem 4) quickly follow. The potentials themselves are smooth increasing potentials which are not too singular at $r = 0$ [see (1.3)].

Lemma 4:

(a) $(\overline{f^{(1)} + f^{(2)}})(s) \geqslant \bar{f}^{(1)}(s) + \bar{f}^{(2)}(s)$.

(b) $f(r) = \sum_{i=1}^{k} \alpha_i \bar{f}^{(i)}(r), \quad \alpha_i > 0,$ implies

$$\bar{f}(s) \geqslant \sum_{i=1}^{k} \alpha_i \bar{f}^{(i)}(s).$$

(c) $f(r) = \int_0^\infty \alpha(\nu) f^{(\nu)}(r)\, d\nu, \quad \alpha(\nu) \geqslant 0,$ implies

$$\bar{f}(s) \geqslant \int_0^\infty \alpha(\nu) \bar{f}^{(\nu)}(s)\, d\nu.$$

(d) Suppose g is increasing; then g convex implies $\overline{g(f)} \geqslant g(\bar{f})$, and g concave implies $\overline{g(f)} \leqslant g(\bar{f})$.

Proof of Lemma 4: Lemma 4(a) follows directly from the definition of the kinetic potential for if $f(r) = f^{(1)}(r) + f^{(2)}(r)$, then

$$\bar{f}(s) = \inf_{\substack{\psi \in \mathscr{D} \\ \|\psi\| = 1 \\ (\psi, -\Delta\psi) = s}} (\psi, (f^{(1)} + f^{(2)})\psi) \geqslant \bar{f}^{(1)}(s) + \bar{f}^{(2)}(s)$$

and this immediately implies Lemmas 4(b) and 4(c). Lemma 4(d) is a consequence of Jensen's inequality (for example, Ref. 10, p. 153). In the case that the kinetic potentials are labelled by a trial function ϕ, the inequalities in 4(a), (b), (c) become equalities because in definition (3.7) ψ is determined up to a factor of modulus 1; in 4(d) the corresponding inequalities follow from Jensen's inequality provided the "bar" is understood in terms of definition (3.8) with the same shape ϕ of the trial wave function on both sides. The integral mixtures which we have accommodated by Lemma 4(c) allow examples like

$$f(r) = -\int_0^\infty \alpha(\nu)(e^{\nu r} - 1)^{-1}\, d\nu,$$

which represents a mixture of Hulthén potentials with different ranges. Of course, the weight function $\alpha(\nu)$ will have to be controlled so that $f(r)$ meets condition (1.3).

We now proceed to our main result. For conciseness we state the result in terms of *sums* of potentials; the corresponding result for integral mixtures is, of course, exactly similar. Positive weights like the $\{\alpha_i\}$ in Lemma 4(b) are now omitted because they are allowed for by the scaling Theorem 2(d).

Theorem 4: Suppose the functions $g^{(i)}$ are all increasing and

$$f(r) = \sum_{i=1}^{k} g^{(i)}(f^{(i)}(r)), \tag{4.1}$$

$$\bar{f}_A(s) = \sum_{i=1}^{k} g^{(i)}(\bar{f}^{(i)}(s)), \tag{4.2}$$

$$G(u) = \min_{s>0} (us + \bar{f}(s)), \tag{4.3}$$

and

$$G_A(u) = \min_{s>0} (us + \bar{f}_A(s)), \tag{4.4}$$

then

(a) If the $\{\bar{f}^{(i)}\}$ are all exact and the $\{g^{(i)}\}$ are all convex, $G_A(u) \leqslant G(u)$.

(b) If $\bar{f}_i^{(i)} = \bar{f}_\phi^{(i)}, i = 1,2,\ldots,k$, and the $\{g^{(i)}\}$ are all concave, $G_A(u) \geqslant G_\phi(u) \geqslant G(u)$.

(c) If there is only one term in (4.1) so that $f(r) = g^{(1)}(f^{(1)}(r))$ and if $\bar{f}^{(1)}$ is exact and $g^{(1)}$ is concave, then $G_A(u) \geqslant G(u)$.

Proof of Theorem 4: For 4(a) we apply Lemmas 4(b), (d) to give $\bar{f}_A(s) \leqslant \bar{f}(s)$; the ordering Theorem 2(e) then yields the required result. For the upper bound via the trial function ϕ we have by the definition (3.8) of $\bar{f}_\phi(s)$:

$$G(u) \leqslant \inf_{\substack{\psi \in \mathscr{D}_\phi \\ \|\psi\| = 1}} (\psi, \{-u\Delta + f\}\psi),$$

i.e.,

$$G(u) \leqslant \inf_{s>0} (su + \bar{f}_\phi(s)). \tag{4.5}$$

But by definition (3.7) and Lemma 4(d) we have

$$\bar{f}_\phi(s) \leqslant \bar{f}_A(s) = \sum_{i=1}^{k} g^{(i)}(\bar{f}_\phi^{(i)}(s)). \tag{4.6}$$

Hence

$$G(u) \leqslant \inf_{s>0} (su + \bar{f}_A(s)) = G_A(s). \tag{4.7}$$

This establishes Theorem 4(b). Similarly, Theorem 4(c) immediately follows from Lemma 4(d); this upper bound is distinct from Theorem 4(b) and does not require a trial function ϕ.

Since, for the potentials we are considering, the minima always exist for u sufficiently small and the $\bar{f}_A(s)$ are smooth, we may use the following recipe for the trajectory bounds:

$$G_A(u) = us + \bar{f}_A(s),$$
$$u = -\frac{d}{ds}\bar{f}_A(s), \tag{4.8}$$

where

$$\bar{f}_A(s) = \sum_{i=1}^{k} g^{(i)}(\bar{f}^{(i)}(s)).$$

Hence $(v, G_A(v))$ is a Legendre transformation[11] of $(s, -\bar{f}_A(s))$.

In the special case of linear combinations of powers and the log potential, for example, we find that, for the potential

$$f(r) = \sum_p \alpha_p \operatorname{sgn}(p) r^p + \alpha \ln r, \quad \alpha_p \geqslant 0, \quad \alpha \geqslant 0, \quad -1 \leqslant p \neq 0, \tag{4.9}$$

the trajectory bounds provided by Eq. (4.8) become

$$G_A(u) = \sum_p \alpha_p [(2+p)/p] G_A^{(p)} t^p + \tfrac{1}{2}\alpha \ln(v_A t^2/2)$$
$$u = \sum_p \alpha_p G_A^{(p)} t^{p+2} + \tfrac{1}{2}\alpha t^2, \quad t > 0, \tag{4.10}$$

where we have reparameterized in terms of $t = s^{-1/2}$, and simplified. For a lower bound, $G_A^{(p)} = G^{(p)}$ and $v_A = v_{10}$; for upper bounds, $G_A^{(p)} = G_\phi^{(p)}$ and $v_A = v_\phi$ with $\phi = g$ or e, corresponding to Gaussian or exponential trial functions; the coefficients are listed in Table II.

V. THE N-BOSON PROBLEM

Consider N identical bosons each of mass m interacting via a central pair potential of the form

$$V_{ij} = V_0 f(r_{ij}/a), \tag{5.1}$$

where V_0 is a positive coupling constant, a is a positive range parameter, and f is a potential shape of the type we have been studying in this paper. The Hamiltonian H for the relative motion of this system may be written[4]

$$H = \sum_{1 \leqslant i < j}^{N} \left[\frac{(\mathbf{p}_i - \mathbf{p}_j)^2}{2mN} + V_{ij} \right]. \tag{5.2}$$

In terms of a set $\{\boldsymbol{\rho}_2, \boldsymbol{\rho}_3, \ldots, \boldsymbol{\rho}_N\}$ of Jacobi orthogonal relative coordinates[4] with $\boldsymbol{\rho}_2 = (\mathbf{r}_1 - \mathbf{r}_2)/\sqrt{2}$ we have for expectations with respect to boson functions of these variables

$$\langle H \rangle = \langle \mathscr{H} \rangle, \tag{5.3}$$

where

$$\mathscr{H} = (N-1)\left[-\frac{\hbar^2}{2m}\Delta_{\boldsymbol{\rho}_2} + \frac{NV_0}{2} f\left(\frac{|\boldsymbol{\rho}_2|\sqrt{2}}{a}\right)\right]. \tag{5.4}$$

We now define the dimensionless variables

$$v = NV_0 a^2 m/2\hbar^2 = 1/u,$$
$$E = mE_N a^2/(N-1)\hbar^2,$$
$$\hat{G} = -u\Delta_{\mathbf{r}} + f(r), \quad r = |\mathbf{r}|$$
$$\mathbf{r} = \boldsymbol{\rho}_2\sqrt{2}/a, \tag{5.5}$$

where E_N is the lowest energy of the N-boson problem. It then follows (see Paper I and the references therein) that

$$G(u) \leqslant uE = \frac{E_N}{V_0}\binom{N}{2}^{-1} \leqslant G_g(u), \tag{5.6}$$

where $G(u)$ is the lowest eigenvalue of $\hat{G}$ and $G_g(u)$ is the minimum of $\langle \hat{G} \rangle$ with respect to Gaussian trial functions $g = \phi$ in r; it is known that $G(u) = G_g(u)$ iff $f(r) = kr^2$ (see Ref. 12).

Equation (5.6) makes the principal results of this paper much more interesting and explains why we have worked to obtain bounds on $G(u)$ and $G_g(u)$ having the same mathematical *form*: we are interested in general recipes for estimating the binding energy per pair interaction of the N-boson problem.

The special case where

$$f(r) = y(h(r)), \ y \text{ concave}, \tag{5.7}$$

yields an upper bound to $\bar{f}_g(s)$ via the inequality

$$\bar{f}_g(s) \leqslant y(\bar{h}_g(s)), \tag{5.8}$$

where the *suffix* g indicates the Gaussian trial function $\phi = g$ in Eq. (3.8); Eq. (5.8) follows from Theorem 4(b). Thus we have the useful special case [see (3.8) and (3.11)]

$$f(r) = y(r^2) \Rightarrow \bar{f}(s) \leqslant \bar{f}_g(s) \leqslant y(9/4s) = f(3/2s^{1/2}). \tag{5.9}$$

That is to say, the approximation to $\bar{f}$ is given *directly* in terms of f itself.

VI. HIGHER EIGENVALUES

We now return to the one (or two)-body problem. If we restrict the problem to an angular momentum or symmetry subspace, then the theory of Sec. IV immediately applies to the bottom of the spectrum of H in this subspace. However, we do not yet have a satisfactory theory for the higher eigenvalues *within* such a subspace when the potential is represented as a *sum* of basis potentials. Therefore, in this section we restrict the discussion to central potentials of the form

$$f(r) = g(h(r)), \tag{6.1}$$

where $h(r)$ is a basis potential, g is increasing and either convex or concave on the range of h, and $f(r)$ satisfies the restrictions (1.3). We suppose in addition that for a sufficiently large coupling constant v the operator $H = -\Delta + vf(r)$ has more than one discrete eigenvalue below the essential spectrum: general conditions sufficient to guarantee this may be found in Reed and Simon[2]; for example, if we add to (1.3) the condition $f(r) \uparrow \infty$ with r, then the entire spectrum is discrete. In place of the domain $\mathscr{D}$ we use the subset $\mathscr{D}_l \subset \mathscr{D}$ defined by the projector onto angular momentum states (eigenfunctions of $\hat{L}^2$) corresponding to the spherical harmonic $Y_l^0(\theta, \phi)$, $l = 0,1,2,\cdots$. We are interested in the eigenvalues E_{ln} of H in $\mathscr{D}_l$, where n is the radial quantum number and

the eigenvalues are ordered according to $E_{ln'} \geqslant E_{ln}$, $n' > n$, $n = 1,2,3,\cdots$. Each of the eigenvalues so labelled will have a degeneracy of exactly $(2l + 1)$. In Paper II we mentioned briefly the problem of higher eigenvalues and indicated the proof of an approximation method based on the "potential envelope" concept. We now state and prove this result using the integral inequality approach characteristic of the present article.

Each eigenvalue of H in $\mathscr{D}_l$ will have an energy trajectory F_{ln} and a corresponding kinetic potential $\bar{f}_{ln}(s)$. The variational definition of $\bar{f}_{ln}(s)$ is a little cumbersome but reads, for fixed l and n, as follows:

$$\bar{f}_{ln}(s) = \inf_{\mathscr{D}_l^{(n)}} \sup_{\substack{\psi \in \mathscr{D}_l^n \\ \|\psi\| = 1 \\ (\psi, -\Delta\psi) = s}} (\psi, f\psi), \tag{6.2}$$

where $\mathscr{D}_l^{(n)}$ is a *finite* n-dimensional *subspace* of $L^2(R^3)$ and $\mathscr{D}_l^{(n)} \subset \mathscr{D}_l$. With this definition we prove

Theorem 6: Suppose that g is monotone increasing. Then

$$\begin{aligned} &\text{(a)} \quad g \text{ convex} \Rightarrow \bar{f}_{ln}(s) \geqslant g(\bar{h}_{ln}(s)), \\ &\text{(b)} \quad g \text{ concave} \Rightarrow \bar{f}_{ln}(s) \leqslant g(\bar{h}_{ln}(s)). \end{aligned} \tag{6.3}$$

Proof: Suppose that g is convex. Then by Jensen's inequality, we have

$$(\psi, g(h)\psi) \geqslant g((\psi, h\psi)). \tag{6.4}$$

Let $C(nls)$ denote all three conditions $\{\psi \in \mathscr{D}_l^{(n)}, \|\psi\| = 1, (\psi, -\Delta\psi) = s\}$ for a given $\mathscr{D}_l^{(n)}$; then by (6.4),

$$\sup_{C(nls)} (\psi, g(h)\psi) \geqslant \sup_{C(nls)} g((\psi, h\psi)). \tag{6.5}$$

Since g is monotone increasing, we have

$$\sup_{C(nls)} g((\psi, h\psi)) = g\left(\sup_{C(nls)} (\psi, h\psi)\right) \geqslant g(\bar{h}_{nl}(s)).$$

Consequently, (6.5) becomes

$$\sup_{C(nls)} (\psi, g(h)\psi) \geqslant g(\bar{h}_{nl}(s)). \tag{6.6}$$

If we now use the fact that the right-hand side of (6.6) is independent of the particular subspace $\mathscr{D}_l^{(n)}$ of fixed dimension n, it follows that

$$\inf_{\mathscr{D}_l^{(n)}} \sup_{C(nls)} (\psi, g(h)\psi) \geqslant g(\bar{h}_{nl}(s)),$$

which result establishes Theorem 6(a). We now suppose that g is concave and obtain, again by Jensen's inequality and the monotonicity of g,

$$\sup_{C(nls)} (\psi, g(h)\psi) \leqslant g\left(\sup_{C(nls)} (\psi, h\psi)\right). \tag{6.7}$$

Now

$$\inf_{\mathscr{D}_l^{(n)}} g\left(\sup_{C(nls)} (\psi, h\psi)\right) = g\left(\inf_{\mathscr{D}_l^{(n)}} \sup_{C(nls)} (\psi, h\psi)\right)$$

because g is monotone increasing. Consequently, by applying inf over $\mathscr{D}_l^{(n)}$ to both sides of (6.7), we establish Theorem 6(b).

In order to make use of Theorem 6, we need to have some exact higher kinetic potentials $\bar{h}_{nl}(s)$. Just as for the ground state, we derive $\bar{h}_{nl}(s)$ from the known $G_{nl}(u)$ by using Eqs. (3.2). Rather than giving an extended table of higher kinetic potentials, we shall illustrate the theory by giving just two families, those corresponding to the Hydrogenic atom and the harmonic oscillator; other such families may be found by using standard handbooks like Flügge.[13] We have the following kinetic potentials in which the angular-momentum quantum number $l = 0,1,2,\cdots$ and the radial quantum number $n = 1,2,3,\cdots$:

$$\begin{aligned} &h(r) = -1/r, \\ &\bar{h}_{nl}(s) = -s^{1/2}/(n + l) \end{aligned} \tag{6.8}$$

and

$$\begin{aligned} &h(r) = r^2, \\ &\bar{h}_{nl}(s) = (2n + l - \tfrac{1}{2})^2/s. \end{aligned} \tag{6.9}$$

As a consistency check we can immediately analyze the harmonic oscillator in terms of the Coulomb potential and obtain

$$\begin{aligned} &h(r) = -1/r, \quad \bar{h}_{nl}(s) = -s^{1/2}/(n + l), \\ &\text{and} \\ &f(r) = r^2 = g(h(r)) = (h(r))^{-2}. \end{aligned} \tag{6.10}$$

Now g in Eq. (6.10) is a convex increasing function of $h(r) = -1/r < 0$ so that, by Theorem 6, we have

$$\bar{f}_{nl}(s) \geqslant g(\bar{h}_{nl}(s)) = (n + l)^2/s, \tag{6.11}$$

whereas $\bar{f}_{nl}(s) = (2n + l - \frac{1}{2})^2/s$ by Eq. (6.9). This rather unlikely approximation is surprisingly good for large l. In terms of the exact eigenvalues $F_{nl}(v)$ of $H = -\Delta + vr^2$, we have by using (6.11) in the general trajectory equations (4.8):

$$2(2n + l - \tfrac{1}{2})v^{1/2} = F_{nl}(v) > 2(n + l)v^{1/2}. \tag{6.12}$$

We shall give more interesting examples of Theorem 6 in Sec. VIII. Our main motivation for the result, however, is the N-fermion problem, which we discuss in the following section.

VII. THE N-FERMION PROBLEM

We consider a system of N identical fermions interacting via central pair potentials of the form

$$V_{ij} = V_0 f(r_{ij}/a), \quad V_0 > 0, \ a > 0, \tag{7.1}$$

where the potential shape f satisfies (1.3). The lowest energy E_N of this system is related[7-9] to the energy trajectories of a two-body problem with Hamiltonian

$$\mathscr{H} = \left\{ -\frac{\hbar^2}{2m\lambda} \Delta_{\rho_2} + \frac{N}{2} V_0 f\left(\frac{\sqrt{2}|\rho_2|}{a}\right)\right\}, \tag{7.2}$$

where $\{\rho_2, \rho_3, \ldots, \rho_N\}$ is a set of $(N - 1)$ relative coordinates with "coefficient of orthogonality"[7-9] $\lambda \geqslant 1$, $\rho_2 = (\mathbf{r}_1 - \mathbf{r}_2)/\sqrt{2}$, and $\rho_3 = P_{23}\rho_2$, where P_{23} exchanges $\mathbf{r}_2$ and $\mathbf{r}_3$. We note that $\mathscr{H}$ differs from the corresponding two-body Hamiltonian used for N-boson systems in Sec. V by the presence of λ and the omission of the overall factor of $(N - 1)$. We now

define the dimensionless quantities:

$$v = N V_0 a^2 \lambda m/2\hbar^2 = 1/u,$$
$$E = E_N a^2 \lambda m/\hbar^2,$$
$$\hat{G} = -u\Delta_{\mathbf{r}} + f(r), \tag{7.3}$$

and

$$\mathbf{r} = \rho_2\sqrt{2}/a, \quad r = |\mathbf{r}|.$$

If $G_\nu(u)$ are the trajectories of $\hat{G}$ with $G_\nu(u) \leqslant G_{\nu'}(u)$, $\nu' > \nu$, (we do not now restrict the problem to an angular momentum subspace and the quantum number ν enumerates *all* linearly independent states), then the theorem of Ref. 7 gives the following lower bound to E_N:

$$\frac{2E_N}{V_0 N} = Eu \geqslant \sum_{\nu=1}^{N-1} G_\nu(u). \tag{7.4}$$

If, for a given u, only $k < (N-1)$ discrete eigenvalues exist, then the sum in (7.4) runs only up to k. We have optimized this result[7] with respect to the allowed class of relative coordinates and this yields $\lambda = \frac{4}{3}$. The result has been extended to higher N-body states[8,9] (using different relative coordinates and a different λ), but here one must be very careful about the Efimov effect[14] and other possible complexities and surprises in the N-body energy spectrum[2]: In the present article we restrict our considerations only to the bottom of the N-body spectrum.

Since we have approximations for the kinetic potentials $\bar{f}_\nu(s)$, we now try to express the bound (7.4) in these terms. We suppose that $f(r)$ is given by

$$f(r) = g(h(r)), \quad g \text{ convex increasing}. \tag{7.5}$$

We then use Theorem 6(a) to generate lower bounds G^{L}_ν to G_ν via the relations

$$G^{\mathrm{L}}_\nu(u) = su + g(\bar{h}_\nu(s)),$$
$$u = -\frac{d}{ds} g(\bar{h}_\nu(s)). \tag{7.6}$$

Since $g' > 0$, $g'' > 0$, and $\bar{h}_\nu''(s) > 0$ [Theorem 2(c)], it follows that the expression for u in (7.6) is a monotone decreasing function of s which in principle can be inverted. That is to say, for each g and h_ν we can find Q_ν such that

$$u = -\frac{d}{ds} g(\bar{h}_\nu(s)) \Leftrightarrow s = Q_\nu(u). \tag{7.7}$$

Hence the lower trajectory bound for the N-fermion system becomes

$$\frac{2E_N}{V_0 N} = Eu \geqslant \sum_{\nu=1}^{N-1} [uQ_\nu(u) + g(\bar{h}_\nu(Q_\nu(u)))]. \tag{7.8}$$

A class of examples for which the details can easily be carried out is power-law potentials. Suppose we consider the basis potential

$$h(r) = \operatorname{sgn}(p) r^p, \quad p \geqslant -1, \ p \neq 0$$

and consider

$$f(r) = g(h(r)) = \operatorname{sgn}(q)|h(r)|^{q/p} = \operatorname{sgn}(q) r^q, \quad p \leqslant q \neq 0. \tag{7.9}$$

For these problems g is convex and increasing and, by scaling arguments, we have

$$\bar{h}_\nu(s) = A_\nu s^{-p/2}, \tag{7.10}$$

where the coefficients $\{A_\nu\}$, which depend on p, are known in certain cases (completely for $p = -1$ and $p = 2$). Now from Eq. (7.7) we have

$$u = -\operatorname{sgn}(q)\frac{d}{ds}|A_\nu s^{-p/2}|^{q/p} \Leftrightarrow s = Q_\nu(u)$$
$$= \{|A_\nu|^{q/p} q/2u\}^{2/(q+2)}. \tag{7.11}$$

Hence Eq. (7.8) becomes in this case

$$E_N \geqslant \tfrac{1}{2} N V_0 u^{q/(q+2)} (q/2)^{2/(q+2)} (1 + 2/q) \times \sum_{\nu=1}^{N-1} |A_\nu|^{2q/p(q+2)}, \tag{7.12}$$

where $u = (2\hbar^2/NV_0 a^2 \lambda m)$. We can summarize this collection of power-law comparison examples as follows: the pair potential is

$$V(r) = \gamma \operatorname{sgn}(q) r^q \tag{7.13}$$

and the lower bound to E_N is given by

$$E_N \geqslant \tfrac{3}{4}(1 + 2/q)(N\gamma q/3)^{2/(q+2)} (\hbar^2/m)^{q/(q+2)} \times \sum_{\nu=1}^{N-1} |A_\nu|^{2q/p(q+2)}, \tag{7.14}$$

where we have set $\gamma = V_0 a^{-q}$ and $\lambda = \frac{4}{3}$, the $\{A_\nu\}$ are given by (7.10), and $p \leqslant q \neq 0$, $0 \neq p \geqslant -1$. As a partial check on the algebra we immediately look at the exactly soluble "pseudofermion" case $p = q = 2$ in *one* spatial dimension (i.e., the lowest *spatially* antisymmetric state). In this case $h_\nu(s) = (\nu - \frac{1}{2})^2/s$, $\nu = 1,2,3,\cdots$, and the lower bound E^{L}_N given by (7.14) becomes $E^{\mathrm{L}}_N = (6\gamma N)^{1/2}(N-1)^2(\hbar^2/m)^{1/2}/4$, whereas the exact energy[15,16] of this problem is given by $E_N = (2N\gamma)^{1/2}(N^2-1)(\hbar^2/m)^{1/2}/2$. Hence we obtain $E^{\mathrm{L}}_N/E_N = \sqrt{3}(N-1)/2(N+1)$, or 86% of the exact energy for large N, as we found in Ref. 7. We shall look at more interesting examples in Sec. VIII.

VIII. SOME EXAMPLES

We consider potentials given by the formula

$$f(r) = \sum_{i=1}^{k} g^{(i)}(f^{(i)}(r)).$$

If there is only one term, Theorem 4(c) is equivalent to the "method of potential envelopes" of Paper I. In this special case one seeks a dual representation for the potential in the form

$$f(r) = g^{(1)}(f^{(1)}(r)) = g^{(2)}(f^{(2)}(r)),$$

where $g^{(1)}$ is convex, leading to a lower bound, and $g^{(2)}$ is concave, leading to an upper bound without the use of a trial function. If exact trajectories are also known for the excited states of $(-\Delta + vf^{(1)})$ and $(-\Delta + vf^{(2)})$, the *nmutatis mutandis* the general trajectory formulas (4.8) yield bounds on each excited state as we saw in Sec. VI. If there is more than one term in the sum (6.1) then the theory applies only to the bottom of the spectrum of H in each angular momentum or symmetry subspace: the general theory *does* apply to these

special excited states. If the $\{g^{(i)}\}$ are all identity maps, then the results of the present paper are equivalent to those of Paper III. One of the results of Paper III is the general formula [Eq. (4.10) here] for linear combinations of powers and the log potential, which, for example, determines the lowest eigenvalue of $-\Delta + v(r^2 + \lambda r^4)$ to within 1% for all $\lambda \geqslant 0$: It is possible to establish such error claims independently because the trajectory formula yields *both* upper and lower bounds. In Paper II we explored the situation in which one knows exact *partial* trajectories, that is to say, $E(\alpha)$ for $-\Delta + \alpha f^{(1)} + f^{(2)}$; one can then treat potentials of the form $\alpha g(f^{(1)}) + f^{(2)}$, where g is convex or concave. Some of these various special cases of the present theory have been illustrated by the problems considered in the earlier papers: below we present some fresh examples.

If we consider a *one*-particle problem with central potential

$$V(r) = V_0 f(r/a), \quad V_0 > 0, \ a > 0,$$

particle mass m, and energy E', then the Schrödinger eigenvalue problem is equivalent to that generated by Eq. (1.1), i.e.,

$$H = -\Delta + v f(r), \quad v > 0$$

if we let

$$E = 2mE'a^2/\hbar^2$$

and

$$v = 2mV_0a^2/\hbar^2.$$

Thus various choices of range a are already allowed for in the trajectories $E = F(v)$. In considering linear combinations of potentials we shall usually introduce more parameters than are logically required; this redundancy allows one to see at a glance how the eigenvalues depend on the components. When we consider the N-body problem, the variables E and v must be interpreted according either to Eqs. (5.5) or (7.3), depending on whether the identical particles are bosons or fermions. In the examples, we consider the problem to have been solved once the upper or lower kinetic potentials are determined: The trajectories are then given by substitution into the general formula (4.8); numerical values are easily obtained with the aid of a programmable calculator.

A. Dual Coulombic-harmonic transformations

Suppose

$$f(r) = g^{(1)}(h^{(1)}(r)) = g^{(2)}(h^{(2)}(r)), \tag{8.1}$$

where

$$h^{(1)}(r) = -1/r, \quad h^{(2)}(r) = r^2, \tag{8.2}$$

$g^{(1)}$ is convex, $g^{(2)}$ is concave, and both the functions are increasing. Then by Theorem 6 and Eqs. (6.7) and (6.8) we have

$$g^{(1)}(-s^{1/2}/(n+l)) \leqslant \bar{f}_{nl}(s) \leqslant g^{(2)}((2n+l-\tfrac{1}{2})^2/s). \tag{8.3}$$

But $g^{(1)}(x) = f(-1/x)$ and $g^{(2)}(x) = f(\sqrt{x})$ so that we have the following general rule for this class of potentials:

$$f((n+l)/s^{1/2}) \leqslant \bar{f}_{nl}(s) \leqslant f((2n+l-\tfrac{1}{2})/s^{1/2}), \tag{8.4}$$

where $l = 0,1,2,\cdots$ and $n = 1,2,3,\cdots$. This is fascinating because the bounds on the kinetic potentials $\bar{f}_{nl}$ are given directly in terms of f itself: If $f(r) = -1/r$, the lower bound is exact; if $f(r) = r^2$, the upper bound is exact. The case ($n = 1$, $l = 0$) was discussed in Paper I, Sec. VB. We now have a situation (dual power-law representations *always* lead to this) where the upper and lower kinetic potentials are scale transformations of each other. Suppose we consider

$$\bar{f}_b(s) = h(s/b), \quad b > 0. \tag{8.5}$$

Then from (4.8) we have

$$\begin{aligned} G_b(u) &= su + h(s/b) \\ u &= -h'(s/b)/b, \end{aligned} \tag{8.6}$$

that is to say, using $t = s/b$,

$$\begin{aligned} G_b(u) &= t(ub) + h(t), \\ u_b &= -h'(t). \end{aligned} \tag{8.7}$$

Hence

$$G_b(u) = G_1(ub), \tag{8.8}$$

and, since, by Theorem 2(a), G is monotone increasing, we have

$$G_{b'}(u) > G_b(u), \quad b' > b. \tag{8.9}$$

Using this framework, we have from (8.4)

$$f((n+l)/s^{1/2})) \leqslant \bar{f}_{nl}(s) \leqslant f((n+l)/(bs)^{1/2}), \tag{8.10}$$

where

$$b = [(2n+l-\tfrac{1}{2})/(n+l)]^2 > 1. \tag{8.11}$$

We can therefore summarize our solution to this class of problems as follows:

$$\begin{aligned} G^{\mathrm{L}}_{nl}(u) &= su + f((n+l)/s^{1/2}) \\ u &= -\frac{d}{ds} f((n+l)/s^{1/2}) \\ G^{\mathrm{U}}_{nl}(u) &= G^{\mathrm{L}}_{nl}(u(2n+l-\tfrac{1}{2})^2/(n+l)^2), \end{aligned} \tag{8.12}$$

where f is the potential and the superscripts U and L stand for "upper" and "lower" ($l = 0,1,2,\cdots$, $n = 1,2,3,\cdots$). The class of potentials (8.1) includes, for example, the following combinations:

$$\begin{aligned} f(r) = \sum_{i=1}^{k} \alpha_i \operatorname{sgn}(p_i) r^{p_i} - \beta_1/(1+br) + \beta_2 \ln r \\ + \beta_3 r^2[1 + \lambda/(1+cr^2)], \end{aligned} \tag{8.13}$$

in which the coefficients $\{\alpha_i\}$ and $\{\beta_j\}$ are nonnegative and not all zero, $b > 0$, $c > 0$, $0 \leqslant \lambda \leqslant 3$, and $-1 \leqslant p_i \leqslant 2$ with $p_i \neq 0$, $i = 1,2,\ldots,k$. By Eq. (5.9) we know that, for the ground state ($l = 0$, $n = 1$), Eqs. (8.12) immediately apply to the N-boson problem. For the excited states of the one-particle problem this very general result yields vanishing percentage error as l increases.

B. Power-law potentials

For the special case of power-law interactions which have recently been of interest,[17] the potential is given by

$$V(r) = \operatorname{sgn}(q) V_0 (r/a)^q \equiv \operatorname{sgn}(q) \gamma r^q, \quad -1 \leqslant q \leqslant 2, \ q \neq 0, \tag{8.14}$$

and from (8.12) we immediately obtain the bounds

$$G^{A}_{nl}(u) = (1 + q/2)(|q|/2)^{2/(q+2)} \nu^{2q/(q+2)} u^{q/(q+2)} \tag{8.15}$$

with $\nu = (n + l)$ for the lower bound and $\nu = (2n + l - \frac{1}{2})$ for the upper bound.

For N identical bosons this implies [see Eq. (5.6) and set $l = 0, n = 1$]

$$E_N^A = (1 + q/2)(\gamma|q|/4)^{2/(q+2)}(\hbar^2/m)^{q/(q+2)}\nu^{2q(q+2)} \times (N-1)N^{2/(q+2)}, \tag{8.16}$$

in which $\nu = 1$ for the lower bound and $\nu = 3/2$ for the upper bound to the ground-state energy. Of course, the *lower* bound (for the N-boson system) is also valid for *all* $q \geqslant -1$, $q \neq 0$; for an upper bound valid for all $q \geqslant -1$, we simply use the Gaussian kinetic potential (see Table II), which yields

$$\nu_u = (\tfrac{3}{2})^{1/2}[\Gamma((3+q)/2)/\Gamma(3/2)]^{1/q}; \tag{8.17}$$

if $q = 2$, Eq. (8.17) implies $\nu_u = \frac{3}{2}$ and for $-1 \leqslant q < 2$ the value $\nu = \nu_u$ yields a better upper bound than $\nu = 3/2$ via Eq. (8.16). A better lower bound can always be found for a given q by use of the corresponding kinetic potential $\bar{h}^{(1)}(s)$ in Eq. (8.3).

We now turn to the N-fermion problem. We consider N identical spin-$\frac{1}{2}$ fermions, where

$$(N-1) = \sum_{\nu=1}^{k} 2\nu^2 = k(k+1)(2k+1)/3. \tag{8.18}$$

This choice of N allows us to fill exactly k "shells" of the basis Coulomb problem for which each eigenvalue labelled by $\nu = (n + l)$ has degeneracy $2\nu^2$. Hence in Eq. (7.14) we set $p = -1, q \geqslant p, q \neq 0$, and each distinct *value* of $A_\nu = -\nu^{-1}$ is repeated $2\nu^2$ times as we sum over the first $(N-1)$ eigenvalues. With this understanding we can write the sum over ν in Eq. (7.14) in the form

$$\sum_{\nu=1}^{k} (2\nu^2)\nu^{2q/(q+2)}$$

so that the lower bound to the ground-state energy of the N-fermion system becomes

$$E_N \geqslant E_N^L = \tfrac{3}{2}(1 + 2/q)(N\gamma|q|/3)^{2/(q+2)}(\hbar^2/m)^{q/(q+2)} \times \sum_{\nu=1}^{k} \nu^{4(q+1)/(q+2)}, \quad q \geqslant -1. \tag{8.19}$$

Consequently, for large N we have for all $q \geqslant -1$

$$E_N \geqslant E_N^L \sim N^Q, \quad Q = 5/3 + 2/(3q+6), \tag{8.20}$$

and, therefore, $5/3 < Q \leqslant 7/3$.

C. The Coulomb plus log potential

We have already treated the ground-state energy of the N-boson system in which the pair potential is a linear combination of powers $\mathrm{sgn}(p)r^p, p = -1,1,2,3,6,8$, and the log potential in Paper III [see Eq. (4.10) here]. Now we should like to look at the excited states, and the N-fermion problem, with a pair potential of this type. Suppose, for example, that

$$f(r) = -\alpha/r + \beta \ln r, \quad \alpha > 0, \quad \beta > 0. \tag{8.21}$$

The potential $f(r)$ satisfies the conditions of the dual representation (8.1) so that we have by (7.7) and (8.12)

$$G_{nl}^A(u) = su - \alpha s^{1/2}/\nu + \beta \ln(\nu/s^{1/2}),$$
$$s = Q_\nu(u) = \beta^2\{[(\alpha/2\nu)^2 + 2u\beta]^{1/2} - (\alpha/2\nu)\}^{-2}, \tag{8.22}$$
lower $\nu = (n + l)$, upper $\nu = (2n + l - \frac{1}{2})$.

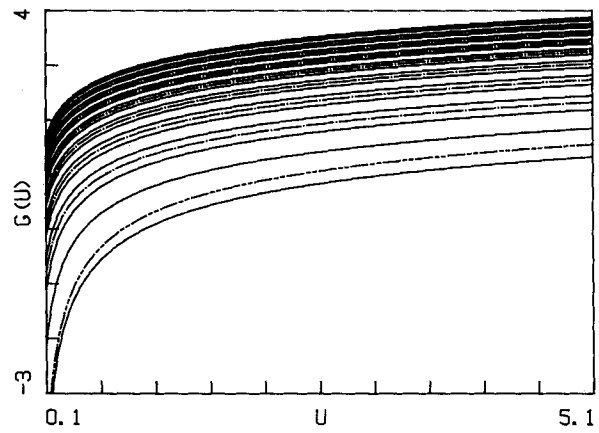

FIG. 1. Trajectories for the Coulomb plus log potential. Bounds on the eigenvalues $G(u)$ of $H = -u\Delta - 1/r + \ln r$ are shown for $n = 1$ and $l = 0$–8. Upper and lower bounds given by Eq. (8.22) are in the full line; the "center lines" for $l > 0$ are obtained by using a mean value $\nu = 3n/2 + l - 1/4$; the dashed line between the $l = 0$ bounds is essentially the exact result given by Eq. (8.23).

This recipe for the trajectories is in explicit form since we have solved for s in terms of u. The trajectories with $n = 1$ and $l = 0$–8 are shown in Fig. 1 for the case $\alpha = \beta = 1$; for $l \geqslant 1$ the "center lines" are means of the bounds obtained by using $\nu = 3n/2 + l - \frac{1}{4}$; for $l = 0$, the "center line" is the exact result. The exact values for $l = 0$ were obtained by using Eq. (4.10) of Sec. IV, which yields the following explicit formula:

$$G(u) = \inf\langle -u\Delta - \alpha/r + \beta \ln r\rangle = -\alpha/2t + \beta \ln(v^* t^2/2)/2, \tag{8.23}$$

where $t = [(8u\beta + \alpha^2)^{1/2} - \alpha]/2\beta$, and $v^* = v_{10} = 8.07$ for the lower bound and $v^* = v_e = 8.6057$ for an upper bound via the exponential trial function (see Table II). If we use the average value $(v_{10} + v_e)/2$, then the error in $G(u)$ is strictly less than 1.6% of β for all $\alpha > 0, \beta > 0$, and $u > 0$; for $\beta = 1$, this error is less than the resolution of Fig. 1.

The pure log potential is obtained by setting $\alpha = 0$ and $u = 1$ so that the eigenvalues E_{nl} of the operator $-\Delta + v \ln r$ (we now use v in place of β) are bounded by

$$-\tfrac{1}{2}v \ln(v/v_{nl}^L) \leqslant E_{nl} = -\tfrac{1}{2}v \ln(v/v_{nl}) \leqslant -\tfrac{1}{2}v \ln(v/v_{nl}^U), \tag{8.24}$$

where $v_{nl}^L = 2e(n + l)^2$ and $v_{nl}^U = 2e(2n + l - \frac{1}{2})^2$. We derived the general form of E_{nl} in Ref. 18; we had probably misread the graphs in Quigg and Rosner's article[19] for we find now, by numerical integration, $v_{10} = 8.07$ (rather than 7.63). Quigg and Rosner[19] find for S states the WKB approximation $v_{n0} \approx \pi(2n - \frac{1}{2})^2$. If we represent the potential $\ln r$ as a concave transformation of the linear potential (rather than the harmonic oscillator), then we get[4] $v_{n0} < 8e(|a_n|/3)^3$, where $\{a_n\}$ are the zeros of the Airy function ($a_1 = -2.338$, $a_2 = -4.088$, etc.[20]). We have indulged in this level of detail for a special case of the very general approximation (8.12) in order to put the result in some numerical perspective.

Now we turn to the N-fermion problem. By exactly similar reasoning to our treatment of the power-law potentials

in Sec. VIIIB we have by Eq. (7.8)

$$E_n \geqslant E_N^L = N\sum_{\nu=1}^{k}[u\nu^2 Q_\nu - \alpha\nu Q_\nu^{1/2} + \beta\nu^2 \ln(\nu/Q_\nu^{1/2})],$$
$$Q_\nu = \beta^2\{[(\alpha/2\nu)^2 + 2u\beta]^{1/2} - (\alpha/2\nu)\}^{-2}, \quad (8.25)$$
$$u = 3\hbar^2/2Nm,$$
$$N - 1 = k(k+1)(2k+1)/3, \quad k = 1,2,3,\cdots,$$

where the complete potential is given by (8.21) for we have set $V_0 = a = 1$. Once again, the pure log potential is recovered as the special case $\alpha = 0$.

D. The Yukawa plus linear potential

Consider the potential

$$f(r) = -\alpha e^{-r}/r + \beta r, \quad \alpha > 0, \ \beta > 0. \quad (8.26)$$

Now this function is concave and therefore also a concave function of r^2. Provided $2\beta \geqslant \alpha$, f is at the same time a convex function of $h(r) = -1/r$ and, therefore, under these conditions (the linear component sufficiently large), we can again immediately apply the recipe (8.12) to find bounds on all the energy eigenvalues. However, we wish in this final example to illustrate some other approaches.

Suppose $h_b(r)$ is the Hulthén potential,

$$h_b(r) = -(e^{br} - 1)^{-1}, \quad b > 0, \quad (8.27)$$

and a transformation g is chosen so that the Yukawa component of $f(r)$ is given by

$$-e^{-r}/r = g_b(h_b(r)), \quad (8.28)$$

where

$$g_b(X) = -\frac{b(1 - 1/X)^{-1/b}}{\ln(1 - 1/X)}, \quad X < 0. \quad (8.29)$$

We then find (after a tedious calculation) that g_b is convex for $b \geqslant \sqrt{6}$ and concave for $b \leqslant 1$; for values of b between these two bounds ($1 < b < \sqrt{6}$), the convexity of g_b is not definite. Now, by Table I and the scaling Theorem (2d) we have

$$\bar{h}_b(s) = -\tfrac{1}{2}[(1 + 4s/b^2)^{1/2} - 1], \quad b > 0. \quad (8.30)$$

Hence, by Theorem 4 we have, for a lower bound to the ground-state energy with $b = \sqrt{6}$

$$\bar{f}_L(s) = -\alpha b / W^{1/b} \ln W + 2\beta G_1 s^{-1/2},$$
where (8.31)
$$W(s) = [(1 + 4s/b^2)^{1/2} + 1]/[(1 + 4s/b^2)^{1/2} - 1],$$
$$b = \sqrt{6},$$

and $G_1 = 0.688\,041$ (from Table II). For an upper bound we use an exponential trial function and by Theorem (4b) we have

$$\bar{f}_U(s) = -4\alpha s^{3/2}(2s^{1/2} + 1)^{-2} + 2\beta G_1^E s^{-1/2}, \quad (8.32)$$

where $G_1^E = 0.75$ (from Table II). The resulting trajectories $F(v) = vG(1/v)$ [bounds on the lowest eigenvalue of $H = -\Delta + vf(r)$] are shown in Figs. 2 and 3 for the cases $(\alpha, \beta) = (1,0)$ and $(1,1)$, respectively.

For the pure Yukawa potential $y(r) = -(re^r)^{-1}$ the following simple result may sometimes be useful. We find by

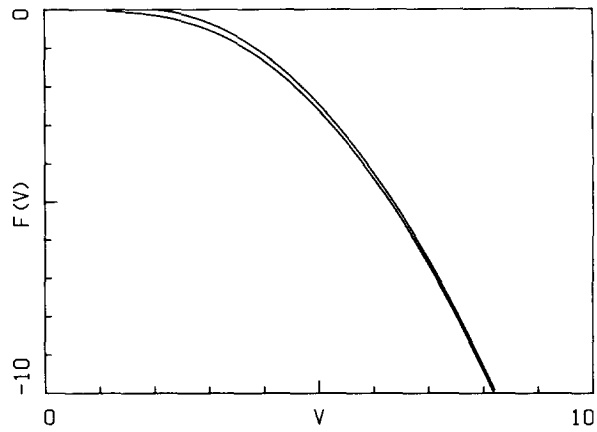

FIG. 2. Trajectories for the Yukawa potential. Upper and lower bounds on the lowest eigenvalue $F(v)$ of $H = -\Delta - ve^{-r}/r$ provided by the kinetic-potential bounds (8.31) and (8.32) with $(\alpha, \beta) = (1,0)$.

elementary methods that

$$0 \leqslant (re^r)^{-1} - [\sinh(r)e^r]^{-1} < 0.056, \quad r > 0. \quad (8.33)$$

Hence

$$-2(2^{2r} - 1)^{-1} - 0.056 < -(re^r)^{-1} \leqslant -2(e^{2r} - 1)^{-1} \quad (8.34)$$

Now the S-state eigenvalues of the Hulthén potential are known exactly[13] and for the potential $h_b(r) = -(e^{br} - 1)^{-1}$ we have the exact trajectories

$$F_{n0}^{(b)}(v) = -(v - n^2b^2)^2/4n^2b^2. \quad (8.35)$$

Consequently, the S state eigenvalues $F_{n0}(v)$ of the operator $H = -\Delta - v(re^r)^{-1}$ are bounded by

$$-(v - 2n^2)^2/4n^2 - (0.056)v \leqslant F_{n0}(v) \leqslant -(v - 2n^2)^2/4n^2. \quad (8.36)$$

It turns out that if we apply the convexity estimates of the first part of this section to the excited S states of the pure Yukawa potential by using Eq. (8.35) with $b = \sqrt{6}$ and 1, the results are not sufficiently better than (8.36) to justify the extra complication. Accurate results for the pure Yukawa

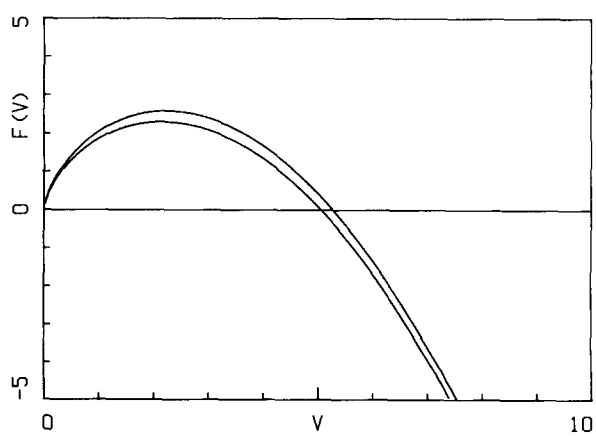

FIG. 3. Trajectories for the Yukawa plus linear potential. Upper and lower bounds on the lowest eigenvalue $F(v)$ of $H = -\Delta + v(-e^{-r}/r + r)$ provided by the kinetic-potential bounds (8.31) and (8.32) with $(\alpha, \beta) = (1,1)$.

potential may be found in Ref. 21; recent results[22] for the Hulthén potential with $l>0$ may also allow us to estimate the higher angular-momentum energy of eigenvalues corresponding to the Yukawa potential by the above methods.

IX. CONCLUSION

We have presented a variety of results concerning bounds on Schrödinger eigenvalues for central potentials. The bounds are analytical recipes which exhibit the dependence of the eigenvalues on the potential parameters. The separations of the bounds are typically a few percent for ground states, rather larger for higher radial quantum numbers, and vanishingly small (as a percentage) for large angular momenta. Such results are useful in the exploratory stages of modelling and as checks on other techniques such as perturbation methods. Our principal motivation, however, has been the N-identical-particle problem whose energy is intimately related to the two-body energy trajectories via the necessary permutation symmetry of the N-body wavefunction.

The introduction of the kinetic potential $\bar{f}(s)$ corresponding to a potential $f(r)$ has allowed us to unify our earlier methods for convex transformations of potentials and for sums of potentials. The optimization of "weights" feature of the potential-sum method of Paper III is also now automatically incorporated in the new general formulation: the problem of finding the optimal w, $0<w<1$, in the representation

$$-\Delta+v(f_1+f_2)=w\{-\Delta+(v/w)f_1\}+(1-w)\{-\Delta+(v/(1-w))f_2\}$$

is now solved by the minimization with respect to the parameter $s>0$. Moreover, s is now identified as the mean kinetic energy for both upper and lower bounds. There is also a certain harmony between the functional analysis[1-3] which regards the potential as a perturbation of $-\Delta$ and our analytical methods in which $\langle -\Delta \rangle$ is set equal to s and then the energy trajectory $(v, F(v))$, $v>0$, is given in terms of a Legendre transformation of the kinetic potential $(s, f(s))$, $s>0$. The Λ transform which yields $F=\Lambda(f)$ has been factored in the form $\Lambda=\Lambda^{(2)}\circ\Lambda^{(1)}$, where $\bar{f}=\Lambda^{(1)}(f)$ and $F=\Lambda^{(2)}(\bar{f})$. In general, $\Lambda^{(1)}$ is labelled by n and l.

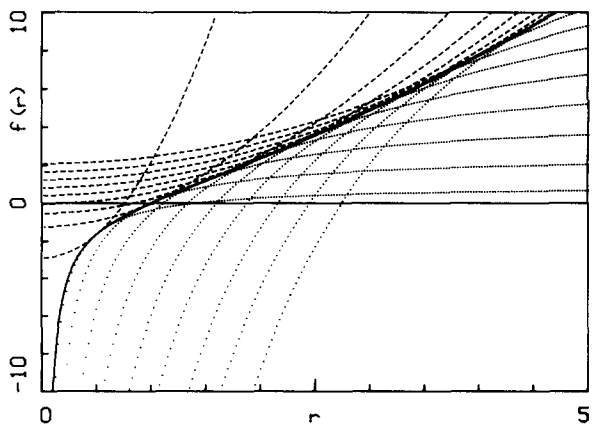

FIG. 4. Dual envelope representations for the potential $f(r)=-1/r+r^{3/2}$. The potential is represented as the envelope of a family of Coulomb potentials (hyperbolas $h^{(t)}$) below, and of a family of harmonic-oscillator potentials (parabolas $p^{(t)}$) above.

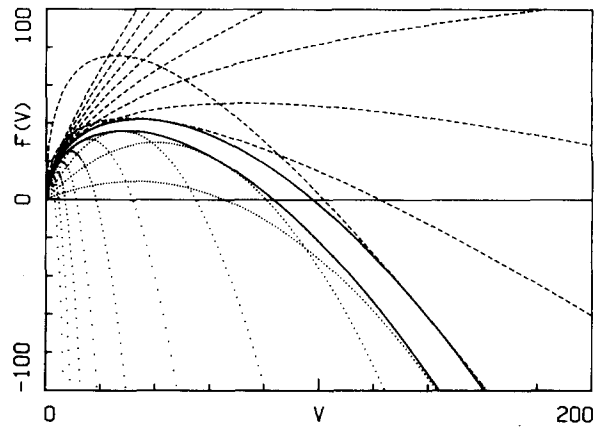

FIG. 5. Trajectory bounds by the potential envelope method. The Λ_{nl} transform for $(n, l)=(1,5)$ is applied to the families $\{h^{(t)}\}$ and $\{p^{(t)}\}$ of hyperbolas and parabolas of Fig. 4. The envelopes of the *new* lower and upper families do not now coincide but split into the lower F^{L}_{nl} and upper F^{U}_{nl} bounds to the unknown exact energy trajectory $\Lambda_{nl}(f)=F_{nl}$. $F^{U}_{15}(v)$ is a magnification of the curve $F^{L}_{15}(v)$ with magnification factor $\mu=(2n+l-\frac{1}{2})^2/(n+l)^2=(13/12)^2$.

In the present paper we have used min–max arguments along with Jensen's inequality to establish our results. Jensen's inequality is a perfectly natural tool, but the idea of the envelope representation from with the potential-envelope method of Paper I originally came may provide the richer conceptual framework. In Fig. 4 we show the potential $f(r)=-1/r+r^{3/2}$ along with its dual $(-1, 2)$ power-law envelope representations

$$f=\underset{t}{\text{envelope}}\{h^{(t)}\}=\underset{t}{\text{envelope}}\{p^{(t)}\} \tag{9.1}$$

by hyperbolas $\{h^{(t)}\}$ below, and parabolas $\{p^{(t)}\}$ above, where $t>0$ is the point of contact. In Fig. 5 we exhibit, for $n=1$ and $l=5$, the curves $\Lambda_{nl}(h^{(t)})$ and $\Lambda_{nl}(p^{(t)})$ along with their envelopes. We have, as a consequence of Theorem 6,

$$\underset{t}{\text{envelope}}\{\Lambda_{nl}(h^{(t)})\}\leqslant F_{nl}\leqslant\underset{t}{\text{envelope}}\{\Lambda_{nl}(p^{(t)})\}. \tag{9.2}$$

The transformation Λ_{nl} applied to the envelope components of f has given rise to a *splitting* of the unknown exact trajectory $F_{nl}=\Lambda_{nl}(f)$ into upper and lower bounds.

Dual power-law representations lead to the concept of *magnification*: the curve $(\mu v, \mu F(v))$ is a magnification of the curve $(v, F(v))$, where $\mu>0$. The *entire* collections of upper $\{F^{U}_{nl}\}$ and lower $\{F^{L}_{nl}\}$ curves given, for example, in the case of $(-1, 2)$ dual representations, by Eq. (8.12) are magnifications of a *single* curve, namely, the curve $(v, F^{*}(v))$ given in parametric form by

$$v^{-1}=-\frac{d}{ds}f(s^{-1/2}),\quad s>0,$$
$$F^{*}(v)=s+vf(s^{-1/2}). \tag{9.3}$$

This result is trivial when f itself is a power law but is otherwise more interesting. In Fig. 5, for example, the upper tra-

jectory is a magnification of the lower trajectory, where $\mu = [(2n + l - \frac{1}{2})/(n + l)]^2 = (13/12)^2$. For a given dual power-law representation *all* the relative magnifications are fixed; they do not depend on the details of f provided only that f does have the given representation.

The restriction in this paper to increasing potentials $f'(r) > 0$ is not essential. This is merely a simple device which, along with nice behavior at $r = 0$ and $r = \infty$, guarantees the existence of eigenvalues for sufficiently large coupling constants v (or small u). Molecular potentials can be treated along the same lines provided the various Hamiltonians which are reached by transformations and sums can be defined on essentially the same domain. For the N-body problem there is reason to expect that our lower bounds will be weaker in the case of saturating potentials.[23]

Similarly, our variational arguments remain valid for potentials which are not smooth. Useful results can be obtained with families of square wells: a monotone increasing potential, for example, is bounded above by an enveloping family of infinite square wells and (if the potential is nonsingular) at the same time below, by a family of finite square wells.

The energies of the excited states of many-particle systems can be treated with the aid of the present results by the method of Ref. 9 provided there are independent reasons for believing that the objects approximated by min–max arguments are in fact eigenvalues. The bounds which we have found in the present paper for the N-body problem simply concern the bottom of the energy spectrum, whether or not this object is an eigenvalue. If one is interested in a particular symmetry or angular-momentum state of the N-body system, then a lower bound to the bottom of the spectrum of H restricted to the corresponding subspace may be found by the methods of this paper with the aid also of group theory.[24]

If we consider the N-boson problem with the linear pair potential

$$f(r) = V_0 r/a = \gamma r,$$

then we find by Eq. (5.6) that

$$A_1(N-1)(N\gamma)^{2/3}(\hbar^2/4m)^{1/3} \leqslant E_N \leqslant A_2(N-1)(N\gamma)^{2/3}(\hbar^2/4m)^{1/3}, \tag{9.4}$$

where $-A_1 \doteqdot -2.338\ 107$ is the first zero of the Airy function, and $A_2 = 3(3/2\pi)^{1/3}$. Thus (9.4) determines the lowest energy of this N-body problem to 0.15% for all $\gamma > 0$ and all $N \geqslant 2$.

Results like this are not difficult to obtain by the methods of this paper: Our theory is complementary to conventional many-body theory[25] which has been developed primarily to tackle "physical problems," where F_N/N approaches a finite limit as N increases. In the conventional theory the interaction is regarded as a perturbation and, just as in QED, one is rarely able to discuss the value of the sum of the entire perturbation series. Because of this, definite bounds like (9.4) and the large variety of similar results which are made possible by Schrödinger Lego may prove to be very useful.

ACKNOWLEDGMENT

The author would like to thank the Natural Sciences and Engineering Research Council Canada for partial financial support of this work by Grants A3438 and E6447.

[1] E. Prugovečki, *Quantum Mechanics in Hilbert Space* (Academic, New York, 1981).
[2] M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV: Analysis of Operators* (Academic, New York, 1978).
[3] W. Thirring, *A Course in Mathematical Physics 3: Quantum Mechanics of Atoms and Molecules* (Springer-Verlag, New York, 1981).
[4] R. L. Hall, Phys. Rev. D **22**, 2062 (1980); Paper I.
[5] R. L. Hall and M. Satpathy, J. Phys. A: Math. Gen. **14**, 2645 (1981); Paper II.
[6] R. L. Hall, Phys. Rev. D **23**, 1421 (1981); Paper III.
[7] R. L. Hall, Proc. Phys. Soc. **91**, 16 (1967).
[8] R. L. Hall, Phys. Lett. B **30**, 320 (1969).
[9] R. L. Hall, Phys. Rev. C **20**, 1155 (1979).
[10] W. Feller, *An Introduction to Probability Theory and its Applications* (Wiley, New York, 1971), Vol. II.
[11] I. M. Gel'fand and S. V. Fomin, *Calculus of Variations* (Prentice-Hall, Englewood Cliffs, N.J., 1963).
[12] R. L. Hall, Aequ. Math. **8**, 281 (1972); Can. J. Phys. **50**, 305 (1972).
[13] S. Flügge, *Practical Quantum Mechanics* (Springer-Verlag, New York, 1974).
[14] V. Efimov, Phys. Lett. B **33**, 563 (1970); R. D. Amado and J. V. Noble, Phys. Rev. D **8**, 1992 (1972); R. L. Hall, Z. Phys. A **291**, 255 (1979).
[15] H. R. Post, Proc. Phys. Soc. A **66**, 649 (1954); A **69**, 936 (1956); A **79**, 819 (1962).
[16] R. L. Hall and B. Schwesinger, J. Math. Phys. **20**, 2481 (1979).
[17] F. T. Hioe and E. W. Montroll, J. Math. Phys. **16**, 1945 (1975); F. T. Hioe, D. MacMillen, and E. W. Montroll, J. Math. Phys. **17**, 1320 (1976); H. Turschner, J. Phys. A: Math. Gen. **12**, 451 (1979); J. Killingbeck, J. Phys. A: Math. Gen. **13**, 49 (1980); B. J. B. Crowley and T. F. Hill, J. Phys. A: Math. Gen. **12**, L223 (1979); L. K. Sharma, J. Choubey, and H. J. W. Müller-Kirsten, J. Math. Phys. **21**, 1533 (1980); M. S. Ashbaugh and J. D. Morgan III, J. Phys. A: Math. Gen. **14**, 809 (1981); R. E. Crandall and M. Hall Reno, J. Math. Phys. **23**, 64 (1982).
[18] R. L. Hall and C. Kalman, Phys. Lett. B **83**, 80 (1979).
[19] C. Quigg and J. L. Rosner, Phys. Lett. B **71**, 153 (1977).
[20] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. Stegun (Dover, New York, 1970).
[21] C. S. Lam and Y. P. Varshni, Phys. Rev. A **4**, 1875 (1971); F. J. Rogers, H. C. Graboske, Jr., and D. J. Harwood, Phys. Rev. A **1**, 1577 (1970); N. Bessis, G. Bessis, G. Corbel, and B. Dakhel, J. Chem. Phys. **63**, 3744 (1975).
[22] C. S. Lai and W. C. Lin, Phys. Lett. A **78**, 335 (1980).
[23] L. W. Bruch and I. J. McGee, J. Chem. Phys. **59**, 409 (1973); T. Brady, E. Harms, L. Laroze, and J. S. Levinger, Nucl. Phys. A **168**, 509 (1971); T. K. Lim and A. Zuniga, J. Chem. Phys. **63**, 2245 (1975); R. N. Hill, J. Math. Phys. **21**, 1070 (1980).
[24] R. L. Hall, J. Phys. A: Math. Nucl. Gen. **1**, 468 (1968); F. Cabal and L. W. Bruch, J. Chem. Phys. **70**, 4669 (1979); L. W. Bruch, J. Chem. Phys. **72**, 5511 (1980).
[25] D. J. Thouless, *The Quantum Mechanics of Many-Body Systems* (Academic, New York, 1972); A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971).

Exact results for the diffusion in a class of asymmetric bistable potentials

M.-O. Hongler[a),b)] and W. M. Zheng[c),d)]
Center for Studies in Statistical Mechanics, The University of Texas at Austin, Austin, Texas 78712

(Received 3 June 1982; accepted for publication 20 August 1982)

We solve the Fokker–Planck equations with drifts deriving from a class of asymmetric nonharmonic potentials which include bistable cases. An analytical expression for the probability current over the potential barrier is obtained. Finally, we compare our exact results with those obtained by Kramers' approximation.

PACS numbers: 05.40. + j, 05.30. − d

In a recent communication,[1] we proposed exactly solved models for the diffusion in a class of nonharmonic, symmetric potentials which includes bistable cases. Here, we give an extension to asymmetric situations.

The diffusion problem we are considering reads

$$\frac{\partial}{\partial t}P(x,t\,|x_0,0) = \frac{\partial}{\partial x}\left\{\left(\frac{d}{dx}U(x)\right)P(x,t\,|x_0,0) + \frac{\partial}{\partial x}P(x,t\,|x_0,0)\right\}, \quad x\in\mathbb{R} \tag{1a}$$

with

$$P(x,t\,|x_0,0) = \delta(x-x_0), \tag{1b}$$

$$P(x,t\,|x_0,0)\geqslant 0, \quad \forall t\in\mathbb{R}^+, \ \forall x\in\mathbb{R}, \tag{1c}$$

$$\int_{\mathbb{R}} P(x,t\,|x_0,0)\,dx = 1, \quad \forall t\in\mathbb{R}^+, \tag{1d}$$

$$\begin{aligned} U(a,x) &= 2\ln\phi(a,x) \\ &= 2\ln\{y_1(a,x)+\beta y_2(a,x)\} \\ &= -\frac{x^2}{2} + \ln\left\{{}_1F_1\left(\frac{a}{2}+\frac{1}{4},\frac{1}{2},\frac{x^2}{2}\right) + \beta x\,{}_1F_1\left(\frac{a}{2}+\frac{3}{4},\frac{3}{2},\frac{x^2}{2}\right)\right\}, \end{aligned} \tag{1e}$$

$$a\in\left]-\tfrac{1}{2},\infty\right[\tag{1f}$$

and

$$\beta<\beta_c = \sqrt{2}\,\Gamma\left(\frac{a}{2}+\frac{4}{3}\right)\left[\Gamma\left(\frac{a}{2}+\frac{1}{4}\right)\right]^{-1}. \tag{1g}$$

The condition (1f) guarantees that the confluent hypergeometric function ${}_1F_1(\alpha,\beta,z)$ in Eq. (1e) is positively defined and the condition (1g) combined with (1f) implies that $\phi(a,x)$ is itself positively defined. The function $\phi(a,x)$ itself is a solution of the Weber equation[2]:

$$\frac{d^2}{dx^2}\phi(a,x) = \left(\frac{x^2}{4}+a\right)\phi(a,x). \tag{2}$$

The potential $U(a,x)$ in Eq. (1e) is asymmetric when $\beta\neq 0$ (the symmetric case $\beta=0$ has been studied in Ref. 1). Moreover, for $a\in\left]-\frac{1}{2},0\right[$ and $\beta<\beta_c$, $U(a,x)$ exhibits asymmetric bistable shapes (see Fig. 1).

To solve the diffusion problem (1) by van Kampen's method,[3] let us write an associated Schrödinger equation in the form

$$\frac{d^2}{dx^2}\psi(x) + [E-V(x)]\psi(x) = 0, \tag{3}$$

with

$$\begin{aligned} V(x) &= \frac{1}{4}\left[\frac{d}{dx}U(x)\right]^2 - \frac{1}{2}\frac{d^2}{dx^2}U(x) \\ &= -\frac{x^2}{4} - a + 2\left[\frac{d}{dx}\{\ln\phi(a,x)\}\right]^2. \end{aligned} \tag{4}$$

By using the transformation discussed in Ref. 4,

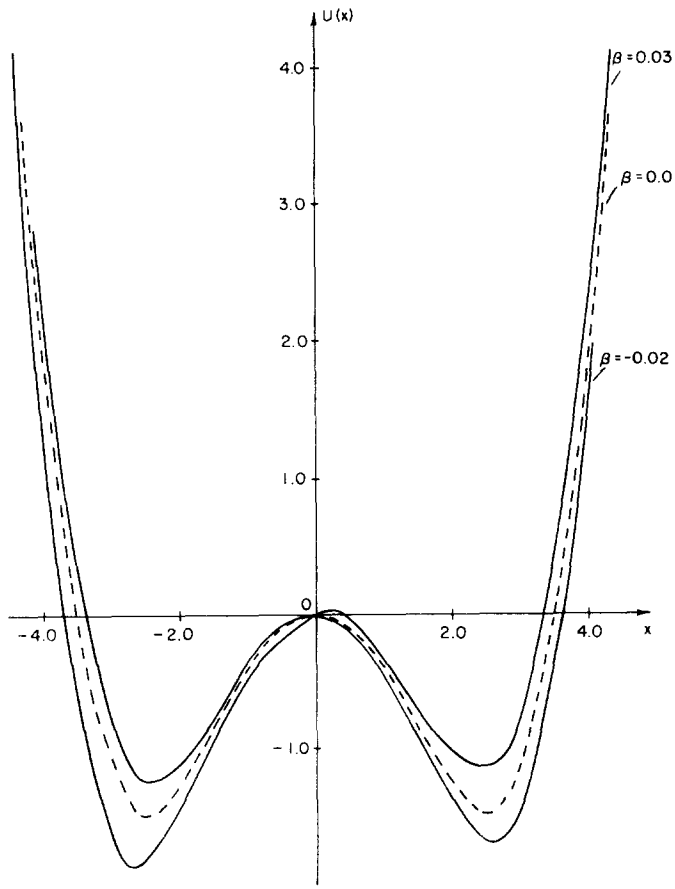

FIG. 1. Shape of the potential $U(x)$ for $a = -0.4$ and various asymmetries $\beta = -0.02;\ 0;\ +0.03$.

[a)] On leave from the University of Geneva, Department of Theoretical Physics, Bvd. d'Yvoy, 1211 Geneva, 4, Switzerland.

[b)] Supported by the Swiss National Science Foundation.

[c)] On leave from the Institute of Theoretical Physics, Academia Sinica, Beijing, China.

[d)] Supported by the Robert A. Welch Foundation.

 0022-2488/83/020336-05$02.50

$$\mathscr{W}\psi(x) = \chi(x) = [\phi(a,x)]^{-1}\frac{d}{dx}\{\phi(a,x)\psi(x)\}, \tag{5}$$

the Schrödinger problem (3) with the potential (4) (see Fig. 2) can be solved exactly.

Indeed, from Eqs. (2)–(5), we obtain

$$\frac{d^2}{dx^2}\chi(x) = \left(\frac{x^2}{4} + a - E\right)\chi(x). \tag{6}$$

Taking into account that $\psi(x)$ is square-integrable, the solution of the Schrödinger problem (6) reads

$$\chi_n(x) = D_n(x) = \exp\{-x^2/4\}H_n(x/\sqrt{2}), \tag{7}$$

$$E_n = n + a + \tfrac{1}{2}, \quad n = 0,1,2,... \tag{8}$$

and the ground state has the form[4]

$$\psi_{E=0}(x) = [\phi(a,x)]^{-1} = \phi(a,x)\frac{d}{dx}\left\{\frac{\xi_1 D_{-a-1/2}(x) + \xi_2 D_{-a-1/2}(-x)}{\phi(a,x)}\right\}, \tag{9}$$

where ξ_1 and ξ_2 are real constants.

From Eqs. (7)–(9), we can write the solution of the diffusion problem (1) in the form

$$P(x,t\,|x_0,0) = N^{-1}\{\phi(a,x)\}^{-2} + \frac{\phi(a,x_0)}{\phi(a,x)} \times \sum_{n=0}^{\infty} e^{-E_n t}\psi_n(x)\psi_n(x_0), \tag{10}$$

where

$$\psi_n(x) = -E_n C_n \phi^{-1}(a,x)\int \phi(a,x)D_n(x)\,dx \tag{11a}$$

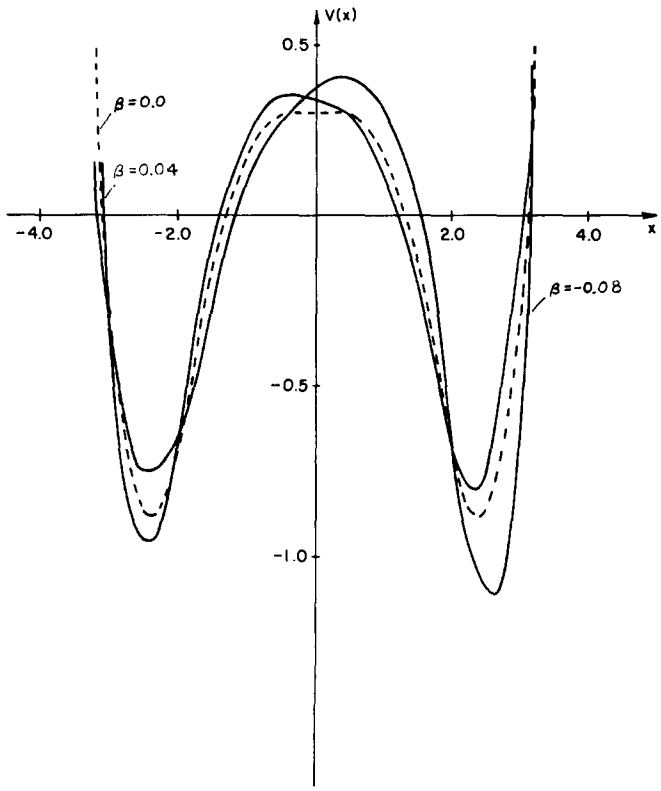

FIG. 2. Shape of the potential $V(x)$ for $a = -0.2$ and various asymmetries $\beta = -0.04; 0; +0.08$.

$$\psi_n(x) = C_n\phi(a,x)\frac{d}{dx}\{D_n(x)\phi^{-1}(a,x)\}. \tag{11b}$$

The last equality in Eq. (11b) has been established in Ref. 4.

The normalization factors N and C_n occurring in Eqs. (10) and (11) are calculated in the Appendices A and B, where we find

$$N = 2\beta_c[\beta_c^2 - \beta^2]^{-1}, \tag{12}$$

$$C_n^{-2} = (n + a + \tfrac{1}{2})n!(2\pi)^{1/2}. \tag{13}$$

From the solution (10), we can immediately calculate the probability current over the potential barrier located at $x = x_B$ (see Fig. 1). We obtain

$$J(x = x_B, x_0, t) = -\frac{\partial}{\partial x}P(x,t\,|x_0,0)\Big|_{x=x_B} = -[\phi(a,x_B)]^{-1}\sum_{n=0}^{\infty}\frac{E_n\exp\{-E_n t\}}{(2\pi)^{1/2}n!} \times D_n(x_B)\int^{x_0}\phi(a,x)D_n(x)\,dx \tag{14a}$$

$$= \sum_{n=0}^{\infty} E_n C_n^2 \exp\{-E_n t\} \times\left[\phi(a,x)\frac{d}{dx}D_n(x) - D_n(x)\frac{d}{dx}\phi(a,x)\right]\Big|_{x=x_B} \times D_n(x_B)[\phi(a,x_B)]^{-1}, \tag{14b}$$

where Eqs. (14a) and (14b), respectively, follow from the use of Eqs. (11a) and (11b).

For small asymmetries ($\beta \ll \beta_c$), the location of x_B can be approximately obtained by keeping only the first-order terms in the expansion of ${}_1F_1(\alpha,\beta,z)$. Namely, we have

$$\frac{d}{dx}U(a,x)\Big|_{x=x_B} = \frac{d}{dx}\left\{e^{-x^2/4}\left[\left(1 + (a+\tfrac{1}{2})\frac{x^2}{2} + \cdots\right) + \beta x\left(1 + (a+\tfrac{3}{2})\frac{x^2}{6} + \cdots\right)\right]\right\}\Big|_{x=x_B} = 0, \tag{15}$$

and hence

$$x_B = -\beta/a + O(\beta^2). \tag{16}$$

Hence by introducing the value (16) into Eq. (14a) and considering $t \gg 1$ such that only the first eigenvalue $n = 0$ contributes, the current $J(x = x_B, x_0, t \gg 1)$ up to first order in β reads

$$J(x = x_B, x_0, t \gg 1) = -\frac{(a+\tfrac{1}{2})\exp\{(a+\tfrac{1}{2})t\}}{(2\pi)^{1/2}}\int^{x_0} e^{-x^2/4}\phi(a,x)\,dx$$

$$= -\frac{(a+\tfrac{1}{2})}{(2\pi)^{1/2}}\exp\{-(a+\tfrac{1}{2})t\}\left[x_0\,{}_1F_1\left(\frac{a}{2}+\frac{5}{4},\frac{3}{2},\frac{x_0^2}{2}\right) + \beta\,{}_1F_1\left(\frac{a}{2}+\frac{3}{4},\frac{1}{2},\frac{x_0^2}{2}\right)\Big/\sqrt{2}(a+\tfrac{1}{2})\right]e^{-x_0^2/2}. \tag{17}$$

From Ref. 5, the current given in Eq. (17) permits us to calculate the large time scale (τ_l) characterizing the decay process:

$$\tau_l^{-1} = -\frac{d}{dt}\ln\{J(x=x_B,x_0,t\gg 1)\} = a+\tfrac{1}{2}. \tag{18}$$

According to Kramers $(\tau_l)^{-1}$ can be expressed in terms of the extreme of the potential located at $x=x_{\pm}$ (minima) and $x=x_B$ (maximum) (see Fig. 1), and in terms of the corresponding curvatures $\omega_\alpha = d^2U(x)/dx^2|_{x=x_\alpha}$. By Ref. 6, we have

$$\begin{aligned}(\tau_l^{-1})_{\mathrm{KRA}} &= (2\pi)^{-1}[(\omega_+|\omega_B|)^{1/2}\exp\{U(x_+)-U(x_B)\} \\ &\quad + (\omega_-|\omega_B|)^{1/2}\exp\{U(x_-)-U(x_B)\}] \\ &= \pi^{-1}\left\{\left[\left(\frac{x_+^2}{4}+a\right)\left|\frac{x_B^2}{4}+a\right|\right]^{1/2}\left[\frac{\phi(a,x_+)}{\phi(a,x_B)}\right]^2\right. \\ &\quad \left.+\left[\left(\frac{x_-^2}{4}+a\right)\left|\frac{x_B^2}{4}+a\right|\right]^{1/2}\left[\frac{\phi(a,x_-)}{\phi(a,x_B)}\right]^2\right\}.\end{aligned} \tag{19}$$

As we have (see Appendix C)

$$x_{\mp} \simeq 2\ln\left\{\frac{\beta_c}{(a+\frac{1}{2})\pi(\beta_c \mp \beta)}\right\} \tag{20}$$

and

$$\phi(a,x_{\mp}) = 2\frac{\beta_c \mp \beta}{x_{\mp}}, \tag{21}$$

Eq. (19) reduces to

$$(\tau^{-1})_{\mathrm{KRA}} = \frac{2}{\pi^{1/2}}(a+\tfrac{1}{2}), \tag{22}$$

which is the same as the result (18) up to the factor $2/\pi^{1/2}\sim 1.1$.[7]

We close this paper by mentioning that the diffusion problem (1) exhibits a situation where the two wells of the potential are not well separated. Indeed, the eigenvalues of the associated quantum mechanical problem are here equally spaced whereas for well separated wells, the spectrum presents a structure of close pairs.[8] Therefore, the approximation schemes based on the WKB method for the corresponding Schrödinger problem are likely to fail for the class of models presented here.

ACKNOWLEDGMENTS

We would like to thank Professor L. E. Reichl and Professor I. Prigogine for their hospitality at the Center for Studies in Statistical Mechanics of the University of Texas at Austin.

NOTATIONS AND FORMULAS

We introduce here the notations and formulas which we shall use in the Appendices.

$$\lambda_0 = \frac{2^{a/2-3/4}}{\pi^{1/2}}\Gamma\left(\frac{a}{2}+\frac{3}{4}\right), \tag{N1}$$

$$\lambda_{\mp} = \lambda_0(1 \pm \beta/\beta_c). \tag{N2}$$

We use the notations of Ref. 2 and have

$$\begin{aligned}\phi(a,x) &= y_1(a,x)+\beta y_2(a,x) \\ &= \lambda_+ D_{-a-1/2}(x)+\lambda_- D_{-a-1/2}(-x) \\ &= \lambda_+ U(a,x)+\lambda_- U(a,-x).\end{aligned} \tag{N3}$$

The following integral is given in Ref. 9:

$$I_n = \int_{-\infty}^{+\infty}[D_n(x)]^2\,dx = (2\pi)^{1/2}n!. \tag{N4}$$

We shall use the asymptotic developments[2]

$$U(-\nu-\tfrac{1}{2},|x|) = D_\nu(|x|) \simeq e^{-x^2/4}|x|^\nu[1+O(x^{-2})], \tag{N5}$$

$$\begin{aligned}&U(-\nu-\tfrac{1}{2},-|x|) \\ &\quad = D_\nu(-|x|) \\ &\quad \simeq e^{x^2/4}|x|^{-\nu-1}\frac{(2\pi)^{1/2}}{\Gamma(-\nu)}[1+O(x^{-2})] \\ &\qquad + e^{-x^2/4}|x|^\nu\cos\nu\pi[1+O(x^{-2})],\end{aligned} \tag{N6}$$

and therefore

$$\frac{d}{dx}D_\nu(|x|) \simeq -\frac{|x|}{2}D_\nu(|x|), \tag{N7}$$

$$\frac{d}{dx}D_\nu(-|x|) = \frac{|x|}{2}D_\nu(-|x|). \tag{N8}$$

From Ref. 2 we have

$$y_1(a,x) = [\cos\alpha U(a,x)+\Gamma(\tfrac{1}{2}-a)\sin\alpha V(a,x)]k_1, \tag{N9}$$

$$y_2(a,x) = [\sin\alpha U(a,x)+\Gamma(\tfrac{1}{2}-a)\cos\alpha V(a,x)]k_2, \tag{N10}$$

with

$$\alpha = \left(\frac{a}{2}+\frac{1}{4}\right)\pi, \tag{N11}$$

$$k_1 = \pi^{1/2}2^{a/2+1/4}\left[\Gamma\left(\frac{1}{4}-\frac{a}{2}\right)\right]^{-1}, \tag{N12}$$

$$k_2 = \pi^{1/2}2^{a/2-1/4}\left[\Gamma\left(\frac{3}{4}-\frac{a}{2}\right)\right]^{-1}. \tag{N13}$$

Finally, for $a<0$, $x^2+4a\gg|a|$, and $x>0$, the Darwin expansion reads[2]

$$U(a,x) \simeq k_3 x^{-1/2}\exp\{-x^2/4\}, \tag{N14}$$

$$V(a,x) \simeq k_4 x^{-1/2}\exp\{x^2/4\}, \tag{N15}$$

with

$$k_3 = [\Gamma(\tfrac{1}{2}-a)]^{1/2}(2\pi)^{-1/4}, \tag{N16}$$

$$k_4 = \{[\Gamma(\tfrac{1}{2}-a)]^{1/2}(2\pi)^{1/4}\}^{-1}. \tag{N17}$$

APPENDIX A

We calculate the normalization factor N in Eq. (10).[10] We have

$$N^{-1} = \int_{-\infty}^{\infty}\frac{dx}{[\phi(a,x)]^2} = \int_{-\infty}^{\infty}\frac{dx}{[y_1(a,x)+\beta y_2(a,x)]^2}. \tag{A1}$$

Using Eqs. (9) and (N3), Eq. (A1) takes the form

$$N^{-1} = \left.\frac{\xi_1 U(a,x)+\xi_2 U(a,-x)}{\lambda_+ U(a,x)+\lambda_- U(a,-x)}\right|_{-\infty}^{+\infty}. \tag{A2}$$

Then using Eqs. (N5) and (N6), Eq. (A2) gives

$$N^{-1} = \frac{\xi_2}{\lambda_-}-\frac{\xi_1}{\lambda_+} = \frac{\xi_2\lambda_+-\xi_1\lambda_-}{\lambda_+\lambda_-}. \tag{A3}$$

To calculate the numerator in Eq. (A3) we use Eq. (9) for which we can write

$$[\lambda_+ U(a,x) + \lambda_- U(a,-x)]^{-1}|_{x=\infty}$$
$$= \frac{d}{dx}\left[\xi_1 U(a,x) + \xi_2 U(a,-x)\right]|_{x=\infty}$$
$$- \frac{\xi_1 U(a,x) + \xi_2 U(a,-x)}{\lambda_+ U(a,x) + \lambda_- U(a,-x)}$$
$$\times\left[\frac{d}{dx}\{\lambda_+ U(a,x) + \lambda_- U(a,-x)\}|_{x=\infty}\right]. \quad \text{(A4)}$$

We finally use the asymptotic developments (N5), (N6), (N7), and (N8) to obtain

$$(\xi_2\lambda_+ - \xi_1\lambda_-)xU(a,x) = [U(a,-x)]^{-1}, \quad \text{(A5)}$$

and hence from (A3) we have

$$N^{-1} = 2\beta_c[\beta_c^2 - \beta^2]^{-1}, \quad \text{(A6)}$$

where

$$\beta_c = \sqrt{2}\Gamma(a/2 + \tfrac{3}{4})[\Gamma(a/2 + \tfrac{1}{4})]^{-1}. \quad \text{(A7)}$$

The normalization constant N being positively defined, Eq. (A6) is meaningful only when $\beta < \beta_c$. This last condition guarantees that $\phi(a,x) > 0$, $\forall x \in \mathbb{R}$ (in terms of quantum mechanics, the ground state does not present nodes). Let us illustrate this last point by the following example.

Example: $a = \frac{1}{2}$,

$$\phi(\tfrac{1}{2},x) = y_1(\tfrac{1}{2},x) + \beta y_2(\tfrac{1}{2},x)$$
$$= \exp\{x^2/4\}[1 + (\beta\pi^{1/2}/\sqrt{2})\,\mathrm{erf}\,(x/\sqrt{2})]. \quad \text{(A8)}$$

It is clear from Eq. (A8) that the function $\phi(\frac{1}{2},x)$ remains strictly positive provided $\beta < 2\pi^{-1/2}$ which precisely is the value of β_c given in Eq. (A7) for $a = \frac{1}{2}$.

APPENDIX B

Here, we calculate the coefficients C_n in Eq. (11).[10] Using the representations (11a) and (11b), we have

$$C_n^{-2} = \int_{\mathbf{R}} \psi_n^2(x)\,dx = -E_n \int_{\mathbf{R}} \left(\frac{d}{dx}\{D_n(x)/\phi(a,x)\}\right)\left(\int \phi(a,x)D_n(x)\,dx\right)dx$$
$$= -E_n\left\{\left[\frac{D_n(x)}{\phi(a,x)}\int_0^x \phi(a,x)D_n(x)\,dx\right]\Big|_{-\infty}^{+\infty} - \int_{\mathbf{R}}[D_n(x)]^2\,dx\right\}. \quad \text{(B1)}$$

Then using Eqs. (11a) and (11b) and (N4), Eq. (B1) can be written in the form

$$C_n^{-2} = -E_n\left\{\left[D_n(x)\frac{\phi(a,x)}{(-E_n)}\frac{d}{dx}(D_n(x)/\phi(a,x))\right]\Big|_{-\infty}^{+\infty} - n!(2\pi)^{1/2}\right\}$$
$$= \left[D_n(x)\phi(a,x)\frac{d}{dx}(D_n(x)/\phi(a,x))\right]\Big|_{-\infty}^{+\infty} + E_n n!(2\pi)^{1/2}. \quad \text{(B2)}$$

Finally, using Eqs. (N2), (N5), and (N6), we obtain

$$C_n^{-2} = E_n n!(2\pi)^{1/2} = (n + a + \tfrac{1}{2})n!(2\pi)^{1/2}. \quad \text{(B3)}$$

APPENDIX C

We calculate the location of the minima $x_\mp$ of $\phi(a,x)$ when $a \sim -\frac{1}{2}$. When $a \sim -\frac{1}{2}$, $x_\mp$ are expected to be large and therefore we use the Darwin expansion Eqs. (N14), (N15), (N16), and (N17) to obtain

$$\phi(a,x) = y_1(a,x) \mp \beta y_2(a,x)$$
$$\simeq (k_1\cos\alpha \mp \beta k_2 \sin\alpha)k_3 x^{-1/2}e^{-x^2/4}$$
$$+ (k_1\sin\alpha \mp \beta k_2\cos\alpha)k_4\Gamma(\tfrac{1}{2} - a)x^{-1/2}e^{-x^2/4}$$
$$(x \to \mp\infty). \quad \text{(C1)}$$

From Eq. (C1), the condition $(d/dx)\phi(a,x)|_{x=x_\mp} = 0$ reads

$$x_\mp \simeq 2\ln\left\{\frac{\beta_c \mp \tan^2\alpha}{2(\beta_c \mp \beta)\tan\alpha}\right\}, \quad \text{(C2)}$$

where α is defined in Eq. (N11) and β_c in Eq. (A7).

Finally, using (C2) and the Darwin expansion, we have

$$y_1(a,x_\mp) \mp \beta y_2(a,x_\mp)$$
$$\simeq \frac{k_3k_1}{x_\mp^{1/2}}\left(\cos\alpha \mp \frac{\beta\tan\alpha}{\beta_c}\sin\alpha\right)e^{-x_\mp^2/4}$$
$$+ 2\left(\sin\alpha \mp \frac{\beta}{\beta_c}\tan\alpha\cos\alpha\right)e^{x_\mp^2/4}$$
$$\simeq 2\cos\alpha\left[\frac{\beta_c \mp \beta\tanh^2\alpha(\beta_c \mp \beta)}{\beta_c x_\mp}\right]^{1/2}. \quad \text{(C3)}$$

For $a \sim -\frac{1}{2}$, we end with

$$y_1(a,x_\mp) + \beta y_2(a,x_\mp) \simeq 2\left(\frac{\beta_c \mp \beta}{x_\mp}\right)^{1/2}. \quad \text{(C4)}$$

[1] M.-O. Hongler and W. M. Zheng, J. Stat. Phys. **29**, 317 (1982).
[2] M. Abramovitz and I. Segun, *Handbook of Mathematical Functions* (Dover, New York, 1964).
[3] N. G. van Kampen, J. Stat. Phys. **17**, 71 (1977).
[4] W. M. Zheng, "The Darboux transformation and solvable double-well potential models for Schrödinger equations" (to appear in J. Math. Phys.).

[5]R. B. Griffiths, C. Y. Weng, and J. S. Langer, Phys. Rev. **149**, 301 (1966).
[6]E. M. Montroll, in *Energetics in Metallurgical Phenomena*, edited by W. M. Müller (Gordon and Breach, New York, 1967), Vol. III.
[7]W. M. Zheng, "On the extra factor $\pi^{1/2}/2$ in the Kramers Approximation" (to appear in Physica A).
[8]D. ter Haar, *Problems in Quantum Mechanics* (Academic, New York, 1974), p. 144.
[9]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965).
[10]W. M. Zheng, "Associated Harmonic Oscillator Potential Models for Schrödinger Equations" (to appear in J. Phys. A).

Quantization of spinor fields. III. Fermions on coherent (Bose) domains

Piotr Garbaczewski [a)]
Department of Physics, Brown University, Providence, Rhode Island 02912

(Received 21 July 1981; accepted for publication 7 December 1981)

A formulation of the c-number classics-quanta correspondence rule for spinor systems requires all elements of the quantum field algebra to be expanded into power series with respect to the generators of the canonical commutation relation (CCR) algebra. On the other hand, the asymptotic completeness demand would result in the (Haag) expansions with respect to the canonical anticommutation relation (CAR) generators. We establish the conditions under which the above correspondence rule can be reconciled with the existence of Haag expansions in terms of asymptotic free Fermi fields. Then, the CAR become represented on the state space of the Bose (CCR) system.

PACS numbers: 11.10. – z

1. MOTIVATION

Our basic purpose is to deal with quantum field theory models (irrespective of the space-time dimensionality) whose elements of the field algebra admit a reconstruction in terms of one or more quantum free fields. By free we understand the field solutions of standard sourceless field equations like, e.g., the Klein–Gordon, Dirac, Maxwell, etc., ones. In addition, we require the equal-time canonical (anti)commutation relations to be satisfied on appropriate domains. The latter are, however, not required to belong to the Fock space.

For quantum fields with well defined asymptotics, the above reconstruction is realized in the form of the Haag series. In what follows, by Haag series we understand any power series in terms of the normal ordered products of the CCR or CAR algebra generators, denoted: $F(a^*, a)$:, $:F(b^*, b):$, respectively.

As is well known, the asymptotic condition is not an obvious notion even for the simplest Fermi system; compare, e.g., Ref. 1 and references therein. In this connection we admit the Haag series reconstruction of quantum fields in terms of free fields which are not the asymptotic series in the usual sense of the word.[2]

In $1+1$ dimensions, for all models solvable via the Bethe *ansatz* technique, the construction of the eigenstates of the Hamiltonian explicitly involves the fundamental free fields; compare, e.g., Refs. 3–5. We know, for example,[6] that in case of the sine-Gordon system the underlying field is the massive neutral scalar. In case of the massive Thirring model the free massive Dirac field is used to construct the energy eigenstates. However, to relate this quantum model to its completely integrable c-number (semiclassical) relative, one is forced to adopt a "bosonization" in terms of the massive neutral vector boson.[1]

A quite analogous situation appears in the infrared QED, where a bosonization of the quantum Dirac field weakly coupled to the photon field is realized in terms of the Coulomb gauge free Maxwell field potential.[1]

A common property of both the Fermi and Bose models mentioned above is that to relate quantum and classical (c-number) levels of a given field theory model, one starts from the Haag-like expansions $\hat{F} = F(a^*, a)$ in terms of the fundamental CCR algebra generators. Then one makes a boson transformation $a \to a + \bar{\lambda}$, $a \to a + \lambda$, where λ is a c-number function, and finally calculates the Fock vacuum expectation value in the tree approximation

$$(0|\hat{F}_\lambda|0) \to (0|:\hat{F}_\lambda:|0) = (0|:F(a^* + \bar{\lambda}, a + \lambda):|0) = F(\bar{\lambda}, \lambda). \tag{1.1}$$

The functional power series $F(\bar{\lambda}, \lambda)$ stand for classical, c-number relatives of the quantum objects $\hat{F} = F(a^*, a)$, to which $:F(a^*, a):$ corresponds in the tree approximation. One knows that the tree approximation prescription can be used to recover the classical Euler analogs of the quantum equations of motion.

It is of special importance to know these boson transformation parameters λ, which in the tree approximation give rise to the classical solitons. This problem was partially solved (for solitons) for the Korteweg–de Vries[7] and $\lambda\Phi^4$ models,[8,9] and more generally for the sine-Gordon system.[6,10,11] The latter case, using the Orfanidis' formulas,[12] allows an identification of at least some soliton solutions of the massive Thirring model. For a few other models in connection with a coherent state description of hadrons, see Ref. 13. The tree approximation procedure can be described as follows:

$$\hat{F} = F(a^*, a) \to F(a^* + \bar{\lambda}, a + \lambda) = \hat{F}_\lambda,$$

$$(0|F(a^* + \bar{\lambda}, a + \lambda)|0) := (\lambda|F(a^*, a)|\lambda) = (\lambda|\hat{F}|\lambda), \tag{1.2}$$

$$(\lambda|:F(a^*, a):|\lambda) = F(\bar{\lambda}, \lambda),$$

where $|\lambda)$ stands for a generalized coherent state for the field (CCR) algebra. In general $|\lambda)$ is not an element of the Fock space and hence gives rise to its own $|\lambda)$th Hilbert space irreducibility sector for the CCR algebra, incomplete direct product space IDPS $(|\lambda)) \subset H$ in the general Hilbert space H. For the particular case of Fermi models one can start from the Haag expansions in terms of the CAR generators: $F(b^*, b)$: but then the bosonization enters via $b = b(a^*, a)$,

[a)]Permanent address: Institute of Theoretical Physics, University of Wroclaw, 50-205 Wroclaw, Poland.

 0022-2488/83/020341-06$02.50

$b^* = b^*(a^*, a)$,[14–16] so that

$$F(b^*, b) = F[b^*, b](a^*, a) = G(a^*, a),$$
$$(0|G(a^* + \bar{\lambda}, a + \lambda)|0) = (\lambda|F[b^*, b](a^*, a)|\lambda), \quad (1.3)$$
$$(\lambda|:G(a^*, a):|\lambda) = G(\bar{\lambda}, \lambda).$$

In particular $(\lambda|b(a^*,a)|\lambda) = b(\bar{\lambda},\lambda)$, $(\lambda|b^*(a^*,a)|\lambda) = b(\bar{\lambda},\lambda) = b(\lambda,\bar{\lambda})$ correspond to b, b^*, respectively. Here the CAR generators are by construction acting on the Bose domain, hence we *are confronted with a serious problem of representations of the CAR algebra living in the non-Fock representations of the CCR algebra the latter being based on generalized coherent states.*

Let us recall that the case of Fock representation has been investigated and solved in Ref. 14, while the non-Fock case was not considered in full generality. We know only[17] that the CAR do allow a local representation in the Hilbert space of the Bose system, i.e., that the CAR hold true (while on a lattice) for a finite number of degrees of freedom, but may not hold true for almost all would-be Fermi degrees of freedom, upon bosonization.

As we show below, only a very special class of (Bose) coherent states allows the existence of fermions (representation of the CAR) on subspaces of IDPS($|\lambda\rangle$) and that in general the CAR are prohibited. In the latter case, the interacting spinor field does not possess an asymptotic spinor partner ("confinement" property), and this role is played by the fundamental boson(s) affiliated with the underlying representation of the CCR algebra. More precisely, it means that in the von Neumann–Hilbert space H of the Bose system we can find irreducibility domains for the CCR algebra such that the CAR can be irreducibly represented on a subspace. On these subspaces an asymptotic expansion of the interacting spinor field $\hat{\Psi} = \Psi(\hat{\psi}_{\text{in}})$ in terms of the free fermion $\hat{\psi}_{\text{in}}$ makes sense. Whenever the CCR algebra irreducibility sector in H does not carry an irreducible CAR algebra representation, the underlying expansion makes no sense, and $\hat{\Psi}$ should be expanded with respect to the free boson: The free fermion is then "confined" and $\hat{\Psi}$ does not possess an asymptotic spinor partner.

2. MAIN THEOREM

For clarity, we shall abandon the explicitly continuous case and restrict considerations to the product representations of the CCR and CAR algebras.[17–20] We refer to Ref. 20 in connection with the role of coherent states in this case.

Let $H = \Pi_k^{\otimes} h_k$, $h_k = h \forall k = 1, 2, \ldots$ be the von Neumann infinite direct product Hilbert space. It is an infinitely reducible carrier space for the representation of the CCR algebra generated by a countable sequence of Schrödinger representations $\{a^*, a\}_j$:

$$[a_i, a_j]_-|\psi) = 0 = [a_i^*, a_j^*]_-|\psi),$$
$$[a_i, a_j^*]_-|\psi) = |\psi)\delta_{ij}, \quad (2.1)$$
$$\forall ij, |\psi) \in H.$$

Let $|\psi) = \Pi_k^{\otimes} f_k, f_k \in h_k$ be a product vector with the property $\|f_k\| = 1 \forall k$. With each $|\psi)$ we have associated a separable Hilbert space IDPS($|\psi)$) on which a representation of the CCR algebra acts irreducibly. Among all possible product vectors in H, we shall distinguish the coherent states, which can formally be obtained from the Fock vacuum $|0) \in H$, $a_i|0) = 0 \forall i$, $|0) = \Pi_k^{\otimes} f_k^0$, by applying a product mapping $U_\lambda^{\otimes}$:

$$U_\lambda^{\otimes}|0) = \Pi_k^{\otimes}(U_\lambda f^0)_k := \Pi_k^{\otimes}|\lambda)_k = |\lambda),$$
$$U_\lambda = \exp(\lambda a^* - \bar{\lambda} a), \quad \lambda_k \in C \forall k. \quad (2.2)$$

Here $|\lambda)$ is determined by fixing a denumerable sequence (λ) of complex parameters. We have $a_i|\lambda) = \lambda_i|\lambda)$, $(\lambda|\lambda) = 1$. The incomplete direct product space based on $|\lambda)$ we denote IDPS ($|\lambda)$). Two coherent product states are equivalent: $|\lambda) \sim |\gamma)$ if and only if the series

$$\sum_k (\bar{\lambda}_k \gamma_k - \tfrac{1}{2}|\lambda_k|^2 - \tfrac{1}{2}|\gamma_k|^2) \quad (2.3)$$

converges.[20] When $\Sigma_k |1 - |(\lambda_k|\gamma_k)|| < \infty$, we talk about a weak equivalence $|\lambda) \underset{w}{\simeq} |\gamma)$. One knows[20] that the weakest condition for the CCR algebra representations acting in IDPS ($|\lambda)$), IDPS ($|\gamma)$), respectively, to be unitarily equivalent is that $|\lambda) \underset{w}{\simeq} |\gamma)$. In particular $|\gamma) \sim |\lambda) \not\Rightarrow |\gamma) \underset{w}{\simeq} |\lambda)$. Notice that if $\Sigma_j |\lambda_j|^2 \not< \infty$ then $|0) \underset{w}{\neq} |\gamma)$. If $\Sigma_j |\lambda_j - \gamma_j|^2 \not< \infty$ then $|\lambda) \underset{w}{\neq} |\gamma)$.

Let us denote

$$P = :\exp(-a^*a): + a^* :\exp(-a^*a):a \quad (2.4)$$

a projection on a two-dimensional subspace h_F of h spanned by vectors f^0 and $a^* f^0 = f^1$. For a countable sequence $\{a^*, a\}_j$ we introduce a corresponding countable sequence $\{P_j\}$, and observe that the operators

$$\sigma_j^+ = a_j^* :\exp(-a_j^* a_j): \equiv P_j a_j^* P_j,$$
$$\sigma_j^- = :\exp(-a_j^* a_j): a_j \equiv P_j a_j P_j \quad (2.5)$$

satisfy the following commutation relations on the Hilbert space:

$$\text{IDPS}_F(|0)) = 1_F \text{IDPS}(|0)), \quad 1_F = \Pi_k^{\otimes} P_k:$$
$$[\sigma_i^+, \sigma_j^-]_- = 0 = [\sigma_i^+, \sigma_j^+]_- = [\sigma_i^-, \sigma_j^-]_-, \quad i \neq j \quad (2.6)$$
$$[\sigma_i^-, \sigma_i^+]_+ = P_i :, \quad P_i|\psi) = |\psi) \forall i, \quad \forall |\psi) \in \text{IDPS}_F(|0)).$$

By applying the Jordan–Wigner transformation to the set $\{\sigma^+, \sigma^-\}_j$ one can easily reproduce a sequence $\{b^+, b\}_j$ of the related CAR algebra generators. *We wish to emphasize that the condition*

$$[\sigma_i^-, \sigma_i^+]_+|\psi) = |\psi) \forall i \quad (2.7)$$

is a crucial requirement, to have the CAR algebra represented on a domain to which a vector $|\psi)$ belongs. Notice that the relations (2.7) are immediate if $|\psi)$ appears in the form of the product vector:

$$|\psi) = \Pi_k^{\otimes}(\alpha f^0 + \beta f^1)_k,$$
$$|\alpha_k|^2 + |\beta_k|^2 = 1 \forall_k, \quad (2.8)$$
$$\alpha f^0 = 0, \quad \alpha^* f^0 = f^1.$$

Vectors of this form are the conventional product ones used to investigate representations of the CAR algebra.[18,19] Notice that (2.7) does not hold true if applied to a coherent product state $|\lambda)$. We relate the above mentioned representation of the CAR algebra to that of the spin $\frac{1}{2}$ algebra (2.6) via

the Jordan–Wigner trick:

$$b_k^* = \exp\left(i\pi \sum_{j=1}^{k-1} \sigma_j^+ \sigma_j^-\right)\sigma_k^+, \qquad k = 1, 2, \ldots \quad (2.9)$$

$$b_k = \exp\left(i\pi \sum_{j=1}^{k-1} \sigma_j^+ \sigma_j^-\right)\sigma_k^-.$$

It is easy to verify that (2.7) reads

$$[b_k, b_k^*]_+|\psi) = [\sigma_k^-, \sigma_k^+]_+|\psi) \forall k, \quad (2.10)$$

and that (2.6) implies

$$\begin{aligned} &[b_k, b_j]_+|\psi) = 0, \quad k \neq j \\ &[b_k, b_j^+]_+|\psi) = 0, \\ &[b_k^+, b_j^+]_+|\psi) = 0. \end{aligned} \quad (2.11)$$

Moreover, if $\sigma_i^+, b_i^*, a_i^*, \sigma_i^-, b_i, a_i$ are applied to the Fock state $|0)$ we find

$$\sigma_i^+|0) = b_i^*|0) = a_i^*|0) \forall i, \quad b_i|0) = 0 = a_i|0) = \sigma_i^-|0), \quad (2.12)$$

i.e., the basic property of the Fock representation constructed in Ref. 14.

Theorem: Suppose we have given IDPS($|\lambda)$), where $|\lambda)$ is a coherent product state determined by a complex sequence $(\lambda) = \{\lambda_1, \lambda_2, \ldots\}$, where $\lambda_k = |\lambda_k|\exp(i\delta_k)$, $|\lambda_k|, \delta_k \in \mathbf{R}^1$. In addition to the sequences $(|\lambda)|$ and (δ) let us introduce the three additional real ones (ϕ), (ψ), (α). Assume that

$$(1) \sum_k |\lambda_k|^2 = \infty, \quad (2) \sum_k |\lambda_k|^4 < \infty,$$

$$(3) \lim_{k\to\infty} \frac{\phi_k}{|\lambda_k|^4} = A \neq 0, \infty,$$

$$(4) \lim_{k\to\infty} \frac{\psi_k - \sigma_k}{|\lambda_k|^2} = B \neq 0, \infty,$$

$$(5) \lim_{k\to\infty} \frac{\alpha_k}{|\lambda_k|} = 1.$$

Then a product vector $|\psi) = \Pi_k^{\otimes}(uf^0 + vf^1)_k$, with

$$u_k = \cos\alpha_k \exp(i\phi_k), \quad v_k = \sin\alpha_k \exp(i\psi_k) \quad (2.13)$$

is an element of IDPS($|\lambda)$).

Proof: It suffices to prove that vectors $|\lambda)$ and $|\psi)$ are equivalent. The equivalence cirterion is $\Sigma_k |z_k| < \infty$, where

$$z_k = 1 - [\cos\alpha_k \exp(i\phi_k) + |\lambda_k| \times \exp i(\psi_k - \delta_k)\sin\alpha_k]\exp(-|\lambda_k|^2/2). \quad (2.14)$$

Let us consider $k \geqslant k_0 \gg 1$, when all the parameters are close to 0. Then, upon expanding z_k into a Taylor series about 0, we have

$$\begin{aligned} \operatorname{Re} z_k &\simeq 1 - \left(1 - \frac{|\lambda_k|^2}{2}\right)\left[\left(1 - \frac{\alpha_k^2}{2}\right)\left(1 - \frac{\phi_k^2}{2}\right) + |\lambda_k|\left(1 - \frac{(\psi_k - \delta_k)^2}{2}\right)\alpha_k\right], \\ \operatorname{Im} z_k &\simeq \left(1 - \frac{|\lambda_k|^2}{2}\right)\left[\left(1 - \frac{\alpha_k^2}{2}\right)\phi_k + |\lambda_k|(\psi_k - \delta_k)\alpha_k\right], \end{aligned} \quad (2.15)$$

i.e., by virtue of (1)–(5),

$$\operatorname{Re} z_k \simeq |\lambda_k|^4, \quad \operatorname{Im} z_k \simeq (A + B)|\lambda_k|^4. \quad (2.16)$$

Consequently,

$$\lim_{k\to\infty} \frac{|z_k|}{|\lambda_k|^4} = [1 + (A + B)^2]^{1/2} \neq 0, \infty. \quad (2.17)$$

Because of (2) the equivalence criterion holds true, and $|\psi) \sim |\lambda)$. Consequently,

$$|\psi) \in \text{IDPS}(|\lambda)).$$

It is worth emphasizing that we must have here $\lim_{k\to\infty} |z_k| = 0$. It leads to $\operatorname{Re} z_k \underset{k\to\infty}{\to} 0$, i.e.,

$$[\cos\alpha_k \cos\phi_k + |\lambda_k|\sin\alpha_k \cos(\psi_k - \delta_k)] \to \exp(|\lambda_k|^2/2),$$

which holds true if and only if $|\lambda_k| \underset{k\to\infty}{\to} 0$.

Remark 1: Notice that in the above, at a fixed choice of parameters $|\lambda_k| \in R^+$, we still have a freedom in the choice of phases (δ) in the complex sequence (λ), which is furthermore reflected in the appropriate freedom of choice of the phases (ψ) in the product vector $|\psi)$. The latter is obviously regulated by

$$\lim_{k\to\infty} \frac{\psi_k - \delta_k}{|\lambda_k|^2} = B. \quad (2.18)$$

A consequence of this is that if we have two sequences (λ), (λ'),

$$\lambda_k = |\lambda_k|\exp(i\delta_k), \quad \lambda_k' = |\lambda_k|\exp(i\delta_k') \forall k,$$

then the condition

$$\sum_k |\lambda_k|^2[\cos(\delta_k - \delta_k') - 1] < \infty \quad (2.19)$$

is a sufficient and necessary condition for the product vectors $|\psi) \in \text{IDPS}(|\lambda))$, $|\psi') \in \text{IDPS}(|\lambda'))$ to be weakly equivalent. To see this, it is enough to notice that product vectors $|\lambda)$, $|\lambda')$ are weakly equivalent if and only if the real part of (2.3) converges. In fact

$$(\lambda|\lambda') = \exp\left\{-\frac{1}{2}\sum_k |\lambda_k - \lambda_k'|^2 + i\sum_k \operatorname{Im}(\bar{\lambda}_k \lambda_k')\right\} \quad (2.20)$$

and $\Sigma_k |\lambda_k - \lambda_k'|^2 < \infty$ is just the same as (2.19). Obviously, if $\Sigma_k |\lambda_k - \lambda_k'|^2 = \infty$, then $|\lambda) \underset{w}{\perp} |\lambda')$.

Remark 2: The above theorem can also be deduced as a special case of a more general theory of Ref. 19. Namely, if h is a Hilbert space with an orthonormal basis $(e_k)_0^\infty$, and p a projection on a linear span of $e_0, \ldots, e_N$ so that $P_N = p_1 \cdots p_N$ is a projection in IDPS($|\lambda)$), then

(1) there exists a limiting projection $P = \lim_{N\to\infty} P_N$ in IDPS($|\lambda)$);

(2) by expanding $|\lambda)_i = \Sigma_k \gamma_i^k e_k = \Sigma_k \lambda_i^k/(k!)^{1/2} e_k$, we arrive at the following conclusion:

$$P \neq 0 \text{ if and only if } \sum_i \left[1 - \left(\sum_{k=0}^N |\gamma_i^k|^2\right)^{1/2}\right] < \infty;$$

(3) the vector $|\psi)$, $P|\psi) \neq 0$ can be constructed as follows:

$$|\lambda) = \Pi_i^{\otimes}|\psi_i), \quad |\psi_i) = \left(\sum_{k=0}^N |\bar{\gamma}_i^k|^2\right)^{-1/2} \sum_{k=0}^N \bar{\gamma}_i^k e_k. \quad (2.21)$$

In the special case of $N=1$, we have $\gamma_i^0=\exp(-|\lambda_i|^2/2)$ and $\lambda_i^1=\lambda_i\exp(-|\lambda|^2/2)$, and $\Sigma_{k=1}^{\infty}|\lambda|_k^4<\infty$, $\Sigma_{k=1}^{\infty}|\lambda_k|^2=\infty$, is a necessary and sufficient condition for a projection P to exist in IDPS($|\lambda\rangle$).

Remark 3: Notice that states $|\psi\rangle=\Pi_i^{\otimes}|\psi_i\rangle$ in (2.21), (2.13) have exactly the structure required by the spin 1/2 approximation procedure of Ref. 4 for quantum Bose systems. The above (Remark 2) statement is more general, however, and allows a construction of quantum spin chain states (with a fixed finite spin) in the Hilbert space of an interacting (non-Fock) Bose system; see in this connection also Ref. 15. The Holstein–Primakoff SU(2) generators

$$\begin{aligned} S_i^+ &= (2s)^{1/2}a_i^*(1-a_i^*a_i/(2s))^{1/2},\\ S_i^- &(2s)^{1/4}(1-a_i^*a_i(2s))^{1/2}a, \\ S_i^3 &= s-a_i^*a_i, \end{aligned} \tag{2.22}$$

provide us with an irreducible (at each ith site) representation of the SU(2) group Lie algebra corresponding to spin $s=N/2$, given by

$$\mathbf{S}_P=P\mathbf{S}P, \tag{2.23}$$

where P is a limiting projection of Remark 1.

3. DISCUSSION

Let us notice that the existence of $|\psi\rangle$ in IDPS($|\lambda\rangle$) guarantees that all vectors equivalent to $|\psi\rangle$, of the form $\Pi_k^{\otimes}(\alpha f^0+\beta f^1)_k$, $|\alpha_k|^2+|\beta_k|^2=1\forall_K$, are elements of IDPS($|\lambda\rangle$). A Hilbert space closure of the set of all linear combinations of such equivalent product vectors, $\mathrm{IDPS}_F(|\psi\rangle)$ is a subspace of IDPS($|\lambda\rangle$). The CAR are irreducibly represented on $\mathrm{IDPS}_F(|\psi\rangle)$ provided $\{b^*, b\}_j$ are constructed from $\{a^*, a\}_j$ according to Ref. 14. Let us also observe[19] that once we have any product vector $|\gamma\rangle\in$IDPS($|\lambda\rangle$) with the basic property $[\sigma_j^-, \sigma_j^+]_+|\gamma\rangle=|\gamma\rangle\forall_j$ then the following two properties cannot be simultaneously satisfied: (1) $\sigma_i^-|\gamma\rangle=0\ \forall_i$, (2) $|\gamma\rangle\neq 0$ under an additional restriction (3) $|\lambda\rangle\overset{\pm}{w}|0\rangle$, where $|0\rangle$ is a Fock state in H, and $|\lambda\rangle$ is a coherent product state. Consequently, there exists a unitary inequivalence of the CCR algebra representations associated with IDPS($|\lambda\rangle$), IDPS($|\lambda'\rangle$), where $|\lambda\rangle\overset{\pm}{w}|\lambda'\rangle$ implies a unitary inequivalence of the related CAR algebra representations in $\mathrm{IDPS}_F(|\psi\rangle)$, $\mathrm{IDPS}_F(|\psi'\rangle)$, respectively. Let us here emphasize that a particular form of the boson transformation parameter for a concrete field theory model follows from its equations of motion. This severe restriction may violate, and in general it does, the condition (2) of the Theorem of Section 2. In this case the bosonic semiclassic (i.e., the CCR representation based on the coherent product state) prevents us from having represented the CAR on the appropriate domain. *The "semiclassical Hilbert space" allows at most a local representation of the* CAR *on a subspace*,[17] i.e., with a property $[b_i, b_i^*]_+|\gamma\rangle=|\gamma\rangle$ for a *finite*, though arbitrarily large, number of modes, $|\gamma\rangle$ belonging to this subspace. Notice that by defining an arbitrary polynomial $W_{(j)}(b^*, b)$ in terms of "bosonized" Fermi generators $\{b^*, b\}_{j\in(j)}$, (j) being a finite set of indices, we arrive at the following definition of *locally Fermi, but globally coherent (Bose) quantum states*:

$$|\lambda\rangle_{BF}=|\lambda\rangle_{(j)}=W_{(j)}(b^*, b)|\lambda\rangle. \tag{3.1}$$

One can easily verify that on $|\gamma\rangle_{BF}$ the CAR hold true for all $j\in(j)$, but not for $j\notin(j)$, albeit $[b_i, b_j^*]_+=0$ for all $i\neq j$; compare, e.g., Ref. 17. Suppose now that the coherent product state $|\lambda\rangle$ obeys the restrictions of the theorem of Sec. 2. Then, the *semiclassical Hilbert space IDPS* ($|\lambda\rangle$)*does carry a Fermi system on a subspace*: the CCR algebra possesses the manifestly Fermi states in IDPS($|\lambda\rangle$); compare, e.g., also Ref. 21. In this case, we can say that *both fundamental free bosons and fermions can exist in the same state space on an equal footing*. However, if the restrictions of the theorem are not satisfied by $|\lambda\rangle$, then the only fundamental free field that remains is the Bose one. No fundamental free fermions are allowed. In the case of interacting Fermi systems such a phenomenon would correspond to a "confinement" of their fundamental free excitations (absence of asymptotic free fermions).

Example 1: *Sine-Gordon versus massive Thirring model.*

(1) Both the Mandelstam[22] construction and the Orfanidis[12] observations allow a bosonization of the massive Thirring field in terms of the interacting sine-Gordon field under appropriate constraints. Namely, we can symbolically write an operator identity:

$$\hat{\Psi}=\Psi(\hat{\Phi}),\quad \hat{\Phi}=\Phi(\hat{\phi}_{\mathrm{in}}),\quad (\Box-m^2)\hat{\phi}_{\mathrm{in}}=0, \tag{3.2}$$

so that according to the tree approximation scheme, we should have calculated a coherent state expectation value:

$$\langle\lambda|:\Psi(\Phi)[\hat{\phi}_{\mathrm{in}}]:|\lambda\rangle=\Psi(\Phi)[\phi]=\Psi(\phi), \tag{3.3}$$

where ϕ is a free classical field (the scalar neutral one) of Ref. 6, $\hat{\phi}_{\mathrm{in}}$ in the above is the plane-wave solution of the Klein–Gordon equation in 1 + 1 dimensions, and the normal ordering refers to its (plane-wave solution) creation–annihilation generators. Classically,[12] one knows that if $\Phi=\Phi(\phi)$ is the sine-Gordon 1-soliton, then $\Psi(\phi)$ introduced according to

$$\begin{aligned} \Psi_1=\Psi_1^a &= ia^{-1/2}\left(\tfrac{1}{2}\sin\frac{\Phi_a}{2}\right)^{1/2}\exp(-i\Phi_a/r),\\ \Psi_2=\Psi_2^a &= a^{1/2}\left(\tfrac{1}{2}\sin\frac{\Phi_a}{2}\right)^{1/2}\exp(i\Phi_a/4),\\ \Phi_a &= \Phi(\phi_a), \end{aligned} \tag{3.4}$$

satisfies the massive (mass 1) Thirring model equations of motions, which are the classical (c-number) ones:

$$\begin{aligned} -i\partial_x\Psi_1 &= \tfrac{1}{2}\Psi_2-2\Psi_2^+\Psi_2\Psi_1,\\ i\partial_t\Psi_2 &= \tfrac{1}{2}\Psi_1-2\Psi_1^+\Psi_1\Psi_2. \end{aligned} \tag{3.5}$$

The underlying coherent 1-soliton states were constructed in Ref. 6, and their boson transformation parameters satisfy

$$\frac{1}{2\pi}\int_R\frac{dk}{(k^2+m^2)^{1/2}}\bar{\lambda}(k)\lambda(k)=\int dx[\phi(x)]^2=\infty, \tag{3.6}$$

where $\phi(x)=\phi_a(x)=\exp m\gamma_a x$, $\gamma_a=(a^2+1)/2a$; hence Condition (2) of the theorem of Sec. 2 is manifestly violated. As a consequence *no free fermion is allowed in the 1-soliton Hilbert space IDPS* ($|\lambda\rangle$) *for the sine-Gordon system.*

(2) On the other hand, the spectral solution of the mas-

sive Thirring model given in Ref. 23 proves that the fundamental free field, to be used in the Haag expansions of the model, is the massive Dirac one in 1 + 1 dimensions. Its creation and annihilation operators are required to satisfy the CAR:

$$[b_i(p), b_j^*(q)]_+ = \delta_{ij}\delta(p-q),$$
$$[b_i(p), b_j(q)]_+ = 0 = [b_i^*(p), b_j^*(q)]_+, \quad ij = 1, 2 \tag{3.7}$$

and $\hat{\Psi} = \Psi(\hat{\psi}_{\text{in}}) = \Psi(b^*, b)$. Notice that in 1 + 1 dimensions one can introduce both Bose and Fermi fields on the common Hilbert space domain, without bothering about any spin-statistics problems (this is not the case in 1 + 3 dimensions). A bosonization of $\{b^*, b\}_{i=1,2}$ involves the corresponding Bose degrees of freedom $\{a^*, a\}_{i=1,2}$ (see Refs. 1 and 14) so that

$$b^* = b^*(a^*, a), \quad b = b(a^*, a),$$
$$\hat{\Psi} = \Psi(\hat{\psi}_{\text{in}}) = \Psi(b^*, b) \equiv \Psi(a^*, a) = \Psi(\hat{U}_\mu), \tag{3.8}$$

where $\hat{U}_\mu$ is the massive vector field in 1 + 1 dimensions with no Proca condition imposed.[1] If the construction of semiclassical domains IDPS($|\lambda\rangle$), i.e., of coherent states $|\lambda\rangle$, respects the coexistence of fermions and bosons on a common domain, both $\hat{\psi}_{\text{in}}$ and $\hat{U}_\mu$ are *equally fundamental* and give rise to equivalent Haag series expansions of the quantum fields on the subspace of IDPS($|\lambda\rangle$).

(3) The above picture breaks down if the coherent state $|\lambda\rangle$ does not respect restrictions of the theorem. Then the CAR are no longer satisfied by $\hat{\psi}_{\text{in}}$, and an appropriate (and then unique) fundamental free field is $\hat{U}_\mu$, i.e., the Bose one. In particular, if we impose a Proca condition we arrive at Case (1), where the fundamental free field is a massive neutral scalar $\hat{\phi}_{\text{in}}$, i.e., a boson again.

To summarize: The massive Thirring model always admits a bosonization in terms of $\hat{U}_\mu$. Nevertheless, the notion of a free fundamental fermion can still be saved if coherent states $|\lambda\rangle$ obey the theorem. Otherwise, either $\hat{U}_\mu$ or $\hat{\phi}_{\text{in}}$ plays the role of fundamental field in the model. Consequently, this special Fermi model admits in principle the three different types of Haag expansions—in terms of $\hat{\psi}_{\text{in}}$, $\hat{\phi}_{\text{in}}$, or $\hat{U}_\mu$, *depending on the choice of the state space in H.* Let us once more emphasize that an expansion in terms of $\hat{\psi}_{\text{in}}$ can always be rewritten as an equivalent expansion in terms of $\hat{U}_\mu$. This is obviously a peculiarity of the 1 + 1 dimensional space-time, where the spin-statistics theorem does not apply. The inverse statement in general is not true, because once having specified a domain for $\Psi(\hat{U}_\mu) = \hat{\Psi}$ in H, we may have prohibited the existence of the CAR on it. Then, even having started from an expansion $\hat{\Psi} = \Psi(\hat{\psi}_{\text{in}})$ one must realize that $\hat{\psi}_{\text{in}}$ is no longer a free Fermi field in the conventional sense of the word. It is worth mentioning at this point that quite a variety of spinor models in 1 + 1 dimensions do not meet the requirement of asymptotic completeness; the asymptotic spinor field related to a given interacting spinor field does not exist on the state space of the latter, see, e.g., Ref. 24, but also Refs. 1 and 25–27, where the spinor field asymptotic in 1 + 3 dimensions is considered.

Example 2: QED *in the infrared domain, or the gauge field transcription of the Dirac-photon system.*

The statement of Ref. 1, that the correspondence principle allowing us to relate the classical (*c*-number) and quantum levels of spinor systems in 1 + 1 and 1 + 3 dimensions, involves free Bose systems with unbounded-from-below Hamiltonians. With any element of the spinor field algebra in hand, upon bosonization we can calculate its coherent state expectation value in the tree approximation, thus arriving at the corresponding semiclassical entity.

In 1 + 1 dimensions, the free (asymptotic) fermion can in prinicple coexist with the subsidiary (background) boson on the same state space in H. Then an *interacting fermion can have its free asymptotic Fermi partner.* However, in 1 + 3 dimensions, the spin-statistics theorem must be taken into account. By using a chain of heuristic arguments, we demonstrated in Ref. 1 that a Dirac field, if weakly coupled to the photon field (a nonlinear system of coupled Maxwell and Dirac equations), allows a bosonization in terms of the pure gauge field itself. We use here the Maxwell potential in the Coulomb gauge

$$\hat{\psi} = \psi(\hat{A}_\mu) \leftrightarrow \psi(A_\mu) = (\lambda\,|\colon \psi(\hat{A}_\mu)\colon |\lambda\,), \tag{3.9}$$

where $|\lambda\,)$ is an appropriate coherent photon state, A_μ being a solution of the sourceless Maxwell equations. A really striking peculiarity of (3.9) is that an interacting spin $\frac{1}{2}$ field appears as a nonlinear and nonlocal excitation in the spin 1 free field algebra. This observation can hardly be reconciled with the traditional wisdom about the (perturbative) QED, and its asymptotic problem solution.[25–27] Namely, in the latter case the interacting fields, both Bose and Fermi, have expansions in terms of free Bose and Fermi fields via the Haag series. The Haag series is written in terms of free Fermi and Bose fields commuting among themselves, which is distinct from the bosonization recipe, as discussed in (2.4)–(2.17). The asymptotic infraparticle states of QED found in Ref. 27 require both free bosons and fermions to commute among each other.

In the bosonized case, while using (2.5) and (2.9), we find that, for example,

$$[b_k, a_k^*]_\pm = \exp i\pi \sum_{j=1}^{k-1} \sigma_i^+ \sigma_j^- \, [\sigma_k^-, a_k^*]_\pm, \tag{3.10}$$

hence neither commutation nor anticommutation occurs.

On the other hand, the observation (3.9) is fully consistent with the attempts of Righi and Venturi[28–30] to construct charged fermion fields from extended particlelike solutions in their nonlinear approach to quantum electrodynamics. An example of the fully bosonized interacting spinor field which satisfies the CAR, and does not at all commute with the electromagnetic field, is given in Ref. 29. An analogy with the previously considered sine-Gordon/Thirring case appears to be striking.

Obviously the field $\hat{A}_\mu$ is not free, but its Haag series do apparently fit in our framework. Hence a construction of the appropriate coherent photon states is quite in order. In the case of the relativistic field theory, we expect that the presence of free fermions should be forbidden in the fully bosonized Fermi system. Hence one should look for coherent states which do not conflict with this theorem. We still cannot propose a final solution to this problem; let us, however, indicate that the coherent photon states invented by

Chung[25] in the conventional approach to the QED do not allow the existence of free fermions on any subspace of the semiclassical (photon) Hilbert space. The coherent states of interest read (a single electron case)

$$|\lambda\rangle_p = U_\lambda^* |0\rangle:$$

$$= \exp\left\{\frac{e}{(2\pi)^{3/2}} \int \sum_{i=1,2} \left[F^i(k,p) a_i^*(\mathbf{k}) - \overset{*}{F}{}^i(k,p) a_i(\mathbf{k})\right] \frac{d^3k}{(2k_0)^{1/2}}\right\} |0\rangle, \tag{3.11}$$

where

$$F^i(k,p) = \frac{p\cdot\epsilon^i}{p\cdot k}\,\phi(k,p) \tag{3.12}$$

and p, k, ϵ^i are the four-vectors, $p\cdot k$ being the corresponding scalar product formula. Here p stands for the four-momentum of the electron to which the state $|\lambda\rangle_p$ is assigned. The function $\phi(k,p)$ equals 1 in the vicinity of $k=0$. By also taking into account a factor $1/(2k_0)^{1/2}$, $k_0 = |\mathbf{k}|$, one easily verifies that the coherent photon state $|\lambda\rangle_p$ violates Condition (2) of the main theorem due to the singularity of $|F^i(k,p)|$ at $k=0$. Let us mention that in analogy to $|\lambda\rangle_p$, the soliton states of the massive Thirring–sine-Gordon example did exhibit a manifest parametrization $|\lambda\rangle = |\lambda\rangle_a$ in terms of the 1-soliton parameter a; compare, e.g., (3.6). Because the 1-soliton total momentum reads $k = 8m(|a|^2 - 1)/2|a|$, $|\lambda\rangle_a$ provides us with a momentum parametrization as well.

ACKNOWLEDGMENTS

I would like to express my warm thanks to Professors H. M. Fried and A. Jevicki for arranging my visit to Brown University, where the present paper was completed.

[1]P. Garbaczewski, "Quantization of spinor fields II. Meaning of bosonization in 1 + 1 and 1 + 3 dimensions," J. Math. Phys. **23**, 442 (1982).
[2]N. N. Bogolubov, A. A. Logunov, and I. I. Todorov, *Axiomatic quantum field theory* (Nauka, Moscow, 1969), in Russian.
[3]H. B. Thacker, Rev. Mod. Phys. **53**, 253 (1981).
[4]J. Honerkamp, "Quantization of exactly integrable field theoretic models," in *Les Houches Colloquim Proceedings* (Springer, Berlin, 1980).
[5]P. Garbaczewski, J. Math. Phys. **22**, 574 (1981); see also **21**, 2670 (1980).
[6]P. Garbaczewski, "On quantum solitons...II. Fermion-boson reciprocity and the classical versus quantum problem for the sine-Gordon system," J. Math. Phys. **22**, 1272 (1981).
[7]Y. H. Ichikawa, N. Yajima, and K. Takano, Prog. Theor. Phys. **55**, 1723 (1976).
[8]H. Matsumoto, P. Sodano, and H. Umezawa, Phys. Rev. D **19**, 511 (1979).
[9]M. S. Swanson, Phys. Rev. D **22**, 1497 (1980).
[10]G. Oberlechner, M. Umezawa, and Ch. Zenses, Lett. Nuovo Cimento **23**, 697 (1978).
[11]H. Matsumoto, N. J. Papastamatiou, H. Umezawa, and M. Umezawa, Phys. Rev. D **23**, 1339 (1981).
[12]S. J. Orfanidis, Phys. Rev. D **14**, 472 (1976).
[13]K. Ishikawa, Prog. Theor. Phys. **55**, 588 (1976).
[14]P. Garbaczewski, Commun. Math. Phys. **43**, 131 (1975).
[15]P. Garbaczewski, Phys. Rep. **36C**, 65 (1978).
[16]P. Garbaczewski, J. Math. Phys. **19**, 642 (1978).
[17]P. Garbaczewski and Z. Popowicz, Rep. Math. Phys. **11**, 69 (1977).
[18]G. G. Emch, *Algebraic methods in statistical mechanics and quantum field theory* (Wiley, New York, 1972).
[19]P. Garbaczewski, "Representations of the CAR generated by representations of the CCR. Non-Fock case," Wroclaw Univ. prepring, 1981.
[20]T. W. B. Kibble, J. Math. Phys. **9**, 315 (1968).
[21]R. F. Streater and I. F. Wilde, Nucl. Phys. B **24**, 561 (1970).
[22]S. Mandelstam, Phys. Rev. D **11**, 3026 (1975).
[23]H. Bergknoff and H. B. Thacker, Phys. Rev. D **19**, 3666 (1979).
[24]N. Nakanishi, Prog. Theor. Phys. **59**, 607 (1978).
[25]V. Chung, Phys. Rev. **140**, B1110 (1965).
[26]J. K. Storrow, Nuovo Cimento **54A**, 15 (1978).
[27]P. P. Kulish and L. D. Faddeev, Teor. Math. Fiz. **4**, 153 (1970).
[28]R. Righi and G. Venturi, Nuovo Cimento A **52**, 166 (1979).
[29]R. Righi and G. Venturi, Nuovo Cimento A **52**, 511 (1979).
[30]R. Righi and G. Venturi, Lett. Nuovo Cimento **31**, 487 (1981).

Verification of the global Markov property in some class of strongly coupled exponential interactions

Roman Gielerak[a)]
Forschungsinstitut für Mathematik ETH, Zürich, Switzerland and University of Wrocław, Wrocław, Poland

(Received 29 July 1981; accepted for publication 23 October 1981)

We verify the global Markov property in some class of strongly coupled exponential interactions in two-dimensional space-time. To obtain this result we apply the Albeverio and Høegh-Krohn strategy. The basic ingredients we use in order to employ this strategy are the Fortuin–Kastelyn–Ginibre correlation inequalities.

PACS numbers: 11.10. – z, 02.50.Ga

I. INTRODUCTION

Let $d\mu_0$ be the free Euclidean field measure on $S'(R^2)$, i.e., the Gaussian with mean zero and covariance $(-\Delta + m_0^2)^{-1}$, where Δ is the two-dimensional Laplacian and $m_0 > 0$ is a free field mass. Let $\{U_\Lambda\}$ be a local space-time cutoff interaction, i.e, the map $R^2 \supset \Lambda \to U_\Lambda$ is an additive functional of the free, Euclidean field such that $U_\Lambda \in L^2(d\mu_0)$ and $e^{-U_\Lambda} \in L^p(d\mu_0)$ for any $1 \leqslant p < \infty$. By Σ_Λ we denote the σ-algebras generated by the fields with support within Λ. The measures considered in the Euclidean field theory are of the form

$$d\mu_\Lambda(\varphi) = [E_0(e^{-U_\Lambda})]^{-1} \cdot e^{-U_\Lambda} \cdot d\mu_0(\varphi), \tag{1.1}$$

$$E_0(\cdot) \equiv \int d\mu_0(\varphi)(\cdot) \equiv \langle(\cdot)\rangle_0. \tag{1.2}$$

Of special interest are the infinite-volume limits of the quantities like (1.1).

A measure $d\mu$ on $S'(\mathbb{R}^2)$, being a Gibbs measure corresponding to the multiplicative functional $\{e^{-U_\Lambda}\}$, is any measure which is locally absolutely continuous with respect to $d\mu_0$ and such that the associated conditional expectation values of functions measurable within Λ with respect to the σ algebra Σ_Λ coincides with those computed with $d\mu_\Lambda$ instead of $d\mu$:

$$E_\mu^{\partial\Lambda}[f] \equiv E\mu(f|\partial\Lambda) \equiv E\mu_\Lambda(f|\partial\Lambda). \tag{1.3}$$

A Gibbs measure $d\mu$ is *pure* iff there is no other Gibbs measure corresponding to the same multiplicative functional which is absolutely continuous with respect to $d\mu$.

We recall now what the global Markov property (GMP) is. Let C be any piecewise-C^1 curve such that $\mathbb{R}^2 \setminus C$ consists of two components Ω^+ and Ω^-. Let f_+ and f_- be any bounded functions of the field φ_μ associated with the Gibbs measure $d\mu$ which are measurable in Ω^+ and Ω^-, respectively. A Gibbs measure $d\mu$ has a global Markov property iff for any C, f_+, f_- as above

$$E_\mu(f_+ f_-|C) = E_\mu(f_+|C)E_\mu(f_-|C). \tag{1.4}$$

In many cases considered in the Euclidean field theory it is known that the property (1.4) holds for every *bounded*, piecewise C^1 curve. It is called the local Markov property (LMP).

Only in the case of the weakly coupled trigonometric interactions has the global Markov property been verified recently by Albeverio and Høegh-Krohn (AHK).[1] The ingenious paper by Albeverio and Høegh-Krohn includes, among other interesting things, a general strategy for the proof of GMP in a more general context than the sine–Gordon[2] model. Let us define

$$\psi_\eta^{\partial\Lambda}(x) = \int_{\partial\Lambda} P^{\partial\Lambda}(x,z)\eta(z)dz, \tag{1.5}$$

where $P^{\partial\Lambda}$ is the Poisson kernel associated with the following Dirichlet problem:

$$\begin{aligned} &(-\Delta + m_0^2)\psi_\eta^{\partial\Lambda}(x) = 0 \text{ for } x \in \mathbb{R}^2 \setminus \partial\Lambda, \\ &\psi_\eta^{\partial\Lambda}(x) = \eta(x) \text{ for } x \in \partial\Lambda \end{aligned} \tag{1.6}$$

for any $\partial\Lambda$ which is piecewise C^1 in $\mathbb{R}^2$.

Now let $d\mu_\eta^C$ be a Gibbs measure μ conditioned by $\xi = \eta$ on a given piecewise-C^1 curve C in $\mathbb{R}^2$. Then we have the following strategy for proof of GMP(AHK strategy):

If $d\mu_\eta^C$ is a *pure* Gibbs measure for the interaction $U_\Lambda(\varphi + \psi_\eta^C)$ for almost every η with respect to $d\mu$ and $d\mu_\Lambda$ has the global Markov property then the Gibbs measure $d\mu$ also has the global Markov property (see Ref. 1).

In this paper we employ this strategy for the class of exponential interactions

$$U_\Lambda(\varphi) = \lambda \int_\Lambda dx \int d\nu(\alpha) :e^{\alpha\varphi}:(x), \tag{1.7}$$

where $\lambda \in \mathbb{R}_+$, $d\nu(\alpha)$ is some bounded measure with support on $(0,\alpha^*)$, $\alpha^* < 2\pi^{1/2}$, and α^* is sufficiently small. In this paper we will not give the precise bound on α^*.

The Gibbs measures $d\mu$ associated with (1.7) are the so-called exponential interaction. They were discussed previously in Refs. 2–4. Our main result is the following theorem:

Theorem 1.1: For α^* sufficiently small the exponential interactions defined by (1.7) have the global Markov property.

Consequences of the GMP have been discussed in several papers; see, e.g., Refs. 5–10. It can be easily shown that these discussions are applicable to the models (1.7). We summarize some standard consequences of Theorem 1.1.

Theorem 1.2: For α^* sufficiently small the quantum fields corresponding to the exponential interactions defined by (1.7) have the following properties:

(i) They are canonical fields in the sense of Ref. 8.

[a)] On leave of absence from the Institute of Theoretical Physics, University of Wrocław, Poland. Supported by the Forschungsinstitut für Mathematik, E. T. H. Zürich.

 0022-2488/83/020347-09$02.50

(ii) Time-zero fields generate the corresponding Hilbert spaces.

(iii) Physical Hamiltonians corresponding to (1.7) are the second order *elliptic* variational operators in the sense of Ref. 7.

It is worth noticing that the AHK strategy of the proof of GMP has been applied in Ref. 11 to a large class of discrete spin systems on a lattice in the region of coupling where the Dobrushin uniqueness theorem can be applied. A similar proof to that in Ref. 11 was given by Föllmer for the pure phases of the Ising model.[12] A very attractive alternative strategy to the AHK strategy for the proof of GMP has been proposed and applied to some lattice spin systems and the continuum Widom–Rowlinson model by Goldstein.[13] Some uniqueness results for the Dobrushin–Lanford–Ruelle equations for the continuous spin systems have been established in Refs. 14 and 15.

We close this Introduction with some comment on the organization of this paper. Section 2 contains some preparations for the proof of GMP, which is given in Sec. 3. The rest of this paper is of a technical character. In Sec. 4 we prove Fortuin–Kastelyn–Ginibre (FKG) correlation inequalities for the conditioned measures. Appendixes A and B contain some technicalities necessary to complete the proof of the crucial Theorem 2.4 in which we prove that the Gibbs measures corresponding to the conditioned exponential interactions (1.7) are the pure Gibbs measures almost surely with respect to $d\mu$. Throughout this paper we always assume for simplicity

$$d\nu(\alpha) = \delta(\alpha - \alpha^*).$$

2. Preparations for the proof of GMP

Our proof is modeled on an original paper by Albeverio and Høegh-Krohn,[1] where they proved the global Markov property (GMP) for the case of the weakly coupled trigonometric interactions. One of the basic ingredients of their proof is that the following local boundness properties of the solution of the Dirichlet problem (1.6) with the boundary conditions η are chosen randomly from the space $\{S'(\mathbb{R}^2), \Sigma, d\mu\}$, where $d\mu$ is the infinite volume measure corresponding to the exponential interactions (1.7) and Σ is a σ algebra generated by the field φ associated with $d\mu$.

Proposition 2.1 (Local Boundness Properties = LBP)[1]:

1. Let $\varphi(x)$ be a field associated with the exponential interactions (1.7). Let C and C_0 be two piecewise C^1 curves in $\mathbb{R}^2$ and let

$$\psi_\varphi^{C,C_0}(x) = \psi_\varphi^{C\cup C_0}(x) - \psi_\varphi^{C_0}(x) \text{ for } x \notin C \cup C_0,$$

where $\psi_g^C(x)$ is the solution of $(-\Delta + m_0^2)\psi_g^C(x) = 0$ in $\mathbb{R}^2 \setminus C$ and $\psi_g^C(x) = g(x)$ on C. Then there exists a constant a such that for almost every φ with respect to $d\mu$ and $x \in \mathbb{R}^2 \setminus C \cup C_0$ with $d(x,C) \geqslant 1$ we have, with respect to $d\mu$

$$\int |\psi_\varphi^{C,C_0}(x)|^2 d\mu(\varphi) \leqslant a e^{-m_0' d(x,C)}, \quad m_0' < m_0. \tag{2.1}$$

Moreover if Λ is a compact subset of $\mathbb{R}^2 \setminus C_0 \cup C$ such that $d(\Lambda, C) \geqslant 1$, then we have

$$\int_\Lambda |\psi_\varphi^{C,C_0}(x)|^2 dx \leqslant a e^{-m_0 d(\Lambda,C)}. \tag{2.2}$$

2. Now let C_0 be any fixed C^1 curve and let C_n be any sequence of piecewise-C^1 curves in $\mathbb{R}^2$ which tends to infinity in the sense that $d(0,C_n) \to \infty$. Let Λ_n be any sequence of bounded sets in $\mathbb{R}^2$ such that $\Lambda_n \subset \mathbb{R}^2 \setminus C_0 \cup C_n)$ and $\operatorname{dist}(\Lambda_n, C_n) \to \infty$ as $n \to \infty$. Then for any $\alpha < m_0$ there is a subsequence n' such that

$$e^{(\alpha/2)d(\Lambda_{n'}, C_{n'})} \sup_{x \in \Lambda_{n'}} |\psi_\varphi^{C_0 \cup C_{n'}}(x) - \psi_\varphi^{C_0}(x)| \underset{n'}{\to} 0$$

for almost every φ with respect to $d\mu$.

Let C^0 be an unbounded connected piecewise-C^1 curve such that $\mathbb{R}^2 \setminus C^0$ consists of two components Ω_+ and Ω_-. Let Λ_n be any sequence of bounded sets in $\mathbb{R}^2$ which tends to $\mathbb{R}^2$ in the sense to be specified below. Let us define $\tilde{\partial}\Lambda_n = \partial\Lambda_n \setminus C^0$. We also assume that for any $n \in N$, $\partial\Lambda_n \cap C^0$ consists of at most a finite number of points. If f is Σ_{Λ_n} measureble then we have

$$E_\mu^{C^0 \cup \partial\tilde{\Lambda}_n}[f](\eta) = \int_{S'(\mathbb{R}^2)} f(\varphi + \psi_\eta^{C^0 \cup \partial\tilde{\Lambda}_n}) d\mu_{\lambda,\eta,\Lambda_n}^{C^0,\partial\tilde{\Lambda}_n}(\varphi), \tag{2.3}$$

where

$$\begin{aligned} &d\mu_{\lambda,\eta,\Lambda_n}^{C^0,\partial\tilde{\Lambda}_n}(\varphi) \\ &\quad \equiv (Z_{\lambda,\eta,\Lambda_n}^{C^0 \cup \partial\tilde{\Lambda}_n})^{-1} \\ &\qquad \times \exp\left[-\lambda \int d\nu(\alpha) \int_{\Lambda_n} :\exp[\alpha(\varphi + \psi_\eta^{C^0 \cup \partial\tilde{\Lambda}_n})]:(x) dx \right] \\ &\qquad \cdot d\mu_0^{C_0 \cup \partial\tilde{\Lambda}_n}(\varphi), \end{aligned} \tag{2.4}$$

where $d\mu_0^{C_0 \cup \partial\tilde{\Lambda}_n}(\varphi)$ is the free field Gaussian measure with Dirichlet boundary condition on $C_0 \cup \partial\tilde{\Lambda}_n$ and

$$\begin{aligned} Z_{\lambda,\eta,\Lambda_n}^{C_0 \cup \partial\tilde{\Lambda}_n} = \int d\mu_0^{C_0 \cup \partial\tilde{\Lambda}_n}(\varphi) \exp\Big[-\lambda \int d\nu(\alpha) \\ \times \int_{\Lambda_n} :\exp[\alpha(\varphi + \psi_\eta^{C_0 \cup \partial\tilde{\Lambda}_n})]:(x) dx \Big]. \end{aligned} \tag{2.5}$$

Because φ and $\psi_\eta^{C_0 \cup \partial\tilde{\Lambda}_n}$ are independent Gaussian processes we can write

$$:\exp[\alpha(\varphi + \psi^{C^0 \cup \partial\tilde{\Lambda}_n})]:(x) = :e^{\alpha\varphi}:(x) :: \exp(\alpha\psi_\eta^{C_0 \cup \partial\tilde{\Lambda}_n}):(x), \tag{2.6}$$

where

$$\begin{aligned} &:\exp(\alpha\psi_\eta^{C_0 \cup \partial\tilde{\Lambda}_n}):(x) \\ &\quad = \exp\left[-(\alpha^2/2) K_{(x,x)}^{C_0 \cup \partial\tilde{\Lambda}_n} \right] \exp(\alpha\psi_\eta^{C_0 \cup \partial\tilde{\Lambda}_n})(x), \end{aligned} \tag{2.7}$$

$$K_{(x,x)}^{C_0 \cup \partial\tilde{\Lambda}_n} \equiv (-\Delta + m_0^2)^{-1}(x,x) - (-\Delta_{C_0 \cup \partial\tilde{\Lambda}_n} + m_0^2)^{-1}(x,x), \tag{2.8}$$

where Δ_C is a Laplacian with Dirichlet boundary condition on C. We summarize the crucial properties of the measure $d\mu_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}_n}(\varphi)$ in the following proposition.

Proposition 2.2:

1. For almost every η with respect to $d\mu$, $\Lambda_n \subset \mathbb{R}^2$ bounded, the measures $d\mu_{\lambda,\eta,\Lambda_n}^{C_0,\partial\tilde{\Lambda}_n}$ are well defined probabilistic measures on $S'(\mathbb{R}^2)$.

2. For almost every η with respect to $d\mu$ the measures $d\mu_{\lambda,\eta,\Lambda_n}^{C_0,\partial\tilde{\Lambda}_n}(\varphi)$ fulfill Fortuin–Kastelyn–Ginibre correlation inequalities.

We prove this proposition in Sec. 3. Now we note the following important consequence of Proposition 2.2.

Corollary 2.3: 1. For almost every η with respect to $d\mu$ the unique thermodynamic limit

$$w\text{-}\lim_{\Lambda\uparrow\mathbb{R}^2} d\mu_{\lambda,\eta,\Lambda}^{C_0\cup\partial\tilde{\Lambda}} \equiv d\mu_{\lambda,\eta}^{C_0}(\varphi) \tag{2.9}$$

as a weak limit of measures exists.

2. For almost every η with respect to $d\mu$, measures $d\mu_{\lambda,\eta}^{C_0}$ have the exponential cluster property uniformly in λ, η, and C_0.

Proof of Corollary 2.3:

ad1. By applying the FKG correlation inequalities we have that for $f\geqslant 0$

$$0\leqslant \int e^{\varphi(f)} d\mu_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}} \leqslant \int e^{\varphi(f)} d\mu_0^{C_0\cup\partial\tilde{\Lambda}}(\varphi)$$

(by conditioning)

$$\leqslant \int e^{\varphi(f)} d\mu_0(\varphi), \tag{2.10}$$

and moreover $\int e^{\varphi(f)} d\mu_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}}$ is monotonically decreasing in Λ.

ad2. Let us consider the λ dependence of the truncated two-point function of the measure $d\mu_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}}$,

$$\begin{aligned}
\frac{d}{d\lambda}\Big(&\int \varphi(x)\varphi(y) d\mu_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}} \\
&- \int \varphi(x) d\mu_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}} \cdot \int \varphi(y) d\mu_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}}\Big) \\
&= -\int dz :\exp(\alpha)\Psi_\eta^{C_0,\partial\tilde{\Lambda}}:(z)\langle\varphi(x)\varphi(y);:e^{\alpha\varphi}:(z)\rangle_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}} \\
&\quad + \int dz :\exp(\alpha\Psi_\eta^{C_0,\partial\tilde{\Lambda}}):(z)\langle\varphi(x);:e^{\alpha\varphi}:(z)\rangle_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}} \\
&\quad \times\langle\varphi(y)\rangle_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}} \\
&\quad + \int dz :\exp(\alpha\Psi_\eta^{C_0,\partial\tilde{\Lambda}}):(z)\cdot\langle\varphi(y);:e^{\alpha\varphi}:(z)\rangle_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}} \\
&\quad \times\langle\varphi(x)\rangle_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}}.
\end{aligned} \tag{2.11}$$

By means of the Ginibre and FKG correlation inequalities we have

$$\langle\varphi(x)\varphi(y);:e^{\alpha\varphi}:(z)\rangle_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}} \geqslant 0, \tag{2.12}$$

$$\langle\varphi(x);:e^{\alpha\varphi}:(z)\rangle_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}} \geqslant 0. \tag{2.13}$$

Applying the integration by parts formula on function space we have

$$\begin{aligned}
&\langle\varphi(x)\rangle_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}} \\
&\quad = -\lambda\int_\Lambda dz :\exp(\alpha\Psi_\eta^{C_0,\partial\tilde{\Lambda}}):(z) S_{(x,z)}^{C_0\cup\partial\tilde{\Lambda}} \langle:e^{\alpha\varphi}:(z)\rangle_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}} < 0.
\end{aligned} \tag{2.14}$$

So finally we have

$$\frac{d}{d\lambda}\langle\varphi(x);\varphi(y)\rangle_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}} < 0, \tag{2.15}$$

from which follows

$$0\leqslant\langle\varphi(x);\varphi(y)\rangle_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}} \leqslant \int \varphi(x)\varphi(y)\, d\mu_0^{C_0\cup\partial\Lambda}(\varphi)$$

(by conditioning)

$$\leqslant \int \varphi(x)\varphi(y) d\mu_0(\varphi). \tag{2.16}$$

We conclude that for almost every η with respect to $d\mu$ the two-point moment of the measures $d\mu_{\lambda,\eta,\Lambda}^{\partial\tilde{\Lambda},C_0}$ clusters exponentially fast with decay rate bounded from below by the free field mass m_0. This cluster property holds uniformly in $\Lambda, C_0, \lambda, \eta$, for almost every η with respect to $d\mu$. Another application of Proposition 2.2 and the Lebowitz–Simon[16] theorem concludes the proof of statement 2. Q.E.D.

The crucial step in the verification of the GMP is the following theorem.

Theorem 2.4: Almost surely with respect to $d\mu$

$$\text{w-}\lim_{\Lambda\nearrow\mathbb{R}^2} d\mu_{\lambda,\eta,\Lambda}^{C_0,\partial\Lambda}(\varphi) = \text{w-}\lim_{\Lambda\nearrow\mathbb{R}^2} d\mu_{\lambda,\eta,\Lambda}^{C_0}(\varphi). \tag{2.17}$$

Proof: Let us take without loss of generality $S(\mathbb{R}^2)\epsilon f \geqslant 0$ with compact support. Let us consider the following interpolation between $d\mu_{\lambda,\eta}^{C_0,\partial\Lambda}(\varphi)$ and $d\mu_{\lambda,\eta}^{C_0}(\varphi)$. Let Λ_n be some sequence of regular sets in $\mathbb{R}^2$ such that $\Lambda_n\uparrow\mathbb{R}^2$ by inclusion. Then we define

$$\begin{aligned}
d\mu_{\lambda,\eta,\Lambda_n}^{C_0,\partial\tilde{\Lambda}_n,\sigma}(\varphi) = &\left[Z_{\lambda,\eta,\Lambda_n}^{C_0,\partial\tilde{\Lambda}_n,\sigma}\right]^{-1} \\
&\times\exp\Big[-\lambda\int_{\Lambda_n} :e^{\alpha\varphi}:(x):\Big] \\
&\times\exp\left[\alpha\Psi_\eta^{C_0}(x) + \sigma\Psi_\eta^n\right]:(x)\Big] \\
&\times d\mu_0^{C_0\cup\partial\tilde{\Lambda}_n}(\varphi)
\end{aligned} \tag{2.18}$$

for $\sigma\in[0,1]$ and

$$\begin{aligned}
Z_{\lambda,\eta,\Lambda_n}^{C_0,\partial\tilde{\Lambda}_n,\sigma} \equiv \int d\mu_0^{C_0\cup\partial\tilde{\Lambda}_n}(\varphi)\exp\Big[&-\lambda\int_{\Lambda_n} :e^{\alpha\varphi}:(x) \\
&:\exp\left[\alpha(\Psi_\eta^{C_0} + \sigma\Psi_\eta^n)\right]:(x)\Big].
\end{aligned} \tag{2.19}$$

Here we are using the following notation:

$$:\exp(\alpha\Psi_\eta^{C_0} + \alpha\sigma\Psi_\eta^n):(x) = :\exp(\alpha\Psi_\eta^{C_0}:(x):e^{\alpha\sigma\Psi_\eta^n}:(x), \tag{2.20}$$

where

$$:e^{\alpha\sigma\Psi_\eta^n}: = \exp\left(-\frac{\alpha^2\sigma^2}{2}K_{(x,x)}^{C_0,\partial\tilde{\Lambda}_n}\right)\exp\left[\alpha\sigma\Psi_\eta^n(x)\right], \tag{2.21}$$

$$\Psi_\eta^n(x) \equiv \Psi_\eta^{C_0,\partial\tilde{\Lambda}_n}(x) - \Psi_\eta^{C_0}(x), \tag{2.22}$$

$$\begin{aligned}
K_{(x,x)}^{C_0,\partial\tilde{\Lambda}_n} &= \int d\mu_0(\varphi)(\Psi_\eta^n)^2(x) \\
&= K_{(x,x)}^{C_0\cup\partial\tilde{\Lambda}_n} - 2\int_{C_0\cup\partial\tilde{\Lambda}_n} dz_1 \int_{C_0} dz_2 P_{(x,z_1)}^{C_0\cup\partial\tilde{\Lambda}_n} \\
&\quad\times P_{(x,z_2)}^{C_0} S_0^2(z_1,z_2) + K_{(x,x)}^{C_0},
\end{aligned} \tag{2.23}$$

$$K^C(x,x) = (-\Delta + m_0^2)^{-1}(x,x) - (-\Delta_C + m_0^2)^{-1}(x,x). \tag{2.24}$$

With this notation we have the fundamental theorem of

calculus

$$\int e^{\varphi(f)} d\mu_{\lambda,\eta,\Lambda_n(\varphi)}^{C_0,\partial\tilde{\Lambda}_n,\sigma=1} - \int e^{\varphi(f)} d\mu_{\lambda,\eta,\Lambda_n(\varphi)}^{C_0,\partial\tilde{\Lambda}_n,\sigma=0}$$

$$= -\lambda \int_{\Lambda_n} dx \int_0^1 d\sigma (-\sigma\alpha^2 K_{(x,x)}^{C_0,\partial\tilde{\Lambda}_n} + \alpha \Psi_\eta^n(x)) : e^{\alpha\sigma\Psi_\eta^n} : (x) \times F_{\lambda,\eta,\Lambda_n}^{C_0,\partial\tilde{\Lambda}_n,\sigma}(f,x), \tag{2.25}$$

where we have defined

$$F_{\lambda,\eta,\Lambda_n}^{C_0,\partial\tilde{\Lambda}_n,\sigma}(f,x) = \langle e^{\varphi(f)} ; : e^{\alpha\varphi} : (x) \rangle_{\lambda,\eta,\Lambda_n}^{C_0,\partial\tilde{\Lambda}_n,\sigma}. \tag{2.26}$$

Now let Y_n be another sequence of sets in $\mathbb{R}^2$ such that $Y_n \uparrow \mathbb{R}^2$ as $n \to \infty$, $Y_n \subset \Lambda_n$, $\lim_{n\to\infty} d(Y_n, \partial\Lambda_n) = \infty$. Furthermore, we will impose some other geometrical conditions on Y_n which will become clear by consideration of Appendix A. Let us now define

$$J_{n,\eta}^1 \equiv \alpha^2 \int_{Y_n} dx \int_0^1 d\sigma\, \sigma K_{(x,x)}^{C_0,\partial\tilde{\Lambda}_n} : e^{\alpha\sigma\Psi_\eta^n} : (x) : \exp(\alpha \Psi_\eta^{C_0}) : (x) \times F_{\lambda,\eta,\Lambda_n(f;x)}^{C_0,\partial\tilde{\Lambda}_n,\sigma}, \tag{2.27}$$

$$J_{n,\eta}^2 \equiv \alpha^2 \int_{\Lambda_n \setminus Y_n} dx \int_0^1 d\sigma\, \sigma K_{(x,x)}^{C_0,\partial\tilde{\Lambda}_n} : \exp(\alpha\sigma\Psi_\eta^{C_0}) : (x) : e^{\alpha\sigma\Psi_\eta^n} : (x) \times F_{\lambda,\eta,\Lambda_n(f;x)}^{C_0,\partial\tilde{\Lambda}_n,\sigma}, \tag{2.28}$$

$$J_{n,\eta}^3 \equiv \alpha \int_{Y_n} dx \int_0^1 d\sigma\, \sigma : \exp(\alpha\Psi_\eta^{C_0} : (x) \Psi_\eta^n(x) : e^{\alpha\sigma\Psi_\eta^n} : (x) \times F_{\lambda,\eta,\Lambda_n(f;x)}^{C_0,\partial\tilde{\Lambda}_n,\sigma}, \tag{2.29}$$

$$J_{n,\eta}^4 \equiv \alpha \int_{\Lambda_n \setminus Y_n} dx \int_0^1 d\sigma\, \sigma : \exp(\alpha\Psi_\eta^{C_0}) : (x) \Psi_\eta^n(x) : e^{\alpha\sigma\Psi_\eta^n} : (x) \times F_{\lambda,\eta,\Lambda_n(f;x)}^{C_0,\partial\tilde{\Lambda}_n,\sigma}. \tag{2.30}$$

In the four lemmas contained in Appendix A we prove that almost surely with respect to $d\mu$ there exists a subsequence $\{n'\} \subset \{n\}$ such that

$$\lim_{n'\to\infty} J_{n',\eta}^i = 0 \quad \text{for } i = 1,2,3,4. \tag{2.31}$$

To control $\lim_{n'\to\infty} J_{n',\eta}^i = 0$ for $i = 1,3$ we use LBP summarized in Proposition 2.1 and to ensure $\lim_{n'\to\infty} J_{n',\eta}^i$ for $i = 2,4$ we use the uniform exponential cluster property given by Corollary 2.3. So we have proved that

$$\text{w-lim}_{\Lambda_n' \nearrow \mathbb{R}^2} d\mu_{\lambda,\eta,\Lambda_{n'}}^{C_0,\partial\tilde{\Lambda}_{n'},\sigma=1} = \text{w-lim}_{\Lambda_n \nearrow \mathbb{R}^2} d\mu_{\lambda,\eta,\Lambda_{n'}}^{C_0,\partial\tilde{\Lambda}_{n'},\sigma=0}.$$

By the 2ϵ argument it remains only to prove

$$\text{w-lim}_{\Lambda_{n'} \nearrow \mathbb{R}^2} d\mu_{\lambda,\eta,\Lambda_{n'}}^{C_0,\partial\tilde{\Lambda}_n} = \text{w-lim}_{\Lambda_{n'} \nearrow \mathbb{R}^2} d\mu_{\lambda,\eta,\Lambda_{n'}}^{C_0}, \tag{2.32}$$

i.e., the independence of the half-Dirichlet boundary condition in the thermodynamic limit of the conditioned measure $d\mu_{\lambda,\eta,\Lambda_{n'}}^{C_0,\sigma=0}$. Appendix B contains a detailed proof of this missing statement. Q.E.D.

Let us note finally the following consequence of the proof of Theorem 2.4.

Corollary 2.5: Exponential interactions corresponding to the interactions (1.7) are pure Gibbs measures.

Remark: Using the kind of arguments (essentially due to Fröhlich and Simon[17] given in Appendix B we are able to prove the independence of certain classical boundary conditions for a class of exponential interactions described by

$$U_\Lambda(\varphi) = \lambda \int_\Lambda dx \int d\nu(\alpha) : e^{\alpha\varphi} : (x), \tag{2.33}$$

where now $d\nu(\alpha)$ is an even, bounded measure supported on $(-2/\sqrt{\pi}, 2/\sqrt{\pi})$. However, we are unable to extend Corollary 2.5 to this case, because conditioning destroys the crucial Griffiths–Kelly–Sherman correlation inequalities.

3. PROOF OF GMP

This section is included only for the reader's convenience. Having proven Theorem 2.4 we follow exactly the original arguments given by Albeverio and Høegh-Krohn.

Proof of Theorem 1.1: From the definition of $d\mu_{\lambda,\eta,\Lambda}^{C_0}$ we see that it depends on η only through the field $\psi_\eta^{C_0}(x)$, which be definition is Σ_{C_0} measurable. Hence the functions

$$\eta \to \int e^{\varphi(f)} d\mu_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}}(\varphi) \tag{3.1}$$

are all Σ_{C_0} measurable. On the other hand we have

$$\int e^{\varphi(f)} d\mu_{\lambda,\eta,\Lambda}^{C_0,\partial\tilde{\Lambda}_n,\sigma=1}(\varphi)$$

$$= \exp[-\Psi_\eta^{C_0}(f)] \exp[-\Psi_\eta^n(f)] E_\mu^{C_0\cup\partial\tilde{\Lambda}_n}[e^{\varphi(f)}](\eta)$$

(by the local Markov property) $\tag{3.2}$

$$= \exp[-\Psi_\eta^{C_0}(f)] \exp[-\Psi_\eta^n(f)] \times E^{C_0\cup\mathbb{R}^2\setminus\Lambda_n}[e^{\varphi(f)}](\eta),$$

assuming n is so large that $\text{supp} f \subset \Lambda_n$. But the conditional expectation $E^{C_0\cup\mathbb{R}^2\setminus\Lambda_n}[\cdot]$ is by definition a martingale in n so by the martingale convergence theorem we have that $\lim_{n\to\infty} E_\mu^{C_0\cup\mathbb{R}^2\setminus\Lambda_n}[\cdot](\eta)$ exists for almost every η with respect to $d\mu$. Hence it follows that $\lim_{n\to\infty} \int e^{\varphi(f)} d\mu_{\lambda,\eta,\Lambda_n}^{C^0,\partial\tilde{\Lambda}_n,\sigma=1}(\varphi)$ converges for almost all η. In particular we obtain by the martingale convergence theorem that it is enough to investigate the limit $\lim_{n\to\infty} \int e^{\varphi(f)} d\mu_{\lambda,\eta,\Lambda_n}^{C^0,\partial\tilde{\Lambda}_n,\sigma=1}(\varphi)$ by choosing subsequence as we did in Theorem 2.4. This limit is equal to

$$E_\mu^{\Sigma_{C_0}^\infty} \times [\exp[(\varphi - \Psi_\eta^{C_0})(f)]](\eta),$$

where $\Sigma_{C_0}^\infty \equiv \cap_n [\Sigma_{C_0\cup(\mathbb{R}^2\setminus\Lambda_n)}]$.

Taking into account that

$$E_\mu^{\Sigma_{C_0}^\infty}[\exp\{(\varphi - \Psi_\eta^{C_0})(f)\}](\eta) = \exp[\Psi_\eta^{C_0}(f)] E_\mu^{\Sigma_{C_0}^\infty}[e^{\varphi(f)}](\eta), \tag{3.3}$$

we conclude that $E_\mu^{\Sigma_{C_0}^\infty}[e^{\varphi(f)}]$ is Σ_{C^0} measurable as a function of η. This enables us to prove that for any $F \in L^\infty[S'(\mathbb{R}^2), \Sigma, d\mu]$, $E^{\Sigma_{C_0}^\infty}[F]$ is Σ_{C^0} measurable. By Theorem 2.4 we have that the σ algebra "at infinity" corresponding to the Gibbs measure associated with the interaction $U_\Lambda(\varphi + \psi_\eta^{C_0})$ is trivial for almost every η with respect to $d\mu$.

Let $\mathbb{R}^2 \setminus C^0 = \Omega_+ \cup \Omega_-$, where $\Omega_\pm$ are connected components of the set $\mathbb{R}^2 \setminus C^0$, and let F_+ and F_- be bounded functions which are $\Sigma_{\Omega_+} \cap \Sigma_{\Lambda_n}$ and $\Sigma_{\Omega_-} \cap \Sigma_{\Lambda_n}$, respectively, measurable. By Theorem 2.4 and the discussion above we have then that for almost every η with respect to $d\mu$

$$E_\mu^{C_0}[F_+\cdot F_-](\eta)$$

$$= \lim_{n'\to\infty} E_\mu^{C_0\cup \mathbf{R}^2\setminus\Lambda_{n'}}[F_+\cdot F_-](\eta) \qquad \text{(by Theorem 2.4)}$$

$$= \lim_{n'\to\infty} E_\mu^{C_0\cup\partial\tilde{\Lambda}_{n'}}[F_+\cdot F_-](\eta) \qquad \text{(by LMP)}$$

$$= \lim_{n'\to\infty} E_\mu^{C_0\cup\partial\tilde{\Lambda}_{n'}}[F_+](\eta)E_\mu^{C_0\cup\partial\tilde{\Lambda}_{n'}}[F_-](\eta)$$

$$= \lim_{n'\to\infty} E^{C_0\cup\mathbf{R}^2\setminus\tilde{\Lambda}_{n'}}[F_+](\eta)E^{C_0\cup\mathbf{R}^2\setminus\Lambda_{n'}}[F_-](\eta)$$

(by Theorem 2.4)

$$= E_\mu^{C_0}[F_+](\eta)E_\mu^{C_0}[F_-](\eta),$$

which is exactly the GMP. Q.E.D.

4. FKG CORRELATION INEQUALITIES

In this section we prove Proposition 2.2. Our idea is the following. First we introduce some auxiliary regularization for the Dirichlet problem (1.6) in such a way that the lattice approximation for the regularized conditioned measure is convergent. For this approximation we check immediately the FKG condition by applying the Avron–Herbst–Simon criterium. Then we prove that for almost every η with respect to $d\mu$ this auxiliary regularization can be removed.

Let $\chi_\epsilon\in\mathfrak{d}(\mathbf{R}^2)$, $\chi_\epsilon>0$, be an arbitrary sequence weakly regularizing the original problem (1.6); i.e., instead of (1.6) we consider the following two-sided Dirichlet problem. Let C be any piecewise C^1 curve in $\mathbf{R}^2$. We define $\psi_{\eta_\epsilon}^C$ as a solution of

$$\begin{cases}(-\Delta+m_0^2)\Psi_{\eta_\epsilon}^C(x)=0, & x\notin C,\\ \Psi_{\eta_\epsilon}^C(x)\equiv\eta_\epsilon(x)\equiv(\eta^*\chi_\epsilon(x), & x\in C,\end{cases} \tag{4.1}$$

where $w-\lim_{\epsilon\searrow 0}\chi_\epsilon=\delta$. This is a well-posed Dirichlet problem because by the fundamental principle $\eta_\epsilon(x)\in C^\infty(\mathbf{R}^2)$. Let us consider the following measure:

$$d\mu_{\lambda,\eta_\epsilon\Lambda}^C(\varphi)=[Z_{\lambda,\eta_\epsilon\Lambda}^C]^{-1}\times\exp\left[-\lambda\int_\Lambda dx{:}e^{\alpha\varphi}{:}(x)\exp(\alpha\Psi_{\eta_\epsilon}^C){:}(x)\right]\cdot d\mu_0^C(\varphi), \tag{4.2}$$

where now

$${:}\exp\alpha\Psi_{\eta_\epsilon}^C{:}(x)=\exp[-(\alpha^2/2)K_\epsilon^C(x,x)]\exp[\alpha\Psi_{\eta_\epsilon}^C(x)], \tag{4.3}$$

$$K_\epsilon^C(x,x)=((\chi_\epsilon\otimes\chi_\epsilon)*K^C)(x,x). \tag{4.4}$$

Let us note the following simple fact.

Lemma 4.1: (a) For every $\eta\in S'(\mathbf{R}^2)$, $\Lambda\subset\mathbf{R}^2$ bounded,

${:}\exp\alpha\psi_{\eta_\epsilon}^C{:}\in L^1(\Lambda)$.

(b) For almost every η with respect to $d\mu$, $\Lambda\subset\mathbf{R}^2$ bounded

${:}e^{\alpha\psi_\eta^C}{:}(x)\in L^1(\Lambda)$.

Proof:

(a) This is trivial.

(b) We have

$$\int d\mu(\eta)\|{:}e^{\alpha\Psi_\eta^C}{:}(x)\|_{L^1(\Lambda)}$$

$$=\int_\Lambda dx\int d\mu(\eta){:}e^{\alpha\Psi_\eta^C}{:}(x)$$

[by positivity of P^C and FKG inequalities for $d\mu(\eta)$],

$$\leqslant\int_\Lambda dx\int d\mu_0(\eta){:}e^{\alpha\psi_\eta^C}{:}(x) \tag{4.5}$$

$$=\int_\Lambda dx\, e^{(\alpha^2/2)K^C(x,x)}<\infty$$

for small α because local singularities of K^C are of logarithmic type. Q.E.D.

Remark: From part (b) of this lemma we get immediately the proof of statement 1 of Proposition 2.2. To see this we apply the Jensen inequality

$$Z_{\Lambda,\eta,\lambda}^C\geqslant\exp\left[-\lambda\int_\Lambda{:}e^{\alpha\Psi_\eta^C}{:}e^{(\alpha^2/2)K_{(x,x)}^{C\cup\partial\Lambda}}\right]. \tag{4.6}$$

For the proof of convergence of the lattice approximation for the measure $d\mu_{\lambda,\eta_\epsilon,\Lambda}^C$ we refer to Refs. 3 and 4. Using the Avron–Herbst–Simon criterion we immediately conclude that the FKG correlation inequalities for $d\mu_{\lambda,\eta_\epsilon,\Lambda}^C$ hold. Now we prove

Theorem 4.2: For almost every η with respect to $d\mu$

$$\text{w-}\lim_{\epsilon\searrow 0} d\mu_{\lambda,\eta,\Lambda}^C=d\mu_{\lambda,\eta,\Lambda}^C.$$

Proof: Let us consider the Laplace transform of the measure $d\mu_{\lambda,\eta_\epsilon,\Lambda}^C$

$$\langle e^{\varphi(f)}\rangle_{\lambda,\eta_\epsilon,\Lambda}^C\equiv\int e^{\varphi(f)}d\mu_{\lambda,\eta_\epsilon,\Lambda}^C(\varphi). \tag{4.7}$$

We take $f\geqslant 0$ without loss of generality. By FKG correlation inequalities

$$\langle e^{\varphi(f)}\rangle_{\lambda,\eta_\epsilon,\Lambda}^C\leqslant\int d\mu_0^C(\varphi)e^{\varphi(f)}. \tag{4.8}$$

Hence by the Lebesgue dominated convergence theorem

$$\lim_{\epsilon\searrow 0}\int d\mu(\eta)\langle e^{\varphi(f)}\rangle_{\lambda,\eta_\epsilon,\Lambda}^C$$

$$\leqslant\int d\mu(\eta)\lim_{\epsilon\searrow 0}\langle e^{\varphi(f)}\rangle_{\lambda,\eta_\epsilon,\Lambda}^C$$

$$\leqslant\int d\mu_0^C(\varphi)e^{\varphi(f)}. \tag{4.9}$$

Therefore almost surely with respect to $d\mu$ the limit $\lim_{\epsilon\downarrow 0}\langle e^{\varphi(f)}\rangle_{\lambda,\eta_\epsilon,\Lambda}^C$ exists. Now we show that this limit is equal to $\langle e^{\varphi(f)}\rangle_{\lambda,\eta,\Lambda}^C$ for almost every η with respect to $d\mu$. For this is enough to prove (almost surely with respect to $d\mu$)

$$\text{(i)}\quad \lim_{\epsilon\searrow 0}\int e^{\varphi(f)}\exp[-U_\Lambda(\varphi+\Psi_{\eta_\epsilon}^C)]d\mu_0^C(\varphi)$$

$$=\int e^{\varphi(f)}e^{-U_\Lambda(\varphi+\Psi_\eta^C)}d\mu_0^C(\varphi) \tag{4.10}$$

and

$$\text{(ii)}\quad Z_{\lambda,\eta_\epsilon,\Lambda}^C>0 \text{ uniformly in } \epsilon. \tag{4.11}$$

Step (ii) follows immediately from Lemma 4.1 and the Jensen inequality. We proceed to prove step (i).

By application of the Duhamel formula it follows that it is enough to show that

$$\lim_{\epsilon\searrow 0}\int d\mu(\eta) \times \int \left[U_\Lambda(\varphi+\Psi^C_{\eta_\epsilon})-U_\Lambda(\varphi+\Psi^C_\eta)\right]^2 d\mu_0^C(\varphi)=0. \tag{4.12}$$

With help of the FKG correlation inequalities we have

$$\begin{aligned}&\int d\mu(\eta)\int d\mu_0^C(\varphi)\left[U_\Lambda(\varphi+\Psi^{C_0}_{\eta_\epsilon})\right]^2\\ &=\lambda^2\int_\Lambda\int_\Lambda dx\,dy\int d\mu(\eta)d\mu_0^C(\varphi):e^{\alpha\varphi}:(x):e^{\alpha\varphi}\\ &\quad :(y):\exp(\alpha\Psi^C_{\eta_\epsilon}):(x):\exp(\alpha\Psi^C_{\eta_\epsilon}):(y)\\ &\leqslant\lambda^2\int_\Lambda\int_\Lambda dx\,dy\int d\mu_0^C(\varphi):e^{\alpha\varphi}:(x):e^{\alpha\varphi}:(y)\\ &\quad\times\int d\mu_0(\eta):\exp(\alpha\Psi^C_{\eta_\epsilon}):(x):\exp(\alpha\Psi^C_{\eta_\epsilon}):(y).\end{aligned} \tag{4.13}$$

By simple Gaussian computation and application of the dominated convergence theorem we have

$$\begin{aligned}&\lim_{\epsilon\searrow 0}\int d\mu(\eta)\int d\mu_0^C(\varphi)U_\Lambda^2(\varphi+\Psi^C_{\eta_\epsilon})(x)\\ &\quad=\int d\mu(\eta)\int d\mu_0^C(\varphi)U_\Lambda^2(\varphi+\Psi^C_\eta)(x)\end{aligned} \tag{4.14}$$

provided α is sufficiently small. Q.E.D.

Extension to the case of $U_\Lambda(\varphi+\psi^C_{\eta_\epsilon}+\sigma\psi^n_{\eta_\epsilon})$ is straightforward. So we have proved that $d\mu^{C,\partial\Lambda}_{\lambda,\eta,\Lambda}$ is a weak limit of a measure for which the FKG correlation inequalities hold. Therefore the limiting measure also obeys these inequalities. This ends the proof of Theorem 4.2 and Proposition 2.2. Q.E.D.

ACKNOWLEDGMENTS

I am happy to have an occasion to express my deep gratitude to Professor K. Osterwalder, who gave me the opportunity to spend the academic year 1979/80 at the Forschungsinstitute für Mathematik, E. T. H. Zürich, where this work was done. Additionally I wish to thank him for numerous discussions and advice. During the course of this work I have also benefited greatly from discussions with Professor T. Spencer.

APPENDIX A

Without loss of generality we always assume $f\geqslant 0$. Let Y_n and Λ_n be a set in $\mathbb{R}^2$ as explained in the course of the proof of Theorem 2. $\{n'\}$ below always denotes the subsequence mentioned in Proposition 2.1.

Lemma A.1: For almost every η with respect to $d\mu$

$$\lim_{n'\to\infty} J^1_{n',\eta}=0.$$

Proof: By FKG correlation inequalities we have

$$\begin{aligned}&|F^{C_0,\partial\tilde\Lambda_{n'},\sigma}_{\lambda,\eta,\Lambda_{n'}}(f;x)|\\ &\leqslant\langle e^{\varphi(f)}:e^{\alpha\varphi}:(x)\rangle^{C_0,\partial\tilde\Lambda_{n'},\sigma}_{\lambda,\eta,\sigma}\\ &\quad+\langle e^{\varphi(f)}\rangle^{C_0,\partial\tilde\Lambda_{n'},\sigma}_{\lambda,\eta,\Lambda_{n'}}\langle:e^{\alpha\varphi}:(x)\rangle^{C_0,\partial\tilde\Lambda_{n'},\sigma}_{\lambda,\eta,\Lambda_{n'}}\\ &\leqslant\langle e^{\varphi(f)}:e^{\alpha\varphi}:(x)\rangle_0+\langle e^{(f)\varphi}\rangle_0\langle:e^{\alpha\varphi}:(x)\rangle_0\\ &=\langle e^{\varphi(f)}\rangle_0\left[e^{\alpha f*S^2_0(x)}+1\right],\end{aligned}$$

$$\begin{aligned}&\int|J^1_{n',\eta}|\,d\mu(\eta)\\ &\leqslant\alpha^2\int_{Y_{n'}}dx\int_0^1 d\sigma\,\sigma\langle e^{\varphi(f)}\rangle_0\left[e^{\alpha f*S^2_0(x)}+1\right]\\ &\quad\times K^{C_0,\partial\tilde\Lambda_{n'}}_{(x,x)}\int d\mu(\eta):\exp(\alpha\sigma\Psi^{n'}_\eta):(x):\exp(\alpha\Psi^{C_0}_\eta):(x).\end{aligned}$$

By FKG,

$$\begin{aligned}&\leqslant\alpha^2\int_{Y_{n'}}dx\int_0^1 d\sigma\,\sigma\langle e^{\varphi(f)}\rangle_0\left[e^{\alpha f*S^2_0(x)}+1\right]\\ &\quad\times K^{C_0,\partial\tilde\Lambda_{n'}}_{(x,x)}\int d\mu_0(\eta):\exp(\alpha\sigma\Psi^{n'}_\eta):(x):\exp(\alpha\Psi^{C_0}_\eta):(x).\end{aligned}$$

By LBP for sufficiently large n'

$$\begin{aligned}&\leqslant\alpha^2(1+\epsilon_{n'})\int_{Y_{n'}}dx\int_0^1 d\sigma\,\sigma\langle e^{\varphi(f)}\rangle_0\left[e^{\alpha f*S^2_0(x)}+1\right]\\ &\quad\times K^{C_0,\partial\tilde\Lambda_{n'}}_{(x,x)}\int d\mu_0(\eta):\exp(\alpha\Psi^{C_0}_\eta):(x)\\ &\leqslant\alpha^2(1+\epsilon_{n'})\int_{Y_{n'}}dx\int_0^1 d\sigma\,\sigma\langle e^{\varphi(f)}\rangle_0\left[e^{\alpha f*S^2_0(x)}+1\right]\\ &\quad\times K^{C_0,\partial\tilde\Lambda_{n'}}_{(x,x)}\exp\left[-(\alpha^2/2K^{C_0}(x,x)\right].\end{aligned}$$

It remains to use the exponential decay of $K^{C_0,\partial\tilde\Lambda_n}_{(x,x)}$ as $\Lambda_{n'}\uparrow\infty$ in order to finish the proof that $\lim_{n'\to\infty}\int d\mu(\eta)J^1_{n',\eta}=0$. Q.E.D.

Lemma A.2: For almost every η with respect to $d\mu$

$$\lim_{n'\to\infty} J^2_{n',\eta}=0.$$

Proof: Here we use the uniform exponential cluster property proved in Corollary 2.3,

$$\begin{aligned}&\int d\mu(\eta)|J^2_{n',\eta}|\\ &\leqslant\alpha^2\int_0^1 d\sigma\,\sigma\int_{\Lambda_{n'}\setminus Y_{n'}}dxK^{C_0,\partial\tilde\Lambda_{n'}}(x,x)e^{-m_0 d(\mathrm{supp}f,\Lambda_{n'}\setminus Y_{n'})}\\ &\quad\times\int d\mu(\eta):\exp(\alpha\Psi^{C_0}_\eta):(x):\exp(\alpha\sigma\Psi^{n'}_\eta):(x).\end{aligned}$$

By FKG

$$\begin{aligned}&\leqslant e^{-m_0\mathrm{dist}(\mathrm{supp}f,\Lambda_{n'}\setminus Y_{n'})}\alpha^2\\ &\quad\times\int_0^1 d\sigma\,\sigma\int dx\,K^{C_0,\partial\tilde\Lambda_{n'}}_{(x,x)}\exp[(-\alpha^2/2)K^{C_0}(x,x)]\\ &\quad\times e^{-(\alpha^2\sigma^2/2)K^n(x,x)}\exp[+(\alpha^2\sigma^2/2)K^{C_0}(x,x)].\end{aligned}$$

If $\Lambda_{n'}\setminus Y_{n'}\to\infty$ sufficiently fast then

$$\lim_{n'\to\infty}\int d\mu(\eta)J^2_{n',\eta}=0.$$ Q.E.D.

Lemma A.3: For almost every η with respect to $d\mu(\eta)$

$$\lim_{n'\to\infty} J^3_{n',\eta} = 0.$$

Proof: Now we use LBP. For sufficiently large n' we have

$$|J^3_{n',\eta}| \leqslant \alpha \int_{Y_{n'}} dx \int_0^1 d\sigma\, \sigma(1+\epsilon_{n'}) \times e^{-(\alpha/2)\mathrm{dist}(Y_{n'},\partial\Lambda_{n'})} |F^{C_0,\partial\tilde{\Lambda}_{n'},\sigma}_{\lambda,\eta,\Lambda_{n'}}(f,x)|.$$

By FKG

$$\leqslant \alpha(1+\epsilon_{n'}) \int_{Y_{n'}} dx \int_0^1 d\sigma\, \sigma e^{-(\alpha/2)\mathrm{dist}(Y_{n'},\partial\Lambda_{n'})} \times \langle e^{\varphi(f)}\rangle_0 \left[e^{\alpha f * S^2_0(x)} + 1\right].$$

Because $\left[e^{\alpha f * S^2_0(x)} + 1\right]$ is locally L^p integrable ($p \geqslant 1$) we conclude

$$\lim_{n'\to\infty} J^3_{n',\eta} = 0$$

for almost every η with respect to $d\mu$.

Lemma A.4: For almost every η with respect to $d\mu$

$$\lim_{n'\to\infty} J^4_{n',\eta} = 0.$$

Proof: (by uniform exponential cluster property)

$$\int d\mu(\eta)|J^4_{n',\eta}| \leqslant \alpha \int_0^1 d\sigma\, \sigma \int_{\Lambda_{n'}\setminus Y_{n'}} dx\, e^{-m_0 \mathrm{dist}(\mathrm{supp} f, \Lambda_{n'}\setminus Y_{n'})} \times \int d\mu(\eta) {:}\exp(\alpha \Psi^{C_0}_\eta){:}(x) |\Psi^{n'}_\eta(x)| {:}\exp(x\sigma\Psi^{n'}_\sigma){:}(x).$$

Now we use Cauchy–Schwartz's inequality

$$\int d\mu(\eta) {:}\exp(\alpha\Psi^{C_0}_\eta){:}(x)|\Psi^{n'}_\eta(x)| {:}\exp(\alpha\sigma\Psi^{n'}_\eta){:}(x) \leqslant \left[\int d\mu(\eta)({:}\exp(\alpha\Psi^{C_0}_\eta){:}(x){:}\exp(\alpha\sigma\Psi^{n'}_\eta){:}(x))^2\right]^{1/2} \times \left[\int d\mu(\eta)(\Psi^{n'}_\eta(x))^2\right]^{1/2}$$

(by FKG)

$$\leqslant \left[\int d\mu_0(\eta)(\Psi^{n'}_\eta)^2(x)\right]^{1/2} \times \left[\int d\mu_0(\eta)({:}\exp(\alpha\Psi^{C_0}_\eta){:}(x))^2({:}e^{\alpha\sigma\Psi^n_\eta}{:}(x))^2\right]^{1/2}.$$

Both

$$\left[\int d\mu_0(\eta)({:}\exp(\alpha\Psi^{C_0}_\eta){:}(x))^2\right]^{1/2}$$

and

$$\left[\int d\mu_0(\eta)(\Psi^{n'}_\eta(x))^2\right]^{1/2}$$

are locally L^p integrable ($p \geqslant 1$) as can be shown by simple Gaussian computations. This enables us to conclude that

$$\lim_{n'\to\infty} \int d\mu(\eta)|J^4_{n',\eta}| = 0.$$

if $\mathrm{dist}(\mathrm{supp} f, \Lambda_{n'}\setminus Y_{n'}) \to \infty$ sufficiently fast. Q.E.D.

APPENDIX B

Theorem B.1: For almost every η with respect to $d\mu$

$$\text{w-}\lim_{\Lambda_{n'}\uparrow\mathbb{R}^2} d\mu^{C_0,\partial\Lambda_{n'}}_{\Lambda_{n'},\eta}(\varphi) = d\mu^{C_0}_\eta(\varphi),$$

where $d\mu^{C_0}_\eta$ is the measure $d\mu$ conditioned by η on the given piecewise C^1 curve $C_0 \subset \mathbb{R}^2$ and $\Lambda_{n'}$ is a subsequence as in Appendix A.

Proof: Let us consider the following identity:

$$\int e^{\varphi(f)} d\mu^{C_0,\partial\Lambda}_{\lambda,\eta,\Lambda}(\varphi) = \int_0^1 d\sigma \frac{d}{d\sigma} e^{\varphi(f)} d\mu^{C_0,\partial\Lambda,\sigma}_{\lambda,\eta,\Lambda}, \tag{B1}$$

where

$$d\mu^{C_0,\partial\Lambda,\sigma}_{\lambda,\eta,\Lambda} \equiv [Z^{C_0,\partial\Lambda,\sigma}_{\lambda,\eta,\Lambda}]^{-1} \times \exp\left(-\lambda\int_\Lambda dx {:}e^{\alpha\varphi}{:}(x){:}\exp(\alpha\Psi^{C_0}_\eta){:}(x)\right) d\mu_0^{C_0\cup\partial\Lambda,\sigma}(\varphi) \tag{B2}$$

and $d\mu_0^{C_0\cup\partial\Lambda}$ is a Gaussian measure with mean zero and covariance $(-\Delta_{C_0\cup\partial\Lambda} + m_0^2)^{-1}(x,y)$. Using well-known formulas we have

$$\frac{d}{d\sigma}\int e^{\varphi(f)} d\mu^{C_0,\partial\Lambda,\sigma}_{\lambda,\eta,\Lambda}(\varphi) = \frac{1}{2}\int\int dx\, dy \frac{\delta}{\delta\varphi(x)} K^{C_0}(x,x) \frac{\delta}{\delta\varphi(y)} \times \int e^{\varphi(f)} d\mu^{C_0,\partial\Lambda,\sigma}_{\lambda,\eta,\Lambda}(\varphi). \tag{B3}$$

Taking the functional derivatives in the last formula we get

$$\begin{aligned}\frac{d}{d\sigma}\int e^{\varphi(f)} d\mu^{C_0,\partial\Lambda,\sigma}_{\lambda,\eta,\Lambda}(\varphi) &= -\lambda\alpha^2 \int dx\, K^{C_0\cup\partial\Lambda}_{(x,x)} {:}\exp(\alpha\Psi^{C_0}_\eta){:}(x) \\ &\times \langle {:}e^{\alpha\varphi}{:}(x); e^{\varphi(f)}\rangle^{C_0,\partial\Lambda,\sigma}_{\lambda,\eta,\Lambda} \\ &+ \lambda^2\alpha^2 \int\int K^{C_0\cup\partial\Lambda}_{(x,y)} {:}\exp(\alpha\Psi^{C_0}_\eta){:}(x){:}\exp(\alpha\Psi^{C_0}_\eta{:}(y) \\ &\times \langle {:}e^{\alpha\varphi}{:}(x){:}e^{\alpha\varphi}{:}(y); e^{\varphi(f)}\rangle^{C_0,\partial\Lambda,\sigma}_{\lambda,\eta,\Lambda} dx\, dy \\ &+ \langle e^{\varphi(f)}\rangle^{C_0,\partial\Lambda,\sigma}_{\lambda,\eta,\Lambda} \int\int f(x)f(y) K^{C_0\cup\partial\Lambda}_{(x,y)} dx\, dy \\ &- \lambda\cdot\alpha \int\int K^{C_0\cup\partial\Lambda}_{(x,y)} \\ &\times \langle {:}e^{\alpha\varphi}{:}(x) e^{\varphi(f)}\rangle^{C_0,\partial\Lambda,\sigma}_{\lambda,\eta,\Lambda} \\ &\times f(y){:}\exp(\alpha\Psi^{C_0}_\eta){:}(x) dx\, dy \\ &- \lambda\cdot x \int\int dx\, dy\, K^{C_0\cup\partial\Lambda}_{(x,y)} \\ &{:}\exp(\alpha\Psi^{C_0}_\eta){:}(y) \\ &\times \langle {:}e^{\alpha\varphi}{:}(y) e^{\varphi(f)}\rangle^{C_0,\partial\Lambda,\sigma}_{\lambda,\eta,\Lambda} \cdot f(x).\end{aligned} \tag{B4}$$

We are going to show that for almost every η with respect to

$d\mu$ all the terms which appear in the right-hand side of (B4) give vanishing contribution to the limit $\Lambda = \mathbb{R}^2$. We start from the first term:

Lemma B.1: For almost every η with respect to $d\mu$, uniformly in σ

$$\lim_{\Lambda \nearrow \mathbb{R}^2} \int [K^{C_0 \cup \partial\Lambda}_{(x,x)} - K^{C_0}(x,x)] :\exp(\alpha \Psi^{C_0}_\eta):(x) \times \langle :e^{\alpha\varphi}:(x); e^{\varphi(f)} \rangle^{C_0, \partial\bar{\Lambda}, \sigma}_{\lambda,\eta,\Lambda} = 0. \tag{B5}$$

Proof: Let us choose $\Xi \subset \mathbb{R}^2$ such that $\operatorname{supp} f \subset \Xi$ and $\Xi \subset \Lambda$. Then we have

$$\begin{aligned} &\int d\mu(\eta) \int_\Lambda [K^{C_0 \cup \partial\Lambda}_{(x,x)} - K^{C_0}(x,x)] \\ &\quad :\exp(\alpha \Psi^{C_0}_\eta):(x) \langle :e^{\alpha\varphi}:(x); e^{\varphi(f)} \rangle^{C_0, \partial\Lambda, \sigma}_{\lambda,\eta,\Lambda} dx \\ &= \int d\mu(\eta) \int_\Xi [K^{C_0 \cup \partial\Lambda}_{(x,x)} - K^{C_0}(x,x)] \\ &\quad :\exp(\alpha \Psi^{C_0}_\eta):(x) \langle :e^{\alpha\varphi}:(x); e^{\varphi(f)} \rangle^{C_0, \partial\Lambda, \sigma}_{\lambda,\eta,\Lambda} dx \\ &\quad + \int d\mu(\eta) \int_{\Lambda \setminus \Xi} [K^{C_0 \cup \partial\Lambda}_{(x,x)} - K^{C_0}(x,x)] \\ &\quad :\exp(\alpha \Psi^{C_0}_\eta):(x) \langle :e^{\alpha\varphi}:(x); e^{\varphi(f)} \rangle^{C_0, \partial\Lambda, \sigma}_{\lambda,\eta,\Lambda} dx. \end{aligned} \tag{B6}$$

By FKG correlation inequalities we have for every bounded Ξ

$$\int d\mu(\eta) \int_\Xi :\exp(\alpha \Psi^{C_0}_\eta):(x) \langle :e^{\alpha\varphi}:(x); e^{\varphi(f)} \rangle^{C_0, \partial\Lambda, \sigma}_{\lambda,\eta,\Lambda} dx < \infty, \tag{B7}$$

from which it follows, using the exponential decay of $K^{C_0 \cup \partial\Lambda}(x,x) - K^{C_0}(x,x)$, that

$$\lim_{\Lambda \nearrow \mathbb{R}^2} \int d\mu(\eta) \int_\Xi [K^{C_0 \cup \partial\Lambda}(x,x) - K^{C_0}(x,x)] :\exp(\alpha \Psi^{C_0}_\eta):(x) \langle :e^{\alpha\varphi}:(x); e^{\varphi(f)} \rangle^{C_0, \partial\Lambda, \sigma}_{\lambda,\eta,\Lambda} dx = 0. \tag{B8}$$

On the other hand, by using the exponential decay of $\langle :e^{\alpha\varphi}:(x); e^{\varphi(f)} \rangle^{C_0, \partial\Lambda, \sigma}_{\lambda,\eta,\Lambda}$ we have

$$\begin{aligned} &\int d\mu(\eta) \int_{\Lambda \setminus \Xi} [K^{C_0 \cup \partial\Lambda}(x,x) - K^{C_0}(x,x)] :\exp(\alpha \Psi^{C_0}_\eta):(x) \\ &\quad \langle :e^{\alpha\varphi}:(x); e^{\varphi(f)} \rangle^{C_0, \partial\Lambda, \sigma}_{\lambda,\eta,\Lambda} dx \\ &\leqslant e^{-m_0 \operatorname{dist}(\Xi, \Lambda \setminus \Xi)} \int d\mu(\eta) \\ &\quad \times \int_{\Lambda \setminus \Xi} [K^{C_0 \cup \partial\Lambda}(x,x) - K^{C_0}(x,x)] :\exp(\alpha \Psi^{C_0}_\eta):(x) \\ &\leqslant e^{-m_0 \operatorname{dist}(\Xi, \Lambda \setminus \Xi)} \int_{\Lambda \setminus \Xi} [K^{C_0 \cup \partial\Lambda}(x,x) - K^{C_0}(x,x)] dx. \end{aligned} \tag{B9}$$

In the last step we have again used FKG correlation inequalities. Q.E.D.

Lemma B.2: For almost every η with respect to $d\mu$, uniformly in σ

$$\begin{aligned} &\lim_{\Lambda \nearrow \mathbb{R}^2} \int_\Lambda \int_\Lambda dx\, dy [K^{C_0 \cup \partial\Lambda}(x,y) - K^{C_0}(x,y)] \\ &\quad :\exp(\alpha \Psi^{C_0}_\eta):(x); \exp(\alpha \Psi^{C_0}_\eta):(y) \\ &\quad \times \langle :e^{\alpha\varphi}:(x) :e^{\alpha\varphi}:(y); e^{\varphi(f)} \rangle^{C_0, \partial\Lambda, \sigma}_{\lambda,\eta,\Lambda} = 0. \end{aligned} \tag{B10}$$

Proof: Let $\Xi \subset \mathbb{R}^2$ be as in Lemma B.1. Then we have

$$\int_\Lambda \int_\Lambda \cdots = \int_\Xi \int_\Xi \cdots + \int_\Xi \int_{\Lambda \setminus \Xi} \cdots + \int_{\Lambda \setminus \Xi} \int_\Xi \cdots + \int_{\Lambda \setminus \Xi} \int_{\Lambda \setminus \Xi} \cdots . \tag{B11}$$

By FKG correlation inequalities and exponential decay of $[K^{C_0 \cup \partial\Lambda} - K^{C_0}](x,y)$ we have

$$\begin{aligned} &\lim_{\Lambda \nearrow \mathbb{R}^2} \int_\Xi \int_\Xi dx\, dy |K^{C_0 \cup \partial\Lambda}(x,y) - K^{C_0}(x,y)| \\ &\quad \cdot \int d\mu(\eta) :\exp(\alpha \Psi^{C_0}_\eta):(x) \\ &\quad \times :\exp(\alpha \Psi^{C_0}_\eta):(y) | \langle :e^{\alpha\varphi}:(x) :e^{\alpha\varphi}:(y); e^{\varphi(f)} \rangle^{C_0, \partial\Lambda, \sigma}_{\lambda,\eta,\Lambda} | \\ &\leqslant \lim_{\Lambda \nearrow \mathbb{R}^2} \int_\Xi \int_\Xi dx\, dy |K^{C_0 \cup \partial\Lambda}(x,y) - K^{C_0}(x,y)| \\ &\quad \cdot \int d\mu(\eta) :\exp(\alpha \Psi^{C_0}_\eta):(x) \\ &\quad \times :\exp(\alpha \Psi^{C_0}_\eta):(y) | \langle :e^{\alpha\varphi}:(x) :e^{\alpha\varphi}:(y) \rangle^{C_0, \partial\Lambda, \sigma}_{\lambda,\eta,\Lambda} \cdot \langle e^{\varphi(f)} \rangle^{C_0, \partial\Lambda, \sigma}_{\lambda,\eta,\Lambda} \\ &\quad + (\langle :e^{\alpha\varphi}:(x) :e^{\alpha\varphi}:(y) e^{\varphi(f)} \rangle^{C_0, \partial\Lambda, \sigma}_{\lambda,\eta,\Lambda}) \\ &\leqslant \lim_{\Lambda \nearrow \mathbb{R}^2} \int_\Xi \int_\Xi dx\, dy |K^{C_0 \cup \partial\Lambda}(x,y) - K^{C_0}(x,y)| \\ &\quad \cdot \int d\mu(\eta) :\exp(\alpha \Psi^{C_0}_\eta):(x) \\ &\quad \times :\exp(\alpha \Psi^{C_0}_\eta):(y) (\langle :e^{\alpha\varphi}:(x) :e^{\alpha\varphi}:(y) e^{\varphi(f)} \rangle_0 \\ &\quad + \langle :e^{\alpha\varphi}:(x) :e^{\alpha\varphi}:(y) \rangle_0 \langle e^{\alpha(f)} \rangle_0) \\ &\leqslant \lim_{\Lambda \nearrow \mathbb{R}^2} \int_\Xi \int_\Xi dx\, dy |K^{C_0 \cup \partial\Lambda}(x,y) - K^{C_0}(x,y)| \\ &\quad \times \int d\mu_0(\eta) :\exp(\alpha \Psi^{C_0}_\eta):(x) \\ &\quad :\exp(\alpha \Psi^{C_0}_\eta):(y) (\langle :e^{\alpha\varphi}:(x) :e^{\alpha\varphi}:(y) e^{\varphi(f)} \rangle_0 \\ &\quad + \langle :e^{\alpha\varphi}:(x) :e^{\alpha\varphi}:(y) \rangle_0 \langle e^{\varphi(f)} \rangle_0) \\ &= \lim_{\Lambda \nearrow \mathbb{R}^2} \int_\Xi \int_\Xi dx\, dy \\ &\quad \times |K^{C_0 \cup \partial\Lambda}(x,y) - K^{C_0}(x,y)| \exp[(\alpha^2/2) K^{C_0}(x,y)] \\ &\quad \times (\langle :e^{\alpha\varphi}:(x) :e^{\alpha\varphi}:(y) e^{\varphi(f)} \rangle_0 \\ &\quad + \langle :e^{\alpha\varphi}:(x) :e^{\beta\varphi}:(y) \rangle_0 \langle e^{\varphi(f)} \rangle_0). \end{aligned} \tag{B12}$$

Logarithmic singularities of the functions which appear in the last expression and exponential decay of $K^{C_0 \cup \partial\Lambda}(x,y) - K^{C_0}(x,y)$ enables us to conclude that the limit $\Lambda \uparrow \mathbb{R}^2$ gives contribution zero for α sufficiently small.

The integral $\int_{\Lambda \setminus \Xi} \int_{\Lambda \setminus \Xi} \cdots$ is then controlled in a similar fashion using exponential decay of correlations. It remains to analyze the integrals $\int_{\Lambda \setminus \Xi} \int_\Xi \cdots$ and $\int_\Xi \int_{\Lambda \setminus \Xi} \cdots$. Let $\Xi_1 \cup \Lambda \setminus (\Xi_1 \cup \Xi) = \Lambda \setminus \Xi$. Then we can write

$$\int_{\Lambda \setminus \Xi} \int_\Xi dx\, dy \cdots = \int_{\Xi_1} \int_\Xi dx\, dy \cdots + \int_{\Lambda \setminus (\Xi_1 \cup \Xi)} \int_\Xi dx\, dy \cdots .$$

Now we proceed in the same way as before, using the exponential decay on $\int_{\Xi_1} \int_\Xi dx\, dy \cdots$ of $[K^{C_0 \cup \partial\Lambda} - K^{C_0}]$ and uniform exponential clustering in order to control the integral $\int_{\Lambda \setminus (\Xi_1 \cup \Xi)} \int_\Xi dx\, dy \cdots$. Q.E.D.

Lemma B.3: For almost every η with respect to $d\mu$, uniformly in σ

$$\lim_{\Lambda \nearrow \mathbf{R}^2} \int\int [K^{C_0 \cup \partial\Lambda}(x,y) - K^{C_0}(x,y)]\, f(x) {:}\exp(\alpha \Psi_\eta^{C_0}){:}(y) \times \langle {:}e^{\alpha\varphi}{:}(x) e^{\varphi(f)} \rangle_{\lambda,\eta,\Lambda}^{C_0,\partial\Lambda,\sigma} = 0. \tag{B13}$$

Proof: As before we use positiveness of $:\exp(\alpha\psi_\eta^{C_0}):(x)$ in order to show that the $L^1(d\mu)$ norm of the expression under the $\lim_{\Lambda \uparrow \mathbf{R}^2}$ goes to zero by divisions of integrations, exponential decay of $K^{C_0 \cup \partial\Lambda}(x,y) - K^{C_0}(x,y)$, and the local integrability property of $f(x):\exp(\alpha\psi_\eta^{C_0}):(y)\langle :e^{\alpha\varphi}:(x) e^{\varphi(f)}\rangle_0$. Q.E.D.

Remark: Similar arguments for the proof of independence of the pure states for some classical boundary conditions in the pure phases of $P(\varphi)_2$ theories have been used before by Fröhlich and Simon.[17]

[1]S. Albeverio and R. Høegh-Krohn, Commun. Math. Phys. **68**, 95 (1979).
[2]S. Albeverio and R. Høegh-Krohn, J. Funct. Anal. **16**, 39 (1974).
[3]J. Fröhlich and Y. M. Park, "Remark on Exponential Interactions and the Quantum Sine-Gordon Equation in Two Space-Time Dimensions," Preprint ZIF Bielefeld (1976) (unpublished).
[4]B. Simon, *The $P(\varphi)_2$ Euclidean (Quantum) Field Theory* (Princeton U.P, Princeton, NJ, 1974).
[5]S. Albeverio and R. Høegh-Krohn, "Topics in infinite dimensional analysis, in *Mathematical Problems in Theoretical Physics*, edited by a G. F. Dell'Antonio, S. Doplicher, and G. Jona-Lasinio, *Lecture Notes in Physics, Vol. 80* (Springer, Berlin, 1978), pp. 279–302.
[6]S. Albeverio and R. Høegh-Krohn, "Local and global Markoff fields", lectures given at Karpacz Winter School, 1978 (unpublished).
[7]S. Albeverio and R. Høegh-Krohn, Z. Wahrscheinlichkeitsth. verw. Geb. **40**, (1977).
[8]S. Albeverio and R. Høegh-Krohn, "Canonical relativistic quantum fields," Ann. Inst. Henri Poincaré, **41**, 32 (1979).
[9]F. Guerra, L. Rosen, and B. Simon, Ann. Math. **101**, 111 (1975).
[10]E. Nelson, J. Funct. Anal. **12**, 97 (1973).
[11]S. Albeverio, R. Høegh-Krohn, and G. Olsen, "The Global Markov property for lattice fields," Bielefeld-Marseille, Preprint, Nov. 1978 (unpublished).
[12]H. Föllmer, "Global Markov property," Preprint E. T. H. Zürich (1978) (unpublished).
[13]S. Goldstein, Commun. Math. Phys. **74**, 223 (1980).
[14]G. Benfatto, E. Pressutti, A. Pulvireuti, Z. Wahrscheinlichkeitsth. verw. Geb. **41**, 305 (1978).
[15]M. Cassandro, E. Olivieri, A. Pellegrinotti, and E. Pressuti, Z. Wahrscheinlichkeitsth. verw. Geb. **41**, 313 (1978).
[16]B. Simon, Commun. Math. Phys. **31**, 127 (1973).
[17]J. Fröhlich and B. Simon, Ann. Math. **105**, 493 (1977).

Multiple fiber bundles and gauge theories of higher order

Richard Kerner
Département de Mécanique, Université Pierre et Marie Curie, 4, Place Jussieu, Paris, France

(Received 1 October 1981; accepted for publication 20 November 1981)

We investigate a theory of gauge fields over multiple bundles, i.e., principal bundles constructed over base spaces which are principal bundles themselves. The Higgs–Kibble field is introduced geometrically, together with the quartic potential. Results of Forgács and Manton are interpreted in this scheme. We also discuss a spherically symmetric ansatz which yields the quartic potential introduced by 't Hooft.

PACS numbers: 11.10.Np, 11.30.Ly

1. INTRODUCTION AND NOTATIONS

Let V be a (pseudo)-Riemannian manifold of dimension n, with the metric form g_V; let G be a compact and semisimple Lie group of dimension N. We denote by g_G the Killing–Cartan invariant metric form on G. Let $P(V,G)$ be a principal fiber bundle over the base space V with the structure group G. We denote by A a left-invariant 1-form over $P(V,G)$, with values in the Lie algebra $\mathscr{A}_G$ of G; the 1-form A defines a *connection* in $P(V,G)$. Let $T_pP(V,G)$ denote a tangent space to $P(V,G)$ at a point $p\in P(V,G)$; by $TP(V,G) = [\cup_{p\in P} T_pP(V,G)]$ we denote the tanget bundle over $P(V,G)$. Let $X\in TP(V,G)$; the vector field X is called *horizontal* if $A(X)=0$. Any $X\in TP(V,G)$ can be uniquely decomposed into the horizontal and vertical parts:

$$X = \text{hor}\, X + \text{ver}\, X \tag{1}$$

such that

$$A(\text{ver}\, X) = A(X), \quad A(\text{hor}\, X) = 0. \tag{2}$$

Let θ be a p-form over $P(V,G)$; let d be its exterior differential. The covariant differential of θ is defined as

$$D\theta(X_1,X_2,\ldots,X_{p+1}) = d\theta(\text{hor}\, X_1,\ldots,\text{hor}\, X_{p+1}). \tag{3}$$

The covariant differential of the connection 1-form A is called *curvature* and denoted by F

$$F = DA. \tag{4}$$

The Maurer–Cartan identity enables us to write

$$F(X,Y) = DA(X,Y) = dA(X,Y) + \tfrac{1}{2}[A(X),A(Y)]. \tag{5}$$

The bracket means the skew-symmetric product in the Lie algebra $\mathscr{A}_G$.

The existence of the canonical projection $\pi: P(V,G)\to V$ and of the connection form A: $TP(V,G)\to\mathscr{A}_G$ enables us to construct a unique metric form on $P(V,G)$:
for any X,Y, and $TP(G,V)$,

$$g_P(X,Y) = g_V(d\pi(X),d\pi(Y)) + g_G(A(X),A(Y)). \tag{6}$$

The principal fiber bundle $P(V,G)$ with the connection A now becomes a Riemannian (or pseudo-Riemannian) manifold; we denote it by $P(V,G,A,g_P)$. [There is another possible choice for defining a canonical metric on $P(V,G)$, namely $g'_P = g_V - g_G$. For physical reasons we do not consider this case, because the resulting signature can no longer imply an interpretation in terms of one "temporal" direction and all the remaining directions as "spatial" ones.]

Let us now give some explicit expressions in local coordinates. We recall that dim $V=n$, dim $G=N$; let $i,j = 1,2,\ldots,n$; $a,b,\ldots = 1,2,\ldots,N$; and $\alpha,\beta,\ldots = 1,2,\ldots,N+n$. Symbolically we denote $\alpha = (j,a)$. Consider an open set $U\subset P(V,G)$ isomorphic with the direct product of an open set $\pi(U)\subset V$ and the group G:

$$U\sim\pi(U)\times G.$$

We can always choose a coordinate system in U such that for $p\in U$, $p=\{p^\alpha\}=\{x^i,\xi^a\}$, where the coordinates x^i correspond to the points $\pi(p)=x\in\pi(U)$, whereas the coordinates ξ^a denote an element of G. We choose a basis L_a in the Lie algebra $\mathscr{A}_G$ such that

$$[L_a,L_b] = C^c_{ab}L_c, \tag{7}$$

C^a_{bc} being the structure constants of G. In the same local coordinates the Killing–Cartan metric tensor is equal to

$$\mathring{g}_{ab} = C^c_{ad}C^d_{cb}, \tag{8}$$

whereas the metric tensor g_V has the components $\mathring{g}_{ij}$.

Now we can decompose the connection 1-form and the curvature 2-form

$$A = A^aL_a = A^a_\beta L_a dp^\beta = (A^a_j dx^j + A^a_b d\xi^b)L_a, \tag{9}$$

$$F = F^aL_a = F^a_{\alpha\beta}(dp^\alpha\wedge dp^\beta)L_a. \tag{10}$$

By definition F does vanish when applied to vertical vectors; therefore, the components F^c_{ab} and $F^c_{ib} = -F^c_{bi}$ vanish identically

$$\partial_a A^d_b - \partial_b A^d_a + C^d_{gh}A^g_a A^h_b = 0, \tag{11}$$

$$\partial_j A^d_b - \partial_b A^d_j + C^d_{gh}A^g_j A^h_b = 0, \tag{12}$$

and the only nonvanishing components of the curvature form F are

$$F^a_{ij} = \partial_i A^a_j - \partial_j A^a_i + C^a_{bd}A^b_i A^d_j. \tag{13}$$

The scalar product can be defined for any two forms of the same degree. We denote by $(F^a,F^b)_{g_V}$ the scalar product defined by

$$(F^a,F^b)_{g_V} = \mathring{g}^{ij}\mathring{g}^{kl}F^a_{ik}F^b_{jl}. \tag{14}$$

Finally, for any two elements of $\mathscr{A}_G$ we denote by $\text{Tr}_G(P,Q)$ the expression

$$\text{Tr}_G(P,Q) = \mathring{g}_{ab}P^aQ^b, \tag{15}$$

where $P=P^aL_a$, $Q=Q^bL_b$. Therefore

$$\text{Tr}_G(F,F)_{g_V} = \mathring{g}_{ab}\mathring{g}^{ij}\mathring{g}^{kl}F^a_{ik}F^b_{jl}. \tag{16}$$

 0022-2488/83/020356-05$02.50

We close this paragraph by expressing the Laplace–Beltrami operator on the bundle $P(V,G,A;g_P)$ induced by the metric g_P: If in local coordinates g_P has the components $g_{\alpha\beta}$, then for a scalar function on P we have

$$(d\delta + \delta d)f = g^{\alpha\beta}\nabla_\alpha \nabla_\beta f, \tag{17}$$

where ∇_α is the covariant derivative with respect to the Christoffel connection of g_P; for a 1-form $\phi = \phi_\alpha\, dp^\alpha$ we have

$$\begin{aligned}[(d\delta + \delta d)\phi]_\gamma &= g^{\alpha\beta}[\nabla_\alpha \nabla_\beta \phi_\gamma - \nabla_\alpha \nabla_\gamma \phi_\beta + \nabla_\gamma \nabla_\alpha \phi_\beta] \\ &= g^{\alpha\beta}[\nabla_\alpha \nabla_\beta \phi_\gamma + R^{\ \delta}_{\alpha\gamma\beta}\phi_\delta],\end{aligned} \tag{18}$$

$R^{\ \delta}_{\alpha\gamma\beta}$ being the Riemannian curvature tensor of the Christoffel connection induced by g_P.

2. DEVELOPMENTS

With the notations established as above we can formulate some useful results.

Theorem 1: Let $P(V,G,A;g_P)$ be a principal fiber bundle over a Riemannian manifold V with the semisimple and compact structure group G, endowed with a connection A and the canonical metric g_P defined by (6). The scalar curvature R_P is a sum of the following three terms:

$$R_P = R_V - \tfrac{1}{4}\mathrm{Tr}_G(F,F)_{g_V} + \tfrac{1}{4}g_G^2, \tag{19}$$

where $g_G^2 = g_{ab}g^{ab} = N$ is the square of the Killing–Cartan metric of the structure group G. R_V is the scalar curvature of the base manifold V; the second term is identified with the Lagrangian of the gauge field.

Proof: By calculus. The first formulation of this result can be found in Ref. 1, though with some errors in calculations. For correct proofs, see later papers.[2,3]

Corollary 1.1: Consider a structure group G which is a direct product of several compact and semisimple Lie groups, $G = G_1 \times G_2 \times \cdots \times G_k$, and construct a principal fiber bundle P as before: $P(V,G_1\times G_2\times\cdots G_k, A;g_P)$. The scalar curvature of the canonical metric g_P is then the sum of the following terms:

$$\begin{aligned}R_P = R_V &- \tfrac{1}{4}\mathrm{Tr}_{G_1}(\overset{(1)}{F},\overset{(1)}{F})_{g_V} - \tfrac{1}{4}\mathrm{Tr}_{G_2}(\overset{(2)}{F},\overset{(2)}{F})_{g_V} \\ &- \cdots - \tfrac{1}{4}\mathrm{Tr}_{G_k}(\overset{(k)}{F},\overset{(k)}{F})_{g_V} + \tfrac{1}{4}g_{G_1}^2 + \tfrac{1}{4}g_{G_2}^2 + \cdots + \tfrac{1}{4}g_{G_k}^2.\end{aligned} \tag{20}$$

The notations are obvious; $\overset{(i)}{F}$ means the projection of the curvature form F onto the Lie algebra $\mathscr{A}_{G_i}$.

Proof: Obvious, because

$$\mathscr{A}_G = \mathscr{A}_{G_1} \oplus \mathscr{A}_{G_2} \oplus \cdots \oplus \mathscr{A}_{G_k}. \tag{21}$$

Example 1: Let $V = M_4$, the Minkowskian space-time; consider the bundle $P(M_2, \mathrm{SU}(2)\times\mathrm{U}(1))$. Then

$$R_P = -\tfrac{1}{4}\mathring{g}_{ab}\mathring{g}^{ij}\mathring{g}^{kl}F^a_{ik}F^b_{jl} - \tfrac{1}{4}\mathring{g}^{ij}\mathring{g}^{kl}F_{ik}F_{jl} + \text{const}, \tag{22}$$

where $a,b = 1,2,3$; $\mathring{g}_{ab} = \delta_{ab}$, $\mathring{g}^{ij} = \mathrm{diag}(-,+,+,+)$,

$$F^a_{ij} = \partial_i B^a_j - \partial_j B^a_i + C^a_{bc}B^b_i B^c_j \tag{23}$$

is the field tensor of the SU(2) Yang–Mills field, whereas

$$F_{ij} = \partial_i A_j - \partial_j A_i \tag{24}$$

is the field tensor of the electromagnetic (abelian) field.

Theorem 2: Consider the principal fiber bundle $P(V,G,A;g_P)$ with a connection A and the canonical metric g_P. Denote by $\Delta_P = (d\delta + \delta d)_P$ the Laplace–Beltrami operator on P induced by the metric g_P. Then

(a) Let f be a scalar function on $P(V,G,A;g_P)$. Then

$$\Delta_P f = \mathring{g}^{ij}\mathscr{D}_i\mathscr{D}_j f + \mathring{g}^{ab}L_aL_b f, \tag{25}$$

where

$$\mathscr{D}_j f = \partial_j f - A^a_j L_a f, \tag{26}$$

L_a being the basis of the generators of $\mathscr{A}_G$ and $g^{ab}L_aL_b$ is the Casimir operator of $\mathscr{A}_G$.

(b) Let ϕ be a 1-form over $P(V,G,A;g_P)$; by means of the connection A and metric g_P it can be decomposed into horizontal and vertical parts, which in local coordinates will be denoted by (W_j,ϕ_a). If the form ϕ satisfies the covariance conditions

$$L_a\phi_b = C^d_{ab}\phi_d, \tag{27}$$

$$L_a W_j = 0, \tag{28}$$

then the equation $\Delta_P\phi = 0$ is equivalent to the following system:

$$\mathring{g}^{ij}\mathscr{D}_i\mathscr{D}_j\phi_c - \tfrac{1}{2}\mathring{g}^{ij}\mathring{g}^{kl}\mathring{g}_{bc}F^b_{ik}(\partial_j W_l - \partial_l W_j) = 0, \tag{29}$$

$$\mathring{g}^{ij}\partial_i\partial_j W_k + \mathring{g}^{ij}\mathscr{D}_j(F^b_{ik}\phi_b) = 0, \tag{30}$$

where

$$\mathscr{D}_j\phi_c = \partial_j\phi_c - A^b_j C^d_{bc}\phi_d. \tag{31}$$

(c) The equation

$$\Delta_P A = g_{ab}g^{ab}A = NA \tag{32}$$

$(N = \dim G)$

is equivalent to the Yang–Mills field equations

$$\mathring{g}^{ij}\mathscr{D}_i F^a_{jk} \equiv \mathring{g}^{ij}(\partial_i F^a_{jk} + C^a_{bc}A^b_i F^c_{jk}) = 0. \tag{33}$$

Proof: By calculus, which is quite cumbersome. The best way to simplify it is to use the an-holonomic system of coordinates on $P(V, G; A; g_P)$ in which g_P is diagonal, and the only nonvanishing components of the Christoffel connection are

$$\begin{aligned}&\Gamma^a_{bc} = \tfrac{1}{2}C^a_{bc}, \quad \Gamma^i_{aj} = \Gamma^i_{ja} = \tfrac{1}{2}\mathring{g}^{im}\mathring{g}_{ab}F^b_{jm}, \\ &\Gamma^a_{ij} = -\tfrac{1}{2}F^a_{ij}.\end{aligned} \tag{34}$$

The details may be found in other papers.[3–5] The case when the principal bundle is itself a Lie group and $V = P/G$ is a homogeneous space is discussed in Ref. 6.

Corollary 2.1: If the structure group is a product of several compact semisimple Lie groups, $G = G_1\times G_2\times\cdots\times G_k$, then for a function f on $P(V,G_1\times\cdots\times G_k,A;g_P)$

$$\Delta_P f = \mathring{g}^{ij}\mathscr{D}_i\mathscr{D}_j f + C_1 f + C_2 f + \cdots + C_k f, \tag{35}$$

where C_i is the Casimir operator of the group G_i.

Proof: Obvious, as $\mathscr{A}_G = \mathscr{A}_{G_1} \oplus \mathscr{A}_{G_2} \oplus \cdots \oplus \mathscr{A}_{G_k}$.

We see that with this formalism we can obtain the nonlinear equations by applying formally linear operations as Δ_P. Now let us proceed to further generalization.

3. MULTIPLE FIBER BUNDLES

The above construction can be repeated *ad infinitum* if we replace the original Riemannian manifold V by a principal fiber bundle endowed with a connection and the corresponding canonical metric g_P. In fact, when we consider vector fields or forms over $P(V,G)$, we are already working in the associate fiber bundle of frames, or a tangent bundle, over $P(V,G)$. What we shall investigate now is the case when the bundle over $P(V,G)$ is also a *principal* fiber bundle.

Symbolically, we write

$$P_2(P_1(V,G_1),G_2), \tag{36}$$

i.e., a principal fiber bundle P_2 with the structure group G_2 over a principal fiber bundle P_1 with the structure group G_1 over a Riemannian space V. We call this manifold a *double (principal) fiber bundle* over V with first structure group G_1 and second structure group G_2. [An interesting question arises immediately: under which conditions the bundle $P_2(P_1(V,G_1),G_2)$ has the structure of a principal fiber bundle over V with some structural group H? (dim H = dim G_1 + dim G_2). We know the answer only for the trivial case $(V\times G_1)\times G_2 = V\times(G_1\times G_2)$; the same is probably true for any $P(V,G_1)$ with V simply connected.]

Theorem 1 can be applied to the double fiber bundle (36), giving the following.

Corollary 1.2: Consider a double principal fiber bundle $P_2(P_1(V,G_1;A_1,g_{P_1}), G_2;A_2;g_{P_2})$ with the two connections A_1 and A_2, and two canonical metrics g_{P_1} on P_1 and g_{P_2} on P_2. The scalar curvature of the metric g_{P_2} decomposes as follows:

$$\begin{aligned} R_{P_2} &= R_{P_1} - \tfrac{1}{4}\mathrm{Tr}_{G_2}\overset{(2)}{(F},\overset{(2)}{F})_{g_{P_1}} + \tfrac{1}{4}g_{G_2}^2 \\ &= R_V - \tfrac{1}{4}\mathrm{Tr}_{G_1}\overset{(1)}{(F},\overset{(1)}{F})_{g_V} - \tfrac{1}{4}\mathrm{Tr}_{G_2}\overset{(2)}{(F},\overset{(2)}{F})_{g_{P_1}} + \tfrac{1}{4}g_{G_2}^2 + \tfrac{1}{4}g_{G_1}^2. \end{aligned} \tag{37}$$

This formula is easily generalized for a multiple fiber bundle $P_m(P_{m-1}(\cdots(P_1(V,G_1))\cdots),G_m)$.

$\overset{(1)}{F}$ is the curvature 2-form of the connection A_1 in the fiber bundle $P_1(V,G_1)$, whereas $\overset{(2)}{F}$ is the curvature 2-form of the connection A_2 in the fiber bundle $P_2(P_1(V,G_1),G_2)$. The horizontal components of the connection A_2 will be identified with new physical fields interacting with the gauge field A_1. As an example, we shall see how the Higgs–Kibble scalar multiplet may be introduced and interpreted geometrically.

Example 2: Consider the simplest case when $G_1 = \mathrm{U}(1)$, and G_2 is any compact and semisimple Lie group of dimension N. Then $\dim(P_1(M_4,\mathrm{U}(1)) = 5$, $\dim P_2(P_1(M_4,\mathrm{U}(1)),G_2) = N + 5$.

The Lie algebra of U(1) is isomorphic with R^1; there is no canonical metric on U(1), but we can replace it by any constant. Let us introduce the local coordinates in an open set of $P_1(M_4,\mathrm{U}(1))$ such that for $p\in P_1(M_4,\mathrm{U}(1))$

$$p = \{p^\alpha\} = \{x^i,x^5\}, \quad i = 0,1,2,3, \quad \alpha = 1,2,3,4,5.$$

In these coordinates the connection form in $P_1(M_4,\mathrm{U}(1))$ has the components A^5_i, A^5_5; its vertical part A^5_5 is just a constant (identified with the elementary charge e). The nonvanishing components of the curvature tensor are

$$F^5_{ij} = \partial_i A^5_j - \partial_j A^5_i \tag{38}$$

and

$$R_{P_1} = -\tfrac{1}{4}\mathring{g}^{ik}\mathring{g}^{jl}F^5_{ij}F^5_{kl} + \text{const.} \tag{39}$$

This is the well-known Kaluza–Klein theory.[7,8] Let us denote the connection form in $P_2(P_1(M_4,\mathrm{U}(1)),G_2)$ by B; in a local coordinate system it has the following components:

$$B^a_\beta = \{B^a_b, B^a_j, B^a_5\}, \quad \beta = \{b,j,5\}. \tag{40}$$

The components B^a_β are functions of x^i, x^5, and ξ^a.

The vanishing of the vertical components of $F^a_{\alpha\beta}$ means that

$$\begin{aligned} &\partial_a B^c_b - \partial_b B^c_a + C^c_{gh}B^g_a B^h_b = 0, \\ &\partial_a B^c_j - \partial_j B^c_a + C^c_{gh}B^g_a B^h_j = 0, \\ &\partial_a B^c_5 - \partial_5 B^c_a + C^c_{gh}B^g_a B^h_5 = 0. \end{aligned} \tag{41}$$

Local triviality of the bundle enables us to choose a coordinate system in which

$$\partial_i B^c_a = 0, \quad \partial_5 B^c_a = 0. \tag{42}$$

We assume that these relations are satisfied in what follows.

The nonvanishing (horizontal) components of the curvature are

$$F^a_{ij} = \partial_i B^a_j - \partial_j B^a_i + C^a_{gh}B^g_i B^h_j \tag{43}$$

and

$$F^a_{i5} = -F^a_{5i} = \partial_i B^a_5 - \partial_5 B^a_i + C^a_{bc}B^b_i B^c_5.$$

We assume that nothing depends explicitly on the cyclic variable x^5, i.e., $\partial_5 B^a_i = 0$. Therefore, denoting B^a_5 by ϕ^a we can rewrite

$$F^a_{i5} = \partial_i\phi^a + C^a_{bc}B^b_i\phi^c = \mathscr{D}_i\phi^a, \tag{44}$$

the "gauge covariant derivative" of the Higgs–Kibble scalar multiplet.

Suppose now that the first bundle $P_1(M_4,\mathrm{U}(1))$ is flat, i.e., that the connection form A_1 in $P_1(M_4,\mathrm{U}(1))$ was identically null. Then

$$\begin{aligned} R_{P_2} = &-\tfrac{1}{4}\mathring{g}_{ab}\mathring{g}^{ij}\mathring{g}^{kl}\overset{(2)}{F}{}^a_{ik}\overset{(2)}{F}{}^b_{jl} - \tfrac{1}{2}\mathring{g}_{ab}\mathring{g}^{ij}\mathscr{D}_i\phi^a\mathscr{D}_j\phi^b \\ &+ \text{const} \end{aligned} \tag{45}$$

(we put $g^{55} = 1$). This is the Lagrangian of the nonabelian gauge field interacting invariantly with the massless Higgs–Kibble scalar multiplet.

If the connection A_1 is not flat, our fields will interact with the abelian (electromagnetic) gauge field. The components of A_1 being denoted by A_j, we have

$$(g_{P_1}) = \begin{pmatrix} \mathring{g}^{ij} & -\mathring{g}^{ij}A_j \\ -\mathring{g}^{ij}A_i & 1+\mathring{g}^{ij}A_iA_j \end{pmatrix}, \tag{46}$$

therefore (45) is modified into

$$\begin{aligned} R_{P_2} = &-\tfrac{1}{4}\mathring{g}^{ij}\mathring{g}^{kl}\overset{(1)}{F}_{ik}\overset{(1)}{F}_{jl} - \tfrac{1}{4}\mathring{g}_{ab}\mathring{g}^{ij}\mathring{g}^{kl}\overset{(2)}{F}{}^a_{ik}\overset{(2)}{F}{}^b_{jl} \\ &- \tfrac{1}{2}\mathring{g}_{ab}\mathring{g}^{ij}\mathscr{D}_i\phi^a\mathscr{D}_j\phi^b - \tfrac{1}{2}\mathring{g}_{ab}\mathring{g}^{ij} \\ &\times A_iA_j\mathring{g}^{kl}\mathscr{D}_k\phi^a\mathscr{D}_l\phi^b + \text{const}. \end{aligned} \tag{47}$$

The nonlinear interaction modifies the characteristics; e.g., the abelian field does not propagate along the light cones when $\phi^a \neq 0$:

$$\Box A_j = (\mathring{g}_{ab}\mathring{g}^{kl}\mathscr{D}_k\phi^a\mathscr{D}_l\phi^b)A_j. \tag{48}$$

Let us consider a case of double fiber bundle with flat connection in the first bundle:

$$P_2(M_4\times G_1;G_2;A_2,g_{P_2}). \tag{49}$$

Let $\{\xi^a\}$ be the local coordinates in G_1, $a,b = 1,2,\ldots,N_1$, the generators of $\mathscr{A}_{G_1}$ being L_a,

$$[L_a,L_b] = C^c_{ab}L_c.$$

The horizontal components of the connection in P_2 will be denoted by B^A_i, B^A_a; the vertical components being B^A_D.

The nonvanishing components of the curvature tensor are now

$$\begin{aligned} F^A_{ij} &= \partial_i B^A_j - \partial_j B^A_i + C^A_{BC}B^B_iB^C_j, \\ F^A_{ib} &= -F^A_{bi} = \partial_i B^A_b - \partial_b B^A_i + C^A_{BC}B^B_iB^C_b, \\ F^A_{bd} &= \partial_b B^A_d - \partial_d B^A_b + C^A_{BC}B^B_bB^C_d, \end{aligned} \tag{50}$$

and the Lagrangian becomes

$$\begin{aligned} R_{P_2} = &-\tfrac{1}{4}\mathring{g}_{AB}\mathring{g}^{ij}\mathring{g}^{kl}F^A_{ik}F^B_{jl} - \tfrac{1}{2}\mathring{g}_{AB}\mathring{g}^{ij}\mathring{g}^{bd}F^A_{ib}F^B_{jd} \\ &-\tfrac{1}{4}\mathring{g}_{AB}\mathring{g}^{bc}\mathring{g}^{ad}F^A_{ba}F^B_{cd} + \text{const.} \end{aligned} \tag{51}$$

We assume that $\partial_b B^A_i = 0$ (the local triviality of the bundle plus the invariance properties of B with respect to the action of G_2 on P_2). On the other hand, different properties of B may be assumed with respect to the action of G_1 on $P_1(M_4,G_1)$. We shall suppose the simplest choice:
either

$$\partial_b B^A_d = 0,$$

or

$$\partial_b B^A_d = C^f_{gd}A^g_bB^A_f, \tag{52}$$

with A^g_b satisfying

$$\partial_a A^g_b - \partial_b A^g_a + C^g_{fh}A^f_aA^h_b = 0 \tag{53}$$

so that the operators $S^a_b\partial_a = L_b$ span $\mathscr{A}_{G_1}$; $S^a_bA^b_c = \delta^a_c$.

4. A MODEL WITH SPHERICAL SYMMETRY

Consider the case when $G_1 = G_2 = O(3)$, i.e.,

$$P_2(P_1(M_4,O(3)),O(3);A_2,g_{P_2}). \tag{54}$$

We shall suppose that $P_1(M_4,O(3)) = M_4\times O(3)$ and is flat (i.e., $A_1 \equiv 0$). A similar case has been considered by Forgács and Manton,[9] with a particular interpretation: the group G_2 was the gauge group generating the Yang–Mills field, and the group G_1 was the symmetry group of the subset of solutions under consideration.

Let $i,j = 0,1,2,3$; $a,b = 1,2,3$; $A,B = 1,2,3$. The structure constants of both Lie groups are the same, but the scale factor (coupling constant) may be different; therefore, let us put

$$C^A_{BD} = e\epsilon_{ABD}, \quad C^a_{bc} = \lambda\epsilon_{abc}, \tag{55}$$

where ϵ_{ABD} denotes the Levi–Civitá antisymmetric tensor.

We call the components B^A_j the Yang–Mills potential, and the components $B^A_b = \phi^A_b$ the Higgs–Kibble *tensor potential*. We want to see what the static and spherically symmetric solutions are like; in order to do so, let us consider the following ansatz (see, e.g., Refs. 10 and 11):

$$\begin{aligned} &B^A_0 \equiv 0, \quad B^A_j = \frac{\epsilon_{Ajk}x^k}{er^2}(1-K(r)), \quad j,k = 1,2,3 \\ &B^A_a = \frac{x^Ax_a}{er^3}H(r) + \frac{\delta^A_a}{er}G(r). \end{aligned} \tag{56}$$

We suppose that

$$\partial_aB^A_b = \lambda\epsilon_{cdb}A^d_aB^A_c \tag{57}$$

and A^d_a are left-invariant forms on G_1; we may as well write

$$F^A_{bc} = 2\lambda\epsilon_{bcd}B^A_d + e\epsilon_{ABD}B^B_bB^D_c. \tag{58}$$

The two cases of (52) correspond to putting λ equal to 0 or 1. The Lagrangian density (up to a constant)

$$-\mathscr{L} = r^2[\tfrac{1}{4}F^A_{ij}F^{ij}_A + \tfrac{1}{2}F^A_{ib}F^{ib}_A + \tfrac{1}{4}F^A_{bd}F^{bd}_A] \tag{59}$$

is equal to the following expression:

$$\begin{aligned} -\mathscr{L} = \Bigg[&\left(\frac{dK}{dr}\right)^2 + \frac{(1-K^2)^2}{2r^2} + \frac{(r\,dH/dr - H)^2}{2r^2} \\ &+ \frac{3(r\,dG/dr - G)^2}{2r^2} + \frac{(r\,dH/dr - H)(r\,dG/dr - G)}{r^2} \\ &+ \frac{H^2(1+K^2)}{r^2} + \frac{2G^2(1-K)^2}{r^2} + \frac{2GH(1-K)^2}{r^2} \\ &+ \frac{3G^4}{2r^2} + \frac{G^2H^2}{r^2} + \frac{2G^3H}{r^2} + \frac{6\lambda(G^2H+G^3)}{r} \\ &+ \lambda^2[H^2 + 2GH + 6G^2]\Bigg]. \end{aligned} \tag{60}$$

This expression is obviously positive definite ($\lambda > 0$). The Lagrangian is considerably simplified if $G = -H$; then (60) reduces to

$$\begin{aligned} -\mathscr{L} = \Bigg[&\left(\frac{dK}{dr}\right)^2 + \frac{(1-K)^2}{2r^2} + \frac{(r\,dH/dr - H)^2}{r^2} \\ &+ \frac{H^2(1+K^2)}{r^2} + \frac{H^4}{2r^2} + 5\lambda^2H^2\Bigg]. \end{aligned} \tag{61}$$

The condition of linear dependence between G and H may be interpreted as the fact that only the trace of B^A_a interacts with the gauge field B^A_j, the traceless part of B^A_a being irrelevant.

There is a striking similarity between (61) and the Lagrangian proposed by 't Hooft[12]:

$$L = -\tfrac{1}{4}F^A_{ij}F^{ij}_A - \tfrac{1}{2}\mathscr{D}_i\phi^A\mathscr{D}^i\phi_A - V(\phi), \tag{62}$$

where B^A_j was the same as in (56), and $\phi^A = (x^A/er^2)H(r)$; the potential $V(\phi)$ was introduced *ad hoc* and had the form

$$V(\phi) = \frac{\mu^2}{2}\phi^A\phi_A - \frac{\beta}{4e^2}(\phi^A\phi_A)^2. \tag{63}$$

In that case

$$\begin{aligned} -\mathscr{L} = \Bigg[&\left(\frac{dK}{dr}\right)^2 + \frac{(1-K^2)^2}{2r^2} + \frac{(r\,dH/dr - H)^2}{2r^2} \\ &+ \frac{H^2K^2}{r^2} + \frac{\beta}{4r^2}H^4 - \frac{\mu^2}{2}H^2\Bigg]. \end{aligned} \tag{64}$$

Our geometric construction of a gauge field over a multiple fiber bundle enabled us to introduce in a natural way the fourth-power terms in $H(r)$. In order to interpret our Lagrangian in the same spirit as 't Hooft's, we have to make λ pure

imaginary, $\lambda^2 < 0$, and identify $2\mu^2/\beta$ with $10\lambda^2$. Introducing an imaginary λ corresponds to the change in the signature in the Killing–Cartan metric of G_1. Another difference is the presence of the term H^2/r^2 in our Lagrangian. If we want the action integral (energy) to be finite, $H(r)$ has to behave asymptotically as $H(r) \approx Cr + h(r), h(r) \to 0$ when $r \to \infty$, with $C^2 = 2\mu^2/\beta = 10\lambda^2$. Then the term H^2/r^2 at infinity behaves like a constant C. In order to remove the linear divergence $\int C^2\, dr$ we are obliged to introduce a constant surface contribution, the physical interpretation of such a counter-term not being clear. The problem does not arise, of course, if the base manifold is compact.

It seems worthwhile to investigate more deeply the properties of the gauge fields over the multiple fiber bundles, including the topological properties of the solutions.

[1]R. Kerner, Ann. Inst. Henri Poincaré **9**, 143 (1968).
[2]Y. M. Cho, J. Math. Phys. **16**, 2029 (1975).
[3]R. Kerner, Ann. Inst. Henri Poincaré **34**, 437 (1981).
[4]R. Kerner, C. R. Acad. Sci. **272**, 760 (1971); see also J. Math. Phys. **21**, 2553 (1980).
[5]J. P. Bourguignon and H. B. Lawson, Jr., Commun. Math. Phys. **79**, 189 (1981).
[6]K. Parthasarathy, Ann. Math. **96**, 1 (1968).
[7]T. Kaluza, Sitr. Ber. Preuss. Akad. Wiss., 966 (1921).
[8]O. Klein, Z. Phys. **37**, 895 (1926).
[9]P. Forgács and N. Manton, Commun. Math. Phys. **72**, 15 (1980).
[10]R. Kerner, Lett. Math. Phys. **2**, 49 (1978).
[11]C. Gu and H. Hu, Commun. Math. Phys. **79**, 75 (1981).
[12]G. 't Hooft, Nucl. Phys. B **79**, 276 (1974).

Graded gauge theories over supersymmetric space

Richard Kerner
Université de Paris VI, Département de Mécanique, 4, Place Jussieu, 75005 Paris, France
Ema Manuela da Silva Maia
Departamento de Matemática, Universidade de Coimbra, 3000 Coimbra, Portugal

(Received 19 October 1981; accepted for publication 8 January 1982)

We generalize the usual gauge theories, as well as the supergauge theories, in the following way. We construct a graded group associated with a compact semisimple Lie group G. This graded group contains G and the linear space of anticommuting G-spinors on which G acts through a highly reducible representation. The graded group generalizes the notion of the super-Poincaré group. Next we construct a fiber bundle the basis of which is the superspace, the structural group being the graded group. Then we introduce the connection, curvature, and calculate the corresponding Yang–Mills Lagrangian. The nontrivial content of such a theory is put forward if we impose the Grassmann parity condition on our connection and curvature; we supposed here that both Grassmann parities (i.e., the one in the superspace and that in the graded group) add up to define the Grassmann parity of the corresponding field components. Together with the Hermiticity condition this supergauge leaves almost no room for arbitrariness in the expansion of the superconnection; it contains only the usual gauge field, the adjoint Higgs multiplet, and the spinor multiplet belonging to the spinorial representation of G. The conformal symmetry of the Lagrangian is broken, and the mass terms appear for the Higgs scalar and the spinor multiplet. The Yukawa and current–current interactions are also obtained, together with the Fermi four-point interaction term. The theory yields the ratio of the Higgs scalar mass versus the bare spinor mass equal to 27/40; the strengths of other couplings depend on the group via the decomposition of the spinor multiplet into the irreducible representations.

PACS numbers: 11.10.Np, 11.30.Pb, 02.20. + b

1. INTRODUCTION

Soon after the supersymmetric theories had been introduced by Wess, Zumino, Akulov, and others,[1–6] many authors proposed different extensions in order to include naturally the gauge field theories. Fiber bundles and connections have been constructed over the superspace[7,8]; the supergravity can also be viewed as a gauge theory with the super-Poincaré group taken as the structural group of a bundle over a Riemannian manifold V_4.[9,10] Usually a complete theory contained so many kinds of fields, all interacting between each other, that it was almost impossible to get any useful information without drastically reducing their number, namely by imposing more or less natural supergauge conditions which eliminated at least some of the ghost fields and simplified the couplings between the remaining fields.

Here we investigate a general graded gauge theory, in which the gauge field is considered as a connection in a principal fiber bundle over a superspace with structure group being a well-defined Z_2-graded extension of a compact, semisimple Lie group. The graded Lie groups (or supergroups) have been considered already by many authors (cf. Refs. 11 and 12). However, with the exception of the supergravity, no serious attempt has been made yet in order to calculate an effective Lagrangian of a gauge-field theory over the superspace, with structural group replaced by a Z_2-graded Lie group. We have carried out such a calculus as far as we could, imposing some supergauge conditions in order to keep a well-defined Grassmann parity of the fields as well as their Hermiticity. This had led to the breaking of the conformal symmetry and mass generation for the fermions and the Higgs bosons; another interesting feature appearing here is the purely geometrical deduction of highly nonlinear Yukawa and current–current couplings.

2. SUPERGROUP ASSOCIATED WITH A COMPACT, SEMISIMPLE LIE GROUP

Let G be a compact, semisimple Lie group of dimension N. Its structure constants being denoted by $C^c_{ab}, a,b = 1,2,...,N$, there is a nondegenerate and positive definite metric on G, the Killing–Cartan form $g_{ab} = C^d_{ac}\, C^c_{db}$; this metric defines a length of any vector from the Lie algebra $\mathscr{A}_G$ of the group G. The metric space $(\mathscr{A}_G, g_{ab})$ has its isometry group, which is SO(N), the real orthogonal group in N dimensions. The right action of G on itself induces the representation of G on the linear space $\mathscr{A}_G$, called the adjoint representation; this action preserves the Killing–Cartan metric too. The covering group of SO(N) is called Spin(N).[13] The lowest-dimensional faithful unitary representation of Spin(N) is often called the spin (or "double-valued") representation of SO(N). It also provides a (reducible) representation of G. Here is how this representation is constructed.

Consider N generators of the Clifford algebra associated with the Killing–Cartan metric on G, satisfying

$$\gamma_a\, \gamma_b + \gamma_b\, \gamma_a = 2g_{ab}\,\mathrm{Id}, \tag{1}$$

where Id means the identity matrix in the representation space. It is well known that the lowest-dimensional real matrix realization of γ_a 's is in $\mathrm{K}\times\mathrm{K}$ matrices, with $\mathrm{K} = 2^{[(N+1)/2]}$ ([l] meaning the entire part of l), whereas the

 0022-2488/83/020361-08$02.50

lowest-dimensional complex Hermitian realization is in J×J matrices, J = 2[N/2].

The $N(N-1)/2$ matrices $\sigma_{ab} = -\sigma_{ba}$, defined by

$$\sigma_{ab} = \tfrac{1}{8}(\gamma_a \gamma_b - \gamma_b \gamma_a), \tag{2}$$

generate the J-dimensional Hermitian (or K-dimensional real) representation of the Lie algebra $\mathscr{A}_{SO(N)} = \mathscr{A}_{Spin(N)}$. These matrices satisfy the commutation relations of $\mathscr{A}_{SO(N)}$:

$$[\sigma_{ab},\sigma_{cd}] = g_{ac}\sigma_{bd} + g_{bd}\sigma_{ac} - g_{bc}\sigma_{ad} - g_{ad}\sigma_{bc}. \tag{3}$$

We can raise and lower the indices a,b by means of the metric tensor g_{ab} and its inverse g^{ab}; we define $C_{abc} = g_{ad} C^{d}_{bc}$; for any compact and semisimple group G, C_{abc} are totally antisymmetric. It is easy to check that the matrices

$$\tau_a = \tfrac{1}{2} C_{abc} \sigma^{bc} \tag{4}$$

satisfy the commutation relations of $\mathscr{A}_G$:

$$[\tau_a,\tau_b] = C^{d}_{ab} \tau_d. \tag{5}$$

They provide a reducible K-dimensional representation of $\mathscr{A}_G$; by exponentiating them one obtains the corresponding representation of G. Let the indices $A,B,...$run from 1 to K; then, in matrix notation

$$[\tau_a] = \tau_a{}^{A}{}_{B}.$$

The element of the linear K-dimensional space will be called a G-spinor (group-spinor) and denoted by $u = u_A \chi^A$, χ^A being a local basis. From now on, by analogy with the supersymmetry formalism, we suppose that χ^A's are anticommuting quantities:

$$\chi^A \chi^B + \chi^B \chi^A = 0. \tag{6}$$

In the cases which we shall investigate for physical reasons, i.e., when $G = SU(2)$ or $G = SU(2)\times U(1)$, in the spinor space there exists a skew-symmetric tensor invariant under the action of isometries and defining a skew-symmetric inner product of two spinors. There always exists a choice of coordinates in which this antisymmetric form has the components

$$\epsilon_{12} = -\epsilon_{21} = 1, \quad \epsilon_{34} = -\epsilon_{43} = 1, \quad \text{etc.}, \tag{7}$$

all other components vanishing. (The dimension is always even.) ϵ_{AB} being nonsingular, there exists an inverse matrix ϵ^{BD} such that

$$\epsilon_{AB}\,\epsilon^{BD} = \delta^{D}_{A}. \tag{8}$$

We can now raise and lower the spinorial indices by means of the tensors ϵ_{AB} and ϵ^{AB}.[14]

Now we can introduce, along with $\chi_A = \epsilon_{AB}\chi^B$, the tensor

$$\tau^{a}_{AB} = g^{ab}\epsilon_{AD}\tau_b{}^{D}{}_{B}. \tag{9}$$

By construction,

$$T_2(\tau_a \tau_b) = K g_{ab}. \tag{10}$$

The matrices τ_a generalize in an obvious way the Pauli matrices in the case of $G = SU(2)$.

Because of the anticommutation relations (6) any function of χ^A is a finite polynomial of order $< K$. In order to define the derivation of functions depending on χ^A it is sufficient to put

$$\partial_A \chi^B = \delta^{B}_{A} \tag{11}$$

together with the (anti)-Leibniz rule

$$\partial_A(\chi^B \chi^C) = \delta^{B}_{A}\chi^C - \delta^{C}_{A}\chi^B. \tag{12}$$

Then

$$\partial_A \partial_B + \partial_B \partial_A = 0. \tag{13}$$

The group G acts on itself on the right: Any $h\in G$ induces a mapping $R_h: G\to G$ given by $R_h\, g = gh, g,h\in G$. This action generates the left-invariant of $\mathscr{A}_G$. Let us denote these fields by $S_a, a = 1,2,...,N$; in a local basis

$$S_a = S^{b}_{a}\partial_b. \tag{14}$$

The Lie brackets of S_a, S_b give

$$[S_a,S_b] = C^{d}_{ab} S_d \tag{15a}$$

or, more explicitly,

$$[S_a,S_b] = \pounds_{S_a} S_b = [S^{g}_{a}\partial_g S^{d}_{b} - S^{g}_{b}\partial_g S^{d}_{a}]\partial_d$$
$$= C^{g}_{ab} S^{d}_{g}\partial_d. \tag{15b}$$

Consider an infinitesimal transformation given by a linear superposition of the N generators:

$$\delta_g = \delta g^a S_a = \delta g^a S^{b}_{a}\partial_b. \tag{16}$$

There is a one-to-one correspondence between this generator and the following transformation in the space of spinors:

$$\chi^A \to \delta g^a \tau_a{}^{A}{}_{B}\chi^B. \tag{17}$$

The finite transformations generated by any $g\in G$ are obtained by exponentiation: If g_t belongs to a 1-parameter subgroup of G such that $g_0 = e$, $dg/dt|_{t=0} = \epsilon^a L_a$, then formally

$$D(g_t) = \exp[t\epsilon^a \tau_a]. \tag{18}$$

Let us denote the linear space of spinors by $\{\chi\}$. With the action of group G defined on $\{\chi\}$ via the representation (18), the direct product $G\times\{\chi\}$ acquires the structure of a semidirect product of G with the abelian group $\{\chi\}$ [let us remember that the linear space $\{\chi\}$ is isomorphic with the group of translations in $\{\chi\}$]. We denote this semidirect product by $G\square\{\chi\}$. The composition rule in $G\square\{\chi\}$ is given by the following formula: If (g_1,χ_1) and (g_2,χ_2) belong to $G\square\{\chi\}$, then their product is

$$(g_1,\chi_1)\cdot(g_2,\chi_2) = (g_1 g_2, \chi_1 + D(g_1)\chi_2). \tag{19}$$

$G\square\{\chi\}$ defined as above will be called the supergroup associated with a compact and semisimple Lie group G.

3. DIFFERENTIAL OPERATORS OVER $\mathscr{A}_{G\square\{\chi\}}$ AND SUPERALGEBRA

Let M_4 denote the Minkowskian space-time with the metric $g_{ij} = \text{diag}(-+++)$, $i,j = 0,1,2,3$. From now on we use the notation introduced by Wess and Zumino.[2,3] The Pauli matrices are denoted by $\sigma^{j\alpha}{}_{\dot\beta}, \alpha = 1,2; \dot\beta = \dot1,\dot2$; the spin representation of the Lorentz group acts in the linear space of spinors $\{\theta\}$, spanned by θ^α, $\bar\theta^{\dot\beta}$, satisfying

$$\theta^\alpha\theta^\beta + \theta^\beta\theta^\alpha = 0, \quad \theta^\alpha\bar\theta^{\dot\beta} + \bar\theta^{\dot\beta}\theta^\alpha = 0,$$
$$\bar\theta^{\dot\alpha}\bar\theta^{\dot\beta} + \bar\theta^{\dot\beta}\bar\theta^{\dot\alpha} = 0. \tag{20}$$

The direct product $M_4 \times \{\theta\}$ is called the *superspace*. The spinors $\{\theta\}$ transform under the isometries of M_4 via the spinorial representation, which leaves invariant the skew-symmetric product defined by the antisymmetric tensors $\epsilon_{\alpha\beta}, \epsilon_{\dot\alpha\dot\beta}$:

$$\epsilon_{12} = -\epsilon_{21} = 1, \quad \epsilon_{\dot1\dot2} = -\epsilon_{\dot2\dot1} = 1. \tag{21}$$

The indices $\alpha, \dot\beta$ are raised and lowered by means of $\epsilon_{\alpha\beta}, \epsilon_{\dot\alpha\dot\beta}$ and their inverse matrices $\epsilon^{\alpha\beta}, \epsilon^{\dot\alpha\dot\beta}$.

The derivations of the functions depending on $\{\theta\}$ are defined by

$$\begin{aligned} &\partial_\alpha\, \theta^\beta = \delta_\alpha^\beta, \quad \partial_\alpha\, \theta_\beta = \epsilon_{\alpha\beta}, \quad \partial_\alpha \bar\theta^{\dot\beta} = 0, \\ &\partial_{\dot\alpha}\, \bar\theta^{\dot\beta} = \delta_{\dot\alpha}^{\dot\beta}, \quad \partial_{\dot\alpha}\, \bar\theta_{\dot\beta} = \epsilon_{\dot\alpha\dot\beta}, \\ &\partial_\alpha\, (\theta^\beta \theta^\alpha) = \delta_\alpha^\beta \theta^\gamma - \delta_\alpha^\gamma\, \theta^\beta, \quad \text{etc.} \end{aligned} \tag{22}$$

Introducing the differential operators acting on functions on $M_4 \times \{\theta\}$ (considered as $\{\theta\}$-Grassmann algebra valued functions on M_4) as follows:

$$\begin{aligned} &\mathscr{D}_\alpha = \partial_\alpha + \sigma^j_{\alpha\dot\beta}\, \bar\theta^{\dot\beta} \partial_j, \\ &\overline{\mathscr{D}}_{\dot\beta} = \partial_{\dot\beta} + \sigma^j_{\alpha\dot\beta}\, \theta^\alpha\, \partial_j, \end{aligned} \tag{23}$$

which satisfy the anticommutation relations

$$\begin{aligned} &\{\mathscr{D}_\alpha, \mathscr{D}_\beta\}_+ = 0, \quad \{\overline{\mathscr{D}}_{\dot\alpha}, \overline{\mathscr{D}}_{\dot\beta}\}_+ = 0, \\ &\{\mathscr{D}_\alpha, \overline{\mathscr{D}}_{\dot\beta}\}_+ = 2\sigma^j_{\alpha\dot\beta}\, \partial_j, \end{aligned} \tag{24}$$

together with the Poincaré algebra generators defined as

$$\begin{aligned} &P_k = \frac{\partial}{\partial x^k}, \\ &T^{kl} = x^k \frac{\partial}{\partial x_l} - x^l \frac{\partial}{\partial x_k} + \sigma^{kl\alpha}{}_{\dot\beta}\, \bar\theta^{\dot\beta} \partial_\alpha + \sigma^{kl}{}_\alpha{}^{\dot\beta}\, \theta^\alpha\, \partial_{\dot\beta}, \end{aligned} \tag{25}$$

where

$$\sigma^{kl} = \tfrac{1}{8}(\gamma^k \gamma^l - \gamma^l \gamma^k), \tag{26}$$

and the Dirac matrices γ^k have the components given by

$$\begin{aligned} &(\gamma^k)^\alpha{}_\beta = 0, \quad (\gamma^k)^\alpha{}_{\dot\beta} = -\sigma^{k\alpha}{}_{\dot\beta}, \\ &(\gamma^k)^{\dot\alpha}{}_\beta = \sigma^k{}_\beta{}^{\dot\alpha}, \quad (\gamma^k)^{\dot\alpha}{}_{\dot\beta} = 0. \end{aligned} \tag{27}$$

We obtain the full graded Poincaré algebra (often called superalgebra)

$$\begin{aligned} &[P_k, P_m] = 0, \\ &[T^{kl}, P_m] = \delta^k_m\, P^l - \delta^l_m\, P^k, \\ &[T^{kl}, T^{mn}] = g^{km}\, T^{ln} + g^{ln}\, T^{km} - g^{lm}\, T^{kn} - g^{kn}\, T^{lm}, \\ &[T^{mn}, \mathscr{D}_\alpha] = \sigma^{mn}{}_\alpha{}^{\dot\beta} \overline{\mathscr{D}}_{\dot\beta}, \\ &[T^{mn}, \overline{\mathscr{D}}_{\dot\beta}] = \sigma^{mn\alpha}{}_{\dot\beta}{}^{\dot\alpha}, \\ &[P_k, \mathscr{D}_\alpha] = [P_k, \overline{\mathscr{D}}_{\dot\beta}] = 0, \\ &\{\mathscr{D}_\alpha, \mathscr{D}_\beta\}_+ = \{\overline{\mathscr{D}}_{\dot\alpha}, \overline{\mathscr{D}}_{\dot\beta}\}_+ = 0, \quad \{\mathscr{D}_\alpha, \overline{\mathscr{D}}_{\dot\beta}\}_+ = 2\sigma^j_{\alpha\dot\beta} P_j. \end{aligned} \tag{28}$$

By analogy, we shall construct the graded algebra of differential operators "tangent" to the supergroup $G \square \{\chi\}$.

In local coordinates the generators of $\mathscr{A}_G$ were represented by $L_a = S^b_a\, \partial_b$, with $[L_a, L_b] = C^d_{ab}\, L_d$. Let us denote by A^b_d the inverse matrix of S^b_a; then, as a direct consequence of (16),

$$\partial_a A^b_d - \partial_d A^b_a + C^b_{gf} A^g_a A^f_d = 0 \tag{29}$$

(the Maurer–Cartan identity). A general differential operator of first order defined over $G \square \{\chi\}$ has the following form:

$$\begin{aligned} &\mathscr{D}_{\dot a} = U^b_a\, \partial_b + U^A_a\, \partial_A, \\ &\mathscr{D}_B = U^b_B\, \partial_b + U^A_B\, \partial_A. \end{aligned} \tag{30}$$

We want our operators to have the well-defined Grassmann parity, i.e., $\mathscr{D}_a$ has to be even (only even powers of χ^A or ∂_A), $_{\dot B}$ has to be odd. Therefore, U^a_b, U^A_B have to be commuting quantities, whereas U^b_B and U^A_a are anticommuting quantities. Therefore, we can always choose a coordinate system in which

$$\begin{aligned} &\mathscr{D}_a = L_a + U_a{}^A{}_B\, \chi^B \partial_A, \\ &\mathscr{D}_A = \partial_A + U^b{}_{AD}\, \chi^D L_b, \end{aligned} \tag{31}$$

where $U_a{}^A{}_B, U^b{}_{AD}$ are commuting quantities (depending on $g \in G$). If we now put

$$U_a{}^A{}_B = \tau_a{}^A{}_B, \tag{32}$$

then

$$[\mathscr{D}_a, \mathscr{D}_b] = C^d_{ab}\, \mathscr{D}_d \tag{33}$$

because $\tau_a{}^A{}_B$ are representation of $\mathscr{A}_G$, too. Also

$$\mathscr{D}_A\, \mathscr{D}_B + \mathscr{D}_B\, \mathscr{D}_A = (U^b_{AB} + U^b_{BA}) L_b = C^b_{AB}\, L_b. \tag{34}$$

To close our algebra, we have to calculate

$$\mathscr{D}_a\, \mathscr{D}_B - \mathscr{D}_B\, \mathscr{D}_a. \tag{35}$$

We postulate

$$\mathscr{D}_a\, \mathscr{D}_B - \mathscr{D}_B\, \mathscr{D}_a = -\tau_a{}^D{}_B\, \mathscr{D}_D. \tag{36}$$

In other words, we have to define a set of generalized structure constants which would provide us with the adjoint representation of the graded algebra: $\mathscr{A}_{G \square \{\chi\}}$.

The only nonvanishing structure constants are

$$C^a_{bc} = -C^a_{cb}, \tag{37}$$

$$C_a{}^D{}_B = -C^D_{Ba} = -\tau_a{}^D{}_B, \tag{38}$$

$$C^a_{BC} = C^a_{CB} = (U^a_{BC} + U^a_{CB}) = g^{ab}\, \epsilon_{BD}\, \tau_b{}^D{}_C. \tag{39}$$

If we introduce generalized indices ϕ, ψ denoting both a, B, and if the Grassmann parity $\pi(\phi)$ is defined as $\pi(a) = 0$, $\pi(B) = 1$, then the generalized Jacobi identity (the adjoint representation of $\mathscr{A}_{G \square \{\chi\}}$) reads

$$C^\phi_{\chi\psi}\, C^\psi_{\Omega\Delta} - (-1)^{\pi(\chi)\pi(\Omega)}\, C^\phi_{\Omega\psi}\, C^\psi_{\chi\Delta} = C^\psi_{\chi\Omega}\, C^\phi_{\psi\Delta}. \tag{40}$$

Explicitly, this will give only one new identity

$$C^d_{ab}\, C^b_{BD} - C^d_{BG}\, C_a{}^G{}_D = C_a{}^G{}_B\, C_G{}^d{}_D, \tag{41}$$

which defines implicitly the spinor metric ϵ_{BD} when combined with the definition

$$C^a_{BC} = 2g^{ab}\, \epsilon_{BD}\, \tau_b{}^D{}_C, \tag{42}$$

$$g^{df}(\epsilon_{DE}\, \tau_f{}^E{}_G\, \tau_a{}^G{}_B + \epsilon_{BE}\, \tau_f{}^E{}_G\, \tau_a{}^G{}_D) = C^d_{ab}\, g^{bf}\, \epsilon_{BE}\, \tau_f{}^E{}_D. \tag{43}$$

We leave aside the problem of the uniqueness of the spinor "metric" ϵ_{BE}; we know how to construct it in the simplest case when $G = \mathrm{SU}(2)$ or $\mathrm{SU}(2) \times \mathrm{U}(1)$. We close this paragraph with a table which gives the correspondence between the notions of supergroup and its graded Lie algebra associated with a compact and semisimple Lie group G, and the usual notions of the Poincaré superalgebra and the su-

TABLE I. Correspondence between the superspace and the graded Poincaré algebra and the graded group G.

M_4, Minkowskian space time	G, a semisimple compact Lie group
g_{ij} Minkowskian metric	$g_{ab} = C^d_{ac} C^c_{db}$, the Cartan–Killing metric
γ_i Dirac matrices	γ_a generators of the Clifford algebra induced by g_{ab}
$\theta^\alpha, \bar{\theta}^{\dot{\beta}}$ Anticommuting Dirac spinors	χ^A, anticommuting G-spinors
$\mathscr{P}^0$ Poincaré group acting on the Minkowskian space–time	G acting on itself
P_k, J_{km} generators of the Poincaré algebra	$L_a = S^b_a \partial_b$ generators of the left-invariant vector fields on G
$\sigma_{km} = \frac{1}{8}(\gamma_k\gamma_m - \gamma_m\gamma_k)$ generators of the spinorial representation of $\mathscr{P}^0$	$\tau_a = \frac{1}{2} C_{abc} \sigma^{bc}$ generators of the spinorial representation of $\mathscr{A}_G$
$P_k, J_{km}, \mathscr{D}_\alpha, \bar{\mathscr{D}}_{\dot{\beta}}$ generators of the graded Poincaré algebra $\mathscr{P}^0 \oplus \mathscr{P}^1$	$\mathscr{D}_a, \mathscr{D}_B$, generators of the graded algebra $\mathscr{A}_{G\square\{\chi\}}$
$M^4 \times \{\theta\}$, the superspace	Graded group (or supergroup) $G\square\{\chi\}$
$\epsilon_{\alpha\beta}, \epsilon_{\dot{\alpha}\dot{\beta}}$, invariant "metric" in the space of Dirac spinors.	ϵ_{AB}, invariant "metric" in the space of G-spinors

perspace $M_4 \times \{\theta\}$.

The essential difference between the Minkowskian superspace and the graded group is the fact that our Lie group G has the double role of the symmetry group of the space and the space on which it acts itself, whereas the Minkowskian space has no group structure, being only the space on which the isometry (Poincaré) group acts.[15]

4. GRADED FIBER BUNDLES, GRADED SUPERCONNECTIONS

The classical theory of nonabelian gauge fields is formulated as the theory of connections in a principal fiber bundle $P(M_4,G)$. Consider an open set $U \subset P(M_4,G)$ isomorphic with a direct product of $\pi(U) \subset M_4$ and G; we can introduce local coordinates in U such that $p \in U$ has the components (x^k, ξ^a) with $(x^k) \in \pi(U)$, $(\xi^a) \in G$.

A connection in $P(M_4,G)$ is defined by a left-invariant 1-form A of type ad with values in the Lie algebra $\mathscr{A}_G$; in our coordinate system we can decompose A as follows:

$$A = A^a L_a = (A^a_j \, dx^j + A^a_b \, d\xi^b) L_a. \tag{44}$$

The covariant exterior differential of A is a 2-form of type ad called the curvature:

$$F = DA = dA + \tfrac{1}{2}[A, A]_{\mathscr{A}_G}. \tag{45}$$

F is horizontal, which in our coordinates means that

$$\begin{aligned} &\partial_a A^b_d - \partial_d A^b_a + C^b_{gf} A^g_a A^f_d = 0, \\ &\partial_j A^b_d - \partial_d A^b_j + C^b_{gf} A^g_j A^f_d = 0 \end{aligned} \tag{46}$$

$(a,b = 1,2,\ldots,N;\ i,j,k = 0,1,2,3)$. (In a local trivialization we can always put $\partial_j A^b_d \equiv 0$.)

The only nonvanishing (horizontal) components of F are

$$F^a_{ij} = \partial_i A^a_j - \partial_j A^a_i + C^a_{bc} A^b_i A^c_j \tag{47}$$

and are identified with the Yang–Mills field tensor.

The Lagrangian of the theory is given by

$$\mathscr{L} = -\tfrac{1}{4} g_{ab} \, g^{ij} g^{kl} F^a_{ik} F^b_{jl}, \tag{48}$$

and the variational principle $\delta \int \mathscr{L} \, d^4x = 0$ yields the equation

$$g^{ij} (\partial_i F^a_{jk} + C^a_{bc} A^b_i F^c_{jk}) = 0. \tag{49}$$

Introducing the superspace and the supergroup defined in Sec. 2 leads to the following three natural generalizations of this scheme:

$$\begin{array}{ccc} P(M_4 \times \{\theta\}, G\square\{\chi\}) & \to & P(M_4, G\square\{\chi\}) \\ \downarrow & & \downarrow \\ P(M_4 \times \{\theta\}, G) & \to & P(M_4, G) \\ \downarrow & & \downarrow \\ M_4 \times \{\theta\} & \to & M_4 . \end{array}$$

We may call the cases represented here an ordinary principal bundle over space-time, $P(M_4,G)$; ordinary principal bundle over the superspace, $P(M_4 \times \{\theta\}, G)$; a graded fiber bundle over space–time, $P(M_4, G\square\{\chi\})$; and a graded fiber bundle over the superspace, $P(M_4 \times \{\theta\}, G\square\{\chi\})$.

Our final goal is the theory of connections (gauge fields) in $P(M_4 \times \{\theta\}, G\square\{\chi\})$; but first we shall have a glance at the connections in $P(M_4 \times \{\theta\}, G)$, because already here we encounter the main characteristics of such a generalization.[16]

In $P(M_4 \times \{\theta\}, G)$, we decompose all geometrical objects in the local frame in which the basis directions are

$$e_j = \partial_j, \quad e_\alpha = \mathscr{D}_\alpha, \quad e_{\dot{\beta}} = \bar{\mathscr{D}}_{\dot{\beta}}, \quad e_d = L_d. \tag{50}$$

The connection form A is naturally decomposed, as before, into the vertical and horizontal components:

$$\begin{aligned} &\text{vertical part} \quad A^a_b, \\ &\text{horizontal part} \quad A^d_m = \{A^d_j, A^d_\alpha, A^d_{\dot{\beta}}\}, \end{aligned} \tag{51}$$

where we use the capital latin indices K,L,M to denote the indices $(j,\alpha,\dot{\beta})$. The generalized curvature form has the following nonvanishing (horizontal) components:

$$F^a_{KL} = \mathscr{D}_K A^a_L - (-1)^{\pi(K)\pi(L)} \mathscr{D}_L A^a_K + C^a_{bd} A^b_K A^d_L, \tag{52}$$

where $\mathscr{D}_K$ mean ∂_j, $\mathscr{D}_\alpha$, $\bar{\mathscr{D}}_{\dot{\beta}}$ and $\pi(K)$ is the Grassmann parity of K, $\pi(j) = 0$, $\pi(\alpha) = \pi(\dot{\beta}) = 1$.

The definition (52) has to be implemented by the condition on the Grassmann parity of A^b_K, i.e., $\pi(A^b_K) = \pi(K)$, in order to preserve the correlation between the spin and statistics and eliminate the so-called ghost fields. In that case,

$$\begin{aligned} &F^a_{jk} = -F^a_{kj}, \\ &F^a_{\alpha j} = -F^a_{j\alpha}, \quad F^a_{j\dot{\beta}} = -F^a_{\dot{\beta}j} \\ &F^a_{\alpha\beta} = F^a_{\beta\alpha}, \quad F^a_{\dot{\alpha}\dot{\beta}} = F^a_{\dot{\beta}\dot{\alpha}}, \quad F^a_{\dot{\alpha}\beta} = F^a_{\dot{\beta}\alpha}. \end{aligned} \tag{53}$$

The Lagrangian of the theory now becomes

$$\mathscr{L} = -\tfrac{1}{4} g_{ab} g^{KL} g^{MN} F^a_{KM} F^b_{LN}, \tag{54}$$

where g^{KL} stays for the matrix:

$$[g^{KL}] = \begin{pmatrix} g^{ij} & 0 & 0 \\ 0 & \epsilon^{\alpha\beta} & 0 \\ 0 & 0 & \epsilon^{\dot\alpha\dot\beta} \end{pmatrix}. \tag{55}$$

Finally, $\mathscr{L}$ has to be integrated over $P(M_4\times\{\theta\},G)$; the integration with respect to the Haar measure dG will yield only a constant because of the invariance properties of an *ad*-type form; we have to integrate then over $M_4\times\{\theta\}$. The integration with respect to the anticommuting variables has been introduced by Berezin[17] and is defined by

$$\int d\theta^{\alpha}=0,\quad \int d\bar\theta^{\dot\beta}=0,\quad \int \theta_{\alpha}d\theta^{\beta}=\delta_{\alpha}^{\beta},\quad \int \bar\theta_{\dot\alpha}d\bar\theta^{\dot\beta}=\delta_{\dot\alpha}^{\dot\beta}. \tag{56}$$

This definition enables us to calculate the integral of any polynomial in θ's. By construction, our Lagrangian is an even polynomial; therefore, symbolically,

$$\mathscr{L}=\mathscr{L}_0+\mathscr{L}_{\alpha\beta}\theta^{\alpha}\theta^{\beta}+\mathscr{L}_{\alpha\dot\beta}\theta^{\alpha}\bar\theta^{\dot\beta}+\mathscr{L}_{\dot\alpha\dot\beta}\bar\theta^{\dot\alpha}\bar\theta^{\dot\beta} + \mathscr{L}_{\alpha\beta\dot\gamma\dot\delta}\theta^{\alpha}\theta^{\beta}\bar\theta^{\dot\gamma}\bar\theta^{\dot\delta}. \tag{57}$$

After integration over the "volume element" of $\{\theta\}$, i.e., with respect to $d\theta^{1}d\theta^{2}d\bar\theta^{1}d\bar\theta^{2}$, only the last term will leave a nonvanishng contribution.

The components A_K^a are polynomials in the variables $\theta^{\alpha},\bar\theta^{\dot\beta}$ with coefficients depending on x^k; the dependence on the group variables is completely determined by the equations generalizing (46). We can develop

$$A_j^a(x,\theta)=B_j^a(x)+B_{j\alpha}^a(x)\theta^{\alpha}+B_{j\dot\beta}^a(x)\bar\theta^{\dot\beta}+B_{j\alpha\beta}^a\theta^{\alpha}\theta^{\beta} + \cdots+B_{j\alpha\beta\dot\gamma\dot\delta}^a\theta^{\alpha}\theta^{\beta}\bar\theta^{\dot\gamma}\bar\theta^{\dot\delta}, \tag{58a}$$

$$A_{\alpha}^d(x,\theta)=\psi_{\alpha}^d(x)+\phi^d(x)\theta_{\alpha}+V_{j\alpha}^d\sigma_{\beta\dot\gamma}^j\theta^{\beta}\bar\theta^{\dot\gamma}+\cdots,\text{etc.} \tag{58b}$$

In order to eliminate the wrong relations between spin and statistics we impose the condition on the Grassmann parity of components[18]:

$$\pi(A_K^d)=\pi(K). \tag{59}$$

This automatically excludes a lot of terms, e.g., in $A_j^a(x,\theta)$ we can have the first term $B_j^a(x)$ but not the next one, $B_{j\alpha}^a\theta^{\alpha}$, which is odd; similarly, in $A_{\alpha}^a(x,\theta)$ the term $\phi^a(x)\theta_{\alpha}$ has the required parity, whereas the spinor multiplet $\psi_{\alpha}^a(x)$ has not, etc.

Moreover, some of the terms will not contribute to the Lagrangian because their order in θ is already too high [e.g., the last term in (58a)].

Any particular choice of the development of A_I^a which does not contain all possible powers of θ's will break the supersymmetry invariance; however, such a "supergauge" fixing conserves both the classical gauge and the Lorentz invariance. Some choices break the conformal invariance and may introduce a mass of the gauge field, or at least of some of its components. Let us illustrate this by an example. The simplest connection A_I^a that is Hermitian and obeys the Grassmann parity rule (59) is

$$A_j^b\equiv 0,\quad A_{\alpha}^b=\phi^b(x)\theta_{\alpha},\quad A_{\dot\beta}^b=\phi^b(x)\bar\theta_{\dot\beta}. \tag{60}$$

Then the only nonvanishing components of the curvature tensor are

$$F_{j\alpha}^a=\partial_j\phi^a\theta_{\alpha},\quad F_{j\dot\beta}^a=\partial_j\phi^a\bar\theta_{\dot\beta},\quad F_{\alpha\beta}^a=\sigma_{\alpha\dot\delta}^j\bar\theta^{\dot\delta}\partial_j\phi^a\theta_{\beta}, \quad F_{\alpha\dot\beta}^a=\sigma_{\alpha\dot\delta}^j\bar\theta^{\dot\delta}\partial_j\phi^a\bar\theta_{\dot\beta},\quad F_{\alpha\dot\beta}^a=\sigma_{\gamma\dot\alpha}^j\theta^{\gamma}\partial_j\phi^a\bar\theta_{\dot\beta}, \tag{61}$$

and the Lagrangian is equal to

$$\mathscr{L}=-\tfrac{1}{2}g_{ab}g^{jk}\partial_j\phi^a\partial_k\phi^b\theta^1\theta^2\bar\theta^1\bar\theta^2 + \text{lower power terms in }\theta\text{'s.} \tag{62}$$

If we choose $A_j^a=B_j^a(x)$, then the gauge invariance will impose the modification in A_{α}^a, $A_{\dot\beta}^a$, namely,

$$A_j^a=B_j^a(x),\quad A_{\alpha}^a=\phi^a\theta_{\alpha}+\sigma_{\alpha\dot\beta}^j\bar\theta^{\dot\beta}B_j^a,\quad A_{\dot\beta}^a=\phi^a\bar\theta_{\dot\beta}+\sigma_{\alpha\dot\beta}^j\theta^{\alpha}B_j^a, \tag{63}$$

and our Lagrangian is equal to

$$\mathscr{L}=[-\tfrac{1}{4}G_{ij}^aG_{kl}^bg_{ab}g^{ik}g^{jl}-\tfrac{1}{2}\nabla_i\phi^a\nabla_j\phi^bg_{ab}g^{ij}]\theta^1\theta^2\bar\theta^1\bar\theta^2 + \text{lower order terms in }\theta, \tag{64}$$

where

$$G_{ij}^a=\partial_iB_j^a-\partial_jB_i^a+C_{bc}^aB_i^bB_j^c,\qquad \nabla_j\phi^a=\partial_j\phi^a+C_{bc}^aB_j^b\phi^c. \tag{65}$$

Of course, the ansatz (63) is not the most general one satisfying the conditions of Hermiticity and of Grassmann parity; we shall discuss later more general forms of connections. We proceed now to the definition of a connection form in $P(M_4\times\{\theta\},G\square\{\chi\})$, its curvature form, and the Lagrangian.

Let us denote by greek letters ϕ,χ,ψ the "vertical" indices, i.e., (a,b,c,A,B); the Grassmann parity is $\pi(\phi)=0$ if $\phi=a,b$ and $\pi(\phi)=1$ if $\phi=A,B$. By capital latin letters we denote the "horizontal" indices $\alpha,\dot\beta$, and j,k; $\pi(L)=0$ if $L=j,k$ and $\pi(L)=1$ if $L=\alpha$ or $\dot\beta$. The connection 1-form has the components A_{ψ}^{ϕ} and A_L^{ϕ}; more explicitly,

$$A_{\psi}^{\phi}=\{A_b^a,A_B^a,A_D^B,A_b^B\} \tag{66}$$

and

$$A_K^{\phi}=\{A_j^a,A_{\alpha}^a,A_{\dot\beta}^a,A_j^B,A_{\alpha}^B,A_{\dot\beta}^B\}. \tag{67}$$

The generalized left-invariance property of this 1-form leads to the vanishing of the vertical components of the curvature 2-form, which is defined as

$$F_{\psi\Omega}^{\phi}=\mathscr{D}_{\psi}A_{\Omega}^{\phi}-(-1)^{\pi(\psi)\pi(\Omega)[1+\pi(\phi)]}\mathscr{D}_{\Omega}A_{\psi}^{\phi} + C_{\chi\Delta}^{\phi}A_{\psi}^{\chi}A_{\Omega}^{\Delta}=0. \tag{68}$$

The nonvanishing structure constants $C_{\Omega\Delta}^{\phi}$ are given by (43).

The vertical components of $F_{\chi\psi}^{\phi}$ are the ones in which one of the lower indices χ,ψ is vertical, i.e., takes on one of the values a, A, or B. The fact that F is horizontal gives the equations of the type

$$F_{ab}^{\phi}=\mathscr{D}_aA_b^{\phi}-\mathscr{D}_bA_a^{\phi}+C_{\chi\psi}^{\phi}A_a^{\chi}A_b^{\psi}=0, \tag{69}$$

$$F_{Ab}^{\phi}=\mathscr{D}_AA_b^{\phi}-(-1)^{\pi(\phi)}\mathscr{D}_bA_A^{\phi}+C_{\chi\psi}^{\phi}A_A^{\chi}A_b^{\psi},$$

etc., which may be regarded as definition of the "vertical" components of A, i.e., (66). The nonvanishing components of

F are the following:

$$F^a_{ij},\quad F^a_{\alpha j},\quad F^a_{\alpha j},\quad F^a_{\alpha\beta},\quad F^a_{\alpha\dot\beta},\quad F^a_{\dot\alpha\dot\beta},$$
$$F^A_{ij},\quad F^A_{\alpha j},\quad F^A_{\alpha\beta},\quad F^A_{\alpha\dot\beta},\quad F^A_{\dot\alpha j},\quad F^A_{\dot\alpha\dot\beta}, \tag{70}$$

with

$$F^\phi_{KL} = \mathscr{D}_K A^\phi_L - (-1)^{\pi(K)\pi(L)[1+\pi(\phi)]} \times \mathscr{D}_L A^\phi_K + C^\phi_{\chi\psi} A^\chi_K A^\psi_L. \tag{70a}$$

The Lagrangian will have the same form as previously, except for the normalization factor which we shall fix after:

$$\mathscr{L} = -\frac{1}{2^s} g_{\phi\psi} g^{KL} g^{PS} F^\phi_{KP} F^\psi_{LS}, \tag{71}$$

the "horizontal" indices α,β, j being denoted by K, P, L, etc.

Now we can proceed farther to compute $\mathscr{L}$ effectively.

5. SUPERGAUGE CONDITIONS AND THE CONFORMAL SYMMETRY BREAKING

Let us generalize the Grassmann-parity conditions for our superconnection. As a matter of fact, these conditions are contained implicitly in the definition of the supercurvature (68), namely, we suppose

$$\pi(A^\phi_\psi) = \pi(\phi) + \pi(\psi),$$
$$\pi(A^\phi_L) = \pi(\phi) + \pi(L), \tag{72}$$

where $\pi(a) = \pi(j) = 0$, $\pi(A) = \pi(B) = \pi(\alpha) = \pi(\dot\beta) = 1$. Therefore, the components A^B_j are odd, A^B_α are even, A^a_α odd, A^a_j even, etc. The left-invariance conditions generalized for the graded group $G\square\{\chi\}$ mean that there exists a coordinate system in which the components of A^ψ_L do not depend on χ (which corresponds to their invariance with respect to the translations in χ − space). Together with the conditions (72) this eliminates most of the terms in the generalized expansion (58). If we require the Hermiticity of the components of our connection, i.e., $(A^a_\alpha)^+ = A^a_\alpha$, etc., then only the following terms will remain:

$$A^a_j = B^a_j(x) + \frac{1}{l}\phi^a \sigma_{j\alpha\dot\beta}\theta^\alpha\bar\theta^{\dot\beta},$$
$$A^a_\alpha = \phi^a(x)\theta_\alpha + A^a_j \sigma^j_{\alpha\dot\beta}\bar\theta^{\dot\beta},$$
$$A^a_{\dot\beta} = \phi^a(x)\bar\theta_{\dot\beta} + A^a_j \sigma^j_{\alpha\dot\beta}\theta^\alpha = (A^a_\beta)^+, \tag{73}$$
$$A^B_j = 0,$$
$$A^B_\alpha = \psi^B_\alpha(x) + \frac{\lambda}{l}(\psi^B_\gamma\theta^\gamma + \bar\psi^B_{\dot\gamma}\bar\theta^{\dot\gamma})\theta_\alpha + \frac{\nu}{l^2}\psi^B_\alpha\theta^1\theta^2\bar\theta^{\dot1}\bar\theta^{\dot2}.$$

Before proceeding farther let us note that we have introduced a dimensional constant l (a length scale), as well as two dimensionless parameters λ and ν. The length scale l occurs in a natural way, because we want A^a_j to have a definite dimension (namely, 1/cm); our potentials B^a_j and ϕ^a have the dimension 1/cm too, whereas the variables θ^α have the dimension $\mathrm{cm}^{1/2}$; the σ-matrices are dimensionless. The spinor field ψ has the dimension $\mathrm{cm}^{-1/2}$. After computing the components of the curvature tensor F^ϕ_{KL} we shall use the length factors when adding different components squared forming the Lagrangian density, e.g.,

$$\dim[F^a_{ij}] = \mathrm{cm}^{-2},$$
$$\dim[F^a_{ij}F^{ij}_a] = \mathrm{cm}^{-4},$$
$$\dim[F^a_{\alpha\beta}] = \mathrm{cm}^{-1},$$
$$\dim[F^a_{\alpha\beta}F^{\alpha\beta}_a] = \mathrm{cm}^{-2}, \tag{74}$$

so we have to take $F^a_{ij}F^{ij}_a + (1/l^2)F^a_{\alpha\beta}F^{\alpha\beta}_a$, etc. With the simplest ansatz (63) the resulting Lagrangian density is homogeneous in $1/l^2$, and the corresponding equations are conformally invariant. This is not the case for our supergauge (73); the conformal invariance is broken.

The Lagrangian density we postulate is then

$$\mathscr{L} = -\frac{1}{2^s}\left[F^a_{ij}F^{ij}_a + \frac{2}{l}F^a_{\alpha j}F^{\alpha j}_a + \frac{1}{l^2}F^a_{\alpha\beta}F^{\alpha\beta}_a + F^A_{ij}F^{ij}_A + \frac{2}{l}F^A_{j\alpha}F^{j\alpha}_A + \frac{1}{l^2}F^A_{\alpha\beta}F^{\alpha\beta}_A\right], \tag{75}$$

where, for simplicity, we did not distinguish between the dotted and undotted indices. All the indices are raised or lowered by means of the corresponding "metrics," i.e., g^{ij} for the space-time vectors and tensors, $\epsilon^{\alpha\beta}$ for the Lorentz spinors, g^{ab} for the group-algebra vectors, and ϵ^{AB} for the group spinors. In the final stages of the calculus, only the terms proportional to $\theta^1\theta^2\bar\theta^{\dot1}\bar\theta^{\dot2}$ in $\mathscr{L}$ will be of importance, because all the lower powers of θ will vanish when integrated over the volume element $d\theta^1 d\theta^2 d\bar\theta^{\dot1} d\bar\theta^{\dot2}$. This relevant term in the Lagrangian density turns out to be the following:

$$\mathscr{L}_{(4)} = -\frac{1}{2^s}\left[8G^a_{ij}G^{ij}_a + 64\nabla_j\phi^a\nabla^j\phi_a + \frac{16}{l^2}\phi^a\phi_a + \frac{4\lambda^2+8\lambda-4\nu}{l}\nabla_j\bar\psi\nabla^j\psi + \frac{8\lambda\nu}{l^3}\bar\psi\psi + \frac{(8\lambda^2+4\nu)}{l^2}(\bar\psi\gamma^j\nabla_j\psi - (\gamma^j\nabla_j\bar\psi)\psi) - \frac{8\lambda}{l}\nabla_j\phi^a C_{aBD}\bar\psi^B\gamma^j\psi^D - \frac{4\nu}{l^2}\phi^a C_{aBD}\bar\psi^B\psi^D + \frac{8\nu-8\lambda^2}{l^2}C^a_{BD}C_{aEF}\bar\psi^B\psi^D\bar\psi^E\psi^F\right]\theta^1\theta^2\bar\theta^{\dot1}\bar\theta^{\dot2}. \tag{76}$$

The following abbreviated notations have been used here:

$$G^a_{ij} = \partial_i B^a_j - \partial_j B^a_i + C^a_{bc}B^b_i B^c_j,$$
$$\nabla_j\phi^a = \partial_j\phi^a + C^a_{bd}B^b_j\phi^d,$$
$$\nabla_j\psi^A_\alpha = \partial_j\psi^A_\alpha + \tau_a{}^A{}_D B^a_j\psi^D_\alpha,$$
$$\bar\psi\psi = \epsilon_{AB}(\epsilon^{\alpha\beta}\psi^A_\alpha\psi^B_\beta + \epsilon^{\dot\alpha\dot\beta}\bar\psi^A_{\dot\alpha}\bar\psi^B_{\dot\beta}). \tag{77}$$

In what follows, we shall use even more symbolical notation when there is no risk of misinterpretation.

The remarkable thing about this Lagrangian is that the supergauge conditions, while keeping the Lorentz and the gauge invariances intact, have broken the conformal symmetry, introducing masses of the Higgs multiplet and of the spinor multiplets. The price to pay is the presence of the d'Alembert-type term for spinors $\nabla_j\bar\psi\nabla^j\psi$, as well as the four-point interactions, which may lead to the unrenormalizability of the theory.

Before identifying the masses of ϕ and ψ fields we have to fix the scale factors for our fields. As we want to have the term $G^a_{ij}G^{ij}_a$ enter with the usual factor $-\frac{1}{4}$, we must put $2^s = 32$; then we have to interpret the Higgs multiplet as given by the expression $\tilde{\phi}^a = 2\phi^a$. We can also absorb one of the two dimensionless parameters λ and ν into the definition of the Dirac spinors, just by fixing $(2\lambda^2 + \nu)/8 = 1$, which amounts to the rescaling of ψ. Then the following expression is obtained:

$$\begin{aligned}\mathscr{L}_{(4)} = -\Big[&\tfrac{1}{4}G^a_{ij}G^{ij}_a + \tfrac{1}{2}\nabla_j\tilde{\phi}^a\nabla^j\tilde{\phi}_a + \frac{1}{8l^2}\tilde{\phi}^a\tilde{\phi}_a \\ &+ \frac{(3\lambda^2+2\lambda-8)}{8l}\nabla_j\bar{\psi}\nabla^j\psi + \frac{\lambda(4-\lambda^2)}{2l^3}\bar{\psi}\psi \\ &+ \frac{1}{l^2}(\bar{\psi}\gamma^j\nabla_j\psi - (\gamma^j\nabla_j\bar{\psi})\psi) - \frac{\lambda}{8l}\nabla_j\tilde{\phi}^a C_{aBD}\bar{\psi}^B\psi^D \\ &+ \frac{(\lambda^2-4)}{8l^2}\tilde{\phi}^a C_{aBD}\bar{\psi}^B\psi^D \\ &+ \frac{8-3\lambda^2}{4l^2}C^a_{BD}C_{aEF}\bar{\psi}^B\psi^D\bar{\psi}^E\psi^F\Big]. \end{aligned} \tag{78}$$

It seems reasonable to restrain our theory by eliminating the second-order derivatives of spinors from the equations of motion; in other words, we want to make $8 - 3\lambda^2 - 2\lambda$ disappear. That gives us two solutions: $\lambda_1 = -2, \lambda_2 = \frac{4}{3}$.

The corresponding Lagrangians are

$$\begin{aligned}\mathscr{L} = -\Big[&\tfrac{1}{4}G^a_{ij}G^{ij}_a + \tfrac{1}{2}\nabla_j\tilde{\phi}^a\nabla^j\tilde{\phi}_a + \frac{1}{8l^2}\tilde{\phi}^a\tilde{\phi}_a \\ &+ \frac{1}{l^2}(\bar{\psi}\gamma^j\nabla_j\psi - (\gamma^j\nabla_j\bar{\psi})\psi) + \frac{1}{4l}\nabla_j\tilde{\phi}^a C_{aBD}\bar{\psi}^B\gamma^j\psi^D \\ &- \frac{1}{l^2}C^a_{EF}C_{aBD}\bar{\psi}^E\psi^F\bar{\psi}^B\psi^D\Big] \quad \text{for } \lambda_1 = -2 \end{aligned} \tag{79}$$

and

$$\begin{aligned}\mathscr{L} = -\Big[&\tfrac{1}{4}G^a_{ij}G^{ij}_a + \tfrac{1}{2}\nabla_j\tilde{\phi}^a\nabla^j\tilde{\phi}_a + \frac{1}{8l^2}\tilde{\phi}^a\tilde{\phi}_a \\ &+ \frac{40}{27l^3}\bar{\psi}\psi + \frac{1}{l^2}(\bar{\psi}\gamma^j\nabla_j\psi - (\gamma^j\nabla_j\bar{\psi})\psi) \\ &- \frac{1}{6l}\nabla_j\tilde{\phi}^a C_{aBD}\bar{\psi}^B\gamma^j\psi^D - \frac{5}{18l^2}\tilde{\phi}^a C_{aBD}\bar{\psi}^B\psi^D \\ &+ \frac{2}{3l^2}C^a_{BD}C_{aEF}\bar{\psi}^B\psi^D\bar{\psi}^E\psi^F\Big] \quad \text{for } \lambda_2 = \tfrac{4}{3}. \end{aligned} \tag{80}$$

In the first case the Lagrangian describes an invariant interaction between (massless) gauge fields, the massive Higgs field $\tilde{\phi}^a$ whose mass is $\mu_1 = 1/2l$, and a reducible multiplet of massless Dirac spinors, which interact with $\tilde{\phi}$ via the current–current coupling; there is a four-point interaction present, as was often postulated for the weak interaction neutrinos.

In the second case spinors acquire the mass equal to $\mu_\psi = 20/27l$; the current–current coupling and the four-point interaction are still there, but with different coefficients than before; finally, the Yukawa coupling between $\tilde{\phi}$ and ψ appears, too.

6. CONCLUSIONS

We have a geometrical method of deriving a gauge and Lorentz-invariant Lagrangian which describes the interaction between the gauge bosons, an adjoint representation multiplet of Higgs scalars, and a reducible multiplet of Dirac spinors. By imposing a supergauge condition which seems reasonable enough, we eliminate all the ghost fields, breaking at the same time the conformal symmetry and introducing masses for Higgs scalars and spinors. Still, we are quite far away from any physical interpretation if we do not perform a group representation-theory analysis of our expressions. Although that is not in the scope of this paper, we shall at least show what we mean by this.

Probably the most interesting feature of the graded gauge presented here is the fact that the fermion multiplets belong to some *well-defined representations* which are imposed by the geometry itself, whereas usually the choice of the representations to which the fermions belonged was exterior to the geometrical content of the theory. The gauge fields and the Higgs field always belong to the adjoint representation of G by construction, whereas no constraint was imposed on the representations of spinor multiplets.

Let us consider the simplest case $G = \mathrm{SU}(2)$. As $\dim \mathrm{SU}(2) = 3$, $K = 2^{[3/2]} = 2$; therefore we have an irreducible doublet of Dirac spinors. In this case the Lagrangian (80) may be interpreted as the isospin-invariant Lagrangian of the nuclear forces, the couple of Dirac spinors representing proton and neutron, ϕ^a representing three pions.

The case of $G = \mathrm{SU}(3)$ is more interesting. Then $\dim \mathrm{SU}(3) = 8$, $K = 2^{[8/2]} = 16$; so the spinor multiplet belongs to a 16-dimensional reducible representation. By the construction $\mathscr{A}_G \to ad\,\mathscr{A}_G \to \mathrm{SO}(N)$, this representation decomposes into two octets $8 \oplus \bar{8}$. (The invariant metrics ϵ_{AB} have to be compared with the invariant Cartan–Killing metrics in *ad* spaces). In general, our representation will contain $2^{[h/2]}$ times the representation δ_G, where h is the dimension of the Cartan subalgebra of G, δ_G being the highest-weight ($\frac{1}{2}$ sum of the roots) irreducible representation of G found in Spin $[\mathrm{SO}(N)]$.[19] As we see, the quark representations are not found here. Further generalization, including more irreducible representations of G, may probably be obtained by enlarging the notion of the supergroup so that it would contain not only $G \square \{\chi\}$, but also all the polynomials of $\{\chi\}$, i.e., the whole Grassmann algebra of the anticommuting variables χ^A: $G \square \Lambda\{\chi\}$.

We cannot take very seriously the "universal length" parameter l; however, the mass ratio $\mu_\phi/\mu_\psi = \frac{27}{40}$ seems to be encouraging if we think of the simplest example of nuclear forces, i.e., if we identify ψ^A with the couple proton–neutron and ϕ^a with the three pions. The orders of magnitude of the Yukawa couplings and of the current–current coupling seem to be then quite realistic, too. In either case it must be underlined that the masses are just unrenormalized quantities, and have to be modified if we take into account the dynamical terms. Only then will we obtain some more realistic picture in which the masses will split.

ACKNOWLEDGMENTS

One of the authors (R.K.) acknowledges a very enlight-

ening discussion with B. Zumino at CERN, as well as very useful talks with V. Rittenberg, Y. Ne'eman, and R. C. King. Both authors are very grateful to Y. M. Cho for many enlightening talks and friendly advice.

[1]D. V. Volkov and V. P. Akulov, Phys. Lett. B **46**, 109 (1973).
[2]J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974).
[3]J. Wess and B. Zumino, Nucl. Phys. B **78**, 1 (1974).
[4]A. Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974).
[5]A. Kotecky, Rep. Math. Phys. **8** (1975).
[6]K. Gawedzki, Ann. Inst. Henri Poincaré **27**, 335 (1977).
[7]R. Kerner, in *Group Theoretical Methods in Physics*, edited by Sharp and Kolman (Academic, New York, 1977), p. 321.
[8]R. Arnowitt, P. Nath, and B. Zumino, Phys. Lett. B **56**, 81 (1975).
[9]D. Z. Freedman, P. Van Nieuwenhuizen, and S. Ferrara, Phys. Rev. D **15**, 3214 (1976).
[10]S. Deser and B. Zumino, Phys. Lett. B **62**, 335 (1976).
[11]M. S. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1627 (1976).
[12]A. Rogers, J. Math. Phys. **22**, 939 (1981); see also L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975); R. W. Tucker, J. Math. Phys. **22**, 422 (1981).
[13]Spin(3) = SU(2), Spin(4) = SU(2)×SU(2), Spin(5) = U Sp(4), Spin(6) = SU(4), but for $N > 6$ Spin(N) is not isomorphic with any classical group (i.e., unitary, unimodular, orthogonal, or symplectic matrices).
[14]For higher-dimensional groups we can also construct the tensor like (7), but the invariance of the induced interior product with respect to the group acting through its spinorial representation is by no means obvious.
[15]The superspace $M_4 \times \{\theta\}$ and the supergroup $G \square \{\chi\}$ are not the same as the superspace of Refs. 1–4 or the supergroups of Ref. 12, the difference being that the even nilpotent part of the Grassmann algebra is not used in the approach of this paper, but is conventionally included.
[16]E. M. da Silva Maia, Lett. Math. Phys. **5(6)** (1981).
[17]F. A. Berezin and G. I. Kac, Math. USSR-Sb. **11**, 311 (1970).
[18]Suggestions concerning this postulate appear in R. Haag, J. Lopuszanski, and M. Sohnius, Nucl. Phys. B **88**, 257 (1975); C. N. Ktorides, in *Group Theoretical Methods in Physics*, edited by Sharp and Kolman (Academic, New York, 1977), p. 329. P. van Nieuwenhuizen, J. Math. Phys. **21**, 2562 (1980).
[19]R. C. King and P. Sorba (private communication).

Normalization of the three-body Bethe–Salpeter wavefunction for protons

Yukio Tomozawa
Randall Laboratory of Physics, University of Michigan, Ann Arbor, Michigan 48109

(Received 3 August 1982; accepted for publication 23 September 1982)

We discuss the normalization condition for a three-body Bethe–Salpeter amplitude and apply the result to the relativistic wavefunction for protons.

PACS numbers: 11.10.St, 11.10.Qr

I. INTRODUCTION

The normalization of a Bethe–Salpeter (BS) wavefunction[1] which describes a relativistic bound system is uniquely determined and has been a subject of many investigations for two-body bound states.[2,3] While the generalization to many particle bound states is straightforward, the importance of the normalization condition for three- or more-body BS wavefunctions can hardly be over emphasized in light of rapid developments in the quark model of hadrons. In fact, it was an important ingredient in a computation of the proton decay rate in grand unified gauge theories.[4]

In this article, we formulate the normalization condition for the BS wavefunction of three-quark bound states in Sec. II and apply it to the proton wavefunction in Sec. III. In Appendices A and B, the residue formula for the bound state is obtained, and the normalization for the three-body BS wave function for constituents with unequal masses is derived in Appendix C.

II. NORMALIZATION CONDITION FOR THREE-BODY BS WAVEFUNCTIONS

In this section, we shall formulate the normalization condition for a three-body bound system, which serves to define our notation. In doing so, we shall closely follow the derivation of Ref. 3.

The three-body propagator for fermion fields $\psi^A(x)$, $\psi^B(x)$, and $\psi^C(x)$ (with masses m_A, m_B, and m_C, respectively),

$$K(x_1,x_2,x_3,x_4,x_5,x_6)\equiv K(1,2,3;4,5,6) = -\langle 0|T\psi^A(x_1)\psi^B(x_2)\psi^C(x_3)\bar{\psi}^A(x_4)\bar{\psi}^B(x_5)\bar{\psi}^C(x_6)|0\rangle \tag{2.1}$$

satisfies the integral equation[5]

$$K(1,2,3;4,5,6) = S_F'^A(1,4)S_F'^B(2,5)S_F'^C(3,6) - \int \prod_{k=7}^{12} d^4x_k\, S_F'^A(1,7)S_F'^B(2,8)S_F'^C(3,9) \times G(7,8,9;10,11,12)K(10,11,12;4,5,6), \tag{2.2}$$

where

$$S_F'^A(1,2) = \langle 0|T\psi^A(x_1)\bar{\psi}^A(x_2)|0\rangle = \frac{1}{(2\pi)^4}\int S_F'^A(p)e^{ip(x_1-x_2)}d^4p, \tag{2.3}$$

and $G(1,2,3;4,5,6)$ is the irreducible kernel for the three-body propagator. Inserting a complete set of states $\{|p,\alpha\rangle\}$, we have

$$K(1,2,3;4,5,6) = -\sum_{p,\alpha}\chi_{p\alpha}(1,2,3)\bar{\chi}_{p\alpha}(4,5,6) \tag{2.4}$$

for $t_1,t_2,t_3 > t_4,t_5,t_6$, where

$$\chi_{p\alpha}(1,2,3) = \langle 0|T\psi^A(x_1)\psi^B(x_2)\psi^C(x_3)|p,\alpha\rangle \tag{2.5}$$

and

$$\bar{\chi}_{p\alpha}(1,2,3) = \langle p,\alpha|T\bar{\psi}^A(x_1)\bar{\psi}^B(x_2)\bar{\psi}^C(x_3)|0\rangle = -\chi^*_{p\alpha}(1,2,3)(\gamma_4)^A(\gamma_4)^B(\gamma_4)^C \tag{2.6}$$

are the BS amplitudes. For the bound state wavefunction $\chi_{p\alpha}(1,2,3)$ with momentum p ($p^2 = -M^2$), we have the BS equation

$$\chi_{p\alpha}(1,2,3) = -\int \prod_{k=7}^{12} d^4x_k\, S_F'^A(1,7)S_F'^B(2,8)S_F'^C(3,9) \times G(7,8,9;10,11,12)\chi_{p\alpha}(10,11,12). \tag{2.7}$$

In order to separate the center of mass coordinate and the internal relative coordinates, we use the following variables (assuming that the three particles have the same mass for simplicity):

$$X = \tfrac{1}{3}(x_1+x_2+x_3),\quad \xi = x_1-x_2,\quad \eta = \tfrac{1}{2}(x_1+x_2-2x_3), \tag{2.8a}$$

and their conjugate momenta

$$p = p_1+p_2+p_3,\quad p_\xi = \tfrac{1}{2}(p_1-p_2),\quad p_\eta = \tfrac{1}{3}(p_1+p_2-2p_3). \tag{2.8b}$$

These variables satisfy the condition

$$p_1x_1+p_2x_2+p_3x_3 = pX+p_\xi\xi+p_\eta\eta, \tag{2.9}$$

and the Jacobian of the transformations (2.8) is unity. The discussion for the unequal mass case will be given in Appendix C.

Using translational invariance, we define the Fourier transforms of K, G, and $\chi_{p\alpha}$ by

$$K(1,2,3;1',2',3') = (2\pi)^{-20}\int K(p_\xi,p_\eta;p'_\xi,p'_\eta;p) \times\exp\{i[p(X-X')+p_\xi\xi+p_\eta\eta-p'_\xi\xi'-p'_\eta\eta'] \times d^4pd^4p_\xi d^4p_\eta d^4p'_\xi d^4p'_\eta\}, \tag{2.10}$$

$$G(1,2,3;1',2',3') = (2\pi)^{-20}\int G(p_\xi,p_\eta;p'_\xi,p'_\eta;p) \times\exp\{i[p(X-X')+p_\xi\xi+p_\eta\eta-p'_\xi\xi'-p'_\eta\eta'] \times d^4pd^4p_\xi d^4p_\eta d^4p'_\xi d^4p'_\eta\}, \tag{2.11}$$

and

 0022-2488/83/020369-06$02.50

$$\chi_{p\alpha}(1,2,3) = \sqrt{M/E_p}\, e^{ipX}\chi_{p\alpha}(\xi,\eta) = \sqrt{M/E_p}\, e^{ipX}(2\pi)^{-8} \times \int \chi_{p\alpha}(p_\xi, p_\eta) e^{i(p_\xi\xi + p_\eta\eta)}\, d^4p_\xi d^4p_\eta, \tag{2.12}$$

where

$$E_p = \sqrt{\mathbf{p}^2 + M^2}. \tag{2.13}$$

Then, Eqs. (2.2) and (2.7) can be written as

$$\int \frac{d^4p''_\xi d^4p''_\eta}{(2\pi)^8} [I(p_\xi, p_\eta; p''_\xi p''_\eta; p) + G(p_\xi, p_\eta; p''_\xi, p''_\eta; p)] \times K(p''_\xi, p''_\eta; p'_\xi, p'_\eta; p) = (2\pi)^8 \delta(p_\xi - p'_\xi)\delta(p_\eta - p'_\eta) \tag{2.14}$$

and

$$\int \frac{d^4p''_\xi d^4p''_\eta}{(2\pi)^8} [I(p_\xi, p_\eta; p''_\xi, p''_\eta; p) + G(p_\xi, p_\eta; p''_\xi, p''_\eta; p)] \times \chi_{p\alpha}(p''_\xi, p''_\eta) = 0 \tag{2.15}$$

or, in short,

$$[I(p) + G(p)]K(p) = 1 \tag{2.16}$$

and

$$[I(p) + G(p)]\chi_p = 0, \tag{2.17}$$

where

$$I(p_\xi, p_\eta; p'_\xi, p'_\eta; p) = (2\pi)^8 \delta(p_\xi - p'_\xi)\delta(p_\eta - p'_\eta) \times [S_F^{A'}(\tfrac{1}{3}p + p_\xi + \tfrac{1}{2}p_\eta) S_F^{B'}(\tfrac{1}{3}p - p_\xi + \tfrac{1}{2}p_\eta) \times S_F^{C'}(\tfrac{1}{3}p - p_\eta)]^{-1}. \tag{2.18}$$

We also have the equations conjugate to Eqs. (2.16) and (2.17),

$$K(p)[I(p) + G(p)] = 1 \tag{2.19}$$

and

$$\bar{\chi}_p [I(p) + G(p)] = 0. \tag{2.20}$$

As is derived in Appendix A, Eq. (2.4) for the bound state can be written as

$$\lim_{p_0 \to E_p} (p_0 - E_p) K(p_\xi, p_\eta; p'_\xi, p'_\eta; p) = -i(M/E_p)\chi_{p\alpha}(p_\xi, p_\eta)\bar{\chi}_{p\alpha}(p'_\xi, p'_\eta) \tag{2.21}$$

or, in short,

$$\lim_{p_0 \to E_p} (p_0 - E_p) K(p) = -i(M/E_p)\chi_p \bar{\chi}_p. \tag{2.22}$$

Again following the method of Ref. 3, we define

$$Q(p) = \lim_{p_0 \to E_p} (p_0 - E_p) K(p) \frac{\partial}{\partial p_0} [I(p) + G(p)] = 1 - \lim_{p_0 \to E_p} \left(\frac{\partial}{\partial p_0} [(p_0 - E_p)K(p)] \right) [I(p) + G(p)], \tag{2.23}$$

where Eq. (2.19) has been used. The use of Eqs. (2.20) and (2.22) enables us to obtain

$$Q(p)\chi_p = \chi_p \tag{2.24}$$

$$= -i \frac{M}{E_p} \chi_p \bar{\chi}_p \left(\frac{\partial}{\partial p_0} [I(p) + G(p)] \right) \chi_p. \tag{2.25}$$

Thus, we get

$$-i\bar{\chi}_p \left(\frac{\partial}{\partial p_0} [I(p) + G(p)] \right) \chi^p = \frac{p_0}{M} \quad (p_0 = E_p) \tag{2.26}$$

or, in the full expression,

$$-\frac{i}{(2\pi)^{16}} \int d^4p_\xi d^4p_\eta d^4p'_\xi d^4p'_\eta\, \bar{\chi}_{p\alpha}(p_\xi, p_\eta) \times \left(\frac{\partial}{\partial p_0} [I(p_\xi, p_\eta; p'_\xi, p'_\eta; p) + G(p_\xi, p_\eta; p'_\xi, p'_\eta; p)] \right) \times \chi_{p\alpha}(p'_\xi, p'_\eta) = \frac{p_0}{M} \quad (p_0 = E_p). \tag{2.27}$$

For the ladder approximation

$$S_F^{A'}(q)^{-1} = S_F^A(q)^{-1} = i(i\gamma q + m_A) \tag{2.28}$$

and

$$\frac{\partial}{\partial p_0} G(p) = 0, \tag{2.29}$$

and hence

$$\frac{\partial}{\partial p_0} I(p_\xi, p_\eta; p'_\xi, p'_\eta; p) = \frac{i}{3}(2\pi)^8 \delta(p_\xi - p'_\xi)\delta(p_\eta - p'_\eta) J(p_\xi, p_\eta; p), \tag{2.30}$$

where

$$J_{abc,a'b'c'}(p_\xi, p_\eta; p) = (\gamma_4)_{aa'}(i\gamma(\tfrac{1}{3}p - p_\xi + \tfrac{1}{2}p_\eta) + m_B)_{bb'}(i\gamma(\tfrac{1}{3}p - p_\eta) + m_C)_{cc'} + (i\gamma(\tfrac{1}{3}p + p_\xi + \tfrac{1}{2}p_\eta) + m_A)_{aa'}(\gamma_4)_{bb'}(i\gamma(\tfrac{1}{3}p - p_\eta) + m_C)_{cc'} + (i\gamma(\tfrac{1}{3}p + p_\xi + \tfrac{1}{2}p_\eta) + m_A)_{aa'}(i\gamma(\tfrac{1}{3}p - p_\xi + \tfrac{1}{2}p_\eta) + m_B)_{bb'} \times (\gamma_4)_{cc'}, \tag{2.31}$$

a, b, c and a', b', c' being spinor indices. The final form of the normalization condition is then given by

$$\frac{1}{3} \int \frac{d^4p_\xi d^4p_\eta}{(2\pi)^8} \bar{\chi}_{p\alpha}(p_\xi, p_\eta) J(p_\xi, p_\eta; p) \chi_{p\alpha}(p_\xi, p_\eta) = \frac{p_0}{M} \quad (p_0 = E_p). \tag{2.32}$$

The normalization for a three-body BS wavefunction, Eq. (2.27) or (2.32), may be compared with that for the two-body case, which is given by

$$-i \int \frac{d^4q d^4q'}{(2\pi)^8} \bar{\chi}_{p\alpha}(q) \frac{\partial}{\partial p_0} [I_2(q,q'; p) + G_2(q,q'; p)] \chi_{p\alpha}(q') = 2p_0 (p_0 = E_p) \tag{2.33}$$

or

$$-\frac{i}{2} \int \frac{d^4q}{(2\pi)^4} \bar{\chi}_{p\alpha}(q) J_2(q;p) \chi_{p\alpha}(q) = 2p_0 \quad (p_0 = E_p), \tag{2.34}$$

where

$$I_2(q,q'; p) = (2\pi)^4 \delta(q - q') [S_F^{A'}(\tfrac{1}{2}p + q) S_F^{B'}(\tfrac{1}{2}p - q)]^{-1} \tag{2.35}$$

and

$$J_2(q,p)_{ab,a'b'} = (\gamma_4)_{aa'}(i\gamma(\tfrac{1}{2}p-q)+m_B)_{bb'} + (i\gamma(\tfrac{1}{2}p+q)+m_A)_{aa'}(\gamma_4)_{bb'}. \tag{2.36}$$

III. APPLICATION TO THREE-QUARK WAVEFUNCTIONS FOR OCTET BARYONS

The BS wavefunction for an octet baryon is expressed as[6]

$$\langle 0|T(\psi^i_{a\alpha}(x_2)\psi^j_{b\beta}(x_2)\psi^k_{c\gamma}(x_3))|p\rangle = \sqrt{M/E_p}\,\epsilon^{ijk}\tfrac{1}{2}(\chi^{(\xi)}_{abc}U^{(\xi)}_{\alpha\beta\gamma}+\chi^{(\eta)}_{abc}U^{(\eta)}_{\alpha\beta\gamma})\psi_p(\xi,\eta)e^{ipX}, \tag{3.1}$$

where i,j,k are $SU_c(3)$ color indices (running from 1 to 3), a,b c are spinor indices (running from 1 to 4), and α,β,γ are ordinary SU(3) indices (running from 1 to 3). The Levi-Civita symbol ϵ^{ijk} represents the color singlet nature of hadrons and the spin wavefunctions

$$\chi^{(\xi)}_{abc} = ([(-i\gamma p+M)/2M]\gamma_5 C)_{ab}u_c(p), \tag{3.2}$$

$$\chi^{\eta}_{abc} = (1/\sqrt{3})(\chi^{(\xi)}_{bca}-\chi^{(\xi)}_{cab})$$

and the SU(3) wavefunctions

$$U^{(\xi)}_{\alpha\beta\gamma} = \epsilon_{\alpha\beta\delta}B^{\delta}_{\gamma}, \tag{3.3}$$

$$U^{(\eta)}_{\alpha\beta\gamma} = (1/\sqrt{3})(U^{(\xi)}_{\beta\gamma\alpha}-U^{(\xi)}_{\gamma\alpha\beta})$$

are constructed in order to make the baryon behave as an SU(3) octet and satisfy the Bargmann–Wigner equation.[7] In Eq. (3.2), C is the charge conjugation matrix and satisfies the conditions

$$C^+C=1, \quad C^T=-C, \quad C^{-1}\gamma_\mu C=-\gamma_\mu^T, \tag{3.4}$$

and B^{β}_{α} in Eq. (3.3) is a symbolic notation for the 3×3 octet matrix (e.g., B^3_1 = proton). By construction, the spin and SU(3) wavefunctions satisfy the relations

$$\chi^{(\xi)}_{abc} = -\chi^{(\xi)}_{bac}, \quad \chi^{(\eta)}_{abc}=\chi^{(\eta)}_{bac}, \tag{3.5}$$

$$\chi^{(\rho)}_{abc}+\chi^{(\rho)}_{bca}+\chi^{(\rho)}_{cab}=0, \quad \rho=\xi,\eta,$$

and

$$U^{(\xi)}_{\alpha\beta\gamma} = -U^{(\xi)}_{\beta\alpha\gamma}, \quad U^{(\eta)}_{\alpha\beta\gamma}=U^{(\eta)}_{\beta\alpha\gamma}, \tag{3.6}$$

$$U^{(\rho)}_{\alpha\beta\gamma}+U^{(\rho)}_{\beta\gamma\alpha}+U^{(\rho)}_{\gamma\alpha\beta}=0, \quad \rho=\xi,\eta.$$

The BS wavefunction for the proton [notice that $U^{(\xi)}_{121}=B^3_1$ and $U^{(\eta)}_{121}=(1/\sqrt{3})B^3_1$] is given by

$$\langle 0|T(\psi^i_{a1}(x_1)\psi^j_{b2}(x_2)\psi^k_{c1}(x_3)|p\rangle \equiv \langle 0|T(u^i_a(x_1)d^j_b(x_2)u^k_c(x_3))|p\rangle = \sqrt{M/E_p}\,\chi^{ijk}_{p,abc}(\xi,\eta)e^{ipX}, \tag{3.7}$$

where

$$\chi_{p\alpha}(\xi,\eta)\equiv\chi^{ijk}_{p,abc}(\xi,\eta) = \epsilon^{ijk}\tfrac{1}{2}[\chi^{(\xi)}_{abc}-\tfrac{1}{3}(\chi^{(\xi)}_{bca}-\chi^{(\xi)}_{cab})]\,\psi_p(\xi,\eta) = \epsilon^{ijk}\tfrac{1}{3}(\chi^{(\xi)}_{abc}-\chi^{(\xi)}_{bca})\,\psi_p(\xi,\eta). \tag{3.8}$$

Defining the Fourier transform

$$\psi_p(\xi,\eta)=\frac{1}{(2\pi)^8}\int\phi_p(p_\xi,p_\eta)e^{ip_\xi\xi+ip_\eta\eta}\,d^4p_\xi d^4p_\eta \tag{3.9}$$

and noticing that

$$\bar{\chi}^{ijk}_{p,abc}(p_\xi,p_\eta) = -(\chi^{ijk}_{p,a'b'c'}(p_\xi,p_\eta))^*(\gamma_4)_{a'a}(\gamma_4)_{b'b}(\gamma_4)_{c'c} = e^{ijk}\tfrac{1}{2}\left[\left(C^{-1}\gamma_5\frac{-i\gamma p+M}{2M}\right)_{ab}\bar{u}_c(p) - \left(C^{-1}\gamma_5\frac{-i\gamma p+M}{2M}\right)_{bc}\bar{u}_a(p)\right]\cdot\phi^*_p(p_\xi,p_\eta), \tag{3.10}$$

we can compute the integrand of Eq. (2.32) as follows:

$$\begin{aligned}
&\bar{\chi}_{p\alpha}(p_\xi,p_\eta)J(p_\xi,p_\eta,p)\chi_{p\alpha}(p_\xi,p_\eta)\\
&= \tfrac{6}{8}|\phi_p(p_\xi,p_\eta)|^2\left[\left(C^{-1}\gamma_5\frac{-i\gamma p+M}{2M}\right)_{ab}\bar{u}_c(p)-\left(C^{-1}\gamma_5\frac{-i\gamma p+M}{2M}\right)_{bc}\bar{u}_a(p)\right]\\
&\quad\times[(\gamma_4)_{aa'}(i\gamma(\tfrac{1}{3}p-p_\xi+\tfrac{1}{2}p_\eta)+m_B)_{bb'}(i\gamma(\tfrac{1}{3}p-p_\eta)+m_C)_{cc'}\\
&\quad+(i\gamma(\tfrac{1}{3}p+p_\xi+\tfrac{1}{2}p_\eta)+m_A)_{aa'}(\gamma_4)_{bb'}(i\gamma(\tfrac{1}{3}p-p_\eta)+m_C)_{cc'}\\
&\quad+(i\gamma(\tfrac{1}{3}p+p_\xi+\tfrac{1}{2}p_\eta)+m_A)_{aa'}(i\gamma(\tfrac{1}{3}p-p_\xi+\tfrac{1}{2}p_\eta)+m_B)_{bb'}(\gamma_4)_{cc'}]\\
&\quad\times\left[\left(\frac{-i\gamma p+M}{2M}\gamma_5C\right)_{a'b'}u_{c'}(p)-\left(\frac{-i\gamma p+M}{2M}\gamma_5C\right)_{b'c'}u_{a'}(p)\right] \qquad (3.11)\\
&= -4\frac{E}{M}|\phi_p(p_\xi,p_\eta)|^2\left[\left(\frac{p(\frac{1}{3}p-p_\xi+\frac{1}{2}p_\eta)}{M}+m_B\right)\left(\frac{p(\frac{1}{3}p-p_\eta)}{M}+m_C\right)\right.\\
&\quad+\left(\frac{p(\frac{1}{3}p+p_\xi+\frac{1}{2}p_\eta)}{M}+m_A\right)\left(\frac{p(\frac{1}{3}p-p_\eta)}{M}+m_C\right)\\
&\quad\left.+\left(\frac{p(\frac{1}{3}p+p_\xi+\frac{1}{2}p_\eta)}{M}+m_A\right)\left(\frac{p(\frac{1}{3}p-p_\xi+\frac{1}{2}p_\eta)}{M}+m_B\right)\right]\\
&= 3\frac{E}{M}|\phi_p(p_\xi,p_\eta)|^2\left[\frac{4}{3}\frac{(pp_\xi)^2}{M^2}+\frac{(pp_\eta)^2}{M^2}-4\left(\frac{M}{3}-m_q\right)^2\right], \qquad (3.12)
\end{aligned}$$

where the factor 6 in Eq. (3.11) is due to the sum over the color index ($\epsilon_{ijk}\epsilon_{ijk}=6$) and all quark masses are set equal:

$$m_A = m_B = m_C = m_q. \tag{3.13}$$

The normalization condition is then given by

$$\frac{1}{(2\pi)^8}\int d^4p_\xi d^4p_\eta\ |\phi_p(p_\xi,p_\eta)|^2 \times\left[\frac{4}{3}\frac{(pp_\xi)^2}{M^2}+\frac{(pp_\eta)^2}{M^2}-4\left(\frac{M}{3}-m_q\right)^2\right]=1. \tag{3.14}$$

For the condition for the case of unequal masses, see Appendix C.

In order to see an explicit form of the normalization factor, we use the relativistic wavefunction $\psi_p(p_\xi,p_\eta)$ for the ground state in a relativistic harmonic oscillator potential, as an example, namely,

$$\psi_p(\xi,\eta)=N\exp\left\{-\frac{\alpha}{6}\left[2\left(\frac{\widehat{p\xi}}{M}\right)^2+\hat{\xi}^2+2\left(\frac{p\hat{\eta}}{M}\right)^2+\hat{\eta}^2\right]\right\}, \tag{3.15}$$

where[6]

$$\hat{\xi}=(1/\sqrt{2})\xi \quad\text{and}\quad \hat{\eta}=\sqrt{\tfrac{3}{2}}\eta \tag{3.16}$$

and N is the normalization factor. The empirical value for α is given by[6]

$$\alpha = 0.4\sim 0.5\ (\mathrm{GeV})^2. \tag{3.17}$$

In the center-of-mass reference frame, the wavefunction in Eq. (3.15) becomes

$$\psi_0(\xi,\eta)=N\exp\left(-\frac{\alpha}{12}(\xi^2+\xi_0^2)-\frac{\alpha}{9}(\eta^2+\eta_0^2)\right) \tag{3.18}$$

and its Fourier transform is given by

$$\phi_0(p_\xi,p_\eta)=N\left(\frac{9\pi}{\alpha}\right)^2\left(\frac{12\pi}{\alpha}\right)^2 \times\exp\left(-\frac{3}{\alpha}(\mathbf{p}_\xi^2+p_\xi^{02})-\frac{9}{4\alpha}(\mathbf{p}_\eta^2+p_\eta^{02})\right). \tag{3.19}$$

Substituting Eq. (3.19) in Eq. (3.14), we obtain

$$N=\left(\frac{\alpha}{3\pi}\right)^2\frac{1}{\sqrt{2\alpha-4(M-3m_q)^2}}. \tag{3.20}$$

If we assume that

$$M\approx 3m_q, \tag{3.21}$$

we get

$$\psi(0,0;p)\equiv N=(1/\sqrt{6\pi})(\alpha/3\pi)^{3/2}. \tag{3.22}$$

This normalization factor has been used in a computation of the proton decay rate in the SU(5) grand unified gauge model.[4]

ACKNOWLEDGMENTS

The author is grateful to David Williams for reading the manuscript. The work is supported in part by the U. S. Department of Energy.

APPENDIX A: THE RESIDUE AT THE BOUND STATE POLE FOR THE THREE-BODY PROPAGATOR

Equation (2.4) can be written as

$$K(1,2,3;4,5,6)=-\sum\chi_{p\alpha}(1,2,3)\bar{\chi}_{p\alpha}(4,5,6) \theta(\mathrm{s}(t_1t_2t_3)-\mathrm{l}(t_4t_5t_6))+\cdots, \tag{A1}$$

where s and l stands for smallest and largest, respectively. Using the result of Appendix B, we have

$$\begin{aligned}&\theta(\mathrm{s}(t_1t_2t_3)-\mathrm{l}(t_4t_5t_6))\\ &=\theta\left[\tfrac{1}{3}(t_1+t_2+t_3)-\tfrac{1}{12}(|t_1-t_2|+|t_2-t_3|+|t_3-t_1|)\right.\\ &-\tfrac{1}{12}(|2t_1-t_2-t_3+|t_2-t_3||+|2t_2-t_3-t_1+|t_3-t_1||\\ &+|2t_3-t_1-t_2+|t_1-t_2||)\\ &-\tfrac{1}{3}(t_4+t_5+t_6)-\tfrac{1}{12}(|t_4-t_5|+|t_5-t_6|+|t_6-t_4|)\\ &\left.-\tfrac{1}{12}(|2t_4-t_5-t_6-|t_5-t_6|\ |+|2t_5-t_6-t_4-|t_6-t_4|\ |\right.\\ &\left.+|2t_6-t_4-t_5-|t_4-t_5|\ ||)\right].\end{aligned} \tag{A2}$$

By explicitly singling out the bound state contribution, we obtain

$$K(1,2,3;4,5,6)=-\frac{2M}{(2\pi)^3}\int d^4k\,\chi_{k\alpha}(\xi,\eta)\bar{\chi}_{k\alpha}(\xi',\eta') \times e^{ik(X-X')}\theta(k_0)\delta(k^2+M^2)\theta(\mathrm{s}(t_1t_2t_3)-\mathrm{l}(t_4t_5t_6)) +\cdots, \tag{A3}$$

where

$$\langle 0|T\psi_A(x_1)\psi_B(x_2)\psi_c(x_3)|k,\alpha\rangle=\sqrt{M/E}\ e^{ikX}\chi_{k\alpha}(\xi,\eta),$$

$|k,\alpha\rangle$ being a bound state of spin $\frac{1}{2}$, mass M, and energy momentum $k_\mu=(\mathbf{k}, i\sqrt{k^2+M^2}\equiv iE_k)$. The remainders in Eqs. (A1) and (A3) vanish in the limit shown in Eq. (2.21). Using the variables X, ξ, η defined in Eq. (2.8a) and the integral representation of $\theta(t)$,

$$\theta(t)=-\frac{1}{2\pi i}\int dp_0\frac{1}{p_0+i\epsilon}e^{-ip_0t_0}, \tag{A4}$$

we obtain

$$\begin{aligned}&K(1,2,3;4,5,6)\\ &=-\frac{M}{(2\pi)^3}\int\frac{d^3k}{E_k}\chi_{k\alpha}(\xi,\eta)\bar{\chi}_{k\alpha}(\xi',\eta')e^{i\mathbf{k}(\mathbf{X}-\mathbf{X}')-iE_k(X_0-X_0')}\\ &\times\theta(X_0-X_0'-g(\xi_0,\eta_0)-g'(\xi_0',\eta_0'))+\cdots\\ &=-\frac{M}{(2\pi)^3}\int\frac{d^3k}{E_k}\chi_{k\alpha}(\xi,\eta)\bar{\chi}_{k\alpha}(\xi',\eta')e^{ik(X-X')-iE_k(X_0-X_0')}\\ &\times\left(-\frac{1}{2\pi i}\right)\int dp_0\frac{1}{p_0+i\epsilon}e^{-ip_0[X_0-X_0'-g(\xi_0,\eta_0)-g'(\xi_0',\eta_0')}+\cdots\\ &=-\frac{iM}{(2\pi)^4}\int\frac{d^4k}{E_k}e^{i\mathbf{k}(\mathbf{X}-\mathbf{X}')-ik_0(X_0-X_0')}\frac{1}{k_0-E_k+i\epsilon}\\ &\times\chi'_{k\alpha}(\xi,\eta)\bar{\chi}''_{k\alpha}(\xi',\eta'),+\cdots,\end{aligned} \tag{A5}$$

where

$$k_0 = p_0 + E_k,$$

$$g(\xi_0,\eta_0) = \tfrac{1}{12}(|\xi_0| + |\eta_0 - \tfrac{1}{2}\xi_0| + |\eta_0 + \tfrac{1}{2}\xi_0| + |\eta_0 + \tfrac{3}{2}\xi_0 + |\eta_0 - \tfrac{1}{2}\xi_0|| + |\eta_0 - \tfrac{3}{2}\xi_0 + |\eta_0 + \tfrac{1}{2}\xi_0|| + |-2\eta_0 + |\xi_0||), \tag{A6}$$

$$g'(\xi'_0,\eta'_0) = g(-\xi'_0,-\eta'_0),$$

and

$$\chi'_{k\alpha}(\xi,\eta) = e^{i(k_0 - E_k)g(\xi_0,\eta_0)}\chi_{k\alpha}(\xi,\eta),$$
$$\overline{\chi''_{k\alpha}}(\xi',\eta') = e^{i(k_0 - E_k)g'(\xi'_0,\eta'_0)}\overline{\chi_{k\alpha}}(\xi',\eta'). \tag{A7}$$

Notice that in the limit $k_0 \to E_k$, $\chi'_{k\alpha}(\xi,\eta)$, $\chi''_{k\alpha}(\xi,\eta)$ reduces to $\chi_{k\alpha}(\xi,\eta)$.

Defining the Fourier transform

$$\chi'_{k\alpha}(\xi,\eta) = \frac{1}{(2\pi)^8}\int d^4p_\xi d^4p_\eta e^{i(p_\xi\xi + p_\eta\eta)}\chi'_{k\alpha}(p_\xi,p_\eta),$$
$$\chi''_{k\alpha}(\xi',\eta') = \frac{1}{(2\pi)^8}\int d^4p'_\xi d^4p'_\eta e^{i(p'_\xi\xi' + p'_\eta\eta')}\chi''_{\alpha k\alpha}(p'_\xi,p'_\eta) \tag{A8}$$

and recalling the Fourier transform for K (123,45,6), Eq. (10), we get

$$K(p_\xi,p_\eta;p'_\xi,p'_\eta;k) = -i\frac{M}{E_k}\frac{1}{k_0 - E_k + i\epsilon} \times\chi_{k\alpha}(p_\xi p_\eta)\overline{\chi}_{k\alpha}(p'_\xi,p'_\eta) + \text{finite terms in the limit } k_0 \to E_k, \tag{A9}$$

which gives (2.21).

APPENDIX B: THE SMALLEST AND THE LARGEST OF THREE NUMBERS

The smaller of two numbers y and z is expressed as

$$s(y,z) = \tfrac{1}{2}(y + z - |y - z|), \tag{B1}$$

and the larger of the two is expressed

$$l(y,z) = \tfrac{1}{2}(y + z + |y - z|). \tag{B2}$$

Then the smallest of three numbers (x,y,z) is given by

$$s(x,y,z) = \tfrac{1}{2}[x + s(y,z) - |x - s(y,z)|] = \tfrac{1}{2}[x + \tfrac{1}{2}|y + z| - \tfrac{1}{2}|y - z| - |x - \tfrac{1}{2}(y + z - |y - z|)|]. \tag{B3}$$

Symmetrizing Eq. (B3) by the permutation $(x \to y \to z)$ and taking the average, we obtain the symmetric expression

$$s(x,y,z) = \tfrac{1}{3}[x + y + z - \tfrac{1}{4}(|y - z| + |z - x| + |x - y|) - \tfrac{1}{4}(|2x - y - z + |y - z|| + |2y - z - x + |z - x|| + |2z - x - y + |x - y||)]. \tag{B4}$$

Similarily, the largest of (x,y,z) is given by

$$l(x,y,z) = \tfrac{1}{2}[x + l(y,z) + |x - l(y,z)|] = \tfrac{1}{2}[x + \tfrac{1}{2}(y + z + |y + z|) + |x - \tfrac{1}{2}(y + z + |y - z|)|] = \tfrac{1}{3}[x + y + z + \tfrac{1}{4}(|x - z| + |y - z| + |z - x|) + \tfrac{1}{4}(|2x - y - z - |y - z|| + |2y - z - x - |z - x|| + |2z - x - y - |x - y||).$$

APPENDIX C: NORMALIZATION OF THE BS WAVEFUNCTION (CASE OF UNEQUAL MASSES)

The appropriate variables for the unequal mass case are

$$X = \frac{m_1x_1 + m_2x_2 + m_3x_3}{m_1 + m_2 + m_3},$$

$$\xi = x_1 - x_2, \quad \eta = \tfrac{1}{2}(x_1 + x_2 - 2x_3) \tag{C1}$$

and

$$p = p_1 + p_2 + p_3,$$

$$p_\xi = \tfrac{1}{2}(p_1 - p_2) - \frac{m_1 - m_2}{2(m_1 + m_2 + m_3)}p = \frac{1}{2(m_1 + m_2 + m_3)}[(2m_2 + m_3)p_1 - (2m_1 + m_3)p_2 - (m_1 - m_2)p_3],$$

$$p_\eta = \tfrac{1}{3}(p_1 + p_2 - 2p_3) - \frac{m_1 + m_2 - 2m_3}{3(m_1 + m_2 + m_3)}p = \frac{1}{m_1 + m_2 + m_3}[m_3(p_1 + p_2) - (m_1 + m_2)p_3]. \tag{C2}$$

The Jacobians of transformations (C1) and (C2) are unity, and these variables satisfy Eq. (2.9). The inverse of transformation (C2) is given by

$$p_1 = \frac{p}{3} + \left(p_\xi + \frac{1}{2}\frac{m_1 - m_2}{m_1 + m_2 + m_3}p\right) + \frac{1}{2}\left(p_\eta + \frac{m_1 + m_2 - 2m_3}{3(m_1 + m_2 + m_3)}p\right) = \frac{m_1}{m_1 + m_2 + m_3}p + p_\xi + \tfrac{1}{2}p_\eta,$$

$$p_2 = \frac{p}{3} - \left(p_\xi + \frac{1}{2}\frac{m_1 - m_2}{m_1 + m_2 + m_3}p\right) + \frac{1}{2}\left(p_\eta + \frac{m_1 + m_2 - 2m_3}{3(m_1 + m_2 + m_3)}p\right) = \frac{m_2}{m_1 + m_2 + m_3}p - p_\xi + \tfrac{1}{2}p_\eta, \tag{C3}$$

and

$$p_3 = \frac{p}{3} - \left(p_\eta + \frac{m_1 + m_2 - 2m_3}{3(m_1 + m_2 + m_3)}p\right) = \frac{m_3}{m_1 + m_2 + m_3}p - p_\eta.$$

Then all formulas in the text will be valid if one replaces $p/3 + p_\xi + \frac{1}{2}p_\eta$, $p/3 - p_\xi + \frac{1}{2}p_\eta$, and $p/3 - p_\eta$ by $[m_1/(m_1+m_2+m_3)]p + p_\xi + \frac{1}{2}p_\eta$, $[m_2/(m_1+m_2+m_3)]p - p_\xi + \frac{1}{2}p_\eta$, and $[m_3/(m_1+m_2+m_3)]p - p_\eta$ throughout the text [namely, in Eqs. (2.18), (2.31), (3.11), and (3.12), where $m_A = m_1$, $m_B = m_2$, and $m_C = m_3$]. The factor in the parentheses in Eq. (3.12) becomes

$$\begin{aligned}
&\left(\frac{p}{M}\left(\frac{m_2}{m_1+m_2+m_3}p - p_\xi + \tfrac{1}{2}p_\eta\right) + m_2\right)\left(\frac{p}{M}\left(\frac{m_3}{m_1+m_2+m_3}p - p_\eta\right) + m_3\right)\\
&\quad + \left(\frac{p}{M}\left(\frac{m_1}{m_1+m_2+m_3}p + p_\xi + \tfrac{1}{2}p_\eta\right) + m_1\right)\left(\frac{p}{M}\left(\frac{m_3}{m_1+m_2+m_3}p - p_\eta\right) + m_3\right)\\
&\quad + \left(\frac{p}{M}\left(\frac{m_1}{m_1+m_2+m_3}p + p_\xi + \tfrac{1}{2}p_\eta\right) + m_1\right)\left(\frac{p}{M}\left(\frac{m_2}{m_1+m_2+m_3}p - p_\xi + \tfrac{1}{2}p_\eta\right) + m_2\right)\\
&= \left(m_2\left(1 - \frac{M}{m_1+m_2+m_3}\right) - \frac{pp_\xi}{M} + \frac{pp_\eta}{2M}\right)\left(m_3\left(1 - \frac{M}{m_1+m_2+m_3}\right) - \frac{pp_\eta}{M}\right)\\
&\quad + \left(m_1\left(1 - \frac{M}{m_1+m_2+m_3}\right) + \frac{pp_\xi}{M} + \frac{pp_\eta}{2M}\right)\left(m_3\left(1 - \frac{M}{m_1+m_2+m_3}\right) - \frac{pp_\eta}{M}\right)\\
&\quad + \left(m_1\left(1 - \frac{M}{m_1+m_2+m_3}\right) + \frac{pp_\xi}{M} + \frac{pp_\eta}{2M}\right)\left(m_2\left(1 - \frac{M}{m_1+m_2+m_3}\right) - \frac{pp_\xi}{M} + \frac{pp_\eta}{2M}\right) \qquad \text{(C4)}\\
&= -\left[\frac{(pp_\xi)^2}{M^2} + \frac{3}{4}\frac{(pp_\eta)^2}{M^2} - (m_1m_2 + m_2m_3 + m_3m_1)\left(1 - \frac{M}{m_1+m_2+m_3}\right)^2\right.\\
&\quad \left. + \left(1 - \frac{M}{m_1+m_2+m_3}\right)\left((m_1 - m_2)\frac{pp_\xi}{M} + \frac{1}{2}(m_1 + m_2 - 2m_3)\frac{pp_\eta}{M}\right)\right]. \qquad \text{(C5)}
\end{aligned}$$

The normalization for the three-body BS wavefunction for unequal mass constituents is given by

$$\begin{aligned}
&\frac{1}{(2\pi)^4}\int d^4p_\xi d^4p_\eta |\phi_p(p_\xi, p_\eta)|^2\left[\frac{4}{3}\frac{(pp_\xi)^2}{M^2} + \frac{(pp_\eta)^2}{M^2} - \frac{4}{3}(m_1m_2 + m_2m_3 + m_3m_1)\left(1 - \frac{M}{m_1+m_2+m_3}\right)^2\right.\\
&\quad \left. + \frac{4}{3}\left(1 - \frac{M}{m_1+m_2+m_3}\right)\left((m_1 - m_2)\frac{pp_\xi}{M} + \frac{1}{2}(m_1 + m_2 - 2m_3)\frac{pp_\eta}{M}\right)\right] = 1. \qquad \text{(C6)}
\end{aligned}$$

[1]H. A. Bethe and E. E. Salpeter, Phys. Rev. **82**, 309 (1951).
[2]K. Nishijima, Prog. Theor. Phys. **10**, 549 (1953); **12**, 279 (1954); **13**, 305 (1955); S. Mandelstam, Proc. R. Soc. (London) A **233**, 248 (1955); A. Klein and C. Zemach, Phys. Rev. **108**, 126 (1957); D. Lurie, A. J. MacFarlane, and Y. Takahashi, Phys. Rev. **140**, B1091 (1965); C. H. Llewellyn Smith, Nuovo Cimento A **60**, 348 (1969).
[3]D. Lurie, *Particles and Fields* (Wiley, New York, 1968), p. 433; N. Nakanishi, Prog. Theor. Phys. Suppl. **43**, 1 (1969), and references quoted there.
[4]Y. Tomozawa, Phys. Rev. Lett. **46**, 463 (1981); **49**, 507(E) (1982).
[5]Here and hereafter, appropriate sums over the spinor indices and the SU(3) color indicies are implied by the integral sign.
[6]R. P. Feynman, M. Kislinger, and F. Ravndal, Phys. Rev. D **3**, 2706 (1971); R. Lipes, Phys. Rev. D **5**, 2849 (1972).
[7]V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. USA **34**, 211 (1948).

On the dimensional regularization procedure for massless Feynman integrals

M. W. Kalinowski
Institute of Philosophy and Sociology, Polish Academy of Sciences, Nowy Świat 72, 00-330 Warsaw, Poland

M. Seweryński[a] and L. Szymanowski[b]
Institute of Nuclear Research, Hoza 69, 00-681 Warsaw, Poland

(Received 30 July 1980; accepted for publication 30 October 1981)

It is shown that the dimensional regularization procedure for massless Feynman integrals, proposed by Capper and Leibbrandt, is unsuitable for practical calculations. This is due to the fact that the procedure yields a logarithmic singularity for some specific massless Feynman integral.

PACS numbers: 11.90. + t

Trying to solve the problem of infrared divergencies arising from massless particles in QFT Capper and Leibbrandt[1] have proposed the following redefinition of the generalized Gaussian integral in $2w$-dimensional Euclidean space:

$$G(w) = \int \frac{d^{2w}q}{(2\pi)^{2w}} \exp(-xq^2 + 2bq) = (4\pi)^{-w} \exp[b^2/x - x f(w)], \quad x>0, \ w\in\mathbb{C}, \tag{1}$$

where the vector b_μ is also defined over $2w$-space and x behaves like a c number. The function $f(w)$ is an entire function which satisfies the conditions

(i) $f^{(k)}(w) = 0$ for $w = n/2$, $n\in N_0^\infty$, $k\in N_0^\infty - N_{k_0+1}^\infty$, $k_0\geqslant 2, k_0 < \infty$,

(ii) $\operatorname{Re} f(w) > 0$ for any $w \neq n/2$ and some $\operatorname{Im} w = 0$ (see Refs. 1 and 2, for details).

It was claimed by some authors[1,3,4] that such an extension of a Gaussian integral allows one to develop a reasonable dimensional regularization scheme for massless theories. In this comment we re-examine the claim and show that the regularization scheme proposed in Ref. 1 is unsuitable for practical calculations. First we calculate the integral

$$\int \frac{d^{2w}q}{(2\pi)^{2w}} (q^2 + 2pq + m^2)^{-z}, \quad w, z\in\mathbb{C},$$

with the help of (1). Substituting $b = -xp$ into (1) one obtains

$$\int \frac{d^{2w}q}{(2\pi)^{2w}} \exp[-x(q^2 + 2pq + m^2)] = (4\pi)^{-w} x^{-w} \exp[-x(f + m^2 - p^2)]. \tag{2}$$

Multiplying both sides of Eq. (2) by x^{z-1}, $z\in\mathbb{C}$ and integrating over x one finds

$$\int \frac{d^{2w}q}{(2\pi)^{2w}} (q^2 + 2pq + m^2)^{-z} = \frac{(4\pi)^{-w}}{\Gamma(z)} \int_0^\infty dx\, x^{z-w-1} \exp[-x(f + m^2 - p^2)] = (4\pi)^{-w} (f + m^2 - p^2)^{w-z} \Gamma(z-w)/\Gamma(z), \quad w,z\in\mathbb{C}. \tag{3}$$

It is now obvious that for an integral of the type

$$\int \frac{d^{2w}q}{(2\pi)^{2w}} (q^2)^{-z}$$

the Capper–Leibbrandt regularization scheme consists of introducing a complex mass square $f(w)$ at an intermediate stage of the calculations. One may expect then that such a procedure will lead to trouble with gauge invariance of the theory and this is indeed the case (see Refs. 1 and 2). Now let us consider the simplest one closed-loop integral

$$I_{SE} = \int d^{2w}q [q^2 (p-q)^2]^{-1}.$$

Using the standard α representation $1/q^2 = \int_0^\infty dx \exp(-\alpha q^2)$, one finds

$$I_{SE} = \pi^w \int_0^\infty dx \int_0^\infty dy\, e^{-yp^2} \int d^{2w}q \times \exp[-(x+y)q^2 + 2y(pq)]. \tag{4}$$

Applying (1) and introducing new variables $u = x + y$, $uw = y$ one gets

$$I_{SE} = \pi^w \int_0^\infty du\, u^{z-1} e^{-uf} \int_0^1 dv \exp[-uv(1-v)p^2], \tag{5}$$

where $z = 2 - w$.

Introducing a new variable $t^2 = v(1-v)$ one obtains

$$K(u) \equiv \int_0^1 dv \exp[-uv(1-v)p^2] = \int_0^{1/2} dt\,(1-4t^2)^{-1/2} \exp(-bt^2) - \int_{1/2}^0 dt\, 2t\,(1-4t^2)^{-1/2} \exp(-bt^2) = \frac{1}{2}\int_0^1 dx (1-x)^{-1/2} \exp(bx/4) = {}_1F_1(1;\tfrac{3}{2}; -p^2u/4), \tag{6}$$

where $b^2 = pu$ [see Ref. 5, 4.2(1)].

Hence

$$I_{SE} = \pi^w \int_0^\infty du\, u^{z-1} {}_1F_1(1,\tfrac{3}{2}; -p^2u/4) \exp(-fu)$$

and

$$I_{SE} = \pi^w f^{w-2} \Gamma(2-w) {}_2F_1(1,2-w;\tfrac{3}{2}; -p^2/4f), \quad |4f| \geqslant p^2, \ \operatorname{Re} f > 0 \tag{7}$$

[a]Supported by MNSzWT Grant No. 04.3.14.02.05.2A-1K8E and partially supported by INT Grant No. 73-20002 A01.
[b]Supported by MNSzWT Grant No. 04.03.14.02.05.2A-1K7E.

 0022-2488/83/020375-02$02.50

[see, e.g., Ref. 5, 3.6(13)].

Now we apply Kummer's theorem [see, e.g., Ref. 6, 2.1.4(22)] to ${}_2F_1(1,2-w;\frac{3}{2};-p^2/4f)$ and deduce that

$$I_{SE}=\pi^w\Gamma(2-w)(f+p^2/4)^{w-2}\,{}_2F_1\left(\tfrac{1}{2},2-w;\tfrac{3}{2};\frac{p^2}{p^2+4f}\right), \quad (8)$$

$$\mathrm{Re}\,f\geqslant 0, \quad 1<\mathrm{Re}\,w<2.$$

Since

$${}_2F_1(\tfrac{1}{2},2-w;\tfrac{3}{2};1)=4^{w-2}\frac{\Gamma(w-1)\Gamma(w-1)}{\Gamma[2(w-1)]}$$

provided $\mathrm{Re}\,w>1$, one may rewrite (8) in a more convenient form

$$I_{SE}=I(0)(1+4f/p^2)^{w-2}\,{}_2F_1\left(\tfrac{1}{2},2-w;\tfrac{3}{2};\frac{p^2}{p^2+4f}\right)\Big/\,{}_2F_1(1), \quad (9)$$

$$1<\mathrm{Re}\,w<2$$

where

$$I(0)\equiv\pi^w(p^2)^{w-2}\Gamma(2-w)\frac{\Gamma(w-1)\Gamma(w-1)}{\Gamma[2(w-1)]},$$

$${}_2F_1(1)\equiv{}_2F_1(\tfrac{1}{2},2-w;\tfrac{3}{2};1).$$

Performing an analytic continuation of the right-hand side of (9) in the variable w one gets I_{SE} for $w\in\mathbb{C}$. It is trivial to see that Eq. (9) yields the usual pole singularity of I_{SE} at the physical point $w=2$ (i.e., $n=4$).

Now let us consider an integral associated with the pure graviton triangle diagram

$$J_3=\int d^{2w}k\,[k^2(k-p_2)^2(k+p_3)^2]^{-1}$$

and assume further that $p_2=-p_3$, $p_2^2=p_3^2=p^2\neq 0$. Applying as usual the α representation one obtains after some calculations [with help of Eq. (1) and Kummer's theorem]:

$$J_3=\pi^w\int_0^\infty dx\,x^{(3-w)-1}e^{-xf}\,{}_1F_1(1,\tfrac{3}{2};-p^2x/4)$$

$$=\pi^w\Gamma(3-w)(f+p^2/4)^{w-3}\,{}_2F_1\left(\tfrac{1}{2},3-w;\tfrac{3}{2};\frac{p^2}{p^2+4f}\right),$$

$$p_2=-p_3=p,$$

so a singularity structure of J_3 at the physical point $w=2$ is determined by the singularity structure of ${}_2F_1$ at this point. Since it is known that if $c-a-b=0$ then ${}_2F_1(a,b,c;z)$ has a logarithmic singularity (see, e.g., Ref. 7, p. 18) then J_3 has a logarithmic singularity at $w=2$ when $p_2=-p_3$ and $p_2^2\neq 0$. This behavior of J_3 at $w=2$ sharply contrasts with the standard pole singularity one gets with the help of the standard definition of a Gaussian integral [i.e., when $f(w)\equiv 0, w\in\mathbb{C}$]:

$$J_3(0)=\pi^w\int_0^\infty dx\,x^{(3-w)-1}\,{}_1F_1(1,\tfrac{3}{2};-p^2z/4)$$

$$=\pi^{w+1/2}(p^2/4)^{w-3}\Gamma(3-w)\Gamma(w-2)/2\Gamma(w-\tfrac{3}{2}),$$

$$p_2=-p_3, \quad p_2^2=p^2.$$

Unfortunately, this fact remained unnoticed in Refs. 1–4.

Summarizing, we have shown that although the Capper–Leibbrandt redefinition of the Gaussian integral (1) yields for one closed-loop integral I_{SE} the standard pole singularity at the physical point $w=2$ nevertheless it results in the logarithmic singularity for J_3 at $w=2$ (for a particular momenta configuration). This means that unless accidental cancellations occur, the pure graviton triangle diagram has a logarithmic singularity at $w=2$ which would render the Capper–Leibbrandt regularization procedure unsuitable for the needs of a renormalization of theory.

[1] D. M. Capper and G. Leibbrandt, J. Math. Phys. **15**, 82 (1974).
[2] G. Leibbrandt, Rev. Mod. Phys. **47**, 849 (1975).
[3] D. M. Capper and G. Leibbrandt, J. Math. Phys. **15**, 86 (1974).
[4] D. M. Capper and G. Leibbrandt, J. Math. Phys. **15**, 795 (1974).
[5] Y. L. Luke, *The Special Functions and Their Approximations* (Academic, New York, 1969), Vol. I.
[6] *Higher Transcendental Functions*, edited by A. Erdelyi (McGraw-Hill, New York, 1953), Vols. I–III.
[7] F. Klein, *Vorlesungen über die Hypergeometrische Funktion* (Springer, Berlin, 1933).

High frequency asymptotic solutions of Yang–Mills and associated fields

Yvonne Choquet-Bruhat
Institut de Mecanique Teorique et Appliquée, Université Paris VI, 75230 Paris, France

Antonio Greco
Istituto Matematico, Universita di Palermo, Palermo, Italy

(Received 19 April 1982; accepted for publication 1 July 1982)

We establish the differential equations which rule the propagation of the high-frequency waves, disturbances of a given background, for the coupled Yang–Mills, scalar and spinor field equations. We discuss their interaction.

PACS numbers: 12.20. – m, 03.50. – z, 03.40.Kf, 02.30.Jr

1. INTRODUCTION

We shall construct asymptotic, high-frequency solutions of the Yang–Mills equations, coupled with the wave equations for scalar and spinor multiplets via the gauge covariant derivative defined with the Yang–Mills potential (connection). We shall write the equations on an arbitrary given space time M with a hyperbolic metric g; this space time can be, for instance, Minkowski M_4.

The method we use (extension of the WKB or "two-timing" method) is the general method of Ref. 1, extended in Ref. 2, to nonlinear equations. The (classical) fields that we construct are also generalizations of what is called simple waves. They depend on the point x of space-time where they are evaluated, on the one hand, directly, and on the other hand through a product $\omega\varphi$, where φ is a scalar function on M (called phase) and $\omega \gg 1$.

We prove that for nontrivial solutions to exist it is necessary and sufficient that the phase satisfy the eikonal equation of the hyperbolic metric. We establish the differential equations which rule the propagation of the high frequency waves along the rays, and show that it induces their mixing in generic backgrounds (cf. in particular the propagation laws 5–8 of the two polarization modes of the spinor waves). We also give the propagation laws of the energies of the boson waves and discuss their conservation.

Some aspects of the high frequency YM field, and of the scattering of a scalar field on a given YM potential had been studied in Ref. 3. The interaction of an electromagnetic and a charged field (in this context) had been given in Ref. 4, together with their interaction with a gravitational field.

2. FIELD EQUATIONS

The equations are, with $F_{\lambda\mu} \equiv \nabla_\lambda A_\mu - \nabla_\mu A_\lambda + c[A_\lambda, A_\mu]$,

$$\nabla_\lambda F^{\lambda\mu} + c[A_\lambda, F^{\lambda\mu}] = 2k\Re\{\phi^* S^\# (\nabla^\mu \phi + SA^\mu \phi) + i\bar{\psi}\gamma^\mu T^\# \Psi\}, \tag{2.1}$$

$$\hat{\Box}\phi \equiv \nabla^\lambda \nabla_\lambda \phi + 2SA^\lambda \nabla_\lambda \phi + (S(\nabla_\lambda A^\lambda) + (SA^\lambda)(SA_\lambda)) = K(\phi, \psi), \tag{2.2}$$

$$\hat{\nabla}\psi \equiv \gamma^\alpha (\nabla_\alpha \psi + TA_\alpha \Psi) = H(\phi, \Psi). \tag{2.3}$$

The notations are as follows (the same as in Ref. 5 $\Re$ and k are constants). x^λ, $\lambda = 0,1,2,3$ are coordinates on M. Indices are raised with g. ∇_λ is the metric covariant derivative (i.e., the partial derivative with respect to x^λ if M is Minkowski).

A_λ is a 1-form on M with values of $\mathfrak{G}$, Lie algebra of a Lie group G admitting a bi-invariant nondegenerate metric (for instance a compact Lie group). [,] is the Lie bracket in $\mathfrak{G}$.

Φ, scalar multiplet, is a mapping $M \to \mathbb{C}^N$, where $\mathbb{C}^N$ is the representation space of G by $N \times N$ unitary matrices. S is the induced representation of their Lie algebras, i.e., some constant linear map $\mathfrak{G} \to \mathfrak{U}(N)$; $S^\#$ is the element of $\mathfrak{U}(N) \otimes \mathfrak{G}$ deduced from S by duality and the isomorphism of $\mathfrak{G}$ with its dual defined by the metric of G. The * is the Hermitian conjugate.

Ψ, spinor multiplet, is a mapping $M \to \mathbb{C}^M \times \mathbb{C}^4$, $\mathbb{C}^m$, being the representation space of G by $m \times m$ unitary matrices, $T: \mathfrak{G} \to \mathfrak{U}(m)$ induced representation, $T^\#$ deduced from T as $S^\#$ from S. The γ^α are the Dirac matrices.

H and K are given analytic functions of ϕ and Ψ, compatible with the gauge transformations of $\hat{\Box}\phi$ and $\hat{\nabla}\psi$, and such that the Eqs. (2.2) and (2.3) imply $\nabla_\mu J^\mu + c[A_\mu, J^\mu] = 0$ (For physical examples of such H and K cf. Ref. 6, some are quoted in Ref. 5), J^μ given by the right-hand side of Eq. (2.1).

3. ASYMPTOTIC HIGH FREQUENCY WAVES

Following the method of Refs. 1 and 2, we consider the unknowns as functions on $M \times \mathbb{R}$ and write them as formal series:

$$\phi(x,\xi) = \overset{0}{\phi}(x) + \frac{1}{\omega}\overset{1}{\phi}(x,\xi) + \frac{1}{\omega^2}\overset{2}{\phi}(x,\xi) + \cdots, \tag{3.1}$$

$$\Psi(x,\xi) = \overset{0}{\Psi}(x) + \frac{1}{\omega}\overset{1}{\Psi}(x,\xi) + \frac{1}{\omega^2}\overset{2}{\Psi}(x,\xi) + \cdots, \tag{3.2}$$

$$A_\lambda(x,\xi) = \overset{0}{A}_\lambda(x) + \frac{1}{\omega}\overset{1}{A}_\lambda(x,\xi) + \frac{1}{\omega^2}\overset{2}{A}_\lambda(x,\xi) + \cdots. \tag{3.3}$$

We set $\xi = \omega\varphi(x)$, where φ is a function $M \to \mathbb{R}$ called the phase. For a general function $f(x,\xi)$ we set

$$f'(x,\omega\varphi(x)) = \frac{\partial f}{\partial \xi}(x,\xi)|_{\xi = \omega\varphi(x)},$$

$$\partial_\lambda f(x,\omega\varphi(x)) = \frac{\partial}{\partial x^\lambda} f(x,\xi)|_{\xi = \omega\varphi(x)}.$$

Thus

 0022-2488/83/020377-03$02.50

$$\nabla_\lambda f(x,\omega\varphi(x)) = \omega\varphi_\lambda f'(x,\omega\varphi(x)) + \partial_\lambda f(x,\omega\varphi(x)) \text{ with } \partial_\lambda \equiv \frac{\partial}{\partial x^\lambda},$$

$$\varphi_\lambda \equiv \frac{\partial\varphi}{\partial x^\lambda}.$$

We say that (3.1)–(3.3) is an asymptotic wave of order p if when we substitute ϕ, Ψ, A with these formal series in (2.2)–(2.4) the terms in $\omega,\ldots,\omega^{-p+1}$ vanish. Obviously an asymptotic wave of order $p>0$ furnishes approximate solutions, by taking only a finite sum of terms and making sure that the remainder, $\omega^{-p}K$, is such that K is bounded for all x and ξ.

The condition that a function $f(x,\xi)$ to be uniformly bounded with respect to ξ requires that the average of its derivative with respect to ξ vanish:

$$\lim_{T=\infty} \frac{1}{T}\int_0^T f'(x,\xi)\,d\xi = 0.$$

This condition will be taken into account in the following—notice that it forces the background $\overset{0}{A}, \overset{0}{\phi}, \overset{0}{\psi}$ to be an exact solution, as we shall assume from the beginning.

4. DETERMINATION OF THE PHASE

The coefficients of the higher powers of ω (respectively ω^1, ω^1, and ω^0) obtained by substituting (3.1)–(3.3) in (2.1)–(2.3) equated to zero give respectively:

$$\overset{1}{A}{}''^{\mu}\varphi^\lambda\varphi_\lambda - \overset{1}{A}{}''^{\lambda}\varphi_\lambda\varphi^\mu = 0, \tag{4.1}$$

$$\overset{1}{\phi}{}''\varphi^\lambda\varphi_\lambda = 0, \tag{4.2}$$

$$\gamma^\alpha\varphi_\alpha \overset{1}{\Psi}{}' = 0. \tag{4.3}$$

A necessary condition for all these equations to be satisfied with nonzero $\overset{1}{A}, \overset{1}{\phi}, \overset{1}{\Psi}$ is that φ_α be a solution of the eikonal equation

$$\varphi^\alpha\varphi_\alpha = 0.$$

The coefficients $\overset{1}{A}_\lambda$ and $\overset{1}{\Psi}$ are then restricted by the conditions (we choose the arbitrary functions of x which come by integration, with respect to ξ, to be zero)

$$\overset{1}{A}{}^\lambda\varphi_\lambda = 0, \tag{4.4}$$

$$\gamma^\alpha\varphi_\alpha \overset{1}{\Psi} = 0. \tag{4.5}$$

The condition (4.4) expresses that the Lorentz condition for $\overset{0}{A}$ and $\overset{0}{A} + (1/\omega)\overset{1}{A}$ are the same at first order. Indeed,

$$\nabla_\lambda\left(\overset{0}{A}{}^\lambda + \frac{1}{\omega}\overset{1}{A}{}^\lambda\right) = \nabla_\lambda \overset{0}{A}{}^\lambda + \overset{1}{A}{}'^{\lambda}\varphi_\lambda + \frac{1}{\omega}\partial_\lambda \overset{1}{A}{}^\lambda. \tag{4.6}$$

Remark: Equation (4.1) alone does not imply $\varphi^\lambda\varphi_\lambda = 0$. But if $\varphi^\lambda\varphi_\lambda \neq 0$, $\overset{1}{A}{}^\mu = f\varphi^\mu$, and can be made to vanish by a gauge transformation $U = \text{identity} + (1/\omega^2)\overset{1}{U}$ which preserves the background $\overset{0}{A}$.

Equation (4.5), considered as an equation on $\mathbb{C}^4$, has a two-dimensional vector space of solutions, the mapping $\gamma^\alpha\varphi_\alpha : \mathbb{C}^4 \to \mathbb{C}^4$ being of rank 2, when $\varphi^\alpha\varphi_\alpha = 0$.

If $\mathbf{h}_r$, $r = 1,2$, is a basis of this space, the general solution of (4.5) is

$$\Psi^1 = \lambda_1\mathbf{h}_1 + \lambda_2\mathbf{h}_2, \tag{4.7}$$

where λ_1 and λ_2 are $\mathbb{C}^m$ valued functions on M (the products are in fact tensor products).

5. PROPAGATION EQUATIONS

Equating to zero the terms of the next order $(\omega^0,\omega^0,\omega^{-1})$ in (2.1)–(2.3) after the substitution, taking into account $\overset{1}{A}{}'^{\lambda}\varphi_\lambda = 0$, and the fact that $\varphi^\lambda\partial_\lambda\varphi_\mu = \varphi^\lambda\partial_\mu\varphi_\lambda = 0$, we obtain

$$2\varphi^\lambda\partial_\lambda \overset{1}{A}{}'^{\mu} + \overset{1}{A}{}'^{\mu}\partial_\mu\varphi^\lambda + 2c\left[\overset{0}{A}{}^\lambda\varphi_\lambda, \overset{1}{A}{}'^{\mu}\right] - c\varphi^\mu\left[\overset{0}{A}_\lambda, \overset{1}{A}{}'^{\lambda}\right] - \varphi^\mu\left(\partial_\lambda \overset{1}{A}{}'^{\lambda} + \overset{2}{A}{}''^{\lambda}\varphi_\lambda\right) = 2k\varphi^\mu\,\Re\left(\overset{0}{\phi}{}^*\overset{1}{\phi}{}'\right), \tag{5.1}$$

$$2\varphi^\lambda\partial_\lambda\overset{1}{\phi}{}' + \overset{1}{\phi}{}'\partial_\lambda\varphi^\lambda + 2S\overset{0}{A}{}^\lambda\varphi_\lambda\overset{1}{\phi}{}' = 0, \tag{5.2}$$

$$\gamma^\alpha\varphi_\alpha\overset{2}{\Psi} + \gamma^\alpha\partial_\alpha\overset{1}{\Psi} + \gamma^\alpha T\left(\overset{0}{A}_\alpha\overset{1}{\Psi} + \overset{1}{A}_\alpha\overset{0}{\Psi}\right) = H'_\Psi\left(\overset{0}{\phi},\overset{0}{\Psi}\right) + H'_\phi\left(\overset{0}{\phi},\overset{0}{\Psi}\right). \tag{5.3}$$

For Eq. (5.1) we first note, multiplying by φ_μ, that it implies the propagation of $\overset{1}{A}{}'^{\mu}\varphi_\mu$ along the rays φ^λ:

$$2\varphi^\lambda\partial_\lambda\left(\overset{1}{A}{}'^{\mu}\varphi_\mu\right) + \overset{1}{A}{}'^{\mu}\varphi_\mu\partial_\lambda\varphi^\lambda + 2c\left[\overset{0}{A}{}^\lambda\varphi_\lambda, \overset{1}{A}{}'^{\mu}\varphi_\mu\right] = 0.$$

Therefore if $\overset{1}{A}{}'^{\mu}$ satisfies (5.1), and $\overset{1}{A}{}'^{\mu}\varphi_\mu = 0$ on a submanifold transversal to the rays, it will satisfy $\overset{1}{A}{}'^{\mu}\varphi_\mu = 0$ on the region regularly spanned by the rays.

We then note that Eq. (5.1) gives for any choice of $\overset{2}{A}{}^\lambda$, a propagation for the metric scalar product (with values in $\mathfrak{G}$) of $\overset{1}{A}{}^\mu$ with any given vector field v^μ orthogonal to the rays, i.e., such that $\varphi^\mu v_\mu = 0$. We obtain a propagation equation for all components of $\overset{1}{A}{}^\mu$ if we impose on $\overset{2}{A}{}^\mu$ (still unrestricted) the condition

$$\overset{2}{A}{}''^{\lambda}\varphi_\lambda + \partial_\lambda\overset{1}{A}{}'^{\lambda} = 0, \tag{5.4}$$

which expresses that the Lorentz condition is preserved at order 2. Equation (5.1) then shows clearly that the significant part of $\overset{1}{A}{}^{\mu}$ is influenced only by the background YM potential, and only if $\overset{0}{A}{}_{\lambda}{}^{\lambda} \neq 0$. It reads

$$2\varphi^{\lambda}\partial_{\lambda}\overset{1}{A}{}'^{\mu} + \overset{1}{A}{}'^{\mu}\partial_{\lambda}\varphi^{\lambda} + 2c\left[\overset{0}{A}{}^{\lambda}\varphi_{\lambda}, \overset{1}{A}{}'^{\mu}\right] = \varphi^{\mu}\left(c\left[\overset{0}{A}{}_{\lambda}, \overset{1}{A}{}'^{\lambda}\right] + 2k\Re\left(\overset{0}{\phi}{}^{*}\overset{1}{\phi}{}'\right)\right). \tag{5.5}$$

We deduce from this equation by multiplication with $\overset{1}{A}{}'_{\mu}$ contracted in ⓢ

$$\partial_{\lambda}\left(\overset{1}{A}{}'^{\mu}\overset{1}{A}{}'^{\mu}\varphi^{\lambda}\right) + 2c\left[\overset{0}{A}{}^{\lambda}\varphi_{\lambda}, \overset{1}{A}{}'^{\mu}\right]\overset{1}{A}{}'_{\mu} = 0,$$

the second term, written explicitly is

$$\left[\overset{0}{A}{}^{\lambda}\varphi_{\lambda}, \overset{1}{A}{}'^{\mu}\right]\overset{1}{A}{}'_{\mu} = c^{a}_{bc}\overset{0}{A}{}^{\lambda,b}\varphi_{\lambda}\overset{1}{A}{}'^{\mu c}\overset{1}{A}{}'_{\mu a}, \tag{5.6}$$

which vanishes if $c^{a}_{bc} + c^{c}_{ba} = 0$, thus, if G is compact. The specific energy of the YM disturbances is then conserved during the propagation. In the other case we may have creation or dissipation of this energy by the background (cf. a similar phenomenon in another context in Trautman[6]).

Note: $\overset{1}{A}{}'^{\mu}\overset{1}{A}{}'_{\mu}$ is the (positive) energy of the high-frequency disturbance, associated to any direction u, transversal to the rays, and normalized by the condition $\varphi^{\lambda}u_{\lambda} = 1$. Indeed the energy of the field at first order is

$$E_{\mathrm{YM}} \equiv \tfrac{1}{4} g^{\lambda\mu}u_{\lambda}u_{\mu}\overset{0}{F}{}^{\alpha\beta}\overset{0}{F}{}_{\alpha\beta} - \overset{0}{F}{}^{\alpha\lambda}u_{\lambda}\overset{0}{F}{}_{\alpha}{}^{\mu}u_{\mu},$$

with

$$\overset{0}{F}{}_{\alpha\beta} = \partial_{\alpha}\overset{0}{A}{}_{\beta} - \partial_{\beta}\overset{0}{A}{}_{\alpha} + c\left[\overset{0}{A}{}_{\alpha}, \overset{0}{A}{}_{\beta}\right] + \overset{1}{A}{}'_{\beta}\varphi_{\alpha} - \overset{1}{A}{}'_{\alpha}\varphi_{\beta},$$

that is, up to a linear term which disappears by averaging in ξ,

$$E_{\mathrm{YM}} \simeq E_{\mathrm{YM}}\left(\overset{0}{A}{}_{\lambda}\right) - \overset{1}{A}{}'^{\beta}\overset{1}{A}{}'_{\beta}.$$

Equation (5.2) gives the propagation of the perturbation $\overset{1}{\phi}$ of the scalar field along the rays φ^{λ} associated with the wave fronts $\varphi = cte$. This propagation is accompanied by a mixing of the components of the multiplet if the background potential is such that $\overset{0}{A}{}^{\lambda}\varphi_{\lambda} \neq 0$.

Note: If we did not require A to be also an asymptotic solution of Yang–Mills equations, we would have, added to the left-hand side of (5.2), the term $\overset{1}{A}{}'^{\lambda}\varphi_{\lambda}\overset{0}{\phi}$, and therefore a generation of a perturbation in ϕ by the perturbation in A if $\overset{1}{A}{}^{\lambda}\varphi_{\lambda} \neq 0$.

Using Eq. (5.2), its Hermitian transpose, and the fact that $S + S^{*} = 0$ since the representation is unitary, we obtain the conservation law for the energy of the disturbance $\overset{1}{\phi}$:

$$\partial_{\lambda}\left(\left[\overset{1}{\phi}{}'\right]^{2}\varphi^{\lambda}\right) = 0.$$

We shall now deduce from (5.3) a propagation equation for the two modes of polarization λ_r, $r = 1,2$ or $\overset{1}{\psi}$. For an arbitrary vector $\mathbf{p}$, and the standard choice of γ matrices, the matrix $\gamma^{\alpha}p_{\alpha}$ reads

$$\gamma^{\alpha}p_{\alpha} = \begin{pmatrix} p_0 & 0 & p_3 & -ip_1 + p_2 \\ 0 & p_0 & ip_1 + p_2 & -p_3 \\ -p^3 & ip_1 - p_2 & -p_0 & 0 \\ -(ip_1 + p_2) & p_3 & 0 & -p_0 \end{pmatrix},$$

where $p_{\alpha} = \varphi_{\alpha}$ satisfies $\varphi^{\alpha}\varphi_{\alpha} = 0$, it is of rank 2. The vectors $\mathbf{h}_r$, $r = 1,2$, depending on p,

$$\mathbf{h}_1 = (p_3, ip^1 + p^2, p^0, 0), \quad \mathbf{h}_2 = (ip^1 - p^2, p^3, 0, p^0), \tag{5.7}$$

satisfy the linear system

$$\gamma^{\alpha}p_{\alpha}\mathbf{h}_r = \mathbf{k}_r p^{\alpha}p_{\alpha}, \quad \mathbf{k}_1 = (0,0,1,0), \quad \mathbf{k}_2 = (0,0,0,1).$$

The covectors (Dirac adjoints) $\bar{\mathbf{h}}_r = \mathbf{h}_r^{*}\gamma^0$ satisfy

$$\bar{\mathbf{h}}_r\gamma^{\alpha}p_{\alpha} = \mathbf{k}_r p^{\alpha}p_{\alpha}, \quad r = 1,2.$$

Thus we obtain (forms which can be foreseen from the general theory[2,8]) by derivation in p, and then in x with $p_{\alpha} = \varphi_{\alpha}$,

$$\bar{\mathbf{h}}_r\gamma^{\alpha}\mathbf{h}_s = 2\varphi^{0}\varphi^{\alpha}\delta^{s}_{r}, \quad \bar{\mathbf{h}}_r\gamma^{\alpha}\partial_{\alpha}\mathbf{h}_s = \varphi^{0}\partial_{\alpha}\varphi^{\alpha}\delta^{s}_{r}. \tag{5.8}$$

Note also that $\bar{h}_r h_s = 0$, for all r, s if $p_{\alpha} = \varphi_{\alpha}$ (but $\bar{h}_r\gamma^0 h_s \neq 0$). Inserting (4.7) in (5.3) and using (5.8) we get

$$\varphi^{0}(2\varphi^{\alpha}\partial_{\alpha}\lambda_r + \lambda_r\partial_{\alpha}\varphi^{\alpha}) + \bar{h}_r\gamma\left(\overset{0}{TA}{}_{\alpha}\right)\left(\sum_{s=1}^{2}\lambda_s h_s\right) + \bar{h}_r\gamma^{\alpha}\left(\overset{1}{TA}{}_{\alpha}\right)\overset{0}{\psi} = \bar{h}_r\left(H'_{\psi}\left(\overset{0}{\phi},\overset{0}{\psi}\right)\sum_{s=1}^{2}\lambda_s h_s + H'_{\phi}\left(\overset{0}{\phi},\overset{0}{\psi}\right)\overset{1}{\phi}\right). \tag{5.9}$$

We thus see that the two amplitudes $\lambda_r \in \mathbb{C}^m$ propagate along the rays φ^{α} but the background potential induces a mixing of these amplitudes and so do the background scalar and spinor fields if $\bar{h}_r H'_{\psi}\left(\overset{0}{\phi},\overset{0}{\psi}\right)h_s \neq 0$ when $s \neq r$. Moreover the wave $\overset{1}{A}{}_{\alpha}$, and the scalar wave $\overset{1}{\phi}$ act as source of this spinor wave, but in exceptional background configurations.

We shall give applications of these general formulas to current physical models in a forthcoming paper.

ACKNOWLEDGMENT

This research has been supported by the C.N.R. (Gruppo Nazionale per la Fisica Matematica).

[1] L. Garding, T. Kotake, and J. Leray, Bull. Soc. Math. France **92**, 263–361 (1964).
[2] Y. Choquet-Bruhat, J. Math. Pures Appl. **48**, 117–158 (1969).
[3] Y. Choquet-Bruhat and A. H. Taub, Gen. Relativ. Gravit. **8**, 561 (1977).
[4] R. Kerner, C. R. Acad. Sci. Paris **280**, 1463 (1975); Rep. Math. Phys. **9**(1), 95–105 (1976).
[5] Y. Choquet-Bruhat and D. Christodoulou, Ann. Sci. École Normale Supérieure (to be published).
[6] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967); A. Salam, *Elementary particle theory* (Almquist, Stockholm, 1968); L. Faddeev and A. Slavnov, *Gauge fields* (Benjamin, London, 1980).
[7] A. Trautman, preprint, Varsovie.
[8] A. M. Anile and A. Greco, Ann. Inst. Henri Poincaré **29**(3), 257–272 (1978).

Expansions around the Gaussian and Lorentzian limits for randomly perturbed spectral lines and their shift, width, and asymmetry

Antoine Royer
Centre de Recherche de Mathématiques Appliquées, Université de Montréal, Montréal, H3C 3J7, Canada

(Received 15 September 1981; accepted for publication 30 October 1981)

A normalized and centered spectral profile $P(\omega)$ is conveniently expressed as $P(\omega) = (2\pi)^{-1}\int_{-\infty}^{\infty} d\tau \exp[-i\omega\tau - \int_0^\tau dt\,(\tau - t)\Psi(t)]$, which defines $\Psi(t)$. We consider cases satisfying certain conditions, in particular $|\int_0^\infty dt\Psi(t)| < \infty$. A "broadening strength" $\Lambda = \Omega\Theta$ is defined, where Ω^2 and Θ are the amplitude and characteristic time scale of $\Psi(t)$, respectively. If we let, formally, Λ vary freely, $P(\omega)$ tends to a Lorentzian when $\Lambda \to 0$ (weak broadening or strong narrowing limit), and to a Gaussian when $\Lambda \to \infty$ (strong broadening limit). To deal with situations where one of these limit shapes is only approached, for Λ not being small or large enough, we obtain for $P(\omega)$ and its shift, width, and asymmetry two kinds of expansions: one in powers of Λ^2, starting with the weak broadening limit; the other in powers of Λ^{-1}, starting with the strong broadening limit. Such expansions should allow one to describe spectral profiles over much wider ranges of physical conditions than does the use of just the Lorentzian and Gaussian limit shapes.

PACS numbers: 32.70.Jz, 02.30.Mv

1. INTRODUCTION

Spectral lines broadened by random perturbations often have, in certain limiting conditions, one of two very simple shapes: Lorentzian or Gaussian.[1-7] In general the Lorentzian shape corresponds, in some sense, to a weak broadening, or narrowing, condition, while the Gaussian shape is usually associated with strong broadening conditions.

Situations may occur wherein a spectral line only approaches one of these two shapes, without the physical conditions being extreme enough for that shape to be assumed exactly. Our purpose here is to deal with such situations, by constructing two kinds of expansions for the line shape and its characteristic shape parameters, shift, width, and asymmetry; one kind of expansions is applicable in the neighborhood of the Lorentzian limit, the other in that of the Gaussian limit.

Such expansions should allow one to describe line shapes over much wider ranges of physical conditions than does the use of just the Lorentzian and Gaussian, for the approach to these limit shapes is often slow in function of the relevant physical parameters.[7] A welcome feature of these expansions is that some of their coefficients are mutually interrelated, thus permitting some measure of quantitative comparison with experiment, without need for explicit computations.

Expansions of the above kind have recently been found very useful for analyzing pressure broadened spectral lines.[7] In the present paper, we wish to render them more readily accessible for other applications, and present more general results than those given for the specific needs of pressure broadening theory.

Such expansions were hinted at by Anderson,[8] who calculated the first corrections to the Lorentzian and Gaussian limit widths, for a special case, and interpolated between the two. Such interpolation, however, may not always be feasible.[9]

In Sec. 2, we set down the basic expressions we shall be dealing with, while Sec. 3 briefly reviews the basic theory of randomly perturbed spectral lines relevant here. The weak and strong broadening expansions are established in Secs. 4 and 5, respectively, and discussed in Sec. 6.

The Gaussian limit naturally evokes the Central Limit Theorem (CLT) of probability theory.[10,11] In Sec. 7, we consider the case that the spectrum depends in a symmetric manner on N random variables, as is the case in pressure broadening.[7] The Gaussian limit is then a case of the Central Limit Theorem, or a generalization thereof when the N variables are not mutually independent; this latter case may be of interest *per se*, and will be discussed in more detail elsewhere.

A. Notation and conventions

Given any function of time $f(t)$, its Fourier transform (FT) will be identified with a hat:

$$\hat{f}(\omega) = \mathrm{FT}\{f(t)\} = (2\pi)^{-1}\int_{-\infty}^{\infty} dt\, e^{-i\omega t} f(t). \tag{1.1}$$

Derivatives with respect to time are indicated by dots:

$$\dot{f}(t) = df/dt, \quad f^{k\cdot}(t) = (d/dt)^k f(t), \tag{1.2}$$

and with respect to frequency by primes:

$$\hat{f}'(\omega) = df/d\omega. \tag{1.3}$$

Convolution is denoted

$$(f^* g)(\omega) = \int_{-\infty}^{\infty} d\omega' f(\omega - \omega') g(\omega'). \tag{1.4}$$

$f^{*N}(\omega)$ is the Nth convolution power of $f(\omega)$.

$f(+0)$ and $f(-0)$ signify the limits of $f(t)$ as $t \to 0$ from above and below, respectively.

$f(t) \sim t^{-m}$ as $t \to \infty$ means that $f(t)$ behaves like t^{-m} as $t \to \infty$.

 0022-2488/83/020380-17$02.50

$f(t) \lesssim t^{-m}$ as $t \to \infty$ means that $f(t)$ vanishes faster than t^{-m} as $t \to \infty$.

$f(t) \sim t^{-\infty}$ as $t \to \infty$ means that $f(t)$ vanishes faster than any power of t^{-1} as $t \to \infty$.

$f(t)$ is said to be C^k at t_0 if all its derivatives of order $\leqslant k$ are continuous at t_0.

The Cauchy principal part is understood whenever the generalized function ω^{-1} appears inside an integral; $(\omega \pm i0)^{-1} = \omega^{-1} \mp i\pi\delta(\omega)$, where δ is the Dirac function; $(\omega \pm i0)^{-2} = -(d/d\omega)(\omega \pm i0)^{-1}$, in the sense of generalized functions.[12]

Complex conjugation is indicated with a star.

2. BASIC EXPRESSIONS

Although the results we shall obtain are of more general applicability, it will be useful, for intuition, to explicitly consider a simple model, that of a randomly modulated oscillator[2,3]

$$X(t) = \exp\left[i\nu_0 t + i\int^t dt' U(t')\right]. \tag{2.1}$$

Here, ν_0 is the natural frequency of the oscillator and $U(t)$ the random frequency modulation. As a typical example, $X(t)$ might be an atomic dipole randomly perturbed by interaction with its surroundings.

The basic quantity we are interested in is the power spectrum

$$\mathscr{P}(\nu) = \operatorname*{Lim}_{T\to\infty}(4\pi T)^{-1}\langle\left|\int_{-T}^{T} dt\, e^{-i\nu t}X(t)\right|^2\rangle, \tag{2.2}$$

where $\langle\ \rangle$ denotes a stochastic average, with respect to which we assume time translation invariance (stationarity[13]):

$$\langle U(t_1)U(t_2)\cdots U(t_k)\rangle = \langle U(t_1+\tau)U(t_2+\tau)\cdots U(t_k+\tau)\rangle. \tag{2.3}$$

Physically, $P(\nu)$ might be the power absorbed from incident light of frequency ν by our model dipole.

It is convenient to use the relative frequency

$$\omega = \nu - \nu_0 \tag{2.4}$$

and normalized spectrum

$$P(\omega) = \mathscr{P}(\nu_0+\omega)\Big/\int_{-\infty}^{\infty} d\nu\, P(\nu), \tag{2.5}$$

$$\int_{-\infty}^{\infty} d\omega\, P(\omega) = 1. \tag{2.6}$$

In view of (2.6) and $P(\omega) \geqslant 0$ [obvious from (2.2)], it will be convenient to regard $P(\omega)$ as the probability density of some random variable μ, which is thus defined by

$$\text{Prob}\{\mu = \omega\} = P(\omega) = \langle\delta(\mu-\omega)\rangle, \tag{2.7}$$

where $\langle\ \rangle$ denotes the associated stochastic average:

$$\langle f(\mu)\rangle = \int_{-\infty}^{\infty} f(\omega)P(\omega)d\omega. \tag{2.8}$$

The moments of $P(\omega)$ are thus

$$\langle\mu^k\rangle = \int_{-\infty}^{\infty} \omega^k P(\omega)d\omega. \tag{2.9}$$

We assume that the first moment

$$\langle\mu\rangle = \langle U\rangle = 0. \tag{2.10}$$

This entails no loss in generality, since a nonzero $\langle\mu\rangle$ can always be absorbed in $\nu_0 \to \nu_0 + \langle\mu\rangle$, which is then defined as the *mean* frequency.

Let us introduce the characteristic function of μ,[10,11]

$$C(\tau) = \int_{-\infty}^{\infty} d\omega\, e^{i\omega\tau}P(\omega) = \langle e^{i\mu\tau}\rangle. \tag{2.11}$$

Reciprocally

$$P(\omega) = (2\pi)^{-1}\int_{-\infty}^{\infty} d\tau\, e^{-i\omega\tau}C(\tau). \tag{2.12}$$

By the Wiener–Khintchin theorem,[13] $C(\tau)$ is a normalized autocorrelation function:

$$C(\tau) = \langle x(0)^* x(\tau)\rangle / \langle |x|^2\rangle \tag{2.13}$$

$$= \left\langle \exp\left[i\int_0^\tau dt\, U(t)\right]\right\rangle, \tag{2.14}$$

where we introduced an "interaction representation"

$$x(t) = X(t)e^{-i\nu_0 t} = \exp\left[i\int^t ds\, U(s)\right], \tag{2.15}$$

wherein the unperturbed (or mean) time evolution is factored out. In the special case that $U(t) \equiv U(0) \equiv U$ is static,

$$C_{\text{static}}(\tau) = \langle e^{i\tau U}\rangle, \tag{2.16}$$

$$P_{\text{static}}(\omega) = \langle\delta(U-\omega)\rangle = \text{Prob}\{U=\omega\} \tag{2.17}$$

are just the characteristic function and probability density of U, respectively [in this case, $(\mu, \langle\ \rangle)$ may be identified with $(U, \langle\ \rangle)$].

The fact that $P(\omega)$ is real and positive in (2.11) implies

$$C(-\tau) = C(\tau)^*, \tag{2.18}$$

$$|C(\tau)| \leqslant C(0) = 1. \tag{2.19}$$

Because of (2.18), we may rewrite (2.12) as

$$P(\omega) = \pi^{-1}\text{Re}\, P_+(\omega), \tag{2.20}$$

where $P_+(\omega)$ is the "Fourier–Laplace" transform

$$P_+(\omega) = \int_0^\infty d\tau\, e^{-i\omega\tau}C(\tau). \tag{2.21}$$

The imaginary part of $P_+(\omega)$ is often also of physical interest. It is related to the real part (2.20) by the dispersion relation[14]

$$\text{Im}\, P_+(\omega) = -\int_{-\infty}^{\infty} d\omega' P(\omega')/(\omega-\omega'). \tag{2.22}$$

Usually, $P(\omega) \sim \omega^{-\infty}$ as $\omega \to \pm\infty$ (see Sec. 3), while $\text{Im}\, P_+(\omega) \sim \omega^{-1}$ is slowly decaying [as is obvious from (2.22)].

Equations (2.12) and (2.13) are the basic expressions we shall be dealing with. As mentioned initially, we use the model (2.1) mostly for heuristic purposes; in fact, the autocorrelation (2.13) can be imagined of a much more general form, classical or quantum, than the simple expression (2.14) obtaining for model (2.1). For instance,[7] $x(t)$ might represent a quantum operator in Heisenberg representation, and $\langle(\ )\rangle$ signify $\text{Tr}\,\rho(\ )$, where ρ is some statistical operator (in this case, * indicates Hermitian conjugation). We shall always regard (2.13) as of such a general nature, and all expressions

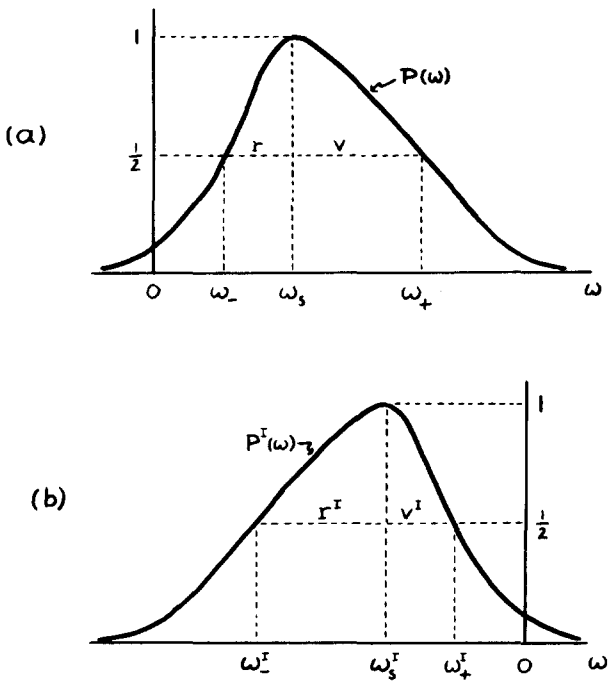

FIG. 1. (a) Definition of the shape parameters: shift $= \omega_s$, width $= r + v$, asym $= (r/v) - 1$. (b) $P^I(\omega)$ is the inversion of $P(\omega)$ about $\omega = 0$.

wherein $U(t)$ does not appear explicitly are to be so understood [of course, expressions containing $U(t)$ pertain to (2.1)].

Rather than $P(\omega)$ itself, its shift, width, and asymmetry are often more practical characterizations of the line shape, especially if it is desired to study the evolution of the spectrum under changing physical conditions. We use the following definitions [Fig. 1(a)].

The shift ω_s is the value of ω maximizing $P(\omega)$:

$$(dP/d\omega)_{\omega=\omega_s} = 0. \tag{2.23a}$$

Half-height frequencies ω_+ and ω_- are defined by

$$P(\omega_\pm) = \tfrac{1}{2}P(\omega_s), \tag{2.23b}$$

with $\omega_- < \omega_s < \omega_+$.

$$\text{width} = r + v = \omega_+ - \omega_-,$$

$$\text{asym} = \left(\frac{r}{v}\right) - 1 = \frac{\omega_s - \omega_-}{\omega_+ - \omega_s} - 1, \tag{2.24a}$$

where $r = \omega_s - \omega_-$ and $v = \omega_+ - \omega_s$ are the "red" and "violet" "half-widths," respectively. One may also add to this list

$$\text{height} = \mathrm{P}(\omega_s). \tag{2.24b}$$

The above definitions may be ambiguous when $P(\omega)$ has a complicated shape (e.g., if it has several local extrema); but for the situations we shall consider, i.e., the vicinities of Lorentzian and Gaussian profiles, there is no difficulty.

It will prove useful to consider the inversion transformation

$$I\colon\ P(\omega) \to P^I(\omega) \equiv P(-\omega). \tag{2.25}$$

We have [see Fig. 1(b)]: $I(r) = v$, $I(v) = r$, and

$$I\omega_s = -\omega_s, \quad I\omega_+ = -\omega_-, \quad I\omega_- = -\omega_+, \tag{2.26a}$$

$$I(\text{shift}) = -\text{shift}, \quad I(\text{width}) = \text{width}, \tag{2.26b}$$

$$I(\text{asym}) = (r^I/v^I) - 1 = (v/r) - 1 = (\text{asym} + 1)^{-1} - 1. \tag{2.26c}$$

We will find in general that the inversion (2.25) can be realized by means of simple modifications of parameters or constants upon which $P(\omega)$ and its shape parameters depend. Thus, e.g., once an expression or expansion is obtained for ω_+, the corresponding result for ω_- can be deduced by applying I [in view of (2.26a)]. Also, if one prefers to define asymmetry as

$$\text{asym}' = (v/r) - 1, \tag{2.27}$$

the latter can be deduced from asym, Eq. (2.24), by applying I [in view of (2.26c)].[15] As a first alternative realization of (2.25), we have, in view of (2.12) and (2.18),

$$I\colon\ C(\tau) \to C(\tau)^*. \tag{2.28}$$

3. GENERAL THEORY

It proves advantageous to express the correlation function (2.13) as

$$C(\tau) = e^{G(\tau)}, \quad G(\tau) = \ln C(\tau). \tag{3.1}$$

In general, physical systems are free of discontinuities, so that $C(\tau)$ and $G(\tau)$ are C^∞ on $(-\infty, \infty)$. This implies, by general properties of Fourier transforms,[12] that $P(\omega) \sim \omega^{-\infty}$ as $\omega \to \pm\infty$. Then, all the moments of $P(\omega)$ exist, and are generated by the Taylor expansion of $C(\tau)$:

$$C(\tau) = \langle e^{i\mu\tau} \rangle = 1 + \sum_{k=1}^{\infty} \frac{(i\tau)^k \langle \mu^k \rangle}{k!}. \tag{3.2}$$

One then has[10]

$$G(\tau) = \langle e^{i\mu\tau} - 1 \rangle_c = \sum_{k=1}^{\infty} \frac{(i\tau)^k \langle \mu^k \rangle_c}{k!}, \tag{3.3}$$

where $\langle \mu^k \rangle_c$ denote the cumulants (or semi-invariants) of $P(\omega)$; these are essentially defined by (3.2) and (3.3), each $\langle \mu^k \rangle_c$ being a *real* polynomial in moments of order $\leq k$; the first few have the explicit expressions [taking account of (2.10)]

$$\langle \mu \rangle_c = \langle \mu \rangle = 0, \quad \langle \mu^2 \rangle_c = \langle \mu^2 \rangle, \quad \langle \mu^3 \rangle_c = \langle \mu^3 \rangle,$$
$$\langle \mu^4 \rangle_c = \langle \mu^4 \rangle - 3\langle \mu^2 \rangle^2, \tag{3.4}$$
$$\ldots.$$

Although, strictly speaking, $G(\tau)$ is C^∞ for real physical systems, there are systems which closely mimick a discontinuous behavior, and true discontinuities may in fact appear mathematically when, e.g., a bulk limit, volume $\to\infty$, number of particles $\to\infty$, is taken; also, certain models may contain discontinuities (e.g., the square well model in pressure broadening[7]). In order to cover such cases, we will allow for possible discontinuities in the third and higher derivatives of $G(\tau)$. Because of (2.18), implying

$$G(-\tau) = G(\tau)^*, \tag{3.5}$$

the points of discontinuity are disposed symmetrically about the origin $\tau = 0$, which may itself be such a point. We shall assume, however, that $\tau = 0$ is not an accumulation point of points of discontinuity, so that there are (symmetric) open

intervals $(-\delta,0)$ and $(0,\delta)$, $\delta>0$, wherein $G(\tau)$ may be Taylor expanded:

$$G(\tau\gtrless 0) = -\tfrac{1}{2}\tau^2\langle\mu^2\rangle_c + \sum_{k=3}^{\infty}\frac{(i\tau)^k\langle\mu^k\rangle_c^{\pm}}{k!}, \tag{3.3'}$$

where

$$\langle\mu^k\rangle_c^{\pm} = G^{k)}(\pm 0)/i^k. \tag{3.6}$$

Discontinuities at $\tau=0$ translate as complex valued "cumulants" (3.6): indeed, (3.5) implies $\langle\mu^k\rangle_c^- = \langle\mu^k\rangle_c^{+*}$, so that (3.6) is real if and only if $G^{k)}(\tau)$ is continuous at $\tau=0$ [i.e., $G^{k)}(+0)=G^{k)}(-0)\Leftrightarrow\langle\mu^k\rangle_c^+ = \langle\mu^k\rangle_c^- = \langle\mu^k\rangle_c^{+*}$]. Of course, $\langle\mu^k\rangle_c^{\pm} = \langle\mu^k\rangle_c$ if $G(\tau)$ is C^k at $\tau=0$.

Because $P(\omega)$ is normalized and centered [Eqs. (2.6) and (2.10)], implying $G(0)=\dot G(0)=0$, $G(\tau)$ is completely determined by its second derivative:

$$G(\tau) = -\int_0^{\tau} d\tau'(\tau-\tau')\Psi(\tau'), \tag{3.7}$$

where we denote

$$\Psi(\tau) = -\ddot G(\tau) = -(d/d\tau)^2\ln C(\tau). \tag{3.8}$$

The quantity $\Psi(\tau)$ completely determines the spectrum $P(\omega)$, and plays a central role in the theory. It has, in certain cases, a direct physical meaning (see below).

In order to motivate certain properties which will be assumed of $\Psi(\tau)$, let us refer to model (2.1) and (2.14), for which case

$$G(\tau) = \left\langle \exp\left[i\int_0^{\tau} dt\, U(t)\right] - 1\right\rangle_c = \sum_{k=1}^{\infty}\frac{i^k}{k!}\int_0^{\tau}dt_1\int_0^{\tau}dt_2\cdots\int_0^{\tau}dt_k\,\langle U(t_1)U(t_2)\cdots U(t_k)\rangle_c, \tag{3.9}$$

where the "generalized cumulants" $\langle U_1U_2\cdots U_k\rangle_c$ are essentially defined[2,16] by (2.14) and (3.9); the first few have the explicit expressions [taking account of (2.10)]

$$\begin{aligned}&\langle U_1\rangle_c = \langle U_1\rangle = 0, \quad \langle U_1U_2\rangle_c = \langle U_1U_2\rangle,\\ &\langle U_1U_2U_3\rangle_c = \langle U_1U_2U_3\rangle,\\ &\langle U_1U_2U_3U_4\rangle_c = \langle U_1U_2U_3U_4\rangle - \langle U_1U_2\rangle\langle U_3U_4\rangle - \langle U_1U_3\rangle\langle U_2U_4\rangle - \langle U_1U_4\rangle\langle U_2U_3\rangle.\end{aligned} \tag{3.10}$$

Such "mixed" cumulants have the notorious *cluster property* of vanishing whenever their arguments separate into two or more statistically independent, i.e., uncorrelated, subsets.[16,17]

Consider now[18]

$$\Psi(\tau) = \left\langle U(0)\exp\left[i\int_0^{\tau} dt\, U(t)\right] U(\tau)\right\rangle_c = \langle U(0)U(\tau)\rangle + \sum_{k=1}^{\infty} i^k\Psi^{(k)}(\tau), \tag{3.11}$$

where

$$\Psi^{(k)}(\tau) = \int_0^{\tau}dt_k\int_0^{t_k}dt_{k-1}\cdots\int_0^{t_2}dt_1\langle U(0)U(t_1)\cdots U(t_k)U(\tau)\rangle_c. \tag{3.12}$$

Note that if the frequency modulation $U(t)$ is Gaussian,[19] $\Psi(\tau)=\langle U(0)U(\tau)\rangle$ is just the autocorrelation of $U(t)$, and is thus of direct physical import.

We make the important assumption that $U(t)$ has a finite correlation time θ.[20] Then, if $\tau>(k+1)\theta$ in (3.12), the set $\{U(0),U(t_1),\ldots,U(t_k),U(\tau)\}$ separates into at least two uncorrelated subsets, as there is at least one gap larger than θ in the sequence $0<t_1<\cdots<t_k<\tau$; this implies, by the cluster property of cumulants, that (3.12) vanishes. We thus conclude that

$$\Psi(\tau)\to 0 \quad \text{as } \tau\to\pm\infty. \tag{3.13}$$

We argued (3.13) on the basis of model (2.1); but (3.13) largely transcends that model, and is presumed to usually be the case whenever $P(\omega)$ consists of a single line, or of several mutually "interacting" lines (i.e., the physical system undergoes transitions between the different line frequencies as time proceeds).[21] How fast $\Psi(\tau)$ tends to zero depends on each particular problem and is usually not easy to determine. We shall assume that

$$\left|\int_0^{\infty} d\tau\,\Psi(\tau)\right| < \infty, \tag{3.14}$$

which happens, e.g., if $\Psi(\tau)\lesssim\tau^{-1}$ as $\tau\to\infty$. Assumption (3.14) is instrumental for the weak broadening limit shape to be Lorentzian, and is justified *a posteriori* whenever this shape is effectively observed to be approached experimentally. In cases where (3.14) does not hold, the weak broadening limit and expansions discussed in this paper do not apply.[22]

A. The broadening strength Λ and scaled function $\psi(t)$

It is useful to distinguish between the size and shape of $\Psi(\tau)$, as these two qualities reflect on $P(\omega)$ in quite different manners.

The size of $\Psi(\tau)$ may be characterized (vertically) by its amplitude Ω^2 taken, e.g., as the initial value

$$\Omega^2 \equiv \Psi(0) = \langle\mu^2\rangle = \langle U^2\rangle, \tag{3.15}$$

and (horizontally) by its characteristic time scale Θ. The latter may really be defined from two different points of view: Θ may be the decay time of $\Psi(\tau)$ taken, e.g., as[23]

$$\Theta_d = \int_0^{\infty} d\tau\left|\frac{\Psi(\tau)}{\Psi(0)}\right|^m \quad \text{(decay time scale)}, \tag{3.16d}$$

with $m>0$ chosen such that the integral exists; or, Θ may be a typical time interval over which $\Psi(\tau)$ varies significantly given, e.g., by[24]

$$\Theta_v^{-1} = \mathrm{Max}\left|\dot\Psi(\tau)/\Psi(0)\right| \quad \text{(variation time scale)}. \tag{3.16v}$$

The times Θ_d and Θ_v will usually be comparable, and we shall not distinguish between the two for simplicity; in cases where they are very different [which may happen, e.g., if $\Psi(\tau)$ is strongly oscillatory, or if (3.14) does not hold, usually implying $\Theta_d=\infty$], it is better to take $\Theta=\Theta_d$ for dealing with weak broadening [if (3.14) holds] and $\Theta=\Theta_v$ for dealing with strong broadening. At any rate, since Θ is introduced mostly for formal purposes, i.e., to get dimensionless quantities and make relative orders of magnitude self-apparent, a rough estimate of it suffices in practice (see Sec. 6 A).

Within model (2.1), Θ may sometimes be comparable to the correlation time θ of $U(t)$ [in particular if $U(t)$ is Gaussian or approximately so], but these two times may also be quite different: e.g., in the static case $U(t) \equiv U(0)$, $\theta = \infty$ while Θ could be anything, finite or infinite. In general, we tend to expect $\Theta \leqslant \theta$.[25]

Taken individually, Ω and Θ are not very determinant, as their numerical values depend on the choice of units (i.e., they only set the scale). What is significant is the dimensionless product

$$\Lambda = \Omega\Theta, \tag{3.17}$$

which may be regarded as an absolute size parameter characterizing $\Psi(\tau)$ [this interpretation seems especially appropriate when m can be chosen $\frac{1}{2}$ in (3.16d), for then $\Lambda = \int_0^\infty d\tau |\Psi(\tau)|^{1/2}$ is just the "area" of $|\Psi(\tau)|^{1/2}$].

As to the shape of $\Psi(\tau)$, it may be represented by

$$\psi(t) = \Psi(\Theta t)/\Psi(0), \tag{3.18}$$

a dimensionless function of dimensionless time, of unit size [i.e., of unit amplitude and characteristic time scale].[26] We shall also need the Fourier transform of $\psi(t)$, $\hat{\psi}(v)$, which is real [since $\psi(-t) = \psi(t)^*$ by (3.5)] and normalized:

$$\int_{-\infty}^{\infty} dv\, \hat{\psi}(v) = \psi(0) = 1. \tag{3.19}$$

In the case of model (2.1) with a Gaussian modulation $U(t)$, $\psi(t)$ and $\hat{\psi}(v)$ are the (normalized and scaled) autocorrelation and power spectrum of $U(t)$, respectively—very important objects indeed.

Defining [compare (3.7)]

$$g(t) = -\int_0^t ds(t-s)\psi(s) = \Lambda^{-2} G(t\Theta), \tag{3.20}$$

we rewrite (2.21) and (3.1) as

$$P_+(\omega) = \Theta \int_0^\infty dt \exp[-i(\omega\Theta)t + \Lambda^2 g(t)] \tag{3.21}$$

or, equivalently,

$$P_+(\omega) = \Omega^{-1} \int_0^\infty dt \exp[-i(\omega/\Omega)t + \Lambda^2 g(t/\Lambda)]. \tag{3.22}$$

The shape of $P_+(\omega)$ is seen to be entirely determined by Λ and ψ; Θ or Ω^{-1} merely scale $P_+(\omega)$, and the explicit appearance of either may be eliminated by simply taking it as unit of time.

Weak and strong broadening correspond to small and large values of the "broadening strength" Λ, respectively, and our main purpose is to construct expansions in powers of Λ^2 and of Λ^{-1} for $P(\omega)$ and its shift, width, and asymmetry. In so doing, we shall formally treat Λ as a free parameter that can be varied at will, independently of ψ; in particular, we shall contrive weak and strong broadening limits $\Lambda \to 0$ and $\Lambda \to \infty$. However, one should be aware that in reality, Λ does not always enjoy such freedom: for one thing, Λ will often not be an experimentally controllable parameter; more so, even in theory Λ cannot always be varied arbitrarily, for this is likely to destroy the positivity of $P(\omega)$. It is only in those special cases where $\hat{\Psi}(v)$ is itself positive that Λ may be varied freely without risk as to the positivity of $P(\omega)$ (see Sec. 6 C); and indeed, in the familiar examples where Λ is variable (the Gaussian approximation in magnetic resonance,[1-4] the Anderson–Talman approximation in pressure broadening[5,7]), $\hat{\psi}(v)$ is the power spectrum of a physical observable, and evidently positive. Still, for convenience of discussion, we will allow Λ to vary freely, as already said; but we keep in mind that in any concrete case, Λ has a specific value, and only in special cases is it actually variable, experimentally or theoretically.

B. Small and large time behaviors of $g(t)$

In weak ($\Lambda \ll 1$) and strong ($\Lambda \gg 1$) broadening conditions, $P(\omega)$ is mostly determined by the behaviors of $g(t)$ at large and small times, respectively.

Let us start with the large time behavior. Assumption (3.14) implies $g(t) \sim t$ as $t \to \pm\infty$. To display this explicity, let us rewrite (3.20) as (the following results are for $t \geqslant 0$)[27]

$$g(t) = g^\#(t) + \beta t, \tag{3.23}$$

where

$$\beta \equiv id - b = -\int_0^\infty dt\, \psi(t) = \dot{g}(\infty), \tag{3.24}$$

b and d real, and

$$g^\#(t) = \int_0^\infty s\, ds[\psi(s) - \psi(s+t)] \tag{3.25}$$

increases slower than t, i.e., $g^\#(t)/t \to 0$ as $t \to \infty$. Inequality (2.19) implies $b \geqslant 0$; we shall assume more restrictively that

$$b > 0. \tag{3.26}$$

When this is not the case, the weak broadening results discussed in this paper do not apply.

If the integral of each term in the integrand of (3.25) exists separately [which happens, e.g., if $\psi(t) \lesssim t^{-2}$ as $t \to \infty$], we have

$$g(t) = h(t) + \alpha + \beta t, \tag{3.27}$$

where

$$\alpha \equiv ia - c = \int_0^\infty s\, ds\, \psi(s) = -h(+0), \tag{3.28}$$

a and c real, and

$$h(t) = -\int_0^\infty s\, ds\, \psi(s+t). \tag{3.29}$$

Because $\psi(s)$ is of unit size, $\alpha, \beta, g^\#$, and h are of order 1 in magnitude, and $h(t) \to 0$ with decay time or order 1.

It is useful to have some of the above functions and constants expressed directly in terms of $\hat{\psi}(v)$, especially since the latter may in certain cases be directly accessible experimentally (see Sec. 6 C). We have, firstly,[28]

$$g(t) = \int_{-\infty}^{\infty} dv \frac{(e^{ivt} - 1 - ivt)\hat{\psi}(v)}{v^2}. \tag{3.30}$$

Note that this is already in the form (3.27), if only we interpret v^{-2} as the generalized function $(v + i0)^{-2}$, thereby giving meaning to the separate integral of each of the three terms in the integrand [provided $\hat{\psi}'(v)$ is not singular at $v = 0$].[29] We deduce from (3.30) with v^{-2} interpreted as above, or directly from (3.24) and (3.28) ($\hat{\psi}' \equiv d\hat{\psi}/dv$),

$$b = \pi\hat{\psi}(0), \quad d = -\int_{-\infty}^{\infty} dv\, \frac{\hat{\psi}(v)}{v}, \tag{3.31}$$

$$a = \pi\hat{\psi}'(0), \quad c = \int_{-\infty}^{\infty} dv\, \frac{\hat{\psi}'(v)}{v}.$$

Let us note, for later use, the following realizations of the inversion (2.25):

$$I:\ g \to g^*;\quad d \to -d,\quad (g^{\#} \to g^{\#*}),$$
$$\text{or}\quad (a \to -a,\quad h \to h^*). \tag{3.32}$$

As to the small time behavior of $g(t)$, it is conveniently exposed as

$$g(t) = -\tfrac{1}{2}t^2 + t^3\Gamma(t) \tag{3.33a}$$

$$= -\tfrac{1}{2}t^2 + \sum_{k=3}^{\infty} \frac{(it)^k \Gamma_{k-2}}{k!}, \tag{3.33b}$$

where

$$\Gamma(t) = -\tfrac{1}{2}\int_0^1 ds(1-s)^2\dot{\psi}(st) \tag{3.34}$$

is of non-negative order in t. In (3.33b), equivalent to (3.3'), we denoted

$$\Gamma_k = \psi^{k\cdot}(+0)/i^k. \tag{3.35a}$$

If $\psi(t)$ is C^k at $t = 0$,

$$\Gamma_k = \int_{-\infty}^{\infty} dv\, v^k\hat{\psi}(v) \tag{3.35b}$$

is the k th moment of $\hat{\psi}(v)$, and real [in concordance with (3.3)]. Note that expansion (3.33b) with (3.35b) is immediate from (3.30).

Condition (2.19) implies $\mathrm{Re}\, g(t) < \mathrm{Re}\, g(0) = 0$. We shall assume more restrictively that

$$\mathrm{Re}\, g(t) < 0 \quad \text{for } t > 0, \tag{3.36}$$

i.e., the value 0 is attained only at $t = 0$. This is a rather weak assumption, as it is clear, on considering (2.14), that in only rather special cases will $C(\tau)$ reassume the value 0 outside $\tau = 0$.[30] Whenever (3.36) is not the case, the strong broadening results discussed in this paper do not apply.

The behaviors $g(t) \sim t$ as $t \to \infty$ and $g(t) \sim t^2$ as $t \to 0$ [and (3.36)] lead to Lorentzian and Gaussian spectral profiles in the respective limits $\Lambda \to 0$ and $\Lambda \to \infty$.

C. $\Lambda \to 0$: Weak broadening or strong narrowing limit

Consider (3.21) with $g(t)$ given by (3.23) or (3.27). As $\Lambda \to 0$, $\Lambda^2 g(t) \to \Lambda^2 \beta t$ and [in view of (2.20)]

$$P(\omega) \to P_{\mathrm{WB}}(\omega) = \pi^{-1}\Theta\, \frac{\Lambda^2 b}{(\omega\Theta - \Lambda^2 d)^2 + (\Lambda^2 b)^2}, \tag{3.37a}$$

with corresponding shape parameters

$$\text{shift} = \Lambda\Omega d, \quad \text{width} = 2\Lambda\Omega b, \quad \text{asym} = 0. \tag{3.37b}$$

When (3.27) holds, $P_{\mathrm{WB}}(\omega)$ constitutes a good approximation to $P(\omega)$ for

$$\Lambda^2 \ll 1, \quad |\omega\Theta| \ll 1; \tag{3.38}$$

the first condition ensures that only $\Lambda^2\beta t$ in $\Lambda^2 g(t)$ ever becomes appreciable [recall that all quantities in (3.27) are of order 1], the second that the Fourier operator $\int dt\, e^{-i\omega\Theta t}$ does not "see" details on a time scale less than 1, so that it is blind to the fact $g(t) \neq \beta t$ when $t < 1$. Note that the frequency range of validity, $|\omega| \ll \Theta^{-1}$, covers the important part of the spectral profile, since width $\simeq \Lambda\Omega = \Lambda^2\Theta^{-1} \ll \Theta^{-1}$ when $\Lambda^2 \ll 1$; this often prompts omission of the second of conditions (3.38). Validity conditions for the approximation $P(\omega) \simeq P_{\mathrm{WB}}(\omega)$ when (3.27) does not hold are given in Sec. 4.

In (3.37), the width is much less (since $\Lambda \ll 1$) than the mean deviation Ω; this means that most of the "energy" is spread out in the wings of $P(\omega)$, which are indeed far extending in (3.37a). Recall though, from the discussion following Eq. (3.1), that for real physical systems, $P(\omega) \sim \omega^{-\infty}$ as $|\omega| \to \infty$, i.e., $P(\omega)$ eventually decays much faster than (3.37a) in the far wings $|\omega\Theta| > 1$.

D. $\Lambda \to \infty$: Strong broadening limit

When $\Lambda \to \infty$, $e^{\Lambda^2 g(t/\Lambda)}$ in (3.22) becomes vanishingly small, because of (3.36), except in the vicinity of $t = 0$, where $\Lambda^2 g(t/\Lambda) \to -\tfrac{1}{2}t^2$ in view of (3.33); thus

$$P(\omega) \to P_{\mathrm{SB}}(\omega) = (2\pi\Omega)^{-1/2}\, e^{-(1/2)(\omega/\Omega)^2} \tag{3.39a}$$

with shape parameters

$$\text{shift} = 0, \quad \text{width} = 2\kappa\Omega, \quad \text{asym} = 0, \tag{3.39b}$$

where[31]

$$\kappa = (2\ln 2)^{1/2} = 1.17741\cdots. \tag{3.40}$$

Here, the width is roughly equal to the mean deviation Ω. $P_{\mathrm{SB}}(\omega)$ constitutes a good approximation to $P(\omega)$ if (loosely)[32]

$$\Lambda \gg \mathrm{Max}|\dot{\psi}(t)| \simeq 1. \tag{3.41}$$

The main purpose of the paper is to extend the limit results (3.37) and (3.39) by means of two kinds of expansions for $P(\omega)$ and its shape parameters: one in powers of Λ^2 starting with (3.37), the other in powers of Λ^{-1} starting with (3.39).

4. WEAK BROADENING EXPANSIONS

In this section, we construct expansions in powers of

$$\gamma = \Lambda^2 \tag{4.1}$$

for $P(\omega)$ and its shape parameters. As we here deal with the weak broadening regime $\Lambda \ll 1$, it is appropriate to use the weak decay time $(\Lambda\Omega b)^{-1}$ of $C(\tau)$ as basic time scale. We accordingly re-express (3.21) and (3.27) as

$$P_+(\omega) = (\Lambda\Omega b)^{-1}\hat{P}_+(v), \quad v \equiv \frac{\omega - \Lambda\Omega d}{\Lambda\Omega b}, \tag{4.2}$$

where

$$\hat{P}_+(v) = \int_0^{\infty} dt\, \exp[-(1+iv)t + \gamma\alpha + \gamma h(t/\gamma b)]. \tag{4.3}$$

For simplicity of notation, we use (3.27) whether α and $h(t)$ exist or not, it being understood that in the latter case, $\alpha = 0$ and $h(t)$ stands for $g^{\#}(t)$, Eq. (3.25).

In the weak broadening limit $\gamma \to 0$,

$$\hat{P}_+(v) \to (1+iv)^{-1} \tag{4.4}$$

and the (scaled) spectrum

$$\hat{P}(v)\equiv\pi^{-1}\mathrm{Re}\,\hat{P}_{+}(v)\rightarrow\hat{P}_{\mathrm{WB}}(v)=\pi^{-1}(1+v^{2})^{-1} \tag{4.5}$$

equivalent to (3.37). Our intention is to express $\hat{P}_{+}(v)$ and $\hat{P}(v)$ as products of their weak broadening limits times bivariate expansions in powers of γ and v. As a first step, we rewrite (4.3) as

$$\hat{P}_{+}(v)=(1+iv)^{-1}\{e^{\gamma a}[1+(1+iv)T(v)]\}, \tag{4.6}$$

where

$$T(v)=\gamma b\int_{0}^{\infty}dt\,e^{-\gamma b(1+iv)t}[e^{\gamma h(t)}-1]. \tag{4.7}$$

There remains to expand $\{\cdots\}$ in γ and v.

A. Case $h(t)\sim t^{-\infty}$ as $t\rightarrow\infty$

It is conceptually simpler to first assume that $\psi(t)$, thence $h(t)$, vanish faster than any power of t^{-1} as $t\rightarrow\infty$. We may then expand the exponentials in (4.7), obtaining

$$T(v)=\gamma\sum_{k=1}^{\infty}\sum_{j=0}^{\infty}\frac{(\gamma v-i\gamma)^{j}\gamma^{k}T_{kj}}{j!k!}, \tag{4.8}$$

where the

$$T_{kj}\equiv T^{r}_{kj}+iT^{i}_{kj}=b^{j+1}\int_{0}^{\infty}dt\,(-it)^{j}h(t)^{k} \tag{4.9}$$

are of order 1, since b is, as well as $h(t)$ in both amplitude and decay time.

1. Expansion of $\hat{P}(v)$

Introducing (4.8) into (4.6), expanding $e^{i\gamma a}$ [recall: $\alpha=ia-c$, Eq. (3.28)], and rearranging, we obtain $\hat{P}_{+}(v)$ as $e^{-\gamma c}(1+iv)^{-1}$ times a bivariate expansion in powers of γ and γv. The corresponding expansion of $\hat{P}(v)$ is readily deduced; to second order in γ (see Appendix B for the third order terms)

$$\hat{P}(v)=\pi^{-1}(1+v^{2})^{-1}e^{-\gamma c}\times[1+(\gamma v)a+\gamma^{2}(T^{r}_{10}-\tfrac{1}{2}a^{2})+(\gamma v)^{2}T^{r}_{10}+\cdots]. \tag{4.10}$$

The validity conditions of the weak broadening approximation $(1+v^{2})^{-1}$ are again seen to be

$$\gamma\ll 1,\quad |\gamma v|\ll 1, \tag{4.11}$$

equivalent to (3.38). In (4.10), we kept $e^{-\gamma c}$ unexpanded, as it does not affect the shape of $\hat{P}(v)$, and moreover, there are cases where $e^{-\gamma c}(1+v^{2})^{-1}$ is a much better approximation than $(1+v^{2})^{-1}$, valid outside conditions (4.11).[33,34]

By retaining in (4.10) terms up to orders k and j in γ and γv, respectively, we obtain higher order weak broadening approximations $\hat{P}^{(k,j)}_{\mathrm{WB}}$, valid over wider ranges in γ and v; e.g.,

$$\hat{P}^{(0,1)}_{\mathrm{WB}}(v)=\pi^{-1}e^{-\gamma c}\frac{1+\gamma va}{1+v^{2}} \tag{4.12}$$

and[35,36]

$$\hat{P}^{(0,\infty)}_{\mathrm{WB}}(v)=e^{-\gamma c}\frac{b^{-1}\hat{\psi}(\gamma bv)}{1+v^{2}}. \tag{4.13}$$

The numerator of (4.13) may also be viewed as the first term of an expansion in γ alone. Approximation (4.13) is valid at all values of v, to lowest order in γ, in contrast to (4.5) or (4.12) valid only in the line center. However, one should be aware that, although $\hat{P}(v)\rightarrow\hat{P}^{(0,\infty)}_{\mathrm{WB}}(v)$ rigorously as $\gamma\rightarrow 0$, there may be, at *finite* values of γ however small, frequency ranges wherein $\hat{P}(v)$ is dominated by higher order terms in γ; this is because $\hat{\psi}(v)$ may vanish very rapidly beyond some frequency, while higher order terms, which involve convolution powers of $\hat{\psi}(v)$ (see Sec.6 D), may extend much farther in frequency and thus eventually dominate, however small γ.

2. Expansions of shape parameters

We next construct expansions in powers of γ for the shape parameters of $\hat{P}(v)$. We first postulate for v_{s} and $v_{\pm}$ [see (2.23)] expansions

$$v_{s,\pm}=v^{(0)}_{s,\pm}+\gamma v^{(1)}_{s,\pm}+\gamma^{2}v^{(2)}_{s,\pm}+\cdots, \tag{4.14}$$

where $v^{(0)}_{s}=0$ and $v^{(0)}_{\pm}=\pm 1$, as determined from (4.5). To determine the $v^{(j)}_{s,\pm}$, we introduce (4.10) into (2.23), then (4.14), and expand throughout in γ. Setting to zero the coefficients of successive powers of γ, we obtain relations between the $v^{(j)}_{s,\pm}$ which are solved recursively. The expansions of width, asymmetry, and height are then deduced by use of (2.24). We get[37,38]

$$\begin{aligned}
\mathrm{shift}_{v}&=\tfrac{1}{2}\gamma a+(\gamma^{3}/24)(a^{3}+12T^{r}_{11})+\cdots,\\
\mathrm{width}_{v}&=2+\tfrac{1}{2}\gamma^{2}(a^{2}+4T^{r}_{10})\\
&\quad+\gamma^{3}(T^{r}_{20}-2aT^{i}_{10}+2T^{i}_{11})+\cdots,\\
\mathrm{asym}_{v}&=-\gamma a+\tfrac{1}{2}\gamma^{2}a^{2}\\
&\quad-(\gamma^{3}/12)(a^{3}+12aT^{r}_{10}+36T^{r}_{11})+\cdots,\\
\mathrm{height}_{v}&=\pi^{-1}e^{-\gamma c}[1+\gamma^{2}(-\tfrac{1}{4}a^{2}+T^{r}_{10})+\cdots]\\
&=\pi^{-1}[1-\gamma c+\gamma^{2}(\tfrac{1}{2}c^{2}-\tfrac{1}{4}a^{2}+T^{r}_{10})+\cdots].
\end{aligned} \tag{4.15}$$

The subscripts v indicate that these are the shape parameters of $\hat{P}(v)$, from which those of $P(\omega)$ are trivially deduced.

The above results were obtained under the assumption $h(t)\sim t^{-\infty}$; more generally, they are valid provided the asymptotic behavior of $h(t)$ is such that the T_{kj}, Eq. (4.9), exist for all k and j.[39]

B. General case

But if, e.g.,

$$h(t)\sim t^{-\sigma}\quad\text{as } t\rightarrow\infty\quad(-1<\sigma<\infty) \tag{4.16}$$

(if $-1<\sigma<0$, h stands for $g^{\#}$ and $\alpha=0$ as we convened), the T_{kj} are ill defined whenever $j-k\sigma\geqslant -1$. This means that only the finite form of the Taylor expansion (4.8) can be used [up to order $k\sigma-1$ in $\gamma(v-i)$ for each value of k], and likewise for (4.10). The remainders may be expanded in γ and v, but the result is no longer in integral powers of γ and γv; rather, we get nonanalytic expansions in γ and v; likewise, expansions (4.15) are valid only up to order $\sigma+1$ in γ,[40] the remainders again being expressible as nonanalytic expansions in γ. Let us see this explicitly.

1. Expansion of $\hat{P}(v)$

Let us go back to (4.7) and, keeping $e^{-\gamma bt}$ unexpanded, obtain in lieu of (4.8),

$$T(v)=\sum_{j=0}^{\infty}\sum_{k=1}^{\infty}\frac{v^{j}\gamma^{k}S^{0}_{kj}}{k!j!}, \tag{4.17}$$

where we define (the dependence on x is introduced for later use)

$$S^x_{kj} = (\gamma b)^{j+1}\int_0^\infty dt\, e^{-\gamma bt(1+ix)}(-it)^j h(t)^k. \tag{4.18}$$

The factor $e^{-\gamma bt}$ guarantees the existence of S^x_{kj} for all k and j, but at the cost of a generally nonanalytic γ dependence: indeed, it is shown in Appendix A that if $h(t)$ behaves asymptotically like (4.16), then (for $k \geqslant 1$)

$$S^x_{kj} = \gamma^{j+1}\mathrm{Pol}(\gamma) + (\text{const})\gamma^{k\sigma} \times \begin{cases} 1, & \text{if } j - k\sigma > -1 \text{ or not integer} \\ \ln(\gamma), & \text{if } j - k\sigma = \text{integer} \leqslant -1 \end{cases}, \tag{4.19}$$

where $\mathrm{Pol}(\gamma)$ denotes some polynomial in γ. If more generally, $h(t)$ behaves asymptotically as a sum of terms t^{-r} [and $t^{-r}\ln(t)$, etc.], then

$$S^x_{kj} = \gamma^{j+1}(S^x_{kj})' + \gamma^{k\sigma}(S^x_{kj})'', \tag{4.20}$$

where σ is the lesser of the exponents r, and $(S^x_{kj})'$ and $(S^x_{kj})''$ are of non-negative orders in γ, $(S^x_{kj})''$ generally being nonanalytic in γ. Expansion (4.17) thus becomes

$$T(v) = \sum_{j=0}^{\infty}\sum_{k=1}^{\infty} \frac{(\gamma v)^j \gamma^{k+1}(S^0_{kj})' + v^j\gamma^{(1+\sigma)k}(S^0_{kj})''}{k!\,j!}. \tag{4.21}$$

The first sum is similar to (4.8), to which it must become identical if $\sigma = \infty$ and the S^0_{kj} are expanded in γ [for the second sum is then of order $\gamma^\infty = 0$ (since $\gamma < 1$ here)]. The second sum is of a very different nature, being in powers of $\gamma^{1+\sigma}$ and v rather than of γ and γv.

Equation (4.21) is to be substituted into (4.6), the S_{kj} replaced by their (nonanalytic) expansions in γ, and the result rearranged; this will yield $\widehat{P}_+(v)$ and $\widehat{P}(v)$ as products of their weak broadening limits times nonanalytic expansions in γ and v [of course identical to (4.10) up to terms containing ill-defined T_{kj}'s].

The validity conditions of the weak broadening approximation (4.5) are again (4.11) if $\sigma \geqslant 0$; but if $-1 < \sigma < 0$, we have the more stringent requirements[41]

$$\gamma^{1+\sigma} \ll 1, \quad |v|\gamma^{1+\sigma} \ll 1 \tag{4.22}$$

as one would expect, since βt then has a harder time dominating $g(t)$ [consider (3.27) with (4.16) and $-1 < \sigma < 0$].

2. Expansions of the shape parameters

We now construct the expansions of the shape parameters for the above case that the T_{kj} do not all exist. Because of the second sum in (4.21), the expansion of $\widehat{P}(v)$ deduced from it [akin to (4.10)] is no longer convenient for deducing the expansions of $v_\pm$.[42] We rather use the following expansion in γ and $v - x$, where x is to be assigned values in the neighborhoods of which we need to evaluate $\widehat{P}(v)$[43]:

$$\widehat{P}(v) = \pi^{-1}\sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \frac{(v-x)^j\rho^k K^x_{kj}}{k!\,j!}, \tag{4.23}$$

where we define

$$\rho = \gamma, \quad K^x_{kj} = \mathrm{Re}\, S^{x\#}_{kj}, \quad \text{if } \sigma \geqslant 0,$$
$$\rho = \gamma^{1+\sigma}, \quad K^x_{kj} = \gamma^{-k\sigma}\mathrm{Re}\, S^{x\#}_{kj}, \quad \text{if } -1 < \sigma < 0, \tag{4.24}$$

the $S^{x\#}_{kj}$ being given by (4.18), but with h replaced by $g^\#$ [or by $ia + h$ if (3.27) holds]. The K^x_{kj} are hereby defined of non-negative order in γ.

We may now expand each term of (2.23) about the value taken by its argument as $\gamma \to 0$, i.e., we use (4.23) with $x = 0$ for expanding $\widehat{P}(v_s)$, and with $x = \pm 1$ for expanding $\widehat{P}(v_\pm)$.[44] We then introduce (4.14) with γ replaced by ρ, expand throughout in ρ, and set to zero the coefficients of successive powers of ρ. We thereby obtain[37,45]

$$\begin{aligned} \mathrm{shift}_v &= \tfrac{1}{2}\rho K^0_{11} + \rho^2(K^0_{11}K^0_{12} + K^0_{21}) + \cdots, \\ \mathrm{width}_v &= 2 + 2\rho(K^+_{10} + K^-_{10} - K^0_{10}) + \cdots, \\ \mathrm{asym}_v &= \rho(K^0_{11} - 2K^+_{10} + 2K^-_{10}) + \cdots, \end{aligned} \tag{4.25}$$

where $K^\pm_{kj} \equiv K^{\pm 1}_{kj}$. There remains to replace each K^x_{kj} by its own expansion in γ, and regroup terms of same order. Note that if $-1 < \sigma < 0$, the expansion parameter $\rho = \gamma^{1+\sigma}$ is a fractional power of γ [compare (4.22)].

In case (3.27) holds, and moreover $h(t) \lesssim t^{-1}$ as $t \to \infty$, it is more convenient to have the contributions from a and $h(t)$ separated out; also, the K_{kj} or S_{kj} are then of needlessly high order in γ.[46] Referring back to (4.18) and (4.19), we rather define

$$\Gamma^x_{kj} = \Gamma^{xr}_{kj} + i\Gamma^{xi}_{kj} = \gamma^{-1}S^x_{kj} \tag{4.26}$$

of non-negative order in γ (since $\sigma \geqslant 1$ here) and obtain, in lieu of (4.25),[47]

$$\begin{aligned} \mathrm{shift}_v &= \tfrac{1}{2}\gamma a + \tfrac{1}{2}\gamma^2\Gamma^{0r}_{11} + (\gamma^3/24)(a^3 + 6\Gamma^{0r}_{21} \\ &\quad + 6a\Gamma^{0r}_{12} - 12a\Gamma^{0i}_{11}) + \cdots, \\ \mathrm{width}_v &= 2 + \tfrac{1}{2}\gamma^2(a^2 - 4\Gamma^{0r}_{10} + 4\Gamma^{+r}_{10} + 4\Gamma^{-r}_{10}) + \cdots, \\ \mathrm{asym}_v &= -\gamma a + \tfrac{1}{2}\gamma^2(a^2 + 2\Gamma^{0r}_{11} - 4\Gamma^{+r}_{10} + 4\Gamma^{-r}_{10}) + \cdots. \end{aligned} \tag{4.27}$$

Clearly, once the S^x_{kj} are replaced by their (nonanalytic) expansions in γ, and (4.27) is properly rearranged, the latter must become identical to (4.15) up to terms containing ill-defined T_{kj}'s [i.e., up to order $\sigma + 1$ in γ][40]; of course, if $h(t) \sim t^{-\infty}$, (4.25) and (4.27) are equivalent to (4.15).

Let us note finally the realizations of the inversion (2.25) relevant here [I below is inversion about $v = 0$; to get inversion about $\omega = 0$, add $d \to -d$ in view of (4.2)[48,49]]:

$$I: a \to -a, \quad T_{kj} \to (-)^j T^*_{kj} \quad \text{or} \quad S^x_{kj} \to (-)^j \bar{S}^{x*}_{kj}$$
$$\text{or} \quad \Gamma^x_{kj} \to (-)^j \bar{\Gamma}^{x*}_{kj}; \quad K^x_{kj} \to (-)^j \bar{K}^x_{kj}. \tag{4.28}$$

These may be used to the ends mentioned in Sec. 2.

5. STRONG BROADENING EXPANSIONS

We will now obtain expansions in powers of

$$\epsilon = \Lambda^{-1} \tag{5.1}$$

for $P(\omega)$ and its shape parameters. As we here deal with the strong broadening regime $\Lambda \gg 1$, it is appropriate to use the strong decay time Ω^{-1} of $C(\tau)$ as basic time scale. We accordingly re-express (3.22) and (3.33) as

$$P_+(\omega) = \Omega^{-1}\tilde{P}_+(\omega/\Omega), \tag{5.2}$$

where

$$\tilde{P}_+(y) = \int_0^\infty dt \exp[-iyt - \tfrac{1}{2}t^2 + \epsilon t^3 \Gamma(\epsilon t)]. \tag{5.3}$$

The (scaled) spectrum

$$\tilde{P}(y) = \pi^{-1}\mathrm{Re}\,\tilde{P}_+(y) = \mathrm{Prob}\{\tilde{\mu} = y\} \tag{5.4}$$

is the probability density of the normalized variable [recall (2.7)]

$$\tilde{\mu} = (\mu - \langle\mu\rangle)/\langle\mu^2\rangle_c^{1/2} \tag{5.5}$$

satisfying

$$\int_{-\infty}^{\infty} dy\, \tilde{P}(y) = 1, \quad \langle\tilde{\mu}\rangle = 0, \quad \langle\tilde{\mu}^2\rangle = 1. \tag{5.6}$$

In the strong broadening limit $\epsilon \to 0$

$$\tilde{P}_+(y) \to E(y), \quad \tilde{P}(y) \to (2\pi)^{-1/2} e^{-(1/2)y^2} \tag{5.7}$$

equivalent to (3.39); we denoted[50]

$$\begin{aligned} E(y) &= \int_0^\infty dt\, e^{-iyt - (1/2)t^2} \\ &= (\pi/2)^{1/2} e^{-(1/2)y^2} - i2^{1/2} D(y/2^{1/2}) \\ &\equiv E^r(y) + iE^i(y), \end{aligned} \tag{5.8}$$

where

$$D(x) = -D(-x) = e^{-x^2}\int_0^x dt\, e^{t^2}$$

is Dawson's integral (tabulated[51]). Note that $E^r(y) \sim y^{-\infty}$ and $E^i(y) \sim y^{-1}$ as $y \to \pm\infty$.

A. Expansion of $\tilde{P}(y)$

Let us now keep ϵ finite and, using (3.33b), expand

$$F(it) \equiv e^{\epsilon t^3 \Gamma(\epsilon t)} = \exp\left[\sum_{k=1}^{\infty} \frac{\epsilon^k \Gamma_k (it)^{k+2}}{(k+2)!}\right] \tag{5.9}$$

$$= 1 + \epsilon b_1(it) + \epsilon^2 b_2(it) + \epsilon^3 b_3(it) + \cdots, \tag{5.10}$$

where the $b_j(z)$ are polynomials in z (extensively studied in Refs. 10):

$$b_0 = 1, \quad b_1(z) = (\Gamma_1/3!)z^3, \tag{5.11}$$

$$b_2(z) = (\Gamma_2/4!)z^4 + \tfrac{1}{2}(\Gamma_1/3!)^2 z^6,$$

We thereby obtain $\tilde{P}_+(y)$ as an expansion in powers of ϵ:

$$\begin{aligned} \tilde{P}_+(y) &= F(-d/dy)E(y) \\ &= \sum_{j=0}^{\infty} \epsilon^j b_j\left(\frac{-d}{dy}\right) E(y). \end{aligned} \tag{5.12}$$

The derivatives of $E^r(y)$ and $E^i(y)$ contained in (5.12) are expressible as[52]

$$(-d/dy)^k E^r(y) = H_k(y)E(y), \tag{5.13a}$$

$$(-d/dy)^k E^i(y) = H_k(y)E^i(y) + B_{k-1}(y), \tag{5.13b}$$

where

$$H_k(y) = e^{(1/2)y^2}(-d/dy)^k e^{-(1/2)y^2} \tag{5.14}$$

are Hermite polynomials, and $B_j(y)$ are real polynomials of order j in y, e.g.,

$$H_3(y) = y^3 - 3y, \quad B_2(y) = -y^2 + 2. \tag{5.15}$$

Note that (5.13b) is useful at moderate values of y, as we require below; but for large y, at which $(d/dy)^k E^i(y) \sim y^{-k-1}$ is small, (5.13b) is clearly not practical, as it then expresses a small number as a difference of large numbers.

If $\psi(t)$ is C^∞ at $t = 0$, as is the case for real physical systems [see discussion following Eq. (3.1)], the Γ_k are real [Eq. (3.35b)], and $F(-d/dy)$ is a real operator, so that $E^i(y)$ does not contribute to the real part of (5.12). the ϵ expression of the spectrum $\tilde{P}(y)$ is then

$$\begin{aligned} \tilde{P}(y) &= (2\pi)^{-1/2} e^{-(1/2)y^2}\{1 + \sum_{j=1}^{\infty} \epsilon^j b_j[z^n \to H_n(y)]\} \\ &= (2\pi)^{-1/2} e^{-(1/2)y^2}\{1 + \epsilon(\Gamma_1/3!)(y^3 - 3y) + \cdots\}, \end{aligned} \tag{5.16}$$

where $b_j[z^n \to H_n]$ is obtained from $b_j(z)$ by replacing each z^n by the Hermite polynomial $H_n(y)$, i.e., writing (5.11) as $b_j(z) = \Sigma_n b_{jn} z^n$, $b_j[z^n \to H_n] = \Sigma b_{jn} H_n$.

If $\psi(t)$ has discontinuous derivatives at points $t_i \neq 0$, (5.16) still applies; but note that by general properties of Fourier transforms,[12] $\tilde{P}(y)$ then has slowly decaying oscillatory tails $\sim y^{-m}\cos[yt_i + (\text{const})]$, not visible on (5.16) [which rather gives the impression that $\tilde{P}(y) \sim y^{-\infty}$; this, however, is the case only if $\psi(t)$ is C^∞ on $(-\infty, \infty)$].

If $\psi(t)$ has discontinuous derivatives at $t = 0$, this leads to complex Γ_k's [see discussion following Eq. (3.6)], so that in (5.12), the slowly decaying $E^i(y)$ contributes to $\tilde{P}(y) = \mathrm{Re}\,\tilde{P}_+(y)$ as well as to $\mathrm{Im}\,\tilde{P}_+(y)$, i.e., $\tilde{P}(y)$ has tails $\sim y^{-m}$, as of course expected.[12]

B. Expansions of shape parameters

We now obtain expansions in powers of ϵ for the shape parameters of $\tilde{P}(y)$. The procedure is quite similar to that used for obtaining (4.25). We first require an expansion of $\tilde{P}(y)$ in powers of ϵ and $(y - Y)$, where Y will be assigned values in the neighborhoods of which we need to evaluate $\tilde{P}(y)$. We have, from (5.12),[53]

$$\tilde{P}(y) = \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \epsilon^k (y - Y)^j P_{kj}^Y, \tag{5.17}$$

where

$$P_{kj}^Y = (1/j!)(d/dY)^j \pi^{-1}\mathrm{Re}\{b_k(-d/dY)E(Y)\}. \tag{5.18}$$

We next postulate for y_s, y_+, and y_- [see (2.23)] expansions

$$y_{s,\pm} = y_{s,\pm}^{(0)} + \epsilon y_{s,\pm}^{(1)} + \epsilon^2 y_{s,\pm}^{(2)} + \cdots, \tag{5.19}$$

with $y_s^{(0)} = 0$ and $y_\pm^{(0)} = \pm\kappa$, as determined from (5.7) [and of course leading to (3.39b)]. We insert (5.17) into (2.23), using $Y = 0$ and $Y = \pm\kappa$ for expanding $\tilde{P}(y_s)$ and $\tilde{P}(y_\pm)$, respectively [using (5.13)], then substitute (5.19), and expand throughout in ϵ. There result relations between the $y_{s,\pm}^{(j)}$ which are solved recursively. We finally obtain, by use of (2.24),[54,37]

$$\begin{aligned} \mathrm{shift}_y &= -\tfrac{1}{2}\epsilon\Gamma_1^r + \epsilon^2(-0.531922\Gamma_1^r\Gamma_1^i \\ &\quad + 0.265961\Gamma_2^i) + \cdots, \\ \mathrm{width}_y &= 2.35482 - \epsilon(0.825413\Gamma_1^i) + \epsilon^2(0.407344\Gamma_1^{r2} \\ &= 0.452683\Gamma_2^r - 0.485461\Gamma_1^{i2} + \cdots, \\ \mathrm{asym}_y &= -\epsilon(0.392470\Gamma_1^r) + \epsilon^2(0.0770164\Gamma_1^{r2} \\ &\quad - 0.800242\Gamma_1^r\Gamma_1^i + 0.290650\Gamma_2^i) + \cdots, \end{aligned} \tag{5.20}$$

where Γ_k^r and Γ_k^i denote the real and imaginary parts of Γ_k. The subscripts y indicate that these are the shape parameters of $\tilde{P}(y)$, from which those of $P(\omega)$ are trivially deduced.

Let us also display these expansions for the case of real Γ_k's [so the $\Gamma_k^i \to 0$ in (5.20)], this time not replacing irrational constants (κ in the present case) by their numerical values:

$$\begin{aligned}
\text{shift}_y &= -\tfrac{1}{2}\epsilon\Gamma_1 - \epsilon^3\left(-\tfrac{1}{4}\Gamma_1^3 + \tfrac{5}{12}\Gamma_1\Gamma_2 - \tfrac{1}{8}\Gamma_3\right) + \cdots, \\
\text{width}_y &= 2\kappa + \epsilon^2\left[\Gamma_1^2\left(-\tfrac{1}{9}\kappa^3 + \tfrac{1}{2}\kappa\right) + \Gamma_2\left(\tfrac{1}{12}\kappa^3 - \tfrac{1}{2}\kappa\right)\right] + \cdots, \\
\text{asym}_y &= -\epsilon\Gamma_1(\kappa/3) + \epsilon^2\Gamma_1^2(\kappa^2/18) + \cdots, \\
\text{height}_y &= (2\pi)^{-1/2}\left[1 + \epsilon^2\left(-\tfrac{19}{24}\Gamma_1^2 + \tfrac{1}{8}\Gamma_2\right) + \cdots\right].
\end{aligned} \tag{5.21}$$

Expansions (5.12) and (5.20) formally apply to any function $\tilde{P}(y)$ satisfying (5.6) (which can always be achieved by suitable rescalings) and (3.36). In practice, ϵ and Γ_k need not always be defined as in (5.1) and (3.35); they need only satisfy the open relation

$$\epsilon^k\Gamma_k = \tilde{\Psi}^k(+0)/i^k = \langle\tilde{\mu}^{k+2}\rangle_c^+, \tag{5.22}$$

where $\tilde{\Psi}(\tau) = -(d/d\tau)^2\ln\langle e^{i\tilde{\mu}\tau}\rangle$ [compare (3.8) and (2.11)] and $\langle\tilde{\mu}^j\rangle_c^+$ are the cumulants of $\tilde{P}(y)$ (real or complex as the case may be); e.g., one might have $\epsilon = 1$, $\Gamma_k = \langle\tilde{\mu}^{k+2}\rangle_c$. In principle, one would like ϵ to be small and the Γ_k of order 1 over some appropriate range of physical conditions [ideally, the Γ_k to remain bounded and $\epsilon \to 0$ as some physical limit is aproached].

Let us note finally the realizations of the inversion (2.25) relevant here:[55]

$$I: \Gamma_k \to (-)^k\Gamma_k^* \quad \text{or} \quad (\epsilon \to -\epsilon, \Gamma_k \to \Gamma_k^*). \tag{5.23}$$

In the case of real Γ_k's, I: $\epsilon \to -\epsilon$, implying, in view of (2.26b), that the expansion (5.21) of shift (width) contains only odd (even) powers of ϵ.

6. DISCUSSION

The expansions obtained in Secs. 4 and 5 are our main results of practical interest. Let us add a few remarks.

A. Concerning Θ

Let us stress again that the time Θ need not be taken as in (3.16) (given as appealing choices, but possibly hard to evaluate), but may be just a crude estimate of the time scale characteristic of $\Psi(\tau)$. In fact, since all results are evidently independent of Θ (all factors Θ would cancel away if reinserted), one may in practice simply do away with Θ, i.e., take $\Theta = 1$, in whichever units are being used (note that the latter are usually adapted to the time scales involved); then, $\Lambda = \langle\mu^2\rangle^{1/2} = \Psi(0)^{1/2}$ and $\psi(t) = \Psi(t)/\Psi(0)$ (Λ and t are here dimensional). This is the procedure used in Ref. 7.

B. A likely form of $G(\tau)$

In many concrete cases, $G(\tau) = \ln C(\tau)$ is only known approximately, e.g., as the first few terms of an expansion in powers of some parameter p (e.g., an interaction strength or a particle number density):

$$G(\tau) = pG_1(\tau) + p^2G_2(\tau) + p^3G_3(\tau) + \cdots \tag{6.1a}$$

and correspondingly

$$\Psi(\tau) = -\ddot{G}(\tau) = p\Psi_1(\tau) + p^2\Psi_2(\tau) + p^3\Psi_3(\tau) + \cdots, \tag{6.1b}$$

where $\Psi_k(\tau) = -\ddot{G}_k(\tau)$. This situation will be met in Sec. 7, and is that prevailing in pressure broadening (p there being the gas pressure).[7] The weak and strong broadening regimes here correspond to small and large values of p, respectively.

Clearly, the characteristic time scale Θ of $\Psi(\tau)$ depends on p. However, it is more convenient to take for Θ some (average, say) value independent of p; for simplicity, we choose $\Theta = 1$ in the present discussion.

As concerns the weak broadening expansions, it may be useful, especially if p is an experimentally accessible parameter against which it is feasible to plot data, to transform the γ expansions (4.10), (4.15), etc., into expansions in p; this, however, may not always be easy, for although $G_1(\tau)$ will usually have the same asymptotic properties as we presumed of $G(\tau)$ (simply because $G \to pG_1$ as $p \to 0$), this may not be the case for $G_2, G_3, \ldots$, and certain delicate resummations (renormalizations) may have to be performed.[56]

As to the strong broadening expansions, their relevant parameters are here (with $\Theta = 1$)

$$\begin{aligned}
\epsilon &= [\Psi(0)]^{-1/2} \\
&= [p\Psi_1(0) + p^2\Psi_2(0) + p^3\Psi_3(0) + \cdots]^{-1/2}
\end{aligned} \tag{6.2a}$$

$$\Gamma_k = \frac{i^{-k}\Psi^{k\cdot}(0)}{\Psi(0)} = i^{-k}\,\frac{p\Psi_1^{k\cdot}(0) + p^2\Psi_2^{k\cdot}(0) + \cdots}{p\Psi_1(0) + p^2\Psi_2(0) + \cdots}. \tag{6.2b}$$

One can see that the strong broadening expansions are now usable only if $\Psi(\tau)$ is well approximated by the first few terms of (6.1b) even when p is large enough that the strong broadening regime obtains. In particular, if the latter obtains while $\Psi \simeq p\Psi_1$, then $P(\omega)$ is a Gaussian of width $\sim\epsilon^{-1}$ proportional to $p^{1/2}$; if the second term of (6.1b) strongly dominates over part of the strong broadening regime, then width $\sim p$ there; likewise, there can be a p range wherein width $\sim p^{3/2}$, etc. One may also have, e.g., width $\sim[p\Psi_1(0) + p^2\Psi_2(0)]^{1/2}$ over some range, with both terms inside the square root important. Some of the above behaviors are actually observed on certain pressure broadened spectral lines.[7] The behavior of the width here provides clues as to the relative importance of the terms of (6.1).

C. Infinitely divisible cases

As was already mentioned in Sec. 3 A, Λ (or γ or ϵ) is not, in general, a free parameter than can be varied at will, independently of the function $\psi(t)$. In concrete cases, it is often a parameter such as p in (6.1) which can be varied experimentally, and one can see that varying p alters both Λ and ψ. But even in principle, Λ cannot always be varied arbitrarily. For let us consider [taking $\Theta = 1$ in (3.21)]

$$\begin{aligned}
P(\omega) &= (2\pi)^{-1}\int_{-\infty}^{\infty} dt\, e^{-i\omega t + \gamma g(t)} \\
&\equiv P(\omega;\gamma,\hat{\psi}),
\end{aligned} \tag{6.3}$$

$g(t)$ being given by (3.30) in terms of $\hat{\psi}$; the fact that

$$P(\omega;\gamma,\hat{\psi}) \geqslant 0 \tag{6.4}$$

is a probability density constrains $\gamma g(t)$: in particular, the positivity (6.4) is preserved under arbitrary variation of γ, i.e.,

$$P(\omega;\gamma,\hat{\psi})\geqslant 0 \quad \text{for any } \gamma>0,\ \hat{\psi} \text{ fixed} \tag{6.5}$$

if and only if $\hat{\psi}(v)$ is itself a probability density, i.e., iff

$$\hat{\psi}(v)\geqslant 0. \tag{6.6}$$

This is a well known result of probability theory, usually stated as[11]

Theorem: A normalized and centered probability density (p.d.) is *infinitely divisible* (i.d.) if and only if the logarithm of its characteristic function is expressible as (3.30) with (6.6) and $\int_{-\infty}^{\infty} dv\, \hat{\psi}(v)<\infty$.

A p.d. is said to be i.d. if it is the N th convolution power of another p.d., for any N. Clearly, infinite divisibility is equivalent to (6.5): for if (6.5) is satisfied, then for any value of γ and any N, $P(\omega;\gamma,\hat{\psi})=P(\omega;\gamma/N,\hat{\psi})^{*N}$, where $P(\omega;\gamma/N,\hat{\psi})\geqslant 0$ is also a p.d.; conversely, if $P(\omega;\gamma,\hat{\psi})$ is an i.d.p.d., it is expressible as (6.3) and (3.30) with (6.6) and no restriction on γ, i.e., it satisfies (6.5).

A simple way of seeing the necessity of (6.6) [in the case (3.14)] is to note that as $\gamma\to 0$ (ψ fixed), $\hat{P}(v)$ tends rigorously to $\hat{P}_{\mathrm{WB}}^{(0,\infty)}(v)$, Eq. (4.13), and for this to be positive requires (6.6). The sufficiency is proved[11] by expressing the integral (3.30) as a Riemann sum, and then noting that this corresponds to a multiple convolution of Poisson and normal p.d.'s, both of which are known to be i.d. (by direct verification).[57] This also shows that the Poisson and normal laws are the basic elements comprising every i.d.p.d.

A simple i.d. example is provided by model (2.1) with a Gaussian $U(t)$, for which case $\hat{\psi}(v)$ is the (scaled) power spectrum of $U(t)$, and evidently positive [here $\Lambda=\langle U^2\rangle^{1/2}\theta$, where θ is the correlation time of $U(t)$]. Another well known i.d. example is the Anderson–Talman–Baranger model of pressure broadening[5]: there, γ is essentially the gas pressure and $\hat{\psi}(v)$ the power spectrum of $\dot{d}_1(t)$, where $d_1(t)$ is the dipole moment of the radiator in the presence of a single perturber.[7,33]

Also, case (6.1) with p freely variable, experimentally or theoretically, is approximately i.d. (rigorously so as $p\to 0$) over the range in p wherein only $p\Psi_1$ is sizable in (6.1b); this implies

$$\hat{\Psi}_1(v)\geqslant 0 \tag{6.7}$$

[alternatively, as $p\to 0$, $\hat{P}(v)\to\hat{P}_{\mathrm{WB}}^{(0,\infty)}(v)$, Eq. (4.13), with $\hat{\psi}$ replaced by $\hat{\Psi}_1(v)/\Psi_1(0)$, again implying (6.7) since $\hat{P}(v)\geqslant 0$].

In i.d. (or approximately i.d.) cases, $\hat{\psi}(v)$ [or $\hat{\Psi}_1(v)$] can be determined directly by observing $\hat{P}(v)$ at small γ (or p) and using (4.13); one may therefrom deduce $\hat{P}(v)$ at any value of γ [or of p over the approximately i.d. range].

D. Intermediate values of Λ

Outside the regimes $\Lambda\ll 1$ and $\Lambda\gg 1$, $P(\omega)$ can have a fairly complicated shape. The following expansion, obtained by expanding $e^{\gamma h}$ in (4.3), may be useful for analyzing the structure of $\hat{P}(v)$, and visualizing its shape in terms of that of $\hat{\psi}(v)$:

$$\hat{P}(v)=e^{-\gamma c}\hat{P}_{as}(./\gamma b)^{*}[\delta+\gamma\hat{h}+\tfrac{1}{2}\gamma^2\hat{h}^{*2}+\cdots](\gamma b v), \tag{6.8}$$

where[58]

$$\hat{P}_{as}(v)=\pi^{-1}\mathrm{Re}\int_0^{\infty} dt\, e^{-ivt+i\gamma a-t}=\pi^{-1}\frac{\cos(\gamma a)+v\sin(\gamma a)}{1+v^2} \tag{6.9}$$

and $\hat{h}(v)$ has the following expression in terms of $\hat{\psi}(v)$:

$$\hat{h}(v)=v^{-2}[\hat{\psi}(v)-\hat{\psi}(0)-v\hat{\psi}'(0)] \tag{6.10}$$

[$f(./\gamma b)$ is the function whose value at v is $f(v/\gamma b)$; δ is the Dirac function]. Note the Poisson factor $e^{-\gamma c}\gamma^k/k!$ weighting each convolution power $\hat{h}^{*k}$. In i.d. cases, (6.8) allows one to mentally visualize the evolution of $\hat{P}(v)$ as γ varies. The above expansion has a nice physical meaning in pressure broadening, besides its mathematical and heuristic utility.[7,33]

E. Finite $\langle\mu\rangle$

We assumed the first moment $\langle\mu\rangle=\int\omega P(\omega)d\omega=0$ [Eq. (2.10)]. When this is not the case, $P(\omega)$ is simply shifted by $\langle\mu\rangle$. This mean shift may sometimes be important; e.g., if in model (2.1) U is small (weak broadening), then $\langle\mu\rangle=\langle U\rangle$, being of first order in U, may be much larger than the terms displayed in expansions (4.15), which are of second or higher order in U (since Ψ is); in particular, the shift $\langle\mu\rangle$ may be much more important than the width, as is sometimes observed.[59] In case (6.1), $i\langle\mu\rangle=p\dot{G}_1(0)+p^2\dot{G}_2(0)+\cdots$ is itself an expansion in p.

F. Remark

A welcome feature of the shape parameter expansions is that some of their coefficients are interrelated, as is apparent on inspection of (4.15) and (5.21). This can be very useful for analyzing and understanding experimental data.[7]

7. ILLUSTRATION: N-BODY

To illustrate the results of the preceding sections, we consider the case that $P(\omega)$ depends on N "things", e.g., particles, in a symmetric manner. We will first assume, within model (2.1), that

$$U(t)=u_1(t)+u_2(t)+\cdots+u_N(t), \tag{7.1}$$

where the $u_i(t)$ are identical mutually independent random processes. We will then progressively complicate by relaxing, first, independence, then the strict additivity of U in the u_i, and finally consider the general case, transcending model (2.1), wherein $U(t)$ need not be defined.

A. Independent additive case

We suppose $U(t)$ given by (7.1) with the $u_i(t)$ identical and mutually independent. We let P_1, C_1, G_1, Λ_1, etc., denote the same objects as P, C, G, Λ, etc., but with $U(t)$ replaced by $u(t)$, the subscript 1 indicating that a single "particle" is involved. For instance

$$C_1(\tau)=\langle\exp[i\textstyle\int_0^{\tau}dt\, u(t)]\rangle \tag{7.2}$$

(because the u_i are all identical, the subscript on u may be omitted when a single u_i appears inside $\langle\ \rangle$). The additivity

(7.1) and independence imply the following relations between the N-particle and single particle quantities:

$$C(\tau) = [C_1(\tau)]^N, \quad G(\tau) = NG_1(\tau),$$
$$P(\omega) = P_1^{*N}(\omega), \quad \Psi(\tau) = N\Psi_1(\tau), \tag{7.3}$$

where P_1^{*N} is the N th convolution power of P_1.[60] There follows that $\Theta = \Theta_1$, $\Omega = N^{1/2}\Omega_1$ [see (3.15) and (3.16)], whence the relation

$$\Lambda = N^{1/2}\Lambda_1 \tag{7.4}$$

connecting the "broadening strengths" of $U(t)$ and $u(t)$.

Usually, Λ_1 is a given fixed quantity; (7.4) thus implies that the weak broadening expansions (4.10) and (4.15) are here in powers of N, and the strong broadening expansions (5.12) and (5.20) in powers of $N^{-1/2}$.

B. Dependence

Let again $U(t)$ be given by (7.1) where the $u_i(t)$ are identical, but not independent, as expressed by the nonvanishing of mixed cumulants. For simplicity, we assume symmetry (or uniformity) in the u_i, in the sense that a moment or cumulant involving $u_{i_1}, u_{i_2}, \ldots, u_{i_k}$ is independent of the particular set of k indices $i_1, i_2, \ldots, i_k$; e.g.,

$$\langle u_{i_1}(t_1)u_{i_1}(t_1')u_{i_2}(t_2)u_{i_3}(t_3)\rangle_c = \langle u_1(t_1)u_1(t_1')u_2(t_2)u_3(t_3)\rangle_c. \tag{7.5}$$

We then have, from (3.9) and (7.1) [notation: $\phi_j = \int_0^\tau dt\, u_j(t)$],

$$\begin{aligned} G(\tau) &= \Big\langle \prod_{j=1}^N e^{i\phi_j} - 1 \Big\rangle_c \\ &= \Big\langle \prod_{j=1}^N (1+f_j) - 1 \Big\rangle_c \\ &= N\langle f_1\rangle_c + \tfrac{1}{2}N(N-1)\langle f_1 f_2\rangle_c + (1/3!) \\ &\quad \times N(N-1)(N-2)\langle f_1 f_2 f_3\rangle_c + \cdots, \end{aligned} \tag{7.6}$$

where we denoted $f_j = e^{\phi_j} - 1$ (Mayer trick). In the independent case, all the mixed cumulants vanish, so only the first term of (7.6) survives, yielding (7.3). In the dependent case, the important new feature is that $G(\tau)$ is now an expansion in powers of N, which brings us to the situation envisaged in Sec. 6 B, with N playing the role of p.

C. Nonadditivity

Now let $U(t)$ be only approximately additive in the $u_i(t)$, i.e.,

$$U(t) = \sum_{i=1}^N u_i(t) + \sum_{i<j} u_{ij}(t) + \sum_{i<j<k} u_{ijk}(t) + \cdots, \tag{7.7}$$

where the $u_{i_1 i_2 \cdots i_k}(t)$ are nonadditive corrections. We might here proceed similarly to (7.6); but let us rather directly pass on to a more general case, covering (7.7) as well as (7.1).

D. General case

We now no longer limit ourselves to model (2.1). We assume that $P(\omega)$ depends on N "particles" in a symmetric manner. We denote $P_{i_1 i_2 \cdots i_k}$, $C_{i_1 i_2 \cdots i_k}$, etc. quantities corresponding to the presence of only particles $i_1, i_2, \ldots, i_k$. The symmetry assumed means that $C_1 = C_2 = \cdots = C_N$, $C_{i_1 i_2} = C_{12}$, etc. We suppose that to a first approximation

$$C(\tau) \cong \prod_{i=1}^N C_i(\tau) = [C_1(\tau)]^N,$$
$$G(\tau) \cong \sum_{i=1}^N G_i(\tau) = NG_1(\tau). \tag{7.8}$$

This may be based on physical intuition, or perhaps motivated by a model of the type discussed in Sec. 7 A approximating the physical system considered.

We now seek corrections to (7.8) in the form of an expansion in powers of N for $G(\tau)$. This is constructed as follows. We first write

$$G_{12} = G_1 + G_2 + K_{12}, \tag{7.9a}$$

where K_{12} is the correction to the additive approximation (7.8). We next set

$$G_{123} = G_1 + G_2 + G_3 + K_{12} + K_{13} + K_{23} + K_{123}, \tag{7.9b}$$

which expresses the correction to (7.8) as a sum of pairwise corrections plus a final triplewise correction. Continuing in this manner, we ultimately obtain

$$\begin{aligned} G(\tau) &= G_{12\cdots N} \\ &= \sum_i G_i + \sum_{i<j} K_{ij} + \sum_{i<j<k} K_{ijk} + \cdots + K_{12\cdots N} \\ &= NG_1 + \tfrac{1}{2}N(N-1)K_{12} + (1/3!) \\ &\quad \times N(N-1)(N-2)K_{123} + \cdots + K_{12\cdots N}. \end{aligned} \tag{7.10}$$

The K's are obtained by inverting (7.9):

$$K_{12} = G_{12} - G_1 - G_2,$$
$$K_{123} = G_{123} - G_{12} - G_{13} - G_{23} + G_1 + G_2 + G_3, \tag{7.11}$$

The $K_{ij\ldots}$ may be shown to be a sort of "additive cumulants", adapted to the case that we have additivity rather than factorization as some "independence" condition. If a bulk limit, $N \to \infty$, Volume $\to \infty$, is taken, the K's become expressible as ordinary cumulants.[7]

In (7.10), $G(\tau)$ appears as an expansion in powers of N, so the situation is that discussed in Sec. 6B, with N playing the role of p. Pressure broadening provides an interesting example, in that the correlations between, and nonadditivity in, the effects of different perturbing atoms, are of a size such that the various possible behaviors for the width mentioned at the end of Sec. 6B [proportional to $N^{1/2}$, or N, or $(aN + bN^2)^{1/2}$, etc.] are actually observed.[7]

E. Discussion

Gaussians, as appear in the strong broadening limit, naturally evoke the Central Limit Theorem (CLT) of probability theory[10]: *The probability density of a sum of N independent random variables tends to a Gaussian as* $N \to \infty$, *of width* $\sim N^{1/2}$. Now, the decomposition (7.1) implies the like for the (static) random variable μ [see (2.7)]: $\mu = \mu_1 + \mu_2 + \cdots + \mu_N$. It is then easy to see that in the independent additive case discussed in Sec. 7A, the strong broadening Gaussian limit is just a case of the CLT; the corresponding expansion (5.16) of $\tilde{P}(y)$, in powers of $N^{-1/2}$, is well known in probability theory (Edgeworth–Cramer asymptot-

ic expansion[10]). As to the strong broadening Gaussian limits in the nonindependent cases discussed in Sec. 7B–7D, they may be considered generalizations of the Central Limit Theorem. Such generalizations may be of interest in themselves, especially since they effectively materialize in pressure broadening; they will be discussed in more detail in a separate paper.

APPENDIX A

In this appendix, we demonstrate (4.19). We first rewrite (4.18) as

$$S^{x}_{kj} = \gamma^{j+1}\int_0^\infty dt\, e^{-\gamma z t} h(t)^k t^j, \tag{A1}$$

where $z = b(1+ix)$. Since we are only interested in the γ dependence of S_{kj}, we do not keep track of overall multiplicative factors independent of γ.

We will assume various simple behaviors of $h(t)$ for t larger than some time T, which may be taken as 1 for simplicity. We presume that a realistic $h(t)$ behaves asymptotically as a combination of the cases considered below.

Case (i): $h(t>T)=0$, or $h(t)\sim t^{-\infty}$, or $h(t)\sim e^{iWt}t^{-\sigma}$. Here, we can expand S_{kj} in powers of γ, and get

$$S^{x}_{kj} = \gamma^{j+1}\mathrm{Pol}(\gamma), \tag{A2}$$

where $\mathrm{Pol}(\gamma)$ denotes any polynomial or series in non-negative powers of γ [in the case $h(t)\sim e^{iWt}t^{-\sigma}$, we get integrals of the form $\int_0^\infty dt e^{-\lambda t}t^m = m!/\lambda^{m+1}$, where $\lambda = \gamma z + ikW$, which may then be expanded in γ, provided $W\neq 0$]. When (A2) is substituted in (4.17) and (4.27), we recover (4.8) and (4.15) [e.g., to order γ^0, $S^0_{11}=0$, and $S^+_{10}=S^-_{10}$, so the second order term of asym in (4.27) becomes $\frac{1}{2}\gamma^2a^2$, as in (4.15)].

To discuss other cases, we set $\int_0^\infty = \int_0^1 + \int_1^\infty$ in (A1); the term $\int_0^1$ yields (A2), so that we have

$$S^{x}_{kj} = \gamma^{j+1}\mathrm{Pol}(\gamma) + S^{>}_{kj}, \tag{A3}$$

where

$$S^{>}_{kj} = \gamma^{j+1}\int_1^\infty dt\, e^{-\gamma z t} t^j h(t)^k. \tag{A4}$$

Case (ii): $h(t>1) = t^{-\sigma}$. We write (A4) as

$$S^{>}_{kj} = \gamma^{j+1} I_{j-k\sigma}(1,\infty;\gamma z), \tag{A5}$$

where

$$I_m(a,b;\lambda) = \int_a^b dt\, e^{-\lambda t} t^m. \tag{A6}$$

The γ dependence of I_m depends on the value of m: If $m > -1$,

$$\begin{aligned} I_m(1,\infty;\gamma z) &= I_m(0,\infty;\gamma z) - I_m(0,1;\gamma z) \\ &= \gamma^{-m-1} I_m(0,\infty;z) - \mathrm{Pol}(\gamma) \\ &= \gamma^{-m-1} + \mathrm{Pol}(\gamma). \end{aligned} \tag{A7}$$

If $m \leqslant -1$, an integration by parts gives

$$I_m(1,\infty;\gamma z) = [-e^{-\gamma z} + \gamma z I_{m+1}(1,\infty;\gamma z)]/(m+1). \tag{A8}$$

Using this $-(m-1)$ times if m is an integer, $-[m]$ times ($[\,]$ = integral part) if not, we get

$$I_m(1,\infty;\gamma z) = \mathrm{Pol}(\gamma) + \gamma^{-m-\alpha} I_{-\alpha}(1,\infty;\gamma z), \tag{A9}$$

where $\alpha = 1$ if m is an integer, $\alpha = m - [m]$ if not. For m not an integer, we have $I_{-\alpha} = \mathrm{Pol}(\gamma) + \gamma^{\alpha-1}$, i.e.,

$$I_m(1,\infty;\gamma z) = \mathrm{Pol}(\gamma) + \gamma^{-m-1}. \tag{A10}$$

For m an integer, we must evaluate $I_{-\alpha}$ for $\alpha = 1$: an integration by parts gives

$$\begin{aligned} I_{-1}(1,\infty;\gamma z) &= \gamma z\int_1^\infty dt\, e^{-\gamma z t}\ln t \\ &= \gamma z\left[\int_0^\infty - \int_0^1\right] dt\, e^{-\gamma z t}\ln t \\ &= \gamma z\int_0^\infty \gamma^{-1} du\, e^{-zu}(\ln u - \ln\gamma) + \mathrm{Pol}(\gamma) \\ &= (\mathrm{const})\ln\gamma + \mathrm{Pol}(\gamma). \end{aligned} \tag{A11}$$

Collecting results, we have

$$I_m(1,\infty;\gamma z) = \mathrm{Pol}(\gamma) + \gamma^{-m-1}\begin{cases} 1 & \text{if } m > -1 \text{ or not an integer} \\ \ln\gamma & \text{if } m = \text{integer} \leqslant -1 \end{cases}, \tag{A12}$$

whence (4.19), in view of (A3)–(A5) [note that we may add factors such as $\ln(t)$ to the $t^{-\sigma}$ dependence without altering (4.20)].

APPENDIX B

This appendix contains additional details concerning the weak and strong broadening expansions. In particular, we here display the expansions of $v_\pm$ and $y_\pm$, and also higher order terms excluded from the main text to avoid cluttering. Some of the definitions are repeated for easy reference.

1. Weak broadening expansions

[We here denote $\bar{g}(t)\equiv g(t) - (idt - c)$.]

$$\hat{P}(v) = \pi^{-1}e^{-\gamma c}\mathrm{Re}\left\{\int_0^\infty dt\, e^{-ivt + \gamma\bar{g}(t/\gamma b)}\right\}, \tag{B1}$$

where c,b,γ are real constants, and $\bar{g}(t)\sim -bt$ as $t\to\infty$, $b>0$. We set

$$\hat{P}(v) = \pi^{-1}\frac{e^{-\gamma c}}{1+v^2}B(v) \tag{B2}$$

which defines $B(v)$.

a. $\bar{g}(t) = -bt + ia + h(t)$

$$B(v) = \mathrm{Re}\{e^{i\gamma a}[1 - iv + (1+v^2)T(v)]\}, \tag{B3}$$

where

$$T(v) = \sum_{k=1}^{\infty}\sum_{j=0}^{\infty}\gamma^{k+j+1}(v-i)^j T_{kj}/(k!j!), \tag{B4}$$

$$T_{kj} \equiv T^r_{kj} + iT^i_{kj} = b^{j+1}\int_0^\infty dt\,(-it)^j h(t)^k. \tag{B5}$$

The following expansions are valid up to terms containing divergent T_{kj}'s.

Expansion (4.10) *(to order* γ^3*).*

$$\begin{aligned} B(v) = {} & 1 + \gamma v a + \gamma^2(T^r_{10} - \tfrac{1}{2}a^2) + \gamma^2 v^2 T^r_{10} + \gamma^3(-aT^i_{10} + \tfrac{1}{2}T^r_{20} + T^i_{11}) \\ & + \gamma^3 v(-\tfrac{1}{6}a^3 + T^r_{11}) + \gamma^3 v^2(-aT^i_{10} + \tfrac{1}{2}T^r_{20} + T^i_{11}) + \gamma^3 v^3 T^r_{11} + \cdots. \end{aligned} \tag{B6}$$

Expansions (4.15) *($v_\pm$ and higher order terms).*

$$\begin{aligned} \mathrm{shift}_v &= \ldots + \gamma^4(-\tfrac{1}{2}aT^i_{11} + \tfrac{1}{4}T^i_{12} + \tfrac{1}{4}T^r_{21}) + (\gamma^5/240)(a^5 + 180aT^r_{12} + 60aT^i_{21} - 60T^r_{13} + 60T^i_{22} + 20T^r_{31}) + \cdots, \\ v_\pm &= \pm 1 + \gamma a \pm \gamma^2(\tfrac{1}{4}a^2 + T^r_{10}) + (\gamma^3/12)(a^3 + 12aT^r_{10} \mp 12aT^i_{10} \pm 12T^i_{11} \pm 6T^r_{20} + 24T^r_{11}) + \cdots, \\ \mathrm{width}_v &= \ldots + (\gamma^4/48)(a^4 + 24a^2T^r_{10} + 432aT^r_{11} - 96aT^i_{20} + 48T^{r2}_{10} + 48T^r_{12} - 48T^i_{21} + 16T^r_{30}) + \cdots, \\ \mathrm{asym}_v &= \ldots - \frac{\gamma^4}{24}(a^4 - 24a^2T^r_{10} - 24a^2T^i_{10} - 48aT^i_{11} + 12aT^r_{20} - 72aT^r_{11} + 72T^i_{12} + 36T^r_{21}) + \cdots. \end{aligned} \tag{B7}$$

Inversion *I*: $a \to -a$, $T_{kj} \to (-)^j T^*_{kj}$. (B8)

An alternative procedure. In lieu of (B3), we may use

$$B(v) = \cos(\gamma a) + v\sin(\gamma a) + (1+v^2)D(v), \tag{B9}$$

where

$$\begin{aligned} D(v) &= \mathrm{Re}\Big\{\int_0^\infty dt\, e^{-ivt}[e^{\gamma\bar{g}(t/\gamma b)} - e^{i\gamma a - t}]\Big\} \\ &= \sum_{k=1}^{\infty}\sum_{j=0}^{\infty}\gamma^{k+j+1}v^j D_{kj}, \end{aligned} \tag{B10}$$

$$D_{kj} = b^{j+1}(k!j!)^{-1}\mathrm{Re}\left\{\int_0^\infty dt\,(-it)^j[\bar{g}(t)^k - (ia - bt)^k]\right\} \tag{B11}$$

(note that $D_{1j} = T^r_{1j}/j!$). the expansion of $b(v)$, immediate from (B9)–(B10), and those of the shape parameters in terms of the D_{kj}'s, are somewhat simpler than (B6) – (B7):

$$\begin{aligned} \mathrm{shift}_v &= \tfrac{1}{2}\gamma a + (\gamma^3/24)(a^3 + 12D_{11}) + \tfrac{1}{2}\gamma^4 D_{21} + (\gamma^5/240)(a^5 + 60a^2D_{11} + 120aD_{12} + 120D_{31}) + \cdots, \\ v_\pm &= \pm 1 + \gamma a \pm \gamma^2(\tfrac{1}{4}a^2 + D_{10}) + (\gamma^3/12)(a^3 + 12aD_{10} + 24D_{11} \pm 12D_{20}) + \cdots, \\ \mathrm{width}_v &= 2 + \tfrac{1}{2}\gamma^2(a^2 + 4D_{10}) + 2\gamma^3 D_{20} + (\gamma^4/48)(a^4 + 72a^2D_{10} + 336aD_{11} + 192D_{12} + 48D^2_{10} + 96D_{30}) + \cdots, \\ \mathrm{asym}_v &= -\gamma a + \tfrac{1}{2}\gamma^2 a^2 - (\gamma^3/12)(a^3 + 12aD_{10} + 36D_{11}) - (\gamma^4/24)(a^4 - 24a^2D_{10} - 72aD_{11} + 24aD_{20} + 72D_{21}) + \cdots, \end{aligned} \tag{B12}$$

Inversion *I*: $a \to -a$, $D_{kj} \to (-)^j D_{kj}$. (B13)

In case the T_{kj} *do not all exist.*

$$T(v) = \sum_{k=1}^{\infty}\sum_{j=0}^{\infty}\gamma^{k+1}v^j\Gamma^0_{kj}/(k!j!), \tag{B14}$$

$$\Gamma^x_{kj} = \gamma^j b^{j+1}\int_0^\infty dt\, e^{-\gamma bt(1+ix)}(-it)^j h(t)^k. \tag{B15}$$

Expansions (4.27) *($v_\pm$ and higher order terms).*

$$\begin{aligned} v_\pm &= \gamma a + \gamma^2(\pm\tfrac{1}{4}a^2 \mp \Gamma^{0r}_{10} \pm 2\Gamma^{\pm r}_{10}) + (\gamma^3/12)(a^3 - 12a\Gamma^{0r}_{10} \mp 6\Gamma^{0r}_{20} \mp 6a\Gamma^{0r}_{11} \pm 12a\Gamma^{0i}_{10} \\ &\quad + 24a\Gamma^{\pm r}_{10} \pm 12\Gamma^{\pm r}_{20} \pm 24a\Gamma^{\pm r}_{11} \mp 24a\Gamma^{\pm i}_{10}) + \cdots, \\ \mathrm{width}_v &= \cdots - \gamma^3[\Gamma^{0r}_{20} + a\Gamma^{0r}_{11} - 2a\Gamma^{0i}_{10} - 2a(\Gamma^{+r}_{10} - \Gamma^{-r}_{10}) - (\Gamma^{+r}_{20} + \Gamma^{-r}_{20}) \\ &\quad - 2a(\Gamma^{+r}_{11} + \Gamma^{-r}_{11}) + 2a(\Gamma^{+i}_{10} + \Gamma^{-i}_{10})] + \cdots, \\ \mathrm{asym}_v &= \cdots + (\gamma^3/12)(-a^3 + 12a\Gamma^{0r}_{10} - 12a\Gamma^{0r}_{11} + 6\Gamma^{0r}_{21} + 6a\Gamma^{0r}_{12} - 12a\Gamma^{0i}_{11} + 12a\Gamma^{+r}_{10} - 12\Gamma^{+r}_{20} - 24a\Gamma^{+r}_{11} \\ &\quad + 24a\Gamma^{+i}_{10} - 36a\Gamma^{-r}_{10} + 12\Gamma^{-r}_{20} + 24a\Gamma^{-r}_{11} - 24a\Gamma^{-i}_{10}) + \cdots, \end{aligned} \tag{B16}$$

Inversion I: $a \to -a$, $\Gamma^x_{kj} \to (-)^j \Gamma^{-x*}_{kj}$. (B17)

b. $\bar{g}(t) = -bt + g^{\#}(t)$, $g^{\#}(t) \sim t^{-\sigma}$ *as* $t \to \infty$, $-1 < \sigma < \infty$ *(here* $c = 0$*)*

$$\hat{P}(v) = \pi^{-1}\mathrm{Re}\left\{\int_0^\infty dt \exp[-ivt - t + \gamma g^{\#}(t/\gamma b)]\right\}. \tag{B18}$$

We define

$$\rho = \gamma, \quad K^x_{kj} = \mathrm{Re} S^{x\#}_{kj} \quad \text{if } \sigma \geqslant 0,$$
$$\rho = \gamma^{1+\sigma}, \quad K^x_{kj} = \gamma^{-k\sigma} \mathrm{Re}\, S^{x\#}_{kj} \quad \text{if } -1 < \sigma < 0, \tag{B19}$$

where

$$S^{x\#}_{kj} = (\gamma b)^{j+1} \int_0^\infty dt\, e^{-\gamma b t(1+ix)} (-it)^j g^{\#}(t)^k. \tag{B20}$$

Expansions (4.25) [notation: $K^{\pm}_{kj} \equiv K^{\pm 1}_{kj}$, $K^{02}_{kj} \equiv (K^0_{kj})^2$, $K^{\pm 2}_{kj} \equiv (K^{\pm}_{kj})^2$, etc.].

$$\begin{aligned}
\mathrm{shift}_v &= \ldots + (\rho^3/8)(2K^{03}_{11} + \tfrac{1}{2}K^{02}_{11}K^0_{13} + K^0_{11}K^{02}_{12} + K^0_{11}K^{02}_{22} + K^0_{21}K^0_{12} + \tfrac{2}{3}K^0_{31}) + \cdots, \\
v_{\pm} &= \rho(\mp K^0_{10} \pm 2K^{\pm}_{10}) + \tfrac{1}{4}\rho^2(\pm 2K^{02}_{10} \mp 8K^0_{10}K^{\pm}_{10} - 8K^0_{10}K^{\pm}_{11} \mp 2K^0_{20} \mp K^{02}_{11} \pm 8K^{\pm 2}_{10} + 16K^{\pm}_{10}K^{\pm}_{11} \pm 4K^{\pm}_{20}) + \cdots, \\
\mathrm{width}_v &= \ldots + \tfrac{1}{2}\rho^2[2K^{02}_{10} - 4K^0_{10}(K^+_{10} + K^-_{10}) - 4K^0_{10}(K^+_{11} - K^-_{11}) - 2K^0_{20} - K^{02}_{11} \\
&\quad + 4(K^{+2}_{10} + K^{-2}_{10}) + 8(K^+_{10}K^+_{11} - K^-_{10}K^-_{11}) + 2(K^+_{20} + K^-_{20})] + \cdots,
\end{aligned} \tag{B21}$$

$$\begin{aligned}
\mathrm{asym}_v = \ldots &+ \tfrac{1}{2}\rho^2(2K^0_{10}K^0_{11} + 4K^0_{10}K^+_{11} + 4K^0_{10}K^-_{11} + K^{02}_{11} + K^0_{11}K^0_{12} - 6K^0_{11}K^+_{10} + 2K^0_{11}K^-_{10} \\
&+ K^0_{21} + 4K^{+2}_{10} - 8K^+_{10}K^+_{11} - 8K^+_{10}K^-_{10} - 2K^+_{20} + 4K^{-2}_{10} - 8K^-_{10}K^-_{11} + 2K^-_{20}) + \cdots,
\end{aligned}$$

Inversion I: $K^x_{kj} \to (-)^j K^{-x}_{kj}$. (B22)

2. Strong broadening expansions

$$\tilde{P}(y) = \pi^{-1}\mathrm{Re}\left\{\int_0^\infty dt\, e^{-iyt - (1/2)t^2} \exp\left[\sum_{k=1}^\infty \epsilon^k \Gamma_k (it)^{k+2}/(k+2)!\right]\right\}. \tag{B23}$$

If the Γ_k are real:

Expansion (5.16) *(to order* ϵ^2*).*

$$\tilde{P}(y) = (2\pi)^{-1/2} e^{(-1/2)y^2}\{1 + \epsilon(\Gamma_1/3!)(y^3 - 3y) + \epsilon^2[(\Gamma_2/4!)(3 - 6y^2 + y^4) + (\Gamma_1/3!)^2(-15 + 45y^2 - 15y^4 + y^6)] + \cdots\}. \tag{B24}$$

Expansions (5.21) ($y_{\pm}$ and higher order terms).

$$\begin{aligned}
y_{\pm} &= \pm\kappa + \epsilon\Gamma_1(\tfrac{1}{6}\kappa^2 - \tfrac{1}{2}) \pm \tfrac{1}{2}\epsilon^2[\Gamma_1^2(-\tfrac{1}{9}\kappa^3 + \tfrac{1}{2}\kappa) + \Gamma_2(\tfrac{1}{12}\kappa^3 - \tfrac{1}{2}\kappa)] \\
&\quad + (\epsilon^3/3)[\Gamma_1^3(\tfrac{1}{9}\kappa^4 - \tfrac{3}{4}\kappa^2 + \tfrac{3}{4}) + \Gamma_1\Gamma_2(-\tfrac{1}{8}\kappa^4 + \kappa^2 - \tfrac{5}{4}) + \Gamma_3(\tfrac{1}{40}\kappa^4 - \tfrac{1}{4}\kappa^2 + \tfrac{3}{8})] + \cdots, \\
\mathrm{asym}_y &= \ldots - \epsilon^3[\Gamma_1^3(\tfrac{11}{108}\kappa^3 - \tfrac{7}{12}\kappa) + \Gamma_1\Gamma_2(-\tfrac{7}{72}\kappa^3 + \tfrac{3}{4}\kappa) + \Gamma_3(\tfrac{1}{60}\kappa^3 - \tfrac{1}{6}\kappa)] + \cdots,
\end{aligned} \tag{B25}$$

where

$$\kappa = (2\ln 2)^{1/2} = 1.177410\cdots, \tag{B26}$$

Inversion I: $\epsilon \to -\epsilon$. (B27)

In case the $\Gamma_k \equiv \Gamma^r_k + i\Gamma^i_k$ are complex:

Expansions (5.20).

$$y_{\pm} = \pm\kappa + \epsilon[\Gamma^r_1(\tfrac{1}{6}\kappa^2 - \tfrac{1}{2}) \pm \Gamma^i_1(\tfrac{1}{3}ps\kappa^2 - ps - \tfrac{1}{3}p\kappa + \tfrac{1}{3}p)] + \epsilon^2(28 \text{ terms}) + \cdots, \tag{B28}$$

where

$$p = (2/\pi)^{1/2}, \quad s = 2^{1/2}D(\kappa/2^{1/2}) = 2^{1/2}(0.536196), \quad D(x) = e^{-x^2}\int_0^x dt\, e^{t^2}, \tag{B29}$$

$$\begin{aligned}
\mathrm{shift}_y &= \ldots - \epsilon^3(-\tfrac{1}{4}\Gamma^{r3}_1 + \tfrac{5}{12}\Gamma^r_1\Gamma^r_2 + 0.892449\Gamma^r_1\Gamma^{i2}_1 - \tfrac{1}{8}\Gamma^r_3 - 0.446224\Gamma^i_1\Gamma^i_2) + \cdots, \\
\mathrm{width}_y &= \ldots - \epsilon^3(-1.23104\Gamma^{r2}_1\Gamma^i_1 + 0.636072\Gamma^r_1\Gamma^i_2 + 0.758722\Gamma^r_2\Gamma^i_1 + 0.530579\Gamma^{i3}_1 - 0.209756\Gamma^i_3) + \cdots, \\
\mathrm{asym}_y &= \ldots - \epsilon^3(-0.520574\Gamma^{r3}_1 - 0.314070\Gamma^{r2}_1\Gamma^i_1 + 0.724368\Gamma^r_1\Gamma^r_2 + 1.81145\Gamma^r_1\Gamma^{i2}_1 \\
&\quad + 0.114072\Gamma^r_1\Gamma^i_2 - 0.169031\Gamma^r_3 - 0.767361\Gamma^i_1\Gamma^i_2) + \cdots,
\end{aligned} \tag{B30}$$

Inversion I: $\epsilon \to -\epsilon$, $\Gamma_k \to \Gamma^*_k$. (B31)

[1]R. Kubo and K. Tomita, J. Phys. Soc. Jpn. **9**, 888 (1954).

[2]R. Kubo, in *Fluctuations, relaxation and resonance in magnetic systems*, edited by D. ter Haar (Oliver and Boyd, Edinburgh, 1962).

[3]R. Kubo, J. Math. Phys. **4**, 174 (1963).

[4](a)P. W. Anderson and P. R. Weiss, Rev. Mod. Phys. **25**, 269 (1953); (b) P. W. Anderson, J. Phys. Soc. Jpn. **9**, 316 (1954).

[5]P. W. Anderson, Phys. Rev. **86**, 809 (1952); P. W. Anderson and J. D. Talman, Bell Syst. Tech. Publ. No. 3117; M. Baranger, Phys. Rev. **111**, 481 (1958).

[6]M. Tokuyama and H. Mori, Prog. Theor. Phys. **55**, 411 (1976).

[7]A. Royer. Phys. Rev. A **22**, 1625 (1980).

[8]Appendix III in Ref. 4(b). For related work in particle physics, see D. B. Lichtenberg, Phys. Rev. D **10**, 3865 (1974).

[9]See, e.g., Figs. 7,8,14–16 in Ref. 7.

[10]E.g., H. Cramer, *Random Variables and Probability Distributions*, 3rd ed. (Cambridge U. P., Cambridge, England, 1970); R. N. Bhattacharya and R. Ranga Rao, *Normal Approximation and Asymptotic Expansions* (Wiley, New York, 1976).

[11]E.g., B. Gnedenko, *The Theory of Probability* (MIR, Moscow, 1976).

[12]M. J. Lighthill, *Introduction to Fourier Analysis and Generalized Functions* (Cambridge U.P., Cambridge, England, 1964).

[13]E.g., M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. **17**, 323 (1945), reprinted in *Noise and Stochastic Processes*, edited by N. Wax (Dover, New York, 1954); S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, 2nd ed. (Academic, New York, 1975); P. Resibois and M. De-Leener, *Classical Kinetic Theory of Fluids*, Part A (Wiley, New York, 1977).

[14]Let us recall a few well-known results concerning Fourier transforms and dispersion relations: given

$$\hat{f}(\omega)=(2\pi)^{-1}\int_{-\infty}^{\infty}dt\,e^{-i\omega t}f(t),\quad \hat{f}_+(\omega)=\int_0^{\infty}dt\,e^{-i\omega t}f(t),$$

we have the dispersion relation

$$\hat{f}_+(\omega)=-i\int_{-\infty}^{\infty}d\omega'\hat{f}(\omega')/(\omega-\omega'-i0)$$
$$=-i\int_{-\infty}^{\infty}d\omega'\hat{f}(\omega-\omega')/(\omega'-i0),$$

where

$$(\omega-i0)^{-1}=\omega^{-1}+i\pi\delta(\omega).$$

If $f(-t)=f(t)^*$ so that $\hat{f}(\omega)$ is real, we get

$$\mathrm{Re}\hat{f}_+(\omega)=\pi\hat{f}(\omega),\quad \mathrm{Im}\hat{f}_+(\omega)=-\int_{-\infty}^{\infty}d\omega'\hat{f}(\omega')/(\omega-\omega').$$

Note that ($\hat{f}'\equiv d\hat{f}/d\omega$)

$$\hat{f}'_+(\omega)=-i\int_0^{\infty}tdt\,e^{-i\omega t}f(t)$$
$$=i\int_{-\infty}^{\infty}d\omega'\hat{f}(\omega')/(\omega-\omega'-i0)^2=-i\int_{-\infty}^{\infty}d\omega'\hat{f}'(\omega-\omega')/(\omega'-i0),$$

where $(\omega-i0)^{-2}=-(d/d\omega)(\omega-i0)^{-1}$ is essentially defined by the last equality.

[15]Also, relations (2.26) will provide simple ways of testing the correctness of the various expansions we shall obtain for $P(\omega)$ and its shape parameters. Such checks are most welcome, in view of the tediousness of the algebra involved (mostly performed by computer).

[16]R. Kubo, J. Phys. Soc. Jpn. **17**, 1100 (1962); A. Royer, Phys. Rev. A **6**, 1741 (1972), Appendix A.

[17]Cumulants are defined, in general, by posing $\ln\langle\exp(\Sigma_i\lambda_iU_i)\rangle$ $=\langle\exp(\Sigma_i\lambda_iU_i)-1\rangle_c$ (λ_i arbitrary constants), expanding both sides and equating corresponding terms. If we let the variables U_i be independent of the variables V_j, so that $\langle\exp(\Sigma U+\Sigma V)\rangle=\langle\exp(\Sigma U)\rangle\langle\exp(\Sigma V)\rangle$, we get $\ln\langle\exp(\Sigma U+\Sigma V)\rangle=\ln\langle\exp(\Sigma U)\rangle+\ln\langle\exp(\Sigma V)\rangle$, whence $\langle\exp(\Sigma U+\Sigma V)-1\rangle_c=\langle\exp(\Sigma U)-1\rangle_c+\langle\exp(\Sigma V)-1\rangle_c$, showing that all cumulants mixing U's and V's vanish (cluster property).

[18]This is obtained by using time-translation invariance inside $\langle\ \rangle_c$, which follows from Eq. (2.3). We have [denoting $\phi(a,b)=\int_a^b dtU(t)$]: $\dot{G}(\tau)=i\langle e^{i\phi(0,\tau)}U(\tau)\rangle_c=i\langle e^{i\phi(-\tau,0)}U(0)\rangle_c$, whence $\ddot{G}(\tau)=-\langle U(-\tau)e^{i\phi(-\tau,0)}U(0)\rangle_c=-\langle U(0)e^{i\phi(0,\tau)}U(\tau)\rangle_c$.

[19]A stochastic process $U(t)$ is called Gaussian if all the cumulants $\langle U(t_1)U(t_2)\cdots U(t_k)\rangle_c$ vanish for $k\geq 3$; see, e.g., Ref. 13.

[20]As a counterexample, suppose that $U(t)$ can assume values in the vicinities of two well-separated frequencies ν_1 and ν_2, without transiting between the two, i.e., if $U(t)$ is near ν_i initially, it stays so forever; $P(\omega)$ then consists of two "noninteracting" lines. Let us denote $\nu_i+\Delta U_i(t)$ histories in the vicinity of ν_i, call $\langle\ \rangle_i$ the average over such histories, and assume $\langle\Delta U_i\rangle_i=0$. We then have $\langle U\rangle=\frac{1}{2}\Sigma_{i=1}^{2}\langle U\rangle_i$ $=\frac{1}{2}\Sigma\langle\nu_i+\Delta U_i\rangle_i=\frac{1}{2}(\nu_1+\nu_2)=0$ if we assume $\nu_2=-\nu_1$. The autocorrelation $\langle U(0)U(t)\rangle=\frac{1}{2}\Sigma\langle U(0)U(t)\rangle_i$ $=\frac{1}{2}\Sigma\langle[\nu_i+\Delta U_i(0)][\nu_i+\Delta U_i(t)]\rangle_i$ $=\frac{1}{2}\Sigma[\nu_i^2+\langle\Delta U_i(0)\Delta U_i(t)\rangle_i]\to\frac{1}{2}(\nu_1^2+\nu_2^2)\neq 0$ as $t\to\infty$, i.e., the correlation time is infinite.

[21]This means that $x(t)$ in (2.13) must be suitably chosen, e.g., as a dipole operator sandwiched between projection operators eliminating undesired frequency components (see, e.g., Ref. 7).

[22]If we had, e.g., $\Psi(\tau)\sim$ constant as $\tau\to\infty$, then $G(\tau)\sim\tau^2$ as $\tau\to\infty$, and the weak broadening limit shape would be Gaussian. Or, if $\Psi(\tau)\sim\tau^{-\epsilon}$, $0<\epsilon<1$, then $G(\tau)\sim\tau^{2-\epsilon}$ and the weak limit profile would be $\int_{-\infty}^{\infty}dt\exp[-i\omega t+(\mathrm{const})t^{2-\epsilon}]$. We do not consider such possibilities.

[23]Other possible definitions are (provided the integrals exist): (i) $\Theta_1=\int_0^{\infty}\tau d\tau|\Psi(\tau)|^m/\int_0^{\infty}d\tau|\Psi(\tau)|^m$; (ii) $\Theta_2=|\int_0^{\infty}\tau d\tau\,\Psi(\tau)|/|\int_0^{\infty}d\tau\,\Psi(\tau)|$, in which case $|\alpha|/|\beta|=1$ [α,β defined in (3.24) and (3.28)]; (iii) $\Theta_3=\mathrm{Re}\int_0^{\infty}d\tau\,\Psi(\tau)/\Psi(0)$, in which case $b=1$ [see Eq. (3.24); if $\Theta_3=0$, $b=0$ and the weak broadening limit as discussed in this paper does not exist]; (iv) $\Theta^{-1}=\int_{-\infty}^{\infty}d\omega|\omega\hat{\Psi}(\omega)|/\Psi(0)$, etc. The optimal choice should make most constants ($\alpha,\beta,T_{kj},\Gamma_k,\ldots$) of order 1; e.g., Θ_3 is sometimes used (e.g., Ref. 3); however, if $\Psi(\tau)$ is strongly oscillatory, Θ_3 will be much smaller than the decay time of $\Psi(\tau)$, and some of the T_{kj}, Eq. (4.9), may then be inordinately large. [To see this, denote $H(\tau)=\int_0^{\infty}tdt\,\Psi(t+\tau)/\Psi(0)$; then, in view of (3.18), (3.29), and (4.9), $h(t)=\Theta^{-2}H(t\Theta)$, $T_{kj}=b^{j+1}\Theta^{-2k-j-1}\int_0^{\infty}\tau^j d\tau\,H(\tau)^k$. Suppose $\Psi(\tau)$ oscillates with a period Θ_v and vanishes beyond $\tau=\Theta_d$, with $\Theta_d\gg\Theta_v$. Then $H(\tau)$ is of order Θ_v^2 in amplitude and Θ_d in decay time. If $\Psi(\tau)$, thence $H(\tau)$, are real, $H(\tau)^k$ does not oscillate for k even, so that $\int_0^{\infty}\tau^j d\tau\,H(\tau)^k\sim\Theta_v^{2k}\Theta_d^{j+1}$; thus, if $\Theta=\Theta_3\simeq\Theta_v$ ($\Rightarrow b=1$), $T_{kj}\sim(\Theta_d/\Theta_v)^{j+1}\gg 1$].

[24]That is, $\mathrm{Max}|d\Psi/d\tau|=\Psi(0)/\Theta_v$, so that $\Psi(\tau)$ varies by roughly $\Psi(0)$ during an interval of order Θ_v. It might be more proper to use $\mathrm{Max}|\Psi(\tau)|$ instead of $\Psi(0)$ in (3.15)–(3.18); however, this should make little difference for we generally expect $|\Psi(\tau)|\lesssim|\langle U(0)U(\tau)\rangle|\lesssim\Psi(0)$, because of the exponential in (3.11).

[25]This is because we expect the exponential in (3.11) to cause $\Psi(\tau)$ to vary and decay more rapidly than $\langle U(0)U(\tau)\rangle$.

[26]That is, $\psi(0)=1$; decay time $=1$ or $\mathrm{Max}|\dot{\psi}(t)|=1$, according to whether (3.16d) or (3.16v) is used.

[27]Equation (3.23) is obtained by setting $\int_0^t=\int_0^{\infty}-\int_t^{\infty}$ in (3.20). Each term in (3.23)–(3.29) may be continued to negative times in two different manners: (i) analytically, which will be indicated by a subscript "a"; thus $g_a^*(t)$, $h_a(t)$, α_a, $(\beta t)_a$ are given by (3.24)–(3.29) for both positive and negative values of t; (ii) by requiring that the symmetry $f(-t)=f(t)^*$ be obeyed; this is what will always be understood except when the subscript a appears); thus, e.g., $h(t)$ at negative t is defined by $h(-t)=h(t)^*$, $t>0$; likewise for α, βt, etc. Note that dg^*/dt and $h(t)$ are discontinuous at $t=0$, so that their FT's have slowly decaying tails [see (6.10) for $\hat{h}(v)$]; by contrast, $h_a(t)$ is continuous at $t=0$, but does not satisfy $h_a(-t)=h_a(t)^*$. The decompositions (3.23) and (3.27) with analytic continuation to negative t are of no interest, since, e.g., $h_a(t)$ does not vanish as $t\to\infty$ [contrary to $h(t)$], as is desirable, but rather becomes linear in t [to see this, assume for simplicity that $\psi(t)$ vanishes outside $(-1,1)$; then, for $T\gg 1$, $h_a(-T)\simeq\int_{T-1}^{T+1}s\,ds\,\psi(s-T)\simeq T\int_{-1}^{1}ds\,\psi(s)$.]

[28]More generally, for any function $f(t)=\int_{-\infty}^{\infty}dv\,e^{ivt}\hat{f}(v)$:

$$f(t)=\sum_{j=0}^{n-1}t^jf^{j}(0)/j!+\int_{-\infty}^{\infty}dv\,\hat{f}(v)\left[e^{ivt}-\sum_{j=0}^{n-1}(ivt)^j/j!\right],$$

wherein $\hat{f}(v)$ may be replaced by $\hat{f}^k(v)/(iv)^k$ for any $k\leq n$ [since $f^k(t)=\int_{-\infty}^{\infty}dv\,e^{ivt}(iv)^k\hat{f}(v)\Rightarrow\hat{f}^k(v)\equiv(2\pi)^{-1}\int_{-\infty}^{\infty}dt\,e^{-ivt}f^k(t)=(iv)^k\hat{f}(v)$]. In (3.30), $g(0)=\dot{g}(0)=0$ and $\psi(t)=-\ddot{g}(t)\Rightarrow\hat{\psi}(v)=v^2\hat{g}(v)$.

[29]Defining $\hat{\psi}_+(v)=\int_0^{\infty}dt\,e^{-ivt}\psi(t)$, and using the formulas in Ref. 14, we have, from (3.24)–(3.29),

$$\text{(A)}\ \beta=-\hat{\psi}_+(0)=-i\int_{-\infty}^{\infty}dv\,\hat{\psi}(v)/(v+i0)=-i\int_{-\infty}^{\infty}dv\,\hat{\psi}(v)/v-\pi\hat{\psi}(0),$$

$$\text{(B)}\ h(t)=-i\hat{\psi}'_+{}'(0)=\int_{-\infty}^{\infty}dv\,e^{ivt}\hat{\psi}(v)/(v+i0)^2$$

$$= \int_{-\infty}^{\infty} dv\, e^{ivt}\,[\hat{\psi}'(v) + it\hat{\psi}(v)]/(v + i0),$$

(C) $\alpha = -h(+0) = i\hat{\psi}'_{+}(0) = i\pi\hat{\psi}'(0) - \int_{-\infty}^{\infty} dv\, \hat{\psi}'(v)/v.$

We denoted $\psi'(s) = \psi(s + t)$ the translate of $\psi(s)$ by t, so that $\hat{\psi}'(v) \equiv (2\pi)^{-1}\int_{-\infty}^{\infty} ds\, e^{-ivs}\psi'(s) = e^{ivt}\hat{\psi}(v)$. Equations (3.31) follow from (A)–(C). Equations (A)–(C), together with (3.30), yield (3.27). Note that (B) and (C) exist only if $\hat{\psi}'(v)$ is nonsingular at $v = 0$, i.e., if $|\int_0^{\infty} t dt\, \psi(t)| < \infty$, as was assumed for obtaining (3.27); if this is not the case, but $\hat{\psi}(v)$ is nonsingular at $v = 0$, as implied by (3.14), we rather have (3.23) with $g^{\#}(t) = \int_{-\infty}^{\infty} dv(e^{ivt} - 1)\,\hat{\psi}(v)/(v + i0)$ not separable into $h(t) + \alpha$. As said in Ref. 27, the above expressions are of interest only if continued to negative t by requiring $f(-t) = f(t)^*$, i.e., by replacing $(v + i0)$ by $(v - i0)$. Note from (B) that $\hat{h}_a(v) \equiv \mathrm{FT}\{h_a(t)\} = \hat{\psi}(v)/(v + i0)^2$, the singularity at $v = 0$ reflecting the linear divergence of $h_a(t)$ as $t \to -\infty$ (see Ref. 27); by contrast, $\hat{h}(v) \equiv \mathrm{FT}\{h(t)\} = \pi^{-1}\,\mathrm{Re}\int_0^{\infty} dt\, e^{-ivt}h(t)$ is regular [see (6.10)], since $h(t) = h(-t)^*$ vanishes as $t \to \pm\infty$.

[30]An example where this occurs is the square-well model of pressure broadening (Ref. 7) in the static approximation — a rather artificial case indeed [there, $\psi(t) = V_0 W^2 e^{iWt}$ where W is the well depth, V_0 the interaction volume].

[31]In Ref. 7, we defined $\kappa = 2(2\ln 2)^{1/2}$, i.e., twice (3.40).

[32]The approximation (3.39) is good if, loosely, $|\Lambda^2(t/\Lambda)^3\Gamma(t/\Lambda)| \ll 1$ for $|t| \lesssim 1$, i.e., within the interval wherein $e^{-(1/2)t^2}$ is sizable. Now, $|\Gamma(t)| < \mathrm{Max}|\dot{\psi}(t)|/3! < 1$ [since $\mathrm{Max}|\dot{\psi}(t)| = 1$ if Θ is defined by (3.16v); see Ref. 26]. We thus get (3.41). Much more sophisticated and rigorous validity conditions for the related "normal approximation" in probability theory may be found in Refs. 10.

[33]A. Royer, Acta Phys. Pol. A **54**, 805 (1978).

[34]According to (4.10), the validity conditions of the approximation $e^{-\gamma c}(1 + v^2)^{-1}$ are $|\gamma v| \ll 1$ and $\gamma^2 \ll 1$ [rather than $\gamma \ll 1$ in (4.11), required to make $e^{-\gamma c} \simeq 1$].

[35]J. Szudy and W. E. Baylis, J. Quant. Spectrosc. Radiat. Trans. **15**, 641 (1975); see also Ref. 33.

[36]To first order in v, $\hat{\psi}(v) = \hat{\psi}(0) + v\hat{\psi}'(0) = \pi^{-1}(b + va)$, in view of (3.31), leading back to (4.12).

[37]The tedious algebra was performed by computer in the *ALTRAN* language [W. S. Brown, *ALTRAN User's Manual*, 3rd ed. (Bell Laboratories, Murray Hill, NJ, 1975). Higher order terms are displayed in Appendix B, to avoid cluttering the main text.

[38]Numerical evaluation of the coefficients up to 8th order in γ, using the square-well model of pressure broadening (Ref. 7), indicates a radius of convergence of order 1 (as one would expect since a and the T_{kj} are of order 1).

[39]We may have, e.g., $h(t) \sim t^{-\sigma}e^{iWt}$ with $\sigma \geqslant 0$ and $W \neq 0$ (see Appendix A).

[40]Let $\sigma > 1$ in (4.16); T_{kj}, Eq. (4.9), exists if $j < k\sigma - 1$, i.e., if $j \leqslant \sigma - 1$ since $k > 1$. In (4.15), T_{kj} appears only in terms of order $\geqslant j + 2$, since T_{kj} always appears in the combination $\gamma^{k+j+1}T_{kj}$, $k \geqslant 1$, in (4.8). Thus, ill-defined T_{kj}'s appear only above order $\sigma + 1$ in (4.15). (This remains true if $-1 < \sigma < 1$).

[41]To determine the validity conditions of (4.5), let us write the real part of (4.3) as

$$\hat{P}(v) = (1 + v^2)^{-1}[1 + (1 + v^2)\mathrm{Re}\, T^{\#}(v)],$$

where $T^{\#}(v)$ is given by (4.7), but with $h(t)$ replaced by $g^{\#}(t) = \alpha + h(t)$. We have $g^{\#}(t) \sim t^{-\epsilon}$ as $t \to \infty$, where $\epsilon = 0$ if $\sigma > 0$ in (4.16), and $\epsilon = \sigma$ if $-1 < \sigma < 0$. Defining $S^{*\#}_{kj}$ by (4.18), but with $h(t)$ replaced by $g^{\#}(t)$, we have again (4.21) with the replacements $T \to T^{\#}$, $S \to S^{\#}$, $\sigma \to \epsilon$. There follows, to lowest orders in γ and v, and neglecting constants of order 1,

$$T^{\#}(v) = (\gamma S^{0\#\prime}_{10} + \gamma^2 v S^{0\#\prime}_{11} + \cdots) + (\gamma^{1+\epsilon}S^{0\#\prime\prime}_{10} + v\gamma^{1+\epsilon}S^{0\#\prime\prime}_{11} + \cdots) \simeq \gamma^{1+\epsilon}(1 + |v|),$$

whence $\hat{P}(v) \simeq (1 + v^2)^{-1}[1 + \gamma^{1+\epsilon}(1 + |v|)]$. Thus $\hat{P}(v) \simeq (1 + v^2)^{-1}$ if $\gamma^{1+\epsilon} \ll 1$ and $\gamma^{1+\epsilon}|v| \ll 1$; these become (4.11) if $\sigma \geqslant 0 \Rightarrow \epsilon = 0$, and (4.22) if $-1 < \sigma < 0 \Rightarrow \epsilon = \sigma$.

[42]If we were to substitute (4.14) and (4.6) with (4.21) into (2.23) then, because (4.21) contains an expansion in powers of v [rather than of γv as in (4.8)], and $v^{(0)}_{\pm} \neq 0$, each power of γ would appear in infinitely many terms of (2.23b). Although such series could presumably be evaluated, the above method is clearly not pleasant.

[43]Equation (4.23) is obtained by setting $v = x + (v - x)$ in (4.3), and expanding in $v - x$ and $g^{\#}$. Note that such expansions do not provide practical approximations to $\hat{P}(v)$ itself, as they are usable over frequency ranges $|v - x| \ll 1$ much smaller than the width of $\hat{P}(v)$ [unlike (4.10) which preserves $(1 + v^2)^{-1}$ unexpanded].

[44]We thereby obtain: $\Sigma_{k,j} j v^j_s \rho^k K^0_{kj}/(k!j!) = 0$ and

$$\Sigma_{k,j}[v_{\pm} - (\pm 1)^j]\rho^k K^{\pm}_{kj}/(k!j!) = \tfrac{1}{2}\Sigma_{kj} v^j_s \rho^k K^0_{kj}/(k!j!).$$

[45]Note that $S^{*\#}_{0j} = \mathrm{Re}\int_0^{\infty} dt\, e^{-t(1+ix)}(-it)^j = \mathrm{Re}\{i^j!/[(i - x)^{j+1}k!]^{-1}\}$ are just numbers. In all results, the $S^{*\#}_{0j}$ are given their numerical values.

[46]This makes them appear needlessly early in the expansions; we rather want such quantities to appear in terms of the order in γ at and beyond which they actually contribute, once they are themselves expanded in γ.

[47]Equations (4.27) are deduced from (4.25) by substituting $K^x_{kj} = \gamma[\cos(\gamma a)\Gamma^{xr}_{kj} - \sin(\gamma a)\Gamma^{xi}_{kj}]$, which follows from expanding (4.3) as $\hat{P}(v) = \pi^{-1}e^{-\gamma c}\Sigma^{\infty}_{kj=0}(v - x)^j\gamma^k\mathrm{Re}\{e^{i\gamma a}\gamma\Gamma^x_{kj}\}$, and comparing with (4.23).

[48]These follow by applying (3.32) in (4.9), (4.18), etc. The realization $K^x_{kj} \to (-)^j K^{-x}_{kj}$ also follows directly from expansions (4.23), since we have, setting $x = -y$, $\hat{P}(v) = \Sigma(v + y)^j\rho^k K^{-y}_{kj}$, and $I\hat{P}(v) = \hat{P}(-v) = \Sigma(-v + y)^j\rho^k K^{-y}_{kj} = \Sigma(v - y)^j(-)^j K^{-y}_{kj}$, whence $I: K^x_{kj} \to (-)^j K^{-x}_{kj}$ on comparing with (4.23). There are still other realizations, e.g., $I: \gamma \to -\gamma$, $T_{kj} \to (-)^{k+j+1}T^{*}_{kj}$.

[49]As mentioned in Ref. 15, these realizations, together with (2.26), provide welcome tests for the exactness of our expansions. They imply, e.g., that in (4.25), *shift* consists of terms $\rho^N \Pi_i K_{k_i j_i}$ with Σj_i odd, and that in *width*, the $K^{\pm}_{kj}$ appear in combinations, e.g. $K^{+}_{kj} \pm K^{-}_{kj}$, of definite parity under I. There are some other useful tests: e.g., since only the combination $\rho^k K_{kj}$ appears in (4.23), all terms in (4.25) must be of the form $\rho^N \Pi_i K_{k_i j_i}$ with $\Sigma k_i = N$; likewise, all terms in (4.15) must be of the form $\gamma^N a^m \Pi_i T_{k_i j_i}$ with $m + \Sigma(k_i + j_i + 1) = N$.

[50]We used [see Ref. 51, p. 302, Eqs. (7.4.6) and 7.4.7)]
$\int_0^{\infty} dt\, e^{-(1/2)t^2}\cos(yt) = (\pi/2)^{1/2}e^{-(1/2)y^2}$,
$\int_0^{\infty} dt\, e^{-(1/2)t^2}\sin(yt) = e^{-(1/2)y^2}\int_0^y ds\, e^{(1/2)s^2} = 2^{1/2}D(y/2^{1/2})$.

[51]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).

[52]Equation (5.13b) follows from repeated use of $(d/dy)E^j(y) = -yE^j(y) + 1$, as compared to $(d/dy)E'(y) = -yE'(y)$.

[53]$\hat{P}_+(y) = \int_0^{\infty} dt \exp[-iYt - \tfrac{1}{2}t^2 - i(y - Y)t]F(it)$
$= F(-d/dY)e^{(y-Y)d/dY}E(Y) = \Sigma^{\infty}_{k=0}\epsilon^k b_k(-d/dY)\Sigma^{\infty}_{j=0}(y - Y)^j(d/dY)^j E(Y)/j!$, whence (5.17)–(5.18).

[54]Numerical examples of these expansions, using the square-well model of pressure broadening (for which the Γ_k are complex) are found in Ref. 7.

[55]As mentioned in Ref. 15, these provide welcome tests for the exactness of our expansions. There are other useful tests: e.g., since only the combination $\epsilon^k\Gamma_k$ appears in (5.9), all expansions must consist of terms $\epsilon^N \Pi_i \Gamma_{k_i}$ with $\Sigma k_i = N$.

[56]See, e.g., Sec. VI of Ref. 7.

[57]See Ref. 11, Chap. 9. Poisson laws have $g(t) = c_1(e^{ic_2 t} - 1) + ic_3 t$, and normal laws have $g(t) = c_4 t^2$, c_i constants. A normal term arises in the Riemann sum (3.30) if $\hat{\psi}(v)$ has a δ singularity at $v = 0$, since $\mathrm{Lim}_{\epsilon \to 0}\int_{-\epsilon}^{\epsilon} dv(e^{ivt} - 1 - ivt)\hat{\psi}(v)/v^2 = -t^2\mathrm{Lim}\int_{-\epsilon}^{\epsilon} dv\,\hat{\psi}(v)$. Note that the weak broadening limit profile is a Poisson law, while the strong broadening limit profile is a normal law.

[58]$\hat{P}_{as}(v)$ is sometimes proposed (Ref. 5) as an improvement to the pure Lorentzian $(1 + v^2)^{-1}$. The shape parameters of $\hat{P}_{as}(v)$ can be found in closed form:

$$\mathrm{shift}_{as} = A, \quad (v_{\pm})_{as} = 2A(1 \pm \tfrac{1}{2}B), \quad \mathrm{width}_{as} = 2AB, \quad \mathrm{asym}_{as} = [(B + 1)/(B - 1)] - 1,$$

where $A = [1 - \cos(\gamma a)]/\sin(\gamma a)$, $B = \{[3 - \cos(\gamma a)]/[1 - \cos(\gamma a)]\}^{1/2}$. These may be expanded in powers of γa, of course yielding (4.15) wherein $h(t)$, i.e., all the T_{kj}, are set to zero. The approximation $\hat{P}_{as}(v)$ is useful if $h(t)$ is small, making the T_{kj} (and also a and d) small. But in general, it is more consistent to use (4.12), since T_{kj}'s appear at order γ^2 in (4.10) (see Ref. 33). Conversely, one might choose not to expand $e^{i\gamma a}$ in obtaining (4.10) and (4.15), which will then contain $\cos(\gamma a)$ and $\sin(\gamma a)$ instead of powers of a.

[59]E.g., pressure shifted hyperfine lines; see, e.g., R. R. Freeman, D. E. Pritchard, and D. Kleppner, Phys. Rev. A **13**, 907 (1976).

[60]Note that $P(\omega) = P_1^{*N}(\omega)$ is infinitely divisible (i.d.) if and only if $P_1(\omega)$ is. But if a bulk limit, $N \to \infty$, Volume $\to \infty$, N/Volume $= n$ finite, is taken, as, e.g., in the Anderson–Talman–Baranger treatment of pressure broadening (Ref. 5), there of course results an i.d. spectrum.

Recursive formulas for Morse-oscillator matrix elements of arbitrary powers of $1 - \exp[-a(r - r_e)]$

Loc Binh Tran and J. N. Huffaker
Department of Physics and Astronomy, Oklahoma University, Norman, Oklahoma 73019

(Received 14 October 1981; accepted for publication 6 November 1981)

The variable $y = 1 - \exp[-a(r - r_e)]$ is a natural one to use in connection with a Morse oscillator. A simple formula is derived via the factorization method relating the Morse matrix element $\langle 0|y^n|0\rangle$ to $\langle 0|y^{n-1}|0\rangle$ and $\langle 0|y^{n-2}|0\rangle$. Another simple formula is derived, with which all matrix elements $\langle v'|y^n|v\rangle$ can be calculated recursively, starting from values of $\langle 0|y^n|0\rangle$.

PACS numbers: 34.20.Be

I. INTRODUCTION

Eigenfunctions of the Morse oscillator are solutions of the Schrödinger equation

$$d^2\psi_v/dr^2 + (2\mu/\hbar^2)(E_v - V_e y^2)\psi_v = 0, \tag{1a}$$

where

$$y = 1 - \exp(-aq), \quad q = r - r_e, \tag{1b}$$

and μ, V_e, a, and r_e are parameters describing the oscillator. The standard way to solve this eigenvalue problem is to treat it as a class II, B-type factorization problem.[1] When we define the dimensionless Morse parameter[2]

$$\sigma = (2\mu V_e)^{1/2}/(a\hbar), \tag{2}$$

the energy eigenvalues are

$$E_v{}^\sigma = V_e[2(v + \tfrac{1}{2})/\sigma - (v + \tfrac{1}{2})^2/\sigma^2], \tag{3}$$

the physically normalized ground-state eigenfunctions are given by[3]

$$\psi_0{}^\sigma(r) = [a/\Gamma(2s)]^{1/2}(2\sigma)^s e^{-saq}\exp(-\sigma e^{-aq}), \tag{4}$$

where $s = \sigma - \frac{1}{2}$, and the normalized excited-state eigenfunctions may be calculated recursively from lower-state eigenfunctions for *other* Morse oscillators by the raising and lowering operators $\mathscr{B}_v{}^{\pm}(\sigma)$, as follows:

$$\psi_v{}^\sigma = \mathscr{B}_v{}^-(\sigma)\psi_{v-1}{}^{\sigma-1}, \tag{5a}$$

$$\psi_{v-1}{}^{\sigma-1} = \mathscr{B}_v{}^+(\sigma)\psi_v{}^\sigma, \tag{5b}$$

where the operators may be written in the form

$$\mathscr{B}_v{}^{\pm}(\sigma) = [v(2\sigma - v - 1)]^{-1/2}[\tfrac{1}{2}e^x - s \pm d/dx], \tag{5c}$$

where

$$e^x = 2\sigma e^{-aq}. \tag{5d}$$

Using Eqs. (5a)–(5d) one can also obtain the recursion relation

$$e^x\psi_v{}^\sigma = C^\sigma\psi_v{}^\sigma + D_v{}^\sigma\psi_{v-1}{}^{\sigma-1} + D_{v+1}{}^{\sigma+1}\psi_{v+1}{}^{\sigma+1}, \tag{6a}$$

where

$$C^\sigma = 2\sigma, \quad D_v{}^\sigma = v(2\sigma - v - 1). \tag{6b}$$

Because Eq. (5c) contains an external positive sign, the relative phases of the eigenfunctions are fixed: they are all positive at the *inner* "classical turning point."

An alternative way to solve Eq. (1) is via class I, F-type factorization[4]: Eqs. (3) and (4) are again obtained, but now normalized excited-state eigenfunctions are calculated from lower-state eigenfunctions of the *same* Morse oscillator by the raising and lowering operators $\mathscr{G}_v{}^{\pm}(\sigma)$, where

$$\psi_{v+1}{}^\sigma = \mathscr{G}_v{}^+(\sigma)\psi_v{}^\sigma, \tag{7a}$$

$$\psi_v{}^\sigma = \mathscr{G}_v{}^-(\sigma)\psi_{v+1}{}^\sigma, \tag{7b}$$

and where the operators are given by

$$\mathscr{G}_v{}^{\pm}(\sigma) = \left(\frac{(\sigma - v - 1 \mp \frac{1}{2})(\sigma - v - 1)^2}{(\sigma - v - 1 \pm \frac{1}{2})(2\sigma - v - 1)(v + 1)}\right)^{1/2} \times \frac{1}{2\sigma}\left(\frac{(\sigma - v - 1 \pm \frac{1}{2})}{1 - y} - \frac{\sigma^2}{(\sigma - v - 1)} \mp \frac{d}{dy}\right), \tag{7c}$$

where y was given in Eq. (1b). In addition, one obtains the recursion relation

$$(e^{aq} - 1)\psi_v{}^\sigma = A_v{}^\sigma\psi_v{}^\sigma + B_v{}^\sigma\psi_{v-1}{}^\sigma + B_{v+1}{}^\sigma\psi_{v+1}{}^\sigma, \tag{8a}$$

where

$$A_v{}^\sigma = \frac{2\sigma(v + \frac{1}{2}) - v(v + 1)}{(\sigma - v)(\sigma - v - 1)}, \tag{8b}$$

$$B_v{}^\sigma = \frac{\sigma}{2(\sigma - v)}\left(\frac{v(2\sigma - v)}{(\sigma - v - \frac{1}{2})(\sigma - v + \frac{1}{2})}\right)^{1/2}. \tag{8c}$$

The use of Eq. (7c) produces eigenfunctions all of which are positive at the *outer* "classical turning point."

The variable y of Eq. (1b) is a very appropriate one to use in connection with a Morse oscillator. For example, expansions in powers of y can provide very accurate models of actual vibrational potentials for diatomic molecules.[2,3,5–9] It has also been suggested[10] that dipole moment functions be expanded in powers of y. Morse matrix elements of powers of y can thus be useful for several different purposes: First, where a realistic vibrational potential is expressed as a Morse potential plus terms in higher powers of y, energy eigenvalues can be obtained, along with eigenfunctions expressed as linear combinations of Morse eigenfunctions, by matrix diagonalization.[9] Second, using such eigenvectors, one can calculate matrix elements of any function that is a power series in y via matrix multiplication.

In Ref. 9, Morse matrix elements of y were obtained by calculating the matrix for e^{aq} using Eq. (8), then subtracting the inverse matrix from the unit matrix. Matrices for higher powers of y were then obtained by matrix multiplication. All matrix operations were performed in a truncated Morse basis, which introduced some error. Reference 10 gives formulas for all Morse matrix elements $\langle v'|y^n|v\rangle$ for $n \leq 4$.

 0022-2488/83/020397-02$02.50

In this paper we derive recursion formulas that permit the step-by-step recursive calculation of $\langle v'|y^n|v\rangle$ for arbitrary power n. The derivation divides into two parts: that for $\langle 0|y^n|0\rangle$, and that for $\langle v'|y^n|v\rangle$ itself.

II. RECURSIVE FORMULA FOR $\langle 0|y^n|0\rangle$

Our derivation of the formula for $\langle 0|y^n|0\rangle$ makes use of the technique of class II, B-type factorization.[1] Defining a variable

$$W = 2\sigma - e^x = 2\sigma y, \tag{9}$$

we obtain from Eq. (6a) the recursion formula

$$W\psi_0{}^\sigma = -D_1{}^{\sigma+1}\psi_1{}^{\sigma+1}. \tag{10}$$

Multiplying by W^{n-1} where $(n \geqslant 1)$, then by $\psi_0{}^\sigma$, and integrating, we obtain

$$\langle 0|W^n|0\rangle = -D_1{}^{\sigma+1}\int \psi_0{}^\sigma W^{n-1}\psi_1{}^{\sigma+1}\,dr. \tag{11}$$

We next replace $\psi_1{}^{\sigma+1}$ with $\mathscr{B}_1{}^{-}(\sigma+1)\psi_0{}^\sigma$, using Eq. (5a), to produce

$$\langle 0|W^n|0\rangle = -\int \psi_0{}^\sigma W^{n-1}\,[\tfrac{1}{2}e^x - (s+1) - d/dx]\psi_0{}^\sigma dr. \tag{12}$$

Integrating the right-hand side of Eq. (12) by parts and rearranging the terms, we obtain

$$\langle 0|W^n|0\rangle = -\int \psi_0{}^\sigma W^{n-1}(\tfrac{1}{2}e^x - s + d/dx)\psi_0{}^\sigma\,dr + \int \psi_0{}^\sigma(W^{n-1} - dW^{n-1}/dx)\psi_0{}^\sigma\,dr. \tag{13}$$

The first integral on the right-hand side vanishes, as the integrand contains the application of a lowering operator to $\psi_0{}^\sigma$. Using the identity

$$dW^{n-1}/dx = (n-1)(W^{n-1} - 2\sigma W^{n-2}), \tag{14}$$

we obtain from Eq. (13) the recursive formula,

$$\langle 0|W^n|0\rangle = 2\sigma(n-1)\langle 0|W^{n-2}|0\rangle - (n-2)\langle 0|W^{n-1}|0\rangle. \tag{15}$$

Dividing by $(2\sigma)^n$, we obtain the desired result,

$$\langle 0|y^n|0\rangle = [(n-1)\langle 0|y^{n-2}|0\rangle - (n-2)\langle 0|y^{n-1}|0\rangle]/2\sigma. \tag{16}$$

III. RECURSIVE FORMULA FOR $\langle v'|y^n|v\rangle$

We now apply techniques of class I, F-type factorization[4] to obtain the formula for $\langle v'|y^n|v\rangle$. Multiplying Eq. (8a) by e^{-aq} and rearranging terms, we obtain

$$y|v\rangle = A_v|v\rangle + B_v|v-1\rangle + B_{v+1}|v+1\rangle - A_v y|v\rangle - B_v y|v-1\rangle - B_{v+1}y|v+1\rangle. \tag{17}$$

(Since all quantities in this equation refer to the same Morse oscillator, we suppress the label $\pm\square\sigma$.) Multiplying by y^{n-1} and again rearranging terms, we can write

$$B_{v+1}y^n|v+1\rangle = B_{v+1}y^{n-1}|v+1\rangle + A_v y^{n-1}|v\rangle + B_v y^{n-1}|v-1\rangle - (1+A_v)y^n|v\rangle - B_v y^n|v-1\rangle. \tag{18}$$

Replacing $v+1$ by v, multiplying by $\langle v'|$, and solving for $\langle v'|y^n|v\rangle$, we obtain the desired recursive formula, which holds for $v > 0$:

$$\langle v'|y^n|v\rangle = \langle v'|y^{n-1}|v\rangle + [A_{v-1}\langle v'|y^{n-1}|v-1\rangle + B_{v-1}\langle v'|y^{n-1}|v-2\rangle - (1+A_{v-1})\langle v'|y^n|v-1\rangle - B_{v-1}\langle v'|y^n|v-2\rangle]/B_v. \tag{19}$$

If we think of the matrix element $\langle v'|y^n|v\rangle$ as lying on the $(v'+1)$th row and $(v+1)$th column of the $(n+1)$th sheet of a three-dimensional matrix, the recursive procedure for calculating the elements of the matrix can be outlined as follows. First, the bottom sheet (for $n = 0$) is simply the orthonormality condition,

$$\langle v'|y^0|v\rangle = \delta_{v',v}. \tag{20}$$

Starting with $\langle 0|y^0|0\rangle = 1$, one can then use Eq. (16) repeatedly to produce values of $\langle 0|y^n|0\rangle$ for successively higher sheets. If the nth sheet is already known (for whatever numbers of rows and columns are desired, subject to the restriction $v_{\max} < \sigma - \frac{1}{2}$), then the first row of the $(n+1)$th sheet can be calculated from Eq. (19), starting with the known value of $\langle 0|y^n|0\rangle$. Since each sheet is a real, symmetric matrix, one can produce the first column of this sheet by reflecting the first row. Elements of each additional row may now be obtained by use of Eq. (19) (or by reflection, if $v' > v$).

Since the derivation of Eq. (19) is based on class I, F-type factorization, the phases of the matrix elements produced by the above procedure are fixed by the convention that all eigenfunctions are positive at the outer turning points.[10] Matrix elements corresponding to the other convention differ by a relative phase $(-1)^{v+v'}$. Of course, Eq. (16) is independent of convention.

As a verification of Eqs. (16) and (19), we used them to calculate a number of matrix elements for $n \leqslant 4$, which were found to agree with the formulas of Ref. 10.

[1] L. Infeld and T. E. Hull, Rev. Mod. Phys. **23**, 21 (1951).
[2] J. N. Huffaker, J. Chem. Phys. **64**, 3175 (1976).
[3] P. H. Dwivedi and J. N. Huffaker, J. Chem. Phys. **66**, 1726 (1977).
[4] J. N. Huffaker and P. H. Dwivedi, J. Math. Phys. **16**, 862 (1975).
[5] J. N. Huffaker, J. Chem. Phys. **64**, 4564 (1976).
[6] J. N. Huffaker and P. H. Dwivedi, J. Chem. Phys. **68**, 1303 (1978).
[7] J. N. Huffaker, J. Mol. Spectrosc. **65**, 1 (1977); **71**, 160 (1978).
[8] J. N. Huffaker, J. Chem. Phys. **70**, 2720 (1979); **72**, 2601 (1980).
[9] J. N. Huffaker, J. Chem. Phys. **74**, 1217 (1981).
[10] J. N. Huffaker and L. B. Tran, J. Chem. Phys. **76**, 3838 (1982).

Path integral approach to multiple scattering

Eugene P. Gross
Physics Department, Brandeis University, Waltham, Massachusetts 02254

(Received 6 October 1981; accepted for publication 13 November 1981)

Multiple scattering of a wave in a system of scatterers with random, uncorrelated positions is studied with path integrals. The Edwards–Gulyaev expression for the position averaged Green's function is used to find the density expansion of the complex optical potential. The expansion is in terms of exact medium propagators and scattering matrices in the medium. The first term is the coherent potential approximation. A source dependent generalization of the path integral is used to derive a functional equation for the optical potential. This leads to a hierarchy for correlation functions that involves exact medium propagators and scattering matrices. The simplest truncations yield new integral equations that are generalizations of the coherent potential approximation and are compatible with the density expansion.

PACS numbers: 42.20. – y, 03.40.Kf, 03.65.Nk

I. INTRODUCTION

Our interest is in the characteristics of waves interacting with discrete scatterers with random site positions. For concreteness we deal with a Schrödinger wave, i.e., a quantum mechanical particle in a random medium. There is a vast literature dealing with this problem, particularly from a multiple scattering viewpoint.[1] This approach uses the site or "atomic" scattering matrices and makes possible a succinct treatment of strong scattering (including hard core interactions) and resonant scattering (including bound state effects). In addition, in the coherent potential approximation, one uses the scattering matrix in a medium characterized by the exact site averaged propagator. The result is particularly simple in the case that the sites are completely uncorrelated.[2] If v is the potential operator and $G(\mathbf{k}|E)$ the momentum diagonal site averaged propagator in energy space, we have for the atomic scattering operator.

$$t = v + vGt. \tag{1.1}$$

The complex optical potential $\Sigma(k|E)$ enters into the averaged Green's function as

$$G^{-1} = G_0^{-1} - \Sigma. \tag{1.2}$$

It is given in the coherent potential approximation by

$$\Sigma(\mathbf{k}|E) = n < k|t|\mathbf{k}). \tag{1.3}$$

This is just the density n of scatterers times for forward scattering amplitude.

It is a nontrivial matter to improve the coherent potential approximation in a controlled manner, even for the completely uncorrelated case. One definite question is to ask for the density expansion of $\Sigma(k|E)$ in a form involving atomic scattering matrices and the exact averaged propagator. The result for the first term beyond the coherent potential approximation is contained in the two body additive approximation in our recent work on multiple scattering.[3] Of course the main focus in that work and in other condensed matter studies[4] is on the case of arbitrarily correlated scatterers in dense systems. Still the density expansion provides a test for such theories and the results should have practical implications for dilute impurity systems in solid state physics, for electromagnetic wave propagation in gases, and other such problems.

Here we study the density expansion from the very different path integral viewpoint. The path integral representation emphasizes the space-time aspects of the propagator, in contrast to the momentum-energy emphasis of the multiple scattering approach. However, the natural approximation schemes treat strong potentials in only indirect, cumbersome ways. The role of complex optical potentials and exact medium propagators is unclear. On the other hand the path integral representation has definite virtues. If the characteristic function for a random process is explicitly known, one can do the impurity averaging first, and one obtains a single, multitime path integral for the averaged propagator.[5] This is the case for the Gaussian random process. It is also the case for the Poisson process describing uncorrelated sites. The result is a path integral studied by Edwards and co-workers,[6] and by Jones and Lukes[7] for the real time case and by Friedberg and Luttinger[8] in the temperature density matrix form. It has been used to study the Lipschitz deep traps that arise from large density fluctuations,[9] and lead to an infinite tail in the density of states. In the Friedberg–Luttinger paper, a systematic cumulant expansion based on a trial potential was developed. (See, however, our article[10] for a discussion of the implications of breaking translation invariance.) The deep traps are not treated in the multiple scattering approach. Even in cluster extensions of the coherent potential there are only finite low energy tails in the density of states.

The present work is devoted to developing a formalism that combines the strong points of the two approaches. We find that progress can be made in the uncorrelated site problem. There is a strong incentive to pursue this. In the path integral representation it is easy to write explicit multipath integrals for the averages of products of Green's functions. These quantities are needed in the study of transport phenomena. This has been exploited in the theory of random continuum fluctuations,[11] as in sound transmission in fluctuating media, or laser scattering from turbulent eddies. It has been possible to study the so-called strong fluctuation regime, i.e., transmission over long paths.

Essentially arbitrary time and space correlations of the wave field can be computed. It is true that these results rely

 0022-2488/83/020399-07$02.50

on special features of the problem. In particular the wavelength is short compared to the size of the fluctuation. There is negligible back scatter and strong forward scatter so that the parabolic approximation is valid. It is also true that the same results can be obtained somewhat more cumbersomely by other methods. Still the results for the continuous fluctuations are impressive, and one would like to have a corresponding theory for discrete scatterers.

Here we deal with the simplest situation, viz., the determination of the average propagator for the case of uncorrelated discrete scatterers. The detailed outline is as follows.

In Sec. 2 we construct the Edwards–Gulyaev path integral. Noting that the density appears as an explicit parameter, we find the density expansion of the average propagator in terms of bare Green's functions and bare atomic scattering matrices. By explicit inversion, term by term, one finds the optical potential Σ in terms of the same quantities. Continuing in the most straightforward way, the bare scattering matrices are expressed in terms of scattering matrices in a medium characterized by the exact site averaged propagator. This yields the explicit representation of the optical potential in terms of medium scattering matrices, exact propagators, and in powers of the density. The usual coherent potential approximation is the first term in the series.

In Sec. 3 we generalize the path integral to include a space dependent source $J(x)$. In the limit of zero density the source generates the bare atomic scattering matrices. An integration by parts technique is used. It converts the path integral into a hierarchy of equations for correlation functions. With the source generating function one finds an equivalent functional equation for the source dependent, site averaged, Green's function. The source function is different from the one introduced in the standard Schwinger approach to quantum field theory. It uses the explicit characteristic function of the Poisson process. There follow two illustrations of the use of the functional equation. First we check that successive functional derivatives yield the expansion of the optical potential in terms of the bare scattering matrices obtained in Sec. 2. Second, the space-time cumulant expansion is obtained.

In Sec. 4 the standard machinery of quantum field theory is put to work to yield a functional equation for the source dependent optical potential. It involves the exact propagator. A hierarchy is obtained by taking successive functional derivatives and evaluating them in the limit that the source vanishes. The hierarchy has the feature that a truncation involving neglect of a given order functional derivative yields results accurate to a corresponding power of the density. One obtains a number of integral equations in terms of exact propagators and scattering matrices, depending on the precise truncation. These may have some validity at higher densities in the same sense that the coherent potential approach may be more accurate than is justified by its agreement with the perturbation expansion to only low order.

II. DENSITY EXPANSION OF THE PATH INTEGRAL

Let the particle impurity interaction be $v(\mathbf{x}-\mathbf{R}_1)$, with the impurities at site positions $\mathbf{R}_1,\ldots,\mathbf{R}_N$. The Green's function for a given configuration is given by the path integral

$$\langle \mathbf{x}|\hat{g}(t\,|\mathbf{R}_1\cdots\mathbf{R}_N)|\mathbf{x}^1\rangle = -i\theta(t)\int D_t\mathbf{x}\delta(\mathbf{x}(t)-\mathbf{x})\delta(\mathbf{x}(0)-\mathbf{x}^1) \times\exp\left(-i\sum_{g=1}^{N}\int_0^t v(\mathbf{x}(u)-\mathbf{R}_g)du\right). \tag{2.1}$$

Here

$$D_t x = Dx\exp\left(\frac{i}{2}\int_0^t \mathbf{x}^2 du\right). \tag{2.2}$$

The end point conditions are included by using the delta functions. $\theta(t)$ is a step function and units have been chosen so that $\hbar = m = 1$. Time dependent Green's functions have a caret.

We have the abbreviated notation

$$P\left\{\exp\left(-i\sum_{g=1}^{N}\int_0^t v(\mathbf{x}(u)-\mathbf{R}_g)du\right)\right\} \tag{2.3}$$

for the right hand side.

The average oversite position is performed with the weight function $\pi_{g=1}^{N}(d\mathbf{R}_g/\Omega)$, corresponding to uncorrelated scatters. The site averaged Green's function is

$$\langle \mathbf{x}|\hat{G}(t)|\mathbf{x}^1\rangle = P\left\{\left(\int\exp\left[-i\int_0^t v(\mathbf{x}(u)-\mathbf{R})du\right]\frac{d\mathbf{R}}{\Omega}\right)^N\right\} \tag{2.4}$$

by virtue of the interchangeability of the site averaging and path integral operations. Let

$$\int\exp\left(-i\int_0^t v(\mathbf{x}(u)-\mathbf{R})du\right)\frac{d\mathbf{R}}{\Omega} = 1+\frac{n}{N}F(t) \tag{2.5}$$

with

$$F(t) = \int\left(\exp\left[-i\int_0^t v(\mathbf{x}(u)-\mathbf{R})du\right]-1\right)d\mathbf{R}, \qquad n = N/\Omega. \tag{2.6}$$

In the limit of large N, with n finite, one finds the Edwards–Gulyaev functional

$$\langle \mathbf{x}|\hat{G}(t)|\mathbf{x}^1\rangle = P\{\exp(nF(t))\}. \tag{2.7}$$

The density of impurities appears as an explicit parameter so that the density series for the Green's function can be found by direct expansion of the exponential.

The path integral

$$P\left\{\exp\left(-i\int_0^t v(\mathbf{x}(u)-\mathbf{R}_1)du\right)\right\} \equiv \langle x|\hat{h}(t|\mathbf{R}_1)|\mathbf{x}^1\rangle \tag{2.8}$$

is the Green's function for a particle with a single impurity at $\mathbf{R}_1$. Introduce the energy representation as the Fourier transform

$$\hat{A}(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-iEt}A(E)\,dE. \tag{2.9}$$

In the energy representation, the single site Green's function obeys

$$\langle \mathbf{x}|h(E\,|\mathbf{R}_1)|\mathbf{x}_1\rangle = \langle \mathbf{x}|G_0(E)|\mathbf{x}^1\rangle + \int\langle \mathbf{x}|G_0(E)|\mathbf{x}_1\rangle v(\mathbf{x}_1-\mathbf{R}_1) \times\langle \mathbf{x}_1|h(E\,|\mathbf{R}_1)|\mathbf{x}|\rangle d\mathbf{x}_1. \tag{2.10}$$

We use a matrix notation with

$$v_1 \to v(\mathbf{x}-\mathbf{R}_1)\delta(\mathbf{x}-\mathbf{x}^1),$$
$$h_1 \to \langle \mathbf{x}|h(E|\mathbf{R}_1)|\mathbf{x}^1\rangle. \tag{2.11}$$

Then

$$h_1 = G_0 + G_0 v_1 h_1. \tag{2.12}$$

The bare scattering matrix $t_1^0(E)$ is defined by

$$h_1 = G_0 + G_0 t_1^0 G_0,$$
$$t_1^0 = v_1 + v_1 G_0 t_1^0. \tag{2.13}$$

The superscript 0 indicates that it is the scattering matrix associated with the free particle propagator G_0. (We hope that no confusion results from using the symbol t for both a scattering matrix and the time variable.)

The first order in the density, the path integral yields

$$\langle \mathbf{x}|\hat{G}(t)|\mathbf{x}^1\rangle = \langle \mathbf{x}|\hat{G}_0(t) + n\int(\hat{h}(t|\mathbf{R}_1) - \hat{G}_0(t))d\mathbf{R}_1|\mathbf{x}^1\rangle. \tag{2.14}$$

In energy space

$$G(E) = G_0 + nG_0\int t(\mathbf{R}_1)d\mathbf{R}_1 G_0. \tag{2.15}$$

We next examine the n^2 terms. The path integral involves

$$\int\int d\mathbf{R}_1 d\mathbf{R}_2 \langle \mathbf{x}|\hat{h}_{12}(t)|\mathbf{x}^1\rangle$$
$$= P\left\{\int\int d\mathbf{R}_1 d\mathbf{R}_2 \exp\left[-i\int_0^t v(\mathbf{x}(u)-\mathbf{R}_1)du - i\int_0^t v(\mathbf{x}(u)-\mathbf{R}_2)du\right]\right\}. \tag{2.16}$$

$\hat{h}_{12}(t)$ is the Green's function when there are two impurities, one at $\mathbf{R}_1$ and one at $\mathbf{R}_2$. In energy space the two site Green's function obeys

$$h_{12}(E) = G_0 + G_0(v_1+v_2)h_{12} \tag{2.17}$$

and the associated bare scattering matrix satisfies

$$t_{12}^0 = (v_1+v_2) + (v_1+v_2)G_0 t_{12}^0. \tag{2.18}$$

Multicenter bare scattering matrices were used by Luttinger and Kohn[12] in a study of the density expansion of the conductivity.

To order n^2

$$G(E) = G_0 + nG_0\int t_1^0 d\mathbf{R}_1 G_0 + \frac{n^2}{2}G_0 \times\int\int(t_{12}^0 - t_1^0 - t_2^0)d\mathbf{R}_1 d\mathbf{R}_2 G_0. \tag{2.19}$$

It is straightforward to extend this expansion in powers of density with bare propagators and scattering matrices to general order. However, the preceding expression involves cancellations from large regions of space where there is no interaction. To avoid this and still have expressions in terms of scattering matrices, one uses the identity

$$[1-(v_1+v_2)G_0]^{-1}$$
$$= [1-v_1G_0]^{-1}[1-t_2^0G_0t_1^0g_0]^{-1}[1-v_2G_0]^{-1}$$
$$= [1+t_1^0G_0][1-t_2^0G_0t_1^0G_0]^{-1}[1+t_2^0G_0]. \tag{2.20}$$

Using also $(1-x)^{-1} = 1 + x(1-x)^{-1}$, it is seen that there are terms cancelling the term $t_1^0 + t_2^0$ in the n^2 coefficient

$$G_0 = G_0 + nG_0\int t_1 dR_1 G_0 + n^2 G_0\int\int d\mathbf{R}_2 \times\left\{t_1^0G_0t_2^0G_0 + t_1^0G_0\frac{t_2^0G_0t_1^0G_0}{[1-t_2^0G_0t_1^0G_0]}(1+t_1^0G_0)\right\}. \tag{2.21}$$

The individual terms are now independent of the volume of integration.

The next step is to expand the spatial potential $\Sigma = G_0^{-1} - G^{-1}$ as a density series in terms of bare propagators and bare scattering matrices. Write the above series as

$$G = G_0 + nG_0\int t_1^0 d\mathbf{R}_1 G_0 + n^2\int\int G_0 A_{12}^0 G_0 d\mathbf{R}_1 d\mathbf{R}_2. \tag{2.22}$$

Then by direct expansion of G^{-1},

$$\Sigma = n\int t_1^0 d\mathbf{R}_1 + n^2\int\int(A_{12}^0 - t_1^0G_0t_2^0)d\mathbf{R}_1 d\mathbf{R}_2 + \cdots. \tag{2.23}$$

Finally we express Σ in terms of the scattering matrices in the medium and in terms of G. To order n^2 it is only necessary to expand t_1^0 in the first term to first order in the density. Since

$$t_1^0 = \frac{1}{1-v_1G}v_1 - \frac{1}{1-v_1G}(G-G_0)t_1^0,$$
$$G - G_0 = nG\int t_2 d\mathbf{R}_2 G_0 + \cdots, \tag{2.24}$$

we have

$$t_1^0 = t_1 - nt_1G\int t_2 d\mathbf{R}_2 G t_1. \tag{2.25}$$

This leads to

$$\Sigma = n\int t_1 d\mathbf{R}_1 + n^2\int\int\left[t_1G\frac{1}{[1-t_2Gt_1G]} \times t_2Gt_1(1+Gt_2) - t_1Gt_2Gt_1\right]d\mathbf{R}_1 d\mathbf{R}_2. \tag{2.26}$$

In fact the third order term in t is seen to cancel, so that the n^2 term starts with the fourth power in t.

We now exhibit the explicit expression for the case of a one dimensional delta function and make contact with our earlier results from the multiple scattering approach.

For the one center case when $v(x-R_1) = v\delta(x-R_1)$ the scattering matrix is

$$\langle x|t_1|x'\rangle = t\delta(x-x')\delta(x-R_1), \tag{2.27}$$

where t is the energy dependent quantity

$$t = v[1 - v\langle 0|G|0\rangle]. \tag{2.28}$$

In view of the overall translation invariance $\langle 0|G|0\rangle = \langle x|G|x\rangle$ for any x.

The two center scattering matrix is given by

$$D(R_2-R_1)\langle x|t_{12}|x'\rangle = [v(1-v\langle 0|G|0\rangle) + v^2\langle R_1|G|R_2\rangle|^2]\delta(x-R_1)\delta(x'-R_2) + 1\leftrightarrows 2]. \tag{2.29}$$

Here

$$D(R_2-R_1) = [1-v\langle 0|G|0\rangle]^2[1-t^2|\langle R_1|G|R_2\rangle|^2]. \tag{2.30}$$

With $R = R_2 - R_1$, we find

$$\tfrac{1}{2}\int\int dR_1 dR_2 \langle x|t_{12} - t_1 - t_2|x'\rangle$$
$$= t^3\delta(x - x_1)\int dR\, \frac{1\langle 0|G|R\rangle|^2}{1 - t^2|\langle 0|G|R\rangle|^2}$$
$$+ \frac{t^2\langle 0|G|x - x'\rangle}{1 - t^2|\langle 0|G|x - x'\rangle|^2}. \tag{2.31}$$

Subtracting the other terms needed to construct Σ,

$$\Sigma = nt\delta(x - x') + n^2\left\{\delta(x - x')t^5\int dR\, \frac{G^4(R)}{1 - t^2|G^2(R)|}\right.$$
$$\left. + t^4 G\frac{\langle x|G|x'\rangle|G|x'\rangle|^2}{1 - t^2|\langle 0|G|x - x'\rangle|^2}\right\}. \tag{2.32}$$

The Fourier transform $\Sigma(k|E)$ is

$$\Sigma(k|E) = nt + n^2\left\{t^5\int \frac{dRG^4(R)}{1 - t^2|G(R)|^2}\right.$$
$$\left. + t^4\int e^{ikx}\frac{G(x)|G(x)|^2}{1 - t^2|G(x)|^2}dx\right\}, \tag{2.33}$$

where the second term is k dependent. This result can be obtained from Eq. (73) of the third paper of Ref. 3.

III. INTEGRAL EQUATION FOR THE GENERATING FUNCTIONAL

Consider the functional,

$$\hat{G}(\mathbf{x}|t) = -i\theta(t)\int_0^t D_t x e^{I(t)}, \tag{3.1}$$

where

$$I(t) = \int\{n + J(\mathbf{R})\}Q(t|\mathbf{R})d\mathbf{R},$$
$$Q(t|\mathbf{R}) = -1 + \exp\left(i\int_0^t v(\mathbf{x}u) - \mathbf{R})\,du\right). \tag{3.2}$$

At zero density the J dependent term generates combinations of the bare atomic scattering matrices.

Note that

$$\frac{\delta}{\delta J(\xi)}e^{I(t)} = e^{I(t)}Q(t|\xi). \tag{3.3}$$

There is an identity

$$\int\frac{\delta\hat{G}(\mathbf{x}|t)}{\delta J(\xi)}d\xi = \frac{d\hat{G}(\mathbf{x}|t)}{dn} \tag{3.4}$$

and further identifies of the form

$$\int\frac{\delta^2\hat{G}(\mathbf{x}|t)}{\delta J(\xi_1)\delta J(\xi)}d\xi = \frac{d}{dn}\frac{\delta\hat{G}(\mathbf{x}|t)}{\delta J(\xi_1)} \tag{3.5}$$

which hold even for $J \neq 0$.

The functional equation for G is obtained by using the integration by parts[13]

$$e^{I(t)} = e^{I(\phi)} = \int_0^t ds\frac{\delta}{\delta s}\{e^{I(s)}\}$$
$$= \int_0^t ds\frac{\delta I}{\delta s}\{e^{I(s)}\}. \tag{3.6}$$

In our case

$$\frac{\delta I(s)}{\delta s} = -i\int d\mathbf{R}\{n + J(\mathbf{R})\}v(\mathbf{x}(s) - \mathbf{R})$$
$$\times[1 + Q(S|\mathbf{R})]. \tag{3.7}$$

The integral is broken up into paths running from $\mathbf{0}, 0$ *to* $\mathbf{y}, s$ and others running from $\mathbf{y}, s$ to $\mathbf{x}, t$. The part from $\mathbf{y}, s$ to $\mathbf{x}, t$ is governed by the free particle action. The part from $\mathbf{0}, 0$ to $\mathbf{y}, s$ is a correlation function expressible in terms of $\delta\hat{G}/\delta J(\xi)$.

We find the functional equation

$$\hat{G}(\mathbf{x}\,\mathbf{0}|t\,0) - \hat{G}_0(\mathbf{x}\,\mathbf{0}|t\,0)$$
$$= \int\int d\mathbf{y}\,d\mathbf{R}_1\{n + J(\mathbf{R}_1)\}\int_0^t \hat{G}_0(\mathbf{x}\,\mathbf{y}|t - s)v(\mathbf{y} - \mathbf{R}_1)$$
$$\times\left\{1 + \frac{\delta}{\delta J(\mathbf{R}_1)}\right\}\hat{G}(\mathbf{y}\,\mathbf{0}|s0)ds. \tag{3.8}$$

In matrix form, with * denoting time convolution,

$$\hat{G} - \hat{G}_0 = (n + J_1)\hat{G}_0 v_1 *\left(1 + \frac{\delta}{\delta J_1}\right)\hat{G}. \tag{3.9}$$

We use a summation convention. An index such as 2 (standing for $\mathbf{R}_2$), that does not occur on the left hand side, is to be integrated. The Fourier transforms of time varying quantities (energy representation) are denoted by ordinary letters without carets.

$$G(E) = \int_{-\infty}^{+\infty} e^{iEt}\hat{G}(t)dt. \tag{3.10}$$

In the energy representation, the functional equation is

$$G(\mathbf{x}\,\mathbf{0}|E) - G(\mathbf{x}\,\mathbf{0}|E)$$
$$= \int\int d\mathbf{y}\,d\mathbf{R}_1\{n + J(\mathbf{R}_1)\}\left\{1 + \frac{\delta}{\delta J(\mathbf{R}_1)}\right\}$$
$$\times G_0(\mathbf{x}\mathbf{y}|E)v(\mathbf{y} - \mathbf{R}_1)G(\mathbf{y}\,\mathbf{0}|E), \tag{3.11}$$

or more succinctly,

$$G - G_0 = (n + J_1)\left\{1 + \frac{\delta}{\delta J_1}\right\}G_0 v_1 G. \tag{3.12}$$

As a first application of the functional equation we derive the density expansion of G in terms of the bare scattering matrices. Introduce

$$H(E) = G(E)\exp\left(\int J(\xi)d\xi\right). \tag{3.13}$$

The functional equation becomes

$$H - G_0\exp\left(\int J(\xi)d\xi\right) = (n + J_1)G_0 v_1\frac{\delta H}{\delta J_1}. \tag{3.14}$$

There is a hierarchy of equations. The typical one is

$$\{1 - G_0(v + ...v_m)\}\frac{\delta^m H}{\delta J_1\cdots\delta J_m} - G_0\exp\left(\int J(\xi)d\xi\right)$$
$$= (n + J_{m+1})G_0 v_{m+1}\frac{\delta^{m+1}H}{\delta J_1\cdots\delta J_{m+1}}. \tag{3.15}$$

On the right hand side, only the index $m + 1$ is summed. This yields the density expansion. For example, at $J = 0$, neglecting the third derivative

$$\frac{\delta^2 H}{\delta J_1\delta J_2} \to [1 - G_0(v_1 + v_2)]^{-1}G_0. \tag{3.16}$$

Since G and H coincide at $J=0$,

$$G\to G_0+nG_0v_1\frac{\delta H}{\delta J_1}$$
$$\to G_0+nG_0v_1\frac{1}{1-G_0v_1}G$$
$$+n^2G_0v_1(1-G_0v_1)^{-1}G_0v_2\frac{\delta^2H}{\delta J_2\delta J_1} \tag{3.17}$$

or

$$G\to G_0+nG_0t_1^0G_0+n^2t_1^0G_0v_2[1-G_0(v_2+v_1)]^{-1}G_0+\cdots. \tag{3.18}$$

Using the identity (2.20)

$$G\to G_0+nG_0t_1^0G_0$$
$$+n^2G_0t_1^0G_0t_2^0(1-G_0t_1^0G_0t_2^0)^{-1}(1+G_0t_1^0)G_0. \tag{3.19}$$

Moving the denominator to the far right

$$G=G_0+nG_0t_1^0G_0$$
$$+n^2(1+G_0t_2^0)G_0t_1^0G_0t_2^0(1-G_0t_1^0G_0t_2^0)^{-1}G_0. \tag{3.20}$$

This is the same result obtained in Sec. 2 direct expansion of the path integral.

As another illustration of the use of the functional equation we obtain the cumulant expansion of the space-time site averaged Green's function $\hat{G}(\mathbf{x}|t)$. Of course it can also be obtained directly from the path integral in the absence of a generating source. In the present problem the cumulant form

$$\hat{G}(\mathbf{x}|t)=\hat{G}_0(x|\mathbf{t})e^{F(\mathbf{x}|t)} \tag{3.21}$$

has $F(\mathbf{x}|t)$ as a series in the density

$$F(\mathbf{x}|t)=\sum_{g=1}^{\infty}n^gF_g(\mathbf{x}|t). \tag{3.22}$$

The $F_g(\mathbf{x}|t)$ are in terms of the free particle $\hat{G}_0$ and bare scattering matrices. So the series can be constructed by brute force from the direct expansion of Sec. 2. However, it is worthwhile to exhibit how the series arises from the generating functional equation so that we may regard the latter as the common starting point for a variety of approximations.

The differential form of the functional equation is

$$(i\frac{\partial}{\partial t}+\tfrac{1}{2}\nabla^2)\hat{G}(\mathbf{x}\,0|t)-\delta(t)\delta(\mathbf{x})$$
$$=\int\{n+J(\xi)\}\left\{1+\frac{\delta}{\delta J(\xi)}\right\}v(\mathbf{x}-\xi)\hat{G}(\mathbf{x}\,\mathbf{0}|t). \tag{3.23}$$

Write

$$\hat{G}(\mathbf{x}\,\mathbf{0}|t)=\hat{G}_0(\mathbf{x}\,\mathbf{0}|t)\exp\hat{L} \tag{3.24}$$

with the condition $\hat{L}(t\to 0)=0$. The delta function contribution is accounted for by $\hat{G}_0(\mathbf{x},\mathbf{0}|t)$. For $t\neq 0$

$$(i\frac{\partial}{\partial t}+\tfrac{1}{2}\nabla^2+\nabla\ln\hat{G}_0\nabla)\hat{L}+\tfrac{1}{2}(\nabla\hat{L})^2$$
$$=\int K(\xi)v(\mathbf{x}-\xi)\left\{1+\frac{\delta}{\delta K(\xi)}\right\}\hat{L}(\mathbf{x}|t)d\xi. \tag{3.25}$$

Here

$$K(\xi)=J(\xi)+n. \tag{3.26}$$

We look for a solution of the form

$$\hat{L}=\int K(\xi)Q_1(\mathbf{x}t|\xi)d\xi$$
$$+\frac{1}{2}\int\int K(\xi_1)Q_2(\mathbf{x}t|\xi_1\xi_2)K(\xi_2)d\xi_1d\xi_2+\cdots. \tag{3.27}$$

Since $K(\xi)\to n$ as $J(\xi)\to 0$, this becomes the cumulant density expansion. Matching powers of K yields a first equation

$$\left(i\frac{\partial}{\partial t}+\tfrac{1}{2}\nabla^2\ln G_0\nabla\right)Q_1(\mathbf{x}\,t|\xi)=v(\mathbf{x}-\xi)\{1+Q_1(\mathbf{x}\,t|\xi)\}. \tag{3.28}$$

With

$$\Phi_1(\mathbf{x}\,t|\xi)=\hat{G}_0(\mathbf{x}\,0|\mathbf{t})Q_1(\mathbf{x}\,t|\xi), \tag{3.29}$$

$$\left\{i\frac{\partial}{\partial t}+\tfrac{1}{2}\nabla^2-v(\mathbf{x}-\xi)\right\}\Phi_1(\mathbf{x}\,t|\xi)=v(\mathbf{x}-\xi)G_0(\mathbf{x}\,t). \tag{3.30}$$

The solution in terms of the single site bare Green's function is

$$Q_1(\mathbf{x}\,t|\xi)$$
$$=\frac{1}{\hat{G}_0(\mathbf{x}\,\mathbf{0}|t)}\int\int_0^t\hat{h}(\mathbf{x}\,\mathbf{y}|\xi|t-s)v(\mathbf{y}-\xi)\hat{G}_0(\mathbf{y}\,\mathbf{0}|s)dsd\mathbf{y}. \tag{3.31}$$

The first cumulant approximation is n times the integral of this expression over ξ. The direct path integral evaluation without a source gives the expression

$$n\frac{1}{G_0(\mathbf{x}\,\mathbf{0}|t)}\int\{\hat{h}(\mathbf{x}\,\mathbf{0}|\xi|t)-\hat{G}_0(\mathbf{x}\,\mathbf{0}|\xi|t)\}d\xi. \tag{3.32}$$

The two expressions are equal in view of the integral equation obeyed by $\hat{h}$.

The next equation comes from matching quadratic terms in K. It is

$$\left\{i\frac{\partial}{\partial t}+\frac{1}{2}(\nabla^2+\nabla\ln\hat{G}_0\cdot\nabla)-[(\mathbf{v}-\xi_1)+v(\mathbf{x}-\xi_2)]\right\}$$
$$\times Q_2(x\,t|\xi_1\xi_2)=-\nabla Q_1(\mathbf{x}\,\mathbf{0}|t|\xi_1)\nabla Q_1(\mathbf{x}\,\mathbf{0}|t|\xi_2). \tag{3.33}$$

Again this simplifies with

$$\Phi_2=\hat{G}_0(\mathbf{x}\,\mathbf{0}|t)Q_2(\mathbf{x}\,t|\xi_1\xi_2) \tag{3.34}$$

to

$$\left\{i\frac{\partial}{\partial t}+\frac{1}{2}\nabla^2-(v(x-\xi_1)+v(x-\xi_2))\right\}\Phi_2$$
$$=-\frac{1}{\hat{G}_0(\mathbf{x}\,\mathbf{0}|t)}\nabla Q_1(\mathbf{x}\,t|\xi_1)\nabla Q_1(\mathbf{x}\,t|\xi_2). \tag{3.35}$$

The solution is in terms of the two site Green's function. After multiplying by $n^2/2$ and integrating over ξ_1 and ξ_2 one has the second cumulant contribution. This shows explicitly that all of the specific path integral techniques needed have been incorporated into the functional equation. The series is similar to the improved perturbation theory of Fradkin[14] for quantum field theory that he derived with operator techniques. However, we have a density rather than a potential expansion.

For comparison with the theory of the next section in terms of the self-energy (in energy space), we note the type of truncation implied by the cumulant method. We have

$$\hat{L}=\ln\{\hat{G}(\mathbf{x}\,\mathbf{0}|t)/\hat{G}_0(\mathbf{x}\,\mathbf{0}|t)\} \tag{3.36}$$

and

$$Q_1(\mathbf{x}\,0|t\,|\xi) = \frac{\delta \ln \hat{G}(\mathbf{x}\,0|t)}{\delta J(\xi)}\bigg|_{J=0},$$

$$Q_2(\mathbf{x}\,0|t\,|\xi_1\xi_2) = \frac{\delta^2 \ln \hat{G}(\mathbf{x}\,0|t)}{\delta J(\xi_1)\delta J(\xi_2)}\bigg|_{J=0}. \tag{3.37}$$

Thus the first cumulant approximation is equivalent to the truncation

$$\hat{G}(\mathbf{x}\,0|t)\frac{\delta^2\hat{G}(\mathbf{x}\,0|t)}{\delta J(\xi_1)\delta J(\xi_2)} = \frac{\delta\hat{G}(\mathbf{x}\,0|t)}{\delta J(\xi_1)}\frac{\delta\hat{G}(x\,0|t)}{\delta J(\xi_2)}. \tag{3.38}$$

The natural truncations in the cumulant method are simple in the space-time description. This method gives a good description (including some but not all features of traps) in the limiting case of almost constant potentials. The self-energy approach of the next section leads to truncations of functional derivatives in energy momentum space and does not treat traps. The relation between the two approaches has never been adequately clarified. Perhaps the functional equation may provide a good language to explore the question.

IV. EXPANSION OF THE SELF-ENERGY

We start with the functional equation for G, i.e., Eq. (3.12), and multiply on the left by G_0^{-1} and on the right by G^{-1}. Using the definition $\Sigma = G_0^{-1} - G^{-1}$,

$$\Sigma = ((n + J_1)v_1 + (n + J_1)v_1)\frac{\delta G}{\delta J_1}G^{-1}. \tag{4.1}$$

We next use the relation $GG^{-1} = 1$ to obtain

$$\frac{\delta}{\delta J_1}(GG^{-1}) = 0, \tag{4.2}$$

$$\frac{\delta G}{\delta J_1} = -G\frac{\delta G^{-1}}{\delta J_1}G = G\frac{\delta\Sigma}{\delta J_1}G. \tag{4.3}$$

This gives the basic functional equation for Σ in terms of the exact propagator G,

$$\Sigma = v_1\{n + J_1\}\{1 + G\frac{\delta\Sigma}{\delta J_1}\}. \tag{4.4}$$

The first functional derivative of this equation yields

$$(1 - v_2G)\frac{\delta\Sigma}{\delta J_2} - v_2 = J_1v_1\frac{\delta}{\delta J_2}\left(G\frac{\delta\Sigma}{\delta J_1}\right) + nv_1\frac{\delta}{\delta J_2}\left(G\frac{\delta\Sigma}{\delta J_1}\right). \tag{4.5}$$

The zero density limit of $\delta\Sigma/\delta J_2$ at $J = 0$ is

$$\frac{\delta\Sigma}{\delta J_2}\bigg|_{J=0} = t_2. \tag{4.6}$$

Then, at $J = 0$,

$$\Sigma = nv_2(1 + Gt_2) = nt_2. \tag{4.7}$$

The second derivative yields

$$(1 - v_2G - v_3G)\frac{\delta^2\Sigma}{\delta J_3\delta J_2} - v_3\frac{\delta}{\delta J_2}G\left(\frac{\delta\Sigma}{\delta J_3}\right) - v_2\frac{\delta G}{\delta J_3}\frac{\delta\Sigma}{\delta J_2} = J_1v_1\frac{\delta^2}{\delta J_3\delta J_2}\left(G\frac{\delta\Sigma}{\delta J_1}\right) + nv_1\frac{\delta^2}{\delta J_3\delta J_2}\left(G\frac{\delta\Sigma}{\delta J_1}\right). \tag{4.8}$$

To find the self-energy to order n^2 one needs $\delta\Sigma/\delta J_1$ to order n. In turn one needs $(\delta/\delta J_2)(G\,\delta\Sigma/\delta J_1)$ in the limit of vanishing density. There are a number of possible truncations that yield integral equations that have a self-energy accurate to order n^2.

One truncation (at $J = 0$) is

$$(1 - v_2G - v_3G)\frac{\delta^2\Sigma}{\delta J_3\delta J_2} \approx v_3\frac{\delta G}{\delta J_2}\frac{\delta\Sigma}{\delta J_3} + v_2\frac{\delta G}{\delta J_3}\frac{\delta\Sigma}{\delta J_2}, \tag{4.9}$$

which is accurate to n°. Here we have neglected the entire n dependent term on the right hand side of Eq. (4.8). In terms of the quantity

$$\Phi_2 = G\frac{\delta\Sigma}{\delta J_2}, \tag{4.10}$$

the truncation is

$$\frac{\delta^2\Phi_1}{\delta J_3\delta J_2} \approx 0. \tag{4.11}$$

We then have, at $J = 0$,

$$(1 - v_2G)\frac{\delta\Sigma}{\delta J}\bigg|_{J=0} - v_2 = nv_1\Phi_2\Phi_1 + nv_1G\frac{1}{1 - v_2G - v_1G}(v_1\Phi_2\Phi_1 + v_2\Phi_1\Phi_2). \tag{4.12}$$

Inserting $\Phi_1 = Gt_1$, $\Phi_2 = Gt_2$ gives Φ_2 to order n when inserted in $\Sigma = nv_1(1 + \Phi_1)$. This leads to an expression algebraically equivalent to the result of Sec. 2. The n^2 correction to the self-energy starts as the fourth power of the atomic scattering matrix.

Another truncation is obtained by forming an equation for Φ_2,

$$(1 - Gv_2)\Phi_2 - Gv_2 = J_1Gv\frac{\delta\Phi_1}{\delta J_2} + nGv_1\frac{\delta\Phi_1}{\delta J_1}. \tag{4.13}$$

One functional derivative yields

$$(1 - Gv_2)\frac{\delta\Phi_2}{\delta J_3} - Gv_3\frac{\delta\Phi_3}{\delta J_2} - \Phi_3v_2(1 + \Phi_2) = (n + J_1)\frac{\delta}{\delta J_3}\left(Gv_1\frac{\delta\Phi_1}{\delta J_2}\right). \tag{4.14}$$

Write the equation with 3 and 2 interchanged, and eliminate $\delta\Phi_3/\delta J_2$. The result at $J = 0$ is

$$(1 - Gv_3 - Gt_3Gv_2)\frac{\delta\Phi_2}{\delta J_3} - \Phi_3Gv_2(1 + \Phi_2) + Gv_3(1 + \Phi_3) \quad n\frac{\delta}{\delta J_3}\left(Gv_1\frac{\delta\Phi_1}{\delta J_2}\right) + nGt_3\frac{\delta}{\delta J_2}\left(Gv_1\frac{\delta\Phi_1}{\delta J_3}\right). \tag{4.15}$$

Neglect of the term proportional to n on the right hand side means

$$\frac{\delta^2\Phi_1}{\delta J_3\delta J_2} + \Phi_3G\frac{\delta\Phi_1}{\delta J_2} \approx 0, \tag{4.16}$$

showing that the truncation is different from the first one. Inserting $\delta\Phi_2/\delta J_3$ into Eq. (4.13) gives a new integral equation for Φ_2.

One can also make more accurate truncations, based on retaining part of the n proportional terms. Thus

$$\frac{\delta}{\delta J_3}\left(Gv_1\frac{\delta\Phi_1}{\delta J_2}\right) = \Phi_3Gv_1\frac{\delta\Phi_1}{\delta J_2} + Gv_1\frac{\delta^2\Phi_1}{\delta J_3\delta J_2}. \tag{4.17}$$

One can retain the first term, since it still leads to a closed equation for $\delta\Phi_2/\delta J_3$. In the second term a zero density

accurate expression can be used, viz.,

$$(1 - Gv_3 - Gt_3Gv_2)\frac{\delta^2\Phi_2}{\delta J_4\delta J_3}$$

$$\approx \frac{\delta}{\delta J_4}\{\Phi_3 Gv_2(1+\Phi_2) + Gt_3\Phi_2 Gv_3(1+\Phi_3)\}. \quad (4.18)$$

Then the error in $\delta\Phi_2/\delta J_3$ is $\sim n^2$, that of Φ_2 is $\sim n^3$, and Σ is of order n^4. The price paid is a higher degree of nonlinearity in the Φ.

We have not studied any of the integral equations in detail. In particular, as emphasized by Dallacusa[15] (for the Gaussian random case), it is desirable to maintain agreement with identities such as (3.4) and (3.5). The relation to the multiple scattering theory of Ref. 3 is unclear.

V. SUMMARY

We have shown that the functional equation for the source dependent Green's function can be used as a bridge between the path integral and multiple scattering approaches to the uncorrelated impurity problem. The key point is that a path integral has been used as a starting point in which the characteristics of the random process are fully incorporated. This is a unique feature of path integrals that is not possessed by the standard multiple scattering formulation. The source is introduced so that it generates scattering matrices of various orders. Thus there is no difficulty in treating strong potentials and scattering resonances. With the standard field theory machinery we are then able to obtain equations for the complex optical potential. This has been tested by using the first nontrivial truncation beyond the coherent potential approximation to find the next term in the expansion of the optical potential in powers of the density.

In subsequent papers we exploit these results to find transport equations for the averages of higher order products of Green's functions. This is relatively straightforward. Another line of development is to use Feynman's method of trials actions for path integrals as explored by Edwards, Jones, and Lukes and by Friedberg and Luttinger so as to obtain theories that include both the Lipschitz deep traps as well as the detailed multiple scattering effects.

[1] Cf. L. Schwartz and H. Ehrenreich, Ann. Phys. (N.Y.) **64**, 100 (1971); M. Lax, in *Wave Propagation and Conductivity in Random Media* (SIAM, Am. Math. Soc., Providence, R. I., 1973).

[2] J. R. Klauder, Ann. Phys. (N.Y.) **14**, 43 (1961); J. S. Faulkner, Phys. Rev. B **1**, 934 (1970); M. Lax, in *Wave Propagation and Conductivity in Random Media* (SIAM, Am. Math. Soc., Providence, R. I., 1973).

[3] E. P. Gross, J. Stat. Phys. **25**, 585, 605 (1981); **26**, 241 (1981).

[4] L. M. Roth, Phys. Rev. B **9**, 2476 (1974).

[5] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965); L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley-Interscience, New York, 1981).

[6] S. F. Edwards and V. B. Gulyaev, Proc. Phys. Soc. **83**, 495–6 (1964); S. F. Edwards, "Electronic Structure of Amorphous Materials," in *New Developments in Semiconductors*, edited by P. R. Wallace (Noordhoff, Leyden, 1973), p. 251; M. Saitoh and S. F. Edwards, J. Phys. C **7**, 3937 (1974); R. A. Abram and S. F. Edwards, *ibid.* C **5**, 1183, 1196 (1972); K. Freed, Phys. Rev. B **5**, 4802–26 (1972).

[7] R. Jones and T. Lukes, Proc. R. Soc. London Ser. A **309**, 457–72 (1969); R. Jones, J. Phys. C **2**, 1198 (1969).

[8] R. Friedberg and J. M. Luttinger, Phys. Rev. B **12**, 4460 (1975).

[9] **B. I. Halperin and M. Lax, Phys. Rev. 148, 722 (1966); 153, 802 (1967); J.** Zittarz and J. S. Langer, *ibid.* **148**, 741 (1966); V. Samathiyakanit, J. Phys. C **7**, 2849 (1974); V. Sa-Yakanit, Phys. Rev. B **19**, 2266 (1979); E. P. Gross, J. Stat. Phys. **17**, 265 (1977).

[10] E. P. Gross, "Applications of Path Integrals," in *Stochastic Processes in Non-Equilibrium Systems, Lecture Notes in Physics No. 84*, edited by L. Garrido, P. Seglar, and P. J. Sheperd (Springer, Berlin, 1978).

[11] *Sound Transmission in a Fluctuating Ocean*, edited by S. M. Flatté (Cambridge U. P., Cambridge, 1979); J. W. Strohbehn, "Laser Propagation," *Topics in Applied Physics* (Springer, Berlin, 1978), Vol. 25; V. I. Tatarski and V. V. Zavortnyi, in *Progress in Optics*, edited by E. Wolf (North-Holland, Amsterdam, 1980), Vol. XVIII, p. 207; R. Dashen, J. Math. Phys. **20**, 894 (1979); K. Furutsu, *ibid.* **17**, 1252 (1976); **20**, 617 (1979); C. M. Rose and I. M. Besieris, J. Math. Phys. **20**, 1530 (1979); **21**, 2114 (1980).

[12] J. M. Luttinger and W. Kohn, Phys. Rev. **109**, 1892 (1956); R. L. Mills, Phys. Rev. B **8**, 3650 (1973).

[13] M. D. Donsker, "On Function Space Integrals," in *Analysis in Function Space*, edited by W. T. Martin and I. Segal (MIT, Cambridge, MA, 1964); P. L. Chow, J. Math.Phys. **13**, 1224 (1972); J. Stat. Phys. **12**, 93 (1975).

[14] E. S. Fradkin, U. Esposito, and S. Termini, Rivista del Nuovo Cimento **2**, 498 (1970).

[15] V. Dallacusa, Phys. Lett. A **47**, 81 (1974); J. Phys. C **8**, 114 (1974).

Exact solutions to the valley problem in inverse scattering

Kay R. Pechenick[a] and Jeffrey M. Cohen
Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104-3859

(Received 21 October 1981; accepted for publication 25 November 1981)

Standard approximate methods involving the Abel integral equation do not allow the ionospheric electron density to be determined in the "valley" between two electron density peaks. Here we present analytic solutions to the Gel'fand–Levitan equation, which occurs in the exact full-wave inverse scattering theory. These exact analytic solutions exhibit multiple peaks in the electron density as a function of height and provide a solution to the valley problem.

PACS numbers: 94.20. – y, 41.10.Hv, 03.40.Kf

I. INTRODUCTION

If a projectile is slid up a hill with different velocities, and a chart is made of the initial velocity versus time to return, it is possible to determine the shape of the hill from the delay time, by solving the Abel integral equation.[1] Unfortunately, no information can be obtained beyond the peak of the hill because, once the projectile passes the peak, it does not return.

Similarly, the shape of the ionosphere (electron density as a function of height above the earth's surface) below the electron density peak can be determined by reflecting electromagnetic waves from the ionosphere. When the Abel integral equation involving the reflection coefficient is solved, the approximate electron density profile is obtained.[2] To obtain information beyond the first ionospheric electron density peak, topside scattering has been used. With this method, signals from a satellite are reflected from the top of the ionosphere back to the satellite. Unfortunately, if the ionosphere has more than one electron density peak, no information can be obtained in the valley(s) between the peaks.

To obtain exact information for all heights, even beyond the first electron density peak, a full-wave method has been developed[3–6] which involves the Gel'fand–Levitan equation for a potential corresponding to the electron density. Because of difficulties in solving this equation, the full-wave method has not been extensively utilized. Recently, our generalization[7–9] of Kay's method[3–6] has allowed an exact analytic solution to the problem.[7–9] In this paper we employ our method to treat the valley problem by exactly solving the Gel'fand–Levitan equation even for potentials with multiple peaks.

II. EXACT INVERSE SCATTERING SOLUTIONS

In previous communications[7–9] we presented procedures for finding exact solutions to the Gel'fand–Levitan[6] equation in inverse scattering theory, for the case in which the reflection coefficient $r(k)$ is a rational function of the wave number k.

The reflection coefficient is assumed to satisfy the following requirements[3,5,10]:

(a) $r(0) = -1$,

(b) $[r(k)]^* = r(-k)$ for all real k,

(c) $|r(k)| \leqslant 1$ for all real k,

and (d) $r(k)$ is analytic for all k in the upper half-plane.

In Refs. 7 and 8 we assumed that

$$r(k) = N(k)/D(k), \tag{1}$$

where $N(k)$ and $D(k)$ are polynomials in k, with

$$D(k) = \prod_{i=1}^{n} (k - k_i), \tag{2}$$

where the k_i are distinct complex numbers in the lower half-plane; the k_i are the n poles of the reflection coefficient. In Refs. 7 and 8 we used our procedure to calculate the potential $V(x)$ for a variety of reflection coefficients. Here we give a brief summary of the procedure.

One first solves the following equation for the a_α ($\alpha = \pm 1, \pm 2, ..., \pm n$; $a_{-\alpha} = -a_\alpha$)

$$r(ia_\alpha)r(-ia_\alpha) = 1. \tag{3}$$

Next, for $j = 1, 2, ..., n$, and $\alpha = 1, 2, ..., n$, one defines

$$F_{j\alpha} = (a_\alpha - ik_j)^{-1} \tag{4}$$

$$G_{j\alpha} = r(-ia_\alpha)/(a_\alpha + ik_j), \tag{5}$$

and solves for the $f_\alpha(x)$ ($\alpha = 1, 2, ..., n$) in the following set of n simultaneous equations (one equation for each value of j):

$$1 + \sum_{\alpha=1}^{n} (F_{j\alpha} e^{a_\alpha x} + G_{j\alpha} e^{-a_\alpha x}) f_\alpha(x) = 0. \tag{6}$$

Then, $K(x, x)$ is given by

$$K(x, x) = \left\{ \sum_{\alpha=1}^{n} [e^{a_\alpha x} - r(-ia_\alpha) e^{-a_\alpha x}] f_\alpha(x) \right\} \theta(x), \tag{7}$$

in which $\theta(x)$ is the step function defined by $\theta(x) = 1$ for $x > 0$, $\theta(x) = 0$ for $x < 0$. Finally,

$$V(x) = 2 \frac{d}{dx} K(x, x). \tag{8}$$

The potential $V(x)$ is zero for negative x, as can be seen from Eqs. (7) and (8).

When $r(k)$ is written in the form of Eq. (1), the zeros of the numerator $N(k)$ are the zeros of $r(k)$, and, because of requirement (a), $r(k)$ is completely determined by its poles and zeros. If the zeros are denoted by l_i, $i = 1, 2, ..., m$, then

$$r(k) = \frac{C(k - l_1)\cdots(k - l_m)}{(k - k_1)\cdots(k - k_n)} \tag{9}$$

[a] Present address: General Electric Co., Space Systems Division, Valley Forge, PA.

 0022-2488/83/020406-04$02.50

$$C = \frac{-(-k_1)\cdots(-k_n)}{(-l_1)\cdots(-l_m)}. \tag{10}$$

If $r(k)$ is a rational function reflection coefficient satisfying (a), then if we multiply the poles and zeros of $r(k)$ by a common factor ξ, assumed to be a positive real number, the potential changes in a simple way. In determining this effect on the potential, we will insert new arguments of k_i and l_i into $r(k)$, a_α, and other quantities, to indicate their dependence on the poles and zeros of $r(k)$. Thus we write $r(k; k_i, l_i)$, $a_\alpha(k_i, l_i)$, ...

From Eqs. (9) and (10), we see that

$$r(\xi k; \xi k_i, \xi l_i) = r(k; k_i, l_i). \tag{11}$$

It follows from Eqs. (3) and (11) that

$$a_\alpha(\xi k_i, \xi l_i) = \xi a_\alpha(k_i, l_i). \tag{12}$$

Because of Eqs. (4), (5), and (12),

$$F_{j\alpha}(\xi k_i, \xi l_i) = \xi^{-1} F_{j\alpha}(k_i, l_i) \tag{13}$$

and

$$G_{j\alpha}(\xi k_i, \xi l_i) = \xi^{-1} G_{j\alpha}(k_i, l_i). \tag{14}$$

From Eqs. (6) and (12)–(14),

$$f_\alpha(x/\xi; \xi k_i, \xi l_i) = \xi f_\alpha(x; k_i, l_i). \tag{15}$$

Then, from Eqs. (7), (12), and (15),

$$K(x/\xi, x/\xi; \xi k_i, \xi l_i) = \xi K(x, x; k_i, l_i). \tag{16}$$

Finally, from Eqs. (8) and (16),

$$V(x/\xi; \xi k_i, \xi l_i) = \xi^2 V(x; k_i, l_i). \tag{17}$$

Thus, if the poles and zeros move farther away from the origin by a common factor of ξ, then any particular feature of the potential V, such as a maximum or minimum, moves to a value of x which is smaller by a factor ξ, and also becomes higher by a factor of ξ^2.

It is important to notice that if the original reflection coefficient satisfies the requirements (a)–(d), then the new reflection coefficient also has these properties.

To use our procedure to find $V(x)$ we choose a set of poles and zeros for $r(k)$. Because of requirement (d), the poles must be in the lower half-plane. Requirement (a) will automatically be satisfied provided that $r(k)$ has the form given by Eqs. (9) and (10). We can easily ensure that requirement (b) is satisfied by requiring that each pole is either purely imaginary or is one of a pair of poles: k_i, k_j, with $k_i^* = -k_j$, and similarly for each zero. Thus, the set of poles of $r(k)$ is symmetrical with respect to a reflection in the imaginary axis, and the set of zeros is similarly symmetrical.

It is more difficult to determine whether requirement (c) holds for a given set of poles and zeros. However, there are several methods for finding particular reflection coefficients satisfying (c). For instance, we have already pointed out that if a specific reflection coefficient is known to satisfy requirements (a)–(d), then the reflection coefficient obtained by multiplying all poles and zeros by ξ, a positive real number, will also satisfy (a)–(d).

It is easy to see that if $r(k)$ has no zeros, all poles are purely imaginary, and (a), (b), and (d) are satisfied, then (c) is always satisfied.

In addition, if we are given two reflection coefficients, $r_1(k)$ and $r_2(k)$, each satisfying (a)–(d), then the reflection coefficient $r(k) = -r_1(k)r_2(k)$ will also satisfy (a)–(d). Furthermore, if A is a positive real number with $0 \leqslant A \leqslant 1$, then

$$r(k) = Ar_1(k) + (1 - A)r_2(k) \tag{18}$$

will also satisfy (a)–(d).

If $r(k)$ has exactly two poles and no zeros and satisfies (a), (b), and (d), and if $k_2 = -k_1^*$, then requirement (c) is satisfied if and only if

$$|\mathrm{Im}\, k_1| \geqslant |\mathrm{Re}\, k_1|. \tag{19}$$

Equation (19) defines an "allowed region" for k_1.

If $r(k)$ has exactly three poles and no zeros and satisfies (a), (b), and (d), with $k_1 = -k_1^*$ and $k_3 = -k_2^*$, then for each (purely imaginary) value of k_1, there will be an allowed region for k_2. Requirement (c) is satisfied only for k_2 in the allowed region. With no loss of generality, as we have seen, we may let $k_1 = -i$; then the allowed region for k_2 consists of the region

$$y \leqslant -\tfrac{1}{2} \tag{20}$$

($y \equiv \mathrm{Im}\, k_2$), together with the region between the lemniscate of Bernoulli:

$$r^2 = 2\cos 2\theta \tag{21}$$

and the straight line

$$y = -\tfrac{1}{2}, \tag{22}$$

which is tangent to the lemniscate.

Using the above theorems, we find that the 3-pole reflection coefficient, $r_3(k)$, with the following poles (and no zeros) satisfies the necessary requirements:

$$k_1 = -0.4300i$$
$$k_{2,3} = \pm 0.7871i - 0.2203i.$$

Furthermore, the 10-pole reflection coefficient, $r_{10}(k)$, with the following poles (and no zeros) also has the required properties:

$$k_1 = -1.1083i$$
$$k_2 = -1.0100i$$
$$k_3 = -1.1900i$$
$$k_4 = -1.2513i$$

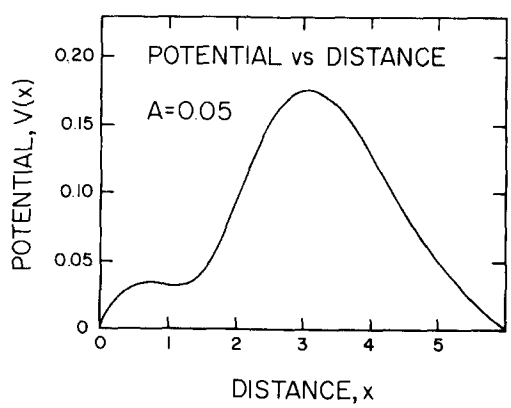

FIG. 1. Potential $V(x)$ vs distance x for $A = 0.05$.

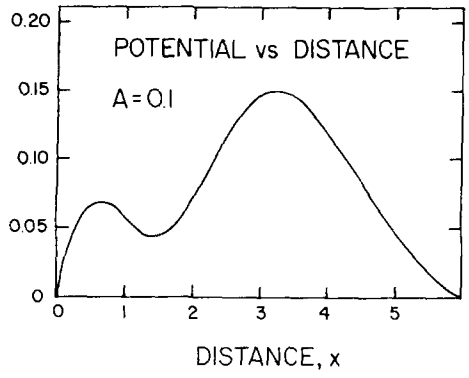

FIG. 2. Potential $V(x)$ vs distance x for $A = 0.1$.

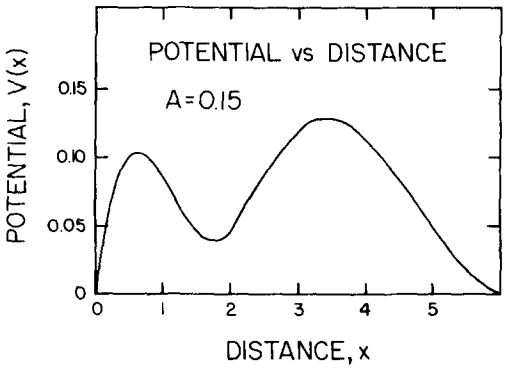

FIG. 3. Potential $V(x)$ vs distance x for $A = 0.15$.

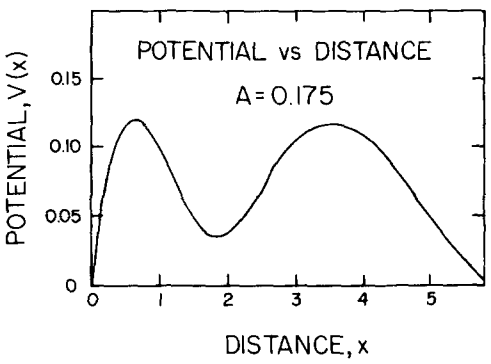

FIG. 4. Potential $V(x)$ vs distance x for $A = 0.175$.

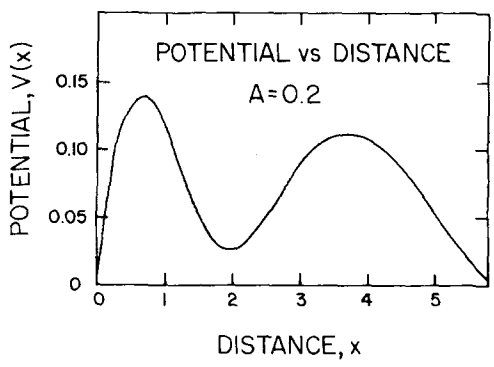

FIG. 5. Potential $V(x)$ vs distance x for $A = 0.2$.

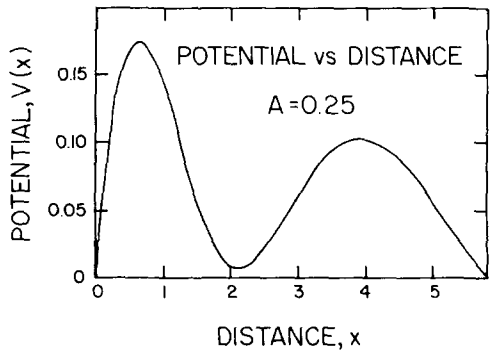

FIG. 6. Potential $V(x)$ vs distance x for $A = 0.25$.

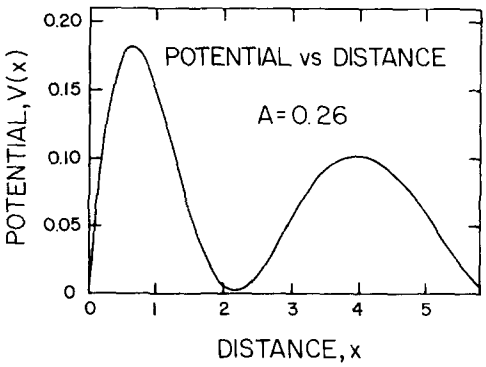

FIG. 7. Potential $V(x)$ vs distance x for $A = 0.26$.

$$k_{5,6} = \pm 0.0950 - 1.0502i$$
$$k_{7,8} = \pm 0.1504 - 1.0960i$$
$$k_{9,10} = \pm 0.1030 - 1.1510i.$$

We now take a linear combination of $r_3(k)$ and $r_{10}(k)$:

$$r(k) = Ar_3(k) + (1 - A)r_{10}(k). \tag{23}$$

The reflection coefficient defined by Eq. (23) has the required properties, provided that $0 \leqslant A \leqslant 1$. By applying our procedure[7,8] to this reflection coefficient (with several values of A), we obtain graphs of potential versus distance (Figs. 1–7) which are similar to graphs of electron density (in the ionosphere) versus height above the earth's surface. At night the ionosphere often contains an $F1$ and $F2$ layer, while in the daytime it often has an E and F layer[2]; similarly, each of our graphs has two peaks. In Fig. 1, where $A = 0.05$, the first peak is much smaller than the second. As A increases ($A = 0.1$ in Fig. 2 and $A = 0.15$ in Fig. 3), the first peak becomes higher and the second becomes lower until, for $A = 0.175$ (Fig. 4), the two peaks have nearly equal heights. As A increases further (in Figs. 5–7, $A = 0.2$, 0.25, and 0.26 respectively) the first peak continues to grow and the second continues to shrink; for $A = 0.26$, the first peak is 80 percent higher than the second. In addition, as A increases from 0.1

to 0.26, the potential at the relative minimum between the two peaks decreases; for $A = 0.26$ the potential almost goes to zero between the peaks.

III. CONCLUSIONS

In this paper we have used the exact full-wave theory of inverse scattering to treat reflection coefficients which yield multiple peaks. As can be seen from the graphs, our method allows the potential to be determined even in the valley between the peaks. Our method should prove useful in determining the ionospheric electron density even in the valley between the nighttime $F1$ and $F2$ layers, or in the valley between the daytime E and F layers.

ACKNOWLEDGMENTS

We are indebted to Dr. Angelo J. Skalafuris for his hospitality at the Mathematics Research Center, and to Professor Peter Lax for his hospitality at the Courant Institute of Mathematical Sciences.

This work was supported in part by ONR, AFOSR, and NSF.

[1]Niels Henrik Abel, *Oeuvres Complètes*, Nouvelle Édition (Grøndahl & Søn, Christiania, Norway, 1881), Tome Premier, pp. 11, 97.
[2]K. G. Budden, *Radio Waves in the Ionosphere* (Cambridge U. P., Cambridge, England, 1961).
[3]I. Kay and H. E. Moses, Nuovo Cimento **3**, 276 (1956).
[4]I. Kay and H. E. Moses, J. Appl. Phys. **27**, 1503 (1956).
[5]Irvin Kay, Commun. Pure Appl. Math. **13**, 371 (1960).
[6]I. M. Gel'fand and B. M. Levitan, Izv. Akad. Nauk SSSR, Ser. Math. **15**, 309 (1951); Am. Math. Soc. Transl. **1**, 253 (1955).
[7]Kay R. Pechenick and Jeffrey M. Cohen, Phys. Lett. A **82**, 156 (1981).
[8]Kay R. Pechenick and Jeffrey M. Cohen, J. Math. Phys. **22**, 1513 (1981).
[9]Kay R. Pechenick and Jeffrey M. Cohen, "Inverse Scattering with Coinciding-Pole Reflection Coefficients," J. Math. Phys. **24**, 115 (1983).
[10]Saeyoung Ahn and Arthur K. Jordan, IEEE Trans. Antennas Propag. **24**, 879 (1976).