Orthonormality properties of double coset matrix elements

John J. Sullivan

Department of Physics, University of New Orleans, New Orleans, Louisiana 70122
(Received 19 December 1974)

The orthogonality and completeness conditions of group representation theory are shown to provide complementary orthonormality relations for weighted double coset matrix elements. The double coset matrix elements are appropriate for symmetry adaption of the basis to two subgroup sequences using the double coset decomposition $\mathbf{H}\backslash\mathbf{G}/\mathbf{K}$. When applied to the symmetric and unitary groups, the double coset matrix elements assume the role of recoupling coefficients. The orthonormality properties give nontrivial relations between coefficients coupling tensors of different rank and/or different dimensions.

1. INTRODUCTION

In a series of papers,[1,2] hereafter referred to as I and II, we have been investigating the properties of double coset matrices in irreducible representations of a group symmetry adapted to different subgroup sequences. Double coset decomposition of a group is not only important for establishing mathematically unique labels, but is often dictated by the physical significance of the subgroups. With the shell model of identical particle systems in mind, we have directed our attention in particular to the symmetric group. Due to the intertwining of the symmetric group algebra with the algebra of the general linear group via Nth rank tensor representations, several nontrivial relations between recoupling coefficients in S_N and $GL(n)$ for different rank N and dimension n have been shown to follow from the orthogonality properites associated with these double coset matrix elements (DCME). The orthogonality properties of the DCME as used in I express the assumed (for compact or finite groups) unitarity of the matrix representation and the group orthogonality condition. The group completeness condition, although noted, was restricted to a consideration of class characters in representations symmetry adapted to identical subgroup sequences. The purpose of this paper is to show the completeness condition requires no such restriction and leads to complementary orthogonality properties that express the unitarity of the transformation between induced matrix representations known to be equivalent as shown by Mackey.[3]

Any finite (or by extension any compact) group can be decomposed into a union of disjoint double cosets with respect to any two of its subgroups as

$$\mathbf{G}=\bigcup_q \mathbf{HqK}=\bigcup_q \mathbf{Kq}^{-1}\mathbf{H}.$$

In a representation symmetry adapted to the subgroup sequences

$$G\begin{cases}\supset \mathbf{K}\supset \mathbf{q}^{-1}\mathbf{Hq}\cap\mathbf{K}\equiv {}^q\mathbf{L}\\ \supset \mathbf{H}\supset H\cap qKq^{-1}\equiv L_q\end{cases}$$

for the left (lower) and right (upper) indices, the double coset matrix elements have been shown to have the form

$$\begin{bmatrix}\lambda & & q\\ {}_i\lambda\ {}_{i'}\lambda'_{j'} m' & \lambda_j\ {}_i\lambda_j m\end{bmatrix}=\delta^{i\lambda'_{j'}\ i\lambda_j}\delta_{m'm}\begin{bmatrix}\lambda & \lambda_j\\ {}_i\lambda & {}_i\lambda_j\end{bmatrix}. \tag{1.1}$$

Here λ, ${}_i\lambda$, λ_j, ${}_i\lambda_j$ label the irreps of the corresponding groups **G**, **H**, **K**, and **L**.

The orthogonality relation over the group **G** was shown in I to require orthonormality of the form

$$\left\{\frac{hk[\lambda][{}_i\lambda_j]}{gd_q[{}_i\lambda][\lambda_j]}\right\}^{1/2}\begin{bmatrix}\lambda & \lambda_j\\ {}_i\lambda & {}_i\lambda_j\end{bmatrix} \tag{1.2}$$

when summed over q, ${}_i\lambda_j$, where $[\lambda]$ denotes the dimension of the irrep and g, h, k, and d_q are the orders of the groups **G**, **H**, **K**, and **L**, respectively. This is in addition to the orthonormality implied by the unitarity of the form Eq. (1.1) as given in Eq. (2.1).

The well-known orthogonality and completeness conditions in the representation theory of finite or compact groups requires the unitarity of the g-dimensional square matrix

$$\left(\frac{[\lambda]}{g}\right)^{1/2}\begin{bmatrix}\lambda & g_i\\ m & n\end{bmatrix}$$

with rows and columns designated by the sets $(\lambda mn, g_i)$. This is in addition to the unitarity of the $[\lambda]$-dimensional square matrix $[{}^{\lambda}_{m}\ {}^{g_i}_{n}]$ with rows and columns designated by the sets (m,n). Similarly we show the completeness condition along with the group orthogonality relation requires the unitarity of the weighted DCME in Eq. (1.2) with rows and columns designated by the sets $(\lambda, q\ {}_i\lambda_j)$. The weighted double coset matrix can be considered as the unitary transformation between representations equivalent by Mackey's theorem:

$$[\lambda_j\uparrow \mathbf{G}]\downarrow \mathbf{H}\approx\sum_q [\lambda_j\downarrow {}^q\mathbf{L}\approx \mathbf{L}_q]\uparrow \mathbf{H}.$$

The theorems and their proofs are given in Sec. 2, while Sec. 3 compares the result given here to a much earlier result by Frame.[4] To the author's knowledge this is the only other published report dealing explicitly with special properties of matrices representing double coset elements.

2. ORTHOGONALITY RELATIONS

Theorem: The double coset matrix (DCM) $[{}^{\lambda}_{i\lambda}\ {}^{\lambda_j}_{i\lambda_j}]$ is unitary with the irrep labels $({}_i\lambda, \lambda_j)$ designating the rows and columns respectively.

Proof: The theorem follows directly from the unitarity of the matrix representation of finite groups. From Eq. (1.1) we obtain

$$\sum_{{}_i\lambda\,{}_i\lambda''_j m''} \begin{bmatrix} \lambda & q \\ {}_i\lambda\,{}_i\lambda''_j m'' & \lambda_j\,{}_i\lambda_j m \end{bmatrix} \begin{bmatrix} \lambda & q \\ {}_i\lambda\,{}_i\lambda''_j m'' & \lambda'_j\,{}_i\lambda'_j m' \end{bmatrix}^*$$
$$= \delta^{{}_i\lambda_j\,{}_i\lambda_j}\delta_{m'm} \sum_{{}_i\lambda} \begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \end{bmatrix} \begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \end{bmatrix}^*$$
$$= \delta^{{}_i\lambda_j\,{}_i\lambda_j}\delta_{m'm}\delta^{\lambda'_j\lambda_j}.$$

Therefore

$$\sum_{{}_i\lambda} \begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \end{bmatrix} \begin{bmatrix} \lambda & \lambda'_j \\ {}_i\lambda & {}_i\lambda_j \end{bmatrix}^* = \delta^{\lambda_j\lambda'_j}. \quad \text{QED} \tag{2.1}$$

Theorem: The weighted DCM

$$\left\{\frac{hk[\lambda][{}_i\lambda_j]}{gd_q[{}_i\lambda][\lambda_j]}\right\}^{1/2} \begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \end{bmatrix}$$

is a square unitary matrix with the labels $(\lambda,\ q\ {}_i\lambda_j)$ designating the rows and columns respectively.

Proof: The group orthogonality relation Eq. (2.3) of I requires that

$$\sum_{{}_i\lambda_j q} \frac{hk[\lambda][{}_i\lambda_j]}{gd_q[{}_i\lambda][\lambda_j]} \begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \end{bmatrix} \begin{bmatrix} \lambda' & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \end{bmatrix}^* = \delta^{\lambda'\lambda}. \tag{2.2}$$

The group completeness condition requires the remaining orthonormality. Under the decomposition

$$\mathbf{G} \begin{cases} \nearrow \mathbf{K} \rightarrow \mathbf{q}^{-1}\mathbf{H}\mathbf{q} \cap \mathbf{K} \equiv {}^q\mathbf{L} \\ \searrow \mathbf{H} \rightarrow \mathbf{H} \cap \mathbf{q}\mathbf{K}\mathbf{q}^{-1} \equiv \mathbf{L}_q \end{cases}$$

a general element $g \in \mathbf{G}$ belongs to a unique double coset but the specific $g = hqk$ is not unique in the (h,k) pair. However, for a given choice of (left, right) coset representatives $\mathbf{H}/\mathbf{L}_q = \sigma$, ${}^q\mathbf{L}\backslash\mathbf{K} = \tau$ the specification $g = \sigma(l_q)q\tau = \sigma q({}^q l)\tau$ with $l_q = q({}^q l)q^{-1} \equiv l$ is unique. The completeness relation requires

$$\sum_{\lambda\,{}_i\lambda' m'\lambda'_j n'} \left(\frac{[\lambda]}{g}\right) \begin{bmatrix} \lambda & \sigma l q \tau \\ {}_i\lambda' m' & \lambda'_j n' \end{bmatrix} \begin{bmatrix} \lambda & \sigma' l' q' \tau' \\ {}_i\lambda' m' & \lambda'_j n' \end{bmatrix}^*$$
$$= \delta_{\sigma'\sigma}\delta_{i'i}\delta_{q'q}\delta_{\tau'\tau}.$$

Multiply both sides by

$$\begin{bmatrix} {}_i\lambda & \sigma l \\ m & {}_i\lambda_j m'' \end{bmatrix}^* \begin{bmatrix} \lambda_j & l'\tau' \\ {}_i\lambda'_j n'' & n \end{bmatrix}$$

and sum on σ, l, l', and τ'. Using orthogonality in $\mathbf{H}$ and $\mathbf{K}$, one obtains on the left

$$\sum_{\lambda} \frac{hk[\lambda]}{g[{}_i\lambda][\lambda_j]} \begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \end{bmatrix} \begin{bmatrix} \lambda_j & \tau \\ {}_i\lambda_j m'' & n \end{bmatrix} \begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda'_j \end{bmatrix}^*.$$

Using orthogonality in $\mathbf{L}$, one obtained on the right

$$\delta^{{}_i\lambda'_j\,{}_i\lambda_j}\delta_{qq'} \frac{d_q}{[{}_i\lambda_j]} \begin{bmatrix} {}_i\lambda & \sigma' \\ m & {}_i\lambda_j n' \end{bmatrix}^* \begin{bmatrix} \lambda_j & \tau \\ {}_i\lambda'_j m' & n \end{bmatrix}.$$

Since these expressions are equal for general σ' and τ, one obtains

$$\sum_{\lambda} \frac{hk[\lambda][{}_i\lambda_j]}{gd_q[{}_i\lambda][\lambda_j]} \begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda_j \end{bmatrix} \begin{bmatrix} \lambda & \lambda_j \\ {}_i\lambda & {}_i\lambda'_j \end{bmatrix}^* = \delta_{q'q}\delta^{{}_i\lambda'_j\,{}_i\lambda_j}. \tag{2.3}$$

By setting $q = q'$, ${}_i\lambda_j = {}_i\lambda'_j$, and summing Eq. (2.3) over ${}_i\lambda_j$, q and using Eq. (2.2) for a fixed ${}_i\lambda$, λ_j one obtains an equality for intertwining numbers

$$\mathcal{J}({}_i\lambda, \lambda_j \downarrow \mathbf{G}) = \sum_q \mathcal{J}({}_i\lambda, \lambda_j \downarrow \mathbf{L}),$$

showing the adjusted DCM to be square, and the theorem is proved.

It is interesting to note that various combinations of the above orthogonality relations give some well-known results in the theory of induced representations: E.g.,

A. for $\lambda = \lambda'$, $\sum_{\lambda_j}[\lambda_j] \times$ Eq. (2.2) gives

$$\sum_{q\,{}_i\lambda_j} \frac{hk[\lambda][{}_i\lambda_j]}{gd_q[{}_i\lambda]} = [\lambda];\ \text{i.e.,}\ \sum_q \frac{hk}{gd_q} = 1;$$

B. for ${}_i\lambda_j = {}_i\lambda'_j$, $\sum_{\lambda_j}[\lambda_j] \times$ Eq. (2.3) gives

$$\frac{hk[{}_i\lambda_j]}{gd_q[{}_i\lambda]} \sum_{\lambda}[\lambda] = \sum_{\lambda_j}[\lambda_j]$$

or

$$\sum_{\lambda}[\lambda] = \frac{g}{h}[{}_i\lambda] \quad \text{and} \quad \sum_{\lambda_j}[\lambda_j] = \frac{k}{d_q}[{}_i\lambda_j]$$

as required by the Frobenius reciprocity relation, and

C. for $\lambda_j = \lambda'_j$, $\sum_{\lambda_j} \times$ Eq. (2.1) gives

$$\mathcal{J}(\lambda\downarrow, {}_i\lambda_j \uparrow \mathbf{H}) = \mathcal{J}(\lambda\downarrow, {}_i\lambda_j \uparrow \mathbf{K}),$$

as required by Mackey's theorem.

3. DISCUSSION

Frame has demonstrated certain modified double coset coefficients $\rho^{ts}_{\alpha\zeta}(f^*_s/n^{ts}_\alpha)^{1/2}$ form a μ^{st} dimensional square matrix with rows and columns designated by (α, ζ). Here we wish to show that this is equivalent to Eq. (2.2) and Eq. (2.3) for the case ${}_i\lambda = {}_i\lambda^0$, $\lambda_j = \lambda^0_j$. The superscript λ^0 indicates the one-dimensional totally symmetric irrep of the corresponding group. For this purpose we give the following identification between the notation adopted by Frame and that used here:

Frame: $\mathbf{H}^t$, $\mathbf{H}^s$, α, $\zeta \equiv j\rho\sigma$, f_{ρ^*}, η^{ts}_α, $\rho^{ts}_{\alpha\zeta}$, μ^{st}.

This work: $\mathbf{H}$, $\mathbf{K}$, q, λ (and multiplicity labels—see below), $[\lambda]$,

$$\frac{g}{d_q},\ \frac{(hk)^{1/2}}{d_q} \begin{bmatrix} \lambda & \lambda^0_j \\ {}_i\lambda^0 & {}_i\lambda^0_j \end{bmatrix},\ \sum_q = \text{number of } DC.$$

With this change in notation the unitarity expressed in Frame's article is the same as stating

$$\left\{\frac{[\lambda]hk}{gd_q}\right\}^{1/2} \begin{bmatrix} \lambda & \lambda^0_j \\ {}_i\lambda^0 & {}_i\lambda^0_j \end{bmatrix}$$

is a square unitary matrix of dimension $\sum_q$ with rows and columns designated by (λ, q). The restriction to the one-dimensional totally symmetric irreps of the subgroups is due to the fact that symmetry adaption to the subgroups as used in Frame's article is determined by diagonalizing only the idempotents associated with these irreps [Eqs. (4.7) and (4.8) of Ref. 4]. Frame allows for possible multiple occurrence of these irreps in the subduction of the irrep F_j on the subgroups; i.e., $F_j \downarrow \mathbf{H}^t$ contains the totally symmetric irrep of that group μ^t_j times. The multiplicity labels thus take values $\rho \leq \mu^t_j$, $\sigma \leq \mu^s_j$ with $\sum_j \mu^t_j f_j = g/h^t$ and $\sum_j \mu^s_j f_j = g/h^s$. These multiplicity labels have been suppressed in our notation and, of course, would have to be included as discussed in I.

The DCME of the symmetric group under the decomposition $\otimes_i S_{f_i N} \backslash S_N / \otimes_j S_{Nj}$ are recoupling coefficients of the outer product and in I were shown to be identical

with the recoupling coefficients of the inner product in the unitary unimodular group. In II these same coefficients have also been shown to be DCME of the unitary group under the decomposition $\otimes_i U(_i n)\backslash U(n)/\otimes_j U(n_j)$. The double coset analysis used in II is similar to that used by Wigner[5] in giving a generalized Euler angle parametrization to certain subgroups of $GL(n)$ that contain $U(n)$. An analogous extension of the proofs in II to these groups can be made. The orthonormality relations relations reported here can be applied to these recoupling coefficients with respect to double coset decomposition of the symmetric group and the unitary group. With certain choices the sums can be limited so as to allow direct evaluation of some of the DCME. The results of such considerations will be reported in a future communication.

[1]J.J. Sullivan, J. Math. Phys. **14**, 387 (1973), referred to as I.

[2]J.J. Sullivan, J. Math. Phys. **16**, 756 (1975), referred to as II.

[3]G.W. Mackey, "The Theory of Group Representations," University of Chicago Lecture Notes 1955.

[4]J.S. Frame, Bull. Am. Math. Soc. **54**, 740 (1948); Proc. Symp. Pure Math. **6**, 89 (1962).

[5]E.P. Wigner "On a Generalization of Euler's Angles," in *Group Theory and Its Applications, Vol. I*, edited by E.M. Loebl (Academic, New York, 1969).

Generalized *P*-representation for bounded operators

Henri-François Gautrin

Département de Mathématiques, Université de Montréal, Montréal, Canada
(Received 16 April 1974; revised manuscript received 8 April 1975)

We generalize the notion of *P*-representation, introducing a space of generalized functions, such that every Hilbert–Schmidt operator has a *P*-representation. Then we extend that notion to another space of generalized function, such that every bounded operator has a *P*-representation.

I. INTRODUCTION

Let $\{|n\rangle\}$ denotes an orthonormal basis in an infinite dimension separable Hilbert space H; the coherent states $|\alpha\rangle$ in H are defined by

$$|\alpha\rangle = \exp\left(-\frac{|\alpha|^2}{2}\right)\left(\sum_{n=0}^{\infty}\frac{\alpha^n}{(n!)^{1/2}}|n\rangle\right),$$

α being a complex number.

Let A be a linear operator in H, a P-representation of A is a weak integral over the complex plane such that $A = {}_w\int P(\alpha)|\alpha\rangle\langle\alpha| d^2\alpha$, where $|\alpha\rangle\langle\alpha|$ is the projector on the one-dimensional subspace generated by $|\alpha\rangle$, $P(\alpha)$ is a complex valued function, $d^2\alpha = dx\,dy$, if $\alpha = x + iy$, and dx is the Lebesgue measure on the line.

This representation was firstly established by Sudarshan[1] for certain "density matrix" operators. Then it has been proved that every "density matrix" operator does not have a P-representation but is a uniform limit of operators having a P-representation.[2–4]

In what sense can we generalize this representation in order that every operator has a "generalized" P-representation?

Miller and Miskhin[5] gave a partial answer to that question. They considered the function $P(\alpha)$, in the P-representation, as a generalized function in Z' (the space of Fourier transform of compact support distributions) and proved that every bounded operator has a P-representation. They established their results using Fourier transform techniques; they considered the expression $\langle\phi|\alpha\rangle\langle\alpha|\psi\rangle P(\alpha)$ as the product of a function by a distribution the whole acting as a functional on Z. But in that case we do not have a weak equality of the type

$$\langle\phi, A\psi\rangle = \int \langle\phi|\rangle\langle\alpha|\psi\rangle P(\alpha)d^2\alpha;$$

indeed the function $\langle\phi|\alpha\rangle\langle\alpha|\psi\rangle$ is not in Z. And as Cahill[6] pointed out this representation is mainly related to the kernel representation of Glauber[7] and cannot be considered as a generalization of the P-representation.

In this work we consider the linear space generated by the functions $\langle\phi|\alpha\rangle\langle\alpha|\psi\rangle$; we study the possible Hilbertian topology on that space, and, using the techniques of Hilbert space with negative norm, we show that for every Hilbert–Schmidt operator ρ there exists a generalized function $P(\alpha)$ such that

$$\langle\phi, |\rho\psi\rangle = \int \langle\phi|\alpha\rangle\langle\alpha|\psi\rangle P(\alpha)d^2\alpha$$

(this integral having a symbolic sense when $P(\alpha)$ is not an ordinary function).

Then we extend these results to bounded operators using a topology of Banach type.

II. COHERENT STATES AND FISCHER SPACE

The proper framework to study the problems of P-representation is the theory of Fischer space.[8,9]

A Fischer space is a space of entire functions of a complex variable, which are square integrable with respect to the measure $d\mu(z)$ on the complex plane:

$$d\mu(z) = \frac{1}{\pi}\exp(-|z|^2)d^2z,$$

where $d^2z = dx\,dy$, and $z = x + iy$. This space $\mathcal{J}$ is a separable Hilbert space with scalar product

$$(f,g)_{\mathcal{J}} = \int f(z)g(z)\,d\mu(z)$$

and norm

$$\|f\|_{\mathcal{J}} = \left(\int |f(z)|^2 d\mu(z)\right)^{1/2};$$

from this equality we obtain the following inequality:

$$|f(z)| \leq \exp(\tfrac{1}{2}|z|^2)\|f\|_{\mathcal{J}}$$

Therefore the functional on $\mathcal{J}$ which associates with each function f its value at point Z is a bounded functional; then by the Riesz–Frechet theorem

$$\forall z\ \exists e_z \in \mathcal{J} \text{ such that } (e_z, f)_{\mathcal{J}} = f(z).$$

In particular we obtain

$$(e_z, e_{z'})_{\mathcal{J}} = \exp(\bar{z}'z).$$

Moreover, if π_z is the operator defined on $\mathcal{J}$ by

$$\pi_z f = f(z)e_z,$$

one can prove by direct calculation the weak equality

$$I = {}_w\int \pi_z\, d\mu(z).$$

Furthermore, if H is a separable Hilbert space and $|\alpha\rangle$ the coherent states in H, there exists an isometric isomorphism i between H and $\mathcal{J}$ such that

$$i[|\alpha\rangle\exp(-|\alpha|^2/2)] = e_\alpha.$$

Let A be a bounded operator in H and let A' be the corresponding operator in $\mathcal{J}$, defined by

$$\forall f \in \mathcal{J},\quad A'(f) = i\mathbf{A}i^{-1}(f).$$

Definition: A has a P-representation if there exists a complex valued function $P(z)$ such that we have the weak equality

$$A' = {}_w\int \pi_z P(z)\, d\mu(z);$$

we have in that case

$$(f, A'g)_{\mathcal{J}} = \int f(z)\overline{g(\bar{z})}P(z)\, d\mu(z).$$

III. TENSOR PRODUCT OF FISHER SPACE AND SPACE $\mathcal{G}$

Let $\mathcal{L}$ be the linear space generated by formal product $g(z')f(z)$, f and $g \in \mathcal{J}$.

Let $\mathcal{G}$ be the space of functions of one complex variable obtained from each function in $\mathcal{L}$ by restriction to the manifold $z = \bar{z}'$.

Lemma 1: $\mathcal{L}$ is isomorphic to $\mathcal{G}$.

Proof: Let $h(z', z)$ be an element in $\mathcal{L}$; $h(z', z)$ is an entire function of each complex variable z' and z. Therefore, it is an entire function of the two complex variables z' and z.

Let h and h' be two functions in $\mathcal{L}$ such that $h(\bar{z}, z) = h'(\bar{z}, z)$; then the function $h - h'$ is an entire function of z' and z which is equal to zero on the manifold $z' = \bar{z}$. By the analytic continuation theorem[10] $h - h'$ is identically equal to zero; and $h = h'$. Therefore, the application which associates with $h(z', z)$ in $\mathcal{L}$ the function $h(\bar{z}, z)$ in $\mathcal{G}$ is an injection. It is trivial that this application is also a surjection and an homomorphism.

QED

Consider now the bounded operator A in $\mathcal{J}$, and let F_A be the application which assigns to each element $\Sigma_i g_i(\bar{z})$ in $\mathcal{G}$ the complex number $\Sigma_i (g_i, A f_i)_{\mathcal{J}}$.

Corollary 1: F_A is a linear functional.

This is obvious from the construction of F_A and the isomorphism between $\mathcal{L}$ and $\mathcal{G}$.

Corollary 2: $\mathcal{G}$ can be equipped with a pre-Hilbertian topology.

Indeed we can define on $\mathcal{L}$ a pre-Hilbertian structure with scalar product

$$\left(\sum_i g_i^{(1)}(z')f_i^{(1)}(z),\ \sum_j g_j^{(2)}(z')f_j^{(2)}(z)\right)_{\mathcal{L}} = \sum_{i,j} (g_i^{(1)}(z'), g_j^{(2)}(z'))_{\mathcal{J}} (f_i^{(1)}(z), f_j^{(2)}(z))_{\mathcal{J}},$$

which according to Lemma 1 induced on $\mathcal{G}$ a pre-Hilbertian structure with norm

$$\left\|\sum_i g_i(\bar{z})f_i(z)\right\|^2_{\mathcal{G}} = \sum_i \left[\int |g_i(z)|^2 d\mu(z)\right]\left[\int |f_i(z)|^2 d\mu(z)\right].$$

Let $\bar{\mathcal{G}}_1$ be the Hilbert space obtained from $\mathcal{G}$ by completion with respect to the norm $\|\cdot\|_{\mathcal{G}_1}$.

Lemma 2: The functions in $\mathcal{G}$ are square integrable with respect to the measure

$$d\mu'(z) = (1/\pi)\exp(-|z|^2)\, d\mu(z).$$

Proof: The functions in $\mathcal{G}$ are linear combination of product $f(z)g(\bar{z})$, where f and g are in $\mathcal{J}$; therefore, they are continuous functions of z. Furthermore, the functions $(1/\pi)[\Sigma_i f_i(z)g_i(\bar{z})]\exp(-|z|^2)$ are integrables with respect to the measure d^2z, and then square integrables. Therefore, the functions in $\mathcal{G}$ are square integrable with respect to the measure

$$d\mu'(z) = (1/\pi)\exp(-|z|^2)\, d\mu(z).$$

QED

Corollary[3]: $\mathcal{G}$ can be equipped with another pre-Hilbertian topology with norm

$$\|f(z)g(\bar{z})\|^2_{\mathcal{G}_2} = \int |f(z)|^2 |g(z)|^2 d\mu'(z).$$

Let $\bar{\mathcal{G}}_1$ and $\bar{\mathcal{G}}_2$ be the Hilbert space obtained respectively by completion of $\mathcal{G}$ in the norm $\|\cdot\|_{\mathcal{G}_1}$ and $\|\cdot\|_{\mathcal{G}_2}$.

IV. P-REPRESENTATION OF HILBERT SCHMIDT OPERATOR

Lemma 3: Let A be a Hilbert–Schmidt operator in $\mathcal{J}$; then the linear functional F_A is a bounded linear functional in $\mathcal{G}_1$.

Proof: Let $\{e_i(z)\}$ be an orthonormal basis in $\mathcal{J}$; we have

$$\sum_{i,j} |(e_i(z), Ae_j(z))_{\mathcal{J}}|^2 = C < \infty,$$

and we have also for every element $\Sigma_{i,j} a_{ij} e_i(z) \otimes e_j(z')$ in $\mathcal{J} \otimes \mathcal{J}$

$$\|\sum_{i,j} a_{ij} e_i(z) \otimes e_j(z')\|^2_{\mathcal{J}\otimes\mathcal{J}} = \sum_{i,j} |a_{ij}|^2.$$

Then

$$\|\sum_{i,j} a_{ij} e_i(z) e_j(\bar{z})\|^2_{\mathcal{G}_1} = \sum_{i,j} |a_{ij}|^2$$

and

$$\left|F_A\left(\sum_{i,j} a_{ij} e_i(z) e_j(\bar{z})\right)\right|^2 = \left|\sum_{i,j} a_{ij} (e_i(z), Ae_j(z))_{\mathcal{J}}\right|^2$$
$$\leq \sum_{i,j} |a_{ij}|^2 \sum_{i,j} |(e_i(z), Ae_j(z))_{\mathcal{J}}|^2$$
$$\leq C\left\|\sum_{i,j} a_{ij} e_i(z) e_j(\bar{z})\right\|^2_{\mathcal{G}_1}$$

QED

It is worth to note that this lemma induce a kernel representation of A. Indeed in that case $\exists K(z, z') \in \mathcal{J} \otimes \mathcal{J}$ such that

$$(f, Ag) = \int\int K(z, z') f(\bar{z}) g(z')\, d\mu(z)\, d\mu(z').$$

In order to have a P-representation we need, in a certain sense, F_A to be bounded in $\bar{\mathcal{G}}_2$. But, unfortunately, this is not always the case. To overcome this difficulty, we shall build on $\mathcal{G}_2$ a "Hilbert space with negative norm" equipment.

V. GENERALIZED P-REPRESENTATION FOR HILBERT-SCHMIDT OPERATORS

Let us define on $\mathcal{G}$ a third norm

$$\|\cdot\|_{\mathcal{G}_3} = \|\cdot\|_{\mathcal{G}_1} + \|\cdot\|_{\mathcal{G}_2}.$$

$\|\cdot\|_{\mathcal{G}_3}$ is again a Hilbertian norm and $\|\cdot\|_{\mathcal{G}_3} > \|\cdot\|_{\mathcal{G}_2}$. Let $\bar{\mathcal{G}}_3$ be the completion of $\mathcal{G}$ with respect to the norm $\|\cdot\|_{\mathcal{G}_3}$; then $\bar{\mathcal{G}}_3 \subset \bar{\mathcal{G}}_2$ and $\bar{\mathcal{G}}_3$ is dense in $\bar{\mathcal{G}}_2$.

Following Berezanskii,[11] we shall consider $\bar{\mathcal{G}}_3$ as a

Hilbert space with positive norm, and $\bar{\mathcal{G}}_2$ as a Hilbert space with zero norm, and then construct the corresponding Hilbert space with negative norm. A vector g in $\bar{\mathcal{G}}_2$ generates a linear functional l_g on $\bar{\mathcal{G}}_3$ such that

$$\forall\ f\in\bar{\mathcal{G}}_3\ l_g(f)=(g,f)_{\mathcal{G}_2}.$$

This functional is bounded; indeed

$$|l_g(f)|\le\|g\|_{\mathcal{G}_2}\|f\|_{\mathcal{G}_2}\le\|g\|_{\mathcal{G}_2}\|f\|_{\mathcal{G}_3}.$$

Consider the norm

$$\|g\|_{\mathcal{G}_-}=\|l_g\|=\sup_{f\in\bar{\mathcal{G}}_3}[|(g,f)_{\mathcal{G}_2}|/\|f\|_{\mathcal{G}_3}];$$

we have $\|g\|_{\mathcal{G}_-}<\|g\|_{\mathcal{G}_2}$.

Definition: Let $\mathcal{G}_-$ be the completion of $\mathcal{G}$ with respect to the norm $\|\cdot\|_{\mathcal{G}_-}$.

Lemma 4: There is a "scalar product" $[\alpha,f]_{\mathcal{G}_2}$ defined for $\alpha\in\mathcal{G}_-$ and $f\in\bar{\mathcal{G}}_3$ which coincides with the scalar product in $\mathcal{G}_2$ for $\alpha\in\bar{\mathcal{G}}_2$; such that

$$|[\alpha,f]_{\mathcal{G}_2}|\le\|\alpha\|_{\mathcal{G}_-}\|f\|_{\mathcal{G}_3}.$$

Furthermore if l is a bounded functional on $\bar{\mathcal{G}}_3$, there exists a vector $\alpha_l\in\mathcal{G}_-$ such that

$$l(f)=[\alpha_l,f]_{\mathcal{G}_2}\quad\forall f\in\bar{\mathcal{G}}_3.$$

The proof of this lemma can be found in Berezanskii.[11]

Definition: A bounded operator A in a Hilbert space $\mathcal{H}$ is said to have a generalized P-representation if there exists a generalized function $P(z)$ in $\mathcal{G}_-$ such that we have the weak equality

$$(f,A'g)=[P,fg]_0=\int f(\bar{z})g(z)\exp(-|z|^2)P(z)\,d\mu(z),$$

this integral being considered in a generalized sense if $P(z)$ is not in $\bar{\mathcal{G}}_2$. P is called the generator of the P-representation.

Theorem 1: Every Hilbert–Schmidt operator in a Hilbert space has a generalized P-representation.

Proof: Let A be a Hilbert–Schmidt operator in $\mathcal{H}$ and let A' be the corresponding operator in the Fischer space $\mathcal{J}$; by Lemma 3 the linear functional $F_{A'}$ is a bounded linear functional on $\bar{\mathcal{G}}_1$.

Therefore, $F_{A'}$ is a bounded linear functional on $\bar{\mathcal{G}}_3$. Then by Lemma 4 there exists a generalized function P in $\mathcal{G}_-$ such that

$$F_{A'}\Big(\sum_{i,j}a_ie_i(z)e_j(\bar{z})\Big)=\Big[P,\sum_{i,j}a_ie_i(z)e_j(\bar{z})\Big]_{\mathcal{G}_2},$$

and in particular we have

$$(f,A'g)=[P,fg]_{\mathcal{G}_2}$$

and, if $P\in\bar{\mathcal{G}}_2$, we have

$$(f,Ag)_{\mathcal{J}}=\frac{1}{\pi}\int f(\bar{z})g(z)P(z)\exp(-|z|^2)\,d\mu(z).\qquad\text{QED}$$

Theorem 2: To every generalized function P in $\mathcal{G}_-$ there corresponds a bounded operator A such that A has a generalized P-representation with generator P.

Proof: Let P be an element in $\mathcal{G}_-$; P is a bounded linear functional on $\bar{\mathcal{G}}_3$; there exists a bilinear function B on $\mathcal{H}$ such that

$$[P,f(\bar{z})g(z)]_0=B(f,g).$$

Let us show that this bilinear functional is continuous in each variable.

Let $\{g_n\}$ be a sequence in $\mathcal{H}$ such that

$$\lim_{n\to\infty}\|g_n-g\|_{\mathcal{H}}=0;$$

we have

$$|B(f,g_n-g)|=[P,f(\bar{z})(g_n(z)-g(z))]_0.$$

But

$$\begin{aligned}|[P,f(\bar{z})(g_n(z)-g(z))]_0|&\le\|f(g_n-g)\|_{\mathcal{G}_3}\\&\le(\int|f|^2d\mu)^{1/2}(\int|g-g_n|^2d\mu)^{1/2}\\&+[(1/\pi)\int|f|^2|g-g_n|^2\exp(-|z|^2)d\mu]^{1/2}\end{aligned}$$

The function $|f|^2\exp(-|z|^2)$ is a continuous function of z which is bounded. Let

$$M=\sup[|f|^2\exp(-|z|^2)];$$

we have

$$|[P,f(\bar{z})(g_n(z)-g(z))]_0|\le(\|f\|+M)\|g-g_n\|.$$

Then the bilinear functional is continuous in each variable. Therefore, $\exists$ a bounded operator A such that

$$B(f,g)=(f,Ag).$$

This operator has a generalized P-representation with generator P. QED

We obtain also, as a direct consequence a result of Klauder–Rocca type.[3,4]

Theorem 3: Every Hilbert–Schmidt operator is the uniform limit of operators having a P representation.

Proof: Let A be a Hilbert–Schmidt operator with P as a generator of its P-representation. P is in $\mathcal{G}_-$, and $\mathcal{G}_2$ is dense in $\mathcal{G}_-$; then there exists a sequence P_i of elements in $\mathcal{G}_2$, which converges, in $\|\cdot\|_{\mathcal{G}_-}$ norm, to P. Let the A_i be the bounded operators, which have a P-representation, with P_i as a generator (it is a P-representation, and not a generalized P-representation, because P_i is in $\bar{\mathcal{G}}_2$).

Let us show that $\|A-A_i\|$ converges to 0. We have

$$\begin{aligned}\|A-A_i\|&=\sup_{\|f\|_{\mathcal{H}}=1\|g\|_{\mathcal{H}}=1}|(f,(A-A_i)g)_{\mathcal{H}}|\\&=\sup_{\|f\|_{\mathcal{J}}=1\|g\|_{\mathcal{J}}=1}|[P-P_i,fg]_{\bar{\mathcal{G}}_2}|\\&\le\sup_{\|f\|_{\mathcal{J}}=1\|g\|_{\mathcal{J}}=1}(\|P-P_i\|_{\mathcal{G}_-}\|fg\|_{\mathcal{G}_3})\\&\le\|P-P_i\|_{\mathcal{G}_-}\sup_{\|f\|_{\mathcal{J}}=1\|g\|_{\mathcal{J}}=1}\|fg\|_{\mathcal{G}_3},\end{aligned}$$

but

$$\begin{aligned}\|fg\|_{\mathcal{G}_3}&=(\int|f|^2d\mu)^{1/2}(\int|g|^2d\mu)^{1/2}\\&+[(1/\pi)\int|f|^2|g|^2\exp(-|z|^2)d\mu]^{1/2}\\&=\|f\|_{\mathcal{J}}\|g\|_{\mathcal{J}}+[(1/\pi)\int|f|^2|g|^2\exp(-|z|^2)d\mu]^{1/2}.\end{aligned}$$

We have

$$|f(z)| \leq \exp(|z|^2/2)\|f\|_{\mathcal{J}}.$$

Therefore,

$$|f(z)|^2 \leq \exp(|z|^2)\|f\|^2_{\mathcal{J}}$$

and

$$[(1/\pi)\int |f|^2|g|^2\exp(-|z|^2)d\mu]^{1/2} \leq [(\|f\|^2_{\mathcal{J}}/\pi\int |g|^2 d\mu]^{1/2} \leq \|f\|_{\mathcal{J}}\|g\|_{\mathcal{J}}/\sqrt{\pi}$$

so that

$$\sup_{\|f\|_{\mathcal{J}}=1\|g\|_{\mathcal{J}}=1} \|fg\|_{\mathcal{G}_3} \leq \sup_{\|f\|_{\mathcal{J}}=1,\|g\|_{\mathcal{J}}=1} (\|f\|\,\|g\|+\|f\|\,\|g\|/\sqrt{\pi}) = 1 + 1/\sqrt{\pi}$$

and

$$\|A - A_t\| \leq (1 + 1/\sqrt{\pi})\|P - P_t\|_{\mathcal{G}_-}. \qquad \text{QED}$$

VI. GENERALIZED P-REPRESENTATION FOR BOUNDED OPERATORS

In order to construct the P-representation for bounded operators, we have to introduce a new topology on the tensor product of $\mathcal{J}$ space.

Let us define the norm on $\mathcal{J}\otimes\mathcal{J}$,

$$\left\|\sum_i f_i\otimes g_i\right\| = \inf_{(\Sigma f_i\otimes g_i = \Sigma f'_i\otimes g'_i)} \sum \|f_i\|_{\mathcal{J}}\|g_i\|_{\mathcal{J}},$$

and let $\|\cdot\|_{\mathcal{G}_4}$ be the corresponding norm on $\mathcal{G}$. Then the completion $\mathcal{G}_4$ of $\mathcal{G}$ with respect to that norm is a Banach space.

Lemma 5: Let A be a bounded operator in $\mathcal{J}$; then the linear functional F_A is a bounded linear functional in $\mathcal{G}_4$.

Proof:

$$[F_A(\Sigma f_i\bar{g}_i)] = |\Sigma(f_i, Ag_i)| \leq \|A\|\|\Sigma\|f_i\|_{\mathcal{J}}\|g_i\|_{\mathcal{J}},$$

but, if

$$\Sigma f_i\otimes g_i = \Sigma f'\otimes g'_i,$$

then

$$F_A(\Sigma f_i\bar{g}_i) = F_A(\Sigma f'_i\bar{g}_i)$$

and

$$F_A(\Sigma f_i\bar{g}_i) \leq \|A\| \inf_{\Sigma f_i\otimes g_i = \Sigma f'_i\otimes g'_i} \Sigma \|f_i\|\,\|g_i\|.$$

Let us now define an equipment of the Hilbert space $\mathcal{G}_2$; let us introduce the norm on $\mathcal{G}$

$$\|\cdot\|_5 = \|\cdot\|_2 + \|\cdot\|_4.$$

Let $\bar{\mathcal{G}}_5$ be the completion of $\mathcal{G}$ with respect to that norm then. $\bar{\mathcal{G}}_5$ is a Banach space densely included in $\bar{\mathcal{G}}_2$. Let $\mathcal{G}'$ be the dual of $\bar{\mathcal{G}}_5$. $\bar{\mathcal{G}}_2$ can be densely included in $\mathcal{G}'$.

Theorem 4: Every bounded operator A in a Hilbert space H has a generalized P representation the generator of which is in $\mathcal{G}'$.

Proof: By Lemma 5 the corresponding linear functional F_A is bounded on $\bar{\mathcal{G}}_5$. Therefore, there exists a generalized vector P in $\mathcal{G}'$ such that

$$F_A(\Sigma f_i g_i) = [P, \Sigma \bar{f}_i g_i]$$

In particular, when P is in $\mathcal{G}_2$, we have

$$F_A(\Sigma \bar{f}_i g_i) = (P, \Sigma \bar{f}_i g_i)_{\mathcal{G}_2}$$

and

$$(f, Ag) = \int P(z)f(\bar{z})g(z)\exp(-|z|^2)\,d\mu(z).$$

Theorem 5: To every generalized function P in $\mathcal{G}'$ there corresponds a bounded operator A such that A has a generalized P-representation.

Proof: The proof is the same as the one for Theorem 2; indeed for functions in $\mathcal{G}$ of the type $f(\bar{z})g(z)$ the norm $\|\cdot\|_3$ is equal to the norm $\|\cdot\|_5$. QED

For the same reason, taking into account the proof of Theorem 3, we obtain a theorem of Klauder–Rocca type for bounded operators:

Theorem 6: Every bounded operator is the uniform limit of operators having a P-representation.

[1] E.C.G. Sudarshan, Phys. Rev. Lett. **10**, 277 (1963).
[2] J.R. Klauder, J.M. McKenna, and D. Currie, J. Math. Phys. **6**, 734 (1965).
[3] F. Rocca, C.R. Acad. Sci. (Paris) A **262**, 547 (1966).
[4] J.R. Klauder, Phys. Rev. Lett. **16**, 534 (1966).
[5] N.M. Miller and E.A. Miskhin, Phys. Rev. **164**, 1610 (1967).
[6] K.E. Cahill, Phys. Rev. **180**, 1239, 1244 (1969).
[7] R.J. Glauber, Phys. Rev. **131**, 2766 (1963).
[8] D.J. Newman and H.S. Shapiro, in *Proceedings of the Symposium in Pure Mathematics*, Vol. 11 (Am. Math. Soc., Providence, R.I., 1966), p. 360.
[9] V. Bargmann, Comm. Pure and Appl. Math. **14**, 187 (1961).
[10] V. Wladimorov, *Methods of the Theory of Functions of Many Complex Variables* (M.I.T. Press, Cambridge, Mass., 1963).
[11] Ju. M. Berezanskii, *Expansion in Eigenfunctions of Self-Adjoint Operators* (Translation of Mathematical Monograph, Vol. 17 (Am. Math. Soc., Providence, R.I., 1968).

Jacobi's principle for the case of time-varying Hamiltonian

Robert G. Cawley

Naval Surface Weapons Center, White Oak, Silver Spring, Maryland 20910
(Received 10 March 1975)

We present a generalization of Jacobi's principle to the case where the particle Hamiltonian varies with time, illustrating the formalism with an example from particle mechanics. It can be applied also to Stueckelberg mechanics to the case of variable mass, and we give an example.

1. INTRODUCTION

In particle mechanics, the equations of motion can be derived from Hamilton's principle that the first order, fixed end point variation δA, of the action

$$A=\int_{t_1}^{t_2} dt\, L(q,\dot{q},t), \tag{1.1}$$

vanish. Here q is a compact notation for the N generalized coordinates $q_1,\ldots,q_N$, $\dot{q}$ means the time (t) derivative, and L is the system's Lagrangian. The variation of A is produced by variations δq and $\delta\dot{q}$, of q and $\dot{q}$, with $\delta\dot{q}$ defined by $[d/dt,\delta]=0$. Jacobi's principle, which is a variation condition giving equations of the system trajectory, is arrived at by first going over to the extended configuration space in which the time is treated as a coordinate and, in the case where the time is ignorable, eliminating it from the formalism by Routh's method. That t is ignorable means that its conjugate momentum, p_t, is constant and hence, by the identity,

$$p_t+H\equiv 0, \tag{1.2}$$

where H is the Hamiltonian derived from L, that the Hamiltonian is constant. So Jacobi's principle applies only to the case in which H is constant in time; and the system trajectories are labeled by this constant value,

$$H=E. \tag{1.3}$$

In the present paper we extend Routh's "ignoration" formalism for ignorable coordinates to the case of a nonignorable time coordinate. The main points of the paper are the realization that the derivation of Jacobi's principle works for the time dependent case and an explication of the meaning of this generalization.

In Sec. 2 we produce the main result of the paper, in Sec. 3 we give an illustrative example from classical mechanics, in Sec. 4 we give an example from Stueckelberg mechanics, and in Sec. 5 we conclude the paper.

2. EXTENSION OF JACOBI'S PRINCIPLE

In the extended configuration space q and t are functions of a parameter τ and the action is

$$A=\int_{\tau_1}^{\tau_2} d\tau\, \bar{L}(q,t;q',t'), \tag{2.1}$$

where primes denote differention by τ,

$$\bar{L}(q,t;q',t')=t'L(q,q'/t',t), \tag{2.2}$$

and where we assume $t'>0$, as this entails no loss of generality. The fixed end point first order variation condition,

$$\delta' A=0, \tag{2.3}$$

where $\delta' q'$ and $\delta' t'$ are defined from $\delta' q$ and $\delta' t$ by $[\delta', d/d\tau]=0$, produces equations of motion of q and t equivalent to those stemming from $\delta A=0$; the additional equation stemming from (2.1), for the time, is realized in the equations deriving from (1.1) as the t-derivative of Eq. (1.2).

We assume first that t is ignorable, then

$$\frac{\partial \bar{L}}{\partial t}=0 \tag{2.4}$$

so that

$$\frac{d}{d\tau}\left(\frac{\partial \bar{L}}{\partial t'}\right)=p_t'=0, \tag{2.5}$$

and p_t is a constant, given by Eqs. (1.2) and (1.3) as $(-E)$. To derive Jacobi's principle, one eliminates the ignorable coordinate t as follows: (i) Replace L with

$$L_J(q,q')=\bar{L}(q,q',t')-t'p_t, \tag{2.6}$$

where t' is regarded as the function of q,q'

$$t'=t'(q,q';p_t), \tag{2.7}$$

determined as the solution to the defining equation of p_t,

$$F(q,q',t';p_t)=p_t-\frac{\partial}{\partial t'}\bar{L}(q,q',t')=0; \tag{2.8}$$

(ii) replace A with

$$A_J=\int_{\tau_1}^{\tau_2} d\tau L_J(q,q'); \tag{2.9}$$

(iii) replace Eq. (2.3) with

$$\delta'' A_J=0, \tag{2.10}$$

where the δ'' variation is produced in Eq. (2.9) by independent fixed end point variations[1] of the $q(\tau)$, with $[\delta'', d/d\tau]=0$ giving $\delta'' q'(\tau)$. To prove Eq. (2.10), one normally invokes the following argument: Evaluate $\delta'' A$ by regarding q and t as intermediate coordinates in Eq. (2.9), using Eq. (2.3), which holds for *arbitrary* (fixed end point) infinitesimal variations of q and t and hence those restricted by Eq. (2.8) and by $[\delta'', d/d\tau]t(\tau)=0$,[2] to get only an end point contribution,

$$\delta'' A=\frac{\partial \bar{L}}{\partial t'}\delta'' t\Big|_{\tau_1}^{\tau_2}=p_t\delta'' t\Big|_{\tau_1}^{\tau_2}. \tag{2.11}$$

The right-hand side of Eq. (2.11) need not vanish because Eq. (2.7) or Eq. (2.8) may not be consistent with the vanishing of $\delta''t$ at τ_1 and τ_2. Equation (2.10) follows at once from Eq. (2.11) if p_t is independent of t, for then we can write

$$p_t \delta'' t|_{\tau_1}^{\tau_2} = \delta'' \int_{\tau_1}^{\tau_2} d\tau\, t' p_t. \tag{2.12}$$

We suppose now that p_t is not a constant but is allowed to possess t dependence,

$$p_t = p_t(t). \tag{2.13}$$

We assume that Eq. (2.13) is available to us in Eq. (2.8), which now becomes

$$F(q, q', t, t', p_t(t)) = p_t(t) - \frac{\partial}{\partial t'} \bar{L}(q, q', t, t') = 0, \tag{2.14}$$

where L now must have explicit t dependence; for, from Eq. (2.8), which defines p_t,

$$p_t = L - \dot{q} \frac{\partial L}{\partial \dot{q}}, \tag{2.15}$$

so that, using the equations of motion for the $q(t)$, we have

$$\frac{dp_t}{dt} = \frac{\partial L(q, q'/t', t)}{\partial t}, \tag{2.16}$$

nonvanishing by Eq. (2.13), unless p_t is constant. Solving Eq. (2.14) for t', we get the analog of Eq. (2.7),

$$t' = t'(q, q', t, p_t(t)) = t'(q, q', t). \tag{2.17}$$

We now regard L_J as having explicit t dependence from Eq. (2.6) via that of L, and Eqs. (2.13) and (2.17). Reasoning exactly as before leads again to Eq. (2.11). We rewrite the new version of that equation to get

$$\begin{aligned} \delta'' A &= \int_{\tau_1}^{\tau_2} d\tau (p_t' \delta'' t + p_t (\delta'' t)') \\ &= \int_{\tau_1}^{\tau_2} d\tau\, \delta''(t' p_t) + \int_{\tau_1}^{\tau_2} d\tau (p_t' \delta'' t - t' \delta'' p_t). \end{aligned} \tag{2.18}$$

We use Eq. (2.13) to evaluate $\delta'' p_t$

$$\delta'' p_t = \delta'' t \frac{dp_t}{dt}, \tag{2.19}$$

so that the second integral in Eq. (2.18) vanishes and Eq. (2.10) follows by transposing the first term from the right side of Eq. (2.18)—but with a new meaning:

$$\delta'' A_J = 0, \tag{2.20}$$

where now

$$A_J = \int_{\tau_1}^{\tau_2} d\tau L_J(q, q', t). \tag{2.21}$$

This needs some explaining because explicit t-dependence still remains in Eq. (2.21)! Equation (2.17) is a differential equation for $t(\tau)$ when the $q(\tau)$ are given, and, as it is satisfied by the varied q in Eq. (2.20), variations of the q lead to changes, $\delta''t$, of t. The contribution to $\delta'' A_J$ from this effect is

$$\int_{\tau_1}^{\tau_2} d\tau \frac{\partial L_J(q, q', t)}{\partial t} \delta'' t. \tag{2.22}$$

But by Eqs. (2.6), (2.13), (2.14), and (implicitly) (2.17),

$$\begin{aligned} &\frac{\partial L_J(q, q', t)}{\partial t} \\ &= \frac{\partial \bar{L}(q, q', t, t')}{\partial t} + \left(\frac{\partial \bar{L}(q, q', t, t')}{\partial t'} - p_t(t) \right) \frac{\partial t'}{\partial t} \\ &\quad - t' \frac{dp_t(t)}{dt} \end{aligned} \tag{2.23}$$

$$= t' \left(\frac{\partial L(q, q'/t', t)}{\partial t} - \frac{dp_t(t)}{dt} \right). \tag{2.24}$$

The N equations of the $q(\tau)$ issuing formally from Eq. (2.20) when the contribution from Eq. (2.22) is ignored involve the additional unknown function $t(\tau)$. When these represent the trajectories for the motions stemming from $\delta A = 0$, where A is given by Eq. (1.1), Eq. (2.16) provides the condition that selects the allowed class of solutions from what results when $t(\tau)$ is assumed arbitrary. To maintain the connection between the two formalisms, we assume Eq. (2.16) in Eq. (2.24), and this makes the δ'' variation of the action caused by induced changes in t, namely Eq. (2.22), zero. So in computing $\delta'' A_J$ in Eq. (2.20), we may regard the t dependence as parametric and treat it passively.

With this interpretation of Eqs. (2.20) and (2.21) we find N Euler—Lagrange equations of the q, which involve the additional unknown $t(\tau)$. In fact, $p_t(t)$ also is unknown in general as this comes from the solution to the Euler—Lagrange equations deriving from Eq. (1.1)! But two additional equations already are available to us to close the system and specify the problem for the trajectory.

The first is the "*interpretation*" of the function $p_t(t)$ (or, inversely, of t) provided by Eq. (2.15), which follows from its defining equation (2.8). To secure this, we assume that the function $L(q, \dot{q}, t)$ is given. [3] The second equation is implicit in the equations of the q issuing from Eq. (2.20) and is associated with the arbitrariness of τ. We can obtain it as follows. From Eqs. (2.2), (2.6), and (2.14) it follows that $L_J(q, q', t)$ is homogeneous of the first degree in the q' as long as Eq. (2.14) can be solved for t'. By Euler's theorem on homogeneous functions, the Hamiltonian resulting from the usual Legendre transformation of L_J vanishes identically, corresponding to which there must exist a relation (constraint) among the variables of the canonical formalism, [4]

$$K(q, p; t, p_t) \approx 0, \tag{2.25}$$

where the wavy equals denotes weak equality in the sense of Dirac. [5] Here K is the total Hamiltonian on the space of the q's and p's, and p is given as

$$p \approx \partial L_J(q, q', t)/\partial q'. \tag{2.26}$$

The consistency condition on the constraint [Eq. (2.25)] is

$$\frac{dK}{d\tau} \approx 0. \tag{2.27}$$

Evaluating the left side, we have

$$\left(\frac{\partial K}{\partial q}q' + \frac{\partial K}{\partial p}p'\right) + \left(\frac{\partial K}{\partial p_t}\frac{dp_t}{dt} + \frac{\partial K}{\partial t}\right)t' \approx 0. \tag{2.28}$$

Since the quantity in the first set of parentheses vanishes by Hamilton's equations, and since $t' \neq 0$, we have

$$\frac{dp_t}{dt} = -\frac{\partial K/\partial t}{\partial K/\partial p_t}, \tag{2.29}$$

and if we substitute Eq. (2.26) into the right-hand side of the last equation, we obtain an equation of the form

$$\frac{dp_t}{dt} = \overline{K}(q, q'; t, p_t). \tag{2.30}$$

Equation (2.30) is distinct from Eq. (2.15) because the equations of the $q(\tau)$ (actually the canonical equations, which are equivalent to these) were needed to derive it. Equation (2.30) has to be reducible to Eq. (2.16), although we do not display this explicitly.

To summarize: The generalized Jacobi principle is embodied in Eqs. (2.20) and (2.21), where the δ'' variation is produced by independent fixed endpoint variations of the q's with t treated passively. Two additional equations are available to close the system resulting from this, the interpretation of $p_t(t)$, Eq. (2.15), provided by L and assumed to be given, plus a consistency equation to the equations of the q's, i.e., Eq. (2.30), which follows from these, and which must be reducible to Eq. (2.16). The complete system gives the function $p_t(t)$, or by Eqs. (1.2) and (1.3), $E(t)$, which we may regard as classifying the trajectories.

It is worthwhile to examine the way these equations are solved, especially as regards the time $t = t(\tau)$, which appears in the equations of the $q(\tau)$ due to the explicit t dependence L_J. Equation (2.15) must be invoked to deal with this since there are only N equations of the q's. Once a parameter τ is chosen, the equations determine $t(\tau)$ in addition to the $q(\tau)$ and $p_t(t)$. In practice it is often, but not always, the case that Eqs. (2.15) and (2.30) reintroduce t' into the equations of the q's in the combination $q'/t' = \dot{q}(t)$, and the time history then may be found as an intermediate stage in getting the trajectory! Where this is not the case, the trajectory may be found without determining the time history first.

We remark finally that if p_t is constant, t does not appear in the equations of the q's and Eq. (2.15) no longer is needed to solve these. Instead, their solutions determine $t(\tau)$ by quadrature from Eq. (2.15). In the ordinary Jacobi principle Eq. (2.15) is not regarded as given, and so $t(\tau)$ is never found.

3. EXAMPLE FROM MECHANICS

We consider a particle of mass m moving in a time-varying potential field, taking

$$L = \tfrac{1}{2}m\dot{\mathbf{r}}^2 - V(\mathbf{r}, t), \tag{3.1}$$

so that

$$\overline{L} = \tfrac{1}{2}m\mathbf{r}'^2 t'^{-1} - V(\mathbf{r}, t)t', \tag{3.2}$$

$$p_t = -\tfrac{1}{2}m\mathbf{r}'^2 t'^{-2} - V(\mathbf{r}, t), \tag{3.3}$$

which gives

$$t' = |\mathbf{r}'|\{(2/m)[-p_t - V(\mathbf{r}, t)]\}^{-1/2} \tag{3.4}$$

whence

$$L_J(\mathbf{r}, \mathbf{r}', t) = \{2m\mathbf{r}'^2[-p_t(t) - V(\mathbf{r}, t)]\}^{1/2}. \tag{3.5}$$

The Euler–Lagrange equations for the variation condition, Eq. (2.20), are as many as the dimensionality of r,

$$\frac{d}{d\tau}\left\{\left[\left(\frac{2}{m}\right)[-p_t(t) - V(\mathbf{r}, t)]\right]^{1/2}\frac{\mathbf{r}'}{|\mathbf{r}'|}\right\} = -m^{-1}|\mathbf{r}'|\left[\left(\frac{2}{m}\right)[-p_t(t) - V(\mathbf{r}, t)]\right]^{-1/2}\frac{\partial V(\mathbf{r}, t)}{\partial \mathbf{r}}. \tag{3.6}$$

The equation of the interpretation of $p_t(t)$ is

$$-p_t = -L(\mathbf{r}, \dot{\mathbf{r}}, t) + \dot{\mathbf{r}}\cdot\frac{\partial L(\mathbf{r}, \dot{\mathbf{r}}, t)}{\partial \dot{\mathbf{r}}} = \tfrac{1}{2}m\dot{\mathbf{r}}^2 + V(\mathbf{r}, t). \tag{3.7}$$

To get the consistency condition, we formulate the canonical constraint [Eq. (2.25)], which is

$$K = p_t + (2m)^{-1}\mathbf{p}^2 + V(\mathbf{r}, t), \tag{3.8}$$

so that Eq. (2.29) gives

$$\frac{dp_t}{dt} = -\frac{\partial V(\mathbf{r}, t)}{\partial t}, \tag{3.9}$$

already in the form of Eq. (2.30). Equations (3.6), (3.7), and (3.9) form the closed system we want.

We consider the case of three-dimensional motion in a potential $V = V(\rho, t)$, where $\rho = x\hat{x} + y\hat{y}$, so that V is independent of z. $\hat{x}$ and $\hat{y}$ are unit vectors along the chosen coordinate axes. Combining Eq. (3.4) with the z component of Eq. (3.6) gives

$$z' = ct', \tag{3.10}$$

where c is a constant for any choice of τ. We exploit this to obtain the equations of the trajectory, expressing this in the form $\rho = \rho(z)$, by choosing

$$\tau = z, \tag{3.11}$$

so that

$$t' = c^{-1}. \tag{3.12}$$

With this choice the remaining components of Eq. (3.6) can be written

$$\frac{d}{dz}\left(c\frac{d\rho}{dz}\right) = -m^{-1}c^{-1}\frac{\partial V(\rho, t)}{\partial \rho}, \tag{3.13}$$

so that by choosing $t = 0$ and $z = 0$ to coincide and substituting the solution of Eq. (3.10), these become

$$mc^2\frac{d^2\rho}{dz^2} = -\frac{\partial V(\rho, z)}{\partial \rho}. \tag{3.14}$$

And finally, manipulating Eq. (3.13) in the standard fashion to secure a first integral, using Equation (3.9) along the way and Equations (3.3) and (3.10) to fix the constant of integration, we find[6]

$$-p_t(z) = \tfrac{1}{2}mc^2\left[1 + \left(\frac{d\rho}{dz}\right)^2\right] + V(\rho, z). \tag{3.15}$$

4. EXAMPLE FROM STUECKELBERG MECHANICS[7]: λ-DEPENDENT COUPLING TO A CONSTANT VECTOR FIELD

We consider the Lagrangian

$$L_S(x, dx/d\lambda) = \tfrac{1}{2}(dx/d\lambda)^2 + \tfrac{1}{2}E(\lambda)\tilde{\omega}(dx/d\lambda), \tag{4.1}$$

where $\tilde{\omega}$ is a Minkowski[8] four-vector and $E(\lambda)$ is a prescribed function whose properties we specify below. The Stueckelberg action is

$$A_s = \int_{\lambda_1}^{\lambda_3} d\lambda\, L_s(x(\lambda), dx/d\lambda, \lambda), \tag{4.2}$$

and the fixed end point variation condition on A_s results in Euler–Langrange equations of the curves $x^\mu = x^\mu(\lambda)$, $\lambda_1 < \lambda < \lambda_3$. In addition, we have the analogue to Eq. (2.15),

$$-p_\lambda(\lambda) = (dx/d\lambda)\cdot\frac{\partial L_s}{\partial(dx/d\lambda)} - L_s$$
$$= \tfrac{1}{2}(dx/d\lambda)^2 \equiv -\tfrac{1}{2}m(\lambda)^2, \tag{4.3}$$

where the fact of λ dependence follows from

$$\frac{dp_\lambda}{d\lambda} = \frac{\partial L_s}{\partial\lambda} = \Delta(\lambda)\tilde{\omega}\cdot\frac{dx}{d\lambda} \tag{4.4}$$

with the notation

$$\Delta(\lambda) \equiv \frac{1}{2}\frac{dE(\lambda)}{d\lambda}. \tag{4.5}$$

The generalized Jacobi principle gives equations of the spacetime trajectories corresponding to $x^\mu = x^\mu(\lambda)$. Introducing a parameter α, analogous to τ in the previous section, we must express x^μ and λ as functions of α and solve the analog of Eq. (2.8), for $\lambda'(\alpha) \equiv d\lambda/d\alpha$, to obtain the trajectory Lagrangian $L_{SJ}(x, x', \lambda)$. The equation giving λ' is [cf. Eq. (4.3)]

$$\lambda'^2 = -x'^2/m(\lambda)^2, \tag{4.6}$$

so the signs of $-x'^2$ and $m(\lambda)^2$ must be correlated; moreover, for a given correlation there are two solutions to Eq. (4.6), corresponding to the two signs of the square root. We want to be able to consider the case of curves that remain everywhere timelike, so that $-x'^2 > 0$, with $m(\lambda)$ real; then

$$\lambda'(\alpha) = \pm\, m(\lambda)^{-1}[-x'(\alpha)^2]^{1/2}, \tag{4.7}$$

where here $m(\lambda)$ means the positive square root of $m(\lambda)^2$, and

$$L_{SJ}(x, x', \lambda) = \mp\, m(\lambda)(-x'^2)^{1/2} + \tfrac{1}{2}E(\lambda)\tilde{\omega}\cdot x'. \tag{4.8}$$

Defining an "observer" Lagrangian,[3] $L_\Omega = \pm L_{SJ}$, we get

$$L_\Omega(x, x', \lambda = -m(\lambda)(-x'^2)^{1/2} + \tfrac{1}{2}E(\lambda)\epsilon^0\tilde{\omega}\cdot x' \tag{4.9}$$

with $\epsilon^0 = +1$ for the upper sign and $\epsilon^0 = -1$ for the lower. Independently of the choice of α, ϵ^0 is the sign of $dx^0/d\lambda$, which is constant if $dx^\mu/d\lambda$ is everywhere timelike. Before elaborating the spacetime trajectory equations, we want to explicate the basis of the formalism for this example, namely the condition that $dx^\mu/d\lambda$ be everywhere timelike; to do this, we must solve the Stueckelberg equations first.

The equations of the curve, stemming from L_S, give

$$\frac{dP^\mu}{d\lambda} = 0 \tag{4.10}$$

where

$$P^\mu = q^\mu + \tfrac{1}{2}E(\lambda)\tilde{\omega}^\mu, \tag{4.11}$$

with

$$q^\mu = dx^\mu/d\lambda. \tag{4.12}$$

The functions $x^\mu(\lambda)$ are gotten by a quadrature, which gives us the curve when a set of end point values is specified, such as $\{x^\mu(\lambda_1)\}$ or $\{x^\mu(\lambda_3)\}$.

We take $E(\lambda)$ to be a smooth function (constant on disjoint segments containing λ_1 and λ_3) with

$$E(\lambda_3) = -E(\lambda_1) = 1. \tag{4.13}$$

Equations (4.11) and (4.12) give

$$q(\lambda) = q(\lambda_1) - \Theta_+(\lambda)\tilde{\omega} \tag{4.14a}$$
$$= q(\lambda_3) + \Theta_-(\lambda)\tilde{\omega}, \tag{4.14b}$$

where

$$\Theta_\pm(\lambda) \equiv \tfrac{1}{2}[1 \pm E(\lambda)]. \tag{4.15}$$

Also, from Eqs. (4.3) and (4.4) we get

$$\frac{dm^2}{d\lambda} = 2\Delta(\lambda)\tilde{\omega}\cdot\frac{dx}{d\lambda} \tag{4.16}$$
$$= 2\Delta(\lambda)\tilde{\omega}\cdot[P - \tfrac{1}{2}E(\lambda)\tilde{\omega}], \tag{4.17}$$

with the aid of Eqs. (4.11) and (4.12). Integrating Eq. (4.17) gives the solution in either of two forms,

$$m(\lambda)^2 = m(\lambda_1)^2 + 2P\cdot\tilde{\omega}\Theta_+(\lambda) + \tilde{\omega}^2\Theta_+(\lambda)\Theta_-(\lambda) \tag{4.18a}$$
$$= m(\lambda_3)^2 - 2P\cdot\tilde{\omega}\Theta_-(\lambda) + \tilde{\omega}^2\Theta_+(\lambda)\Theta_-(\lambda). \tag{4.18b}$$

From Eq. (4.14),

$$q(\lambda_1) - q(\lambda_3) = \tilde{\omega}, \tag{4.19}$$

and from Eqs. (4.18),

$$m(\lambda_1)^2 - m(\lambda_3)^2 = -2P\cdot\tilde{\omega}. \tag{4.20}$$

We re-express the content of Eqs. (4.14) and (4.18)–(4.20) with the help of a notation change, considering two cases: (i) $dx^\mu/d\lambda$ timelike with $dx^0/d\lambda > 0$, at $\lambda = \lambda_1$, and (ii) $dx^\mu/d\lambda$ timelike with $dx^0/d\lambda < 0$, at $\lambda = \lambda_3$. We introduce the constants, $m > 0$, m', p, p', u and u', defining

$$\left.\begin{aligned} &m(\lambda_1) = m, \quad q(\lambda_1) = p = mu \\ &m(\lambda_3) = m', \quad q(\lambda_3) = p' = m'u' \end{aligned}\right\}, \quad \text{case (i)}; \tag{4.21a}$$

in addition, we assume

$$\left.\begin{aligned} &m(\lambda_3) = m, \quad -q(\lambda_3) = p = mu \\ &m(\lambda_1) = m', \quad -q(\lambda_1) = p' = m'u' \end{aligned}\right\}, \quad \text{case (ii)}. \tag{4.21b}$$

Since $m > 0$, u is a future pointing timelike vector. By defining $\epsilon = +$ in case (i) and $\epsilon = -$ in case (ii), Eqs. (4.14) become

$$p(\lambda) \equiv \epsilon q(\lambda) = p - \Theta_\epsilon(\lambda)\tilde{\omega}, \tag{4.22}$$

for both cases. In place of Eqs. (4.18) we write, again for both cases,

$$m(\lambda)^2 = -p^2 + 2\epsilon P\cdot\tilde{\omega}\Theta_\epsilon(\lambda) + \tilde{\omega}^2\Theta_+(\lambda)\Theta_-(\lambda), \tag{4.23}$$

whence substituting a similarly recast version of Eq. (4.11),

$$\epsilon P = p - \tfrac{1}{2}\tilde{\omega}, \tag{4.24}$$

we get

$$m(\lambda)^2 = m^2 + 2p\cdot\tilde{\omega}\Theta_\epsilon(\lambda) - \tilde{\omega}^2\Theta_\epsilon(\lambda)^2$$
$$= -[p - \Theta_\epsilon(\lambda)\tilde{\omega}]^2. \quad (4.25)$$

Equation (4.25) also can be gotten by squaring Eq. (4.22), so that it also is a consequence of Eqs. (4.3) and (4.10). Finally, from Eqs. (4.19) or (4.22) we have

$$p' = p - \tilde{\omega}, \quad (4.26a)$$
$$m'u' = mu - \tilde{\omega}, \quad (4.26b)$$

while from Eq. (4.25) or Eqs. (4.26), we have

$$m'^2 = -(p - \tilde{\omega})^2 = m^2 + 2p\cdot\tilde{\omega} - \tilde{\omega}^2. \quad (4.27)$$

The condition that $dx^\mu/d\lambda = q^\mu(\lambda)$ be everywhere timelike simply is that $m(\lambda)^2$ be everywhere positive, which assures also the constancy of the sign of $dx^0/d\lambda$ and allows ϵ to be replaced by ϵ^0 above. Equation (4.25) shows that this condition is met only if $\epsilon(\lambda)$ belongs to a restricted class of functions depending on the end point value p and on ϵ^0. The spacetime trajectory equations deriving from Eqs. (4.8) or (4.9) will determine the (e.g., time) evolution of m via that of λ. Particle and antiparticle solutions, corresponding to $\epsilon^0 = +1$ and $\epsilon^0 = -1$, are obtained together, the example being that of a smoothly varying mass and four-momentum, in which the mass changes invariantly from m to m', and the four-momentum goes from p to $p' = p - \tilde{\omega}$.

By choosing α to be the proper time, so that $d\alpha = ds = (-dx^\mu\, dx^\nu \eta_{\mu\nu})^{1/2}$, the Lagrangian (4.9) gives the equations of the trajectory,

$$\left(\frac{d}{ds}\; m(\lambda)\frac{dx^\mu}{ds} + \tfrac{1}{2}\epsilon^0 E(\lambda)\tilde{\omega}^\mu\right) = 0; \quad (4.28)$$

the equation of the interpretation of m, Eq. (4.3), is

$$m(\lambda)^2 = -(dx/d\lambda)^2 = (d\lambda/ds)^{-2}, \quad m(\lambda) > 0, \quad (4.29)$$

and the consistency condition, which comes from

$$K = [p_\Omega - \tfrac{1}{2}\epsilon^0 E(\lambda)\tilde{\omega}]^2 + m^2, \quad (4.30)$$

with p_Ω defined from L_Ω, is

$$\frac{dm^2}{d\lambda} = -\frac{\partial K/\partial\lambda}{\partial K/\partial m^2}$$
$$= 2\Delta(\lambda)\tilde{\omega}\cdot\epsilon^0 m(\lambda)\frac{dx}{ds}, \quad (4.31)$$

which we recognize as Eq. (4.16).

The special case of the improper "limit" that $E(\lambda) \to \epsilon(\lambda - \lambda_2)$, $\lambda_1 < \lambda_2 < \lambda_3$, is of some interest; here the particle (or antiparticle) experiences a sudden change of four-momentum and mass. The former is determined as $\tilde{\omega}^\mu$ by the equations of motion, but the latter is *not* determined by Eq. (4.17) for $dm/d\lambda$. In this "limit" $\Theta_\epsilon(\lambda) \to \theta(\pm(\lambda - \lambda_2))$, where θ is the Heaviside unit step function,[10] and $\Delta(\lambda) \to \delta(\lambda - \lambda_2)$, where δ is the Kirchoff–Dirac δ function.[10,11] Thus $dx/d\lambda$ lacks a value at $\lambda = \lambda_2$, where the argument of the δ function vanishes, and Eq. (4.17) cannot be integrated. So this could signal a problem with the condition that $dx^\mu/d\lambda$ be everywhere timelike. On the other hand, the equations of motion give $dx^\mu/d\lambda$ as constant on the segments (λ_1, λ_2) and (λ_2, λ_3), so that Eqs. (4.26) are all that is needed to verify the constancy of ϵ^0. This requires $p'^0 > 0$ and hence, by Lorentz invariance $m'^2 > 0$ which is a kind of threshold condition,

$$-(p - \tilde{\omega})^2 > 0. \quad (4.32)$$

Note that if $\tilde{\omega}$ is spacelike, Eq. (4.32) is enough to assure the constancy of ϵ^0. In the present improper "limit" the consistency condition, Eq. (4.31), no longer is in force through $\lambda = \lambda_2$.

5. CONCLUSION

We have shown that Jacobi's principle holds for the case of time-varying Hamiltonian, and we have given the general interpretation of the meaning of the formalism. We have illustrated the formalism with two simple examples, one from the mechanics of point particles and one from the mechanics of Stueckelberg, where the spacetime trajectory equations result.

ACKNOWLEDGMENT

I wish to thank Egon Marx for his comments on the paper.

[1]There is a slight *abus de langage* in our presentation, for there should be a part (iv) to the procedure, namely, choice of a parameter $\tau = T$, some function of the q's. So then there are only $N-1$ independent degress of freedom, with an Euler–Lagrange system of only $N-1$ independent equations. However, we can treat *this* system the way we did that of the N $q(t)$'s, passing to an "extended space" where the parameter T is treated as a coordinate, with T and the $N-1$ q's all functions of an arbitrary parameter, which we may as well just call τ. So the prescription of the text is that of the Jacobi principle if it is realized that it represents an "extended (trajectory) space" formulation.

[2]This follows from Eq. (2.7), for, examining the solution to that equation,

$$t(\tau) = t(\tau_1) + \int_{\tau_1}^{\tau} d\tau' t'(q(\tau'), q'(\tau')), \quad (a)$$

under $q \to q + \delta'' q$, $q' \to q' + \delta'' q'$, we find that $t \to t + \delta'' t$, where

$$\delta'' t(\tau) = \delta'' t(\tau_1) + \int_{\tau_1}^{\tau} d\tau'\, \delta'' t'(q(\tau'), q'(\tau')). \quad (b)$$

On the other hand,

$$\delta'' t(\tau) = \delta'' t(\tau_1) + \int_{\tau_1}^{\tau} d\tau' \frac{d}{d\tau'}\, \delta'' t(\tau'), \quad (c)$$

identically. Subtracting (c) from (b) and differentiating by τ gives the assertion of the text.

[3]Note that t', given by Eq. (2.17), is not needed here, although $q = q'/t'$ must be used in Eq. (2.15) when the latter is adjoined to the set deriving from Eq. (2.20).

[4]See, for example, C. Lanczos, *The Variational Principles of Mechanics* (University of Toronto Press, Toronto, 1957), Chap. 6, Sec. 10.

[5]P. A. M. Dirac, Can. J. Math. 2, 147 (1950).

[6]In Eqs. (3.14) and (3.15) we ought strictly to have written $V(\rho, c^{-1}z)$ and $p_t(c^{-1}z)$ rather than $V(\rho, z)$ and $p_t(z)$, as we have done for simplicity.

[7]E. C. G. Stueckelberg, Helv. Phys. Acta 14, 588 (1941); 15, 23 (1942).

[8]We use the space-favoring metric, whose nonvanishing Cartesian components are $\eta_{11} = \eta_{22} = \eta_{33} = -\eta_{00} = +1$, we put the speed of light equal to one and we use the notation $a\cdot b = \eta_{\mu\nu}a^\mu b^\nu$.

[9]R. G. Cawley, Int. J. Theor. Phys. 3, 483 (1970); 7, 77 (1973) (E).

[10]Generalized function.

[11]This is commonly called the Dirac δ function, but see also G. Kirchoff, *Vorlesungen über Mathematische Optik* (Leipzig, 1891), 2*te* Vorlesung, § 1 [By A. Sommerfeld, K. Akad. Wet. Amsterdam, Proc. 8, 346 (1904)].

On the solution of the phase retrieval problem*

B. J. Hoenders†

Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627
(Received 23 August 1974; revised manuscript received 18 February 1975)

It is shown that the intensity in the image plane of a microscope determines *uniquely* the phase of the corresponding image wavefunction up to an over-all phase. This result is obtained using the *a priori* information that both the image wavefunction and the unperturbed wavefunction in the Fraunhofer plane are band-limited and that we have some *a priori* knowledge about the intensity at the rim of the diaphragm in the Fraunhofer plane. If we have no useful *a priori* information about the wavefunction in the Fraunhofer plane, unique phase reconstruction is possible from two exposures, corresponding to two different values of the defocusing.

INTRODUCTION

Since only the modulus of a wavefunction is usually a measurable quantity rather than the wavefunction itself, the phase of a wavefunction is, as a rule, lost during measurement. The problem whether or not the phase of a function can be unambiguously determined from the values of its modulus is known as the phase retrieval problem.

The loss of phase information leads to serious difficulties in the interpretation of measurements and also to ambiguities. It is well known, for example, that the structure of a crystal cannot in general be uniquely determined from intensity measurements on the scattered x rays.

Another example arises in the theory of partial coherence where it is known that in general, the complex degree of coherence cannot be determined from its modulus without additional information, Wolf,[1] Nussenzveig.[2] Another example of this type is provided by the theory of image formation in microscopes, where it is shown that several distributions in the Fraunhofer plane might lead to the same intensity distribution in the Gaussian image plane, Walther.[3]

However, most functions occurring in physics are analytic, which implies that both phase and modulus are connected by the Cauchy–Riemann equations which considerably narrows the class of phase functions which might be assigned to a given modulus. Much work has been done to determine to which degree the phase of an analytic function is determined by its modulus. For an extensive survey see Mandel and Kohler.[4] From the calculations of Wolf,[1] Dialetis and Wolf,[5] Walther,[3] and Nussenzveig[2] it became clear that in several cases of physical interest the phase of an analytic function is not uniquely determined by its modulus. It is the aim of this paper to give an extension of the analysis of the phase problem as developed by Walther[3] and to provide an answer to the following phase problem occurring in the theory of image formation in a microscope: Suppose that an object is imaged through a microscope free of aberrations and the intensity in the image plane is measured. We then investigate whether the intensity distribution determines the phase of the image wavefunction unambiguously. As already shown by Walther,[3] several wavefunctions in the Fraunhofer plane might lead to the same intensity distribution in the image plane. However, it should be stressed that this does *not* prove that the phase cannot be reconstructed. This comes from our *a priori* knowledge that the *unperturbed* wavefunction $g(y)$, viz. the wavefunction just before the diaphragm in the Fraunhofer plane, is band-limited, like the image wavefunction, i.e.,

$$g(y)=\int_{-1}^{+1}\exp(ix_0y)\,\psi(x_0)\,dx_0, \tag{Ia}$$

$$h(x)=\int_{-1}^{+1}\exp(ixy)\,g(y)\,dy, \tag{Ib}$$

where $\psi(x_0)$ is the distribution in the object plane, $g(y)$ the unperturbed distribution in the Fraunhofer plane, and $h(x)$ the distribution in the Gaussian image plane, Born, Wolf.[6] It might so happen that only *one* of the various distributions in the Fraunhofer plane is consistent with our *a priori* knowledge that it must be a *perturbed* band-limited function, viz. a function which, relying on Kirchoff's boundary conditions is part of a band-limited function on the transparent part of the diaphragm and zero on the nontransparent part. This would imply that due to this *a priori* knowledge arg $h(x)$ is uniquely determined by $|h(x)|$. However, unfortunately this conjecture is not true, e.g., the modulus of the band-limited function $h^{(1)}(z)\equiv h^*(z^*)$

$$h^{(1)}(z)=\int_{-1}^{+1}\exp(izy)\,g^*(-y)\,dy,$$

is equal to the modulus of $h(z)$ on the real axis. Therefore the unperturbed wavefunction $g^*(-y)$ yields the same intensity distribution in the image plane as $g(y)$. Moreover, it follows from Eq. (1a) that $g^*(-y)$ is a band-limited function thus obtaining at least two unperturbed bandlimited distributions in the Fraunhofer plane leading to the same intensity in the image plane. In holography $g(y)$ and $g^*(-y)$ would correspond to the reconstructed object and its twin object, "twin images."

But, fortunately, uniqueness can be obtained by using a little more *a priori* information, i.e., if we know *a priori* that either the number of zeros of $h(z)$ is finite in the upper half or the lower half of the complex plane, uniqueness is obtained. The uniqueness is obtained by showing that only one (the band-limited) unperturbed wavefunction in the Fraunhofer plane decays if y tends to plus or minus infinity whereas all the other possible wavefunctions diverge. Moreover it will be shown how to calculate all these possible unperturbed wavefunctions from the intensity. The above mentioned *a priori*

knowledge concerning the distribution of the zeros of $h(z)$ is readily obtained in light—or electron microscopy and, as will be shown in the section "application to microscopy," boils down, to the *a priori* knowledge whether $|g(-1)| > |g(1)|$ or $|g(-1)| < |g(1)|$. This knowledge can be obtained from experiment or *a fortiori* by illuminating the object with a plane wave incident with a certain angle with the optical axis of the microscope. From now on it will be assumed that the number of zeros of $h(z)$ located in the upper half of the complex plane is finite. A theory completely analogous to the one developed in this paper can be derived if a finite number of zeros are located in the lower half of the complex plane. It would be interesting to find whether a similar analysis also holds if the object is imaged into the Fraunhofer plane by a microscope suffering from aberrations. It might so happen that in that case also only *one* wavefunction can be identified as the image in the Fraunhofer plane of an object distribution. Although the author regards this to be highly probable a proof of this conjecture has not been constructed so far. Therefore, let us suppose that in this case all the various wavefunctions $g(y)$ are admissible and hence that the phase cannot be reconstructed unambiguously from the modulus. In order to get rid of this ambiguity more information has to be put in. Recently, Gerchberg and Saxton[7] propose as one possible choice for such additional information the knowledge of the intensity distribution in the Fraunhofer plane.

Another way of obtaining additional information is to measure several intensity distributions in the image plane corresponding to different values of parameters which are at our disposal such as the location of the focus, illumination, or the size of the aperture. Misell[8] proposed a method of phase reconstruction for weak objects from two exposures, obtained by using two diaphragms, one which transmits only positive spatial frequencies, and one which transmits only negative spatial frequencies.

Another idea will be investigated in more detail in this paper. It will be shown that from two exposures, corresponding to two different values of the defocusing, the phases of the two wavefunctions in the image plane can be reconstructed unambiguously, up to a constant. To prove this statement we shall derive in the next section a dispersion relation between the phase and the modulus of a bandlimited function. Moreover it will be shown that all the wavefunctions $g_1(y)$ in the Fraunhofer plane leading to the same intensity in the image plane are related to the true wavefunction in the Fraunhofer plane by a linear Volterra integral equation of the second kind. Comparing the two sets of wavefunctions $\{g_1^{(1)}(y)\}$ and $\{g_1^{(2)}(y)\}$ which correspond to the two exposures it is shown that only one wave function is consistent with both intensity distributions. Hence, using (1) $\arg h(x)$ can be calculated.

A DISPERSION RELATION FOR BAND LIMITED FUNCTIONS

In the introduction we already indicated the need for a dispersion relation between phase and amplitude of a band limited function. Such a dispersion relation may be obtained by means of the following lemma:

Lemma: Let the complex valued function $g(y)$ be defined in the interval $-1 \leq y \leq +1$. Suppose $g'(y) = (d/dy) \times g(y)$ exists everywhere in that interval and is of bounded variation. Suppose that the entire function $h(z)$ is band limited to the interval $-1 \leq y \leq +1$, i.e., it has a representation of the form

$$h(z) = \int_{-1}^{+1} \exp(izy)\, g(y)\, dy. \tag{1}$$

Then, if the number of zeros $a_{n'}$ $(n' = 1, 2, \cdots)$ of $h(z)$ in the upper half of the complex plane is finite, the phase and the modulus of $h(z)$ are related by the dispersion relation

$$\frac{1}{\pi} \oint_{-\infty}^{\infty} \ln|h(x')x'| \, \frac{dx'}{x' - x}$$

$$= -\arg h(x) - \arg\left(x \exp(ix) \prod_{n'} \frac{x - a_{n'}^*}{x - a_{n'}}\right)$$

$$+ \frac{1}{2}\pi + \arg g(-1), \tag{2}$$

the integral on the left being interpreted as the Cauchy principal value and both x and x' are real numbers.

Proof: Repeated integration by parts yields

$$\int_{-1}^{+1} \exp(izy)\, g(y)\, dy = \frac{\exp(izy)}{iz}\, g(y)\Big|_{-1}^{+1} - \frac{\exp(izy)}{(iz)^2}\, g'(y)\Big|_{-1}^{+1}$$

$$+ \frac{1}{(iz)^2} \int_{-1}^{+1} \exp(izy)\, dg'(y). \tag{3}$$

Introducing complex numbers α and β by the relations

$$\exp(i\alpha) = g(1), \quad \exp(-i\beta) = g(-1), \tag{4}$$

we may readily derive from (2) the asymptotic formula

$$h(z) = \frac{2}{z} \exp\left(\frac{i}{2}(\alpha - \beta)\right) \sin[z + \tfrac{1}{2}(\alpha + \beta)] \left(1 + O\left\{\frac{1}{z}\right\}\right),$$

$$0 \leq \arg z \leq 2\pi. \tag{5}$$

Therefore the zeros a_n of $h(z)$ are distributed according to the asymptotic formula (Titchmarsh[9])

$$a_n \sim n\pi - \tfrac{1}{2}(\alpha + \beta), \tag{6a}$$

or Cartwright[10]

$$a_n \sim n\pi + \frac{i}{2} \ln\left(\frac{g(-1)}{g(1)}\right). \tag{6b}$$

Consider the contour integral

$$I(x, R) = \frac{1}{\pi i} \int_C \ln\left(\prod_{n'} \frac{z - a_{n'}^*}{z - a_{n'}}\, h(z)\, z \exp(iz)\right) \frac{dz}{z - x}, \tag{7}$$

for large values of the parameter R defined below.

In Eq. (7) the contour C consists of the part of the real axis between $-R$ and $+R$, indented at $z = x$ and at the possible zeros of $h(z)$ at the real axis with semicircles with radii ϵ in the upper half of the complex plane, and a semicircle in the upper half of the complex plane of radius R and centre at the origin. Using Eq. (3)

$$\exp(iz)\int_{-1}^{1}\exp(izy)\,g(y)\,dy=\frac{ig(-1)}{z}+O\left\{\frac{1}{z^2}\right\},$$

$$\theta<\arg z<\pi. \tag{8}$$

On using the asymptotic formula (8) we derive the formula

$$\frac{1}{\pi i}\int_{\text{semicircle}}\ln\left(\prod_{n'}\frac{z-a_{n'}^*}{z-a_n}\,h(z)z\exp(iz)\right)\frac{dz}{z-x}$$

$$=\frac{1}{\pi i}\int_{\text{semicircle}}\left(\ln\left(ig(-1)\right)+O\left\{\frac{1}{z}\right\}\right)\frac{dz}{z-x}$$

$$=\ln\,ig(-1)\;+O\left\{\frac{1}{R}\right\}. \tag{9}$$

We inserted the asymptotic expansion (8) into the lhs of (9) although (8) is not valid for real values of z. However, Eq. (3) shows that $\exp(iz)zh(z)$ is bounded for all values of $\arg z$ in the closed interval $[0,\pi]$. Hence, recalling that the number of zeros of $h(z)$ in the upper half of the complex plane is finite, which means using (6b) that $|g(-1)/g(1)|<1$, there exists a positive number M, independent of R, such that for sufficiently large values of R

$$\left|\ln\left(\prod_{n'}\frac{z-a_{n'}^*}{z-a_{n'}}\,h(z)z\exp(iz)\right)\right|<M,\quad z=R\exp(i\phi),$$

$$0\leq\phi\leq\pi. \tag{10}$$

Using (10), we observe that the contributions to (9) from those parts of the semicircle corresponding to values of $\arg z$ lying in the intervals

$$0\leq\arg z\leq\delta,\quad \pi-\delta\leq\arg z\leq\pi,$$

can be made arbitrarily small by choosing the positive number δ to be small enough. This proves that

$$\lim_{R\to\infty}\frac{1}{\pi i}\int_{\text{semicircle}}\ln\left(\prod_{n'}\frac{z-a_{n'}^*}{z-a_{n'}}\,h(z)z\exp(iz)\right)\frac{dz}{z-x}$$

$$=\ln(ig(-1)). \tag{11}$$

The argument of the logarithm of the integrand of (7) is an analytic function which by construction has no zeros in the upper half of the complex z-plane but possibly on the real axis. Hence the logarithm is an analytic function within the domain with boundary C and Cauchy's theorem applied to (7), yields the result

$$\lim_{R\to\infty}I(x,R)=0. \tag{12}$$

Letting the radii ϵ of the semicircles tend to zero and using the property that the possible zeros of $h(z)$ at the real axis gives no contribution to the integral (7), Eqs. (7), (11), and (12) leads to

$$\lim_{R\to\infty}I(x,R)=0=\ln(ig(-1))+\frac{1}{\pi i}\oint_{-\infty}^{+\infty}\ln\left(\prod_{n'}\frac{x-a_n^{*\prime}}{x'-a_{n'}}\right.$$

$$\left.\times h(x')x'\exp(ix')\right)\frac{dx'}{x'-x}-\ln\left(\prod_{n'}\frac{x-a_{n'}^*}{x-a_{n'}}\,h(x)x\exp(ix)\right). \tag{13}$$

Equating the imaginary parts of Eq. (13) leads to the desired dispersion relation

$$\frac{1}{\pi}\oint_{-\infty}^{+\infty}\ln|h(x')x'|\,\frac{dx'}{x'-x}=-\arg h(x)-\arg\left(xe^{ix}\prod_{n'}\frac{x-a_{n'}^*}{x-a_{n'}}\right)$$

$$+\frac{\pi}{2}+\arg(g(-1)). \tag{14}$$

RELATIONS BETWEEN POSSIBLE SOLUTIONS OF THE PHASE RETRIEVAL PROBLEM

Equation (14) reveals that, knowing the location of the zeros a_n of $h(z)$ in the upper half of the complex z-plane, the phase of $h(z)$ can be calculated from the values of its modulus along the real axis up to an overall phase constant. This result is similar to the well-known relation between the modulus and phase of the complex degree of coherence (Wolf[1], Nussenzveig[2]). Hence the question now arises whether or not the zeros a_n are determined from the knowledge of $|h(x)|$ along the real axis.

To answer this question we consider the function $h(z)h^*(z^*)$. First we observe that if $h(z)$ is an entire function so is $h^*(z^*)$, as follows immediately from the Cauchy–Riemann equations. Hence $h(z)\,h^*(z^*)$ is the analytical continuation of $h(x)h^*(x)$ into the whole complex plane and can be calculated with any desired tolerance from the values of $|h(x)|$ at a sufficiently large number of points.[11] Let $\{a_n\}$ denote the set of zeros of $g(z)$ and $\{b_n\}$ denote the set of zeros of $g^*(z^*)$. Therefore the set of zeros of $h(z)\,h^*(z^*)$ is the union $\{a_n\}U\{b_n\}$, which can in principle, be determined from the knowledge of $|h(x)|$ at the real axis.[11,12]

In order to apply the dispersion relation (2) we should be able to decide whether a particular zero in the upper half of the complex plane belongs to the set $\{a_n\}$ of zeros of $h(z)$ or to the set $\{b_n\}$ of zeros of $h^*(z^*)$. Unfortunately, as already observed by Walther,[3] it is impossible to make such a distinction. This can be seen by "flipping" one of the zeros of $h(z)$ about the real axis, which is equivalent to multiplying it by a so-called Blaschke factor $(z-a_n^*)/(z-a_n)$. We then obtain a new entire function, one zero of which has been replaced by its complex conjugate and the value of its modulus along the real axis has not changed because

$$\left|\frac{x-a_n^*}{x-a_n}\right|=1,\quad x\text{ real}.$$

Furthermore, it can be shown that multiplying a band limited function $h(z)$ by a Blaschke factor transforms it into another band limited function[15] (Walther[3]). (A proof different from Walther's will be given in Theorem 1.)

Observing that the set $\{b_n\}$ of zeros of $g(z)$ are the complex conjugates of the set $\{a_n\}$ of zeros of $g^*(z^*)$, it is clear that multiplication of the original function $h(z)$ with a suitable product of Blaschke factors yields a new band limited function, the zeros of which may be any finite number of combinations of all those elements of

the union $\{a_n\}U\{b_n\}$ lying in the upper half of the complex plane. The reason that we allow only a finite number of combinations is due to our *a priori* knowledge that only a *finite* number of zeros of $h(z)$ are located in the upper half of the complex plane as is discussed in the introduction. Moreover, the moduli of all these functions are the same along the real axis. Therefore any combination of Blaschke factors, generated by all possible combinations of those elements of the union $\{a_n\}$ $U\{b_n\}$ which lie in the lower half of the complex plane, yields, after insertion of these factors into the dispersion relation (2), a possible solution for the phase problem. All the possible band limited functions $h_1(z)$, labeled by the index 1, having the same modulus along the real axis, and a finite number of zeros in the upper half of the complex plane, can be represented by the formula

$$h_1(z)=\int_{-1}^{+1}\exp(izy)\,g_1(y)\,dy, \tag{15}$$

where

$$h_1(z)=\prod_{\{n_1\}}\frac{z-a_n^*}{z-a_n}\,h(z) \tag{16}$$

and

$$h(z)=\int_{-1}^{+1}\exp(izy)\,g(y)\,dy. \tag{17}$$

The product in Eq. (16) is taken over any subset $\{n_1\}$ containing a finite number of elements of the union $\{a_n\}U\{b_n\}$ in the lower half of the complex plane.

The following theorem will show that all the functions $g_1(y)$ are related to $g(y)$ by a linear Volterra integral equation of the second kind.

Theorem 1: Let the function $g(y)$ be defined on the interval $-1\leq y\leq+1$ and suppose that $g'(y)$ exists everywhere in this closed interval and is of bounded variation. Suppose the complex numbers a_n denote the zeros of the function

$$h(z)=\int_{-1}^{+1}\exp(izy)\,g(y)\,dy, \tag{18}$$

and let

$$h_1(z)=\prod_{\{n_1\}}\frac{z-a_n^*}{z-a_n}\,h(z), \tag{19}$$

where the index n_1 labels all the finite number of possible combinations of Blaschke factors. Then the functions $h_1(z)$ are band limited,[15] i. e., there exist functions $g_1(y)\in L^2$ such that

$$h_1(z)=\int_{-1}^{+1}\exp(izy)\,g_1(y)\,dy, \tag{20}$$

and $g(y)$ and $g_1(y)$ are related by the following Volterra integral equation of the second kind:

$$g_1(y)=g(y)-i\sum_{n''}(a_{n''}-a_{n''}^*)\prod_{\substack{\{n_1\}\\ n_1\neq n''}}\frac{a_{n''}-a_n^*}{a_{n''}-a_n}\times\int_{-1}^{y}e^{ia_{n''}(y-y')}g(y')\,dy'+i\sum_{n}(a_n-a_n^*)\prod_{\substack{\{n_1\}\\ n_1\neq n'}}\frac{a_{n'}-a_n^*}{a_{n'}-a_n}\times\int_{y}^{1}e^{ia_{n'}(y-y')}g(y')\,dy'. \tag{21}$$

In Eq. (21) the index n'' labels all the poles in the lower half of the complex plane of the Blaschke products occurring in (19). Similarly the index n' labels all the poles a_n in the upper half of the complex plane.

Proof: Consider the two integrals

$$I_1(y,c)=\frac{1}{2\pi}\int_{-c}^{+c}\exp(-iyx)\left(\prod_{\{n_1\}}\frac{x-a_n^*}{x-a_n}\times\int_{-1}^{y}\exp(ix\tau)\,g(\tau)\,d\tau\right)dx\quad\text{if } -1<y<+1, \tag{22}$$

$$I_2(y,c)=\frac{1}{2\pi}\int_{-c}^{+c}\exp(-iyx)\left(\prod_{\{n_1\}}\frac{x-a_n^*}{x-a_n}\times\int_{y}^{1}\exp(ix\tau)\,g(\tau)\,d\tau\right)dx\quad\text{if } -1<y<+1, \tag{23}$$

for large values of the real parameter c. On adding Eqs. (22) and (23) we obtain at once

$$I_1(y,c)+I_2(y,c)=\frac{1}{2\pi}\int_{-c}^{+c}\exp(-iyx)\,h_1(x)\,dx. \tag{24}$$

Let us close the contour of I_1 by a semicircle C^- in the lower half of the complex plane, with radius c, and centered at the origin, and let us close the contour of I_2 by a semicircle C^+ in the upper half of the complex plane, with radius c and centered at the origin.

Using the asymptotic expansions

$$\int_{-1}^{y}\exp[iz(-y+\tau)]\,g(\tau)\,d\tau=\frac{g(y)}{iz}+O\left\{\frac{1}{z^2}\right\},\quad \pi<\arg z<2\pi,\quad -1<y<+1, \tag{25}$$

$$\int_{y}^{+1}\exp[iz(-y+\tau)]\,g(\tau)\,d\tau=-\frac{g(y)}{iz}+O\left\{\frac{1}{z^2}\right\},\quad 0<\arg z<\pi,\quad -1<y<+1, \tag{26}$$

that follows from (3), we obtain from (22) and (25), with the help of the residue theorem,[13]

$$I_1(y,c)=-i\sum_{n''}(a_{n''}-a_{n''}^*)\prod_{\substack{\{n_1\}\\ n_1\neq n''}}\frac{a_{n''}-a_n^*}{a_{n''}-a_n}\times\int_{-1}^{y}\exp[ia_{n''}(-y+\tau)]\,g(\tau)\,d\tau+\frac{1}{2\pi}\int_{\pi}^{2\pi}\left(\frac{g(y)}{ice^{i\phi}}+O\left\{\frac{1}{c^2}\right\}\right)\times ce^{i\phi}\,id\phi,\quad -1<y<+1, \tag{27}$$

and

$$I_2(y,c)=i\sum_{n'}(a_{n'}-a_{n'}^*)\prod_{\substack{\{n_1\}\\ n_1\neq n'}}\frac{a_{n'}-a_n^*}{a_{n'}-a_n}\times\int_y^1\exp[ia_{n'}(-y+\tau)]\,g(\tau)\,d\tau+\int_\pi^0\left(\frac{g(y)}{-ice^{i\phi}}+O\left\{\frac{1}{c^2}\right\}\right)cie^{i\phi}d\phi,\quad -1<y<+1, \tag{28}$$

Hence if c tends to infinity we derive from (27) and (28)

$$\begin{aligned}\lim_{c\to\infty}\{I_1(y,c)+I_2(y,c)\}&=g(y)-i\sum_{n''}(a_{n''}-a_{n''}^*)\times\prod_{\substack{\{n_1\}\\ n_1\neq n''}}\frac{a_{n''}-a_n^*}{a_{n''}-a_n}\int_{-1}^y\exp[ia_{n''}(-y+\tau)]\,g(\tau)\,d\tau\\&+i\sum_{n'}(a_{n'}-a_{n'}^*)\times\prod_{\substack{\{n_1\}\\ n_1\neq n'}}\frac{a_{n'}-a_n^*}{a_{n'}-a_n}\int_y^1\exp[ia_{n'}(-y+\tau)]\,g(\tau)\,d\tau,\\&\qquad -1<y<+1.\end{aligned} \tag{29}$$

The relation (24) shows that the left-hand side of (29) is the Fourier transform of the entire function $h_1(z)$. Moreover, defining $g(\tau)$ to have the value zero if $|\tau|>1$ we derive in a similar way the result

$$\begin{aligned}\lim_{c\to\infty}\{I_1(y,c)+I_2(y,c)\}&=-i\sum_{n''}(a_{n''}-a_{n''}^*)\prod_{\substack{\{n_1\}\\ n_1\neq n''}}\frac{a_{n''}-a^*}{a_{n''}-a_n}\times\int_{-1}^{+1}\exp[ia_{n''}(-y+\tau)]\,g(\tau)\,d\tau=0\quad\text{if } y>1,\\&=i\sum_{n'}(a_{n'}-a_{n'}^*)\prod_{\substack{\{n_1\}\\ n_1\neq n'}}\frac{a_{n'}-a_n^*}{a_{n'}-a_n}\int_{-1}^{+1}\exp[ia_{n'}(-y+\tau)]\times g(\tau)d\tau=0\quad\text{if } y<-1.\end{aligned} \tag{30}$$

Hence the Fourier transform of the function $h_1(z)$ vanishes outside the interval $|x|\leq+1$. Therefore, the entire function $h_1(z)$ can be represented on the real axis and by, analytical continuation everywhere in the complex plane, by the formula

$$h_1(z)=\int_{-1}^{+1}\exp(izy)\,g_1(y)\,dy, \tag{31}$$

where

$$g_1(y)=\int_{-\infty}^{+\infty}\exp(-ixy)\,h_1(y)\,dy. \tag{32}$$

Combination of (24), (29), and (32) yields

$$\begin{aligned}g_1(y)=g(y)&-i\sum_{n''}(a_{n''}-a_{n''}^*)\prod_{\substack{\{n_1\}\\ n_1\neq n''}}\frac{a_n-a_n^*}{a_{n''}-a_n}\times\int_{-1}^y\exp[ia_{n''}(-y+\tau)]\,g(\tau)\,d\tau\\&+i\sum_{n'}(a_{n'}-a_{n'}^*)\prod_{\substack{\{n_1\}\\ n_1\neq n'}}\frac{a_{n'}-a_n^*}{a_{n'}-a_n}\times\int_y^1\exp[ia_{n'}(-y+\tau)]\,g(\tau)\,d\tau.\end{aligned} \tag{33}$$

APPLICATION TO MICROSCOPY

The preceding calculations have shown that the phase of a band limited function is not uniquely determined by its modulus along the real axis. Hence *a priori* information has to be used in order to obtain an unambiguous relation between the modulus and phase. We will now show that if we apply the preceding theorems to microscopy we have such a required *a priori* information, especially about the function $g(y)$ of Eq. (1).

Let us consider image formation by a microscope, free of aberrations, of a monochromatically illuminated object.

We know from the preceding analysis that several field distributions $g(y)$ in the Fraunhofer plane lead to the same intensity distribution $|h(x)|$ in the Gaussian image plane. However, we have the *a priori* information that, according to (1a), $g(y)$ is *band limited* and that the number of zeros of $h(z)$ located in the upper half of the complex plane is finite. The band limitation of the unperturbed wavefunction $g(y)$ is due to the imaging properties of the microscope [Eqs. (1a) and (1b)] and is therefore valid for any object imaged by a microscope whereas in general the number of zeros of $h(z)$ may be finite or infinite in the upper half of the complex plane [e.g. consider the example discussed in the introduction which shows that if a finite number of the zeros of $h(z)$ are located in the upper half of the complex plane, the band limited function $h^*(z^*)$ having the same modulus as $h(z)$ on the real axis, has an infinite number of zeros located in the upper half of the complex plane]. Recalling Eq. (6b),

$$a_n\sim n\pi+\frac{i}{2}\ln\left(\frac{g(-1)}{g(1)}\right), \tag{6b}$$

we deduce that the number of zeros of $h(z)$ located in the upper half of the complex plane is finite if $|g(-1)|<|g(1)|$. Hence, in microscopy, just by measuring the intensity in the end points of the Fraunhofer plane we can decide whether the condition $|g(-1)|<|g(1)|$ is valid or not. If $|g(-1)|>|g(1)|$, we can still apply the theory of this paper considering the function $h(-x)$ instead of $h(x)$, because using (1b)

$$h(-x)=\int_{-1}^{+1}\exp(ixy)\,g(-y)\,dy.$$

It is even possible to ensure that the condition $|g(-1)|<|g(1)|$ is valid *a fortiori* if, as usual in electron microscopy, we are dealing with *weak* objects, i.e., objects which only slightly perturbs the illuminating plane wave. Because in this case we can choose the angle of

incidence of the illuminating wave in such a way that the maximum of the diffraction spot coincides with the rim of the diaphragm. It will now be shown that only one function out of set of all possible distributions in the Fraunhofer plane is consistent with our *a priori* knowledge and, moreover, can be easily determined.

According to (34), $g(y)$ is an entire function. Therefore the rhs of (33) is an entire function, which is equal to the entire function $g_1(y)$ in the interval $-1 < y < +1$. Hence, by the principle of analytical continuation, Eq. (33) holds everywhere in the complex plane. Recalling that the numbers $a_{n'}$ and $a_{n''}$ have nonzero imaginary parts, it is apparent from (33) that in general $g_1(y)$ diverges if y tends to either $+\infty$ or $-\infty$. However, one and only one of all the possible field distributions [viz. $g(y)$] is bounded, as required by (34) and the asymptotic expansion (3), if y tends to either $+\infty$ or $-\infty$. Inspection of (33) shows that this will be the case if $g_1(y) = g(y)$.

Hence just by inspection of the asymptotic behavior of all the possible field distributions in the Fraunhofer plane and by using the *a priori* information of band limitation and hence boundedness, we can uniquely determine the field distribution in the Fraunhofer plane. Inserting this uniquely determined function in (35) yields the unique solution to the phase retrieval problem.

If *a priori* information of $g(y)$ is not available or at least not in a form which can be treated analytically, additional information could be obtained by making a second exposure, for different defocusing. We will now show that these two measurements are sufficient to allow us to determine $g(y)$ up to an overall phase. In one dimension the relation between the image wavefunction $h(x)$ and the wavefunction $g(y)$ in the Fraunhofer plane is

$$h_k(x) = \int_{-1}^{+1} \exp\left(ixy + \frac{\Delta z_k y^2}{f^2}\right) g(y)\, dy, \quad k = 1, 2, \tag{34}$$

where Δz_k is the distance between the defocused observation plane and the Gaussian image plane.

Inserting all possible combinations of Blaschke factors into the dispersion relation (3) we obtain from both measurements ($k = 1, 2$) two sets of functions $\{h_1^{(1)}(y)\}$ and $\{h_1^{(2)}(y)\}$ and by Fourier inversion two sets of functions $\{g_1^{(1)}(x)\}$ and $\{g_1^{(2)}(x)\}$. We know from (34) that only those functions for which

$$\frac{g_1^{(1)}(y)}{g_1^{(2)}(y)} = \exp\frac{i}{f^2}(\Delta z_1 - \Delta z_2)y^2, \tag{35}$$

are consistent with our *a priori* knowledge that both exposures are taken with two different values of the defocusing.

Using (33), condition (35) yields

$$g(y)\exp\left(\frac{i\Delta z_1}{f^2} y^2\right) - i\sum_{n''}(a_{n''} - a_{n''}^*) \prod_{\substack{\{n_1\} \\ n_1 \neq n''}} \frac{a_{n''} - a_n^*}{a_{n''} - a_n}$$

$$\times \int_{-1}^{y} \exp\left(ia_{n''}(-y + \tau) + \frac{i\Delta z_1}{f^2}\tau^2\right) g(\tau)\, d\tau$$

$$+ i\sum_{n'}(a_{n'} - a_{n'}^*)$$

$$\times \prod_{\substack{\{n_1\} \\ n_1 \neq n'}} \frac{a_{n'} - a_n^*}{a_{n'} - a_n} \int_{y}^{1} \exp\left(ia_{n'}(-y + \tau) + i\frac{\Delta z_1}{f^2}\tau^2\right) g(\tau)\, d\tau$$

$$= \left(\exp\frac{i}{f^2}(\Delta z_1 - \Delta z_2)y^2\right)\left[g(y)\exp\left(i\frac{\Delta z_2}{f^2} y^2\right)\right.$$

$$- i\sum_{n''}(b_{n''} - b_{n''}^*) \prod_{\substack{\{n_1\} \\ n_1 \neq n''}} \frac{b_{n''} - b_n^*}{b_{n''} - b_n} \int_{-1}^{y} \exp\left(ib_{n''}(-y + \tau)\right.$$

$$\left. + i\frac{\Delta z_2}{f^2}\tau^2\right) g(\tau)\, d\tau + i\sum_{n'}(b_{n'} - b_{n'}^*) \prod_{\substack{\{n_1\} \\ n_1 \neq n'}} \frac{b_{n'} - b_n^*}{b_{n'} - b_n}$$

$$\left. \times \int_{y}^{1} \exp\left(ib_{n'}(-y + \tau) + i\frac{\Delta z_2}{f^2}\tau^2\right) g(\tau)\, d\tau\right], \tag{36}$$

where the numbers $a_{n'}$ and $b_{n'}$ denote the zeros of the functions $h_1(z)$ and $h_2(z)$ in the upper half of the complex plane, and the numbers $a_{n''}$ and $b_{n''}$ the zeros of the functions $h_1(z)$ and $h_2(z)$ in the lower half of the complex plane.

Equation (36) is identically satisfied if none of the summations appear, i.e., if $g_1(y) \equiv g(y)$. Otherwise, as in the discussion following Eq. (33), we derive from Eq. (3) that if y tends to infinity the left-hand side of Eq. (36) is $O\{\exp(+c_1 y)\}$ if $y \to +\infty$, or $O\{\exp(+c_2 y)\}$ if $y \to -\infty$, where $c_1 = \max\{\mathrm{Im}\, a_{n'}\}$ and $c_2 = \max\{\mathrm{Im}\, a_{n''}\}$, whereas the right-hand side of Eq. (36) is $O\{\exp i(\Delta z_1 - \Delta z_2)y^2 + c_3 y\}$ if $y \to \infty$ or $O\{\exp(i(\Delta z_1 - \Delta z_2)y^2 + c_4 y)\}$ if $y \to -\infty$, where $c_3 = \max\{\mathrm{Im}(b_{n'})\}$ and $c_4 = \max\{\mathrm{Im}(b_{n''})\}$. Hence, Eq. (36) only can be satisfied if none of the summations appear. Therefore, only $g(y)$ satisfies condition (35), and can be determined up to a constant by the following procedure. Divide each function belonging to the set of functions $\{g_1^{(1)}(y)\}$ by every function belonging to the set of functions $\{g_1^{(2)}(y)\}$ and test if condition (35) is satisfied. Then one and only one pair of functions will be found satisfying (37), and these are the functions which determine the unknown function $g(y)$. Having determined $g(y)$, which is the main goal of image reconstruction, $h(y)$ and $\arg h(y)$ can be calculated. This provides the required solution to the phase reconstruction problem.

DISCUSSION

The preceding calculations not only prove the uniqueness of phase reconstruction of optical images but also provides an explicit procedure for calculating the phase. However, the question of stability is not considered and the procedure might be very sensitive to noise and errors in measurements. Hence computer simulated calculations should perhaps be employed to indicate the feasibility of the procedure or the need of another algorithm.

Uniqueness was obtained by using *a priori* information about the unperturbed wavefunction $g(y)$ in the Fraunhofer plane, namely that this wavefunction is band

limited and that we know that either $|g(-1)| < |g(1)$ or $|g(-1)| > |g(1)|$. Another *a priori* bit of information which could be used would be the knowledge of the quotient of two functions calculated from two images corresponding to two different settings of the defocusing. Gerchberg and Saxton[7] suggested that knowledge of the modulus of the wavefunction in the Fraunhofer plane determines uniquely the phases of both this wavefunction and of the image wavefunction. Computer simulated calculations sustained their claim. The results of this paper give additional support to their hypothesis.

Our theory is one dimensional. However, since micrographs are essentially two dimensional, a two dimensional extension of our theory is required. Clearly such a theory will be more complicated than the one presented here. For example flipping of zeros expressed by the use of Blaschke factors will then in general not apply, as is obvious from Weierstrass's preparation theorem, Osgood.[14]

These points will be discussed in a future publication.

ACKNOWLEDGMENT

The author wishes to thank Professor E. Wolf for the many stimulating discussions and critical comments and Dr. H. A. Ferwerda for his aid in revising this paper.

*This research was carried out during the tenure of a fellowship from the Netherlands Organization for the Advancement of Pure Research, (Z.W.O.) and was also supported in part by the Army Research Office (Durham).

†Present address: Technical Physical Laboratories, State University at Groningen, Nijenborgh 18, Groningen 8002, The Netherlands.

[1]E. Wolf, Proc. Phys. Soc. 80, 1269 (1962).
[2]H.M. Nussenzveig, J. Math. Phys. 8, 561 (1967).
[3]A. Walther, Opt. Acta 10, 41 (1963).
[4]D. Kohler and L. Mandel, J. Opt. Soc. Am. 63, 126 (1973).
[5]D. Dialetis and E. Wolf, Nuovo Cimento 47, 113 (1967).
[6]M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York, 1965), especially Sec. 8.6.3.
[7]R.W. Gerchberg and W.O. Saxton, Optik 34, 275 (1971).
[8]D.L. Misell, R.E. Burge, and A.H. Greenaway, J. Phys. D 7, 27 (1974). D.L. Misell, J. Phys. D 7, 69 (1974), 7, 832 (1974).
[9]E.C. Titchmarsh, Proc. Lond. Math. Soc. 25, 283 (1926).
[10]M.L. Cartwright, Q. J. Math. Oxford, Ser. (1) 1, 38 (1930) and 2, 113 (1931).
[11]H.A. Ferwerda and B.J. Hoenders, Optik 37, 542 (1973).
[12]For an explicit procedure see H.A. Ferwerda and B.J. Hoenders, Optik 40, 14 (1974).
[13]We substituted the asymptotic expansions (25) and (26) into (27) and (28) although (25) is not valid in the sectors $\pi \leq \arg z \leq \pi + \delta$ and $2\pi - \delta \leq \arg z \leq 2\pi$ and (26) is not valid in the sectors $0 \leq \arg z \leq \delta$ and $\pi - \delta \leq \arg z \leq \pi$, where δ is an arbitrarily small positive number. However, it is an immediate consequence of (3) that the left-hand sides of (25) and (26) are $O\{1/z\}$ if $|z|$ tends to infinity in the *closed* sectors $\pi \leq \arg z \leq 2\pi$, resp. $0 \leq \arg z \leq \pi$, and that therefore the error made by substituting (25) and (26) into (27) and (28) as if they were valid for *closed* sectors can be made arbitrarily small.
[14]W.F. Osgood, *Lehrbuch der Funktionentheorie*, Vol. II, 1, Chap. 2, Sec. 2.
[15]This follows immediately from the following theorem which can be found in R.A.C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain* (Am. Math. Soc., New York, 1934), especially Chap. 3, Sec. 6: Suppose that an entire function $f(z) \in L^2$ along the real axis and suppose there exists a real number A such that $f(z) = 0$ $(\exp A|z|)$. Then there exists a function $F(x) \in L^2$ such that

$$f(z) = \int_{-A}^{+A} \exp(izx) F(x)\, dx.$$

Remarks on the Green's functions for the cubic lattices

Koichi Mano

Air Force Cambridge Research Laboratories, Hanscom AFB, Bedford, Massachusetts 01731
(Received 21 April 1975)

A simpler way of finding the Green's function $G(2p,0,0)$ for the body-centered and face-centered cubic lattices as the square and product, respectively, of the pth derivatives of two complete elliptic integrals of the first kind is pointed out. The key relations required for these results are the Clausen's and Brafman's formulas for the body-centered and face-centered cubic lattices, respectively.

The Green's functions for the cubic lattices have been the object of an extensive study for many years. As a result of such a study it is known that the value of a Green's function at an arbitrary site can be expressed in terms of the values of the function at less general sites through the use of the recurrence formula for that function.

Recently it has been shown by Inoue[1,2] that an essential role is played by $G(2p,0,0)$ [see Eq. (1) below] in evaluating the value of the Green's function at an arbitrary site for the case of the face-centered cubic (fcc) and body-centered cubic (bcc) lattices. Based on this observation she has gone through the evaluation of $G(2p,0,0)$ for these lattices with a considerable effort. In the case of the fcc lattices the final result is expressed in a rather complicated triple sum which involves the F_4 function of Appell, while that for the bcc case is expressed in a form which is also fairly cumbersome. For the fcc case it has been pointed out by the present author[3] that the triple sum can be carried out in closed form to give a simple form of being the product of the pth derivatives of two complete elliptic integrals of the first kind.

The purpose of this paper is to present a simpler alternative procedure for obtaining the result found in Ref. 3 for the fcc case and also to point out a method by which the closed general formula for the bcc case can be obtained.

Let us begin with the relevant Green's functions which are defined by

$$G_C(l,m,n)=\frac{1}{\pi^3}\int_0^\pi\int\int\frac{\cos lx\cos my\cos nz\,dxydz}{E-i\delta-\omega(x,y,z)},$$

$$\omega(x,y,z)=\begin{cases}\cos x\cos y\cos z & \text{for } C=b,\\ \cos x\cos y+\cos y\cos z+\cos z\cos x & \text{for } C=f,\\ \cos x+\cos y+\cos z & \text{for } C=s,\end{cases}\tag{1}$$

where b, f, and s refer to the bcc, fcc, and sc (simple cubic) lattices, respectively. In the discussion that follows we will restrict ourselves to the regions $E\geq 1$, $E\geq 3$, and $E\geq 3$ for the bcc, fcc, and sc cases, respectively, where the functions defined by Eq. (1) will all be real.

When $l=m=n=0$ holds, it is well known that G_b with $E=1$ can be represented as the square of a complete elliptic integral of the first kind, while G_f and G_s with $E=3$ as the product of two complete elliptic integrals of the first kind. In the cases of G_f and G_s, however, they can also be reduced to the square of one such integral due to the degeneracy that results from the special choice $l=m=n=0$ for these functions.

Focusing our attention to $G_C(2p,0,0)$ for the reason stated in the foregoing, we remind ourselves that the function can be represented in the following form[4]:

$$G_C(2p,0,0)=\frac{1}{\pi^2}\int_0^\pi dx\cos 2px H_C(k_C),\tag{2}$$

$$H_C(k_C)=\begin{cases}2K(k_b)/E, & k_b=\cos x/E \quad \text{for } C=b,\\ 2K(k_f)/(1+E), & k_f=2(E+\cos^2x)^{1/2}/(1+E) \quad \text{for } C=f,\\ k_sK(k_s), & k_s=2/(E-\cos x) \quad \text{for } C=s,\end{cases}\tag{3}$$

where $K(k)$ stands for the complete elliptic integral of the first kind.

To evaluate the integral in Eq. (2), let us first work with the bcc case. It is readily seen that $G_b(2p,0,0)$ can be represented by

$$G_b(2p,0,0)=\frac{1}{E}\sum_{n=0}^{\infty}\frac{(\frac{1}{2})_n(\frac{1}{2})_n}{(1)_n(1)_n}\left(\frac{1}{E^2}\right)^n\frac{1}{\pi}\int_0^\pi dx\cos 2px\cos^{2n}x,\tag{4}$$

as shown in Ref. 2. The well-known integral in Eq. (4) can be evaluated as[5]

$$\frac{\pi\,\Gamma(1+2p)}{2^{2n+1}\,\Gamma(1+n+p)\,\Gamma(1+n-p)},\tag{5}$$

and in view of the presence of the factor $1/\Gamma(1+n-p)$ in Eq. (5) it is important to realize that the summation in Eq. (4) starts effectively from $n=p$. Moreover, the factor $\Gamma(1+2n)$ of Eq. (5) may be rewritten by the use of the duplication formula as $\Gamma(1+2n)=2^{2n}\Gamma(n+1)\Gamma(n+\frac{1}{2})/\sqrt{\pi}$. Then, upon introduction of a new summation index by $n-p\to n$ and use of the identity $(a)_{n+p}=(a+p)_n(a)_p$ we see that

$$G_b(2p,0,0)=\frac{1}{2^{2p}E^{2p+1}}\left(\frac{(\frac{1}{2})_p}{p!}\right)^2{}_3F_2\left[\begin{matrix}\frac{1}{2}+p,\ \frac{1}{2}+p,\ \frac{1}{2}+p;\\ 1+p,\ 1+2p\end{matrix}\ 1/E^2\right].\tag{6}$$

This result should be compared with Eq. (3.2) of Ref. 2. If we combine Clausen's formula[6] with Eq. (10) on p. 111 of Ref. 5 we obtain

$${}_3F_2\left[\begin{matrix}2a,\ a+b,\ 2b;\\ a+b+\frac{1}{2},\ 2a+2b\end{matrix}\ z\right]=\left\{{}_2F_1\left[\begin{matrix}2a,\ 2b;\\ a+b+\frac{1}{2}\end{matrix}\ (1-\sqrt{1-z})/2\right]\right\}^2,\tag{7}$$

which is, of course, equivalent to Eq. (2.10) of Ref. 2. Then by applying the result of Eq. (7) for $2a=\frac{1}{2}+p=2b$ to Eq. (5) we finally obtain

$$G_b(2p,0,0) = \frac{1}{2^{2p}E^{2p+1}}\left(\frac{(\frac{1}{2})_p}{p!}\right)^2 \left\{ {}_2F_1\left[\begin{matrix}\frac{1}{2}+p,\ \frac{1}{2}+p \\ 1+p\end{matrix}; (1-\sqrt{1-E^{-2}})/2\right]\right\}^2. \tag{8}$$

If it is desired, one may rewrite the ${}_2F_1$ in Eq. (8) as the pth derivative of ${}_2F_1(\frac{1}{2},\frac{1}{2};1;k_0^2)=(2/\pi)K(k_0)$ with respect to $k_0^2=[1-(1-E^{-2})^{1/2}]/2$. It should be noted that the result in Eq. (8) is precisely identical to that obtained by Joyce[7] by a different method for his function $P(0,2m,2n)$, which coincides with our $G_b(2p,0,0)$ when $m=0$, $n=p$, and $z=E^{-1}$. As already mentioned in Ref. 2, the results of Eq. (8) and Joyce's for $p=0,\ldots,3$ produces the same expressions as given in Eqs. (3.9)–(3.12) of Ref. 2. One observes from the expression for $P(0,2m,2n)$ found by Joyce given as the product of two ${}_2F_1$ functions that the degeneracy of the bcc lattice Green's function from the *product* for $P(0,2m,2n)$ to the *square* for $P(0,0,2p)$ of the ${}_2F_1$ functions takes place as a direct consequence of our choice $m=0$.

Returning to the fcc lattice case we recall that the necessary integration for $G_f(2p,0,0)$ has already been carried out in Ref. 1. However, we shall see shortly that carrying the integration all the way through as it was done in Ref. 1 complicates the matter unnecessarily. In spite of the fact that the complicated final result obtained in Ref. 1 can actually be summed in closed form as shown in Ref. 3, it is advisable to cease the evaluation of Ref. 1 at some appropriate stage in order to obtain the desired result in a quicker way. Thus, we stop our calculation at the stage where after evaluating the integral in Eq. (2) in terms of Legendre polynomials $P_n^p(\xi)$ we express them as multiples of Jacobi polynomials $P_{n-p}^{(p,p)}(\xi)$. By doing so we obtain

$$G_f(2p,0,0) = \frac{1}{1+E}\,\frac{(\xi^2-1)^{p/2}}{2^p}\sum_{n=0}^{\infty}\frac{(\frac{1}{2})_n(\frac{1}{2})_n}{(1)_n(1)_n}\,P_{n-p}^{(p,p)}(\xi) \times\left[\left(\frac{2}{1+E}\right)^2[E(1+E)]^{1/2}\right]^n, \tag{9}$$

where

$$\xi=(E+\tfrac{1}{2})/[E(1+E)]^{1/2}. \tag{10}$$

Since it can be shown here again that the summation in Eq. (9) starts effectively from $n=p$ by essentially the same reason as for Eq. (4), we introduce a new summation index by $n-p\to n$. In contrast to Eq. (6) for $G_b(2p,0,0)$, this gives rise to the following expression for $G_f(2p,0,0)$:

$$G_f(2p,0,0) = \frac{1}{(1+E)^{2p+1}}\left(\frac{(\frac{1}{2})_p}{p!}\right)^2\sum_{n=0}^{\infty}\frac{(\frac{1}{2}+p)_n(\frac{1}{2}+p)_n}{(1+p)_n(1+p)_n} \times P_n^{(p,p)}(\xi)\left[\left(\frac{2}{1+E}\right)^2[E(1+E)]^{1/2}\right]^n. \tag{11}$$

Here we note the following formula due to Brafman[8]

$$\sum_{n=0}^{\infty}\frac{(q)_n(1+\alpha+\beta-q)_n}{(1+\alpha)_n(1+\beta)_n}P_n^{(\alpha,\beta)}(x)t^n = {}_2F_1\left[\begin{matrix}q,\ 1+\alpha+\beta-q \\ 1+\alpha\end{matrix}; (1-t-w)/2\right] \times {}_2F_1\left[\begin{matrix}q,\ 1+\alpha+\beta-q \\ 1+\beta\end{matrix}; (1+t-w)/2\right], \tag{12}$$

where $w=(1-2xt+t^2)^{1/2}$ and q is arbitrary. If we set $x\to\xi=(E+\frac{1}{2})/[E(1+E)]^{1/2}$ and $t=[2/(1+E)]^2[E(1+E)]^{1/2}$, we find that

$$\tfrac{1}{2}(1\mp t-w)=\tfrac{1}{2}\left[1\mp\left(\frac{2}{1+E}\right)^2[E(1+E)]^{1/2} - \frac{(E-1)(1+E)^{1/2}(E-3)^{1/2}}{(1+E)^2}\right], \tag{13}$$

which are identical to $k_\pm^2$ of Ref. 1. By substituting the result of Eq. (12) for $\alpha=\beta=p$ and $q=\frac{1}{2}+p$ with the above values of $(1\mp t-w)/2$ into Eq. (11) we finally obtain

$$G_f(2p,0,0)=\frac{1}{(1+E)^{2p+1}}\left(\frac{(\frac{1}{2})_p}{p!}\right)^2 {}_2F_1\left[\begin{matrix}\frac{1}{2}+p,\ \frac{1}{2}+p \\ 1+p\end{matrix}; k_+^2\right] \times {}_2F_1\left[\begin{matrix}\frac{1}{2}+p,\ \frac{1}{2}+p \\ 1+p\end{matrix}; k_-^2\right], \tag{14}$$

which coincides with the result that has been found previously in Ref. 3. Naturally, we may express the ${}_2F_1$ functions as the pth derivatives with respect to the respective variables of more basic ${}_2F_1$ functions, as was done in Ref. 3.

Lastly, let us consider the sc case. Here the procedure analogous to the foregoing leads to

$$G_s(2p,0,0)=\frac{1}{\pi E}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(\frac{1}{2})_m(\frac{1}{2})_m(2m+1)_n}{(1)_m(1)_m n!}\left(\frac{2}{E}\right)^{2m}\left(\frac{1}{E}\right)^n \times\int_0^{\pi}dx\cos 2px\cos^n x. \tag{15}$$

To evaluate the integral in Eq. (15), we split the range of integration into $[0,\pi/2]$ and $[\pi/2,\pi]$. The result for the former follows directly from the same formula as that used for the bcc case. The latter, upon change of variable $x\to\pi-x$, gives $(-1)^n$ times the former, leading to the result that only $n=$ even contributes to the sum in Eq. (15). With this in mind we introduce a new summation index by $n\to 2n$ which allows us to express $G_s(2p,0,0)$ as

$$\frac{1}{\pi E}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{\Gamma(m+\frac{1}{2})\Gamma(m+\frac{1}{2})\Gamma(2m+2n+1)}{m!\,\Gamma(m+1)\Gamma(2m+1)\Gamma(1+n+p)\Gamma(1+n-p)} \times\left(\frac{2}{E}\right)^{2m}\left(\frac{1}{2E}\right)^{2n}.$$

By replacing $n-p\to n$ and rewriting the factors of the form $\Gamma(2k+1)$ which appear both in the numerator and denominator we obtain

$$G_s(2p,0,0) = \frac{1}{\pi E^{2p+1}}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{\Gamma(m+\frac{1}{2})\Gamma(m+n+p+\frac{1}{2})\Gamma(m+n+p+1)}{m!\,n!\,[\Gamma(m+1)]^2\,\Gamma(n+2p+1)} \times\left(\frac{2}{E}\right)^{2m}\left(\frac{1}{E}\right)^{2n}. \tag{16}$$

We note that Eq. (16) above represents a generalization of the result for $G_s(0,0,0)$ which was found earlier by a different method.[9] We observe also that the direct rewriting of the right-hand side of Eq. (16) in terms of the Pochhammer symbol does not identify it with any one of the functions in the Horn's list of functions in two variables.[10]

It can readily be shown that Eq. (16) may be rewritten as

$$G_s(2p,0,0)=\frac{1}{2^{2p}E^{2p+1}}\sum_{m=0}^{\infty}\frac{(\frac{1}{2})_m(\frac{1}{2}+p)_m(1+p)_m}{m!(1)_m(1)_m}\left(\frac{2}{E}\right)^{2m}\times {}_2F_1\left[\begin{matrix}m+\frac{1}{2}+p,\ m+1+p;\\ 1+2p\end{matrix}\ (1/E)^2\right] \tag{17}$$

$$=\frac{1}{2^{2p}E^{2p+1}}\sum_{n=0}^{\infty}\frac{(\frac{1}{2}+p)_n(1+p)_n}{n!\,(1+2p)_n}\left(\frac{1}{E}\right)^{2n}\times {}_3F_2\left[\begin{matrix}\frac{1}{2},\ n+\frac{1}{2}+p,\ n+1+p;\\ 1,\ 1\end{matrix}\ (2/E)^2\right]. \tag{18}$$

Unfortunately, however, neither the right-hand side of Eq. (17) nor that of (18) seems to be summed up in a closed form which represents a generalization of $G_s(0,0,0)$ that can be represented by the product of two $K(k)\propto {}_2F_1(\frac{1}{2},\frac{1}{2};1;k^2)$ functions [or the square of a $K(k)$ function], as was mentioned in the beginning. If the right-hand side of Eq. (17) or (18) were equal to some function, such as the F_i function, $i=1,2,3,4$, of Appell, for which there exists a known identity by which it can be expressed as the type ${}_2F_1\times {}_2F_1$, then the situation would be rather interesting. However, it seems very unlikely that such is the case.

Thus, we have to conclude at this stage that no specially interesting representation for $G_s(2p,0,0)$ has been found as yet.

Note added in proof: After completing the present manuscript we succeeded in evaluating $G(2p_1,2p_2,2p_3)$ for $C=b$ and s (to be submitted for publication). According to the formula we found it is easy to check that Eq. (16) for $G_s(2p,0,0)$ equals $E^{-1}(2E)^{-2p}F_C(\frac{1}{2}+p,1+p;1+2p,1,1;E^{-2},E^{-2},E^{-2})$, where F_C stands for the Lauricella's function of three variables. [See P. Appell and J. Kampé de Fériet, *Fonctions hypergéométriques et hypersphériques-polynomes d'Hermite* (Gauthier-Villars, Paris, 1926), Chap. VII.]

[1]M. Inoue, J. Math. Phys. **15**, 704 (1974).
[2]M. Inoue, J. Math. Phys. **16**, 809 (1975).
[3]K. Mano, J. Math. Phys. **15**, 2175 (1974).
[4]This follows, for example, from I. Mannari and C. Kawabata, Department of Physics, Okayama University, Research Notes, No. 15, 1964, quoted in T. Morita and T. Horiguchi, J. Math. Phys. **12**, 981 (1971).
[5]See, for example, A. Erdélyi, Ed., *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I, p. 12.
[6]See p. 185 of Ref. 5 with the correction that $a+2b$ is to be replaced by $2a+2b$.
[7]G. S. Joyce, J. Phys. C **1**, 1510 (1971).
[8]A. Erdélyi, Ed., *Higher Transcendental Functions* (McGraw-Hill, New York, 1955), Vol. III, p. 265.
[9]S. Katsura, S. Inawashiro, and Y. Abe, J. Math. Phys. **12**, 895 (1971).
[10]Page 224 of Ref. 5.

Canonical transformations and path integrals

Robert Fanelli

Brooklyn College of the City University of New York, Brooklyn, New York 11210
(Received 5 February 1975)

A limited class of canonical transformations is introduced into the Lagrangian path integral method of quantization. Path integral quantization in different representations is discussed and a simple example is given.

I. INTRODUCTION

The path integral method of quantization introduced by Feynman[1] has always seemed to suffer from two limitations, first that it can only be implemented for Lagrangians quadratic in the generalized velocity and second that canonical transformations cannot naturally be incorporated into the scheme. Although it is a simple matter, for example, to formally introduce the momentum representation by means of the Fourier transform, justifying the identification of the argument of the transformed wavefunction with the momentum is not so easy.

In this article we discuss how at least a limited class of canonical transformations can be defined and carried out in the context of path integral quantization. The above-mentioned identification of the canonical momentum is then easily done. Furthermore, path integral quantization can then be carried out in representations other than the coordinate representation. For Lagrangians that are not quadratic in the velocities this allows propagators to be found indirectly.

We restrict ourselves here to the Lagrangian path integral method rather than the canonical approach involving integration over paths in phase space.[2]

The "ordering" problem, that is, the reflection of the operator ordering ambiguity of canonical quantization by corresponding ambiguities in path integral quantization, is not a major concern of this work although it has been discussed in several recent papers. It will be mentioned only briefly in the last section.

As a preliminary matter, the first section contains a short discussion of the relationship of classical canonical transformations to Hamilton's principle. In the second section canonical transformations of the wavefunction are defined and interpreted by means of a version of Dirac's classical limit argument.[3] Path integral quantization in different representations is discussed in the third section, and a simple example is presented.

2. CLASSICAL CANONICAL TRANSFORMATIONS

We wish to define canonical transformations directly in terms of Hamilton's principle for a classical system. Thus given the variational principle

$$\delta \int_{t_0}^{t_f} L(q,\dot{q},t)\,dt=0, \tag{1}$$

where q refers to a set of m coordinates, we wish to transform to new coordinates $Q(q,\dot{q},t)$ with a variational principle

$$\delta \int_{t_0}^{t_f} L'(Q,\dot{Q},t)\,dt=0 \tag{2}$$

so that both give the same extremum path. The variations about the extremum path are different since different generalized coordinates are held fixed at the end points in each case. Thus $L' \neq L$ and we write

$$L'(Q,\dot{Q},t)=L(q,\dot{q},t)-f(q,\dot{q},Q,\dot{Q},t). \tag{3}$$

The new variation then yields

$$\begin{aligned}\delta\int L'\,dt=\int_{t_0}^{t_f}\Big\{&\left[\frac{\partial L}{\partial q_i}-\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right)-\frac{\partial f}{\partial q_i}+\frac{d}{dt}\left(\frac{\partial f}{\partial \dot{q}_i}\right)\right]\delta q_i\\&-\left[\frac{\partial f}{\partial Q_i}-\frac{d}{dt}\left(\frac{\partial f}{\partial \dot{Q}_i}\right)\right]\delta Q_i\Big\}\,dt\\&+\left[\left(\frac{\partial L}{\partial \dot{q}_i}-\frac{\partial f}{\partial \dot{q}_i}\right)\delta q_i\right]_{t_0}^{t_f}=0,\end{aligned} \tag{4}$$

and a sufficient but not necessary condition (since the δq_i are not independent of the δQ_i) for the new variation to give the same extremum path as the old is

$$f=\frac{d}{dt}F(q,Q,t),\quad \frac{\partial L}{\partial \dot{q}_i}=\frac{\partial F}{\partial q_i}. \tag{5}$$

By similarly considering the old variation we obtain

$$\frac{\partial L'}{\partial \dot{Q}_i}=-\frac{\partial F}{\partial Q_i}. \tag{6}$$

These conditions define the usual canonical transformation with our F as the F_1 generating function, and the new action integral is related to the old by

$$\int_{t_0}^{t_f} L(q,\dot{q},t)\,dt=\int_{t_0}^{t_f} L'(Q,\dot{Q},t)\,dt+[F(q,Q,t)]_{t_0}^{t_f}. \tag{7}$$

If we are given $F(q,Q,t)$, we can deduce $Q(q,\dot{q},t)$ from (5) and the old Lagrangian in the standard fashion.

3. TRANSFORMATIONS OF THE PATH INTEGRAL

Since the change in the action integral due to a canonical transformation is equal to the difference between the end point values of the generating function, we might, in quantum mechanics, represent the modification of a Feynman path integral in a corresponding manner. That is, if we represent the time evolution of a wavefunction from t_0 to t_f in the q representation in terms of the usual propagator

$$\psi(q_f,t_f)=\int \exp\left(\frac{i}{\hbar}\int_{t_0}^{t_f} L\,dt\right) Dq\,\psi(q_0,t_0)\,d^m q_0, \tag{8}$$

where $q_0 = q(t_0)$ etc., then the time evolution in the Q representation would be given by

$$\chi(Q_f,t_f) = \int \exp\left(\frac{i}{\hbar}\left[-F(q_f,Q_f,t_f) + \int_{t_0}^{t_f} L(q,\dot{q},t)\,dt + F(q_0,Q_0,t_0)\right]\right) \bar{D}q\,\chi(Q_0,t_0)\,\frac{d^mQ_0}{A'}$$

$$= \int \exp\left(\frac{-i}{\hbar}F(q_f,Q_f,t_f)\right) \int \exp\left(\frac{i}{\hbar}\int L dt\right) Dq \times \exp\left(\frac{i}{\hbar}F(q_0,Q_0,t_0)\right)\chi(Q_0,t_0)\,\frac{d^mq_f}{A}\,\frac{d^mq_0}{A}\,\frac{d^mQ_0}{A'}, \tag{9}$$

leading us to postulate

$$\chi(Q,t) = \int \exp\left(\frac{-i}{\hbar}F(q,Q,t)\right)\psi(q,t)\,\frac{d^mq}{A},$$

$$\psi(q,t) = \int \exp\left(\frac{i}{\hbar}F(q,Q,t)\right)\chi(Q,t)\,\frac{d^mQ}{A'}. \tag{10}$$

We must then have

$$\int \exp\left(\frac{i}{\hbar}F(q,Q,t)\right)\exp\left(\frac{-i}{\hbar}F(q',Q,t)\right)\frac{d^mQ}{A'A} = \delta^{(m)}(q-q'). \tag{11}$$

We have restricted ourselves to coordinates having continuous ranges of possible values. Requirement (11) can be satisfied if we restrict $F(q,Q,t)$ to be of the form

$$F(q,Q,t) = a_{ij}q_iQ_j + g(q) + G(Q), \quad A = A' = [(2\pi\hbar)^m/\det a_{ij}]^{1/2}. \tag{12}$$

Just as the Lagrangian is restricted to forms quadratic in the generalized velocity for the path integral formulation, the generating functions must be restricted in form for the incorporation of canonical transformations into the method to be possible.

We can interpret the above in the spirit of the Feynman approach in the following fashion. In classical probability theory we can say that, if two variables x and y are statistically related, the probability for y to have the value y_1, $P(y_1)$, can be written as

$$P(y_1) = \sum_n P(x_n)\,P_c(y_1,x_n), \tag{13}$$

where $P(x_n)$ is the probability that x has the value x_n and $P_c(y_1,x_n)$ is the conditional probability that y has the value y_1 if x has the value x_n. In quantum mechanics we must use amplitudes so that

$$P(y_1) = |A(y_1)|^2,$$
$$A(y_1) = \sum_n A(x_n)\,A_c(y_1,x_n). \tag{14}$$

Thus we are postulating that the conditional probability amplitude for Q to have the value Q_1 if q has the value q_1 is essentially

$$(1/A)\exp[(-i/\hbar)F(q_1,Q_1,t)]. \tag{15}$$

We now wish to establish the correspondence between the quantum mechanical variables, Q_i, defined by (10), and the classical transformed coordinates $Q_i(q,\dot{q},t)$. We will do this by transforming the Feynman propagator by (10) and repeating Dirac's well-known classical limit argument to obtain (5). If the original propagator is

$$K(q_f,t_f,q_0,t_0) = \int \exp\left(\frac{i}{\hbar}\int_{t_0}^{t_f} L\,dt\right) Dq, \tag{16}$$

the transformed propagator is

$$K(Q_f,t_f,Q_0,t_0) = \frac{1}{A}\int \exp\left[\frac{i}{\hbar}\left(-F(q_f,Q_f,t_f) + \int_{t_0}^{t_f} L\,dt + F(q_0,Q_0,t_0)\right)\right] Dq\,d^mq_f\,d^mq_0. \tag{17}$$

We can express this as a new path integral $\bar{D}q$, where the q_i are varied at the end points as well as along the path. So

$$K(Q_f,t_f,Q_0,t_0) = (1/A)\int \exp\{(i/\hbar)[-F(q_f,Q_f,t_f) + \int L\,dt + F(q_0,Q_0,t_0)]\}\,\bar{D}q. \tag{18}$$

Now, if we let $\hbar$ become very small, we must get phase cancellation of the integrand for all paths except those in the neighborhood of the path that makes

$$\int_{t_0}^{t_f} L(q,\dot{q},t)\,dt - F(q_f,Q_f,t_f) + F(q_0,Q_0,t_0) \tag{19}$$

an extremum. This means that

$$\delta\left[\int_{t_0}^{t_f} L\,dt - F(q_f,Q_f,t_f) + F(q_0,Q_0,t_0)\right] = 0, \tag{20}$$

where the δ variation includes variation of the q_i at the end points. The variation (20) leads immediately to (5) at times t_0 and t_f and to the usual Lagrange equations for the q_i. In a similar manner consideration of the inverse transformation of the propagator leads to (6).

Although the class of transformations we can treat in the above manner is restricted, it does include two important cases, namely the transformation to the momentum representation and the infinitesimal time translation generated by the action Ldt. The first is obvious and corresponds to

$$F(q,Q,t) = q_iQ_i, \quad Q_i = P_i = \frac{\partial L}{\partial \dot{q}_i}. \tag{21}$$

The second is given by

$$F(q,Q,t) = \frac{\epsilon}{2}L\left(\frac{Q-q}{\epsilon},Q\right) + \frac{\epsilon}{2}L\left(\frac{Q-q}{\epsilon},q\right) \tag{22}$$

for Cartesian coordinates.[4] Equation (12) leads directly to the condition that L be quadratic in the velocity.

4. PATH INTEGRALS IN DIFFERENT REPRESENTATIONS

We now turn to another aspect of our discussion. It is easy to find examples of canonical transformations of the form (12) which connect Lagrangians which are not quadratic in the velocity with Lagrangians simpler in form. While the path integral quantization method cannot be applied directly to the former, if a connecting canonical transformation exists the simple Lagrangian

can be used in the quantization procedure and the propagator corresponding to the nonquadratic Lagrangian obtained by use of the corresponding unitary transformation given by (10). One could thus obtain, in an indirect fashion, propagators which are not of the Feynman form (16) and which correspond to Lagrangians which are not quadratic in the velocity.

We now give a simple illustration of this. Consider a Lagrangian of the form

$$L(q,\dot{q},t)=W(\dot{q})-V(q), \tag{23}$$

where W is an even but not quadratic function of the m coordinates q_i. This Lagrangian is simple enough to avoid the ordering problem, which we will discuss further on.

The canonical transformation which we consider is that which gives the momentum representation, that is,

$$F(q,p,t)=q_ip_i, \tag{24}$$

$$p_i=\frac{\partial W}{\partial \dot{q}_i}. \tag{25}$$

The new Lagrangian is then

$$\begin{aligned} &L'(p,\dot{p},q,t)=-[q_i\dot{p}_i+T(p)+V(q)],\\ &T(p)=p_i\dot{q}_i-W. \end{aligned} \tag{26}$$

We note that the q_i have been carried over as auxiliary variables, characteristic of Lagrangians giving first order equations of motion.

To better understand the role of the q_i in the path integral method, we consider the classical variational principle in some detail. The variation of the classical action is

$$\delta S=-\int_{t_0}^{t_f}\left[\left(\dot{p}_i+\frac{\partial V}{\partial q_i}\right)\delta q_i+\left(\frac{\partial T}{\partial p_i}-\dot{q}_i\right)\delta p_i\right]dt - [q_i\delta p_i]_{t_0}^{t_f}. \tag{27}$$

If we seek the classical trajectory between specified initial and final momenta, the variation of the p_i at the end points must vanish. However, as we shall see, the variation of the q_i cannot vanish at the end points. Since (27) contains no end point term involving δq_i, we can immediately write down the Lagrange equations which are

$$\dot{p}_i+\frac{\partial V}{\partial q_i}=0,\quad \dot{q}_i=\frac{\partial T}{\partial p_i}. \tag{28}$$

Specifying the p_i at the end points determines the solution to these equations including the $q_i(t)$. Thus the q_i cannot be independently specified at the end points and the correct classical trajectory can only be obtained by varying the q_i at the end points as well as along the path.

When we define the propagator as a sum over classical histories, we thus consider it as the amplitude that the system will go from an initial point to a final point in the space of the p_i's only, including paths going from all possible initial q_i's to all final ones. With that understanding, we write the propagator as

$$K(p_f,t_f,p_0,t_0)=\int \exp\left\{\frac{-i}{\hbar}\int_{t_0}^{t_f}[q_i\dot{p}_i+T(p)+V(q)]\,dt\right\}Dq\,Dp. \tag{29}$$

When we break the time interval into infinitesimal segments in the usual fashion to evaluate the propagator, we cannot approximate the action by its classical value for an arbitrary path because of (28). Instead, we write the action in the following form

$$\begin{aligned} &S[p(t+\epsilon),t+\epsilon,p(t),t]\\ &\quad=\int_t^{t+\epsilon}[q_i\,dp_i+T(p)\,dt+V(q)\,dt]\\ &\quad\simeq\bar{q}_i[p_i(t+\epsilon)-p_i(t)]+[\bar{T}(p)+V(\bar{q})]\epsilon, \end{aligned} \tag{30}$$

where $\bar{q}_i$ and $\bar{T}$ are some average values for the time interval which we leave unspecified for the moment. The propagator is then

$$\begin{aligned} K(p_f,t_f,p_0,t_0)&=\lim_{\epsilon\to 0}\int \exp\left(\frac{-i}{\hbar}\sum_{n=0}^{N-1}\bar{q}_i^{(n)}(p_i^{(n+1)}-p_i^{(n)})\right.\\ &\left.+\epsilon[\bar{T}(p^{(n+1)},p^{(n)})+V(\bar{q}^{(n)})]\right)\prod_{n=0}^{N-1}d^mp^{(n)}\,\frac{d^mq^{(n)}}{A}\,\frac{d^mq_0}{A}, \end{aligned} \tag{31}$$

where $p_i^{(0)}=p_i(t_0), p_i^{(n)}=p_i(t_n)$, etc. It is sufficient for our purposes to consider just the propagator for an infinitesimal time interval which is

$$\begin{aligned} K(p,t+\epsilon,p',t)&=\int \exp\{-i/\hbar)[\bar{q}_i(p_i-p_i')+\epsilon\bar{T}(p,p')+\epsilon V(\bar{q})]\}\,d\bar{q}/A\\ &=\int \exp[(-i/\hbar)\bar{q}_ip_i]\exp\{-i/\hbar)[\bar{T}(p,p')+V(\bar{q})]\epsilon\}\\ &\quad\times\exp[(i/\hbar)\bar{q}_ip_i']\,d\bar{q}/A. \end{aligned} \tag{32}$$

Choosing $T=\frac{1}{2}[T(p)+T(p')]$ and noting that $\bar{q}$ is now a dummy variable, we get

$$\begin{aligned} K(p,t+\epsilon,p',t)&=\exp[(-i\epsilon/2\hbar)\,T(p)]\int \exp[(-i/\hbar)\,\bar{q}_ip_i)\exp[-i\epsilon/\hbar)V(\bar{q})]\\ &\quad\times\exp[(i/\hbar)\,\bar{q}_ip_i'][d\bar{q}/(2\pi\hbar)^m]\exp[(-i\epsilon/2\hbar)\,T(p')], \end{aligned} \tag{33}$$

where unitarity dictates that $A=(2\pi\hbar)^m$.

Returning to the coordinate representation via (10), which, of course, is now just the Fourier transform, we obtain

$$\begin{aligned} K(q,t+\epsilon,q',t)&=\int \exp[(i/\hbar)\,p_iq_i]\exp[(-i\epsilon/2\hbar)\,T(p)]\exp[(-i/\hbar)\,p_i\bar{q}_i]\\ &\quad\times\exp[(-i\epsilon/\hbar)\,V(\bar{q})]\exp[(i/\hbar)\,\bar{q}_ip_i']\exp[(-i\epsilon/2\hbar)\,T(p')]\\ &\quad\times\exp[(i/\hbar)\,p_i'q_i']\,d^mp\,d^mp'\,d^m\bar{q}/(2\pi\hbar)^{2m}. \end{aligned} \tag{34}$$

This propagator is not of the Feynman form and can only be put in that form when $T(p)$ is a quadratic function. In that case, the integrations in (34) can easily be done, and the procedure is straightforward.

We note here that since, in the above calculation, the action for the infinitesimal time interval was not approximated by its classical value (indeed, could not be), the results (33) and (34) could be obtained even for the free particle case, i.e., $V(q)=0$, where a classical action does not in general exist.[5] What is implied here is that the notion of an integral over paths can be implemented even in cases where there is no path that makes the action an extremum. For such cases, resolution of the ordering ambiguity by postulating that the action must be approximated by its classical value[6] is impossible. The ambiguities that occur when the above momentum space quantization procedure is applied to Lagrangians more general than (23) correspond to those characteristic of the canonical quantization procedure. This can be seen by noting the formal resemblance of the propagator (29) to the canonical phase space path integral which is often used.[2,7] The analysis of Cohen[7] then essentially applies here also.

[1]R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948); G. Rosen, *Formulation of Classical and Quantum Dynamical Theory* (Academic, New York, 1969); R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).

[2]A. Katz, *Classical Mechanics, Quantum Mechanics, Field Theory* (Academic, New York, 1965).

[3]P. A. M. Dirac, *Principles of Quantum Mechanics* (Oxford, U. P., London, 1958), 4th ed.

[4]R. P. Feynman, Ref. 1.

[5]B. S. DeWitt, Rev. Mod. Phys. **29**, 377 (1957).

[6]K. S. Cheng, J. Math. Phys. **14**, 980 (1973).

[7]E. Kerner and W. Sutcliffe, J. Math. Phys. **11**, 391 (1970); L. Cohen, J. Math. Phys. **11**, 3296 (1970); I. W. Mayes and J. S. Dowker, J. Math. Phys. **14**, 434 (1973).

A variational principle for the Boltzmann equation for hot electrons in a semiconductor

W. A. Schlup

IBM Zürich Research Laboratory, 8803 Rüschlikon, Switzerland
(Received 10 September 1974; revised manuscript received 4 February 1975)

It is shown that the well-known Kohler variational principle for the small electric-field solution of the Boltzmann equation can be extended to arbitrary fields, if a generalization of Hamilton's principle proposed by Djukic and Vujanovic is used.

Variational principles determine the solution of a problem given by an equation by extremumizing an action function $S=\int d\Omega\, \mathcal{L}$ with certain conditions on the boundaries, which make partially integrated terms vanish. The infinitesimal volume $d\Omega$ goes over all independent variables of the equation; in the general case considered here it is $d\Omega = dt\, d^3k\, d^3x$.

Kohler's variational principle[1-3] deals with the first-order solution $\phi_{\mathbf{k}}$ of the Boltzmann equation for small electric fields $\mathbf{F}$,

$$g_{\mathbf{k}} = \int d^3k' (W_{\mathbf{k}\mathbf{k}'}\, \phi_{\mathbf{k}'} - W_{\mathbf{k}'\mathbf{k}}\, \phi_{\mathbf{k}}), \tag{1}$$

where $g_{\mathbf{k}} = \mathbf{F}(\partial h_{\mathbf{k}}/\partial \mathbf{k})$, $h_{\mathbf{k}} = C\exp(-E_{\mathbf{k}}/T)$ is the Maxwell–Boltzmann distribution and the transition-rate product $W_{\mathbf{k}\mathbf{k}'}\, h_{\mathbf{k}'}$ is symmetric. Extensions for a magnetic field and for frequency-dependent conductivity[4-6] are known (for a Lagrangian formulation see the Appendix). Upper and lower bounds for the transport coefficients can be found for Hermitian[7] and non-Hermitian[8] collision operators.

By using a power-series expansion with respect to the electric field the variational solution of the Boltzmann equation has been formulated[9] and discussed for the lowest field-dependent term in the conductivity[10] by Adawi.

It is not possible to find a classical Lagrangian (CL) for the hot-electron Boltzmann equation, since it is of first order in $\partial/\partial t$ for the transient case and also in $\partial/\partial \mathbf{k}$ for the stationary case. For first-order equations, there exist no Lagrangians, or the calculus of variations generates even-order equations only.[11]

There are limit Lagrangians (LL), which produce odd-order equations, e.g., for $\dot{x}+x=0$, no Cl exist, but $L_\mu(x,\dot{x},t) = (\mu\dot{x}^2/2 + x^2/2)\cdot\exp(-t/\mu)$ gives the upper equation in the limit $\mu \to 0$. It is highly probable that there exists no LL, neither for the time-dependent nor for the stationary hot-electron problem [see remark after Eq. (14)].

Djukic and Vujanovic[12] proposed a formal method, which allows to find Lagrangians, hereafter shortly named DV Lagrangians (DVL) for a much wider class of equations. The method is especially adapted for first-order equations, or second-order equations, which include (complicated) first-order terms (damping terms) like the Navier–Stokes equations, for which no Cl or LL is known.

The DVL for a hot-electron system is ($\alpha,\ \beta = 1, 2, 3$)

$$\mathcal{L} = \frac{1}{2h_{\mathbf{k}}}\left[\psi_0(t,\lambda)\left(\frac{\partial f}{\partial t}\right)^2 + \sum_\alpha \psi_{1\alpha}(\mathbf{k},\lambda)\, F_\alpha \left(\frac{\partial f}{\partial \mathbf{k}}\right)^2 + \sum_\alpha \psi_{2\alpha}(\mathbf{x},\lambda)\, v_{\mathbf{k}\alpha}\left(\frac{\partial f}{\partial \mathbf{x}}\right)^2 + \int d^3k'\, W_{\mathbf{k}\mathbf{k}'}\, f_{\mathbf{k}} f_{\mathbf{k}'} - \nu_{\mathbf{k}} f^2\right]. \tag{2}$$

The first derivatives are

$$\frac{\partial \mathcal{L}}{\partial(\partial f/\partial t)} = \frac{1}{h_{\mathbf{k}}}\,\psi_0(t,\lambda)\,\frac{\partial f}{\partial t}, \tag{3}$$

$$\frac{\partial \mathcal{L}}{\partial(\partial f/\partial k_\beta)} = \frac{1}{h_{\mathbf{k}}}\sum_\alpha \psi_{1\alpha}(\mathbf{k},\lambda) F_\alpha \frac{\partial f}{\partial k_\beta}, \tag{4}$$

$$\frac{\partial \mathcal{L}}{\partial(\partial f/\partial x_\beta)} = \frac{1}{h_{\mathbf{k}}}\sum_\alpha \psi_{2\alpha}(\mathbf{x},\lambda)\, v_{\mathbf{k}\alpha}\frac{\partial f}{\partial x_\beta}. \tag{5}$$

The functional derivative neglecting all terms with f derivatives $\delta(S_{\partial f=0})/\delta f$ ($=\partial\mathcal{L}/\partial f$ for Lagrangians without integrals) is

$$\frac{\delta(S_{\partial f=0})}{\partial f} = \frac{1}{h_{\mathbf{k}}}\left(\int d^3k'\, W_{\mathbf{k}\mathbf{k}'} f_{\mathbf{k}'} - \nu_{\mathbf{k}} f\right). \tag{6}$$

If we assume $\psi_1(\mathbf{k},\lambda) = [\psi_{11}(k_1,\lambda),\ \psi_{12}(k_2,\lambda),\ \psi_{13}(k_3,\lambda)]$ and analogously $\psi_2(\mathbf{x},\lambda) = [\psi_{21}(x_1,\lambda),\ \psi_{22}(x_2,\lambda),\ \psi_{23}(x_3,\lambda)]$, their derivatives become in the limit

$$\lim_{\lambda\to 0}\frac{\partial\psi_{1\alpha}(\mathbf{k},\lambda)}{\partial k_\beta} = \delta_{\alpha\beta}, \quad \lim_{\lambda\to 0}\frac{\partial\psi_{2\alpha}(\mathbf{x},\lambda)}{\partial x_\beta} = \delta_{\alpha\beta}, \tag{7}$$

since according to Djukic and Vujanovic the auxiliary functions ψ have the properties

$$\lim_{\lambda\to 0}\psi(z,\lambda) = 0, \quad \lim_{\lambda\to 0}\frac{\partial\psi}{\partial z}(z,\lambda) = 1. \tag{8}$$

The second derivatives therefore become in the limit $\lambda \to 0$:

$$\lim_{\lambda\to 0}\frac{d}{dt}\,\frac{\partial\mathcal{L}}{\partial(\partial f/\partial t)} = \frac{1}{h_{\mathbf{k}}}\,\frac{\partial f}{\partial t}, \tag{9}$$

$$\lim_{\lambda\to 0}\frac{d}{d\mathbf{k}}\,\frac{\partial\mathcal{L}}{\partial(\partial f/\partial \mathbf{k})} = \frac{1}{h_{\mathbf{k}}}\,\mathbf{F}\cdot\frac{\partial f}{\partial \mathbf{k}}, \tag{10}$$

$$\lim_{\lambda\to 0}\frac{d}{d\mathbf{x}}\,\frac{\partial\mathcal{L}}{\partial(\partial f/\partial \mathbf{x})} = \frac{1}{h_{\mathbf{k}}}\,\mathbf{v}_{\mathbf{k}}\cdot\frac{\partial f}{\partial \mathbf{x}}. \tag{11}$$

Finally, the Euler–Lagrange equation in the limit $\lambda \to 0$ yields

$$\frac{\partial f}{\partial t} + \mathbf{F}\frac{\partial f}{\partial \mathbf{k}} + \mathbf{v}_{\mathbf{k}}\frac{\partial f}{\partial \mathbf{x}} = \int d^3k' W_{\mathbf{k}\mathbf{k}'} f_{\mathbf{k}'} - \nu_{\mathbf{k}} f, \tag{12}$$

where $\nu_{\mathbf{k}} = \int d^3k' W_{\mathbf{k}'\mathbf{k}}$ is the total relaxation frequency. This is the hot-electron Boltzmann equation. A DV Lagrangian for the space-independent equation is obtained formally by putting $\psi_2(\mathbf{x}, \lambda) \equiv 0$. A DVL for the stationary solution is similarly found by putting $\psi_0(t, \lambda) \equiv 0$ and $\psi_2(\mathbf{x}, \lambda) \equiv 0$. In the zero-field case $F = 0$ a limit Lagrangian can be constructed:

$$\begin{aligned} \mathcal{L} = \left[\exp\left(\frac{t}{\beta} + \frac{x_1}{\gamma_1} + \frac{x_2}{\gamma_2} + \frac{x_3}{\gamma_3}\right) \Big/ 2h_{\mathbf{k}}\right] \left[\beta\left(\frac{\partial f}{\partial t}\right)^2 + \gamma_1 v_{1\mathbf{k}}\left(\frac{\partial f}{\partial x_1}\right)^2 \right. \\ + \gamma_2 v_{2\mathbf{k}}\left(\frac{\partial f}{\partial x_2}\right)^2 + \gamma_3 v_{3\mathbf{k}}\left(\frac{\partial f}{\partial x_3}\right)^2 \\ \left. + \int d^3k' W_{\mathbf{k}\mathbf{k}'} f_{\mathbf{k}} f_{\mathbf{k}'} - \nu_{\mathbf{k}} f^2\right]. \end{aligned} \tag{13}$$

By standard differentiation the Euler–Lagrange equation becomes

$$\begin{aligned} \beta\frac{\partial^2 f}{\partial t^2} + \frac{\partial f}{\partial t} + \mathbf{v}_{\mathbf{k}}\frac{\partial f}{\partial \mathbf{x}} + \gamma_1 v_{1\mathbf{k}}\frac{\partial^2 f}{\partial x_1^2} + \gamma_2 v_{2\mathbf{k}}\frac{\partial^2 f}{\partial x_2^2} + \gamma_3 v_{3\mathbf{k}}\frac{\partial^2 f}{\partial x_3^2} \\ = \int d^3k' W_{\mathbf{k}\mathbf{k}'} f_{\mathbf{k}'} - \nu_{\mathbf{k}} f, \end{aligned} \tag{14}$$

which in the limit $\beta \to 0$, γ_1, γ_2, $\gamma_3 \to 0$ goes into the Boltzmann equation (12) with $\mathbf{F} = 0$. This procedure is not possible for $\mathbf{F} \neq 0$, since an additional factor $\exp(\mathbf{F}\cdot\mathbf{k}/\alpha)$ would be necessary in order to produce $\mathbf{F}(\partial f/\partial \mathbf{k})$ in the limit $\alpha \to 0$. But such a $\mathbf{k}$-dependent factor would destroy the symmetry of the $W_{\mathbf{k}\mathbf{k}'}/h_{\mathbf{k}}$ and generate a field-dependent collision term in the corresponding Euler–Lagrange equation.

It is easy to show that for a certain class of Lagrangians the variational method of Djukic and Vujanovic coincides with the classical calculus of variations.[13] In other words, there are equations for which it is possible to find DVL's and also LL's (e.g., the heat equation $\dot{u} = \overset{\shortparallel}{u}$) or even CL's (e.g., the damped-wave equation $\mu\ddot{u} + \dot{u} = \overset{\shortparallel}{u}$).

Other variational methods for transport problems have been discussed by Cercignani[14] for the linearized Boltzmann equation in kinetic theory (no field) and by Pomraning and Clark[15] for the monoenergetic neutron-transport equation. Lagrangians (LL's for the transient and CL's for steady state) are also known for the diffusion equation and the (formally equivalent) Fokker–Planck equation. A very special variational method has been developed by Biot[16] for the heat-transport equation, which does not make use of any limits $\mu \to 0$ (LL's) or $\lambda \to 0$ (DVL's) in order to produce the first-order term in time. It would be interesting to find out whether other first-order transport equations can also be treated with Biot's method.

APPENDIX

In a magnetic field B the Boltzmann equation is

$$\frac{\partial f}{\partial t} + (\mathbf{F} + \mathbf{v}_{\mathbf{k}} \times \mathbf{B})\frac{\partial f}{\partial \mathbf{k}} + \mathbf{v}_{\mathbf{k}}\frac{\partial f}{\partial \mathbf{k}} = \int d^3k' W_{\mathbf{k}\mathbf{k}'} f_{\mathbf{k}'} - \nu_{\mathbf{k}} f. \tag{A1}$$

For a small periodically space- and time-dependent electric field $\mathbf{F} \propto \exp(i\mathbf{q}\mathbf{x} - i\omega t)$ the solution $f_{\mathbf{k}} = h_{\mathbf{k}} + \phi_{\mathbf{k}} \times \exp(i\mathbf{q}\mathbf{x} - i\omega t)$ follows from

$$\int d^3k' W_{\mathbf{k}\mathbf{k}'} \phi_{\mathbf{k}'} - \nu_{\mathbf{k}}\phi + \mathbf{B}\mathbf{v}_{\mathbf{k}} \times \frac{\partial \phi}{\partial \mathbf{k}} + i(\omega - \mathbf{q}\mathbf{v}_{\mathbf{k}})\phi = F\frac{\partial h_{\mathbf{k}}}{\partial \mathbf{k}}. \tag{A2}$$

The Hermitian adjoint $\psi = \phi^*$ follows from

$$\int d^3k' W_{\mathbf{k}\mathbf{k}'} \psi_{\mathbf{k}'} - \nu_{\mathbf{k}}\psi - \mathbf{B}\mathbf{v}_{\mathbf{k}} \times \frac{\partial \psi}{\partial \mathbf{k}} - i(\omega - \mathbf{q}\mathbf{v}_{\mathbf{k}})\psi = F\frac{\partial h_{\mathbf{k}}}{\partial \mathbf{k}}. \tag{A3}$$

Equations (A2) and (A3)* are the Euler–Lagrange equations of the Lagrangian

$$\begin{aligned} \mathcal{L} = \frac{1}{h_{\mathbf{k}}}\left[\psi_{\mathbf{k}}^*\left(\int d^3k' W_{\mathbf{k}\mathbf{k}'} \phi_{\mathbf{k}'} - \nu_{\mathbf{k}}\phi_{\mathbf{k}} + \mathbf{B}\mathbf{v}_{\mathbf{k}} \times \frac{\partial \phi_{\mathbf{k}}}{\partial \mathbf{k}}\right.\right. \\ \left.\left. + i(\omega - \mathbf{q}\mathbf{v}_{\mathbf{k}})\phi_{\mathbf{k}}\right) - \mathbf{F}\frac{\partial h_{\mathbf{k}}}{\partial \mathbf{k}}(\phi_{\mathbf{k}} + \psi_{\mathbf{k}}^*)\right] + \text{c.c.} \end{aligned} \tag{A4}$$

Kohler's variational principle in Lagrangian form is: Extremumize S with respect to the complex functions $\phi_{\mathbf{k}}$, $\psi_{\mathbf{k}}$:

$$S = \int d^3k\, \mathcal{L}(\phi_{\mathbf{k}}, \psi_{\mathbf{k}}^*),$$

$$\delta S = \int d^3k (\mathcal{L}_{\phi_{\mathbf{k}}}\delta\phi_{\mathbf{k}} + \mathcal{L}_{\psi_{\mathbf{k}}^*}\delta\psi_{\mathbf{k}}^*) = 0.$$

Then $\mathcal{L}_{\psi_{\mathbf{k}}} = 0$ implies Eq. (A2) and $(\mathcal{L}_{\phi_{\mathbf{k}}})^* = 0$ implies (A3).

[1] J. M. Ziman, *Electrons and Phonons* (Oxford U. P., London, 1960), pp. 275 and 508.
[2] M. Kohler, Z. Physik 124, 772 (1948).
[3] E. H. Sondheimer, Proc. Roy. Soc. (London) A 203, 75 (1950).
[4] F. Garcia-Moliner and S. Simons, Proc. Cambridge Phil. Soc. 53, 848 (1957).
[5] N. K. Hindley and F. Garcia-Moliner, Phys. Rev. A 140, 343 (1965).
[6] M. Bailyn, Phys. Rev. 126, 2040 (1962).
[7] M. Højgaard-Jensen, H. Smith, and J. W. Wilkins, Phys. Rev. 185, 323 (1969).
[8] L. E. G. Ah-Sam, H. Højgaard-Jensen, and H. Smith, J. Stat. Phys. 3, 17 (1971).
[9] I. Adawi, Phys. Rev. 115, 1152 (1959).
[10] I. Adawi, Phys. Rev. 120, 118 (1960).
[11] B. A. Finlayson, *The Method of Weighted Residuals and Variational Principles* (Academic, New York, 1972).
[12] D. Djukic and B. Vujanovic, Z. Angew. Math. Mech. 51, 611 (1971).
[13] W. A. Schlup (to be published).
[14] C. Cercignani, J. Stat. Phys. 1, 297 (1969).
[15] G. C. Pomraning and M. Clark, Jr., Nucl. Sci. Eng. 16, 144, 147 (1963).
[16] M. A. Biot, *Variational Principle in Heat Transfer* (Oxford U. P., London, 1970).

On Ursell's combinatorial problem

A. D. MacGillivray

Department of Mathematics, State University of New York at Buffalo, Amherst, New York 14226
(Received 13 January 1975)

A combinatorial problem considered by H. D. Ursell in his seminal paper on cluster theory [Proc. Camb. Phil. Soc. 23, 685 (1927)] is studied. Ursell's analysis, which is not rigorous, is described by Fowler and Guggenheim as being far from simple. In this paper we arrive at Ursell's result using a method which is straightforward, yet completely rigorous.

INTRODUCTION

In 1927, H.D. Ursell[1] wrote a paper which came to be accepted as a seminal paper in cluster theoretic methods in statistical mechanics.[2–5]

In Ursell's paper, combinatorial problems associated with the evaluation of the phase integral are set up, leading to polynomials of large degree (comparable with Avogadro's number) which are to be summed. The simplest problem of this sort treated by Ursell (and for the present we consider only this problem) was considered earlier by others, notably Jeans[6] and Fowler.[7] These earlier analyses are however described by Fowler and Guggenheim[8] as being "fallacious," and these latter authors attribute the correct solution to Ursell.

Ursell's method of summing the polynomial, or rather of finding an appropriate asymptotic representation of the sum, is to show first that the function represented by the series satisfies a certain linear, second order differential equation, and then to find an appropriate approximate solution of this differential equation.

Fowler and Guggenheim[8] describe Ursell's as being "far from simple." This is probably due to the fact that the analysis includes steps which are not justified. We can identify some of these questionable steps by noting that Ursell's procedure is to treat the problem as a perturbation problem in the small parameter $\epsilon = 1/N$, where N is the number of systems in the statistical mechanical model. Now, when Ursell neglects certain terms in the differential equation because $1/N$ is small, one of the terms so eliminated is the second derivative. This is one of the best known means of arriving at a singular perturbation problem.[9] The usual circumstance is that the original second order differential equation involves two auxiliary conditions, usually boundary or initial conditions, whereas the first order differential equation can require only one auxiliary condition. Now, Ursell's differential equation has a singularity at the origin, the nature of which is such that precisely one solution is analytic and satisfies one initial condition; any second initial condition is necessarily redundant. Hence the omission of the second derivative in the present problem does not seem to eliminate an auxiliary condition. Nevertheless, the known complications that arise in the usual case when the highest derivative is dropped should move us to look for justification for doing so in Ursell's problem.

The referee of the present paper has shown very nicely how the eikonal approximation can be used to clarify the reason why Ursell's original neglect of the second derivative term did not prevent his obtaining a correct answer. This approach, although not completely rigorous, clarifies Ursell's derivation substantially.

Professor A. Erdélyi has kindly communicated to me that Ursell's polynomial, Eq. (1), can, after some changes of variable, be expressed as a Hermite polynomial, and known asymptotic representations can be used to obtain rigorous justification for Ursell's final answer. This approach is perhaps less than satisfying since it gives little insight into the fundamental nature of Ursell's equation, unless of course one feels such insight can be gained by going through the extremely intricate derivation of the asymptotic expansions needed.[10]

The advantages of the present analysis are that it proceeds directly from Ursell's differential equation, it is short and straightforward, and yet it is completely rigorous. In addition to these features, the fact that the differential equation happens to be linear is not exploited. This suggests that our approach has wider application for the study of differential equations. Indeed, from a fundamental point of view, our method is more in the spirit of some recent work on nonlinear problems.[11]

We return to the remarks from the referee and Erdélyi after our analysis, to which we now turn.

ANALYSIS

The polynomial obtained by Ursell[1,8] is

$$F_N(x) = \sum_{r=0}^{N/2} \frac{N!}{r!\,(N-2r)!}\,\frac{x^r}{N^r}\,. \tag{1}$$

In Eq. (1), x is negative and, being related to the volume of influence of a single system and to the density of the assembly, is regarded as small. N is the number of such systems in the assembly. Thus, for the statistical mechanical application, one is interested in the behavior of $F_N(x)$ as $N \to \infty$, with x small but fixed.

As mentioned in the Introduction, Jeans[6] and Fowler[7] had earlier approximated Eq. (1). They did this by keeping only the first two terms,

$$\begin{aligned} F_N(x) &\approx 1 + \frac{N!}{(N-2)!}\,\frac{x}{N} \\ &= 1 + (N-1)x. \end{aligned} \tag{2}$$

If now N is large but is regarded as fixed, and x, instead of being small but fixed is permitted to approach zero, one obtains from (2) the asymptotic estimate

$$(1/N)\ln F_N(x) \approx (1/N)(N-1)x \approx x. \tag{3}$$

However, as Ursell pointed out, the ratio of the neglected third term to the second term retained in (2) is

$$\tfrac{1}{2}(1-2/N)(N-3)x, \tag{4}$$

which, for fixed x and $N\to\infty$ (i.e., for the physically relevant case) forces one to admit that the third term cannot be ignored. Ursell then proceeds to derive an approximate solution of Eq. (1) which leads, surprisingly, to the same estimate, Eq. (3).

Our point of departure is the differential equation Ursell showed is satisfied by F_N. It can be easily verified that F_N satisfies

$$4x^2\frac{d^2F}{dx^2} = \frac{dF}{dx}(N+4Nx-6x) - N(N-1)F, \quad x<0, \tag{5}$$

with initial conditions

$$F(0)=1, \tag{6}$$

$$\left.\frac{dF}{dx}\right|_{x=0} = N-1. \tag{7}$$

As was stated in the Introduction, Eq. (7) can be regarded as redundant. Letting $\epsilon=1/N$, we arrive at

$$4x^2\epsilon^2\frac{d^2F}{dx^2} = \frac{dF}{dx}(\epsilon+4\epsilon x-6x\epsilon^2) - (1-\epsilon)F, \quad x<0, \tag{8}$$

$$F(0)=1, \tag{9}$$

$$\left.\frac{dF}{dx}\right|_{x=0} = \frac{1-\epsilon}{\epsilon}. \tag{10}$$

Ursell's analysis leads to (see Fowler and Guggenheim,[8] Eq. 703,11)

$$F(x,N) = \{1+x+O(x^2)\}^N, \tag{11}$$

whose behavior is very different from the approximation described by Eq. (2).

Our analysis, which we now begin, will show that Ursell's result (11) is correct in a sense to be made precise.

We begin by letting $F(x;\epsilon)$ denote the solution of Eqs. (8), (9), (10), and define $g(x;\epsilon)$ as follows:

$$F(x;\epsilon) = \exp[(1-\epsilon)x/\epsilon]\,[1+g(x;\epsilon)]. \tag{12}$$

We find that $g(x;\epsilon)$ must satisfy the following differential equation and boundary conditions:

$$\frac{d^2g}{dx^2} = \frac{1}{4x^2\epsilon^2}\left(\epsilon\frac{dg}{dx}[1+x(4-6\epsilon)-8x^2(1-\epsilon)] + \frac{(1+g)}{\alpha}(1-\epsilon)[\alpha x(4-6\epsilon)-\alpha 4x^2(1-\epsilon)]\right), \quad x<0, \tag{13}$$

$$g(0)=0, \tag{14}$$

$$\left.\frac{dg}{dx}\right|_{x=0} = 0, \tag{15}$$

where α is an arbitrary constant exceeding unity. Its purpose will be explained shortly.

In what follows, we shall always assume N is an integer at least as large as 2, and hence $\epsilon=1/N$ will be a number in the interval $(0,\frac{1}{2}]$. Eventually, of course, we are interested in ϵ being small.

Proposition 1: For all sufficiently small $|x|$, depending on $\alpha>1$ but independent of $\epsilon=1/N$,

$$g(x;\epsilon) \geq -1+\exp[(1-\epsilon)x/\epsilon\alpha], \quad x<0. \tag{16}$$

Proof: Choose $|x_\alpha|$ small enough to ensure that $x\in[x_\alpha,0)$ implies

$$1+x(4-6\epsilon)-8x^2(1-\epsilon) > \tfrac{1}{2}, \tag{17}$$

$$\alpha x(4-6\epsilon)-\alpha 4x^2(1-\epsilon) > \tfrac{1}{4}, \tag{18}$$

and

$$1/16x^2 > 1. \tag{19}$$

Note that x_α can be chosen independently of ϵ.

For convenience, let

$$f_\alpha(x;\epsilon) = -1+\exp[(1-\epsilon)x/\epsilon\alpha]. \tag{20}$$

Next, assume tentatively that there exists a point $\bar{x}_\alpha$, $x_\alpha \leq \bar{x}_\alpha < 0$, where

$$g(\bar{x}_\alpha;\epsilon) < -1+\exp[(1-\epsilon)\bar{x}_\alpha/\epsilon\alpha] = f_\alpha(\bar{x}_\alpha;\epsilon). \tag{21}$$

Without loss of generality, we can assume that, at $\bar{x}_\alpha$, g has a slope exceeding that of f_α. That is,

$$\left.\frac{dg}{dx}\right|_{\bar{x}_\alpha} > \left.\frac{df_\alpha}{dx}\right|_{\bar{x}_\alpha}. \tag{22}$$

This follows from the fact that the boundary conditions on g together with (21) force the graph of g to cross that of f_α. One can then apply the mean value theorem to find $\bar{x}_\alpha$, and indeed such considerations allow us to assume that

$$1+g(\bar{x}_\alpha) > 0. \tag{23}$$

Combining (21) and (22), we have

$$\epsilon\left.\frac{dg}{dx}\right|_{\bar{x}_\alpha} > \epsilon\left.\frac{df_\alpha}{dx}\right|_{\bar{x}_\alpha} = \frac{1-\epsilon}{\alpha}[1+f_\alpha(\bar{x}_\alpha)] > \frac{1-\epsilon}{\alpha}[1+g(\bar{x}_\alpha)]. \tag{24}$$

Then, using (17), (18), (22), (23), we have, at $\bar{x}_\alpha$,

$$\epsilon\frac{dg}{dx}[1+x(4-6\epsilon)-8x^2(1-\epsilon)] + \frac{(1+g)}{\alpha}(1-\epsilon)[\alpha x(4-6\epsilon)-\alpha 4x^2(1-\epsilon)] > \frac{\epsilon}{2}\frac{dg}{dx} - \frac{(1+g)(1-\epsilon)}{4\alpha}. \tag{25}$$

By combining (25) with (24), and using (19), Eq. (13) yields

$$\left.\frac{d^2g}{dx^2}\right|_{\bar{x}_\alpha} > \frac{1}{4\epsilon^2\bar{x}_\alpha^2}\left(\frac{\epsilon}{2}\left.\frac{dg}{dx}\right|_{\bar{x}_\alpha} - \frac{\epsilon}{4}\left.\frac{dg}{dx}\right|_{\bar{x}_\alpha}\right) > \frac{1}{\epsilon}\left.\frac{dg}{dx}\right|_{\bar{x}_\alpha}. \tag{26}$$

We now restrict ourselves to the interval to the right of $\bar{x}_\alpha$. Certainly (24) and continuity imply

$$\epsilon \frac{dg}{dx} - \frac{1-\epsilon}{\alpha}[1+g(x)] \geq 0 \tag{27}$$

is valid on *some* interval to the right of $\bar{x}_\alpha$. Furthermore, so long as (27) holds, we shall have

$$\frac{d^2g}{dx^2} > \frac{1}{\epsilon}\frac{dg}{dx} \tag{28}$$

and

$$\frac{dg}{dx} > 0. \tag{29}$$

Let us therefore study the behavior of

$$\epsilon \frac{dg}{dx} - \frac{(1-\epsilon)}{\alpha}[1+g(x)]. \tag{30}$$

So long as (30) is nonnegative, we shall have

$$\begin{aligned}
&\epsilon \frac{dg}{dx} - \frac{1-\epsilon}{\alpha}[1+g(x)] \\
&= \left(\epsilon \frac{dg}{dx}\bigg|_{\bar{x}_\alpha} + \epsilon \int_{\bar{x}_\alpha}^{x} \frac{d^2g}{d\xi^2}\, d\xi\right) \\
&\quad - \frac{1-\epsilon}{\alpha}\left(1+g(\bar{x}_\alpha) + \int_{\bar{x}_\alpha}^{x} \frac{dg}{d\xi}\, d\xi\right) \\
&\geq \left(\epsilon \int_{\bar{x}_\alpha}^{x} \frac{1}{\epsilon}\frac{dg}{d\xi}\, d\xi - \frac{1-\epsilon}{\alpha}\int_{\bar{x}_\alpha}^{x} \frac{dg}{d\xi}\, d\xi\right) \\
&\quad + \left(\epsilon \frac{dg}{dx}\bigg|_{\bar{x}_\alpha} - \frac{1-\epsilon}{\alpha}[1+g(\bar{x}_\alpha)]\right) \\
&> \frac{\alpha-1+\epsilon}{\alpha}\int_{\bar{x}_\alpha}^{x} \frac{dg}{d\xi}\, d\xi \\
&> 0.
\end{aligned} \tag{31}$$

[This last step uses (29) and $\alpha > 1$.]

Thus, so long as the *weak* inequality (27) holds, the *strict* inequality

$$\epsilon \frac{dg}{dx} - \frac{1-\epsilon}{\alpha}[1+g(x)] > 0 \tag{32}$$

holds, which, of course, means (32) holds for *all* x to the right of $\bar{x}_\alpha$, in particular at $x=0$. But the boundary conditions (14) and (15) force (32) to fail at $x=0$, and we then have the desired contradiction to our tentative assumption. This completes the proof.

Proposition 2: For all sufficiently small $|x|$, depending on $\alpha > 1$ but independent of $\epsilon = 1/N$,

$$g(x;\epsilon) \leq 0. \tag{33}$$

Proof: First choose x_α as in Proposition 1. We then assume tentatively that there exists a point, denoted $\bar{x}_\alpha$, such that $x_\alpha \leq \bar{x}_\alpha < 0$ and

$$g(\bar{x}_\alpha;\epsilon) > 0. \tag{34}$$

In fact, because of the boundary condition $g(0)=0$, we may in addition assume $\bar{x}_\alpha$ is chosen such that

$$\frac{dg}{dx}\bigg|_{\bar{x}_\alpha} < 0. \tag{35}$$

We then see that Eq. (13) yields

$$\frac{d^2g}{dx^2}\bigg|_{\bar{x}_\alpha} < 0, \tag{36}$$

and clearly dg/dx will remain negative as x increases to the right of $\bar{x}_\alpha$ at least so long as $g \geq -1$. Hence either $x=0$ is reached before $g(x)$ reaches -1, in which case the boundary condition (15) would be inconsistent with (35), or else there is some x_0, $x_\alpha < x_0 < 0$, where $g(x_0) = -1$, which is impossible because of Proposition 1. Hence (34) is an untenable assumption and our proof is finished.

From (12), (16), (33), we have immediately

Proposition 3: For all sufficiently small $|x|$ depending on $\alpha > 1$ but independent of $\epsilon = 1/N$ we have

$$\{\exp[1-\epsilon)x(1+1/\alpha)]\}^{1/\epsilon} \leq F(x;\epsilon) \leq \{\exp[1-\epsilon)x]\}^{1/\epsilon}. \tag{37}$$

We discuss the application of (37) in the following section.

DISCUSSION

We can now give a completely rigorous interpretation of Ursell's result, Eq. (11).

Proposition 4: $F(x;N)^{1/N} = \exp[(1-1/N)x] + O(x^2)$, uniformly in N. (38)

Proof: From (37) we find

$$\begin{aligned}
&|F(x;\epsilon)^\epsilon - \exp[(1-\epsilon)x]| \\
&\quad \leq \exp[(1-\epsilon)x]\{1 - \exp[(1-\epsilon)x/\alpha]\}
\end{aligned} \tag{39}$$

$$\leq -[(1-\epsilon)/\alpha]x - [(1-\epsilon)/\alpha]^2x^2 - \cdots, \quad \text{uniformly in } \epsilon. \tag{40}$$

Dividing by x^2, $x \neq 0$, yields

$$\left|\frac{F(x;\epsilon)^\epsilon - \exp[(1-\epsilon)x]}{x^2}\right| \leq \frac{1-\epsilon}{|\alpha x|} + \left(\frac{1-\epsilon}{\alpha}\right)^2 + o(|x|), \quad \text{uniformly in } \epsilon. \tag{41}$$

We note that the left side of (41) is independent of α. This suggests we attempt to make use of α which up until now has been restricted only by $\alpha > 1$. To do this, we refer to (18), which, for $|x| < 1$, will hold for some *fixed* value of $|\alpha x|$. And for such a fixed value of $|\alpha x|$, we see that as $|x| \to 0$, α increases without bound.

Using the above observations, we see that the right-hand side of (41) is bounded as $x \to 0$, and the bound can be chosen independent of ϵ. In the usual convention, this says

$$F(x;\epsilon)^\epsilon = \exp[(1-\epsilon)x] + O(x^2) \quad \text{uniformly in } \epsilon. \tag{42}$$

To obtain (38) of the proposition, we set $\epsilon = 1/N$.

We now observe that (38) implies

$$F(x;N)^{1/N} = 1 + (1-1/N)x + O(x^2) \tag{43}$$

$$= 1 + x - x/N + O(x^2). \tag{44}$$

Because this result is uniformly valid in N, if x has been chosen we may certainly choose N sufficiently large to make x/N negligible, and, with this understanding, write (44) as

$$F(x;N)^{1/N} \approx 1 + x + O(x^2) \tag{45}$$

and, with the same understanding,

$$F(x;N) \approx \{1 + x + O(x^2)\}^N. \tag{46}$$

This is our interpretation of Ursell's result.

The statistical mechanical application requires an estimate for $1/N \ln F$, and can easily be obtained directly from (37). Thus, taking the logarithm of each member of (37) yields

$$\frac{(1-\epsilon)}{\epsilon} x\left(1 + \frac{1}{\alpha}\right) \leq \ln F(x;N) \leq \frac{(1-\epsilon)x}{\epsilon}. \tag{47}$$

Multiplying by $\epsilon = 1/N$ and substracting x yields

$$-\left|\frac{x}{\alpha}\right| + \left|\frac{x}{N}\right| + \left|\frac{x}{N\alpha}\right| \leq \frac{1}{N}\ln F(x;N) - x \leq -\frac{x}{N} = \left|\frac{x}{N}\right| \tag{48}$$

so that

$$\left|\frac{(1/N)\ln F(x;N) - x}{x}\right| < \left|\frac{1}{\alpha}\right| + \left|\frac{1}{N}\right|. \tag{49}$$

Recalling that when x is small α can be large, we see that if x is small and N is large, the fractional error will be small. That is, in the above sense,

$$(1/N)\ln F(x,N) \approx x, \tag{50}$$

and (49) gives us precise knowledge of just how this approximation depends on α and on N, and thereby, via the dependence of α on x, how the approximation depends on x and N, the parameters of physical interest.

We return now to the suggestions from the referee and from Professor Erdélyi. The former suggests making the following eikonal-type substitution into Eq. (8):

$$F = \exp[(1/\epsilon)A_0 + A_1 + \epsilon A_2 + \cdots]. \tag{51}$$

Then

$$4x^2(A_0')^2 - (1+4x)A_0' + 1 = 0. \tag{52}$$

The appropriate root is

$$A_0' = (1 + 4x - \sqrt{1+8x})/8x^2, \tag{53}$$

leading to a solution which for small x is

$$A_0 \simeq x - 2x^2. \tag{54}$$

This leads immediately to Ursell's result.

A further point made by the referee is that the appropriate root of Eq. (52) is close to that obtained by neglecting the $(A_0')^2$ term, whose origin is the second derivative of Eq. (8), and this to some extent at least justifies Ursell's omission of the second derivative.

Professor Erdélyi points out that Ursell's solution can be represented using a Hermite polynomial. Using the notation of Olver,[10] he gives

$$F_N(x) = \left(-\frac{x}{N}\right)^{N/2} H_N\left(\frac{1}{2}\left(-\frac{N}{x}\right)^{1/2}\right)$$
$$= \left(\frac{-N}{4x}\right)^{-N/2} U\left(-\frac{N}{2}, \frac{1}{2}, \frac{-N}{4x}\right). \tag{55}$$

Olver[10] (on p. 403) gives an asymptotic form for Hermite polynomials from which the behavior of F should follow.

This is an interesting observation, for we see that our direct approach to solving Ursell's combinatorial problem has incidentally provided a method for finding an asymptotic representation for a Hermite polynomial of large order, and indeed a method seems to be considerably simpler and more direct than the standard methods.

[1] H.D. Ursell, Proc. Camb. Phil. Soc. **23**, 685 (1927).
[2] J.E. Mayer, J. Chem. Phys. **5**, 67 (1936).
[3] M. Born and K. Fuchs, Proc. Roy. Soc. Lond. A **166**, 391 (1938).
[4] B. Kahn and G.E. Uhlenbeck, Physica **5**, 399 (1938).
[5] W.G. McMillan and J.E. Mayer, J. Chem. Phys. **13**, 276 (1945).
[6] J. Jeans, *Dynamical Theory of Gases* (Cambridge U.P., Cambridge, 1916).
[7] R.H. Fowler, Proc. Camb. Phil. Soc. **22**, 861 (1925).
[8] R.H. Fowler and E.A. Guggenheim, *Statistical Thermodynamics* (Cambridge U.P., Cambridge, 1939).
[9] J.D. Cole, *Perturbation Methods in Applied Mathematics* (Ginn-Blaisdel, Boston, 1968).
[10] F.W.J. Olver, *Asymptotics and Special Functions* (Academic, New York, 1974).
[11] A.D. MacGillivray, J.Chem. Phys. **56**, 80, 83 (1972); **57**, 4071, 4075 (1972).

Discrete space quantum mechanics

David P. Vasholz

Naval Coastal Systems Laboratory, Panama City, Florida 32401
(Received 1 April 1975)

The general problem of finding exactly soluble quantum systems is considered. It is argued that discrete space quantum mechanics emerges in a natural way as an avenue of approach. Discrete space quantum mechanics is formulated and applied to one-dimensional quantum systems with emphasis on single-channel models. It is found that a large variety of systems are exactly soluble in the sense that they only require the inversion of a finite-dimensional matrix. The interaction may in general be both nonlocal and non-time-reversal-invariant. The analytic structure of the resolvent is worked out in detail for a simple class of examples. It is shown that a slightly modified version of the usual continuous space Schrödinger equation may in principle be solved exactly for any finite range local potential by writing the solution in terms of corresponding discrete space solutions. It is also shown that from an algebraic viewpoint the models constructed are realizations of generalized versions of the Weyl relations.

I. INTRODUCTION

One of the manifestations of the complexity of quantum as opposed to classical mechanics is that there are very few quantum systems that can be solved exactly. There is a shortage of even highly idealized and simplified systems which, while nontrivial, are nonetheless tractable enough to serve as useful "theoretical laboratories" in which one may investigate within specific contexts some of the many techniques, approximations, and conjectures which have accrued over the years, and perhaps also in which entirely new ideas may be developed. It is the purpose of this work to propose and illustrate a general method which appears capable of generating a very wide variety of such systems.

All quantum systems are represented by a vector space S and a Hamiltonian H, which acts upon the elements of S. All aspects of the dynamics are known when one has calculated the operator e^{-iHt}. The basic goal of mechanics is to describe the motions of various objects. To do this, the concepts of position and momentum have proved indispensable. It is well known, however, that the canonical commutation relation $[P, Q] = i\hbar$ cannot be realized[1] if S is finite dimensional. Thus it would appear that if one wishes to construct models relevant to the motion of objects, S must be infinite dimensional.[2] In addition, finite-dimensional systems are limited in the sense that there can be only a finite number of energy eigenvalues. This means that any such system must be quasiperiodic, which precludes a direct study of phenomena such as decay and scattering processes. Initially, therefore, we consider systems for which S is infinite dimensional; more specifically, S is taken to be a separable Hilbert space. Let an orthonormal basis be specified by

$$\{|n\rangle;\ n=0,\ \pm1,\ \pm2\cdots\} \tag{1.1}$$

$$\langle m|n\rangle=\delta_{mn}. \tag{1.2}$$

Mathematically, the simplest conceivable nontrivial Hamiltonian is one that connects nearest-neighbor states in a homogeneous fashion. We denote such a Hamiltonian by T'. Since T' must be Hermitian, we have

$$T'|n\rangle=-|n+1\rangle-|n-1\rangle,\quad n=0,\ \pm1,\cdots, \tag{1.3}$$

the minus sign being chosen for later convenience. Note that

$$T|n\rangle\equiv(T'+2I)|n\rangle \tag{1.4}$$
$$=-|n+1\rangle+2|n\rangle-|n-1\rangle,$$

where I is the identity operator, which has no physical effect when added to a Hamiltonian. The bottom line of (1.4) is the negative of the finite difference version of a second derivative. This immediately suggests that our model be interpreted as a discrete-space analog of the usual formulation of single-particle quantum mechanics in one dimension, with the correspondences

$$T\longleftrightarrow p^2/2m \tag{1.5}$$

and

$$|n\rangle\longleftrightarrow|x\rangle \tag{1.6}$$

where

$$x=n\epsilon \tag{1.7}$$

with ϵ some constant equal to the "lattice spacing." In conclusion, DSQM (discrete space quantum mechanics) suggests itself almost as a matter of course if one starts from first principles and seeks out the simplest possible nontrivial quantum systems. The remainder of this paper is largely an application of DSQM to single-channel systems.

Mathematically, the present work is closely related to the codiagonal bordering models considered by Stey and Gibberd,[3] to some work of Case, and Case and Kac on the inverse scattering problem,[4,5] and to the work of the author on the exponential decay problem.[6] Some elementary versions of the models considered here have also appeared in a text by Feynman, Leighton, and Sands.[7]

In Sec. II the relevant operators are defined which act upon our basis of position eigenkets. The general method of solution is introduced in Sec. III and applied in Sec. IV to a wide class of single-channel systems. In Sec. V a very basic symmetry which gives some insight into the spectral structure of DSQM is formulated and discussed. A readily solved class of examples is considered in some detail in Sec. VI. Section VII is devoted to showing how a slightly modified version of the

usual continuous space Schrödinger equation may in principle be solved exactly using solutions to corresponding discrete systems, while in Sec. VIII DSQM is considered from an algebraic point of view. Section IX is given over to general remarks.

II. BASIC OPERATORS

The position operator X is of course defined by

$$X|n\rangle \equiv n|n\rangle. \tag{2.1}$$

Consider now the continuous set defined by

$$\left|p\right\rangle \equiv (2\pi)^{-1/2}\sum_{n=-\infty}^{\infty} e^{-inp}\left|n\right\rangle, \tag{2.2}$$

with the normalization

$$\langle p|p'\rangle = \delta(p-p'). \tag{2.3}$$

Equation (2.2) may be inverted to yield

$$|n\rangle = (2\pi)^{-1/2}\int_{-\pi}^{\pi} dp e^{inp}|p\rangle. \tag{2.4}$$

It follows immediately that

$$U_k|p\rangle = e^{ikp}|p\rangle, \tag{2.5}$$

where k is an arbitrary integer and U_k is the associated space translation operator defined by

$$U_k|n\rangle \equiv |n+k\rangle. \tag{2.6}$$

Hence P is the momentum operator, where P is defined by

$$P|p\rangle = p|p\rangle. \tag{2.7}$$

It is easily shown that

$$T = \frac{\hbar^2}{m\epsilon^2}\left(I - \cos\frac{\epsilon P}{\hbar}\right), \tag{2.8}$$

$$mX = \frac{\hbar}{\epsilon}\sin\frac{\epsilon P}{\hbar}, \tag{2.9}$$

$$\sin\frac{\epsilon P}{\hbar}|n\rangle = \frac{1}{2i}(|n+1\rangle - |n-1\rangle), \tag{2.10}$$

where the quantities $2m$, ϵ, and $\hbar$ had previously been set equal to unity (as they will be for most of what follows). Equations (2.8) through (2.10) indicate that DSQM becomes equivalent to the conventional CSQM (continuous space quantum mechanics) for small ϵ, or equivalently if one is dealing with momenta that are small compared with $\hbar/\epsilon$.

A unitary operator $U(\theta)$ which performs translations in momentum space may be defined by

$$U(\theta)|p\rangle \equiv |p+\theta\rangle. \tag{2.11}$$

It follows immediately from Eq. (2.4) that

$$U(\theta)|n\rangle = e^{-in\theta}|n\rangle. \tag{2.12}$$

Define the antiunitary time reversal operator $\mathcal{T}$ by

$$\mathcal{T}(\alpha|m\rangle + \beta|n\rangle) \equiv \alpha^*|m\rangle + \beta^*|n\rangle. \tag{2.13}$$

The parity operator $\mathcal{P}$ is defined by

$$\mathcal{P}|m\rangle \equiv |-m\rangle. \tag{2.14}$$

It is trivial to verify all the usual relations, such as

$$\mathcal{P}P\mathcal{P}^{-1} = \mathcal{T}P\mathcal{T}^{-1} = -P,$$

$$\mathcal{T}X\mathcal{T}^{-1} = -\mathcal{P}X\mathcal{P}^{-1} = X, \tag{2.15}$$

$$\mathcal{P}^2 = \mathcal{T}^2 = I.$$

III. THE LIPPMANN–SCHWINGER EQUATION: FREE PARTICLE SOLUTION

If one knows the resolvent of a quantum system one knows all aspects of the dynamics. Consider a Hamiltonian consisting of a term H_0, whose resolvent is known, plus some interaction V:

$$H = H_0 + V. \tag{3.1}$$

Then

$$E - H_0 = E - H + V, \tag{3.2}$$

where E is any complex number. Multiply Eq. (3.2) from the left by $(E-H_0)^{-1}$ and from the right by $(E-H)^{-1}$ to obtain

$$G = F + FVG, \tag{3.3}$$

where

$$F \equiv (E-H_0)^{-1} \quad G \equiv (E-H)^{-1}. \tag{3.4}$$

Equation (3.3) is the Lippmann–Schwinger equation. In CSQM Eq. (3.3) is an integral equation, typically insoluble in closed form. In DSQM Eq. (3.3) becomes an equation involving summations over discrete indices, which often may be solved exactly.

An important example, the result of which we shall require later, is furnished by the calculation of the resolvent of the kinetic energy operator T, defined in Eq. (2.8). Setting $H_0 = 0$ and $V = T$, Eq. (3.3) requires

$$\langle m|(E-T)^{-1}|n\rangle = E^{-1}\delta_{mn} + E^{-1}\langle m|T(E-T)^{-1}|n\rangle. \tag{3.5}$$

Let

$$F_{mn}(E) \equiv \langle m|(E-T)^{-1}|n\rangle. \tag{3.6}$$

From translational and parity invariance,

$$F_{mn}(E) = F_{|m-n|}(E). \tag{3.7}$$

Equation (3.5) therefore becomes

$$(E-2)F_{|m-n|} = \delta_{mn} - F_{|m-n+1|} - F_{|m-n-1|}. \tag{3.8}$$

Postulating a solution of the form

$$F_{mn} = af^{|m-n|}, \tag{3.9}$$

one readily obtains

$$F_{mn}(E) = \left(\frac{1}{2}\right)^{|m-n|}\frac{[2-E+\sqrt{E(E-4)}\,]^{|m-n|}}{\sqrt{E(E-4)}}. \tag{3.10}$$

The precise meaning of $\sqrt{E(E-4)}$ must now be specified. For the $F_{mn}(E)$ to be the matrix of a bounded linear operator defined on the entire Hilbert space it is necessary and sufficient that[8]

$$\tfrac{1}{2}|2-E+\sqrt{E(E-4)}\,| < 1. \tag{3.11}$$

Note also that

$$\tfrac{1}{4}[2-E+\sqrt{E(E-4)}][2-E-\sqrt{E(E-4)}] = 1, \tag{3.12}$$

so that for each value of E there is only one way the square root can be evaluated which satisfies (3.11). Let E and $E-4$ be written in the forms

$$E=r_1e^{i\theta_1}, \quad 0\leq\theta_1<2\pi, \quad r_1\geq 0,$$
$$E-4=r_2e^{i\theta_2}, \quad 0\leq\theta_2<2\pi, \quad r_2\geq 0.$$

Then it may be verified that (3.11) is satisfied if and only if the square root in Eq. (3.10) is defined by

$$\sqrt{E(E-4)}\equiv\sqrt{r_1r_2}\,e^{i(\theta_1+\theta_2)/2}. \tag{3.13}$$

Thus in the complex E plane $F_{mn}(E)$ is an analytic function on a two-sheeted Riemann surface having a square root branch cut running along the real axis from $E=0$ to $E=4$ (measured in units of $\hbar^2/2m\epsilon^2$). Relation (3.11) is satisfied on the first (or physical) sheet, and violated on the second sheet. Denoting by the superscripts I and II this function as evaluated on the first and second sheets, we have

$$F^{\mathrm{I}}_{mn}(E)=\left(\frac{1}{2}\right)^{|m-n|}\frac{[2-E+\sqrt{E(E-4)}]^{|m-n|}}{\sqrt{E(E-4)}} \tag{3.14a}$$

$$F^{\mathrm{II}}_{mn}(E)=-\left(\frac{1}{2}\right)^{|m-n|}\frac{[2-E-\sqrt{E(E-4)}]^{|m-n|}}{\sqrt{E(E-4)}}. \tag{3.14b}$$

It is often convenient to work instead in the complex z-plane, where

$$E=2-z-z^{-1}. \tag{3.15}$$

Let

$$\mathcal{J}_{mn}(z)\equiv z^{|m-n|}/(z-z^{-1}). \tag{3.16}$$

Then it may readily be verified that

$$\mathcal{J}_{mn}(z)=F^{\mathrm{I}}_{mn}(E), \quad |z|\leq 1, \tag{3.17a}$$

$$\mathcal{J}_{mn}(z)=F^{\mathrm{II}}_{mn}(E), \quad |z|\geq 1. \tag{3.17b}$$

Thus the entire first and second sheets of the complex E plane are mapped respectively into the interior and exterior of the unit circle in the complex z-plane. The branch cut itself is mapped onto the unit circle.

IV. SINGLE CHANNEL SYSTEMS

Consider a single channel system. Then

$$H=T+V \tag{4.1}$$

where

$$T=-\frac{\hbar^2}{2m\epsilon^2}\sum_{n=-\infty}^{\infty}[\,|n\rangle\langle n+1|-2|n\rangle\langle n|+|n+1\rangle\langle n|\,], \tag{4.2}$$

$$V=\sum_{k=-N}^{N}\sum_{m=-N}^{N}|k\rangle A_{km}\langle m|. \tag{4.3}$$

The matrix A must be Hermitian for H to be Hermitian. In general, the interaction may be both nonlocal and non-time-reversal invariant. The system is time-reversal invariant if and only if A is symmetric. If A is also diagonal, one has a local potential. It is assumed that the interaction is of finite range, in the sense that N is a finite integer.

The Lippman–Schwinger equation becomes

$$G_{mn}=F_{mn}+\sum_{k=-N}^{N}\sum_{r=-N}^{N}F_{mk}A_{kr}G_{rn} \tag{4.4}$$

where F_{mn} is defined in Eq. (3.14) and where

$$G_{mn}(E)\equiv\langle m|(E-H)^{-1}|n\rangle. \tag{4.5}$$

Define the $(2N+1)\times\infty$ matrix B and the $(2N+1)\times(2N+1)$ matrix C by truncating sectors of the $\infty\times\infty$ matrix F according to

$$B_{mn}=F_{mn}, \quad -N\leq m\leq N, \quad -\infty<n<\infty,$$
$$C_{mn}=F_{mn}, \quad -N\leq n\leq N, \quad -N\leq m\leq N. \tag{4.6}$$

It may readily be shown that, suppressing matrix indices and denoting the transpose of B by $\tilde{B}$,

$$G=F+\tilde{B}(I-AC)^{-1}AB. \tag{4.7}$$

Thus the system may be completely solved from a knowledge of F and the inverse of the finite-dimensional matrix $I-AC$.

In order to obtain the G^I_{mn}, that is the matrix elements of $(E-H)^{-1}$, one must of course use F^{I} on the right-hand sides of Eqs. (4.6) and (4.7); likewise the second sheet continuations, the G^{II}_{mn}, are obtained by using F^{II}.

Expressing the above results in terms of the complex variable z defined in Sec. III, one obtains

$$\mathcal{G}(z)=\mathcal{J}(z)+\tilde{\mathcal{B}}(z)[I-A\mathcal{C}(z)]^{-1}A\mathcal{B}(z), \tag{4.8}$$

where $\mathcal{J}(z)$ is as defined in Eq. (3.16), while $\mathcal{B}$ and $\mathcal{C}$ are appropriate truncations of $\mathcal{J}$ completely analogous to B and C. In analogy with Eq. (3.17) we have

$$\mathcal{G}_{mn}(z)=G^{\mathrm{I}}_{mn}(E), \quad |z|\leq 1, \tag{4.9a}$$

$$\mathcal{G}_{mn}(z)=G^{\mathrm{II}}_{mn}(E), \quad |z|\geq 1. \tag{4.9b}$$

From Eqs. (3.16) and (4.8) it is clear that G may be written in the form

$$\mathcal{G}_{mn}(z)=N_{mn}(z)/D(z), \tag{4.10}$$

where N and D are finite order polynomials in z. Hence in DSQM there exist a great variety of systems whose resolvents have very simple analytic structures.

Once the resolvent is known, the S operator may readily be calculated via the relation[9]

$$S=I-i\int_{-\infty}^{\infty}dt\,e^{iTt}Ve^{-iTt}-\int_{-\infty}^{\infty}dt\int_{-\infty}^{t}dt'\,e^{iTt}Ve^{-iH(t-t')}Ve^{-iTt'}. \tag{4.11}$$

After some manipulation, we obtain for the S-matrix elements

$$\langle p|S|p'\rangle=\delta(p-p')-i\delta(E-E')\sum_{m,n}e^{imp}A_{mn}e^{-inp'} - i\delta(E-E')\sum_{m,n,k,r}e^{imp}A_{mn}G_{nk}(E+i\eta)A_{kr}e^{-irp'}, \tag{4.12}$$

where

$$\eta\to 0^+, \quad E=2(1-\cos p), \quad E'=2(1-\cos p').$$

V. REFLECTION SYMMETRY

Mathematically, DSQM is a formulation of quantum mechanics in terms of the Hilbert space l^2. In this section we formulate and apply a general symmetry, which we call reflection symmetry, that is exhibited by resolvents of operators in l^2.

Let M be an arbitrary linear operator in l^2. Then M may be expressed as a function of two complex numbers α and β as follows:

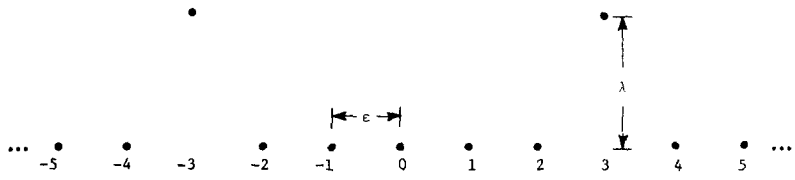


FIG. 1. The potential (6.1) for $N=3$.

$$M(\alpha,\beta)=\alpha\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}|m\rangle a_{mn}[1-(-1)^{m+n}]\langle n| + \beta\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}|m\rangle b_{mn}[1+(-1)^{m+n}]\langle n|. \tag{5.1}$$

Consider now the unitary operator $U(\theta)$ defined in Equation (2.12) with $\theta=\pi$. One immediately obtains

$$U^{-1}(\pi)M(\alpha,\beta)U(\pi)=-M(\alpha,-\beta). \tag{5.2}$$

Define the resolvent

$$R_{mn}(\alpha,\beta;E)\equiv\langle m|[E-M(\alpha,\beta)]^{-1}|n\rangle. \tag{5.3}$$

Then Eq. (5.2) implies

$$R_{mn}(\alpha,-\beta;E)=(-1)^{m+n+1}R_{mn}(\alpha,\beta;-E). \tag{5.4}$$

As an application of this symmetry, let the Hamiltonian of a system be given by

$$H=T+\lambda V, \tag{5.5}$$

where T is given in Equation (4.2) and where V is a local potential. Specializing to this particular Hamiltonian, Eq. (5.4) implies for the resolvent of H the condition

$$G_{mn}(-\lambda;E)=(-1)^{m+n+1}G_{mn}(\lambda;4-E). \tag{5.6}$$

Therefore, in DSQM, the resolvent is essentially unchanged when the sign of a local potential is reversed. All that is involved is a reflection through the point $E=\hbar^2/m\epsilon^2$ in the complex E plane. Thus if an attractive potential has a bound state pole at $E<0$, the corresponding repulsive potential has a bound state at $E>2\hbar^2/m\epsilon^2$, beyond the upper limit of the continuum. This is in marked contrast to the situation in CSQM, where only attractive local potentials can have bound states. As $\epsilon\to 0$, these repulsive potential bound states are "chased" off to infinity by the lengthening branch cut. For small momenta, which corresponds to E near the origin, repulsive potential bound states should have little effect, since they are "faraway" singularities. The result (5.6) is of course independent of any restriction to finite N as in Eq. (4.3).

VI. A SIMPLE EXAMPLE

We restrict ourselves to local potentials, that is to situations where A is diagonal. The simplest possible class of parity-conserving potentials occurs when there are only two nonvanishing elements of A, given by

$$A_{NN}=A_{-N-N}=\lambda. \tag{6.1}$$

Such potentials correspond to two lattice points being "out of place," as shown in Fig. 1.

The $N=0$ case has been considered previously[7] in a pedagogical context. It is of interest because it specifies the simplest possible interacting unbounded quantum system. The $N=1$ case has also been treated[6]; it specifies the simplest posssible quantum system to exhibit exponential decay.

For the particularly simple potential of Eq. (6.1), it is trivial to solve Eq. (4.4) directly. Taking advantage of parity invariance, one obtains

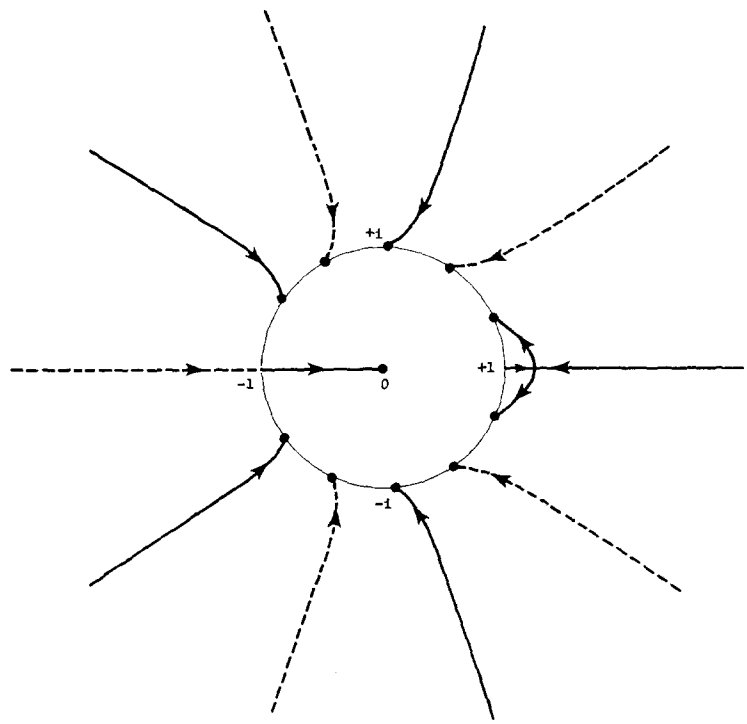


FIG. 2. Trajectories of the poles of $\mathcal{G}$ for $N=3$ with respect to the unit circle in the complex z plane as λ varies from 0 to ∞. The arrows denote directions of increasing λ, while solid and dashed lines correspond respectively to trajectories of positive and negative parity poles. All complex poles lie outside the unit circle, as they must since they correspond to second sheet poles in the complex E plane. The dots represent limit points as λ becomes infinite.

$$(z-z^{-1})\mathcal{G}^{(\pm)}_{mn}(z)=z^{|m-n|}\pm z^{|m+n|}$$
$$-\frac{\lambda z(z^{|m-N|}\pm z^{|m+N|})(z^{|N-n|}\pm z^{|N+n|})}{\pm\lambda z^{2N+1}-z^2+\lambda z+1}, \tag{6.2}$$

where

$$\mathcal{G}^{(\pm)}_{mn}(z)=\mathcal{G}_{mn}(z)\pm\mathcal{G}_{m-n}(z). \tag{6.3}$$

From Sec. V it follows that only the case $\lambda>0$ need be considered.

The poles of the resolvent thus correspond to the roots of $D^{\pm}(z)$, where

$$D^{\pm}(z)=\pm\lambda z^{2N+1}-z^2+\lambda z+1. \tag{6.4}$$

Except for the two roots at $z=\pm1$ of D^-, any given root z_0 corresponds to a pole on the first or second sheet in the complex E plane, according to whether z_0 lies inside or outside the unit circle. For $\lambda>0$ it is not difficult to establish that D^+ has one and only one real root in the interval $(-\infty,1]$. This root z_0 must satisfy $-1<z_0<0$; it has multiplicity 1. If λ is sufficiently small, there are two more real roots, each of multiplicity 1 and each greater than 1. As λ increases they eventually coalesce into a double root and go off the real axis, becoming a conjugate pair. The remaining $2N-2$ roots are all complex and occur in conjugate pairs. The negative parity polynomial D^- may be written as

$$D^-(z)=(1-z^2)[\lambda(z+z^3+z^5+\cdots+z^{2N-1})+1]. \tag{6.5}$$

It follows from Eq. (6.2) that the roots of D^- at $z=\pm1$ are trivial in the sense that they do not produce poles in the resolvent. From Eq. (6.5) it is apparent that D^- has one and only one nontrivial real root which is negative for $\lambda>0$ and which lies inside or outside the unit circle according to whether λ is larger or smaller than $1/N$. The remaining $2N-2$ nontrivial roots are all complex and occur in conjugate pairs. The situation for $N=3$ is shown graphically in Fig. 2.

For large values of λ, the roots may be expanded in powers of $1/\lambda$. To first order the positive parity roots may be expressed as

$$z_0^+=-1/\lambda$$
$$z_k^+=\exp[\pi i(k+1/2)/N]\left(1-\frac{i}{\lambda N}\sin\frac{\pi(k+1/2)}{N}+\cdots\right),\quad k=1,2,\ldots,2N$$

while the negative parity poles are given by

$$z_0^-=-1/\lambda$$
$$z_k^-=e^{\frac{\pi ik}{N}}\left(1-\frac{i}{\lambda N}\sin\frac{\pi k}{N}+\cdots\right),\quad k=1,2,\ldots,2N. \tag{6.7}$$

Let $E_k^{\pm}$, $k=1,2,\ldots,2N$ denote the positive and negative parity complex poles in the complex E plane in the limit $\lambda\to\infty$. Then from (6.6) and (6.7)

$$E_k^+=\frac{\hbar^2}{m\epsilon^2}\left(1-\cos\frac{\pi(k+1/2)}{N}\right), \tag{6.8}$$

$$E_k^-=\frac{\hbar^2}{m\epsilon^2}\left(1-\cos\frac{\pi k}{N}\right). \tag{6.9}$$

The $E_k^{\pm}$ are easily shown to be the energy eigenvalues of a particle confined to the $2N-1$ lattice points lying between $x=-(N-1)\epsilon$ and $X=(N-1)\epsilon$. Thus the E_k, which populate the second sheet for finite λ, simply represent virtual states corresponding to a particle in an infinite well. Writing $L=(2N-1)\epsilon$ and taking the limits of $E_k^{\pm}$ as $\epsilon\to0$ with L constant, one immediately obtains

$$\lim_{\epsilon\to0}E_k^+=2\pi^2\hbar^2(k+1/2)^2/mL^2 \tag{6.10}$$

$$\lim_{\epsilon\to0}E_k^-=2\pi^2\hbar^2k^2/mL^2, \tag{6.11}$$

which are the positive and negative parity energy levels obtained in CSQM for an infinite well of width L.

VII. CONTINUOUS SPACE QUANTUM MECHANICS AS A COMPOSITION OF DISCRETE SOLUTIONS

As we have formulated it, CSQM differs in two essential respects from DSQM: The eigenvalues of the position operator are discrete and the kinetic energy operator T is given by

$$T=I-\cos\epsilon P. \tag{7.1}$$

In this section we remove the first of these differences, letting space be continuous but retaining (7.1) as the kinetic energy. It then follows that

$$T|x\rangle=-\frac{1}{2\epsilon^2}|x-\epsilon\rangle+\frac{1}{\epsilon^2}|x\rangle-\frac{1}{2\epsilon^2}|x+\epsilon\rangle \tag{7.2}$$

where

$$\langle x|x'\rangle=\delta(x-x'). \tag{7.3}$$

Let $H=T+V$ with the operator V signifying a local potential so that

$$V|x\rangle=v(x)|x\rangle. \tag{7.4}$$

It follows from Equations (7.2) and (7.4) that the original space may be decomposed into an uncountable sum of subspaces, each of which is invariant with respect to H. This decomposition is completely determined by specifying that two basis kets $|x\rangle$ and $|x'\rangle$ are members of the same subspace if and only if x and x' differ by an integral number of ϵ's. A given subspace may be denoted by S_r where

$$0\leq r<\epsilon \tag{7.5}$$

and where $|x\rangle\in S_r$ if and only if

$$x=r+n\epsilon \tag{7.6}$$

with n an integer. Let H as it operates within the subspace S_r be denoted by H_r. Then, if $v(x)$ is of finite range, the results of Sec. IV may be used to obtain a solution within each S_r and the results combined to produce the solution for the entire problem. The result is

$$\left\langle x'\left|\frac{1}{E-H}\right|x\right\rangle=\sum_{m=-\infty}^{\infty}\delta(x-x'+m\epsilon-n\epsilon)\left\langle m\left|\frac{1}{E-H_r}\right|n\right\rangle, \tag{7.7}$$

where r and n are as specified in Eq. (7.6). Therefore, given a kinetic energy of the form (7.2), one can in principle solve the one-dimensional Schrödinger equation for any finite range local potential. These results can also trivially be extended to three dimensions, although a kinetic energy which is a sum of three terms similar to (7.2) suffers from the unappealing feature of singling out a preferred set of axes in space (such effects should, however, disappear as $\epsilon\to0$).

All that remains to recover the usual formulation of CSQM is to take the limit $\epsilon \to 0$. The effect of taking this limit is clearly to increase the complexity of the analytic structure of the resolvents within their invariant subspaces. Beyond a certain point, the increased complexity occurs at such high energies as to be irrelevant to nonrelativistic quantum mechanics. In this regard it is amusing to note that by setting

$$\epsilon = \sqrt{3}\hbar/mc \tag{7.8}$$

one has

$$T = \frac{p^2}{2m} - \frac{p^4}{8m^3c^2} + \frac{p^6}{80m^5c^4} - \cdots \tag{7.9}$$

which is identical with $\sqrt{p^2c^2 + m^2c^4} - mc^2$ through fourth order in p. Thus one could argue that at low enough energies and for this value of ϵ, a kinetic energy given by Eq. (7.2) is actually an improvement over $T = p^2/2m$.

VIII. ALGEBRAIC ASPECTS

In the introduction DSQM was motivated essentially on the basis of mathematical simplicity. In this section we outline a completely different approach which encompasses DSQM and which gives some insight into its relationships with other structures.

All systems in nonrelativistic quantum mechanics consist of realizations of the Weyl relations[10] which for a single pair of conjugate variables are given by

$$e^{itP}e^{isQ} = e^{ist}e^{isQ}e^{itP} \tag{8.1}$$

where s and t may be any real numbers. Let such systems be designated as Class I. Suppose that the Weyl relations are weakened—that they are not required to hold as s and t are varied continuously. The natural way to do this which preserves e^{itP} and e^{isQ} as one-parameter groups is to require that Eq. (8.1) need be valid only if

Class II: t is an integral multiple of some constant; s may be any real number.

Class III: s and t are both integral multiples of some constant. We have of course

$$\text{Class I} \subset \text{Class II} \subset \text{Class III}. \tag{8.2}$$

It may be readily verified that the discrete space systems which we have studied so far are examples of Class II systems[11] which are not Class I.

It is not difficult to find systems which are uniquely Class III. A basis may be obtained from Eq. (1.1) simply by selecting a fixed integer M and requiring

$$|n+M\rangle = |n\rangle \tag{8.3}$$

for all n. The orthonormality condition is still $\langle m|n\rangle = \delta_{mn}$ except the Kronecker delta is now defined by

$$\delta_{mn} = 1 \quad \text{if } m = n \pmod M$$
$$= 0 \text{ otherwise.} \tag{8.4}$$

In other words, the underlying Hilbert space is made finite-dimensional. Let a complementary othonormal basis set $|\bar{k}\rangle$ be defined by

$$|\bar{k}\rangle \equiv \frac{1}{\sqrt{M}} \sum_{n=1}^{M} e^{\frac{2\pi i k n}{M}} |n\rangle, \quad k = 0, \pm 1, \pm 2, \cdots. \tag{8.5}$$

Note that for this set one also has $|\overline{k+M}\rangle = |\bar{k}\rangle$. Define the operators P and Q by

$$Q|n\rangle \equiv n\sqrt{2\pi/M}\,|n\rangle, \quad P|\bar{k}\rangle \equiv k\sqrt{2\pi/M}\,|\bar{k}\rangle. \tag{8.6}$$

It may then be verified that Q and P satisfy Eq. (8.1) whenever s and t are any integer multiples of $\sqrt{2\pi/M}$. Thus although canonical commutation relations cannot be realized on a finite-dimensional vector space, a rather natural modification of the Weyl relations can be. These particular Class III systems have also been proposed by Schwinger[12] from the point of view of unitary operator bases.

In spite of the limitations mentioned in the Introduction, there are evidently a number of model-building possibilities utilizing these finite dimensional quantum systems. If the kinetic energy is as defined in Eq. (1.4), one has the picture of a particle hopping around a ring consisting of M different sites. Equations (2.8) through (2.10) are still valid. It is clearly possible to take tensor products, hypothesize various interactions, and form many-body systems in the usual way. Assuming spinless fermions, for example, the solution to an N particle system would consist of diagonalizing an $\binom{M}{N} \times \binom{M}{N}$ Hamiltonian matrix. It is also possible to second quantize and allow terms in the Hamiltonian which change the number of bare particles. This would permit one to study very simple models of quantum field theories in finite dimensional vector spaces, where there is no possibility of divergence difficulties.

Thus we see that one may use the Weyl relations to embed both DSQM and CSQM into a scheme which yields in addition a third class of possibilities. A theorem of von Neumann[13] assures us that there is essentially only one representation of the Weyl relations for Class I systems. It is not known by the author whether similar results hold for Class II and Class III systems.

IX. CONCLUSIONS

In this work we have attempted to indicate some of the possibilities inherent in discrete space quantum mechanics. Our main point is that discrete space quantum mechanics produces a great simplification in mathematical structure while at the same time retaining much of the qualitative content of quantum mechanics as it is usually formulated. This raises the possibility that interesting models of increasingly sophisticated structure may be constructed and carried much further toward complete solutions than their continuous space counterparts.[14]

ACKNOWLEDGMENTS

The author is happy to acknowledge useful and stimulating conversations with Professor Michael Parkinson and Professor James Dufty of the University of Florida Physics Department.

[1]To see this, simply take the trace.
[2]See, however, Sec. VIII.
[3]G. C. Stey and R. W. Gibberd, Physica **60**, 1 (1972).
[4]K. M. Case, J. Math. Phys. **15**, 143 (1974).
[5]K. M. Case and M. Kac, J. Math. Phys. **14**, 594 (1973).
[6]D. P. Vasholz, Physica **74**, 577 (1974).
[7]R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison-Wesley, Reading, 1965), Vol. III, p. 13-13.
[8]See, for example, N. I. Akhiezer and I. M. Glazman, *Theory of Linear Operators in Hilbert Space* (Ungar, New York, 1961), p. 53.
[9]See, for example, R. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966), p. 164.
[10]See, for example, M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1972), Vol. I. p. 275.
[11]There is of course a fourth class where s and t of Class II are interchanged. This simply corresponds to the familiar case of periodic boundary conditions.
[12]J. Schwinger, Proc. Natl. Acad. Sci. **46**, 570 (1960).
[13]See, for example, Ref. 10.
[14]This has been accomplished for some types of one-dimensional multichannel systems. D. P. Vasholz (to be published).

Ray-optical analysis of fields on shadow boundaries of two parallel plates

S. W. Lee*

Department of Electrical Engineering, University of Illinois, Urbana, Illinois 61801

J. Boersma

Department of Mathematics, Technological University of Eindhoven, Eindhoven, Netherlands
(Received 17 October 1974)

The electromagnetic diffraction by two parallel plates of semi-infinite length is treated by ray methods. Two special problems are considered: (i) calculation of the fields in the forward and backward directions due to diffraction of a normally incident plane wave by two nonstaggered parallel plates; (ii) calculation of the field due to a line source in the presence of two staggered parallel plates when the source, the two edges, and the observation point are on a straight line. The crucial step in the ray-optical analysis is the calculation of the interaction between the plates. This calculation is performed by two methods, namely, the uniform asymptotic theory of edge diffraction and the method of modified diffraction coefficient. The relative merits of the two methods are discussed. The ray-optical solution of problem (i) agrees with the asymptotic expansion (plate separation large compared to wavelength) of the exact solution.

PACS numbers: 42.10.H

I. INTRODUCTION

This paper is concerned with the solution by ray methods, of some electromagnetic diffraction problems for a set of two perfectly conducting, parallel plates of semi-infinite length. More specifically, the paper consists of three parts dealing with:

(i) The calculation of the electromagnetic fields in the forward and backward directions in the case of diffraction of a normally incident plane wave by two nonstaggered parallel plates (Sec. II). This calculation is based on the uniform asymptotic theory of edge diffraction,[1–3] and its extension as utilized in Refs. 4, 5.

(ii) The study of the same problem as in (i) by the method of modified diffraction coefficient[6,7] (Sec. III).

(iii) The calculation of the electromagnetic field due to a line source in the presence of two staggered parallel plates when the source, the two edges and the observation point are on a straight line (Sec. IV). The limiting case of plane wave excitation in a direction parallel to the line through the edges is discussed as well. The calculation is based on a combination of the uniform asymptotic theory and the method of modified diffraction coefficient.

The motivations and conclusions of our investigation are stated below.

First, the physical problems themselves are of interest as they relate to the wave propagation over sharp ridges; see the introduction of Ref. 8 and the literature quoted there.

Our second, and main, motivation is to show that ray methods provide an effective tool for the (high-frequency) asymptotic analysis of diffraction problems involving parallel-plate configurations. The analysis for such configurations is by no means trivial. In order to explain the difficulties encountered, we present a brief outline of the ray-optical approach to the diffraction problems stated above. In both problems, the incident wave when hitting the first plate, generates a primary diffracted field. The latter field is a cylindrical wave centred at the diffracting edge and as such is determined by Keller's geometrical theory of diffraction.[9,10] The primary diffracted field in turn acts as an incident wave on the second plate and gives rise to secondary diffraction. The secondary diffracted field will interact again with the first plate thus leading to higher-order diffractions. The actual calculation of the secondary diffracted field is complicated by the fact that the second edge lies on the geometrical-optics shadow boundary of the incident wave, due to the first plate. In the case of diffraction by two nonstaggered plates, an additional and similar difficulty comes up at the calculation of the higher-order interaction fields. In the case of multiple diffraction the backscattered direction coincides with the shadow boundary of the specularly reflected wave or, in other words, each edge lies on the ray-optical reflection boundary of the opposite plate. Now, as is well known, Keller's theory is not valid along shadow boundaries.

In order to overcome this difficulty, three different methods have been proposed in recent years, namely, the method of Yee, Felsen, and Keller (YFK),[11] the method of modified diffraction coefficient (MDC),[6,7] and the uniform asymptotic theory of edge diffraction (UAT).[1–3] In the approach by YFK each interaction field is approximated by the field of an equivalent set of isotropic line sources, the source strengths being such as to provide the correct interaction field in the direction toward the opposite edge. Then the interaction fields are determined recursively by means of a special asymptotic formula for scattering of an isotropic cylindrical wave by a half-plane. Originally, YFK was devised in connection with a ray-optical treatment of reflection in an open-ended parallel-plate waveguide. In view of the approximate character of YFK, it is not surprising that the final ray-optical solution of the reflection problem fails to agree with the asymptotic expansion (width of waveguide large com-

pared to wavelength) of the exact solution. A corrected ray-optical solution, based on UAT and in complete agreement with the asymptotic form of the exact solution, was recently derived in Refs. 4, 5.

In the present paper, the successive diffracted fields are calculated by means of MDC and UAT. The first method, MDC, employs a modified diffraction coefficient for diffraction by a half-plane in the presence of a second parallel half-plane. This modified coefficient, which automatically includes the interaction between the diffracting edge and the second half-plane, is derived from the solution of a canonical problem. The second method, UAT, is applicable to diffraction of an arbitrary incident wave by a plane screen. UAT provides an asymptotic solution of the diffraction problem that is uniformly valid near the edge and the shadow boundaries. Away from these regions the solution reduces to an expansion for the diffracted field which contains Keller's result as its leading term. Higher-order terms are obtained as well whereas Keller's theory is incapable of determining these terms.

In the ray-optical analysis of the parallel-plate diffraction problems, both MDC and UAT turn out to be effective methods, although not to the same extent (see the discussion below). For the case of nonstaggered parallel plates, an exact solution to the diffraction problem is obtainable by the Wiener–Hopf technique[12,13]; see Appendix A for a brief discussion of this exact solution. Our ray-optical solution given in (II. 68), (II. 70) and based on UAT, agrees exactly with the asymptotic expansion (plate separation large compared to wavelength) of the exact solution. A second ray-optical solution, given in (III. 11), (III. 12) and based on MDC, precisely recovers the exact far field solution. For the case of staggered parallel plates, a partial solution ignoring interaction between the plates was recently derived by Jones.[8] Excluding interaction terms, our ray-optical solution (IV. 27), (IV. 30) is found to agree with Jones' rigorous asymptotic result.

The ray-optical analysis of this paper also provides a clear insight into the relative merits of MDC and UAT. Our conclusions are: (i) As Keller's theory, UAT describes a general method which in principle can be applies to all edge diffraction problems. On the other hand, MDC is designed to attack diffraction by special configurations involving two parallel plates, and those only. For example, in the diffraction problem for two staggered parallel plates (Sec. IV), the ray-optical solution can be obtained by UAT alone, but not by MDC alone. (ii) When both methods apply, MDC appears simpler than UAT, as demonstrated by the example in Secs. II and III.

Finally we list some conventions to be used throughout this paper: (i) The time factor is $\exp(-i\omega t)$ and is suppressed. (ii) All problems are two-dimensional (no z variation). Both the TM case (nonzero field components H_z, E_x, E_y) and the TE case (nonzero field components E_z, H_x, H_y) are treated simultaneously, with the help of two symbols u and τ such that

for TM $u = H_z$, $\tau = +1$,

for TE $u = E_z$, $\tau = -1$.

It is convenient to associate τ with the reflection coefficient of the field u from a perfectly conducting plane. (iii) The total field u^t is the sum of the incident field u^i and the scattered field u. Additional subscripts in u^t and u (e.g., u^t_{cy}, u_3, etc.) are employed to identify the sequence of fields arising in the multiple interaction between the parallel plates.

II. NONSTAGGERED PARALLEL PLATES: SOLUTION BY UNIFORM ASYMPTOTIC THEORY

A. Statement of problem and approach

The configuration of a pair of nonstaggered parallel plates and our choice of coordinates are sketched in Fig. 1. The polar coordinates $\{r_m, \phi_m\}$; $m = 0, \pm 1, \pm 2, \cdots$ have origins at $\{x = 0,\ y = ma\}$. The angle ϕ_m is measured in a counterclockwise sense when m is positive, and clockwise when $m = 0$ or m is negative; futhermore, $0 \leq \phi_m \leq 2\pi$. Let the incident plane wave propagate in the negative y direction and be given by

$$u^i(x, y) = \exp(-iky). \tag{II. 1}$$

The problem at hand is to derive a high-frequency approximation for the far field in the forward direction ($x = 0$, $ky \to -\infty$) and the backward direction ($x = 0$, $ky \to \infty$) of the incident plane wave.

Our approach is outlined below. The incident field (II. 1) first reaches the upper plate $x \leq 0$, $y = a$, and scattering produces a total field $u^t_1(r_1, \phi_1)$ that is written as

$$u^t_1(r_1, \phi_1) = \exp(-iky) + u_1(r_1, \phi_1), \tag{II. 2}$$

where u_1 denotes the scattered field. The field u^t_1 in turn acts as an incident field on the lower plate $x \leq 0$, $y = 0$. Scattering of u^t_1 at the lower plate gives rise to a scattered field $u_2(r_0, \phi_0)$, which will interact again with the upper plate and yield a scattered field $u_3(r_1, \phi_1)$. In this manner there results a sequence of scattered fields

$$u_1(r_1, \phi_1), u_2(r_0, \phi_0), u_3(r_1, \phi_1), u_4(r_0, \phi_0), \cdots. \tag{II. 3}$$

Note that $u_n(r_1, \phi_1)$ with n odd arises from a scattering at the upper plate; whereas $u_n(r_0, \phi_0)$ with n even arises

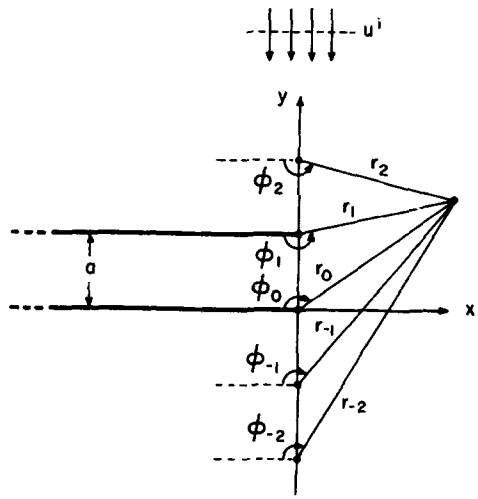


FIG. 1. Two nonstaggered parallel plates illuminated by a normally incident plane wave.

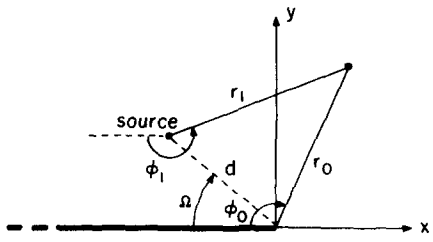


FIG. 2. A half-plane illuminated by a line source at $(x=-d\cos\Omega,\ y=d\cos\Omega)$.

from a scattering at the lower plate. A useful property of the scattered fields is

$$u_n(r_1,\phi_1) = -\tau u_n(r_1, 2\pi-\phi_1), \quad n \text{ odd}, \tag{II. 4a}$$

$$u_n(r_0,\phi_0) = -\tau u_n(r_0, 2\pi-\phi_0), \quad n \text{ even}. \tag{II. 4b}$$

This symmetry relation is a consequence of the fact that u_n is the scattered field from a single plate, as if the other plate were absent. For the sequence of scattered fields in (II. 3) we will determine them recursively instead of consecutively. A special form of u_n is assumed, and it is used to derive u_{n+1} by the uniform asymptotic theory, which is summarized in Sec. II B. Comparing the expression of u_{n+1} thus obtained with the assumed form of u_n after replacing n by $n+1$ in the latter, we obtain two recurrence relations in Sec. II C. Next we solve the recurrence relations in Sec. II D, and present the final results for the scattered fields on the shadow boundaries of the incident and reflected fields in Sec. II E.

B. Summary of uniform asymptotic theory

The uniform asymptotic theory of edge diffraction was developed in Refs. 1 and 3 for the scalar wave, and in Ref. 2 for the vectorial wave. Here we summarize its explicit formulas for a two-dimensional problem, and they constitute a theoretical basis for our analysis in Secs. II and IV of this paper.

Referring to Fig. 2, let the half-plane $x \leq 0,\ y=0$ be illuminated by a cylindrical wave due to a line source located at $x=-d\cos\Omega,\ y=d\sin\Omega,\ 0<\Omega<\pi$. Polar coordinates $\{r_1,\phi_1\}$ with origin at the source point, and $\{r_0,\phi_0\}$ with origin at the edge $\{x=0, y=0\}$ will be employed. We assume the incident cylindrical wave is given by the asymptotic representation:

$$u^i(r_1,\phi_1) \sim \exp(ikr_1) z^i(r_1,\phi_1), \quad k\to\infty, \tag{II. 5a}$$

$$= \exp(ikr_1)\sum_{m=0}^{\infty}(ik)^{-m} z_m^i(r_1,\phi_1). \tag{II. 5b}$$

Then the total field u^t is found to be

$$u^t(r_0,\phi_0) = U(r_0,\phi_0) + \tau U(r_0, 4\pi-\phi_0), \tag{II. 6a}$$

where the double-valued function U is represented by a uniform asymptotic expansion:

$$U(r_0,\phi_0) \sim U^G(r_0,\phi_0) + U^d(r_0,\phi_0), \quad k\to\infty, \tag{II. 6b}$$

where

$$U^G(r_0,\phi_0) \sim \exp[ik(r_0+d)][F(k^{1/2}\xi_0^*) - \hat{F}(k^{1/2}\xi_0^*)] z^i(r_1,\phi_1),$$

$$U^d(r_0,\phi_0) \sim \exp[ik(r_0+d)] k^{-1/2}\sum_{m=0}^{\infty}(ik)^{-m}\hat{v}_m(r_0,\phi_0).$$

The various notations which appeared in (II. 6) are explained below. The Fresnel integral $F(x)$ is defined by

$$F(x) = \pi^{-1/2}\exp(-i\pi/4)\exp(-ix^2)\int_{-\infty}^{x}\exp(it^2)\,dt. \tag{II. 7}$$

Its asymptotic expansion for large x is

$$F(x) \sim \exp(-ix^2)H(x) + \hat{F}(x), \quad x\to\pm\infty. \tag{II. 8}$$

Here $H(x)$ is the unit step function, i. e., $H(x)=1$ for $x>0$ and $H(x)=0$ for $x<0$, and

$$\hat{F}(x) = -\frac{\exp(i\pi/4)}{2\pi x}\sum_{m=0}^{\infty}\Gamma(m+\tfrac{1}{2})(ix^2)^{-m}, \tag{II. 9}$$

where the Gamma function $\Gamma(m+\frac{1}{2})$ is given by

$$\Gamma(m+\tfrac{1}{2}) = \sqrt{\pi}(\tfrac{1}{2})(\tfrac{3}{2})\cdots((2m-1)/2). \tag{II. 10}$$

The Taylor expansion of $F(x)$ around $x=0$ is

$$F(x) = \tfrac{1}{2}\sum_{q=0}^{\infty}\frac{\exp(-iq\pi/4)}{\Gamma(q/2+1)}x^q, \tag{II. 11}$$

which is convergent for each x. The function ξ_0^* in (II. 6) is defined by

$$\xi_0^* = (r_0+d-r_1)^{1/2}\,\mathrm{sgn}[\cos\tfrac{1}{2}(\phi_0-\Omega)]$$

$$= [4dr_0/(r_0+d+r_1)]^{1/2}\cos\tfrac{1}{2}(\phi_0-\Omega). \tag{II. 12}$$

Note that $\xi_0^*=0$ along the shadow boundary $\phi_0=\Omega+\pi$ of the incident wave. The sign of ξ_0^* is such that $\xi_0^*>0$ ($\xi_0^*<0$) when the observation point (r_0,ϕ_0) is in the illuminated region (shadow region) of the incident wave. Note that $(\xi_0^*)^2$ measures the excessive ray path from the source to the observation point via the edge of the half-plane. The two leading coefficients of the series in (II. 6) have been generally determined in Ref. 4 and in the present case are given by

$$\hat{v}_0(r_0,\phi_0) = -\frac{\exp(i\pi/4)}{2(2\pi)^{1/2}} z_0^i(r_1=d, \phi_1=\pi-\Omega)$$

$$\times r_0^{-1/2}\sec\tfrac{1}{2}(\phi_0-\Omega), \tag{II. 13}$$

$$\hat{v}_1(r_0,\phi_0) = -\frac{\exp(i\pi/4)}{2(2\pi)^{1/2}}\Big[z_1^i(d,\pi-\Omega)r_0^{-1/2}\sec\tfrac{1}{2}(\phi_0-\Omega)$$

$$+\frac{1}{4}\Big(\frac{3}{2d}z_0^i(d,\pi-\Omega)\sin^2\tfrac{1}{2}(\phi_0-\Omega)$$

$$-\frac{\partial z_0^i(d,\pi-\Omega)}{\partial r_1}\cos(\phi_0-\Omega)$$

$$+\frac{1}{d}\frac{\partial z_0^i(d,\pi-\Omega)}{\partial\phi_1}\sin(\phi_0-\Omega)\Big)r_0^{-1/2}\sec^3\tfrac{1}{2}(\phi_0-\Omega)$$

$$+\tfrac{1}{4}z_0^i(d,\pi-\Omega)r_0^{-3/2}\sec^3\tfrac{1}{2}(\phi_0-\Omega)\Big]. \tag{II. 14}$$

There exists a recursive formula for the determination of higher order $\hat{v}_m$.[3] They are not needed here since throughout this paper we are only interested in terms up to the order of $k^{-3/2}$.

The expression in (II. 6) for the total field is uniformly valid for all $0<r_0<\infty$ and $0\leq\phi_0\leq 2\pi$. It is convenient to interpret the first term $U(r_0,\phi_0)$ in (II. 6a) as a contribution to the total field associated with the incident field, while the second term $U(r_0, 4\pi-\phi_0)$ as that associated with the reflected field. Let us concentrate on $U(r_0,\phi_0)$ given in (II. 6b), and consider the following two cases:

(i) Away from the shadow boundary and the edge $|k^{1/2}\xi_0^*| \gg 1$: The use of (II. 8) into (II. 6b) leads to the conclusion that U^G recovers the classical geometrical

optics field, and U^d gives the edge-diffracted field with its first term identical to Keller's results.[9,10]

(ii) In the immediate neighborhood of the shadow boundary and/or edge $k^{1/2}\xi_0^* \to 0$: Both U^G and U^d become infinite, but the singularities in $\hat{F}(k^{1/2}\xi_0^*)$ cancel exactly those in $\{\hat{v}_m\}$. Consequently, U is continuous and well defined across $k^{1/2}\xi_0^* = 0$.

Between the above two extreme cases, $U(r_0, \phi_0)$ in (II. 6a) provides a smooth transition and, therefore, is called a uniform asymptotic expansion for large k. Similar comments apply to $U(r_0, 4\pi - \phi_0)$, the second term in (II. 6a).

C. Multiple scattering between plates

In this section we consider the multiple scattering of the incident field (II. 1) between the two parallel plates in Fig. 1 and derive recurrence relations for the multiply scattered fields.

First let us determine the total field $u_1^t(r_1, \phi_1)$ due to the scattering of the incident field (II. 1) at the upper plate. The solution of this Sommerfeld half-plane problem is well known, and can be written as (see Ref. 3)

$$u_1^t(r_1, \phi_1) = \exp[ik(r_1 - a)]\{F[-\sqrt{2kr_1}\cos\tfrac{1}{2}(\phi_1 + \tfrac{1}{2}\pi)] + \tau F[-\sqrt{2kr_1}\cos\tfrac{1}{2}(\phi_1 - \tfrac{1}{2}\pi)]\}. \quad \text{(II. 15)}$$

The latter result can be also derived by means of the uniform asymptotic theory. In the backward direction of the incident field $\phi_1 = 3\pi/2$, we may replace the first Fresnel integral in (II. 15) by its asymptotic expansion (II. 8) and the second Fresnel integral becomes equal to $F(0) = \frac{1}{2}$. Retaining only the leading terms we have

$$u_1^t(r_1, 3\pi/2) = u^i + \exp[ik(r_1 - a)] \times\{\tfrac{1}{2}\tau - [\exp(i\pi/4)/2\sqrt{2\pi}]k^{-1/2}r_1^{-1/2} + O(k^{-3/2})\}. \quad \text{(II. 16)}$$

Furthermore, in the interior region $0 \leq \phi_1 \leq \pi$, the use of (II. 8) in the second Fresnel integral in (II. 15) leads to

$$u_1^t(r_1, \phi_1) = \exp[ik(r_1 - a)]\{F[-\sqrt{2kr_1}\cos\tfrac{1}{2}(\phi_1 + \tfrac{1}{2}\pi)] + [\tau\exp(i\pi/4)/2\sqrt{2\pi kr_1}]\sec\tfrac{1}{2}(\phi_1 - \tfrac{1}{2}\pi) + O(k^{-3/2})\}, \quad 0 \leq \phi_1 \leq \pi. \quad \text{(II. 17)}$$

The field u_1^t acts as an incident field on the lower plate, and the resultant scattered field $u_2(r_0, \phi_0)$ is to be determined by means of the uniform asymptotic theory. However, the uniform theory cannot be immediately applied because of the fact that the incident field u_1^t in (II. 17) is not locally a cylindrical wave in the direction of $\phi_1 = \pi/2$. To circumvent this difficulty, we follow the method in Ref. 4: we replace the Fresnel integral $F(x)$ by its Taylor expansion in (II. 11), and (II. 17) becomes

$$u_1^t(r_1, \phi_1) = \exp[ik(r_1 - a)]\left\{\tfrac{1}{2}\sum_{q=0}^{\infty}\frac{\exp(-iq\pi/4)}{\Gamma(\frac{1}{2}q+1)}(-1)^q(2kr_1)^{q/2} \times\cos^q\tfrac{1}{2}(\phi_1 + \tfrac{1}{2}\pi) + \frac{\tau\exp(i\pi/4)}{2\sqrt{2\pi kr_1}}\sec\tfrac{1}{2}(\phi_1 - \tfrac{1}{2}\pi) + O(k^{-3/2})\right\}, \quad 0 \leq \phi_1 \leq \pi. \quad \text{(II. 18)}$$

The representation in (II. 18) comprises an infinite sum of cylindrical waves centered at the upper edge $r_1 = 0$, and is convergent throughout the interior region $0 \leq \phi_1 \leq \pi$. We now perform a term-by-term application of the uniform theory. To each cylindrical-wave term in (II. 18) the uniform theory is applied, and the corresponding scattered field constituent may be evaluated. Collecting the latter constituents, we obtain the scattered field $u_2(r_0, \phi_0)$. We do not perform this computation in detail, since later on we will derive a general result for the scattered field u_n which includes u_2 as a special case.

Consider now the scattered field $u_n(r_1, \phi_1)$, n odd, arising at the upper edge, and $u_n(r_0, \phi_0)$, n even, arising at the lower edge. Uniform expansions for these fields will be derived valid in the interior region $0 \leq \phi_1 \leq \pi$, $0 \leq \phi_0 \leq \pi$. Similar to the discussion in Sec. 7 of Ref. 4, we introduce the following ansatz for the uniform expansions:

$$u_n(r_1, \phi_1) = \tfrac{1}{2}\exp\{ik[r_1 + (n-2)a]\}\left(\sum_{q=0}^{\infty}\exp(-iq\pi/4)u_{n,q}(r_1, \phi_1) \times(k^{1/2}\xi_1)^q + \frac{\exp(i\pi/4)}{\sqrt{2\pi}}k^{-1/2}\sum_{q=0}^{\infty}\exp(-iq\pi/4) \times v_{n,q}(r_1, \phi_1)(k^{1/2}\xi_1)^q + O(k^{-1})\right) - \delta_{n1}u^i, \quad n \text{ odd}, \ 0 \leq \phi_1 \leq \pi, \quad \text{(II. 19)}$$

$$u_n(r_0, \phi_0) = \frac{\tau}{2}\exp\{ik[r_0 + (n-2)a]\}\left(\sum_{q=0}^{\infty}\exp(-iq\pi/4)u_{n,q}(r_0, \phi_0) \times(k^{1/2}\xi_0)^q + \frac{\exp(i\pi/4)}{\sqrt{2\pi}}k^{-1/2}\sum_{q=0}^{\infty}\exp(-iq\pi/4) \times v_{n,q}(r_0, \phi_0)(k^{1/2}\xi_0)^q + O(k^{-1})\right), \quad n \text{ even}, \ 0 \leq \phi_0 \leq \pi, \quad \text{(II. 20)}$$

where $\delta_{n1} = 1$ if $n = 1$ and $\delta_{n1} = 0$ if $n \neq 1$, and ξ_1 and ξ_0 are given by

$$\xi_1 = (r_1 + a - r_2)^{1/2}\,\mathrm{sgn}[\cos\tfrac{1}{2}(\phi_1 + \tfrac{1}{2}\pi)] = \left(\frac{4ar_1}{r_1 + a + r_2}\right)^{1/2}\cos\tfrac{1}{2}(\phi_1 + \tfrac{1}{2}\pi), \quad \text{(II. 21)}$$

$$\xi_0 = (r_0 + a - r_{-1})^{1/2}\,\mathrm{sgn}[\cos\tfrac{1}{2}(\phi_0 + \tfrac{1}{2}\pi)] = \left(\frac{4ar_0}{r_0 + a + r_{-1}}\right)^{1/2}\cos\tfrac{1}{2}(\phi_0 + \tfrac{1}{2}\pi). \quad \text{(II. 22)}$$

The ansatz in (II. 19) and (II. 20) describes the first and second terms of a high-frequency expansion in inverse powers of k. Each of these terms is represented by a convergent Taylor series with coefficients $\{u_{n,q}\}$ and $\{v_{n,q}\}$, respectively, which are to be determined. It should be emphasized that each of these Taylor series is to be considered in its entirety and should not be looked at as a series that can be truncated after several terms. Once the scattered fields $\{u_n\}$ are determined in the interior region from (II. 19) and (II. 20), those in the exterior region $\pi \leq \phi_1 \leq 2\pi$, $\pi \leq \phi_0 \leq 2\pi$ follow immediately from the symmetry relation in (II. 4).

For $n = 1$, the expansion (II. 19) should agree with (II. 2) and (II. 18), thus yielding

$$u_{1,q}(r_1,\phi_1)=\frac{(-1)^q}{\Gamma(q/2+1)}\frac{(2r_1)^{q/2}\cos^q\frac{1}{2}(\phi_1+\frac{1}{2}\pi)}{\xi_1^q}$$
$$=\frac{(-1)^q}{\Gamma(\frac{1}{2}q+1)}\left(\frac{r_1+a+r_2}{2a}\right)^{q/2},\qquad\text{(II. 23)}$$
$$v_{1,q}(r_1,\phi_1)=\delta_{q0}\tau r_1^{-1/2}\sec\tfrac{1}{2}(\phi_1-\tfrac{1}{2}\pi).\qquad\text{(II. 24)}$$

Scattering of the incident field u_n at the upper or lower plates gives rise to the scattered field u_{n+1}. The field u_{n+1} can be determined by a term-by-term application of the uniform asymptotic theory as summarized in Sec. IIB. The result for u_{n+1} thus obtained is to be compared with the ansatz (II. 19) and (II. 20) with n replaced by $(n+1)$. By equating corresponding terms we are led to a set of recurrence relations for the coefficients $u_{n,q}$ and $v_{n,q}$. It is found that the recurrence relations are exactly the same as those given in Refs. 4 and 5. Upon specializing to $\phi_0=\phi_1=\pi/2$, the recurrence relations become

$$u_{n+1,m}(r_0,\pi/2)=\tfrac{1}{2}\sum_{q=0}^{\infty}\frac{u_{n,q}(r_0+a,\pi/2)}{\Gamma(m/2-q/2+1)}\left(\frac{r_0}{r_0+2a}\right)^{q/2},\qquad\text{(II. 25)}$$

$$v_{n+1,m}(r_0,\pi/2)=\tfrac{1}{2}\sum_{q=0}^{\infty}\frac{v_{n,q}(r_0+a,\pi/2)}{\Gamma(m/2-q/2+1)}\left(\frac{r_0}{r_0+2a}\right)^{q/2}$$
$$-\frac{\tau}{2}\delta_{m0}u_{n,0}(a,\pi/2)r_0^{-1/2}-\delta_{m0}\pi^{1/2}$$
$$\times\frac{r_0^{1/2}}{a^{1/2}(r_0+a)^{1/2}}\sum_{q=0}^{\infty}\frac{1}{\Gamma(\frac{1}{2}-q)}$$
$$\times\frac{\partial u_{n,2q}(r_0+a,\pi/2)}{\partial\phi_0}\left(\frac{r_0}{r_0+2a}\right)^q,\qquad\text{(II. 26)}$$

where $m=0,1,2,\cdots$ and $n=1,2,\cdots$, provided that the following "finiteness condition" is satisfied:

$$u_{n,0}(a,\pi/2)=\pi^{1/2}\left(\frac{r_0+a}{a}\right)^{1/2}\sum_{q=0}^{\infty}\frac{u_{n,2q}(r_0+a,\pi/2)}{\Gamma(\frac{1}{2}-q)}\left(\frac{r_0}{r_0+2a}\right)^q.\qquad\text{(II. 27)}$$

In Sec. IID, it will be shown that coefficients $\{u_{n,q}\}$ do indeed satisfy (II. 27). The recurrence relations (II. 25) and (II. 26) are accompanied by the initial values:

$$u_{1,q}(r_0,\pi/2)=\frac{(-1)^q}{\Gamma(q/2+1)}\left(\frac{r_0+a}{a}\right)^{q/2},$$
$$v_{1,q}(r_0,\pi/2)=\delta_{q0}\tau r_0^{-1/2},\qquad\text{(II. 28)}$$

which are taken from (II. 23) and (II. 24). Furthermore, according to Ref. 4, the derivative $\partial u_{n,q}/\partial\phi_0$, which appeared in (II. 26), is determined by an additional recurrence relation

$$\frac{\partial u_{n+1,m}(r_0,\pi/2)}{\partial\phi_0}=\frac{1}{2}\frac{r_0}{r_0+a}\sum_{q=0}^{\infty}\frac{1}{\Gamma(m/2-q/2+1)}$$
$$\times\frac{\partial u_{n,q}(r_0+a,\pi/2)}{\partial\phi_0}\left(\frac{r_0}{r_0+2a}\right)^{q/2},\qquad\text{(II. 29)}$$

subject to the initial conditions

$$\frac{\partial u_{1,q}(r_0,\pi/2)}{\partial\phi_0}=0,\qquad\text{(II. 30)}$$

which is obtained by differentiation of (II. 23). Hence all derivatives $\{\partial u_{n,q}/\partial\phi_0\}$ vanish and (II. 26) simplifies to

$$v_{n+1,m}(r_0,\pi/2)=\frac{1}{2}\sum_{q=0}^{\infty}\frac{v_{n,q}(r_0+a,\pi/2)}{\Gamma(m/2-q/2+1)}\left(\frac{r_0}{r_0+2a}\right)^{q/2}$$
$$-\tfrac{1}{2}\tau\delta_{m0}u_{n,0}(a,\pi/2)r_0^{-1/2}.\qquad\text{(II. 31)}$$

The latter recurrence relation holds for $n=1,2,\cdots$, $m=0,1,\cdots$. By defining

$$u_{0,0}(a,\pi/2)=-2,\quad v_{0,q}(r_0,\pi/2)=0,\quad\text{for } q=0,1,2,\cdots,$$

it is easily seen by comparing with (II. 28) that (II. 31) is also valid for $n=0$.

Let us summarize the results obtained so far. The coefficients $\{u_{n,q}(r_0,\pi/2)\}$ are determined by the recurrence relation and initial conditions

$$\begin{cases} u_{n+1,m}(r_0,\pi/2)=\frac{1}{2}\sum_{q=0}^{\infty}\frac{u_{n,q}(r_0+a,\pi/2)}{\Gamma(m/2-q/2+1)}\left(\frac{r_0}{r_0+2a}\right)^{q/2}, \\ \qquad n=1,2,\cdots,\ m=0,1,\cdots, \\ u_{1,q}(r_0,\pi/2)=\frac{(-1)^q}{\Gamma(q/2+1)}\left(\frac{r_0+a}{a}\right)^{q/2},\quad q=0,1,\cdots. \end{cases}\qquad\text{(II. 32)}$$

The coefficients $\{v_{n,q}(r_0,\pi/2)\}$ are determined by the recurrence relation and initial conditions:

$$\begin{cases} v_{n+1,m}(r_0,\pi/2)=\frac{1}{2}\sum_{q=0}^{\infty}\frac{v_{n,q}(r_0+a,\pi/2)}{\Gamma(m/2-q/2+1)}\left(\frac{r_0}{r_0+2a}\right)^{q/2} \\ \qquad -\frac{1}{2}\tau\delta_{m0}u_{n,0}(a,\pi/2)r_0^{-1/2}, \\ \qquad n,m=0,1,\cdots, \\ v_{0,q}(r_0,\pi/2)=0,\quad q=0,1,\cdots,\quad u_{0,0}(a,\pi/2)=-2. \end{cases}\qquad\text{(II. 33)}$$

The solution of the recurrence relations (II. 32) and (II. 33) will be given in Sec. IID.

Once the recurrence relations are solved, we may calculate the desired field solutions as below. Setting $\phi_1=\pi/2$ in (II. 19) and $\phi_0=\pi/2$ in (II. 20), we have

$$u_n(r_1,\pi/2)=\tfrac{1}{2}\exp\{ik[r_1+(n-2)a]\}\{u_{n,0}(r_1,\pi/2)$$
$$+[\exp(i\pi/4)/\sqrt{2\pi}\,]k^{-1/2}v_{n,0}(r_1,\pi/2)$$
$$+O(k^{-1})\}-\delta_{n1}\exp[ik(r_1-a)],\ n\text{ odd},\qquad\text{(II. 34)}$$
$$u_n(r_0,\pi/2)=\tfrac{1}{2}\tau\exp\{ik[r_0+(n-2)a]\}\{u_{n,0}(r_0,\pi/2)$$
$$+[\exp(i\pi/4)/\sqrt{2\pi}\,]k^{-1/2}v_{n,0}(r_0,\pi/2)+O(k^{-1})\},$$
$$n\text{ even}.\qquad\text{(II. 35)}$$

The total fields in the forward direction $\phi_0=3\pi/2$ and backward direction $\phi_1=3\pi/2$ of the incident field are given by

$$u^t(r_0,\phi_0=3\pi/2)=u_1^t(r_1=r_0+a,\phi_1=\pi/2)+u_2(r_0,\phi_0=3\pi/2)$$
$$+\sum_{n=2}^{\infty}[u_{2n-1}(r_1=r_0+a,\phi_1=\pi/2)+u_{2n}(r_0,\phi_0=3\pi/2)],\qquad\text{(II. 36)}$$

$$u^t(r_1,\phi_1=3\pi/2)=u_1^t(r_1,\phi_1=3\pi/2)+u_2(r_0=r_1+a,\phi_0=\pi/2)$$
$$+\sum_{n=2}^{\infty}[u_{2n-1}(r_1,\phi_1=3\pi/2)$$
$$+u_{2n}(r_0=r_1+a,\phi_0=\pi/2)].\qquad\text{(II. 37)}$$

Let us consider the first terms in (II. 36) and (II. 37) in a little more detail. Since $u_1^t=u^t+u_1$, it follows from (II. 34) with $n=1$ and the symmetry relation in (II. 4) that

$$u_1^t(r_1 = r_0 + a,\ \phi_1 = \pi/2)$$
$$= \tfrac{1}{2}\exp(ikr_0)\{u_{1,0}(r_0 + a, \pi/2) + [\exp(i\pi/4)/\sqrt{2\pi}\,]k^{-1/2}$$
$$\times v_{1,0}(r_0 + a, \pi/2) + O(k^{-1})\}, \tag{II. 38}$$
$$u_1^t(r_1, \phi_1 = 3\pi/2)$$
$$= \exp[-ik(r_1 + a)] + u_1(r_1, \phi_1 = 3\pi/2)$$
$$= \exp[-ik(r_1 + a)] - \tau u_1(r_1, \phi_1 = \pi/2)$$
$$= \exp[-ik(r_1 + a)] + \tau\exp[ik(r_1 - a)] - \tfrac{1}{2}\tau\exp[ik(r_1 - a)]$$
$$\times \{u_{1,0}(r_1, \pi/2) + [\exp(i\pi/4)/\sqrt{2\pi}\,]k^{-1/2}v_{1,0}(r_1, \pi/2)$$
$$+ O(k^{-1})\}. \tag{II. 39}$$

When (II. 4), (II. 34), (II. 35), (II. 38), and (II. 39) are used in (II. 36) and (II. 37), we have the expressions for the total field in the forward and backward directions:

$$u^t(r_0, \phi_0 = 3\pi/2)$$
$$= \tfrac{1}{2}\exp(ikr_0)\sum_{n=0}^{\infty}\exp(i2nka)\{[u_{2n+1,0}(r_0 + a, \pi/2)$$
$$- u_{2n+2,0}(r_0, \pi/2)] + [\exp(i\pi/4)/\sqrt{2\pi}\,]k^{-1/2}$$
$$\times [v_{2n+1,0}(r_0 + a, \pi/2) - v_{2n+2,0}(r_0, \pi/2)] + O(k^{-1})\}, \tag{II. 40}$$

$$u^t(r_1, \phi_1 = 3\pi/2)$$
$$= \exp[-ik(r_1 + a)] + \tau\exp[ik(r_1 - a)]$$
$$- \tfrac{1}{2}\tau\exp[ik(r_1 - a)]\{u_{1,0}(r_1, \pi/2) + [\exp(i\pi/4)/\sqrt{2\pi}\,]k^{-1/2}$$
$$\times v_{1,0}(r_1, \pi/2)\} + \tfrac{1}{2}\tau\exp[ik(r_1 - a)]\sum_{n=1}^{\infty}\exp(i2nka)$$
$$\times \{[u_{2n,0}(r_1 + a, \pi/2) - u_{2n+1,0}(r_1, \pi/2)]$$
$$+ [\exp(i\pi/4)/\sqrt{2\pi}\,]k^{-1/2}[v_{2n,0}(r_1 + a, \pi/2)$$
$$- v_{2n+1,0}(r_1, \pi/2)] + O(k^{-1})\}. \tag{II. 41}$$

It is interesting to note that the total field in the forward and backward directions depends on $\{u_{n,0}\}$ and $\{v_{n,0}\}$ only.

D. Solution of recurrence relations

Consider first the recurrence relation in (II. 32). The same recurrence relation, subject to a different initial condition, was discussed in Ref. 4, Appendix C, where it was solved by a generating-function technique. Employing the same technique, we introduce the generating function

$$F_n(r_0; z) = \sum_{q=0}^{\infty} u_{n,q}(r_0, \pi/2)(iz)^q, \tag{II. 42}$$

where z is a complex variable. Thus, it was shown in Ref. 4, Appendix C, that (II. 32) can be reduced to a recurrence relation for F_n expressed in terms of F_{n-1}, namely,

$$F_n(r_0; z) = \frac{1}{2\pi i}\int_{-\infty+i\alpha}^{\infty+i\alpha}\frac{\exp(-t^2)}{t - z}F_{n-1}\left(r_0 + a; t\left(\frac{r_0}{r_0 + 2a}\right)^{1/2}\right)dt \tag{II. 43}$$

where $\alpha < \mathrm{Im}z$. By repeated application of (II. 43), F_n can be expressed in terms of F_1:

$$F_n(r_0; z) = \left(\frac{1}{2\pi i}\right)^{n-1}\int_{-\infty+i\alpha_1}^{\infty+i\alpha_1}\cdots\int_{-\infty+i\alpha_{n-1}}^{\infty+i\alpha_{n-1}}\exp\left(-\sum_{m=1}^{n-1}t_m^2\right)$$
$$\times (t_1 - z)^{-1}\prod_{m=2}^{n-1}\left[t_m - t_{m-1}\left(\frac{r_0 + (m-2)a}{r_0 + ma}\right)^{1/2}\right]^{-1}$$
$$\times F_1\left[r_0 + (n-1)a; t_{n-1}\left(\frac{r_0 + (n-2)a}{r_0 + na}\right)^{1/2}\right]$$
$$\times dt_1\, dt_2\cdots dt_{n-1}, \tag{II. 44}$$

where $\alpha_1 < \mathrm{Im}z$, $\alpha_m < \alpha_{m-1}[r_0 + (m-2)a]^{1/2}[r_0 + ma]^{-1/2}$, $m = 2, 3, \ldots, (n-1)$. From the initial condition in (II. 32) and the definition of F_1 in (II. 42), we find

$$F_1(r_0; z) = \sum_{q=0}^{\infty}\frac{(-1)^q}{\Gamma(q/2+1)}\left(\frac{r_0 + a}{a}\right)^{q/2}(iz)^q$$
$$= 2F\left[\exp(-i\pi/4)z\left(\frac{r_0 + a}{a}\right)^{1/2}\right]$$

according to (II. 11). Using a well-known integral representation for the Fresnel integral F, we have

$$F_1(r_0; z) = \frac{1}{\pi i}\int_{-\infty+i\alpha}^{\infty+i\alpha}\frac{\exp(-t^2)}{t + z[(r_0 + a)/a]^{1/2}}\,dt, \tag{II. 45}$$

where $\alpha < -\mathrm{Im}z\sqrt{(r_0 + a)/a}$, which is to be substituted in (II. 44). After simplification in a manner similar to that given in Ref. 4, Appendix C, we have the desired expression of F_n:

$$F_n(r_0; z) = 2\pi^{-n/2}\left(\frac{a}{r_0}\right)^{1/2}\int_0^{\infty}\cdots\int_0^{\infty}\exp\left[2ix_1 z\left(\frac{r_0 + a}{r_0}\right)^{1/2}\right.$$
$$\left. - \frac{r_0 + a}{r_0}x_1^2 - 2\sum_{m=2}^{n-1}x_m^2 + 2\sum_{m=1}^{n-2}x_m x_{m+1} - 2x_{n-1}x_n - x_n^2\right]$$
$$\times dx_1\, dx_2\cdots dx_n. \tag{II. 46}$$

The result in (II. 46) can easily be expanded in a power series of (iz), comparable to (II. 42). Then it is found that the solution of $u_{n,q}(r_0, \pi/2)$ is given by

$$u_{n,q}(r_0, \pi/2) = \frac{2^{q+1}}{q!}\left(\frac{a}{r_0}\right)^{1/2}\left(\frac{r_0 + a}{r_0}\right)^{q/2}J_{n,q}(r_0),$$
$$n = 2, 3, \cdots,\ q = 0, 1, 2, \cdots, \tag{II. 47}$$

where $J_{n,q}$ is an n-fold integral defined by

$$J_{n,q}(r_0) = \pi^{-n/2}\int_0^{\infty}\cdots\int_0^{\infty}x_1^q\exp\left(-\frac{r_0 + a}{r_0}x_1^2\right.$$
$$\left. - 2\sum_{m=2}^{n-1}x_m^2 + 2\sum_{m=1}^{n-2}x_m x_{m+1} - 2x_{n-1}x_n - x_n^2\right)$$
$$\times dx_1\, dx_2\cdots dx_n. \tag{II. 48}$$

The result in (II. 47) and (II. 48) together with the initial coefficient $u_{1,q}(r_0, \pi/2)$ in (II. 32) constitutes the solution for the recurrence relation in (II. 32). It can be shown that this solution satisfies the "finiteness condition" in (II. 27).

Next let us turn to the second recurrence relation in (II. 33). Except for the inhomogeneous term, this relation is identical to Eq. (C4) in Ref. 4. Hence, its solution can be derived in exactly the same manner with the result

$$v_{n,q}(r_0, \pi/2) = -\frac{\tau 2^q}{q!\sqrt{r_0}}\left(\frac{r_0 + a}{r_0}\right)^{q/2}\sum_{m=1}^{n}J_{m-1,0}(a)I_{n-m,q}(r_0),$$
(II. 49)

where

$$J_{0,0}(a) = -1, \quad J_{1,0}(a) = \tfrac{1}{2}, \quad I_{0,q}(r_0) = \delta_{q0}, \tag{II. 50}$$

$$I_{n,q}(r_0) = \pi^{-n/2} \int_0^\infty \cdots \int_0^\infty x_1^q \exp\left(-\frac{r_0 + a}{r_0} x_1^2 - 2\sum_{m=2}^{n} x_m^2 + 2\sum_{m=1}^{n-1} x_m x_{m+1}\right) dx_1\, dx_2 \cdots dx_n,$$

$$n = 1, 2, \cdots, \quad q = 0, 1, 2, \cdots. \tag{II. 51}$$

According to (II. 40) and (II. 41), the total field in the directions $\phi_0 = 3\pi/2$ and $\phi_1 = 3\pi/2$ only depends on the coefficients $\{u_{n,0}(r_0, \pi/2)\}$ and $\{v_{n,0}(r_0, \pi/2)\}$. Therefore, we present the special results of (II. 47) and (II. 49):

$$u_{1,0}(r_0, \pi/2) = 1; \quad u_{n,0}(r_0, \pi/2) = 2(a/r_0)^{1/2} J_{n,0}(r_0), \quad n = 2, 3, \cdots, \tag{II. 52}$$

$$v_{n,0}(r_0, \pi/2) = -\tau (1/r_0)^{1/2} \sum_{m=1}^{n} J_{m-1,0}(a) I_{n-m,0}(r_0), \quad n = 1, 2, \cdots. \tag{II. 53}$$

For later use, we derive simple closed-form results for $u_{2,0}$ and $v_{2,0}$. We evaluate the double integral

$$J_{2,0}(r_0) = \frac{1}{\pi} \int_0^\infty \int_0^\infty \exp\left(-\frac{r_0 + a}{r_0} x_1^2 - 2x_1 x_2 - x_2^2\right) dx_1\, dx_2 \tag{II. 54}$$

by introducing the new variables $y_1 = (a/r_0)^{1/2} x_1$, $y_2 = x_1 + x_2$; then, (II. 54) passes into

$$J_{2,0}(r_0) = (1/\pi)(r_0/a)^{1/2} \int\int_S \exp(-y_1^2 - y_2^2)\, dy_1\, dy_2, \tag{II. 55}$$

where S is a sector described by $y_1 \geq 0$, $y_2 \geq (r_0/a)^{1/2} y_1$. The sector S has an interior angle $(\pi/2) - \tan^{-1}(r_0/a)^{1/2}$. Thus, we find easily

$$J_{2,0}(r_0) = \left(\frac{r_0}{a}\right)^{1/2} \frac{(\pi/2) - \tan^{-1}(r_0/a)^{1/2}}{2\pi} \tag{II. 56}$$

and, consequently,

$$u_{2,0}(r_0, \pi/2) = \tfrac{1}{2} - (1/\pi) \tan^{-1}(r_0/a)^{1/2}. \tag{II. 57}$$

The latter result has been checked by a direct computation based on (II. 32). The coefficient $v_{2,0}$ in (II. 49) becomes

$$v_{2,0}(r_0, \pi/2) = -\tau r_0^{-1/2} [J_{0,0}(a) I_{1,0}(r_0) + J_{1,0}(a) I_{0,0}(r_0)] = \tfrac{1}{2}\tau (r_0 + a)^{-1/2} - \tfrac{1}{2}\tau r_0^{-1/2}, \tag{II. 58}$$

which was also checked by a direct computation based on (II. 33).

Furthermore, we need the values of $J_{n,0}(r_0)$ and $I_{n,0}(r_0)$ as $r_0 \to \infty$, and $J_{n,0}(a)$. Their evaluations are given in Ref. 14. Here we list the final results:

$$J_{n,0}(\infty) = \frac{1}{2\pi\sqrt{n-1}}, \quad n = 2, 3, \cdots, \tag{II. 59}$$

$$I_{n,0}(\infty) = \frac{1}{n!} \frac{\Gamma(n + \frac{1}{2})}{\Gamma(\frac{1}{2})}, \quad n = 0, 1, 2, \cdots, \tag{II. 60}$$

$$J_{n,0}(a) = \frac{-1}{n!} \frac{\Gamma(n - \frac{1}{2})}{\Gamma(-\frac{1}{2})}, \quad n = 0, 1, 2, \cdots. \tag{II. 61}$$

A comparison of (II. 59), (II. 61) with (II. 56) shows that they agree for $n = 2$.

In summary, the solutions of the recurrence relations in (II. 32) and (II. 33) are given, respectively, in (II. 47) and (II. 49). The explicit solutions for $\{u_{n,0}(r_0, \pi/2)\}$ and $\{v_{n,0}(r_0, \pi/2)\}$ as $r_0 \to \infty$ are given in (II. 50), (II. 52), (II. 53), and (II. 59)—(II. 61).

E. Far fields in the forward and backward directions

Consider first the total field $u^t(r_0, 3\pi/2)$ in the forward direction $\phi_0 = 3\pi/2$, as given in (II. 40). On substitution of results in (II. 52), (II. 53), (II. 57), and (II. 58) for $\{u_{n,0}\}$ and $\{v_{n,0}\}$, we obtain

$$u^t\left(r_0, \frac{3\pi}{2}\right) = \exp(ikr_0)\left\{\left[\frac{1}{4} + \frac{1}{2\pi}\tan^{-1}\left(\frac{r_0}{a}\right)^{1/2}\right] + a^{1/2}\sum_{n=1}^{\infty} \exp(i2nka)\left(\frac{J_{2n+1,0}(r_0 + a)}{(r_0 + a)^{1/2}} - \frac{J_{2n+2,0}(r_0)}{r_0^{1/2}}\right) + k^{-1/2}\frac{\tau \exp(i\pi/4)}{2\sqrt{2\pi}}\left[\frac{1}{2\sqrt{r_0 + a}} + \frac{1}{2\sqrt{r_0}} + \sum_{n=1}^{\infty} \exp(i2nka)\left(\frac{-1}{\sqrt{r_0 + a}} \sum_{m=1}^{2n+1} J_{m-1,0}(a) \times I_{2n+1-m,0}(r_0 + a) + \frac{1}{\sqrt{r_0}} \sum_{m=1}^{2n+2} J_{m-1,0}(a) \times I_{2n+2-m,0}(r_0)\right)\right] + O(k^{-1})\right\}. \tag{II. 62}$$

It has been verified that the first term in (II. 62), i. e.,

$$u^t(r_0, 3\pi/2)\big|_{\text{first term}} = \exp(ikr_0)[\tfrac{1}{4} + (1/2\pi)\tan^{-1}(r_0/a)^{1/2}] \tag{II. 63}$$

agrees with the result that is obtained by specialization of a rigorous asymptotic expansion for the field due to Jones.[8] Jones did not take into account the interaction between the edges of the two plates, and hence did not obtain the other terms in (II. 62). For large values of r_0, (II. 62) can be simplified, and we obtain the total far field in the forward direction

$$u^t\left(r_0, \frac{3\pi}{2}\right) = \exp(ikr_0)\left\{\frac{1}{2} - \frac{1}{2\pi}\left(\frac{a}{r_0}\right)^{1/2} + \left(\frac{a}{r_0}\right)^{1/2} \times \sum_{n=1}^{\infty} \exp(i2nka)[J_{2n+1,0}(\infty) - J_{2n+2,0}(\infty)] + \frac{\tau \exp(i\pi/4)}{2\sqrt{2\pi k r_0}}\left[1 + \sum_{n=1}^{\infty} \exp(i2nka)\left(-\sum_{m=1}^{2n+1} J_{m-1,0}(a) I_{2n+1-m,0}(\infty) + \sum_{m=1}^{2n+2} J_{m-1,0}(a) I_{2n+2-m,0}(\infty)\right)\right] + O(k^{-1}) + O(r_0^{-3/2})\right\}. \tag{II. 64}$$

The term of order $k^{-1/2}$ in (II. 64) can be considerably simplified. From (II. 60) and (II. 61) it follows that

$$-\sum_{m=1}^{p+1} J_{m-1,0}(a) I_{p+1-m,0}(\infty) = -\sum_{m=0}^{p} J_{m,0}(a) I_{p-m,0}(\infty) = \sum_{m=0}^{p} \frac{\Gamma(m - \frac{1}{2})\Gamma(p - m + \frac{1}{2})}{m!\,\Gamma(-\frac{1}{2})(p - m)!\,\Gamma(\frac{1}{2})}, \tag{II. 65}$$

where $p = 2n$ or $p = 2n + 1$. The latter sum in (II. 65) is just the coefficient of t^p in the power-series expansion of the product

$$\left(\sum_{q=0}^{\infty}\frac{\Gamma(q-\frac{1}{2})}{q!\,\Gamma(-\frac{1}{2})}t^q\right)\left(\sum_{q=0}^{\infty}\frac{\Gamma(q+\frac{1}{2})}{q!\,\Gamma(\frac{1}{2})}t^q\right)=(1-t)^{1/2}(1-t)^{-1/2}=1. \tag{II. 66}$$

Note that both series in (II. 66) are binomial series which have been explicitly summed. Since the coefficient of t^p in (II. 66) is equal to δ_{p0}, it follows immediately that

$$-\sum_{m=1}^{p+1}J_{m-1,0}(a)I_{p+1-m,0}(\infty)=\delta_{p0}. \tag{II. 67}$$

The use of (II. 67) and (II. 59) in (II. 64) leads to the final expression for the total far field in the forward direction:

$$u^t\left(r_0,\phi_0=\frac{3\pi}{2}\right)=\exp(ikr_0)\left[\frac{1}{2}-\frac{1}{2\pi}\left(\frac{a}{r_0}\right)^{1/2}+\frac{1}{2\pi}\left(\frac{a}{r_0}\right)^{1/2}\times\sum_{n=1}^{\infty}\left(\frac{1}{\sqrt{2n}}-\frac{1}{\sqrt{2n+1}}\right)\exp(i2nka)+\frac{\tau e^{i\pi/4}}{2\sqrt{2\pi kr_0}}+O(k^{-1})+O(r_0^{-3/2})\right], \tag{II. 68}$$

which agrees exactly with (A6) in Appendix A, which is an asymptotic expansion of the exact solution derived by the Wiener–Hopf technique.

Next consider the total field u^t in the backward direction $\phi_1=3\pi/2$, as given by (II. 41). On substitution of the results of (II. 52) and (II. 53) for the coefficients $\{u_{n,0}\}$, $\{v_{n,0}\}$, we obtain

$$u^t\left(r_1,\frac{3\pi}{2}\right)=\exp[-ik(r_1+a)]+\exp[ik(r_1-a)]\times\left[\frac{\tau}{2}+\tau a^{1/2}\sum_{n=1}^{\infty}\exp(i2nka)\left(\frac{J_{2n,0}(r_1+a)}{\sqrt{r_1+a}}-\frac{J_{2n+1,0}(r_1)}{\sqrt{r_1}}\right)\right]+\frac{\exp(i\pi/4)}{2\sqrt{2\pi}}k^{-1/2}\exp[ik(r_1-a)]\times\left[-r_1^{-1/2}+\sum_{n=1}^{\infty}\exp(i2nka)\times\left(-\frac{1}{\sqrt{r_1+a}}\sum_{m=1}^{2n}J_{m-1,0}(a)I_{2n-m,0}(r_1+a)+\frac{1}{\sqrt{r_1}}\sum_{m=1}^{2n+1}J_{m-1,0}(a)I_{2n+1-m,0}(r_1)\right)\right]+O(k^{-1}). \tag{II. 69}$$

As $r_1\to\infty$, (II. 69) can be simplified in a similar manner as the reduction of (II. 62). The final expression for the total far field in the backward direction is given by

$$u^t\left(r_1,\phi_1=\frac{3\pi}{2}\right)=\exp[-ik(r_1+a)]+\exp[ik(r_1-a)]\times\left[\frac{\tau}{2}+\frac{\tau}{2\pi}\left(\frac{a}{r_1}\right)^{1/2}\sum_{n=1}^{\infty}\left(\frac{1}{\sqrt{2n-1}}-\frac{1}{\sqrt{2n}}\right)\times\exp(i2nka)-\frac{\exp(i\pi/4)}{2\sqrt{2\pi kr_1}}+O(k^{-1})+O(r_1^{-3/2})\right] \tag{II. 70}$$

which again agrees with the asymptotic expansion of the exact solution given in (A3), Appendix A.

Let us now comment on several key steps in the derivation of the final solution in (II. 68) and (II. 70):

(i) In the calculation of multiple scattering between edges, the term-by-term application of the uniform asymptotic theory to the incident field in (II. 19) or (II. 20) is a formal procedure. As other formal procedures in ray-optical methods, its "justification" is its correct final result.

(ii) The derivation of the recurrence relations in (II. 25) and (II. 26) depends critically on the fact that the qth constituent of the incident field in (II. 19) [or (II. 20)] is proportional to ξ_1^q (or ξ_0^q), and ξ_1 is identically zero at the observation point, the location of the lower edge. Had the two plates been slightly staggered, simple recurrence relations as those in (II. 25) and (II. 26) could not have been derived.

(iii) The evaluations of the integral $J_{n,q}$ in (II. 48) and $I_{n,q}$ in (II. 51) are themselves interesting mathematical problems. In Ref. 14, two methods are used for their evaluations: one is elementary and involves transformation of variables in n-dimensional space and generating-function techniques, while the other uses integral equations, Fourier transforms, and Wiener–Hopf technique.

(iv) In two occasions in our derivation, the argument of analytical continuation was resorted to for extending the domain of convergence of the series involved. One occurs in the derivation of (II. 57) by a direct computation from (II. 32):

$$u_{2,0}\left(r_0,\frac{\pi}{2}\right)=\frac{1}{2}\sum_{q=0}^{\infty}\frac{u_{1,q}(r_0+a,\pi/2)}{\Gamma(1-\frac{1}{2}q)}\left(\frac{r_0}{r_0+2a}\right)^{q/2}=\frac{1}{2}\sum_{q=0}^{\infty}\frac{(-1)^q}{\Gamma(1-\frac{1}{2}q)\Gamma(1+\frac{1}{2}q)}\left(\frac{r_0+2a}{a}\right)^{q/2}\left(\frac{r_0}{r_0+2a}\right)^{q/2}=\frac{1}{2}-\frac{1}{\pi}\sum_{q=0}^{\infty}\frac{(-1)^q}{2q+1}\left(\frac{r_0}{a}\right)^{q+(1/2)}. \tag{II. 71}$$

Note that this series converges to the right-hand side of (II. 57) only in the range $0\le r_0\le a$. To show that (II. 57) also holds for $r_0>a$, one may invoke some analytical continuation argument. The other similar situation arises in the verifications of (II. 25) and (II. 27) by a direct substitution from (II. 47). [Yet another occurs later in the derivation of (IV. 19) in Sec. IV, where three series converge only when $|\eta|<1$.]

Some numerical results calculated from (II. 68) will be presented in Sec. III C.

III. NONSTAGGERED PARALLEL PLATES: SOLUTION BY MODIFIED DIFFRACTION COEFFICIENT

A. Outline of approach

In this part of the paper, the same problem sketched in Fig. 1, namely, the diffraction of a normally incident plane wave by two nonstaggered parallel plates is attacked by a different ray method—the method of modified diffraction coefficient described in Refs. 6 and 7. The solution so obtained turns out to be in complete agreement with the exact far field solution given in Appendix A.

First let us outline the general approach. From the symmetry of the problem it follows (see pp. 137–38 of Ref. 13) that the original problem sketched in Fig. 1 can be replaced by two auxiliary ones: (i) a problem with a perfect electric wall (where the tangential electric field is zero) at $y=a/2$ (Fig. 3a), and (ii) a problem with a

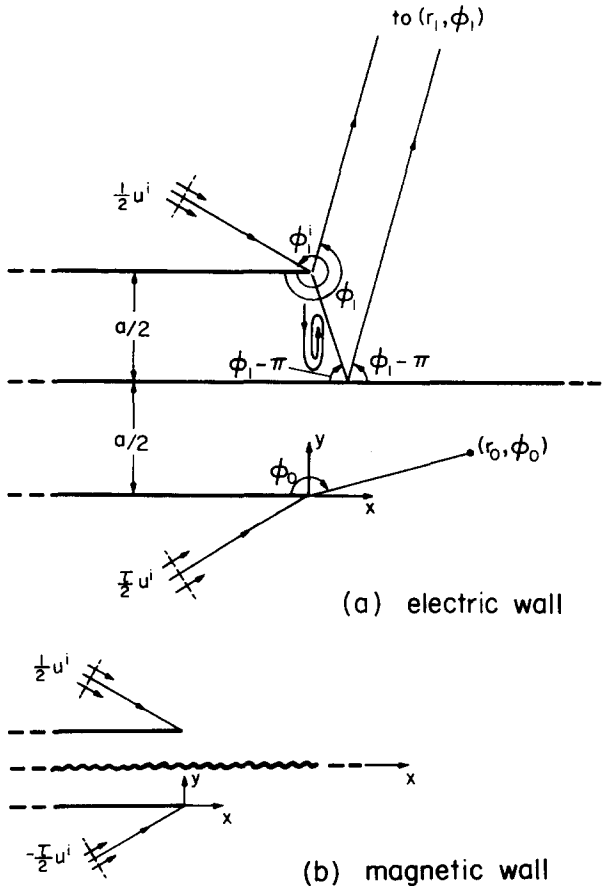


FIG. 3. Two auxiliary problems for the problem sketched in Fig. 1.

magnetic wall (where the tangential magnetic field is zero) at $y = a/2$ (Fig. 3b). Once these two auxiliary problems are solved their solutions will be properly superimposed to yield the solution of the original problem. For the convenience of applying the method of modified diffraction coefficient, we generalize the problem by letting the incident field u^i come from the direction ϕ_1^i, where $(3\pi/2) < \phi_1^i \leq 2\pi$:

$$u^i(r_1, \phi_1) = \exp[-ikr_1 \cos(\phi_1 - \phi_1^i)] \exp(-ika), \quad \text{(III. 1)}$$

where $\{r_1, \phi_1\}$ are polar coordinates with origin at $\{x = 0,\ y = a\}$ (Fig. 3a). The problem is to determine the total field at an observation point (r_1, ϕ_1), where $kr_1 \to \infty$ and $\pi \leq \phi_1 \leq 2\pi$. After the field is derived, we will set $\phi_1 = 3\pi - \phi_1^i$, and let ϕ_1^i go to $3\pi/2$ in order to obtain the desired field solutions in the forward and backward directions.

Let us concentrate on the problem with an electric wall (Fig. 3a). The incident field u^i in (III. 1) reaches the upper edge $x = 0$, $y = a$, and diffraction there produces a scattered field $u_1(r_1, \phi_1)$. The field u_1 propagates along diffracted rays emanating from the edge. Then the field along the diffracted ray in the direction $\phi_1 = \pi/2$ is specularly reflected from the electric wall at $y = a/2$. The reflected field u_2 strikes the upper edge again, and diffraction there produces a scattered field $u_3(r_1, \phi_1)$. Such a diffraction and reflection sequence continues. The total field at (r_1, ϕ_1) is then given by

$$u^t(r_1, \phi_1) = u^i(r_1, \phi_1) + [u_1(r_1, \phi_1) + u_{\text{int}}(r_1, \phi_1)] \times [1 - H(x) \exp(ika\,|\sin\phi_1|)]. \quad \text{(III. 2)}$$

Here u_{int} is the contribution from the interaction between the upper edge and the electric wall

$$u_{\text{int}}(r_1, \phi_1) = \sum_{n=1}^{\infty} u_{2n+1}(r_1, \phi_1). \quad \text{(III. 3)}$$

The term with the unit step function $H(x)$ in (III. 2) is to account for the possible specular reflection at the electric wall of the outgoing diffracted rays emanating from the upper edge. When the observation point (r_1, ϕ_1) has a negative x coordinate, i.e., $3\pi/2 < \phi_1 \leq 2\pi$, $H(x) = 0$ in agreement with the fact that there is no such a specular reflection. When $x > 0$, the same factor $\exp(ika\,|\sin\phi_1|)$ accounts for the contribution of the specular reflection for both TM case ($\tau = +1$) and TE case ($\tau = -1$). This independence of τ is due to the combination of the facts that (i) the scattered field u_n satisfies the symmetry relation in (II. 4) and (ii) the reflection coefficient of u_n from the electric wall is τ.

The central step is to determine u_{int}. In the present approach, instead of determining $u_2, u_3, \cdots$ successively, we will introduce a diffraction coefficient for the upper edge, a modified version of Keller's diffraction coefficient, and write down u_{int} in a single step.

B. Far fields in the forward and backward directions

Let us consider $u_1(r_1, \phi_1)$, the scattered field from the upper plate $x < 0$, $y = a$ due to an incident field (III. 1) (as if the electric wall at $y = a/2$ and lower plate $x < 0$, $y = 0$ were absent). Following Keller's geometrical theory of diffraction,[9,10] the far field solution of u_1 is the sum of the usual geometrical optics field and a diffracted field u_1^d. The latter is

$$u_1^d(r_1, \phi_1) \sim \frac{\exp[i(kr_1 + \pi/4)]}{2\sqrt{2\pi k r_1}} D(\phi_1, \phi_1^i) u^i(r_1 = 0), \quad kr_1 \to \infty, \quad \text{(III. 4)}$$

where $D(\phi_1, \phi_1^i)$ is known as Keller's diffraction coefficient

$$D(\phi_1, \phi_1^i) = -\left(\sec\frac{\phi_1 - \phi_1^i}{2} + \tau \sec\frac{\phi_1 + \phi_1^i}{2}\right). \quad \text{(III. 5)}$$

The result in (III. 4) and (III. 5) is not valid in the neighborhood of shadow boundary of the incident field $\phi_1 = \phi_1^i - \pi$, or that of reflected field $\phi_1 = 3\pi - \phi_1^i$. In those neighborhoods we may use the exact Sommerfeld half-plane solution for the scattered field:

$$u_1(r_1, \phi_1) = \exp(ikr_1)\left[-F\left(-\sqrt{2kr_1}\cos\frac{\phi_1 - \phi_1^i}{2}\right) + \tau F\left(\sqrt{2kr_1}\cos\frac{\phi_1 + \phi_1^i}{2}\right)\right] u^i(r_1 = 0), \quad \text{(III. 6)}$$

where F is the Fresnel integral defined in (II. 7). The result in (III. 6) can be also derived, of course, by the uniform asymptotic theory described in Sec. II B. When $\phi_1 \neq \phi_1^i - \pi$ and $\phi_1 \neq 3\pi - \phi_1^i$, the Fresnel integral in (III. 6) can be replaced by its asymptotic expansion according to (II. 8) and (II. 9). Retaining only the leading term, we recover (III. 4) and (III. 5), plus the usual geometrical optics field.

According to Ref. 6 and 7, the interaction term u_{int} in (III. 2) can be written in a similar form as (III. 4), and is given by

$$u_{int}(r_1, \phi_1) \sim \frac{\exp[i(kr_1 + \pi/4)]}{2\sqrt{2\pi k r_1}} [\bar{D}(\phi_1, \phi_1^i) - D(\phi_1, \phi_1^i)] \times u^i(r_1 = 0), \quad kr_1 \to \infty, \tag{III. 7a}$$

where $\bar{D}(\phi_1, \phi_1^i)$ is a modified diffraction coefficient and is related to Keller's diffraction coefficient in (III. 5) by

$$\bar{D}(\phi_1, \phi_1^i) = D(\phi_1, \phi_1^i) f(\phi_1) f(\phi_1^i), \tag{III. 7b}$$

$$f(\phi_1) = \begin{cases} 1/G_+(k|\cos\phi_1|), & \pi/2 < \phi_1 < 3\pi/2, \\ G_+(k|\cos\phi_1|), & 0 \le \phi_1 < \pi/2, \text{ or } 3\pi/2 < \phi_1 \le 2\pi. \end{cases} \tag{III. 7c}$$

The function $G_+(\alpha)$ is described in Appendix B. Several remarks about the formula in (III. 7) are in order:

(i) $\bar{D}$ is the exact diffraction coefficient for the edge diffraction by a perfectly conducting half-plane in the presence of a parallel, infinite electric wall at distance $a/2$. It was derived from the rigorous solution of a canonical problem.

(ii) In case that the infinite electric wall (Fig. 3a) is replaced by an infinite magnetic wall (Fig. 3b), (III. 7) remains valid after replacing $G_+(\alpha)$ by $\tilde{G}_+(\alpha)$. The function $\tilde{G}_+(\alpha)$ is also described in Appendix B.

(iii) The formula (III. 7) is valid for both TM and TE cases. The difference in these two cases enters through D in (III. 5).

(iv) Apparently, $f(\phi_1)$ and hence $\bar{D}(\phi_1, \phi_1^i)$ are not continuous across $\phi_1 = 3\pi/2$, since $G_+(0) = [1 - \exp(ika)]^{1/2} \neq 1$. However, in (III. 2) this discontinuity is compensated by the term with unit step function $H(x)$, and as a result the total field u^t is continuous across $\phi_1 = 3\pi/2$.

(v) In Refs. 6 and 7, u_1^q in (III. 4) and u_{int} in (III. 7) are combined in a single term. For the present application it is more convenient to separate out u_1^q, which is the component that becomes infinite on shadow boundaries and should be replaced by u_1 in (III. 6).

Concerning the result in (III. 7), we are particularly interested in the field exactly on the reflected shadow boundary. Letting $\phi_1 \to (3\pi - \phi_1^i)$ in (III. 7), we obtain in the limit

$$u_{int}(r_1, \phi_1 = 3\pi - \phi_1^i) \sim \frac{\exp[i(kr_1 + \pi/4)]}{2\sqrt{2\pi k r_1}} \left(2\tau \sin\phi_1^i \frac{kG_+'(k|\cos\phi_1^i|)}{G_+(k|\cos\phi_1^i|)}\right) u^i(r_1 = 0), \tag{III. 8}$$

where $G_+'(\alpha)$ means the derivative of $G_+(\alpha)$ with respect to α. $G_+'(\alpha)$ is also discussed in Appendix B.

In summary, for the problem sketched in Fig. 3a with an incident field in (III. 1), the total far field solution ($kr_1 \to \infty$) is given by (III. 2), (III. 6), and (III. 7) when $3\pi/2 < \phi_1^i \le 2\pi$, $\pi \le \phi_1 \le 2\pi$. For the special case $\phi_1 = 3\pi - \phi_1^i$ and $\phi_1^i \to 3\pi/2$, we obtain the total far field on shadow boundary of the reflected field from (III. 2), (III. 6), and (III. 8), namely,

electric wall: $u^t(r_1, \phi_1 = 3\pi/2)$

$$\sim \exp[-ik(r_1 + a)] + \exp[ik(r_1 - a)]G(0) \times \left[\frac{\tau}{2} - \frac{\exp(i\pi/4)}{2\sqrt{2\pi k r_1}}\left(1 + 2\tau \frac{kG_+'(0)}{G_+(0)}\right)\right], \quad kr_1 \to \infty, \tag{III. 9}$$

where $G(\alpha) = G_+(\alpha)G_+(-\alpha)$ is defined in Appendix B and we have written the factor $[1 - \exp(ika)]$ as $G(0)$. In the above derivation the case $\phi_1^i = 3\pi/2$ is obtained as a limit $\phi_1^i = 3\pi/2 + \delta$, $\delta \to 0+$. It can be shown that the identical result is obtained when the limit is approached from the other side $\phi_1^i = 3\pi/2 - \delta$, $\delta \to 0+$.

Following exactly the same procedure we can solve the problem sketched in Fig. 3b. For the special case $\phi_1 = 3\pi - \phi_1^i$ and $\phi_1^i \to 3\pi/2$, the total far field is found to be

magnetic wall: $u^t(r_1, \phi_1 = 3\pi/2)$

$$\sim \exp[-ik(r_1 + a)] + \exp[ik(r_1 - a)]\tilde{G}(0) \times \left[\frac{\tau}{2} - \frac{\exp(i\pi/4)}{2\sqrt{2\pi k r_1}}\left(1 + 2\tau \frac{k\tilde{G}_+'(0)}{\tilde{G}_+(0)}\right)\right], \quad kr_1 \to \infty. \tag{III. 10}$$

Note that (III. 10) is identical to (III. 9) except for the replacement of $[G(0), G_+(0), G_+'(0)]$ by $[\tilde{G}(0), \tilde{G}_+(0), \tilde{G}_+'(0)]$, as discussed in (ii) following (III. 7).

Now let us return to the original problem sketched in Fig. 1, with incident field given in (II. 1). The *scattered* far field in the forward direction $\phi_0 = 3\pi/2$ is simply $(\tau/2)$ times the difference of (III. 9) and (III. 10) after replacing (r_1, ϕ_1) by (r_0, ϕ_0). This is evident from the sketch in Fig. 3. Including the incident field (II. 1), we have the total far field in the forward direction:

$$u^t\left(r_0, \phi_0 = \frac{3\pi}{2}\right) \sim \tfrac{1}{2}\exp(ikr_0) + \frac{\exp[i(kr_0 + \pi/4)]}{2\sqrt{2\pi k r_0}} \left[\tau + [1 - \exp(-ika)]\frac{kG_+'(0)}{G_+(0)} + [1 + \exp(-ika)]\frac{k\tilde{G}_+'(0)}{\tilde{G}_+(0)}\right], \quad kr_0 \to \infty. \tag{III. 11}$$

The total far field in the backward direction $\phi_1 = 3\pi/2$ is simply $\frac{1}{2}$ times of the sum of (III. 10) and (III. 11), and the result is

$$u^t\left(r_1, \phi_1 = \frac{3\pi}{2}\right) \sim \exp[-ik(r_1 + a)] + \tfrac{1}{2}\tau\exp[ik(r_1 - a)] - \frac{\exp[ik(r_1 - a) + i(\pi/4)]}{2\sqrt{2\pi k r_1}} \times \left[1 + \tau[1 - \exp(ika)]\frac{kG_+'(0)}{G_+(0)} + \tau[1 + \exp(ika)]\frac{k\tilde{G}_+'(0)}{\tilde{G}_+(0)}\right], \quad kr_1 \to \infty. \tag{III. 12}$$

The results in (III. 11) and (III. 12) are in complete agreement with the rigorous far field solutions given by (A5) and (A2) in Appendix A. We emphasize that (III. 11) and (III. 12) are valid for arbitrary values of ka. When ka is large, we may use the asymptotic formulas for $G_+(\alpha)$, $\tilde{G}_+(\alpha)$, etc., in (III. 11) and (III. 12). Retaining the leading terms up to $O(k^{-1}a^{-1})$, we recover (II. 68) and (II. 70) exactly.

Compared with the derivation given in Sec. II, we arrive at the solution in (III. 11) and (III. 12) in fewer steps. The key to this simplification is that the interaction field u_{int} is calculated from (III. 7), instead of from (III. 3). Looking from a different viewpoint, it is rather satisfactory that the use of the uniform asymptotic theory in Sec. II also recovered the exact asymptotic solution. This was done without introducing a new canonical problem, with the interaction between two edges being "built up" from the local consideration of a single edge. In more general edge diffraction problems, formula (III. 7) may not be applicable, while the uniform asymptotic theory can always be employed. One such example is given in Sec. IV.

C. Numerical results and discussion

For the problem sketched in Fig. 1 with incident field given in (II. 1), the solutions for the total far field ($kr_0 \to \infty, kr_1 \to \infty$) in the forward and backward directions are given in (III. 11) and (III. 12), respectively. When ka is large, the solutions reduce to those in (II. 68) and (II. 70). Some remarks concerning the numerical evaluations of those results are in order.

First let us concentrate on (II. 68), and normalize it with respect to the incident field:

$$\left.\frac{u^t}{u^i}\right|_{\phi_0=3\pi/2} \approx \tfrac{1}{2} + \tau\frac{\exp(i\pi/4)}{2\sqrt{2\pi k r_0}} - \frac{1}{2\pi}\left(\frac{a}{r_0}\right)^{1/2}\left[1 - S\left(\frac{ka}{\pi}\right)\right], \tag{III. 13}$$

where $S(x)$ is a short notation for the infinite series

$$S(x) = \sum_{n=1}^{\infty}\left(\frac{1}{\sqrt{2n}} - \frac{1}{\sqrt{2n+1}}\right)\exp(i2n\pi x). \tag{III. 14}$$

The latter series is slowly convergent. It is advantageous to transform it into an integral:

$$\begin{aligned} S(x) &= \frac{1}{\sqrt{\pi}}\sum_{n=1}^{\infty}\exp(i2n\pi x)\left(\int_0^{\infty}\exp(-2nt)t^{-1/2}\,dt\right. \\ &\quad \left. - \int_0^{\infty}\exp[-(2n+1)t]t^{-1/2}\,dt\right) \\ &= \frac{2}{\sqrt{\pi}}\int_0^{\infty}\frac{\exp(i2\pi x - 2t^2)[1-\exp(-t^2)]}{1-\exp(i2\pi x - 2t^2)}\,dt, \end{aligned} \tag{III. 15}$$

which is rapidly convergent and can be easily evaluated by numerical integration. The series $S(x)$ is periodic with period 1, and in fact a Fourier series. For later use we examine the behavior of $S(x)$ in the vicinity of $x = 0$. Referring to Section 1.11 in Ref. 15, $S(x)$ can be expressed in terms of Lerch's transcendent $\Phi(z, s, v)$, viz.,

$$\begin{aligned} S(x) &= 2^{-1/2}\exp(i2\pi x)\{\Phi[\exp(i2\pi x), \tfrac{1}{2}, 1] \\ &\quad - \Phi[\exp(i2\pi x), \tfrac{1}{2}, \tfrac{3}{2}]\}. \end{aligned} \tag{III. 16}$$

By means of formula 1.11(8) in Ref. 15, we obtain the Taylor expansion of $S(x)$, and its leading terms are

$$S(x) = S(0) - \tfrac{1}{2}\pi\exp(-i\pi/4)x^{1/2} + O(x), \tag{III. 17}$$

valid around $x = 0$, where $x^{1/2} = i|x|^{1/2}$ when $x < 0$. The initial constant term in (III. 17) is equal to

$$S(0) = (1/\sqrt{2})[\zeta(\tfrac{1}{2}) - \zeta(\tfrac{1}{2}, \tfrac{3}{2})] = 0.\,395\,101\,3566\cdots, \tag{III. 18}$$

where $\zeta(s)$ and $\zeta(s, v)$ are, respectively, ordinary, and generalized zeta functions, and the numerical values were taken from a table in Ref. 16. Since $S(x)$ is periodic with period 1, the expansion in (III. 17) is also valid after replacing x by $(x - m)$, where m is an arbitrary integer. When this result is used in (III. 13), we have the normalized total field in the vicinity of $(ka/\pi) = m$, $m = 1, 2, 3, \cdots$ (i.e., the width a between the plates being a multiple of half wavelength):

$$\begin{aligned} \left.\frac{u^t}{u^i}\right|_{\phi_0=3\pi/2} &\approx \frac{1}{2} + \frac{1}{2\sqrt{2\pi k r_0}}\left[\exp(i\pi/4)\tau - 0.\,605\sqrt{2m}\right. \\ &\quad \left. - \exp(-i\pi/4)\frac{m\pi}{\sqrt{2}}\left(\frac{ka}{m\pi} - 1\right)^{1/2} + O\left(\frac{ka}{m\pi} - 1\right)\right]. \end{aligned} \tag{III. 19}$$

From (III. 13), (III. 15), and (III. 19), it follows that u^t/u^i is a smooth function of ka, except at $ka = m\pi$. At the latter locations, the amplitude and phase plots of u^t/u^i vs ka exhibit vertical tangents.

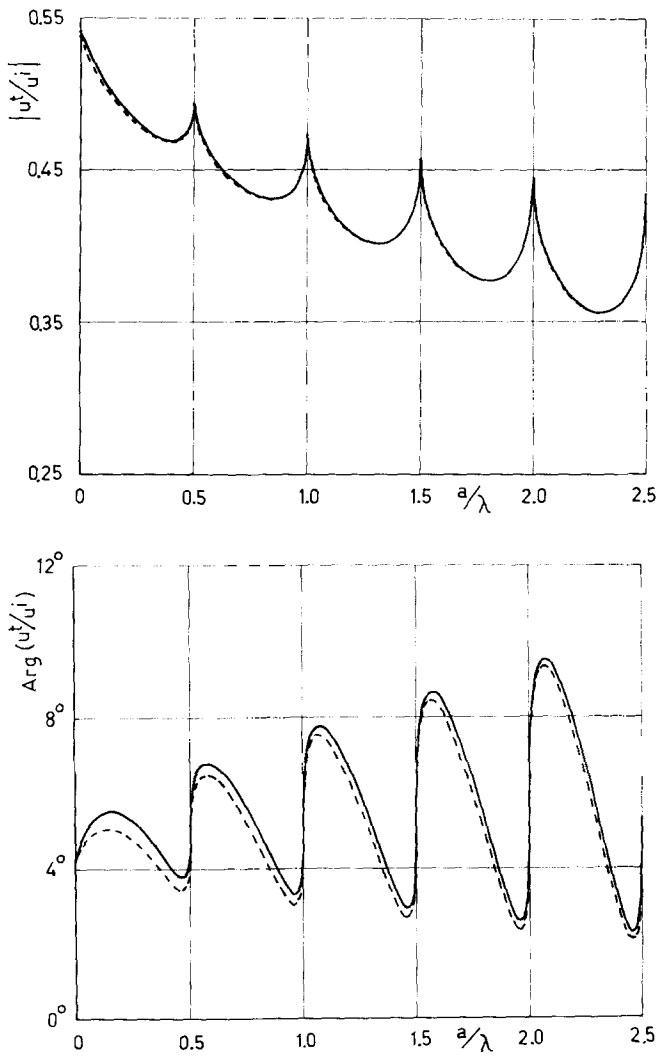


FIG. 4. Normalized total field on the incident shadow boundary of two nonstaggered parallel plates (Fig. 1) for TM case. The solid curves are calculated from (III. 20), and the dashed curves from (III. 13).

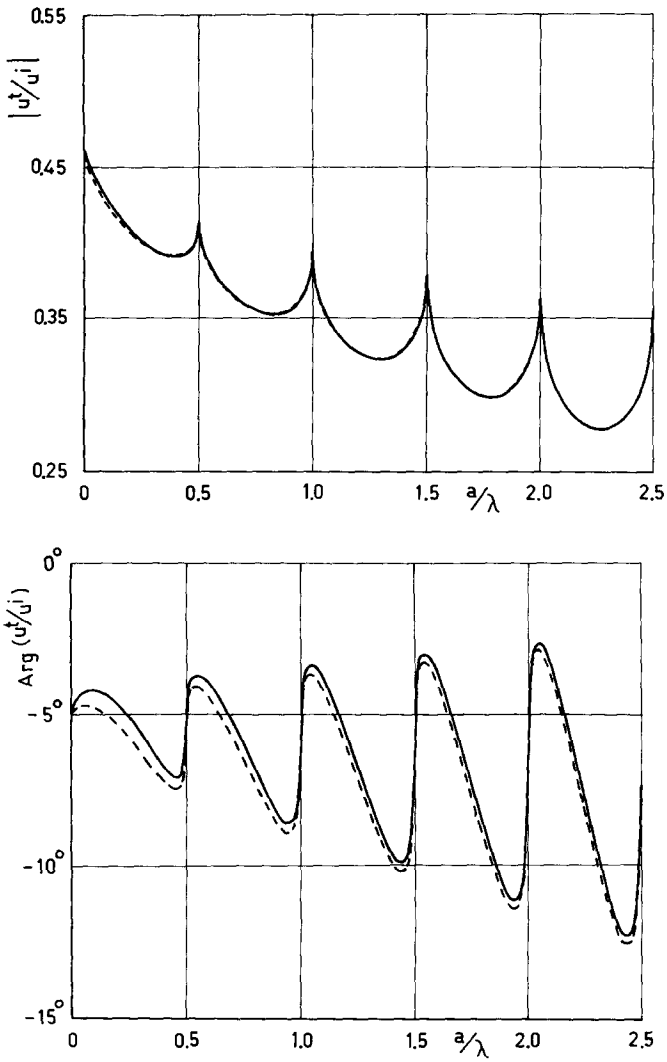


FIG. 5. Normalized total field on the incident shadow boundary of two nonstaggered parallel plates (Fig. 1) for TE case. The solid curves are calculated from (III. 20), and the dashed curves from (III. 13).

Next consider (III. 11), which after normalization becomes

$$\left.\frac{u^t}{u^i}\right|_{\phi_0=3\pi/2} \approx \frac{1}{2} + \frac{\exp(i\pi/4)}{2\sqrt{2\pi kr_0}}\left(\tau + [1-\exp(-ika)]\frac{kG_+'(0)}{G_+(0)} + [1+\exp(-ika)]\frac{k\tilde{G}_+'(0)}{\tilde{G}_+(0)}\right). \quad \text{(III. 20)}$$

The evaluation of the last two terms in (III. 20) is most easily done by numerical integration of the following representations, cf. Appendix B,

$$\frac{kG_+'(0)}{G_+(0)} = \frac{\exp(-i\pi/4)}{\pi\sqrt{2}}\, ka \int_{-\infty}^{\infty} \frac{\exp(ika - kat^2)}{1-\exp(ika-kat^2)} \times \frac{dt}{(1+\frac{1}{2}it^2)^{1/2}}, \quad \text{(III. 21)}$$

$$\frac{k\tilde{G}_+'(0)}{\tilde{G}_+(0)} = -\frac{\exp(-i\pi/4)}{\pi\sqrt{2}}\, ka \int_{-\infty}^{\infty} \frac{\exp(ika - kat^2)}{1+\exp(ika-kat^2)} \times \frac{dt}{(1+\frac{1}{2}it^2)^{1/2}}. \quad \text{(III. 22)}$$

Some caution should be taken when $ka = m\pi$, $m = 1, 2, 3, \cdots$. When m is even, the integral (III. 21) is divergent; when m is odd, the integral in (III. 22) is divergent. However, these divergent integrals are compensated by the factors $[1 \mp \exp(-ika)]$ in (III. 20) such that their combined values become zero in the respective cases. It can be shown that the amplitude and phase plots of u^t/u^i vs ka, based on (III. 20), exhibit the same behavior at "resonance values" $ka = m\pi$ as the previous curves based on (III. 13).

In Figs. 4 and 5, numerical results for the total far field in the forward direction are presented as a function of the plate separation-to-wavelength ratio a/λ, with the observation point at a fixed distance from the lower edge $r_0 = 2\lambda$. The solid curves are calculated from (III. 20)–(III. 22), while the dashed curves stem from (III. 13) and (III. 15). Note that these two sets of curves are in good agreement even for a is about half wavelength.

IV. STAGGERED PARALLEL PLATES

A. Statement of problem and approach

In this part of the paper, we consider the diffraction by two perfectly conducting, parallel plates staggered a length l. We assume l to be positive, finite, and not close to zero. The separation of the plates is $a/2$, which is written as b hereafter (Fig. 6). The incident field is that from an isotropic line source:

$$u^i(r_2, \phi_2) = (i/4)H_0^{(1)}(kr_2) = \frac{\exp[i(kr_2+\pi/4)]}{2\sqrt{2\pi kr_2}}[1 + (1/8ikr_2) + O(k^{-2})]. \quad \text{(IV. 1)}$$

The polar coordinates $\{r_0, \phi_0\}$, $\{r_1, \phi_1\}$, and $\{r_2, \phi_2\}$ have origins at the lower edge, the upper edge, and the source point, respectively. We are interested in the case when the line source, the two edges, and the observation point are exactly on a straight line (Fig. 6), i. e.,

$$\text{line source: } r_0 = c + d, \quad \phi_0 = \Omega, \quad \text{(IV. 2a)}$$

$$\text{observation point: } r_0 = r_0, \quad \phi_0 = \pi + \Omega. \quad \text{(IV. 2b)}$$

Except for the special situations $l = 0$ or $l = \infty$, rigorous analytical solution to this problem is not known. In Ref. 6, two coupled Wiener–Hopf equations were formulated and an approximate method for solving them valid for large kl was presented. However, for the case described in (IV. 2) (the most difficult one), no explicit result was obtained. Recently Jones[8] studied the same problem with a plane wave incidence (instead of incidence from a line source). He first considered the

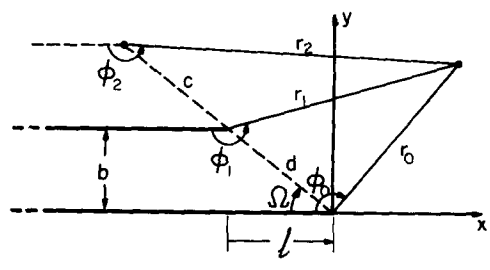


FIG. 6. Two staggered parallel plates illuminated by an incident cylindrical wave from a line source at $r_2 = 0$.

scattering of the incident plane wave at the upper plate and obtained an exact result for the scattered field in terms of Fresnel integrals. Then the diffraction of this scattered field at the second (lower) plate is treated by the conventional Wiener–Hopf technique. The final result thus obtained may be considered as the field scattered by two parallel plates when the interaction between the plates is ignored. Jones' analysis includes the special case when the incident plane wave propagates in a line through the edges of the two plates. It is this special case that is comparable to our result to be derived next. Excluding interaction terms, Jones' result and our result are found in agreement.

To attack the problem sketched in Fig. 6 with incident field in (IV. 1), we will use a combination of the uniform asymptotic theory (cf. Sec. II) and the method of modified diffraction coefficient (cf. Sec. III). Again our solution is asymptotic for large k, and contains terms up to and including the order of $k^{-3/2}$. The steps of solution are described below. The incident field u^i in (IV. 1) reaches the upper plate $x<-l$, $y=b$, and diffraction there produces a scattered field $u_1(r_1,\phi_1)$. The field u_1 on the diffracted ray traveling in the direction $\phi_1=\pi/2$ is bounced back and forth between the lower plate and the upper edge. This multiple interaction is accounted for by a single scattered field $u_{\rm int}(r_1,\phi_1)$ emanating from the upper edge. The calculation of u_1 and $u_{\rm int}$ follows a procedure similar to that used in Sec. III. For the diffraction at the lower edge $x=0$, $y=0$, the incident field is taken to be

$$\tilde{u}^i(r_1,\phi_1)=u^i+u_1+u_{\rm int}=\tilde{u}^i_{\rm cy}+\tilde{u}^i_{\rm no}. \tag{IV. 3}$$

In the neighborhood of the lower edge, $\tilde{u}^i$ is further divided into two components: cylindrical wave component $\tilde{u}^i_{\rm cy}$ and noncylindrical component $\tilde{u}^i_{\rm no}$. Their respective diffractions give rise to $u^t_{\rm cy}$ and $u^t_{\rm no}$, which are calculated by the uniform asymptotic theory described in Sec. II. The further successive diffraction of $u^t_{\rm no}$ by the upper edge results in $u^t_{\rm up}$. Successive diffraction of $u^t_{\rm cy}$ by the upper edge gives rise to a field of order k^{-2}, and hence this contribution is ignored. The total field solution in the direction $\phi_0=\pi+\Omega$, correct to the order $k^{-3/2}$, is then given by the sum of $u^t_{\rm cy}, u^t_{\rm no}$, and $u^t_{\rm up}$.

B. Far field solution in forward direction

The scattering of u^i in (IV. 1) at the upper plate gives rise to a scattered field $u_1(r_1,\phi_1)$. To derive an asymptotic expression of u_1 valid in the region $\pi/2\leq\phi_1\leq\pi$, we may use the uniform asymptotic theory summarized in Sec. II B. The result is

$$\begin{aligned}u_1(r_1,\phi_1)=&\frac{-\exp[i(kr_1+kc+\pi/4)]}{2\sqrt{2\pi kr_2}}\left(1+\frac{1}{8ikr_2}\right)F(-k^{1/2}\xi_1^*)\\&+\frac{\exp[i(kr_1+\pi/4)]}{2\sqrt{2\pi kr_1}}\,\frac{\exp[i(kc+\pi/4)]}{2\sqrt{2\pi kc}}\\&\times\left[\left(\frac{2r_1c}{r_2}\right)^{1/2}(\xi_1^*)^{-1}-\sec\frac{\phi_1-(2\pi-\Omega)}{2}\right.\\&\left.-\tau\sec\frac{\phi_1+(2\pi-\Omega)}{2}\right]+O(k^{-2}),\end{aligned} \tag{IV. 4}$$

where F is the Fresnel integral defined in (II. 7) and

$$\begin{aligned}\xi_1^*&=(r_1+c-r_2)^{1/2}\,\mathrm{sgn}\left(\cos\frac{\phi_1-(2\pi-\Omega)}{2}\right)\\&=\left(\frac{4r_1c}{r_1+c+r_2}\right)^{1/2}\cos\frac{\phi_1-(2\pi-\Omega)}{2}.\end{aligned} \tag{IV. 5}$$

For $\pi/2\leq\phi_1\leq\pi$ and $r_1\neq 0$, $u_1(r_1,\phi_1)$ given in (IV. 4) is finite and continuous everywhere.

The field u_1 on the diffracted ray traveling in the direction $\phi_1=\pi/2$ is bounced back and forth between the upper edge and the lower plate, resulting in a scattered field $u_{\rm int}$. Since l is assumed to be positive and not close to zero, this interaction is locally the same as that discussed in Sec. III. Thus, using (III. 7), one obtains

$$\begin{aligned}u_{\rm int}(r_1,\phi_1)=&\frac{\exp[i(kr_1+\pi/4)]}{2\sqrt{2\pi kr_1}}\,\frac{\exp[i(kc+\pi/4)]}{2\sqrt{2\pi kc}}\\&\times[\bar{D}(\phi_1,2\pi-\Omega)-D(\phi_1,2\pi-\Omega)]+O(k^{-2}).\end{aligned} \tag{IV. 6}$$

In the region $\pi/2\leq\phi_1\leq\pi$, $u_{\rm int}$ can be written more explicitly

$$\begin{aligned}u_{\rm int}(r_1,\phi_1)=&\frac{\exp[i(kr_1+\pi/4)]}{2\sqrt{2\pi kr_1}}\,\frac{\exp[i(kc+\pi/4)]}{2\sqrt{2\pi kc}}\\&\times\left(1-\frac{G_+[k\cos(2\pi-\Omega)]}{G_+(k|\cos\phi_1|)}\right)\\&\times\left(\sec\frac{\phi_1-(2\pi-\Omega)}{2}+\tau\sec\frac{\phi_1+(2\pi-\Omega)}{2}\right)\\&+O(k^{-2}).\end{aligned} \tag{IV. 7}$$

Assuming that Ω or ϕ_1 is not close to $\pi/2$, we may use the asymptotic expansion for $G_+(\alpha)$ given in (B11), Appendix B (remembering $a=2b$), and (IV. 7) passes into

$$\begin{aligned}u_{\rm int}(r_1,\phi_1)=&\frac{\exp[i(kr_1+\pi/4)]}{2\sqrt{2\pi kr_1}}\,\frac{\exp[i(kc+\pi/4)]}{2\sqrt{2\pi kc}}\\&\times\left(\frac{\exp(i\pi/4)}{2\sqrt{\pi kb}}\sum_{n=1}^{\infty}\frac{\exp(i2nkb)}{n^{3/2}}\right)\\&\times[\sec(2\pi-\Omega)+\sec\phi_1]\left(\sec\frac{\phi_1-(2\pi-\Omega)}{2}\right.\\&\left.+\tau\sec\frac{\phi_1+(2\pi-\Omega)}{2}\right)+O(k^{-2}),\end{aligned} \tag{IV. 8}$$

valid for $\pi/2<\phi_1\leq\pi$, away from $r_1=0$. It should be remarked that the result in (IV. 8) can be also derived by using the uniform asymptotic theory described in Sec. II. Such a derivation, however, is quite involved, whereas the use of (III. 7) enables us to write down (IV. 8) readily as we did above.

Next consider the diffraction at the lower edge $x=0$, $y=0$. The solutions of u_1 and $u_{\rm int}$ having been found, the incident field $\tilde{u}^i$ defined in (IV. 3) now can be written explicitly as $\tilde{u}^i=\tilde{u}^i_{\rm cy}+\tilde{u}^i_{\rm no}$, which consists of a cylindrical wave component $\tilde{u}^i_{\rm cy}$ and a noncylindrical wave component $\tilde{u}^i_{\rm no}$. From (IV. 3), (IV. 4), and (IV. 8) we find the cylindrical wave component to be

$$\tilde{u}^i_{\rm cy}(r_1,\phi_1)=\frac{\exp[i(kr_1+\pi/4)]}{2\sqrt{2\pi kr_1}}\,C(r_1,\phi_1)+O(k^{-2}),\qquad \pi/2<\phi_1\leq\pi, \tag{IV. 9}$$

where

$$C(r_1,\phi_1)=\frac{\exp[i(kc+\pi/4)]}{2\sqrt{2\pi kc}}\left[\left(\frac{2r_1c}{r_2}\right)^{1/2}(\xi_1^*)^{-1}-\sec\frac{\phi_1-(2\pi-\Omega)}{2}\right.$$
$$-\tau\sec\frac{\phi_1+(2\pi-\Omega)}{2}+\left(\frac{\exp(i\pi/4)}{2\sqrt{\pi kb}}\sum_{n=1}^{\infty}\frac{\exp(i2nkb)}{n^{3/2}}\right)$$
$$\times[\sec(2\pi-\Omega)+\sec\phi_1]$$
$$\left.\times\left(\sec\frac{\phi_1-(2\pi-\Omega)}{2}+\tau\sec\frac{\phi_1+(2\pi-\Omega)}{2}\right)\right]. \quad \text{(IV. 10)}$$

The noncylindrical wave component in (IV. 9) is found to be

$$\tilde{u}^i_{no}(r_1,\phi_1)=\frac{\exp[i(kr_1+kc+\pi/4)]}{2\sqrt{2\pi kr_2}}\left(1+\frac{1}{8ikr_2}\right)F(k^{1/2}\xi_1^*). \quad \text{(IV. 11)}$$

The diffraction of $\tilde{u}^i_{cy}, \tilde{u}^i_{no}$ at the lower edge gives rise to total field components u^t_{cy}, u^t_{no}, respectively. Their calculations are considered below.

For an incident cylindrical wave $\tilde{u}^i_{cy}$, we can apply formula (II. 6) directly. Retaining only the leading terms, we have

$$u^t_{cy}(r_0,\phi_0)=\frac{\exp[i(kr_0+kd+\pi/4)]}{2\sqrt{2\pi kr_1}}F(k^{1/2}\xi_0^*)C(r_1,\phi_1)$$
$$+\frac{\exp[i(kr_0+\pi/4)]}{2\sqrt{2\pi kr_0}}\frac{\exp[i(kd+\pi/4)]}{2\sqrt{2\pi kd}}$$
$$\times\left[\left(\frac{2r_0d}{r_1}\right)^{1/2}(\xi_0^*)^{-1}C(r_1,\phi_1)-\left(\sec\frac{\phi_0-\Omega}{2}\right.\right.$$
$$\left.\left.+\tau\sec\frac{\phi_0+\Omega}{2}\right)C(d,\pi-\Omega)\right]+O(k^{-2}), \quad \text{(IV. 12)}$$

valid for $\pi\leq\phi_0\leq 2\pi$, away from the lower edge $r_0=0$. The function ξ_0^* was defined in (II. 12). Of particular interest is the field in the forward direction $\phi_0=\pi+\Omega$. In this direction, $\xi_0^*=0$ and $\sec(\phi_0-\Omega)/2$ becomes infinite. However, the resultant singularities do cancel and u^t_{cy} remains finite as shown below. Let us assume that ϕ_0 deviates from $(\pi+\Omega)$ by a small number δ:

$$\phi_0=(\pi+\Omega)-\delta. \quad \text{(IV. 13)}$$

Then, it follows from simple geometry in Fig. 7 and the definitions in (II. 12) and (IV. 5) that

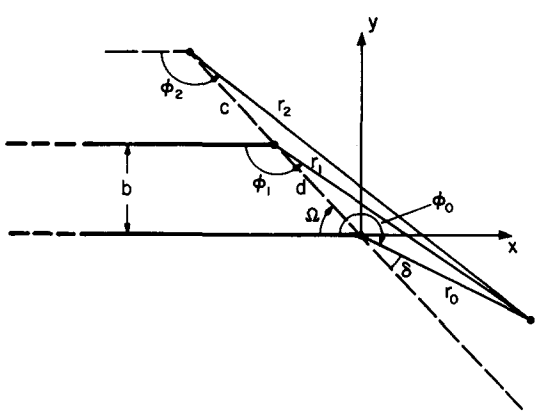


FIG. 7. Limiting case when the observation point falls on the line through the source point and two edges: r_0 is fixed and $\phi_0\to\pi+\Omega$ (or equivalently $\delta\to 0$).

$$\phi_1=(\pi-\Omega)+\frac{r_0}{r_0+d}\delta[1+A_0\delta^2+O(\delta^4)],\quad A_0=\frac{d(r_0-d)}{6(r_0+d)^2},$$

$$r_1=(r_0+d)[1+A_1\delta^2+O(\delta^4)],\quad A_1=-\frac{r_0d}{2(r_0+d)^2},$$

$$r_2=(r_0+d+c)[1+A_2\delta^2+O(\delta^4)],\quad A_2=-\frac{r_0(d+c)}{2(r_0+d+c)^2},$$

$$\xi_0^*=\delta\left(\frac{r_0d}{2(r_0+d)}\right)^{1/2}[1+A_3\delta^2+O(\delta^4)],\quad A_3=\frac{1}{8}\left(\frac{r_0d}{(r_0+d)^2}-\frac{1}{3}\right),$$

$$\xi_1^*=\delta\frac{r_0}{r_0+d}\left(\frac{(r_0+d)c}{2(r_0+d+c)}\right)^{1/2}[1+A_4\delta^2+O(\delta^4)],$$

$$A_4=\frac{1}{8}\left[\frac{r_0}{r_0+d+c}\left(\frac{d}{r_0+d}+\frac{d+c}{r_0+d+c}\right)-\frac{r_0^2+2r_0d+4d^2}{3(r_0+d)^2}\right] \quad \text{(IV. 14)}$$

Substituting (IV. 13) and (IV. 14) into (IV. 10) and (IV. 12) and letting $\delta\to 0$, we obtain

$$u^t_{cy}(r_0,\phi_0=\pi+\Omega)$$
$$=\frac{\exp[i(kr_0+kd+\pi/4)]}{2\sqrt{2\pi k(r_0+d)}}\frac{\exp[i(kc+\pi/4)]}{2\sqrt{2\pi kc}}$$
$$\times\left(\frac{\tau}{2\sin\Omega}+\frac{\exp(i\pi/4)\sin\Omega}{2\sqrt{\pi kb}\cos^2\Omega}\sum_{n=1}^{\infty}\frac{\exp(i2nkb)}{n^{3/2}}\right)$$
$$+\frac{\exp[i(kr_0+\pi/4)]}{2\sqrt{2\pi kr_0}}\frac{\exp[i(kd+\pi/4)]}{2\sqrt{2\pi kd}}\frac{\exp[i(kc+\pi/4)]}{2\sqrt{2\pi kc}}$$
$$\times\left(\frac{cr_0}{2(r_0+d+c)^2}+\frac{1}{\sin^2\Omega}+\frac{\tau r_0}{r_0+d}\frac{\cos\Omega}{\sin^2\Omega}\right)+O(k^{-2}). \quad \text{(IV. 15)}$$

We note also that the successive diffraction of u^t_{cy} by the upper edge (including interaction) leads to terms of $O(k^{-2})$ for the field in the direction $\phi_0=\pi+\Omega$. Hence they are ignored.

It remains to calculate u^t_{no}, the total field component due to the incidence of $\tilde{u}^t_{no}$. Because of the rapid variation of the Fresnel function across $\xi_1^*=0$, $\tilde{u}^i_{no}$ cannot be regarded as a cylindrical wave, and the uniform asymptotic theory cannot be directly applied to calculate its diffraction at the lower edge. Following the method in Ref. 4, we expand the Fresnel integral in a Taylor series around $\xi_1^*=0$, viz.,

$$\tilde{u}^i_{no}(r_1,\phi_1)=\exp(ikr_1)\left(\sum_{q=0}^{\infty}z^{(q)}(r_1,\phi_1)k^{(q-1)/2}\right.$$
$$\left.+\frac{1}{8i}\sum_{q=0}^{\infty}r_2^{-1}z^{(q)}(r_1,\phi_1)k^{(q-3)/2}\right), \quad \text{(IV. 16)}$$

where $z^{(q)}$ is determined from (II. 11) and is given by

$$z^{(q)}(r_1,\phi_1)=\frac{\exp[i(kc+\pi/4)]}{4\sqrt{2\pi r_2}}\frac{\exp(-iq\pi/4)}{\Gamma(q/2+1)}(\xi_1^*)^q,$$
$$q=0,1,2,\cdots. \quad \text{(IV. 17)}$$

Each term in (IV. 16) is now considered as a cylindrical wave constituent. We apply the formula of uniform asymptotic theory in (II. 6) to each constituent separately and then sum up the resultant fields to obtain u^t_{no}, namely,

$$u^t_{no}(r_0,\phi_0)$$
$$=\exp[ik(r_0+d)]\left[\sum_{q=0}^{\infty}\left(F(k^{1/2}\xi_0^*)+\frac{\exp(i\pi/4)}{2\pi}k^{-1/2}(\xi_0^*)^{-1}\right.\right.$$

$$\times \sum_{m=0}^{[(q+1)/2]} \Gamma(m+\tfrac{1}{2})(ik)^{-m}(\xi_0^*)^{-2m}\Big) z^{(q)}(r_1,\phi_1)k^{(q-1)/2}$$

$$+\frac{1}{8i}\sum_{q=0}^{\infty}\Big(F(k^{1/2}\xi_0^*)+\frac{\exp(i\pi/4)}{2\pi}\,k^{-1/2}(\xi_0^*)^{-1}$$

$$\times \sum_{m=0}^{[(q-1)/2]} \Gamma(m+\tfrac{1}{2})(ik)^{-m}(\xi_0^*)^{-2m}\Big) r_2^{-1} z^{(q)}(r_1,\phi_1)k^{(q-3)/2}\Big]$$

$$-\frac{\exp[i(kr_0+kd+\pi/4)]}{2\sqrt{2\pi r_0}}\Big[k^{-1}z^{(0)}(d,\pi-\Omega)\Big(\sec\frac{\phi_0-\Omega}{2}$$

$$+\tau\sec\frac{\phi_0+\Omega}{2}\Big)-ik^{-3/2}\frac{1}{4d}\frac{\partial z^{(1)}(d,\pi-\Omega)}{\partial\phi_1}$$

$$\times\Big(\sin(\phi_0-\Omega)\sec^3\frac{\phi_0-\Omega}{2}-\tau\sin(\phi_0+\Omega)\sec^3\frac{\phi_0+\Omega}{2}\Big)\Big]$$

$$+O(k^{-2}), \qquad \text{(IV. 18a)}$$

where ξ_0^* was defined in (II. 12), $[(q\pm1)/2]$ is the largest integer $\leq (q\pm1)/2$, and

$$z^{(0)}(d,\pi-\Omega)=\frac{\exp[i(kc+\pi/4)]}{4\sqrt{2\pi(c+d)}}, \qquad \text{(IV. 18b)}$$

$$\frac{\partial z^{(1)}(d,\pi-\Omega)}{\partial\phi_1}=\frac{\exp(ikc)}{4\pi}\,\frac{\sqrt{cd}}{c+d}. \qquad \text{(IV. 18c)}$$

In deriving (IV. 18a), we have made use of the fact that, due to the factor $(\xi_1^*)^q$, the incident field amplitude $z^{(q)}(r_1,\phi_1)$ and its first $(q-1)$ derivatives vanish at the lower edge $r_1=d$, $\phi_1=\pi-\Omega$. The result in (IV. 18) can be simplified considerably. The steps follow exactly those in Ref. 4, Sec. 6, except that terms of $O(k^{-3/2})$ in (IV. 18a) were not present in Ref. 4. After simplification, (IV. 18a) becomes

$$u_{no}^t(r_0,\phi_0)=\frac{\exp[i(kr_0+kc+kd+\pi/4)]}{2\sqrt{2\pi kr_2}}\Big\{\Big[F(k^{1/2}\xi_0^*)F(k^{1/2}\xi_1^*)$$

$$+\tfrac{1}{2}F(k^{1/2}\xi_0^*(1+\eta^2)^{1/2})-\tfrac{1}{2}\exp(-ik\xi_1^{*2})F(k^{1/2}\xi_0^*)$$

$$+G(\eta,k^{1/2}\xi_0^*)\Big]\Big(1+\frac{1}{8ikr_2}\Big)+\frac{\exp(i\pi/4)}{4\sqrt{\pi k}}(\xi_0^*)^{-1}$$

$$\times\frac{1}{(1+\eta^2)^{1/2}}-\frac{i}{4\pi k}(\xi_0^*)^{-2}\frac{\eta}{1+\eta^2}\Big\}$$

$$-\frac{\exp[i(kr_0+\pi/4)]}{2\sqrt{2\pi kr_0}}\,\frac{\exp[i(kc+kd+\pi/4)]}{2\sqrt{2\pi k(c+d)}}$$

$$\times\frac{1}{2}\Big(\sec\frac{\phi_0-\Omega}{2}+\tau\sec\frac{\phi_0+\Omega}{2}\Big)+\frac{\exp[i(kr_0+\pi/4)]}{2\sqrt{2\pi kr_0}}$$

$$\times\frac{i\exp[ik(c+d)]}{16\pi k(c+d)}\Big(\frac{c}{d}\Big)^{1/2}\Big(\sin(\phi_0-\Omega)\sec^3\frac{\phi_0-\Omega}{2}$$

$$-\tau\sin(\phi_0+\Omega)\sec^3\frac{\phi_0+\Omega}{2}\Big)$$

$$+O(k^{-2}), \qquad \text{(IV. 19)}$$

valid for $\pi\leq\phi_0\leq2\pi$, away from the edge $r_0=0$. In (IV. 19), the following notations were used:

$$\eta=\frac{\xi_1^*}{\xi_0^*}, \qquad \text{(IV. 20)}$$

$$G(\eta,t)=\frac{\exp(-i\eta^2t^2)}{2\pi}\int_0^{\eta}\frac{\exp(i\sigma^2t^2)}{1+\sigma^2}\,d\sigma. \qquad \text{(IV. 21)}$$

On the shadow boundary $\phi_0=\pi+\Omega$, $\xi_0^*=\xi_1^*=0$, and $\sec(\phi_0-\Omega)/2$ becomes infinity. As before, the resultant singularities do cancel, and a finite u_{no}^t is obtained, namely,

$$u_{no}^t(r_0,\phi_0=\pi+\Omega)$$

$$=\frac{\exp[i(kr_0+kd+kc+\pi/4)]}{2\sqrt{2\pi k(r_0+d+c)}}\Big[\frac{1}{4}+\frac{1}{2\pi}\tan^{-1}\Big(\frac{r_0c}{(r_0+d+c)d}\Big)^{1/2}\Big]$$

$$\times\Big(1+\frac{1}{8ik}\,\frac{1}{r_0+d+c}\Big)+\frac{\exp[i(kr_0+\pi/4)]}{2\sqrt{2\pi kr_0}}\,\frac{\exp[i(kc+kd+\pi/4)]}{2\sqrt{2\pi k(c+d)}}$$

$$\times\frac{\tau}{2\sin\Omega}+\frac{\exp[i(kr_0+\pi/4)]}{2\sqrt{2\pi kr_0}}\,\frac{\exp[i(kc+kd+\pi/2)]}{8\pi k(c+d)}\Big(\frac{c}{d}\Big)^{1/2}$$

$$\times\Big(-\tau\frac{\cos\Omega}{\sin^2\Omega}+A_5\Big)+O(k^{-2}), \qquad \text{(IV. 22)}$$

where

$$A_5=4\Big(3A_3-A_4+\tfrac{1}{2}A_2+\frac{2r_0c(A_4-A_3)}{(r_0+d)(c+d)}\Big)-\frac{1}{6}$$

$$=-\frac{1}{2}-\frac{(c+d)(r_0^2-cd-d^2)}{2(r_0+d)(r_0+c+d)^2}. \qquad \text{(IV. 23)}$$

The previous expansions (IV. 14) were intermediate in deriving these results. It remains to determine the diffraction of u_{no}^t by the upper edge. The field on the diffracted ray of u_{no}^t traveling in the direction $\phi_0=\Omega$ is

$$\frac{\exp[i(kr_0+\pi/4)]}{2\sqrt{2\pi kr_0}}\,\frac{\exp[i(kc+kd+\pi/4)]}{2\sqrt{2\pi k(c+d)}}\Big(\frac{1}{2}\Big)$$

$$\times(-1-\tau\sec\Omega)+O(k^{-3/2}), \qquad \text{(IV. 24)}$$

which was calculated from (III. 4) with $u^t(r_1=0)$ replaced by $\tilde{u}_{no}^t(d,\pi-\Omega)$ as given by (IV. 11). After diffraction at the upper edge, it gives rise to a field in the direction $\phi_1=\pi-\Omega$:

$$\frac{\exp[i(kr_1+\pi/4)]}{2\sqrt{2\pi kr_1}}\,\frac{\exp[i(kd+\pi/4)]}{2\sqrt{2\pi kd}}\,\frac{\exp[i(kc+kd+\pi/4)]}{2\sqrt{2\pi k(c+d)}}$$

$$\times(\tfrac{1}{2})(-1-\tau\sec\Omega)[-1-\tau\sec(\pi-\Omega)]+O(k^{-2}), \qquad \text{(IV. 25)}$$

which was calculated from (III. 4) with u^t equal to the quantity in (IV. 24). Diffraction of the field in (IV. 25) at lower edge simply reduces the dominant term by a half. Thus, the successive diffractions of u_{no}^t at the upper and lower edges result in a contribution to the total field in the direction $\phi_0=\pi+\Omega$:

$$u_{up}^t(r_0,\phi_0=\pi+\Omega)$$

$$=\frac{\exp[i(kr_0+kd+\pi/4)]}{2\sqrt{2\pi k(r_0+d)}}\,\frac{\exp[i(kd+\pi/4)]}{2\sqrt{2\pi kd}}$$

$$\times\frac{\exp[i(kc+kd+\pi/4)]}{2\sqrt{2\pi k(c+d)}}\Big(-\frac{\tan^2\Omega}{4}\Big)+O(k^{-2}). \qquad \text{(IV. 26)}$$

Further interaction of u_{up}^t with the upper edge is of no interest because it produces fields of $O(k^{-5/2})$ in the direction $\phi_0=\pi+\Omega$.

Summing up (IV. 15), (IV. 22), and (IV. 26), we obtain the desired total field in the direction $\phi_0=\pi+\Omega$ due to an incidence given in (IV. 1), viz.,

$$u^t(r_0,\phi_0=\pi+\Omega)=k^{-1/2}P+k^{-1}Q+k^{-3/2}R+O(k^{-2}), \qquad \text{(IV. 27a)}$$

where

$$P=\frac{\exp[i(kr_0+kd+kc+\pi/4)]}{2\sqrt{2\pi(r_0+d+c)}}\Big[\frac{1}{4}+\frac{1}{2\pi}\tan^{-1}\Big(\frac{r_0c}{(r_0+d+c)d}\Big)^{1/2}\Big], \qquad \text{(IV. 27b)}$$

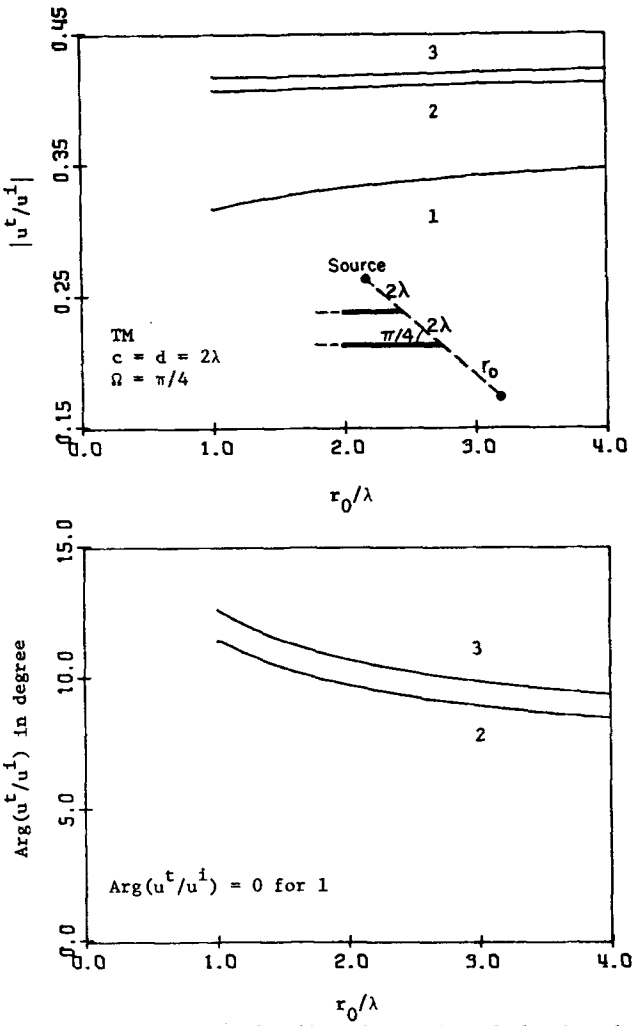


FIG. 8. Normalized total field on the incident shadow boundary of two staggered parallel plates (Fig. 6) for TM case. u^t is calculated from Eq. (IV.27) with one, two, and three terms (indicated by 1, 2, and 3, respectively, next to the curves); $u^i = \exp[ikr_2 + i\pi/4]/2\sqrt{2\pi kr_2}$ where $r_2 = c + d + r_0$.

$$Q = \frac{\exp[i(kr_0 + kd + kc + \pi/2)]}{8\pi}\left(\frac{1}{\sqrt{r_0(c+d)}} + \frac{1}{\sqrt{(r_0+d)c}}\right)\frac{\tau}{2\sin\Omega} \tag{IV.27c}$$

$$R = \frac{\exp[i(kr_0 + kd + kc + 3\pi/4)]}{8(2\pi)^{3/2}}\left\{\frac{\sqrt{2}}{\sqrt{(r_0+d)bc}}\frac{\sin\Omega}{\cos^2\Omega}\right.$$
$$\times\sum_{n=1}^{\infty}\frac{\exp(i2nkb)}{n^{3/2}} - \frac{\pi}{(r_0+d+c)^{3/2}}\left[\frac{1}{4} + \frac{1}{2\pi}\right.$$
$$\left.\times\tan^{-1}\left(\frac{r_0 c}{(r_0+d+c)d}\right)^{1/2}\right] + \frac{1}{\sqrt{r_0 dc}}\left(-\frac{c}{2(c+d)}\right.$$
$$+\frac{cd}{2(r_0+d)(r_0+d+c)} + \frac{1}{\sin^2\Omega} + \frac{d(r_0-c)}{(r_0+d)(c+d)}\frac{\tau\cos\Omega}{\sin^2\Omega}\Big)$$
$$\left. - \frac{\exp(i2kd)}{\sqrt{(r_0+d)(d+c)d}}\frac{\tan^2\Omega}{4}\right\}, \tag{IV.27d}$$

which is valid for Ω not close to $\pi/2$ (or $l \neq 0$). Some numerical results calculated from (IV.27) are presented in Figs. 8 and 9 pertaining to a configuration with $\Omega = \pi/4$, $c = d = 2\lambda$.

From the result in (IV.27) we can also obtain the total field $u^t(r_0, \phi_0 = \pi + \Omega)$ when the incident field is a plane wave coming from the direction $\phi_0 = \Omega$. To this end, let us multiply (IV.1) by the factor

$$2\sqrt{2\pi kc}\exp[-ik(c+d) - i(\pi/4)]. \tag{IV.28}$$

In the limit $c \to \infty$, the incident field in (IV.1) then becomes a plane wave given by

$$u^i = \exp[-ikr_0\cos(\phi_0 - \Omega)]. \tag{IV.29}$$

Multiplying the final result (IV.27) by the same factor (IV.28) and letting $c \to \infty$, we obtain the total field on the incident shadow boundary when the incident field is given by (IV.29), namely,

$$u^t(r_0, \phi_0 = \pi + \Omega) = \bar{P} + k^{-1/2}\bar{Q} + k^{-1}\bar{R} + O(k^{-3/2}), \tag{IV.30a}$$

where

$$\bar{P} = \exp(ikr_0)\left[\frac{1}{4} + \frac{1}{2\pi}\tan^{-1}\left(\frac{r_0}{d}\right)^{1/2}\right], \tag{IV.30b}$$

$$\bar{Q} = \frac{\exp[i(kr_0 + \pi/4)]}{2\sqrt{2\pi}}\left(\frac{1}{\sqrt{r_0}} + \frac{1}{\sqrt{r_0+d}}\right)\frac{\tau}{2\sin\Omega}, \tag{IV.30c}$$

$$\bar{R} = \frac{i\exp(ikr_0)}{8\pi\sqrt{r_0 d}}\left(-\frac{r_0}{2(r_0+d)} + \frac{1}{\sin^2\Omega} - \frac{d}{r_0+d}\frac{\tau\cos\Omega}{\sin^2\Omega}\right) + \bar{R}_{\mathrm{int}}, \tag{IV.30d}$$

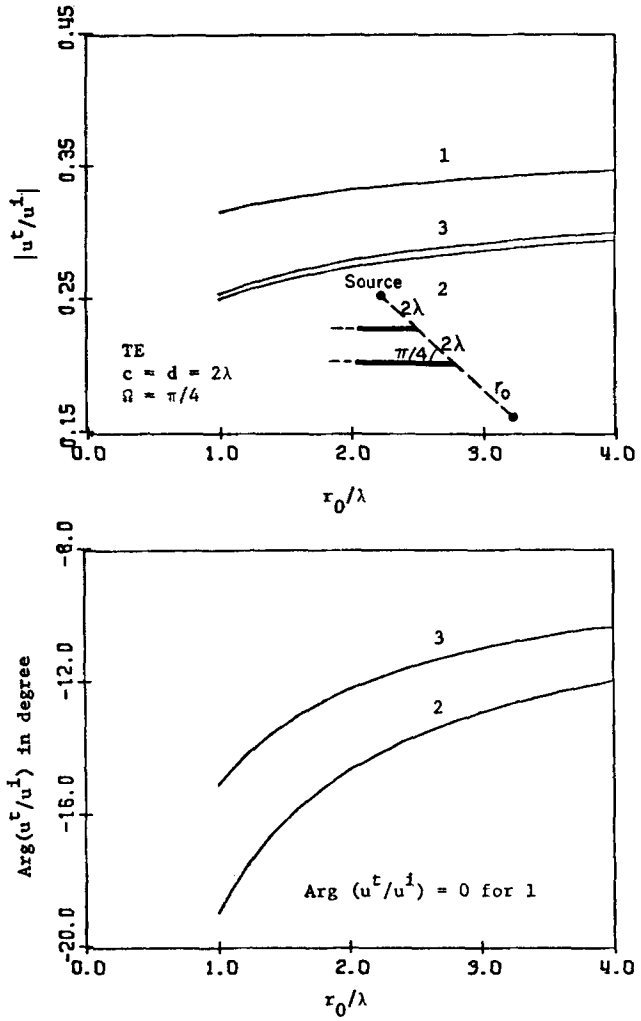


FIG. 9. Normalized total field on the incident shadow boundary of two staggered parallel plates (Fig. 6) for TE case. u^t is calculated from Eq. (IV. 27) with one, two, and three terms (indicated by 1, 2, and 3, respectively, next to the curves); $u^i = \exp[ikr_2 + i\pi/4]/2\sqrt{2\pi kr_2}$, where $r_2 = c + d + r_0$.

$$\bar{R}_{\rm int}=\frac{i\exp(ikr_0)}{8\pi\sqrt{r_0+d}}\left[\left(\frac{2}{b}\right)^{1/2}\frac{\sin\Omega}{\cos^2\Omega}\sum_{n=1}^{\infty}\frac{\exp(i2nkb)}{n^{3/2}}-\exp(i2kd)\times\frac{\tan^2\Omega}{4\sqrt{d}}\right]. \tag{IV.30e}$$

The final result in (IV.30) may be compared with the rigorous solution derived by Jones in Ref. 8. Jones' solution is a uniform asymptotic expansion for the total field $u^t(r_0,\phi_0)$ in the forward region $\pi\leq\phi_0\leq 2\pi$; this expansion contains terms up to $O(k^{-1})$ [term of $O(k^{-3/2})$ is not included]. In Ref. 17, Jones' solution is specialized to the case $\phi_0=\pi+\Omega$ (a step that involves some tedious limit computations). It is found that the result agrees precisely with (IV.30), except that $\bar{R}_{\rm int}$ in (IV.30e) does not appear in Jones' solution. The term $\bar{R}_{\rm int}$ describes the interactions between the parallel plates: (i) the first term in (IV.30e) accounts for the multiple reflections and diffractions along the shadow boundary $x=-l$, $0\leq y\leq b$ (Fig. 6), and (ii) the second term results from the interaction between the edges along their connecting line. Jones did not consider these two types of interactions which explains the absence of $\bar{R}_{\rm int}$ in his solution.

It is also interesting to observe that the interaction contribution described by $\bar{R}_{\rm int}$ is of order $O(k^{-1})$ provided that $\Omega\neq\pi/2$ and Ω is not close to $\pi/2$. When $\Omega=\pi/2$ precisely, the interaction contribution is of order $O(k^0)$, as can be seen by comparison of the results in Sec. II and that of Jones. Thus, as $\Omega\to\pi/2$, the interaction contribution increases from $O(k^{-1})$ to $O(k^0)$. It would be desirable to derive a result which is uniform in Ω; however, such a derivation seems beyond our means for the moment.

APPENDIX A: EXACT SOLUTION OF DIFFRACTION BY TWO NONSTAGGERED PARALLEL PLATES

For diffraction of an incident plane wave by two nonstaggered parallel plates, the problem has been solved exactly by the Wiener–Hopf technique (see Ref. 12, Sec 3.2, or Ref. 13, Sec. 3-12). The solution is given in terms of an inverse Fourier transform, and it is exact. We have evaluated the inverse Fourier transform and obtained the far field solution. When the incident field is given by (II.1) (normal incidence), the total field u^t in the upper half-space ($y\geq a$), far away from the edges, is found to be (Fig. 1)

$$u^t(r_1,\phi_1)\sim u_u^t(r_1,\phi_1)+\exp(-ika)\frac{\exp[i(kr_1+\pi/4)]}{2\sqrt{2\pi kr_1}}D(\phi_1,3\pi/2)\times\left(\frac{G_+(k\cos\phi_1)G_+(k\cos 3\pi/2)+\tilde{G}_+(k\cos\phi_1)\tilde{G}_+(k\cos 3\pi/2)}{2}-1\right),\qquad kr_1\to\infty,\ \pi\leq\phi_1\leq 2\pi, \tag{A1a}$$

where u_u^t is the (exact) total field due to the upper plate when the lower plate is removed, and is given by

$$u_u^t=\exp[ik(r_1-a)]\ \{F[\sqrt{2kr_1}\cos\tfrac{1}{2}(\phi_1-3\pi/2)]+\tau F[\sqrt{2kr_1}\cos\tfrac{1}{2}(\phi_1+3\pi/2)]. \tag{A1b}$$

Here the Fresnel integral F is defined in (II.7), Keller's diffraction coefficient D in (III.5), and $G_+(\alpha)$ and $\tilde{G}_+(\alpha)$ in Appendix B. The solution in (A1) is valid uniformly for all ϕ_1, between π and 2π, and for an arbitrary ka.

In deriving (A1) from the said inverse Fourier transform, we have used the following procedure: First, the integrand of the inverse Fourier transform is decomposed into a term which exhibits a pole singularity and a second term which has no such a pole singularity but contains a branch singularity. Evaluation of the first constituent yields the field u_u^t in (A1b). Saddle point integration of the second constituent yields the remainder of (A1a).

We are interested particularly in the total field exactly on the reflected shadow boundary $\phi_1=3\pi/2$. Setting $\phi_1=(3\pi/2)-\delta$, where $\delta\to 0$ and making use of the relations

$$D(\phi_1,3\pi/2)=(2\tau/\delta)[1+O(\delta)],$$
$$G_+(k\cos\phi_1)=G_+(0)-\delta kG_+'(0)+O(\delta^2),$$
$$G_+(0)=[G(0)]^{1/2},$$

we obtain

$$u^t(r_1,\phi_1=3\pi/2)\sim\exp[-ik(r_1+a)]+\tfrac{1}{2}\tau\exp[ik(r_1-a)]-\frac{\exp[ik(r_1-a)+i(\pi/4)]}{2\sqrt{2\pi kr_1}}\left(1+\tau[1-\exp(ika)]\frac{kG_+'(0)}{G_+(0)}+\tau[1+\exp(ika)]\frac{k\tilde{G}_+'(0)}{\tilde{G}_+(0)}\right),\qquad kr_1\to\infty. \tag{A2}$$

For large ka, formulas (B20) and (B21), Appendix B, may be used in (A2); then u^t becomes

$$u^t\left(r_1,\phi_1=\frac{3\pi}{2}\right)\sim\exp[-ik(r_1+a)]+\exp[ik(r_1-a)]\times\left[\frac{\tau}{2}+\frac{\tau}{2\pi}\left(\frac{a}{r_1}\right)^{1/2}\sum_{n=1}^{\infty}\left(\frac{1}{\sqrt{2n-1}}-\frac{1}{\sqrt{2n}}\right)\times\exp(i2nka)-\frac{\exp(i\pi/4)}{2\sqrt{2\pi kr_1}}\right],\qquad kr_1\to\infty\ \text{and}\ \ ka\to\infty. \tag{A3}$$

Corresponding to (A1), the solution of u^t in the lower half-space ($y\leq 0$), far away from the edges, is found to be (Fig. 1)

$$u^t(r_0,\phi_0)\sim u_l^t(r_0,\phi_0)-\exp(-ika)\frac{\exp[i(kr_0+\pi/4)]}{2\sqrt{2\pi kr_0}}D(\phi_0,\pi/2)\times\left(\frac{G_+(k\cos\phi_0)G_+(k\cos\pi/2)-\tilde{G}_+(k\cos\phi_0)\tilde{G}_+(k\cos\pi/2)}{2}+\exp(ika)\right),\qquad kr_0\to\infty,\ \pi\leq\phi_0\leq 2\pi, \tag{A4a}$$

where u_l^t is the (exact) total field due to the lower plate when the upper plate is removed, and is given by

$$u_l^t=\exp(ikr_0)\{F[\sqrt{2kr_0}\cos\tfrac{1}{2}(\phi_0-\pi/2)]+\tau F[\sqrt{2kr_0}\cos\tfrac{1}{2}(\phi_0+\pi/2)]\}. \tag{A4b}$$

Exactly on the incident shadow boundary $\phi_0=3\pi/2$, we obtain the total field u^t from (A4), namely,

$$u^t\left(r_0,\phi_0=\frac{3\pi}{2}\right)\sim\tfrac{1}{2}\exp(ikr_0)+\frac{\exp[i(kr_0+\pi/4)]}{2\sqrt{2\pi kr_0}}\times\left(\tau+[1-\exp(-ika)]\frac{kG_+'(0)}{G_+(0)}+[1+\exp(-ika)]\times\frac{k\tilde{G}_+'(0)}{\tilde{G}_+(0)}\right),\qquad kr_0\to\infty. \tag{A5}$$

For large ka, the use of (B20) and (B21) in (A5) leads to

$$u^t\left(r_0, \phi_0 = \frac{3\pi}{2}\right) \sim \exp(ikr_0)\left[\frac{1}{2} - \frac{1}{2\pi}\left(\frac{a}{r_0}\right)^{1/2} + \frac{1}{2\pi}\left(\frac{a}{r_0}\right)^{1/2} \times \sum_{n=1}^{\infty}\left(\frac{1}{\sqrt{2n}} - \frac{1}{\sqrt{2n+1}}\right)\exp(i2nka) + \tau\frac{\exp(i\pi/4)}{2\sqrt{2\pi kr_0}}\right], \quad kr_0 \to \infty \text{ and } ka \to \infty. \tag{A6}$$

APPENDIX B: FUNCTIONS $G_+(\alpha)$ AND $\tilde{G}_+(\alpha)$

We are interested here in the factorization of two functions

$$G(\alpha) = 1 - \exp[-a(\alpha^2 - k^2)^{1/2}], \tag{B1}$$

$$\tilde{G}(\alpha) = 1 + \exp[-a(\alpha^2 - k^2)^{1/2}], \tag{B2}$$

in the manner

$$G(\alpha) = G_+(\alpha)G_+(-\alpha), \tag{B3}$$

$$\tilde{G}(\alpha) = \tilde{G}_+(\alpha)\tilde{G}_+(-\alpha), \tag{B4}$$

where $G_+(\alpha)$ and $\tilde{G}_+(\alpha)$ are regular in the upper half complex α-plane ($\mathrm{Im}\,\alpha > 0$) and have algebraic behavior at infinity. These two factorizations have been studied extensively in the literature. Here we simply list several useful final results.

(i) Infinite product forms[6, 7, 12, 13]

$$G_+(\alpha) = [2(1+\alpha/k)\sin(ka/2)]^{1/2}\exp(-i\pi/4) \times \exp[(i\alpha a/2\pi)(1 - C + \ln(4\pi/ka) + \tfrac{1}{2}\pi i)] \times \exp\{(ia\gamma/2\pi)\ln[(\alpha-\gamma)/k]\}\prod_{n=1}^{\infty}(1+\alpha/i\gamma_n) \times \exp(i\alpha a/2n\pi), \tag{B5}$$

$$\tilde{G}_+(\alpha) = [2\cos(ka/2)]^{1/2}\exp[(i\alpha a/2\pi)(1 - C + \ln(\pi/ka) + \tfrac{1}{2}\pi i)] \times \exp\{(ia\gamma/2\pi)\ln[(\alpha-\gamma)/k]\}\prod_{n=1}^{\infty}(1+\alpha/i\gamma_{n-1/2}) \times \exp[i\alpha a/(2n-1)\pi], \tag{B6}$$

where $C = 0.57721\cdots$ is Euler's constant, $\gamma = (\alpha^2 - k^2)^{1/2} = -i(k^2 - \alpha^2)^{1/2}$, and $\gamma_m = [(2m\pi/a)^2 - k^2]^{1/2} = -i[k^2 - (2m\pi/a)^2]^{1/2}$ with $m = n$ or $n - \frac{1}{2}$.

(ii) Relation to Weinstein's functions[18, 19]:

$$G_+(\alpha) = \exp[U(\alpha/k, ka)], \tag{B7}$$

$$\tilde{G}_+(\alpha) = \exp[\tilde{U}(\alpha/k, ka)]. \tag{B8}$$

Here $U(s, ka)$ denotes the exact Weinstein function

$$U(s, ka) = \frac{1}{2\pi i}\int_{-\infty}^{\infty}\ln[1 - \exp(ika - kat^2)] \times \frac{(1+it^2)(1+\frac{1}{2}it^2)^{-1/2}}{t(1+\frac{1}{2}it^2)^{1/2} - 2^{-1/2}s\exp(i\pi/4)}\,dt, \tag{B9}$$

and $\tilde{U}(s, ka)$ is again given by (B9) after replacing $\ln[1 - \exp(ika - kat^2)]$ by $\ln[1 + \exp(ika - kat^2)]$. In Ref. 18, an approximation for $U(s, ka)$ was introduced,

$$U_{\mathrm{app}}(s, ka) = \frac{1}{2\pi i}\int_{-\infty}^{\infty}\ln[1 - \exp(ika - kat^2)] \times \frac{1}{t - 2^{-1/2}s\exp(i\pi/4)}\,dt, \tag{B10}$$

and various properties of $U_{\mathrm{app}}(s, ka)$ were discussed. It has been shown[19] that the error of this approximation is $O(k^{-1}a^{-1})$, uniformly in s.

(iii) Asymptotic expansion for large ka when α/k is not close to zero[18, 19]:

$$G_+(\alpha) = 1 - \frac{\exp(i\pi/4)}{(2\pi ka)^{1/2}}\frac{k}{\alpha}\sum_{n=1}^{\infty}\frac{\exp(inka)}{n^{3/2}} + O(k^{-1}a^{-1}), \tag{B11}$$

$$\tilde{G}_+(\alpha) = 1 - \frac{\exp(i\pi/4)}{(2\pi ka)^{1/2}}\frac{k}{\alpha}\sum_{n=1}^{\infty}\frac{(-1)^n\exp(inka)}{n^{3/2}} + O(k^{-1}a^{-1}), \tag{B12}$$

which can be derived starting from either (B9) or (B10).

(iv) Logarithmic derivative at $\alpha = 0$:

$$\frac{kG_+'(0)}{G_+(0)} = \frac{1}{2} + \frac{ika}{2\pi}\left[-C + \ln\left(\frac{4\pi}{ka}\right) + i\frac{\pi}{2}\right] + \frac{ika}{2\pi}\sum_{n=1}^{\infty}\left(\frac{1}{n} - \frac{1}{\sqrt{n^2 - (ka/2\pi)^2}}\right), \tag{B13}$$

$$\frac{k\tilde{G}_+'(0)}{\tilde{G}_+(0)} = \frac{ika}{2\pi}\left[-C + \ln\left(\frac{\pi}{ka}\right) + i\frac{\pi}{2}\right] + \frac{ika}{2\pi}\sum_{n=1}^{\infty}\left(\frac{1}{(n-\frac{1}{2})} - \frac{1}{[(n-\frac{1}{2})^2 - (ka/2\pi)^2]^{1/2}}\right), \tag{B14}$$

where $[m^2 - (ka/2\pi)^2]^{1/2} = -i[(ka/2\pi)^2 - m^2]^{1/2}$ with $m = n$ or $n - \frac{1}{2}$. The derivations of (B13) and (B14) follow from (B5) and (B6), respectively. An alternative representation of the derivatives reads

$$\frac{kG_+'(0)}{G_+(0)} = \tfrac{1}{2}ka\sum_{n=1}^{\infty}H_0^{(1)}(nka), \tag{B15}$$

$$\frac{k\tilde{G}_+'(0)}{\tilde{G}_+(0)} = \tfrac{1}{2}ka\sum_{n=1}^{\infty}(-1)^nH_0^{(1)}(nka). \tag{B16}$$

To establish (B15), let us start with the logarithmic differentiation of (B7), viz.,

$$\frac{kG_+'(0)}{G_+(0)} = \frac{\partial}{\partial s}U(s, ka)\Big|_{s=0} = \frac{2^{-1/2}\exp(i\pi/4)}{2\pi i}\int_{-\infty}^{\infty}\ln[1 - \exp(ika - kat^2)] \times \frac{(1+it^2)(1+\frac{1}{2}it^2)^{-1/2}}{[t(1+\frac{1}{2}it^2)^{1/2} - 2^{-1/2}s\exp(i\pi/4)]^2}\,dt\Big|_{s=0}.$$

Integrating by parts, we find

$$\frac{kG_+'(0)}{G_+(0)} = \left(\frac{2^{-1/2}\exp(i\pi/4)}{2\pi i}\ln[1 - \exp(ika - kat^2)] \times \frac{(-1)}{t(1+\frac{1}{2}it^2)^{1/2} - 2^{-1/2}s\exp(i\pi/4)}\Big|_{t=-\infty}^{\infty} + \frac{2^{-1/2}\exp(i\pi/4)}{2\pi i}\int_{-\infty}^{\infty}\frac{\exp(ika - kat^2)}{1 - \exp(ika - kat^2)} \times \frac{2kat}{t(1+\frac{1}{2}it^2)^{1/2} - 2^{-1/2}s\exp(i\pi/4)}\,dt\right)_{s=0} = \frac{\exp(-i\pi/4)ka}{\sqrt{2}\pi}\int_{-\infty}^{\infty}\frac{\exp(ika - kat^2)}{1 - \exp(ika - kat^2)} \times \frac{dt}{(1+\frac{1}{2}it^2)^{1/2}}. \tag{B17}$$

The integrand can be expanded in a geometric series yielding

$$\frac{kG_+'(0)}{G_+(0)} \simeq \frac{\exp(-i\pi/4)ka}{\sqrt{2\pi}} \sum_{n=1}^{\infty} \exp(inka) \int_{-\infty}^{\infty} \frac{\exp(-nkat^2)}{(1+\frac{1}{2}it^2)^{1/2}}\, dt. \tag{B18}$$

By applying the substitution $t^2 = 2s$, the integral in (B18) becomes

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{\exp(-nkat^2)}{(1+\frac{1}{2}it^2)^{1/2}}\, dt &= \exp(-i\pi/4)\sqrt{2} \int_0^{\infty} \frac{\exp(-2nkas)}{(s^2-is)^{1/2}}\, ds \\ &= \frac{\pi}{\sqrt{2}} \exp(i\pi/4) \exp(-inka) H_0^{(1)}(nka), \end{aligned} \tag{B19}$$

according to Equation 4.3 (16) in Ref. 20. On substitution of (B19) into (B18) we obtain (B15). A similar proof holds for (B16).

(v) Logarithmic derivative at $\alpha = 0$ for large ka:

$$\frac{kG_+'(0)}{G_+(0)} = \frac{\exp(-i\pi/4)}{\sqrt{2\pi}} (ka)^{1/2} \sum_{n=1}^{\infty} \frac{\exp(inka)}{n^{1/2}} + O(k^{-1/2}a^{-1/2}), \tag{B20}$$

$$\frac{k\tilde{G}_+'(0)}{\tilde{G}_+(0)} = \frac{\exp(-i\pi/4)}{\sqrt{2\pi}} (ka)^{1/2} \sum_{n=1}^{\infty} \frac{(-1)^n \exp(inka)}{n^{1/2}} + O(k^{-1/2}a^{-1/2}) \tag{B21}$$

which are obtained from (B15) and (B16) after replacing the Hankel functions by their asymptotic expansions.

*Supported by NSF Grant No. GK37839.

[1]R.M. Lewis and J. Boersma, J. Math. Phys. 10, 2291 (1969).
[2]P.H.M. Kersten, "Diffraction of an electromagnetic wave by a plane screen," Tech. Rept., Department of Mathematics, Tech. University of Eindhoven, Eindhoven, Netherlands (in Dutch), (1967).
[3]D.S. Ahluwalia, R.M. Lewis, and J. Boersma, SIAM J. Appl. Math. 16, 783 (1968).
[4]J. Boersma, "Ray-optical analysis of reflection in an open-ended parallel-plane waveguide, I. TM case," to appear in SIAM J. Appl. Math. 29, July 1975.
[5]J. Boersma, IEEE Proc. 62, 1475 (1974).
[6]S.W. Lee, J. Math. Phys. 11, 2830 (1970).
[7]S.W. Lee, J. Math. Phys. 13, 656 (1972).
[8]D.S. Jones, Quart. J. Mech. Appl. Math. 26, 1 (1973).
[9]J.B. Keller, J. Opt. Soc. Am. 52, 116 (1962).
[10]R.M. Lewis and J.B. Keller, "Asymptotic methods for partial differential equations: The reduced wave equation and Maxwell's equations," Res. Rep. EM-194, Courant Institute of Mathematics, New York University, New York, New York, (1964).
[11]H.Y. Yee, L.B. Felsen, and J.B. Keller, SIAM J. Appl. Math. 16, 268 (1968).
[12]B. Noble, *Methods Based on the Wiener-Hopf Technique* (Pergamon, London, 1958).
[13]R. Mittra and S.W. Lee, *Analytical Techniques in the Theory of Guided Waves* (Macmillan, New York, 1971).
[14]J. Boersma, "On certain multiple integrals occurring in a waveguide scattering problem," to appear.
[15]A. Erdélyi, W. Magnus, F. Oberhettinger, and F.G. Tricomi, *Higher Transcendental Functions, Vol. I* (McGraw-Hill, New York, 1953).
[16]E.O. Powell, Quart. J. Mech. Appl. Math. 5, 116 (1952).
[17]J. Boersma, "Diffraction by two parallel half-planes," to appear.
[18]L.A. Weinstein, *The Theory of Diffraction and the Factorization Method* (Golem, Boulder, 1969).
[19]J. Boersma, Philips Res. Repts. 30, 161* (1975).
[20]A. Erdélyi, W. Magnus, F. Oberhettinger, and F.G. Tricomi, *Tables of Integral Transforms, Vol. I* (McGraw-Hill, New York, 1954).

Gauge dependence of Green's functions in quantum electrodynamics from parallel translation

Kurt Meetz

Physikalisches Institut der Universität Bonn, Germany
(Received 1 April 1975)

The classical concept of parallel translation is extended to scalar quantum electrodynamics in order to give a gauge-independent definition of differentiation. This is achieved by a suitable definition of time ordering for operator products. However, due to some essentially nonlocal commutation relations, the differential equations for the Green's functions are still gauge-dependent. The gauge dependence of the commutators can be removed by parallelism at large. Since this is not considered to be a physically reasonable concept, the gauge dependence of the Green's functions is discussed for general linear gauges in space-time.

1. INTRODUCTION

In classical field theory one is used to define the gauge-covariant derivative of charged fields by means of gauge fields or potentials. In scalar electrodynamics, e.g., the gauge-covariant derivative ∇_μ of the charged scalar field $\phi(x)$ is defined by

$$\nabla_\mu \phi := (\partial_\mu - ieA_\mu)\phi, \tag{1.1}$$

where A_μ is the electromagnetic potential and e is the coupling constant. This quantity is clearly covariant under the gauge transformation

$$\phi(x) \to \phi(x)\exp[ie\Lambda(x)], \tag{1.2a}$$

$$A_\mu(x) \to A_\mu(x) + \partial_\mu \Lambda(x) \tag{1.2b}$$

for any function $\Lambda(x)$. The coupling with the electromagnetic field strength $F_{\mu\nu}$ is provided by the relations

$$(\nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu)\phi = -ie(\partial_\mu A_\nu - \partial_\nu A_\mu)\phi =: -ieF_{\mu\nu}\phi, \tag{1.3}$$

from which we obtain the homogeneous Maxwell equations

$$\partial_\rho F_{\mu\nu} + \partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} = 0 \tag{1.4}$$

by a further derivation ∇_ρ and application of the Jacobi identity. We further have the field equations

$$(\nabla^\mu \nabla_\mu + m^2)\phi = 0, \tag{1.5a}$$

$$\partial^\mu F_{\mu\nu} = ie(\phi(\nabla_\nu\phi)^* - \phi^*\nabla_\nu\phi) =: j_\nu, \tag{1.5b}$$

where ϕ^* means the complex conjugate of ϕ and j_ν is the electromagnetic current.

This simple setup can also be considered from a more sophisticated point of view. Let us regard the value $\phi(x)$ of the complex scalar field at x as the component of a complex vector attached to x. By analogy with differential geometry this suggests to introduce the concept of parallel translation of vectors by the requirement that the Hermitian scalar product of any two vectors with components $\psi(x)$ and $\phi(x)$ is conserved under parallel translation along a curve $\xi(s)$ in space–time. The condition

$$\frac{d}{ds}[\psi^*(\xi(s))\phi(\xi(s))] = 0 \tag{1.6}$$

implies that $\psi(s)$ and $\phi(s)$ obey linear homogeneous differential equations along $\xi(s)$,

$$\frac{d\phi}{ds} - ieA_\mu \frac{d\xi^\mu}{ds}\phi = 0, \tag{1.7a}$$

$$\frac{d\psi^*}{ds} + ieA_\mu \frac{d\xi^\mu}{ds}\psi^* = 0, \tag{1.7b}$$

where the coefficients A_μ are components of an arbitrary but real Lorentz-vector field. The introduction of the potential is now motivated by the requirement to give a definition of physical equivalence for vectors at different points in space—time. The answer is that vectors, which can be mapped onto each other by parallel translation, are to be considered as physically equivalent. Along a curve $\xi(s)$ with $\xi(0) = x$ such a mapping is provided by the solution of (1.7a):

$$\phi(x) \to \phi(x)\exp\left(ie\int_0^s A_\mu(x+\xi)\,d\xi^\mu\right) =: \phi(x)\tau_s(x). \tag{1.8}$$

Parellelism at large is of physical relevance only, if it does not depend on the path choosen. This is not the case in the presence of an electromagnetic field if we relate the field strength $F_{\mu\nu}$ to the potential A_μ according to (1.3). Keeping the end point fixed, we obtain for a variation $\delta\xi(s)$ of the path from (1.8)

$$\delta_\xi \tau_s(x) = -ie\int_0^s F_{\mu\nu}(x+\xi)\,\delta\xi^\nu\,d\xi^\mu. \tag{1.9}$$

Locally, however, the notion of parallelism enables us to define the gauge-covariant derivative unambiguously in terms of the potential:

$$\lim_{s\to 0}\frac{\tau_s^{-1}(x)\phi(s) - \phi(x)}{s} =: \left(\frac{d\xi^\mu}{ds}\nabla_\mu\phi\right)_x = (\partial_\mu - ieA_\mu)\phi(x). \tag{1.10}$$

This is sufficient to give the field equations a gauge-independent meaning because they involve only local properties.

The quantum theory of gauge fields suffers from the fact that quantization as we know it is a gauge-dependent concept. Without adopting a particular gauge, we can only derive commutation relations for gauge-independent field variables from the gauge-invariant Lagrangian, e.g., by the method of Peierls.[1] To circumvent this difficulty, Mandelstam[2,3] has introduced gauge-independent field variables that depend on the path instead connecting the field point with infinity. In scalar electrodynamics the path-dependent scalar field $\Phi(x, P)$ is con-

structed from the scalar field $\phi(x)$ and the potential $A_\mu(x)$ in an arbitrary gauge according to

$$\Phi(x, P) := \phi(x) \exp(-ie\int_p A_\mu(\xi)\, d\xi^\mu), \tag{1.11}$$

or, as we would say, by parallel translation from x to infinity. Such an approach, reminiscent of the early attempts of Einstein[4] to introduce something like parallelism at distance, cannot be considered physically satisfactory as mentioned above. Instead of using parallelism at large to eliminate the gauge-dependence, we suggest transfering only the local implications of parallelism to the quantized theory. This requires a local definition of parallel translation for time-ordered products of field operators by means of potential operators. In Sec. 2 we shall derive the commutation relations for field variables and potentials insofar as they are needed for a local parallel translation of time-ordered products of scalar field variables and electromagnetic field strengths. By starting from the commutation relations for gauge-independent variables, this can be done quite generally without adopting a particular gauge. For calculational purposes it is, however, advisable to restrict ourselves to linear gauges. Parallel translation and gauge-covariant derivation of the Green's functions will be introduced in Sec. 3. In Sec. 4 we shall summarize the corresponding field equations using the condensed notation introduced by Mandelstam.[3] Their solution by perturbation theory yields the Green's functions in an arbitrary linear gauge. An extension of the method proposed in this paper to non-Abelian gauge theories is in preparation.

2. COMMUTATION RELATIONS

The classical Poisson brackets for gauge-independent field variables can be derived from the gauge-invariant Lagrangian

$$L = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + (\nabla_\mu\phi)^*\nabla^\mu\phi - m^2\phi^*\phi \tag{2.1}$$

by the method of Peierls[1] if we relate the gauge-covariant derivative and the fields strength to the potential according to (1.1) and (1.3). We first state the corresponding commutation relations between two field-strength operators,

$$[F_{0i}(\mathbf{x}, x_0), F_{0j}(\mathbf{y}, x_0)] = 0, \tag{2.2a}$$

$$[F_{ij}(\mathbf{x}, x_0), F_{kl}(\mathbf{y}, x_0)] = 0, \tag{2.2b}$$

$$[F_{0i}(\mathbf{x}, x_0), F_{jk}(\mathbf{y}, x_0)] = i\left(\delta_{ij}\frac{\partial}{\partial y^k} - \delta_{ik}\frac{\partial}{\partial y^j}\right)\delta(\mathbf{x}-\mathbf{y}). \tag{2.2c}$$

Here and in the following we use Latin super- and subscripts for the spatial components of Lorentz tensors. The only nonvanishing commutators between two scalar field operators are

$$[\phi(\mathbf{x}, x_0), (\nabla_0\phi(\mathbf{y}, x_0))^*] = i\delta(\mathbf{x}-\mathbf{y}), \tag{2.3a}$$

$$[\phi^*(\mathbf{x}, x_0), \nabla_0\phi(\mathbf{y}, x_0)] = i\delta(\mathbf{x}-\mathbf{y}). \tag{2.3b}$$

Next we consider the commutation relations between electromagnetic and scalar field operators. The commutators between the spatial components F_{jk} and scalar field variables vanish because these variables are dynamically independent. However, the commutators between the space–time components F_{0i} and scalar variables must not vanish if we require that a suitable version of the classical field equations holds in the quantized theory. The 0-component of (1.5b) is a condition of constraint,

$$\partial^i F_{i0} = j_0 = ie[\phi(\nabla_0\phi)^* - \phi^*\nabla_0\phi]. \tag{2.4}$$

(We should, of course, use normal products of operators, but we shall not write them). The commutation relations (2.3a)–(2.3b) imply, e.g.,

$$[j_0(\mathbf{x}, x_0), \phi(\mathbf{y}, x_0)] = e\delta(\mathbf{x}-\mathbf{y})\phi(\mathbf{y}, x_0). \tag{2.5}$$

This requires

$$[F_{0i}(\mathbf{x}, x_0), \phi(\mathbf{y}, x_0)] = e\Gamma_i(\mathbf{x}, \mathbf{y})\phi(\mathbf{y}, x_0) \tag{2.6a}$$

with

$$\partial^i\Gamma_i(\mathbf{x}, \mathbf{y}) = -\delta(\mathbf{x}-\mathbf{y}).$$

The method of Peierls tells us that the commutator of F_{0i} and $\phi^*\phi$ vanishes. Hence,

$$[F_{0i}(\mathbf{x}, x_0), \phi^*(\mathbf{y}, x_0)] = -e\phi^*(\mathbf{y}, x_0)\,\Gamma_i(\mathbf{x}, \mathbf{y}). \tag{2.6b}$$

The same argument leads to analogous commutation relations between F_{0i} and $\nabla_0\phi$ or $(\nabla_0\phi)^*$. In general the quantity $\Gamma_i(\mathbf{x}, \mathbf{y})$ is an operator depending on the field variables. A comparison of Eqs. (2.6a) and (2.6b) shows that it must be an Hermitian operator. Furthermore, the operator has to obey the conditions imposed by the Jacobi identities.

To derive the commutation relations between F_{0i} and the spatial covariant derivatives $\nabla_j\phi$ of the scalar field, we first determine the commutator of F_{0i} and the gauge-invariant quantity $\phi^*\nabla_j\phi$ by the method of Peierls. The result is

$$[F_{0i}(\mathbf{x}, x_0), \phi^*(\mathbf{y}, x_0)\nabla_j\phi(\mathbf{y}, x_0)] = -e\delta_{ij}\delta(\mathbf{x}-\mathbf{y})\phi^*(\mathbf{y}, x_0)\phi(\mathbf{y}, x_0). \tag{2.7}$$

We now observe (2.6b) and obtain

$$[F_{0i}(\mathbf{x}, x_0), \nabla_j\phi(\mathbf{y}, x_0)] = -e\delta_{ij}\delta(\mathbf{x}-\mathbf{y}) + e\Gamma_i(\mathbf{x}, \mathbf{y})\nabla_j\phi(\mathbf{y}, x_0). \tag{2.8}$$

Finally we express the covariant derivation in terms of the potential,

$$\nabla_j\phi = (\partial_j - ieA_j)\phi$$

and conclude from (2.8) and (2.6a)

$$[F_{0i}(\mathbf{x}, x_0), A_j(\mathbf{y}, x_0)] = -i\left(\delta_{ij}\delta(\mathbf{x}-\mathbf{y}) + \frac{\partial}{\partial y^j}\Gamma_i(\mathbf{x}, \mathbf{y})\right). \tag{2.9}$$

These commutation relations are consistent with Eq. (2.20) for

$$F_{jk} = \partial_j A_k - \partial_k A_j$$

and with the condition of constraint (2.4).

If Γ_i is a c-number, i.e., a real function, we see from (2.9) and (2.7) that the operator

$$\int d\tau(\mathbf{y}')A^j(\mathbf{y}', y_0)\Gamma_j(\mathbf{y}', \mathbf{y})$$

commutes with all the field variables. Hence we must have

$$\int d\tau(\mathbf{y}')A^j(\mathbf{y}', y_0)\Gamma_j(\mathbf{y}', \mathbf{y}) = B(\mathbf{y}, y_0), \tag{2.10}$$

where B is also a c-number function. On the other hand, any linear gauge condition that is consistent with the commutation relations and does involve only the spatial components of the potential can be written in the form (2.10). A generalization to gauge conditions linear in the four components of the potential will be given in Sec. 4. Taking into account the field equation

$$\partial_0 A_i - \partial_i A_0 = F_{0i},$$

we obtain by means of (2.10)

$$A_0(\mathbf{y}, y_0) = -\int d\tau(\mathbf{y}')F_{0i}(\mathbf{y}', y_0)\Gamma^i(\mathbf{y}', \mathbf{y}) + \partial_0 B. \tag{2.11}$$

Together with (2.7) and (2.9), respectively (2.6a) and (2.6b), we compute the commutators

$$[A_0(\mathbf{x}, x_0), A_j(\mathbf{y}, x_0)] = -i\left(\Gamma_j(\mathbf{y}, \mathbf{x}) - \frac{\partial}{\partial y^j}\int d\tau(\mathbf{y}')\Gamma^i(\mathbf{y}', \mathbf{x})\Gamma_i(\mathbf{y}', \mathbf{y})\right), \tag{2.12a}$$

$$[A_0(\mathbf{x}, x_0), \phi(\mathbf{y}, x_0)] = -e\int d\tau(\mathbf{y}')\Gamma^i(\mathbf{y}', \mathbf{x})\Gamma_i(\mathbf{y}', \mathbf{y})\phi(\mathbf{y}, x_0), \tag{2.12b}$$

$$[A_0(\mathbf{x}, x_0), \phi^*(\mathbf{y}, x_0)] = e\int d\tau(\mathbf{y}')\Gamma^i(\mathbf{y}', \mathbf{x})\Gamma_i(\mathbf{y}', \mathbf{y})\phi^*(\mathbf{y}, x_0). \tag{2.12c}$$

We emphasize that the commutation relations (2.12a)–(2.12c) hold only for linear gauges which satisfy the condition (2.10). For the Coulomb gauge, e.g., we have

$$\Gamma_i(\mathbf{x}, \mathbf{y}) = -\frac{\partial}{\partial x^i}\frac{1}{4\pi|\mathbf{x}-\mathbf{y}|}. \tag{2.13}$$

Another example is given by what might be called a path gauge, defined by the condition that the integral over a spatial path $\xi(s)$ from the field point x to infinity should vanish. If $\xi(0) = 0$ and $\xi(s)$ tends to infinity for $s \to -\infty$, we may write

$$\int_{-\infty}^{0} ds \frac{d\xi^i}{ds} A_i(\mathbf{x} + \xi) = 0. \tag{2.14}$$

The function Γ_i is

$$\Gamma_i(\mathbf{x}, \mathbf{y}) = -\int_{-\infty}^{0} d\xi_i \delta(\mathbf{x} - \mathbf{y} + \xi). \tag{2.15}$$

A special case is the axial gauge with $\xi(s) = \mathbf{n}s$, $|\mathbf{n}| = 1$, which has been used by Arnowitt and Fickler[5] and Schwinger[6] in context with non-Abelian gauge fields.

3. GAUGE-COVARIANT DERIVATION OF TIME-ORDERED PRODUCTS

In this section we shall give a definition of time ordering that is compatible with local parallel translation and gauge-covariant derivation. We start with the time-ordered products of scalar field and field strength operators,

$$\tilde{T}(\phi(x_1)\phi(x_2)\cdots\phi^*(y_1)\phi^*(y_2)\cdots F_{\mu\nu}(z_1)F_{\rho\sigma}(z_2)\cdots).$$

The time-ordering $\tilde{T}$ is the usual one for products of scalar variables alone. We shall denote the latter by the symbol T, e.g.,

$$\tilde{T}(\phi(x)\phi(y)) := T(\phi(x)\phi(y)) = \theta(x_0 - y_0)\phi(x)\phi(y) + \theta(y_0 - x_0)\phi(y)\phi(x). \tag{3.1}$$

On the other side it is well known that we have to define

$$\tilde{T}(F_{\mu\nu}(z_1)F_{\rho\sigma}(z_2)) := T(F_{\mu\nu}(z_1)F_{\rho\sigma}(z_2)) + i(g_\mu^0 g_\nu^i - g_\nu^0 g_\mu^i)(g_\rho^0 g_\sigma^j - g_\sigma^0 g_\rho^j)\delta_{ij}\delta(z_1 - z_2), \tag{3.2}$$

in order to get Lorentz-covariant field equations. It is easily checked by means of the commutation relations (2.26) that the definition (3.2) satisfies

$$\partial^\mu \tilde{T}(F_{\mu\nu}(z_1)F_{\rho\sigma}(z_2)) = T(\partial^\mu F_{\mu\nu}(z_1)F_{\rho\sigma}(z_2)) + i(g_{\rho\nu}\partial_\sigma - g_{\sigma\nu}\partial_\rho)\delta(z_1 - z_2), \tag{3.3a}$$

$$\partial_{[\lambda}\tilde{T}(F_{\mu\nu]}(z_1)F_{\rho\sigma}(z_2)) = T(\partial_{[\lambda}F_{\mu\nu]}(z_1)F_{\rho\sigma}(z_2)), \tag{3.3b}$$

where the symbol $[\lambda\mu\nu]$ indicates cyclical permutation. The time ordering $\tilde{T}$ shall, therefore, include contributions in accordance with (3.2) for any pairing of field strength operators.

We next turn to the definition of $\tilde{T}$ ordering for the product of a potential and a field strength operator. Observing the commutation relations (2.9), we can compute

$$\partial'_\rho T(F_{\mu\nu}(z)A_\sigma(z')) - \partial'_\sigma T(F_{\mu\nu}(z)A_\rho(z')) = T(F_{\mu\nu}(z)F_{\rho\sigma}(z')) + i(g_\mu^0 g_\nu^i - g_\nu^0 g_\mu^i)(g_\rho^0 g_\sigma^j - g_\sigma^0 g_\rho^j) \times\left(\delta_{ij}\delta(z - z') + \frac{\partial}{\partial z^j}\Gamma_i(\mathbf{z}, \mathbf{z}')\delta(z_0 - z_0')\right). \tag{3.4}$$

Hence the definition

$$\tilde{T}(F_{\mu\nu}(z)A_\sigma(z')) := T(F_{\mu\nu}(z)A_\sigma(z')) + i(g_\mu^0 g_\nu^i - g_\nu^0 g_\mu^i)g_\sigma^0 \times\Gamma_i(\mathbf{z}, \mathbf{z}')\delta(z_0 - z_0') \tag{3.5}$$

secures that

$$\partial'_\rho\tilde{T}(F_{\mu\nu}(z)A_\sigma(z')) - \partial'_\sigma\tilde{T}(F_{\mu\nu}(z)A_\sigma(z')) = \tilde{T}(F_{\mu\nu}(z)F_{\rho\sigma}(z')). \tag{3.6}$$

This is sufficient to extend the $\tilde{T}$ ordering to products of field operators containing one potential operator, in the sense that in addition to extra terms of type (3.2) we have to write an extra term of type (3.5) for each pairing of the potential with a field strength operator. We now define the gauge-covariant derivative of the $\tilde{T}$-ordered product by

$$\tilde{\nabla}_\mu\tilde{T}(\phi(x)\cdots F_{\rho\sigma}(z)\cdots) := \partial_\mu\tilde{T}(\phi(x)\cdots F_{\rho\sigma}(z)\cdots) - ie\tilde{T}(A_\mu(x)\phi(x)\cdots F_{\rho\sigma}(z)\cdots). \tag{3.7}$$

Performing the differentiation, we see that the contributions according to the commutation relation (2.6a) cancel against the extra terms according to (3.5). Hence we obtain

$$\tilde{\nabla}_\mu\tilde{T}(\phi(x)\cdots F_{\rho\sigma}(z)\cdots) = \tilde{T}((\partial_\mu - ie\hat{A}_\mu(x))\phi(x)\cdots F_{\rho\sigma}(z)\cdots) = \tilde{T}(\hat{\nabla}_\mu\phi(x)\cdots F_{\rho\sigma}(z)\cdots), \tag{3.8}$$

where the symbol $\hat{A}_\mu$ indicates that the extra terms according to (3.5) have to be dropped.

Let us now consider second derivatives. There is no problem if the first derivation is timelike and the second is spacelike. In accordance with (3.8) we define

$$\tilde{\nabla}_i\tilde{\nabla}_0\tilde{T}(\phi(x)\cdots F_{\rho\sigma}(z)\cdots) := \tilde{T}(\nabla_i\hat{\nabla}_0\phi(x)\cdots F_{\rho\sigma}(z)\cdots). \tag{3.9}$$

For the reversed order we refer to the definition (3.7) and apply the commutation relations (2.8),

$$\begin{aligned}\tilde{\nabla}_0\nabla_i\tilde{T}(\phi(x)\cdots F_{\rho\sigma}(z)\cdots)&\\ =\partial_0\tilde{T}(\nabla_i\phi(x)\cdots F_{\rho\sigma}(z)\cdots)&\\ -ie\tilde{T}(A_0(x)\nabla_i\phi(x)\cdots F_{\rho\sigma}(z)\cdots)&\\ =-e(g^0_\rho g^j_\sigma - g^0_\sigma g^j_\rho)\delta_{ij}\delta(x-z)\tilde{T}(\cdots)&\\ +\tilde{T}(\hat{\nabla}_0\nabla_i\phi(x)\cdots F_{\rho\sigma}(z)\cdots)+\cdots.&\end{aligned} \tag{3.10}$$

In general we may write

$$\begin{aligned}(\tilde{\nabla}_\mu\tilde{\nabla}_\nu-\tilde{\nabla}_\nu\tilde{\nabla}_\mu)\tilde{T}(\phi(x)\cdots F_{\rho\sigma}(z)\cdots)&\\ =e(g^0_\mu g^i_\nu - g^0_\nu g^i_\mu)(g^0_\rho g^j_\sigma - g^0_\sigma g^j_\rho)\delta_{ij}\delta(x-z)\tilde{T}(\cdots)&\\ +\cdots+\tilde{T}((\hat{\nabla}_\mu\hat{\nabla}_\nu-\hat{\nabla}_\nu\hat{\nabla}_\mu)\phi(x)\cdots F_{\rho\sigma}(z)\cdots).&\end{aligned} \tag{3.11}$$

If we assume the field equation (1.3) also in quantum theory, the latter term should be

$$-ie\tilde{T}(\hat{F}_{\mu\nu}(x)\phi(x)\cdots F_{\rho\sigma}(z)\cdots).$$

Thus we see that the extra terms required in accordance with (3.2) to include the operator $F_{\mu\nu}(x)$ into $\tilde{T}$ ordering are just supplied by the former terms of (3.11). The field equation (1.3) also holds for $\tilde{T}$-ordered products without additional terms,

$$\begin{aligned}(\tilde{\nabla}_\mu\tilde{\nabla}_\nu-\tilde{\nabla}_\nu\tilde{\nabla}_\mu)\tilde{T}(\phi(x)\cdots F_{\rho\sigma}(z)\cdots)&\\ =-ie\tilde{T}(F_{\mu\nu}(x)\phi(x)\cdots F_{\rho\sigma}(z)\cdots).&\end{aligned} \tag{3.12}$$

The related homogeneous Maxwell equations (1.4) have already been shown to be true for $\tilde{T}$-ordered products [see (3.3b)].

It should be emphasized that the relations (3.12) have been obtained without making any assumption on the gauge defining operator Γ_i. Formally this is due to the fact that in order to derive (3.12) it is sufficient to define $\tilde{T}$-ordered products containing only one timelike component of the potential. If we want to go further and to transfer the field equations (1.5a) to quantum theory, we have to introduce $\tilde{T}$ ordering for products including at least two timelike components of the potential. To do so, we start from the relation

$$\begin{aligned}\partial_\mu T(A_\nu(x)A_\sigma(y))-\partial_\nu T(A_\mu(x)A_\sigma(y))&\\ =T(F_{\mu\nu}(x)A_\sigma(y))+(g^0_\mu g^i_\nu - g^0_\nu g^i_\mu)g^0_\sigma[A_i(x),A_0(y)]&\\ \times\delta(x_0-y_0).&\end{aligned} \tag{3.13}$$

The commutator involved can only be computed if we make assumptions on the gauge. We shall adopt the linear gauge condition (2.10) and apply the commutation relations (2.12a). In order to be in agreement with Eq. (3.5), we have to define

$$\begin{aligned}\tilde{T}(A_\nu(x)A_\sigma(y)) := T(A_\nu(x)A_\sigma(y))-ig^0_\nu g^0_\sigma\int d\tau(\mathbf{y}')&\\ \times\Gamma^i(\mathbf{y}',\mathbf{x})\Gamma_i(\mathbf{y}',\mathbf{y})\delta(x_0-y_0).&\end{aligned} \tag{3.14}$$

The definition of the second gauge-covariant derivative $\tilde{\nabla}_0\tilde{\nabla}_0$ is then straightforward. Observing (3.8), we are led to write

$$\begin{aligned}\tilde{\nabla}_0\tilde{\nabla}_0\tilde{T}(\phi(x)\cdots\phi^*(y)\cdots F_{\mu\nu}(z)\cdots):&\\ =\partial_0\tilde{T}(\hat{\nabla}_0\phi(x)\cdots\phi^*(y)\cdots F_{\mu\nu}(z)\cdots)&\\ -ie\tilde{T}(A_0(x)\hat{\nabla}_0\phi(x)\cdots\phi^*(y)\cdots F_{\mu\nu}(z)\cdots).&\end{aligned} \tag{3.15}$$

This can be evaluated by means of the commutation relations (2.6a) with ϕ replaced by $\nabla_0\phi$ and (2.3b). Again the equal time contribution from (2.6a) cancels against the additional term included in the second expression according to the definition (3.5) of $\tilde{T}$ ordering. We obtain

$$\begin{aligned}\tilde{\nabla}_0\tilde{\nabla}_0\tilde{T}(\phi(x)\cdots\phi^*(y)\cdots F_{\mu\nu}(z)\cdots)&\\ =-i\delta(x-y)\tilde{T}(\cdots F_{\mu\nu}(z)\cdots)+\cdots&\\ +\tilde{T}(\hat{\nabla}_0\hat{\nabla}_0\phi(x)\cdots\phi^*(y)\cdots F_{\mu\nu}(z)\cdots).&\end{aligned} \tag{3.16}$$

In the latter term both potential operators are excluded from $\tilde{T}$ ordering. The same relation holds for $\tilde{T}$-ordered products with an arbitrary number of potential operators if we extend $\tilde{T}$ ordering to products of more than two potential operators by the agreement to add extra terms in accordance with (3.14) for each pairing of potential operators. Bearing in mind that $\tilde{T}$ ordering is not affected by spacelike gauge-covariant derivation, we can now transfer the field equation (1.5a) to $\tilde{T}$-ordered products of scalar field operators and potential operators in a linear gauge:

$$\begin{aligned}\tilde{\nabla}_\mu\tilde{\nabla}^\mu\tilde{T}(\phi(x)\cdots\phi^*(y)\cdots A_\nu(z)\cdots)&\\ =-i\delta(x-y)\,\tilde{T}(\cdots A_\nu(z)\cdots)+\cdots&\\ -m^2\tilde{T}(\phi(x)\cdots\phi^*(y)\cdots A_\nu(z)\cdots).&\end{aligned} \tag{3.17}$$

Our final task is to extend the field equation (1.5b) to $\tilde{T}$-ordered products. Together with (2.9) and (2.11) we first derive from (3.5)

$$\begin{aligned}\partial^\mu\tilde{T}(F_{\mu\nu}(z)A_\sigma(z')) = T(\partial^\mu F_{\mu\nu}(z)A_\sigma(z'))&\\ +i\left(g_{\nu\sigma}\delta(z-z')-\frac{\partial}{\partial z^\sigma}g^i_\nu\Gamma_i(\mathbf{z},\mathbf{z}')\right.&\\ \left.\times\delta(z_0-z_0')\right).&\end{aligned} \tag{3.18}$$

For the general product we have also to take into account contributions according to (2.6a) and (2.6b):

$$\begin{aligned}\partial^\mu\tilde{T}(\phi(x)\cdots\phi^*(y)\cdots F_{\mu\nu}(z)A_\sigma(z')\cdots)&\\ =i\left(g_{\nu\sigma}\delta(z-z')-\frac{\partial}{\partial z^\sigma}g^i_\nu\Gamma_i(\mathbf{z},\mathbf{z}')\delta(z_0-z_0')\right)&\\ \times\tilde{T}(\phi(x)\cdots\phi^*(y)\cdots)+\cdots+eg^i_\nu\Gamma_i(\mathbf{z},\mathbf{x})\delta(z_0-x_0)&\\ \times\tilde{T}(\phi(x)\cdots\phi^*(y)\cdots A_\sigma(z')\cdots)+\cdots-eg^i_\nu\Gamma_i(\mathbf{z},\mathbf{y})&\\ \times\delta(z_0-y_0)\tilde{T}(\phi(x)\cdots\phi^*(y)\cdots A_\sigma(z')\cdots)+\cdots&\\ +\tilde{T}(\phi(x)\cdots\phi^*(y)\cdots\partial^\mu F_{\mu\nu}(z)A_\sigma(z')\cdots).&\end{aligned} \tag{3.19}$$

For later reference we remark that the gauge function appears only in the formation

$$\Gamma_\nu(x,y) := g^i_\nu\Gamma_i(\mathbf{x},\mathbf{y})\delta(x_0-y_0), \tag{3.20}$$

which is a particular solution of the differential equation

$$\partial_x^\mu \Gamma_\nu(x,y) = -\delta(x-y). \tag{3.21}$$

The field equation (1.5b) requires

$$\begin{aligned}\tilde{T}(\phi(x)\cdots\phi^*(y)\cdots\partial^\mu F_{\mu\nu}(z)A_\sigma(z')\cdots)\\ = ie\tilde{T}(\phi(x)\cdots\phi^*(y)(\phi(z)(\hat{\nabla}_\nu\phi(z))^* - \phi^*(z)\\ \times\nabla_\nu\phi(z))A_\sigma(z')\cdots),\end{aligned} \tag{3.22}$$

where the symbol $\hat{\nabla}_\nu$ is to be understood as in (3.8). In accordance with (3.8) we can also write the rhs of (3.22) as a gauge-covariant derivation of a $\tilde{T}$-ordered product

$$\begin{aligned}\tilde{T}(\phi(x)\cdots\phi^*(y)\cdots j_\nu(z')A_\sigma(z')\cdots\\ = ie(\hat{\nabla}_\nu^w - \tilde{\nabla}_\nu^z)\tilde{T}(\phi(x)\phi(z)\cdots\phi^*(y)\\ \times\phi^*(w)\cdots A_\sigma(z'))\big|_{w=z}.\end{aligned} \tag{3.23}$$

Finally we mention that we may also extend the gauge-independent definition (1.10) of the covariant derivative to $\tilde{T}$-ordered products by

$$\begin{aligned}\lim_{s\to 0}\frac{1}{s}\tilde{T}\left(\left(\exp\left(-ie\int_0^s A_\mu(x+\xi)\,d\xi^\mu\right)\phi(s)-\phi(x)\right)\right.\\ \left.\times\cdots\phi^*(y)\cdots A_\rho(z)\cdots\right) =: \frac{d\xi^\mu}{ds}\tilde{\nabla}_\mu\tilde{T}(\phi(x)\cdots\\ \times\phi^*(y)\cdots A_\rho(z)\cdots)\end{aligned} \tag{3.24}$$

because we have defined $\tilde{T}$ ordering for products with an arbitrary number of potential operators.

4. GREEN'S FUNCTIONS

In this section we wish to derive the field equations for the Green's functions following from the field equations for $\tilde{T}$-ordered products we have obtained in the last section. The Green's functions are defined by

$$\begin{aligned}G_{\mu\ldots}(x\cdots y\cdots z\cdots):\\ = \langle 0|\tilde{T}(\phi(x)\cdots\phi^*(y)\cdots A_\mu(z)\cdots)|0\rangle.\end{aligned} \tag{4.1}$$

The general structure of the differential equations for the Green's functions is best displayed in terms of Mandelstam's condensed notation,[3] which we summarize shortly.

Consider the set of functions

$$F_{\mu\ldots}(x_1\cdots y_1\cdots z_1\cdots),$$

depending arbitrarily on the space–time variables $(x_1\cdots y_1\cdots z_1\cdots)$. We may look upon these functions as linear forms over a vector space V, i.e., special bilinear forms over $V^*\times V$, where V^* is the dual space. This view is indicated by the notation

$$F_{\mu\ldots}(x_1\cdots y_1\cdots z_1\cdots) =: (e_{\mu\ldots}(x_1\cdots y_1\cdots z_1\cdots)|F). \tag{4.2}$$

The symbol (|) denotes the canonical bilinear form over $V^*\times V$ in the sense of linear algebra. F is any element of V while the quantities $e_{\mu\ldots}(x_1\cdots y_1\cdots z_1\cdots)$ have to be considered as basis elements of V^*. The Green's functions (4.1) are obtained for a particular vector $F=G\in V$.

Next we define linear operators $\tilde{\phi}(x)$, $\tilde{\phi}(y)$, and $\tilde{A}_\nu(z)$ that act on the basis elements of the dual space as follows:

$$e_{\mu\ldots}(x_1\cdots y_1\cdots z_1\cdots)\tilde{\phi}(x) := e_{\mu\ldots}(x_1\cdots x, y_1\cdots z_1\cdots), \tag{4.3a}$$

$$e_{\mu\ldots}(x_1\cdots y_1\cdots z_1\cdots)\tilde{\phi}(y) := e_{\mu\ldots}(x_1\cdots y_1\cdots y, z_1\cdots), \tag{4.3b}$$

$$e_{\mu\ldots}(x_1\cdots y_1\cdots z_1\cdots)\tilde{A}_\nu(z) := e_{\mu\ldots\nu}(x_1\cdots y_1\cdots z_1\cdots z). \tag{4.3c}$$

The operators are denoted by a tilde in order to distinguish them from Hilbert space operators. The action of the operators on elements of V is defined by transposition according to the rule

$$\begin{aligned}(e_{\mu\ldots}(x_1\cdots y_1\cdots z_1\cdots)\tilde{\phi}(x)|F)\\ =: (e_{\mu\ldots}(x_1\cdots y_1\cdots z_1\cdots)|\tilde{\phi}(x)F).\end{aligned} \tag{4.4}$$

A second set of operators $\eta(x)$, $\bar{\eta}(x)$, and $Z_\mu(z)$ is needed to produce the source terms of the field equations. We define them by means of the commutation relations

$$\begin{aligned}[\eta(x),\tilde{\phi}^*(y)] = i\delta(x-y),\quad [\eta(x),\tilde{\phi}(x')] = 0,\\ [\eta(x),\tilde{A}_\nu(z)] = 0,\end{aligned} \tag{4.5a}$$

$$\begin{aligned}[\bar{\eta}(y),\tilde{\phi}(x)] = i\delta(x-y),\quad [\bar{\eta}(y),\tilde{\phi}^*(y')] = 0,\\ [\bar{\eta}(x),\tilde{A}_\nu(z)] = 0,\end{aligned} \tag{4.5b}$$

$$[Z_\mu(z),\tilde{A}_\nu(z')] = -i\left(g_{\mu\nu}\delta(z-z') - \frac{\partial}{\partial z'^\nu}\Gamma_\mu(z,z')\right) \tag{4.5c}$$

$$\begin{aligned}[Z_\mu(z),\tilde{\phi}(x)] = -e\Gamma_\mu(z,x)\tilde{\phi}(x),\\ [Z_\mu(z),\tilde{\phi}^*(y)] = e\Gamma_\mu(z,y)\tilde{\phi}^*(y),\end{aligned} \tag{4.5d}$$

where Γ_ν is defined as in (3.20). Furthermore, we require

$$e_0\eta(x) = 0,\quad e_0\bar{\eta}(y) = 0,\quad e_0 Z_\mu(z) = 0. \tag{4.6}$$

e_0 is the basis element of the one-dimensional subspace of V,

$$(e_0|F)\in\mathbb{C}.$$

The action of the operators on elements of V^* is then completely determined and may be transferred to elements of V by transposition.

The field equations for the Green's functions can easily be formulated within this framework. Gauge-covariant derivation of the basis elements is defined by

$$\begin{aligned}\tilde{\nabla}_\nu e_\mu(x\cdots y_1\cdots z_1\cdots) := \partial_\nu e_\mu(x\cdots y_1\cdots z_1\cdots)\\ - iee_\mu(x\cdots y_1\cdots z_1)\tilde{A}_\nu(x),\end{aligned} \tag{4.7}$$

or, equivalently, by the operator equation

$$\tilde{\nabla}_\nu\tilde{\phi}(x) = [\partial_\nu - ie\tilde{A}_\nu(x)]\tilde{\phi}(x). \tag{4.8}$$

This is in agreement with the covariant derivation of Green's functions as determined by (3.7). The gauge-independent field equations derived from (3.12) are represented by the operator equation

$$(\tilde{\nabla}_\mu\tilde{\nabla}_\nu - \tilde{\nabla}_\nu\tilde{\nabla}_\mu)\tilde{\phi}(x) = -ie\tilde{F}_{\mu\nu}(x)\tilde{\phi}(x), \tag{4.9}$$

where

$$\tilde{F}_{\mu\nu}(z) = \partial_\mu \tilde{A}_\nu(z) - \partial_\nu \tilde{A}_\mu(z). \tag{4.10}$$

Consequently, the field equations corresponding to the homogeneous Maxwell equations (1.4) are also represented by operator equations on V^*.

The situation is different for the field equations that follow from (3.17) and (3.19) for the Green's functions. The latter give rise to the following vector equations for the Green's vector G:

$$[\tilde{\nabla}^\mu \tilde{\nabla}_\mu \tilde{\phi}(x) + m^2 \tilde{\phi}(x) - \eta(x)]\, G = 0, \tag{4.11a}$$

$$((\tilde{\nabla}^\mu \tilde{\nabla}_\mu \tilde{\phi}(y))^* + m^2 \tilde{\phi}^*(y) - \bar{\eta}(y))G = 0, \tag{4.11b}$$

$$[\partial^\mu \tilde{F}_{\mu\nu}(z) - \tilde{j}_\nu(z) - Z_\nu(z)]G = 0. \tag{4.11c}$$

Here $\tilde{\nabla}_\mu \tilde{\phi}$ and $\tilde{F}_{\mu\nu}$ are defined by (4.3) and (4.10) respectively. The operator $\tilde{j}_\nu$ is given by

$$\tilde{j}_\nu = ie[\tilde{\phi}(\tilde{\nabla}_\nu \tilde{\phi})^* - \tilde{\phi}^* \nabla_\nu \tilde{\phi}] \tag{4.12}$$

in accordance with (3.23). To obtain the differential equations for the Green's functions from (4.11), one has to multiply with the basis elements

$$e_\mu \ldots (x_1 \cdots y_1 \cdots z_1 \cdots) = e_0 \tilde{\phi}(x_1) \ldots \tilde{\phi}^*(y_1) \cdots \tilde{A}_\mu(z_1) \cdots, \tag{4.13}$$

where the multiplication is defined by the bilinear form $(\ |\ )$, and to shift the operators η, $\bar{\eta}$, and Z_ν, respectively, to the left by means of the commutation relations (4.5a)—(4.5d) until the conditions (4.6) apply.

We now deduce from the commutation relations (4.5a)—(4.5d) that the operator

$$\partial^\nu Z_\nu + ie(\bar{\eta}\tilde{\phi} - \eta\tilde{\phi}^*)$$

commutes with $\tilde{\phi}$, $\tilde{\phi}^*$, and $\tilde{A}_\mu$. Furthermore, we see from (4.6) that

$$e_0[\partial^\nu Z_\nu + ie(\bar{\eta}\tilde{\phi} - \eta\tilde{\phi}^*)] = 0. \tag{4.14}$$

These statements imply the operator equation

$$\partial^\nu Z_\nu + ie(\bar{\eta}\tilde{\phi} - \eta\tilde{\phi}) = 0 \tag{4.15}$$

and show that the Ward identity

$$(\partial^\nu \tilde{j}_\nu + \partial^\nu Z_\nu)G = 0 \tag{4.16}$$

holds in consequence of the field equations because it follows from (4.11a), (4.11b), and (4.12) that

$$[\partial^\nu \tilde{j}_\nu + ie(\eta\tilde{\phi}^* - \bar{\eta}\tilde{\phi})]G = 0. \tag{4.17}$$

Next we introduce the operator $\zeta_\mu(z)$ defined on V^* by the commutation relations

$$[\zeta_\mu(z), \tilde{A}_\nu(z')] = -ig_{\mu\nu}\delta(z - z'), \quad [\zeta_\mu(z), \tilde{\phi}(x)] = 0,$$
$$[\zeta_\mu(z), \tilde{\phi}^*(y)] = 0 \tag{4.18}$$

and the condition

$$e_0 \zeta_\mu(z) = 0. \tag{4.19}$$

The commutation relations (4.5c) and (4.5d) and the operator equation (4.15) show that Z_μ is to be expressed in terms of ζ_μ as follows:

$$Z_\mu(z) = \zeta_\mu(z) + \int d^4w \Gamma_\mu(z, w) \times \{\partial^\nu \zeta_\nu(w) + ie[\bar{\eta}(w)\tilde{\phi}(w) - \eta(w)\tilde{\phi}^*(w)]\}. \tag{4.20}$$

If we insert the expression (4.20) into the field equation (4.11c), we can use (4.17) and obtain with (4.10)

$$\{\Box \tilde{A}_\nu(z) - \partial_\nu[\partial^\mu \tilde{A}_\mu(z)]\}G = \left(\zeta_\nu(z) + \tilde{j}_\nu(z) - \int d^4w \frac{\partial \Gamma_\nu}{\partial w^\rho}(z, w)[\zeta^\rho(w) + \tilde{j}^\rho(w)]\right) G. \tag{4.21}$$

To rewrite (4.21) in integral form, we use the particular solution under Feynman boundary conditions,

$$\tilde{A}_\nu(z)G = \int d^4w D_F(z - w)\left(g_{\nu\rho}\delta(z - w) - \frac{\partial \Gamma_\nu}{\partial w^\rho}(z, w)\right) \times [\zeta^\rho(w) + \tilde{j}^\rho(w)]G, \tag{4.22}$$

where

$$\Box D_F(z - w) = \delta(z - w)$$

and add a gradient term, which can be determined from the gauge condition (2.10). Together with (3.20) the latter can be written as

$$\int d^4y' A^\nu(y')\Gamma_\nu(y', y)G = B(y')G. \tag{4.23}$$

Let us put $B = 0$ for the sake of simplicity. We then obtain

$$\tilde{A}_\nu(z)G = \int d^4z' D_{\nu\tau}(z, z' \mid \Gamma)[\zeta^\tau(z') + \tilde{j}^\tau(z')]G, \tag{4.24}$$

where

$$D_{\nu\tau}(z, z' \mid \Gamma) := \int d^4w \int d^4w' \left(g^\rho_\nu \delta(z - w) - \frac{\partial \Gamma^\rho}{\partial z^\nu}(z, w)\right) \times g_{\rho\sigma} D_F(w - w') \left(g^\sigma_\tau \delta(w' - z') - \frac{\partial \Gamma^\sigma}{\partial z'^\tau}(w', z')\right) \tag{4.25}$$

is the propagation kernel for the gauge defined by (4.23). The integration of the scalar field equations (4.11a) and (4.11b) is straightforward. Introducing the potential $\tilde{A}_\nu$ in accordance with (4.8), Eq. (4.11a), e.g., can be rewritten as

$$\tilde{\phi}(x)G = \int d^4x' \Delta_F(x - x')\{\eta(x') + ie\tilde{A}^\mu(x')\partial'_\mu \tilde{\phi}(x') + ie\partial'_\mu[\tilde{A}_\mu(x')\tilde{\phi}(x')] + e^2 \tilde{A}^\mu(x')\tilde{A}_\mu(x')\tilde{\phi}(x')\}G, \tag{4.26}$$

where $(\Box + m^2)\Delta_F(x - x') = \delta(x - x')$.

The solution of the integrated equations (4.24), (4.26) and the corresponding equation for $\tilde{\phi}^*(y)G$ produces all terms summarized by the usual Feynman rules with (4.25) as photon propagation kernel.

Hitherto we have assumed that the gauge function Γ_ν has the particular structure given by (3.20), due to the fact that we started from equal-time commutation relations. But having obtained the field equations for the Green's functions, there is no reason not to admit any solution of the differential equation (3.21). This is just the class of gauges considered by Zumino[7] by means of functional techniques. The equation (3.21) is Lorentz-invariant, but the solution generally is not. The only covariant solution

$$\Gamma_\nu(x,y)=\frac{\partial}{\partial y^\nu}D_F(x-y) \tag{4.27}$$

leads to the Landau gauge, which is, however, not determined uniquely. Noncovariant examples are given by a generalization of the spatial path gauge (2.15) to space–time,

$$\Gamma_\nu(x,y)=\int_{-\infty}^{0}d\xi_\nu\,\delta(x-y+\xi), \tag{4.28}$$

or the generalized Coulomb gauge.[7]

In conclusion we emphasize that it is not possible to eliminate the gauge function Γ_ν from the field equations for the Green's functions in quantum theory. This is in contrast to the classical field equations, where the concept of gauge-covariant derivation is sufficient to remove any reference to a particular gauge. We could have used field strength operators $\tilde{F}_{\mu\nu}$ instead of the potential operators $\tilde{A}_\mu$. But this device only eliminates the gauge function from the commutation relations (4.5c). We are still left with (4.5d), where the gauge function appears due to the quantum mechanical commutators (2.6a) and (2.6b). These commutators are essentially nonlocal. Hence, it is not possible to remove their gauge-dependence by local concepts such as gauge-covariant derivation. This can only be achieved by parallelism at large as in Mandelstam's treatment, which in our opinion is not a physically reasonable concept.

[1]R.E. Peierls, Proc. Roy. Soc. (London) 213, 143 (1952).
[2]S. Mandelstam, Ann. Phys. (N.Y.) 19, 1 (1962).
[3]S. Mandelstam, Phys. Rev. 175, 1580 (1968).
[4]A. Einstein, Sitz. Preuss. Akad. Wiss., Phys.-Math. Kl., 217 (1928); 2 (1929).
[5]R.L. Arnowitt and S.I. Fickler, Phys. Rev. 127, 1821 (1962).
[6]J. Schwinger, Phys. Rev. 130, 402 (1963).
[7]B. Zumino, J. Math. Phys. 1, 1 (1960).

Statistical theory of effective electrical, thermal, and magnetic properties of random heterogeneous materials. VI. Comment on the notion of a cell material

Motoo Hori

Department of Applied Physics, Tokyo Institute of Technology, Meguroku, Tokyo 152, Japan
(Received 21 February 1975; revised manuscript received 26 March 1975)

The concept of a cell material introduced by Miller is reinvestigated in connection with Brown's assertion that an asymmetric cell material is not self-consistent [J. Math. Phys. **15**, 1516 (1974)]. We construct a simple example of the asymmetric cell material which is in fact self-consistent. The misleading interpretation of the asymmetric cell material is due to Miller rather than to Brown.

In previous papers[1-5] we dealt with the effective permittivity of random heterogeneous media such as cell materials or completely random materials. The concept of a cell material was first introduced by Miller,[6,7] who defined symmetric and asymmetric cell materials. A completely random material may be regarded as a limiting case of a symmetric cell material.[3] Recently, however, Brown[8] has claimed that while the symmetric cell material is self-consistent, the asymmetric cell materials or complete random materials. The con- - tion would require some modification. The aim of this note is to check the validity of Brown's arguments and to decide definitely whether or not the asymmetric cell material is inconsistent.

According to Miller,[6,7] a cell material is defined as a random multiphase material that fulfills the following requirements:

(i) The space is completely covered by nonoverlapping cells within which the material property is constant;

(ii) cells are distributed in a manner such that the material is statistically homogeneous and isotropic;

(iii) the material property of a cell is statistically independent of that of any other cell.

Of course, the assumption of statistical isotropy in (ii) should be omitted when we treat statistically anisotropic materials.[1] Furthermore, cell materials are classified as symmetric or asymmetric, according as the following additional requirement is satisfied or not:

(iv) The conditional probabilities of n points being and n' points not being in the same cell of a particular material, given that one point is in a cell of that material, are the same for each material.

Henceforth we shall confine ourselves to a cell material composed of two phases A and B. By way of illustration consider two-point probabilities associated with points $\mathbf{r}_1$ and $\mathbf{r}_2$. Then the independence hypothesis (iii) leads to:

(iii′) The event of a cell containing material A or B is statistically independent of the event of another cell containing A or B.

As a substitute for (iii′), let us tentatively adopt a plausible assumption:

(v) Under the condition that points $\mathbf{r}_1$ and $\mathbf{r}_2$ are in different cells, the event of point $\mathbf{r}_1$ being in phase A or B is statistically independent of the event of point $\mathbf{r}_2$ being in A or B.

For phase A, this statement is expressed by

$$\begin{aligned}&\mathrm{Prob}\{\mathbf{r}_1\in A\cap \mathbf{r}_2\in A\,|\,\mathbf{r}_1,\mathbf{r}_2\notin \mathrm{S.C.}\}\\&\quad=\mathrm{Prob}\{\mathbf{r}_1\in A\,|\,\mathbf{r}_1,\mathbf{r}_2\notin \mathrm{S.C.}\}\mathrm{Prob}\{\mathbf{r}_2\in A\,|\,\mathbf{r}_1,\mathbf{r}_2\notin \mathrm{S.C.}\}.\end{aligned}\tag{1}$$

Here, for instance, the left-hand member denotes the conditional probability that points $\mathbf{r}_1$ and $\mathbf{r}_2$ belong to phase A, given that they are not in the same cell. An alternative form of Eq. (1) is

$$\begin{aligned}&\mathrm{Prob}\{\mathbf{r}_1\in A\,|\,\mathbf{r}_2\in A\cap \mathbf{r}_1,\mathbf{r}_2\notin \mathrm{S.C.}\}\\&\quad=\mathrm{Prob}\{\mathbf{r}_1\in A\,|\,\mathbf{r}_1,\mathbf{r}_2\notin \mathrm{S.C.}\}.\end{aligned}\tag{2}$$

In the process of calculation of three-point probabilities, Miller substantially supposed[6]

$$\begin{aligned}&\mathrm{Prob}\{\mathbf{r}_1\in A\cap \mathbf{r}_2\in A\cap \mathbf{r}_1,\mathbf{r}_2\notin \mathrm{S.C.}\}\\&\quad=\mathrm{Prob}\{\mathbf{r}_1\in A\}\,\mathrm{Prob}\{\mathbf{r}_2\in A\}\,\mathrm{Prob}\{\mathbf{r}_1,\mathbf{r}_2\notin \mathrm{S.C.}\}.\end{aligned}\tag{3}$$

Brown pointed out that Miller had implicitly interpreted the independence requirement (iii′) as follows[8]:

(vi) The conditional probability that point $\mathbf{r}_1$ is in A or B, given that point $\mathbf{r}_2$ is not in the same cell with it, is independent of which phase point $\mathbf{r}_2$ is in, and is equal to the absolute probability that point $\mathbf{r}_1$ is in A or B.

We remark that this interpretation is equivalent to postulating not only Eq. (2) but also the relation[9]

$$\mathrm{Prob}\{\mathbf{r}_1\in A\,|\,\mathbf{r}_1,\mathbf{r}_2\notin \mathrm{S.C.}\}=\mathrm{Prob}\{\mathbf{r}_1\in A\}.\tag{4}$$

Indeed, combination of Eqs. (2) and (4) yields

$$\begin{aligned}&\mathrm{Prob}\{\mathbf{r}_1\in A\cap \mathbf{r}_2\in A\cap \mathbf{r}_1,\mathbf{r}_2\notin \mathrm{S.C.}\}\\&\quad=\mathrm{Prob}\{\mathbf{r}_1\in A\,|\,\mathbf{r}_2\in A\cap \mathbf{r}_1,\mathbf{r}_2\notin \mathrm{S.C.}\}\\&\qquad\times\mathrm{Prob}\{\mathbf{r}_2\in A\,|\,\mathbf{r}_1,\mathbf{r}_2\notin \mathrm{S.C.}\}\mathrm{Prob}\{\mathbf{r}_1,\mathbf{r}_2\notin \mathrm{S.C.}\}\\&\quad=\mathrm{Prob}\{\mathbf{r}_1\in A\}\mathrm{Prob}\{\mathbf{r}_2\in A\}\mathrm{Prob}\{\mathbf{r}_1,\mathbf{r}_2\notin \mathrm{S.C.}\},\end{aligned}\tag{5}$$

which is nothing but Eq. (3).

On the other hand, it is obvious that a symmetric cell material obeys the condition

$$\text{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \in \text{S. C.} \mid \mathbf{r}_1 \in A\} = \text{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \in \text{S. C.} \mid \mathbf{r}_1 \in B\}, \tag{6}$$

whence

$$\text{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \in \text{S. C.} \mid \mathbf{r}_1 \in A\} = \text{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \in \text{S. C.}\}. \tag{7}$$

Namely, the symmetry condition (iv) is rewritten as:

(iv′) The event of point $\mathbf{r}_1$ being in A or B is statistically independent of the event of points $\mathbf{r}_1$ and $\mathbf{r}_2$ being in the same cell.

In Ref. 8, Brown showed that Eq. (6) necessarily follows from the assumptions (ii) and (vi), and concluded that the concept of the asymmetric cell material is not self-consistent. However, Eq. (6) or (7) can be derived without assuming the statistical homogeneity and isotropy of the medium. After some manipulation of Eq. (4) we have

$$\text{Prob}\{\mathbf{r}_1 \in A \cap \mathbf{r}_1, \mathbf{r}_2 \notin \text{S. C.}\} = \text{Prob}\{\mathbf{r}_1 \in A\}\text{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \notin \text{S. C.}\}, \tag{8}$$

so that

$$\text{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \notin \text{S. C.} \mid \mathbf{r}_1 \in A\} = \text{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \notin \text{S. C.}\}, \tag{9}$$

which reduces to Eq. (6).

A serious question arises as to the above reasoning by Miller and Brown: Does the independence hypothesis (iii′) really imply the statement (v) or (vi)? In (iii′) we speak of properties of cells, while in (v) and (vi) we speak of properties of points.[9] Therefore, it is natural to expect that (iii′), (v), and (vi) are not always equivalent to one another.

As the simplest couterexample consider a one-dimensional cell material on the x axis where cells of lengths a and b are arranged at random. Here the term "random arrangement" means that the process of placing cells on the x axis is an independent trial and the probability of choosing cells of each length has a common value 1/2. Denote a cell of length a by C_a and that of length b by C_b. Moreover, phase A or B is independently assigned to C_a cells with probability p_a or $q_a = 1 - p_a$, and to C_b cells with probability p_b or $q_b = 1 - p_b$, respectively. The one-dimensional two-phase material thus defined is clearly a kind of cell material, because it satisfies the requirements (i)—(iii). Especially for one-point probabilities, we easily obtain

$$\text{Prob}\{\mathbf{r}_1 \in C_a\} = a/(a+b), \tag{10}$$

$$\text{Prob}\{\mathbf{r}_1 \in C_b\} = b/(a+b) \tag{11}$$

and

$$\text{Prob}\{\mathbf{r}_1 \in A\} = p_a \frac{a}{a+b} + p_b \frac{b}{a+b}, \tag{12}$$

$$\text{Prob}\{\mathbf{r}_1 \in B\} = q_a \frac{a}{a+b} + q_b \frac{b}{a+b}, \tag{13}$$

which are constant in conformity with the postulate of statistical homogeneity in (ii).

For convenience let us suppose that $a < b$ and $p_a \neq p_b$. Proceeding in the same way as we did in Sec. 3B of Ref. 1, we can evaluate two-point conditional probabilities as

$$\text{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \notin \text{S. C.} \mid \mathbf{r}_1 \in C_a\} = \begin{cases} |x_{12}|/a & \text{for } |x_{12}| \leq a \\ 1 & \text{for } |x_{12}| \geq a, \end{cases} \tag{14}$$

$$\text{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \notin \text{S. C.} \mid \mathbf{r}_1 \in C_b\} = \begin{cases} |x_{12}|/b & \text{for } |x_{12}| \leq b \\ 1 & \text{for } |x_{12}| \geq b, \end{cases} \tag{15}$$

where $x_{12} = x_2 - x_1$. Multiplication of Eq. (14) or (15) by Eq. (10) or (11) gives

$$\text{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \notin \text{S. C.} \cap \mathbf{r}_1 \in C_a\} = \begin{cases} |x_{12}|/(a+b) & \text{for } |x_{12}| \leq a \\ a/(a+b) & \text{for } |x_{12}| \geq a, \end{cases} \tag{16}$$

$$\text{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \notin \text{S. C.} \cap \mathbf{r}_1 \in C_b\} = \begin{cases} |x_{12}|/(a+b) & \text{for } |x_{12}| \leq b \\ b/(a+b) & \text{for } |x_{12}| \geq b, \end{cases} \tag{17}$$

whence

$$\text{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \notin \text{S. C.}\} = \begin{cases} 2|x_{12}|/(a+b) & \text{for } |x_{12}| \leq a \\ (a+|x_{12}|)/(a+b) & \text{for } a \leq |x_{12}| \leq b \\ 1 & \text{for } |x_{12}| \geq b. \end{cases} \tag{18}$$

Straightforward calculation shows

$$\text{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \notin \text{S. C.} \cap \mathbf{r}_1 \in A\} = \begin{cases} (p_a + p_b)[|x_{12}|/(a+b)] & \text{for } |x_{12}| \leq a \\ p_a[a/(a+b)] + p_b[|x_{12}|/(a+b)] & \text{for } a \leq |x_{12}| \leq b \\ p_a[a/(a+b)] + p_b[b/(a+b)] & \text{for } |x_{12}| \geq b, \end{cases} \tag{19}$$

$$\text{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \notin \text{S. C.} \cap \mathbf{r}_1 \in B\} = \begin{cases} (q_a + q_b)[|x_{12}|/(a+b)] & \text{for } |x_{12}| \leq a \\ q_a[a/(a+b)] + q_b[|x_{12}|/(a+b)] & \text{for } a \leq |x_{12}| \leq b \\ q_a[a/(a+b)] + q_b[b/(a+b)] & \text{for } |x_{12}| \geq b, \end{cases} \tag{20}$$

which lead to

$$\text{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \notin \text{S. C.} \mid \mathbf{r}_1 \in A\} = \begin{cases} (p_a + p_b)|x_{12}|/(p_a a + p_b b) & \text{for } |x_{12}| \leq a \\ (p_a a + p_b|x_{12}|)/(p_a a + p_b b) & \text{for } a \leq |x_{12}| \leq b \\ 1 & \text{for } |x_{12}| \geq b, \end{cases} \tag{21}$$

$$\text{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \notin \text{S. C.} \mid \mathbf{r}_1 \in B\} = \begin{cases} (q_a + q_b)|x_{12}|/(q_a a + q_b b) & \text{for } |x_{12}| \leq a \\ (q_a a + q_b|x_{12}|)/(q_a a + q_b b) & \text{for } a \leq |x_{12}| \leq b \\ 1 & \text{for } |x_{12}| \geq b. \end{cases} \tag{22}$$

From Eqs. (18)—(20) we find

$$\mathrm{Prob}\{r_1 \in A \mid r_1, r_2 \notin \mathrm{S.C.}\}$$
$$= \begin{cases} (p_a + p_b)/2 & \text{for } |x_{12}| \le a \\ p_a[a/a + |x_{12}|)] + p_b[|x_{12}|/a + |x_{12}|)] & \text{for } a \le |x_{12}| \le b \\ p_a[a/(a+b)] + p_b[b/(a+b)] & \text{for } |x_{12}| \ge b, \end{cases} \tag{23}$$

$$\mathrm{Prob}\{r_1 \in B \mid r_1, r_2 \notin \mathrm{S.C.}\}$$
$$= \begin{cases} (q_a + q_b)/2 & \text{for } |x_{12}| \le a \\ q_a[a/a + |x_{12}|)] + q_b[|x_{12}|/(a + |x_{12}|)] & \text{for } a \le |x_{12}| \le b \\ q_a[a/(a+b)] + q_b[b/(a+b)] & \text{for } |x_{12}| \ge b. \end{cases} \tag{24}$$

Notice that joint and conditional two-point probabilities calculated above do not depend upon the absolute positions x_1 and x_2, but instead depend only upon the relative distance $|x_{12}|$. Needless to say, this is a direct consequence of the statistical homogeneity and isotropy postulated in (ii). Since Eqs. (21) and (22) contradict Eq. (6), our model material belongs to the category of an asymmetric cell material. In other words, there exists an example of the asymmetric cell material which is in fact self-consistent. Thus we arrive at the conclusion that the asymmetric cell material is self-consistent as well as the symmetric cell material. The cause of the error made by Miller (and Brown) lies in the misinterpretation of the independence hypothesis (iii). Their assumption (vi) substituted for (iii′) turns out to include an unnecessary statement that should rather be regarded as an equivalent for the symmetry condition (iv′). Actually, Eqs. (21)—(24) indicate that neither Eq. (4) nor (6) holds for the present model.

Next we will show that the cell material under consideration does not meet the requirement (v). The proof described below was suggested by Brown.[9] For simplicity we restrict ourselves to the case where $a < x_{12} < b < 2a$. Assume that point r_1 is in C_a; then point r_2 is in a different cell. Let a random variable ξ represent the distance from point r_1 to the right end of its cell; obviously $0 \le \xi \le a$. The event of $\xi < x_{12} - a$ occurs with probability $(x_{12} - a)/a$. Then point r_2 is in the second cell to the right if the first is a C_a cell, and in the first if it is a C_b cell. The probability that point r_2 is in C_a is therefore $\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot 0 = \frac{1}{4}$. When $x_{12} - a < \xi < a$ [with probability $(2a - x_{12})/a$], point r_2 is in the first cell and has probability 1/2 of being in C_a. Thus,

$$\mathrm{Prob}\{r_2 \in C_a \cap r_1, r_2 \notin \mathrm{S.C.} \mid r_1 \in C_a\}$$
$$= \frac{1}{4}\frac{x_{12} - a}{a} + \frac{1}{2}\frac{2a - x_{12}}{a} = \frac{3a - x_{12}}{4a}. \tag{25}$$

Combining Eq. (25) with Eq. (10), we get

$$\mathrm{Prob}\{r_1 \in C_a \cap r_2 \in C_a \cap r_1, r_2 \notin \mathrm{S.C.}\} = (3a - x_{12})/4(a+b). \tag{26}$$

Similarly,

$$\mathrm{Prob}\{r_1 \in C_a \cap r_2 \in C_b \cap r_1, r_2 \notin \mathrm{S.C.}\} = (a + x_{12})/4(a+b), \tag{27}$$

$$\mathrm{Prob}\{r_1 \in C_b \cap r_2 \in C_a \cap r_1, r_2 \notin \mathrm{S.C.}\} = (a + x_{12})/4(a+b), \tag{28}$$

$$\mathrm{Prob}\{r_1 \in C_b \cap r_2 \in C_b \cap r_1, r_2 \notin \mathrm{S.C.}\} = (3x_{12} - a)/4(a+b). \tag{29}$$

It follows from Eqs. (26)—(29) that

$$\mathrm{Prob}\{r_1 \in A \cap r_2 \in A \cap r_1, r_2 \notin \mathrm{S.C.}\}$$
$$= p_a{}^2\frac{3a - x_{12}}{4(a+b)} + 2p_a p_b\frac{a + x_{12}}{4(a+b)} + p_b{}^2\frac{3x_{12} - a}{4(a+b)}, \tag{30}$$

$$\mathrm{Prob}\{r_1 \in A \cap r_2 \in A \mid r_1, r_2 \notin \mathrm{S.C.}\}$$
$$= \frac{p_a{}^2(3a - x_{12}) + 2p_a p_b(a + x_{12}) + p_b{}^2(3x_{12} - a)}{4(a + x_{12})}. \tag{31}$$

It is readily seen that Eqs. (23) and (31) are incompatible with Eq. (1).

Now we reconsider the meaning of the independence assumptions. Miller's definition (iii′) can be interpreted as follows:

(vii) Let the subdivision of the material space into cells be fixed; under the condition that points r_1 and r_2 are in different cells, then, the event of point r_1 being in phase A or B is statistically independent of the event of point r_2 being in A or B.

If we use (vii) instead of (v), Eq. (1) is replaced by

$$\mathrm{Prob}\{r_1 \in A \cap r_2 \in A \mid r_1, r_2 \notin \mathrm{S.C.} \cap \alpha\}$$
$$= \mathrm{Prob}\{r_1 \in A \mid r_1, r_2 \notin \mathrm{S.C.} \cap \alpha\}\mathrm{Prob}\{r_2 \in A \mid r_1, r_2 \notin \mathrm{S.C.} \cap \alpha\}, \tag{32}$$

or, equivalently,

$$\mathrm{Prob}\{r_1 \in A \cap r_2 \in A \cap r_1, r_2 \notin \mathrm{S.C.} \mid \alpha\}$$
$$= \mathrm{Prob}\{r_1 \in A \mid r_1, r_2 \notin \mathrm{S.C.} \cap \alpha\}\mathrm{Prob}\{r_2 \in A \cap r_1, r_2 \notin \mathrm{S.C.} \mid \alpha\}, \tag{33}$$

where α refers to a particular pattern of the subdivision into cells. Hence we obtain[10]

$$\mathrm{Prob}\{r_1 \in A \cap r_2 \in A \cap r_1, r_2 \notin \mathrm{S.C.}\}$$
$$= \sum_\alpha \mathrm{Prob}\{\alpha\}\mathrm{Prob}\{r_1 \in A \cap r_2 \in A \cap r_1, r_2 \notin \mathrm{S.C.} \mid \alpha\}$$
$$= \sum_\alpha \mathrm{Prob}\{\alpha\}\mathrm{Prob}\{r_1 \in A \mid r_1, r_2 \notin \mathrm{S.C.} \cap \alpha\} \times \mathrm{Prob}\{r_2 \in A \cap r_1, r_2 \notin \mathrm{S.C.} \mid \alpha\}, \tag{34}$$

which does not necessarily imply Eq. (1).

As asserted by Brown,[8,9] the most practical procedure for guaranteeing the independence property is that we first divide the space by some random procedure into statistically equivalent cells and then allot each cell independently to material A or B. His method of constructing cell materials is characterized by the rule:

(viii) The process of assigning phase A or B to each cell is statistically independent of the subdivision of the space into cells.

This statement is stronger than Miller's symmetry condition (iv′); for it implies

$$\mathrm{Prob}\{r_1 \in A \mid r_1, r_2 \notin \mathrm{S.C.} \cap \alpha\} = \mathrm{Prob}\{r_1 \in A \mid r_1, r_2 \notin \mathrm{S.C.}\} = \mathrm{Prob}\{r_1 \in A\}. \tag{35}$$

In such a case, Eq. (34) is transformed into

$$\mathrm{Prob}\{\mathbf{r}_1 \in A \cap \mathbf{r}_2 \in A \cap \mathbf{r}_1, \mathbf{r}_2 \notin \mathrm{S.C.}\}$$
$$= \mathrm{Prob}\{\mathbf{r}_1 \in A \,|\, \mathbf{r}_1, \mathbf{r}_2 \notin \mathrm{S.C.}\}\mathrm{Prob}\{\mathbf{r}_2 \in A \cap \mathbf{r}_1, \mathbf{r}_2 \notin \mathrm{S.C.}\}, \tag{36}$$

which can be reduced to Eq. (1). To sum up, Miller's definition (iii) implies our independence requirement (v), provided that Brown's symmetry condition (viii) is fulfilled. Unfortunately, we have not yet produced any asymmetric model that satisfies both of the two independence criteria (iii') and (v).

Finally we wish to correct some expressions for the effective permittivity of asymmetric cell materials derived in Refs. 1 and 2. Let the permittivity $\epsilon(\mathbf{r})$ be ϵ_A in phase A and ϵ_B in phase B. Denote ensemble averaging by angular brackets and put $\epsilon'(\mathbf{r}) = \epsilon(\mathbf{r}) - \langle\epsilon\rangle$. Then the two-point moment $\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\rangle$ becomes

$$\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\rangle = \epsilon_A'^2\mathrm{Prob}\{\mathbf{r}_1 \in A \cap \mathbf{r}_1, \mathbf{r}_2 \in \mathrm{S.C.}\}$$
$$+ \epsilon_B'^2\mathrm{Prob}\{\mathbf{r}_1 \in B \cap \mathbf{r}_1, \mathbf{r}_2 \in \mathrm{S.C.}\}$$
$$+ \epsilon_A'^2\mathrm{Prob}\{\mathbf{r}_1 \in A \cap \mathbf{r}_2 \in A \cap \mathbf{r}_1, \mathbf{r}_2 \notin \mathrm{S.C.}\}$$
$$+ \epsilon_A'\epsilon_B'\mathrm{Prob}\{\mathbf{r}_1 \in A \cap \mathbf{r}_2 \in B \cap \mathbf{r}_1, \mathbf{r}_2 \notin \mathrm{S.C.}\}$$
$$+ \epsilon_A'\epsilon_B'\mathrm{Prob}\{\mathbf{r}_1 \in B \cap \mathbf{r}_2 \in A \cap \mathbf{r}_1, \mathbf{r}_2 \notin \mathrm{S.C.}\}$$
$$+ \epsilon_B'^2\mathrm{Prob}\{\mathbf{r}_1 \in B \cap \mathbf{r}_2 \in B \cap \mathbf{r}_1, \mathbf{r}_2 \notin \mathrm{S.C.}\}, \tag{37}$$

where $\epsilon_A' = \epsilon_A - \langle\epsilon\rangle$ and $\epsilon_B' = \epsilon_B - \langle\epsilon\rangle$. For a symmetric cell material satisfying (viii), substitution of Eqs. (3) and (8) into Eq. (37) yields

$$\langle\epsilon'(\mathbf{r}_1)\epsilon'(\mathbf{r}_2)\rangle$$
$$= \langle\epsilon'^2\rangle\mathrm{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \in \mathrm{S.C.}\} + \langle\epsilon'\rangle^2\mathrm{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \notin \mathrm{S.C.}\}$$
$$= \langle\epsilon'^2\rangle\mathrm{Prob}\{\mathbf{r}_1, \mathbf{r}_2 \in \mathrm{S.C.}\}. \tag{38}$$

This is equivalent to Eq. (3.5) in Ref. 1, so that previous results need no revision. In the asymmetric case, nevertheless, Eq. (37) does not ensure the validity of simple averaging formulas such as Eq. (3.3) of Ref. 1, because the last four terms on the right-hand side of Eq. (37) can give nonzero contributions. Similarly, Eqs. (5.18) and (5.19) in Ref. 2 do not hold true. In general, it is difficult to compute two- or more-point moments for asymmetric cell materials.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge enlightening discussions with Professor W. F. Brown, Jr., who especially pointed out the nonequivalence of the independence assumptions (iii') and (v). Thanks are also due to Y. Akeda and T. Aoki for their beneficial suggestions and constructive criticism, and to Miss K. Kameya for her aid in typing the manuscript.

[1] M. Hori, J. Math. Phys. **14**, 514 (1973).
[2] M. Hori, J. Math. Phys. **14**, 1942 (1973).
[3] M. Hori and F. Yonezawa, J. Math. Phys. **15**, 2177 (1974).
[4] M. Hori and F. Yonezawa, J. Math. Phys. **16**, 352 (1975).
[5] M. Hori and F. Yonezawa, J. Math. Phys. **16**, 365 (1975).
[6] M.N. Miller, J. Math. Phys. **10**, 1988 (1969).
[7] M.N. Miller, J. Math. Phys. **10**, 2005 (1969); **12**, 1057 (1971).
[8] W. F. Brown Jr., J. Math. Phys. **15**, 1516 (1974).
[9] W. F. Brown Jr. (private communication).
[10] Generally speaking, the sums in Eq. (34) should be replaced by the corresponding Lebesgue–Stieltjes integrals.

Operator algebra of dual resonance models

M. Arik,* D. D. Coon,† and Yuk-ming Lam‡

Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260
(Received 18 April 1975)

Properties of a set of operators introduced by Baker, Coon, and Yu are discussed. The operators involve generalizations of harmonic oscillator operators and facilitate the construction of a family of dual resonance models which includes the Veneziano model as a limiting case. Matrix representations of the operators are constructed, and it is shown that the operators have finite norm in contrast with the unboundedness of creation and annihilation operators of the Veneziano model.

I. INTRODUCTION

Recently, Baker, Coon, and Yu[1] (hereafter referred to as I) introduced an operator formulation of a one-parameter family of dual resonance models[2-4] which includes the Veneziano model as a limiting case. The operator formulation (I) is similar to the operator formulation of the Veneziano model which was found by Fubini, Gordon, and Veneziano[5] and Nambu.[6] An interesting difference is that, in I, a *finite* number of operators "replace" the infinite number of operators of the Veneziano model formulations.[5,6] This gain is only partially offset by the fact that the new operators are not quite as simple as harmonic oscillator operators because they have additional advantage over the corresponding Veneziano model operators in that they are bounded. There is apparently a highly nontrivial element involved here since there seems to be no simple way to re-express the new operators in terms of harmonic oscillator operators even apart from the fact that re-expressing bounded operators in terms of unbounded operators is not advantageous. From the point of view of construction of dual models it is clear that the new operators represent a considerable mathematical improvement since a class of models which is *wider* than the Veneziano model can be expressed in terms of a smaller (finite rather than infinite) number of operators with better properties.

To our knowledge, the new operators have not previously appeared in the literature. The purpose of this paper is to provide some mathematical information about the operators themselves and to point out some of their interesting properties. Apart from their usefulness in formulating dual resonance theories, the new operators may have some significance from a purely mathematical point of view. The parameter which characterizes the operators and the associated dual resonance models[2-4] is directly related to the parameter which occurs in elliptic theta functions.[7,8] There are also connections with number theory[9] and the permutation group.[4]

In Sec. II, the equations satisfied by the operators of I are given and explicit infinite matrix representations of the operators are constructed. The set of creation and annihilation operators is not closed under a commutation relation. The infinite algebra generated by these operators is analyzed in Sec. III, and it is shown that all elements of the algebra have finite norm. We refer to the algebra as a q-algebra since it is characterized by a parameter q with $0<q<1$. In the limit $q\to 1$, the creation and annihilation operators of the q-algebra become the usual unbounded harmonic oscillator operators.

II. REPRESENTATIONS OF THE OPERATORS

The creation operators $a^{\dagger\mu}$ and the annihilation operators a_μ satisfy the relation:

$$a_\mu a^{\dagger\nu}=qa^{\dagger\nu}a_\mu+\delta_\mu^\nu \quad \text{with} \quad \mu,\nu=1,2,3,\ldots,D \tag{1}$$

and[10] $0<q<1$. The inverse Hamiltonian satisfies

$$a_\mu H^{-1}=qH^{-1}a_\mu, \quad H=H^\dagger \tag{2}$$

and the vacuum $|0\rangle$ with $\langle 0|0\rangle=1$ satisfies

$$a_\mu|0\rangle=0$$

and

$$H^{-1}|0\rangle=|0\rangle. \tag{3}$$

For each integer l, the "l-particle" states and the conjugate states are defined by

$$\begin{aligned}&|\mu_1\mu_2\cdots\mu_l\rangle\equiv a^{\dagger\mu_l}a^{\dagger\mu_{l-1}}\cdots a^{\dagger\mu_1}|0\rangle,\\ &\langle\mu_1\cdots\mu_l|\equiv\langle 0|a_{\mu_1}a_{\mu_2}\cdots a_{\mu_l}.\end{aligned} \tag{4}$$

There are D^l linearly independent l-particle states.

It follows from (2), (3), and (4) that

$$H^{-1}|\mu_1\mu_2\cdots\mu_l\rangle=q^l|\mu_1\mu_2\cdots\mu_l\rangle \tag{5}$$

and from (1) and (3) that

$$\langle\nu_1\nu_2\cdots\nu_l|\mu_1\mu_2\cdots\mu_l\rangle\equiv N_{\nu_1\nu_2\cdots\nu_l}^{\mu_1\mu_2\cdots\mu_l}$$

with

$$N_{\nu_1\nu_2\cdots\nu_l}^{\mu_1\mu_2\cdots\mu_l}=\sum_P q^T\delta_{\nu_{i_1}}^{\mu_1}\delta_{\nu_{i_2}}^{\mu_2}\cdots\delta_{\nu_{i_l}}^{\mu_l}, \tag{6}$$

where the summation is over all permutations $(i_1i_2\cdots i_l)$ of $(1,2\cdots l)$ and T is the minimum number of adjacent transpositions needed to generate the permutation. At this point we introduce the notation

$$\lambda=(\lambda_1\cdots\lambda_l)$$

for brevity and define the inverse N^{-1} of the tensor N by[11]

$$N_\mu^{-1\lambda}N_\nu^\mu=N_\nu^{-1\mu}N_\mu^\lambda=\delta_{\nu_1}^{\lambda_1}\delta_{\nu_2}^{\lambda_2}\cdots\delta_{\nu_l}^{\lambda_l}.$$

We define

$$\langle\tilde{\nu}|=N_\nu^{-1\mu}\langle\mu| \quad \text{(summation over } \mu_i\text{'s)}$$

and

$$|\tilde{\mu}\rangle=N_\mu^{-1\lambda}|\lambda\rangle.$$

It then follows that

$$\langle \widetilde{\nu_1 \cdots \nu_{l'}} | \mu_1 \cdots \mu_l \rangle = \delta^{\mu_1}_{\nu_1} \delta^{\mu_2}_{\nu_2} \cdots \delta^{\mu_l}_{\nu_{l'}} \delta_{ll'} \tag{7}$$

and

$$\langle \widetilde{\nu_1 \cdots \nu_{l'}} | H^{-1} | \mu_1 \cdots \mu_l \rangle = q^l \delta_{l'l} \delta^{\mu_1}_{\nu_1} \cdots \delta^{\mu_2}_{\nu_2} \cdots \delta^{\mu_l}_{\nu_{l'}}. \tag{8}$$

We take these relations as the starting point for our infinite matrix representation.

We first identify the states $|\mu_1 \cdots \mu_l\rangle$ with infinite component column vectors:

$$|0\rangle \equiv \begin{pmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}, \qquad |\mu_1 \cdots \mu_l\rangle \equiv \begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \end{pmatrix}$$

with 1 as the nth component where $n = \mu_1 + D\mu_2 + \cdots + D^{l-1}\mu_l$. The remaining components are all zero. Similarly, we identify the conjugate tilde states $\langle \widetilde{\nu_1 \cdots \nu_l} |$ with infinite component row vectors $\langle \widetilde{\nu_1 \cdots \nu_l} | = (00 \cdots 010 \cdots)$ with 1 as the nth component with $n = \nu_1 + D\nu_2 + \cdots + D^{l-1}\nu_l$. The scalar product as defined by (7), instead of (6), becomes matrix multiplication.

To construct the infinite matrix representation of operators, it is most convenient to write an infinite matrix A as a supermatrix consisting of submatrices in the following block form. The matrix A is written as

$$A = (\alpha_{ij}) = \begin{bmatrix} \alpha_{00} & \alpha_{01} & \cdots \\ \alpha_{10} & \alpha_{11} & \cdots \\ \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdots \end{bmatrix}.$$

where α_{ij} is a $D^i \times D^j$ submatrix for $i, j = 0, 1, 2, \cdots$. With this convention, it follows from (8) that

$$H^{-1} = (h^{-1}_{ij}) \quad \text{with} \quad h^{-1}_{ij} = q^i \delta_{ij} (\otimes^j I), \tag{9}$$

where I is the $D \times D$ unit matrix and $\otimes^j I$ is the direct product of I with itself j times. Note that H^{-1} is diagonal:

$$H^{-1} = \begin{bmatrix} 1 & 0 & 0 & \cdot & \cdot \\ 0 & q & 0 & \cdot & \cdot \\ 0 & 0 & q^2 & 0 & \cdot \\ \cdot & \cdot & \cdot & q^3 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix},$$

where each block stands for the appropriate submatrix so that q^l represents q^l times the $D^l \times D^l$ unit matrix.

The matrices $a^{\dagger\mu}$ are determined by [see (4)]

$$a^{\dagger\mu} | \mu_1 \mu_2 \cdots \mu_l \rangle = | \mu_1 \mu_2 \cdots \mu_l \mu \rangle,$$

and they are seen to be of the form

$$a^{\dagger\mu} = (a^{\dagger\mu}_{ij}), \quad \text{where } a^{\dagger\mu}_{ij} \text{ is nonzero only for } i = j + 1,$$
$$a^{\dagger\mu}_{j+1,j} = \otimes^j I \otimes \psi^\mu, \tag{10}$$

where ψ^μ is $D \times 1$ column with 1 in the μth place and zeros everywhere else. The direct product notation is such that $|\mu_1 \cdots \mu_l\rangle \to \psi^{\mu_1} \otimes \cdots \otimes \psi^{\mu_l}$ and

$$A \otimes B = \begin{pmatrix} b_{00}A & b_{01}A & \cdots \\ b_{10}A & b_{11}A & \cdots \\ \cdots & \cdots & \\ \cdots & \cdots & \end{pmatrix}.$$

The general form of the matrices a_μ is fixed by manipulating the commutation relation (2)

$$a_\mu H^{-1} = q H^{-1} a_\mu.$$

By using Eq. (9),

$$\sum_j (a_\mu)_{ij} h^{-1}_{jk} = q \sum_j h^{-1}_{ij} (a_\mu)_{jk}$$

yields

$$(a_\mu)_{ik} q^k = q\, q^i (a_\mu)_{ik} = q^{i+1} (a_\mu)_{ik}$$

and $0 < q < 1$ implies $(a_\mu)_{ik} = 0$ unless $k = i + 1$.

To find the form of the submatrix $(a_\mu)_{i,i+1}$, we use the commutation relations

$$a_\mu a^{\dagger\nu} = q\, a^{\dagger\nu} a_\mu + \delta^\nu_\mu \tag{1}$$

which gives

$$(a_\mu)_{i,i+1} a^{\dagger\nu}_{i+1,i} = q\, a^{\dagger\nu}_{i,i-1} (a_\mu)_{i-1,i} + \delta^\nu_\mu (\otimes^i I).$$

To simplify the notation, we define

$$a_{\mu i} \equiv (a_\mu)_{i,i+1} \quad \text{and} \quad a^{\dagger\nu}_i \equiv (a^{\dagger\nu})_{i+1,i}.$$

Then

$$a_{\mu i} a^{\dagger\nu}_i = q\, a^{\dagger\nu}_{i-1} a_{\mu,i-1} + \delta^\nu_\mu (\otimes^i I)$$

and substituting $a^{\dagger\nu}_i = \otimes^i I \otimes \psi^\nu$ [Eq. (10)] into this relation yields

$$a_{\mu i} (\otimes^i I \otimes \psi^\nu) = q (\otimes^{i-1} I \otimes \psi^\nu) a_{\mu,i-1} + \delta^\nu_\mu (\otimes^i I).$$

Multiplying both sides from the left by $\otimes^i I \otimes \psi_\nu$ (where ψ_ν is the transpose of ψ^ν) and summing over ν, we get

$$a_{\mu i} = q (\otimes^{i-1} I \otimes \psi^\nu) a_{\mu,i-1} (\otimes^i I \otimes \psi_\nu) + \otimes^i I \otimes \psi_\mu.$$

This equation can be iterated to solve for $a_{\mu i}$ solely in terms of direct products of I and ψ_μ:

$$a_{\mu 0} = \psi_\mu,$$
$$a_{\mu 1} = q \psi^{\lambda_1} \psi_\mu (I \otimes \psi_{\lambda_1}) + I \otimes \psi_\mu$$
$$= q \psi^{\lambda_1} \psi_\mu \otimes \psi_{\lambda_1} + I \otimes \psi_\mu,$$
$$\cdots,$$
$$a_{\mu n} = \sum_{i=1}^{n} q^i (\otimes^{n-i} I) \otimes \psi^{\lambda_i} \psi_\mu \otimes \psi^{\lambda_{i-1}} \psi_{\lambda_i} \otimes \cdots \otimes \psi^{\lambda_1} \psi_{\lambda_2} \otimes \psi_{\lambda_1}$$
$$+ \otimes^n I \otimes \psi_\mu$$

or

$$a_{\lambda_0 n} = \sum_{i=0}^{n} q^i (\otimes^{n-i} I) \otimes \psi^{\lambda_1} \psi_{\lambda_0} \otimes \psi^{\lambda_2} \psi_{\lambda_1} \otimes \psi^{\lambda_3} \psi_{\lambda_2} \otimes \cdots \otimes \psi^{\lambda_i} \psi_{\lambda_{i-1}} \otimes \psi_{\lambda_i}. \tag{11}$$

This completes the construction of matrix representations of a_μ and $a^{\dagger\nu}$, which are given by Eqs. (11) and (10).

III. STRUCTURE OF THE Q-ALGEBRA

Let β be the set of all double finite (possibly empty)

sequences of integers between 1 and N (inclusively). We associate

$$a^{\dagger\mu_1}\cdots a^{\dagger\mu_n}a_{\nu_1}\cdots a_{\nu_m}$$

with

$$(\mu_1,\ldots,\mu_n;\nu_1,\ldots,\nu_m)\in\beta, \tag{12}$$

and let $\mathcal{A}$ be the set of all formal linear combinations of a finite number of elements of β over a field $\mathcal{F}$:

$$c_1\alpha_1+\cdots+c_n\alpha_n\in\mathcal{A}, \tag{13}$$

where $c_i\in\mathcal{F}$ and $\alpha_i\in\beta$. (For the family of dual resonance models $\mathcal{F}$ is the field of complex numbers.) Thus the product of an element of $\mathcal{A}$ with an element of $\mathcal{F}$ is naturally defined in $\mathcal{A}$. Also, under addition $\mathcal{A}$ is an Abelian group. The zero element 0 is the empty linear combination, and the additive inverse of $A\in\mathcal{A}$ is $(-1)A$. Therefore, $\mathcal{A}$ is a vector space over $\mathcal{F}$ and β is a basis of $\mathcal{A}$.

Let ϕ be the empty sequence and denote the element $(\phi;\phi)$ of $\mathcal{A}$ by $\mathbb{I}$. Multiplication in $\mathcal{A}$ is defined by

$$\text{(i)}\ (\{\alpha\};\{\beta\})(\phi;\{\nu\})\equiv(\{\alpha\};\{\beta\},\{\nu\}), \tag{14}$$

$$\text{(ii)}\ (\{\alpha\};\phi)(\{\nu\};\{\beta\})\equiv(\{\alpha\},\{\nu\};\{\beta\}), \tag{15}$$

$$\text{(iii)}\ (\phi;\lambda)(\mu;\phi)\equiv q(\mu;\lambda)+\delta^{\mu}_{\lambda}\,\mathbb{I}, \tag{16}$$

(iv) multiplication is associative,

(v) multiplication is distributive,

$$(aA+bB)(cC)=(ac)(AC)+(bc)(BC),$$
$$(cC)(aA+bB)=(ca)(CA)+(cb)(CB) \tag{17}$$

for $a,b,c\in\mathcal{F}$ and $A,B,C\in\mathcal{A}$.

Here $\{\alpha\}$, $\{\beta\}$, and $\{\nu\}$ are possibly empty sequences of integers, $\{\alpha\}$, $\{\beta\}\equiv\{\alpha_1\cdots\alpha_n\beta_1\cdots\beta_m\}$, δ^{μ}_{ν} is the Kronecker delta in $\mathcal{F}$, and q is a fixed element of $\mathcal{F}$. (For the family of dual resonance models q is real and $0<q<1$.) The $\mathcal{A}$ becomes an *associative algebra*, and we will call a *q-algebra*. The identity of multiplication is $\mathbb{I}$.

It follows from the definition of multiplication that

$$(\phi;\nu)(\mu_1,\ldots,\mu_n;\phi)$$
$$=\sum_{i=1}^{n}q^{i-1}\delta^{\mu_i}_{\nu}(\mu_1,\ldots,\mu_{i-1},\mu_{i+1},\ldots,\mu_n;\phi)$$
$$+q^n(\mu_1,\ldots,\mu_n;\nu) \tag{18}$$

Natural representation of $\mathcal{A}$

We define a linear projection operator P on $\mathcal{A}$ by

$$P((\{\alpha\};\phi))=(\{\alpha\};\phi),$$

and

$$P(\{\alpha\};\{\beta\})=0 \tag{19}$$

for any nonempty sequence $\{\beta\}$. Let

$$\mathcal{H}_1\equiv P\mathcal{A}. \tag{20}$$

$P\beta$ is a basis of $\mathcal{H}_1$.

We now restrict $\mathcal{F}$ to the field of complex numbers. For any $(\{\alpha\};\{\beta\})\in\mathcal{A}$, define an operator $[\{\alpha\};\{\beta\}]$ on $\mathcal{H}_1$ by

$$\text{(i)}\ [\{\alpha\};\{\beta\}]\ \text{is linear on}\ \mathcal{H}_1,$$
$$\text{(ii)}\ [\{\alpha\};\{\beta\}](\{\mu\};\phi)\equiv P((\{\alpha\};\{\beta\})(\{\mu\};\phi))\in\mathcal{H}_1 \tag{21}$$

for any $(\{\mu\};\phi)$ in $P\beta$.

For any $A=c_1(\{\alpha_1\};\{\beta_1\})+\cdots+c_n(\{\alpha_n\};\{\beta_n\})\in\mathcal{A}$, define a linear operator $\mathcal{O}(A)$ on $\mathcal{H}_1$ by

$$\mathcal{O}(A)\equiv\sum_{i=1}^{n}c_i[\{\alpha_i\};\{\beta_i\}]. \tag{22}$$

Thus the operators $\mathcal{O}(A)$ acting on the space $\mathcal{H}_1$ form a representation of $\mathcal{A}$ via the mapping $\mathcal{O}$. It is easy to show that

$$\mathcal{O}(A)\mathcal{O}(B)=\mathcal{O}(AB) \tag{23}$$

for $A,\ B\in\mathcal{A}$.

Let us define a scalar product [in the spirit of Eq. (18)] in $\mathcal{H}_1$ by

$$\text{(i)}\ ((\mu_1,\ldots,\mu_n);(\nu_1,\ldots,\nu_m))=\delta^{n}_{m}\delta^{\mu_1}_{\nu_1}\cdots\delta^{\mu_n}_{\nu_n}, \tag{24}$$

$$\text{(ii)}\ (\Phi;\Phi)=1, \tag{25}$$

and

$$\text{(iii)}\ (aA,bB)=a^*b(A,B)\quad\text{for } a,b\in\mathcal{F}\text{ and } A,\ B\in\mathcal{H}_1, \tag{26}$$

where we have used the shorthand notation

$$(\mu_1,\ldots,\mu_n)\equiv(\mu_1,\ldots,\mu_n;\phi) \tag{27}$$

for a basis vector of $\mathcal{H}_1$ and

$$\Phi\equiv(\phi,\phi). \tag{28}$$

The Hilbert space $\mathcal{H}$ obtained by the completion of $\mathcal{H}_1$ with respect to this scalar product is also a representation of $\mathcal{A}$. An element of $\mathcal{H}$ is an infinite series

$$A=\sum_{\{\mu\}}A_{\{\mu\}}(\{\mu\}), \tag{29}$$

where the series $\sum_{\{\mu\}}|A_{\{\mu\}}|^2$ is Cauchy convergent. The scalar product of the above A with another element

$$B=\sum_{\{\nu\}}B_{\{\nu\}}(\{\nu\})$$

of $\mathcal{H}$ is induced by the scalar product in $\mathcal{H}_1$:

$$(A,B)\equiv\sum_{\{\mu\}}A^*_{\{\mu\}}B_{\{\mu\}}. \tag{30}$$

The norm of A is

$$\|A\|\equiv\sum_{\{\mu\}}|A_{\{\mu\}}|^2, \tag{31}$$

which converges for every $A\in\mathcal{H}$.

Bound of operators

For all A in $\mathcal{H}$ as given by Eq. (31)

$$[\lambda;\phi]A=\sum_{\{\mu\}}A_{\{\mu\}}(\lambda,\{\mu\}), \tag{32}$$
$$\|[\lambda;\phi]A\|=\|A\|;$$

hence

$$\|[\lambda;\phi]\|=1. \tag{33}$$

Similarly

$$\|[\{\mu\};\phi]\| = 1, \tag{34}$$

for a possibly empty sequence $\{\mu\}$.

To find the bound of $[\phi;\lambda]$, let

$$A^{(n)} \equiv \sum_{\mu_1 \cdots \mu_n} A_{\mu_1 \cdots \mu_n}(\mu_1 \cdots \mu_n)$$

be a vector in that subspace of $\mathcal{H}$ spanned by basis vectors of n integers. Then, by Eq. (18),

$$[\phi;\lambda]A^{(n)} = \sum_{\mu_1 \cdots \mu_{n-1}} \left(\sum_{i=1}^{n} q^{i-1} A_{(\mu,i)} \right)(\mu),$$

where

$$(\mu) = (\mu_1 \mu_2 \cdots \mu_{n-1})$$

and

$$A_{(\mu,i)} \equiv A_{\mu_1 \cdots \mu_{i-1} \lambda \mu_i \cdots \mu_{n-1}}.$$

The square of its norm is

$$\|[\phi;\lambda]A^{(n)}\|^2 = \sum_{\{\mu\}} \left| \sum_{i=1}^{n} q^{i-1} A_{(\mu,i)} \right|^2$$

$$= \sum_{\{\mu\}} \sum_{i,j=1}^{n} q^{i+j-2} A^*_{(\mu,i)} A_{(\mu,j)}.$$

On the other hand,

$$\|A^{(n)}\|^2 = \sum_{\mu_1 \cdots \mu_n} |A_{\mu_1 \cdots \mu_n}|^2 \geq \sum_{\mu_1 \cdots \mu_{n-1}} |A_{(\mu,i)}|^2,$$

so that for $q \geq 0$

$$(1 + q + \cdots + q^n)^2 \, \|A^{(n)}\|^2 \geq \sum_{\{\mu\}} \sum_{i,j=1}^{n} q^{i+j-2} |A_{(\mu,i)}|^2.$$

Therefore,

$$(1 + q + \cdots + q^n)^2 \|A^{(n)}\|^2 - \|[\phi;\lambda]A^{(n)}\|^2$$

$$\geq \tfrac{1}{2} \sum_{\{\mu\}} \sum_{i,j=1}^{n} q^{i+j-2} [\, |A_{(\mu,i)}|^2 + |A_{(\mu,j)}|^2 - A^*_{(\mu,i)} A_{(\mu,j)}$$

$$- A_{(\mu,i)} A^*_{(\mu,j)}]$$

$$= \tfrac{1}{2} \sum_{\{\mu\}} \sum_{i,j=1}^{n} q^{i+j-2} |A_{(\mu,i)} - A_{(\mu,j)}|^2 \geq 0.$$

Thus,

$$\|[\phi;\lambda]A^{(n)}\| \leq (1 + q + \cdots + q^{(n)}) \|A^{(n)}\|. \tag{35}$$

Any vector $A \in \mathcal{H}$ has the form

$$A = \sum_{n=0}^{\infty} A^{(n)}$$

Since $([\phi;\lambda]A^{(n)}, [\phi;\lambda]A^{(m)}) = 0$ for $n \neq m$,

$$\|[\phi;\lambda]A\|^2 = \sum_{n=0}^{\infty} \|[\phi;\lambda]A^{(n)}\|^2$$

$$\leq \sum_{n=0}^{\infty} (1 + \cdots + q^n)^2 \|A^{(n)}\|^2 \leq [1/(1-q)^2] \, \|A\|^2.$$

Therefore, for $0 \leq q < 1$

$$\|[\phi;\lambda]A\| \leq [1/(1-q)] \, \|A\| \tag{36}$$

for any vector $A \in \mathcal{H}$, from which we conclude

$$\|[\phi;\lambda]\| \leq 1/(1-q). \tag{37}$$

Let $B^{(n)} \in \mathcal{H}$ be given by

$$B^{(n)} = (\lambda, \ldots, \lambda)$$

where there are n λ's. Then

$$\|B^{(n)}\| = 1$$

and

$$\|[\phi;\lambda]B^{(n)}\| / \|B^{(n)}\| = 1 + q + \cdots + q^{n-1}. \tag{38}$$

Given any $\epsilon > 0$, we can always find an integer n_0 such that

$$\left| 1/(1-q) - (1 + \cdots + q^{n-1}) \right| < \epsilon, \quad \text{for } n < n_0, \; 0 \leq q < 1.$$

Therefore, combining this with (37) and (38), we have

$$\|[\phi;\lambda]\| = 1/(1-q). \tag{39}$$

Thus we have proven

$$\|[\{\mu\};\lambda_1 \cdots \lambda_l]\| \leq 1/(1-q)^l$$

and it is easy to show that actually equality holds. Since the q-algebra consists of all finite linear combination of $[\{\mu\},\{\lambda\}]$, all the elements of the q-algebra have finite norm.

*Present address: Brookhaven National Laboratory, Upton, New York 11973.

†Work supported in part by the National Science Foundation under Grant No. GP-43713.

‡Present address: institute for Theoretical Physics, Freie Universität Berlin, Arnimallee 3, D-1000 Berlin 33, Germany.

[1]M. Baker, D.D. Coon, and S. Yu, Phys. Rev. D **5**, 1429 (1972).
[2]M. Baker and D.D. Coon, Phys. Rev. D **2**, 2349 (1970).
[3]M. Baker and D.D. Coon, Phys. Rev. D **4**, 1234 (1971); D.D. Coon, Phys. Lett. B **29**, 669 (1969).
[4]S. Yu, M. Baker, and D.D. Coon, Phys. Rev. D **5**, 3108 (1972); D **6**, 538 (1972).
[5]S. Fubini, D. Gordon, and G. Veneziano, Phys. Lett. B **29**, 679 (1969).
[6]Y. Nambu, in *Symmetries and Quark Models* edited by R. Chand (Gordon and Breach, New York, 1970).
[7]E.T. Whittaker and G.N. Watson, *Modern Analysis* (Cambridge U.P., London, 1963).
[8]W.N. Bailey, *Generalized Hypergeometric Series* (Cambridge U.P., London, 1935); L.J. Slater, *Generalized Hypergeometric Functions* (Cambridge U.P., London, 1966).
[9]*High Transcendental Functions*, Bateman Manuscript Project, edited by A. Erdelyi (McGraw-Hill, New York, 1953), Vol. 3.
[10]For the family of dual resonance models, [2-4] $D = 6$.
[11]We have been able to show the existence of the inverse of N up to $l = 3$ and we conjecture that N^{-1} exists for all l and $|q| < 1$. For any finite l, N^{-1} can fail to exist at only a finite number of values of q.

Null geodesic deviation. I. Conformally flat space–times

P. C. Peters

Department of Physics, University of Washington, Seattle, Washington 98195
and Battelle Seattle Research Center, 4000 N.E. 41st Street, Seattle, Washington 98105
(Received 20 March 1975)

The equation of geodesic deviation is solved in conformally flat space–time in a covariant manner. The solution is given as an integral equation for general geodesics. The solution is then used to evaluate second derivatives of the world function and derivatives of the parallel propagator, which need to be known in order to find the Green's function for wave equations in curved space–time. A method of null geodesic limits of two-point functions is discussed, and used to find the scalar Green's function as an iterative series.

PACS numbers: 04.20.C, 02.40.H

1. INTRODUCTION

The equation of geodesic deviation is fundamental in the operational definition of the curvature of space–time in a coordinate independent manner.[1] Although this definition depends only on the local deviation of geodesics, one many wish to find the deviation of geodesics after a finite change in the affine parameter, given suitable initial conditions. For example, a study of null geodesic deviation, applicable to photon trajectories, would describe much of the optical phenomena associated with curved space–time. A knowledge of geodesic deviation is also needed to evaluate, or simplify, the first and higher derivatives of two-point geometrical quantities,[2] e.g., the world function[2] (or geodetic interval[3]), the parallel propagator[2] (or parallel displacement bivector[3]), and other geometric quantities related to these. Since these geometrical quantities exist independent of a coordinate system, it would be desirable if such derivatives could be expressed in a coordinate independent manner, i.e., in terms of other geometrical quantities. The particular application envisioned here is to the study of Green's functions for perturbation equations (scalar, electromagnetic, and gravitational) in selected space–times. For this study it suffices to solve the equation of geodesic deviation only for null or near null geodesics.

If one can solve the geodesic equation in some coordinate system, one can compute directly the geodesic deviation for particular initial values. The deviation vector will not, of course, be expressed in coordinate independent geometrical quantities, and the procedure may be poor for numerical computations, since one would need to evaluate small differences in numerically evaluated quantities. A more direct approach has been discussed by Synge.[2] Here one defines an orthogonal tetrad at one space–time point and establishes the tetrad at other space–time points by parallel transport along geodesics. By considering tetrad components of the deviation vector and Riemann tensor, the equation of geodesic deviation can be written as an integral equation with specified end values of the deviation vector. For a general space–time this can be solved only by iteration, giving, in effect, an expansion of the solution in powers of the Riemann tensor. This solution is applied by Synge[2] to the evaluation of covariant derivatives of geometrical quantities, given as a similar expansion. For weak gravitational fields such an expansion has been useful in consideration of Green's functions for perturbation fields.[4]

The tetrad method does not take advantage of possible symmetries in the Riemann tensor for particular geometries, since although the original tetrad may be chosen to exhibit these symmetries, the parallel propagated tetrad, in general, will not. We consider a covariant method of solution of the equation of geodesic deviation which is useful when the Riemann tensor can be simply expressed in terms of geometrically defined, lower rank tensor fields. The deviation vector, in effect, is resolved in terms of field projections rather than coordinate components or tetrad components. A trial solution of the deviation vector is constructed out of the vector and its projections, using available geometrical fields. Substitution of the trial solution into the equation of geodesic deviation leads to a series of differential equations for the coefficients of the various terms, which can then be solved by standard techniques. The solution can then be applied to the calculations of covariant derivatives of the two-point geometrical quantities, following Synge.

In this paper we illustrate the method by applying the formalism to any conformally flat space–time. The most reasonable physical geometries of this type are the Friedmann cosmological models; we will, however, keep our solution general, expressing quantities in terms of the geometric scalar field that relates the conformally flat metric to a flat metric by a conformal transformation.[2] We will illustrate some of the applications of this method by calculating covariant derivatives of selected geometrical quantities, which would otherwise be given as a series expansion in powers of the Riemann tensor, and by constructing the scalar Green's function for the scalar wave equation in this geometry.

2. GEODESIC DEVIATION

Consider a family of geodesics with special parameter u emanating from a fixed point x_1 (with $u=u_1$) and terminating at a variable point $x_2(v)$ (with $u=u_2$) on an arbitrary curve $C(v)$. This defines a two-space $x^\mu(u,v)$ with tangent vectors to the geodesics $U^\mu=\partial x^\mu/\partial u$ and deviation vector $V^\mu=\partial x^\mu/\partial v$. The equation of geodesic deviation is then[2]

$$\frac{\delta^2 V^\mu}{\delta u^2}+R^\mu{}_{\alpha\beta\gamma}U^\alpha V^\beta U^\gamma=0. \tag{2.1}$$

We wish to solve (2.1) for specified end values of the deviation vector V^μ; here we choose $V^{\mu 1} = V^\mu(u_1) = 0$ and $V^{\mu 2} = V^\mu(u_2)$, which is an arbitrary vector at x_2. The solution for $V^{\mu 1} \neq 0$ can be obtained by a superposition of the solution generated here and this solution with u_1 and u_2 interchanged.

One integral along the geodesic of (2.1) is immediately found by multiplying (2.1) by U_μ and using the fact that $\delta U^\mu/\delta u = 0$, since U^μ is the tangent vector along the geodesic. This gives the result that $\delta^2(U_\mu V^\mu)/\delta u^2 = 0$, which can be integrated with the assigned end values to give

$$U_\mu V^\mu = U_{\mu_2} V^{\mu 2}(u - u_1)/(u_2 - u_1). \tag{2.2}$$

For the case of a general Riemann tensor no other such integrals are found.

We now specialize to the case of a conformally flat space–time for which geometry the Weyl tensor $C_{\mu\nu\alpha\beta} = 0$ everywhere. For such a metric there is a scalar function ψ which relates the conformally flat metric to a flat space–time metric by a conformal transformation

$$g_{\mu\nu} = e^{-\psi} g^F_{\mu\nu} \equiv G^2 g^F_{\mu\nu}. \tag{2.3}$$

In terms of ψ the Riemann tensor can be written as

$$R^\mu{}_{\alpha\beta\gamma} = -\delta^\mu{}_\gamma P_{\alpha\beta} + \delta^\mu{}_\beta P_{\alpha\gamma} + g_{\alpha\gamma} P^\mu{}_\beta - g_{\alpha\beta} P^\mu{}_\gamma, \tag{2.4}$$

where $P_{\mu\nu}$ is defined in terms of ψ as

$$P_{\mu\nu} = \tfrac{1}{2}\psi_{;\mu;\nu} - \tfrac{1}{4}\psi_{;\mu}\psi_{;\nu} + \tfrac{1}{8} g_{\mu\nu}\psi_{;\alpha}\psi^{;\alpha} \tag{2.5}$$

and where covariant differentiation in (2.5) is carried out with respect to the conformally flat metric $g_{\mu\nu}$.

Substitution of (2.4) and (2.5) into (2.1) yields the equation of geodesic deviation for a conformally flat space–time

$$\frac{\delta^2 V^\mu}{\delta u^2} + \frac{U^\mu}{G}\left(\frac{\delta G_{;\alpha}}{\delta u}\right) V^\alpha - \frac{V^\mu}{G}\frac{d^2 G}{du^2} = F^\mu, \tag{2.6}$$

where $G(u) = G(x(u))$ is defined by (2.3) and F^μ is given by

$$F^\mu = Q^\mu{}_\alpha [U^\alpha (U_\beta V^\beta) - V^\alpha (U_\beta U^\beta)] \tag{2.7}$$

with

$$Q_{\mu\nu} = P_{\mu\nu} + \tfrac{1}{8} g_{\mu\nu}\psi_{;\alpha}\psi^{;\alpha}. \tag{2.8}$$

Given the fiducial geodesic $x^\mu(u)(v = 0)$, $G(u)$, $\psi^{;\mu}$, and $\psi^{;\mu}{}_{;\nu}$ are all known functions of u. The only unknown is $V^\mu(u)$. Note that from (2.2) the first term of (2.7) is a known function of u and thus its contribution in (2.6) can be treated as an inhomogeneous term in the differential equation. If the fiducial geodesic is null, even though the test geodesics may be nonnull, the second term of (2.7) vanishes. If the fiducial geodesic is nearly null, we can consider, to first order in $U_\alpha U^\alpha$, the second term in (2.7) to be known, since one can use for V^μ the solution derived for the null case.

Consider the equation (2.6) with $F^\mu = 0$, which we call the homogeneous equation. The solution of the homogeneous equation is given as a superposition of two vectors $g^\mu{}_{\alpha_2} V^{\alpha 2}$ and U^μ with coefficients which are functions of u. With $F^\mu \neq 0$, we add a term which we call the inhomogeneous solution V^μ_I, which is assumed to be determined from F^μ. The solution of (2.6) can then be given as

$$V^\mu = g(u) g^\mu{}_{\alpha_2} V^{\alpha 2} + h(u) U^\mu + V^\mu_I, \tag{2.9}$$

where $g(u)$ and $h(u)$ are the scalar coefficients subject to the boundary conditions $g(u_1) = h(u_1) = h(u_2) = 0$, $g(u_2) = 1$; $g^\mu{}_{\alpha_2}$ is the parallel propagator, which parallel transports the deviation vector at x_2 to the point $x(u)$ along the geodesic; and the boundary conditions of the inhomogeneous solution V^μ_I are chosen so that $V^{\mu 1}_I = V^{\mu 2}_I = 0$. Substituting (2.9) into (2.6) and requiring that the coefficients of like vectors vanish gives the series of differential equations for g, h, and V^μ_I:

$$\frac{d^2 g}{du^2} - \frac{d^2 G}{du^2}\frac{g}{G} = 0; \tag{2.10}$$

$$\frac{d^2 h}{du^2} = -\frac{1}{G}\left(\frac{\delta G_{;\alpha}}{\delta u}\right) T^\alpha, \tag{2.11}$$

$$\frac{\delta^2 V^\mu_I}{\delta u^2} - \frac{d^2 G}{du^2}\frac{V^\mu_I}{G} = F^\mu, \tag{2.12}$$

where

$$T^\lambda = g(u) g^\lambda{}_{\alpha_2} V^{\alpha 2} + V^\lambda_I. \tag{2.13}$$

In deriving these differential equations we have made use of the fact that $\delta g^\mu{}_{\alpha_2}/\delta u = 0$.

The two independent solution of (2.10) are $G(u)$ and $G(u)\int^u du'/G^2(u')$. Imposing the boundary conditions $g(u_1) = 0$, $g(u_2) = 1$ gives the solution of (2.10)

$$g(u) = G(u)A(u)/G_2 A_2, \tag{2.14}$$

where

$$A(u) = \int_{u_1}^u du'/G^2(u') \tag{2.15}$$

and $G_2 = G(u_2)$, $A_2 = A(u_2)$. The same solutions of the homogeneous equation are used to generate the solution of (2.12) subject to the boundary conditions $V^{\mu 1}_I \equiv V^\mu_I(u_1) = 0$ and $V^{\mu 2}_I \equiv V^\mu_I(u_2) = 0$:

$$V^\mu_I(u) = G(u)[H^\mu(u) - A(u) g^\mu{}_{\alpha_2} H^{\alpha 2}/A_2], \tag{2.16}$$

where

$$H^\mu(u) = \int_{u_1}^u \frac{du'}{G^2(u')} \int_{u_1}^{u'} du'' G(u'') g^\mu{}_{\alpha''} F^{\alpha''} \tag{2.17}$$

and $H^{\mu 2} = H^\mu(u_2)$. Given $g(u)$ from (2.14) and V^μ_I from (2.16), we then known T^λ and thus the right side of (2.11). Subject to the boundary conditions $h(u_1) = h(u_2) = 0$, this gives the expression for $h(u)$:

$$h(u) = -K(u) + \frac{u - u_1}{u_2 - u_1} K_2, \tag{2.18}$$

where

$$K(u) = \int_{u_1}^u du' \int_{u_1}^{u'} du'' \frac{1}{G}\left(\frac{\delta G_{;\alpha''}}{\delta u''}\right) T^{\alpha''}. \tag{2.19}$$

The sum of the three terms in (2.9), using our solutions (2.14), (2.16), and (2.18), define the deviation vector over the range $u_1 \leq u \leq u_2$. If we multiply (2.9) by U_μ, it can be shown directly that our solution satisfies (2.2).

If the fiducial geodesic is null, then our solution gives the deviation vector once the end value $V^{\mu 2}$ is specified. If the fiducial geodesic is not null, then the deviation vector V^μ appears in F^μ, which is needed to generate V_I^μ, giving an integral equation for V^μ. Since the term with V^μ appears multiplied by the factor $U_\alpha U^\alpha$, we can generate a series solution for V^μ in powers of the parameter $U_\alpha U^\alpha$. This is unlike the expansion of Synge,[2] since our result is exact for null fiducial geodesics. We note also the fact that the solution for $g(u)$ does not depend on such an expansion in powers of the parameter $U_\alpha U^\alpha$.

For our later use it is desirable to have V^μ explicitly in terms of the end point deviation $V^{\mu 2}$. Define the two-point tensor $S^\mu{}_{\nu_2}$ by

$$V^\mu = S^\mu{}_{\nu_2} V^{\nu 2}. \tag{2.20}$$

Then $S^\mu{}_{\nu_2}$ has the explicit form

$$\begin{aligned} S^\mu{}_{\nu_2} &= g(u) g^\mu{}_{\nu_2} + G(u)\left(W^\mu - \frac{A(u) g^\mu{}_{\alpha_2} W^{\alpha 2}}{A_2}\right) U_{\nu_2} \\ &\quad - U^\mu \left(X_{\nu_2}(u) - \frac{u-u_1}{u_2-u_1} X_{\nu_2}(u_2)\right) \\ &\quad - U^\mu U_{\nu_2}\left(J(u) - \frac{u-u_1}{u_2-u_1} J_2\right) \\ &\quad - U_\alpha U^\alpha \left[G(u)\left(M^\mu{}_{\nu_2} - \frac{A(u) g^\mu{}_{\beta_2} M^{\beta 2}{}_{\nu_2}}{A_2}\right)\right. \\ &\quad \left. + U^\mu \left(Y_{\nu_2}(u) - \frac{u-u_1}{u_2-u_1} Y_{\nu_2}(u_2)\right)\right], \end{aligned} \tag{2.21}$$

where

$$W^\mu = \int_{u_1}^{u} \frac{du'}{G^2} \int_{u_1}^{u'} du'' G g^\mu{}_{\alpha''} Q^{\alpha''}{}_{\beta''} U^{\beta''} \frac{u''-u_1}{u_2-u_1},$$

$$X_{\nu_2}(u) = \int_{u_1}^{u} du' \int_{u_1}^{u'} \frac{du''}{G}\left(\frac{\delta G_{;\alpha''}}{\delta u}\right) g^{\alpha''}{}_{\nu_2} g(u''),$$

$$J(u) = \int_{u_1}^{u} du' \int_{u_1}^{u'} du'' G_{;\lambda''}\left(W^{\lambda''} - \frac{A(u'') g^{\lambda''}{}_{\nu_2} W^{\alpha 2}}{A_2}\right),$$

$$M^\mu{}_{\nu_2} = \int_{u_1}^{u} \frac{du'}{G^2} \int_{u_1}^{u'} du'' G g^\mu{}_{\alpha''} Q^{\alpha''}{}_{\beta''} S^{\beta''}{}_{\nu_2},$$

$$Y_{\nu_2}(u) = \int_{u_1}^{u} du' \int_{u_1}^{u'} du'' \times G_{;\lambda''}\left(M^{\lambda''}{}_{\nu_2} - \frac{A(u'') g^{\lambda''}{}_{\alpha_2} M^{\alpha 2}{}_{\nu_2}}{A_2}\right).$$

Again note that if $U_\alpha U^\alpha = 0$, the solution is given in closed form, but if $U_\alpha U^\alpha \neq 0$, the solution is given in iterated form, since the solution appears in the expression for the matrix $M^\mu{}_{\nu_2}$.

3. GEOMETRICAL RELATIONS

We now show how the solution to the equation of geodesic deviation is used to evaluate first and higher derivatives of two-point geometrical quantities, following Synge.[2] Let Ω be one-half the square of the proper time between x_1 and x_2 along the geodesic joining x_1 and x_2. We assume that x_1 and x_2 are not conjugate points, i.e., that there is only one geodesic which joins them. Ω is called the world function by Synge[2] and the geodetic interval by DeWitt.[3] The covariant derivatives of Ω at the end points are related to the tangent vectors by

$$\begin{aligned} \Omega_{;\mu_1} &= -(u_2-u_1) U_{\mu_1}, \\ \Omega_{;\mu_2} &= (u_2-u_1) U_{\mu_2}, \end{aligned} \tag{3.1}$$

from which it follows that $\Omega_{;\alpha}\Omega^{;\alpha} = 2\Omega$, and $g^\mu{}_{\alpha'}\Omega^{;\alpha'} = -\Omega^{;\mu}$. To evaluate second derivatives of Ω, we take the variation of (3.1) with respect to v, giving

$$\Omega_{;\mu_1;\nu_2} V^{\nu 2} = -(u_2-u_1)\left(\frac{\delta V_\mu}{\delta u}\right)_1, \tag{3.2a}$$

$$\Omega_{;\mu_2;\nu_2} V^{\nu 2} = (u_2-u_1)\left(\frac{\delta V_\mu}{\delta u}\right)_2, \tag{3.2b}$$

where we have used the fact that $\delta U_\mu/\delta v = \delta V_\mu/\delta u$. From Sec. 2 we have the solution for $V^\mu(u)$ and therefore we can express $\delta V^\mu/\delta u$ at each of the end points in terms of $V^{\nu 2}$, the arbitrary end-point deviation. In (3.2a) and (3.2b) this means that coefficients of $V^{\nu 2}$ on both sides of this equation must be equal, giving the desired derivatives of Ω explicitly in terms of known quantities. Formally this gives

$$\Omega_{;\mu_1;\nu_2} = -(u_2-u_1)\left[\frac{\delta}{\delta u} S_{\mu\nu_2}\right]_{u=u_1}, \tag{3.3}$$

$$\Omega_{;\mu_2;\nu_2} = (u_2-u_1)\left[\frac{\delta}{\delta u} S_{\mu\nu_2}\right]_{u=u_2}. \tag{3.4}$$

We shall evaluate these derivatives only for null fiducial geodesics, $U_\alpha U^\alpha = 0$. Higher order terms in the parameter $U_\alpha U^\alpha$ will be denoted by $O(\Omega)$. From (3.3) and (3.4) we have

$$\begin{aligned} \Omega_{;\mu_1;\nu_2} &= -\frac{(u_2-u_1)}{G_1 G_2 A_2} g_{\mu_1\nu_2} + \frac{(u_2-u_1)}{G_1 A_2} g_{\mu_1\alpha_2} W^{\alpha 2} U_{\nu_2} \\ &\quad - U_{\mu_1} X_{\nu_2}(u_2) - U_{\mu_1} U_{\nu_2} J_2 + O(\Omega), \end{aligned} \tag{3.5}$$

$$\begin{aligned} \Omega_{;\mu_2;\nu_2} &= (u_2-u_1)\left[\left(\frac{G_2'}{G_2} + \frac{1}{G_2^2 A_2}\right) g_{\mu_2\nu_2} - (B_{\mu_2} U_{\nu_2} + U_{\mu_2} B_{\nu_2})\right. \\ &\quad \left. + \left(L_2 - J'(u_2) + \frac{J_2}{u_2-u_1}\right) U_{\mu_2} U_{\nu_2}\right] + O(\Omega), \end{aligned} \tag{3.6}$$

where

$$B_{\mu_2} = \frac{1}{G_2 A_2} \int_{u_1}^{u_2} du A(u)\left(\frac{\delta}{\delta u} G_{;\alpha}\right) g^\alpha{}_{\mu_2}\left(\frac{u-u_1}{u_2-u_1}\right),$$

$$L_2 = \frac{1}{G_2 A_2} \int_{u_1}^{u_2} du \frac{A(u)}{G} G_{;\alpha} G^{;\alpha}\left(\frac{u-u_1}{u_2-u_1}\right),$$

and prime indicates differentiation with respect to u. It is straightforward to verify that $U^{\mu 1}\Omega_{;\mu_1;\nu_2} = -U_{\nu_2}$, $U^{\nu 2}\Omega_{;\mu_1;\nu_2} = -U_{\mu_1}$ and $U^{\mu 2}\Omega_{;\mu_2;\nu_2} = U_{\nu_2}$, as is required from differentiating the identity $\Omega_{;\alpha}\Omega^{;\alpha} = 2\Omega$ with respect to either end point. The contraction of (3.6) is especially simple

$$\Omega_{;\mu_2}{}^{;\mu 2} = 2\left[(u_2-u_1)\left(\frac{G_2'}{G_2} + \frac{1}{G_2^2 A_2}\right) + 1\right] + O(\Omega). \tag{3.7}$$

Higher derivatives of Ω are found by varying (3.5) or (3.6) with respect to the parameter v. This requires a knowledge of the variation of the parallel propagator,

which therefore implies a knowledge of the covariant derivative of the parallel propagator. Following Synge,[2] the variation in $g_{\mu_1\nu_2}$ for an arbitrary end-point derivation vector V^{λ_2} is given by

$$g_{\mu_1\nu_2;\lambda_2}V^{\lambda_2}=\int_{u_1}^{u_2} g_{\mu_1\alpha}g_{\nu_2\beta}R^{\alpha\beta}{}_{\gamma\delta}V^{\gamma}U^{\delta}\,du. \tag{3.8}$$

Writing V^{γ} in the form (2.20) and equating coefficients of V^{λ_2} gives

$$g_{\mu_1\nu_2;\lambda_2}=\int_{u_1}^{u_2} g_{\mu_1\alpha}g_{\nu_2\beta}R^{\alpha\beta}{}_{\gamma\delta}S^{\gamma}{}_{\lambda_2}U^{\delta}\,du, \tag{3.9}$$

where $S^{\mu}{}_{\nu_2}$ is given in (2.21). Further derivatives of the parallel propagator are evaluated by varying (3.9) with respect to v. As in the case of computing the higher than second derivatives of the world function, the resulting expressions are in general quite complicated. One must be careful, too, in making these evaluations since the terms of $O(\Omega)$, which do not contribute to lower order derivatives for null fiducial geodesics, may become important for higher order derivatives. This fact arises because $\delta(U_\alpha U^\alpha)/\delta v=2\delta(U_\alpha V^\alpha)/\delta u=2U_{\alpha_2}V^{\alpha_2}/(u_2-u_1)$. Thus the variation with respect to v of a term of order Ω gives a term of order 1 in general. In application to Green's functions, however, there is much simplicity in that contributions will often be found multiplied by $\delta(\Omega)$, which automatically eliminates any terms of order Ω, and also in that only particular combinations of derivatives need to be found. We illustrate this by considering the Green's function-for the scalar wave equation, using the relations we have derived here.

4. SCALAR GREEN'S FUNCTION

The scalar Green's function is defined as the solution of the inhomogeneous scalar wave equation

$$\psi_{;\alpha}{}^{;\alpha}+aR\psi=\delta^4(x,z), \tag{4.1}$$

where a is a dimensionless constant and δ^4 is a scalar two-point function which vanishes unless $x=z$ and which satisfies the integral relation $\int\delta^4(x,z)\sqrt{-g}\,d^4x=1$. The covariant solution of (4.1) in flat space–time is

$$\psi^{(0)}=(1/4\pi)\delta_R(\Omega), \tag{4.2}$$

where Ω is the world function and δ_R gives a contribution only from the retarded root of $\Omega=0$. If we assume that the leading contribution in curved space–time is of the form (4.2) (with implied summation over multiple null geodesics), then a series solution in powers of the Riemann tensor of (4.1) can be generated from the integral equation

$$\psi(x,z)=\psi^{(0)}(x,z)-(1/4\pi)\int[aR\delta_R(\Omega(x',z))$$
$$+\delta_R'(\Omega(x',z))(\Omega_{;\alpha}{}^{;\alpha}-4)]\psi(x,x')\sqrt{-g}\,d^4x'. \tag{4.3}$$

Although the series generated from (4.3) may be useful in the case of weak gravitational fields, it is clear on physical grounds that it fails if the geometry because highly curved, as would be the case near a black hole. As an example of this, (4.2) signifies that the leading contribution is given equal weight for all null geodesics, where it is clear on physical grounds that null geodesics which wrap around the black hole many times must contribute a much smaller amount than do the principal null geodesics. Further we see that the right side of (4.3) involves a derivative of the δ function, which means that terms of order Ω will have to be included in the evaluation of $\Omega_{;\alpha}{}^{;\alpha}$.

The structure of curved space–time wave equations has been investigated by DeWitt[3] following earlier studies by Hadamard.[5] The important feature for our investigation is that the sharply propagated contribution is, in general, not (4.2), but rather

$$\tilde{\psi}^{(0)}=(\Delta^{1/2}/4\pi)\delta_R(\Omega), \tag{4.4}$$

where the two-point scalar Δ is defined by

$$\Delta=-\det(-\Omega_{;\mu;\nu'})/\sqrt{-g}\,\sqrt{-g'}. \tag{4.5}$$

From (4.5) it follows that $\Delta^{1/2}$ satisfies the differential equation

$$\frac{d}{du}\ln(\Delta^{1/2})=-\frac{1}{2(u-u_1)}(\Omega_{;\alpha}{}^{;\alpha}-4), \tag{4.6}$$

where u is the affine parameter on the geodesic connecting the two points, and implied summation over multiple geodesics again holds. The remaining contributions to ψ are smeared out inside the light cone and are generally called the tail of the Green's function. In flat space–time $\Delta^{1/2}=1$, so that (4.4) is then in agreement with (4.2). In curved space–time we can use (4.4) to generate an integral equation, analogous to (4.3), which exhibits the tail term explicitly. Thus the solution of (4.1) is given by

$$\psi(x,z)=\tilde{\psi}^{(0)}(x,z)-(1/4\pi)\int[aR\Delta^{1/2}(x',z)+(\Delta^{1/2})_{;\alpha}{}^{;\alpha}]$$
$$\times\delta_R(\Omega(x',z))\psi(x,x')\sqrt{-g}\,d^4x'. \tag{4.7}$$

Note that with the leading term given by (4.4), the corrections involve the undifferentiated δ function, so that the quantity $(\Delta^{1/2})_{;\alpha}{}^{;\alpha}$ needs to be evaluated only up to order Ω^0.

The solution (4.7), when iterated, gives a graphical representation in space–time of the generation of the Green's function from the source. The dominant contribution is that which is propagated sharply along null geodesics. The next order represents a contribution which sharply propagates along null geodesics to an intermediate point x', scatters off the curvature [represented by R or $(\Delta^{1/2})_{;\alpha}{}^{;\alpha}$], and then sharply propagates from the intermediate point to the observer point x. There is then a coherent sum (integral) over all intermediate points x' of these single scattered contributions. The next order indicates sharp propagation to a first scattering, sharp propagation from that to a second scattering, and sharp propagation from that to the observer, with summations (integrations) over all intermediate points. The iterated solution (4.7) thus shows how the signal in a curved space–time gets smeared out inside the light cone.

The calculation of the scattering strength, which involves $(\Delta^{1/2})_{;\alpha}{}^{;\alpha}$, depends on evaluating (4.5) or solving (4.6), which requires a knowledge of the second derivatives of the world function. In a general space–time

this would have to be calculated using the methods of Synge,[2] generating a solution as a power series in the Riemann tensor. Thus one could characterize the order of the scattering by the power of the Riemann tensor which appears in each term of the series. The general iterated solution would then involve a double infinite series. One could arrange this series in a triangular fashion, e.g., combining the term representing two first order scattering with the term representing a single second order scattering.

For conformally flat space–times we can calculate $\Delta^{1/2}$ along the null geodesic, either directly from (4.5), or by solving the differential equation (4.6). If we use (4.5), it is easiest to choose a parallel propagated tetrad and evaluate the determinant from the tetrad components. If the null vector U^μ had tetrad components $U^{(0)}=U^{(1)}$, $U^{(2)}=U^{(3)}=0$, then, from the requirement that $\Omega_{;\mu\;;\nu'}\Omega^{;\nu'}=\Omega_{;\mu}$ and $\Omega_{;\mu\;;\nu'}\Omega^{;\mu}=\Omega_{;\nu'}$, Δ must be given in terms of the tetrad components by

$$\Delta=\Omega_{(2)(2')}\Omega_{(3)(3')}-\Omega_{(2)(3')}\Omega_{(3)(2')}. \tag{4.8}$$

On the other hand, comparing (4.8) with our derived expression (3.5), we see that, to zeroth order in Ω, $\Omega_{(2)(3')}=\Omega_{(3)(2')}=0$ and $\Omega_{(2)(2')}=\Omega_{(3)(3')}=-(u_2-u_1)/G_1G_2A_2$, which implies that $\Delta^{1/2}$ is

$$\Delta^{1/2}=\frac{u_2-u_1}{G_1G_2\int_{u_1}^{u_2}du/G^2} \tag{4.9}$$

if the points x_1 and x_2 are separated by a null geodesic. This same relation can be derived from integrating the differential equation (4.6), using (3.7) and taking care to distinguish between derivatives with respect to u keeping the end points fixed and derivatives with respect to a variable end point u_2. We see that the only way in which $\Delta^{1/2}$ can become infinite, generally the condition for a caustic surface, is for G to be zero at one of the end points. The expression (4.9) therefore gives the explicit determination of the dominant term in the scalar Green's function, (4.4). There is no need here to worry about off-null geodesic contribution since the terms we have been ignoring are of order Ω, which give 0 when multiplied by $\delta(\Omega)$.

The calculation is not so easily done for the correction term in (4.7), which involves the quantity $(\Delta^{1/2})_{;\alpha}{}^{;\alpha}$ evaluated along a null geodesic. Suppose that $\Delta^{1/2}$ is given as an expansion in powers of Ω; i.e.,

$$\Delta^{1/2}=A+B\Omega+C\Omega^2+\cdots.$$

Then $(\Delta^{1/2})_{;\alpha}{}^{;\alpha}$ would be given as an expansion

$$(\Delta^{1/2})_{;\alpha}{}^{;\alpha}=A_{;\alpha}{}^{;\alpha}+2B_{;\alpha}\Omega^{;\alpha}+O(\Omega)$$

so that an evaluation of $(\Delta^{1/2})_{;\alpha}{}^{;\alpha}$ along the null geodesic requires a knowledge of $\Delta^{1/2}$ to first order in Ω off the null geodesic. Therefore, it does not suffice to compute the d'Alembertian of (4.9) directly for use in (4.7).

5. NULL GEODESIC LIMITS

The brute force calculation of derivatives of geometrical quantities is not always the most efficient method to use, as we now demonstrate. If we make use of the symmetries of the Riemann tensor and the form of the second derivatives of the world function, we can extract much information by considering the limits of two-point geometrical quantities as the points x and x' become separated by a null geodesic. We call such limits null geodesic limits (NGL), which form a generalization of the coincidence limits $(x\to x')$ considered by Synge.

From (3.6) we have shown that the second derivatives of the world function for conformally flat space–times have the form

$$\Omega_{;\mu\;;\nu}=Dg_{\mu\nu}+E_\mu\Omega_{;\nu}+\Omega_{;\mu}E_\nu+f_{\mu\nu}\Omega, \tag{5.1}$$

where D and E_μ could be explicitly determined, to order Ω^0, from (3.6). The identity $\Omega^{;\nu}\Omega_{;\mu\;;\nu}=\Omega_{;\mu}$ applied to (5.1) implies the relations

$$E^\alpha\Omega_{;\alpha}=1-D+l\Omega \tag{5.2}$$

where l is an undetermined function at this point, and

$$\Omega^{;\nu}f_{\mu\nu}+2E_\mu+l\Omega_{;\mu}=0. \tag{5.3}$$

It should be noted that the form of (5.1) does not give a unique prescription for D, E_μ, or $f_{\mu\nu}$. In fact $\Omega_{;\mu\;;\nu}$ remains the same under the transformation

$$\begin{aligned} &f_{\mu\nu}\to f_{\mu\nu}+\alpha g_{\mu\nu}+\beta_\mu\Omega_{;\nu}+\Omega_{;\mu}\beta_\nu\\ &D\to D-\alpha\Omega,\quad E_\mu\to E_\mu-\beta_\mu\Omega. \end{aligned} \tag{5.4}$$

We will find it convenient later to choose a particular guage.

If we differentiate $\Omega^{;\alpha}\Omega_{;\mu\;;\alpha}=\Omega_{;\mu}$, we find

$$\Omega_{;\mu;\nu}=\Omega_{;\mu\;;\alpha}\Omega^{;\alpha}{}_{;\nu}+\Omega_{;\mu\;;\nu;\alpha}\Omega^{;\alpha}+R^\sigma{}_{\mu\alpha\nu}\Omega_{;\sigma}\Omega^{;\alpha}. \tag{5.6}$$

We then substitute (5.1) and (2.4) into (5.6). In the NGL the last term of (5.1) does not appear, and the result of the substitution in the NGL gives the equalities

$$D_{;\alpha}\Omega^{;\alpha}=D-D^2-P_{\alpha\beta}\Omega^{;\alpha}\Omega^{;\beta},\qquad \text{(NGL)} \tag{5.7}$$

$$E_{\mu;\alpha}\Omega^{;\alpha}=-(D+1)E_\mu-\tfrac{1}{2}E_\alpha E^\alpha\Omega_{;\mu}+P_{\mu\alpha}\Omega^{;\alpha},\qquad \text{(NGL)} \tag{5.8}$$

which represent the differential equations for the fields D and E_μ along the null geodesic. It may be verified that the functions found in (3.6), when expressed explicitly in the form of D and E_μ, satisfy (5.7) and (5.8) in the NGL.

Although (5.6) gives one relation involving third derivatives of Ω, a more detailed relation can be obtained by differentiating (5.1) and taking the NGL. This gives

$$\begin{aligned}\Omega_{;\mu\;;\nu;\lambda}=&D_{;\lambda}g_{\mu\nu}+DE_\mu g_{\nu\lambda}+DE_\nu g_{\mu\lambda}+\Omega_{;\nu}(E_{\mu\;;\lambda}+E_\mu E_\lambda)\\&+\Omega_{;\mu}(E_{\nu;\lambda}+E_\nu E_\lambda)+\Omega_{;\lambda}(f_{\mu\nu}+2E_\mu E_\nu)\end{aligned}\qquad \text{(NGL)} \tag{5.9}$$

The requirement $\Omega_{;\mu\;;\nu;\lambda}-\Omega_{;\mu\;;\lambda;\nu}=R^\sigma{}_{\mu\nu\lambda}\Omega_{;\sigma}$ leads through (5.9) and (2.4), to the relations

$$D_{;\lambda}-DE_\lambda+P_{\lambda\alpha}\Omega^{;\alpha}=\phi\Omega_{;\lambda},\qquad \text{(NGL)} \tag{5.10}$$

$$E_{\nu;\lambda}-E_{\lambda;\nu}=\Omega_{;\nu}\xi_\lambda-\Omega_{;\lambda}\xi_\nu,\qquad \text{(NGL)} \tag{5.11}$$

$$E_{\mu;\nu}-E_\mu E_\nu-f_{\mu\nu}-P_{\mu\nu}-\phi g_{\mu\nu}+\Omega_{;\mu}\xi_\nu=0,\qquad \text{(NGL)} \tag{5.12}$$

where ϕ and ξ_ν are undetermined functions at this point. In the NGL $\Omega^{;\lambda}$ times (5.10) reproduces (5.7). If we multiply (5.12) by $\Omega^{;\nu}$ and use (5.2), (5.3), and (5.8), we obtain a consistency condition on the undetermined functions

$$l-\phi+\xi^\alpha\Omega_{;\alpha}=\tfrac{1}{2}E_\alpha E^\alpha. \qquad \text{(NGL)} \quad (5.13)$$

We next impose the transformation (5.4), choosing $\alpha=-\phi$ and $\beta_\mu=-\xi_\mu$, which makes the right sides of (5.10) and (5.11) vanish. Also (5.12) becomes

$$E_{\mu;\nu}-E_\mu E_\nu-f_{\mu\nu}-P_{\mu\nu}+\Omega_{;\mu}\xi_\nu+\xi_\mu\Omega_{;\nu}=0. \quad \text{(NGL)} \quad (5.14)$$

The symmetry in μ and ν is obvious since $E_{\mu;\nu}-E_{\nu;\mu}=0$, implying that in this gauge E_μ can be written as the gradient of a scalar function. In this same gauge from (5.2) $l\to l+\phi-\xi^\alpha\Omega_{;\alpha}$ so that (5.13) becomes

$$l+2\xi^\alpha\Omega_{;\alpha}=\tfrac{1}{2}E_\alpha E^\alpha. \qquad \text{(NGL)} \quad (5.15)$$

Further, if we differentiate (5.2) and make use of (5.1), (5.2), (5.8), and (5.10), we find that l is determined to be

$$l=\tfrac{1}{2}E_\alpha E^\alpha, \qquad \text{(NGL)} \quad (5.16)$$

which, from (5.15), implies that in this gauge $\xi^\alpha\Omega_{;\alpha}=0$. The trace of $f_{\mu\nu}$ can then be found from (5.14) as

$$f_\alpha{}^\alpha=E_\alpha{}^{;\alpha}-P_\alpha{}^\alpha-E_\alpha E^\alpha. \qquad \text{(NGL)} \quad (5.17)$$

Then from (5.9) and the other relations we find that

$$\Omega_{;\alpha}{}^{;\alpha}{}_{;\mu}=2D_{;\mu}+\Omega_{;\mu}(E_\alpha{}^{;\alpha}-P_\alpha{}^\alpha). \qquad \text{(NGL)} \quad (5.18)$$

Note that (5.18) implies a knowledge of $\Omega_{;\alpha}{}^{;\alpha}$ to first order in Ω off the null geodesic, which is all that is needed to evaluate $\Delta^{1/2}$ to first order off the null geodesic and $(\Delta^{1/2})_{;\lambda}{}^{;\lambda}$ along the null geodesic.

From (4.6) we have

$$(\ln\Delta^{1/2})_{;\alpha}\Omega^{;\alpha}=\tfrac{1}{2}(4-\Omega_{;\alpha}{}^{;\alpha}), \qquad (5.19)$$

which in the NGL implies

$$(\ln\Delta^{1/2})_{;\alpha}\Omega^{;\alpha}=(1-D). \qquad \text{(NGL)} \quad (5.20)$$

Differentiating (5.19) yields the differential equation for $(\ln\Delta^{1/2})_{;\mu}$

$$(\ln\Delta^{1/2})_{;\mu;\alpha}\Omega^{;\alpha}+D(\ln\Delta^{1/2})_{;\mu}+E^\alpha(\ln\Delta^{1/2})_{;\alpha}\Omega_{;\mu}$$
$$+E_\mu(1-D)+\tfrac{1}{2}(\Omega_{;\alpha}{}^{;\alpha}{}_{;\mu})=0 \qquad \text{(NGL)} \quad (5.21)$$

with the last term given in (5.18). The solution to this is given by

$$(\ln\Delta^{1/2})_{;\mu}=E_\mu+\zeta\Omega_{;\mu}, \qquad \text{(NGL)} \quad (5.22)$$

where ζ satisfies the differential equation

$$\zeta_{;\alpha}\Omega^{;\alpha}+2\zeta+\tfrac{1}{2}[E_\alpha{}^{;\alpha}-P_\alpha{}^\alpha+E_\alpha E^\alpha]=0. \quad \text{(NGL)} \quad (5.23)$$

Now consider the term of order Ω that is neglected in (5.22); i.e., let

$$(\ln\Delta^{1/2})_{;\mu}=E_\mu+\zeta\Omega_{;\mu}+\pi_\mu\Omega. \qquad (5.24)$$

The requirement that $(\ln\Delta^{1/2})_{;\mu;\nu}$ be symmetric in μ and ν implies, in the NGL, that

$$\pi_\mu=\zeta_{;\mu}+\epsilon\Omega_{;\mu}, \qquad \text{(NGL)} \quad (5.25)$$

where ϵ is not determined. Then the divergence of (5.24) becomes, in the NGL,

$$(\ln\Delta^{1/2})_{;\mu}{}^{;\mu}=E_\mu{}^{;\mu}+2\zeta_{;\mu}\Omega^{;\mu}+\zeta\Omega_{;\mu}{}^{;\mu}. \quad \text{(NGL)} \quad (5.26)$$

Solving (5.26) for $(\Delta^{1/2})_{;\alpha}{}^{;\alpha}$ yields the simple result

$$(\Delta^{1/2})_{;\alpha}{}^{;\alpha}=P_\alpha{}^\alpha\Delta^{1/2}. \qquad \text{(NGL)} \quad (5.27)$$

Note that, from (2.4), $P_\alpha{}^\alpha=-\frac{1}{6}R$.

We return now to the integral equation for the Green's function, (4.7). Using (5.27), we find that the Green's function ψ is given by

$$\psi(x,z)=\tilde{\psi}^{(0)}(x,z)+\frac{1}{4\pi}\int(\tfrac{1}{6}-a)R\Delta^{1/2}(x',z)\delta_R(\Omega(x',z))$$
$$\times\psi(x,x')\sqrt{-g}\,d^4x'. \qquad (5.28)$$

Note that if a is chosen to be $\frac{1}{6}$ in the scalar wave equation, then there is no scattering; i.e., $\tilde{\psi}^{(0)}(x,z)$ is then the exact solution. This should not be too surprising however, since for $a=\frac{1}{6}$ the wave equation is conformally invariant and the Green's function should be sharply propagated on the light cone.

One can derive more information about higher derivatives of Ω, $\Delta^{1/2}$ and the parallel propagator than we have done here. It appears that in applications the NGL methods yields a cleaner method of evaluation than does a direct manipulation of the explicit solutions.

ACKNOWLEDGMENTS

The author wishes to express appreciation to the Battelle Seattle Research Center and the Institute for Advanced Study for making their facilities available for this research.

[1] C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), Chap. 11.
[2] J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960), Chap. II.
[3] B.S. DeWitt, in *Relativity, Groups and Topology*, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1964).
[4] P.C. Peters, Phys. Rev. **146**, 938 (1966); Phys. Rev. D 7, 368 (1973).
[5] J. Hadamard, *Lectures on Couchy's Problem in Linear Partial Differential Equations* (Yale U.P., New Haven, 1923).

Propagation of high frequency surface waves along cylinders of general cross section

J. A. Morrison

Bell Laboratories, Murray Hill, New Jersey 07974
(Received 8 January 1975)

The propagation of high frequency scalar surface waves along the generators of a homogeneous cylinder which has a cross-sectional boundary of nonconstant curvature is investigated. Asymptotic solutions are obtained to the reduced wave equation, subject to an impedance boundary condition at the surface of the cylinder. In the case of an open boundary curve for which the curvature attains its algebraic maximum at a single point, it is found that modes exist for which the disturbance is essentially confined to a region in the neighborhood of the point of maximum curvature, as well as to the neighborhood of the surface. The amplitude of the disturbance decays rapidly on either side of the point of maximum curvature, and the higher order modes have nulls. The case of closed boundary curves is also discussed. In a companion joint paper by L. O. Wilson and the author, the asymptotic procedures developed for the analysis of the scalar problem will be applied to the investigation of the propagation of elastic surface waves (Rayleigh waves) along a homogeneous isotropic cylinder with stress-free boundary. This problem arises in connection with guided acoustic surface waves.

1. INTRODUCTION

Surface waves are disturbances which propagate along a surface, and have their amplitudes confined to a neighborhood of the surface. They occur, in particular, in electromagnetics, acoustics and elasticity. In this paper we investigate the high frequency propagation of surface waves along the generators of a homogeneous cylinder which has a cross-sectional boundary of nonconstant curvature. The boundary curve may be open or closed. We confine our attention here to scalar waves, but in a companion joint paper with Wilson[1] it will be shown that analogous techniques apply to the propagation of (Rayleigh) surface waves along a homogeneous elastic cylinder, with stress-free boundary. The elastic problem may be analyzed in terms of a scalar and a vector potential, but since this leads to considerable algebraic complexities, it is desirable to illustrate the techniques for the simpler scalar problem.

We consider a disturbance described by the three-dimensional reduced wave equation

$$\nabla^2 u + k^2 u = 0, \tag{1.1}$$

and suppress the time dependence $\exp(-ikct)$, where c is the velocity in the medium under consideration. We impose the impedance boundary condition on the surface of the cylinder

$$\frac{\partial u}{\partial n} + k\alpha u = 0 \quad \text{for } n = 0, \tag{1.2}$$

where $0 < \alpha < \infty$ is constant. Here n measures distance from the surface along the inward normal. It is assumed that, with an appropriate unit of length, the constant wave number k is large, so that (1.2) corresponds to a disturbance which decreases rapidly away from the surface. We are interested in solutions in which u varies with z, the distance along the generators of the cylinder, as $\exp(i\beta z)$, where β is a real constant which is to be determined.

The propagation of high frequency surface waves over curved surfaces has been investigated by Keller and Karal,[2] who used a complex ray extension of geometrical optics, and by Grimshaw,[3] who developed an asymptotic theory. They allowed for a medium with variable refractive index, and for a variable surface impedance which is, in general, complex, but we will not consider such generalizations here. Grimshaw assumes an asymptotic expansion of the form

$$u = \exp(ik\varphi)\sum_{\nu=0}^{\infty} A_\nu (ik)^{-\nu}, \tag{1.3}$$

where φ and A_ν are complex functions of position. One of the examples he considers is that of a circular bore, in which the propagation is in the z direction, parallel to the axis of the bore. However, (1.3) is not an appropriate form of expansion for propagation along the generators of cylinders with cross-sectional boundaries that do not have constant curvature.

In the next section we formulate the problem in terms of the coordinate system depicted in Fig. 1, where n measures distance from the surface along the inward normal, and s is signed arc length along the cross-sectional boundary. A solution to (1.1) and (1.2) is sought in the form

$$u = \exp[ik\lambda^{1/2} z - k\alpha n + k^{1/2}\theta(s)] f(n, s, k), \tag{1.4}$$

where f and λ are expanded in inverse powers of $k^{1/2}$. Attention is first confined to the case of an open boundary curve for which the curvature $\kappa(s)$ attains its algebraic maximum at a single point, taken to be $s = 0$, with $\kappa'(0) = 0$ and $\kappa''(0) < 0$. It is shown that, with $\lambda \sim 1 + \alpha^2 + k^{-1}\alpha\kappa(0)$, there is a solution with $\theta(s) \leq \theta(0) = 0$, and $\theta(s)$ tending to $-\infty$ as $|s| \to \infty$. For small s,

$$\theta(s) \sim -\tfrac{1}{2}\delta s^2, \quad |s| \ll 1, \quad \delta = [-\tfrac{1}{2}\alpha\kappa''(0)]^{1/2}. \tag{1.5}$$

The wave is therefore essentially confined to the region $|s| = O(k^{-1/4})$, as well as being confined close to the surface in the region $n = O(k^{-1})$.

The evaluation of higher order terms in the expansions of f and λ may be simplified by means of a boundary layer analysis. This is done in Sec. 3, where the stretched variable $\xi = kn$ is introduced. A particular example is also considered.

Higher order modes are analyzed in Sec. 4. Since the lowest order mode is confined to the region $|s| = O(k^{-1/4})$, we first also rescale the variable s, and introduce the stretched variable $\zeta = k^{1/4}s$. Then we seek solutions to (1.1) and (1.2) in the form

$$u = \exp(ik\lambda^{1/2}z - \alpha\xi)v(\xi, \zeta, k), \tag{1.6}$$

where v and λ are expanded in inverse powers of $k^{1/4}$, although it turns out that λ may be expanded in inverse powers of $k^{1/2}$. It is found that

$$\lambda = 1 + \alpha^2 + k^{-1}\alpha\kappa(0) - k^{-3/2}(2m+1)\delta + k^{-2}\lambda_2 + \cdots \tag{1.7}$$

and

$$v \sim D_m(\eta) = \exp(-\tfrac{1}{4}\eta^2)He_m(\eta); \quad \eta = (2\delta)^{1/2}\zeta, \tag{1.8}$$

where m is a non-negative integer. Here D_m and He_m denote parabolic cylinder functions[4] and Hermite polynomials,[5] respectively. Thus $m = 0$ corresponds to the lowest order mode considered previously. An expression for λ_2 is given in (4.12), the quantities therein being defined by (3.11). From (1.7) the propagation constant is

$$\beta = k\lambda^{1/2} = k(1+\alpha^2)^{1/2} + \tfrac{1}{2}(1+\alpha^2)^{-1/2}[\alpha\kappa(0) - k^{-1/2}(2m+1)\delta + \cdots], \tag{1.9}$$

so that the modes are not dispersionless.

Since the expansion in terms of the stretched variable ζ is not very suitable for representing the modes for moderately large values of ζ, an expansion in terms of parabolic cylinder functions with argument $\rho = 2k^{1/4} \times [-\theta(s)]^{1/2}$ is considered in the latter part of Sec. 4. This expansion is analogous to that obtained for the lowest order mode.

In Sec. 5 attention is turned to closed boundary curves. It is first assumed that the curvature attains its algebraic maximum at a single point on the boundary curve, and it is argued that modes exist which differ by only an exponentially small amount from those derived for an open boundary, because of the confinement to an interval of s of length $O(k^{-1/4})$ on either side of the point of maximum curvature. Moreover, the corresponding values of λ should differ by only an exponentially small amount.

Attention is then turned to closed boundary curves with the symmetry of an ellipse, and for which the curvature attains its algebraic maximum at two points. The region under consideration may be either interior to the boundary cylinder, corresponding to a rod, or exterior to it, corresponding to a bore, but the points of maximum algebraic curvature of the boundary curve are different for the rod and for the bore. It is argued that modes exist which are essentially symmetric and antisymmetric combinations of the modal shapes derived in the neighborhood of the points of maximum algebraic curvature, and that the values of λ for the two combinations differ by an exponentially small quantity.

These assertions are verified for a particular example, at least in so far as an approximation to the surface wave is concerned. With

$$u = \exp(ik\lambda^{1/2}z - k\alpha n)w(n, s, k), \quad \left.\frac{\partial w}{\partial n}\right|_{n=0} = 0 \tag{1.10}$$

this approximation is

$$w|_{n=0} \sim W(s, k), \quad \lambda \sim 1 + \alpha^2 + k^{-1}\mu, \tag{1.11}$$

where

$$\frac{d^2W}{ds^2} + k[\alpha\kappa(s) - \mu]W = 0. \tag{1.12}$$

A term of order W has been neglected in obtaining (1.12). The example considered is that of a boundary of circumference 2π for which $\kappa(s) = 1 + 2c\cos 2s$, with $0 < c < \frac{1}{2}$. Periodic solutions of (1.12) may be expressed in terms of Mathieu functions[6] with parameter $q = k\alpha c \gg 1$, for which asymptotic results may be used. The case of a bore, for which $\kappa(s) = -(1 - 2c\cos 2s)$, is also briefly considered.

We remark that the analysis of other cases, in which the curvature does not satisfy the properties assumed so far, hinges upon the asymptotic behavior of appropriate solutions of (1.12) for $k \gg 1$. Asymptotic solutions of equations of the form (1.12) have been considered in some generality by Lynn and Keller.[7]

Finally, in Sec. 6, in order to clarify the transition from the case of constant curvature to one in which the curvature has a strict maximum, we consider a family of boundary curves for which $\kappa(s) = \gamma_0 + k^{-1}\gamma_1(s) + \cdots$, where γ_0 is constant. The results are illustrated for the previous example, for which $\gamma_0 = 1$ and $\gamma_1(s) = 2kc\cos 2s$, where now $0 \leq c \ll 1$, and the transition to the case $k\alpha c \gg 1$ is considered. It is also pointed out that a refined approximation to the surface wave is necessary in the case $\gamma_1(s) = O(1)$, since a term of order W has been neglected in obtaining (1.12).

2. FORMULATION OF THE PROBLEM

The coordinate system is depicted in Fig. 1, wherein n is distance from the surface along the inward normal, s is signed arc length along the cross-sectional boundary curve, and z is distance along the generators of the cylinder. The unit vectors n, t, and k, which form a right-handed set, are in the directions of the inward normal and tangent to the boundary curve, and of the positive z axis, respectively. Thus,

$$\mathbf{n} = \mathbf{t} \times \mathbf{k}, \tag{2.1}$$

and[8]

$$\frac{d\mathbf{n}}{ds} = -\kappa(s)\mathbf{t}, \quad \frac{d\mathbf{t}}{ds} = \kappa(s)\mathbf{n}, \quad \frac{d\mathbf{k}}{ds} = 0, \tag{2.2}$$

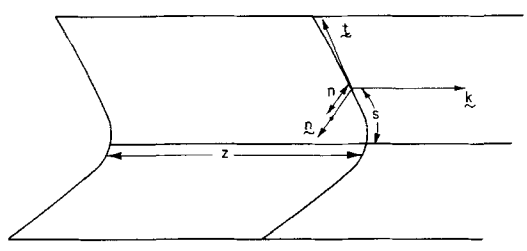


FIG. 1. Coordinate system and unit vectors.

where $\kappa(s)$ is the curvature of the boundary curve, which is assumed to be sufficiently smooth. If the curve is given by $\mathbf{r}(s)$, then

$$\frac{d\mathbf{r}}{ds} = \mathbf{t}, \tag{2.3}$$

and the three-dimensional position vector is

$$\mathbf{x} = \mathbf{r}(s) + n\mathbf{n}(s) + z\mathbf{k}. \tag{2.4}$$

From (2.2)—(2.4), it follows that

$$d\mathbf{x} = \mathbf{n}\,dn + [1 - \kappa(s)n]\mathbf{t}\,ds + \mathbf{k}\,dz. \tag{2.5}$$

The coordinate system becomes singular for $\kappa n = 1$, but, since we are considering surface waves, κn will be small in the region of interest. The line elements are, from (2.5),

$$h_n = 1, \quad h_t = 1 - \kappa(s)n \equiv h, \quad h_z = 1. \tag{2.6}$$

Hence,[9] the reduced wave equation (1.1) takes the form

$$\frac{\partial^2 u}{\partial n^2} - \frac{\kappa}{h}\frac{\partial u}{\partial n} + \frac{1}{h^2}\frac{\partial^2 u}{\partial s^2} + \frac{\kappa' n}{h^3}\frac{\partial u}{\partial s} + \frac{\partial^2 u}{\partial z^2} + k^2 u = 0. \tag{2.7}$$

In view of the boundary condition (1.2), and since we are looking for waves propagating in the z direction, we let

$$u = \exp(i\beta z)\exp(-k\alpha n)w(n, s, k), \tag{2.8}$$

and, for convenience, set

$$\beta^2 = k^2\lambda. \tag{2.9}$$

Then, from (2.7),

$$k^2(1 + \alpha^2 - \lambda)w + k\alpha\left(\frac{\kappa}{h}w - 2\frac{\partial w}{\partial n}\right) + \frac{\partial^2 w}{\partial n^2} - \frac{\kappa}{h}\frac{\partial w}{\partial n} + \frac{1}{h^2}\frac{\partial^2 w}{\partial s^2} + \frac{\kappa' n}{h^3}\frac{\partial w}{\partial s} = 0. \tag{2.10}$$

Also, the boundary condition (1.2) leads to

$$\left.\frac{\partial w}{\partial n}\right|_{n=0} = 0. \tag{2.11}$$

Both w and λ depend on k, which is large by assumption. If we expand them in inverse powers of k, as did Grimshaw,[3] then, with

$$\lambda = \lambda_0 + k^{-1}\lambda_1 + \cdots, \quad w = w_0 + k^{-1}w_1 + \cdots, \tag{2.12}$$

it follows from (2.10) that $\lambda_0 = 1 + \alpha^2$ and

$$-\lambda_1 w_0 + \alpha\left(\frac{\kappa}{h}w_0 - 2\frac{\partial w_0}{\partial n}\right) = 0. \tag{2.13}$$

Thus, from (2.6) and (2.1), setting $n = 0$ in (2.13), we find that $\lambda_1 = \alpha\kappa(s)$, which is possible only if κ is constant. Consideration of how the terms of order k in (2.10) could be balanced when $\kappa(s)$ is not constant leads to setting

$$w(n, s, k) = \exp[k^{1/2}\theta(s)]f(n, s, k), \tag{2.14}$$

and assuming expansions of the form

$$f(n, s, k) = \sum_{r=0}^{\infty} k^{-r/2} f_{r/2}(n, s), \quad \lambda = \sum_{r=0}^{\infty} k^{-r/2}\lambda_{r/2}. \tag{2.15}$$

The boundary condition (2.11) implies that

$$\left.\frac{\partial f}{\partial n}\right|_{n=0} = 0. \tag{2.16}$$

From (2.10), (2.14), and (2.15) it is found that

$$\lambda_0 = 1 + \alpha^2, \quad \lambda_{1/2} = 0. \tag{2.17}$$

The terms of order k in (2.10) then lead to

$$-\lambda_1 f_0 + \alpha\left(\frac{\kappa}{h}f_0 - 2\frac{\partial f_0}{\partial n}\right) + \left(\frac{1}{h}\frac{d\theta}{ds}\right)^2 f_0 = 0. \tag{2.18}$$

Setting $n = 0$ it follows, from (2.6) and (2.16), that

$$\left(\frac{d\theta}{ds}\right)^2 = \lambda_1 - \alpha\kappa(s). \tag{2.19}$$

We consider first the case of an open boundary curve for which $\kappa(s)$ attains its algebraic maximum at a single point, which we take to be at $s = 0$, with

$$\kappa'(0) = 0, \quad \kappa''(0) < 0, \tag{2.20}$$

and

$$\kappa(s) < \kappa(0), \quad 0 < |s| \leq s_0, \quad \kappa(s) \leq \tilde{\kappa} < \kappa(0), \quad s_0 \leq |s| < \infty. \tag{2.21}$$

If $\lambda_1 > \alpha\kappa(0)$ it follows, from (2.19) and (2.21), that $\exp[k^{1/2}\theta(s)]$ is not bounded for $|s| < \infty$. If $\lambda_1 < \alpha\kappa(0)$ then $\exp[k^{1/2}\theta(s)]$ oscillates rapidly for $\alpha\kappa(s) > \lambda_1$, and this is not the type of disturbance we are looking for. On the other hand, if

$$\lambda_1 = \alpha\kappa(0), \tag{2.22}$$

and if we choose $\operatorname{sgn}(d\theta/ds) = -\operatorname{sgn}s$, then $\theta(s)$ tends to $-\infty$ as $|s| \to \infty$. With $\theta(0) = 0$, we have

$$\theta(s) = -\int_0^s \{\alpha[\kappa(0) - \kappa(\sigma)]\}^{1/2}\operatorname{sgn}\sigma\, d\sigma, \tag{2.23}$$

where the positive square root is to be taken. It follows, from (2.20), that for small s

$$\theta(s) \sim -\tfrac{1}{2}\delta s^2, \quad |s| \ll 1, \quad \delta = [-\tfrac{1}{2}\alpha\kappa''(0)]^{1/2}. \tag{2.24}$$

Thus on the boundary curve the solution decays rapidly on either side of the point of maximum algebraic curvature. The solution is therefore essentially confined to the region $|s| = O(k^{-1/4})$, as well as being confined close to the surface in the region $n = O(k^{-1})$.

Now, from (2.18) and (2.19), using (2.6) and (2.22), it follows that

$$f_0 = \frac{b_0(s)}{[1 - \kappa(s)n]^{1/2}}\exp\left(-\frac{\kappa(s)n[1 - \kappa(0)n]}{2[1 - \kappa(s)n]}\right), \tag{2.25}$$

where $b_0(s)$ is yet to be determined. An equation for $b_0(s)$ may be obtained by considering the terms of order $k^{1/2}$ in (2.10), which involve $f_{1/2}$ and $\lambda_{3/2}$, and by applying the boundary condition (2.16). However, since we are interested only in the region $n = O(k^{-1})$, it is simpler to use a boundary layer analysis, which we do in the next section.

In a later section we discuss the applicability of the results to the case of closed boundary curves for which the curvature attains its algebraic maximum at a single point. We also consider closed boundary curves with the symmetry of an ellipse, and for which the curvature attains its algebraic maximum at two points.

3. BOUNDARY LAYER ANALYSIS

We confine our attention here to a region close to the boundary surface and set

$$\xi = kn, \quad f(n, s, k) = g(\xi, s, k). \tag{3.1}$$

Then, from (2.8), (2.9) and (2.14),

$$u = \exp[ik\lambda^{1/2}z - \alpha\xi + k^{1/2}\theta(s)]g(\xi, s, k), \tag{3.2}$$

and the boundary condition (2.16) is now

$$\left.\frac{\partial g}{\partial \xi}\right|_{\xi=0} = 0. \tag{3.3}$$

The partial differential equation satisfied by g is given by (A1) in Appendix A. We assume an expansion for g of the form

$$g(\xi, s, k) = \sum_{r=0}^{\infty} k^{-r/2} g_{r/2}(\xi, s), \tag{3.4}$$

and retain the expansion for λ given in (2.15).

It is shown in Appendix A that

$$g_0 = b_0(s), \quad g_{1/2} = b_{1/2}(s), \quad g_1 = b_1(s), \quad g_{3/2} = b_{3/2}(s), \tag{3.5}$$

and the form of g_2 is given in (A7). In addition to (2.17) and (2.19), it is found that

$$2\frac{d\theta}{ds}\frac{db_0}{ds} + \left(\frac{d^2\theta}{ds^2} - \lambda_{3/2}\right)b_0(s) = 0, \tag{3.6}$$

and an equation for $b_{1/2}(s)$ is given by (A8). We consider the case $b_0(0) \neq 0$. Then, from (2.24) and (3.6), it follows that

$$\lambda_{3/2} = -\delta. \tag{3.7}$$

Note that if instead of assuming $b_0(0) \neq 0$, we assume that $b_0(s) \sim s^m$ for $|s| \ll 1$, where m is a positive integer, then $\lambda_{3/2} = -(2m+1)\delta$. This corresponds to the higher order modes, which are treated in a different manner in the next section, for reasons which will be apparent later.

If we set

$$b_{1/2}(s) = c_{1/2}(s)b_0(s) \tag{3.8}$$

in (A8), it is found, using (2.22) and (3.6), that

$$2\frac{d\theta}{ds}\frac{dc_{1/2}}{ds} = \lambda_2 - \tfrac{1}{2}\kappa(s)[2\kappa(0) - \kappa(s)] - \frac{1}{b_0(s)}\frac{d^2b_0}{ds^2}. \tag{3.9}$$

In order that $c_{1/2}(s)$ remain bounded for small s, it follows from (2.24) that

$$\lambda_2 = \tfrac{1}{2}[\kappa(0)]^2 + \frac{1}{b_0(0)}\frac{d^2b_0}{ds^2}(0). \tag{3.10}$$

Higher order terms in the asymptotic expansions may be obtained in an analogous manner. For small s, since $\kappa'(0) = 0$, we may write

$$\kappa(s) = \kappa_0 + \kappa_2 s^2 + \kappa_3 s^3 + \kappa_4 s^4 + \cdots, \quad |s| \ll 1, \tag{3.11}$$

where, in particular, $\kappa_2 = \frac{1}{2}\kappa''(0)$. The term involving b_0 in (3.10) may be evaluated with the help of (2.19), (2.22)–(2.24), (3.6), and (3.7), and it is found that

$$\lambda_2 = \frac{\kappa_0^2}{2} - \frac{3\kappa_4}{4\kappa_2} + \frac{11\kappa_3^2}{16\kappa_2^2}. \tag{3.12}$$

The quantity $b_0(s)$ is obtained by quadrature from (3.6), wherein $\theta(s)$ is given by (2.23), and $\lambda_{3/2}$ by (3.7). Then $b_{1/2}(s)$ is given by (3.8), where $c_{1/2}(s)$ is obtained by quadrature from (3.9), with λ_2 given by (3.12). Without loss of generality we may take $c_{1/2}(0) = 0$. As a particular example, we consider

$$\kappa(s) = \gamma \operatorname{sech}^2 s, \quad 0 < \gamma < \pi/2, \tag{3.13}$$

corresponding to the boundary curve

$$x(s) = \int_0^s \cos(\gamma \tanh\sigma)\, d\sigma, \quad y(s) = -\int_0^s \sin(\gamma \tanh\sigma)\, d\sigma. \tag{3.14}$$

In this case it is found that

$$\lambda_1 = \alpha\gamma, \quad \lambda_{3/2} = -(\alpha\gamma)^{1/2}, \quad \lambda_2 = \tfrac{1}{2}(1+\gamma^2), \tag{3.15}$$

and

$$\theta(s) = -(\alpha\gamma)^{1/2}\log(\cosh s), \quad b_0(s) = b_0(0)(\cosh s)^{1/2}. \tag{3.16}$$

Also,

$$8(\alpha\gamma)^{1/2}\frac{dc_{1/2}}{ds} = -\tanh s(1 + 2\gamma^2\tanh^2 s). \tag{3.17}$$

Note that $c_{1/2}(s)$ does not remain bounded for $|s| \to \infty$, so that $k^{-1/2}c_{1/2}(s)$ does not remain small, but since u is already exponentially small for $s = O(1)$, because of the factor $\exp[k^{1/2}\theta(s)]$, this is not important.

4. THE HIGHER ORDER MODES

We have seen that the lowest order mode is essentially confined to the region $|s| = O(k^{-1/4})$. In order to analyze the higher order modes, we rescale the variable s, as well as the variable n. Thus, in (2.8) we let

$$\xi = kn, \quad \zeta = k^{1/4}s, \quad w(n, s) = v(\xi, \zeta, k). \tag{4.1}$$

Then, from (2.10) and (2.11),

$$k^2(1 + \alpha^2 - \lambda)v + k\alpha\left(\frac{\kappa}{h}v - 2k\frac{\partial v}{\partial \xi}\right) + k^2\frac{\partial^2 v}{\partial \xi^2} - \frac{k\kappa}{h}\frac{\partial v}{\partial \xi} + \frac{k^{1/2}}{h^2}\frac{\partial^2 v}{\partial \zeta^2} + \frac{\kappa'\zeta}{k^{3/4}h^3}\frac{\partial v}{\partial \zeta} = 0, \tag{4.2}$$

where h is given by (A2), and

$$\left.\frac{\partial v}{\partial \xi}\right|_{\xi=0} = 0. \tag{4.3}$$

Also, from (3.11),

$$\kappa = \kappa_0 + k^{-1/2}\kappa_2\zeta^2 + k^{-3/4}\kappa_3\zeta^3 + k^{-1}\kappa_4\zeta^4 + \cdots,$$
$$\kappa' = 2k^{-1/4}\kappa_2\zeta + \cdots. \tag{4.4}$$

We assume expansions of the form

$$v(\xi, \zeta, k) = \sum_{r=0}^{\infty} k^{-r/4} v_{r/4}(\xi, \zeta), \quad \lambda = \sum_{r=0}^{\infty} k^{-r/4}\lambda_{r/4}. \tag{4.5}$$

Proceeding in an analogous manner to that in Appendix A, we find that

$$\lambda_0 = 1 + \alpha^2, \quad \lambda_{1/4} = 0, \quad \lambda_{1/2} = 0,$$
$$\lambda_{3/4} = 0, \quad \lambda_1 = \alpha\kappa_0, \quad \lambda_{5/4} = 0 \tag{4.6}$$

and

$$v_{r/4}(\xi, \zeta) = a_{r/4}(\zeta), \quad r = 0, 1, 2, 3, 4, 5. \tag{4.7}$$

From the terms of order $k^{1/2}$ in (4.2), and the boundary condition (4.3), it then follows that

$$\frac{d^2a_0}{d\zeta^2} + (\alpha\kappa_2\zeta^2 - \lambda_{3/2})a_0(\zeta) = 0. \tag{4.8}$$

and that (4.7) holds for $r=6$.

The solution of Eq. (4.8) may be expressed in terms of parabolic cylinder functions.[4] Since $s=k^{-1/4}\zeta$, we want solutions which are bounded for large $|\zeta|$. Hence,

$$\lambda_{3/2}=-(2m+1)\delta, \quad a_0=D_m(\eta), \tag{4.9}$$

where m is a nonnegative integer, and

$$\eta=(2\delta)^{1/2}\zeta, \quad \delta=(-\alpha\kappa_2)^{1/2}. \tag{4.10}$$

In terms of Hermite polynomials[5]

$$D_m(\eta)=\exp(-\tfrac{1}{4}\eta^2)\mathrm{He}_m(\eta), \tag{4.11}$$

and $\mathrm{He}_m(\eta)$ has m zeros. Since $\eta=(2\delta)^{1/2}k^{1/4}s$ and $\mathrm{He}_m(\eta)=\eta^m-\binom{m}{2}\eta^{m-2}+\cdots$, it is now clear why the expansion in the previous section breaks down when $s=O(k^{-1/4})$ if $m\geq 2$.

The terms of order $k^{1/4}$ and 1 in (4.2) are considered in Appendix B and it is found, from an orthogonality condition on $-\infty<\zeta<\infty$, that

$$\lambda_{7/4}=0,$$

$$\lambda_2=\left[\left(\frac{\kappa_0^2}{2}-\frac{3\kappa_4}{4\kappa_2}+\frac{11\kappa_3^2}{16\kappa_2^2}\right)+\tfrac{3}{8}m(m+1)\left(\frac{5\kappa_3^2}{\kappa_2^2}-\frac{4\kappa_4}{\kappa_2}\right)\right], \tag{4.12}$$

and an expression for $a_{1/4}$ is given in (B7). It is also shown that (4.7) holds for $r=7$, and the form of v_2 is given in (B5). Note that for $m=0$ the expression for λ_2 in (4.12) agrees with that in (3.12). Thus we are able to derive the asymptotic expansion of λ by means of an analysis in the stretched variable $\zeta=k^{1/4}s$, and also to obtain the structure of the higher order modes. We remark that, using the variable ζ, we have carried out the asymptotic expansion, far enough to calculate λ_2, with the variable n, rather than the boundary layer variable $\xi=kn$, but this involves more algebra than the above analysis.

The expansion in terms of the stretched variable ζ is not very suitable for representing the modes for moderately large values of ζ. Hence we now consider an expansion for the higher order modes which corresponds to that obtained in Sec. 3 for the lowest order mode. In view of (2.14), (3.1), and (4.11), we assume that

$$w(n,s,k)=G(\xi,s,k)D_m(\rho)+mk^{-1/4}H(\xi,s,k)D_{m-1}(\rho), \tag{4.13}$$

where

$$\xi=kn, \quad \rho=k^{1/4}\chi(s), \quad [\chi(s)]^2=-4\theta(s), \tag{4.14}$$

with $\mathrm{sgn}\chi(s)=\mathrm{sgn}s$, and $\theta(s)$ as in (2.23). The equations satisfied by G and H may be derived from (2.10), with the help of the recurrence relations satisfied by the parabolic cylinder functions, and they are given by (C2) and (C3) in Appendix C. The boundary condition (2.11) leads to (C4).

In view of the form of Eqs. (C2) and (C3), we assume expansions of the form

$$G(\xi,s,k)=\sum_{r=0}^{\infty}k^{-r/2}G_{r/2}(\xi,s), \quad H(\xi,s,k)=\sum_{r=0}^{\infty}k^{-r/2}H_{r/2}(\xi,s), \tag{4.15}$$

and it is now evident that we may expand λ in powers of $k^{-1/2}$, as in (2.15), rather than assume an expansion in powers of $k^{-1/4}$, as in (4.5). As before, we have

$$\lambda_0=1+\alpha^2, \quad \lambda_{1/2}=0, \quad \lambda_1=\alpha\kappa_0, \tag{4.16}$$

and, proceeding in an analogous manner to that in Appendix A, it is found that

$$G_{r/2}(\xi,s)=p_{r/2}(s), \quad H_{r/2}(\xi,s)=q_{r/2}(s), \quad r=0,1,2. \tag{4.17}$$

We have used the relationship

$$\left(\frac{\chi}{2}\frac{d\chi}{ds}\right)^2=\alpha[\kappa(0)-\kappa(s)], \tag{4.18}$$

which follows from (2.23) and (4.14).

From the terms of order $k^{1/2}$ in (C2) and (C3), and the boundary conditions (C4), it follows that (4.17) holds for $r=3$, and that

$$\chi\frac{d\chi}{ds}+\frac{dp_0}{ds}\left[(m+\tfrac{1}{2})\left(\frac{d\chi}{ds}\right)^2+\frac{\chi}{2}\frac{d^2\chi}{ds^2}+\lambda_{3/2}\right]p_0(s)=0, \tag{4.19}$$

and the equation for $q_0(s)$ given by (C5) holds. Note that, in view of (4.14), the equation for $p_0(s)$ is equivalent to that for $b_0(s)$ in (3.6) when $m=0$. Now, for small s we have, from (2.24) and (4.14),

$$\chi(s)\sim(2\delta)^{1/2}s, \quad |s|\ll 1. \tag{4.20}$$

The requirement that $p_0(0)\neq 0$ thus implies that

$$\lambda_{3/2}=-(2m+1)\delta. \tag{4.21}$$

It is shown in Appendix C that the requirement that $q_0(0)$ be finite determines $q_0(s)$ uniquely. We also obtain an explicit expression for $q_0(s)$ in terms of $p_0(s)$ and $\chi(s)$, as given by (C6).

The terms of order 1 in (C2) are considered in Appendix C, and the boundary condition (C4) leads to the form of G_2 given in (C7), and to the equation for $p_{1/2}(s)$ given in (C8). Substituting $p_{1/2}(s)=r_{1/2}(s)p_0(s)$, using (4.19), and requiring $r_{1/2}(s)$ to remain bounded for small s, we obtain the expression for λ_2 given in (C9). The terms involving p_0 and q_0 in (C9) may be evaluated with the help of (2.23), (4.14), (4.19), and (C5), and we have verified that the resulting expression for λ_2 is that given in (4.12). We omit the rather tedious details.

5. CLOSED BOUNDARY CURVES

We have confined our attention, so far, to open boundary curves for which the curvature attains its algebraic maximum at a single point. We have seen that on the boundary curve the modes are essentially confined to an interval of length $O(k^{-1/4})$, on either side of the point of maximum algebraic curvature, and they decay exponentially outside this interval. If the boundary curve is closed, and the curvature attains its algebraic maximum at a single point, we would intuitively expect that modes would exist which differ by only an exponentially small amount from those derived in the previous section. Moreover, the corresponding values of λ would presumably also differ by an exponentially small quantity.

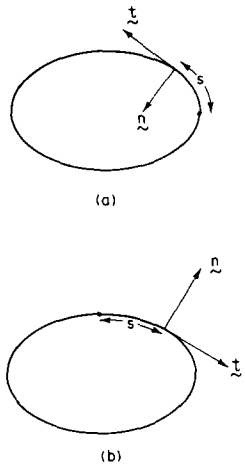


FIG. 2. Cross-sectional boundary curves, arc length s and unit vectors for (a) a rod and (b) a bore.

We now turn our attention to closed boundary curves with the symmetry of an ellipse, and for which the curvature attains its algebraic maximum at two points. The region under consideration may be either interior to the boundary cylinder, corresponding to a rod, or exterior to it, corresponding to a bore. The curvature at a given point on the boundary curve for the bore is just the negative of that for the rod, because of our convention that $\mathbf{n}$ is directed into the region, as depicted in Figs. 2(a) and (b). Note that the points of maximum algebraic curvature of the boundary curve are different for the rod and for the bore.

Asymptotic expansions may be developed, as in the previous section, about each of the two points of maximum algebraic curvature, with appropriate translation of the origin of the coordinate s. Because of the symmetry, both expansions will lead to the same expressions for λ, and to the same modal shape. Analogously to the previous discussion, we intuitively expect modes to exist with shapes differing from those corresponding to the asymptotic expansions by an exponentially small amount, and with values of λ differing by an exponentially small quantity. However, care must be taken as to what should be called a mode.

Thus, from the symmetry, it is to be expected that the true modes are either symmetric or antisymmetric. Consider, for instance, the lowest order modal shape, corresponding to $m=0$ in (4.13). Corresponding to this there should be a symmetric mode without nulls, and an antisymmetric mode with two nulls, as depicted in Figs. 3(a) and (b), respectively. (Note that the end points should be identified because of the periodicity.) Moreover, it is to be expected that the values of λ for the symmetric and the antisymmetric mode will differ by only an exponentially small amount. Analogous comments apply to the higher order modal shapes in (4.13), and the corresponding symmetric and antisymmetric modes will have $2m$ and $2m+2$ nulls, respectively.

The above assertions may be verified for a particular example, at least in so far as an approximation to the surface wave is concerned. Thus, setting $n=0$ in (2.10), and using (2.6) and (2.11), we find that

$$w\big|_{n=0} \sim W(s,k), \quad \lambda-(1+\alpha^2)\sim k^{-1}\mu, \tag{5.1}$$

where, neglecting the term $\partial^2 w/\partial n^2|_{n=0}$, which is of order W,

$$\frac{d^2W}{ds^2}+k[\alpha\kappa(s)-\mu]W=0. \tag{5.2}$$

As the particular example, we consider

$$\kappa(s)=1+2c\cos 2s, \quad 0<c<\tfrac{1}{2}, \tag{5.3}$$

corresponding to the boundary curve

$$x(s)=\int_0^s \cos(\sigma+c\sin 2\sigma)\,d\sigma, \quad y(s)=-\int_0^s \sin(\sigma+c\sin 2\sigma)\,d\sigma. \tag{5.4}$$

The length of the circumference is 2π.

When $\kappa(s)$ is given by (5.3), the solution of (5.2) may be expressed in terms of Mathieu functions, and periodic solutions, and corresponding eigenvalues, are[6]

$$\begin{aligned} &W=ce_m(s+\pi/2,q), \quad k(\alpha-\mu)=a_m(q),\\ &W=se_{m+1}(s+\pi/2,q), \quad k(\alpha-\mu)=b_{m+1}(q), \end{aligned} \tag{5.5}$$

where m is a nonnegative integer, and

$$q=k\alpha c\gg 1. \tag{5.6}$$

The solutions in (5.5) correspond to modes which are symmetric and antisymmetric respectively, about $s=\pi/2$. But,[6] for $q\gg 1$,

$$a_m(q)\sim b_{m+1}(q)\sim -2q+2(2m+1)q^{1/2}+\cdots \tag{5.7}$$

and

$$b_{m+1}(q)-a_m(q)\sim \frac{2^{4m+5}}{m!}\left(\frac{2}{\pi}\right)^{1/2} q^{(m+3/2)/2}\exp(-4q^{1/2}). \tag{5.8}$$

Thus the values of μ for the symmetric and antisymmetric modes differ by an exponentially small quantity, and, from (5.1) and (5.5)–(5.7),

$$\lambda\sim 1+\alpha^2+\alpha(1+2c)k^{-1}-2(2m+1)(\alpha c)^{1/2}k^{-3/2}+\cdots. \tag{5.9}$$

The approximate expression (5.9) is consistent with (4.16) and (4.21), as is seen from (3.11), (4.10), and (5.3), but higher order terms will differ, since terms of order W have been neglected in obtaining (5.2). Also from (4.13)–(4.15) and (4.17)–(4.19), it is found that, for $|s|<\pi/2$,

$$w\big|_{n=0}\sim p_0(0)[\sec(s/2)]^{m+1}D_m[4(k\alpha c)^{1/4}\sin(s/2)]. \tag{5.10}$$

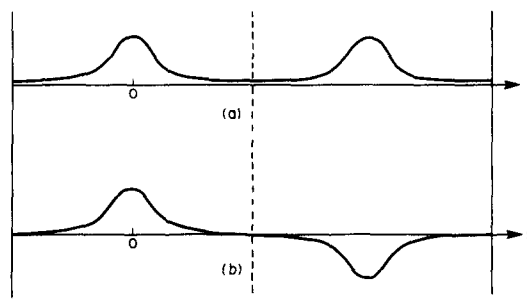


FIG. 3. (a) Symmetric, and (b) antisymmetric, modes corresponding to the lowest order modal shape.

This is consistent with the asymptotic expressions[6] for the Mathieu functions for $|\sin s| < 2^{1/4}(k\alpha c)^{-1/8}$,

$$ce_m(s+\pi/2,q) \sim (-1)^m C_m D_m[2(k\alpha c)^{1/4}\sin s],$$
$$se_{m+1}(s+\pi/2,q) \sim (-1)^m S_m D_m[2(k\alpha c)^{1/4}\sin s]. \quad (5.11)$$

Moreover,[4] for $4(k\alpha c)^{1/4}|\sin(s/2)| \gg m$,

$$D_m[4(k\alpha c)^{1/4}\sin(s/2)] \sim [4(k\alpha c)^{1/4}\sin(s/2)]^m \exp[2(k\alpha c)^{1/2}(\cos s - 1)], \quad (5.12)$$

and (5.10) is consistent with the asymptotic expressions[6] for the Mathieu functions for $|s|$ between 0 and $\pi/2$, and not close to either.

The above example corresponds to a cylindrical rod, but a similar analysis applies to the case of a bore with the same boundary curve. In this case, changing the sign of the curvature, and replacing s by $\pi/2 - s$ in (5.3),

$$\kappa(s) = -(1 - 2c\cos 2s), \quad 0 < c < \tfrac{1}{2}. \quad (5.13)$$

The eigenvalues in (5.5) are thus replaced by

$$k(\alpha+\mu) = -a_m(q), \quad k(\alpha+\mu) = -b_{m+1}(q), \quad (5.14)$$

while the eigenfunctions are unaltered. From (5.1), (5.6), (5.7), and (5.14), we now have

$$\lambda \sim 1 + \alpha^2 - \alpha(1-2c)k^{-1} - 2(2m+1)(\alpha c)^{1/2}k^{-3/2} + \cdots. \quad (5.15)$$

6. BOUNDARY CURVES WITH NEARLY CONSTANT CURVATURE

In Sec. 2 we showed that expanding λ and w in inverse powers of k is appropriate only when the prescribed curvature is constant. The results obtained in the preceding sections, which involve expansions in fractional inverse powers of k, are applicable to the case when the curvature has a strict maximum with nonvanishing second derivative. It is clear that there is a nonuniform behavior as $k \to \infty$, and in order to clarify the transition from the case of constant curvature to one in which the curvature has a strict maximum, we now consider boundary curves with nearly constant curvature, the deviation from constancy being of order k^{-1}.

Thus, for $k \gg 1$, we consider a family of boundary curves for which

$$\kappa(s) = \gamma_0 + k^{-1}\gamma_1(s) + k^{-2}\gamma_2(s) + \cdots, \quad (6.1)$$

where γ_0 is constant. As before, for simplicity, we confine our attention to the boundary layer, and let

$$\xi = kn, \quad w = \omega(\xi, s, k). \quad (6.2)$$

Then, from (2.10),

$$k^2(1+\alpha^2-\lambda)\omega + k\alpha\left(\frac{\kappa}{h}\omega - 2k\frac{\partial\omega}{\partial\xi}\right) + k^2\frac{\partial^2\omega}{\partial\xi^2} - \frac{k\kappa}{h}\frac{\partial\omega}{\partial\xi} + \frac{1}{h^2}\frac{\partial^2\omega}{\partial s^2} + \frac{\kappa'\xi}{kh^3}\frac{\partial\omega}{\partial s} = 0, \quad (6.3)$$

where h is given by (A2), and the boundary condition (2.11) implies that

$$\left.\frac{\partial\omega}{\partial\xi}\right|_{\xi=0} = 0. \quad (6.4)$$

We assume expansions of the form

$$\lambda = \sum_{r=0}^{\infty} k^{-r}\lambda_r, \quad \omega(\xi, s, k) = \sum_{r=0}^{\infty} k^{-r}\omega_r(\xi, s). \quad (6.5)$$

Proceeding as previously, the terms of order k^2 and k in (6.3), and the boundary condition (6.4), lead to

$$\lambda_0 = 1 + \alpha^2, \quad \lambda_1 = \alpha\gamma_0, \quad (6.6)$$

and

$$\omega_r(\xi, s) = \tau_r(s), \quad r = 0, 1. \quad (6.7)$$

From the terms of order 1 in (6.3), and the boundary condition (6.4), it is found that

$$\omega_2(\xi, s) = \tau_2(s) + \tfrac{1}{4}\xi^2\gamma_0^2\tau_0(s) \quad (6.8)$$

and

$$\frac{d^2\tau_0}{ds^2} + \left(\frac{\gamma_0^2}{2} + \alpha\gamma_1(s) - \lambda_2\right)\tau_0(s) = 0. \quad (6.9)$$

Asymptotic solutions of (6.9) may be obtained if $\alpha\gamma_1(s)$ is large, and perturbation expansions may be derived if $\alpha\gamma_1(s)$ is small, but in the transition region, where $\alpha\gamma_1(s)$ is $O(1)$, no approximation may be made in Eq. (6.9) for $\tau_0(s)$. Consequently, we consider a particular example for which this equation may be solved analytically, namely the closed boundary curve of circumference 2π with $\kappa(s)$ given by (5.3), where now $0 \leq c \ll 1$. Then, from (6.1),

$$\gamma_0 = 1, \quad \gamma_1(s) = 2kc\cos 2s. \quad (6.10)$$

Hence (6.9) has the form of Mathieu's equation, and the eigenfunctions are as in (5.5), with corresponding eigenvalues

$$\tfrac{1}{2} - \lambda_2 = a_m(q), \quad b_{m+1}(q), \quad q = k\alpha c. \quad (6.11)$$

Let us consider the case $q \gg 1$, which holds, for example, if $c = O(k^{-1/2})$. Then,[6]

$$a_m(q) \sim b_{m+1}(q) \sim -2q + 2(2m+1)q^{1/2} - \tfrac{1}{2}(\tfrac{1}{2} + m + m^2) + \cdots. \quad (6.12)$$

Hence, from (6.5), (6.6), and (6.10)–(6.12),

$$\lambda \sim 1 + \alpha^2 + k^{-1}\alpha + k^{-2}[2k\alpha c - 2(2m+1)(k\alpha c)^{1/2} + \tfrac{1}{2}(\tfrac{3}{2} + m + m^2) + \cdots] + k^{-3}\lambda_3 + \cdots. \quad (6.13)$$

But, from (3.11) and (5.3),

$$\kappa_0 = 1 + 2c, \quad \kappa_2 = -4c, \quad \kappa_3 = 0, \quad \kappa_4 = \tfrac{4}{3}c. \quad (6.14)$$

Thus, from (4.5), (4.6), (4.9), (4.10), and (4.12),

$$\lambda = 1 + \alpha^2 + k^{-1}\alpha(1+2c) - 2k^{-3/2}(2m+1)(\alpha c)^{1/2} + k^{-2}[\tfrac{1}{2}(1+2c)^2 + \tfrac{1}{4} + \tfrac{1}{2}m(m+1)] + \cdots. \quad (6.15)$$

Note that (6.13) and (6.15) are consistent through $O(k^{-2})$ for $c = o(1)$, $k\alpha c \gg 1$.

For small q the Mathieu functions and corresponding eigenvalues may be expanded in powers of q.[6] The limiting case $c = 0$, i.e., $q = 0$, corresponds to a circular rod with circumference 2π. It follows from (6.9) and (6.10) that the solutions of period 2π are simple

harmonic functions, and $\lambda_2 = \frac{1}{2} - l^2$, where l is an integer. Hence, from (6.5) and (6.6),

$$\lambda = 1 + \alpha^2 + k^{-1}\alpha + k^{-2}(\tfrac{1}{2} - l^2) + \cdots . \tag{6.16}$$

The exact solution to (1.1) and (1.2) may be obtained in the case of a circular rod, by separating variables in cylindrical coordinates. We have verified that the asymptotic solution of the eigenvalue equation for λ, which involves modified Bessel functions, leads to (6.16).

We conclude by commenting on the approximation to the surface wave considered in the previous section. The approximate equation (5.2) for $w|_{n=0}$ was obtained by setting $n=0$ in (2.10), and using (2.6) and (2.11), and neglecting the term $\partial^2 w/\partial n^2|_{n=0}$, which is of order W. Hence this approximation is not useful in the case of boundary curves with nearly constant curvature, when $\kappa(s)$ is given by (6.1) and $\alpha\gamma_1(s)$ is $O(1)$. However, a refined approximation to the surface wave may be obtained which is useful in this case. Thus, if (2.10) is differentiated with respect to n and then n set equal to zero, the equation obtained by setting $n=0$ in (2.10) may be used to eliminate $\partial^2 w/\partial n^2|_{n=0}$. The refined approximation involves dropping the term $\partial^3 w/\partial n^3|_{n=0}$, which is of order W, but is multiplied by a factor $O(k^{-1})$ times the coefficient of d^2W/ds^2. With $\kappa(s)$ given by (6.1), the term $\frac{1}{2}\gamma_0^2$ in (6.9) is contained in the refined approximation to the surface wave.

ACKNOWLEDGMENTS

The author is grateful to L. O. Wilson who suggested investigating high frequency waves in elastic cylinders, which prompted the analysis of the simpler scalar problem considered in this paper, and who checked some of the details of the analysis. The author is grateful to both J. McKenna and L. O. Wilson for suggesting some improvements in the presentation.

APPENDIX A

We give here some details of the boundary layer analysis in a region close to the surface. From (2.10), (2.14), and (3.1), it follows that

$$\begin{aligned} &k^2(1+\alpha^2-\lambda)g + k\alpha\left(\frac{\kappa}{h}g - 2k\frac{\partial g}{\partial \xi}\right) + k^2\frac{\partial^2 g}{\partial \xi^2} - \frac{k\kappa}{h}\frac{\partial g}{\partial \xi} \\ &+ \frac{\kappa'\xi}{kh^3}\left(k^{1/2}\frac{d\theta}{ds}g + \frac{\partial g}{\partial s}\right) + \frac{1}{h^2}\left[k\left(\frac{d\theta}{ds}\right)^2 g + k^{1/2}\left(2\frac{d\theta}{ds}\frac{\partial g}{\partial s} \right.\right. \\ &\left.\left. + \frac{d^2\theta}{ds^2}g\right) + \frac{\partial^2 g}{\partial s^2}\right] = 0, \end{aligned} \tag{A1}$$

where, from (2.6),

$$h = 1 - k^{-1}\kappa(s)\xi. \tag{A2}$$

The boundary condition is that given in (3.3), and λ and g are expanded in inverse powers of $k^{1/2}$, as in (2.15) and (3.4).

From the previous analysis, $\lambda_0 = 1 + \alpha^2$, and the terms of order k^2 in (A1) imply that

$$-2\alpha\frac{\partial g_0}{\partial \xi} + \frac{\partial^2 g_0}{\partial \xi^2} = 0. \tag{A3}$$

Rejecting the exponentially growing solution, it follows that

$$g_0 = b_0(s). \tag{A4}$$

Next, from the terms of order $k^{3/2}$ in (A1),

$$-\lambda_{1/2}g_0 - 2\alpha\frac{\partial g_{1/2}}{\partial \xi} + \frac{\partial^2 g_{1/2}}{\partial \xi^2} = 0. \tag{A5}$$

Hence, from (A4) and (A5),

$$g_{1/2} = b_{1/2}(s) - \lambda_{1/2}b_0(s)\xi/2\alpha. \tag{A6}$$

The boundary condition (3.3) implies that $\lambda_{1/2} = 0$, as before. Similarly, from the terms of order k and $k^{1/2}$ in (A1), and from the boundary condition (3.3), it is found that (2.19) and (3.6) hold, and that $g_1 = b_1(s)$ and $g_{3/2} = b_{3/2}(s)$. Finally, from the terms of order 1 in (A1),

$$g_2 = b_2(s) + \frac{\kappa\xi^2}{4\alpha}\left[\alpha\kappa + 2\left(\frac{d\theta}{ds}\right)^2\right]b_0(s), \tag{A7}$$

and, using (2.19),

$$\begin{aligned} &2\frac{d\theta}{ds}\frac{db_{1/2}}{ds} + \left(\frac{d^2\theta}{ds^2} - \lambda_{3/2}\right)b_{1/2}(s) \\ &= \left(\lambda_2 - \frac{\kappa}{2\alpha}(2\lambda_1 - \alpha\kappa)\right)b_0(s) - \frac{d^2 b_0}{ds^2}. \end{aligned} \tag{A8}$$

APPENDIX B

We consider here some of the higher order terms in the expansions in (4.5). The terms of order $k^{1/4}$ in (4.2) imply, using (4.4), (4.6), (4.7), and (A2), that

$$\begin{aligned} &2\alpha\frac{\partial v_{7/4}}{\partial \xi} - \frac{\partial^2 v_{7/4}}{\partial \xi^2} \\ &= \frac{d^2 a_{1/4}}{d\zeta^2} + (\alpha\kappa_2\zeta^2 - \lambda_{3/2})a_{1/4}(\zeta) + (\alpha\kappa_3\zeta^3 - \lambda_{7/4})a_0(\zeta). \end{aligned} \tag{B1}$$

Rejecting the solution which grows exponentially with ξ, the boundary condition (4.3), using (4.9) and (4.10), leads to

$$\begin{aligned} &\frac{d^2 a_{1/4}}{d\zeta^2} + [(2m+1)\delta - \delta^2\zeta^2]a_{1/4}(\zeta) \\ &= \left(\lambda_{7/4} + \frac{\kappa_3}{\kappa_2}\delta^2\zeta^3\right)D_m[(2\delta)^{1/2}\zeta], \end{aligned} \tag{B2}$$

and to $v_{7/4} = a_{7/4}(\zeta)$. Multiplying equation (B2) by $D_m[(2\delta)^{1/2}\zeta]$, which is a solution of the homogeneous equation, and integrating with respect to ζ from $-\infty$ to ∞, it follows that

$$\lambda_{7/4} = 0. \tag{B3}$$

Next, from the terms of order 1 in (4.2), and from the boundary condition (4.3), it is found that

$$\begin{aligned} &\frac{d^2 a_{1/2}}{d\zeta^2} + (\alpha\kappa_2\zeta^2 - \lambda_{3/2})a_{1/2}(\zeta) \\ &= \left(\lambda_2 - \frac{\kappa_0^2}{2} - \alpha\kappa_4\zeta^4\right)a_0(\zeta) - \alpha\kappa_3\zeta^3 a_{1/4}(\zeta) \end{aligned} \tag{B4}$$

and

$$v_2 = a_2(\zeta) + \tfrac{1}{4}\kappa_0^2\zeta^2 a_0(\zeta). \tag{B5}$$

From (4.9), (4.10), and (B4), it follows that

$$\int_{-\infty}^{\infty}\left(\lambda_2-\frac{\kappa_0^2}{2}+\frac{\kappa_4}{4\kappa_2}\eta^4\right)[D_m(\eta)]^2\,d\eta + \frac{(2\delta)^{1/2}\kappa_3}{4\kappa_2}\int_{-\infty}^{\infty}\eta^3 a_{1/4}D_m(\eta)\,d\eta=0. \tag{B6}$$

Using properties of the parabolic cylinder functions,[4] it is found, from (4.10), (B2), and (B3), that

$$a_{1/4}=\frac{-\kappa_3}{4(2\delta)^{1/2}\kappa_2}\left[\tfrac{1}{3}D_{m+3}(\eta)+3(m+1)D_{m+1}(\eta) - 3m^2D_{m-1}(\eta)-\tfrac{1}{3}m(m-1)(m-2)D_{m-3}(\eta)\right]. \tag{B7}$$

Then the expression for λ_2 given in (4.12) follows from (B6) and (B7), using (4.11) and the orthogonality relationships for the Hermite polynomials[5]

$$\int_{-\infty}^{\infty}D_l(\eta)D_n(\eta)\,d\eta=(2\pi)^{1/2}n!\,\delta_{ln}, \tag{B8}$$

where $\delta_{ln}=1$ if $l=n$, and is 0 otherwise.

APPENDIX C

We consider here the representation (4.13) of the higher order modes. Now,[4]

$$\begin{aligned}D_m'(\rho)&=mD_{m-1}(\rho)-\tfrac{1}{2}\rho D_m(\rho),\\ D_{m-1}'(\rho)&=\tfrac{1}{2}\rho D_{m-1}(\rho)-D_m(\rho),\end{aligned} \tag{C1}$$

where the prime denotes differentiation with respect to the argument. Substituting (4.13) into (2.10), using (4.14) and (C1), and equating to zero the coefficients of $D_m(\rho)$ and $mk^{-1/4}D_{m-1}(\rho)$, it is found that

$$\begin{aligned}&k^2(1+\alpha^2-\lambda)G+k\alpha\left(\frac{\kappa}{h}G-2k\frac{\partial G}{\partial\xi}\right)+k^2\frac{\partial^2G}{\partial\xi^2}\\ &-\frac{k\kappa}{h}\frac{\partial G}{\partial\xi}+\frac{\kappa'\xi}{kh^3}\left(-\frac{k^{1/2}}{2}\chi\frac{d\chi}{ds}G+\frac{\partial G}{\partial s}-m\frac{d\chi}{ds}H\right)\\ &+\frac{1}{h^2}\left\{k\left(\frac{\chi}{2}\frac{d\chi}{ds}\right)^2G+\left[\frac{\partial^2G}{\partial s^2}-m\left(\frac{d^2\chi}{ds^2}H+2\frac{d\chi}{ds}\frac{\partial H}{\partial s}\right)\right]\right.\\ &\left.-\frac{k^{1/2}}{2}\left[(2m+1)\left(\frac{d\chi}{ds}\right)^2G+\chi\frac{d^2\chi}{ds^2}G+2\chi\frac{d\chi}{ds}\frac{\partial G}{\partial s}\right]\right\}=0\end{aligned} \tag{C2}$$

and

$$\begin{aligned}&k^2(1+\alpha^2-\lambda)H+k\alpha\left(\frac{\kappa}{h}H-2k\frac{\partial H}{\partial\xi}\right)+k^2\frac{\partial^2H}{\partial\xi^2}\\ &-\frac{k\kappa}{h}\frac{\partial H}{\partial\xi}+\frac{\kappa'\xi}{kh^3}\left[k^{1/2}\frac{d\chi}{ds}\left(\frac{\chi}{2}H+G\right)+\frac{\partial H}{\partial s}\right]\\ &+\frac{1}{h^2}\left\{k\left(\frac{\chi}{2}\frac{d\chi}{ds}\right)^2H+\frac{\partial^2H}{\partial s^2}+k^{1/2}\left(\frac{d^2\chi}{ds^2}G+2\frac{d\chi}{ds}\frac{\partial G}{\partial s}\right)\right.\\ &\left.+\frac{k^{1/2}}{2}\left[\chi\frac{d^2\chi}{ds^2}H-(2m-1)\left(\frac{d\chi}{ds}\right)^2H+2\chi\frac{d\chi}{ds}\frac{\partial H}{\partial s}\right]\right\}=0,\end{aligned} \tag{C3}$$

where h is given by(A2). From (2.11), (4.13), and (4.14), the boundary conditions are

$$\left.\frac{\partial G}{\partial\xi}\right|_{\xi=0}=0,\quad\left.\frac{\partial H}{\partial\xi}\right|_{\xi=0}=0. \tag{C4}$$

From the terms of order k^2, $k^{3/2}$, and k in (C2) and (C3), and the boundary conditions (C4), it is found, using (4.18), that (4.16) and (4.17) hold. Next, from the terms of order $k^{1/2}$ in (C2) and (C3), it is found that (4.17) holds for $r=3$, that $p_0(s)$ satisfies (4.19), and that

$$\chi\frac{d\chi}{ds}\frac{dq_0}{ds}+\left[\frac{\chi}{2}\frac{d^2\chi}{ds^2}-(m-\tfrac{1}{2})\left(\frac{d\chi}{ds}\right)^2-\lambda_{3/2}\right]q_0(s) =-\left(2\frac{d\chi}{ds}\frac{dp_0}{ds}+\frac{d^2\chi}{ds^2}p_0(s)\right). \tag{C5}$$

From (4.20) and (4.21) it follows that, for small s, the left-hand side of Eq. (C5) is approximately $2\delta\,d/ds\times[sq_0(s)]$, so that $q_0(s)$ is determined uniquely by the requirement that $q_0(0)$ be finite. It is readily verified, with the help of (4.19), that a particular solution of (C5) is $\tilde{q}_0(s)=-p_0(s)/\chi(s)$, but this does not remain finite at $s=0$. However, it may also be verified that $[p_0(s)\chi\,d\chi/ds]^{-1}$ is a solution of the homogeneous equation, and thus, from (4.20),

$$q_0(s)=\frac{p_0(0)}{\chi(s)}\left(\frac{(2\delta)^{1/2}p_0(0)}{p_0(s)\,d\chi/ds}-\frac{p_0(s)}{p_0(0)}\right). \tag{C6}$$

From the terms of order 1 in (C2), and the boundary condition (C4), it is found, using (4.18), that

$$G_2=p_2(s)+\tfrac{1}{4}\xi^2\kappa(s)[2\kappa(0)-\kappa(s)]p_0(s), \tag{C7}$$

and

$$\begin{aligned}&\chi\frac{d\chi}{ds}\frac{dp_{1/2}}{ds}+\left[(m+\tfrac{1}{2})\left(\frac{d\chi}{ds}\right)^2+\frac{\chi}{2}\frac{d^2\chi}{ds^2}+\lambda_{3/2}\right]p_{1/2}(s)\\ &=\frac{d^2p_0}{ds^2}+\left\{\tfrac{1}{2}\kappa(s)[2\kappa(0)-\kappa(s)]-\lambda_2\right\}p_0(s)\\ &\quad-m\left(\frac{d^2\chi}{ds^2}q_0(s)+2\frac{d\chi}{ds}\frac{dq_0}{ds}\right).\end{aligned} \tag{C8}$$

Substituting $p_{1/2}(s)=r_{1/2}(s)p_0(s)$, the left-hand side of (C8) is equal to $\chi\,d\chi/ds\,dr_{1/2}/ds\,p_0(s)$, in view of (4.19). In order than $r_{1/2}(s)$ remain bounded for small s, it follows from (4.20) that

$$\lambda_2=\frac{1}{p_0(0)}\frac{d^2p_0}{ds^2}(0)+\frac{\kappa_0^2}{2}-\frac{m}{p_0(0)}\left(\frac{d^2\chi}{ds^2}(0)q_0(0)+2\frac{d\chi}{ds}(0)\frac{dq_0}{ds}(0)\right). \tag{C9}$$

[1] L. O. Wilson and J. A. Morrison, J. Math. Phys. 16, 1795 (1975).
[2] J. B. Keller and F. C. Karal, Jr., J. Appl. Phys. 31, 1039 (1960).
[3] R. Grimshaw, J. Inst. Math. Appl. 4, 174 (1968).
[4] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D.C., 1964), pp. 686–700.
[5] Ref. 4, p. 775.
[6] Ref. 4, pp. 721–743.
[7] R. Y. S. Lynn and J. B. Keller, Commun. Pure Appl. Math. 23, 379 (1970).
[8] C. E. Weatherburn, *Elementary vector analysis* (Bell, London, 1949), pp. 84–86.
[9] C. E. Weatherburn, *Advanced vector analysis* (Bell, London, 1949), p. 16.

Propagation of high frequency elastic surface waves along cylinders of general cross section

L. O. Wilson and J. A. Morrison

Bell Laboratories, Murray Hill, New Jersey 07974
(Received 24 March 1975)

The propagation of high frequency elastic surface waves along the generators of a homogeneous isotropic cylinder which has a cross-sectional boundary of nonconstant curvature is investigated. The boundary surface is stress-free and the surface waves, or Rayleigh waves, are disturbances whose amplitudes decay rapidly with depth into the cylinder. In the case of an open boundary curve for which the curvature attains its algebraic maximum at a single point, it is found that modes exist for which the disturbance is essentially confined to a region in the neighborhood of the point of maximum curvature, as well as to the neighborhood of the surface. The amplitude of the disturbance decays rapidly on either side of the point of maximum curvature. The application of these high frequency asymptotic results to the case of a closed boundary curve is discussed. Particular cases will be investigated in more detail in a subsequent paper.

1. INTRODUCTION

Elastic surface waves, or Rayleigh waves, are disturbances which travel over the stress-free surface of an elastic solid and whose amplitudes decay rapidly with depth into the solid. We investigate here the propagation of high frequency elastic surface waves down a homogeneous isotropic cylinder which has a cross-sectional boundary of nonconstant curvature. The boundary curve may be open or closed. Rosenfeld and Keller[1] treated a similar problem, but their analysis does not cover the type of modes we discuss in this paper. Gregory[2] also has considered the propagation of high frequency Rayleigh waves over curved surfaces, but his formalism does not apply when the waves travel down the generators of a cylinder whose cross-sectional boundary has nonconstant curvature.

The analysis involves a scalar wave equation, a vector wave equation, and rather complicated boundary conditions. Since the analysis tends to become cumbersome algebraically, a simpler scalar "model problem" was first studied by one of the authors (J. A. M.).[3] The analytical techniques developed for that problem have counterparts for the more complicated elastic surface wave problem to be treated here.

In the next section, we formulate the problem in terms of the coordinate system depicted in Fig. 1 of Ref. 3, where n measures distance from the surface along the inward normal and s is signed arc length along the cross-sectional boundary. Using appropriate units for length and frequency, we shall be interested only in the region $n=O(\omega^{-1})$, where ω is the frequency, and shall perform a boundary layer analysis, with $\xi=\omega n$.

The displacement $\mathbf{u}$ is assumed to be given in terms of a scalar potential and a vector potential: $\mathbf{u}=\nabla\varphi+\nabla\times\mathbf{A}$. We seek solutions to the resulting scalar and vector wave equations of the form

$$\varphi=\exp[-i\beta z-\alpha_L\xi+\omega^{1/2}\psi_L(s)]g(\xi,s;\omega),$$

$$\mathbf{A}=\exp[-i\beta z-\alpha_T\xi+\omega^{1/2}\psi_T(s)]\mathbf{C}(\xi,s;\omega),$$

where g, $\mathbf{C}$, and β are expanded in inverse powers of $\omega^{1/2}$, and α_L and α_T are positive numbers.

In Sec. 3, the lowest order mode is analyzed for the case of an open boundary curve for which the curvature $\kappa(s)$ attains its algebraic maximum at a single point, $s=0$, with $\kappa'(0)=0$ and $\kappa''(0)<0$. It is shown that $\psi_L(s)=\psi_T(s)\equiv\psi(s)$ and that, for small s,

$$\psi(s)\sim-\tfrac{1}{2}\delta s^2,\quad \delta=[-\tfrac{1}{2}\gamma\kappa''(0)]^{1/2},$$

where $\gamma>0$ is a constant given by (3.18), where α_L, α_T, and β_0 satisfy (3.3) and (3.5). Furthermore, $\psi(s)\leq\psi(0)=0$, and $\psi(s)\to-\infty$ as $|s|\to\infty$. Hence the disturbance is essentially confined to the region $|s|=O(\omega^{-1/4})$, as well as being confined close to the surface in the region $n=O(\omega^{-1})$.

An inductive procedure is developed, which allows us to determine the expansions of g, $\mathbf{C}$, and β to any order desired. The solution is compared to that for Rayleigh waves on a plane infinite half-space. The first five terms in the expansion of β in inverse powers of $\omega^{1/2}$ are also obtained. As we shall see, the first term β_0 in the expansion of β satisfies the classical secular equation (3.5) for Rayleigh waves on an infinite half-space. We will find that if we retain terms of orders ω and $\omega^{1/2}$ in the expression for the displacement, we have

$$\mathbf{u}=\omega b^{(0)}(s)\exp[-i\beta z+\omega^{1/2}\psi(s)]\Big[\mathbf{n}[1+\omega^{-1/2}c^{(1/2)}(s)]\left(-\alpha_L\exp(-\alpha_L\xi)+\frac{(\beta_0^2+\alpha_T^2)}{2\alpha_T}\exp(-\alpha_T\xi)\right)$$
$$+\left(-i\beta_0\mathbf{k}[1+\omega^{-1/2}c^{(1/2)}(s)]+\mathbf{t}\omega^{-1/2}\frac{d\psi}{ds}\right)\left(\exp(-\alpha_L\xi)-\frac{(\beta_0^2+\alpha_T^2)}{2\beta_0^2}\exp(-\alpha_T\xi)\right)\Big]. \tag{1.1}$$

where $b^{(0)}(s)$, $c^{(1/2)}(s)$, and $\psi(s)$ are given by (3.25), (3.30), and (3.19). On the other hand, the solution for Rayleigh waves travelling on the surface of a plane infinite half-space is[4]

$$\mathbf{u}=B\exp(-i\beta_0 z)\ \mathbf{n}\left(-\alpha_L\exp(-\alpha_L\xi)+\frac{(\beta_0^2+\alpha_T^2)}{2\alpha_T}\exp(-\alpha_T\xi)\right)$$
$$-i\beta_0\mathbf{k}\left(\exp(-\alpha_L\xi)-\frac{(\beta_0^2+\alpha_T^2)}{2\beta_0^2}\exp(-\alpha_T\xi)\right), \tag{1.2}$$

where B is a constant. Thus the first order effect of curvature is to multiply the plane infinite half-space solution by a factor proportional to $b^{(0)}(s)\exp[\omega^{1/2}\psi(s)]$. The second order effect is to multiply this solution by an additional factor and to add a t component of displacement which is related to the first order k component of displacement.

In a subsequent paper, we intend to investigate the behavior of the approximation (1.1) in greater detail for specific cylinders of physical interest. We will then emphasize the physical implications of our analysis and will stress less strongly the mathematical aspects which we treat here.

Higher order modes are investigated in Sec. 4. Since the lowest order mode is confined to the region $|s| = O(\omega^{-1/4})$, we introduce the stretched variable $\zeta = \omega^{1/4}s$. We look for solutions of the form

$$\varphi = \exp(-i\beta z - \alpha_L \xi)\, h(\xi, \zeta; \omega),$$

$$\mathbf{A} = \exp(-i\beta z - \alpha_T \xi)\mathbf{E}(\xi, \zeta; \omega).$$

In general, h, $\mathbf{E}$, and β would be expanded in inverse powers of $\omega^{1/4}$. Because of the algebraic complexity this involves, we make the further assumption that the boundary curve is symmetric, i.e., $\kappa(s) = \kappa(-s)$. It is then possible to expand the unknown functions in inverse powers of $\omega^{1/2}$. The solutions are expressed in terms of parabolic cylinder functions,[5] or, equivalently, in terms of Hermite polynomials.[6] It is shown that

$$\beta = \omega\beta_0 + \beta_1 + \omega^{-1/2}\beta_{3/2} + \cdots,$$

where β_0 satisfies the classical secular equation for Rayleigh waves and

$$2\beta_0\beta_1 = \gamma\kappa(0), \quad 2\beta_0\beta_{3/2} = -(2m+1)\delta.$$

The case $m = 0$ corresponds to the lowest order mode considered in Sec. 3.

Finally, in Sec. 5, we turn our attention to closed boundary curves. The region under consideration may be either interior to the boundary cylinder, corresponding to a rod, or exterior to it, corresponding to a bore. Conclusions analogous to those for the scalar model problem are reached.

2. FORMULATION

The coordinate system is described in Sec. 2 of Ref. 3. The displacement can be written as $\mathbf{u} = \nabla\varphi + \nabla\times\mathbf{A}$, where φ and A satisfy[7]

$$\nabla^2\varphi + (\omega/c_L)^2\varphi = 0, \tag{2.1}$$

$$\nabla(\nabla\cdot\mathbf{A}) - \nabla\times\nabla\times\mathbf{A} + (\omega/c_T)^2\mathbf{A} = 0, \tag{2.2}$$

$$\nabla\cdot\mathbf{A} = 0. \tag{2.3}$$

A time dependence $\exp(i\omega t)$ has been assumed. The longitudinal and transverse velocities are given by $c_L = [(\lambda+2\mu)/\rho]^{1/2}$, $c_T = (\mu/\rho)^{1/2}$, where λ and μ are Lamé's constants and ρ is the density.

That the surface of the cylinder is stress-free is expressed by the boundary conditions

$$-\omega^2\left(\frac{1}{c_T^2} - \frac{2}{c_L^2}\right)\varphi\mathbf{n} - \frac{\omega^2}{c_T^2}\mathbf{A}\times\mathbf{n} + 2\frac{\partial}{\partial n}(\nabla\varphi + \nabla\times\mathbf{A}) = 0 \quad \text{at } n = 0. \tag{2.4}$$

These are obtained from the boundary conditions for[8] $\mathbf{u}$ by using (2.1)–(2.3).

We are interested in high frequency waves, for which the wavelength is small compared to the smallest radius of curvature of the cylinder. The propagation of Rayleigh waves over curved surfaces at high frequency was studied by Gregory,[2] who assumed an asymptotic expansion essentially of the form

$$\varphi(\mathbf{r}, \omega) = \exp[i\omega s(\mathbf{r})]\sum_{m=0}^{\infty}\varphi_m(\mathbf{r})\omega^{-m},$$

$$\mathbf{A}(\mathbf{r}, \omega) = \exp[i\omega S(\mathbf{r})]\sum_{m=0}^{\infty}\mathbf{A}_m(\mathbf{r})\omega^{-m},$$

where s, S, φ_m, and $\mathbf{A}_m$ are complex functions of position. One may show that Gregory's formalism cannot be used to describe the propagation of Rayleigh waves along the generators of cylinders of nonconstant curvature.

Proceeding by analogy with the treatment of the scalar problem,[3] we conjecture that a suitable expansion for the lowest order mode is

$$\varphi(n, s, z; \omega) = \exp[-i\beta z - \omega\alpha_L n + \omega^{1/2}\psi_L(s)]f(n, s; \omega),$$

$$\mathbf{A}(n, s, z; \omega) = \exp[-i\beta z - \omega\alpha_T n + \omega^{1/2}\psi_T(s)]\mathbf{B}(n, s; \omega),$$

where

$$f(n, s; \omega) = \sum_{m=0}^{\infty}\omega^{-m/2}f^{(m/2)}(n, s),$$

$$\mathbf{B}(n, s; \omega) = \sum_{m=0}^{\infty}\omega^{-m/2}\mathbf{B}^{(m/2)}(n, s), \tag{2.5}$$

and where α_L and α_T are positive numbers. We are thus seeking waves which propagate in the z direction without attenuation and which decay away from the surface of the cylinder. The propagation constant is also a function of ω and is assumed to take the form

$$\beta = \omega\sum_{m=0}^{\infty}\omega^{-m/2}\beta_{m/2}. \tag{2.6}$$

It is indeed possible to carry through the analysis under the above assumptions. However, since we are interested only in the region $n = O(\omega^{-1})$, our task is simplified if we restrict ourselves to a boundary layer analysis. For this purpose, we set

$$\xi = \omega n, \quad f(n, s; \omega) = g(\xi, s; \omega), \quad \mathbf{B}(n, s; \omega) = \mathbf{C}(\xi, s; \omega), \tag{2.7}$$

so that

$$\varphi = \exp[-i\beta z - \alpha_L\xi + \omega^{1/2}\psi_L(s)]g(\xi, s; \omega),$$

$$\mathbf{A} = \exp[-i\beta z - \alpha_T\xi + \omega^{1/2}\psi_T(s)]\mathbf{C}(\xi, s; \omega), \tag{2.8}$$

with

$$g(\xi, s; \omega) = \sum_{m=0}^{\infty}\omega^{-m/2}g^{(m/2)}(\xi, s),$$

$$\mathbf{C}(\xi, s; \omega) = \sum_{m=0}^{\infty} \omega^{-m/2} \mathbf{C}^{(m/2)}(\xi, s). \tag{2.9}$$

Then the line element $h = 1 - \kappa n = 1 - \kappa\xi/\omega$, and $\partial/\partial n = \omega\partial/\partial\xi$. Upon substituting (2.8) into (2.1) and (2.2), and using the curvilinear expressions for the curl, divergence, and ∇^2,[9] we obtain the differential equations

$$\omega^2\left[\left(\alpha_L^2 + \frac{1}{c_L^2} - \frac{\beta^2}{\omega^2}\right) g - 2\alpha_L \frac{\partial g}{\partial \xi} + \frac{\partial^2 g}{\partial \xi^2}\right]$$
$$+ \omega\left[\frac{\kappa\alpha_L g}{h} - \frac{\kappa}{h}\frac{\partial g}{\partial \xi} + \frac{1}{h^2}\left(\frac{d\psi_L}{ds}\right)^2 g\right]$$
$$+ \omega^{1/2}\left(\frac{2}{h^2}\frac{d\psi_L}{ds}\frac{\partial g}{\partial s} + \frac{1}{h^2}\frac{d^2\psi_L}{ds^2} g\right)$$
$$+ \frac{1}{h^2}\frac{\partial^2 g}{\partial s^2} + \omega^{-1/2}\frac{\kappa'\xi}{h^3}\frac{d\psi_L}{ds} g$$
$$+ \omega^{-1}\frac{\kappa'\xi}{h^3}\frac{\partial g}{\partial s} = 0,$$

$$\omega^2\left[\left(\alpha_T^2 + \frac{1}{c_T^2} - \frac{\beta^2}{\omega^2}\right) C_t - 2\alpha_T \frac{\partial C_t}{\partial \xi} + \frac{\partial^2 C_t}{\partial \xi^2}\right]$$
$$+ \omega\left[\frac{\kappa\alpha_T C_t}{h} - \frac{\kappa}{h}\frac{\partial C_t}{\partial \xi} + \frac{1}{h^2}\left(\frac{d\psi_T}{ds}\right)^2 C_t\right]$$
$$+ \omega^{1/2}\left(\frac{2}{h^2}\frac{d\psi_T}{ds}\frac{\partial C_t}{\partial s} + \frac{1}{h^2}\frac{d^2\psi_T}{ds^2} C_t - \frac{2\kappa}{h^2}\frac{d\psi_T}{ds} C_n\right)$$
$$+ \left(\frac{1}{h^2}\frac{\partial^2 C_t}{\partial s^2} - \frac{\kappa^2}{h^2} C_t - \frac{\kappa' C_n}{h^3} - \frac{2\kappa}{h^2}\frac{\partial C_n}{\partial s}\right)$$
$$+ \omega^{-1/2}\frac{\kappa'\xi}{h^3}\frac{d\psi_T}{ds} C_t + \omega^{-1}\frac{\kappa'\xi}{h^3}\frac{\partial C_t}{\partial s} = 0,$$

$$\omega^2\left[\left(\alpha_T^2 + \frac{1}{c_T^2} - \frac{\beta^2}{\omega^2}\right) C_n - 2\alpha_T \frac{\partial C_n}{\partial \xi} + \frac{\partial^2 C_n}{\partial \xi^2}\right]$$
$$+ \omega\left[\frac{\kappa\alpha_T C_n}{h} - \frac{\kappa}{h}\frac{\partial C_n}{\partial \xi} + \frac{1}{h^2}\left(\frac{d\psi_T}{ds}\right)^2 C_n\right]$$
$$+ \omega^{1/2}\left(\frac{2}{h^2}\frac{d\psi_T}{ds}\frac{\partial C_n}{\partial s} + \frac{1}{h^2}\frac{d^2\psi_T}{ds^2} C_n + \frac{2\kappa}{h^2}\frac{d\psi_T}{ds} C_t\right)$$
$$+ \left(\frac{1}{h^2}\frac{\partial^2 C_n}{\partial s^2} - \frac{\kappa^2}{h^2} C_n + \frac{\kappa' C_t}{h^3} + \frac{2\kappa}{h^2}\frac{\partial C_t}{\partial s}\right)$$
$$+ \omega^{-1/2}\frac{\kappa'\xi}{h^3}\frac{d\psi_T}{ds} C_n + \omega^{-1}\frac{\kappa'\xi}{h^2}\frac{\partial C_n}{\partial s} = 0,$$

$$\omega^2\left[\left(\alpha_T^2 + \frac{1}{c_T^2} - \frac{\beta^2}{\omega^2}\right) C_k - 2\alpha_T \frac{\partial C_k}{\partial \xi} + \frac{\partial^2 C_k}{\partial \xi^2}\right]$$
$$+ \omega\left[\frac{\kappa\alpha_T C_k}{h} - \frac{\kappa}{h}\frac{\partial C_k}{\partial \xi} + \frac{1}{h^2}\left(\frac{d\psi_T}{ds}\right)^2 C_k\right]$$
$$+ \omega^{1/2}\left(\frac{2}{h^2}\frac{d\psi_T}{ds}\frac{\partial C_k}{\partial s} + \frac{1}{h^2}\frac{d^2\psi_T}{ds^2} C_k\right) + \frac{1}{h^2}\frac{\partial^2 C_k}{\partial s^2}$$
$$+ \omega^{-1/2}\frac{\kappa'\xi}{h^3}\frac{d\psi_T}{ds} C_k + \omega^{-1}\frac{\kappa'\xi}{h^3}\frac{\partial C_k}{\partial s} = 0. \tag{2.10}$$

We next substitute (2.8) into the boundary conditions (2.4) and the divergence condition (2.3) to get

$$\psi_L(s) = \psi_T(s) = \psi(s), \tag{2.11}$$

$$\omega^2\left[\left(\alpha_L^2 + \frac{1}{c_L^2} - \frac{1}{2c_T^2}\right) g - 2\alpha_L \frac{\partial g}{\partial \xi} + \frac{\partial^2 g}{\partial \xi^2}\right.$$
$$\left. - \frac{i\beta\alpha_T}{\omega} C_t + \frac{i\beta}{\omega}\frac{\partial C_t}{\partial \xi}\right] + \omega^{3/2}\frac{d\psi}{ds}\left(-\alpha_T C_k + \frac{\partial C_k}{\partial \xi}\right)$$
$$+ \omega\left[-\alpha_T\frac{\partial C_k}{\partial s} + \frac{\partial^2 C_k}{\partial \xi \partial s}\right] + \omega^{1/2}\kappa\frac{d\psi}{ds} C_k + \kappa\frac{\partial C_k}{\partial s} = 0$$
$$\text{at } \xi = 0, \tag{2.12}$$

$$\omega^2\left[\left(\alpha_T^2 + \frac{1}{2c_T^2}\right) C_t - 2\alpha_T\frac{\partial C_t}{\partial \xi} + \frac{\partial^2 C_t}{\partial \xi^2} + \frac{i\beta\alpha_L}{\omega} g - \frac{i\beta}{\omega}\frac{\partial g}{\partial \xi}\right]$$
$$+ \omega^{3/2}\frac{d\psi}{ds}\left(\alpha_T C_n - \frac{\partial C_n}{\partial \xi}\right)$$
$$+ \omega\left(\alpha_T\frac{\partial C_n}{\partial s} - \frac{\partial^2 C_n}{\partial \xi \partial s} + \alpha_T\kappa C_t - \kappa\frac{\partial C_t}{\partial \xi}\right) - \omega^{1/2}\kappa\frac{d\psi}{ds} C_n$$
$$- \left(\kappa^2 C_t + \kappa\frac{\partial C_n}{\partial s}\right) = 0 \quad \text{at } \xi = 0, \tag{2.13}$$

$$\omega^2\left[-\left(\alpha_T^2 + \frac{1}{2c_T^2}\right) C_k + 2\alpha_T\frac{\partial C_k}{\partial \xi} - \frac{\partial^2 C_k}{\partial \xi^2} + \frac{i\beta\alpha_T}{\omega} C_n\right.$$
$$\left. - \frac{i\beta}{\omega}\frac{\partial C_n}{\partial \xi}\right] + \omega^{3/2}\frac{d\psi}{ds}\left(-\alpha_L g + \frac{\partial g}{\partial \xi}\right) + \omega\left(-\alpha_L\frac{\partial g}{\partial s} + \frac{\partial^2 g}{\partial \xi \partial s}\right)$$
$$+ \omega^{1/2}\kappa\frac{d\psi}{ds} g + \kappa\frac{\partial g}{\partial s} = 0 \quad \text{at } \xi = 0, \tag{2.14}$$

$$\omega\left(\alpha_T C_n - \frac{\partial C_n}{\partial \xi} + \frac{i\beta}{\omega} C_k\right) - \omega^{1/2}\frac{d\psi}{ds} C_t + \kappa C_n - \frac{\partial C_t}{ds} = 0$$
$$\text{at } \xi = 0. \tag{2.15}$$

Although the divergence condition holds throughout the medium, it will suffice to apply it at the boundary $\xi = 0$.

3. ANALYSIS OF THE LOWEST ORDER MODE

Substitute (2.6) and (2.9) in the differential equations (2.10) and extract terms of order ω^2 to obtain

$$\left(\alpha_L^2 + \frac{1}{c_L^2} - \beta_0^2\right) g^{(0)} - 2\alpha_L\frac{\partial g^{(0)}}{\partial \xi} + \frac{\partial^2 g^{(0)}}{\partial \xi^2} = 0,$$

$$\left(\alpha_T^2 + \frac{1}{c_T^2} - \beta_0^2\right) \mathbf{C}^{(0)} - 2\alpha_T\frac{\partial \mathbf{C}^{(0)}}{\partial \xi} + \frac{\partial^2 \mathbf{C}^{(0)}}{\partial \xi^2} = 0. \tag{3.1}$$

It follows from assumptions (2.5) and (2.7) that $g^{(0)}$ and $\mathbf{C}^{(0)}$ are independent of ξ. This can be seen by expanding $f^{(0)}(n,s)=f^{(0)}(\xi/\omega,s)$ in inverse powers of ω. So

$$g^{(0)}(\xi,s)=b^{(0)}(s), \quad \mathbf{C}^{(0)}(\xi,s)=\mathbf{F}^{(0)}(s). \tag{3.2}$$

We wish the first terms in the expansion of φ and $\mathbf{A}$ to be nontrivial, i.e., that $b^{(0)}\neq 0$ and that at least one component of $\mathbf{F}^{(0)}\neq 0$. It then follows from (3.1) and (3.2) that

$$\alpha_L^2+\frac{1}{c_L^2}-\beta_0^2=0, \quad \alpha_T^2+\frac{1}{c_T^2}-\beta_0^2=0. \tag{3.3}$$

Next, we use (3.2), (3.3), and the expansions (2.5) and (2.7) in the first two boundary conditions (2.12) and (2.13). Setting the terms of order ω^2 to zero gives

$$\begin{pmatrix} \beta_0^2-1/2c_T^2 & -i\beta_0\alpha_T \\ i\beta_0\alpha_L & \beta_0^2-1/2c_T^2 \end{pmatrix}\begin{pmatrix} b^{(0)} \\ F_t^{(0)} \end{pmatrix}=0. \tag{3.4}$$

Since at least $b^{(0)}$ is nontrivial, the determinant of the 2×2 matrix in (3.4) must be zero:

$$\left(\beta_0^2-\frac{1}{2c_T^2}\right)^2-\beta_0^2\alpha_T\alpha_L=0. \tag{3.5}$$

Equations (3.3) determine α_T and α_L. When combined with (3.5), they yield the familiar secular equation for Rayleigh waves on a plane infinite half space. The secular equation determines β_0.

Using the expansions again and considering terms of order ω^2 in the third boundary condition (2.14) and of order ω in the divergence condition (2.15), we find

$$\begin{pmatrix} i\beta_0\alpha_T & -(\beta_0^2-1/2c_T^2) \\ \alpha_T & i\beta_0 \end{pmatrix}\begin{pmatrix} F_n^{(0)} \\ F_k^{(0)} \end{pmatrix}=0.$$

Since $\alpha_T>0$, this equation implies that

$$F_n^{(0)}=F_k^{(0)}=0. \tag{3.6}$$

Turning now to terms of order $\omega^{3/2}$ in the differential equations (2.10) and using (3.2), (3.3) and (3.6), we obtain

$$\frac{\partial^2 g^{(1/2)}}{\partial\xi^2}-2\alpha_L\frac{\partial g^{(1/2)}}{\partial\xi}-2\beta_0\beta_{1/2}b^{(0)}=0,$$

$$\frac{\partial^2 C_t^{(1/2)}}{\partial\xi^2}-2\alpha_T\frac{\partial C_t^{(1/2)}}{\partial\xi}-2\beta_0\beta_{1/2}F_t^{(0)}=0,$$

$$\frac{\partial^2 C_n^{(1/2)}}{\partial\xi^2}-2\alpha_T\frac{\partial C_n^{(1/2)}}{\partial\xi}=0,$$

$$\frac{\partial^2 C_k^{(1/2)}}{\partial\xi^2}-2\alpha_T\frac{\partial C_k^{(1/2)}}{\partial\xi}=0.$$

Upon rejecting solutions which grow exponentially with ξ, we have

$$g^{(1/2)}(\xi,s)=b^{(1/2)}(s)-(\beta_0\beta_{1/2}\xi/\alpha_L)\,b^{(0)}(s),$$
$$C_t^{(1/2)}(\xi,s)=F_t^{(1/2)}(s)-(\beta_0\beta_{1/2}\xi/\alpha_T)\,F_t^{(0)}(s),$$
$$C_n^{(1/2)}(\xi,s)=F_n^{(1/2)}(s),$$
$$C_k^{(1/2)}(\xi,s)=F_k^{(1/2)}(s). \tag{3.7}$$

From (3.3), (3.6), (3.7), and the expansions (2.5) and (2.7) applied to the first two boundary conditions (2.12) and (2.13), it follows, upon setting terms of order $\omega^{3/2}$ to zero, that

$$\begin{pmatrix} \beta_0^2-1/2c_T^2 & -i\beta_0\alpha_T \\ i\beta_0\alpha_L & \beta_0^2-1/2c_T^2 \end{pmatrix}\begin{pmatrix} b^{(1/2)} \\ F_t^{(1/2)} \end{pmatrix}$$
$$+\beta_{1/2}\begin{pmatrix} 2\beta_0 & -i(\alpha_T+\beta_0^2/\alpha_T) \\ i(\alpha_L+\beta_0^2/\alpha_L) & 2\beta_0 \end{pmatrix}\begin{pmatrix} b^{(0)} \\ F_t^{(0)} \end{pmatrix}=0. \tag{3.8}$$

We next premultiply (3.8) by $((\beta_0^2-1/2c_T^2)\ \ i\beta_0\alpha_T)$. The first term vanishes because of the secular equation (3.5). Then with the help of (3.3)–(3.5) we find that

$$\beta_{1/2}\left[2\beta_0\left(\beta_0^2-\frac{1}{2c_T^2}\right)+2\beta_0\left(\beta_0^2-\frac{1}{2c_T^2}\right)\right.$$
$$\left.-\beta_0\alpha_L\left(\alpha_T+\frac{\beta_0^2}{\alpha_T}\right)-\beta_0\alpha_T\left(\alpha_L+\frac{\beta_0^2}{\alpha_L}\right)\right]$$
$$=-\beta_0\beta_{1/2}\left[\frac{\beta_0^2(\alpha_L-\alpha_T)^2}{\alpha_L\alpha_T}+2\alpha_T(\alpha_L-\alpha_T)\right]=0,$$

and thus, since $\alpha_L-\alpha_T>0$,

$$\beta_{1/2}=0. \tag{3.9}$$

It follows from (3.7) that

$$g^{(1/2)}(\xi,s)=b^{(1/2)}(s), \quad C_t^{(1/2)}(\xi,s)=F_t^{(1/2)}(s). \tag{3.10}$$

This is consistent with assumptions (2.5) and (2.7), which also imply that $g^{(1/2)}$ and $\mathbf{C}^{(1/2)}$ are independent of ξ. From (3.8) we find that

$$\begin{pmatrix} \beta_0^2-1/2c_T^2 & -i\beta_0\alpha_T \\ i\beta_0\alpha_L & \beta_0^2-1/2c_T^2 \end{pmatrix}\begin{pmatrix} b^{(1/2)} \\ F_t^{(1/2)} \end{pmatrix}=0. \tag{3.11}$$

Now consider terms of order $\omega^{3/2}$ in the third boundary condition (2.14) and of order $\omega^{1/2}$ in the divergence condition (2.15). With the aid of (3.2), (3.7), (3.10), and the expansions (2.6) and (2.9), these yield

$$\begin{pmatrix} i\beta_0\alpha_T & -(\beta_0^2-1/2c_T^2) \\ \alpha_T & i\beta_0 \end{pmatrix}\begin{pmatrix} F_n^{(1/2)} \\ F_k^{(1/2)} \end{pmatrix}$$
$$=\frac{d\psi}{ds}\begin{pmatrix} \alpha_L & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} b^{(0)} \\ F_t^{(0)} \end{pmatrix}. \tag{3.12}$$

We then use (3.4) to get

$$F_n^{(1/2)}=0, \quad F_k^{(1/2)}=\frac{1}{i\beta_0}\frac{d\psi}{ds}F_t^{(0)}. \tag{3.13}$$

Before proceeding to the analysis of terms of lower order in ω, we use the expansions (2.6) and (2.9) in the differential equations, boundary conditions, and di-

vergence condition to write down general forms of these equations. Our preliminary analysis has led to certain results, such as $\beta_{1/2}=0$, which simplify the writing of the general forms, which appear in the Appendix. The differential equations appear in (A1), and the boundary conditions and divergence condition appear in (A6) and (A7). The notation used in these latter equations is introduced in (A2).

We next proceed to the analysis of terms of order ω, i.e., with $p=2$, according to the notation of the Appendix. Using (3.2), (3.6), and (A1), and rejecting solutions which grow exponentially with ξ, we get

$$g^{(1)}(\xi,s)=b^{(1)}(s)+\frac{b^{(0)}(s)\xi}{2\alpha_L}\left[\kappa\alpha_L+\left(\frac{d\psi}{ds}\right)^2-2\beta_0\beta_1\right],$$

$$C_t^{(1)}(\xi,s)=F_t^{(1)}(s)+\frac{F_t^{(0)}(s)\xi}{2\alpha_T}\left[\kappa\alpha_T+\left(\frac{d\psi}{ds}\right)^2-2\beta_0\beta_1\right],$$

$$C_n^{(1)}(\xi,s)=F_n^{(1)}(s),\quad C_k^{(1)}(\xi,s)=F_k^{(1)}(s). \tag{3.14}$$

Then from the first two boundary conditions (A6) and from (3.14), it follows (for $p=2$) that

$$\begin{pmatrix}\beta_0^2-1/2c_T^2 & -i\beta_0\alpha_T\\ i\beta_0\alpha_L & \beta_0^2-1/2c_T^2\end{pmatrix}\begin{pmatrix}b^{(1)}\\ F_t^{(1)}\end{pmatrix}$$

$$+\begin{pmatrix}0 & -i\beta_1\alpha_T\\ i\beta_1\alpha_L & \kappa\alpha_T\end{pmatrix}\begin{pmatrix}b^{(0)}\\ F_t^{(0)}\end{pmatrix}+\begin{pmatrix}-1 & i\beta_0/2\alpha_T\\ -i\beta_0/2\alpha_L & -1\end{pmatrix}$$

$$\times\begin{pmatrix}[\kappa\alpha_L+(d\psi/ds)^2-2\beta_0\beta_1] & 0\\ 0 & [\kappa\alpha_T+(d\psi/ds)^2-2\beta_0\beta_1]\end{pmatrix}$$

$$\times\begin{pmatrix}b^{(0)}\\ F_t^{(0)}\end{pmatrix}+\alpha_T\frac{d\psi}{ds}\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}\begin{pmatrix}F_n^{(1/2)}\\ F_k^{(1/2)}\end{pmatrix}=0. \tag{3.15}$$

Premultiplication of the entire equation by $(\beta_0^2-1/2c_T^2\ \ i\beta_0\alpha_T)$ causes the first term to drop out because of the secular equation (3.5). The last term can be evaluated by means of (3.12), or alternatively, by means of the general equation (A7) with $p=1$. Upon doing this all, and using (3.4), we find that

$$\left[\left(\frac{d\psi}{ds}\right)^2-2\beta_0\beta_1\right]\left[\frac{\beta_0^2(\alpha_L-\alpha_T)^2}{2\alpha_L\alpha_T}+\alpha_T(\alpha_L-\alpha_T)\right]$$

$$+\frac{\kappa\alpha_T}{2}(\beta_0^2-\alpha_L\alpha_T)=0. \tag{3.16}$$

It follows that

$$\left(\frac{d\psi}{ds}\right)^2=2\beta_0\beta_1-\gamma\kappa(s),$$

$$\gamma=\frac{\alpha_L\alpha_T^2(\beta_0^2-\alpha_L\alpha_T)}{\beta_0^2(\alpha_L-\alpha_T)^2+2\alpha_T^2\alpha_L(\alpha_L-\alpha_T)}>0. \tag{3.17}$$

Now suppose that the boundary curve of the cylinder is open and that $\kappa(s)$ attains its algebraic maximum at a single point, which we define to be $s=0$. We assume that $\kappa''(0)<0$ and that $\kappa(s)$ is bounded away from $\kappa(0)$ as $|s|\to\infty$. Then, by following the argument of Morrison,[3] we have

$$2\beta_0\beta_1=\gamma\kappa(0), \tag{3.18}$$

with $\operatorname{sgn}(d\psi/ds)=-\operatorname{sgn}s$. With $\psi(0)=0$, we have

$$\psi(s)=-\int_0^s\{\gamma[\kappa(0)-\kappa(\sigma)]\}^{1/2}\operatorname{sgn}\sigma\,d\sigma, \tag{3.19}$$

where the positive square root is to be taken. Note that $\psi(s)\to-\infty$ as $|s|\to\infty$, so that $\exp[\omega^{1/2}\psi(s)]\to 0$ as $|s|\to\infty$. Furthermore, for small s we have

$$\psi(s)\sim-\tfrac{1}{2}\delta s^2,\quad |s|\ll 1,\quad \delta=[-\tfrac{1}{2}\gamma\kappa''(0)]^{1/2}, \tag{3.20}$$

so that on the boundary curve the solution decays rapidly on either side of the point of maximum algebraic curvature. The disturbance is thus essentially confined to the region $|s|=O(\omega^{-1/4})$, as well as being confined close to the surface in the region $n=O(\omega^{-1})$.

The case in which the boundary curve of the cylinder is closed will be discussed later.

Next consider terms of order ω in the third boundary condition and of order 1 in the divergence condition. By setting $p=2$ in the general equation (A7) and using (3.6) and (3.14), we get

$$\begin{pmatrix}i\beta_0\alpha_T & -(\beta_0^2-1/2c_T^2)\\ \alpha_T & i\beta_0\end{pmatrix}\begin{pmatrix}F_n^{(1)}\\ F_k^{(1)}\end{pmatrix}-\frac{d\psi}{ds}\begin{pmatrix}\alpha_L & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}b^{(1/2)}\\ F_t^{(1/2)}\end{pmatrix}$$

$$-\frac{d}{ds}\begin{pmatrix}\alpha_L & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}b^{(0)}\\ F_t^{(0)}\end{pmatrix}=0. \tag{3.21}$$

Thus, upon completion of the analysis of terms of order ω or higher in the differential equations and boundary conditions, and of terms of order 1 and higher in the divergence condition, we have determined β_0, $\beta_{1/2}$, β_1, and $\psi(s)$. We have yet to determine $b^{(0)}(s)$, $b^{(1/2)}(s)$, and $b^{(1)}(s)$, but know how to express all the components of $\mathbf{F}^{(p/2)}(s)$, $p=0,1,2$, in terms of these three functions.

Proceeding to terms of order $\omega^{1/2}$ (i.e., $p=3$) in the differential equations (A1), we obtain, after using (3.2), (3.6), (3.7), and (3.10), and rejecting solutions which grow exponentially with ξ,

$$g^{(3/2)}(\xi,s)=b^{(3/2)}(s)+\frac{\xi}{2\alpha_L}\left\{\left[\kappa\alpha_L+\left(\frac{d\psi}{ds}\right)^2-2\beta_0\beta_1\right]b^{(1/2)}(s)\right.$$

$$\left.+\left[2\frac{d\psi}{ds}\frac{d}{ds}+\frac{d^2\psi}{ds^2}-2\beta_0\beta_{3/2}\right]b^{(0)}(s)\right\},$$

$$C_t^{(3/2)}(\xi,s)=F_t^{(3/2)}(s)+\frac{\xi}{2\alpha_T}\left\{\left[\kappa\alpha_T+\left(\frac{d\psi}{ds}\right)^2-2\beta_0\beta_1\right]F_t^{(1/2)}(s)\right.$$

$$\left.+\left[2\frac{d\psi}{ds}\frac{d}{ds}+\frac{d^2\psi}{ds^2}-2\beta_0\beta_{3/2}\right]F_t^{(0)}(s)\right\},$$

$$C_n^{(3/2)}(\xi,s)=F_n^{(3/2)}(s)+\frac{\kappa\xi}{\alpha_T}\frac{d\psi}{ds}F_t^{(0)}(s),$$

$$C_k^{(3/2)}(\xi,s)=F_k^{(3/2)}(s)+\frac{\xi}{2\alpha_T}\left[\kappa\alpha_T+\left(\frac{d\psi}{ds}\right)^2-2\beta_0\beta_1\right]F_k^{(1/2)}(s). \tag{3.22}$$

Next, set $p=3$ in the boundary conditions (A6) and use (3.6) to get

$$\begin{pmatrix} \beta_0^2 - 1/2c_T^2 & -i\beta_0\alpha_T \\ i\beta_0\alpha_L & \beta_0^2 - 1/2c_T^2 \end{pmatrix} \begin{pmatrix} b^{(3/2)} \\ F_t^{(3/2)} \end{pmatrix}$$

$$+ \begin{pmatrix} 0 & -i\beta_1\alpha_T \\ i\beta_1\alpha_L & \kappa\alpha_T \end{pmatrix} \begin{pmatrix} b^{(1/2)} \\ F_t^{(1/2)} \end{pmatrix} + \begin{pmatrix} -1 & i\beta_0/2\alpha_T \\ -i\beta_0/2\alpha_L & -1 \end{pmatrix}$$

$$\times \begin{pmatrix} [\kappa\alpha_L + (d\psi/ds)^2 - 2\beta_0\beta_1] & 0 \\ 0 & [\kappa\alpha_T + (d\psi/ds)^2 - 2\beta_0\beta_1] \end{pmatrix}$$

$$\times \begin{pmatrix} b^{(1/2)} \\ F_t^{(1/2)} \end{pmatrix} + \left(2\frac{d\psi}{ds}\frac{d}{ds} + \frac{d^2\psi}{ds^2} - 2\beta_0\beta_{3/2}\right)$$

$$\times \begin{pmatrix} -1 & i\beta_0/2\alpha_T \\ -i\beta_0/2\alpha_L & -1 \end{pmatrix} \begin{pmatrix} b^{(0)} \\ F_t^{(0)} \end{pmatrix} + i\beta_{3/2} \begin{pmatrix} 0 & -\alpha_T \\ \alpha_L & 0 \end{pmatrix}$$

$$\times \begin{pmatrix} b^{(0)} \\ F_t^{(0)} \end{pmatrix} + \alpha_T \frac{d\psi}{ds} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} F_n^{(1)} \\ F_k^{(1)} \end{pmatrix}$$

$$+ \alpha_T \frac{d}{ds} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} F_n^{(1/2)} \\ F_k^{(1/2)} \end{pmatrix} = 0. \tag{3.23}$$

In arriving at this equation, we have, as before, used our knowledge about the functional forms of $g^{(p/2)}$ and $\mathbf{C}^{(p/2)}$, with $p=0, 1, 2, 3$. We premultiply the equation by $(\beta_0^2 - 1/2c_T^2 \;\; i\beta_0\alpha_T)$. With the aid of the secular equation (3.5), the first term drops out. The last two terms can be evaluated with the aid of (A7) or, alternatively, (3.12) and (3.21). By performing the analysis and using (3.3), (3.4), (3.12), and (3.16), we eventually obtain

$$\left(\frac{\beta_0^2(\alpha_L - \alpha_T)^2 + 2\alpha_T^2\alpha_L(\alpha_L - \alpha_T)}{2\alpha_T\alpha_L}\right)$$

$$\times \left(2\frac{d\psi}{ds}\frac{d}{ds} + \frac{d^2\psi}{ds^2} - 2\beta_0\beta_{3/2}\right) b^{(0)}(s) = 0. \tag{3.24}$$

It follows that

$$2\frac{d\psi}{ds}\frac{db^{(0)}}{ds} + \left(\frac{d^2\psi}{ds^2} - 2\beta_0\beta_{3/2}\right) b^{(0)}(s) = 0. \tag{3.25}$$

If $b^{(0)}(0) \neq 0$, then by using (3.20) we obtain

$$2\beta_0\beta_{3/2} = -\delta, \quad \delta = [-\tfrac{1}{2}\gamma\kappa''(0)]^{1/2}. \tag{3.26}$$

On the other hand, if we assume that $b^{(0)}(s) \sim s^m$ for $|s| \ll 1$, where m is a positive integer, then

$$2\beta_0\beta_{3/2} = -(2m+1)\delta, \tag{3.27}$$

which corresponds to higher order modes. These will be treated in Sec. 4 by another method. Similar results were obtained by Morrison[3] in his treatment of the scalar model problem.

The quantity $b^{(0)}(s)$ is obtained by quadrature from (3.25), with $\psi(s)$ given by (3.19) and $\beta_{3/2}$ by (3.26).

Note that in the Appendix we introduce the notation

$$X^{p/2} = \begin{pmatrix} g^{(p/2)} \\ C_t^{(p/2)} \end{pmatrix}, \quad Y^{p/2} = \begin{pmatrix} C_n^{(p/2)} \\ C_k^{(p/2)} \end{pmatrix}. \tag{A2}$$

To complete the analysis of terms of order $\omega^{1/2}$, we would need to use (A7), with $p=3$, to express $Y^{3/2}$ in terms of $Y^{1/2}$, X^1, $X^{1/2}$, and X^0. By virtue of (3.12), we then would know $Y^{3/2}$ in terms of X^1, $X^{1/2}$, and X^0. We omit the details.

Having thus finished the analysis of terms of order $\omega^{1/2}$ and higher in the differential equations and boundary conditions, and of order $\omega^{-1/2}$ and higher in the divergence condition, we have determined β_0, $\beta_{1/2}$, β_1, $\beta_{3/2}$, $\psi(s)$, X^0, Y^0, and $Y^{1/2}$. We have yet to determine $X^{3/2}$, X^1, and $X^{1/2}$, but know how the components of each are related. We know how to express $Y^{3/2}$ in terms of X^1, $X^{1/2}$, and X^0 and Y^1 in terms of $X^{1/2}$ and X^0.

At this point it is possible to proceed inductively. Consider the general differential equations, boundary conditions, and divergence condition (A1), (A6), and (A7). Assume we are at the pth step in the analysis, with $p \geq 4$. Suppose that we know β_0, $\beta_{1/2}$, $\ldots$, $\beta_{(p-1)/2}$, X^0, $X^{1/2}$, $\ldots$, $X^{(p-4)/2}$, and Y^0, $Y^{1/2}$, $\ldots$, $Y^{(p-3)/2}$ and that we can express $Y^{(p-2)/2}$ in terms of $X^{(p-3)/2}$ and these known functions, and can express $Y^{(p-1)/2}$ in terms of $X^{(p-2)/2}$, $X^{(p-3)/2}$, and these known functions. Then at the pth step we will be able to determine $\beta_{p/2}$, $X^{(p-3)/2}$, $Y^{(p-2)/2}$, and will be able to express $Y^{p/2}$ in terms of $X^{(p-1)/2}$, $X^{(p-2)/2}$, and known functions, and, of course, $Y^{(p-1)/2}$ in terms of $X^{(p-2)/2}$ and known functions. We omit the details of the inductive proof, since the procedures used are quite similar to those demonstrated in the preceding analysis. The only thing which might not be clear from examination of (A1), (A6), and (A7) is how $X^{p/2}$ and $X^{(p-2)/2}$ are both eliminated from (A6). After substitution for $Y^{(p-1)/2}$ and premultiplication of (A6) by $(\beta_0^2 - 1/2c_T^2 \;\; i\beta_0\alpha_T)$, the $X^{p/2}$ term drops out by means of the secular equation (3.5). At the same time, the term involving $X^{(p-2)/2}$ will drop out because it is multiplied by a factor which vanishes when the expression (3.17) for $(d\psi/ds)^2$ is substituted in it.

In order to complete the analysis for all terms in the expansion with superscript 1/2, we must go to the step $p=4$ in the inductive procedure. This eventually yields the rather messy equation

$$(\alpha_L - \alpha_T)[\beta_0^2(\alpha_L - \alpha_T) + 2\alpha_L\alpha_T^2]$$

$$\times \left\{\left[2\frac{d\psi}{ds}\frac{db^{(1/2)}}{ds} + \left(\frac{d^2\psi}{ds^2} - 2\beta_0\beta_{3/2}\right) b^{(1/2)}\right]\right.$$

$$\left. + \left(\frac{d^2b^{(0)}}{ds^2} - (2\beta_0\beta_2 + \beta_1^2)b^{(0)}\right)\right\} + \rho(s)b^{(0)} = 0, \tag{3.28}$$

with

$$\rho(s)=\frac{(\alpha_L-\alpha_T)^2}{4\alpha_L^2\alpha_T^2}\left[\beta_0^2(\alpha_L+\alpha_T)^2-4\alpha_L^2\alpha_T^2\right]\left[\left(\frac{d\psi}{ds}\right)^2-2\beta_0\beta_1\right]^2+\tfrac{1}{4}[\kappa(s)]^2\left[\beta_0^2(\alpha_T^2-3\alpha_L^2)+2\alpha_L\alpha_T^3\right]$$

$$+\kappa(s)\beta_0\beta_1[\alpha_L(\beta_0^2-\alpha_L\alpha_T)-\alpha_T(\alpha_L-\alpha_T)^2]+\frac{\kappa(s)}{2\alpha_L\alpha_T}\left(\frac{d\psi}{ds}\right)^2\left[\beta_0^2(2\alpha_L^3+2\alpha_T^3-3\alpha_L^2\alpha_T)+\alpha_L\alpha_T^2(6\alpha_L^2+\alpha_T^2-8\alpha_L\alpha_T)\right].$$

We now set

$$b^{(1/2)}(s)=c^{(1/2)}(s)b^{(0)}(s) \tag{3.29}$$

and use (3.25) to obtain

$$2\frac{d\psi}{ds}\frac{dc^{(1/2)}}{ds}+\frac{1}{b^{(0)}}\frac{d^2b^{(0)}}{ds^2}+\frac{\rho(s)}{\beta_0^2(\alpha_L-\alpha_T)^2+2\alpha_T^2\alpha_L(\alpha_L-\alpha_T)}$$

$$-(2\beta_0\beta_2+\beta_1^2)=0, \tag{3.30}$$

from which $c^{(1/2)}(s)$ may be determined. There is no loss of generality in taking $c^{(1/2)}(0)=0$. The trick of setting, in general, $b^{((p-3)/2)}(s)=c^{((p-3)/2)}(s)b^{(0)}(s)$ is similarly useful in the pth stage of the analysis, for $p>4$. In order to have $c^{(1/2)}(s)$ remain bounded as $s\to 0$, we must set

$$2\beta_0\beta_2=\frac{1}{b^{(0)}(0)}\frac{d^2b^{(0)}(0)}{ds^2}-\beta_1^2$$

$$+\frac{\rho(0)}{\beta_0^2(\alpha_L-\alpha_T)^2+2\alpha_L\alpha_T^2(\alpha_L-\alpha_T)}, \tag{3.31}$$

with

$$\rho(0)=\tfrac{1}{4}[\kappa(0)]^2\{[\gamma^2(\alpha_L-\alpha_T)^2/\alpha_L^2\alpha_T^2][\beta_0^2(\alpha_L+\alpha_T)^2-4\alpha_L^2\alpha_T^2]$$

$$+[\beta_0^2(\alpha_T^2-3\alpha_L^2)+2\alpha_L\alpha_T^3]$$

$$+2\gamma[\alpha_L(\beta_0^2-\alpha_L\alpha_T)-\alpha_T(\alpha_L-\alpha_T)^2]\},$$

and

$$\beta_1^2=[\kappa(0)]^2\gamma^2/4\beta_0^2,$$

where γ is defined in (3.17). Now if we write for small s

$$\kappa(s)=\kappa_0+\kappa_2s^2+\kappa_3s^3+\kappa_4s^4+\cdots, \quad |s|\ll 1,$$

then we can obtain from (3.17)—(3.20), (3.25), and (3.26)

$$\frac{1}{b^{(0)}(0)}\frac{d^2b^{(0)}(0)}{ds^2}=\frac{11\kappa_3^2}{16\kappa_2^2}-\frac{3\kappa_4}{4\kappa_2}. \tag{3.32}$$

Equation (3.32), when combined with (3.31), gives an explicit expression for β_2.

By using (2.8), (2.9), and the information we have obtained from the analysis in this section, we can now write the first two terms in the expansion of $\mathbf{u}=\nabla\varphi+\nabla\times\mathbf{A}$. Retaining terms of orders ω and $\omega^{1/2}$, we have (1.1) as presented in Sec. 1. Here $b^{(0)}(s)$, $c^{(1/2)}(s)$, and $d\psi/ds$ are given by (3.25), (3.30), and (3.17) respectively. We know how to expand β as far as the term involving β_2:

$$\beta=\omega\beta_0+\beta_1+\omega^{-1/2}\beta_{3/2}+\omega^{-1}\beta_2, \tag{3.33}$$

with β_0, β_1, $\beta_{3/2}$, and β_2 given by (3.5), (3.18), (3.26), and (3.31), respectively.

4. HIGHER ORDER MODES

In this section we investigate higher order surface waves. We have seen that the disturbance corresponding to the mode analyzed in the previous section is essentially confined to the region $|s|=O(\omega^{-1/4})$. In order to simplify the analysis of the higher order modes, we rescale the variable s, as well as the variable n, and introduce the stretched variable $\zeta=\omega^{1/4}s$. An analogous procedure was adopted by Morrison[3] for the scalar problem. We now look for solutions of the form

$$\varphi=\exp(-i\beta z-\alpha_L\xi)h(\xi,\zeta;\omega),$$
$$\mathbf{A}=\exp(-i\beta z-\alpha_T\xi)\mathbf{E}(\xi,\zeta;\omega). \tag{4.1}$$

The equations satisfied by h and $\mathbf{E}$ are obtained from (2.10)—(2.14) by omitting the terms involving $\psi(s)$, and by substituting

$$s=\omega^{-1/4}\zeta, \quad \frac{\partial}{\partial s}=\omega^{1/4}\frac{\partial}{\partial\zeta}. \tag{4.2}$$

We restrict our attention to the case of a symmetric boundary curve, for which $\kappa(-s)=\kappa(s)$, so that

$$\kappa=\kappa_0+\omega^{-1/2}\kappa_2\zeta^2+\omega^{-1}\kappa_4\zeta^4+\cdots,$$
$$\kappa'=2\omega^{-1/4}\kappa_2\zeta+\cdots. \tag{4.3}$$

It is then possible to expand the unknown functions in (4.1) in inverse powers of $\omega^{1/2}$. In the unsymmetric case it is necessary to expand in inverse powers of $\omega^{1/4}$, which leads to considerably more algebra. Specifically, we assume expansions of the form

$$h(\xi,\zeta;\omega)=\sum_{l=0}^{\infty}\omega^{-l/2}h^{(l/2)}(\xi,\zeta),$$
$$E_t(\xi,\zeta;\omega)=\sum_{l=0}^{\infty}\omega^{-l/2}E_t^{(l/2)}(\xi,\zeta), \tag{4.4}$$

and

$$E_n(\xi,\zeta;\omega)=\sum_{l=0}^{\infty}\omega^{-(2l+1)/4}E_n^{(l/2+1/4)}(\xi,\zeta),$$
$$E_k(\xi,\zeta;\omega)=\sum_{l=0}^{\infty}\omega^{-(2l+1)/4}E_k^{(l/2+1/4)}(\xi,\zeta). \tag{4.5}$$

As before, the expansion of β is given by (2.6).

Proceeding as previously, we deduce that

$$h^{(0)}(\xi,\zeta)=a^{(0)}(\zeta), \quad E_t^{(0)}(\xi,\zeta)=G_t^{(0)}(\zeta), \tag{4.6}$$

and that $\alpha_L>0$ and $\alpha_T>0$ are given by (3.3). The terms of order ω^2 in the boundary conditions corresponding to (2.12) and (2.13) lead to

$$\begin{pmatrix}\beta_0^2-1/2c_T^2 & -i\beta_0\alpha_T\\ i\beta_0\alpha_L & \beta_0^2-1/2c_T^2\end{pmatrix}\begin{pmatrix}a^{(0)}\\ G_t^{(0)}\end{pmatrix}=0, \tag{4.7}$$

and hence to the secular equation (3.5) for β_0. Similarly,

$$E_n^{(1/4)}(\xi,\zeta)=G_n^{(1/4)}(\zeta),\quad E_k^{(1/4)}(\xi,\zeta)=G_k^{(1/4)}(\zeta), \tag{4.8}$$

and the boundary conditions lead to

$$\begin{pmatrix} i\beta_0\alpha_T & -(\beta_0^2-1/2c_T^2) \\ \alpha_T & i\beta_0 \end{pmatrix} \begin{pmatrix} G_n^{(1/4)} \\ G_k^{(1/4)} \end{pmatrix} =0, \tag{4.9}$$

and hence to

$$G_n^{(1/4)}=G_k^{(1/4)}=0. \tag{4.10}$$

Since the procedure for determining the higher order terms is analogous to that used in the previous section, we merely state the results. It is found that $\beta_{1/2}=0$, as before, and that

$$h^{(1/2)}(\xi,\zeta)=a^{(1/2)}(\zeta),\quad E_t^{(1/2)}(\xi,\zeta)=G_t^{(1/2)}(\zeta), \tag{4.11}$$

where

$$\begin{pmatrix} \beta_0^2-1/2c_T^2 & -i\beta_0\alpha_T \\ i\beta_0\alpha_L & \beta_0^2-1/2c_T^2 \end{pmatrix} \begin{pmatrix} a^{(1/2)} \\ G_t^{(1/2)} \end{pmatrix} =0. \tag{4.12}$$

Also,

$$E_n^{(3/4)}(\xi,\zeta)=G_n^{(3/4)}(\zeta),\quad E_k^{(3/4)}(\xi,\zeta)=G_k^{(3/4)}(\zeta), \tag{4.13}$$

where

$$\begin{pmatrix} i\beta_0\alpha_T & -(\beta_0^2-1/2c_T^2) \\ \alpha_T & i\beta_0 \end{pmatrix} \begin{pmatrix} G_n^{(3/4)} \\ G_k^{(3/4)} \end{pmatrix} = \frac{d}{d\zeta} \begin{pmatrix} \alpha_L & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a^{(0)} \\ G_t^{(0)} \end{pmatrix}. \tag{4.14}$$

With the aid of (4.7) this gives

$$G_n^{(3/4)}=0,\quad G_k^{(3/4)}=\frac{1}{i\beta_0}\frac{dG_t^{(0)}}{d\zeta}. \tag{4.15}$$

Next, it is found that

$$h^{(1)}(\xi,\zeta)=a^{(1)}(\zeta)+(\xi/2\alpha_L)(\alpha_L\kappa_0-2\beta_0\beta_1)a^{(0)}(\zeta), \tag{4.16}$$

$$E_t^{(1)}(\xi,\zeta)=G_t^{(1)}(\zeta)+(\xi/2\alpha_T)(\alpha_T\kappa_0-2\beta_0\beta_1)G_t^{(0)}(\zeta),$$

where

$$\begin{pmatrix} \beta_0^2-1/2c_T^2 & -i\beta_0\alpha_T \\ i\beta_0\alpha_L & \beta_0^2-1/2c_T^2 \end{pmatrix} \begin{pmatrix} a^{(1)} \\ G_t^{(1)} \end{pmatrix} + \begin{pmatrix} 0 & -i\beta_1\alpha_T \\ i\beta_1\alpha_L & \kappa_0\alpha_T \end{pmatrix} \begin{pmatrix} a^{(0)} \\ G_t^{(0)} \end{pmatrix} + \begin{pmatrix} -1 & i\beta_0/2\alpha_T \\ -i\beta_0/2\alpha_L & -1 \end{pmatrix} \times \begin{pmatrix} (\kappa_0\alpha_L-2\beta_0\beta_1) & 0 \\ 0 & (\kappa_0\alpha_T-2\beta_0\beta_1) \end{pmatrix} \begin{pmatrix} a^{(0)} \\ G_t^{(0)} \end{pmatrix} =0. \tag{4.17}$$

Premultiplication of the entire equation by $(\beta_0^2-1/2c_T^2\ \ i\beta_0\alpha_T)$ leads, with the help of (3.5) and (4.7), to

$$2\beta_0\beta_1=\gamma\kappa_0, \tag{4.18}$$

where γ is as defined in (3.17). This is consistent with (3.18). It is also found that

$$E_n^{(5/4)}(\xi,\zeta)=G_n^{(5/4)}(\zeta),\quad E_k^{(5/4)}(\xi,\zeta)=G_k^{(5/4)}(\zeta), \tag{4.19}$$

and the boundary conditions lead to

$$G_n^{(5/4)}=0,\quad G_k^{(5/4)}=\frac{1}{i\beta_0}\frac{dG_t^{(1/2)}}{d\zeta}. \tag{4.20}$$

Finally, it is found that

$$h^{(3/2)}(\xi,\zeta)-a^{(3/2)}(\zeta)=\frac{\xi}{2\alpha_L}\left((\alpha_L\kappa_0-2\beta_0\beta_1)a^{(1/2)}(\zeta) +(\alpha_L\kappa_2\zeta^2-2\beta_0\beta_{3/2})a^{(0)}(\zeta)+\frac{d^2a^{(0)}}{d\zeta^2}\right), \tag{4.21}$$

and

$$E_t^{(3/2)}(\xi,\zeta)-G_t^{(3/2)}(\zeta)=\frac{\xi}{2\alpha_T}\left((\alpha_T\kappa_0-2\beta_0\beta_1)G_t^{(1/2)}(\zeta) +(\alpha_T\kappa_2\zeta^2-2\beta_0\beta_{3/2})G_t^{(0)}(\zeta)+\frac{d^2G_t^{(0)}}{d\zeta^2}\right). \tag{4.22}$$

The boundary conditions, with the aid of (3.5), (3.17), (4.7), (4.12), (4.15), and (4.18), lead to

$$\frac{d^2a^{(0)}}{d\zeta^2}+(\gamma\kappa_2\zeta^2-2\beta_0\beta_{3/2})a^{(0)}(\zeta)=0. \tag{4.23}$$

The solution of the differential equation (4.23) may be expressed in terms of parabolic cylinder functions.[4] Since $s=\omega^{-1/4}\zeta$, we want solutions which are bounded for large $|\zeta|$. Hence,

$$2\beta_0\beta_{3/2}=-(2m+1)\delta,\quad a_0=D_m[(2\delta)^{1/2}\zeta], \tag{4.24}$$

where m is a nonnegative integer and $\delta=(-\gamma\kappa_2)^{1/2}$, which is consistent with the definition in (3.20). The case $m=0$ corresponds to the lowest order mode considered in the previous section, since from (3.20), for $s=\omega^{-1/4}\zeta$,

$$b_0(s)\exp[\omega^{1/2}\psi(s)]\sim b_0(0)\exp(-\tfrac{1}{2}\delta\zeta^2), \tag{4.25}$$

and $D_0(\eta)=\exp(-\frac{1}{4}\eta^2)$.

The approach used in this section was carried to one higher order in the simpler scalar problem,[3] and it was found that the next term in the expansion of the propagation constant is determined by an orthogonality condition on $-\infty<\zeta<\infty$. However, because of the algebraic complexity of the present problem, we have not carried the analysis of the higher order modes any further. We also remark that a more uniform expansion for the higher order modes was developed for the scalar problem,[3] and the corresponding expansion for the present problem would involve parabolic cylinder functions with argument $2\omega^{1/4}[-\psi(s)]^{1/2}$. However, two unknown functions were introduced in the asymptotic expansion of the single dependent variable in the scalar problem, and presumably eight functions would be required in the present problem, so that we refrain from developing the more uniform expansion.

5. CLOSED BOUNDARY CURVES

We have confined our attention, so far, to open boundary curves for which the curvature attains its algebraic maximum at a single point, designated by $s=0$. We have seen that the disturbance, in addition to being confined close to the surface in the region $n=O(\omega^{-1})$, is also essentially confined to the region $|s|=O(\omega^{-1/4})$, and decays exponentially outside this region. As was argued for the scalar problem,[3] if the boundary curve is closed, and the curvature attains its algebraic maximum at a single point, we would intuitively expect that surface wave modes would exist which differ by only an exponentially small amount from those derived in the previous sections. Moreover, the corresponding values of the propagation constant β would presumably also differ by an exponentially small quantity.

We now turn our attention to closed boundary curves which are symmetric, and for which the curvature attains its algebraic maximum at two points, designated by $s=0$ and $s=2\tau$. Thus, we assume that

$$\kappa(\tau+\sigma)=\kappa(\tau-\sigma), \quad 0\leq\sigma\leq l/2, \tag{5.1}$$

where l is the length of the circumference of the boundary curve. Asymptotic expansions may be developed, as in the previous sections, about $s=0$ and about $s=2\tau$. Because of the symmetry, both expansions will lead to the same expressions for the propagation constant β and to the same modal shape. Analogously to the previous discussion, we intuitively expect modes to exist with shapes differing from those corresponding to the asymptotic expansions by an exponentially small amount, and with values of β differing by an exponentially small quantity. However, care must be taken as to what should be called a mode.

Thus, from the symmetry and from the differential equations and boundary conditions, it is to be expected that to each modal expansion about $s=0$ there correspond two true modes with the symmetries

$$\begin{aligned}
&\varphi(n,\tau+\sigma,z)=\pm\varphi(n,\tau-\sigma,z),\\
&A_t(n,\tau+\sigma,z)=\pm A_t(n,\tau-\sigma,z),\\
&A_n(n,\tau+\sigma,z)=\mp A_n(n,\tau-\sigma,z),\\
&A_k(n,\tau+\sigma,z)=\mp A_k(n,\tau-\sigma,z).
\end{aligned} \tag{5.2}$$

Moreover, it is to be expected that the values of β for these two modes differ by only an exponentially small amount. The corresponding assertions for the scalar problem[3] were verified for a particular example, at least in so far as an approximation to the surface wave is concerned. We conclude by noting that the symmetries for the displacement corresponding to (5.2) are

$$\begin{aligned}
&u_n(n,\tau+\sigma,z)=\pm u_n(n,\tau-\sigma,z),\\
&u_k(n,\tau+\sigma,z)=\pm u_k(n,\tau-\sigma,z),\\
&u_t(n,\tau+\sigma,z)=\mp u_t(n,\tau-\sigma,z).
\end{aligned} \tag{5.3}$$

ACKNOWLEDGMENTS

We are grateful to G. D. Boyd for posing the problem which led to this investigation and to J. McKenna for encouragement and many fruitful discussions.

APPENDIX

We present here general forms of the differential equations (2.10), boundary conditions (2.12)–(2.14), and divergence condition (2.15). First we write the differential equations. We shall make use of (2.6), (2.9), (3.2), (3.6), and the fact that $h=1-\kappa\xi/\omega$. If we were to set terms of order $\omega^{2-p/2}$ to zero in (2.10), we would get

$$\begin{aligned}
&\left(-2\alpha_L\frac{\partial}{\partial\xi}+\frac{\partial^2}{\partial\xi^2}\right)g^{(p/2)}+\left[\kappa\left(\alpha_L-\frac{\partial}{\partial\xi}\right)+\left(\frac{d\psi}{ds}\right)^2-2\beta_0\beta_1\right]\\
&\times g^{((p-2)/2)}+\left(2\frac{d\psi}{ds}\frac{\partial}{\partial s}+\frac{d^2\psi}{ds^2}-2\beta_0\beta_{3/2}\right)g^{((p-3)/2)}\\
&-2\beta_0\beta_{p/2}b^{(0)}+(\dagger)=0,\\
&\left(-2\alpha_T\frac{\partial}{\partial\xi}+\frac{\partial^2}{\partial\xi^2}\right)C_t^{(p/2)}+\left[\kappa\left(\alpha_T-\frac{\partial}{\partial\xi}\right)+\left(\frac{d\psi}{ds}\right)^2-2\beta_0\beta_1\right]\\
&\times C_t^{((p-2)/2)}+\left(2\frac{d\psi}{ds}\frac{\partial}{\partial s}+\frac{d^2\psi}{ds^2}-2\beta_0\beta_{3/2}\right)C_t^{((p-3)/2)}\\
&-2\kappa\frac{d\psi}{ds}C_n^{((p-3)/2)}-2\beta_0\beta_{p/2}F_t^{(0)}+(\dagger)=0,\\
&\left(-2\alpha_T\frac{\partial}{\partial\xi}+\frac{\partial^2}{\partial\xi^2}\right)C_n^{(p/2)}+\left[\kappa\left(\alpha_T-\frac{\partial}{\partial\xi}\right)+\left(\frac{d\psi}{ds}\right)^2-2\beta_0\beta_1\right]\\
&\times C_n^{((p-2)/2)}+\left(2\frac{d\psi}{ds}\frac{\partial}{\partial s}+\frac{d^2\psi}{ds^2}-2\beta_0\beta_{3/2}\right)C_n^{((p-3)/2)}\\
&+2\kappa\frac{d\psi}{ds}C_t^{((p-3)/2)}+(\dagger)=0,\\
&\left(-2\alpha_T\frac{\partial}{\partial\xi}+\frac{\partial^2}{\partial\xi^2}\right)C_k^{(p/2)}+\left[\kappa\left(\alpha_T-\frac{\partial}{\partial\xi}\right)+\left(\frac{d\psi}{ds}\right)^2-2\beta_0\beta_1\right]\\
&\times C_k^{((p-2)/2)}+\left(2\frac{d\psi}{ds}\frac{\partial}{\partial s}+\frac{d^2\psi}{ds^2}-2\beta_0\beta_{3/2}\right)C_k^{((p-3)/2)}\\
&+(\dagger)=0.
\end{aligned} \tag{A1}$$

The dagger denotes all terms which have superscripts less than or equal to $(p-4)/2$. If the superscript is negative, then by convention that term is zero. It will turn out that if we are, as above, considering terms of order $\omega^{2-p/2}$, then all those denoted by daggers will have been determined explicitly by earlier analysis of higher order terms. We do, however, write out explicitly those terms involving $\beta_{p/2}$, since $\beta_{p/2}$ remains to be determined. If $p\leq 3$, then these terms are redundant and should not be included again.

We will return to these equations shortly to do a little more analysis, but first let us look at the boundary conditions. To simplify notation slightly, write

$$X^{p/2}=\begin{pmatrix}g^{(p/2)}\\C_t^{(p/2)}\end{pmatrix}, \quad Y^{p/2}=\begin{pmatrix}C_n^{(p/2)}\\C_k^{(p/2)}\end{pmatrix}. \tag{A2}$$

In the boundary conditions and in other equations which follow, any terms which involve $\beta_{p/2}$ appear explicitly, and an asterisk denotes terms whose superscripts are of lower order than the other terms which appear explicitly. The order may differ for the X terms and the Y terms. Then, taking terms of order $\omega^{2-p/2}$ in the first two boundary conditions (2.12) and (2.13) gives

$$\begin{pmatrix} \beta_0^2 - 1/2c_T^2 & -i\beta_0\alpha_T \\ i\beta_0\alpha_L & \beta_0^2 - 1/2c_T^2 \end{pmatrix} X^{p/2}$$

$$+ \begin{pmatrix} -2\alpha_L\partial/\partial\xi + \partial^2/\partial\xi^2 & i\beta_0\partial/\partial\xi \\ -i\beta_0\partial/\partial\xi & -2\alpha_T\partial/\partial\xi + \partial^2/\partial\xi^2 \end{pmatrix} X^{p/2}$$

$$+ \frac{d\psi}{ds}\begin{pmatrix} \alpha_T - \frac{\partial}{\partial\xi} \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} Y^{(p-1)/2}$$

$$+ \begin{pmatrix} 0 & -i\beta_1(\alpha_T - \partial/\partial\xi) \\ i\beta_1(\alpha_L - \partial/\partial\xi) & \kappa(\alpha_T - \partial/\partial\xi) \end{pmatrix} X^{(p-2)/2}$$

$$+ \frac{\partial}{\partial s}\begin{pmatrix} \alpha_T - \frac{\partial}{\partial\xi} \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} Y^{(p-2)/2}$$

$$+ i\beta_{3/2}\begin{pmatrix} 0 & -(\alpha_T - \partial/\partial\xi) \\ \alpha_L - \partial/\partial\xi & 0 \end{pmatrix} X^{(p-3)/2}$$

$$+ i\beta_{p/2}\begin{pmatrix} 0 & -\alpha_T \\ \alpha_L & 0 \end{pmatrix} X^0 + (*) = 0 \quad \text{at} \quad \xi = 0. \tag{A3}$$

The term involving $\beta_{p/2}$ is redundant if $p \leq 3$.

Terms of order $\omega^{2-p/2}$ in the third boundary condition (2.14) and of order $\omega^{1-p/2}$ in the divergence condition (2.15) give

$$\begin{pmatrix} i\beta_0\alpha_T & -(\beta_0^2 - 1/2c_T^2) \\ \alpha_T & i\beta_0 \end{pmatrix} Y^{p/2}$$

$$- \begin{pmatrix} i\beta_0\partial/\partial\xi & -2\alpha_T\partial/\partial\xi + \partial^2/\partial\xi^2 \\ \partial/\partial\xi & 0 \end{pmatrix} Y^{p/2}$$

$$- \frac{d\psi}{ds}\begin{pmatrix} \alpha_L - \partial/\partial\xi & 0 \\ 0 & 1 \end{pmatrix} X^{(p-1)/2}$$

$$- \frac{\partial}{\partial s}\begin{pmatrix} \alpha_L - \partial/\partial\xi & 0 \\ 0 & 1 \end{pmatrix} X^{(p-2)/2}$$

$$+ \begin{pmatrix} i\beta_1(\alpha_T - \partial/\partial\xi) & 0 \\ \kappa & i\beta_1 \end{pmatrix} Y^{(p-2)/2} + (*) = 0 \quad \text{at} \quad \xi = 0. \tag{A4}$$

We now return to the differential equations (A1) in order to replace the second matrix in (A3) and (A4) by terms involving lower order superscripts. Because we reject solutions which grow exponentially with ξ and because of the form of (A1), it is possible to show inductively that for all p,

$$\frac{\partial g^{(p/2)}}{\partial\xi} = \frac{1}{2\alpha_L}\left\{\left[\kappa\alpha_L + \left(\frac{d\psi}{ds}\right)^2 - 2\beta_0\beta_1\right] g^{((p-2)/2)} + \left(2\frac{d\psi}{ds}\frac{\partial}{\partial s} + \frac{d^2\psi}{ds^2} - 2\beta_0\beta_{3/2}\right) g^{((p-3)/2)} - 2\beta_0\beta_{p/2}b^{(0)}\right\} + (*) \quad \text{at} \quad \xi = 0. \tag{A5}$$

As before, the term involving $\beta_{p/2}$ is redundant if $p \leq 3$. Similar results hold for $\partial\mathbf{C}^{(p/2)}/\partial\xi\,|_{\xi=0}$. Notice that $\partial g^{(p/2)}/\partial\xi\,|_{\xi=0}$ and $\partial\mathbf{C}^{(p/2)}/\partial\xi\,|_{\xi=0}$ involve only terms whose superscripts are less than or equal to $(p-2)/2$.

We use (A1) and (A5) to rewrite the boundary conditions and divergence condition (A3) and (A4) in the following forms. Again remember that the term involving $\beta_{p/2}$ is redundant for $p \leq 3$.

$$\begin{pmatrix} \beta_0^2 - 1/2c_T^2 & -i\beta_0\alpha_T \\ i\beta_0\alpha_L & \beta_0^2 - 1/2c_T^2 \end{pmatrix} X^{p/2} + \begin{pmatrix} 0 & -i\beta_1\alpha_T \\ i\beta_1\alpha_L & \kappa\alpha_T \end{pmatrix} X^{(p-2)/2} + \begin{pmatrix} -1 & i\beta_0/2\alpha_T \\ -i\beta_0/2\alpha_L & -1 \end{pmatrix}$$

$$\times \begin{pmatrix} [\kappa\alpha_L + (d\psi/ds)^2 - 2\beta_0\beta_1] & 0 \\ 0 & [\kappa\alpha_T + (d\psi/ds)^2 - 2\beta_0\beta_1] \end{pmatrix} X^{(p-2)/2}$$

$$+ \left(2\frac{d\psi}{ds}\frac{\partial}{\partial s} + \frac{d^2\psi}{ds^2} - 2\beta_0\beta_{3/2}\right)\begin{pmatrix} -1 & i\beta_0/2\alpha_T \\ -i\beta_0/2\alpha_L & -1 \end{pmatrix} X^{(p-3)/2}$$

$$+ i\beta_{3/2}\begin{pmatrix} 0 & -\alpha_T \\ \alpha_L & 0 \end{pmatrix} X^{(p-3)/2} + i\beta_{p/2}\begin{pmatrix} 0 & -\alpha_T \\ \alpha_L & 0 \end{pmatrix} X^0$$

$$- 2\beta_0\beta_{p/2}\begin{pmatrix} -1 & i\beta_0/2\alpha_T \\ -i\beta_0/2\alpha_L & -1 \end{pmatrix} X^0$$

$$+ \alpha_T\frac{d\psi}{ds}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} Y^{(p_1-2)/2} + \alpha_T\frac{\partial}{\partial s}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} Y^{(p-2)/2} + (*) = 0 \text{ at } \xi = 0. \tag{A6}$$

$$\begin{pmatrix} i\beta_0\alpha_T & -(\beta_0^2 - 1/2c_T^2) \\ \alpha_T & i\beta_0 \end{pmatrix} Y^{\flat/2} + \begin{pmatrix} i\beta_1\alpha_T & 0 \\ \kappa & i\beta_1 \end{pmatrix} Y^{(\flat-2)/2}$$

$$+ \left[\kappa\alpha_T + \left(\frac{d\psi}{ds}\right)^2 - 2\beta_0\beta_1\right] \begin{pmatrix} -i\beta_0/2\alpha_T & 1 \\ -1/2\alpha_T & 0 \end{pmatrix} Y^{(\flat-2)/2}$$

$$- \frac{\partial\psi}{\partial s} \begin{pmatrix} \alpha_L & 0 \\ 0 & 1 \end{pmatrix} X^{(\flat-1)/2} - \frac{\partial}{\partial s} \begin{pmatrix} \alpha_L & 0 \\ 0 & 1 \end{pmatrix} X^{(\flat-2)/2}$$

$$+ (*) = 0 \text{ at } \xi = 0. \qquad \text{(A7)}$$

[1]G. Rosenfeld and J.B. Keller, J. Acoust. Soc. Am. 55, 555 (1974).
[2]R.D. Gregory, Proc. Camb. Phil. Soc. 70, 103 (1971).
[3]J.A. Morrison, J. Math. Phys. 16, 1786 (1975) (preceding article).
[4]B.A. Auld, *Acoustic Fields and Waves in Solids* (Wiley, New York, 1973), Vol. II, p. 93.
[5]M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D.C., 1964), pp. 686–700.
[6]Ref. 5, p. 775.
[7]Ref. 4, p. 66.
[8]C.E. Weatherburn, *Advanced Vector Analysis* (Bell, London, 1949), p. 146.
[9]Ref. 8, p. 16.

All stationary axisymmetric rotating dust metrics*

Jeffrey Winicour

Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260
(Received 9 May 1975)

The Einstein equations for stationary axisymmetric space-times with a rotating dust source are systematically reduced to quadratures. The general solution depends upon an arbitrary axisymmetric solution of the flat three-dimensional Laplace equation and upon an arbitrary function of one variable.

1. INTRODUCTION

The formalisms of Ernst[1] and of Geroch[2] reduce the vacuum Einstein equations for stationary axisymmetric space—times to what appears to be their simplest form. However, the problem of obtaining the general solution remains intrinsically difficult. In this paper, I shall describe how the same problem for the Einstein equation with a rotating dust source may be systematically reduced to quadratures.

The dust follows trajectories $T^a + \Omega\Phi^a$, where T^a is the time translational Killing vector, Φ^a is the rotational Killing vector, and Ω is the angular velocity of the flow. The special case of rigid motion, $\Omega = \text{const}$, has been solved by van Stockum[3] in terms of one arbitrary solution of the Laplace equation. In the differentially rotating case treated here, the solution depends upon one arbitrary function of one variable in addition to an arbitrary solution of the Laplace equation. Thus the freedom in the solution is still less than in the vacuum case which depends upon two arbitrary functions of two variables. The hydrodynamical equation governing the rotational motion of the dust gives a constraint between the angular momentum and energy distributions which limits this freedom.

2. THE EQUATIONS

The solutions were obtained using the space-of-trajectories formalism of Geroch[2] as modified to include sources.[4] All notation used here is strictly consistent with Ref. 4. For vanishing convective circulation and pressure, the field equations for rotating dust reduce to[4,5]

$$D^m(\tau^{-1}D_m\lambda_\alpha) = \tau^{-3}\lambda_\alpha(D^m\lambda^\beta)D_m\lambda_\beta + 8\pi\tau^{-1}\mu(\lambda_\alpha + \tau^2\psi^{-1}S_\alpha) \tag{2.1}$$

and

$$S^\alpha D_m\lambda_\alpha = 0. \tag{2.2}$$

Here the index α corresponds to symmetric index pairs (AB) in the two-dimensional space of Killing vectors. The λ_α are scalars formed from the Killing vectors,

$$\tau^2 = -\lambda^\alpha\lambda_\alpha, \tag{2.3}$$

$$S^\alpha = S^{AB} = \begin{pmatrix} 1 & \Omega \\ \Omega & \Omega^2 \end{pmatrix}, \tag{2.4}$$

and

$$\psi = S^\alpha\lambda_\alpha. \tag{2.5}$$

Each solution of Eqs. (2.1) and (2.2) determines a space—time metric by means of straightforward quadratures.[2,4] (The Appendix gives some details of this procedure which are relevant to the present case.) These equations are invariant under conformal rescaling of the geometry of the two dimensional space of trajectories. Furthermore, multiplication of Eq. (2.1) by λ^α gives the harmonic condition

$$D^mD_m\tau = 0 \tag{2.6}$$

so that it is possible, when convenient, to pick a conformally flat frame for which the twodimensional metric h_{mn} has the form

$$h_{mn} = e^{2\Phi}\delta_{mn} \tag{2.7}$$

in terms of coordinates τ and its conjugate harmonic function σ.

In order to analyze the remaining content of Eqs. (2.1) and (2.2), it is helpful to introduce a null triad for the metric $G^{\alpha\beta}$ of the space of symmetry labels.[3,4] We define N^α and R^α by

$$D_mS^\alpha = 2N^\alpha D_m\Omega \tag{2.8}$$

and

$$D_mN^\alpha = R^\alpha D_m\Omega \tag{2.9}$$

so that

$$N^\alpha = N^{AB} = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & \Omega \end{pmatrix}, \tag{2.10}$$

$$R^\alpha = R^{AB} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{2.11}$$

and

$$G^{\alpha\beta} = 2R^{(\alpha}S^{\beta)} - 2N^\alpha N^\beta. \tag{2.12}$$

The Killing scalars can now be written in terms of ψ, η, and ρ^2, where

$$\eta = N^\alpha\lambda_\alpha \tag{2.13}$$

and

$$\rho^2 = R^\alpha\lambda_\alpha, \tag{2.14}$$

with the identity

$$\tau^2 = -\lambda^\alpha\lambda_\alpha = -G^{\alpha\beta}\lambda_\alpha\lambda_\beta = 2(\eta^2 - \rho^2\psi). \tag{2.15}$$

Equation (2.2) now takes the form

$$D_m\psi = 2\eta D_m\Omega, \tag{2.16}$$

which implies that ψ, η, and Ω are mutually dependent functions. The content of Eq. (2.1) not contained in Eq. (2.6) can be extracted by contracting (2.1) with $N^{[\alpha}S^{\beta]}\lambda_\beta$ and with S^α. The Killing scalars in the resulting two

equations can then be rewritten in terms of η, ψ, and τ by using Eqs. (2.3)–(2.15). Also, terms involving $D_m\Omega$ can be eliminated by using Eq. (2.16). For instance,

$$(D_m\lambda^\alpha)D_n\lambda_\alpha = (2R^{(\alpha}{}_{\beta)} - 2N^\alpha N^\beta)(D_m\lambda_\beta)D_n\lambda_\alpha$$
$$= -2(N^\beta D_m\lambda_\beta)N^\alpha D_n\lambda_\alpha$$

and

$$N^\alpha D_m\lambda_\alpha = D_m\eta - \rho^2 D_m\Omega$$
$$= D_m\eta + [(\tau^2 - 2\eta^2)/4\eta\psi]D_m\psi.$$

This procedure leads to the two equations:

$$D^m\left[\frac{\tau}{2\eta\psi}D_m\psi + \frac{\psi}{\eta\tau}D_m\left(\frac{\eta^2}{\psi}\right)\right] + \frac{(D^m\psi)D_m\tau}{\eta\psi} = 0 \tag{2.17}$$

and

$$16\pi\tau^2\eta^2\psi^{-2}\mu = \left[D^m\left(\frac{\eta^2}{\psi}\right)\right]D_m\left(\frac{\eta^2}{\psi}\right) - \frac{\tau^4}{4\psi^4}(D^m\psi)D_m\psi. \tag{2.18}$$

3. THE SOLUTIONS

The field equations have now been reduced to Eqs. (2.6), (2.16), (2.17), and (2.18). Equation (2.6) is satisfied by introducing τ as a harmonic coordinate, and Eq. (2.16) implies that ψ and η are functionally dependent. Equation (2.18) determines the dust density μ in terms of η, ψ, τ and the conformal scale of the 2-geometry (see the Appendix). That leaves Eq. (2.17) to be solved.

In order to solve Eq. (2.17), the functional dependence of η and ψ is used to define a function β by

$$D_m\beta = D_m\psi/\eta\psi. \tag{3.1}$$

By using the harmonic property of τ, Eq. (2.17) can then be put in the form

$$D^m\left\{\frac{1}{\tau}\left[\tfrac{1}{2}D_m(\beta\tau^2) + \frac{\psi}{\eta}D_m\left(\frac{\eta^2}{\psi}\right)\right]\right\} = 0. \tag{3.2}$$

This last equation implies the existence of a potential ω defined by

$$*D_m\omega = \frac{1}{\tau}\left[\tfrac{1}{2}D_m(\beta\tau^2) + \frac{\psi}{\eta}D_m\left(\frac{\eta^2}{\psi}\right)\right], \tag{3.3}$$

where a star denotes the dual operator so that

$$*D^m = \epsilon^{mn}D_n$$

in terms of the twodimensional alternating tensor ϵ^{mn}. The potential ω must satisfy the equation

$$*D^m *D_m\omega = *D^m\left\{\frac{1}{\tau}\left[\tfrac{1}{2}D_m(\beta\tau^2) + \frac{\psi}{\eta}D_m\left(\frac{\eta^2}{\psi}\right)\right]\right\}$$
$$= *D^m\left(\frac{1}{\tau}\right)\left[\tfrac{1}{2}D_m(\beta\tau^2) + \frac{\psi}{\eta}D_m\left(\frac{\eta^2}{\psi}\right)\right]$$
$$= -\tau^{-1}(*D^m\tau)*D_m\omega$$

or

$$\tau^{-1}D^m(\tau D_m\omega) = 0. \tag{3.4}$$

But Eq. (3.4) is equivalent to a flat threedimensional Laplace equation!

The general solution thus depends upon an arbitrary axisymmetric solution of the Laplace equation ω and an arbitrary function of one variable, say $\eta(\psi)$. Then Eq. (3.1) determines the function $\beta(\psi)$, and Eq. (3.3) gives

$$\tau *D_m\omega = \tfrac{1}{2}D_m(\beta\tau^2) + \frac{\psi}{\eta}D_m\left(\frac{\eta^2}{\psi}\right). \tag{3.5}$$

Integration of Eqs. (3.5) determines the function

$$\tfrac{1}{2}\beta\tau^2 + \alpha,$$

where α is defined by

$$\alpha = 2\eta - \int(\eta/\psi)\,d\psi.$$

Thus ω and $\eta(\psi)$ determine ψ by a sequence of quadratures. Finally, Ω is determined by integration of (2.14) and the metric is determined by a further quadrature (see the Appendix).

For special choices of the functions ω and $\eta(\psi)$. The physical properties of the solution depend upon the density μ determined by Eq. (2.18). In particular, the physically relevant domain of the solution is restricted to the region where $(-\psi)$, μ, and ρ^2 are positive. Explicit details for special cases will be presented elsewhere.

ACKNOWLEDGMENTS

I am grateful to R. O. Hansen and C. V. Vishveshwara for numerous helpful discussions.

APPENDIX

In addition to the analogs of Eqs. (2.1) and (2.2), there are auxiliary field equations given in Ref. 4 which can be solved by straightforward methods. In the present case, these equations reduce to

$$\mathcal{R} = \tfrac{1}{2}\tau^{-2}(D^m\lambda^\alpha)D_m\lambda_\alpha + 8\pi\mu \tag{A1}$$

and

$$E_{mn} \equiv 2D_mD_n\tau + \tau^{-1}(D_m\lambda^\alpha)D_n\lambda_\alpha - \tfrac{1}{2}\tau^{-1}h_{mn}(D^p\lambda^\alpha)D_p\lambda_\alpha = 0, \tag{A2}$$

where $\mathcal{R}$ is the scalar curvature of the 2-metric h_{ab}.

When Eqs. (2.1), (2.2), and (2.6) hold, the divergence of E_{mn} satisfies

$$D^mE_{mn} = D_n\tau[\mathcal{R} - \tfrac{1}{2}\tau^{-2}(D^m\lambda^\alpha)D_m\lambda_\alpha - 8\pi\mu]$$

so that Eq. (A1) is extraneous.

Also, E_{mn} is trace-free and symmetric so that Eq. (A2) is equivalent to

$$(D^n\tau)E_{mn} = 0$$

or

$$-D_m[(D^n\tau)D_n\tau]$$
$$= \tau^{-1}(D^n\tau)(D_m\lambda^\alpha)D_n\lambda_\alpha - \tfrac{1}{2}\tau^{-1}(D_m\tau)(D^p\lambda^\alpha)D_p\lambda_\alpha. \tag{A3}$$

To analyze the content of Eq. (A3), it is convenient to introduce the harmonic coordinates, described by Eq. (2.7), for which

$$\partial_m\partial_n\tau = 0,$$

where ∂_m is the derivative with respect to the metric δ_{ab}. Then Eq. (A3) reduces to

$$2\partial_m\phi = \tau^{-1}(\partial^n\tau)(\partial_m\lambda^\alpha)\partial_n\lambda_\alpha - \tfrac{1}{2}\tau^{-1}(\partial_m\tau)(\partial^n\lambda^\alpha)\partial_n\lambda_\alpha. \quad \text{(A4)}$$

Equation (A4) determines the conformal scale factor in terms of a quadrature provided that the right-hand side is consistent with the integrability condition

$$\epsilon^{mp}\partial_p\partial_m\phi = 0.$$

we have

$$\begin{aligned}
&\epsilon^{mp}\partial_p[\tau^{-1}(\partial^n\tau)(\partial_m\lambda^\alpha)(\partial_n\lambda_\alpha) - \tfrac{1}{2}\tau^{-1}(\partial_m\tau)(\partial^n\lambda^\alpha)\partial_n\lambda_\alpha] \\
&\quad = \epsilon^{mp}(\partial_n\tau)(\partial_m\lambda^\alpha)\partial_p(\tau^{-1}\partial^n\lambda_\alpha) - \epsilon^{mp}(\partial_m\tau)(\partial_n\lambda^\alpha)\partial_p(\tau^{-1}\partial^n\lambda_\alpha) \\
&\quad = \epsilon^{mp}\epsilon_{nm}\epsilon^{qr}(\partial_q\tau)(\partial_r\lambda^\alpha)\partial_p(\tau^{-1}\partial^n\lambda_\alpha) \\
&\quad = -\epsilon^{qr}(\partial_q\tau)(\partial_r\lambda^\alpha)\partial_n(\tau^{-1}\partial^n\lambda_\alpha).
\end{aligned}$$

But the final side of this equation is conformally invariant and vanishes because of Eqs. (2.1) and (2.2).

*This research was supported in part by NSF Grant MPS74 18020.

[1]F. Ernst, Phys. Rev. **168**, 1415 (1968).
[2]R. Geroch, J. Math. Phys. **13**, 394 (1972).
[3]W. van Stockum, Proc. Roy. Soc. Edinburgh A **57**, 135 (1937).
[4]R. O. Hansen and J. Winicour, J. Math. Phys. **16**, 804 (1975).
[5]Certain constants c_A in Refs. 2 and 4 vanish because of the assumption of a rotation axis (or asymptotic flatness).

Correlation inequalities for two-dimensional vector spin systems

James L. Monroe*

Department of Physics, Northwestern University, Evanston, Illinois 60201
(Received 25 February 1975)

A set of correlation function inequalities including Griffiths, Kelly, Sherman type inequalities are proven for a lattice system of N sites where on each site there is a vector spin $[\mathbf{s}=(s^x,s^z), |\mathbf{s}|=1]$ whose distribution of values over the unit circle is given by $f(\mathbf{s})$, where $f(\mathbf{s})=f(-\mathbf{s})$. The spins interact through two-body, anisotropic, ferromagnetic interactions. Also an external field $\mathbf{h}$, $h^x\geq 0$, and $h^z\geq 0$, is present. The proof uses Gaussian random variables.

I. INTRODUCTION

Correlation inequalities have received a great amount of interest in recent years initially because of their applications in statistical mechanics[1] and more recently because of their applications in quantum field theory.[2] In this paper we extend the method of Gaussian random variables used previously[3-5] to prove a number of correlation inequalities for systems with one-dimensional spins to prove a set of inequalities for certain lattice spin systems having two-dimensional spins.

Ginibre[6] has proven a set of Griffiths, Kelly, and Sherman[7,8] (hereafter GKS) inequalities for a certain class of systems with two-dimensional vector spins. The class of systems allowed in the following results complements that of Ginibre's by being in some respects less restrictive, e.g., the amount of allowed anisotropy being greater, while in other respects being more restrictive, e.g., only allowing two-body interactions.

Besides complementing the class of systems for which the inequalities hold new inequalities are proven. A negative GKS inequality is proven which shows that any correlation average of a product of x components of the spin decreases as a function of the interactions in the z direction and vice versa [see Eq. (23)]. Also a set of new inequalities of more complicated forms of thermal averages is proven.

In Sec. II we describe the general model system and discuss a number of preliminaries before proving in Sec. III the existence of a general class of inequalities for these model systems. Section IV contains explicit examples of the general inequalities proven in the preceding section and some application of these inequalities to statistical mechanics. Applications of these type inequalities in quantum field theory are also now being made.[9]

II. THE MODEL SYSTEM

The model system consist of set of N sites in ν-dimensional space. Each site has associated with it a two-dimensional vector spin, $\mathbf{s}=(s^x,s^z)$, with $|\mathbf{s}|=1$, whose distribution over the unit circle is given by $f(\mathbf{s})$ where $f(\mathbf{s})$ is assumed to be even, i.e., $f(\mathbf{s})=f(-\mathbf{s})$. For the plane rotator system $f(\mathbf{s})=1$. Discrete rotators with p allowed states [p even since $f(\mathbf{s})=f(-\mathbf{s})$] can be constructed by taking $f(\mathbf{s})=\sum_{n=1}^{p}\delta(\theta-[2\pi/p]n)$. Models of this type were proposed and studied by Potts.[10] The system is assumed to have a Hamiltonian of the form

$$H=-\tfrac{1}{2}\sum_{i,j}J(i,j)\,(s_i{}^x s_j{}^x+\epsilon^2 s_i{}^z s_j{}^z)-\sum_{i=1}^{N}(h^x s_i{}^x+h^z s_i{}^z) \tag{1}$$

where $J(i,j)$, ϵ, h^x, and h^z are nonnegative for all i and j. The external fields, h^x and h^z, could be made to vary from site to site if desired.

The method of the proof is based on the identity[11]

$$\exp\left(\tfrac{1}{2}\sum_{i,j}\xi_i\alpha_{ij}\xi_j\right)=(2\pi)^{-N/2}(\det\alpha)^{-1/2}\times\int\cdots\int\exp\left(-\tfrac{1}{2}\sum_{i,j}x_i(\alpha^{-1})_{ij}x_j+\sum_{j=1}^{N}\xi_j x_j\right)\prod_{i=1}^{N}dx_i \tag{2}$$

valid for any real, symmetric, and positive definite matrix α, and for any N complex variables ξ_K. The sign of $(\det\alpha)^{-1/2}$ is to be chosen positive. The right-hand side of Eq. (2) can be considered as the expected value $E_{\mathbf{x}}(\exp\sum_{i=1}^{N}x_i\xi_i)$ with respect to the Gaussian density function

$$W_N(\mathbf{x})=(2\pi)^{-N/2}(\det\alpha)^{-1/2}\exp\left(-\tfrac{1}{2}\sum_{i,j}x_i(\alpha^{-1})_{ij}x_j\right), \tag{3}$$

where $\mathbf{x}=(x_1,x_2,\ldots,x_N)$. If α is a nonnegative matrix, then one can show

$$E_{\mathbf{x}}\left\{\prod_{i=1}^{N}(x_i)^{n_i}\right\}\begin{cases}=0 & \text{if } \sum_{i=1}^{N}n_i \text{ is odd}\\ \geq 0 & \text{if } \sum_{i=1}^{N}n_i \text{ is even,}\end{cases} \tag{4}$$

where the n_i are nonnegative integers.

The identity (2) can be used to rewrite the Boltzmann factor $\exp(-\beta H)$, where hereafter we set $\beta=1$, by identifying the variable ξ_j with the spin variables $S_j{}^x$ or $\epsilon S_j{}^z$ and forming a matrix $J=\alpha$ with off-diagonal elements $J(i,j)$ and all diagonal elements equal to a number $J_0\equiv J(i,i)$ large enough to guarantee that J is positive definite. The Boltzmann factor with Hamiltonian (1) is then

$$e^{-H}=\int\cdots\int W_N(\mathbf{x})W_N(\mathbf{z})\exp\left(\sum_{i=1}^{N}(x_i s_i{}^x+h^x s_i{}^x)\right)\times\exp\left(\sum_{i=1}^{N}(z_i\epsilon s_i{}^z+h^z s_i{}^z)\right) \tag{5}$$

$$\equiv E_{\mathbf{x},\mathbf{z}}\left\{\exp\left(\sum_{i=1}^{N}(x_i s_i{}^x+h^x s_i{}^x\right)\exp\left(\sum_{i=1}^{N}(z_i\epsilon s_i{}^z+h^z s_i{}^z)\right)\right\}.$$

Here we only assume J to be positive definite and $J(i,j)$ real and symmetric $J(i,j)=J(j,i)$. The partition function, Z_N, is given by

$$Z_N = \int\cdots\int \prod_{i=1}^{N} ds_i f(\mathbf{s}_i) e^{-H}$$
$$= E_{\mathbf{x},\mathbf{z}} \left\{ \int\cdots\int \prod_{i=1}^{N} ds_i f(\mathbf{s}_i) \times \exp\left(\sum_{i=1}^{N} (x_i s_i{}^x + h^x s_i{}^x + z_i \epsilon s_i{}^z + h^z s_i{}^z) \right) \right\}. \quad (6)$$

III. PROOF OF THE CORRELATION INEQUALITIES

Using the re-expression of the partition function in terms of Gaussian random variables, Eq. (6), one can write for the correlation functions expressions such as

$$\langle fg\rangle - \langle f\rangle\langle g\rangle = E\left\{ \int\cdots\int \prod_{i=1}^{N} ds_i f(\mathbf{s}_i) \prod_{j=1}^{N} ds_j' f(\mathbf{s}_j') \times (FG - FG') \exp\left(\sum_{i=1}^{N} s_i{}^x (x_i + h^x) + s_i{}^z (\epsilon z_i + h^z) + \sum_{j=1}^{N} \epsilon_j' {}^x (x_j' + h^x) + s_j{}^z (\epsilon z_j' + h_j{}^z) \right) \right\}, \quad (7)$$

where f and g are products of spin variables $\mathbf{s}_i$, $i = 1,\ldots,N$; and F, G, and G' are the corresponding derivative operators in terms of x_i, ϵz_i, x_i', and $\epsilon z_i'$ with $i = 1, 2, \ldots N$ and where by $E\{\cdots\}$ we mean the average taken with respect to the product measure $W_N(\mathbf{x})\, W_N(\mathbf{z})\, W_N(\mathbf{x}')\, W_N(\mathbf{z}')\, d\mathbf{x}\, d\mathbf{z}\, d\mathbf{x}'\, d\mathbf{z}'$.

The crux of the proof is to rewrite Eq. (7) in terms of the new variables

$$\alpha_i = (1/\sqrt{2})(x_i + x_i'), \quad \beta_i = (\epsilon/\sqrt{2})(\mathbf{z}_i + \mathbf{z}_i'),$$
$$\gamma_i = (1/\sqrt{2})(x_i - x_i'), \quad \delta_i = (\epsilon/\sqrt{2})(\mathbf{z}_i - \mathbf{z}_i'), \quad (8)$$

which are obtained by an orthogonal transformation of the x_i's, x_i''s, z_i's, and z_i''s. The F, G, and G' now become expressions involving partial derivatives with respect to the α's, β's, γ's, and δ's. For a large number of F's, G's, and G''s a simplification occurs. $(FG - FG')$ has a minus sign preceding the second term, while the transformed derivative operator will be a sum of products of partial derivatives all of which are preceded by only a plus sign. This orthogonal transformation is such that $W_N(\mathbf{x})\, W_N(z)\, W_N(\mathbf{x}')\, W_N(z')$ becomes simply $W_N(\alpha)\, W_N(\beta)\, W_N(\gamma)\, W_N(\delta)$ and we shall use $E\{\cdots\}$ to denote either average as they are equivalent.

For the statement of the theorem we focus our attention on the transformed derivative operators, i.e., those involving the α's, β's, γ's, and δ's. We consider the general form

$$D = \prod_{i,j,k,l=1}^{N} \left(\frac{\partial}{\partial\alpha_i}\right)^{n_i{}^\alpha} \left(\frac{\partial}{\partial\beta_j}\right)^{n_j{}^\beta} \left(\frac{\partial}{\partial\gamma_k}\right)^{n_k{}^\gamma} \left(\frac{\partial}{\partial\delta_l}\right)^{n_l{}^\delta} \quad (9)$$

where $i, j, k,$ and l range over all lattice sites $1, 2, \ldots, N$ and $n_i{}^\alpha, n_j{}^\beta, n_k{}^\gamma,$ and $n_l{}^\delta$ are nonnegative integers.

Theorem 1: Define

$$N_1 = \sum_{i=1}^{N} n_i{}^\alpha, \quad N_2 = \sum_{j=1}^{N} n_j{}^\beta, \quad N_3 = \sum_{k=1}^{N} n_k{}^\gamma, \quad N_4 = \sum_{l=1}^{N} n_l{}^\delta. \quad (10)$$

We then have for the model system of Sec. II:

(a) $E\{D\} \geq 0$ if N_3 and N_4 are both even;

(b) $E\{D\} \leq 0$ if N_3 and N_4 are both odd;

(c) $E\{D\} = 0$ in all other cases.

Note: If $h^x(h^z) = 0$, then we have the following more restrictive statements:

(a′) $E\{D\} \geq 0$ if each $N_1(N_2)$, N_3, and N_4 is even;

(b′) $E\{D\} \leq 0$ if each $N_1(N_2)$, N_3, and N_4 is odd;

(c′) $E\{D\} = 0$ in all other cases.

Furthermore, if both $h^x = 0$ and $h^z = 0$, then we have:

(a″) $E\{D\} \geq 0$ only if N_1, N_2, N_3, and N_4 are all even;

(b″) $E\{D\} \leq 0$ only if N_1, N_2, N_3, and N_4 are all odd;

(c″) $E\{D\} = 0$ in all other cases.

We shall only prove (a), (b), and (c). The others will be easily seen to follow.

Proof: From Eq. (7) and the definitions of the rotated variables, Eq. (8), one has

$$E\{D\} = E\left\{ D \int\cdots\int \prod_{i=1}^{N} ds_i f(\mathbf{s}_i) \prod_{j=1}^{N} ds_j' f(\mathbf{s}_j') \times \exp\left[\sum_{k=1}^{N} \left(\frac{1}{\sqrt{2}}\alpha_k + h^x\right)(s_k{}^x + s_k'{}^x) + \left(\frac{\epsilon}{\sqrt{2}}\beta_k + h^z\right) \times (s_k{}^z + s_k'{}^z) + \frac{1}{\sqrt{2}}\gamma_k (s_k{}^x - s_k'{}^x) + \frac{\epsilon}{\sqrt{2}}\delta_k (s_k{}^z - s_k'{}^z) \right] \right\} \quad (11)$$

Now integrating over $-\mathbf{s}_i$ and $-\mathbf{s}_j'$ is equivalent to the integration over $\mathbf{s}_i$ and $\mathbf{s}_j'$, and therefore we can write

$$E\{D\} = E\left\{ D \int\cdots\int \prod_{i=1}^{N} ds_i f(\mathbf{s}_i) \prod_{j=1}^{N} ds_j' f(\mathbf{s}_j') \times \prod_{k=1}^{N} \cosh\left[\left(\frac{1}{\sqrt{2}}\alpha_k + h^x\right)(s_k{}^x + s_k'{}^x) + \left(\frac{\epsilon}{\sqrt{2}}\beta_k + h^z\right)(s_k{}^z + s_k'{}^z) + \frac{1}{\sqrt{2}}\gamma_k (s_k{}^x - s_k'{}^x) + \frac{\epsilon}{\sqrt{2}}\delta_k (s_k{}^z - s_k'{}^z) \right] \right\}. \quad (12)$$

Furthermore, using a common identity of hyperbolic trig functions, we have

$$E\{D\} = E\left\{ D \int\cdots\int \prod_{i=1}^{N} ds_i f(\mathbf{s}_i) \prod_{j=1}^{N} ds_j' f(\mathbf{s}_j') \times \prod_{k=1}^{N} \cosh\left[\left(\frac{1}{\sqrt{2}}\alpha_k + h^x\right)(s_k{}^x + s_k'{}^x) + \left(\frac{\epsilon}{\sqrt{2}}\beta_k + h^z\right)(s_k{}^z + s_k{}^z) \right] \times \cosh\left(\frac{1}{\sqrt{2}}\gamma_k (s_k{}^x - s_k'{}^x) + \frac{\epsilon}{\sqrt{2}}\delta_k (s_k{}^z - s_k'{}^z) \right) + \sinh\left[\left(\frac{1}{\sqrt{2}}\alpha_k + h^x\right)(s_k{}^x + s_k'{}^x) + \left(\frac{\epsilon}{2}\beta_k + h^z\right)(s_k{}^z + s_k'{}^z) \right] \times \sinh\left(\frac{1}{\sqrt{2}}\gamma_k (s_k{}^x - s_k'{}^x) + \frac{\epsilon}{\sqrt{2}}\delta_k (s_k{}^z - s_k'{}^z) \right) \right\}. \quad (13)$$

For every pair of states $\mathbf{s}=\mathbf{A}$ and $\mathbf{s}'=\mathbf{B}$ there is the reciprocal pair of states $\mathbf{s}=\mathbf{B}$ and $\mathbf{s}'=\mathbf{A}$. Going from one of these pairs of states to the other leaves $\mathbf{s}+\mathbf{s}'$ unchanged while $\mathbf{s}-\mathbf{s}'$ changes sign. Therefore, since $\sinh(\mathbf{x})$ is an odd function of $\mathbf{x}$ the integrations over the sinh functions is zero, and we are left only with the integrations over the products of the cosh functions.

Expanding the cosh terms and considering each site separately (hence we drop the subscripts), we have for the single site integration in Eq. (13)

$$\int\cdots\int f(\mathbf{s})f(\mathbf{s}')\,d\mathbf{s}\,d\mathbf{s}'$$
$$\times\sum_{n=0}^{\infty}\frac{\{[(1/\sqrt{2})\alpha+h^x](s^x+s'^x)+[(\epsilon/\sqrt{2})\beta+h^z](s^z+s'^z)\}^{2n}}{(2n)!}$$
$$\times\sum_{m=0}^{\infty}\frac{[(1/\sqrt{2})\gamma(s^x-s'^x)+(\epsilon/2)\delta(s^z-s'^z)]^{2m}}{(2m)!}\,. \qquad (14)$$

Expanding the separate terms by the binomial theorem and using the angle variables θ and φ, where $s^x=\cos\theta$, $s^z=\sin\theta$, $s'^x=\cos\varphi$, and $s'^z=\sin\varphi$,

$$\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\sum_{p=0}^{2n}\sum_{q=0}^{2m}\frac{(\frac{1}{2})^{n+m}}{(2n)!(2m)!}\binom{2n}{p}\binom{2m}{q}(\alpha+\sqrt{2}\,h^x)^{2n-p}(\epsilon\beta+\sqrt{2}\,h^z)^p$$
$$\times(\gamma)^{2m-q}(\epsilon\delta)^q\int_{-\pi}^{\pi}d\theta\int_{-\pi}^{\pi}d\phi\,(\cos\theta+\cos\phi)^{2n-p}$$
$$\times(\sin\theta+\sin\phi)^p(\cos\theta-\cos\phi)^{2m-q}(\sin\theta-\sin\phi)^q. \qquad (15)$$

Considering now only the θ and φ integrations and expanding each term by the binomial theorem, we have

$$\sum_{a=0}^{2n-p}\sum_{b=0}^{p}\sum_{c=0}^{2m-q}\sum_{d=0}^{q}\binom{2n-p}{a}\binom{p}{b}\binom{2m-q}{c}\binom{q}{d}(-1)^{c+d}$$
$$\times\int_{-\pi}^{\pi}d\theta\,(\cos\theta)^{(2m-q)-c+(2n-p)-a}\,(\sin\theta)^{p-b+q-d}$$
$$\times\int_{-\pi}^{\pi}d\phi\,(\cos\phi)^{a+c}(\sin\phi)^{b+d}. \qquad (16)$$

For the φ integral, if $b+d=$ odd number, then the integral is zero. Therefore, we require $b+d=$ even number, and hence b and d are both even or both odd. Similar reasoning for the θ integral shows $(q+p)-(b+d)$ must be an even number. Since $b+d$ is even, $q+p$ is even, and therefore q and p are either both odd or both even. Therefore, also, $(2m-q)$ and $(2n-p)$ are both even (odd) if p and q are both even (odd). Hence the $(\alpha+\sqrt{2}\,h^x)$'s, γ's, $(\epsilon\beta+\sqrt{2}\,h^z)$'s, and δ's which are to the powers $2n-p$, $2m-q$, p, and q respectively are either all to an odd power or all to an even power. However, as yet we know nothing of the sign of the coefficients given by the integrals over θ and φ.

Using the addition formulas of sin and cos functions, one can rewrite the θ and φ integral in Eq. (15) as

$$(-1)^{2m-q}\int d\theta\int d\phi\,[\cos\tfrac{1}{2}(\theta+\phi)]^{2n-p+q}[\cos\tfrac{1}{2}(\theta-\phi)]^{2n}$$
$$\times[\sin\tfrac{1}{2}(\theta+\phi)]^{2m-q+p}[\sin\tfrac{1}{2}(\theta-\phi)]^{2m}, \qquad (17)$$

where due to the restrictions on p and q found above each square bracketed term is to an even power and hence the integral is positive. Therefore the coefficients are composed of positive terms, binomial coefficients, factorials, etc., except for the $(-1)^{2m-q}$ found in Eq. (17). If $(2m-q)=$ odd number (forcing p to be odd), then we have a negative coefficient, or if $2m-q$ is an even number (meaning p also even), we have a positive coefficient. For each site then we have a general series of terms of the form

$$\Theta(\alpha+\sqrt{2}\,h^x)^a(\epsilon\beta+\sqrt{2}\,h^z)^b(\gamma)^c(\epsilon\delta)^d$$
$$-\Psi(\alpha+\sqrt{2}\,h^x)^e(\epsilon\beta+\sqrt{2}\,h^z)^f(\gamma)^g(\epsilon\delta)^h, \qquad (18)$$

where a, b, c, and d are even nonnegative integers, e, f, g, and h are odd nonnegative integers, and where Θ and ψ are nonnegative. This same form is retained when taking the product over all sites i, $i=1,2,\ldots,N$. Therefore, using the basic relation expressed in Eq. (4), one has the results (a), (b), and (c) of the theorem.

IV. EXPLICIT INEQUALITIES AND APPLICATIONS

Theorem 1 is stated in terms of the expectation value, $E\{D\}$, of a product of derivative operators in the rotated variables α, β, γ, and δ. We now show how these results when written in terms of the initial x, x', z, and z' variables give GKS type inequalities as well as additional new correlation inequalities.

First we prove the following inequalities using Theorem 1:

$$\langle\prod_{i\in A}(s_i^x)\prod_{j\in B}(s_j^x)\rangle-\langle\prod_{i\in A}(s_i^x)\rangle\langle\prod_{j\in B}(s_j^x)\rangle\geq 0, \qquad (19)$$

$$\langle\prod_{i\in A}(s_i^z)\prod_{j\in B}(s_j^z)\rangle-\langle\prod_{i\in A}(s_i^z)\rangle\langle\prod_{j\in B}(s_j^z)\rangle\geq 0. \qquad (20)$$

The derivative operator terms of Eq. (7), $FG-FG'$, which give these correlation functions are respectively

$$\prod_{i\in A}\frac{\partial}{\partial x_i}\left[\prod_{j\in B}\frac{\partial}{\partial x_j}-\prod_{j\in B}\frac{\partial}{\partial x_j'}\right]=\prod_{i\in A}\left[\frac{1}{\sqrt{2}}\left(\frac{\partial}{\partial\alpha_i}+\frac{\partial}{\partial\gamma_i}\right)\right]$$
$$\times\left\{\prod_{j\in B}\left[\frac{1}{\sqrt{2}}\left(\frac{\partial}{\partial\alpha_j}+\frac{\partial}{\partial\delta_j}\right)\right]-\prod_{j\in B}\left[\frac{1}{\sqrt{2}}\left(\frac{\partial}{\partial\alpha_j}-\frac{\partial}{\partial\gamma_j}\right)\right]\right\}, \qquad (21)$$

$$\prod_{i\in A}\frac{\partial}{\partial z_i}\left(\prod_{j\in B}\frac{1}{\epsilon}\frac{\partial}{\partial z_j}-\prod_{j\in B}\frac{1}{\epsilon}\frac{\partial}{\partial z_j'}\right)=\prod_{i\in A}\left[\frac{1}{\epsilon\sqrt{2}}\left(\frac{\partial}{\partial\beta_i}+\frac{\partial}{\partial\delta_i}\right)\right]$$
$$\times\left\{\prod_{j\in B}\left[\frac{1}{\epsilon\sqrt{2}}\left(\frac{\partial}{\partial\beta_j}+\frac{\partial}{\partial\delta_j}\right)\right]-\prod_{j\in B}\left[\frac{1}{\epsilon\sqrt{2}}\left(\frac{\partial}{\partial\beta_j}-\frac{\partial}{\partial\delta_j}\right)\right]\right\}, \qquad (22)$$

where the right-hand side of Eqs. (21) and (22) are in terms of the rotated variables of Theorem 1. Equation (21) [(22)] consists of a summation of terms of the form of Eq. (9) each with a plus sign (all negative terms cancel). Furthermore, each remaining term in the expression for Eq. (21) [(22)] has $N_4=0$ [$N_3=0$], and therefore we have that each term is covered by either statement (a) [(a)] or (c) [(c)] of the theorem. Therefore the total expression for Eq. (21) [(22)] consists of a sum of nonnegative terms and therefore is itself nonnegative.

One can also prove the following type inequalities:

$$\langle\prod_{i\in A}(s_i^x)\prod_{j\in B}(s_j^z)\rangle-\langle\prod_{i\in A}(s_i^x)\rangle\langle\prod_{j\in B}(s_j^z)\rangle\leq 0. \qquad (23)$$

The proper derivative expression for this correlation is

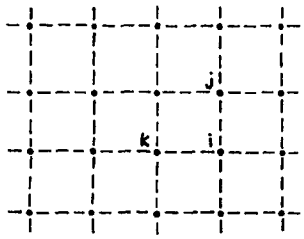


FIG. 1.

$$\prod_{i\in A}\frac{\partial}{\partial x_i}\left(\prod_{j\in B}\frac{\partial}{\epsilon\partial z_j}-\prod_{j\in B}\frac{\partial}{\epsilon\partial z_j'}\right)=\prod_{i\in A}\frac{1}{\sqrt{2}}\left(\frac{\partial}{\partial\alpha_i}+\frac{\partial}{\partial\gamma_i}\right)$$
$$\times\left\{\prod_{j\in B}\left[\frac{1}{\epsilon\sqrt{2}}\left(\frac{\partial}{\partial\gamma_j}+\frac{\partial}{\partial\delta_j}\right)\right]-\prod_{j\in B}\left[\frac{1}{\epsilon\sqrt{2}}\left(\frac{\partial}{\partial\gamma_j}-\frac{\partial}{\partial\delta_j}\right)\right]\right\}, \tag{24}$$

where again on the right-hand side all terms proceeded by a minus sign cancel. Looking at the terms in the curly brackets, we have only terms remaining which contain an odd number of $(\partial/\partial\delta_j)$'s, i.e., N_4 = odd number. Since there are no other $\partial/\partial\delta_j$ type terms, only statements (b) and (c) apply to the terms of Eq. (24), and the total expression therefore has a nonpositive expectation value.

These inequalities can be used as the original GKS inequalities to prove the existence of the thermodynamic limit of the correlation functions. The boundary conditions obtainable with these model systems extends those previously obtainable for the similar systems of Ginibre.[6] Now an external field in both x and z direction can be applied. Thus, for example, for the plane rotator, one can fix the spin to be in any direction in the x–z plane, not simply the z direction.

One of the physical consequences of Eqs. (19), (20), and (23) when taking $A=\{i\}$ and $B=\{j,k\}$ is that increasing or adding any interaction along the same direction as that of the magnetization, s_i^x or $s_i^{x'}$ increases the magnetization while increasing or adding any interaction in the opposite direction decreases the magnetization. Therefore, if one can show, for example, there exists a spontaneous magnetization in the x direction, m^{x*} for any decrease of ϵ in Eq. (1), one has the guaranteed continued existence of m^{x*}.

Besides the GKS type inequalities we can obtain new correlation inequalities by writing any derivative operator which has the form of Eq. (9) but which does not directly correspond to a GKS inequality when written in terms of the x's, x''s, z's, and z''s. In the case of a derivative product involving only two partial derivatives one has only first or second GKS inequalities. With three partial differentiations one has a number of new inequalities. For example, $(1/\sqrt{2})^3(\partial/\partial\alpha_i)(\partial/\partial\gamma_j)(\partial/\partial\gamma_k)$ gives one the inequality

$$\langle s_i^x s_j^x s_k^x\rangle-\langle s_k^x\rangle\langle s_i^x s_j^x\rangle-\langle s_j^x\rangle\langle s_i^x s_k^x\rangle+\langle s_i^x\rangle\langle s_i^x s_k^x\rangle\geq 0. \tag{25}$$

A similar expression for the z components of the spin can be obtained by interchanging β_i with α_i, δ_j with γ_j, and δ_k with γ_k. Inequalities of this form may in some cases be stronger than the usual GKS inequalities. One can select sites i, j, and k as in Fig. 1. Then if one thinks of a system with nearest neighbor interactions only or any interaction decreasing with the distance between sites, one is intuitively lead to believe for nonzero temperatures

$$\langle s_i^x s_j^x\rangle>\langle s_j^x s_k^x\rangle,\quad \langle s_i^x s_k^x\rangle>\langle s_j^x s_k^x\rangle, \tag{26}$$

(For the two-dimensional Onsager–Ising model this type behavior can be shown explicitly.[12]) Also one imagines $\langle s_i^x\rangle=\langle s_j^x\rangle=\langle s_k^x\rangle$. By defining $m=\langle s_i^x\rangle$, inequality (25) can be written as

$$\langle s_i^x s_j^x s_k^x\rangle-\langle s_k^x\rangle\langle s_i^x s_j^x\rangle\geq m[\langle s_i^x s_k^x\rangle-\langle s_j^x s_k^x\rangle]. \tag{27}$$

Hence, when $m>0$ and Eq. (26) holds, the right-hand side of (27) is a positive number rather than the zero of the GKS inequality.

Other inequalities involving three partial differentiations can be found. Corresponding to $[(1/\epsilon^2)(1/\sqrt{2})^3(\partial/\partial\alpha_i)(\partial/\partial\delta_j)(\partial/\partial\delta_k)]$, one has

$$\langle s_i^x s_j^{x'} s_k^{x'}\rangle-\langle s_i^x s_j^{x'}\rangle\langle s_k^{x'}\rangle-\langle s_i^x s_k^{x'}\rangle\langle s_j^{x'}\rangle+\langle s_i^x\rangle\langle s_j^{x'} s_k^{x'}\rangle\geq 0, \tag{28}$$

or, taking $[(1/\epsilon^2)(1/\sqrt{2})^3(\partial/\partial\gamma_i)(\partial/\partial\beta_j)(\partial/\partial\delta_k)]$, one has

$$\langle s_i^x s_j^{x'} s_k^{x'}\rangle-\langle s_i^x s_j^{x'}\rangle\langle s_k^{x'}\rangle+\langle s_i^x s_k^{x'}\rangle\langle s_j^{x'}\rangle-\langle s_i^x\rangle\langle s_j^{x'} s_k^{x'}\rangle\leq 0. \tag{29}$$

Together (28) and (29) give

$$\langle s_i^x\rangle\langle s_j^{x'} s_k^{x'}\rangle-\langle s_j^{x'}\rangle\langle s_i^x s_k^{x'}\rangle\geq 0. \tag{30}$$

Other inequalities involving four, five, etc., sites can be generated in a similar manner.

*Supported by National Science Foundation Grant GP-36564-X1.

[1] See the review article R. B. Griffiths, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, London, 1972), Vol. I.
[2] B. Simon, *The $P(\Phi)_2$ Euclidean (Quantum) Field Theory* (Princeton U. P., Princeton, N. J., 1974).
[3] J. L. Monroe and A. J. F. Siegert, J. Stat Phys. 10, 237 (1974).
[4] J. L. Monroe, J. Math. Phys. 15, 998 (1974).
[5] R. S. Ellis and J. L. Monroe, "A Simple Proof of the GHS and Further Inequalities," to appear in Commun. Math. Phys.
[6] J. Ginibre, Commun. Math. Phys. 16, 310 (1970).
[7] R. B. Griffiths, J. Math. Phys. 8, 478, 484 (1967).
[8] D. G. Kelly and S. Sherman, J. Math. Phys. 9, 466 (1968).
[9] C. Newman (private communication).
[10] R. B. Potts, Proc. Camb. Phil. Soc. 48, 106 (1952).
[11] H. Cramer, *Mathematical Methods in Statistics* (Princeton U. P., Princeton, N. J., 1951), p. 118.
[12] B. M. McCoy and T. T. Wu, *The Two Dimensional Ising Model* (Harvard U. P., Cambridge, Mass., 1973), p. 202.

Dilatations and the Poincaré group

Luis J. Boya

Departamento de Física Teórica, Facultad de Ciencias, Universidad de Zaragoza, G.I.F.T., Zaragoza, Spain

Jose F. Cariñena and Mariano Santander

Departamento de Física Teórica, Facultad de Ciencias, Universidad de Valladolid, G.I.F.T., Valladolid, Spain
(Received 8 May 1975)

We discuss the projective unitary representations of the Weyl group (Poincaré group enlarged with dilatations).

1. INTRODUCTION

Conformal invariance is actively investigated nowadays; it implies, besides Poincaré invariance, dilatations and "special conformal transformations."[1] Since scale invariance guarantees by itself the conservation of the special generators (in some sense; see Refs. 1 and 2), a first step to understand a conformal quantum theory is to develop a "Wigner program" for the "Weyl group" of dilatations plus Poincaré group, i.e., the search of projective unitary irreducible representations of this group; this is the aim of the present paper.

In Sec. 2 we define the group; in Sec. 3 and 4 we study and characterize the pertinent representations; Sec. 5 contains some comments on the explicit realizations.

2. THE GROUP $\mathcal{Y}$

Let D be the one-parametric Lie group of multiplication of a four-vector by a positive real number. It is an Abelian, connected and simply connected group, but a noncompact group, which is isomorphic to the additive group $\mathbb{R}$ of the real numbers.

Let $\Delta = D \otimes \mathcal{L}_0$ be the direct product of the homogeneous connected Lorentz group $\mathcal{L}_0$ by the former group D of dilatations.

If T_4 denotes the four-dimensional translation group and we consider the natural action of $D \otimes \mathcal{L}_0 = \Delta$ over T_4, we will call $\mathcal{Y}$ the semidirect product of groups: $\mathcal{Y} = T_4 \odot (D \otimes \mathcal{L}_0)$. This group $\mathcal{Y}$ does not keep distances, but does operate causally, i.e., keeps the nature of spacelike, timelike, and lightlike vectors, as well as future and past (when compatible); in fact it is a "maximal causal group" for relativity (Zeeman[3]).

Then, when an origin is fixed, the coordinates of a point and the corresponding image by the transformation (a, λ, Λ) of $\mathcal{Y}$ are related by

$$x'^{\mu} = \lambda \Lambda^{\mu}{}_{\nu} x^{\nu} + a^{\mu}.$$

3. THE PROJECTIVE UNITARY REPRESENTATIONS OF $\mathcal{Y}$

If $\mathcal{Y}$ is considered as the symmetry group in a quantum theory, one must find the projective (anti-) unitary irreducible representations of $\mathcal{Y}$. But as $\mathcal{Y}$ is a connected Lie group, a continuous representation cannot be (projective) antiunitary. The method of studying the projective unitary irreducible representations (PUIR) of a connected Lie group G is developed in Ref. 4, where it is shown that all of the PUIR of G can be deduced from the UIR of a group $\bar{G}$ called projective covering group of G, which is the middle group of a well-defined central extension of the universal covering group G^* of G, by the second cohomology group $H^2{}_0(LG, \mathbb{R})$ of the Lie algebras LG and $\mathbb{R}$ of the groups G and $U(1)$, relative to the trivial action of LG over $\mathbb{R}$.

In the simplest cases in which $H^2{}_0(LG, \mathbb{R}) = 0$, the projective covering group coincides with the universal covering group. This is the case that we are considering as we shall justify.

In fact, let $\mathcal{Y}$ be the group defined in Sec. 2. The corresponding Lie algebra $L\mathcal{Y}$ is generated by the infinitesimal generators (non-Hermitian) $M_{\mu\nu}, P_{\mu}$, and D. The commutation relations in $L\mathcal{Y}$ are those of the Poincaré Lie algebra, plus the following ones:

$$[M_{\mu\nu}, D] = 0, \quad [P_{\mu}, D] = -P_{\mu}.$$

All central extensions of $L\mathcal{Y}$ by $\mathbb{R}$ can be obtained in a simple way by using the method proposed by Levy-Leblond.[5] All the extensions of Poincaré Lie algebra by $\mathbb{R}$ are trivial extensions, so that we can say that any central extension of $L\mathcal{Y}$ by $\mathbb{R}$ is generated by $M_{\mu\nu}, P_{\mu}, D$, and a new generator I. We must consider the same commutation relations that in the Poincaré case, plus the new ones:

$$[M_{\mu\nu}, D] = m_{\mu\nu} I, \quad [P_{\mu}, D] = -P_{\mu} + p_{\mu} I,$$

where $m_{\mu\nu} = \Xi(M_{\mu\nu}, D)$ and $p_{\mu} = \Xi(P_{\mu}, D)$ are the values of the "cocycle" Ξ associated to the extension. As a consequence of the associative character of the composition law in the algebra (d is the boundary operator),

$$d\Xi(A, B, C) = \Xi([A, B], C) + \Xi([B, C], A) + \Xi([C, A], B) = 0$$

holds for all $A, B, C \in L\mathcal{Y}$

If we consider $d\Xi(D, M_{\mu\nu}, M_{\lambda\sigma}) = 0$, we obtain easily

$$\Xi([M_{\mu\nu}, M_{\lambda\sigma}], D) = 0.$$

Here, the explicit form of this commutator is irrelevant, but such a commutator is a linear combination of the generators $M_{\alpha\beta}$ which is reduced to a single $M_{\nu\sigma}$ (up to a constant) when $\mu = \lambda$. Then we obtain in this case $\Xi(M_{\nu\sigma}, D) = 0$, that is to say, $m_{\nu\sigma} = 0$. An analogous argument, with $d\Xi([D, M_{\lambda\mu}], P_{\nu}) = 0$ leads to $p_{\nu} = 0$. Then every extension of $L\mathcal{Y}$ by $\mathbb{R}$ is a trivial one, and we can summarize as follows:

Theorem 3.1: Every PUR of the group $\mathcal{Y}$ can be lifted

to an UR of the universal covering group $\mathcal{L}^*$ of $\mathcal{L}$. Conversely those unitary representations of $\mathcal{L}^*$ mapping the kernel of the covering homomorphism in $U(1)$ give rise to projective unitary representations of $\mathcal{L}$.

4. CHARACTERIZATION OF UNITARY REPRESENTATIONS OF $\mathcal{L}$

Now, we must characterize all IUR of the universal covering group of $\mathcal{L}$ because any IUR of $\mathcal{L}^*$ does satisfy the condition "mapping the kernel of the covering homomorphism in $U(1)$." In fact, the Schur lemma tells us that this kernel, being a central subgroup of $\mathcal{L}^*$, is necessarily mapped onto $U(1)$, for every IUR.

This problem can be solved following the Wigner–Mackey method which can be applied directly in the search of the IUR of any regular semidirect product with an Abelian kernel.[6] In fact, the group $\mathcal{L}^*$ is similar to $\mathcal{L}$, but $\mathcal{L}_0$ must be "replaced" by $SL(2,\mathbb{C})$, this last acting on T_4 via covering homomorphism $SL(2,\mathbb{C}) \to \mathcal{L}_0$. Therefore, $\mathcal{L}^*$ is a regular semidirect product

$$\mathcal{L}^* = T_4 \odot (D \otimes SL(2,\mathbb{C}))$$

with the Abelian kernel $T_4 \approx \mathbb{R}^4$. The dual group $\hat{T}_4$ of T_4 is also (isomorphic to) $\mathbb{R}^4$. Under the action of $D \otimes SL(2,\mathbb{C})$, T_4 breaks up into six orbits: $\{0\}$, $V^{\pm}$, $\Omega^{\pm}$, Ω_i, where $V^{\pm}$ is the future (past) lightcone, $\Omega^{\pm}$ the inside and Ω_i the outside of V.

For the orbits $\Omega^{\pm}$, Ω_i the little groups can be easily calculated, and we find

$$G_{\Omega^{\pm}} \approx SU(2), \quad G_{\Omega_i} \approx SU(1,1).$$

They are the same isotopy groups as for the Poincaré case. This result is that we have to expect: Any dilatation $\lambda \neq 1$ does change the (Minkowsky) length of any spacelike or timelike vector; therefore, the corresponding little groups cannot contain any dilatation with $\lambda \neq 1$. But for the light cone, the preceding argument does not exclude dilatations of the little group, because they keep the length of the lightlike vectors. In fact, let $\hat{p} = (0,0,1,1)$ the standard point on V^+. The general form of an element in $D \otimes SL(2,\mathbb{C})$ can be taken as a 2×2 matrix M

$$M = \begin{pmatrix} \sqrt{\lambda} & \\ & \sqrt{\lambda} \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix},$$

where $\lambda \in \mathbb{R}^+$, $\alpha, \beta, \gamma, \delta \in \mathbb{C}$, and $\alpha\delta - \beta\gamma = 1$.

If M leaves $\hat{p}$ invariant, we can easily conclude that M is as follows:

$$M = \begin{pmatrix} \exp(i\varphi/2) & \lambda \exp(-i\varphi/2) \\ 0 & \lambda \exp(-i\varphi/2) \end{pmatrix}$$

with $\lambda \in \mathbb{R}^+$, $a \in \mathbb{C}$, $\varphi \in \mathbb{R}$.

The structure of this little group can be easily seen: Its law of composition is

$$M(a,\lambda,\varphi)\, M(b,\mu,\theta) = M(a + (1/\lambda)\exp(-i\varphi)\, b, \lambda\mu, \varphi + \theta)$$

Because of this law, it is clear that the subgroup $T_2 = \{M(a,1,0)\}$ is an invariant one, $A = \{M(0,\lambda,\varphi)\}$ is an subgroup and both determines a semidirect product structure on $G_{\{V^+\}}$ whose action is (λ,φ): $a \mapsto \lambda^{-1} e^{-i\varphi} a$, where a is a shorthand symbol for $(a,1,0)$ and (λ,φ) for $(0,\lambda,\varphi)$. Then, the little group of the light cone is $G_{\{V^+\}} \approx T_2 \odot (\mathbb{R}^+ \otimes \widetilde{U(1)})$, where $\widetilde{U(1)}$ stands for the first covering (twofold) of $U(1)$ and $\mathbb{R}^+$ is the multiplicative group of positive real numbers. To see more clearly the relation of this little group with the first covering of the Euclidean group $E(2)$ which appears in the case of Poincaré group, the little group $G_{\{V^+\}}$ can be seen to be isomorphic to the semidirect product $\widetilde{E(2)} \odot \mathbb{R}^+$ with respect to the action λ: $(a,\varphi) \mapsto (\lambda^{-1} a, \varphi)$.

To visualize this little group, we can say that, on the standard lightlike vector, some Lorentz transformations act as a dilatation, and can be "compensated" by a suitable real dilatation. The real number $\lambda \in \mathbb{R}^+$ parametrizes such a "compensated" dilatation.

Obviously, the little group of the trivial orbit $\{0\}$ is the direct product $D \otimes SL(2,\mathbb{C})$ and its representations are product of the any two corresponding to D and $SL(2,\mathbb{C})$, which are well known.

The following step is to find the irreducible unitary representations of these little groups.

We comment only the case of $G_{\{V^+\}}$. As this group is itself a regular semidirect product with Abelian kernel, the Wigner–Mackey technique is to be applied. It is clear that $\widetilde{U(1)} \otimes \mathbb{R}^+$ produces in $\hat{T}_2$ only two orbits, which we call $\{0\}$ and $C = \hat{T}_2 - \{0\}$. In the first case, the little group is $\widetilde{U(1)} \otimes \mathbb{R}^+$; in the second, Z_2 [because the double covering $\widetilde{U(1)} \to U(1)$]. The irreducible unitary representations of $\widetilde{U(1)} \otimes \mathbb{R}^+$ are well known, and are characterized by $n/2$ ($n \in \mathbb{Z}$) and $p \in \mathbb{R}$; in the case of Z_2, there are only two representations which are denoted by $\epsilon = \pm$.

All IUR of $G_{\{V^+\}}$ are given below.

Orbit	Representations	
$\{0\}$	$[p, n/2]$	$p \in \mathbb{R}$, $n \in \mathbb{Z}$
C	$[\epsilon]$	$\epsilon = \pm$

Then, all the IUR of $\mathcal{L}^*$ are given by

Orbit	Representations	
$\Omega^{\pm}$	$[\pm; n/2]$	$n \in \mathbb{Z}$
Ω_i	no physical sense	
$\{0\}$	no physical sense, but presumably include the vacuum	
$V^{\pm}$	$[\pm; \{0\}; p, n/2]$ $[\pm; C; \epsilon]$	$p \in \mathbb{R}$, $n \in Z$ $\epsilon = \pm$

As stated in Theorem 3.1, these are also the PUIR of the group $\mathcal{L}$. As we can see, for the $\Omega^{\pm}, \Omega_i$ orbits, the label $m \in \mathbb{R}^+$ dissappears; for $V^{\pm}$ there appears a "dimension label" $p \in \mathbb{R}$ in one series and the label Ξ of "continuous spin" disappears in the other series; finally for the orbit $\{0\}$ there appears also a "dimension label" $p \in \mathbb{R}$.

5. REALIZATION OF THE REPRESENTATIONS

As is well known, the Mackey–Wigner technique pro-

vides not only a characterization, but also a *construction* of the induced representations. In our case we can take the same procedure that for the Poincaré group with only this difference: the invariant Haar measure induce now only a *quasi-invariant* measure, so that the representation acquires a factor (the square root of the Radon–Nikodym derivative; see Simms.[7])

As an example take the Ω^+ case; the Hilbert space is $L^2(\Omega^+ \to V_{2s+1}(\mathbb{C}))$, with scalar product

$$\langle \psi_1, \psi_2 \rangle = \int \frac{d^3\mathbf{p}}{p^0} \psi_1^*(p) \psi_2(p)$$

and the representation is

$$[U(a, \lambda, \Lambda)\psi](p) = (1/\lambda) \exp(-ipa) D_s(A_0) \psi(\lambda^{-1}\Lambda^{-1}p)$$

so that unitarity is automatic [A_0 is in the little group $SU(2)$].

[1]A. Salam and G. Mack, Ann. Phys. (N.Y.) **53**, 174 (1969).
[2]S. Coleman and R. Jackiw, Ann. Phys. (N.Y.) **67**, 552 (1971).
[3]E.C. Zeeman, J. Math. Phys. **5**, 490 (1964).
[4]J.F. Carinena and M. Santander, J. Math. Phys. **16**, 1416 (1975); see also references therein.
[5]J.M. Levy-Leblond "Galilei group and Galilean invariance," in *Group Theory and Its Applications, Vol. II*, edited by E.M. Loebl (Academic, New York, 1971).
[6]M.A. Naimark, *Les rêpresentations linéaires du groupe de Lorentz* (Dunod, Paris, 1962).
[7]D.J. Simms, *Lie Groups and Quantum Mechanics*, Lecture Notes in Math., Vol. 52 (Springer, New York, 1968).

Properties of linear representations with a highest weight for the semisimple Lie algebras*

B. Gruber†

Physics Department, Southern Illinois University, Carbondale, Illinois 62901
and Physikalisches Institut, Universität Würzburg, D-87 Würzburg, Germany

A. U. Klimyk

Institute for Theoretical Physics, Academy of Sciences of the Ukrainian SSR, Kiev, USSR
(Received 13 January 1975)

A theory for the representations with a highest and/or lowest weight is given for the semisimple complex Lie algebras (and their real forms). These representations are either irreducible finite-dimensional, irreducible infinite-dimensional or reducible, but not completely reducible, infinite-dimensional (called elementary representations), depending upon the property of the associated highest (or lowest) weight Λ. No restriction is made to those representations of the semisimple Lie algebras which can be integrated to form representations of the corresponding Lie group. The algebra A_1 is chosen (Sec. III) as a simple and familiar example upon which, however, much of the proof for the results obtained for the theory of representations with a highest (and/or lowest) weight for the general case of a semisimple Lie algebra rests (Sec. IV). It is demonstrated that the irreducible representations $D(\Lambda)$ with a highest (and/or lowest) weight Λ of the semisimple Lie algebras decompose with respect to any (regularly) embedded subalgebra of the type A_1 in the manner that either (a) the subrepresentations subduced on A_1 are all irreducible finite-dimensional, or (b) all infinite-dimensional. If for case (b) the complex number $M^\alpha \equiv 2(M,\alpha)/(\alpha,\alpha)$, α the (simple) root of A_1 and M a weight of $D(\Lambda)$ extremal with respect to A_1, is not a nonnegative integer, then the representation subduced on A_1 is irreducible. If, however, M^α is a nonnegative integer, then a reducible but not completely reducible, representation is subduced on A_1. Based upon the results of Sec. IV a generalization of Freudenthal's formula is obtained in Sec. V, valid for irreducible infinite-dimensional representations with highest (or lowest) weight. In Sec. VI generalizations are given of Racah's recurrence relation for the multiplicity of weights, Weyl's character formula and Kostant's formula for the multiplicity of weights for infinite-dimensional irreducible representations with a highest (or lowest) weight of the semisimple Lie algebras. These formulas are derived utilizing theorems and lemmas obtained by Verma, I. M. Gel'fand, S. I. Gel'fand, Bernstein, Harish-Chandra and the results of Sec. IV. In Sec. VII some of the infinite-dimensional representations of the algebra A_2 are discussed as examples, employing the geometrical methods developed by Antoine and Speiser and by Biedenharn and others.

I. INTRODUCTION

The motivation for writing this article stems basically from the fact that, over the past few years, infinite-dimensional representations with highest weights have become of increased importance in physics. Among the applications of infinite-dimensional representations with highest weights of semisimple Lie algebras are the hydrogen atom,[1] the N-dimensional harmonic oscillator.[2] Also the so-called ladder representations,[3a] applied in particle physics,[3b] represent an example for the use of this type of representations in physics. Still other interesting applications are discussed in Ref. (4).

It was felt that a systematic study of the theory of infinite-dimensional representations with highest weights of the semisimple Lie algebras would be of interest in view of their increasing importance in physics. From this point of view, in addition to results obtained by the authors, results obtained by other scientists have been included in this article, in order to give as complete as possible a comprehensive theory for the infinite-dimensional representations with highest weights for the complex semisimple Lie algebras.

Infinite-dimensional representations with highest weights for the semisimple Lie algebras have been studied extensively by Harish-Chandra.[5] However, Harish-Chandra restricted his attention to those representations with a highest weight of real semisimple Lie algebras that can be integrated to form a representation of the corresponding Lie group. These representations have the property that $2(\Lambda, \alpha_i)/(\alpha_i, \alpha_i) = n_i$, n_i integers ≥ 0 or < 0, with Λ the highest weight of an irreducible representation and the α_i the simple roots of a semisimple Lie algebra G. This type of representation forms only a small subset of the set of representations with a highest weight. And, in fact, it is only those representations of this subset which have at most one of the integers $n_i < 0$ that can be integrated to representations of the corresponding group. This restriction is not maintained in this article and *all* representations with a highest weight of a semisimple complex Lie algebra will be considered, independent of whether they can be integrated to representations of the corresponding Lie group or whether they can be continued to merely a local representation of the corresponding Lie group.

In 1968 an article was published by Verma[6] which contains essential results for a theory of infinite-dimensional representations with a highest weight of semisimple Lie algebras. Some of these results will be quoted and utilized below. Verma, in proving some of his theorems had to use conjectures, based upon which the theorems would hold. One of these conjectures was shown to be not valid by Gel'fand and collaborators.[7] In fact, it was demonstrated by them that Theorem 1 of Verma, based on that conjecture, could not hold in the form in which it was formulated. Below it will be proved that *part* of Verma's Theorem 1, reformulated as Lem-

ma 5, nevertheless holds. Based on another conjecture, Verma derived a formula for the multiplicity of weights of irreducible representations with a highest weight of the semisimple Lie algebras. However, for the case of highest weights Λ, which have the property that $S(\Lambda+R)=\Lambda+R$, for some element S of the Weyl group W, $S\neq 1$, this formula is not correct (for infinite-dimensional representations Λ is *not* restricted to the fundamental domain). Counter examples can be easily constructed from the algebras of rank 2. Other formulas for the multiplicity of weights of representations with highest weights will be derived in Sec. V and Sec. VI.

Listed in some greater detail, this article consists in a study of the infinite-dimensional representations with highest weight of the complex semisimple Lie algebras, with the aim of

(1) determining the *invariance properties* of their weight diagrams with respect to the Weyl group,

(2) obtaining *recurrence relations* of the type of the formulas by Freudenthal and Kostant for the case of finite-dimensional representations,

(3) introducing *characters*, as elements of a commutative domain of integrity,

(4) the derivation of a formula for the character, which corresponds to Weyl's character formula for the case of finite-dimensional representations,

(5) the derivation of a formula for the multipicity of weights which corresponds to Kostant's formula (involving the partition function) for the case of finite-dimensional representations.

In Sec. II of this article definitions are given concerning Lie algebras and representations, not necessarily irreducible, with a highest weight. In particular, the notion of an elementary representation is introduced. Further, a theorem on the decomposition of an infinite-dimensional representation with highest weight with respect to its weight subspaces is quoted, in view of its use in later sections. Finally, the complex Lie algebra $L(3, C)$ and its real form $L(2, 1)$ are chosen in order to give some examples of infinite-dimensional representations with a highest weight.

Section III gives a summary of the theory of representations with a highest weight for the algebra A_1, corresponding to the Lie group $SU(2)$ and $SO(3)$. The purpose of this summary is twofold. First, it serves as another, and complete, example for the theory of representations with a highest weight of a complex Lie algebra through the derivation of *all* its representations with highest weight. Second, this summary is of vital importance for later applications, in particular in Sec. IV.

Section IV is the central part of this article and much of the rest of this article is based upon it. In particular, the results of Sec. IV permit the derivation of Freudenthal's formula for (infinite-dimensional) irreducible representations with a highest weight of a simple Lie algebra G. Theorem 3 of Sec. IV describes the decomposition of irreducible representations with a highest weight with respect to *all* subalgebras of the type $G^0+G^\alpha+G^{-\alpha}$, where G^0 is the Cartan subalgebra of the algebra G and G^α and $G^{-\alpha}$ are the subspaces of the Cartan decomposition of the algebra G corresponding to the roots α and $(-\alpha)$. Knowing this decomposition for any such subalgebra, with α corresponding to a "compact root" or a "noncompact root," yields in particular complete information concerning the multiplicity structure of these representations.

In Sec. V Freudenthal's formula is derived for irreducible representations with a highest weight for simple Lie algebras. The proof follows closely Jacobson's proof[8a] of Freudenthal's formula for the case of finite-dimensional representations and is based upon results obtained in Sec. IV.

In Sec. VI the concept of character is introduced for infinite-dimensional representations with a highest weight. Theorems derived by Gel'fand and Verma, pertinent for the derivation of results that are to follow, are quoted. Based upon these theorems the *recurrence relations* for the multiplicity of weights of the type of Racah and Kostant are proved, with some modifications, for the case of infinite-dimensional representations with a highest weight. Weyl's character formula is derived and, finally, Kostant's formula for the multiplicity of weights (involving the partition function) is obtained for infinite-dimensional representations with a highest weight. Some of the proofs for the theorems of this section have also been given by Dixmier.[8b]

In Sec. VII the algebra A_2 is chosen to demonstrate the meaning of the formulas obtained in Sec. VI on hand of an example. The graphical method employed is a continuation of the geometrical methods developed by Antoine and Speiser[9] for finite-dimensional representations and by Biedenharn and others[10] for infinite-dimensional representations.

II. REPRESENTATIONS WITH A HIGHEST (LOWEST) WEIGHT

Let G be a complex semisimple Lie algebra of rank l. Let $G=G^0+\sum_{\alpha\in\Gamma}G^\alpha$ denote the canonical decomposition of this algebra, where the sum goes over all roots α of the root system Γ. Then the algebra G can be generated by elements h_i, e_i, f_i, $i=1,2,\ldots,l$. Thereby the elements h_i form a basis for the Cartan subalgebra G^0 of the algebra G, while the e_i are elements of the (one-dimensional) subspaces G^{α_i} of G, corresponding to the simple roots α_i, $i=1,2,\ldots,l$, and the f_i are the elements of the root spaces $G^{-\alpha_i}$ of G. The elements h_i, e_i, and f_i can be chosen in a manner such that the following commutation relations hold,

$$[h_i, h_j]=0,$$
$$[h_i, e_j]=\alpha_j(h_i)e_j,$$
$$[h_i, f_j]=-\alpha_j(h_i)f_j,$$
$$[f_i, e_i]=h_i,$$
$$[f_i, e_j]=0 \text{ for } i\neq j.$$

Let ρ denote a linear representation of the algebra G in a linear space V. A representation ρ of an algebra G defines uniquely a representation $\rho(\Omega)$ of the universal enveloping algebra Ω of the algebra G.

A representation ρ of an algebra G is called a representation with a highest weight on the space V if the following conditions hold:

(1) There exists a vector x in the representation space V such that $\rho(h)x=\Lambda(h)x$ for every h, $h\in G$, where $\Lambda(h)$ is a linear form on G^0.

(2) $\rho(e_i)x=0, \quad i=1,2,\cdots,l.$

(3) $V=\{\rho(a)x/a\in\Omega\}$, i.e., the set of vectors $\rho(a)x$, for given x, coincides with V as the element a goes over Ω.

The linear form $\Lambda(h)$ is called the highest weight of the representation ρ.

The finite-dimensional irreducible representations of the semisimple complex Lie algebras are examples for representations with a highest weight.

In the following the definition for an *elementary representation* is given. For this purpose let $\Lambda\equiv\Lambda(h)$ denote some fixed linear form on G^0. For given Λ a representation ρ of the algebra G is constructed in the following manner:

Let G^- denote that subalgebra of G which is generated by the elements f_i of G. Let Ω_- denote the subalgebra of Ω which is generated by the subalgebra G^-. Then Ω_- is the universal enveloping algebra of G^-. The enveloping algebra Ω_- forms a linear space and the space is taken as the carrier space of a linear representation ρ of the algebra G. As a basis for the space Ω_- the following set of elements is chosen

$$1, f_{i_1}\times f_{i_2}\times\cdots\times f_{i_r}\equiv f_{i_1}f_{i_2}\cdots f_{i_r},$$

$$i_j=1,2,\ldots,l, \quad r=1,2,\ldots$$

with the condition that every pair of elements

$$f_{i_1}\cdots f_{i_{n-1}}f_{i_n}f_{i_{n+1}}f_{i_{n+2}}\cdots f_{i_r},$$

$$f_{i_1}\cdots f_{i_{n-1}}f_{i_{n+1}}f_{i_n}f_{i_{n+2}}\cdots f_{i_r},$$

is identified for which $[f_{i_n},f_{i_{n+1}}]=0$ holds. A representation ρ of the algebra G on the space Ω_- can then be defined as follows:

$$\rho(h)1=\Lambda 1,$$

$$\rho(f_i)1=f_i,$$

$$\rho(e_i)1=0,$$

$$\rho(h)f_{i_1}f_{i_2}\cdots f_{i_r}=(\Lambda-\alpha_{i_1}-\alpha_{i_2}-\cdots-\alpha_{i_r})\times(h)f_{i_1}f_{i_2}\cdots f_{i_r}, \tag{1}$$

$$\rho(f_i)f_{i_1}f_{i_2}\cdots f_{i_r}=f_if_{i_1}f_{i_2}\cdots f_{i_r},$$

$$\rho(e_i)f_{i_1}f_{i_2}\cdots f_{i_r}=f_{i_1}(\rho(e_i)f_{i_2}\cdots f_{i_r}),$$

$$-\delta_{i_1i}(\Lambda-\alpha_{i_2}-\cdots-\alpha_{i_r})(h_i)f_{i_2}\cdots f_{i_r},$$

$$i=1,2,\cdots,l,$$

where the δ_{ij} is the Kronecker Delta.

It is not difficult to verify that the formulas (1) indeed define a representation of the algebra G on the space Ω_-. This representation will be called an *elementary representation with a highest weight* and will be denoted by d_Λ.

There exists a one-to-one correspondence between complex linear forms Λ on G^0 and elementary representations d_Λ. The representations d_Λ are in general reducible, but not completely reducible. This implies that condition (2) of the definition of representations with a highest weight will, in general, hold also for other vectors of the space V. Thus the representation space V of d_Λ may have an invariant subspace V' such that the representation induced on the *factorspace* V/V' is an *irreducible* representation with highest weight $\Lambda(h)$.

Elementary representations play a useful and important role in the investigation of irreducible representations with a highest weight.

Below an important theorem on certain properties of linear representations with a highest weight is quoted for subsequent use in later sections.[11]

Theorem 1: Let ρ denote a linear representation of a semisimple complex Lie algebra G with highest weight Λ which acts on the vector space V. Then the space V decomposes into a direct sum of finite-dimensional subspaces V_M where V_M is the set of all vectors x of V for which $\rho(h)x=M(h)x$ for all $h\in G^0$. The subspace V_Λ is one-dimensional. The highest weight Λ is unique and every other weight of the representation ρ can be expressed in the form $\Lambda-\sum_{i=1}^{l}k_i\alpha_i$, where the α_i are the positive simple roots of G and the k_i nonnegative integers. Moreover, to every linear form Λ on G^0 there exists, up to equivalence, one and only one irreducible representation of G with Λ as highest weight.

An *irreducible* representation with highest weight Λ will be denoted by D_Λ. Representations with a lowest weight are related to irreducible representations with a highest weight through symmetry considerations. If $2(\Lambda,\alpha_i)/(\alpha_i,\alpha_i)$ is a nonnegative integer for every simple root α_i, then and only then is D_Λ a finite-dimensional representation (i.e., a representation with a highest as well as a lowest weight).

The dimension of the subspace V_M of Theorem 1 is called the multiplicity of the weight M in the representation D_Λ.

So far linear representations of *complex* semisimple Lie algebras have been considered. This restriction is, however, not necessary since there exists a one-to-one correspondence between the representations with a highest weight of a complex Lie algebra G and its real forms. This one-to-one correspondence will be demonstrated in what follows, taking the general linear group $GL(n,C)$ as an example.

The algebra of $GL(n,C)$ is denoted by $L(n,C)$. As basis for $L(n,C)$ the matrices e_{ik}, $i,k=1,2,\ldots n$, can be chosen with the matrix elements $(e_{ik})_{lm}=\delta_{il}\delta_{km}$, $l, m=1,2,\ldots,n$. The complex linear combinations of the e_{ik} constitute the algebra $L(n,C)$. The elements e_{ik} and $\sqrt{-1}e_{ik}$ are linearly dependent elements of this algebra, and in a representation of $L(n,C)$ the operators E_{ik} and $\sqrt{-1}E_{ik}$ correspond to these two elements. This

is in distinction to the case when the algebra $L(n, C)$ is considered to be real. Then the dimension of the algebra is twice the dimension of the "complex $L(n, C)$" with e_{ik} and $\sqrt{-1}\,e_{ik}$ as *basis elements*. The algebra consists in this case of all real linear combinations of the e_{ik} and $\sqrt{-1}\,e_{ik}$. However, in a representation the two operators representing the elements e_{ik} and $\sqrt{-1}\,e_{ik}$ are not necessarily related by $\sqrt{-1}$ as it is the case above. Representations of the complex Lie algebra will be called *complex* representations.

A complex Lie algebra G may contain real subalgebras with the property that their complex extension yields the complex Lie algebra G. Such subalgebras are called real forms of the complex algebra G. Therefore, a complex representation of the (complex) Lie algebra G will induce a representation of any of its real forms. Conversely, a representation of a real Lie algebra will generate a complex representation of the complex Lie algebra corresponding to it. The irreducibility of representations is conserved in this correspondence between the complex representation of the complex Lie algebras and the representations of its real forms.

It follows that to (irreducible) representations with a highest weight of a complex Lie algebra correspond (irreducible) representations of their real forms. These representations of the real forms are also called representations with a highest weight.

A simple example for representations with a highest weight is supplied by considering the (real) Lie algebra $L(2, 1)$ of the group $U(2, 1)$. Some of the representations of the discrete series of representations of the algebra $L(2, 1)$ are representations with a highest weight. The representations of the discrete series of $L(2, 1)$ are obtained by partitioning the number 2 into two nonnegative integers, p, q. To the three possible partitions (2, 0), (1, 1), and (0, 2) correspond the following three types of generalized Gel'fand Zetlin patterns,

$$\left\langle \begin{array}{ccccc} & m_{13} & & m_{23} & m_{33} \\ m_{12} & & m_{22} & & \\ & m_{11} & & & \end{array} \right\rangle \qquad \begin{array}{l} m_{12} \geq m_{13}+1 \\ m_{13}+1 \geq m_{22} \geq m_{33}+1 \\ m_{12} \geq m_{11} \geq m_{22} \end{array}$$

$$\left\langle \begin{array}{ccccc} & m_{13} & m_{23} & & m_{33} \\ m_{12} & & & m_{22} & \\ & & m_{11} & & \end{array} \right\rangle \qquad \begin{array}{l} m_{12} \geq m_{13}+1 \\ m_{33}-1 \geq m_{22} \\ m_{12} \geq m_{11} \geq m_{22} \end{array}$$

$$\left\langle \begin{array}{ccccc} m_{13} & & m_{23} & & m_{33} \\ & m_{12} & & m_{22} & \\ & & m_{11} & & \end{array} \right\rangle \qquad \begin{array}{l} m_{23}-1 \geq m_{12} \geq m_{33}-1 \\ m_{33}-1 \geq m_{22} \\ m_{12} \geq m_{11} \geq m_{22} \end{array}$$

with the m_{ik} integers. If a pattern is denoted by the symbol Θ and the basis vector corresponding to it by $\xi(\Theta)$, then the operators E_{kk}, $E_{k-1,k}$, $E_{k,k-1}$ which correspond to the elements e_{kk}, e_{k-1}, $e_{k,k-1}$ of $L(2,1)$ [and of $L(3, C)$, the complex extension of $L(2, 1)$] are given by

$$E_{k-1,k}\xi(\Theta) = \sum_{i=1}^{k-1} a_{k-1}^{i}(\Theta)\xi(\Theta_{k-1}^{i}) \tag{2}$$

$$E_{kk}\xi(\Theta) = \left(\sum_{i=1}^{k} m_{ik} - \sum_{i=1}^{k-1} m_{i,k-1}\right)\xi(\Theta), \tag{3}$$

$$E_{k,k-1}\xi(\Theta) = \sum_{i=1}^{k} b_{k-1}^{i}(\Theta)\xi(\Theta_{k-1}^{-i}), \tag{4}$$

where $a_{k-1}^{i}(\Theta)$, $b_{k-1}^{i}(\Theta)$ are numbers depending on the values m_{ik} and Θ_{k-1}^{i} is the pattern obtained from the pattern Θ by replacing $m_{i,k-1}$ by $m_{i,k-1}+1$ while Θ_{k-1}^{-i} is the pattern obtained from the pattern Θ by replacing $m_{i,k-1}$ by $m_{i,k-1}-1$. The formulas (2)—(4) have been given for the general case $L(p, q)$, $p+q=n$, with $i=1, 2, \ldots, n$ [and thus also hold for $L(n, C)$]. For $n=3$ and $p=2$, $q=1$, the example $L(2, 1)$ is obtained.

The matrices e_{kk}, $k=1, 2, \ldots, n$, form a basis for the Cartan subalgebra of the algebra $L(n, C)$. Therefore, the vectors $\xi(\Theta)$ are weight vectors with respect to this Cartan subalgebra. Analyzing the inequalities imposed on the numbers m_{ik} it is easy to recognize that the representations which correspond to the partitions $p=2+0$ and $p=0+2$ of $L(2, 1)$ are representations with a highest weight and a lowest weight, respectively. In general it can be observed that as a consequence of the conditions imposed on the m_{ik}, the representations of the discrete series of the algebra $L(p, q)$ which correspond to the partitions $p=p+0$ and $p=0+p$ are representations with a highest and lowest weight, respectively.

The "ladder representations" of the algebra $L(p, q)$, which are not representations of the discrete series,[3] are also representations with a highest (lowest) weight. The Lie algebra of the group $SO(p, q)$ contains among its representations of the discrete series also representations with a highest (lowest) weight.

III. REPRESENTATIONS WITH A HIGHEST WEIGHT OF THE SIMPLE ALGEBRAS OF RANK 1

In this section some known results are discussed concerning the theory of linear representations of the simple algebras of rank 1. These results are basic for the further development of the theory of representations of the semisimple Lie algebras as given in this article. Therefore, below a brief outline is given of the properties of representations of simple algebras of rank 1 upon which much of the following will rely upon.

Let G be a three-dimensional complex simple Lie algebra with basis e, f, h for which the following commutation relations hold:

$$[h, e]=2e, \quad [h, f]=-2f, \quad [f, e]=h. \tag{5}$$

Since the Cartan subspace of this algebra is one-dimensional, the linear form $\Lambda(h)$ is uniquely characterized by the complex number Λ.

Let ρ_Λ be a representation of G with Λ as highest weight and let x be the vector corresponding to this highest weight, $\rho_\Lambda(h)x=\Lambda x$. Then, since

$$\rho_\Lambda(h)(\rho_\Lambda(f)x) = (\rho_\Lambda(f)\rho_\Lambda(h) - 2\rho_\Lambda(f))x = (\Lambda-2)\rho_\Lambda(f)x, \tag{6}$$

the vector $\rho_\Lambda(f)x$ is again an eigenvector of $\rho_\Lambda(h)$ and thus is a weight vector corresponding to the weight $(\Lambda-2)$. Similarly, the vectors $(\rho_\Lambda(f))^2x$, $(\rho_\Lambda(f))^3x, \cdots$ are weight vectors corresponding to the weights $\Lambda-4$,

$\Lambda - 6, \cdots$, respectively. Continuing in this manner a sequence of weight vectors is obtained. This sequence may be finite or infinite.

Setting $(\rho_\Lambda(f))^i x \equiv x_i$, $i = 0, 1, 2, \cdots$, the following property holds:

$$\rho_\Lambda(e)x_j = (-j\Lambda + j(j-1))x_{j-1}. \tag{7}$$

Proof: This formula is correct for $j = 0$. Suppose it is correct for all values $j \leq i$. Then for $i = j + 1$,

$$\begin{aligned}\rho_\Lambda(e)x_i &= \rho_\Lambda(e)x_{j+1}\\ &= \rho_\Lambda(e)\rho_\Lambda(f)x_j\\ &= (\rho_\Lambda(f)\rho_\Lambda(e) - \rho_\Lambda(h))x_j\\ &= \rho_\Lambda(f)(-j\Lambda + j(j-1))x_{j-1} - (\Lambda - 2j)x_j\\ &= (-(j+1)\Lambda + j(j+1))x_j\\ &= (-i\Lambda + i(i-1))x_{i-1}.\end{aligned}$$

The operators $\rho_\Lambda(h)$, $\rho_\Lambda(e)$, $\rho_\Lambda(f)$ acting on a vector x_i of the set $\{x_i / i = 0, 1, 2 \cdots\}$ yield a multiple of another element of this set. For representations with a highest weight Λ therefore this set is invariant under the operators $\rho(h)$, $\rho(e)$, $\rho(f)$ and the most general element of this space which carries the representation ρ_Λ is obtained as a linear combination over the elements of this set.

If for some value i it holds that $\rho_\Lambda(f)x_i = 0$, then the representation is finite-dimensional and necessarily irreducible. In order that a representation is finite-dimensional a necessary (but not sufficient) condition is that the highest weight Λ is a nonnegative integer. (It should be emphasized that we are not restricting our attention to irreducible representations. Thus ρ_Λ may be a representation which is reducible but not completely reducible. For *irreducible* representations the condition is necessary *and* sufficient.)

A representation ρ_Λ, with Λ a complex number, may be infinite-dimensional (and is so necessarily except for Λ a nonnegative integer). A necessary and sufficient condition for an infinite-dimensional representation ρ_Λ to be *reducible* is that Λ is a nonnegative integer. This is shown as follows. If for every value $i = 0, 1, 2, \cdots$ it holds that $\rho_\Lambda(f)x_i \neq 0$, then the representation is infinite-dimensional. If in equation (7) the coefficient on the right-hand side becomes zero, namely if

$$-j\Lambda + j(j-1) = 0, \tag{8}$$

for some value $j \neq 0$, then the representation ρ_Λ is reducible. In this case the vectors x_j, $x_{j+1}, \cdots$ form the basis of an invariant subspace $\bar{V}$ of V. This is however the case for $\Lambda = j - 1$. The restriction of ρ_Λ on the invariant subspace $\bar{V}$ forms an *irreducible* infinite-dimensional representation with highest weight $-(\Lambda + 2)$. The weight vector corresponding to the highest weight is $x_{\Lambda+1}$. If, on the other hand, Λ is not a nonnegative integer, then an infinite-dimensional representation is irreducible. This can be seen as follows. An arbitrary vector of the representation space V is given by

$$y = \beta_{i_1}x_{i_1} + \beta_{i_2}x_{i_2} + \cdots + \beta_{i_n}x_{i_n},$$

where the β_{i_j} are complex numbers. Acting on such a vector successively with the operator $\rho_\Lambda(e)$ yields, after a finite number of steps, a vector βx_0 with β some complex number not equal to zero and x_0 the basis vector corresponding to the highest weight Λ. Then, acting on x_0 by operators of the form $(\rho_\Lambda(f))^n$ it is possible to obtain an arbitrary basis vector of V. Thus the following theorem holds:

Theorem 2: *Every* representation with a *highest weight* of a complex simple Lie algebra of rank 1 is equivalent to one of the following representations:

(1) Irreducible, finite-dimensional representation ρ_Λ. For this type of representations Λ is a nonnegative integer m. If the basis for the finite-dimensional vector space V is given by the vectors $x_0, x_1, x_2, \ldots, x_m$, then it holds

$$\begin{aligned}&\rho_\Lambda(h)x_i = (m - 2i)x_i, \quad i = 1, 2, \ldots, m,\\ &\rho_\Lambda(f)x_i = x_{i+1}, \quad i = 0, 1, 2, \ldots, m,\\ &\rho_\Lambda(f)x_m = 0,\\ &\rho_\Lambda(e)x_0 = 0,\\ &\rho_\Lambda(e)x_i = (-im + i(i-1))x_{i-1}, \quad i = 1, 2, \ldots, m.\end{aligned}$$

(2) Infinite-dimensional representation ρ_Λ. For these representations Λ is a complex number. If the basis for the infinite-dimensional vector space is given by $x_0, x_1, x_2, \cdots$, then it holds

$$\begin{aligned}&\rho_\Lambda(h)x_i = (\Lambda - 2i)x_i, \quad i = 0, 1, 2, \cdots,\\ &\rho_\Lambda(f)x_i = x_{i+1}, \quad i = 0, 1, 2, \cdots,\\ &\rho_\Lambda(e)x_0 = 0,\\ &\rho_\Lambda(e)x_i = (-i\Lambda + i(i-1))x_{i-1}, \quad i = 1, 2, \cdots.\end{aligned}$$

If Λ is a nonnegative integer then the infinite-dimensional representation is reducible. If Λ is not a nonnegative integer then the representation is irreducible.

A reducible infinite-dimensional representation ρ_Λ, $\Lambda = m$ a nonnegative integer, has in a basis $x_0, x_1, x_2, \cdots$ the matrix form

$$\begin{pmatrix}\gamma_1 & 0\\ * & \gamma_2\end{pmatrix}.$$

The submatrices γ_1 form a finite-dimensional irreducible representation with highest weight m while the submatrices γ_2 form an infinite-dimensional representation with a highest weight $-(m + 2)$.

IV. PROPERTIES OF IRREDUCIBLE REPRESENTATIONS WITH A HIGHEST WEIGHT

Let G denote a complex semisimple Lie algebra. Let D_Λ denote an irreducible representation of G with a highest weight Λ. Moreover, let G^0 denote an arbitrary, but fixed, Cartan subalgebra of G with basis elements h_i, $i = 1, 2, \ldots, l$. The basis elements h_i are chosen in such a manner that for every element $h \in G^0$ holds $\alpha_i(h) = (h, h_i)$, $i = 1, 2, \ldots, l$, where the bracket denotes the scalar product of the elements h, h_i, defined by the Killing—Cartan form, and where the α_i, $i = 1, 2, \ldots, l$, are a system of simple roots of G. We define $h_i' = 2h_i/(\alpha_i, \alpha_i)$. The Weyl group of G is denoted by W, the subset of all positive roots of G is denoted

by Δ. Let $\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_r}$ be a subset of roots of the set of simple roots for which holds $\Lambda(h'_{i_j}) = n_{ij}$, $j = 1, 2, \ldots, r$, with n_{ij} nonnegative integers. Then the subset of roots of Δ which are linearly dependent on these roots will be denoted by Δ_Λ. The subgroup of the Weyl group W which is generated by the reflections $S_{i_1}, S_{i_2}, \ldots, S_{i_r}$ on the hyperplanes perpendicular to the roots $\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_r}$ is denoted by W_Λ.

Lemma 1: Let D_Λ denote an irreducible representation of G with highest weight Λ. Let M denote an arbitrary weight of D_Λ. Then, if $\Lambda(h'_i)$ is a nonnegative integer, $S_i M = M - M(h'_i)\alpha_i$ is a weight of D_Λ.

The proof of this lemma is analogous to the proof of *Lemma* 7.3 in Ref. 8 and we refer to this reference.

Lemma 2: Let D_Λ denote an irreducible representation of G with highest weight Λ. If M is a weight of D_Λ, then SM, $S \in W$, is a weight of D_Λ.

The proof follows directly from lemma 1.

The operators which correspond in a representation D_Λ of the algebra G to the elements e_j, e_α (both, α_j and α denote positive roots), f_j, f_α, h are denoted by E_j, E_α, F_j, F_α, H, respectively.

In the following, a theorem will be proved. This theorem is the basis for the investigation of the properties of representations with a highest weight, reducible as well as irreducible, of the semisimple Lie algebras.

Theorem 3: Let D_Λ denote an irreducible representation of G with a highest weight Λ on a space R. For every α of Δ root vectors e_α, f_α are selected in such a manner that $[f_\alpha, e_\alpha] = h_\alpha$, where h_α is defined by $(h, h_\alpha) = \alpha(h)$, $h \in G^0$. If the representation D_Λ is restricted to the subalgebra $G_{(\alpha)} = G^0 + G^\alpha + G^{-\alpha}$, with G^α, $G^{-\alpha}$ root subspaces of G corresponding to the roots α and $-\alpha$, then the following two cases hold for the decomposition of the representation D_Λ with respect to the subalgebra $G_{(\alpha)}$:

(1) For $\alpha \in \Delta$, but $\alpha \notin \Delta_\Lambda$, the space R decomposes into a direct sum of infinite-dimensional subspaces, which are invariant under $G_{(\alpha)}$ (but not necessarily irreducible). For each of the subspaces exists a basis $y_0, y_1, y_2, \ldots$, such that

$$\begin{aligned} &Hy_i = (M - i\alpha)(h)y_i, \quad i = 0, 1, 2, \cdots, \\ &F_\alpha y_i = y_{i+1}, \quad i = 0, 1, 2, \ldots, \\ &E_\alpha y_0 = 0, \\ &E_\alpha y_i = \frac{(\alpha, \alpha)}{2} i(i - 1 - m)y_{i-1}, \quad i = 1, 2, \ldots . \end{aligned} \tag{9}$$

(2) For $\alpha \in \Delta$ *and* $\alpha \in \Delta_\Lambda$, the space R decomposes into a direct sum of finite-dimensional subspaces, which are invariant and irreducible under $G_{(\alpha)}$. For each of the subspaces exists a basis $y_0, y_1, \ldots, y_m$, such that

$$\begin{aligned} &Hy_i = (M - i\alpha)(h)y_i, \quad i = 1, 2, \ldots, m, \\ &F_\alpha y_i = y_{i+1}, \quad i = 0, 1, 2, \ldots, m - 1, \\ &F_\alpha y_m = 0, \\ &E_\alpha y_0 = 0, \\ &E_\alpha y_i = \frac{(\alpha, \alpha)}{2} i(i - 1 - m)y_{i-1}, \quad i = 1, 2, \ldots, m. \end{aligned} \tag{10}$$

In Eqs. (9) and (10) the symbol M denotes the weight of the vector y_0, and it holds $(\alpha, \alpha) = (h_\alpha, h_\alpha)$ and $m = 2(M, \alpha)/(\alpha, \alpha)$.

Unfortunately, a simple proof of this theorem has eluded us and we find it necessary to break up the proof in three main parts. It has also been found convenient to prove first Lemmas 3 to 5 before proving Theorem 3 itself.

Lemma 3: Let α and β be positive roots of G and s the largest nonnegative integer for which $s\alpha + \beta$ is a root of G. Then for $p \geq s$ the relation holds

$$F_\alpha{}^p F_\beta = F_\beta F_\alpha{}^p + \gamma_1 F_{\alpha+\beta} F_\alpha{}^{p-1} + \cdots + \gamma_s F_{s\alpha+\beta} F_\alpha{}^{p-s},$$

with the γ_i, $i = 1, 2, \ldots, s$, as numerical factors.

Proof: Up to a numerical factor it holds $[F_\alpha, F_\beta] = F_{\alpha+\beta}$. Hence, ignoring coefficients,

$$\begin{aligned} F_\alpha{}^p F_\beta &= F_\alpha{}^{p-1} F_\beta F_\alpha + F_\alpha{}^{p-1} F_{\alpha+\beta} \\ &= (F_\alpha{}^{p-2} F_\beta F_\alpha + F_\alpha{}^{p-2} F_{\alpha+\beta}) F_\alpha \\ &\quad + (F_\alpha{}^{p-2} F_{\alpha+\beta} F_\alpha + F_\alpha{}^{p-2} F_{2\alpha+\beta}). \end{aligned} \tag{11}$$

Continuing to substitute the first line of Eq. (11) for $F_\alpha{}^{p-i} F_\beta$ yields the desired result, since $F_{r\alpha+\beta} = 0$ for $r > s$.

Lemma 4: Let y denote a vector of D_Λ such that $Hy = M(h)y$, M some weight of D_Λ. Then the following relation holds:

$$\begin{aligned} &F_\alpha{}^{m+k} E_\alpha{}^k y \\ &\quad = (c_0 F_\alpha{}^m + c_1 E_\alpha F_\alpha{}^{m+1} + c_2 E_\alpha^2 F_\alpha{}^{m+2} + \cdots + c_k E_\alpha^k F_\alpha{}^{m+k}) y, \end{aligned} \tag{12}$$

where the c_i, $i = 0, 1, 2, \ldots, k$ are numerical factors.

Proof:

(1) The relation holds,

$$\begin{aligned} (F_\alpha E_\alpha^k) y &= (E_\alpha F_\alpha E_\alpha^{k-1} + E_\alpha^{k-1}) y \\ &= (E_\alpha^k F_\alpha + E_\alpha^{k-1}) y, \quad k \geq 1, \end{aligned} \tag{13}$$

as can be easily verified by successively substituting the left side of Eq. (13) into the right side. In this equation multiplicative factors are ignored since they are irrelevant for later considerations.

(2) The relation holds,

$$\begin{aligned} F_\alpha{}^m E_\alpha^k y &= F_\alpha{}^{m-1} (E_\alpha^k F_\alpha + E_\alpha^{k-1}) y \\ &= F_\alpha{}^{m-2} (F_\alpha E_\alpha^k F_\alpha + F_\alpha E_\alpha^{k-1}) y \\ &= F_\alpha{}^{m-2} (E_\alpha^k F_\alpha + E_\alpha^{k-1}) F_\alpha \\ &\quad + (E_\alpha^{k-1} F_\alpha + E_\alpha^{k-2}) y \\ &= F_\alpha{}^{m-2} E_\alpha^k F_\alpha^2 + E_\alpha^{k-1} F_\alpha + E_\alpha^{k-2} y \\ &\quad \cdot \\ &\quad \cdot \\ &\quad \cdot \\ &= (E_\alpha^k F_\alpha^m + E_\alpha^{k-1} F_\alpha^{m-1} + \cdots + E_\alpha^{m-k} F_\alpha) y, \end{aligned} \tag{14}$$

with k, $m \geq 1$ and $E^r = 0$ for $r < 0$. Again, multiplicative factors are ignored. For $m > k$, Eq. (12) is obtained.

Lemma 5: If a subspace R' of R is invariant under the algebra $G_{(\alpha)}$, then R' is also an invariant subspace with respect to the three-dimensional simple algebra $\tilde{G}_{(\alpha)} = G_\alpha^0 + G^\alpha + G^{-\alpha}$, where G_α^0 is the subspace of G^0 spanned by the vector h_α.

Proof: $\tilde{G}_{(\alpha)}$ is a subalgebra of $G_{(\alpha)}$.

Proof of Theorem 3

The proof of Theorem 3 is given in three parts. Part 1 is for the case that $2(\Lambda, \alpha)/(\alpha, \alpha) \neq n$, $n \geq 0$, integer. Part 2 considers the case with $2(\Lambda, \alpha)/(\alpha, \alpha) = n$, $n \geq 0$, integer and $\alpha \notin \Delta_\Lambda$. Part 3 treats the case with $2(\Lambda, \alpha)/(\alpha, \alpha) = n$, $n \geq 0$, integer and $\alpha \in \Delta_\Lambda$. These parts are subdivided into smaller units in order to keep the proof as transparent as possible.

An ordering is introduced into the set of weights of the (irreducible) representation D_Λ in the following manner. Let the set of simple positive roots $\Pi = \{\alpha_1, \alpha_2, \ldots, \alpha_l\}$, l the rank of the algebra, be ordered in some arbitrary manner. Then a weight $M = \Lambda - \sum_{i=1}^{l} k_i \alpha_i$ is called higher (lower) than a weight $M' = \Lambda - \sum_{i=1}^{l} k_i' \alpha_i$, if for the first nonvanishing difference $k_i - k_i' > 0$ (< 0) holds.

A vector y which corresponds to some weight M is called *extremal* (with respect to the operator E_α) if it holds that $E_\alpha y = 0$.

Throughout the proof α denotes a (not necessarily simple) root of the system Δ of positive roots.

Part 1: $2(\Lambda, \alpha)/(\alpha, \alpha) \neq n$, $n \geq 0$, integer

In this case it necessarily holds that $\alpha \notin \Delta_\Lambda$. This is seen as follows. The set of roots Δ_Λ is defined as $\Delta_\Lambda = \{\alpha \in \Delta \mid \alpha = r_i \alpha_i, \alpha_i \in \Pi, r_i$ nonnegative integers, $2(\Lambda, \alpha_i)/(\alpha_i, \alpha_i) = n_i, n_i \geq 0$, integer$\}$. Thus, if $\alpha \in \Delta_\Lambda$, then $2(\Lambda, \alpha)/(\alpha, \alpha) = n$, $n \geq 0$, integer. It should be noted that the converse is not true (parts 2 and 3 below).

(a) $2(\Lambda, \alpha)/(\alpha, \alpha) \neq n$, $n \geq 0$, integer.

Let x denote a vector of highest weight, $Hx = \Lambda(h)x$. A subspace R_Λ of R is constructed which is spanned by the vectors

$$y_0 = x, \; y_1 = F_\alpha x, \ldots, y_n = F_\alpha^n x, \cdots. \tag{15}$$

R_Λ is invariant under $G_{(\alpha)}$ (see proof of Theorem 2) as well as under $\tilde{G}_{(\alpha)}$ (Lemma 5). Since $2(\Lambda, \alpha)/(\alpha, \alpha)$ is not a nonnegative integer if follows from theorem 2 that the space R_Λ is irreducible and thus carries an *irreducible infinite-dimensional* representation of the algebra $\tilde{G}_{(\alpha)}$.

(b) $2(M, \alpha)/(\alpha, \alpha) \neq n$, $n \geq 0$, integer

Let $M = \Lambda - \sum k_i \alpha_i$, not all $k_i = 0$, denote the second weight according to the order introduced into D_Λ. We assume that $2(M, \alpha)/(\alpha, \alpha) \neq n$, $n \geq 0$, integer. Let V_M denote the (finite-dimensional) subspace of D_M consisting of the elements y such that $Hy = M(h)y$, with $\dim V_M = r$, r some integer > 0. In V_M there may exist elements y such that $E_\alpha y = 0$. For any element $y \in V_M$ which is annihilated by E_α we form the sequence

$$y, F_\alpha y, \ldots, F_\alpha^n y, \cdots. \tag{16}$$

It follows from (a) that we obtain an invariant subspace R_M which carries an *irreducible infinite-dimensional* representation of the algebra $\tilde{G}_{(\alpha)}$. The number of invariant and irreducible subspaces R_M obtained in this manner is equal to the number of linearly independent elements $y \in V_M$ for which $E_\alpha y = 0$.

(c) $2(M, \alpha)/(\alpha, \alpha) = n$, $n \geq 0$, integer

By proceeding consecutively to lower and lower weights M, according to the ordering introduced into D_Λ, the construction (a) holds as long as $2(M, \alpha)/(\alpha, \alpha)$ is not a nonnegative integer. It will in general, however, happen that $2(\sum k_i \alpha_i, \alpha)/(\alpha, \alpha)$, for $k_i > 0$, integer, becomes negative. In this case, if $2(\Lambda, \alpha)/(\alpha, \alpha)$ happens to be a *negative integer*, $2(\Lambda - \sum k_i \alpha_i, \alpha)/(\alpha, \alpha)$, not all $k_i = 0$, may become a positive integer or zero.

Let M again denote a weight of the form $M = \Lambda - \sum k_i \alpha_i$, not all $k_i = 0$, however with the property that $2(M, \alpha)/(\alpha, \alpha) = n$, $n \geq 0$, integer. If there exists an element $y \in V_M$ such that $E_\alpha y = 0$, then we form the sequence

$$y_0 = y, \quad y_1 = F_\alpha y, \quad y_2 = F_\alpha^2 y, \cdots, \quad y_n = F_\alpha^n y, \cdots. \tag{17}$$

The sequence, Eq. (17), does not terminate and we obtain a *reducible infinite-dimensional* representation of $\tilde{G}_{(\alpha)}$.

Proof: Assume that the sequence, Eq. (17), terminates, i.e., there exists a nonnegative integer m such that $F_\alpha^m y \neq 0$, $F_\alpha^{m+1} y = 0$. If $\alpha_1 \neq \alpha$ is a simple root we form $E_1 y$, $E_1 \equiv E_{\alpha_1}$. Two cases can happen:

Case (c1): $E_1 y = 0$. If $E_1 y = 0$, we have under the assumption made above,

$$E_1 y = 0, \; E_\alpha y = 0, \; F_\alpha^{m+1} y = 0, \tag{18}$$

with

$$2(M, \alpha)/(\alpha, \alpha) = n, \quad n \geq 0, \quad \text{integer}. \tag{19}$$

Case (c2): $E_1 y \neq 0$. If $E_1 y \neq 0$, a vector z is constructed from y such that z is extremal with respect to E_1 and Eqs. (18) and (19) hold for the vector z with weight M' for some value of m.

The vector z is obtained as follows. Let k denote a positive integer such that $E_\alpha^k E_1 y \neq 0$, $E_\alpha^{k+1} E_1 y = 0$. Such a k exists since *any* weight of D_Λ is of the form $\Lambda - \sum k_i' \alpha_i$, $\alpha_i \in \Pi$, k_i nonnegative integers. Acting with E_α, $\alpha = \sum r_i \alpha_i$, r_i nonnegative integers, onto the vector $E_1 y$ yields the weight $\Lambda - \sum k_i \alpha_i + \alpha_1 + t\alpha$, $t = 1, 2, 3, \cdots$, with $M + \alpha_1 = \Lambda - \sum k_i \alpha_i + \alpha_1$ the weight of the vector $E_1 y$. For a finite value $t = k$ a vector will be obtained which is no longer a weight of the representation D_Λ (in order to express this vector in the form $\Lambda - \sum k_i'' \alpha_i$, $\alpha_i \in \Pi$, some of the k_i must become negative integers).

At this point it is convenient to prove the following lemma before proceeding further with case (c2).

Lemma 6: Under the assumption that the sequence, Eq. (17), terminates, it holds $F^{m'}E^kE_1y \neq 0$, $F^{m'+1}E^kE_1y = 0$, m' some positive integer. If $[F_\alpha, E_1]=0$ it follows from Lemma 4 that Lemma 6 is true and that $n' = m + k$. If $[F_\alpha, E_1] = aF_{\alpha-\alpha_1}$, $a \neq 0$, one obtains with $[F_\alpha, E_\alpha] = H_\alpha$, and ignoring coefficients,

$$\begin{aligned} F_\alpha^{m'}E_1y &= F_\alpha^{m'-1}E_1F_\alpha y + F_\alpha^{m'-1}F_{\alpha-\alpha_1}y \\ &= F_\alpha^{m'-2}E_1F_\alpha^2 y + F_\alpha^{m'-2}F_{\alpha-\alpha_1}F_\alpha y + F_\alpha^{m'-1}F_{\alpha-\alpha_1}y \\ &= E_1F_\alpha^{m'}y + F_{\alpha-\alpha_1}F_\alpha^{m'-1}y + F_\alpha F_{\alpha-\alpha_1}F_\alpha^{m'-2} \\ &\quad + \cdots + F_\alpha^{m'-1}F_{\alpha-\alpha_1}y. \end{aligned} \tag{20}$$

Let $s-1$ be the largest integer for which $F_{(s-1)\alpha-\alpha_1} \neq 0$. It then follows from Lemma 3, Lemma 4, and the assumption $F_\alpha^m y \neq 0$, $F_\alpha^{m+1}y = 0$, that Lemma 6 holds with $n' = m + k + s$. This completes the proof of Lemma 6.

Proceeding further with case (c2), let $M' = \Lambda - \sum k_i'\alpha_i$, k_i' nonnegative integers, be the weight of the vector $E_\alpha^k E_1 y$, then it must hold, due to Lemma 6, that $2(M', \alpha)/(\alpha, \alpha) = n$, $n > 0$, integer, since according to Theorem 2 the sequence equation (17) does not terminate for any other value. The vector $E_\alpha^k E_1 y$ has therefore the property $E_\alpha(E_\alpha^k E_1 y) = 0$, $F_\alpha^{m'+1}(E_\alpha^k E_1 y) = 0$, with n' some positive integer, and $2(M', \alpha)/(\alpha, \alpha) = n > 0$, integer.

If it now holds $E_1(E_\alpha^k E_1 y) = 0$, then $E_\alpha^k E_1 y$ satisfies Eq. (18) with $y \to E_\alpha^k E_1 y$, $m \to n'$ and we have case (c1). Thus $z = E_\alpha^k E_1 y$. If $E_1(E_\alpha^k E_1 y) \neq 0$, then $y' = E_\alpha^k E_1 y$ is not yet the desired vector z and we have again case (c2) with $y \to y'$. The same construction as given in case (c2) is now applied to the vector y'. For the vector y'' obtained in this manner it either holds $E_1 y'' = 0$, case (c1), and $y'' = z$. Or it holds $E_1 y'' \neq 0$, case (c2), and the construction as given for case (c2) is repeated for the vector $y''' = E_1 y''$. After a finite number of steps a nonzero vector z is obtained, since every time case (c2) is repeated the positive root α_1 is added to the weight of M of the vector y.

With the vector z obtained in this manner the same construction as given above for $\alpha_1 \neq \alpha$ is repeated for the simple root $\alpha_2 \neq \alpha_1, \alpha$. A vector v with some weight M' is obtained satisfying Eqs. (18) and (19) with $\alpha_1 \to \alpha_2$.

In constructing v it may occur that while a vector extremal with respect to E_{α_2} is obtained, this vector is no longer extremal with respect to E_{α_1}. In this case this vector has again to be made extremal with respect to E_{α_1}, which may in turn result in a vector no longer extremal with respect to E_{α_2}. However, after a finite number of steps a vector z is obtained which is extremal with respect to E_α, E_{α_1}, and E_{α_2}.

The procedure described above is repeated for all simple roots $\alpha_i \neq \alpha$. A vector w with weight $\tilde{M}$ is obtained, satisfying

$$E_i w = 0, \quad 2(\tilde{M}, \alpha)/(\alpha, \alpha) = n \quad (n \geq 0, \text{ integer}) \tag{21}$$

for all simple roots α_i, $i = 1, 2, \cdots$. This implies that under the assumption that the sequence equation (17) terminates, a highest weight $\tilde{M}$ is obtained for the irreducible representation D_Λ, which satisfies Eq. (21). This would imply that the *irreducible* representation has two different highest weights. This is, however, impossible. Thus the assumption that the sequence equation (17) terminates leads to a contradiction.

Part II: $2(\Lambda, \alpha)/(\alpha, \alpha) = n$, $n \geq 0$, integer, $\alpha \notin \Delta_\Lambda$

As in Part I let R_M denote a subspace of R invariant with respect to the algebra $\tilde{G}_{(\alpha)}$. In the following, Theorem 3 will be proved for the case that $2(M, \alpha)/(\alpha, \alpha) = n$, $n \geq 0$, integer, with $\alpha \notin \Delta_\Lambda$.

If $\alpha \notin \Delta_\Lambda$, then in the decomposition $\alpha = \sum r_i\alpha_i$ (α_i simple roots) there is at least one simple root α_j such that $\alpha_j \notin \Delta_\Lambda$. Let $y \in R_M$ denote an extremal vector for E_α satisfying $Hy = M(h)y$, i.e., $E_\alpha y = 0$, with $2(M, \alpha)/(\alpha, \alpha) = n$, $n \geq 0$, integer. The sequence

$$y_0 = y, \; y_1' = F_j y, \ldots, y_m' = F_j^m y, \cdots \tag{22}$$

does not terminate since $2(\Lambda, \alpha_j)/(\alpha_j, \alpha_j) \neq n$, $n \geq 0$ integer, as follows from Part I. It will be shown that the sequence

$$y_0 = y, \; y_1 = F_\alpha y, \ldots, y_n = F_\alpha^n y, \cdots \tag{23}$$

does not terminate either.

In the following three cases will be distinguished, namely $(\alpha, \alpha_j) > 0$, $(\alpha, \alpha_j) < 0$, and $(\alpha, \alpha_j) = 0$.

(a) $(\alpha, \alpha_j) > 0$

Suppose the sequence, Eq. (23), terminates, i.e., there exists a positive integer n' such that

$$F_\alpha^{n'} y \neq 0, \quad F_\alpha^{n'+1} y = 0. \tag{24}$$

Then, for arbitrary m, the sequence

$$y_m' = F_j^m y, \; F_\alpha y_m', \; F_\alpha^2 y_m', \ldots, F_\alpha^n y_m' \cdots \tag{25}$$

terminates too. This follows from Lemma 3. However, the vector y_m' may not be a vector of highest weight for the algebra $\tilde{G}_{(\alpha)}$. For this reason the sequence is formed

$$y_m', E_\alpha y_m', E_\alpha^2 y_m', \ldots, E_\alpha^n y_m', \cdots \tag{26}$$

as a continuation of the sequence equation (25) to vectors of higher and higher weight (with respect to $\tilde{G}_{(\alpha)}$). The sequence (26) obviously terminates since α is a positive root. [See Part I, case (c2).] Moreover, the two sets together, Eq. (25) and Eq. (26) are invariant under $\tilde{G}_{(\alpha)}$ and form a basis for a finite-dimensional (and irreducible) representation of $\tilde{G}_{(\alpha)}$.

The weight of $E_\alpha^n y_m'$ is $M - m\alpha_j + n\alpha$, where M is the weight of y_0. Now it holds $\alpha \neq \alpha_j$. Thus $\alpha = \sum r_i\alpha_i$, $\alpha_i \in \Pi$, contains at least one other simple root $\alpha_k \neq \alpha_j$. E_α acting on a vector with weight $M - m\alpha_j + n\alpha$ yields a vector with a weight $M - m\alpha_j + (n+1)\alpha$. Since there exists at least one other simple root $\alpha_k \neq \alpha_j$, it follows that there exists a positive integer n_0, such that $M - m\alpha_j + n_0\alpha$ is no longer a weight of D_Λ for all m. Moreover, n_0 is independent of the value m. Therefore, under the condition that $n < n_0$ it holds for some value m,

$$2(M - m\alpha_j + n\alpha, \alpha)/(\alpha, \alpha) < 0, \tag{27}$$

since $(\alpha_j, \alpha) > 0$. According to Theorem 2 this contradicts the fact that $M - m\alpha_j + n\alpha$ is the highest weight of a finite-dimensional space which is invariant under

$\tilde{G}_{(\alpha)}$. Thus the sequence equation (23) does not terminate if $(\alpha, \alpha_j) > 0$. As a consequence we obtain the following lemma:

Lemma 7: If β is a simple root of G for which the sequence

$$y_0, y'_1 = F_\beta y_0, \ldots, y'_n = F_\beta^n y_0, \cdots$$

does not terminate, and if α is another positive root, $\alpha \notin \Delta_\Lambda$, such that $(\alpha, \beta) > 0$, then the sequence

$$y_0, y_1 = F_\alpha y_0, \ldots, y_n = F_\alpha^n y_0, \cdots$$

does not terminate either.

(b) $(\alpha, \alpha_j) \leq 0$

It is sufficient to prove that the sequence equation (23) does not terminate for $M = \Lambda$, Λ the highest weight of D_Λ. Indeed, if x_0 is a vector which corresponds to the weight Λ of D_Λ then it holds $x_0 \in \Omega y_0$, i.e., x_0 is a finite linear combination of vectors $A_1 A_2 \cdots A_k y_0$ with the A_i either E_γ or F_γ, $\gamma \in \Delta$. This follows from the irreducibility of D_Λ. If the sequence equation (23) terminates, then it follows from Lemmas 3 and 4 that there exists an integer n, $n \geq 0$, such that $F_\alpha^n (A_1 A_2 \cdots A_k y_0) = 0$. Hence, there exists an integer n', $n' \geq 0$, such that $F_\alpha^{n'} x_0 = 0$. Thus, if the sequence equation (23) does not terminate for $M = \Lambda$, then it does also not terminate for $M \neq \Lambda$.

Since α is not a simple root [if it were then the condition $2(\Lambda, \alpha)/(\alpha, \alpha) = n$, $n \geq 0$, integer, implies that $\alpha \in \Delta_\Lambda$], the root α can be represented as a sum of 2 positive roots β_1, β_2, one of which is a simple root. Moreover, at least one of these two roots does not belong to Δ_Λ. Consider now the set of roots of G which is generated by the application of the Weyl reflections S_α. S_{β_1}, S_{β_2} and their products to the roots α, β_1, and β_2. The set of roots obtained is a subsystem $\Gamma^{(2)}$ of the root system Γ. Moreover, $\Gamma^{(2)}$ is a system of roots of a simple Lie algebra of rank 2, denoted by $G^{(2)}$. Let $\Delta^{(2)} = \Gamma^{(2)} \cap \Delta$. Then $\Delta^{(2)}$ is a system of positive roots for the algebra $G^{(2)}$. Let $\Omega^{(2)}$ denote a universal enveloping algebra of $G^{(2)}$. The universal enveloping algebra $\Omega^{(2)}$ can be embedded in the universal enveloping algebra Ω in a canonical manner. In the following we shall assume that $\Omega^{(2)}$ is embedded canonically in Ω, i.e., $\Omega^{(2)} \subset \Omega$.

Consider the subspace $R^{(2)} \equiv \Omega^{(2)} y_0$ of R. It is clear that the reduction of the representation D_Λ with respect to the subalgebra $G^{(2)}$ contains a representation of the algebra $G^{(2)}$ on the space $R^{(2)}$ which has a highest weight Λ. This representation will be denoted by $D_\Lambda^{(2)}$. It is also clear that the vectors of Eq. (23) belong to the space $R^{(2)}$. Thus, the sequence equation (23) terminates for the representation D_Λ of G if and only if this sequence terminates for the representation $D_\Lambda^{(2)}$ of $G^{(2)}$. Similarly, if the sequences $y_0, F_{\beta_i} y_0, \ldots, F_{\beta_i}^n y_0, \cdots$, $i = 1, 2$, do not terminate for the representation D_Λ they do not terminate for the representation $D_\Lambda^{(2)}$.

Thus, in order to prove that the sequence equation (23) does not terminate it is sufficient to prove it for the simple Lie algebras $G^{(2)}$ of rank 2, i.e., for the case of A_2, B_2 and G_2, respectively. Let us first consider the case that the algebra $G^{(2)}$ is the simple algebra B_2. Let α_1 and α_2 denote two simple roots of B_2 such that $|\alpha_1| > |\alpha_2|$, i.e., α_1 is the longer of the two simple roots. The other positive roots of B_2 are then given as $\alpha_1 + \alpha_2$, $\alpha_1 + 2\alpha_2$. There exist only two possibilities for a realization of the roots α, β_1, β_2 in B_2, namely (b1) $\alpha = \alpha_1 + 2\alpha_2$, $\beta_1 = \alpha_2$, $\beta_1 = \alpha_2$, $\beta_2 = \alpha_1 + \alpha_2$ and (b2) $\alpha = \alpha_1 + \alpha_2$, $\beta_1 = \alpha_1$, $\beta_2 = \alpha_2$.

Case (*b*1): $\alpha = \alpha_1 + 2\alpha_2$, $\beta_1 = \alpha_2$, $\beta_2 = \alpha_1 + \alpha_2$. It is clear that the root α_2 is a simple root of the system Δ of G, while β_2 is not a simple root. If the sequence $y_0, F_{\beta_1} y_0, \ldots, F_{\beta_1}^n y_0, \ldots$ does not terminate, then the sequence equation (23) does not terminate either, according to Lemma 7, since $(\alpha_2, \alpha_1 + 2\alpha_2) > 0$.

Now let us assume that the sequence y_0, $F_{\beta_1} y_0, \ldots, F_{\beta_1}^n y_0, \cdots$ does terminate (hence $\beta_1 \in \Delta_\Lambda$, and for some integer $m \geq 0$ it holds $F_{\beta_1}^m y_0 \neq 0$, $F_{\beta_1}^{m+1} y_0 = 0$). Then the sequence y_0, $F_{\beta_2} y_0, \ldots F_{\beta_2}^n y_0, \cdots$ does not terminate, as a consequence of Lemma 7 [$(\alpha_1, \beta_2) > 0$], since the sequence y_0, $F_{\alpha_1} y_0, \ldots, F_{\alpha_1}^n y_0, \cdots$ can not terminate. If this last sequence would terminate, then it would hold $\alpha_1, \alpha_2 \in \Delta_\Lambda$ and thus also $\alpha \in \Delta_\Lambda$, contrary to our assumption.

Consider the sequence $F_{\beta_1}^m y_0 \equiv y'_m$, $F_\alpha y'_m, \ldots$, $F_\alpha^n y'_m, \cdots$. According to Lemma 3 this sequence terminates if the sequence equation (23) terminates. It will be demonstrated that this sequence does not terminate. Consider the system of roots $S_{\alpha_2} \Delta = \Delta'$. The system Δ' is obtained from the system Δ by replacing α_2 by $-\alpha_2$ (Ref. 8, Sec. 8, § 1, Lemma 1). The system Δ' can again be considered as a system of positive roots of G. According to Lemma 1 (applying it to the root $\alpha_2 = \beta_1$) the representation D_Λ is a representation with highest weight vector y'_m with respect to the system of positive roots Δ' (it holds $E'_\beta y'_m = 0$ for all $\beta \in \Delta'$, $E'_\beta = E_\beta$ for $\beta \neq \alpha_2$, and $E'_{\alpha_2} = F_{\alpha_2}$). All considerations following Lemma 7 can now be repeated by substituting $\Delta \to \Delta'$, (highest weight y_0 for Δ) $\to$ (highest weight y'_m for Δ'), (roots β) $\to$ (roots $S_{\alpha_2} \beta$), (sequence $y_0, F_\beta y_0, \ldots$, $F_\beta^n y_0, \cdots$) $\to$ (sequence $y'_m, F'_{\beta'} y'_m, \ldots, F'^n_{\beta'} y'_m, \cdots$, where $\beta' = S_{\alpha_2} \beta$, $F'_{\beta'} = F_{\beta'}$ if $\beta' > 0$ and $F'_{\beta'} = E_{\beta'}$ if $\beta' < 0$). Obviously, the subalgebra $G^{(2)}$ is the same for both cases. According to Theorem 2 it holds that $E_{\alpha_2}^m y'_m = F'^m_{-\alpha_2} y'_m = c y_0$, $c \neq 0$. Hence $\Omega^{(2)} y'_m = \Omega^{(2)} y_0 = R^{(2)}$. Thus the representations $D_\Lambda^{(2)}$ for the two cases coincide. Thus, from the fact that for the case of the system of positive roots Δ the sequence $y_0, F_{\alpha_1} y_0, \ldots, F_{\alpha_1}^n y_0, \cdots$ does not terminate, it follows that the sequence $y'_m, F_\alpha y'_m, \ldots$, $F_\alpha^n y'_m, \cdots$ does not terminate, since $S_{\alpha_2} \alpha_1 = \alpha$. Above it was shown that if the last sequence does not terminate then the sequence equation (23) does not terminate either.

Case (*b*2): $\alpha = \alpha_1 + \alpha_2$, $\beta_1 = \alpha_1$, $\beta_2 = \alpha_2$. The proof for this case is quite similar to the proof of case (b1) and we delete it. Again it follows that the sequence equation (23) does not terminate.

So far the proof has been given for the special case that $G^{(2)} = B_2$. However, for the other algebras of rank 2(namely A_2 and G_2) the proof follows analogous lines and we do not give it.

Consequently, it has been proved that the sequence equation (23) does not terminate for $2(\Lambda, \alpha)/(\alpha, \alpha)$ an integer ≥ 0 and $\alpha \notin \Delta_\Lambda$. Equation (9) shows how the operators H, F_α, E_α act upon the vectors of the sequence equation (23). The proof of Eq. (9) follows the proof given in Theorem 2 and will not be repeated here.

The sequence equations (16) and (23) consist of linearly independent vectors. The operators H, F_α, E_α transform the vectors of these sequences into vectors of the same sequence. Thus, in the cases considered so far the space R is decomposed into a direct sum of infinite-dimensional subspaces which are invariant under $G_{(\alpha)}$. Each subspace forms the space for a representation with a highest weight of the algebra $G_{(\alpha)}$.

Part III: $2(\Lambda, \alpha)/(\alpha, \alpha) = n$, $n > 0$, integer, $\alpha \in \Delta_\Lambda$

According to Lemma 2 all sequences of type equation (23) are finite *under the conditions given above*. From Theorem 2 follows that every sequence of this type forms a basis for a finite-dimensional irreducible representation of $\tilde{G}_{(\alpha)}$.

The operators H_α, E_α, F_α act on the elements of the basis according to the formulas given by Eq. (10). Thus, for this case the space R decomposes into a direct sum of finite-dimensional subspaces, each of which is invariant under the algebra $G_{(\alpha)}$. This completes the proof of Theorem 3.

Corollary 1: Let D_Λ be an irreducible representation of G with a highest weight Λ. If α is a positive root and $\alpha \in \Delta_\Lambda$, M a weight of D_Λ, then the multiplicity of the weights M and $M - 2(M, \alpha)/(\alpha, \alpha)$ is the same.

Proof: According to Theorem 3, the space R of D_Λ decomposes into a direct sum of finite-dimensional subspaces which are invariant under $G_{(\alpha)}$. The multiplicity of a weight M is equal to the number of invariant subspaces which contain a vector of weight M. From Lemma 2 follows that the finite-dimensional irreducible representations of $G_{(\alpha)}$ contain with a weight M also the weight $M - 2(M, \alpha)/(\alpha, \alpha)$. This proves Corollary 1.

Corollary 2: Under the conditions of Lemma 2 the multiplicity of the weights M and SM, $S \in W_\Lambda$, is equal in a representation D_Λ. The proof follows from Corollary 1.

Theorem 3 shows that the symmetry of Corollary 2 is the maximal symmetry of the weight diagram of the irreducible representation D_Λ.

V. FREUDENTHAL'S FORMULA FOR INFINITE-DIMENSIONAL REPRESENTATIONS WITH A HIGHEST WEIGHT

The derivation of Freudenthal's formula for the case of *finite*-dimensional representations of the simple Lie algebras G is well known. Below it is demonstrated that Freudenthal's formula also holds for *infinite* dimensional irreducible representations with a highest weight of the simple Lie algebras G. The proof rests on Theorem 3 of Sec. IV. The notation employed is that of Sec. IV.

Let α denote a positive root of a simple Lie algebra G. Let M' denote a weight of an irreducible representation D_Λ of G with highest weight Λ. Then the set of elements

$$\cdots, M' + 2\alpha,\ M' + \alpha,\ M', M' - \alpha, M' - 2\alpha, \cdots$$

is called an α-sequence containing the weight M'. Such a sequence forms a subset of weights of D_Λ (to elements of the sequence which do not correspond to weights the multiplicity zero is assigned).

Let $\alpha \notin \Delta_\Lambda$. Then, if M is a weight of D_Λ such that $M + \alpha \notin D_\Lambda$, the α-sequence containing M is given as

$$M, M - \alpha, M - 2\alpha, \cdots . \tag{28}$$

Let R denote a linear space which carries the irreducible representation D_Λ and let the subspace $\mathcal{M}_{M-p\alpha}$ be defined in the same manner as the subspace $V_{M'-p\alpha}$ of the space V of Theorem 1, Sec. II. From among the representations of the subalgebra $G_{(\alpha)}$ of G, as constructed in Theorem 3, those with the highest weights $M, M - \alpha, \ldots, M - p\alpha$ contribute a subspace to $\mathcal{M}_{M-p\alpha}$, such that the direct sum of these subspaces forms the space $\mathcal{M}_{M-p\alpha}$. If m_j denotes the number of representations of $G_{(\alpha)}$ with highest weight $M - j\alpha$, then

$$m_j = n_{M-j\alpha} - n_{M-(j-1)\alpha} . \tag{29}$$

Here and in the following n_M denotes the multiplicity of the weight M of D_Λ. Let y_{p-j} denote a vector of weight $M - p\alpha$ of the space which carries a representation of $G_{(\alpha)}$ with highest weight $M - j\alpha$, $j \leq p$, as given by equation (9). Since

$$m' = 2(M - j\alpha, \alpha)/(\alpha, \alpha) = m - 2j,$$

it holds that

$$E_\alpha F_\alpha y_{p-j} = \tfrac{1}{2}(p - j + 1)(p - m + j)(\alpha, \alpha) y_{p-j}$$

and

$$F_\alpha E_\alpha y_{p-j} = \tfrac{1}{2}(p - j)(p - m + j - 1)(\alpha, \alpha) y_{p-j}. \tag{30}$$

In the following the traces of the operators $E_\alpha F_\alpha$ and $F_\alpha E_\alpha$, restricted to the subspace $\mathcal{M}_{M-p\alpha}$, are evaluated. From Eqs. (29) and (30) it follows that

$$\begin{aligned}
\mathrm{Tr}_{\mathcal{M}_{M-p\alpha}} F_\alpha E_\alpha &= \sum_{j=0}^{p} m_j \tfrac{1}{2}(p - j)(p - m + j - 1)(\alpha, \alpha) \\
&= \sum_{j=0}^{p} (n_{M-j\alpha} - n_{M-(j-1)\alpha}) \tfrac{1}{2}(p - j)(p - m + j - 1) \\
&\quad \times (\alpha, \alpha) \\
&= \sum_{j=0}^{p-1} n_{M-j\alpha} \tfrac{1}{2}(2j - m)(\alpha, \alpha) \\
&= -\sum_{j=1}^{p-1} n_{M-j\alpha}(M - j\alpha, \alpha)
\end{aligned}$$

and

$$\mathrm{Tr}_{\mathcal{M}_{M-p\alpha}} E_\alpha F_\alpha = -\sum_{j=0}^{p} n_{M-j\alpha}(M - j\alpha, \alpha).$$

Setting $n_{M'} = 0$ if M' is not a weight of D_Λ and inverting the sums, one obtains

$$\begin{aligned}
\mathrm{Tr}_{\mathcal{M}_M} F_\alpha E_\alpha &= -\sum_{j=1}^{\infty} n_{M+j\alpha}(M + j\alpha, \alpha), \\
\mathrm{Tr}_{\mathcal{M}_M} E_\alpha F_\alpha &= -\sum_{j=0}^{\infty} n_{M+j\alpha}(M + j\alpha, \alpha),
\end{aligned} \tag{31}$$

where M now denotes an *arbitrary* weight of the se-

quence equation (28). If $\alpha \in \Delta_\Lambda$, then from considerations analogous to those given in Ref. (8), Sec. 8, § 2, it follows that

$$\mathrm{Tr}_{\mathcal{M}_M} F_\alpha E_\alpha = -\sum_{j=0}^{\infty} n_{M-j\alpha}(M-j\alpha, \alpha), \tag{32}$$

$$\mathrm{Tr}_{\mathcal{M}_M} E_\alpha F_\alpha = -\sum_{j=0} n_{M+j\alpha}(M+j\alpha, \alpha).$$

The Casimir operator for a simple Lie algebra G is given as

$$\Gamma = \sum_{i=1}^{l} H_i H^i - \sum_{\alpha\in\Delta} F_\alpha E_\alpha - \sum_{\alpha\in\Delta} E_\alpha F_\alpha$$

$$= \sum_{i=1}^{l} H_i H^i - \sum_{\alpha\in\Delta} H_\alpha - 2\sum_{\alpha\in\Delta} F_\alpha E_\alpha,$$

where the H_i and H^i are defined as in Ref. 8, Sec. 8, § 2. On the space R the Casimir operator Γ is a multiple γ of the identity operator with $\gamma = (\Lambda + R, \Lambda + R) - (R, R)$. This can be seen as follows. The space R can symbolically be written as $R \sim \Omega x \sim \Omega - x$, where Ω characterizes the enveloping algebra of G (see Sec. II) and where x is the vector corresponding to the highest weight Λ. This means that every element of R can be expressed as a linear combination of elements of the form

$$y = F_{i_1} F_{i_2} \cdots F_{i_n} x. \tag{33}$$

Since $E_\alpha x = 0$, it holds that

$$\Gamma x = \sum_{i=1}^{l} H_i H^i x + \sum_{\alpha\in\Delta} H_\alpha x.$$

Using the fact that[8]

$$\sum_{i=1}^{l} H_i H^i x = (\Lambda, \Lambda)x$$

and that

$$\sum_{\alpha\in\Delta} H_\alpha x = \sum_{\alpha\in\Delta} \Lambda(h_\alpha)x = \left(\Lambda, \sum_{\alpha\in\Delta} \alpha\right)x,$$

it follows

$$\Gamma x = (\Lambda, \Lambda)x + \left(\Lambda, \sum_{\alpha\in\Delta} \alpha\right)x$$

$$= (\Lambda, \Lambda)x + (\Lambda, 2R)x$$

$$= ((\Lambda + R, \Lambda + R) - (R, R))x$$

with $2R = \sum_{\alpha\in\Delta} \alpha$. Since it holds $[\Gamma, F_j] = 0$, it follows that this relation is true not merely for the vector x but for any vector $y \in R$. Thus

$$\Gamma = \gamma \mathbb{1} = ((\Lambda + R, \Lambda + R) - (R, R))\mathbb{1}. \tag{34}$$

The trace of Γ, restricted to the subspace $\mathcal{M}_M$ is γn_M and thus, from the definition of the Casimir operator, it follows

$$\gamma n_M = \sum_{i=1}^{l} \mathrm{tr}_{\mathcal{M}_M} H_i H^i - \sum_{\alpha\in\Delta} \mathrm{tr}_{\mathcal{M}_M}(F_\alpha E_\alpha - E_\alpha F_\alpha). \tag{35}$$

With the help of the relation[8]

$$\sum_{i=1}^{l} \mathrm{Tr}_{\mathcal{M}_M} H_i H^i = (M, M) n_M$$

and with Eqs. (31), (32), it follows from Eq. (35) that

$$\gamma n_M = (M, M) n_M$$

$$+ \sum_{\substack{\alpha\in\Delta \\ \alpha\notin\Delta_\Lambda}} \left(\sum_{j=0}^{\infty} n_{M+j\alpha}(M+j\alpha, \alpha) + \sum_{j=1}^{\infty} n_{M+j\alpha}(M+j\alpha, \alpha) \right)$$

$$+ \sum_{\alpha\in\Delta_\Lambda} \left(\sum_{j=0}^{\infty} n_{M-j\alpha}(M-j\alpha, \alpha) \right.$$

$$\left. + \sum_{j=0}^{\infty} n_{M+j\alpha}(M+j\alpha, \alpha) \right). \tag{36}$$

Using the relation[8]

$$\sum_{j=1}^{\infty} n_{M-j\alpha}(M-j\alpha, \alpha) = n_M(M, \alpha) + \sum_{j=1}^{\infty} n_{M+j\alpha}(M+j\alpha, \alpha),$$

Eq. (36) goes over into

$$\gamma n_M = (M, M) n_M + \sum_{\alpha\in\Delta} n_M(M, \alpha) + 2\sum_{\alpha\in\Delta} \sum_{j=1}^{\infty} n_{M+j\alpha}(M+j\alpha, \alpha),$$

and by substituting into this equation the identity

$$(M, M) n_M + n_M \sum_{\alpha\in\Delta} (M, \alpha) = ((M + R, M + R) - (R, R)) n_M ,$$

Freudenthal's formula is obtained as

$$((\Lambda + R, \Lambda + R) - (M + R, M + R)) n_M = 2\sum_{\alpha\in\Delta} \sum_{j=1}^{\infty} n_{M+j\alpha}(M+j\alpha, \alpha).$$

Thus, Freudenthal's formula for the case of infinite-dimensional irreducible representations having a highest weight is identical in form to Freudenthal's formula for finite-dimensional irreducible representations. However, in distinction to the finite-dimensional case there may exist weights M with the property $(M + R, M + R) = (\Lambda + R, \Lambda + R)$. In this case the coefficient of n_M is zero and Freudenthal's formula does not yield the multiplicity of such weights.

VI. MULTIPLICITIES OF WEIGHTS AND CHARACTERS FOR ELEMENTARY REPRESENTATIONS AND FOR INFINITE-DIMENSIONAL IRREDUCIBLE REPRESENTATIONS WITH A HEIGHEST WEIGHT

Let $P(M)$ denote Kostant's partition function. For given M its value equals to the number of partitions of the linear form M into a sum of positive roots of the algebra G, namely $M = \sum_{\alpha\in\Delta} k_\alpha \alpha$, with nonnegative integer coefficients k_α. Thus, for $M = 0$, it holds that $P(M) = 1$, and for elements M for which one or more of the integers k_α are negative it holds that $P(M) = 0$.

In the following some theorems and lemmas are given. These theorems and lemmas are needed for the derivation of both, characters as well as recurrence relations for the multiplicities of weights of infinite-dimensional representations with highest weight of the semisimple Lie algebras G. Some of these theorems and lemmas have been obtained previously. In this case the theorems and lemmas are simply quoted without proof, except when it was felt that an alternate or simpler proof would be desirable.

The following theorem is due to Verma.[6]

Theorem 4: The multiplicity of the weight Λ' of an elementary representation d_Λ is equal to $P(\Lambda - \Lambda')$.

Proof: The elementary representation d_Λ is defined on the space of the enveloping algebra Ω_- (see Sec. II). According to the Poincaré—Birkhoff—Witt theorem[8] a basis for Ω_- can be defined in the following manner. The set of elements f_α, with $\alpha \in \Delta$, forms a basis of the subalgebra G^- of the algebra G. An ordering is introduced into the set of positive roots Δ. For any pair α, α' of positive roots it then holds with respect to this ordering that either $\alpha < \alpha'$ or $\alpha > \alpha'$. The number 1 and all elements of the form $f_{\alpha'} f_{\alpha''} \cdots f_{\alpha^{(n)}}$, such that $\alpha' \leq \alpha'' \leq \cdots \leq \alpha^{(n)}$, make up a basis for the enveloping algebra Ω. According to Eq. (1) the operators of the representation d_Λ act upon this basis in the following manner [it should be noted that the two basis for Ω_-, the one given now and the one of Eq. (1), Sec. II, are different],

$$\rho(h)1 = \Lambda(h)1, \tag{38a}$$

$$\rho(f_\alpha)1 = f_\alpha, \tag{38b}$$

$$\rho(e_\alpha)1 = 0, \tag{38c}$$

$$\rho(h) f_{\alpha'} f_{\alpha''} \cdots f_{\alpha^{(n)}} = (\Lambda - \alpha' - \alpha'' - \cdots - \alpha^{(n)})(h) \times f_{\alpha'} f_{\alpha''} \cdots f_{\alpha^{(n)}} \tag{38d}$$

$$\rho(f_i) f_{\alpha'} f_{\alpha''} \cdots f_{\alpha^{(n)}} = f_i f_{\alpha'} f_{\alpha''} \cdots f_{\alpha^{(n)}}, \tag{38e}$$

$$\rho(e_i) f_{\alpha'} f_{\alpha''} \cdots f_{\alpha^{(n)}} = f_{\alpha'} (\rho(e_i) f_{\alpha''} \cdots f_{\alpha^{(n)}}) + [e_i, f_{\alpha'}] f_{\alpha''} \cdots f_{\alpha^{(n)}}. \tag{38f}$$

Obviously, the right-hand side of Eqs. (38e) and (38f) can be reexpressed as linear combinations over the basis elements introduced above for the enveloping algebra Ω_-.

From Eq. (38d) the multiplicity of a weight M of the representation d_Λ can be read off easily. It is equal to the number of basis elements $f_{\alpha'} f_{\alpha''} \cdots f_{\alpha}(n)$ of the space Ω_- for which holds $\alpha' + \alpha'' + \cdots + \alpha^{(n)} = \Lambda - \Lambda'$. This is, however, precisely the value of $P(\Lambda - \Lambda')$. This proves the theorem.

The following lemma is also due to Verma.[6] Let Z denote the center of the enveloping algebra Ω. Each representation ρ_Λ of an algebra G with highest weight Λ generates also a representation of its enveloping algebra Ω. The representation of Ω obtained in this manner is also denoted by the symbol ρ_Λ. In the representation ρ_Λ of the enveloping algebra Ω the elements of Z are represented by operators which are multiples of the identity operator. This is seen as follows. Let x denote the vector of highest weight. Then, since $[z, h] = 0$ for $z \in Z$ and $h \in G^0$, it holds

$$\rho_\Lambda(h)\rho_\Lambda(z)x = \rho_\Lambda(z)\rho_\Lambda(h)x = \Lambda(h)\rho_\Lambda(z)x.$$

Since the subspace V_Λ is one-dimensional (Theorem 1, Sec. II) it holds $\rho_\Lambda(z)x = \gamma_z x$, where $\gamma_z = \gamma_z(\Lambda)$ is a complex number. As for the case of the Casimir operator Γ (Sec. V) it can be demonstrated that $\rho_\Lambda(z)y = \gamma_z y$ for every element y of the representation space. Thus $\rho_\Lambda(z)1 = \gamma_z(\Lambda)1$.

The map $z \to \gamma_z(\Lambda)$ is a homomorphism of Z into C, the field of complex numbers. According to Theorem 5 of Ref. 12 it holds $\gamma_z(\Lambda) = \chi_\lambda(z)$, where λ is some linear form on G^0 and $\chi_\Lambda(z)$ is the character on Ω (see Ref. 12).

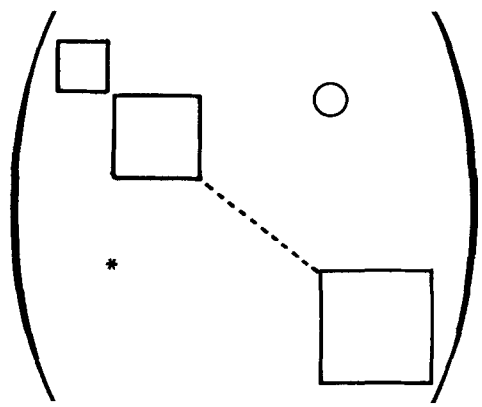
Lemma 4: It holds $\gamma_z(\Lambda) = \chi_\Lambda(z)$, i.e., $\lambda = \Lambda$.

Proof: From Lemma 36 in Ref. 12 it follows that $\chi_\Lambda(z)$ is an entire analytic function of the $\lambda_i = 2(\Lambda, \alpha_i)/(\alpha_i, \alpha_i)$, $i = 1, 2, \ldots, l$. Equation (1) shows that for every $z \in Z$ the expression $\gamma_z(\Lambda)$ is an entire analytic function of the λ_i, $i = 1, 2, \ldots, l$. For finite-dimensional representations D_Λ the proof of $\gamma_z(\Lambda) = \chi_\Lambda(z)$ has been given in Ref. 12, § III. Since the lemma holds for highest weights of finite-dimensional representations, it must hold identically due to the fact that $\gamma_z(\Lambda)$ and $\chi_\Lambda(z)$ are entire functions.

Lemma 5: A representation ρ_Λ with highest weight Λ contains only irreducible subrepresentations which have a highest weight. Moreover, the highest weights of the irreducible subrepresentations are of the form $S(\Lambda + R) - R$, $S \in W$.

The first part of this lemma is a consequence of Theorem 1 of Verma in Ref. 6. However, Bernstein, Gel'fand, and Gel'fand have given a counterexample to this theorem in Ref. 7. In the following it will be proved that while Verma's Theorem 1 is not correct, Lemma 5 nevertheless holds.

Proof: A reducible representation ρ_Λ with a highest weight Λ can be decomposed into a semidirect sum of irreducible representations. The operators are then of the form



where every box corresponds to an irreducible representation (finite- or infinite-dimensional). The star indicates that there are nonzero matrix elements below the block-diagonal. Above the block diagonal all matrix elements are zero.

Let $\tilde{V}$ denote an invariant irreducible subspace of the space V which carries the representation ρ_Λ. The subspace $\tilde{V}$ may be finite-dimensional or infinite-dimensional. It will be demonstrated that the subspace $\tilde{V}$ carries an irreducible representation *with a highest weight.*

Let $y = \beta_1 x_1 + \ldots + \beta_n x_n$ be an arbitrary element of the subspace $\tilde{V}$, with $x_1, \ldots, x_n$ weight vectors corresponding to *different* weights. Consider a sequence of elements,

$$y, \rho_\Lambda(e_{\alpha_{i_1}})y, \ \rho_\Lambda(e_{\alpha_{i_2}})\rho_\Lambda(e_{\alpha_{i_1}})y, \ldots, \rho_\Lambda(e_{\alpha_{i_n}}) \cdots \rho_\Lambda(e_{\alpha_{i_1}})y, \tag{39a}$$

such that none of these elements is equal to zero, while the last element satisfies

$$\rho_\Lambda(e_{\alpha_j})(\rho_\Lambda(e_{\alpha_{i_n}})\cdots\rho_\Lambda(e_{\alpha_{i_1}}))y=0,\quad j=1,2,\ldots,l. \tag{39b}$$

Such a sequence can be constructed for any element y of the space $\tilde{V}$, since the action of the operators $\rho_\Lambda(e_{\alpha_i})$ increases a weight by the vector α_i. However, no weight can become larger than Λ, the highest weight of the representation ρ_Λ. Since $\tilde{V}$ is an irreducible subspace, every vector of the sequence equation (39a) belongs to the space $\tilde{V}$. The last vector of the sequence equation (39a) is a linear combination $\gamma_1 z_1+\cdots+\gamma_m z_m$ of weight vectors $z_1,\ldots,z_m$ which belong to different weights. According to Theorem 1 weight vectors which belong to different weights are linearly independent. Now, for arbitrary i, $i=1,2,\ldots,l$, the vectors $\rho_\Lambda(e_{\alpha_i})z_1,\ldots,\rho_\Lambda(e_{\alpha_i})z_m$ correspond to different weights. Therefore, it follows from Eq. (39b) that for every j, $j=1,2,\ldots,m$, it holds

$$\rho(e_{\alpha_i})z_j=0,\quad i=1,2,\ldots,l. \tag{40}$$

Thus, in the linear combination $\gamma_1 z_1+\cdots+\gamma_m z_m$ each term vanishes separately, *independent* of the coefficients γ_j. Since the vectors $z_1, z_2, \ldots, z_m$ belong to different weights it follows that the z_i, $i=1,2,\ldots,m$, *themselves* are elements of the space $\tilde{V}$.

For each of the elements z_j the space $V_{z_j}=\rho_\Lambda(\Omega)z_j$ is constructed. Since $z_j\in\tilde{V}$, it holds $V_{z_j}\subset\tilde{V}$. From Eq. (40) it follows that V_{z_j} is an invariant subspace of the space V with a vector of highest weight, namely z_j. Now, $V_{z_j}\neq V_{z_i}$, $i\neq j$, since the two vectors z_i and z_j correspond to *different* weights. On the other hand, it has been assumed that the space $\tilde{V}$ is an *irreducible* subspace. This leads to a contradiction unless there is only one irreducible space V_{z_k} and it holds that $\tilde{V}=V_{z_k}$, with z_k the highest weight of the subspace $\tilde{V}$.

Now, the same considerations as given above are repeated for the factor-space $V'=V/\tilde{V}$. This space is invariant under ρ_Λ. Continuing in this manner the first part of the lemma is proved by taking into account that the representation ρ_Λ can contain only a finite number of irreducible representations with a highest weight. This follows from the fact that every weight subspace V_M of the representation space V is finite-dimensional (Theorem 1) and from the second part of the lemma. The second part of the lemma is a consequence of Theorem 5 of Ref. 12 concerning the character $\chi_\Lambda(z)$, $z\in Z$, and of Lemma 4.

Theorem 5: Let d_Λ be an elementary representation with highest weight Λ. If for every element S of the Weyl group W the element $S(\Lambda+R)-R$ is *not* a weight of d_Λ, then d_Λ is irreducible.

The proof is a consequence of Lemma 5.

A representation d_Λ for which elements $S(\Lambda+R)-R$, $S\neq 1$, are weights of d_Λ may contain invariant subspaces. Before a theorem can be formulated concerning this type of representations the following definition has to be made. Let Λ and Λ' denote linear forms on G^0 and $\beta_1, \beta_2, \ldots, \beta_n$ a sequence of positive roots. As before it holds

$$S_{\beta_i}\Lambda=\Lambda-\frac{2(\Lambda,\beta_i)}{(\beta_i,\beta_i)}\beta_i.$$

The sequence $\beta_1,\beta_2,\ldots,\beta_n$ is said to satisfy condition (a) for a pair of weights $(\Lambda';\Lambda)$, if

$$\text{(1)}\ \ \Lambda'=S_{\beta_n}S_{\beta_{n-1}}\cdots S_{\beta_1}\Lambda; \tag{41}$$

(2) for every j, $j=1,2,\ldots,n$, it holds

$$2(S_{\beta_{j-1}},S_{\beta_{j-2}}\cdots S_{\beta_1}\Lambda,\beta_j)/(\beta_j,\beta_j)=m,$$

with $m\geq 0$, integer $(S_{\beta_0}\equiv 1)$.

Theorem 6: An elementary representation d_Λ contains the elementary representation $d_{\Lambda'}$ as a subrepresentation if and only if a sequence of positive roots $\beta_1,\beta_2,\ldots,\beta_n$ exists such that condition (A) is satisfied for the pair $(\Lambda'+R;\Lambda+R)$.

The proof of this theorem is given in ref. (7).

Theorem 7: If an elementary representation d_Λ contains a subrepresentation $d_{\Lambda'}$ then it contains $d_{\Lambda'}$ only once.

This theorem is quoted in Ref. 6 and proved in Verma's Yale University dissertation (1966).

In Ref. 6 a conjecture was made, which reads as follows (conjecture 1 of Ref. 6):

Conjecture: Condition (A) is a necessary condition for d_Λ to contain $d_{\Lambda'}$, as a subrepresentation.

This conjecture made by Verma has subsequently been proved to be correct, as is evident from Theorem 6. Verma has shown[6] that assuming the validity of this conjecture the following theorem holds.

Theorem 8: An elementary representation d_Λ contains those and only those irreducible representations D_M with highest weights M for which the elementary representation d_M is a subrepresentation of d_Λ.

Corollary: An elementary representation d_Λ contains an irreducible representation D_M, at most once.

In the following expressions are derived for the multiplicity of weights of (infinite-dimensional) elementary and irreducible representations with a highest weight. These expressions are the analog to the formulas obtained by Racah[13] and Kostant[14] for the multiplicity of weights of finite-dimensional irreducible representations.

In order to derive expressions for the multiplicity of weights an algebra U over the field of complex numbers C is defined in the following manner.[8] A one to one correspondence is established between the linear forms M on G^0 and formal exponents $e(M)$. In this set of formal exponents a multiplication is introduced,

$$e(M)e(M')=e(M+M'). \tag{42}$$

Then the elements of the algebra U are defined as (in general infinite) sums over the exponents $e(M)$ with coefficients in C,

$$\sum_M a_M e(M),\quad a_M\in C, \tag{43}$$

with the condition that this sum belongs to U if and only if there exists a weight such that any weight M of the

sum can be written as

$$M=\Lambda-\sum_{\alpha_i\in\Delta} m_i\alpha_i,\quad m_i\geq 0,\ \text{integer},$$

where the sum extends over all positive roots α_i. The multiplication in the algebra U is given by Eq. (42) and it is easy to see that $e(0)$ is the unit element for U. It can be shown that the algebra U forms a commutative integral domain.

An element of U,

$$\chi=\sum_M n_M e(M), \tag{44}$$

is called a *character of* a representation ρ_Λ with highest weight Λ of an algebra G, if the n_M are the multiplicities of the weights M of the representation ρ_Λ. For convenience the summation in Eq. (44) is taken over *all* weights, setting $n_M=0$ if M is not a weight of the representation ρ_Λ.

According to Theorem 4 the multiplicity of a weight M of an elementary representation d_Λ is given as $P(\Lambda-M)$. Thus, the character of an elementary representation d_Λ is given as

$$\chi'_\Lambda=\sum_M P(\Lambda-M)e(M)=e(\Lambda)\sum_M P(\Lambda-M)e(M-\Lambda).$$

Since it holds that[8]

$$\sum_{\Lambda'} P(\Lambda')e(-\Lambda')\prod_{\alpha\in\Delta}(1-e(-\alpha))=1$$

and that

$$\sum_{S\in W}(\det S)e(SR)=e(R)\prod_{\alpha\in\Delta}(1-e(-\alpha)),$$

the character χ'_Λ for the elementary representation d_Λ is obtained as

$$\chi'_\Lambda=\frac{e(\Lambda)}{\prod_{\alpha\in\Delta}(1-e(-\alpha))}=\frac{e(\Lambda+R)}{\sum_{S\in W}(\det S)e(SR)}. \tag{45}$$

On the other hand, the character χ'_Λ can be written as

$$\chi'_\Lambda=\sum_M n_M e(M). \tag{46}$$

Equating Eqs. (45) and (46) and comparing coefficients of $e(M)$, the following recurrence relation is obtained for the multiplicities n_M^Λ of the weights of an elementary representation d_Λ,

$$\begin{aligned} n_M^\Lambda&=-\sum_{\substack{S\in W\\ S\neq 1}}(\det S)n_{M+R-SR},\quad M\neq\Lambda,\\ n_\Lambda^\Lambda&=1. \end{aligned} \tag{47}$$

This result is not surprising since the multiplicity structure of the elementary representations d_Λ is precisely that of the partition function. This recurrence relation for the partition function has first been obtained by Kostant and proving it is one of the exercises given by Jacobson (Ref. 8, p. 263). Since, however, elementary representations are basic to what follows a short proof of Kostant's result has been given above. A generalization of this formula has been obtained by one of the authors in Ref. 15.

In order to obtain similar recurrence relations for *irreducible* representations D_Λ, Theorems 6, 7, and 8 have to be taken into account. For a given weight M of D_Λ the multiplicity n_M is obtained from the multiplicity n_M^Λ of the elementary representation d_Λ by subtracting the multiplicity of the weight M of all subrepresentations of d_Λ. First a special case is considered. Let d_Λ denote an elementary representation and $d_{\Lambda'}$ its only subrepresentation. Then it must hold, from Theorem 6, that $\Lambda'=S(\Lambda+R)-R$ for some $S\in W$. Let n_M^Λ denote the multiplicity of the weight M of the representation d_Λ and $n_m^{\Lambda'}$ the multiplicity of the weight M of the representation $d_{\Lambda'}$. Then it holds

$$\begin{aligned} n_M^\Lambda&=-\sum_{\substack{S\in W\\ S\neq 1}}(\det S)n_{M+R-SR}^\Lambda,\quad M\neq\Lambda,\\ n_\Lambda^\Lambda&=1,\\ n_M^{\Lambda'}&=-\sum_{\substack{S\in W\\ S\neq 1}}(\det S)n_{M+R-SR}^{\Lambda'},\quad M\neq\Lambda',\\ n_{\Lambda'}^{\Lambda'}&=1. \end{aligned} \tag{48}$$

Thus, the multiplicity n_M of the weights M of the irreducible representation D_Λ is, for this special case, given as

$$\begin{aligned} n_M&=n_M^\Lambda-n_M^{\Lambda'}=-\sum_{\substack{S\in W\\ S\neq 1}}(\det S)n_{M+R-SR}-\delta_{M,\Lambda'},\quad M\neq\Lambda,\\ n_\Lambda&=1. \end{aligned} \tag{49}$$

It should be remembered that the multiplicity $n_M^{\Lambda'}$ of weights M of d_Λ which are not weights of $d_{\Lambda'}$ is zero. For $M=\Lambda'$, the third of the Eqs. (48) yields zero and thus gives a zero contribution to $n_{\Lambda'}$. The actual multiplicity $n_{\Lambda'}^{\Lambda'}=1$ is subtracted through the Kronecker symbol.

From equations (47) and (49) it is apparent that the formula for the multiplicity of a weight M of an elementary representation d_Λ, containing one subrepresentation $d_{\Lambda'}$, is identical to the formula for the multiplicity of the weight M of the irreducible representation D_Λ, except for the weight $M=\Lambda'$ of D_Λ. For this particular weight the value obtained from Eq. (47) has to be decreased by one in order to obtain the multiplicity $n_{\Lambda'}$ of the weight Λ' of D_Λ.

Theorem 9: If D_Λ is an irreducible representation of the algebra G with highest weight Λ, then the multiplicity n_M of a weight M of D_Λ is obtained as follows:

(a) if M is a weight of D_Λ such that

$$M+R\neq S'(\Lambda+R),\quad \text{any } S'\in W,\quad S'\neq 1,$$

or if

$$M+R=S'(\Lambda+R),\quad \text{for some } S'\in W,\quad S'\neq 1,$$

but there exists no sequence of positive roots $\beta_1,\beta_2,\ldots,\beta_n$ which satisfies condition (A) for the pair of weights $(M+R;\ \Lambda+R)$, then it holds

$$n_M=-\sum_{\substack{S\in W\\ S\neq 1}}(\det S)n_{M+R-SR};$$

(b) if M is a weight of D_Λ such that

$$M+R=S'(\Lambda+R),\quad \text{for some } S'\in W,\quad S'\neq 1,$$

and there exists a sequence $\beta_1,\beta_2,\ldots,\beta_n$ of positive roots which satisfies condition (A) for the pair of weights $(M+R;\ \Lambda+R)$, then

$$n_M = -\sum_{\substack{S\in W\\ S\neq 1}} (\det S) n_{M+R-SR} + \gamma_{S'},$$

with

$$\gamma_{S'} = -\sum_T \gamma_T - 1. \tag{50}$$

The sum in Eq. (50) goes over all elements T of the factor space W/W^Λ, where $W^\Lambda = \{S \in W, S(\Lambda + R) = \Lambda + R\}$, for which $T \notin S'W^\Lambda$, $T \notin W^\Lambda$ and for which a sequence of positive roots $\beta_1, \beta_2, \ldots, \beta_n$ exists which satisfies condition (A) for the pair of weights $(M+R, T(\Lambda+R))$. The Eq. (50) is a recurrence relation for the determination of the integers γ_T.

Proof: Part (a) of the theorem is correct since Eq. (47) holds for all weights of D_Λ which are *not* highest weights M of elementary subrepresentations d_M of the elementary representation d_Λ. The proof of the validity of part (b) is as follows. Let d_m be an elementary subrepresentation of d_Λ with highest weight M and let $M'' = T(\Lambda+R) - R$ be the lowest highest weight which corresponds to an elementary subrepresentation $d_{M''}$ of d_Λ, such that $d_{M''}$ contains d_M as a subrepresentation. Then it holds $\gamma_T = -1$.

In general it holds for the case of any subrepresentation $d_{M''}$ of d_Λ which contains d_M in turn as a subrepresentation, and does not contain any $d_{\tilde{M}}$ which contains d_M, that $\gamma_T = -1$. The same argument is now repeated for the subrepresentations $d_{M''}$ in place of the subrepresentation d_M. For given subrepresentations $d_{M''}$ all subrepresentations $d_{M'''}$ of d_Λ are considered which contain $d_{M''}$ (and its subrepresentations) as a subrepresentation, but no others. The information obtained in the first step determines the value of the corresponding γ_T. Continuing in this manner the value of $\gamma_{S'}$ is obtained. This completes the proof of Theorem 9.

Theorem 10. Let D_Λ denote an irreducible representation of an algebra G with highest weight Λ. Then its character and the multiplicity of its weight are given as

$$\chi_\Lambda = \frac{\sum_{S\in W/W_\Lambda} \gamma_S e(S(\Lambda+R))}{\sum_{S\in W} (\det S) e(SR)},$$

$$n_M = \sum_{S\in W/W_\Lambda} \gamma_S P(S(\Lambda+R) - (M+R))$$

with $\gamma_S = 0$ if $d_{S(\Lambda+R)-R}$ is *not* a subrepresentation of d_Λ.

Proof: The proof of this Theorem rests the theorems for the characters and the multiplicity of weights of elementary representations by following the lines of argumentation as given in the proof of Theorem 9.

VII. GEOMETRICAL INTERPRETATION OF CHARACTER AND MULTIPLICITY OF WEIGHTS FOR INFINITE-DIMENSIONAL REPRESENTATIONS WITH A HIGHEST WEIGHT

A simple geometrical interpretation is possible for the characters and the multiplicity of weights of elementary as well as irreducible representations with a highest weight. Such a geometrical description has been given by Antoine and Speiser[9] for the case of finite-dimensional irreducible representations, while Biedenharn, Gruber, and Weber and Klimyk[10] treated the special case of infinite-dimensional representations of the algebra A_2 along similar lines. For algebras of low rank the geometrical description lends itself to a graphical interpretation. In the following the algebra A_2 is chosen to demonstrate graphically the meaning of the formulas for the character and for the multiplicity of weights for its infinite-dimensional representations with a highest weight.

In Fig. 1 the root system of the algebra A_2 is plotted, with α_1 and α_2 simple (positive) roots. With respect to a conveniently chosen basis the simple roots can be represented as vectors $\alpha_1 = (1, -1, 0)$ and $\alpha_2 = (0, 1, -1)$. Then $R = (1, 0, -1)$. [This embedding of the weight space of the algebras A_l in an $(l+1)$-dimensional space is conventional. In this $(l+1)$-dimensional space the action of the Weyl group takes on its simplest form.] The three lines 1, 2, and 3, perpendicular to the positive roots α_1, α_2, and R, are reflection planes. All possible reflections and products of reflections on these lines (planes, in general) generate the Weyl group W of A_2. For the representation of the roots (and weights) chosen the action of the Weyl group consists simply in all possible permutations of the components of a root (weight). A set of weights related through the Weyl group is called a set of equivalent weights. Such a set of equivalent weights has been indicated in Fig. 1 by points. The shaded domain A is called the fundamental domain. It contains the highest weight of every set of equivalent weights and all other domains B-F are related to it through the Weyl group. If for a weight Λ holds that $\Lambda_i = 2(\Lambda, \alpha_i)/(\alpha_i, \alpha_i) = n$, $n \geq 0$, integer, for all simple roots α_i, $i = 1, 2, \ldots l$, then the irreducible representation which has Λ as highest weight is finite-dimensional. Its outer contour is obtained by joining the points in A and B, B and $C, \ldots, F$ and A, consecutively by straight lines. All highest weights of finite-dimensional irreducible representations lie in the fundamental domain A (including the boundary). A typical example is $\Lambda = (4, -1, -3)$ with $\Lambda_1 = 5$, $\Lambda_2 = 2$.

Figure 2 illustrates the case of an elementary representation d_Λ with highest weight $\Lambda = (4, -3, -1)$. It

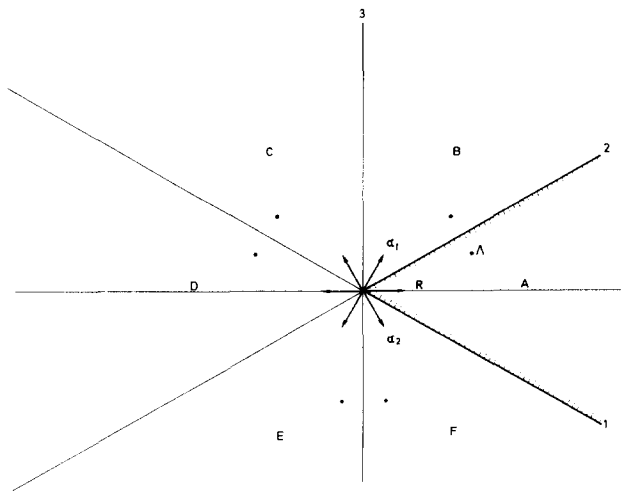


FIG. 1.

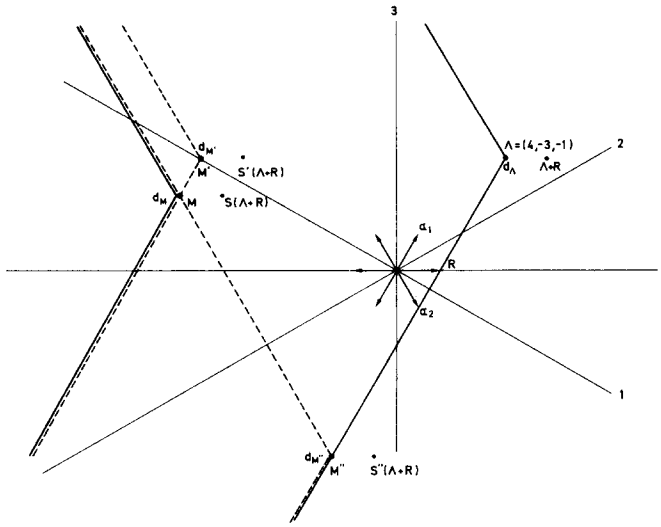


FIG. 2.

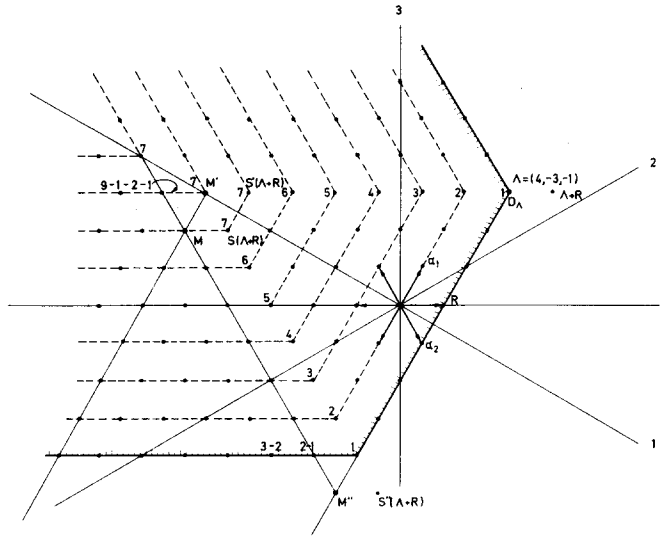


FIG. 4.

holds that $\Lambda_1 = 7$, $\Lambda_2 = -2$. This elementary representation is not irreducible, but contains $d_{M'}$, $d_{M''}$ and d_M as subrepresentations, with $M' = (-3,-3,6)$, $M'' = (-4,5,-1)$, and $M = (-4,-2,6)$. This is a consequence of Theorem 6, since condition A is satisfied for the pairs of weights $(M'+R;\ \Lambda+R)$, $(M''+R;\ \Lambda+R)$, and $(M+R;\ \Lambda+R)$. Property (1) of condition A holds obviously. While (2) of condition A, since $(\alpha_i,\ \alpha_i) = 2$, $(R, R) = 2$, takes on the form $(\Lambda+R, R) = 6$, $((\Lambda+R),\ \alpha_2) = 8$, and $(S'(\Lambda+R),\ \alpha_1) = ((-2,-3,5),\ \alpha_1) = 1$ for the three pairs obviously. In turn, *both* subrepresentations $d_{M'}$ and $d_{M''}$ contain d_M as a subrepresentation, while $d_M \equiv D_M$ is irreducible according to Theorem 5. The outer contour of the infinite-dimensional representations d_Λ, $d_{M'}$, $d_{M''}$, and D_M is indicated by the two straight lines (solid for d_Λ and d_M and broken for $d_{M'}$ and $d_{M''}$) which emerge from the highest weights and extend to infinity. The weights of the representations lie on the lines and to the left of the lines. In each of the subspaces $R_{M'}$, $R_{M''}$ and R_M of the space R, which carries the representation d_Λ, there exists a vector y which satisfies $\rho_\Lambda(e_i)y = 0$, $i = 1, 2, \ldots, l$.

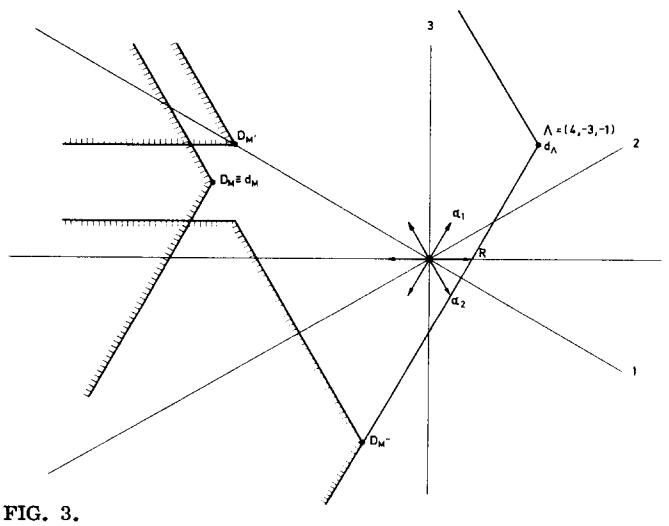


FIG. 3.

This is simply a consequence of the fact that d_M, $d_{M''}$ and d_M are subrepresentations. It should, however, be noted that, while d_Λ is reducible, it is not completely reducible. The subrepresentation d_M is a subrepresentation of both, d_M and $d_{M''}$.

Figure 3 shows the (infinite-dimensional) *irreducible* representations $D_{M'}$, $D_{M''}$, and $D_M \equiv d_M$ which are subquotients of the elementary representation d_Λ. According to Theorem 8 the highest weights are the same as for the elementary subrepresentations $d_{M'}$, $d_{M''}$, and d_M. It holds $M'_1 = 0$, $M'_2 = -9$ for $D_{M'}$, $M''_1 = -9$, $M''_2 = 6$ for $D_{M''}$ and $M_1 = -2$, $M_2 = -8$ for D_M. The positive values (more precisely, the nonnnegative integer values) of these projections of the highest weights onto the two simple roots α_1 and α_2 are directly related to the invariance (if any) of the weight diagram under Weyl reflections. Thus, the weight diagram of $D_{M'}$ is invariant with respect to reflections on the line perpendicular to the root α_1, the weight diagram of $D_{M''}$ is invariant with respect to reflections on the line perpendicular to the root α_2, while the weight diagram of D_M is not invariant under any operation of the Weyl group. Using a symbolic notation, Theorem 6 shows that $D_{M'} = d_{M'} - d_M$, $D_{M''} = d_{M''} - d_M$, and $D_M \equiv d_M$.

Figure 4 shows the irreducible representation D_Λ with its multiplicity structure. Again symbolically writing, the irreducible representation D_Λ is related to the elementary representation d_Λ and its elementary subrepresentations according to Theorem 8 as $D_M = d_\Lambda - (d_{M'} - d_M) - (d_{M''} - d_M) - d_M$. It is worth noticing that for irreducible representations D_Λ which have their highest weight within the fundamental domain, the familiar formula of Kostant for the multiplicity of weights is obtained in this manner. (The multiplicity within each elementary representation is given by the partition function.) Racah's formula for the case of infinite-dimensional irreducible representations has to be modified, according to Theorem 9, at the encircled weights, namely at M', M'' and M. Consecutively determining the γ_T it holds that $\gamma_{s'} = -1$, $\gamma_{s''} = -1$, and $\gamma_s = -(\gamma_{s'} + \gamma_{s''}) - 1 = 1$. Thus, at the weights M' and M'' the multiplicity obtained by Racah's formula has,

according to Theorem 9, to be decreased by 1, while at the weight M the multiplicity has to be increased by 1.

ACKNOWLEDGMENTS

One of the authors (A. U. K.) wishes to thank Southern Illinois University for the hospitality extended to him during a three months visit during which part of this article was written. The other author (B. G.) wishes to thank the Academy of Sciences of the Ukrainian SSR for the hospitality extended to him at Kiev as well as the authorities of the Dublin Institute for Advanced Studies, School of Theoretical Physics, for a visit during which the manuscript was completed. Finally, both authors wish to thank Professor J. Dixmier for his valuable comments.

*This work was partially supported through a travel grant (GF-37265) by the National Science Foundation, Washington, D.C.

[1]A. O. Barut, Phys. Rev. **135**, B839 (1964); Y. Dothan, M. Gell-Mann, and Y. Ne'eman, Phys. Rev. Lett. **17**, 148 (1965); N. Mukunda, L. O. 'Raifeartaigh, and E. C. G. Sudarshan, Phys. Rev. Lett. **19**, 322 (1965).

[2]A. O. Barut, Phys. Rev. **139**, B1433 (1965); R. C. Hwa and J. Nuyts, Phys. Rev. **145**, 1188 (1966).

[3a]I. M. Gel'fand and M. I. Graev, Izv. Akad. Nauk SSSR, Ser. Mat. **29**, 1329 (1965); I. T. Todorov, "Discrete Series of Hermitian Representations of the Lie Algebra of U (p, q)" ICTP, Trieste, Preprint i. c./66/71 (1966).

[3b]E. Majorana, Nuovo Cimento **9**, 335 (1932); I. M. Gel'fand and A. M. Yaglom, Zh. Exp. Theor. Fiz. **18**, 703, 1096, and 1105 (1965); Y. Nambu, Phys. Rev. **160**, 1171 (1967); A. O. Barut, "Application of the Dynamical Group Theory to the Structure of Hadrons," in *Lectures in Theoretical Physics* (Gordan & Breach, New York, 1968), Vol. X-B.

[4]A. O. Barut, Siam J. Appl. Math. **25**, 247 (1973).

[5]Harish-Chandra, Am. J. Math. **77**, 743 (1955).

[6]Daya-Nand Verma, Bull. Am. Math. Soc. **74**, 160 (1968).

[7]I. N. Bernstein, I. M. Gel'fand and S. I. Gel'fand, Funct. Anal. Appl. **5**, 1 (1971).

[8a]N. Jacobson, *Lie Algebras* (Interscience, New York, 1962).

[8b]J. Dixmier, *Algebres enveloppantes* (Gauthier-Villars, Paris, 1974), Collection: "Cabiers Scientifiques".

[9]J. P. Antoine and D. Speiser, J. Math. Phys. **5**, 1226 and 1560 (1965).

[10]L. C. Biedenharn, J. Nuyts, and N. Straumann, Ann. Inst. Henri Poincaré, III, 13 (1965); L. C. Biedenharn, B. Gruber, and H. J. Weber, Proc. Roy. Ir. Acad. 67, A1, 1 (1968); A. U. Klimyk, "Semisimple Lie Groups and Algebras and their Representations," preprint, Institute of Theoretical Physics, Kiev, 1968.

[11]Seminar "Sophus Lie" de l'Ecole Norm. Sup., 1954-55, Vol. 1, Secrétariat Mathématique, Paris.

[12]Harish-Chandra, Trans. Am. Math. Soc. **70**, 28 (1951).

[13]G. Racah, in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gürsey (Gordon and Breach, New York, 1964).

[14]B. Kostant, Trans. Am. Math. Soc. **93**, 53 (1959).

[15]B. Gruber, J. Math. Phys. **11**, 1783 (1970).

Path integral method for superfields

M. Huq*

Department of Mathematical Physics, University College, Belfield, Dublin 4, Ireland
(Received 26 December 1974)

A path integral method for superfields has been developed. This is made possible using a suitable definition for functional differentiations in the 8-dimensional space of x and θ. The Ward–Takahashi identities arising from the invariance under supersymmetry transformations are shown to take very neat and compact forms for the Green's functions of the superfields.

1. INTRODUCTION

Recently Wess and Zumino[1] introduced Fermi—Bose supersymmetry. The transformations of this are generated by infinitesimal coordinate dependent, anticommuting Majorana spinors $\alpha(x)$ in four dimensions. Later on[2] the coordinate dependence of α is dropped to avoid problems arising from scale and conformal anomalies. Salam and Strathdee[3,4] have shown that the constant parameter supersymmetry transformations can be considered as operations on superfields defined on the eight-dimensional space of coordinates x_μ, and the anticommuting Majorana spinors θ_α.

A remarkable feature of supersymmetry is that both bosons and fermions are accommodated in the same multiplet. Iliopoulos and Zumino[5] have studied extensively the simple case of the superfield containing a scalar field A, a pseudoscalar field B, a Majorana spinor field ψ, and two auxiliary fields F and G. These authors have shown that a nontrivial Langrangian model constructed out of these fields is not only renormalizable, but with the feature that only one wavefunction renormalization common to all the fields is required. There are no mass and coupling constant renormalizations. This latter feature is due to the specific structure of the Lagrangian. Tsao[6] has shown this can be shown to arise also from Ward—Takahashi (WT) identities for broken γ_5-invariance. Iliopoulos and Zumino[5] and Ferrara, Iliopoulos, and Zumino[7] have written down WT identities for invariance under supersymmetry transformations. Because of these we have remarkable cancellations between the Green's functions of the different fields. When considering perturbation theory, this would manifest as cancellations between different Feynman diagrams, hence getting rid of many bad divergences. By proceeding this way the study of divergences of multiloop diagrams and their possible cancellations can turn out to be very messy and rather tricky. If it is possible to deal with the superfields directly without decomposing them into component boson and fermion fields, then we shall be treating a host of Feynman diagrams collectively. This is expected to make everything much simpler. The very first step in this direction was taken in Appendix C of Ref. 4. Using this as a starting point, Capper[8] has written down Feynman rules for superfields directly and used them to show that the study of divergences of multiloop diagrams can indeed be quite straightforward.

In this paper, our aim is to show that a path integral method for superfields can be developed and the WT identities written in neat forms. For the purpose of illustration we shall consider the model of Iliopoulos and Zumino. Here one is dealing with the chiral and scalar superfields $\Phi_\pm(x,\theta)$. If we introduce suitable external sources $J_\pm(x,\theta)$, then it should be possible to generate all Green's functions by functional differentiations with respect to $J_\pm(x,\theta)$ of the generating functional. As it is pointed out in Appendix C of Ref. 4, a clarification of the meaning of $\delta/\delta J_\pm(x,\theta)$ is necessary before we can proceed any further. There this has been by-passed by making use of the invariance of the Green's functions under supersymmetry transformations. Salam and Strathdee have shown in a more recent paper[9] that a clarification of the meaning of $\delta/\delta J_\pm(x,\theta)$ is easily obtained. We have adopted this definition of $\delta/\delta J_\pm(x,\theta)$, using which it is possible to develop a consistent path integral method and derive WT identities directly in terms of superfields.

The plan of this paper is as follows. In Sec. 2 we have developed the path integral method. The Ward—Takahashi identities for supersymmetry transformations are derived in Sec. 3. Section 4 contains our conclusions. Some useful formulas are given in the Appendix.

2. PATH INTEGRAL METHOD

We shall adopt the notations of Ref. 4. These and some useful formulas for manipulating with the anticommuting Majorana spinors are given in the Appendix.

A supersymmetry transformation is

$$\begin{aligned} x_\mu &\to x_\mu + \tfrac{1}{2}i\bar{\epsilon}\gamma_\mu\theta, \\ \theta_\alpha &\to \theta_\alpha + \epsilon_\alpha, \end{aligned} \tag{2.1}$$

where ϵ_α is an anticommuting infinitesimal Majorana spinor. Under these the scalar superfields $\Phi_\pm(x,\theta)$ transforms as

$$\Phi_\pm(x,\theta) \to \Phi_\pm(x,\theta) + \bar{\epsilon}_\alpha \mathcal{D}_\alpha(x,\theta)\Phi_\pm(x,\theta) \tag{2.2}$$

with

$$\mathcal{D}_\alpha(x,\theta) \equiv \frac{\partial}{\partial\bar{\theta}_\alpha} + \frac{i}{2}(\gamma_\mu\theta)_\alpha \frac{\partial}{\partial x_\mu}. \tag{2.3}$$

We have for $\Phi_\pm(x,\theta)$ the following expansions:

$$\begin{aligned} \Phi_\pm(x,\theta) = \exp(\mp\tfrac{1}{4}\bar{\theta}\partial\gamma_5\theta) \\ \times\{A_\pm(x) + \bar{\theta}\psi_\pm(x) + \tfrac{1}{4}\bar{\theta}(1\pm i\gamma_5)\theta F_\pm(x)\}. \end{aligned} \tag{2.4}$$

To go over to the notations of Iliopoulos and Zumino we define

$$A_\pm = \tfrac{1}{2}(A \pm iB),$$
$$F_\pm = \tfrac{1}{2}(F \mp iG), \tag{2.5}$$
$$\psi_\pm = \tfrac{1}{2}(1 \pm i\gamma_5)\psi.$$

The Lagrangian constructed from superfields should transform like a scalar superfield. But for invariance under supersymmetry transformations it is necessary for the action to be invariant which require that all θ-dependent terms in the Lagrangian should be surface terms. By keeping this in mind, the following simple Lagrangian can be constructed using $\Phi_\pm(x,\theta)$:

$$\mathcal{L}(x,\theta) = \tfrac{1}{2}\bar{D}D\{\mathcal{L}_0 + \mathcal{L}_M + \mathcal{L}_g\}, \tag{2.6}$$

where

$$\mathcal{L}_0 = \tfrac{1}{4}\bar{D}D\{\Phi_+(x,\theta)\Phi_-(x,\theta)\}, \tag{2.7a}$$
$$\mathcal{L}_M = -\tfrac{1}{2}M\{\Phi_+^2(x,\theta) + \Phi_-^2(x,\theta)\}, \tag{2.7b}$$
$$\mathcal{L}_g = -\tfrac{1}{3}g\{\Phi_+^3(x,\theta) + \Phi_-^3(x,\theta)\}. \tag{2.7c}$$

The differential operator D_α is defined by

$$D_\alpha \equiv \frac{\partial}{\partial\bar{\theta}_\alpha} - \frac{i}{2}(\gamma_\mu\theta)_\alpha \frac{\partial}{\partial x_\mu} \tag{2.8}$$

and

$$\bar{D}_\alpha = C^{-1}_{\alpha\beta}D_\beta.$$

Now we introduce the external sources $J_\pm(x,\theta)$ with the expansions

$$J_\pm(x,\theta) = \exp(\mp\tfrac{1}{4}\bar{\theta}\partial\!\!\!/\gamma_5\theta) \times\{J_{F_\pm}(x) - \bar{\theta}J_{\psi_\pm}(x) + \tfrac{1}{4}\bar{\theta}(1 \pm i\gamma_5)\theta J_{A_\pm}(x)\}. \tag{2.9}$$

Then the generating functional $W[J]$ for Green's functions in presence of external sources may be written down except for an over-all constant factor as follows:

$$W[J] = \int d[\Phi]\exp\{iS[\Phi] + i\rho(\theta)\int d^4x[J_+(x,\theta)\Phi_+(x,\theta) + J_-(x,\theta)\Phi_-(x,\theta)]\}, \tag{2.10}$$

where $d[\Phi]$ is an invariant measure over the fields $\Phi_+(x,\theta)$ and $\Phi_-(x,\theta)$ and the invariant action $S[\Phi]$ is defined by

$$S[\Phi] = -\rho(\theta)\int d^4x\{\mathcal{L}_0(x,\theta) + \mathcal{L}_M(x,\theta) + \mathcal{L}_g(x,\theta)\} \tag{2.11}$$

and the differential operator $\rho(\theta)$ is defined by

$$\rho(\theta) = -\tfrac{1}{2}\bar{D}D,$$
$$\rho(\theta)\int d^4x \equiv \int d^4x[-\tfrac{1}{2}\bar{D}D]. \tag{2.12}$$

Now, to enable us to generate all the Green's functions from Eq. (2.10), we shall need the definition of $\delta/\delta J_\pm(x,\theta)$. We shall find the following definition, which is the same as that given by Salam and Strathdee,[9] to be appropriate:

$$\frac{\delta}{\delta J_\pm(x,\theta)} = \exp(\mp\tfrac{1}{4}\bar{\theta}\partial\!\!\!/\gamma_5\theta) \times\left(\frac{\delta}{\delta J_{A_\pm}(x)} + \bar{\theta}\frac{\delta}{\delta J_{\mp}(x)} + \tfrac{1}{4}\bar{\theta}(1 \pm i\gamma_5)\theta\frac{\delta}{\delta J_{F_\pm}(x)}\right). \tag{2.13}$$

Using Eq. (2.9), we can immediately calculate

$$\frac{\delta J_\pm(y,\theta_2)}{\delta J_\pm(x,\theta_1)} = \exp(\mp\tfrac{1}{4}\bar{\theta}_1\partial\!\!\!/\gamma_5\theta_1) \times\exp(\mp\tfrac{1}{4}\bar{\theta}_2\partial\!\!\!/\gamma_5\theta_2)\tfrac{1}{4}(\bar{\theta}_1 - \bar{\theta}_2)(1 \pm i\gamma_5)(\theta_1 - \theta_2)\,\delta^4(x-y) \equiv \delta_\pm(x,\theta_1;y,\theta_2) \tag{2.14}$$

and

$$\delta J_\mp(y,\theta_2)/\delta J_\pm(x,\theta_1) = 0; \tag{2.15}$$

we have the property

$$\rho(\theta)\int d^4x\,\Phi_\pm(x,\theta)\delta_\pm(x,\theta;y,\theta') = \Phi_\pm(y,\theta'). \tag{2.16}$$

The change of variable for functional differentiation is effected by

$$\frac{\delta}{\delta J_\pm(x,\theta_1)} = \rho(\theta)\int d^4y\left(\frac{\delta\Psi_+(y,\theta)}{\delta J_\pm(x,\theta_1)}\frac{\delta}{\delta\Psi_+(y,\theta)} + \frac{\delta\Psi_-(y,\theta)}{\delta J_\pm(x,\theta_1)}\frac{\delta}{\delta\Psi_-(y,\theta)}\right), \tag{2.17}$$

where

$$\Psi_\pm(x,\theta) = \exp(\mp\tfrac{1}{4}\bar{\theta}\partial\!\!\!/\gamma_5\theta) \times\{\mathcal{A}_\pm(x) + \bar{\theta}\mathcal{N}_\pm(x) + \tfrac{1}{4}\bar{\theta}(1 \pm i\gamma_5)\theta f_\pm(x)\} \tag{2.18}$$

and

$$\frac{\delta}{\delta\Psi_\pm(x,\theta)} = \exp(\mp\tfrac{1}{4}\bar{\theta}\partial\!\!\!/\gamma_5\theta) \times\left(\frac{\delta}{\delta f_\pm(x)} - \bar{\theta}\frac{\delta}{\delta\mathcal{N}_\mp(x)} + \tfrac{1}{4}\bar{\theta}(1 \pm i\gamma_5)\theta\frac{\delta}{\delta\mathcal{A}_\pm(x)}\right) \tag{2.19}$$

The consistency of Eq. (2.17) may be checked by choosing $\Psi_\pm = J_\pm$

$$\rho(\theta)\int d^4y\left(\frac{\delta J_+(y,\theta)}{\delta J_\pm(x,\theta_1)}\frac{\delta}{\delta J_+(y,\theta)} + \frac{\delta J_-(y,\theta)}{\delta J_\pm(x,\theta_1)}\frac{\delta}{\delta J_-(y,\theta)}\right) = \rho(\theta)\int d^4y\,\delta_\pm(x,\theta_1;y,\theta)\frac{\delta}{\delta J_\pm(y,\theta)} = \frac{\delta}{\delta J_\pm(x,\theta_1)}. \tag{2.20}$$

Now, since

$$\rho(\theta)\int d^4x\{J_+(x,\theta)\Phi_+(x,\theta) + J_-(x,\theta)\Phi_-(x,\theta)\} = \int d^4x\{J_{A_+}(x)A_+(x) + J_{F_+}(x)F_+(x) + \bar{J}_{\psi_-}(x)\psi_+(x) + J_{A_-}(x)A_-(x) + J_{F_-}(x)F_-(x) + \bar{J}_{\psi_+}(x)\psi_-(x)\}, \tag{2.21}$$

it is trivial to check using Eq. (2.13) that

$$\frac{\delta W[J]}{\delta J_\pm(x,\theta)} = i\langle 0|\Phi_\pm(x,\theta)|0\rangle. \tag{2.22}$$

To generate Green's functions of arbitrary order, all we have to do is to differentiate appropriate number of times. Let us now introduce the generating functional $Z[J]$ for connected Green's functions by

$$Z[J] = -i\ln W[J]. \tag{2.23}$$

Then

$$i^{N+M}\langle 0|T\{\Phi_+(x_1,\theta_1)\ldots\Phi_+(x_N,\theta_N)\,\Phi_-(y_1,\theta_1')\ldots\Phi_-(y_M,\theta_M')\}|0\rangle \tag{2.24}$$

$$= i \frac{\delta^{N+M} Z[J]}{\delta J_+(x_1,\theta_1)\ldots\delta J_+(x_N,\theta_N)\delta J_-(y_1,\theta_1')\ldots\delta J_-(y_M,\theta_M')}\Big|_{J_\pm=0}.$$

Now we can go ahead to introduce the generating functional $\Gamma[\Psi]$ for one particle irreducible (OPI) vertices. To do this, we have to introduce the vacuum expectation value of $\Phi_\pm(x,\theta)$ in presence of the sources $J_\pm(x,\theta)$:

$$\Psi_\pm(x,\theta) = \frac{\delta Z[J]}{\delta J_\pm(x,\theta)} = \langle 0|\Phi_\pm(x,\theta)|0\rangle_J. \tag{2.25}$$

We can use Eq. (2.18) and (2.19) to define expansion of $\Psi_\pm(x,\theta)$ and functional derivative with respect to $\Psi_\pm(x,\theta)$ respectively. Then observe $A_\pm(x)=\langle 0|A_\pm(x)|0\rangle$, $f_\pm(x)=\langle 0|F_\pm(x)|0\rangle_J$, and $\lambda_\pm(x)=\langle 0|\psi_\pm(x)|0\rangle_J$.
We have

$$\Gamma[\Psi] = Z[J] - \rho(\theta)\int d^4x \times\{J_+(x,\theta)\Psi_+(x,\theta)+J_-(x,\theta)\Psi_-(x,\theta)\}. \tag{2.26}$$

To generate OPI vertices, we have to differentiate with respect to $\Psi_\pm(x,\theta)$ sufficient number of times and then set $\Psi_\pm=0$. As, for example, for the inverse propagators

$$\Gamma_{\pm\pm}(x,\theta_1;y,\theta_2) = \frac{\delta^2\Gamma[\Psi]}{\delta\Psi_\pm(x,\theta_1)\delta\Psi_\pm(y,\theta_2)}, \tag{2.27a}$$

$$\Gamma_{\pm\mp}(x,\theta_1;y,\theta_2) = \frac{\delta^2\Gamma[\Psi]}{\delta\Psi_\pm(x,\theta_1)\delta\Psi_\mp(y,\theta_2)}. \tag{2.27b}$$

We introduce the Fourier transforms

$$\Gamma_{\pm\pm}(p;\theta_1,\theta_2) = \int d^4z\, e^{ipz}\Gamma_{\pm\pm}(x,\theta_1;y,\theta_2), \tag{2.28}$$

where $z=x-y$. A similar expression holds for $\Gamma_{\pm\mp}(p;\theta_1,\theta_2)$.

3. WARD–TAKAHASHI IDENTITIES

Following a method used by Lee and Zinn-Zustin[10] in the case of gauge theories, it is straightforward to derive the WT identities for supersymmetry transformations. Observe that the functional integral in Eq. (2.10) remains invariant under the change of integration variable $\Phi_\pm(x,\theta)\to\Phi_\pm(x,\theta)+\epsilon_\alpha D_\alpha(x,\theta)\Phi_\pm(x,\theta)$. Since $S[\Phi]$ is invariant under this, only the source terms in Eq. (2.10) are altered, and we are led to

$$\begin{aligned}&\int d[\Phi]\exp\{iS[\Phi]+i\rho(\theta)\int d^4x[J_+(x,\theta)\Phi_+(x,\theta)\\&\quad+J_-(x,\theta)\Phi_-(x,\theta)]\}\\&\times\{+i\bar\epsilon_\alpha\rho(\theta)\int d^4y[J_+(y,\theta)D_\alpha(y,\theta)\Phi_+(y,\theta)\\&\quad+J_-(y,\theta)D_\alpha(y,\theta)\Phi_-(y,\theta)]\}=0.\end{aligned} \tag{3.1}$$

The second curly bracket may be taken outside of the functional integration if we replace $\Phi_\pm(y,\theta)$ by $\delta/i\delta J_\pm(y,\theta)$. Then we have

$$\rho(\theta)\int d^4x\left\{J_+(x,\theta)D_\alpha(x,\theta)\frac{\delta}{\delta J_+(x,\theta)} + J_-(x,\theta)D_\alpha(x,\theta)\frac{\delta}{\delta J_-(x,\theta)}\right\}W[J]=0. \tag{3.2}$$

This can be expressed in terms of $Z[J]$:

$$\rho(\theta)\int d^4x\left\{J_+(x,\theta)D_\alpha(x,\theta)\frac{\delta Z[J]}{\delta J_+(x,\theta)} + J_-(x,\theta)D_\alpha(x,\theta)\frac{\delta Z[J]}{\delta J_-(x,\theta)}\right\}=0. \tag{3.3}$$

This is the WT identity satisfied by $Z[J]$ in presence of the sources $J_\pm(x,\theta)$. Now observe for any chiral fields $\Phi_\pm(x,\theta)$ we have the result

$$\rho(\theta)\int d^4x\frac{\partial}{\partial\bar\theta_\alpha}\Phi_\pm(x,\theta)=0 \quad\text{(surface term)}. \tag{3.4}$$

Using this, one can easily verify that

$$\begin{aligned}&\rho(\theta)\int d^4x\Phi_{1\pm}(x,\theta)D_\alpha(x,\theta)\Phi_{2\pm}(x,\theta)\\&\quad=-\rho(\theta)\int d^4x\Phi_{2\pm}(x,\theta)D_\alpha(x,\theta)\Phi_{1\pm}(x,\theta).\end{aligned} \tag{3.5}$$

Then the WT identity Eq. (3.3) may be rewritten in the form

$$\rho(\theta)\int d^4x\left\{\frac{\delta Z[J]}{\delta J_+(x,\theta)}D_\alpha(x,\theta)J_+(x,\theta) + \frac{\delta Z[J]}{\delta J_-(x,\theta)}D_\alpha(x,\theta)J_-(x,\theta)\right\}=0. \tag{3.6}$$

By working through the actions of $D_\alpha(x,\theta)$ and $\rho(\theta)$, it is straightforward to check that Eq. (3.6) is identical to Eq. (17) of Iliopoulos and Zumino,[5] which is the WT identity in terms of component fields. To do this, we observe that

$$\begin{aligned}D_\alpha(x,\theta)J_\pm(x,\theta) &= \exp(\mp\tfrac14\bar\theta\partial\!\!\!/\gamma_5\theta)\\&\{-J_{\psi\pm\alpha}(x)+[(J_{A_\pm}(x)+i\partial\!\!\!/J_{F_\pm}(x))\tfrac12(1\pm i\gamma_5)\theta]_\alpha\\&+\tfrac14\bar\theta(1\pm i\gamma_5)\theta(i\partial\!\!\!/J_{\psi_\pm}(x))_\alpha\}.\end{aligned} \tag{3.7}$$

Next making use of the rule of multiplication of two (+) and two (−) fields, we obtain

$$\begin{aligned}&\int d^4x\left(i\frac{\delta Z[J]}{\delta J_{A_+}(x)}\partial\!\!\!/J_{\psi_+}(x)-\frac{\delta Z[J]}{\delta J_{F_+}(x)}J_{\psi_+}(x)\right.\\&+[J_{A_+}(x)+i\partial\!\!\!/J_{F_+}(x)]\frac{1+i\gamma_5}{2}\frac{\delta Z[J]}{\delta\bar J_{\psi_-}(x)}\\&+i\frac{\delta Z[J]}{\delta J_{A_-}(x)}\partial\!\!\!/J_{\psi_-}(x)-\frac{\delta Z[J]}{\delta J_{F_-}(x)}J_{\psi_-}(x)\\&\left.+[J_{A_-}(x)+i\partial\!\!\!/J_{F_-}(x)]\frac{1-i\gamma_5}{2}\frac{\delta Z[J]}{\delta\bar J_{\psi_+}(x)}\right)=0.\end{aligned} \tag{3.8}$$

Using Eqs. (2.5), we have

$$\begin{aligned}&\frac{\delta Z[J]}{\delta J_A(x)}i\partial\!\!\!/J_\psi(x)-\frac{\delta Z[J]}{\delta J_B(x)}i\partial\!\!\!/\gamma_5J_\psi(x)\\&-\frac{\delta Z[J]}{\delta J_F(x)}J_\psi(x)-\frac{\delta Z[J]}{\delta J_G(x)}\gamma_5J_\psi(x)\\&+[J_A(x)+\gamma_5J_B(x)+i\partial\!\!\!/(J_F(x)-\gamma_5J_G(x))]\frac{\delta Z[J]}{\delta\bar J_\psi(x)}=0.\end{aligned} \tag{3.9}$$

This is precisely Eq. (17) of Ref. 5 in our notation.

Now returning to Eq. (3.3), let us differentiate it with respect to $J_\pm(y,\theta_1)$ and then set $J_\pm=0$ to obtain

$$\begin{aligned}&\rho(\theta)\int d^4x\exp(\mp\tfrac14\bar\theta_1\partial\!\!\!/_y\gamma_5\theta_1)\exp(\mp\tfrac14\bar\theta\partial\!\!\!/_x\gamma_5\theta)\\&\times\{\tfrac14\bar\theta_1(1\pm i\gamma_5)\theta_1+\tfrac14\bar\theta(1\pm i\gamma_5)\theta-\tfrac12\bar\theta_1(1\pm i\gamma_5)\theta\}\delta^4(x-y)\\&\times D_\alpha(x,\theta)\frac{\delta Z[J]}{\delta J_\pm(x,\theta)}=0.\end{aligned} \tag{3.10}$$

By working out the actions of $D_\alpha(x,\theta)$ and $\rho(\theta)$, it is possible to perform the x-integration using $\delta^4(x-y)$.

Then we obtain

$$\exp(\mp\tfrac{1}{4}\bar{\theta}_1\not{\partial}\gamma_5\theta_1)\left[\frac{\delta Z[J]}{\delta\bar{J}_{\psi_\mp}(y)}+\left(\frac{\delta Z[J]}{\delta J_{F_\pm}(y)}+i\not{\partial}\frac{\delta Z[J]}{\delta J_{A_\pm}(y)}\right)\frac{1\pm i\gamma_5}{2}\theta_1\right.$$

$$\left.-\tfrac{1}{4}\bar{\theta}_1(1\pm i\gamma_5)\theta_1 i\not{\partial}\frac{\delta Z[J]}{\delta\bar{J}_{\psi_\mp}(y)}\right]=0. \tag{3.11}$$

But this is precisely

$$D_\alpha(y,\theta_1)\frac{\delta Z[J]}{\delta J_\pm(y,\theta_1)}=0. \tag{3.12}$$

Similarly, if we perform functional differentiations with respect to $J_+(x_1,\theta_1)\cdots J_+(x_N,\theta_N)$ and $J_-(y_1,\theta_1')\cdots J_-(y_M,\theta_M')$ and then set $J_\pm=0$, we shall obtain the WT identity for the $(N+M)$th order connected Green's function

$$\left(\sum_{i=1}^{N}D_\alpha(x_i,\theta_i)+\sum_{j=1}^{M}D_\alpha(y_j,\theta_j)\right) \tag{3.13}$$

$$\times\frac{\delta^{N+M}Z[J]}{\delta J_+(x_1,\theta_1)\cdots\delta J_+(x_N,\theta_N)\delta J_-(y_1,\theta_1')\cdots\delta J_-(y_M,\theta_M')}=0.$$

It is easy to see that for OPI vertices we shall have WT identities of the form

$$\left(\sum_{i=1}D_\alpha(x_i,\theta_i)+\sum_{j=1}D_\alpha(y_j,\theta_j')\right)$$

$$\times\frac{\delta^{N+M}\Gamma[\Psi]}{\delta\Psi_+(x_1,\theta_1)\cdots\delta\Psi_+(x_N,\theta_N)\delta\Psi_-(y_1,\theta_1')\cdots\delta\Psi_-(y_M,\theta_M')}=0. \tag{3.14}$$

4. CONCLUSIONS

To bring out the usefulness of the WT identities for the superfields, we shall first show the consequences of WT identity for the vacuum expectation value of $\Phi_\pm(x,\theta)$, i.e., Eq. (3.12). Note that, using the fact that the vacuum expectation values in the absence of external sources are coordinate independent, we find that Eq. (3.12) takes the form

$$\frac{\partial}{\partial\bar{\theta}_\alpha}\{\langle 0|A_\pm(x)|0\rangle$$

$$+\bar{\theta}\langle 0|\psi_\pm(x)|0\rangle+\tfrac{1}{4}\bar{\theta}(1\pm i\gamma_5)\theta\langle 0|F_\pm(x)|0\rangle\}=0. \tag{4.1}$$

which immediately leads to

$$\langle 0|\psi_\pm(x)|0\rangle=0, \tag{4.2}$$

$$\langle 0|F_\pm(x)|0\rangle=0. \tag{4.3}$$

Equation (4.2) is simply a consequence of the spinoral nature of $\psi_\pm(x)$, while Eq. (4.3) is a consequence of invariance under supersymmetry transformations.

Next, taking $N=M=1$ in Eq. (3.14), we have WT identities for the inverse propagators

$$[D_\alpha(x,\theta_1)+D_\alpha(y,\theta_2)]\Gamma_{+\pm}(x,\theta_1,y,\theta_2) \tag{4.4}$$

Using Eq. (2.28), we have in momentum space

$$\left[\frac{\partial}{\partial\bar{\theta}_{1\alpha}}+\frac{\partial}{\partial\bar{\theta}_{2\alpha}}+\tfrac{1}{2}\{\not{p}(\theta_1-\theta_2)\}_\alpha\right]\Gamma_{+\pm}(p;\theta_1,\theta_2)=0. \tag{4.5}$$

Because of chiral nature of $\Gamma_{++}(x,\theta_1;y,\theta_2)$ we have [see Eq. (A8)]

$$\left(\frac{1-i\gamma_5}{2}\right)_{\alpha\beta}\left\{\frac{\partial}{\partial\bar{\theta}_{1\beta}}-\frac{i}{2}(\gamma_\mu\theta_1)_\beta\frac{\partial}{\partial x_\mu}\right\}\Gamma_{++}(x,\theta_1;y,\theta_2)=0, \tag{4.6}$$

$$\left(\frac{1-i\gamma_5}{2}\right)_{\alpha\beta}\left\{\frac{\partial}{\partial\bar{\theta}_{2\beta}}-\frac{i}{2}(\gamma_\mu\theta_2)_\beta\frac{\partial}{\partial y_\mu}\right\}\Gamma_{++}(x,\theta_1;y,\theta_2)=0.$$

Or in momentum space

$$\left(\frac{1-i\gamma_5}{2}\right)_{\alpha\beta}\left\{\frac{\partial}{\partial\bar{\theta}_{1\beta}}-\tfrac{1}{2}(\not{p}\theta_1)_\beta\right\}\Gamma_{++}(p;\theta_1,\theta_2)=0,$$

$$\left(\frac{1-i\gamma_5}{2}\right)_{\alpha\beta}\left\{\frac{\partial}{\partial\bar{\theta}_{2\beta}}+\tfrac{1}{2}(\not{p}\theta_2)_\beta\right\}\Gamma_{++}(p;\theta_1,\theta_2)=0. \tag{4.7}$$

A first look at Eqs. (4.5) and (4.7) suggests that the most general form for $\Gamma_{++}(p;\theta_1,\theta_2)$ can be taken to be

$$\Gamma_{++}(p;\theta_1,\theta_2)$$

$$=\exp(\tfrac{1}{4}a\bar{\theta}_1 i\not{p}\gamma_5\theta_1+\tfrac{1}{4}b\bar{\theta}_2 i\not{p}\gamma_5\theta_2+\tfrac{1}{2}c\bar{\theta}_1\not{p}\theta_2$$

$$+\tfrac{1}{2}d\bar{\theta}_1 i\not{p}\gamma_5\theta_2)\times[C(p^2)+(\bar{\theta}_1-\bar{\theta}_2)$$

$$\times\tfrac{1}{2}(1+i\gamma_5)(\theta_1-\theta_2)D(p^2)]. \tag{4.8}$$

Inserting this in Eqs. (4.5) and (4.7) and using Eq. (A7), we can easily show that $a=-b=1$, $c=d=0$, $C(p^2)=0$. Just for the sake of convenience we shall take out a factor $\frac{1}{2}M$ and write

$$\Gamma_{++}(p;\theta_1,\theta_2)=\exp[\tfrac{1}{4}\bar{\theta}_1 i\not{p}\gamma_5\theta_1-\tfrac{1}{4}\bar{\theta}_2 i\not{p}\gamma_5\theta_2]$$

$$\times\tfrac{1}{2}M(\bar{\theta}_1-\bar{\theta}_2)\tfrac{1}{2}(1+i\gamma_5)(\theta_1-\theta_2)D(p^2). \tag{4.9}$$

For the propagator $\Gamma_{+-}(p;\theta_1,\theta_2)$ Eqs. (4.7) are replaced

$$\left(\frac{1-i\gamma_5}{2}\right)_{\alpha\beta}\left\{\frac{\partial}{\partial\bar{\theta}_{1\beta}}-\tfrac{1}{2}(\not{p}\theta_1)_\beta\right\}\Gamma_{+-}(p;\theta_1,\theta_2)=0,$$

$$\left(\frac{1+i\gamma_5}{2}\right)_{\alpha\beta}\left\{\frac{\partial}{\partial\bar{\theta}_{2\beta}}+\tfrac{1}{2}(\not{p}\theta_2)_\beta\right\}\Gamma_{+-}(p;\theta_1,\theta_2)=0. \tag{4.10}$$

Then Γ_{+-} can be shown to have the form

$$\Gamma_{+-}(p;\theta_1,\theta_2)$$

$$=\exp[\tfrac{1}{4}\bar{\theta}_1 i\not{p}\gamma_5\theta_1+\tfrac{1}{4}\bar{\theta}_2 i\not{p}\gamma_5\theta_2+\tfrac{1}{2}\bar{\theta}_1\not{p}(1-i\gamma_5)\theta_2]C(p^2). \tag{4.11}$$

We introduce the renormalized propagators by

$$\Gamma^r_{+\pm}(p;\theta_1,\theta_2)=Z\Gamma_{+\pm}(p;\theta_1,\theta_2) \tag{4.12}$$

and take

$$\Gamma^r_{++}(0;\theta_1,\theta_2)=\tfrac{1}{2}M_r(\bar{\theta}_1-\bar{\theta}_2)\tfrac{1}{2}(1+i\gamma_5)(\theta_1-\theta_2),$$

$$\Gamma^r_{+-}(0;\theta_1,\theta_2)=1. \tag{4.13}$$

It has been shown by Iliopoulos and Zumino[5] that because of the special relation

$$\frac{\partial}{\partial M}\mathcal{L}_M=\frac{1}{2g}\left(\frac{\partial}{\partial\Phi_+}+\frac{\partial}{\partial\Phi_-}\right)\mathcal{L}_g \tag{4.14}$$

being satisfied by the Lagrangian Eq. (2.6), the following condition is satisfied by $\Gamma[\Psi]$:

$$\frac{\partial}{\partial M}\Gamma[\Psi]=-\frac{M}{2g}\rho(\theta)\int d^4y[\Psi_+(y,\theta)+\Psi_-(y,\theta)]$$

$$+\frac{1}{2g}\rho(\theta)\int d^4y\left(\frac{\delta\Gamma[\Psi]}{\delta\Psi_+(y,\theta)}+\frac{\delta\Gamma[\Psi]}{\delta\Psi_-(y,\theta)}\right). \tag{4.15}$$

This is Eq. (40) of Ref. 5 in our notation. Taking functional derivatives with respect to $\Psi_+(z,\theta_1)$ and setting

$\Psi_\pm = J_\pm = 0$, we obtain

$$\rho(\theta)\int d^4y\{\Gamma_{++}(z,\theta_1;y,\theta)+\Gamma_{+-}(z,\theta_1;y,\theta)\} = +M\exp(-\tfrac{1}{4}\bar{\theta}_1\partial\!\!\!/_z\gamma_5\theta_1) \times\int d^4y\,\delta^4(y-z) = +M. \quad (4.16)$$

Going over to momentum space, we have

$$\rho(\theta)\{\Gamma_{++}(0;\theta_1,\theta)+\Gamma_{+-}(0;\theta_1,\theta)\} = +M. \quad (4.17)$$

Using Eqs. (4.12) and (4.13), we arrive at the result

$$M_r = ZM, \quad (4.18)$$

showing that no mass renormalization counterterms is necessary. It is left to the interested readers to show that no coupling constant renormalizations are necessary as well.

In this paper we restricted ourselves to chiral fields $\Phi_\pm(x,\theta)$, but all the results can be easily extended to deal with general nonchiral fields $\Phi(x,\theta) = \Phi_+(x,\theta) + \Phi_-(x,\theta) + \Phi_1(x,\theta)$. The notable difference is that now the operator $\rho(\theta)$ should be replaced by the operator $\frac{1}{2}(\bar{D}D)^2$.

Salam and Strathdee[9] have derived in their paper the Feynman rules for superfields using the path integral method, while in this paper our main objective has been the derivation of Ward–Takahashi identities. These should be extremely valuable in the studies of renormalizability of superfield theories.

APPENDIX

We shall take the γ-matrices to satisfy

$$\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2g_{\mu\nu} \quad (A1)$$

with

$$g_{00} = -g_{11} = -g_{22} = -g_{33} = 1, \quad \gamma_5^2 = -1 \quad (A2)$$

and

$$\gamma^0\gamma_\mu^\dagger\gamma^0 = \gamma_\mu. \quad (A3)$$

The Majorana spinor ψ is such that

$$\psi = \psi^c = C\bar{\psi}^T \quad (A4)$$

with

$$C^{-1} = C^T = -C, \quad C^{-1}\gamma_\mu C = -\gamma_\mu^T. \quad (A5)$$

A very important identity satisfied by the Majorana spinor θ_α is

$$\theta_\alpha\bar{\theta}_\beta = -\tfrac{1}{4}\delta_{\alpha\beta}\bar{\theta}\theta + \tfrac{1}{4}(\gamma_5)_{\alpha\beta}\bar{\theta}\gamma_5\theta + \tfrac{1}{4}(i\gamma^\mu\gamma_5)_{\alpha\beta}\bar{\theta}i\gamma_\mu\gamma_5\theta. \quad (A6)$$

The following identity is very useful:

$$\left(\frac{1\pm i\gamma_5}{2}\theta\right)_\alpha\bar{\theta}\,\frac{1\pm i\gamma_5}{2}\theta = 0. \quad (A7)$$

The chiral fields $\Phi_\pm(x,\theta)$ satisfy

$$\left(\frac{1\mp i\gamma_5}{2}D\right)_\alpha\Phi_\pm(x,\theta) = 0 \quad (A8)$$

It is quite straightforward to prove these identities and the proofs can be found in Refs. 4 and 8.

*Present address: Department of Physics, University of Nigeria, Nsukka, Nigeria.

[1]J. Wess and B. Zumino, Nucl. Phys. B 70, 39 (1974).
[2]J. Wess and B. Zumino, Phys. Lett. 49B, 52 (1974).
[3]A. Salam and J. Strathdee, Nucl. Phys. B 76, 477 (1974).
[4]A. Salam and J. Strathdee, Trieste Preprint IC/74/42 (1974), submitted to Phys. Rev.
[5]J. Iliopoulos and B. Zumino, Nucl. Phys. B 76, 310 (1974).
[6]Hung-Sheng Tsao, Brandeis preprint (1974).
[7]S. Ferrara, J. Iliopoulos, and B. Zumino, Nucl. Phys. B 77, 413 (1974).
[8]D. M. Capper, Trieste Preprint IC/74/66 (1974).
[9]A. Salam and J. Strathdee, Nucl. Phys. B 86, 142 (1975).
[10]B. W. Lee and J. Zinn-Zustin, Phys. Rev. D 7, 1049 (1972).

Random function theory revisited: Exact solutions versus the first order smoothing conjecture

I. Lerche and E. N. Parker

Enrico Fermi Institute, University of Chicago, Chicago, Illinois
(Received 6 May 1975)

We remark again that the mathematical conjecture known as first order smoothing or the quasilinear approximation does not give the correct dependence on correlation length (time) in many cases, although it gives the correct limit as the correlation length (time) goes to zero. In this sense, then, the method is unreliable.

In a recent paper Roberts and Soward[1] urge upon the reader the concept that earlier papers by ourselves are riddled with mathematical errors. Insofar as there is a central point to their paper, it seems to be the assertion that we[2] have provided no valid mathematical examples in which first order smoothing theory (quasilinear theory, or the adiabatic approximation) gives erroneous results. We are flattered by their fixation on our work on this question, but in all modesty we must point out that Kraichnan,[3] Frisch,[4] Herring,[5] and others earlier had provided several examples of the invalidity of first order smoothing theory (FOST). Indeed Frisch[4] has put the point succinctly: "Using dishonest methods is like gambling: one does not know in advance whether the result will be valid or not." It is curious, in fact, that Roberts and Soward purport to quote from our paper, but delete from the text of the quotation our reference to the earlier examples of error in FOST provided by other authors.

To illustrate the difficulty with FOST, we[2] worked out the normal modes for two different equations

$$\frac{\partial^2 B}{\partial t^2} - [1 + \epsilon\, \delta v(t)] \frac{\partial B}{\partial x} = 0$$

and

$$\frac{\partial^2 B}{\partial t^2} - [1 + \epsilon\, \delta v(x)] \frac{\partial B}{\partial x} = 0,$$

where δv is a random function of either t or x. We used a method developed by Uhlenbeck and Orstein[6] to obtain *exact* solutions (normal modes), the dispersion relation appearing then as an infinite determinant which can be summed to any desired order in ϵ. We also obtained the dispersion relation for the normal modes using the truncation of the original equations as prescribed by FOST.

Comparison showed that if δv is a function of t, FOST gives a dispersion relation for the normal modes which disagrees with the exact theory in the first correction term $O(\epsilon^2)$, i.e., it agrees in the limit for ϵ rigorously zero, but FOST dispersion relation does *not* correctly include the effect of nonvanishing ϵ. This was enough for us to declare FOST untrustworthy. We should emphasize that in our exact solution, to which the result of FOST compared unfavorably, we obtained the *exact* solution of the original equations in the form of an infinite determinant. The determinant can then be summed to any order. We summed it to $O(\epsilon^2)$. Roberts and Soward in their solution first *approximate* the original equations by a moment scheme correct to $O(\epsilon^2)$. Then they obtain the *exact* solution of those *approximate* equations. The solution of their approximate equations agrees with the result of FOST but not with our solution of the exact equations. They have established, then, that the exact solution of approximate equations (such as FOST, or the moment equations) is not the same as an approximate solution of the exact equations. Their work underscores again the dangers of schemes such as FOST that start out by approximating the equations. The only way to be sure of the result is to solve the *exact* equations with a systematic mathematical approximation scheme. Indeed, Roberts and Soward attack the problem correctly in their Sec. III, employing an asymptotic matching method to the exact equations. There they find an eigenvalue approximately 8 percent different from the value obtained from FOST and from their exact solution of the truncated moment equations. Thus they join us in showing the unreliability of exact solutions of approximate equations, such as FOST.

Now, when δv is a function of x we found that FOST gives the *correct* contribution $O(\epsilon^2)$ to the dispersion relation for the individual modes. Thus FOST is correct in one case, $\delta v(x)$, but wrong in the other, $\delta v(t)$.

Curiously enough Soward and Roberts criticize our solution for the normal modes when δv is a function of x (the one which agreed with FOST) on the grounds that the solution is unphysical. They assert that the probability must be restricted to $x > 0$, just as in an initial value problem one works out the solution for $t > 0$. They are correct, of course, that any one normal mode is indeed unphysical. As is well known, physical solutions a are made up of a suitable superposition of normal modes. We went only as far as the normal modes, which is, of course, a legitimate mathematical exercise.

But suppose that our solution for δv a function of x is incorrect, as Roberts and Soward assert. Then the *incorrect* solution agrees with FOST, implying that the *correct* solution must *disagree* with FOST. Then in *both* cases FOST gives the *wrong* answer.

As a matter of fact, we[7] have reason to believe that FOST may be reliable when applied to equations that are fully self adjoint both before and after the truncation. If our conjecture is correct, then the agreement when δv is a function of x is not at all fortuitous. Altogether, it appears that Roberts and Soward have joined the ranks of those who have independently demonstrated

errors inherent in FOST. They would be interested, we are sure, in a recent paper by Jones and Birmingham[8] discussing in general terms when FOST is, and is not, valid.

Finally we should remark on Footnote 27 in the paper of Roberts and Soward, in which they assert that we[9] incorrectly evaluated a certain integral. It is evident from their cursory inspection of the bound on the integrand that they have overlooked the phase factor, which is rapidly varying and which accomplishes the necessary convergence in a standard manner (Jordan's lemma). We will be glad to supply them the complete and elementary, but tedious, details of the evaluation of the integral.

In conclusion we suggest that Roberts and Soward's statement, "If the claims of Lerche and Parker prove well founded, turbulent dynamo theory will... suffer a serious setback," taken by itself is too harsh a judgment of the earlier work of Steenbeck and others. As we pointed out,[2] the work of Steenbeck, Krause, and Radler[10] is entirely correct, so far as we are aware, in the short sudden limit $\epsilon \to 0$, in which we first worked out the dynamo equations twenty years ago.[11] Consequently, their derivations of the dynamo equations and their discussion[10,12–14] of the physical consequences of cyclonic turbulence (helicity) in generating magnetic fields are correct and introduce a number of important new physical points, as well as providing a broad look at a variety of dynamo models.

ACKNOWLEDGMENTS

This work was supported by the National Aeronautics and Space Administration under Grant NGL 14-001-001, and by the United States Air Force, Air Force Systems Command under Contract F-19628-75-C-0101.

[1] P.H. Roberts and A.M. Soward, J. Math. Phys. **16**, 609 (1975).
[2] I. Lerche, and E.N. Parker, J. Math. Phys. **14**, 1949 (1973).
[3] R.H. Kraichnan, J. Math. Phys. **2**, 124 (1961).
[4] U. Frisch, "Wave Propagation in Random Media," in *Probabilistic Methods in Applied Mathematics*, Vol. 1, edited by A.T. Rharucha-Reid (Academic, New York, 1968).
[5] J.R. Herring, Phys. Fluids **12**, 39 (1969).
[6] G.E. Uhlenbeck and L.S. Ornstein, Phys. Rev. **35**, 823 (1930).
[7] I. Lerche, J. Math. Phys. **15**, 1967 (1974).
[8] F.C. Jones and T.J. Birmingham, Plasma Phys. **17**, 15 (1975).
[9] I. Lerche, J. Math. Phys. **14**, 1381 (1973).
[10] M. Steenbeck, F. Krause, and K.H. Radler, Z. Naturforsch **21a**, 369 (1966).
[11] E.N. Parker, Astrophys. J. **122**, 293 (1955).
[12] I. Lerche, Astrophys. J. **166**, 627 (1971); **181**, 993 (1973); and references cited therein.
[13] M. Steenbeck and F. Krause, Astron. Nach. **291**, 49 (1969).
[14] F. Krause, Z. Angew. Math. Mech. **48**, 333 (1968).

A three-dimensional neutron transport problem*

Luigi Mangiarotti

Istituto Matematico "U. Dini," Università di Firenze, Viale Morgani 67/A, 50134 Firenze, Italy
(Received 2 January 1975)

We study the stationary neutron transport Boltzmann equation as applied to a three-dimensional system D, made by a rectangular multiplying core surrounded by a finite reflector, in both the Lebesgue space $L_2(D)$ and the space $C(D)$ (with the sup norm). As a result of this analysis, we prove some basic properties, such as the continuous dependence of the neutron flux on the parameters characterizing both the geometrical and the material properties of the system and the continuous and monotonic dependence of the average number of secondary neutrons per collision in the core on these parameters.

1. INTRODUCTION

In a recent work,[1] some techniques based on functional analysis were used to study the properties of the solution of the neutron transport Bolztmann equation for slabs and spheres with finite reflectors. In this paper, we are interested in investigating the case of a three-dimensional rectangular multiplying core surrounded by equal finite reflectors on opposite sides. To be definite, let us consider in the three-dimensional Euclidean space R^3 a finite closed rectangular system, with the center at the origin and embedded in the vacuum, whose material properties are characterized by c ($c \geq 1$) and by γ ($0 < \gamma < 1$), the average number of secondary neutrons per collision respectively in the core and in the reflector, and by $\sum$ ($\sum > 0$), the total macroscopic cross section for all processes both in the core and in the reflector (fission, scattering, and absorption in the core; scattering and absorption in the reflector). Neutrons are supposed monoenergetic, and the processes are taken to be spherically symmetric in the laboratory system.

Let now τ_i ($\tau_i > 0$) be the optical half-thickness ($\Sigma \times$ geometrical length) of the system along the coordinate axis x_i ($i = 1, 2, 3$); if α_i ($\alpha_i > 0$) and β_i ($\beta_i \geq 0$) respectively are the optical half-thickness of the core and the optical thickness on each side of the reflector, along the axis x_i ($i = 1, 2, 3$), then

$$\tau_i = \alpha_i + \beta_i, \quad \alpha_i > 0, \ \beta_i \geq 0, \quad i = 1, 2, 3. \tag{1}$$

By making use of the optical units τ_i along the axis x_i ($i = 1, 2, 3$), the domain of the system is $D = [-1, 1]^3 \subset R^3$; on the other hand, if C is the domain of the core, then

$$\begin{aligned} &C = [-\chi_1\chi_1] \times [-\chi_2, \chi_2] \times [-\chi_3, \chi_3], \\ &\chi_i = \alpha_i/\tau_i, \quad 0 < \chi_i \leq 1, \quad i = 1, 2, 3. \end{aligned} \tag{2}$$

Finally, let

$$\begin{aligned} &\tau = (\tau_1, \tau_2, \tau_3), \quad \sigma = (\chi_1, \chi_2, \chi_3, \gamma, c), \ p = (\tau, \sigma) \in P, \\ &P = (0, +\infty)^3 \times (0, 1]^3 \times (0, 1) \times [1, +\infty) \subset R^8; \end{aligned} \tag{3}$$

p is the parameter characterizing both the geometrical and the material properties of the system.

For the physical situation illustrated above and in the absence of external sources, the stationary neutron total flux $\phi_p(x)$ in the system must satisfy the linear integral Boltzmann equation[2]

$$\phi_p(x) = \int_D h_\sigma(x') K_\tau(x, x') \phi_p(x') \, dx', \tag{4}$$

where

$$K_\tau(x, x') = \tau_1\tau_2\tau_3 \exp\left[-\left(\sum_{i=1}^{3} \tau_i^2 (x_i - x_i')^2\right)^{1/2}\right] \Big/ 4\pi \sum_{i=1}^{3} \tau_i^2 (x_i - x_i')^2 \tag{5}$$

and

$$h_\sigma(x) = \begin{cases} c, & x \in C, \\ \gamma, & x \in D \setminus C. \end{cases} \tag{6}$$

We shall study the Eq. (4) by making use of both the integral operator valued functions T_p and S_p, $p \in P$, whose kernels (using the same symbol for the operator and for its kernel) respectively are

$$T_p(x, x') = h_\sigma(x') K_\tau(x, x'), \quad (x, x') \in D \times D, \tag{7}$$

and the symmetrized one

$$\begin{aligned} &S_p(x, x') = k_\sigma(x, x') K_\tau(x, x') \\ &k_\sigma(x, x') = [h_\sigma(x) h_\sigma(x')]^{1/2} \end{aligned} \quad , \quad (x, x') \in D \times D. \tag{8}$$

2. PROPERTIES OF THE OPERATORS T_p AND S_p

A. The first step is to choose the spaces to define the operators T_p and S_p, $p \in P$; since T_p and S_p are integral operators with a weak singularity [we recall that the kernels (7) and (8), in which there is an unbounded term such as $K_\tau(x, x')$, see (5), are said to have a weak singularity[3]], it turns out that it is possible to define them as follows[4]: For any $p \in P$, T_p is defined on all the space $C(D)$ of the real valued functions defined and continuous on D (endowed with the sup norm); for any $p \in P$, S_p is defined on all the space $L_2(D)$ of the real valued functions defined and square integrable (in the Lebesgue sense) on D [endowed with the inner product $(f, g) = \int_D f(x) g(x) \, dx$, $f, g \in L_2(D)$ and the L_2-norm $\|f\| = (f, f)^{1/2}$]. It turns out also that,[5] for any $p \in P$, the linear operator T_p and the symmetric linear operator S_p are completely continuous respectively in $C(D)$ and in $L_2(D)$ [but S_p is not a Fredholm operator, that is its kernel is not square integrable on $D \times D$, see (5) and (8)].

Let us emphasize that we have defined T_p in the space $C(D)$ [also it also would have been possible to define it in $L_2(D)$] and S_p in the space $L_2(D)$: This choice is motived by the fact that the eigenfunctions of S_p are

functions of the type $\sqrt{h_\sigma}\times f$, $f\in C(D)$, and hence T_p and S_p have the same eigenvalues and eigenfunctions (except the factor $\sqrt{h_\sigma}$). What we have said about the eigenfunctions of S_p can be seen by the two following arguments: In the first place, the iterated kernels of a kernel with a weak singularity, beginning at a certain one, are bounded[6] and, hence, the eigenfunctions of S_p are also bounded; in the second place, the operator S_p maps[7] the subspace of bounded functions of $L_2(D)$ into that of the functions of the type $\sqrt{h_\sigma}\times f$, $f\in C(D)$, and, hence, the eigenfunctions of S_p are also functions of this type.

Since the eigenfunctions of S_p are bounded, while in the kernel $S_p(x,x')$ there is an unbounded term such as $K_\tau(x,x')$, see (5) and (8), it follows at once that S_p is not degenerate, $p\in P$. We now show that S_p is positive definite, $p\in P$. Since the Fourier transform of $\exp(-|x|)/4\pi|x|^2$, $x\in R^3$, is $\tan^{-1}|\omega|/|\omega|$, $\omega\in R^3$, from (5) and (8), we get

$$(S_pf,f)=(2\pi)^{-3}\int_{R^3}\frac{\tan^{-1}H_\tau(\omega)}{H_\tau(\omega)}\,|F_\sigma(\omega)|^2\,d\omega, \tag{9}$$

where $f\in L_2(D)$ and

$$H_\tau(\omega)=\left(\sum_{i=1}^{3}\omega_i^2/\tau_i^2\right)^{1/2}, \tag{10}$$

$$F_\sigma(\omega)=\int_D\sqrt{h_\sigma(x)}\exp(i\omega x)f(x)\,dx. \tag{11}$$

Since $|F_\sigma(\omega)|^2\not\equiv 0$ if $f\in L_2(D)$ is not zero almost everywhere in D, the result follows.

At this point, we may state[8] that S_p (and hence T_p), $p\in P$, has a denumerably infinite set of positive eigenvalues forming a sequence $\lambda_1(p)\geq\lambda_2(p)\geq\cdots$ converging to zero (but zero is not an eigenvalue) and each eigenvalue is of finite multiplicity; the first eigenvalue is given by

$$\lambda_1(p)=\max_{\|f\|=1}(S_pf,f),\quad f\in L_2(D). \tag{12}$$

Finally, we give a simple proof of the fact (generally known) that the first eigenvalue $\lambda_1(p)$ is nondegenerate, $p\in P$. Since the kernel $S_p(x,x')$ is >0 in $D\times D$, $p\in P$, see (5), (6), and (8), it follows that $(S_pf,f)<(S_p|f|,|f|)$, if $f\in L_2(D)$ takes values of opposite signs over sets of nonzero measure in D. Therefore, from the maximum property (12) of $\lambda_1(p)$, we get that an eigenfunction of S_p corresponding to $\lambda_1(p)$ must be >0 or <0 almost everywhere in D and hence $\lambda_1(p)$ is necessarily simple. Likewise, an eigenfunction of T_p corresponding to $\lambda_1(p)$ must be >0 or <0 in D.

B. Afterwards we must frequently consider the case in which only one parameter, say τ_i, χ_i and so on [see (3)], is supposed to vary in $p\in P$: we agree that, instead of writing τ', σ', or p', we shall write merely τ_i', χ_i' and so on. We now prove

Theorem 1: T_p and S_p depend continuously on $p\in P$, that is, if $p'\to p$, then $\|T_{p'}-T_p\|\to 0$ and $\|S_{p'}-S_p\|\to 0$.

Proof: Let $f\in C(D)$, $\|f\|=1$; beginning from the parameters τ_i $(i=1,2,3)$ and recalling (7), we write

$$[(T_{\tau_i'}-T_p)f](x)$$

$$=\int_D h_\sigma(x')[K_{\tau_i'}(x,x')-K_\tau(x,x')]f(x')\,dx'. \tag{13}$$

Since from the definition (5) of $K_\tau(x,x')$ it follows that $\tau_i'K_\tau(x,x')>\tau_iK_{\tau_i'}(x,x')$, if $\tau_i'>\tau_i$, and that

$$\int_D K_\tau(x,x')\,dx'\leq\int_{R^3}K_\tau(x,x')\,dx'=1 \tag{14}$$

for any $x\in D$ and any τ, we get the inequalities

$$\begin{aligned}&\int_D|K_{\tau_i'}(x,x')-K_\tau(x,x')|\,dx'\\&\leq(1-\tau_i/\tau_i')\int_D K_{\tau_i'}(x,x')\,dx'\\&\quad+\int_D[K_\tau(x,x')-(\tau_i/\tau_i')K_{\tau_i'}(x,x')]\,dx'\\&\leq 2(1-\tau_i/\tau_i')=2|\tau_i'-\tau_i|/\tau_i',\end{aligned} \tag{15}$$

which is true for any $x\in D$ and also for $\tau_i'\geqq\tau_i$. Since $|h_\sigma(x)|\leq c$ for any $x\in D$ [see (6)], from the well-known definition of the norm of an operator and from (13) and (15), it follows at once that

$$\|T_{\tau_i'}-T_p\|\leq 2c|\tau_i'-\tau_i|/\tau_i',\ i=1,2,3. \tag{16}$$

Let us now deal with the parameters χ_i, $i=1,2,3$; by recalling (6) and (7), we write

$$|[T_{\chi_i'}-T_p)f](x)|\leq\int_D|h_{\chi_i'}(x')-h_\sigma(x')|K_\tau(x,x')\,dx'. \tag{17}$$

By putting

$$\Delta_\tau=2\left(\sum_{i=1}^{3}\tau_i^2\right)^{1/2}, \tag{18}$$

if $2<s<3$, from the definition (5) of $K_\tau(x,x')$ we get the following inequality:

$$\begin{aligned}\int_D|K_\tau(x,x')|^{s/2}\,dx'&\leq(\tau_1\tau_2\tau_3)^{(s/2)-1}(4\pi)^{-s/2}\int_{|x'|\leq\Delta_\tau}\frac{dx'}{|x'|^s}\\&=(\tau_1\tau_2\tau_3/4\pi)^{(s/2)-1}(\Delta_\tau)^{3-s}/(3-s),\end{aligned} \tag{19}$$

for any $x\in D$. By applying now the Holder's inequality to (17) and by taking into account (2), (6), and (19), we have

$$\begin{aligned}&|[(T_{\chi_i'}-T_p)f](x)|\\&\leq(\int_D|h_{\chi_i'}(x')-h_\sigma(x')|^r\,dx')^{1/r}(\int_D|K_\tau(x,x')|^{s/2}\,dx')^{2/s}\\&\leq c(8|\chi_i'-\chi_i|)^{1/r}(\tau_1\tau_2\tau_3/4\pi)^{[(s/2)-1]2/s}\\&\quad\times[(\Delta_\tau)^{3-s}/(3-s)]^{2/s},\end{aligned} \tag{20}$$

which is true for any $x\in D$ and any $f\in C(D)$, $\|f\|=1$, and where $1/r+2/s=1$. From (20) we deduce that

$$\|T_{\chi_i'}-T_p\|\leq c(2|\chi_i'-\chi_i|\,\tau_1\tau_2\tau_3/\pi)^{1/r}[(\Delta_\tau)^{3-s}/(3-s)]^{2/s},\quad i=1,2,3. \tag{21}$$

Finally, it is easily seen that

$$\|T_{c'}-T_p\|\leq|c'-c|,\quad\|T_{\gamma'}-T_p\|\leq|\gamma'-\gamma|. \tag{22}$$

In view of the inequalities (16), (21), and (22) it follows that T_p depends continuously on $p\in P$. We deal now with S_p. Let $f\in L_2(D)$, $\|f\|=1$; since $|h_\sigma(x)|\leq c$ for any $x\in D$, see (6), from the definition (8) we get

$$|[(S_{\tau_i'} - S_p)f](x)|^2$$
$$\leq c^2\left(\int_D |K_{\tau_i'}(x,x') - K_\tau)(x,x')|\,|f(x')|\,dx'\right)^2$$
$$\leq c^2 \int_D |K_{\tau_i'}(x,x') - K_\tau(x,x')|\,dx'$$
$$\times \int_D |K_{\tau_i'}(x,x') - K_\tau(x,x')|\,|f(x')|^2\,dx', \quad (23)$$

where we have used the Schwartz inequality. By recalling the inequality (15), if we integrate (23) over D and take into account (14), we have

$$\|(S_{\tau_i'} - S_p)f\|^2$$
$$\leq 2c^2(|\tau_i' - \tau_i|/\tau_i') \int_D dx' |f(x')|^2$$
$$\times \int_D [K_{\tau_i'}(x,x') + K_\tau(x,x')]\,dx$$
$$\leq 4c^2|\tau_i' - \tau_i|/\tau_i', \quad (24)$$

which is true for any $f \in L_2(D)$, $\|f\| = 1$. From the inequality (24), we get

$$\|S_{\tau_i'} - S_p\| \leq 2c(|\tau_i' - \tau_i|/\tau_i')^{1/2}, \quad i = 1,2,3. \quad (25)$$

By means of analogous calculations, we can deduce that

$$\|S_{\chi_i'} - S_p\|$$
$$\leq c(8|\chi_i' - \chi_i|)^{1/2r}(\tau_1\tau_2\tau_3/\pi)^{1/r}[(\Delta_\tau)^{3-s}/(3-s)]^{2/s},$$
$$i = 1,2,3, \quad (26)$$
$$\|S_{c'} - S_p\| \leq |c' - c|, \quad \|S_{\gamma'} - S_p\| \leq |\gamma' - \gamma|.$$

In view of the inequalities (25) and (26), it follows that S_p depends continuously on $p \in P$.

C. We now introduce the parameter

$$q = (\alpha, \beta, \gamma, c) \in Q, \quad \alpha = (\alpha_1, \alpha_2, \alpha_3), \quad \beta = (\beta_1, \beta_2, \beta_3),$$
$$Q = (0, +\infty)^3 \times [0, +\infty)^3 \times (0,1) \times [1, +\infty) \subset R^8. \quad (27)$$

q, as p, characterizes both the geometrical and the material properties of the system. By taking into account (1), (2), (3), and (27), it follows at once that $p \in P$ depends continuously on $q \in Q$: Therefore, Theorem 1 implies that also the operator valued functions T_q and S_q depend continuously on $q \in Q$. Moreover, as is well known, since, for any $q \in Q$, S_q is a positive definite, symmetric, and completely continuous linear operator of the Hilbert space $L_2(D)$ into itself, we have the inequality[9]

$$|\lambda_n(q') - \lambda_n(q)| \leq \|S_{q'} - S_q\|, \quad n \geq 1, \ q, q' \in Q, \quad (28)$$

and, hence, the eigenvalues of S_q (or T_q) are also continuous functions of $q \in Q$. We now prove

Theorem 2: The first eigenvalue $\lambda_1(q)$ is a strictly increasing function of each variable α_i, β_i, γ, and c, $i = 1,2,3$.

Proof: For any $q \in Q$ let $\psi_q \in L_2(D)$ be an eigenfunction of S_q corresponding to $\lambda_1(q)$ and normalized to unity; we know that necessarily $\psi_q > 0$ or $\psi_q < 0$ almost everywhere in D. We begin by considering the variables α_i $(i = 1,2,3)$; from the definition (8) of $S_p(x,x')$ and recalling also (1), (2), (3), and (27), we write

$$((S_{\alpha_i'} - S_q)\psi_q, \psi_q)$$
$$= \int_D\int_D k_{\chi_i'}(x,x')[K_{\tau_i'}(x,x') - K_\tau(x,x')]\,\psi_q(x)\,\psi_q(x')\,dx\,dx'$$
$$+ \int_D\int_D [k_{\chi_i'}(x,x') - k_q(x,x')]K_\tau(x,x')\,\psi_q(x)\,\psi_q(x')\,dx\,dx'. \quad (29)$$

If $\alpha_i' > \alpha_i$, from (1) and (2) we deduce that $\chi_i' > \chi_i$ and hence, from (2), (6), and (8), it follows that $k_{\chi_i'}(x,x') \geq k_q(x,x')$, $(x,x') \in D \times D$, so that the second term in (29) is surely > 0. Coming to the first term in (29), by applying the same procedure used to prove that S_p is positive definite, see (9), (10), and (11), we get

$$\int_D\int_D k_{\chi_i'}(x,x')[K_{\tau_i'}(x,x') - K_\tau(x,x')]\,\psi_q(x)\,\psi_q(x')\,dx\,dx'$$
$$= (2\pi)^{-3}\int_{R^3} d\omega \left(\frac{\tan^{-1}H_{\tau_i'}(\omega)}{H_{\tau_i'}(\omega)} - \frac{\tan^{-1}H_\tau(\omega)}{H_\tau(\omega)}\right)$$
$$\times \left|\int_D \sqrt{h_{\chi_i'}(x)}\exp(i\omega x)\,\psi_q(x)\,dx\right|^2. \quad (30)$$

Since from $\alpha_i' > \alpha_i$ it follows that $\tau_i' > \tau_i$, see (1), and hence that $H_{\tau_i'}(\omega) < H_\tau(\omega)$, $\omega \in R^3$, see the definition (10) of $H_\tau(\omega)$, by taking into account that $\tan^{-1}|\omega|/|\omega|$ is a strictly decreasing function of $|\omega|$, (30) assures us that also the first term in (29) is > 0. Finally, from the maximum property (12) of the first eigenvalue $\lambda_1(\alpha_i')$, we we deduce that

$$\lambda_1(\alpha_i') \geq (S_{\alpha_i'}\psi_q, \psi_q) > \lambda_1(q), \quad \alpha_i' > \alpha_i, \quad i = 1,2,3. \quad (31)$$

We consider now the variables β_i $(i = 1,2,3)$ and, in particular, the case $i = 1$: The same procedure is valid for $i = 2,3$. Let $\beta_1' > \beta_1$ and hence $\tau_1' > \tau_1$, see (1), and let $D_1' = [-\tau_1/\tau_1', \tau_1/\tau_1'] \times [-1,1] \times [-1,1] \subset D$. By putting

$$f(x) = \begin{cases} (\tau_1'/\tau_1)^{1/2}\psi_q(\tau_1'x_1/\tau_1, x_2, x_3), & x \in D_1', \\ 0, & x \in D \setminus D_1', \end{cases} \quad (32)$$

we get that $f \in L_2(D)$, $\|f\| = 1$, because $\psi_q \in L_2(D)$ and $\|\psi_q\| = 1$; moreover, it is easily seen that

$$(S_{\beta_1'}f, f) = (S_q\psi_q, \psi_q) = \lambda_1(q). \quad (33)$$

Now, let us note that $f \equiv 0$ in the subset $D \setminus D_1'$ which is of measure $\neq 0$, see (32); on the other hand, we know that $\psi_{\beta_1'}$ must be $\neq 0$ almost everywhere in D and that $\lambda_1(\beta_1')$ is simple. Therefore, (12) and (33) imply that necessarily $\lambda_1(\beta_1') > \lambda_1(q)$. Finally, by considering the variables γ and c, it is easily seen that

$$(S_{\gamma'}\psi_q, \psi_q) > (S_q\psi_q, \psi_q) = \lambda_1(q), \quad \gamma' > \gamma,$$
$$(S_{c'}\psi_q, \psi_q) > (S_q\psi_q, \psi_q) = \lambda_1(q), \quad c' > c, \quad (34)$$

and hence, from the maximum property (12), we deduce that $\lambda_1(\gamma') > \lambda_1(q)$ and that $\lambda_1(c') > \lambda_1(q)$.

3. PROPERTIES OF THE SOLUTION OF EQ. (4)

A. Let us now consider the original Eq. (4). The main results are summarized in

Theorem 3: Let $\mu = (\alpha, \beta, \gamma) \in M = (0, +\infty)^3 \times [0, +\infty)^3 \times (0,1)$; then:

(i) For any $\mu \in M$ there is one and only one critical value $c(\mu) > 1$ in the core and one and only one neutron flux $N_\mu \in C(D)$ in the system such that $N_\mu(x) > 0$ for any $x \in D$, $\| N_\mu \| = 1$.

(ii) The function $c(\mu)$ is continuous in M and strictly decreasing in each variable α_i, β_i, and γ $(i = 1, 2, 3)$; when $\beta_i = 0$ for any i, c is independent of γ.

(iii) If $\mu' \to \mu$, then $\| N_{\mu'} - N_\mu \| \to 0$ and hence, the neutron flux $N_\mu(x)$ is a continuous function in $M \times D$.

Proof: Let us consider the equation

$$\lambda_1(q)\, \phi_q(x) = \int_D h_q(x')\, K_\tau(x, x')\, \phi_q(x')\, dx', \tag{35}$$

where, for any $q \in Q$, $\phi_q \in C(D)$, $\| \phi_q \| = 1$; by choosing $\phi_q(x) > 0$ for any $x \in D$, then ϕ_q is uniquely determined. Since $| h_q(x) | \leq 1$ for any $x \in D$, if $c = 1$, see (6), by taking into account (5), from Eq. (35) we get

$$\lambda_1(\mu, 1) \leq \max_{x \in D} \left(\int_D K_\tau(x, x')\, dx' \right) \leq \int_{|x'| \leq \Delta_\tau} \frac{\exp(-|x'|)}{4\pi |x'|^2}\, dx' = 1 - \exp(-\Delta_\tau) < 1, \tag{36}$$

for any $\mu \in M$ and where Δ_τ is given by (18). On the other hand, by means of the function $f \in L_2(D)$ such that

$$f(x) = \begin{cases} (1/8\chi_1\chi_2\chi_3)^{1/2}, & x \in C, \\ 0, & x \in D \setminus C, \end{cases} \tag{37}$$

it is easily seen that $\lim_{c \to +\infty} (S_q f, f) = +\infty$, for any $\mu \in M$; therefore, since $\| f \| = 1$ [see (37)], (12) implies that also

$$\lim_{c \to +\infty} \lambda_1(\mu, c) = +\infty, \tag{38}$$

for any $\mu \in M$. Now, let us consider the equation

$$\lambda_1(\mu, c) = 1, \quad (\mu, c) \in Q = M \times [1, +\infty). \tag{39}$$

By recalling that $\lambda_1(q)$ is a continuous function of q and strictly increasing in each variable α_i, β_i, γ_i, and c $(i = 1, 2, 3)$, the points (i) and (ii) follow at once from the implicit function theorem and from (36) and (38), by putting

$$N_\mu = \phi_{(\mu, c(\mu))}, \quad \mu \in M. \tag{40}$$

Evidently, if $\beta = 0$, the results are that of a critical bare rectangular system. We now prove the point (iii). For this, it is sufficient to show that

$$\phi_{q'} \to \phi_q, \quad \text{if } q' \to q; \tag{41}$$

indeed, the result then follows from (40), by putting $q = (\mu, c(\mu))$, $q' = (\mu', c(\mu'))$, $c(\mu)$ being a continuous function of μ. In order to prove (41), let (q_n) be a sequence in Q converging to q. Since T_q is completely continuous, the sequence $(T_q \phi_{q_n})$ then contains a convergent subsequence $(T_q \phi_{\tilde{q}_n})$, that is,

$$T_q \phi_{\tilde{q}_n} \to \lambda_1(q)\, \tilde{\phi}_q, \quad \tilde{\phi}_q \in C(D). \tag{42}$$

By recalling that $\lambda_1(q)$ and T_q are continuous functions of q, from the inequality

$$\lambda_1(q) \| \phi_{\tilde{q}_n} - \tilde{\phi}_q \| \leq \| [\lambda_1(q)/\lambda_1(\tilde{q}_n)]\, T_{\tilde{q}_n} - T_q \| + \| T_q \phi_{\tilde{q}_n} - \lambda_1(q)\, \tilde{\phi}_q \|, \tag{43}$$

and, from (42), we get that $\phi_{\tilde{q}_n} \to \tilde{\phi}_q$. But $\lambda_1(q)$ is simple and, hence, (42) implies that also

$$\tilde{\phi}_q = \phi_q, \quad T_q \phi_{q_n} \to \lambda_1(q)\, \phi_q. \tag{44}$$

Since the sequence (q_n) is arbitrary, (44) assures us that

$$T_q \phi_{q'} \to \lambda_1(q)\, \phi_q, \quad \text{if } q' \to q. \tag{45}$$

Finally, from (45), (41) follows at once.

B. To finish, we now give three properties of $c(\mu)$ which have an obvious physical meaning. They are:

(i) $\lim_{\Delta_\alpha \to +\infty} c(\mu) = 1$,

(ii) $\lim_{\Delta_\alpha \to 0} c(\mu) = +\infty$,

(iii) $\lim_{\beta_1, \beta_2, \beta_3 \to +\infty} c(\mu) > 1$,

where Δ_α is defined as Δ_τ, see (18), (i) and (ii) being true for any β and γ, (iii) for any α and γ. In order to prove (i), let $\beta = 0$; then, by recalling (1), (2), and (6), from Theorem 3 and Eq. (35), it follows that

$$1 < c(\mu) \leq c(\alpha, 0, \gamma) = 1/\lambda_1(\alpha, 0, \gamma, 1), \tag{46}$$

for any $\mu \in M$. Therefore, it is sufficient to show that

$$\lim_{\Delta_\alpha \to +\infty} \lambda_1(\alpha, 0, \gamma, 1) = 1, \quad \text{for any } \gamma. \tag{47}$$

For this, from the inequality (36) and the maximum property (12), we get

$$(S_q f, f) \leq \lambda_1(q) < 1, \quad q = (\alpha, 0, \gamma, 1), \tag{48}$$

where $f(x) = \sqrt{1/8}$, $x \in D$, $\| f \| = 1$. Now, by writing explicitly $(S_q f, f)$, it is easily seen that $\lim_{\Delta_\alpha \to +\infty} (S_q f, f) = 1$, for any γ; hence, (48) assures us that (47) is true. In order to prove (ii) and (iii), by taking into account Eq. (4) and definitions (6) and (40), it follows that the neutron flux $N_\mu(x)$ must satisfy the equation

$$N_\mu(x) = [c(\mu) - \gamma] \int_C K_\tau(x, x')\, N_\mu(x')\, dx' + \gamma \int_D K_\tau(x, x')\, N_\mu(x')\, dx', \tag{49}$$

$\mu \in M$. By appling to Eq. (49) the same procedure as used to prove the inequality (36), we get the following inequality:

$$1 \leq c(\mu)[1 - \exp(-\Delta_\alpha)] + \gamma[\exp(-\Delta_\alpha) - \exp(-\Delta_\tau)]. \tag{50}$$

From this, (ii) and (iii) are easily proved.

*Work performed under contract C.N.R.

[1]G. Busoni, G. Frosali, and L. Mangiarotti, J. Math. Anal. Appl. (to be published).
[2]See, for example, B. Davison, *Neutron Transport Theory* (Clarendon, Oxford, 1958), Part II, Chap. 4.
[3]S. Mikhlin, *Mathematical Physics, An Advanced Course* (North-Holland, Amsterdam, 1970), p. 158.
[4]Reference 3, pp. 158, 162.
[5]Reference 3, pp. 160, 162.
[6]S. Mikhlin, *Integral Equations* (Pergamon, Oxford, 1957), p. 61.
[7]Reference 3, p. 164.
[8]F. Riesz and B. Sz-Nagy, *Functional Analysis* (Ungar, New York, 1965), pp. 233—34.
[9]Reference 8, pp. 238—39.

Hamiltonian formulation of the classical two-charge problem in straight-line approximation

Frederick James Kennedy*

Theoretical Physics Institute, University of Alberta, Edmonton, Alberta, Canada
(Received 24 July 1974)

The Newtonian equations of motion expressing the interaction of electric charges correct up to terms of order e^2, the product of the charges, is cast into Hamiltonian form. Lorentz transformations are canonically represented to order e^2, but as anticipated by the zero-interaction theorem, there is no canonical transformation from the canonical variables to the charges' physical positions.

I. INTRODUCTION

The equations

$$m_1\dot{u}_1 = e^2(1 - v_2^2/c^2)\frac{\hat{r} - c^{-2}v_1\times(\hat{r}\times v_2)}{r^2[1 - c^{-2}(\hat{r}\times v_2)^2]^{3/2}}, \tag{1a}$$

$$m_2\dot{u}_2 = -e^2(1 - v_1^2/c^2)\frac{\hat{r} - c^{-2}v_2\times(\hat{r}\times v_1)}{r^2[1 - c^{-2}(\hat{r}\times v_1)^2]^{3/2}}, \tag{1b}$$

where the m's are rest masses, r the separation of the particles, and $\dot{u}_i$, the time derivative of $(1 - v_i^2/c^2)^{-1/2}v_i$, are the equations of motion for a pair of interacting electric charges correctly written to terms of order e^2, the product of the charges. The approximation is sometimes called the "straight line" approximation because the force on each charge would be the exact result if the other charge were constrained to move uniformly.[1-3]

Since many-body forces must always be of second order in the coupling constant, the forces in (1) also correctly describe, to first order in e^2, the interaction of a particular pair of charges within a system of N charges.[4] Thus, for example, (1) can be used to provide relativistic corrections to the mutual Coulomb and Darwin forces [the first two terms in the expansion in c^{-2} of the forces in (1)[5]] of the interelectron interactions in the classical atom. To pass from such a Newtonian description to a quantum mechanical correction of atomic energy levels—to introduce the improved Coulomb forces in (1) into any quantum mechanical calculation—is another matter altogether. Whatever snares await in the maze leading to quantization, however, the unavoidable first step must be the rendition of the classical dynamics (1), consistently to order e^2, in acceptable Hamiltonian form; that is, a form in which the transformations of the inhomogeneous Lorentz group are canonical to order e^2.[6] That first step will be taken in this paper.

The Hamiltonian formulation of (1) cannot be obtained by any textbook recipe, nor is there any property of the forces that suggests a successful *anzatz*. Indeed, as will be seen, and as foreseen by the zero-interaction theorem,[7] the forces in (1) are sufficiently relativistic that the existence of an acceptable Hamiltonian of the form $H(r_i, p_i)$ is out of the question. Still, with the proper definition of the canonical position variables, an Hamiltonian formulation always exists for an even-order system such as (1),[8] and a method exists for the determination of that formulation.

In what follows, after some necessary background, the method is explained and applied to Eqs. (1). Fortunately, the only substantial problem which appears, the determination of a Lagrangian which produces (1) in the form of twelve first-order differential equations, is disposed of gratuitously by certain prior results from generalized mechanics. In effect, the solution consists in the recognition and rearrangement of quantities already available. A transformation of the position variables in this Lagrangian then produces the required Hamiltonian formulation of (1). The result is tested, then extended to N particles. Canonical transformations, and expansion of the Hamiltonian and canonical variables in inverse powers of c^2 are the subject of the concluding discussion.

II. BACKGROUND

The equations of motion (1) may be either induced directly from Coulomb's law or deduced from the full apparatus of classical electrodynamics.[9] The second approach begins with a separate Lagrangian for the motion of each charge in the field of the other. The motion of the first charge, for example, is implied by

$$L_1(r_1, v_1) = -m_1(1 - v_1^2)^{1/2} - e^2\sum_{p=0}^{\infty}\frac{(-D_2)^p}{p!}(1 - v_1\cdot v_2)r^{p-1}, \tag{2}$$

where the potential is just the Lienard–Wiechert potential originating at the second charge and expanded about the present time t, and D_2 denotes a time differentiation which acts on the variables of the second charge only. The speed of light has been set equal to unity. Exchange of indices produces the companion L_2. When the exact force is derived from L_1 its terms fall naturally into two classes according to whether or not they contain accelerations and higher derivatives of the second charge's variables. Terms which do are not less than e^4 order; those which do not are of order e^2 and their sum is the force in (1a).[10] Since the force of radiation damping, not contained in (2), belongs in any case with the higher order terms, the Newtonian-type forces in (1) constitute a consistent relativistic correction to the Coulomb interaction of the two charges.

Only derivatives of v_2 are set to zero in obtaining the approximation to the exact equation of motion of the first charge. The approximation may therefore be made directly in the Lagrangian; that is,

$$\Big[\text{straight line limit}\left(D\frac{\partial}{\partial \mathbf{v}_1}-\frac{\partial}{\partial \mathbf{r}_1}\right)-\left(D\frac{\partial}{\partial \mathbf{v}_1}-\frac{\partial}{\partial \mathbf{r}_1}\right)$$
$$\text{straight line limit}\Big]L_1=0, \tag{3}$$

where the limit operation consists in replacing D_2 in the operand of the limit with $\mathbf{v}_2\cdot\partial/\partial\mathbf{r}_2$. Now let a new Lagrangian L_1' be prepared equal to L_1 but with odd p terms deleted. Since these odd p terms vanish when D_2 becomes $\mathbf{v}_2\cdot\partial/\partial\mathbf{r}_2$, the difference between these otherwise distinct Lagrangians, and between their implicit equations of motion, vanishes in the straight line limit. The new Lagrangian, however, can be symmetrized, for $D_2^{2p}=D_2^p(D-D_1)^p=(-D_1D_2)^p$ plus a disposable total time derivative D. Consequently, the private Lagrangians L_1' and the corresponding one for the other particle L_2' may be replaced jointly by the single Lagrangian

$$L=-m_1(1-v_1^2)^{1/2}-m_2(1-v_2^2)^{1/2}$$
$$-e^2\sum_{p=0}^{\infty}\frac{(-D_1D_2)^p}{(2p)!}(1-\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-1} \tag{4}$$

and the equations of motion (1) looked upon as the straight line approximation to an electrodynamics with an authentic action-at-a-distance Lagrangian, hence canonical, formulation. In fact, the electrodynamics implied by (4) is just that of the half advanced, half retarded interaction whose e^2-order congruence with the fully retarded interaction is well known.[11] The equations of motion are the Euler–Lagrange equations of "generalized mechanics," or Ostrogradski equations,

$$\sum_{l=0}^{\infty}(-D)^l\frac{\partial L}{\partial(D^l\mathbf{r}_i)}=0,\quad i=1,2, \tag{5}$$

and the system (1) emerges just as originally, by setting to zero accelerations and higher derivatives which occur in the forces of (5).

The usual advantage associated with the Lagrangian formulation of mechanics, the effortless construction of constants of the motion, is also present when the Lagrangian depends on higher time derivatives. For instance, if time and space translational, and space rotational, invariance are canonically represented with respect to a Lagrangian depending on derivatives through $D^n\mathbf{r}_1$ and $D^n\mathbf{r}_2$, then the corresponding energy, total linear and angular momentum are respectively[12]

$$E=\sum_{i=1}^{2}\sum_{l=0}^{n-1}\mathbf{P}_i^{(l)}\cdot(D^l\mathbf{v}_i)-L(\mathbf{r}_1,\mathbf{r}_2,\mathbf{v}_1,\mathbf{v}_2,\ldots,D^n\mathbf{r}_1,D^n\mathbf{r}_2), \tag{6a}$$

$$\mathbf{P}=\mathbf{P}_1^{(0)}+\mathbf{P}_2^{(0)}, \tag{6b}$$

$$\mathbf{L}=\sum_{i=1}^{2}\sum_{l=0}^{n-1}(D^l\mathbf{r}_i)\times\mathbf{P}_i^{(l)}, \tag{6c}$$

where the n "Ostrogradski momenta" are defined as

$$\mathbf{P}_i^{(l)}=\sum_{s=l}^{n-1}(-D)^{s-l}\frac{\partial L(\mathbf{r}_j,\ldots,D^n\mathbf{r}_j)}{\partial(D^s\mathbf{v}_i)},$$
$$l=0,1,\ldots,n-1,\quad i=1,2 \tag{7}$$

If n is extended to infinity and the joint Lagrangian (4) together with its partial derivatives

$$\frac{\partial L}{\partial(D^s\mathbf{v}_1)}=\frac{m_1\mathbf{v}_1}{(1-v_1^2)^{1/2}}\delta_{so}-e^2\sum_{p=s}^{\infty}\frac{(-D_2)^{p+1}D_1^{p-s}}{(2p+2)!}(2p+1)\binom{p+1}{s+1}$$
$$\times(1-\mathbf{v}_1\cdot\mathbf{v}_2)r^{2p-1}\mathbf{r}$$
$$+e^2\sum_{p=s}^{\infty}\frac{(-D_2)^pD_1^{p-s}}{(2p)!}\binom{p}{s}\mathbf{v}_2r^{2p-1}, \tag{8}$$

where δ_{so} is the Kronecker δ and $\binom{q}{s}$ a binomial coefficient, for $i=1$, for example, are substituted in these expressions, then the quantities in (6) are integrals of the dynamics inhering in (4). Moreover, the components of these quantities not depending on accelerations or higher derivatives are integrals of the system (1) consistent to order e^2.[13]

The joint Lagrangian (4) is thus useful for studying the equations of motion (1), though it is a Lagrangian not for those equations but for a much more complex system. The Hamiltonian formulation aimed at in this work, on the other hand, is a function of four independent vectors only and *properly* contains the equations of motion (1) to terms of order e^2. As it turns out, the problem of determining such a Hamiltonian is effectively solved with the introduction of the joint Lagrangian (4). The solution is not a simple matter of dropping accelerations and higher derivatives in (4) and then forming a "Hamiltonian" by Legendre transformation of the remainder, for (4) does not admit the latitude expressed by the relation (3): The straight line limit of L is *not* a Lagrangian $L(\mathbf{r},\mathbf{v}_1,\mathbf{v}_2)$ for the straight line limit of the equations of motion (5) implied by L. Nevertheless, a kind of true Lagrangian for the system (1) alone is embedded in (4), and the first step which must be taken before that system of equations can be Hamiltonized is to extract it.

III. THE LAGRANGIAN

Kerner has elaborated a method for casting a system of differential equations such as (1) into Hamiltonian form.[14] The first step is to find a Lagrangian linear in the independent "velocities" $\dot{\mathbf{r}}_i$ and $\dot{\mathbf{v}}_i$ which will imply the differential equations in the first-order form

$$\sum_{j=1}^{12}A_{ij}(q_l)\dot{q}_j=B_i(q_l),\quad i=1\cdots12,$$

where the twelve "coordinates" q_l are the components of $\mathbf{r}_1$ and $\mathbf{r}_2$, $\mathbf{v}_1$ and $\mathbf{v}_2$. In this regime $\mathbf{r}_i=\mathbf{v}_i$ is not an identity, so $\dot{\mathbf{r}}_i$ and $\mathbf{v}_i$ must not be confused. The general form of *any* Lagrangian for a first-order system of twelve equations must be

$$L'(\mathbf{r}_i,\mathbf{v}_i,\dot{\mathbf{r}}_i,\dot{\mathbf{v}}_i)=\sum_{i=1}^{2}(\boldsymbol{\alpha}_i\cdot\dot{\mathbf{r}}_i+\boldsymbol{\beta}_i\cdot\dot{\mathbf{v}}_i)-h \tag{9}$$

where $\boldsymbol{\alpha}_i$, $\boldsymbol{\beta}_i$ and h depend only on the r's and v's. It is easy to show that h has to be constant and must be the total energy; the total linear momentum of the system is $\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2$, and the total angular momentum is $\mathbf{r}_1\times\boldsymbol{\alpha}_1+\mathbf{r}_2\times\boldsymbol{\alpha}_2+\mathbf{v}_1\times\boldsymbol{\beta}_1+\mathbf{v}_2\times\boldsymbol{\beta}_2$.

Now these blanket requirements bear an obvious resemblance to the generalized mechanics results quoted in (6). Moreover, the Lagrangian from (6a)

$$L(\mathbf{r}_i, \mathbf{v}_i, \ldots, D^n\mathbf{r}_i) = \sum_{i=1}^{2}\sum_{l=0}^{n-1} \mathbf{P}_i^{(l)} \cdot (D^l\mathbf{v}_i) - E$$

where $\mathbf{v}_i$, conventionally, is $\dot{\mathbf{r}}_i$, and the P's and E depend on all the various derivatives of $\mathbf{r}_i$, may equally well be taken as

$$L'(\mathbf{r}_i, \mathbf{v}_i, \mathbf{a}_i, \ldots, \dot{\mathbf{r}}_i, \dot{\mathbf{v}}_i, \dot{\mathbf{a}}_i, \ldots) = \sum_{i=1}^{2}(\mathbf{P}_i^{(0)} \cdot \dot{\mathbf{r}}_i + \mathbf{P}_i^{(1)} \cdot \dot{\mathbf{v}}_i + \mathbf{P}_i^{(2)} \cdot \dot{\mathbf{a}}_i + \cdots) - E$$

where $\mathbf{r}_i, \mathbf{v}_i, \mathbf{a}_i, \ldots, D^{2n-1}\mathbf{r}_i$ are reckoned independent "coordinates," and the quantities with dots their "velocities." Half these velocities appear linearly while the P's and E depend only on the coordinates. The sets of equations

$$\sum_{l=0}^{n}(-D)^l \frac{\partial L}{\partial (D^l \mathbf{r}_i)} = 0, \quad i = 1, 2,$$

and

$$D\frac{\partial L'}{\partial \dot{\mathbf{q}}_i^{(l)}} - \frac{L'}{\partial \mathbf{q}_i^{(l)}} = 0, \quad \mathbf{q}_i^{(0)} = \mathbf{r}_i, \qquad \mathbf{q}_i^{(1)} = \mathbf{v}_i, \ldots,$$
$$i = 1, 2,$$

are then alternate descriptions of the same dynamics. The equivalence of these descriptions is proved for a certain broad class of Lagrangians $L(\mathbf{r}_i, \ldots, D^n\mathbf{r}_i)$ in Appendix A.

Without getting into the question of the soundness of such a dual view in the case in which n extends to infinity, it is certainly true, formally at least, that a structure matching (9), required by the Lagrangian for the first-order rewrite of the straight line approximation to the second-order equations of motion implied by (4), stands out quite clearly in the alternative, linearized, expression of (4). It is L' with all coordinates and velocities beyond $\mathbf{r}_i$, $\mathbf{v}_i$ and $\dot{\mathbf{r}}_i$, $\dot{\mathbf{v}}_i$ set to zero:

$$\begin{aligned} &L'(\mathbf{r}_i, \mathbf{v}_i, 0, \ldots, \dot{\mathbf{r}}_i, \dot{\mathbf{v}}_i, 0, \ldots) \\ &= \sum_{i=1}^{2}[\mathbf{P}_i^{(0)}(\mathbf{r}_j, \mathbf{v}_j, 0, \ldots) \cdot \dot{\mathbf{r}}_i \\ &\quad + \mathbf{P}_i^{(1)}(\mathbf{r}_j, \mathbf{v}_j, 0, \ldots) \cdot \dot{\mathbf{v}}_i] - E(\mathbf{r}_i, \mathbf{v}_i, 0, \ldots) \end{aligned} \tag{10}$$

where $f(\mathbf{r}_j, \mathbf{v}_j, 0, \ldots)$ is of course the straight line limit of the quantity $f(\mathbf{r}_j, \mathbf{v}_j, \mathbf{a}_j, \ldots)$. In the following the straight line limit will be understood when $f(\mathbf{r}_j, \mathbf{v}_j, \mathbf{a}_j, \ldots)$ is written as $f(\mathbf{r}_j, \mathbf{v}_j)$. The energy in (10) is given by

$$E(\mathbf{r}_j, \mathbf{v}_j) = \sum_{i=1}^{2}\mathbf{P}_i^{(0)}(\mathbf{r}_j, \mathbf{v}_j) \cdot \mathbf{v}_i - L(\mathbf{r}_j, \mathbf{v}_j) \tag{11}$$

and the $\mathbf{P}_i^{(0)}$ and $\mathbf{P}_i^{(1)}$ are determined from (7) with n extended to infinity. When L is substituted from (4), and its partial derivatives from (8), the Lagrangian in (10) is explicitly written. It will be confirmed later through the Hamiltonian that this proposal does work, that the $L'(\mathbf{r}_i, \mathbf{v}_i)$ which results gives back twelve first-order differential equations which reexpress the equations of motion (1). What remains in this section is to obtain the straight line limit of the first two Ostrogradski momenta and the energy. Since the calculation concerns the *form* of these quantities, and not their dynamical content, it is not yet necessary to implement the distinction between $\dot{\mathbf{r}}_i$ and $\mathbf{v}_i$, $\dot{\mathbf{v}}_i$ and $\mathbf{a}_i$, and so on which was mentioned above.

The Ostrogradski momentum $\mathbf{P}_i^{(0)}$ with $i = 1$ chosen for illustration will be studied first. When (8), the derivative of (4) with respect to $D^s\mathbf{v}_i$, is substituted in (7) and n is set to infinity, the exact $\mathbf{P}_i^{(0)}$ appears as a complicated, double infinite series. Some manipulation reduces it to

$$\begin{aligned} \mathbf{P}_i^{(0)} &= m_1\mathbf{u}_1 + e^2\sum_{p=0}^{\infty}\frac{D_2^{2p}}{(2p)!}\mathbf{v}_2\gamma^{2p-1} \\ &\quad + e^2\frac{1}{D}\sum_{p=0}^{\infty}\frac{D_2^{2p}}{(2p)!}(2p-1)(1-\mathbf{v}_1\cdot\mathbf{v}_2)\gamma^{2p-3}\mathbf{r} \\ &\quad - e^2\frac{1}{D}\sum_{p=0}^{\infty}\frac{(-D_1D_2)^p}{(2p)!}(2p-1)(1-\mathbf{v}_1\cdot\mathbf{v}_2)\gamma^{2p-3}\mathbf{r}. \end{aligned} \tag{12}$$

The inverse D's, which stand for time integrations, are the cost of eliminating $\sum$'s in the last two interaction pieces.[15] Since it is the straight line limit $\mathbf{P}_1^{(0)}(\mathbf{r}, \mathbf{v}_1, \mathbf{v}_2)$ of (12) which is required, velocities may be treated as constants and placed to the left or right of D operators as convenient. The third interaction piece of (12) may therefore be written

$$-e^2(1-\mathbf{v}_1\cdot\mathbf{v}_2)\int\frac{\partial}{\partial\mathbf{r}}\left(\sum_{p=0}^{\infty}\frac{(-D_1D_2)^p}{(2p)!}\gamma^{2p-1}\right)dt. \tag{13}$$

The result derived in Appendix B that

$$\begin{aligned} &\sum_{p=0}^{\infty}\frac{1}{(2p)!}\left(-\mathbf{v}_1\cdot\frac{\partial}{\partial\mathbf{r}_1}\right)^p\left(\mathbf{v}_2\cdot\frac{\partial}{\partial\mathbf{r}_2}\right)^p\gamma^{2p-1} \\ &= \frac{1}{[1-\mu_1\cdot\mu_2-\frac{1}{4}(\mu_1\times\mu_2)^2]^{1/2}\gamma} \end{aligned} \tag{14}$$

where $\mu_i = \hat{\mathbf{r}}\times\mathbf{v}_i$, shows that (13) may then be replaced with

$$-e^2(1-\mathbf{v}_1\cdot\mathbf{v}_2)\int\frac{\partial}{\partial\mathbf{r}}\left(\frac{1}{[1-\mu_1\cdot\mu_2-\frac{1}{4}(\mu_1\times\mu_2)^2]^{1/2}\gamma}\right)dt. \tag{15}$$

It can also be seen from comparing the third interaction term in (12) with the interaction term $L^{(\mathrm{int})}$ of the Lagrangian (4) that the former is

$$\int\frac{\partial L^{(\mathrm{int})}}{\partial\mathbf{r}}dt, \tag{16}$$

while, from (14), the straight line limit of $L^{(\mathrm{int})}$ is

$$L^{(\mathrm{int})}(\mathbf{r}, \mathbf{v}_1, \mathbf{v}_2) = -e^2\frac{(1-\mathbf{v}_1\cdot\mathbf{v}_2)}{[1-\mu_1\cdot\mu_2-\frac{1}{4}(\mu_1\times\mu_2)^2]^{1/2}\gamma}. \tag{17}$$

If D_i is approximated by $\mathbf{v}_i\cdot\partial/\partial\mathbf{r}_i$, then $D_1f(r)$ becomes $-D_2f(r)$ upon changing $\mathbf{v}_1$ into $\mathbf{v}_2$. Consequently, from (14),

$$\sum_{p=0}^{\infty}\frac{1}{(2p)!}\left(\mathbf{v}_2\cdot\frac{\partial}{\partial\mathbf{r}_2}\right)^{2p}\gamma^{2p-1} = \frac{1}{(1-\mu_2^2)^{1/2}\gamma} \tag{18}$$

with the inclusion of the factor $e^2\mathbf{v}_2$, becomes the straight line approximation to the first interaction term in (12). Finally, comparison of the second and third interaction terms in (12) shows that the straight line approximation to the former is

$$e^2(1-\mathbf{v}_1\cdot\mathbf{v}_2)\int\frac{\partial}{\partial\mathbf{r}}\left(\frac{1}{(1-\mu_2^2)^{1/2}\gamma}\right)dt. \tag{19}$$

There are any number of ways of performing the integrations in (15) and (19) each of which may lead to a different conclusion. This is an inessential ambiguity, however, because, just as the integrations have been introduced (to buy a little simplicity in writting $P_1^{(0)}$), so they can be eliminated. A time derivative D can be factored out of the last two interaction pieces of (12) and "cancelled" with the denominator D to recover the original double series representation of $P_1^{(0)}$, the touchstone for the admissibility of any proposed alternative representation. The test is as follows. The integrations in (15) and (19) are carried out; the two results are added, the missing $1/c^2$'s are restored (by dimensional analysis), and the whole is expanded in powers of $1/c^2$. Each coefficient in this expansion must then be identical with the straight line component of the term of corresponding order (labelled by p), in the sum of the last two interaction pieces of (12). It may be mentioned here that the straight line approximation to the companion $P_1^{(1)}$, which will be found below, is also checked in this way; its unambiguous series representation is derived just as for $P_1^{(0)}$ except, of course, l must be 1 in (7). In this way the authenticity of the final $P_1^{(0)}(\mathbf{r}, \mathbf{v}_1, \mathbf{v}_2)$ and $P_1^{(1)}(\mathbf{r}, \mathbf{v}_1, \mathbf{v}_2)$, which will appear shortly, has been established and the validity of the following rules and results confirmed.

The same rule applies to both (15) and (19). Replace $\mathbf{r}$ in the square brackets with $\mathbf{r}+\mathbf{v}t$, where $\mathbf{v}=\mathbf{v}_1-\mathbf{v}_2$. Then, in this sequence: integrate, take the gradient, and set t equal to zero. If A stands for the square of the bracketed denominator in either (15) or (19), then, after the first step, the integrals of both expressions have the form

$$\int \frac{\partial}{\partial \mathbf{r}}\left(\frac{1}{(A+Bt+Ft^2)^{1/2}}\right)dt. \tag{20}$$

Though this may not be obvious in the case of (15) where it might seem that a biquadratic in t should appear, it is true, nonetheless, owing to the fact that $r^2(\mu_1\times\mu_2)^2$ is invariant under $\mathbf{r}\to\mathbf{r}+\mathbf{v}t$. When the remaining sequential steps are applied to (20), there emerges the formula

$$\int \frac{\partial}{\partial \mathbf{r}}\left(\frac{1}{A^{1/2}}\right)dt=\frac{1}{(4AF-B^2)A^{1/2}}\left(2A\frac{\partial B}{\partial \mathbf{r}}-B\frac{\partial A}{\partial \mathbf{r}}\right), \tag{21a}$$

or, equivalently,

$$\int \frac{\partial}{\partial \mathbf{r}}\left(\frac{1}{A^{1/2}}\right)dt=\frac{\partial}{\partial \mathbf{r}}\left[\frac{1}{F^{1/2}}\ln\left(\frac{2(AF)^{1/2}+B}{(4AF-B^2)^{1/2}}\right)\right]. \tag{21b}$$

When $\mathbf{v}$ is dotted through the latter expression, one has

$$\frac{1}{A^{1/2}}=\mathbf{v}\cdot\frac{\partial}{\partial \mathbf{r}}\left[\frac{1}{F^{1/2}}\ln\left(\frac{2(AF)^{1/2}+B}{(4AF-B^2)^{1/2}}\right)\right]; \tag{22}$$

and from either of the last two expressions

$$\int \frac{1}{A^{1/2}}dt=\frac{1}{F^{1/2}}\ln\left(\frac{2(AF)^{1/2}+B}{(4AF-B^2)^{1/2}}\right) \tag{23}$$

which will give the correct straight line approximation to the integral of the square brackets of either (15) or (19).

The above formulas are perfected by substitution of the particular A, B, F generated by the transformation $\mathbf{r}\to\mathbf{r}+\mathbf{v}t$ of A in either (15) or (19). These are

$$A=A_s=[1-\mu_1\cdot\mu_2-\tfrac{1}{4}(\mu_1\times\mu_2)^2]r^2, \tag{24a}$$

$$B=B_s=2r[\hat{\mathbf{r}}\cdot\mathbf{v}-\tfrac{1}{2}(\mu_1+\mu_2)\cdot(\mathbf{v}_1\times\mathbf{v}_2)], \tag{24b}$$

$$F=v^2-(\mathbf{v}_1\times\mathbf{v}_2)^2 \tag{24c}$$

in the case of (15); and

$$A=A_2=(1-\mu_2^2)r^2 \tag{25a}$$

$$B=B_2=2r[\hat{\mathbf{r}}\cdot\mathbf{v}-\mu_2\cdot(\mathbf{v}_1\times\mathbf{v}_2)], \tag{25b}$$

$$F=v^2-(\mathbf{v}_1\times\mathbf{v}_2)^2 \tag{25c}$$

in the case of (19). The discriminant formed from (24) is

$$4A_sF-B_s^2=4r^2[1-\mathbf{v}_1\cdot\mathbf{v}_2-\tfrac{1}{4}(\mathbf{v}_1\times\mathbf{v}_2)^2][\mu^2-(\mu_1\times\mu_2)^2], \tag{26a}$$

and from (25) is

$$4A_2F-B_2^2=4r^2(1-v_2^2)[\mu^2-(\mu_1\times\mu_2)^2], \tag{26b}$$

where $\mu=\hat{\mathbf{r}}\times\mathbf{v}=\mu_1-\mu_2$.

The straight line approximation to $P_1^{(0)}$ may now be obtained by adding to $m_1\mathbf{u}_1$ the following pieces: (18) multiplied by $e^2\mathbf{v}_2$; either version of (21) multiplied by $e^2(1-\mathbf{v}_1\cdot\mathbf{v}_2)$, with A, B, F given by (25); and either version of (21) multiplied by $-e^2(1-\mathbf{v}_1\cdot\mathbf{v}_2)$ with A, B, F given by (24). The net result is

$$\begin{aligned}\mathbf{P}_1^{(0)}(\mathbf{r},\mathbf{v}_1,\mathbf{v}_2)=&\,m_1\frac{\mathbf{v}_1}{(1-v_1^2)^{1/2}}+e^2\mathbf{v}_2\frac{1}{A_2^{1/2}}\\&+e^2(1-\mathbf{v}_1\cdot\mathbf{v}_2)\frac{\partial}{\partial \mathbf{r}}\left[\frac{1}{F^{1/2}}\ln\left(\frac{2(A_2F)^{1/2}+B_2}{(4A_2F-B_2^2)^{1/2}}\right)\right]\\&-e^2(1-\mathbf{v}_1\cdot\mathbf{v}_2)\frac{\partial}{\partial \mathbf{r}}\left[\frac{1}{F^{1/2}}\ln\left(\frac{2(A_sF)^{1/2}+B_s}{(4A_sF-B_s^2)^{1/2}}\right)\right],\end{aligned} \tag{27}$$

or, substituting from (24), (25), and (26),

$$\begin{aligned}\mathbf{P}_1^{(0)}(\mathbf{r},\mathbf{v}_1,\mathbf{v}_2)=&\,m_1\frac{\mathbf{v}_1}{(1-v_1^2)^{1/2}}+e^2\mathbf{v}_2\frac{1}{r(1-\mu_2^2)^{1/2}}\\&+e^2\frac{1-\mathbf{v}_1\cdot\mathbf{v}_2}{[v^2-(\mathbf{v}_1\times\mathbf{v}_2)^2]^{1/2}}\\&\times\frac{\partial}{\partial \mathbf{r}}\ln\left(\frac{(1-\mu_2^2)^{1/2}[v^2-(\mathbf{v}_1\times\mathbf{v}_2)^2]^{1/2}+\hat{\mathbf{r}}\cdot\mathbf{v}-\mu_2\cdot(\mathbf{v}_1\times\mathbf{v}_2)}{[1-\mu_1\cdot\mu_2-\frac{1}{4}(\mu_1\times\mu_2)^2]^{1/2}[v^2-(\mathbf{v}_1\times\mathbf{v}_2)^2]^{1/2}+\hat{\mathbf{r}}\cdot\mathbf{v}-\frac{1}{2}(\mu_1+\mu_2)\cdot(\mathbf{v}_1\times\mathbf{v}_2)}\right).\end{aligned} \tag{28}$$

Both (28) and $P_2^{(0)}(\mathbf{r},\mathbf{v}_2,\mathbf{v}_1)$ may be summarized with the single expression

$$P_i^{(0)}(r_j, v_j) = m_i u_i + U_i + \frac{\partial}{\partial r_i}\Omega, \quad i = 1, 2. \tag{29}$$

Thus

$$U_1 = e^2 v_2 \frac{1}{A_2^{1/2}} + e^2(1 - v_1 \cdot v_2)\frac{\partial}{\partial r}\left[\frac{1}{F^{1/2}} \ln\left(\frac{2(A_2F)^{1/2} + B_2}{(4A_2F - B_2^2)^{1/2}}\right)\right] \tag{30}$$

with U_2 gotten by exchange of indices; and

$$\Omega = -e^2 \frac{(1 - v_1 \cdot v_2)}{F^{1/2}} \ln\left(\frac{2(A_sF)^{1/2} + B_s}{(4A_sF - B_s^2)^{1/2}}\right). \tag{31}$$

It is apparent from (17), (22), and (24) that

$$L^{(\mathrm{int})}(r, v_1, v_2) = v \cdot \frac{\partial \Omega}{\partial r}. \tag{32}$$

The U's, defined by (30), have a property which must be recorded. Take the straight line time derivative of (30):

$$v \cdot \frac{\partial U_1}{\partial r} = e^2 v_2 v \cdot \frac{\partial}{\partial r}\frac{1}{A_2^{1/2}} + e^2(1 - v_1 \cdot v_2)\frac{\partial}{\partial r} v \cdot \frac{\partial}{\partial r} \times \left[\frac{1}{F^{1/2}} \ln\left(\frac{2(A_2F)^{1/2} + B_2}{(4A_2F - B_2^2)^{1/2}}\right)\right].$$

The formula (22) shows this to be

$$v \cdot \frac{\partial U_1}{\partial r} = e^2[v_2 v + \mathbf{I}(1 - v_1 \cdot v_2)] \cdot \frac{\partial}{\partial r}\frac{1}{A_2^{1/2}}$$

$$= -e^2 \frac{1}{A_2^{3/2}}(1 - v_2^2)[\mathbf{I}(1 - v_1 \cdot v_2) + v_2 v_1] \cdot r$$

which is the negative of the force F_1 on the right-hand side of (1a). Therefore, to terms of order e^2

$$\sum_{j=1}^{2}\left(\frac{1}{m_j} F_j \cdot \frac{\partial}{\partial u_j} + v_j \cdot \frac{\partial}{\partial r_j}\right)(m_i u_i + U_i) = 0, \tag{33}$$

$i = 1, 2$

which is to say that the first two terms of $P_i^{(0)}(r_j, v_j)$ form a constant of the motion in straight line approximation. These are peculiar constants, however, because, in the case $i = 1$ for example, the constancy of $m_1 u_1 + U_1$ is independent of F_2; but more will be said about this later.

The comparable treatment of $P_1^{(1)}$ is simplified by beginning with an identity which follows directly from (7),

$$DP_i^{(1)} \equiv -P_i^{(0)} + \frac{\partial L}{\partial v_i}, \quad i = 1, 2 \tag{34}$$

When the exact $P_i^{(0)}$ is replaced by (29), and the Lagrangian (4) by its straight line limit,

$$-m_1(1 - v_1^2)^{1/2} - m_2(1 - v_2^2)^{1/2} + L^{(\mathrm{int})}(r, v_1, v_2),$$

then the identity (34) becomes

$$\left(v_1 \cdot \frac{\partial}{\partial r_1} + v_2 \cdot \frac{\partial}{\partial r_2}\right)P_i^{(1)}(r_j, v_j) = v \cdot \frac{\partial}{\partial r} P_i^{(1)}(r_j, v_j)$$

$$= -U_i - \frac{\partial \Omega}{\partial r_i} + \frac{\partial}{\partial v_i} L^{(\mathrm{int})}(r, v_1, v_2). \tag{35}$$

It is evident from (35) that there is no part of $P_i^{(1)}(r_j, v_j)$ without a factor e^2.

Now, from (32),

$$\frac{\partial}{\partial v_i} L^{(\mathrm{int})}(r, v_1, v_2) - \frac{\partial \Omega}{\partial r_i} = v \cdot \frac{\partial}{\partial r}\left(\frac{\partial \Omega}{\partial v_i}\right). \tag{36}$$

Next, multiplication of (22) by $e^2(1 - v_1 \cdot v_2)$, with A, B, F as in (25), gives

$$e^2 \frac{(1 - v_1 \cdot v_2)}{A_2^{1/2}} = v \cdot \frac{\partial}{\partial r}\left[e^2 \frac{(1 - v_1 \cdot v_2)}{F^{1/2}} \ln\left(\frac{2(A_2F)^{1/2} + B_2}{(4A_2F - B_2^2)^{1/2}}\right)\right]$$

partial differentation of which with respect to v_1 gives

$$-e^2 \frac{v_2}{A_2^{1/2}} = \left(\frac{\partial}{\partial r} + v \cdot \frac{\partial}{\partial r}\frac{\partial}{\partial v_1}\right)\left[e^2 \frac{(1 - v_1 \cdot v_2)}{F^{1/2}} \times \ln\left(\frac{2(A_2F)^{1/2} + B_2}{(4A_2F - B_2^2)^{1/2}}\right)\right]. \tag{37}$$

If V_i is defined such that

$$V_i = \frac{\partial}{\partial v_1}\left[e^2 \frac{(1 - v_1 \cdot v_2)}{F^{1/2}} \ln\left(\frac{2(A_2F)^{1/2} + B_2}{(4A_2F - B_2^2)^{1/2}}\right)\right], \tag{38}$$

for $i = 1$, for example, then from (30) and (37)

$$U_i = -v \cdot \frac{\partial V_i}{\partial r}. \tag{39}$$

When (36) and (39) are used in (35) the result is

$$v \cdot \frac{\partial}{\partial r}\left(P_i^{(1)}(r_j, v_j) - V_i - \frac{\partial \Omega}{\partial v_i}\right) = 0. \tag{40}$$

The comparison test $P_i^{(1)}(r_j, v_j)$ mentioned above confirms that the solution of (40) is

$$P_i^{(1)}(r_j, v_j) = V_i + \frac{\partial \Omega}{\partial v_i} \tag{41}$$

which for $i = 1$ comes out to be

$$P_1^{(1)}(r, v_1, v_2) = \frac{\partial}{\partial v_1}\left\{e^2 \frac{(1 - v_1 \cdot v_2)}{[v^2 - (v_1 \times v_2)^2]^{1/2}} \ln\left[\left(\frac{1 - v_1 \cdot v_2 - \frac{1}{4}(v_1 \times v_2)^2}{1 - v_2^2}\right)^{1/2} \times \left(\frac{(1 - \mu_2^2)^{1/2}[v^2 - (v_1 \times v_2)^2]^{1/2} + \hat{r} \cdot v - \mu_2 \cdot (v_1 \times v_2)}{[1 - \mu_1 \cdot \mu_2 - \frac{1}{4}(\mu_1 \times \mu_2)^2]^{1/2}[v^2 - (v_1 \times v_2)^2]^{1/2} + \hat{r} \cdot v - \frac{1}{2}(\mu_1 + \mu_2) \cdot (v_1 \times v_2)}\right)\right]\right\}. \tag{42}$$

Finally the energy, given by (11), is

$$E(r, v_1, v_2) = \sum_{i=1}^{2}\left(m_i u_i + U_i + \frac{\partial \Omega}{\partial r_i}\right) \cdot v_i - [-m_1(1 - v_1^2)^{1/2} - m_2(1 - v_2^2)^{1/2} + L^{(\mathrm{int})}(r, v_1, v_2)]$$

$$= \sum_{i=1}^{2}\left(\frac{m_i}{(1 - v_i^2)^{1/2}} + v_i \cdot \left(U_i + \frac{\partial \Omega}{\partial r_i}\right)\right) - L^{(\mathrm{int})}(r, v_1, v_2) \tag{43}$$

which, because of (32), is also

$$\sum_{i=1}^{2}\left(\frac{m_i}{(1-v_1^2)^{1/2}}+\mathbf{v}_i\cdot\mathbf{U}_i\right). \tag{44}$$

The square of $\mathbf{P}_i^{(0)}(\mathbf{r}_j,\mathbf{v}_j)$, from (29), is

$$m_i^2u_i^2+2m_i\mathbf{u}_i\cdot\left(\mathbf{U}_i+\frac{\partial\Omega}{\partial\mathbf{r}_i}\right)+O(e^4),\quad i=1,2,$$

whence, to order e^2,

$$[P_i^{(0)}(\mathbf{r}_j,\mathbf{v}_j)]^2+m_i^2=\frac{m_i^2}{1-v_i^2}+2\frac{m_i}{(1-v_i^2)^{1/2}}\mathbf{v}_i\cdot\left(\mathbf{U}_i+\frac{\partial\Omega}{\partial\mathbf{r}_i}\right).$$

To within terms of order e^2 the right-hand side of this expression is the square of

$$\frac{m_i}{(1-v_i^2)^{1/2}}+\mathbf{v}_i\cdot\left(\mathbf{U}_i+\frac{\partial\Omega}{\partial\mathbf{r}_i}\right),$$

so, from (43), the energy may also be written

$$E(\mathbf{r},\mathbf{v}_1,\mathbf{v}_2)=\sum_{i=1}^{2}\{[\mathbf{P}_i^{(0)}(\mathbf{r}_j,\mathbf{v}_j)]^2+m_i^2\}^{1/2}-L^{(\mathrm{int})}(\mathbf{r},\mathbf{v}_1,\mathbf{v}_2). \tag{45}$$

To summarize thus far: Equations (29) and (41) with the U's, V's, and Ω defined through (24), (25), (30), (31), and (38) are the first two Ostrogradski momenta, $\mathbf{P}_i^{(0)}$ and $\mathbf{P}_i^{(1)}$, corresponding to the joint Lagrangian (4), with accelerations and all higher derivatives set to zero. The energy E in the same limit, corresponding to (4), is given by either (44) or (45).

It may now be proven that when these straight line limits of the $\mathbf{P}_i^{(0)}$, $\mathbf{P}_i^{(1)}$ and E are substituted in (10), the resulting L' contains a system of twelve first-order equations equivalent to the second-order system of six equations (1). Instead of carrying out the proof here, however, the Hamiltonian formulation of (1) will be obtained first. The advantage in waiting is that a proof via the Hamiltonian will not only confirm that the equations (1) are implied as a first-order system, but that they are implied through Hamilton's equations.

IV. THE CANONICAL VARIABLES AND HAMILTONIAN

The usual scheme for fashioning a Hamiltonian from a Lagrangian implying a system of second-order equations involves replacing each velocity with a new independent variable. The dynamics may thus be expressed by twice as many first-order Hamilton equations as second-order Euler–Lagrange equations. Since these characteristics of a Hamiltonian formulation are already present when a system like (1) has been packaged in the form of the Lagrangian (9), the usual Legendre transformation, which aims at halving the order and doubling the size of the system, is inapplicable. Instead, the Hamiltonian formulation of the dynamics implied by (9) is obtained by a point transformation of the independent coordinates $\mathbf{r}_i$, $\mathbf{v}_i$:

$$\mathbf{r}_i=\mathbf{r}_i(\rho_j,\pi_j),\quad i=1,2,$$

$$\mathbf{v}_i=\mathbf{v}_i(\rho_j,\pi_j).$$

The transformed Lagrangian (9)

$$\sum_{j=1}^{2}\left[\left(\sum_{i=1}^{2}\alpha_i\cdot\frac{\partial\mathbf{r}_i}{\partial\rho_j}+\sum_{i=1}^{2}\beta_i\cdot\frac{\partial\mathbf{v}_i}{\partial\rho_j}\right)\cdot\dot{\rho}_j+\left(\sum_{i=1}^{2}\alpha_i\cdot\frac{\partial\mathbf{r}_i}{\partial\pi_j}+\sum_{i=1}^{2}\beta_i\cdot\frac{\partial\mathbf{v}_i}{\partial\pi_j}\right)\cdot\pi_j\right]-h,$$

where the dots join vectors with the same index symbol, and the α's, β's, and h are now written in terms of the new coordinates ρ_i and π_i, is certainly a Lagrangian for the transformed first-order equations of motion since the Euler–Lagrange equation is true in any coordinate system. If it can be arranged, however, that

$$\sum_{i=1}^{2}\left[\alpha_i\cdot\frac{\partial\mathbf{r}_i}{\partial\rho_j}+\beta_i\cdot\frac{\partial\mathbf{v}_i}{\partial\rho_j}\right]=\pi_j,$$

$$\sum_{i=1}^{2}\left[\alpha_i\cdot\frac{\partial\mathbf{r}_i}{\partial\pi_j}+\beta_i\cdot\frac{\partial\mathbf{v}_i}{\partial\pi_j}\right]=0; \tag{46}$$

then the Euler–Lagrange equations are, in fact, Hamilton's equations with the energy $h(\mathbf{r}_i(\rho_j,\pi_j),\mathbf{v}_i(\rho_j,\pi_j))$ being the Hamiltonian. The determination of a transformation satisfying (46) is the second and last step in Kerner's Hamiltonization method.

Of course the system (46) is a restatement in twelve dimensions of Pfaff's problem: to find a transformation $x_i=x_i(y_j)$ between sets of $2N$ coordinates such that for functions f_i of the x's

$$\sum_{i=1}^{2N}f_i(x_j)\,dx_i=\sum_{i=1}^{N}y_{N+i}\,dy_i.$$

The reduction is always possible and a method exists for finding transformations which will effect it.[16] The formalism is not necessary in the present instance, however, since the reduction can be carried out consistently to order e^2 almost by inspection. The solution is as follows.

The Lagrangian for the first-order version of (1) is

$$L'=\sum_{i=1}^{2}[\mathbf{P}_i^{(0)}(\mathbf{r}_j,\mathbf{v}_j)\cdot\dot{\mathbf{r}}_i+\mathbf{P}_i^{(1)}(\mathbf{r}_j,\mathbf{v}_j)\cdot\dot{\mathbf{v}}_i]-E(\mathbf{r},\mathbf{v}_1,\mathbf{v}_2).$$

Between $\dot{\mathbf{v}}_i$ and $\mathbf{P}_i^{(1)}$ insert the unit dyad in the form

$$\mathbf{I}=[(1-v_i^2)^{1/2}(\mathbf{I}-\mathbf{v}_i\mathbf{v}_i)]\cdot\left[\frac{\mathbf{I}(1-v_i^2)+\mathbf{v}_i\mathbf{v}_i}{(1-v_i^2)^{3/2}}\right].$$

Now, the time derivative of $\mathbf{u}_i$ is just

$$D\frac{\mathbf{v}_i}{(1-v_i^2)^{1/2}}=\left[\frac{\mathbf{I}(1-v_i^2)+\mathbf{v}_i\mathbf{v}_i}{(1-v_i^2)^{3/2}}\right]\cdot\dot{\mathbf{v}}_i,$$

so

$$\begin{aligned}\mathbf{P}_i^{(1)}(\mathbf{r}_j,\mathbf{v}_j)\cdot\dot{\mathbf{v}}_i&=\mathbf{P}_i^{(1)}(\mathbf{r}_j,\mathbf{v}_j)\cdot[(1-v_i^2)^{1/2}(\mathbf{I}-\mathbf{v}_i\mathbf{v}_i)]\cdot\dot{\mathbf{u}}_i\\&=D[(1-v_i^2)^{1/2}\mathbf{P}_i^{(1)}(\mathbf{r}_j,\mathbf{v}_j)\cdot(\mathbf{I}-\mathbf{v}_i\mathbf{v}_i)\cdot\mathbf{u}_i]\\&\quad-\mathbf{u}_i\cdot D[(1-v_i^2)^{1/2}(\mathbf{I}-\mathbf{v}_i\mathbf{v}_i)\cdot\mathbf{P}_i^{(1)}(\mathbf{r}_j,\mathbf{v}_j)].\end{aligned}$$

When this latter expression is substituted for $\mathbf{P}_i^{(1)}(\mathbf{r}_j,\mathbf{v}_j)\cdot\dot{\mathbf{v}}_i$ in the above Lagrangian, the first term, the exact time derivative, may be discarded. The Lagrangian L' may therefore be written equivalently as

$$\begin{aligned}L'=\sum_{i=1}^{2}\Bigg[&\left(m_i\mathbf{u}_i+\mathbf{U}_i+\frac{\partial\Omega}{\partial\mathbf{r}_i}\right)\cdot\dot{\mathbf{r}}_i\\&-\mathbf{u}_i\cdot D[(1-v_i^2)^{1/2}(\mathbf{I}-\mathbf{v}_i\mathbf{v}_i)\cdot\mathbf{P}_i^{(1)}(\mathbf{r}_j,\mathbf{v}_j)]\Bigg]\\&-E(\mathbf{r},\mathbf{v}_1,\mathbf{v}_2)\end{aligned} \tag{47}$$

where the expression for $\mathrm{P}_i^{(0)}(\mathrm{r}_j, \mathrm{v}_j)$ has been introduced from (29). Since, however, the construction need only be consistent to terms in e^2, and since U_i, Ω, and $\mathrm{P}_i^{(1)}$ each have e^2 for a factor, u_i in the second part of the summation may be replaced with

$$\frac{1}{m_i}\left(m_i\mathrm{u}_i + \mathrm{U}_i + \frac{\partial\Omega}{\partial \mathrm{r}_i}\right);$$

that is $\mathrm{P}_i^{(0)}(\mathrm{r}_j, \mathrm{v}_j)$ may be factored from both parts of the summation in (47) to give

$$L' = \sum_{i=1}^{2} \mathrm{P}_i^{(0)}(\mathrm{r}_j, \mathrm{v}_j) \cdot D[\mathrm{r}_i - m_i^{-1}(1-v_i^2)^{1/2} \times (\mathbf{I} - \mathrm{v}_i\mathrm{v}_i) \cdot \mathrm{P}_i^{(1)}(\mathrm{r}_j, \mathrm{v}_j)] - E(\mathrm{r}, \mathrm{v}_1, \mathrm{v}_2) \tag{48}$$

which completes the reduction. The canonical variables are immediately identifiable.

Canonical momenta:

$$\mathrm{P}_i = \mathrm{P}_i^{(0)}(\mathrm{r}_j, \mathrm{v}_j), \quad i = 1, 2. \tag{49a}$$

Canonical position:

$$\mathrm{Q}_i = \mathrm{r}_i - \frac{1}{m_i}(1-v_i^2)^{1/2}(\mathbf{I} - \mathrm{v}_i\mathrm{v}_i) \cdot \mathrm{P}_i^{(1)}(\mathrm{r}_j, \mathrm{v}_j), \quad i = 1, 2. \tag{49b}$$

To obtain the Hamiltonian it is necessary to express the energy in terms of the Q's and P's. From (45)

$$E = \sum_{i=1}^{2} (P_i^2 + m_i^2)^{1/2} - L^{(\mathrm{int})}(\mathrm{r}, \mathrm{v}_1, \mathrm{v}_2),$$

so that it remains only to transform the interaction term; but since $L^{(\mathrm{int})}(\mathrm{r}, \mathrm{v}_1, \mathrm{v}_2)$ has coefficient e^2, only the zero-order part of the inverse of (49),

$$\mathrm{v}_i = \frac{\mathrm{P}_i}{(P_i^2 + m_i^2)^{1/2}} + O(e^2) \tag{50a}$$

$$\mathrm{r}_i = \mathrm{Q}_i + O(e^2), \tag{50b}$$

is necessary in this last step. When (50) are used in $L^{(\mathrm{int})}(\mathrm{r}_1, \mathrm{v}_1, \mathrm{v}_2)$, which is given by (17), the energy becomes the *Hamiltonian*,

$$H = \sum_{i=1}^{2} (P_i^2 + m_i^2)^{1/2} + e^2 \frac{[1 - (P_1^2 + m_2^2)^{1/2}(P_2^2 + m_2^2)^{1/2}]}{\{Q^2 - [(\mathrm{Q}\times\mathrm{P}_1)/(P_1^2 + m_1^2)^{1/2}] \cdot [(\mathrm{Q}\times\mathrm{P}_2)/(P_2^2 + m_2^2)^{1/2}] - M^2/4\}^{1/2}} \tag{51}$$

where

$$M^2 = r^2(\mu_1 \times \mu_2)^2 = \frac{[(\mathrm{Q}\times\mathrm{P}_1)\times(\mathrm{Q}\times\mathrm{P}_2)]^2}{Q^2(P_1^2 + m_1^2)(P_2^2 + m_2^2)}. \tag{52}$$

This latter quantity has been provided with a distinct label because, as noted earlier, it is invariant under $\mathrm{r} \to \mathrm{r} + \mathrm{v}t$. Within an interaction term it therefore becomes a constant of the motion.

The statement made earlier about the constants of the motion which can be derived from a Lagrangian whose form is (9) applies also to (48). The total linear mechanical momentum is therefore the sum of the canonical momenta $\mathrm{P}_1 + \mathrm{P}_2$, and the total angular momentum is $\mathrm{Q}_1 \times \mathrm{P}_1 + \mathrm{Q}_2 \times \mathrm{P}_2$. It has been verified that these quantities are respectively the total linear and angular momentum for the system (1) which have been worked out in an earlier study.[13]

V. TWO TESTS

First, it will be certified that the equations of motion (1) are properly contained in the Hamiltonian (51) just given. The Hamiltonian is, in abbreviated form,

$$H = \sum_{i=1}^{2} (P_i^2 + m_i^2)^{1/2} - L^{(\mathrm{int})}(\mathrm{Q}, \mathrm{P}_1, \mathrm{P}_2), \tag{53}$$

and its derivative with respect to P_i is

$$\frac{\partial H}{\partial \mathrm{P}_i} = \frac{\mathrm{P}_i}{(P_i^2 + m_i^2)^{1/2}} - \frac{\partial}{\partial \mathrm{P}_i} L^{(\mathrm{int})}(\mathrm{Q}, \mathrm{P}_1, \mathrm{P}_2). \tag{54}$$

From (49a) and (29) the first piece on the right, to terms in e^2, is

$$\mathrm{v}_i + \frac{1}{m_i}(1-v_i^2)^{1/2}(\mathbf{I} - \mathrm{v}_i\mathrm{v}_i) \cdot \left(\mathrm{U}_i + \frac{\partial\Omega}{\partial \mathrm{r}_i}\right).$$

Since

$$\frac{\partial}{\partial \mathrm{P}_i} = \frac{1}{m_i}(1-v_i^2)^{1/2}(\mathbf{I} - \mathrm{v}_i\mathrm{v}_i) \cdot \frac{\partial}{\partial \mathrm{v}_i} + O(e^2), \tag{55}$$

then, to terms of order e^2, the right-hand side of (54), written in the original variables $\mathrm{r}, \mathrm{v}_1, \mathrm{v}_2$, is

$$\mathrm{v}_i + \frac{1}{m_i}(1-v_i^2)^{1/2}(\mathbf{I} - \mathrm{v}_i\mathrm{v}_i) \cdot \left[\mathrm{U}_i + \frac{\partial\Omega}{\partial \mathrm{r}_i} - \frac{\partial}{\partial \mathrm{v}_i} L^{(\mathrm{int})}(\mathrm{r}, \mathrm{v}_1, \mathrm{v}_2)\right].$$

The use of (35) gives

$$\frac{\partial H}{\partial \mathrm{P}_i} = \left(\mathrm{v}_1 \cdot \frac{\partial}{\partial \mathrm{r}_1} + \mathrm{v}_2 \cdot \frac{\partial}{\partial \mathrm{r}_2}\right)\left(\mathrm{r}_i - \frac{1}{m_i}(1-v_i^2)^{1/2} \times (\mathbf{I} - \mathrm{v}_i\mathrm{v}_i) \cdot \mathrm{P}_i^{(1)}(\mathrm{r}_j, \mathrm{v}_j)\right)$$

which is

$$\frac{\partial H}{\partial \mathrm{P}_i} = \left(\mathrm{v}_1 \cdot \frac{\partial}{\partial \mathrm{r}_1} + \mathrm{v}_2 \cdot \frac{\partial}{\partial \mathrm{r}_2}\right)\mathrm{Q}_i.$$

Since the additional piece $(m_1^{-1}\mathrm{F}_1 \cdot \partial/\partial \mathrm{u}_1 + m_2^{-1}\mathrm{F}_2 \cdot \partial/\partial \mathrm{u}_2)\mathrm{Q}_i$, where the F's are the forces in (1), is of order e^4, the final results may be written

$$\frac{\partial H}{\partial \mathrm{P}_i} = \sum_{j=1}^{2}\left(\frac{1}{m_j}\mathrm{F}_j \cdot \frac{\partial}{\partial \mathrm{u}_j} + \mathrm{v}_j \cdot \frac{\partial}{\partial \mathrm{r}_j}\right)\mathrm{Q}_i. \tag{56}$$

Next, the derivative of H with respect to Q_i is taken. The derivative with respect to Q_i, however, is sufficiently represented by the derivative with respect to r_i when the operand has a factor e^2. So,

$$\frac{\partial H}{\partial \mathrm{Q}_i} = -\frac{\partial}{\partial \mathrm{r}_i} L^{(\mathrm{int})}(\mathrm{r}, \mathrm{v}_1, \mathrm{v}_2).$$

But, from (32) and (33) the right-hand side may be replaced with

$$-\sum_{j=1}^{2}\left(\frac{1}{m_j}\mathbf{F}_j\cdot\frac{\partial}{\partial \mathbf{u}_j}+\mathbf{v}_j\cdot\frac{\partial}{\partial \mathbf{r}_j}\right)\left(m_i\mathbf{u}_i+\mathbf{U}_i+\frac{\partial\Omega}{\partial \mathbf{r}_i}\right)$$

without altering its e^2-order value. Hence, from the definition of the canonical momentum (49a),

$$\frac{\partial H}{\partial \mathbf{Q}_i}=-\sum_{j=1}^{2}\left(\frac{1}{m_j}\mathbf{F}_j\frac{\partial}{\partial \mathbf{u}_j}+\mathbf{v}_j\frac{\partial}{\partial \mathbf{r}_j}\right)\mathbf{P}_i. \tag{57}$$

Thus, with the Q's and P's expressed now in terms of the r's and u's, Hamilton's equations,

$$\dot{\mathbf{Q}}_i=\frac{\partial H}{\partial \mathbf{P}_i},\quad \dot{\mathbf{P}}_i=-\frac{\partial H}{\partial \mathbf{Q}_i}$$

become

$$\sum_{j=1}^{2}\left[\left(\dot{\mathbf{u}}_j-\frac{1}{m_j}\mathbf{F}_j\right)\cdot\frac{\partial \mathbf{Q}_i}{\partial \mathbf{u}_j}+(\dot{\mathbf{r}}_j-\mathbf{v}_j)\cdot\frac{\partial \mathbf{Q}_i}{\partial \mathbf{r}_j}\right]=0, \tag{58a}$$

$$\sum_{j=1}^{2}\left[\left(\dot{\mathbf{u}}_j-\frac{1}{m_j}\mathbf{F}_j\right)\cdot\frac{\partial \mathbf{P}_i}{\partial \mathbf{u}_j}+(\dot{\mathbf{r}}_j-\mathbf{v}_j)\cdot\frac{\partial \mathbf{P}_i}{\partial \mathbf{r}_j}\right]=0. \tag{58b}$$

Since

$$\frac{\partial \mathbf{Q}_i}{\partial \mathbf{r}_j}\approx\frac{1}{m_j}\frac{\partial \mathbf{P}_i}{\partial \mathbf{u}_j}=\delta_{ij}+O(e^2),$$

while the other coefficients are of order e^2, the matrix of coefficients in (58) has an inverse consistent to that order of approximation. The original equations of motion (1) are thus *uniquely* implied by the canonical formulation presented in the last section.

Second, it must be proven that the canonical formulation is not tied to one Lorentz frame. For interacting particles it is possible to express the Lorentz transformation connecting primed and unprimed sets of particle variables, such that each set is taken at a *single* instant of time in its own inertial frame. This expression of the Lorentz transformation, which requires that the particles motions be used as a retardation condition, is easily obtained for the infinitesimal transformation. For in that instance, the separation of the time coordinates in the new reference frame of world points simultaneous in the old is infinitesimal, and world lines connecting those ϵ-separated time coordinates may be approximated by straight lines. Thus, for the case of two particles, the infinitesimal Lorentz transformation of their positions and velocities may be written

$$\begin{aligned}
\mathbf{r}_1'&=\mathbf{r}_1+\epsilon[\mathbf{v}_1(\hat{\mathbf{n}}\cdot\mathbf{r})-\hat{\mathbf{n}}t],\\
\mathbf{r}_2'&=\mathbf{r}_2-\epsilon\hat{\mathbf{n}}t,\\
\mathbf{v}_1'&=\mathbf{v}_1+\epsilon\{\mathbf{v}_1(\hat{\mathbf{n}}\cdot\mathbf{v}_1)-\hat{\mathbf{n}}\\
&\quad+[m_1^{-1}(1-v_1^2)^{1/2}(\mathbf{I}-\mathbf{v}_1\mathbf{v}_1)\cdot\mathbf{F}_1](\hat{\mathbf{n}}\cdot\mathbf{r})\},\\
\mathbf{v}_2'&=\mathbf{v}_2+\epsilon[\mathbf{v}_2(\hat{\mathbf{n}}\cdot\mathbf{v}_2)-\hat{\mathbf{n}}],
\end{aligned} \tag{59}$$

where $t'=t-\epsilon\hat{\mathbf{n}}\cdot\mathbf{r}_2$ and $\epsilon\hat{\mathbf{n}}$ is the velocity of frame S' with respect to S.[17] All quantities on the left are reckoned at t', all those on the right at t. Since $\mathbf{F}_1$ is the force on the first particle, the quantity in the square brackets of the third equation is just the acceleration of the first particle at time t.

The equations of motion (1) are form invariant in straight line approximation under the Lorentz transformation (59), and may therefore be derived in different inertial frames from the same form of Hamiltonian. Nevertheless, these formally equivalent Hamiltonians will not be canonically equivalent unless they admit the Lorentz transformations as canonical transformations. What must be shown then is that (59) is canonically represented with respect to the Hamiltonian formulation described in the previous section. To demonstrate the canonicity of the Lorentz transformation and, consequently, the covariance of the Hamiltonian formulation, it suffices to show that under (59) the differential of the action is form invariant to within an exact differential up to first order in e^2 and ϵ.

From (48) the differential of the action may be written

$$\sum_{i=1}^{2}\mathbf{P}_i\cdot d\mathbf{Q}_i-H dt, \tag{60a}$$

or, in the equivalent Pfaffian-expanded form,

$$\sum_{i=1}^{2}[\mathbf{P}_i^{(0)}(\mathbf{r}_j\cdot\mathbf{v}_j)\cdot d\mathbf{r}_i+\mathbf{P}_i^{(1)}(\mathbf{r}_j\cdot\mathbf{v}_j)\cdot d\mathbf{v}_i]-E(\mathbf{r},\mathbf{v}_1,\mathbf{v}_2)\,dt. \tag{60b}$$

The latter of these is the easier to work with. When (29), (41), and (44) are substituted for the coefficients of the d's, (60b) becomes

$$\sum_{i=1}^{2}\left[\left(m_i\mathbf{u}_i+\mathbf{U}_i+\frac{\partial\Omega}{\partial \mathbf{r}_i}\right)\cdot d\mathbf{r}_i+\left(\mathbf{V}_i+\frac{\partial\Omega}{\partial \mathbf{v}_i}\right)\cdot d\mathbf{v}_i-\left(\frac{m_i}{(1-v_i^2)^{1/2}}+\mathbf{v}_i\cdot\mathbf{U}_i\right)dt\right],$$

which is

$$\sum_{i=1}^{2}\{(m_i\mathbf{u}_i+\mathbf{U}_i)\cdot d\mathbf{r}_i+\mathbf{V}_i\cdot d\mathbf{v}_i-[m_i(1-v_i^2)^{-1/2}+\mathbf{v}_i\cdot\mathbf{U}_i]dt\}+d\Omega. \tag{61}$$

Since $d\Omega$ will transform to another exact differential, it is only necessary to transform the term in braces. The task is simplified by proceeding in stages. For example, if the transformations of A_2, B_2, F and $\partial/\partial\mathbf{r}$ induced by (59) are worked out beforehand, then from (30)

$$\begin{aligned}
m_1\mathbf{u}_1'+\mathbf{U}_1'&=m_1\mathbf{u}_1+\mathbf{U}_1+\epsilon[(\hat{\mathbf{n}}\cdot\mathbf{r})\mathbf{F}_1-\hat{\mathbf{n}}m_1(1-v_1^2)^{-1/2}]\\
&\quad+e^2\frac{\{\mathbf{v}_2+\epsilon[\mathbf{v}_2(\hat{\mathbf{n}}\cdot\mathbf{v}_2)-\hat{\mathbf{n}}]\}}{\{A_2+\epsilon[(\hat{\mathbf{n}}\cdot\mathbf{r})B_2+2(\hat{\mathbf{n}}\cdot\mathbf{v}_2)A_2]\}^{1/2}}\\
&\quad+\left(\frac{\partial}{\partial\mathbf{r}}-\epsilon\hat{\mathbf{n}}\mathbf{v}_1\cdot\frac{\partial}{\partial\mathbf{r}}\right)e^2\frac{(1-\mathbf{v}_1\cdot\mathbf{v}_2)}{F^{1/2}}\\
&\quad\times\ln\left\{\frac{2(A_2F)^{1/2}+B_2}{(4A_2F-B_2^2)^{1/2}}\left[1+\epsilon(\hat{\mathbf{n}}\cdot\mathbf{r})\left(\frac{F}{A_2}\right)^{1/2}\right]\right\}.
\end{aligned}$$

When all terms are expanded to first order in ϵ, and the expression for $\mathbf{F}_1$ is substituted from the right-hand side of (1a), the result is

$$m_1\mathbf{u}_1'+\mathbf{U}_1'=m_1\mathbf{u}_1+\mathbf{U}_1-\epsilon\hat{\mathbf{n}}[m_1(1-v_1^2)^{-1/2}+\mathbf{v}_1\cdot\mathbf{U}_1]. \tag{62a}$$

Similarly,

$$m_2\mathbf{u}_2' + \mathbf{U}_2' = m_2\mathbf{u}_2 + \mathbf{U}_2 - \epsilon\hat{\mathbf{n}}[m_2(1-v_2^2)^{-1/2} + \mathbf{v}_2\cdot\mathbf{U}_2]. \tag{62b}$$

The V's, given by (38), are transformed next:

$$\mathbf{V}_1' = \mathbf{V}_1 - \epsilon[(\hat{\mathbf{n}}\cdot\mathbf{v}_1)\mathbf{V}_1 + \hat{\mathbf{n}}(\mathbf{v}_1\cdot\mathbf{V}_1) + (\hat{\mathbf{n}}\cdot\mathbf{r})\mathbf{U}_1], \tag{62c}$$

$$\mathbf{V}_2' = \mathbf{V}_2 - \epsilon[(\hat{\mathbf{n}}\cdot\mathbf{v}_2)\mathbf{V}_2 + \hat{\mathbf{n}}(\mathbf{v}_2\cdot\mathbf{V}_2)]. \tag{62d}$$

Finally,

$$m_i(1-v_i'^2)^{-1/2} + \mathbf{v}_i'\cdot\mathbf{U}_i' = m_i(1-v_i^2)^{-1/2} + \mathbf{v}_i\cdot\mathbf{U}_i - \epsilon\hat{\mathbf{n}}\cdot(m_i\mathbf{u}_i + \mathbf{U}_i), \tag{62e}$$

$i = 1, 2.$

The Lorentz transformation of the braces of (61) is now a matter of replacing the terms in the braces using (62) and the differentials of (59). The status of $\mathbf{r}_i$ and $\mathbf{v}_i$ as independent coordinates in the differential form (60b) must be recalled here to forestall erroneously cancelling $\mathbf{U}_i\cdot(d\mathbf{r}_i - \mathbf{v}_i dt)$. The "simplification" is not admissable in the action since $d\mathbf{r}_i = \mathbf{v}_i\,dt$ is not an identity. Each term in the braces of (61) is retained and transformed to give

$$\sum_{i=1}^{2}\{(m_i\mathbf{u}_i' + \mathbf{U}_i')\cdot d\mathbf{r}_i' + \mathbf{V}_i'\cdot d\mathbf{v}_i - [m_i(1-v_i'^2)^{-1/2} + \mathbf{v}_i'\cdot\mathbf{U}_i']dt'\}$$
$$= \sum_{i=1}^{2}\{(m_i\mathbf{u}_i + \mathbf{U}_i)\cdot d\mathbf{r}_i + \mathbf{V}_i\cdot d\mathbf{v}_i - [m_i(1-v_i^2)^{-1/2} + \mathbf{v}_i\cdot\mathbf{U}_i]dt - d[\epsilon(\hat{\mathbf{n}}\cdot\mathbf{r})m_1(1-v_1^2)^{1/2}].$$

This result confirms that the Hamiltonian formulation of the straight line electrodynamics presented in the foregoing section is an inertial-frame independent description.

It is obvious from (49) and (51) that the other transformations of the inhomogenous Lorentz group, time translation, etc., are also canonical transformations.

VI. EXTENSION TO N CHARGES

The forces in (1) obey the principle of superposition. If F_{ij} denotes the force on the ith charge due to the jth charge alone,

$$\mathbf{F}_{ij} = e^2(1-v_j^2)\frac{\hat{\mathbf{r}}_{ij} - \mathbf{v}_i\times(\hat{\mathbf{r}}_{ij}\times\mathbf{v}_j)}{r_{ij}^2[1-(\hat{\mathbf{r}}_{ij}\times\mathbf{v}_j)^2]^{3/2}},$$

then

$$m_i\dot{\mathbf{u}}_i = \mathbf{F}_i = \sum_{j=1}^{N}{}' \mathbf{F}_{ij}, \quad i = 1\cdots N, \tag{63}$$

is the system of equations for N interacting charges correct to first order in e^2. The primed $\sum$ means $j\neq i$ in the summation.

When all the formulas and expressions which have arisen so far are similarly relabelled, i, j replacing 1, 2 in the arguments of all quantities, then it is a straightforward exercise to show that if

$$\mathbf{P}_i = m_i\mathbf{u}_i + \sum_{j=1}^{N}{}'\mathbf{U}_{ij} + \frac{\partial}{\partial\mathbf{r}_i}\left(\frac{1}{2}\sum_{i,j=1}^{N}{}'\Omega_{ij}\right), \tag{64a}$$

$$\mathbf{Q}_i = \mathbf{r}_i - \frac{1}{m_i}(1-v_i^2)^{1/2}(\mathbf{I} - \mathbf{v}_i\mathbf{v}_i)\cdot\left[\sum_{j=1}^{N}{}'\mathbf{P}_{ij}{}^{(1)}\right], \quad i = 1\cdots N, \tag{64b}$$

and

$$H = \sum_{i=1}^{N}(P_i^2 + m_i^2)^{1/2} - \frac{1}{2}\sum_{i,j=1}^{N}{}' L_{ij}{}^{(\mathrm{int})}, \tag{65}$$

comprise a Hamiltonian formulation of (63) for $N=2$, then they comprise a Hamiltonian formulation for all N. The quantities in (64) and (65) are defined through the $N=2$ case, $\mathbf{U}_{12}$ is the $\mathbf{U}_1(\mathbf{r}, \mathbf{v}_1, \mathbf{v}_2)$ in (30), Ω_{12} is the Ω in (31), and so on.

The infinitesimal Lorentz transformation of the positions and velocities of the N particles can be written

$$\mathbf{r}_i' = \mathbf{r}_i + \epsilon[\mathbf{v}_i(\hat{\mathbf{n}}\cdot\mathbf{r}_{ik}) - \hat{\mathbf{n}}t],$$
$$\mathbf{v}_i' = \mathbf{v}_i + \epsilon\{\mathbf{v}_i(\hat{\mathbf{n}}\cdot\mathbf{v}_i) - \hat{\mathbf{n}} + [m_i^{-1}(1-v_i^2)^{1/2} \times(\mathbf{I} - \mathbf{v}_i\mathbf{v}_i)\cdot\mathbf{F}_i](\hat{\mathbf{n}}\cdot\mathbf{r}_{ik})\}, \tag{66}$$
$$i = 1\cdots N,$$

where the world point of the kth particle is "unshifted." That is, the transformation of the time is given by

$$t' = t - \epsilon\hat{\mathbf{n}}\cdot\mathbf{r}_k,$$

while the motion of each of the other particles is used to express $\mathbf{r}_i(t_i)$ in terms of $\mathbf{r}_i(t_k) = \mathbf{r}_i(t)$. If this transformation is applied to the differential of the action for the system (63),

$$\sum_{i=1}^{N}\left[\left(m_i\mathbf{u}_i + \sum_{j=1}^{N}{}'\mathbf{U}_{ij}\right)\cdot d\mathbf{r}_i + \left(\sum_{j=1}^{N}{}'\mathbf{V}_{ij}\right)\cdot d\mathbf{v}_i - \left(m_i(1-v_i^2)^{-1/2} + \mathbf{v}_i\cdot\sum_{j=1}^{N}{}'\mathbf{U}_{ij}\right)dt\right] + d\left(\frac{1}{2}\sum_{i,j=1}^{N}{}'\Omega_{ij}\right),$$

then the term in brackets is form invariant, to terms in ϵ and e^2, to within the exact differential

$$-d\sum_{i=1}^{N}[\epsilon(\hat{\mathbf{n}}\cdot\mathbf{r}_{ik})m_i(1-v_i^2)^{1/2}].$$

Thus the Lorentz transformations are canonical transformations with respect to (65) for arbitrary values of N.

VII. DISCUSSION

If the speed of light is restored to its proper locations in the formulas which comprise the canonical formulation, then these formulas may be expanded in inverse powers of c^2. Thus the variables $\mathbf{P}_1$ and $\mathbf{Q}_1$ to terms of order c^{-4} are given by

$$\mathbf{P}_1 \approx m_1\mathbf{v}_1\left(1 + \frac{v_1^2}{2c^2} + \frac{3}{8}\frac{v_1^4}{c^4}\right) + \frac{e^2}{2rc^2}[\mathbf{v}_2 + \hat{\mathbf{r}}(\hat{\mathbf{r}}\cdot\mathbf{v}_2)]$$
$$+ \frac{e^2}{8rc^4}\{[(\hat{\mathbf{r}}\cdot\mathbf{v}_2)^2 - v_2^2]\mathbf{v}_1 + [2(\mathbf{v}_1\cdot\mathbf{v}_2) + 2(\hat{\mathbf{r}}\cdot\mathbf{v}_1)(\hat{\mathbf{r}}\cdot\mathbf{v}_2)$$
$$+ v_2^2 - (\hat{\mathbf{r}}\cdot\mathbf{v}_2)^2]\mathbf{v}_2 + [v_2^2(\hat{\mathbf{r}}\cdot\mathbf{v}_1) + 3v_2^2(\hat{\mathbf{r}}\cdot\mathbf{v}_2) - 2(\hat{\mathbf{r}}\cdot\mathbf{v}_2)(\mathbf{v}_1\cdot\mathbf{v}_2)$$
$$- 3(\hat{\mathbf{r}}\cdot\mathbf{v}_2)^2(\hat{\mathbf{r}}\cdot\mathbf{v}_1 + \hat{\mathbf{r}}\cdot\mathbf{v}_2)]\hat{\mathbf{r}}\}, \tag{67a}$$

$$\mathbf{Q}_1 \approx \mathbf{r}_1 - \frac{e^2}{8m_1c^4}\{2(\hat{\mathbf{r}}\cdot\mathbf{v}_2)\mathbf{v}_2 + [(\hat{\mathbf{r}}\cdot\mathbf{v}_2)^2 - v_2^2]\hat{\mathbf{r}}\}. \tag{67b}$$

These same results may also be obtained directly from the relation of the canonical variables to the primitive series representations of the Ostrogradski momenta $\mathbf{P}_i{}^{(0)}$ and $\mathbf{P}_i{}^{(1)}$. For example, the canonical momentum of the first particle is given to all orders by the straight line component of (12). The expansion of the Hamiltonian

to c^{-4} is

$$H \approx \sum_{i=1}^{2}\left(\frac{P_i^2}{2m_i}-\frac{P_i^4}{8m_i^3c^2}+\frac{P_i^6}{16m_i^5c^4}\right)+\frac{e^2}{Q}-\frac{e^2}{2m_1m_2c^2Q}[(\mathbf{P}_1\cdot\mathbf{P}_2)$$
$$+(\hat{\mathbf{Q}}\cdot\mathbf{P}_1)(\hat{\mathbf{Q}}\cdot\mathbf{P}_2)]+\frac{e^2}{8m_1^2m_2^2c^4Q}\left(P_1^2P_2^2-P_2^2(\hat{\mathbf{Q}}\cdot\mathbf{P}_1)^2\right.$$
$$-P_1^2(\hat{\mathbf{Q}}\cdot\mathbf{P}_2)^2+3(\hat{\mathbf{Q}}\cdot\mathbf{P}_1)^2(\hat{\mathbf{Q}}\cdot\mathbf{P}_2)^2-2(\mathbf{P}_1\cdot\mathbf{P}_2)^2$$
$$\left.+4\frac{m_2}{m_1}P_1^2(\mathbf{P}_1\cdot\mathbf{P}_2)+4\frac{m_1}{m_2}P_2^2(\mathbf{P}_1\cdot\mathbf{P}_2)\right), \tag{68}$$

where the rest energy has been dropped.

The first two terms in this expansion in inverse powers of c^2 of the canonical formulation of the e^2-accurate, two-charge problem will be recognized as the approximately relativistic treatment originally presented by Darwin.[18] Though it turns out in the Darwin approximation that physical position can be taken for canonical position, the zero-interaction theorem suggests that that possibility can in no order of relativistic approximation be taken for granted. Certainly, in the next order of approximation beyond Darwin's, the physical positions $\mathbf{r}_i$ *cannot* be canonical; that is to say, a c^{-4}-order Hamiltonian equivalent to (68) for which the contrary would be the case does not, in fact, exist. If it were otherwise a canonical transformation could be found which would remove the interaction piece in (67b); but there is no such transformation. A proof goes as follows. Let a canonical transformation be induced by the addition of an exact derivative $d\Lambda$ within the action integral for the problem so that

$$\sum_{i=1}^{2}\left[\left(\mathbf{P}_i^{(0)}(\mathbf{r}_j,\mathbf{v}_j)+\frac{\partial\Lambda}{\partial\mathbf{r}_i}\right)\cdot d\mathbf{r}_i+\left(\mathbf{P}_i^{(1)}(\mathbf{r}_j,\mathbf{v}_j)+\frac{\partial\Lambda}{\partial\mathbf{v}_i}\right)\cdot d\mathbf{v}_i\right]$$
$$-\left(H-\frac{\partial\Lambda}{\partial t}\right)dt,$$

for example, replaces (60b). Since the Pffafian reduction to canonical variables follows the same rule, the new Q's are, from (49b),

$$\mathbf{Q}_i'=\mathbf{r}_i-\frac{1}{m_i}(1-v_i^2)^{1/2}(\mathbf{I}-\mathbf{v}_i\mathbf{v}_i)\cdot\left(\mathbf{P}_i^{(1)}(\mathbf{r}_j,\mathbf{v}_j)+\frac{\partial\Lambda}{\partial\mathbf{v}_i}\right)$$
$$=\mathbf{Q}_i-\frac{1}{m_i}(1-v_i^2)^{1/2}(\mathbf{I}-\mathbf{v}_i\mathbf{v}_i)\cdot\frac{\partial\Lambda}{\partial\mathbf{v}_i}. \tag{69}$$

If Λ has the coefficient e^2 then to first-order in e^2 (69) may also be written [using (55)] as

$$\mathbf{Q}_i'=\mathbf{Q}_i-\frac{\partial\Lambda}{\partial\mathbf{P}_i},$$

which is the familiar rule for the infinitesimal canonical transformation of the Q's. Suppose, now, that $\mathbf{Q}_1$ is given by (67b) and $\mathbf{Q}_1'$ is to be $\mathbf{r}_1$. Since the difference between these has a factor of c^{-4}, it is only necessary that

$$\frac{\partial\Lambda}{\partial\mathbf{v}_1}=-\frac{e^2}{8c^4}\{2(\hat{\mathbf{r}}\cdot\mathbf{v}_2)\mathbf{v}_2+[(\hat{\mathbf{r}}\cdot\mathbf{v}_2)^2-v_2^2]\hat{\mathbf{r}}\} \tag{70a}$$

in order that $\mathbf{r}_1$ be canonical to order c^{-4}. Similarly, the condition for transforming the index-exchanged (67b) to $\mathbf{r}_2$ is

$$\frac{\partial\Lambda}{\partial\mathbf{v}_2}=\frac{e^2}{8c^4}\{2(\hat{\mathbf{r}}\cdot\mathbf{v}_1)\mathbf{v}_1+[(\hat{\mathbf{r}}\cdot\mathbf{v}_1)^2-v_1^2]\hat{\mathbf{r}}\}. \tag{70b}$$

The nonexistence of a generating function Λ, thus, of *canonical* variables $\mathbf{r}_i$ to order c^{-4}, follows now from the incompatibility of these two equations for Λ.

Even for equal masses, the fact remains that there is no Λ. This obvious point is of interest because a function of the form

$$\frac{e^2}{c^4}f(\mathbf{r},\mathbf{v}_1,\mathbf{v}_2)+\frac{e^4}{c^4}g(\mathbf{r},\mathbf{v}_1,\mathbf{v}_2)+\frac{e^6}{c^4}h(r)+\frac{m}{16c^4}(v_1^4+v_2^4),$$

whose first term is the c^{-4} component of (17), was in 1956 offered, and continues to be referred to, as $L^{(4)}(\mathbf{r},\mathbf{v}_1,\mathbf{v}_2)$, the fourth-order correction to the classical Lagrangian for a pair of identical charges.[19] Since the identical-charge condition promotes the radiation to a c^{-5} effect, it was thought, correctly, that an extension of the Darwin Lagrangian to embrace all c^{-4}-order terms would be a consistent calculation in this instance. The fault lay in the construction of the extension where the lower order equations of motion of the particles were used to *replace* accelerations in the c^{-4} term of a valid Lagrangian equivalent to (4). The upshot is that while, for example, e^4-order terms replaced terms of the type $e^2\dot{\mathbf{v}}_i\cdot\boldsymbol{\phi}(\mathbf{r},\mathbf{v}_1,\mathbf{v}_2)$ in the original Lagrangian, they also annihilated e^2-order terms in the correct equations of motion through loss of the $D^2\partial/\partial\dot{\mathbf{v}}_i$ contribution from the Ostrogradski operator. The attempt points up the pitfalls of constructing approximations to equations of motion indirectly through the Lagrangian.

The mention of canonical transformation brings us now to the final question: Why not drop the Ω, given by (31), which has been carried from the very beginning of this discussion? Even though Ω has not been introduced "artificially" but derives from terms that already appear in each of the Ostrogradski momenta for the complete dynamics implied by (4)—the gradient $\partial\Omega/\partial\mathbf{r}$, recall, is the straight line approximation to the antisymmetric, last term in (12)—still, the tendency is usually to drop such a quantity when it becomes clear, as it is here even before (61), that it plays no part in determining the equations of motion to the required order of approximation. An answer to this question is more easily and effectively presented in the light of some consequences, so, for this purpose, let Ω be discarded.

If the new, Ω-less canonical variables are distinguished by overbars, then

$$\bar{\mathbf{P}}_i=m_i\mathbf{u}_i+\mathbf{U}_i \tag{71a}$$

which, for $i=1$, for instance, is

$$\bar{\mathbf{P}}_1=m_1\mathbf{u}_1+\frac{e^2}{r}\frac{\mathbf{v}_2}{(1-\mu_2^2)^{1/2}}$$
$$+\frac{\partial}{\partial\mathbf{r}}\left[e^2\frac{(1-\mathbf{v}_1\cdot\mathbf{v}_2)}{[v^2-(\mathbf{v}_1\times\mathbf{v}_2)^2]^{1/2}}\right.$$
$$\left.\times\ln\left(\frac{(1-\mu_2^2)^{1/2}[v^2-(\mathbf{v}_1\times\mathbf{v}_2)^2]^{1/2}+\hat{\mathbf{r}}\cdot\mathbf{v}-\mu_2\cdot(\mathbf{v}_1\times\mathbf{v}_2)}{(1-v_2^2)^{1/2}[\mu^2-(\boldsymbol{\mu}_1\times\boldsymbol{\mu}_2)^2]^{1/2}}\right)\right]. \tag{72}$$

Corresponding to $\bar{\mathbf{P}}_i$ is the canonical coordinate

$$\bar{\mathbf{Q}}_i=\mathbf{r}_i-\frac{1}{m_i}(1-v_i^2)^{1/2}(\mathbf{I}-\mathbf{v}_i\mathbf{v}_i)\cdot\frac{\partial}{\partial\mathbf{v}_i}\{\ \}_i \tag{71b}$$

where $\{\}_1$ is that given in (72). These canonical variables look no different in kind from their counterparts with Ω which have already been given. The new Hamiltonian, on the other hand, clearly suggests what has happened. The new $\bar{H}$, which may be obtained by either transforming (51) or, more easily, by rederiving (45) with Ω suppressed, is simply

$$\bar{H} = (\bar{P}_1^2 + m_1^2)^{1/2} + (\bar{P}_2^2 + m_2^2)^{1/2}, \tag{73}$$

a pair of free-particle Hamiltonians![20] This sort of result, though, has been anticipated. It has already been observed in (33) that each of the momenta $\bar{P}_i$ is a constant of the motion to order e^2, and Eqs. (62) show that under an infinitesimal Lorentz transformation

$$\bar{P}_i' = \bar{P}_i - \epsilon\hat{n}(\bar{P}_i^2 + m_i^2)^{1/2},$$
$$(\bar{P}_i'^2 + m_i^2)^{1/2} = (\bar{P}_i^2 + m_i^2)^{1/2} - \epsilon\hat{n}\cdot\bar{P}_i,$$

the usual free-particle relations.

The two-charge problem is thus uncoupled somewhat, and quite legitimately, but this uncoupling may not be very useful. The act of "dropping" Ω, after all, amounts to an infinitesimal canonical transformation of P_i, Q_i generated by $-\Omega$; for from (29) and (49a), (41), (49b), and (55)

$$\bar{P}_i = P_i + \frac{\partial}{\partial Q_i}(-\Omega), \tag{74}$$

$$\bar{Q}_i = Q_i - \frac{\partial}{\partial P_i}(-\Omega).$$

Inasmuch as the new Hamiltonian function $\bar{H}(\bar{Q}_i, \bar{P}_i)$ will be just $H(Q_i, \bar{P}_i)$ with an additional, "infinitesimal" (i.e., e^2-order) piece (the generator not being a constant of the motion), it is evident that the interaction term in the original H is in a position to be cancelled.

The following illustration will be useful. Here is the simplest of canonical formulations:

$$H(x,p) = \frac{p^2}{2m} + \kappa\phi(x), \quad p = m\dot{x}, \tag{75}$$

where κ denotes the strength of the force. If $\bar{p} = p + \kappa(m/p)\phi(x)$, then $H(\bar{p}) = \bar{p}^2/2m$ to first order in κ. The latter function is still a Hamiltonian, $\bar{H}$ now, because the transformation from p to $\bar{p}$ is an infinitesimal canonical transformation. The corresponding transformation of position is $\bar{x} = x + \kappa(m/p^2)\int\phi(x)\,dx$, and the generator is $\kappa(m/p)\int\phi(x)\,dx$. It can be verified by inspection that, to first order in κ, the new momentum $\bar{p}$ is constant and equal to $m\dot{\bar{x}}$, as required by Hamilton's equations.

This instructive example also points up the fact that in first order the free-particle format is a procrustean bed; the description of a particle's behavior will always fit providing most of it is chopped off. Since $\bar{x} = (\bar{p}/m)t + \bar{x}_0$, the transformation equation for position immediately provides $x(t)$ as a first-order correction to *uniform* motion. This kind of truncation is precisely what happens with the two-charge problem treated in this paper when the Ω is dropped, only in that case it happens in the nonrelativistic limit. Since Ω does not vanish in the limit $c \to \infty$, the transformation (74) does not then reduce to the identity. Thus, for example, while P_1 goes to m_1v_1, the transformed $\bar{P}_1$, from (72), goes to $m_1v_1 - (e^2/r)\mu^{-2}(\hat{r}\times\mu)$. The results exactly parallel those illustrated from (75); the exact nonrelativistic Hamiltonian

$$H(r, p_1, p_2) = \frac{p_1^2}{2m_1} + \frac{p_2^2}{2m_2} + \frac{e^2}{r}$$

is replaced by the approximate one,

$$\bar{H} = \frac{\bar{p}_1^2}{2m_1} + \frac{\bar{p}_2^2}{2m_2} = H(r, p_1, p_2) + O(e^4),$$

and so forth.

To sum up: The Hamiltonian given by (51) with P_i, Q_i given by (49) is an approximation to order e^2 implying the approximate dynamical equations (1). The nonrelativstic limit of that canonical formulation *happens*, however, to imply exactly, i.e., to *all* orders of e^2, the nonrelativistic limit of (1), the Coulomb interaction. Canonical transformations, which can be carried out consistently to order e^2, lead to new e^2-order canonical formulations, which still imply the dynamics (1), but then in the limit, the Coulomb interaction may only be implied with $O(e^4)$ error. The Hamiltonian (51), therefore, or any canonically equivalent Hamiltonian which also contains the Coulomb limit exactly, provides an e^2-order correction to the Coulomb interaction. Hamiltonians canonically equivalent to (51) but which do not contain the Coulomb limit exactly can only consistently provide an e^2-order correction to the uncoupled, uniform motion of the charges.

ACKNOWLEDGMENTS

The research reported here was carried out at the Theoretical Physics Institute of the University of Alberta. I wish to thank Professors M. Razavy, member, and D.D. Betts, director of the Institute, for their generous support and hospitality.

APPENDIX A

Here it will be shown that there can be two ways to construe a Lagrangian.

It will be convenient to begin with some additional remarks on the generalized mechanics generated by $L(q_i, \ldots, D^n q_i)$[21]. The Ostrogradski momenta

$$P_i^{(l)} = \sum_{s=l}^{n-1}(-D)^{s-l}\frac{\partial L}{\partial(D^{s+1}q_i)}, \quad l = 0, \ldots, n-1, \tag{A1}$$

satisfy the identities

$$\dot{P}_i^{(l+1)} \equiv -P_i^{(l)} + \frac{\partial L}{\partial(D^{l+1}q_i)}, \quad l = 0, \ldots, n-2, \tag{A2}$$

$$0 \equiv -P_i^{(n-1)} + \frac{\partial L}{\partial(D^n q_i)}, \tag{A3}$$

where of course $\dot{P}$ means DP. The equations of motion are

$$\dot{P}_i^{(0)} = \frac{\partial L}{\partial q_i}. \tag{A4}$$

If the dynamics is nondegenerate (so that L implies $2n$-order differential equations for each dimension i), then it follows from (A2) and (A3) that $P_i^{(n-1)}$ depends on the same variables as the Lagrangian, that

$$P_i^{(l)} = P_i^{(l)}(q_j, \ldots, D^{2n-l-1}q_j), \quad l = 0, \ldots, n-1, \tag{A5}$$

and that the Hessian of L does not vanish, that is,

$$\det\left\| (-1)^{n-l-1} \frac{\partial^2 L}{\partial(D^n q_j)(D^n q_i)} \right\| = \det\left\| \frac{\partial P_j^{(l)}}{\partial(D^{2n-l-1}q_j)} \right\| \neq 0, \quad l = 0, \ldots, n-1. \tag{A6}$$

The coordinate $\eta_i^{(l)}$ from the first half of the set of independent coordinates

$$\eta_i^{(k)} = D^k q_i, \quad k = 0, \ldots, 2n-1, \tag{A7}$$

is the canonical mate of $P_i^{(l)}$. When the Hamiltonian is written

$$H = \sum_{l=0}^{n-1} P_i^{(l)} \eta_i^{(l+1)} - L(\eta_i^{(0)}, \ldots, \eta_i^{(n-1)}, \eta_i^{(n)}), \tag{A8}$$

repeated subscripts denoting summation, and the noncanonical $\eta_i^{(n)}$ is eliminated through (A3), then Hamilton's equations

$$\dot{P}_i^{(l)} = -\frac{\partial H}{\partial \eta_i^{(l)}}, \tag{A9}$$

$$\dot{\eta}_i^{(l)} = \frac{\partial H}{\partial P_i^{(l)}}, \quad l = 0, \ldots, n-1, \tag{A10}$$

are equivalent to

$$\dot{\eta}_i^{(l)} = \eta_i^{(l+1)}, \quad l = 0, \ldots, 2n-2$$

$$\dot{P}_i^{(0)}(\eta_j^{(0)}, \ldots, \eta_j^{(2n-1)}) = \frac{\partial L}{\partial \eta_i^{(0)}},$$

an alternative expression of the dynamics implied by L.

It can now easily be shown that the Lagrangian

$$L'(\eta_i^{(l)}, \dot{\eta}_i^{(l)}) = \sum_{k=0}^{n-1} P_i^{(k)} \dot{\eta}_i^{(k)} - H(P_i^{(l)}, \eta_i^{(0)}, \ldots, \eta_i^{(n-1)}), \tag{A11}$$

where H is the one in (A8), embodies yet another expression of that same dynamics. The equation of motion for each $\eta_i^{(l)}$ is the Euler—Lagrange equation; the set is

$$\dot{P}_i^{(l)} - \sum_{k=0}^{n-1} \frac{\partial P_j^{(k)}}{\partial \eta_i^{(l)}} \dot{\eta}_j^{(k)} + \sum_{k=0}^{n-1} \frac{\partial H}{\partial P_j^{(k)}} \frac{\partial P_j^{(k)}}{\partial \eta_i^{(l)}} + \frac{\partial H}{\partial \eta_i^{(l)}},$$

$$l = 0, \ldots, n-1,$$

where the derivative on the right is only taken with respect to H's explicit η dependence, and

$$-\sum_{k=0}^{n-1} \frac{\partial P_j^{(k)}}{\partial \eta_i^{(l)}} \dot{\eta}_j^{(k)} + \sum_{k=0}^{n-1} \frac{\partial H}{\partial P_j^{(k)}} \frac{\partial P_j^{(k)}}{\partial \eta_i^{(l)}} = 0, \quad l = n, \ldots, 2n-1.$$

These are

$$\sum_{k=0}^{n-1} \frac{\partial P_j^{(k)}}{\partial \eta_i^{(l)}} \left(\dot{\eta}_j^{(k)} - \frac{\partial H}{\partial P_j^{(k)}} \right) = \dot{P}_i^{(l)} + \frac{\partial H}{\partial \eta_i^{(l)}}, \quad l = 0, \ldots, n-1, \tag{A12}$$

$$\sum_{k=0}^{n-1} \frac{\partial P_j^{(k)}}{\partial \eta_i^{(l)}} \left(\dot{\eta}_j^{(k)} - \frac{\partial H}{\partial P_j^{(k)}} \right) = 0, \quad l = n, \ldots, 2n-1. \tag{A13}$$

Clearly, if the Hamiltonian equations (A9) and (A10) are satisfied, then so are these equations; but this is merely necessary. The reverse must also be true if $L'(\eta_i^{(l)}, \dot{\eta}_i^{(l)})$ is to be a Lagrangian equivalent to the original $L(q_i, \ldots, D^n q_i)$. That this is indeed the case can be seen by writing out the set (A13) in reverse order starting with the highest l, and recalling from (A5) that the derivative of $P_j^{(k)}$ with respect to $\eta_i^{(l)}$ vanishes when k exceeds $2n-l-1$. Thus,

$$\frac{\partial P_j^{(0)}}{\partial \eta_i^{(2n-1)}} \left(\dot{\eta}_j^{(0)} - \frac{\partial H}{\partial P_j^{(0)}} \right) = 0,$$

$$\frac{\partial P_j^{(0)}}{\partial \eta_i^{(2n-2)}} \left(\dot{\eta}_j^{(0)} - \frac{\partial H}{\partial P_j^{(0)}} \right) + \frac{\partial P_j^{(1)}}{\partial \eta_i^{(2n-2)}} \left(\dot{\eta}_j^{(1)} - \frac{\partial H}{\partial P_j^{(1)}} \right) = 0, \tag{A14}$$

and so on down to $l = n$. Hence, for the matrix of differential coefficients in (A13) to have an inverse it is only necessary that the submatrices which appear in the *last* term of each of the n equations of this triangular system (A14) have inverses. That is,

$$\det\left\| \frac{\partial P_j^{(l)}}{\partial \eta_i^{(2n-l-1)}} \right\| \neq 0, \quad l = 0, \ldots, n-1,$$

which is just the condition (A6). The system (A13) therefore implies (A10), and (A12) in turn implies (A9).

The result then is this: Suppose

$$L(q_j, \ldots, D^n q_j) = \sum_{l=0}^{n-1} P_i^{(l)}(q_j, \ldots, D^{2n-l-1}q_j) D^{l+i} q_i - E(q_j, \ldots, D^{2n-1}q_j)$$

implies one $2n$-order differential equation for each dimension i. It follows that

$$L'(\eta_j^{(l)}, \dot{\eta}_j^{(l)}) = \sum_{l=0}^{n-1} P_i^{(l)}(\eta_j^{(0)}, \ldots, \eta_j^{(2n-l-1)}) \dot{\eta}_i^{(l)} - E(\eta_j^{(0)}, \ldots, \eta_j^{(2n-1)}),$$

where the P's are the Ostrogradski momenta (A1) generated by L, and the η's independent coordinates given by (A7), implies the same dynamics through $2n$ first-order equations for each i.

The proof given here is quite simple owing to the assumption (A6). A look at some examples, however, will show that the conclusion can hold even if (A6) does not.

APPENDIX B

The object here is to sum the straight-line components of the series

$$\sum_{p=0}^{\infty} \frac{(-D_1 D_2)^p}{(2p)!} r^{2p-1}. \tag{B1}$$

If A is r^{2p-1} and B is r^2 in Leibniz's expansion

$$(-D_1 D_2)^{p+2} AB = (-1)^{p+2} \sum_{l=0}^{p+2} \sum_{k=0}^{p+2} \binom{p+2}{l} \binom{p+2}{k} [D_1^{p+2-k} D_2^{p+2-l} A] \times [D_1^k D_2^l B]$$

and Φ_p denotes the pth term in (B1), then

$$(2p+1)(2p+2) D_1 D_2 \Phi_{p+1} = -r^2 D_1^2 D_2^2 \Phi_p - 2(p+2)(\mathbf{r} \cdot \mathbf{v}_1) D_1 D_2^2 \Phi_p + 2(p+2)(\mathbf{r} \cdot \mathbf{v}_2) D_1^2 D_2 \Phi_p - (p+1)(p+2) v_1^2 D_2^2 \Phi_p - (p+1)(p+2) v_2^2 D_1^2 \Phi_p + 2(p+2)^2 (\mathbf{v}_1 \cdot \mathbf{v}_2) D_1 D_2 \Phi_p + \text{terms involving the higher derivatives of } \mathbf{v}_1, \mathbf{v}_2. \tag{B2}$$

Now, direct calculation of the opening terms of (B1) indicates that the straight line portion of Φ_p, which may be called ϕ_p, is given by

$$\phi_p = \frac{1}{r}\sum_{l=0}^{[p/2]} a_p(l)(\mu_1^2\mu_2^2)^l(\mu_1\cdot\mu_2)^{p-2l}, \tag{B3}$$

where $[p/2]$ has the customary meaning, $p/2$ for even p, or the lower of $p/2$'s neighboring integers if p is odd. As in the test, μ_i stands for $\hat{\mathbf{r}}\times\mathbf{v}_i$. The substitution of (B3) in (B2) demands exceedingly tedious calculations but the result confirms the conjectured form of ϕ_p, and provides a recursion relation in which $a_p(l)$ is given in terms of the a_p's with lower integers l, and the neighboring set $a_{p-1}(0)$, $a_{p-1}(1)\cdots a_{p-1}([p/2])$. Observations of the actual values of these coefficients from the first few terms in (B1) leads to the solution

$$a_p(l) = \frac{p!}{2^{(p+2l)}(l!)^2(p-2l)!}. \tag{B4}$$

Substitution of (B4) in (B3), and (B3) for the pth term of (B1) gives then

$$\frac{1}{r}\sum_{p=0}^{\infty}\sum_{l=0}^{[p/2]}\frac{p!}{2^{(p+2l)}(l!)^2(p-2l)!}(\mu_1^2\mu_2^2)^l(\mu_1\cdot\mu_2)^{p-2l}, \tag{B5}$$

so the problem reduces to determining the sum of (B5) which is quite easy to do. First (B5) may be rearranged as

$$\frac{1}{r}\sum_{p=0}^{\infty}\sum_{l=0}^{\infty}\frac{(p+2l)!}{2^{(p+4l)}(l!)^2 p!}(\mu_1^2\mu_2^2)^l(\mu_1\cdot\mu_2)^p, \tag{B6}$$

which contains standard forms for the sums on p and l. If the sum on p is taken first, (B6) becomes

$$\frac{1}{r}\sum_{l=0}^{\infty}\frac{(2l)!}{2^{4l}(l!)^2}\frac{(\mu_1^2\mu_2^2)^l}{[1-\frac{1}{2}(\mu_1\cdot\mu_2)]^{2l+1}}$$

whose sum is then

$$\frac{1}{r[1-\frac{1}{2}(\mu_1\cdot\mu_2)]}\left(1-\frac{1}{4}\frac{\mu_1^2\mu_2^2}{[1-\frac{1}{2}(\mu_1\cdot\mu_2)]^2}\right)^{-1/2}.$$

The final result may therefore be written

$$\sum_{p=0}^{\infty}\frac{1}{(2p)!}\left(-\mathbf{v}_1\cdot\frac{\partial}{\partial\mathbf{r}_1}\right)^p\left(\mathbf{v}_2\cdot\frac{\partial}{\partial\mathbf{r}_2}\right)^p r^{2p-1} = \frac{1}{[1-\mu_1\cdot\mu_2-\frac{1}{4}(\mu_1\times\mu_2)^2]^{1/2}r}. \tag{B7}$$

*Present address: MacDonald Science Library, Dalhousie University, Halifax, Nova Scotia.

[1]E.H. Kerner, J. Math. Phys. **6**, 1218 (1965). This paper together with those cited in Refs. 5, 6, 10, 11, 16 have been gathered in the reprint collection *The Theory of Action-at-a-Distance in Particle Dynamics*, edited by E.H. Kerner (Gordon and Breach, New York, 1972).

[2]F.J. Kennedy, J. Math. Phys. 10, 1349 (1969).

[3]F.J. Kennedy, Am. J. Phys. **40**, 63 (1972).

[4]Ref. 3, Sec. III.

[5]Actually the Darwin Lagrangian (whose interaction term is of order e^2 and c^{-2}—see Ref. 18) produces an additional term of order e^4 and c^{-2} in the normal-form particle equations of motion.

[6]See, for example, the introduction to the paper by R.N. Hill, J. Math. Phys. 8, 1756 (1967).

[7]D.G. Currie, T.F. Jordan, and E.C.G. Sudarshan, Rev. Mod. Phys. **35**, 350 (1963). Their proof in the case of two particles has been extended to any finite number of particles by H. Leutwyler, Nuovo Cimento 37, 556 (1965).

[8]E.H. Kerner, Ref. 1. Contrary to what seems to be popularly held, *any* even order system of equations has a Hamiltonian formulation. The assertion, called the Lie—Konig's theorem, is proved in E.T. Whittaker's *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge U.P., New York, 1960), p. 275.

[9]See Kerner, Ref. 1, Sec. II, for example, for a more extensive discussion of the method given here. A knowledge of the exact force on m_1 when m_2 is stationary is all that is needed to determine the forces of interaction between m_1 and m_2 correct to first order in the coupling constant, however, so the apparatus of the classical field theory for the interaction is not necessary to find the forces in straight line approximation. See Kennedy, Ref. 3.

[10]Dimensional analysis shows that an expansion of the forces in powers of e^2 must be an expansion in the dimensionless parameters $e^2/(m_i c^2)r$, that is the classical charge radius of m_i divided by the separation of the two charges.

[11]J.A. Wheeler and R.P. Feynmen, Rev. Mod. Phys. **17**, 157 (1945); **21**, 425 (1949). See also Kennedy, Ref. 2, Sec. II.

[12]E.H. Kerner, J. Math. Phys. **3**, 35 (1962).

[13]F.J. Kennedy, Ref. 2, Sec. III. In this paper the results of generalized mechanics do not appear explicitly. Instead, formulas equivalent to (6) are developed from (4) via the infinite-dimensional Noether's theorem, but the approaches are really identical.

[14]E.H. Kerner, Ref. 1; also Bull. Math. Biophys. **26**, 333 (1964).

[15]F.J. Kennedy, Ref. 2, Sec. III.

[16]Hill, in Ref. 6, Sec. II, describes the method and gives further references.

[17]The extension to N particles is given in (66). See Hill, Ref. 6, Sec. IV, Part B. Also D.G. Currie, Phys. Rev. **142**, 817 (1966).

[18]See, for example, L.D. Landau and E.M. Lifshitz, *The Classical Theory of Fields*, translated by M. Hamermesh (Addison-Wesley, Reading, Mass., 1971), 3rd ed., Sec. 65, p. 165. The original work is C.G. Darwin's which was reported in the Phil. Mag. and J. Sci. **39**, 537 (1920).

[19]V.N. Golubenkov and Ia.A. Smorodinskii, Sov. Phys. JETP **4**, 442 (1957). Their result is derived by Landau and Lifshitz, Ref. 18, p. 207, and is mentioned by W.R. Davis in his *Classical Fields, Particles and the Theory of Relativity* (Gordon and Breach, New York, 1970), p. 215. That the c^{-4} "Lagrangian" does not imply the equations of motion was first noticed by Kerner, Ref. 1, footnote 9.

[20]The "free-particle-like formulation of Newtonian instantaneous action-at-a-distance electrodynamics" is discussed by D. Hirondel in J. Math. Phys. **15**, 1689 (1974).

[21]See, for instance, Whittaker, Ref. 8, pp. 266—7.

On properties of the singularity of the ground state in certain classical Heisenberg models

Walter F. Wreszinski

Instituto de Física, Universidade de São Paulo, São Paulo, Brazil
(Received 25 March 1975)

We prove that for both the classical ferro- and antiferromagnetic Heisenberg models the infinite volume limit of the ground state energy per unit volume of the system (Hamiltonian plus λ times an operator) is not differentiable at zero in λ for some operators. This characterization of the singularity at $T=0$, which corresponds to Fisher's for positive temperature, adds to a number of others, which are to some extent analogous to the several characterizations of phase transitions at $T>0$. A comment is made upon a related open problem concerning the ground state of the quantum antiferromagnetic Heisenberg chain.

As far as we know no explicit examples of Fisher's[1] characterization of a phase transition in terms of the nondifferentiability of certain infinite volume correlation functions with respect to external parameters exist. In this note we study the analogous characterization for $T=0$, in the case of some classical Heisenberg (anti-) ferromagnets[2] and prove that it holds. This result on ground states of classical systems (for general ground-state representations, see Ref. 5) adds to some other features of the singularity at $T=0$, known for the one-dimensional chain with nearest-neighbor interactions, namely divergence of the susceptibility χ_β as β^2 as $\beta\to\infty$,[6] existence of long-range order,[7] (infinite) asymptotic degeneracy of the highest eigenvalue of the transfer matrix as $\beta\to\infty$,[7,8] which are to some extent analogous to some of the several alternative determinations of a phase transition at $T>0$ (see, e.g., Ref. 9). To display one more property in this set of alternate descriptions, whose interrelation is not entirely clear, and the clarification of which is a major problem in the theory of phase transitions, is the motivation of this paper. For notational simplicity, we write out the proof for the one-dimensional case and nearest neighbor interactions. However, the result and proof of the forthcoming theorem hold in any number of dimensions, with a Hamiltonian for the region $\Lambda\subset\mathbb{Z}^\nu$ (ν arbitrary integer) given by

$$H_\Lambda=-\sum_{i,j\in\Lambda}J(i-j)\,\mathbf{t}_i\cdot\mathbf{t}_j \tag{1}$$

where $\mathbf{t}_i$, $i\in\Lambda$, are unit vectors, $\sum_{i\in\Lambda}|J(i)|<\infty$ by stability[10] such that Λ may be divided into two "sublattices" A and B ($A\cup B=\Lambda$), with $J(i-j)\le 0$ if i,j both belong to either A or B and $J(i-j)\ge 0$ if $i\in A$ and $j\in B$ or vice versa. If A is the set of nearest neighbors of B, the above conditions correspond to antiferromagnetism, and if A or B are empty, we have a ferromagnetic system (see Ref. 11).

Let $\underline{t}_j$, $i\in[0,N-1]$, be vectors in $S=\{\mathbf{x}\in\mathbb{R}^3;\mathbf{x}^2=1\}$, with $\underline{t}_0=\underline{t}_N$ (periodic boundary conditions), with components $[\Omega_i\equiv(\theta_i,\phi_i)]$

$$t_i^1(\Omega_i)=t_i^1=\sin\theta_i\cos\phi_i,$$

$$t_i^2(\Omega_i)=t_i^2=\sin\theta_i\sin\phi_i,$$

$$t_i^3(\Omega_i)=t_i^3=\cos\theta_i.\quad(0\le\theta_i<\pi,\ 0\le\phi_i<2\pi)$$

On $\mathcal{H}=\otimes_{i=0}^{N-1}L^2(S,d\Omega_i)$, with $d\Omega_i=\sin\theta_i\,d\theta_i\,d\phi_i$, let

$$C_N^+(r,\Omega)=\sum_{i=0}^{N-1}t_i^3t_{i+2r}^3 \tag{2a}$$

$$C_N^-(r,\Omega)=\sum_{i=0}^{N-1}t_i^3t_{i+r}^3 \tag{2b}$$

$$H_N^{(\pm)}(\Omega)=\pm 2J\sum_{i=0}^{N-1}\underline{t}_{j+1},\quad \underline{t}_0=\underline{t}_N,\ J>0 \tag{3}$$

with + (resp. −) corresponding to antiferro- (resp. ferro-) magnetism. The precise way in which a large class of classical spin systems [including (3) and the more general Hamiltonian (1)] is the limit, "as the spin tends to infinity," of the corresponding quantum spin systems is described in Ref. 12.

Let

$$g_{N,r}^{(\pm)}(\lambda)=\min_{\Omega\in S^N}\frac{1}{N}\left[H_N^{(\pm)}(\Omega)+\lambda C_N^{(\pm)}(\Omega)\right] \tag{4}$$

and let

$$g_r^{(\pm)}(\lambda)=\lim_{N\to\infty}g_{N,r}^{(\pm)}(\lambda). \tag{5}$$

For fixed r, this limit exists by a simple adaptation of the proof in Ref. 5.

Note that $g_{N,r}^{(\pm)}$ is a concave function of λ. Hence $g_r^{(\pm)}$ is also a concave function of λ, whence (e.g., Ref. 13) it has both a right-hand derivative $d^+g_r^{(\pm)}(\lambda)/d\lambda$ and a left-hand one $d^-g_r^{(\pm)}(\lambda)/d\lambda$.

Theorem: $[d^+g_r^{(\pm)}(\lambda)/d\lambda]\ (\lambda=0)\neq[d^-g_r^{(\pm)}(\lambda)/d\lambda]\ (\lambda=0)$.

Proof: A possible choice for our "sublattices" is $A=\{0,2,\ldots,N\}$ and $\{B=1,3,\ldots,(N-1)\}$ if N is even, or $A=\{0,2,\ldots,(N-1)\}$ and $B=\{1,3,\ldots,N\}$ if N is odd. Clearly, for all N,

$$g_{N,r}^{(\pm)}(0)=-2J.$$

Now consider the state given by $t_i^2=t_i^3=0$, $\forall i\in[0,N-1]$, and $t_i^1=+1$, $\forall i\in[0,N-1]$, in the − (ferromagnetic) case, and $t_i^1=1$, $\forall i\in A$, and $t_i^1=-1$, $\forall i\in B$, in the + (antiferromagnetic) case. In this state and any λ, $(1/N)[H_N^{(\pm)}(\Omega)+\lambda C_N^{(\pm)}(r,\Omega)]$ takes the value $(-2J)$; hence $g_{N,r}^{(\pm)}(\lambda)\le-2J$ for any λ and for all N. Hence,

$$(1/\lambda)[g_{N,r}^{\pm}(\lambda)-g_{N,r}^{\pm}(0)]\le 0\quad\forall\lambda>0,\ \forall N.$$

Hence, $(1/\lambda)[g_r^{\pm}(\lambda)-g_r^{\pm}(0)]\le 0$, $\forall\lambda>0$, from which we get

$$\frac{d^+ g_r^{\pm}(\lambda)}{d\lambda}\,(\lambda=0) \leq 0. \tag{6}$$

Now,

$$\frac{d^- g_r^{(\pm)}(\lambda)}{d\lambda}\,(\lambda=0)$$

$$= \lim_{\lambda \to 0} \frac{[g_r^{(\pm)}(0) - g_r^{(\pm)}(-\lambda)]}{\lambda}$$

$$= \lim_{\lambda \to 0_+} \frac{1}{\lambda} \left[\lim_{N\to\infty} \min_{\Omega \in \mathcal{S}^N} \left(\frac{1}{N} H_N^{(\pm)}(\Omega) \right) \right.$$

$$\left. - \lim_{N\to\infty} \min_{\Omega \in \mathcal{S}^N} \left(\frac{1}{N} [H_N^{(\pm)}(\Omega) - \lambda C_N^{(\pm)}(r, \Omega)] \right) \right] \tag{7}$$

Now, we clearly have

$$\min_{\Omega \in \mathcal{S}^N} \frac{1}{N} [H_N^{(\pm)}(\Omega)] = -1 \tag{8}$$

while

$$(1/N)[H_N^{(\pm)}(\Omega) - \lambda C_N^{(\pm)}(r, \Omega)] \geq -1-\lambda,$$

and this minimum value is attained, e.g., for $t_i^1 = t_i^2 = 0 \ \forall i \in [0, N-1]$, and $t_i^3 = 1 \ \forall i \in [0, N-1]$ in the $-$ case, and $t_i^3 = 1 \ \forall i \in A$, and $t_i^3 = -1 \ \forall i \in B$, in the $+$ case. Hence, we have

$$\min_{\Omega \in \mathcal{S}^N} \frac{1}{N} [H_N^{(\pm)}(\Omega) - \lambda C_N^{(\pm)}(r, \Omega)] = -1-\lambda. \tag{9}$$

By (8) and (9) in (7) we get immediately

$$\frac{d^- g_r^{(\pm)}(\lambda)}{d\lambda}\,(\lambda=0) = \lim_{\lambda \to 0} \frac{1}{\lambda} [-1-(-1-\lambda)] = 1. \tag{10}$$

The results (6) and (10) imply the assertion of the theorem. ■

Remark: Consider the one-dimensional isotropic antiferromagnetic Heisenberg chain for spin S and periodic boundary conditions, described by the Hamiltonian (on $\mathcal{H} = \otimes_{i=0}^{N-1} \mathbb{C}_i^{2S+1}$,

$$H_N^S \equiv \frac{2J}{S^2} \sum_{i=0}^{N-1} (\underline{S}_i \cdot \underline{S}_{i+1}), \quad \underline{S}_0 = \underline{S}_N, \quad S_i^{(k)} = \tfrac{1}{2}\sigma_i^{(k)},$$

$\sigma_i^{(k)}$, $k \in [1, 3]$, being spin matrices for spin S, and $J > 0$. The ground state Ω_N^S of H_N^S is unique,[11] and we define

$$L_s \equiv \liminf_{r\to\infty} \liminf_{N\to\infty} (\Omega_N^s, S_0^3 S_{2r}^3 \Omega_N^s)/S^2.$$

If $L_s \geq \gamma > 0$, we may take this to mean that the one-dimensional antiferromagnet exhibits "long-range order" in the ground state.

Define

$$C_N(r) \equiv \sum_{i=0}^{N-1} S_i^3 S_{i+2r}^3, \quad \underline{S}_0 = \underline{S}_N,$$

$$C_N(\delta, r, \Omega) \equiv \delta^2 C_N^{(+)}(r, \Omega),$$

$$H_N(\delta, \Omega) \equiv \delta^2 H_N^{(+)}(\Omega),$$

$$f_{N,r}(\lambda, \delta)$$

$$\equiv \min_{\Omega \in \mathcal{S}^N} \{(1/N)[H_N(\delta, \Omega) + \lambda C_N(\delta, r, \Omega)]\}$$

$$- \min_{\Omega \in \mathcal{S}^N} (1/N) H_N(1, \Omega).$$

Now $f_{N,r}(\lambda, \delta)$ is a double sequence of functions concave in λ. Hence

$$f(\lambda, \delta) \equiv \liminf_{r\to\infty} \liminf_{N\to\infty} f_{N,r}(\lambda, \delta)$$

is also concave in λ. Now, $f_{N,r}(\lambda, \delta)$ is uniformly continuous (in (r, N) as a function of δ. Hence,

$$f(\lambda) = f(\lambda) - f(0) \equiv \lim_{\delta \to 1} f(\lambda, \delta) = \liminf_{r\to\infty} \liminf_{N\to\infty} f_{N,r}(\lambda, 1)$$

$$= \liminf_{r\to\infty} \liminf_{N\to\infty} [f_{N,r}(\lambda, 1) - f_{N,r}(0, 1)].$$

From the rhs of inequality (6.5) of Ref. 12, transcribed to vacuum expectation values, we easily get

$$\liminf_{S\to\infty} \liminf_{r\to\infty} \liminf_{N\to\infty} \left(\Omega_N^s, \frac{C_N(r)}{N} \Omega_N^s \right) \Big/ S^2$$

$$\geq \lambda^{-1} f(\lambda) = \liminf_{r\to\infty} \liminf_{N\to\infty} \left\{ \lambda^{-1} f_{N,r}(\lambda, 1) = \liminf_{r\to\infty} \liminf_{N\to\infty} \right\}$$

$$\times \lim_{\beta\to\infty} \lambda^{-1} f_{N,r,\beta}(\lambda, 1), \tag{11}$$

where

$$f_{N,r,\beta}(\lambda, 1) \equiv -\frac{1}{N\beta}$$

$$\times \log \frac{\int d\Omega^N \{\exp -\beta[H_N(1, \Omega) + \lambda C_N(1, r, \Omega)]\}}{\int d\Omega^N \exp[-\beta H_N(1, \Omega)]}.$$

If

$$\lim_{\lambda\to 0_+} \liminf_{r\to\infty} \liminf_{N\to\infty} \lim_{\beta\to\infty} \lambda^{-1} f_{N,r,\beta}(\lambda, 1)$$

$$\overset{(?)}{=} \liminf_{r\to\infty} \liminf_{N\to\infty} \lim_{\beta\to\infty} \lim_{\lambda\to 0_+} \lambda^{-1} f_{N,r,\beta}(\lambda, 1) = \tfrac{1}{3}$$

by Refs. 6, 7, which would follow if, e.g., for λ in a sufficiently small neighborhood of zero one had

$$[\beta \langle (C_N(1, r, \Omega) - \langle C_N(1, r, \Omega) \rangle_\lambda)^2 \rangle_\lambda]/N \geq \text{const}$$

(independent of r, N, β), (13)

where

$$\langle A_N(r, \Omega) \rangle_\lambda$$

$$\equiv \frac{\int d\Omega^N \{\exp -\beta[H_N(1, \Omega) + \lambda C_N(1, r, \Omega)\} A_N(r, \Omega)}{\int d\Omega^N \{\exp -\beta[H_N(1, \Omega) + \lambda C_N(1, r, \Omega)]\}},$$

then we would clearly have $L_s \geq \gamma > 0$ for sufficiently large S, on putting (12) into (11). Unfortunately, we have have been unable to prove (12) [or (13)] to date. ■

ACKNOWLEDGMENTS

I am greatly indebted to Professor E. H. Lieb for his encouragement, patience, and various remarks and corrections in a correspondence about the problem described in the Remark. I am also indebted to P. A. Vuillermot for useful conversations relating to the existence of the limit (5). Finally, I should like to thank the referee for several extremely helpful suggestions, which led to a much simpler proof of the main result than my original one.

[1]M.E. Fisher, J. Math. Phys. **6**, 1643 (1965).
[2]Some references on classical spin systems, to which we refer for additional literature, are Refs. 3, 4.
[3]P.A. Vuillermot and M.V. Romerio, Comm. Math. Phys. (to appear) and work in preparation.
[4]A. Bortz and R.B. Griffiths, Comm. Math. Phys. **26**, 102 (1972).
[5]D. Ruelle, Comm. Math. Phys. **11**, 339 (1969).
[6]M.E. Fisher, Am. Jour. Phys. **32**, 343 (1964).
[7]G.S. Joyce, Phys. Rev. **155**, 458 (1967).
[8]The eigenvalues $\{\lambda_l(\beta)\}$, $l=0,1,2,\cdots$, of the transfer matrix, of which the largest $\lambda_0(\beta)$ is simple and the remainder ones are $(2l+1)$-fold degenerate, become all degenerate with the largest eigenvalue for l odd, i.e., $\lim_{\beta\to\infty}[\lambda_l(\beta)/\lambda_0(\beta)]=(-1)^l$.[7]
[9]M. Marinaro and G. L. Sewell, Comm. Math. Phys. **24**, 310 (1972).
[10]D. Ruelle, *Statistical Mechanics—Rigorous Results* (Benjamin, New York, 1969).
[11]E.H. Lieb and D.C. Mattis, J. Math. Phys. **3**, 749 (1962).
[12]E.H. Lieb, Comm. Math. Phys. **31**, 327 (1973).
[13]Hardy, Littlewood, and Pólya, *Inequalities* (Cambridge U.P., Cambridge, 1934).

A toroidal solution of the vacuum Einstein field equations*

Kip S. Thorne

Kellogg Radiation Laboratory, California Institute of Technology, Pasadena, California 91125
(Received 6 May 1975)

This paper presents a new solution to the vacuum Einstein field equations for the static external gravitational field of a toroidal singularity. This solution is unique among known toroidal solutions in that the singularity is locally cylindrically symmetric; near it the spacetime geometry becomes that of an infinite line mass ("Levi-Civita metric").

I. INTRODUCTION

Infinitely long, cylindrically symmetric systems have played a useful role, since 1919, as tools for gaining insight into general relativistic phenomena. For example, much of the pioneering work on gravitational-wave theory dealt with cylindrical systems;[1] and in recent years cylindrical systems have been used as a testing ground for ideas about highly nonspherical gravitational collapse.[2]

A key difficulty with all cylindrical analyses is the fact that spacetime is not asymptotically Minkowskiian far outside a cylindrical system: Just as the Newtonian potential of a cylinder diverges logarithmically at large radii ($\Phi = \text{const} \times \ln r$), so its general relativistic analog, $\Psi \equiv \frac{1}{2} \ln |g_{00}|$, diverges logarithmically. As a result, the physical interpretation of cylindrical spacetimes is often uncertain.

One way to remedy this problem is to deal with systems that are locally cylindrical, but are confined to a finite region of space—e.g., needles (finite cylinders) and thin rings (toruses). Unfortunately, such systems are far more difficult to analyze than are infinitely long cylinders. The purpose of this paper is to present a tool that may be helpful in future analyses of bounded, locally cylindrical systems. That tool is a static, two-parameter solution of the vacuum Einstein field equations representing the external gravitational field of a torus. Unlike other toroidal solutions, very near the ring singularity this one is cylindrically symmetric.

II. THE SOLUTION IN GENERAL

A. The Weyl formalism

In presenting the new solution, we shall use Weyl's formalism[3] for axially symmetric, vacuum solutions of the Einstein field equations. The Weyl formalism is couched in the mathematical language of a flat "background space" with cylindrical coordinates (ρ, z, ϕ) and with metric

$$d\sigma^2 = d\rho^2 + dz^2 + \rho^2 d\phi^2. \tag{1}$$

Two gravitational potentials with axial symmetry reside in the background space: $\psi(\rho, z)$ and $\gamma(\rho, z)$. They satisfy the field equations

$$\psi_{,\rho\rho} + \rho^{-1}\psi_{,\rho} + \psi_{,zz} = 0, \tag{2a}$$

$$\gamma_{,\rho} = \rho(\psi_{,\rho}{}^2 - \psi_{,z}{}^2), \tag{2b}$$

$$\gamma_{,z} = 2\rho\psi_{,\rho}\psi_{,z}, \tag{2c}$$

where commas denote partial derivatives. It is often useful to rewrite these field equations in terms of the gradient operator ∇ and Laplacian ∇^2 of the flat background space (1):

$$\nabla = \mathbf{e}_\rho \frac{\partial}{\partial\rho} + \mathbf{e}_z \frac{\partial}{\partial z}, \quad \nabla^2 = \rho^{-1}\frac{\partial}{\partial\rho}\frac{\rho\partial}{\partial\rho} + \frac{\partial^2}{\partial z^2}, \tag{2a'}$$

$$\nabla^2\psi = 0, \quad |\nabla\gamma| = \rho(\nabla\psi)^2, \tag{2b'}$$

If $\nabla\psi$ makes an angle θ_0 with the radial ($\mathbf{e}_\rho$) direction, then $\nabla\gamma$ makes an angle $2\theta_0$ with the radial direction (2c')

Corresponding to any solution of the field equations (2) or (2') in the flat background space (1), there exists a static, axially symmetric solution of the vacuum Einstein field equations with the metric

$$ds^2 = -\exp(2\psi)\,dt^2 + \exp[2(\gamma - \psi)](d\rho^2 + dz^2) + \rho^2\exp(-2\psi)\,d\phi^2. \tag{3}$$

Different solutions are obtained by choosing different singular souces for ψ in the background space (point sources, line sources, surface sources). If the sources are confined to a finite region of the background space, then both ψ and γ will approach constants as $(\rho^2 + z^2)^{1/2} \to \infty$; those constants can be chosen zero without loss of generality, and the resulting physical spacetime (3) is asymptotically Minkowskiian.

B. Toroidal solutions that are not locally cylindrical

The easiest way to construct toroidal solutions is to choose, as the source of ψ in the background space, a singularity at $\rho = b$, $z = 0$ (ring singularity around axis of symmetry). The simplest ring singularity is a pure "line monopole," for which[4]

$$\psi = \text{const} \times \ln[(\rho - b)^2 + z^2]^{1/2} \text{ near singularity, i.e., at } [(\rho - b)^2 + z^2]^{1/2} \ll b. \tag{4}$$

Unfortunately, when ψ has this locally cylindrical form, γ and the physical metric are not locally cylindrical near the singularity; Eq. (2c') forbids it. One cannot remedy this problem by any other choice for the ring source of ψ (any superposition of line multipoles at $\rho = b$, $z = 0$).[5]

This situation is analogous to the case of spherical symmetry: No type of point singularity in the background space (no superposition of point multipoles) can lead to a spherical physical metric; Eq. (2c′) forbids it. To get a spherical metric (the Schwarzschild solution), one must choose as the source of ψ a *"line* mass" on the axis of symmetry, with "mass per unit length" $\frac{1}{2}$ (so $\psi = \frac{1}{2}\ln\rho$ near it), and with finite length $\Delta z = 2M$ = ("Schwarzschild radius").[6]

C. The potentials ψ and γ for the new toroidal solution

It turns out that the background-space source for a locally cylindrical, globally toroidal metric is even more peculiar than that for the Schwarzschild solution. The desired source is best understood by thinking of the background space as filled with an incompressible fluid that undergoes steady-state potential flow with potential ψ and with momentum density $\rho_0 \mathbf{v} = \nabla\psi$ (ρ_0, not to be confused with ρ, is the mass density of the fluid). The fluid is created in a line singularity on the axis of symmetry (Fig. 1), and flows outward from there. The singularity has a finite height, $z = 2a$; and it pours out fluid at a constant rate $\dot{m}$. Once created, the fluid does not freely expand into the background space. Rather, its flow is constrained by two solid disks that are attached to the ends of the source ($z = \pm a$) and that have radii b (Fig. 1).

By the time the flowing fluid gets far from the constraining disks, $r \equiv (\rho^2 + z^2)^{1/2} \gg b$, its flow has become nearly spherical with mass flow rate

$$\dot{m} = 4\pi r^2 \rho_0 v^r = 4\pi r^2 \psi_{,r} \tag{5}$$

and potentials

$$\psi = -(\dot{m}/4\pi) r^{-1}, \quad \gamma = O(\dot{m}^2 r^{-2}). \tag{6}$$

Hence, the physical spacetime metric (3) has the asymptotic form

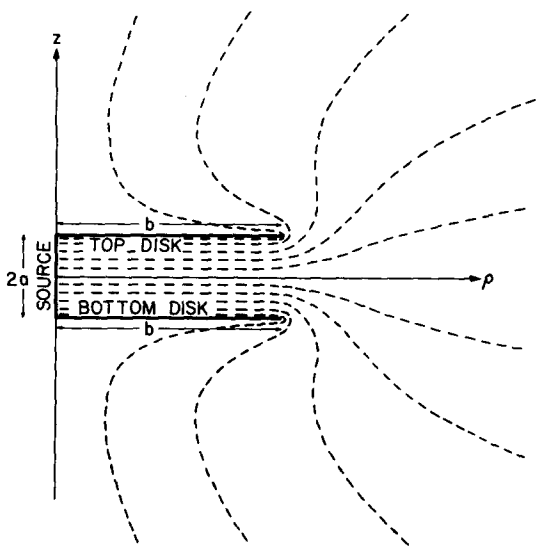


FIG. 1. The flow of fluid in the (fictitious) background space. The flow lines (trajectories of $\nabla\psi$) are shown dashed.

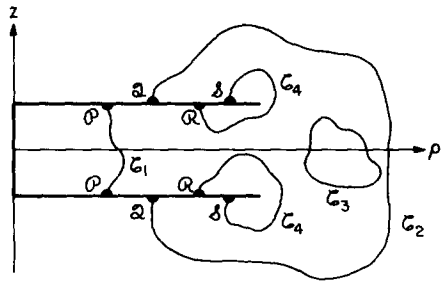


FIG. 2. The toroidal topology of physical spacetime. The events P and R are located, in coordinate space, on the inside faces of both the upper disk and the lower disk; Q and S are located on both outside faces. The closed curves C_1 and C_2 are topologically linked through the ring singularity; the singularity prevents them from being contracted to a point. The closed curves C_3 C_4 do not link the singularity; they can be contracted to a point.

$$ds^2 = -[1 - (\dot{m}/2\pi)r^{-1}]\,dt^2 + [1 + (\dot{m}/2\pi)r^{-1}] \times (d\rho^2 + dz^2 + \rho^2\,d\phi^2), \tag{7}$$

from which we can read off the total mass-energy M of the gravitating system in terms of the mass flow rate $\dot{m}$ in the (fictitious) background metric:

$$M = \dot{m}/4\pi. \tag{8}$$

Near the singular source the flow is in the $\mathbf{e}_\rho$ direction (see Fig. 1), with

$$4\pi M = \dot{m} = (2a)(2\pi\rho)\psi_{,\rho} \tag{9}$$

and thus with

$$\psi = (M/a)\ln\rho + \text{const} \quad \text{at} \quad |z| < a, \quad \rho \ll \max(a, b). \tag{10a}$$

The solution for ψ can be summarized mathematically as follows: (i) ψ has the asymptotic form (10a) near the singularity; (ii) ψ satisfies the boundary conditions

$$\psi_{,z} = 0 \quad \text{at} \quad z = \pm a, \quad \text{for } 0 < \rho < b \tag{10b}$$

("fluid flow constrained by disk"); (iii) everywhere ψ satisfies

$$\nabla^2\psi = 0 \tag{10c}$$

("potential flow"); (iv) ψ vanishes at spatial infinity

$$\psi = -M/r \quad \text{as} \quad r = (\rho^2 + z^2)^{1/2} \to \infty. \tag{10d}$$

The corresponding solution for γ can be summarized by: (v) γ satisfies Eqs. (2b, c) everywhere; and (vi) γ vanishes at spatial infinity.

D. Topology of the new solution

The above discussion fixes the metric coefficients of physical spacetime [Eq. (3)] but does not determine the topology. The topology is fixed by two identifications: (i) the outside face of the upper disk consists of the same events as the outside face of the lower disk:

$$\lim_{\epsilon\to 0} (t, \rho, z = a + \epsilon, \phi) \text{ is same event as } \lim_{\epsilon\to 0} (t, \rho, z = -a - \epsilon, \phi) \quad \text{if } 0 \leqslant \rho \leqslant b; \tag{11a}$$

(ii) the inside face of the upper disk consists of the same events as the inside face of the lower disk:

$$\lim_{\epsilon\to 0}(t,\rho,z=a-\epsilon,\phi) \text{ is same event as } \lim_{\epsilon\to 0}(t,\rho,z=-a+\epsilon,\phi) \quad \text{if } 0<\rho\le b. \tag{11b}$$

These two identifications endow the singularity of physical spacetime with a toroidal topology; see Fig. 2.

E. Local cylindrical symmetry near the singularity

Near the singularity, i.e., for $\rho \ll \max(a,b)$, ψ and γ have the form

$$\psi=(M/a)\ln\rho+\psi_0,$$

$$\gamma=(M/a)^2\ln\rho+\gamma_0,\ \psi_0 \text{ and } \gamma_0 \text{ constant;} \tag{12}$$

cf. Eqs. (10a) and (2b). The corresponding spacetime metric (3) is

$$ds^2=-\exp(2\psi_0)\rho^{2M/a}dt^2 + \exp(-2\psi_0)[\exp(2\gamma_0)\rho^{2(M/a)(M/a-1)}(d\rho^2+dz^2) + \rho^{2-2M/a}d\phi^2]. \tag{13a}$$

In this region of spacetime, z is a periodic coordinate that encircles the singularity

$$-a\le z\le +a,\quad z=-a \text{ is same set of events as } z=+a, \tag{13b}$$

and ϕ is a "longitudinal coordinate" stretching along the singularity. Since the metric coefficients depend only on the radial coordinate ρ, the geometry is cylindrically symmetric. In fact, except for topological closure of the ring (periodicity of longitudinal coordinate ϕ), the spacetime geometry (13) is that of an infinitely long, cylindrically symmetric line mass (Levi-Civita's[7] solution of the Einstein field equations).

F. Free parameters in the solution

At first sight there are three free parameters in the solution: M, b, and a. However, for arbitrary choices of M, b, a there exists a singularity at the common edge of the disks ($\rho=b$, $z=\pm a$). One can see this as follows: The field equation (10c) and boundary condition (10b) guarantee that near ($\rho=b$, $z=\pm a$) ψ has the form

$$\psi=A+B\,\bar r^{1/2}\cos\bar\theta/2, \tag{14a}$$

where $\bar r$ and $\bar\theta$ are polar coordinates centered on the edge of the disks (Fig. 3):

$$\bar r=[(\rho-b)^2+(z-a)^2]^{1/2},\quad \bar\theta=\tan^{-1}[(a-z)/(b-\rho)] \text{ near } (b,a); \qquad \bar r=[(\rho-b)^2+(z+a)^2]^{1/2},\quad \bar\theta=\tan^{-1}[(-a-z)/(b-\rho)]+2\pi \text{ near } (b,-a). \tag{15}$$

The form of γ near the edge of the disks, as fixed by Eqs. (14a) and (2b′, c′), is

$$\gamma=C-\tfrac{1}{4}B^2 b\ln\bar r. \tag{14b}$$

The constants A, B, C are unique functions of M, b, a—functions which one can determine by fully solving Eqs. (10) and (2). Expressions (14) and (15), when inserted into the physical metric (3), yield

$$ds^2=-\exp(2A)[1+2B\bar r^{1/2}\cos(\bar\theta/2)]dt^2+\exp(-2A)\times[1-2B\bar r^{1/2}\cos(\bar\theta/2)]\times[\exp(2C)\bar r^{-B^2b/2}(d\bar r^2+\bar r^2 d\bar\theta^2)+(b-\bar r\cos\bar\theta)^2 d\phi^2]. \tag{16}$$

This metric with its square roots and half angles is ugly. However, the coordinate transformation

$$R=\bar r^{1/2},\quad \Theta=\bar\theta/2 \quad (\text{so } \Theta \text{ runs from 0 to } 2\pi) \tag{17}$$

brings it into the nicer form

$$ds^2=-\exp(2A)[1+2BR\cos\Theta]dt^2+\exp(-2A)[1-2BR\cos\Theta]\times[4\exp(2C)R^{2-B^2b}(dR^2+R^2 d\Theta^2)+(b-R^2\cos2\Theta)^2 d\phi^2]. \tag{18}$$

The spacetime geometry described by this metric is perfectly well behaved if $B^2=2/b$; otherwise it possesses a physical singularity at $R=\bar r=0$—i.e., on the edge of the disks.

Thus, by demanding that spacetime be nonsingular at the common edge of the disks, we impose the constraint

$$[B(M,b,a)]^2=2/b \tag{19}$$

and thereby reduce the number of free parameters from 3 to 2. It is easy to verify that in this case spacetime is completely free of singularities, except for the locally cylindrical ring source at $\rho=0$, $|z|<a$.

III. THE SPECIAL CASE OF A THIN-RING TORUS

We now specialize our solution to the case

$$b\gg a \tag{20}$$

i.e., (radius of constraining disks in background space) $\gg$ (separation between disks). The spacetime geometry in this special case will turn out to be that of a thin-ring torus with (total mass-energy) $=M\ll$ (radius of ring) $=b$; see Sec. IVA, below.

In this special case we shall solve explicitly but approximately for the metric coefficients. The errors in our solution will vanish in the limit $a/b\to 0$. Our solution will have different forms in three different regions (see Fig. 4):

Region I: $[(\rho-b)^2+z^2]^{1/2}\gtrsim(ab)^{1/2}$ always, and $|z|>a$ when $\rho<b$, (21a)

Region II: $[(\rho-b)^2+z^2]^{1/2}\lesssim(ab)^{1/2}$ (21b)

Region III: $|z|<a$, $(b-\rho)\gtrsim(ab)^{1/2}$. (21c)

Note that Regions I and II overlap and Regions II and III overlap.

A. Region I

Region I is the "external region" that lies outside the constraining disks and is bounded away from their edges.

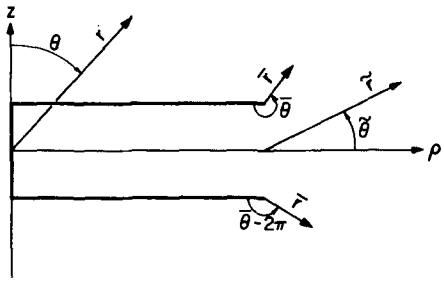


FIG. 3. Various coordinate systems used in the background space. Note that

$$\rho = r\sin\theta = b + \tilde{r}\cos\tilde{\theta} = b - \bar{r}\cos\bar{\theta},$$

$$z = r\cos\theta = \tilde{r}\sin\tilde{\theta} = \begin{cases} a - \bar{r}\sin\bar{\theta} & \text{for } z>0,\ 0<\bar{\theta}<2\pi, \\ -a - \bar{r}\sin\bar{\theta} & \text{for } z<0,\ 2\pi<\bar{\theta}<4\pi. \end{cases}$$

In solving for ψ and γ here, we pretend that the disks are fitted tightly together so that the "fluid" in the background space emerges from a ring singularity at $\rho=b$, $z=0$. This approximation produces

$$\begin{aligned} &(\text{fractional errors in } \psi) \lesssim a/(ab)^{1/2} = (a/b)^{1/2}, \\ &(\text{fractional errors in } \gamma) \lesssim (a/b)^{1/2}[\ln(a/b)^{1/2}]^2. \end{aligned} \tag{22}$$

The solution to the potential-flow equation $\nabla^2\psi=0$ with a ring source at $\rho=b$, $z=0$ and with asymptotic form (10d) is

$$\psi = \frac{-(2/\pi)M}{[(\rho+b)^2+z^2]^{1/2}}K(k), \quad k \equiv \left(\frac{4b\rho}{(\rho+b)^2+z^2}\right)^{1/2}, \tag{23a}$$

$$\approx -\frac{M}{r}\left[1 + O\left(\frac{b^2}{r^2}\right)\right] \quad \text{if } r \equiv (\rho^2+z^2)^{1/2} \gg b \tag{23b}$$

$$\approx \frac{-M}{\pi b}\ln\left(\frac{8b}{\tilde{r}}\right)\left[1 + O\left(\frac{\tilde{r}}{b}\right)\right] \quad \text{if } \tilde{r} \equiv [(\rho-b)^2+z^2]^{1/2} \ll b. \tag{23c}$$

Here $K(k)$ is the complete elliptic function. The corresponding solution to Eq. (2) for γ is[9]

$$\gamma = \frac{M^2k^4}{4\pi^2 b\rho}\left[-K^2 + 4(1-k^2)K\dot{K} + 4k^2(1-k^2)\dot{K}^2\right] + \frac{M^2k^4}{4\pi^2b^2}\left[-K^2 + 4(1-k^2)K\dot{K} - 4(1-k^2)(2-k^2)\dot{K}^2\right],$$

$$\dot{K} \equiv dK/dk^2, \tag{24a}$$

$$\approx -\tfrac{1}{2}M^2\left[\frac{\sin^2\theta}{r^2} + O\left(\frac{b^2}{r^4}\right)\right] \quad \text{if } r \gg b \tag{24b}$$

$$\approx -\frac{M^2}{\pi^2 b}\left\{\frac{\cos\tilde{\theta}}{\tilde{r}} + O\left[\frac{1}{b}\left(\ln\frac{b}{\tilde{r}}\right)^2\right]\right\} \quad \text{if } \tilde{r} \ll b. \tag{24c}$$

The coordinates (r, θ) used near infinity and $(\tilde{r}, \tilde{\theta})$ used near the ring are shown in Fig. 3. The metric is obtained by inserting expressions (23) and (24) into Eq. (3).

B. Region II

Region II is the "intermediate region" near the common edge of the constraining disks. When solving for ψ and γ in Region II we shall pretend that the edges of the disks in background space are straight rather than curved; i.e., we shall replace the axially-symmetric potential-flow equation $\psi_{,\rho\rho} + \rho^{-1}\psi_{,\rho} + \psi_{,zz} = 0$ by the plane-symmetric potential-flow equation

$$\psi_{,\rho\rho} + \psi_{,zz} = 0; \tag{25a}$$

and we shall set $\rho=b$ in the derivatives of γ:

$$\gamma_{,\rho} = b(\psi_{,\rho}{}^2 - \psi_{,z}{}^2), \quad \gamma_{,z} = 2b\psi_{,\rho}\psi_{,z}. \tag{25b}$$

In doing so we make

$$\begin{aligned} &(\text{fractional errors in } \psi) \lesssim (ab)^{1/2}/b = (a/b)^{1/2}, \\ &(\text{fractional errors in } \gamma) \lesssim (a/b)^{1/2}[\ln(a/b)^{1/2}]^2. \end{aligned} \tag{26}$$

Equation (25a) for ψ must be solved subject to the "flow-around-the-edge-of-the-disks" constraint (10b). The solution can be found by using the conformal transformation

$$\rho + iz = b + (a/\pi)[1 + u + iv + \exp(u+iv)], \quad |v| \le \pi. \tag{27}$$

More specifically, in terms of the function $u(\rho, z)$ the solution is

$$\psi = (M/\pi b)[u - \ln(8\pi b/a)] \tag{28a}$$

$$\approx -\frac{M}{\pi b}\ln\left(\frac{8b}{\tilde{r}}\right)\left[1 + O\left(\frac{a}{\tilde{r}}\right)\right] \quad \text{if } \tilde{r} \gg a \text{ and } |z| > \frac{b-\rho}{|b-\rho|}a \tag{28b}$$

$$\approx -\frac{M}{\pi b}\ln\left(\frac{8\pi b}{a}\right) + \frac{M}{b}\left(\frac{2\bar{r}}{\pi a}\right)^{1/2}\left(\cos\frac{\bar{\theta}}{2}\right)\left[1 + O\left(\frac{\bar{r}^{1/2}}{a^{1/2}}\right)\right] \quad \text{if } \bar{r} \ll a \tag{28c}$$

$$\approx -\frac{M}{a}\left(\frac{b-\rho}{b}\right) - \frac{M}{\pi b}\left[\ln\left(\frac{8\pi b}{a}\right) + 1 + O(\exp[-\pi(b-\rho)/a])\right] \quad \text{if } (b-\rho) \gg a, \quad |z| < a. \tag{28d}$$

See Fig. 3 for definitions of the coordinates $\tilde{r}, \tilde{\theta}, \bar{r}, \bar{\theta}$.

By comparing Eq. (28c) with Eq. (14a), we obtain the explicit form of condition (19), which makes the physical geometry nonsingular at the common edge of the disks:

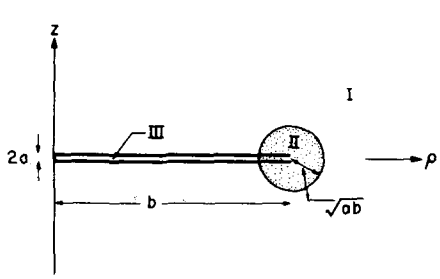


FIG. 4. Three regions, I, II, III, in which three different approximate solutions are valid for the case of a thin-ring torus.

$$B^2 = 2/b \Leftrightarrow M^2 = \pi ab. \tag{29}$$

Henceforth we shall regard M and b as independent variables, and a as the algebraic combination

$$a \equiv M^2/\pi b. \tag{30}$$

The solution for γ can be obtained from Eq. (28) for ψ by integrating Eq. (25b) and imposing the boundary condition (24c) at the outer edge of Region II:

$$\gamma = \frac{M^2}{\pi ab}\,\mathrm{Re}\left[\ln\left(\frac{\exp(u+iv)}{1+\exp(u+iv)}\right)\right] \tag{31a}$$

$$\approx -\frac{M^2}{\pi^2 ab}\left[\frac{a\cos\tilde{\theta}}{\tilde{r}} + O\left(\frac{a^2}{\tilde{r}^2}\ln\frac{\tilde{r}}{a}\right)\right] \text{ if } \tilde{r} \gg a, \ |z| > \frac{b-\rho}{|b-\rho|}\,a \tag{31b}$$

$$\approx \frac{M^2}{2\pi ab}\left[\ln\left(\frac{a}{2\pi\bar{r}}\right) + O\left(\frac{\bar{r}^{1/2}}{a^{1/2}}\right)\right] \quad \text{if } \bar{r} \ll a \tag{31c}$$

$$\approx -\frac{M^2}{a^2}\left\{\frac{b-\rho}{b} + \frac{a}{\pi b} + O\left[\frac{a}{b}\exp\left(-\frac{(b-\rho)\pi}{a}\right)\right]\right\} \quad \text{if } b-\rho \gg a, \quad |z| < a. \tag{31d}$$

C. Region III

Region III is the "inner region" between the disks and bounded away from their edges. In solving for ψ and γ here we ignore the existence of the edges, thereby making

$$(\text{fractional errors in } \psi \text{ and } \gamma) \lesssim a/(ba)^{1/2} = (a/b)^{1/2} \tag{32}$$

and thereby obtaining the cylindrically symmetric expressions (12). The constants ψ_0 and γ_0 in those expressions are fixed by matching onto Region II [Eqs. (28d) and (31d)]:

$$\psi = (M/a)\ln(\rho/b) - (M/\pi b)\ln(8\pi b/a), \tag{33}$$

$$\gamma = (M/a)^2 \ln(\rho/b) - M^2/\pi ab. \tag{34}$$

IV. DISCUSSION OF THE SOLUTION

A. The vacuum solution

The asymptotically flat region of spacetime (the region of redshifts small compared to unity and of nearly globally Minkowski geometry) is that region in which $|\psi| \ll 1$ and $|\gamma| \ll 1$. For the thin-ring case (Sec. III, where $M^2 = \pi ab$ and $a \ll b$) all of Region I is asymptotically flat; the strong-field regime begins in Region II. This allows one to perform Newtonian analyses in Region I, using $\psi = \frac{1}{2}\ln|g_{00}|$ as the Newtonian gravitational potential. Straightforward examination of Eq. (23) shows that a Newtonian observer in Region I will regard the source as a thin ring of total mass-energy M and ring radius b.

Notice that the relation $M^2 = \pi ab$ can be rewritten as

$$\frac{2M}{b} = \frac{(\text{"Schwarzschild radius" of ring})}{(\text{"actual radius" of ring})} = \left(\frac{4\pi a}{b}\right)^{1/2} \sim (\text{fractional errors in thin-ring formulas}). \tag{35}$$

This says that, for rings of fixed mass M and ever decreasing ring radius b, the "thin-ring approximation" $a \ll b$ breaks down when b becomes of order the Schwarzschild radius $2M$ of the ring. In this limit the general solution of Sec. II remains valid, but the thin-ring formulas of Sec. III fail.

B. The join to an interior solution

The author's PhD thesis[2] develops mathematical tools for the analysis of infinitely long, cylindrically symmetric systems. Those tools should be applicable, with fractional errors $\lesssim O[(a/b)^{1/2}] = O[M/b]$, in Region III of our thin-ring toroidal solution. One tool of particular interest is the following theorem, which can be inferred from Sec. 8-M of the author's thesis:

Consider an infinitely long, nonsingular material cylinder which is momentarily static and which has, as its external gravitational field, the Levi-Civita line-mass metric with "spacetime character" $D^{(+)}$.[9,10] Demand that the cylinder have nonnegative energy density T^{00} on its hypersurface of time symmetry. Then at the surface of the cylinder (point where $T^{\alpha\beta} \to 0$) the "C-energy" scalar U must be positive.[10,11]

In the $D^{(+)}$ Levi-Civita metric, U is $-\infty$ at the singularity and increases monotonically as one moves radially outward. At some radius ρ_c, U becomes zero; and thereafter it continues to increase, approaching $+\frac{1}{8}$ as $\rho \to \infty$. The above theorem says that *any material cylinder with $T^{00} > 0$, which generates the $D^{(+)}$ Levi-Civita metric, must have its surface outside the "critical radius" ρ_c at which $U = 0$.*

Region III of the thin-ring toroidal solution is endowed with a Levi-Civita metric of character $D^{(+)}$. The C-energy scalar at radius ρ can be calculated by combining that metric [Eqs. (3), (33), (34)] with Eq. (7.8) of the author's thesis[10]; the result is

$$U = \tfrac{1}{8}[1 - (b/\rho)^2]. \tag{36}$$

Thus, the critical radius is

$$\rho_c = b. \tag{37}$$

But this radius lies outside Region III—i.e., it is so large that the line element is already showing noticeable deviations from that of Levi-Civita! Thus, one is forced to conclude that *any nonsingular, momentarily (or permanently) static torus which generates the thin-ring metric and which has nonnegative energy density must have its surface outside Region III—i.e., in Region II or Region I.*

This surprising (and, to me, unhappy) result is intimately tied to the fact that the thin-ring toroidal metric of this paper has only two independent parameters. Since the *general* Levi-Civita solution has two free parameters ("mass parameter" and "canonical radius"),[12] one might hope to construct a locally cylindrical, globally toroidal vacuum metric with three independent parameters—two characterizing the Levi-Civita singularity and one characterizing the radius of the ring. By adjusting one of the singularity parameters appropriately, one would then be able to build interior solutions with given M and b and with arbitrarily small

surface radii. However, such solutions will not be possible unless one succeeds in adding a new free parameter to the two-parameter vacuum metric of this paper. I have tried, and failed.

ACKNOWLEDGMENTS

For helpful discussions I wish to thank Dieter Brill, Achilles Papapetrou, and John A. Wheeler.

*Research carried out in 1964—65, while the author was an NSF Predoctoral Fellow at Princeton University. Publication supported by NSF Grant MPS75-01398 at Caltech.

[1]A. Einstein and N. Rosen, J. Franklin Inst. **223**, 43 (1937); N. Rosen, Bull. Res. Council Israel **3**, 328 (1953); J. Weber and J.A. Wheeler, Rev. Mod. Phys. **29**, 509 (1957); W.B. Bonnor, J. Math. Mech. **6**, 203 (1957); L. Marder, Proc. Roy. Soc. (London) **244**, 524 (1958); **246**, 133 (1958).

[2]K.S. Thorne, "Geometrodynamics of Cylindrical Systems," unpublished Ph.D thesis, Princeton University (1965), available from University Microfilms, Ann Arbor, Michigan; W.J. Cocke, J. Math. Phys. **7**, 1171 (1966); K.S. Thorne, in *Magic Without Magic: John Archibald Wheeler*, edited by J. Klauder (Freeman, San Francisco 1972); E.P. Liang, Phys. Rev. D **10**, 447 (1974).

[3]H. Weyl, Ann. Physik, **54**, 117 (1917). For a pedagogical treatment see J.L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960), pp. 309—17. Our notation, terminology, and sign conventions are those of C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[4]H. Weyl, Ref. 3, and R. Bach and H. Weyl, Math. Z. **13**, 134 (1922) have derived and studied the metric produced by such a toroidal line monopole.

[5]M. Misra, Proc. Nat. Inst. Sci. India **27A**, 373 (1961).

[6]H. Weyl, Ref. 3; G. Darmois, *Les équations de la gravitation einsteinienne* (Gauthier-Villars, Paris, 1927).

[7]T. Levi-Civita, Rend. Acc. Lincei **28**, 101 (1919).

[8]See Bach and Weyl, Ref. 4.

[9]For the concept of "spacetime character" see Sec. 7-J of the author's thesis (Ref. 2).

[10]Note that the signature of the metric in the author's thesis is opposite to that used in this paper.

[11]For the concept of "C-energy" see Chap. 7 of the author's thesis

[12]See, e.g., Chap. 8 of the author's thesis.

Canonical transformations and accidental degeneracy. IV. Problems with continuous spectra

M. Moshinsky* and J. Patera†

Instituto de Física, Universidad de México, México 20, D.F., Mexico
(Received 13 January 1975; revised manuscript received 29 April 1975)

In the present series of papers we have been trying to bring out the unifying role of groups of canonical transformations in the understanding of problems of accidental degeneracy in quantum mechanics. In Paper III of this series we achieved our purposes for two-dimensional problems with discrete spectra. In the present paper we turn our attention to problems with continuous spectra. There is the well-known case of the free particle in the full plane in which the accidental degeneracy is due to the Euclidean group in two dimensions, $E(2)$. We show that in this problem the accidental degeneracy can also be explained by an $O(2,1)$ group of canonical transformations which provides a clue of the approach to more general problems. We also derive explicitly the group $O(2,1)$, and not only its Lie algebra, associated with the accidental degeneracy of the Coulomb problem in two dimensions. The procedure followed in the above problems is "*sui generis*" and does not provide a general approach. For the latter we discuss two new problems with continuous spectra that have accidental degeneracy: the free particle in a sector of angle π/q, q integer, of the plane and the Calogero problem with continuous spectrum. For both of these problems we find the canonical transformations that map them on the free particle in the full plane. It turns out that their accidental degeneracy is explained then by the $O(2,1)$ group of the latter problem, that we mentioned above, rather than by $E(2)$. The procedures developed seem general enough to encompass other problems of accidental degeneracy in configuration spaces of two or more dimensions.

1. INTRODUCTION

In the present series of papers[1,2,3] we have been trying to bring out the unifying role of groups of canonical transformations, in the understanding of problems of accidental degeneracy in quantum mechanics. In particular, in paper III of this series,[3] we indicated the general approach one could follow in obtaining these groups for two-dimensional problems whose spectra was of the type

$$E_{nN} = C(k_1 n + k_2 N) + D. \tag{1.1}$$

In (1.1) C, D are arbitrary constants and k_1, k_2 relatively prime integers. The energy E_{nN} is then a linear function of the quantum numbers n and N. Most two-dimensional problems with accidental degeneracy,[1-6] that have a discrete spectrum, can be reduced to ones in which the latter takes the form (1.1).

While the detailed discussion in paper III concerned the problem of the oscillator in a sector of angle π/q, q integer, or the Calogero problem,[5] the structure of the analysis clearly showed its validity for all spectra of the type (1.1).

There are cases though, in which we have problems with a continuous spectrum that show a remarkable degeneracy. Again, for the sake of keeping the physical ideas more clearly in focus, we shall restrict ourselves to the lowest possible number of dimensions in which degeneracy is present, i.e., two. Two problems with continuous spectrum come immediately to our attention. The first one is the free particle in which for a given energy we have an infinite number of levels corresponding to the angular momentum $m = 0, \pm 1, \pm 2, \cdots$. The second is the Coulomb problem where for positive energies we have the same type of degeneracy.

In the case of the free particle the degeneracy cannot be termed accidental[7] as there is in fact a group of point transformations in the space x_1, x_2, the Euclidean group $E(2)$, responsible for it. We shall show in the next section that there is a simple group of transformations in momentum space, which is a realization of $O(2, 1)$, that can be also associated with the degeneracy problem of the free particle in the plane. The latter group will be particularly significant when we proceed to develop a more systematic approach to problems with continuous spectra that present degeneracies.

For positive energies, the two-dimensional Coulomb problem has a symmetry group $O(2, 1)$ whose Lie algebra is well known.[8] A simple way of deriving the group itself will be given in Sec. 3 where first by the standard dilatation technique we reduce the problem to what we call the "pseudo-Coulomb" form[9] and then by a point transformation it becomes a two-dimensional repulsive oscillator. The latter admits a symmetry group $O(2, 1)$ which is a subgroup of the symplectic group $Sp(4)$ of linear canonical transformations related with the dynamical group of the problem.[10]

The procedure followed in the literature[7,8] for the analysis of these two problem has been "sui generis" and we are left in the dark on what to do in other cases. But are there other cases? We wish to draw on the experience in article III of this series to immediately suggest two problems of interest.

The first one concerns a free particle not in the plane but in a sector of angle π/q, q integer, whose wavefunction vanishes at the edges of the sector. The solution is elementary as it coincides with the function describing the vibrations of a membrane in a wedge. For a given energy we have an infinite number of states associated with the eigenvalues $\mu^2 q^2$ (we take $\hbar = 1$) of the square of the angular momentum p_φ^2 where $\mu = 1, 2, \cdots$. Thus degeneracy is present and its explanation is not as trivial as that of the free particle in the full plane. The problem is discussed in detail in Sec. 4

where we show how to find a canonical transformation that maps the problem on the full plane, but then it is not the $E(2)$ of the latter that is responsible for accidental degeneracy but rather the $O(2,1)$ that is discussed in Sec. 2.

The second concerns a particular case of the Calogero problem[5] discussed in III. It had to do with three particles in one dimension interacting through potentials proportional and inversely proportional to the square of their distances. Eliminating the center of mass we have a problem in the plane with discrete spectrum. But if we also eliminate the force proportional to the square of the distance, i.e., the harmonic oscillator part, the eigenstates of the Hamiltonian form a continuous spectrum.[11] As we show in Sec. 5 the levels have an infinite degeneracy again characterized by a quantum number μ that can take the values $\mu = 1, 2, 3, \cdots$. In a similar way that we did for the sector problem, we proceed to derive the generators of the Lie algebra of $O(2,1)$ which relates all the degenerate states and thus provides us with an explanation of the accidental degeneracy of this problem.

The approach and techniques followed in Secs. 4 and 5 seem to be applicable to other problems with accidental degeneracy and, in the last section of the present paper, we outline what could be a general procedure for situations of this type.

2. SYMMETRY GROUPS FOR A FREE PARTICLE IN THE PLANE

The eigenstates of the free particle in the plane can be expressed as

$$|km\rangle \equiv i^m J_m(kr)\exp(im\varphi), \quad m = 0, \pm 1, \pm 2, \cdots, \tag{2.1}$$

where J_m is a Bessel function and, in units in which $\hbar$ and the mass of the particle are 1, the energy is

$$E = \tfrac{1}{2}k^2. \tag{2.2}$$

The degeneracy in this case cannot be termed accidental as one normally attributes it to the Euclidean group $E(2)$ of point transformations in the space x_1, x_2. The generators of this group are

$$p_\varphi \equiv x_1 p_2 - x_2 p_1, \quad p_\pm \equiv p_1 \pm i p_2,$$
$$p_j = \frac{1}{i}\frac{\partial}{\partial x_j}, \quad j = 1, 2, \tag{2.3}$$

which satisfy the commutation relations

$$[p_\varphi, p_\pm] = \pm p_\pm, \quad [p_+, p_-] = 0. \tag{2.4}$$

In fact, using polar coordinates, we can write

$$p_\pm = \exp(\pm i\varphi)(p_r \pm i r^{-1} p_\varphi), \quad p_r = \frac{1}{i}\frac{\partial}{\partial r}, \quad p_\varphi = \frac{1}{i}\frac{\partial}{\partial \varphi},$$
$$x_\pm \equiv \tfrac{1}{2}(x_1 \mp i x_2) = \tfrac{1}{2} r \exp(\mp i\varphi), \tag{2.5}$$

and from the properties of Bessel functions[12] we have

$$p_\pm |km\rangle = k|km \pm 1\rangle, \quad p_\varphi |km\rangle = m|km\rangle, \tag{2.6}$$

thus seeing that the generators $p_\pm$ of $E(2)$ will connect all states (2.1) associated with a given energy $\frac{1}{2}k^2$. We note also, for later use, that the states can be written as

$$|k \pm |m|\rangle = k^{-|m|} p_\pm^{|m|} |k_0\rangle, \tag{2.7}$$

which is the reason for the choice of phase in (2.1).

The Euclidean group $E(2)$ whose generators are (2.3) gives then the following transformation in classical phase space

$$x'_\pm = \exp(\pm i\alpha)x_\pm - a_\pm, \quad p'_\pm = \exp(\mp i\alpha)p_\pm, \tag{2.8}$$

where α, $a_\pm$ are the parameters associated with elements of the group. In quantum mechanics we are interested in the matrix representation of the generators (2.3) of the group $E(2)$ with respect to the states (2.1) of definite energy. To indicate a problem that we will have to face in the sector and Calogero case, we write these matrices explicitly taking for convenience $k = 1$ and denoting by m', m the row and column indices. We have then

$p_+ = \| \delta_{m',m+1} \|$

m' \ m	−3	−2	−1	0	1	2	3	
−3	·	·	·	·	·	·	·	
−2	·	·	·	·	·	·	·	
−3	·	·	·	·	·	·	·	
−2	·	·	1	0	0	0		
= −1	·	·	0	1	0	0		(2.9a)
0	·	·	0	0	1	0		
1	·	·	0	0	0	1	0	0 0 · · ·
2						0	1	0 0 · · ·
3						0	0	1 0 · · ·
							·	· · · · ·
							·	· · · · ·

$p_- = \| \delta_{m',m-1} \|$

m' \ m	−3	−2	−1	0	1	2	3	
	·	·	·	·	·	·		
−3	·	·	1	0	0	0		
−2	·	·	0	1	0	0		
−1	·	·	0	0	1	0		
= 0	·	·	0	0	0	1	0	0 (2.9b)
1						0	1	0 · · ·
2						0	0	1 · · ·
3						0	0	0 · · ·
							·	· · · · ·
							·	· · · · ·

$p_\varphi = \| m\delta_{m',m} \|$

m' \ m	−3	−2	−1	0	1	2	3	
−3	−3	0	0	0				
−2	0	−2	0	0				
−1	0	0	−1	0				
= 0	0	0	0	0				(2.9c)
1					1	0	0	
2					0	2	0	
3					0	0	3	

The matrices p_+, p_- of course commute, as must be the case from the operator relations (2.4), but had we taken the submatrices in the lower right corner for which $m', m = 1, 2, 3, \cdots$, they would not commute as in fact they give a 1 for $m' = m = 1$ and zero for all other terms. This is as it should be, because otherwise the states (2.1) would not constitute a basis for an *irreducible* representation (BIR) of $E(2)$. But when we go to the sector problem, where the quantum number μ takes only positive integer values, the possibility of introducing the $E(2)$ group as a symmetry group raises paradoxes such as the lack of commutation of p_+, p_-.

This situation leads us to ask the question whether $E(2)$ is the only symmetry group we can associate with the free particle in the plane. The answer is that there is in fact an infinite number of ways in which we can introduce symmetry groups in this problem, but some are particularly simple. The one we wish to discuss is the $O(2, 1)$ group whose generators are given in the quantum mechanical picture by

$$I_+ \equiv I_1 + iI_2 = p_\varphi p_\varphi (p_+ p_-)^{-1/2}, \quad I_- \equiv I_1 - iI_2 = (p_+ p_-)^{-1/2} p_\varphi p_-,$$
$$I_3 = p_\varphi - \tfrac{1}{2}. \tag{2.10}$$

we note that $p_+ p_-$ is both the Hamiltonian of the free particle and the Casimir operator of $E(2)$ and thus it commutes with the generators (2.3) of this group and therefore also with the operators (2.10) which are essentially a part of its enveloping algebra. The matrix representation of the operators (2.10) with respect to the states (2.1) is given by:

$I_+ = \| m\delta_{m',m+1} \|$

m' \ m	-3	-2	-1	0	1	2	3
-3							
-2		-3	0	0	0		
-1		0	-2	0	0		
0		0	0	-1	0		
1					0	0	0
2					1	0	0
3					0	2	0

(2.11a)

$I_- = \| (m-1)\delta_{m',m-1} \|$

m' \ m	-3	-2	-1	0	1	2	3
-3		0	-3	0	0		
-2		0	0	-2	0		
-1		0	0	0	-1		
0		0	0	0	0		
1					0	1	0
2					0	0	2
3					0	0	0

(2.11b)

$I_3 = \| (m - \tfrac{1}{2})\delta_{m'm} \|$

m' \ m	-3	-2	-1	0	1	2	3
-3	$-\frac{7}{2}$	0	0	0			
-2	0	$-\frac{5}{2}$	0	0			
-1	0	0	$-\frac{3}{2}$	0			
0	0	0	0	$-\frac{1}{2}$			
1					$\frac{1}{2}$	0	0
2					0	$\frac{3}{2}$	0
3					0	0	$\frac{5}{2}$

(2.11c)

These matrices clearly satisfy the commutation relations

$$[I_3, I_\pm] = \pm I_\pm, \quad [I_+, I_-] = -2I_3, \tag{2.12}$$

and I_1, I_2, I_3 are Hermitian. Thus they are associated with the Lie algebra of $O(2, 1)$. We note though that, contrary to what happens to p_+, p_-, p_φ of the Euclidean group, they are reducible into the upper left and lower right blocks. This we expect as the Casimir operator becomes

$$I^2 = I_- I_+ - I_3(I_3 + 1) = \tfrac{1}{4}, \tag{2.13}$$

and thus it contains[13] the two representations $D_+^{1/2}$, $D_-^{1/2}$.

One can then argue that while the group $O(2, 1)$, whose generators are the operators (2.10), is a symmetry group, it is not the one that explains accidental degeneracy as it does not associate a *single* irreducible representation with all the states of a given energy. This problem can be avoided if to the infinitesimal transformation associated with the generators $I_\pm$, I_3 we add the finite reflection transformation R on the $x_2 = 0$ axis, i.e., the change $\varphi \to -\varphi$. The matrix representation of R on the basis (2.1) has then 1 on the antidiagonal and 0 elsewhere, and thus the states (2.1) from a *single* BIR of the group generated by $I_\pm$, I_3, and R. This is in entire analogy with the situation of the orthogonal group[14] $O(2)$ which includes reflections and whose BIR are two-dimensional, and its $O'(2)$ subgroup under which the representation is reducible.

We have found an alternative way of explaining the accidental degeneracy present in the problem of the free particle in the plane. Its interest lays not so much in its relevance for this problem but rather in the fact that if we disregard the reflection R the representation is reducible. Thus when we discuss the particle in the sector π/q, q integer, where the quantum number $\mu = 1, 2, 3, \cdots$ is restricted to positive integers, we will be able to use the lower right matrix in (2.11), to show that the states belong to the *single* irreducible representation $D_+^{1/2}$ of $O(2, 1)$.

Before turning our attention to the other problems we want to discuss in this article, we actually like to obtain explicitly the $O(2, 1)$ group whose generators are given by (2.10), as a group of canonical transformations in classical phase space. For this purpose we note that, using the correspondance principle, we have to think of p_φ as associated with large quantum num-

bers and thus we disregard the $\frac{1}{2}$ in the definition of I_3. We have then

$$I_1 = p_1 p_\varphi (p_+ p_-)^{-1/2}, \quad I_2 = p_2 p_\varphi (p_+ p_-)^{-1/2}, \quad I_3 = p_\varphi, \tag{2.14}$$

and the group can be obtained by applying the operator[15]

$$\exp[\alpha(I_3)_{op}] \exp[\beta(I_2)_{op}] \exp[\gamma(I_3)_{op}] \tag{2.15}$$

to the vector in phase space. The classical operator $(I_j)_{op}$, $j = 1, 2, 3$ is defined by

$$(I_j)_{op} = \sum_{k=1}^{2} \left(\frac{\partial I_j}{\partial x_k} \frac{\partial}{\partial p_k} - \frac{\partial I_j}{\partial p_k} \frac{\partial}{\partial x_k} \right), \tag{2.16}$$

and thus

$$(I_j)_{op} F = \{I_j, F\}, \tag{2.17}$$

where the last parenthesis is the standard Poisson bracket.

The application of $\exp[\alpha(I_3)_{op}]$ is trivial and gives for the momenta the transformation

$$p_1' = p_1 \cos\alpha + p_2 \sin\alpha,$$
$$p_2' = -p_1 \sin\alpha + p_2 \cos\alpha. \tag{2.18}$$

For $\exp[\beta(I_2)_{op}]$ we note that from the Poisson bracket relations

$$\{I_1, I_2\} = -I_3, \quad \{I_3, I_1\} = I_2, \quad \{I_2, I_3\} = I_1, \tag{2.19}$$

we have

$$\begin{bmatrix} I_1' \\ I_2' \\ I_3' \end{bmatrix} = \exp[\beta(I_2)_{op}] \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} c & 0 & s \\ 0 & 1 & 0 \\ s & 0 & c \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix}, \tag{2.20a}$$

where

$$c = \cosh\beta, \quad s = \sinh\beta. \tag{2.20b}$$

Taking the definitions (2.14) both for the primed and the original I_j, we finally obtain

$$p_1' = (I_1'/I_3')(p_1'^2 + p_2'^2)^{1/2} = [cp_1 + s(p_1^2 + p_2^2)^{1/2}] \times [sp_1(p_1^2 + p_2^2)^{-1/2} + c]^{-1}, \tag{2.21}$$
$$p_2' = (I_2'/I_3')(p_1'^2 + p_2'^2)^{1/2} = p_2[sp_1(p_1^2 + p_2^2)^{-1/2} + c]^{-1},$$

where we made use of the fact that as the Poisson bracket of $p_+ p_-$ with I_j is zero we have $p_1'^2 + p_2'^2 = p_1^2 + p_2^2$. We can check this directly both in (2.18) and (2.21).

The canonical transformation associated with the generators (2.14) of $O(2, 1)$ is then a rather simple point transformation in momentum space. To get the corresponding transformation in the configuration variables we only note that we can use the quantum relation

$$x_l' = i\frac{\partial}{\partial p_l'} = \sum_{k=1}^{2} \frac{\partial p_k}{\partial p_l'} i\frac{\partial}{\partial p_k} = \sum_{k=1}^{2} \frac{\partial p_k}{\partial p_l'} x_k, \quad l = 1, 2, \tag{2.22}$$

and thus derive them straightforwardly.

We have obtained the group of classical canonical transformations associated with the symmetry group $O(2, 1)$ of the free particle. We now proceed to derive the corresponding group for the Coulomb case in the continuous spectra, before turning to the new problems that suggest a general type of attack in these situations.

3. SYMMETRY GROUP FOR THE COULOMB PROBLEM WITH CONTINUOUS SPECTRUM

The two-dimensional Coulomb problem with continuous spectrum requires solving the Schrödinger equation

$$(\tfrac{1}{2}P^2 - R^{-1})\psi = (2\nu^2)^{-1}\psi, \tag{3.1}$$

where R, P are coordinates and momenta in atomic units in which the energy can be denoted by

$$E = (2\nu^2)^{-1} \tag{3.2}$$

with ν being any real number.

As is customary for this problem[8,9] we introduce the dilatation transformation

$$\rho = \nu^{-1}R, \quad \pi = \nu P, \tag{3.3}$$

so that the equation (3.1) becomes

$$\tfrac{1}{2}\rho(\pi^2 - 1)\psi = \nu\psi. \tag{3.4}$$

We have denoted the operator on the left-hand side of (3.4) as the Hamiltonian of the pseudo-Coulomb problem. Introducing polar coordinates ρ, ϑ, for the vector ρ, Eq. (3.4) becomes

$$\tfrac{1}{2}\rho\left(-\frac{1}{\rho}\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho} - \frac{1}{\rho^2}\frac{\partial^2}{\partial\vartheta^2} - 1\right)\psi = \nu\psi, \tag{3.5}$$

which by the point transformation

$$\rho = \tfrac{1}{2}r^2, \quad \vartheta = 2\varphi, \tag{3.6}$$

reduces to that of the repulsive harmonic oscillator, i.e.,

$$H\psi \equiv \frac{1}{2}\left(-\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial}{\partial r} - \frac{1}{r^2}\frac{\partial^2}{\partial\varphi^2} - r^2\right)\psi = 2\nu\psi. \tag{3.7}$$

Thus we concentrate on the latter problem when we discuss accidental degeneracy and the symmetry group responsible for it.

From (3.7) it is clear that for each value of ν [that determine through (3.2) the energy of the Coulomb problem] we have an infinite number of states characterized by the eigenvalues $0, \pm 1, \pm 2, \cdots$ of the angular momentum p_φ.

What is then the symmetry group of the repulsive oscillator and, more specifically, what are the generators of the Lie algebra of this group?

In Refs. 10 and 15 we showed that dynamical group of the two-dimensional attractive oscillator was the four-dimensional symplectic group $Sp(4)$ whose generators are

$$x_i x_j, \quad p_i p_j, \quad \tfrac{1}{2}(x_i p_j + p_j x_i), \quad i, j = 1, 2. \tag{3.8}$$

As these operators from a Lie algebra,[10,15] and as linear combinations of them give $p^2 - r^2$ (the Hamiltonian of the repulsive oscillator), we expect that the generators of the Lie algebra of the symmetry group of our problem will also be linear combinations of them.

To construct these generators explicitly we first introduce creation and annihilation operators in spherical components by the definitions

$$\eta_\pm = x_\mp - i\tfrac{1}{2}p_\pm,$$
$$\xi_\pm = x_\pm + i\tfrac{1}{2}p_\mp, \tag{3.9}$$

where we use the expressions (2.3) and (2.5) for $p_\pm$ and $x_\pm$. The η's and ξ's commute among themselves and furthermore

$$[\xi_\pm, \eta_\pm]=1, \quad [\xi_\pm, \eta_\mp]=0. \tag{3.10}$$

From the definition (3.7) of the Hamiltonian H for the repulsive oscillator, we have also

$$H=-(\eta_+\eta_- + \xi_+\xi_-). \tag{3.11}$$

We can now show trivially, using (3.10), that the operators

$$T_+ \equiv T_1 + iT_2 = \tfrac{1}{2}(\eta_+^2 + \xi_-^2), \quad T_- \equiv T_1 - iT_2 = \tfrac{1}{2}(\eta_-^2 + \xi_+^2),$$
$$T_3 = \tfrac{1}{2}(\eta_+\xi_+ - \eta_-\xi_-) = \tfrac{1}{2}p_\varphi, \tag{3.12}$$

satisfy the following commutation relations

$$[H, T_\pm]=0, \quad [H, T_3]=0,$$
$$[T_3, T_\pm]=\pm T_\pm, \quad [T_+, T_-]=-2T_3. \tag{3.13}$$

As furthermore T_1, T_2, T_3 are Hermitian, we conclude that they are the generators of a Lie algebra of $O(2,1)$ corresponding to the symmetry group of the repulsive oscillator.

To find out to which irreducible representation of $O(2,1)$ belong the eigenkets of the Hamiltonian, we first define them fully by the equations

$$H|\nu m\rangle = 2\nu|\nu m\rangle, \quad T_3|\nu m\rangle = m|\nu m\rangle. \tag{3.14}$$

As T_3 is half the angular momentum we can expect m to be integer or half integer, but because of the relation $\varphi=(\vartheta/2)$ given in (3.6), we see that the dependence on ϑ will be precisely $\exp(im\vartheta)$ and thus the pseudo-Coulomb problem will be single valued only if m takes the *integer* values $m=0, \pm 1, \pm 2, \cdots$. Furthermore, we normalize the states through the relation

$$\langle \nu' m'|\nu m\rangle = \delta(\nu'-\nu)\delta_{m'm}. \tag{3.15}$$

To find out the effect of the operator $T_\pm$ on the kets $|\nu m\rangle$, we note that from the commutation relations (3.13) we have

$$T_\pm|\nu m\rangle = A_\pm(\nu, m)|\nu m\pm 1\rangle, \tag{3.16}$$

where we still have to determine the coefficients $A_\pm(\nu, m)$. For this purpose we note from (3.11), (3.12) that the Casimir operator of the $O(2,1)$ group is given by

$$T^2 \equiv T_\mp T_\pm - T_3(T_3 \pm 1) = \tfrac{1}{4}(H^2+1).$$

We obtain then

$$\langle \nu' m'|T_\mp T_\pm|\nu m\rangle = |A_\pm(\nu, m)|^2\delta(\nu'-\nu)\delta_{m'm}$$
$$= \langle \nu' m'|T^2 + T_3(T_3\pm 1)|\nu m\rangle$$
$$= [(\nu^2+\tfrac{1}{4}) + m(m\pm 1)]\delta(\nu'-\nu)\delta_{m'm}. \tag{3.17}$$

Thus, making the usual choice of phase factor,[13] we have

$$A_\pm(\nu, m) = [(\nu^2+\tfrac{1}{4}) + m(m\pm 1)]^{1/2}, \tag{3.18}$$

and identifying $\nu^2+\frac{1}{4}$ with[13]

$$\nu^2 + \tfrac{1}{4} \equiv \lambda(1-\lambda), \tag{3.19}$$

we obtain for λ the value

$$\lambda = \tfrac{1}{2} \pm i\nu, \tag{3.20}$$

which indicates that the states $|\nu m\rangle$ are BIR of $O(2,1)$ in the principal series.[13]

We have proved that the Lie algebra of $O(2,1)$ whose generators are the T_i, $i=1,2,3$, is responsible for the accidental degeneracy of the repulsive oscillator and thus also of the pseudo-Coulomb problem. What is the group itself of canonical transformations in phase space? Again we have to apply the classical operator (2.15) to a vector in phase space replacing the I_i by T_i. It is more convenient to discuss the transformation of the vector $(\eta_+, \xi_+, \eta_-, \xi_-)$ of creation and annihilation variables. Expressing $(T_i)_{op}$ in terms of these variables [see Eq. (3.1) of Ref. 2] we immediately obtain that the application $\exp[\alpha(T_3)_{op}]$ to the vector gives

$$\eta_\pm' = \exp(\mp i\alpha/2)\eta_\pm, \quad \xi_\pm' = \exp(\pm i\alpha/2)\xi_\pm. \tag{3.21}$$

For $\exp[\beta(T_2)_{op}]$ we note that the operator can be written as[2]

$$(T_2)_{op} = \frac{1}{2i}[(T_+)_{op} - (T_-)_{op}]$$
$$= \frac{1}{2}\left(\eta_+\frac{\partial}{\partial\xi_+} + \xi_+\frac{\partial}{\partial\eta_+}\right) - \frac{1}{2}\left(\eta_-\frac{\partial}{\partial\xi_-} + \xi_-\frac{\partial}{\partial\eta_-}\right), \tag{3.22}$$

and thus we obtain

$$\begin{pmatrix}\eta_+'\\ \xi_+'\\ \eta_-'\\ \xi_-'\end{pmatrix} = (T_2)_{op}\begin{pmatrix}\eta_+\\ \xi_+\\ \eta_-\\ \xi_-\end{pmatrix} = \frac{1}{2}\begin{bmatrix}0&1&0&0\\1&0&0&0\\0&0&0&1\\0&0&1&0\end{bmatrix}\begin{pmatrix}\eta_+\\ \xi_+\\ \eta_-\\ \xi_-\end{pmatrix}. \tag{3.23}$$

The exponentiation of $(T_2)_{op}$ is then immediate and combining the effect of all the operators in (2.15), we obtain

$$\begin{pmatrix}\eta_+'\\ \xi_+'\end{pmatrix} = M(\alpha,\beta,\gamma)\begin{pmatrix}\eta_+\\ \xi_+\end{pmatrix}, \quad (\eta_-', \xi_-') = (\eta_-, \xi_-)M^{-1}(\alpha,\beta,\gamma), \tag{3.24}$$

where the matrix $M(\alpha,\beta,\gamma)$ is given by

$$M(\alpha,\beta,\gamma) = \begin{bmatrix}\exp(-i\gamma/2)\cosh\tfrac{1}{2}\beta\exp(-i\alpha/2) & \exp(i\gamma/2)\sinh\tfrac{1}{2}\beta\exp(-i\alpha/2)\\ \exp(-i\gamma/2)\sinh\tfrac{1}{2}\beta\exp(i\alpha/2) & \exp(i\gamma/2)\cosh\tfrac{1}{2}\beta\exp(i\alpha/2)\end{bmatrix}. \tag{3.25}$$

Clearly, when we write the Hamiltonian (3.11) as

$$H = -(\eta_-, \xi_-)\begin{pmatrix}\eta_+\\ \xi_+\end{pmatrix}, \tag{3.26}$$

we obtain its invariance under the linear canonical transformations (3.24). It is interesting to note that the transformation affects (η_+, ξ_+), (η_-, ξ_-), independently and for each of them it corresponds to the $O(2,1)$ [or equivalently the $SU(1,1)$ or $Sp(2)$] dynamical group of the one dimensional oscillator.[16]

From (3.24) and (3.9) we can express the symmetry

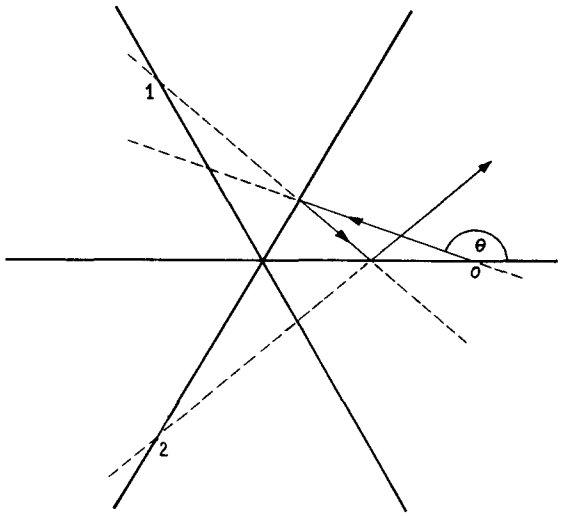


FIG. 1. Trajectory of a free particle in a sector of angle $\pi/3$. The actual trajectory is given by the full lines. The straight lines formed by a full segment and its extension by dotted segments, represent the trajectories given by the method of images. All three of these straight lines map on a single line in the plane X_1, X_2 related to the motion of a free particle in the *full* plane. Point 1 is the image point of 0 with respect to the reflection line at angle $\pi/3$. Point 2 is the image point of 1 with respect to the horizontal reflection line.

group in terms of x_+, x_-, p_+, p_- and, using (2. 5), in terms of r, φ, p_r, p_φ. If we want to have this group in terms of the phase space vectors of the pseudo-Coulomb problem, we have to make use of the point transformation (3. 6) and the corresponding relation for the momenta[9]

$$\pi_\vartheta = \tfrac{1}{2} p_\varphi, \quad \rho\pi_\rho = \tfrac{1}{2} r p_r. \tag{3.27}$$

4. THE FREE PARTICLE IN A SECTOR

We consider a free particle in a sector of the plane of angle π/q, q integer. In Fig. 1 we draw the sector for the case $q=3$ and also show a classical trajectory (the full line) which was obtained by using the method of images.

The eigenkets of this problem are given by

$$\begin{aligned} |k\mu) &= i^{\mu q+1} J_{\mu q}(kr)\sin(\mu q\varphi) \\ &= \tfrac{1}{2} k^{-\mu q}(p_+^{\mu q} - p_-^{\mu q})|k0\rangle, \quad \mu = 1, 2, \cdots. \end{aligned} \tag{4.1}$$

We designate them by round brackets to distinguish them from the angular kets associated with the states in the full plane. The last expression in (4. 1) comes from (2. 1), (2. 7) when we remember that $J_m(x) = (-1)^m J_{-m}(x)$ for m integer.

To each energy $\frac{1}{2}k^2$ we have again an infinite number of states corresponding to the values $\mu = 1, 2, 3, \cdots$. We note that because of collisions with the edge of the sector the angular momentum changes sign there, but p_φ^2 remains an integral of motion and in fact

$$p_\varphi^2 |k\mu) = q^2\mu^2 |k\mu). \tag{4.2}$$

We wish now to find the symmetry group that is responsible for the accidental degeneracy of this problem and, more specifically, the generators of the corresponding Lie algebra.

As a first step in this direction we proceed to derive operators, which we designate by $\bar{p}_+$, $\bar{p}_-$, $\bar{p}_\varphi$, whose effect on $|k\mu)$ is similar to that of p_+, p_-, p_φ on $|km\rangle$, i. e. ,

$$\bar{p}_\pm |k\mu) = k|k\mu \pm 1), \quad \bar{p}_\varphi |k\mu) = \mu |k\mu). \tag{4.3}$$

To achieve this purpose we introduce the auxiliary operator

$$\alpha^+ \equiv p_+^q + p_-^q, \tag{4.4}$$

and note from (4. 1) that

$$\alpha^+ |k\mu) = k^q\{|k\mu+1) + |k\mu-1)\}, \tag{4.5}$$

where this equation holds for all $\mu = 1, 2, 3, \cdots$ if we interpret the undefined round ket $|k0) = 0$.

Combining (4. 2) and (4. 5) we get that

$$[p_\varphi^2, \alpha^+]|k\mu) = k^q q^2(2\mu+1)|k\mu+1) - k^q q^2(2\mu-1)|k\mu-1). \tag{4.6}$$

Using (2. 4), the commutator on the left-hand side can also be written as

$$\begin{aligned} [p_\varphi^2, \alpha^+] &= p_\varphi[p_\varphi, \alpha^+] + [p_\varphi, \alpha^+]p_\varphi \\ &= [p_\varphi, [p_\varphi, \alpha^+]] + 2[p_\varphi, \alpha^+]p_\varphi \\ &= q^2\alpha^+ + 2q(p_+^q - p_-^q)p_\varphi, \end{aligned} \tag{4.7}$$

so it follows from (4. 5), (4. 6) that

$$\tfrac{1}{2}[(p_+^q + p_-^q) \pm (p_+^q - p_-^q)(\mu q)^{-1} p_\varphi]|k\mu) = k^q |k\mu \pm 1). \tag{4.8}$$

As in paper III[3] we can now define the operator $|p_\varphi|$ as the one whose eigenfunction is $|k\mu)$ but with eigenvalue μq, i. e. ,

$$|p_\varphi||k\mu) = q\mu |k\mu). \tag{4.9}$$

Furthermore, as k^{q-1} is the eigenvalue of $(p_+p_-)^{(q-1)/2}$, we can write

$$\bar{p}_\pm = \tfrac{1}{2}\{(p_+^q + p_-^q) \pm (p_+^q - p_-^q)|p_\varphi|^{-1} p_\varphi\}(p_+p_-)^{(1-q)/2}, \tag{4.10a}$$

$$\bar{p}_\varphi = q^{-1}|p_\varphi|. \tag{4.10b}$$

At first sight it would seem that the set of states $|k\mu)$ of the sector problem are BIR of the Euclidean group $E(2)$ as from construction the Eqs. (4. 3) should hold. But this is not true, because from the remark after Eq. (4. 5) we have that for $\mu = 1$

$$\bar{p}_- |k1) = 0, \tag{4.11}$$

and thus with the operators (4. 10) the Eqs. (4. 3) are not satisfied in all cases. The reason for the paradox can be seen immediately when we turn to the matrices p_+, p_-, p_φ of (2. 9) when we are dealing with the particle in the full plane. As we indicated in Sec. 2, the lower right submatrices associated with indices $m, m' = 1, 2, 3, \cdots$ are *not* a representation for the Lie algebra of $E(2)$.

At this point we can turn though to the $O(2,1)$ group discussed also in Sec. 2. The submatrices there do follow the commutation relations associated with the Lie algebra of $O(2,1)$ as the representation is reducible. Thus if we introduce now the operators

$$\bar{I}_+ = \bar{p}_+\bar{p}_\varphi(\bar{p}_+\bar{p}_-)^{-1/2}, \quad \bar{I}_- = (\bar{p}_+\bar{p}_-)^{-1/2}\bar{p}_\varphi\bar{p}_-, \quad \bar{I}_3 = \bar{p}_\varphi - \tfrac{1}{2}, \tag{4.12}$$

their representation on the states $|k\mu)$ will be given precisely by the lower left submatrices of (2.11), i.e., the representation would be $D_+^{1/2}$.

Classically the group of canonical transformations responsible for accidental degeneracy of the sector problem is then given by (2.18), (2.21), (2.22) when we replace p_1, p_2, x_1, x_2 and p_1', p_2', x_1', x_2', by the corresponding variables with bars above. In turn the relations between $\bar{p}_+$, $\bar{p}_-$, $\bar{x}_+$, $\bar{x}_-$ and the p_+, p_-, x_+, x_- can be obtained from (4.10) when we go to the classical picture. In there we have two cases to consider $p_\varphi = |p_\varphi| > 0$ and $p_\varphi = -|p_\varphi| < 0$. In the first case we get

$$\bar{p}_+ = p_+^{(1+q)/2} p_-^{(1-q)/2}, \quad \bar{p}_- = p_+^{(1-q)/2} p_-^{(1+q)/2} \tag{4.13a}$$

(4.13b)

and, as these are functions of p_+, p_- only, we can make use of the relations $\bar{x}_\pm = i\partial/\partial\bar{p}_\pm$, $x_\pm = i\partial/\partial p_\pm$, to obtain

$$\bar{x}_+ = \left(\frac{q+1}{2q}\right)\left(\frac{p_-}{p_+}\right)^{(q-1)/2} x_+ + \left(\frac{q-1}{2q}\right)\left(\frac{p_-}{p_+}\right)^{(q+1)/2} x_-, \tag{4.13c}$$

$$\bar{x}_- = \left(\frac{q-1}{2q}\right)\left(\frac{p_+}{p_-}\right)^{(q+1)/2} x_+ + \left(\frac{q+1}{2q}\right)\left(\frac{p_+}{p_-}\right)^{(q-1)/2} x_-. \tag{4.13d}$$

For the case $p_\varphi < 0$ a similar analysis gives

$$\bar{p}_+ = p_+^{(1-q)/2} p_-^{(1+q)/2}, \tag{4.14a}$$

$$\bar{p}_- = p_+^{(1+q)/2} p_-^{(1-q)/2}, \tag{4.14b}$$

$$\bar{x}_+ = \left(\frac{q-1}{2q}\right)\left(\frac{p_+}{p_-}\right)^{(1+q)/2} x_+ + \left(\frac{q+1}{2q}\right)\left(\frac{p_+}{p_-}\right)^{(q-1)/2} x_-, \tag{4.14c}$$

$$\bar{x}_- = \left(\frac{q+1}{2q}\right)\left(\frac{p_-}{p_+}\right)^{(q-1)/2} x_+ + \left(\frac{q-1}{2q}\right)\left(\frac{p_-}{p_+}\right)^{(q+1)/2} x_-. \tag{4.14d}$$

Equations (4.13) and (4.14) provide, respectively, the canonical transformations that maps the problem of a free particle in a sector on the problem of a free particle in the full plane, for $p_\varphi > 0$ and $p_\varphi < 0$.

We shall show explicitly for the case $q = 3$ that the mapping has all the features we expect. Without loss of generality we can start our classical motion at a point (indicated as 0) on the $x_2 = 0$ axis a distance r_0 from the origin. The angle of the momentum with this axis could be denoted by θ and thus our initial conditions at point 0 will be

$$x_{\pm 0} = \tfrac{1}{2} r_0, \quad p_{\pm 0} = k \exp(\pm i\theta). \tag{4.15a}$$

To use the method of images, we need to know the motion starting at the point 1, which is a reflection of the point 0 with respect to the line $\varphi = \pi/3$, and also at point 2 which is a reflection of point 1 with respect to the line $\varphi = 0$. The initial condition at point 1 is then

$$x_{\pm 0}' = \tfrac{1}{2} r_0 \exp(\mp i 2\pi/3), \quad p_{\pm 0}' = k \exp(\mp i\theta) \exp(\pm i 2\pi/3), \tag{4.15b}$$

and at point 2 it becomes

$$x_{\pm 0}'' = \tfrac{1}{2} r_0 \exp(\pm i 2\pi/3), \quad p_{\pm 0}'' = k \exp(\pm i\theta) \exp(\mp i 2\pi/3). \tag{4.15c}$$

At the starting point 0, $p_\varphi > 0$ and so we have to use the transformation (4.13) to get the initial values of the new coordinates and momenta

$$\bar{p}_{\pm 0} = k \exp(\pm i 3\theta), \quad \bar{x}_{\pm 0} = [\tfrac{2}{3} \exp(\mp i 2\theta) + \tfrac{1}{3} \exp(\mp i 4\theta)] \tfrac{1}{2} r_0. \tag{4.16}$$

Now $p_\pm$ continues to conserve its initial value in its motion until the particle hits the wall at $\varphi = \pi/3$. Then the momentum is the one indicated at the point 1, but as now we use formula (4.14) as $p_\varphi < 0$ we continue to obtain $\bar{p}_\pm = k \exp(\pm i 3\theta)$. Finally the particle hits the wall $\varphi = 0$ and its motion originates now from point 2. We use then (4.13), with the values of (4.15c), to get again the same value for $\bar{p}_\pm$. Through the motion, and independently of the collisions with the walls, we get

$$\bar{p}_\pm = k \exp(\pm i 3\theta). \tag{4.17}$$

Furthermore as we see from (4.13) and (4.14) that

$$H = \tfrac{1}{2} p_+ p_- = \tfrac{1}{2} \bar{p}_+ \bar{p}_-, \tag{4.18}$$

we have from the Hamiltonian equations of motion

$$\frac{d\bar{x}_\pm}{dt} = \frac{\partial H}{\partial \bar{p}_\pm} = \tfrac{1}{2} \bar{p}_\mp \quad \text{or} \quad \bar{x}_\pm = \tfrac{1}{2} \bar{p}_\mp t + \bar{x}_{\pm 0}. \tag{4.19}$$

Thus the trajectory illustrated by the solid line in Fig. 1, translates in the configuration space $\bar{x}_1$, $\bar{x}_2$ into the straight line given by the parametric equations

$$\begin{aligned} \bar{x}_1 &= kt \cos 3\theta + (\tfrac{2}{3} \cos 2\theta + \tfrac{1}{3} \cos 4\theta) r_0, \\ \bar{x}_2 &= kt \sin 3\theta + (\tfrac{2}{3} \sin 2\theta + \tfrac{1}{3} \sin 4\theta) r_0. \end{aligned} \tag{4.20}$$

As a last point we consider some relations between polar coordinates and momenta in the old and new systems. Using Eqs. (4.13) or (4.14), we have

$$\bar{r}\bar{p}_r = \bar{x}_+ \bar{p}_+ + \bar{x}_- \bar{p}_- = x_+ p_+ + x_- p_- = r p_r, \tag{4.21}$$

$$\bar{p}_\varphi = -i(\bar{x}_+ \bar{p}_+ - \bar{x}_- \bar{p}_-) = \begin{cases} q^{-1}(-i)(x_+ p_+ - x_- p_-) & \text{for } p_\varphi > 0 \\ q^{-1}(-i)(x_- p_- - x_+ p_+) & \text{for } p_\varphi < 0 \end{cases}$$

$$= q^{-1} |p_\varphi| \tag{4.22}$$

$$\bar{r}^2 = r^2 + (1 - q^2)(2q^2)^{-1} p_\varphi^2 (p_+ p_-)^{-1}. \tag{4.23}$$

The transformation that maps the free particle in a sector into a free particle in the full plane is thus a fairly complicated one. There is though a particularly simple case when $q = 1$, i.e., the half plane. We have then from (4.13) and (4.14) that

$$\bar{p}_+ = p_+, \quad \bar{p}_- = p_-, \quad \bar{x}_+ = x_+, \quad \bar{x}_- = x_-, \quad \text{for } p_\varphi > 0, \tag{4.24a}$$

$$\bar{p}_+ = p_-, \quad \bar{p}_- = p_+, \quad \bar{x}_+ = x_-, \quad \bar{x}_- = x_+, \quad \text{for } p_\varphi < 0. \tag{4.24b}$$

Having found a realization of the $O(2,1)$ group on the states $|k\mu)$ and the canonical transformation that maps the problem of the sector on the full plane, we turn to the Calogero problem.

5. THE CALOGERO PROBLEM WITH CONTINUOUS SPECTRUM

As a last example of accidental degeneracy we discuss the Calogero problem with continuous spectrum that was mentioned in the Introduction. The Hamiltonian, once we eliminate the center of mass motion and the oscillator force, takes the operator form[3,11]

$$H = -\frac{1}{2}\left(\frac{1}{r}\frac{\partial}{\partial r} r \frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2}\right) + \frac{9\tau(\tau-1)}{2r^2 \sin^2 3\varphi}, \tag{5.1}$$

which still admits the integral of motion

$$M^2 \equiv p_\varphi^2 + \frac{9\tau(\tau-1)}{\sin^2 3\varphi}, \quad p_\varphi = \frac{1}{i}\frac{\partial}{\partial\varphi}. \tag{5.2}$$

In (5.1), (5.2) the strength of the inverse square potential, denoted by g in Ref. 3, is replaced by $\tau(\tau-1)$ where τ is any real number larger than 1.

The nonnormalized eigenstates of H, M^2 corresponding respectively to the eigenvalues[3,11]

$$\tfrac{1}{2}k^2, \quad 9(N+\tau)^2, \tag{5.3}$$

are then given by

$$|kN] = J_{3N+3\tau}(kr)\sin^\tau 3\varphi C_N^\tau(\cos 3\varphi), \quad N=0,1,2,\cdots, \tag{5.4}$$

where C_N^τ is a Gegenbauer polynomial[12] and $J_{3N+3\tau}$ a Bessel function of the orders indicated. We denote the state by a square bracket to distinguish it both from the free particle in the full plane and in the sector.

From (5.4) we see that associated with the energy $E=\frac{1}{2}k^2$ we have an infinite number of states characterized by $N=0,1,2,\cdots$ where this last quantum number is related, through (5.3), with the eigenvalue of M^2. From the experience that we had in the previous problems, it is clear that an essential step for finding in this case the group responsible for its accidental degeneracy (and specifically the generators of its Lie algebra), is to determine operators that take us from the state $|kN]$ to states $|kN\pm1]$.

In the previous section we found for the sector problem an operator a^+ that gave a linear combination of $|k\mu+1)$ and $|k\mu-1)$. It is therefore necessary to find a corresponding operator for the Calogero problem. Fortunately in the bound state Calogero problem this type of operator was provided by Perelomov[6] and we designated it by B^+ in Eq. (4.12) of III. The present problem differs from the bound one by the fact that there is no harmonic oscillator potential. Thus it seems convenient to write the operator B^+ not in the dimensionless units used in III but only with $\hbar=m=1$, keeping the frequency ω as a parameter. This implies only replacing r by $\omega^{1/2}r$, p_r by $\omega^{-1/2}p_r$ and φ, p_φ remain the same. Also using the notation for $p_\pm$, $x_\pm$ given in Sec. 2 of the present article, we can write B^+ as

$$\begin{aligned} B^+ &= (\omega^{1/2}x_- - i\tfrac{1}{2}\omega^{-1/2}p_+)^3 + (\omega^{1/2}x_+ - i\tfrac{1}{2}\omega^{-1/2}p_-)^3 \\ &+ \frac{27}{4}\frac{\tau(\tau-1)}{\omega r^2\sin^2 3\varphi}\left[(\omega^{1/2}r - i\omega^{-1/2}p_r)\cos 3\varphi\right. \\ &\left. + i(3r)^{-1}\omega^{-1/2}p_\varphi \sin 3\varphi\right], \end{aligned} \tag{5.5}$$

where again we replaced the strength g of the inverse square potential by $\tau(\tau-1)$ where τ is any real number larger than 1. In the present units the Hamiltonian H of the bound state Calogero problem can be written as

$$H = \tfrac{1}{2}\omega^{-1}p_+p_- + \tfrac{1}{2}\omega r^2 + 9\tau(\tau-1)(2\omega r^2\sin^2 3\varphi)^{-1}, \tag{5.6}$$

and as shown by Perelomov[3,6]

$$[H, B^+] = 3B^+. \tag{5.7}$$

If we multiply H by ω and B^+ by $(-4\omega^{3/2})$, of course, still we have, the relation

$$[\omega H, -4\omega^{3/2}B^+] = 3\omega(-4\omega^{3/2}B^+), \tag{5.8}$$

but now passing to the limit $\omega\to 0$ we see that

$$\lim_{\omega\to 0}\omega H = H \tag{5.9}$$

and

$$\begin{aligned} \beta^+ &\equiv \lim_{\omega\to 0}(-4\omega^{3/2}B^+) \\ &= -i\tfrac{1}{2}(p_+^3 + p_-^3) - 27\tau(\tau-1)(r^2\sin^2 3\varphi)^{-1}[-ip_r\cos 3\varphi \\ &+ i(3r)^{-1}p_\varphi \sin 3\varphi], \end{aligned} \tag{5.10}$$

where H is the Hamiltonian (5.1). From (5.8) we conclude then that

$$[H, \beta^+] = 0, \tag{5.11}$$

and thus we have found a new constant of the motion for the Hamiltonian of the present problem.

The application of β^+ to $|kN]$ then should give necessarily a linear combination, with respect to N, of the same type of states. As the radial derivatives do not exceed order 3 we have from the properties of the Bessel function[12] that we only expect at most combinations of $|kN+1]$, $|kN]$, $|kN-1]$. In fact a straightforward though lengthy analysis, using properties of the Bessel functions[12] and Gegenbauer polynomials,[12] leads to the equation

$$\beta^+|kN] = k^3\left[-\left(\frac{N+1}{2N+2\tau}\right)|kN+1] + \left(\frac{N+2\tau-1}{2N+2\tau}\right)|kN-1]\right]. \tag{5.12}$$

In (5.12) we have an expression that is the equivalent for the present problem of Eq. (4.5) of the sector problem. In fact, it essentially reduces to it for the case $q=3$ when $\tau=1$.

The next step also parallels the procedure of the preceeding section. From the integral of motion M^2 of (5.2) we have[3,5]

$$M^2|kN] = 9(N+\tau)^2|kN], \tag{5.13}$$

and thus we immediately obtain

$$\begin{aligned} [M^2, \beta^+]|kN] &= 9\beta^+|kN] \\ &- 9(2N+2\tau)k^3\left[\left(\frac{N+1}{2N+2\tau}\right)|kN+1]\right. \\ &\left. + \left(\frac{N+2\tau-1}{2N+2\tau}\right)|kN-1]\right] \end{aligned} \tag{5.14}$$

Combining then (5.12) and (5.14), we finally have

$$\{9(2N+2\tau-1)\beta^+ + [M^2, \beta^+]\}|kN] = -18k^3(N+1)|kN+1], \tag{5.15a}$$

$$\{9(2N+2\tau+1)\beta^+ - [M^2, \beta^+]\}|kN] = 18k^3(N+2\tau-1)|kN-1]. \tag{5.15b}$$

Before proceeding to determine from (5.15) operators whose effect on the ket $|kN]$ are similar to those of $\bar{p}_\pm$, $\bar{p}_\varphi$ on $|k\mu)$ of the sector problem, we note that the $|kN]$ are not normalized in the angular part. We shall denote by $|k\mu\}$ the normalized ket and from III[3,12] we have

$$|k\mu\} \equiv 2^\tau\Gamma(\tau)\left[\frac{3}{2}\frac{(\mu-1)!(\mu-1+\tau)}{\Gamma(\mu-1+2\tau)}\right]^{1/2}|k\mu-1], \tag{5.16}$$

where now we use an index μ instead of N that takes the values $\mu=1,2,3,\cdots$.

We transform Eqs. (5.15) relating $|kN]$ to $|kN\pm1]$, into similar equations relating $|k\mu\}$ to $|k\mu\pm1\}$. We introduce the operator $|M|$ as in III, i.e., its eigenfunctions are $|k\mu\}$ and its eigenvalues are given by[3]

$$|M|\,|k\mu\}=3(\mu-1+\tau)\,|k\mu\}. \tag{5.17}$$

We make use of this last operator to define

$$\bar{p}_+\equiv-(6p_+p_-)^{-1}(|M|)^{1/2}[(|M|-3\tau)(|M|+3\tau-3)\times(|M|-3)]^{-1/2}[3\beta^*(2|M|-3)+i\{M^2,\beta^*\}], \tag{5.18a}$$

$$\bar{p}_-\equiv(6p_+p_-)^{-1}(|M|)^{1/2}[(|M|-3\tau+3)(|M|+3)\times(|M|+3\tau)]^{-1/2}[3\beta^*(2|M|+3)-i\{M^2,\beta^*\}], \tag{5.18b}$$

$$\bar{p}_\varphi\equiv\tfrac{1}{3}(|M|-3\tau)+1, \tag{5.18c}$$

where $\{M^2,\beta^*\}$ is the Poisson bracket which quantum mechanically is

$$\{M^2,\beta^*\}=-i[M^2,\beta^*], \tag{5.19}$$

but that also has the standard classical meaning if we are in this last picture. In that case though by the correspondance principle[3] we can think of $|M|$ as a large number and then all terms $|M|+a$ where a is any number, reduce to $|M|$. Furthermore, in the classical picture $|M|$ is the square root of the M^2 given by (5.2).

From their construction, the effect of the operators $\bar{p}_\pm$, $\bar{p}_\varphi$ on $|k\mu\}$ is exactly the same as that of the operators (4.10) on $|k\mu)$, i.e., it is given by (4.3). Thus again we are tempted to consider the $E(2)$ group as the symmetry group of the Calogero problem. But the paradoxical property

$$\bar{p}_-\,|k1\}=0, \tag{5.20}$$

also holds in this case and thus Eqs. (4.3) with $|k\mu)$ replaced by $|k\mu\}$ are not satisfied in all cases. From here on the analysis of the Calogero problem parallels exactly the discussion of the sector problem given in the previous section after Eq. (4.11). We conclude that the symmetry group responsible for accidental degeneracy in the present case is again $O(2,1)$ and its generators are given by (4.12) where $\bar{p}_\pm$, $\bar{p}_\varphi$ now take the form (5.18).

Having determined the Lie algebras and groups of canonical transformations responsible for accidental degeneracy in the problems analyzed in this paper, we proceed to discuss critically the results and outline what could be a general procedure in these situations.

6. CONCLUSION

The discussion of the previous sections raises a number of questions in relation with groups of canonical transformations responsible for accidental degeneracy for problems with continuous spectra. We restrict our remarks to problems in a two-dimensional configuration space though they seem, at least in principle, generalizable to more.

We found that in all the examples we discussed we could introduce a symmetry group $O(2,1)$ or, equivalently, the $SU(1,1)$ group homomorphic to it.[13] This situation raises the question whether, at least classically, the $SU(1,1)$ is a kind of universal symmetry group related to problems with continuous spectra that have accidental degeneracy, in the same way as $SU(2)$ plays this role in the case of discrete spectra.[17]

At the same time we found problems, such as the free particle in the plane, where more than one type of symmetry group can explain the accidental degeneracy, in this case $E(2)$ and the $O(2,1)$ with reflections. This raises the question of the uniqueness of the groups of canonical transformations that were obtained. The answer seems to be that they are not unique, i.e., there are many ways, in some cases a possible infinite number, in which we can construct the Lie algebras of these groups. As an example we consider, for the free particle in the plane, the operators

$$I_+=p_+[p_\varphi(p_\varphi+1)+f^2(p_+p_-)+\tfrac{1}{4}]^{1/2}(p_+p_-)^{-1/2},\qquad I_-=I_+^\dagger,\quad I_3=p_\varphi, \tag{6.1}$$

where f is an arbitrary function. From the analysis preceeding (3.19) we see that the matrix representation of these operators, when acting on the states (2.1) of the free particle, will give an irreducible representation of the $O(2,1)$ group in the principal series, associated with

$$\lambda=\tfrac{1}{2}\pm if(k^2). \tag{6.2}$$

In the classical limit, by the correspondance principle, we can disregard the $\frac{1}{4}$, 1 appearing in (6.1) and we have

$$I_\pm=I_1\pm I_2=p_\pm[p_\varphi^2+f^2(p_+p_-)]^{1/2}(p_+p_-)^{-1/2},\quad I_3=p_\varphi, \tag{6.3}$$

whose Poisson brackets are given by (2.19) and thus the corresponding group of classical canonical transformations could still be obtained with the help of the operator (2.15). The application of the latter, except in the case $f=0$ discussed in Sec. 2, will be complicated, but at least it is feasible in principle. Thus we can obtain for each function f a different group $O(2,1)$ of canonical transformations. It seems therefore that when speaking of a *group of canonical transformations* responsible for accidental degeneracy we should add the qualifying word *simple*.

How can we systematically obtain groups of canonical transformations, as well as their corresponding Lie algebras, for problems other than those discussed here? We start from a given Hamiltonian that has a continuous spectrum and whose states are degenerate. We first find an operator, function of course of the coordinates and momenta, that characterizes the different states. We call this operator the weight operator and examples are p_φ for the free particle and repulsive oscillator, $|p_\varphi|$ for the particle in the sector, and $|M|$ for the Calogero problem. We then look for operators, which we can denote as ladder operators, that take us from one eigenstate of the weight operator to the next one above or below. We then take functions of these weight and ladder operators that form a Lie algebra, both from their matrix representation on the states of our problem and from the standpoint of their classical Poisson brackets. From the latter by exponentiation, as was done for example in (2.15), we can obtain the classical group of canonical transformations responsible for accidental degeneracy.

The program seems feasible in principle but, as MacIntosh[17] points out, there may exist problems in which the weight and ladder operators exist but that we may not form from them a Lie algebra.

In any case it would be of interest to find other examples of accidental degeneracy in problems with continuous spectra, particularly in the simple case of two-dimensional configuration space, to test the general procedure suggested here.

ACKNOWLEDGMENT

One of the authors (J. Patera) wishes to thank the Consejo Nacional de Ciencia y Tecnologia (CONACYT) of Mexico that has made his stay in Mexico possible. He also wishes to thank the Instituto de Fisica, Universidad de Mexico for its hospitality. Many persons contributed with fruitful discussions to this paper and we want to mention in particular C. Boyer, P. A. Mello, J. Plebanski, K. B. Wolf from Mexico and P. Winternitz from Montreal. We are indebted also to the referee for pointing out that $E(2)$ could not explain the accidental degeneracy of the sector problem essentially because Eq. (4.3) must be supplemented with Eq. (4.11).

*Member of the Institute Nacional de Energia Nuclear and El Colegio Nacional.
†Permanent address: Centre de Recherches Mathématique, Université de Montréal, Montréal, Canada.

[1]J. D. Louck, M. Moshinsky, and K. B. Wolf, J. Math. Phys. 14, 692 (1973).
[2]J. D. Louck, M. Moshinsky, and K. B. Wolf, J. Math. Phys. 14, 696 (1973).
[3]M. Moshinsky, J. Patera, and P. Winternitz, J. Math. Phys. 16, 82 (1975).
[4]M. Moshinsky, *The Harmonic Oscillator in Modern Physics: From Atoms to Quarks* (Gordon and Breach, New York, 1969); M. J. Engelfield, *Group Theory and the Coulomb Problem* (Wiley Interscience, New York, 1972); P. Kramer, M. Moshinsky, and T. Seligman, "Complex Extensions of Canonical Transformations and Quantum Mechanics," in *Group Theory and Applications*, edited by E. Loebl (Academic, New York), Vol. III (to be published 1975). V. A. Fock, Z. Phys. 98, 145 (1935); V. Bargmann, Z. Phys. 99, 576 (1936); J. M. Jauch and E. L. Hill, Phys. Rev. 57, 641 (1940); Yu. M. Demkov, Zh. Eksp. Teor. Fiz. 44, 2007 (1963) [Sov. Phys. JETP 17, 1349 (1963)]; I. Fris, V. Mandrosov, Ya. A. Smorodinskii, M. Uhlir and P. Winternitz, Phys. Lett. 16, 354 (1965); F. Duimio and G. Zambotti, Nuovo Cim. A 43, 1203 (1966); P. Winternitz, Ya. A. Smorodinskii, M. Uhlir, and I. Fris, Yad. Fiz. 4, 625 (1966) [Sov. J. Nucl. Phys. 4, 444 (1967)]. A. A. Makarov, Ya. A. Smorodinskii, Kh. V. Valiev, and P. Winternitz, Nuovo Cim. 52A, 1061 (1967); A. Cisneros and V. McIntosh, J. Math. Phys. 11, 870 (1970).
[5]F. Calogero, J. Math. Phys. 10, 2191 (1969).
[6]A. M. Perelomov, Teor. Mat. Fiz. 6, 364 (1971). [Teor. Math. Phys. 6, 263 (1971)].
[7]J. D. Talman, *Special Functions. A Group Theoretical Approach* (Benjamin, New York, 1968), pp. 189—214.
[8]K. B. Wolf, Suppl. Nuovo Cim. 5, 1041 (1967).
[9]M. Moshinsky, T. H. Seligman, and K. B. Wolf, J. Math. Phys. 13, 901 (1972).
[10]M. Moshinsky and C. Quesne, J. Math. Phys. 12, 1772 (1971).
[11]F. Calogero and C. Marchioro, J. Math. Phys. 15, 1425 (1974).
[12]W. Magnus, F. Oberhettinger, and P. P. Soni, *Formulas and Theorems of Special Functions of Mathematical Physics* (Springer, New York, 1966), pp. 16, 77.
[13]V. Bargmann, Ann. Math. 48, 568 (1947).
[14]E. P. Winer, *Group Theory* (Academic, New York, 1953), pp. 142—148.
[15]M. Moshinsky, SIAM J. Appl. Math. 25, 193 (1973).
[16]M. Moshinsky and C. Quesne, *Proceedings of the XV Solvay Conference on Symmetry in Nuclei* (Gordon and Breach, London, 1974).
[17]H. McIntosh "Symmetry and Degeneracy" in *Group Theory and Its Applications*, edited by E. M. Loebl (Academic, New York, 1971), Vol. II.

Equivalence of a class of Wigner coefficients of $SU(1,1)$ with those of $SU(2)$

E. Chacón,* D. Levi,† and M. Moshinsky‡

Instituto de Física, Universidad de México (UNAM), México, D.F., Mexico
(Received 29 April 1975)

The purpose of this note is to establish the relation $(\lambda_1\lambda_2\mu_1\mu_2|\Lambda M)_{nc} = (-1)^{\mu_2-\lambda_2}\langle(1/2)(M+\lambda_1-\lambda_2-1), (1/2)(M+\lambda_2-\lambda_1-1), (1/2)(\mu_1-\mu_2+\lambda_1+\lambda_2-1), (1/2)(\mu_2-\mu_1+\lambda_1+\lambda_2-1)|\Lambda-1,\lambda_1+\lambda_2-1\rangle$, where the left-hand side is a Wigner coefficient of the noncompact group $SU(1,1)$ and in the right-hand side appears a standard Wigner coefficient of $SU(2)$. The parameters λ_1, λ_2, Λ characterize unitary irreducible representations in the positive discrete series of $SU(1,1)$, and thus they take positive integer or half-integer values. The other parameters are restricted by $\mu_1 = \lambda_1,\lambda_1+1,\lambda_1+2,\cdots$, $\mu_2 = \lambda_2,\lambda_2+1,\lambda_2+2,\cdots$, $M = \Lambda,\Lambda+1,\Lambda+2,\cdots$. Besides we have $\mu_1+\mu_2 = M$ and $\Lambda = \lambda_1+\lambda_2,\lambda_1+\lambda_2+1,\cdots$. A similar result is obtained [cf. our Eq. (3.28)] for Wigner coefficients involving unitary irreducible representations of the negative discrete series of $SU(1,1)$.

1. INTRODUCTION

In a recent paper[1] two of the authors discussed radial one and two body matrix elements using the Wigner–Eckart theorem for the $SU(1,1)$ group. For two body Coulomb matrix elements the Wigner coefficients of $SU(1,1)$ that appear[1] contain only discrete positive representations[2] of the $SU(1,1)$ group, i.e., the lowest weight of the states involved is a positive integer or half integer. In the present paper we shall proceed to show that these particular Wigner coefficients of $SU(1,1)$ are standard Wigner coefficients of $SU(2)$ where the irreducible representations (IR) of the latter are characterized by numbers which are some linear combinations of those appearing in the coefficients of $SU(1,1)$. Thus the Coulomb two body radial matrix elements[1] have all the well-known selection rules and symmetries of the $SU(2)$ Wigner coefficients.

To carry out our identifications we shall first indicate the double meaning we can associate with the eigenstates of the n-dimensional oscillator, i.e., that the states are characterized by the IR of either the chain of groups

$$U(n) \supset O(n), \tag{1.1}$$

or

$$Sp(2)\times O(n), \quad Sp(2)\supset O(2). \tag{1.2}$$

We then particularize our results to $n=4$ and discuss the chains of groups

$$\underset{N}{U(4)} \supset \underset{K}{O(4)} \supset \underset{m_1\quad m_2}{\begin{bmatrix} O(2) & 0 \\ 0 & O(2) \end{bmatrix}}, \tag{1.3}$$

and

$$\underset{N=N_1+N_2}{U(4)} \supset \underset{N_1\quad N_2}{\begin{bmatrix} U(2) & 0 \\ 0 & U(2) \end{bmatrix}} \supset \underset{m_1\quad m_2}{\begin{bmatrix} O(2) & 0 \\ 0 & O(2) \end{bmatrix}}, \tag{1.4}$$

where underneath each group we have put the quantum number that characterizes its irreducible representation. The kets for the chain (1.3) can be denoted by $|NKm_1m_2\rangle$ and those for chain (1.4) by $|N_1N_2m_1m_2\rangle$. We shall prove, through the equivalence for characterization purposes of four-dimensional oscillator states of the chains of groups (1.1) and (1.2), that the transformation bracket

$$\langle N_1N_2m_1m_2|NKm_1m_2\rangle, \tag{1.5}$$

is in fact a Wigner coefficient of $Sp(2)$ and also, therefore, of $SU(1,1)$.[3] We then proceed to evaluate (1.5) showing that it actually reduces to a Wigner coefficient of $SU(2)$. Thus we establish the connection of some of the Wigner coefficients of $SU(1,1)$ (associated with discrete positive IR) with those of $SU(2)$. We note incidentally that m_1, m_2 are nonnegative integers as they are related with IR of $O(2)$ and not $O^+(2)$.

We start our analysis by discussing in terms of creation and annihilation operators the generators of the chains of groups (1.1), (1.2) and then proceed to characterize the states of n-dimensional oscillators by their irreducible representations.

2. STATES OF THE n-DIMENSIONAL OSCILLATOR AND GENERATORS OF THE GROUPS ASSOCIATED WITH THEM

Following Moshinsky and Quesne[4] we introduce the creation and annihilation operators

$$\eta_j = \frac{1}{\sqrt{2}}(x_j - ip_j), \quad \xi_k = \frac{1}{\sqrt{2}}(x_k + ip_k), \quad j,k = 1,\ldots,n. \tag{2.1}$$

The generators of the dynamical group $Sp(2n)$ of the n-dimensional oscillator become then[4]

$$\eta_j\eta_k, \quad \tfrac{1}{2}(\eta_j\xi_k + \xi_k\eta_j), \quad \xi_j\xi_k, \tag{2.2}$$

which form a Lie algebra as can be immediately checked from the commutation relation

$$[\xi_j, \eta_k] = \delta_{jk}. \tag{2.3}$$

The group $Sp(2n)$ admits, among others, the following chains of subgroups:

$$Sp(2n)\supset U(n)\supset O(n), \tag{2.4}$$

$$Sp(2n)\supset Sp(2)\times O(n), \quad Sp(2)\supset O(2). \tag{2.5}$$

For the chain (2.4) the generators are[4]

$$U(n) \to \tfrac{1}{2}(\eta_j\xi_k + \xi_k\eta_j) = \eta_j\xi_k + \tfrac{1}{2}\delta_{jk}, \quad O(n) \to \eta_j\xi_k - \eta_k\xi_j, \tag{2.6}$$

while for $Sp(2)$ they become[1,4]

$$T_+ = -\tfrac{1}{2}\eta\cdot\eta, \quad T_3 = \tfrac{1}{4}(\eta\cdot\xi + \xi\cdot\eta), \quad T_- = -\tfrac{1}{2}\xi\cdot\xi, \tag{2.7}$$

where the dot indicates the scalar product in n-dimensional space. The generator of the subgroup $O(2)$ of $Sp(2)$ is T_3, and $2T_3$ is also the first order Casimir operator of $U(n)$.

The states of the n-dimensional oscillator characterized by the IR N of $U(n)$ and K of $O(n)$ are the eigenkets $|NK\rangle$ of the operators[1,4]

$$2T_3|NK\rangle=(N+n/2)\,|NK\rangle, \tag{2.8a}$$

$$\mathcal{L}^2|NK\rangle=K(K+n-2)\,|NK\rangle, \tag{2.8b}$$

where $\mathcal{L}^2$, the Casimir operator of $O(n)$, is given by[1]

$$\mathcal{L}^2=-\tfrac{1}{2}\sum_{i,j}(\eta_i\xi_j-\eta_j\xi_i)(\eta_i\xi_j-\eta_j\xi_i)$$
$$=-(\eta\cdot\eta)(\xi\cdot\xi)+(\eta\cdot\xi)^2+(n-2)(\eta\cdot\xi).$$

In turn the Casimir operator of $Sp(2)$ is

$$T^2\equiv T_+T_--T_3(T_3-1)$$
$$=\tfrac{1}{4}\{(\eta\cdot\eta)(\xi\cdot\xi)-(\eta\cdot\xi+n/2)[\eta\cdot\xi+(n-4)/2)]\}$$
$$=-\tfrac{1}{4}[\mathcal{L}^2+n(n-4)/4], \tag{2.9}$$

which is related to $\mathcal{L}^2$. Thus the state $|NK\rangle$ is then also an eigenket of T^2, T_3 and we have shown that the chains of groups (2.4), (2.5) provide two equivalent ways of characterizing the states of the n-dimensional harmonic oscillator.

The states $|NK\rangle$ are homogeneous polynomial of degree N in the creation operators applied to the ground state, i.e.,

$$|NK\rangle=P_{NK}(\eta)\,|0\rangle. \tag{2.10}$$

From the form (2.7) of the generators of $Sp(2)$ and the relation (2.9), we see that the Eq. (2.8b) is satisfied by

$$P_{NK}(\eta)=(\eta\cdot\eta)^{(N-K)/2}\,P_{KK}(\eta), \tag{2.11}$$

where

$$\xi\cdot\xi\;P_{KK}(\eta)\,|0\rangle=\left(\sum_{i=1}^{n}\frac{\partial^2}{\partial\eta_i^2}\,P_{KK}(\eta)\right)|0\rangle=0, \tag{2.12}$$

and $N-K$ is even.

For the state of lowest number of quanta $N=K$ consistent with a given value of K, i.e., $|KK\rangle$, the eigenvalue of T_3 is

$$\tfrac{1}{2}(K+n/2) \tag{2.13}$$

and, because of (2.12), $|KK\rangle$ is the lowest weight state of $Sp(2)$ and thus the eigenvalue (2.13) characterizes the IR of this group. If n is even we note that the eigenvalue will be integer or half integer as K is an integer.

From the discussion carried out above it is clear that the states $|NK\rangle$ can also be written as

$$|NK\rangle\equiv|\tfrac{1}{2}(K+n/2),\ \tfrac{1}{2}(N+n/2), \tag{2.14}$$

where the round bracket ket is characterized by the IR of $Sp(2)$, i.e., its lowest weight (2.13), and the IR of $O(2)$, i.e., the eigenvalue (2.8a) of T_3.

Starting from the results discussed in this section, we show in the next one how we can choose different subgroups of $U(4)$ so that the transformation brackets between the basis of irreducible representations (BIR) related with them will be the Wigner coefficients of $Sp(2)$.

3. SUBGROUPS OF U (4) AND TRANSFORMATION BRACKETS RELATING THEIR BIR

We now consider the chain of subgroups (1.3), (1.4) of $U(4)$ discussed in the introduction and the states $|NKm_1m_2\rangle$, $|N_1N_2m_1m_2\rangle$ related with them. From (2.14) we conclude that we can also write

$$|NKm_1m_2\rangle=|\tfrac{1}{2}(K+2),\ \tfrac{1}{2}(N+2);m_1,m_2), \tag{3.1}$$

so we deal with the IR $\frac{1}{2}(K+2)$, $\frac{1}{2}(N+2)$ of $Sp(2)$ and $O(2)$, respectively. For the chain (1.4) we have

$$|N_1N_2m_1m_2\rangle=|N_1m_1\rangle\,|N_2m_2\rangle, \tag{3.2}$$

in which each ket is related with the $U(2)\supset O(2)$ chain and thus we can also write, if we use $Sp(2)\supset O(2)$, that

$$|N_1N_2m_1m_2\rangle=|\tfrac{1}{2}(m_1+1),\ \tfrac{1}{2}(N_1+1))\,|\tfrac{1}{2}(m_2+1),\ \tfrac{1}{2}(N_2+1))$$
$$\equiv|\tfrac{1}{2}(m_1+1),\ \tfrac{1}{2}(m_2+1),\ \tfrac{1}{2}(N_1+1),\ \tfrac{1}{2}(N_2+1)). \tag{3.3}$$

Thus we have the relation

$$\langle N_1N_2m_1m_2\,|NKm_1m_2\rangle$$
$$=(\tfrac{1}{2}(m_1+1),\ \tfrac{1}{2}(m_2+1),\ \tfrac{1}{2}(N_1+1),\ \tfrac{1}{2}(N_2+1)\,|\tfrac{1}{2}(K+2),$$
$$\tfrac{1}{2}(N+2))_{nc}, \tag{3.4}$$

where $(\ |\ )_{nc}$ is the Wigner coefficient of the noncompact group $Sp(2)$ or equivalently $SU(1,1)$, in which the notation $(\lambda_1\lambda_2\mu_1\mu_2|\lambda\mu)_{nc}$ is followed with the λ's being the IR of $Sp(2)$ characterized by their lowest weights and the μ's being the IR of the subgroup $O(2)$ of $Sp(2)$.

We now turn our attention to the evaluation of the transformation bracket (1.5). For this purpose we introduce creation and annihilation operators

$$\eta_i^{\alpha},\ \xi_j^{\beta},\quad i,j=1,2,\quad \alpha,\beta=1,2, \tag{3.5}$$

where the upper indices distinguish between the two $U(2)$ groups of the chain (1.4) while the lower indices relate to the two components within these groups. The generators of $U(4)$ are now

$$\tfrac{1}{2}(\eta_i^{\alpha}\xi_j^{\beta}+\xi_j^{\beta}\eta_i^{\alpha}). \tag{3.6}$$

Those of $O(4)$ are given by

$$\eta_i^{\alpha}\xi_j^{\beta}-\eta_j^{\beta}\xi_i^{\alpha}, \tag{3.7}$$

while the generators of $U(2)$, $O(2)$ are given, respectively, by (3.6), (3.7) when $\alpha=\beta=1$ or $\alpha=\beta=2$.

Let us introduce creation and annihilation operators in spherical components

$$\eta_{\pm}^{\alpha}=\frac{1}{\sqrt{2}}(\eta_1^{\alpha}\pm i\eta_2^{\alpha}),\quad \xi_{\pm}^{\alpha}=\eta_{\pm}^{\alpha^*}=\frac{1}{\sqrt{2}}(\xi_1^{\alpha}\mp i\xi_2^{\alpha}). \tag{3.8}$$

In terms of them the states $|N_1N_2m_1m_2\rangle$ are expressed as

$$|N_1N_2m_1m_2\rangle$$
$$=\frac{(\eta_+^1)^{(N_1+m_1)/2}(\eta_-^1)^{(N_1-m_1)/2}(\eta_+^2)^{(N_2+m_2)/2}(\eta_-^2)^{(N_2-m_2)/2}}{\{[(N_1+m_1)/2]!\,[(N_1-m_1)/2]!\,[(N_2+m_2/2]!\,[(N_2-m_2)/2]!\}^{1/2}}\,|0\rangle. \tag{3.9}$$

We note that the chain (1.4) of subgroups $U(2)$ of $U(4)$ has the generators

$$J_+^{\alpha}\equiv\eta_+^{\alpha}\xi_-^{\alpha},\quad J_0^{\alpha}\equiv\tfrac{1}{2}(\eta_+^{\alpha}\xi_+^{\alpha}-\eta_-^{\alpha}\xi_-^{\alpha}),\quad J_-^{\alpha}\equiv\eta_-^{\alpha}\xi_+^{\alpha},$$

$$\hat{N}^\alpha \equiv \eta_+^\alpha \xi_+^\alpha + \eta_-^\alpha \xi_-^\alpha, \quad \alpha = 1, 2, \tag{3.10}$$

and that the Casimir operators $(J^\alpha)^2$ of $SU(2)$ are related to the number operators $\hat{N}^\alpha$ through

$$(J^\alpha)^2 \equiv J_-^\alpha J_+^\alpha + J_0^\alpha (J_0^\alpha + 1) = \tfrac{1}{4}\hat{N}^\alpha(\hat{N}^\alpha + 2). \tag{3.11}$$

Thus the states (3.9) are the eigenstates of the operators $(J^1)^2, (J^2)^2, J_0^1, J_0^2$ with eigenvalues that are, respectively,

$$\left(\frac{N_1}{2}\right)\left(\frac{N_1}{2}+1\right), \quad \left(\frac{N_2}{2}\right)\left(\frac{N_2}{2}+1\right), \quad \frac{m_1}{2}, \quad \frac{m_2}{2}. \tag{3.12a}$$

To bring out more explicitly the fact that the states (3.9) are eigenstates of $(J^1)^2$, $(J^2)^2$, J_0^1, J_0^2 we shall also denote them by the square bracket ket

$$\left|\frac{N_1}{2}, \frac{N_2}{2}, \frac{m_1}{2}, \frac{m_2}{2}\right] \equiv |N_1 N_2 m_1 m_2\rangle. \tag{3.12b}$$

We now turn our attention to the states $|NKm_1m_2\rangle$. We easily check that the six independent generators (3.7) can be written as linear combinations of

$$\begin{aligned}
&A_+ = \eta_+^1\xi_+^2 - \eta_-^2\xi_-^1, \quad A_- = \eta_+^2\xi_+^1 - \eta_-^1\xi_-^2,\\
&A_0 = \tfrac{1}{2}(\eta_+^1\xi_+^1 - \eta_-^1\xi_-^1 - \eta_+^2\xi_+^2 + \eta_-^2\xi_-^2),\\
&B_+ = \eta_+^2\xi_-^1 - \eta_+^1\xi_-^2, \quad B_- = \eta_-^1\xi_+^2 - \eta_-^2\xi_+^1,\\
&B_0 = \tfrac{1}{2}(\eta_+^1\xi_+^1 - \eta_-^1\xi_-^1 + \eta_+^2\xi_+^2 - \eta_-^2\xi_-^2),
\end{aligned} \tag{3.13}$$

where we have the commutation relations

$$\begin{aligned}
&[A_-, A_+] = -2A_0, \quad [A_0, A_\pm] = \pm A_\pm,\\
&[B_-, B_+] = -2B_0, \quad [B_0, B_\pm] = \pm B_\pm,
\end{aligned} \tag{3.14}$$

and the generators of type A commute with those of type B. Thus the generators of $O(4)$ in (3.13) correspond to the locally isomorphic group $O(3)\times O(3)$.

The Casimir operators of these $O(3)$ groups are

$$A^2 = A_-A_+ + A_0(A_0+1), \quad B^2 = B_-B_+ + B_0(B_0+1), \tag{3.15}$$

and we can show directly that

$$A^2 = B^2 = \tfrac{1}{4}\mathcal{L}^2. \tag{3.16}$$

The ket $|NKm_1m_2\rangle$ satisfies, from (3.13), (3.16), the following equations:

$$\begin{aligned}
&\hat{N}|NKm_1m_2\rangle = N|NKm_1m_2\rangle,\\
&B^2|NKm_1m_2\rangle = (\tfrac{1}{2}K)(\tfrac{1}{2}K+1)|NKm_1m_2\rangle,\\
&(A_0+B_0)|NKm_1m_2\rangle = m_1|NKm_1m_2\rangle,\\
&(-A_0+B_0)|NKm_1m_2\rangle = m_2|NKm_1m_2\rangle,
\end{aligned} \tag{3.17}$$

where the operator $\hat{N}$ is given by

$$\hat{N} = \hat{N}^1 + \hat{N}^2, \tag{3.18}$$

We shall construct the states (3.17) explicitly through the following reasoning. We start by noticing that besides $U(2)$ groups whose generators are the ones given in (3.10) we have another chain of unitary subgroups of $U(4)$ whose generators take the form

$$\begin{aligned}
&I_+^1 \equiv -\eta_+^1\xi_-^2, \quad I_0^1 \equiv \tfrac{1}{2}(\eta_+^1\xi_+^1 - \eta_-^2\xi_-^2), \quad I_-^1 \equiv -\eta_-^2\xi_+^1,\\
&\hat{M}^1 \equiv \eta_+^1\xi_+^1 + \eta_-^2\xi_-^2,\\
&I_+^2 \equiv \eta_+^2\xi_-^1, \quad I_0^2 \equiv \tfrac{1}{2}(\eta_+^2\xi_+^2 - \eta_-^1\xi_-^1), \quad I_-^2 \equiv \eta_-^1\xi_+^2,\\
&\hat{M}^2 \equiv \eta_+^2\xi_+^2 + \eta_-^1\xi_-^1.
\end{aligned} \tag{3.19}$$

It is easy to check that these generators have the standard commutation relations and that their Casimir operators $(I^\alpha)^2$ are related with the $\hat{M}^\alpha$ through expressions similar to (3.11). The construction of the eigenstates of $(I^1)^2$, $(I^2)^2$, I_0^1, I_0^2 corresponding respectively to the eigenvalues

$$\left(\frac{N'}{2}\right)\left(\frac{N'}{2}+1\right), \left(\frac{N''}{2}\right)\left(\frac{N''}{2}+1\right), \frac{m'}{2}, \frac{m''}{2}, \tag{3.20}$$

follows in a way similar to (3.9) but now, from (3.12b), the corresponding ket takes the form

$$\left|\frac{N'}{2}, \frac{N''}{2}, \frac{m'}{2}, \frac{m''}{2}\right] = \frac{(\eta_+^1)^{(N'+m')/2}(-\eta_-^2)^{(N'-m')/2}(\eta_+^2)^{(N''+m'')/2}(\eta_-^1)^{(N''-m'')/2}}{\{[(N'+m')/2]![(N'-m')/2]![(N''+m'')/2]![(N''-m'')/2]!\}^{1/2}}|0\rangle, \tag{3.21}$$

where it is important to note the phase factor (-1) in front of η_-^2, as the replacement of $\eta_-^2 \to -\eta_-^2$, $\xi_-^2 \to -\xi_-^2$, puts the generators I_+^1, I_0^1 in a canonical form.

The purpose of the states (3.21) becomes immediately clear when we notice that

$$B_\pm = I_\pm^1 + I_\pm^2, \quad B_0 = I_0^1 + I_0^2. \tag{3.22}$$

Thus if we want an eigenstate of B^2, such as discussed in (3.17), we can construct it through the ordinary Wigner coefficients of $SU(2)$, i.e.,

$$\sum_{m'/2, m''/2} \left\langle \frac{N'}{2}, \frac{N''}{2}, \frac{m'}{2}, \frac{m''}{2} \middle| \frac{K}{2}, \frac{m_1+m_2}{2} \right\rangle \left|\frac{N'}{2}, \frac{N''}{2}, \frac{m'}{2}, \frac{m''}{2}\right], \tag{3.23}$$

where, again using (3.17), we have denoted by $\frac{1}{2}(m_1+m_2)$ the eigenvalue of the eigenstate of B_0. Furthermore, we notice that

$$A_0 = \tfrac{1}{2}(\hat{M}_1 - \hat{M}_2), \quad \hat{N} = \hat{M}_1 + \hat{M}_2. \tag{3.24}$$

Therefore the state $|NKm_1m_2\rangle$ which satisfies Eqs. (3.17) is given by (3.23) if

$$\tfrac{1}{2}(N'-N'') = \tfrac{1}{2}(m_1-m_2), \quad N'+N'' = N. \tag{3.25}$$

We are now in a position to derive the relation between the Wigner coefficients of $SU(1,1)$ that correspond to discrete positive representations and those of $SU(2)$. We note from (3.4), (3.12b) that

$$(\tfrac{1}{2}(m_1+1), \tfrac{1}{2}(m_2+1), \tfrac{1}{2}(N_1+1), \tfrac{1}{2}(N_2+1)|\tfrac{1}{2}(K+2), \tfrac{1}{2}(N+2))_{nc} = \langle N_1N_2m_1m_2|NKm_1m_2\rangle$$

$$= \sum_{N'N''m'm''} \left[\frac{N_1}{2}, \frac{N_2}{2}, \frac{m_1}{2}, \frac{m_2}{2} \,\middle|\, \frac{N'}{2}, \frac{N''}{2}, \frac{m'}{2}, \frac{m''}{2}\right] \left\langle \frac{N'}{2}, \frac{N''}{2}, \frac{m'}{2}, \frac{m''}{2} \,\middle|\, \frac{K}{2}, \frac{m_1+m_2}{2} \right\rangle$$

$$= \sum_{N'N''m'm''} \left(\delta_{N_1+m_1, N'+m'} \delta_{N_1-m_1, N''-m''} \delta_{N_2+m_2, N''+m''} \delta_{N_2-m_2, N'-m'} (-1)^{N'-m'/2} \left\langle \frac{N'}{2}, \frac{N''}{2}, \frac{m'}{2}, \frac{m''}{2} \,\middle|\, \frac{K}{2}, \frac{m_1+m_2}{2} \right\rangle \right)$$

$$= (-1)^{(N_2-m_2)/2} \left\langle \tfrac{1}{4}(N_1+N_2+m_1-m_2),\ \tfrac{1}{4}(N_1+N_2+m_2-m_1),\ \tfrac{1}{4}(N_1-N_2+m_1+m_2), \right.$$
$$\left. \times \tfrac{1}{4}(N_2-N_1+m_1+m_2) \,\middle|\, \frac{K}{2}, \frac{m_1+m_2}{2} \right\rangle, \tag{3.26}$$

where we made use of the explicit expressions of the kets

$$\left|\frac{N_1}{2}, \frac{N_2}{2}, \frac{m_1}{2}, \frac{m_2}{2}\right] , \quad \left|\frac{N'}{2}, \frac{N''}{2}, \frac{m'}{2}, \frac{m''}{2}\right]$$

given in (3.9), (3.12b), and (3.21) to evaluate the scalar products.

The expression (3.26) provides us with the relation we are searching. Had we used the more common notation $(\lambda_1\lambda_2\mu_1\mu_2 | \Lambda M)_{nc}$ for the Wigner coefficients of $SU(1,1)$ where $\lambda_1, \lambda_2, \Lambda$ are integers or half integers associated with the lowest weight of the discrete positive representations, then the relation (3.26) becomes

$$(\lambda_1\lambda_2\mu_1\mu_2 | \Lambda M)_{nc}$$
$$= (-1)^{\mu_2-\lambda_2} \langle \tfrac{1}{2}(M+\lambda_1-\lambda_2-1),\ \tfrac{1}{2}(M+\lambda_2-\lambda_1-1),$$
$$\times \tfrac{1}{2}(\mu_1-\mu_2+\lambda_1+\lambda_2-1),\ \tfrac{1}{2}(\mu_2-\mu_1+\lambda_1+\lambda_2-1)$$
$$| \Lambda-1,\ \lambda_1+\lambda_2-1\rangle, \tag{3.27}$$

where we remember that $M = \mu_1 + \mu_2$.

It is easy to check that the phase convention used in $(\lambda_1\lambda_2\mu_1\mu_2 | \Lambda M)_{nc}$ coincides exactly with that proposed by Holman and Biedenharn.[2]

By a procedure essentially identical to the one followed to arrive at Eq. (3.27) we can obtain a corresponding result for unitary irreducible representations of the negative discrete series of $SU(1,1)$. The basic modifications one must do in the analysis are the interchange of the upper indices 1, 2 in the η's and ξ's of Eq. (3.19), and the use of new generators T'_q of $Sp(2)$ related to the T_q of (2.7) by

$$T'_+ = T_-, \quad T'_- = T_+, \quad T'_3 = -T_3.$$

The result that we obtain is

$$(\lambda_1\lambda_2\mu_1\mu_2 | \Lambda M)_{nc}$$
$$= (-1)^{-\mu_1-\lambda_1} \langle \tfrac{1}{2}(-M-\lambda_1+\lambda_2-1),\ \tfrac{1}{2}(-M+\lambda_1-\lambda_2-1),$$
$$\times \tfrac{1}{2}(\mu_1-\mu_2+\lambda_1+\lambda_2-1),\ \tfrac{1}{2}(-\mu_1+\mu_2+\lambda_1+\lambda_2-1)$$
$$| \Lambda-1, \lambda_1+\lambda_2-1\rangle, \tag{3.28}$$

where now

$$\mu_1 = -\lambda_1, -\lambda_1-1, -\lambda_1-2, \cdots,$$
$$\mu_2 = -\lambda_2, -\lambda_2-1, -\lambda_2-2, \cdots,$$
$$M = \mu_1 + \mu_2.$$

The phase convention is again the same of Ref. 2.

ACKNOWLEDGMENT

We want to thank Professor P. Winternitz for suggesting to us to examine also the case of the discrete negative series of IR of $SU(1,1)$.

*Member of the Instituto Nacional de Energía Nuclear, Mexico.
†Fellow of the Mexican-Italian exchange program.
‡Member of the Instituto Nacional de Energía Nuclear and El Colegio Nacional.

[1] D. Levi and M. Moshinsky (unpublished).
[2] W. J. Holman and L. C. Biedenharn, Ann. Phys. 39, 1 (1966); 47, 205 (1968).
[3] M. Moshinsky and C. Quesne, "Oscillator Systems" in *Symmetry Properties of Nuclei*, Proceedings of the Fifteenth Solvay Conference on Physics (Gordon and Breach, London, 1974).
[4] M. Moshinsky and C. Quesne, J. Math. Phys. 12, 1772 (1971).

Marginally singular integral equation with divided-difference kernel: A problem in N/D theory*

Porter W. Johnson and Robert L. Warnock

Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 7 January 1975)

We describe a method for solution of a linear integral equation of the form $\phi(s) = f(s) + (1/\pi) \int_4^\infty \{[B(s) - B(t)]/(s-t)\}\ q(t)\phi(t)dt$, where $q(t) = (1-4/t)^{1/2}$. The specified functions $f(s)$ and $B(s)$ have the asymptotic behaviors $f(s) \sim f_0\ (\ln s)^{-\alpha}$, $\alpha > 1$, $B(s) \sim b(\ln s)^{-1}$, $s \to \infty$; in addition, B is subject to smoothness conditions. The equation is analyzed on a Banach space S of continuous functions $\phi(s)$ which have asymptotes of the form $\phi_0(\ln s)^{-\alpha}$. It is found that the integral operator K is bounded but not compact on the space S, so that the equation is not of Fredholm type on S. We separate K into a noncompact part K_1 and a compact part K_2, and construct explicitly the inverse of $1 - K_1$ by solving an associated differential equation. We then convert the original equation $\phi = f + K\phi$ into an equivalent Fredholm equation $\phi = (1 - K_1)^{-1}(f + K_2\phi)$.

1. INTRODUCTION

In the past, physical models in S matrix theory have usually been such that the linear integral equation of the N/D method is a regular Fredholm equation in some standard Banach space (for instance, L^2 or a space of continuous functions with supremum norm).[1] Recently, we have studied a model of meson–meson scattering in which the N/D equation is irregular.[2] The integral operator is bounded on a Banach space S which naturally suggests itself, but it is not compact on that space, or on any other space which comes to mind. Consequently, we have not been able to apply Fredholm theory to the equation in its original form, but we have been able to find a simple way of transforming the equation to an equivalent regular Fredholm equation on the space S. In the following we describe this transformation, without attention to the particular model that inspired it. It is likely that a similar technique will be necessary in models more general and realistic than that of Ref. 2, especially if "realistic" means that the total cross section rises logarithmically at high energy.[3] As further motivation for this work, we mention that an analogous case of a noncompact operator occurs in a study of the unitarity equation.[4]

Singular N/D equations have been discussed before in the literature (Refs. 5–7 and other papers cited therein). Earlier authors have chosen the same general approach that we have taken. Namely, they separate the troublesome part of the operator from the rest, and try to find its inverse by a non-Fredholm method. The assumed characteristics of the troublesome part, and the methods of finding its inverse, have been quite different from those of the present paper.

Integral equations with divided-difference kernels, having essentially the same formal structure as the N/D equation, arise in the analysis of singular integral equations with Cauchy kernel.[8] Notably, they occur in the theory of the Hilbert problem in several unknown functions,[8] which has been taken as the basis for a proof of existence of the many-channel ND^{-1} representation.[9] The proof of Ref. 9 might be carried out under weaker hypotheses by using the method described here; originally, the scattering amplitude was so restricted as to supply a regular Fredholm equation.

In Sec. 2, we specify the integral operator K of interest, and show that it is bounded on a certain Banach space S. In Sec. 3, we separate K into two parts, K_1 and K_2, and show that K_2 is compact on S, but that K_1 does not enjoy that property. In Sec. 4, we show that the inverse of $1 - K_1$ always exists on S, and that it may be constructed explicitly by solving a simple differential equation. Multiplication of the original equation by $(1 - K_1)^{-1}$ then yields a regular Fredholm equation on S, with a compact kernel.

2. DEFINITION AND BOUNDEDNESS OF THE INTEGRAL OPERATOR

The investigation of Ref. 2 led to an equation of the form

$$\psi(s) = f(s) + \int_4^\infty \frac{B(s) - B(t)}{s - t} q(t)\phi(t)\,dt. \tag{2.1}$$

In the following, we regard f and B as being given, although they are nonlinear functionals of the unknown ϕ in the complete problem of Ref. 2. The function ϕ is related to the numerator function of the N/D representation, and $q(t)$ is the phase-space factor, proportional to the momentum:

$$q(t) = [(t-4)/t]^{1/2}. \tag{2.2}$$

The variable s is the squared energy in units of the squared meson mass, so that $s = 4$ is the threshold for a two-meson scattering state.

The following properties of the real functions f and B will be assumed:

(i) f is continuous, and has the asymptotic behavior

$$f(s) = \frac{f_0}{\ln^\alpha s}[1 + o(1)], \quad s \to \infty, \quad \alpha > 1. \tag{2.3}$$

(ii) B has a continuous second derivative, and as s tends to infinity

$$B(s) = \frac{b}{\ln s} + O(\ln^{-2} s), \quad b > 0, \tag{2.4}$$

$$B'(s) = O(s^{-1}\ln^{-2} s), \tag{2.5}$$

$$B''(s) = O(s^{-2}\ln^{-2} s). \tag{2.6}$$

The constants f_0 and b are arbitrary, aside from the restriction $b>0$. In Ref. 2, one has $b=2\pi/9$ and $\alpha=5/3$.

It seems appropriate to seek a solution of (2.1) in a Banach space S, such that the function f is contained in S. Accordingly, we take S to be the set of all real, continuous functions $\phi(s)$ on the half-line $4\leq s<\infty$ such that the following limit exists:

$$\phi_0=\lim_{s\to\infty}\ln^\alpha s\phi(s). \tag{2.7}$$

The norm in S is taken to be

$$\|\phi\|=\sup_{4\leq s<\infty}|\ln^\alpha s\phi(s)|. \tag{2.8}$$

We now write Eq. (2.1) in abridged notation as

$$\phi=f+K\phi, \tag{2.9}$$

and prove that the integral operator K maps S into itself. The integral defining $K\phi$ obviously exists for any ϕ in S, and $K\phi(s)$ is clearly bounded for all s less than any s_0. We may then restrict attention to the case $\mu s>4$, where μ is some constant less than 1. We also choose a constant λ greater than one and write

$$K\phi(s)=\frac{1}{\pi}\left(\int_4^{\mu s}+\int_{\mu s}^{\lambda s}+\int_{\lambda s}^{\infty}\right)\frac{B(s)-B(t)}{s-t}$$

$$\times q(t)\phi(t)dt=(J_1+J_2+J_3)\phi. \tag{2.10}$$

With the help of (2.4) and (2.5) it is a simple matter to bound the $J_i\phi$, when $\phi\in S$. l'Hospital's rule yields the asymptotes of logarithmic integrals, and we have

$$\frac{|J_1\phi|}{\|\phi\|}\leq\frac{M}{(1-\mu)s}\int_4^{\mu s}\left(\frac{1}{\ln s}+\frac{1}{\ln t}\right)\frac{dt}{\ln^\alpha t}$$

$$\leq M\ln^{-\alpha-1}s, \tag{2.11}$$

$$\frac{|J_3\phi|}{\|\phi\|}\leq\frac{M}{(1-\lambda^{-1})}\int_{\lambda s}^{\infty}\left(\frac{1}{\ln s}+\frac{1}{\ln t}\right)\frac{dt}{t\ln^\alpha t}$$

$$\leq M\ln^{-\alpha}s. \tag{2.12}$$

For $J_2\phi$, we use (2.5) and the mean value theorem to obtain

$$\frac{|J_2\phi|}{\|\phi\|}\leq\frac{M}{\mu s\ln^2\mu s}\int_{\mu s}^{\lambda s}\frac{dt}{\ln^\alpha t}\leq M\ln^{-\alpha-2}s. \tag{2.13}$$

In these equations and elsewhere, M is a generic constant which may have different values in different inequalities. To show that $K\phi\in S$, it remains to show that $K\phi$ is continuous and that $\lim\ln^\alpha s\,K\phi(s)$ exists. The estimates (2.11)—(2.13) establish that the three integrals in (2.10) are uniformly convergent in s for $4\leq s\leq S$, for any $S<\infty$. Since the integrands are continuous functions of s, it follows that $K\phi(s)$ is a continuous function of s, for $4\leq s<\infty$. We shall now prove that $K\phi$ is in fact differentiable, with derivative satisfying the bound

$$(K\phi)'(s)\leq M\|\phi\|/s\ln^{1+\alpha}s. \tag{2.14}$$

The formal derivative of $K\phi$ is

$$\left(\int_4^{\mu s}+\int_{\mu s}^{\lambda s}+\int_{\lambda s}^{\infty}\right)\left(B'(s)-\frac{B(s)-B(t)}{s-t}\right)$$

$$\times\frac{q(t)\phi(t)\,dt}{s-t}=\|\phi\|(I_1+I_2+I_3). \tag{2.15}$$

We shall now prove that (2.15) is uniformly convergent on any finite interval of s, so that (2.15) is in fact equal to the derivative of $K\phi$ and is continuous at finite s. We may assume that $\mu s>4$, and apply (2.4), (2.5) to obtain

$$|I_1|\leq\frac{M}{s}\int_4^{\mu s}\left[\frac{1}{s\ln^2 s}+\frac{1}{s}\left(\frac{1}{\ln s}+\frac{1}{\ln t}\right)\right]\frac{dt}{\ln^\alpha t}$$

$$\leq\frac{M}{s\ln^{1+\alpha}s}. \tag{2.16}$$

To bound I_2, we use the mean-value theorem with (2.5) and (2.6) as follows:

$$|I_2|\leq M\int_{\mu s}^{s}\frac{1}{s-t}|B'(s)-B'(s_1)|\frac{dt}{\ln^\alpha t}$$

$$(t\leq s_1\leq s)$$

$$+M\int_s^{\lambda s}\frac{1}{t-s}|B'(s)-B'(s_2)|\frac{dt}{\ln^\alpha t}$$

$$(s\leq s_2\leq t).$$

$$\leq M\int_{\mu s}^{s}|B''(s_3)|\frac{dt}{\ln^\alpha t}\quad(t\leq s_1\leq s_3\leq s)$$

$$+M\int_s^{\lambda s}|B''(s_4)|\frac{dt}{\ln^\alpha t}\quad(s\leq s_4\leq s_2\leq t)$$

$$\leq M\int_{\mu s}^{s}\frac{dt}{t^2\ln^{2+\alpha}t}+\frac{M}{s^2\ln^2 s}\int_s^{\lambda s}\frac{dt}{\ln^\alpha t}$$

$$\leq\frac{M}{s\ln^{2+\alpha}s}. \tag{2.17}$$

Finally, we estimate I_3 by applying the mean value theorem to the second term in the integral:

$$|I_3|\leq M\int_{\lambda s}^{\infty}\left(\frac{1}{s\ln^2 s}+\frac{1}{s_1\ln^2 s_1}\right)\frac{dt}{(t-s)\ln^\alpha t}$$

$$(s\leq s_1\leq t)$$

$$\leq\frac{M}{s\ln^2 s}\int_{\lambda s}^{\infty}\frac{dt}{(t-s)\ln^\alpha(t-s)}$$

$$=\frac{M}{s\ln^2 s}\frac{\ln^{1-\alpha}(\lambda-1)s}{\alpha-1}\leq\frac{M}{s\ln^{1+\alpha}s}. \tag{2.18}$$

This completes the proof that $(K\phi)'$ is given by (2.15), and satisfies the bound (2.14).

To finish the proof that $K\phi$ belongs to S, we show that $\ln^\alpha s\,K\phi(s)$ approaches a limit. Only the term J_3 in (2.10) contributes to the limit, as is evident from (2.11), (2.12), and (2.13). It is readily seen that only the asymptotic parts of $B(s)$ and $B(t)$ are relevant, so that we analyze the integral

$$J(s)=\int_{\lambda s}^{\infty}\left(\frac{1}{\ln s}-\frac{1}{\ln t}\right)\frac{q(t)\phi(t)\,dt}{t-s}. \tag{2.19}$$

We first note that

$$\ln^\alpha t\,q(t)\phi(t)=\phi_0+o(1), \tag{2.20}$$

so that

$$J(s)=\phi_0\int_{\lambda s}^{\infty}\left(\frac{1}{\ln s}-\frac{1}{\ln t}\right)\frac{dt}{\ln^\alpha t(t-s)}+\frac{\epsilon}{\ln^\alpha s}, \tag{2.21}$$

where ϵ may be made arbitrarily small for sufficiently large s. Thus, only the first term of (2.21) contributes to the limit, and we evaluate its limit by l'Hospital's rule. The first term of the integral is

$$\frac{\phi_0}{\ln s}\int_{(\alpha-1)s}^{\infty}\frac{du}{u}\frac{1}{\ln^{\alpha}(s+u)}. \tag{2.22}$$

By comparison of the integral in (2.22) with $\ln^{1-\alpha}s$ through l'Hospital's rule, we see that (2.22) is asymptotic to $\phi_0(\alpha-1)^{-1}\ln^{-\alpha}s$. The second term of the integral in (2.21) is treated in the same way, and the result is that

$$K\phi(s)\sim\frac{-b\phi_0}{\pi\alpha(\alpha-1)\ln^{\alpha}s}. \tag{2.23}$$

It follows that $K\phi\in S$.

3. SEPARATION OF COMPACT AND NONCOMPACT PARTS OF THE OPERATOR

Noncompactness of the operator K is due to a part of the integral in which t may be much greater than s. Such a part of the integral defines the operator K_1:

$$K_1\phi(s)=\frac{b}{\pi}\int_s^{\infty}\left(\frac{1}{\ln t}-\frac{1}{\ln s}\right)\frac{\phi(t)\,dt}{t}. \tag{3.1}$$

We shall first demonstrate that K_1 is not compact on S, and then show that the remainder, $K_2=K-K_1$, is compact.

Our definition of compactness is the usual one: An operator L mapping S into itself is called compact (or completely continuous) if and only if for every bounded sequence $\{\phi_n\}\subset S$, $\{L\phi_n\}$ has a subsequence convergent in S. As is well known,[10,11] Fredholm theory applies to an equation of the form $\phi=f+L\phi$, where $f\in S$ and L is linear and compact on S, in the sense that Fredholm's "determinant-free" theorems hold. For instance, either $(1-L)^{-1}$ exists, or the homogeneous equation $\psi=L\psi$ has a nonzero solution. A compact operator may be approximated uniformly by an operator of finite rank; that circumstance allows one to compute approximations to the solution of $\phi=f+L\phi$, by solving a finite system of linear algebraic equations.[12]

To prove noncompactness of K_1, we employ the Ascoli–Arzelà criterion. With every function ϕ in S we may associate a function $\psi(x)=\ln^{\alpha}s\,\phi(s)$, where $x=4/s$. The functions ψ are continuous on the closed interval $[0,1]$. The space S is equivalent to the space C of functions ψ continuous on $[0,1]$ with norm

$$\|\psi\|=\sup_{0\le x\le 1}|\psi(x)|. \tag{3.2}$$

We define the operator L_1 on C as

$$L_1\psi(x)=\ln^{\alpha}s\,K_1\phi(s). \tag{3.3}$$

To show that K_1 is noncompact on S is equivalent to showing that L_1 is noncompact on C. The Ascoli–Arzelà theorem[10] asserts that an operator L_1 is compact on C if and only if every bounded sequence $\{\phi_n\}\subset C$ is mapped into a uniformly bounded, equicontinuous sequence $\{L_1\phi_n\}$. Therefore, L_1 is noncompact if there exists a sequence $\{\psi_n\}$ of elements of C with uniformly bounded norm such that the set of all functions $L_1\psi_n(x)$ is not equicontinuous on $[0,1]$. We display a sequence $\{\psi_n\}$ with the following properties:

(i) $\psi_n\in C$, $\|\psi_n\|=1$, and $L_1\psi_n(0)=0$ for all n. (3.4)

(ii) For some $\epsilon>0$ and any $\delta\in(0,1)$ there is an n such that

$$|L_1\psi_n(\delta)-L_1\psi_n(0)|=|L_1\psi_n(\delta)|>\epsilon. \tag{3.5}$$

From existence of such a sequence it follows that L_1 is not compact on C, and hence that K_1 is not compact on S.

The ψ_n are defined as

$$\psi_n(x)=\chi_n(s)=\begin{cases}1, & s\le s_n,\\ (2s_n-s)/s_n, & s_n<s\le 2s_n,\\ 0, & s>2s_n,\end{cases} \tag{3.6}$$

where $\{s_n\}$ is any monotonically increasing sequence of numbers such that $\lim s_n=\infty$. Evidently, the conditions (3.4) are satisfied. For any $\delta\in(0,1)$ we write $s_0=4/\delta$, and choose n for (3.5) in such a way that $s_n>s_0^3$. The integrand of (3.1) is strictly negative when $\phi(s)=\phi_n(s)=\ln^{-\alpha}s\,\chi_n(s)$, so that we have

$$\begin{aligned}|L_1\psi_n(\delta)|&=|\ln^{\alpha}s_0K_1\phi_n(s_0)|\\ &\ge\frac{b\ln^{\alpha}s_0}{\pi}\int_{s_0^2}^{s_0^3}\frac{dt}{t\ln^{\alpha}t}\left(\frac{1}{\ln s_0}-\frac{1}{\ln t}\right)\\ &\ge\frac{b}{2\pi}\frac{1}{\ln^{1-\alpha}s_0}\int_{s_0^2}^{s_0^3}\frac{dt}{t\ln^{\alpha}t}\\ &=\frac{b}{2\pi(\alpha-1)}\left(\frac{1}{2^{\alpha-1}}-\frac{1}{3^{\alpha-1}}\right).\end{aligned} \tag{3.7}$$

This verifies (3.5), and the noncompactness of K_1 is proved.

To establish that the operator K_2 is compact on S, we show that the corresponding operator L_2 maps every bounded subset of C into a relatively compact set. In analogy with Eq. (3.3) above, we have defined L_2 by the relation

$$L_2\psi(x)=\ln^{\alpha}sK_2\phi(s). \tag{3.8}$$

We shall need to establish that

$$|(K_2\phi)'(s)|\le M\|\phi\|/s\ln^{\alpha+1}s \tag{3.9}$$

and

$$|K_2\phi(s)|\le M\|\phi\|/\ln^{\alpha+1}s, \tag{3.10}$$

with the generic constant M independent of ϕ.

We have already obtained the bound (2.14) for $|(K\phi)'|$, and one may easily establish the same bound for $|(K_1\phi)'|$ from the relation

$$(K_1\phi)'(s)=\frac{b}{\pi s\ln^2 s}\int_s^{\infty}\frac{\phi(t)}{t}dt. \tag{3.11}$$

The bound (3.9) follows. In establishing (3.10), we note that, in the decomposition (2.10) of $K\phi$, the terms $J_1\phi$ and $J_2\phi$ are subject to the bounds (2.11) and (2.13), respectively. Consequently, we need only establish an analogous bound for the term $(J_3-K_1)\phi$, which we may write in the form

$$(J_3 - K_1)\phi = \frac{1}{\pi}\int_{\lambda s}^{\infty} dt(q(t)-1)\frac{B(s)-B(t)}{s-t}\phi(t)$$
$$+\frac{s}{\pi}\int_{\lambda s}^{\infty}\frac{dt}{t}\frac{B(s)-B(t)}{s-t}\phi(t)$$
$$+\frac{1}{\pi}\int_{\lambda s}^{\infty}\frac{dt}{t}\left(B(t)-\frac{b}{\ln t}-B(s)+\frac{b}{\ln s}\right)\phi(t)$$
$$+\frac{b}{\pi}\int_{s}^{\lambda s}\frac{dt}{t}\left(\frac{1}{\ln s}-\frac{1}{\ln t}\right)\phi(t)$$
$$=\|\phi\|(H_1+H_2+H_3+H_4). \tag{3.12}$$

From the bounds

$$\left|\frac{B(s)-B(t)}{s-t}\phi(t)\right| \le \frac{M\|\phi\|}{\ln s}\frac{1}{t\ln^{\alpha}t}, \quad t \ge \lambda s, \tag{3.13}$$

and

$$|q(t)-1| \le 4/t, \tag{3.14}$$

we obtain

$$|H_1|,\ |H_2| \le M\frac{s}{\ln s}\int_{\lambda s}^{\infty}\frac{dt}{t^2\ln^{\alpha}t} \le \frac{M}{\ln^{\alpha+1}s}. \tag{3.15}$$

Furthermore, from the asymptotic limit (2.4) we obtain

$$|H_3| \le \frac{M}{\ln^2 s}\int_{\lambda s}^{\infty}\frac{dt}{t\ln^{\alpha}t} \le \frac{M}{\ln^{\alpha+1}s}. \tag{3.16}$$

Finally, we bound H_4 through a change of variable $t = rs$:

$$|H_4| \le M\int_1^{\lambda}\frac{dr}{r}\frac{1}{\ln s\ln^{\alpha+1}rs} \le \frac{M}{\ln^{\alpha+2}s}. \tag{3.17}$$

The result (3.10) is thereby established.

From the bounds (3.9) and (3.10) concerning $K_2\phi$, one may infer the following bounds involving $L_2\psi$:

$$|L_2\psi(x)| < N\|\psi\|/(1+|\ln x|) \tag{3.18}$$

and

$$|(L_2\psi)'(x)| < N\|\psi\|/x(1+|\ln x|). \tag{3.19}$$

Let D be any bounded subset of the space C; i.e., there is a d such that $\psi \in D$ implies $\|\psi\| < d$. We shall use the estimates (3.18) and (3.19) to demonstrate that the set $L_2D = \{L_2\phi(x) : \phi \in D\}$ is equicontinuous on $[0,1]$. In other words, we shall show that, for any $\epsilon > 0$, there exists a number δ such that if $x_1, x_2 \in [0,1]$ and $|x_1 - x_2| < \delta$, then

$$|L_2\phi(x_1) - L_2\phi(x_2)| < \epsilon, \tag{3.20}$$

the number $\delta(\epsilon)$ being independent of ϕ. (It then follows from the Ascoli–Arzelà criterion that L_2 is a compact operator on C.)

We shall verify (3.20) with $\delta(\epsilon)$ given by

$$\delta(\epsilon) = \min[(\epsilon/Nd)^2, \tfrac{1}{4}\exp(-4Nd/\epsilon)]. \tag{3.21}$$

We make the choice $x_1 \le x_2$ and treat separately two distinct regimes of x_1.

(i) $x_1 < \delta^{1/2}$, so that $x_2 < \delta + \delta^{1/2}$. From (3.18),

$$|L_2\psi(x_1) - L_2\psi(x_2)| \le Nd\left(\frac{1}{|\ln x_1|}+\frac{1}{|\ln x_2|}\right)$$
$$< \frac{4Nd}{|\ln 4\delta|} \le \epsilon. \tag{3.22}$$

(ii) $x_1 \ge \delta^{1/2}$. Let us use the bound (3.19) along with the mean-value theorem to obtain

$$|L_2\psi(x_1) - L_2\psi(x_2)| \le Nd|x_1 - x_2|/x_1 < Nd\delta^{1/2} \le \epsilon. \tag{3.23}$$

The inequality (3.20) follows.

4. CONSTRUCTION OF INVERSE OF 1-K_1

We shall construct the inverse of $1 - K_1$ as a bounded linear operator on S. Once the inverse is available, our integral equation (2.9) may be cast into the form

$$\phi = (1-K_1)^{-1}(f + K_2\phi). \tag{4.1}$$

The equation (4.1) is susceptible to Fredholm theory, and it is equivalent to (2.9). Thus, (4.1) is the regularized integral equation by which our problem is solved.

To find the inverse, we obtain an explicit solution of the equation

$$(1-K_1)\psi = h, \tag{4.2}$$

for an arbitrary h in S, by means of an associated differential equation. We let

$$\chi(s) = K_1\psi(s), \tag{4.3}$$

take note of the definition (3.1), and differentiate formally to obtain

$$\chi'(s) = \frac{\beta}{s\ln^2 s}\int_s^{\infty}\frac{\psi(t)\,dt}{t}, \tag{4.4}$$

$$(s\ln^2 s\chi'(s))' = \frac{-\beta\psi(s)}{s}, \tag{4.5}$$

$$\beta = \frac{b}{\pi}. \tag{4.6}$$

Substitution of these results in (4.2) yields

$$s(s\ln^2 s\chi'(s))' + \beta\chi(s) = -\beta h(s). \tag{4.7}$$

We shall find that among the solutions χ of (4.7), there is one that gives the unique solution of (4.2) in S through the formula

$$\psi = \chi + h. \tag{4.8}$$

The homogeneous form of (4.7),

$$s(s\ln^2 s\chi'(s))' + \beta\chi(s) = 0, \tag{4.9}$$

has two linearly independent solutions of the form

$$\chi(s) = \ln^{\gamma} s, \tag{4.10}$$

the exponents γ being the roots of the equation

$$\gamma(\gamma+1) + \beta = 0. \tag{4.11}$$

Since β is positive, the roots $\gamma_{\pm}$ of (4.11) may be either real or complex, but their real parts are always greater than -1:

$$\gamma_+, \gamma_- = -\tfrac{1}{2}[1 \pm (1-4\beta)^{1/2}]. \tag{4.12}$$

In the case that arises in Ref. 2, $\beta = 2/9$ and the roots are real. We can now see that if (4.2) has a solution in S, it must be unique. Equation (4.7) is a necessary condition on any solution of (4.2) in S. If (4.2) had two solutions $\psi_1, \psi_2 \in S$ the homogeneous equation (4.9) would have the nontrivial solution

$$\zeta(s) = K_1(\psi_1 - \psi_2)(s). \tag{4.13}$$

[It is clear that ζ is not identically zero, since $s(s\ln^2 s\,\zeta'(s))' = -\beta[\psi_1(s) - \psi_2(s)] \not\equiv 0$]. Since ζ belongs to S, it cannot in fact be a solution of (4.9), since the general solution of the latter is

$$\chi(s) = a_+ \ln^{\gamma_+} s + a_- \ln^{\gamma_-} s, \tag{4.14}$$

where $\mathrm{Re}\,\gamma_+ > -1$. A nonzero solution of the form (4.14) is not in S, since it does not vanish rapidly enough at infinity. It follows that a solution of (4.2) in S is necessarily unique. Notice that the situation would be different if β were negative; then (4.9) would have a solution in S if γ_+ were greater than or equal to α, and our uniqueness argument would fail.

To solve the inhomogeneous equation (4.7), we apply the method of variation of parameters. That is, we seek a solution of the form

$$\chi(s) = \ln^{\gamma_+} s\,\Gamma_+(s) + \ln^{\gamma_-} s\,\Gamma_-(s), \tag{4.15}$$

where the $\Gamma_\pm$ are restricted so that

$$\ln^{\gamma_+} s\,\Gamma_+'(s) + \ln^{\gamma_-} s\,\Gamma_-'(s) = 0. \tag{4.16}$$

When (4.15) is substituted in (4.7), and Eqs. (4.11) and (4.16) are invoked, we obtain a second linear equation for the $\Gamma_\pm'$. By solving the latter equation and (4.16) together, we find (provided $\gamma_- \neq \gamma_+$)

$$\Gamma_\pm'(s) = \pm \beta h(s)/(\gamma_- - \gamma_+) s(\ln s)^{1+\gamma_\pm}. \tag{4.17}$$

When the $\gamma_\pm$ are complex, the function Γ_+' is the complex conjugate of Γ_-'. The case $\gamma_- = \gamma_+$ ($\beta = 1/4$) requires a different calculation, which we defer for the moment. We can now obtain a solution of (4.7) as

$$\chi(s) = (\ln s)^{\gamma_+}\left[\int_\infty^s \Gamma_+'(u)\,du + c_+\right] + (\ln s)^{\gamma_-}\left[\int_\infty^s \Gamma_-'(u)\,du + c_-\right]. \tag{4.18}$$

There is a unique choice of the constants of integration c_+, c_- such that $\chi \in S$ for an arbitrary $h \in S$; namely,

$$c_+ = c_- = 0. \tag{4.19}$$

To show that (4.19) is sufficient for $\chi \in S$, we simply apply l'Hospital's rule to obtain the asymptotes of the integrals:

$$\int_\infty^s \Gamma_\pm'(u)\,du \sim \frac{\pm\beta h_0}{\gamma_- - \gamma_+}\int_\infty^s \frac{du}{u(\ln u)^{1+\alpha+\gamma_\pm}} = \frac{\mp\beta h_0}{(\gamma_- - \gamma_+)(\alpha+\gamma_\pm)}\,\frac{1}{(\ln s)^{\alpha+\gamma_\pm}}, \tag{4.20}$$

where

$$h(s) = (h_0/\ln^\alpha s)[1 + o(1)].$$

Thus, $\chi(s) = (\chi_0/\ln^\alpha s)[1 + o(1)]$, and χ is clearly real, even if the $\gamma_\pm$ are complex. Hence, $\chi \in S$ for any $h \in S$, when (4.19) holds. Equations (4.19) are clearly necessary as well as sufficient for $\chi \in S$, since $\mathrm{Re}\,\gamma_\pm > -1$.

To complete the argument, we must show that the following function actually satisfies (4.2):

$$\psi(s) = (\ln s)^{\gamma_+}\int_\infty^s \Gamma_+'(u)\,du + (\ln s)^{\gamma_-}\int_\infty^s \Gamma_-'(u)\,du + h(s). \tag{4.21}$$

Heretofore, we have only demonstrated that if (4.2) has a solution in S, it must be the function (4.21). We must prove that

$$\beta\int_s^\infty \frac{dt}{t}\left(\frac{1}{\ln t} - \frac{1}{\ln s}\right)\left((\ln t)^{\gamma_+}\int_\infty^t \Gamma_+'(u)\,du + (\ln t)^{\gamma_-}\int_\infty^t \Gamma_-'(u)\,du + h(t)\right) = (\ln s)^{\gamma_+}\int_\infty^s \Gamma_+'(u)\,du + (\ln s)^{\gamma_-}\int_\infty^s \Gamma_-'(u)\,du. \tag{4.22}$$

The proof consists of reversing the order of u and t integrations on the left-hand side of (4.22). After the reversal, the t integration may be performed explicitly. The identities $\gamma_+\gamma_- = \beta$ and $\gamma_+ + \gamma_- = -1$ are used in a subsequent calculation to verify (4.22).

We have now proved (in the case $\beta \neq 1/4$) that $(1 - K_1)^{-1}$ exists and has the explicit representation

$$(1 - K_1)^{-1}h(s) = \frac{\beta}{\gamma_+ - \gamma_-}\int_s^\infty \frac{dt}{t}\left(\frac{(\ln s)^{\gamma_+}}{(\ln t)^{1+\gamma_+}} - \frac{(\ln s)^{\gamma_-}}{(\ln t)^{1+\gamma_-}}\right)h(t) + h(s), \tag{4.23}$$

where

$$\gamma_\pm = -\tfrac{1}{2}[1 \pm (1 - 4\beta)^{1/2}], \quad \beta = b/\pi \neq \tfrac{1}{4}.$$

The inverse of $1 - K_1$ exists just as well in the case $\beta = 1/4$. It is easy to guess the formula for this case, by formally taking the limit $\gamma_+ \to \gamma_- = -\frac{1}{2}$ in (4.23). We find

$$(1 - K_1)^{-1}h(s) = \frac{1}{4\ln^{1/2} s}\int_s^\infty \frac{dt}{t\ln^{1/2} t} \times (\ln\ln s - \ln\ln t)h(t) + h(s), \quad \beta = \tfrac{1}{4}. \tag{4.24}$$

One easily verifies that the right side of (4.24) is a member of S. One can also show, by changing integration order as before, that this function satisfies (4.2). It remains to show that this solution of (4.2) is unique. We demonstrate uniqueness as before, by proving that (4.9) has no nonzero solution in S. To find the general solution of (4.9), we must find a solution in addition to $\chi(s) = \ln^{1/2} s$. We substitute

$$\chi(s) = \ln^{1/2} s\;\xi(s) \tag{4.25}$$

in (4.9) to obtain

$$(1 + \ln s)\,\xi'(s) + s\ln s\,\xi''(s) = 0, \tag{4.26}$$

which is satisfied by $\xi(s) = \ln\ln s$. The general solution of (4.9) is then

$$\chi(s) = a_1 \ln^{1/2} s + a_2 \ln^{1/2} s \cdot \ln\ln s. \tag{4.27}$$

Since the only function of this form in S is identically

zero, it follows that the inverse of $1-K_1$ is indeed given by (4.24).

Our explicit construction of the inverse of $1-K_1$ depended on the special circumstance that there was an equivalent differential equation (4.7), which had a simple solution in closed form. It would be worthwhile to have a more general method, which could be used for singular integral equations not precisely the same as the one discussed here. One method that comes to mind immediately is a simple iterative solution of (4.2):

$$\psi_0 = h, \quad \psi_1 = K_1\psi_0 + h, \quad \cdots, \quad \psi_n = K_1\psi_{n-1} + h, \quad \cdots. \tag{4.28}$$

According to the contraction mapping theorem,[10] the sequence (4.28) converges to the unique solution of (4.2) in S, provided that

$$\|K_1\| < 1, \tag{4.29}$$

where

$$\|K_1\| = \sup_{\phi \in S} (\|K_1\phi\|/\|\phi\|). \tag{4.30}$$

To find a sufficient condition for (4.29) to hold, observe that

$$\begin{aligned}\|K_1\phi\| &= \sup_{s\geq 4} \left| \beta \ln^\alpha s \int_s^\infty \left(\frac{1}{\ln t} - \frac{1}{\ln s}\right) \frac{\phi(t)\,dt}{t} \right| \\ &\leq \beta \|\phi\| \sup \left| \ln^\alpha s \int_s^\infty \left(\frac{1}{\ln t} - \frac{1}{\ln s}\right) \frac{dt}{t \ln^\alpha t} \right| \\ &= \frac{\beta}{\alpha(\alpha-1)} \|\phi\|. \end{aligned} \tag{4.31}$$

The iterative method certainly succeeds if

$$\beta/(\alpha(\alpha-1) < 1. \tag{4.32}$$

In the problem of Ref. 2, one has $\beta/\alpha(\alpha-1) = 1/5$.

Finally, let us return to the rearranged integral equation (4.1). The inhomogeneous term, $(1-K_1)^{-1}h$, is a member of S. Also, the operator $(1-K_1)^{-1}K_2$ is compact on S, being the product of a compact operator and a bounded linear operator. Consequently, the Fredholm alternative theorem[10] applies to (4.1), so that the latter has a unique solution unless the corresponding homogeneous equation has a nontrivial solution. The homogeneous equation is equivalent to

$$\phi = K\phi = (K_1 + K_2)\phi, \tag{4.33}$$

so that we can say that at least part of the Fredholm alternative theorem is true for the non-Fredholm equation (2.9). Namely, (2.9) has a unique solution unless the corresponding homogeneous equation (4.33) has a nontrivial solution.

*Supported in part by the National Science Foundation.

[1]For instance, a Regge theory in which the total cross section tends to a constant at high energy leads to a regular Fredholm N/D equation. See R. L. Warnock, in *Lectures in Theoretical High Energy Physics*, edited by H. H. Aly (Wiley-Interscience, New York, 1968), and G.R. Bart and R. L. Warnock, Phys. Rev. D 3, 1429 (1971).

[2]P.W. Johnson and R. L. Warnock, to be published.

[3]Such behavior of the total cross section, suggested by the trend of experiments at the highest available energies, will probably demand a less rapid decrease of partial waves at high energy than that discussed in Ref. 1. Singularity of the N/D equation is a corollary of weak decrease of the partial waves.

[4]The study in question has to do with determination of the scattering amplitude from a given differential cross section and the unitarity condition, under the constraint that the amplitude be analytic in the cut $\cos\theta$ plane (θ being the scattering angle). See G.R. Bart, P.W. Johnson, and R. L. Warnock, Nucl. Phys. B 72, 329 (1974).

[5]D. Atkinson and A. P. Contogouris, Nuovo Cimento 39, 1082, 1102 (1965).

[6]A. P. Contogouris and A. Martin, Nuovo Cimento A 49, 61 (1967).

[7]C. E. Jones, Nuovo Cimento A 40, 761 (1965).

[8]N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, Holland, 1953).

[9]R. L. Warnock, Nuovo Cimento 50, 894 (1967), and Erratum, *ibid.* A 52, 637 (1967).

[10]L.A. Liusternik and V.J. Sobolev, *Elements of Functional Analysis* (Ungar, New York, 1961), §29, §10.

[11]L.V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces* (Pergamon, Oxford, 1964).

[12]M.A. Krasnosel'skii, G.M. Vainikko, P.P. Zabreiko, Ya.B. Rutitskii, and V.Ya. Stetsenko, *Approximate Solution of Operator Equations* (Walters-Noordhoff, Groningen, 1972), p. 151.

Regge amplitudes through solution of *S*-matrix equations*

Jørgen S. Frederiksen

Research School of Physical Sciences, The Australian National University, Canberra, Australia 2600

Porter W. Johnson and Robert L. Warnock

Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 28 May 1975)

This work is a first step in a program for construction of meson–meson scattering amplitudes with analyticity, crossing symmetry, and unitarity. The construction is to be carried out by solving a nonlinear integral equation for the partial wave amplitude $a(l,s)$ at complex l and physical s. The program is intended to overcome the difficulties encountered in the traditional approach based on the Mandelstam iteration of double-spectral functions. An important initial step is to analyze nonrelativistic potential scattering from this autonomous S-matrix viewpoint, in which the Schrödinger equation is replaced by a nonlinear equation for the partial wave amplitude. In the present paper, it is demonstrated that the partial-wave equation has a locally unique solution, provided the potential is of a suitably restricted Yukawa type. This result indicates the feasibility of a pure S-matrix approach to dynamics. In the present report, the potential is so restricted in strength as to preclude bound states or resonances. The extension of the method to the case of strong potentials will be pursued in a later publication.

1. INTRODUCTION

This paper pertains to the construction of scattering amplitudes with Mandelstam analyticity and unitarity. The construction proceeds by solving a certain nonlinear functional equation for the partial-wave scattering amplitude $a(l,s)$, in which the angular momentum l is complex, but the squared-energy s is restricted to its physical region.[1–3] In Ref. 2 and the present work, the partial-wave equation for nonrelativistic potential scattering is studied. The discussion of potential scattering by the partial-wave equation is more complicated than the usual treatment based on the Schrödinger equation, but it is interesting as a preamble to a discussion of crossing-symmetric, relativistic scattering. In the relativistic case, the approach via the Schrödinger equation is, of course, not available, but the S matrix equation for the partial wave can still be formulated. The equation is quite similar to the corresponding equation of potential scattering, so that it is reasonable to study the simpler case of potential scattering first. Indeed, one of the most difficult terms in the relativistic equation differs only by a kinematical factor from the corresponding term of potential theory.

The program of constructing crossing-symmetric unitary amplitudes through partial-wave equations was initiated in Refs. 1 and 3, and will be continued in a forthcoming series of papers.[4,5] The long range aims of the study are: (a) to clarify questions of principle (for instance, the question of existence of crossing-symmetric unitary amplitudes with Regge behavior) and (b) to provide a practical means of computing realistic hadronic amplitudes from "first principles." In this context, a computation from first principles is understood to be one in which general requirements such as analyticity, crossing, and unitarity are satisfied precisely, and in which phenomenological input is held to a bare minimum. The partial-wave program evolved from an earlier program in which the complete scattering amplitude $A(s,t)$ was constructed by solving a nonlinear integral equation for the double-spectral function, namely, the Mandelstam unitarity equation.[6] The earlier program was not able to accommodate realistic amplitudes. In particular, it faltered in the realistic case in which Regge poles make large excursions into the right-half l plane. In the presence of Regge poles, the double-spectral function takes on an oscillatory behavior. The oscillations are undoubtedly essential in modulating the high-energy behavior of the unitarity equation, but at the same time they make the analysis of the equation extraordinarily difficult. It turns out that the oscillations are much more tractable if one works with partial-wave amplitudes at complex l, rather than the double-spectral function. The advantage is similar to that which is sometimes gained in working with the Fourier transform of a function, rather than the function itself. In fact, the Watson–Sommerfeld representation of the double-spectral function in terms of partial waves is a kind of Fourier representation, in which a quadratic unitarity product of partial waves is like the Fourier transform. The oscillatory function $P_l(z)$, for $\mathrm{Re}\, l = -\epsilon$, plays the role of the exponential in this transform.

Our method of studying potential scattering consists of the following steps:

(a) Deduce an equation which must be satisfied by the partial waves of any scattering amplitude which satisfies unitarity and has an unsubtracted Mandelstam representation.

(b) Show that the partial-wave equation has a solution which may be constructed by iteration, and which is unique in a certain set $\mathcal{S}$ of functions. This is done under appropriate restrictions on the potential: notably, that it be sufficiently weak, and such that an unsubtracted Mandelstam representation is valid.

(c) Show that the usual partial-wave amplitude obtained from the Schrödinger equation lies in the set $\mathcal{S}$, and coincides, therefore, with the solution obtained in step (b).

Thus, the partial-wave equation may replace the Schrödinger equation as the fundamental dynamical

system of the theory, at least in the case of weak, suitably restricted potentials.

Our real interest, however, is in the regime of strong interactions, in which the Mandelstam representation requires subtractions and in which it is difficult to carry out the existence proof of step (b). Fortunately, we know that the solution of the Schrödinger equation is a continuous function of the potential strength. A strength parameter λ can be inserted as a factor in the potential, and continuation of the partial-wave amplitude from small to large λ presents no difficulty. Correspondingly, one expects that it should be possible to do the continuation to large λ in a pure S-matrix scheme based on partial-wave equations. The partial-wave scheme cannot be exactly the one studied here, since when the potential strength reaches a certain value, Regge poles of $a(l, s)$ enter the right-half l plane. The Mandelstam representation then requires subtractions, the Watson–Sommerfeld representation has Regge pole terms as well as a "background" integral, and the equations require modification. Whatever the difficulties of the strong-coupling case, the weak-coupling study carried out here is a necessary preliminary exercise, since it is prudent to start in a regime where the problem can be analyzed completely, and since certain problems of asymptotic behavior in l are probably common to the strong- and weak-coupling cases.

One possible avenue to the strong-coupling regime, which is being studied by Johnson and Warnock,[4] is to replace the partial-wave dispersion relation of the present paper by an equivalent N/D equation. That method promises to allow a smooth continuation to large λ, thus providing an autonomous S-matrix theory for arbitrary potential strength, in which scattering amplitudes and Regge trajectories could be computed without appeal to the Schrödinger equation. One expects that a similar continuation from weak to strong coupling will be possible in the relativistic case, where coupling strength is gauged in terms of an appropriate generalized potential.[4] Another approach is being followed by Atkinson, Frederiksen, and Kaekebeke,[5] who investigate the existence of crossing-symmetric, unitary amplitudes with Regge trajectories and residues which are specified *ab initio*.

In Ref. 2, the potential was required to have energy dependence, in such a way that it became weaker at high energy. In the present work we remove that restriction, and deal with an ordinary energy-independent potential of Yukawa type. We were able to remove the restriction through a technical improvement in our treatment of the oscillations of Legendre functions. This improvement, which we think to be quite crucial for the further development of Regge theory, should also allow one to remove some of the restrictions imposed in Ref. 3. In Ref. 3, crossing-symmetric partial-wave equations for relativistic meson–meson scattering were analyzed. The method of analysis required a cutoff of the elastic unitarity term at high energy. Such a cutoff does not destroy crossing symmetry or elastic unitarity, the latter being required only at low energy. It would be interesting, nevertheless, to see if the cutoff could be removed. It is likely that our new method of handling Legendre functions will allow removal of the cutoff and also lead to certain simplifications in the discussion of Ref. 3.

In Sec. 2 we establish notation and carry out step (a) mentioned above; i.e., we derive the dynamical equation for the partial-wave amplitude. The equation is a partial-wave dispersion relation, in which the left cut term is expressed as a nonlinear functional of the partial-wave amplitude itself. Our expression for the left cut term is simpler than the conventional one,[7] and involves fewer integrations.

Section 3 is devoted to step (b); that is, we prove the existence of a locally unique solution of the equation, with a specified potential subject to certain conditions.

In Sec. 4, we show that the solution of the partial-wave equation actually leads to a unitary amplitude with a Mandelstam representation, and discuss briefly the relation of this work to earlier work on potential scattering.[8,9]

2. DYNAMICAL EQUATION FOR THE PARTIAL-WAVE AMPLITUDE

We discuss the non-relativistic interaction of two spinless particles of equal mass m, interacting through a potential $V(r)$ which depends only on the inter-particle distance r. The potential will have the conventional Yukawa form,

$$rV(r)=\frac{1}{2m}\int_{4\mu^2}^{\infty} dt\,\rho(t)\exp(-rt^{1/2}). \tag{2.1}$$

(Note: $\hbar=1$ throughout this work.) The range of the force corresponds to that of a two-particle exchange force, each of the exchanged particles having mass μ. Equivalently, the range is the Compton wavelength of a particle of mass 2μ. Note that we do not require $m=\mu$. The weight function $\rho(t)$ is subject to conditions specified in Sec. 3.

The reason for writing $4\mu^2$ rather than μ^2 as the limit in (2.1) is that we wish our equations to resemble the analogous relativistic equations as much as possible. For the same reason, we define a squared-energy variable s by

$$s=4(q^2+\mu^2), \tag{2.2}$$

where q is the momentum in the center-of-mass frame,

$$q=(2mE)^{1/2}. \tag{2.3}$$

Of course, $s\neq E^2$, in general. Henceforth, we take units such that $\mu=1$. The combination $s-4$ occurs so frequently that we adopt a special notation,

$$\sigma=s-4. \tag{2.4}$$

The squared-momentum-transfer variable t is defined in the usual way,

$$t=-2q^2(1-z),\quad z=\cos\theta, \tag{2.5}$$

where θ is the scattering angle.

It is known[9] that when $\rho(t)$ is suitably restricted, the scattering amplitude $A(s, t)$ obeys a Mandelstam representation. If the potential is so weak that no Regge poles of the partial-wave amplitude enter the right-half

l plane, then the Mandelstam representation requires no subtractions and has the form

$$A(s,t) = \frac{1}{\pi}\int_4^\infty \frac{\rho(t')\,dt'}{t'-t} + \frac{1}{\pi^2}\int_4^\infty ds'\int_4^\infty dt'\,\frac{\rho(s',t')}{(s'-s)(t'-t)}. \quad (2.6)$$

The single-spectral function $\rho(t)$ is identical with the weight function of the Yukawa potential. The region in which $\rho(s,t)$ may be nonzero is determined by the requirement of unitarity. The boundary of the region is often deduced from Mandelstam's form of the unitarity condition for the double-spectral function. It may be obtained directly from our partial-wave viewpoint as follows. By reversing the order of the t' and z integrals, one obtains the Froissart–Gribov form of the partial-wave amplitude:

$$a(l,s) = \frac{1}{2}\int_{-1}^{1} dz\,P_l(z)\,A(s,t) = \frac{2}{\pi\Delta}\int_4^\infty dt\,Q_l\left(1+\frac{2t}{\Delta}\right) D(s,t), \quad (2.7)$$

$$D(s,t) = \rho(t) + \frac{1}{\pi}\int_4^\infty \frac{ds'\rho(s',t)}{s'-s}. \quad (2.8)$$

The reversal in integration order is justified for the amplitudes we construct (see Sec. 4). From the known behavior of $Q_l(z)$ at large positive l and a suitable bound on the t-discontinuity $D(s,t)$, one concludes that the partial wave is bounded as follows at positive l:

$$|a(l,s)| \le \kappa[z_0 + (z_0^2-1)^{1/2}]^{-l}, \quad (2.9)$$

$$z_0 = 1 + 8/\Delta.$$

Here and in the following work, κ is a generic positive constant which may have different values at different points in the argument. In Sec. 3, we show that (2.9) holds for our amplitudes. By unitarity, the absorptive part of the amplitude has the form

$$A_s(s,t) = (1/2i)[A(s_+,t) - A(s_-,t)] = \sum_{l=0}^{\infty}(2l+1)\,P_l(1+2t/\Delta)\,q(s)\,a(l,s_+)\,a(l,s_-), \quad (2.10)$$

where

$$f(s_+) = \lim_{\omega\to s+i0} f(\omega).$$

By (2.6), A_s may also be written as

$$A_s(s,t) = \frac{1}{\pi}\int_4^\infty \frac{dt'\rho(s,t')}{t'-t}. \quad (2.11)$$

Since

$$|P_l(z)| \le \kappa[z + (z^2-1)^{1/2}]^l \quad (2.12)$$

at positive l, we see from (2.9) that the series (2.10) converges absolutely and uniformly in any closed region interior to the unifocal ellipse defined by the equation

$$z + (z^2-1)^{1/2} = [z_0 + (z_0^2-1)^{1/2}]^2, \quad (2.13)$$

$$z = 1 + 2t/\Delta.$$

That is to say, $A_s(s, -2q^2(1-z))$ is analytic in a unifocal ellipse in the z plane, with semimajor axis $2z_0^2 - 1$. When this assertion of analyticity is compared with (2.11), we see that the cut of A_s in the t plane must start at $t \ge 16s/\Delta$, since $z < 2z_0^2 - 1$ is equivalent to $t < 16s/\Delta$. Thus, $\rho(s,t)$ is zero unless

$$s \ge 4, \quad t \ge \tau(s) = 16s/\Delta. \quad (2.14)$$

We shall now derive the dynamical equation by a series of formal steps. The manipulations will be formal in the sense that we shall not pay attention, initially, to all questions of convergence, etc. Having obtained the equation, we shall then prove rigorously that it has a unique solution in a certain set. From this solution we construct the complete scattering amplitude $A(s,t)$, and show that the latter is unitary and satisfies an unsubtracted Mandelstam representation. That is, the amplitude $A(s,t)$ is the solution of our original problem.

The derivation of the equation starts from the observation that the double-spectral function may be expressed in terms of partial waves at complex l. If a Watson–Sommerfeld transformation is applied to (2.10), one obtains the expression

$$A_s(s,t) = \frac{i}{2}\int_{\mathrm{Re}\,l=w} \frac{dl(2l+1)}{\sin\pi l}\,P_l(-z)\,q(s)\,a(l,s_+)\,a(l,s_-). \quad (2.15)$$

We take $-\frac{1}{2} < w < 0$. According to (2.11), the discontinuity of (2.15) over its cut in the t plane is the double-spectral function,[9]

$$\rho(s,t) = \frac{1}{2i}\int_{\mathrm{Re}\,l=w} dl(2l+1)\,P_l(z)\,q(s)\,a(l,s_+)\,a(l,s_-). \quad (2.16)$$

If $A(s,t)$ is real analytic in s and t, it follows from (2.7) that

$$a(l,s_-) = [a(l^*,s_+)]^*. \quad (2.17)$$

Henceforth, it will be understood that $a(l,s_-)$ is defined by (2.17). Consequently, the function $\rho(s,t)$ as expressed in (2.16) will be real. Equations (2.7), (2.8), and (2.16) taken together constitute a nonlinear system satisfied by $a(l,s)$. We take advantage of the support of ρ, as specified in (2.14), to replace the lower limit of the integral (2.8) by

$$\sigma(t) = 4t/(t-16). \quad (2.18)$$

(The function σ is the inverse of τ). The Froissart–Gribov representation (2.7) can now be stated as

$$a(l,s_+) = a_B(l,s) + \frac{2}{\pi\Delta}\int_4^\infty dt\,Q_l\left(1+\frac{2t}{\Delta}\right) \times\frac{1}{\pi}\int_{\sigma(t)}^\infty \frac{ds'}{s'-s_+}\,\frac{1}{2i}\int_{\mathrm{Re}\,l'=-\epsilon} dl'(2l'+1) \times P_{l'}\left(1+\frac{2t}{\Delta'}\right) q(s')\,a(l',s'_+)\,a(l',s'_-), \quad (2.19)$$

where a_B is the Born term,

$$a_B(l,s) = \frac{2}{\pi\Delta}\int_4^\infty dt\,Q_l\left(1+\frac{2t}{\Delta}\right)\rho(t), \quad (2.20)$$

and $0 < \epsilon < \frac{1}{2}$. Equation (2.19) may be regarded as a dynamical equation for $a(l,s_+)$, with $\mathrm{Re}\,l = -\epsilon$ and $s \ge 4$. The equation as it stands is difficult to analyze, however, so that we have chosen to study an equivalent rearranged

form of the equation. In the new arrangement we carry out the t integration first and obtain (for $\mathrm{Re}\,l > -\epsilon$),

$$a(l,s_+) = a_B(l,s) + \frac{1}{\pi}\int_4^\infty \frac{ds' q(s')}{s'-s_+}\int_{\mathrm{Re}\,l'=-\epsilon} dl'(2l'+1) \times a(l',s'_+)\,a(l',s'_-)\,\Lambda(l,l',s,s'), \tag{2.21}$$

$$\Lambda(l,l',s,s') = \frac{1}{i\pi\Delta}\int_{\tau(s')}^\infty dt\, Q_l\left(1+\frac{2t}{\Delta}\right) P_{l'}\left(1+\frac{2t}{\Delta'}\right), \tag{2.22}$$

where $\tau(s')$ was defined in Eq. (2.14). Equation (2.21) is suitable for our purposes in the open region $\mathrm{Re}\,l > -\epsilon$, but for $\mathrm{Re}\,l = -\epsilon$ we must use a different expression since the integral (2.22) diverges when $l = l'$. We proceed by separating the part of Λ which diverges at $l = l'$. We write $\Lambda = \Lambda_1 + \Lambda_2$, where

$$\Lambda_1 = \frac{1}{i\pi\Delta}\left(\frac{\Delta}{\Delta'}\right)^{l+1}\int_{\tau(s')}^\infty dt\, Q_l\left(1+\frac{2t}{\Delta'}\right) P_{l'}\left(1+\frac{2t}{\Delta'}\right), \tag{2.23}$$

$$\Lambda_2 = \frac{1}{i\pi\Delta}\int_{\tau(s')}^\infty dt\, P_{l'}\left(1+\frac{2t}{\Delta'}\right)\left[Q_l\left(1+\frac{2t}{\Delta}\right) - \left(\frac{\Delta}{\Delta'}\right)^{l+1} Q_l\left(1+\frac{2t}{\Delta'}\right)\right]. \tag{2.24}$$

The divergent part Λ_1 may be calculated in terms of Legendre functions, by using the Legendre differential equation and partial integration. The result is

$$\Lambda_1 = \frac{1}{2\pi i}\left(\frac{\Delta}{\Delta'}\right)^{l}\frac{1-z_1^2}{(l'-l)(l'+l+1)} \times [Q_l(z_1)\,\partial_z P_{l'}(z_1) - P_{l'}(z_1)\,\partial_z Q_l(z_1)], \quad z_1 = 1 + 2\tau(s')/\Delta'. \tag{2.25}$$

Standard asymptotic estimates on Q_l show that the integral (2.24) converges for $\mathrm{Re}\,l = \mathrm{Re}\,l'$; see Appendix C. For the term Λ_1, the l' integration in (2.21) may be carried out by closing the l' contour by an infinite semicircle in the right half-plane. The contribution of the semicircle vanishes if the functions

$$a(l',s'_+)\,a(l',s'_-)\,P_{l'}(z_1) \tag{2.26}$$

and

$$a(l',s'_+)\,a(l',s'_-)\,\partial_z P_{l'}(z_1) \tag{2.27}$$

vanish, uniformly in direction, as $|l'|$ tends to infinity. This requirement is satisfied for any $a(l,s_+)$ of the class considered in Sec. 3. The contribution of Λ_1 to the integral (2.21) may be evaluated as the residue of the pole at $l = l'$; if one notes the value of the Wronskian[10] of P_l and Q_l, one obtains

$$\int_4^\infty \frac{ds'}{s'-s_+}\left(\frac{\Delta}{\Delta'}\right)^{l} q(s')\,a(l,s'_+)\,a(l,s'_-). \tag{2.28}$$

This is the familiar right-cut term of a dispersion relation for the reduced partial wave amplitude[11]

$$b(l,s) = a(l,s)\Delta^{-1}. \tag{2.29}$$

The new form of (2.21), obtained by the decomposition of Λ, is best stated in terms of the reduced amplitude. We write

$$b_+(l,s) = F(b_+;l,s), \tag{2.30}$$

where

$$b_+(l,s) = b(l,s_+) \tag{2.31}$$

and

$$F(b_+;l,s) = b_B(l,s) + \frac{1}{\pi}\int_4^\infty \frac{ds'}{s'-s_+} q(l,s')\,b_+(l,s')\,b_-(l,s') + \frac{1}{\pi}\int_4^\infty \frac{ds'}{s'-s}\int_{\mathrm{Re}\,l'=-\epsilon} dl'(2l'+1)\,q(l',s') \times b_+(l',s')\,b_-(l',s')\,G(l,l',s,s'), \quad \mathrm{Re}\,l \geq -\epsilon, \quad q(l,s) = q(s)\Delta^{-1}, \quad b_-(l,s) = [b_+(l^*,s)]^*. \tag{2.32}$$

The function G differs from Λ_2 just by a kinematical factor:

$$G(l,l',s,s') = \frac{\Delta'^{l'}}{i\pi}\int_{\tau(s')}^\infty dt\, P_{l'}\left(1+\frac{2t}{\Delta'}\right)\left[\frac{1}{\Delta^{l+1}} Q_l\left(1+\frac{2t}{\Delta}\right) - \frac{1}{\Delta'^{l+1}} Q_l\left(1+\frac{2t}{\Delta'}\right)\right]. \tag{2.33}$$

For the open region $\mathrm{Re}\,l > -\epsilon$, we have the alternative form of $F(b_+;l,s)$ by (2.21):

$$F(b_+;l,s) = b_B(l,s) + \frac{\Delta^{-l}}{\pi}\int_4^\infty \frac{ds' q(s')}{s'-s_+}\int_{\mathrm{Re}\,l'=-\epsilon} dl' \times (2l'+1)\,(\Delta')^{2l'}\,b_+(l',s')\,b_-(l',s')\,\Lambda(l,l',s,s'). \tag{2.34}$$

In Sec. 3 we shall determine a solution of (2.30), which is to say that we shall find a fixed point of the nonlinear mapping F in a certain function space. The space will be such that when b_+ is one of its elements, the two forms of F, (2.32) and (2.34), are equivalent for $\mathrm{Re}\,l > -\epsilon$.

Equations (2.30), with F in the form (2.32), is a conventional partial-wave dispersion relation for b, although the left-cut term, the third term in F, appears in an unfamiliar form. The function G is a "known" function, in the sense that it is independent of dynamics and may be computed once and for all. In computing a solution of the equation only two integrations (over s' and l') are required to evaluate the left-cut term. In the conventional form of the left-cut contribution,[7] four integrations are required.

3. EXISTENCE AND CONSTRUCTION OF SOLUTIONS

In this section we prove that the dynamical equation (2.30) has a locally unique solution in a Banach space $\mathcal{B}$, when the potential is suitably restricted. The space $\mathcal{B}$ consists of all complex functions $f(l,s)$ that are analytic in l for $\mathrm{Re}\,l > -\epsilon$, continuous in l for $\mathrm{Re}\,l = -\epsilon$, continuous in s for $s > 4$, and such that the following number, the norm in $\mathcal{B}$, is finite:

$$\|f\| = \|f\|_1 + \|f\|_2. \tag{3.1}$$

The subnorms over the Region 1 ($-\epsilon \leq \mathrm{Re}\,l \leq -\zeta$) and the Region 2 ($-\zeta < \mathrm{Re}\,l$) are defined by

$$\|f\|_1 = \sup_{-\epsilon \le w \le -\zeta} \left(\sup_{s \ge 4} s^{\lambda+w} l_+^{\nu} |f(l,s)| + \sup_{s' > s \ge 4} s^{\lambda+w+\delta} l_+^{\nu-\delta} \frac{|f(l,s') - f(l,s)|}{|s' - s|^{\delta}} \right), \tag{3.2}$$

and

$$\|f\|_2 = \sup_{w > -\zeta} \sup_{s \ge 4} s^{\lambda} l_+^{1/2} [p(s)]^{w} |f(l,s)|. \tag{3.3}$$

We have defined the quantities

$$w = \mathrm{Re} l, \tag{3.4}$$

$$l_+ = 1 + |l|, \tag{3.5}$$

and

$$p(s) = (s^{1/2} + 2)^2. \tag{3.6}$$

Also, we shall define and use the function

$$u(\sigma, t) = (\sqrt{t} + \sqrt{\sigma + t})^2; \tag{3.7}$$

note that

$$p(s) = u(s, 4). \tag{3.8}$$

Note that the elements of β decrease as $l_+^{-\nu}$ as $l \to \infty$ within Region 1 and as $l_+^{-1/2}$ as $l \to \infty$ within Region 2; in addition they decrease exponentially as $\mathrm{Re} l \to \infty$.

The positive quantities ϵ, ν, λ, ζ, and δ appearing in the norm (3.1) are subject to the following constraints:

$$\epsilon + 2\delta < \tfrac{1}{2}, \quad \zeta + 2\delta < \epsilon, \quad \tfrac{3}{4} + \delta/2 < \nu < 1, \quad \tfrac{1}{2} < \lambda < \tfrac{5}{8} - \delta/4, \quad \nu/2 + \lambda \le 1. \tag{3.9}$$

In other respects, they may be chosen arbitrarily. By contrast, in Ref. 2 it was required to take ν ($\nu + \frac{1}{2}$ in that notation) to be greater than 1, and the Born term b_B was not an element of the Banach space for an energy-independent potential. Our Born term will belong to β. In fact, the last of the conditions (3.9) is imposed by the requirement that $b_B \in \beta$. When $b \in \beta$, the functions (2.26) and (2.27) vanish as $|l'|$ tends to infinity, uniformly in direction in the right-half l' plane. This follows from the upper bounds on $P_{l'}(z)$ and $\partial_z P_{l'}(z)$ given in Appendix A of Ref. 2. Consequently, the two forms of F, stated in (2.32) and (2.34), are equivalent when $b \in \beta$.

In proving various central results of this paper, we shall need to use certain special properties of Legendre functions of the second kind. To describe these properties, let us define a function $\overline{Q}_l(\sigma, t)$ by

$$Q_l(1 + 2t/\sigma) = \sigma^{l+1} \overline{Q}_l(\sigma, t). \tag{3.10}$$

Using the standard Laplace representation[12] of Q_l, we may represent $\overline{Q}_l$ as

$$\overline{Q}_l(\sigma, t) = \int_0^{\infty} d\theta \{\sigma + 2t + 2[t(\sigma + t)]^{1/2} \cosh\theta\}^{-(l+1)}. \tag{3.11}$$

It is convenient to factor the value of the integrand at $\theta = 0$ from the integral representation (3.11) to obtain

$$\overline{Q}_l(\sigma, t) = u^{-(l+1)} R_l(x), \tag{3.12}$$

where $u(\sigma, t)$ is given by (3.7),

$$x(\sigma, t) = (2/u)[t(\sigma + t)]^{1/2}, \tag{3.13}$$

and

$$R_l(x) = \int_0^{\infty} d\theta [1 + x(\cosh\theta - 1)]^{-(l+1)}. \tag{3.14}$$

Note that for σ and t positive, x lies between 0 and $\frac{1}{2}$. We show in Appendix A that, for $\mathrm{Re} l > -1$ and $0 < x \le \frac{1}{2}$, the function $R_l(x)$ and its x derivative satisfy the bounds

$$|R_l(x)| \le \kappa (l_+ x)^{-1/2}, \tag{3.15}$$

$$|R_l'(x)| \le \kappa (l_+ x^3)^{-1/2}, \tag{3.16}$$

and

$$|R_l''(x)| \le \kappa (l_+ x^5)^{-1/2}. \tag{3.17}$$

Let us point out that, whereas the z derivatives of $Q_l(z)$ become unbounded at large l_+, the transformations (3.10) and (3.12) permit us to separate out an oscillating part, such that the remainder $R_l(x)$ has bounded x derivatives. The Legendre function of the first kind, $P_l(z)$, may be expressed in terms of $Q_l(z)$ and $Q_{-l-1}(z)$ by a standard identity quoted in Appendix C, Eq. (C2). This identity yields the bound

$$|P_l(1 + 2t/\sigma)| \le \kappa (u(\sigma, t)/\sigma)^{w} (l_+ x)^{-1/2}, \quad -\tfrac{1}{2} \le w \le -\zeta. \tag{3.18}$$

Similarly, one may obtain bounds on the derivatives of P_l.

We shall verify that F maps a certain ball of β into itself, in a weak-coupling regime in which certain restrictions are placed upon the potential $V(r)$. Specifically, let us require that the real-valued function $\rho(t)$ be differentiable for $t \ge 4$ and be subject there to bounds

$$|\rho(t)| \le c t^{-\alpha} |(t-4)/t|^{1/2} \tag{3.19}$$

and

$$|\rho'(t)| \le c t^{-\alpha-1} |t/(t-4)|^{1/2}, \tag{3.20}$$

with $\alpha > \frac{3}{4}$. Furthermore, we shall place a restriction to be specified presently upon the magnitude of the constant c. The form of the bounds upon $|\rho|$ and $|\rho'|$ is chosen to include the case in which $\rho(t) \sim \kappa(t-4)^{1/2}$ as $t \to 4+$. One might expect that an effective potential based upon two-particle exchanges would lead to a Yukawa density ρ with such a crossed-channel threshold factor. We are not able to handle the case of "one-particle exchange," in which $\rho(t) = \delta(t-4)$, by the current procedure.

Under conditions (3.9), (3.19), and (3.20), one may easily demonstrate that $b_B(l,s)$ and $\partial_s b_B(l,s)$ are analytic in l for $\mathrm{Re} l > -\epsilon$, continuous in l along $\mathrm{Re} l = -\epsilon$, and continuous in $s \ge 4$. In fact, we establish the following bounds in Appendix B for $s \ge 4$ and $w = \mathrm{Re} l \ge -\epsilon$:

$$|b_B(l,s)| \le \frac{\kappa c}{l_+^{1/2} s^{3/4}} \frac{1}{[p(s)]^{w}}, \tag{3.21}$$

$$|b_B(l,s)| \le \frac{\kappa c}{l_+^{3/2} s^{1/4}} \frac{1}{[p(s)]^{w}}, \tag{3.22}$$

and

$$|\partial_s b_B(l,s)| \le \frac{\kappa c}{l_+^{1/2} s^{5/4}} \frac{1}{[p(s)]^{w}}. \tag{3.23}$$

The constant κ appearing in (3.21)–(3.23) is independent of ρ, l, and s. The bounds (3.21)–(3.23) are sufficient to guarantee that the function $b_B(l,s)$ is in the Banach

space β, with any choice of $(\epsilon, \lambda, \nu, \zeta, \delta)$ subject to (3.9); and that its norm is subject to the bound

$$\|b_B\| \leq \kappa_B c, \tag{3.24}$$

the constant κ_B being independent of ρ, and therefore of c.

Let us express Eq. (2.30) as

$$b_+ = b_B + K(b_+). \tag{3.25}$$

We shall show that the operator K maps the space β into itself, in such a way that

$$\|K(b_+)\| \leq \kappa \|b_+\|^2, \tag{3.26}$$

where κ is independent of b_+. We define a closed ball S, centered at the origin in β, as the set of all b_+ such that

$$\|b_+\| \leq r. \tag{3.27}$$

By an analysis similar to that which yields (3.26), one may show that, for b_{+1}, $b_{+2} \in S$,

$$\|K(b_{+1}) - K(b_{+2})\| \leq \kappa r \|b_{+1} - b_{+2}\|. \tag{3.28}$$

From (3.24)—(3.28), one sees that if c is sufficiently small, one may choose a correspondingly small ball radius r such that $F = b_B + K$ is a contraction mapping of S into itself. It then follows from the contraction mapping theorem[13] that Eq. (2.30) has a unique solution in S and that the solution may be computed by iteration.

In the remainder of this section, we shall establish that K maps β into itself in such a way that (3.26) is valid. We shall prove first that $K(b_+; l, s)$ is analytic in l and continuous in s, for $\mathrm{Re}\, l > -\epsilon$, and that $\|K\|_2 \leq \kappa \|b_+\|^2$. This will be done by using the second form of K, as given in (2.34). We shall then demonstrate, with the help of the first form of K, (2.32), that $K(b_+; l, s)$ is continuous in l and s for $\mathrm{Re}\, l = -\epsilon$, and that the subnorm over Region 1 obeys the bound $\|K\|_1 \leq \kappa \|b_+\|^2$. The bound (3.26) will then follow.

The analysis of this section employs information upon $b_+(l, s)$ itself only for l in Region 1. But, of course, the mappings (2.32) and (2.34) are equivalent only if b_+ decreases exponentially in w in Region 2, as is the case when b_+ is in the Banach space β.

We shall make use of the following bounds upon $b_+(l, s) \in \beta$ for l in Region 1 $(-\epsilon \leq w \leq -\zeta)$:

$$|b_+(l, s)| \leq \frac{\|b_+\|}{l_+^{\nu} s^{\lambda + w}} \tag{3.29}$$

for $s \geq 4$, and

$$|b_+(l, s') - b_+(l, s)| \leq \|b_+\| \, |s' - s|^{\delta} / l_+^{\nu - \delta} s^{\lambda + w + \delta} \tag{3.30}$$

for $s' \geq s \geq 4$.

To analyze $K(b_+)$, it is convenient to define the function

$$C(l, s) = q(l, s)\, b_+(l, s)\, b_-(l, s). \tag{3.31}$$

It follows from (3.29) and (3.30) that C is subject to the following bounds for l in Region 1:

$$|C(l, s)| \leq \frac{\kappa \|b_+\|^2}{l_+^{2\nu}} \left(\frac{\delta}{s}\right)^{1/2 + w} \frac{1}{s^{2\lambda - 1/2 + w}}, \quad s \geq 4, \tag{3.32}$$

$$|C(l, s') - C(l, s)| \leq \frac{\kappa \|b_+\|^2}{l_+^{2\nu - \delta}} \left(\frac{\delta'}{s'}\right)^{1/2 + w - \delta} \frac{|s' - s|^{\delta}}{s^{2\lambda - 1/2 + w + \delta}}, \quad s' \geq s \geq 4. \tag{3.33}$$

Let us cast the expression (2.34) for $K(b_+)$ into the form

$$K(b_+) = \frac{P}{\pi \delta'} \int_4^{\infty} \frac{ds'}{s' - s} H(l, s, s') + \frac{i}{\delta'} H(l, s, s), \tag{3.34}$$

where

$$H(l, s, s') = \int_{\mathrm{Re}\, l' = -\epsilon} dl' (2l' + 1) C(l', s') (\delta')^{l'} \Lambda(l, l', s, s'), \tag{3.35}$$

with Λ defined by (2.22). The integral (2.22) is absolutely convergent when $w = \mathrm{Re}\, l > -\epsilon$ and $\mathrm{Re}\, l' = -\epsilon$. We may directly apply (3.10)—(3.18) to establish the following bounds:

$$|\delta'^{l'} \Lambda(l, l', s, s')| \leq \frac{\kappa}{(w + \epsilon)(l_+' l_+)^{1/2}} \left[1 + \left(\frac{s'}{s}\right)^{1/4 - \epsilon}\right] \frac{1}{s^{\epsilon}} \left(\frac{\delta}{p(s)}\right)^{w}, \tag{3.36}$$

$$|\partial_{s'}(\delta'^{l'} \Lambda(l, l', s, s'))| \leq \frac{\kappa}{w + \epsilon} \left(\frac{l_+'}{l_+}\right)^{1/2} \frac{1}{s^{\epsilon}} \left(\frac{\delta}{p(s)}\right)^{w} \frac{s'}{(\delta')^2} \left[1 + \left(\frac{s'}{s}\right)^{1/4 - \epsilon}\right]. \tag{3.37}$$

We use (3.32), (3.33), and (3.35)—(3.37) to get the following bounds for the function $H(l, s, s')$:

$$|H(l, s, s')| \leq \frac{\kappa \|b_+\|^2}{l_+^{1/2}(w + \epsilon)} \frac{1}{s^{\epsilon}} \left(\frac{\delta}{p(s)}\right)^{w} \left[1 + \left(\frac{s'}{s}\right)^{1/4 - \epsilon}\right] \times \left(\frac{\delta'}{s'}\right)^{1/2 - \epsilon} \frac{1}{(s')^{2\lambda - \epsilon - 1/2}}. \tag{3.38}$$

and, for $s_1' < s_2'$,

$$|H(l, s, s_1') - H(l, s, s_2')| \leq \frac{\kappa \|b_+\|^2}{l_+^{1/2}(w + \epsilon)} \frac{1}{s^{\epsilon}} \left(\frac{\delta}{p(s)}\right)^{w} \left[1 + \left(\frac{s_1'}{s}\right)^{1/4 - \epsilon} + \left(\frac{s_2'}{s}\right)^{1/4 - \epsilon}\right] \times \left(\frac{\delta_2'}{s_2'}\right)^{1/2 - \epsilon - 2\delta} \frac{1}{(s_1')^{2\lambda - \epsilon - 1/2 + \delta}} |s_1' - s_2'|^{\delta}. \tag{3.39}$$

We may use the bounds (3.38) and (3.39) on the principal value integral in Eq. (3.34), along with Lemma II of Appendix D, to establish that $K(b_+, l, s)$ is continuous in s for $s \geq 4$ and subject to the bound

$$|K(b_+, l, s)| \leq \frac{\kappa}{w + \epsilon} \frac{\|b_+\|^2}{l_+^{1/2}} \frac{1}{s^{2\lambda - 1/2}} [p(s)]^{-w}. \tag{3.40}$$

We have thus established the required bound $\|K\|_2 < \kappa \|b_+\|^2$, and have shown that K is continuous in s. It remains to show that $K(b_+; l, s)$ is analytic in l for $\mathrm{Re}\, l > -\epsilon$. One first notes that $\Lambda(l, l', s, s')$ is analytic in l, for $\mathrm{Re}\, l > -\epsilon$, $\mathrm{Re}\, l' = -\epsilon$, by virtue of uniform convergence of the integral (2.22) that defines Λ. Further,

(3.35) converges uniformly in l, so that $H(l,s,s')$ is analytic for $\mathrm{Re}l > -\epsilon$. To show that the principal-value integral in (3.34) is analytic in l, we write the integral as

$$\frac{1}{\pi s^{1}}\int_4^\infty \frac{ds'}{s'-s}\left[H(l,s,s')-\frac{s}{s'}H(l,s,s)\right] + \frac{s}{\pi s^{1}}H(l,s,s)P\int_4^\infty \frac{ds'}{s'(s'-s)}. \tag{3.41}$$

The first term in (3.41) is convergent uniformly in l, thanks to (3.38) and (3.39), so that the required analyticity is apparent.

We shall now study the expression (2.32) for $K(b_+,l,s)$, which we write as

$$K(b_+)=K_R(b_+)+K_L(b_+), \tag{3.42}$$

for l in Region 1. We shall show that $K(b_+)$ is continuous in l and s for $\mathrm{Re}l=-\epsilon$. In addition, we shall establish these bounds:

$$|K_{R,L}(b_+,l,s)| \le \frac{\kappa\|b_+\|^2}{l_+^{\nu}}\frac{1}{s^{\lambda+w}}, \tag{3.43}$$

and for $s' \ge s$,

$$|K_{R,L}(b_+,l,s')-K_{R,L}(b_+,l,s)| \le \frac{\kappa\|b_+\|^2}{l_+^{\nu-\delta}}\frac{|s'-s|^{\delta}}{s^{\lambda+w+\delta}}. \tag{3.44}$$

That is to say, the subnorm $\|K(b_+)\|_1$ obeys the bound

$$\|K(b_+)\|_1 \le \kappa\|b_+\|^2. \tag{3.45}$$

Let us establish (3.43) and (3.44) for K_R, which is given by the expression

$$K_R(l,s)=\frac{P}{\pi}\int_4^\infty \frac{ds'}{s'-s}C(l,s')+iC(l,s). \tag{3.46}$$

For $\lambda > \frac{1}{2}$, we obtain bounds (3.43) and (3.44) for the second term in (3.46), from (3.32) and (3.33). By a standard lemma on principal-value integrals, which is stated as Lemma I of Appendix D, we may conclude that the first term is continuous in s and subject to the bounds (3.43) and (3.44). Furthermore, one easily establishes that $K_R(l,s)$ is continuous in l and s for $\mathrm{Re}l = -\epsilon$, by uniform convergence.

To analyze K_L, let us define the auxiliary function

$$E(l,s,s') = \int_{\mathrm{Re}l'=\mathrm{Re}l=w} dl'(2l'+1)C(l',s')G(l,l',s',s), \tag{3.47}$$

with G defined by (2.33). Note that the integral (3.47) is equal to the l' integral appearing in (2.32). The contour displacement from $\mathrm{Re}l'=-\epsilon$ to $\mathrm{Re}l'=\mathrm{Re}l$ is permitted by analyticity of the $b_\pm(l,s)$. It is evident from (2.33) that G vanishes for $s=s'$, so that the expression for K_L,

$$K_L(l,s)=\frac{1}{\pi}\int_4^\infty \frac{ds'}{s'-s}E(l,s,s'), \tag{3.48}$$

involves a nonsingular integration over $s=s'$. We establish in Appendix C that G is subject to the following bounds [see Eqs. (C63), (C69), and (C70)]:

$$|G(l,l',s,s')| \le \kappa(l_+l_+')^{-1/2}\ln(s+s')[1+(s'/s)^{1/2+w}] \times[1+\ln l_+l_+'][|y'-y|^{-1/2}+|y'+y|^{-1/2} + |y'-(s'/s)^{1/2}y|^{-1/2}+|y'+(s'/s)^{1/2}y|^{-1/2}], \tag{3.49}$$

$$|\partial_s G(l,l',s,s')| \le [\kappa(l_+l_+')^{1/2}/(l_++l_+')][1+(s'/s)^{1/4+w}] \times[\ln(s'+s)/s](1+\ln l_+l_+'), \tag{3.50}$$

$$|\partial_{s'}G(l,l',s,s')| < [\kappa(l_+l_+')^{1/2}/(l_++l_+')](1+\ln l_+l_+') \times[\ln(s+s')/s'][1+(s'/s)^{3/4+w}]. \tag{3.51}$$

We have adopted the notation $y=\mathrm{Im}l$, $y'=\mathrm{Im}l'$ in writing (3.49). It follows from (3.32), (3.33), (3.47), and (3.49)–(3.51) that E is subject to the following bounds:

$$|E(l,s,s')| \le \frac{\kappa\|b_+\|^2(1+\ln l_+)}{l_+}\left(\frac{s'}{s'}\right)^{1/2+w}\frac{\ln(s+s')}{(s')^{2\lambda+w-1/2}} \times\left[\left(\frac{s}{s'}\right)^{1/4}+\left(\frac{s'}{s}\right)^{1/2+w}\right], \tag{3.52}$$

$$|E(l,s_1,s')-E(l,s_2,s')| \le \frac{\kappa\|b_+\|^2(1+\ln l_+)}{l_+^{1+\delta-3/2\delta'}}\left(\frac{s'}{s'}\right)^{1/2+w}\frac{\ln(s_2+s')}{(s')^{2\lambda+w-1/2}} \times\left[\left(\frac{s_2}{s'}\right)^{1/4}+\left(\frac{s'}{s_1}\right)^{1/2+w}\right]\left|\frac{s_1-s_2}{s_2}\right|^{\delta'}, \quad s_1 \le s_2, \tag{3.53}$$

$$|E(l,s,s_1')-E(l,s,s_2')| \le \frac{\kappa\|b_+\|^2(1+\ln l_+)}{l_+^{1-\delta/2}}\left(\frac{s_2'}{s_2'}\right)^{1/2+w-\delta}\frac{\ln(s+s_2')}{(s_1')^{2\lambda+w-1/2}} \times\left[\left(\frac{s}{s_1'}\right)^{1/4}+\left(\frac{s_2'}{s}\right)^{1/2+w+\delta/4}\right]\left|\frac{s_1'-s_2'}{s_2'}\right|^{\delta}, \quad s_1' \le s_2'. \tag{3.54}$$

In (3.53), the Hölder exponent δ' is required to satisfy the constraint

$$\delta < \delta' < \min(\delta+\tfrac{1}{6}, 8\delta). \tag{3.55}$$

In order to obtain Hölder continuity of the integral (3.48), with respect to s and with exponent δ, we require a Hölder continuity of $E(l,s,s')$ with respect to s with an exponent δ' bigger than δ; cf. Ref. 14.

We may now apply Lemma III of Appendix D, in conjunction with (3.48) and (3.52)–(3.54), to infer that K_L obeys the bounds (3.43) and (3.44). [Even though the integral (3.48) does not require a principal-value definition, the lemma may still be used.]

This completes the proof of inequality (3.26). The proof of the Lipschitz inequality (3.28) is carried out by almost identical considerations. We have thus finished the proof of existence of a locally unique solution of the dynamical equation (2.30).

4. MANDELSTAM ANALYTICITY AND UNITARITY

We have established the existence of a solution to the dynamical equation (2.30), under the conditions that the potential (2.1) be sufficiently weak and satisfy inequalities (3.19) and (3.20). The solution is a member of the space β, defined in the first paragraph of Sec. 3, and is unique in a subset $\mathcal{S}$ of β, defined by (3.27).

We shall now show that the solution $b_+(l,s)$ of (2.30) leads to a solution of our original problem, namely, the problem of finding an amplitude $A(s,t)$ with an un-

subtracted Mandelstam representation and unitarity. We shall show, in fact, that the required amplitude is

$$A(s,t)=\frac{1}{\pi}\int_4^\infty \frac{dt'\rho(t')}{t'-t}+\frac{1}{\pi^2}\int_4^\infty\int_4^\infty \frac{ds'\,dt'\rho(s',t')}{(t'-t)(s'-s)}, \quad (4.1)$$

where $\rho(s,t)$ is computed as follows in terms of the solution $b_+(l,s)=\Delta^{-l}a_+(l,s)$ of (2.30):

$$\rho(s,t)=(1/2i)\int_{\mathrm{Re}l'=-\epsilon} dl'(2l'+1)P_{l'}(1+2t/\Delta)\,q(s) \times a_+(l',s)\,a_-(l',s), \quad (4.2)$$

$$a_-(l,s)=[a_+(l^*,s)]^*. \quad (4.3)$$

We shall first demonstrate that

(a) $\rho(s,t)$ has the proper support, as defined in Eq. (2.14),

(b) $\rho(s,t)$ is Hölder-continuous in both s and t,

(c) the integral (4.1) converges for all s and t interior to the cut s and t planes, and exists as a limit when s or t approaches its cut.

To prove (a), we simply close the integration contour in (4.2) by an infinite semicircle in the right half-plane. Since $b_+\in\beta$, the contribution of the semicircle vanishes if $t\leq\tau(s)$; see Eq. (2.14). The integrand is analytic inside the closed contour, so that $\rho(s,t)\equiv 0$ for $t\leq\tau(s)$.

We prove (b) and (c) by applying the bound (3.8) on $|P_l|$ and a similar bound on $|P_l'|$ to obtain the following inequalities:

$$|\rho(s,t)|\leq\kappa\|b_+\|^2\left(\frac{\Delta}{s}\right)^{1/2-\epsilon}\frac{(s+t)^{1/4-\epsilon}}{s^{2\lambda-1/2-\epsilon}t^{1/4}}; \quad (4.4)$$

$$|\rho(s_1,t)-\rho(s_2,t)| \leq\kappa\|b_+\|^2\left(\frac{\Delta}{s_2}\right)^{1/2-\epsilon-\delta}\frac{(s_1+t)^{1/4-\epsilon}}{s_1^{2\lambda-1/2-\epsilon}t^{1/4}}\left|\frac{s_1-s_2}{s_1}\right|^\delta, \quad s_1\leq s_2; \quad (4.5)$$

$$|\rho(s,t_1)-\rho(s,t_2)| \leq\kappa\|b_+\|^2\left(\frac{\Delta}{s}\right)^{1/2-\epsilon}\frac{(s+t_1)^{1/4-\epsilon-\delta/2}}{s^{2\lambda-1/2-\epsilon}t_1^{1/4-\delta/2}}\left|\frac{t_1-t_2}{t_1}\right|^\delta, \quad t_1\leq t_2. \quad (4.6)$$

The conclusions (b) and (c) follow immediately. The integrations over s' and t' may be carried out in either order.

Our next step is to take the Legendre projection of (4.1), and then reverse integration order to put the result in Froissart–Gribov form. The Legendre projection, call it $\hat{a}(l,s)$, is given by

$$\hat{a}(l,s)=\frac{1}{2}\int_{-1}^1 dz\,P_l(z)A(s,t) \quad (4.7)$$

$$=\frac{2}{\pi\Delta}\int_4^\infty dt\,Q_l\left(1+\frac{2t}{\Delta}\right)D(s,t), \quad (4.8)$$

$$D(s,t)=\rho(t)+\frac{1}{\pi}\int_{\sigma(t)}^\infty\frac{ds'\rho(s',t)}{s'-s}. \quad (4.9)$$

The reversal of integration order is easily justified through use of the above bounds on ρ. Similarly, one may justify a further change in integration order, to show that the right side of (4.8) is equal to the right side of the dynamical equation (2.34), for $s\to s_+$ and $\mathrm{Re}l\geq-\xi$. (Recall that $\epsilon>\xi>0$, so that the difficult situation $\mathrm{Re}l=\mathrm{Re}l'$ does not arise). Consequently, $\hat{a}(l,s_+)=a_+(l,s)$; i.e., our solution of (2.30) is identical with the Legendre projection of $A(s_+,t)$ for $s\geq 4$ and $\mathrm{Re}l\geq-\xi$.

Note that $\hat{a}(l,s)$ is defined in the cut s-plane and that, for $\mathrm{Re}l\geq-\xi$,

$$a_-(l,s)=[a_+(l^*,s)]^*=[\hat{a}(l^*,s_+)]^*=\hat{a}(l,s_-). \quad (4.10)$$

Henceforth, we may replace $a_\pm(l,s)$ by $a(l,s_\pm)$ in Eq. (4.2). We are now in a position to prove unitarity:

$$(1/2i)[\hat{a}(l,s_+)-\hat{a}(l,s_-)]=q(s)\,\hat{a}(l,s_+)\,\hat{a}(l,s_-). \quad (4.11)$$

This statement will hold even for complex l, with $\mathrm{Re}l\geq-\xi$. By applying (4.8), we see that the left side of (4.11) is given by

$$\frac{2}{\pi\Delta}\int_0^\infty dt\,Q_l(1+2t/\Delta)\,\rho(s,t), \quad (4.12)$$

where we have extended the integration region to $t=0$. The latter step is permissible since the integral (4.2) is identically zero for $0\leq t\leq 4$. We may now reverse the order of the t integration in (4.12) and l' integration in the definition (4.2) of ρ to obtain

$$\frac{1}{i\pi\Delta}\int_{\mathrm{Re}l'=-\epsilon} dl'(2l'+1)\,q(s)\,\hat{a}(l',s_+)\,\hat{a}(l',s_-) \times\int_0^\infty dt\,Q_l\left(1+\frac{2t}{\Delta}\right)P_{l'}\left(1+\frac{2t}{\Delta}\right)$$

$$=\frac{1}{2\pi i}\int_{\mathrm{Re}l'=-\epsilon} dl'(2l'+1)\,q(s) \times\hat{a}(l',s_+)\,\hat{a}(l',s_-)\,\frac{1}{l-l'}\,\frac{1}{l+l'+1}$$

$$=q(s)\,\hat{a}(l,s_+)\,\hat{a}(l,s_-). \quad (4.13)$$

This concludes the proof that the amplitude A of (4.1), which was constructed from the solution of Eq. (2.30), is unitary and has an unsubtracted Mandelstam representation.

To show that the amplitude we have constructed coincides with that which would be obtained by solving the Schrödinger equation, we must establish that the Schrödinger reduced partial-wave amplitude, $b^s(l,s_+)$, lies in the subset $\mathcal{S}$ of the Banach space β. To this end, we may appeal to the work of Bessis,[9] who has obtained bounds on b^s for a class of potentials $V(r)$, analytic in r for $\mathrm{Re}r>0$, and restricted for $\mathrm{Re}r>0$ as follows:

$$|V(r)|\leq m|r|^{-\rho},\quad \rho<2,\quad |r|\leq 1,$$
$$|V(r)|\leq m\exp(-\mu_0\,\mathrm{Re}r)\,|r|^{-\gamma},\quad \gamma>7/4,\ \mu_0>0,\ |r|>1. \quad (4.14)$$

One may show that our potential, as specified in Eqs. (2.1), (3.19), and (3.20), satisfies the conditions of Bessis, with $\rho=\frac{3}{2}$, $\gamma=2$; also, $\mu_0=2$ with our choice of units. One may apply Eqs. (2.3)–(2.5) and (2.13)–(2.14) of Ref. 9 to show that $b^s(l,s_+)$ is bounded as follows in our Region 1 ($-\epsilon\leq\mathrm{Re}l\leq-\xi$):

$$|b^s(l,s_+)|\leq\kappa m\ln l_+/\Delta^{1/2+w}(l_++s^{1/2}). \quad (4.15)$$

This result meets the requirements for membership in the set $\mathcal{S}$ (for sufficiently small m), as far as behavior

at large l_+ and large s is concerned. The right side of (4.15) becomes infinite as $s \to 4$, whereas the members of $\mathcal{S}$ are bounded and decrease at large l_+ even when s is close to 4. Bessis has also proved [see his Eq. (2.21)] that

$$|b^s(l, s_+)| \leq \kappa m, \quad \epsilon \leq 1. \tag{4.16}$$

This shows that b^s is finite at $s=4$, but it does not give the required simultaneous vanishing of b^s at large l_+. It seems likely, however, that this difficulty could be overcome by a more careful analysis at small ϵ.

There is a further difficulty in deriving the required properties of b^s from the analysis of Bessis, in that Ref. 9 contains no discussion of Hölder-continuity with respect to s. This means not only that boundedness of the second term in our norm $\|f\|_1$ of (3.2) is not verified, but also that $\rho(s,t)$ is not proved to be Hölder-continuous in s. The latter shortcoming means that Bessis has actually not shown that $A(s,t)$, as given by the Mandelstam representation, approaches a limit as s tends to its cut. Again, we think it likely that Hölder-continuity could be proved with some additional effort.

Bessis has not discussed the behavior of b^s at large Rel (i.e., l in our Region 2); we cannot infer, therefore, that b^s decreases exponentially at large Rel, or that $\|b_s\|_2$ is finite. Undoubtedly, this omission could be corrected with further work. We note that without this exponential decrease, the proper support of the double spectral function cannot be established either.

We see that step (c) of our study of potential scattering, as mentioned in Sec. 1 is not quite complete. There seems to be no serious doubt, however, that the remaining steps could be filled in. The most difficult step in verifying that $b^s \in \mathcal{S}$ is to show that b^s has suitable bounds simultaneously at large s and large l_+ in Region 1. This much has been accomplished by Bessis.

ACKNOWLEDGMENTS

We wish to thank Professor David Atkinson for profitable discussions and for his hospitality at the Institute of Theoretical Physics, University of Groningen. Visits to Groningen were partly supported through a NATO Research Grant. We also thank Professor Stanley Mandelstam for useful comments on the Watson-Sommerfeld method during the early stages of our program. One of us (J. S. F.) wishes to thank FOM (Foundation for Fundamental Research on Matter) for the receipt of a grant and Kothman's University Endowment for the award of a fellowship.

APPENDIX A

Here it is shown that the function $R_l(x)$, defined by Eq. (3.14), is subject to the bounds (3.15)–(3.17) for Re$l > -1+\zeta$, $\zeta > 0$, $0 < x \leq \frac{1}{2}$. Before analyzing R_l, let us change the variable of integration in (3.14) to $v = \cosh\theta - 1$, and define $m = l+1$, to obtain

$$R_l(x) = \int_0^\infty dv\,[v(v+2)]^{-1/2}(1+vx)^{-m}. \tag{A1}$$

We shall prove the bounds for Im$l \geq 0$, since they would then follow for Im$l < 0$ from the relation

$$R_l(x) = [R_{l^*}(x)]^*. \tag{A2}$$

Let us distort the v-integration contour so that it lies along the negative imaginary axis. The result is

$$R_l(x) = \int_0^\infty dt\,[t(t+2i)]^{-1/2}(1-itx)^{-m}. \tag{A3}$$

The integrand in (A3) may then be replaced by its modulus, and we find

$$|R_l(x)| < \int_0^\infty dt\, t^{-1/2}(t^2+4)^{-1/4}(1+t^2x^2)^{-m_0/2} \times \exp[-y\arctan(tx)], \tag{A4}$$

where we have written $m = m_0 + iy$ in (A4), with $m_0 \geq \zeta > 0$ and $y \geq 0$. Let us split the integral on the right side of (A4) into parts for which $xt \geq 1$ and $xt \leq 1$, respectively. For the first part, the integral is less than

$$\int_{1/x}^\infty \frac{dt}{t}\,(tx)^{-m_0} \exp[-y\arctan(tx)] = \int_1^\infty \frac{dr}{r^{1+m_0}} \exp[-y\arctan(r)]. \tag{A5}$$

The arctangent in (A5) takes on its minimum value at $r=1$, so that (A5) is less than

$$\kappa \exp(-\pi y/4). \tag{A6}$$

For the second part of the integral we obtain the bound

$$\int_0^{1/x} dt\, t^{-1/2} \exp[-y\arctan(tx)]. \tag{A7}$$

It follows from elementary considerations that, for $0 \leq r \leq 1$,

$$\arctan(r) \geq \pi r/4, \tag{A8}$$

so that an upper bound for (A7) is given by

$$(xy)^{-1/2} \int_0^y dr\, r^{-1/2} \exp(-\tfrac{1}{4}\pi r) \leq \kappa (xy)^{-1/2}. \tag{A9}$$

We may also bound the second term of (A4) by

$$\int_0^{1/x} dt\, t^{-1/2}(t^2+4)^{-1/4} \leq \kappa \ln(1/x). \tag{A10}$$

To obtain an estimate on $|R_l|$ that is useful at large m_0, we work directly with (A1), which we replace by its modulus to obtain

$$|R_l(x)| \leq \int_0^\infty dr\,[r(r+2x)]^{-1/2} \exp[-m_0 \ln(1+r)]. \tag{A11}$$

We split the integral in (A11) into the region $r \leq 1$, for which $\log(1+r) \geq r\log 2$, and $r \geq 1$. We obtain for $m_0 > \zeta$ that

$$|R_l(x)| \leq x^{-1/2} \int_0^1 dr\, r^{-1/2} \exp(-rm_0 \ln 2) + \int_1^\infty \frac{dr}{r^{1+m_0}} \leq \kappa (xm_0)^{-1/2}. \tag{A12}$$

From (A6), (A9), and (A12) we conclude that

$$|R_l(x)| \leq \kappa (l_+ x)^{-1/2}, \tag{A13}$$

which is the desired bound (3.15).

To obtain the result (3.16) for $|R_l'|$, we differentiate the representation (A1) with respect to x, and then integrate by parts to obtain

$$R_l'(x) = -\frac{1}{x} \int_0^\infty dv\, v^{-1/2}(v+2)^{-3/2}(1+vx)^{-m}. \tag{A14}$$

One may majorize the integral in (A14) by the same procedures used in the case of (A1). As a result we establish that

$$|R_l'(x)| \leq \kappa l_+^{-1/2} x^{-3/2}, \tag{A15}$$

which is the inequality (3.16).

To get (3.17), we differentiate (A14) with respect to x and integrate by parts to obtain

$$R_l''(x) = \frac{3}{x^2}\int_0^\infty dv\, v^{-1/2}(v+2)^{-5/2}(1+vx)^{-m}. \tag{A16}$$

The bound (3.17) is readily established from (A16) by techniques similar to the above; viz., we obtain

$$|R_l''(x)| \leq \kappa l_+^{-1/2} x^{-5/2}. \tag{A17}$$

Our analysis leads also to the results

$$|R_l(x)| \leq \kappa \ln(1/x),$$

$$|xR_l'(x)| \leq \kappa,$$

and

$$|xR_l''(x)| \leq \kappa. \tag{A18}$$

These bounds are not needed, however, in the present work.

APPENDIX B

We shall show that the partial wave Born amplitude $b_B(l,s)$, which may be determined from the Yukawa density ρ by Eq. (2.20), is subject to bounds (3.21)–(3.23) when ρ satisfies conditions (3.19) and (3.20). Let us substitute (3.10) and (3.12) into (2.20) to obtain

$$b_B(l,s) = \frac{2}{\pi}\int_4^\infty dt\, u(s,t)^{-l-1} R_l(x(s,t))\rho(t). \tag{B1}$$

We have used (3.7) and (3.13) to define the functions $u(s,t)$ and $x(s,t)$. To demonstrate (3.21), we replace the integrand in (B1) by its modulus and use (3.15) and (3.19) to obtain

$$\begin{aligned}|b_B(l,s)| &\leq \frac{\kappa}{l_+^{1/2}}\int_4^\infty dt\,|\rho(t)|\,\frac{1}{x^{1/2}u^{w+1}}\\ &\leq \frac{\kappa}{l_+^{1/2}}\int_4^\infty \frac{dt}{t^{\alpha+1/4}}\,\frac{1}{u^{w+3/4}},\\ &\leq \frac{\kappa}{l_+^{1/2}s^{3/4}p(s)^w}.\end{aligned} \tag{B2}$$

Here l_+ and w are defined as in Sec. 3. We have used the inequalities $\alpha > \frac{3}{4}$, $u \leq 4(s+t)$, $s \leq p(s) \leq u(s,t)$, $w + \frac{1}{2} > 0$, in obtaining (B2).

To demonstrate (3.22), let us integrate the expression (B1) by parts to obtain

$$\begin{aligned}b_B(l,s) = \frac{1}{\pi}\frac{1}{l+1}\int_4^\infty dt\, u(s,t)^{-l-1}\\ \times \partial_t[\rho(t)\,u(s,t)\,x(s,t)\,R_l(x(s,t))].\end{aligned} \tag{B3}$$

We have used the fact that $\rho(4)=0$ and the relation

$$x(s,t)\frac{\partial u}{\partial t}(s,t) = 2 \tag{B4}$$

in getting (B3). It is straightforward to show by using (3.10)–(3.16) and (3.19)–(3.20) that

$$|\partial_t(\rho u x R_l)| \leq \frac{\kappa}{l_+^{1/2}}\left(\frac{t}{t-4}\right)^{1/2}\frac{(s+t)^{3/4}}{t^{3/4+\alpha}}. \tag{B5}$$

Let us apply the bound (B5) in (B3) to obtain

$$\begin{aligned}|b_B(l,s)| &\leq \frac{\kappa}{l_+^{3/2}}\int_4^\infty dt\left(\frac{t}{t-4}\right)^{1/2}\frac{(s+t)^{3/4}}{u^{1+w}t^{3/4+\alpha}}\\ &\leq \frac{\kappa}{l_+^{3/2}p(s)^{1/2+w}}\int_4^\infty dt\left(\frac{t}{t-4}\right)^{1/2}\frac{(s+t)^{1/4}}{t^{3/4+\alpha}}\\ &\leq \frac{\kappa}{l_+^{3/2}s^{1/4}p(s)^w}.\end{aligned} \tag{B6}$$

In the last step of (B6), one shows that the integral is of order $s^{1/4}$, by splitting the integral into the parts with $t \leq s$ and $t > s$.

To obtain the bound (3.23), let us change the variable of integration in Eq. (B1) to u and use relation (B4) to obtain

$$b_B(l,s) = \frac{1}{\pi}\int_{p(s)}^\infty du\, u^{-l-1}\rho(t)\,x(s,t)\,R_l(x(s,t)). \tag{B7}$$

The function $t(s,u)$ is defined implicitly through (3.7). We differentiate (B7) with respect to s to obtain

$$\begin{aligned}&\partial_s b_B(l,s)\\ &= \frac{1}{\pi}\int_{p(s)}^\infty du\, u^{-l-1}\,\partial_s[\rho(t)\,x(s,t)\,R_l(x(s,t))].\end{aligned} \tag{B8}$$

The contribution from differentiating the lower limit of (B7) vanishes because $\rho(4)=0$. We may show by routine estimates that

$$|\partial_s(\rho x R_l)| \leq \frac{\kappa}{l_+^{1/2}}\left(\frac{t}{t-4}\right)^{1/2}\frac{1}{u^{3/4}t^{1/4+\alpha}}. \tag{B9}$$

We insert (B9) into the integral in (B8), and obtain the following bound upon $\partial_s b_B$:

$$\begin{aligned}|\partial_s b_B| &\leq \frac{\kappa}{l_+^{1/2}}\int_4^\infty dt\left(\frac{t}{t-4}\right)^{1/2}\frac{1}{u^{5/4+w}t^{3/4+\alpha}}\\ &\leq \frac{\kappa}{l_+^{1/2}s^{5/4}p(s)^w}.\end{aligned} \tag{B10}$$

Finally, we note in passing that if $\alpha > 1$ in (3.19), one can obtain an upper limit on $|b_B|$ which does not decrease when $\mathrm{Im}\,l \to \infty$; namely,

$$|b_B(l,s)| < \frac{\kappa}{s}\ln\frac{1}{p(s)^w}. \tag{B11}$$

It is an intrinsic feature of the Born term that bounds reflecting more rapid decrease at large $|\mathrm{Im}\,l|$ are obtained at the expense of correspondingly less rapid decrease at large s. The results (3.21), (3.22), and (B11) illustrate this property of b_B. The large s and l_+ dependence of these estimates cannot simultaneously be improved, if only conditions (3.19) and (3.20) are assumed for ρ.

APPENDIX C

The purpose of this appendix is to establish the bounds (3.49)–(3.51) for the function $G(l,l',s,s')$, which is defined by Eq. (2.33). We carry through the analysis for $\mathrm{Re}\,l' = \mathrm{Re}\,l = w$, where $-\epsilon \leq w \leq -\zeta$, $\epsilon < \frac{1}{2}$, $\zeta > 0$. The special case $\mathrm{Im}\,l \geq 0$ shall be analyzed here; the bounds for $\mathrm{Im}\,l < 0$ follow from the symmetry relation:

$$G(l,l',s,s') = [G(l^*,l'^*,s,s')]^*. \tag{C1}$$

The lower limit $\tau(s')$ of the integral is defined by Eq.

(2.14); we shall frequently use the fact that $\tau(s')$ is greater than 16.

Let us first use the identity

$$P_{l'}(z)=\frac{\tan\pi l'}{\pi}\,[Q_{l'}(z)-Q_{-l'-1}(z)], \tag{C2}$$

along with (3.7) and (3.10)–(3.13), to write

$$G=\frac{\tan\pi l'}{i\pi^2}\left(s'^{1+2l'}\,G_1-G_2\right), \tag{C3}$$

where

$$G_2=\int_\tau^\infty dt\,u(s',t)^{l'}R_{-l'-1}(x(s',t))\times[u(s,t)^{-l-1}R_l(x(s,t))-u(s',t)^{-l-1}R_l(x(s',t))]. \tag{C4}$$

The expression for G_1 is obtained by interchanging l' and $-l'-1$ in (C4). We shall carry through estimates upon G_2, and then indicate how the results are altered for G_1.

One may show that the separate terms in the integrand (C4) behave at large t as $1/t$, so that the integrals of the separate terms do not converge absolutely. There is a cancellation in the integrand of (C4) at large t, however, so that the integral (C4) is absolutely convergent. By contrast, each of the separate terms in G_1 gives an absolutely convergent integral. To exploit this cancellation in the integrand, let us decompose the t integral in (C4):

$$G_2=G_{21}+G_{22}\equiv\int_\tau^T+\int_T^\infty, \tag{C5}$$

where

$$T=(s+s')\,l_+'l_+^2. \tag{C6}$$

[We shall write out the analysis as though T is greater than $\tau(s')$; if $\tau>T$, we have only the second term of (C5).] The mean-value theorem, together with the bounds (3.15) and (3.16), may be applied to the last factor of (C4) to obtain

$$|u(s,t)^{-l-1}R_l(x(s,t))-u(s',t)^{-l-1}R_l(x(s',t))|\le\kappa|s'-s|\,l_+^{1/2}t^{-1/4}(s_<+t)^{-w-l/4}, \tag{C7}$$

where $s_<=\min(s,s')$. Let us apply the bound (C7) to G_{22} to obtain

$$|G_{22}|\le\kappa|s'-s|\left(\frac{l_+}{l_+'}\right)^{1/2}\int_T^\infty dt\,(s_<+t)^{-3/2}t^{-1/2}\le\kappa\,(l_+l_+')^{-3/2}\,\frac{|s'-s|}{s'+s}. \tag{C8}$$

The estimate (C8) verifies that G_{22} becomes small at large l_+ and, independently, at large l_+'; we obtain such a bound by having chosen T to be large at large s, s', l, or l'.

By contrast, we choose *not* to exploit the cancellation at large t in G_{21}; rather we write

$$|G_{21}|\le|I(s',s)|+|I(s',s')|, \tag{C9}$$

where

$$I(s,s')=\int_\tau^T dt\,\frac{u(s',t)^{l'}}{u(s,t)^{l+1}}R_{-l'-1}(x(s',t))\,R_l(x(s,t)). \tag{C10}$$

We shall obtain bounds upon $|I(s,s')|$; the bounds for $|I(s',s')|$ follow as a special case.

Let us replace each factor in (C10) by its modulus, and use the bound (3.15) upon the R's to obtain

$$|I(s,s')|\le\kappa(l_+l_+')^{-1/2}\int_\tau^T\frac{dt}{[t(s+t)]^{1/2}}\left(\frac{s'+t}{s+t}\right)^{1/4+w}. \tag{C11}$$

The integral on the right side of (C11) is conveniently analyzed by separating it into an integral from τ to $s'+s$ and an integral from $s'+s$ to T. We may thereby establish that the integral in (C11) is bounded by $\kappa\mathcal{E}(l,l',s,s',\frac14+w)$, where we have adopted the notation

$$\mathcal{E}(l,l',s,s',\beta)=\left[1+\left(\frac{s'}{s}\right)^\beta\right]\ln(s'+s)(1+\ln l_+l_+'). \tag{C12}$$

In turn, $|I(s,s')|$ satisfies the bound

$$|I(s,s')|\le\kappa(l_+l_+')^{-1/2}\mathcal{E}(l,l',s,s',\tfrac14+w). \tag{C13}$$

[Actually, the factor $\ln(s+s')$ may be replaced by 1 in (C13)]. Alternately, we may integrate (C10) by parts, taking $u^{-l-1}\,du/dt$ as the term to be integrated, and then bound the resulting expression to obtain

$$|I(s,s')|\le\kappa(l_+')^{1/2}l_+^{-3/2}\times\left[1+\left(\frac{s'}{s}\right)^{1/4+w}+\int_\tau^T\frac{dt}{t}\left(\frac{s'+t}{s+t}\right)^{1/4+w}\right]. \tag{C14}$$

We may treat the integral in (C14) as we did the one in (C11) to obtain the result

$$|I(s,s')|\le\kappa(l_+')^{1/2}l_+^{-3/2}\mathcal{E}(l,l',s,s',\tfrac14+w). \tag{C15}$$

The bound (C15) is more stringent than (C11) at large l_+, but it does diverge at large l_+'. We may integrate (C10) by parts, taking $u^{l'}\,du/dt$ as the integrated term, to obtain

$$|I(s,s')|\le\kappa\,l_+^{1/2}(l_+')^{-3/2}\times\left[1+\left(\frac{s'}{s}\right)^{3/4+w}+\int_\tau^T\frac{dt}{t}\left(\frac{s'+t}{s+t}\right)^{3/4+w}\right]. \tag{C16}$$

A bound of the form

$$|I(s,s')|\le\kappa l_+^{1/2}(l_+')^{-3/2}\,\mathcal{E}(l,l',s,s',\tfrac34+w) \tag{C17}$$

is readily established from (C16). We emphasize that the bounds (C13), (C15), and (C17) are all valid when $-\epsilon\le w\le-\zeta$.

To make an improvement in these bounds when both l_+ and l_+' are large, we must take into account the oscillations in the integrand (C10) when both $|\mathrm{Im}\,l|$ and $|\mathrm{Im}\,l'|$ are large. We handle the oscillations in $[u(s',t)]^{l'}$ and $[u(s,t)]^{-l-1}$ by writing (C10) as

$$I=\int_\tau^T dt\,f(t)[g(t)]^{2iy}, \tag{C18}$$

where

$$f(t)=\left(\frac{u(s',t)}{u(s,t)}\right)^w\frac{1}{u(s,t)}\,R_l(x(s,t))\,R_{-l'-1}(x(s',t)) \tag{C19}$$

and

$$g(t)=\frac{(\sqrt{t}+\sqrt{s'+t})^{\lambda}}{(\sqrt{t}+\sqrt{s+t})}, \tag{C20}$$

with

$$y=\mathrm{Im}l, \quad y'=\mathrm{Im}l', \quad \lambda=y'/y. \tag{C21}$$

We shall use the bounds (C13), (C15), and (C17) if either $|y|$ or $|y'|$ is not large; the analysis of (C18) applies to the case in which both are large.

Our procedure is to integrate (C18) by parts in order to obtain more stringent bounds at large l_+, which are not divergent at large l_+'. For the straightforward, though lengthy, analysis we shall need bounds on f, which are easily obtained through (3.15) and (3.16); namely,

$$|f(t)|, |tf'(t)|$$
$$\leq \kappa[t(s+t)l_+'l_+]^{-1/2}\left(\frac{s'+t}{s+t}\right)^{1/4+w}. \tag{C22}$$

The derivative of $g(t)$, which plays a crucial role in the analysis of (C18), may be written as

$$g'(t)=g(t)m(t), \tag{C23}$$

where

$$m(t)=\frac{1}{2t^{1/2}}\left(\frac{\lambda}{(s'+t)^{1/2}}-\frac{1}{(s+t)^{1/2}}\right). \tag{C24}$$

In the analysis we shall distinguish three cases, depending upon whether, for $4\leq t\leq\infty$, $m(t)$ is everywhere positive, everywhere negative, or neither:

Case I (positive): $\lambda\geq\lambda_M=\sup[1,(s'/s)^{1/2}]$; (C25)

Case II (negative): $\lambda\leq\lambda_m=\inf[1,(s'/s)^{1/2}]$; (C26)

and

Case III (mixed): $\lambda_m<\lambda<\lambda_M$. (C27)

Note that if $s=s'$, Case III does not arise. We shall handle Cases I and II first; they are similar in that for either case one may integrate by parts, with $g^{2iy}g'$ being integrated, to obtain

$$(1+2iy)I(s,s')=g^{2iy+1}\frac{f}{g'}\Big|_{\tau}^{T}-\int_{\tau}^{T}dt\,g^{2iy+1}\partial_t\left(\frac{f}{g'}\right). \tag{C28}$$

We may write the derivative in (C28) as

$$\partial_t(f/g')=(f'-fm'/m)/mg-f/g. \tag{C29}$$

Then (C29) may be used in (C28) to give

$$2iyI(s,s')$$
$$=g^{2iy}\frac{f}{m}\Big|_{\tau}^{T}-\int_{\tau}^{T}dt\,g^{2iy}\left[\frac{f'}{m}+f\partial_t\left(\frac{1}{m}\right)\right]. \tag{C30}$$

To handle (C30), it is convenient to note that

$$\frac{1}{m(t)}=2[t(s'+t)]^{1/2}\left[\lambda-\left(\frac{s'+t}{s+t}\right)^{1/2}\right]^{-1}, \tag{C31}$$

$$\left|\partial_t\frac{1}{m(t)}\right|\leq 2[t(t+s')]^{1/2}\left|\partial_t\left[\lambda-\left(\frac{s'+t}{s+t}\right)^{1/2}\right]^{-1}\right|$$
$$+\frac{2}{t|m(t)|}. \tag{C32}$$

Thus, (C30) yields

$$|yI|\leq\left|\frac{f(T)}{m(T)}\right|+\left|\frac{f(\tau)}{m(\tau)}\right|+\int_{\tau}^{T}\frac{dt}{|m(t)|}\left(|f'(t)|+\frac{1}{t}|f(t)|\right)$$
$$+\int_{\tau}^{T}dt\,[t(t+s')]^{1/2}|f(t)|\left|\partial_t\left[\lambda-\left(\frac{s'+t}{s+t}\right)^{1/2}\right]^{-1}\right|. \tag{C33}$$

Now notice that the factor

$$\chi=\left[\lambda-\left(\frac{s'+t}{s+t}\right)^{1/2}\right]^{-1} \tag{C34}$$

is monotonic in t (increasing or decreasing, depending on the sign of $s-s'$) on the interval $(4,\infty)$; i.e.,

$$|\partial_t\chi|=\pm\partial_t\chi, \tag{C35}$$

with a definite choice of the sign for fixed $s-s'$. This observation allows us conveniently to integrate by parts in the last term of (C33), *after* replacing $|f|$ by its upper bound (C22). After integrating partially, and using (C22) and (C31) to treat the first three terms of (C33), we find that

$$|yI|\leq\frac{\kappa}{(l_+l_+')^{1/2}}\left\{\left[1+\left(\frac{s'}{s}\right)^{3/4+w}\right][|\chi(\tau)|+|\chi(T)|]\right.$$
$$\left.+\int_{\tau}^{T}\frac{dt}{t}\left(\frac{t+s'}{t+s}\right)^{3/4+w}|\chi(t)|\right\}. \tag{C36}$$

To complete the estimate of I, we note that $|\chi|$, being monotonic in t, is majorized either by its value at $t=4$ or by its value at $t=\infty$. Consequently, for any intermediate value of t it is majorized by the sum of those values:

$$|\chi(t)|\leq|\lambda-1|^{-1}+|\lambda-(s'/s)^{1/2}|^{-1}. \tag{C37}$$

When this result is applied in (C36), we obtain

$$|yI|\leq\kappa(l_+l_+')^{-1/2}\left[|\lambda-1|^{-1}+\left|\lambda-\left(\frac{s'}{s}\right)^{1/2}\right|^{-1}\right]\left[1+\left(\frac{s'}{s}\right)^{3/4+w}\right.$$
$$\left.+\int_{\tau}^{T}\frac{dt}{t}\left(\frac{s'+t}{s+t}\right)^{3/4+w}\right]. \tag{C38}$$

By comparison with (C16) and (C17), this yields the required bound for cases I and II:

$$|I(s,s')|\leq\kappa(l_+l_+')^{-1/2}\left[|y'-y|^{-1}+\left|y'-\left(\frac{s'}{s}\right)^{1/2}y\right|^{-1}\right]$$
$$\times\mathcal{E}(l,l',s,s',\tfrac{3}{4}+w). \tag{C39}$$

This brings us to Case III, $\lambda_m<\lambda<\lambda_M$, in which we must face the complication that $g'(t)$ has a simple zero at $t=t_0\in(4,\infty)$, with

$$\lambda=[(s'+t_0)/(s+t_0)]^{1/2}. \tag{C40}$$

We shall give the detailed analysis of $I(s,s')$ for the case in which $\tau<t_0<T$. The remaining cases, $t_0\geq T$, $t_0\leq\tau$, may be treated in a similar way. Let us divide the integral (C18) into three parts, as follows:

$$I=\int_{\tau}^{T}=\int_{\tau}^{t_1}+\int_{t_1}^{t_2}+\int_{t_2}^{T}=I_1+I_2+I_3. \tag{C41}$$

Here

$$t_1=\max[\tau,t_0(1-\eta)], \quad t_2=\min[t_0(1+\eta),T], \tag{C42}$$

where the parameter η, to be specified presently, is

restricted so that $0 \leq \eta \leq \frac{1}{2}$. We integrate I_1 and I_3 by parts, to show that they vanish appropriately at large y, while the smallness of I_2 at large y will be ensured through a suitable choice of η. Partial integration of I_2 is not possible, since the integral would then diverge at $t = t_0$.

For I_2, we simply use the estimate (C. 22) of f to obtain

$$|I_2| < \int_{t_1}^{t_2} dt\,|f(t)|$$

$$\leq \kappa\,(l_+ l_+')^{-1/2} \int_{t_0(1-\eta)}^{t_0(1+\eta)} \frac{dt}{[t(\delta+t)]^{1/2}} \left(\frac{\delta'+t}{\delta+t}\right)^{1/4+w}$$

$$\leq \kappa\eta\,(l_+ l_+')^{-1/2} \left[1 + \left(\frac{s'}{s}\right)^{1/4+w}\right]$$

$$\leq \kappa\eta\,(l_+ l_+')^{-1/2}\,\mathcal{E}(l, l', s, s', \tfrac{1}{4}+w). \tag{C43}$$

For the remaining integrals, we note that the right side of (C36) is an upper bound for I_1 or I_3, if the limits $[\tau, T]$ are replaced by $[\tau, t_1]$, $[t_2, T]$, respectively. The requisite integration by parts may be carried through on each of these subintervals as χ is monotonic and has a definite sign on each subinterval. We shall use this bound, after finding a new estimate for the factor χ that occurs under the integral.

To treat χ, we substitute (C40) in the definition (C34), and rearrange the quotient to obtain

$$\chi = \frac{(\delta+t_0)(\delta+t)}{(s'-s)(t-t_0)} \left[\left(\frac{\delta'+t_0}{\delta+t_0}\right)^{1/2} + \left(\frac{\delta'+t}{\delta+t}\right)^{1/2}\right]. \tag{C44}$$

Next, we make the essential observation that both $\lambda - 1$ and $\lambda - (s'/s)^{1/2}$, the quantities that occur in our previous estimate (C37) of χ, are proportional to $s - s'$. In fact, a little calculation shows that

$$s' - s = (\lambda - 1)(\delta + t_0)\left[1 + \left(\frac{\delta'+t_0}{\delta+t_0}\right)^{1/2}\right], \tag{C45}$$

and

$$s - s' = \left[\lambda - \left(\frac{s'}{s}\right)^{1/2}\right] \frac{s(\delta+t_0)}{t_0 - 4} \left[\left(\frac{s'}{s}\right)^{1/2} + \left(\frac{\delta'+t_0}{\delta+t_0}\right)^{1/2}\right]. \tag{C46}$$

By substituting one or the other of these expressions for $s - s'$ in (C44), the choice depending on the disposition of the variables t_0 and s, we obtain the required bound on χ. One uses (C45) when $s \leq t_0$ and (C46) when $s > t_0$. The analysis is straightforward, being based on monotonicity in t of $[(\delta'+t)/(\delta+t)]^{1/2}$, but it involves a separate consideration of each of the possible inequalities between s, s', t, and t_0. The result is that

$$|\chi| \leq \frac{\kappa \max(t, t_0)}{|t - t_0|} \left(\frac{1}{|\lambda - 1|} + \frac{1}{|\lambda - (s'/s)^{1/2}|}\right). \tag{C47}$$

For the term I_1 of (C41) we have $t \leq t_0(1-\eta)$ and

$$\frac{\max(t, t_0)}{|t - t_0|} \leq \frac{1}{\eta}, \tag{C48}$$

whereas for I_3 we have $t \geq t_0(1+\eta)$ and

$$\frac{\max(t, t_0)}{|t - t_0|} \leq \frac{1+\eta}{\eta}. \tag{C49}$$

We can now majorize I_1 or I_3 by the right side of (C36), after replacing the limits $[\tau, T]$ by $[\tau, t_1]$ or $[t_2, T]$, respectively. We introduce (C47)–(C49) to obtain

$$|yI_1| + |yI_3| \leq \frac{\kappa}{(l_+ l_+')^{1/2}} \frac{1}{\eta} \left(\frac{1}{|\lambda - 1|} + \frac{1}{|\lambda - (s'/s)^{1/2}|}\right)$$

$$\times \left[1 + \left(\frac{s'}{s}\right)^{3/4+w} + \int_\tau^T \frac{dt}{t} \left(\frac{t+\delta'}{t+\delta}\right)^{3/4+w}\right]. \tag{C50}$$

By combining this result with (C43), we arrive at the following bound of the integral I:

$$|I(s, s')| \leq \frac{\kappa}{(l_+ l_+')^{1/2}}\,\mathcal{E}(l, l', s, s', \tfrac{1}{4}+w)$$

$$\times \left\{\eta + \frac{1}{\eta}\left[1 + \left(\frac{s'}{s}\right)^{1/2}\right]\left[|y' - y|^{-1} + \left|y' - \left(\frac{s'}{s}\right)^{1/2} y\right|^{-1}\right]\right\}. \tag{C51}$$

It remains to choose η. We define

$$\delta = [1 + (s'/s)^{1/4}][\,|y' - y|^{-1/2} + |y' - (s'/s)^{1/2} y|^{-1/2}]. \tag{C52}$$

In the case $\delta < 1$ we put $\eta = \delta$, which yields

$$|I(s, s')| \leq \frac{\kappa\delta\,\mathcal{E}(l, l', s, s', \frac{1}{4}+w)}{(l_+ l_+')^{1/2}}$$

$$\leq \frac{\kappa}{(l_+ l_+')^{1/2}}\left[|y' - y|^{-1/2} + \left|y' - \left(\frac{s'}{s}\right)^{1/2} y\right|^{-1/2}\right]$$

$$\times \mathcal{E}(l, l', s, s', \tfrac{1}{2}+w). \tag{C53}$$

When $\delta \geq 1$, the inequality (C53) is a trivial consequence of (C13). Thus, we have shown that (C53) holds in Case III, if $\tau < t_0 < T$. It is not difficult to establish (C53) also for the cases $t_0 \leq \tau$, $t_0 \geq T$. For instance, if $t_0 \leq \tau$, we integrate by parts over the entire interval $[\tau, T]$ if $\tau - t_0 \geq \tau\delta$, but if $\tau - t_0 < \tau\delta$, we integrate by parts only on the interval $[t_0 + \tau\delta, T]$, while integrating the bound of $|f|$ straightaway on the interval $[\tau, t_0 + \tau\delta]$.

The result (C53) is valid for Cases I and II as well as III. Indeed, it is an immediate consequence of (C39) and (C13) in those cases. The bound is good in the event $s = s'$, a situation which occurs only in Cases I and II, so that we see from (C9) that G_{21} is also majorized by the expression in (C53). By using (C8), it is easy to show that G_{22}, defined in (C5), is subject to the same bound as G_{21}. By (C5), the final result for G_2 is

$$|G_2(s, s')| < \frac{\kappa}{(l_+ l_+')^{1/2}}\left[|y' - y|^{-1/2} + \left|y' - \left(\frac{s'}{s}\right)^{1/2} y\right|^{-1/2}\right]$$

$$\times \mathcal{E}(l, l', s, s', \tfrac{1}{2}+w). \tag{C54}$$

The function $\delta'^{1+2l'} G_1$ of Eq. (C3) has an upper bound the same as (C54), except that y' is replaced by $-y'$. The proof of this assertion requires only slight modifications of the arguments presented above. We again make a decomposition of the integral as in (C5):

$$G_1 = G_{11} + G_{12} = \int_\tau^T + \int_\tau^\infty. \tag{C55}$$

We easily find that $\delta'^{1+2l'} G_{12}$ obeys (C8). As in (C9), we write

$$|\sigma'^{l+2l'} G_{11}| \leq |J(s,s')| + |J(s',s')|, \tag{C56}$$

where

$$J(s,s') = \sigma'^{l+2l'} \int_\tau^T dt\, \frac{R_{l'}(x(\sigma',t))\, R_l(x(\sigma,t))}{u(\sigma',t)^{l'+1} u(\sigma,t)^{l+1}}. \tag{C57}$$

The problems in bounding $J(s,s')$ are easily reduced to those encountered in bounding $I(s,s')$. We find immediately that J obeys (C13), (C15), and (C17). To establish (C53), we write in place of (C18) the expression

$$J = \int_\tau^T dt\, \hat{f}(t) [\hat{g}(t)]^{-2iy}, \tag{C58}$$

$$\hat{f}(t) = \frac{\sigma'^{l+2l'}}{[u(\sigma',t)\, u(\sigma,t)]^{1+w}}\, R_l(x(\sigma,t))\, R_{l'}(x(\sigma',t)), \tag{C59}$$

$$\hat{g}(t) = (\sqrt{t} + \sqrt{\sigma'+t})^\lambda\, (\sqrt{t} + \sqrt{\sigma+t}). \tag{C60}$$

Now $\hat{f}(t)$ satisfies (C22), and, in analogy to (C23),

$$\hat{g}'(t) = \hat{g}(t)\, \hat{m}(t), \tag{C61}$$

$$\hat{m}(t) = \frac{1}{2t^{1/2}} \left(\frac{\lambda}{(\sigma'+t)^{1/2}} + \frac{1}{(\sigma+t)^{1/2}} \right). \tag{C62}$$

If $m(t)$ has a zero at $y' = y_0$, then $\hat{m}(t)$ has a zero at $y' = -y_0$. By reviewing the discussion following (C18), one sees that the argument required for (C58) is the same as that for (C18), except that $-y'$ plays the role that y' played before. We can then assert that the right-hand side of (C53) is larger than $J(s,s')$, if $-y'$ is substituted for y'. Since $|\tan \pi l'|$ is uniformly bounded for $-\epsilon \leq w \leq 0$, $\epsilon < \frac{1}{2}$, our final bound for G, obtained from (C3), is as follows:

$$\begin{aligned} &|G(l,l',s,s')| \\ &\quad \leq \kappa (l_+ l_+')^{-1/2}\, \mathcal{E}(l,l',s,s',\tfrac{1}{2}+w) \\ &\qquad \times [\,|y'-y|^{-1/2} + |y'+y|^{-1/2} + |y' - (s'/s)^{1/2} y|^{-1/2} \\ &\qquad + |y' + (s'/s)^{1/2} y|^{-1/2}], \end{aligned} \tag{C63}$$

where

$$\mathrm{Re} l' = \mathrm{Re} l = w, \quad -\epsilon \leq w \leq \zeta, \quad \epsilon < \tfrac{1}{2}, \quad \zeta > 0,$$

$$\mathrm{Im} l' = y', \quad \mathrm{Im} l = y,$$

the function $\mathcal{E}$ being defined in (C12). The bound (C63) is equivalent to (3.49), the bound to be established for $|G|$.

In the remainder of this appendix, we shall obtain the estimates (3.50) and (3.51) of the derivatives of G. We first notice that $\partial_s G$ and $\partial_{s'} G$ are related as follows:

$$\begin{aligned} \partial_{s'} G + \left(\frac{\sigma}{\sigma'}\right) \partial_s G = \frac{l'-l}{\sigma'}\, G + \frac{1}{\pi i}\, \partial_{s'} \left(\frac{\tau(s')}{\sigma'}\right) P_{l'} \left(1 + \frac{2\tau(s')}{\sigma'}\right) \\ \times \left[Q_l \left(1 + \frac{2\tau(s')}{\sigma'}\right) - \left(\frac{\sigma'}{\sigma}\right)^{l+1} Q_l \left(1 + \frac{2\tau(s')}{\sigma}\right) \right]. \end{aligned} \tag{C64}$$

We majorize $\partial_s G$, and then use (C64) to majorize $\partial_{s'} G$. By (C3) we have

$$\partial_s G = \frac{\tan \pi l'}{i\pi^2}\, (\sigma'^{l+2l'}\, \partial_s G_1 - \partial_s G_2), \tag{C65}$$

where

$$\begin{aligned} \partial_s G_2 = \int_{\tau(s')}^\infty dt\, u(\sigma',t)^{l'}\, R_{-l'-1}(x(\sigma',t)) \\ \times \partial_s [u(\sigma,t)^{-l-1} R_l(x(\sigma,t))]. \end{aligned} \tag{C66}$$

The formula for $\partial_s G_1$ is obtained by interchanging l' and $-l'-1$ in (C66).

The integral in (C66) is absolutely convergent, and we may employ (3.15) and (3.16) to show that

$$|\partial_s G_2| \leq \kappa (l_+/l_+')^{1/2} (1/s)\, \mathcal{E}(l,l',s,s',\tfrac{1}{4}+w). \tag{C67}$$

Also, we may integrate (C66) by parts and use (3.15)–(3.16) to show that

$$|\partial_s G_2| \leq \kappa (l_+'/l_+)^{1/2} (1/s)\, \mathcal{E}(l,l',s,s',\tfrac{1}{4}+w). \tag{C68}$$

Bounds the same as these are true for $|\sigma'^{l+2l'}\, \partial_s G_1|$, and are proved in the same way. Hence,

$$|\partial_s G| \leq \kappa\, \frac{(l_+ l_+')^{1/2}}{l_+ + l_+'}\, \frac{1}{s}\, \mathcal{E}(l,l',s,s',\tfrac{1}{4}+w). \tag{C69}$$

Finally, we use this result in (C64), to bound $\partial_{s'} G$. We use (C2), (3.15), and (3.16) to handle the Legendre functions, and the result is

$$|\partial_{s'} G| \leq \kappa\, \frac{(l_+ l_+')^{1/2}}{l_+ + l_+'}\, \frac{1}{\sigma'}\, \mathcal{E}(l,l',s,s',\tfrac{3}{4}+w). \tag{C70}$$

The inequalities (C69) and (C70) are equivalent to (3.50) and (3.51), respectively.

It is worth mentioning that the estimates (C69) and (C70) are not optimal, since we have not fully exploited the oscillations in the integrand of (C66). Nevertheless, our results are sufficient for the purposes of Sec. 3.

APPENDIX D

We shall state three lemmas concerning principal-value integrals, which are used at several stages of the analysis. These lemmas are proved by arguments in the spirit of Muskhelishvili's book.[15] For Lemma III, see also Pogorselski.[14]

Lemma I

Let $f(s)$ be a complex continuous function for $s \in [4,\infty)$ which satisfies the following bounds:

$$|f(s)| \leq c \left(\frac{\sigma}{s}\right)^\theta \frac{1}{s^\alpha} \tag{D1}$$

and, for $s' > s$,

$$|f(s') - f(s)| \leq \frac{c}{s^\alpha} \left(\frac{\sigma'}{s'}\right)^{\theta-\delta} \left|\frac{s'-s}{s}\right|^\delta. \tag{D2}$$

The positive constants c, α, δ, and θ are subject to the constraints $\delta < \theta$ and $\alpha + \delta < 1$.

Then the function $g(s)$, defined by

$$g(s) = \frac{P}{\pi} \int_4^\infty \frac{ds'}{s'-s}\, f(s'), \tag{D3}$$

satisfies the bounds

$$|g(s)| \leq \frac{\kappa c}{s^\alpha}, \tag{D4}$$

and, for $s' > s$,

$$|g(s') - g(s)| \leq \frac{\kappa c}{s^\alpha} \left|\frac{s'-s}{s}\right|^\delta. \tag{D5}$$

The constant κ in (D4) and (D5) is independent of the particular choice of the function f.

Lemma II

Let $f(s,t)$ be a complex continuous function for $(s,t)\in[4,\infty)\times[4,\infty)$, which satisfies the constraints

$$|f(s,t)|\leq c\left(\frac{4}{s}\right)^{\theta}\left[1+\left(\frac{s}{t}\right)^{\beta}\right]\frac{1}{s^{\alpha}}, \tag{D6}$$

and, for $s_1<s_2$,

$$|f(s_1,t)-f(s_2,t)| \leq c\left[1+\left(\frac{s_1}{t}\right)^{\beta}+\left(\frac{s_2}{t}\right)^{\beta}\right]\left(\frac{4}{s_2}\right)^{\theta-2\delta}\frac{|s_1-s_2|^{\delta}}{s_1^{\alpha+\delta}}. \tag{D7}$$

The constants α, β, θ, and δ are subject to the constraints $\alpha>0$, $\delta>0$, $\alpha+\delta<1$, $0<\alpha-\beta<1-\delta$, and $\theta-2\delta>0$.

We define the function $g(s)$ by the principal value integral

$$g(s)=\frac{P}{\pi}\int_4^{\infty} ds'\,\frac{f(s',s)}{s'-s}. \tag{D8}$$

Then for $s\in[4,\infty)$ it may be shown that $g(s)$ is continuous and subject to the bound (D4) above, with the constant c the same as in (D6)–(D7) and with the constant κ independent of f.

Lemma III

Let $f(s,t)$ be a complex continuous function for $(s,t)\in[4,\infty)\times[4,\infty)$ which meets the conditions

$$|f(s,t)|\leq c\,\frac{\ln t}{s^{\alpha}}\left[\left(\frac{s}{t}\right)^{\beta}+\left(\frac{t}{s}\right)^{\gamma}\right]\left(\frac{4}{s}\right)^{\theta}, \tag{D9}$$

$$|f(s,t_1)-f(s,t_2)| \leq c\,\frac{\ln t_2}{s^{\alpha}}\left[\left(\frac{s}{t_1}\right)^{\beta}+\left(\frac{t_2}{s}\right)^{\gamma}\right]\left(\frac{4}{s}\right)^{\theta}\left|\frac{t_1-t_2}{t_1}\right|^{\delta'}, \quad t_2>t_1, \tag{D10}$$

$$|f(s_1,t)-f(s_2,t)| \leq \frac{c\ln t}{s_1^{\alpha}}\left[\left(\frac{s_2}{t}\right)^{\beta+\theta/4}+\left(\frac{t}{s_1}\right)^{\gamma}\right] \times\left(\frac{4}{s_2}\right)^{\theta-\delta}\left|\frac{s_1-s_2}{s_1}\right|^{\delta}, \quad s_2\geq s_1, \tag{D11}$$

The positive constants α, β, γ, δ, δ', θ are subject to the constraints

$$\delta'>\delta, \quad \theta>\delta, \quad \alpha>\beta, \quad \alpha+\gamma<1, \quad \alpha+\delta<1. \tag{D12}$$

Then

$$|g(s,s)|\leq\kappa c\,\frac{\ln s}{s^{\alpha}}, \tag{D13}$$

$$|g(s_2,s_2)-g(s_1,s_1)|\leq\kappa c\,\frac{\ln s_1}{s_1^{\alpha}}\left|\frac{s_2-s_1}{s_1}\right|^{\delta}, \quad s_2\geq s_1, \tag{D14}$$

where

$$g(s,t)=\frac{P}{\pi}\int_4^{\infty}\frac{f(s',t)\,ds'}{s'-s}.$$

The constant κ is independent of f.

*Supported in part by the National Science Foundation.

[1] D. Atkinson, P.W. Johnson, and R.L. Warnock, Phys. Rev. D **6**, 2966 (1972). This paper contains the first formulation of partial-wave equations with exact crossing symmetry. The analytic continuation of absorptive parts out of the physical region was accomplished by a somewhat awkward conformal mapping method, which we have now replaced by the more suitable Sommerfeld–Watson continuation based on complex l.

[2] D. Atkinson and J.S. Frederiksen, Comm. Math. Phys. **40**, 55 (1975).

[3] J.S. Frederiksen, "Sommerfeld–Watson Representation for Double-Spectral Functions II: Crossing-Symmetric Pion–Pion Scattering Amplitude without Regge Poles," Canberra preprint, October, 1974.

[4] P.W. Johnson and R.L. Warnock (to be published).

[5] D. Atkinson, J.S. Frederiksen, and M. Kaekebeke (to be published).

[6] S. Mandelstam, Phys. Rev. **112**, 1344 (1958); **115**, 1741, 1752 (1959). D. Atkinson, Nucl. Phys. B **7**, 375 (1968); B **8**, 377 (1968); B **13**, 415 (1969); B **23**, 397 (1970).

[7] See, for instance, P.D.B. Collins and E.J. Squires, *Regge Poles in Particle Physics*, Springer Tracts in Modern Physics, Vol. 45 (Springer-Verlag, Berlin, 1968), Sec. II.4.

[8] A proof of the Mandelstam representation for potential scattering was first outlined by R. Blankenbecler, M.L. Goldberger, N.N. Khuri, and S.B. Treiman, Ann. Phys. (N.Y.) **10**, 62 (1960). These authors also suggested that S matrix equations could replace the Schrödinger equation as a means of treating potential scattering. They did not succeed in proving that the Schrödinger equation could be replaced, however, since there are serious omissions in their treatment of convergence questions connected with low partial waves when Regge poles are present. This matter will be discussed in detail in Ref. 4.

[9] D. Bessis, J. Math. Phys. **6**, 637 (1965).

[10] *Higher Transcendental Functions, Vol.* 1, edited by A. Erdélyi (McGraw-Hill, New York, 1953), Formula 3.6.2(21), p. 152.

[11] S. Mandelstam, Ann. Phys. **21**, 302 (1963). In this paper, Mandelstam first emphasized the importance of working with the reduced amplitude $b(l,s)$; see his Eq. (3.9). The function Δ^l in our Eq. (2.29) is defined in a cut Δ plane, with the cut running from $\Delta=0$ to $\Delta=-\infty$, in such a way that Δ^l is positive for positive Δ and real l.

[12] E.T. Whittaker and G.N. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, 1962), §15.33, p. 319.

[13] L.V. Kantorovich and G.P. Akilov, *Functional Analysis in Normed Spaces* (Pergamon, Oxford, 1964).

[14] W. Pogorselski, *Integral Equations and Their Applications, Vol. I* (Pergamon, Oxford, 1966), Chap. 15, p. 443.

[15] N.I. Muskhelishivili, *Singular Integral Equations* (Noordhoff, Groningen, 1953).

Second-quantization representation for a nonrelativistic system of composite particles. I. Generalized Tani transformation and its iterative evaluation*

M. D. Girardeau

Department of Physics and Institute of Theoretical Science, University of Oregon, Eugene, Oregon 97403
(Received 31 January 1975)

A method of constructing "atomic second quantization" representations is described, based on the introduction of redundant modes ("ideal atom variables") which are then given physical content by carrying out a suitable unitary transformation, the "generalized Tani transformation." In such a representation bound atoms or molecules are described by elementary Bose or Fermi operators, the field operators for nuclei and electrons referring only to unbound particles. The Hamiltonian thus obtained contains not only nucleus–nucleus, electron–electron, and nucleus–electron Coulomb interactions, but also atom–atom, atom–nucleus, and atom–electron Coulomb and exchange interactions, including breakup and recombination terms. All possible scattering and reaction channels are exhibited simultaneously in this transformed Hamiltonian. The method is applicable to any species or mixture of species of composite particles, but for simplicity the derivation is restricted here to the case of atomic hydrogen.

1. INTRODUCTION

There are many problems involving systems of composite particles in which their internal degrees of freedom cannot validly be ignored. Examples are high-temperature gases and partially ionized plasmas, molecular gases, chemical and nuclear reactions, Cooper pairs in superconductors, ferromagnetism, and phenomena in quantum liquids and solids involving real or virtual molecular excitation. In these and other problems, a representation in which the existence of the composite particles is treated kinematically, through use of appropriate composite-particle dynamical variables, is desirable. Recently several different methods[1–5] have been developed for obtaining such representations, in which *bound* composite particles are described by "ideal atom" annihilation and creation operators satisfying elementary Bose or Fermi commutation or anticommutation relations, and the *unbound* constituents are described by the usual elementary field operators. Of these various methods, it seems that the one[4] based on redundant modes and a generalization of Tani's canonical transformation[6] is both the simplest and the most easily generalized. Although the second-quantized Hamiltonian thus obtained is unitarily equivalent to the standard one in which only the field operators for the elementary constituents appear, it is more convenient for calculations, since known information about the atomic and/or molecular structure and exchange is kinematically built into the representation via the ideal atom operators and matrix elements involving atomic and/or molecular wavefunctions. The derivation was previously sketched in a highly abbreviated version.[4] The purpose of this paper is to develop a systematic method of evaluation of physical operators in this new representation, to exhibit the various terms in the transformed Hamiltonian in explicit form, and to discuss their physical significance. We shall limit ourselves here to the case of atomic hydrogen, for simplicity. The application to N-electron atoms has been discussed in connection with the theory of ferromagnetism,[7] but only those terms in the transformed Hamiltonian relevant to the ferromagnetism problem were considered there.

2. FORMULATION

Consider a system composed of hydrogen atoms, unbound protons, and unbound electrons (e.g., a partially ionized atomic hydrogen plasma). The standard "first-principles" approach to the study of such a system would be to formulate the quantum-mechanical many-body problem for the system in terms of states and observables expressed in terms of the Schrödinger dynamical variables of protons and electrons only, or, equivalently, in terms of the second-quantized proton field operators $\psi(X)$ and $\psi^\dagger(X)$ and electron field operators $\psi(x)$ and $\psi^\dagger(x)$. Here $X=(\mathbf{R},\sigma_p)$ where $\mathbf{R}$ is the proton position vector and σ_p its spin variable ($=\uparrow$ or $\downarrow$); similarly, $x=(\mathbf{r},\sigma_e)$ with $\mathbf{r}$ the electron position vector and σ_e its spin variable (again $\uparrow$ or $\downarrow$). In this representation the nonrelativistic Hamiltonian H is

$$
\begin{aligned}
&H = T_p + T_e + V_{pp} + V_{ee} + V_{pe},\\
&T_p = \int dX \psi^\dagger(X) T(X) \psi(X),\\
&T_e = \int dx \psi^\dagger(x) T(x) \psi(x),\\
&V_{pp} = \tfrac{1}{2}\int dX\, dX' \psi^\dagger(X)\psi^\dagger(X') V(XX')\psi(X')\psi(X),\\
&V_{ee} = \tfrac{1}{2}\int dx\, dx' \psi^\dagger(x)\psi^\dagger(x') V(xx')\psi(x')\psi(x),\\
&V_{pe} = \int dX\, dx\, \psi^\dagger(X)\psi^\dagger(x) V(Xx)\psi(x)\psi(X),
\end{aligned}
\tag{1}
$$

where $\int dX$ stands for $\sum_{\sigma_p}\int d^3R$, $\int dx$ stands for $\sum_{\sigma_e}\int d^3r$, $T(X)$ and $T(x)$ are the single-proton and single-electron kinetic energy operators plus external potentials (if present), and $V(XX')$, $V(xx')$, and $V(Xx)$ are the proton–proton, electron–electron, and proton–electron Coulomb interaction potentials. If such a representation is employed, the presence and properties of bound composite particles (here hydrogen atoms) are not explicit in the algebra of observables, but must be inserted essentially as boundary conditions on the state vectors.

For many-particle state vectors representing a system of nonzero density, this is a highly nontrivial problem. It is possible, however, to find a unitary transformation U such that $U^{-1}HU$, although in principle containing the same physical information as H, contains explicit "atomic dynamical variables" representing bound atoms, as well as proton and electron variables for the *unbound* constituents. Such a representation is more convenient for calculations since dynamical processes such as atom–atom, atom–proton, and atom–electron scattering, atomic ionization and recombination, etc. are exhibited explicitly in the Hamiltonian and are thus amenable to physical interpretation and physically motivated approximation procedures. Furthermore, representation of atoms by their own Bose of Fermi variables facilitates application of standard many-body calculational techniques (Green's functions, Wick's theorem, etc.) to problems where both bound atoms and their unbound constituents play an essential role.

As a preliminary to construction of the appropriate U, let us define the "physical-atom annihilation and creation operators" A_α and $A_\alpha^\dagger$. Let $\{\phi_\alpha(Xx)\}$ be the orthonormal but not complete[8] set of *bound* hydrogen-atom wavefunctions. The second-quantized state vector $|\phi_\alpha\rangle$ representing a single hydrogen atom in state ϕ_α is

$$|\phi_\alpha\rangle = A_\alpha^\dagger|0\rangle \tag{2}$$

where $|0\rangle$ is the normalized vacuum (no-particle state) and $A_\alpha^\dagger$ is the physical-atom creation operator

$$A_\alpha^\dagger \equiv \int dX\,dx\,\phi_\alpha(Xx)\psi^\dagger(X)\psi^\dagger(x). \tag{3}$$

The corresponding annihilation operator A_α is defined as $A_\alpha \equiv (A_\alpha^\dagger)^\dagger$. More generally, a state containing only bound atoms is a linear combination of atomic product states $A_{\alpha_1}^\dagger\cdots A_{\alpha_n}^\dagger|0\rangle$. In order to write down the commutation relations satisfied by these operators, it is necessary to adopt a definite convention regarding the commutation properties between the proton and electron fields. Although our treatment is nonrelativistic, it is convenient nevertheless to adopt the standard dogma of relativistic field theory: "Kinematically independent fermion fields anticommute."[7] Thus we take the proton and electron field operators to anticommute with each other.[9] Then with the standard anticommutation relations for the proton and electron fields separately, we have

$$\begin{aligned}
&[\psi(X),\psi(X')]_+ = 0, \quad [\psi(X),\psi^\dagger(X')]_+ = \delta(X-X'),\\
&[\psi(x),\psi(x')]_+ = 0, \quad [\psi(x),\psi^\dagger(x')]_+ = \delta(x-x'),\\
&[\psi(X),\psi(x)]_+ = [\psi(X),\psi^\dagger(x)]_+ = 0.
\end{aligned} \tag{4}$$

It is then straightforward to verify that the A_α and $A_\alpha^\dagger$ operators satisfy the commutation relations

$$\begin{aligned}
&[A_\alpha, A_\beta] = 0, \quad [A_\alpha, A_\beta^\dagger] = \delta_{\alpha\beta} + C_{\alpha\beta},\\
&[\psi(X), A_\alpha] = [\psi(x), A_\alpha] = 0,\\
&[\psi(X), A_\alpha^\dagger] = \int dx\,\phi_\alpha(Xx)\psi^\dagger(x),\\
&[\psi(x), A_\alpha^\dagger] = -\int dX\,\phi_\alpha(Xx)\psi^\dagger(X)
\end{aligned} \tag{5}$$

where the bracket $[A,B]$ with no subscript denotes the commutator, and

$$\begin{aligned}
C_{\alpha\beta} = &-\int dX\,dX' K_{\alpha\beta}(X,X')\psi^\dagger(X)\psi(X')\\
&-\int dx\,dx' K_{\alpha\beta}(x,x')\psi^\dagger(x)\psi(x')
\end{aligned} \tag{6}$$

where the "exchange kernels" $K_{\alpha\beta}$ are defined as[10]

$$\begin{aligned}
K_{\alpha\beta}(X,X') &= \int \phi_\alpha^*(X'x)\,\phi_\beta(Xx)\,dx,\\
K_{\alpha\beta}(x,x') &= \int \phi_\alpha^*(Xx')\,\phi_\beta(Xx)\,dX.
\end{aligned} \tag{7}$$

The presence of the operator $C_{\alpha\beta}$ in the commutation relation between A_α and $A_\beta^\dagger$ is a kinematical manifestation of the composite nature and internal structure of hydrogen atoms. Similarly, the fact that the commutators $[\psi(X),A_\alpha^\dagger]$ and $[\psi(x),A_\alpha^\dagger]$ are nonvanishing is a consequence of the lack of kinematical independence of hydrogen atoms from proton and electrons. We note, parenthetically, that the *anti*commutativity of the proton and electron fields was used in deriving the *commutation* relations of the proton and electron fields with the A_α and $A_\alpha^\dagger$ operators. It would also be consistent[9] to take the proton fields to *commute* with the electron fields. However, the second, third, and fourth lines of (5) would then be replaced by *anti*commutation relations, which would be less convenient for the subsequent development.

The interpretation of the kernels $K_{\alpha\beta}$ as arising from exchange of electrons or protons between atoms α and β is intuitively clear from their definition. Since interatomic exchange can only occur when the atomic wavefunctions overlap, we expect the $K_{\alpha\beta}$, and hence $C_{\alpha\beta}$, to vanish in the limit of vanishing overlap. To see that this is indeed the case, suppose that two atoms with wavefunctions ϕ_α and ϕ_β $(\alpha\neq\beta)$ are localized in disjoint regions $\mathcal{R}_\alpha$ and $\mathcal{R}_\beta$, in the sense that ϕ_α vanishes unless both $\mathbf{R}\in\mathcal{R}_\alpha$ and $\mathbf{r}\in\mathcal{R}_\alpha$, and similarly for ϕ_β, where $\mathcal{R}_\alpha$ and $\mathcal{R}_\beta$ are disjoint. It is then easy to see that both of the kernels $K_{\alpha\beta}$ [Eq. (7)] vanish identically in their arguments, and hence the operator $C_{\alpha\beta}$ also vanishes. This confirms our physical expectation that the commutator $[A_\alpha,A_\beta^\dagger]$ should vanish for $\alpha\neq\beta$ when the atoms α and β do not overlap. On the other hand, the diagonal elements $K_{\alpha\alpha}$ and hence the operator $C_{\alpha\alpha}$ do not vanish; they represent exchange effects of strongly overlapping atomic wavefunctions. Similarly, one can interpret the nonzero expressions (5) for the commutators $[\psi(X),A_\alpha^\dagger]$ and $[\psi(x),A_\alpha^\dagger]$ as representing the kinematical effects of exchange of an unbound proton or electron with the atomic proton or electron. It is clear from (5) that $[\psi(X),A_\alpha^\dagger]$ vanishes identically if ϕ_α is localized in $\mathcal{R}_\alpha$ and the proton position $\mathbf{R}$ is outside $\mathcal{R}_\alpha$ and similarly $[\psi(x),A_\alpha^\dagger]$ vanishes identically if the electron position $\mathbf{r}$ is outside $\mathcal{R}_\alpha$.

The nontrivial commutation relations (5) would lead to computational difficulties if one were to employ the A_α and $A_\alpha^\dagger$ operators as atomic dynamical variables. E.g., the atomic product states $A_{\alpha_1}^\dagger\cdots A_{\alpha_n}^\dagger|0\rangle$ are neither orthonormal nor independent of the proton–electron product states $\psi^\dagger(X_1)\cdots\psi^\dagger(X_n)\psi^\dagger(x_1)\cdots\psi^\dagger(x_n)|0\rangle$. A related difficulty is that Wick's theorem only applies to operators satisfying elementary Bose or Fermi commutation or anticommutation relations. To overcome these difficulties, we employ the familiar "redundant-mode" technique. Define the "Schrödinger state space" $\mathcal{S}$ as the space of all normalizable[11] linear combinations

of proton–electron product states $\psi^\dagger(X_1)\cdots\psi^\dagger(X_m)\psi^\dagger(x_1)\cdots\psi^\dagger(x_l)|0\rangle$ with arbitrary m and l. Also define a *completely independent* Hilbert space $\mathcal{A}$, the "ideal atom space," as the space of all normalizable linear combinations of "ideal atomic product states"[12] $a^\dagger_{\alpha_1}\cdots a^\dagger_{\alpha_n}|0\rangle$ with arbitrary n, where, *by definition*, the a_α and $a^\dagger_\alpha$ satisfy *elementary* Bose commutation relations

$$[a_\alpha, a_\beta]=0, \quad [a_\alpha, a^\dagger_\beta]=\delta_{\alpha\beta}. \tag{8}$$

Finally, define the "ideal state space" $\mathcal{G}$ to be the direct product space $\mathcal{G}=\mathcal{S}\otimes\mathcal{A}$. Since the ideal atom operators are kinematically independent of the proton and electron fields, they satisfy

$$[a_\alpha, \psi(X)]=[a_\alpha, \psi^\dagger(X)]=[a_\alpha, \psi(x)]=[a_\alpha, \psi^\dagger(x)]=0 \tag{9}$$

on $\mathcal{G}$. The commutation relations (8), initially defined on $\mathcal{A}$, are also valid on $\mathcal{G}$, as are the commutation relations (5).

The basic idea is now to establish a one-to-one correspondence between "physical state vectors" in $\mathcal{S}$ and "ideal state vectors" contained in a certain subspace of $\mathcal{G}$. On the one hand, such an approach is in the same spirit as Dyson's treatment of the Heisenberg model,[13] wherein such an "ideal state space" was introduced in order to circumvent the problem of the non-Bose commutation relations of spin-wave annihilation and creation operators. On the other hand, the method we shall use to set up the correspondence between $\mathcal{S}$ and $\mathcal{G}$, namely that of initially interpreting the ideal atoms as "redundant modes" and then giving them physical content by carrying out an appropriate unitary transformation U, is closely related to the Bohm–Pines theory of plasma oscillations[14] and to some treatments of collective modes in nuclei.

We note first that the Schrödinger space $\mathcal{S}$ is trivially isomorphic with the subspace $\mathcal{G}_0$ of $\mathcal{G}$ consisting of those $|\psi\rangle\in\mathcal{G}$ which satisfy the constraints

$$a_\alpha|\psi\rangle=0, \quad \text{all } \alpha, \quad |\psi\rangle\in\mathcal{G}_0, \tag{10}$$

i.e., they are eigenstates of all the ideal atom occupation number operators $N_\alpha=a^\dagger_\alpha a_\alpha$ with eigenvalues zero. Equivalently, $\mathcal{G}_0$ is the subspace of $|\psi\rangle\in\mathcal{G}$ satisfying

$$N_a|\psi\rangle=0, \quad |\psi\rangle\in\mathcal{G}_0 \tag{11}$$

where N_a is the ideal atom total number operator

$$N_a=\sum_\alpha a^\dagger_\alpha a_\alpha. \tag{12}$$

Thus in $\mathcal{G}_0$ the ideal atoms are "redundant modes" (totally unoccupied), which makes the isomorphism with $\mathcal{S}$ trivial and obvious.

We next try to find a unitary transformation U which in some sense shifts the description of bound atomic states to the a_α and $a^\dagger_\alpha$ operators, so that they acquire physical content (note, again, the close analogy with the Bohm–Pines theory).[14] For the case of one-atom states, this is easy to do. The physical atom state (2) is transformed into the ideal atom state $a^\dagger_\alpha|0\rangle$ by the operator $a^\dagger_\alpha A_\alpha$:

$$(a^\dagger_\alpha A_\alpha)A^\dagger_\alpha|0\rangle=a^\dagger_\alpha|0\rangle, \tag{13}$$

as is easily verified by Eqs. (5)–(9). Although the operator $a^\dagger_\alpha A_\alpha$ is not unitary, one can construct a unitary operator U which does the same job, as follows. Define the antihermitian operator F by

$$F=\sum_\alpha (a^\dagger_\alpha A_\alpha - A^\dagger_\alpha a_\alpha). \tag{14}$$

Then one readily verifies that

$$FA^\dagger_\alpha|0\rangle=a^\dagger_\alpha|0\rangle, \quad Fa^\dagger_\alpha|0\rangle=-A^\dagger_\alpha|0\rangle. \tag{15}$$

Define the unitary operator $U(\epsilon)$ by

$$U(\epsilon)=\exp(\epsilon F). \tag{16}$$

Then by expansion of the exponential and iteration of (15) one can verify that

$$\begin{aligned} U(\epsilon)A^\dagger_\alpha|0\rangle &= \left(1-\frac{\epsilon^2}{2!}+\frac{\epsilon^4}{4!}-\cdots\right)A^\dagger_\alpha|0\rangle \\ &\quad+\left(\epsilon-\frac{\epsilon^3}{3!}+\frac{\epsilon^5}{5!}-\cdots\right)a^\dagger_\alpha|0\rangle \\ &=(\cos\epsilon)A^\dagger_\alpha|0\rangle+(\sin\epsilon)a^\dagger_\alpha|0\rangle. \end{aligned} \tag{17}$$

If we define U by choosing $\epsilon=\pi/2$,

$$U=\exp[(\pi/2)F], \tag{18}$$

then our goal of constructing a unitary U such that

$$UA^\dagger_\alpha|0\rangle=a^\dagger_\alpha|0\rangle \tag{19}$$

is achieved. We call U a "generalized Tani transformation" since it is an obvious generalization of a transformation employed by Tani[6] to formulate the theory of single-particle scattering by a potential with a bound state. There is also a chose connection with the "quasi-chemical equilibrium" theory of Blatt and Matsubara.[15] However, we work with a unitary transformation U, whereas Blatt and Matsubara employed a nonunitary transformation which, in our notation, is $\exp(\sum_\alpha a^\dagger_\alpha A_\alpha)$. The unitary transformation has the advantage of preserving hermiticity of observables, the eigenvalue spectrum of the Hamiltonian, normalization and more generally matrix elements, etc. Bohm and Pines also employed a unitary transformation.[14]

The effect of U on a many-body state is more complicated than (19). However, in the approximation in which $C_{\alpha\beta}$ in (5) may be neglected (we have seen already that this is the case in the zero-density limit), U converts a physical-atom product state into an ideal-atom product state. To see this, note first that

$$UA^\dagger_{\alpha_1}\cdots A^\dagger_{\alpha_n}|0\rangle=(UA^\dagger_{\alpha_1}U^{-1})(UA^\dagger_{\alpha_2}U^{-1})\cdots(UA^\dagger_{\alpha_n}U^{-1})|0\rangle, \tag{20}$$

since[16]

$$F|0\rangle=0, \quad U|0\rangle=|0\rangle. \tag{21}$$

The unitary transforms in (20) can in principle be evaluated from the multiple commutator expansion

$$\begin{aligned} UA^\dagger_\alpha U^{-1} &= \exp\left(\frac{\pi}{2}F\right)A^\dagger_\alpha\exp\left(-\frac{\pi}{2}F\right) \\ &=A^\dagger_\alpha+\sum_{j=1}^{\infty}\frac{(-\pi/2)^j}{j!}[A^\dagger_\alpha, F]_j \end{aligned} \tag{22}$$

where $[A^\dagger_\alpha, F]_j$ denotes the multiple commutator defined recursively by

$$[A^\dagger_\alpha, F]_1=[A^\dagger_\alpha, F], \quad [A^\dagger_\alpha, F]_{j+1}=[[A^\dagger_\alpha, F]_j, F]. \tag{23}$$

It follows from (14) and (5)–(9) that

$$[A_\alpha^\dagger, F] = -a_\alpha^\dagger - \sum_\beta a_\beta^\dagger C_{\beta\alpha},$$

$$[A_\alpha^\dagger, F] = A_\alpha^\dagger. \tag{24}$$

In the approximation in which $C_{\alpha\beta}$ in (5) is neglected, only the term $-a_\alpha^\dagger$ in the top line of (24) is to be retained, so that (22) becomes

$$UA_\alpha^\dagger U^{-1} \approx A_\alpha^\dagger \cos(\pi/2) + a_\alpha^\dagger \sin(\pi/2) = a_\alpha^\dagger. \tag{25}$$

In the same approximation, (20) reduces to

$$UA_{\alpha_1}^\dagger \cdots A_{\alpha_n}^\dagger |0\rangle \approx a_{\alpha_1}^\dagger \cdots a_{\alpha_n}^\dagger |0\rangle. \tag{26}$$

In fact, this approximation becomes an exact equality if the atoms are all localized in a set of mutually disjoint regions. This is proved in Appendix A. The physical interpretation is that in the zero-density limit, bound hydrogen atoms behave like elementary bosons, and the physical-atom product states behave like orthonormal product states of elementary bosons.[17] This is already physically obvious without introducing the transformation U. The advantage of the use of such a transformation is that it will enable us to introduce Bose dynamical variables for bound hydrogen atoms into the problem in a physically meaningful way even at realistic densities, where the internal atomic structure and exchange effects are important.

Since the transformation U is constructed in such a way as to replace physical-atom variables $A_\alpha^\dagger$ by ideal-atom variables $a_\alpha^\dagger$, one expects that in a physical situation where no bound atoms ϕ_α can form, the transformation U will have no effect, i.e., it will behave like the unit operator. In fact, suppose that $|\psi\rangle$ is a state of electrons only, or of protons only [so that bound atoms $\phi_\alpha(Xx)$ cannot be present]. Then it follows trivially from (3), (14), and (18) that

$$F|\psi\rangle = 0, \quad U|\psi\rangle = |\psi\rangle. \tag{27}$$

More generally, suppose that all the α indices occurring in (14) refer to atoms ϕ_α localized in some region $\mathcal{R}$, in the sense defined in Appendix A, and suppose that $|\psi\rangle$ is any state in which all the protons and electrons are localized in a region $\mathcal{R}'$, where $\mathcal{R}$ and $\mathcal{R}'$ are disjoint. Then it is easy to show that (27) still holds. This justifies the interpretation that after the transformation U, the proton and electron field operators refer only to *unbound* protons and electrons.

In addition to the ideal-atom number operator (12), consider the proton and electron number operators N_p and N_e:

$$N_p = \int dX \psi^\dagger(X)\psi(X), \quad N_e = \int dx \psi^\dagger(x)\psi(x). \tag{28}$$

Let $|\psi\rangle$ be any m-proton, l-electron state in the subspace[18] $\mathcal{G}_0$ [Eq. (11)]:

$$N_a|\psi\rangle = 0, \quad N_p|\psi\rangle = m|\psi\rangle, \quad N_e|\psi\rangle = l|\psi\rangle. \tag{29}$$

Let $|\psi)$ be the image of such a state under the generalized Tani transform:

$$|\psi\rangle = U|\psi), \quad |\psi) = U^{-1}|\psi\rangle. \tag{30}$$

It is clear from (14), (18), (12), and (27)–(29) that

$$[(N_a + N_p), U] = [(N_a + N_e), U] = 0,$$

$$(N_a + N_p)|\psi) = m|\psi), \quad (N_a + N_e)|\psi) = l|\psi). \tag{31}$$

The physical interpretation is that in $|\psi)$ the *total* number of protons is the sum of the number of atoms (each containing one proton) and the number of *free* (unbound) protons, and similarly for the total number of electrons. More generally, if A is any physical observable on the Schrödinger state space $\mathcal{S}$, expressed in terms of proton and electron field operators, one has

$$[A, N_a] = [A, N_p] = [A, N_e] = 0, \tag{32}$$

the first following trivially because A does not contain a_α and $a_\alpha^\dagger$ operators, whereas the second and third follow from conservation of the numbers of protons and electrons. Then with (30) one has

$$[(N_a + N_p), U^{-1}AU] = [(N_a + N_e), U^{-1}AU] = 0. \tag{33}$$

This establishes two superselection rules: For any physical observable A, its generalized Tani transform $U^{-1}AU$ has nonvanishing matrix elements only between states with the same eigenvalue of $N_a + N_p$, and also the same eigenvalue of $N_a + N_e$. These superselection rules are merely the transforms, under U, of the usual particle number selection rules. They have important consequences for the structure of transformed observables $U^{-1}AU$. E.g., we shall find that $U^{-1}HU$ has terms representing the breakup of an atom into a proton and an electron, but, as expected physically, none with the "wrong" numbers of outgoing particles of various species.

States $|\psi)$ related to states $|\psi\rangle \in \mathcal{G}_0$ by (29) define a subspace of $\mathcal{G}$ which we shall denote by $\mathcal{G}_{\text{phys}}$. Symbolically, one can write

$$\mathcal{G}_{\text{phys}} = U^{-1}\mathcal{G}_0. \tag{34}$$

$\mathcal{G}_{\text{phys}}$ is the subspace of those states $|\psi) \in \mathcal{G}$ satisfying subsidiary conditions which are the transform of (10):

$$(U^{-1}a_\alpha U)|\psi) = 0, \quad \text{all } \alpha, \quad |\psi) \in \mathcal{G}_{\text{phys}}, \tag{35}$$

or equivalently, by (11),

$$(U^{-1}N_a U)|\psi) = 0, \quad |\psi) \in \mathcal{G}_{\text{phys}}. \tag{36}$$

For any two physical states $|\psi\rangle$, $|\psi'\rangle \in \mathcal{S}$, there are essentially identical[18] states in $\mathcal{G}_0$, which have images $|\psi)$, $|\psi') \in \mathcal{G}_{\text{phys}}$. Any calculation in the conventional Schrödinger space $\mathcal{S}$ is equivalent to a calculation in $\mathcal{G}_{\text{phys}}$. E.g., for any observable A, one has trivially

$$\langle\psi|A|\psi'\rangle = (\psi|U^{-1}AU|\psi'). \tag{37}$$

As mentioned before, the advantage of carrying out the calculation in $\mathcal{G}_{\text{phys}}$ is that processes involving existence, excitation, and ionization of bound atoms are then built explicitly into the algebra of observables and hence are exhibited explicitly in $U^{-1}AU$. Discussion of the transformed subsidiary condition (36) will be deferred until Sec. 5. We note here only that the subsidiary condition is dynamically consistent, in the sense that the original Hamiltonian (1) commutes trivially with N_a [Eq. (12)] and hence the same holds for the transformed operators:

$$[U^{-1}HU, U^{-1}N_aU] = 0. \tag{38}$$

3. EQUATIONS OF MOTION AND ITERATIVE SOLUTION

The multiple commutator expansion (22) used in evaluation of the approximation (25) is not useful for evaluation of more accurate expressions for the transforms. The reason is that even the "zero order" approximation (25) involved summation of infinite series. As soon as the operator term $C_{\alpha\beta}$ in (5) is taken into account, the infinite series generated by the multiple commutator expansion are more complicated, and there is little hope of recognizing the general terms of the relevant series and of summing them. What is needed is a more efficient method of evaluating the transforms, such that the evaluation of the zero order approximations is trivial. Then one can expect that the problem of finding the low-order corrections due to atomic structure (i. e., due to the $C_{\alpha\beta}$) will be tractable.

The "equation of motion" method provides such a method of calculation. Define $U(\epsilon)$ by (16), and, for any operator A,

$$A(\epsilon) \equiv U^{-1}(\epsilon) A U(\epsilon). \tag{39}$$

Then with (18) one has

$$A(0) = A, \quad A(\pi/2) = U^{-1} A U. \tag{40}$$

Differentiation of (39) yields the "equation of motion"

$$\frac{dA(\epsilon)}{d\epsilon} = [A(\epsilon), F(\epsilon)] = \exp(-\epsilon F)[A, F]\exp(\epsilon F). \tag{41}$$

We need the explicit forms of (41) for the cases $A = a_\alpha$, $A = A_\alpha$, $A = \psi(X)$, and $A = \psi(x)$. Using the commutation relations (5), (8), and (9), one finds with (14)

$$\begin{aligned}
&\frac{da_\alpha(\epsilon)}{d\epsilon} = A_\alpha(\epsilon),\\
&\frac{dA_\alpha(\epsilon)}{d\epsilon} = -a_\alpha(\epsilon) - \sum_\beta C_{\alpha\beta}(\epsilon) a_\beta(\epsilon),\\
&\frac{\partial\psi(X,\epsilon)}{\partial\epsilon} = -\sum_\alpha \int dx\, \phi_\alpha(Xx)\psi^\dagger(x,\epsilon) a_\alpha(\epsilon),\\
&\frac{\partial\psi(x,\epsilon)}{\partial\epsilon} = \sum_\alpha \int dX\, \phi_\alpha(Xx)\psi^\dagger(X,\epsilon) a_\alpha(\epsilon)
\end{aligned} \tag{42}$$

where, by (6),

$$\begin{aligned}
C_{\alpha\beta}(\epsilon) = &-\int dX\, dX' K_{\alpha\beta}(X, X')\psi^\dagger(X,\epsilon)\psi(X',\epsilon)\\
&-\int dx\, dx' K_{\alpha\beta}(x, x')\psi^\dagger(x,\epsilon)\psi(x',\epsilon).
\end{aligned} \tag{43}$$

Equations (42) and their Hermitian conjugates are coupled, nonlinear differential equations for the unknown operator functions $a_\alpha(\epsilon)$, $A_\alpha(\epsilon)$, $\psi(X,\epsilon)$, $\psi(x,\epsilon)$, and their Hermitian conjugates. In view of their nonlinearity, they cannot be solved in closed form. However, it is a straightforward matter to solve them iteratively (method of successive approximations) starting with the zero order approximations discussed previously. In fact, these zero order solutions are themselves most easily found by approximate solution of (42). In the approximation in which the atoms behave like elementary bosons, and hence the term $C_{\alpha\beta}$ in the commutations (5) is dropped, the first two differential equations (42) reduce to

$$\frac{da_\alpha^{(0)}(\epsilon)}{d\epsilon} = A_\alpha^{(0)}(\epsilon), \quad \frac{dA_\alpha^{(0)}(\epsilon)}{d\epsilon} = -a_\alpha^{(0)}(\epsilon). \tag{44}$$

By differentiating the first of these again and substituting from the second, or vice versa, one finds

$$\frac{d^2 a_\alpha^{(0)}(\epsilon)}{d\epsilon^2} = -a_\alpha^{(0)}(\epsilon), \quad \frac{d^2 A_\alpha^{(0)}(\epsilon)}{d\epsilon^2} = -A_\alpha^{(0)}(\epsilon), \tag{45}$$

which have the general solutions

$$\begin{aligned}
&a_\alpha^{(0)}(\epsilon) = c_\alpha \cos\epsilon + d_\alpha \sin\epsilon,\\
&A_\alpha^{(0)}(\epsilon) = f_\alpha \cos\epsilon + g_\alpha \sin\epsilon
\end{aligned} \tag{46}$$

where C_α, d_α, f_α, and g_α are operator "constants of integration" to be determined from the "initial conditions"

$$a_\alpha^{(0)}(0) = a_\alpha, \quad A_\alpha^{(0)}(0) = A_\alpha \tag{47}$$

[note Eq. (40)]. Determining the coefficients in this way, one finds

$$\begin{aligned}
&a_\alpha^{(0)}(\epsilon) = a_\alpha \cos\epsilon + A_\alpha \sin\epsilon,\\
&A_\alpha^{(0)}(\epsilon) = A_\alpha \cos\epsilon - a_\alpha \sin\epsilon,
\end{aligned} \tag{48}$$

which reduces to (25) if $\epsilon = -\pi/2$ [note that $U(-\pi/2) = U^{-1}$]. In the same approximation, the A_α and $A_\alpha^\dagger$ operators should be regarded as kinematically independent of the ψ and $\psi^\dagger$ operators (as they would be if the atoms were elementary), and hence the commutators $[\psi(X), A_\alpha^\dagger]$ and $[\psi(x), A_\alpha^\dagger]$ should be neglected [see Eq. (5)]. Then the differential equations (42) for the ψ operators reduce to

$$\frac{\partial\psi^{(0)}(X,\epsilon)}{\partial\epsilon} = 0 = \frac{\partial\psi^{(0)}(x,\epsilon)}{\partial\epsilon}, \tag{49}$$

with trivial solutions

$$\psi^{(0)}(X,\epsilon) = \psi(X), \quad \psi^{(0)}(x,\epsilon) = \psi(x) \tag{50}$$

satisfying the initial conditions at $\epsilon = 0$.

To proceed to higher order we need a systematic way of classifying the orders of various terms. Since the terms neglected in obtaining (48) and (50) are of positive degree in bound atom wavefunctions ϕ_α and ϕ_α^*, it is natural to order various contributions according to their degree in the ϕ_α and ϕ_α^*. Thus we write

$$\begin{aligned}
&a_\alpha(\epsilon) = \sum_{j=0}^{\infty} a_\alpha^{(j)}(\epsilon), \quad A_\alpha(\epsilon) = \sum_{j=0}^{\infty} A_\alpha^{(j)}(\epsilon),\\
&\psi(X,\epsilon) = \sum_{j=0}^{\infty} \psi^{(j)}(X,\epsilon), \quad \psi(x,\epsilon) = \sum_{j=0}^{\infty} \psi^{(j)}(x,\epsilon)
\end{aligned} \tag{51}$$

where the superscript j denotes that the given contribution is of degree j in the ϕ_α and ϕ_α^*. These expansions are not in powers of the large parameter $\pi/2$ in (18), but instead are essentially in powers of the density of the system, since the bound state wavefunctions enter via the operator terms on the right sides of Eqs. (5), which have negligible effects in the zero density limit (all particles infinitely far apart) in which the atoms behave as elementary. We shall adopt the convention that in counting the orders of terms involving the A_α and/or $A_\alpha^\dagger$, the implicit factors of ϕ_α and/or ϕ_α^* entering via the definition (3) are *not* to be counted; this is consistent with (48), and is desirable so as to preserve the symmetry between a_α and A_α.

The differential equations for the first order corrections $a_\alpha^{(1)}$, etc., are obtained by evaluating the first degree (in the ϕ_α and ϕ_α^*) terms on the right sides of Eqs. (42), making use of the zero order solutions. It is clear from (43) and (7) that $C_{\alpha\beta}(\epsilon)$ involves terms of second and higher degree in factors ϕ_α and ϕ_α^*, the terms of second degree coming from the terms $\psi^{(0)}$ and $\psi^{(0)\dagger}$ in the ψ and $\psi^\dagger$ operators. Hence there are no first order terms on the right sides of the first two differential equations (42), except for the terms linear in $A_\alpha^{(\epsilon)}$ and $a_\alpha(\epsilon)$, so that

$$\frac{da_\alpha^{(1)}(\epsilon)}{d\epsilon} = A_\alpha^{(1)}(\epsilon), \quad \frac{dA_\alpha^{(1)}(\epsilon)}{d\epsilon} = -a_\alpha^{(1)}(\epsilon). \tag{52}$$

Since the initial conditions (47), (50) on the zero order solutions were chosen so that the initial conditions are assigned entirely to the zero order terms, it is clear from (40) and (51) that one must put

$$\begin{aligned} &a_\alpha^{(j)}(0) = 0 = A_\alpha^{(j)}(0), \\ &\psi^{(j)}(X,0) = 0 = \psi^{(j)}(x,0), \quad j \geq 1. \end{aligned} \tag{53}$$

Thus we must pick the trivial solution of (52),

$$a_\alpha^{(1)}(\epsilon) = 0 = A_\alpha^{(1)}(\epsilon), \tag{54}$$

as would also be obtained by putting c_α, d_α, f_α, and g_α all equal to zero in (46). The corresponding first order equations for the proton and electron fields are obtained by replacing the $\psi^\dagger$ and a operators on the right sides of the third and fourth differential equations (42) by their zero order approximations:

$$\begin{aligned} \frac{\partial\psi^{(1)}(X,\epsilon)}{\partial\epsilon} &= -\sum_\alpha \int dx\, \phi_\alpha(Xx)\psi^{(0)\dagger}(x,\epsilon)a_\alpha^{(0)}(\epsilon) \\ &= -\sum_\alpha \int dx\, \phi_\alpha(Xx)\psi^\dagger(x)(a_\alpha\cos\epsilon + A_\alpha\sin\epsilon), \\ \frac{\partial\psi^{(1)}(x,\epsilon)}{\partial\epsilon} &= \sum_\alpha \int dX\, \phi_\alpha(Xx)\psi^{(0)\dagger}(X,\epsilon)a_\alpha^{(0)}(\epsilon) \\ &= \sum_\alpha \int dX\, \phi_\alpha(Xx)\psi^\dagger(X)(a_\alpha\cos\epsilon + A_\alpha\sin\epsilon). \end{aligned} \tag{55}$$

Then with the initial conditions (53) one has, upon integration from 0 to ϵ,

$$\begin{aligned} \psi^{(1)}(X,\epsilon) &= -\sum_\alpha \int dx\, \phi_\alpha(Xx)\psi^\dagger(x)[a_\alpha\sin\epsilon + A_\alpha(1-\cos\epsilon)], \\ \psi^{(1)}(x,\epsilon) &= \sum_\alpha \int dX\, \phi_\alpha(Xx)\psi^\dagger(X)[a_\alpha\sin\epsilon + A_\alpha(1-\cos\epsilon)]. \end{aligned} \tag{56}$$

The second order equations for $a_\alpha(\epsilon)$ and $A_\alpha(\epsilon)$ are obtained by replacing $C_{\alpha\beta}(\epsilon)$ and $a_\beta(\epsilon)$ by $C_{\alpha\beta}^{(2)}(\epsilon)$ and $a_\beta^{(0)}(\epsilon)$ in the second Eq. (42), since, as noted previously, $C_{\alpha\beta}$ starts with a second order term $C_{\alpha\beta}^{(2)}$. Thus

$$\begin{aligned} \frac{da_\alpha^{(2)}(\epsilon)}{d\epsilon} &= A_\alpha^{(2)}(\epsilon), \\ \frac{dA_\alpha^{(2)}(\epsilon)}{d\epsilon} &= -a_\alpha^{(2)}(\epsilon) - \sum_\beta C_{\alpha\beta}^{(2)}(\epsilon)a_\beta^{(0)}(\epsilon). \end{aligned} \tag{57}$$

The expression for $C_{\alpha\beta}^{(2)}(\epsilon)$ is to be obtained by replacing ψ and $\psi^\dagger$ by $\psi^{(0)}$ and $\psi^{(0)\dagger}$ in (43); hence by (50) $C_{\alpha\beta}^{(2)}(\epsilon)$ reduces simply to $C_{\alpha\beta}$, Eq. (6). Then with (48) one finds

$$\frac{da_\alpha^{(2)}(\epsilon)}{d\epsilon} = A_\alpha^{(2)}(\epsilon),$$

$$\frac{dA_\alpha^{(2)}(\epsilon)}{d\epsilon} = -a_\alpha^{(2)}(\epsilon) - \sum_\beta C_{\alpha\beta}(a_\beta\cos\epsilon + A_\beta\sin\epsilon). \tag{58}$$

This is a set of coupled, inhomogeneous first order differential equations, which can be solved by the standard method of variation of constants. The solution satisfying the initial conditions (53) is

$$\begin{aligned} a_\alpha^{(2)}(\epsilon) &= -\tfrac{1}{2}\sum_\beta C_{\alpha\beta}[a_\beta\epsilon\sin\epsilon + A_\beta(\sin\epsilon - \epsilon\cos\epsilon)], \\ A_\alpha^{(2)}(\epsilon) &= -\tfrac{1}{2}\sum_\beta C_{\alpha\beta}[A_\beta\epsilon\sin\epsilon + a_\beta(\sin\epsilon + \epsilon\cos\epsilon)]. \end{aligned} \tag{59}$$

The correctness of (59) can be verified by direct substitution into (58).

The second order equations for the ψ operators are seen from (42) and (54) to be

$$\begin{aligned} \frac{\partial\psi^{(2)}(X,\epsilon)}{\partial\epsilon} &= -\sum_\alpha \int dx\, \phi_\alpha(Xx)\psi^{(1)\dagger}(x,\epsilon)a_\alpha^{(0)}(\epsilon), \\ \frac{\partial\psi^{(2)}(x,\epsilon)}{\partial\epsilon} &= \sum_\alpha \int dX\, \phi_\alpha(Xx)\psi^{(1)\dagger}(X,\epsilon)a_\alpha^{(0)}(\epsilon). \end{aligned} \tag{60}$$

Substituting from (56) and (48) and carrying out the straightforward integrations from 0 to ϵ, one finds

$$\begin{aligned} \psi^{(2)}(X,\epsilon) &= -\sum_{\alpha\beta} \int dx\, dX'\, \phi_\alpha^*(X'x)\phi_\beta(Xx)A_{\alpha\beta}(\epsilon)\psi(X'), \\ \psi^{(2)}(x,\epsilon) &= -\sum_{\alpha\beta} \int dX\, dx'\, \phi_\alpha^*(Xx')\phi_\beta(Xx)A_{\alpha\beta}(\epsilon)\psi(x'), \\ A_{\alpha\beta}(\epsilon) &\equiv \tfrac{1}{2}a_\alpha^\dagger a_\beta\sin^2\epsilon + \tfrac{1}{2}a_\alpha^\dagger A_\beta(\epsilon - \sin\epsilon\cos\epsilon) \\ &\quad + A_\alpha^\dagger a_\beta(\sin\epsilon - \tfrac{1}{2}\epsilon - \tfrac{1}{2}\sin\epsilon\cos\epsilon) \\ &\quad + A_\alpha^\dagger A_\beta(1 - \cos\epsilon - \tfrac{1}{2}\sin^2\epsilon). \end{aligned} \tag{61}$$

The iterative process of solution can in principle be carried to any desired order, although the complexity of the expressions obtained increases rapidly with order. As illustrated by the case $j=2$, $a_\alpha^{(j)}(\epsilon)$ and $A_\alpha^{(j)}(\epsilon)$ for arbitrary $j \geq 2$ can be obtained by solving an inhomogeneous pair of coupled linear differential equations, with homogeneous solution given by (46) with an inhomogeneous term known in terms of the previously obtained lower order solutions. On the other hand, $\psi^{(j)}(X,\epsilon)$ and $\psi^{(j)}(x,\epsilon)$ can always be obtained by direct integration, from 0 to ϵ, of an expression involving only the known lower order solutions, as illustrated by the cases $j=1$ and $j=2$. We shall find in Sec. 4 that the leading terms (for low density) in the transformed Hamiltonian involve only the $\psi^{(j)}$ for $j \leq 3$. Thus the only remaining expressions needed are those for the third order proton and electron field operators $\psi^{(3)}$. Since the method of derivation has already been illustrated by the cases $j=1$ and $j=2$, we shall merely exhibit the final expressions for $j=3$:

$$\begin{aligned} \psi^{(3)}(X,\epsilon) &= \sum_{\alpha\beta\gamma} \int dx\, dX'\, dx'\, \phi_\alpha^*(X'x)\phi_\beta(Xx)\phi_\gamma(X'x')\psi^\dagger(x')A_{\alpha\beta\gamma}(\epsilon) \\ &\quad + \sum_{\alpha\beta} \int dx\, \phi_\alpha(Xx)\psi^\dagger(x)C_{\alpha\beta}[\tfrac{1}{2}a_\beta(\sin\epsilon - \epsilon\cos\epsilon) \\ &\quad + A_\beta(1 - \cos\epsilon - \tfrac{1}{2}\epsilon\sin\epsilon)], \\ \psi^{(3)}(x,\epsilon) &= -\sum_{\alpha\beta\gamma} \int dX\, dX'\, dx'\, \phi_\alpha^*(Xx')\phi_\beta(Xx)\phi_\gamma(X'x')\psi^\dagger(X')A_{\alpha\beta\gamma}(\epsilon) \\ &\quad - \sum_{\alpha\beta} \int dX\, \phi_\alpha(Xx)\psi^\dagger(X)C_{\alpha\beta}[\tfrac{1}{2}a_\beta(\sin\epsilon - \epsilon\cos\epsilon) \end{aligned}$$

$$+A_\beta(1-\cos\epsilon-\tfrac{1}{2}\epsilon\sin\epsilon)],$$

$$A_{\alpha\beta\gamma}(\epsilon)\equiv\tfrac{1}{6}a_\alpha^\dagger a_\gamma a_\beta\sin^3\epsilon-\tfrac{1}{3}a_\alpha^\dagger a_\gamma A_\beta(\cos\epsilon-1+\tfrac{1}{2}\sin^2\epsilon\cos\epsilon)$$
$$+\tfrac{1}{2}a_\alpha^\dagger A_\gamma a_\beta[\sin^2\epsilon+1-\cos\epsilon-\epsilon\sin\epsilon+\tfrac{1}{3}(\cos^3\epsilon-1)]$$
$$-\tfrac{1}{2}a_\alpha^\dagger A_\gamma A_\beta(\sin\epsilon\cos\epsilon-\epsilon+\sin\epsilon-\epsilon\cos\epsilon+\tfrac{1}{3}\sin^3\epsilon)$$
$$+\tfrac{1}{2}A_\alpha^\dagger a_\gamma a_\beta[\cos\epsilon-1+\epsilon\sin\epsilon+\tfrac{1}{3}(\cos^3\epsilon-1)]$$
$$+\tfrac{1}{2}A_\alpha^\dagger a_\gamma A_\beta(\sin\epsilon-\epsilon\cos\epsilon-\tfrac{1}{3}\sin^3\epsilon)$$
$$+A^\dagger A_\gamma a_\beta(\sin\epsilon-\tfrac{1}{2}\sin\epsilon\cos\epsilon-\tfrac{1}{2}\epsilon-\tfrac{1}{6}\sin^3\epsilon)$$
$$+A_\alpha^\dagger A_\gamma A_\beta[\tfrac{2}{3}(1-\cos\epsilon)-\tfrac{1}{2}\sin^2\epsilon+\tfrac{1}{6}\sin^2\epsilon\cos\epsilon]. \quad (62)$$

According to (40), the generalized Tani transforms $U^{-1}AU$ are obtained by substitution of $\epsilon=\pi/2$. Thus

$$U^{-1}a_\alpha U=\sum_{j=0}^{\infty}a_\alpha^{(j)},\quad U^{-1}A_\alpha U=\sum_{j=0}^{\infty}A_\alpha^{(j)},$$
$$U^{-1}\psi(X)U=\sum_{j=0}^{\infty}\psi^{(j)}(X),\quad U^{-1}\psi(x)U=\sum_{j=0}^{\infty}\psi^{(j)}(x) \quad (63)$$

where the jth term in each series (63) is obtained by substitution of $\epsilon=\pi/2$ in the previously obtained jth term [cf. (51)]. One finds thus

$$a_\alpha^{(0)}=A_\alpha,\quad a_\alpha^{(1)}=0,\quad a_\alpha^{(2)}=-\sum_\beta C_{\alpha\beta}(\tfrac{1}{4}\pi a_\beta+\tfrac{1}{2}A_\beta);$$
$$A_\alpha^{(0)}=-a_\alpha,\quad A_\alpha^{(1)}=0,\quad A_\alpha^{(2)}=-\sum_\beta C_{\alpha\beta}(\tfrac{1}{4}\pi A_\beta+\tfrac{1}{2}a_\beta). \quad (64)$$

Similarly,

$$\psi^{(0)}(X)=\psi(X),$$
$$\psi^{(1)}(X)=\sum_\alpha\int dx\,\phi_\alpha(Xx)\psi^\dagger(x)(a_\alpha+A_\alpha),$$
$$\psi^{(2)}(X)=-\sum_{\alpha\beta}\int dx\,dX'\,\phi_\alpha^*(X'x)\phi_\beta(Xx)A_{\alpha\beta}\psi(X'),$$
$$\psi^{(3)}(X)=\sum_{\alpha\beta\gamma}\int dx\,dX'\,dx'\,\phi_\alpha^*(X'x)\phi_\beta(Xx)\phi_\gamma(X'x')\psi^\dagger(x')A_{\alpha\beta\gamma}$$
$$+\sum_{\alpha\beta}\int dx\,\phi_\alpha(Xx)\psi^\dagger(x)C_{\alpha\beta}[\tfrac{1}{2}a_\beta+(1-\tfrac{1}{4}\pi)A_\beta] \quad (65)$$

and

$$\psi^{(0)}(x)=\psi(x),$$
$$\psi^{(1)}(x)=\sum_\alpha\int dX\,\phi_\alpha(Xx)\psi^\dagger(X)(a_\alpha+A_\alpha),$$
$$\psi^{(2)}(x)=-\sum_{\alpha\beta}\int dX\,dx'\,\phi_\alpha^*(Xx')\phi_\beta(Xx)A_{\alpha\beta}\psi(x'),$$
$$\psi^{(3)}(x)=-\sum_{\alpha\beta\gamma}\int dX\,dX'\,dx'\,\phi_\alpha^*(Xx')\phi_\beta(Xx)\phi_\gamma(X'x')\psi^\dagger(X')A_{\alpha\beta\gamma}$$
$$-\sum_{\alpha\beta}\int dX\,\phi_\alpha(Xx)\psi^\dagger(X)C_{\alpha\beta}[\tfrac{1}{2}a_\beta+(1-\tfrac{1}{4}\pi)A_\beta] \quad (66)$$

with

$$A_{\alpha\beta}=\tfrac{1}{2}a_\alpha^\dagger a_\beta+\tfrac{1}{4}\pi a_\alpha^\dagger A_\beta+(1-\tfrac{1}{4}\pi)A_\alpha^\dagger a_\beta+\tfrac{1}{2}A_\alpha^\dagger A_\beta,$$
$$A_{\alpha\beta\gamma}=\tfrac{1}{6}a_\alpha^\dagger a_\gamma a_\beta+\tfrac{1}{3}a_\alpha^\dagger a_\gamma A_\beta+(\tfrac{5}{6}-\tfrac{1}{4}\pi)a_\alpha^\dagger A_\gamma a_\beta$$
$$+(\tfrac{1}{4}\pi-\tfrac{2}{3})a_\alpha^\dagger A_\gamma A_\beta+(\tfrac{1}{4}\pi-\tfrac{2}{3})A_\alpha^\dagger a_\gamma a_\beta+\tfrac{1}{3}A_\alpha^\dagger a_\gamma A_\beta$$
$$+(\tfrac{5}{6}-\tfrac{1}{4}\pi)A_\alpha^\dagger A_\gamma a_\beta+\tfrac{1}{6}A_\alpha^\dagger A_\gamma A_\beta. \quad (67)$$

4. TRANSFORMED HAMILTONIAN

According to (37), the Hamiltonian acting on states $|\psi)\in\mathcal{J}_{\rm phys}$ is given by $U^{-1}HU$ where H is the Hamiltonian (1) on states $|\psi\rangle\in\mathcal{S}$, and $\mathcal{S}$ is the Schrödinger state space. In $\mathcal{S}$, only proton and electron variables occur explicitly, and the existence and properties of bound atoms are not manifest in the algebra of observables. On the other hand, upon carrying out the unitary transformation $U^{-1}HU$, one obtains a transformed Hamiltonian containing not only proton and electron variables, but also the operators a_α and $a_\alpha^\dagger$ representing the bound atoms. The generalized Tani transformation automatically generates explicit terms in $U^{-1}HU$ representing all possible atomic scattering, ionization, and recombination, etc. processes which can occur. The general procedure for evaluation of $U^{-1}HU$ is to substitute the transforms (65)–(67) for the various ψ and $\psi^\dagger$ factors in (1), and then to put all operator products into normal order by use of the commutation relations (4), (8), and (9), or equivalently by application of Wick's theorem. Upon carrying out these operations one obtains a transformed Hamiltonian of the structure

$$U^{-1}HU=T_p+T_e+V_{pp}+V_{ee}+H_{pe}$$
$$+H_a+H_{aa}+H_{ap}+H_{ae}$$
$$+H(pe\leftarrow a)+H(a\leftarrow pe)$$
$$+H(ppee\leftarrow aa)+H(aa\leftarrow ppee)$$
$$+H(pea\leftarrow aa)+H(aa\leftarrow pea)$$
$$+H(ppe\leftarrow pa)+H(pa\leftarrow ppe)$$
$$+H(pee\leftarrow ea)+H(ea\leftarrow pee)+\cdots. \quad (68)$$

Our notation is motivated by the work of Stolt and Brittin,[1] who obtained a similar expression by a different method. The details of the derivation of (68) are described in Appendix B. Here we shall merely list the results and explain their physical interpretations.

First consider the terms T_p, T_e, V_{pp}, and V_{ee}. These terms are the same as the corresponding terms in the untransformed Hamiltonian (1), except that the physical interpretation is different, in that the ψ and $\psi^\dagger$ operators now refer only to *unbound* protons and electrons. These contributions are obtained from the terms $\psi^{(0)}=\psi$ and $\psi^{(0)\dagger}=\psi^\dagger$ in all factors in the transform of (1). The diagramatic representations of the corresponding terms in $U^{-1}HU$ are shown in Fig. 1. The same type of line can be used for both protons and electrons, since the distinction is clear from the labels (X for proton, x for electron). The fact that the labels on the outgoing lines are the same as those on the incoming lines is a result of the locality of the Coulomb interactions, kinetic energy operators, and external potentials (if present) in Schrödinger representation. We use a standard convention: Incoming lines always approach from the right, and stand for annihilation operators on the right in the corresponding term in $U^{-1}HU$; outgoing lines always leave toward the left, and stand for creation operators on the left in the same term; two-line vertices stand

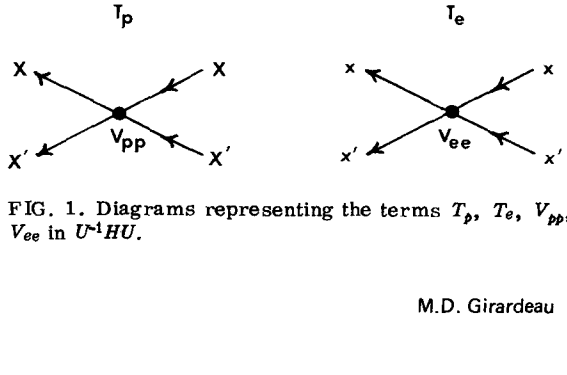


FIG. 1. Diagrams representing the terms T_p, T_e, V_{pp}, and V_{ee} in $U^{-1}HU$.

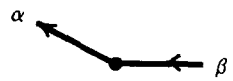


FIG. 2. Diagramatic representation of the single-atom Hamiltonian H_a.

for $T(X)$ or $T(x)$ (kinetic energy operator plus any external potentials and/or magnetic fields, if present); four-line vertices stand for Coulomb interaction potentials $V(XX')$ and $V(xx')$; finally, free indices are to be integrated over, with a combinatorial factor $(2!)^{-1}$ for identical particles. If one were to Fourier transform $\psi(X)$ and $\psi(x)$, one would obtain the usual diagrams with lines labelled by momenta and spins. We have written down the well-known diagrams of Fig. 1 merely to establish notation and to motivate the construction of related but more complicated diagrams for the other terms in (68). All other terms represent processes involving bound atoms. For example, we shall see later that the term H_{pe} in $U^{-1}HU$ differs from the bare proton–electron Coulomb interaction Hamiltonian V_{pe} of Eq. (1), due to the influence of bound atoms (bound proton–electron pairs).

The "single atom Hamiltonian" H_a in (68) has the structure

$$H_a = \sum_{\alpha\beta} a_\alpha^\dagger (\alpha | H | \beta) a_\beta \tag{69}$$

where the single atom matrix elements $(\alpha | H | \beta)$ are, as suggested by the notation, matrix elements of the single atom Hamiltonian between single atom wavefunctions:

$$(\alpha | H | \beta) = \int \phi_\alpha^*(Xx)[T(X) + T(x) + V(Xx)]\phi_\beta(Xx)\, dXdx. \tag{70}$$

This contribution to $U^{-1}HU$ arises from the terms of structure $\psi^{(1)\dagger}(X)T(X)\psi^{(1)}(X)$, $\psi^{(1)\dagger}(x)T(x)\psi^{(1)}(x)$, and $\psi^{(1)\dagger}(X)\psi^{(0)\dagger}(x)V(Xx)\psi^{(0)}(x)\psi^{(1)}(X)$ upon substitution of (65)–(67) into the corresponding terms in (1) and reduction of all operator products to normal order. Terms such as $\psi^\dagger a^\dagger a \psi$, etc. also contribute to $U^{-1}HU$, but will be considered later (as contributions to H_{ap}, H_{ae}, etc.). If the ϕ_α are chosen to be single-atom energy eigenstates, i. e.,

$$[T(X) + T(x) + V(Xx)]\phi_\alpha(Xx) = \epsilon_\alpha \phi_\alpha(Xx), \tag{71}$$

then H_a becomes diagonal,

$$H_a = \sum_\alpha \epsilon_\alpha N_\alpha, \tag{72}$$

with $N_\alpha = a_\alpha^\dagger a_\alpha$ the occupation number operator for atoms in atomic bound state ϕ_α. The diagramatic representation of the more general expression (69) is shown in Fig. 2, in which thick lines stand for (bound) atoms and the vertex represents the single atom matrix element $(\alpha | H | \beta)$. In the special case (71), (72), a single atom state β propagates without decay into other atomic states α (and, as we shall see later, without decay into protons and electrons). On the other hand, even if the ϕ_α are taken to be free-atom energy eigenstates, they will not be energy eigenstates in the presence of an external field [which is included in the definitions of $T(X)$ and $T(x)$]. In such a case Fig. 2 represents real physical effects, namely atomic transitions induced by an external field. The related processes of field-induced ionization and recombination are described by $H(pe \leftarrow a)$ and $H(a \leftarrow pe)$, which are found to be

$$H(pe \leftarrow a) = \sum_\alpha \int dX\, dx \psi^\dagger(X)\psi^\dagger(x)(Xx | H | \alpha) a_\alpha,$$

$$H(a \leftarrow pe) = [H(pe \leftarrow a)]^\dagger = \sum_\alpha \int dX\, dx a_\alpha^\dagger (\alpha | H | Xx)\psi(x)\psi(X) \tag{73}$$

with[19]

$$(Xx | H | \alpha) = -H(Xx)\phi_\alpha(Xx) + \int dX'\, dx' \Delta(Xx, X'x') \times H(X'x')\phi_\alpha(X'x'),$$

$$(\alpha | H | Xx) = (Xx | H | \alpha)^*. \tag{74}$$

Here Δ is the "bound state kernel" defined by (B3) and $H(Xx)$ is the single-atom Hamiltonian

$$H(Xx) = T(X) + T(x) + V(Xx) \tag{75}$$

which also occurs in (70) and (71). In case the ϕ_α are energy eigenstates (71), the matrix elements $(Xx | H | \alpha)$ and $(\alpha | H | Xx)$ vanish identically:

$$(Xx | H | \alpha) = -\epsilon_\alpha \phi_\alpha(Xx) + \epsilon_\alpha \int dX'\, dx' \Delta(Xx, X'x')\phi_\alpha(X'x') = -\epsilon_\alpha \phi_\alpha(Xx) + \epsilon_\alpha \phi_\alpha(Xx) = 0, \tag{76}$$

since it follows from (B3) and orthonormality of the ϕ_α that

$$\int \Delta(Xx, X'x')\phi_\alpha(X'x')\, dX'\, dx' = \phi_\alpha(Xx). \tag{77}$$

The vanishing of $H(pe \leftarrow a)$ and $H(a \leftarrow pe)$ if the ϕ_α are single-atom energy eigenstates is an expression, together with (72), of the stability of the bound atomic states in the absence of external perturbations. On the other hand, if the ϕ_α are not energy eigenstates and hence not stationary states, then the matrix elements (74) will not in general vanish, and there will be a nonzero probability of spontaneous breakup (decay) of such nonstationary states, in which case (73) represents real physical processes. Similarly, if the ϕ_α are taken to be isolated-atom energy eigenstates, they will not be energy eigenstates in the presence of external fields [which should then be included in $T(X)$ and $T(x)$], in which case (73) represents field-induced ionization and recombination processes.[20] The diagramatic representation of such terms is given in Fig. 3.

The term H_{pe} was not included in the diagrams of Fig. 1 since the presence of bound atoms modifies the effective interaction of free protons and electrons, i. e., H_{pe} differs from the bare proton–electron interaction Hamiltonian V_{pe} of (1). In fact, one finds

$$H_{pe} = \int dX\, dx\, dX'\, dx' \psi^\dagger(X)\psi^\dagger(x)(Xx | H | X'x')\psi(x')\psi(X') \tag{78}$$

with

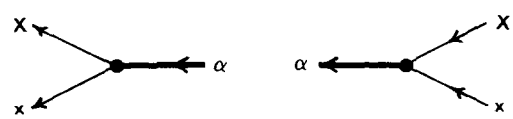


FIG. 3. Representation of single-atom breakup and recombination Hamiltonians.

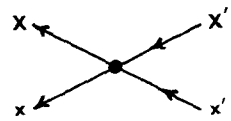


FIG. 4. Diagramatic representation of the effective interaction H_{pe} between unbound protons and electrons.

$$(Xx|H|X'x') = V(Xx)\delta(X-X')\delta(x-x') - H(Xx)\Delta(Xx,X'x') - [H(X'x')\Delta(X'x',Xx)]^* + \int dX''\,dx''\,\Delta(Xx,X''x'')H(X''x'')\Delta(X''x'',X'x'). \quad (79)$$

The diagramatic representation of (78) is shown in Fig. 4. The first term in (79) is the bare proton–electron Coulomb interaction which gives rise to the term V_{pe} of Eq. (1), whereas the remaining terms represent a nonlocal contribution to this interaction arising because of the possibility of binding of proton–electron pairs, the interaction within such bound pairs already being included in those terms in $U^{-1}HU$ which involve a_α and/or $a_\alpha^\dagger$ operators. In contrast, the proton–proton and electron–electron Hamiltonians V_{pp} and V_{ee} in (68), depicted in Fig. 1, are just the bare proton–proton and electron–electron Coulomb interactions (1). The range of the nonlocal term in (79) is of the order of the Bohr radius a_0, as can be seen[21] from the definition (B3) of the bound state kernel. If the ϕ_α are taken to be single-atom energy eigenstates (71), then (79) reduces to

$$(Xx|H|X'x') = V(Xx)\delta(X-X')\delta(x-x') - \sum_\alpha \epsilon_\alpha \phi_\alpha(Xx)\phi_\alpha^*(X'x'). \quad (80)$$

The sum subtracted from the bare Coulomb potential in (80) is just the portion of the spectral representation of $H(Xx)$ associated with all of its bound states. As a result, one expects that the modified potential (80) will not have any bound states, whereas it will be equivalent to the bare Coulomb potential when acting on a continuum (unbound) energy eigenstate. These properties are verfied in Appendix C, where it is shown that if the ϕ_α consist of all the bound energy eigenstates, then $(T_p + T_e + H_{pe})$ is positive semidefinite and is equivalent to $(T_p + T_e + V_{pe})$ when acting on any state orthogonal to all the bound states ϕ_α. This substantiates our interpretation of H_{pe} as the interaction between *unbound* protons and electrons, the interaction between *bound* protons and electrons already being included in those terms in the Hamiltonian involving the a_α and $a_\alpha^\dagger$.

We next consider the two atom interaction Hamiltonian H_{aa}, which is defined as the sum of all the terms in $U^{-1}HU$ of the structure $a^\dagger a^\dagger aa$. Thus

$$H_{aa} = \tfrac{1}{2}\sum_{\alpha\beta\gamma\delta} a_\alpha^\dagger a_\beta^\dagger (\alpha\beta|H|\gamma\delta) a_\delta a_\gamma \quad (81)$$

where the atom–atom interaction matrix element $(\alpha\beta|H|\gamma\delta)$ is the sum of a number of contributions of different physical origins. In the first place, there is a direct Coulomb interaction contribution $(\alpha\beta|H|\gamma\delta)_{\text{Coul}}$ arising from terms of the form $\psi^{(1)\dagger}\psi^{(1)\dagger}\psi^{(1)}\psi^{(1)}$ in the transforms of V_{pp}, V_{pe}, and V_{ee} upon complete contraction[22] so as to remove the factors $\psi\psi\psi^\dagger\psi^\dagger$, leaving only factors $a^\dagger a^\dagger aa$. The corresponding matrix element is found to be

$$(\alpha\beta|H|\gamma\delta)_{\text{Coul}} = \int \phi_\alpha^*(Xx)\phi_\beta^*(X'x')[V(XX') + V(xx') + V(Xx') + V(X'x')]\phi_\gamma(Xx)\phi_\delta(X'x')\,dX\,dx\,dX'\,dx' \quad (82)$$

as one expects for the matrix element of interatomic Coulomb interaction. Note that *internal* Coulomb interaction $V(Xx)$ and $V(X'x')$ are not contained in (82), being already taken into account in the single atom Hamiltonian[23] H_a. There is also a matrix element representing coupling between interatomic Coulomb interactions and interatomic electron[24] exchange, given by

$$(\alpha\beta|H|\gamma\delta)_{\text{Coul-ex}} = -\int \phi_\alpha^*(Xx)\phi_\beta^*(X'x')\{V(XX') + V(xx') + \tfrac{1}{2}[V(Xx) + V(X'x') + V(X'x) + V(Xx')]\} \times \phi_\gamma(Xx')\phi_\delta(X'x')\,dX\,dx\,dX'\,dx'. \quad (83)$$

The first two terms in the curly brackets also arise from terms of the structure $\psi^{(1)\dagger}\psi^{(1)\dagger}V\psi^{(1)}\psi^{(1)}$, there being two ways of completely contracting such a product if all four operators refer to protons or all four to electrons; the remaining terms come from more complicated terms in the Tani transform, as discussed in Appendix B. The first two terms in (83) are recognizable as matrix elements of the operator IV where I is the interatomic electron exchange operator and V the proton–proton and electron–electron interatomic interactions. The third and fourth terms in (83) are matrix elements of the operator IV whereas the fifth and sixth terms are matrix elements of VI, where I is again the interatomic electron exchange operator and V is the proton–electron interatomic interactions. Finally, there is an additional contribution to (81) arising from coupling between *intra*-atomic energy and *inter*atomic electron exchange, given by

$$(\alpha\beta|H|\gamma\delta)_{\text{intra-ex}} = \tfrac{1}{6}\int \phi_\alpha^*(Xx)\phi_\beta^*(X'x')[H(Xx) + H(X'x') + H(Xx') + H(X'x)]\phi_\gamma(Xx')\phi_\delta(X'x)\,dX\,dx\,dX'\,dx'. \quad (84)$$

Here $H(Xx)$ is the single atom Hamiltonian (75). The first two terms in (84) are matrix elements of HI whereas the third and fourth are matrix elements of IH, where H is the sum of $H(Xx)$ and $H(X'x')$, and I is the interatomic electron exchange operator. The matrix elements (83) and (84) are generalizations of the "exchange integrals" which play an important role in the theory of ferromagnetism and quantum chemistry. The implications of the many-electron generalizations of such exchange matrix elements for the theory of ferromagnetism have been discussed elsewhere.[7] In contrast with the direct Coulomb matrix elements (82), the exchange matrix elements (83) and (84) vanish identically if the atoms are nonoverlapping, i.e., if ϕ_α and ϕ_β are

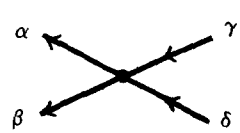


FIG. 5. Diagramatic representation of the two-atom interaction Hamiltonian H_{aa}.

FIG. 6. Representation of processes in which atoms collide with protons or electrons without atomic breakup.

localized in disjoint regions R_α and R_β; compare with the physical discussion of the exchange kernels (7). The diagramatic representation of H_{aa} is shown in Fig. 5.

The terms H_{ap} and H_{ae} in (68) represent processes in which single atoms collide with single protons or electrons without atomic breakup, and have the general structures

$$H_{ap} = \int dX\, dX' a_\alpha^\dagger \psi^\dagger(X)(\alpha X|H|\beta X')\psi(X')a_\beta,$$
$$H_{ae} = \sum_{\alpha\beta} \int dx\, dx' a_\alpha^\dagger \psi^\dagger(x)(\alpha x|H|\beta x')\psi(x')a_\beta. \tag{85}$$

The diagramatic representations are shown in Fig. 6. As in the case of H_{aa}, the matrix elements in (85) have both direct Coulomb and exchange contributions. One finds for the direct Coulomb matrix elements

$$(\alpha X|H|\beta X')_{\text{Coul}} = \delta(X-X')\int \phi_\alpha^*(Yy)[V(XY)+V(Xy)]\phi_\beta(Yy)\,dY\,dy,$$
$$(\alpha x|H|\beta x')_{\text{Coul}} = \delta(x-x')\int \phi_\alpha^*(Yy)[V(xy)+V(Yx)]\phi_\beta(Yy)\,dY\,dy. \tag{86}$$

The delta function prefactors are an expression of the locality of direct Coulomb interactions. As one expects physically, the expressions (86) are matrix elements, between single atom wavefunctions, of the Coulomb interactions between the incident proton or electron and the two particles (proton and electron) in the atom. Since the incident particle can exchange with one of the particles in the atom, there are also exchange contributions, arising from different parings of the contracted operators. One finds for the Coulomb-exchange coupling matrix elements

$$(\alpha X|H|\beta X')_{\text{Coul-ex}} = -\int \phi_\alpha^*(X'x)[V(XX')+V(Xx)+V(X'x)]\phi_\beta(Xx)\,dx,$$
$$(\alpha x|H|\beta x')_{\text{Coul-ex}} = -\int \phi_\alpha^*(Xx')[V(xx')+V(Xx)+V(Xx')]\phi_\beta(Xx)\,dX. \tag{87}$$

$$(XxX'|H|\alpha X'') = -\delta(X'-X'')[V(XX')+V(X'x)]\phi_\alpha(Xx)$$
$$+\delta(X'-X'')\int \Delta(Xx,Yy)[V(X'y)+V(X'Y)]\phi_\alpha(Yy)\,dY\,dy$$
$$-\tfrac{1}{4}\pi\int\{[T(X'')\Delta(X''x'',Xx)]^*$$
$$-\Delta(Xx,X''x'')T(X')\}\phi_\alpha(X'x'')\,dx''$$
$$-\int\{\Delta(Xx,X''x'')H(X'x'')+\tfrac{1}{2}H(Xx)\Delta(Xx,X''x'')\}\phi_\alpha(X'x'')\,d$$
$$-\int\Delta(Xx,X''x'')[\tfrac{1}{2}V(X'x)+\tfrac{1}{2}V(XX')+V(X'X'')$$
$$+V(X''x'')]\phi_\alpha(X'x'')\,dx'',$$

There are also terms representing coupling of kinetic energy and exchange, which are found to be

$$(\alpha X|H|\beta X')_{\text{kin-ex}} = -\int \phi_\alpha^*(X'x)T(x)\phi_\beta(Xx)\,dx$$
$$-\tfrac{1}{2}\int\{\phi_\alpha^*(X'x)T(X)+[T(X')\phi_\alpha(X'x)]^*\}\phi_\beta(Xx)\,dx,$$
$$(\alpha x|H|\beta x')_{\text{kin-ex}} = -\int \phi_\alpha^*(Xx')T(X)\phi_\beta(Xx)\,dX$$
$$-\tfrac{1}{2}\int\{\phi_\alpha^*(Xx')T(x)+[T(x')\phi_\alpha(Xx')]^*\}\phi_\beta(Xx)\,dX. \tag{88}$$

In contrast with the direct Coulomb contributions, the exchange contributions (87) and (88) are nonlocal since they do not contain prefactors $\delta(X-X')$ and $\delta(x-x')$. It is not difficult to see that the range of the nonlocality is small, of order a_0, the Bohr radius of hydrogen. It is also easy to see that these exchange matrix elements vanish if the atomic wavefunctions are localized in regions R_α and R_β and either X is outside R_β or X' is outside R_α (resp. x outside R_β or x' outside R_α), or if R_α and R_β are nonoverlapping.

The remaining terms in (68) are reaction Hamiltonians, representing processes in which hydrogen atoms break up or recombine (ionization and recombination). The simplest such terms, $H(pe \leftarrow a)$ and $H(a \leftarrow pe)$, have already been discussed. The terms $H(ppe \leftarrow pa)$, $H(pee \leftarrow ea)$, and their Hermitian conjugates represent processes in which a single atom collides with a proton or electron and is thereby ionized, together with the inverse recombination processes, and are of the form

$$H(ppe \leftarrow p\alpha) = \sum_\alpha \int dX\, dx\, dX'\, dX''\, \psi^\dagger(X)\psi^\dagger(x)\psi^\dagger(X')(XxX''|H|\alpha X'')\psi(X'')a_\alpha,$$
$$H(pee \leftarrow e\alpha) = \sum_\alpha \int dX\, dx\, dx'\, dx''\, \psi^\dagger(X)\psi^\dagger(x)\psi^\dagger(x')(Xxx'|H|\alpha x'')\psi(x'')a_\alpha,$$
$$H(p\alpha \leftarrow ppe) = [H(ppe \leftarrow p\alpha)]^\dagger, \quad H(e\alpha \leftarrow pee) = [H(pee \leftarrow e\alpha)]^\dagger. \tag{89}$$

The corresponding diagrams are shown in Fig. 7; the diagrams for the inverse recombination processes differ only by left–right inversion. Although such recombination processes are three-body collision terms, they should be included for consistency (otherwise the Hamiltonian would not be Hermitian). The matrix elements in (89) are found to be

$$
\begin{aligned}
&(Xxx'|H|\alpha x'')\\
&\quad = -\delta(x'-x'')[V(Xx')+V(xx')]\phi_\alpha(Xx)\\
&\qquad + \delta(x'-x'')\int \Delta(Xx,Yy)[V(Yx')+V(x'y)]\phi_\alpha(Yy)\,dY\,dy\\
&\qquad - \tfrac{1}{4\pi}\int \{[T(x'')\Delta(X''x'',Xx)]^* - \Delta(Xx,X''x'')T(x')\}\phi_\alpha(X''x')\,dX''\\
&\qquad - \int \{\Delta(Xx,X''x'')H(X''x') + \tfrac{1}{2}H(Xx)\Delta(Xx,X''x'')\}\phi_\alpha(X''x')\,dX''\\
&\qquad - \int \Delta(Xx,X''x'')[\tfrac{1}{2}V(Xx')+\tfrac{1}{2}V(xx')+V(x'x'')+V(X''x'')]\phi_\alpha(X''x')\,dX''. \qquad (90)
\end{aligned}
$$

The first term in each of these matrix elements clearly represents the effect of the direct Coulomb interaction of the incident proton or electron with the two particles (proton and electron) in the atom; the minus signs are merely phase factors and have no physical significance.[25] The second term in each matrix element represents indirect processes in which the Coulomb interaction with the incident particle induces a bound state–bound state transition of the atom, followed by decay[26] of the resultant bound state into its constituents. All of the remaining terms in (90) represent nonlocal exchange with nonlocality of range $\sim a_0$ as a function of $(\mathbf{R}'-\mathbf{R}'')$ or $(\mathbf{r}'-\mathbf{r}'')$; the argument for the range is similar to that previously given[21] for the simpler matrix element (79).

Finally, we consider the terms in (68) representing binary atomic collisions with resultant ionization of one or both atoms, and the inverse recombination terms. The terms representing ionization and recombination of both atoms have the form

$$
\begin{aligned}
&H(ppee \leftarrow aa)\\
&\quad = \tfrac{1}{2}\sum_{\alpha\beta}\int dX\,dx\,dX'\,dx'\psi^\dagger(X)\psi^\dagger(x)\psi^\dagger(X')\psi^\dagger(x')\\
&\qquad \times (XxX'x'|H|\alpha\beta)a_\beta a_\alpha,\\
&H(aa \leftarrow ppee) = [H(ppee \leftarrow aa)]^\dagger, \qquad (91)
\end{aligned}
$$

and are represented by the diagrams of Fig. 8. Although the diagram on the right, and the corresponding Hamiltonian $H(aa \leftarrow ppee)$, represent four-body collisions, they should be included for consistency, as in the case of the analogous three-body collision terms $H(pa \leftarrow ppe)$ and $H(ea \leftarrow pee)$ discussed previously. The matrix elements in (91) are found to be

$$
\begin{aligned}
&(XxX'x'|H|\alpha\beta)\\
&\quad = [V(XX')+V(xx')+V(Xx')+V(X'x)]\phi_\alpha(Xx)\phi_\beta(X'x') \qquad (92)
\end{aligned}
$$

and clearly represent direct[27] Coulomb interactions between the two atoms. Terms representing *partial* ionization as a result of binary atomic collisions, and the inverse recombination processes, have the form

$$
H(pea \leftarrow aa) = \sum_{\alpha\beta\gamma}\int dX\,dx a_\alpha^\dagger \psi^\dagger(X)\psi^\dagger(x)(\alpha Xx|H|\beta\gamma)a_\gamma a_\beta,
$$

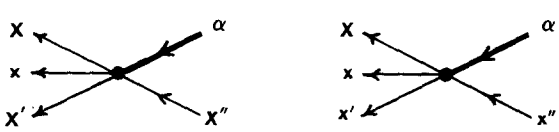


FIG. 7. Diagrams representing atomic ionization due to collision with protons or electrons.

$$
H(aa \leftarrow pea) = [H(pea \leftarrow aa)]^\dagger, \qquad (93)
$$

and are represented by the diagrams of Fig. 9. The matrix element in (93) is a sum of three types of contributions, which are similar to (82)–(84). The direct Coulomb contribution is

$$
\begin{aligned}
&(\alpha Xx|H|\beta\gamma)_{\text{Coul}}\\
&\quad = -\int \phi_\alpha^*(X'x')[V(XX')+V(xx')+V(Xx')+V(X'x)]\\
&\qquad \times \phi_\beta(Xx)\phi_\gamma(X'x')\,dX'\,dx', \qquad (94)
\end{aligned}
$$

whereas the interatomic Coulomb-interatomic exchange coupling contribution is

$$
\begin{aligned}
&(\alpha Xx|H|\beta\gamma)_{\text{Coul-ex}}\\
&\quad = \int \phi_\alpha^*(X'x')\{V(XX')+V(xx')+\tfrac{1}{2}[V(Xx)+V(X'x')\\
&\qquad + V(X'x)+V(Xx')]\}\phi_\beta(Xx')\phi_\gamma(X'x)\,dX'\,dx' \qquad (95)
\end{aligned}
$$

and the intra-atomic energy–interatomic exchange coupling contribution is

$$
\begin{aligned}
&(\alpha Xx|H|\beta\gamma)_{\text{intra-ex}}\\
&\quad = \int \phi_\alpha^*(X'x')[\tfrac{1}{2}H(X'x')+\tfrac{1}{6}H(Xx)]\phi_\beta(Xx')\phi_\gamma(X'x)\,dX'\,dx'. \qquad (96)
\end{aligned}
$$

As in the cases of the previously discussed exchange matrix elements, the sign difference between the direct Coulomb and exchange contributions is a real physical effect; on the other hand, the overall phase is not physically observable, since (93) contributes to observable quantities only in even orders.

This completes the enumeration of the various contributions to (68). The omitted terms "..." are of two kinds. In the first place, all multiple collision terms, i. e., all terms in which *both* the number of incoming *and* the number of outgoing particles is ≥ 3, have been omitted. In the second place, higher order contributions to some of the matrix elements of binary collision terms have been omitted. We shall conclude this section by discussing the nature of such corrections to matrix elements.

In the first place, the matrix elements (70) and (74) of the single-atom scattering Hamiltonian (69) and the single-atom ionization and recombination Hamiltonian

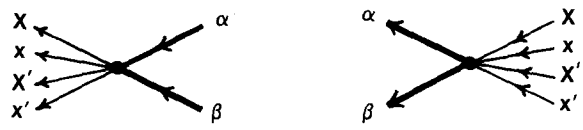


FIG. 8. Diagrams representing complete atomic breakup due to binary atomic collisions, and the inverse recombination processes.

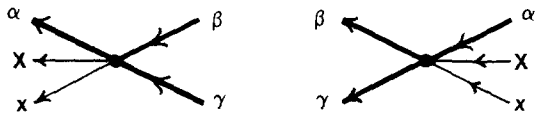


FIG. 9. Diagrams representing partial atomic breakup due to binary atomic collisions, and the inverse recombination processes.

(73) are exact, i.e., there are no higher order corrections. Although the analysis carried out in Appendix B only verifies this through fourth order, a different method of evaluation discussed elsewhere[28,29] enables one to prove that higher order corrections to these matrix elements do not occur. Similarly, it can be shown[28,29] that, as previously noted, the proton–proton and electron–electron interaction Hamiltonians V_{pp} and V_{ee} in (68) have exactly the forms (1), i.e., there are no corrections[30] due to the influence of bound atoms. In addition, the modified proton–electron interaction matrix element (79) is exact.[28,29]

The other matrix elements in (68) have only been evaluated to the lowest few orders. Specifically, there are omitted fourth-order (in wavefunctions) corrections to $(\alpha X|H|\beta X')$ and $(\alpha x|H|\beta x')$ [Eq. (85)], to $(XxX'|H|\alpha X'')$ and $(Xxx'|H|\alpha x'')$ [Eq. (89)], and to $(XxX'x'|H|\alpha\beta)$ [Eq. (91)]. The matrix element $(\alpha Xx|H|\beta\gamma)$ in (93) has been exhibited through third order, and the analysis in Appendix B indicates that there are no fourth order corrections. However, there may be corrections in fifth and higher orders. In addition, there may be corrections to the interatomic interaction matrix elements $(\alpha\beta|H|\gamma\delta)$ [Eq. (81)] in sixth and higher orders.

5. SUBSIDIARY CONDITION

The subsidiary condition (11) on state vectors $|\psi\rangle$ in the subspace $\mathcal{G}_0$ isomorphic to the Schrödinger state space $\mathcal{S}$ is equivalent to the transformed subsidiary condition (36) acting on states $|\psi) = U^{-1}|\psi\rangle$; the subspace $\mathcal{G}_{\rm phys} = U^{-1}\mathcal{G}_0$ [Eq. (34)] consists of those states $|\psi)$ satisfying (36). More generally, if $\mathcal{G}_j$ is the subspace of $\mathcal{G}$ consisting of those states $|\psi\rangle$ satisfying

$$N_a|\psi\rangle = j|\psi\rangle, \tag{97}$$

then one can easily prove that the $\mathcal{G}_j$ are disjoint and that their union (j running from 0 to ∞) is the entire ideal state space $\mathcal{G}$. If the subsidiary condition (11) is dropped, then, since H does not contain a_α and $a^\dagger_\alpha$ operators, it will have the same eigenvalue spectrum[31] as it does on $\mathcal{G}_0$, but a spurious infinite degeneracy of every energy level. Similarly, in addition to $\mathcal{G}_{\rm phys} = U^{-1}\mathcal{G}_0$ [Eq. (34)], there is a whole infinite sequence of disjoint subspaces $U^{-1}\mathcal{G}_j$, whose union is $\mathcal{G}$; each $U^{-1}\mathcal{G}_j$ for $j \geq 1$ can be regarded as a "copy" of $\mathcal{G}_{\rm phys}$, with all states of $\mathcal{G}_{\rm phys}$ imaged with the same energies.

It follows from this that so long as one is interested in energies but not in the density of states, one can ignore the subsidiary condition with impurity. In fact, since the same argument[31] implies to any physical observable, not merely the energy, one can safely ignore the subsidiary condition in evaluating the eigenvalue or expectation value of any observable. On the other hand, the subsidiary condition must be properly incorporated in the evaluation of quantities depending on the density of states, for example the partition function[32]

$$Z(\beta) = \mathrm{Tr}_0 \exp(-\beta U^{-1}HU) \tag{98}$$

where the subscript zero on Tr means that the trace must be restricted to a basis spanning only the physical subspace $\mathcal{G}_{\rm phys} = U^{-1}\mathcal{G}_0$. However, if one defines a generalized partition function

$$\begin{aligned}\Xi(\beta,\xi) &= \mathrm{Tr}[\xi^{U^{-1}N_aU}\exp(-\beta U^{-1}HU)\\ &= \mathrm{Tr}\exp[-\beta U^{-1}(H-\lambda N_a)U]\end{aligned} \tag{99}$$

where

$$\xi = \exp(\beta\lambda) \tag{100}$$

and Tr denotes the unrestricted trace over the whole state space $\mathcal{G}$, then clearly $Z(\beta)$ is the coefficient of the constant term in an expansion of $\Xi(\beta,\xi)$ in powers of ξ.

In case one wishes to project out the component of any state $|\psi)$ which lies in the physical subspace $\mathcal{G}_{\rm phys}$, this can be done by multiplication by the projection operator

$$P_{\rm phys} = U^{-1}P_0U \tag{101}$$

where P_0 is the projector onto $\mathcal{G}_0$. Recalling that the eigenvalues of N_a are integers, one can easily write down a formal expression for P_0:

$$P_0 = (2\pi)^{-1}\int_0^{2\pi} d\vartheta \exp(i\vartheta N_a). \tag{102}$$

This expression is not very useful for actual calculations, since power series expansion of the exponential is term-by-term incompatible with the periodicity which is essential for the validity of (102). What is needed is an expansion of P_0 in terms of normally ordered products of annihilation and creation operators. Such an expansion is easily constructed and is in fact well known in other contexts. The desired expression is

$$P_0 = 1 + \sum_{j=1}^{\infty}[(-1)^j/j!]\sum_{\alpha_1\cdots\alpha_j} a^\dagger_{\alpha_1}\cdots a^\dagger_{\alpha_j}a_{\alpha_j}\cdots a_{\alpha_1}. \tag{103}$$

Using the identity

$$\sum_{\alpha_1\cdots\alpha_j} a^\dagger_{\alpha_1}\cdots a^\dagger_{\alpha_j}a_{\alpha_j}\cdots a_{\alpha_1} = N_a(N_a-1)\cdots(N_a-j+1), \tag{104}$$

one can verify that (103) has the desired projection properties. Low-order approximations to the transformed projector (101) can be constructed using the transformation (64):

$$U^{-1}a_\alpha U = A_\alpha - \sum_\beta C_{\alpha\beta}(\tfrac{1}{4}\pi a_\beta + \tfrac{1}{2}A_\beta) + \cdots. \tag{105}$$

The projector (101) can be used to construct a "projected Hamiltonian"

$$\mathcal{H} = U^{-1}HP_0U = (U^{-1}HU)P_{\rm phys} \tag{106}$$

which has the same eigenvalues and eigenstates as the Hamiltonian (68) on the physical subspace $\mathcal{G}_{\rm phys}$, but which annihilates states orthogonal to $\mathcal{G}_{\rm phys}$ and hence no longer has spurious degeneracies. Such projected Hamiltonians have been discussed previously[3] in a slightly different context. The evaluation of (106) will be discussed elsewhere,[29] where it will be shown that $\mathcal{H}$ differs from (68) only in having additional contribu-

tions to some (but by no means all) of its matrix elements.

6. DISCUSSION

A transformation has been developed which allows states and observables of a nonrelativistic system of composite particles and their constituents to be expressed in terms of elementary Bose or Fermi operators for the bound composites plus the usual elementary field operators for the unbound constituents. In this new representation all possible scattering and reaction channels of the composites and their constituents are exhibited simultaneously. The transformed Hamiltonian is evaluated explicitly for the case of atomic hydrogen through binary collision terms, including the terms in which the number of outgoing or incoming particles (but not both) is greater than two. The matrix elements have qualitative properties expected on the basis of physical arguments. Similar representations have been developed previously; we shall conclude by comparing and contrasting our results with those.

The representation of Brittin and Stolt[1] is qualitatively similar, although the derivation is very different from ours; however, the matrix elements obtained differ in detail. Although their approach is in principle exact, they were forced (as were we) to make approximations in obtaining an explicit expression for the transformed Hamiltonian; the following remarks apply to this approximate expression only. The terms in the Brittin–Stolt Hamiltonian to which we refer here are given on pp. 76–79 of their Boulder Lectures.[1] In the first place, the Brittin–Stolt terms $\tilde{T}_a$, $\tilde{T}_p$, $\tilde{T}_e$, $\tilde{V}_{ee}$, and $\tilde{V}_{pp}$ correspond exactly to (and agree with) our single-atom Hamiltonian H_a [Eq. (69)], the single-proton and single-electron Hamiltonians T_p and T_e, and the proton–proton and electron–electron interaction Hamiltonians V_{pp} and V_{ee}; such exact correspondence is to be expected for such simple and "obvious" terms. On the other hand, their single-atom dissociation term $\tilde{V}(ep \leftarrow a)$ differs from ours $H(pe \leftarrow a)$ [Eq. (73)] not only in an irrelevant sign difference, but also in that there is no term analogous to the term involving Δ in (74). As a result and in contrast with (76), their matrix element fails to vanish in the case of an energy eigenstate (stationary state) ϕ_α, a case in which dissociation does not in fact occur physically. Similarly, their proton–electron interaction Hamiltonian $\tilde{V}_{ep}$ corresponds to only the first term in our effective proton–electron interaction matrix element (79), and hence fails to account for the effect on the unbound proton–unbound electron interaction of the fact that the bound proton–bound electron interaction is already included elsewhere in the Hamiltonian. The direct interatomic Coulomb interaction term $\tilde{V}_{aa}$ is equivalent to the direct Coulomb term (82) in H_{aa} [Eq. (81)], but the exchange contributions $\tilde{E}_{aa}$ appear to differ from (83) and (84). The terms $\tilde{V}_{pa}$ and $\tilde{V}_{ea}$ are equivalent to the direct Coulomb contributions (86) to H_{ap} and H_{ae} [Eq. (85)], and the exchange contributions $\tilde{E}_{pa}$ and $\tilde{E}_{ea}$ also appear to be equivalent to (87) and (88). On the other hand, the corresponding dissociation matrix elements $\tilde{V}(epp \leftarrow pa)$ and $\tilde{V}(eep \leftarrow ea)$ are equivalent to only the first term in the expressions (90) for the matrix elements in $H(ppe \leftarrow pa)$ and $H(pee \leftarrow ea)$ [Eq. (89)]. The direct Coulomb contribution to $\tilde{V}(epa \leftarrow aa)$ is equivalent to the corresponding term (94) in $H(pea \leftarrow aa)$ [Eq. (93)], but the exchange contributions appear to differ from (95) and (96). Finally, in the approximation in which the Brittin–Stolt Hamiltonian is evaluated, there are no terms corresponding to our two-atom total dissociation and recombination terms (91) and (92).

The representation of Sakakura[2] is also superficially similar to ours, but the matrix elements differ more in detail from those in (68) than do those of the Brittin–Stolt representation. The following comments refer to the terms exhibited on pp. 504, 505 of Sakakura's Boulder Lectures.[2] Sakakura's first term is equivalent to the sum of our terms T_p, T_e, V_{pp}, V_{ee}, and V_{pe}; as in the case of the Brittin–Stolt representation, there is no analog of the terms in (79) which effectively subtract the bound proton–bound electron interaction from the bare proton–electron interaction. Sakakura's second term is equivalent to our H_a for the case that the ϕ_α are taken to be single-atom energy eigenstates. In the same case, his third term is equivalent to only the first term of (74); as in the case of the Brittin–Stolt represention, omission of the remainder of (74) leads to an unphysical instability of bound atomic energy eigenstates. Sakakura's fourth and fifth terms correspond to our H_{ap} and H_{ae}, and the direct Coulomb contributions agree; however, Sakakura's exchange contributions differ both from (87), (88), and from those of Brittin and Stolt. His sixth term corresponds to our H_{aa}, and the direct Coulomb contribution agrees with (82); again, his exchange terms differ both from (83), (84) and from those of Brittin and Stolt. In the approximation in which Sakakura evaluated his Hamiltonian, no terms corresponding to $H(ppee \leftarrow aa)$, $H(pea \leftarrow aa)$, $H(ppe \leftarrow pa)$, $H(pee \leftarrow ea)$, and their Hermitian conjugates appear.

A more recent formulation of Brittin and Sakakura[5] was not carried to the point of explicit evaluation of the interaction parts of the Hamiltonian, so we cannot make a comparison at this time.

A previous conceptually more complicated approach[3] by the present author also led to a similar Hamiltonian, but again there are differences in the detailed matrix elements. Precise comparison is made difficult because terms in the matrix elements vanishing by virtue of the subsidiary condition employed were dropped in the course of the derivation. We shall therefore not attempt such a comparison here. However, there are some indications[33] that the representation obtained there is closely related to, but no completely equivalent with, the one of the present paper.

A composite particle second quantization formalism adapted to the theory of nuclear reactions has been given recently by Nishigori.[34] Many approximations (including neglect of all exchange terms) were made in the course of the derivation, so a detailed comparison is not possible. However, one can note, for example, that Nishigori's single-nucleus breakup matrix element[35] omits the second term of (74) necessary for stability of bound energy eigenstates, and in addition

omits the kinetic energy contribution to even the first term of (74).

The formalism of Gilbert[28] is based on the same Tani transformation as ours but uses a different method of evaluation of the transformed Hamiltonian. It therefore necessarily leads to the same transformed Hamiltonian. The details are discussed elsewhere.[28,29]

In addition to these first-principles approaches, there have been many previous approaches which must be classed as semiphenomenological, since they do not start from any generally accepted complete quantum-mechanical representation of states and observables in terms of the constituents of the composites. It would be inappropriate to attempt a comparison with such approaches. In addition, there have been a few recent formulations[36,37] which, although first-principles approaches, employ composite particle annihilation and creation operators similar to our A_α and $A_\alpha^\dagger$ and make no attempt to introduce operators analogous to our a_α and $a_\alpha^\dagger$. A detailed comparison of our formalism with such approaches does not appear to be possible.

ACKNOWLEDGMENTS

I am greatly indebted to Wesley Brittin and Arthur Sakakura for providing the initial stimulus which culminated in this work. The interest of Manuel Berrondo and a resultant three-week visit at the Instituto de Fisica of the Universidad Nacional Autonoma de México made an important contribution to the development. The willingness of my students to discuss earlier incomplete versions of the formalism was also quite helpful. I am particularly indebted to my students D. Gilbert and R. Munro for such discussions and helpful suggestions.

APPENDIX A: BEHAVIOR OF TANI TRANSFORM IN THE LIMIT OF ZERO OVERLAP

The proof presented here is closely related to previous proofs by my students Munro[38] and Gilbert.[39] Suppose that the atoms $\alpha_1 \cdots \alpha_n$ are localized in mutually disjoint regions, in the sense that for each j, $\phi_{\alpha_j}(Xx)$ is only nonzero if both $\mathbf{R} \in \mathcal{R}_j$ and $\mathbf{r} \in \mathcal{R}_j$, where the regions $\mathcal{R}_j$ are mutually disjoint. Suppose furthermore that only nonoverlapping atoms are included in the Tani transformation (14), in the sense that α is only summed over a discrete set of α_j including $\alpha_1 \cdots \alpha_n$ (but perhaps including the other α_j as well) which refer to mutually disjoint regions in the above sense. Then as previously noted, the exchange kernels $K_{\alpha\beta}$ [Eq. (7)] and hence the operators $C_{\alpha\beta}$ [Eq. (6)] vanish for $\alpha \neq \beta$, with the understanding that α and β are both members of the set $\alpha_1, \alpha_2, \cdots$. It then follows with (6), (7), (14), and (22)–(24) that for all integers $j \geq 1$

$$\begin{aligned} &[A_\alpha^\dagger, F]_{2j} = (-1)^j A_\alpha^\dagger + D_\alpha^{(2j)}, \\ &[A_\alpha^\dagger, F]_{2j-1} = (-1)^{j-1} a_\alpha^\dagger + D_\alpha^{(2j-1)} \end{aligned} \tag{A1}$$

where the $D_\alpha^{(j)}$ satisfy

$$\begin{aligned} &D_\alpha^{(j)} |0\rangle = 0, \\ &[D_\alpha^{(j)}, a_\beta] = [D_\alpha^{(j)}, a_\beta^\dagger] = [D_\alpha^{(j)}, A_\beta] \\ &\quad = [D_\alpha^{(j)}, A_\beta^\dagger] = [D_\alpha^{(j)}, D_\beta^{(k)}] = 0, \quad \alpha \neq \beta. \end{aligned} \tag{A2}$$

Then by (22)

$$UA_\alpha^\dagger U^{-1} = a_\alpha^\dagger + D_\alpha \tag{A3}$$

where

$$\begin{aligned} &D_\alpha |0\rangle = 0, \\ &[D_\alpha, a_\beta] = [D_\alpha, a_\beta^\dagger] = [D_\alpha, A_\beta] \\ &\quad = [D_\alpha, A_\beta^\dagger] = [D_\alpha, D_\beta] = 0, \quad \alpha \neq \beta. \end{aligned} \tag{A4}$$

Then

$$\begin{aligned} UA_{\alpha_1}^\dagger \cdots A_{\alpha_n}^\dagger |0\rangle &= UA_{\alpha_1}^\dagger U^{-1} \cdots UA_{\alpha_n}^\dagger U^{-1} |0\rangle \\ &= (a_{\alpha_1}^\dagger + D_{\alpha_1}) \cdots (a_{\alpha_n}^\dagger + D_{\alpha_n}) |0\rangle \\ &= (a_{\alpha_1}^\dagger + D_{\alpha_1}) \cdots (a_{\alpha_{n-1}}^\dagger + D_{\alpha_{n-1}}) a_{\alpha_n}^\dagger |0\rangle \\ &= (a_{\alpha_1}^\dagger + D_{\alpha_1}) \cdots (a_{\alpha_{n-2}}^\dagger + D_{\alpha_{n-2}}) a_{\alpha_{n-1}}^\dagger a_{\alpha_n}^\dagger |0\rangle \\ &= \cdots = a_{\alpha_1}^\dagger \cdots a_{\alpha_n}^\dagger |0\rangle. \end{aligned} \quad \text{QED (A5)}$$

APPENDIX B: DERIVATION OF TANI-TRANSFORMED HAMILTONIAN

Consider first the terms T_p and T_e in (1). Upon substitution from (63) one has

$$\begin{aligned} U^{-1} T_p U = \int dX [&\psi^{(0)\dagger}(X) + \psi^{(1)\dagger}(X) + \psi^{(2)\dagger}(X) + \psi^{(3)\dagger}(X) + \cdots] \\ &\times T(X) [\psi^{(0)}(X) + \psi^{(1)}(X) + \psi^{(2)}(X) + \psi^{(3)}(X) + \cdots]. \end{aligned} \tag{B1}$$

The transform of T_e is expressed similarly. We shall denote by ij the contribution of $\psi^{(i)\dagger} T \psi^{(j)}$. The total order[40] of any such contribution is $(i+j)$. By (65) and (66), the zero order contributions 00 to $U^{-1}T_pU$ and $U^{-1}T_eU$ are merely T_p and T_e. The first order contributions 10 and 01 are found from (65), (66), and (3) to be

$$\begin{aligned} &-\sum_\alpha \int dX\, dx \psi^\dagger(X) \psi^\dagger(x) [T(X) + T(x)] \phi_\alpha(Xx) a_\alpha + \text{h. c.} \\ &\quad - \int dX\, dx\, dX'\, dx' \psi^\dagger(X) \psi^\dagger(x) [T(X) + T(x)] \\ &\quad \times \Delta(Xx, X'x') \psi(x') \psi(X') + \text{h. c.} \end{aligned} \tag{B2}$$

where "h. c." denotes the Hermitian conjugate and Δ is the "bound state kernel"

$$\Delta(Xx, X'x') \equiv \sum_\alpha \phi_\alpha(Xx) \phi_\alpha^*(X'x'). \tag{B3}$$

The second order contributions are of the forms 11, 02, and 20. The contributions 11 are, by (65), (66), and (3) and after rearranging into normal order by use of Wick's theorem,

$$\begin{aligned} &\sum_{\alpha\beta} \int dX\, dx a_\alpha^\dagger \phi_\alpha^*(Xx) [T(X) + T(x)] \phi_\beta(Xx) a_\beta \\ &\quad + \sum_\alpha \int dX\, dx\, dX'\, dx' \psi^\dagger(X) \psi^\dagger(x) \Delta(Xx, X'x') [T(X') + T(x')] \phi_\alpha(X'x') a_\alpha + \text{h. c.} \\ &\quad + \int dX\, dx\, dX'\, dx'\, dX''\, dx'' \psi^\dagger(X) \psi^\dagger(x) \Delta(Xx, X''x'') [T(X'') + T(x'')] \Delta(X''x'', X'x') \psi(x') \psi(X') \end{aligned}$$

$$-\sum_{\alpha\beta}\int dX\,dx\,dX'a^\dagger_\alpha\psi^\dagger(X)\phi^*_\alpha(X'x)T(x)\phi_\beta(Xx)\psi(X')a_\beta-\sum_{\alpha\beta}\int dX\,dx\,dx'a^\dagger_\alpha\psi^\dagger(x)\phi^*_\alpha(Xx')T(X)\phi_\beta(Xx)\psi(x')a_\beta$$

$$-\sum_{\alpha}\int dX\,dx\,dX'\,dx'\,dX''\psi^\dagger(X)\psi^\dagger(x)\psi^\dagger(X')\Delta(Xx,X''x')T(x')\phi_\alpha(X'x')\psi(X'')a_\alpha+\text{h. c.}$$

$$-\sum_{\alpha}\int dX\,dx\,dX'\,dx'\,dx''\psi^\dagger(X)\psi^\dagger(x)\psi^\dagger(x')\Delta(Xx,X'x'')T(X')\phi_\alpha(X'x')\psi(x'')a_\alpha+\text{h. c.}+\cdots. \tag{B4}$$

Here "$\cdots$" stands for terms of the structure $\psi^\dagger\psi^\dagger\psi^\dagger\psi\psi\psi$. Such terms represent three-particle collisions, and are beyond the accuracy of the binary collision approximation within which (68) is evaluated. The contributions 02 and 20 are found to be

$$-\tfrac{1}{2}\sum_{\alpha\beta}\int dX\,dx\,dX'a^\dagger_\alpha\psi^\dagger(X)\{\phi^*_\alpha(X'x)T(X)\phi_\beta(Xx)+[T(X')\phi_\alpha(X'x)]^*\phi_\beta(Xx)\}\psi(X')a_\beta$$

$$-\tfrac{1}{2}\sum_{\alpha\beta}\int dX\,dx\,dx'a^\dagger_\alpha\psi^\dagger(x)\{\phi^*_\alpha(Xx')T(x)\phi_\beta(Xx)+[T(x')\phi_\alpha(Xx')]^*\phi_\beta(Xx)\}\psi(x')a_\beta$$

$$-\sum_{\alpha}\int dX\,dx\,dX'\,dX''\,dx''a^\dagger_\alpha\psi^\dagger(X)\{\tfrac{1}{4}\pi\phi^*_\alpha(X'x)T(X)\Delta(Xx,X''x'')$$

$$+(1-\tfrac{1}{4}\pi)[T(X')\phi_\alpha(X'x)]^*\Delta(Xx,X''x'')\}\psi(x'')\psi(X'')\psi(X')+\text{h. c.}$$

$$-\sum_{\alpha}\int dX\,dx\,dx'\,dX''\,dx''a^\dagger_\alpha\psi^\dagger(x)\{\tfrac{1}{4}\pi\phi^*_\alpha(Xx')T(x)\Delta(Xx,X''x'')$$

$$+(1-\tfrac{1}{4}\pi)[T(x')\phi_\alpha(Xx')]^*\Delta(Xx,X''x'')\}\psi(x')\psi(x'')\psi(X'')+\text{h. c.}+\cdots \tag{B5}$$

where "$\cdots$" again stands for three-particle terms of structure $\psi^\dagger\psi^\dagger\psi^\dagger\psi\psi\psi$. The terms of structures $a^\dagger\psi^\dagger\psi\psi\psi$ and $\psi^\dagger\psi^\dagger\psi^\dagger\psi a$ are viewed as binary collision terms and hence retained in (B5).

The third order contributions are of the forms 03, 30, 12, and 21. The contributions 03 and 30 are

$$\tfrac{1}{6}\sum_{\alpha\beta\gamma}\int dX\,dx\,dX'\,dx'a^\dagger_\alpha\psi^\dagger(X)\psi^\dagger(x)\phi^*_\alpha(X'x')[T(X)+T(x)]\phi_\beta(Xx')\phi_\gamma(X'x)a_\gamma a_\beta+\text{h. c.}$$

$$+(\tfrac{1}{4}\pi-\tfrac{2}{3})\sum_{\alpha\beta}\int dX\,dx\,dX'\,dx'\,dX''\,dx''\psi^\dagger(X)\psi^\dagger(x)\psi^\dagger(X')\psi^\dagger(x')\Delta(X'x',X''x'')[T(X)+T(x)]\phi_\alpha(Xx'')\phi_\beta(X''x)a_\beta a_\alpha+\text{h. c.}$$

$$-\tfrac{1}{2}\sum_{\alpha}\int dX\,dx\,dX'\,dx'\,dX''\psi^\dagger(X)\psi^\dagger(x)\psi^\dagger(X')[T(X)+T(x)]\Delta(Xx,X''x')\phi_\alpha(X'x')\psi(X'')a_\alpha+\text{h. c.}$$

$$-\tfrac{1}{2}\sum_{\alpha}\int dX\,dx\,dX'\,dx'\,dx''\psi^\dagger(X)\psi^\dagger(x)\psi^\dagger(x')[T(X)+T(x)]\Delta(Xx,X'x'')\phi_\alpha(X'x')\psi(x'')a_\alpha+\text{h. c.}+\cdots \tag{B6}$$

where the contributions "$\cdots$" are again three-particle or higher-order collision terms in the sense that they correspond to processes in which the number of incoming particles is three or more *and* the number of outgoing particles is three or more; all terms in which the number of incoming *or* outgoing particles is two or less are included. The contributions 12 and 21 are

$$\tfrac{1}{2}\sum_{\alpha\beta\gamma}\int dX\,dx\,dX'\,dx'a^\dagger_\alpha a^\dagger_\beta\phi^*_\alpha(X'x)\phi^*_\beta(Xx')[T(X')+T(x')]\phi_\gamma(X'x')\psi(x)\psi(X)a_\gamma+\text{h. c.}$$

$$+\tfrac{1}{4}\pi\sum_{\alpha\beta}\int dX\,dx\,dX'\,dx'\,dX''\,dx''a^\dagger_\alpha a^\dagger_\beta\phi^*_\alpha(X''x)\phi^*_\beta(Xx'')[T(X'')+T(x'')]\Delta(X''x'',X'x')\psi(x')\psi(X')\psi(x)\psi(X)+\text{h. c.}$$

$$-2(1-\tfrac{1}{4}\pi)\sum_{\alpha\beta}\int dX\,dx\,dX'\,dx'\,dX''a^\dagger_\alpha\psi^\dagger(X)\Delta(Xx,X'x')\phi^*_\alpha(X''x)T(X'')\phi_\beta(X''x')\psi(X')a_\beta$$

$$-2(1-\tfrac{1}{4}\pi)\sum_{\alpha\beta}\int dX\,dx\,dX'\,dx'\,dx''a^\dagger_\alpha\psi^\dagger(x)\Delta(Xx,X'x')\phi^*_\alpha(Xx'')T(x'')\phi_\beta(X''x')\psi(x')a_\beta$$

$$-(1-\tfrac{1}{4}\pi)\sum_{\alpha}\int dX\,dx\,dX'\,dx'\,dX''\,dX'''\,dx'''\psi^\dagger(X)\psi^\dagger(x)\psi^\dagger(X')\Delta(Xx,X'''x''')\Delta(X'x''',X''x')T(X''')\phi_\alpha(X'''x')\psi(X'')a_\alpha+\text{h. c.}$$

$$-(1-\tfrac{1}{4}\pi)\sum_{\alpha}\int dX\,dx\,dX'\,dx'\,dx''\,dX'''\,dx'''\psi^\dagger(X)\psi^\dagger(x)\psi^\dagger(x')\Delta(Xx,X'''x''')\Delta(X'''x',X'x'')T(x''')\phi_\alpha(X'x''')\psi(x'')a_\alpha+\text{h. c.}$$

$$-\tfrac{1}{2}\sum_{\alpha}\int dX\,dx\,dX'\,dx'\,dX''\,dx''\,dX'''a^\dagger_\alpha\psi^\dagger(X)\Delta(Xx,X'x')\phi^*_\alpha(X'''x)T(X''')\Delta(X'''x',X''x'')\psi(x'')\psi(X'')\psi(X')+\text{h. c.}$$

$$-\tfrac{1}{2}\sum_{\alpha}\int dX\,dx\,dX'\,dx'\,dX''\,dx''\,dx'''a^\dagger_\alpha\psi^\dagger(x)\Delta(Xx,X'x')\phi^*_\alpha(Xx''')T(x''')\Delta(X'x''',X''x'')\psi(x'')\psi(X'')\psi(x')+\text{h. c.}+\cdots. \tag{B7}$$

The fourth order contributions are of the forms 04, 40, 13, 31, and 22. We cannot evaluate the contributions 04 and 40 here since the series (63) for the transformed proton and electron field operators have only been evaluated up to third order. However, the only possible contributions to the terms 04 and 40 which are of binary collision form are those of structures $a^\dagger\psi^\dagger\psi a$, $a^\dagger\psi^\dagger\psi\psi a$, $a^\dagger\psi^\dagger\psi^\dagger\psi a$, $\psi^\dagger\psi^\dagger\psi^\dagger\psi^\dagger aa$, $a^\dagger a^\dagger\psi\psi\psi\psi$, $\psi^\dagger\psi^\dagger\psi^\dagger\psi a$, and $a^\dagger\psi^\dagger\psi\psi\psi$. Such contributions will be found to also occur in other terms [e.g., there are such terms in (B4), (B5), and (B7)] in the second order. Hence we shall, for consistency, drop all fourth and higher order contributions to such terms. Note that this requires dropping the term with coefficient $(\pi/4-\frac{2}{3})$ in (B6), and the terms with coefficients $\pi/4$ and $2(1-\pi/4)$ in (B7). By the same token, the terms of structure $a^\dagger\psi^\dagger\psi\psi\psi$ and $\psi^\dagger\psi^\dagger\psi^\dagger\psi a$ in (B7), which are of fifth order, will also be dropped, as will such fifth order contributions in similar terms which will subsequently be evaluated.

The terms 13 and 31 are found to be

$$-\tfrac{1}{6}\sum_{\alpha\beta\gamma\delta}\int dX\,dx\,dX'\,dx'a^\dagger_\alpha a^\dagger_\beta\phi^*_\alpha(Xx)\phi^*_\beta(X'x')[T(X)+T(x)+T(X')+T(x')]\phi_\gamma(Xx')\phi_\delta(X'x)a_\delta a_\gamma+\cdots \tag{B8}$$

where the omitted terms "$\cdots$" consist of multiple (higher than binary) collision terms, binary collision terms of fifth and sixth order in wavefunctions, and binary collision terms $a^\dagger\psi^\dagger\psi a$ of fourth order in wavefunctions.[41] The contributions 22 consist entirely of multiple collision terms, binary collision terms of fifth order, and binary collision terms $a^\dagger\psi^\dagger\psi a$ of fourth order, and hence will all be dropped.

Next we consider the transforms of the interaction terms V_{pp}, V_{ee}, and V_{pe} in (1), which can be evaluated in analogy with (B1). We shall employ a notation similar to that used in enumerating the various contributions to (B1), denoting by *ijkl* the sum of the contributions of form $\psi^{(i)\dagger}\psi^{(j)\dagger}V\psi^{(k)}\psi^{(l)}$ to the Tani transforms of V_{pp}, V_{ee}, and V_{pe}. The terms 0000 are merely V_{pp}, V_{ee}, and V_{pe} themselves. The first order contributions 0001, 1000, 0010, and 0100 are found in analogy with the derivation of (B2) to be

$$
\begin{aligned}
&-\sum_{\alpha}\int dX\,dx\,\psi^{\dagger}(X)\psi^{\dagger}(x)V(Xx)\phi_{\alpha}(Xx)a_{\alpha}+\text{h. c.}-\int dX\,dx\,dX'\,dx'\,\psi^{\dagger}(X)\psi^{\dagger}(x)V(Xx)\Delta(Xx,X'x')\psi(x')\psi(X')+\text{h. c.}\\
&\quad+\sum_{\alpha}\int dX\,dx\,dX'\,\psi^{\dagger}(X)\psi^{\dagger}(x)\psi^{\dagger}(X')[V(Xx)+V(XX')]\phi_{\alpha}(X'x)\psi(X)a_{\alpha}+\text{h. c.}\\
&\quad+\sum_{\alpha}\int dX\,dx\,dx'\,\psi^{\dagger}(X)\psi^{\dagger}(x)\psi^{\dagger}(x')[V(Xx)+V(xx')]\phi_{\alpha}(Xx')\psi(x)a_{\alpha}+\text{h. c.}+\cdots
\end{aligned}
\tag{B9}
$$

where the terms "$\cdots$" are ternary collision terms of the form $\psi^{\dagger}\psi^{\dagger}\psi^{\dagger}\psi\psi\psi$.

The second order contributions are of the forms 0002, 2000, 0020, 0200, 1001, 0110, 0101, 1010, 0011, and 1100. The terms 0002, 2000, 0020, and 0200 are

$$
\begin{aligned}
&(1-\tfrac{1}{4}\pi)\sum_{\alpha}\int dX\,dx\,dX'\,dX''\,dx''\psi^{\dagger}(X)\psi^{\dagger}(x)\psi^{\dagger}(X')V(Xx)\Delta(X'x,X''x'')\phi_{\alpha}(Xx'')\psi(X'')a_{\alpha}+\text{h. c.}\\
&\quad+\tfrac{1}{2}(1-\tfrac{1}{4}\pi)\sum_{\alpha}\int dX\,dx\,dX'\,dX''\,dx''\psi^{\dagger}(X)\psi^{\dagger}(x)\psi^{\dagger}(X')V(XX')\Delta(X'x,X''x'')\phi_{\alpha}(Xx'')\psi(X'')a_{\alpha}+\text{h. c.}\\
&\quad+\tfrac{1}{2}(1-\tfrac{1}{4}\pi)\sum_{\alpha}\int dX\,dx\,dx'\,dX''\,dx''\psi^{\dagger}(X)\psi^{\dagger}(x)\psi^{\dagger}(x')V(xx')\Delta(Xx',X''x'')\phi_{\alpha}(X''x)\psi(x'')a_{\alpha}+\text{h. c.}+\cdots
\end{aligned}
\tag{B10}
$$

where "$\cdots$" are multiple collision terms. Similarly, the sum of the terms 1001, 0110, 0101, 1010, 0011, and 1100 is

$$
\begin{aligned}
&\sum_{\alpha\beta}\int dX\,dx a_{\alpha}^{\dagger}\phi_{\alpha}^{*}(Xx)V(Xx)\phi_{\beta}(Xx)a_{\beta}+\sum_{\alpha}\int dX\,dx\,dX'\,dx'\psi^{\dagger}(X)\psi^{\dagger}(x)\Delta(Xx,X'x')V(X'x')\phi_{\alpha}(X'x')a_{\alpha}+\text{h. c.}\\
&\quad+\int dX\,dx\,dX'\,dx'\,dX''\,dx''\psi^{\dagger}(X)\psi^{\dagger}(x)\Delta(Xx,X''x'')V(X''x'')\Delta(X''x'',X'x')\psi(x')\psi(X')\\
&\quad+\sum_{\alpha\beta}\int dX\,dx\,dX'a_{\alpha}^{\dagger}\psi^{\dagger}(X)\phi_{\alpha}^{*}(X'x)[V(Xx)+V(XX')]\phi_{\beta}(X'x)\psi(X)a_{\beta}\\
&\quad+\sum_{\alpha\beta}\int dX\,dx\,dx'a_{\alpha}^{\dagger}\psi^{\dagger}(x)\phi_{\alpha}^{*}(Xx')[V(Xx)+V(xx')]\phi_{\beta}(Xx')\psi(x)a_{\beta}\\
&\quad-\sum_{\alpha\beta}\int dX\,dx\,dX'a_{\alpha}^{\dagger}\psi^{\dagger}(X)\phi_{\alpha}^{*}(X'x)[V(Xx)+V(X'x)+V(XX')]\phi_{\beta}(Xx)\psi(X')a_{\beta}\\
&\quad-\sum_{\alpha\beta}\int dX\,dx\,dx'a_{\alpha}^{\dagger}\psi^{\dagger}(x)\phi_{\alpha}^{*}(Xx')[V(Xx)+V(Xx')+V(xx')]\phi_{\beta}(Xx)\psi(x')a_{\beta}\\
&\quad+\sum_{\alpha}\int dX\,dx\,dX'\,dx'\,dX''\psi^{\dagger}(X)\psi^{\dagger}(X')\psi^{\dagger}(x')\Delta(X'x',X''x)[V(Xx)+V(XX'')]\phi_{\alpha}(X''x)\psi(X)a_{\alpha}+\text{h. c.}\\
&\quad+\sum_{\alpha}\int dX\,dx\,dX'\,dx'\,dx''\psi^{\dagger}(x)\psi^{\dagger}(X')\psi^{\dagger}(x')\Delta(X'x',Xx'')[V(Xx)+V(xx'')]\phi_{\alpha}(Xx'')\psi(x)a_{\alpha}+\text{h. c.}\\
&\quad-\sum_{\alpha}\int dX\,dx\,dX'\,dx'\,dX''\psi^{\dagger}(X)\psi^{\dagger}(X')\psi^{\dagger}(x')\Delta(X'x',X''x)[V(Xx)+V(Xx')+\tfrac{1}{2}V(XX')+V(X''x)+V(XX'')]\phi_{\alpha}(Xx)\psi(X'')a_{\alpha}+\text{h. c.}\\
&\quad-\sum_{\alpha}\int dX\,dx\,dX'\,dx'\,dx''\psi^{\dagger}(x)\psi^{\dagger}(X')\psi^{\dagger}(x')\Delta(X'x',Xx'')[V(Xx)+V(Xx'')+\tfrac{1}{2}V(xx')+V(xx'')]\phi_{\alpha}(Xx)\psi(x'')a_{\alpha}+\text{h. c.}\\
&\quad+\tfrac{1}{2}\sum_{\alpha\beta}\int dX\,dx\,dX'\,dx'\psi^{\dagger}(X)\psi^{\dagger}(x)\psi^{\dagger}(X')\psi^{\dagger}(x')[V(XX')+V(xx')+V(Xx')+V(X'x)]\phi_{\alpha}(Xx)\phi_{\beta}(X'x')a_{\beta}a_{\alpha}+\text{h. c.}+\cdots.
\end{aligned}
\tag{B11}
$$

Next consider the third order terms. The sum of terms of structure 0003, 3000, 0030, and 0300 is found to be

$$
\begin{aligned}
&\tfrac{1}{6}\sum_{\alpha\beta\gamma}\int dX\,dx\,dX'\,dx'a_{\alpha}^{\dagger}\psi^{\dagger}(X)\psi^{\dagger}(x)\phi_{\alpha}^{*}(X'x')V(Xx)\phi_{\beta}(Xx')\phi_{\gamma}(X'x')a_{\gamma}a_{\beta}+\text{h. c.}\\
&\quad-\tfrac{1}{4}\sum_{\alpha}\int dX\,dx\,dX'\,dx'\,dX''\psi^{\dagger}(X)\psi^{\dagger}(x)\psi^{\dagger}(X')[2V(Xx)+V(XX')]\Delta(Xx,X''x')\phi_{\alpha}(X'x')\psi(X'')a_{\alpha}+\text{h. c.}\\
&\quad-\tfrac{1}{4}\sum_{\alpha}\int dX\,dx\,dX'\,dx'\,dx''\psi^{\dagger}(X)\psi^{\dagger}(x)\psi^{\dagger}(x')[2V(Xx)+V(xx')]\Delta(Xx,X'x'')\phi_{\alpha}(X'x')\psi(x'')a_{\alpha}+\text{h. c.}\\
&\quad+\tfrac{1}{2}\sum_{\alpha}\int dX\,dx\,dX'\,dx'\,dx''\psi^{\dagger}(X)\psi^{\dagger}(x)\psi^{\dagger}(x')V(Xx)\Delta(Xx',X'x'')\phi_{\alpha}(X'x)\psi(x'')a_{\alpha}+\text{h. c.}+\cdots
\end{aligned}
\tag{B12}
$$

where the terms "$\cdots$" are multiple collision terms and terms of structure ($\psi^{\dagger}\psi^{\dagger}\psi^{\dagger}\psi^{\dagger}aa+\text{h. c.}$) which are of fourth order in wavefunctions, hence of the same order as terms 04 and 40 which were not calculated [see the discussion following Eq. (B7)]. The sum of the terms 1002, 2001, 1020, 0201, 0120, 0210, 0102, 2010, 0012, 2100, 0021, and 1200 is

$$
\begin{aligned}
&\tfrac{1}{2}\sum_{\alpha\beta\gamma}\int dX\,dx\,dX\,dx'a_{\alpha}^{\dagger}a_{\beta}^{\dagger}\phi_{\alpha}^{*}(X'x)\phi_{\beta}^{*}(Xx')[V(X'x)+V(Xx')]\phi_{\gamma}(X'x')\psi(x)\psi(X)a_{\gamma}+\text{h. c.}\\
&\quad-(\tfrac{3}{2}-\tfrac{1}{4}\pi)\sum_{\alpha}\int dX\,dx\,dX'\,dX''\,dx''\psi^{\dagger}(X)\psi^{\dagger}(x)\psi^{\dagger}(X')V(Xx)\Delta(X'x,X''x'')\phi_{\alpha}(Xx'')\psi(X'')a_{\alpha}+\text{h. c.}\\
&\quad-\tfrac{1}{2}(\tfrac{3}{2}-\tfrac{1}{4}\pi)\sum_{\alpha}\int dX\,dx\,dX'\,dX''\,dx''\psi^{\dagger}(X)\psi^{\dagger}(x)\psi^{\dagger}(X')V(XX')\Delta(X'x,X''x'')\phi_{\alpha}(Xx'')\psi(X'')a_{\alpha}+\text{h. c.}\\
&\quad-\tfrac{1}{2}(\tfrac{3}{2}-\tfrac{1}{4}\pi)\sum_{\alpha}\int dX\,dx\,dx'\,dX''\,dx''\psi^{\dagger}(X)\psi^{\dagger}(x)\psi^{\dagger}(x')V(xx')\Delta(Xx',X''x'')\phi_{\alpha}(X''x)\psi(x'')a_{\alpha}+\text{h. c.}\\
&\quad+\tfrac{1}{2}\sum_{\alpha\beta\gamma}\int dX\,dx\,dX'\,dx'a_{\alpha}^{\dagger}\psi^{\dagger}(X)\psi^{\dagger}(x)\phi_{\alpha}^{*}(X'x')V(Xx)\phi_{\beta}(X'x')\phi_{\gamma}(Xx')a_{\gamma}a_{\beta}+\text{h. c.}+\cdots.
\end{aligned}
\tag{B13}
$$

Finally, the terms 0111, 1110, 1011, and 1101 are

$$
\begin{aligned}
&-\sum_{\alpha\beta\gamma}\int dX\,dx\,dX'\,dx'a_{\alpha}^{\dagger}\psi^{\dagger}(X)\psi^{\dagger}(x)\phi_{\alpha}^{*}(X'x')[V(Xx')+V(X'x)+V(XX')+V(xx')]\phi_{\beta}(Xx)\phi_{\gamma}(X'x')a_{\gamma}a_{\beta}+\text{h. c.}\\
&\quad+\sum_{\alpha\beta\gamma}\int dX\,dx\,dX'\,dx'a_{\alpha}^{\dagger}\psi^{\dagger}(X)\psi^{\dagger}(x)\phi_{\alpha}^{*}(X'x')[V(X'x')+V(XX')+V(xx')]\phi_{\beta}(X'x)\phi_{\gamma}(Xx')a_{\gamma}a_{\beta}+\text{h. c.}+\cdots.
\end{aligned}
\tag{B14}
$$

The fourth order terms 0004, 4000, 0040, and 0400 cannot be evaluated here, since the series (63) were only evaluated to third order. However, in analogy with the case of the terms 04 and 40, one concludes that the only binary collision terms which can arise from 0004, 4000, 0040, and 0400 are those of structures $a^\dagger\psi^\dagger\psi^\dagger\psi a$, $a^\dagger\psi^\dagger\psi\psi a$, $\psi^\dagger\psi^\dagger\psi^\dagger\psi^\dagger aa$, $a^\dagger a^\dagger\psi\psi\psi\psi$, $\psi^\dagger\psi^\dagger\psi^\dagger\psi a$, and $a^\dagger\psi^\dagger\psi\psi\psi$. Since we have consistently been dropping fourth or higher order contributions to such terms, we need not concern ourselves further with them here. The remaining fourth order terms are of the forms 1003, 3001, 1030, 0301, 0103, 3010, 0130, 0310, 0013, 3100, 0031, 1300, 2002, 0220, 2020, 0202, 0022, 2200, 0112, 2110, 0121, 1210, 0211, 1120, 1012, 2101, 1021, 1201, 2011, 1102, and 1111. The sum of the terms involving 1 and 3 is

$$-\tfrac{1}{6}\sum_{\alpha\beta\gamma\delta}\int dX\,dx\,dX'\,dx' a_\alpha^\dagger a_\beta^\dagger \phi_\alpha^*(Xx)\phi_\beta^*(X'x')[V(Xx)+V(Xx')]\phi_\gamma(Xx')\phi_\delta(X'x)a_\delta a_\gamma + \cdots \tag{B15}$$

where "$\cdots$" stands for multiple collision and fifth[42] and higher order terms. The terms 2002, 0220, 2020, 0202, 0022, and 2200 are all of multiple collision form or of fifth or higher order, and hence negligible. The terms 0112, 2110, 0121, 1210, 0211, 1120, 1012, 2101, 1021, 1201, 2011, and 1102 yield

$$-\tfrac{1}{2}\sum_{\alpha\beta\gamma\delta}\int dX\,dx\,dX'\,dx' a_\alpha^\dagger a_\beta^\dagger \phi_\alpha^*(Xx)\phi_\beta^*(X'x')[V(Xx)+V(X'x)]\phi_\gamma(X'x)\phi_\delta(Xx')a_\delta a_\gamma + \cdots . \tag{B16}$$

Finally, the term 1111 is found to be

$$\begin{aligned}&\tfrac{1}{2}\sum_{\alpha\beta\gamma\delta}\int dX\,dx\,dX'\,dx' a_\alpha^\dagger a_\beta^\dagger \phi_\alpha^*(Xx)\phi_\beta^*(X'x')[V(Xx')+V(X'x)+V(XX')+V(xx')]\phi_\gamma(Xx)\phi_\delta(X'x')a_\delta a_\gamma \\ &\quad -\tfrac{1}{2}\sum_{\alpha\beta\gamma\delta}\int dX\,dx\,dX'\,dx' a_\alpha^\dagger a_\beta^\dagger \phi_\alpha^*(Xx)\phi_\beta^*(X'x')[V(XX')+V(xx')]\phi_\gamma(X'x)\phi_\delta(Xx')a_\delta a_\gamma + \cdots .\end{aligned} \tag{B17}$$

This completes the evaluation of the transformed Hamiltonian to fourth order. Upon combining terms of like operator structure, one finds the expressions in Eqs. (68) ff.

APPENDIX C: PROPERTIES OF THE UNBOUND PROTON-ELECTRON INTERACTION

Let $|\phi\rangle$ be any one proton—one electron state:

$$|\phi\rangle = \int dX\,dx\,\phi(Xx)\psi^\dagger(X)\psi^\dagger(x)|0\rangle. \tag{C1}$$

Assuming the ϕ_α to be single-atom energy eigenstates (71) so that (80) holds, one easily verifies with (78) that

$$\begin{aligned}H_{pe}|\phi\rangle = \int dX\,dx[&V(Xx)\phi(Xx) \\ &-\textstyle\sum_\alpha \epsilon_\alpha(\phi_\alpha,\phi)\phi_\alpha(Xx)]\psi^\dagger(X)\psi^\dagger(x)|0\rangle.\end{aligned} \tag{C2}$$

Then by (1) and (75)

$$\begin{aligned}&(T_p+T_e+H_{pe})|\phi\rangle \\ &\quad = \int dX\,dx[H(Xx)\phi(Xx) - \textstyle\sum_\alpha \epsilon_\alpha(\phi_\alpha,\phi)\phi_\alpha(Xx)] \\ &\quad\quad \times\psi^\dagger(X)\psi^\dagger(x)|0\rangle.\end{aligned} \tag{C3}$$

If ϕ is one of the bound eigenstates ϕ_α of $H(Xx)$, then one has trivially[43]

$$(T_p+T_e+H_{pe})|\phi_\alpha\rangle = 0. \tag{C4}$$

On the other hand, if $\phi_\perp$ is orthogonal to all the bound states so that $(\phi_\alpha,\phi_\perp)=0$, then one has with (1)

$$\begin{aligned}(T_p+T_e+H_{pe})|\phi_\perp\rangle &= (T_p+T_e+V_{pe})|\phi_\perp\rangle \\ &= \int dX\,dx H(Xx)\phi_\perp(Xx)\psi^\dagger(X)\psi^\dagger(x)|0\rangle.\end{aligned} \tag{C5}$$

Let $\phi_k(Xx)$ be the continuum (unbound, positive energy) eigenstates of $H(Xx)$, with energies $\epsilon_k \geq 0$. One can expand any $\phi(Xx)$ in the form

$$\phi(Xx) = \sum_\alpha c_\alpha\phi_\alpha(Xx) + \sum_k c_k\phi_k(Xx) \tag{C6}$$

where $(\phi_\alpha,\phi_k)=0$. Then by (C4) and (C5)

$$(T_p+T_e+H_{pe})|\phi\rangle = \sum_k c_k\epsilon_k|\phi_k\rangle, \tag{C7}$$

so that

$$\langle\phi|(T_p+T_e+H_{pe})|\phi\rangle = \sum_k |c_k|^2\epsilon_k \geq 0, \tag{C8}$$

i.e., $(T_p+T_e+H_{pe})$ is positive semidefinite and hence has no bound states.

*Supported in part by National Science Foundation Grants GP-11728 and DMR72-03211.

[1]R.H. Stolt and W.E. Brittin, Phys. Rev. Lett. 27, 616 (1971); Wesley E. Brittin, in *Lectures in Theoretical Physics*, edited by Wesley E. Brittin (Colorado Associated University Press, Boulder, 1973), Vol. XIVB, pp. 55 ff.
[2]A.Y. Sakakura, Phys. Rev. Lett. 27, 822 (1971) and in *Lectures in Theoretical Physics*, edited by Wesley E. Brittin (Colorado Associated University Press, Boulder, 1973), Vol. XIVB, pp. 483 ff.
[3]M.D. Girardeau, in *Lectures in Theoretical Physics*, edited by Wesley E. Brittin (Colorado Associated University Press, Boulder, 1973), Vol. XIVB, pp. 147 ff., in particular pp. 167 ff.
[4]M.D. Girardeau, Phys. Rev. Lett. 27, 1416 (1971).
[5]W.E. Brittin and A.Y. Sakakura, in *Physical Reality and Mathematical Description*, edited by C.P. Enz and J. Mehra (Reidel, New York, 1974).
[6]S. Tani, Phys. Rev. 117, 252 (1960).
[7]M.D. Girardeau (to be published).
[8]To obtain a complete set, the continuum atomic states would have to be included.
[9]See, e.g., R.F. Streater and A.S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964), pp. 146 ff., in particular Theorem 4-12.
[10]Such exchange kernels also arise naturally in the formulation of Ref. 3. Their many-electron generalizations are an important ingredient in the investigation of many-electron exchange effects carried out in Ref. 7.
[11]We assume that the system consists of finite numbers of nuclei and electrons enclosed in a box of *finite* volume. This assumption is merely a matter of convenience. There is no reason is principle why the present approach could not be incorporated into recent formulations of many-body quantum and statistical mechanics of *infinite* systems of finite density.
[12]We shall use the same symbol $|0\rangle$ to denote the vacuum state in $\mathcal{S}$, that in $\mathcal{A}$, and the direct product of these two vacua. The distinction should always be clear from the context.
[13]F.J. Dyson, Phys. Rev. 102, 1217 (1956).
[14]D. Bohm and D. Pines, Phys. Rev. 92, 609 (1953).

[15]J. M. Blatt and T. Matsubara, Prog. Theor. Phys. **20**, 553 (1958).

[16]In Eq. (20), $|0\rangle$ stands for the vacuum state in $\mathcal{J}$, which satisfies $A_\alpha|0\rangle = a_\omega|0\rangle = 0$, the vanishing of $A_\alpha|0\rangle$ following from $\psi(X)|0\rangle = \psi(x)|0\rangle = 0$.

[17]This should be qualified by the statement that even in the zero-density limit, the terms arising from $C_{\alpha\beta}$ play an important role in atomic collisions, being closely related to exchange effects (primarily electron exchange but in principle also nuclear exchange) occurring in such collisions. Such contributions will be evaluated when we evaluate the Tani-transformed Hamiltonian; however, our purpose in this section is to introduce and motivate the transformation (14), (18) without obscuring the qualitative ideas by such details.

[18]Such a state $|\psi\rangle$ is of the general form

$$|\psi\rangle = \text{const} \int dX_1 \cdots dX_m dx_1 \cdots dx_l \, \phi(X_1 \cdots X_m x_1 \cdots x_l) \times \psi^\dagger(X_1) \cdots \psi^\dagger(X_m)\psi^\dagger(x_1) \cdots \psi^\dagger(x_l)|0\rangle.$$

The corresponding state (according to the one-one correspondence $\mathcal{S} \longleftrightarrow \mathcal{J}_0$) in $\mathcal{S}$ is of exactly the same form, except that then $|0\rangle$ is the vacuum of $\mathcal{S}$ rather than of $\mathcal{J}$.

[19]The calculation initially gives an operator matrix element $(\alpha|H|Xx)$:

$$(\alpha|H|Xx) = -\phi_\alpha^*(Xx)H(Xx) + \int dX'dx'\phi_\alpha^*(X'x')H(X'x')\Delta(X'x', Xx).$$

However, upon substituting this into (73) and making use of the hermiticity of $H(Xx)$, one can rewrite the expression (73) so that the matrix element is the c-number function $(Xx|H|\alpha)^*$.

[20]In such a case $(Xx|H|\alpha)$ reduces to $(Xx|V_{\text{ext}}|\alpha)$ where V_{ext} is the external potential, since the contribution from the free-atom Hamiltonian vanishes by the previous argument (76).

[21]More precisely, it is of range $\sim (m/M)a_0 \ll a_0$ with respect to the nuclear separation $\mathbf{R}-\mathbf{R}'$, and of range $\sim a_0$ with respect to the electron—nucleus separations $\mathbf{r}-\mathbf{R}$ and $\mathbf{r}'-\mathbf{R}'$, hence also of range $\sim a_0$ with respect to $\mathbf{r}-\mathbf{r}'$. Here m and M are the electron and nuclear masses. To see this, note that the quantum numbers α labeling the atomic bound state ϕ_α can be taken to consist of a wave vector $\mathbf{k}$, where $\hbar\mathbf{k}$ is the total translational momentum of the atom, together with a set v of internal quantum numbers (the usual atomic quantum numbers). Accordingly, one can decompose ϕ_α into a center-of-mass translational wavefunction $\exp(i\mathbf{k}\cdot\mathbf{R}_{\text{c.m.}})$ and an internal wavefunction u_v:

$$\phi_\alpha(Xx) = \Omega^{-1/2}\exp(i\mathbf{k}\cdot\mathbf{R}_{\text{c.m.}})\, u_v(\mathbf{r}-\mathbf{R}, \sigma_p, \sigma_e).$$

Here $\mathbf{R}_{\text{c.m.}} = (M\mathbf{R}+m\mathbf{r})/(M+m)$, Ω is the volume of the system, and σ_p and σ_e are the proton and electron spin variables. We assume periodic boundary conditions in a macroscopic cube of volume Ω, thus determining the allowed wavevectors $\mathbf{k}$ in the usual way. Substitution of this decomposition of the ϕ_α into (B3) yields

$$\Delta(Xx, X'x') = \Omega^{-1}\Sigma_{\mathbf{k},v}\exp[i\mathbf{k}\cdot(\mathbf{R}_{\text{c.m.}} - \mathbf{R}'_{\text{c.m.}})]u_v(\mathbf{r}-\mathbf{R}, \sigma_p, \sigma_e)u_v^*(\mathbf{r}'-\mathbf{R}', \sigma_p', \sigma_e')$$
$$= \delta(\mathbf{R}_{\text{c.m.}} - \mathbf{R}'_{\text{c.m.}})\Sigma_v u_v(\mathbf{r}-\mathbf{R}, \sigma_p, \sigma_e)u_v^*(\mathbf{r}'-\mathbf{R} - (m/M)(\mathbf{r}-\mathbf{r}'), \sigma_p', \sigma_e').$$

Since the U_v have range $\sim a_0$, one sees that Δ is only appreciable when $\mathbf{r}$ and $\mathbf{r}'$ are both within $\sim a_0$ of $\mathbf{R}$. Then by the definition of $\mathbf{R}_{\text{c.m.}}$, $\mathbf{R}$ with be within $\sim (m/M)a_0$ of $\mathbf{R}_{\text{c.m.}}$, and since $\mathbf{R}_{\text{c.m.}} = \mathbf{R}'_{\text{c.m.}}$ because of the delta function, $\mathbf{R}'$ will also be within $\sim (m/M)a_0$ of $\mathbf{R}_{\text{c.m.}}$, so that $\mathbf{R}$ and $\mathbf{R}'$ will be within $\sim (m/M)a_0$ of each other. Thus Δ has range $\sim (m/M)a_0 \ll a_0$ with respect to $\mathbf{R}-\mathbf{R}'$, and range $\sim a_0$ with respect to $\mathbf{r}-\mathbf{R}$ and $\mathbf{r}'-\mathbf{R}'$, hence range $\sim a_0$ with respect to $\mathbf{r}-\mathbf{r}'$.

[22]Partially contracted terms contribute only to multiparticle collision terms $a^\dagger a^\dagger \psi^\dagger \psi a a$, etc.

[23]Such a separation is not *assumed*; it is an automatic consequence of the algebra generating the various terms in $U^{-1}HU$.

[24]Interatomic proton exchange also contributes, but does not produce a distinct matrix element, since interatomic proton exchange is equivalent to interatomic electron exchange followed by exchange of the whole atoms; exchange of whole atoms is already accounted for by the commutation relations of the a_α operators.

[25]Only even powers of these matrix elements contribute to physically observable quantities.

[26]On physical grounds, one expects that if the ϕ_α are chosen to be energy eigenstates, then such spontaneous decay will not in fact occur. This could arise through cancellation of this third order term by a formally) fifth order term, in analogy with the cancellation (76) occurring in the simpler matrix element (74) in case the ϕ_α are energy eigenstates. Evaluation of such fifth order terms in beyond the scope of this paper.

[27]The absence of explicit exchange contributions to the matrix element (92) does not mean that exchange plays no role in such collisions. The point is that the operator product $\psi^\dagger(X)\psi^\dagger(x)\psi^\dagger(X')\psi^\dagger(x')$ is already antisymmetric under proton and electron exchange, so that the operator (91) implicitly includes both direct and exchange processes.

[28]D. Gilbert, Ph.D. dissertation, University of Oregon, 1974 and to be published.

[29]M. D. Girardeau, J. Math. Phys. (to be published).

[30]There will still be corrections to the *states* containing both protons and atoms or electrons and atoms, since such states are not eigenstates of the separate portions V_{pp} or V_{ee} of the Hamiltonian.

[31]$\mathcal{J}_j$ can be constructed as the space of all linear combinations of states of the form $a_{\alpha_1}^\dagger \cdots a_{\alpha_j}^\dagger|\psi\rangle$ where $|\psi\rangle$ ranges over $\mathcal{J}_0$ and $\alpha_1 \cdots a_j$ range over all quantum numbers. But since H does not contain a_α and $a_\alpha^\dagger$ operators, it commutes with the a_α and $a_\alpha^\dagger$; hence if $H|\psi\rangle = E|\psi\rangle$ then

$$Ha_{\alpha_1}^\dagger \cdots a_{\alpha_j}^\dagger|\psi\rangle = Ea_{\alpha_1}^\dagger \cdots a_{\alpha_j}^\dagger|\psi\rangle.$$

[32]For simplicity we give the argument here for the canonical partition function. In practice, one would work with a grand partition function. The generalization of the argument to that case is trivial.

[33]The second quantized Hamiltonian appears on pp. 231—234 of the author's Boulder Lectures (Ref. 3). The terms which were dropped were those vanishing by virture of the "strong" orthogonality" constraint $A_\alpha|c) = 0$ on allowed state vectors $|c)$, where A_α is the equivalent to the Hermitian conjugate of the physical atom creation operator (3) of the present paper. However, such terms were not dropped in some of the preliminary calculations, allowing a partial comparison. Thus, e.g., the sum of the matrix elements $(Xx|T|\alpha)$ and $(Xx|V|\alpha)$ defined by Eqs. (III. 78) and (III. 83) of Ref. 3 is just the negative of the matrix element $(Xx|H|\alpha)$ defined by Eq. (74) of the present paper, and hence shares the desirable physical properties of (74) (no unphysical spontaneous breakup); the sign difference arises from a physically irrelevant difference of choice of phases. Similarly, combination of suitable terms in (III. 76) and (III. 81) of Ref. 3 yields an unbound proton-unbound electron interaction matrix element

$$(Xx|H|X'x') = V(Xx)\delta(X-X')\delta(x-x') - \Delta(Xx, X'x')H(X'x').$$

This is not quite the same as Eq. (79) of the present paper, but does in fact effectively reduce to (80) in the special case that the ϕ_α are energy eigenstates, when use is made of hermiticity. Thus it shares with (80) the desired property of excluding the bound proton—bound electron interaction. On the other hand, there appear to be significant differences in some other matrix elements.

[34]T. Nishigori, Prog. Theor. Phys. **51**, 1387 (1974).

[35]See the unnumbered equation after Eq. (2.19) of Ref. 34.

[36]H. L. Sahlin and J. L. Schwartz, Phys. Rev. **138**, B267 (1965).

[37]A. Goldberg and R.D. Puff, Phys. Rev. Lett. 30, 869 (1973).
[38]R. Munro, private communication.
[39]Pp. 51, 52 of Ref. 28.
[40]At this point we are maintaining the previous convention that A_α and $A_\alpha^\dagger$ are of zero order. In such contributions we shall, after normal ordering, restore any ϕ_α^* and ϕ_α factors coming from the definitions of A_α and $A_\alpha^\dagger$, provided that they have not disappeared in the contractions of $A_\alpha A_\alpha^\dagger$. Thus the true order with respect to ϕ_α and ϕ_α^* factors will be $\geq$ the nominal order defined by our iterative scheme.
[41]Recall that the contributions $a^\dagger\psi^\dagger\psi a$ start with terms of second order, and that there are fourth order terms of structure $a^\dagger\psi^\dagger\psi a$ in the contributions 04 and 40, which we are not evaluating.
[42]Uncontracted terms involving A and $A^\dagger$ factors are of fifth and higher orders.
[43]In fact, (C4) holds even if the ϕ_α are not eigenstates of $H(Xx)$. However, the proof then requires the more general expression (79) and is less trivial.

Borel summability and distribution

P. Hillion

Institut Henri Poincaré, Paris, France
(Received 18 October 1974)

This paper concerns the Borel summability of series expansions $\Sigma_0^\infty f_n\ \psi_n(x) \overset{B}{=} f(x)$, where the polynomials $\psi_n(x)$ defined on the interval I of the real line, are orthogonal in some space $L^2(I)$ and we look for conditions on f_n such that $f(x)$ is a distribution. As an application, a long standing ambiguity in the quantum theory of Coulomb scattering is solved.

I. INTRODUCTION

The Borel transform $f(z)=\beta\{f_n\}$ of a sequence $\{f_n\}$, $n=0,1,2,\cdots$, of real or complex numbers, is defined[1] by the relation:

$$f_n=\frac{1}{2\pi i}\int_\Gamma f(z)\,z^{n-1}\,dz \tag{1}$$

where the contour Γ encloses all the singularities of $f(z)$. Equation (1) can be written in a symbolic way[1]:

$$f_n h_n=f(z)\,\delta_{n0}, \tag{1'}$$

δ_{n0} being the Kronecker symbol, $h_n=I$ for $n\geq 0$, and $\beta\{h_n\}=h(z)=z/(z-1)$. $f(z)$ is the analytic continuation of the generating function $\sum_0^\infty f_n z^{-n}$, so the Borel transform enjoys all the well-known algebraic properties of generating functions; in particular

$$\beta(\{f_n\}*\{g_n\})=f(z)\,g(z)$$

where the symbol $*$ denotes a convolution product.

If $\lim_{z\to 1} f(z)$ exists, the series $\sum_0^\infty f_n$ is said B-summable and we write

$$\sum_0^\infty f_n \overset{B}{=} f(1); \tag{2}$$

of course, a convergent series is also B-summable with the same limit. Let $f(z)$, $g(z)$ be the Borel transforms of the sequences $\{f_n\}$, $\{g_n\}$, defined, respectively, for z outside some domain Ω_f, Ω_g, according to (1) and let Γ be a contour enclosing all the singularities of $p^{-1}f(p)g(z/p)$; if C denotes the infinite circle, then from Cauchy's theorem, it follows

$$\frac{1}{2\pi i}\int_\Gamma p^{-1}f(p)\,g\left(\frac{z}{p}\right)dp+\frac{1}{2\pi i}\int_C p^{-1}f(p)\,g\left(\frac{z}{p}\right)dp=0.$$

Provided that $p^{-1}f(p)$ and $g(z/p)$ have no common singularities, we can always find Γ_f enclosing the singularities of $p^{-1}f(p)$ and for $\hat{\Gamma}_g$ those of $g(z/p)$ such that $\int_\Gamma = \int_{\Gamma_f}+\int_{\hat{\Gamma}_g}$ and since $\int_{\Gamma_F}F(z)\,dz=-\int_{\hat{\Gamma}_F} z\,F(1/z)\,dz$ where Γ_F and $\hat{\Gamma}_F$ enclose, respectively, the singularities of $F(z)$ and those of $zF(1/z)$. The previous relation can be written in the form

$$\frac{1}{2\pi i}\int_{\Gamma_f}\frac{1}{p}f(p)\,g\left(\frac{z}{p}\right)dp+\frac{1}{2\pi i}\int_C\frac{1}{2p}f(p)\,g\left(\frac{z}{p}\right)dp$$

$$=\frac{1}{2\pi i}\int_{\Gamma_g}\frac{g(p)}{p}\,f\left(\frac{z}{p}\right)dp+\frac{1}{2\pi i}\int_C\frac{g(p)}{2p}\,f\left(\frac{z}{p}\right)dp.$$

If we note $f(z)\circledast g(z)$ and $g(z)\circledast f(z)$, respectively, the left- and right-hand side of this equation, we have

$$f(z)\circledast g(z)=g(z)\circledast f(z)$$

with

$$f(z)\circledast g(z)=\frac{1}{2\pi i}\int_{\Gamma_g}\frac{f(p)}{p}\,g\left(\frac{z}{p}\right)dp+\frac{1}{2\pi i}\int_C\frac{f(p)}{2p}\,g\left(\frac{z}{p}\right)dp \tag{3}$$

and similarly for $g(z)\circledast f(z)$. For Borel transforms, in all the cases that we have considered, the value of the integral along the infinite circle tends to zero.

Theorem 1:

$$f(z)\circledast g(z)=\beta\{f_n\,g_n\}. \tag{4}$$

Proof: From the equalities

$$f_n h_n=f(z)\,\delta_{n0}, \tag{5a}$$

$$g_n h_n=g(z)\,\delta_{n0}, \tag{5b}$$

$$h_n h_n\equiv h_n=h(z)\,\delta_{n0},\qquad h(z)=z/(z-1) \tag{6}$$

it is trivial to prove the relation $h(z)\circledast g(z)=g(z)=\beta\{g_n h_n\}$. This last result and Eq. (5b) lead to

$$h_n g_n h_n=g_n h_n=(h(z)\circledast g(z))\,\delta_{n0}=g(z)\,\delta_{n0};$$

so

$$f_n g_n h_n=(f_n h(z)\,\delta_{n0}\circledast g(z))$$

and with (6)

$$f_n g_n h_n=(f_n h_n\circledast g(z)).$$

Finally, using (5a)

$$f_n g_n h_n=(f(z)\circledast g(z))\,\delta_{n0};$$

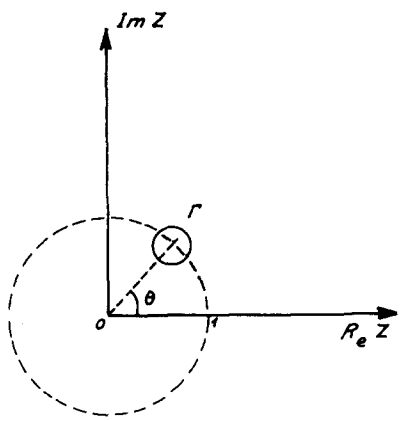


FIG. 1. The Fourier series.

TABLE I. Borel transforms $\psi(x, z)$ of sequences of polynomials $\psi(x)$ defined on the interval I of $\mathbb{R}$. [a]

	I	R	λ_n	$\psi_n(x)$	$\psi(x, z)$
Fourier	$(-\pi, \pi)$	$-iD$	n	$F_n(x) = e^{inx}$	$F(x, z) = \frac{z}{z - e^{ix}}$
Fourier	$(0, \pi)$	D^2	$-n^2$	$F_n(x) = \cos nx$	$F(x, z) = \frac{z(z - \cos x)}{z^2 - 2z\cos x + 1}$
Fourier	$(0, \pi)$	D^2	$-n^2$	$F_n(x) = \sin nx$	$F(x, z) = \frac{z \sin x}{z^2 - 2z\cos x + 1}$
Chebyschef	$(-1, 1)$	$(1-x^2)^{1/4} D (1-x^2)^{1/2} \times D (1-x^2)^{1/4}$	$-n^2$	$T_n(x) = \cos nx$	$T(x, z) = \frac{z(z - \cos x)}{z^2 - 2z\cos x + 1}$
Legendre	$(-1, 1)$	$D(x^2 - 1)D$	$n(n+1)$	$P_n(x) = \frac{1}{2^n} \frac{1}{n!} \frac{d^n}{dx^n} (x^2 - 1)^n$	$P(x, z) = \frac{z}{\sqrt{z^2 - 2z\cos x + 1}}$
Laguerre	$(0, \infty)$	$xD^2 + D - \frac{x}{4}$	$-n$	$\frac{L_n(x)}{n!} = \sum_0^\infty \binom{n}{r} \frac{(-x)^r}{r!}$	$L(x, z) = \frac{z}{z-1} e^{-x/(z-1)}$
Hermite	$(-\infty, \infty)$	$D^2 - x^2 + 1$	$-2n$	$\frac{H_n(x)}{n!} = \sum_0^{[n/2]} \frac{(-1)^{n-k}}{k!} \frac{x^{2n-2k}}{(n-2k)!\, 2^k}$	$H(x, z) = e^{-x/z - 1/2z^2}$
Gegenbauer	$(-1, 1)$	$\frac{D}{W(x)} (1-x^2)^{\alpha+1/2} \frac{D}{W(x)}$ $W(x) = (1-x^2)^{\alpha/2 - 1/4}$	$-n(n+2\alpha)$	$C_n^\alpha(x) = \frac{\Gamma(n+2\alpha)}{\Gamma(n+\alpha)\Gamma(2\alpha)} \times {}_2F_1\left(n+2\alpha, -n, \frac{\alpha+1}{2}; \frac{1-x}{2}\right)$	$G^\alpha(x, z) = \frac{z^{2\alpha}}{(z^2 - 2z\cos x + 1)^\alpha}$

[a] $D = d/dx$, Q is the differential operators with eigenvectors $\psi_n(x)$ and eigenvalues λ_n. In the definition of Gegenbauer polynomials, ${}_2F_1$ is the hypergeometric function.

that is, $f(z) \circledast g(z) = \beta\{f_n g_n\}$. This completes the proof.

Of course, $\beta\{f_n g_n\}$ is computed with the easier of both expressions: $f(z) \circledast g(z)$ or $g(z) \circledast f(z)$. An immediate corollary follows

Corollary:

$$f(z) = \frac{1}{2\pi i} \int_{\Gamma_f} \frac{1}{p} f(p) \frac{z}{z-p}\, dp = \frac{1}{2\pi i} \int_{\Gamma_{f_1}} \frac{1}{p-1} f\left(\frac{z}{p}\right) dp.$$

Γ_h enclosing the point 1 in the complex plane. It is easy to show that in this case the integral along C is zero since for $p \to \infty$, $f(p) = \sum_0^\infty (f_n/p^n)$ and for $|z/p| < 1$, $z/(z-p) = -(z/p)(1 + z/p + z^2/p^2 + \cdots)$. Theorem I and Eq. (2) lead to

$$\sum_0^\infty f_n g_n \overset{B}{=} \lim_{z \to 1} (f(z) \circledast g(z)). \tag{7}$$

Jury[2] has proved Theorem I when $f(z)$ is a meromorphic function and the integral along C zero. He has also given, in this special case, many applications of (7) to the summation of series.

2. ORTHONORMAL B-SUMMABLE SERIES EXPANSIONS

Let $\{\psi_n(x)\}$ be a sequence of polynomials defined on an open interval $I = (a, b)$ of the real line $\mathbb{R}$ with the Borel transform $\psi(x, z) = \beta\{\psi_n(x)\}$.

From now on, we assume that in Eq. (3) the second term is zero (from a practical point of view, it is not a real restriction); then relation (7) for $x \in I$ leads to

$$\sum_0^\infty f_n \psi_n(x) \overset{B}{=} f(x),$$

$$f(x) = \lim_{z \to 1} \frac{1}{2\pi i} \int_{\Gamma_f} \frac{1}{p} f(p) \psi(x, z/p)\, dp. \tag{8}$$

[We assume that $(1/2\pi i) \int_{\Gamma_f} (1/p) f(p) \psi(x, z/p)\, dp$ is holomorphic bounded in the open disk $|z^{-1}| < 1$ so that from a well-known theorem[7] the radial limit exist.] Now the question is: What does the equality (8) mean?

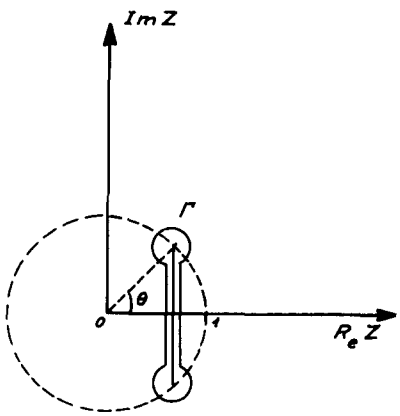


FIG. 2. The Chebyschev, Legendre, and Gegenbauer polynomials.

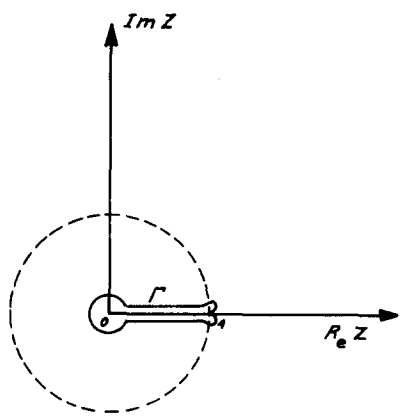


FIG. 3. The Laguerre polynomials.

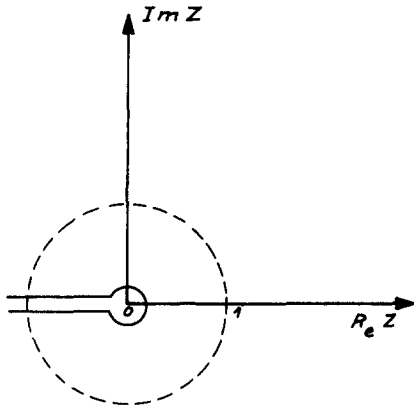


FIG. 4. The Hermite polynomials.

Before answering, we have a look at orthonormal series expansions in distribution theory but first we give in Table I the Borel transform of the most used families of polynomials and Figs. 1 to 4 show the contour Γ_f of definition (1). Γ is the same for Chebyschev, Legendre, Gegenbauer polynomials while for Laguerre polynomials Γ is such that one can reach the point I only along the real axis; functions $\psi(x, z)$ are analytic outside Γ for $x \in I$.

Zemanian[3] seems to be the first to have studied the expansion of distributions into orthogonal polynomials. Here, we only give the main results useful later, without demonstration; but the interested reader can find all the proofs in Zemanian's monograph.[3]

He introduces a particular testing function space $\mathcal{A}$ as the set of functions $\varphi(x)$ with the three following properties:

(i) $\varphi(x)$ is defined, complex valued, continuously differentiable on I;

(ii) for each integer k, the quantity $\alpha_k(\varphi) = [\int_a^b |R^k \varphi(x)|^2 dx]^{1/2}$ exists, R being a differential operator whose eigenvectors are the polynomials $\psi_n(x)$. The expression of R and the corresponding eigenvalues λ_n are also given in Table I,

(iii) for each integer n, k, $(R^k \varphi(x), \psi_n(x)) = (\varphi(x), R^k \psi_n(x))$ with $(f(x), g(x)) = \int_I f(x)\overline{g(x)}\,dx$, where $\overline{g(x)}$ denotes the complex conjugate of $g(x)$.

$\mathcal{A}$ is a linear space with the system of seminorms $\{\alpha_k\}$. Equipped with the topology generated by $\{\alpha_k\}_0^\infty$, $\mathcal{A}$ becomes a countably multinormed space; it is complete and therefore a Fréchet space.

Lemma 1: If $\varphi \in \mathcal{A}$, then $\varphi(x) = \sum_0^\infty (\varphi(x), \psi_n(x))\,\psi_n(x)$ where the series converges in $\mathcal{A}$.

Lemma 2: Let a_n denote complex numbers. Then $\sum_0^\infty a_n \psi_n(x)$ converges if and only if $\sum_0^\infty |\lambda_n|^{2k} |a_n|^2$ converges for every nonnegative integer k.

Lemma 3: $\mathcal{A}$ is a subspace of $L^2(I)$ and $E(I)$ (space of complex, continuously differentiable functions on I) and $D(I) \subset \mathcal{A} \subset E(I)$.

Let $\mathcal{A}'$ be the dual of $\mathcal{A}$ and $\langle f(x), \overline{\varphi(x)}\rangle$ be the number that $f \in \mathcal{A}'$ assigns to $\varphi \in \mathcal{A}$ [so that, if $f \in L^2(I)$, then $\langle f(x), \overline{\varphi(x)}\rangle = (f(x), \varphi(x))$]. $\mathcal{A}'$ has the three following properties:

(1) $E'(I) \subset \mathcal{A}' \subset D'(I)$;

(2) $\mathcal{A} \subset L^k(I) \subset \mathcal{A}'$;

(3) for each $f \in \mathcal{A}'$, there exists a nonnegative real integer and a positive constant c such that

$$|\langle f(x), \overline{\varphi(x)}\rangle| \leq c \max_{0 \leq k \leq r} d_k(\varphi) \quad \forall \varphi \in \mathcal{A},$$

where r and c depend on f but not on φ. Then, Zemanian[3] proves the following theorems:

Theorem 2: If $f \in \mathcal{A}'$, then the series $\sum_0^\infty \langle f(x), \overline{\psi_n(x)}\rangle \psi_n(x)$ converges in $\mathcal{A}'$.

Theorem 3: Let $\{f_n\}$ be a sequence of complex numbers, then $\sum_0^\infty f_n \psi_n(x)$ converges in $\mathcal{A}'$, if and only if there exists a nonnegative integer φ such that $\sum_{\lambda_n \neq 0} |\lambda_n|^{-2\varphi} |f_n|^2$ converges. Furthermore, if $f(x)$ denotes the sum in $\mathcal{A}'$ of the series, then $f_n = \langle f(x), \overline{\psi_n(x)}\rangle$.

For the polynomials in Table I (see the eigenvalues λ_n) Theorem 3 gives:

Corollary: $\sum_0^\infty f_n \psi_n(x)$ converges in $\mathcal{A}'$ if and only if there exists an integer k such that $\lim_{n\to\infty} f_n/n^k \to 0$.

For trigonometric series, this result was proved by Schwartz.[4] From now on, when $\sum_0^\infty f_n \psi_n(x)$ converges in $\mathcal{A}'$, this is, when $\langle f(x), \overline{\psi}(x)\rangle = \sum_0^\infty f_n(\psi_n(x), \varphi(x))$ for any $\varphi \in \mathcal{A}$, we write

$$f(x) \overset{\mathcal{A}'}{=} \sum_0^\infty f_n \psi_n(x).$$

So, the question following relation (8) can be put more precisely: Under what conditions does

$$f(x) \overset{B}{=} \sum_0^\infty f_n \psi_n(x)$$

belong to $\mathcal{A}'$?

3. DISTRIBUTIONS AND BOREL SUMMABILITY

First, we introduce the following theorems for $\chi(x, z) = \mathcal{B}\{\chi_n(x)\}$, that is,

$$\chi_n(x) = \frac{1}{2\pi i} \int_\Gamma \chi(x, z)\, z^{n-1}\, dz, \quad x \in I.$$

(a) If $\lim_{x \to x_0} \chi(x, z) = \chi(x_0, z)$ for every $z \in \Gamma$, if there exists $g_\Gamma \in L(\Gamma)$ such that $|\chi(x, z)| \leq g_\Gamma(z)$ for every $x \in I$ and $z \in \Gamma$; then from Lebesgue's theorem it follows that

$$\lim_{x \to x_0} \chi_n(x) = \frac{1}{2\pi i} \int_\Gamma \lim_{x \to x_0} \chi(x, z)\, z^{n-1}\, dz, \quad n \geq 0,$$

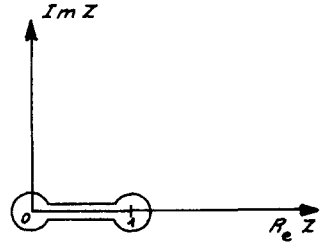


FIG. 5.

or

$$\mathcal{B}\{\lim_{x\mapsto x_0} \chi_n(x)\} = \lim_{x\mapsto x_0} \chi(x,z). \tag{9a}$$

(b) If $\int_\Gamma |\chi(x,z)\, z^{n-1}|\, dz$ is bounded for almost every $x \in I$, if $\int_I dx \int_\Gamma |\chi(x,z) z^{n-1}|\, dx$ is also bounded, then Fubini's theorem implies

$$\int_I \chi_n(x)\, dx = \frac{1}{2\pi i} \int_\Gamma dz\; z^{n-1} \int_I \chi(x,z)\, dx;$$

that is

$$\mathcal{B}\{\int_I \chi_n(x)\, dx\} = \int_I \chi(x,z)\, dx. \tag{9b}$$

(c) If for every $x \in I$, $\chi(x,z)\, z^{n-1}$ is summable on Γ, if for every $z \in \Gamma$, $\chi(x,z)$ is differentiable, and if there exists $g_\Gamma(z) \in L^1(\Gamma)$ such that $|\partial\chi(x,z)/\partial x| \leq g_\Gamma$ for every $x \in I$ and every $z \in \Gamma$, then

$$\frac{\partial \chi_n(x)}{\partial x} = \frac{1}{2\pi i} \int_\Gamma \frac{\partial}{\partial \chi}\, \chi(x,z)\, z^{n-1}\, dz,$$

that is,

$$\mathcal{B}\left\{\frac{\partial}{\partial x}\chi_n(x)\right\} = \frac{\partial}{\partial x}\, \chi(x,z). \tag{9c}$$

Remark 1: Since Γ is arbitrary in the analytic domain Ω of $\chi(x,z)$, the conditions which make valid the three previous theorems are to be fulfilled for any Γ in Ω.

Remark 2: If we consider $\chi(x,z)\,\bar{\xi}(x)$ instead of $\chi(x,z)$, relation (9b) gives [it is assumed that $\chi(x,z)\,\bar{\xi}(x)$ satisfies all the necessary conditions] the following relation, useful later:

$$\mathcal{B}\{\int_I \chi_n(x)\,\bar{\xi}(x)\, dx\} = \int_I \chi(x,z)\,\bar{\xi}(x)\, dx. \tag{9d}$$

Let us now consider

$$\chi(x) = \lim_{z\mapsto 1} \chi(x,z) \quad \text{and} \quad \chi_1(x) = \lim_{z\mapsto 1} \frac{\partial}{\partial x}\, \chi(x,z).$$

Lemma 4: If $|\chi(x,z)|$ and $|(\partial/\partial x)\chi(x,z)|$ are bounded by some integrable functions on I when $z \mapsto 1$, then $\chi_1(x) = (\partial/\partial x)\chi(x)$ in $\mathcal{A}'$.

Indeed, for every $\varphi \in \mathcal{A}$

$$\int_I \varphi(x)\chi_1(x)\, dx = \int_I \varphi(x) \lim_{z\to 1} \frac{\partial}{\partial x}\, \chi(x,z)\, dx$$

$$= -\int_I \varphi'(x) \lim_{z\to 1} \chi(x,z)\, dx,$$

since from the assumptions one may exchange limit and integration.

Corollary: Any B-summable series satisfying the conditions of Lemma 4 can be differentiated term by term in $\mathcal{A}'$ if the derived series is B-summable. In fact, we have

$$\chi(x) = \lim_{z\mapsto 1} \sum_0^\infty \chi_n(x)\, z^{-n}, \quad \chi_1(x) = \lim_{z\mapsto 1} \sum_0^\infty \chi_n'(x)\, z^{-n}$$

[in this last case provided that relation (9c) holds] and from Lemma 4 $\chi_1(x) = \chi'(x)$.

We now discuss the consequences of relation (9d). For $|z| > \sigma$, $\sigma = \sup_{x\in I}^n \sqrt{|f_n \psi_n(x)|}$, $f(z) \circledast \psi(x,z)$ has the power series expansion $\sum_0^\infty f_n \psi_n(x)\, z^{-n}$, so that from (9d) it follows that

$$\sum_0^\infty \frac{f_n}{z_n} (\psi_n(x), \psi_m(x)) = \int_I f(z) \circledast \psi(x,z)\, \bar{\psi}_m(x)\, dx;$$

that is, since $(\psi_n(x),\ \psi_m(x)) = \delta_{nm}$,

$$\frac{f_n}{z_n} = \int_I f(z) \circledast \psi(x,z)\, \bar{\psi}_n(x)\, dx$$

and

$$f_n = \lim_{z\mapsto 1} \int_I f(z) \circledast \psi(x,z)\, \bar{\psi}_n(x)\, dx.$$

Now if (i) $|f(z) \circledast \psi(x,z)|$ is bounded by an integrable function on I when $z \mapsto 1$, then from the Lebesgue theorem

$$f_n = \int_I \lim_{z\mapsto 1} f(z) \circledast \psi(x,z)\, \bar{\psi}_n(x)\, dx = \int_I f(x)\, \bar{\psi}_n(x)\, dx; \tag{10}$$

(ii) $\sum_0^\infty f_n^2$ converges, then from the Riesz–Fisher theorem

$$\sum_0^\infty f_n \psi_n(x) \overset{L^2(I)}{=} f_1(x)$$

and

$$f_n = \int_I f_1(x)\, \bar{\psi}_n(x)\, dx. \tag{10'}$$

The comparison between (10) and (10') gives $f_1(x) = f(x)$, so we can state:

Theorem 4: If conditions (i), (ii) hold, then

$$\sum_0^\infty f_n \psi_n(x) \overset{B}{=} f(x) \in L^2(I).$$

Remark: In the following, we use the fact that the generalized differential operator $\bar{R}'$ on $\mathcal{A}'$ defined through the relationship $\langle f, \overline{R\varphi}\rangle = \langle \overline{R'f},\ \bar{\varphi}\rangle$ satisfies $R = R'$ (see Ref. 3).

Using Zemanian's theorem that a necessary and sufficient condition for f to belong to $\mathcal{A}'$ is that there exists some nonnegative integer k and $g \in L^2(I)$ such that $f = R^k g + \sum_{\lambda_n=0} c_n \psi_n$ where the c_n denote complex constants, we can state:

Theorem 4': If

$$f(x) \overset{B}{=} \sum_0^\infty f_n \psi_n(x)$$

and

$$f(x) = R^k g(x) + \sum_{\lambda_n=0} c_n \psi_n(x), \quad g \in L^2(I),$$

then $f(x) \in \mathcal{A}'$ and from Theorem 2

$$f_n = \langle f(x),\ \bar{\psi}_n(x)\rangle. \tag{11}$$

Theorem 5: If $f(x) \overset{B}{=} \sum_0^\infty f_n \psi_n(x)$ and Eq. (11) holds, then $f \in \mathcal{A}'$.

Indeed, Eq. (11) defines $f(x)$ as a continuous linear functional on the linear subspace $\hat{\mathcal{A}}$ of $\mathcal{A}$: $\hat{\mathcal{A}} = \{\varphi(x) : \varphi(x) = \sum_0^N a_n \varphi_n(x)\}$; so by the Hahn–Banach theorem, f can be extended onto all of $\mathcal{A}$.

Theorem 6: If

$$f(x) \overset{\mathcal{A}'}{=} \sum_0^\infty f_n\, \psi_n(x)$$

and

$$f_1(x) \overset{B}{=} \sum_0^\infty f_n \psi_n(x)$$

with

$$f_n = \langle f_1(x), \bar{\psi}_n(x) \rangle,$$

then

$$f(x) \overset{\mathcal{A}'}{=} f_1(x).$$

From Theorem 3 $f_n = \langle f(x), \bar{\psi}_n(x) \rangle$ and from Theorem 5 $f_1(x) \in \mathcal{A}'$, so

$$f(x) - f_1(x) \overset{\mathcal{A}'}{=} \sum_0^\infty \langle f(x) - f_1(x), \bar{\psi}_n(x) \rangle \psi_n(x)$$

$$\overset{\mathcal{A}'}{=} \sum_0^\infty (\langle f(x), \bar{\psi}_n(x) \rangle - \langle f_1(x), \bar{\psi}_n(x) \rangle) \psi_n(x)$$

$$= 0.$$

Theorem 6 does not imply $f(x) = f_1(x)$, an interesting counter example is given in Sec. 4.

Let us now give some examples, considering first trigonometric series. The Borel transforms of $\{e^{in\theta}\}$, $\{\cos n\theta\}$, $\{\sin n\theta\}$ are given in Table I and, besides, we have

$$\mathcal{B}\left\{\frac{1}{n}\cos n\theta\right\} = \log \frac{z}{(z^2 - 2z\cos\theta + 1)^{1/2}},$$

$$\mathcal{B}\left\{\frac{1}{n}\sin n\theta\right\} = \theta + \arctan \frac{\sin\theta}{z - \cos\theta}, \quad \theta > 0,$$

$$\mathcal{B}\{n\cos n\theta\} = \frac{z[(1+z^2)\cos\theta - 2z]}{(z^2 - 2z\cos\theta + 1)^2},$$

$$\mathcal{B}\{n\sin n\theta\} = \frac{z(z^2-1)\sin\theta}{(z^2 - 2z\cos\theta + 1)^2}.$$

It follows at once for $-\pi < \theta < \pi$ that

$$\sum_1^\infty \frac{\cos n\theta}{n} \overset{B}{=} -\log\sin\theta - \log 2, \quad \sum_1^\infty \sin n\theta \overset{B}{=} \tfrac{1}{2}\operatorname{cotan}\frac{\theta}{2},$$

$$\sum_1^\infty n\cos n\theta \overset{B}{=} \frac{-1}{4\sin^2(\theta/2)} \tag{12a}$$

$$\sum_1^\infty \frac{\sin n\theta}{n} \overset{B}{=} \frac{\pi}{2} - \frac{\theta}{2}, \quad \sum_1^\infty \cos n\theta \overset{B}{=} -\tfrac{1}{2}, \quad \sum_1^\infty n\sin n\theta \overset{B}{=} 0. \tag{12b}$$

These series divide into two classes according whether $f(\theta)$ is periodic (first line) or not (second line). In this last case, the sum in the left-hand side is not $f(\theta)$, but rather a periodic distribution $\hat{f}(\theta)$ which coincides with $f(\theta)$ in the period $(-\pi, \pi)$. In the previous examples, it is easy to obtain $\hat{f}(\theta)$:

$$\sum_1^\infty \frac{\sin n\theta}{n} \overset{B}{=} \hat{f}(\theta) = \begin{cases} \dfrac{\pi}{2} - \dfrac{\theta}{2}, & -\pi < \theta < \pi, \\ \dfrac{\pi}{2} - \dfrac{\theta}{2} - m\pi, & (2m-1)\pi < \theta < (2m+1)\pi; \end{cases}$$

that is,

$$\sum_1^\infty \frac{\sin n\theta}{n} \overset{B}{=} \frac{\pi}{2} - \frac{\theta}{2} - \pi\sum_1^\infty H\{\theta - (2m-1)\pi\}$$

$$+ \pi\sum_{-\infty}^0 H\{(2m-1)\pi - \theta\} \tag{12a'}$$

where $H(\theta)$ is the Heaviside function. Taking the derivative term by term of this expression gives, according to the corollary of Lemma 4,

$$\sum_1^\infty \cos n\theta \overset{B}{=} -\tfrac{1}{2} + \pi\,\delta_{2\pi}(\theta) \tag{12b'}$$

with $\delta_{2\pi}(\theta) = \sum_{-\infty}^{\infty} \delta(\theta + 2\pi m)$ where δ is the Dirac distribution. In the same way, with $\delta'_{2\pi}(\theta) = (\partial/\partial\theta)\,\delta_{2\pi}(\theta)$,

$$\sum_1^\infty n\sin n\theta \overset{B}{=} \pi\,\delta'_{2\pi}(\theta). \tag{12c'}$$

Relations (12b) and (12') differ by some linear combinations of Dirac distribution and of its derivatives at the ends of the interval I. More important, with (12') but not with (12b), relation (11) is valid, so that from Theorem 5 $\hat{f}(\theta) \in \mathcal{A}'$ but not $f(\theta)$.

We generalize this result: Since $f(x)$ is defined only on the open interval I, one can always consider an arbitrary linear combination of $f(x)$ and of Dirac distribution and its derivatives at the ends of I. From now on we call Borel sum this particular combination $\hat{f}(x)$ when it exists which makes relation (11) valid, so that from Theorem 5 $\hat{f}(x) \in \mathcal{A}'$. Not any B-summable series

$$f(x) \overset{B}{=} \sum_0^\infty f_n \psi_n(x)$$

can satisfy (11), for instance, if f_n does not fulfill conditions of Theorem 3.

It is trivial to show that for expansions (12a), relation (11) holds. For instance, from

$$\sum_1^\infty n\cos n\theta \overset{B}{=} -\left(4\sin^2\frac{\theta}{2}\right)^{-1},$$

it follows that (where f. p. means "finite part")

$$n = \left\langle -\frac{1}{4\sin^2(\theta/2)}, \cos n\theta \right\rangle = \text{f. p.}\left\langle -\frac{1}{4\sin^1(\theta/2)}, \cos n\theta \right\rangle$$

$$= -\lim_{\epsilon\to 0}\left(\frac{1}{2}\int_\epsilon^\pi \frac{\cos n\theta}{1-\cos\theta}\,d\theta \quad -\frac{1}{\epsilon}\right)$$

In the same way,

$$\sum_1^\infty n^2\sin n\theta \overset{B}{=} -\frac{1}{4}\frac{\cos\theta/2}{\sin^3\theta/2}$$

leads to

$$n^2 = \left\langle -\frac{1}{4}\frac{\cos(\theta/2)}{\sin^3(\theta/2)}, \sin n\theta \right\rangle = \text{f. p.}\left\langle -\frac{1}{4}\frac{\cos(\theta/2)}{\sin^3(\theta/2)}, \sin n\theta \right\rangle$$

$$= -\lim_{\epsilon\to 0}\left(\frac{1}{2}\int_\epsilon^\pi \frac{\cos(\theta/2)\sin n\theta}{1-\cos\theta}\,d\theta - \frac{2}{\epsilon}\right).$$

Let us now give some results using Legendre polynomials. From the Borel transform of $\{P_n(\cos\theta)\}$ in Table I and from the following equality[1,2] $\mathcal{B}\{n\psi_n(x)\} = -z(\partial/\partial z)\,\psi(x, z)$ we obtain for $x = \cos\theta \in (-1, 1)$

$$\sum_0^\infty P_l(x) \overset{B}{=} \frac{1}{\sqrt{2(1-x)}}, \quad \sum_0^\infty l P_l(x) \overset{B}{=} \frac{-1}{2\sqrt{2(1-x)}},$$

$$\sum_0^\infty (2l+1) P_l(x) \overset{B}{=} 0;$$

but to make (11) valid, these relations have to be changed into

$$\sum_0^\infty P_l(x) \overset{B}{=} \frac{1}{\sqrt{2(1-x)}}, \quad \sum_0^\infty l P_l(x) \overset{B}{=} \frac{-1}{2\sqrt{2(1-x)}} + \delta(1-x),$$

$$\sum_0^\infty (2l+1) P_l(x) \overset{B}{=} 2\,\delta(1-x).$$

To sum up, if the Borel sum is correctly defined at the ends of the interval, that is, provided that Eq. (11) can be satisfied, two different B-summable series $\sum_0^\infty f_n \psi_n(x)$ have different limits which belong to $\mathcal{A}'$.

4. APPLICATION

Zemanian[3] has given many applications of the distribution theory in $\mathcal{A}'$ for solving some partial differential equations. Here we intend to give an application with a view to solve an ambiguity in the physical theory of scattering.

First we look for the Borel transform of the sequence of hypergeometric functions $\{{}_2F_1(-l,\ 2i\alpha,\ 1+i\alpha;\ 1)\}$ where l is a nonnegative integer and α an arbitrary real number. Using the gamma function (Ref. 5, p. 67), one gets

$${}_2F_1(-l,\ 2i\alpha,\ 1+i\alpha;\ 1) = \frac{\Gamma(1+i\alpha)\Gamma(l+1-i\alpha)}{\Gamma(1-i\alpha)\Gamma(l+1+i\alpha)},$$

so that the relation $\Gamma(1+z) = z\Gamma(z)$ leads to

$$\sum_0^\infty {}_2F_1(-l,\ 2i\alpha,\ 1+i\alpha;\ 1)\frac{t^l}{l!} = 1 + \frac{1-i\alpha}{1+i\alpha}\,t + \frac{(1-i\alpha)(2-i\alpha)}{(1+i\alpha)(2+i\alpha)} \times \frac{t^2}{z!} + \cdots$$

$$= {}_1F_1(1-i\alpha,\ 1+i\alpha;\ t) \quad \text{(Ref. 5, p. 248)}$$

where $[{}_1F_1]$ is the confluent hypergeometric function. But in the Borel's polygon of summability, we have[1]

$$\mathcal{B}\{{}_2F_1(-l,\ 2i\alpha,\ 1+i\alpha;\ 1)\}$$

$$= \int_0^\infty e^{-t}\,{}_1F_1(1-i\alpha,\ 1+i\alpha;\ t/z)\,dt$$

$$= {}_2F_1(1-i\alpha,\ 1,\ 1+i\alpha;\ z^{-1}), \quad |z| > 1 \qquad \text{(Ref. 5, p. 269).}$$

The analytic continuation of this result is obtained through the two relations (Ref. 5, pp. 59, 104)

$$c\left(1-\frac{1}{z}\right){}_2F_1\left(a, b, c; \frac{1}{z}\right) - c\,{}_2F_1\left(a, b-1, c; \frac{1}{z}\right) + \frac{c-a}{z} \times {}_2F_1\left(a, b, c+1; \frac{1}{z}\right) = 0,$$

$${}_2F_1\left(a, b, c; \frac{1}{z}\right) = \frac{\Gamma(c)}{\Gamma(a)\,\Gamma(c-a)} \int_0^1 t^{a-1}(1-t)^{c-a-1} \times \left(1-\frac{t}{z}\right)^{-b} dt, \quad \mathrm{Re}\,a > 0, \quad \mathrm{Re}(c-a) > 0,$$

with $a = 1-i\alpha$, $b = 1$, $c = 1+i\alpha$, so that, finally

$$\mathcal{B}\{{}_2F_1(-l, 2i\alpha, 1+i\alpha;1)\} = \frac{1}{z-1}\left(z - \frac{2i\alpha}{(1+i\alpha)B(1-i\alpha, 1+2i\alpha)} \int_0^1 \frac{t^{-i\alpha}(1-t)^{2i\alpha}dt}{1-t/z}\,dt\right). \qquad (13)$$

This function is analytic in the complex plane cut along $(0, 1)$. Using (13), it is easy to verify relation (1) on the contour of the Fig. 5. Let us notice that Eq. (13) yields the known result (Ref. 5, p. 86)

$$\sum_0^\infty {}_2F_1(-l, 2i\alpha, 1+i\alpha;1) \overset{B}{=} i\alpha(2i\alpha-1)^{-1}.$$

With this previous result and the Borel transform of $\{P_l(\cos\theta)\}$ (Table I), we prove in the Appendix the following equality:

$$\mathcal{B}\{(2l+1)\,{}_2F_1(-l, 2i\alpha, 1+i\alpha; 1)P_l(\cos\theta)\} = \lim_{\epsilon\to 0}\lambda_\epsilon(\alpha)\int_0^1 \frac{t^{-i\alpha}(1-t)^{\epsilon+2i\alpha-1}}{(1-2t\cos\theta+t^2/z^2)^{3/2}}\left(1-\frac{t^2}{z^2}\right)dt$$

where ϵ is a small positive number and

$$\lambda_\epsilon(\alpha) = \frac{\Gamma(1+\epsilon+i\alpha)}{\Gamma(1-i\alpha)\Gamma(\epsilon+2i\alpha)}.$$

Exchanging $\lim_{z\mapsto 1}$ and integration, since the absolute value of the integrand is bounded on $(0, 1)$ for $z \mapsto 1$, leads to

$$\sum_0^\infty (2l+1)\,{}_2F_1(-l, 2i\alpha, 1+i\alpha; 1)P_l(\cos\theta) \overset{B}{=} \lim_{\epsilon\mapsto 0}\lambda_\epsilon(\alpha)\int_0^1 \frac{t^{-i\alpha}(1-t)^{\epsilon+2i\alpha-1}(1-t^2)}{(1-2t\cos\theta+t^2)^{3/2}}\,dt. \qquad (14)$$

In the Appendix it is also proved that relation (11) holds in the form

$${}_2F_1(-l, 2i\alpha, 1+i\alpha; 1) = \lim_{\epsilon\mapsto 0}\lambda_\epsilon(\alpha)\,\text{f.p.}\int_{-1}^1 P_l(\cos\theta)\,d\cos\theta \times \int_0^1 \frac{t^{-i\alpha}(1-t)^{\epsilon+i\alpha-1}(1-t^2)}{(1-2t\cos\theta+t^2)^{3/2}}\,dt.$$

So the right-hand side of (14) belongs to $\mathcal{A}'$.

We now prove that we also have

$$\sum_0^\infty \frac{(2l+1)}{2}\,{}_2F_1(-l, 2i\alpha, 1+i\alpha; 1)P_l(\cos\theta) \overset{\mathcal{A}'}{=} \frac{i\alpha}{2^{i\alpha}}(1-\cos\theta)^{i\alpha-1}. \qquad (15)$$

From Theorem 3, convergence can only be in $\mathcal{A}'$, so that Eq. (15) holds if

$${}_2F_1(-l, 2i\alpha, 1+i\alpha; 1) = \frac{i\alpha}{2^{i\alpha}}\,\text{f.p.}\int_{-1}^1 (1-\cos\theta)^{i\alpha-1} \times P_l(\cos\theta)\,d(\cos\theta). \qquad (16)$$

Now equality (16) is trivial for $l = 0$, $l \pm I$, and we prove that if it is true for $l-1$, l, it also holds for $l+I$. Indeed, from (16) and from the well-known relation ($x = \cos\theta$)

$$P_{l+1}(x)=\frac{2l+1}{l+1}\times P_l(x)-\frac{l}{l+1}P_{l-1}(x),$$

we deduce (writing to simplify in the left-hand side ${}_2F_1(-l-1, 2i\alpha, 1+i\alpha; 1)$ instead of

$$\frac{i\alpha}{2^{i\alpha}}\,\text{f.p.}\int_{-1}^{1}(1-\cos\theta)^{i\alpha-1}P_{l+1}(\cos\theta)\,d\cos\theta)$$

that

$${}_2F_1(-l-1, 2i\alpha, 1+i\alpha; 1)$$
$$=\frac{2l+1}{l+1}\,{}_2F_1(-l, 2i\alpha, 1+i\alpha; 1)-\frac{l}{l+1}$$
$$\times{}_2F_1(-l+1, 2i\alpha, 1+i\alpha; 1)$$
$$-\frac{2l+1}{l+1}\,\frac{i\alpha}{2^{i\alpha}}\int_{-1}^{1}(1-x)^{i\alpha}P_l(x)\,dx.$$

The last term can be written in the form

$$-\frac{2l+1}{l+1}\,\frac{2i\alpha}{1+i\alpha}\,{}_2F_1(-l, 2(1+i\alpha), 2+i\alpha; 1)$$

and thus

$${}_2F_1(-l-1, 2i\alpha, 1+i\alpha; 1)$$
$$=\frac{2l+1}{l+1}\,{}_2F_1(-l, 2i\alpha, 1+i\alpha; 1)-\frac{l}{l+1}$$
$$\times{}_2F_1(-l+1, 2i\alpha, 1+i\alpha; 1)-\frac{2l+1}{l+1}\,\frac{2i\alpha}{1+i\alpha}$$
$$\times{}_2F_1(-l, 2(1+i\alpha), 2+i\alpha; 1).$$

Using gamma functions and the relation $\Gamma(1+z)=z\,\Gamma(z)$, leads to

$${}_2F_1(-l-1, 2i\alpha, 1+i\alpha; 1)$$
$$=\frac{\Gamma(1+i\alpha)}{\Gamma(1-i\alpha)}\left\{\frac{2l+1}{l+1}\,\frac{\Gamma(l+1-i\alpha)}{\Gamma(l+1+i\alpha)}-\frac{l}{l+1}\,\frac{\Gamma(l-i\alpha)}{\Gamma(l+i\alpha)}\right.$$
$$\left.-\frac{2l+1}{l+1}\,2\alpha^2\,\frac{\Gamma(l-i\alpha)}{\Gamma(l+2+i\alpha)}\right\}$$
$$=\frac{1}{l+1}\,\frac{\Gamma(1+i\alpha)}{\Gamma(1-i\alpha)}\,\frac{\Gamma(l+1-i\alpha)}{\Gamma(l+1+i\alpha)}$$
$$\times\frac{(2l+1)(l^2+l-\alpha^2-i\alpha)-l(l^2+2il\alpha-\alpha^2+l+i\alpha)}{(l-i\alpha)(l+1+i\alpha)}$$
$$=\frac{\Gamma(1+i\alpha)}{\Gamma(1-i\alpha)}\,\frac{\Gamma(l+1-i\alpha)}{\Gamma(l+1+i\alpha)}\,\frac{l+1-i\alpha}{l+1+i\alpha}$$
$$=\frac{\Gamma(1+i\alpha)}{\Gamma(1-i\alpha)}\,\frac{\Gamma(l+2-i\alpha)}{\Gamma(l+2+i\alpha)}.$$

This completes the proof.

From (14) and (15), according to Theorem 6, it follows that

$$\frac{i\alpha}{2^{i\alpha}}(1-\cos\theta)^{i\alpha-1}\overset{\mathcal{A}'}{=}\lim_{\epsilon\to0}\lambda_\epsilon(\alpha)$$
$$\times\int_0^1\frac{t^{-i\alpha}(1-t)^{\epsilon+2i\alpha-1}(1-t^2)}{(1-2t\cos\theta+t^2)^{3/2}}\,dt. \qquad (17)$$

Now let us come to the physical problem of writing Eq. (15) in the form

$$\alpha(1-\cos\theta)^{i\alpha-1}\,\frac{\Gamma(1-i\alpha)}{\Gamma(1+i\alpha)}$$
$$\overset{\mathcal{A}'}{=}-i\,\frac{2^{i\alpha}}{2}\sum_0^\infty(2l+1)\,\frac{\Gamma(l+1-i\alpha)}{\Gamma(l+1+i\alpha)}\,P_l(\cos\theta). \qquad (15')$$

Physicists were also to prove that the left-hand side of (15′) is a solution, in spherical coordinates, for the scattering amplitude in a Coulomb potential, while the right-hand side is the solution written as a sum of spherical harmonics. Erroneously, some of them[6] inferred the equality of both sides. Moreover, those who used the Borel summability method were unable to obtain the left-hand side of (15′); the reason why is now clear: They could have only found the right-hand side of (14) and from (17) equality holds only in $\mathcal{A}'$.

5. CONCLUSION

The Borel summability appears as an interesting tool for obtaining a class of distributions with orthonormal series expansions. Because of the importance in applied mathematics and physics of distributions on the one hand and of orthogonal polynomials on the other hand, one can think that this method will have more and more application.

APPENDIX

To simplify computations, we do not use relation (13) but since ${}_2F_1(1-i\alpha, 1, 1+i\alpha; z^{-1})$ $=\lim_{\epsilon\to0}{}_2F_1(1-i\alpha, 1, 1+\epsilon+i\alpha; z^{-1})$ where ϵ is an arbitrary small positive number, we can write (Ref. 5, p. 59)

$${}_2F_1(1-i\alpha, 1, 1+i\alpha; z^{-1})=\lim_{\epsilon\to0}\frac{\Gamma(1+\epsilon+i\alpha)}{\Gamma(1-i\alpha)\Gamma(\epsilon+2i\alpha)}$$
$$\int_0^1\frac{t^{-i\alpha}(1-t)^{\epsilon+2i\alpha-1}}{1-t/z}\,dt.$$

As a first step, let us compute with relation (4) and the contour Γ of Fig. 2 $\varphi(z)=\beta\{{}_2F_1(-l, 2i\alpha, 1+i\alpha; 1)P_l(\cos\theta)$

$$\varphi(z)=\lim_{\epsilon\to0}\frac{\lambda_\epsilon(\alpha)}{2\pi i}\int_\Gamma\frac{dp}{(p^2-2p\cos\theta+1)^{1/2}}$$
$$\times\int_0^1\frac{t^{-i\alpha}(1-t)^{\epsilon+2i\alpha-1}}{1-tz/p}\,dt,$$
$$\lambda_\epsilon(\alpha)=\frac{\Gamma(1+\epsilon+i\alpha)}{\Gamma(1-i\alpha)\Gamma(\epsilon+2i\alpha)}. \qquad (A1)$$

The integral is zero on both circles, so that formally exchanging integrations in (A1) it follows that

$$\varphi(z)=\lim_{\epsilon\to0}\frac{\lambda_\epsilon(\alpha)}{2\pi}\int_0^1 t^{-i\alpha}(1-t)^{\epsilon+2i\alpha-1}\,\psi(t, z)\,dt,$$
$$\psi(t, z)=2\int_{-\sin\theta}^{\sin\theta}\frac{dy}{(\sin^2\theta-y^2)^{1/2}}\left(\frac{1}{1-(t/z)(\cos\theta+iy)}\right.$$
$$\left.+\frac{1}{1-(t/z)(\cos\theta-iy)}\right).$$

This last integration is trivial:

$$\psi(t,z)=2\pi\left(1-\frac{2t}{z}\cos\theta+\frac{t^2}{z^2}\right)^{-1/2},$$

so

$$\varphi(z)=\lim_{\epsilon\to 0}\lambda_\epsilon(\alpha)\int_0^1\frac{t^{-i\alpha}(1-t)^{\epsilon+2i\alpha-1}}{(1-(t/z)\cos\theta+t^2/z^2)^{1/2}}\,dt. \tag{A2}$$

To prove that $\varphi(z)$ is actually the Borel transform of $\{{}_2F_1(-l,2i\alpha,1+i\alpha;1)P_l(\cos\theta)\}$, Eq. (1) has to be verified; that is,

$${}_2F_1(-l,2i\alpha,1+i\alpha;1)P_l(\cos\theta)$$

$$=\lim_{\epsilon\to 0}\frac{1}{2\pi i}\lambda_\epsilon(\alpha)\int_\Gamma z^{l-1}\,dz$$

$$\times\int_0^1\frac{t^{-i\alpha}(1-t)^{\epsilon+i\alpha-1}}{(1-(2t/z)\cos\theta+t^2/z^2)^{1/2}}\,dt. \tag{A2'}$$

But equality (A2') is evident; just compute the residue at infinity with

$$\left(1-\frac{2t}{z}\cos\theta+\frac{t^2}{z^2}\right)^{-1/2}=\sum_0^\infty P_l(\cos\theta)\frac{t^l}{z^l}$$

$$\text{for}\quad\left|\frac{t}{z}\right|<1$$

and the definition of the beta function.

As a second step, we obtain $\beta\{l\,{}_2F_1(-l,2i\alpha,1+i\alpha;1)P_l(\cos\theta)\}$ using the relation[1,2] $\beta\{nf_n\}=-z(d/dz)f(z)$:

$$\beta\{l\,{}_2F_1(-l,2i\alpha,1+i\alpha;1)P_l(\cos\theta)\}$$

$$=\lim_{\epsilon\to 0}-\lambda_\epsilon(\alpha)z\frac{d}{dz}\int_0^1\frac{t^{-i\alpha}(1-t)^{\epsilon+2i\alpha-1}}{(1-(2t/z)\cos\theta+t^2/z^2)^{1/2}}\,dt$$

$$=\lim_{\epsilon\to 0}\frac{\lambda_\epsilon(\alpha)}{z}\int_0^1\frac{t^{1-i\alpha}(1-\theta)^{\epsilon+2i\alpha-1}}{(1-(2t/z)\cos\theta+t^2/z^2)^{3/2}}$$

$$\times\left(\cos\theta-\frac{t}{z}\right)dt. \tag{A3}$$

Finally, relations (A2) and (A3) lead to the result

$$\beta\{(2l+1)\,{}_2F_1(-l,2i\alpha,1+i\alpha;1)P_l(\cos\theta)\}$$

$$=\lim_{\epsilon\to 0}\lambda_\epsilon(\alpha)\int_0^1\frac{t^{-i\alpha}(1-t)^{\epsilon+2i\alpha-1}(1-t^2/z^2)}{(1-(2t/z)\cos\theta+t^2/z^2)^{3/2}}\,dt, \tag{A4}$$

which, since (A4) is continuous for $z=1$ and $t\in(0,1)$, gives

$$\sum_0^\infty(2l+1)\,{}_2F_1(-l,2i\alpha,1+i\alpha;1)P_l(\cos\theta)$$

$$\overset{B}{=}\lim_{\epsilon\to 0}\lambda_\epsilon(\alpha)\int_0^1\frac{t^{-i\alpha}(1-t)^{\epsilon+2i\alpha-1}(1-t^2)}{(1-2t\cos\theta+t^2)^{3/2}}\,dt. \tag{A5}$$

Let us now look for condition (11), which gives

$${}_2F_1(-l,2i\alpha,1+i\alpha;1)$$

$$=\lim_{\epsilon\to 0}\lambda_\epsilon(\alpha)\left\langle\int_0^1\frac{t^{-i\alpha}(1-t)^{\epsilon+2i\alpha-1}(1-t^2)}{(1-2t\cos\theta+t^2)^{3/2}}\,dt,\,P_l(\cos\theta)\right\rangle$$

$$=\lim_{\epsilon\to 0}\lambda_\epsilon(\alpha)\,\text{f.p.}\int_{-1}^1 P_l(\cos\theta)\,\alpha(\cos\theta)$$

$$\times\int_0^1\frac{t^{-i\alpha}(1-t)^{\epsilon+i\alpha-1}(1-t^2)}{(1-2t\cos\theta+t^2)^{3/2}}\,dt$$

and from Fubini's theorem

$${}_2F_1(-l,2i\alpha,1+i\alpha;1)$$

$$=\lim_{\epsilon\to 0}\lambda_\epsilon(\alpha)\int_0^1 t^{-i\alpha}(1-t)^{\epsilon+i\alpha-1}(1-t^2)\,dt$$

$$\times\int_{-1}^1\frac{P_l(\cos\theta)\,d(\cos\theta)}{(1-2t\cos\theta+t^2)^{3/2}}.$$

This equality is then a consequence of the following relation easy to prove:

$$\int_{-1}^1\frac{P_l(x)\,dx}{(1-2tx+t^2)^{3/2}}=\frac{2t^l}{1-t^2}.$$

Note added in proof: Dr. L. Marquez has made me aware of J. R. Taylor's work, "A new rigourous approach to Coulomb scattering" [Nuovo Cimento B 23, 313 (1974)] in which similar conclusions are obtained.

[1] P. Hillion, "Symbolic method for the Borel transform" (unpublished).
[2] E. I. Jury, *Theory and Application of the Z-Transform Method* (Wiley, New York, 1964).
[3] A. H. Zemanian, *Generalized Integral Transformation* (Interscience, New York, 1966).
[4] L. Schwartz, *Theorie des distributions* (Hermann, Paris, 1950), Vol. I.
[5] A. Erdelyi, Ed., *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I.
[6] L. Marquez, Am. J. Phys. 40, 1520 (1972). See also some of the well-known works which serve as references herein: N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collision* (Oxford U. P., Oxford, 1965); L. D. Landau and E. M. Lifschitz, *Quantum Mechanics* (Pergamon, New York, 1958); M. L. Goldberger and K. M. Watson, *Collision Theory* (Wiley, New York, 1964).
[7] W. Rudin, *Real and Complex Analysis* (McGraw-Hill, New York, 1966), Theorem 11.20.

Small energy denominators in interacting quantum systems: Bound states

S.Ø. Aks, B.B. Varga, and J.J. Sienicki

Department of Physics, University of Illinois, Chicago, Illinois 60680
(Received 31 January 1974)

A perturbative method is developed for calculating bound states of interacting quantum systems, which is based on an analysis of terms with small energy denominators. An iterative scheme is formulated in a systematic manner which eliminates small energy denominators completely. The method is applied to the ϕ^4 model of interacting bosons. The zeroth order solution of the equation of motion differs significantly from the usual free solution, and satisfies a different equation, the determining or bifurcation equation. The additional information contained in this zeroth approximation is used to calculate properties of bound states.

1. INTRODUCTION

The equations of motion of interacting quantum-mechanical system are given as nonlinear differential equations, whose solutions are operator-valued (i.e., q-number) functions of the space–time variables. Although much attention is given to the operator aspects of the problem, somewhat less emphasis is focused on the nonlinearity. This is further complicated in field theory by the fact that the solutions are not actual q-number functions, but rather distributions (generalized functions), in which case the nonlinear terms may not be well defined.

The basic tools employed in the calculations of such systems are the diagrammatic methods, which are based on the standard method of successive approximations, perturbation theory involving straightforward expansion in a coupling parameter. Such procedures are closely related to the classical method of Picard for treating (c-number) differential equations. In that context the method is used not so much to determine solutions, but more often to provide proofs of existence of solutions, and as a means of calculating approximations. However, the approximations obtained by straightforward perturbation theory (to a given order) are often qualitatively, as well as quantitatively, poor, even when the full expansion, including all orders in the coupling parameter, is known to converge. For example, approximations to solutions of equations which can be proved to be periodic in the time variable are found to be nonperiodic, and blow up at large times. Such problems are referred to as secular behavior of the approximation.[1] Modified versions of perturbation theory differing from standard perturbation theory have been devised to eliminate such difficulties. Examples are the classical methods of Lindstedt and Poincaré, and those workers following them.[2] It is interesting that certain of these procedures resemble methods of renormalization theory as used in the many-body problem and quantum field theory.[2]

When the solution of a nonlinear equation is Fourier transformed with respect to the time variable, into a frequency variable or if $\hbar=1$ an energy variable, the occurrence of secular behavior shows up as a resonance phenomenon wherein a fundamental mode (or oscillator) is driven at, or near, its bare frequency by the nonlinearity. This feedback resonance mechanism manifests itself as the problem of terms with small (or vanishing) energy denominators in the perturbative solution.[3–5] Such terms must be treated with special care. Indeed it is these terms which give rise to significant physical effects in a given order of perturbation theory, since the other terms can be transformed away (by a canonical transformation) to that order.[6]

The diagrammatic methods deal with the problem of small denominators by use of adiabatic switching (i.e., by placing $i\epsilon$ in the denominators.) The implications of adiabatic switching are sometimes sufficiently clear for transient interactions, so that its consequences can be handled by minor modifications, as in the formalism of scattering theory. However, in problems involving persistent effects occurring on a longer time scale,[7] adiabatic switching gives rise to incorrect results. For example, phenomena such as bound states and superconductivity are not exhibited in adiabatically switched systems. In such systems the small energy denominators lead to physical effects which are lost by adiabatic switching.

In this paper a version of perturbation theory is developed, which carefully treats the small energy denominators in the context of the ϕ^4 model of quantum field theory. The method which is presented describes persistent phenomena such as binding of particles.

The method is formulated in the Hilbert space of almost periodic functions[8] (in time), and the problem of small denominators is related to the unboundedness of the inverse of the differential wave operator in this space. The basic idea is that in finding approximate solutions for nonlinear differential equations, by a method of successive approximations, only certain zeroth order approximations can be used. The standard zeroth order approximation leads to (secular) behavior in time, which is inconsistent with the behavior that must be shown by the exact solution. The new method involves finding a suitable zeroth approximation as the solution of a bifurcation[9] or determining equation.[10]

In previous work on the ϕ^4 model,[11] a brute force method, which was called quasisecular perturbation theory, was developed for correcting the inconsistent time dependence of the standard first-order perturbative

solution. This corrected solution was applied to the study of the bound state problem in one, two, and three space dimensions. [12]

The present method is not only a justification of the previous heuristic work; it gives a general technique which can be applied to other systems, and furthermore provides a systematic way to go to higher orders in the coupling parameter. In addition, it helps to resolve the issue of characterizing how small a small denominator must be. [6,11] Finally, the present method embodies relativistic invariance for the ϕ^4 model.

In Secs. 2 and 3 the model and the method are introduced. The determining equation is derived and investigated in Secs. 3 and 4. It is used in the construction of one and two particle states in Sec. 5. These states are set up in just such a way that the determining equations are satisfied on them. Particular attention is directed at the two-particle states of energy $E=2m-E_B$ ($E_B>0$ and small) which correspond to bound states. An eigenvalue equation for the energy of the bound state is developed in Sec. 5 and analyzed in Sec. 6. The bound state energies are evaluated for the cases of one and two space dimensions.

The results agree with those obtained with the quasisecular perturbative scheme. [12] An s-wave bound state appears for arbitrarily weak coupling (of the right sign) in both one and two space dimensions, but in three space dimensions the bound state emerges only for couplings well beyond the range of validity of any perturbative method.

We note that this paper is not intended to be a contribution to the extensive and profound literature on the ϕ^4 model of quantum field theory. It is offered as an attempt at developing new calculational procedures for use in problems involving infinitely many degrees of freedom. The ϕ^4 model is treated because of its relative simplicity. The method should be applicable to a variety of other quantum systems in connection with phenomena occurring on long time scales.

2. THE MODEL

The model considered here is based on the field equation

$$(\Box+m^2)\,\phi(t,x)=\lambda m^{3-N}{:}\phi(t,x)^3{:}=\lambda_N{:}\,\phi(t,x)^3{:}, \quad (2.1)$$

where $\Box$ is the differential wave operator $\partial^2/\partial t^2-\nabla^2$ in $N=1$, 2, or 3 space dimensions ($\hbar=c=1$), λ the dimensionless coupling parameter which is taken to be positive, and the field ϕ is real in the sense $\phi^*=\phi$. The symbol : : denoting normal ordering is explained below in the context of this work. In order to effect stability [12] the right-hand side of this equation can be augmented by a term such as $\frac{1}{4}\lambda^2 m^{4-2N}\phi^5$. This order λ^2 term is not explicitly carried in the following. Vector symbols on coordinates x and momenta k are not indicated explicitly.

In addition, the field is subject to the canonical commutation relations

$$[\phi(t,x),\ \phi(t,x')]=[(\partial/\partial t)\phi(t,x),\ (\partial/\partial t)\phi(t,x')]=0, \quad (2.2a)$$

$$[\phi(t,x),\ (\partial/\partial t)\phi(t,x')]=i\,\delta(x-x'). \quad (2.2b)$$

The field is handled in a box of volume V and can be written in the form

$$\phi(t,x)=V^{-1/2}\sum_k a_k(t)\exp(ik\cdot x),$$

where the sum is taken over all allowed momenta for the periodicity box of volume V. The Fourier amplitudes satisfy the equations

$$\ddot{a}_k+\omega_k^2 a_k=\lambda_N V^{-1}\sum_{pqr}\delta_{k,p+q+r}{:}\,a_p a_q a_r{:}, \quad (2.3)$$

where the overhead dots indicate time derivatives and

$$\omega_k^2=m^2+k^2.$$

It is to be expected that the time dependence of a_k (or its matrix elements) is a linear combination of various discrete frequencies since the system is in a finite box. Thus a_k has the time dependence of an almost periodic function. [8]

We will need only a few properties of almost periodic functions. The continuous almost periodic functions form a vector space in which an inner product

$$(f,g)=\lim_{T\to\infty}(1/T)\int_0^T f^*(t)\,g(t)\,dt$$

can be defined. The resulting inner-product space is *not* complete, but can be formally completed by the standard metrical completion process [8] to give a nonseparable Hilbert space $\mathfrak{A}$. The set of exponential functions $\{\exp(-i\sigma t)\,|\,\sigma\in R\}$ (where R is the set of real numbers) is an uncountable orthonormal basis for $\mathfrak{A}$. Thus an element $f\in\mathfrak{A}$ can be represented by a series $f(t)=\sum_{\sigma\in R}\alpha_\sigma\exp(-i\sigma t)$, where only a denumerable number of the α_σ can be nonvanishing since $\sum_\sigma|\alpha_\sigma|^2<\infty$. Thus an alternative way of writing f in this basis is $f(t)=\sum_{n=1}^{\infty}\alpha_{\sigma(n)}\exp[-i\sigma(n)t]$.

The significance of the space of almost periodic functions $\mathfrak{A}$ in the following is that the differential operator $L_k=d^2/dt^2+\omega_k{}^2$ is self-adjoint (more precisely, it has a self-adjoint extension) in $\mathfrak{A}$.

We close this section with a remark on the time dependence of operators. Let $\{|e_n,\gamma\rangle\}$ be a (complete) basis of energy eigenstates with energy e_n and γ representing the other eigenvalues of a complete commuting set. If $A(t)$ is a Heisenberg picture operator, then we can write

$$\begin{aligned}A(t)&=\sum_{e_n e_m}\sum_{\gamma\gamma'}\exp[-i(e_m-e_n)t]\\&\quad\times|e_n,\gamma\rangle\langle e_n,\gamma|A(0)|e_m,\gamma'\rangle\langle e_m,\gamma'|\\&=\sum_\sigma\exp(-i\sigma t)\sum_{e_n}\sum_{\gamma\gamma'}\langle e_n,\gamma|A(0)|e_n+\sigma,\gamma'\rangle\\&\quad\times|e_n,\gamma\rangle\langle e_n+\sigma,\gamma|.\end{aligned}$$

The state $|e_n+\sigma,\ \gamma'\rangle$ is taken to be zero if $e_n+\sigma$ does not correspond to an energy eigenvalue. Therefore, $A(t)$ can be represented on the exponential basis of $\mathfrak{A}$ as $A(t)=\sum_\sigma\exp(-i\sigma t)\hat{A}(\sigma)$. If $|e,\gamma''\rangle$ is an energy eigenstate, then the state $\hat{A}(\sigma)|e,\gamma''\rangle$, if it is nonvanishing, is also an energy eigenstate but with energy $e-\sigma$. For σ positive, $\hat{A}(\sigma)$ lowers the energy by σ, and, for σ nega-

tive, $\hat{A}(\sigma)$ raises the energy by $|\sigma|$, provided the resulting state is nonzero. If a state of lowest energy exists, then the positive frequency part of any Heisenberg operator maps that state into zero. When we fix the energy scale so that the lowest energy state has zero energy, the energy of states obtained by operating on it with a product of negative energy parts of Heisenberg operators can be read off by adding the arguments σ of the operators $\hat{A}(\sigma)$ and multiplying by minus one.

The normal ordering symbol requires that all positive frequency (energy) quantities appearing within it must appear to the right of all negative frequency quantities, with the ordering of factors of the same sign of frequency being otherwise unaffected.

3. THE METHOD

Our aim in this section is to set up a perturbative procedure for treating Eq. (2.3). The major problem with the standard perturbative procedure based on an expansion of the form $a_k=\sum_m \lambda^m a_k^{(m)}$ is the fact that the restricted inverse of L_k, denoted by L_k^{-1}, is unbounded on the orthogonal complement of the null space of L_k. Restricted inverse refers to the inverse of L_k considered as a mapping from the orthogonal complement of its null space onto its range. As a consequence the higher order terms which should be getting smaller need not. We have examined consequences of this behavior in work on the quasisecular perturbation method[11,12] in which a brute force procedure was proposed for dealing with the small denominators which give rise to the quasisecularity or unboundedness. The procedure there was based on the connection between small and zero denominators on the one hand and quasisecularity and secularity on the other. The quasisecular behavior is handled in the standard texts on periodic behavior in nonlinear systems.[1]

The perturbative procedure developed here is unlike the method given in the standard applied mathematics texts on oscillating motion, but it does resemble a procedure given by Hale[10] for the study of periodic oscillations.

The first step involves writing L_k in the spectral form, $L_k=\sum_{\rho\in R}(\omega_k^2-\rho^2)P_\rho$, where P_ρ is the orthogonal projection defined by $P_\rho \exp(-i\sigma t)=\delta_{\rho,\sigma}\exp(-i\rho t)$. Then L_k can be partitioned into $L_k=L_k^0+S_k$ (where L_k^0 has a bounded restricted inverse)

$$L_k^0=\sum_{\rho\notin N_k}(\omega_k^2-\rho^2)P_\rho, \tag{3.1a}$$

$$S_k=\sum_{\rho\in N_k}(\omega_k^2-\rho^2)P_\rho$$

$$=\lambda_N\sum_{\rho\in N_k}\frac{(\omega_k^2-\rho^2)}{\lambda_N}P_\rho=\lambda_N\tilde{S}_k, \tag{3.1b}$$

using the index set $N_k=\{\rho\in R\mid |\rho^2-\omega_k^2|<\lambda m^2\}$. Eq. (2.3) can be rewritten as

$$L_k^0 a_k=\lambda_N\Big(V^{-1}\sum_{pqr}\delta_{k,p+q+r}a_p a_q a_r-\tilde{S}_k a_k\Big)=\lambda_N F_k(a). \tag{3.2}$$

Each term in the spectral representation of S_k is of order λ (the dimensionless coupling parameter); hence both terms on the right-hand side of Eq. (3.2) are of order λ.

Equation (3.2) cannot be used, as it stands, in a perturbative or iterative treatment. Such a procedure would involve the family of equations

$$L_k^0 a_k^{(n+1)}=\lambda_N F_k(a^{(n)}),\quad n=0,1,2,\cdots, \tag{3.3}$$

where the $a_k^{(0)}$ are solutions of the homogeneous equation. For this family of equations to be meaningful, at each iteration step, $F_k(a^{(n)})$ must be in the range of the linear mapping L_k^0, for otherwise the resulting equation is not solvable. This is known as the solvability condition. Since, for a self-adjoint operator, the range is the orthogonal complement of the null space, $F_k(a^{(n)})$ must therefore be within the orthogonal complement of the null space of L_k^0.

It is possible to set up a meaningful iteration procedure. Using $Q_k=\sum_{\rho\in N_k}P_\rho$, the orthogonal projection on the null space of L_k^0, we can write Eq. (3.2) as

$$L_k^0 a_k=\lambda_N(I-Q_k)F_k(a)+\lambda_N Q_k F_k(a).$$

The pair of equations

$$L_k^0 a_k=\lambda_N(I-Q_k)F_k(a), \tag{3.4a}$$

$$Q_k F_k(a)=0 \tag{3.4b}$$

is then equivalent to Eq. (3.2). Clearly any solution of Eq. (3.2) satisfies Eq. (3.4b). Equation (3.4a) can be iterated without difficulty since the projection $(I-Q_k)$ guarantees that the solvability condition is satisfied at every stage of the iteration procedure given by

$$L_k^0 a_k^{(n+1)}=\lambda_N(I-Q_k)F_k(a^{(n)}). \tag{3.5}$$

If the new iteration scheme is terminated at nth order to obtain an $(n+1)$th order approximation, the question arises as to whether or not the resulting approximation is an $(n+1)$th order approximation of the original Eq. (3.2). We can answer affirmatively if at each stage of the iteration procedure Eq. (3.4b) is satisfied, $Q_k F_k(a^{(j)})=0$, $j=0,1,\ldots,n$ for then Eq. (3.4a) is identical with Eq. (3.2).

However, not every homogeneous solution may be used as a zeroth order iterate. Only those homogeneous solutions, for which Eq. (3.4b) is satisfied at every stage of the iteration procedure, are suitable candidates for zeroth order iterates. Thus Eq. (3.4b) plays the role of a determining (or bifurcation) equation.[9]

If the determining equation is satisfied, the iteration procedure of Eq. (3.3) will satisfy the solvability condition at each stage, and the iteration according to Eq. (3.4a) is the same as the iteration according to Eq. (3.3).

It is worth noting that violations of the solvability condition are responsible for such additional problems as secular behavior (unphysical behavior of solutions as functions of time), where solutions are obtained outside the domain of definitions of the linear operator involved. Many versions of perturbation theory are (or can be) based on satisfying a solvability condition as a means for determining parameters such as eigenvalues.[9]

Equation (3.4a) can be put into the "integral" form using the restricted inverse of L_k^0:

$$a_k(t) = h_k(t) + \lambda_N \sum_{\rho \notin N_k} (\omega_k^2 - \rho^2)^{-1} P_\rho F_k(a). \tag{3.6}$$

The $(I - Q_k)$ factor is taken care of by the summation condition $\rho \notin N_k$. The term $h_k(t)$ is a solution of the homogeneous equation $L_k^0 h_k = 0$ and can therefore be represented in the form $h_k(t) = \sum_{\rho \in N_k} \hat{h}_k(\rho) \exp(-i\rho t)$. The iteration scheme written out in integral form is

$$a_k^{(n+1)}(t) = h_k(t) + \lambda_N \sum_{\rho \notin N_k} (\omega_k^2 - \rho^2)^{-1} P_\rho F_k(a^{(n)}), \tag{3.7a}$$

$$a_k^{(0)}(t) = h_k(t). \tag{3.7b}$$

The zeroth order iterate h_k is found using the determining equation. We note that for each n, $a_k^{(n+1)}$ is, as the result of the iteration procedure, a function of the h's.

4. THE DETERMINING EQUATIONS

If an order λ^{n+1} approximate solution is sought, it follows that n iterations must be performed and the solvability condition must be satisfied each time. If h_k is taken in the form $h_k = \sum_{m>0} \lambda^m h_k^{(m)}$, we need h_k to order λ^{n+1} (inclusive). Such h_k are obtained by introducing the order λ^n solutions, $a^{(n)}$ or Eq. (3.5), into Eq. (3.4b) and solving for the h's to order λ^{n+1}. The solvability conditions for the first through nth iterations will be satisfied and, in addition, we obtain h_k to order λ^{n+1} (inclusive) for use as the first term on the right-hand side of Eq. (3.7).

In particular, if a first order in λ solution is sought, we are required to substitute the a's to order λ^0 (namely $a_k^{(0)} = h_k$) into Eq. (3.4b) and solve for the h's to first order in λ. In this case the equation to be solved is

$$\sum_{\rho \in N_k} \lambda_N^{-1} (\omega_k^2 - \rho^2) P_\rho h_k = V^{-1} Q_k \sum_{pqr} \delta_{k,p+q+r}\ h_p h_q h_r. \tag{4.1}$$

The superscript has been omitted from the h's to keep the notation more tractable. We note that Eq. (4.1) is similar to Eq. (2.3) except for the projection Q_k and the restrictions on the frequencies of the h_k's given by $Q_k h_k = h_k$. We are leaving λ_N on the left side of Eq. (4.1) to stress that both sides are of zeroth order in the dimensionless coupling parameter, thus ruling out the usual approximation which involves setting $\lambda = 0$ on the right side.

Using the representation $h_k(t) = \sum_{\sigma \in N_k} \hat{h}_k(\sigma) \exp(-i\sigma t)$, we can define $h_k^\pm$, the positive and negative frequency parts of h_k:

$$h_k^+(t) = \sum_{\sigma \in N_k} \theta(\sigma)\, \hat{h}_k(\sigma) \exp(-i\sigma t) = \sum_{\sigma \in N_k} U_k(\sigma) \exp(-i\sigma t),$$

$$h_k^-(t) = \sum_{\sigma \in N_k} \theta(-\sigma)\, \hat{h}_k(\sigma) \exp(-i\sigma t) = \sum_{\sigma \in N_k} V_k(\sigma) \exp(-i\sigma t),$$

where $\theta(x) = 1$ for $x > 0$ and 0 for $x \leq 0$. Since the (frequency) supports of the (Fourier transformations of the) $h_k^\pm$'s are in narrow mass bands [where mass means the invariant quantity $(\sigma^2 - k^2)^{1/2}$] about m, it follows that this decomposition is relativistically invariant. The determining equation can be replaced by the equation for the positive frequency parts U,

$$\lambda_N^{-1} (\omega_k^2 - \rho^2) U_k(\rho) = 3V^{-1} \sum_{pqr} \sum_{\sigma\mu\nu} \delta_{k+p,q+r}\, \delta_{\rho+\sigma,\nu+\mu}\, U_p(\sigma)^*\, U_q(\nu) U_r(\mu) \tag{4.2}$$

and its conjugate. The summations are subject to the conditions σ, ν, μ positive and in the N of the corresponding momentum. The reality condition for the field implies $\hat{h}_s^+(\rho)^* = \hat{h}_{-s}^-(-\rho)$. The factors have been normally ordered by placing positive frequency factors (which lower energies) to the right of negative frequency factors.

Our objective at this point is to set up the space of states on which Eq. (4.2) is satisifed.

5. CONSTRUCTION OF ONE- AND TWO-PARTICLE STATES

In this section we construct one- and two-particle states and look for two-particle states of energy E which is less than the sum of the energies of any two one-particle states. Such a two-particle state corresponds to a bound state. The one- and two-particle states are constructed so that Eq. (4.2) is satisfied on these states.

The vacuum state is taken to be the state of lowest energy which, by an adjustment of the energy scale has zero energy. The vacuum state satisfies the condition

$$U_k(\rho)\Omega = 0 \tag{5.1}$$

for all k and $\rho \in N_k$, with $\rho > 0$. This follows from the fact that U_k agrees with the positive frequency part of $\hat{a}_k$ in the frequency gap region N_k and is, therefore, an energy lowering operator.

We note that Eq. (4.2) is trivially satisfied on Ω. The conjugate equation applied to Ω gives

$$(\omega_k^2 - \rho^2)\, U_k(\rho)^*\, \Omega = 0. \tag{5.2}$$

The nonvanishing states of the type $U_k(\rho)^*\Omega$ are interpreted as one-particle states. The only nonvanishing one-particle state of momentum k is of the form

$$\Omega(k) = U_k(\omega_k)^*\, \Omega. \tag{5.3}$$

Such one-particle states are all of mass $m (m^2 = \omega_k^2 - k^2)$ even though $U_k(\rho)^*$ has support over a range of mass values.

Now Eq. (4.2) must be satisfied on the one-particle states $\Omega(k')$:

$$\lambda_N^{-1} (\omega_k^2 - \rho^2)\, U_k(\rho)\, \Omega(k') = 0. \tag{5.4}$$

The right-hand side vanishes since there are two energy lowering operators acting on $\Omega(k')$. We assume that the gap determined by N_k is small compared to the mass. This, of course, means that we are dealing with a weak coupling theory. Actually Eq. (5.4) can be proved even if we do not assume weak coupling. In that case the right-hand side differs from zero by a quantity of order λ which does not affect our result to leading order in λ. It follows that $U_k(\rho)\Omega(k') = 0$, unless $\rho = \omega_k$. For the case $\rho = \omega_k$ we have

$$U_k(\omega_k)\Omega(k') = (2\omega_k)^{-1} \delta_{k,k'}\, \Omega, \tag{5.5}$$

assuming that there is a gap about $E = 0$ in the mass

spectrum (i.e., there are no states whose mass squared is less than $\frac{1}{2}\lambda m^2$, other than the vacuum state). The factor $(2\omega_k)^{-1}$ is introduced for normalization purposes to insure the validity of the commutation relations Eq. (2.2) on the vacuum state Ω. See the Appendix for details.

Finally we apply the conjugate of Eq. (4.2) to $\Omega(k')$:

$$\lambda_N^{-1}(\omega_k^2-\rho^2)U_k(\rho)^*\Omega(k') = 3V^{-1}\sum_{\rho\sigma\tau}\sum_{\sigma\mu\nu}\delta_{k+\rho,q+\tau}\delta_{\rho+\sigma,\mu+\nu} \times U_\tau(\nu)^*U_q(\mu)^*U_\rho(\sigma)\Omega(k'). \quad (5.6)$$

Using Eq. (5.5) and restricting attention to the case $k'=-k$ gives

$$\lambda_N^{-1}(\omega_k^2-\rho^2)U_k(\rho)^*\Omega(-k) = 3(2\omega_k V)^{-1}\sum_{q\tau}\sum_{\sigma\nu}\delta_{0,q+\tau}\delta_{\rho+\omega_k,\sigma+\nu}U_\tau(\sigma)^*U_q^*(\nu)\Omega. \quad (5.7)$$

In terms of the new variable $E=\rho+\omega_k$, Eq. (5.7) becomes

$$\lambda_N^{-1}(E-2\omega_k)U_k(E-\omega_k)^*U_{-k}(\omega_k)^*\Omega = -3(2\omega_k EV)^{-1}\sum_r U_r(E-\omega_r)^*U_{-r}(\omega_r)^*\Omega. \quad (5.8)$$

Note that the states involved on both sides of this equation have momentum zero and energy E. We consider linear combinations of these states to get an eigenvalue equation for the energy E:

$$\sum_k{}' f_k\lambda_N^{-1}(E-2\omega_k)U_k(E-\omega_k)^*U_{-k}(\omega_k)^*\Omega = -3\sum_k{}' f_k(2\omega_k EV)^{-1}\sum_r{}' U_r(E-\omega_r)^*U_{-r}(\omega_r)^*\Omega. \quad (5.9)$$

The primes on the k summations indicate that, for fixed E, k is restricted according to the frequency support condition $(E-\omega_k)\in N_k$. The corresponding condition applies to the r summation where k is replaced everywhere by r. Equation (5.9) can be rewritten as

$$\sum_q{}'[f_q\lambda_N^{-1}(E-2\omega_q)+3\sum_l{}' f_l(2\omega_l EV)^{-1}] U_q(E-\omega_q)^*U_{-q}(\omega_q)^*\Omega=0, \quad (5.10)$$

from which we obtain the eigenvalue equation to be satisfied by E and f:

$$\lambda_N^{-1}(E-2\omega_q)f_q=-3\sum_r(2\omega_r EV)^{-1}\theta(\lambda m^2-|(E-\omega_r)^2-\omega_r^2|)f_r. \quad (5.11)$$

In the next section we look for solutions of this equation subject to the condition $E<2m$, corresponding to the requirement that the rest energy of a bound state is less than the rest energy of its constituent particles, in this case $2m$.

6. BOUND STATE CALCULATION

Equation (5.12) resembles the eigenvalue equation obtained using the quasisecular approximation procedure in first order to determine a bound state.[12]

Equation (4.2) of that paper can be rewritten in the notation of this paper as

$$(E-2\omega_l)g_l=-(3\lambda_N/4V)\sum_k(\omega_l\omega_k)^{-1}\theta(\lambda m-2|\omega_l-\omega_k|)g_k. \quad (6.1)$$

The approximate bound state energies $E_B=2m-E$ found in that paper for the cases $N=1$, 2, 3 (corresponding to 1, 2, and 3 space dimensions) are of order λ or higher order in λ. We look for approximate solutions of the same form for Eq. (5.11).

Solutions of this form can be obtained from Eq. (5.11) by replacing E on the right-hand side by $2m$ to get the approximate equation

$$(E-2\omega_l)f_l=-(3\lambda_N/4V)\sum_k(m\omega_k)^{-1}\theta(\lambda m-4|m-\omega_k|)f_k. \quad (6.2)$$

Since ω_l and ω_k differ from m by terms of order λ, both can be replaced by their nonrelativistic approximations to give

$$(E_B+m^{-1}|l|^2)f_l=(3\lambda_N/4Vm)\sum_k\omega_k^{-1}\theta(\tfrac{1}{2}\lambda m^2-|k|^2)f_k. \quad (6.3)$$

We define $\chi=V^{-1}\sum_k\omega_k^{-1}\theta(\frac{1}{2}\lambda m^2-|k|^2)f_k$ and express Eq. (6.3) in the form $f_l=(3\lambda_N/4m)(E_B+m^{-1}|l|^2)^{-1}\chi$. By substituting this expression for f_k into the definition of χ, we obtain the eigenvalue condition

$$(3\lambda_N/4Vm)\sum_k\omega_k^{-1}(E_B+m^{-1}|k|^2)^{-1}\theta(\tfrac{1}{2}\lambda m^2-|k|^2)=1. \quad (6.4)$$

It is clear that this equation can only have a solution E_B for λ_N positive.

In the case of one space dimension we can pass to the infinite volume limit of Eq. (6.4) to obtain the equation

$$\int_0^{(\lambda/2)^{1/2}m}dk\,(mE_B{}^{(1)}+k^2)^{-1}=4\pi m/3\lambda.$$

This gives the transcendental equation

$$(mE_B{}^{(1)})^{-1/2}\tan^{-1}(\lambda m/2E_B^{(1)})^{1/2}=4\pi/3\lambda m,$$

where λ is the dimensionless coupling parameter. In the weak coupling limit $\tan^{-1}(\lambda m/2E_B{}^{(1)})^{1/2}\to\frac{1}{2}\pi$ and the lowest order contribution to $E_B^{(1)}$ is

$$E_B^{(1)}=(3\lambda/8)^2m.$$

For the case of two space dimensions we again pass to the infinite volume limit of Eq. (6.4) and get

$$\int_0^{(\lambda/2)^{1/2}m}dk\,k(mE_B{}^{(2)}+|k|^2)^{-1}=8\pi/3\lambda,$$

which leads to

$$E_B^{(2)}=\tfrac{1}{2}\lambda m\exp(-16\pi/3\lambda)$$

for weak coupling. These results agree with those of the previous work.[12]

In the case of three space dimensions the bound state occurs only for strong coupling ($\lambda>8.5$) which is well beyond the range of validity of our procedure.

APPENDIX

In the Appendix we show that the normalization of the

one-particle states is given by

$$\|U_k(\omega_k)^*\Omega\| = (2\omega_k)^{-1/2}. \tag{A1}$$

The commutation relations

$$[a_k(t),\ \dot{a}_l(t)] = i\delta_{k,-l}, \tag{A2a}$$

$$[a_k(t), a_l(t)] = [\dot{a}_k(t),\ \dot{a}_l(t)] = 0 \tag{A2b}$$

follow from those for the field equations (2.2). The first order approximations $a_k^{(1)}$ must satisfy these commutation relations with an error of order λ^2 at the most.

To determine the normalization of the one-particle states, we need only the vacuum expectation value of Eq. (A.2a):

$$(\Omega, [a_k^{(1)}(t),\ \dot{a}_l^{(1)}(t)]\Omega) = i\delta_{k,-l} + O(\lambda^2). \tag{A3}$$

Simplifying the notation, we drop the superscript (1), but the reader should keep in mind in the following that we are dealing with the first order approximation. We rewrite Eq. (A3) as

$$(a_k^+(t)^*\Omega, \dot{a}_l^-(t)\Omega) - (\dot{a}_l^+(t)^*\Omega,\ a_k^-(t)\Omega) = i\delta_{k,-l} + O(\lambda^2).$$

Note that the positive frequency parts map Ω into zero. Taking $k = -l$ and using the reality condition, we obtain from Eq. (A2)

$$\sum_{\rho \geq 0} 2\rho(\hat{a}_k^+(\rho)^*\Omega,\ \hat{a}_k^+(\rho)^*\Omega) = 1 + O(\lambda^2). \tag{A4}$$

Introducing

$$\hat{a}_k^+(\rho)^* = U_k(\rho)^* + \lambda_N(\omega_k^2 - \rho^2)^{-1}(I - Q_k)P_\rho^0 F_k(h)^*$$

into Eq. (A4) gives zeroth, first, and second order terms in λ. Here P_ρ^0 indicates that after projecting with P_ρ, t is set to zero. The zeroth order term is

$$2\omega_k(U_k(\omega_k)^*\Omega,\ U_k(\omega_k)^*\Omega),$$

where we have used the result following from Eq. (5.2) that only mass-m one-particle states occur. Concerning the first order terms, these appear with $\hat{h}_k^+(\rho)^*$ acting on Ω on one side of the inner product and with $(\omega_k^2 - \rho^2)^{-1} \times (I - Q_k)P_\rho^0 F_k(h)^*$ acting on the other side. The first quantity is nonvanishing only when $\rho = \omega_k$, whereas the second factor vanishes in that case. Consequently, the order λ contribution from the left-hand side of Eq. (A4) vanishes as it must if that euqation is to be valid. We do not consider the order λ^2 contributions, since the right-hand side of Eq. (A4) admits an order λ^2 error, except to remark that these terms involve states of energy near $4m$ and larger.

It follows that

$$\|U_k(\omega_k)^*\Omega\|^2 = (U_k(\omega_k)^*\Omega,\ U_k(\omega_k)^*\Omega) = (2\omega_k)^{-1}. \tag{A5}$$

Equation (5.5) follows from Eq. (A5) since the vacuum is unique:

$$U_k(\omega_k)\Omega(k) = U_k(\omega_k)U_k(\omega_k)^*\Omega = C_k\Omega,$$

where C_k is a constant. Taking the inner product of this with Ω gives $C_k = (2\omega_k)^{-1}$ through comparison with Eq. (A5).

[1] N.N. Bogoliubov and Y.A. Mitropolsky, *Asymptotic Methods in the Theory of Nonlinear Oscillations* (Hindustani, Delhi, 1961).
[2] R. Bellman, *Perturbation Techniques in Mathematics, Physics and Engineering* (Dover, New York, 1966).
[3] H. Poincaré, *Méthodes nouvelles de la mécanique céleste* (Gauthier-Villars, Paris, 1892), Vol. I. Chap. V.
[4] E.T. Whittaker, *A Treatise on Analytical Dynamics* (Dover, New York, 1944), Chap. XVI.
[5] N.N. Bogoliubov, Zh. Eksp. Teor. Fiz. **34**, 58 (1958) [Sov. Phys. JETP **7**, 41 (1958)].
[6] H. Frölich, Proc. Roy. Soc. A **215**, 291 (1952); S.G. Eckstein, D. Friedlander and C.G. Kuper, Phys. Rev. Lett. **30**, 78 (1973).
[7] L. Van Hove, Physica **21**, 901 (1955); **22**, 343 (1956).
[8] N. Akhiezer and M. Glazman, *Theory of Linear Operators in Hilbert Space*, translated by M. Nestell (Ungar, New York, 1961), Vol. I, p. 132.
[9] J. Keller, in *Bifurcation Theory and Nonlinear Eigenvalue Problems*, edited by J. Keller and S. Antman (Benjamin, New York, 1969), p. 17.
[10] J.K. Hale, *Nonlinear Oscillations* (McGraw-Hill, New York, 1963).
[11] S. Aks, J. Sienicki and B. Varga, Phys. Rev. D **6**, 520 (1972).
[12] S. Aks and B. Varga, Phys. Rev. D **6**, 2773 (1972).

Theory of tachyons for arbitrary spin

D. Shay

Department of Physics and Astronomy, University of Oklahoma, Oklahoma 73069
(Received 23 September 1974)

It is shown that, even though the finite-dimensional representations of the little group for tachyons, particles with spacelike momentum, are not unitary in general, a fully covariant theory for spin-s tachyons can be obtained. This occurs because a helicity dependent factor $(-1)^{\sigma}$, where σ denotes helicity, appears in the expressions for the invariant integral. This factor acts as a metric for the nonunitary representations and a covariant theory results. As has been found for spin-(1/2), the ideas of conserved particle number or total charge do not work for tachyons. The conserved quantities depend on helicity.

I. INTRODUCTION

Dhar and Sudarshan[1] and Feinberg[2] have suggested that a covariant theory of tachyons, particles with spacelike momenta, may be obtained only when the representations of the little group are unitary. Since the little group for the class of particles with spacelike momentum is the Lorentz group in two dimensions, the only finite unitary representation is the one-dimensional representation. This would restrict the covariant theory of tachyons to the spin zero case only.

It was found in Bandukwala and Shay[3] that a covariant spin-$\frac{1}{2}$ theory could be developed. What was obtained was a pseudounitary representation of the little group $O(2,1)$, with the appropriate metric appearing in the invariant integral, a pseudoscalar, and the conserved axial vector. It was the presence of the metric in the expressions for the observables that made them invariant with respect to the little group transformations and made a covariant theory possible. It was also the presence of this metric that prevented the construction of a scalar probability and momentum 4-vector for the tachyons. It was noted, as Tanaka[4] had earlier, that particle number, charge, or energy–momentum were not conserved quantities for tachyons, but rather quantities involving helicity were the only constant observables.

In different approaches Marx[5] and Hamamoto[6] have considered the spin-$\frac{1}{2}$ and general spin cases, respectively. For the spin-$\frac{1}{2}$ case they have obtained spinors and observables similar to those obtained by Bandukwala and Shay[3] and in this paper. Both Marx[5] and Hamamoto[6] eliminate the helicity dependence of the observables by a reinterpretation or by assuming helicity dependent commutation relations, thus obtaining a scalar probability and a momentum 4-vector. Their procedures involve an unjustified disregard for the actual transformation properties of these observables.

In an effort to see what happens in the general spin case for tachyons, the spin-s equations given by Weinberg[7] are applied to tachyons. For the sake of simplicity a simple non-second-quantized theory is developed. The negative energy states are included to ensure covariance as was suggested by Arons and Sudarshan.[8] However, following the convention in Bandukwala and Shay,[3] antitachyon states will not be included since they are not necessary for covariance. There is room in the theory for an independent theory of antitachyons which may be just added on to the particle theory. This is noted at the end of Sec. II.

The tachyon spinors are found by solving the wave equations in the transcendental $k^0=0$ frame and then boosting out of that frame to a general one. The transformation properties of the spinors under the little group are also found, along with the orthogonality relations for the spinors. The invariant integrals, a scalar for integral spin and a pseudoscalar for odd half-integral spin, depend on helicity and are unaffected by the little group transformations. So a covariant theory results in all cases, and, in all cases, helicity and not just charge or particle number is important.

II. THE WAVE EQUATIONS

A tachyon field having an arbitrary spin s satisfies the following wave equation:

$$(\gamma^{\mu_1\cdots\mu_{2s}}\partial_{\mu_1}\cdots\partial_{\mu_{2s}}+m^{2s})\psi(x)=0, \tag{1}$$

with the subsidiary condition that

$$(\partial_\mu\partial^\mu-m^2)\,\psi(x)=0. \tag{2}$$

It should be emphasized here that the metric $g^{\mu\nu}=(+---)$ is used. The $2(2s+1)\times 2(2s+1)$ γ-matrices have been given by Weinberg[7]; and, although the form of the equations here appears identical to Weinberg's, the metric used here ensures that tachyons are described.

The wavefunction, or field operator, may be expanded as a Fourier integral in the manner of Bandukwala and Shay[3]:

$$\begin{aligned}\psi(x)=m^s(2\pi)^{-3/2}\int d\Omega\int_0^\infty dk^0\,[(k^0)^2+m^2]^{1/4}\\ \times\sum_\sigma[a_\sigma(\hat\omega,k^0)u_+^\sigma(\hat\omega,k^0)\exp(-ikx)\\ +a_\sigma(-\hat\omega,-k^0)u_-^\sigma(-\hat\omega,k^0)\exp(+ikx)].\end{aligned} \tag{3}$$

The negative energy particle states are included explicitly as was done by Arons and Sudarshan.[8] The integral $\int d\Omega$ is the solid angle integral over all directions of k,

$$\mathrm{k}=k\hat\omega,$$

and

$$k=\sqrt{(k^0)^2+m^2},$$

where k^0 will be positive whenever it is written. The index σ indicates the helicity and ranges from $-s$ to $+s$

in integral steps.

The spinors $u_\epsilon^\sigma(\hat{\omega}, k^0)$ satisfy the following equations:

$$\hat{\omega}\cdot\Sigma\, u_\epsilon^\sigma(\hat{\omega}, k^0) = \sigma u_\epsilon^\sigma(\hat{\omega}, k^0) \tag{4}$$

and

$$[k_{\mu_1}\cdots k_{\mu_{2s}}\gamma^{\mu_1\cdots\mu_{2s}} + (im)^{2s}]u_\epsilon^\sigma(\hat{\omega}, k^0) = 0. \tag{5}$$

The index ϵ indicates positive and negative energy states and is the sign of the energy. The momentum vector is

$$k^\mu = (\sqrt{(k^0)^2 + m^2}\hat{\omega},\ \epsilon k^0),\quad k^0 \geq 0.$$

The γ-matrices used here are defined in terms of the three general $2s+1\times 2s+1$ spin matrices $\mathbf{S}$. The $2(2s+1)\times 2(2s+1)$ matrices are defined by

$$\Sigma = \begin{pmatrix} \mathbf{S} & 0 \\ 0 & \mathbf{S} \end{pmatrix}, \tag{6}$$

$$\gamma^5 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}, \tag{7}$$

and

$$\beta = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \tag{8}$$

where I is the $2s+1\times 2s+1$ unit matrix.

From Weinberg,[7] allowing for the difference in the metric, the γ-matrices are given in the following form:

$$k_{\mu_1}\cdots k_{\mu_{2s}}\gamma^{\mu_1\cdots\mu_{2s}} = -\,p^{(s)}(k), \tag{9}$$

with

$$p^{(s)}(k) = (k_\mu k^\mu)^s\beta + \sum_{n=0}^{s-1}\frac{(k_\mu k^\mu)^{s-1-n}}{(2n+2)!} \times(2\mathbf{k}\cdot\Sigma)\,[(2\mathbf{k}\cdot\Sigma)^2 - (2\mathbf{k})^2]\cdots[(2\mathbf{k}\cdot\Sigma)^2 - (2n\mathbf{k})^2] \times[(2\mathbf{k}\cdot\Sigma)\beta + (2n+2)(\epsilon k^0)\gamma^5\beta] \tag{10}$$

for integral s and

$$p^{(s)}(k) = -(k_\mu k^\mu)^{s-1/2}[(\epsilon k^0)\beta + (2\mathbf{k}\cdot\Sigma)\gamma^5\beta] - \sum_{n=1}^{s-1/2}\frac{(k_\mu k^\mu)^{s-1/2-n}}{(2n+1)!} \times[(2\mathbf{k}\cdot\Sigma)^2 - \mathbf{k}^2]\cdots[(2\mathbf{k}\cdot\Sigma)^2 - ([2n-1]\mathbf{k})^2] \times[(2n+1)(\epsilon k^0)\beta + (2\mathbf{k}\cdot\Sigma)\gamma^5\beta] \tag{11}$$

for odd half-integral s.

In the transcendental frame k^0 vanishes and $\mathbf{k} = m\hat{\omega}$, so the spinors satisfy the following equations:

$$(F^{(s)}(\sigma)\beta - I)u_\epsilon^\sigma(\hat{\omega}, 0) = 0 \tag{12}$$

for integral s and

$$(G^{(s)}(\sigma)\gamma^5\beta + iI)u_\epsilon^\sigma(\hat{\omega}, 0) = 0 \tag{13}$$

for odd half-integral s. In both cases the helicity equation (4) has been used to reduce the matrix $p^{(s)}(k)$ to the simple form given above. The polynomials which appear in the above equations have the form

$$F^{(s)}(\sigma) = 1 + \sum_{n=0}^{s-1}\frac{2^{2n+2}(-1)^{n+1}}{(2n+2)!}\,\sigma^2[\sigma^2 - 1]\cdots[\sigma^2 - n^2] \tag{14}$$

and

$$G^{(s)}(\sigma) = 2\sigma + \sum_{n=1}^{s-1/2}\frac{2^{2n+1}(-1)^n}{(2n+1)!}\,\sigma[\sigma^2 - \tfrac{1}{2}^2)\cdots[\sigma^2 - (n-\tfrac{1}{2})^2]. \tag{15}$$

These sums are terminated hypergeometric series which may be rewritten as gamma functions and, in turn, as trigonometric functions using relations given in Whittaker and Watson.[9] They may then be found to have the values

$$F^{(s)}(\sigma) = (-1)^\sigma \tag{16}$$

and

$$G^{(s)}(\sigma) = (-1)^{\sigma - 1/2}. \tag{17}$$

The $2(2s+1)$ element spinors $u_\epsilon^\sigma(\hat{\omega}, 0)$ may be written in terms of the $2s+1$ spinors $\chi_\sigma(\hat{\omega})$, where

$$\hat{\omega}\cdot\mathbf{S}\,\chi_\sigma(\hat{\omega}) = \sigma\chi_\sigma(\hat{\omega}).$$

For integral spin the spinors have the form

$$u_\epsilon^\sigma(\hat{\omega}, 0) = \begin{pmatrix} \tfrac{1}{2}(1+(-1)^\sigma)\chi_\sigma(\hat{\omega}) \\ \tfrac{1}{2}(1-(-1)^\sigma)\chi_\sigma(\hat{\omega}) \end{pmatrix}, \tag{18}$$

while for the odd half-integral spin,

$$u_\epsilon^\sigma(\hat{\omega}, 0) = \begin{pmatrix} \tfrac{1}{2}\chi_\sigma(\hat{\omega}) \\ \tfrac{1}{2}i(-1)^{\sigma-1/2}\chi_\sigma(\hat{\omega}) \end{pmatrix}. \tag{19}$$

In both cases the spinors are orthonormal,

$$u_\epsilon^{\sigma\dagger}(\hat{\omega}, 0)\,u_\epsilon^{\sigma'}(\hat{\omega}, 0) = \delta_{\sigma\sigma'},$$

where

$$\chi_\sigma^\dagger(\hat{\omega})\,\chi_{\sigma'}(\hat{\omega}) = \delta_{\sigma\sigma'}.$$

Notice that there is no distinction between positive and negative energy states in the transcendental frame. It is also true that there are only $2s+1$ linearly independent spinors $u_\epsilon^\sigma(\hat{\omega}, 0)$. The other $2s+1$ spinors could be used to describe antitachyon states, but they are not required for the covariance of the theory and do not affect the main theme of this paper. The antitachyon states would satisfy a wave equation like Eq. (1) with a minus sign in front of the mass term.

III. TRANSFORMATION PROPERTIES OF THE TACHYON SPINORS

Following the notation used by Hamermesh,[10] the spinors transform according to the following rule under the homogeneous Lorentz group:

$$D(W)u_\epsilon^\sigma(\hat{\omega}, k^0) = \sum_{\sigma'\epsilon'} u_{\epsilon'}^{\sigma'}(\hat{\omega}', k^{0\prime})[D(W_0)]_{\sigma'\epsilon',\,\sigma\epsilon}, \tag{20}$$

where W is the transformation that takes $(\sqrt{(k^0)^2 + m^2}\hat{\omega},\ \epsilon k^0)$ to $(\sqrt{(k^{0\prime})^2 + m^2}\hat{\omega}',\ \epsilon' k^{0\prime})$. The matrix $[D(W_0)]_{\sigma'\epsilon',\sigma\epsilon}$ is the matrix representation of the little group, the group of transformations that leaves the 4-vector $(m\hat{\omega},\ 0)$ invariant. It is made up of the rotations about an axis parallel to $\hat{\omega}$ and the Lorentz boosts perpendicular to $\hat{\omega}$.

Therefore, a tachyon spinor in an arbitrary Lorentz frame may be written as

$$u_\epsilon^\sigma(\hat{k}, k^0) = D(W) u_\epsilon^\sigma(\hat{\omega}, 0) = \sum_{\sigma' \epsilon'} U_{\epsilon'}^{\sigma'}(\hat{k}, k^0) [D(W_0)]_{\sigma' \epsilon', \sigma \epsilon}, \tag{21}$$

where

$$D(W_0) u_\epsilon^\sigma(\hat{\omega}, 0) = \sum_{\sigma' \epsilon'} u_{\epsilon'}^{\sigma'}(\hat{\omega}, 0) [D(W_0)]_{\sigma' \epsilon', \sigma \epsilon}. \tag{22}$$

The spinors $U_\epsilon^\sigma(\hat{k}, k^0)$ are obtained from the transcendental frame spinors $u_\epsilon^\sigma(\hat{k}, 0)$ by boosting along the $\hat{k}$ direction. The spinors $u_\epsilon^\sigma(\hat{k}, k^0)$, on the other hand, are obtained from the $u_\epsilon^\sigma(\hat{k}, 0)$ by a general Loretnz transformation. In a sense, the little group matrices represent an arbitrariness in the specific form of the spinors $u_\epsilon^\sigma(\hat{k}, k^0)$ in any frame. For tardyons, timelike momentum, and massless particles the matrices $[D(W_0)]_{\sigma' \epsilon', \sigma \epsilon}$ are unitary. As a result, observables like total charge, particle number, or energy–momentum are independent of these matrices. This happens because the orthogonality relations used to calculate these observables are invariant with respect to the little group transformations. So the arbitrariness in the tardyon and massless particle spinors due to the little group has no physical consequence.

For tachyons, the little group is the two-dimensional Lorentz group; it is not compact; and its finite-dimensional representations are not unitary in general. Under these circumstances the usual observables like total charge or probability or energy–momentum will not be invariant with respect to these transformations. It was shown by Bandukwala and Shay[3] that spin-$\frac{1}{2}$ tachyon theory yields the total helicity, a pseudoscalar, and the product of helicity and momentum, an axial vector, as the only conserved quantities. These observables were found to be invariant with respect to the little group because of the unusual orthogonality relations that hold for the spinors. Although the little group matrices were not unitary, the helicity dependent parts of the relevant orthogonality relations acted as a metric for the transformations. The little group representations are found here for any spin $s \geq \frac{1}{2}$, and it is found that what occurs for spin-$\frac{1}{2}$ occurs for any nonzero spin.

The spinors $U_\epsilon^\sigma(\hat{\omega}, k^0)$ may be obtained by boosting out of the transcendental frame along $\hat{\omega}$ so that $(m\hat{\omega}, 0)$ becomes $(\sqrt{(k^0)^2 + m^2}\hat{\omega}, \epsilon k^0)$ and

$$U_\epsilon^\sigma(\hat{\omega}, k^0) = \exp -\left(\hat{\omega}\gamma^5 \cdot \Sigma \sinh^{-1} \frac{\epsilon k^0}{m}\right) u_\epsilon^\sigma(\hat{\omega}, 0), \tag{23}$$

where $\sinh^{-1} \epsilon k^0/m$ appears for tachyons instead of the usual $\sinh^{-1} k/m$ as for tardyons. The form of a general boost is given by Weinberg,[7] for example.

Expanding the exponential as an infinite series and using the fact that the spinors are helicity eigenstates yields

$$\exp\left(-\hat{\omega} \cdot \gamma^5 \Sigma \sinh^{-1} \frac{\epsilon k^0}{m}\right) u_\epsilon^\sigma(\hat{\omega}, 0) = [\tfrac{1}{2}(\kappa_{+1}^\sigma + \kappa_{-1}^\sigma) - \tfrac{1}{2}\epsilon\gamma^5(\kappa_{+1}^\sigma - \kappa_{-1}^\sigma)] u_\epsilon^\sigma(\hat{\omega}, 0), \tag{24}$$

where

$$\kappa_{\pm 1}^\sigma = (k/m \pm k^0/m)^\sigma.$$

This transformation gives the following spinors:

$$U_\epsilon^\sigma(\hat{\omega}, k^0) = \frac{1}{2}\begin{pmatrix} (\kappa_{-\epsilon}^\sigma + (-1)^\sigma \kappa_{+\epsilon}^\sigma)\chi_\sigma(\hat{\omega}) \\ (\kappa_{-\epsilon}^\sigma - (-1)^\sigma \kappa_{+\epsilon}^\sigma)\chi_\sigma(\hat{\omega}) \end{pmatrix}. \tag{25}$$

for integral s and

$$U_\epsilon^\sigma(\hat{\omega}, k^0) = \frac{1}{2\sqrt{2}}(1 + i\epsilon(-1)^{\sigma - 1/2})\begin{pmatrix} (\kappa_{-1}^\sigma - i\epsilon(-1)^{\sigma-1/2}\kappa_{+1}^\sigma)\chi_\sigma(\hat{\omega}) \\ \epsilon(\kappa_{-1}^\sigma + i\epsilon(-1)^{\sigma-1/2}\kappa_{+1}^\sigma)\chi_\sigma(\hat{\omega}) \end{pmatrix} \tag{26}$$

for odd half-integral s.

The little group transformations, from Bargmann and Wigner,[11] are

$$D(W_0) = \exp(-\zeta \cdot \gamma^5 \Sigma) \exp(-i\zeta^0 \hat{\omega} \cdot \Sigma). \tag{27}$$

The right-hand term is a rotation about $\hat{\omega}$ and contributes a phase factor when $D(W_0)$ acts on $u_\epsilon^\sigma(\hat{\omega}, 0)$, so it will be ignored. The remaining term is a pure Lorentz transformation, a boost, in the $\hat{\zeta}$ direction perpendicular to $\hat{\omega}$. The velocity of the boost is

$$v = \tanh\zeta.$$

For an eigenstate of helicity the matrix $\gamma^5 \zeta \cdot \Sigma$, for $\hat{\zeta} \cdot \hat{\omega} = 0$, has the following property:

$$\gamma^5 \hat{\zeta} \cdot \Sigma u_\epsilon^\sigma(\hat{\omega}, 0) = \sum_{\sigma'} M_{\sigma\sigma'} u_\epsilon^{\sigma'}(\hat{\omega}, 0), \tag{28}$$

where

$$M_{\sigma\sigma'} = \exp(i\alpha)\left(\frac{\sqrt{(s-\sigma)(s+\sigma+1)}}{2}\delta_{\sigma+1,\sigma'} + \frac{\sqrt{(s+\sigma)(s-\sigma+1)}}{2} \times \delta_{\sigma-1,\sigma'}\right)$$

and

$$\exp(i\alpha) = \begin{cases} 1 & \text{integral spin} \\ i(-1)^{\sigma-1/2} & \text{odd half-integral spin.} \end{cases}$$

This follows from the properties of raising and lowering operators for angular momentum as discussed in Edmonds,[12] for example.

For an infinitesimal little group transformation it follows that

$$D(W_0) u_\epsilon^\sigma(\hat{\omega}, 0) = \sum_{\sigma'} (\delta_{\sigma\sigma'} - \zeta M_{\sigma\sigma'}) u_\epsilon^{\sigma'}(\omega, 0), \tag{29}$$

so, in general,

$$D(W) u_\epsilon^\sigma(\hat{\omega}, k^0) = \sum_{\sigma'} [\exp(-\zeta M)]_{\sigma\sigma'} u_\epsilon^{\sigma'}(\omega', k^{0'}) \tag{30}$$

where

$$Wk = k'$$

and the choice of ϵ' depends on whether $k^0 - \mathbf{k} \cdot \mathbf{v}$ is positive or negative, v being the velocity associated with the transformation W.

IV. OBSERVABLES AND ORTHOGONALITY RELATIONS

The spinors $U_\epsilon^\sigma(\hat{\omega}, k^0)$ are found by boosting the transcendental frame spinors $u_\epsilon^\sigma(\hat{\omega}, 0)$ along the $\hat{\omega}$ direction.

This is a special set of spinors and $U^{\sigma}_{\epsilon}(\hat{\omega}, k^0)$ will be transformed into $U^{\sigma}_{\epsilon}(\hat{\omega}', k^{0\prime})$, helicity remaining invariant, only by members of the factor group of the little group in the Lorentz group. A general Lorentz transformation, on the other hand, will mix helicity states, the coefficients of this linear combination forming the representation of the little group [see Eq. (30)]. The theory will be fully covariant only if the observables are unaffected by the little group transformations, this is equivalent to saying that the results should be independent of the choice of the $u^{\sigma}_{\epsilon}(\hat{\omega}, k^0)$ or the $U^{\sigma}_{\epsilon}(\hat{\omega}, k^0)$. So all orthogonality relations used to construct the observables should be invariant with respect to the little group.

If an orthogonality relation for the spinors $U^{\sigma}_{\epsilon}(\hat{\omega}, k^0)$ is indicated by

$$O_{\sigma\epsilon,\sigma'\epsilon'} = U^{\sigma\dagger}_{\epsilon}(\hat{\omega}, k^0)\ \Gamma\ U^{\sigma'}_{\epsilon'}(\hat{\omega}, k^0), \tag{31}$$

then it will be invariant with respect to the little group if

$$\sum_{\sigma''} (M^*_{\sigma\sigma''} O_{\sigma''\epsilon,\sigma'\epsilon'} + M_{\sigma'\sigma''} O_{\sigma\epsilon,\sigma''\epsilon'}) = 0, \tag{32}$$

where $M_{\sigma\sigma'}$ is given in Eq. (28). This follows from the requirement of invariance under the infinitesimal transformations. Equation (32) will imply that the orthogonality relation will also hold for the spinors $u^{\sigma}_{\epsilon}(\hat{\omega}, k^0)$. The helicity dependence of the orthogonality relation acts as a metric for the nonunitary transformations. Furthermore, relations simply proportional to $\delta_{\sigma\sigma'}$, with no other σ dependence, will not be invariant; and these are the relations used to calculate total charge, particle number, and momentum.

For integral spin s, Eq. (25) implies the following:

$$U^{\sigma\dagger}_{\epsilon}(\hat{\omega}, k^0) U^{\sigma'}_{\epsilon'}(\hat{\omega}, k^0) = \tfrac{1}{2}\delta_{\sigma\sigma'}\delta_{\epsilon\epsilon'}(\kappa^{2\sigma}_{+1} + \kappa^{2\sigma}_{-1}) + \delta_{\sigma\sigma}(1 - \delta_{\epsilon\epsilon'}), \tag{33}$$

$$\bar{U}^{\sigma}_{\epsilon}(\hat{\omega}, k^0) U^{\sigma'}_{\epsilon'}(\hat{\omega}, k^0) = (-1)^{\sigma}\delta_{\sigma\sigma'}\delta_{\epsilon\epsilon'} + \tfrac{1}{2}(-1)^{\sigma}\delta_{\sigma\sigma'}(1 - \delta_{\epsilon\epsilon'}) \times (\kappa^{2\sigma}_{+1} + \kappa^{2\sigma}_{-1}), \tag{34}$$

$$U^{\sigma\dagger}_{\epsilon}(\hat{\omega}, k^0)\gamma^5 U^{\sigma'}_{\epsilon'}(\hat{\omega}, k^0) = -\tfrac{1}{2}\epsilon\delta_{\sigma\sigma'}\delta_{\epsilon\epsilon'}(\kappa^{2\sigma}_{+1} - \kappa^{2\sigma}_{-1}), \tag{35}$$

and

$$\bar{U}^{\sigma}_{\epsilon}(\hat{\omega}, k^0)\gamma^5 U^{\sigma'}_{\epsilon'}(\hat{\omega}, k^0) = \tfrac{1}{2}\epsilon(-1)^{\sigma}\delta_{\sigma\sigma'}(1 - \delta_{\epsilon\epsilon'})(\kappa^{2\sigma}_{+1} - \kappa^{2\sigma}_{-1}). \tag{36}$$

In the odd half-integral spin case, Eq. (26) implies:

$$U^{\sigma\dagger}_{\epsilon}(\hat{\omega}, k^0) U^{\sigma'}_{\epsilon'}(\hat{\omega}, k^0) = \tfrac{1}{2}\delta_{\sigma\sigma'}\delta_{\epsilon\epsilon'}(\kappa^{2\sigma}_{+1} + \kappa^{2\sigma}_{-1}) + \delta_{\sigma\sigma'}(1 - \delta_{\epsilon\epsilon'}), \tag{37}$$

$$\bar{U}^{\sigma}_{\epsilon}(\hat{\omega}, k^0) U^{\sigma'}_{\epsilon'}(\hat{\omega}, k^0) = \tfrac{1}{2}i\epsilon(-1)^{\sigma-1/2}\delta_{\sigma\sigma'}(1 - \delta_{\epsilon\epsilon'})(\kappa^{2\sigma}_{+1} - \kappa^{2\sigma}_{-1}), \tag{38}$$

$$U^{\sigma\dagger}_{\epsilon}(\hat{\omega}, k^0)\gamma^5 U^{\sigma'}_{\epsilon'}(\hat{\omega}, k^0) = -\tfrac{1}{2}\epsilon\delta_{\sigma\sigma'}\delta_{\epsilon\epsilon'}(\kappa^{2\sigma}_{+1} - \kappa^{2\sigma}_{-1}), \tag{39}$$

and

$$\bar{U}^{\sigma}_{\epsilon}(\hat{\omega}, k^0)\gamma^5 U^{\sigma'}_{\epsilon'}(\hat{\omega}, k^0) = i(-1)^{\sigma-1/2}\delta_{\sigma\sigma'}\delta_{\epsilon\epsilon'} + \tfrac{1}{2}i(-1)^{\sigma-1/2} \times \delta_{\sigma\sigma'}(1 - \delta_{\epsilon\epsilon'})(\kappa^{2\sigma}_{+1} + \kappa^{2\sigma}_{-1}). \tag{40}$$

In both cases, the definition

$$\bar{U}^{\sigma}_{\epsilon}(\hat{\omega}, k^0) = U^{\sigma\dagger}_{\epsilon}(\hat{\omega}, k^0)\beta$$

is used and the expressions $\kappa^{2\sigma}_{\pm 1} = (\kappa^{\sigma}_{\pm 1})^2$ are defined in Eq. (24).

From the wave equation (1) a conserved current may be found such that

$$\partial_{\mu} j^{\mu}(x) = 0, \tag{41}$$

where

$$j^{\mu}(x) = \frac{1}{m^{2s-1}} \sum_{k=0}^{2s-1} (-1)^k \partial_{\mu_1} \cdots \partial_{\mu_k} \bar{\psi}(x)\, \Gamma\gamma^{\mu_1 \cdots \mu_{2s-1}\mu} \times \partial_{\mu_{k+1}} \cdots \partial_{\mu_{2s-1}} \psi(x), \tag{42}$$

and $\Gamma = I, \gamma^5$ for integral, odd half-integral spin, respectively. Also since the wavefunctions satisfy the Klein–Gordon equation, there are two more possible conserved currents for each spin:

$$J_{\mu}(\Gamma, x) = \frac{i}{2m^{2s}}[\bar{\psi}(x)\Gamma\partial_{\mu}\psi(x) - \partial_{\mu}\bar{\psi}(x)\Gamma\psi(x)], \tag{43}$$

where $\Gamma = I, -i\gamma^5$ for each spin. In arriving at Eq. (42) the following properties of the γ-matrices were used:

$$\beta\gamma^{\mu_1 \cdots \mu_{2s}\dagger}\beta = \gamma^{\mu_1 \cdots \mu_{2s}}$$

and

$$\gamma^5\gamma^{\mu_1 \cdots \mu_{2s}}\gamma^5 = -\gamma^{\mu_1 \cdots \mu_{2s}}.$$

The invariant integral, the integral over all 3-space of the zero component of a conserved 4-vector, obtained from Eq. (42) is proportional to that obtained from $J_{\mu}(I, x)$ for integral spin and from $J_{\mu}(-i\gamma^5, x)$ for odd half-integral spin. This can be shown by substituting Eq. (3) into Eq. (42), integrating, and using the orthogonality relations for the $U^{\sigma}_{\epsilon}(\omega, k^0)$ to show that

$$\int d^3\mathbf{x} j^0(x) = i^{2s-1} m s(2s+1) \sum_{\epsilon\sigma} \int d\Omega \int_0^{\infty} \frac{dk^0}{k^0} k_{\mu_1} \cdots k_{\mu_{2s-1}} \times \bar{U}^{\sigma}_{\epsilon}(\hat{\omega}, k^0)\Gamma\gamma^{\mu_1 \cdots \mu_{2s-1}0} U^{\sigma}_{\epsilon}(\hat{\omega}, k^0)\, a^{\dagger}_{\sigma}(\hat{\omega}, \epsilon k^0) a_{\sigma}(\hat{\omega}, \epsilon k^0),$$

where

$$k^{\mu} = (\mathbf{k}, \epsilon k^0).$$

The following relation is true:

$$\bar{U}^{\sigma}_{\epsilon}(\hat{\omega}, k^0)\Gamma k_{\mu_1} \cdots k_{\mu_{2s-1}} \gamma^{\mu_1 \cdots \mu_{2s-1}0} U^{\sigma}_{\epsilon}(\hat{\omega}, k^0) = a^0{}_{\nu}\bar{U}^{\sigma}_{\epsilon}(\hat{\omega}, 0) m^{2s-1}\Gamma\hat{\omega}_{i_1} \cdots \hat{\omega}_{i_{2s-1}} \gamma^{i_1 \cdots i_{2s-1}0} U^{\sigma}_{\epsilon}(\hat{\omega}, 0),$$

where the $a^{\mu}{}_{\nu}$ are the vector transformation coefficients for the boost from the transcendental frame to the general one. The use of the form of $a^0{}_{\nu}$ and the properties of the matrices $\rho^{(s)}(k)$ and the spinors in the transcendental frame completes the proof.

For the case of integral spin, the invariant integral is

$$\int d^3\mathbf{x} J_0(I, x) = \sum_{\sigma} \int d\Omega \int_0^{\infty} dk^0 (-1)^{\sigma} [a^{\dagger}_{\sigma}(\hat{\omega}, k^0) a_{\sigma}(\hat{\omega}, k^0) - a^{\dagger}_{\sigma}(\hat{\omega}, -k^0) a_{\sigma}(\hat{\omega}, -k^0)], \tag{44}$$

and

$$\int d^3\mathbf{x} J_0(-i\gamma^5, x) = 0. \tag{45}$$

For the odd half-integral spin, the invariant integral is

$$\int d^3\mathbf{x} J_0(-i\gamma^5, x) = \sum_\sigma \int d\Omega \int_0^\infty dk^0 (-1)^{\sigma-1/2}$$
$$\times [a_\sigma^\dagger(\hat{\omega}, k^0) a_\sigma(\hat{\omega}, k^0) - a_\sigma^\dagger(\hat{\omega}, -k^0) a_\sigma(\hat{\omega}, -k^0)], \tag{46}$$

while

$$\int d^3\mathbf{x} J_0(I, x) = 0. \tag{47}$$

Both integrals have essentially the same form, but under spatial inversions $\sigma \to -\sigma$ so that for integral spin $(-1)^\sigma$ is a scalar but for odd half-integral spin $(-1)^{\sigma-1/2}$ is a pseudoscalar. The invariant integrals are a product of the number of particles and a helicity factor. In neither case can the conventional interpretation of total number of particles or charge be used. These integrals are not positive definite. Also, for $\sigma = 0$, Eq. (44) reduces to the form given by Dhar and Sudarshan[1] for spinless particles.

It is simple to show that the invariant integrals are unaffected by the little group transformations. Referring back to Eq. (32), it can be shown that the orthogonality relations that contribute to the invariant integral,

$$O_{\sigma\epsilon,\sigma'\epsilon'} = (-1)^\sigma \delta_{\sigma\sigma'} \delta_{\epsilon\epsilon'} \tag{48}$$

for integral spin and

$$O_{\sigma\epsilon,\sigma'\epsilon'} = (-1)^{\sigma-1/2} \delta_{\sigma\sigma'} \delta_{\epsilon\epsilon'} \tag{49}$$

for odd half-integral spin, satisfy the condition for invariance with respect to the little group. What occurs is that the little group, $O(2, 1)$, is found to have pseudounitary representations with respect to the metric $(-1)^\sigma$ or $(-1)^{\sigma-1/2}$. The resulting invariant integrals, and any observables, contain this metric as a factor; and this is sufficient to yield a fully covariant theory for any spin.

V. CONCLUSIONS

It has been shown here that Weinberg's wave equations for particles with any spin can be applied to tachyons. The tachyon spinors form a basis for a pseudounitary representation of the little group $O(2, 1)$ with the metric $(-1)^\sigma \delta_{\sigma\sigma'}$ for integral spin and $(-1)^{\sigma-1/2} \delta_{\sigma\sigma'}$ for odd half-integral spin, where σ denotes helicity. The invariant integral in each case involves the metric as a factor and transforms like a scalar for integral spin and a pseudoscalar for odd half-integral spin. It is precisely because the finite representations of the little group are not unitary that the need for a metric arises, necessarily leading to a helicity dependent invariant integral which cannot have the usual interpretation of probability or charge. So the requirement of Dhar and Sudarshan,[1] that the little group representation be unitary, is not necessary in order to have covariance, and the discussion of tachyons need not be limited to the spinless case.

Positive and negative energy tachyons are taken to be different states of the same particle since they are identical in the transcendental frame and can be transformed into one another. The reinterpretation of negative energy states as antitachyons leads to a noninvariant vacuum. It is possible, on the basis of the transformation properties of momentum and velocity, to consider that negative energy tachyons carry momentum directed oppositely to their velocity. Momentum becomes an internal property of tachyons and may further be considered unobservable. The expression involving the difference of the amplitudes which appears in the invariant integral can now be interpreted as the net number of tachyons with helicity σ traveling in the $\hat{\omega}$ direction.

The alternate addition and subtraction of particles over successive helicity states is a problem to interpret. Certainly, since helicity is an invariant, helicity eigenstates can be constructed; and then the invariant integral is proportional to the net helicity being transported. For the spin-$\frac{1}{2}$ case the integral is the net helicity transported for a general superposition of states. For a general superposition of helicity states for any spin the interpretation of the invariant integral is just not clear.

[1] J. Dhar and E.C.G. Sudarshan, Phys. Rev. **174**, 1808 (1968).
[2] G. Feinberg, Phys. Rev. **159**, 1089 (1967).
[3] J. Bandukwala and D. Shay, Phys. Rev. D **9**, 889 (1974).
[4] S. Tanaka, Prog. Theor. Phys. **24**, 177 (1960).
[5] E. Marx, Int. J. Theor. Phys. **3**, 299 (1970).
[6] S. Hamamoto, Prog. Theor. Phys. **48**, 1037 (1972).
[7] S. Weinberg, Phys. Rev. B **133**, 1318 (1964).
[8] M.E. Arons and E.C.G. Sudarshan, Phys. Rev. **173**, 1622 (1968).
[9] E.T. Whittaker and G.N. Watson, *A Course of Modern Analysis* (Cambridge U.P., New York, 1958), 4th ed.
[10] M. Hamermesh, *Group Theory and its Application to Physical Problems* (Addison-Wesley, Reading, 1964).
[11] V. Bargmann and E.P. Wigner, Proc. Natl. Acad. Sci. USA **34**, 211 (1948).
[12] A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, 1957).

Nonmetrical dynamics. I

C. Martin Pereira*

Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742
(Received 6 August 1974)

Einstein's nonvacuum gravitational field equations are examined. The following nonmetrical existence theorem is proved. It is shown that given any analytic symmetric contravariant energy–momentum tensor density as a function of the space–time coordinates, a solution to the gravitational field equations always exists. Furthermore, this solution is such that the law of conservation of energy–momentum is satisfied. The new proof of the present paper makes no use of the coordinate transformation method used in a previous paper. The Cauchy–Kowalewsky existence theorem is used in the proof, and the literature on the Cauchy–Kowalewsky existence theorem is briefly reviewed. (Riquier's existence theorem is avoided except for a brief discussion in the Appendix.) The initial value problem of Einstein's equations is examined as part of the proof. The physical and mathematical meaning of this new proof is discussed. It is noted that the nonmetrical existence theorem disappears when Einstein's equations are linearized whereas the familiar Lichnerowicz-type existence theorems survive linearization. This suggests that the nonmetrical existence theorem may add a new dimension to our understanding of the physical meaning of nonlinearity in Einstein's equations.

1. INTRODUCTION

In a recent paper,[1] the author has examined Einstein's field equations

$$(-g)^{1/2}G^{\mu\nu}=-8\pi\tilde{T}^{\mu\nu} \tag{1.1}$$

and proven the following theorem.

Theorem 1: *Nonmetrical existence theorem*: Given any symmetric analytic tensor density $\tilde{T}^{\mu\nu}(x)$, there always exists a corresponding metric $g^{\mu\nu}(x)$ which satisfies the field equations (1.1).

J. N. Goldberg has pointed out that it should be possible to prove this theorem without making use of the coordinate transformation method of Ref. 1. In the present paper, it is shown that such a proof is indeed possible.

In addition, the proof of the present paper avoids making explicit use of Riquier's existence theorem. The resulting proof is therefore more elementary than the proof of Ref. 1.

The methods of the present proof help to broaden one's understanding of the physical meaning contained in the original existence theorem of Ref. 1. In addition, the methods of the present proof will allow a more rapid generalization of the result to the Maxwell–Einstein equations.[2]

2. A NEW PROOF OF THE NONMETRICAL EXISTENCE THEOREM

The new proof of Theorem 1 will now be presented. One begins with an analysis of the energy–momentum conservation law

$$\tilde{T}^{\mu\nu}{}_{;\nu}=0. \tag{2.1}$$

Transvect Eq. (2.1) with $g_{\tau\mu}$ and obtain

$$g_{\tau\mu}(\tilde{T}^{\mu\nu}{}_{,\nu}+\Gamma_{\alpha\beta}{}^{\mu}\tilde{T}^{\alpha\beta})=0, \tag{2.2}$$

which becomes

$$g_{\tau\mu}\tilde{T}^{\mu\nu}{}_{,\nu}+\Gamma_{\alpha\beta,\tau}\tilde{T}^{\alpha\beta}=0, \tag{2.3}$$

and this may be written

$$g_{\tau\mu}\tilde{T}^{\mu\nu}{}_{,\nu}+\Gamma_{ij,\tau}\tilde{T}^{ij}+2\Gamma_{i4,\tau}\tilde{T}^{i4}+\Gamma_{44,\tau}\tilde{T}^{44}=0, \tag{2.4}$$

where $\mu=1,\cdots,4$ and $i=1,2,3$. But

$$\Gamma_{\alpha\beta,\tau}=\tfrac{1}{2}(g_{\alpha\tau,\beta}+g_{\beta\tau,\alpha}-g_{\alpha\beta,\tau}). \tag{2.5}$$

Substitute Eq. (2.5) into Eq. (2.4) and obtain

$$\tfrac{1}{2}(g_{4\tau,4}+g_{4\tau,4}-g_{44,\tau})\tilde{T}^{44}+(g_{i\tau,4}+g_{4\tau,i}-g_{4i,\tau})\tilde{T}^{i4}+\Gamma_{ij,\tau}\tilde{T}^{ij}+g_{\tau\mu}\tilde{T}^{\mu\nu}{}_{,\nu}=0. \tag{2.6}$$

Next solve Eqs. (2.6) for derivatives of the form $g_{\tau4,4}$. The first three Eqs. (2.6) (which have $\tau=1,2,3$) can be written

$$g_{4k,4}\tilde{T}^{44}=\tfrac{1}{2}g_{44,k}\tilde{T}^{44}-(g_{ik,4}+g_{4k,i}-g_{i4,k})\tilde{T}^{i4}-\Gamma_{ij,k}\tilde{T}^{ij}-g_{k\mu}\tilde{T}^{\mu\nu}{}_{,\nu}. \tag{2.7a}$$

and the fourth of Eqs. (2.6) (which has $\tau=4$) can be written

$$\tfrac{1}{2}g_{44,4}\tilde{T}^{44}=-g_{44,i}\tilde{T}^{i4}-\Gamma_{ij,4}\tilde{T}^{ij}-g_{4\mu}\tilde{T}^{\mu\nu}{}_{,\nu}. \tag{2.7b}$$

Equations (2.7) may be written, after division by $\tilde{T}^{44}$,

$$g_{4k,4}=\tfrac{1}{2}g_{44,k}-(\tilde{T}^{44})^{-1}[2\Gamma_{i4,k}\tilde{T}^{i4}+\Gamma_{ij,k}\tilde{T}^{ij}+g_{k\mu}\tilde{T}^{\mu\nu}{}_{,\nu}] \tag{2.8a}$$

and

$$g_{44,4}=-2(\tilde{T}^{44})^{-1}[g_{44,i}\tilde{T}^{i4}+\Gamma_{ij,4}\tilde{T}^{ij}+g_{4\mu}\tilde{T}^{\mu\nu}{}_{,\nu}]. \tag{2.8b}$$

Careful examination shows that the right-hand sides of Eqs. (2.8) do not contain any derivatives of the form $g_{4k,4}$ or $g_{44,4}$. One can therefore apply the Cauchy–Kowalewsky existence theorem[3–6] and conclude that one can think of Eqs. (2.8) as defining the unknown functions $g_{4\mu}$ once the functions g_{ij} and $\tilde{T}^{\mu\nu}$ are given. That is, regardless of what analytic choice is made for the functions g_{ij} and $\tilde{T}^{\mu\nu}$, the $g_{4\mu}$ always exist provided that $\tilde{T}^{44}\neq0$. Note that if $\tilde{T}^{44}=0$, one can always transform to a coordinate system in which $T^{44}\neq0$ provided that at least one of the $\tilde{T}^{\mu\nu}$ is nonzero.

One can now proceed to consider Einstein's equations

(1.1). As is well known, Eqs. (1.1) imply the energy–momentum conservation law (2.1). Consider the combined system

$$(-g)^{1/2}G^{\mu\nu} = -8\pi\tilde{T}^{\mu\nu}, \tag{2.9a}$$

$$\tilde{T}^{\mu\nu}{}_{;\nu} = 0. \tag{2.9b}$$

The system (2.9) is equivalent to Einstein's equations (1.1). A proof of existence for the system (2.9) will now be given. Lichnerowicz has shown[7] that the system (2.9) can be replaced by the following equivalent system in which four of Eqs. (2.9a) have been replaced by initial conditions on the spacelike surface $x^4 = 0$:

$$R_{ij} = -8\pi(-g)^{-1/2}[\tilde{T}_{ij} - \tfrac{1}{2}g_{ij}\tilde{T}], \tag{2.10a}$$

$$[(-g)^{1/2}G^4{}_\lambda + 8\pi\tilde{T}^4{}_\lambda]_{x^4=0} = 0, \tag{2.10b}$$

$$\tilde{T}_\mu{}^\nu{}_{;\nu} = 0, \tag{2.10c}$$

where $i, j = 1, 2, 3$ and $\mu = 1, \cdots, 4$.

For a demonstration that Eqs. (2.10) are equivalent to Eqs. (2.9), see Appendix B or Ref. 7.

The final step is to prove the existence of solutions to the system (2.10) by using the Cauchy–Kowalewsky existence theorem.[4,5] To do this, one writes out Eqs. (2.10) more explicitly. Equation (2.10a) may be written

$$\begin{aligned}&\tfrac{1}{2}g^{\alpha\beta}(g_{ij,\alpha\beta} + g_{\alpha\beta,ij} - g_{\alpha j,i\beta} - g_{i\beta,\alpha j}) \\ &\quad + 8\pi(-g)^{-1/2}(\tilde{T}_{ij} - \tfrac{1}{2}g_{ij}\tilde{T}) \\ &\quad = -g^{\mu\nu}(\Gamma_{ij,\nu}\Gamma_{\alpha\beta,\mu} - \Gamma_{i\beta,\nu}\Gamma_{\alpha j,\mu})\, g^{\alpha\beta}\end{aligned} \tag{2.11}$$

Next, separate out the terms in Eq. (2.11) that are multiplied by g^{44} and then divide by g^{44}. Equation (2.11) then becomes

$$\begin{aligned}&g_{ij,44} + g_{44,ij} - g_{4j,i4} - g_{i4,4j} \\ &\quad + 16\pi(g^{44})^{-1}(-g)^{-1/2}(\tilde{T}_{ij} - \tfrac{1}{2}g_{ij}\tilde{T}) = N_{ij},\end{aligned} \tag{2.12}$$

where N_{ij} is an abbreviation for the remaining terms:

$$\begin{aligned}N_{ij} \equiv{}& g^{k4}(g^{44})^{-1}(g_{k4,ij} + g_{ij,k4} - g_{kj,i4} - g_{i4,kj}) \\ &+ g^{4k}(g^{44})^{-1}(g_{4k,ij} + g_{ij,4k} - g_{4j,ik} - g_{ik,4j}) \\ &- g^{kn}(g^{44})^{-1}(g_{kn,ij} + g_{ij,kn} - g_{kj,in} - g_{in,kj}) \\ &- 2g^{\alpha\beta}g^{\mu\nu}(g^{44})^{-1}[\Gamma_{ij,\nu}\Gamma_{\alpha\beta,\mu} - \Gamma_{i\beta,\nu}\Gamma_{\alpha j,\mu}].\end{aligned} \tag{2.13}$$

Note that although N_{ij} contains both first and second derivatives of the metric, it does not contain any of the derivatives $g_{ij,44}$ and it also does not contain any derivatives of the form $g_{\mu 4,4\nu}$. (See Appendix D.)

Using Eqs. (2.12) and Eqs. (2.8), one can rewrite Eqs. (2.10) in the form

$$\begin{aligned}&g_{ij,44} + g_{44,ij} - g_{4j,i4} - g_{i4,4j} \\ &\quad + 16\pi(g^{44})^{-1}(-g)^{-1/2}(\tilde{T}_{ij} - \tfrac{1}{2}g_{ij}\tilde{T}) = N_{ij},\end{aligned} \tag{2.14a}$$

$$[(-g)^{1/2}G^4{}_\lambda + 8\pi\tilde{T}^4{}_\lambda]_{x^4=0} = 0, \tag{2.14b}$$

$$\begin{aligned}g_{4k,4} = {}& \tfrac{1}{2}g_{44,k} \\ &- (\tilde{T}^{44})^{-1}[2\Gamma_{i4,k}\tilde{T}^{i4} + \Gamma_{ij,k}\tilde{T}^{ij} + g_{k\mu}\tilde{T}^{\mu\nu}{}_{,\nu}],\end{aligned} \tag{2.14c}$$

$$g_{44,4} = -2(\tilde{T}^{44})^{-1}[g_{44,i}\tilde{T}^{i4} + \Gamma_{ij,4}\tilde{T}^{ij} + g_{4\mu}\tilde{T}^{\mu\nu}{}_{,\nu}]. \tag{2.14d}$$

In order to obtain from Eqs. (2.14) a set of equations which satisfies the requirements of the Cauchy–Kowalewsky existence theorem, one must eliminate the derivatives $g_{i4,4j}$ from Eqs. (2.14a). To do this, define

$$\begin{aligned}V_k \equiv{}& \tfrac{1}{2}g_{44,k} - (\tilde{T}^{44})^{-1}(g_{ik,4} + g_{4k,i} - g_{i4,k})\tilde{T}^{i4} \\ &- (\tilde{T}^{44})^{-1}[\Gamma_{ij,k}\tilde{T}^{ij} + g_{k\mu}\tilde{T}^{\mu\nu}{}_{,\nu}],\end{aligned} \tag{2.15a}$$

$$\begin{aligned}V_4 \equiv{}& -2(\tilde{T}^{44})^{-1}[g_{44,i}\tilde{T}^{i4} \\ &+ \tfrac{1}{2}(g_{i4,j} + g_{j4,i} - g_{ij,4})\tilde{T}^{ij} + g_{4\mu}\tilde{T}^{\mu\nu}{}_{,\nu}].\end{aligned} \tag{2.15b}$$

By using the abbreviations V_i and V_4, Eqs. (2.14) become

$$\begin{aligned}&g_{ij,44} + g_{44,ij} - g_{4j,i4} - g_{i4,4j} + 16\pi(g^{44})^{-1}(-g)^{-1/2} \\ &\quad \times(g_{i\mu}g_{j\nu} - \tfrac{1}{2}g_{ij}g_{\mu\nu})\tilde{T}^{\mu\nu} = N_{ij},\end{aligned} \tag{2.16a}$$

$$[(-g)^{1/2}G^4{}_\lambda + 8\pi g_{\lambda\mu}\tilde{T}^{4\mu}]_{x^4=0} = 0, \tag{2.16b}$$

$$g_{4k,4} = V_k, \tag{2.16c}$$

$$g_{44,4} = V_4. \tag{2.16d}$$

One can now eliminate the derivatives $g_{i4,4j}$ from Eq. (2.16a) by substitution, using Eqs. (2.16c) and (2.16d). The result is

$$\begin{aligned}g_{ij,44} = {}& -g_{44,ij} + V_{j,i} + V_{i,j} + N_{ij} \\ &- 16\pi(g^{44})^{-1}(-g)^{-1/2}(g_{i\mu}g_{j\nu} - \tfrac{1}{2}g_{ij}g_{\mu\nu})\tilde{T}^{\mu\nu},\end{aligned} \tag{2.17a}$$

$$[(-g)^{1/2}G^4{}_\lambda + 8\pi g_{\lambda\mu}\tilde{T}^{4\mu}]_{x^4=0} = 0 \tag{2.17b}$$

$$g_{4k,4} = V_k, \tag{2.17c}$$

$$g_{44,4} = V_4, \tag{2.17d}$$

where V_k, V_4, and N_{ij} are respectively defined in Eqs. (2.15a), (2.15b), and (2.13). The significance of the manipulations just completed may be summarized as follows. Equations (2.17) have been obtained from Einstein's field equations (2.10) by solving for a particular set of derivatives, namely, $g_{ij,44}$ and $g_{\mu 4,4}$. The resulting system (2.17a), (2.17c), and (2.17d) is now in the form required for the application of Riquier's existence theorem[8–12] and it would be easy at this point to prove existence using Riquier's ordering procedure. However, that path will not be followed here. Instead, a proof of existence using the more familiar Cauchy–Kowalewsky existence theorem will be presented. To place Eqs. (2.17) in the Cauchy–Kowalewsky form, one notes that the terms $V_{j,i} + V_{i,j}$ in Eq. (2.17a) contain derivatives of the form $g_{4k,ij}$. Unfortunately, the Cauchy–Kowalewsky theorem does not allow second derivatives of a function on the right-hand side of a system of equations when any of the first derivatives of this function (in this case g_{4k}) appear on the left-hand side of the system. To remedy this, one must change Eqs. (2.17c) so that the $g_{4k,4}$ do not appear. To do this, simply differentiate Eqs. (2.17c) with respect to x^4 (see below).

A second problem appears with respect to g_{44} since the *first* derivative of g_{44} appears on the right-hand side of Eq. (2.17d) and the *second* derivative of g_{44} appears on the left-hand side of Eq. (2.17a). The second problem is solved by differentiating Eqs. (2.17d) with respect to x^4 (see below). In addition, one adds appropriate initial conditions so that the new system is equivalent to the old system. After these changes, the resulting system is as follows:

$$g_{ij,44} = -g_{44,ij} + V_{j,i} + V_{i,j} + N_{ij}$$
$$- 16\pi(g^{44})^{-1}(-g)^{-1/2}(g_{i\mu}g_{j\nu} - \tfrac{1}{2}g_{ij}g_{\mu\nu})\tilde{T}^{\mu\nu}, \quad (2.18a)$$
$$[(-g)^{1/2}G^4{}_\lambda + 8\pi g_{\lambda\mu}\tilde{T}^{4\mu}]_{x^4=0} = 0, \quad (2.18b)$$
$$[g_{4k,4} - V_k]_{x^4=0} = 0, \quad (2.18c)$$
$$g_{4k,44} = V_{k,4}, \quad (2.18d)$$
$$[g_{44,4} - V_4]_{x^4=0} = 0, \quad (2.18e)$$
$$g_{44,44} = V_{4,4}. \quad (2.18f)$$

Equations (2.18) still fall short of the Cauchy–Kowalewsky form because the derivatives $g_{ik,44}$ appear both on the left-hand side of Eq. (2.18a) and on the right-hand side of Eq. (2.18d). Furthermore, the $g_{ik,44}$ appear on the right-hand side of Eq. (2.18f). To remedy these two circumstances, define a new abbreviation M_{ij} equal to the right-hand side of Eq. (2.18a):

$$M_{ij} \equiv -g_{44,ij} + V_{j,i} + V_{i,j} + N_{ij}$$
$$- 16\pi(g^{44})^{-1}(-g)^{-1/2}(g_{i\mu}g_{j\nu} - \tfrac{1}{2}g_{ij}g_{\mu\nu})\tilde{T}^{\mu\nu}. \quad (2.19)$$

Equation (2.18a) can now be rewritten in the form

$$g_{ij,44} - M_{ij} = 0. \quad (2.20)$$

Next eliminate $g_{ij,44}$ from Eqs. (2.18d) and (2.18f) by subtracting an appropriate multiple of Eq. (2.20) from Eqs. (2.18d) and (2.18f):

$$g_{ij,44} = -g_{44,ij} + V_{j,i} + V_{i,j} + N_{ij}$$
$$- 16\pi(g^{44})^{-1}(-g)^{-1/2}(g_{i\mu}g_{j\nu} - \tfrac{1}{2}g_{ij}g_{\mu\nu})\tilde{T}^{\mu\nu}, \quad (2.21a)$$
$$[(-g)^{1/2}G^4{}_\lambda + 8\pi g_{\lambda\mu}\tilde{T}^{4\mu}]_{x^4=0} = 0, \quad (2.21b)$$
$$[g_{4k,4} - V_k]_{x^4=0} = 0, \quad (2.21c)$$
$$g_{4k,44} = V_{k,4} + (\tilde{T}^{44})^{-1}\tilde{T}^{i4}(g_{ik,44} - M_{ik}), \quad (2.21d)$$
$$[g_{44,4} - V_4]_{x^4=0} = 0, \quad (2.21e)$$
$$g_{44,44} = V_{4,4} - (\tilde{T}^{44})^{-1}\tilde{T}^{ij}(g_{ij,44} - M_{ij}), \quad (2.21f)$$

where M_{ij}, V_i, V_4, and N_{ij} are given respectively by Eqs. (2.19), (2.15a), (2.15b), and (2.13). The system consisting of Eqs. (2.21a), (2.21d), and (2.21f) is now in the form required for application of the Cauchy–Kowalewsky existence theorem. Applying this theorem, one concludes that solutions exist to the system for every analytic choice of initial values on an $x^4 = \text{const}$ surface, and for every given analytic symmetric tensor $\tilde{T}^{\mu\nu}(x)$.

It only remains to prove that solutions exist to the Eqs. (2.21b), (2.21c), and (2.21e) which express the conditions which these initial values must satisfy. A proof of existence for this initial value problem is given in Appendix C. (Note that the initial value problem solved here is not quite the same as the usual initial value problem since one has eight initial value equations rather than four.) This completes the proof of the following theorem.

Theorem 1: *Nonmetrical existence theorem*: Given any symmetric analytic tensor density $\tilde{T}^{\mu\nu}(x)$, there always exists a corresponding metric $g_{\mu\nu}(x)$ which satisfies Einstein's field equations.

This is the theorem which was to be proved. The theorem has been proved for $\mu = 1, \cdots, 4$ and $i = 1, \cdots, 3$. However, the same procedure also holds for $\mu = 1, \cdots, n$ and $i = 1, \cdots, n-1$, $n \geq 3$. For other related results, see Ref. 1 and also Refs. 13–15.

3. CONCLUSION

One advantage of the proof just completed is that it avoids the use of coordinate transformations. The entire proof is carried out in a single coordinate system. The result is an increase in our fund of information on the mathematical and physical significance of the nonmetrical existence theorem.

A second advantage of this new proof is that no use is made of Riquier's existence theorem. Only the Cauchy–Kowalewsky existence theorem is used. It is hoped that this will make the proof more accessible than previously. In addition the present proof serves as an instructive and nontrivial example of the relationship of the Cauchy–Kowalewsky theorem to Riquier's theorem.

One tantalizing feature of the nonmetrical existence theorem is that a complete geometrical object of space–time can be chosen arbitrarily. That object is $\tilde{T}^{\mu\nu}$. Equally important is the fact that this arbitrary choice can be made without any prior knowledge of the metric.

The purpose of nonmetrical dynamics is to separate out those properties of physical theory which can be defined without reference to the metric.[16] In this connection one should keep in mind that the nonmetrical existence theorem would disappear if one were to linearize Einstein's equations. By comparison, the original existence theorem of Lichnerowicz is unaffected by linearization. Thus, the nonmetrical picture differs from the usual picture in a manner that is far more extreme than one might think at first sight.

ACKNOWLEDGMENTS

I wish to thank J. N. Goldberg, J. W. Weinberg, J. R. Stachel, J. W. York, and K. Kuchař for helpful comments.

APPENDIX A: COMMENTS ON THE CAUCHY-KOWALEWSKY EXISTENCE THEOREM

There is a surprising variation in the statements of the Cauchy–Kowalewsky existence theorem which appear in the literature. References 4 and 5 give the Cauchy–Kowalewsky theorem in a form which applies not only to a system of partial differential equations all of which are of the same order m, but also to a system of partial differential equations some of which are of order m_1, some of which are of order m_2, some of which are of order m_3, etc. Reference 6 gives a statement of the Cauchy–Kowalewsky theorem which applies only to a system of partial differential equations all of which are of the same order m. If one is presented with an unfamiliar system of partial differential equations, it is often easier to examine it using the more general Cauchy–Kowalewsky theorems of Ref. 4 or Ref. 5 rather than to convert the system to a form suitable for the Cauchy–Kowalewsky theorem of Ref. 6.

In addition, one must keep in mind that there are many systems of equations to which the Cauchy–Kowalewsky theorem does not apply. If a given system of partial differential equations possesses integrability conditions, then the Cauchy–Kowalewsky theorem does not apply to the system. It is possible, however, to replace some of the given equations by initial conditions and thus build an equivalent system to which the Cauchy–Kowalewsky theorem does apply.

There are also systems (a) which do not possess integrability conditions and (b) to which the Cauchy–Kowalewsky theorem does not apply. For example, consider the following system:

$$\frac{\partial u}{\partial x}+v=0, \quad \frac{\partial v}{\partial y}+u=0. \tag{A1}$$

Equations (A1) possess no integrability conditions. In addition, the Cauchy–Kowalewsky theorem does not apply to Eqs. (A1). Of course, Riquier's existence theorem[8–12] does apply to the system (A1). Follow the notation used in Appendix B of Reference 1 and take

$$\frac{\partial u}{\partial x}>\frac{\partial v}{\partial y}>u>v,$$

$$\frac{\partial u}{\partial x}>\frac{\partial v}{\partial x}>\frac{\partial u}{\partial y}>\frac{\partial v}{\partial y}>u>v.$$

Then the first member of the first equation of the system (A1) is $\partial u/\partial x$ and the first member of the second equation is $\partial v/\partial y$. (See Ref. 1 for a definition of the term "first member.") And the requirements for the applicability of Riquier's theorem are obviously satisfied. Furthermore, one sees by following Riquier's procedure that the system (A1) has no integrability conditions.

APPENDIX B

Lichnerowicz[7] has outlined the main features of the methods needed to show that Eqs. (2.10) are equivalent to Einstein's equations in the form (2.9). However, Lichnerowicz's presentation makes the special assumptions

$$T^{\mu\nu}=\rho v^{\mu}v^{\nu}$$

and

$$T^{\mu\nu}=(\rho+p)v^{\mu}v^{\nu}-pg^{\mu\nu}.$$

It is therefore necessary for the purposes of the present paper to demonstrate that Eqs. (2.10) are equivalent to Eqs. (2.9) in the general case of arbitrary $T^{\mu\nu}$.

Consider the following system of equations:

$$R_{ij}=-8\pi(T_{ij}-\tfrac{1}{2}g_{ij}T), \tag{B1a}$$

$$G^{4}{}_{\lambda}=-8\pi T^{4}{}_{\lambda}. \tag{B1b}$$

[Note that Eq. (B1b) has not yet been replaced by an initial condition.] One wishes to demonstrate that the system (B1) is indeed equivalent to Einstein's equations (1.1). By expanding Eq. (B1b), the system becomes

$$R_{ij}=-8\pi(T_{ij}-\tfrac{1}{2}g_{ij}T), \tag{B2a}$$

$$R^{4}{}_{k}=-8\pi T^{4}{}_{k}, \tag{B2b}$$

$$R^{4}{}_{4}-\tfrac{1}{2}\delta^{4}{}_{4}g^{\mu\nu}R_{\mu\nu}=-8\pi T^{4}{}_{4}. \tag{B2c}$$

For convenience, define

$$P_{\mu\nu}\equiv T_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}T, \tag{B3}$$

where

$$T_{\mu\nu}=(-g)^{-1/2}\tilde{T}_{\mu\nu}. \tag{B4}$$

Contraction of Eq. (B3) gives

$$P=-T, \tag{B5a}$$

$$T^{4}{}_{4}=P^{4}{}_{4}-\tfrac{1}{2}P. \tag{B5b}$$

Equations (B2) can then be written

$$R_{ij}=-8\pi P_{ij}, \tag{B6a}$$

$$g^{44}R_{k4}+g^{4i}R_{ki}=-8\pi T^{4}{}_{k}, \tag{B6b}$$

$$\begin{aligned}&g^{44}[R_{44}-\tfrac{1}{2}g_{44}(g^{44}R_{44}+2g^{4i}R_{4i}+g^{ij}R_{ij})]\\&\quad+g^{4i}[R_{4i}-\tfrac{1}{2}g_{4i}(g^{44}R_{44}+2g^{4i}R_{4i}+g^{ij}R_{ij})]\\&=-8\pi T^{4}{}_{4},\end{aligned} \tag{B6c}$$

where in Eq. (B6a) one has made use of the definition (B3). Substitute Eq. (B6a) into Eq. (B6c), and the system (B6) becomes

$$R_{ij}=-8\pi P_{ij}, \tag{B7a}$$

$$g^{44}R_{k4}-8\pi g^{4i}P_{ki}=-8\pi P^{4}{}_{k}, \tag{B7b}$$

$$\begin{aligned}&g^{44}(1-\tfrac{1}{2}g^{4\alpha}g_{4\alpha})R_{44}-g^{4i}(g^{4\alpha}g_{4\alpha}-1)R_{4i}\\&\quad-\tfrac{1}{2}g^{4\alpha}g_{4\alpha}g^{ij}(-8\pi P_{ij})=-8\pi T^{4}{}_{4}.\end{aligned} \tag{B7c}$$

Simplification of Eqs. (B7) gives

$$R_{ij}=-8\pi P_{ij}, \tag{B8a}$$

$$R_{k4}=-8\pi P_{k4}, \tag{B8b}$$

$$R_{44}=-8\pi P_{44}, \tag{B8c}$$

where we have divided through by g^{44} and thus have assumed that $g^{44}\neq 0$. Rewriting Eqs. (B8), one obtains

$$R_{\mu\nu}=-8\pi P_{\mu\nu}, \tag{B9}$$

which is equivalent to Einstein's equations (1.1). One concludes that Eqs. (B1) are equivalent to Einstein's equations. The proof has been surprisingly long.

The next step is to demonstrate that the initial condition (2.10b) combined with Eqs. (2.10c) implies Eq. (B1b). To do this, one begins by writing down the contracted Bianchi identity[17]:

$$G_{\mu}{}^{\nu}{}_{;\nu}=0. \tag{B10}$$

Combine the identity (B10) with Eq. (2.10c). Then Eq. (2.10c) may be written

$$[G^{\nu}{}_{\mu}+8\pi T^{\nu}{}_{\mu}]_{;\nu}=0. \tag{B11}$$

Equations (2.10) can therefore be written

$$R_{ij}=-8\pi[T_{ij}-\tfrac{1}{2}g_{ij}T], \tag{B12a}$$

$$[G^{4}{}_{\lambda}+8\pi T^{4}{}_{\lambda}]_{x^4=0}=0, \tag{B12b}$$

$$[G^{\nu}{}_{\mu}+8\pi T^{\nu}{}_{\mu}]_{;\nu}=0. \tag{B12c}$$

Using Eq. (B3), one can rewrite Eqs. (B12) in the form

$$R_{ij}+8\pi P_{ij}=0, \tag{B13a}$$

$$[R_{ij}+8\pi P_{ij}]_{x^4=0}=0, \tag{B13b}$$

$$[G^4{}_\lambda + 8\pi T^4{}_\lambda]_{x^4=0} = 0, \tag{B13c}$$

$$[G^\nu{}_\mu + 8\pi T^\nu{}_\mu]_{;\nu} = 0, \tag{B13d}$$

where the redundancy of Eq. (B13b) is intentional.

It has been shown that Eqs. (B1) are equivalent to Eqs. (B9). One can therefore conclude that Eqs. (B13b) and (B13c) are equivalent to Eqs. (B14b) below:

$$R_{ij} + 8\pi P_{ij} = 0, \tag{B14a}$$

$$[R_{\mu\nu} + 8\pi P_{\mu\nu}]_{x^4=0} = 0, \tag{B14b}$$

$$[G^\nu{}_\mu + 8\pi T^\nu{}_\mu]_{;\nu} = 0. \tag{B14c}$$

One may rewrite Eq. (B14b) and obtain

$$R_{ij} + 8\pi P_{ij} = 0, \tag{B15a}$$

$$[G^\nu{}_\mu + 8\pi T^\nu{}_\mu]_{x^4=0} = 0, \tag{B15b}$$

$$[G^\nu{}_\mu + 8\pi T^\nu{}_\mu]_{;\nu} = 0. \tag{B15c}$$

Equation (B15c) may be written

$$[G^4{}_\mu + 8\pi T^4{}_\mu]_{,4} + \Gamma_{\beta\nu}{}^\nu[G^\beta{}_\mu + 8\pi T^\beta{}_\mu] - \Gamma_{\mu\nu}{}^\beta[G^\nu{}_\beta + 8\pi T^\nu{}_\beta] + [G^i{}_\mu + 8\pi T^i{}_\mu]_{,i} = 0. \tag{B16}$$

Evaluate Eq. (B16) at $x^4 = 0$ and use Eq. (B15b) to obtain

$$[(G^4{}_\mu + 8\pi T^4{}_\mu)_{,4}]_{x^4=0} = 0. \tag{B17}$$

Successive differentiations of Eq. (B16) with respect to x^4 lead to the result

$$\left[\frac{\partial^n}{\partial t^n}(G^4{}_\mu + 8\pi T^4{}_\mu)\right]_{t=0} = 0 \tag{B18}$$

for all $n \geq 1$, where we have put $x^4 = t$. Finally one may expand $G_\mu{}^4 + 8\pi T_\mu{}^4$ in Taylor's series:

$$G^4{}_\mu + 8\pi T^4{}_\mu = \sum_{n=0}^{\infty}\left\{\left[\frac{\partial^n}{\partial t^n}(G^4{}_\mu + 8\pi T^4{}_\mu)\right]_{t=0}(t^n)\right\}. \tag{B19}$$

Since by Eq. (B18), each term of the sum on the left-hand side of Eq. (B19) is zero, Eq. (B19) can be written

$$G^4{}_\mu + 8\pi T^4{}_\mu = 0. \tag{B20}$$

This proves then that Eqs. (B15b) and (B15c) together imply Eq. (B20) so that Eqs. (B15) become

$$R_{ij} + 8\pi P_{ij} = 0, \tag{B21a}$$

$$G^4{}_\mu + 8\pi T^4{}_\mu = 0, \tag{B21b}$$

$$[G^\nu{}_\mu + 8\pi T^\nu{}_\mu]_{;\nu} = 0. \tag{B21c}$$

Equations (B21a) and (B21b) are equivalent to Eqs. (B1) and therefore can be rewritten in the form (B9). One obtains

$$R_{\mu\nu} = -8\pi P_{\mu\nu}, \tag{B22a}$$

$$[G^\nu{}_\mu + 8\pi T^\nu{}_\mu]_{;\nu} = 0 \tag{B22b}$$

as the final form of Eqs. (2.10). This completes the proof demonstrating that Eqs. (2.10) are equivalent to Einstein's equations.

APPENDIX C

Consider the following system of initial value equations taken from the system of Eqs. (2.21):

$$[(-g)^{1/2}G^4{}_\lambda + 8\pi g_{\lambda\mu}\tilde{T}^{4\mu}]_{x^4=0} = 0, \tag{C1a}$$

$$[g_{4k,4} - V_k]_{x^4=0} = 0, \tag{C1b}$$

$$[g_{44,4} - V_4]_{x^4=0} = 0. \tag{C1c}$$

Convert Eqs. (C1a) to the notation of Ref. 18. Then Eqs. (C1) take the form

$$N^{(nj)}{}_{;j} - N^{-1}N_{;j}[N^{(nj)} - \gamma^{nj}\gamma_{im}N^{(im)} - e^{nj}] - \bar{R}^n{}_j N^j - e^{nj}{}_{;j} = -8\pi S^n, \tag{C2a}$$

$$(Q^i{}_i)^2 - Q^j{}_i Q^i{}_j - (N)^2\bar{R} = 16\pi(N)^2 g_{4\mu}\tilde{T}^{4\mu}(-g)^{-1/2}, \tag{C2b}$$

$$N_{k,4} - V_k = 0, \tag{C2c}$$

$$g_{44,4} - V_4 = 0, \tag{C2d}$$

where all quantities and equations are evaluated at $x^4 = 0$, where

$$2Q_{ij} = \gamma_{ij,4} - N_{j;i} - N_{i;j}, \tag{C3}$$

$$S^i = \gamma^{ki}(N)^2(-g)^{-1/2}g_{k\mu}\tilde{T}^{4\mu} \tag{C4}$$

with $i = 1, 2, 3$ $\mu = 1, \cdots, 4$, and

$$\gamma_{ij} = g_{ij}, \tag{C5a}$$

$$\gamma^{ij}\gamma_{jk} = \delta_k{}^i, \tag{C5b}$$

$$\gamma^{ij} = g^{ij} - g^{i4}g^{j4}(g^{44})^{-1}, \tag{C5c}$$

$$N_i = g_{4i}, \tag{C5d}$$

$$N = (-g^{44})^{-1/2}, \tag{C5e}$$

where Eq. (C2a) is equivalent to Eq. (5) of Ref. 18, where $\bar{R}_{ij}$ refers to the Ricci tensor formed from the spatial metric γ_{ij}, where

$$2e_{nj} = \gamma_{nj,4} - \gamma_{nj}\gamma^{im}\gamma_{im,4}, \tag{C6}$$

and where

$$\gamma = |\gamma_{nj}|. \tag{C7}$$

For the purposes of the present paper, it is unnecessary to determine the complete range of arbitrariness in the solution of the initial value problem (C2). Instead, existence will be proven for a particular class of solutions. Consider those solutions of Eqs. (C2) such that $N^i = 0$. [Note that since Eqs. (C2) are required to hold only on an $x^4 = 0$ hypersurface, there is no contradiction in the fact that the value we choose for the N^i is quite independent of the value given for $N_{i,4}$ in Eq. (C2c).] With this special assumption, Eqs. (C2) become:

	First members	
$e^{nj}{}_{;j} - N^{-1}N_{;j}e^{nj} = 8\pi S^n,$	$e^{n3}{}_{,3}$	(C8a)
$(Q^i{}_i)^2 - Q_i{}^j Q^i{}_j - (N)^2\bar{R} = 16\pi(N)^2(-g)^{-1/2}g_{4\mu}\tilde{T}^{4\mu},$	$\gamma_{22,33}$	(C8b)
$N_{k,4} = V_k,$	$N_{k,4}$	(C8c)
$g_{44,4} = V_4,$	$g_{44,4}$	(C8d)
$N_k = 0.$	N_k	(C8e)

If one solves each of the equations in Eq. (C8) respectively for the derivatives indicated in the First members column, one obtains a system of equations which is in the form required for the application of the Cauchy–Kowalewsky existence theorem. Applying the theorem, one concludes that solutions exist to the initial value system (C1) from which Eqs. (C8) were obtained. Note that the proof holds regardless of what analytic choice

has been made for the tensor density $\tilde{T}^{\mu\nu}(x)$ on the initial value surface $x^4 = 0$.

APPENDIX D

In explanation of the argument following Eq. (2.13) one should note the following:

$$R_{\alpha\mu\nu\beta} = \tfrac{1}{2}[g_{\mu\nu,\alpha\beta} + g_{\alpha\beta,\mu\nu} - g_{\alpha\nu,\mu\beta} - g_{\mu\beta,\alpha\nu} + (\cdots)], \tag{D1}$$

where the parentheses contain no derivatives of the metric except first derivatives. If a given second derivative of the metric has three equal indices, then it is cancelled by other similar terms. For example, derivatives of the form $g_{\mu 4,44}$ do not appear in the Riemann tensor. A similar result holds for the Ricci tensor.

*Present address: 8601 Manchester Rd., Apt. 322, Silver Spring, Maryland 20901.

[1]C.M. Pereira, J. Math. Phys. **13**, 1542 (1972), especially pp. 1542, 1545–47.
[2]C.M. Pereira, "Nonmetrical Dynamics II. The Einstein–Maxwell Equations" (to be published).
[3]See Appendix A.
[4]E. Goursat, *A Course in Mathematical Analysis, Vol. II, Part 2, Differential Equations* (Dover, New York, 1959, first published in 1917 as a translation from the original French work), pp. 283–287.
[5]A.R. Forsythe, *Theory of Differential Equations, Vol. V* (Dover, New York, 1959, originally published in 1906), pp. 43–54.
[6]R. Courant and D. Hilbert, *Methods of Mathematical Physics, Vol. II* (Interscience, New York, 1962), pp. 39–54.
[7]A. Lichnerowicz, *Théories relativistes de la gravitation et de l'électromagnetisme* (Masson, Paris, 1955), pp. 29–43.
[8]Appendix B of Ref. 1 contains a statement of Riquier's existence theorem.
[9]J.F. Ritt, *Differential Algebra*, Amer. Math. Soc. Colloq. Publ. Vol. 33 (Dover, New York, 1966), Chap. VIII. Also see J.F. Ritt, *Differential Equations from the Algebraic Standpoint*, AMS Colloq. Publ. Vol. 14 (Amer. Math. Soc., Providence, R.I., 1932), Chap. IX.
[10]J.A. Schouten, *Ricci-Calculus* (Springer-Verlag, Berlin, 1954), 2nd ed., p. 88.
[11]J.A. Schouten and W.v.d. Kulk, *Pfaff's Problem* (Oxford U.P., New York, 1949; reprinted by Chelsea, New York, 1969), Chap. X, p. 506. Also contains additional references.
[12]C.M. Pereira, J. Math. Phys. **15**, 269 (1974).
[13]J.R. Stachel, Gen. Rel. Grav. **3**, 257 (1972).
[14]J.R. Stachel, Phys. Rev. **180**, 1256 (1969).
[15]C.M. Pereira, "Two Theorems on Energy-Momentum Conservation," (to be published).
[16]See Ref. 1 for a further discussion of the nonmetrical viewpoint.
[17]In Ref. 7, Lichnerowicz uses the symbol $S_{\mu\nu}$ to refer to the Einstein tensor, whereas, in the present paper, the Einstein tensor is denoted by $G_{\mu\nu}$.
[18]C.M. Pereira, J. Math. Phys. **14**, 1498 (1973).

Quantum theory of anharmonic oscillators. I. Energy levels of oscillators with positive quartic anharmonicity*

F. T. Hioe and E. W. Montroll

Institute for Fundamental Studies, Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627
(Received 1 April 1975)

This is an investigation of the energy levels of an anharmonic oscillator characterized by the potential $(1/2)x^2+\lambda x^4$. Two regions of λ and n are distinguishable (n being the quantum number of the energy level) one in which the harmonic oscillator levels $E_n = n+1/2$ are only slightly distorted and the other in which the purely quartic oscillator form $E_n \simeq c\lambda^{1/3}(n+1/2)^{4/3}$ (c being a constant) is only slightly distorted. Rapidly converging algorithms have been developed, using the Bargmann representation, from which energy levels in any (λ, n) (with $\lambda > 0$) regime can easily be computed. Simple formulas are also derived which give excellent approximations to the energy levels in various (λ, n) regimes.

I. INTRODUCTION

An early quantum mechanical model to which the Rayleigh–Schrödinger perturbation theory was applied was the anharmonic oscillator characterized by the Hamiltonian

$$H \equiv H(\omega, \lambda) = \tfrac{1}{2}(p^2 + x^2\omega^2) + \lambda x^4. \tag{I. 1}$$

In the Schrödinger representation the associated energy levels and wavefunctions are solutions of

$$H\psi = E\psi \text{ with } H = -\tfrac{1}{2}\, d^2/dx^2 + \tfrac{1}{2}x^2\omega^2 + \lambda x^4. \tag{I. 2}$$

We have chosen units in which $\hbar = m = 1$. The formal perturbation theory of this eigenvalue problem yields the power series expansion

$$E_0(\lambda) = \tfrac{1}{2}\omega + \sum_{n=1}^{\infty} \omega A_n (\lambda/\omega^3)^n \tag{I. 3}$$

for the ground state energy. The first few A_n are known to be

$$\begin{aligned} &A_1 = 3/4, \quad A_2 = -21/8, \quad A_3 = 333/16, \\ &A_4 = -30,885/128, \quad A_5 = 916731/256, \text{ etc.} \end{aligned} \tag{I. 4}$$

70 more A_n's have recently been calculated by Bender and Wu.[1] The rapid increase in the $|A_n|$ with n suggests that the series (I. 3) does not converge; Bender and Wu have proven that (I. 3) diverges for any $\lambda > 0$.

Intuition gained from perturbation theory in classical mechanics warns of difficulties in series such as (I. 3). The classical solutions of Newton's equations for model (I. 1) (with $\lambda > 0$) are periodic elliptic functions. However, simple perturbation theory generated from the unperturbed trigonometirc solutions of the harmonic oscillator ($\lambda = 0$) equation yields secular terms which are products of powers of t with the trigonometric functions. When the perturbation series is terminated at finite order the secular terms cause calculated displacements to become arbitrarily large at sufficiently long times thus strongly violating the energy conservation principle. Lindstedt and Poincaré, through a frequency renormalization of the zero order trigonometric terms (i. e., introduction of a new frequency $\Omega = \omega + \lambda\omega_1 + \cdots$) have eliminated secular terms. It would not be surprising if a quantum mechanical perturbation series calculated from a basis set expressed in terms of the original frequency ω did not lead to a convergent series expansion. A second indicator of impending trouble is evident from expressing H in a momentum representation with

$$p \sim p \text{ and } x \sim id/dp,$$

$$(\tfrac{1}{2}p^2 - \tfrac{1}{2}\omega^2 d^2/dp^2 + \lambda d^4/dp^4)\psi = E\psi. \tag{I. 5}$$

In this differential equation the "small" parameter λ appears as the coefficient of the highest derivative. This is a situation analogous to one encountered in the Navier–Stokes equation of hydrodynamics in which the viscosity η, which in certain regimes is considered to be small, is the coefficient of the highest order derivative in the equation. One of the obstacles in the development of a theory of turbulence is associated with the identification of the turbulent regime with small η, and the observation that a perturbation theory in which the $\eta = 0$ solution (that of an equation of lower order) is used as an unperturbed starting function is misleading and generally useless. Hence, we should be reconciled to the possibility that a standard perturbation solution of (I. 1) might yield nonconvergent series.

Bazley and Fox[2] have, by variational methods, derived excellent upper and lower bounds for various energy levels. When $0 < \lambda \leq \frac{1}{2}$ the gap between these bounds may be only a few parts per thousand. However, with increasing $\lambda > \frac{1}{2}$ it widens considerably. Reid[3] has further developed the variational calculations using a method of Löwdin. He also calculated energy levels of high quantum number states. Loeffel, Martin, Simon and Wightman[4,5] made a detailed investigation of the Padé method taking it to 20th order and obtaining excellent numerical results in the range $0 < \lambda \leq \frac{1}{2}$. They have proven the convergence of the method for the problem at hand. Graffi, Grecchi and Simon[6] have proven uniqueness theorems for energy levels which were derived by applying Borel summability methods to the Rayleigh–Schrödinger series.

The most extensive numerical results on anharmonic oscillator energy levels (which extend into the range $0 \leq \lambda \leq 50$) have been obtained by Biswas, Datta, Saxena, Srivastava, and Varma[7] who postulated wavefunctions to be of the form

$$\psi = (\exp -\tfrac{1}{2}x^2) \sum_{n=0}^{\infty} c_n x^{2n}. \tag{I. 6}$$

When they substituted this expression into (I. 2) they obtained a three term difference equation for the $\{c_n\}$ and the energy E. In order to assure the existence of solutions they had to set the infinite determinant of the coefficients equal to zero. The energy levels were then found numerically from the resulting "Hill" determinant. The numerical method used by the above authors was to truncate the determinants, calculate the eigenvalues at different levels of truncation, and search for the limits of successive estimates as the truncated determinants were increased in size. Determinants of orders as high as 100×100 were used for large values of λ. As determinants become very large, machine round off errors can become severe.

Several other interesting papers have recently appeared on the quantum mechanical anharmonic oscillator. They are listed in Refs. 8—13. However, no one seems to have produced simple formulas which give good approximations to the energy levels in either the small or large λ regimes. One of the aims of this paper is to produce such formulas. A scaling argument due to Symanzik will be very useful for the derivation of formulas which are valid in the large λ range, say $\lambda > 2$. Since the argument is very simple, we sketch it here.

Let us write the Hamiltonian of Eq. (I. 2) as $H(\omega, \lambda)$. Then if we let $x = \lambda^{-1/6} y$, we see that

$$H(\omega, \lambda) = \lambda^{1/3} H(\omega\lambda^{-1/3}, 1), \tag{I. 7a}$$

so that as $\lambda \to \infty$

$$H(\omega, \lambda) \sim \lambda^{1/3} H(0, 1). \tag{I. 7b}$$

Since $H(0, 1)$ is independent of λ, we would expect the energy levels of $H(\omega, \lambda)$ to have the asymptotic form

$$E_n(\omega, \lambda) \sim C_n \lambda^{1/3} \quad \text{as } \lambda \to \infty \tag{I. 8}$$

where C_n depends on n and ω. We will show for example that the ground state energy can be represented to better than 1% accuracy when $\lambda > 0.3$ and to better than one part in 10^6 accuracy when $\lambda > 100$ by

$$E_0(\omega, \lambda) \sim \lambda^{1/3}(0.667\ 986\ 259\ 18 + 0.143\ 67\lambda^{-2/3} - 0.0088\lambda^{-4/3} + \ldots). \tag{I. 9}$$

Schiff[14] by a numerical calculation and Schwartz[15] by a variational method which he called a new Tamm—Dancoff method, discussed the Hamiltonian $H(0, \frac{1}{4})$ and obtained a ground state energy which is equivalent to 0.667 986 26 for that of $H(0, 1)$. For small λ, we will also present a simple iteration scheme by which the energies can be obtained to an excellent accuracy.

The second quantization form of the Hamiltonian (I. 1) is obtained by introducing the creation and annihilation operators $a^\dagger$ and a through

$$a^\dagger = (x\omega^{1/2} - ip\omega^{-1/2})2^{-1/2}, \tag{I. 10a}$$

$$a = (x\omega^{1/2} + ip\omega^{-1/2})2^{-1/2}, \tag{I. 10b}$$

which yield

$$H = \omega(a^\dagger a + \tfrac{1}{2}) + \tfrac{1}{4}(\lambda\omega^{-2})(a^\dagger + a)^4. \tag{I. 11}$$

In our consideration of this Hamiltonian we will employ the Bargmann[16,17] representation and follow some of the ideas which we introduced in a previous paper on the interaction of a photon with a two level atom.[18] In the Bargmann representation $a^\dagger$ and a are related to a complex variable z by

$$a^\dagger \equiv z \quad \text{and} \quad a \equiv d/dz. \tag{I. 12}$$

Then the eigenfunctions of $a^\dagger a$ have the remarkably simple form, z^n, as is clear from

$$(zd/dz)z^n = nz^n, \tag{I. 13a}$$

so that the eigenfunctions and eigenvalues of $(a^\dagger a + \frac{1}{2})\omega$ in this representation are obtained from

$$\omega[(zd/dz) + \tfrac{1}{2}]z^n = (n + \tfrac{1}{2})\omega z^n. \tag{I. 13b}$$

Bargmann[17] has constructed a kernel function which transforms the function $f(z)$ into a corresponding function in configurational space.

The operator equation $H\psi = E\psi$ with H given by (I. 1) has the Bargmann representation

$$\left[\omega\left(\frac{1}{2} + z\frac{d}{dz}\right) + \frac{1}{4}(\lambda/\omega^2)\left(z + \frac{d}{dz}\right)^4\right]\psi = E\psi. \tag{I. 14}$$

As with the momentum representation (I. 5), λ appears as a coefficient of the highest derivative in the equation. In Sec. II we will seek solutions of this equation of the form

$$E_n(\lambda) = (\tfrac{1}{2} + n)\omega + A_n(\lambda), \tag{I. 15}$$

$$\psi_n(z) = f_n(\lambda, z) \equiv z^n B_n(\lambda, z) \tag{I. 16}$$

with

$$B_n(\lambda, z) \equiv \sum_{k=-n}^{\infty} [U_k^{(n)}(\lambda) + \delta_{k,0}]z^k \quad \text{and} \quad U_0(\lambda) \equiv 0. \tag{I. 17}$$

In order to develop series such as (I. 9) we rewrite the Hamiltonian $H(\omega\lambda^{-1/3}, 1)$ as

$$H(\omega\lambda^{-1/3}, 1) \equiv \tfrac{1}{2}(p^2 + \omega^2 x^2/\lambda^{2/3}) + x^4$$

$$\equiv \tfrac{1}{2}(p^2 + \omega^2 x^2) - \epsilon\omega^2 x^2 + x^4, \tag{I. 18}$$

$$\epsilon \equiv \tfrac{1}{2}(1 - \lambda^{-2/3}) \quad \text{and} \quad 0 \leq \epsilon \leq \tfrac{1}{2}, \quad \text{if } \lambda \geq 0. \tag{I. 19}$$

We put this in second quantized form by setting

$$x = (a^\dagger + a)/(2\omega)^{1/2} \quad \text{and} \quad p = i(a^\dagger - a)(\omega/2)^{1/2}. \tag{I. 20}$$

Since

$$H(\omega\lambda^{-1/3}, 1) \equiv \omega(a^\dagger a + \tfrac{1}{2}) - \tfrac{1}{2}\epsilon\omega(a^\dagger + a)^2 + (a^\dagger + a)^4/4\omega^2, \tag{I. 21}$$

in the Bargmann representation, we have the eigenvalue equation

$$\left[\omega\left(\frac{1}{2} + z\frac{d}{dz}\right) - \frac{\epsilon}{2}\omega\left(z + \frac{d}{dz}\right)^2 + \left(z + \frac{d}{dz}\right)^4/4\omega^2\right] = E\psi. \tag{I. 22}$$

The parameter ϵ, which might be considered as a perturbation parameter [being restricted to the interval $(0, \frac{1}{2})$] is not a coefficient of the highest order derivative in our equation. Equation (I. 22) will be discussed in Sec. III where formulas such as (I. 9) will be derived.

The detailed program developed in this paper is a combination of analysis and computer evaluation of certain determinants which appear in solving the basic difference equations which are required to solve (I. 14) and (I. 22) using (I. 15)—(I. 17). We present some new re-

presentative tables (see Table II) of energy levels accurate to 7 or more significant figures. They sample the entire range $0 \leq \lambda \leq \infty$ and $0 \leq n \leq \infty$. Our main thrust, however, will be to derive various analytical formulas which when combined, give excellent values of $E_n(\lambda)$ over all positive λ and n. The single formula which seems to have the greatest range of validity is the large n, large λ formula

$$\lambda^{-1/3}E_n(\lambda) = C\{(n+\tfrac{1}{2}) + \delta(n+\tfrac{1}{2})^{-1}\}^{4/3} + a\lambda^{-2/3}\{(n+\tfrac{1}{2}) + \delta(n+\tfrac{1}{2})^{-1}\}^{2/3} + b\lambda^{-4/3} + \cdots,$$

where

$$C = 3^{4/3}\pi^2\{\Gamma(\tfrac{1}{4})\}^{-8/3} = 1.376\ 507\ 40,$$

$$a = 4 \cdot 3^{2/3}\pi^3\{\Gamma(\tfrac{1}{4})\}^{-16/3} = 0.268\ 055\ 493,$$

$$b = 6\pi^4\{\Gamma(\tfrac{1}{4})\}^{-8} - (1/32) = -0.011\ 674\ 983,$$

$$\delta = 0.026\ 50.$$

An important feature of our general numerical algorithims is that each energy level is calculated independently of the others so that small errors in the ground state calculations do not propagate with amplification into the excited state energy calculations.

In the remainder of this paper, the various formulas can be converted into those appropriate for the differential equation

$$\left(\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + (E' - \tfrac{1}{2}m\omega^2x^2 - \lambda'x^4)\right)\psi = 0 \tag{I.23}$$

by noting that if we set

$$E = E'/\hbar\omega \quad \text{and} \quad y = x(m\omega/\hbar)^{1/2}, \tag{I.24}$$

then

$$\left(\frac{1}{2}\frac{d^2}{dy^2} + (E - \tfrac{1}{2}y^2 - \lambda y^4)\right)\psi = 0 \tag{I.25}$$

where

$$\lambda = \lambda'\hbar/m^2\omega^3. \tag{I.26}$$

All our analysis will be based on (I. 25) or its equivalent Bargmann representation. Hence all formulas which relate E and λ can be transformed to relations between E' and λ' by using (I. 24) and (I. 26).

II. ON THE SMALL λ REGIME

In the small λ regime, we start with Eq. (I. 14). After substituting (I. 15)—(I. 17) into (I. 14) and equating coefficients of like powers of z^{k+n}, we obtain [with $U_k \equiv U_k^{(n)}(\lambda)$] when n is even

$$aU_{k-4} + b_{k-2}U_{k-2} + \gamma_k U_k + d_{k+2}U_{k+2} + e_{k+4}U_{k+4} = -a\delta_{k,4} - b_{k-2}\delta_{k,2} + \delta_{k,0}[A(\lambda) - c_0] - d_{k+2}\delta_{k,-2} - e_{k+4}\delta_{k,-4}, \tag{II.1}$$

$$k = -n, -n+1, \ldots, 0, 1, 2, \cdots,$$

where

$$a = \tfrac{1}{4}\lambda, \quad c_k = \tfrac{3}{4}\lambda[1 + 2(n+k)(n+k+1)], \tag{II.2}$$

$$b_k = \lambda(\tfrac{3}{2} + k + n), \quad \gamma_k = k + c_k - A(\lambda), \tag{II.3}$$

$$d_k = \lambda(n+k)(n+k-1)(n+k-\tfrac{1}{2}), \tag{II.4}$$

$$e_k = \tfrac{1}{4}\lambda(n+k)(n+k-1)(n+k-2)(n+k-3). \tag{II.5}$$

All these coefficients except γ_k are proportional to λ as $\lambda \to 0$. The set of Eqs. (II. 1) has a simple matrix representation. It should be noted that only U_k with even subscripts appear in the equations. For example, if $n = 4$ the only U_k which appear are U_{-4}, U_{-2}, U_0, $U_2, \cdots$. One can arbitrarily set $U_{-3} = U_{-1} = U_1 = \cdots = 0$ since they don't appear in any formulas. If $n = 3$, the only U_k which appear are U_{-2}, U_0, $U_2, \cdots$. Generally, for even n the lower index of Eq. (I. 17) starts with $-n$ and for odd n the lower index of Eq. (I. 17) starts with $-(n-1)$. The matrix form of (II. 1) is

$$\begin{pmatrix}
\gamma_{-n} & d_{-n+2} & e_{-n+4} & & & & & & \\
b_{-n} & \cdot & \cdot & \cdot & & & & & \\
a & \cdot & & \cdot & \cdot & \cdot & & & \\
 & \cdot & \cdot & & & & & & \\
 & & \cdot & \cdot & \cdot & \cdot & \cdot & & \\
 & & & a & b_{-4} & \gamma_{-2} & d_0 & e_2 & \\
 & & & & a & b_{-2} & \gamma_0 & d_2 & e_4 \\
 & & & & & a & b_0 & \gamma_2 & d_4 \\
 & & & & & & \cdot & \cdot & \cdot \\
 & & & & & & & \cdot & \cdot
\end{pmatrix}
\begin{pmatrix} U_{-n} \\ \cdot \\ \cdot \\ \cdot \\ U_{-2} \\ U_0 \\ U_2 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}
=
\begin{pmatrix} 0 \\ \vdots \\ 0 \\ -e_0 \\ -d_0 \\ A(\lambda) - c_0 \\ -b_0 \\ -a \\ 0 \\ \vdots \end{pmatrix}. \tag{II.6}$$

All elements in the matrix which do not lie on the main diagonal, or the two off diagonals immediately above or on the two immediately below the main diagonal vanish.

Let us first consider the ground state with $n = 0$. An analysis of other n values follows in a similar way. The basic matrix equation in this case is

$$\begin{pmatrix} \gamma_0 & d_2 & e_4 & 0 & 0 & 0 & \cdots \\ b_0 & \gamma_2 & d_4 & e_6 & 0 & 0 & \cdots \\ a & b_2 & \gamma_4 & d_6 & e_8 & 0 & \cdots \\ 0 & a & b_4 & \gamma_6 & d_8 & e_{10} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \end{pmatrix} \begin{pmatrix} U_0 \\ U_2 \\ U_4 \\ U_6 \\ \cdot \\ \cdot \end{pmatrix} = \begin{pmatrix} A(\lambda) - c_0 \\ -b_0 \\ -a \\ 0 \\ 0 \\ \cdot \end{pmatrix}, \tag{II. 7}$$

where now

$$a = \tfrac{1}{4}\lambda, \quad b_k = \tfrac{1}{2}\lambda(3+2k), \quad d_k = \lambda k(k-1)(k-\tfrac{1}{2}),$$

$$c_k = \tfrac{3}{4}\lambda[1+2k(k+1)], \quad e_k = \tfrac{1}{4}\lambda k(k-1)(k-2)(k-3),$$

$$\gamma_k = k + c_k - A(\lambda). \tag{II. 8}$$

By constructing the inverse of the matrix on the left of Eq. (II. 7) (which we call G), we can solve for $U_0, U_2, \cdots$. Let

$$X = \begin{vmatrix} \gamma_2 & d_4 & e_6 & 0 & \cdots \\ b_2 & \gamma_4 & d_6 & e_8 & \cdots \\ a & b_4 & \gamma_6 & d_8 & \cdots \\ 0 & a & b_6 & \gamma_8 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \end{vmatrix}, \quad Z = \begin{vmatrix} \gamma_2 & d_4 & e_6 & 0 & 0 & \cdots \\ d_2 & e_4 & 0 & 0 & 0 & \cdots \\ a & b_4 & \gamma_6 & d_8 & e_{10} & \cdots \\ 0 & a & b_6 & \gamma_8 & d_{10} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \end{vmatrix}, \tag{II. 9a}$$

$$Y = \begin{vmatrix} d_2 & e_4 & 0 & 0 & 0 & 0 & \cdots \\ b_2 & \gamma_4 & d_6 & e_8 & 0 & 0 & \cdots \\ a & b_4 & \gamma_6 & d_8 & e_{10} & 0 & \cdots \\ 0 & a & b_6 & \gamma_8 & d_{10} & e_{12} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \end{vmatrix}. \tag{II. 9b}$$

Then

$$U_0(\lambda) = (\det G)^{-1}[(A(\lambda) - c_0)\det X + b_0 \det Y + a \det Z]. \tag{II. 10}$$

Since we postulated $U_0(\lambda) \equiv 0$, we find that

$$A(\lambda) = \frac{3}{4}\lambda - \frac{3\lambda}{2}\frac{\det Y}{\det X} - \frac{\lambda}{4}\frac{\det Z}{\det X}. \tag{II. 11}$$

It is to be remembered that each of the determinants X, Y, and Z are functions of $A(\lambda)$ through the dependence of the diagonal elements $\{\gamma_k\}$ of each on $A(\lambda)$.

We have used the following numerical scheme to solve Eq. (II. 11) for the ground state energy level shift $A(\lambda)$ due to the anharmonic term in (I. 2). Each of the determinants X, Y, Z is truncated to some order, say into a 2×2 determinant. A trial value of $A(\lambda)$ is substituted into those determinants and a corrected $A(\lambda)$ is calculated as the left-hand side of (II. 11). The corrected $A(\lambda)$ is then substituted into the determinants X, Y, and Z and a second correction to $A(\lambda)$ is obtained. The process is repeated until $A(\lambda)$ is obtained to the required number of significant figures. The convergence of this process has been experienced to be very rapid. The determinants X, Y, Z are then truncated at a higher order say into 3×3 determinants. The value of $A(\lambda)$ obtained from the 2×2 truncation is used as a first approximation to be inserted in the 3×3 determinants which with (II. 11) are used to find a next approximation to $A(\lambda)$. This approximation is used to obtain a better one until the $A(\lambda)$ from the 3×3 truncation is obtained for the required number of significant figures. The result of the 3×3 truncation for $A(\lambda)$ is used as the starting point for the iteration processes using a 4×4 truncation, etc. The $A(\lambda)$ obtained from successive levels of truncation are compared and the calculation terminated when the successive $A(\lambda)$ agree with each other to the required accuracy.

Table I shows how rapidly these iterations of $A(\lambda)$ converge for some small values of λ used by Loeffel *et al.* The table was produced in less than 10 seconds of computer time. Notice that the results given by using only the 2×2 determinants are already remarkably close to the exact values.

With the aid of the Schweinsian expansion of the ratio of determinants[19] we can apply the above ideas to the derivation of analytical expressions for $A(\lambda)$. A convenient notation in terms of which the Schweinsian expansion can be expressed is

$$|a_1 b_2 c_3 d_4| \equiv \begin{vmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{vmatrix}. \tag{II. 12}$$

In this notation the expansion of the ratio of two determinants which differ only in the first column is

$$\begin{aligned} &|h_1 b_2 c_3 d_4 \cdots| / |a_1 b_2 c_3 d_4 \cdots| \\ &= h_1/a_1 + |h_1 a_2|\, b_1/a_1\, |a_1 b_2| \\ &\quad + |h_1 a_2 b_3|\, |b_1 c_2| / |a_1 b_2|\, |a_1 b_2 c_3| \\ &\quad + |h_1 a_2 b_3 c_4|\, |b_1 c_2 d_3| / |a_1 b_2 c_3|\, |a_1 b_2 c_3 d_4| + \cdots. \end{aligned} \tag{II. 13}$$

The expansion (II. 13) is of course still applicable after interchanging rows and columns in X, Y, and Z.

Application of (II. 13) to terms in (II. 11) yields

$$\det Y/\det X = (d_2/\gamma_2) + \frac{b_2}{\gamma_2}\left\{ \begin{vmatrix} d_2 & \gamma_2 \\ e_4 & d_4 \end{vmatrix} \Big/ \begin{vmatrix} \gamma_2 & b_2 \\ d_4 & \gamma_4 \end{vmatrix} \right\}$$

$$+ \begin{vmatrix} b_2 & a \\ \gamma_4 & b_4 \end{vmatrix} \begin{vmatrix} d_2 & \gamma_2 & b_2 \\ e_4 & d_4 & \gamma_4 \\ 0 & e_6 & d_6 \end{vmatrix} \Big/ \begin{vmatrix} \gamma_2 & b_2 & a \\ d_4 & \gamma_4 & b_4 \\ e_6 & d_6 & \gamma_6 \end{vmatrix} \begin{vmatrix} \gamma_2 & b_2 \\ d_4 & \gamma_4 \end{vmatrix} + \cdots, \tag{II. 14}$$

TABLE I. The convergence of ground state energy shifts $A_0(\lambda)$.

Size of determinant	$\lambda = 0.05$	$\lambda = 0.1$	$\lambda = 0.5$
2×2	0.032 687 83	0.059 385 59	0.197 535 77
4×4	0.032 642 85	0.059 146 65	0.196 685 53
6×6	0.032 642 76	0.059 146 40	0.196 187 47
8×8	0.032 642 75	0.059 146 33	0.196 177 74
10×10	0.032 642 75	0.059 146 33	0.196 176 22
12×12	0.032 642 75	0.059 146 33	0.196 175 82
14×14	0.032 642 75	0.059 146 33	0.196 175 82

$$\det Z/\det X=(e_4/\gamma_4)+\frac{d_4}{\gamma_4}\left\{\begin{vmatrix} e_4 & \gamma_4 \\ d_2 & b_2 \end{vmatrix}\Big/\begin{vmatrix} \gamma_4 & d_4 \\ b_2 & \gamma_2 \end{vmatrix}\right\}+\cdots. \tag{II. 15}$$

The first term in the ratios is, as $\lambda\to 0$ of order λ, the next of order λ^2, etc. From the definition (II. 8) and especially

$$\gamma_2=2+(39\lambda/4)-A(\lambda),$$
$$\gamma_4=4+(123\lambda/4)-A(\lambda), \tag{II. 16}$$

we find that

$$A(\lambda)=\tfrac{3}{4}\lambda-\tfrac{3}{2}\lambda^2[3\gamma_2^{-1}+\gamma_4^{-1}]$$
$$-\frac{63\lambda^3}{2}\frac{[(21\lambda/\gamma_2)+(7\lambda/\gamma_4)-2]}{(\gamma_2\gamma_4-147\lambda^2)}+\cdots. \tag{II. 17}$$

When λ is very small, we can choose as a zeroth approximation $A(\lambda)\sim\frac{3}{4}\lambda$. The next approximation is obtained by substituting this form into γ_2 and γ_4 and retaining only the first and second term in (II. 17). Then

$$A(\lambda)=\tfrac{3}{4}\lambda-\tfrac{3}{2}\lambda^2\left(\frac{3}{2+9\lambda}+\frac{1}{4+30\lambda}\right)+\cdots. \tag{II. 18}$$

Notice that the radius of convergence of $(2+9\lambda)^{-1}=\frac{1}{2}[(1-(9\lambda/2)+\cdots]$ is $|\lambda|<2/9$ while that of $(4+30\lambda)^{-1}=\frac{1}{4}[1-(15\lambda/2)+\cdots]$ is $|\lambda|<2/15$. The term $1/\gamma_6$ still has a smaller radius of convergence. Hence, if expansions such (II. 17) and (II. 18) are to be useful for even small λ, one must refrain from expanding denominators containing λ. This observation is consistent with the result of Simon, that Padé approximant expressions for $A(\lambda)$ converge while the basic perturbation series (I. 3) diverges. A second order approximation can be found for $A(\lambda)$ by using the first order approximation for γ_2 and γ_4 in the term of $O(\lambda^2)$ in (II. 17), and the zeroth order γ_2 and γ_4 (with $A(\lambda)\approx 3\lambda/4$) in the term of $O(\lambda^3)$ in (II. 17). Then we find

$$A(\lambda)=\tfrac{3}{4}\lambda-\tfrac{3}{2}\lambda^2\left\{3/\left[2+9\lambda+\frac{3}{2}\lambda^2\left(\frac{3}{2+9\lambda}+\frac{1}{4+30\lambda}\right)\right]\right.$$
$$\left.+1/\left[4+30\lambda+\frac{3}{2}\lambda^2\left(\frac{3}{2+9\lambda}+\frac{1}{4+30\lambda}\right)\right]\right\}$$
$$-\frac{63\lambda^3}{2}\left(\frac{3\lambda-4}{2+9\lambda}+\frac{7\lambda}{4+30\lambda}\right)/(8+96\lambda+123\lambda^2)+O(\lambda^4). \tag{II. 19}$$

For an arbitrary energy level n, we have

$$A(\lambda)=c_0-\left(-e_0\frac{\det U}{\det X}+d_0\frac{\det V}{\det X}+b_0\frac{\det Y}{\det X}-a\frac{\det Z}{\det X}\right) \tag{II. 20}$$

where U, V, X, Y, Z are matrices obtained from G by striking out the rows and columns of the elements a_0, b_0, γ_0, d_0, and e_0, respectively, G being the matrix on the left-hand side of Eq. (II. 6). The elements a, b, γ, d, and e are defined in (II. 2)—(II. 5) and are, of course, dependent on n. A number of values of $A_n(\lambda)$ for $0\leq\lambda\leq 1$ have been computed for n ranging from 0 to 10. To obtain an accuracy of 8 decimal places for $\lambda=0.05$, for example, the size of determinants required ranges from 7×7 for $n=0$ to about 20×20 for $n=10$, and for $\lambda=1$, it ranges from about 20×20 for $n=0$ to about 40×40 for $n=10$.

Our numerical data suggest, however, that for small values of λ, the values of $A_n(\lambda)$ are given to rather good accuracy by determinants of much smaller sizes. Indeed, we find that the following simple iteration formulas give better than 1% accuracy for $0<\lambda<0.2$ when $n\leq 4$.

For $n=0, 1$,

$$A(\lambda)=c_0-\frac{1}{\begin{vmatrix}\gamma_2 & d_4\\ b_2 & \gamma_4\end{vmatrix}}\left\{b_0\begin{vmatrix} d_2 & e_4\\ b_2 & \gamma_4\end{vmatrix}+a\begin{vmatrix}\gamma_2 & d_4\\ d_2 & e_4\end{vmatrix}\right\}. \tag{II. 21}$$

For $n=2, 3$,

$$A(\lambda)=c_0-\frac{1}{\begin{vmatrix}\gamma_{-2} & e_2 & 0\\ a & \gamma_2 & d_4\\ 0 & b_2 & \gamma_4\end{vmatrix}}\left\{b_0\begin{vmatrix}\gamma_{-2} & e_2 & 0\\ b_{-2} & d_2 & e_4\\ 0 & b_2 & \gamma_4\end{vmatrix}+a\begin{vmatrix}\gamma_{-2} & e_2 & 0\\ a & \gamma_2 & d_4\\ b_{-2} & d_2 & e_4\end{vmatrix}+d_0\begin{vmatrix} b_{-2} & d_2 & e_4\\ a & \gamma_2 & d_4\\ 0 & b_2 & \gamma_4\end{vmatrix}\right\}. \tag{II. 22}$$

For $n\geq 4$,

$$A(\lambda)=c_0-\frac{1}{\begin{vmatrix}\gamma_{-4} & d_{-2} & 0 & 0\\ b_{-4} & \gamma_{-2} & e_2 & 0\\ 0 & a & \gamma_2 & d_4\\ 0 & 0 & b_2 & \gamma_4\end{vmatrix}}\left\{b_0\begin{vmatrix}\gamma_{-4} & d_{-2} & 0 & 0\\ b_{-4} & \gamma_{-2} & e_2 & 0\\ a & b_{-2} & d_2 & e_4\\ 0 & 0 & b_2 & \gamma_4\end{vmatrix}+a\begin{vmatrix}\gamma_{-4} & d_{-2} & 0 & 0\\ b_{-4} & \gamma_{-2} & e_2 & 0\\ 0 & a & \gamma_2 & d_4\\ a & b_{-2} & d_2 & e_4\end{vmatrix}+d_0\begin{vmatrix}\gamma_{-4} & d_2 & 0 & 0\\ a & b_{-2} & d_2 & e_4\\ 0 & a & \gamma_2 & d_4\\ 0 & 0 & b_2 & \gamma_4\end{vmatrix}+e_0\begin{vmatrix} a & b_{-2} & d_2 & e_4\\ b_{-4} & \gamma_{-2} & e_2 & 0\\ 0 & a & \gamma_2 & d_4\\ 0 & 0 & b_2 & \gamma_4\end{vmatrix}\right\}. \tag{II. 23}$$

Since (II. 21), (II. 22), and (II. 23) give cubic, quartic, and quintic equations, respectively, in $A(\lambda)$, we can, in principle, express $A_n(\lambda)$ for small values of λ in closed form expressions in terms of the roots of these cubic or quartic equations, or in terms of some elliptic integrals. We shall not write down the solutions here because in practice, given any small value of λ, $A(\lambda)$ can be readily obtained (to within 1% of the exact values) by iterating (II. 21), (II. 22), or (II. 23). In Table II, we present some examples of these values. The large λ case (say $\lambda>0.2$) will be discussed in the next section. A larger collection of these data has been obtained but we shall not present it here as those presented in Table II are quite representative of these data.

In the same manner as the approximate series expansion (II. 18) was derived for the ground state energy shift, we have, for an arbitrary energy level n, the

TABLE IIA. Comparison of the exact values of $E_n(\lambda)$ for $n=0$, 1, and 2 with those obtained by using Eqs. (II.21), (II.22), or (II.23) or those obtained by using Eq. (III.9).

	$n=0$		$n=1$		$n=2$	
λ	$E_0(\lambda)$	$\frac{\text{Eq. (II.21)}}{\text{Eq. (III.9)}}$	$E_1(\lambda)$	$\frac{\text{Eq. (II.21)}}{\text{Eq. (III.9)}}$	$E_2(\lambda)$	$\frac{\text{Eq. (II.22)}}{\text{Eq. (III.9)}}$
0.002	0.501 489 66	0.501 489 66 –	1.507 419 39	1.507 419 40 –	2.519 202 12	2.519 202 15 –
0.006	0.504 409 71	0.504 409 74 –	1.521 805 65	1.521 806 05 –	2.555 972 30	2.555 974 52 –
0.010	0.507 256 20	0.507 256 44 –	1.535 648 28	1.535 650 77 –	2.590 845 80	2.590 858 44 –
0.050	0.532 642 75	0.532 687 83 –	1.653 436 01	1.653 703 49 –	2.873 979 63	2.874 656 64 –
0.100	0.559 146 33	0.559 385 59 –	1.769 502 64	1.770 324 14 –	3.138 624 31	3.139 625 55 –
0.300	0.637 991 78	0.639 067 97 0.632 453 15	2.094 641 99	2.096 857 59 2.090 199 72	3.844 782 65	3.874 390 36 3.840 416 09
0.500	0.696 175 82	0.697 535 77 0.693 593 90	2.324 406 35	2.338 139 60 2.322 636 27	4.327 524 98	– 4.325 212 17
0.700	0.743 903 50	0.745 822 88 0.742 344 28	2.509 228 10	– 2.508 297 60	4.710 328 10	– 4.708 778 94
1.000	0.803 770 65	0.808 228 54 0.802 856 26	2.737 892 27	– 2.737 444 02	5.179 291 69	– 5.178 265 39
2	0.951 568 47	– 0.951 240 90	3.292 867 82	– 3.292 788 53	6.303 880 57	– 6.303 405 63
50	2.499 708 77	– 2.499 704 46	8.915 096 36	– 8.915 101 83	17.436 9921	– 17.436 9762
200	3.930 931 34	– 3.930 930 56	14.059 2268	– 14.059 2281	27.551 4347	– 27.551 4312
1000	6.694 220 85	– 6.694 220 79	23.972 2061	– 23.972 2062	47.017 3387	– 47.017 3384
8000	13.366 9076	– 13.366 9076	47.890 7687	– 47.890 7687	93.960 6046	– 93.960 6047
20 000	18.137 2291	– 18.137 2291	64.986 6757	– 64.986 6757	127.508 839	– 127.508 839

following approximate expansion:

$$A(\lambda)=\tfrac{3}{4}\lambda\{1+2n(n+1)\}-\lambda^2\left(\frac{(n+1)(n+3/2)^2(n+2)}{2+3\lambda(2n+3)}\right.$$
$$+\frac{(n+1)(n+2)(n+3)(n+4)}{16[4+6\lambda(2n+5)]}-\frac{n(n-\tfrac{1}{2})^2(n-1)}{2+3\lambda(2n-1)}$$
$$\left.-\frac{n(n-1)(n-2)(n-3)}{16[4+6\lambda(2n-3)]}\right)+\cdots. \qquad \text{(II. 24)}$$

We observe, as we shall again do in the following sections, that as n increases, the range of validity of this approximate expression decreases.

III. ON THE LARGE λ REGIME

In the large λ regime we start with (I.22) and proceed with a series expansion for $\psi(z)$ of the same form as that used in the last section. The coefficients $U_k \equiv U_k^{(n)}(\epsilon)$ again satisfy a recurrence formula similar to (II.1) (with $k=-n, -n+1, \ldots, 0, 1, \cdots$)

$$aU_{k-4}+b'_{k-2}U_{k-2}+\gamma'_kU_k+d'_{k+2}U_{k+2}+e_{k+4}U_{k+4} \qquad \text{(III. 1)}$$
$$=a\delta_{k,4}-b'_{k-2}\delta_{k,2}+\delta_{k,0}[A(\lambda)-c'_0]-d'_{k+2}\delta_{k,-2}-e_{k+4}\delta_{k,-4}$$

where now the parameters $a, b'_k, \ldots, e_k$ are just those defined in Eq. (II.2)–(II.5), but with $\lambda\equiv 1$ and the primed parameters related to the unprimed ones through

$$b'_k=-\tfrac{1}{2}\epsilon+b_k,\quad d'_k=-\tfrac{1}{2}\epsilon(k+n)(k+n-1)+d_k,$$
$$\gamma'_k=-\epsilon(\tfrac{1}{2}+n+k)+\gamma_k \quad \text{with } \epsilon=\tfrac{1}{2}(1-\lambda^{-2/3}). \qquad \text{(III. 2)}$$

As in Sec. II, let us first consider the case of the ground state. Then the analog of Eq. (II.11) is

$$A(\epsilon)=\tfrac{3}{4}-\tfrac{1}{2}\epsilon-\tfrac{1}{2}(3-\epsilon)(\det Y'/\det X')-\tfrac{1}{4}\det Z'/\det X' \qquad \text{(III. 3)}$$

where the matrices X', Y', and Z' are obtained from (II.9a) and (II.9b) by replacing those elements $\gamma_j, d_j, \ldots$, etc. by the primed elements $\gamma'_j, d'_j, \cdots$ (with $n=0$). The numerical work for the determination of $A(\epsilon)$ for a fixed value of ϵ follows in exactly the manner discribed in Sec. II after Eq. (II.11). The $\epsilon=\frac{1}{2}$ ($\lambda\to\infty$) results are as follows (using the scaling argument of Eq. (I.7a):

$$E_0(\lambda)\sim\lambda^{1/3}[\tfrac{1}{2}+A_0(\tfrac{1}{2})]. \qquad \text{(III. 4)}$$

The values of $A_0(\frac{1}{2})$ which have been obtained from con-

TABLE IIB. Comparison of the exact values of $E_n(\lambda)$ for $n=3$, 4, 5 with those obtained by using Eqs. (II.22) or (II.23) or those obtained by using Eq. (III.9).

	$n=3$		$n=4$		$n=5$	
λ	$E_2(\lambda)$	Eq. (II.22) Eq. (III.9)	$E_4(\lambda)$	Eq. (II.23) Eq. (III.9)	$E_5(\lambda)$	Eq. (II.23) Eq. (III.9)
0.002	3.536 744 13	3.536 744 29 –	4.559 955 56	4.559 956 06 –	5.588 750 05	5.588 751 42 –
0.006	3.606 186 33	3.606 194 63 –	4.671 800 37	4.671 824 29 –	5.752 230 87	5.752 287 77 –
0.010	3.671 094 94	3.671 137 90 –	4.774 913 12	4.775 023 88 –	5.901 026 67	5.901 257 11 –
0.050	4.176 338 91	4.177 245 52 –	5.549 297 81	5.550 191 03 –	6.984 963 10	6.988 463 30 –
0.100	4.628 882 81	4.631 457 13 –	6.220 300 90	6.240 080 74 6.213 533 84	7.899 767 23	7.988 925 29 7.893 866 50
0.300	5.796 573 63	5.979 729 13 5.793 670 48	7.911 752 73	8.555 269 81 7.910 940 94	10.166 4889	– 10.165 7642
0.500	6.578 401 95	– 6.576 931 56	9.028 778 72	– 9.028 513 74	11.648 7207	– 11.648 4741
0.700	7.193 265 28	– 7.192 309 47	9.902 610 70	– 9.902 498 08	12.803 9297	– 12.803 8175
1.000	7.942 403 99	– 7.941 790 01	10.963 5831	– 10.963 5485	14.203 1394	– 14.203 0966
2	9.727 323 19	– 9.727 053 94	13.481 2759	– 13.481 2880	17.514 1324	– 17.514 1335
50	27.192 6458	– 27.192 6380	37.938 5022	– 37.938 5037	49.516 4187	– 49.516 4171
200	43.005 2709	– 43.005 2693	60.033 9933	– 60.033 9934	78.385 6232	– 78.385 6216
1000	73.419 1140	– 73.419 1139	102.516 157	– 102.516 157	133.876 891	– 133.876 890
8000	146.745 512	– 146.745 512	204.922 711	– 204.922 711	267.628 498	– 267.628 498
20 000	199.145 124	– 199.145 124	278.100 238	– 278.100 238	363.201 843	– 363.201 843

sidering ratios of various truncations of the determinants are given in Table III.

We have calculated the $A_0(\lambda)$ in

$$E_0(\lambda)=\lambda^{1/3}[\tfrac{1}{2}+A_0(\lambda)] \tag{III.5}$$

as well as $E_0(\lambda)$ for the range $1<\lambda<\infty$. Some values are recorded in Table II. These results as well as those in a much larger collection were fitted to the asymptotic formula

$$E_0=\lambda^{1/3}(0.667\,986\,259+0.143\,67\lambda^{-2/3} - 0.0088\lambda^{-4/3}+\cdots)\ . \tag{III.6}$$

The tabulated results were themselves used to obtain the coefficients of $\lambda^{-2/3}$ and $\lambda^{-4/3}$. A sequence of $E_0(\lambda)$ was calculated for $\lambda=2, 10, 50, 100, 200, \ldots, 10^5$. From these the number α_0 defined by

$$\lim_{\lambda\to\infty}[\lambda^{-1/3}E_0(\lambda)-0.667\cdots]\lambda^{2/3}=\alpha_0 \tag{III.7}$$

was calculated to be 0.143 67. With this constant evaluated, we proceed to evaluate the coefficient of $\lambda^{-4/3}$.

For an arbitrary energy level n, the analog of Eq. (II.20) is

$$A(\epsilon)=c_0'-\left(-e_0\frac{\det U'}{\det X'}+d_0'\frac{\det V'}{\det X'}+b_0'\frac{\det Y'}{\det X'}-a\frac{\det Z'}{\det X'}\right). \tag{III.8}$$

The expression

$$E_n(\lambda)=\lambda^{1/3}(\epsilon_n+\alpha_n\lambda^{-2/3}+\beta_n\lambda^{-4/3}+\cdots) \tag{III.9}$$

was constructed in a similar manner for n in the range $n=1, 2, 3, \ldots, 10$. Equation (III.3) was used to evaluate all $A_n(\epsilon)\equiv A_n(\lambda)$ by using the U', V', X', Y', and Z' for the appropriate n using the matrix elements (III.2). The results of these calculations are collected in Table IV.

Putting the values of ϵ_n, α_n, and β_n in Eq. (III.9), the values of the energies so calculated for $0.3\leq\lambda\leq10^5$ are given in Table II, and they are, as it will be observed, remarkably close to the exact values.

Thus for small n, the small λ regime (say $0<\lambda\leq0.2$) is covered well by the simple iteration formulas (II.21)—(II.23), while the large λ regime ($0.2<\lambda\leq\infty$)

TABLE III. The convergence of $A_0(\epsilon)$ for $\epsilon=\frac{1}{2}$, i.e., $\lambda=\infty$ as the sizes of determinants increase.

Size of determinant	$A_0(\frac{1}{2})$
2×2	0.172 798 464
8×8	0.167 998 149
14×14	0.167 986 327
20×20	0.167 986 261
26×26	0.167 986 259

TABLE IIC. Comparison of the exact values of $E_n(\lambda)$ for $n=6$, 7, 8 with those obtained by using Eq. (II.23) or those obtained by using Eq. (III.9).

	$n=6$		$n=7$		$n=8$	
λ	$E_6(\lambda)$	$\frac{\text{Eq. (II.23)}}{\text{Eq. (III.9)}}$	$E_7(\lambda)$	$\frac{\text{Eq. (II.23)}}{\text{Eq. (III.9)}}$	$E_8(\lambda)$	$\frac{\text{Eq. (II.23)}}{\text{Eq. (III.9)}}$
0.002	6.623 044 60	6.623 047 83 –	7.662 759 33	7.662 766 21 –	8.707 817 30	8.707 830 68 –
0.006	6.846 948 47	6.847 064 52 –	7.955 470 29	7.955 677 56 –	9.077 353 66	9.077 680 58 –
0.010	7.048 326 88	7.048 725 00 –	8.215 837 81	8.216 411 26 –	9.402 692 31	9.403 362 81 –
0.050	8.477 397 34	8.495 811 19 8.459 245 93	10.021 9318	10.086 4747 10.005 5658	11.614 7761	11.783 0337 11.599 8776
0.100	9.657 839 99	9.927 270 75 9.652 632 32	11.487 3156	– 11.482 6361	13.378 9698	14.600 5501 13.378 2438
0.300	12.544 2587	– 12.543 6302	15.032 7713	– 15.032 2095	17.622 4482	– 17.621 9566
0.500	14.417 6692	– 14.417 4616	17.320 4242	– 17.320 2394	20.345 1931	– 20.345 0402
0.700	15.873 6836	– 15.873 5937	19.094 5183	– 19.094 4385	22.452 9996	– 22.452 9391
1.000	17.634 0492	– 17.634 0196	21.236 4362	– 21.236 4095	24.994 9457	– 24.994 9237
2	21.790 9564	– 21.790 9635	26.286 125	– 26.286 131	30.979 883	– 30.979 896
50	61.820 3488	– 61.820 3490	74.772 829	– 74.772 829	88.314 328	– 88.314 330
200	97.891 3315	– 97.891 3308	118.427 83	– 118.427 83	139.900 40	– 139.900 40
1000	167.212 258	– 167.212 257	202.311 20	– 202.311 20	239.011 58	– 239.011 58
8000	334.284 478	– 334.284 478	404.468 35	– 404.468 35	477.855 70	– 477.855 70
20 000	453.664 875	– 453.664 875	548.916 14	– 548.916 14	648.515 33	– 648.515 33

is well covered by the simple formula Eq. (III.9). We shall now discuss the large n regime in the next section.

IV. LARGE n REGIME

It is evident from Table IV that ϵ_n and α_n depend on n while as n increases β_n becomes independent of n. In Table V we show that $\epsilon_n(n+\frac{1}{2})^{-4/3}$ and $\alpha_n(n+\frac{1}{2})^{-2/3}$ approach constants as n increases. Since the WKB approximation should be valid for large n, these constants should be related to WKB results. Furthermore, since we are concerned with the large λ expansion, we apply the WKB method to the Hamiltonian $H(\omega\lambda^{-1/3}, 1)$ rather than to $H(\omega, \lambda)$.

Let us first consider the limit $\lambda \to \infty$ which yields the Hamiltonian of a purely quartic potential energy function. This case has been treated by Titchmarsh in his more rigorous adaptation of the WKB type formulas (see Ref. 20, p. 151, Eq. 7.7.4). Let $\mu_0, \mu_1, \cdots$ be the eigenvalues of

$$y'' + [\mu - q(x)]y = 0 \quad \text{with } q(x) \to \infty \text{ as } x \to \pm\infty, \qquad \text{(IV.1)}$$

TABLE IV. Values of ϵ_n, α_n, and β_n for $n=0, 1, \ldots, 10$.

n	ϵ_n	α_n	β_n
0	0.677 986 259	0.143 67	−0.0088
1	2.393 644 02	0.357 80	−0.0140
2	4.696 795 39	0.493 97	−0.0125
3	7.335 730 01	0.618 26	−0.0122
4	10.244 3085	0.730 84	−0.0116
5	13.379 3366	0.835 36	−0.0116
6	16.711 8896	0.933 73	−0.0116
7	20.220 8495	1.027 16	−0.0116
8	23.889 9937	1.116 53	−0.0116
9	27.706 3935	1.202 45	−0.0116
10	31.659 4566	1.285 40	−0.0116

TABLE V. The approach to the WKB results.

n	$\epsilon_n/(n+\frac{1}{2})^{4/3}$	$\alpha_n/(n+\frac{1}{2})^{2/3}$	β_n
0	1.683 219 90	0.228 06	−0.0088
1	1.394 027 11	0.273 05	−0.0140
2	1.384 251 37	0.268 16	−0.0125
3	1.380 445 55	0.268 20	−0.0122
4	1.378 899 55	0.268 13	−0.0116
5	1.378 111 41	0.268 10	−0.0116
6	1.377 656 94	0.268 08	−0.0116
7	1.377 371 35	0.268 08	−0.0116
8	1.377 180 30	0.268 07	−0.0116
9	1.377 046 24	0.268 07	−0.0116
10	1.376 948 59	0.268 06	−0.0116
WKB	1.376 507 40	0.268 055 493	−0.0116 749 83

$$y(\pm\infty)=y'(\pm\infty)=0.$$

Then, if x_n and x_n' are roots of $q(x)=\mu_n$,

$$\frac{1}{\pi}\int_{x_n'}^{x_n}[\mu_n-q(x)]^{1/2}dx=n+\tfrac{1}{2}+O(1/n). \tag{IV.2}$$

When $q(x)=2x^4$, $x_n=(\frac{1}{2}\mu_n)^{1/4}$, and $x_n'=-(\frac{1}{2}\mu_n)^{1/4}$ (see Ref. 20, p. 144, Eq. 7.17)

$$\tfrac{1}{2}\mu_n=\left(3^{4/3}\pi^2/[\Gamma(\tfrac{1}{4})]^{8/3}\right)(n+\tfrac{1}{2})^{4/3}. \tag{IV.3}$$

If (III. 9) is compared with (IV. 3) this implies that as $n\to\infty$ (identifying μ_n with $2E_n$)

$$C\equiv\lim_{n\to\infty}\epsilon_n(n+\tfrac{1}{2})^{-4/3}=3^{4/3}\pi^2/[\Gamma(\tfrac{1}{4})]^{8/3}=1.376\ 507\ 40. \tag{IV.4}$$

This result is consistent with the fact that in Table V, when $n=10$, the value of $\epsilon_n(n+\frac{1}{2})^{-4/3}$ has already become 1.376 948.

Now let us choose $q(x)$ in (IV. 2) to be

$$q(x)=\xi x^2+\eta x^4; \tag{IV.5a}$$

we can discuss the small λ case with the

$$\xi=1 \quad\text{and}\quad \eta=2\lambda; \tag{IV.5b}$$

we can also discuss the large λ case by choosing

$$\xi=\lambda^{-2/3} \quad\text{and}\quad \eta=2. \tag{IV.6}$$

Equation (IV. 2) becomes the elliptic integral

$$n+\tfrac{1}{2}+O(1/n)=(2\eta^{1/2}/\pi)\int_0^{x_0}[(x_0^2-x^2)(x^2+x_1^2)]^{1/2}dx,$$

or

$$n+\tfrac{1}{2}+O(1/n)=(3\pi\eta)^{-1}[(u^{3/4}+\xi u^{1/4})K(k)-2\xi u^{1/4}E(k)], \tag{IV.7}$$

where x_0^2 and x_1^2 are defined by

$$\left.\begin{matrix}x_0^2\\ x_1^2\end{matrix}\right\}=\mp\frac{\xi}{2\eta}+\left(\frac{\xi^2}{4\eta^2}+\frac{\mu_n}{\eta}\right)^{1/2}, \tag{IV.8}$$

$$u=(\xi^2+4\eta\mu_n), \tag{IV.9}$$

$$k^2=\tfrac{1}{2}[1-\xi(\xi^2+4\eta\mu_n)^{-1/2}], \tag{IV.10}$$

and $K(k)$ and $E(k)$ are the complete elliptic integrals of the first and second kinds, respectively. With the choices given by Eqs. (IV. 5) and (IV. 6), $\mu_n\equiv 2E_n$ in our case.

In the large μ regime the elliptic integrals can be expanded yielding a series in μ_n. If only the first few terms are retained one can solve for $\mu_n\equiv 2E_n$ in terms of n as is done in the Appendix. In our case, we find the form (III. 9) with the limits of α_n and β_n as $n\to\infty$ given by

$$a\equiv\lim_{n\to\infty}\alpha_n(n+\tfrac{1}{2})^{-2/3}=4.\,3^{2/3}\pi^3/[\Gamma(\tfrac{1}{4})]^{16/3}=0.268\ 055\,493, \tag{IV.11}$$

$$b\equiv\lim_{n\to\infty}\beta_n=-(1/32)+6\pi^4/[\Gamma(\tfrac{1}{4})]^8=-0.011\ 674\ 983. \tag{IV.12}$$

These numbers are very close to those in Table V for $n=10$.

A still further improvement can be made by noting that in Titchmarsh's formula (IV. 2) the combination $(n+\frac{1}{2})$ is corrected by a term of $O(1/n)$ which he has not calculated. Let us suppose that

$$n+\tfrac{1}{2}+O(1/n)=n+\tfrac{1}{2}+\delta/(n+\tfrac{1}{2})+O(1/n^2) \tag{IV.13}$$

and we use our numerical data to find the coefficient δ. If we let the constant in Eq. (IV. 4) be C, then

$$\delta=\lim_{n\to\infty}(n+\tfrac{1}{2})[(\epsilon_n/C)^{3/4}-(n+\tfrac{1}{2})]. \tag{IV.14}$$

Our numerical data for ϵ_n for $n=0, 1, \ldots, 10$ of Table IV yields the results in Table VI from which we acquire the estimate $\delta\approx 0.026\ 50$.

We thus find that for the quartic oscillator with potential energy $V(x)=\lambda x^4$ the energy levels are

$$E_n=\lambda^{1/3}(3^{4/3}\pi^2/[\Gamma(\tfrac{1}{4})]^{8/3})[(n+\tfrac{1}{2})+0.026\ 50(n+\tfrac{1}{2})^{-1}+\cdots]^{4/3} \tag{IV.15a}$$

$$=\lambda^{1/3}(1.376\ 507\ 40)[(n+\tfrac{1}{2})+0.026\ 50(n+\tfrac{1}{2})^{-1}+\cdots]^{4/3}. \tag{IV.15b}$$

This formula yields two figure accuracy when n is as small as 1, three figure when $n=2$, four when $n=3$, etc. When combined with

$$E_0=0.667\ 986\ 2592\lambda^{1/3},\quad E_1=2.393\ 644\ 02\lambda^{1/3},\quad E_2=4.696\ 795\ 39\lambda^{1/3} \tag{IV.16}$$

we have very accurate results for all energy levels of a quartic oscillator with $\lambda>0$.

The energy levels for our general oscillator with both quadratic and quartic contributions to the potential energy are given in the large n, large λ regime by

$$E_n=\lambda^{1/3}\left[C\left(n+\frac{1}{2}+\frac{\delta}{n+\frac{1}{2}}\right)^{4/3}+a(n+\tfrac{1}{2})^{2/3}\lambda^{-2/3}+b\lambda^{-4/3}+\cdots\right], \tag{IV.16a}$$

where C, a, b and δ are given by (IV. 4), (IV. 11), (IV. 12) and (IV. 14), respectively. While our asymptotic expressions were derived for large values of n (say >8) they are very accurate for n even as small as 3. When λ is small and n is large, series expansions in λ diverge and Padé approximants converge rather slowly as do expressions such as (II. 24) unless $\lambda<0.2$. However, the elliptic integral scheme described below converges rapidly.

The elliptic integral formula (IV. 7) can be used most

TABLE VI. Estimates of δ from Eq. (IV.14).

n	δ
0	0.040 711 18
1	0.021 443 93
2	0.026 352 50
3	0.026 275 77
4	0.026 387 73
5	0.026 433 28
6	0.026 459 77
7	0.026 476 49
8	0.026 487 63
9	0.026 495 36
10	0.026 501 19

effectively by noting that n being a function of E_n is equivalent to E_n being a function of n. Consider an arbitrary sequence of choices of E_n. These give rise to a succession of corresponding values of $n+\frac{1}{2}$ which generally would not be half integers. If one wished for a given λ to determine with some accuracy the mth energy level (with m "large"), one would choose the two values of E_{n_1} and E_{n_2} which yielded values of $n_1 < m < n_2$. Since in our model E_m lies between E_{n_1} and E_{n_2}, a first estimate of m could be determined. The process could be repeated until $m+\frac{1}{2}$ was sufficiently well approximated so that one would have a value of E_m to the accuracy required. Since, this procedure is a numerical one and does not involve an expansion in any parameter, Eq. (IV. 7) can be used with either the choice (IV. 5b) or (IV. 6). If (IV. 5b) is used, the numbers E_n represent the energy levels, while if (IV. 6) is used the energy levels are $\lambda^{1/3}$ multiplied by the appropriate E_n.

For a fixed n the large λ and small λ regimes can be identified by examining the combination $(\xi^2+8\eta E_n)$ which appears in various terms in Eq. (IV. 7) as some function

$$f(\xi^2+8\eta E_n).$$

Since one can write

$$\xi^2+8\eta E_n=\begin{cases}\xi^2[1+8\eta E_n/\xi^2]\\ 8\eta E_n[1+\xi^2/8\eta E_n],\end{cases} \tag{IV. 17}$$

the manner in which various series expansions are to be made depend on whether

$$E_n < \xi^2/8\eta \quad \text{or} \quad E_n > \xi^2/8\eta. \tag{IV. 18}$$

If we identify $\xi=1$ and $\eta=2\lambda$, then the regime of small λ corresponds to

$$E_n < 1/16\lambda, \quad \text{and large } \lambda \text{ to } E_n > 1/16\lambda. \tag{IV. 19}$$

If $\lambda=0.01$, and in the small λ regime $E_n \approx (n+\frac{1}{2})$, the energy level n to which the small λ regime would end would be that for which

$$n+\tfrac{1}{2} < 100/16 \approx 6.25. \tag{IV. 20}$$

would be violated, i.e., when n is about 6. The larger the quantum number n of an excited state the greater the root mean square displacement of the oscillator from equilibrium and the more important the anharmonic term λx^4 becomes in determining the motion. Hence it is natural to expect that with increasing λ the critical number n at which the large λ regime becomes appropriate decreases. These statements are bourne out in Table II where various analytical approximations to the energy levels are compared with exact results.

Thus, to summarize, we have the following simple formulas or iteration schemes for all regimes of (λ, n):

	Small n	Large n
Small λ	Eqs. (II. 21)—(II. 23)	Eq. (IV. 7)
Large λ	Eq. (III. 9)	Eq. (IV. 16a)

ACKNOWLEDGMENT

We wish to thank Robert Helleman for a number of interesting discussions on anharmonic oscillators.

APPENDIX

Consider

$$H=\tfrac{1}{2}(p^2+\xi x^2)+\eta x^4.$$

If we choose

$$\xi=\lambda^{-2/3}, \quad \eta=2,$$

the elliptic integral iteration formula is

$$n+\frac{1}{2}+O\left(\frac{1}{n}\right)=\left(\frac{(\xi^2+4\eta\mu_n)^{3/4}}{3\pi\eta}+\frac{\xi(\xi^2+4\eta\mu_n)^{1/4}}{3\pi\eta}\right)K(k^2) - \frac{2\xi(\xi^2+4\eta\mu_n)^{1/4}}{3\pi\eta}E(k^2) \tag{A1}$$

where $\mu_n=2E_n$, E_n being the nth energy level of H, and

$$k^2=\frac{1}{2}\left(1-\frac{\xi}{(\xi^2+4\eta\mu_n)^{1/2}}\right)\equiv c. \tag{A2}$$

Denoting $K(\frac{1}{2})$ by K_0, and $E(\frac{1}{2})$ by E_0, to the order ξ^2, we have

$$K(k^2)=K_0-\frac{\xi(2E_0-K_0)}{(\xi^2+4\eta\mu_n)^{1/2}}+\frac{\xi^2K_0}{8(\xi^2+4\eta\mu_n)}-\cdots, \tag{A3}$$

where we have used the formulas

$$\frac{\partial K(c)}{\partial c}=[E(c)-(1-c)K(c)]/2c(1-c), \tag{A4}$$

$$\frac{\partial E(c)}{\partial c}=[E(c)-K(c)]/2c. \tag{A5}$$

Similarly,

$$E(k^2)=E_0-\frac{\xi(E_0-K_0)}{2(\xi^2+4\eta\mu_n)^{1/2}}+\cdots. \tag{A6}$$

Substituting (A3) and (A6) into (A1) and expanding in powers of ξ, then retaining terms up to ξ^2 only, we get

$$n+\frac{1}{2}+O\left(\frac{1}{n}\right)=\frac{2^{3/2}K_0}{3\pi\eta^{1/4}}\mu_n^{3/4}-\frac{\xi(2E_0-K_0)^{1/4}}{2^{1/2}\pi\eta^{3/4}}\mu_n^{1/4} + \frac{\xi^2 2^{1/2}K_0}{16\pi\eta^{5/4}}\mu_n^{-1/4}. \tag{A7}$$

Letting $x\equiv\mu_n^{1/4}$, (A7) is a quartic equation in x. Solving for x, letting $\mu_n=2E_n$ in the end and noting that

$$K_0=\tfrac{1}{4}\pi^{-1/2}[\Gamma(\tfrac{1}{4})]^2 \tag{A8}$$

and

$$2E_0-K_0=2\pi^{3/2}[\Gamma(\tfrac{1}{4})]^{-2}, \tag{A9}$$

we find

$$E_n=C\left[(n+\tfrac{1}{2})+O\left(\frac{1}{n}\right)\right]^{4/3}+a\lambda^{-2/3}\left[(n+\tfrac{1}{2})+O\left(\frac{1}{n}\right)\right]^{2/3} + b\lambda^{-4/3}+\cdots \tag{A10}$$

where

$$C=3^{4/3}\pi^2[\Gamma(\tfrac{1}{4})]^{-8/3}=1.376\ 507\ 40\cdots, \tag{A11}$$

$$a=4\cdot 3^{2/3}\pi^3[\Gamma(\tfrac{1}{4})]^{-16/3}=0.268\ 055\ 493\cdots, \tag{A12}$$

$$b=6\pi^4[\Gamma(\tfrac{1}{4})]^{-8}-(1/32)=-0.011\ 674\ 983\cdots. \tag{A13}$$

WKB elliptic integral methods have also been used by Mathews and Eswaran[12] for the quartic anharmonic

oscillator and by Lakshmanan and Prabhakaran[13] for the sextic one. In the quartic case Mathews and Eswaran have explicitly derived the equivalent of our large and small λ limits but they have not obtained series expansions such as (A10).

*This work was partially supported by a National Science Foundation Grant #DID71-04010A02.

[1]C.M. Bender and T.T. Wu, Phys. Rev. **184**, 1231 (1969); Phys. Rev. D **7**, 1620 (1973).
[2]N. Bazley and D. Fox, Phys. Rev. **124**, 483 (1961).
[3]C. Reid, J. Chem. Phys. **43S**, 186 (1965).
[4]J. Loeffel, A. Martin, B. Simon, and A. Wightman, Phys. Lett. **30B**, 656 (1969).
[5]B. Simon, Ann. Phys. **58**, 76 (1970).
[6]S. Graffi, V. Grecchi, and B. Simon, Phys. Lett. **32B**, 631 (1970).
[7]S. Biswas, K. Datta, R. Saxena, P. Srivastava, and V. Varma, J. Math. Phys. **14**, 1190 (1973).
[8]S.N. Biswas, Proc. Indian Acad. Sci. A **78**, 36 (1973).
[9]F.R. Halpern, J. Math. Phys. **14**, 219 (1973).
[10]J.H. Henkel and C.A. Uzes, Phys. Rev. D **8**, 4430 (1973).
[11]F.R. Halpern and T.W. Yonkman, J. Math. Phys. **15**, 1718 (1974).
[12]P.M. Mathews and K. Eswaran, Lett. Nuovo Cimento **5**, 15 (1972).
[13]M. Lakshmanan and J. Prabhakaran, Lett. Nuovo Cimento **7**, 689 (1973); M. Lakshmanan, Lett. Nuovo Cimento **8**, 743 (1973).
[14]L.I. Schiff, Phys. Rev. **92**, 766 (1953).
[15]C. Schwartz, Ann. Phys. **32**, 277 (1965).
[16]V. Bargmann, Rev. Mod. Phys. **34**, 829 (1962).
[17]V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961); **20**, 1 (1967).
[18]F.T. Hioe and E.W. Montroll, J. Math. Phys. **16**, 1259 (1975).
[19]A.C. Aitken, *Determinants and Matrices* (Oliver and Boyd, Edinburgh, 1944).
[20]E.C. Titchmarsh, *Eigenfunction Expansions*, Part I (Oxford U.P., Oxford, 1961), 2nd ed.

Moment inequalities for ferromagnetic Gibbs distributions*

Charles M. Newman

Department of Mathematics, Indiana University, Bloomington, Indiana 47401
(Received 14 November 1974)

Moment inequalities analogous to Khintchine's inequality (for sums of independent Bernoulli random variables) are obtained for a certain class of random variables which naturally arises in the context of ferromagnetic Ising models and ϕ^4 Euclidean (quantum) field models in a positive external field. These results extend ones obtained previously which applied only to the mean zero (vanishing external field) case.

1. INTRODUCTION

Throughout this paper we consider finite collections of (spin) random variables $\{X_j: j=1, \ldots, N\}$ whose joint probability distribution ν on $\mathbb{R}^N$ is of the form:

$$\nu(x_1, \ldots, x_N)=\frac{1}{Z}\exp\left(\sum_{j=1}^{N}h_jx_j+\sum_{j\le k=1}^{N}J_{jk}x_jx_k\right)\prod_{j=1}^{N}\rho_j(x_j) \quad (1.1)$$

with $J_{jk}\ge 0$, $h_j\ge 0$, and each ρ_j an even probability measure such that

$$\int \exp(bx^2)\,d\rho_j(x)<\infty \quad \forall\ b\in\mathbb{R}. \quad (1.2)$$

When $\rho_j(x)=[\delta(x-1)+\delta(x+1)]/2$ for all j, (1.1) is the Gibbs distribution of a spin-$\frac{1}{2}$ Ising model with pair ferromagnetic interactions J_{jk}, in a positive external magnetic field h_j. Distributions of the form (1.1) also arise naturally as lattice approximations to Euclidean (quantum) field theories; for example, the lattice approximation to an $a\phi^4+b\phi^2-\mu\phi$ Euclidean field model yields

$$d\rho_j/dx=\exp(-a_jx^4-b_jx^2) \text{ for each } j.^1$$

We define $F(\mathbf{r})=\log E(\exp[X(\mathbf{r})])$, where $X(\mathbf{r})=\sum r_jx_j$. The Griffiths–Kelly–Sherman inequalities[2–4] state, with no further restrictions on the ρ_j's, that, for any choice of m, n, j_i, k_i,

$$E(X_{j_1}\cdots X_{j_m})\ge 0 \quad (1.3a)$$

and

$$E(X_{j_1}\cdots X_{j_m}X_{k_1}\cdots X_{k_n})\ge E(X_{j_1}\cdots X_{j_m})E(X_{k_1}\cdots X_{k_n}). \quad (1.3b)$$

In particular this shows that $\partial F/\partial r_j\ge 0$ and $\partial^2F/\partial r_j\partial r_k\ge 0$ when $\mathbf{r}\ge 0$ (i.e., when $r_i\ge 0\ \forall\ i$).

For *certain* choices of ρ_j, including both spin-$\frac{1}{2}$ models and $a\phi^4+b\phi^2-\mu\phi$ lattice approximations,[5] the Griffiths–Hurst–Sherman inequality[6] states that for any i, j, k,

$$\frac{\partial^3F}{\partial r_i\partial r_j\partial r_k}\le 0 \quad \text{when } \mathbf{r}\ge 0. \quad (1.4)$$

The following simple proposition will then apply to the random variable $X(\mathbf{r})$ when $\mathbf{r}\ge 0$.

Proposition 1: If X is any random variable such that $E(\exp(rX))\equiv\exp[F(r)]<\infty\ \forall\ r$ and such that $F^{(3)}(r)\le 0\ \forall\ r\ge 0$, then

$$E(\exp(rX))\le\exp[rE(X)+r^2\sigma^2(X)/2] \quad (1.5)$$

for $r\ge 0$, where $\sigma^2(X)=E(X^2)-[E(X)]^2$ is the variance of X.

Proof: By Taylor's theorem

$$F(r)=F(0)+F'(0)r+\tfrac{1}{2}F''(0)r^2+(1/3!)F^{(3)}(s)r^3 \quad (1.6)$$

for some $s\in[0,r]$. Now $F^{(3)}(s)\le 0$ for $s\ge 0$ so that $F(r)\le F(0)+F'(0)r+\frac{1}{2}F''(0)r^2$. Since $F(0)=0$, $F'(0)=E(X)$, and $F''(0)=\sigma^2(X)$, the proof is complete.

In Ref. 7, it was shown by means of the Lee–Yang theorem (see Sec. 3) that when $h_j=0\ \forall j$ in (1.1), then, for any $\mathbf{z}\in\mathbb{C}^N$,

$$|E(\exp[X(\mathbf{z})])|\le\exp[\tfrac{1}{2}E(X(|\mathrm{Re}\mathbf{z}|)^2)], \quad (1.7)$$

where $|\mathrm{Re}\mathbf{z}|$ denotes $(|\mathrm{Re}z_1|, \ldots, |\mathrm{Re}z_N|)\in\mathbb{R}^N$. The following theorem extends this result to nonzero h_j and gives a simpler proof even in the $h_j\equiv 0$ case.

Theorem 2: If the GHS inequality (1.4) is valid for the $\{X_j\}$ defined by (1.1) (with J_{jk}, $h_j\ge 0$), then, for any $\mathbf{z}\in\mathbb{C}^N$,

$$|E(\exp[X(\mathbf{z})])|\le\exp[E(X(|\mathrm{Re}\mathbf{z}|))+\tfrac{1}{2}\sigma^2(X(|\mathrm{Re}\mathbf{z}|))]. \quad (1.8)$$

Proof: The left-hand side of (1.8) is, of course, bounded by $E(\exp[X(\mathrm{Re}\mathbf{z})])$, which, by expanding in terms of the moments of $X(\mathrm{Re}\mathbf{z})$ and using (1.3a), may in turn be bounded by $E(\exp[X(|\mathrm{Re}\mathbf{z}|)])$. The right-hand bound of (1.8) is then obtained by applying Proposition 1 to $X=X(|\mathrm{Re}\mathbf{z}|)$.

Remark 1: A simple symmetry argument shows that Theorem 2 is still correct when $h_j\le 0\ \forall j$, providing that $E(X(|\mathrm{Re}\mathbf{z}|))$ in the right-hand side of (1.8) is replaced by its absolute value.

It was shown in Ref. 7 that for $h_j\equiv 0$, not only is inequality (1.7) correct, but in addition

$$0\le E(X(\mathbf{r})^{2m})\le[(2m)!/2^mm!][E(X(\mathbf{r})^2)]^m \quad (1.9)$$

for $\mathbf{r}\ge 0$ and any $m=2, 3, \cdots$. For a spin-$\frac{1}{2}$ model with $J_{jk}\equiv 0$, (1.9) is a statement about linear combinations of Bernoulli random variables and is just Khintchine's inequality (Ref. 8, Chap. 5). A direct consequence of (1.9) (Ref. 8, Chap. 5) is the existence (for $0<p<\infty$) of positive constants A_p and B_p depending only on p {with $B_{2m}=[(2m)!/2^mm!]^{1/2m}$} such that

$$A_p\|X(\mathbf{r})\|_2\le\|X(\mathbf{r})\|_p\le B_p\|X(\mathbf{r})\|_2, \quad (1.10)$$

where $\|X\|_p$ denotes the p-norm $[E(|X|^p)]^{1/p}$.

Now, a natural *conjecture* from (1.8), is that by analogy with (1.9), we should have for $h_j \geq 0$ that

$$0 \leq E(X(\mathbf{r})^n) \leq \sum_{k=0}^{[n/2]} \frac{n!}{2^k k!(n-2k)!} [E(X(\mathbf{r}))]^{n-2k} [\sigma^2(X(\mathbf{r}))]^k \tag{1.11}$$

for $\mathbf{r} \geq 0$ and any $n = 1, 2, 3, \cdots$, where $[n/2]$ denotes the greatest integer $\leq n/2$; a further conjecture would be that (1.11) could be strengthened to yield point correlation inequalities (such as in Ref. 7, Eq. (3.3)). Unfortunately, we have not been able to obtain either of these very strong results; instead, we do derive that (1.9) [and consequently (1.10)] is still valid in the $h_j \geq 0$ case (see Theorem 4 below). The remainder of this paper is concerned with the derivation of this and other results and an application of them to the construction of ϕ^4 Euclidean field models: in Sec. 2, we define and study random variables of type $\mathcal{L}_0$ and, in Sec. 3, we show that for general Ising models obeying the Lee–Yang theorem, $X(\mathbf{r})$ is of type $\mathcal{L}_0$ for $\mathbf{r} \geq 0$.

2. RANDOM VARIABLES OF TYPE $\mathcal{L}_0$

Definition: A random variable X will be said to be of type $\mathcal{L}_0$ if, for some real C, C',

$$|E(\exp(zX))| \leq C \exp(C'|z|^2) \quad \forall\, z \in \mathbb{C}, \tag{2.1}$$

and the zeroes of $E(\cosh(zX))$ are all pure imaginary.

We recall from Ref. 7 that a random variable X of type $\mathcal{L}$ is one satisfying (2.1) with $E(X) = 0$ and with the zeroes of $E(\exp(zX))$ pure imaginary. It was shown in Ref. 7 that such an X is necessarily even so that random variables of type $\mathcal{L}$ are exactly those of type $\mathcal{L}_0$ which are also even. The following trivial proposition allows us to extend many of the results of Ref. 7 to the class $\mathcal{L}_0$.

Proposition 3: If X is of type $\mathcal{L}_0$, and W is a Bernoulli random variable independent of X, then WX is of type $\mathcal{L}$.

Proof: This follows immediately from the fact that

$$E(\exp(zWX)) = E(\exp(zX) + \exp(-zX))/2 = E(\cosh(zX)). \tag{2.2}$$

The cumulants $u_n(X)$ of a random variable X may be obtained from the relation

$$E(\exp(zX)) = \exp\left(\sum_{n=1}^{\infty} u_n(X) \frac{z^n}{n!}\right). \tag{2.3}$$

We define modified cumulants $u_n^e(X)$ by the relation

$$E(\cosh(zX)) = \exp\left(\sum_{n=1}^{\infty} u_n^e(X) \frac{z^n}{n!}\right); \tag{2.4}$$

thus $u_{2m+1}^e = 0$, $u_2^e = E(X^2)$, $u_4^e = E(X^4) - 3[E(X^2)]^2$, etc. It is, of course, clear that for X even, $u_n^e(X) = u_n(X)$. Almost all the results of Ref. 7 for random variables of type $\mathcal{L}$ can be extended to those of type $\mathcal{L}_0$ (with certain obvious modifications); the following theorem gives only the most interesting of these.

Theorem 4: If X is of type $\mathcal{L}_0$, then $\forall\, z \in \mathbb{C}$

$$|E(\exp(zX))| \leq 2E(\cosh[(\mathrm{Re}z)X]) \leq 2\exp[|\mathrm{Re}z|^2 E(X^2)/2]. \tag{2.5}$$

In addition, for any $m = 1, 2, \ldots$,

$$0 \leq E(X^{2m}) \leq [(2m)!/2^m m!][E(X^2)]^m, \tag{2.6}$$

$$E(X^{2m-1}) \leq E(|X|^{2m-1}) \leq [(2m)!/2^m m!]^{1-1/2m} [E(X^2)]^{(2m-1)/2}, \tag{2.7}$$

and

$$0 \leq (-1)^{m-1} u_{2m}^e(X) \leq [(2m)!/2^m m][E(X^2)]^m. \tag{2.8}$$

Proof: We let $Y = WX$ as in Proposition 3 and note from (2.2) and (2.4) that $E(X^{2m}) = E(Y^{2m})$ while $u_{2m}^e(X) = u_{2m}(Y)$. Thus (2.5), (2.6), and (2.8) follow from the analogous results for Y (Ref. 7, Theorems 4 and 5). (2.7) is a special case of (1.10) and follows from (2.6) together with the fact that $\|X\|_p$ is increasing with p so that $(\|X\|_{2m-1})^{2m-1} \leq (\|X\|_{2m})^{2m-1}$.

The next theorem is a fairly simple corollary of Theorem 4. A more sophisticated version for use in constructing quantum field models is given at the end of Sec. 3. We write $X_m \xrightarrow{w} X$ when the probability distribution of X_m converges weakly to that of X or equivalently when $E(\exp(iyX_m)) \to E(\exp(iyX))$ for all real y.

Theorem 5: If $\{X_m\}$ are random variables of type $\mathcal{L}_0$ such that $E(X_m^2) \leq A$ independent of m and if $X_m \xrightarrow{w} X$, then $E(\exp(zX_m)) \to E(\exp(zX))$ uniformly on compact subsets of $\mathbb{C}$ and X is of type $\mathcal{L}_0$.

Proof: The bound on $E(X_m^2)$ implies by (2.5) that $|E(\exp(zX_m))| \leq 2\exp(A|z|^2/2)$ independent of m. Since $E(\exp(zX_m))$ was already assumed to converge for pure imaginary z, Vitali's theorem implies uniform convergence on compacts for both $E(\exp(zX_m))$ and $E(\cosh(zX_m))$ (Ref. 9, Sec. 15.3) and also the estimate (2.1) for X. By Hurwitz' theorem (Ref. 9, p. 205), the zeroes of $E(\cosh(zX))$ must be pure imaginary, which completes the proof.

3. INEQUALITIES FOR ISING MODELS AND ϕ^4 FIELDS

In this section we consider random variables $\{X_j\}_{j=1}^N$ whose joint probability distribution is given by (1.1) with $J_{j_k} \geq 0$, $h_j \geq 0$, and each ρ_j an even probability measure satisfying (1.2) *and* such that

$$\int \exp(zx) d\rho_j(x) = 0 \Rightarrow z = i\alpha \quad \text{for some real } \alpha. \tag{3.1}$$

A general version of the Lee–Yang theorem (Ref. 10, Theorem 1.1) implies that under these circumstances $E(\exp[zX(\mathbf{r})])$ for $\mathbf{r} \geq 0$ has all its zeroes in the closed left half-plane and that the zeroes are pure imaginary when $h_j \equiv 0$.

It is possible to obtain a factorization for $E(\exp[zX(\mathbf{r})])$ (when $h_j \geq 0$) analogous to that of (Ref. 7, Prop. 2) and to then show directly that $X(\mathbf{r})$ is of type $\mathcal{L}_0$. That technique, however, is quite lengthy, and there is a much simpler method due to Griffiths[2,6] of introducing a "ghost" spin which has the effect of replacing the original $h_j \geq 0$ Ising model with a new $h_j \equiv 0$ model. The following proposition is a version of Griffiths' method.

Proposition 6: Suppose $X_1, \ldots, X_N$ have the joint

distribution ν given by (1.1) (with each ρ_j even) and W is a Bernoulli random variable independent of $\{X_j\}$; if we define $Y_0 = W$ and $Y_j = WX_j$ $(j=1,\ldots,N)$, then $\{Y_j: j=0,\ldots,N\}$ have the joint probability distribution $\tilde{\nu}$ on $\mathbb{R}^{N+1}$ with

$$\tilde{\nu}(y_0,\ldots,y_N) = \frac{1}{Z}\exp\left(\sum_{j\le k=0}^{N} J_{jk}y_jy_k\right)\prod_{j=0}^{N}\rho_j(y_j), \tag{3.2}$$

where $J_{00}=0$, $J_{0k}=h_k$, and $\rho_0(y)=[\delta(y-1)+\delta(y+1)]/2$. Further, $X_j = WY_j$ $(j=1,\ldots,N)$ so that for any $m=1, 2, \cdots$ and any choice of $j_1,\ldots,j_{2m}\in\{1,\ldots,N\}$,

$$E(X_{j_1}\cdots X_{j_{2m}}) = E(Y_{j_1}\cdots Y_{j_{2m}}), \tag{3.3}$$

$$E(X_{j_1}\cdots X_{j_{2m-1}}) = E(Y_0Y_{j_1}\cdots Y_{j_{2m-1}}). \tag{3.4}$$

Proof: From (1.1) and the evenness of the ρ_j's, we have

$$\begin{aligned} E\left(\exp\left(\sum_{j=0}^{N} z_jY_j\right)\right) &= \tfrac{1}{2}\sum_{a=\pm1} E(\exp[z_0a + aX(\mathbf{z})]) \\ &= \tfrac{1}{2}\sum_{a=\pm1}\frac{\exp(z_0a)}{Z}\int \exp\left(\sum_{j=1}^{N} z_j(ax_j)\right. \\ &\quad \left. + \sum_{j=1}^{N} h_ja(ax_j) + \sum_{j\le k=1}^{N}(ax_j)(ax_k)\right)\prod_{j=1}^{N}\rho_j(a(ax_j)) \\ &= \int \exp\left(\sum_{j=0}^{N} z_jy_j\right)d\tilde{\nu}(y_0,\ldots,y_N). \end{aligned}$$

This proves (3.2) while (3.3) and (3.4) follow from the fact that $WY_j = W^2X_j = X_j$ $(j=1,\ldots,N)$.

We note that since $J_{0k}=h_k$ in (3.2), we have $J_{jk}\ge 0$ $(j\le k=0,\ldots,N)$ whenever $h_k\ge 0$ $\forall\, k$ and $J_{jk}\ge 0$ $(j\le k=1,\ldots,N)$. It thus follows from (3.3) that any correlation inequality involving only even correlations which is true in ferromagnetic Ising models with $h_j\equiv 0$ is necessarily also valid with $h_j\ge 0$. We also note that with h_j, $J_{jk}\ge 0$ and with (3.1) satisfied, $\sum\lambda_jY_j$ is of type $\mathcal{L}$ for[7] $\lambda_j\ge 0$ so that by (2.2) we have the following result.

Theorem 7: If $\{X_j: j=1,\ldots,N\}$ have joint distribution (1.1) with $h_j\ge 0$, $J_{jk}\ge 0$ and each ρ_j an even probability measure satisfying (1.2) and (3.1), then for any choice of $\lambda_j\ge 0$ $(j=0,\ldots,N)$, $\lambda_0+\sum\lambda_jX_j$ is of type $\mathcal{L}_0$.

The next theorem is an extension of (Ref. 7, Theorems 8 and 9) to the case of nonzero external field; it is an immediate consequence of Theorems 4 and 7 above together with the proof of Ref. 7, Theorem 9.

Theorem 8: With the same hypotheses as in Theorem 7, it follows that for any $\mathbf{z}$, $\mathbf{z}^1,\ldots,\mathbf{z}^n\in\mathbb{C}^N$ and any $\lambda\in\mathbb{R}^N$ with $\lambda_j\ge 0$ $\forall\, j$,

$$\begin{aligned} |E(\exp[X(\mathbf{z})])| &\le 2E(\cosh[X(|\mathrm{Re}\,\mathbf{z}|)]) \\ &\le 2\exp[E(X(|\mathrm{Re}\,\mathbf{z}|)^2)/2] \end{aligned} \tag{3.5}$$

while

$$\left|E\left(\prod_{j=1}^{n} X(\mathbf{z}^j)\right)\right| \le D_n\prod_{j=1}^{n}[E(X(|\mathbf{z}^j|)^2)]^{1/2}, \tag{3.6}$$

where $D_{2m} = (2m)!/2^m m!$ and $D_{2m-1} = (D_{2m})^{1-1/2m}$; in addition,

$$(-1)^{m-1}u^e_{2m}(X(\lambda)) \ge 0. \tag{3.7}$$

Remark 2: It follows from (3.3) above that the conjectured point correlation inequalities given in Ref. 7, Eqs. (3.3) and (3.4), would immediately extend to the $h_j\ge 0$ situation providing that the Ursell functions in Ref. 7, Eq. (3.4), were replaced by modified Ursell functions defined in analogy with (2.4).

Remark 3: We include (3.5), which is weaker than (1.8), in the statement of the theorem above since it is not known whether all Ising models satisfying the hypotheses of Theorem 7 must necessarily obey the GHS inequality (1.4). The GHS inequality does not follow from the spin-$\frac{1}{2}$ approximation of Ref. 10, Theorem 2.3, because of the $\exp(in_j\pi\sigma_j/2)$ factors used there [see Ref. 10, Eq. (2.3)]. Examples of measures which satisfy (3.1) but for which "physical" $(n_j\equiv 0)$ spin-$\frac{1}{2}$ approximations are not known include $d\rho/dx = \text{const}\,\exp(-\lambda\cosh x)$ for[10] $\lambda>0$ and $d\rho/dx = \text{const}\,(1-x^2)^{(d-2)/2}$ on the the interval $[-1,1]$ for $d>0$. The latter example is, for d an integer, the one-dimensional marginal distribution of uniform surface measure on the d-sphere (in $\mathbb{R}^{d+1}$); a physical spin-$\frac{1}{2}$ approximation is known for[11] $d=2$.

Our final result (Theorem 9) is an application of Theorem 8 to the construction of $a\phi^4+b\phi^2-\mu\phi$ Euclidean field theories as limits of lattice approximations; it strengthens the original results of this type by Glimm and Jaffe[12] and is an extension of Ref. 7, Theorem 10, which only applied to *even* ϕ^4 models $(\mu=0)$. For the sake of convenience, we state the results only for ϕ^4 models rather than for general random fields satisfying the Lee–Yang theorem where a clearly analogous theorem would apply. Theorem 9 follows immediately from Theorems 2, 5, 7, 8, and the proof of Ref. 7, Theorem 10.

By a random field $\phi(f)$ (indexed by f in some topological vector space F) we mean a linear mapping from F to random variables on some probability space such that $\phi(f_\alpha)\to\phi(f)$ in probability as $f_\alpha\to f$ in F. As in Ref. 7 we consider a lattice approximation Ising model as a random field $\phi(f)$ indexed by f in the Schwartz space $\mathcal{S}(\mathbb{R}^d)$ with $\phi(f)=\sum f(w_j)X_j$ for some specified $\{w_j: j=1,\ldots,N\}\subset\mathbb{R}^d$ (see Ref. 7 for further details).

Theorem 9: Suppose $\{\phi_k\}$ is an infinite sequence of random fields which are lattice approximations to a d-dimensional $a\phi^4+b\phi^2-\mu\phi$ Euclidean field theory with $a>0$, $\mu\ge 0$. If

$$|E(\phi_k(f)^2)| \le \|f\|^2 \tag{3.8}$$

for some fixed $\mathcal{S}$-norm $\|\cdot\|$, then (3.8) also holds with $\|\cdot\|$ taken as

$$\|f\| = C\sup_{w\in\mathbb{R}^d}|(1+|w|^2)^sf(w)| \tag{3.9}$$

for some choice of C and s. In addition, there exists a subsequence $k_j\to\infty$ and a random field ϕ indexed by $f\in F$, the Banach space of continuous functions with finite $\|\cdot\|$-norm, such that

$$E(\exp[\phi(f)]) = \lim_{j\to\infty} E(\exp[\phi_{k_j}(f)]) \tag{3.10}$$

and

$$E(\phi(f_1)\cdots\phi(f_n)) = \lim_{j\to\infty} E(\phi_{k_j}(f_1)\cdots\phi_{k_j}(f_n)) \tag{3.11}$$

for any $f, f_1,\ldots,f_n\in F$. Further, $E(\exp[\phi(f)])$ is an

entire functional on F (i.e., for $f=z_1f_1+\cdots+z_mf_m$ with any m and any choice of $f_j\in F$, it is entire in $z_1,\ldots,z_m$) satisfying

$$|E(\exp[\phi(f)])|\leq\exp[E(\phi(|\mathrm{Re}f|))+\sigma^2(\phi(|\mathrm{Re}f|))/2] \quad (3.12)$$

and

$$|E(\phi(f_1)\cdots\phi(f_n))|\leq D_n\prod_{j=1}^{n}[E(\phi(|f_j|)^2)]^{1/2}$$

$$\leq D_n\prod_{j=1}^{n}\|f_j\|, \quad (3.13)$$

where D_n is as in (3.6). Finally, if $f\geq 0$, then $\phi(f)$ is a random variable of type L_0 so that Theorem 4 applies; in particular,

$$(-1)^{m-1}u_{2m}(\phi(f))\geq 0. \quad (3.14)$$

*Supported in part by NSF MPS 74-04870.

[1]F. Guerra, L. Rosen, and B. Simon, "The $P(\phi)_2$ Euclidean quantum field theory as classical statistical mechanics," Ann. Math. (to appear).
[2]R.B. Griffiths, J. Math. Phys. 8, 484 (1967).
[3]D.G. Kelly and S. Sherman, J. Math. Phys. 9, 466 (1968).
[4]J. Ginibre, Comm. Math. Phys. 16, 310 (1970).
[5]B. Simon and R.B. Griffiths, Comm. Math. Phys. 33, 145 (1973).
[6]R.B. Griffiths, C.A. Hurst, and S. Sherman, J. Math. Phys. 11, 790 (1970).
[7]C.M. Newman, "Inequalities for Ising models and field theories which obey the Lee—Yang theorem," Comm. Math. Phys. (to appear).
[8]A. Zygmund, *Trigonometrical Series* (Cambridge U.P., Cambridge, 1959), new ed., Vol. I.
[9]E. Hille, *Analytic Function Theory* (Ginn, Boston, 1962).
[10]C.M. Newman, Comm. Pure Appl. Math. 27, 143 (1974).
[11]R.B. Griffiths, J. Math. Phys. 10, 1559 (1969).
[12]J. Glimm and A. Jaffe, Phys. Rev. Lett. 33, 440 (1974).

Erratum: Double coset analysis for symmetry adapting *N*th rank tensors of *U(n)* to its unitary subgroups [J. Math. Phys. 16, 756 (1975)]

John J. Sullivan

Physics Department, University of New Orleans, New Orleans, Louisiana 70122
(Received 9 May 1975)

Page 756: Schur is not spelled correctly. The first lines of the two lemmas in Sec. II should read Lemma I: Any $n\times m$ matrix A has a Hermitian product.... Lemma II: Any square matrix Y formed as the order....

Page 757: The matrix on the right of Eq. (2a′) should be E_{1^n}. The expression following Eq. (3) should read

$$(j,n-2n_1+j)\exp(B_jI_{j,n-n_1+j})(j,n-2n_1+j).$$

Page 758: In the second line before Eq. (4) the dimension should be ${}_1n_2^0={}_1n-n_1$.

Page 759: The right-hand side of Eq. (9) is missing the bracket

$$\left[\frac{{}_1N_1!\,{}_2N_1!\,\tau_1!}{N_1!\,!\,{}_1\tau_1!\,!\,{}_2\tau_1!}\right]^{1/2}.$$

The right-hand side of Eq. (11) should have $\tau=\lambda$.

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering issues 1 through 12 is published yearly. An (E) *after the page number indicates an Erratum.*

entire functional on F (i.e., for $f=z_1f_1+\cdots+z_mf_m$ with any m and any choice of $f_j\in F$, it is entire in $z_1,\ldots,z_m$) satisfying

$$|E(\exp[\phi(f)])|\leq\exp[E(\phi(|\mathrm{Re}f|))+\sigma^2(\phi(|\mathrm{Re}f|))/2] \quad (3.12)$$

and

$$|E(\phi(f_1)\cdots\phi(f_n))|\leq D_n\prod_{j=1}^{n}[E(\phi(|f_j|)^2)]^{1/2}$$

$$\leq D_n\prod_{j=1}^{n}\|f_j\|, \quad (3.13)$$

where D_n is as in (3.6). Finally, if $f\geq 0$, then $\phi(f)$ is a random variable of type L_0 so that Theorem 4 applies; in particular,

$$(-1)^{m-1}u_{2m}(\phi(f))\geq 0. \quad (3.14)$$

*Supported in part by NSF MPS 74-04870.

[1]F. Guerra, L. Rosen, and B. Simon, "The $P(\phi)_2$ Euclidean quantum field theory as classical statistical mechanics," Ann. Math. (to appear).
[2]R.B. Griffiths, J. Math. Phys. 8, 484 (1967).
[3]D.G. Kelly and S. Sherman, J. Math. Phys. 9, 466 (1968).
[4]J. Ginibre, Comm. Math. Phys. 16, 310 (1970).
[5]B. Simon and R.B. Griffiths, Comm. Math. Phys. 33, 145 (1973).
[6]R.B. Griffiths, C.A. Hurst, and S. Sherman, J. Math. Phys. 11, 790 (1970).
[7]C.M. Newman, "Inequalities for Ising models and field theories which obey the Lee—Yang theorem," Comm. Math. Phys. (to appear).
[8]A. Zygmund, *Trigonometrical Series* (Cambridge U.P., Cambridge, 1959), new ed., Vol. I.
[9]E. Hille, *Analytic Function Theory* (Ginn, Boston, 1962).
[10]C.M. Newman, Comm. Pure Appl. Math. 27, 143 (1974).
[11]R.B. Griffiths, J. Math. Phys. 10, 1559 (1969).
[12]J. Glimm and A. Jaffe, Phys. Rev. Lett. 33, 440 (1974).

Erratum: Double coset analysis for symmetry adapting *N*th rank tensors of *U(n)* to its unitary subgroups [J. Math. Phys. 16, 756 (1975)]

John J. Sullivan

Physics Department, University of New Orleans, New Orleans, Louisiana 70122
(Received 9 May 1975)

Page 756: Schur is not spelled correctly. The first lines of the two lemmas in Sec. II should read Lemma I: Any $n\times m$ matrix A has a Hermitian product.... Lemma II: Any square matrix Y formed as the order....

Page 757: The matrix on the right of Eq. (2a′) should be E_{1^n}. The expression following Eq. (3) should read

$$(j,n-2n_1+j)\exp(B_jI_{j,n-n_1+j})(j,n-2n_1+j).$$

Page 758: In the second line before Eq. (4) the dimension should be ${}_1n_2^0={}_1n-n_1$.

Page 759: The right-hand side of Eq. (9) is missing the bracket

$$\left[\frac{{}_1N_1!\,{}_2N_1!\,\tau_1!}{N_1!\,!\,{}_1\tau_1!\,!\,{}_2\tau_1!}\right]^{1/2}.$$

The right-hand side of Eq. (11) should have $\tau=\lambda$.

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering issues 1 through 12 is published yearly. An (E) *after the page number indicates an Erratum.*